Smarten
Augmented Analytics
Powered by ElegantJ BI

Self-Serve Data Preparation (SSDP) User Manual

Version 5.1

Document Version: 2.0
Product Version: 5.1

www.Smarten.com

Document Information	
Document ID	Smarten-User-Manual-SSDP
Document Version	2.0
Product Version	5.1
Date	28-November-2020
Recipient	NA
Author	EMTPL

Statement of Confidentiality, Disclaimer and Copyright

This document contains information that is proprietary and confidential to EMTPL, which shall not be disclosed, transmitted, or duplicated, used in whole or in part for any purpose other than its intended purpose. Any use or disclosure in whole or in part of this information without the express written permission of EMTPL is prohibited.

Any other company and product names mentioned are used for identification purpose only, may be trademarks of their respective owners and are duly acknowledged.

Disclaimer

This document is intended to support administrators, technology managers or developers using and implementing Smarten. The business needs of each organization will vary and this document is expected to provide guidelines and not rules for making any decisions related to Smarten. The overall performance of Smarten depends on many factors, including but not limited to hardware configuration and network throughput.

Contents

1 About this document

This manual explains how to use various features of Self-Serve Data Preparation (SSDP) in Smarten Advanced Data Discovery Suite.

1.1 Scope and Organisation of Topic Areas

Chapter 1	Introducing ElegantJ BI - Smarten
Chapter 2	Introducing Self-Serve Data Preparation (SSDP)
Chapter 3	Creating a Data Source
Chapter 4	Data Source Management
Chapter 5	Creating a Dataset
Chapter 6	Working with Dataset
Chapter 7	Product and Support Information

1.2 Conventions used

This manual uses typographical conventions in the text to help you distinguish between the names of files, instructions, and other important notes that are relevant during installation. For example:

- Important notes are indicated in a different font colour as shown in the example below.

> Note:
> These are the front-end operations that highlight records in the front-end interface and do not remove any records from the Dataset.

- References to documents are highlighted as below:

Reference: **Self-Serve Data Preparation (SSDP) - Concept Manual > Shape Data > Add Column > Custom**

2 Introducing ElegantJ BI - Smarten

ElegantJ BI is a full-stack Business Intelligence tool that employs the “Smarten” approach to Advanced Data Discovery. The solution comprises a comprehensive set of tools, including Self-Serve Data Preparation, Smart Visualisation, and Plug n’ Play Predictive Analytics. These tools are designed to democratize advanced analytics and transform business users into citizen data scientists.

Self-Serve Data Preparation

Self-Serve Data Preparation allows business users to perform data preparation on their own without the assistance of IT staff or data analysts. Users are not restricted by complex tools or forced to wait for IT to deliver crucial data. Guided by smart suggestions and auto recommendations, business users can prepare, blend, and transform data and create analysis-ready data quickly and accurately without assistance rather than waiting for central metadata prepared by IT.

Smart Data Visualization

Intuitive Smart Data Visualization tools suggest the best options for visualizing and plotting a particular set or type of data based on the nature, dimensions, and trend of data so that business

users can easily select the appropriate method to clearly and quickly visualize data in a way that is meaningful to the task.

Plug n' Play Predictive Analysis

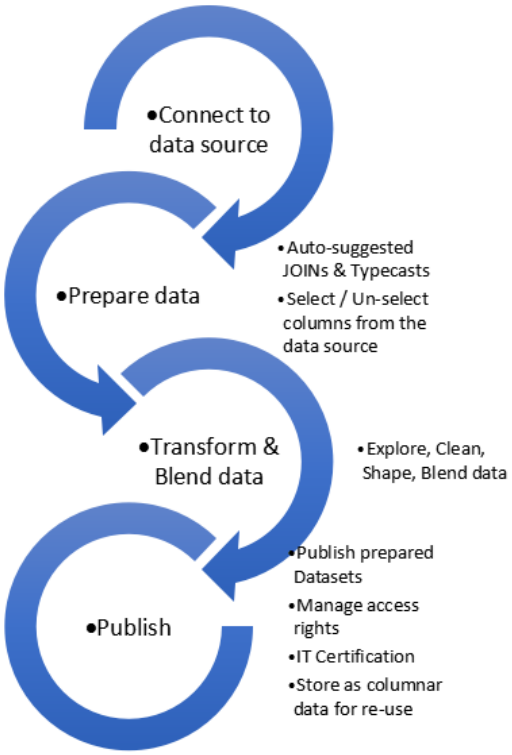
Assisted Predictive Modelling allows business users to explore predictive algorithms and models without the skill or knowledge of a data scientist or statistician. The solution considers the underlying data and use case, suggests the best-fit algorithm, and then displays output results and data visualization accompanied by an interpretation using simple human language.

3 Introducing Self-Serve Data Preparation (SSDP)

The Self-Serve Data Preparation component of the Smarten Advanced Data Discovery solution allows business users to prepare and analyze data with clear results without the assistance of technology staff or analysts.

The Smarten SSDP solution requires minimal training and gives business users the freedom to use simple, intuitive tools to perform sophisticated tasks, including data preparation utilizing machine learning, auto-detection, and auto-suggestion features. The comprehensive toolset allows users to gather, prepare, and analyze with only a basic working knowledge of Excel.

Users can process and work on raw data and convert and transform information into reusable analysis-ready data. The IT staff can certify the data quality so that all users understand the origin and veracity of the data as that data is published and shared with other users.



SELF-SERVE DATA PREPARATION—PROCESS

Here is a snapshot of the Self-Serve Data Preparation (SSDP) process:

Create Data Source profile:
The process of SSDP begins by identifying Data Source(s) and then creating a Data Source Profile. The Data Source contains the data the user wishes to extract, manipulate, and analyze. The Smarten Self-Serve Data Preparation component allows users to extract data from a variety of Data Source types, including Files, Databases, SAP®, R Script, and Google Analytics.

Create Dataset:
The user creates a Dataset by connecting to the Data Source to fetch the desired data. During the data fetching process, the user has the opportunity to select columns for the Dataset. Users can create many Datasets from one Data Source.

Smarten supports both Cache Datasets and Real-Time Dataset architecture. When using Real-time Datasets, the latest data is extracted from the Data Sources as and when required, and all data-related actions performed on the Dataset are performed in real time on the latest source data. For Cache Datasets, the data is not extracted in real time. Rather, it is cached and stored in a columnar data structure. Cache Datasets are updated periodically from the Data Sources with the help of a scheduler.

Process Data:
While creating a Dataset, the user can work with and explore data, and clean, shape, and blend data, employing a wide variety of functions. Users can also combine (JOIN) many Datasets with the help of auto-suggestion to identify possible JOINs and their relative value and strength. Business users can employ these tools with no required specialized skills or scripting or advanced knowledge. The intuitive interface is combined with machine learning capability and auto-detection and auto-suggest features to create analysis-ready data quickly, easily, and clearly.

Manage Data:
Users can manage Datasets, provide access rights and permissions to other users, and IT can certify data quality to help users identify quality Datasets.

Publish Data:
When a user publishes a Dataset, it is made available to other users to create Objects, such as reports, dashboards, visualization, and predictive models. During the publication process, the user can specify a Dataset as a Cache Dataset or a Real-time Dataset. When publishing a Cache Dataset, a Scheduler is created to allow for automatic update of the Dataset from Data Source(s) with a defined frequency.

The Self-Serve Data Preparation component of the Smarten solution will benefit the average organization with tools that are accessible to and suitable for a business user with average Excel skills and will provide clear, swift results that can be shared throughout the organization.

4 Creating a Data Source

You can create a data source from such sources as a database, file, Google Analytics, R Script, and SAP. You create data source profiles to extract the required data from these data sources.

Reference: **Concept Manual > Data Source > Creating Data Source**

4.1 Creating a Database Profile

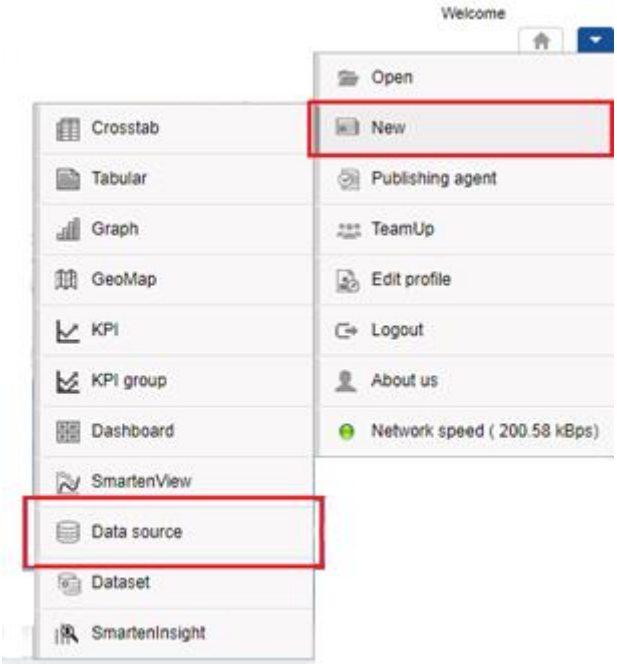
You can use a database as a data source and create a data source profile to extract the required data from the database.

About this task

Use this task to create a database data source profile.

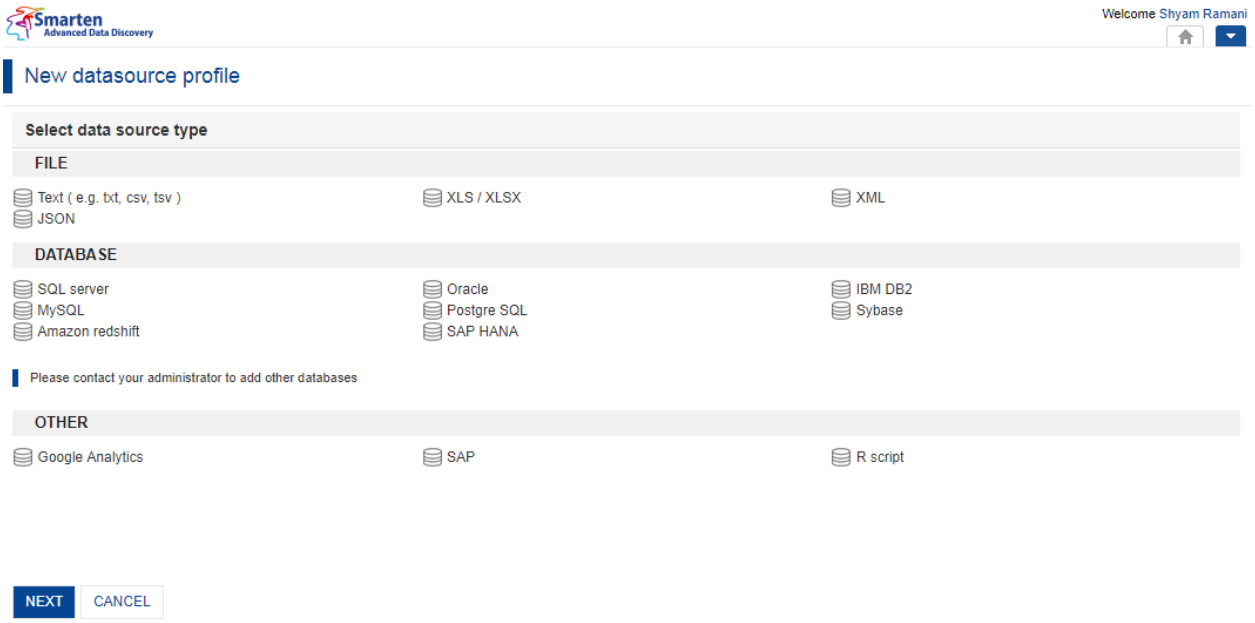
Procedure

1. Click **New** -> **Data source** from the menu.



MENU OPTION—NEW DATA SOURCE

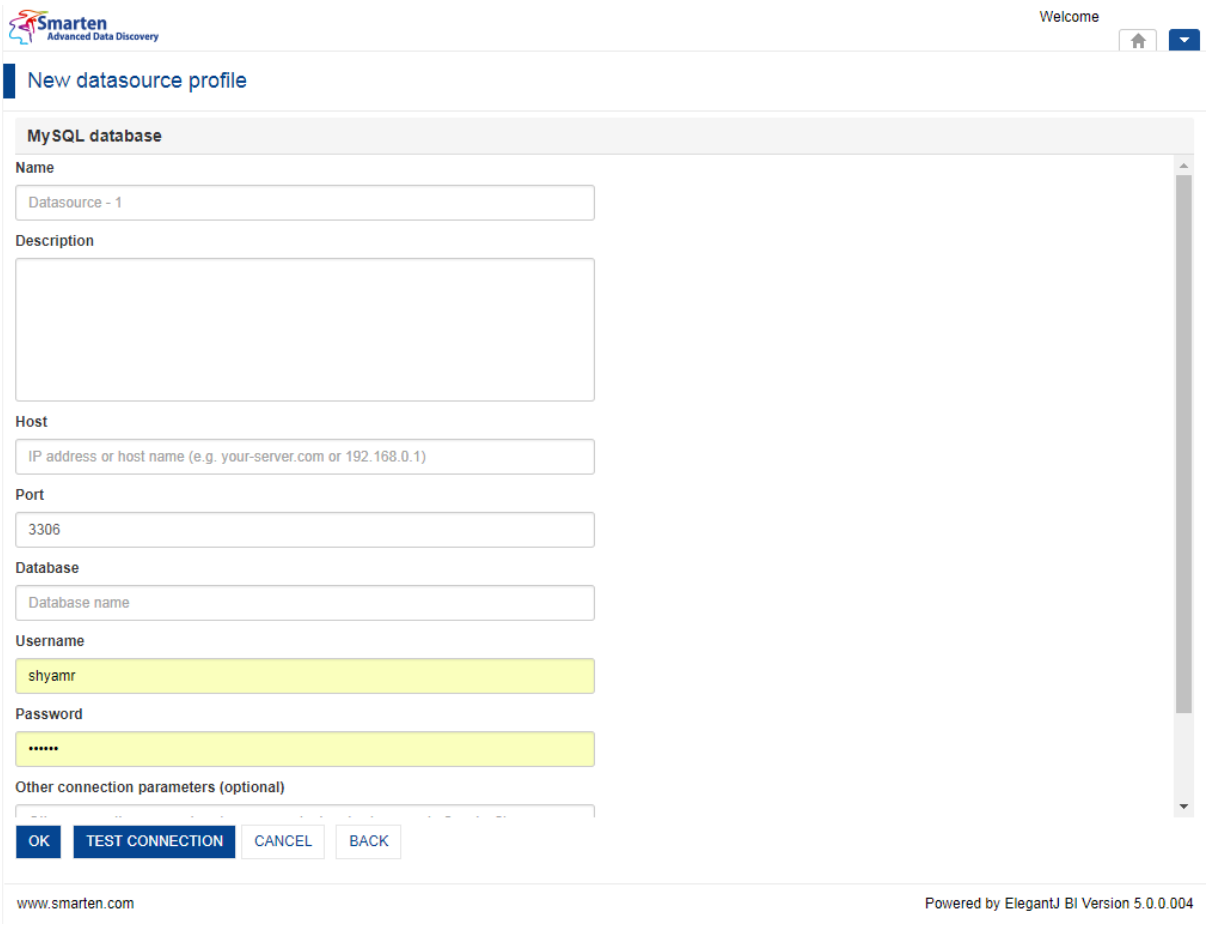
The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select the type of database that you want to use for the data source profile, and then click **NEXT**.

The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING MYSQL DATABASE PROFILE

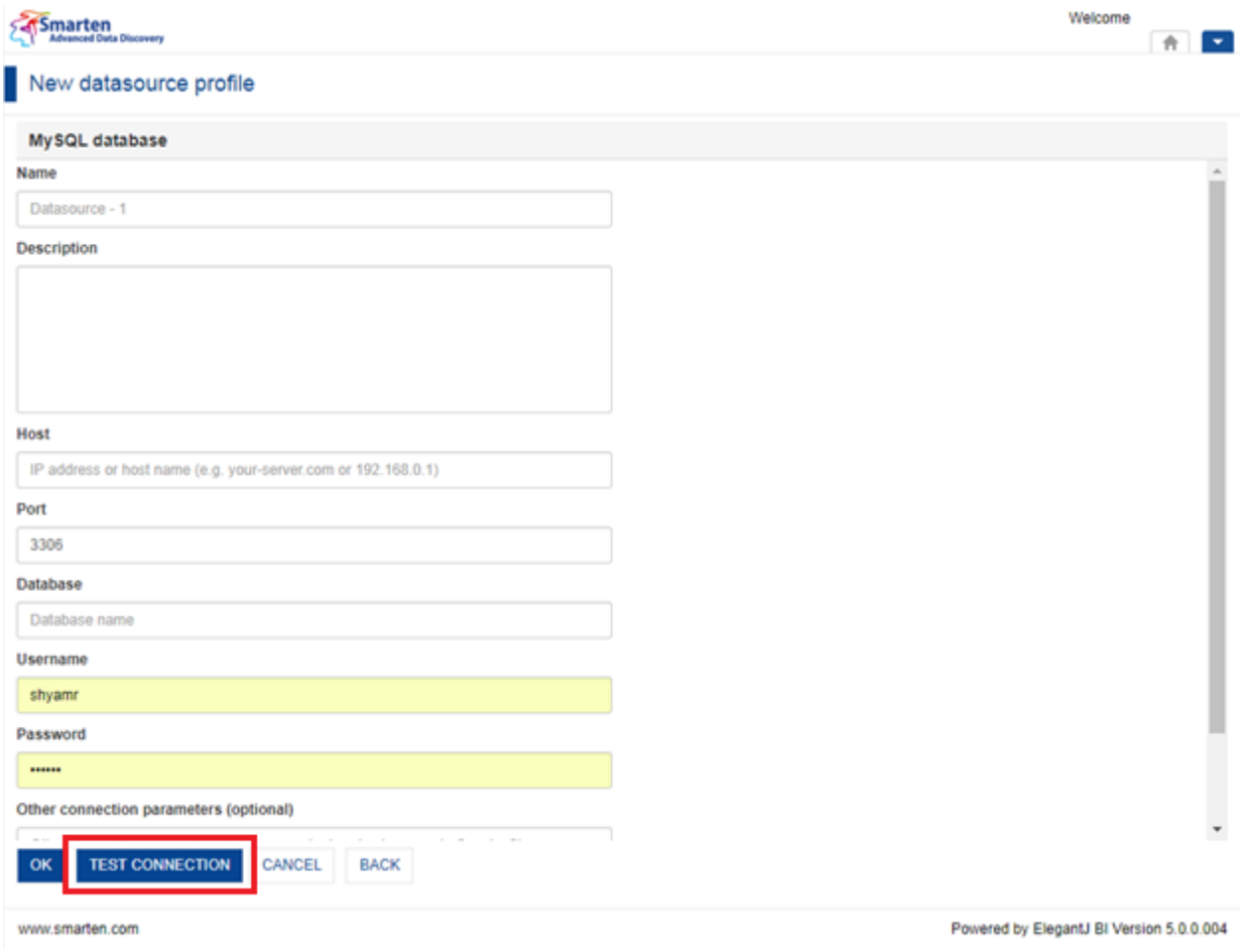
3. Provide information in the following fields:
   - **Name**: Name for the data source profile.
   - **Description**: Description for the data source profile.
   - **Host**: IP address or the hostname of the machine that contains the database that you want to use in the profile.
   - **Port**: Port number that must be used to connect to the database.
   - **Database**: Name of the database that you want to use in the profile.
   - **Username**: Username that will be used to connect to the database.
   - **Password**: Password that will be used to connect to the database.
   - **Other connection parameters**: Provide any other connection parameters that are required to connect with the database.
   - **Number of threads**: Provide a value to specify the number of threads to be established with the database.

4. Click **TEST CONNECTION** to verify that the connection to the source is successful.
5. Click **OK**.
   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

   The data source profile is now available in the repository.

4.1.1 Testing a Database Connection

Once you have provided all the values to set up a connection with a database, you can test to ensure that the connection with the database is established successfully.



Once you have provided all the information for creating a database profile, you can click the **TEST CONNECTION** option to test the connection with the database.

4.2 Creating a File Type Profile

You can use a file as a data source and create a data source profile to extract the required data from those files.

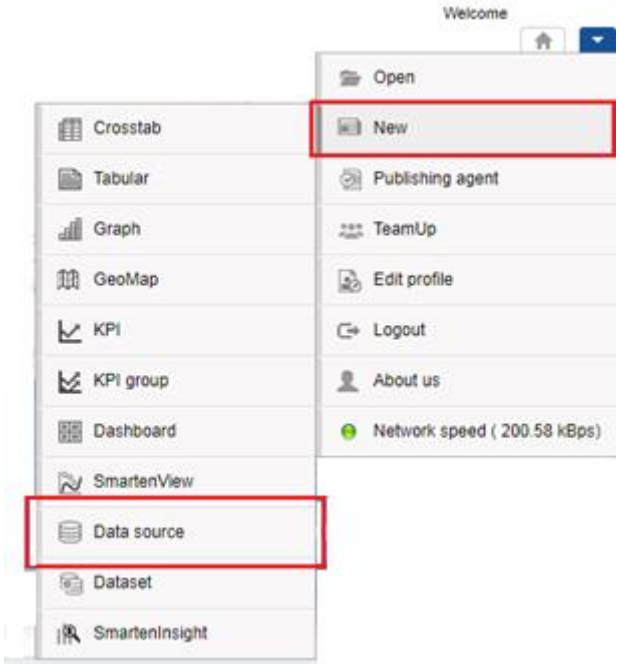
4.2.1 Creating a Text File Type Profile

About this task

Use this task to create a text data source profile. You can use .txt, .csv, .tsv, or any other file formats that contain data separated by a character.

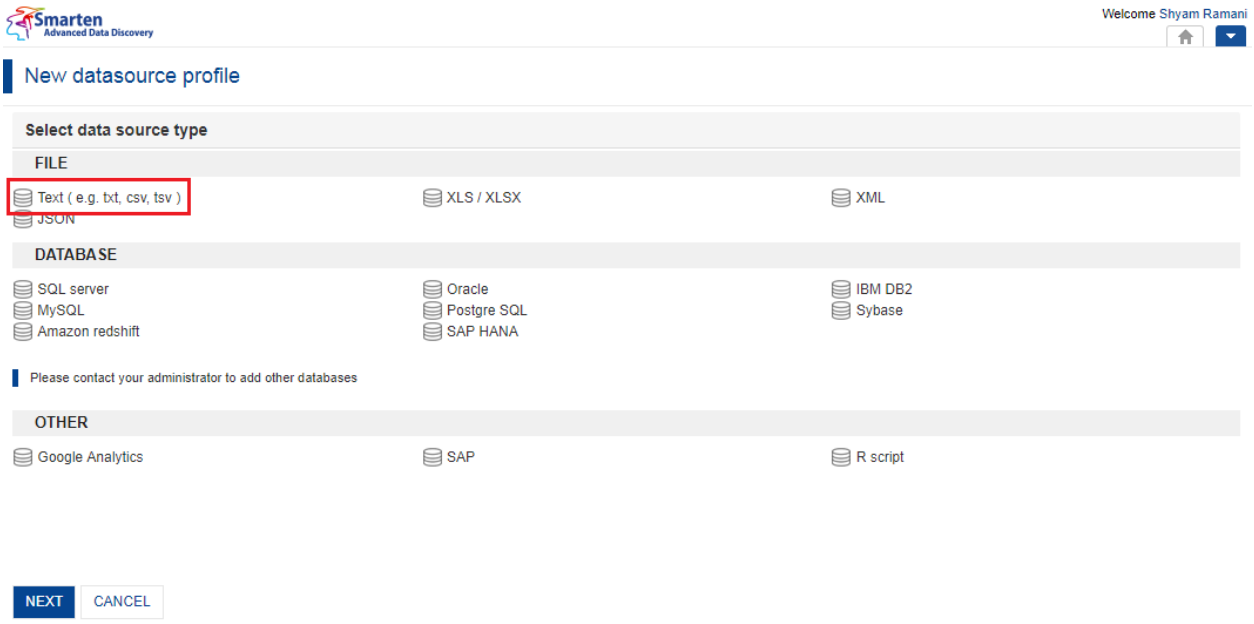
Procedure

1. Click **New** -> **Data source** from the menu.



MENU OPTION—NEW DATA SOURCE

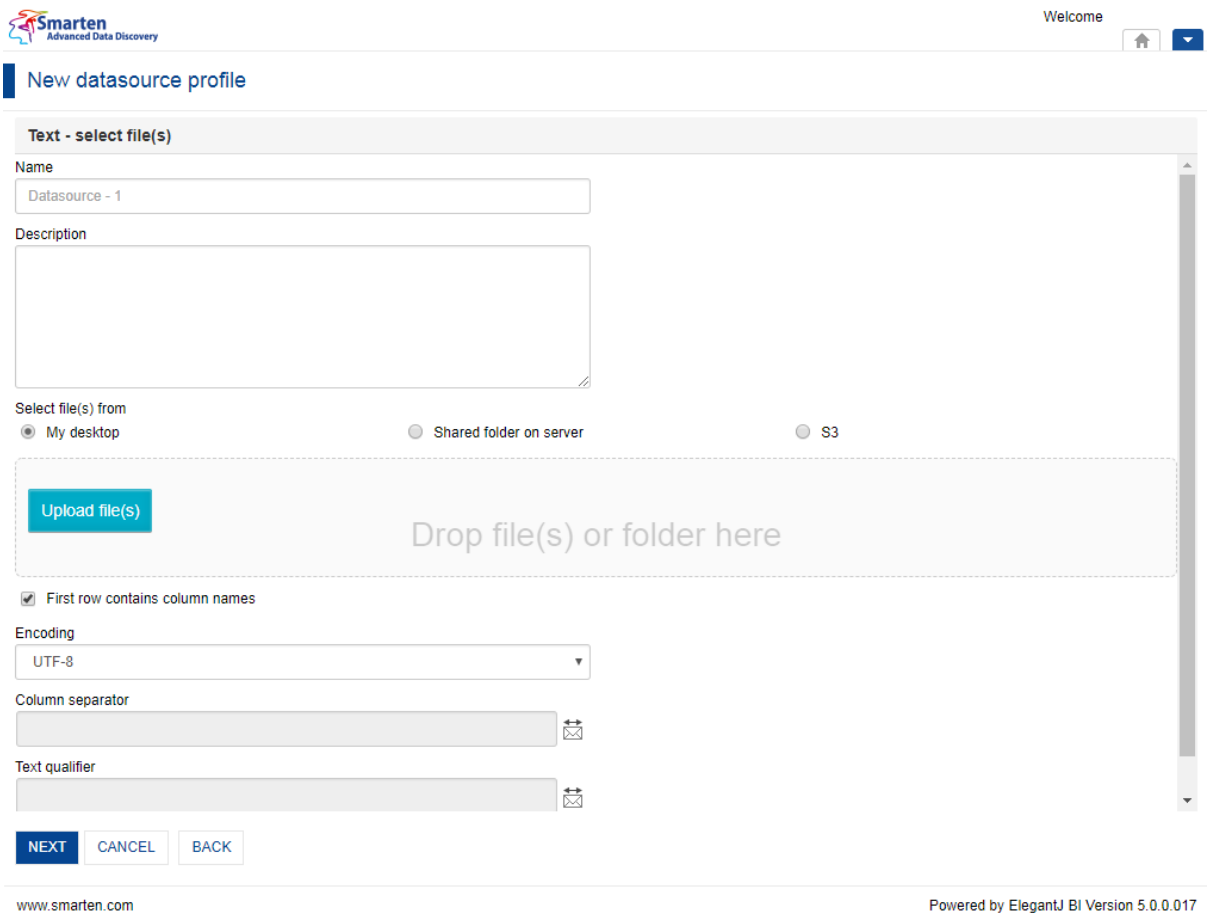
The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select **Text** from the **FILE** section, and then click **NEXT**.

   The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING A TEXT FILE TYPE PROFILE

3. Provide information in the following fields:
    - **Name**: Name for the data source profile.
    - **Description**: Description for the data source profile.
    - **Select file(s) from**: Options to specify the location where the file or folder to be used as the data source is available. The following options are available:
        - **My desktop**: Click **Upload file(s)** to select files you want to use as a data source from your machine.
        - **Shared folder on network**: Click **Select file(s)** to open the **Select folder(s) and file(s)** dialog box, and select file(s) and/or folder(s) from the shared folders on the server.
        - **S3**: Provide the **Access Key** and **Secret Key** for your AWS S3 account credential, and then select the bucket you want to retrieve data from. Click **OK** to open the **Select folder(s) and file(s)** dialog box, and then select the file(s) and/or folder(s) from the S3 bucket.
    - **First row contains column name**: Option to specify if the first row in the file contains the name of the columns.
    - **Encoding**: Select the encoding that is applied on the file from the list.
    - **Column separator**: The system identifies the column separator used in the file and displays it in this field.
    - **Text qualifier**: The system identifies the text qualifier used in the file and displays it in this field.
    - **Column data contains multiline values**: You can select this option if there are any multiline values in a column of the file.

> Note:
> The system automatically detects the column separator and text qualifier used in the file. The system also detects if the first row contains the name of the columns. These fields display the values used for the column separator and text qualifier and highlight the fields once the file is successfully uploaded on the system.

4. Click **NEXT**.
   The system displays the preview of the data available in the files.
5. Click **OK**.

   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

   The data source profile is now available in the repository.

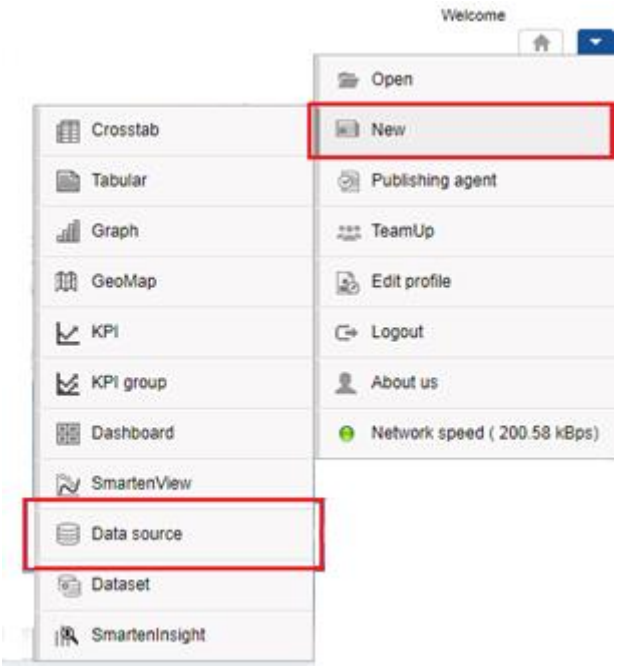
4.2.2 Creating an Excel File Type Profile

About this task

Use this task to create an Excel data source profile. Both xls and xlsx formats can be used for creating an Excel data source profile.

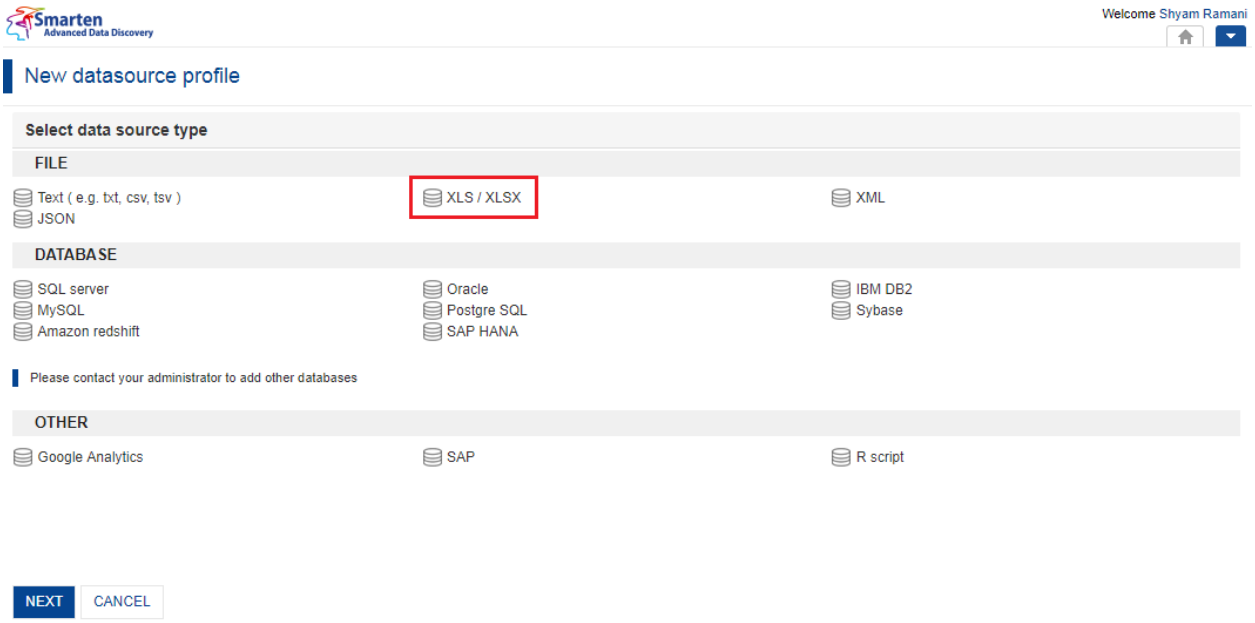
Procedure

1 Click **New** -> **Data source** from the menu.



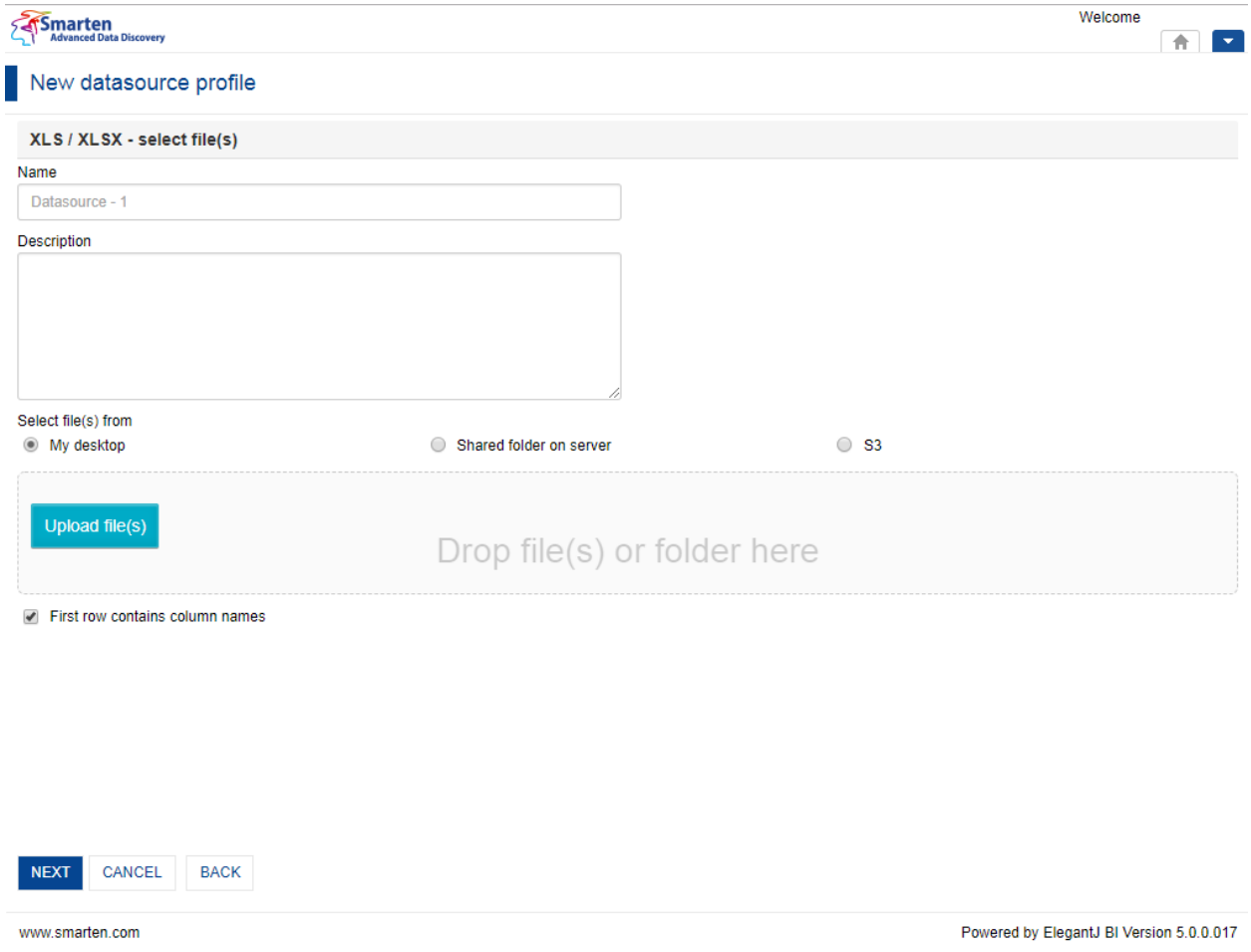
MENU OPTION—NEW DATA SOURCE

The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2 Select **XLS/XLSX** from the **FILE** section, and then click **NEXT**.
The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING AN EXCEL FILE TYPE PROFILE

3 Provide information in the following fields:
- **Name**: Name for the data source profile.
- **Description**: Description for the data source profile.
- **Select file(s) from**: Options to specify the location where the file or folder to be used as the data source is available. The following options are available:
  - **My desktop**: Click **Upload file(s)** to select files you want to use as a data source from your machine.
  - **Shared folder on network**: Click **Select file(s)** to open the **Select folder(s) and file(s)** dialog box, and select file(s) and folder(s) from the shared folders on the server.
  - **S3**: Provide the **Access Key** and **Secret Key** for your AWS S3 account credential, and then select the bucket you want to retrieve data from. Click **OK** to open the **Select folder(s) and file(s)** dialog box, and then select the file(s) and/or folder(s) from the S3 Bucket.
- **First row contains column name**: Option to specify if the first row in the file contains the name of the columns.

> Note:
> The system automatically detects if the first row contains the name of the columns once the file is successfully uploaded on the system.

4 Click **NEXT**.
The system displays the preview of the data available in the files.
5 Click **OK**.
The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

The data source profile is now available in the repository.

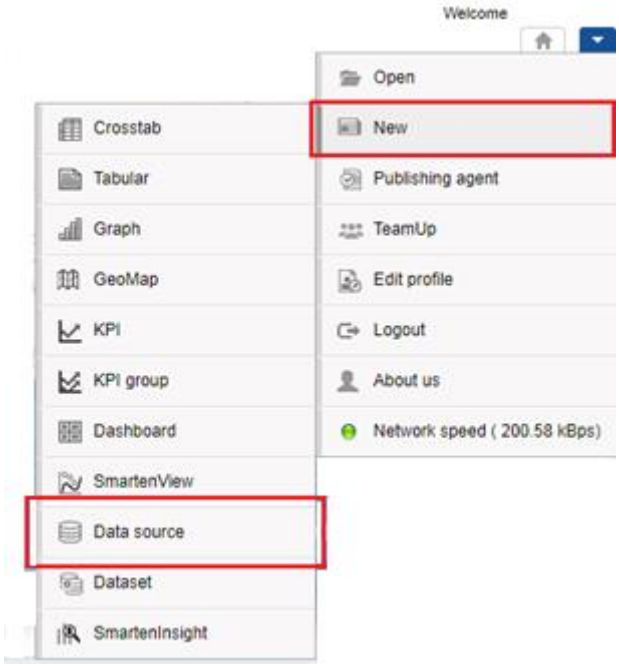
4.2.3 Creating an XML File Type Profile

About this task

Use this task to create an XML data source profile.

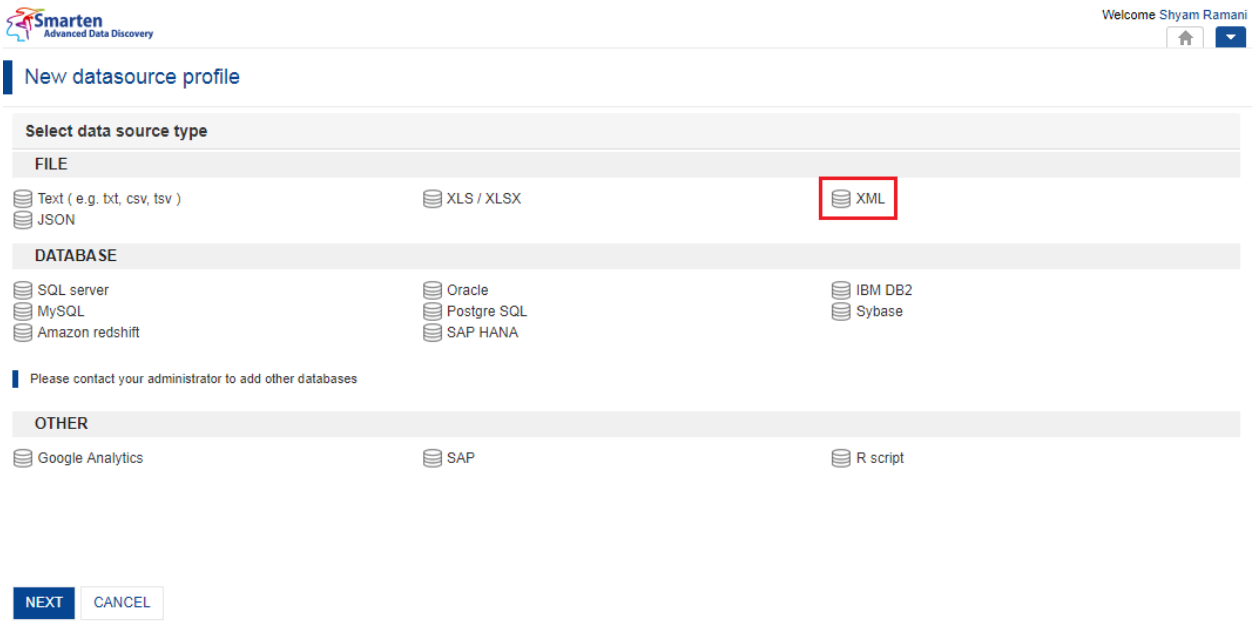
Procedure

1. Click **New** -> **Data source** from the menu.



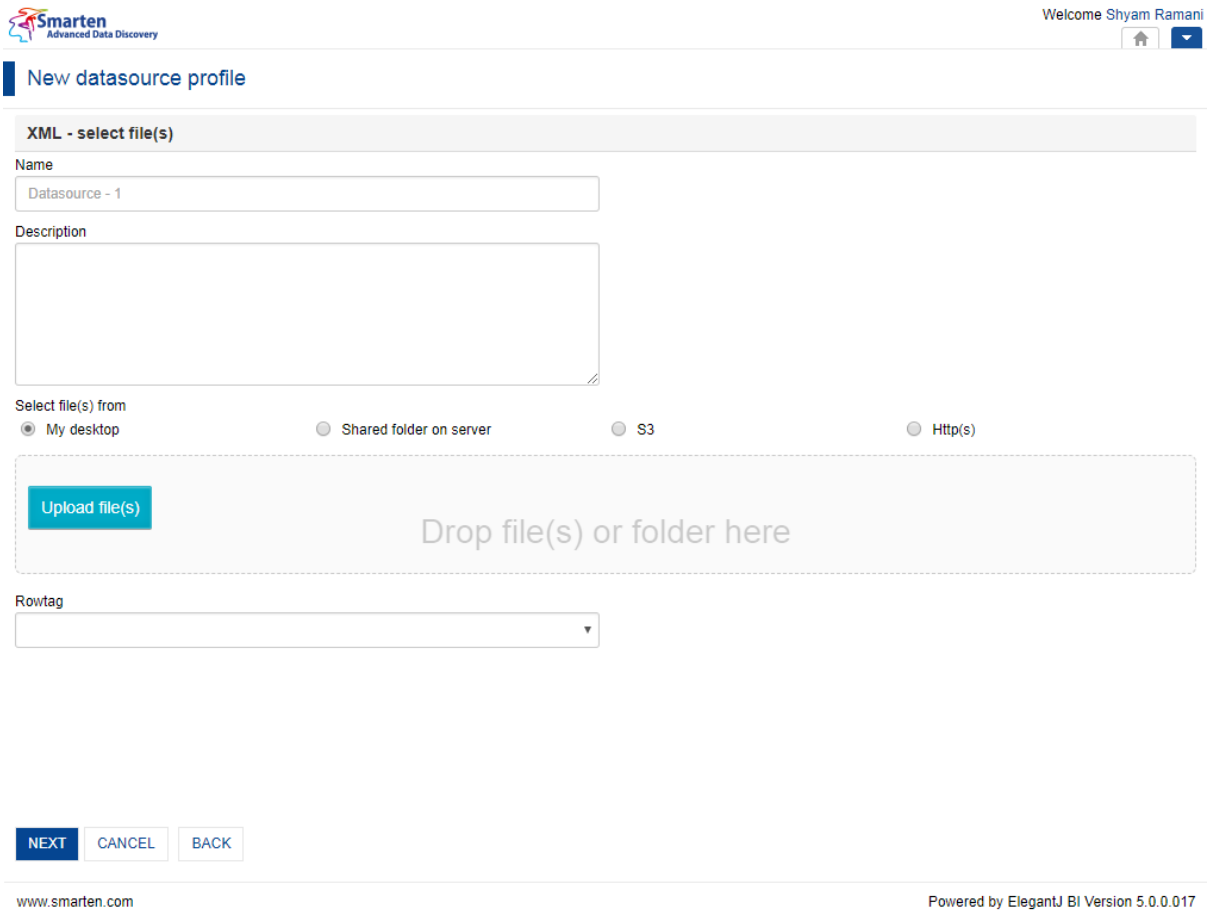
MENU OPTION—NEW DATA SOURCE

The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select **XML** from the **FILE** section, and then click **NEXT**.
   The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING AN XML FILE TYPE PROFILE

3. Provide information in the following fields:
    - **Name**: Name for the data source profile.
    - **Description**: Description for the data source profile.
    - **Select file(s) from**: Options to specify the location where the file or folder to be used as the data source is available. The following options are available:
        - **My desktop**: Click **Upload file(s)** to select files you want to use as a data source from your machine.
        - **Shared folder on network**: Click **Select file(s)** to open the **Select folder(s) and file(s)** dialog box, and select file(s) and folder(s) from the shared folders on the server.
        - **S3**: Provide the **Access Key** and **Secret Key** for your AWS S3 account credential, and then select the bucket you want to retrieve data from. Click **OK** to open the **Select folder(s) and file(s)** dialog box, and then select the file(s) and/or folder(s) from the S3 Bucket.
    - **Rowtag**: This list displays the tags that are available in the XML file. You can select a tag that you want to be considered as a row.
4. Click **NEXT**.
   The system displays the preview of the data available in the files.
5. Click **OK**.
   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

The data source profile is now available in the repository.

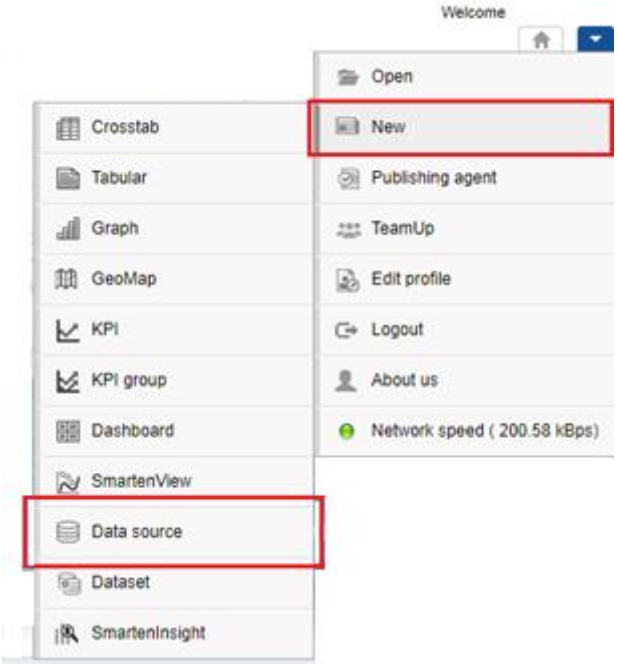
4.2.4 Creating a JSON File Type Profile

About this task

Use this task to create a JSON data source profile.

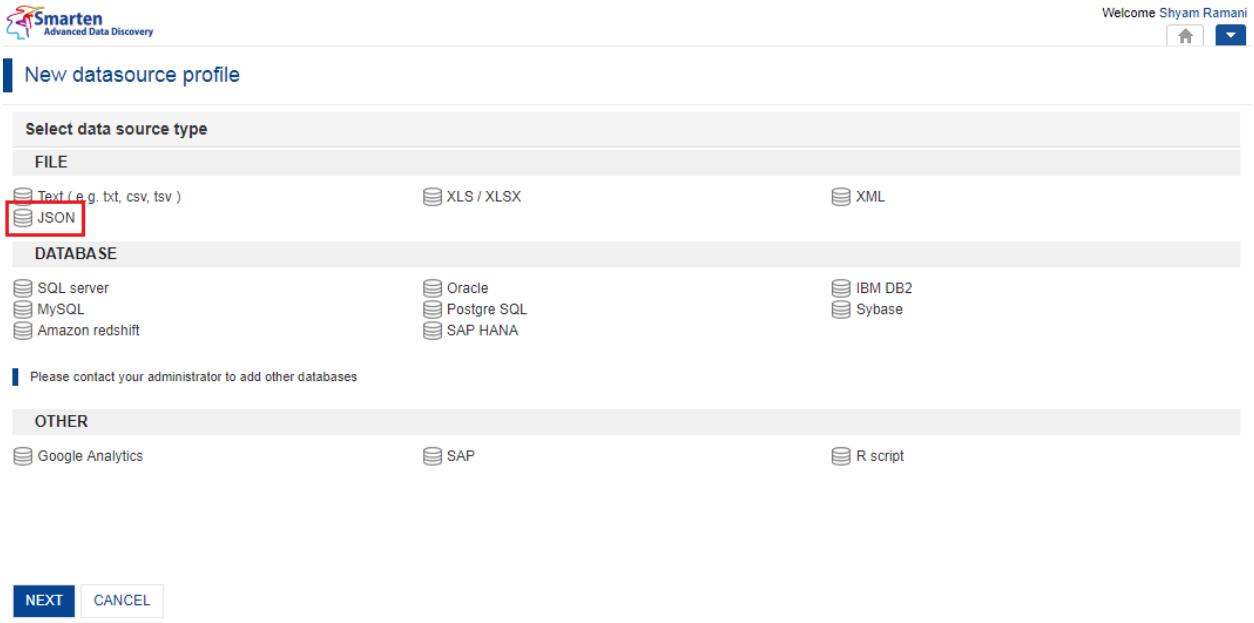
Procedure

1. Click **New** -> **Data source** from the menu.



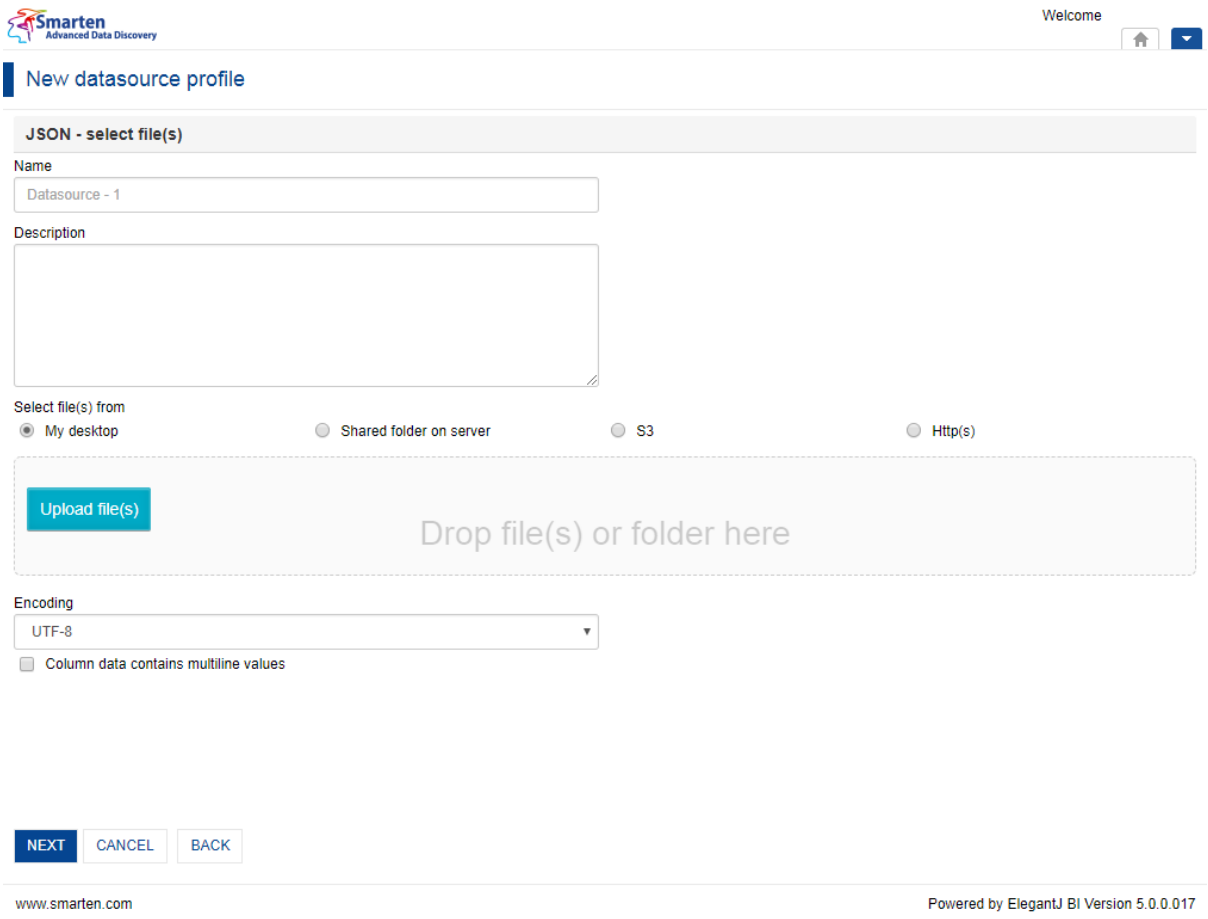
MENU OPTION—NEW DATA SOURCE

The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select **JSON** from the **FILE** section, and then click **NEXT**.
   The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING A JSON FILE TYPE PROFILE

3. Provide information in the following fields:
    - **Name**: Name for the data source profile.
    - **Description**: Description for the data source profile.
    - **Select file(s) from**: Options to specify the location where the file or folder to be used as the data source is available. The following options are available:
        - **My desktop**: Click **Upload file(s)** to select files you want to use as a data source from your machine.
        - **Shared folder on network**: Click **Select file(s)** to open the **Select folder(s) and file(s)** dialog box, and select file(s) and folder(s) from the shared folders on the server.
        - **S3**: Provide the **Access Key** and **Secret Key** for your AWS S3 account credential, and then select the bucket you want to retrieve data from. Click **OK** to open the **Select folder(s) and file(s)** dialog box, and then select the file(s) and/or folder(s) from the S3 Bucket.
    - **Encoding**: Select the encoding that is applied on the file from the list.
4. Click **NEXT**.
   The system displays the preview of the data available in the files.
5. Click **OK**.
   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

The data source profile is now available in the repository.

4.3 Creating a Google Analytics Profile

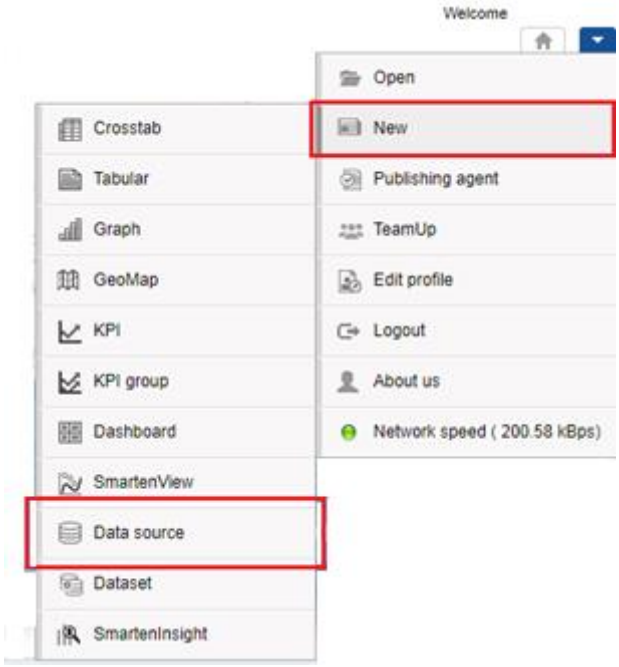
You can use data from Google Analytics associated with your digital asset, such as your website.

About this task

Use this task to create a Google Analytics data source profile.

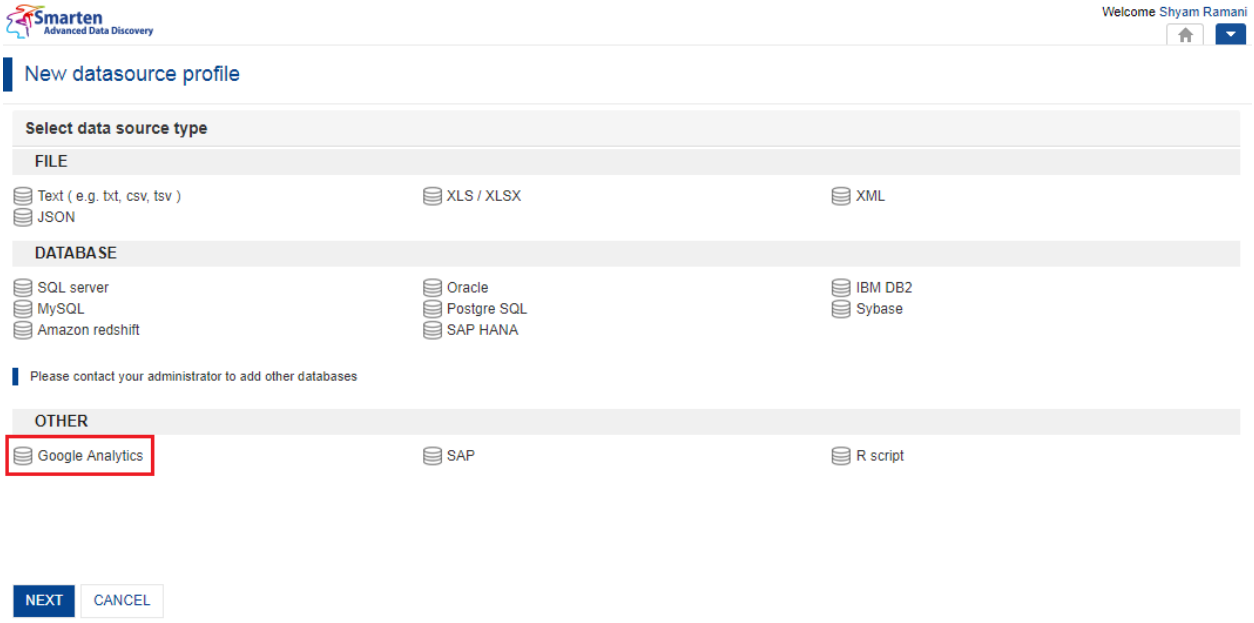
Procedure

1. Click **New** -> **Data source** from the menu.



MENU OPTION—NEW DATA SOURCE

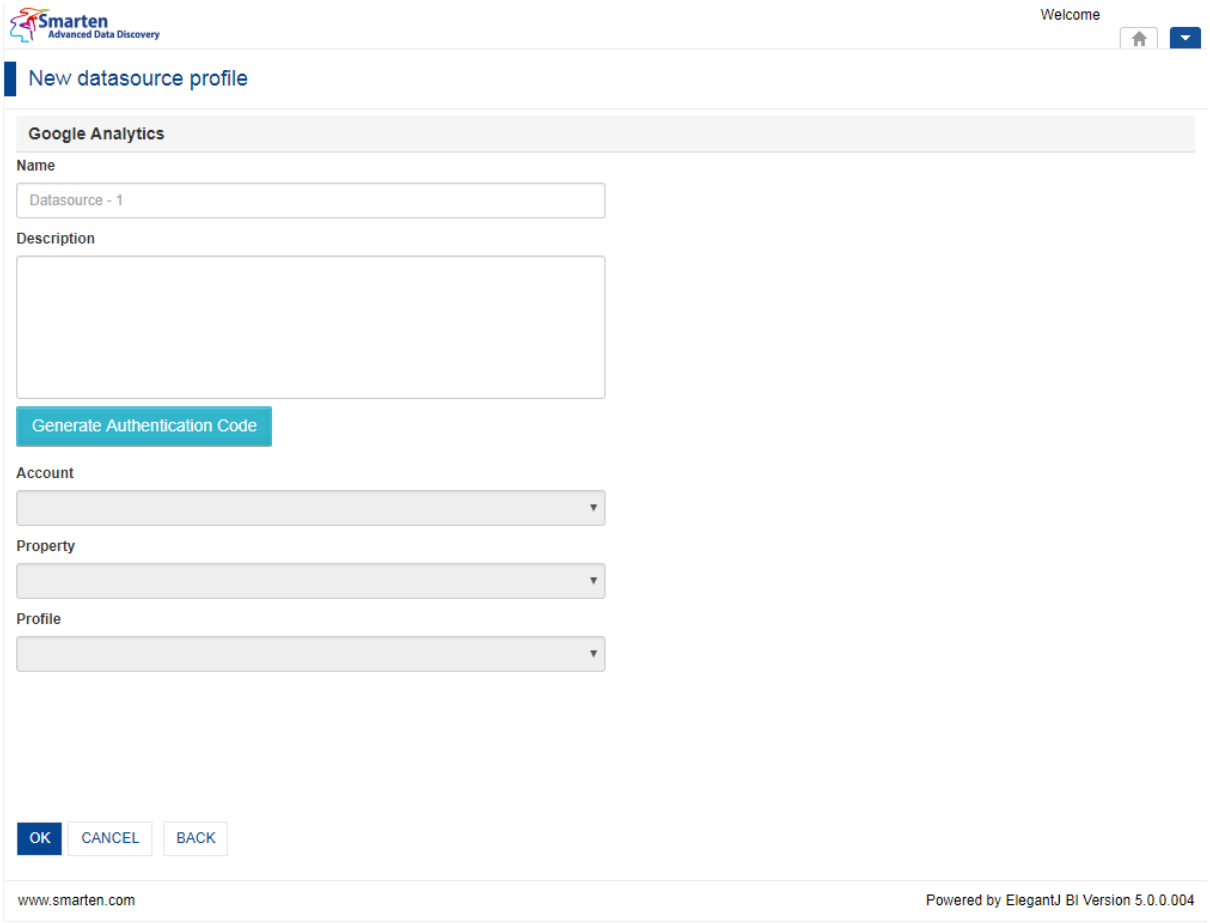
The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select **Google Analytics** from the **OTHER** section, and then click **NEXT**.

The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING A GOOGLE ANALYTICS PROFILE

3. Provide information in the following fields:
    - **Name**: Name for the data source profile.
    - **Description**: Description for the data source profile.
    - **Generate Authentication Code**: Click the button to generate an authentication code for Google Analytics. Copy the code in the field.
    - **Account**: Select the account you want to use to extract data.
    - **Property**: Select the website, mobile application, blog, or any other entity for which you want to extract data.
    - **Profile**: Select the Google Analytics profile you want to use for the selected property.
4. Click **OK**.
   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

The data source profile is now available in the repository.

4.4 Creating an R Script Profile

You can use R script as a data source to create an R cube..

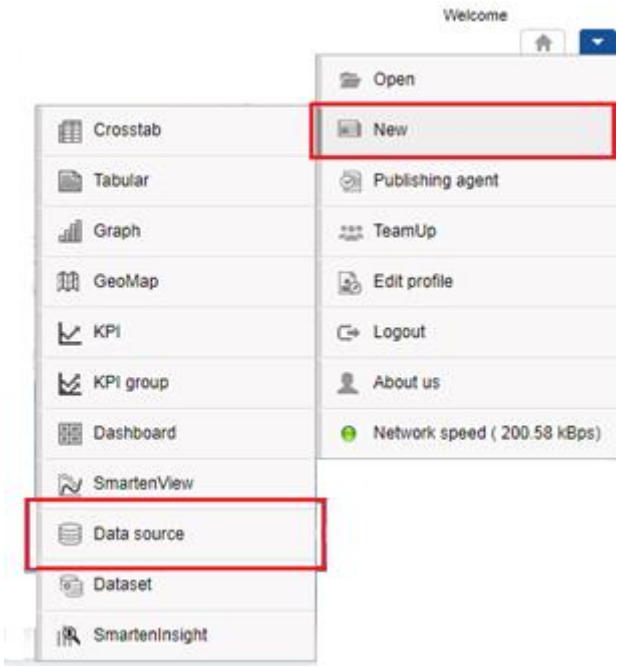
Reference: **Working with R Integration**

About this task

Use this task to create an R script data source profile.

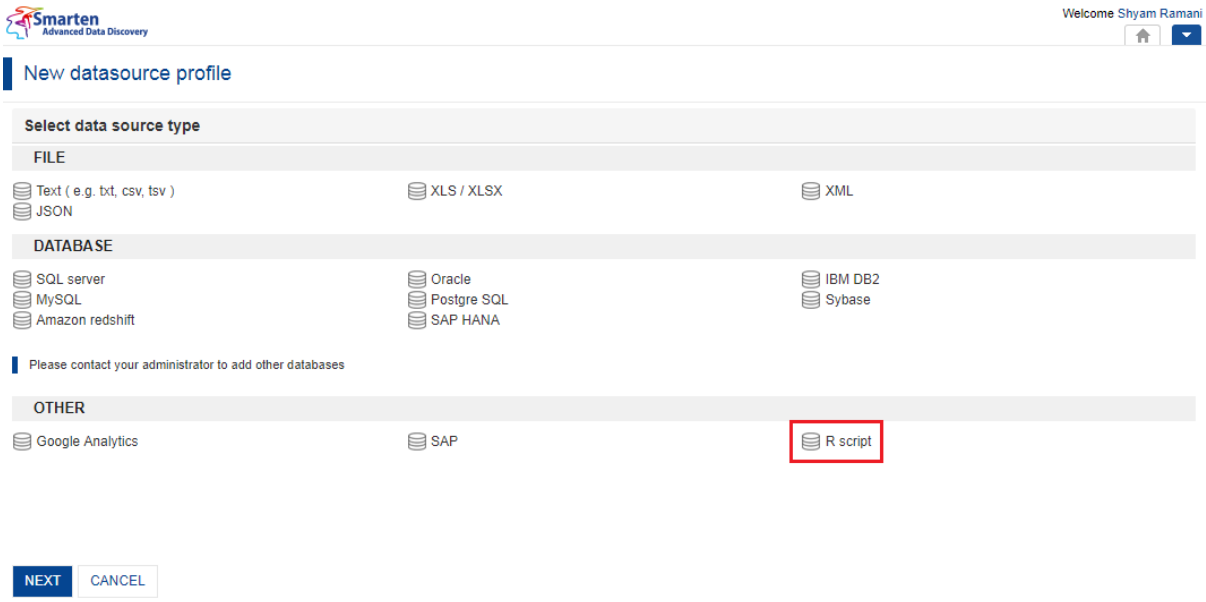
Procedure

1. Click **New** -> **Data source** from the menu.



MENU OPTION—NEW DATA SOURCE

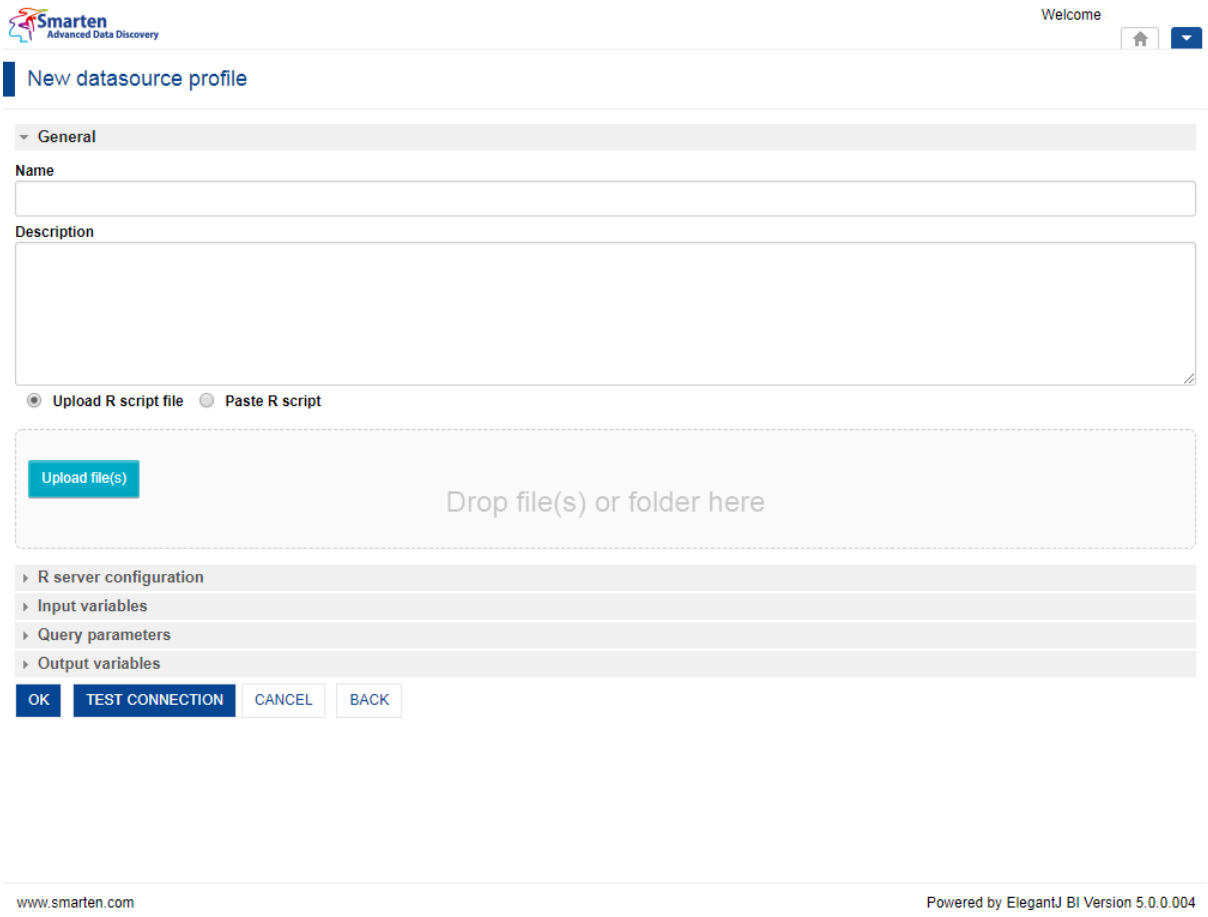
The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select **R script** from the **OTHER** section, and then click **NEXT**.

   The system displays the fields to provide values that will be used to create the data source profile.



THE NEW DATASOURCE PROFILE PAGE—CREATING AN R SCRIPT PROFILE

3. Provide information in the following fields:
   - **Name**: Name for the data source profile.
   - **Description**: Description for the data source profile.
   - **Upload R script file**: Select this option if you want to upload a file that contains an R script. Click **Upload file(s)** to locate and upload the file from your machine.
   - **Paste R script**: Select this option to paste an R script in the field.
4. Click **R server configuration**.
   a. Provide hostname and port number of the R server on which the R script has to be executed.
5. Click **Input variables**.
   a. Provide details about the Input variable used in the R script in the **Variable name (R script)** and **Display name** fields, and select its type from the **Input type** list. The following options are available:
      i. **Single value**: Select this option to allow users to manually enter data to be used as input for the R script variable.
      ii. **Single column**: Select this option to allow users to map only one column of the data source with an Input variable of R script. The data available in the selected column is used as input for the R script variable.

iii. **Multi columns**: Select this option to allow users to map more than one column of the data source with an Input variable of R script. The data available in all the selected columns are used as input for the R script variable.

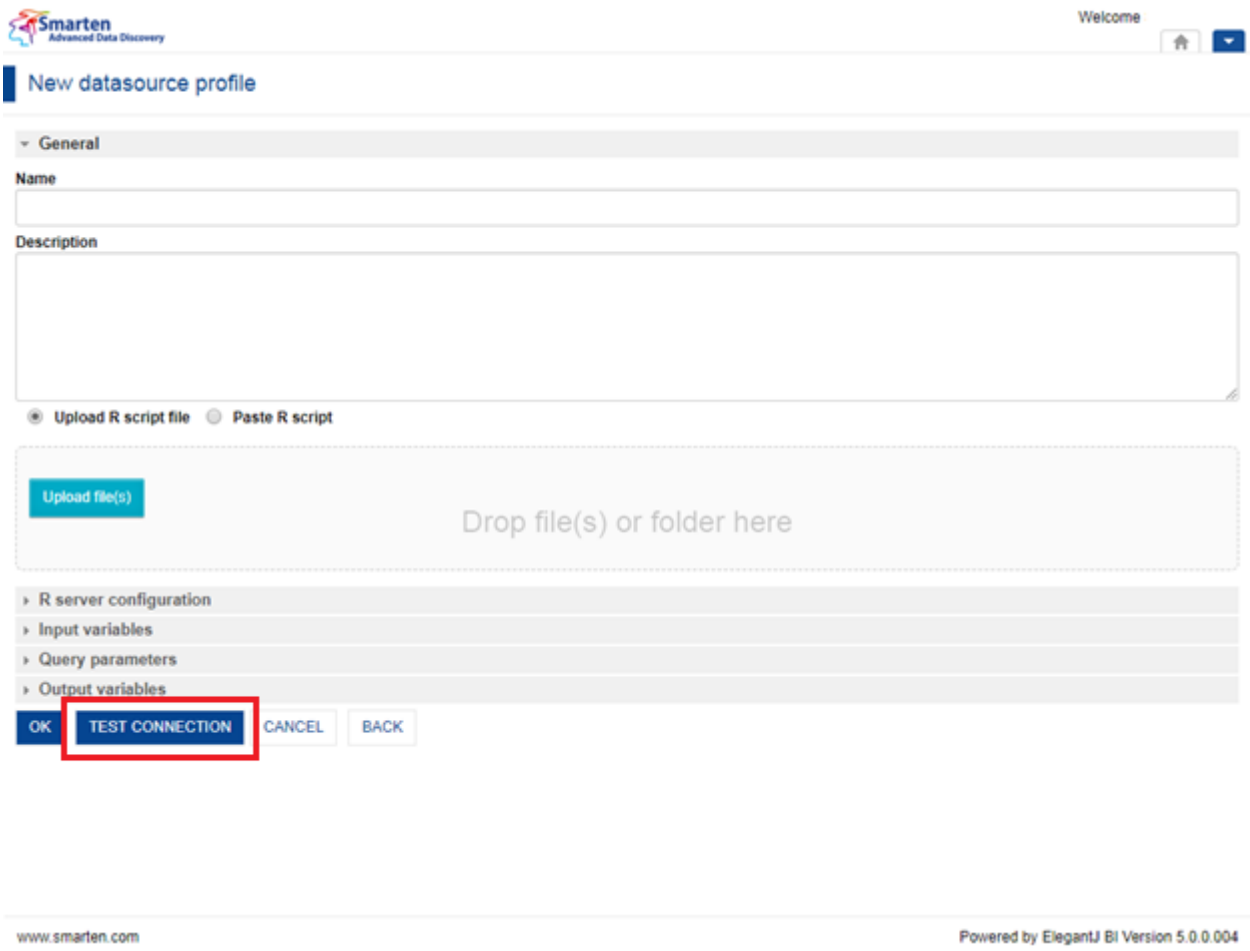
b. You can click the Add icon to add details for multiple input variables.

6. Click **Query parameters**.
   a. Provide details about the Input variable used in the R script in the **Variable name (R script)** and **Display name** fields, and select its type from the **Input type** list.
   b. You can click the Add icon to add details for multiple query parameters.
7. Click **Output variables**.
   a. Provide details about the Input variable used in the R script in the **Variable name (R script)** and **Display name** fields, and select its type from the **Input type** list.
   b. You can click the Add icon to add details for multiple output variables.
8. Click **OK**.
   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

The data source profile is now available in the repository.

4.4.1 Testing an R Integration Connection

Once you have provided all the values to set up a connection with an R integration, you can test to ensure that the connection with the R integration is established successfully.



Once you have provided all the information for creating an R script profile, you can click the **TEST CONNECTION** option to test the connection with the R server.

4.5 Creating an SAP Profile

You can use an SAP profile to connect to SAP and retrieve data through BAPIs on the SAP server.

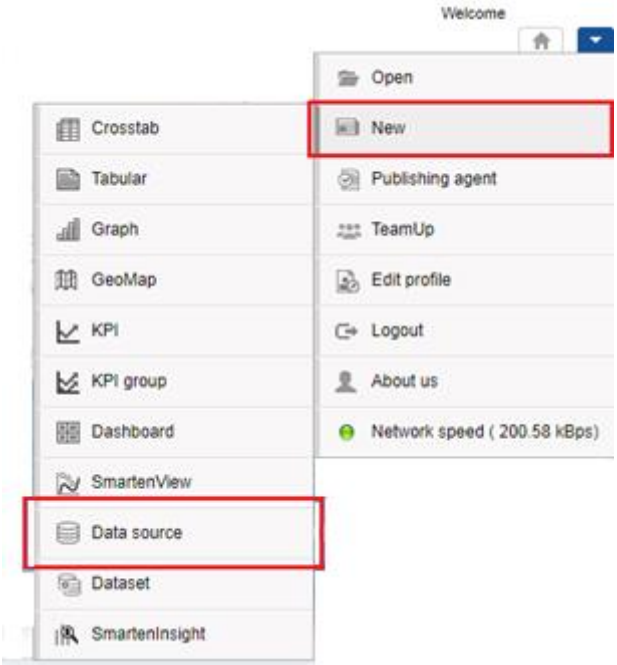
Reference: **Integration with SAP**

About this task

Use this task to create an SAP data source profile.

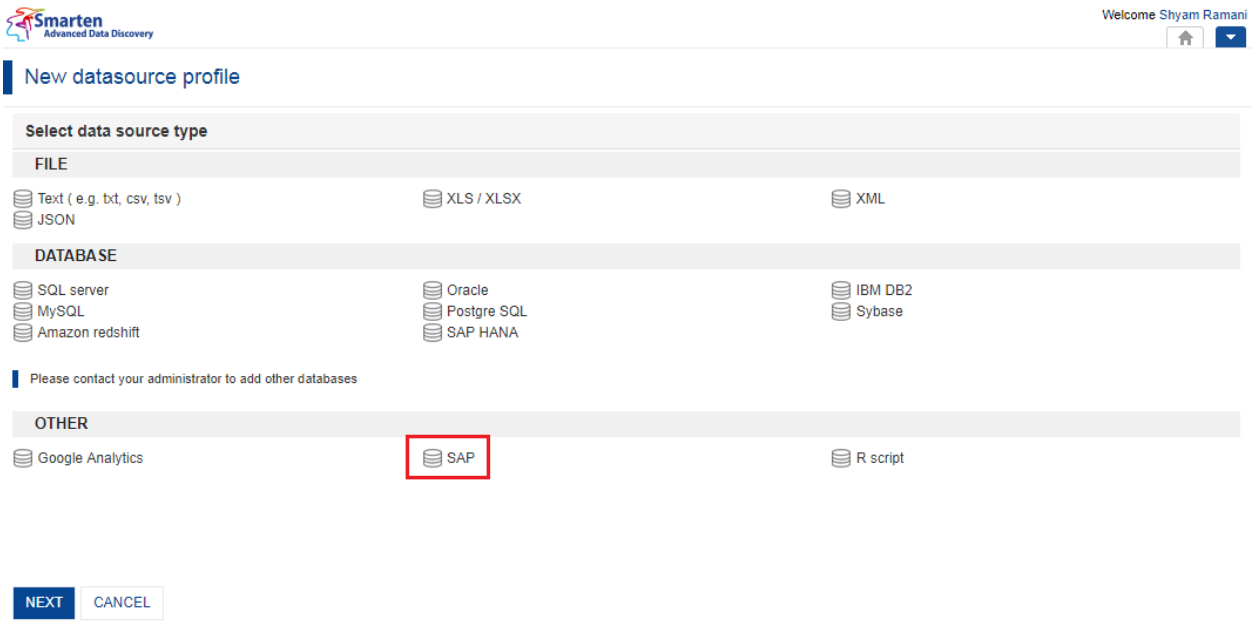
Procedure

1. Click **New** -> **Data source** from the menu.



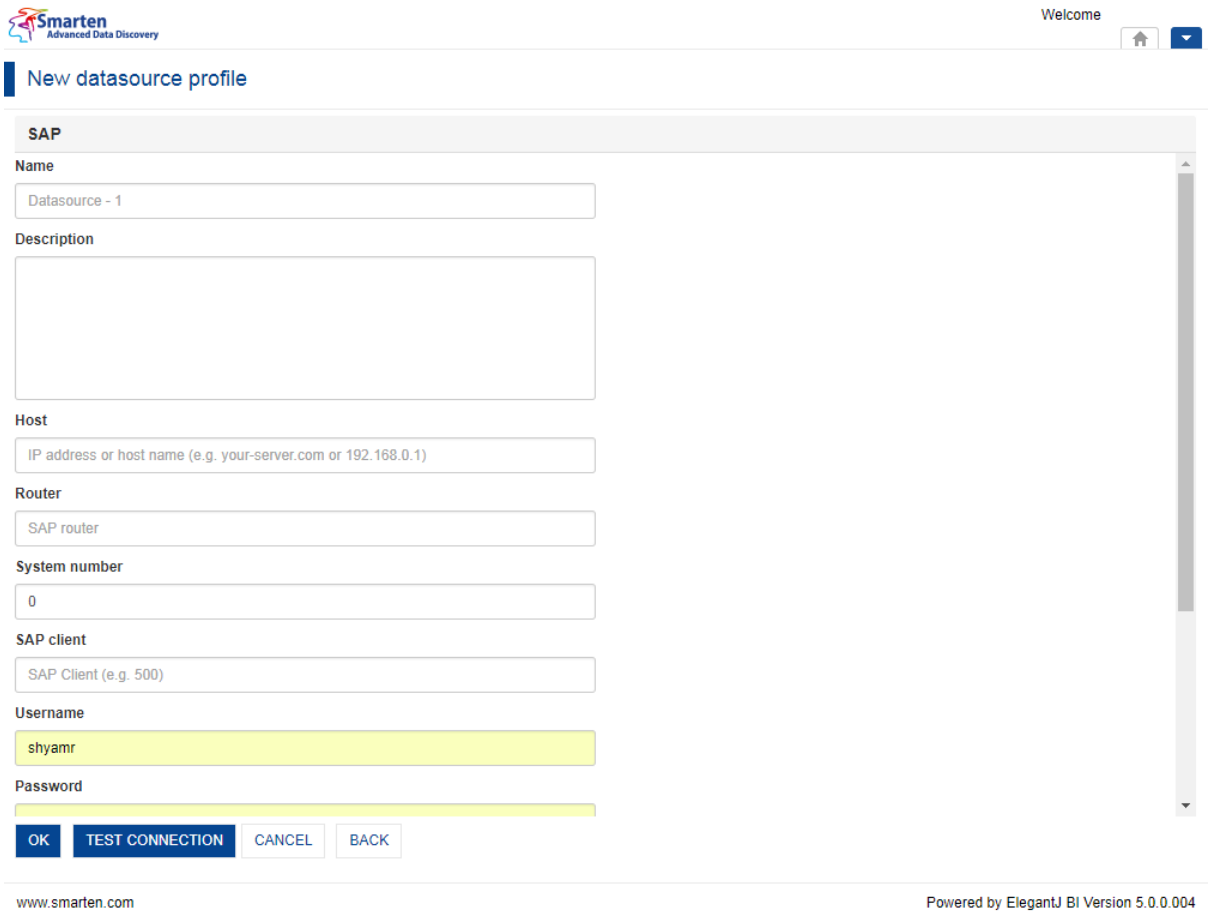
MENU OPTION—NEW DATA SOURCE

The system displays the **New datasource profile** page.



THE NEW DATASOURCE PROFILE PAGE—SELECTING A DATA SOURCE TYPE

2. Select **SAP** from the **OTHER** section, and then click **NEXT**.
   The system displays the fields to provide values that will be used to create the data source profile.



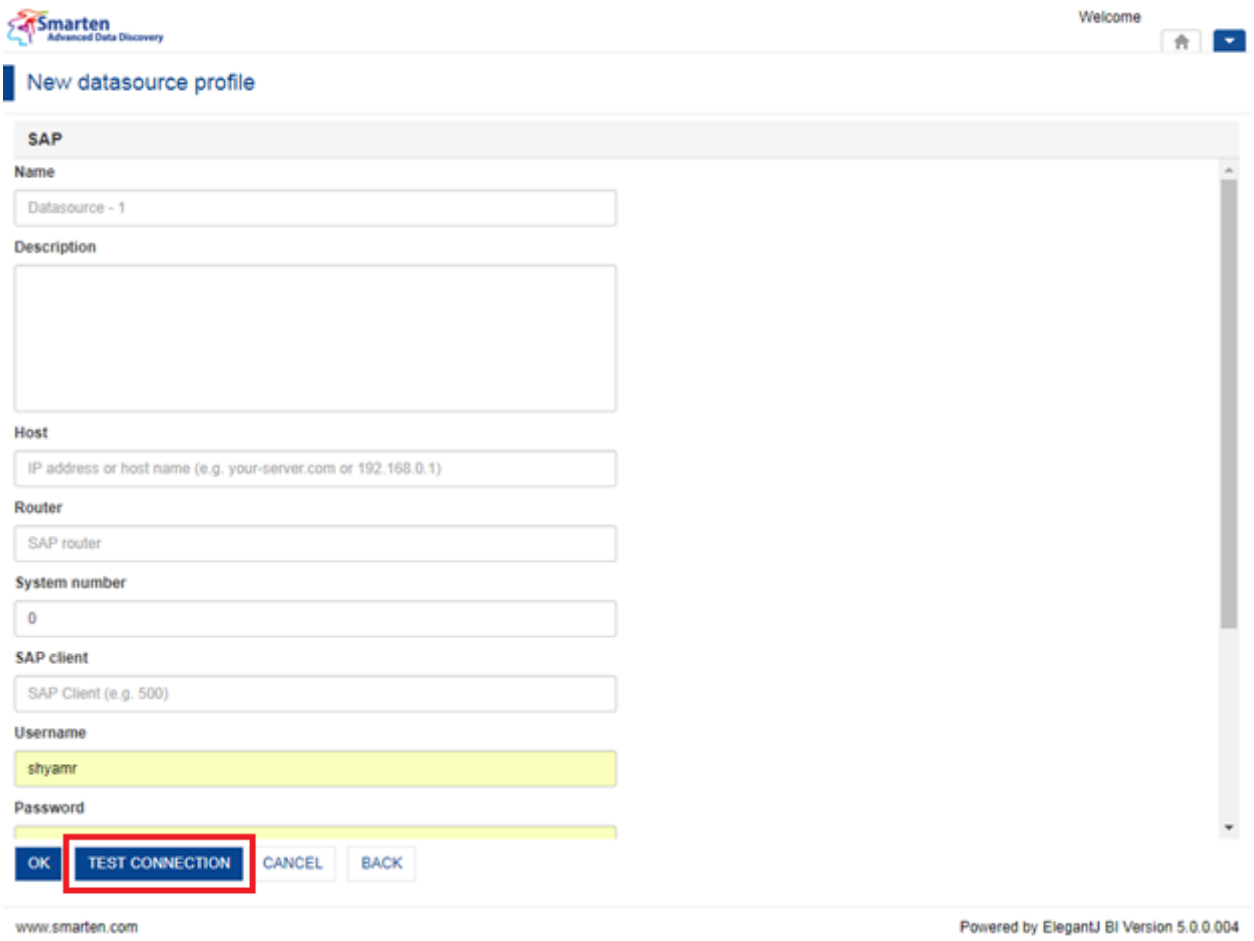
THE NEW DATASOURCE PROFILE PAGE—CREATING AN SAP PROFILE

3. Provide information in the following fields:
   - **Name**: Name for the data source profile.
   - **Description**: Description for the data source profile.
   - **Host**: Provide the hostname or IP address of the SAP system.
   - **Router**: Provide details of the SAP router.
   - **System number**: Provide details of the SAP system number.
   - **SAP client**: Provide details of the SAP client.
   - **Username**: Provide the username that will be used to connect to the SAP system.
   - **Password**: Provide the password that will be used to connect to the SAP system.
4. Click **OK**.
   The system displays a confirmation message after the data source is successfully created. It also allows you to create a dataset using the data source profile.

The data source profile is now available in the repository.

4.5.1 Testing a SAP Connection

Once you have provided all the values to set up a connection with the SAP, you can test to ensure that the connection with the SAP system is established successfully.



Once you have provided all the information for creating an SAP profile, you can click the **TEST CONNECTION** option to test the connection with the SAP system.

5 Data Source Management

Managing a data source involves providing access rights, editing, deleting, copying, and getting a data source IT certified.

5.1 Editing a Data Source

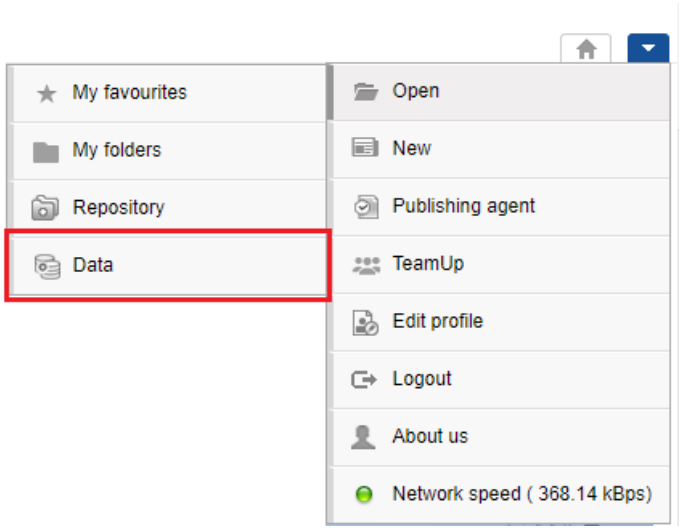
This feature enables authorized users to edit a data source. Users can change the name, description, and other attributes based on the type of data source.

About this task

Use this task to edit a data source.

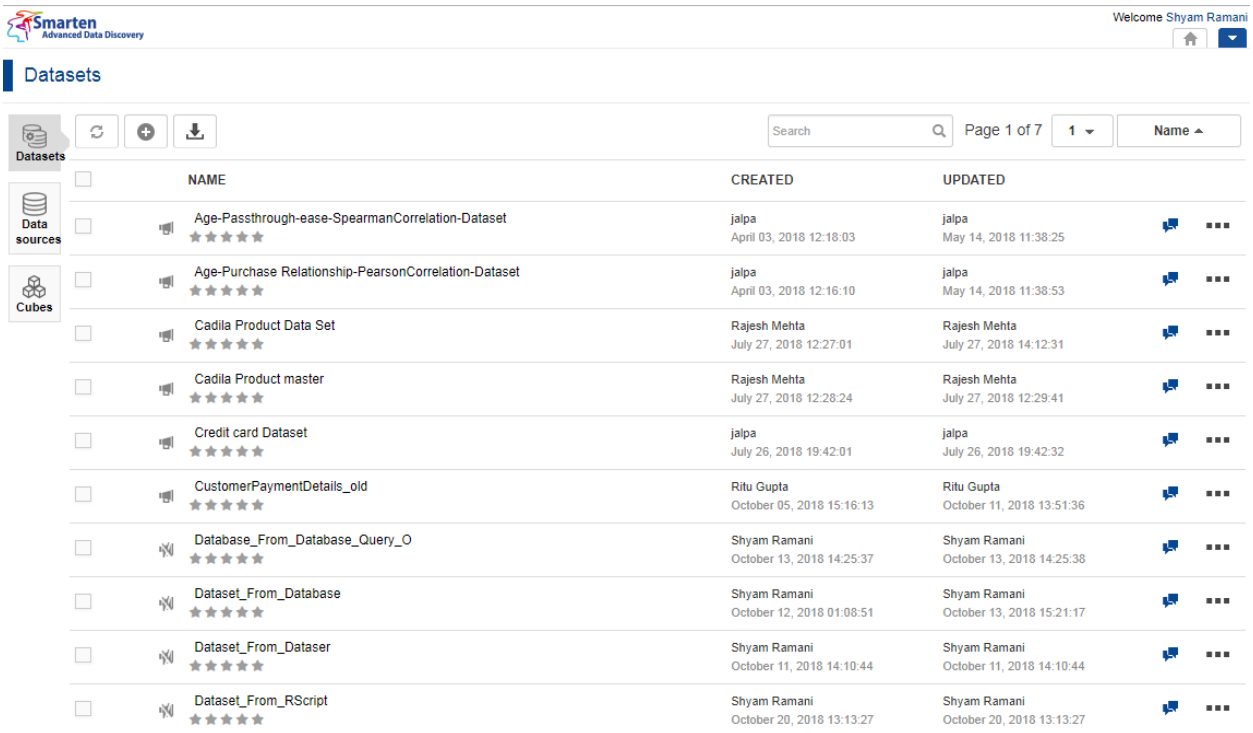
Procedure

1. Click **Open** -> **Data** from the menu.



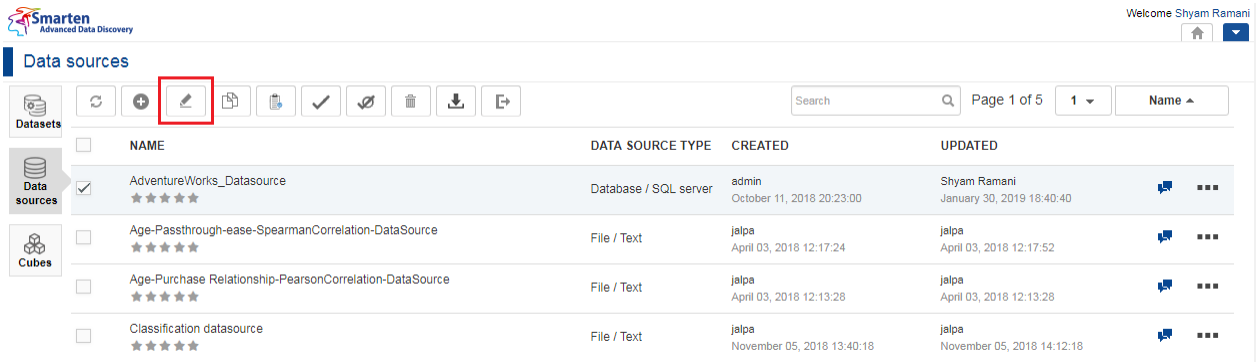
MENU OPTION—OPEN DATA

The system displays the following page.



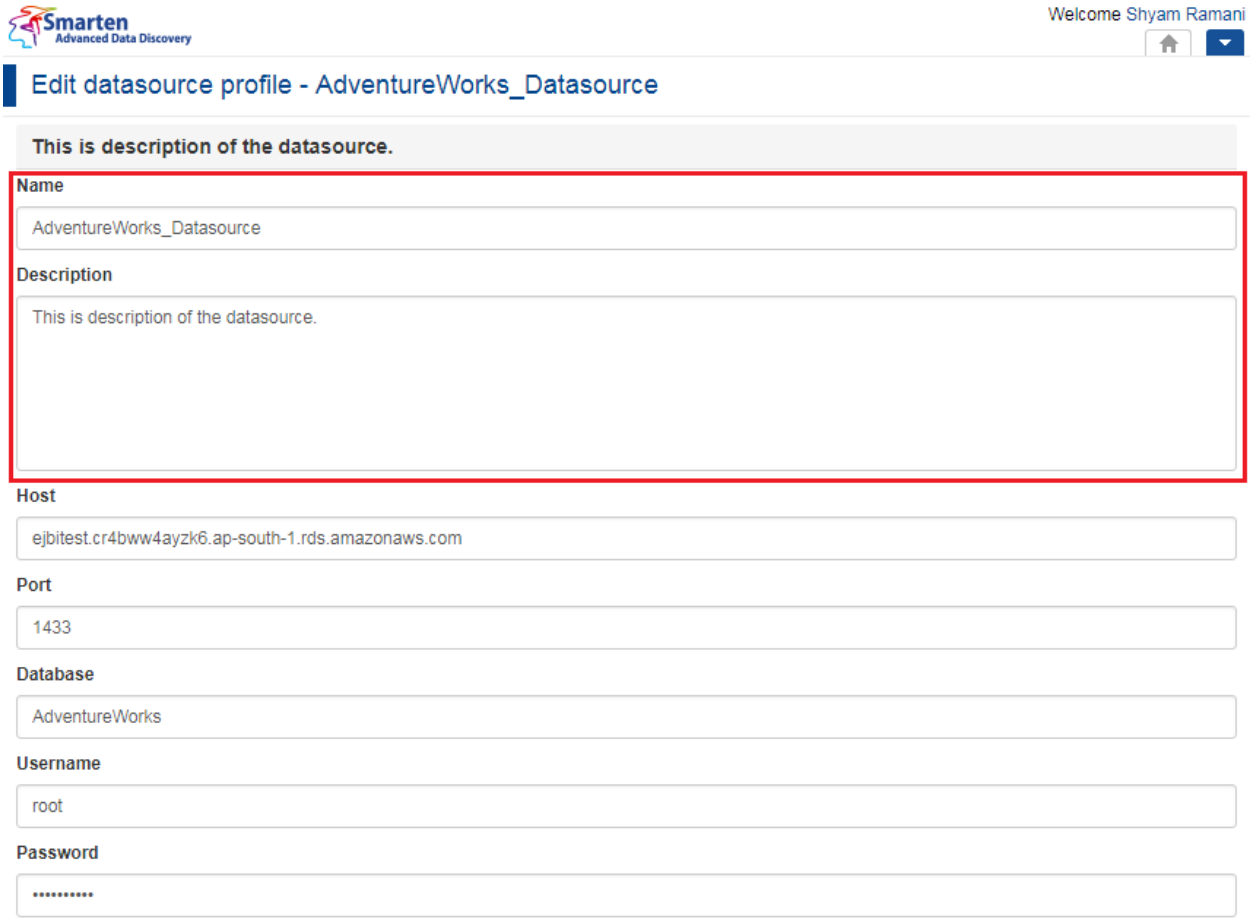
ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Select the check box adjacent to the data source you want to edit.
4. Click the Edit icon.
   The system displays the **Edit datasource** page.



EDITING A DATA SOURCE—THE EDIT ICON

5. You can provide a new name and description for the data source in the **Name** and **Description** boxes.



EDITING A DATA SOURCE—EDITING NAME AND DESCRIPTION

6. You can edit the remainder of the fields as per your requirement and then click **OK**.

> Note:
> The fields available are based on the type of the data source.

5.2 Deleting a Data Source

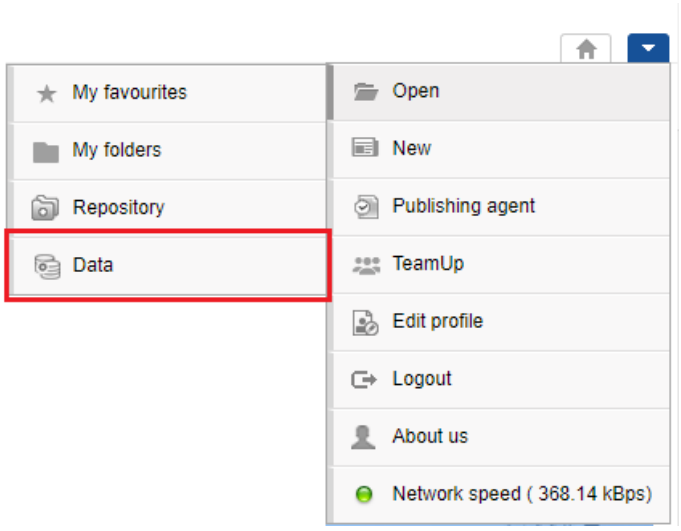
This feature enables authorized users to delete a data source. A deleted data source is no longer available in the system.

About this task

Use this task to delete a data source.

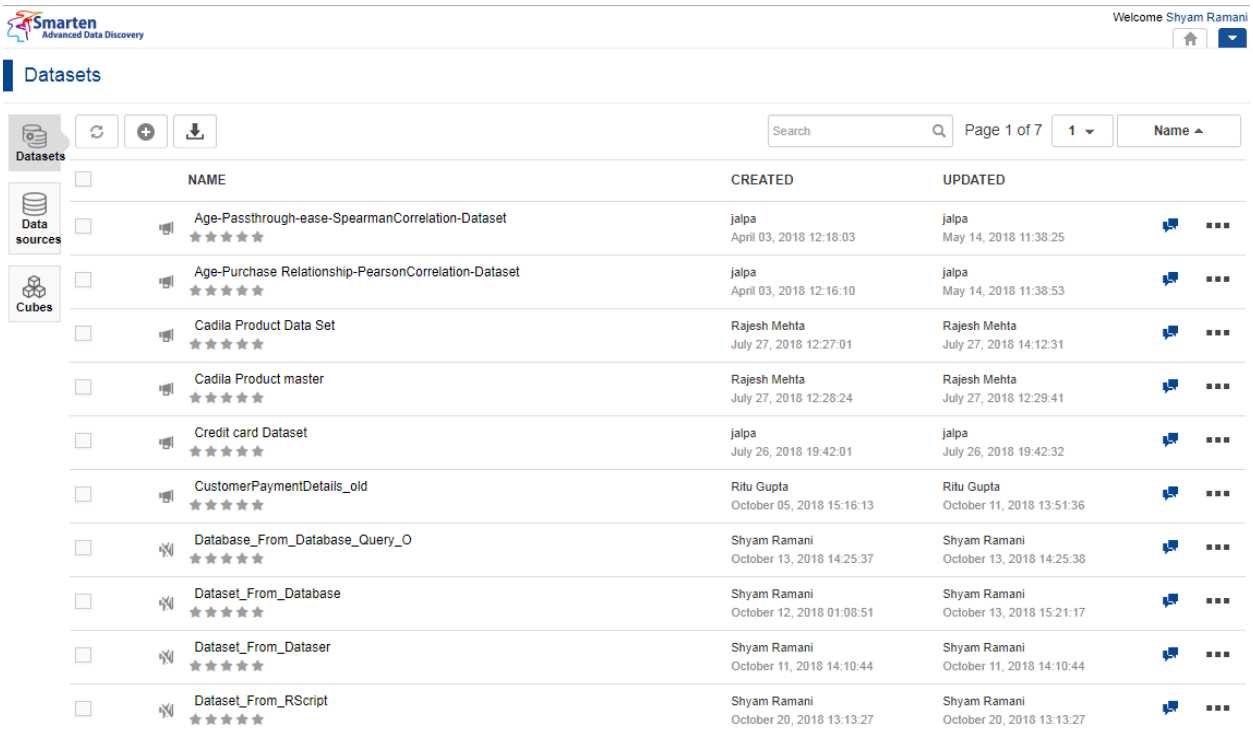
Procedure

1. Click **Open** -> **Data** from the menu.



MENU OPTION—OPEN DATA

The system displays the following page.

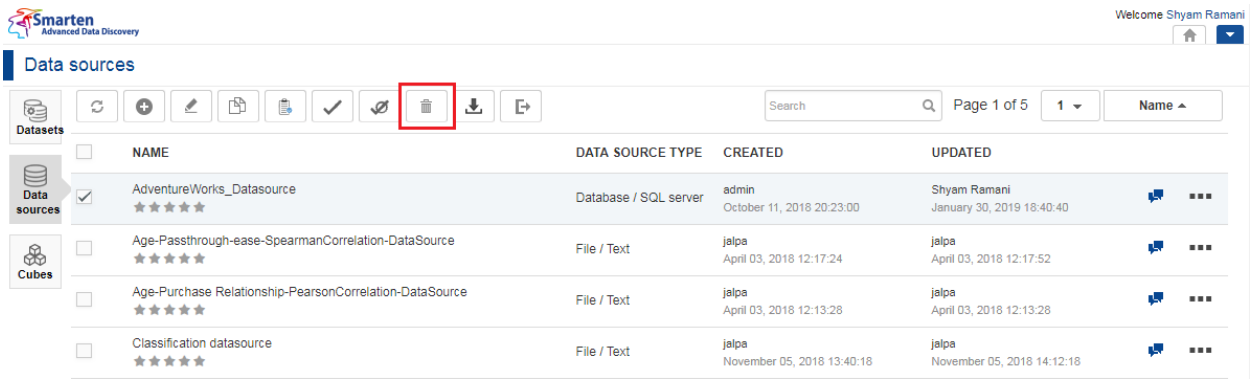


ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Select the check box adjacent to the data source you want to delete.

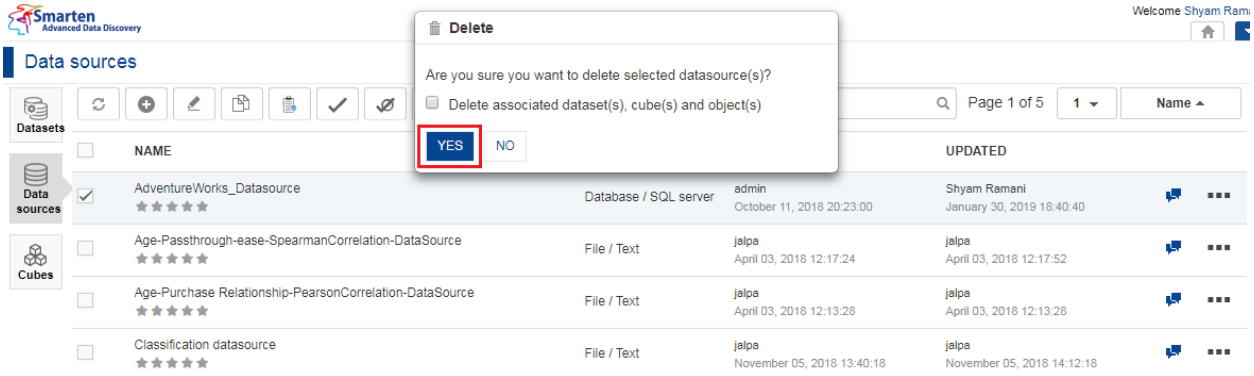
4. Click the Delete icon.

   The system displays the **Delete** dialog box.



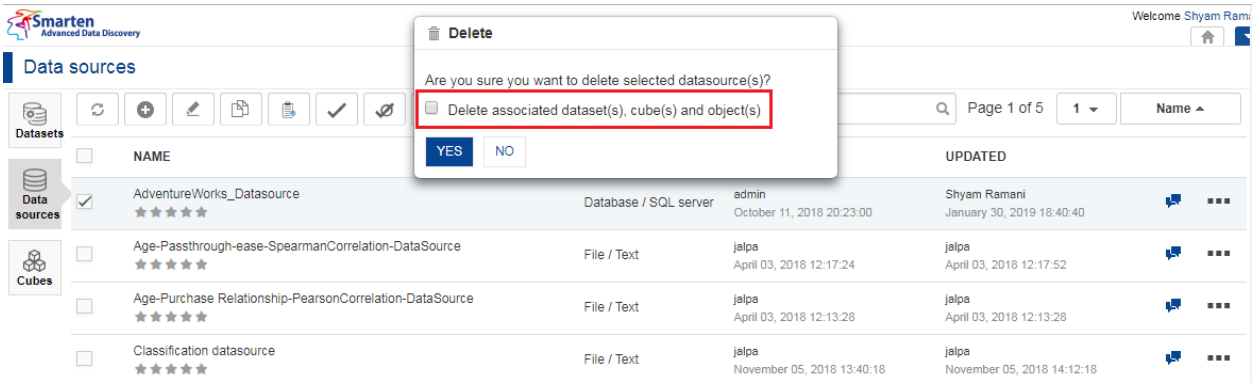
DELETING A DATA SOURCE—CLICKING THE DELETE ICON

5. Click **YES** to delete the selected data source.



DELETING A DATA SOURCE—PROVIDING CONFIRMATION TO DELETE THE DATA SOURCE

If there are datasets, cubes, and objects associated with the data source, the system displays the option asking if you want to delete the associated datasets, cubes, and objects along with the data source. You can select the option if you want to delete the datasets, cubes, and objects.



DELETING A DATA SOURCE—DELETING ASSOCIATED OBJECTS

5.3 Managing Access Rights for a Data Source

Access permission is about granting or restricting access to a data source. Permissions are provided to view, edit, or delete a data source. For example, team members of a sales team have all the rights to perform all actions on the sales-related data sources, whereas only view rights is assigned to members of other teams. Similarly, you can have a data source for a marketing team. Members of the marketing team have all the rights assigned to them, whereas members of the other team have only view rights assigned to them. Access permissions can be given as per Roles or to individual users of Smarten by the user who created the data source or by the Administrator. The access rights provided by the last user whether creator or Administrator are applicable. The following table describes the rights that can be assigned to a user:

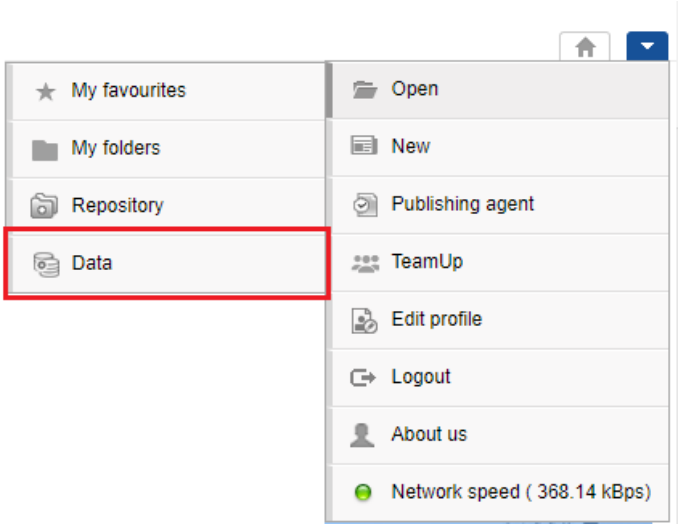
Right	Description
View	This access right grants permission to a user to view and access a data source.
Write	This access right grants permission to a user to edit a data source.
Delete	This access right grants permission to a user to delete a data source.
Export	This access right grants permission to a user to export a data source.

About this task

Use this task to manage access rights for a data source.

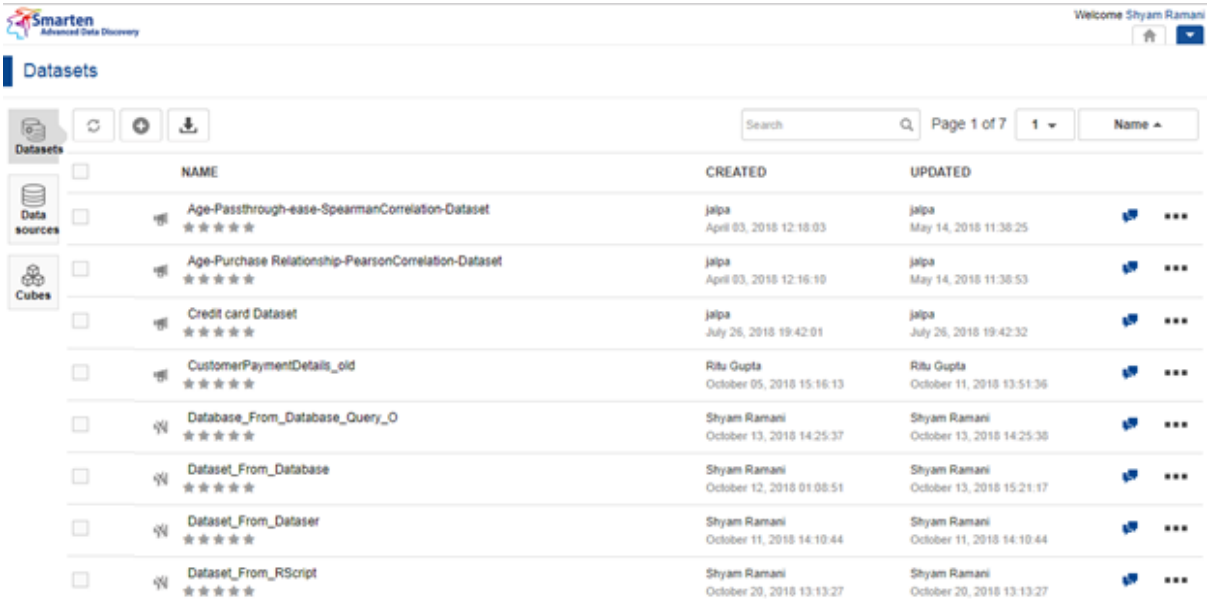
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION—OPEN DATA

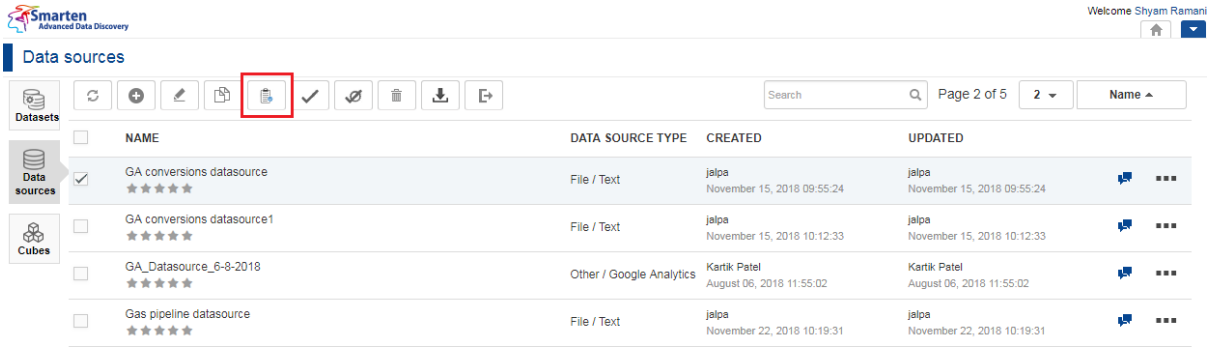
The system displays the following page.



ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

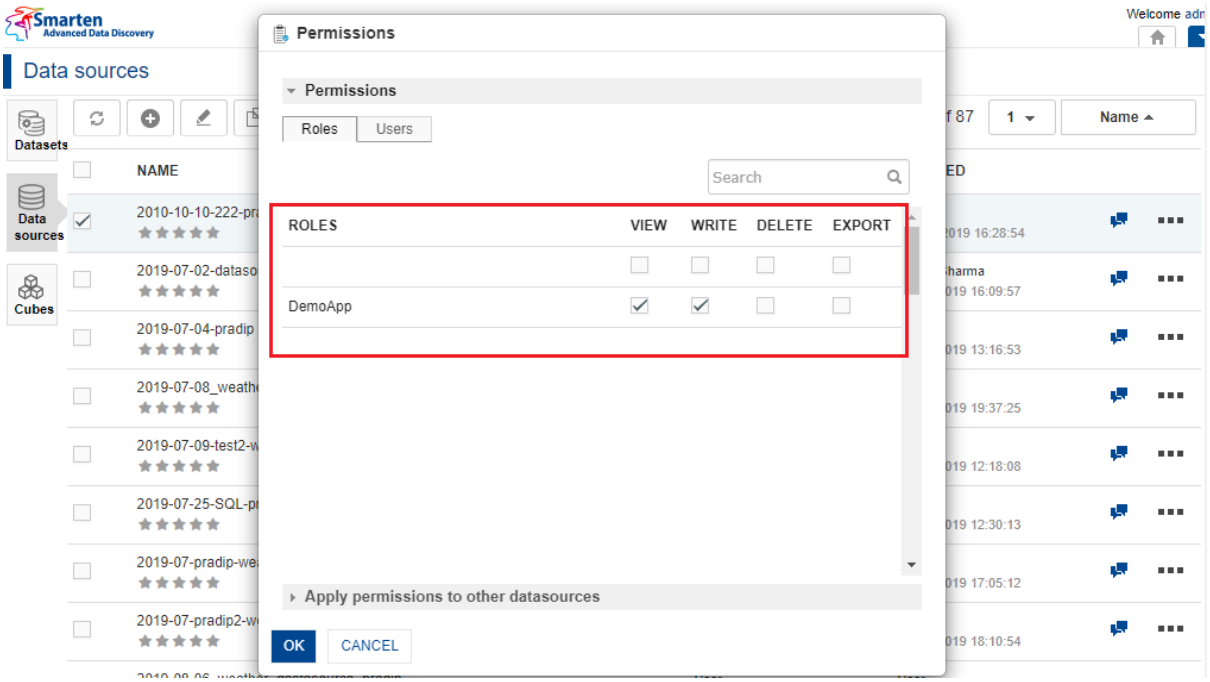
2. Click **Data sources**.
3. Select the check box adjacent to the data source for which you want to manage access rights.
4. Click the Permissions icon.

   The system displays the **Permissions** dialog box.



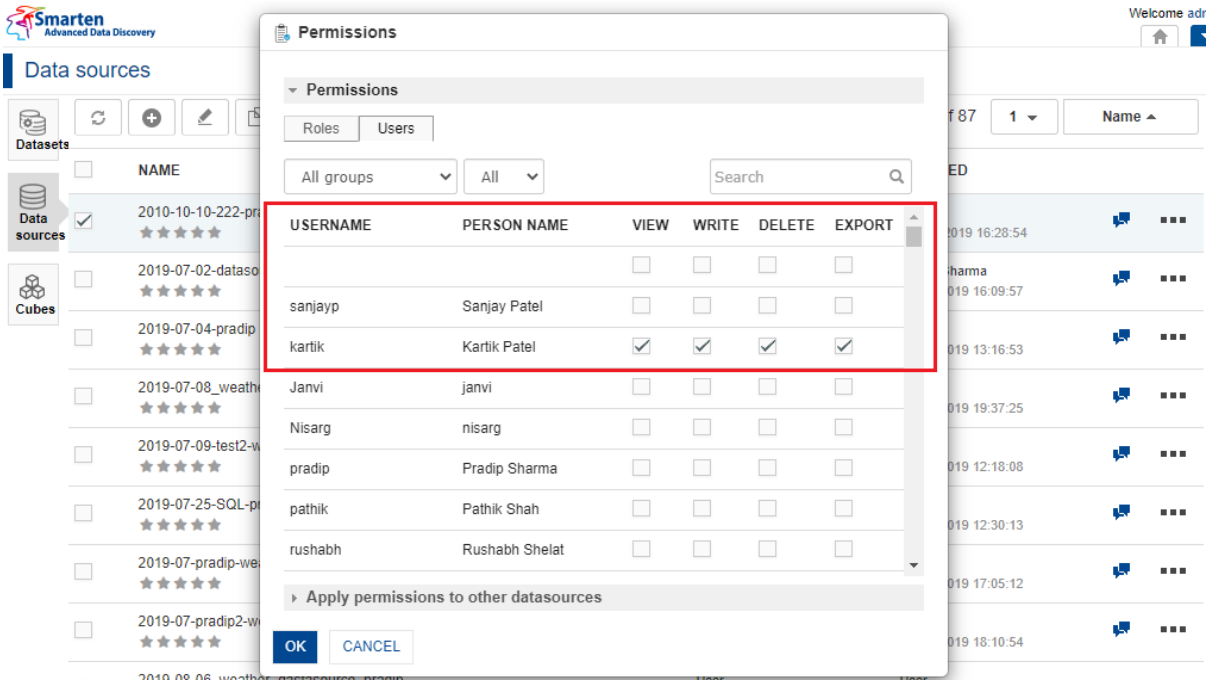
MANAGING ACCESS RIGHTS—CLICKING PERMISSIONS ICON

5. Click the **Roles** tab to assign access rights to various roles.



ASSIGN PERMISSIONS—ACCESS PERMISSIONS FOR ROLES

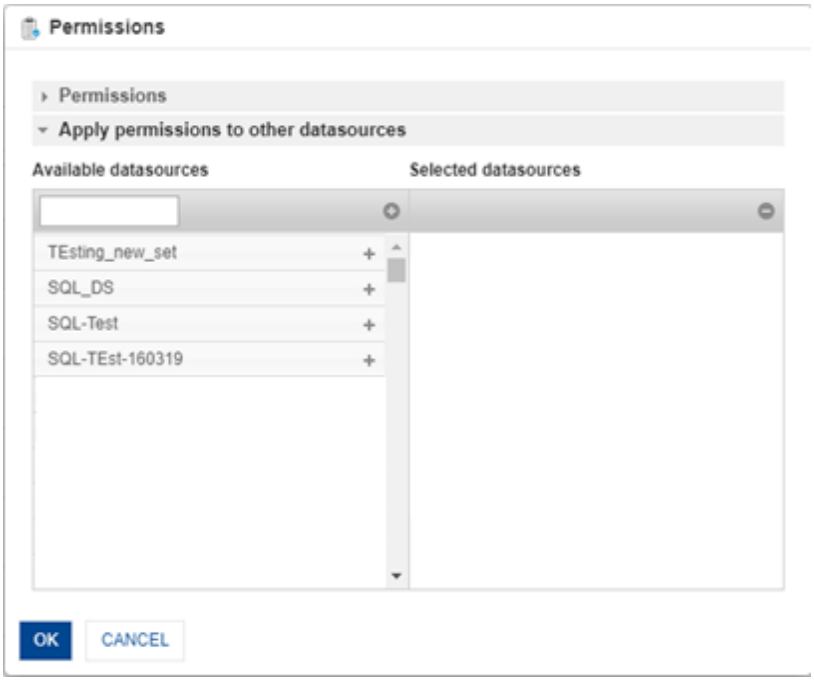
6. Select the check box under the **VIEW**, **WRITE**, **DELETE**, and **EXPORT** permissions columns to assign that access to a role. For example, in the image above, you can select the **VIEW**, **WRITE**, **DELETE,** and **EXPORT** boxes for DemoApp role to allow users with DemoApp role to view, modify, delete, or export the data source.
7. Click the **Users** tab to assign access rights to different users or group of users.



ASSIGN PERMISSIONS—ACCESS PERMISSIONS FOR USERS

8. Select the check box under the **VIEW**, **WRITE**, **DELETE**, and **EXPORT** permissions columns to assign that access to a role. For example, in the image above, you can select the **VIEW**, **WRITE**, **DELETE,** and **EXPORT** boxes to allow “kartik” to view, modify, delete, or export the data source.

9. Click Apply permissions to other datasourcesoption to grant the same permissions to other datasources which you have selectedfor roles and users in the previous step.
This option allows you to grant the same set of permissions you have granted to a role to other data sources instead of granting the same set of permissions to the role for each data source separately. For example, if you have grated view and export permissions to Role 1 and want to grant the same permissions for Datasource1, Datasource2, and Datasource3. You can use the Apply permissions to other datasources option to grant the view and export permissions to Role 1 for Datasource1, Datasource2, and Datasource3.



DATA SOURCE PERMISSION: APPLY PERMISSIONS TO OTHER DATA SOURCE

10. Click the plus sign adjacent to the data sources for which you want to grant the permissions you have granted to the roles in the earlier step.
11. Click **OK** togrant the permissions you have selected for roles and users**.**

5.4 Copying a Data Source

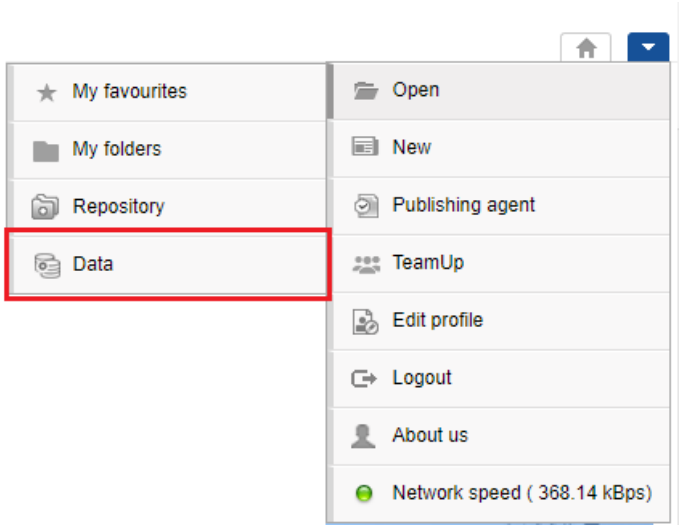
This feature enables authorized users to create a copy of the data source.

About this task

Use this task to copy a data source.

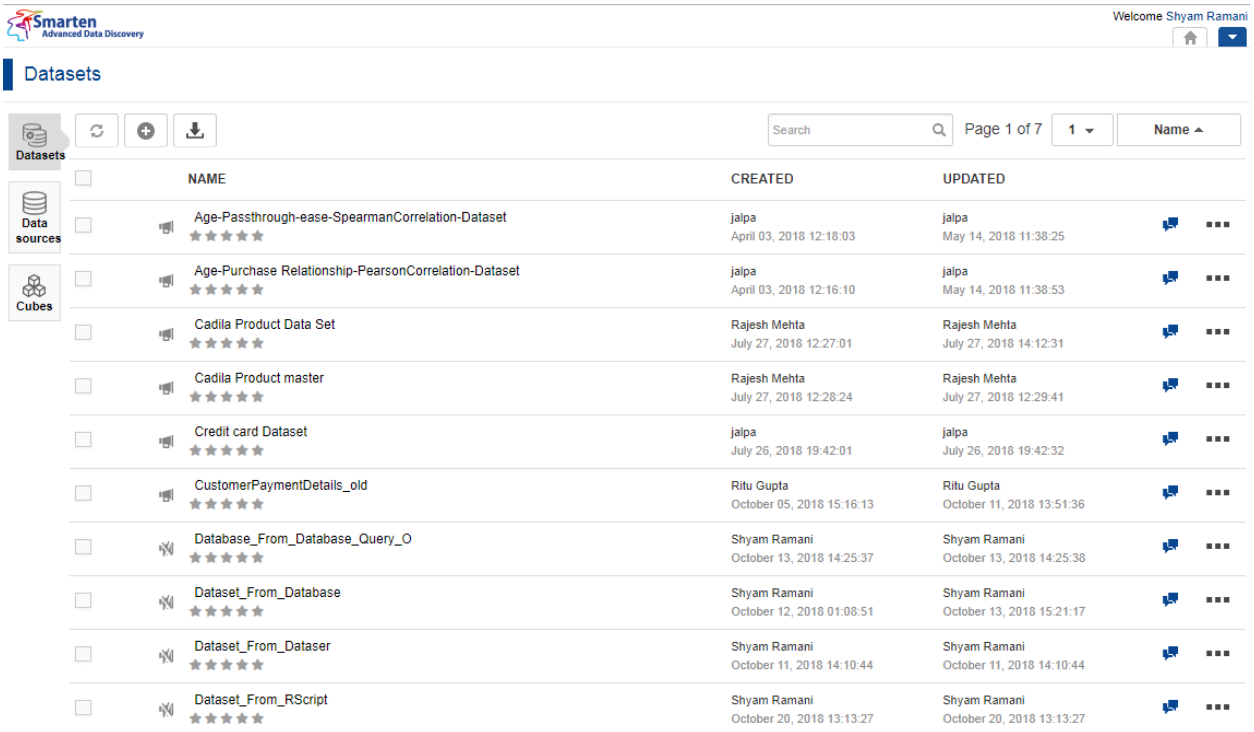
Procedure

1. Click **Open** -> **Data** from the menu.



MENU OPTION—OPEN DATA

The system displays the following page.

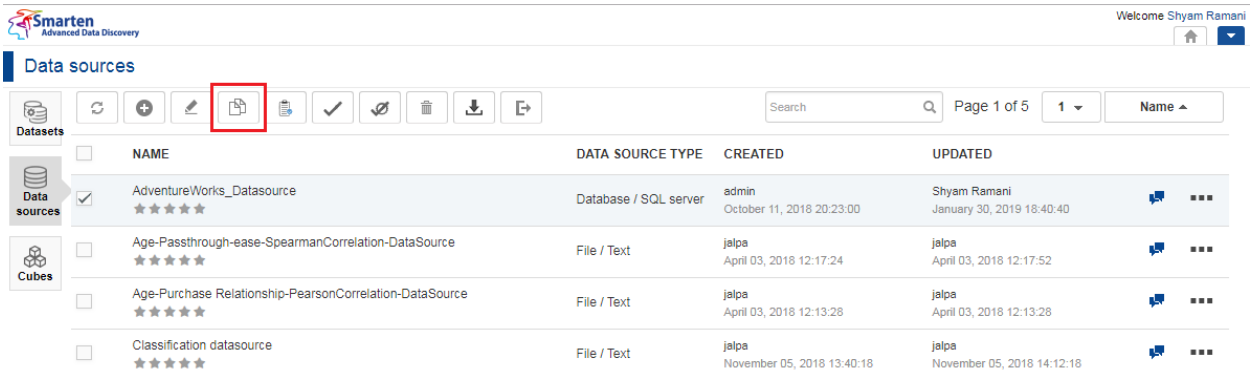


ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Select the check box adjacent to the data source you want to copy.

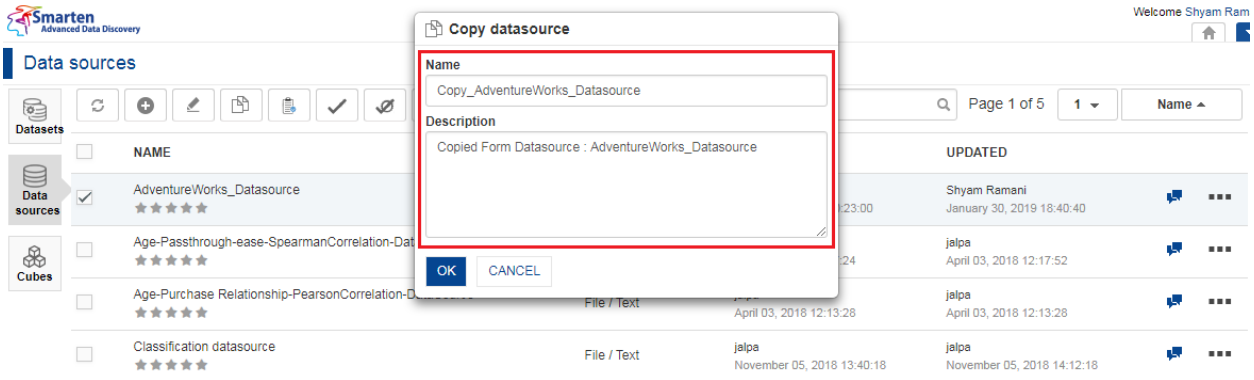
4. Click the Copy icon.

The system displays the **Copy datasource** dialog box.



COPYING A DATA SOURCE—CLICKING THE COPY ICON

5. Specify a name and description for the new data source to be created using the data source you copied in the previous step.



COPYING A DATA SOURCE—PROVIDING A NEW NAME AND DESCRIPTION

6. Click **OK**.

5.5 Exporting a Data Source

You can export a data source definition in XML format. The exported XML file can be imported in the same or another instance of Smarten. Data source export does not include data source permissions and data.

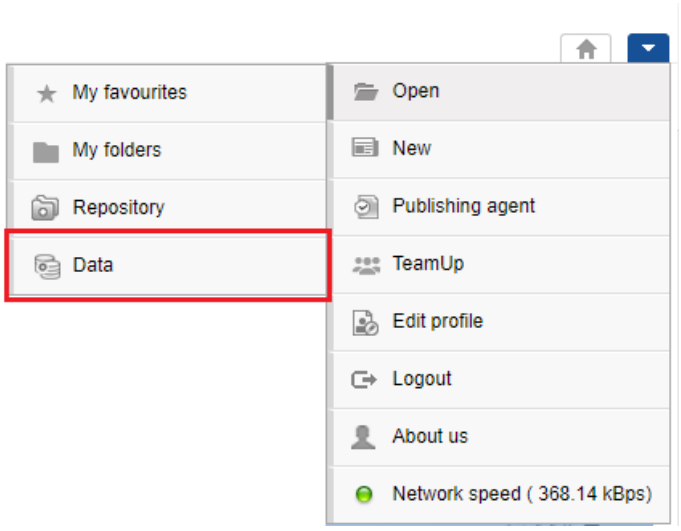
You can use the data source export feature to make a backup of the data source, and you can restore a data source by importing the XML file of that data source.

About this task

Use this task to export a data source.

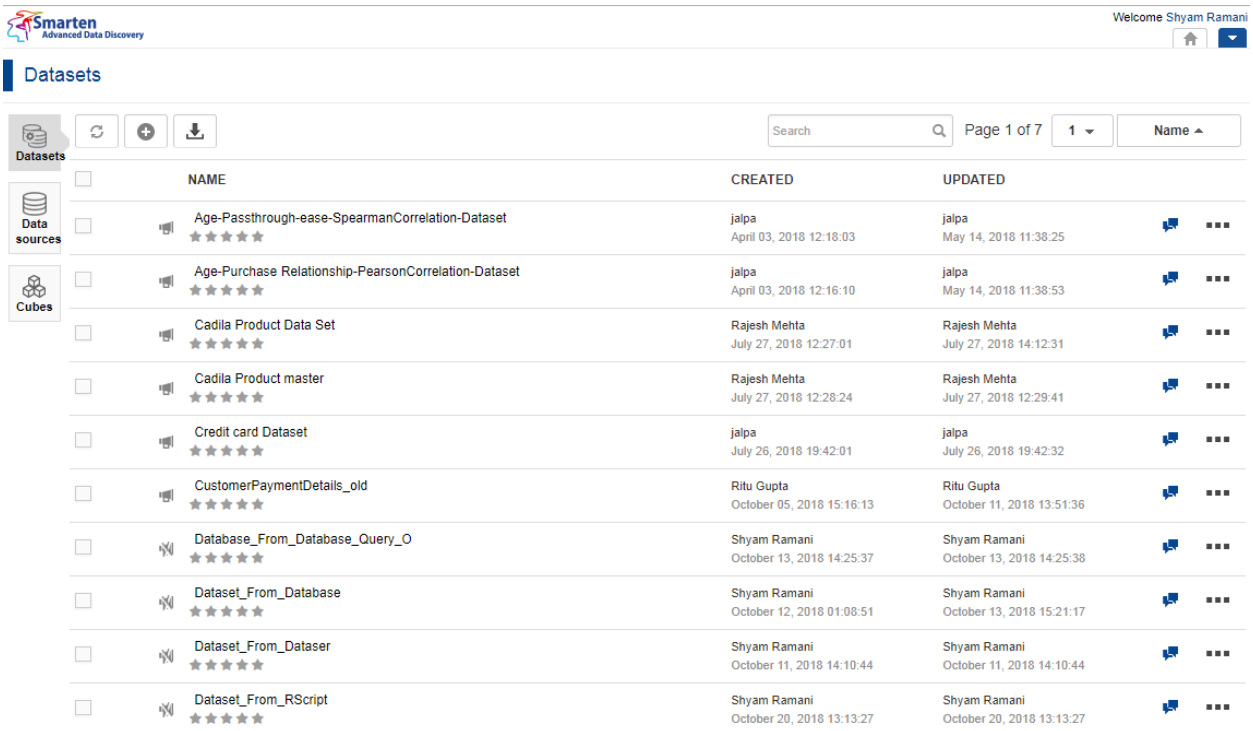
Procedure

1. Click **Open** -> **Data** from the menu.



MENU OPTION—OPEN DATA

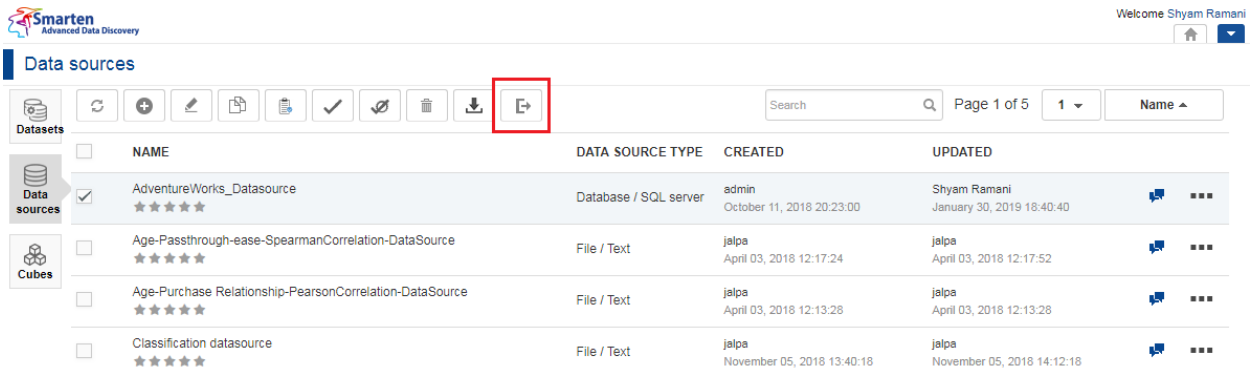
The system displays the following page.



ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Select the check box adjacent to the data source you want to export.
4. Click the Export icon.

   The system downloads the data source in .xml format.



EXPORTING A DATA SOURCE—CLICKING THE EXPORT ICON

5.6 Importing a Data Source

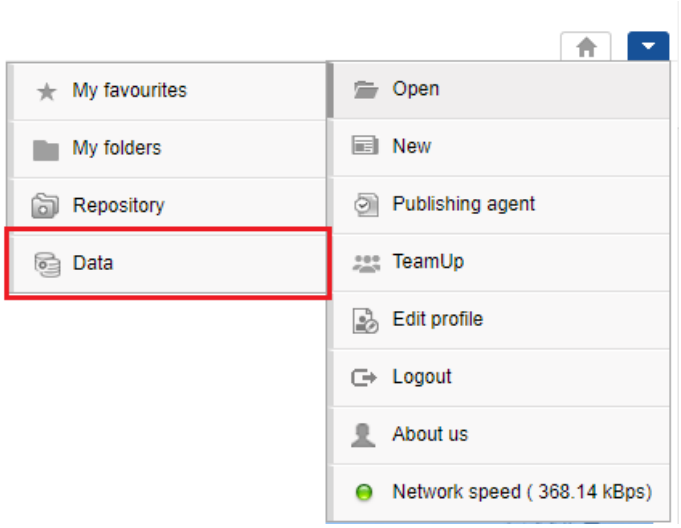
You can import an exported data source in XML format from the same instance or another instance of Smarten.

About this task

Use this task to import a data source.

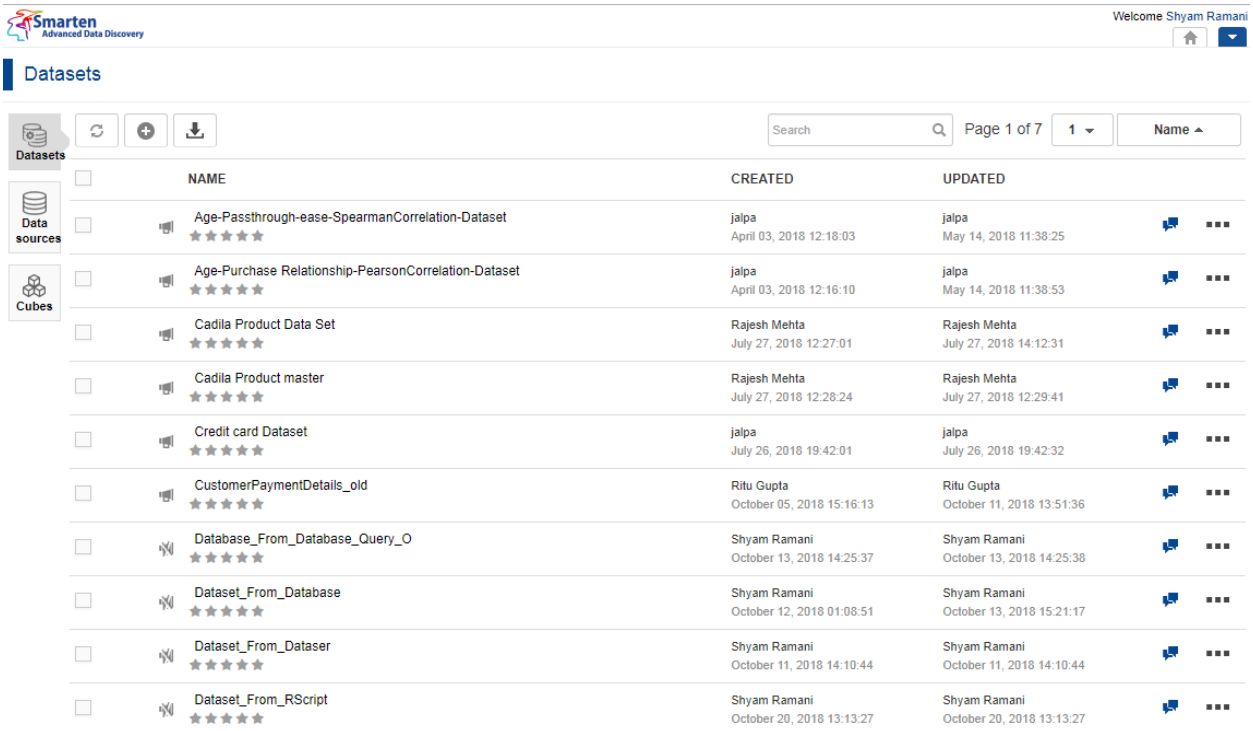
Procedure

1. Click **Open** -> **Data** from the menu.



MENU OPTION—OPEN DATA

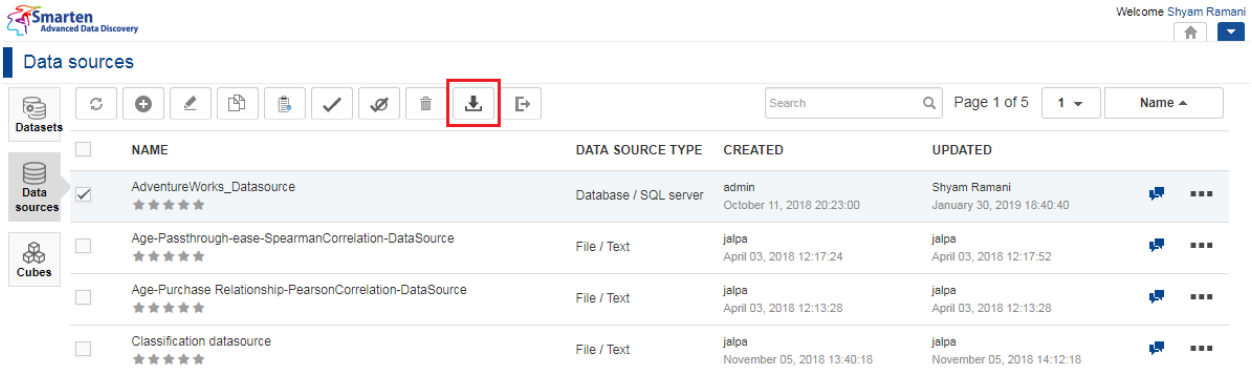
The system displays the following page.



ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Click the Import icon.

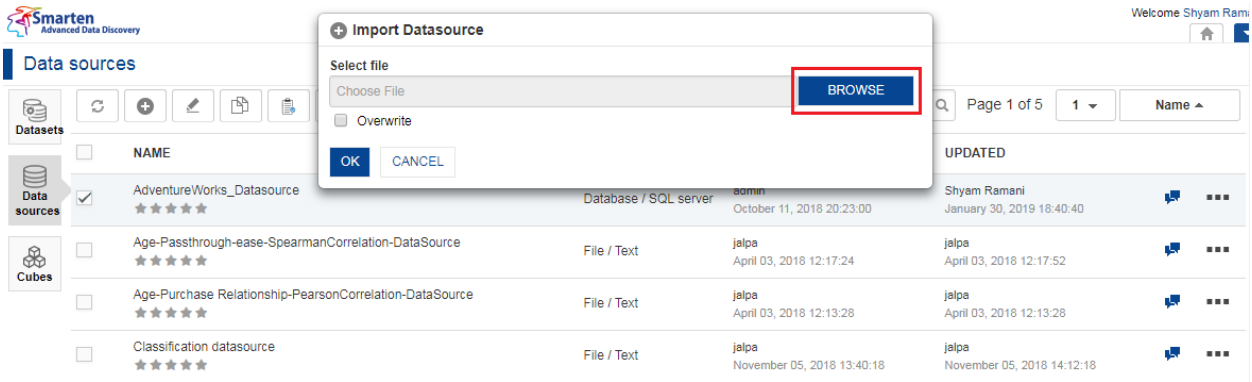
The system displays the **Import Datasource** dialog box.



IMPORTING A DATA SOURCE—CLICKING THE IMPORT ICON

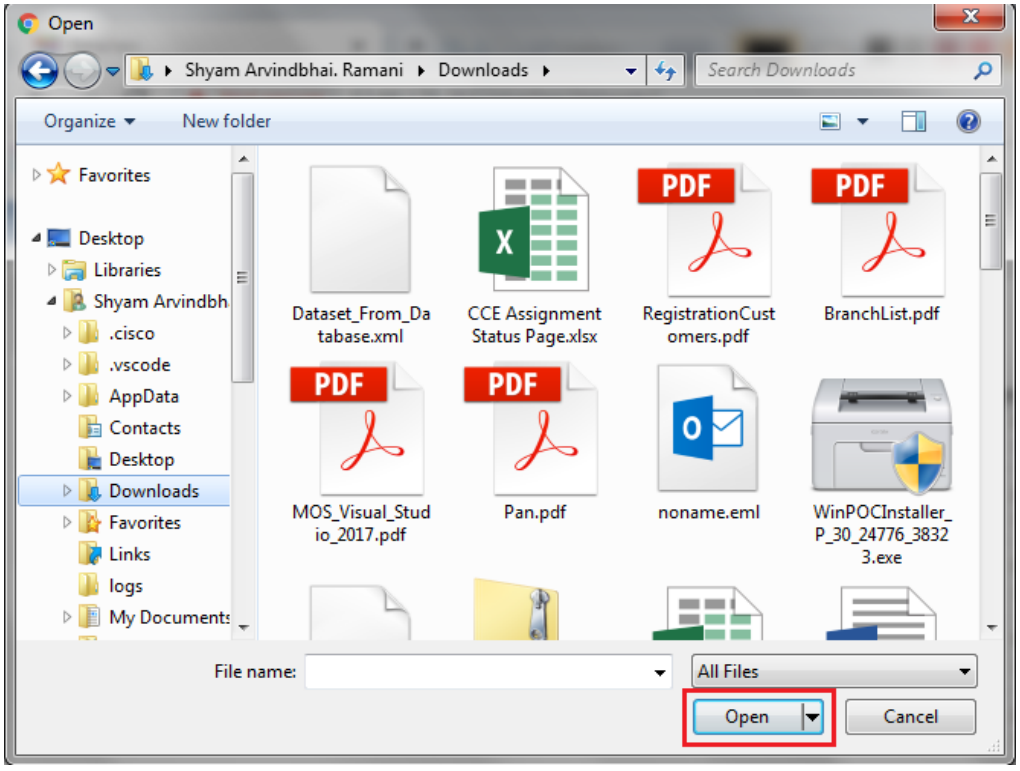
4. Click the **BROWSE** button.

The system displays the **Open** dialog box.



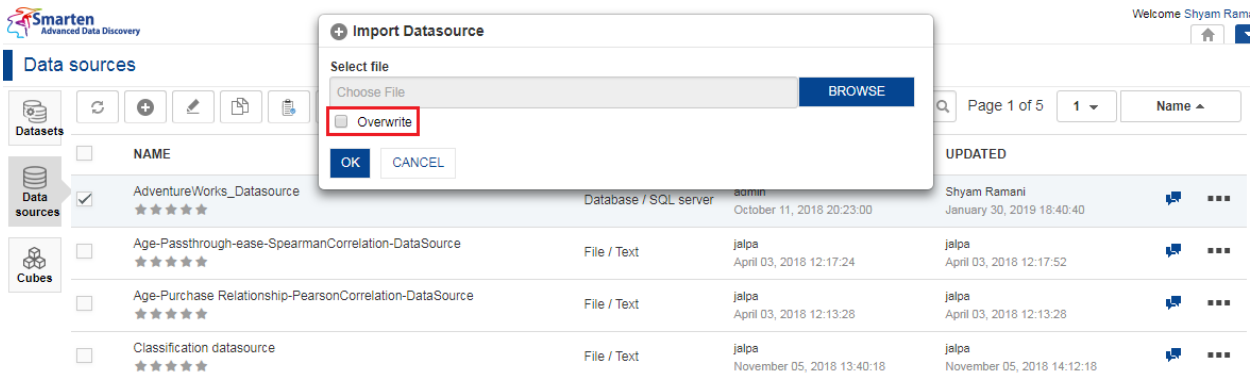
IMPORTING A DATA SOURCE—OPENING THE OPEN DIALOG BOX

5. Select the XML file for the data source that has been exported, and then click **Open**.



IMPORTING A DATA SOURCE—THE OPEN DIALOG BOX

6. If a data source with the same name as the imported data source exists in the system, you can select the **Overwrite** check box to allow the system to overwrite the data source in the system with the imported data source.



IMPORTING A DATA SOURCE—SELECTING THE OVERWRITE OPTION

If the **Overwrite** option is not selected, the system retains the existing data source and adds the imported data source with a new name. The new name of the data source is appended with a numeric value in increasing order. For example, if you import ABC data source and a data source with the same name already exists, the system will rename the imported data source as ABC_1. Similarly, if a data source named ABC_1 already exists, the system will rename the imported data source as ABC_2.

7. Click **OK**.

5.7 Marking a Data Source as IT Approved

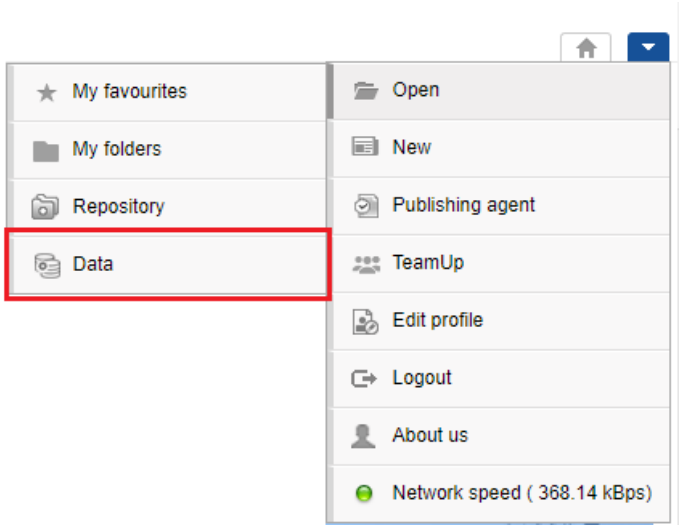
This feature enables IT staff to approve a data source. Marking a data source as IT approved certifies it for data quality and helps users in identifying quality data sources.

About this task

Use this task to mark a data source as IT approved.

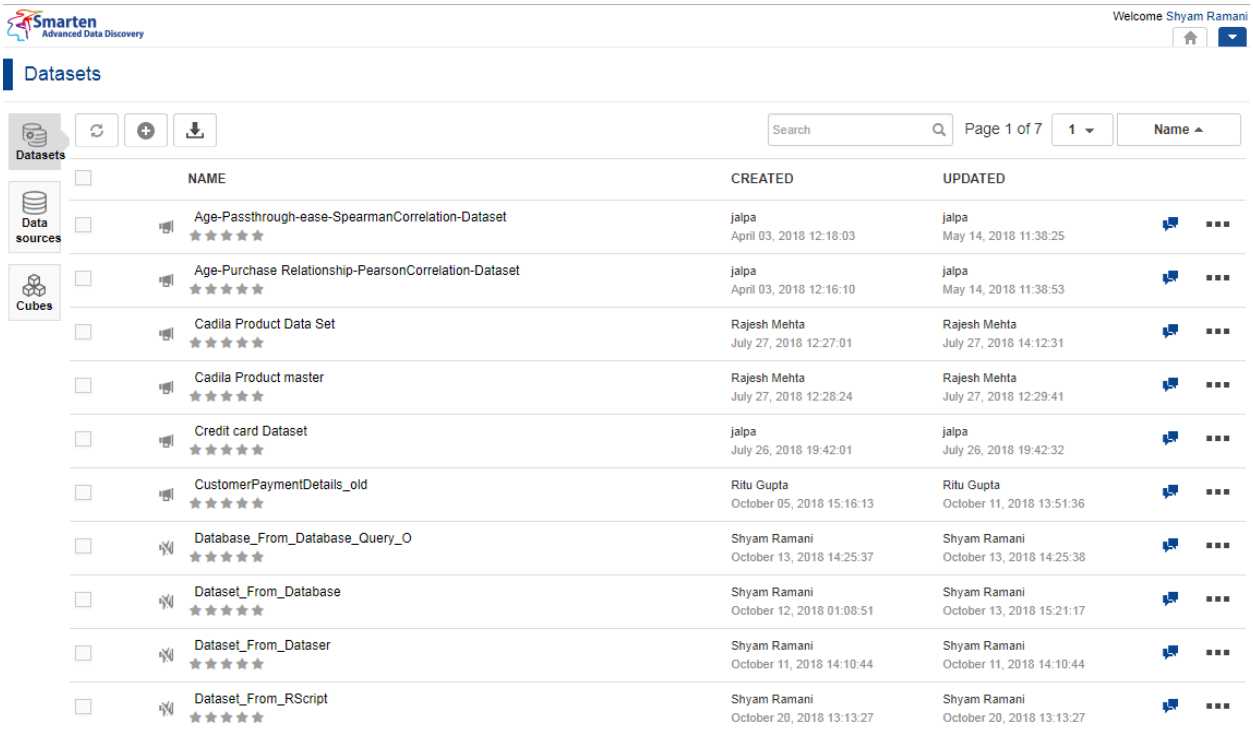
Procedure

1. Click **Open** -> **Data** from the menu.



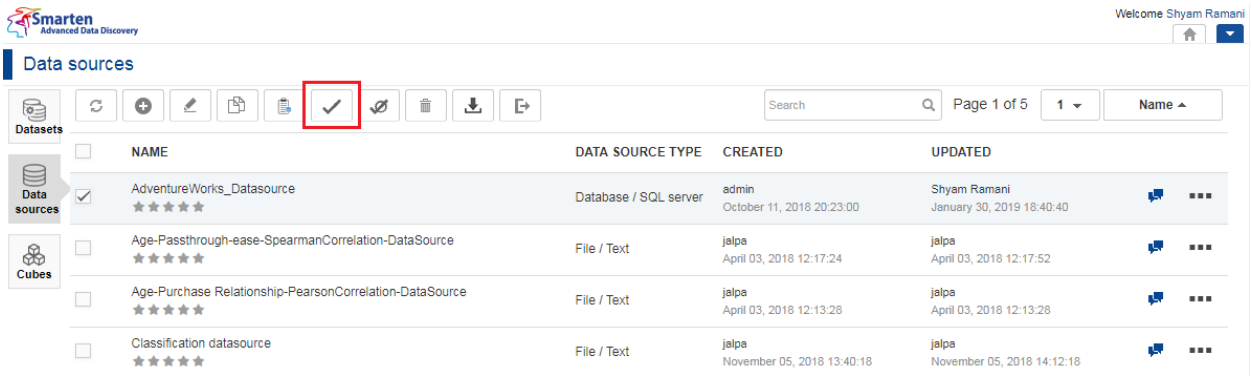
MENU OPTION—OPEN DATA

The system displays the following page.



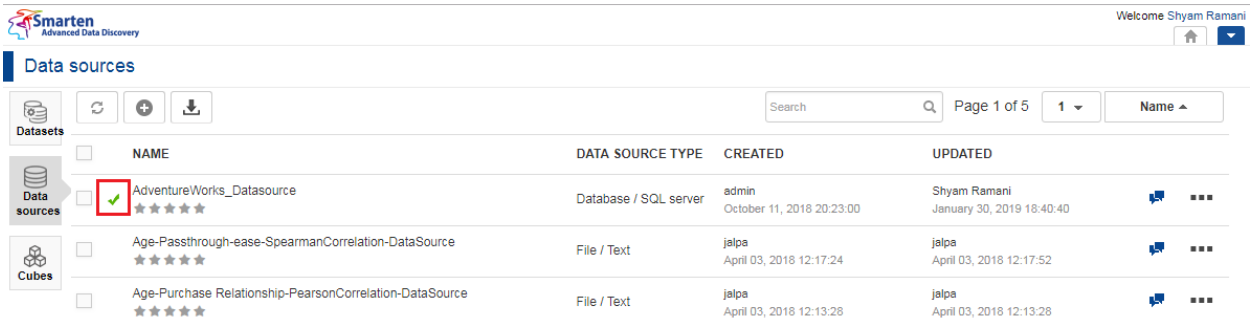
ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Select the check box adjacent to the data source you want to mark as IT approved.
4. Click the Mark IT Approved icon.



MARKING A DATA SOURCE—CLICKING THE MARK IT APPROVED ICON

The system marks the data source as IT approved and displays a check symbol adjacent to the data source.



MARKING A DATA SOURCE—THE CHECK MARK INDICATING THAT THE DATA SOURCE IS IT APPROVED

5.8 Unmarking a Data Source as IT Approved

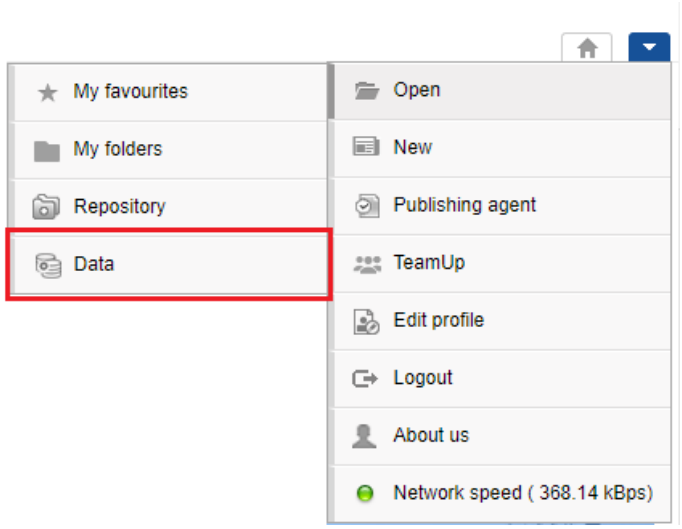
This feature enables IT staff to unmark the data sources that are marked as approved.

About this task

Use this task to unmark a data source that is marked as IT approved.

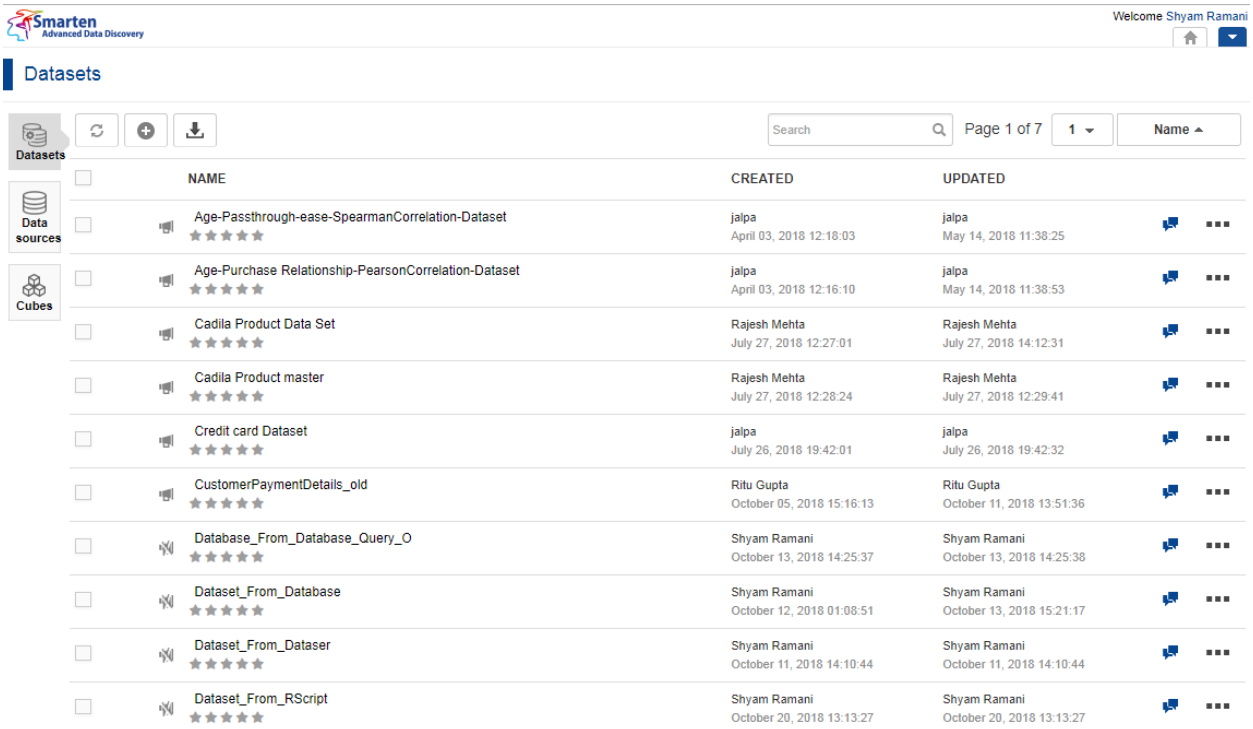
Procedure

1. Click **Open** -> **Data** from the menu.



MENU OPTION—OPEN DATA

The system displays the following page.

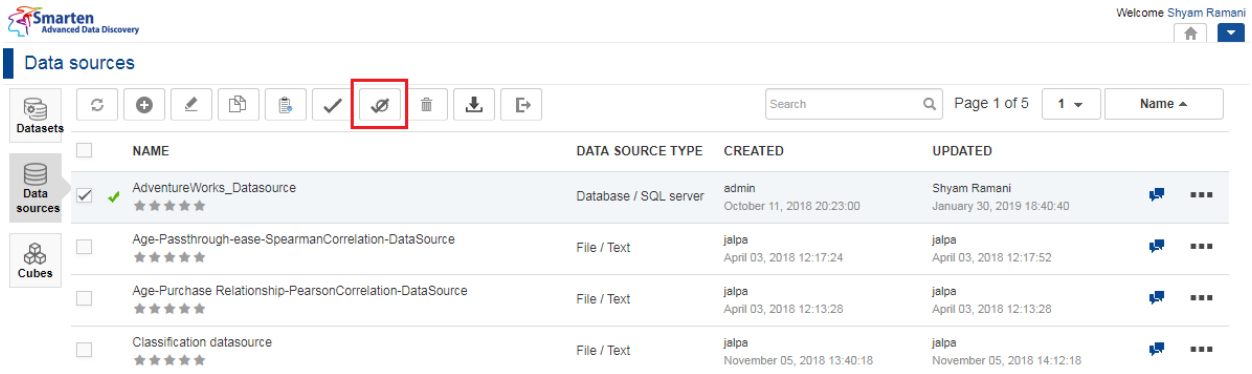


ACCESS A DATA SOURCE—DISPLAYING DATA SOURCES

2. Click **Data sources**.
3. Select the check box adjacent to the data source you want to unmark.

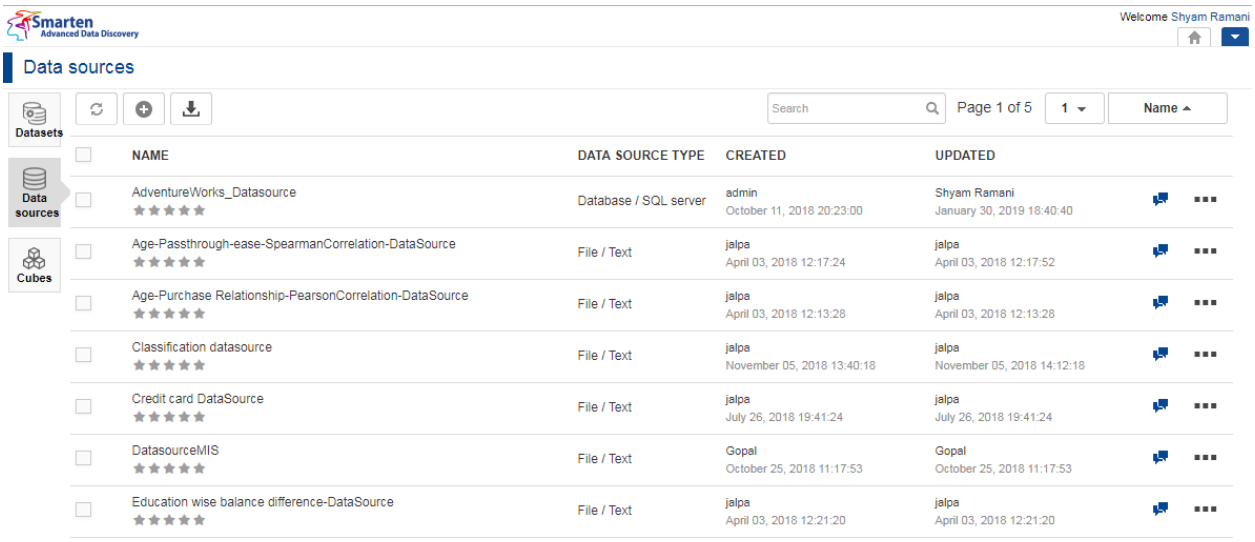
> Note:
> You can only unmark the data sources that are marked as IT approved.

4. Click the Unmark IT Approved icon.



UNMARKING A DATA SOURCE—CLICKING THE UNMARK IT APPROVED ICON

The system unmarks the data source as IT approved, and the check symbol adjacent to the data source is no longer available.



UNMARKING A DATA SOURCE—THE CHECK MARK FOR IT CERTIFIED IS REMOVED

6 Creating a Dataset

A dataset is a source of analysis-ready data in a columnar structure. The process of creating a dataset includes connecting to a data source and extracting raw data from it based on the columns selected.

> Note:
> You can create a dataset from an existing data source or a dataset.

Reference: **Concept Manual > Dataset > Creating Dataset**

6.1 Creating a Dataset Using a Database Profile

You can use a database profile to create a dataset and retrieve data from that database profile. You can retrieve data using either a ready to use query or using a step-by-step wizard.

6.1.1 Creating a Dataset Using a Ready to Use Query

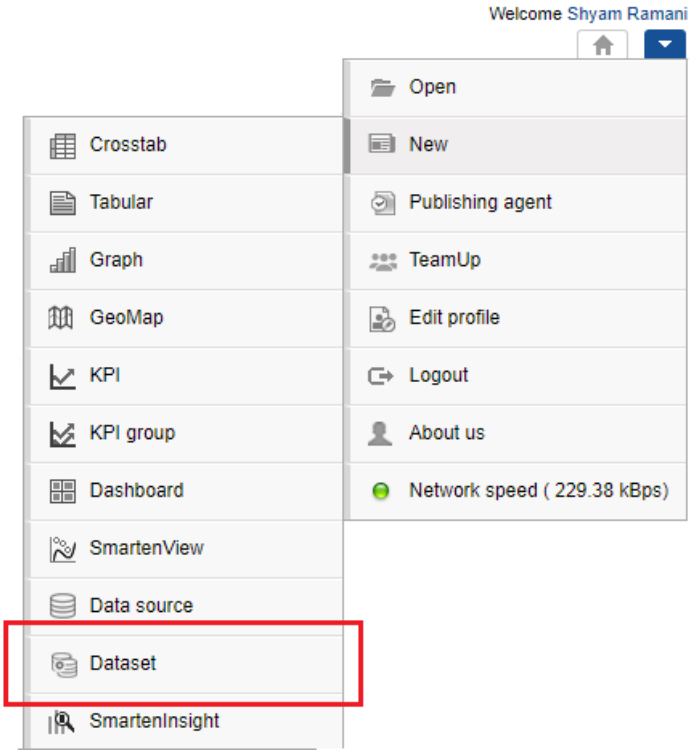
You can use a query to retrieve data from a database profile and use that data to create a dataset.

About this task

Use this task to create a dataset using a query to retrieve data from a database profile.

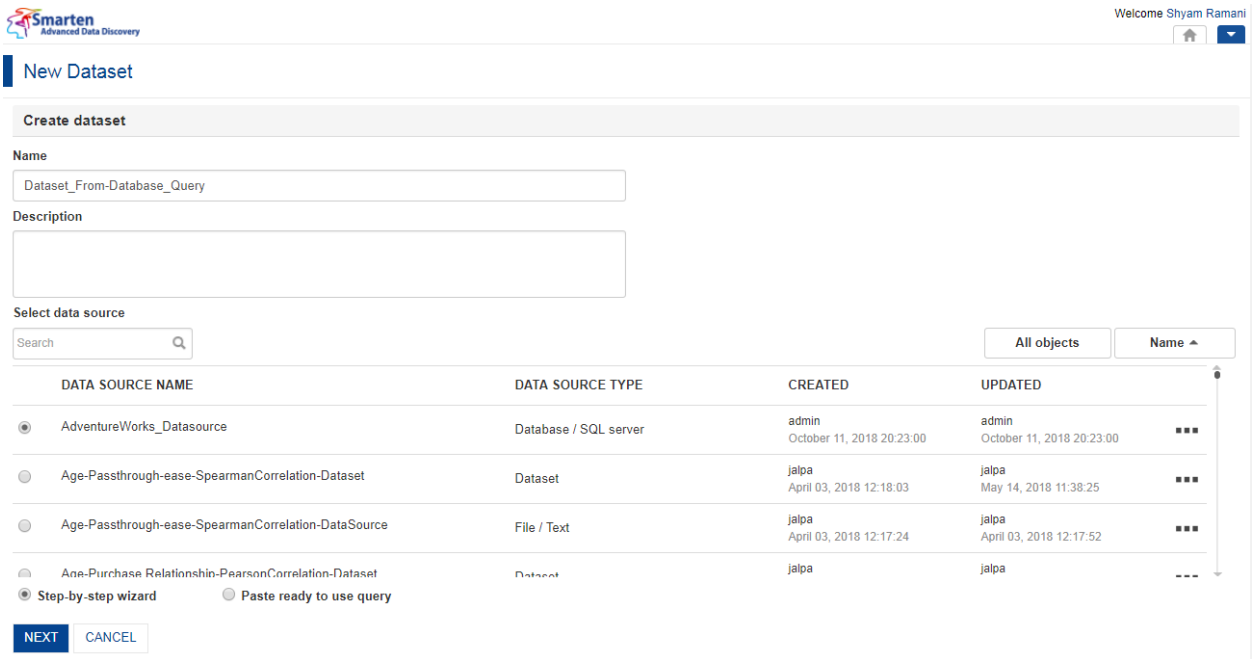
Procedure

1. Click **New** -> **Dataset** from the menu.



MENU OPTION—NEW DATASET

The system displays the **Create dataset** page.`



CREATING A DATASET—THE CREATE DATASET SCREEN

2. Enter a name and description for the dataset in the **Name** and **Description** fields.
3. Select the database profile you want to use to create a dataset from the list.

   Or,

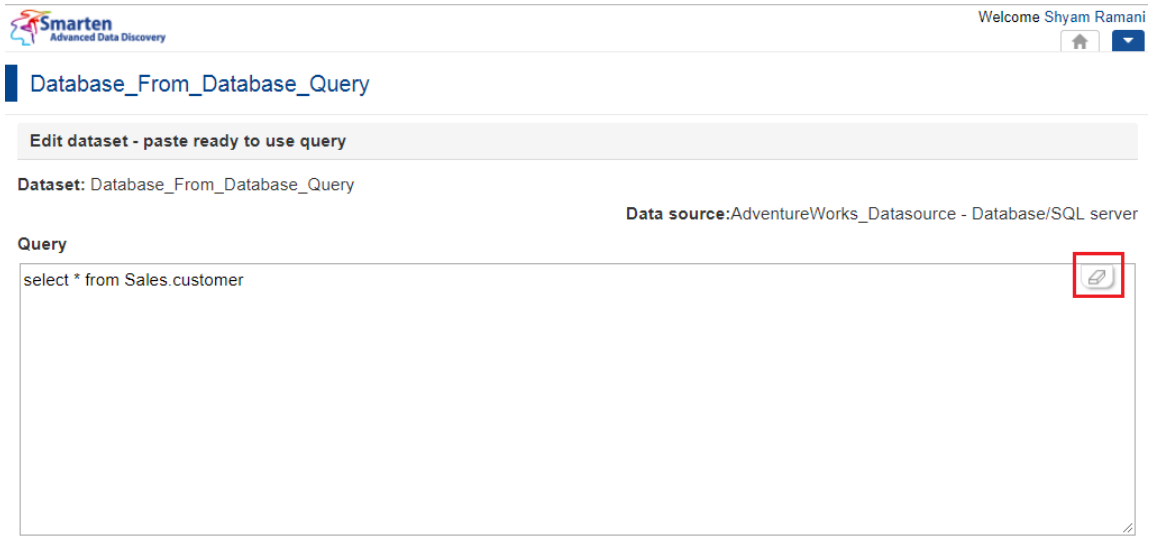
   Enter the keyword to search for the database data source profile based on the keyword provided, and then press **Enter**.

The system displays the database data source profiles that match the keyword you have provided and displays the **Step-by-step wizard** and **Paste ready to use query** options.

4. Select the **Paste ready to use query** option to paste a query that you want to use to retrieve data from the database profile.
5. Click **Next**.
   The system displays the **Paste ready to use query** screen.
6. Paste the query you want to use in the **Query** field.

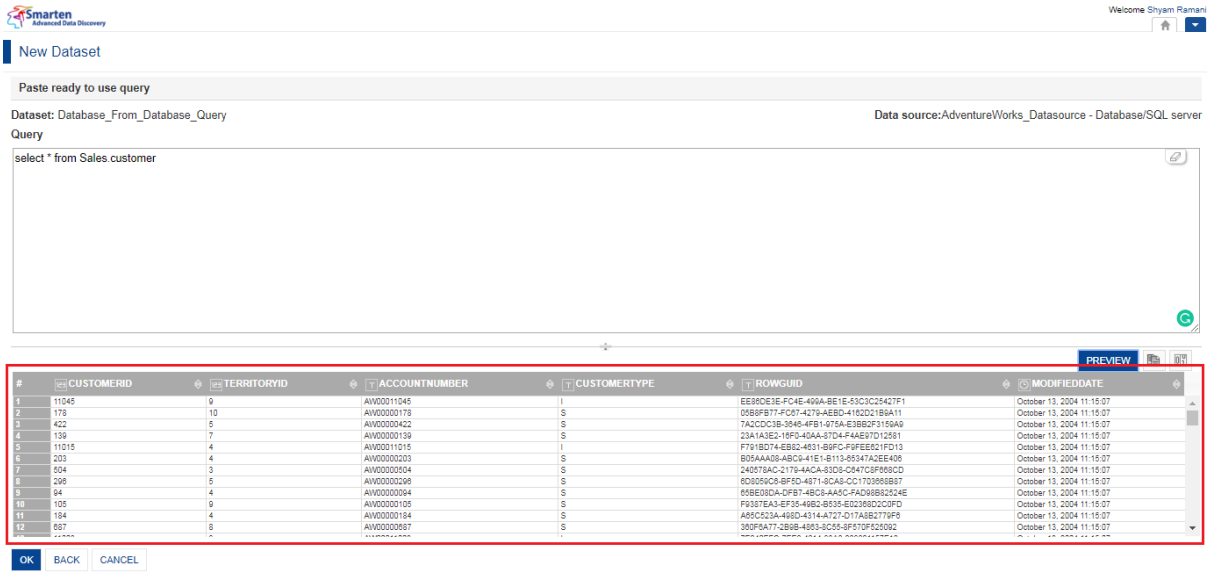
CREATING A NEW DATASET—USING A QUERY TO RETRIEVE DATA FROM A DATABASE PROFILE

7. You can click the Clear Query button to reset the query.

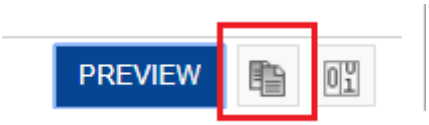


8. Click **PREVIEW**.

   The system displays the data retrieved from the database based on the query you have specified.

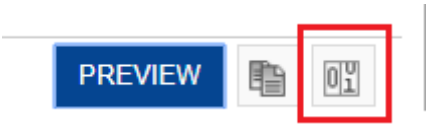


9. You can click the full data mode button to retrieve and display the entire data of a result set. By default, the system displays limited records in the preview mode.



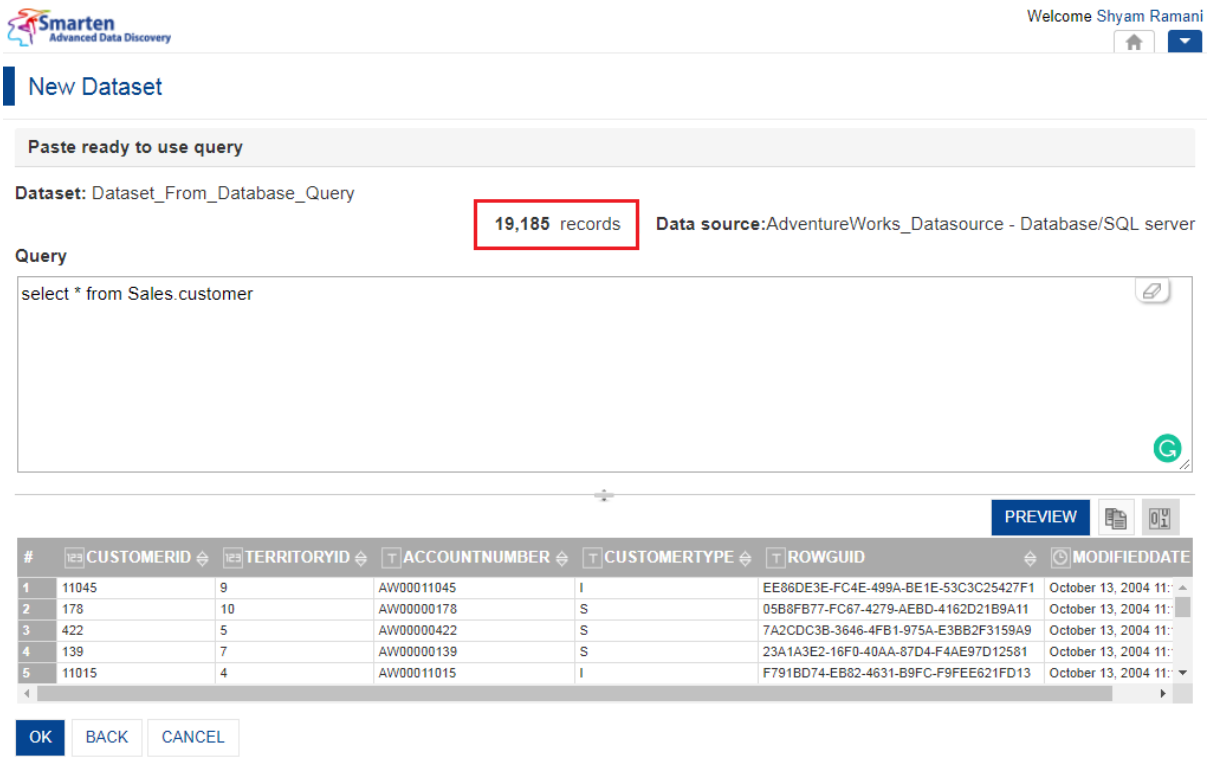
PREVIEW DATA—FULL DATA MODE OPTION

10. You can click the option to enable a record count.



PREVIEW DATA—RECORD COUNT OPTION

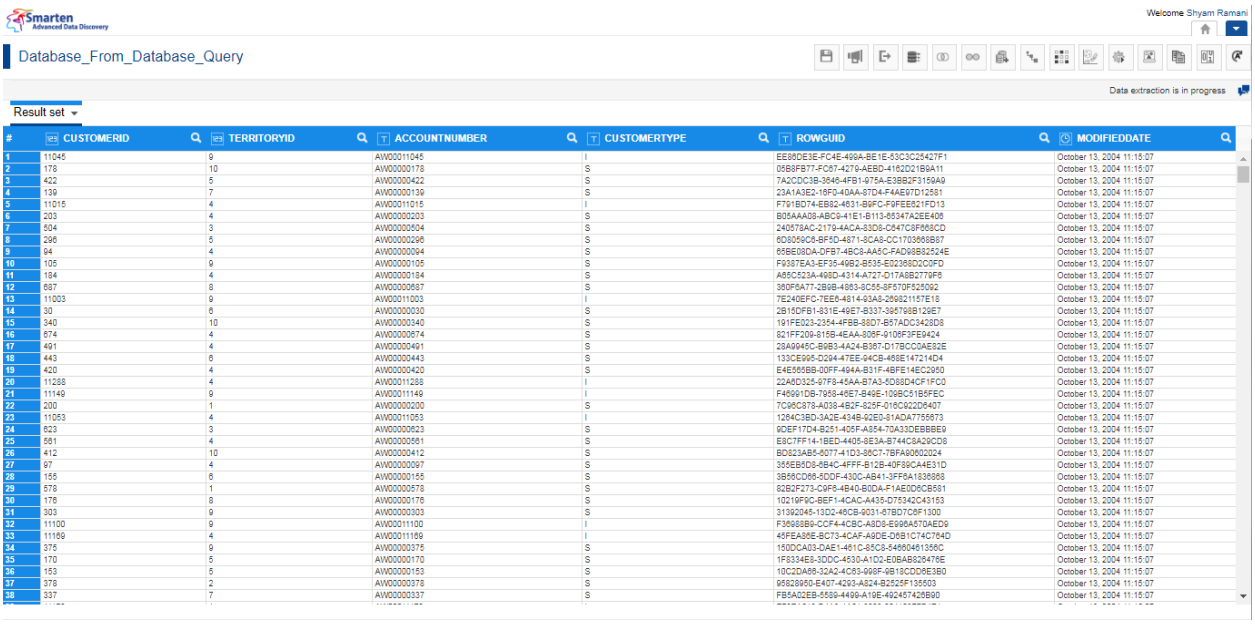
When enabled, this option displays the total number of records available.



PREVIEW DATA—RECORD COUNT ENABLED

11. Click **OK**.

The system retrieves the data from the database and creates the dataset.



PREVIEW—DATASET CREATED USING A DATABASE PROFILE

The dataset is now available in the repository.

6.1.2 Creating a Dataset Using the Step-by-Step Wizard

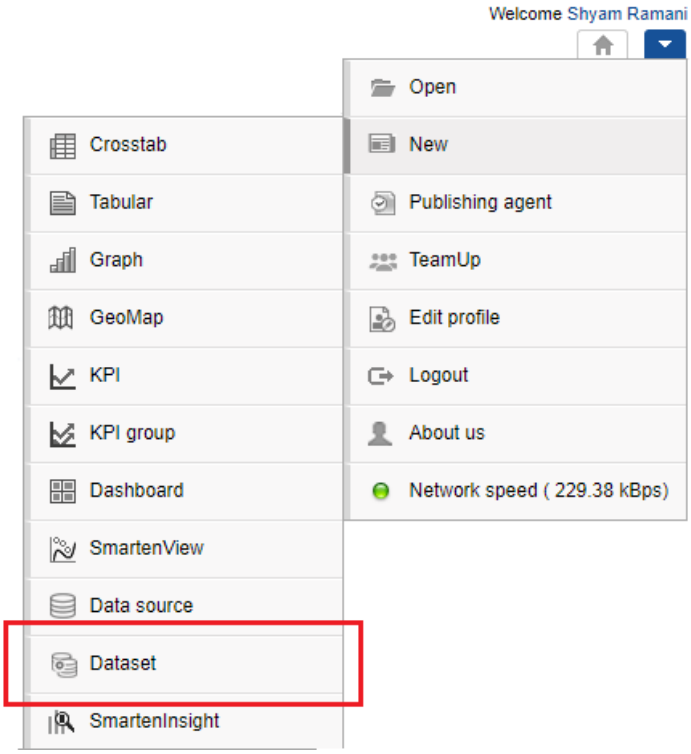
You can create a dataset using the step-by-step wizard that allows you to choose the tables, apply joins, manage columns, and filter data.

About this task

Use this task to create a dataset using the step-by-step wizard to retrieve data from a database profile.

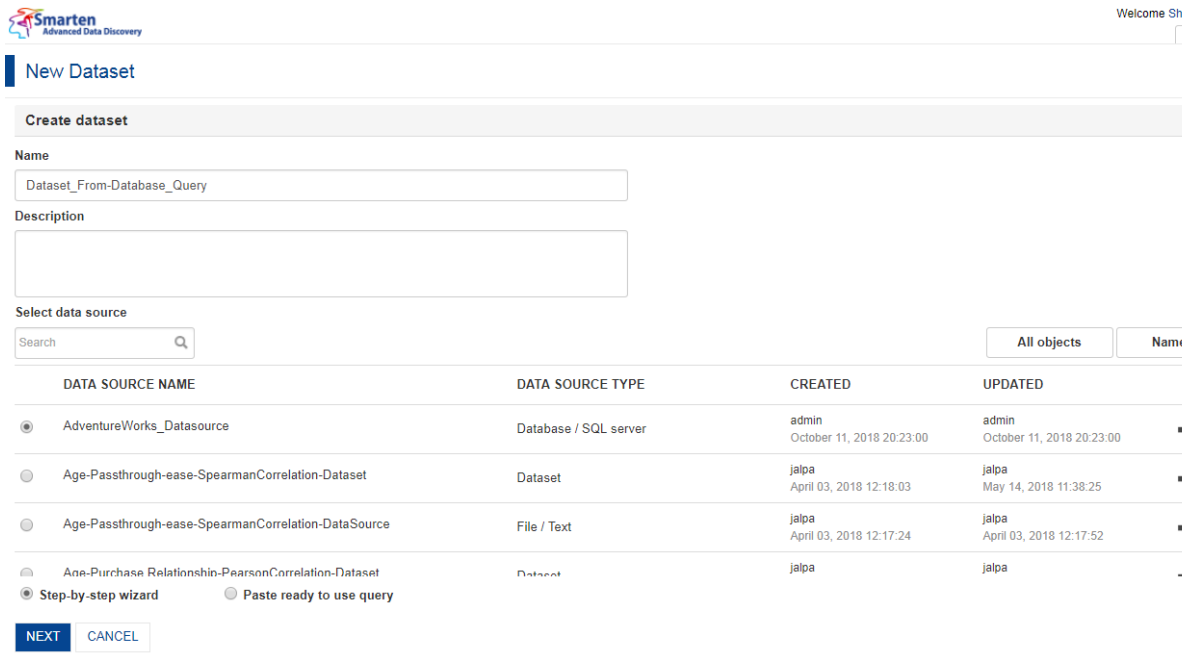
Procedure

1. Click **New** -> **Dataset** from the menu.



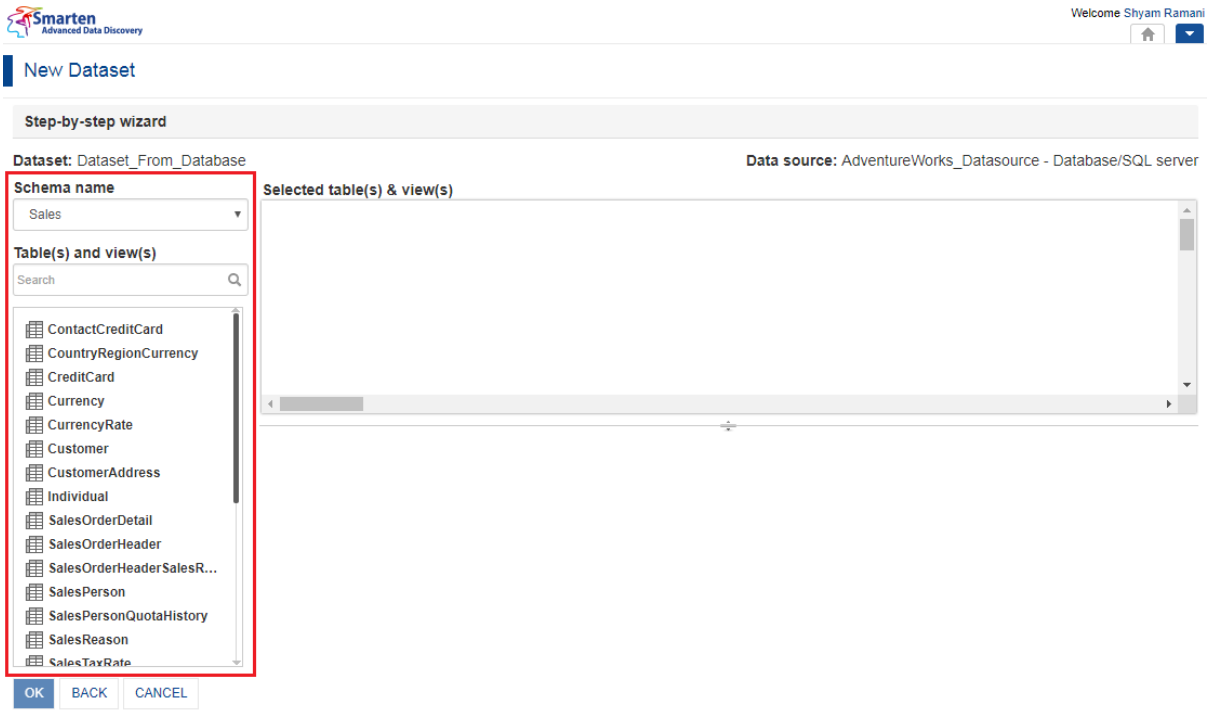
MENU OPTION—NEW DATASET

The system displays the **Create dataset** page.



CREATING A DATASET—THE CREATE DATASET SCREEN

2. Enter a name and description for the dataset in the **Name** and **Description** fields.
3. Select the database profile you want to use to create a dataset from the list.
   Or,
   Enter the keyword to search for the database data source profile based on the keyword provided, and then press **Enter**.
   The system displays the database data source profiles that match the keyword you have provided.
4. Select the **Step-by-step wizard** option to manually select schema, table, and apply joins to retrieve data from the database profile.
5. Click **Next**.
   The system displays the **Step-by-step wizard** screen.
6. Select the schema from which you want to retrieve data from the **Schema name** list.
   The system displays the tables and views available for the schema you have selected in the **Table(s) and view(s)** section.



CREATING A DATASET—SELECTING A SCHEMA

7. Double-click the table(s) and view(s) from the **Table(s) and view(s)** section.

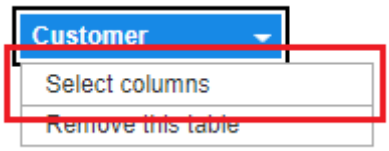
   The system displays the selected table(s) and view(s) in the **Selected table(s) & view(s)** section and retrieves data from these table(s) and view(s). The system displays data available in the result set.

> Note:
> By default, the system displays limited records and limited columns from the result set.

- You can click the full data mode button to retrieve and display entire data of the result set.

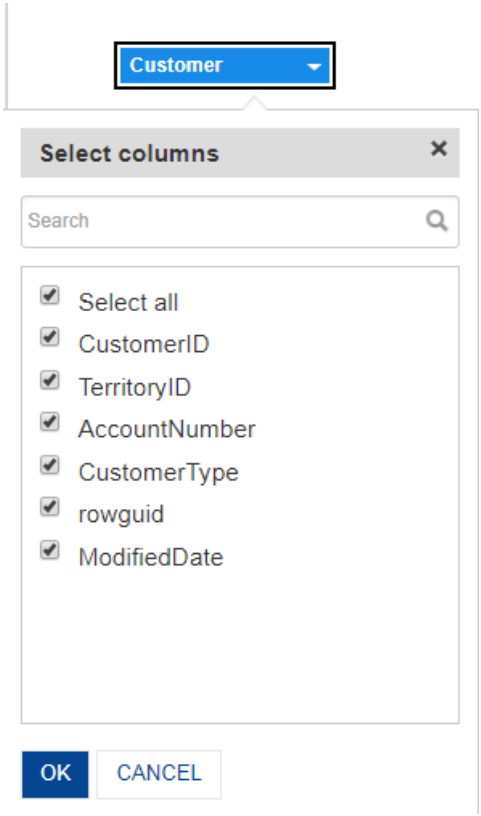
CREATING A DATASET— FULL DATA MODE OPTION

- You can click **Select columns** from the table menu to select or deselect the columns from a table.



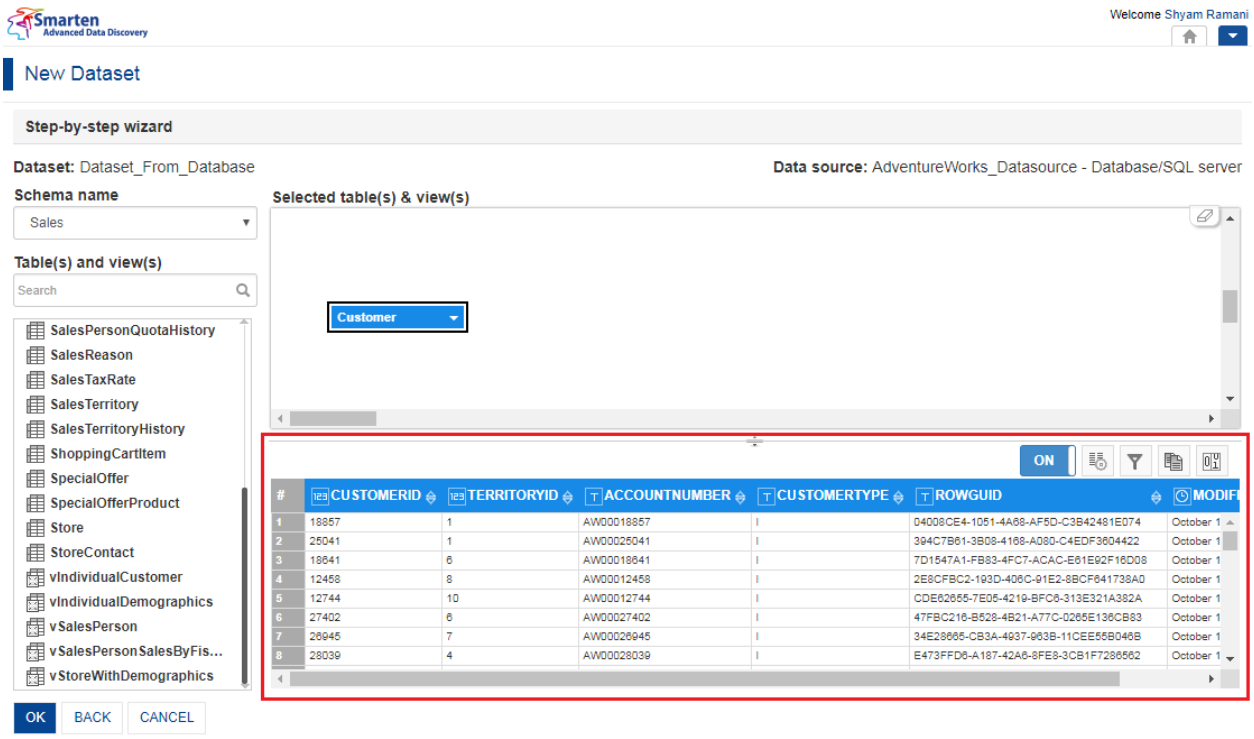
CREATING A DATASET—SELECT COLUMNS OPTION

The system displays the **Select columns** dialog box.



CREATING A DATASET—SELECTING COLUMNS

- Select or deselect the columns, and then click **OK**.

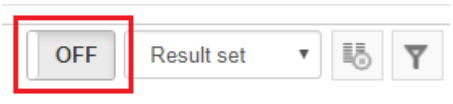


CREATING A DATASET—SELECTING A TABLE OR VIEW

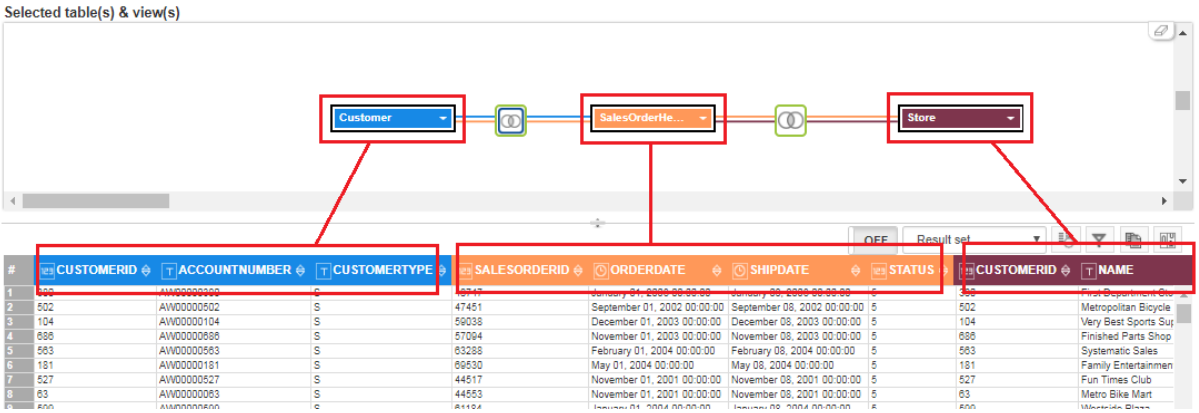
> Note:
> To add a table or view as an alias, you need to use the same table or view twice.

8. Repeat the previous steps to add more than one table or view.
   The system displays the data available in the selected tables or views and performs the following operations:
   - The system automatically applies the join condition if a relationship exists between the table you have added in this step and other table(s) you added in the previous step.

> Note:
> The system does not identify a join automatically if the auto-join option is disabled.



   - If the system cannot find joins in the first scan, it displays the same in a message box and asks whether or not to perform additional scans to identify other possible joins. Once you provide your confirmation, the system performs a detailed scan to identify other joins.
   - If the system cannot identify a join automatically, you have to manually apply to join between those tables.
   - In the preview pane, the system differentiates the data belonging to various tables by providing different background color to the column header. The background color of the column header is the same as the table the columns belong to. For example, in the image below, the Customer table has a blue background, and the Store table has a brown background. In the preview pane, the data for the Customer table has a column header with a blue background, and the Store table has a column header with a brown background.



CREATING A DATASET—COLUMN HEADER WITH SAME BACKGROUND COLOR AS TABLE

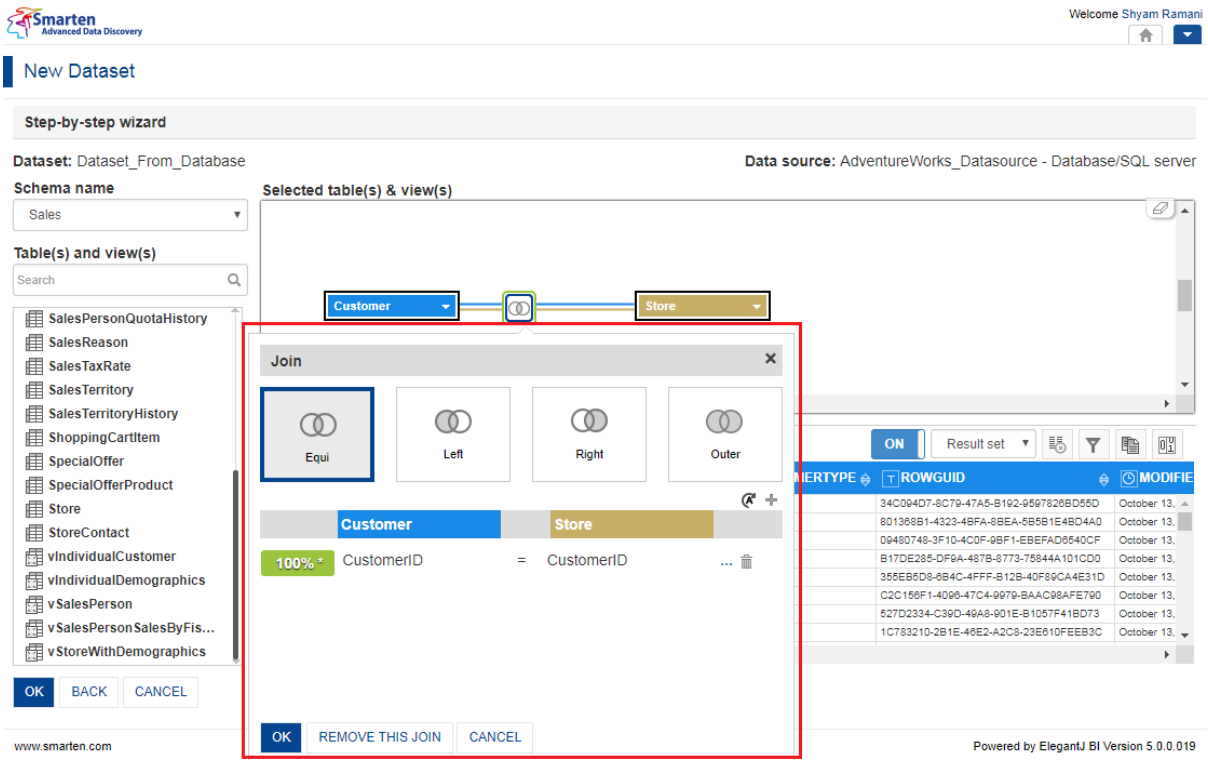
> Note:
> The system applies the equijoin automatically if the option to apply auto-join is enabled.

The system indicates the following characteristic for the join:

CREATING A DATASET—CHARACTERISTICS OF A JOIN

- The innermost symbol indicates the type of join between two tables or views. By default, equijoin is applied.
- The blue border encompassing the innermost symbol indicates that the join is applied automatically.
- The outermost border indicates the identical records between the tables and views in percentage. The border is displayed in the following colors:
    - **Green**: Indicates that more than 80% of records are identical between the two tables or views.
    - **Orange**: Indicates that more than 50% and less than 80% of records are identical between the two tables or views
    - **Red**: Indicates that less than 50% of records are identical between the two tables or views.

9. Click the join symbol between the tables or views to modify that join.
   The system displays the Join dialog box.



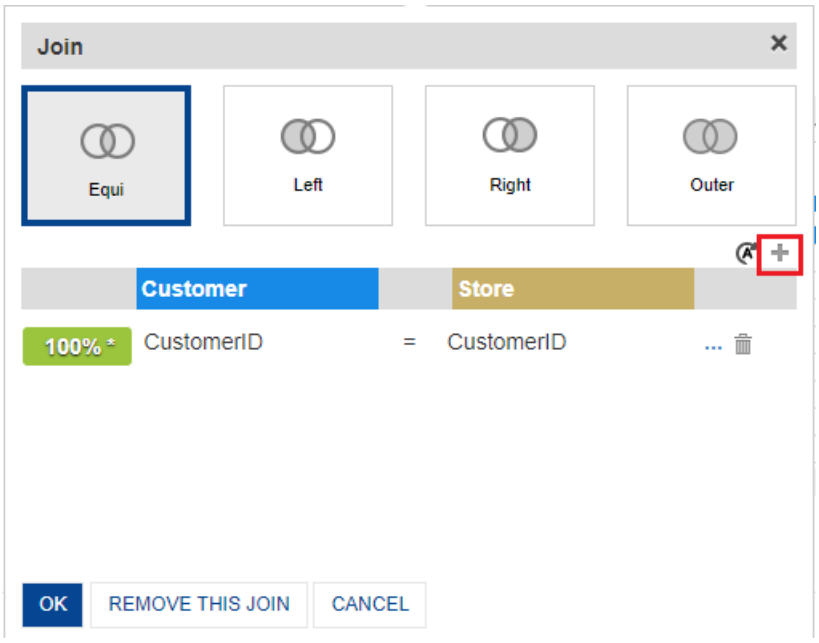
CREATING A DATASET—MODIFYING A JOIN BETWEEN TWO TABLES OR VIEWS

10. Select the join you want to apply. The following options are available:

    - **Equi**: Select this option if you want to apply equijoin between the two tables or views.
    - **Left**: Select this option if you want to apply left join between the two tables or views.
    - **Right**: Select this option if you want to apply right join between the two tables or views.
    - **Outer**: Select this option if you want to apply outer join between the two tables or views.

11. Click the Add join condition icon to add a join condition.

The system displays the list of columns available in the tables that you are applying the join.



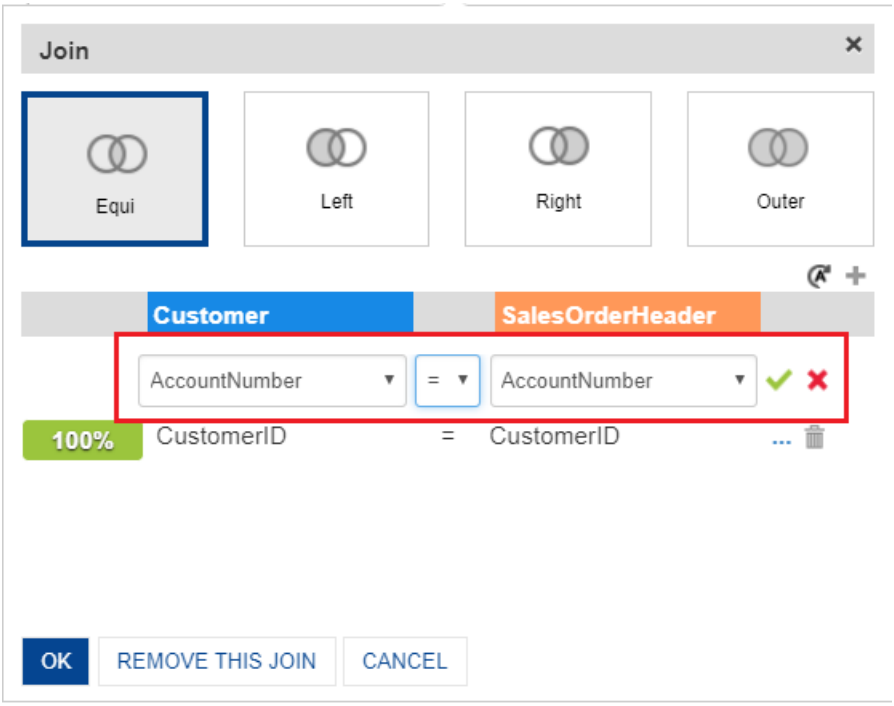
MODIFY JOIN CONDITION—ADD A JOIN CONDITION

- Select the columns for each of the tables from their respective list and the condition you want to apply.

  The system displays the records that matched the join condition in percentage. The background of the percentage value is displayed in the following colors:

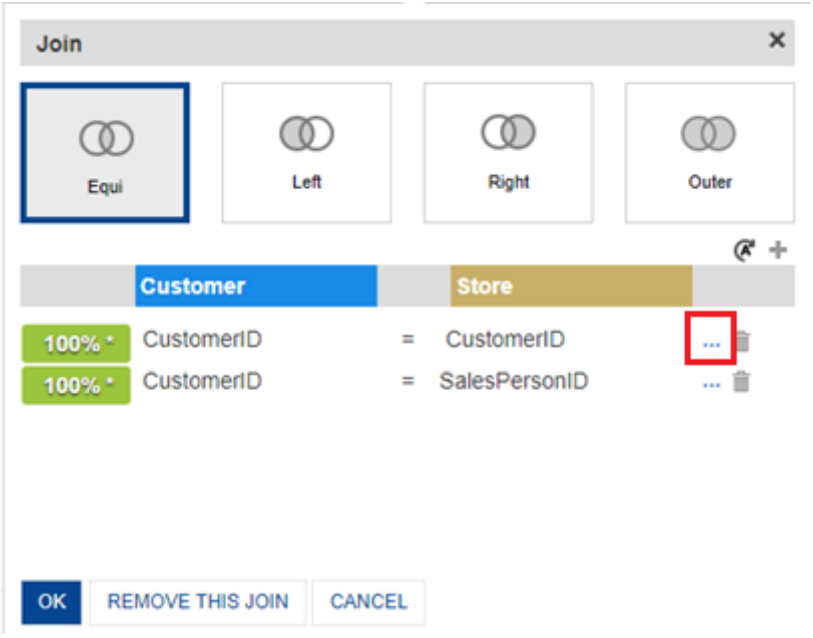
  - **Green**: Indicates that more than 80% of records are identical between the two tables or views.
  - **Orange:** Indicates that more than 50% and less than 80% of records are identical between the two tables or views.
  - **Red:** Indicates that less than 50% of records are identical between the two tables or views.

  > Note:
  > The "*" in the percentage value indicates that the value is calculated based on the sample data. Click the percentage value to calculate the actual percentage match for the entire data.



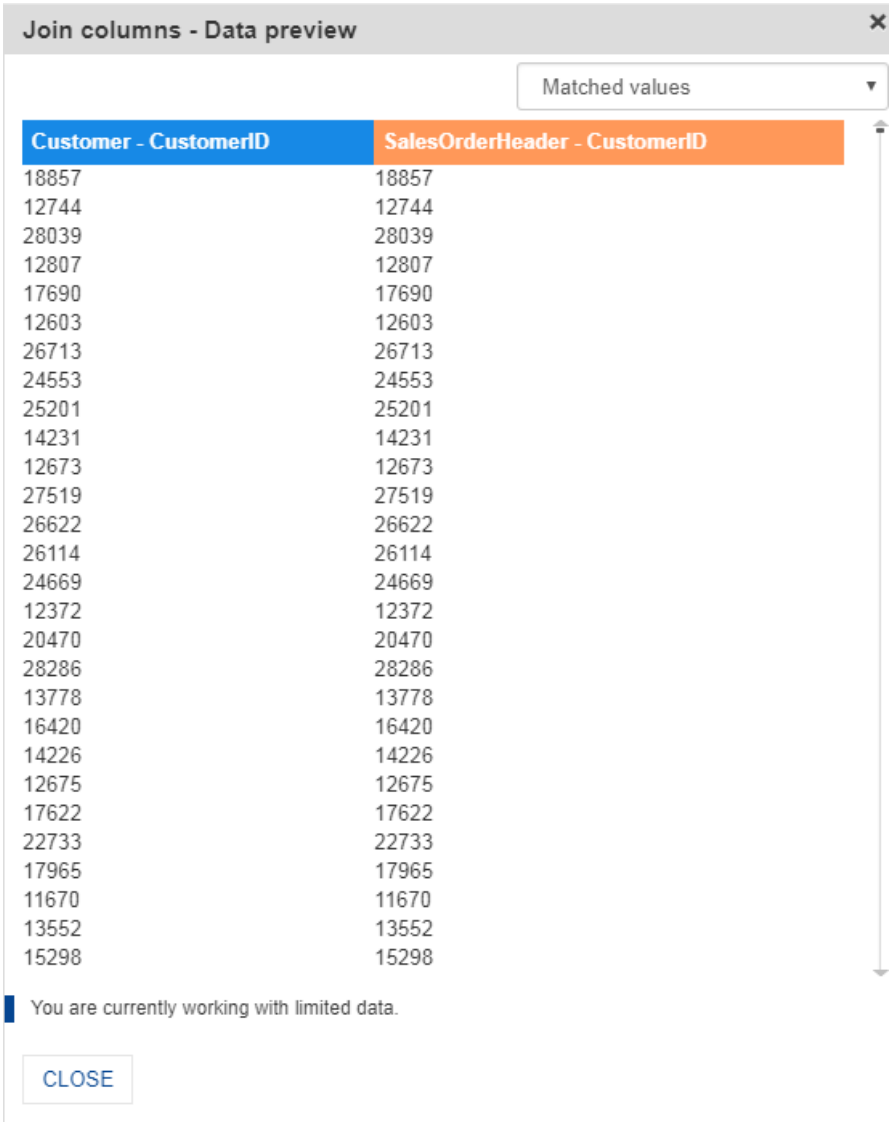
MODIFY JOIN CONDITION—ADDING A JOIN CONDITION

12. You can click the View matching records to view matching values in both columns.



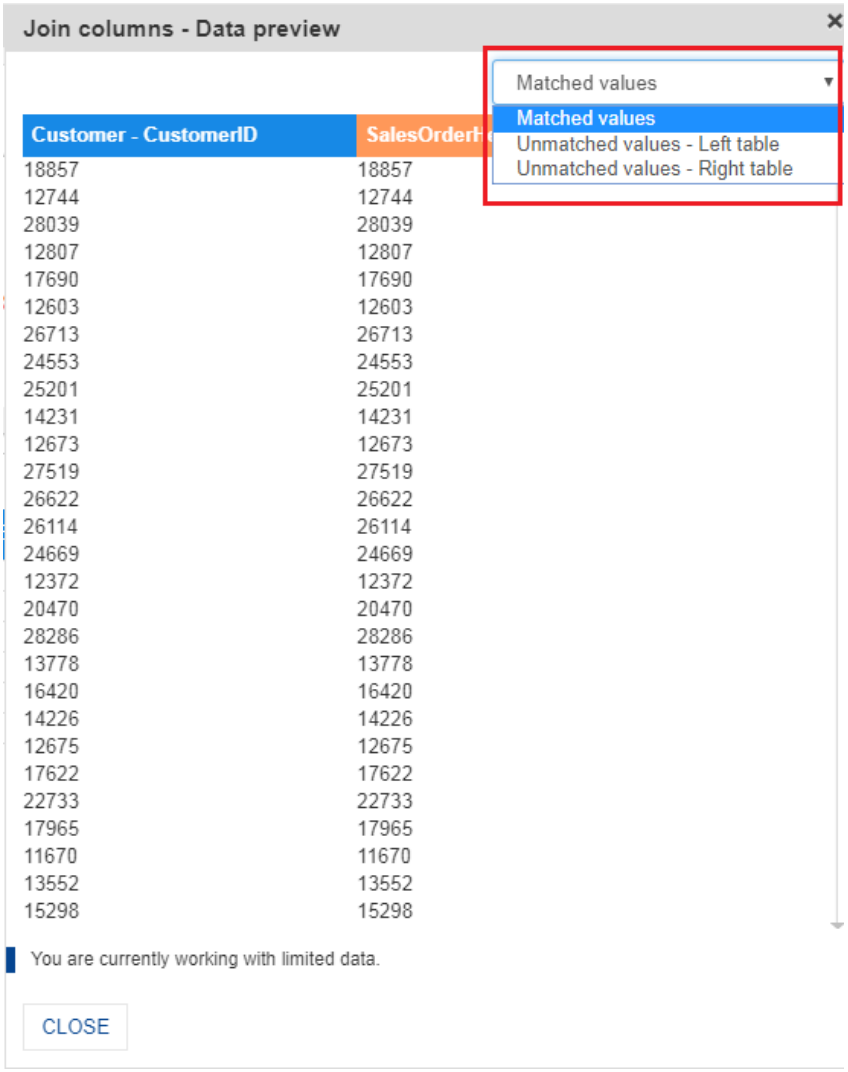
ADDING A JOIN CONDITION—VIEW PREVIEW BUTTON

The system displays the matched values in the **Data preview** dialog box.



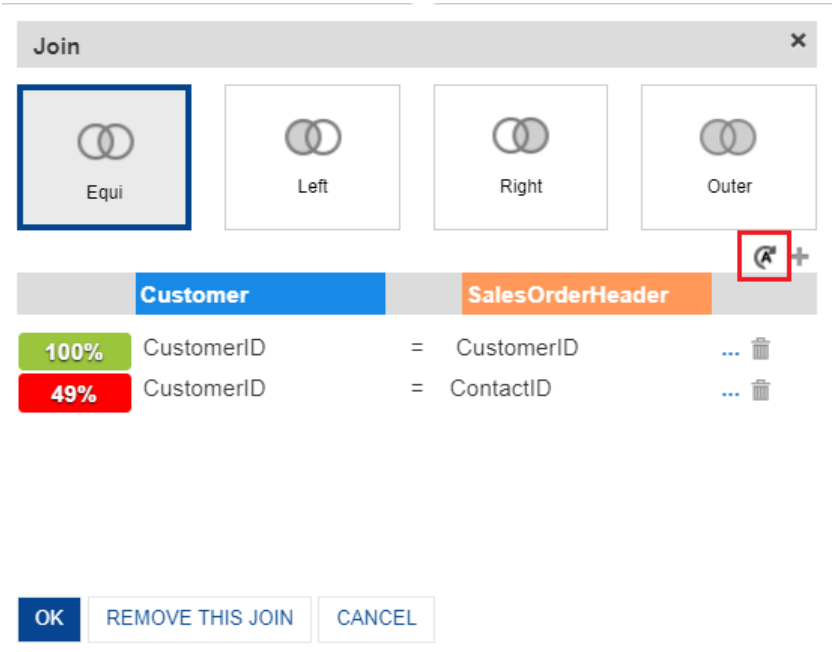
ADDING A JOIN CONDITION—PREVIEW MATCHING VALUES

- You can select Unmatched values—Left table or Unmatched values—Right table to view unmatched values from the left table or the right table.



ADDING A JOIN CONDITION—VIEW UNMATCHED VALUES

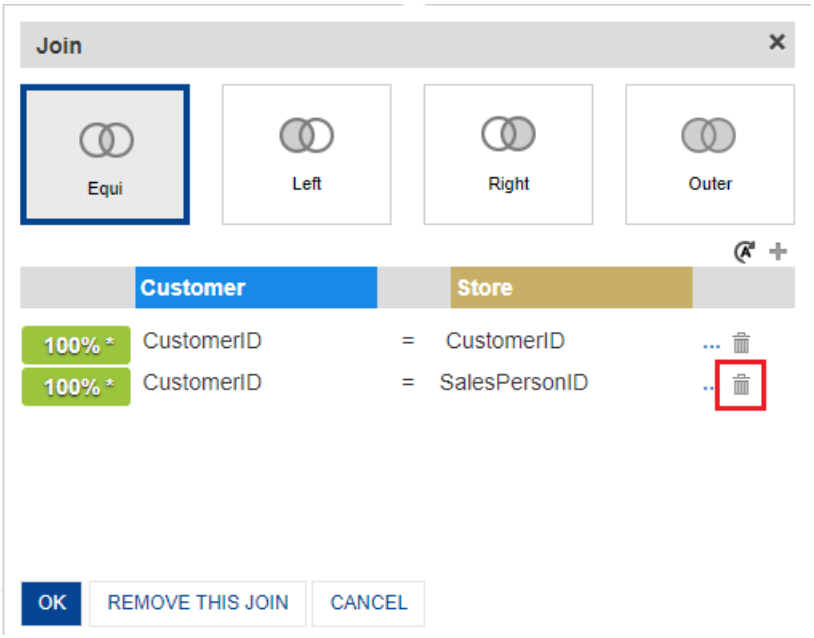
13. You can click the Auto suggested join button to restore the modified or deleted join conditions that were applied by the system.



MODIFY JOIN CONDITION—RESTORE AUTO SUGGESTED JOIN CONDITIONS

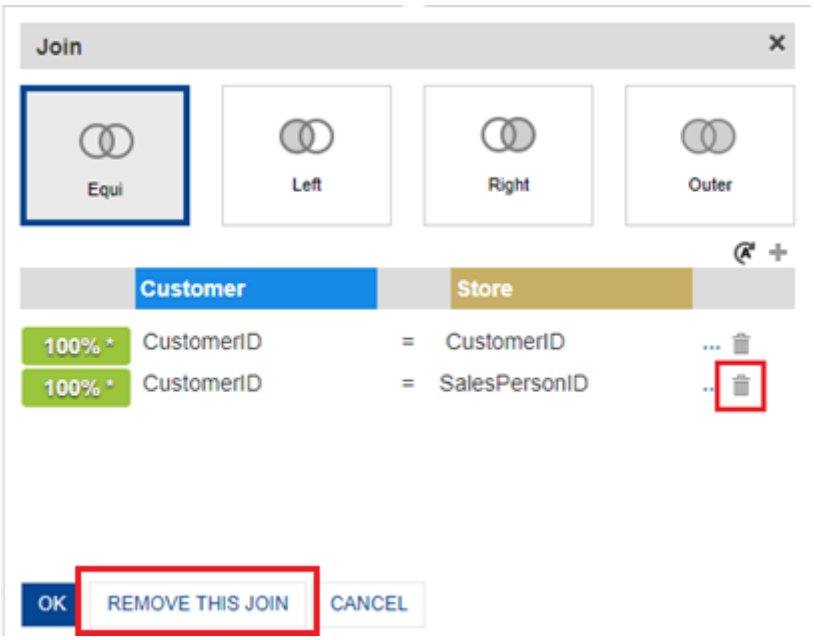
When you join two tables or views, the system automatically applies join conditions. You can modify or delete these join conditions. You can click the Auto suggested join condition to restore the original join condition without affecting the additional join conditions that you have applied.

14. You can click the delete icon next to a join condition to delete that join condition.



MODIFY JOIN CONDITION—DELETING A JOIN CONDITION

15. You can click **REMOVE THIS JOIN** to delete the join between the tables or views.

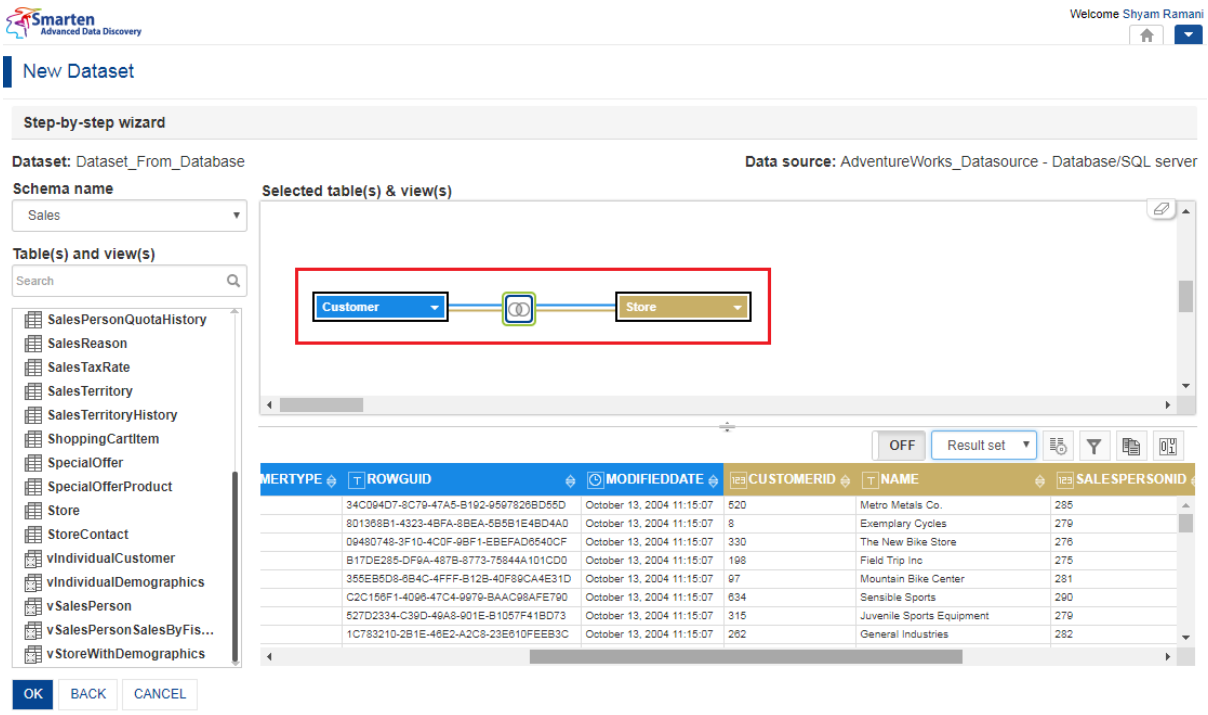


MODIFY JOIN CONDITION—REMOVING A JOIN

16. Click **OK**.

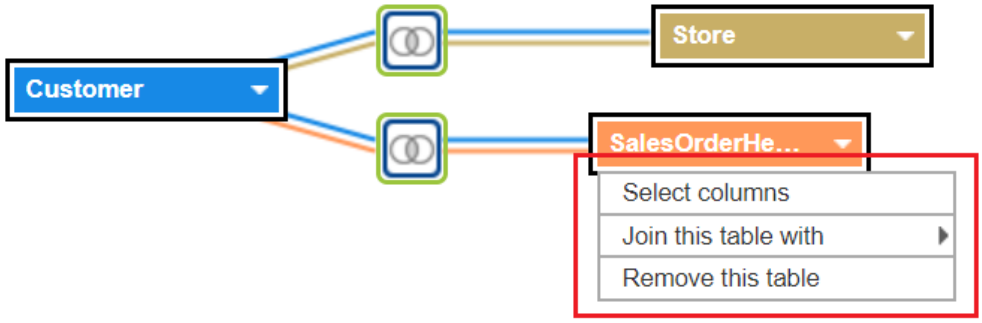
The system displays the data based on the joins applied between the tables and views.

> Note:
> If you have applied a join that results in too many rows, the system displays a message notifying the same and proceeds to create that join after confirmation.



CRATING A DATASET—PREVIEW OF THE DATA BASED ON THE JOIN CONDITION APPLIED

17. Click a table to perform the following operation:



CREATING A DATASET—PERFORMING AN OPERATION ON A TABLE

- **Select columns**: Click this option to select or deselect the columns that you want to use in the data from the **Select columns** dialog box.
- **Join this table with**: Select this option to join the table with other tables from the list. The list displays tables that do not have any join with the table.
  The system displays the **Join** dialog box. Refer to steps 10–13 to add a join condition.

> Note:
> You must specify at least one join condition when you join a table using this option.

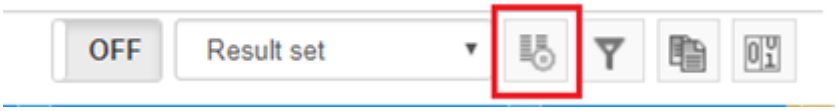
- **Remove this table**: Select this option to remove the table or view.

18. You can perform the following operations:

- **Preview data list**: Select an option from the list to view data for the selected table, view, or the result set.

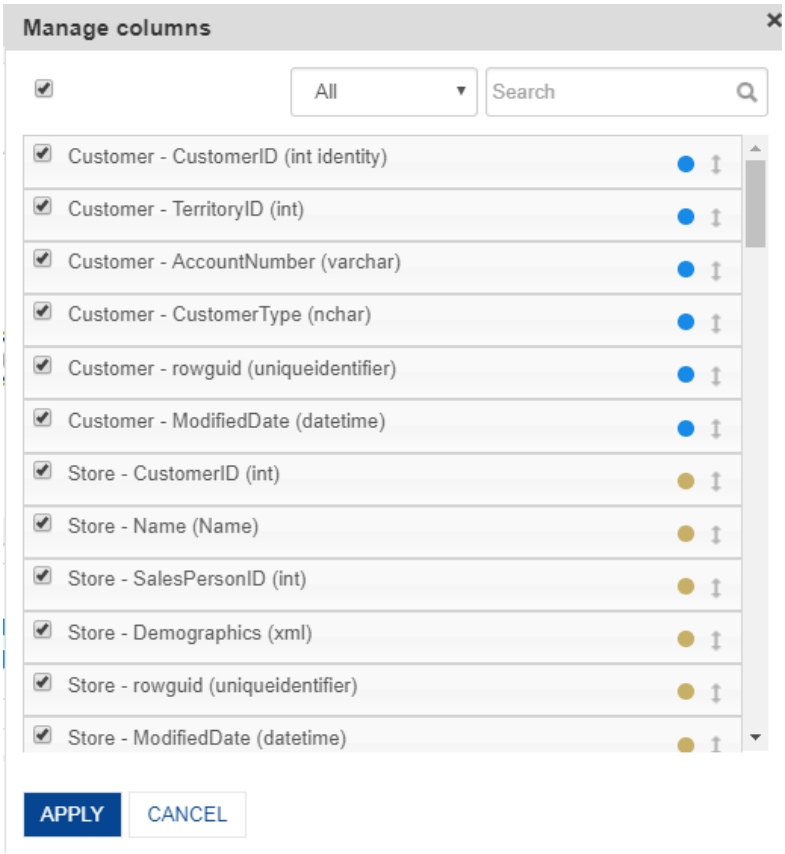
PREVIEW DATA—PREVIEW DATA FROM LIST

- **Manage columns**: Click this option to select the columns you want to use in the result set from each table you have used.



PREVIEW DATA—MANAGE COLUMN OPTION

  - Select the check box adjacent to the columns that you want to include in the dataset.
  - You can select a table from the list to narrow down the list of columns available.
    A circle adjacent to a column is of the same color as the table it belongs to. For example, in the image below, the columns that belong to the Customer table have a blue colored circle adjacent to them.
  - You can drag-and-drop the columns to rearrange their order.



PREVIEW DATA—MANAGE COLUMNS DIALOG BOX

- **Filter**: Select this option to specify the criteria to filter the data that will be retrieved from the database based on the condition applied.

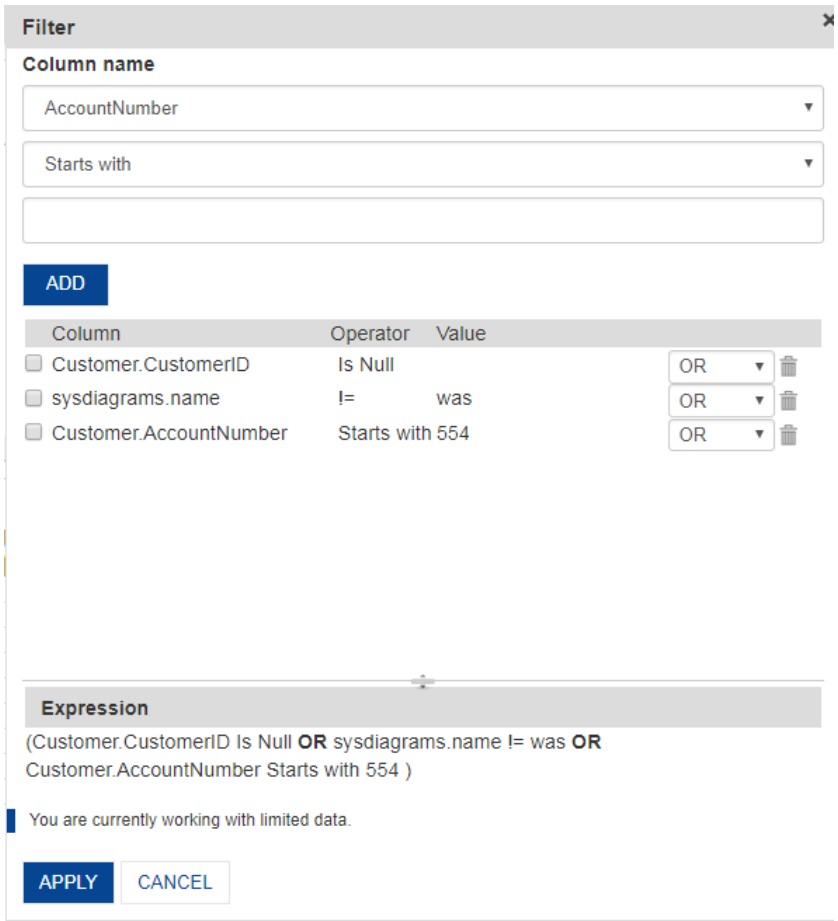
PREVIEW DATA—FILTER OPTION

  - Select a column from the **Column name** list.
  - Select an option from the operations list.
    Based on the option you select from the list, the system displays a box, and you have to provide a value in that box.

> Note:
> The options available in the operations list are based on the data type of the column you have selected from the **Column name** list.

  - Click **ADD**.

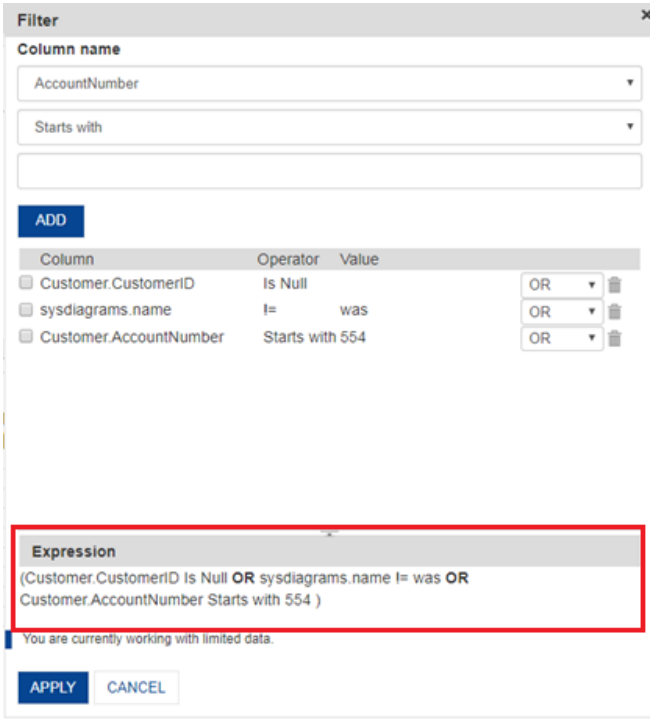
    The system displays the criteria for the column.



PREVIEW DATA—FILTER DIALOG BOX

  - Repeat steps 1–3 to add more criteria for other columns.
  - You can select an option to apply AND or OR condition with that criterion and the next criteria in the list.

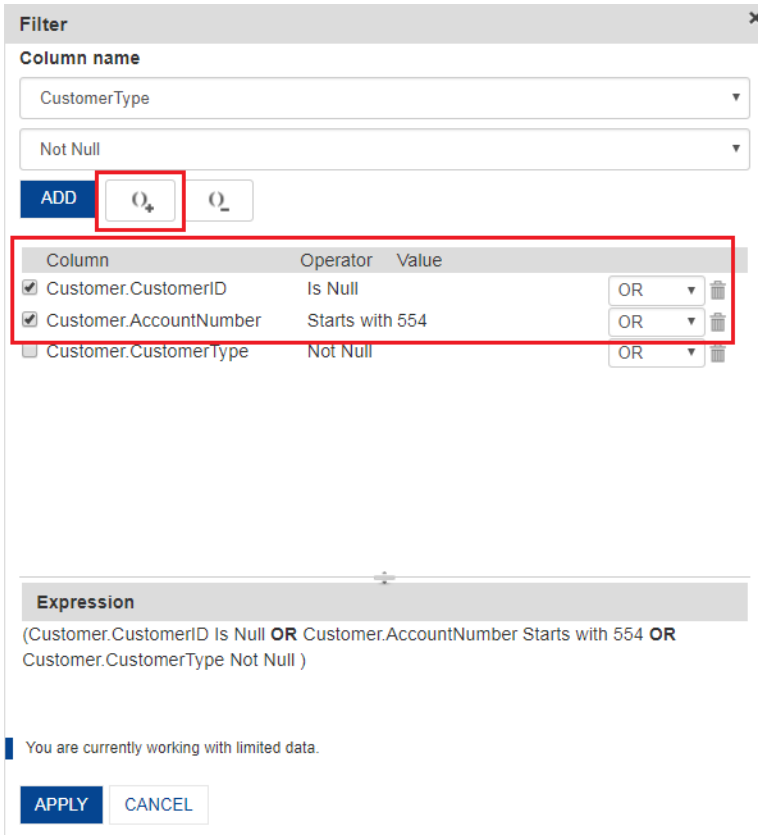
The criteria you add are displayed within the **Expression** section. The criteria in the **Expression** section are similar to the “Where” condition used in the SQL query.



FILTER DIALOG BOX—EXPRESSION SECTION

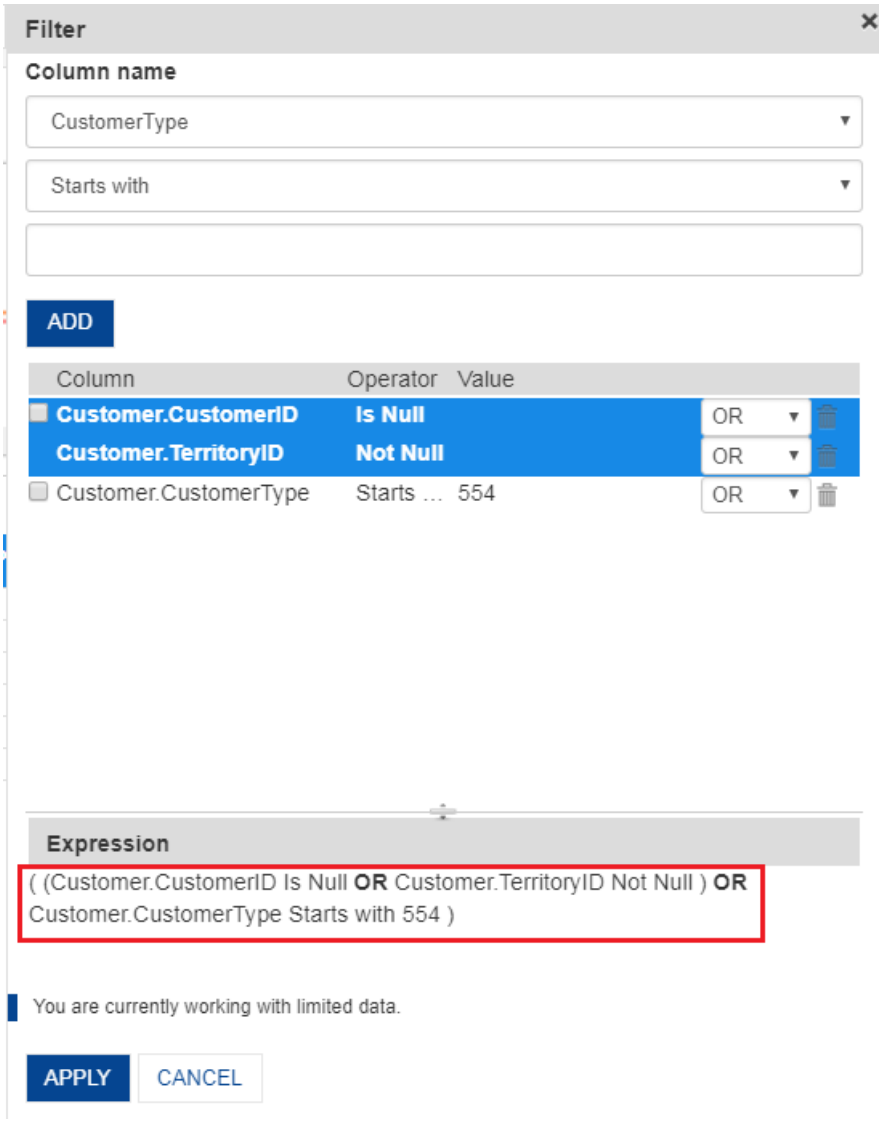
- If there are more than two criteria, you can select criteria that you want to group together and then click the group option.

The selected criteria appear in the parenthesis in the **Expression** section.

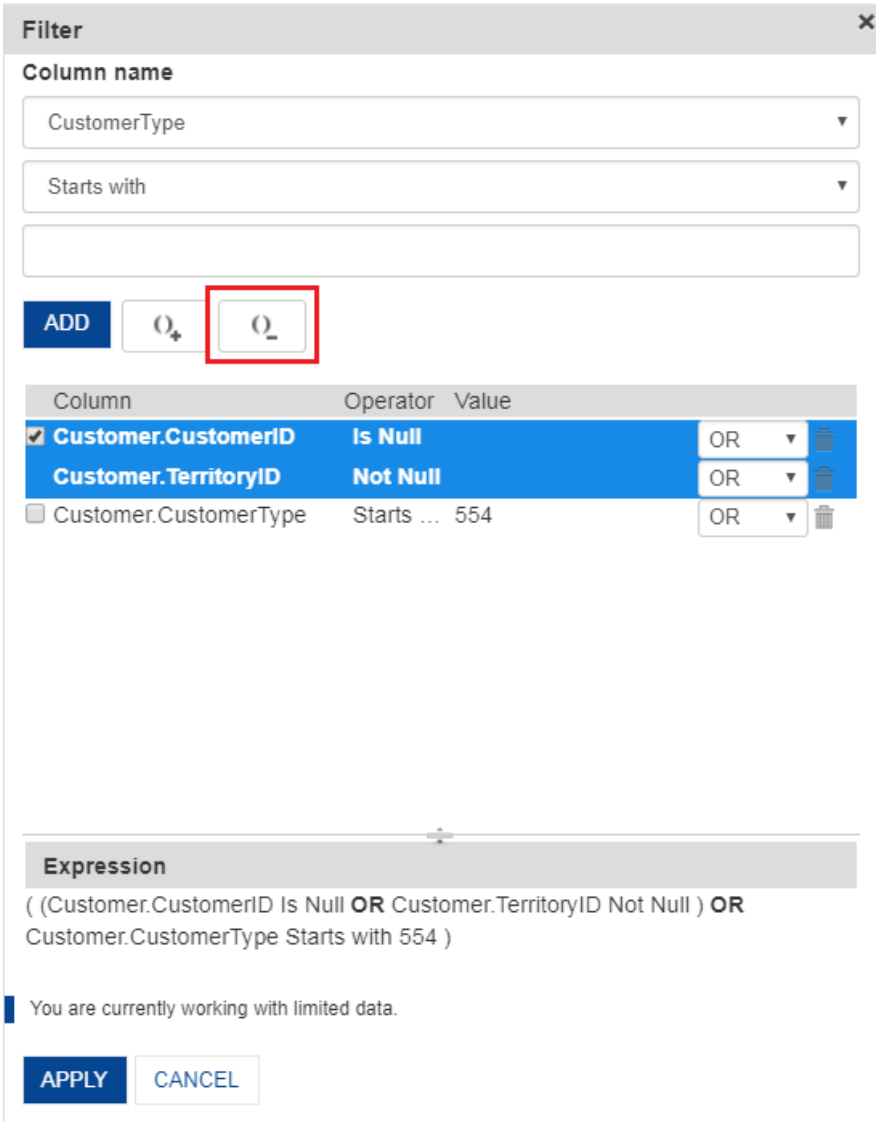


PREVIEW DATA—FILTER DIALOG BOX

The selected criteria are grouped, and the same is reflected in the **Expression** section.



- You can select a grouped criteria and click the ungroup button to ungroup the selected criteria.

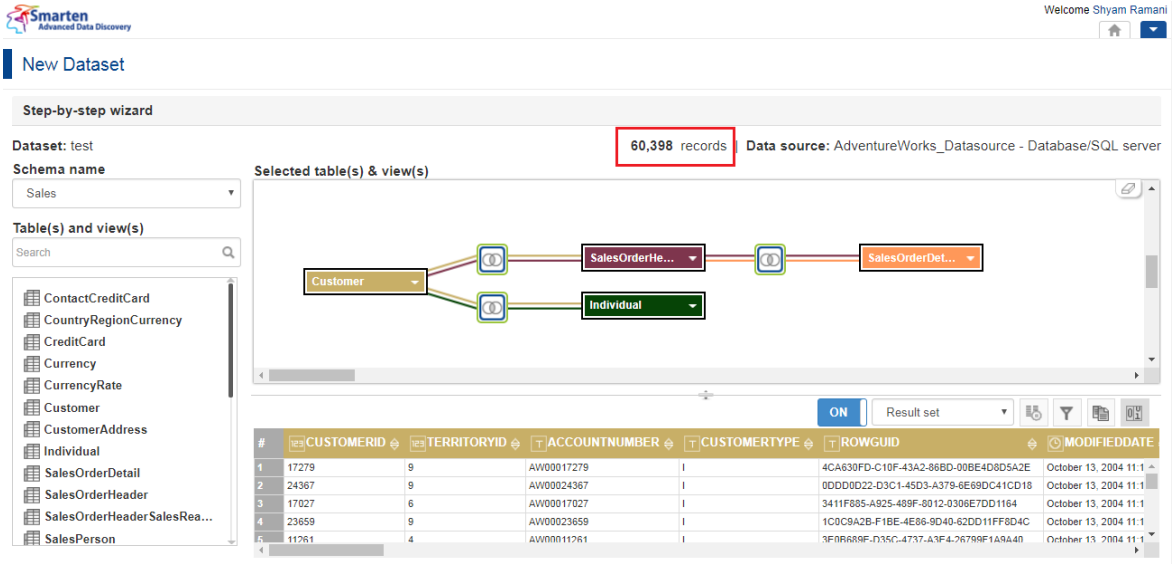


The criteria are no longer in parenthesis in the **Expression** section.

- **Enable record count**: Click this option to enable a record count.

PREVIEW DATA—RECORD COUNT OPTION

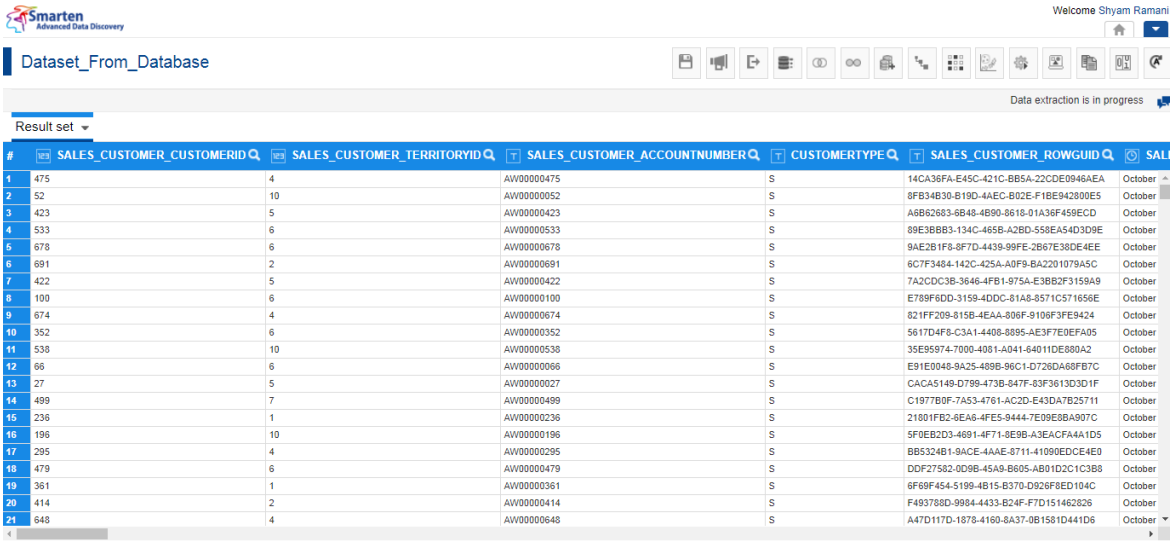
When enabled, this option displays the total number of records available.



PREVIEW DATA—RECORD COUNT ENABLED

19. Click **OK**.

The system retrieves the data from the database and creates the dataset.



PREVIEW—DATASET CREATED USING A DATABASE PROFILE

The dataset is now available in the repository.

6.2 Creating a Dataset Using a File Type Profile

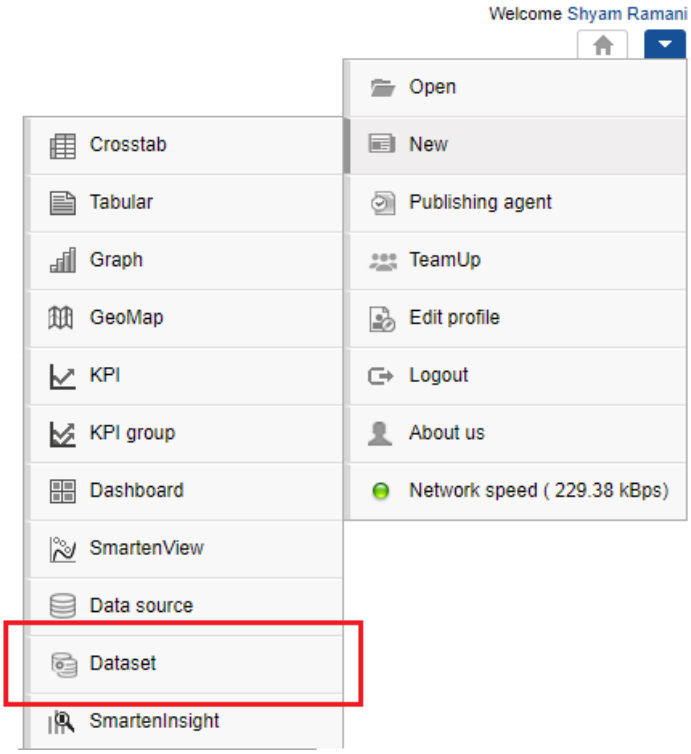
You can use a file as a data source and create a dataset to retrieve the required data from that data source.

About this task

Use this task to create a dataset using a file type profile.

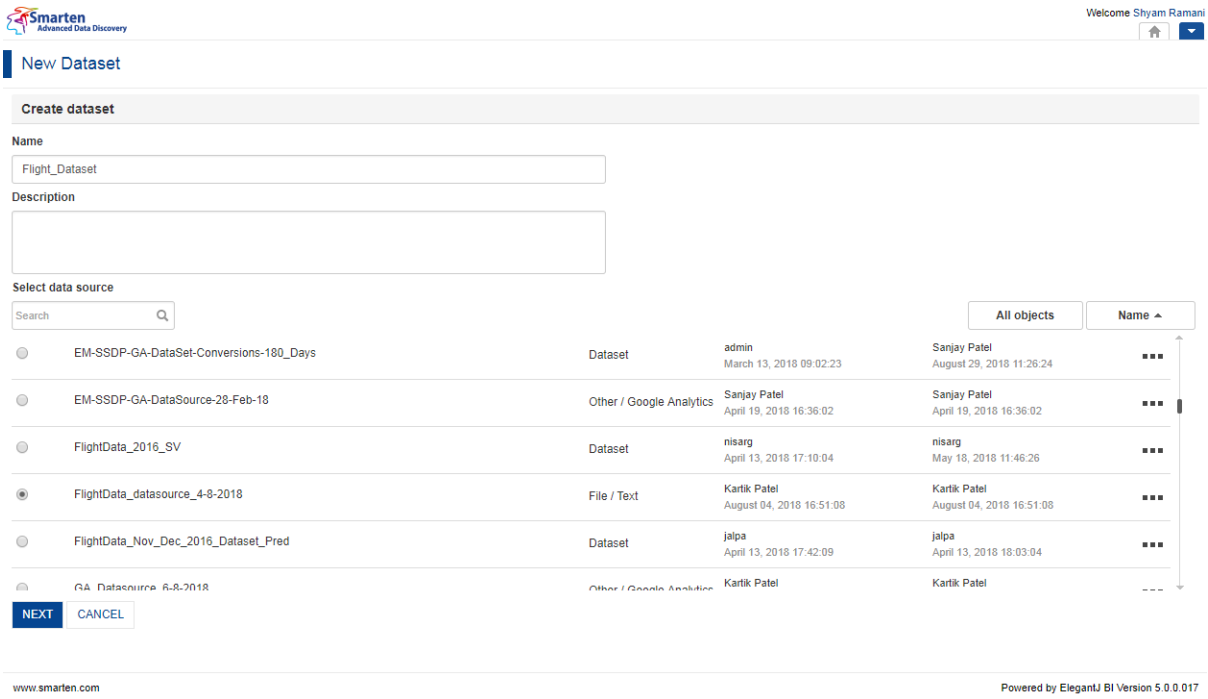
Procedure

1. Click **New** -> **Dataset** from the menu.



MENU OPTION—NEW DATASET

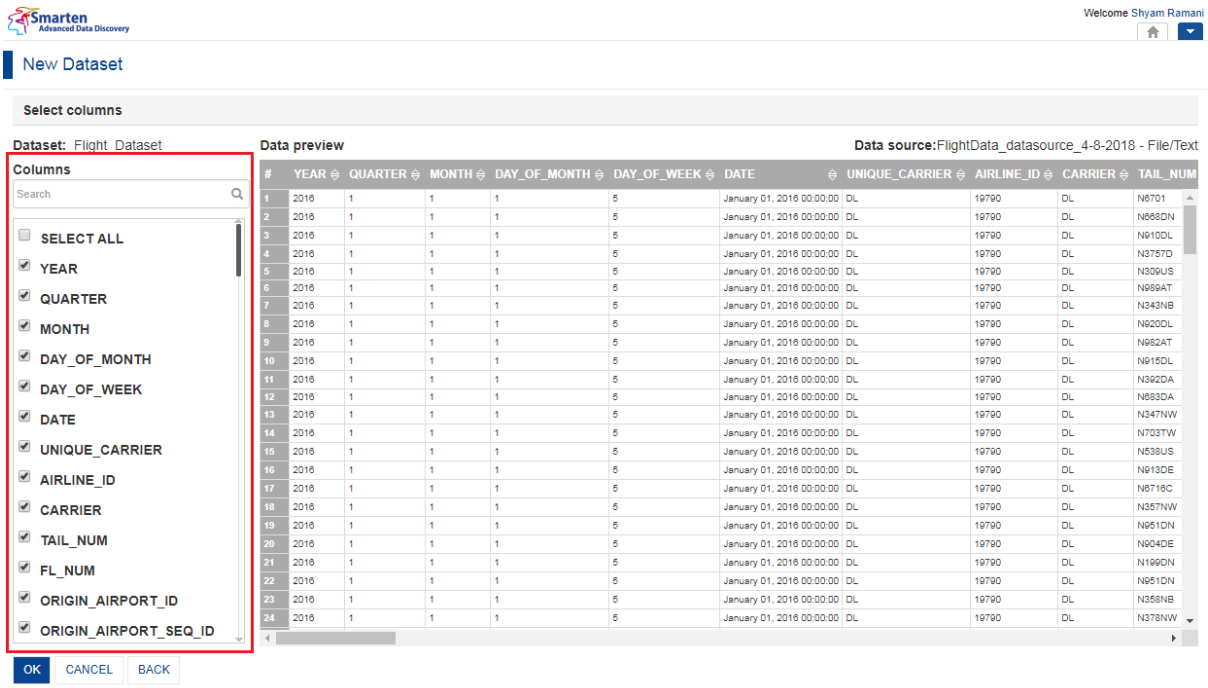
The system displays the **Create dataset** page.



CREATING A NEW DATASET—THE CREATE DATASET SCREEN

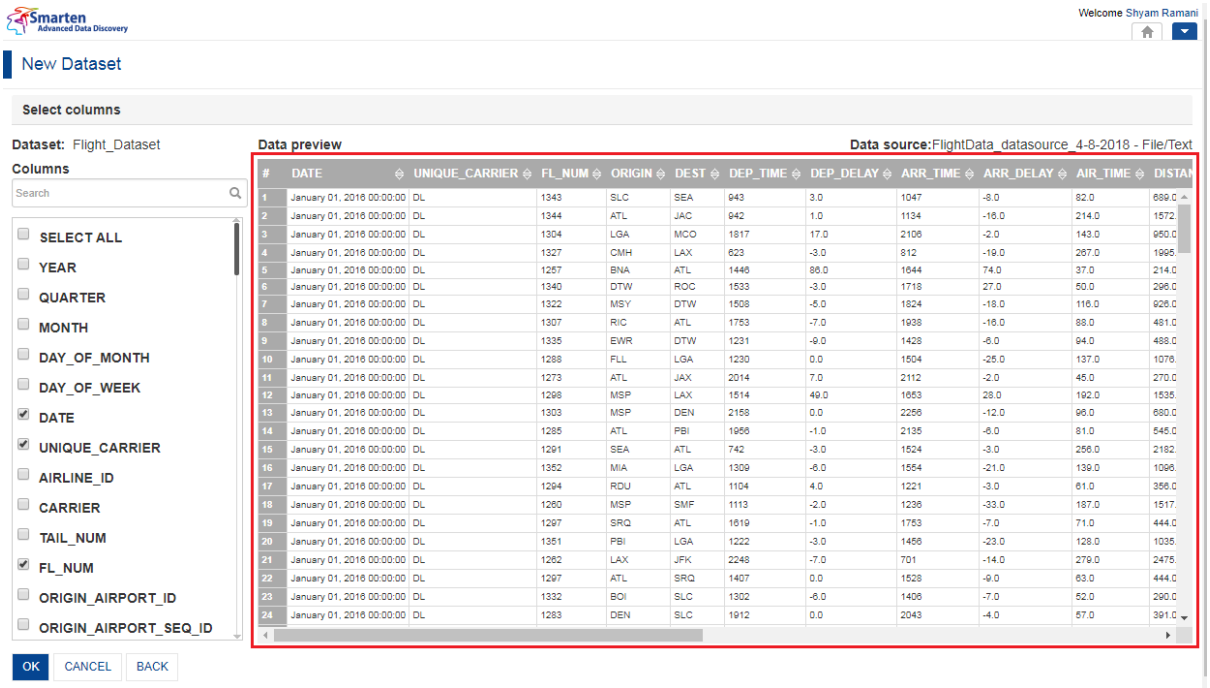
2. Enter a name and description for the dataset in the **Name** and **Description** fields.
3. Select the file type profile you want to use to create a dataset from the list.
   Or,
   Enter the keyword to search for the file type data source profile based on the keyword provided, and then press **Enter**.
   The system displays the file type data source profiles that match the keyword you have provided.
4. Click **Next**.

The system displays the **Select columns** page.



CREATING A DATASET—COLUMNS AVAILABLE IN THE SELECTED FILE TYPE DATA SOURCE

5. Select the columns that you want to use in the dataset from the left pane.
   Or,
   Enter the keyword to search for the columns you want to use in the dataset in the **Columns** field, and then press **Enter**.

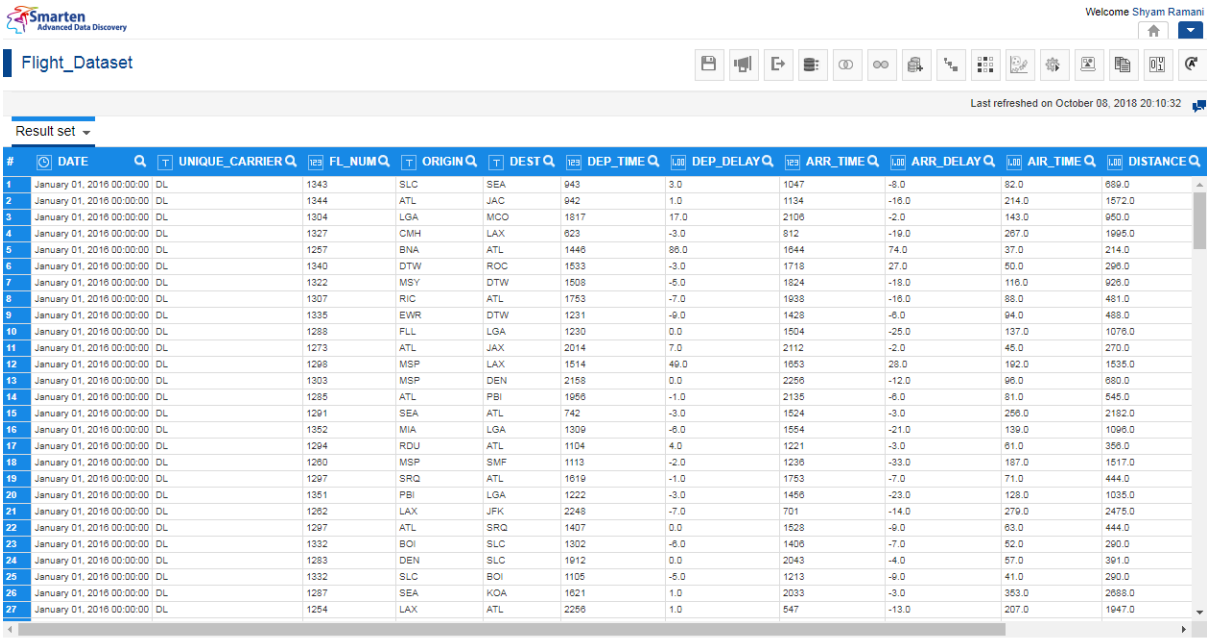


CREATING A DATASET—DATA FROM THE SELECTED COLUMNS

The system displays the columns you have selected along with the data available in those columns in the **Data preview** pane.

6. Click **OK**.

   The system displays a preview of the data based on the columns you have selected.



#	DATE	UNIQUE_CARRIER	FL_NUM	ORIGIN	DEST	DEP_TIME	DEP_DELAY	ARR_TIME	ARR_DELAY	AIR_TIME	DISTANCE
1	January 01, 2016 00:00:00	DL	1343	SLC	SEA	943	3.0	1047	-8.0	82.0	689.0
2	January 01, 2016 00:00:00	DL	1344	ATL	JAC	942	1.0	1134	-16.0	214.0	1572.0
3	January 01, 2016 00:00:00	DL	1304	LGA	MCO	1817	17.0	2106	-2.0	143.0	950.0
4	January 01, 2016 00:00:00	DL	1327	CMH	LAX	623	-3.0	812	-19.0	267.0	1995.0
5	January 01, 2016 00:00:00	DL	1257	BNA	ATL	1446	86.0	1644	74.0	37.0	214.0
6	January 01, 2016 00:00:00	DL	1340	DTW	ROC	1533	-3.0	1718	27.0	50.0	296.0
7	January 01, 2016 00:00:00	DL	1322	MSY	DTW	1508	-5.0	1824	-18.0	116.0	926.0
8	January 01, 2016 00:00:00	DL	1307	RIC	ATL	1753	-7.0	1938	-16.0	88.0	481.0
9	January 01, 2016 00:00:00	DL	1335	EWR	DTW	1231	-9.0	1428	-6.0	94.0	488.0
10	January 01, 2016 00:00:00	DL	1288	FLL	LGA	1230	0.0	1504	-25.0	137.0	1076.0
11	January 01, 2016 00:00:00	DL	1273	ATL	JAX	2014	7.0	2112	-2.0	45.0	270.0
12	January 01, 2016 00:00:00	DL	1298	MSP	LAX	1514	49.0	1653	28.0	192.0	1535.0
13	January 01, 2016 00:00:00	DL	1303	MSP	DEN	2158	0.0	2256	-12.0	96.0	680.0
14	January 01, 2016 00:00:00	DL	1285	ATL	PBI	1956	-1.0	2135	-6.0	81.0	545.0
15	January 01, 2016 00:00:00	DL	1291	SEA	ATL	742	-3.0	1524	-3.0	256.0	2182.0
16	January 01, 2016 00:00:00	DL	1352	MIA	LGA	1309	-6.0	1554	-21.0	139.0	1096.0
17	January 01, 2016 00:00:00	DL	1294	RDU	ATL	1104	4.0	1221	-3.0	61.0	356.0
18	January 01, 2016 00:00:00	DL	1260	MSP	SMF	1113	-2.0	1236	-33.0	187.0	1517.0
19	January 01, 2016 00:00:00	DL	1297	SRQ	ATL	1619	-1.0	1753	-7.0	71.0	444.0
20	January 01, 2016 00:00:00	DL	1351	PBI	LGA	1222	-3.0	1456	-23.0	128.0	1035.0
21	January 01, 2016 00:00:00	DL	1262	LAX	JFK	2248	-7.0	701	-14.0	279.0	2475.0
22	January 01, 2016 00:00:00	DL	1297	ATL	SRQ	1407	0.0	1528	-9.0	63.0	444.0
23	January 01, 2016 00:00:00	DL	1332	BOI	SLC	1302	-6.0	1406	-7.0	52.0	290.0
24	January 01, 2016 00:00:00	DL	1283	DEN	SLC	1912	0.0	2043	-4.0	57.0	391.0
25	January 01, 2016 00:00:00	DL	1332	SLC	BOI	1105	-5.0	1213	-9.0	41.0	290.0
26	January 01, 2016 00:00:00	DL	1287	SEA	KOA	1621	1.0	2033	-3.0	363.0	2688.0
27	January 01, 2016 00:00:00	DL	1254	LAX	ATL	2256	1.0	547	-13.0	207.0	1947.0

PREVIEW—DATASET USING A FILE TYPE PROFILE

The dataset is now available in the repository.

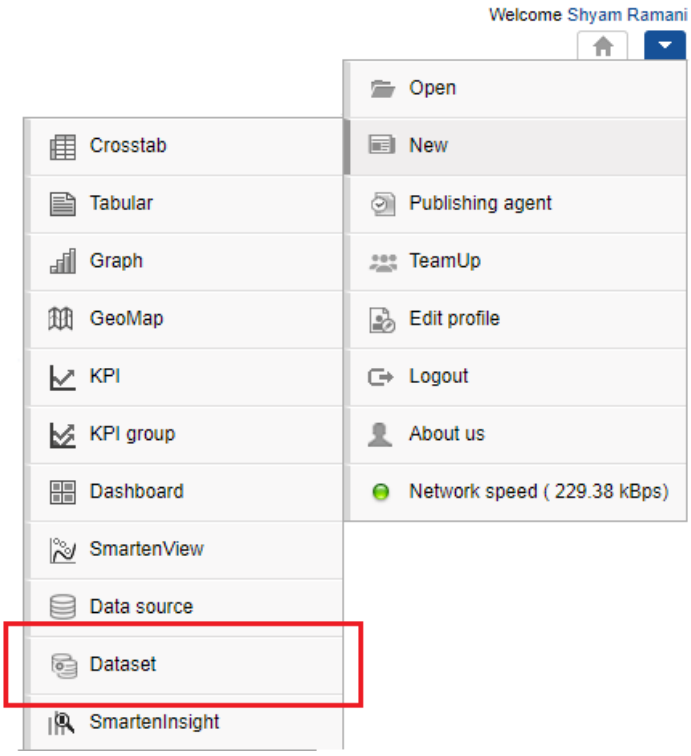
6.3 Creating a Dataset Using a Google Analytics Profile

About this task

Use this task to create a dataset using a Google Analytics profile.

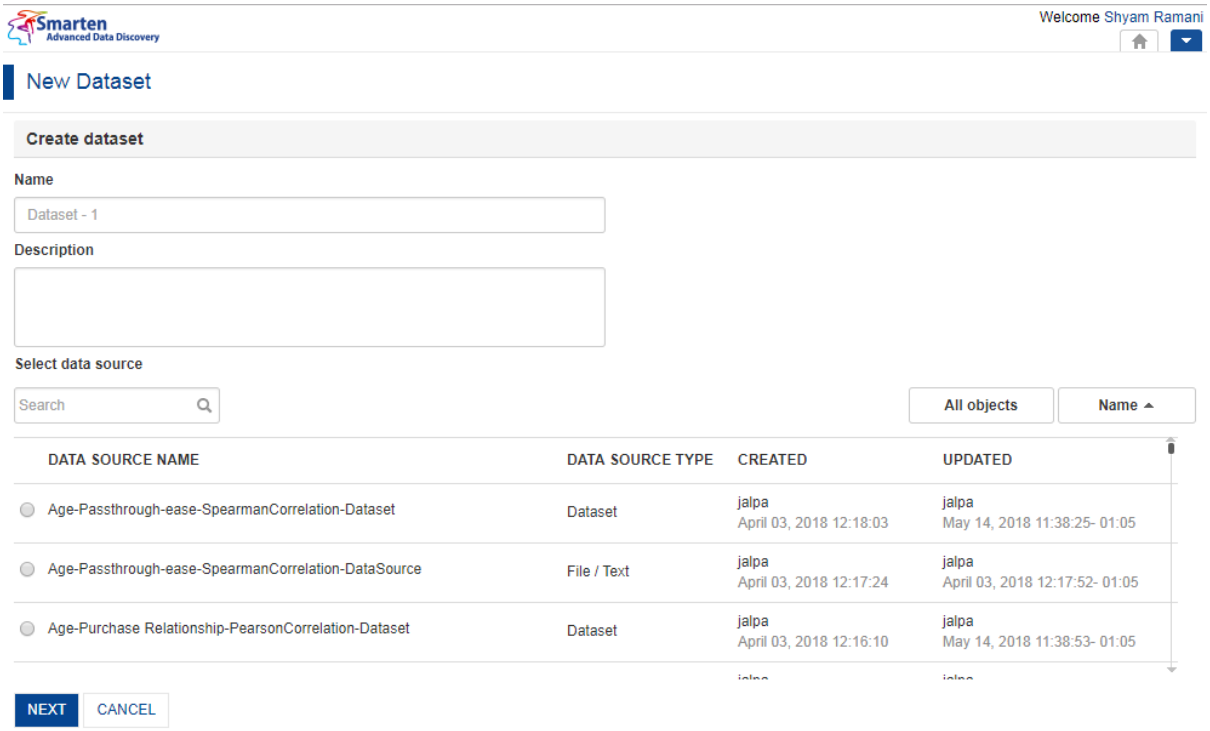
Procedure

1 Click **New** -> **Dataset** from the menu.



MENU OPTION—NEW DATASET

The system displays the **Create dataset** page.



CREATING A NEW DATASET—THE CREATE DATASET SCREEN

2 Enter a name and description for the dataset in the **Name** and **Description** fields.

3 Select the Google Analytics profile you want to use to create a dataset from the list.
Or,
Enter the keyword to search for the Google Analytics data source profile based on the keyword provided, and then press **Enter**.
The system displays the Google Analytics data source profiles that match the keyword you have provided.

4 Click **Next**.

The system displays the **Select Google Analytics parameters** page. The information about the measure group, dimensions, and measures related to the Google Analytics profile is available in the left pane.

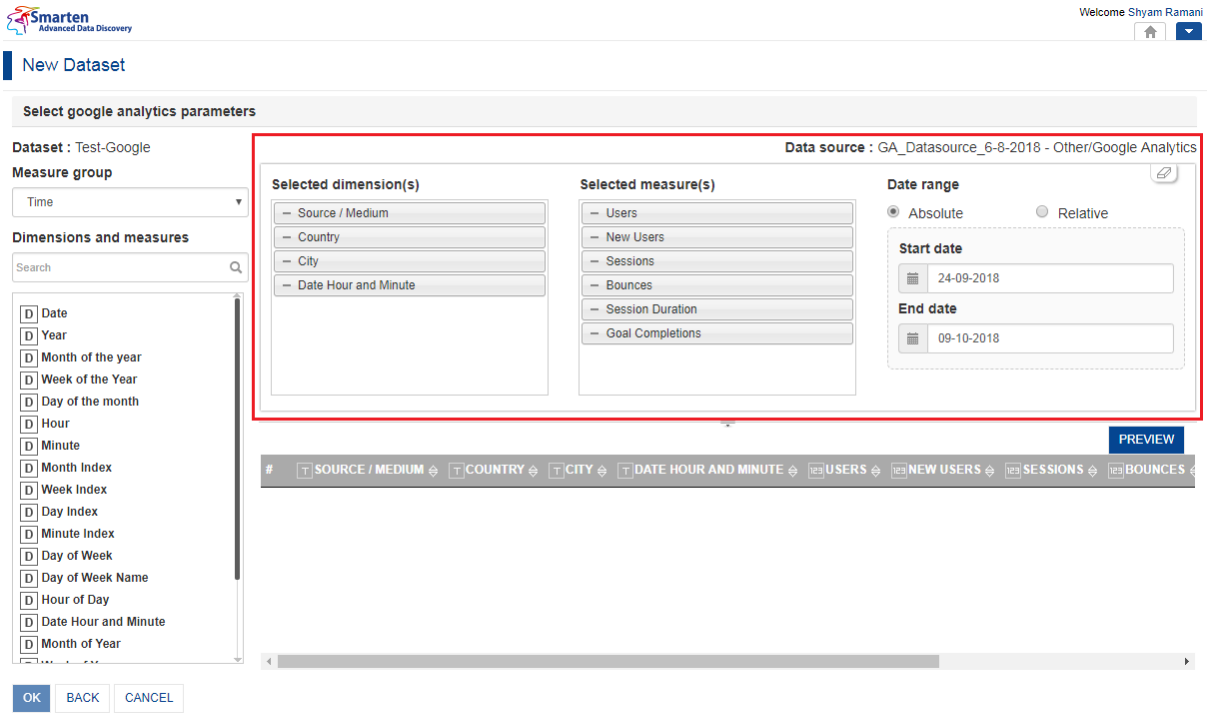
CREATING A DATASET—SELECTING GOOGLE ANALYTICS PARAMETERS

5 Select the measure group you want to use from the **Measure group** list.
The system displays the dimensions and measures available for the selected measure group in the **Dimensions and measures** section.

6 Double-click the dimensions and measures you want to use from the **Dimensions and measures** section.

> Note:
> You can select a maximum of 7 dimensions and 10 measures to retrieve data from the Google Analytics profile due to Google Analytics API restrictions.

The selected dimensions and measures are displayed in the right pane.



CREATING A DATASET—SELECTED DIMENSIONS AND MEASURES FROM THE GOOGLE ANALYTICS PROFILE

7 Repeat steps 5 and 6 to add dimensions and measures from more than one measure group.

8 Select an option to specify a time period for which you want to retrieve data.

The following options are available:

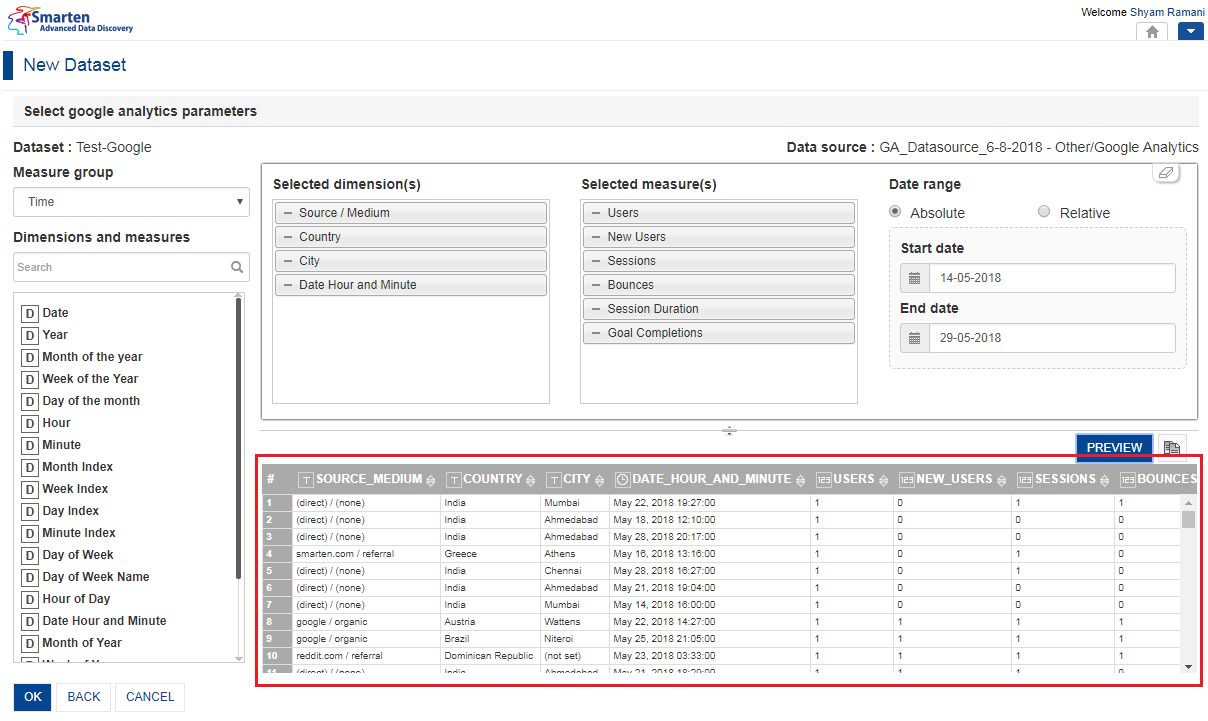
- **Absolute**: Select this option to specify an exact date range. Specify a start date and an end date in the **Start date** and **End date** fields.
- **Relative**: Select this option to specify a relative time period. Select an option from the **Period** list. The following table lists and describes the options available in the **Period** list.

Option	Description
Today	Select this option to retrieve data from the Google Analytics profile for today. For example, if today is 10th August 2018, the data will be retrieved for 10th August 2018.
Yesterday	Select this option to retrieve data from the Google Analytics profile for yesterday. For example, if today is 10th August 2018, the data will be retrieved for 9th August 2018.
This Week	Select this option to retrieve data from the Google Analytics profile for the current week starting from Monday until today. For example, if today is 10th August 2018, the data will be retrieved from Monday, 6th August 2018 until 10th August 2018.
Last Week	Select this option to retrieve data from the Google Analytics profile for the last week. The last week comprises the time period starting from the Monday and ending on Sunday of the

	previous week. For example, if today is 10th August 2018, the data will be retrieved from Monday, 30th July 2018 until 5th August 2018..
This Month	Select this option to retrieve data from the Google Analytics profile for the current month starting from the first day of the month until today. For example, if today is 10th August 2018, the data will be retrieved from 1st August 2018 until 10th August 2018.
Last Month	Select this option to retrieve data from the Google Analytics profile for the last month. The last month comprises the time period starting from the first day and ending on the last day of the previous month. For example, if today is 10th August 2018, the data will be retrieved for the period starting from 1st July until 31st July.
This Quarter	Select this option to retrieve data from the Google Analytics profile for the current quarter. The time period for the current quarter starts from the first day of the quarter until today. For example, if today is 10th August, the data will be retrieved for the period starting from 1st July until 10th August.
Last Quarter	Select this option to retrieve data from the Google Analytics profile for the previous quarter. The time period for the previous quarter starts from the first day of that quarter until the last day of the quarter. For example, if today is 10th August, the data will be retrieved for the period starting from 1st April until 30th June.
Last 3 Months	Select this option to retrieve data from the Google Analytics profile for the last three months. The time period for the current quarter starts from the first day of the three months before the current month until the last day of the previous month. For example, if today is 10th August, the period for the last three months would be from 1st May until 31st July.
Last Year	Select this option to retrieve data from the Google Analytics profile for the last year. The time period for the last year starts from the first day of the previous year to the last day of the previous year. For example, if today is 10th August 2018, the period for last year would be from 1st January 2017 until 31st December 2017.
Last 7 Days	Select this option to retrieve data from the Google Analytics profile for the last seven days. The time period for the last seven days ranges from the seven days before today. For example, if today is 10th August, the period for the last seven days would be from 4th August until 10th August.
Last 30 Days	Select this option to retrieve data from the Google Analytics profile for the last 30 days. The time period for the last 30 days ranges from the 30 days before today. For example, if today is 10th August, the period for the last 30 days would be from 12th July until 10th August.
Last 60 Days	Select this option to retrieve data from the Google Analytics

	profile for the last 60 days. The time period for the last 60 days ranges from the 60 days before today. For example, if today is 10th August, the period for the last 60 days would be from 12th June until 10th August.
Last 90 Days	Select this option to retrieve data from the Google Analytics profile for the last 90 days. The time period for the last 90 days ranges from the 90 days before today. For example, if today is 10th August, the period for the last 90 days would be from 13th May until 10th August.
Last 180 Days	Select this option to retrieve data from the Google Analytics profile for the last 180 days. The time period for the last 180 days ranges from the 180 days before today. For example, if today is 10th August, the period for the last 180 days would be from 11th February until 10th August.
Last 365 Days	Select this option to retrieve data from the Google Analytics profile for the last 365 days. The time period for the last 365 days ranges from the 365 days before today. For example, if today is 10th August 2018, the period for the last 365 days would be from 11th August 2017 until 10th August 2018.

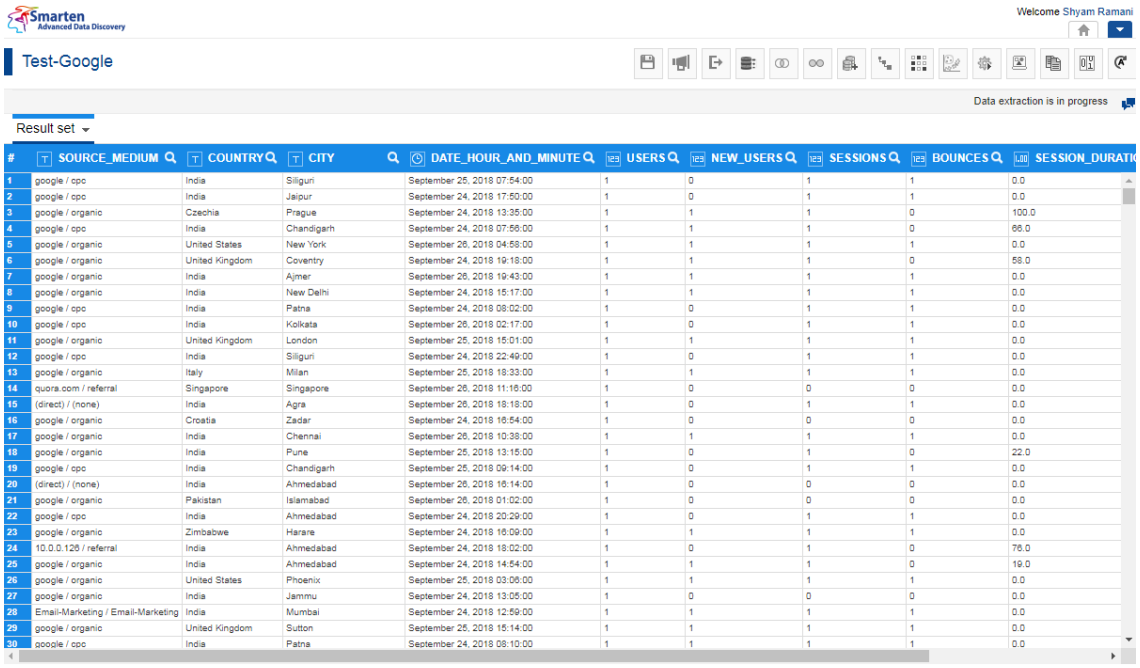
9 Click **PREVIEW** to view a preview of the data retrieved from the Google Analytics profile based on the dimensions and measures you have selected.



CREATING A DATASET—PREVIEW OF THE DATA RETRIEVED FROM THE GOOGLE ANALYTICS PROFILE

10 Click **OK**.

The system retrieves data from the Google Analytics profile.



#	SOURCE_MEDIUM	COUNTRY	CITY	DATE_HOUR_AND_MINUTE	USERS	NEW_USERS	SESSIONS	BOUNCES	SESSION_DURATI
1	google / cpc	India	Siliguri	September 25, 2018 07:54:00	1	0	1	1	0.0
2	google / cpc	India	Jaipur	September 24, 2018 17:50:00	1	0	1	1	0.0
3	google / organic	Czechia	Prague	September 24, 2018 13:35:00	1	1	1	0	100.0
4	google / cpc	India	Chandigarh	September 24, 2018 07:56:00	1	1	1	0	66.0
5	google / organic	United States	New York	September 26, 2018 04:58:00	1	1	1	1	0.0
6	google / organic	United Kingdom	Coventry	September 24, 2018 19:18:00	1	1	1	0	58.0
7	google / organic	India	Ajmer	September 26, 2018 19:43:00	1	1	1	1	0.0
8	google / organic	India	New Delhi	September 24, 2018 15:17:00	1	1	1	1	0.0
9	google / cpc	India	Patna	September 24, 2018 08:02:00	1	0	1	1	0.0
10	google / cpc	India	Kolkata	September 26, 2018 02:17:00	1	0	1	1	0.0
11	google / organic	United Kingdom	London	September 25, 2018 15:01:00	1	1	1	1	0.0
12	google / cpc	India	Siliguri	September 24, 2018 22:49:00	1	0	1	1	0.0
13	google / organic	Italy	Milan	September 25, 2018 18:33:00	1	1	1	1	0.0
14	quora.com / referral	Singapore	Singapore	September 26, 2018 11:16:00	1	0	0	0	0.0
15	(direct) / (none)	India	Agra	September 26, 2018 18:18:00	1	0	1	1	0.0
16	google / organic	Croatia	Zadar	September 24, 2018 16:54:00	1	0	0	0	0.0
17	google / organic	India	Chennai	September 26, 2018 10:38:00	1	1	1	1	0.0
18	google / organic	India	Pune	September 25, 2018 13:15:00	1	0	1	0	22.0
19	google / cpc	India	Chandigarh	September 26, 2018 09:14:00	1	0	1	1	0.0
20	(direct) / (none)	India	Ahmedabad	September 26, 2018 16:14:00	1	0	0	0	0.0
21	google / organic	Pakistan	Islamabad	September 26, 2018 01:02:00	1	0	0	0	0.0
22	google / cpc	India	Ahmedabad	September 24, 2018 20:29:00	1	0	1	1	0.0
23	google / organic	Zimbabwe	Harare	September 24, 2018 16:09:00	1	1	1	1	0.0
24	10.0.0.126 / referral	India	Ahmedabad	September 24, 2018 18:02:00	1	0	1	0	76.0
25	google / organic	India	Ahmedabad	September 24, 2018 14:54:00	1	1	1	0	19.0
26	google / organic	United States	Phoenix	September 25, 2018 03:06:00	1	1	1	1	0.0
27	google / organic	India	Jammu	September 24, 2018 13:05:00	1	0	0	0	0.0
28	Email-Marketing / Email-Marketing	India	Mumbai	September 24, 2018 12:59:00	1	1	1	1	0.0
29	google / organic	United Kingdom	Sutton	September 25, 2018 15:14:00	1	1	1	1	0.0
30	google / cpc	India	Patna	September 24, 2018 08:10:00	1	1	1	1	0.0

PREVIEW—DATASET USING A GOOGLE ANALYTICS PROFILE

The dataset is now available in the repository.

6.4 Creating a Dataset Using an R Script Data Source

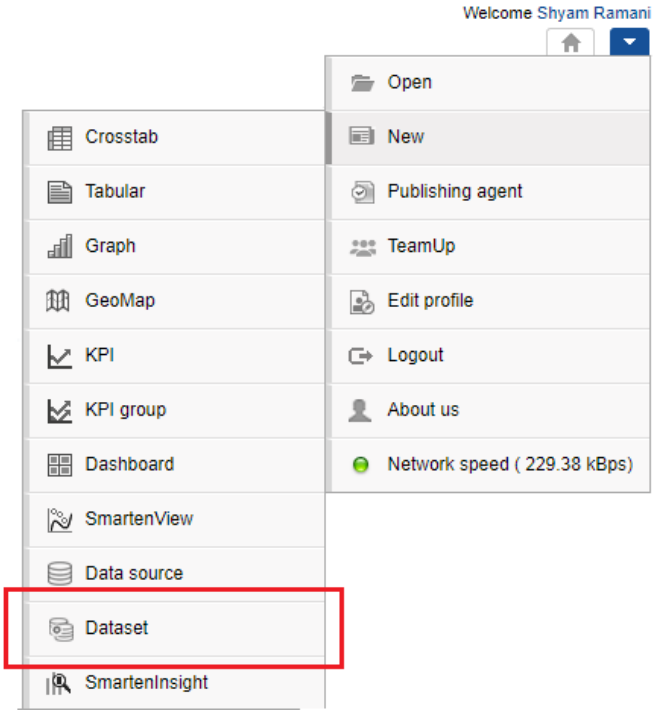
You can use an existing R Script data source and create a dataset to retrieve the required data from that data source.

About this task

Use this task to create a dataset using an existing R Script data source.

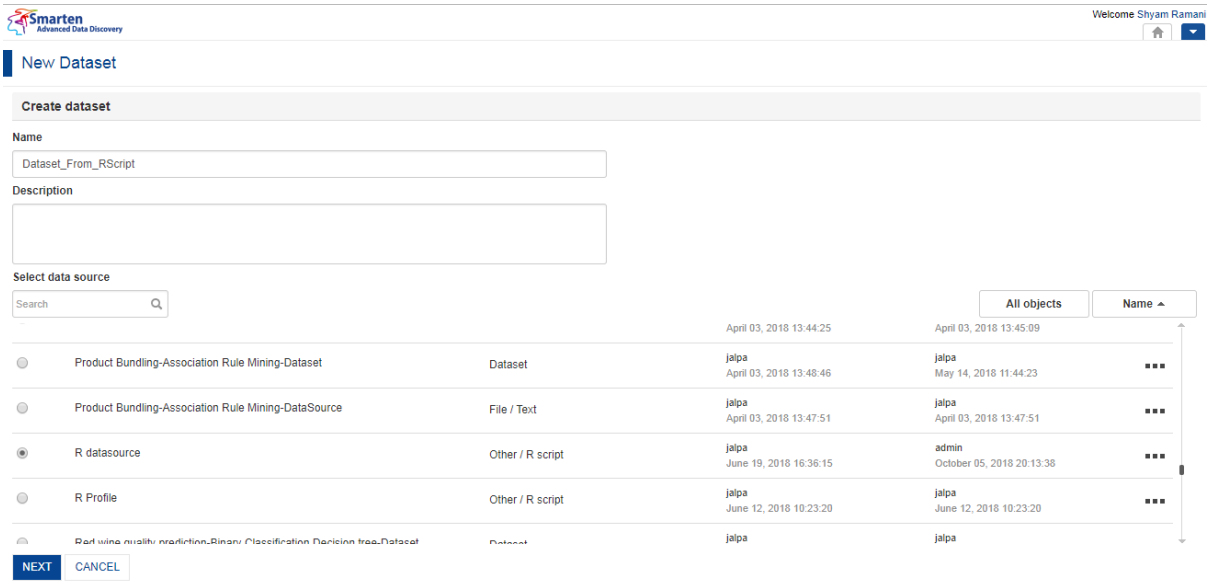
Procedure

1. Click **New** -> **Dataset** from the menu.



MENU OPTION—NEW DATASET

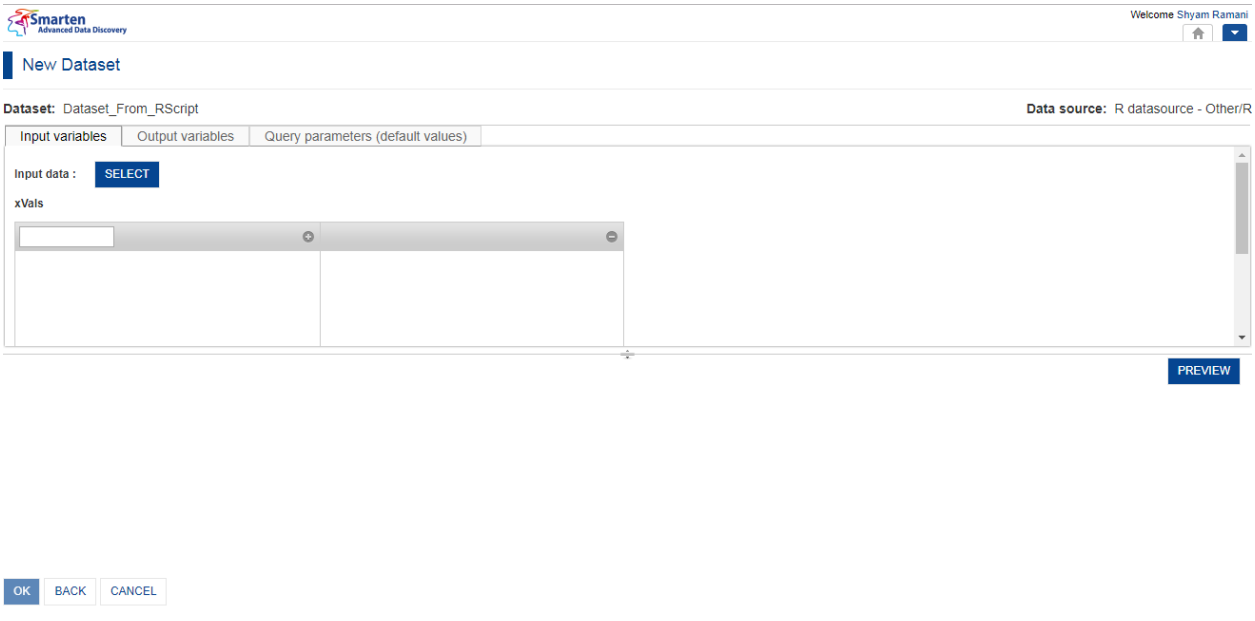
The system displays the **Create dataset** page.



CREATING A NEW DATASET—THE CREATE DATASET SCREEN

2. Enter a name and description for the dataset in the **Name** and **Description** fields.
3. Select the R Script profile you want to use to create a dataset from the list.
   Or,
   Enter the keyword to search for the R Script data source profile based on the keyword provided, and then press **Enter**.
   The system displays the R Script data source profiles that match the keyword you have provided.
4. Click **Next**.

   The system displays the **variable configuration** page. This page is rendered based on the Input variables, Output variables, and Query variables configured in the selected R Script data source.

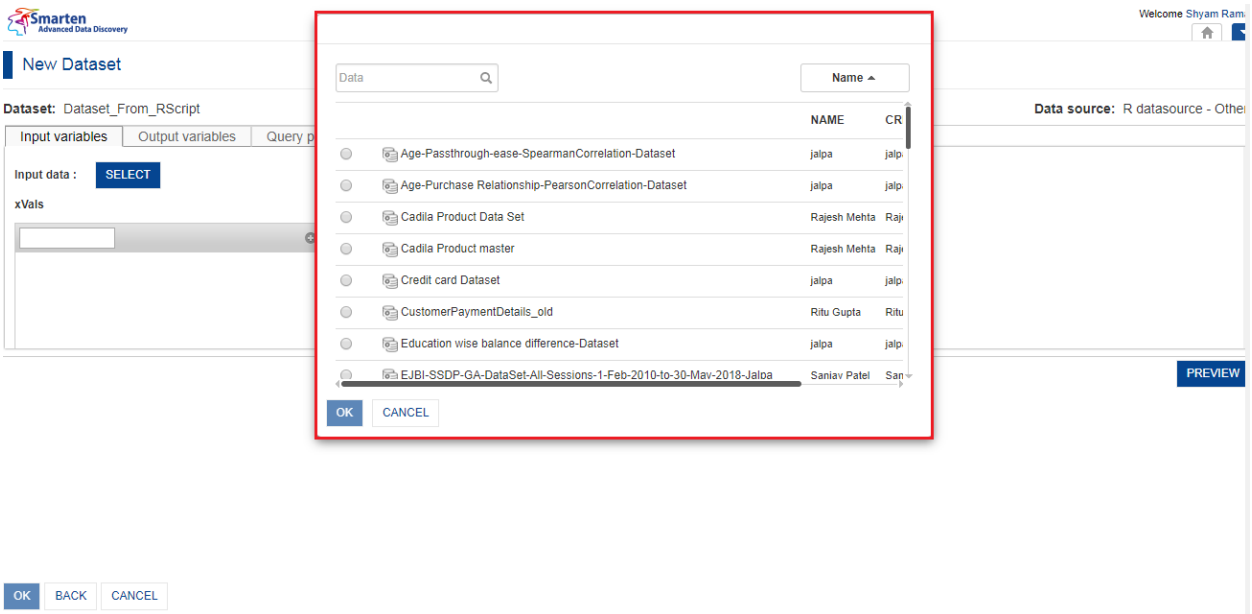


CREATING A DATASET—NEW DATASET PAGE

5. Input variables are the parameters in which values are passed to the R Script from Smarten.

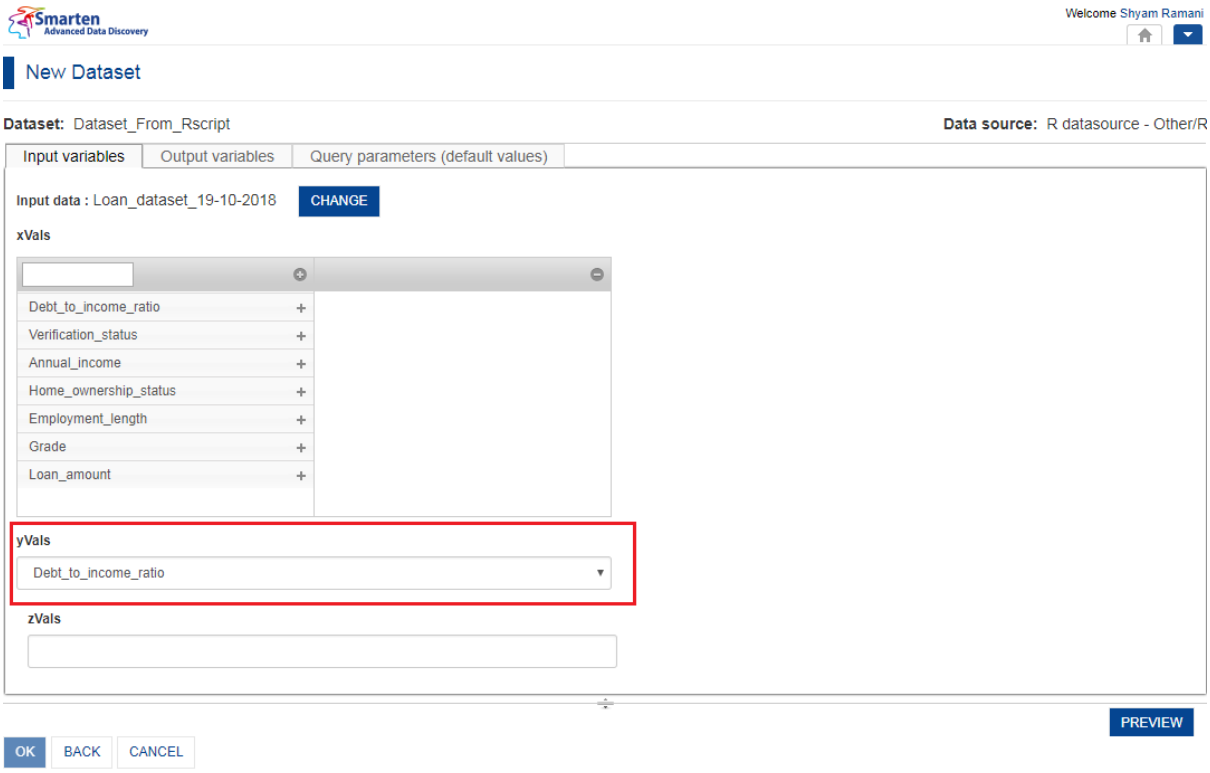
   The options available on the Input variables tab are displayed based on the Input variables configured in the R Script data source. The following elements are available based on the Input variables configured in the data source:

   - **Input data**: This option is available only if the **Single column** or **Multi columns** is configured as Input variables in the data source. You can click the **SELECT** button to select the dataset from which you want to pass values to the R Script.



CREATING A DATASET—SELECTING A DATASET

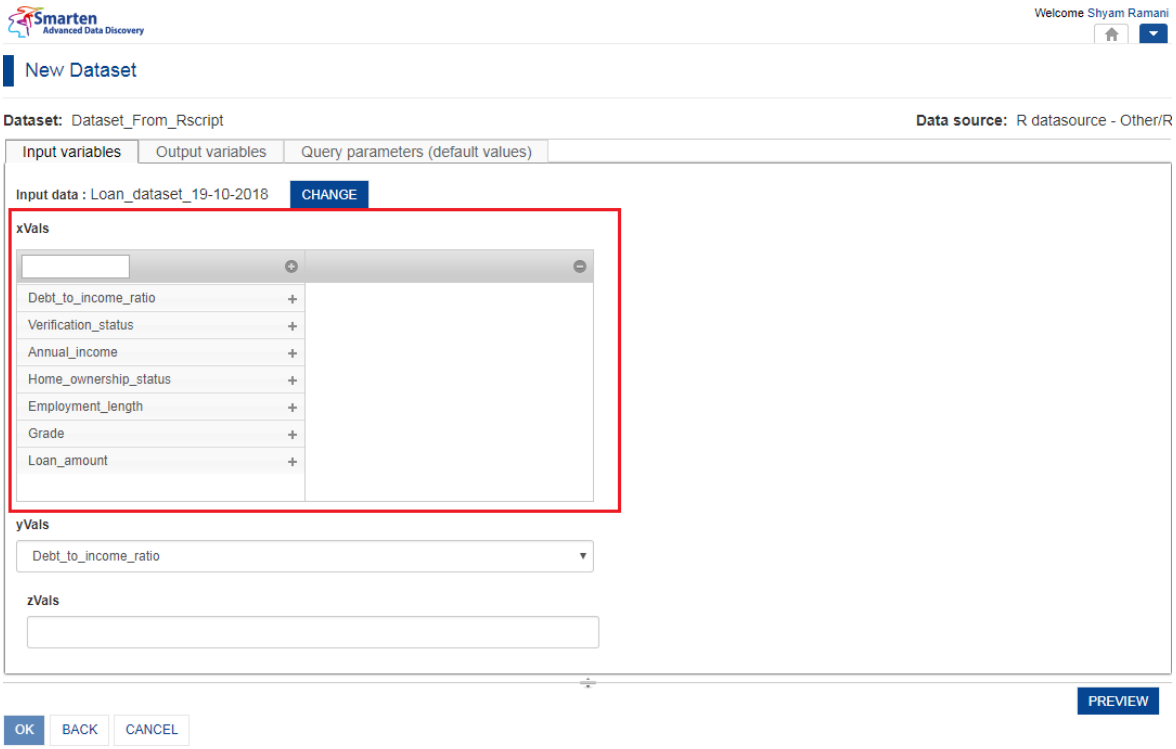
   - **Single column**: If this option is configured in the data source, the system displays a list that is populated with the columns available in the dataset you have selected in the **Input data**. You can select a column, and the system will pass data of that column to the R Script.



INPUT VARIABLES—SINGLE COLUMN

For example, **yVals** in the image above indicates the single column input type for the Input variable configured in the data source.

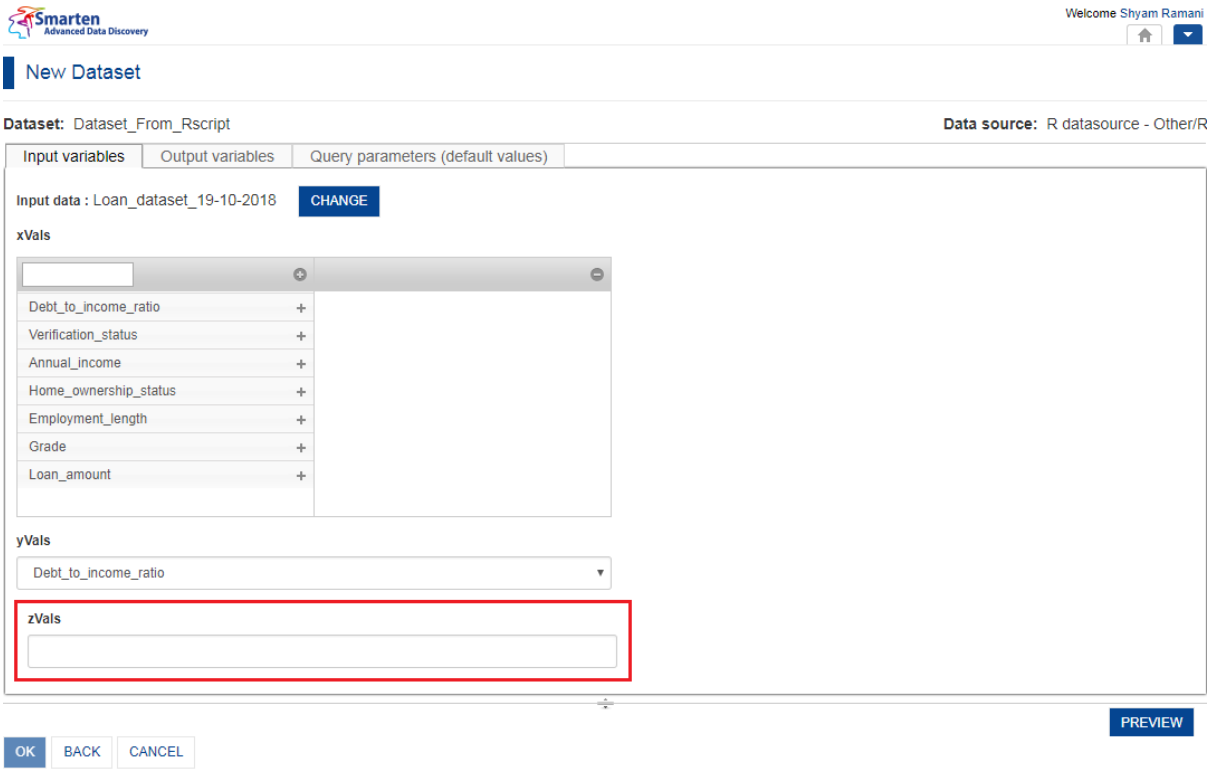
- **Multi columns**: If this option is configured in the data source, the system displays a list that is populated with the columns available in the dataset you have selected in the **Input data**. You can select more than one column, and the system will pass data of those columns to the R Script.



INPUT VARIABLES—MULTI COLUMNS

For example, **xVals** in the image above indicates the multi columns input type for the Input variable configured in the data source.

- **Single value**: If this option is configured in the data source, you can manually enter a value, and the system will pass that value to the R Script.

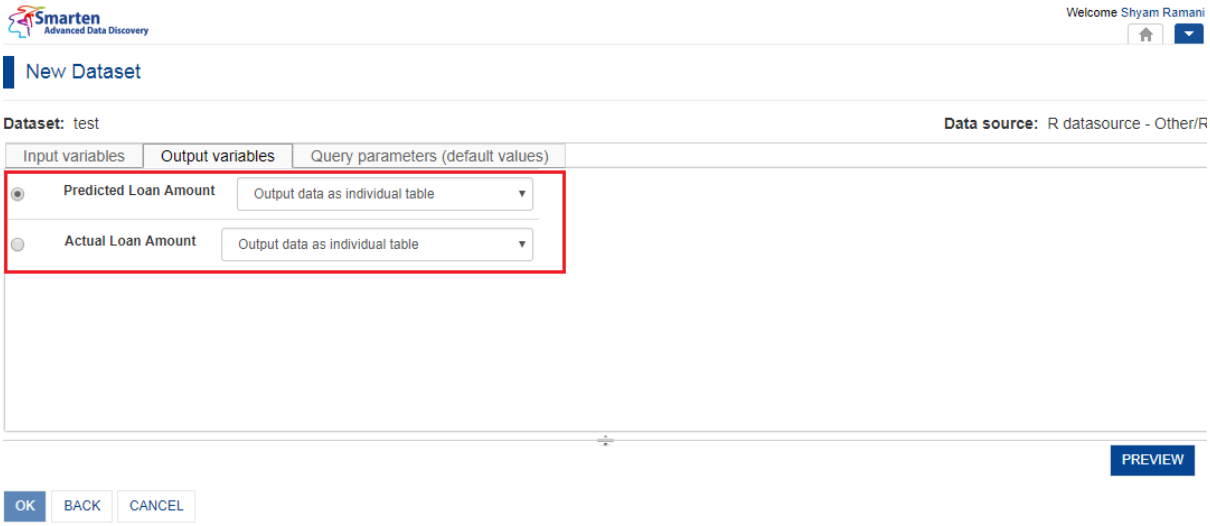


INPUT VARIABLES—SINGLE VALUE

For example, **zVals** in the image above indicates the single value input type for the Input variable configured in the data source.

6. Click the **Output Variables** tab. Output variables are the variables that are defined in the R Script configured in the data source. You can select the appropriate Output variable to create the dataset. The system retrieves data from the R Script based on the Output variable you have specified and creates the dataset. The following options are available for each Output variable configured:

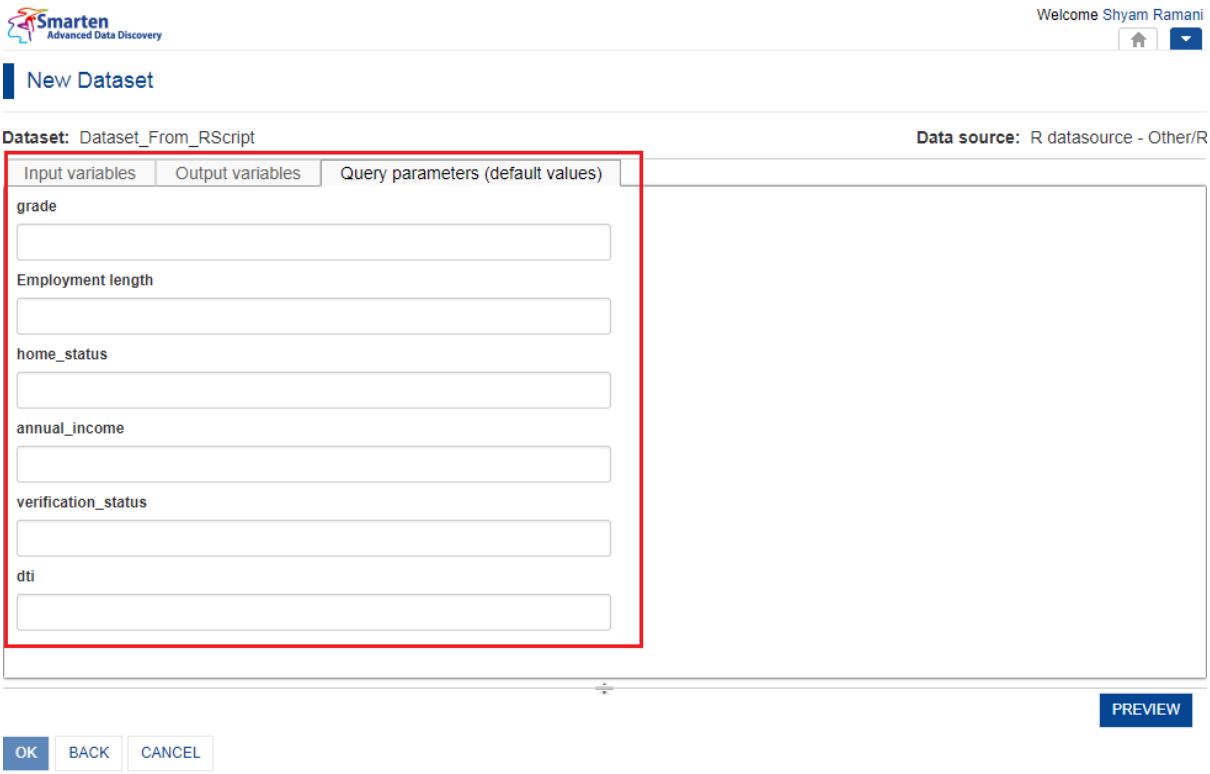
   - **Output data as an individual table**: Select this option to create the dataset from the output table.
   - **Append output data as a column**: Select this option to append the output data as a column along with the input data. The system appends the output data to the input data and then creates the dataset.
   - **Append output data as a row**: Select this option to append the output data as rows after the input data. The system appends the output data rows after the input data and then creates the dataset.



CREATING A NEW DATASET—SELECTING OUTPUT VARIABLES

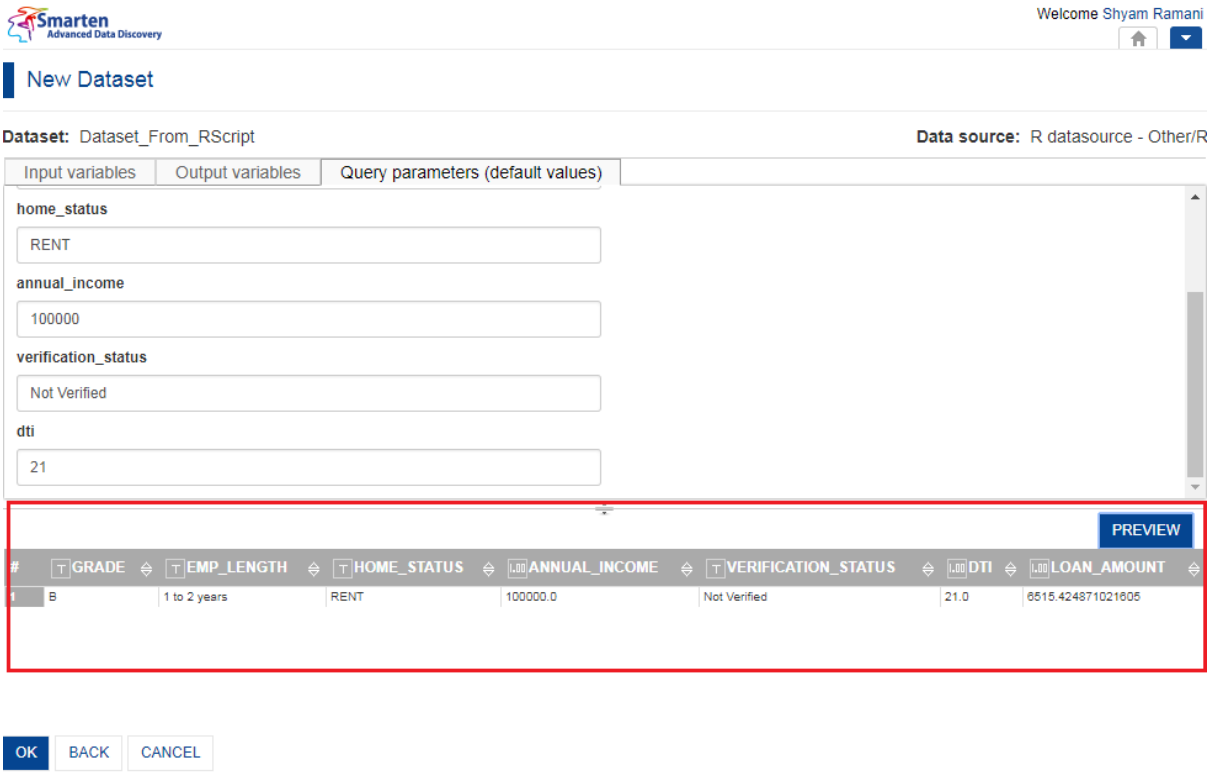
7. Click the **Query parameters** tab. Query variables are parameters that are used to query the R script model to retrieve data based on specific criteria. You can set default values for these parameters while creating a dataset.

   The options displayed in this tab are variables configured for query parameters in the data source that you have used to create the dataset.



CREATING A NEW DATASET—SPECIFYING QUERY PARAMETERS

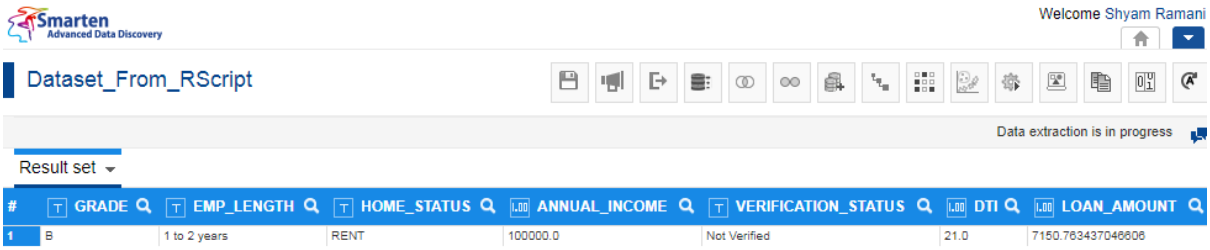
8. Click **PREVIEW** to view a preview of the data retrieved from the R script based on the parameters provided.



CREATING A NEW DATASET—PREVIEW OF THE SAMPLE DATA

9. Click **OK**.

The system displays a preview of the data based on the columns you have selected.



PREVIEW—DATASET CREATED USING AN EXISTING DATASET

The dataset is now available in the repository.

6.5 Creating a Dataset Using an SAP Data Source

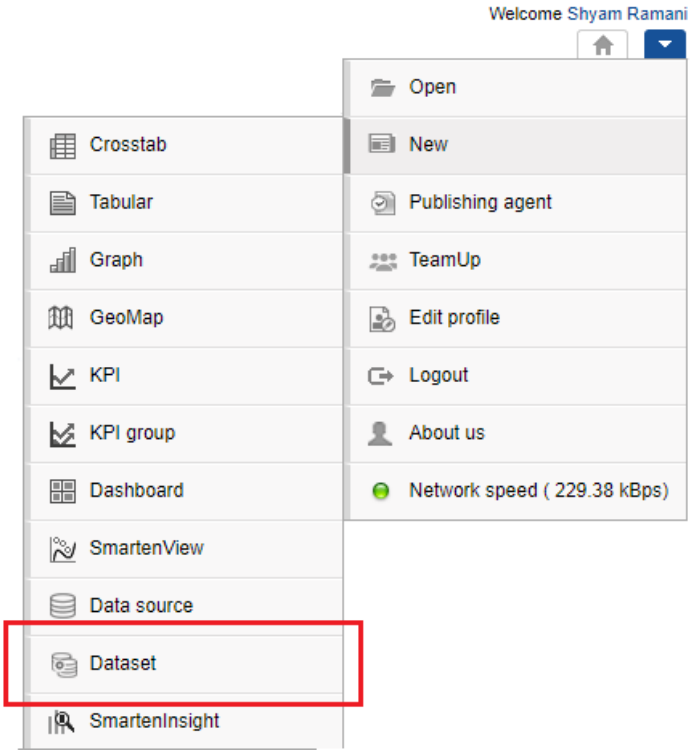
You can use an SAP profile to create a dataset and retrieve data from that profile.

About this task

Use this task to create a dataset using an SAP profile.

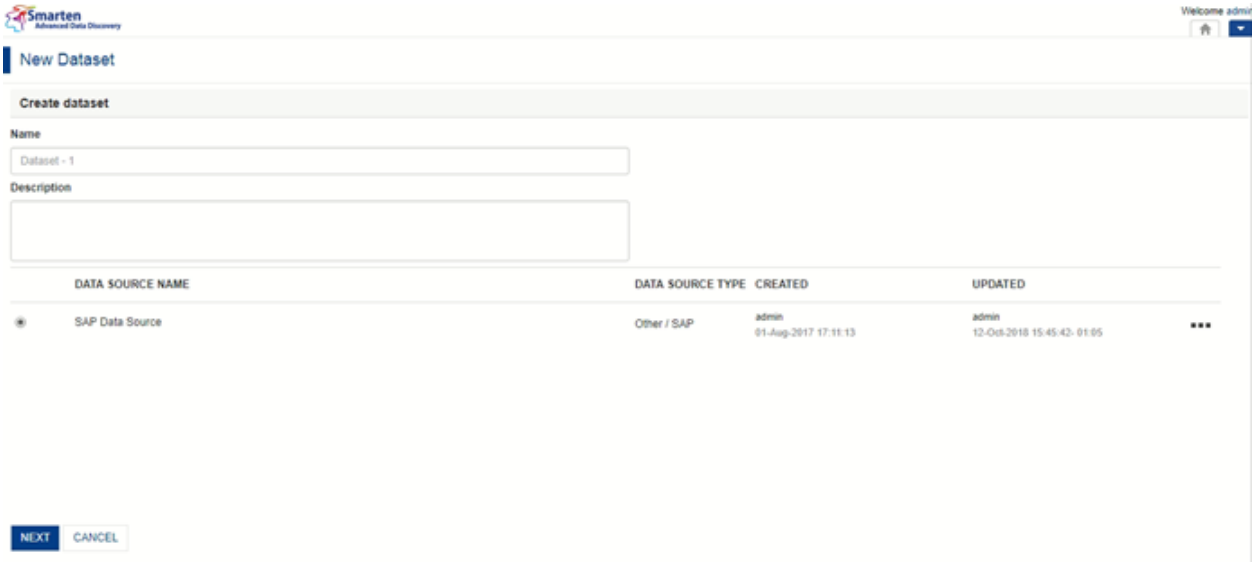
Procedure

1. Click **New** -> **Dataset** from the menu.



MENU OPTION—NEW DATASET

The system displays the **Create dataset** page.

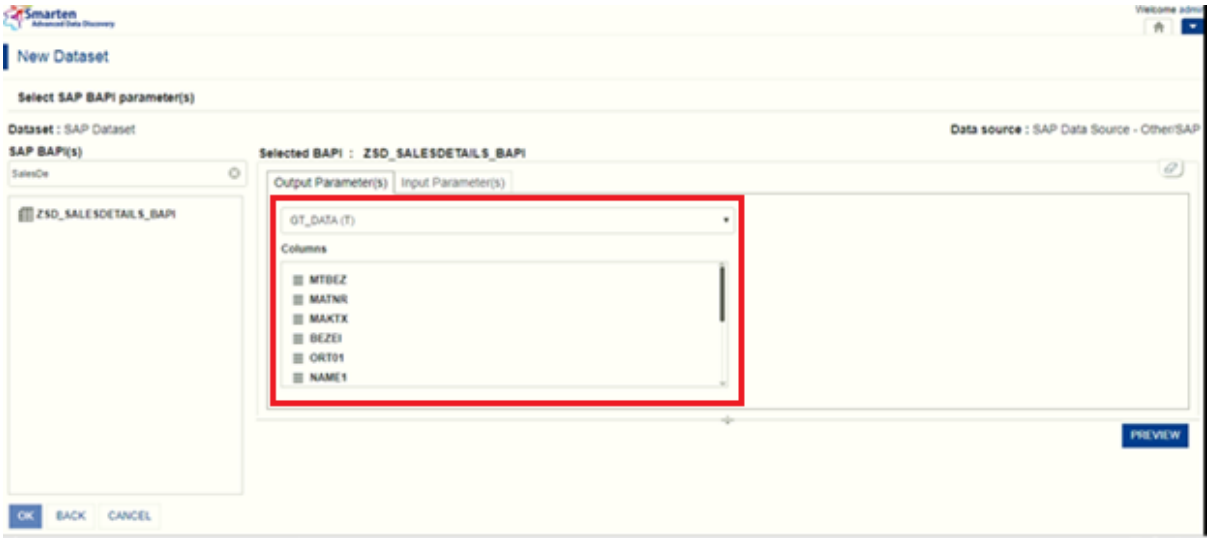


CREATING A NEW DATASET—THE CREATE DATASET SCREEN

2. Enter a name and description for the dataset in the **Name** and **Description** fields.
3. Select the SAP profile you want to use to create a dataset from the list.
   Or,
   Enter the keyword to search for the SAP profile based on the keyword provided, and then press **Enter**.
   The system displays the SAP profile that matches the keyword you have provided.
4. Select a SAP BAPI you want to use in the dataset from the **SAP BAPI(s)** list.

CREATING A NEW DATASET—SELECTING THE SAP BAPI(S)

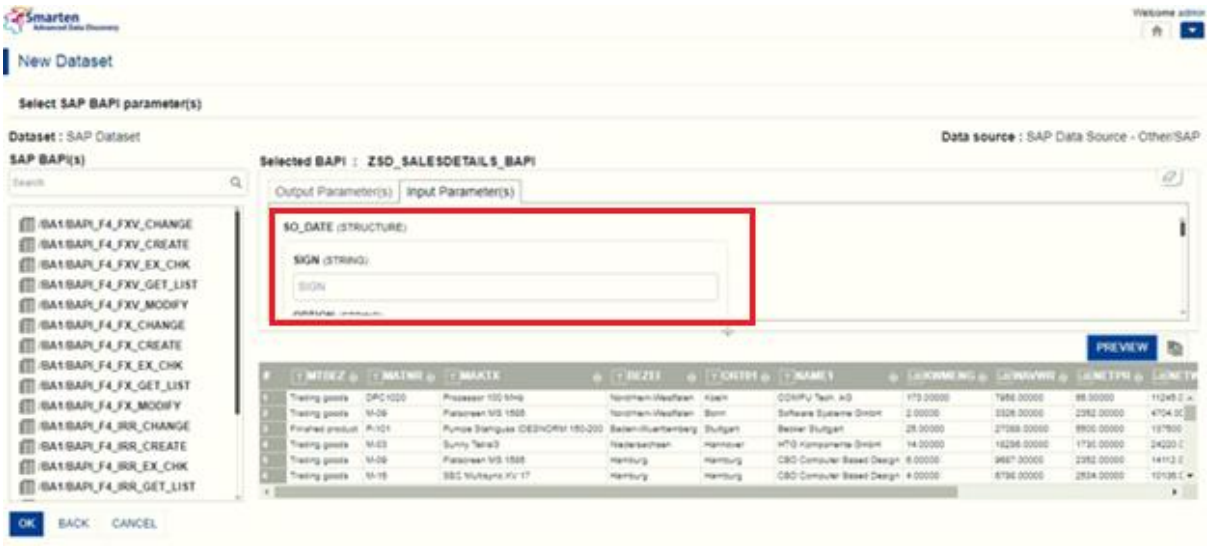
5. Click the **Output Parameter(s)** tab.
   The output parameter is the export parameter that a BAPI returns as an output table. A BAPI can have more than one output parameter. You can select the required output parameters that you want to use to create the dataset.
6. Select output parameters from the list within the **Output Parameter(s)** tab.
   The system displays values in the **Columns** section based on the output parameter you have selected from the list.



CREATE A DATASET—SELECTING OUTPUT PARAMETERS

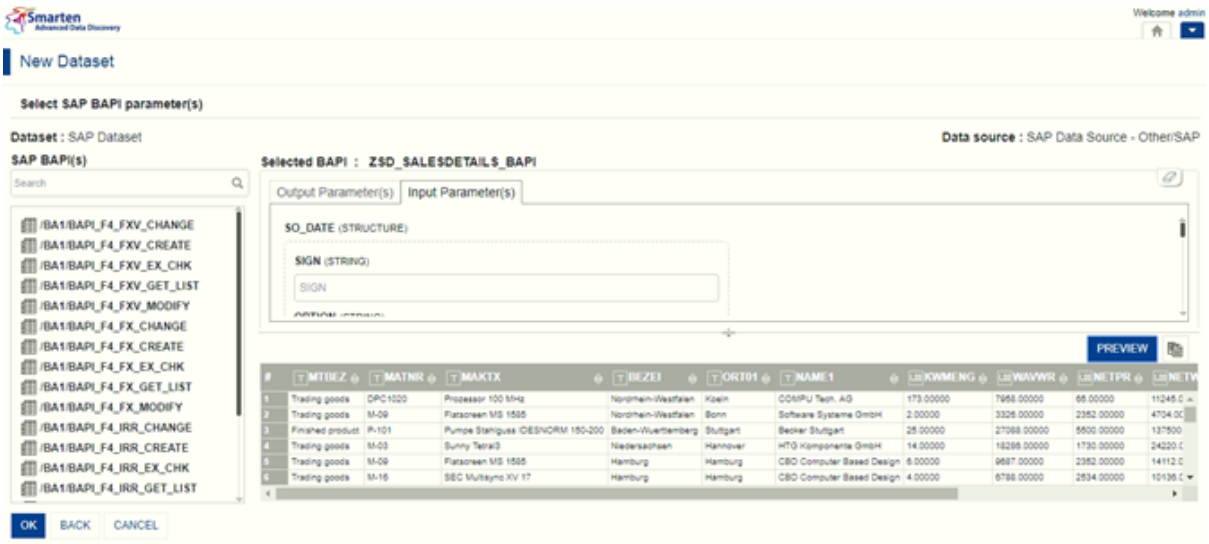
7. Click the **Input Parameter(s)** tab.

   The system passes the input parameters to the BAPI when they are executed, and based on the values of input parameters, the system returns the output. The system shows the appropriate UI control based on the data type of input parameter.
8. Select the input parameters from the **Input Parameter(s)** tab.
   The options available within the tab are based on the SAP BAPI you have selected.



CREATE A DATASET—SELECTING INPUT PARAMETERS

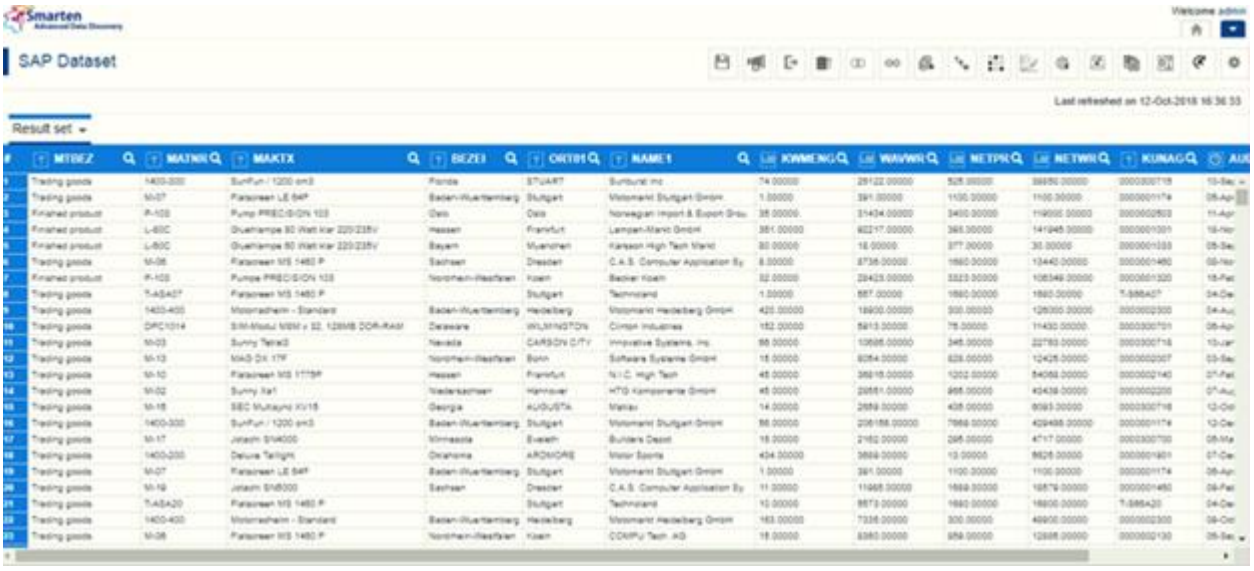
9. Click **Preview** to view a preview of the data retrieved based on the output parameters you have specified for the selected BAPI.



CREATE A DATASET—PREVIEW DATA RETRIEVED

10. Click **OK**.

The system creates a dataset based on the criteria you have provided. You can verify the data and columns available in the dataset.



CREATING A DATASET—COLUMNS AVAILABLE IN THE EXISTING DATASET

The dataset is now available in the repository.

6.5.1 Data Types Supported by Smarten

The following table lists the data types of SAP ABAP script that are compatible with Smarten data types. Smarten typecasts the data type of the BAPI output column as mentioned in the following table. Any data type that is not mentioned in the table is converted to null values.

ABAP Type	Description	Smarten Data Type
C	Character	String
N	Numerical Character	String
P	Binary Coded Decimal	Big Decimal
I	4-byte Integer	Int
B	1-byte Integer	Int
S	2-byte Integer	Int
F	Float	Double
D	Date	Date
T	Time	Date
decfloat16	Decimal floating point 8 bytes (IEEE 754r)	Big Decimal
decfloat34	Decimal floating point 16 bytes (IEEE 754r)	Big Decimal
g	String (variable length)	String

6.6 Creating a Dataset Using a Dataset

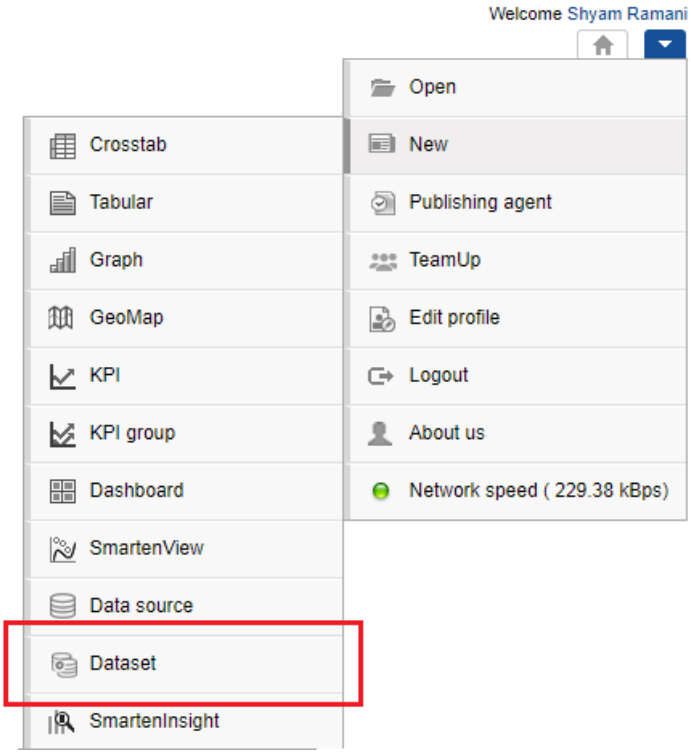
You can use an existing dataset and create a dataset to retrieve the required data from that dataset.

About this task

Use this task to create a dataset using an existing dataset.

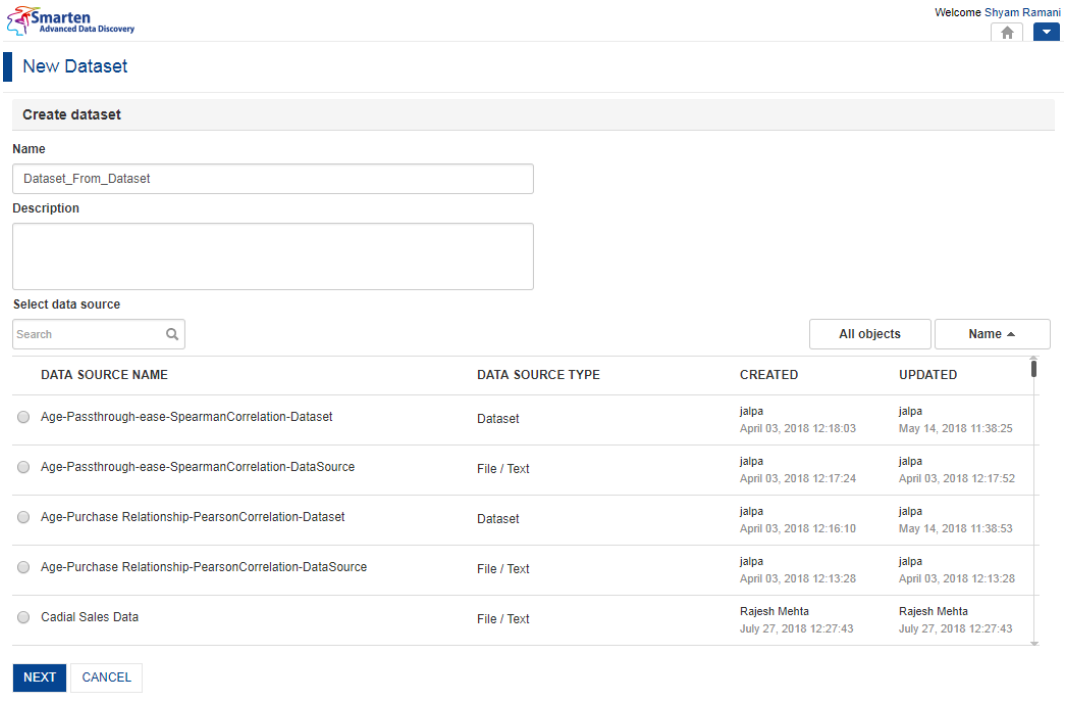
Procedure

1. Click **New** -> **Dataset** from the menu.



MENU OPTION—NEW DATASET

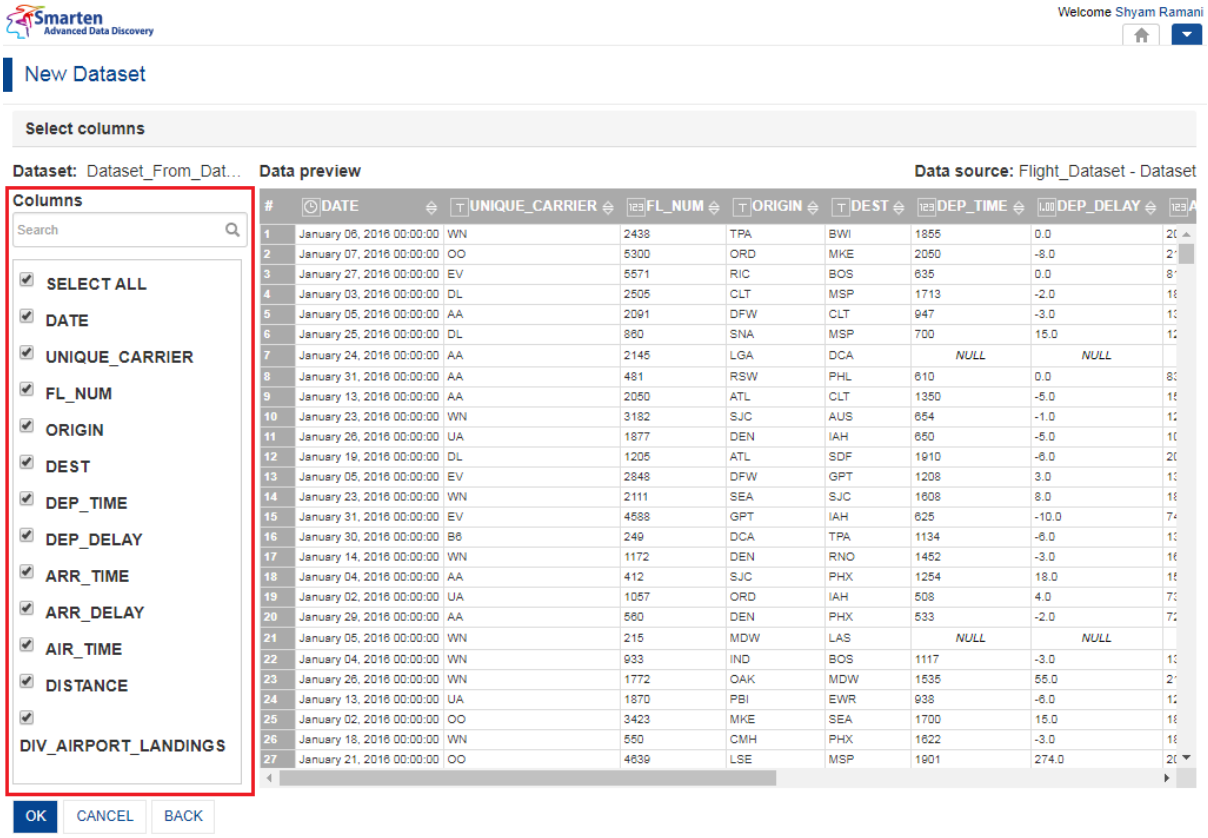
The system displays the **Create dataset** page.



CREATING A NEW DATASET—THE CREATE DATASET SCREEN

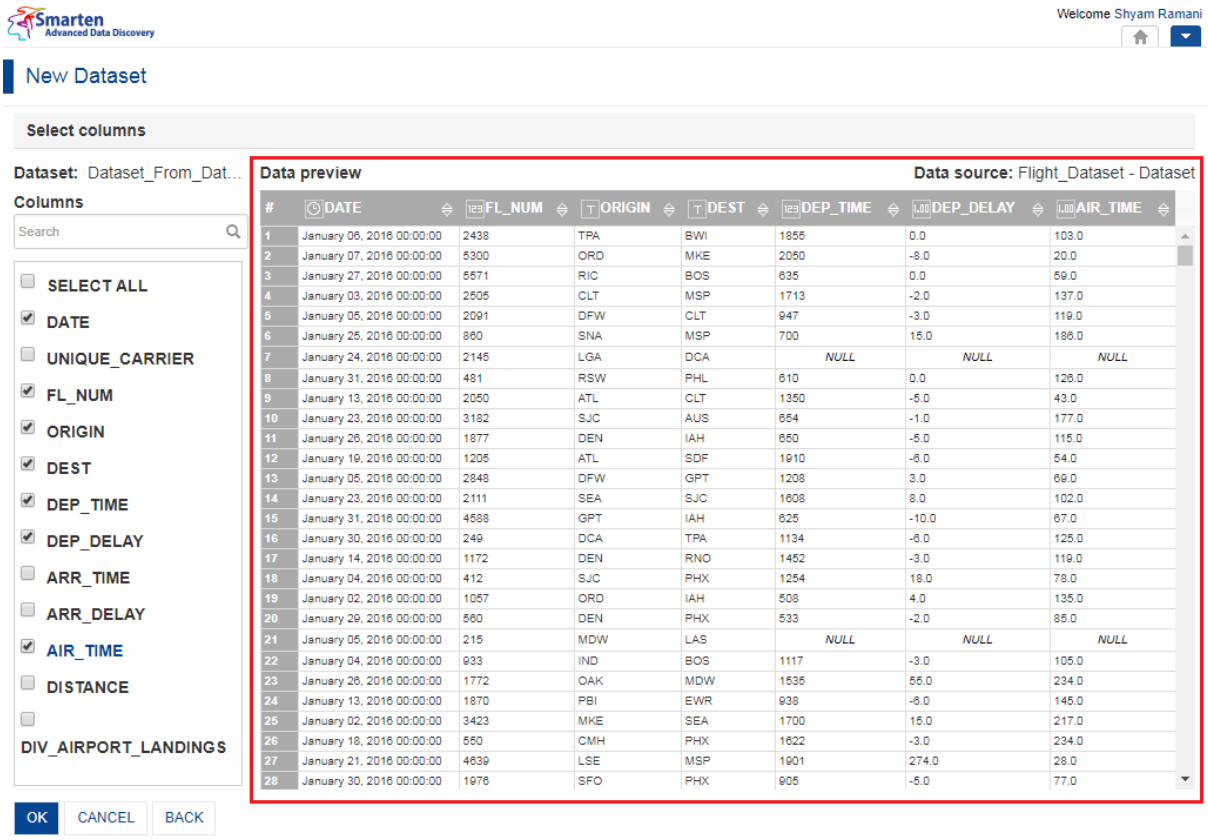
2. Enter a name and description for the dataset in the **Name** and **Description** fields.
3. Select the file type profile you want to use to create a dataset from the list.
   Or,
   Enter the keyword to search for the file type data source profile based on the keyword provided, and then press **Enter**.
   The system displays the file type data source profiles that match the keyword you have provided.
4. Click **Next**.

   The system displays the **Select columns** page.



CREATING A DATASET—COLUMNS AVAILABLE IN THE EXISTING DATASET

5. Select the columns that you want to use in the dataset from the left pane.
   Or,
   Enter the keyword to search for the columns you want to use in the dataset in the **Columns** field, and then press **Enter**.



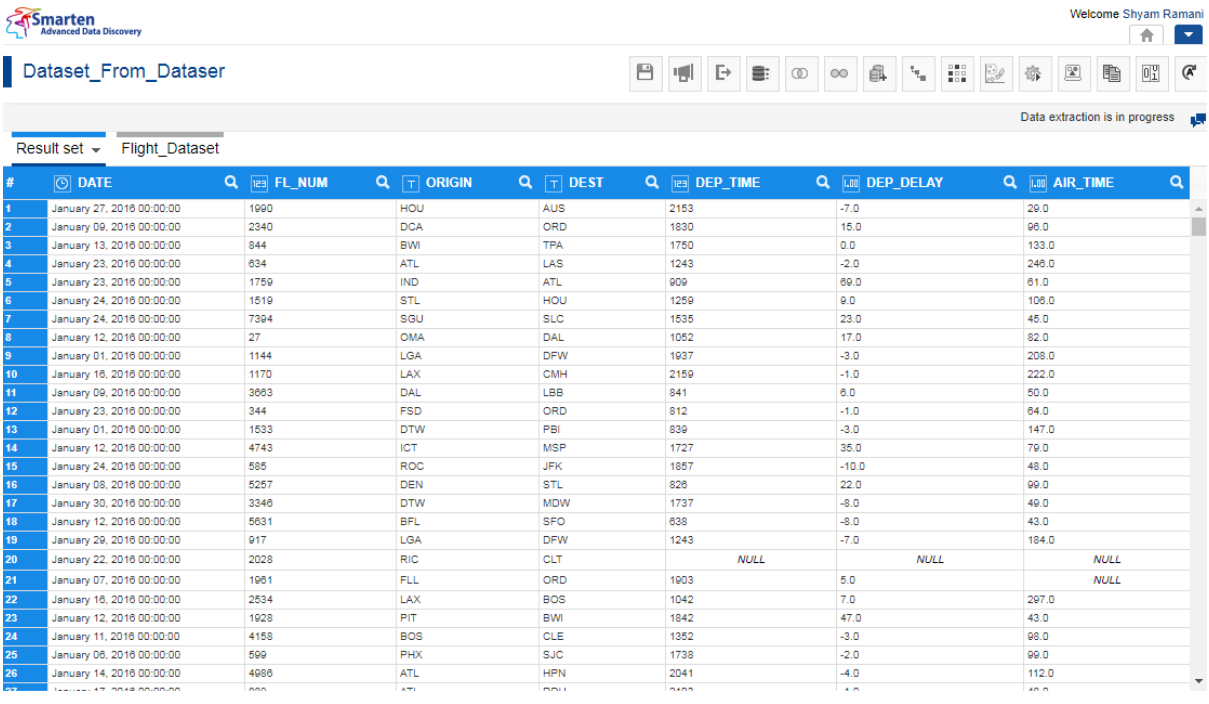
#	DATE	FL_NUM	ORIGIN	DEST	DEP_TIME	DEP_DELAY	AIR_TIME
1	January 06, 2016 00:00:00	2438	TPA	BWI	1855	0.0	103.0
2	January 07, 2016 00:00:00	5300	ORD	MKE	2050	-8.0	20.0
3	January 27, 2016 00:00:00	5571	RIC	BOS	635	0.0	59.0
4	January 03, 2016 00:00:00	2505	CLT	MSP	1713	-2.0	137.0
5	January 05, 2016 00:00:00	2091	DFW	CLT	947	-3.0	119.0
6	January 25, 2016 00:00:00	860	SNA	MSP	700	15.0	186.0
7	January 24, 2016 00:00:00	2145	LGA	DCA	NULL	NULL	NULL
8	January 31, 2016 00:00:00	481	RSW	PHL	610	0.0	126.0
9	January 13, 2016 00:00:00	2050	ATL	CLT	1350	-5.0	43.0
10	January 23, 2016 00:00:00	3182	SJC	AUS	654	-1.0	177.0
11	January 26, 2016 00:00:00	1877	DEN	IAH	650	-5.0	115.0
12	January 19, 2016 00:00:00	1205	ATL	SDF	1910	-6.0	54.0
13	January 05, 2016 00:00:00	2848	DFW	GPT	1208	3.0	69.0
14	January 23, 2016 00:00:00	2111	SEA	SJC	1608	8.0	102.0
15	January 31, 2016 00:00:00	4588	GPT	IAH	625	-10.0	67.0
16	January 30, 2016 00:00:00	249	DCA	TPA	1134	-6.0	125.0
17	January 14, 2016 00:00:00	1172	DEN	RNO	1452	-3.0	119.0
18	January 04, 2016 00:00:00	412	SJC	PHX	1254	18.0	78.0
19	January 02, 2016 00:00:00	1057	ORD	IAH	508	4.0	135.0
20	January 29, 2016 00:00:00	560	DEN	PHX	533	-2.0	85.0
21	January 05, 2016 00:00:00	215	MDW	LAS	NULL	NULL	NULL
22	January 04, 2016 00:00:00	933	IND	BOS	1117	-3.0	105.0
23	January 26, 2016 00:00:00	1772	OAK	MDW	1535	55.0	234.0
24	January 13, 2016 00:00:00	1870	PBI	EWR	938	-6.0	145.0
25	January 02, 2016 00:00:00	3423	MKE	SEA	1700	15.0	217.0
26	January 18, 2016 00:00:00	550	CMH	PHX	1622	-3.0	234.0
27	January 21, 2016 00:00:00	4639	LSE	MSP	1901	274.0	28.0
28	January 30, 2016 00:00:00	1976	SFO	PHX	905	-5.0	77.0

CREATING A DATASET—DATA FROM THE SELECTED COLUMNS

The system displays the columns you have selected along with the data available in those columns in the **Data preview** pane.

6. Click **OK**.

The system displays a preview of the data based on the columns you have selected.



#	DATE	FL_NUM	ORIGIN	DEST	DEP_TIME	DEP_DELAY	AIR_TIME
1	January 27, 2016 00:00:00	1990	HOU	AUS	2153	-7.0	29.0
2	January 09, 2016 00:00:00	2340	DCA	ORD	1830	15.0	96.0
3	January 13, 2016 00:00:00	844	BWI	TPA	1750	0.0	133.0
4	January 23, 2016 00:00:00	634	ATL	LAS	1243	-2.0	246.0
5	January 23, 2016 00:00:00	1759	IND	ATL	909	69.0	61.0
6	January 24, 2016 00:00:00	1519	STL	HOU	1259	9.0	106.0
7	January 24, 2016 00:00:00	7394	SGU	SLC	1535	23.0	45.0
8	January 12, 2016 00:00:00	27	OMA	DAL	1052	17.0	82.0
9	January 01, 2016 00:00:00	1144	LGA	DFW	1937	-3.0	208.0
10	January 16, 2016 00:00:00	1170	LAX	CMH	2159	-1.0	222.0
11	January 09, 2016 00:00:00	3663	DAL	LBB	841	6.0	50.0
12	January 23, 2016 00:00:00	344	FSD	ORD	812	-1.0	64.0
13	January 01, 2016 00:00:00	1533	DTW	PBI	839	-3.0	147.0
14	January 12, 2016 00:00:00	4743	ICT	MSP	1727	35.0	79.0
15	January 24, 2016 00:00:00	585	ROC	JFK	1857	-10.0	48.0
16	January 08, 2016 00:00:00	5257	DEN	STL	826	22.0	99.0
17	January 30, 2016 00:00:00	3346	DTW	MDW	1737	-8.0	49.0
18	January 12, 2016 00:00:00	5631	BFL	SFO	638	-8.0	43.0
19	January 29, 2016 00:00:00	917	LGA	DFW	1243	-7.0	184.0
20	January 22, 2016 00:00:00	2028	RIC	CLT	NULL	NULL	NULL
21	January 07, 2016 00:00:00	1961	FLL	ORD	1903	5.0	NULL
22	January 16, 2016 00:00:00	2534	LAX	BOS	1042	7.0	297.0
23	January 12, 2016 00:00:00	1928	PIT	BWI	1842	47.0	43.0
24	January 11, 2016 00:00:00	4158	BOS	CLE	1352	-3.0	98.0
25	January 06, 2016 00:00:00	599	PHX	SJC	1738	-2.0	99.0
26	January 14, 2016 00:00:00	4986	ATL	HPN	2041	-4.0	112.0

PREVIEW—DATASET CREATED USING AN EXISTING DATASET

The dataset is now available in the repository.

7 Working with Dataset

Dataset is a source of analysis-ready data in the columnar structure. You can perform various operations on a dataset such as highlighting data, apply transformation and cleaning operations, manage columns, and much more.

7.1 Dataset Management

Managing a Dataset involves providing access rights for a Dataset; opening, editing, deleting, and copying a Dataset; changing the Data Source for a Dataset, and getting a dataset IT certified. A Dataset can be managed either by the user who created it or by the Administrator.

Reference: **Concept Manual > Dataset Management**

7.1.1 Opening a Dataset

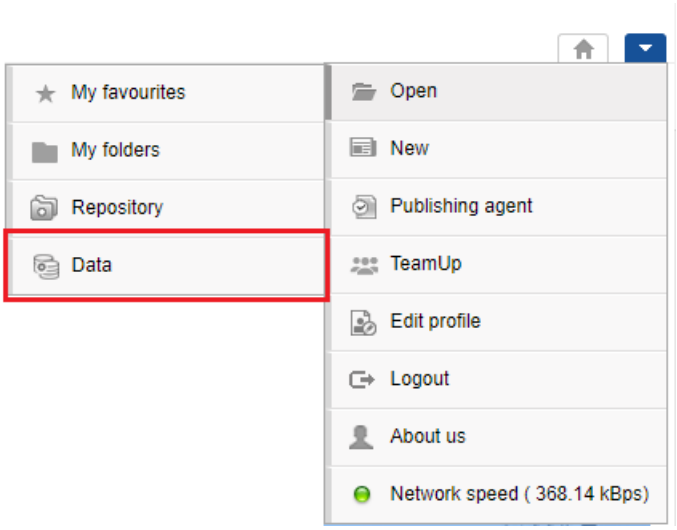
You must open a dataset to be able to perform any operation on that dataset.

About this task

Use this task to open an existing dataset.

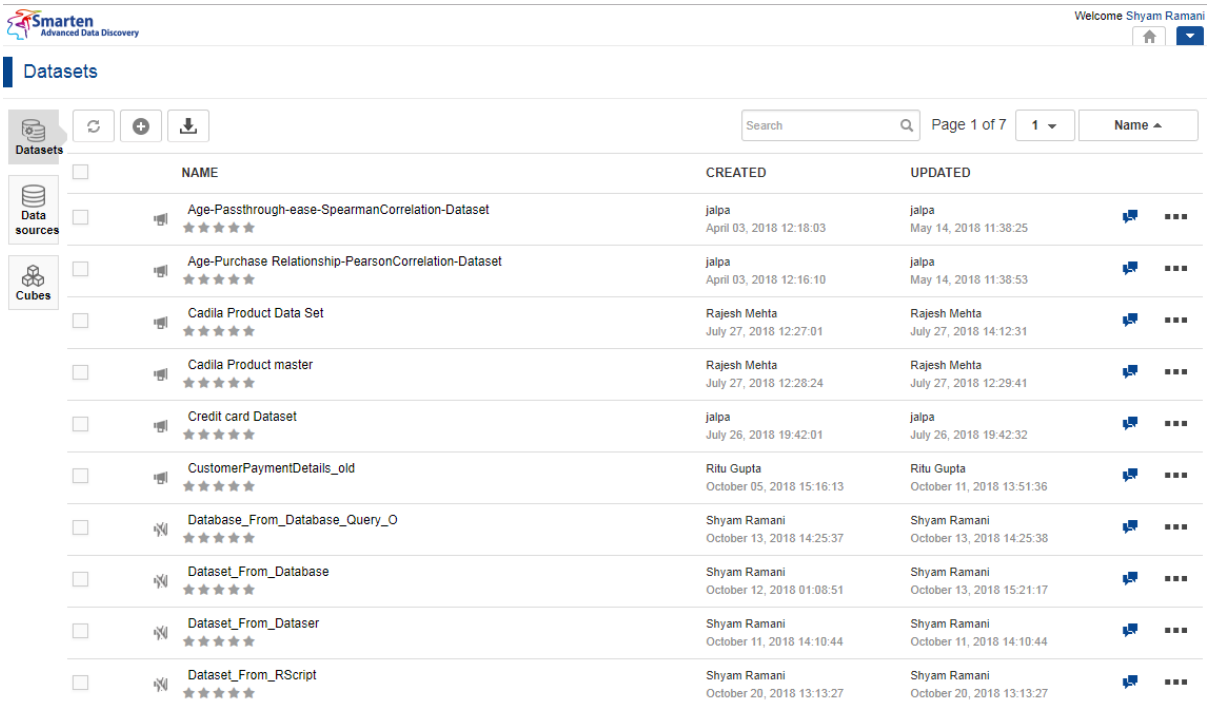
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

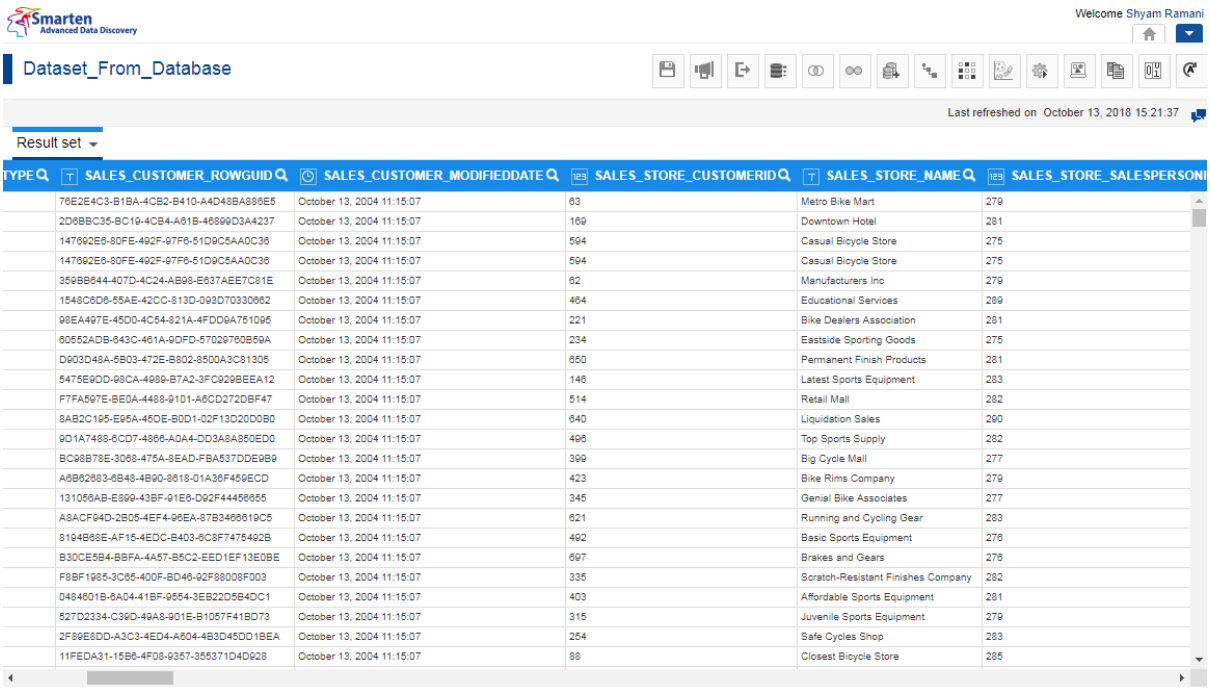
The system displays the following page.



OPEN ADATASET – SELECTING A DATASET

2. Click **Datasets**, and then select the dataset you want to open.

The system displays the dataset along with the data from the data source.



OPEN DATASET – DATA AVAILABLE IN AN EXISTING DATASET

7.1.2 Editing a Dataset

This feature enables authorized users to edit a Dataset. Users can change the name, description, data source, and columns of the dataset. Users can also change the data source of the dataset.

Reference: **Concept Manual > Dataset Management > Editing a Dataset**

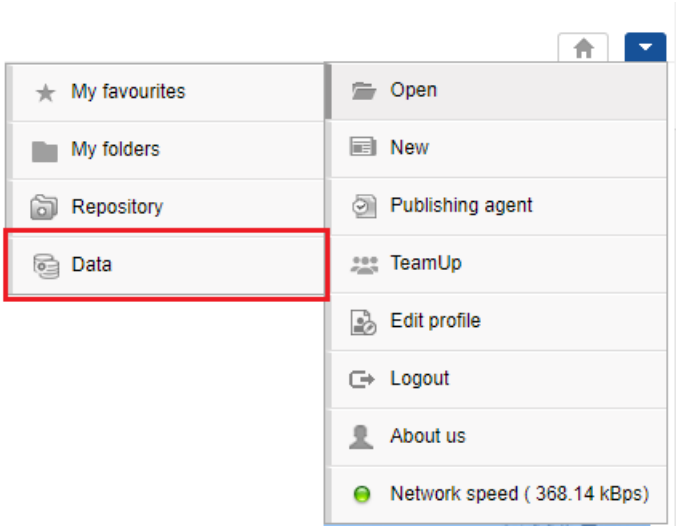
7.1.2.1 Editing Name and Description of a Dataset

About this task

Use this task to editname and description of a dataset.

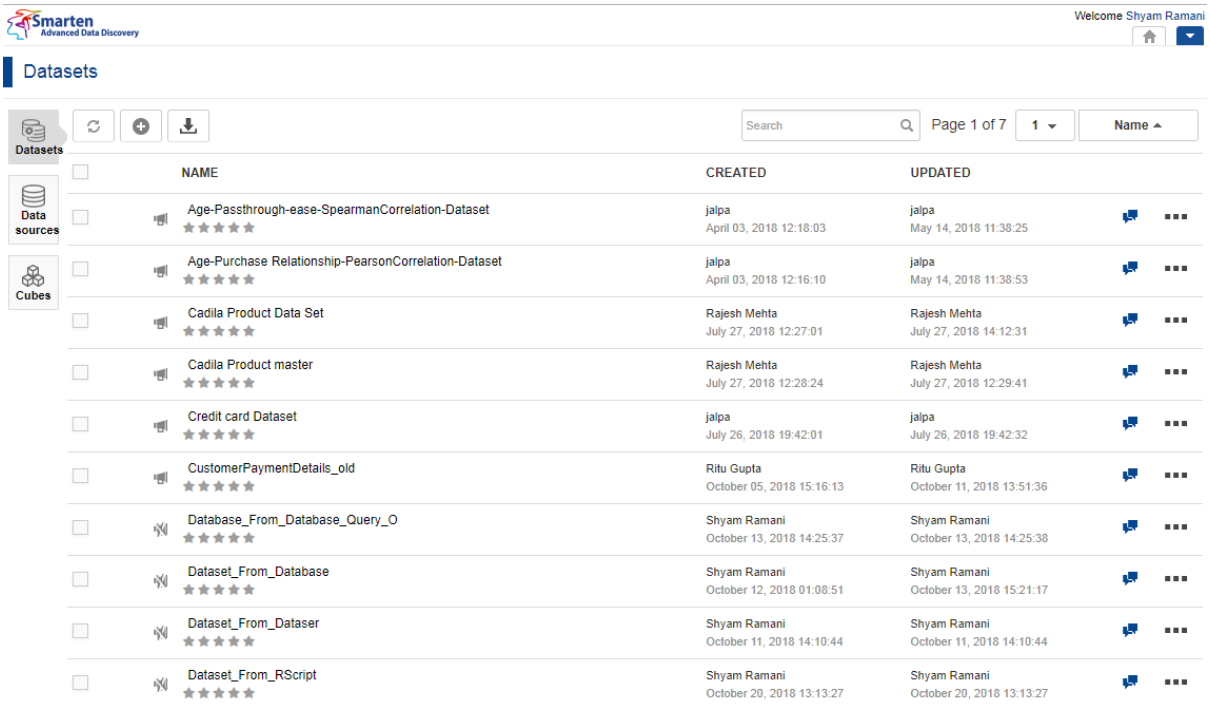
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

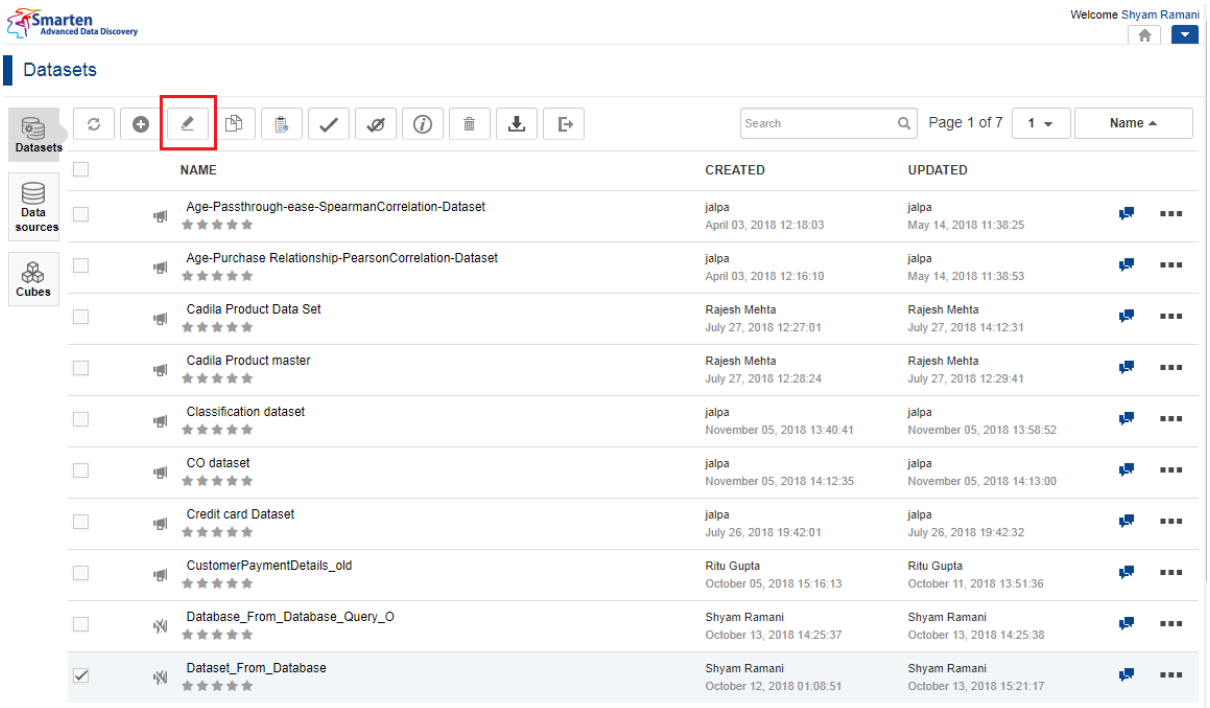
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

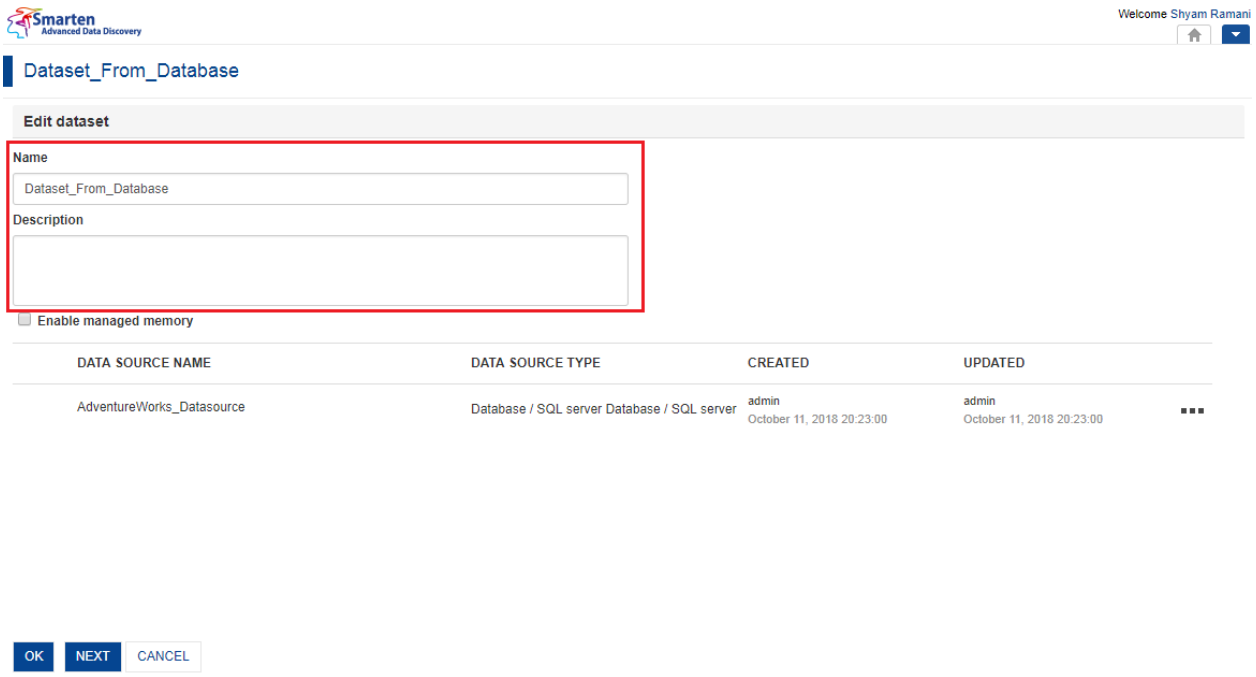
2. Select the checkbox adjacent to the dataset you want to edit.
3. Click the edit icon.

The system displays the **Edit dataset** page.



MANAGING DATASET – EDITING A DATASET

4. You can provide a new name and description for the dataset in the **Name** and **Description** boxes, and then click **OK**.



EDITING A DATASET – EDITING NAME AND DESCRIPTION

7.1.2.2 Enabling Managed Memory for a Dataset

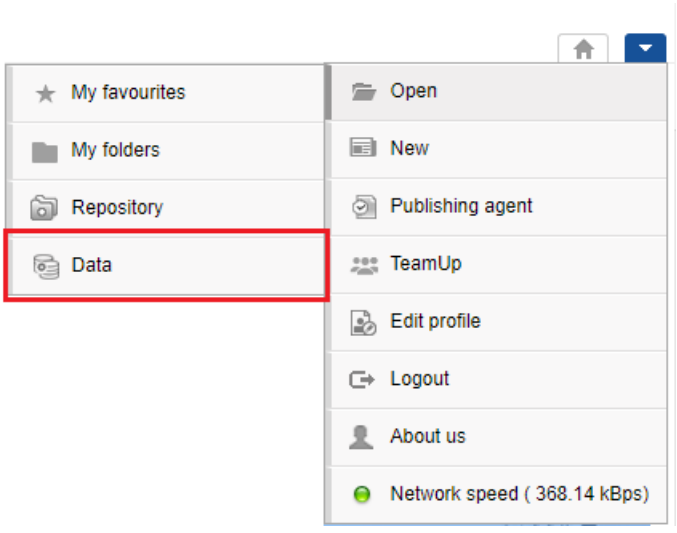
About this task

Use this task to allow the system to use managed memory computing architecture for this dataset.

Reference: **Smarten The Working of Managed Memory Computing**
Reference: **Smarten Managed Memory Computing Concept**

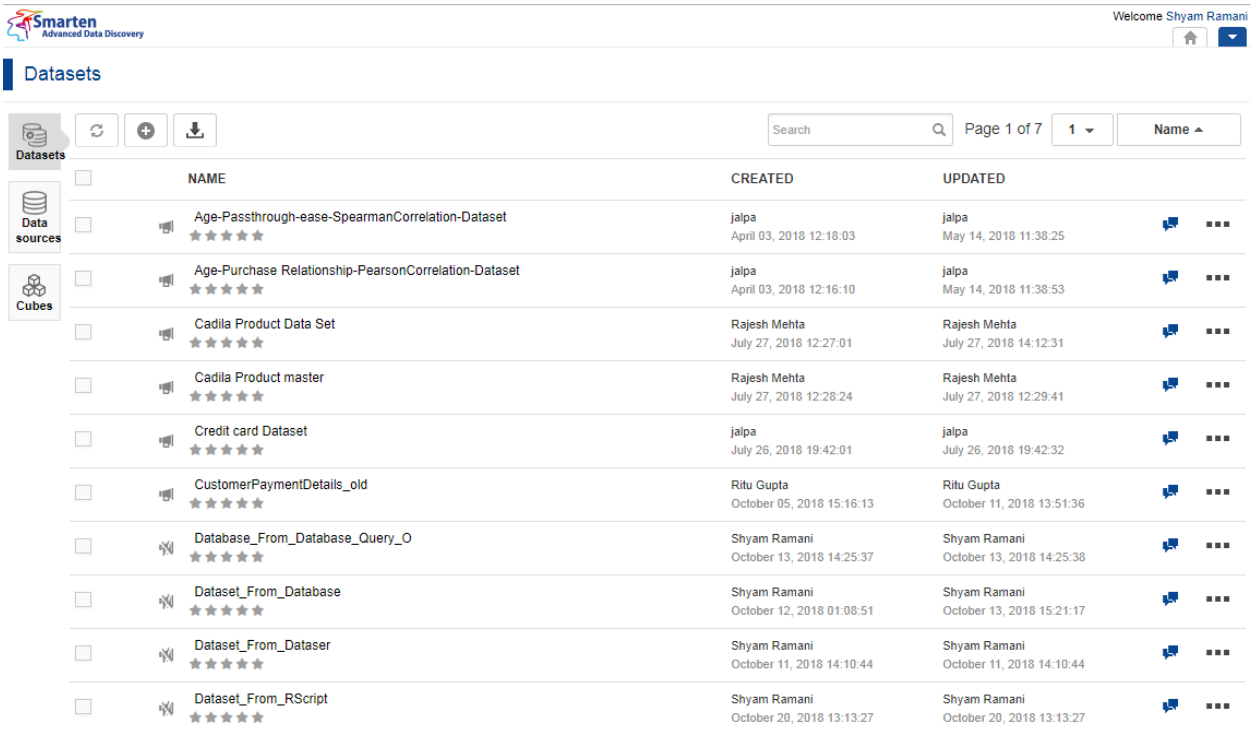
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

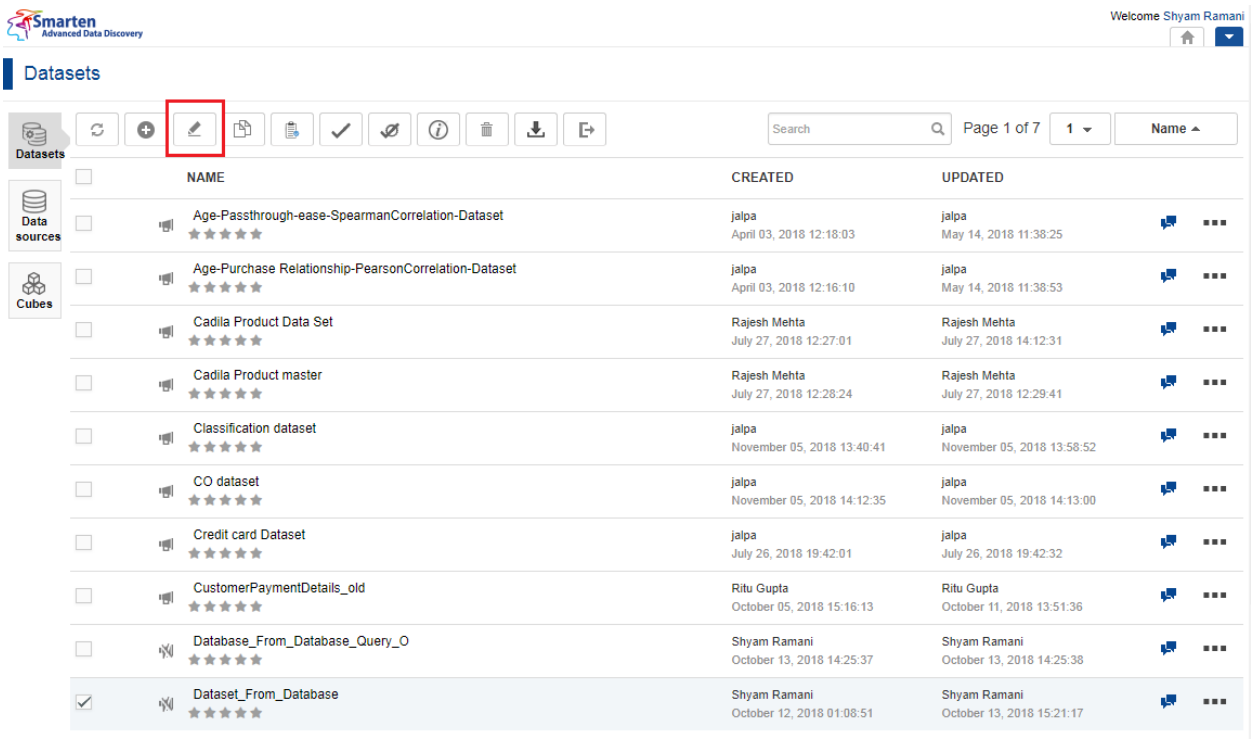
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

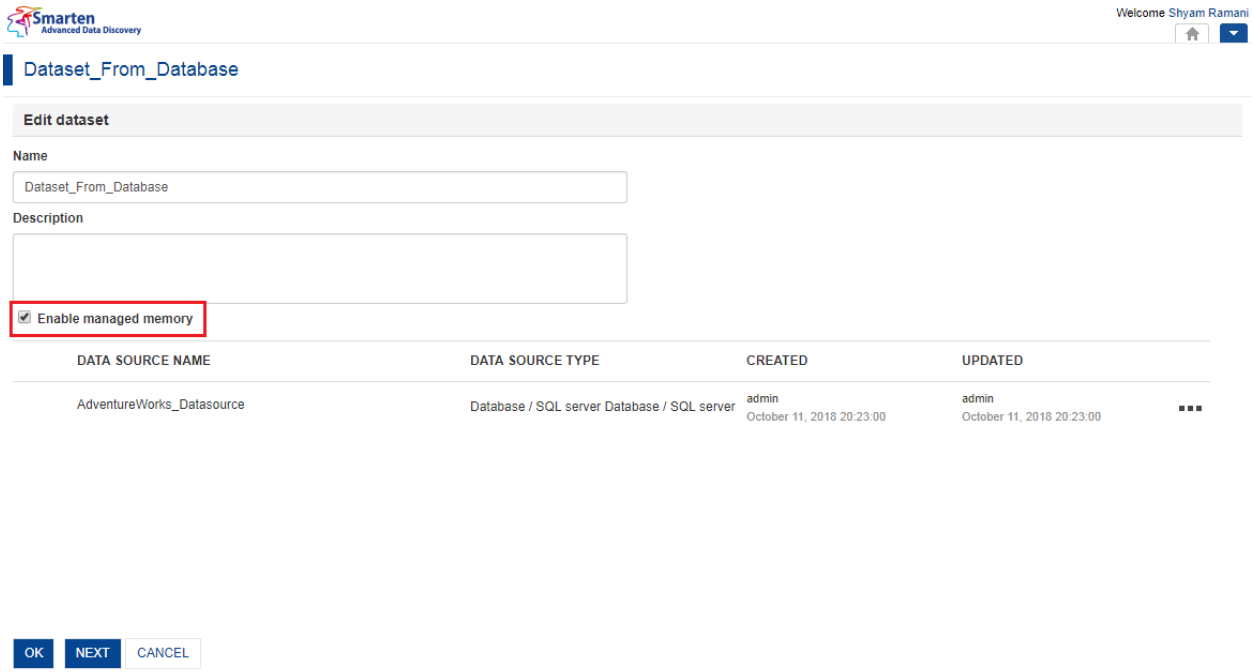
2. Select the checkbox adjacent to the dataset you want to edit.
3. Click the edit icon.

The system displays the **Edit dataset** page.



MANAGING DATASET – EDITING A DATASET

4. You can select the **Enable managed memory**checkbox.



EDITING A DATASET – OPTION TO ENABLE MANAGED MEMORY

7.1.2.3 Changing the Data Source of a Dataset

There are two scenarios when you need to change data source of a dataset.

Scenario 1:

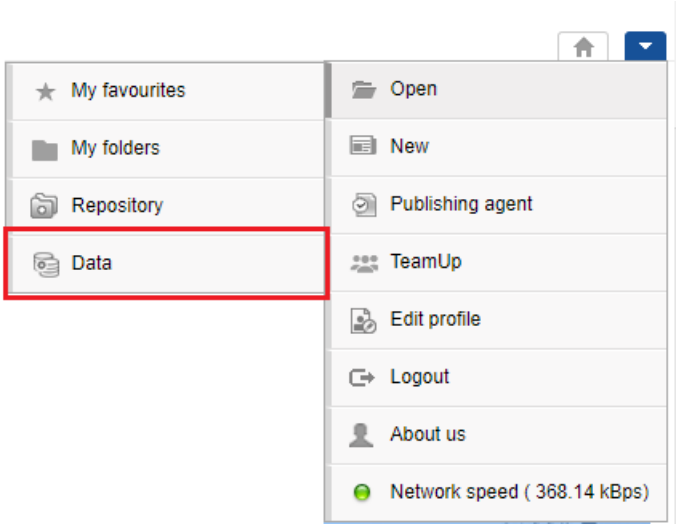
You need to change data source of a dataset if the data source is deleted or not available.

About this task

Use this task to change the data source of a dataset when the data source is deleted or not available.

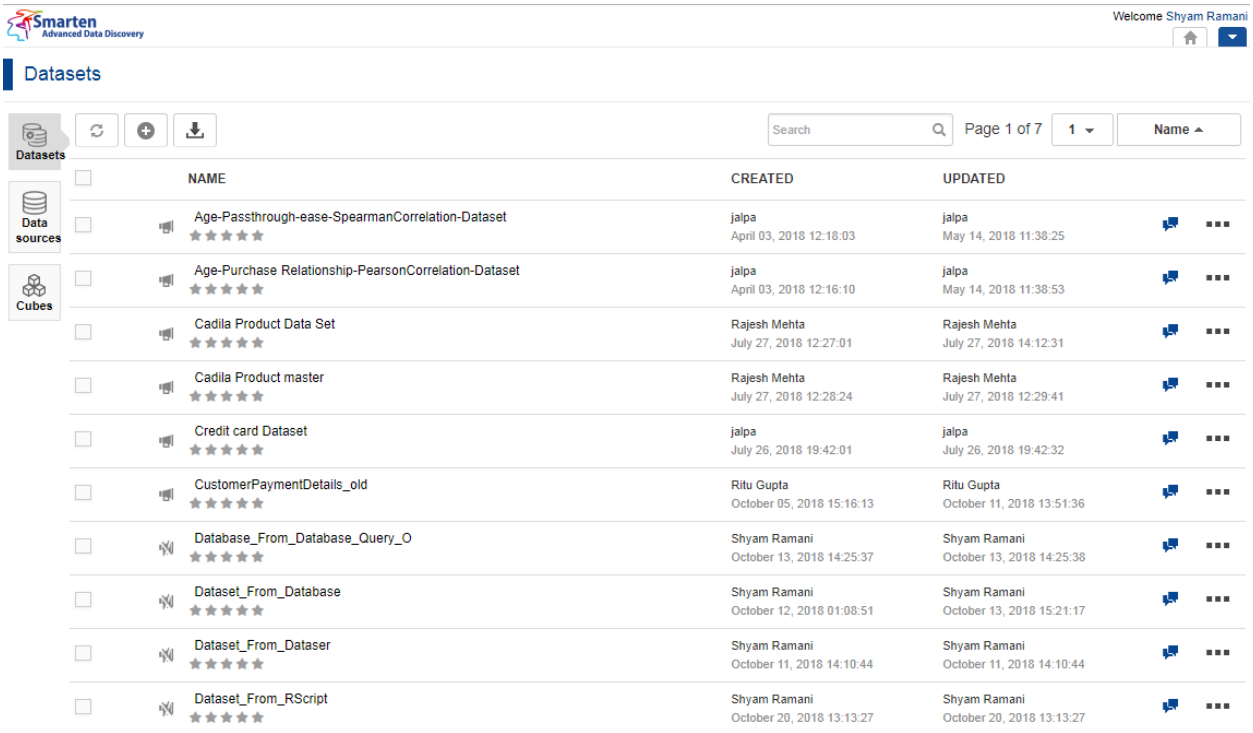
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

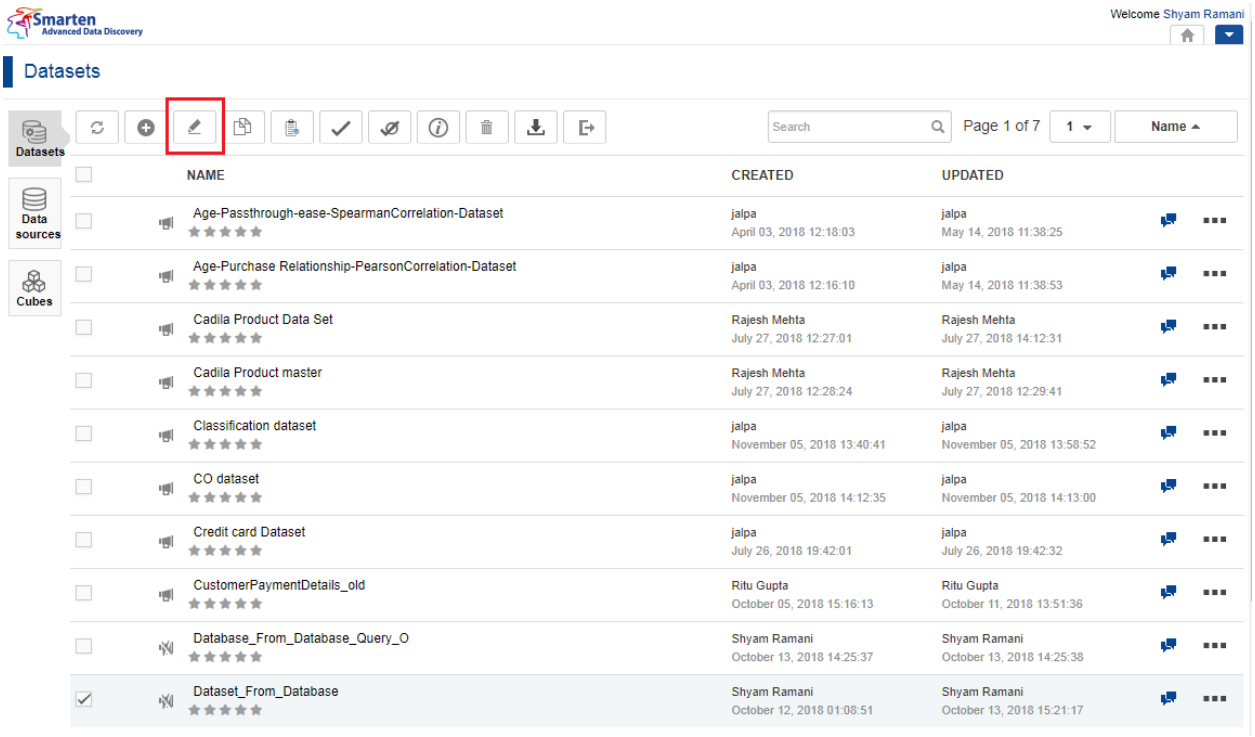
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

2. Select the checkbox adjacent to the dataset for which you want to change the data source.
3. Click the edit icon.

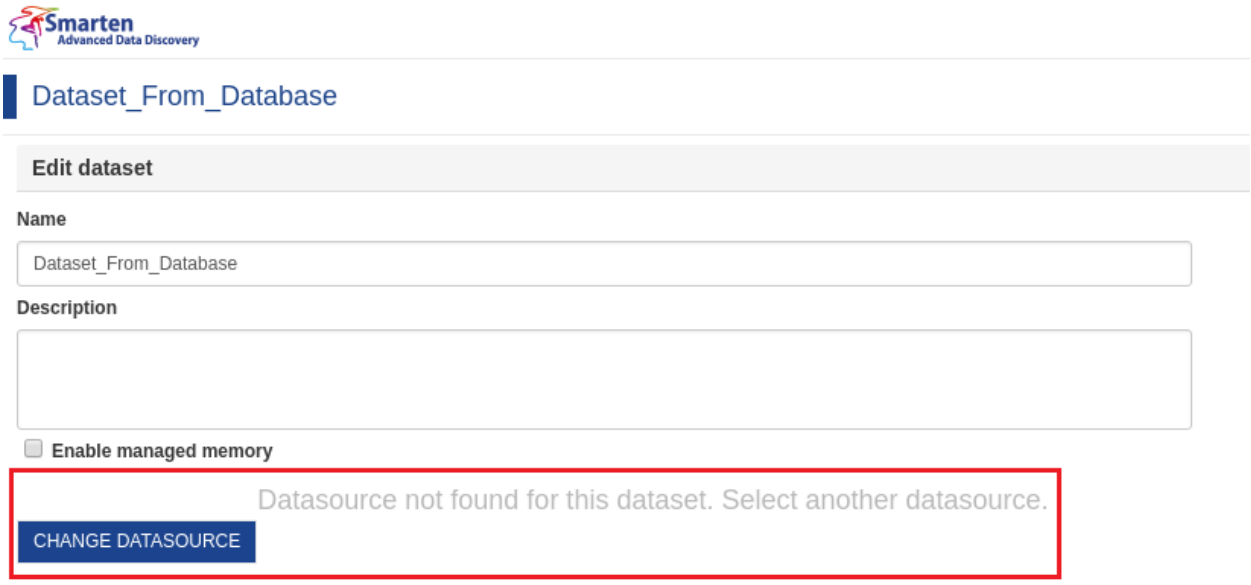
The system displays the **Edit dataset** page.



MANAGING DATASET – EDITING A DATASET

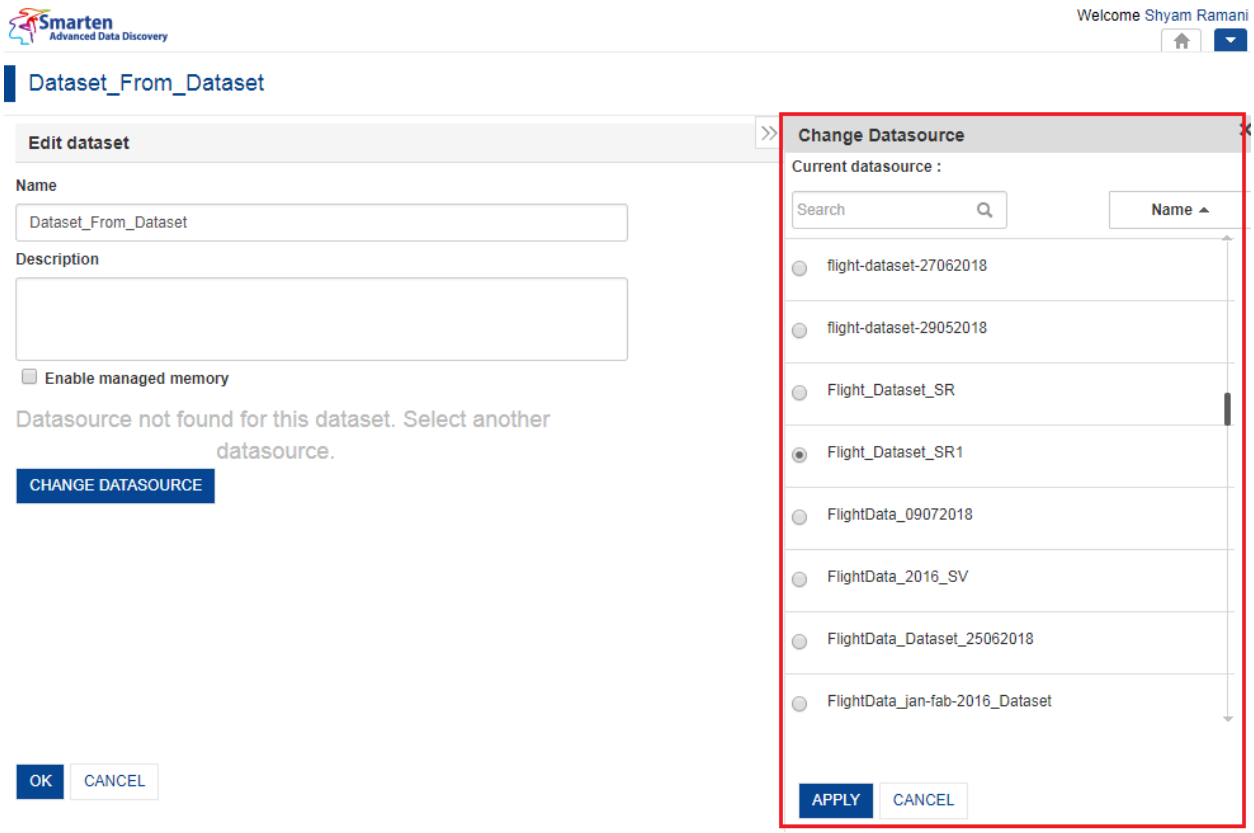
4. Click **CHANGE DATASOURCE**, if the data source used to create the dataset is deleted and no longer available, the system displays the message and provides an option to change the data source.

   The system displays the Change Datasource dialog box.



EDITING A DATASET – CHANGING THE DATA SOURCE

5. Select a data source from the list.



EDITING A DATASET – CHANGING THE DATA SOURCE

6. Click **Apply**.

Scenario 2:

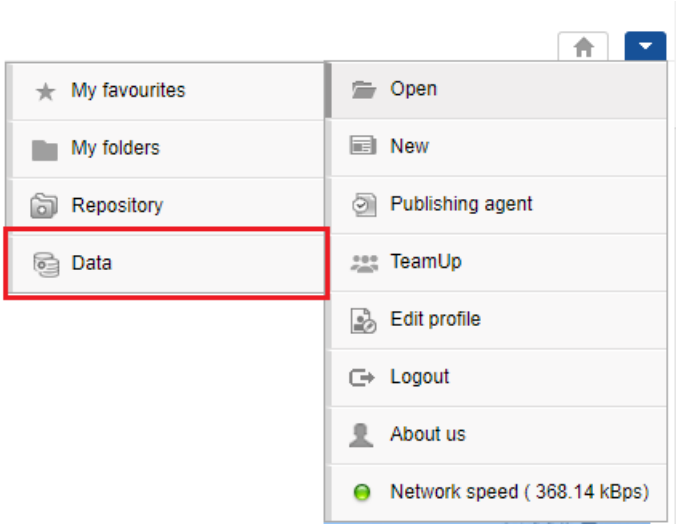
You can change data source associated to a dataset. For example, monthlysales dataset is currently associated with June_Sales data source and you want to change the data source to July_Sales.

About this task

Use this task to change the data source of a dataset.

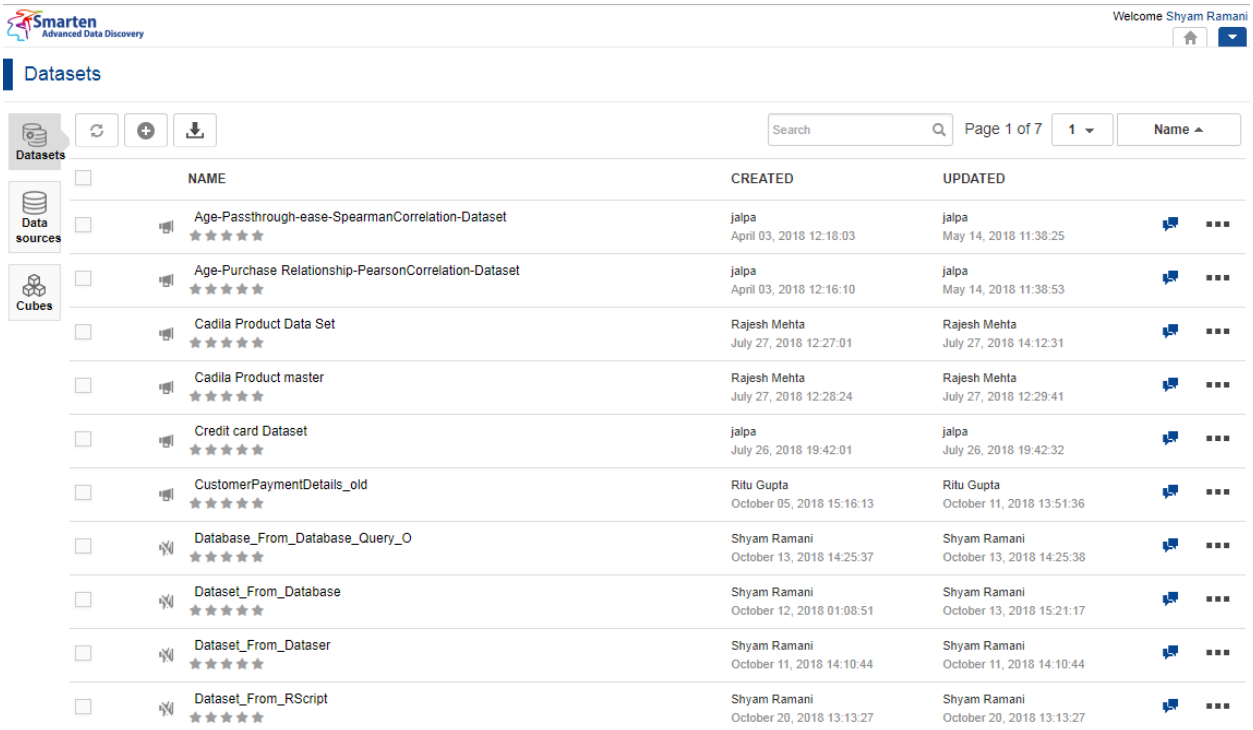
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

The system displays the following page.

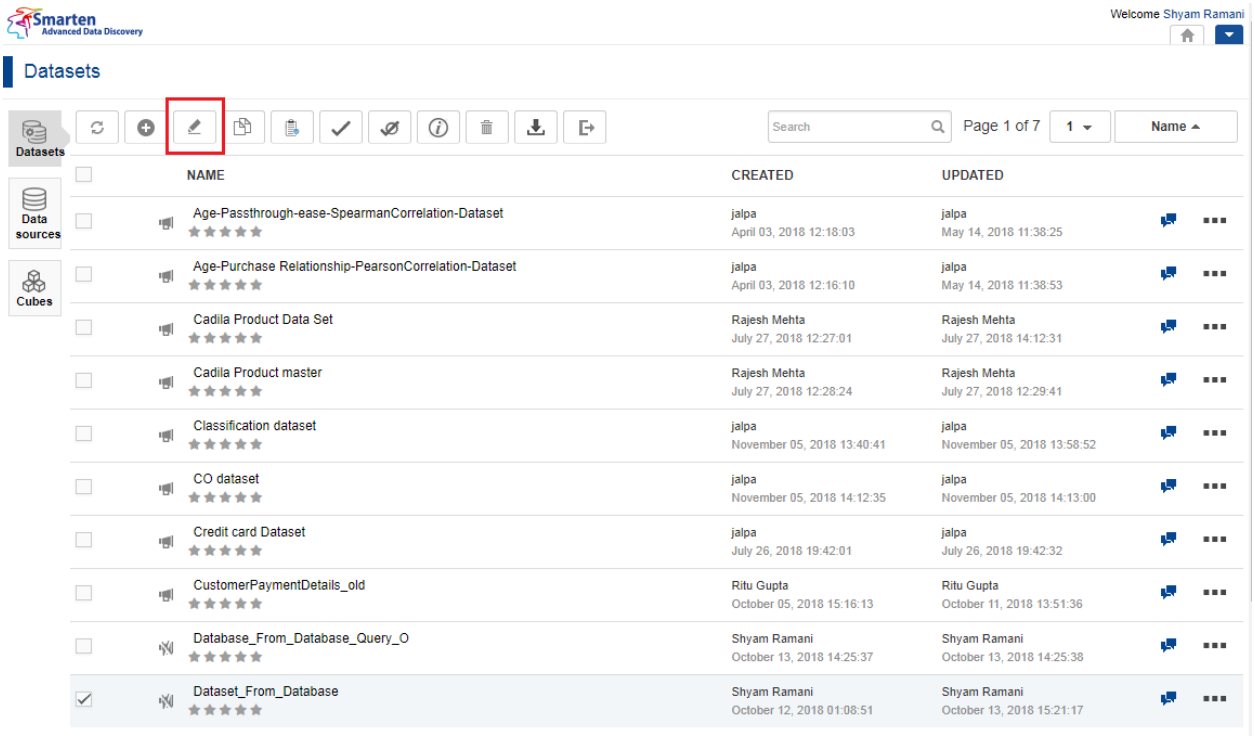


OPEN A DATASET – SELECTING A DATASET

2. Select the checkbox adjacent to the dataset for which you want to change the data source.

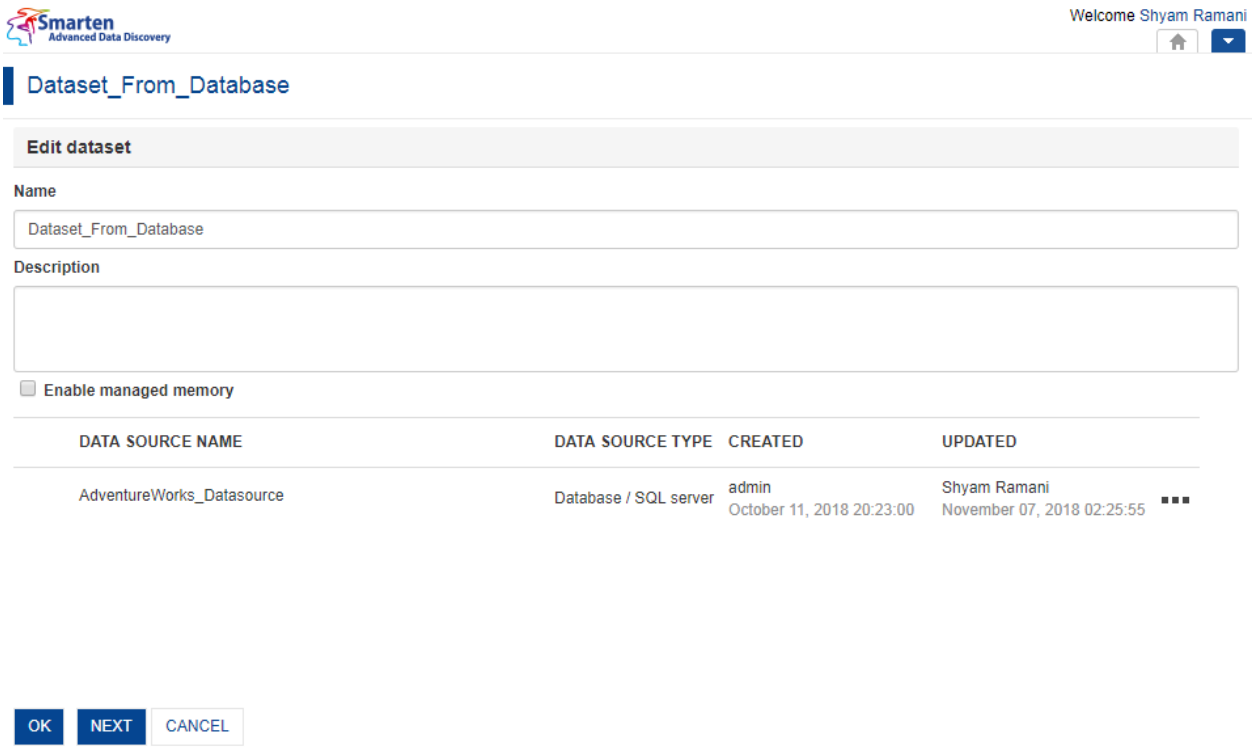
3. Click the edit icon.

The system displays the **Edit dataset** page.



MANAGING DATASET – EDITING A DATASET

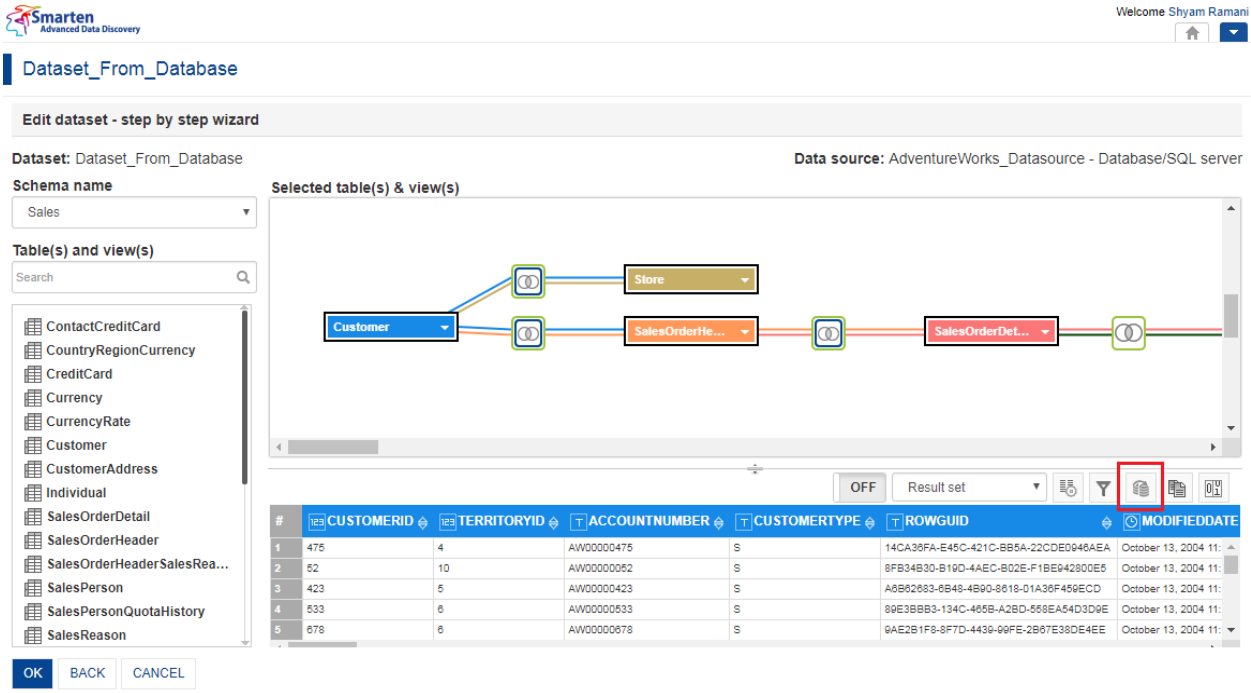
4. You can change the name, description, and managed memory option for the dataset, and then click **NEXT**.



EDITING A DATASET – CHANGING THE DATA SOURCE

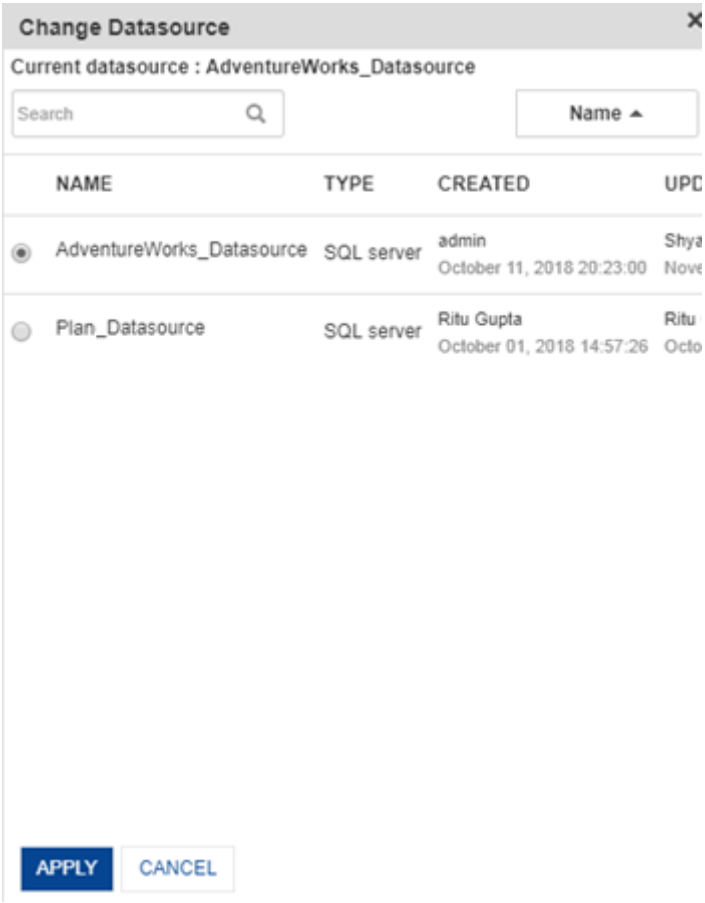
5. Click the Change Datasource button.

The system displays the **Change Datasource** dialog box.



EDITING A DATASET – THE OPTION TO CHANGE DATA SOURCE

6. Select a data source from the list of available data sources.



EDITING A DATASET – THE CHANGE DATASOURCE DIALOG BOX

> Note:
>
> The options available to change datasource are of the same type that of the current data source.For example, the current data source type is a database, hence, the system displays only database data sources. The system does not display a data source that has a different data source type than the current data source.
>
> The new data source schema must be the same as the current data source. If any column in the current dataset is not available in the new data source, the system does not allow you to change the current data source with the new data source.
>
> If the new data source has additional or more columns in the schema, the system allows you to change the data source.

7. Click **Apply**.
8. Click **OK**.

7.1.3 Managing Access Rights for a Dataset

Dataset access permission is about granting or restricting access to Datasets. The permissions are provided to view, edit, or delete a Dataset. For example, team members who are power users require performing all actions on a Dataset and hence should be given all permissions, whereas some team members may just need to view or use the data, and so they should be given the permission to only view the Dataset. Access permissions can be given as per Roles or to individual users of Smarten by the user who created the Dataset or by the Administrator. The access rights provided by the last user whether creator or Administrator are applicable. The following table describes the rights that can be assigned to a user:

Right	Description
View	This access right grants permission to a user to view and access a dataset.
Write	This access right grants permission to a user to edit a dataset.
Delete	This access right grants permission to a user to delete a dataset.
Export	This access right grants permission to a user to export a dataset.

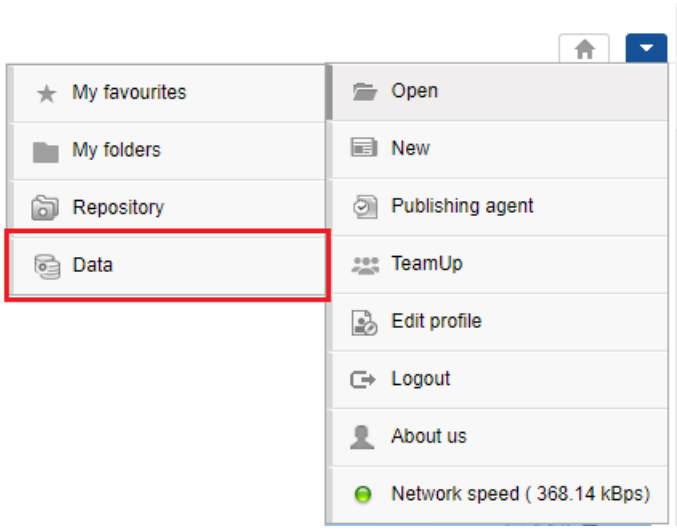
Reference: **Concept Manual > Dataset Management > Managing Access Rights for a Dataset**

About this task

Use this task to manage access rights for a dataset.

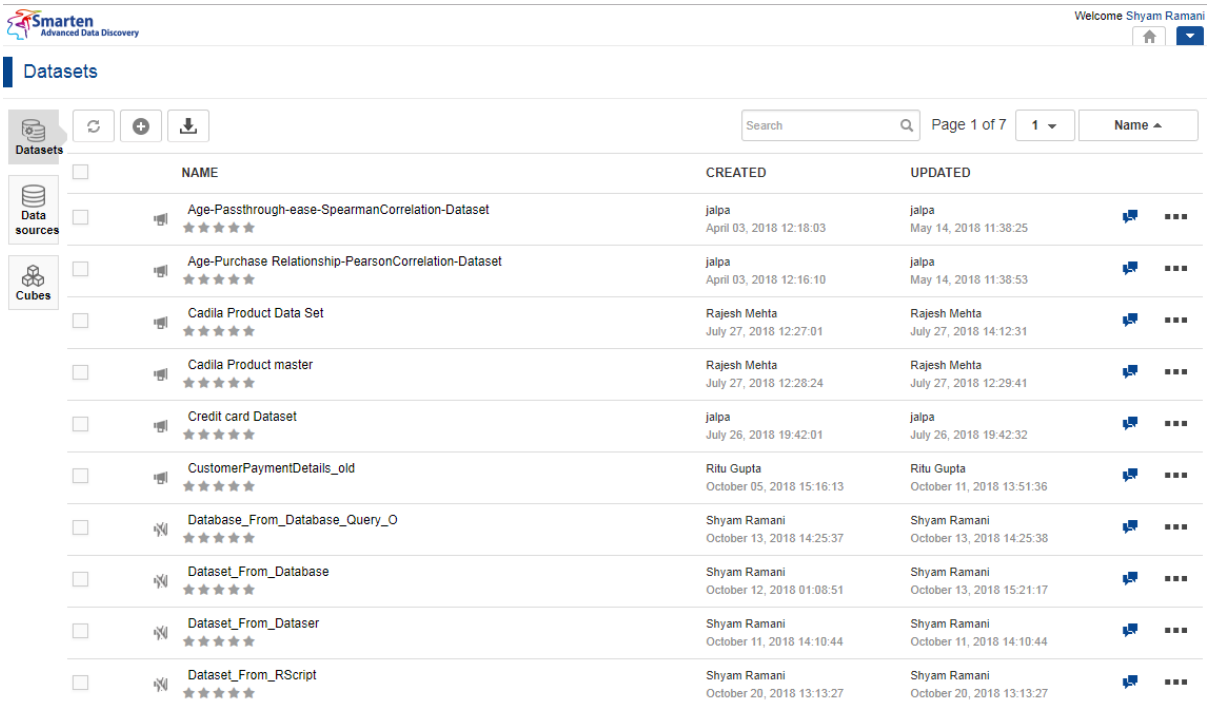
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

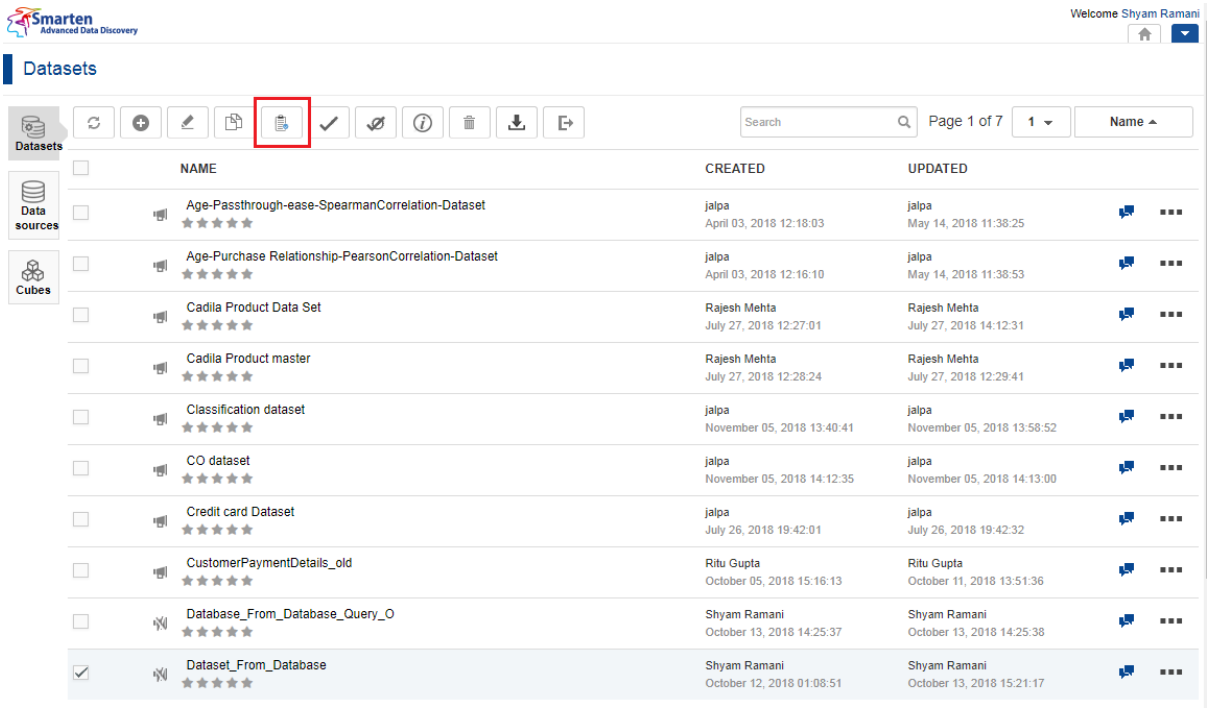
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

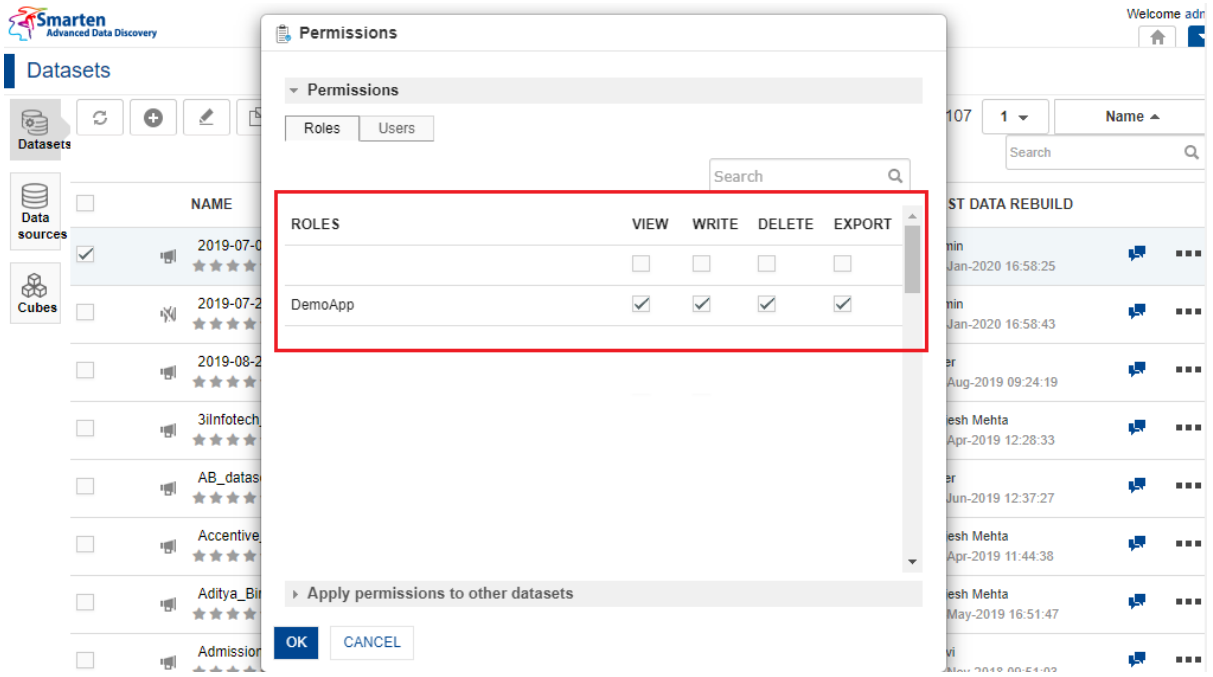
2. Select the checkbox adjacent to the dataset for which you want to manage access rights.
3. Click the permissions icon.

   The system displays the **Permissions** dialog box.



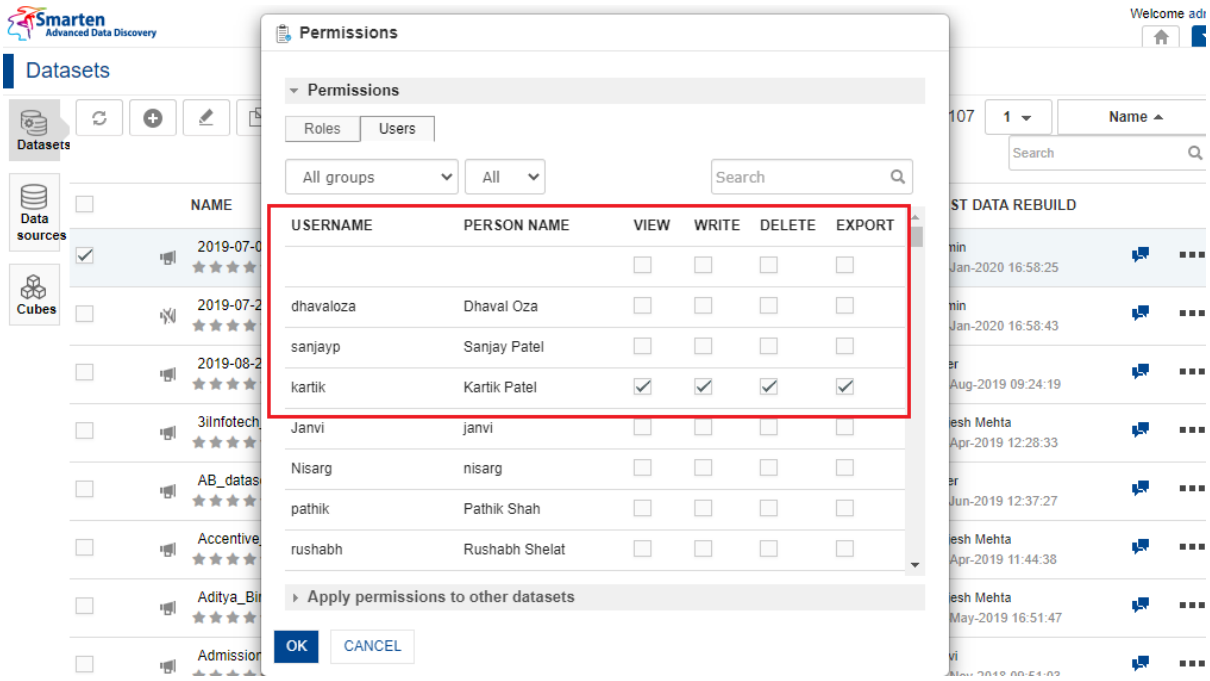
MANAGING ACCESS RIGHTS – CLICKING PERMISSIONS ICON

4. Click the **Roles** tab, to assign access rights to various roles.



ASSIGN PERMISSIONS – ACCESS PERMISSIONS FOR ROLES

5. Select the box under the **VIEW**, **WRITE**, **DELETE**, and **EXPORT** permissions columns to assign that access to a role. For example, in the above image, you can select the **VIEW**, **WRITE**, and **DELETE** boxes for DemoApp role, to allow users with DemoApp role to view, modify, or delete the dataset.
6. Click the **Users** tab, to assign access rights to different users or group of users.

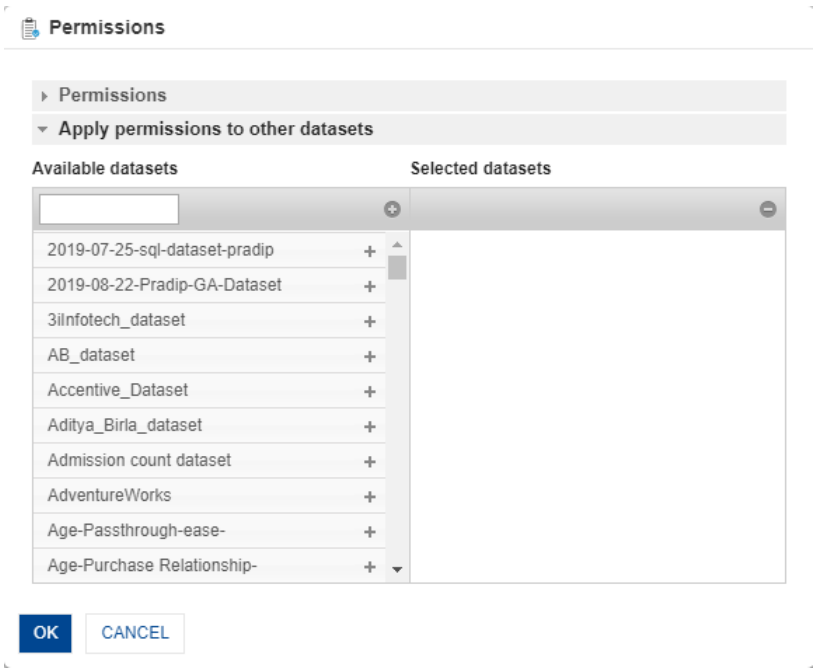


ASSIGN PERMISSIONS – ACCESS PERMISSIONS FOR USERS

7. Select the box under the **VIEW**, **WRITE**, **DELETE**, and **EXPORT** permissions columns to assign that access to a role. For example, in the above image, you can select the **VIEW**, **WRITE**, and **DELETE** boxes for the use 'kartik', to allow 'kartik' to view, modify, or delete the dataset.

8. Click **Apply permissions to other datasets**option to grant the same permissions to other datasets which you have selectedfor roles and users in the previous step.

   This option allows you to grant the same set of permissions you have granted to a role for other datasets instead of granting the same set of permissions to the role for each datasets separately. For example, if you have grated view and export permissions to Role 1 and want to grant the same permissions for Dataset1, Dataset2, and Dataset3. You can use the **Apply permissions to other datasets** option to grant the view and export permissions to Role 1 for Dataset1, Dataset2, and Dataset3.



DATASET PERMISSION: APPLY PERMISSIONS TO OTHER DATASETS

9. Click the plus sign adjacent to the datasets for which you want to grant the permissions you have granted to the roles in the earlier step.
10. Click **OK**togrant the permissions you have selected for roles and users.

7.1.4 Deleting a Dataset

This feature enables authorized users to delete a Dataset. A deleted Dataset is no longer available in the system.

While deleting a Dataset, users may or may not delete Objects associated with that Dataset. Users can reuse these objects by associating them with another Dataset having the same columnar data structure as the deleted one.

> Note:
> If you delete a dataset that is associated with an object, the system displays an error message while accessing the object using deleted dataset.

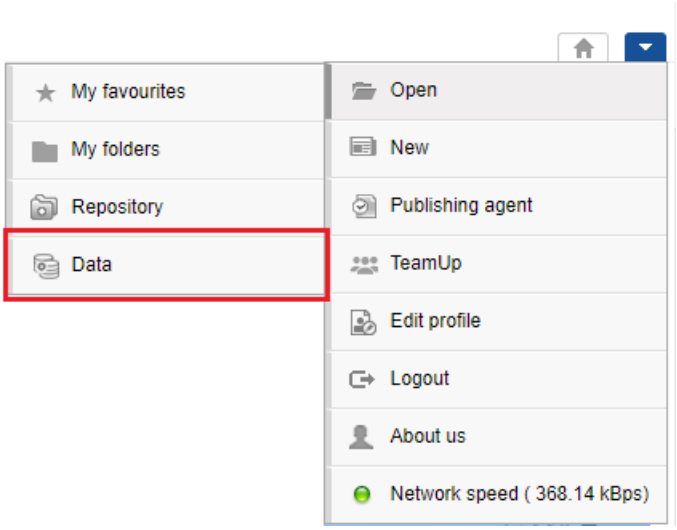
Reference: **Concept Manual > Dataset Management > Deleting a Dataset**

About this task

Use this task to delete a dataset.

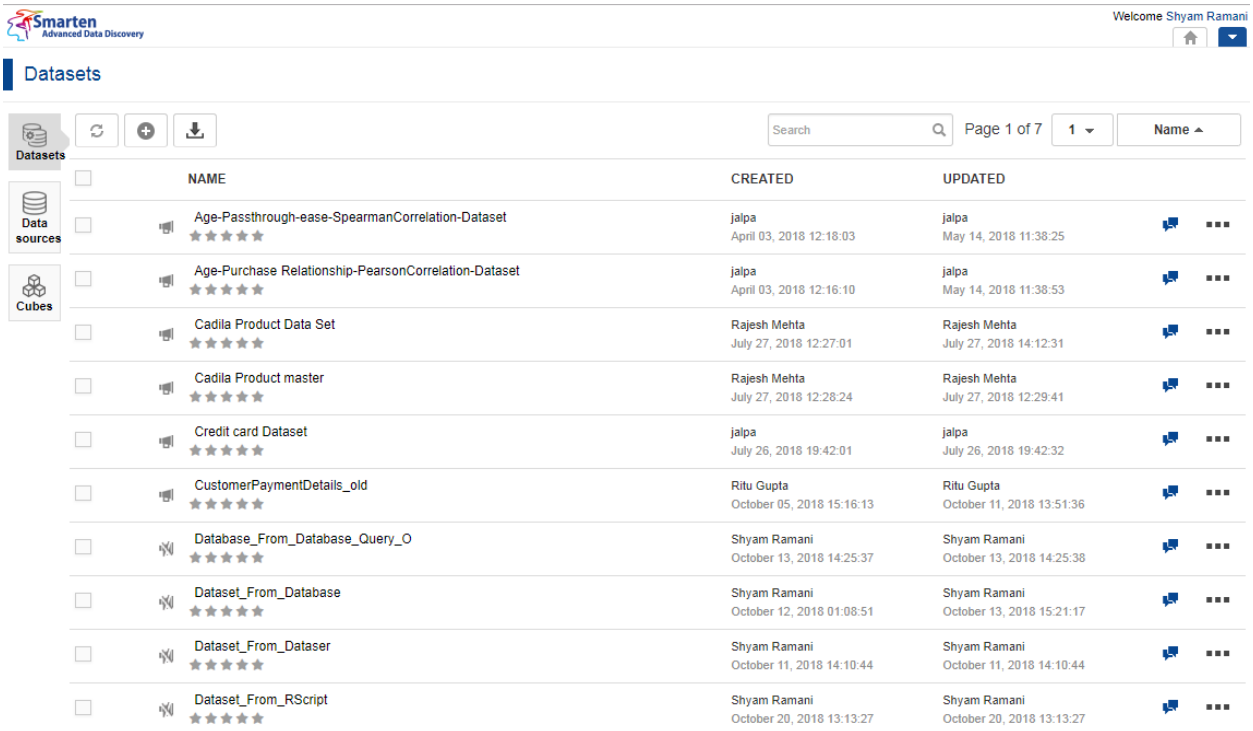
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

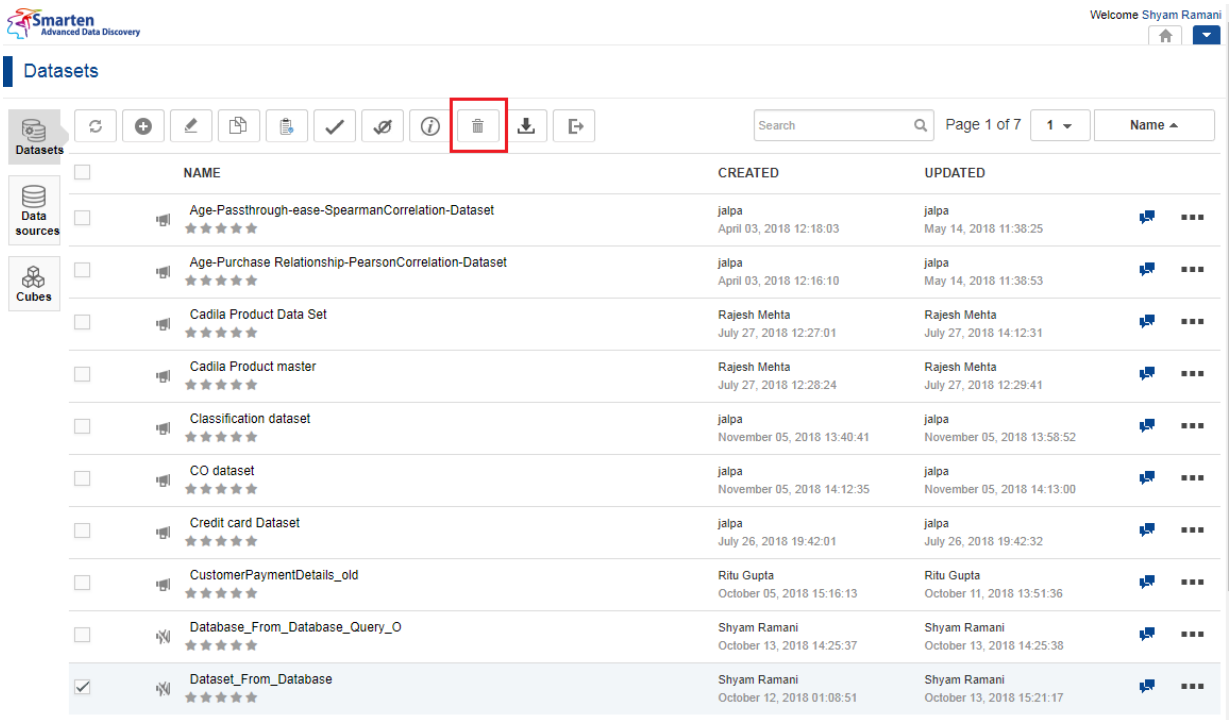
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

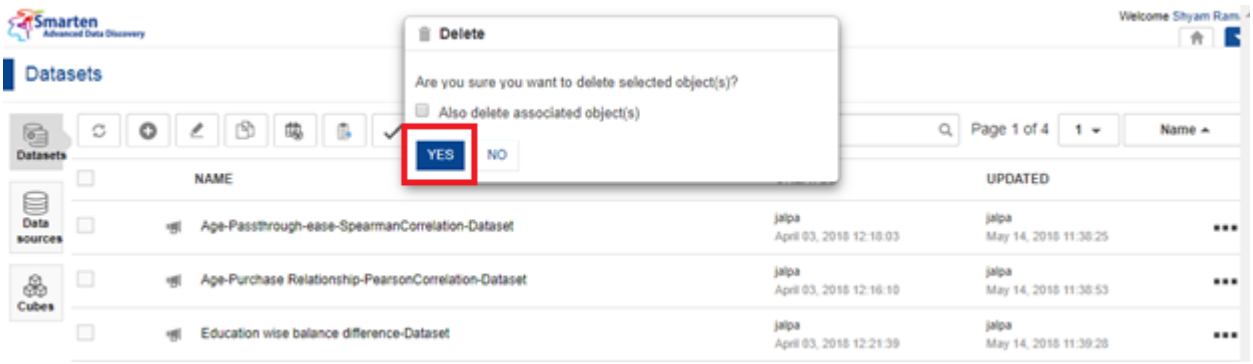
2. Select the checkbox adjacent to the dataset you want to delete.
3. Click the delete icon.

The system displays the **Delete** dialog box.



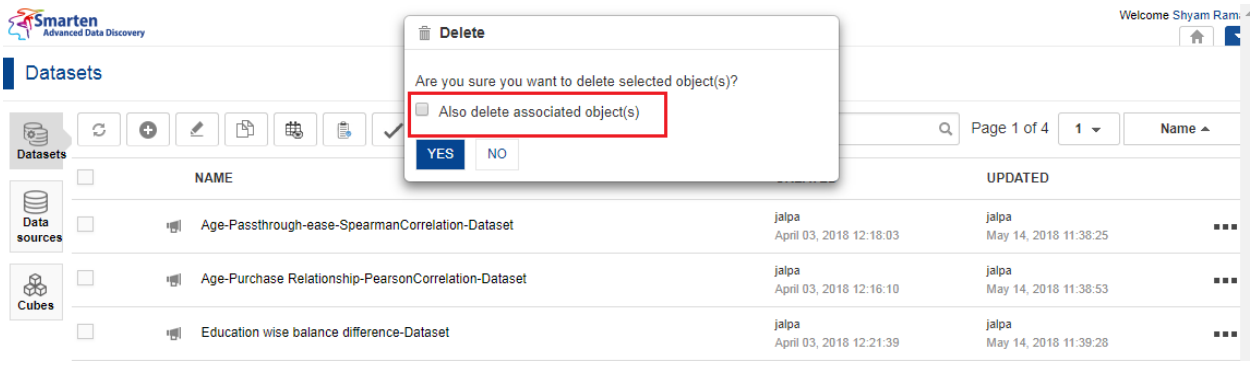
DELETING A DATASET – CLICKING THE DELETE ICON

4. Click **YES**, to delete the selected dataset.



DELETING A DATASET – PROVIDING CONFIRMATION TO DELETE THE DATASET

If there are objects associated with the dataset, the system displays the option asking if you want to delete the associated objects along with the dataset. You can select the option if you want to delete the objects or associate the objects with another dataset.



DELETING A DATASET – DELETING ASSOCIATED OBJECTS

7.1.5 Copying a Dataset

This feature enables authorized users to replicate a Dataset along with its structure, data source, and data from the dataset.

It allows users to create a copy of a Dataset without going through the process of creating a Dataset from scratch.

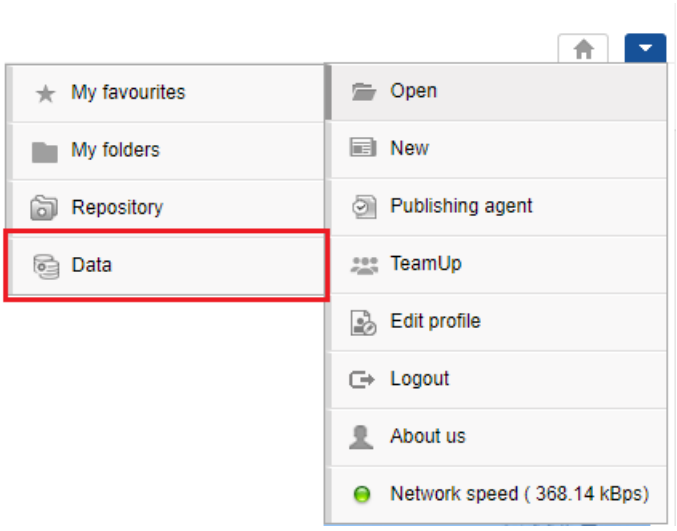
Reference: **Concept Manual > Dataset Management > Copying a Dataset**

About this task

Use this task to copy a dataset.

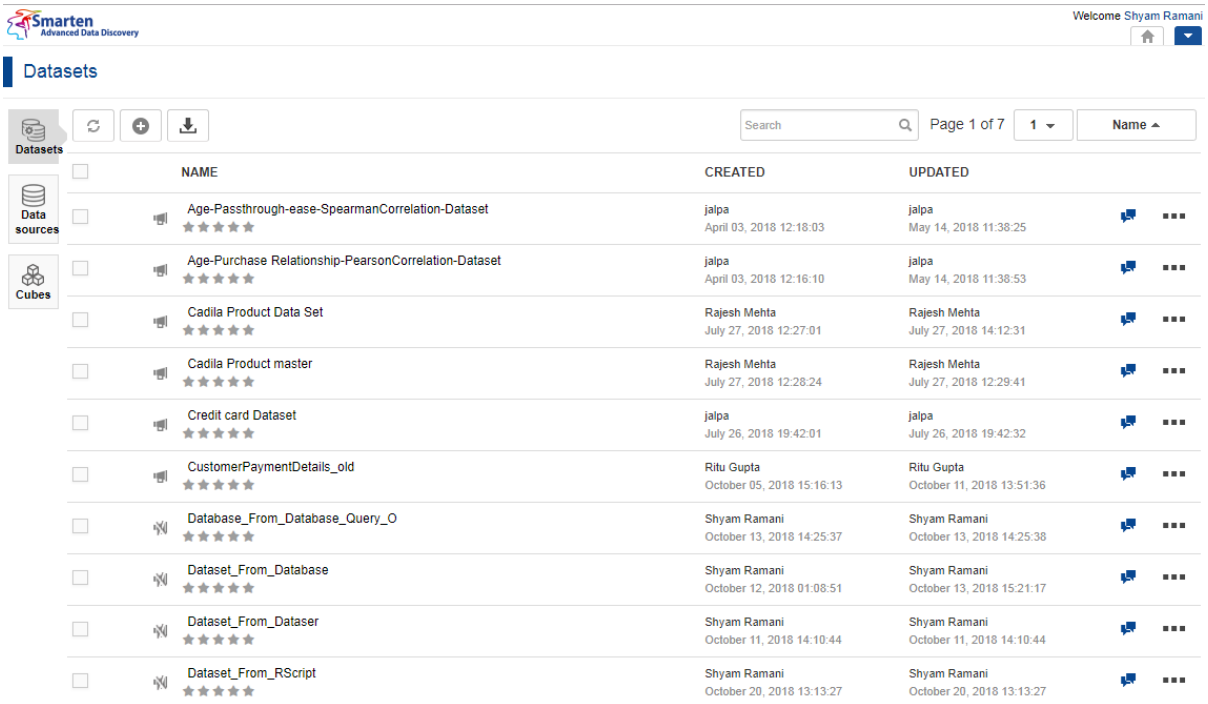
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

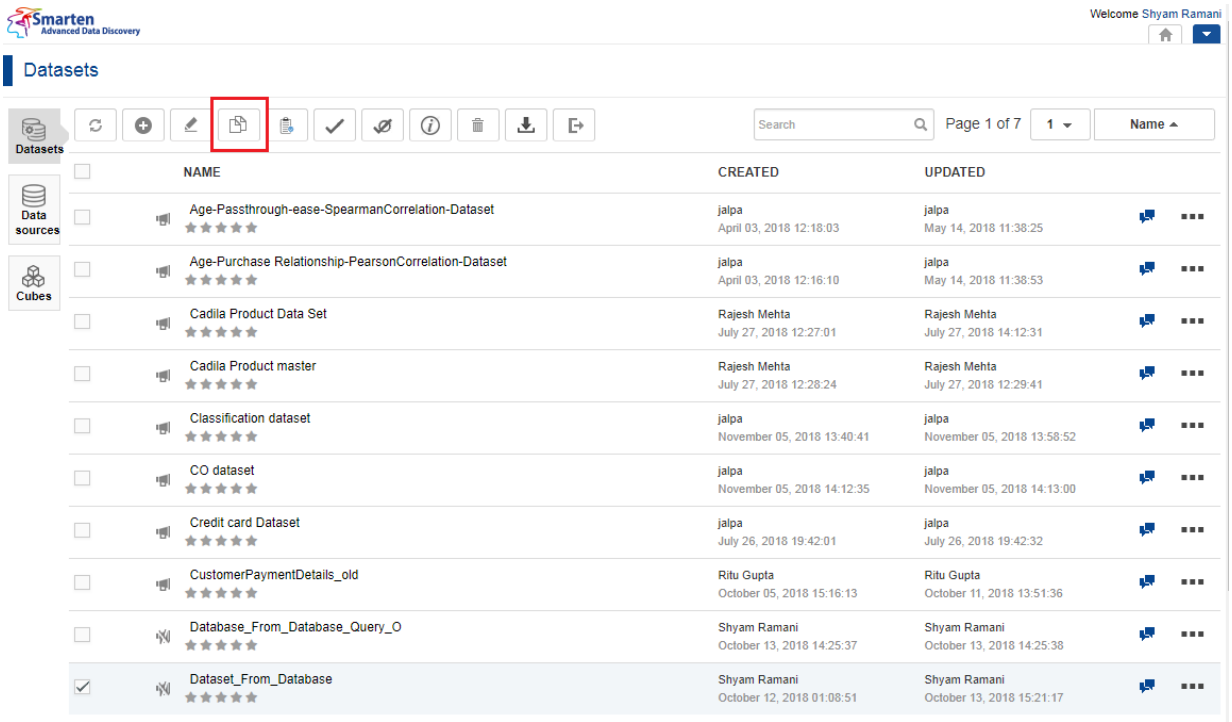
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

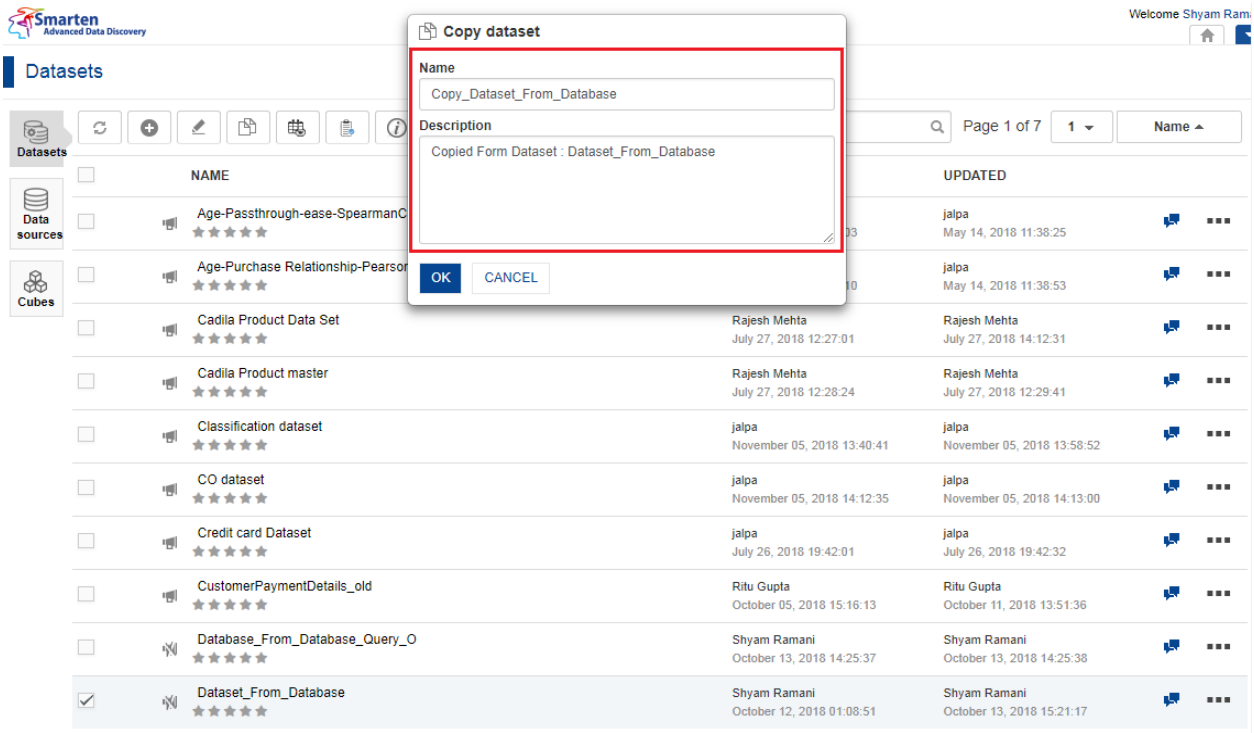
2. Select the checkbox adjacent to the dataset you want to copy.
3. Click the copy icon.

   The system displays the **Copy dataset** dialog box.



COPYING A DATASET – CLICKING THE COPY ICON

4. Specify a name and description for the new dataset to be created using the dataset you copied in the previous step.



COPYING A DATASET – PROVIDING A NEW NAME AND DESCRIPTION

5. Click **OK**.

   The system creates a new dataset using the dataset you have copied with a new name and the description you have provided. The system copies complete dataset metadata definition and data of the selected dataset. The system will not copy permissions from the original dataset.

7.1.6 Exporting a Dataset

You can export a dataset definition in the XML format. The exported XML file can be imported in same or another instance of Smarten. The system exports metadata, data source reference, and preparation actions (as per action editor) used for the dataset in the XML file. Dataset export does not include dataset permissions and data.

You must also import associated data source while importing a dataset. If the associated data source is not available in the system where the dataset is imported, the system will not be able to extract data for that dataset. In such cases, you must change the data source for the imported dataset.

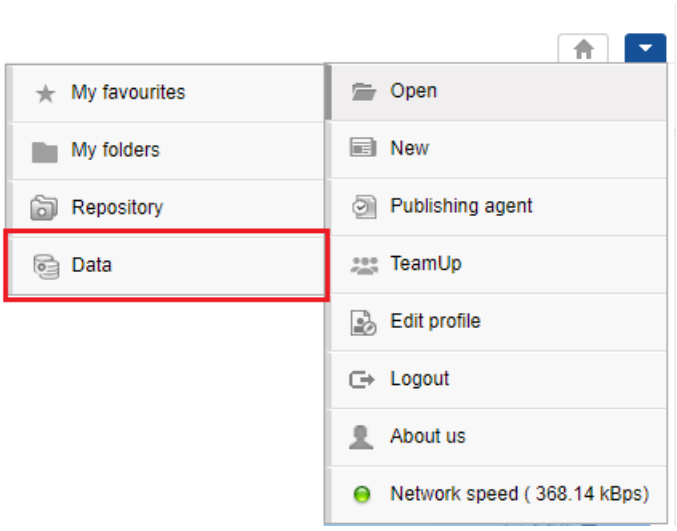
You can use dataset export feature to take backup of the dataset and you can restore a dataset by importing the XML file of that dataset.

About this task

Use this task to export a dataset.

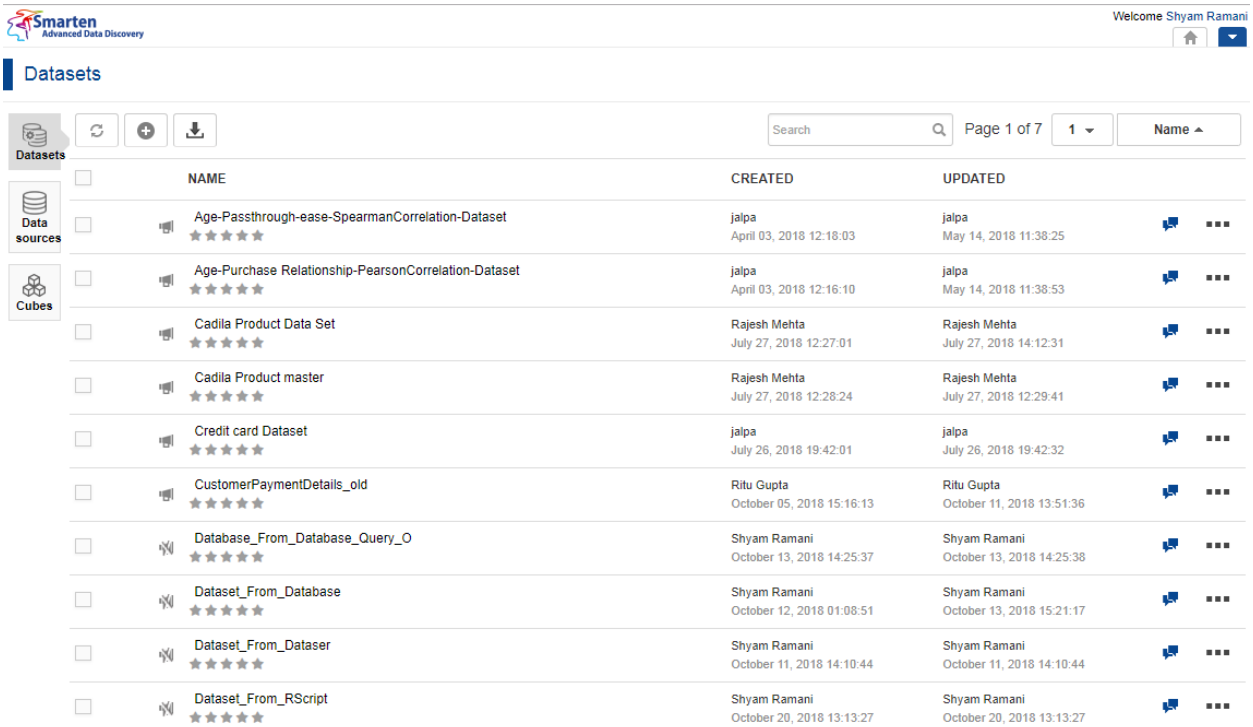
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

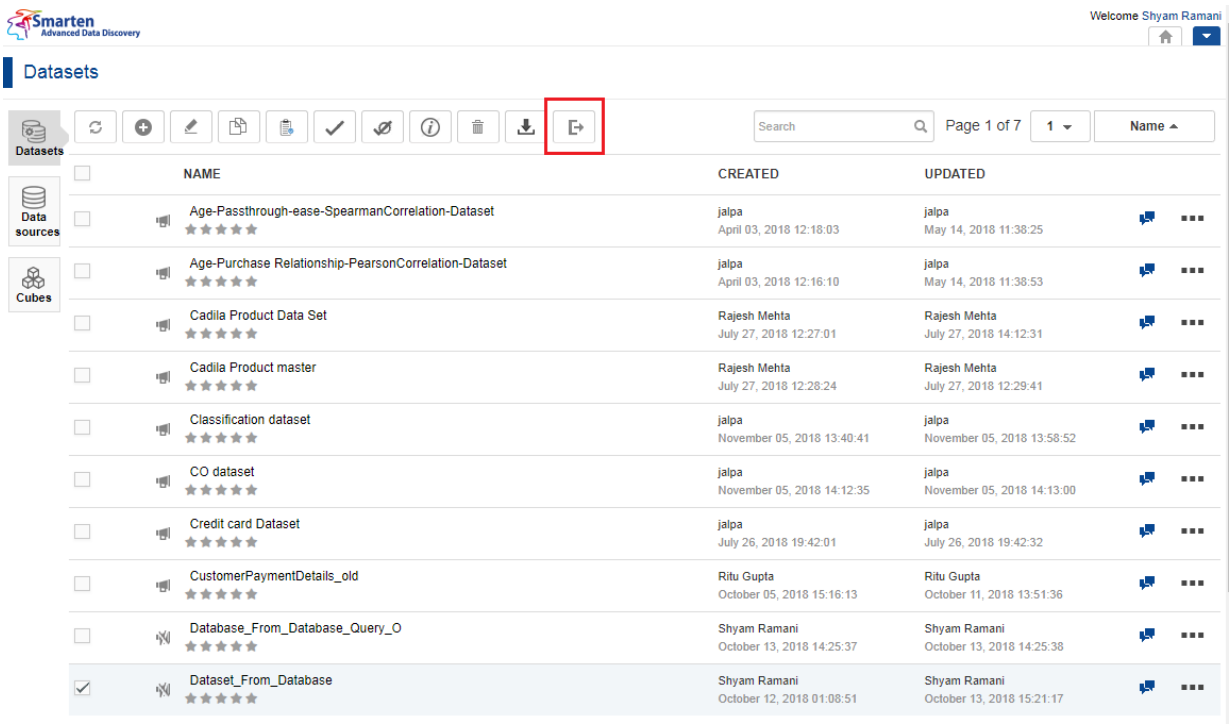
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

2. Select the checkbox adjacent to the dataset you want to export.
3. Click the export icon.

   The system downloads the dataset in .xml format.



EXPORTING A DATASET – CLICKING THE EXPORT ICON

7.1.7 Importing a Dataset

You can import an exported dataset in XMLformat from the same instance or another instance of Smarten.

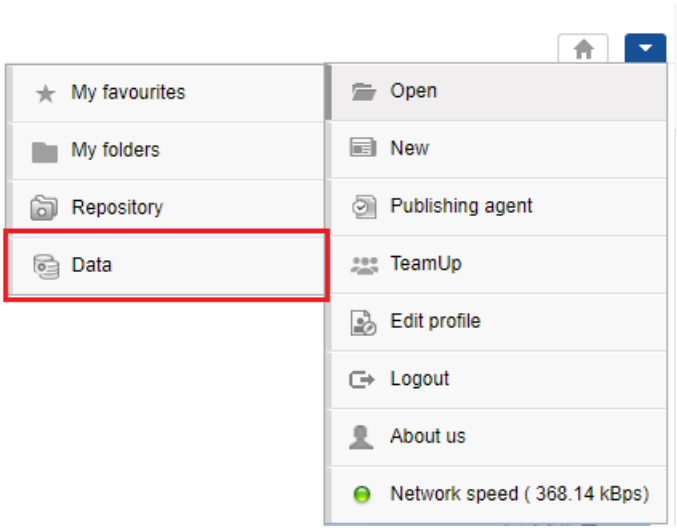
> Note:
>
> Dataset Export does not include data. You must also import the associated data source along with thedataset. If the associated data source is not available in the system, the system will not be able to extract data for that dataset. In such cases, you must change the data source for those datasets.
>
> After importing a dataset, you must open and publish the dataset to extract data from the associated data source.

About this task

Use this task to import a dataset.

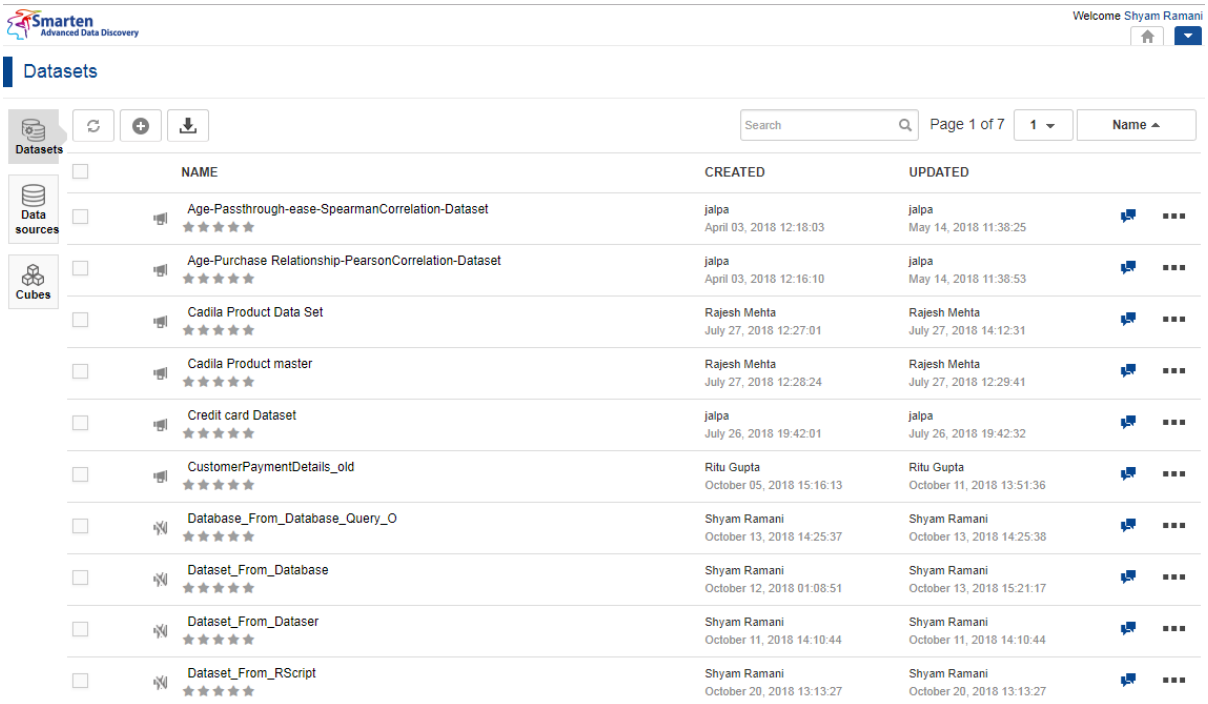
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

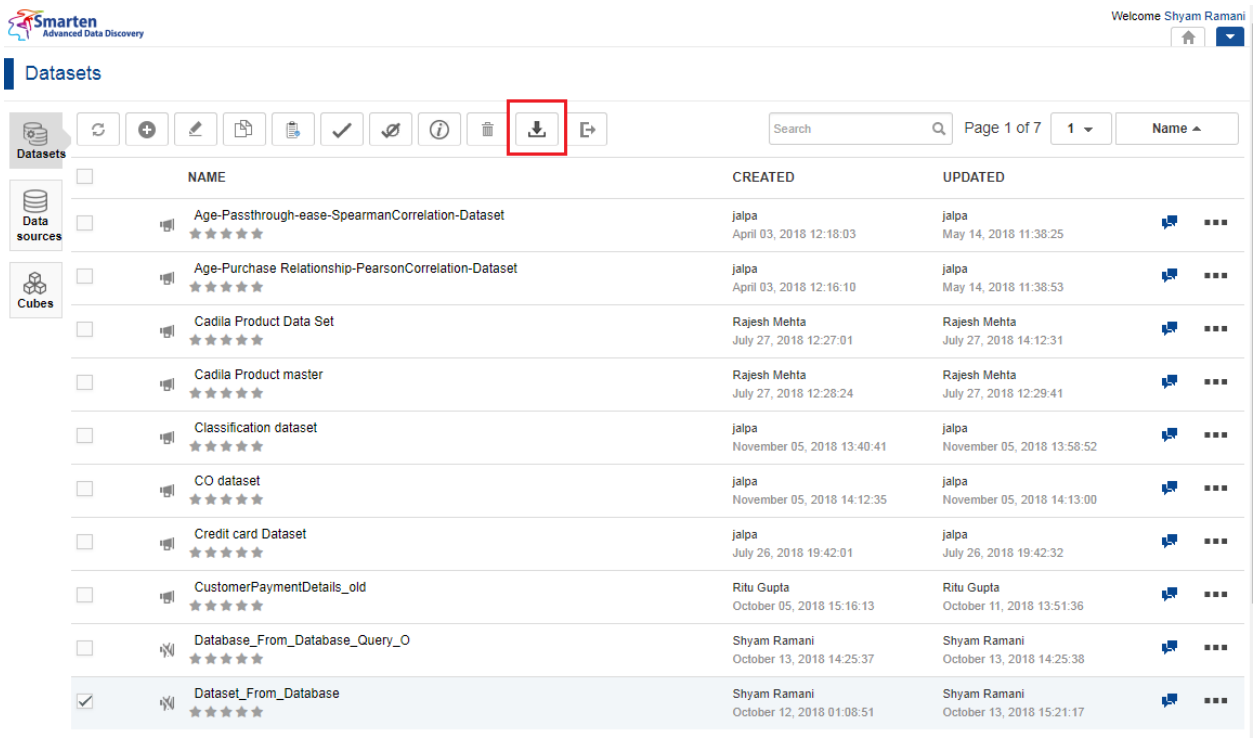
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

2. Click the import icon.

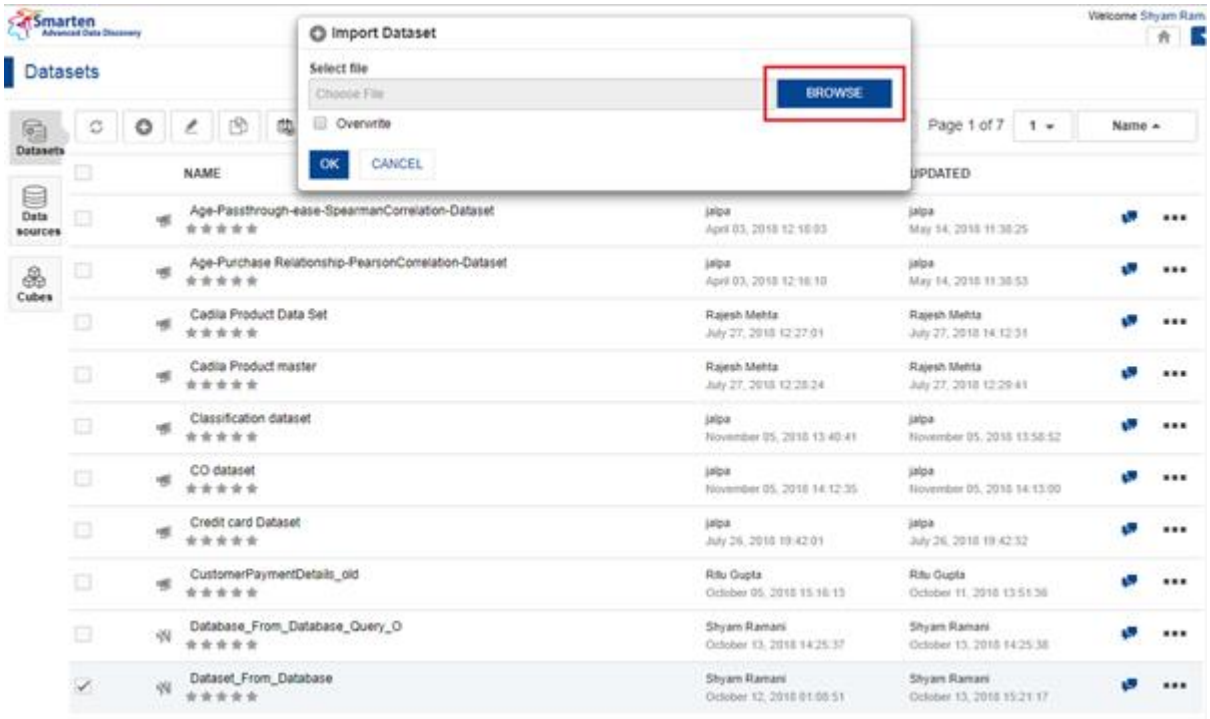
   The system displays the **Import Dataset** dialog box.



IMPORTING A DATASET – CLICKING THE IMPORT ICON

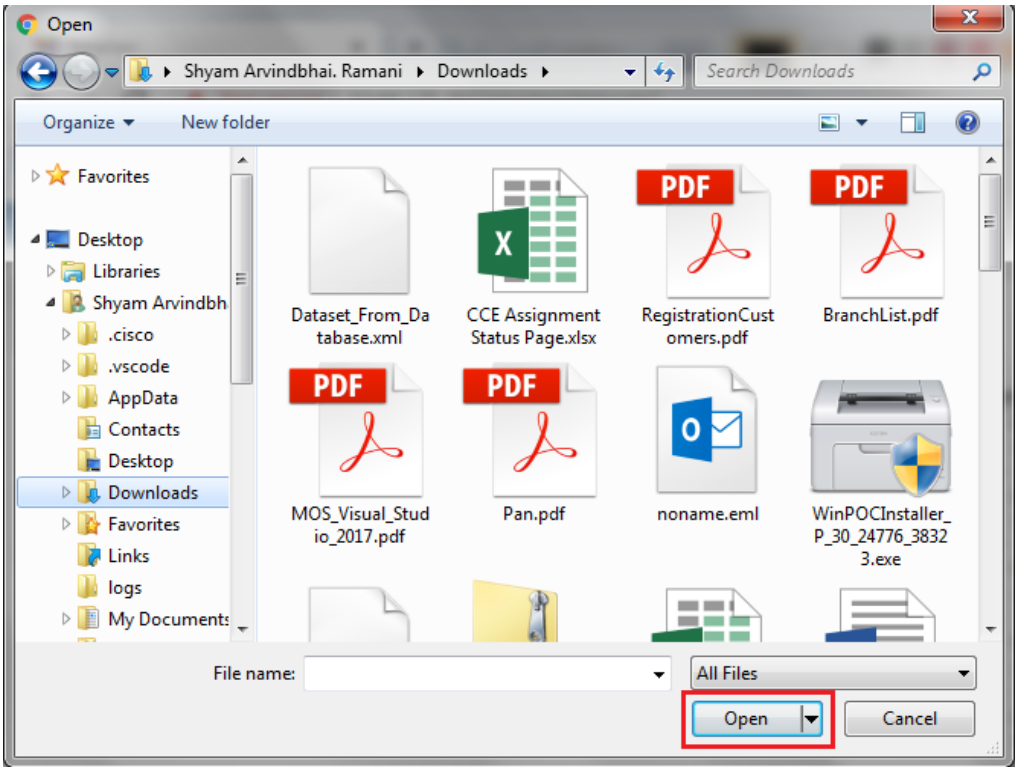
3. Click the **BROWSE** button.

   The system displays the **Open** dialog box.



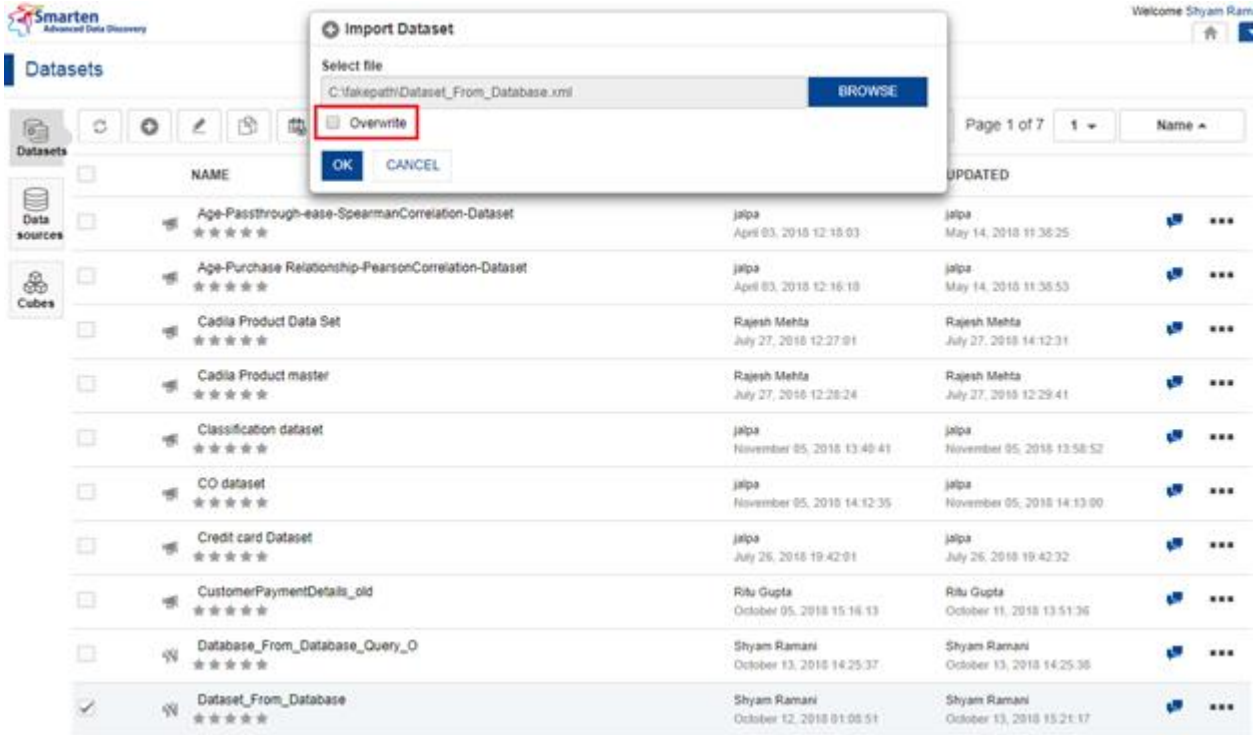
IMPORTING A DATASET – OPENING THE OPEN DIALOG BOX

4. Select the XML file for the dataset that has been exported, and then click **Open**.



IMPORTING A DATASET – THE OPEN DIALOG BOX

5. If a dataset with the same name as of the imported dataset exists in the system, you can select the **Overwrite** checkbox, to allow the system to overwrite dataset in the system with the imported dataset.



IMPORTING A DATASET – SELECTING THE OVERWRITE OPTION

If the **Overwrite** option is not selected, the system retains the existing dataset and adds the imported dataset with a new name. The new name of the dataset is appended with a numeric value in increasing order. For example, if you import ABC dataset and a dataset with the same name already exists, the system will rename the imported dataset as ABC_1.

Similarly, if dataset with name ABC_1 exists, the system will rename the imported dataset as ABC_2.

6. Click **OK**.

7.1.8 Viewing Dataset Information

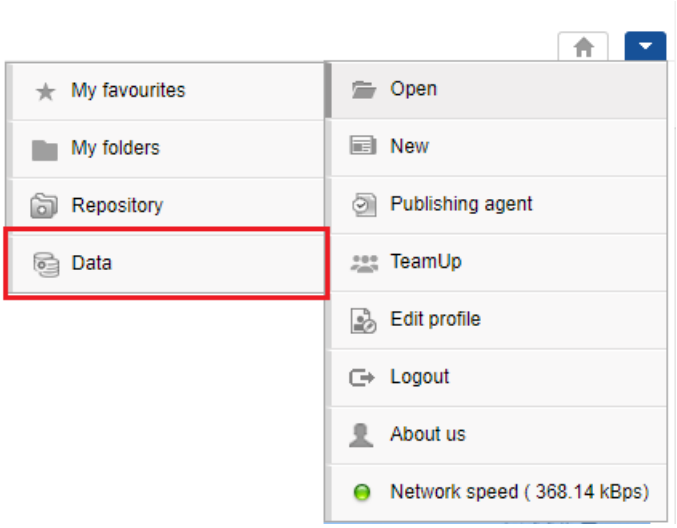
You can view information about a dataset such as a name, last updated, data source, dataset size, and much more. Based on the

About this task

Use this task to view information about a dataset.

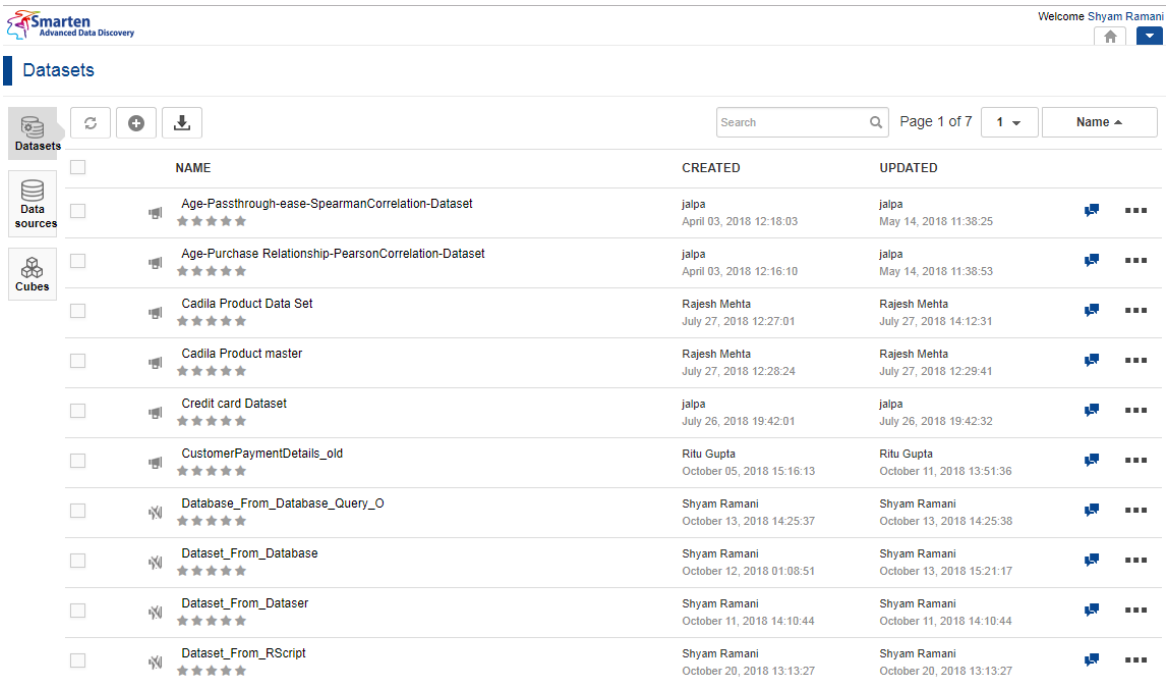
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

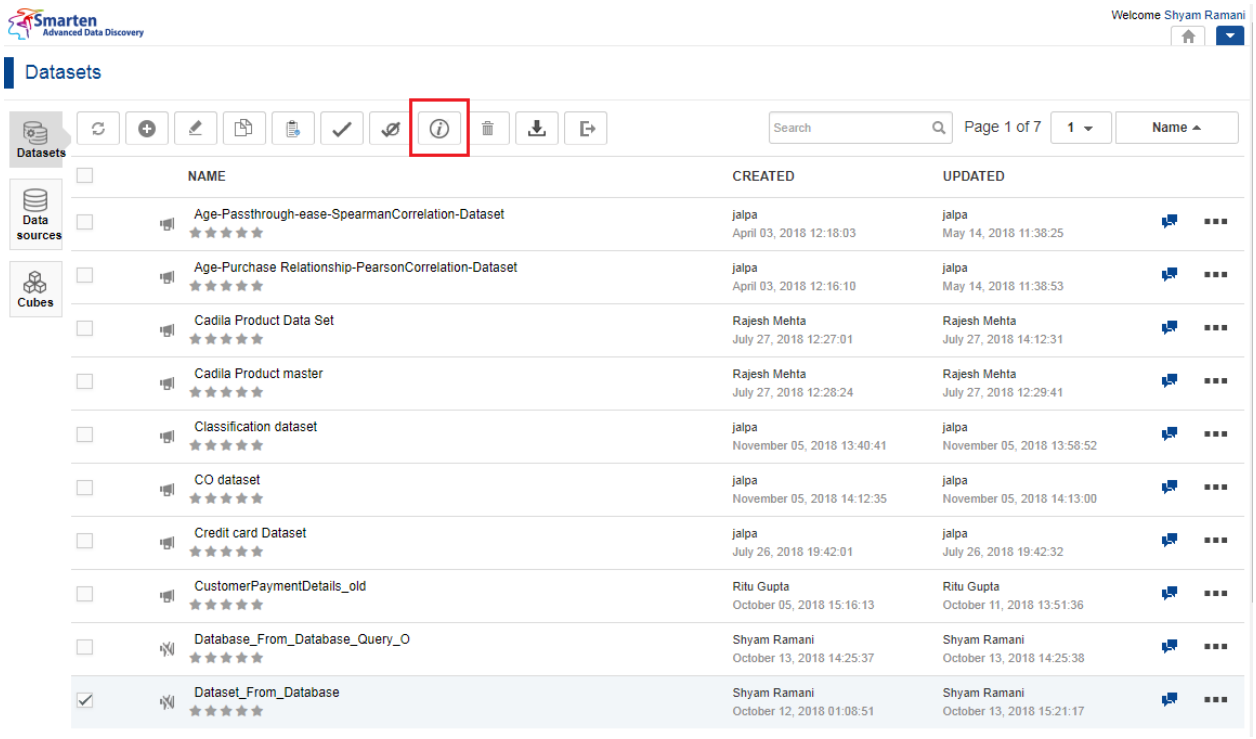
The system displays the following page.



OPEN A DATASET – SELECTING A DATASET

2. Select the checkbox adjacent to the dataset whose information you want to view.
3. Click the information icon.

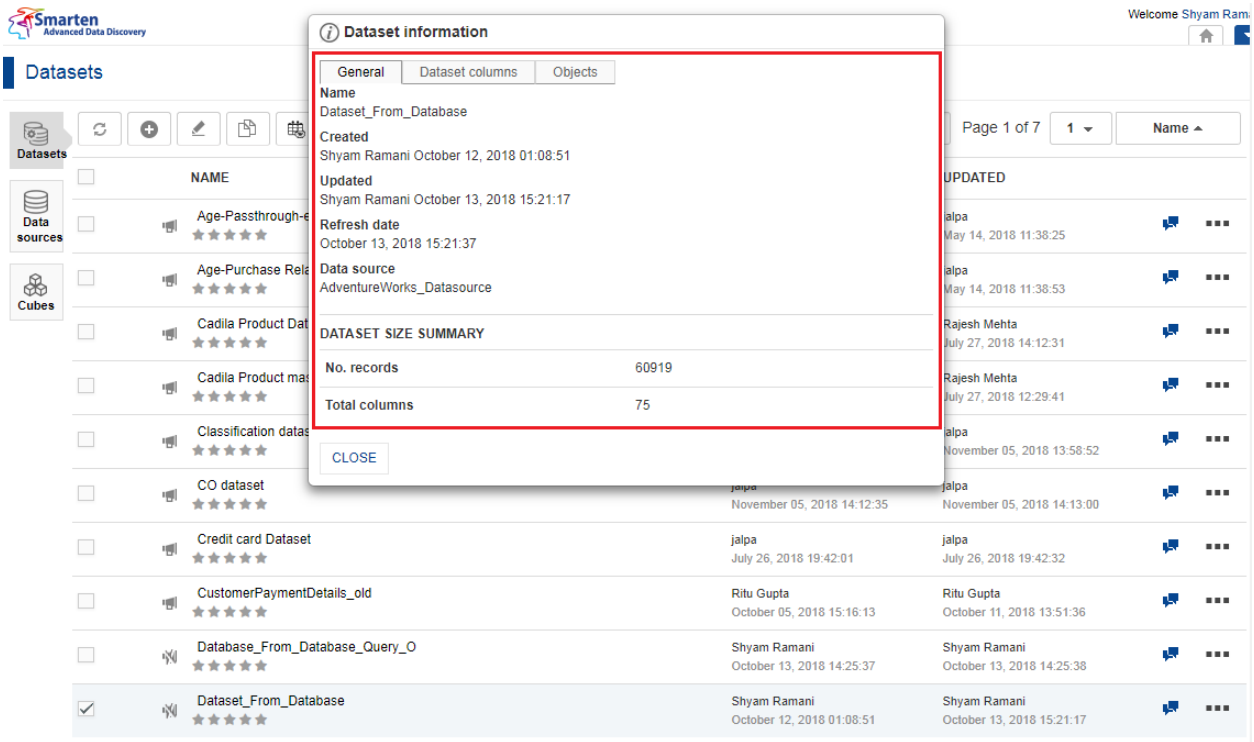
The system displays the **Datasets information** dialog box.



VIEW DATASET INFORMATION – CLICKING THE INFORMATION ICON

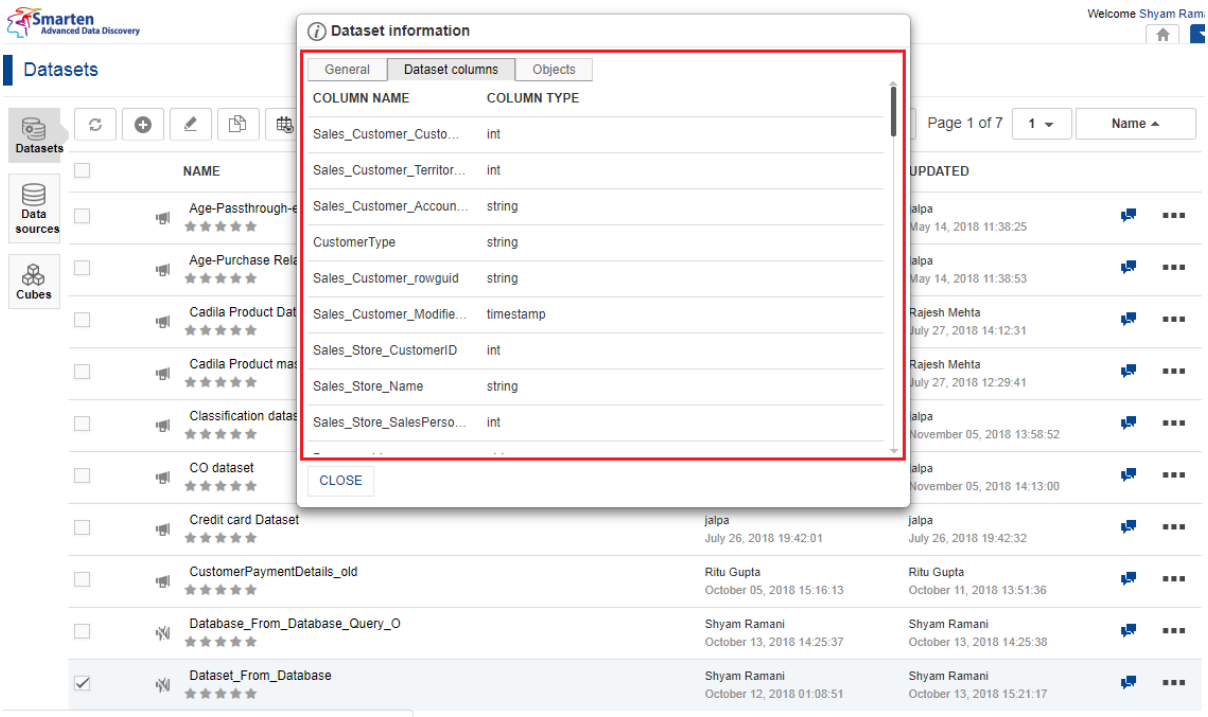
4. Click the **General** tab, to view general information about the dataset. The following information is displayed:

- **Name**: Name of the dataset.
- **Created**: Date and time when the dataset was created.
- **Updated**: Date and time when the dataset was last updated.
- **Refreshdate**: Date and time when the dataset was last refreshed.
- **Datasource**: Name of the data source used in the dataset.
- **No.records**: Number of records available in the dataset.
- **Totalcolumns**: Number of columns available in the dataset.



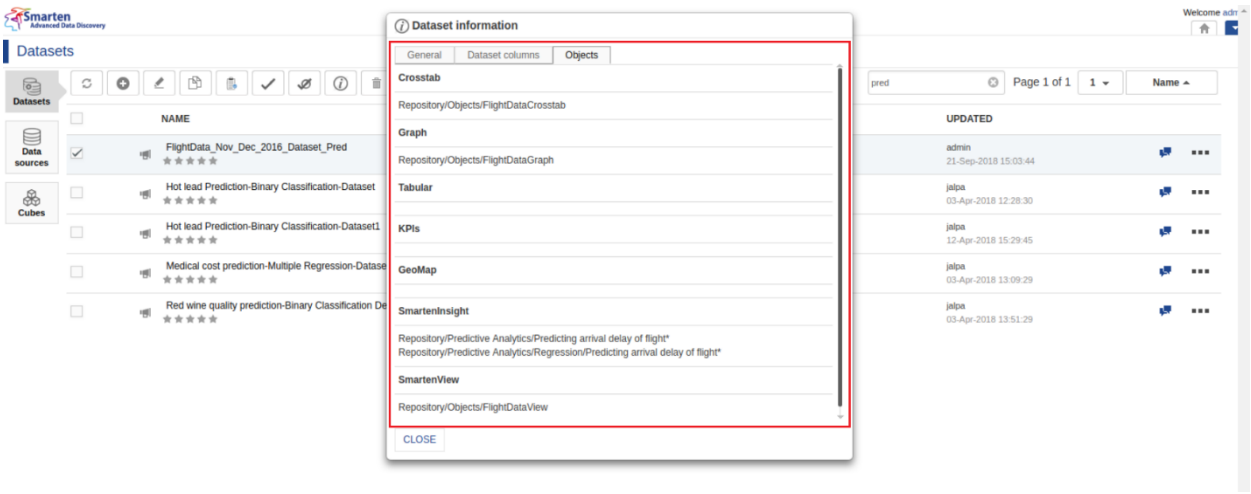
VIEWING DATASET INFORMATION – VIEWING GENERAL INFORMATION

5. Click the **Dataset columns** tab, to view information regarding the dataset columns of the selected dataset. The tab displays the name of the columns and their data type.



VIEWING DATASET INFORMATION – VIEWING THE DATASET COLUMN INFORMATION

6. Click the **Objects** tab, to view information regarding the objects using the selected dataset.



VIEWING DATASET INFORMATION – VIEWING THE OBJECTS RELATED INFORMATION

7. Click **CLOSE**.

7.1.9 Marking a Dataset as IT Approved

This feature enables IT staff to approve a Dataset. Marking a Dataset as IT approved certifies it for data quality and helps users in identifying quality Datasets.

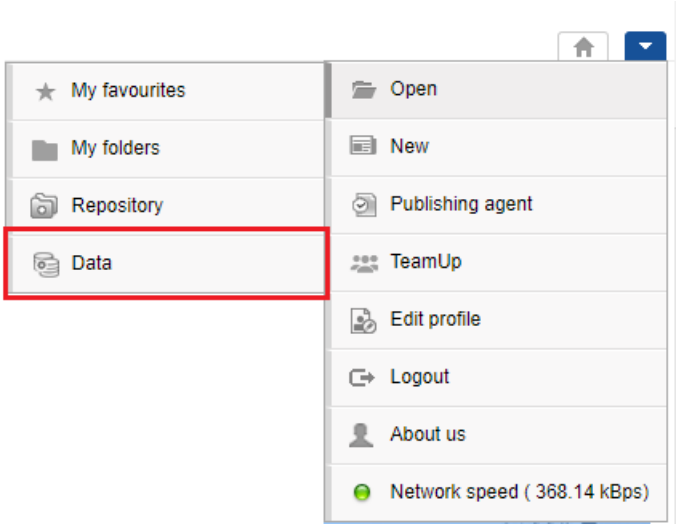
Reference: **Concept Manual > Dataset Management >IT Certification for Dataset**

About this task

Use this task to mark a dataset as IT approved.

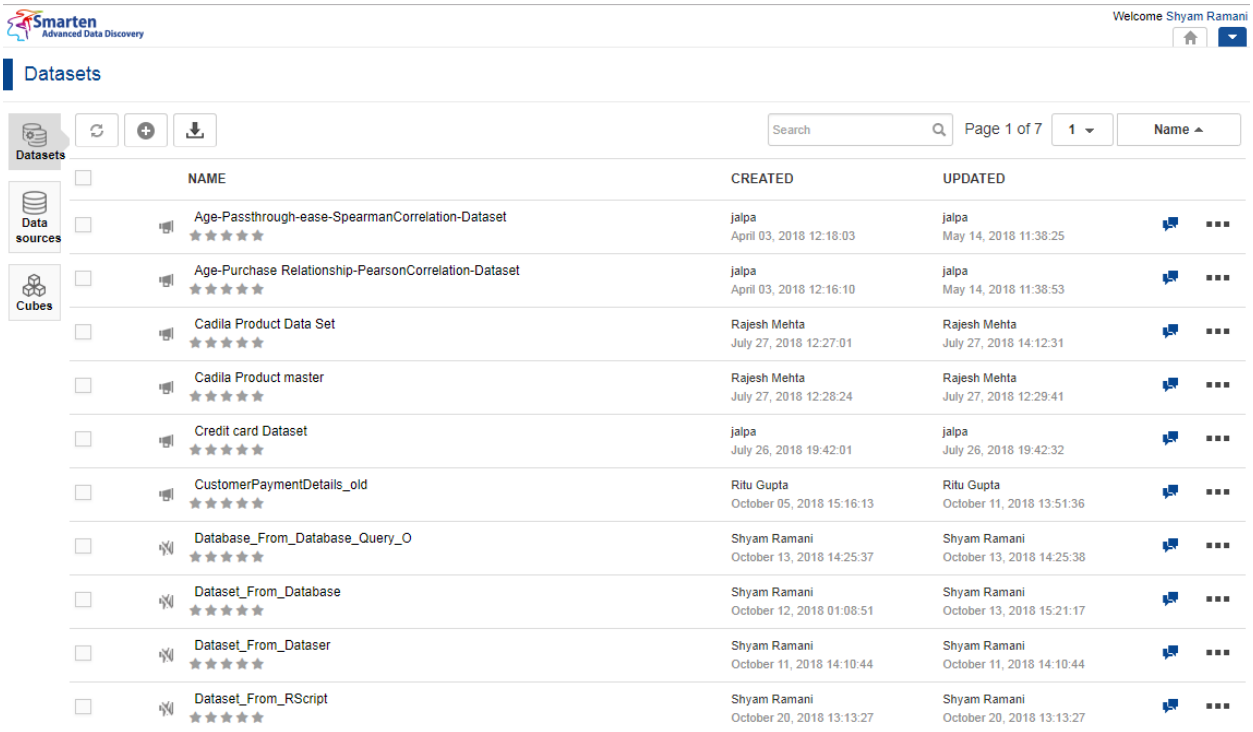
Procedure

1. Click **Open** ->**Data** from the menu.



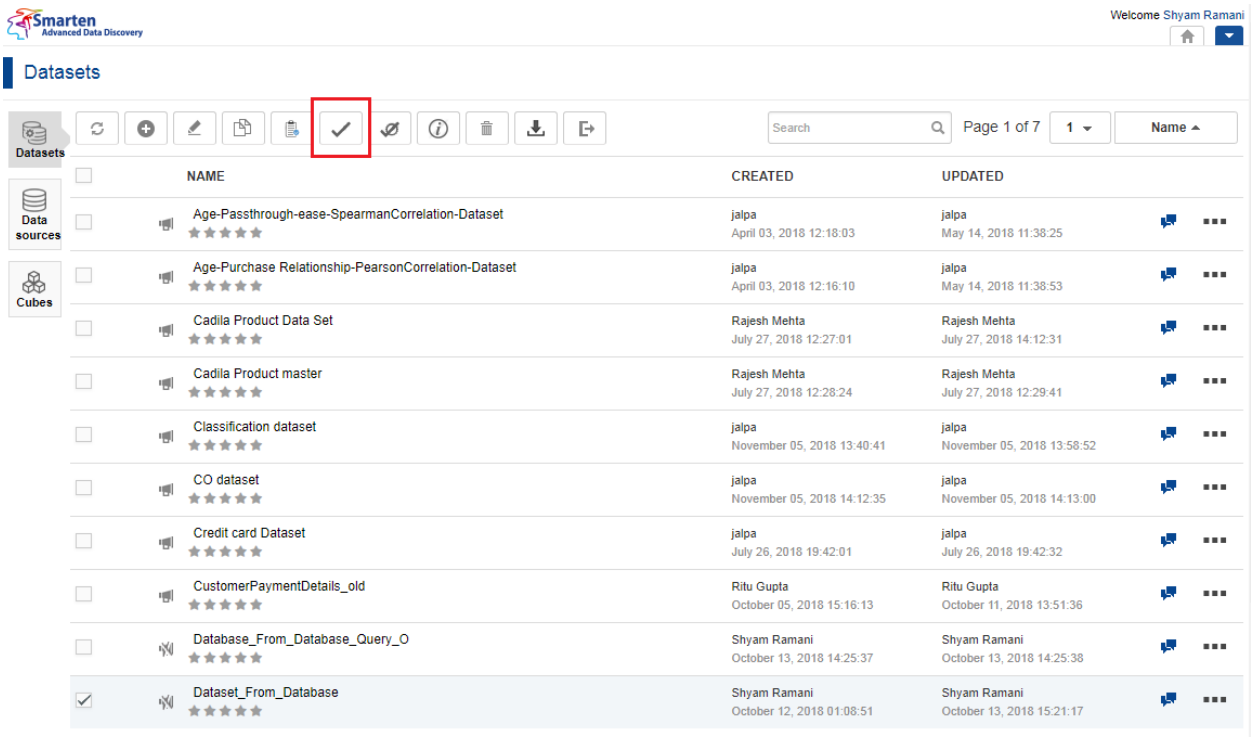
MENU OPTION – OPEN DATA

The system displays the following page.



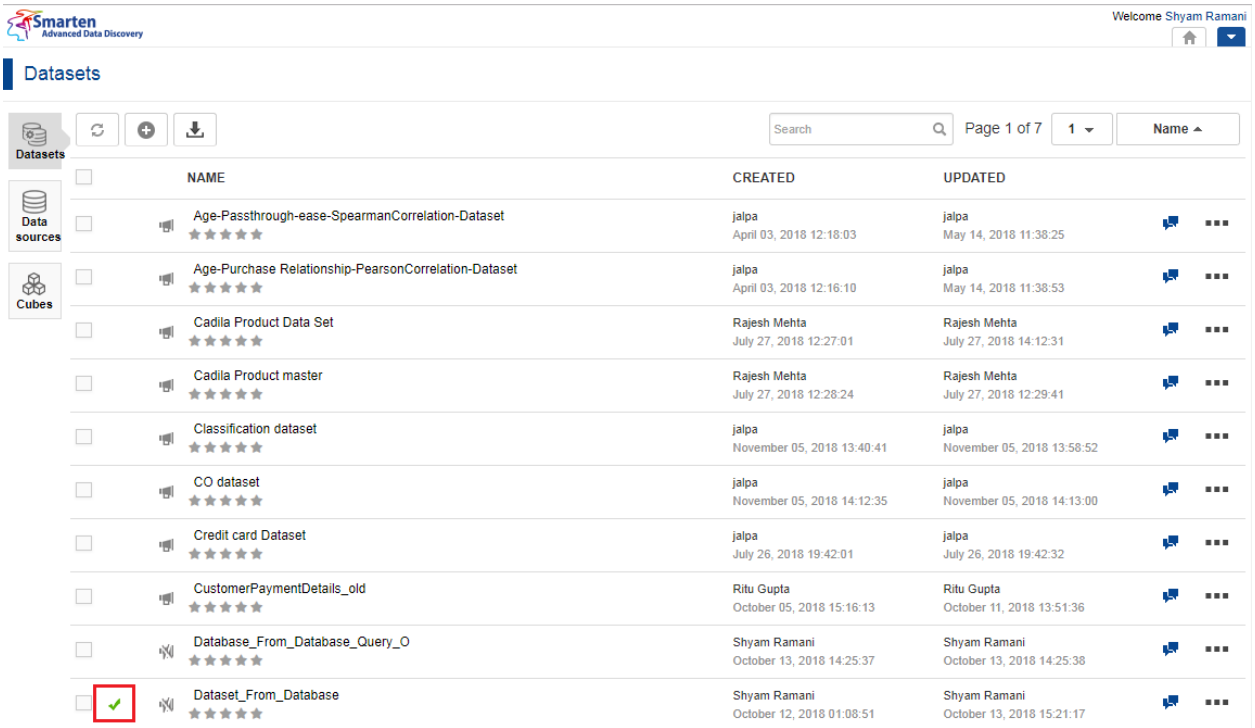
OPEN A DATASET – SELECTING A DATASET

2. Select the checkbox adjacent to the dataset you want to mark as IT approved.
3. Click the Mark IT Approved icon.



MARKING A DATASET – CLICKING THE MARK IT APPROVED ICON

The system marks the dataset as IT approved and displays a check symbol adjacent to the dataset.



MARKING A DATASET – THE CHECK MARK INDICATING THAT THE DATASET IS IT APPROVED

7.1.10 Unmarking a Dataset as IT Approved

This feature enables IT staff to unmark the datasets that are marked as approved.

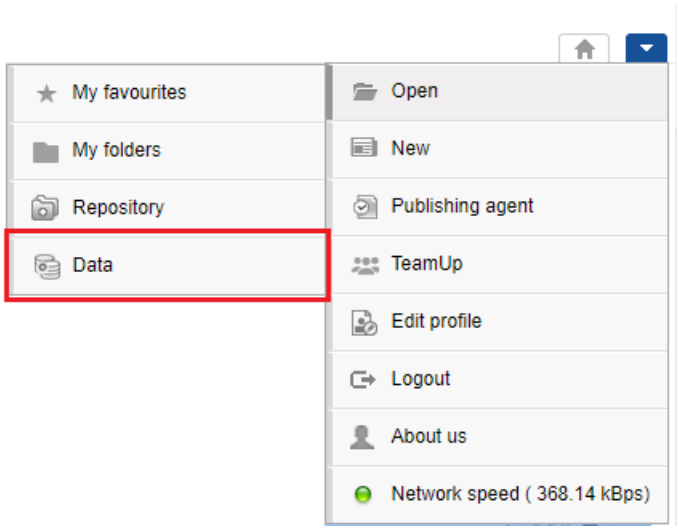
Reference: **Concept Manual > Dataset Management > IT Certification for Dataset**

About this task

Use this task to unmark a dataset that is marked as IT approved.

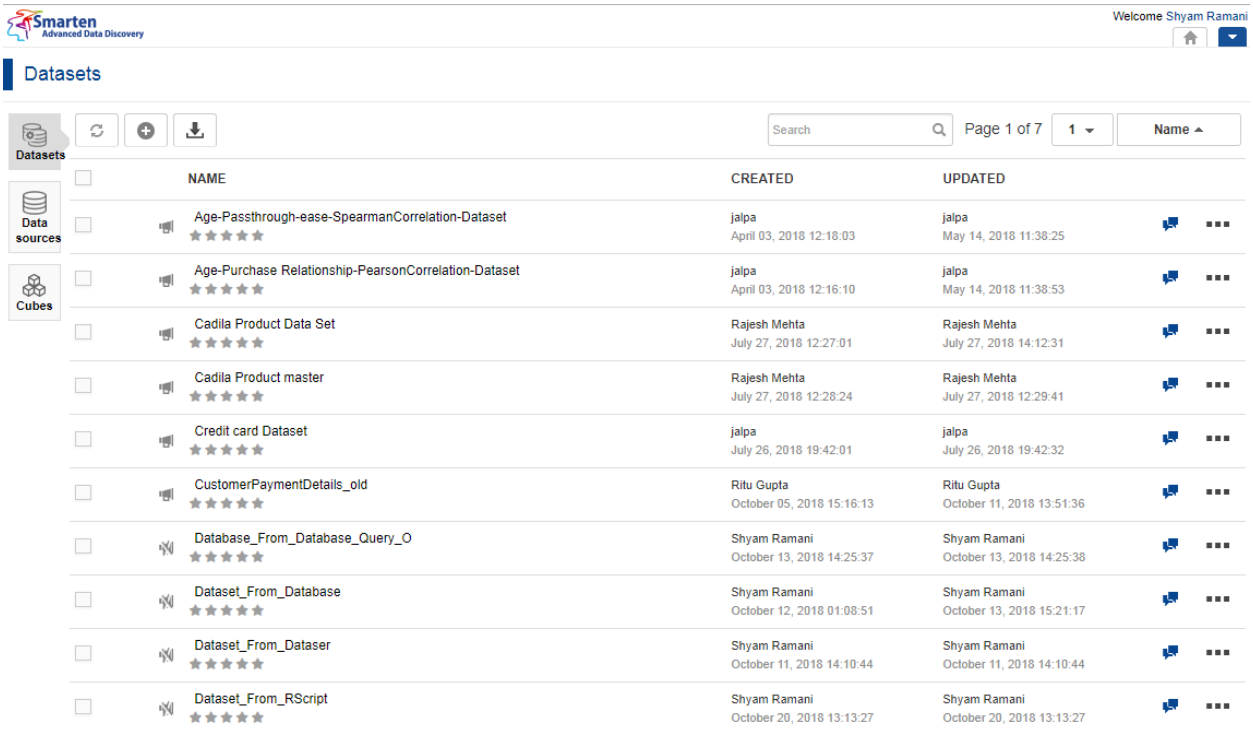
Procedure

1. Click **Open** ->**Data** from the menu.



MENU OPTION – OPEN DATA

The system displays the following page.

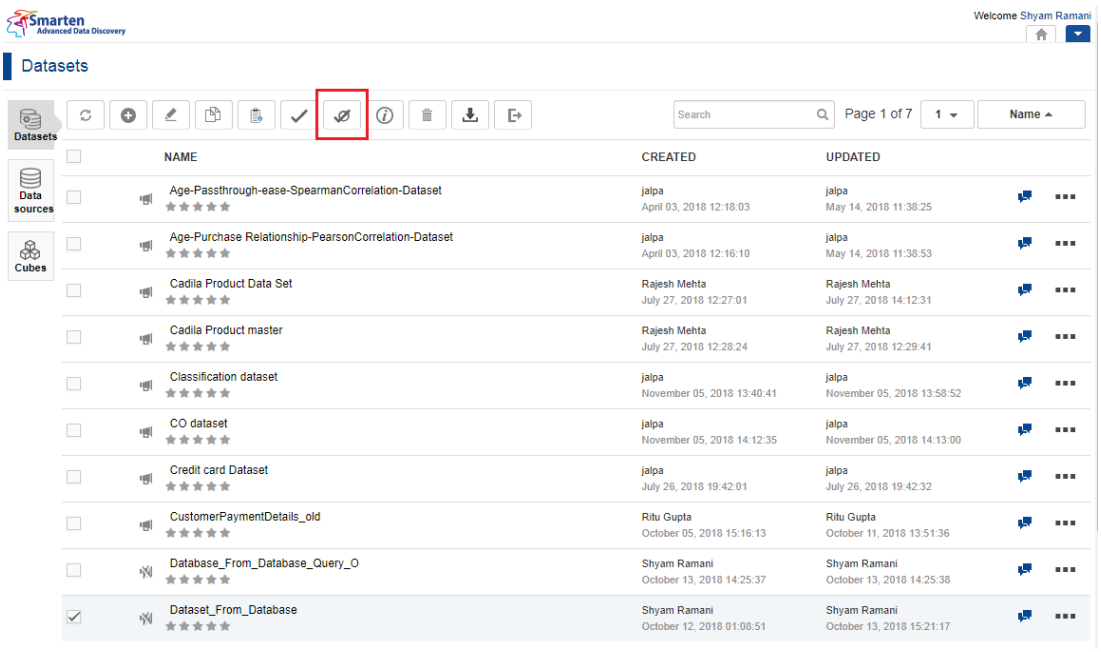


OPEN A DATASET – SELECTING A DATASET

2. Select the checkbox adjacent to the dataset you want to unmark.

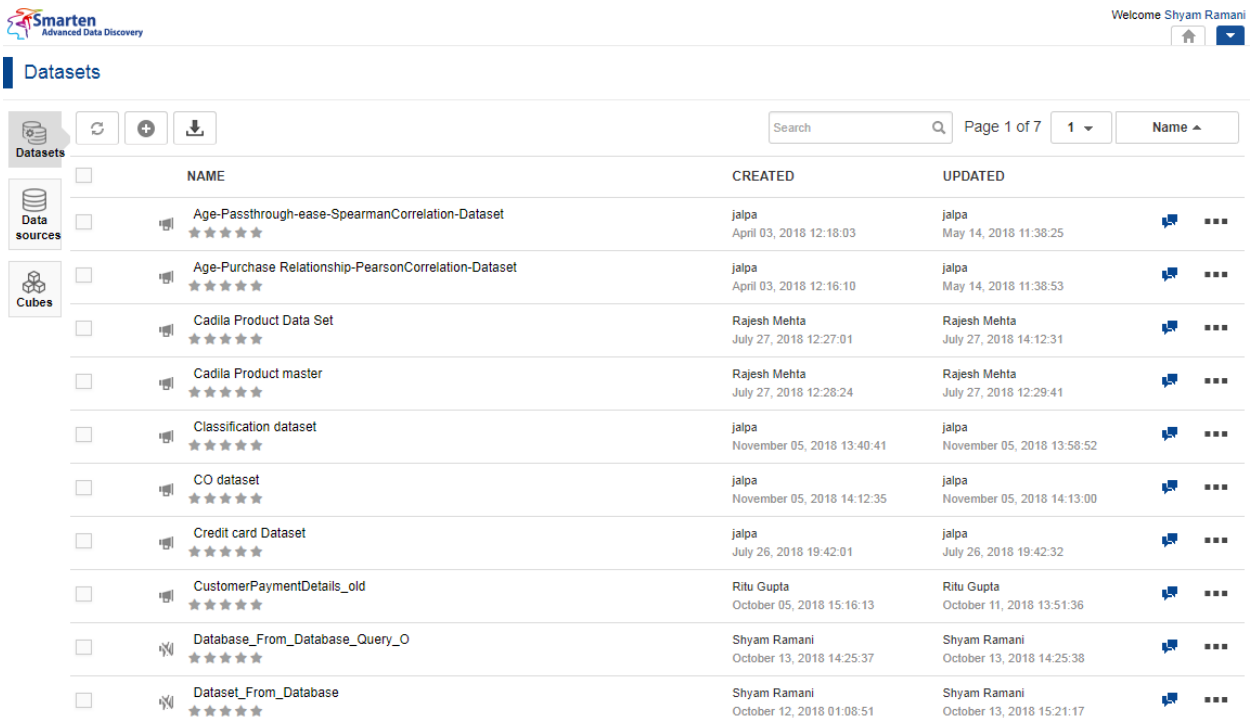
> Note:
> You can only unmark the datasets that are marked as IT approved.

3. Click the Unmark IT Approved icon.



UNMARKING A DATASET – CLICKING THE UNMARK IT APPROVED ICON

The system unmarks the dataset as IT approved and the check symbol adjacent to the dataset is no longer available.



UNMARKING A DATASET – THE CHECK MARK FOR IT CERTIFIED IS REMOVED

7.2 Operations on Dataset

7.2.1 Managing Columns in the Resultset

You can hide the columns that are displayed in a dataset, delete columns, export data from selected columns, and arrange the order in which they are displayed.

7.2.1.1 Displaying Columns in a Dataset

You can display the hidden columns in a dataset.

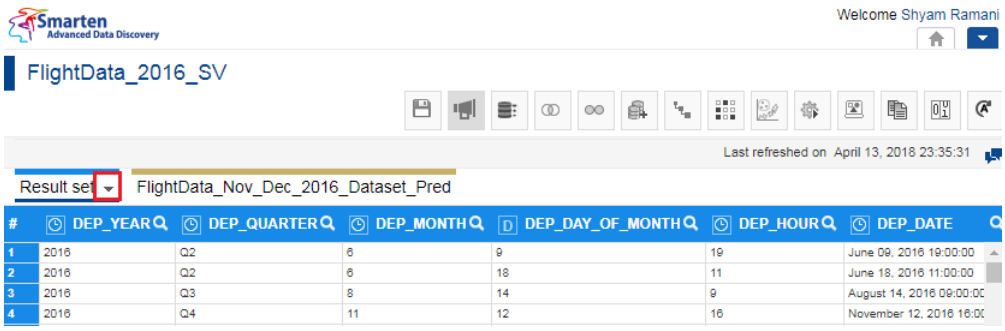
> Note:
> This is a front-end operation and does not affect any data in a dataset.

About this task

Use this task to display the hidden columns in a dataset.

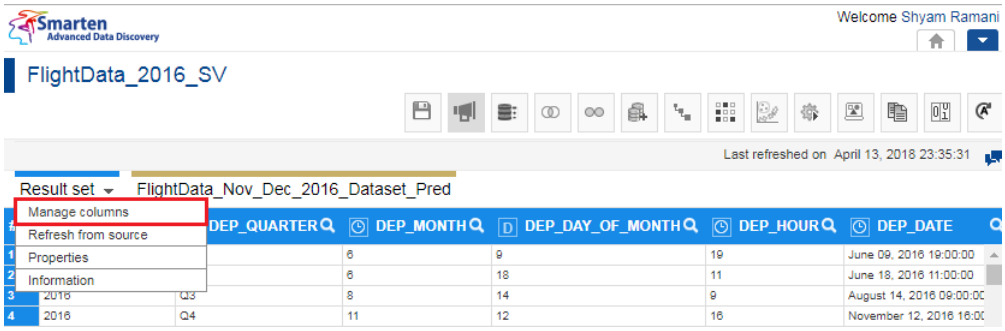
Procedure

1. Open the dataset for which you want to display the hidden columns.
2. Click the Result set menu icon.



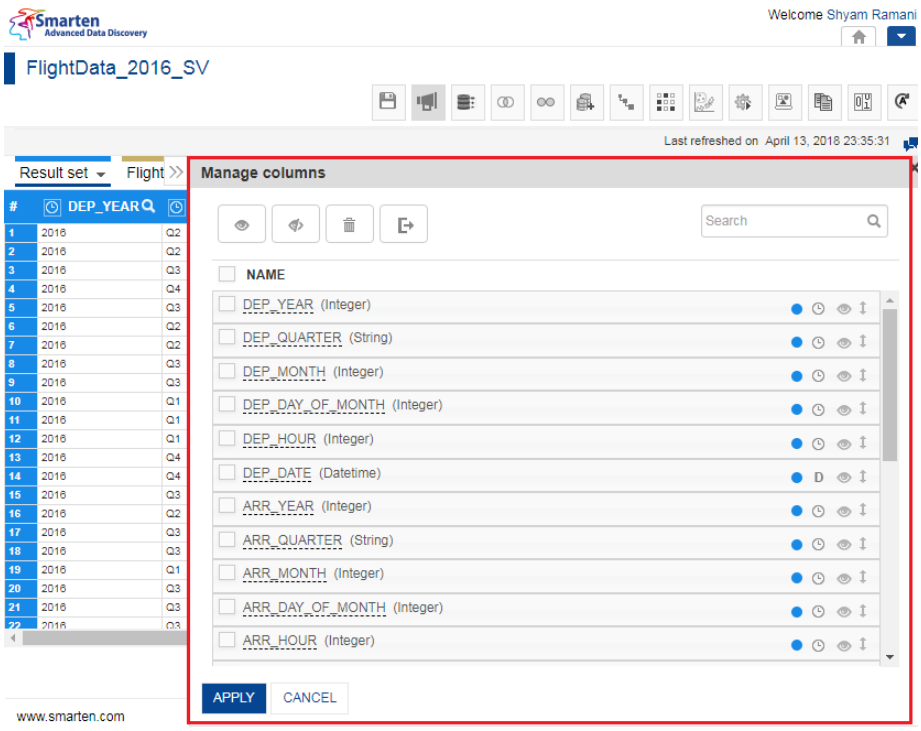
MANAGE COLUMNS—THE RESULT SET MENU ICON

3. Click **Manage columns** from the menu.



MANAGE COLUMNS—THE MANAGE COLUMNS OPTION

The system displays the **Manage columns** dialog box.

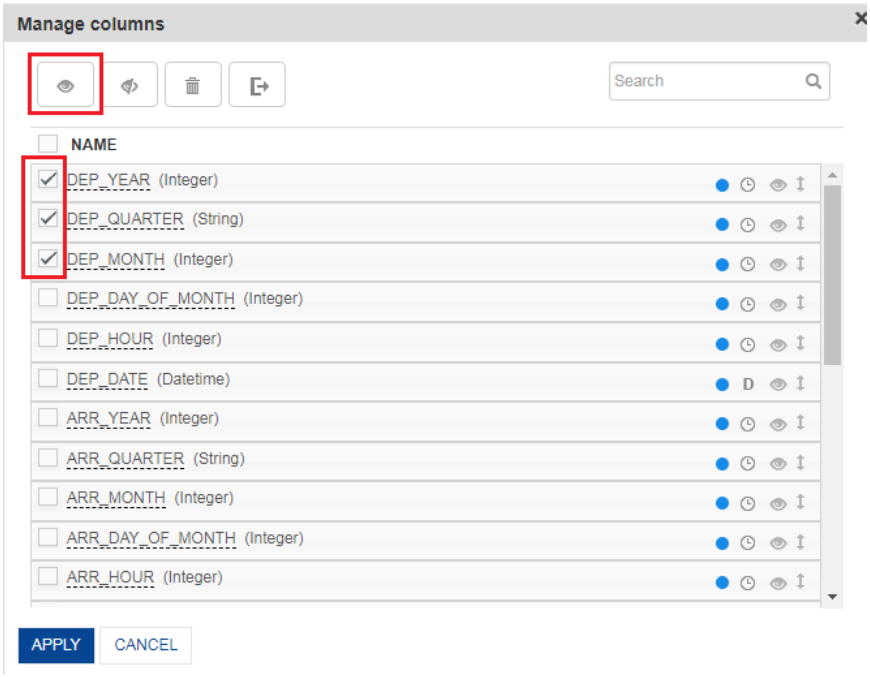


MANAGE COLUMNS—THE MANAGE COLUMNS DIALOG BOX

The dialog box displays the following information about the columns available in the dataset:

- : This icon indicates the color theme of the dataset a column belongs to.
- : This icon indicates that the column is marked as a time dimension.
- : This icon indicates that the column is marked as a dimension.
- : This icon indicates that the column is marked as a measure.
- : This icon indicates that the column is marked as a geo dimension.
- : This icon indicates whether the column is visible or hidden.
- : This icon allows you to move a column up or down to change the position of that column.

4. Select the check box next to the columns to select those columns.
5. Click the Visible icon to display those columns in the dataset.



MANAGE COLUMNS—OPTION TO MAKE A COLUMN VISIBLE

6. Click **APPLY**.

7.2.1.2 Hiding Columns in a Dataset

You can hide the visible columns in a dataset.

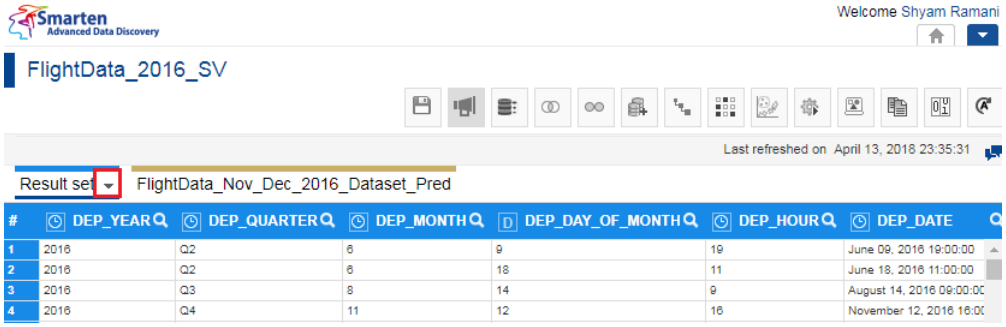
> Note:
> This is a front-end operation and does not affect any data in a dataset.

About this task

Use this task to hide the visible columns in a dataset.

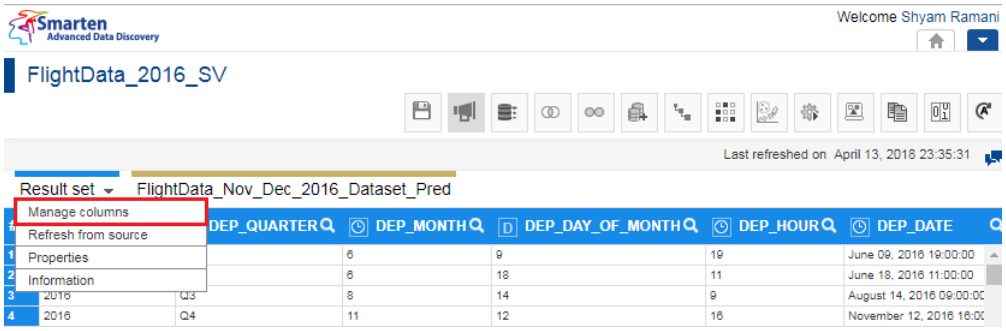
Procedure

1. Open the dataset for which you want to hide the visible columns.
2. Click the Result set menu icon.



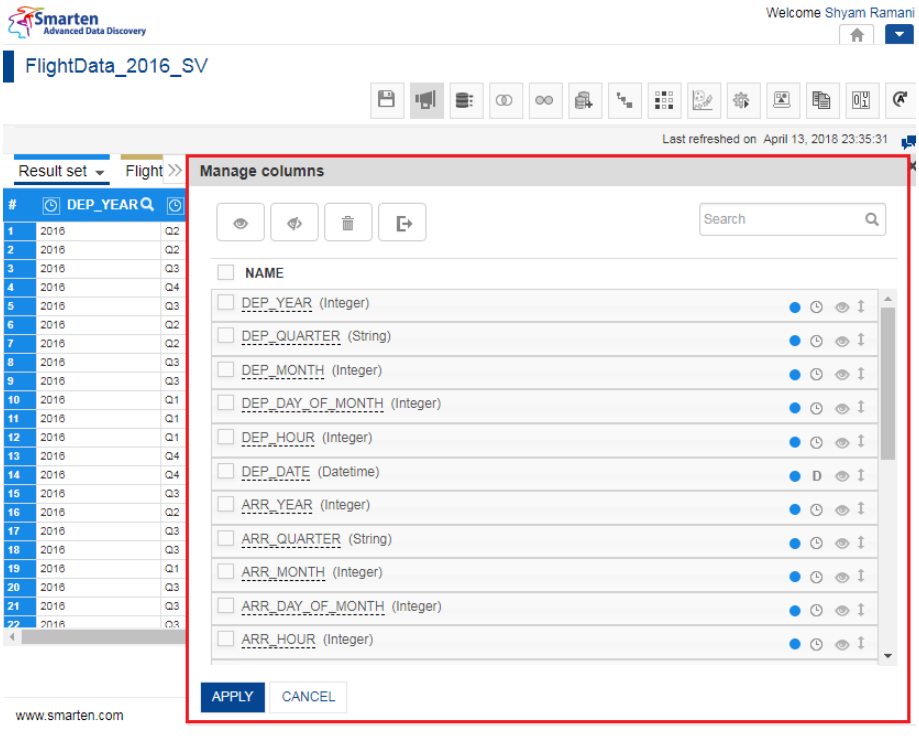
MANAGE COLUMNS—THE RESULT SET MENU ICON

3. Click **Manage columns** from the menu.



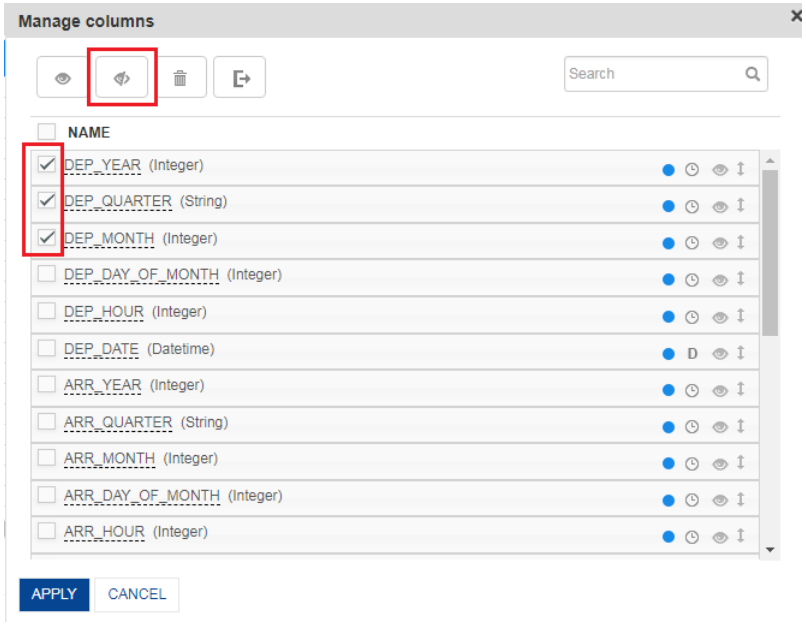
MANAGE COLUMNS—THE MANAGE COLUMNS OPTION

The system displays the **Manage columns** dialog box.



MANAGE COLUMNS—THE MANAGE COLUMNS DIALOG BOX

4. Select the check box next to the columns to select those columns.
5. Click the Hide icon to hide those columns in the dataset.



MANAGE COLUMNS—OPTION TO HIDE A COLUMN

6. Click **APPLY**.

7.2.1.3 Deleting Columns in a Dataset

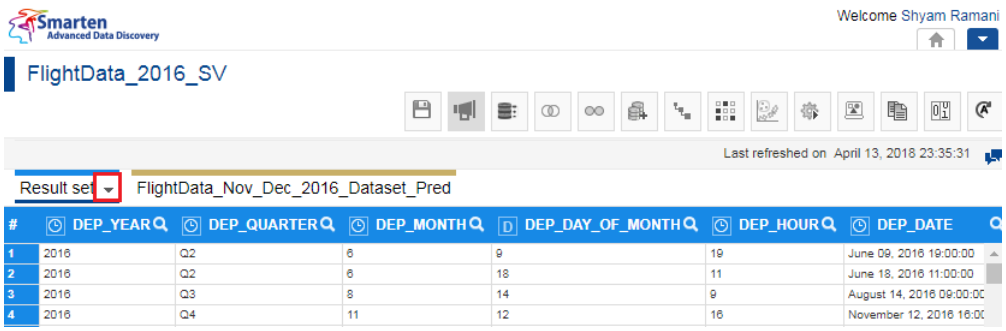
You can delete columns in a dataset. This will delete columns from the dataset.

About this task

Use this task to delete columns in a dataset.

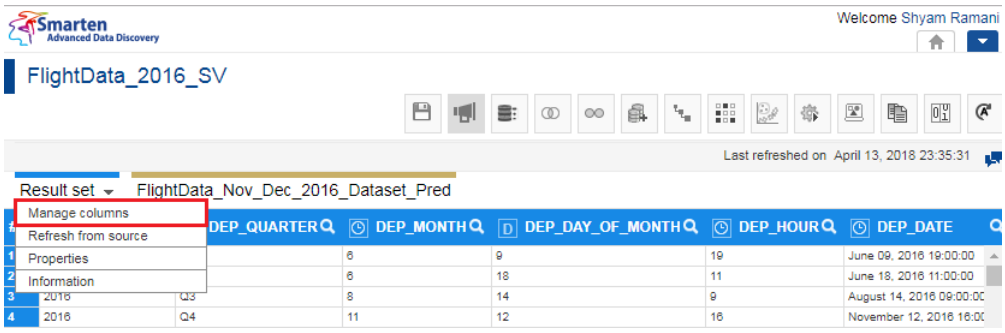
Procedure

1. Open the dataset from which you want to delete the columns.
2. Click the Result set menu icon.



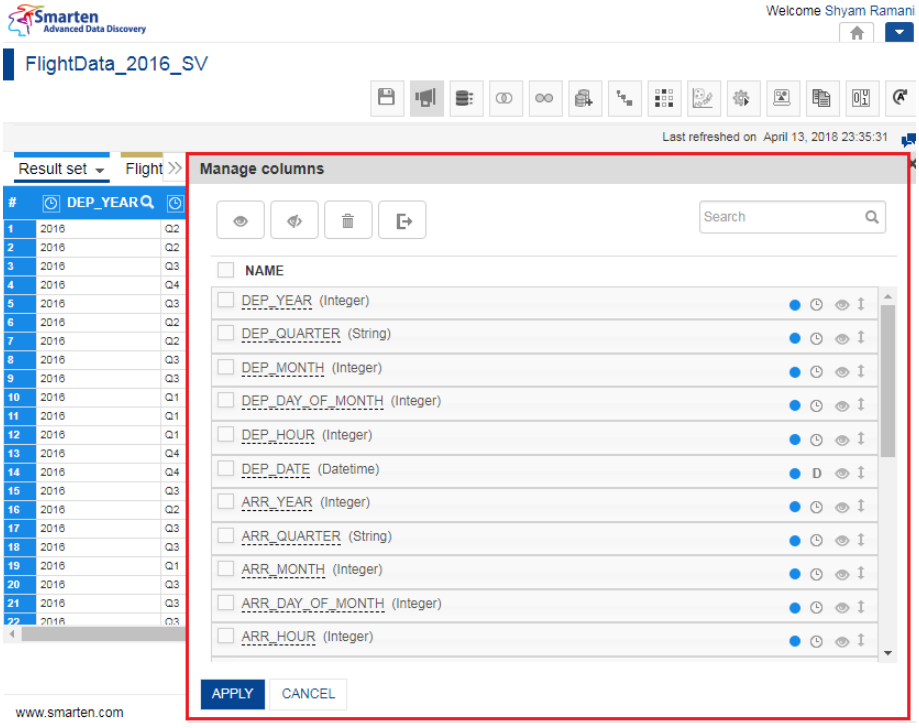
MANAGE COLUMNS—THE RESULT SET MENU ICON

3. Click **Manage columns** from the menu.



MANAGE COLUMNS—THE MANAGE COLUMNS OPTION

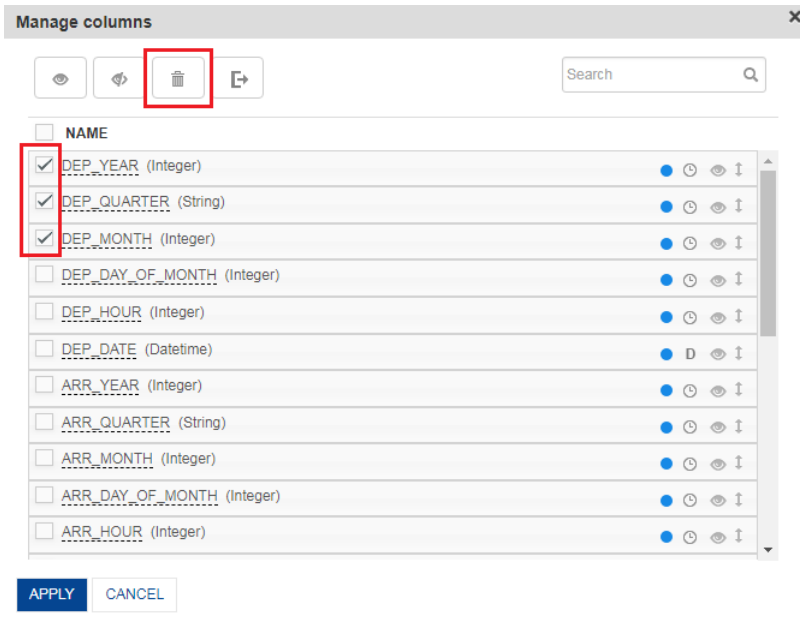
The system displays the **Manage columns** dialog box.



MANAGE COLUMNS—THE MANAGE COLUMNS DIALOG BOX

4. Select the check box next to the columns to select those columns.
5. Click the Delete icon to delete those columns in the dataset.

The system deletes the selected columns after confirmation.



MANAGE COLUMNS—OPTION TO DELETE A COLUMN

6. Click **APPLY**.

7.2.1.4 Exporting dataset metadata information

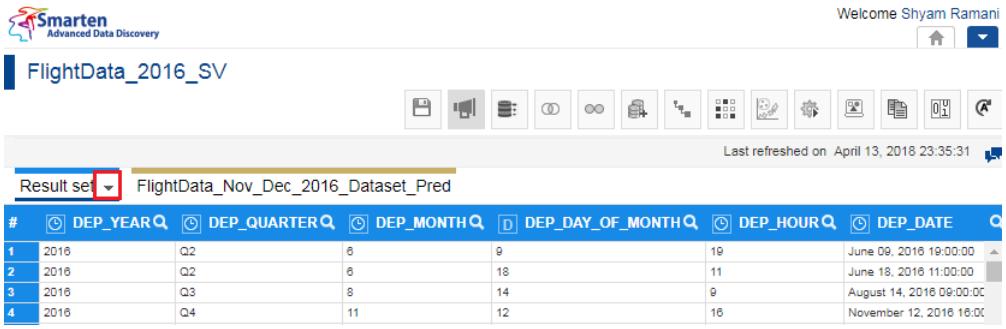
You can export dataset metadata information.

About this task

Use this task to export columns from a dataset.

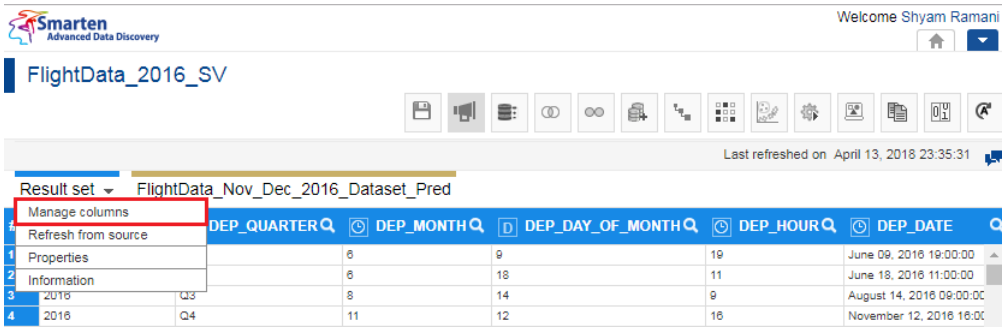
Procedure

1. Open the dataset from which you want to export metadata information of the columns.
2. Click the Result set menu icon.



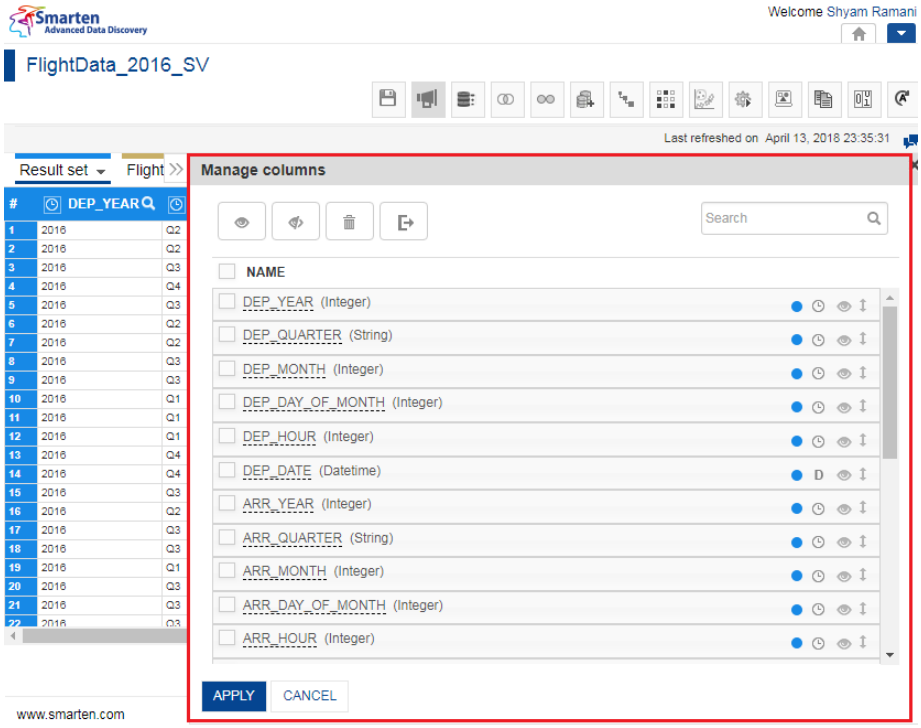
MANAGE COLUMNS—THE RESULT SET MENU ICON

3. Click **Manage columns** from the menu.



MANAGE COLUMNS—THE MANAGE COLUMNS OPTION

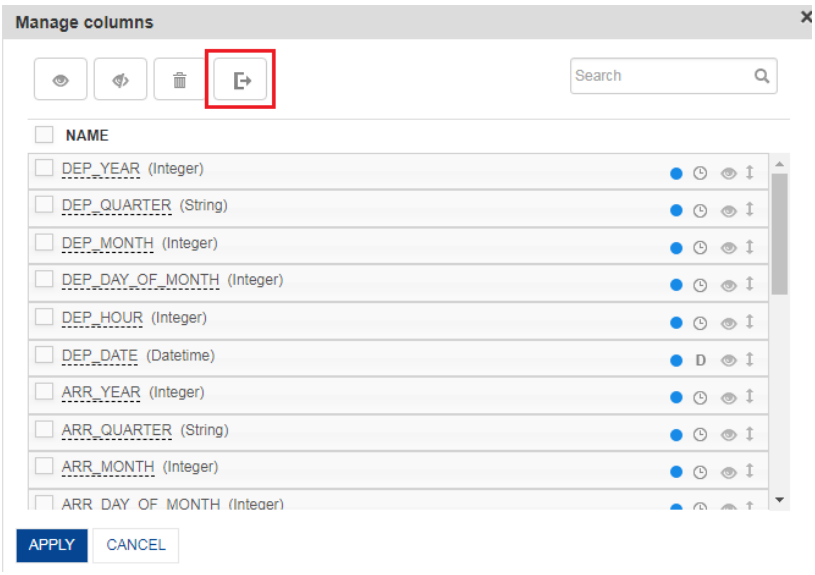
The system displays the **Manage columns** dialog box.



MANAGE COLUMNS—THE MANAGE COLUMNS DIALOG BOX

4. Click the Export icon to hide those columns in the dataset.

The system exports the metadata information of the dataset into an Excel file.



MANAGE COLUMNS—OPTION TO EXPORT COLUMNS

5. Click **APPLY**.

7.2.1.5 Arranging Columns in a Dataset

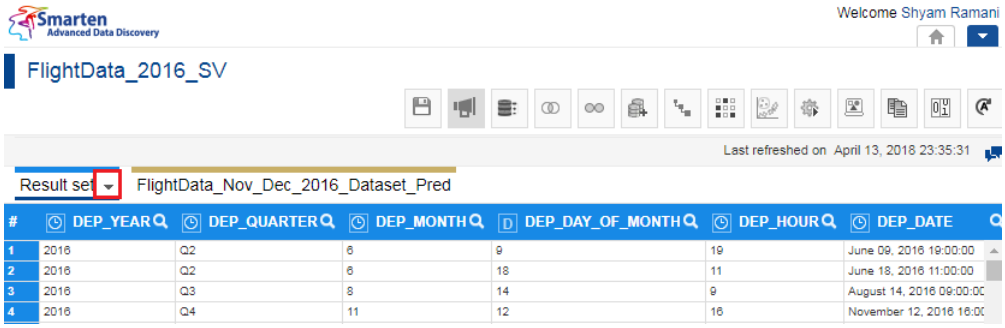
You can arrange columns from a dataset to display them in a specific order.

About this task

Use this task to arrange columns in a dataset.

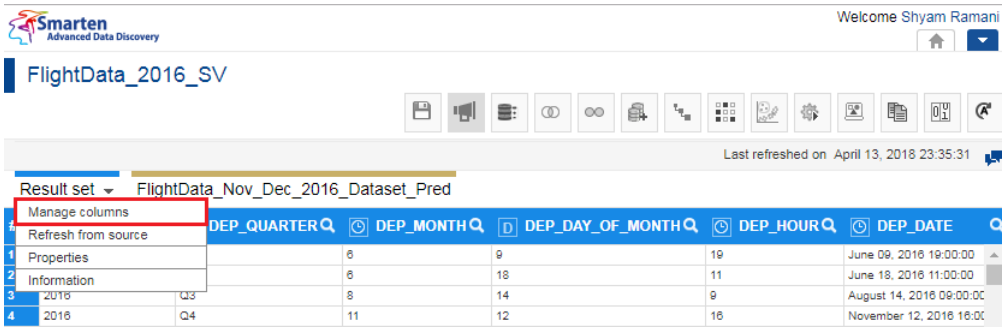
Procedure

1. Open the dataset for which you want to arrange the columns.
2. Click the Result set menu icon.



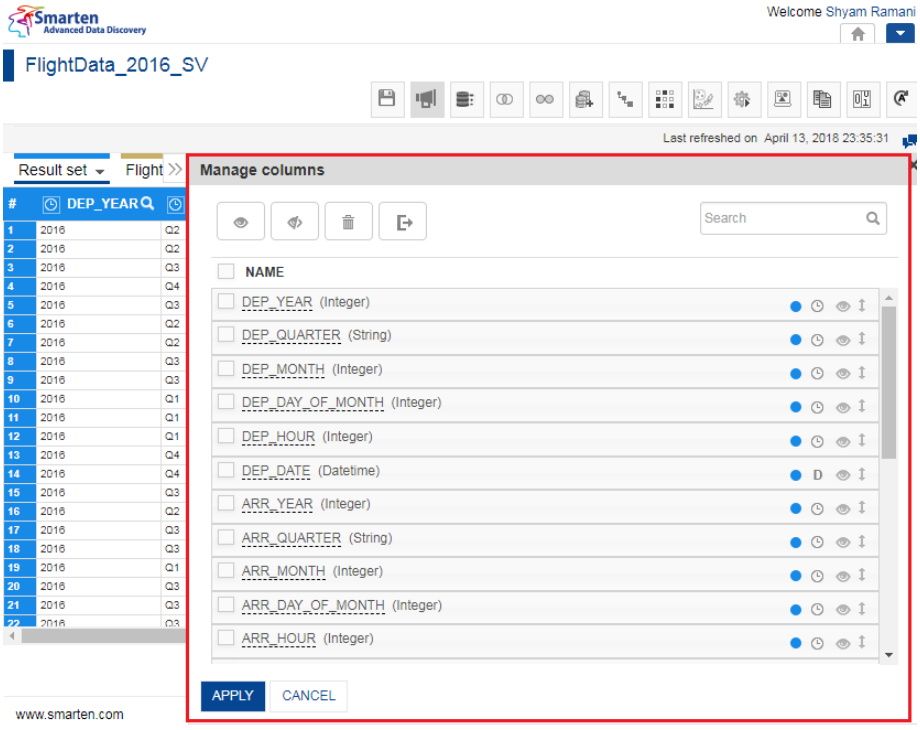
MANAGE COLUMNS—THE RESULT SET MENU ICON

3. Click **Manage columns** from the menu.



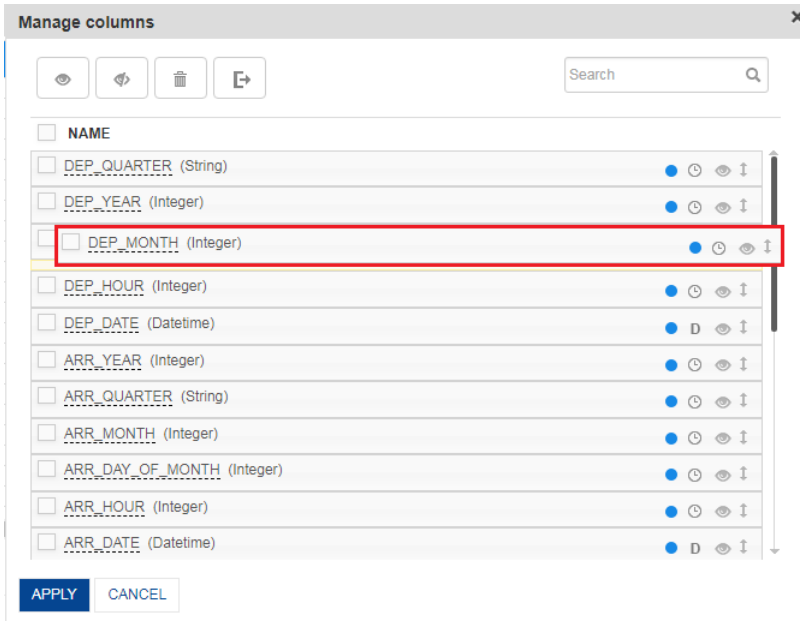
MANAGE COLUMNS—THE MANAGE COLUMNS OPTION

The system displays the **Manage columns** dialog box.



MANAGE COLUMNS—THE MANAGE COLUMNS DIALOG BOX

4. Drag the column that you want to change, and drop it at the desired position.



MANAGE COLUMNS—ARRANGING COLUMNS

5. Click **APPLY**.

7.2.2 Refreshing Data in a Dataset

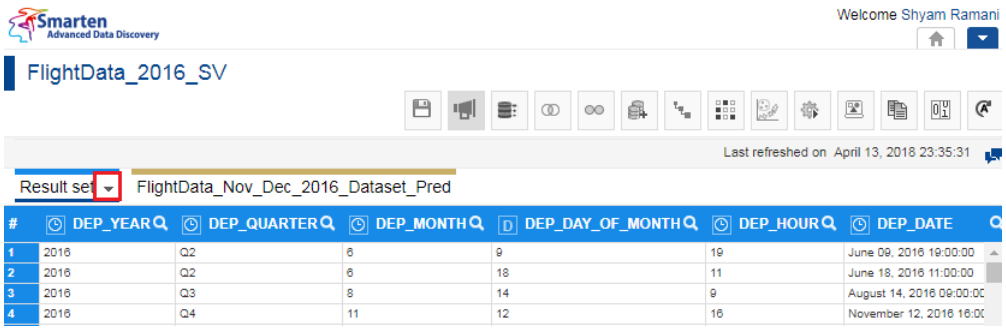
You can refresh the data available in a dataset with the latest data available in the data source.

About this task

Use this task to refresh data in a dataset.

Procedure

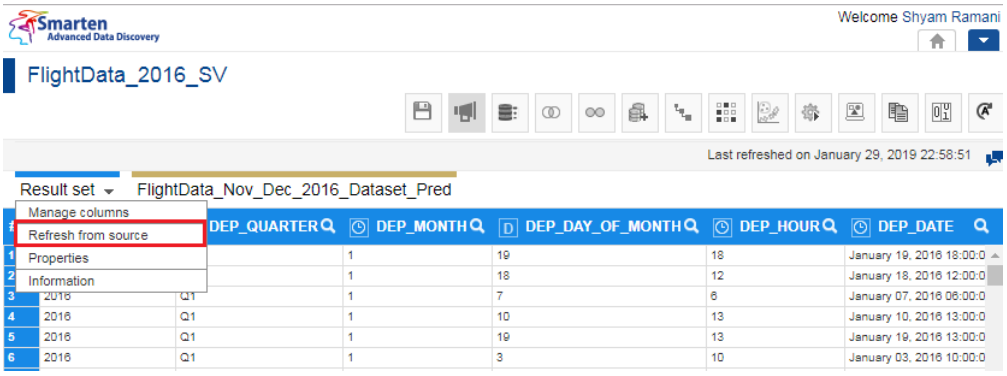
1. Open the dataset for which you want to refresh data.
2. Click the Result set menu icon.



MANAGE COLUMNS—THE RESULT SET MENU ICON

3. Click **Refresh from source** from the menu.

   The system refreshes the data with the latest data in the data source.



REFRESH DATA—THE REFRESH FROM SOURCE OPTION

7.2.3 Editing Dataset Properties

You can edit properties of a dataset. A user can edit the first month of the financial year.

7.2.3.1 Specify the First Month of the Financial Year

You can specify which month will be the first month of the current year. This feature allows you to specify the start of a financial year if it is different from the calendar year. For example, if the financial year is from April to March, you can specify April as the first month of the year.

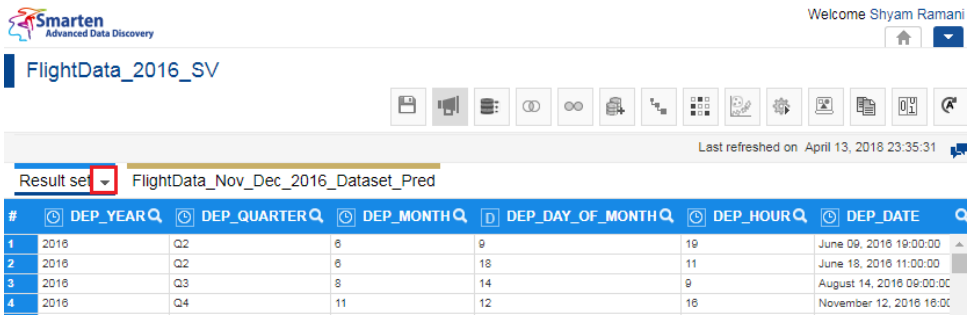
Reference: **Concept Manual > Designing the Data Model > Cube > Time Dimension > Time dimension based on a financial year**

About this task

Use this task to specify the first month of the financial year.

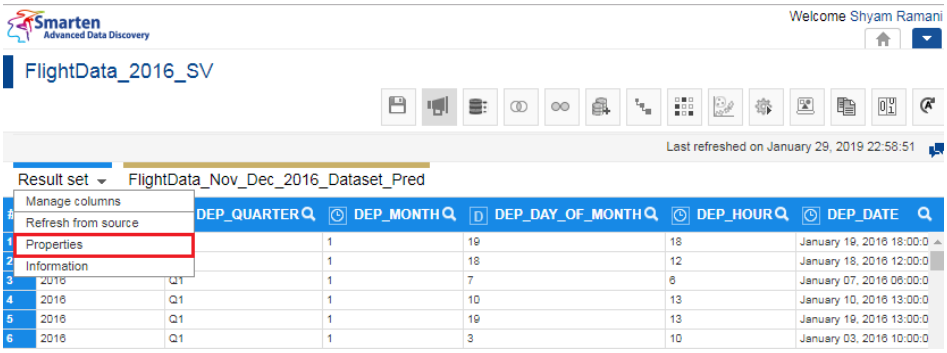
Procedure

1. Open the dataset for which you want to specify the first month of the year.
2. Click the Result set menu icon.



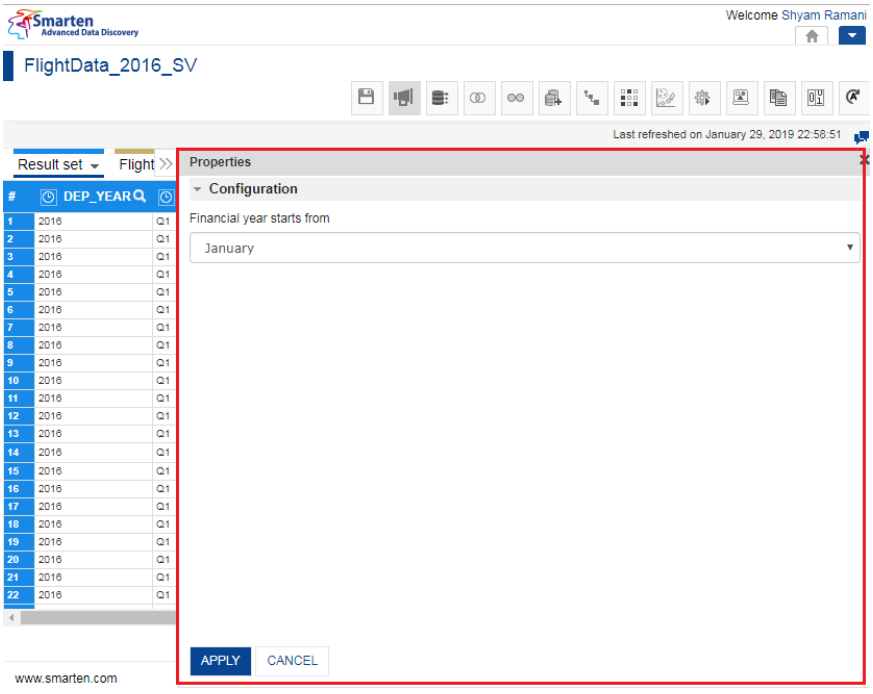
PROPERTIES—THE RESULT SET MENU ICON

3. Click **Properties** from the menu.



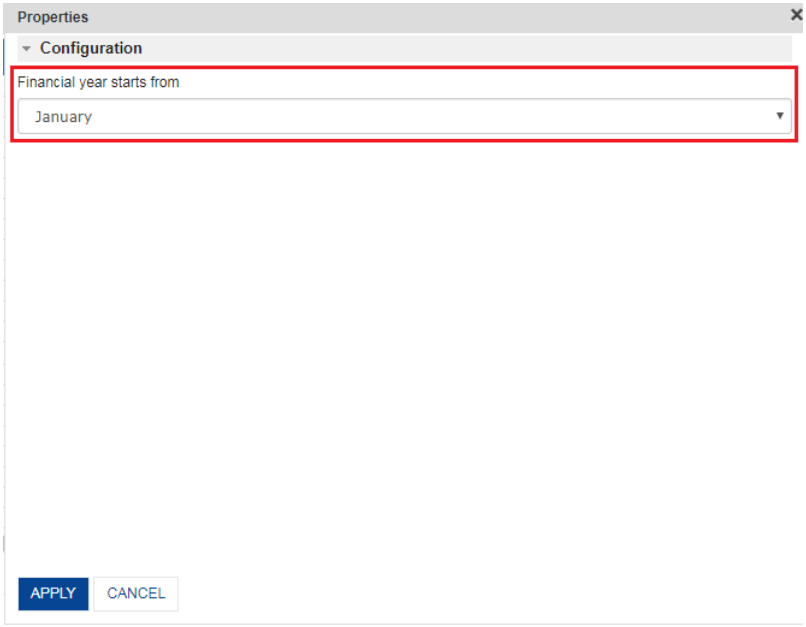
PROPERTIES—THE PROPERTIES OPTION

The system displays the **Properties** dialog box.



MANAGE COLUMNS—THE MANAGE COLUMNS DIALOG BOX

4. Select a month from the list to specify that month as the start of the financial year.



MANAGE COLUMNS—OPTION TO MAKE A COLUMN VISIBLE

5. Click **APPLY**.

7.2.4 Viewing Dataset Information

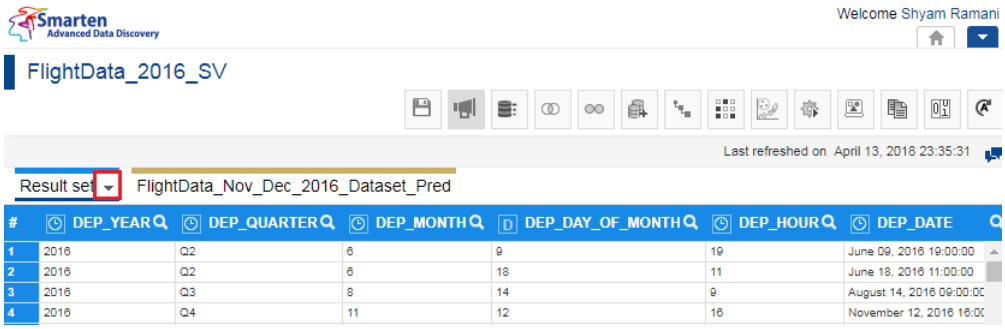
You can view information about a dataset, such as a name, last updated, data source, dataset size, and much more.

About this task

Use this task to view information about a dataset.

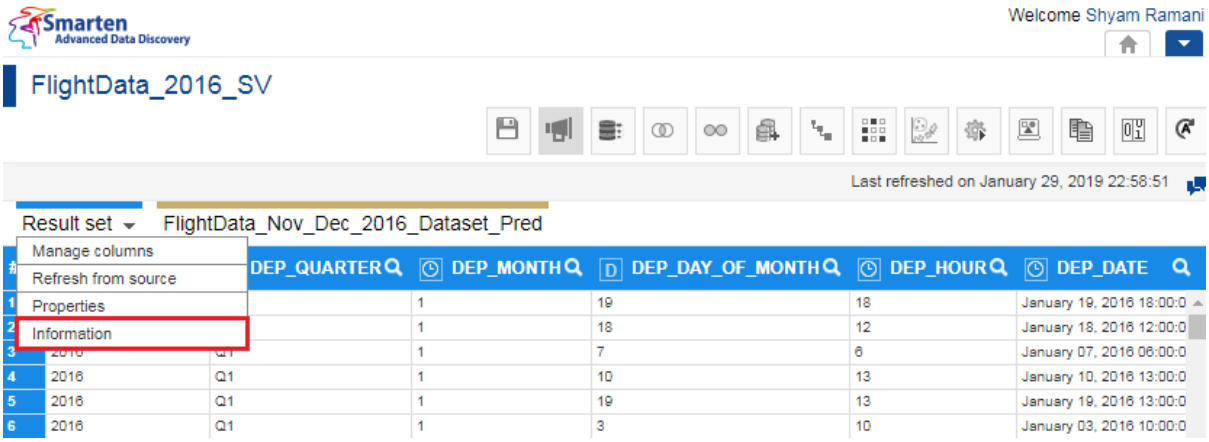
Procedure

1. Open the dataset for which you want to view information.
2. Click the Result set menu icon.



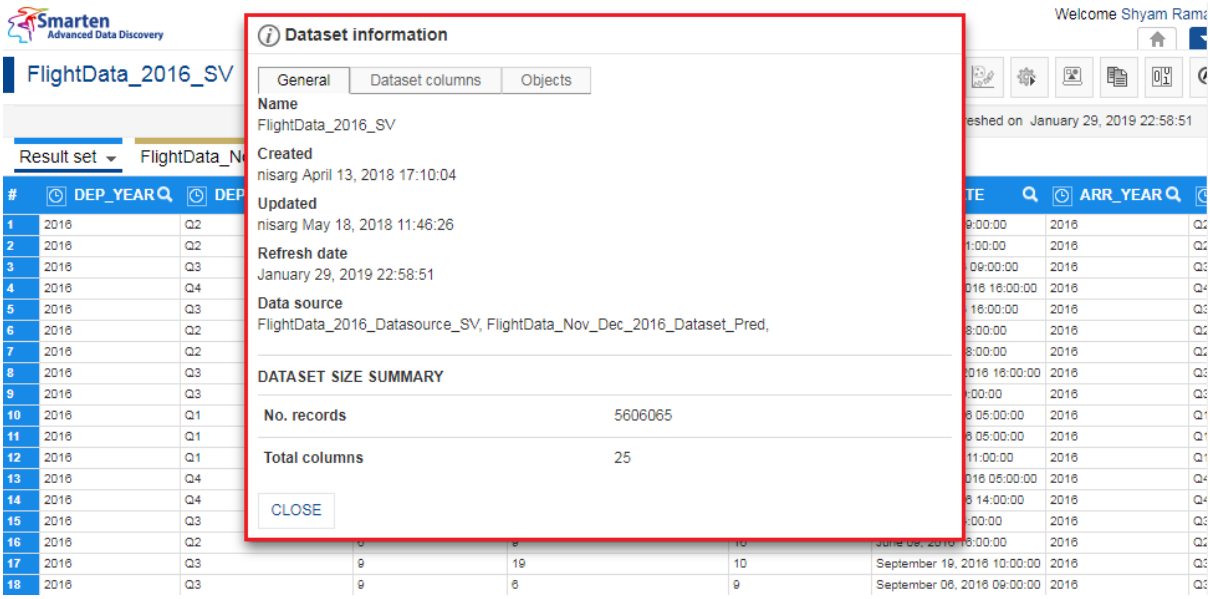
DATASET INFORMATION—THE RESULT SET MENU ICON

3. Click **Information** from the menu.



DATASET INFORMATION—THE INFORMATION OPTION

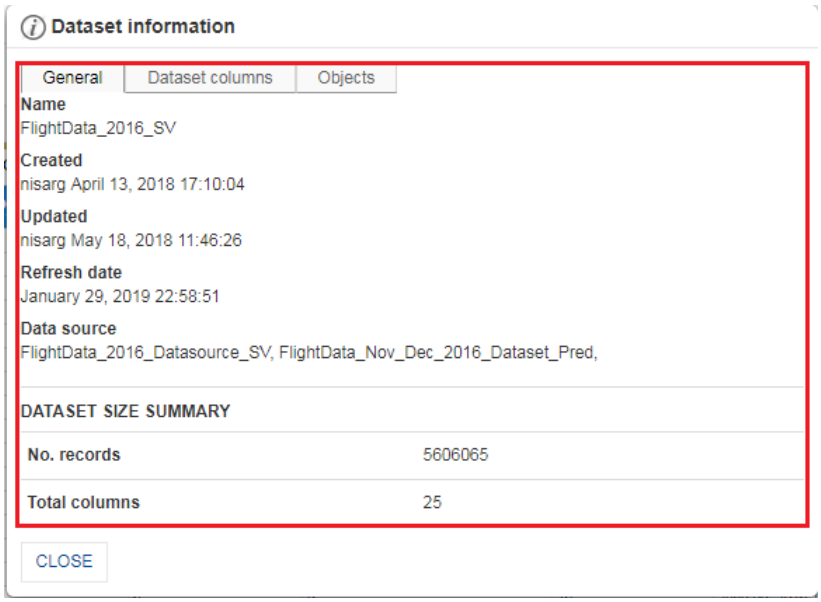
4. The system displays the **Dataset information** dialog box.



VIEW DATASET INFORMATION—THE DATASET INFORMATION DIALOG BOX

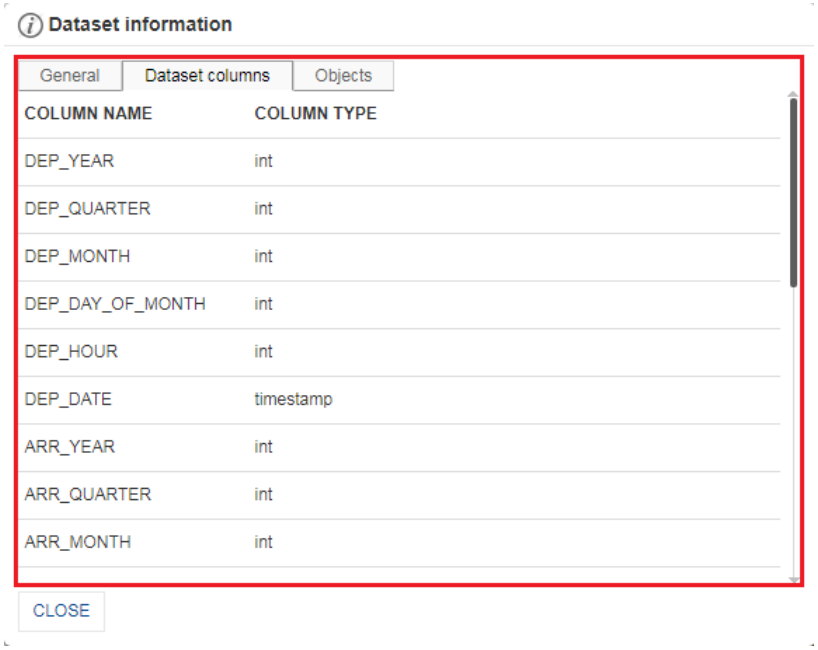
5. Click the **General** tab to view general information about the dataset. The following information is displayed:

   - **Name**: Name of the dataset.
   - **Created**: Date and time when the dataset was created.
   - **Updated**: Date and time when the dataset was last updated.
   - **Refresh date**: Date and time when the dataset was last refreshed.
   - **Data source**: Name of the data source used in the dataset.
   - **No. records**: Number of records available in the dataset.
   - **Total columns**: Number of columns available in the dataset.



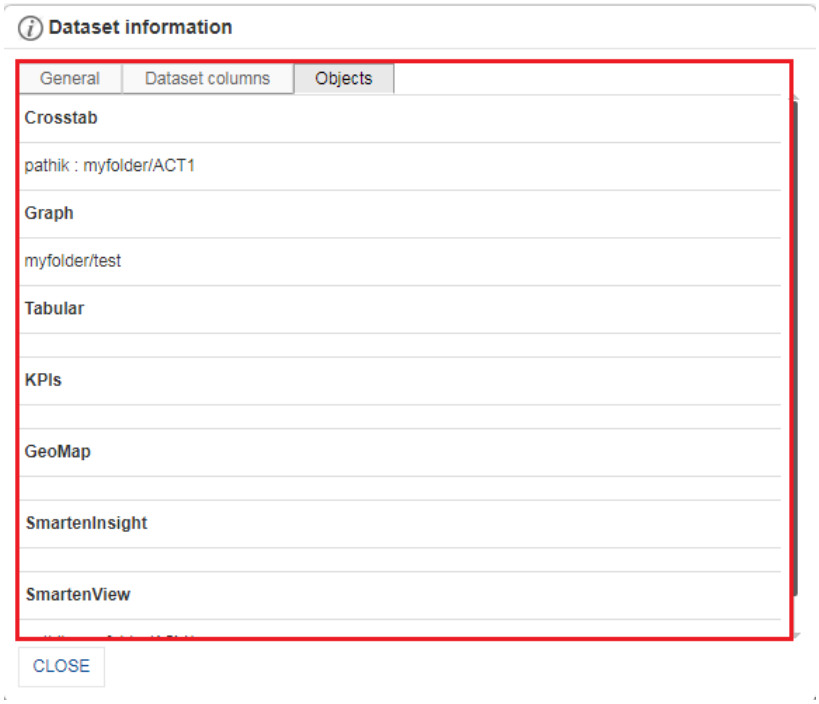
VIEWING DATASET INFORMATION—VIEWING GENERAL INFORMATION

6. Click the **Dataset columns** tab to view information regarding the dataset columns of the selected dataset. The tab displays the name of the columns and their data type.



VIEWING DATASET INFORMATION—VIEWING THE DATASET COLUMN INFORMATION

7. Click the **Objects** tab to view information regarding the objects using the selected dataset.



VIEWING DATASET INFORMATION—VIEWING THE OBJECTS RELATED INFORMATION

8. Click **CLOSE**.

7.2.5 Highlighting Data in a Dataset

Typical steps involved in creating a Data Source Profile for File (txt, csv, tsv) are as follows:

Users can identify the data by highlighting it for missing and inconsistent values, spaces, and duplication of rows and columns and accordingly take action for data correction.

> Note:
> These are the front-end operations that highlight records in the front-end interface and do not remove any records from the dataset.

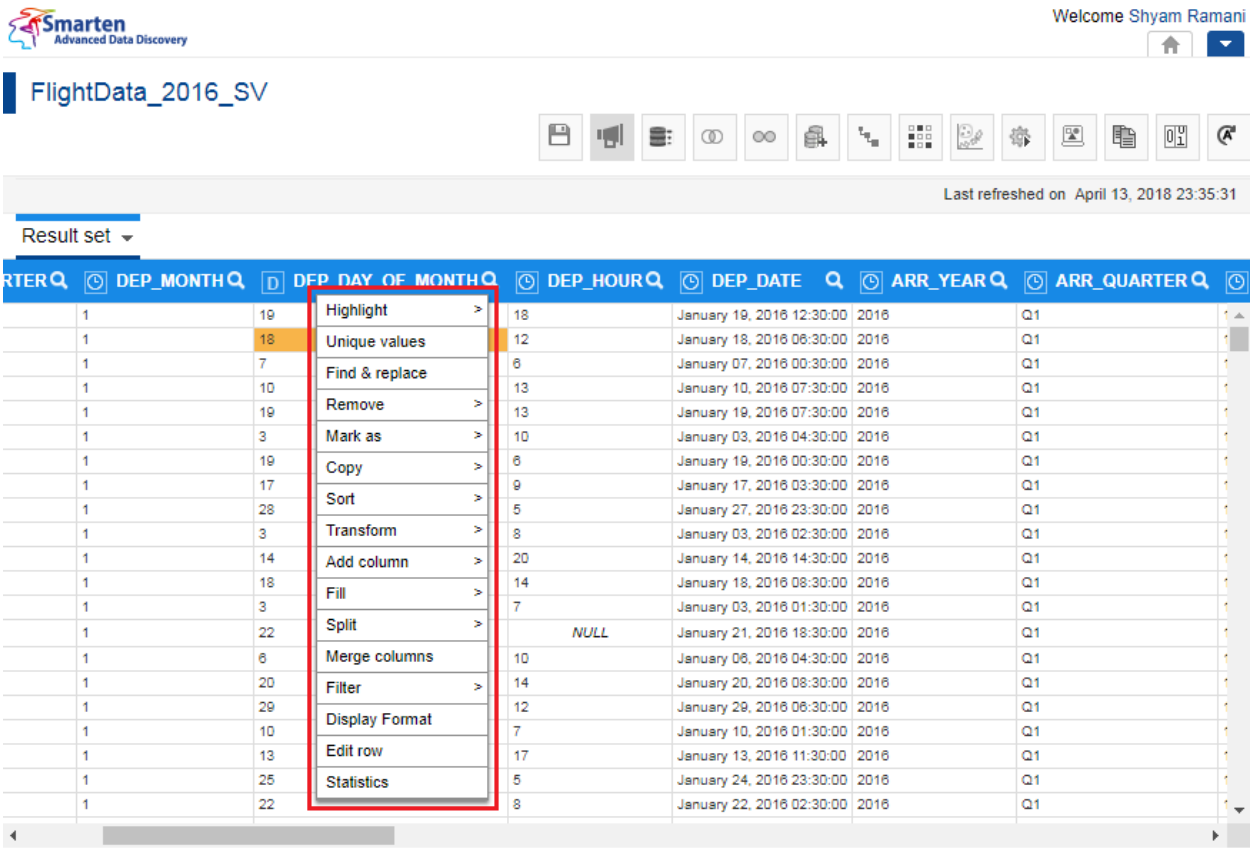
Reference: **Concept Manual > Explore Data > Highlight**

About this task

Use this task to highlight certain data in a dataset.

Procedure

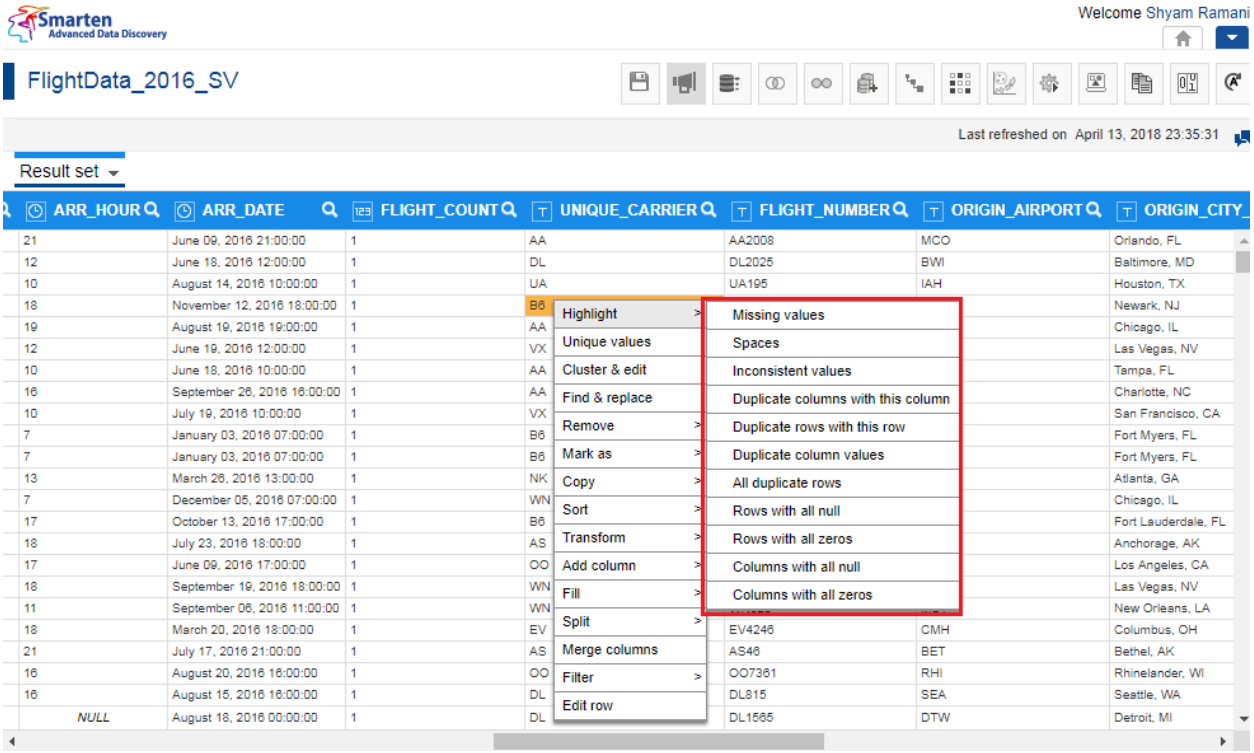
1. Open the dataset for which you want to highlight data.
2. Right-click in the column from which you want to highlight data.
   The system displays the context menu.



HIGHLIGHT VALUES—THE CONTEXT MENU

3. Click **Highlight** from the menu.

   The system displays the options available for highlight based on the data type of the column.



HIGHLIGHT VALUES—OPTIONS AVAILABLE FOR THE HIGHLIGHT MENU

4. Click any of the options in the following table to apply that operation.

Option	Description
Missing values	Use this option to highlight all the cells of the selected column that contain null or blank values.
Spaces	Use this option to highlight all the cells of the selected column that have spaces in the data. Note: This option is only available for the columns with the string data type.
Inconsistent values	Use this option to highlight all the inconsistent values in the selected column, such as a string type of column having integer values and vice versa or capitalization not staying the same throughout and many more.
Duplicate columns with this column	Use this option to highlight all the columns that contain exactly the same data as the selected column.
Duplicate row with this row	Use this option to highlight all the rows that contain exactly the same data as the selected row.
Duplicate column values	Use this option to highlight all duplicate values in the selected column. Each set of duplicate values is highlighted with a different color.
All duplicate rows	Use this option to highlight a set of rows that contain exactly the same data. Each set of duplicate rows is highlighted with a different color.
Rows with all null	Use this option to highlight all the rows that have null in all the columns.
Rows with all zeros	Use this option to highlight all the rows that have zeros in all numeric columns.
Columns with all null	Use this option to highlight all the columns that have null in all rows.
Columns with all zeros	Use this option to highlight all the columns that have zeros in all rows.

7.2.6 Identifying Unique Values

You can identify unique values and the number of times a value is repeated in a particular column. This function allows you to edit or delete a value. If you delete a value, the system deletes all rows containing that value. Similarly, if you edit a value, the system replaces the value in all rows with the new value.

Reference: **Concept Manual > Explore Data > Unique Values**

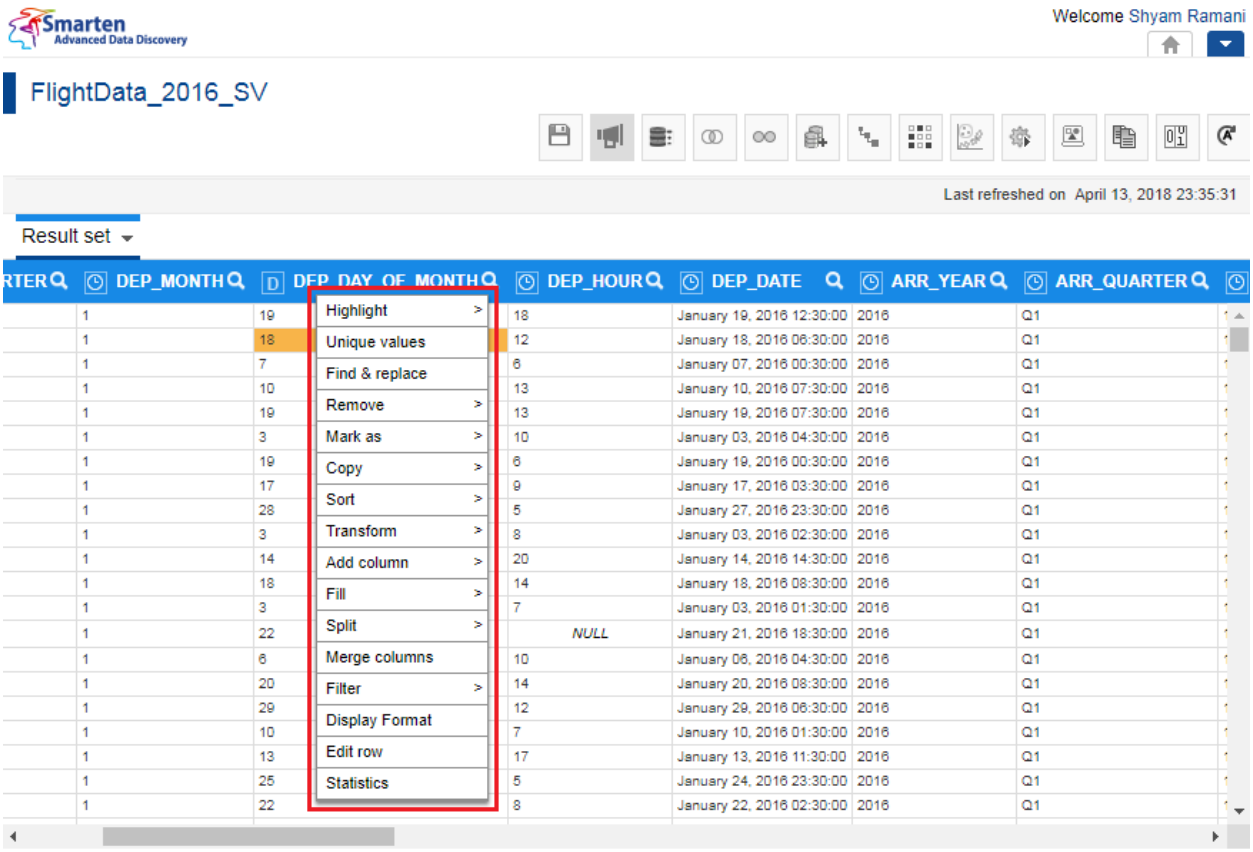
About this task

Use this task to identify unique values in a column.

Procedure

1. Open the dataset in which you want to identify unique values.
2. Right-click in the column from which you want to identify unique values.

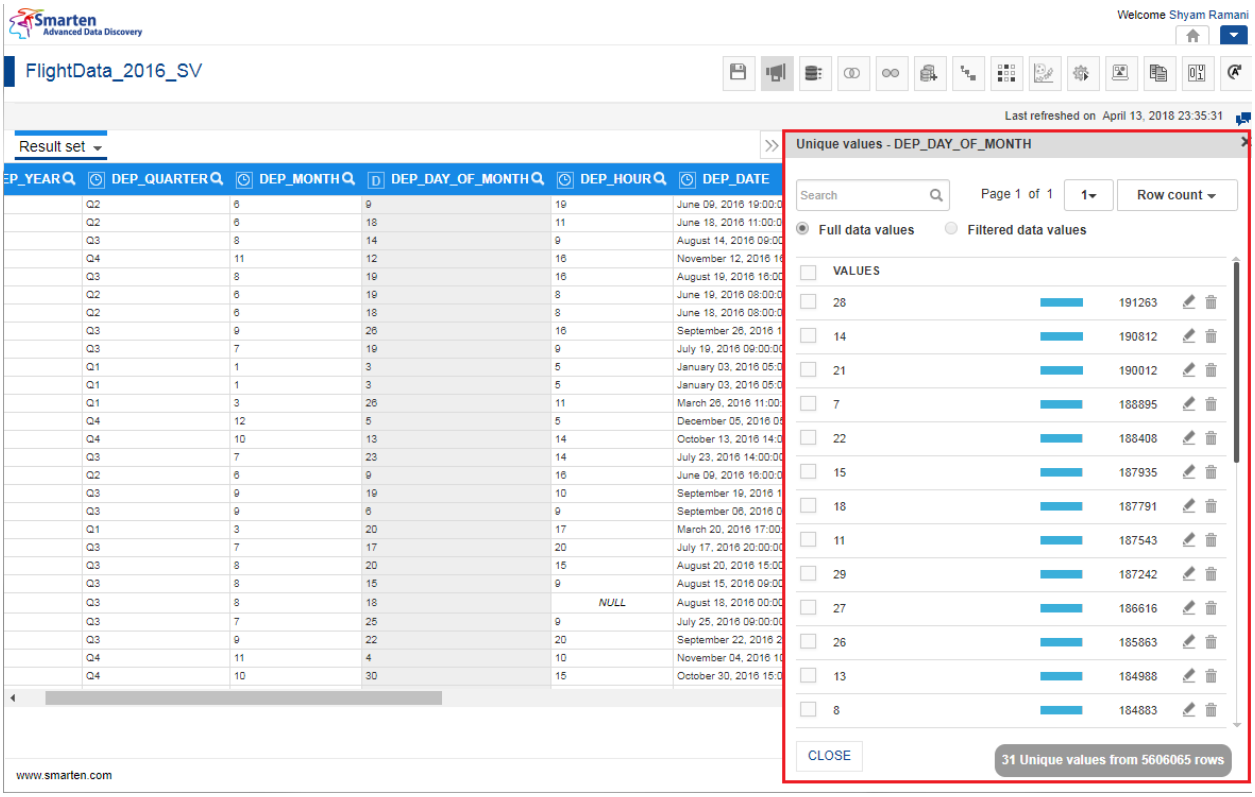
   The system displays the context menu.



UNIQUE VALUES—THE CONTEXT MENU

3. Click **Unique values** from the menu.

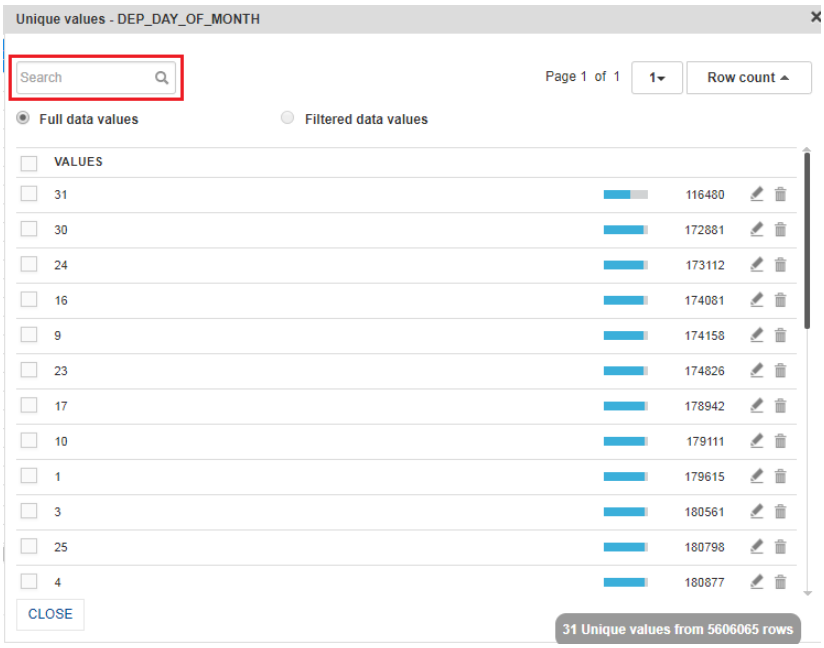
   The system displays the **Unique values** dialog box.



UNIQUE VALUES—DIALOG BOX DISPLAYING UNIQUE VALUES

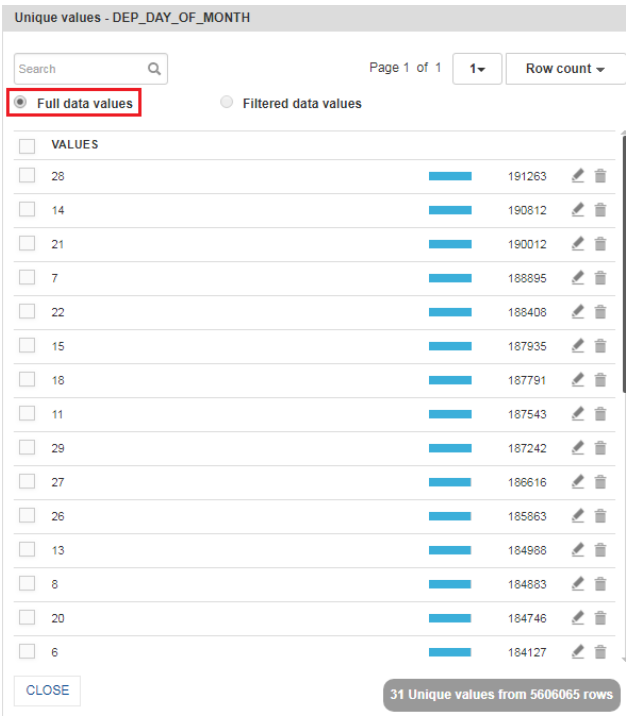
The system displays the unique values in the dataset along with the number of rows a value appears in the dataset and options to edit and delete a unique value. The system also displays a total number of unique values available from the specific number of rows available in a dataset.

4. You can type a keyword in the Search box to search for a unique value.



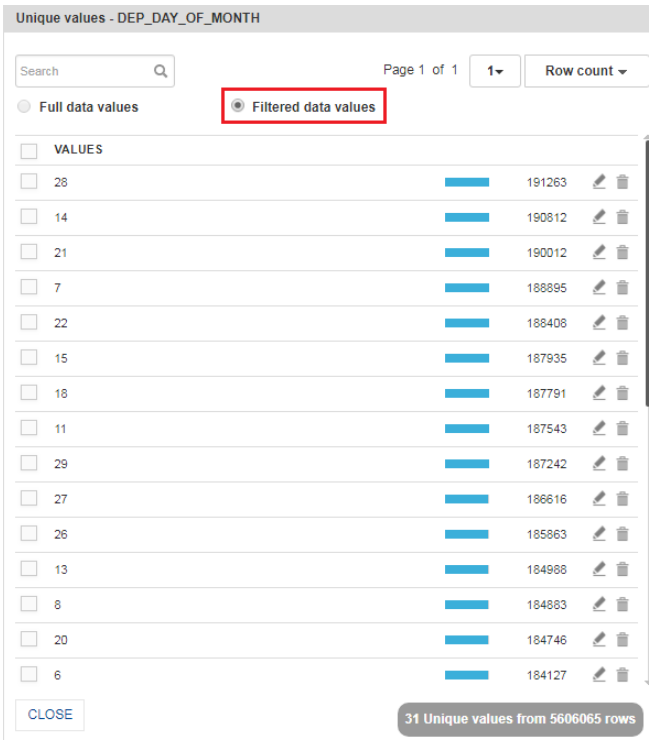
UNIQUE VALUES—SEARCHING A UNIQUE VALUE

5. You can select the Full data values if you want to perform the search in the entire dataset.



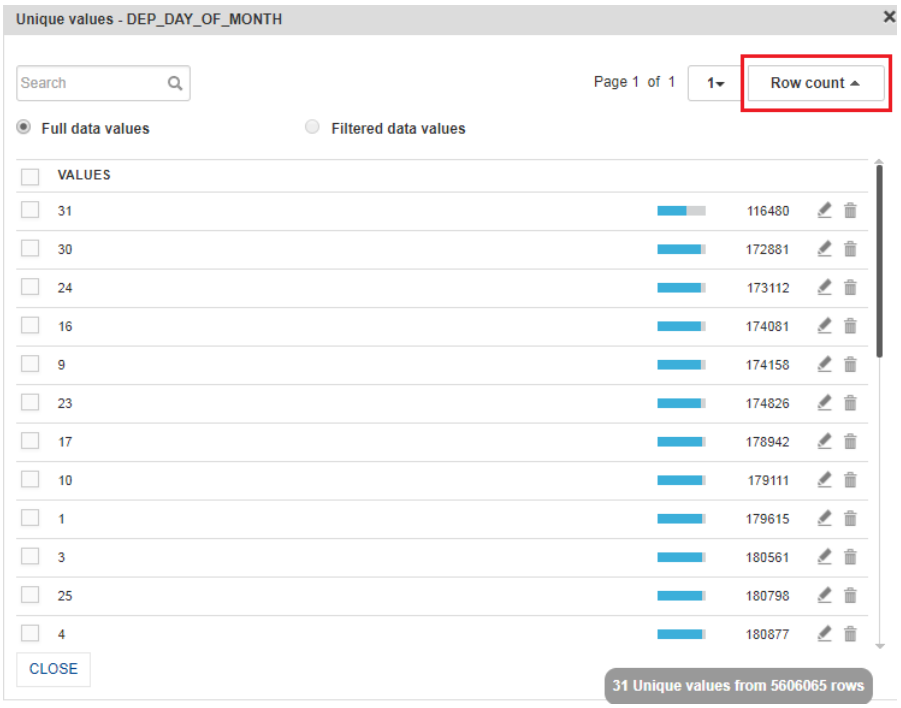
UNIQUE VALUES—SEARCH WITHIN FULL DATA

6. You can select the Filtered data values if you want to perform the search within only the unique values.



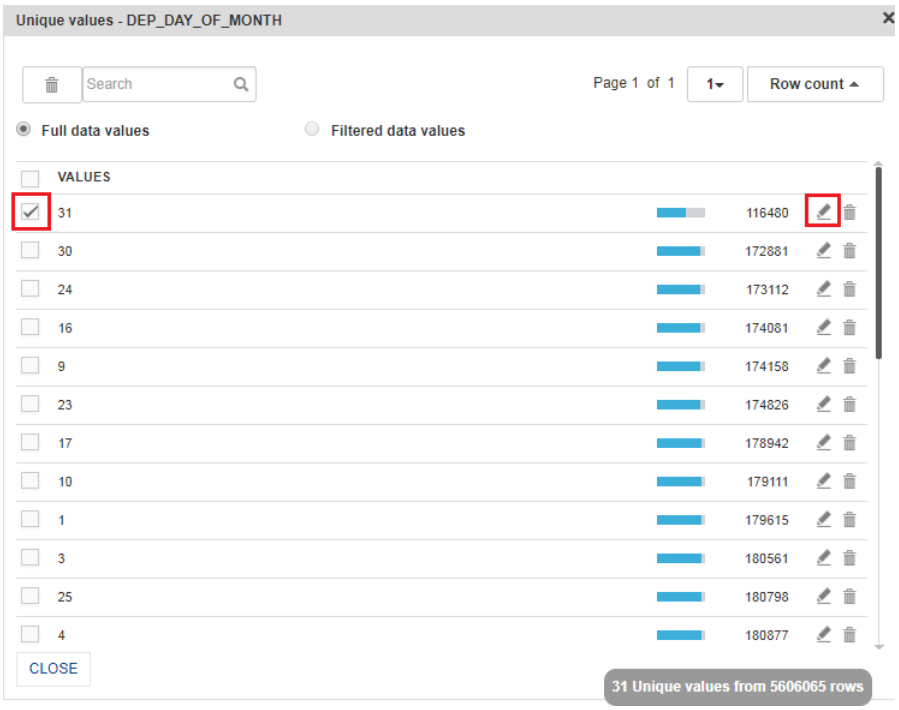
UNIQUE VALUES—SEARCH WITHIN FILTERED DATA

7. You can select an option to sort the unique values listed in the dialog box based on the number of rows the unique value appears in or the unique value itself.



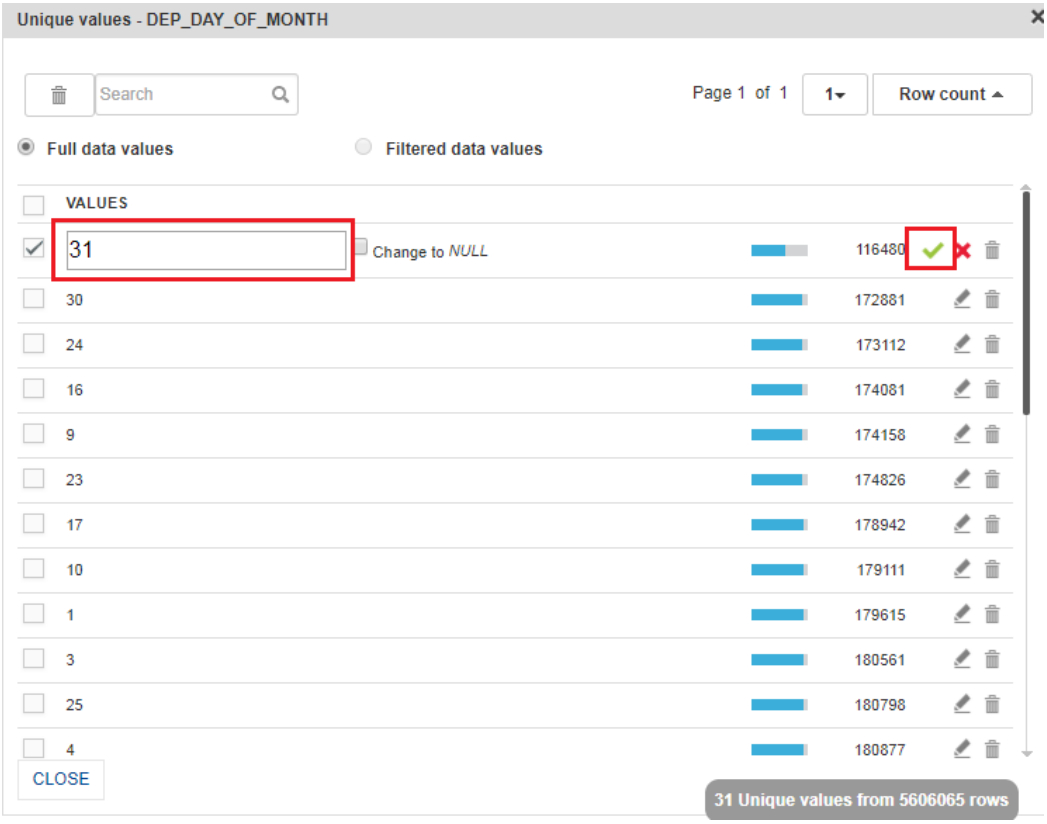
UNIQUE VALUES—SORTING UNIQUE VALUES

8. Select the unique value you want to edit, and then click the Edit icon adjacent to that value.



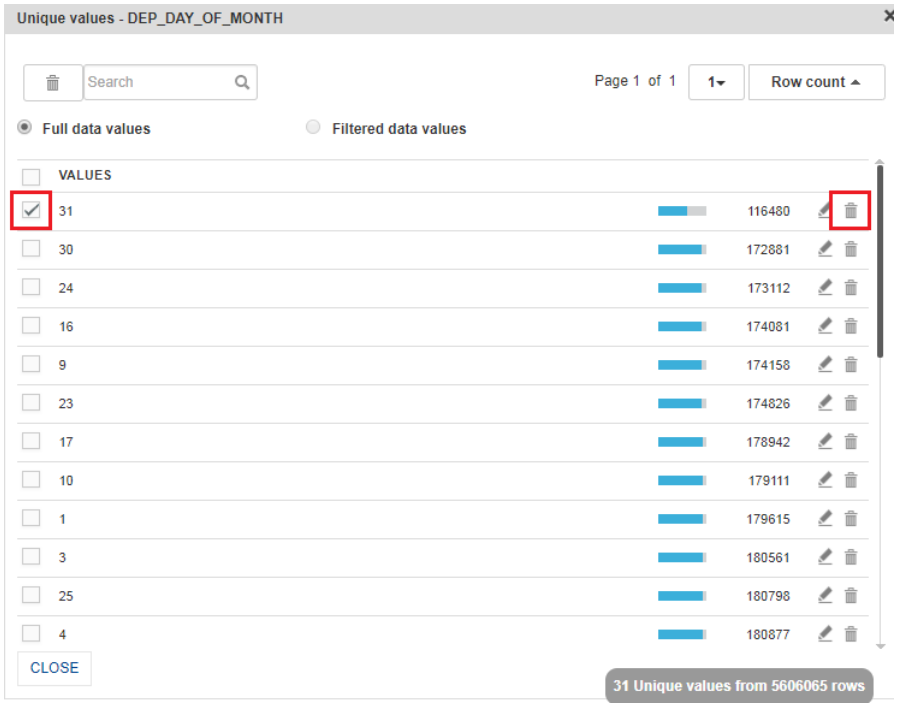
UNIQUE VALUES—EDITING A UNIQUE VALUE

9. Modify the unique value, and then click the Save icon.



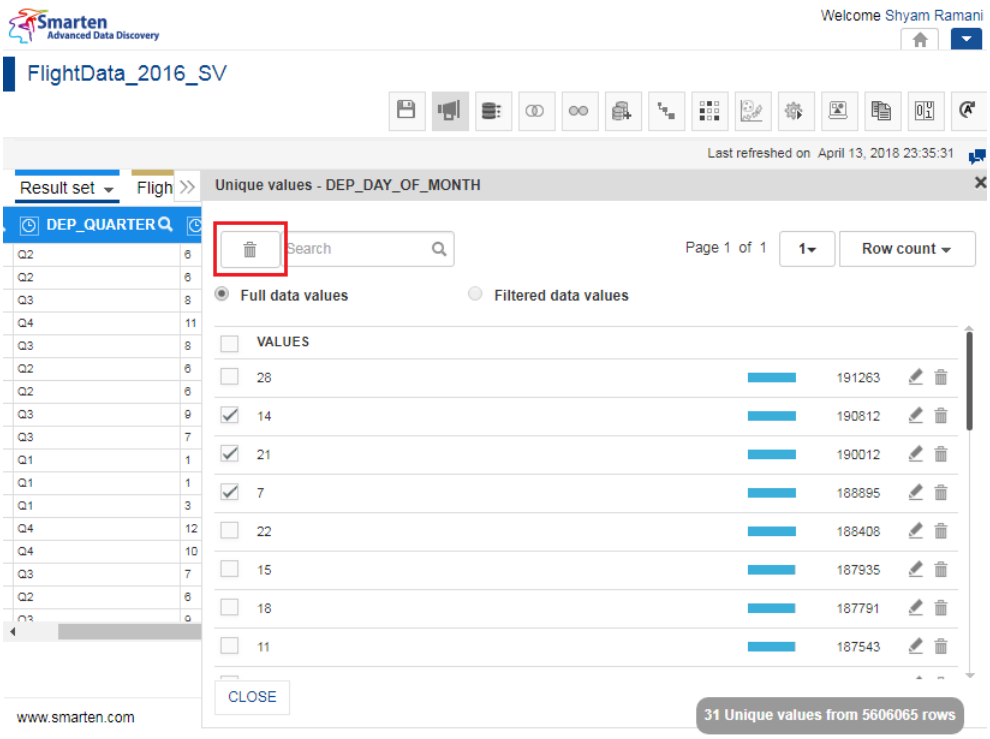
UNIQUE VALUES—EDITING A UNIQUE VALUE

10. You can click the Delete icon adjacent to a unique value to delete that value.



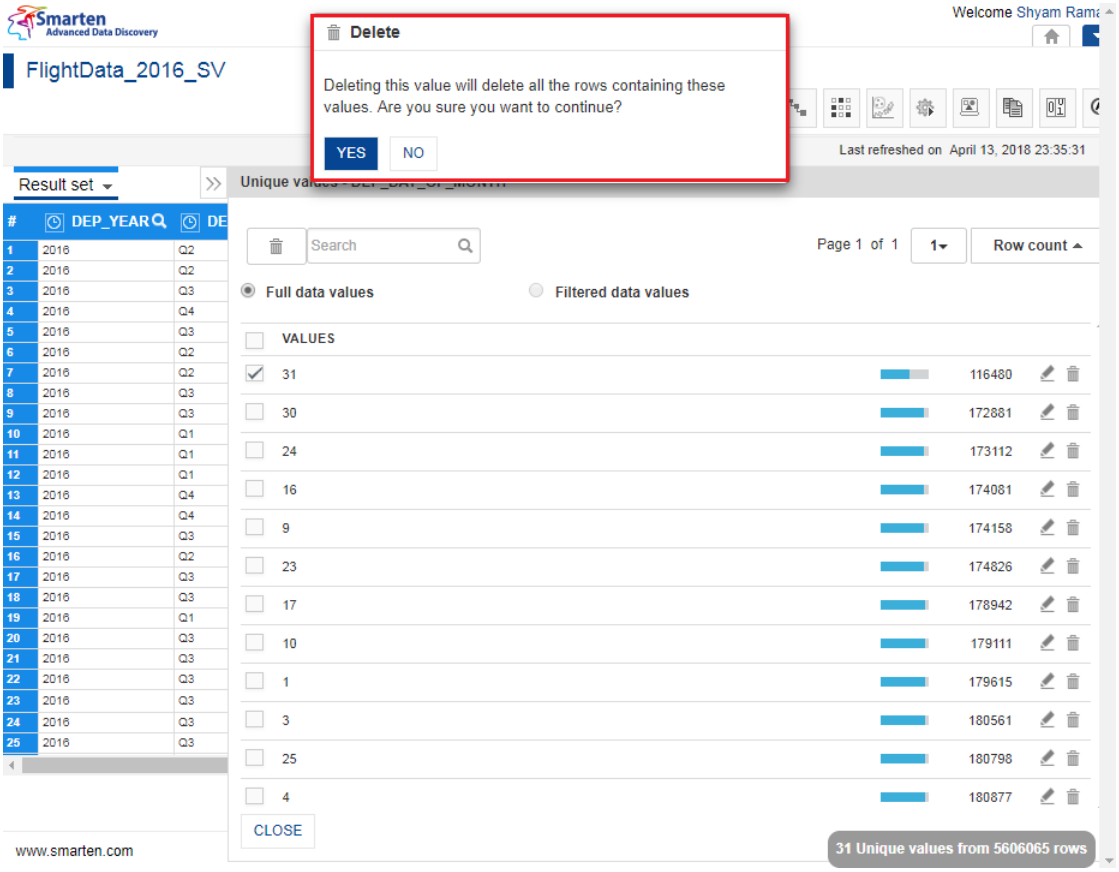
UNIQUE VALUES—DELETING A UNIQUE VALUE

You can also select multiple values you want to delete, and then click the Delete icon next to the Search box.



UNIQUE VALUES—DELETING MULTIPLE UNIQUE VALUES

The system deletes the unique value after confirmation.



UNIQUE VALUES—DELETING UNIQUE VALUE CONFIRMATION

11. Click **CLOSE**.

7.2.7 Clustering and Editing Data in a Dataset

You can use the Cluster and Edit function to edit values in bulk. This function creates groups of similar data and displays them in a cluster that allows you to edit them in bunches.

> Note:
> This option is only available for the columns with the string data type.

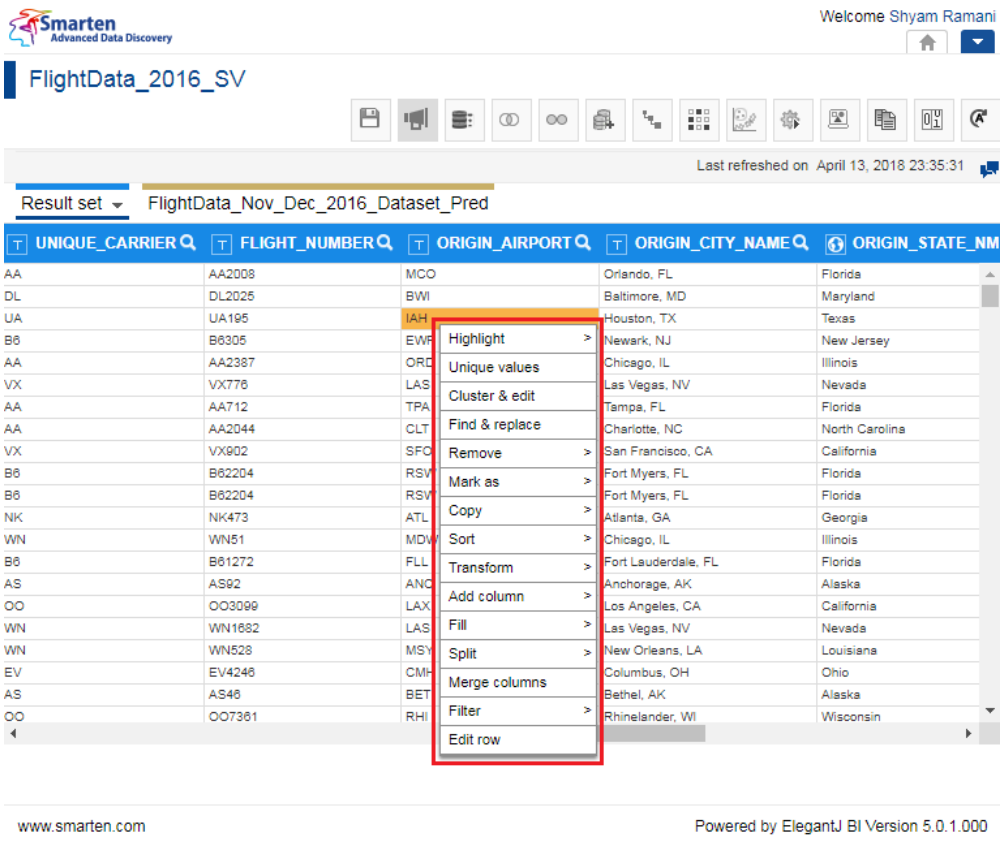
Reference: **Concept Manual > Clean Data > Cluster & Edit**

About this task

Use this task to cluster and edit data in bulk.

Procedure

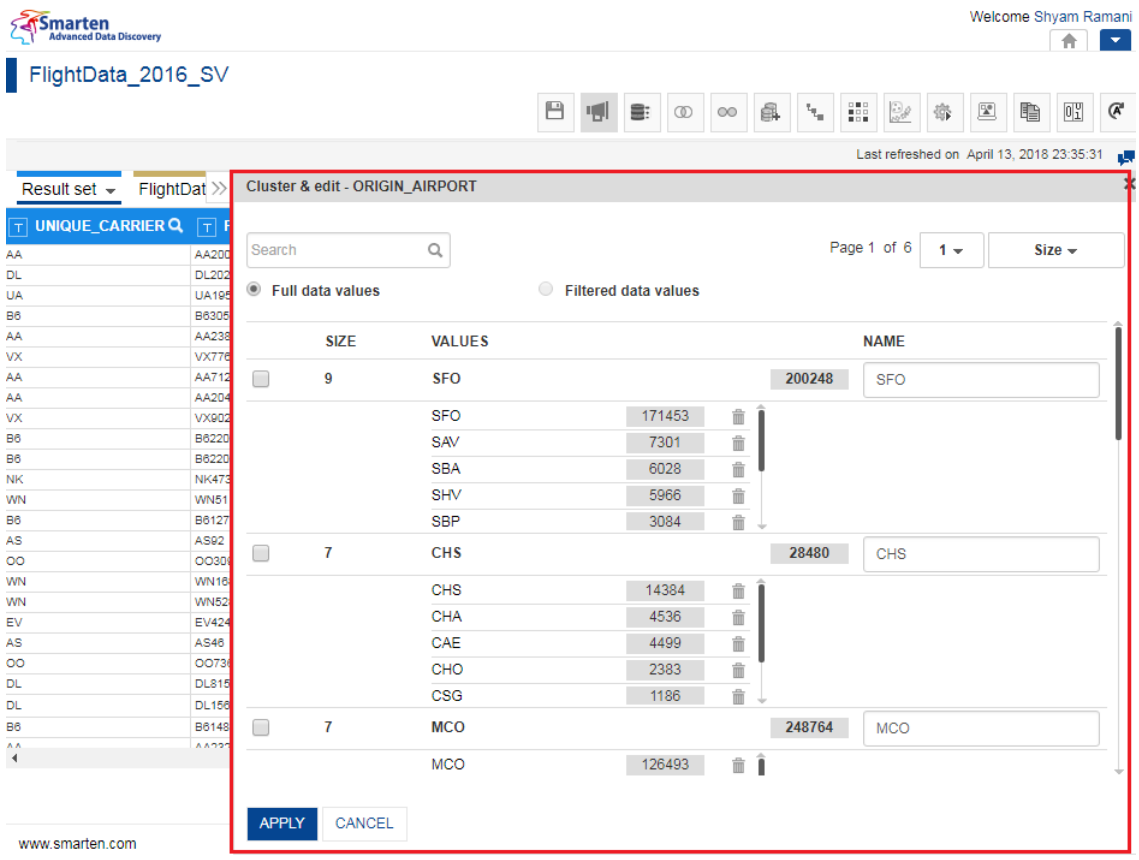
1. Open the dataset for which you want to edit data.
2. Right-click in the column from which you want to edit data.
   The system displays the context menu.



CLUSTER AND EDIT VALUES—THE CONTEXT MENU

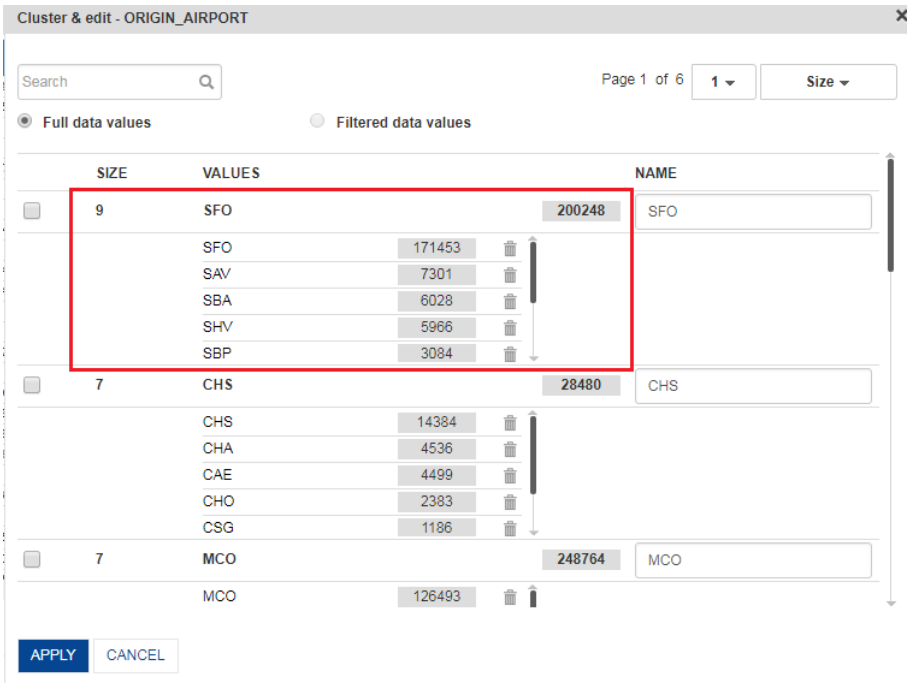
3. Click **Cluster & edit** from the menu.

   The system displays the **Cluster & edit** dialog box. The system creates clusters of values that are similar types of words.



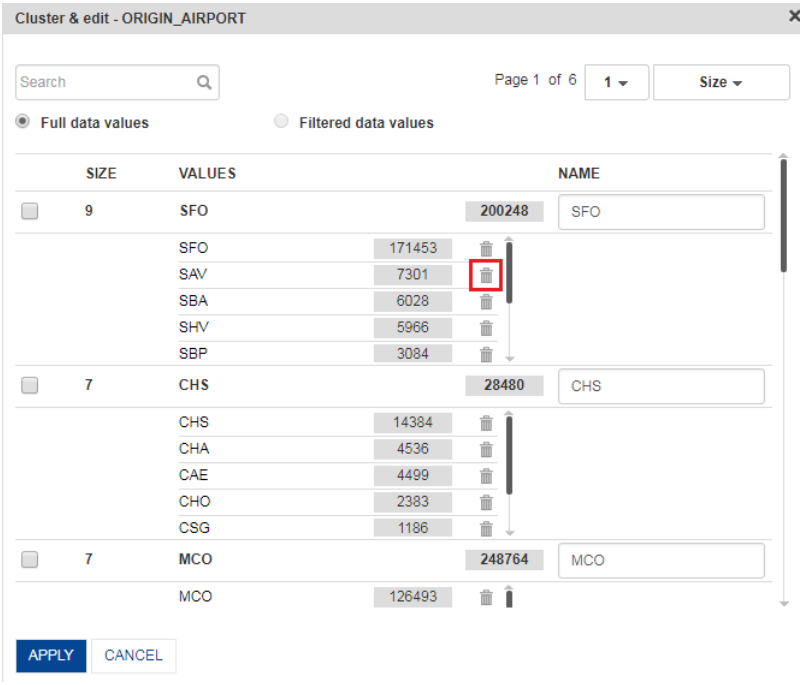
CLUSTER AND EDIT VALUES—CLUSTER AND EDIT DIALOG BOX

The system displays the size of a cluster, which indicates the number of values available in that cluster, the number of rows in which a particular value appears, and the total number of rows for a particular cluster. For example, in the image below, the highlighted region is a cluster. The cluster has 9 values, which indicates there are 9 unique values in that cluster and the total number of rows is 200248. The SFO value in the cluster appears in 171453 rows.



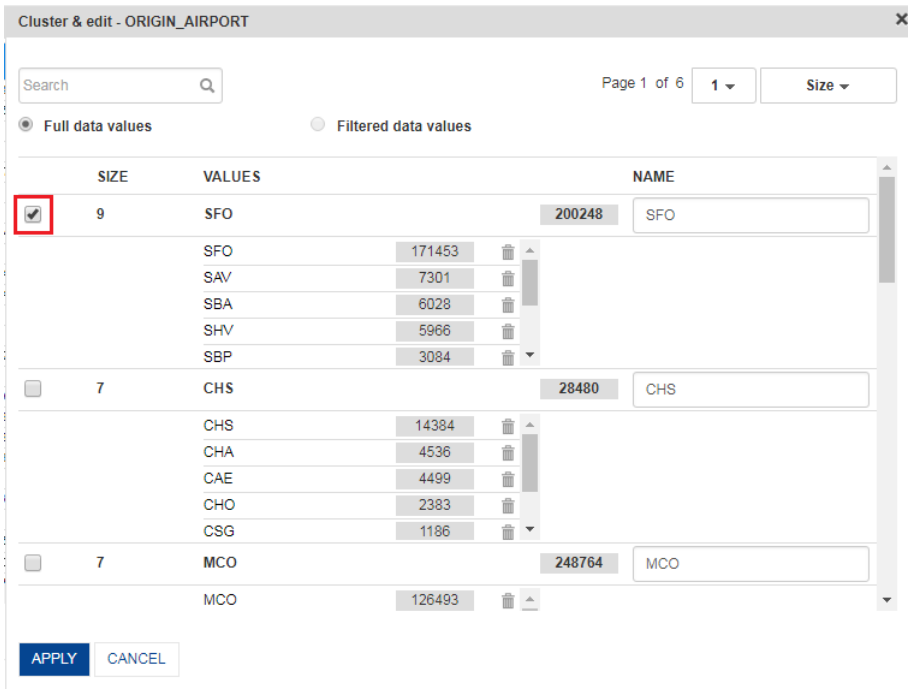
CLUSTER AND EDIT—INFORMATION AVAILABLE IN A CLUSTER

4. You can click the Delete icon adjacent to a value to exclude that value from the cluster. The system excludes the value after confirmation.



CLUSTER AND EDIT—EXCLUDE A VALUE FROM A CLUSTER

5. Select the cluster that contains the value you want to edit, and then specify a new name in the **Name** box.



CLUSTER AND EDIT—SELECTING A CLUSTER

6. Provide a new value in the box adjacent to the selected cluster, and then click **APPLY**. The system updates all values within the selected cluster with the value provided. For example, in the image above, the values in the selected cluster, i.e., SFO, SAV, SBA, SHV, SBP, and so on will be replaced with the value SAR.

7.2.8 Finding and Replacing a Value

You use the find and replace function to find a particular value in a column and replace it with another value.

> Note:
> This option is only available for the columns with the string and numeric data type.

Reference:	**Concept Manual > Clean Data > Find & Replace**

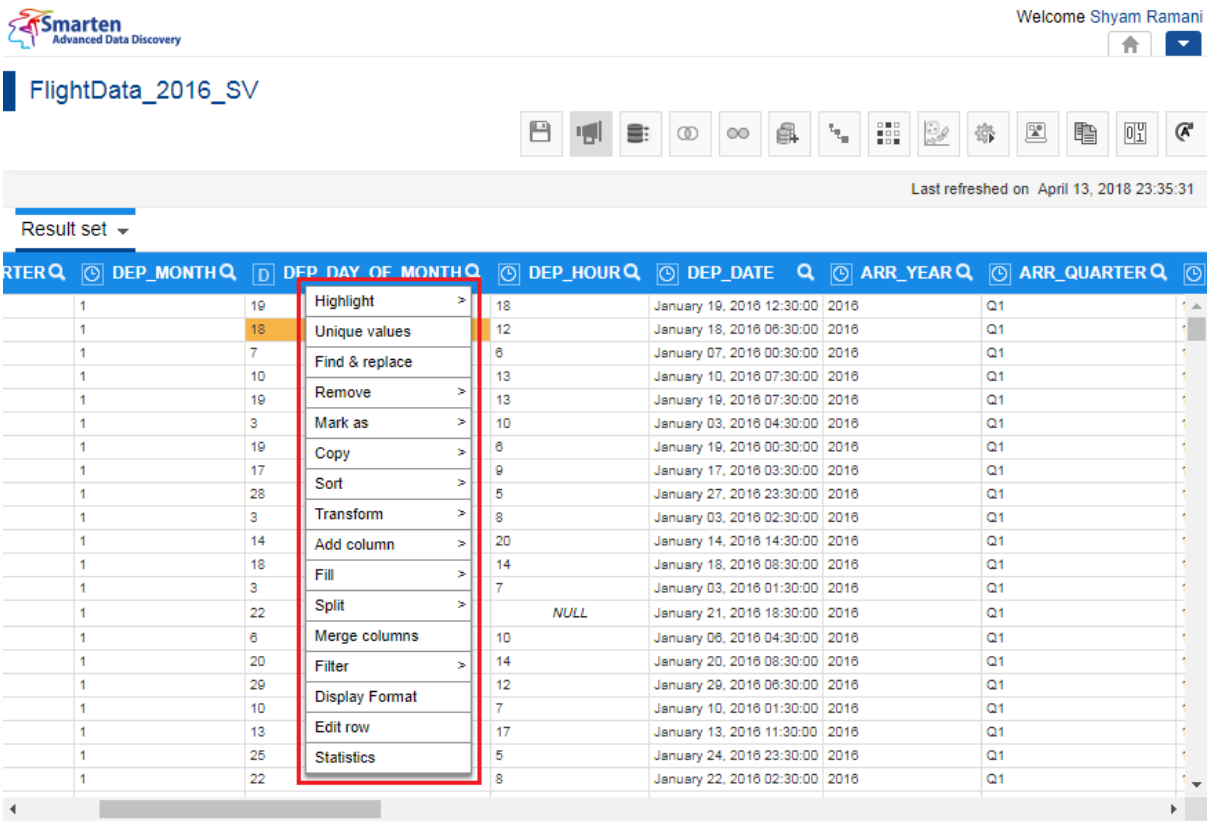
About this task

Use this task to find and replace a value within a column.

Procedure

1. Open the dataset in which you want to find and replace a value.
2. Right-click the column from which you want to find and replace a value.

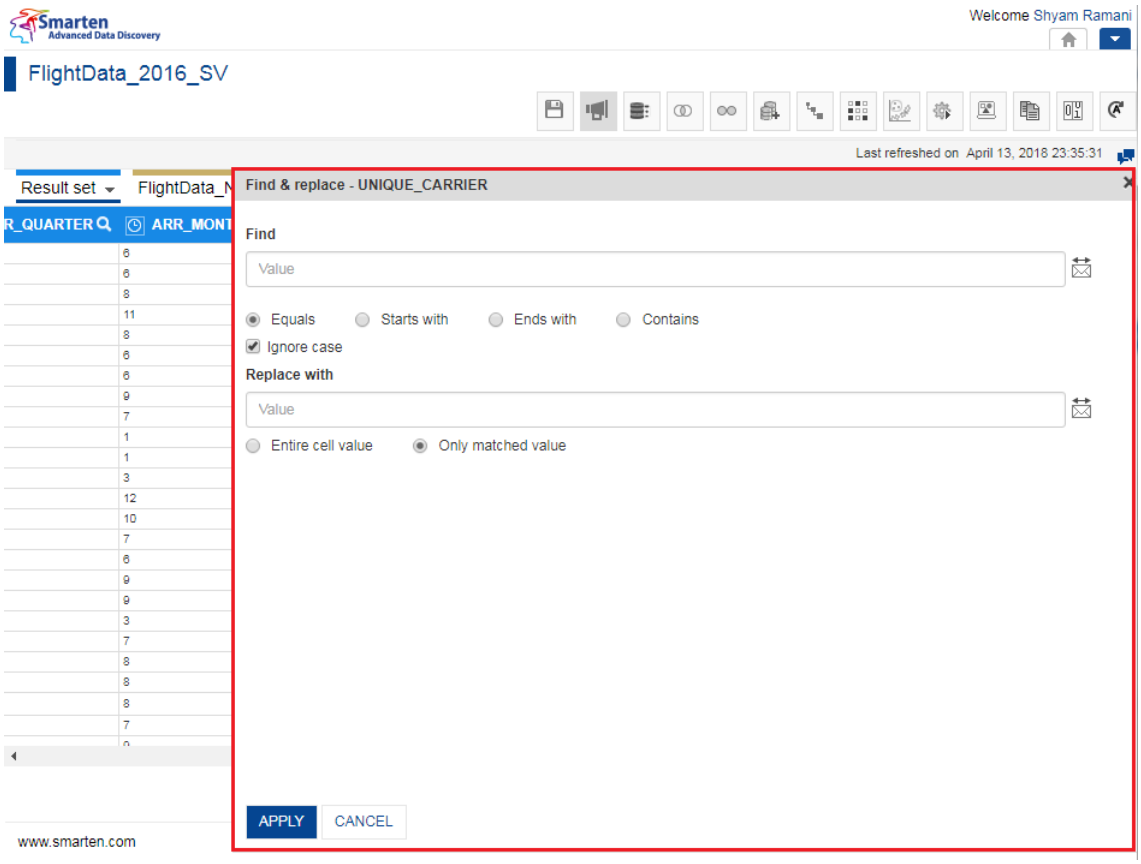
   The system displays the context menu.



FIND AND REPLACE A VALUE—THE CONTEXT MENU

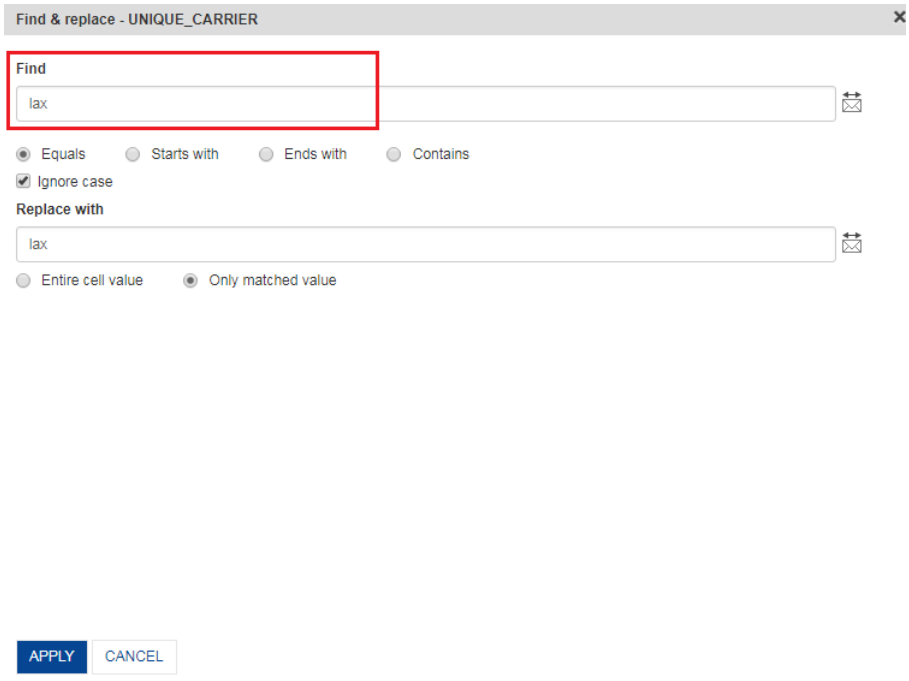
3. Click **Find & replace** from the menu.

   The system displays the **Find & replace** dialog box.



FIND AND REPLACE A VALUE—FIND AND REPLACE DIALOG BOX

4. Enter the value you want to find in the Find box.

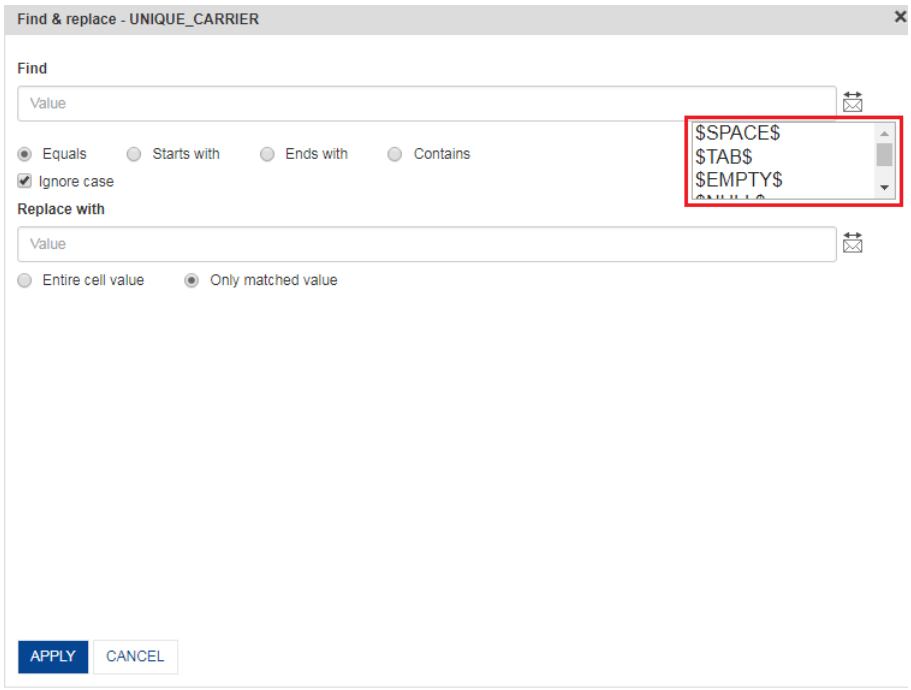


FIND AND REPLACE—PROVIDING THE VALUE TO BE FOUND

You can also click the Envelope icon to select a value from the list.

The following options are available:

- **Space**: Select this option if you want to find values that contain space.
- **Tab**: Select this option if you want to find values that contain tab.
- **Empty**: Select this option if you want to find empty values.
- **Null**: Select this option if you want to find null values.
- **Other**: Select this option if you want to find values other than space, tab, empty, and null.

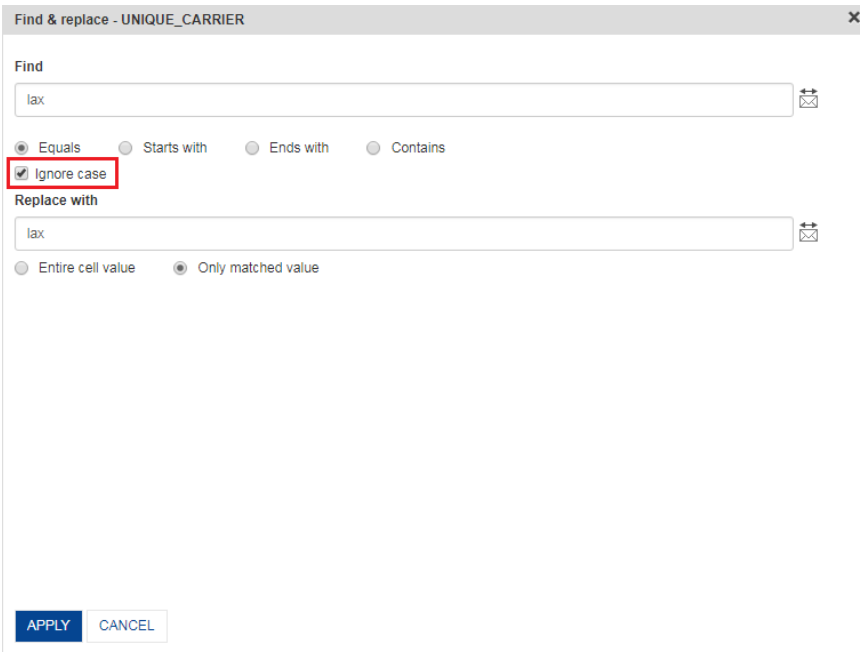


FIND AND REPLACE—SELECTING A VALUE FROM THE LIST

5. Select any of the following options:
    - **Equals**: Select this option to find values that are exactly the same as the value provided in the **Find** box.
    - **Starts with**: Select this option to find values that start with the value provided in the **Find** box.
    - **Ends with**: Select this option to find values that end with the value provided in the **Find** box.
    - **Contains**: Select this option to find values that contain the value provided in the **Find** box.

    > **Note**:
    > These options are only available for the string data type columns.

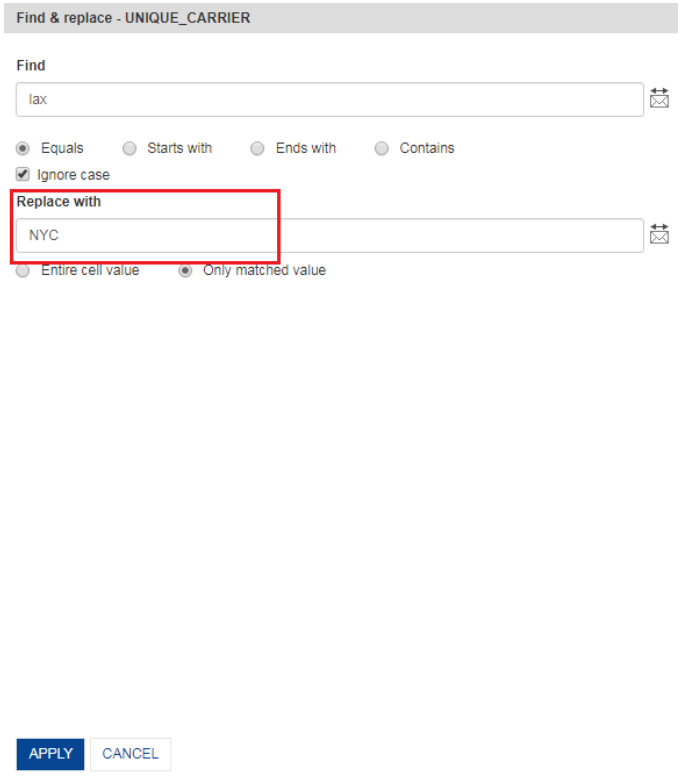
6. Select the **Ignore case** option to ignore the case of the value provided in the **Find** box. For example, while finding “ABCD,” data containing “abcd” or “ABcd” is also considered.



FIND AND REPLACE A VALUE—THE IGNORE CASE OPTION

If you do not select the option, the system finds values that contain the same case of the value provided in the **Find** box. For example, while finding “ABCD,” data containing “ABCD” will only be considered, and “abcd” or “ABcd” will not be considered.

7. Enter the value you want to replace with the found values in the **Replace with** box.



FIND AND REPLACE A VALUE—SPECIFYING THE REPLACEMENT VALUE

8. Select any of the following values:
    - **Entire cell value**: Select this option to replace the entire value of the cell that contains find value with the new value. For example, “N002AA” will be replaced with “99” when finding values containing “00” and replacing them with ”99.”
    - **Only matched value**: Select this option to replace only the matched value with the new value. For example, “N002AA” will be replaced with “N992AA” when finding values containing “00” and replacing them with ”99.”

9. Click **APPLY**.

7.2.9 Removing Data from a Dataset

You can remove unnecessary or duplicate rows, columns, and rows having null or zero values or having a specific value in a column.

Reference: **Concept Manual > Clean Data > Remove**

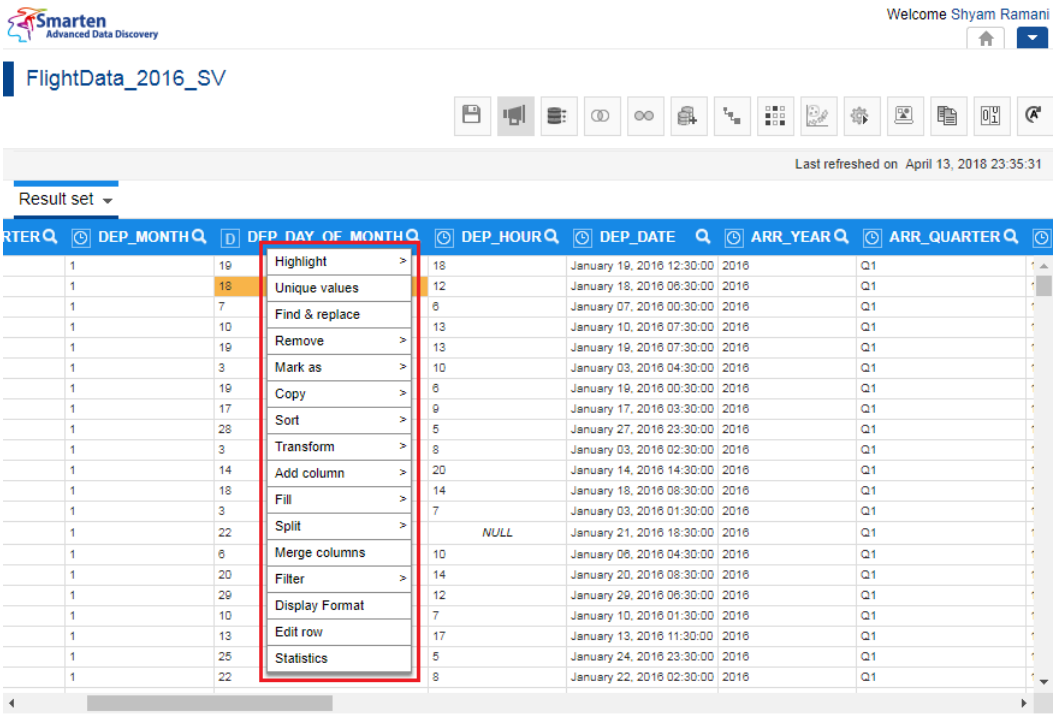
About this task

Use this task to remove certain data in a dataset.

Procedure

1. Open the dataset from which you want to remove data.
2. Right-click in the column from which you want to remove data.

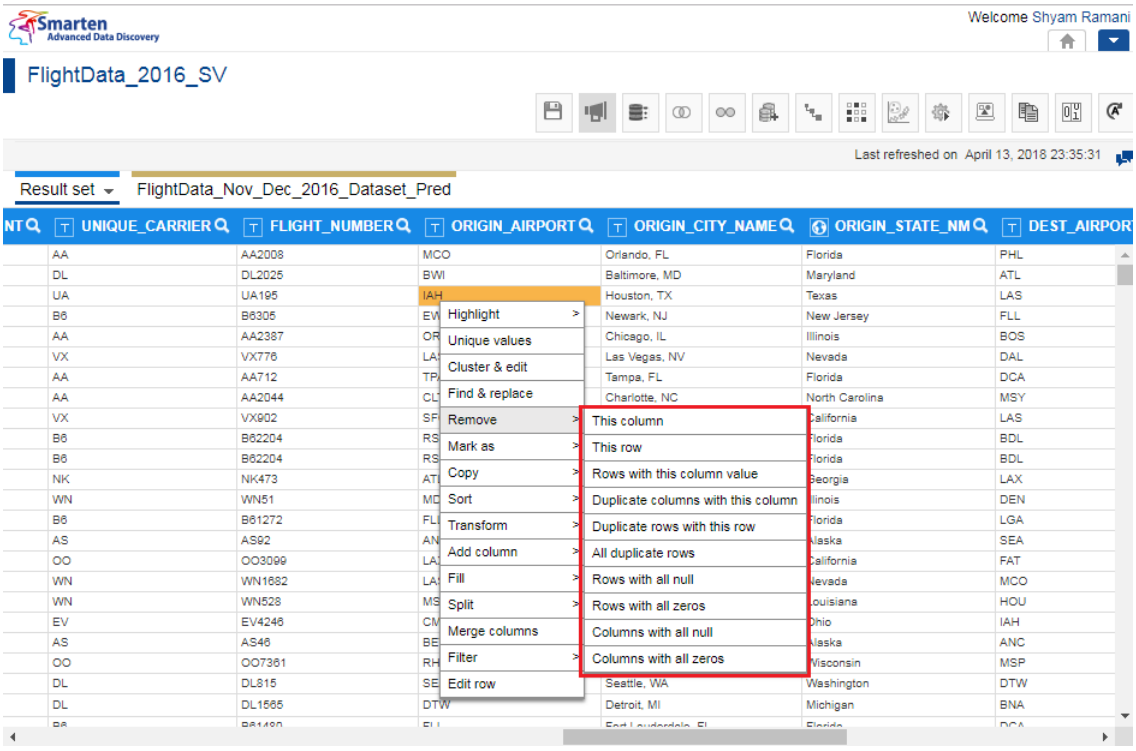
   The system displays the context menu.



REMOVE VALUES—THE CONTEXT MENU

3. Click **Remove** from the menu.

The system displays the options available to remove data from the dataset.



REMOVE VALUES—OPTIONS AVAILABLE FOR THE REMOVE MENU

4. Click any of the options in the following table to apply that operation:

Option	Description
This column	Use this option to remove the selected column.
This row	Use this option to remove the selected row.
Rows with this column value	Use this option to remove all the rows that contain selected value in a particular column.
Duplicate columns with this column	Use this option to remove all the columns that contain exactly the same data as the selected column.
Duplicate row with this row	Use this option to remove all the rows that contain exactly the same data as the selected row.
All duplicate rows	Use this option to remove a set of rows that contain exactly the same data.
Rows with all null	Use this option to remove all the rows that have null in all the columns.
Rows with all zeros	Use this option to remove all the rows that have zeros in all numeric columns.
Columns with all null	Use this option to remove all the columns that have null in all rows.
Columns with all zeros	Use this option to remove all the columns that have zeros in all rows.

7.2.10 Marking Data in a Dataset

You can mark a column as a GeoMap dimension, time dimension, or dimension column depending on the data of that column.

Marking columns as a GeoMap dimension makes them available to be used in GeoMap objects. Similarly, columns marked as time dimension makes them available to be used in objects where they are needed for time-based analysis. Marking a column as dimension allows them to be used as dimension columns.

> Note:
> This function is only applicable for string and numeric data.

Reference: **Concept Manual > Shape Data > Mark as**

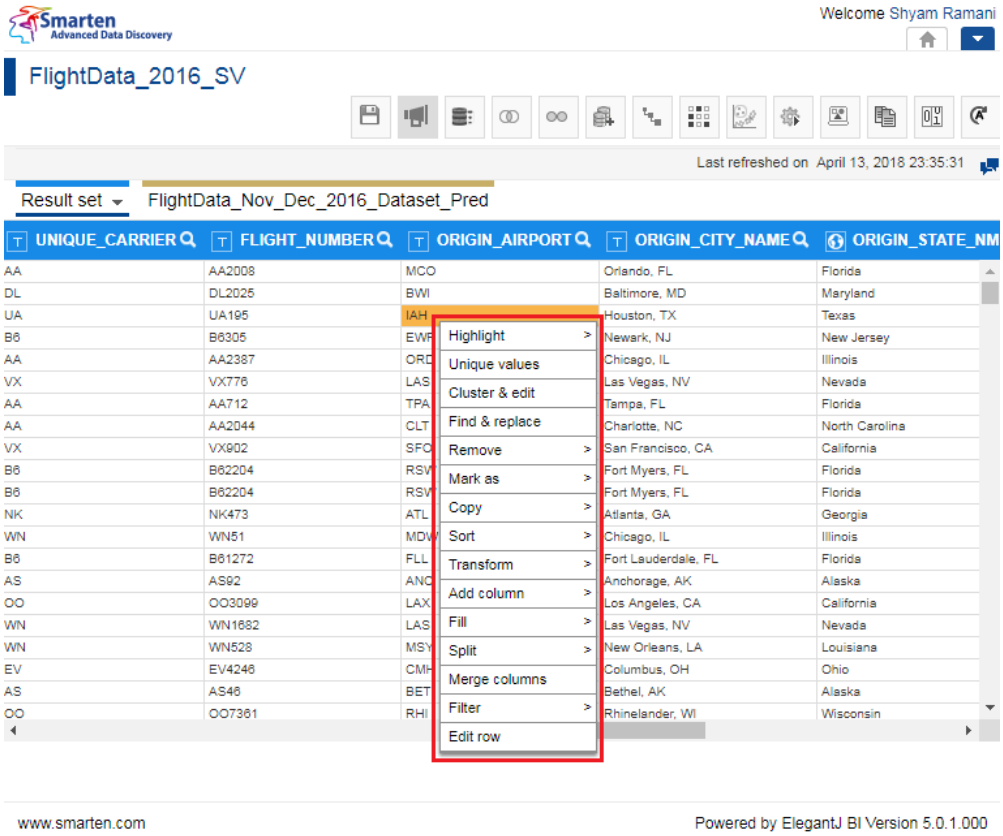
About this task

Use this task to mark data in a column.

Procedure

1. Open the dataset for which you want to mark data.
2. Right-click in the column from which you want to mark data.

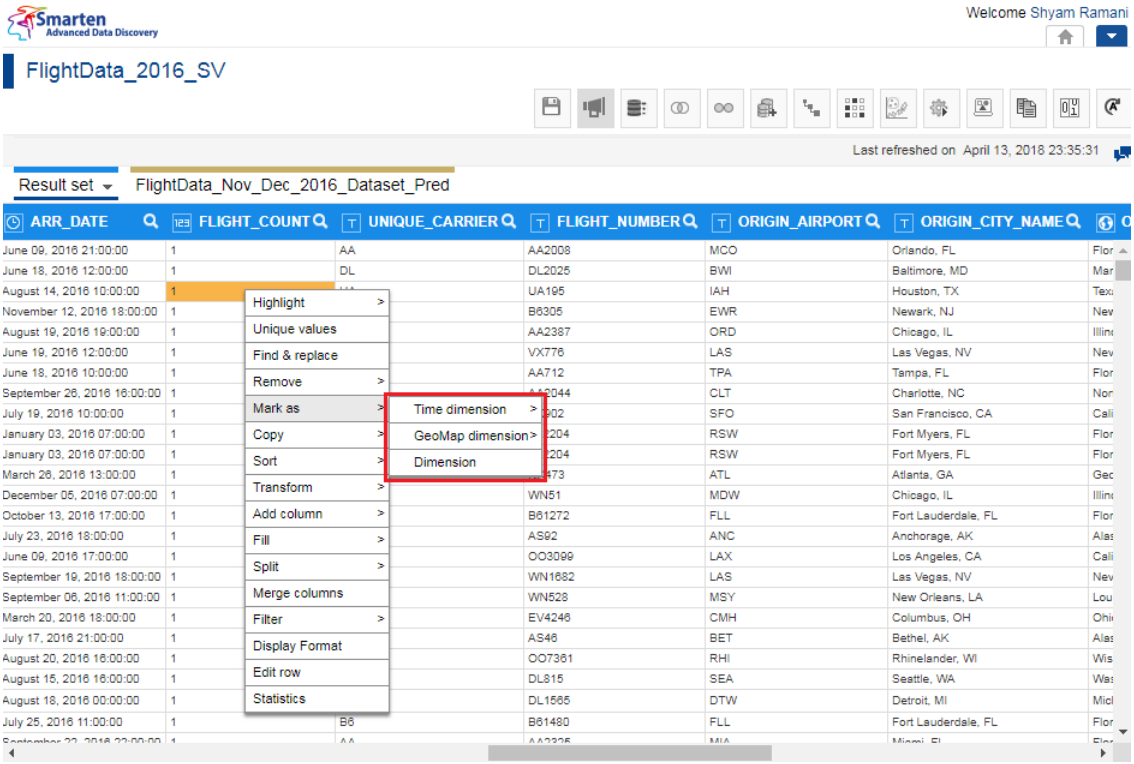
   The system displays the context menu.



MARK DATA—THE CONTEXT MENU

3. Click **Mark as** from the menu.

The system displays the options available to mark data from the dataset.



MARK DATA—OPTIONS AVAILABLE FOR THE MARK MENU

4. Click any of the options in the following table to apply that operation.

Option	Suboption	Description
Time dimension Note: This option is available for the numeric data type.	Year	Use this option to mark the selected column as time dimension for the interval of a year.
	Quarter	Use this option to mark the selected column as time dimension for the interval of a quarter.
	Month	Use this option to mark the selected column as time dimension for the interval of a month.
	Week	Use this option to mark the selected column as time dimension for the interval of a week.
	Day	Use this option to mark the selected column as time dimension for the interval of a day.

	Hour	Use this option to mark the selected column as time dimension for the interval of an hour.
	Minute	Use this option to mark the selected column as time dimension for the interval of a minute.
	Second	Use this option to mark the selected column as time dimension for the interval of a second.
GeoMap dimension Note: The ZIP code, latitude, and longitude options are available for numeric data type, and the country, county, state, city, and area options are available for the string data type.	ZIP code	Use this option to mark the selected column as GeoMap dimension of a ZIP code.
	Latitude	Use this option to mark the selected column as GeoMap dimension of latitude.
	Longitude	Use this option to mark the selected column as GeoMap dimension of longitude.
	Country	Use this option to mark the selected column as GeoMap dimension of a country.
	County	Use this option to mark the selected column as GeoMap dimension of a county.
	State	Use this option to mark the selected column as GeoMap dimension of state.
	City	Use this option to mark the selected column as GeoMap dimension of a city.
	Area	Use this option to mark the selected column as GeoMap dimension of the area of a city.
Dimension Note: This option is available for the numeric data type.		Use this option to mark the selected column as a dimension column.

Similar to marking a column, you can use the above steps to unmark a marked column.

7.2.11 Copying Data in a Dataset

You can copy a particular column or a row and add a duplicate of the copied row or column in the dataset.

Reference: **Concept Manual > Shape Data > Copy**

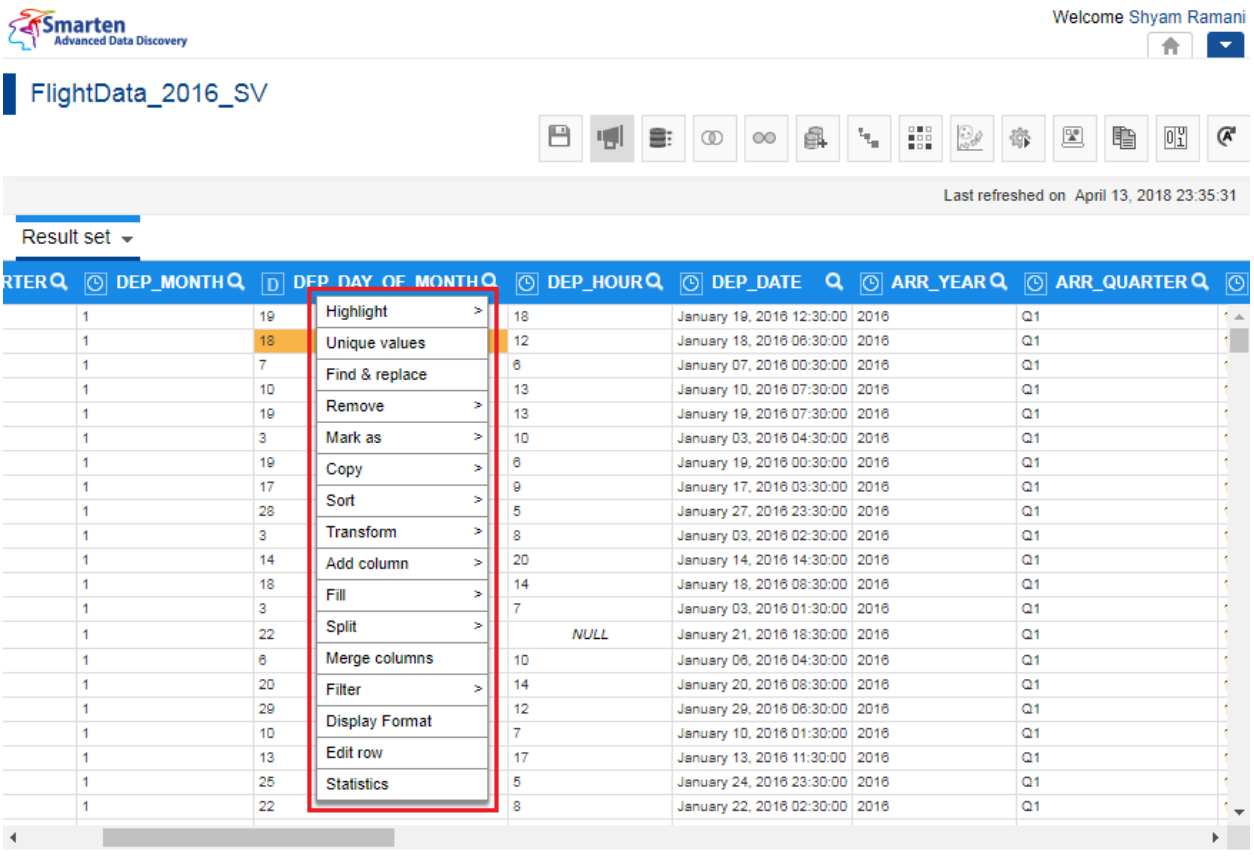
About this task

Use this task to copy data in a dataset.

Procedure

1. Open the dataset in which you want to copy data.
2. Right-click in the column or row you want to copy.

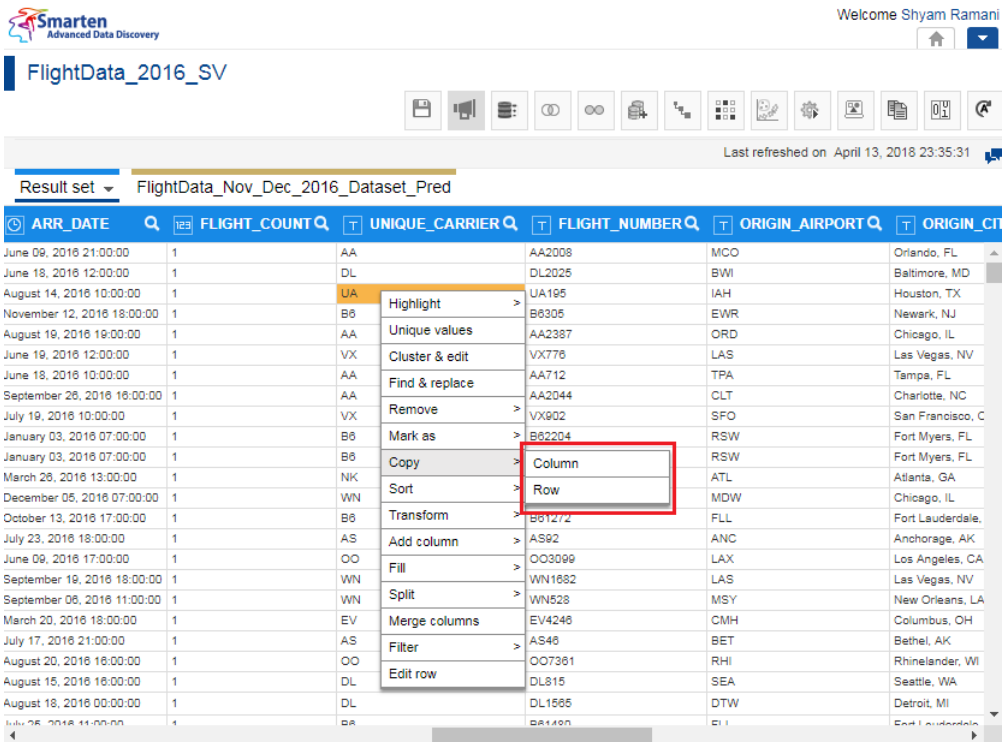
   The system displays the context menu.



COPY DATA—THE CONTEXT MENU

3. Click **Copy** from the menu.

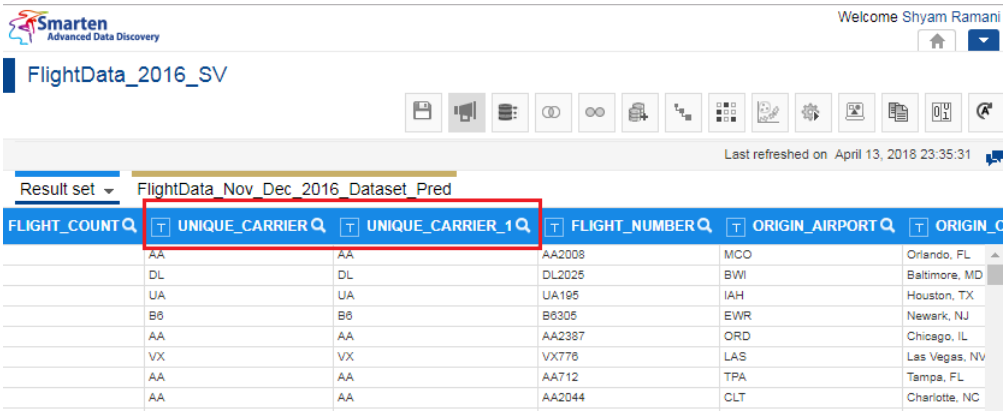
   The system displays the options available to mark data from the dataset.



COPY DATA—OPTIONS AVAILABLE FOR COPYING DATA

4. You can select **Column** if you want to copy the selected column.

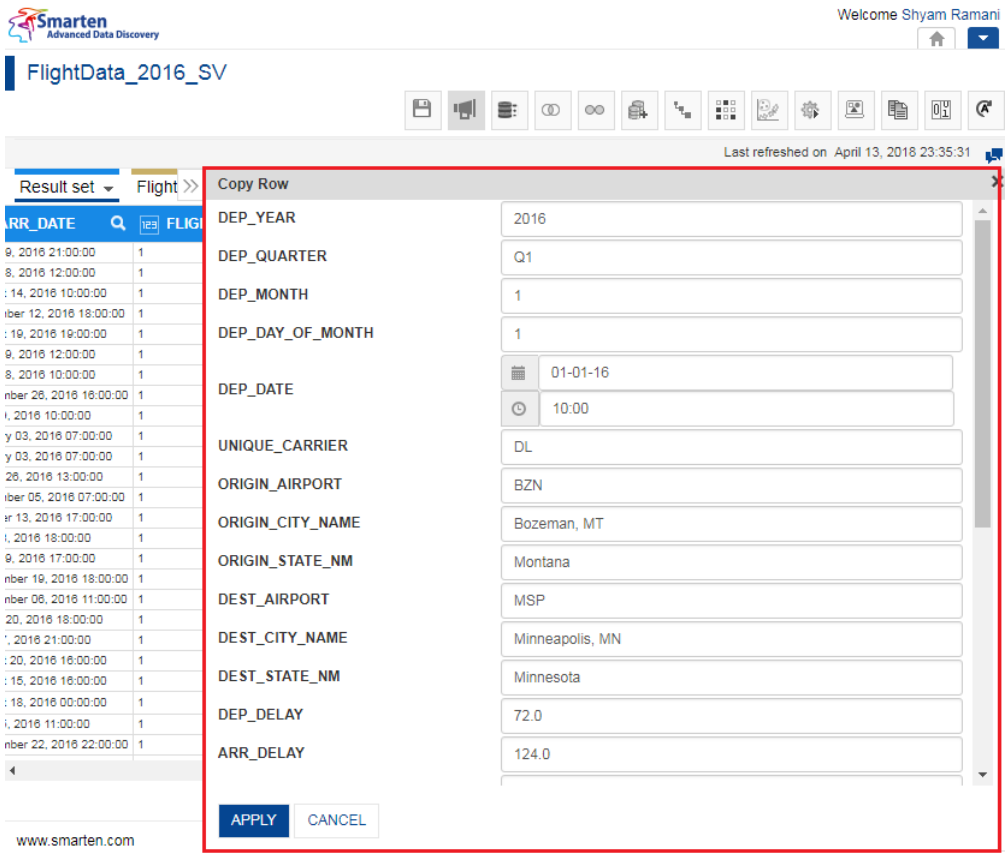
   The system creates a copy of the selected column next to the selected column. The name of the new column is the name of the copied column appended with the numeric value, which increases each time a copy of the column is added. For example, in the image above, we have selected the UNIQUE_CARRIER column. The system creates a copy of the column with the name UNIQUE_CARRIER_1 as shown in the image below.



FIND AND REPLACE—PROVIDING THE VALUE TO BE FOUND

5. You can select **Row** if you want to copy data of the selected row.

   The system displays the **Copy Row** dialog box that allows you to copy the selected row. The user can modify value of any columns in that row, and a new row is inserted into the dataset.



COPY DATA—THE COPY ROW DIALOG BOX

6. Click **APPLY**.

   The system creates a new row with the updated values.

7.2.12 Sorting Data in a Dataset

You can sort data in a column within a dataset in ascending or descending order.

Reference:	**Concept Manual > Shape Data > Sort**

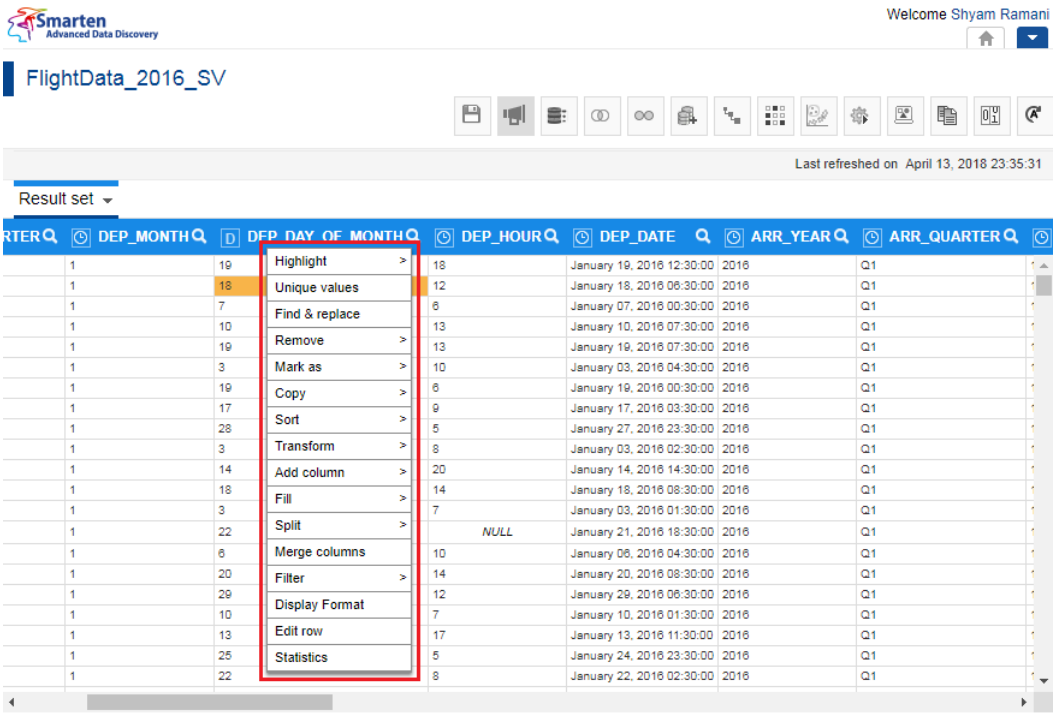
About this task

Use this task to sort data in a particular column.

Procedure

1. Open the dataset in which you want to sort data.
2. Right-click in the column whose data you want to sort.

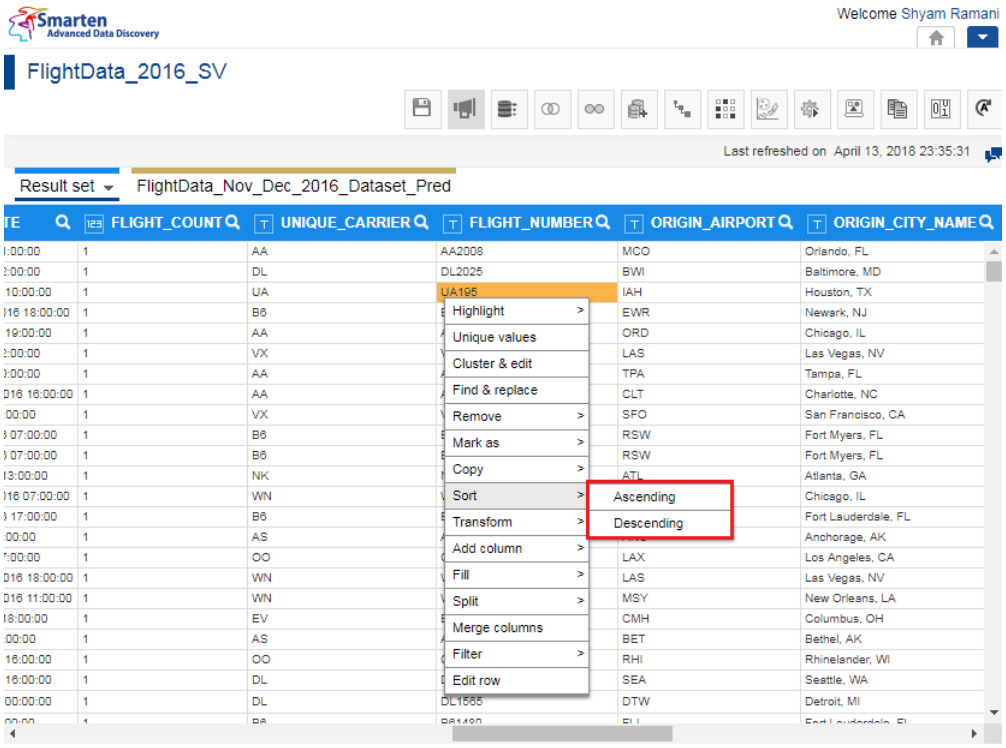
   The system displays the context menu.



SORT DATA—THE CONTEXT MENU

3. Click **Sort** from the menu.

   The system displays the options available to sort data from the dataset.



SORT DATA—OPTIONS AVAILABLE FOR THE SORT MENU

4. Click **Ascending** to sort data in the selected column in ascending order.
   Or,
   Click **Descending** to sort data in the selected column in descending order.

7.2.13 Transforming Data in a Dataset

You can change the data type and format of data within a particular column.

> Note:
> The options available to transform data depend on the data type of the selected column.

Reference:	**Concept Manual > Transform Data**

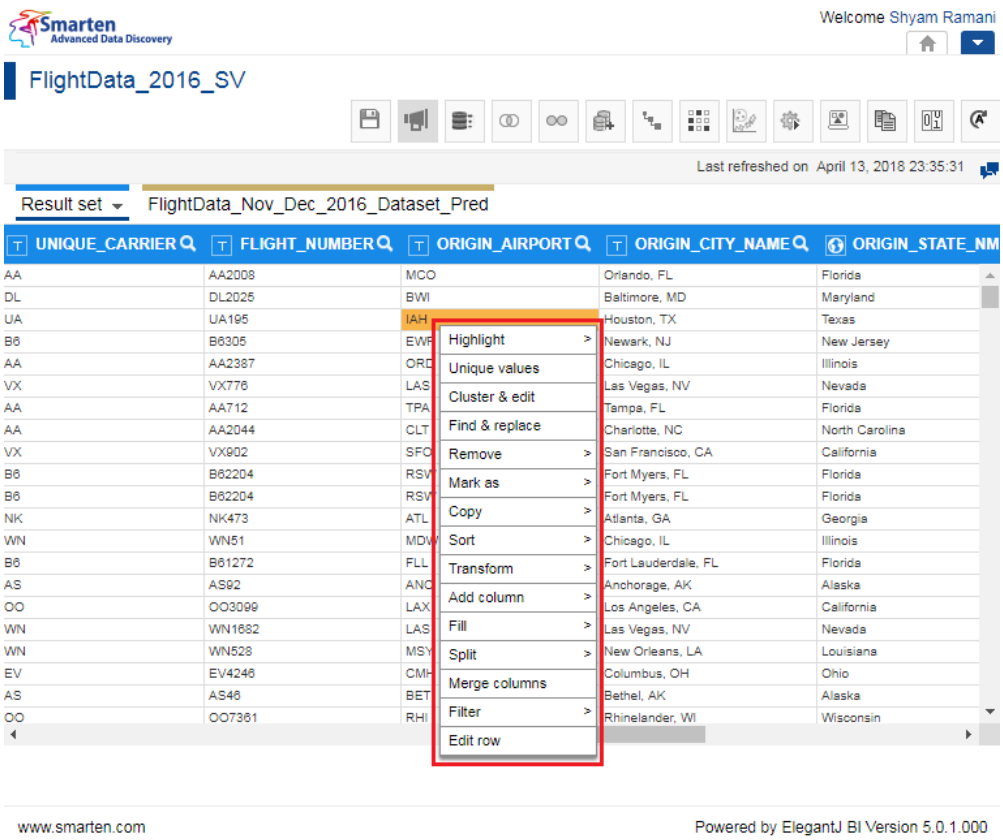
About this task

Use this task to transform data in a column.

Procedure

1. Open the dataset for which you want to transform data.
2. Right-click in the column that you want to transform data.

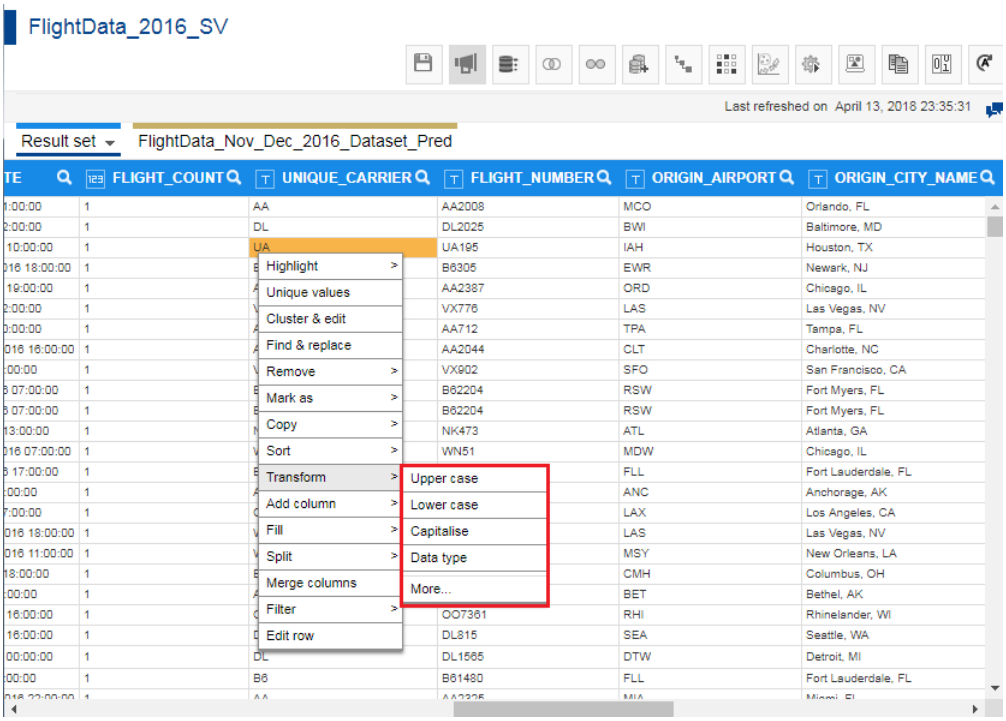
   The system displays the context menu.



TRANSFORM DATA—THE CONTEXT MENU

3. Click **Transform** from the menu.

   The system displays the options available to transform data based on the data type of the selected column.



TRANSFORM DATA—OPTIONS AVAILABLE FOR THE TRANSFORM MENU

4. Click any of the options in the following table to apply that operation:

Data Type	Option	Description
String	Uppercase	Use this option to convert the characters in lowercase to uppercase in the selected column.
	Lowercase	Use this option to convert the characters in uppercase to lowercase in the selected column.
	Capitalize	Use this option to capitalize the data of the selected column.
	Data Type	Use this option to change the data type of the data within the selected column.
	More	Use this option to access the advanced options. The advanced options available depend on the data type of the data within the selected column.
Numeric	Data Type	Use this option to change the data type of the data within the selected column.
	More	Use this option to access the advanced options. The advanced options available depend on the data type of the data within the selected column.
Datetime	Make timepart zero	Use this option to convert the timepart of the datetime type of the selected column to zero.
	First date of month	Use this option to convert all dates of the selected column to the first date of their

		respective month.
	Last date of month	Use this option to convert all dates of the selected column to the last date of their respective month.
	Data type	Use this option to change the data type of the data within the selected column.
	More	Use this option to access the advanced options. The advanced options available depend on the data type of the data within the selected column.

7.2.13.1 Changing Data Type of Data in a Dataset

You can change the data type of data in a particular column.

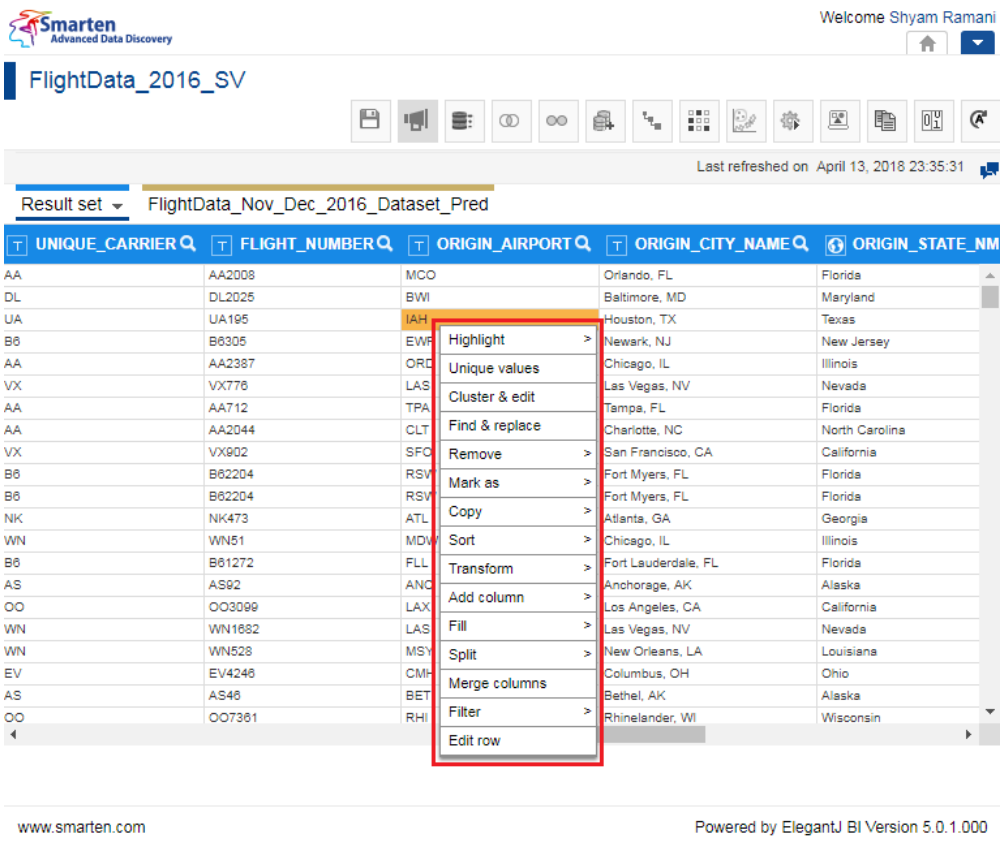
About this task

Use this task to change the data type of a column.

Procedure

1. Open the dataset for which you want to change the data type.
2. Right-click in the column you want to change the data type.

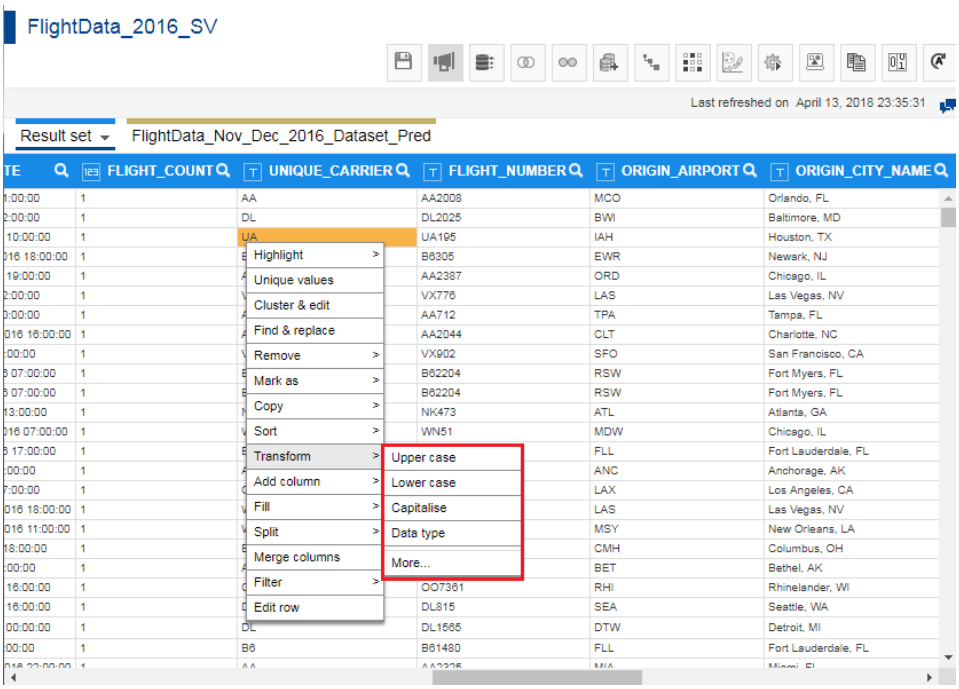
   The system displays the context menu.



TRANSFORM DATA—THE CONTEXT MENU

3. Click **Transform** from the menu.

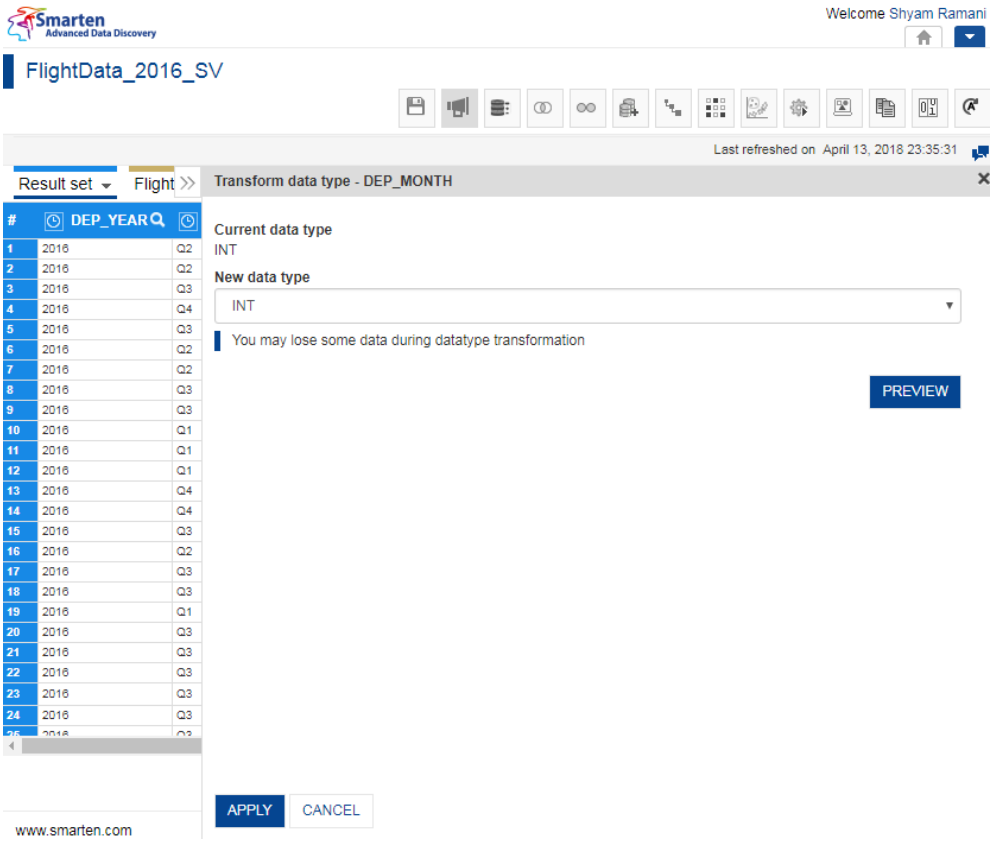
   The system displays the options available to transform data based on the data type of the selected column.



TRANSFORM DATA—OPTIONS AVAILABLE FOR THE TRANSFORM MENU

4. Click **Data type**.

   The system displays the **Transform data type** dialog box.



TRANSFORM DATA—THE TRANSFORM DATA TYPE DIALOG BOX

The system displays the current data type of the selected column.

5. Select an option from the **New data type** list.

   The following options are available based on the data type of the selected column:
   String:
   - STRING
   - DATE
   - TIMESTAMP
   - Double
   - INT
   - BIGINT

   Integer:
   - STRING
   - DOUBLE
   - INT
   - BIGINT

   Datetime:
   - STRING
   - DATE
   - TIMESTAMP

6. You can click **PREVIEW** to view a preview of the transformed data.
7. Click **APPLY**.

7.2.13.2 Advanced Options for Transforming Data in a Dataset

You can use the advanced option to change data in a dataset.

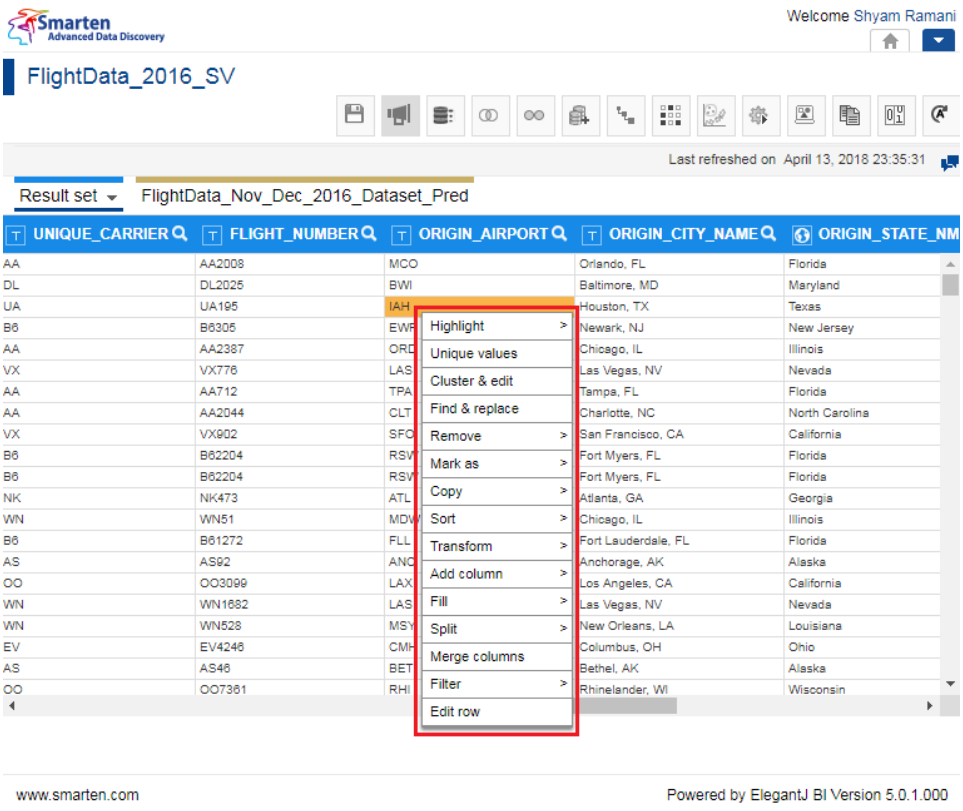
About this task

Use this task to use advanced options to change data in a dataset.

Procedure

1. Open the dataset for which you want to change the data type.
2. Right-click in the column from which you want to change the data type.

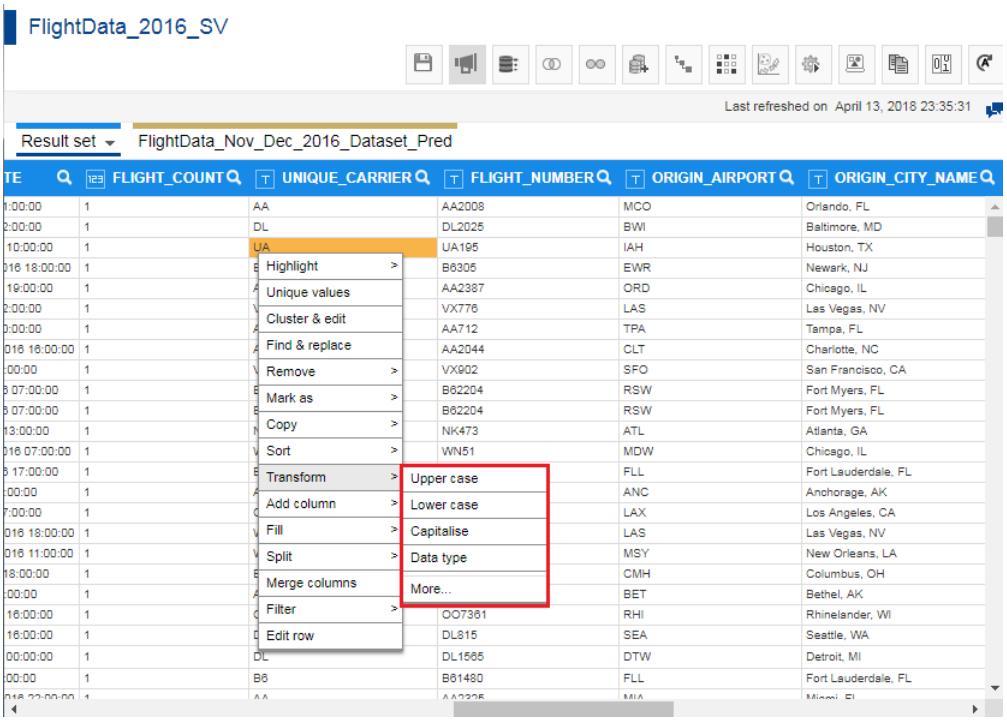
   The system displays the context menu.



TRANSFORM DATA—THE CONTEXT MENU

3. Click **Transform** from the menu.

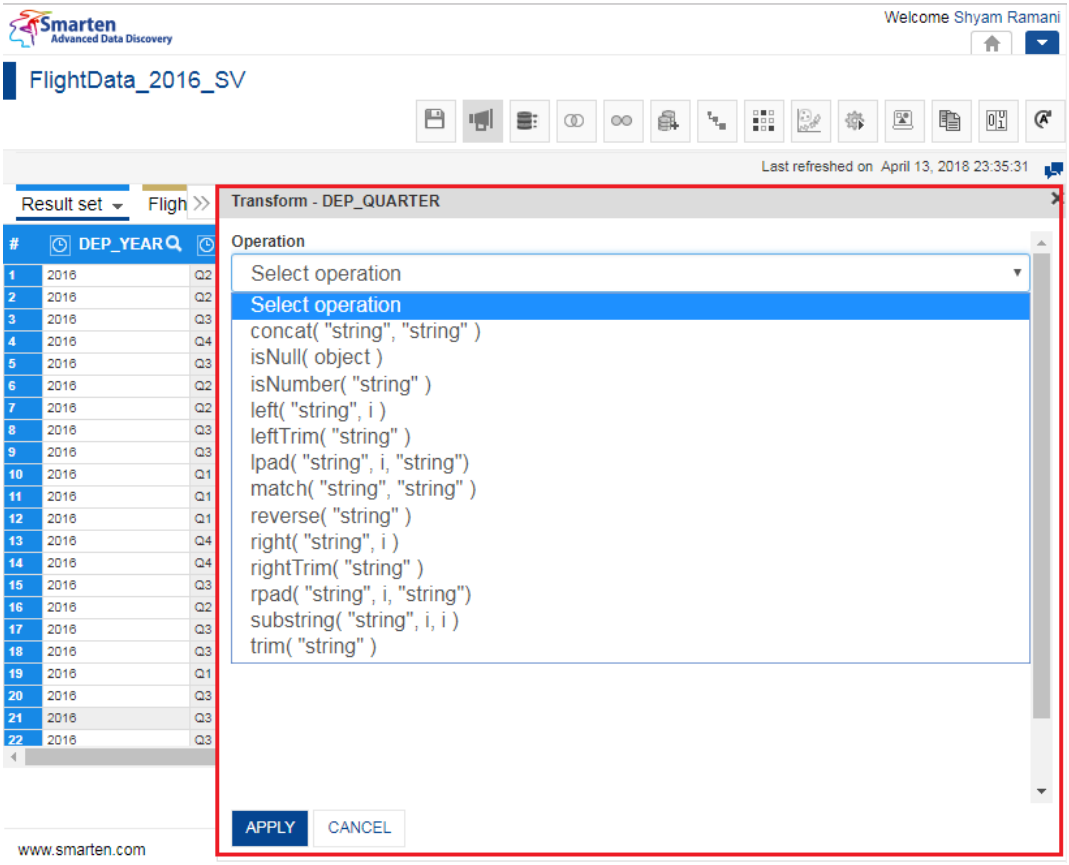
   The system displays the options available to transform data based on the data type of the selected column.



TRANSFORM DATA—OPTIONS AVAILABLE FOR THE TRANSFORM MENU

4. Click **More**.

   The system displays the **Transform** dialog box.



TRANSFORM DATA—THE TRANSFORM DATA ADCVANCED OPTIONS

5. Select an option from the **Operation** list.

The following tables provide information about the advanced options available for transforming data in a dataset:

For string data type:

Operation	Description	Example
concat("string", "string")	Returns a resulting string after concatenating specified strings Argument 1: The text that has to be concatenate with argument 2 Argument 2: The text that has to be concatenate with argument 1 Returns: A string	Argument 1 = "N787" Argument 2 = "AA" Returns "N787AA"
isNull(object)	Determines if the argument is NULL Argument 1: The object that is to be checked Returns: A boolean	Argument 1 = "N787AA" Returns "false" Argument 1 = "NULL" Returns "true"

isNumber("string")	Determines if the specified string contains a number Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = "N787AA" Returns "false"
		Argument 1 = "787" Returns "true"
left("string", i)	Returns a specified number of characters from a string starting with the first character Argument 1: The text from which the partial words are to be returned Argument 2: The number of characters to be extracted from the beginning of the text Returns: A string	Argument 1 = "N787AA" Argument 2 = 2 Returns "N7"
		Argument 1 = "N787AA" Argument 2 = 8 Returns "N787AA"
leftTrim("string")	Returns a copy of a specified string with leading blanks removed Argument 1: The text for which blank spaces are to be removed from left Returns: A string	Argument 1 = "87AA" Returns "87AA"
		Argument 1 = "87AA" Returns "87AA"
		Argument 1 = "87AA" Returns "87AA"
lpad("string", i, "string")	Returns a string with left-pad of specified length Argument 1: The text in which left-pad is to be added Argument 2: The length of string to be returned Argument 3: The text that is to be padded in specified string Returns: A string	Argument 1 = "N787AA" Argument 2 = 10 Argument 3 = "X" Returns "XXXXN787AA"
		Argument 1 = "N787AA" Argument 2 = 3 Argument 3 = "X" Returns "N78"
		Argument 1 = "N787AA" Argument 2 = 6 Argument 3 = "X" Returns "N787AA"
match("string", "string")	Returns a determination of whether or not a string contains a particular pattern of characters Argument 1: The text that has to be searched in argument 2 Argument 2: The text in which argument 1 has to be searched Returns: A number	Argument 1 = "AA" Argument 2 = "N787AA" Returns 1
		Argument 1 = "aa" Argument 2 = "N787AA" Returns 0
		Argument 1 = "AB" Argument 2 = "N787AA" Returns 0
reverse("string")	Reverses the order or characters in a string Argument 1: The text that needs to be reversed Returns: A string	Argument 1 = "N208WN" Returns "NW802N"
right("string", i)	Returns the specified number of characters from the end of a specified string Argument 1: The text from which the	Argument 1 = "N208WN" Argument 2 = 3 Returns "8WN"

	specified number of characters should be returned from the end Argument 2: The number of characters to be returned from the string Returns: A string	Argument 1 = “N208WN” Argument 2 = 8 Returns “N208WN”
rightTrim(“string”)	Returns a copy of the specified string with trailing blanks removed Argument 1: The text from which extra spaces have to be removed from the right Returns: A string	Argument 1 = “N208” Returns “N208”
		Argument 1 = “08WN” Returns “08WN”
		Argument 1 = “208W” Returns “208W”
rpad(“string”, i, “string”)	Returns a string with right-pad of specified length Argument 1: The text in which left-pad is to be added Argument 2: The length of string to be returned Argument 3: The text that is to be padded in specified string Returns: A string	Argument 1 = “N208WN” Argument 2 = 9 Argument 3 = “XY” Returns “N208WNXYX”
		Argument 1 = “N208WN” Argument 2 = 6 Argument 3 = “XY” Returns “N208WN”
		Argument 1 = “N208WN” Argument 2 = 4 Argument 3 = “XY” Returns “N208”
substring(“string”, i, i)	Returns a string containing a character copied (starting at a specified position and ending at a specified position) from a specified string Argument 1: The text from which the characters have to be copied Argument 2: Starting position from which the characters have to be copied considering the position of the first character at 0 Argument 3: Ending position up to which the characters in the text are to be copied Returns: A string	Argument 1 = “N208WN” Argument 2 = 2 Argument 3 = 4 Returns “08”
		Argument 1 = “N208WN” Argument 2 = 2 Argument 3 = 6 Returns “08WN”
trim(“string”)	Returns a string with leading and trailing blanks removed Argument 1: The text from which the extra spaces are to be removed Returns: A string	Argument 1 = “08WN” Returns “08WN”
		Argument 1 = “N208” Returns “N208”
		Argument 1 = “208W” Returns “208W”

For numeric data type:

Operation	Description	Example
abs(number)	Return absolute value of a number, a number without its sign Argument 1: The number for which absolute value is required Returns: A number	Argument 1 = 32 Returns 32 Argument 1 = 67.98 Returns 67.98 Argument 1 = -23 Returns 23
ceil(d)	Returns the smallest whole number that is greater than or equal to a specified number Argument 1: The number that has to be rounded up Returns: A number	Argument 1 = 26 Returns 26 Argument 1 = 26.7 Returns 27 Argument 1 = -26.7 Returns -26
divide(number, number)	Returns the quotient of two numbers Argument 1: Dividend number that is to be divided by the divisor Argument 1: Divisor number Returns: A number	Argument 1 = 551 Argument 2 = 2 Returns 275.50 Argument 1 = -450 Argument 2 = 3 Returns -150.00 Argument 1 = 551 Argument 2 = 0 Returns NULL Argument 1 = 0 Argument 2 = 551 Returns 0.00
exp(d)	Returns the exponential value of a number Argument 1: The exponent applied to base e Returns: A number	Argument 1 = 1145 Returns “Infinity” Argument 1 = 12 Returns 162754.79 Argument 1 = -25 Returns 0.00
fact(i)	Returns the factorial of a number Argument 1: The number for which factorial is to be calculated Returns: A number	Argument 1 = 7 Returns 5040 Argument 1 = -5 Returns NULL
floor(d)	Returns the largest whole number that is smaller than or equal to a specified number Argument 1: The number to be rounded down Returns: A number	Argument 1 = 26 Returns 26 Argument 1 = 26.7 Returns 26 Argument 1 = -26.7 Returns -27
log(d)	Returns natural logarithm (base e) of a number Argument 1: A value greater than 0 for which logarithm is to be calculated Returns: A number	Argument 1 = 551 Returns 6.31 Argument 1 = -551 Returns NULL

		Argument 1 = 551.45 Returns 6.31
logTen(d)	Returns decimal logarithm (base 10) of a number Argument 1: The value greater than 0 for which logarithm is to be calculated Returns: A number	Argument 1 = 551 Returns 2.74
		Argument 1 = -551 Returns NULL
		Argument 1 = 551.45 Returns 2.74
max(number, number)	Returns larger of two numbers Argument 1: First number to find out if it is larger than the second number Argument 2: Second number to find out if it is larger than the first number Returns: A number	Argument 1 = 198 Argument 2 = 1660 Returns 1660.00
		Argument 1 = 198 Argument 2 = -1660 Returns 198.00
min(number, number)	Returns smaller of two numbers Argument 1: First number to find out if it is smaller than the second number Argument 2: Second number to find out if it is smaller than the first number Returns: A number	Argument 1 = 198 Argument 2 = 1660 Returns 198.00
		Argument 1 = 198 Argument 2 = -1660 Returns -1660.00
minus(number, number)	Returns the subtraction of two numbers Argument 1: A base number Argument 2: A number that is to be subtracted from the base number Returns: A number	Argument 1 = -5 Argument 2 = 1237 Returns -1242.00
		Argument 1 = 1237 Argument 2 = 29 Returns 1208.00
mod(number, number)	Returns modulus of two numbers Argument 1: Dividend: The number to be divided Argument 2: Divisor: The number by which the dividend has to be divided Returns: A number	Argument 1 = 460 Argument 2 = 72 Returns 28.00
		Argument 1 = -460 Argument 2 = 72 Returns 44.00
		Argument 1 = 460 Argument 2 = -72 Returns 28.00
		Argument 1 = -460 Argument 2 = -72 Returns -28.00
multiply(number, number)	Returns the product of two numbers Argument 1: A base number Argument 2: A number that is to be multiplied by the base number Returns: A number	Argument 1 = 460 Argument 2 = 72 Returns 33120.00
		Argument 1 = -460 Argument 2 = 72 Returns -33120.00
		Argument 1 = -460 Argument 2 = -72 Returns 33120.00

pi(d)	Returns pi times a number Argument 1: The number Returns: A number	Argument 1 = 641 Returns 2013.76 Argument 1 = -3 Returns -9.42
plus(number, number)	Returns the sum of two numbers Argument 1: A base number Argument 1: A number that is to be added to the base number Returns: A number	Argument 1 = 460 Argument 2 = 72 Returns 532.00 Argument 1 = 460 Argument 2 = -72 Returns 388.00 Argument 1 = -460 Argument 2 = -72 Returns -532.00
round(d, i)	Returns the number rounded to a specified number of decimal places Argument 1: The number to be rounded Argument 2: The number of places to which the number is to be rounded Returns: A number	Argument 1 = 12.356 Argument 2 = 1 Returns 12.40 Argument 1 = -12.356 Argument 2 = 1 Returns -12.40 Argument 1 = 12.356 Argument 2 = 2 Returns 12.36 Argument 1 = 12.356 Argument 2 = 3 Returns 12.35
sign(d)	Returns a number (-1, 0, or 1) indicating the sign of a number Argument 1: The number for which the algebraic sign is to be determined Returns: A number	Argument 1 = -5 Returns -1 Argument 1 = 0 Returns 0 Argument 1 = 29 Returns 1
sqrt(d)	Returns the square root of a number Argument 1: A positive value for which the square root is to be calculated Returns: A number	Argument 1 = 100 Returns 10.00 Argument 1 = 588 Returns 24.24 Argument 1 = -588 Returns NaN (Not a number)

For datetime data type:

Operation	Description	Example
date(Timestamp)	Returns the date part of a timestamp Argument 1: The timestamp for which the date has to be returned Returns: A date	Argument 1 = 2018-02-16 20:38:40 Returns 2018-02-16

dateAdd("string", i , date)	Adds a certain date or time interval to a date Argument 1: The type of interval to be added (where the type of interval can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The interval to be added Argument 3: The date or timestamp to which the specified interval has to be added or subtracted Returns: A timestamp	Argument 1 = "Year" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2020-02-16 20:38:40
		Argument 1 = "Month" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-04-16 20:38:40
		Argument 1 = "Day" Argument 2 = 10 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-26 20:38:40
		Argument 1 = "Hour" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:38:40
		Argument 1 = "Minute" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:40:40
		Argument 1 = "Second" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:38:42
dateDiff("string", date, date)	Returns the number of intervals between two dates or times Argument 1: The type of interval to be calculated (where the type of interval to be calculated can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The start date or time Argument 3: The end date or time Returns: A number	Argument 1 = "Year" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2016-02-16 20:30:20 Returns 2
		Argument 1 = "Month" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-05-16 20:38:40 Returns -3
		Argument 1 = "Day" Argument 2 = 2018-02-20 20:38:40 Argument 3 = 2018-02-16 20:38:40 Returns 4
		Argument 1 = "Hour" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:38:40 Returns 10
		Argument 1 = "Minute" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:18:40 Returns 10
		Argument 1 = "Second" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:38:10 Returns 30

day(date)	Returns the day of a date represented by a number (an integer between 1 and 31) Argument 1: The date or timestamp argument whose day of the month is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 16
dayName (date)	Returns the name of the day of the week Argument 1: The date or timestamp for which day of the week is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns Friday
dayOfWeek(date)	Returns a number (between 1 and 7) representing the day of the week (for example, Monday is 1, Wednesday is 3, and Sunday is 7) Argument 1: The date or timestamp for which day of the week is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 5
dayofYear(date)	Returns a number representing the day of the year (an integer between 1 and 366) Argument 1: The date or timestamp for which day of the year is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 47
daysAfter(date, date)	Returns the count of number of days after specified date Argument 1: The start date Argument 2: The end date Returns: A number	Argument 1 = 2018-02-16 20:38:40 Argument 2 = 2018-02-10 20:38:40 Returns 6
formatDate(date, “string”)	Returns the date or timestamp in specified format as a string datatype Argument 1: The date or timestamp Argument 2: The format in which specified date or timestamp will be transformed (where the format can be user defined, such as “dd-mm-yy hh:mm:ss”) Returns: A string	Argument 1 = 2018-02-16 Argument 2 = “yy/mm/dd” Returns 18/02/16 Argument 1 = 2018-02-16 20:38:40 Argument 2 = “mm/dd/yyyy” Returns 02/16/2018
hour(date)	Returns the hour of a time value (an integer ranging from 0 [12:00 AM] to 23 [11:00 PM]) Argument 1: The timestamp for which hours are to be returned	Argument 1 = 2018-02-16 20:38:40 Returns 20

	Returns: A number	
minute(date)	Returns the minutes of a time value (an integer ranging from 0 to 59) Argument 1: The timestamp for which minutes are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 38
month(date)	Returns the month (an integer between 1 and 12) Argument 1: The date or timestamp for which month is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 2
monthName (date)	Returns the month name for a given date or timestamp Argument 1: The date or timestamp for which month name is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns February
quarter(date)	Returns the quarter corresponding to a date (an integer between 1 and 4) Argument 1: The date or timestamp for which quarter is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 1
second(timestamp)	Returns the seconds of a time value (an integer in the range 0 to 59) Argument 1: The timestamp for which seconds are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 40
time(timestamp)	Returns the time part from a given timestamp as a string datatype Argument 1: The timestamp for which time part is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns “20:38:40”
weekOfMonth (date)	Returns a number (between 1 and 5) representing the week of the month (for example, 1 for the first week and 3 for the third week) Argument 1: The date or timestamp for which week of month is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 3
weekOfYear (date)	Returns a number (between 1 and 52) representing the week of the year (for example, 8 for the eighth week, 19 for the nineteenth week) Argument 1: The date or timestamp for which week of the year is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 7

year(date)	Returns the year part of the date or timestamp (for example, 2001, 2018, 3000) Argument 1: The date or timestamp for which year part is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 2018

7.2.14 Adding a Column in a Dataset

You can add columns to a dataset as per your requirements.

Reference:	**Concept Manual > Shape Data > Add Column**

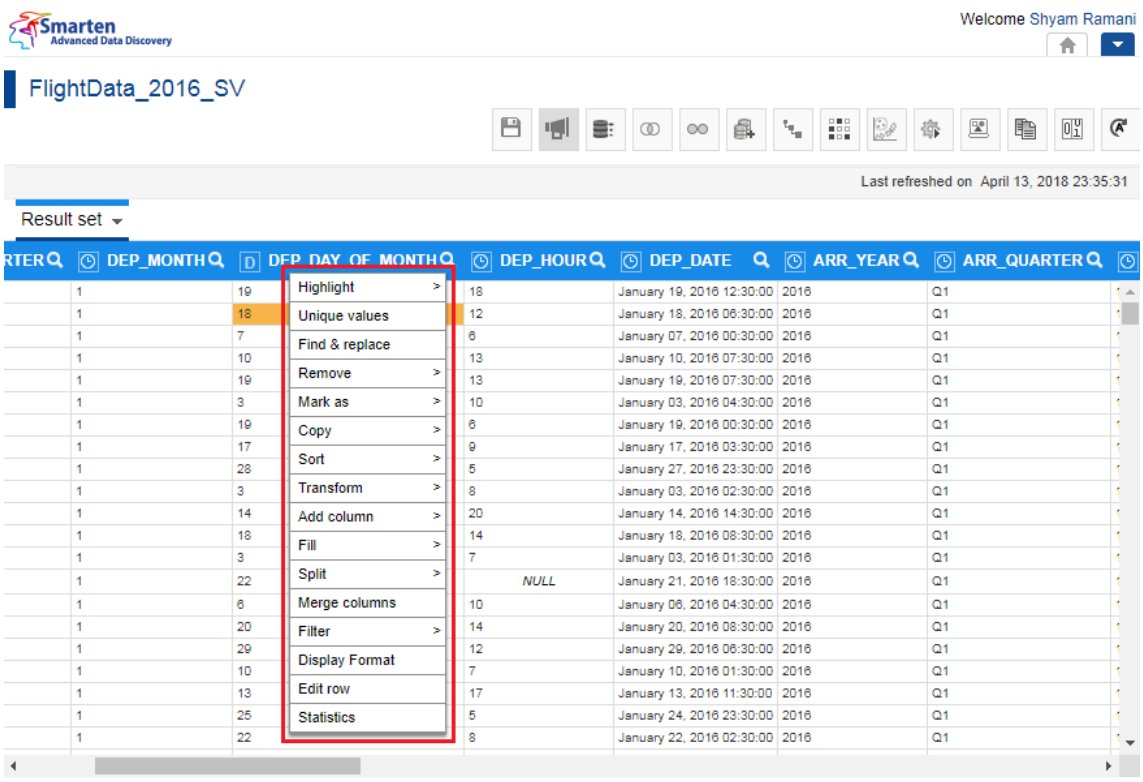
About this task

Use this task to add a column in a dataset.

Procedure

1. Open the dataset in which you want to add a column.
2. Right-click in the column next to which you want to add a column.

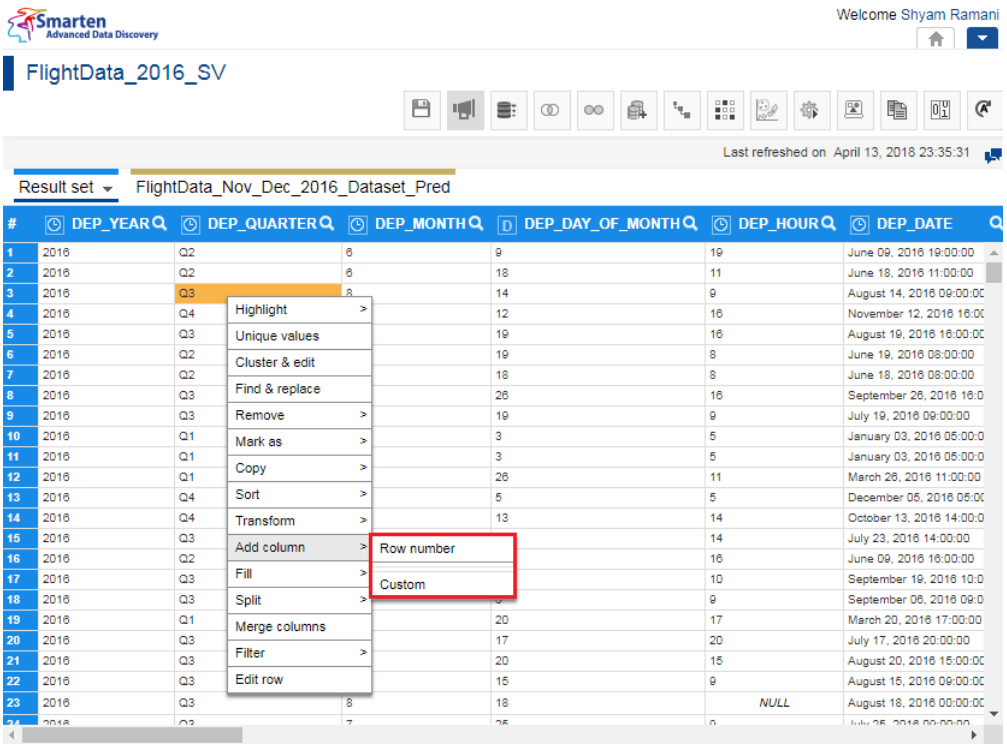
   The system displays the context menu.



ADD A COLUMN—THE CONTEXT MENU

3. Click **Add column** from the menu.

   The system displays the options available to add a column in the dataset.



ADD A COLUMN—OPTIONS AVAILABLE FOR ADDING A COLUMN

4. Click any of the options in the following table to apply that operation.

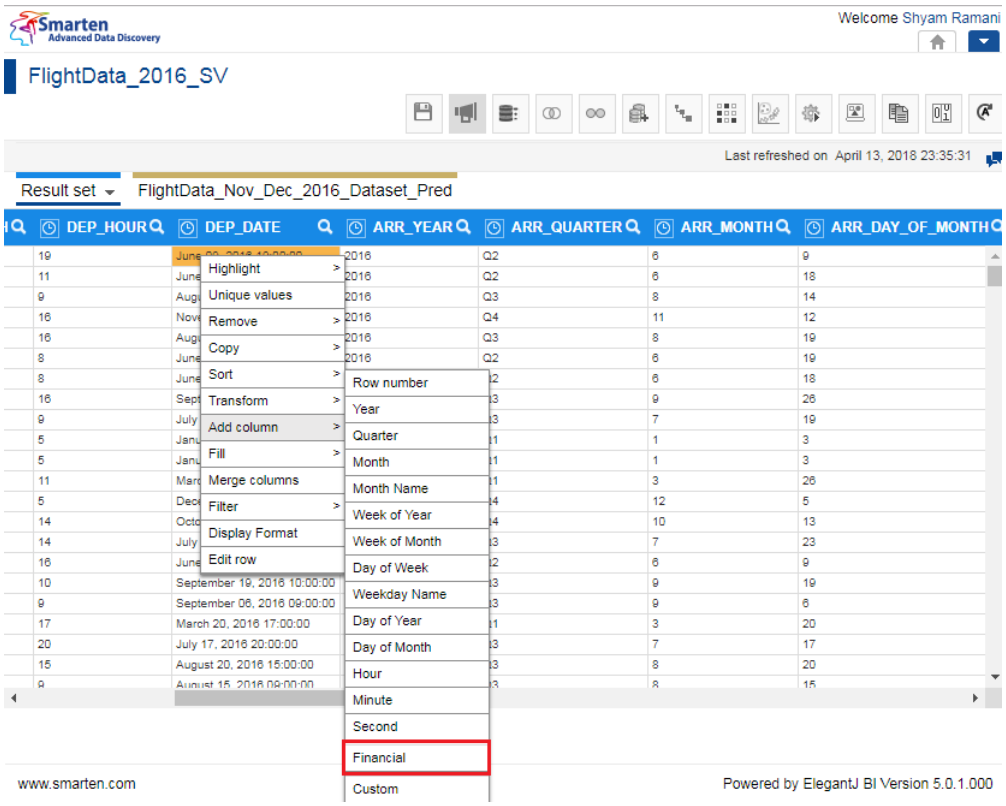
Option	Description
Row number	Use this option to add a column that contains row number in ascending order starting from zero next to the selected column. For example, in the image above, the DEP_DATE column contains 100 records. When you apply Row number operation on that column, the system adds a column after the DEP_DATE column that contains row numbers in ascending order starting from zero. Row numbers sequence is based on the current order of the dataset.
Custom	Use this option to add a custom column based on the expression you have provided.

The table below describes the options available for the DateTime data type.

Function	Description	Example (23-Jan-2015 03:45:00)
Year	Use this option to return the year part of the selected date.	2015
Quarter	Returns the quarter corresponding to the date with the year starting from January.	1

Month	Returns the month number of the date with the year starting from January.	1
Month name	Returns the name of the month.	January
Week of Year	Returns a number (between 1 and 52) representing the week of the year starting from January.	4
Week of Month	Returns a number (between 1 and 5) representing the week of the month with the year starting from January.	4
Day of Week	Returns a number (between 1 and 7) representing the day of the week with the year starting from January.	5
Weekday Name	Returns the name of the weekday.	Friday
Day of Year	Returns a number representing the day of the year (an integer between 1 and 366) with the year starting from January.	23
Day of Month	Returns the day of the month.	23
Hour	Returns the hour of a time value (an integer ranging from 0 (12:00 A.M.) to 23 (11:00 P.M.).	03
Minute	Returns the minutes of a time value (an integer ranging from 0 to 59).	45
Second	Returns the seconds of a time value (an integer in the range from 0 to 59).	00

While adding a column for a datetime data type column, you can click the **Financial** option.



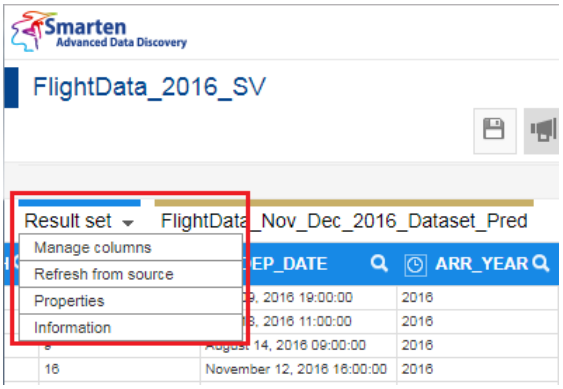
ADD A COLUMN—FINANCIAL OPTION FOR DATETIME DATA TYPE

Shown below are the options available for the Financial suboption.

Financial		
Function	**Description**	**Example (23-Jan-2015 03:45:00) Financial year start - April**
Year	Returns the year part of the date.	2014
Quarter	Returns the quarter corresponding to the date with the year starting from the user's selected month.	4
Month	Returns the month number of the date with the year starting from the user's selected month.	10
Week of Year	Returns a number (between 1 and 52) representing the week of the year starting from the user's selected month.	43
Week of Month	Returns a number (between 1 and 5) representing the week of the month with the year starting from the user's selected month.	4
Day of Month	Returns the day of the month.	23

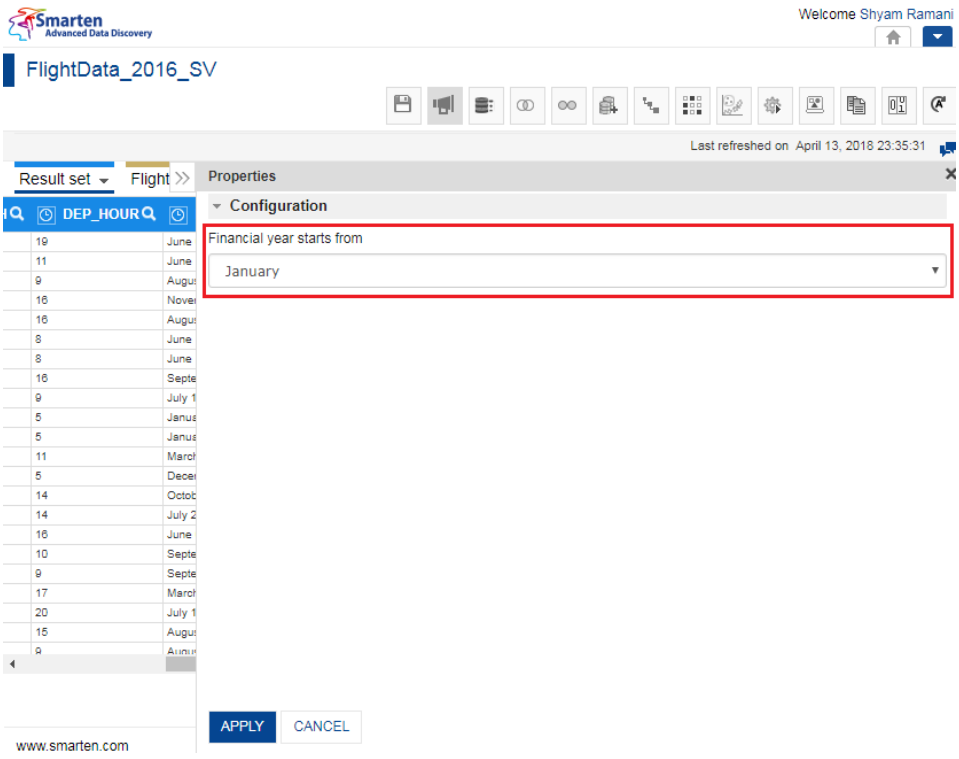
Hour	Returns the hour of a time value (an integer ranging from 0 (12:00 A.M.) to 23 (11:00 P.M.).	03
Minute	Returns the minutes of a time value (an integer ranging from 0 to 59).	45
Second	Returns the seconds of a time value (an integer ranging from 0 to 59).	00

In the example above, you can set any month as the start of the financial year using the **Properties** option of the **Resultset** menu.



RESULTSET MENU – THE PROPERTIES OPTION

The system displays the **Properties** dialog box.



RESULTSET PROPERTIES—SETTING START OF A FINANCIAL YEAR

The system adds a new column based on the option you have selected.

7.2.14.1 Using Custom Options to Add a Column

You can quickly create custom columns by creating expressions based on various string, arithmetic, date, or miscellaneous statements using various arithmetic and comparison operators.

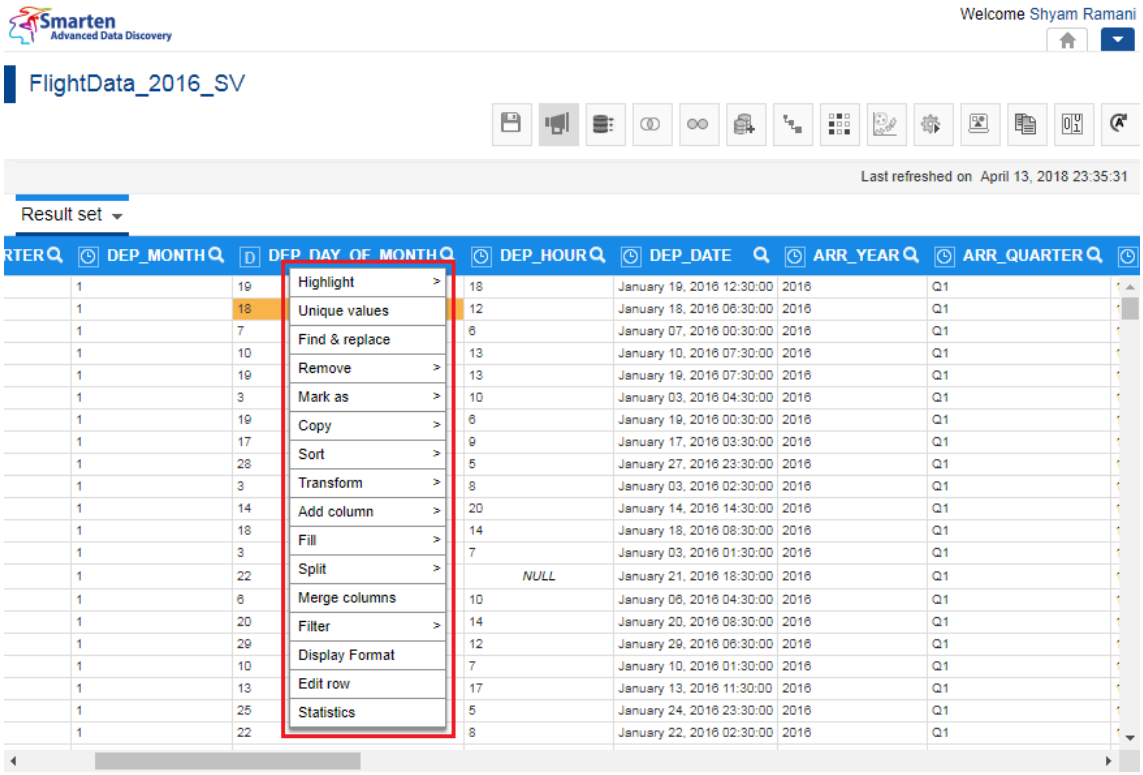
About this task

Use this task to add a column in a dataset using custom options.

Procedure

1. Open the dataset in which you want to add a column.
2. Right-click in the column next to which you want to add a column.

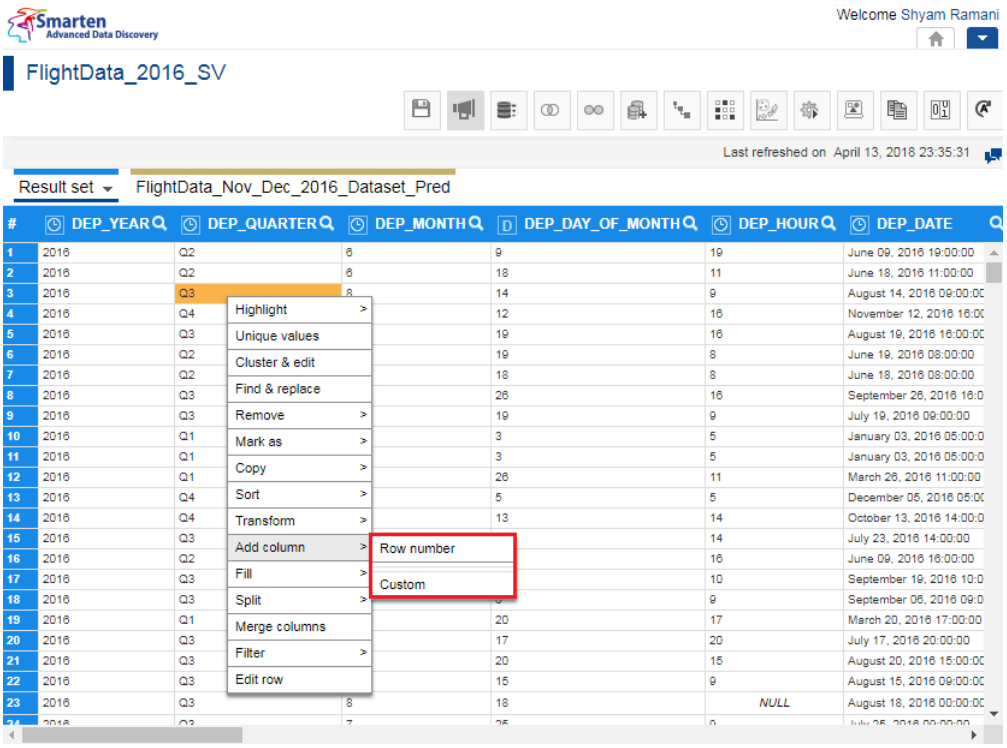
   The system displays the context menu.



ADD A COLUMN—THE CONTEXT MENU

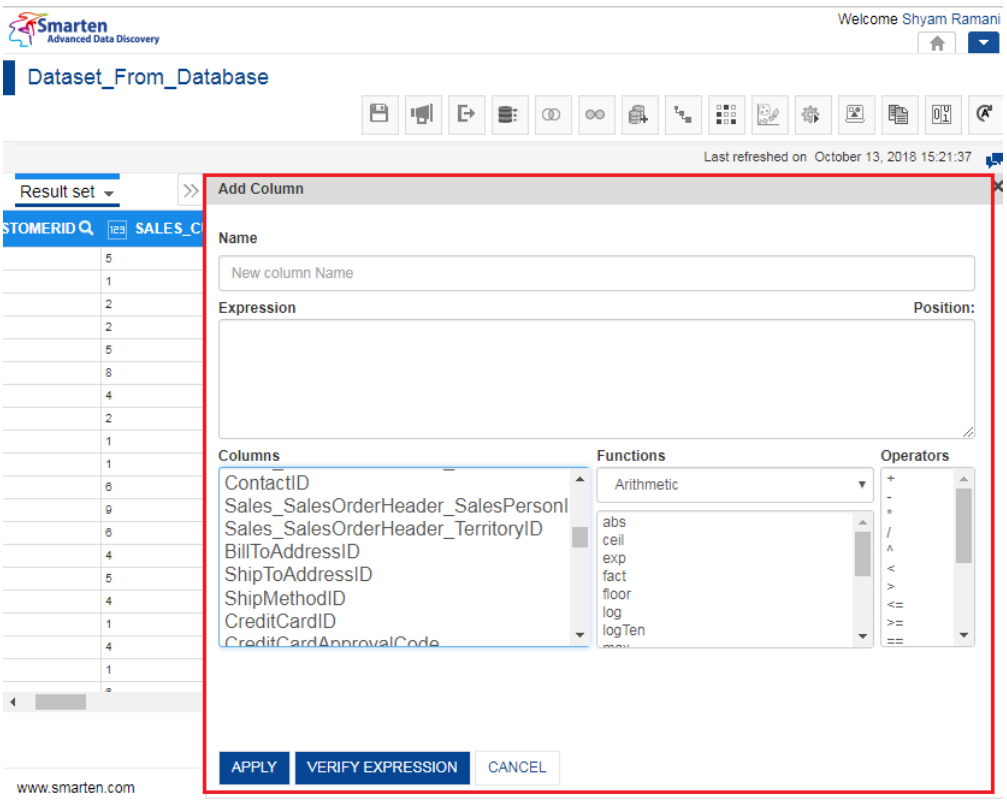
3. Click **Add column** from the menu.

   The system displays the options available to add a column in the dataset.



ADD A COLUMN—OPTIONS AVAILABLE FOR ADDING A COLUMN

4. Click **Custom**.
5. The system displays the **Add Column** dialog box.



ADD A COLUMN—ADDING A COLUMN USING CUSTOM OPTION

6. Enter a name for the new column in the **Name** box.
7. Double-click name of the column you want to use in the expression from the **Columns** section.

8. Select an option from the Functions list.

   The following options are available:

   - Arithmetic
   - Date
   - Miscellaneous
   - String

9. Select an operation.
   The operations available are based on the option you have selected in the above step.
10. Select an operator you want to use from the **Operators** section.
11. Click **APPLY**.

You can refer to the tables below for more information about the operations available for arithmetic, string, date, and miscellaneous functions.

Arithmetic functions:

Function	Description	Example
abs(number)	Returns absolute value of a number, a number without its sign Argument 1: The number for which absolute value is required Returns: A number	Argument 1 = 32 Returns 32 Argument 1 = 67.98 Returns 67.98 Argument 1 = -23 Returns 23
ceil(d)	Returns the smallest whole number that is greater than or equal to a specified number Argument 1: The number that has to be rounded up Returns: A number	Argument 1 = 26 Returns 26 Argument 1 = 26.7 Returns 27 Argument 1 = -26.7 Returns -26
exp(d)	Returns the exponential value of a number Argument 1: The exponent applied to base e Returns: A number	Argument 1 = 1145 Returns “Infinity” Argument 1 = 12 Returns 162754.79 Argument 1 = -25 Returns 0.00
fact(i)	Returns the factorial of a number Argument 1: The number for which factorial is to be calculated Returns: A number	Argument 1 = 7 Returns 5040 Argument 1 = -5 Returns NULL
floor(d)	Returns the largest whole number that is smaller than or equal to a specified number Argument 1: The number to be rounded down Returns: A number	Argument 1 = 26 Returns 26 Argument 1 = 26.7 Returns 26 Argument 1 = -26.7 Returns -27

log(d)	Returns natural logarithm (base e) of a number Argument 1: A value greater than 0 for which logarithm is to be calculated Returns: A number	Argument 1 = 551 Returns 6.31 Argument 1 = -551 Returns NULL Argument 1 = 551.45 Returns 6.31
logTen(d)	Returns decimal logarithm (base 10) of a number Argument 1: The value greater than 0 for which logarithm is to be calculated Returns: A number	Argument 1 = 551 Returns 2.74 Argument 1 = -551 Returns NULL Argument 1 = 551.45 Returns 2.74
max(number, number)	Returns larger of two numbers Argument 1: First number to find out if it is larger than the second number Argument 2: Second number to find out if it is larger than the first number Returns: A number	Argument 1 = 198 Argument 2 = 1660 Returns 1660.00 Argument 1 = 198 Argument 2 = -1660 Returns 198.00
min(number, number)	Returns smaller of two numbers Argument 1: First number to find out if it is smaller than the second number Argument 2: Second number to find out if it is smaller than the first number Returns: A number	Argument 1 = 198 Argument 2 = 1660 Returns 198.00 Argument 1 = 198 Argument 2 = -1660 Returns -1660.00
mod(number, number)	Returns modulus of two numbers Argument 1: Dividend: The number to be divided Argument 2: Divisor: The number by which the dividend has to be divided Returns: A number	Argument 1 = 460 Argument 2 = 72 Returns 28.00 Argument 1 = -460 Argument 2 = 72 Returns 44.00 Argument 1 = 460 Argument 2 = -72 Returns 28.00 Argument 1 = -460 Argument 2 = -72 Returns -28.00
pi(d)	Returns pi times a number Argument 1: The number Returns: A number	Argument 1 = 641 Returns 2013.76 Argument 1 = -3 Returns -9.42
random(number, number)	Returns a random number between two specified numbers Argument 1: The smallest integer value Argument 2: The largest integer value	Argument 1 = 54 Argument 2 = 55 Returns 54.45/54.51 /54.95…

	Returns: A number	Argument 1 = 72 Argument 2 = 80 Returns 72.89/ 73.94/75.20/76.47..
		Argument 1 = 20 Argument 2 = -10 Returns -7.68/-9.75/-2.65/5.97
round(d, i)	Returns a number rounded to a specified number of decimal places Argument 1: The number to be rounded Argument 2: The number of places to which the number is to be rounded Returns: A number	Argument 1 = 12.356 Argument 2 = 1 Returns 12.40
		Argument 1 = -12.356 Argument 2 = 1 Returns -12.40
		Argument 1 = 12.356 Argument 2 = 2 Returns 12.36
		Argument 1 = 12.356 Argument 2 = 3 Returns 12.35
sign(d)	Returns a number (-1, 0, or 1) indicating the sign of a number Argument 1: The number for which the algebraic sign is to be determined Returns: A number	Argument 1 = -5 Returns -1
		Argument 1 = 0 Returns 0
		Argument 1 = 29 Returns 1
sqrt(d)	Returns the square root of a number Argument 1: A positive value for which the square root is to be calculated Returns: A number	Argument 1 = 100 Returns 10.00
		Argument 1 = 588 Returns 24.24
		Argument 1 = -588 Returns NaN (Not a number)
truncate(d, i)	Returns a number truncated to a specified number of decimal places Argument 1: The number to be truncated Argument 2: The scale of the truncation Returns: A number	Argument 1 = 10.54 Argument 2 = 1 Returns 10.50
		Argument 1 = 10.54 Argument 2 = 2 Returns 10.54
		Argument 1 = 10.54 Argument 2 = 0 Returns 11.00
		Argument 1 = 10.25 Argument 2 = 0 Returns 10.00

		Argument 1 = -10.54 Argument 2 = 1 Returns -10.50

String functions:

Function	Description	Examples
asc(c)	Returns ASCII value of a character Argument 1: The character for which the ASCII value is to be returned Returns: A number	Argument 1 = A Returns 65
		Argument 1 = a Returns 97
		Argument 1 = "1" Returns 49
booleanValue ("string")	Returns the content of a string as a boolean Argument 1: The string from which boolean is to be returned Returns: A boolean	Argument 1 = "True" Returns true
		Argument 1 = "TRUE" Returns true
		Argument 1 = "abc" Returns false
		Argument 1 = "748" Returns false
byteValue(object)	Returns the content of a string as a byte Argument 1: The object from which byte is to be returned Returns: A byte	Argument 1 = "N787AA" Returns 0
		Argument 1 = "-128" Returns -128
		Argument 1 = "-129" Returns 0
		Argument 1 = "127" Returns 127
		Argument 1 = "128" Returns 0
		Argument 1 = "120.54" Returns 120
charValue(i)	Returns the content of an integer as a character Argument 1: The number from which character is to be returned Returns: A character	Argument 1 = 65 Returns "A"
		Argument 1 = 97 Returns "a"
		Argument 1 = 49 Returns "1"
doubleValue(object)	Returns the content of a string as double Argument 1: The object from which double is to be returned Returns: A double	Argument 1 = "748" Returns 748.00
		Argument 1 = "748.52" Returns 748.52
		Argument 1 = "-748.52" Returns -748.52

		Argument 1 = “abc” Returns 0.00
		Argument 1 = “ABC” Returns 0.00
fill(“string,” i)	Returns a string of specified length filled with occurrences of a specified string Argument 1: The string that has to be filled Argument 2: The length of the filled string Returns: A string	Argument 1 = “N787AA” Argument 2 = 2 Returns “N7”
		Argument 1 = “N787AA” Argument 2 = 6 Returns “N787AA”
		Argument 1 = “N787AA” Argument 2 = 9 Returns “N787AAN78”
floatValue(object)	Returns the content of a string as a float Argument 1: The object from which float is to be returned Returns: A float	Argument 1 = “748” Returns 748.00
		Argument 1 = “-748.52” Returns -748.52
		Argument 1 = “abc” Returns 0.00
		Argument 1 = “ABC” Returns 0.00
indexOfChar(“string”, c, i)	Returns the starting position of a character within a specified string Argument 1: The string from which the index is to be returned Argument 2: The character to find the index Argument 3: The starting position of the string in number Returns: A number	Argument 1 = “N787AA” Argument 2 = ‘7’ Argument 3 = 1 Returns 1
		Argument 1 = “N787AA” Argument 2 = ‘7’ Argument 3 = 3 Returns 3
		Argument 1 = “N787AA” Argument 2 = ‘A’ Argument 3 = 3 Returns 4
		Argument 1 = “N787AA” Argument 2 = ‘A’ Argument 3 = 6 Returns 5
		Argument 1 = “N787AA” Argument 2 = ‘Y’ Argument 3 = 1 Returns -1
IndexOfString (“string”, “string”, i)	Returns the starting position of a string within a specified string Argument 1: The string from which the index is to be returned	Argument 1 = “N208WN” Argument 2 = ‘208’ Argument 3 = 1 Returns 1

	Argument 2: The string to find index Argument 3: The starting position of the string in number Returns: A number	Argument 1 = “N208WN” Argument 2 = ‘208’ Argument 3 = 4 Returns -1
		Argument 1 = “N208WN” Argument 2 = ‘WN’ Argument 3 = 4 Returns 4
intValue(object)	Returns the contents of a string as an integer Argument 1: The object from which integer is to be returned Returns: An integer	Argument 1 = “N787AA” Returns 0
		Argument 1 = “748” Returns 748
		Argument 1 = “748.52” Returns 748
		Argument 1 = “-748” Returns -748
isDate(“string”)	Determine if the specified string contains a valid date Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = “2015-01-09” Returns “true”
		Argument 1 = “N787AA” Returns “false”
isNull(object)	Determines if the argument is NULL Argument 1: The object that is to be checked Returns: A boolean	Argument 1 = “N787AA” Returns “false”
		Argument 1 = NULL Returns “true”
isNumber(“string”)	Determines if the specified string contains a number Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = “N787AA” Returns “false”
		Argument 1 = “787” Returns “true”
isTime(“string”)	Determines if the specified string contains a valid time Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = “15:30:00” Returns “true”
		Argument 1 = “N787AA” Returns “false”
left(“string”, i)	Returns a specified number of characters from a string starting with the first character Argument 1: The text from which the partial words are to be returned Argument 2: The number of characters to be extracted from the beginning of the text Returns: A string	Argument 1 = “N787AA” Argument 2 = 2 Returns “N7”
		Argument 1 = “N787AA” Argument 2 = 8 Returns “N787AA”
leftTrim(“string”)	Returns a copy of a specified string with leading blanks removed Argument 1: The text for which blank spaces are to be removed from left	Argument 1 = “87AA” Returns “87AA”
		Argument 1 = “87AA” Returns “87AA”

	Returns: A string	Argument 1 = “87AA” Returns “87AA”
length(“string”)	Returns the length of a string Argument 1: The string for which length is to be checked Returns: A number	Argument 1 = “N787AA” Returns 6
		Argument 1 = “748” Returns 3
		Argument 1 = “748.52” Returns 6
		Argument 1 = “-748.52” Returns 7
longValue(object)	Returns the content of a string as long Argument 1: The object from which long is to be returned Returns: A long	Argument 1 = “N787AA” Returns 0
		Argument 1 = “748” Returns 748
		Argument 1 = “748.52” Returns 748
		Argument 1 = “-748.52” Returns -748
match(“string”, “string”)	Returns a determination whether or not a string contains a particular pattern of characters Argument 1: The text that has to be searched in argument 2 Argument 2: The text in which argument 1 has to be searched Returns: A number	Argument 1 = “AA” Argument 2 = “N787AA” Returns 1
		Argument 1 = “aa” Argument 2 = “N787AA” Returns 0
		Argument 1 = “AB” Argument 2 = “N787AA” Returns 0
replace(“string”, i, i, “string”)	Returns a copy of a specified string in which a specified number of characters starting with a specified character have been replaced with characters from another specified string Argument 1: The string to be processed Argument 2: Start index Argument 3: End index Argument 4: The string to be replaced Returns: A string	Argument 1 = “N208WN” Argument 2 = 1 Argument 3 = 2 Argument 4 = “3” Returns “N308WN”
		Argument 1 = “N208WN” Argument 2 = 4 Argument 3 = 6 Argument 4 = “ML” Returns “N208ML”
		Argument 1 = “N208WN” Argument 2 = 0 Argument 3 = 1 Argument 4 = “M” Returns “M208WN”
reverse(“string”)	Reverses the order or characters in a string Argument 1: The text that needs to be reversed Returns: A string	Argument 1 = “N208WN” Returns “NW802N”

right("string", i)	Returns the specified number of characters from the end of a specified string Argument 1: The text from which the specified number of characters should be returned from the end Argument 2: The number of characters to be returned from the string Returns: A string	Argument 1 = "N208WN" Argument 2 = 3 Returns "8WN"
		Argument 1 = "N208WN" Argument 2 = 8 Returns "N208WN"
rightTrim("string")	Returns a copy of a specified string with trailing blanks removed Argument 1: The text from which extra spaces have to be removed from the right Returns: A string	Argument 1 = "N208" Returns "N208"
		Argument 1 = "08WN" Returns "08WN"
		Argument 1 = "208W" Returns "208W"
shortValue(object)	Returns contents of a string as short Argument 1: The object from which short to be returned Returns: A long	Argument 1 = "N787AA" Returns 0
		Argument 1 = "748" Returns 748
		Argument 1 = "748.52" Returns 748
		Argument 1 = "-748.52" Returns -748
space(i)	Returns the string of a specified length filled with a specified number of spaces Argument 1: Number of space Returns: A string	Argument 1 = 5 Returns " "
substring("string", i, i)	Returns a string containing a character copied (starting at a specified position and ending at a specified position) from a specified string Argument 1: The text from which the characters have to be copied Argument 2: Starting position from which the characters have to be copied Argument 3: Ending position up to which the characters in the text are to be copied Returns: A string	Argument 1 = "N208WN" Argument 2 = 2 Argument 3 = 4 Returns "08"
		Argument 1 = "N208WN" Argument 2 = 2 Argument 3 = 6 Returns "08WN"
ToLowerCase ("string")	Returns a copy of a specified string with all uppercase letters converted to lowercase Argument 1: The text for which the uppercase letters are to be converted to lowercase Returns: A string	Argument 1 = "N208WN" Returns "n208wn"
		Argument 1 = "N208wN" Returns "n208wn"
		Argument 1 = "n208wn" Returns "n208wn"
toString(object)	Returns a string representation of a specified object Argument 1: The object for which string is to be returned	Argument 1 = 748 Returns "748"
		Argument 1 = 748.52 Returns "748.52"

	Returns: A string	Argument 1 = -748.52 Returns "-748.52"
		Argument 1 = 16-02-2018 20:38:40 Returns "16-02-2018 20:38:40"
ToUpperCase ("string")	Returns a copy of a specified string with all lowercase letters converted to uppercase Argument 1: The text for which the lowercase letters are to be converted to uppercase Returns: A string	Argument 1 = "n208wn" Returns "N208WN"
		Argument 1 = "n208Wn" Returns "N208WN"
		Argument 1 = "N208WN" Returns "N208WN"
trim("string")	Returns a string with leading and trailing blanks removed Argument 1: The text from which the extra spaces are to be removed Returns: A string	Argument 1 = "08WN" Returns "08WN"
		Argument 1 = "N208" Returns "N208"
		Argument 1 = "208W" Returns "208W"

Miscellaneous functions:

Functions	Description	Examples
ifCase(condition, truevalue, falsevalue)	Returns TRUE if the condition is validated and returns FALSE if invalidated Argument 1: The condition Argument 2: True value Argument 3: False value Returns: An object	Argument 1 = origin=="LAX" Argument 2 = "Los Angeles" Argument 3 = "Others" Returns "Los Angeles" if the value of origin is "LAX" or else returns "Others"
noOfDaysByDate(Start Date, EndDate)	Returns the number of days between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-03-10 Argument 2 = 2014-04-10 Returns 32
noOfHalfYearsByDate(StartDate, EndDate)	Returns a number of half years between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2014-12-31 Returns 2
		Argument 1 = 2014-01-01 Argument 2 = 2014-05-31 Returns 0
		Argument 1 = 2014-01-01 Argument 2 = 2014-08-31 Returns 1

noOfMonthsByDate(StartDate, EndDate)	Returns the number of months between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2014-12-31 Returns 12 Argument 1 = 2014-01-01 Argument 2 = 2014-07-10 Returns 6 Argument 1 = 2014-01-01 Argument 2 = 2014-05-15 Returns 4
noOfQuartersByDate(StartDate, EndDate)	Returns a number of quarters between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2014-12-31 Returns 4 Argument 1 = 2014-01-01 Argument 2 = 2014-08-15 Returns 2
noOfWeeksByDate(StartDate, EndDate)	Returns the number of weeks between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2015-01-01 Returns 52 Argument 1 = 2014-01-01 Argument 2 = 2014-07-01 Returns 25 Argument 1 = 2014-01-01 Argument 2 = 2014-01-03 Returns 0 Argument 1 = 2014-01-01 Argument 2 = 2014-01-12 Returns 1
whenThen(columnname, whenvalue1, thenresult1, whenvalue2, thenresult2, ..., elseresult)	Tests values of a column or expression and returns values based on the results of the test	Argument 1 = Origin Argument 2 = “LAX” Argument 3 = “Los Angeles” Argument 4 = “JFK” Argument 5 = “John F. Kennedy” Argument 6 = “Others” Returns “Los Angeles” if the value of column “origin” is “LAX” and “John F. Kennedy” if the value is “JFK.” For rest of the values, returns “Others”

Date functions:

Function	Description	Example
date(Timestamp)	Returns the date part of a timestamp Argument 1: The timestamp for which the date has to be returned Returns: A date	Argument 1 = 2018-02-16 20:38:40 Returns 2018-02-16
dateAdd (“string”, i , date)	Adds a certain date or time interval to a date Argument 1: The interval of time (where the type of interval can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The number of interval to be added to the time Argument 3: The date and time to add the interval to Returns: A date	Argument 1 = “y” or “Y” Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2020-02-16 20:38:40
		Argument 1 = “m” or “M” Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-04-16 20:38:40
		Argument 1 = “d” or “D” Argument 2 = 10 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-26 20:38:40
		Argument 1 = “h” or “H” Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:38:40
		Argument 1 = “n” or “N” Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:40:40
		Argument 1 = “s” or “S” Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:38:42
dateDiff (“string”, date, date)	Returns the number of intervals between two dates or times Argument 1: The interval of time (where the type of interval to be calculated can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The first date or time Argument 3: The second date or time Returns: A number	Argument 1 = “y” or “Y” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2016-02-16 20:30:20 Returns 2
		Argument 1 = “m” or “M” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-05-16 20:38:40 Returns -3
		Argument 1 = “d” or “D” Argument 2 = 2018-02-20 20:38:40 Argument 3 = 2018-02-16 20:38:40 Returns 4
		Argument 1 = “h” or “H” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:38:40 Returns 10

		Argument 1 = “n” or “N” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:18:40 Returns 10
		Argument 1 = “s” or “S” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:38:10 Returns 30
datePart (“string”, date)	Returns the specified part of a given date Argument 1: The interval of time (where the part of the date can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The date Returns: A number	Argument 1 = “y” or “Y” Argument 2 = 2018-02-16 20:38:40 Returns 2018
		Argument 1 = “m” or “M” Argument 2 = 2018-02-16 20:38:40 Returns 2
		Argument 1 = “d” or “D” Argument 2 = 2018-02-16 20:38:40 Returns 16
		Argument 1 = “h” or “H” Argument 2 = 2018-02-16 20:38:40 Returns 20
		Argument 1 = “n” or “N” Argument 2 = 2018-02-16 20:38:40 Returns 38
		Argument 1 = “s” or “S” Argument 2 = 2018-02-16 20:38:40 Returns 40
dateTime(“string”)	Returns contents of a string as date-time Argument 1: The string for which date-time is to be returned Returns: A date-time	Argument 1 = “2018-02-16 20:38:40” Returns 2018-02-16 20:38:40
day(date)	Returns the day of a date represented by a number (an integer between 1 and 31) Argument 1: The date or timestamp for which day part is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 16
dayName (date)	Returns the name of the day of the week Argument 1: The date or timestamp for which day of the week is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns Friday
dayofWeek(date)	Returns a number (between 1 and 7) representing the day of the week Argument 1: The date or timestamp for which day of the week is to be returned	Argument 1 = 2018-02-16 20:38:40 Returns 5

	Returns: A number	
daysAfter(date, date)	Returns the count of number of days after a specified date Argument 1: The start date Argument 2: The end date Returns: A number	Argument 1 = 2018-02-16 20:38:40 Argument 2 = 2018-02-10 20:38:40 Returns 6
formatDate (date, “string”)	Returns the date format for a given pattern Argument 1: The target date Argument 2: The string (where the format can be user defined, such as “dd-mm-yy hh:mm:ss”) Returns: A date	Argument 1 = 2018-02-16 Argument 2 = “yy/mm/dd” Returns 18/02/16
		Argument 1 = 2018-02-16 20:38:40 Argument 2 = “mm/dd/yyyy” Returns 02/16/2018
hour(date)	Returns the hour of a time value (an integer ranging from 0 [12:00 AM] to 23 [11:00 PM]) Argument 1: The timestamp for which hours are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 20
minute(date)	Returns the minutes of a time value (an integer ranging from 0 to 59) Argument 1: The timestamp for which minutes are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 38
month(date)	Returns the month (an integer between 1 and 12) Argument 1: The date or timestamp for which month is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 2
monthName(i, [b], [i])	Returns the month name for a given month number Argument 1: The number for month Argument 2: True if the month name is abbreviated; otherwise, False (Optional to enter. Default is False) Argument 3: The starting month of year in number (Optional to enter. Default is 1 for January) Returns: A string	Argument 1 = 1 Argument 2 = True Argument 3 = 1 Returns Jan
		Argument 1 = 3 Argument 2 = True Argument 3 = 4 Returns Jun
		Argument 1 = 9 Argument 2 = False Argument 3 = 1 Returns September
		Argument 1 = 2 Argument 2 = False Argument 3 = 12 Returns January

		Argument 1 = 2 Argument 2 = “” Argument 3 = “” Returns February
now()	Returns the current time Returns: A timestamp	Returns 20:38:40
relativeDate (timestamp, i)	Returns the date that occurs n days after a given date Argument 1: The date or timestamp Argument 2: The number of days to be added to the date-timestamp Returns: A timestamp	Argument 1 = 2018-02-16 20:38:40 Argument 2 = 5 Returns 2018-02-21
relativeTime (timestamp, i)	Returns the time that occurs n seconds after a given time Argument 1: The timestamp Argument 2: The number of seconds to be added to the timestamp Returns: A timestamp	Argument 1 = 20:38:40 Argument 2 = 5 Returns 20:38:45
second(timestamp)	Returns the seconds of a time value (an integer in the range 0 to 59) Argument 1: The timestamp for which seconds are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 40
time(timestamp)	Returns the time part from a given timestamp as a string datatype Argument 1: The timestamp for which time part is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns “20:38:40”
today()	Returns the current system date Returns: A date	Returns 2018-02-16
weekdayName(i, [b], [i])	Returns the day name for a given day number of a week Argument 1: The number for day of week Argument 2: True if the day name is abbreviated; otherwise, False (Optional to enter. Default is False) Argument 3: The first day of the week in number (Optional to enter. Default is 1 for Sunday) Returns: A string	Argument 1 = 1 Argument 2 = True Argument 3 = 1 Returns Sun
		Argument 1 = 1 Argument 2 = False Argument 3 = 1 Returns Sunday
		Argument 1 = 5 Argument 2 = False Argument 3 = 1 Returns Thursday
		Argument 1 = 1 Argument 2 = False Argument 3 = 3 Returns Tuesday

		Argument 1 = 1 Argument 2 = False Argument 3 = 5 Returns Thursday
		Argument 1 = 4 Argument 2 = “” Argument 3 = “” Returns Wednesday
year(date)	Returns the year corresponding to a date (an integer between 1000 and 3000) Argument 1: The date or timestamp for which year part is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 2018

7.2.15 Splitting Data in a Dataset

You can split data in a column into multiple columns or rows.

Reference: **Concept Manual > Shape Data > Split**

Note:
You can split data for string, numeric, array, and struct data type only.

7.2.15.1 Splitting Data into Columns

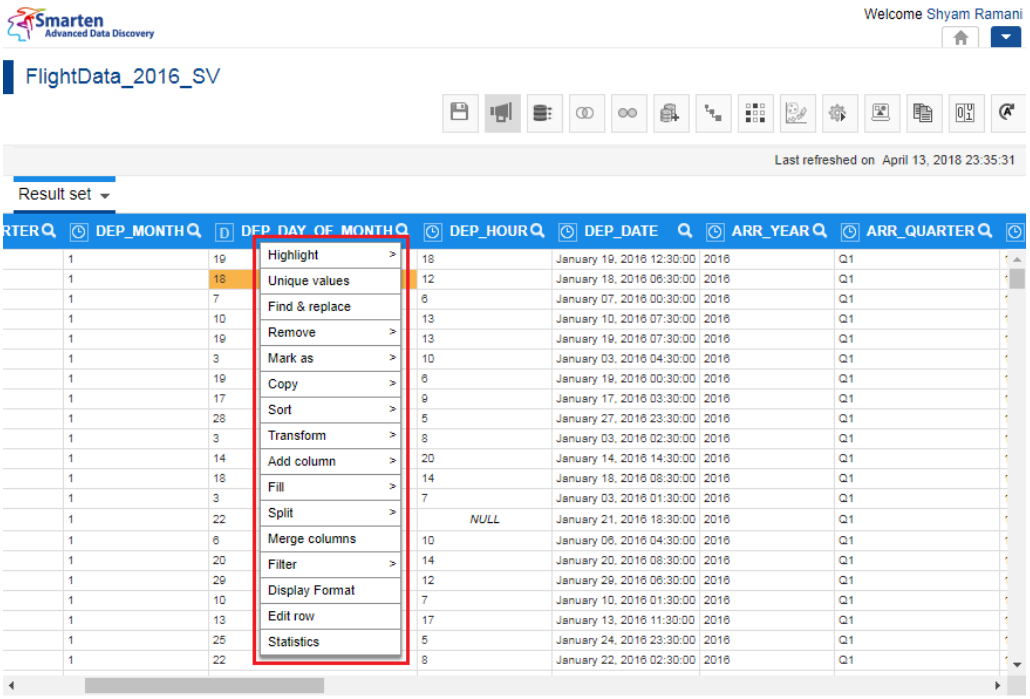
About this task

Use this task to split data in a particular column within a dataset.

Procedure

1. Open the dataset from which you want to split data.
2. Right-click in the column from whose data you want to split.

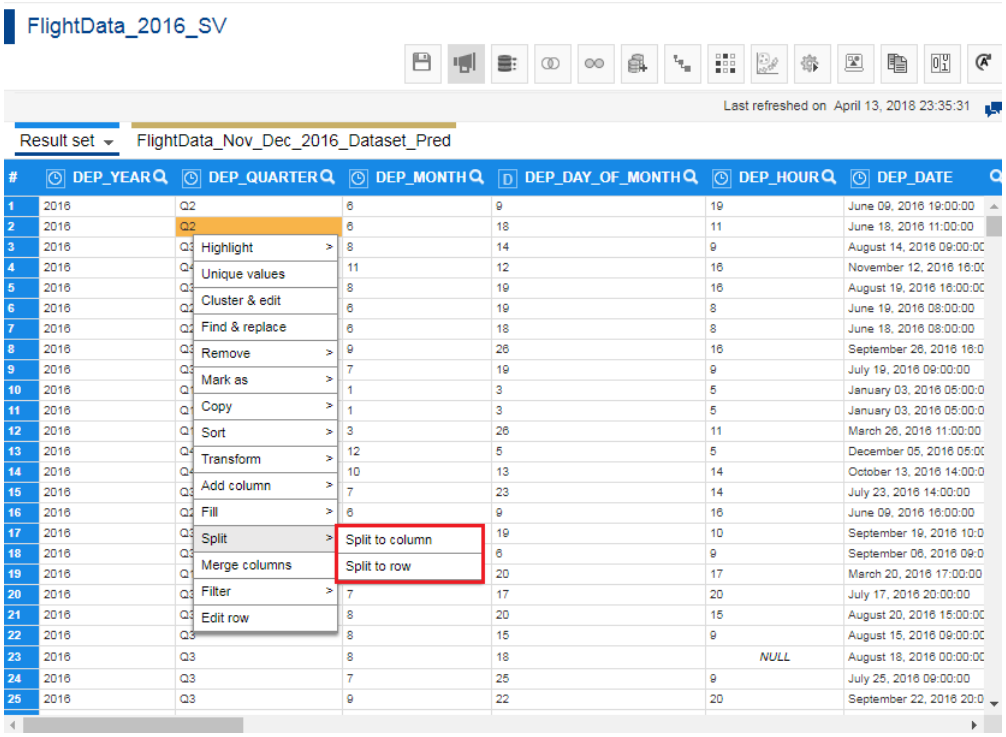
   The system displays the context menu.



SPLIT DATA—THE CONTEXT MENU

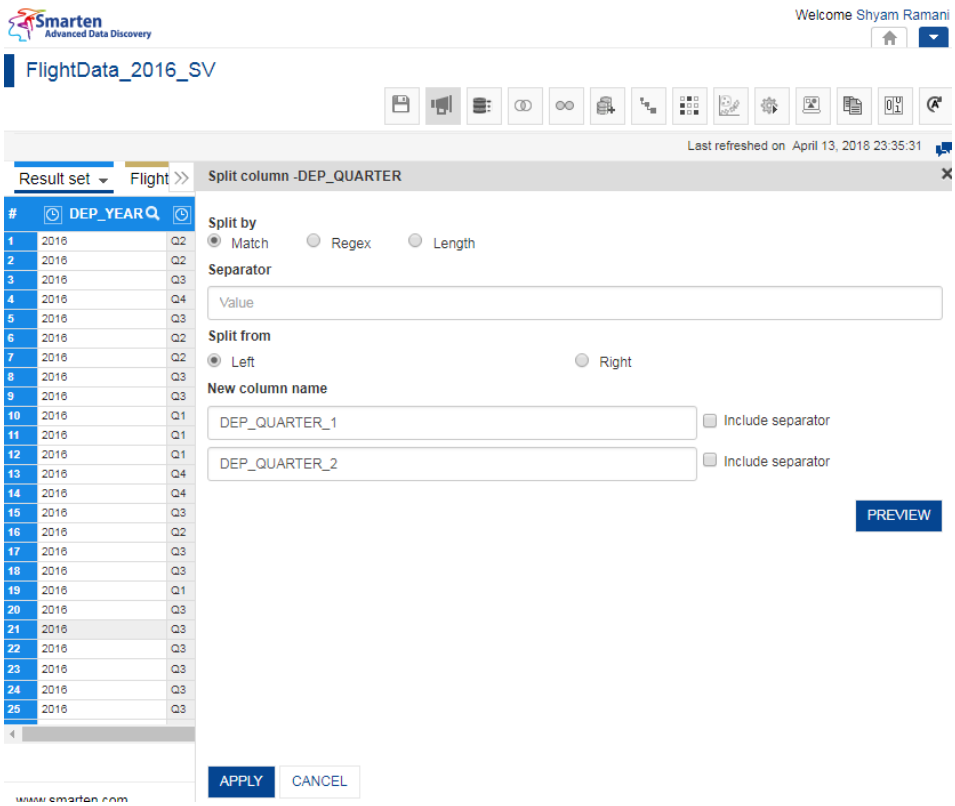
3. Click **Split** from the menu.

   The system displays the options available to split data.



SPLIT DATA—OPTIONS AVAILABLE TO SPLIT DATA

4. You can click **Split to column** to split data of the selected columns into two columns. The system displays the **Split column** dialog box.

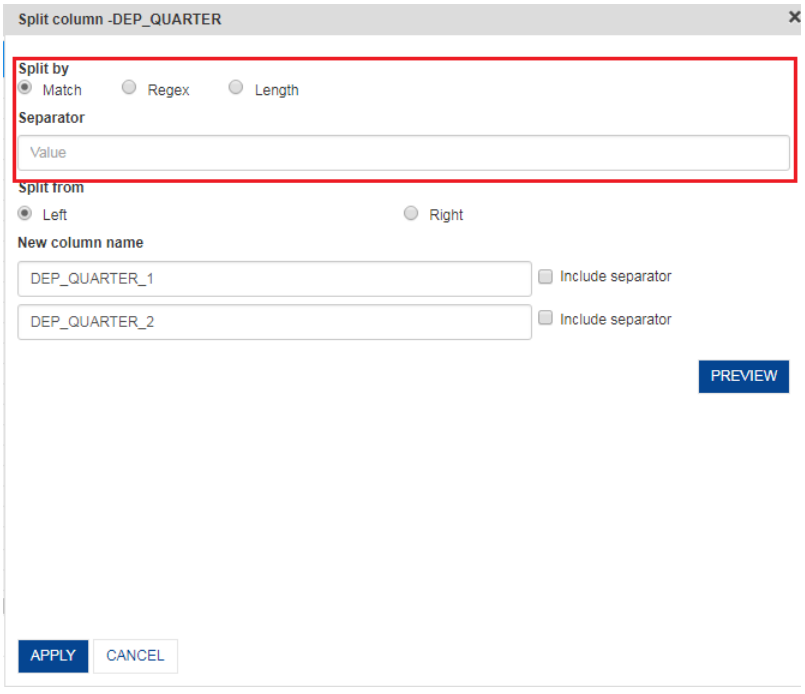


SPLIT DATA—THE SPLIT COLUMN DIALOG BOX

5. Select an option for **Split by** to select an option based on which you want to split the data.

   The following options are available:

   - **Match**: Select this option to split the data based on a separator. For example, colon, semicolon, and comma.
   - **Regex**: Select this option to split the data based on an expression.
   - **Length**: Select this option to split the data based on the length specified.

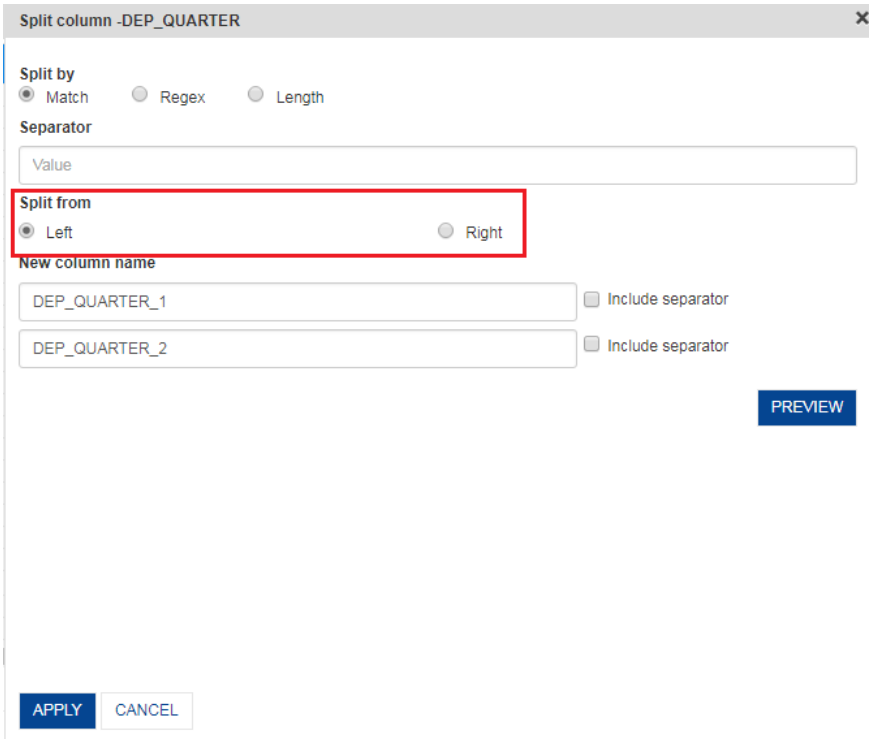


SPLIT DATA—THE SPLIT BY OPTIONS

6. Based on the option you have selected, specify a value for separator, expression, or length in the box.
7. Select an option to specify from where you want to split the data.

   The following options are available:

   - **Left**: Select this option to split the data from the left.
   - **Right**: Select this option to split the data from the right.



SPLIT DATA—THE SPLIT FROM OPTIONS

8. Specify the name of the new column names in the **New column name** boxes.
9. You can select the **Include separator** options **to include t**he separator in the new co**lumns.**
10. Click **APPLY**.

7.2.15.2 Splitting Data into Rows

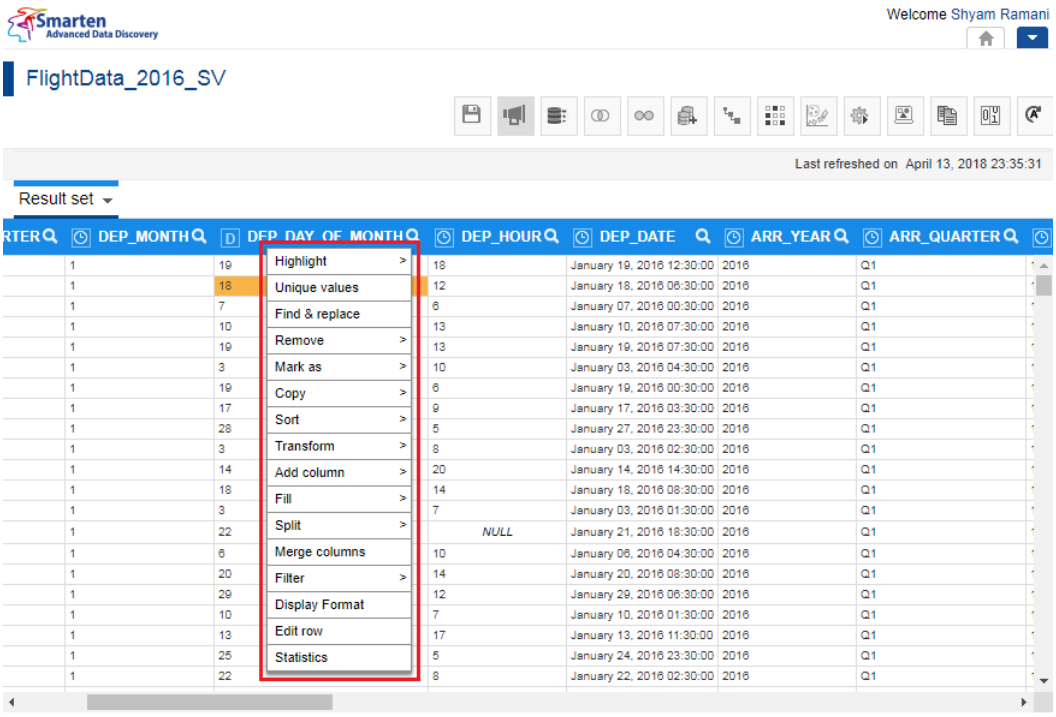
About this task

Use this task to split data in a particular column within a dataset.

Procedure

1. Open the dataset from which you want to split data.
2. Right-click in the column from whose data you want to split.

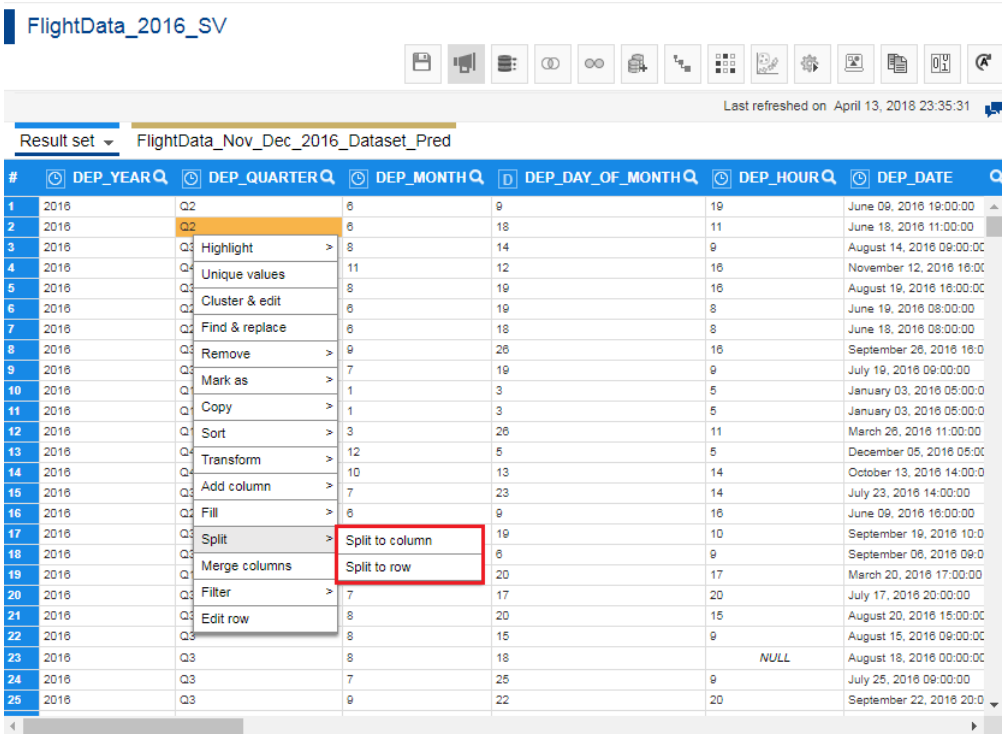
   The system displays the context menu.



SPLIT DATA—THE CONTEXT MENU

3. Click **Split** from the menu.

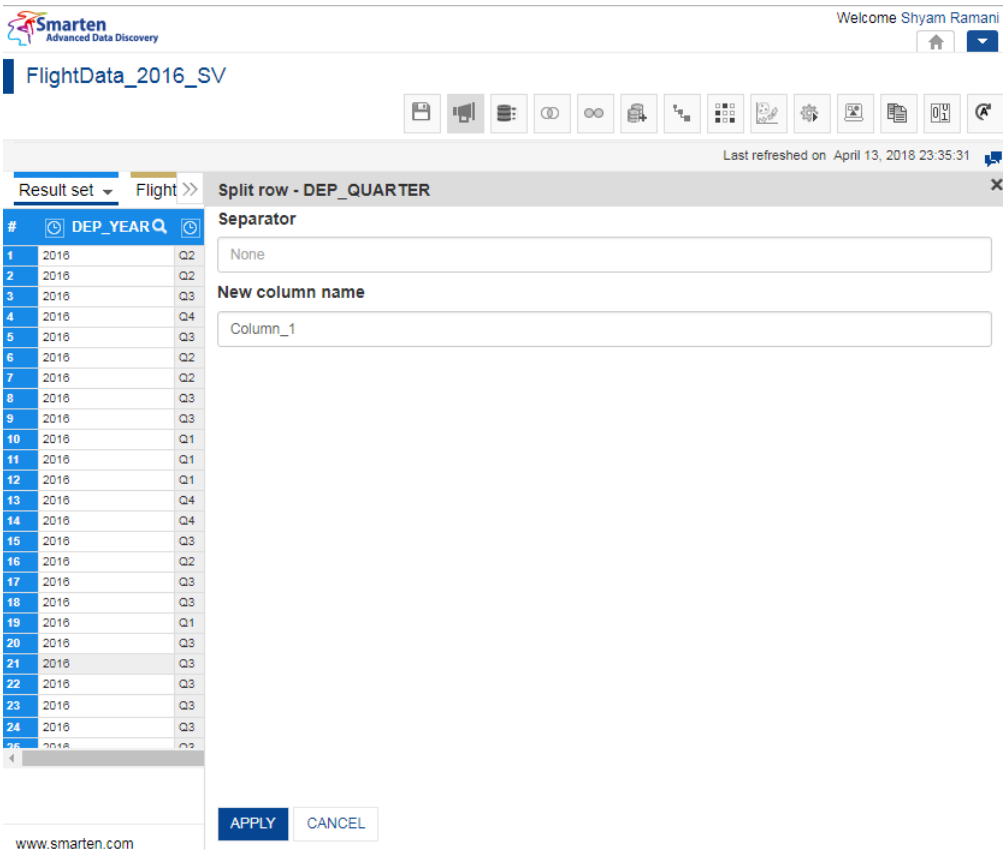
   The system displays the options available to split data.



SPLIT DATA—OPTIONS AVAILABLE TO SPLIT DATA

4. You can click the **Split to row** to split the value of a column into one or more rows.

   The system displays the **Split row** dialog box.



SPLIT DATA—THE SPLIT ROW DIALOG BOX

5. Specify the separator in the **Separator** box based on which you want to split the data.
6. Specify a name for the new column in the **New column name** box.
7. Click **APPLY**.

7.2.16 Merging Columns in a Dataset

You can merge data from multiple columns into a single column.

Reference: **Concept Manual > Shape Data > Merge column**

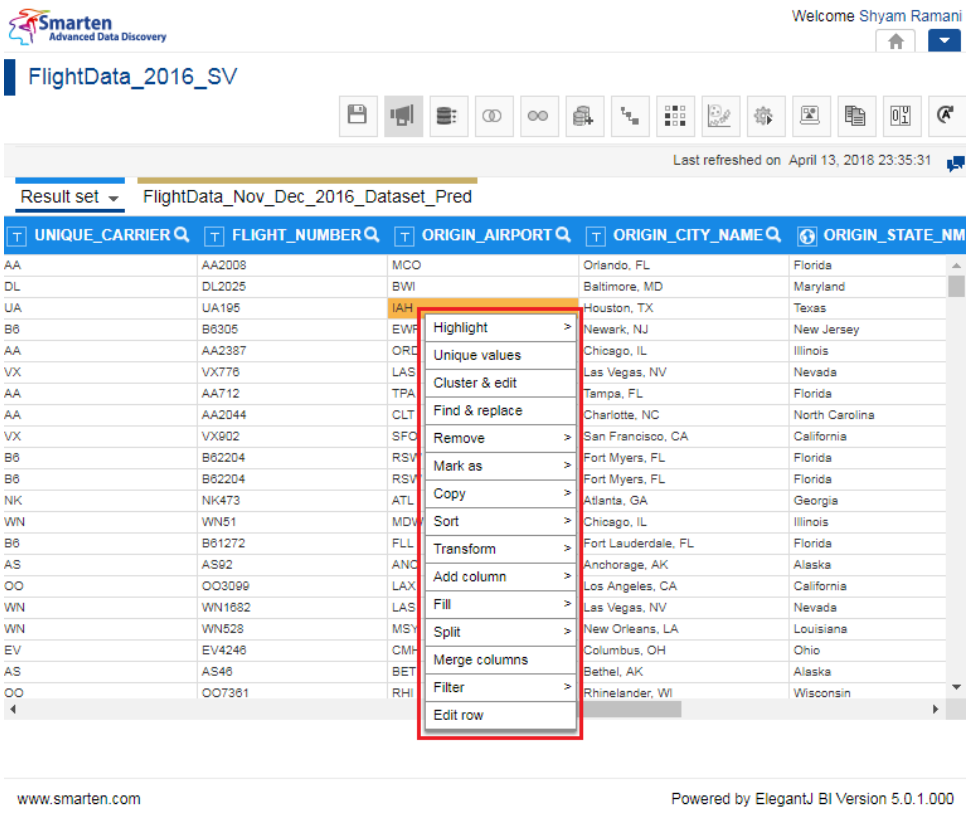
About this task

Use this task to merge data into a single column.

Procedure

1. Open the dataset in which you want to merge data.
2. Right-click in the column in which you want to merge data.

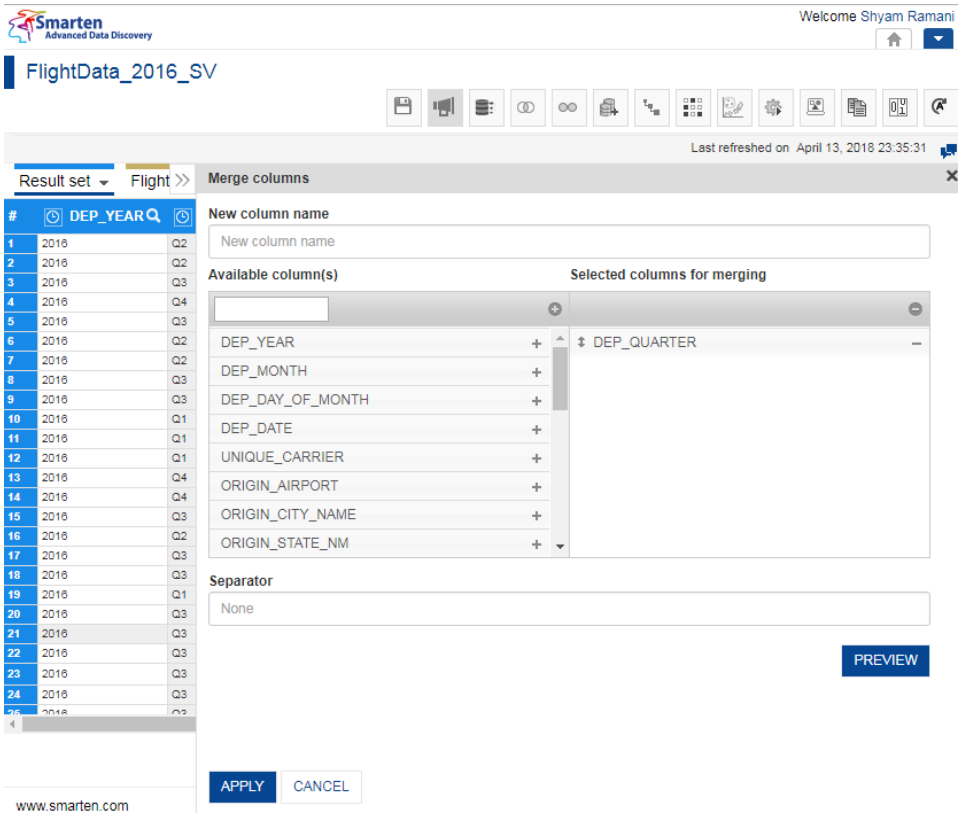
   The system displays the context menu.



MERGE COLUMNS—THE CONTEXT MENU

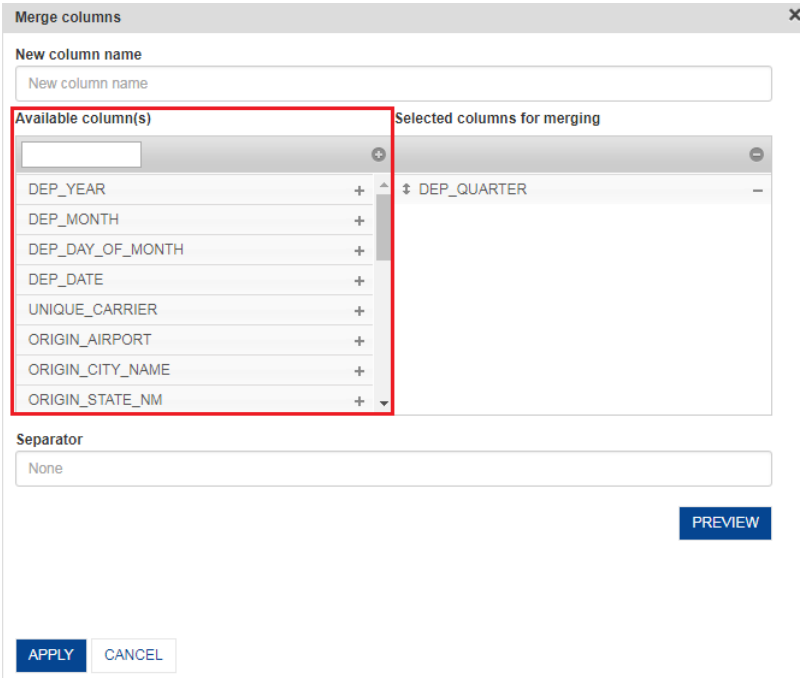
3. Click **Merge columns** from the menu.

The system displays the **Merge columns** dialog box.



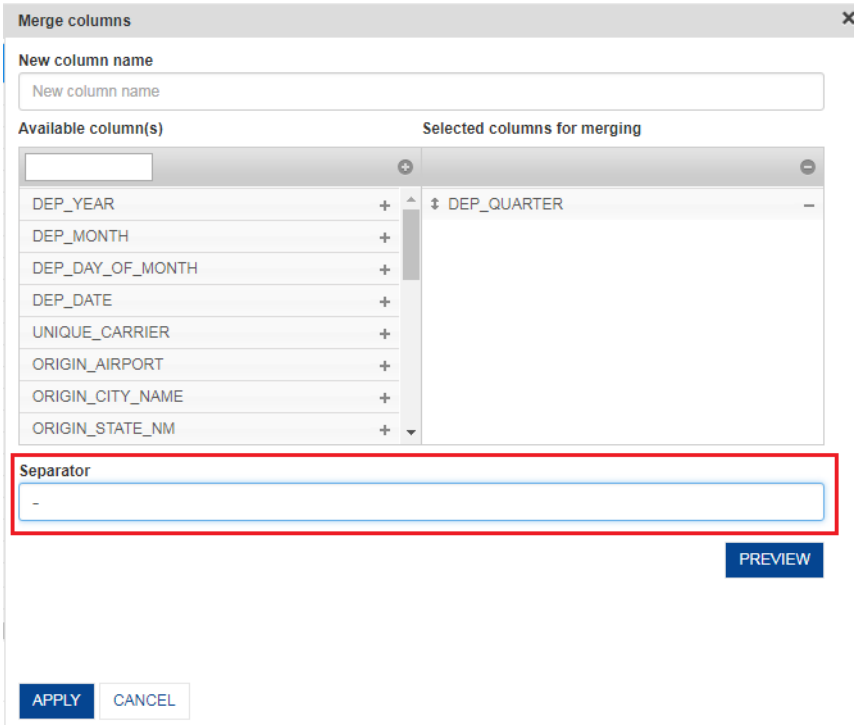
MERGE COLUMNS—THE MERGE COLUMNS DIALOG BOX

4. Specify a name for the new column in the **New column name** box.
5. Click the plus sign adjacent to a column from the **Available column(s)** section.



MERGE COLUMNS—THE LIST OF COLUMNS AVAILABLE FOR MERGING

6. The selected column is now available within the **Selected columns for merging** section.
7. Specify a separator that you want to be used in the merged data.



MERGE COLUMNS—OPTION TO SPECIFY SEPARATOR

8. You can click **PREVIEW** to view a preview of the merged data.
9. Click **APPLY**.

7.2.17 Filtering Data in a Dataset

You can filter data for a particular column value, duplicate rows, or rows with all null and zero values.

Reference: **Concept Manual > Explore Data > Filter**

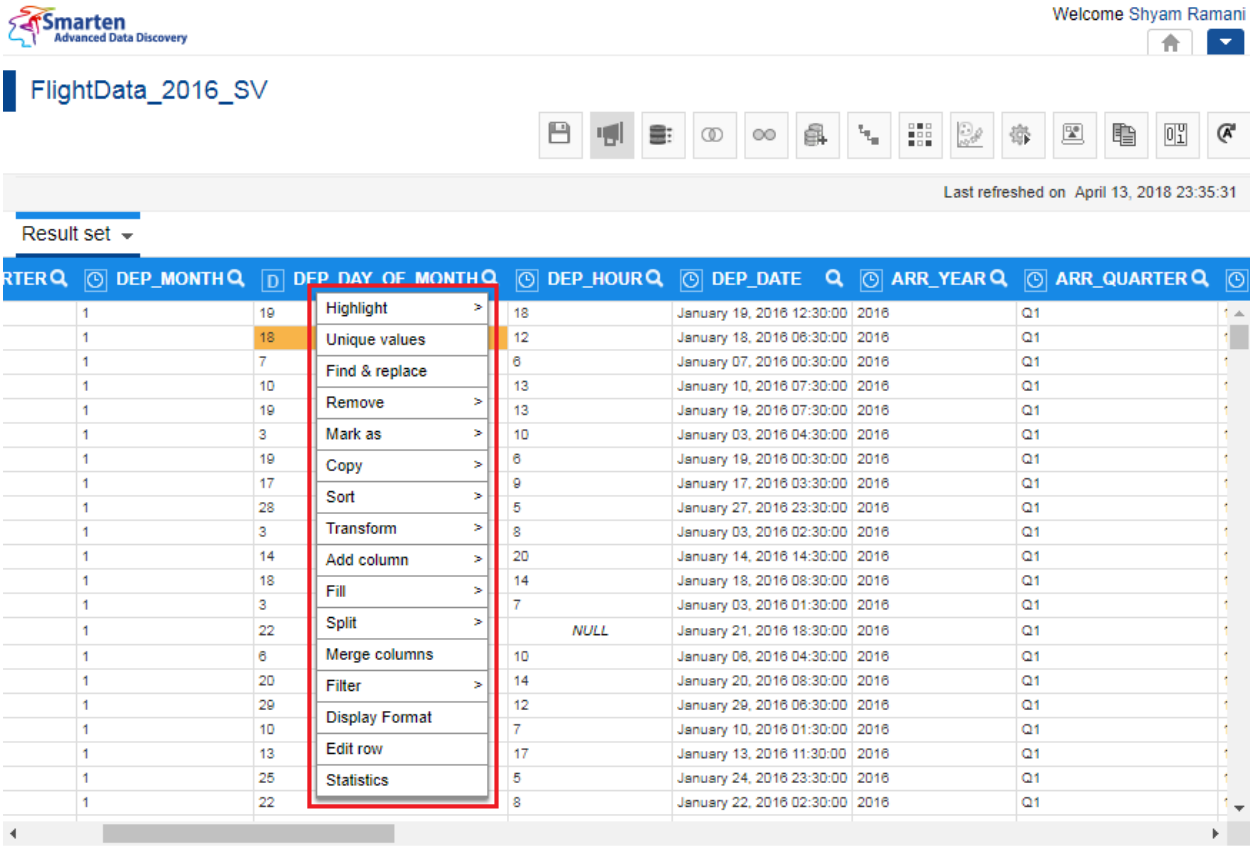
About this task

Use this task to filter data in a dataset.

Procedure

1. Open the dataset in which you want to filter data.
2. Right-click in the column for which you want to filter data.

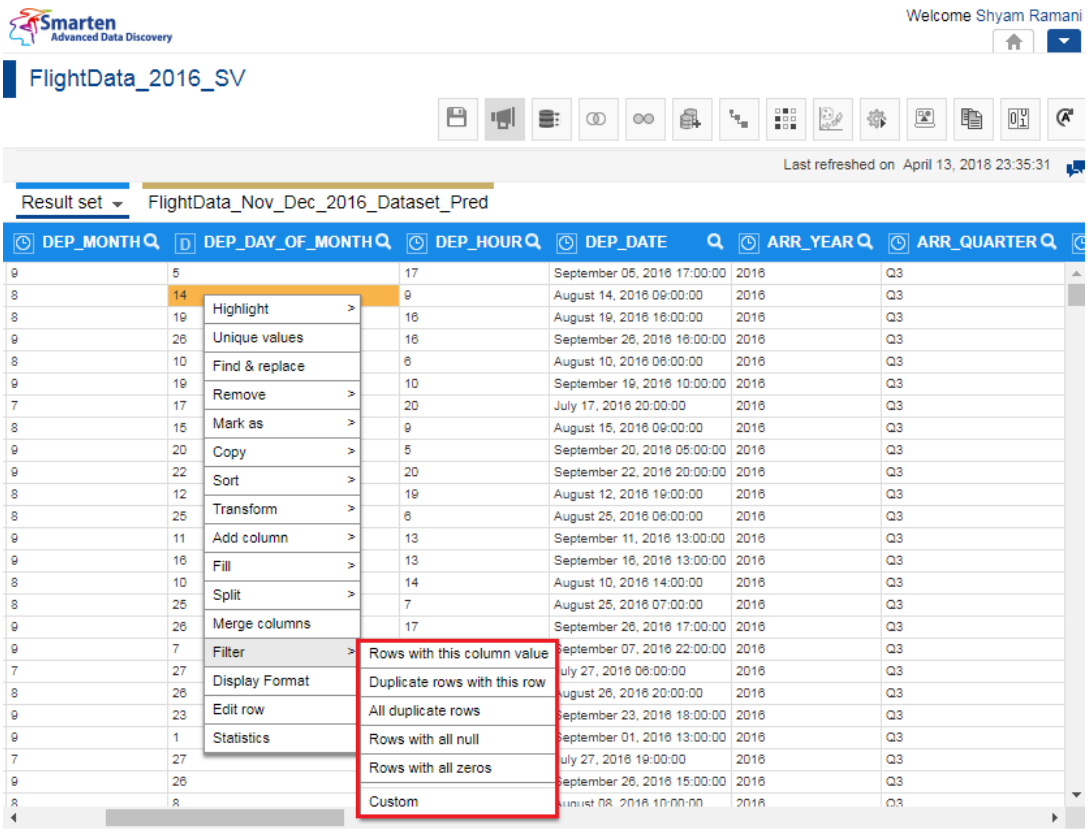
   The system displays the context menu.



FILTER DATA—THE CONTEXT MENU

3. Click **Filter** from the menu.

   The system displays the options available to filter data from the dataset.



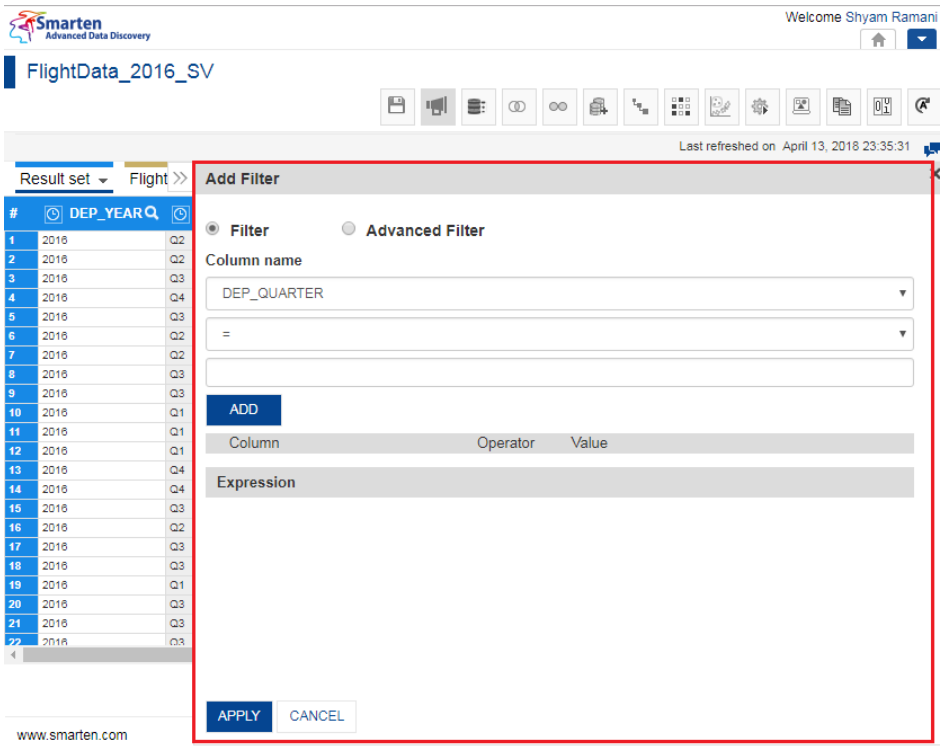
FILTER DATA—OPTIONS AVAILABLE FOR FILTERING DATA

4. Click any of the options in the following table to apply that operation.

Option	Description
Rows with this column value	Use this option to find and display the rows that contain the value available in the selected cell.
Duplicate rows with this row	Use this option to find and display the rows that contain exactly the same data as the selected row.
All duplicate rows	Use this option to find and display all duplicate rows. The duplicate rows are the set of rows that contain exactly the same data.
Rows with all null	Use this option to find and display all rows that have a null value in all columns.
Rows with all zeros	Use this option to find and display all rows that have zeros in all numeric columns.
Custom	Use this option to filter data based on an expression that is created using more than one column.

5. You can click **Custom** to filter data based on an expression.

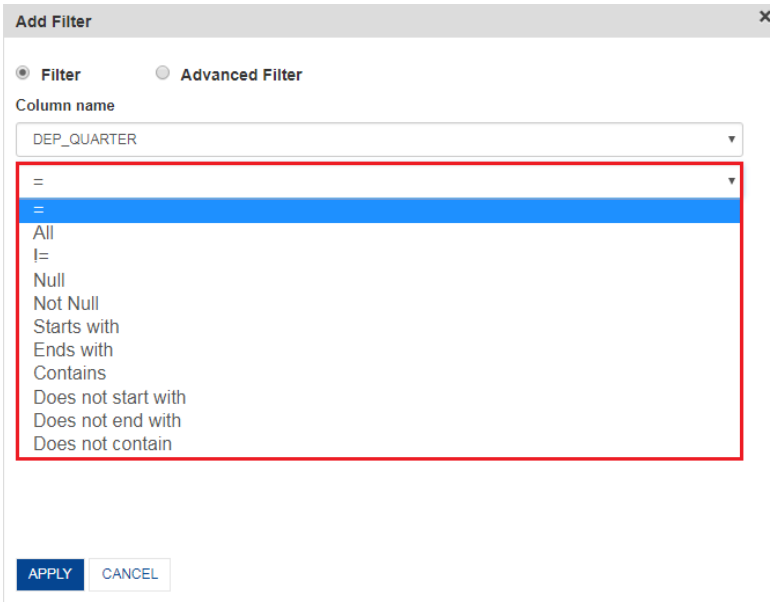
   The system displays the **Add Filter** dialog box. By default, the **Filter** option is selected.



FILTER DATA—THE ADD FILTER DIALOG BOX

6. Select a column that you want to use to build an expression from the **Column name** list.
7. Select an operator from the list.

> Note:
> The operations available in the list depend on the data type of the column you have selected in the above step.



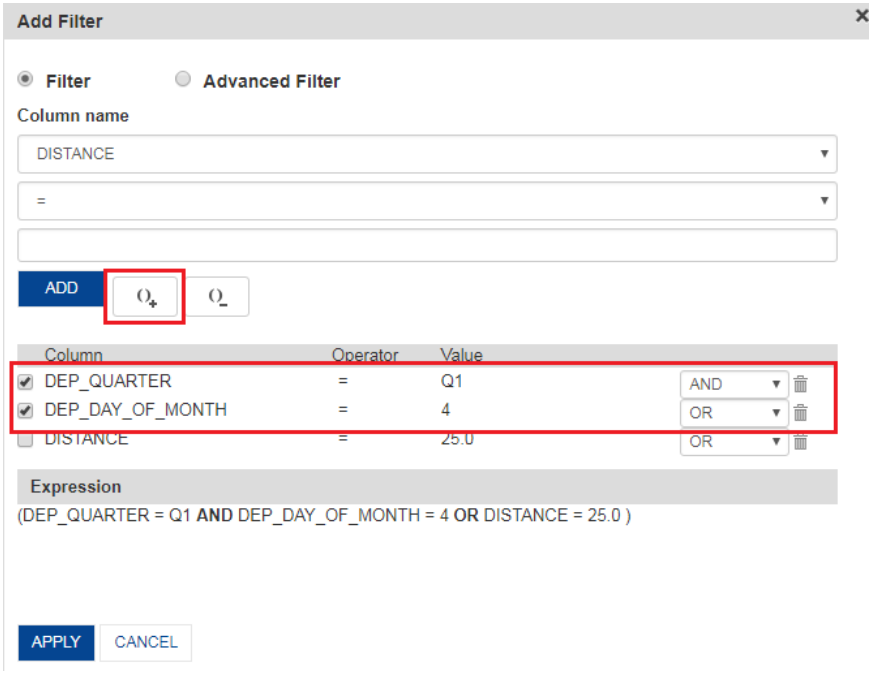
FILTER DATA—BUILDING AN EXPRESSION

8. Select a value from the list.

   The list provides values that are available for the column you have selected from the **Column name** list. You can select multiple values for this field.

9. Click **ADD**.
   The system adds the filter based on the options you have selected.
10. You can perform the above steps to add more than one filter.
11. You can combine more than one filter in a group:

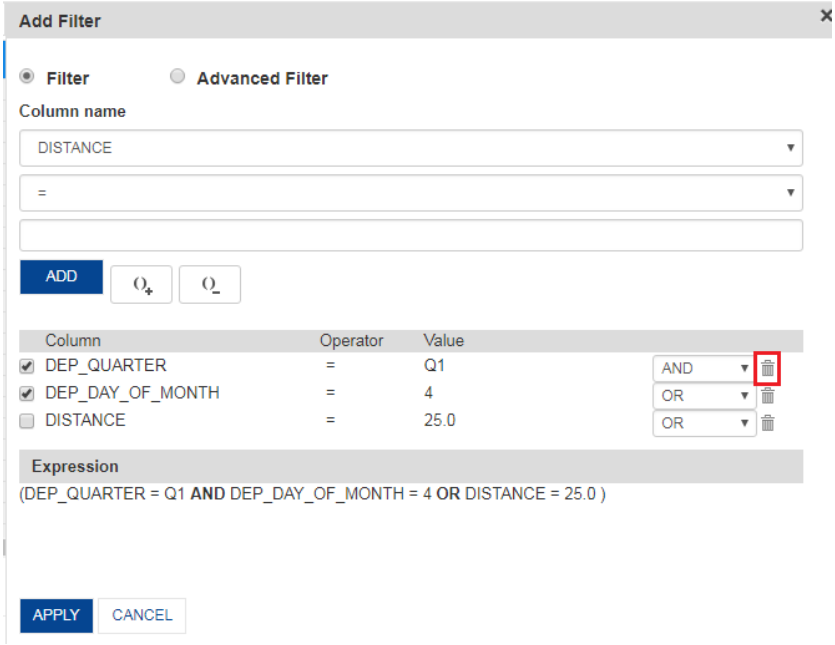
    a) Select the check box adjacent to the filters you want to group.
    b) The system displays the option to group and ungroup.



FILTER DATA—OPTION TO GROUP FILTERS

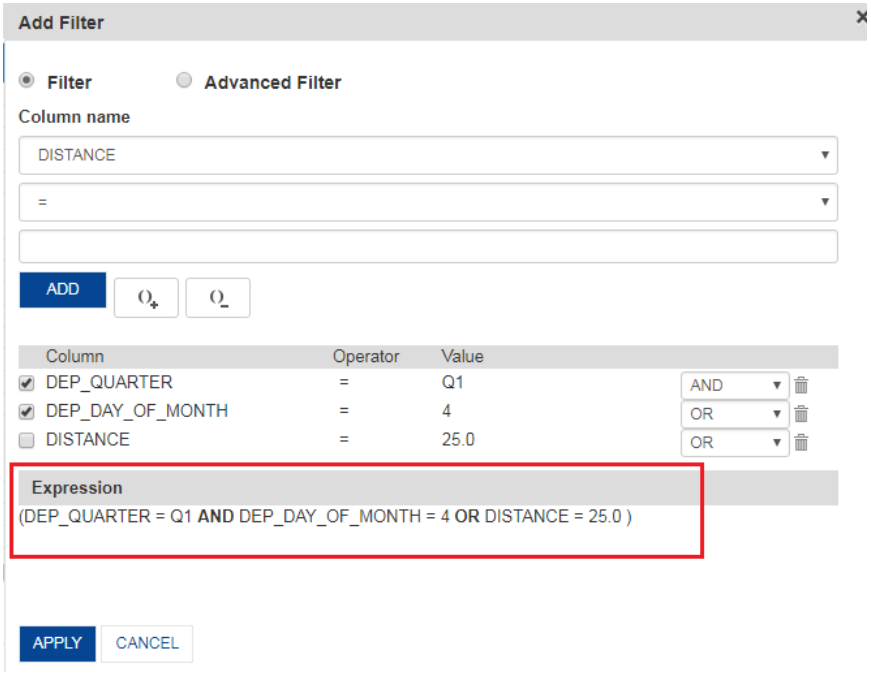
    c) Similarly, you can select a group of filters and click ungroup to ungroup those filters.

12. You can select an option from the list adjacent to a filter to apply **AND** or **OR** operator.
    The operation is applied between the current filter and the next filter.
13. You can click the Delete icon adjacent to a filter to delete that filter.



FILTER DATA—DELETING A FILTER

14. The resultant expression is displayed within the **Expression** section.



FILTER DATA—THE RESULTANT EXPRESSION TO FILTER DATA

15. Click **APPLY**.

7.2.17.1 Filtering Data in a Dataset Using the Advanced Option

You can use advanced options to filter data in a dataset.

Reference: **Concept Manual > Explore Data > Filter**

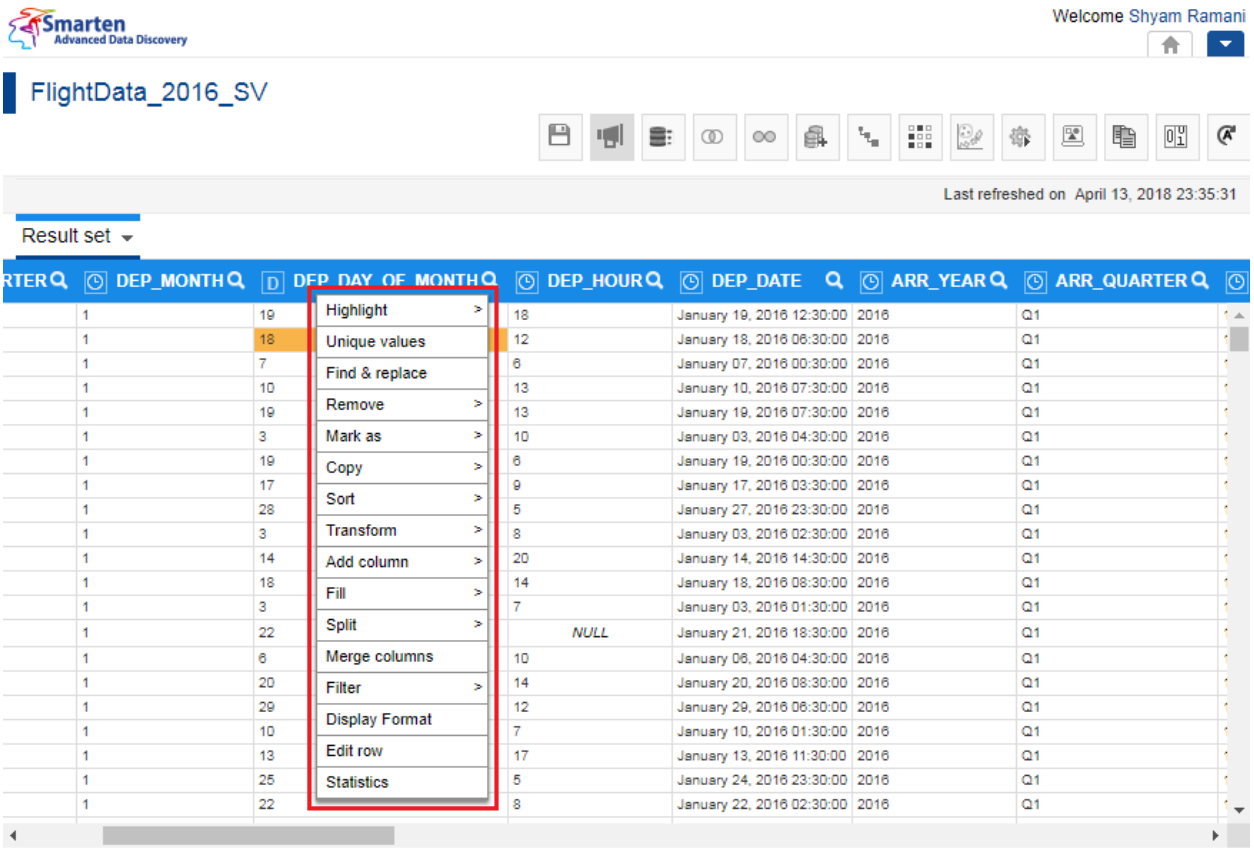
About this task

Use this task to filter data in a dataset using advanced options.

Procedure

1. Open the dataset in which you want to filter data.
2. Right-click in the column for which you want to filter data.

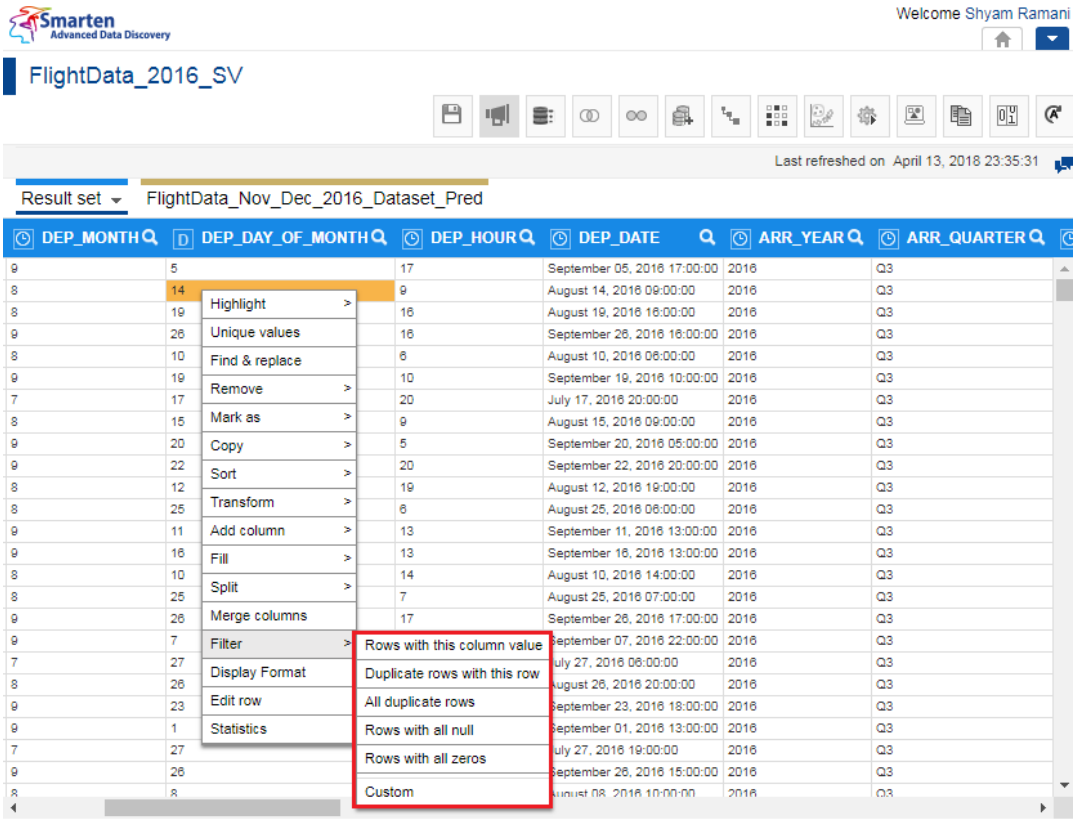
   The system displays the context menu.



FILTER DATA—THE CONTEXT MENU

3. Click **Filter** from the menu.

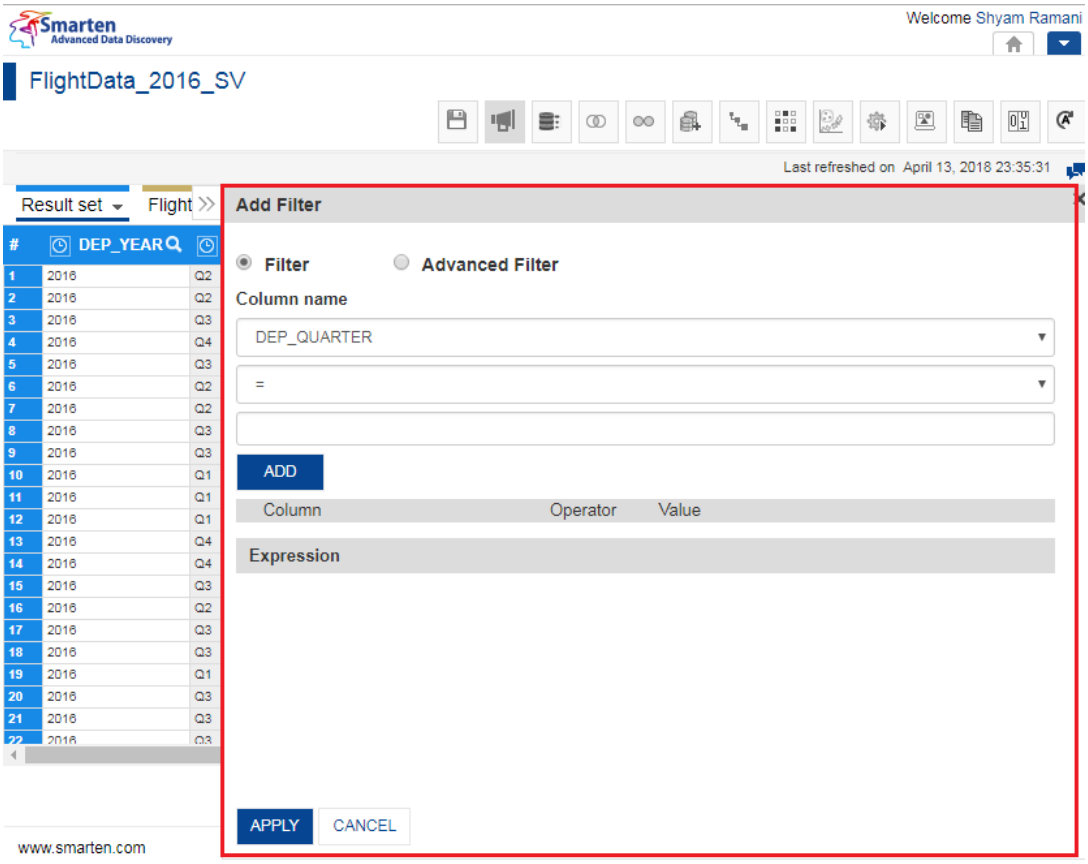
   The system displays the options available to filter data from the dataset.



FILTER DATA—OPTIONS AVAILABLE FOR FILTERING DATA

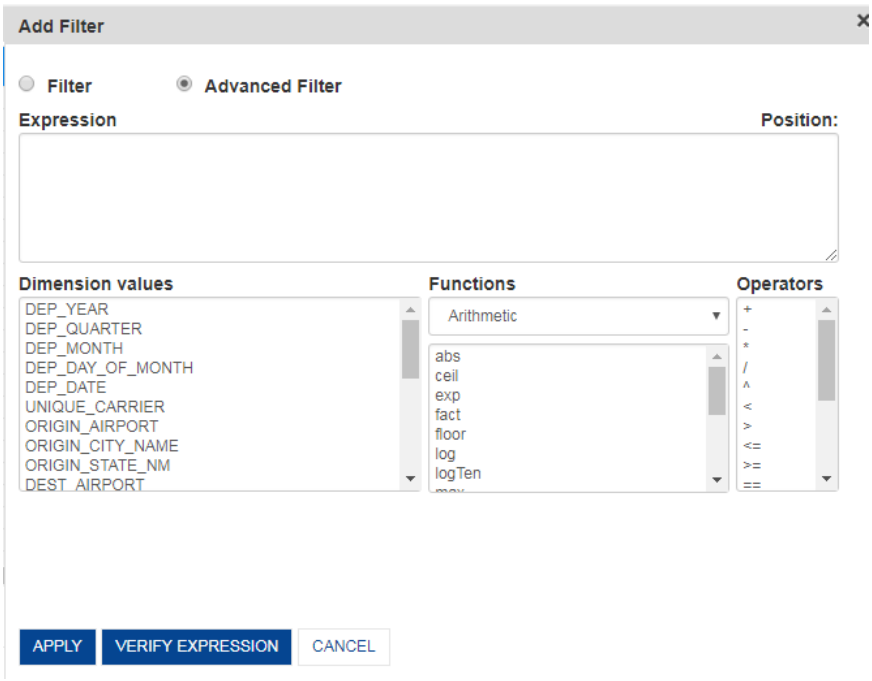
4. Click **Custom**.

   The system displays the **Add Filter** dialog box. By default, the **Filter** option is selected.



FILTER DATA—THE ADD FILTER DIALOG BOX

5. Select the **Advanced Filter** option.



FILTER DATA—BUILDING EXPRESSION FOR ADVANCED FILTER

6. Double-click the name of the column you want to use in the expression from the **Columns** section.
7. Select an option from the Functions list.

   The following options are available:
   - Arithmetic
   - Date
   - Miscellaneous
   - String
8. Select an operation.
   The operations available are based on the option you have selected in the above step.
9. Select an operator you want to use from the **Operators** section.
10. Click **APPLY**.

You can refer to the tables below for more information about the operations available for arithmetic, string, date, and miscellaneous functions.

Arithmetic functions:

Function	Description	Example
abs(number)	Return absolute value of a number, a number without its sign Argument 1: The number for which absolute value is required Returns: A number	Argument 1 = 32 Returns 32 Argument 1 = 67.98 Returns 67.98 Argument 1 = -23 Returns 23
ceil(d)	Returns the smallest whole number that is greater than or equal to a specified number Argument 1: The number that has to be rounded up Returns: A number	Argument 1 = 26 Returns 26 Argument 1 = 26.7 Returns 27 Argument 1 = -26.7 Returns -26
exp(d)	Returns the exponential value of a number Argument 1: The exponent applied to base e Returns: A number	Argument 1 = 1145 Returns “Infinity” Argument 1 = 12 Returns 162754.79 Argument 1 = -25 Returns 0.00
fact(i)	Returns the factorial of a number Argument 1: The number for which factorial is to be calculated Returns: A number	Argument 1 = 7 Returns 5040 Argument 1 = -5 Returns NULL
floor(d)	Returns the largest whole number that is smaller than or equal to a specified number Argument 1: The number to be rounded down Returns: A number	Argument 1 = 26 Returns 26 Argument 1 = 26.7 Returns 26 Argument 1 = -26.7 Returns -27

log(d)	Returns natural logarithm (base e) of a number Argument 1: A value greater than 0 for which logarithm is to be calculated Returns: A number	Argument 1 = 551 Returns 6.31 Argument 1 = -551 Returns NULL Argument 1 = 551.45 Returns 6.31
logTen(d)	Returns decimal logarithm (base 10) of a number Argument 1: The value greater than 0 for which logarithm is to be calculated Returns: A number	Argument 1 = 551 Returns 2.74 Argument 1 = -551 Returns NULL Argument 1 = 551.45 Returns 2.74
max(number, number)	Returns larger of two numbers Argument 1: First number to find out if it is larger than the second number Argument 2: Second number to find out if it is larger than the first number Returns: A number	Argument 1 = 198 Argument 2 = 1660 Returns 1660.00 Argument 1 = 198 Argument 2 = -1660 Returns 198.00
min(number, number)	Returns smaller of two numbers Argument 1: First number to find out if it is smaller than the second number Argument 2: Second number to find out if it is smaller than the first number Returns: A number	Argument 1 = 198 Argument 2 = 1660 Returns 198.00 Argument 1 = 198 Argument 2 = -1660 Returns -1660.00
mod(number, number)	Returns modulus of two numbers Argument 1: Dividend: The number to be divided Argument 2: Divisor: The number by which the dividend has to be divided Returns: A number	Argument 1 = 460 Argument 2 = 72 Returns 28.00 Argument 1 = -460 Argument 2 = 72 Returns 44.00 Argument 1 = 460 Argument 2 = -72 Returns 28.00 Argument 1 = -460 Argument 2 = -72 Returns -28.00
pi(d)	Returns pi times a number Argument 1: The number Returns: A number	Argument 1 = 641 Returns 2013.76 Argument 1 = -3 Returns -9.42
random(number, number)	Returns a random number between two specified numbers Argument 1: The smallest integer value Argument 2: The largest integer value	Argument 1 = 54 Argument 2 = 55 Returns 54.45/54.51 /54.95…

	Returns: A number	Argument 1 = 72 Argument 2 = 80 Returns 72.89/ 73.94/75.20/76.47..
		Argument 1 = 20 Argument 2 = -10 Returns -7.68/-9.75/-2.65/5.97
round(d, i)	Returns a number rounded to a specified number of decimal places Argument 1: The number to be rounded Argument 2: The number of places to which the number is to be rounded Returns: A number	Argument 1 = 12.356 Argument 2 = 1 Returns 12.40
		Argument 1 = -12.356 Argument 2 = 1 Returns -12.40
		Argument 1 = 12.356 Argument 2 = 2 Returns 12.36
		Argument 1 = 12.356 Argument 2 = 3 Returns 12.35
sign(d)	Returns a number (-1, 0, or 1) indicating the sign of a number Argument 1: The number for which the algebraic sign is to be determined Returns: A number	Argument 1 = -5 Returns -1
		Argument 1 = 0 Returns 0
		Argument 1 = 29 Returns 1
sqrt(d)	Returns the square root of a number Argument 1: A positive value for which the square root is to be calculated Returns: A number	Argument 1 = 100 Returns 10.00
		Argument 1 = 588 Returns 24.24
		Argument 1 = -588 Returns NaN (Not a number)
truncate(d, i)	Returns a number truncated to a specified number of decimal places Argument 1: The number to be truncated Argument 2: The scale of the truncation Returns: A number	Argument 1 = 10.54 Argument 2 = 1 Returns 10.50
		Argument 1 = 10.54 Argument 2 = 2 Returns 10.54
		Argument 1 = 10.54 Argument 2 = 0 Returns 11.00
		Argument 1 = 10.25 Argument 2 = 0 Returns 10.00

		Argument 1 = -10.54 Argument 2 = 1 Returns -10.50

String functions:

Function	Description	Examples
asc(c)	Returns ASCII value of a character Argument 1: The character for which the ASCII value is to be returned Returns: A number	Argument 1 = A Returns 65
		Argument 1 = a Returns 97
		Argument 1 = "1" Returns 49
booleanValue ("string")	Returns the content of a string as a boolean Argument 1: The string from which boolean is to be returned Returns: A boolean	Argument 1 = "True" Returns true
		Argument 1 = "TRUE" Returns true
		Argument 1 = "abc" Returns false
		Argument 1 = "748" Returns false
byteValue(object)	Returns the content of a string as a byte Argument 1: The object from which byte is to be returned Returns: A byte	Argument 1 = "N787AA" Returns 0
		Argument 1 = "-128" Returns -128
		Argument 1 = "-129" Returns 0
		Argument 1 = "127" Returns 127
		Argument 1 = "128" Returns 0
		Argument 1 = "120.54" Returns 120
charValue(i)	Returns the content of an integer as a character Argument 1: The number from which character is to be returned Returns: A character	Argument 1 = 65 Returns "A"
		Argument 1 = 97 Returns "a"
		Argument 1 = 49 Returns "1"
doubleValue(object)	Returns the content of a string as double Argument 1: The object from which double is to be returned Returns: A double	Argument 1 = "748" Returns 748.00
		Argument 1 = "748.52" Returns 748.52
		Argument 1 = "-748.52" Returns -748.52

		Argument 1 = “abc” Returns 0.00
		Argument 1 = “ABC” Returns 0.00
fill(“string”, i)	Returns a string of specified length filled with occurrences of a specified string Argument 1: The string that has to be filled Argument 2: The length of the filled string Returns: A string	Argument 1 = “N787AA” Argument 2 = 2 Returns “N7”
		Argument 1 = “N787AA” Argument 2 = 6 Returns “N787AA”
		Argument 1 = “N787AA” Argument 2 = 9 Returns “N787AAN78”
floatValue(object)	Returns the content of a string as a float Argument 1: The object from which float is to be returned Returns: A float	Argument 1 = “748” Returns 748.00
		Argument 1 = “-748.52” Returns -748.52
		Argument 1 = “abc” Returns 0.00
		Argument 1 = “ABC” Returns 0.00
indexOfChar(“string”, c, i)	Returns the starting position of a character within a specified string Argument 1: The string from which the index is to be returned Argument 2: The character to find the index Argument 3: The starting position of the string in number Returns: A number	Argument 1 = “N787AA” Argument 2 = ‘7’ Argument 3 = 1 Returns 1
		Argument 1 = “N787AA” Argument 2 = ‘7’ Argument 3 = 3 Returns 3
		Argument 1 = “N787AA” Argument 2 = ‘A’ Argument 3 = 3 Returns 4
		Argument 1 = “N787AA” Argument 2 = ‘A’ Argument 3 = 6 Returns 5
		Argument 1 = “N787AA” Argument 2 = ‘Y’ Argument 3 = 1 Returns -1
indexOfString(“string”, “string”, i)	Returns the starting position of a string within a specified string Argument 1: The string from which the index is to be returned	Argument 1 = “N208WN” Argument 2 = ‘208’ Argument 3 = 1 Returns 1

	Argument 2: The string to find index Argument 3: The starting position of the string in number Returns: A number	Argument 1 = “N208WN” Argument 2 = ‘208’ Argument 3 = 4 Returns -1
		Argument 1 = “N208WN” Argument 2 = ‘WN’ Argument 3 = 4 Returns 4
intValue(object)	Returns the contents of a string as an integer Argument 1: The object from which integer is to be returned Returns: An integer	Argument 1 = “N787AA” Returns 0
		Argument 1 = “748” Returns 748
		Argument 1 = “748.52” Returns 748
		Argument 1 = “-748” Returns -748
isDate(“string”)	Determine if the specified string contains a valid date Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = “2015-01-09” Returns “true”
		Argument 1 = “N787AA” Returns “false”
isNull(object)	Determines if the argument is NULL Argument 1: The object that is to be checked Returns: A boolean	Argument 1 = “N787AA” Returns “false”
		Argument 1 = NULL Returns “true”
isNumber(“string”)	Determines if the specified string contains a number Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = “N787AA” Returns “false”
		Argument 1 = “787” Returns “true”
isTime(“string”)	Determines if the specified string contains a valid time Argument 1: The string that is to be checked Returns: A boolean	Argument 1 = “15:30:00” Returns “true”
		Argument 1 = “N787AA” Returns “false”
left(“string”, i)	Returns a specified number of characters from a string starting with the first character Argument 1: The text from which the partial words are to be returned Argument 2: The number of characters to be extracted from the beginning of the text Returns: A string	Argument 1 = “N787AA” Argument 2 = 2 Returns “N7”
		Argument 1 = “N787AA” Argument 2 = 8 Returns “N787AA”
leftTrim(“string”)	Returns a copy of a specified string with leading blanks removed Argument 1: The text for which blank spaces are to be removed from left	Argument 1 = “87AA” Returns “87AA”
		Argument 1 = “87AA” Returns “87AA”

	Returns: A string	Argument 1 = “87AA” Returns “87AA”
length(“string”)	Returns the length of a string Argument 1: The string for which length is to be checked Returns: A number	Argument 1 = “N787AA” Returns 6
		Argument 1 = “748” Returns 3
		Argument 1 = “748.52” Returns 6
		Argument 1 = “-748.52” Returns 7
longValue(object)	Returns the content of a string as long Argument 1: The object from which long is to be returned Returns: A long	Argument 1 = “N787AA” Returns 0
		Argument 1 = “748” Returns 748
		Argument 1 = “748.52” Returns 748
		Argument 1 = “-748.52” Returns -748
match(“string”, “string”)	Returns a determination whether or not a string contains a particular pattern of characters Argument 1: The text that has to be searched in argument 2 Argument 2: The text in which the argument 1 has to be searched Returns: A number	Argument 1 = “AA” Argument 2 = “N787AA” Returns 1
		Argument 1 = “aa” Argument 2 = “N787AA” Returns 0
		Argument 1 = “AB” Argument 2 = “N787AA” Returns 0
replace(“string”, i, i, “string”)	Returns a copy of a specified string in which a specified number of characters starting with a specified character have been replaced with characters from another specified string Argument 1: The string to be processed Argument 2: Start index Argument 3: End index Argument 4: The string to be replaced Returns: A string	Argument 1 = “N208WN” Argument 2 = 1 Argument 3 = 2 Argument 4 = “3” Returns “N308WN”
		Argument 1 = “N208WN” Argument 2 = 4 Argument 3 = 6 Argument 4 = “ML” Returns “N208ML”
		Argument 1 = “N208WN” Argument 2 = 0 Argument 3 = 1 Argument 4 = “M” Returns “M208WN”
reverse(“string”)	Reverses the order or characters in a string Argument 1: The text that needs to be reversed Returns: A string	Argument 1 = “N208WN” Returns “NW802N”

right("string", i)	Returns the specified number of characters from the end of a specified string Argument 1: The text from which the specified number of characters should be returned from the end Argument 2: The number of characters to be returned from the string Returns: A string	Argument 1 = "N208WN" Argument 2 = 3 Returns "8WN"
		Argument 1 = "N208WN" Argument 2 = 8 Returns "N208WN"
rightTrim("string")	Returns a copy of a specified string with trailing blanks removed Argument 1: The text from which extra spaces have to be removed from the right Returns: A string	Argument 1 = "N208" Returns "N208"
		Argument 1 = "08WN" Returns "08WN"
		Argument 1 = "208W" Returns "208W"
shortValue(object)	Returns contents of a string as short Argument 1: The object from which short to be returned Returns: A long	Argument 1 = "N787AA" Returns 0
		Argument 1 = "748" Returns 748
		Argument 1 = "748.52" Returns 748
		Argument 1 = "-748.52" Returns -748
space(i)	Returns the string of a specified length filled with a specified number of spaces Argument 1: Number of space Returns: A string	Argument 1 = 5 Returns " "
substring("string", i, i)	Returns a string containing a character copied (starting at a specified position and ending at a specified position) from a specified string Argument 1: The text from which the characters have to be copied Argument 2: Starting position from which the characters have to be copied Argument 3: Ending position up to which the characters in the text are to be copied Returns: A string	Argument 1 = "N208WN" Argument 2 = 2 Argument 3 = 4 Returns "08"
		Argument 1 = "N208WN" Argument 2 = 2 Argument 3 = 6 Returns "08WN"
ToLowerCase("string")	Returns a copy of a specified string with all uppercase letters converted to lowercase Argument 1: The text for which the uppercase letters are to be converted to lowercase Returns: A string	Argument 1 = "N208WN" Returns "n208wn"
		Argument 1 = "N208wN" Returns "n208wn"
		Argument 1 = "n208wn" Returns "n208wn"
toString(object)	Returns a string representation of a specified object Argument 1: The object for which string is to be returned	Argument 1 = 748 Returns "748"
		Argument 1 = 748.52 Returns "748.52"

	Returns: A string	Argument 1 = -748.52 Returns “-748.52”
		Argument 1 = 16-02-2018 20:38:40 Returns “16-02-2018 20:38:40”
toUpperCase(“string”)	Returns a copy of a specified string with all lowercase letters converted to uppercase Argument 1: The text for which the lowercase letters are to be converted to uppercase Returns: A string	Argument 1 = “n208wn” Returns “N208WN”
		Argument 1 = “n208Wn” Returns “N208WN”
		Argument 1 = “N208WN” Returns “N208WN”
trim(“string”)	Returns a string with leading and trailing blanks removed Argument 1: The text from which the extra spaces are to be removed Returns: A string	Argument 1 = “08WN” Returns “08WN”
		Argument 1 = “N208” Returns “N208”
		Argument 1 = “208W” Returns “208W”

Miscellaneous functions:

Functions	Description	Examples
ifCase(condition, truevalue, falsevalue)	Returns TRUE if the condition is validated and returns FALSE if invalidated Argument 1: The condition Argument 2: True value Argument 3: False value Returns: An object	Argument 1 = origin==“LAX” Argument 2 = “Los Angeles” Argument 3 = “Others” Returns “Los Angeles” if the value of origin is “LAX” or else returns “Others”
noOfDaysByDate(Start Date, EndDate)	Returns the number of days between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-03-10 Argument 2 = 2014-04-10 Returns 32
noOfHalfYearsByDate(StartDate, EndDate)	Returns a number of half years between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2014-12-31 Returns 2
		Argument 1 = 2014-01-01 Argument 2 = 2014-05-31 Returns 0
		Argument 1 = 2014-01-01 Argument 2 = 2014-08-31 Returns 1

noOfMonthsByDate(StartDate, EndDate)	Returns the number of months between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2014-12-31 Returns 12 Argument 1 = 2014-01-01 Argument 2 = 2014-07-10 Returns 6 Argument 1 = 2014-01-01 Argument 2 = 2014-05-15 Returns 4
noOfQuartersByDate(StartDate, EndDate)	Returns a number of quarters between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2014-12-31 Returns 4 Argument 1 = 2014-01-01 Argument 2 = 2014-08-15 Returns 2
noOfWeeksByDate(StartDate, EndDate)	Returns the number of weeks between a given start and end date Argument 1: Start date Argument 2: End date Returns: A number	Argument 1 = 2014-01-01 Argument 2 = 2015-01-01 Returns 52 Argument 1 = 2014-01-01 Argument 2 = 2014-07-01 Returns 25 Argument 1 = 2014-01-01 Argument 2 = 2014-01-03 Returns 0 Argument 1 = 2014-01-01 Argument 2 = 2014-01-12 Returns 1
whenThen(columnname, whenvalue1, thenresult1, whenvalue2, then result2, ..., elseresult)	Tests values of a column or expression and returns values based on the results of the test	Argument 1 = Origin Argument 2 = “LAX” Argument 3 = “Los Angeles” Argument 4 = “JFK” Argument 5 = “John F. Kennedy” Argument 6 = “Others” Returns “Los Angeles” if the value of column “origin” is “LAX” and “John F. Kennedy” if the value is “JFK.” For rest of the values, returns “Others”

Date functions:

Function	Description	Example
date(Timestamp)	Returns the date part of a timestamp Argument 1: The timestamp for which the date has to be returned Returns: A date	Argument 1 = 2018-02-16 20:38:40 Returns 2018-02-16
dateAdd ("string", i , date)	Adds a certain date or time interval to a date Argument 1: The interval of time (where the type of interval can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The number of interval to be added to the time Argument 3: The date and time to add the interval to Returns: A date	Argument 1 = "y" or "Y" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2020-02-16 20:38:40
		Argument 1 = "m" or "M" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-04-16 20:38:40
		Argument 1 = "d" or "D" Argument 2 = 10 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-26 20:38:40
		Argument 1 = "h" or "H" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:38:40
		Argument 1 = "n" or "N" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:40:40
		Argument 1 = "s" or "S" Argument 2 = 2 Argument 3 = 2018-02-16 20:38:40 Returns 2018-02-16 22:38:42
dateDiff ("string", date, date)	Returns the number of intervals between two dates or times Argument 1: The interval of time (where the type of interval to be calculated can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The first date or time Argument 3: The second date or time Returns: A number	Argument 1 = "y" or "Y" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2016-02-16 20:30:20 Returns 2
		Argument 1 = "m" or "M" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-05-16 20:38:40 Returns -3
		Argument 1 = "d" or "D" Argument 2 = 2018-02-20 20:38:40 Argument 3 = 2018-02-16 20:38:40 Returns 4
		Argument 1 = "h" or "H" Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:38:40 Returns 10

		Argument 1 = “n” or “N” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:18:40 Returns 10
		Argument 1 = “s” or “S” Argument 2 = 2018-02-16 20:38:40 Argument 3 = 2018-02-16 10:38:10 Returns 30
datePart (“string”, date)	Returns the specified part of a given date Argument 1: The interval of time (where the part of the date can be: Year / Month / Day / Hour / Minute / Second) Argument 2: The date Returns: A number	Argument 1 = “y” or “Y” Argument 2 = 2018-02-16 20:38:40 Returns 2018
		Argument 1 = “m” or “M” Argument 2 = 2018-02-16 20:38:40 Returns 2
		Argument 1 = “d” or “D” Argument 2 = 2018-02-16 20:38:40 Returns 16
		Argument 1 = “h” or “H” Argument 2 = 2018-02-16 20:38:40 Returns 20
		Argument 1 = “n” or “N” Argument 2 = 2018-02-16 20:38:40 Returns 38
		Argument 1 = “s” or “S” Argument 2 = 2018-02-16 20:38:40 Returns 40
dateTime(“string”)	Returns contents of a string as date-time Argument 1: The string for which date-time is to be returned Returns: A date-time	Argument 1 = “2018-02-16 20:38:40” Returns 2018-02-16 20:38:40
day(date)	Returns the day of a date represented by a number (an integer between 1 and 31) Argument 1: The date or timestamp for which day part is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 16
dayName (date)	Returns the name of the day of the week Argument 1: The date or timestamp for which day of the week is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns Friday
dayofWeek(date)	Returns a number (between 1 and 7) representing the day of the week Argument 1: The date or timestamp for which day of the week is to be returned	Argument 1 = 2018-02-16 20:38:40 Returns 5

	Returns: A number	
daysAfter(date, date)	Returns the count of number of days after specified date Argument 1: The start date Argument 2: The end date Returns: A number	Argument 1 = 2018-02-16 20:38:40 Argument 2 = 2018-02-10 20:38:40 Returns 6
formatDate (date, “string”)	Returns the date format for a given pattern Argument 1: The target date Argument 2: The string (where the format can be user defined, such as “dd-mm-yy hh:mm:ss”) Returns: A date	Argument 1 = 2018-02-16 Argument 2 = “yy/mm/dd” Returns 18/02/16 Argument 1 = 2018-02-16 20:38:40 Argument 2 = “mm/dd/yyyy” Returns 02/16/2018
hour(date)	Returns the hour of a time value (an integer ranging from 0 [12:00 AM] to 23 [11:00 PM]) Argument 1: The timestamp for which hours are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 20
minute(date)	Returns the minutes of a time value (an integer ranging from 0 to 59) Argument 1: The timestamp for which minutes are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 38
month(date)	Returns the month (an integer between 1 and 12) Argument 1: The date or timestamp for which month is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 2
monthName(i, [b], [i])	Returns the month name for a given month number Argument 1: The number for month Argument 2: True if the month name is abbreviated; otherwise, False (Optional to enter. Default is False) Argument 3: The starting month of year in number (Optional to enter. Default is 1 for January) Returns: A string	Argument 1 = 1 Argument 2 = True Argument 3 = 1 Returns Jan Argument 1 = 3 Argument 2 = True Argument 3 = 4 Returns Jun Argument 1 = 9 Argument 2 = False Argument 3 = 1 Returns September Argument 1 = 2 Argument 2 = False Argument 3 = 12 Returns January

		Argument 1 = 2 Argument 2 = “” Argument 3 = “” Returns February
now()	Returns the current time Returns: A timestamp	Returns 20:38:40
relativeDate (timestamp, i)	Returns the date that occurs n days after a given date Argument 1: The date or timestamp Argument 2: The number of days to be added to the date-timestamp Returns: A timestamp	Argument 1 = 2018-02-16 20:38:40 Argument 2 = 5 Returns 2018-02-21
relativeTime (timestamp, i)	Returns the time that occurs n seconds after a given time Argument 1: The timestamp Argument 2: The number of seconds to be added to the timestamp Returns: A timestamp	Argument 1 = 20:38:40 Argument 2 = 5 Returns 20:38:45
second(timestamp)	Returns the seconds of a time value (an integer in the range 0 to 59) Argument 1: The timestamp for which seconds are to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 40
time(timestamp)	Returns the time part from a given timestamp as a string datatype Argument 1: The timestamp for which time part is to be returned Returns: A string	Argument 1 = 2018-02-16 20:38:40 Returns “20:38:40”
today()	Returns the current system date Returns: A date	Returns 2018-02-16
weekdayName(i, [b], [i])	Returns the day name for a given day number of a week Argument 1: The number for day of week Argument 2: True if the day name is abbreviated; otherwise, False (Optional to enter. Default is False) Argument 3: The first day of the week in number (Optional to enter. Default is 1 for Sunday) Returns: A string	Argument 1 = 1 Argument 2 = True Argument 3 = 1 Returns Sun Argument 1 = 1 Argument 2 = False Argument 3 = 1 Returns Sunday Argument 1 = 5 Argument 2 = False Argument 3 = 1 Returns Thursday Argument 1 = 1 Argument 2 = False Argument 3 = 3 Returns Tuesday

		Argument 1 = 1 Argument 2 = False Argument 3 = 5 Returns Thursday
		Argument 1 = 4 Argument 2 = “” Argument 3 = “” Returns Wednesday
year(date)	Returns the year corresponding to a date (an integer between 1000 and 3000) Argument 1: The date or timestamp for which year part is to be returned Returns: A number	Argument 1 = 2018-02-16 20:38:40 Returns 2018

7.2.18 Changing Display Format of Data in a Dataset

You can change the format of how data is displayed in a dataset.

> Note:
> You can change the display format of the number and DateTime data type only. This is a front-end operation and does not affect any data in a dataset.

7.2.18.1 Changing Display Format of Numeric Data in a Dataset

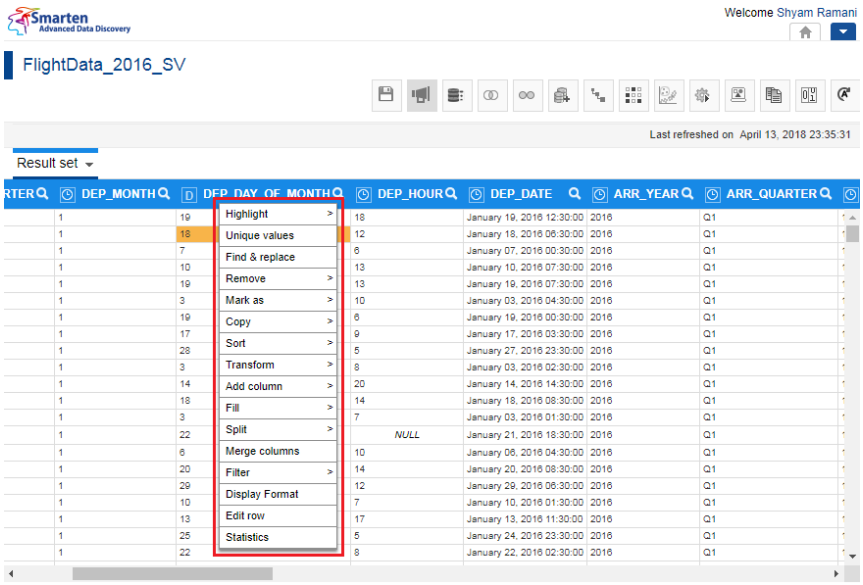
About this task

Use this task to change the display format of numeric data in a particular column.

Procedure

1. Open the dataset in which you want to change the display format.
2. Right-click in the column whose display format you want to change.

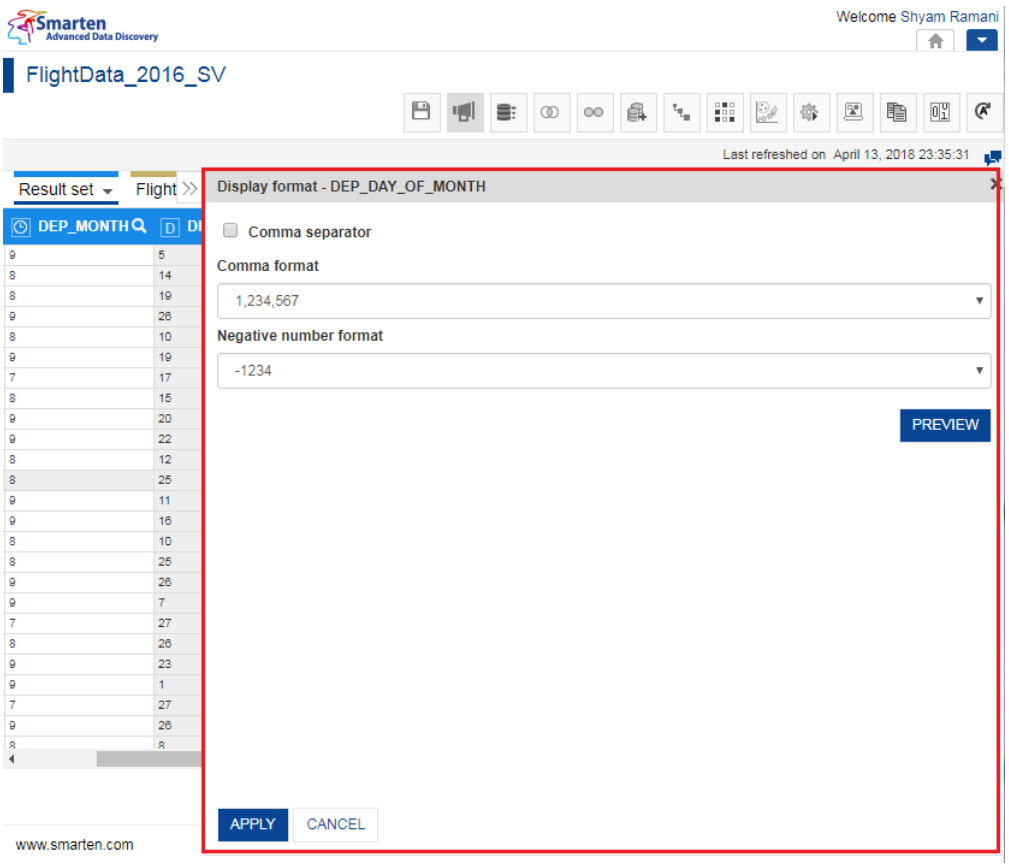
   The system displays the context menu.



DISPLAY FORMAT—THE CONTEXT MENU

3. Click **Display Format** from the menu.

   The system displays the **Display format** dialog box.



DISPLAY FORMAT—THE DISPLAY FORMAT DIALOG BOX

4. Select the Comma separator option to use comma as a separator in the format.
5. Select an option from the **Comma format** list to display the data in that format.
6. Select an option from the **Negative number format** list to display negative values in that format.
7. You can click **PREVIEW** to view a preview of the new format selected.
8. Click **APPLY**.

7.2.18.2 Changing Display Format of Datetime Data in a Dataset

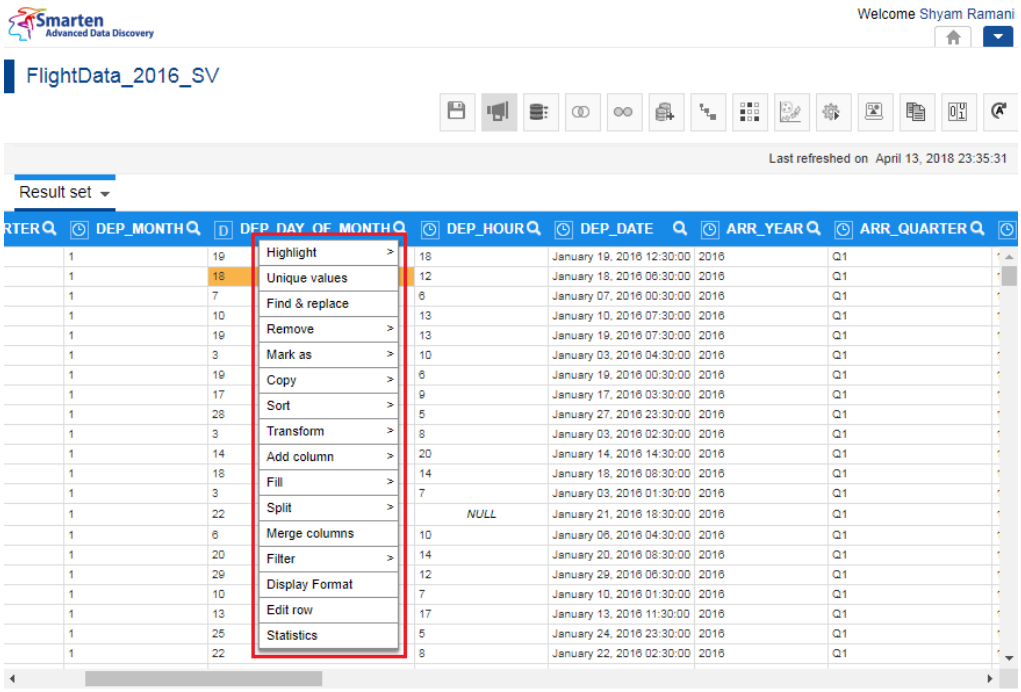
About this task

Use this task to change the display format of numeric data in a particular column.

Procedure

1. Open the dataset in which you want to change the display format.
2. Right-click in the column whose display format you want to change.

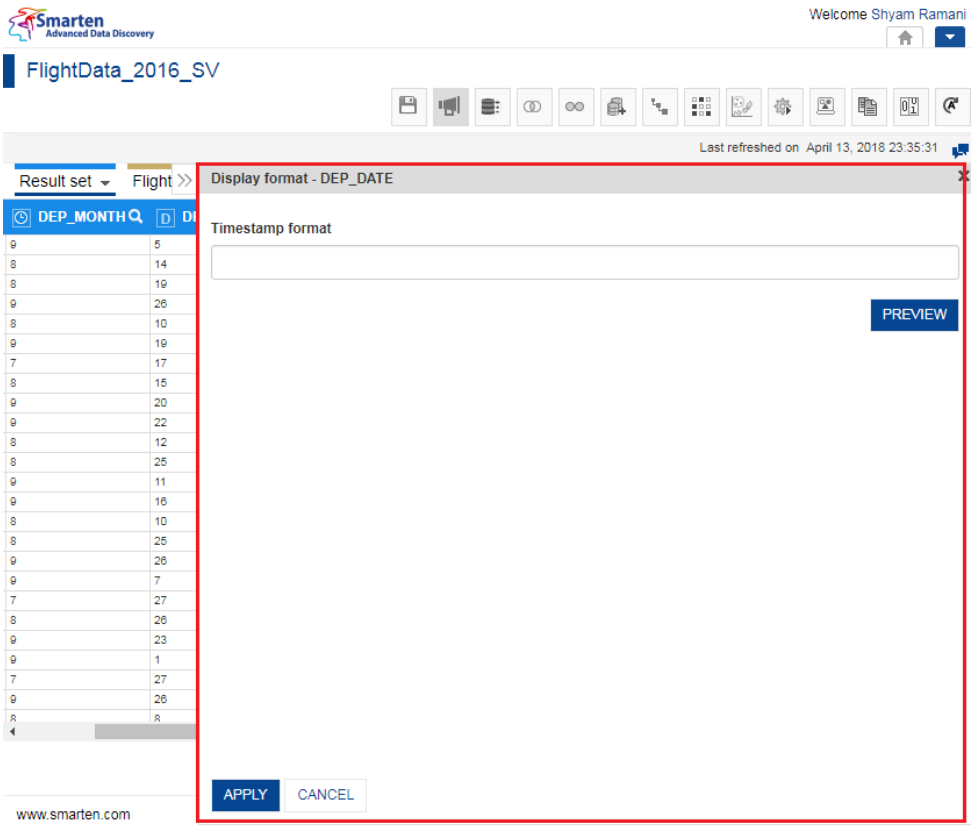
   The system displays the context menu.



DISPLAY FORMAT—THE CONTEXT MENU

3. Click **Display Format** from the menu.

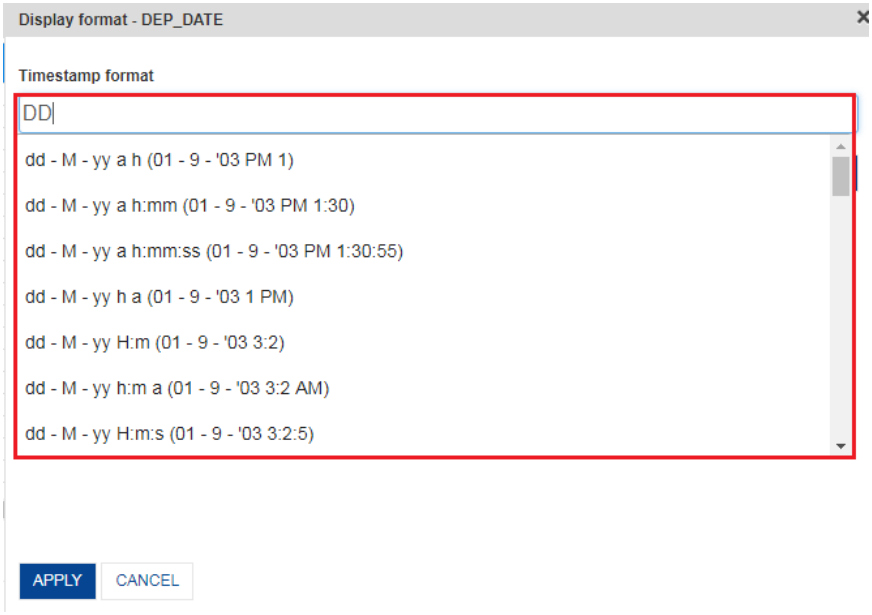
   The system displays the **Display format** dialog box.



DISPLAY FORMAT—DIALOG BOX FOR THE DATETIME DATA TYPE

4. Specify a format in the **Timestamp format** box.

As you start typing, the system displays the available formats, and you can select a format from the list.



DISPLAY FORMAT—SPECIFYING A FORMAT

5. You can click **PREVIEW** to view a preview of the selected format.
   Click **APPLY**.

7.2.19 Editing a row in a Dataset

You can change the value of one or more columns for a particular row if at least one column in the dataset has unique values.

Reference: **Concept Manual > Shape Data > Edit**

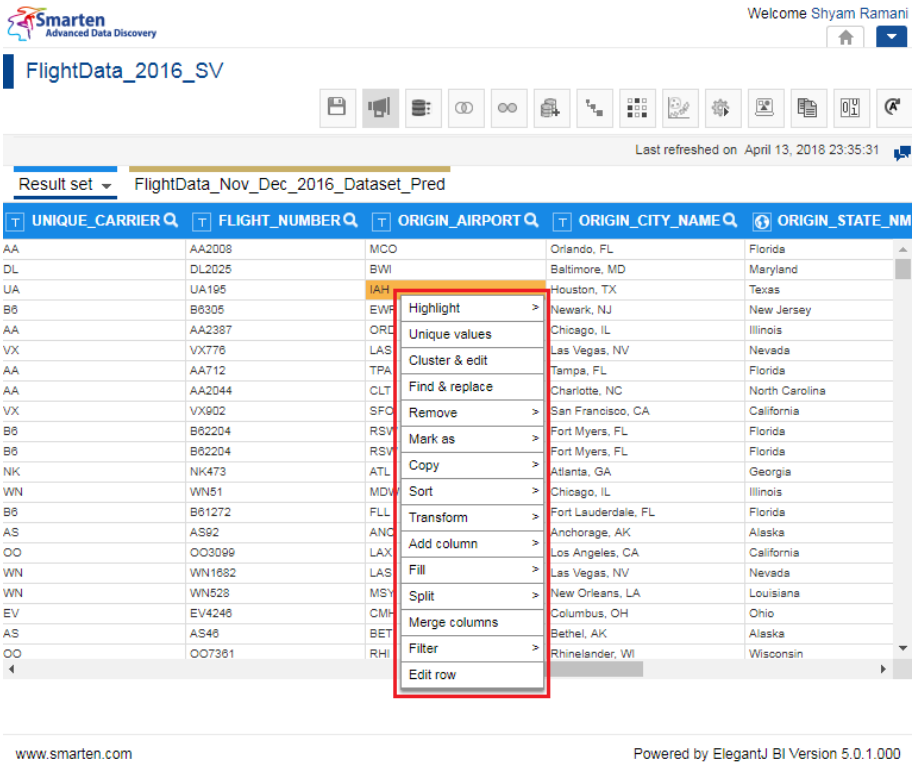
About this task

Use this task to edit data of a row.

Procedure

1. Open the dataset for which you want to edit data.
2. Right-click in the row for which you want to edit data.

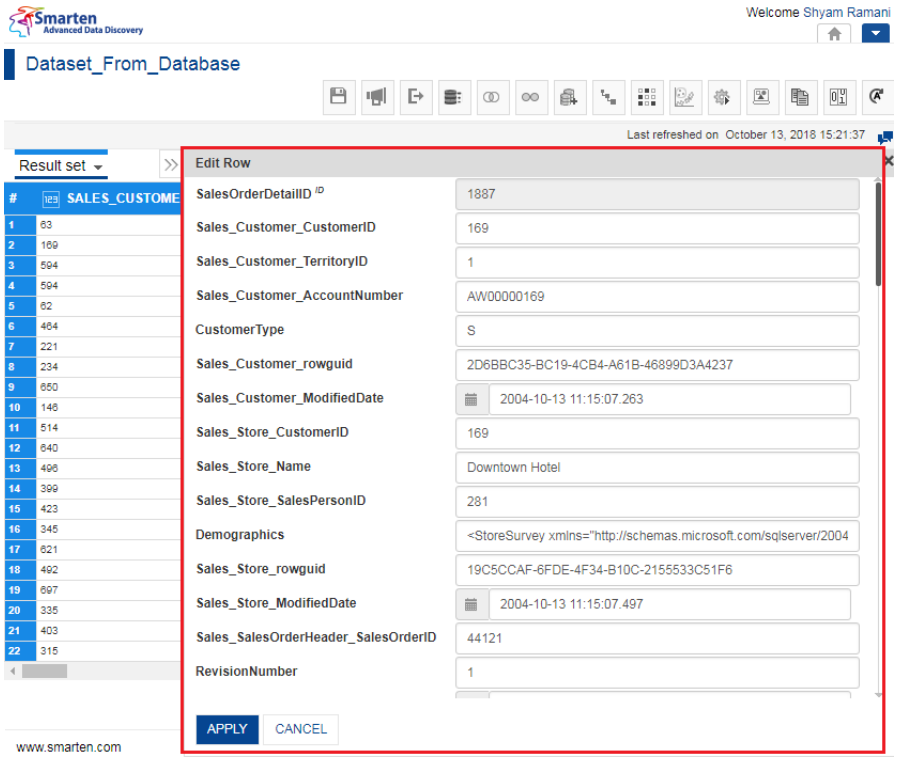
   The system displays the context menu.



EDIT ROW—THE CONTEXT MENU

3. Click **Edit row** from the menu.

   The system displays the **Edit Row** dialog box.



EDIT ROW—THE EDIT ROW DIALOG BOX

All columns and their values are displayed in the dialog box for the selected row.

4. You can edit the value of columns, and then click **APPLY**.

7.2.20 Viewing Statistics of Data in a Dataset

You can explore data with the help of various statistics, including skewness and kurtosis.

Reference: **Concept Manual > Explore Data > Statistics**

About this task

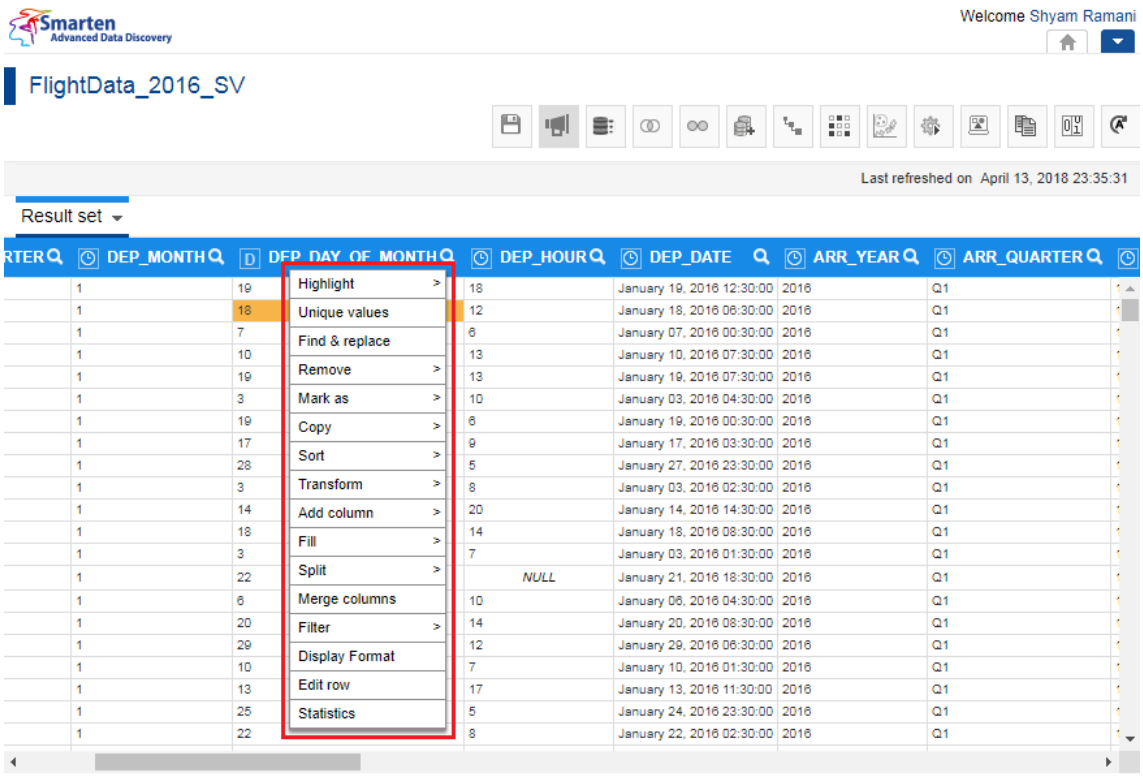
Use this task to view statistics of data in a dataset.

> Note:
> This option is only available for the numeric data type.

Procedure

1. Open the dataset for which you want to view statistics.
2. Right-click in the column for which you want to view statistics.

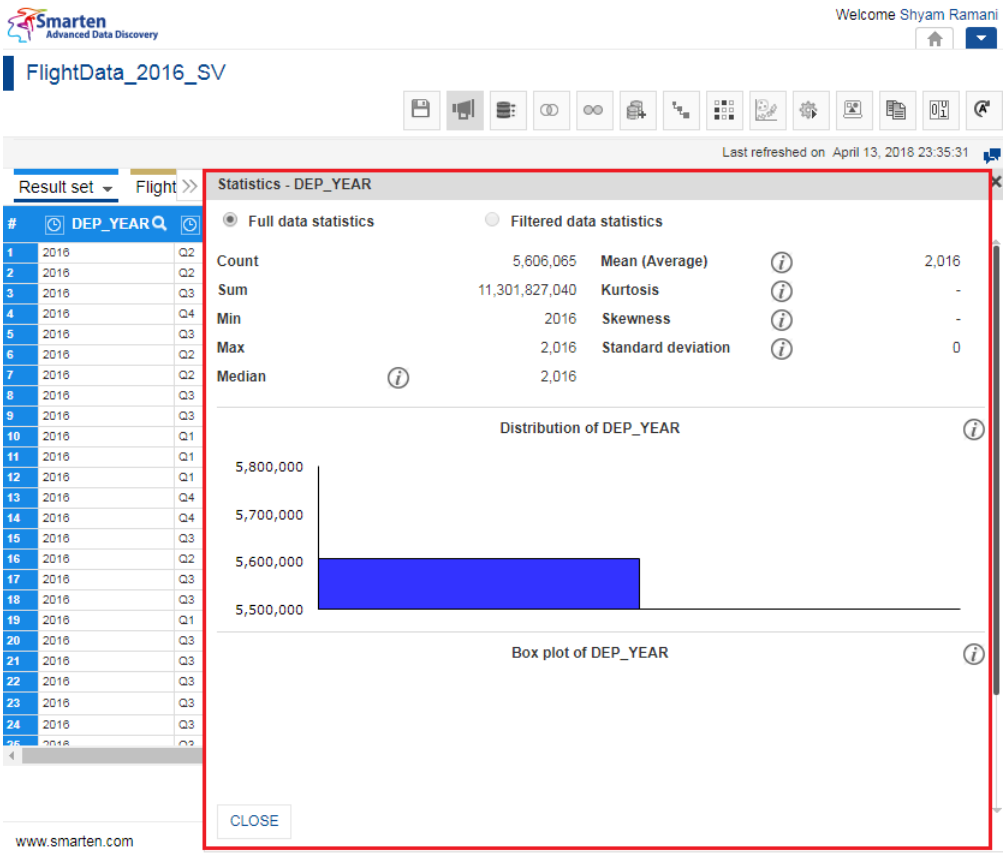
   The system displays the context menu.



STATISTICS—THE CONTEXT MENU

3. Click **Statistics** from the menu.

   The system displays the **Statistics** dialog box.



STATISTICS—THE STATISTICS DIALOG BOX

The dialog box displays the following statistics:

Heading	Description
Count	Displays the count of the column.
Sum	Displays the sum of the column.
Min	Displays the minimum value of the column.
Max	Displays the maximum value of the column.
Median	Displays the value in the middle when the data items are arranged in ascending order.
Mean	Displays the average of all data values of the column.
Kurtosis	Displays the measure of the peakedness of the dataset.
Skewness	Displays the measure of symmetry. A dataset is symmetric if it looks the same to the left and right of the center point.
Standard deviation	Displays the measure of how spread out the dataset is.
Distribution	A graphical display where the data is grouped into buckets and then plotted as bars.
Box Plot	A standardized way of displaying the distribution of data based on the five-number summary: minimum, first quartile, median, third quartile, and maximum.

7.2.21 Filling Data in a Dataset

You can fill null or empty values in rows for a column based on previous row values.

Reference: **Concept Manual > Shape Data > Fill Data**

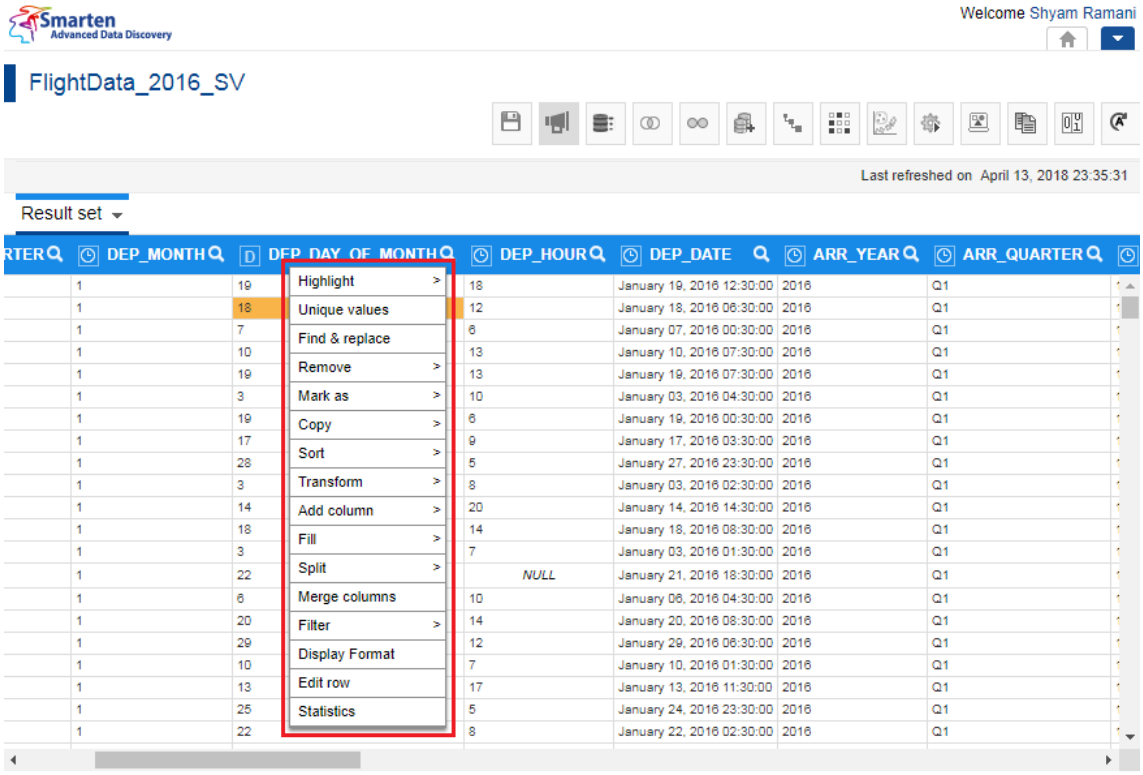
About this task

Use this task to fill null or empty values.

Procedure

1. Open the dataset for which you want to fill null or empty values.
2. Right-click in the row for which you want to fill null or empty values.

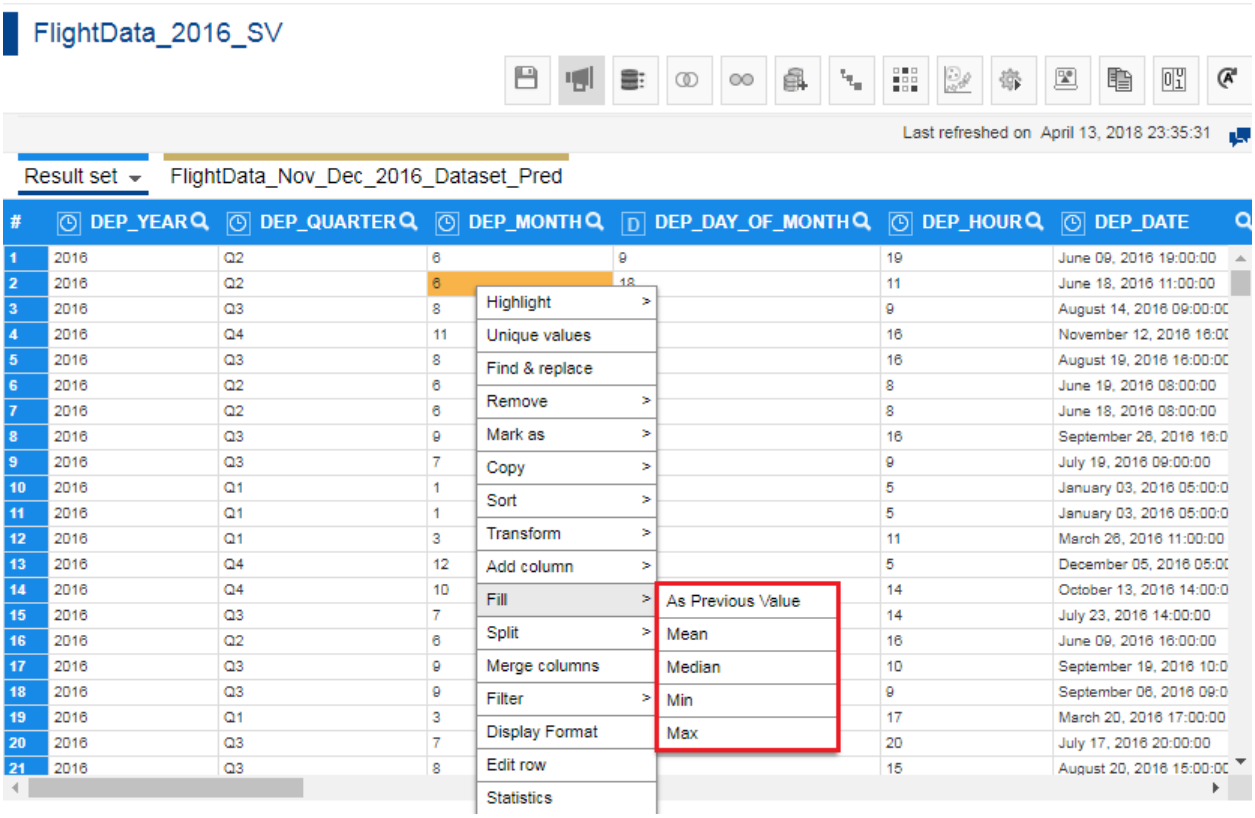
   The system displays the context menu.



FILL DATA—THE CONTEXT MENU

3. Click **Fill** from the menu.

   The system fills the null or empty value with the value from the above cell.



FILL DATA—THE FILL OPTIONS

4. Select **As Previous Value** to fill the null or empty cells with the value available in the previous cell.

   The following options are available for numeric data type:

   - **Mean**: This option fills the null or empty cells with the mean value of all values available in the selected column.
   - **Median**: This option fills the null or empty cells with the median value of all values available in the selected column.
   - **Min**: This option fills the null or empty cells with the lowest value available in the selected column.
   - **Max**: This option fills the null or empty cells with the highest value available in the selected column.

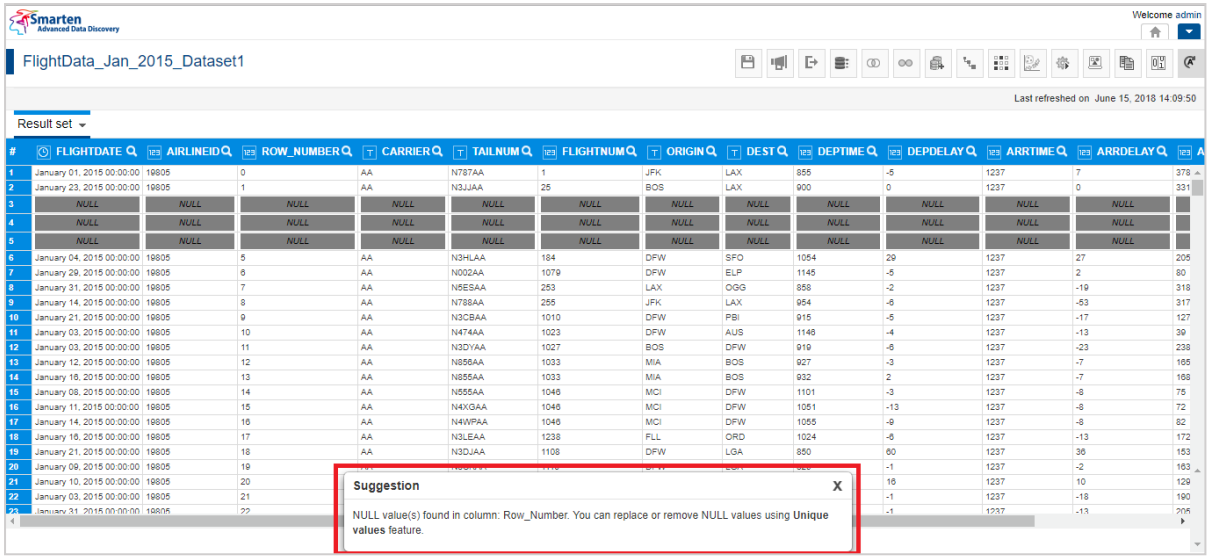
7.2.22 Enable or Disable Auto-Suggestions and Recommendations

The intuitively guided interface of Smarten SSDP provides users with relevant auto-suggestions and recommendations to help users improve the quality of data.

Guided by smart suggestions and auto recommendations, business users can prepare, blend, and transform data and create analysis-ready data quickly and accurately without assistance or any specialized skills or scripting.

For example, Smarten SSDP will provide suggestions when there are blank values in a String column, outlier values in a Numeric column, or when all rows of a String column contain numeric values, and many other suggestions to help users quickly create analysis-ready data.

Shown below is the image of auto-suggestions and recommendations provided by Smarten SSDP for a dataset.



AUTO-SUGGESTIONS AND RECOMMENDATIONS—DATASET

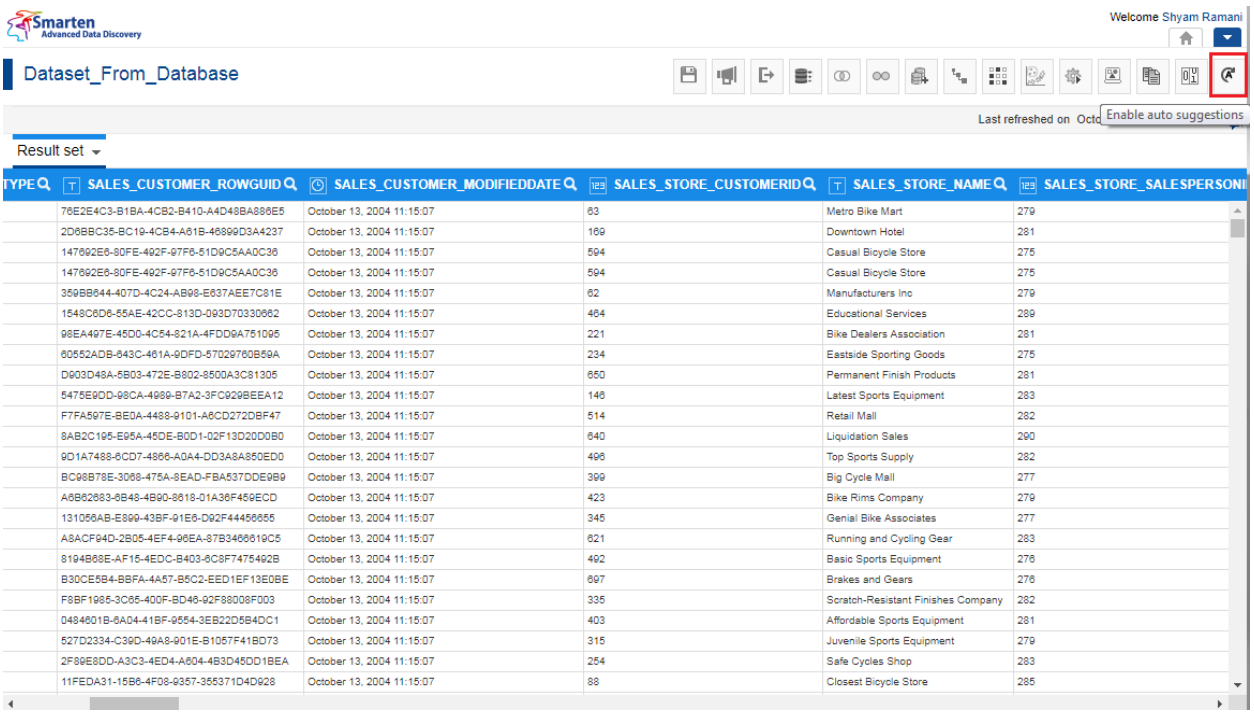
Reference: **Concept Manual > Auto-suggestions and Recommendations**

About this task

Use this task to enable or disable auto-suggestions and recommendations.

Procedure

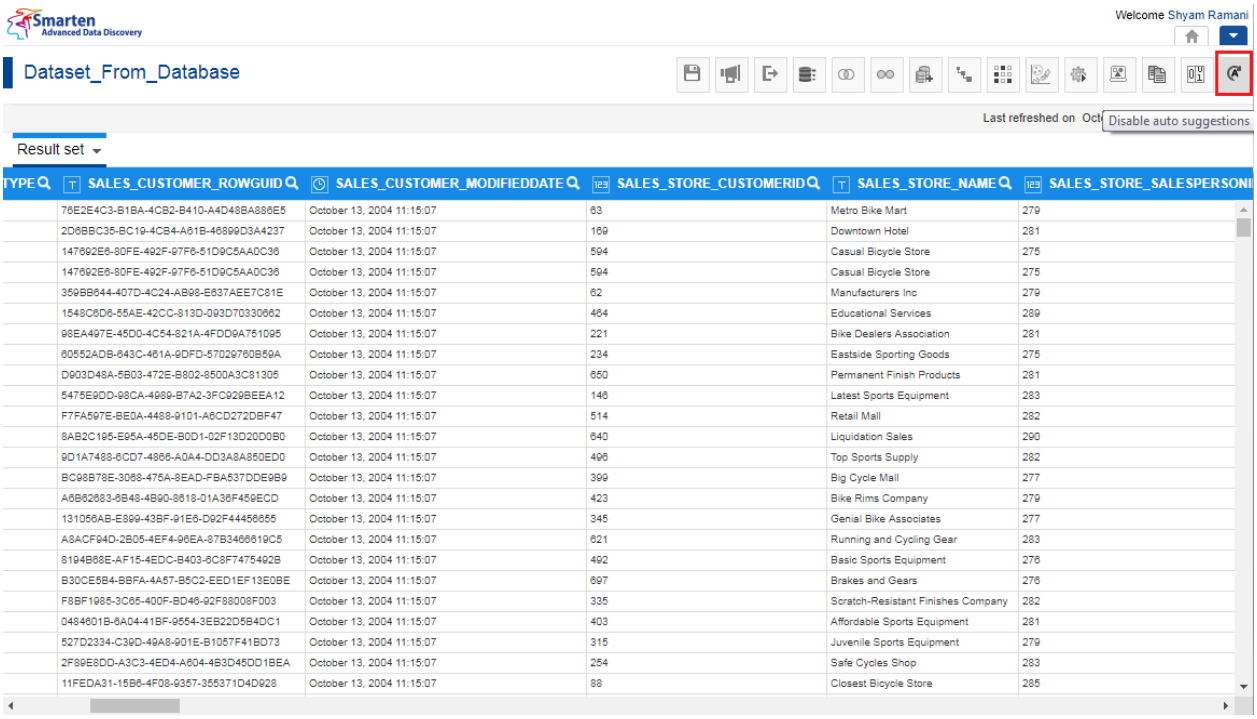
1. Open the dataset for which you want to enable or disable auto-suggestions.
2. Click the Auto-Suggestions icon on the toolbar to enable auto-suggestions and recommendations.



AUTO-SUGGESTIONS—ENABLING AUTO-SUGGESTIONS

> **Note**:
> The system displays a tooltip to enable auto-suggestions when you hover the mouse pointer over the icon if auto-suggestions is disabled.
> Once auto-suggestions is enabled, the system analyses the data in the background and provides suggestions to the user.

3. Similarly, you can click the same icon to disable auto-suggestions.



AUTO-SUGGESTIONS—DISABLING AUTO-SUGGESTIONS

> **Note**:
> The system displays a tooltip to disable auto-suggestions when you hover the mouse pointer over the icon if auto-suggestions is enabled.

7.2.23 Enable or Disable Record Count

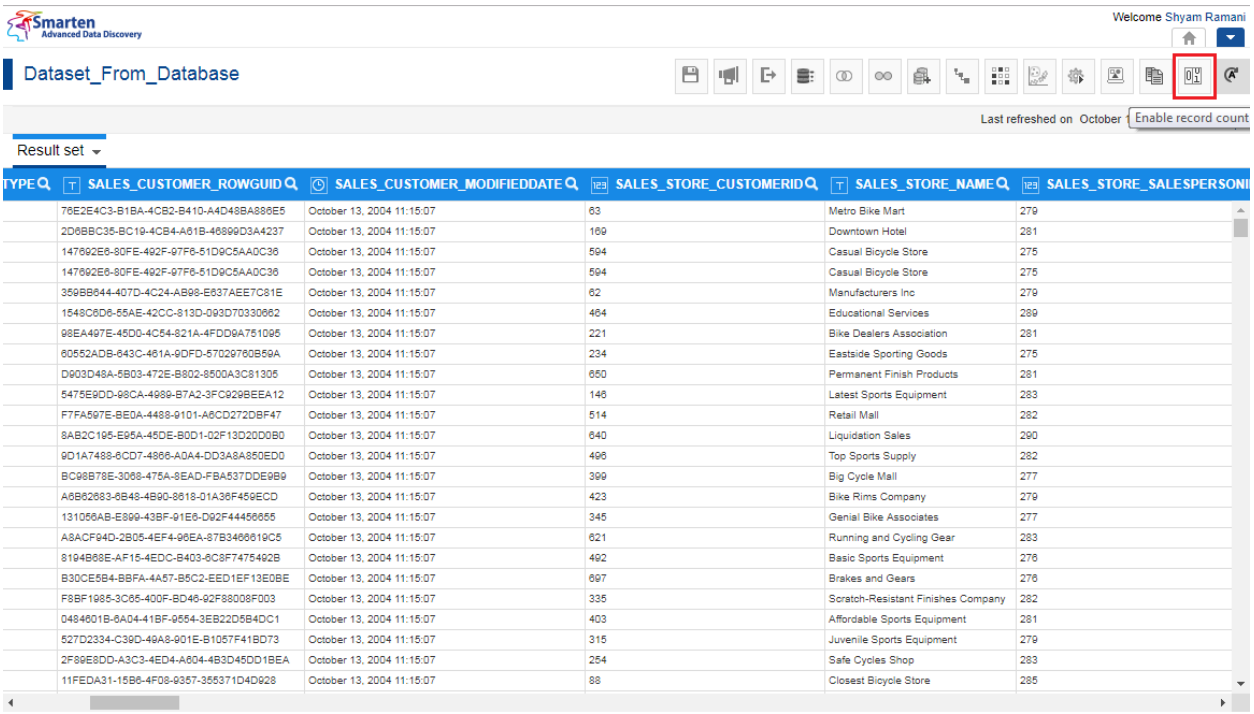
You can display the total number of records available in a dataset.

About this task

Use this task to enable or disable the record count display for a dataset.

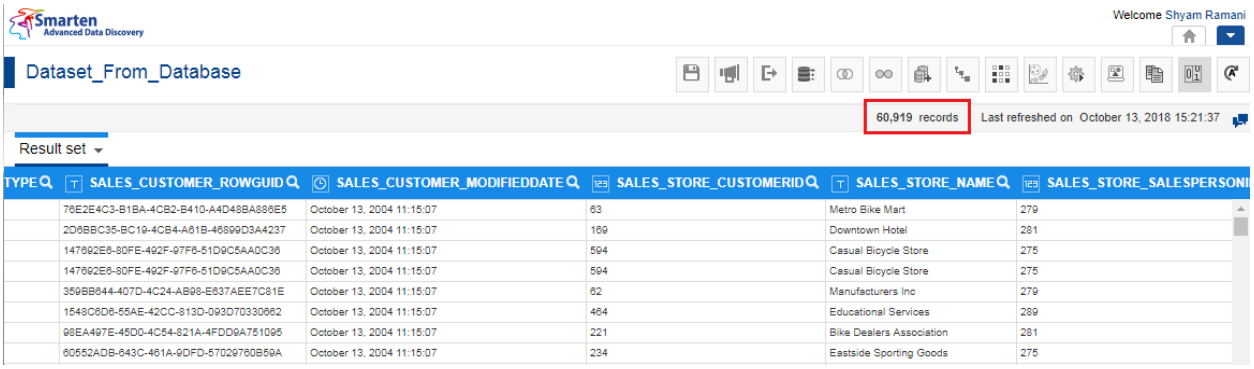
Procedure

1. Open the dataset for which you want to enable or disable record count.
2. Click the record count icon on the toolbar to display the total number of records available for the dataset.



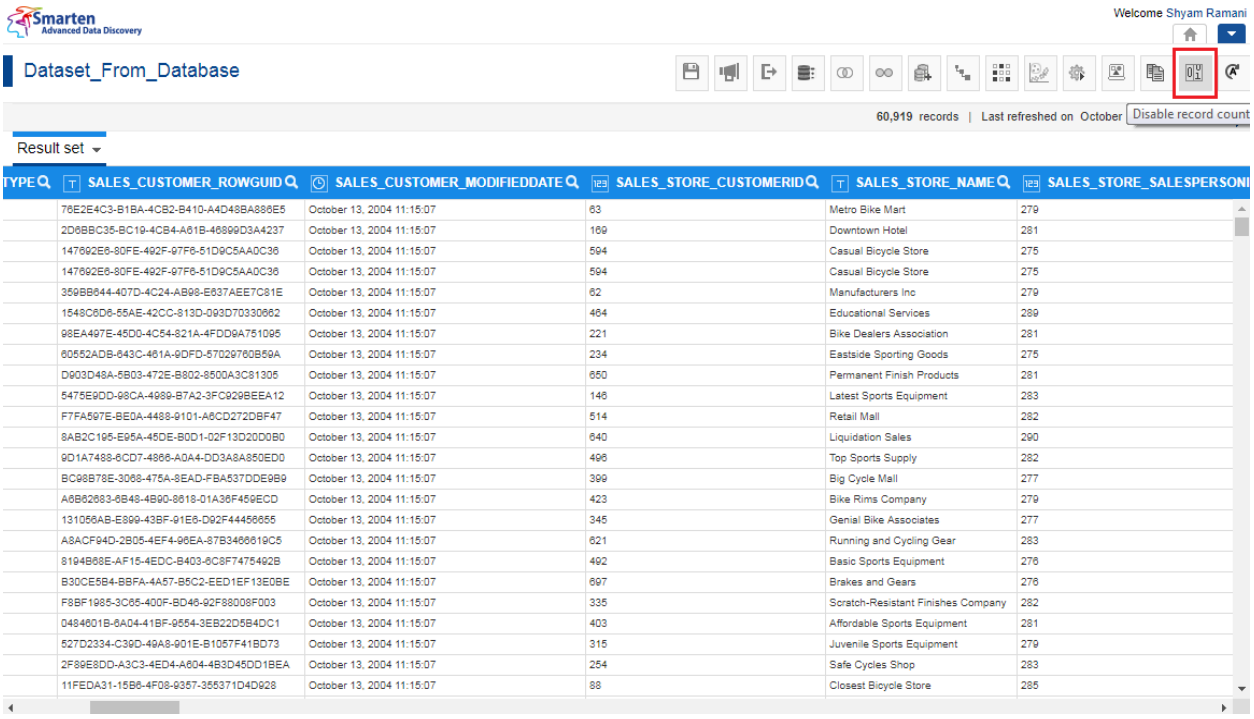
RECORD COUNT—ENABLING RECORD COUNT

The system displays the number of records available for the dataset.



RECORD COUNT—DISPLAYING THE TOTAL NUMBER OF RECORDS

3. If record count is disabled, the system displays a tooltip to enable record count when you hover the mouse pointer over the icon. Similarly, you can click the same icon to disable record count.



AUTO-SUGGESTIONS—DISABLING AUTO-SUGGESTIONS

> **Note**:
> If record count is enabled, the system displays a tooltip to disable record count when you hover the mouse pointer over the icon.

7.2.24 Enable or Disable Full Data Mode

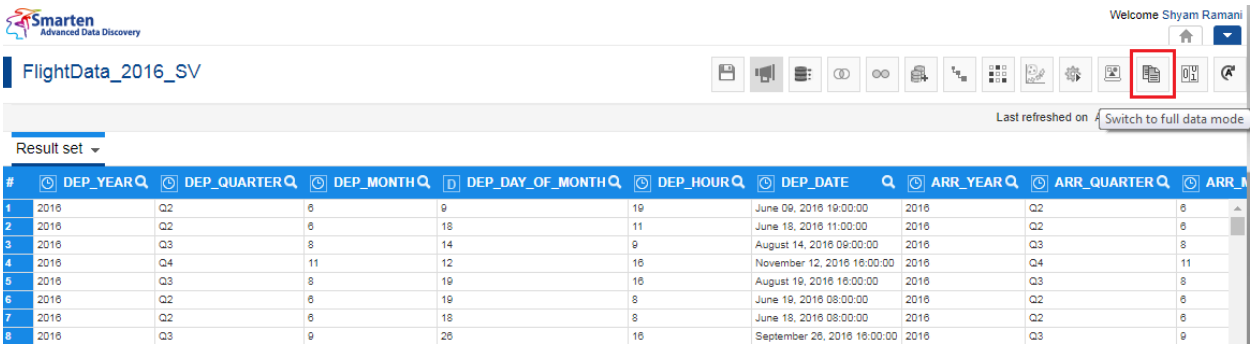
You can enable full data mode to retrieve and display all the data of a dataset. You can also disable full data mode to work on limited data.

About this task

Use this task to enable or disable full data mode for a dataset.

Procedure

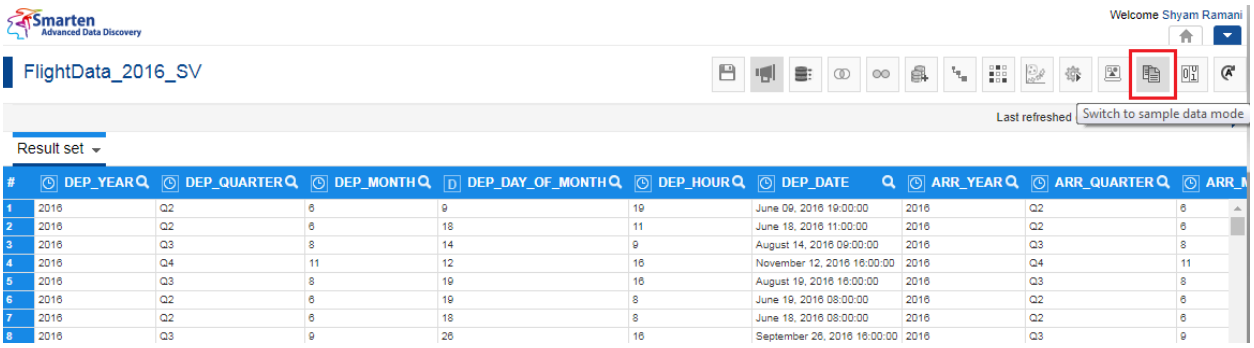
1. Open the dataset for which you want to enable or disable full data mode.
2. Click the full data mode icon on the toolbar to retrieve and display all the data for the dataset.



FULL DATA MODE—ENABLING FULL DATA MODE

> **Note**:
> If full data mode is disabled, the system displays a tooltip to enable full data mode when you hover the mouse pointer over the icon.

3. Similarly, you can click the same icon to disable full data mode.



#	DEP_YEAR	DEP_QUARTER	DEP_MONTH	DEP_DAY_OF_MONTH	DEP_HOUR	DEP_DATE	ARR_YEAR	ARR_QUARTER	ARR_M
1	2016	Q2	6	9	19	June 09, 2016 19:00:00	2016	Q2	6
2	2016	Q2	6	18	11	June 18, 2016 11:00:00	2016	Q2	6
3	2016	Q3	8	14	9	August 14, 2016 09:00:00	2016	Q3	8
4	2016	Q4	11	12	16	November 12, 2016 16:00:00	2016	Q4	11
5	2016	Q3	8	19	16	August 19, 2016 16:00:00	2016	Q3	8
6	2016	Q2	6	19	8	June 19, 2016 08:00:00	2016	Q2	6
7	2016	Q2	6	18	8	June 18, 2016 08:00:00	2016	Q2	6
8	2016	Q3	9	26	16	September 26, 2016 16:00:00	2016	Q3	9

FULL DATA MODE—DISABLING FULL DATA MODE

> **Note**:
> If full data mode is enabled, the system displays a tooltip to disable full data mode when you hover the mouse pointer over the icon.

7.2.25 Viewing the Lineage Diagram

A diagram representation of all the actions performed by a user, a Lineage diagram helps users view the complete data flow and transformation steps in a single view.

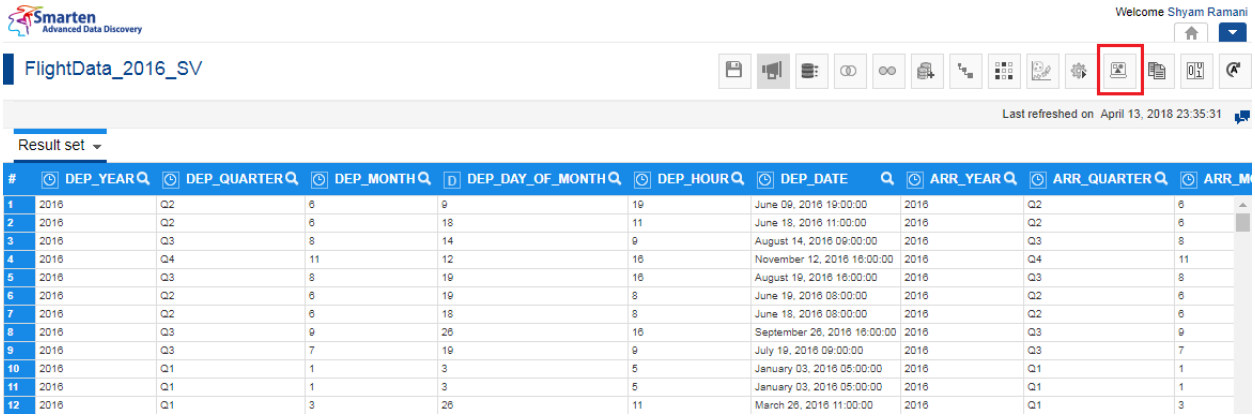
Reference: **Concept Manual > Lineage Diagram**

About this task

Use this task to view the lineage diagram.

Procedure

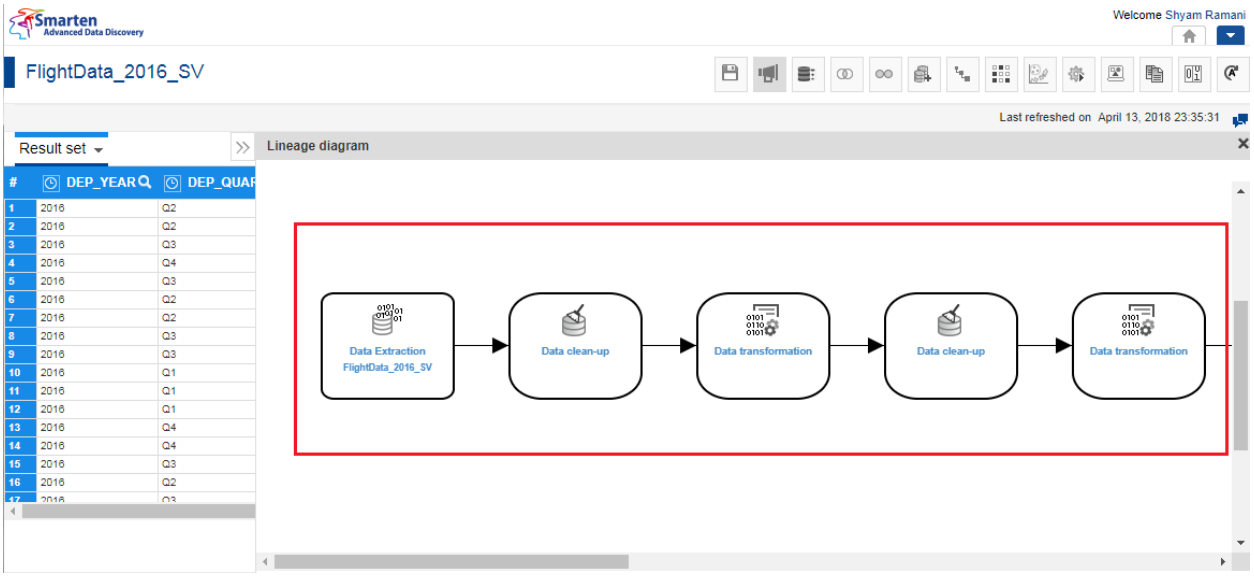
1. Open the dataset for which you want to view the lineage diagram.
2. Click the lineage diagram icon on the toolbar to display the lineage diagram for the dataset.



#	DEP_YEAR	DEP_QUARTER	DEP_MONTH	DEP_DAY_OF_MONTH	DEP_HOUR	DEP_DATE	ARR_YEAR	ARR_QUARTER	ARR_M
1	2016	Q2	6	9	19	June 09, 2016 19:00:00	2016	Q2	6
2	2016	Q2	6	18	11	June 18, 2016 11:00:00	2016	Q2	6
3	2016	Q3	8	14	9	August 14, 2016 09:00:00	2016	Q3	8
4	2016	Q4	11	12	16	November 12, 2016 16:00:00	2016	Q4	11
5	2016	Q3	8	19	16	August 19, 2016 16:00:00	2016	Q3	8
6	2016	Q2	6	19	8	June 19, 2016 08:00:00	2016	Q2	6
7	2016	Q2	6	18	8	June 18, 2016 08:00:00	2016	Q2	6
8	2016	Q3	9	26	16	September 26, 2016 16:00:00	2016	Q3	9
9	2016	Q3	7	19	9	July 19, 2016 09:00:00	2016	Q3	7
10	2016	Q1	1	3	5	January 03, 2016 05:00:00	2016	Q1	1
11	2016	Q1	1	3	5	January 03, 2016 05:00:00	2016	Q1	1
12	2016	Q1	3	26	11	March 26, 2016 11:00:00	2016	Q1	3

LINEAGE DIAGRAM—OPEN LINEAGE DIAGRAM

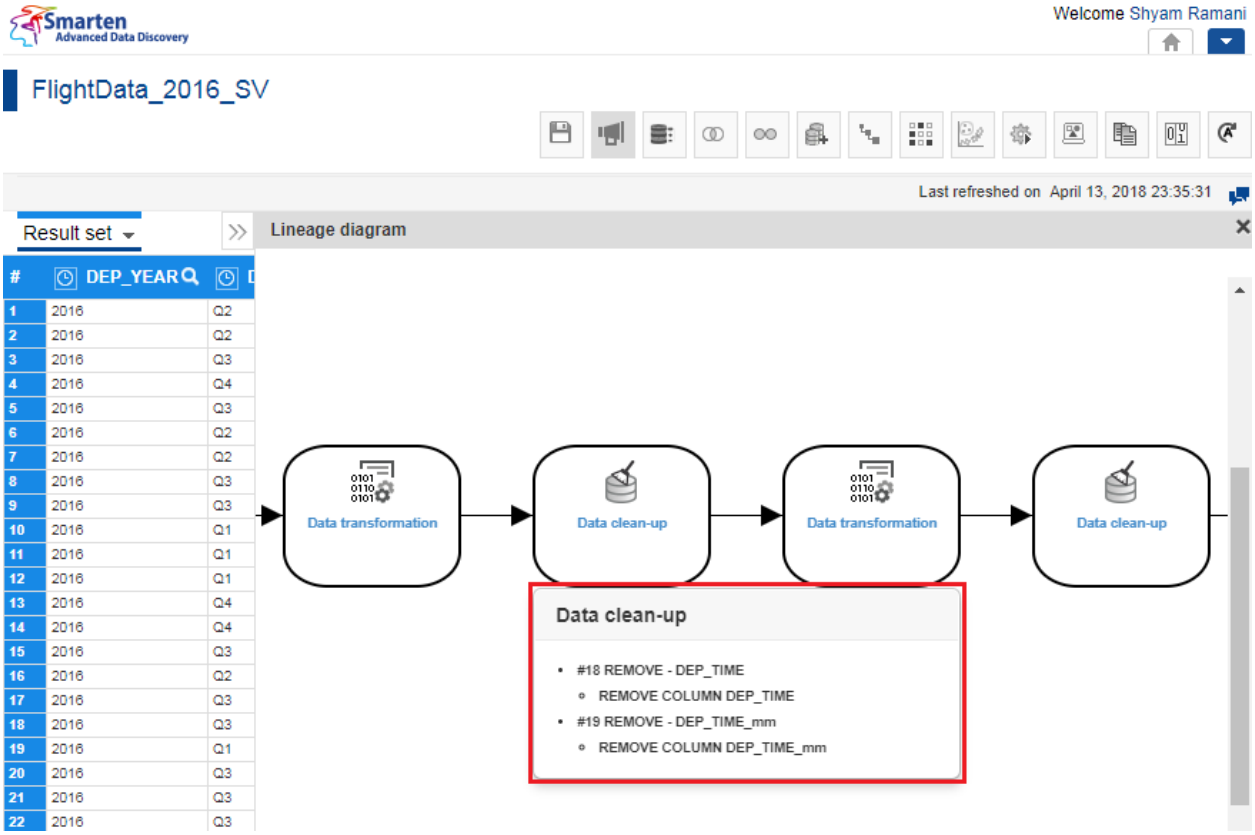
The system displays the lineage diagram in the **Lineage diagram** dialog box.



LINEAGE DIAGRAM—VIEWING LINEAGE DIAGRAM

3. You can click the close icon in the **Lineage diagram** dialog box to close the dialog box.

The system displays information about actions being performed in a step when you hover the mouse pointer over that step.



LINEAGE DIAGRAM—VIEWING ACTION DETAILS FOR A STEP

7.2.26 Opening the Action Editor

Users can view all data-related actions performed on the dataset right from the extraction process to preparing analysis-ready data. Users can roll back the effects of actions by deleting and activating or inactivating the actions. Some actions can also be edited.

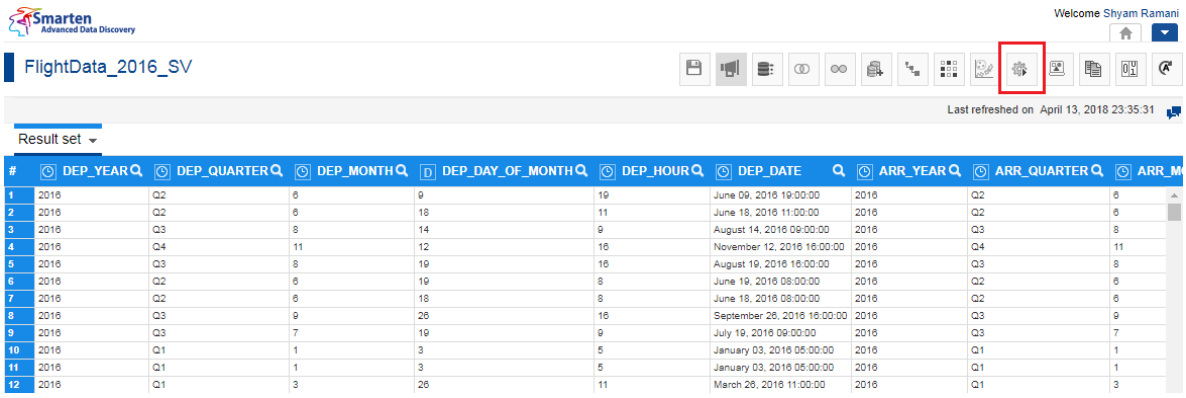
Reference: **Concept Manual > Action Editor**

About this task

Use this task to view the action editor.

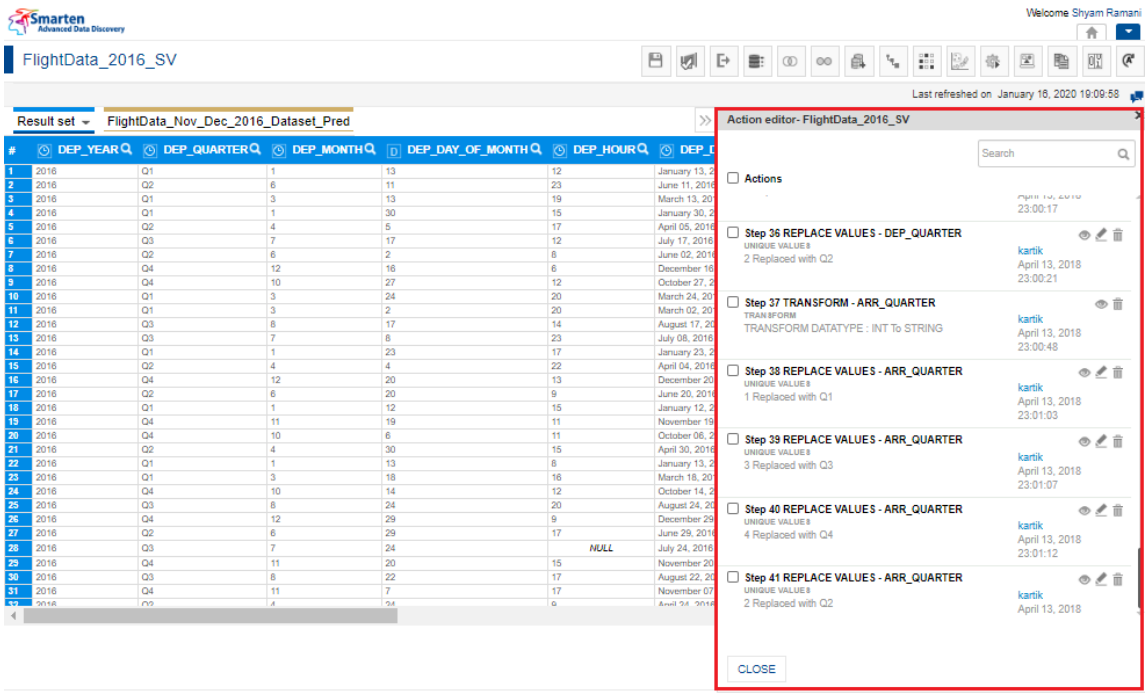
Procedure

1. Open the dataset for which you want to open the action editor.
2. Click the action editor icon on the toolbar to display the actions that are performed on the dataset.



ACTION EDITOR—OPEN ACTION EDITOR

The system displays the actions performed on the dataset in the **Action editor** dialog box.

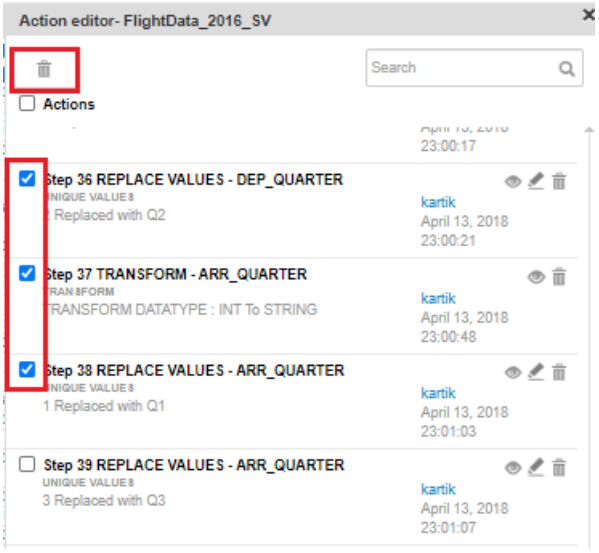


ACTION EDITOR—VIEWING ACTION EDITOR DIALOG BOX

The Action editor dialog box lists the actions in the sequence they are performed on the dataset. For each action, you can view the operations being performed, creator of the action, time and date when the action was created, and options to activate, inactivate, edit, and delete the action. You can also use the search box to search for an action.

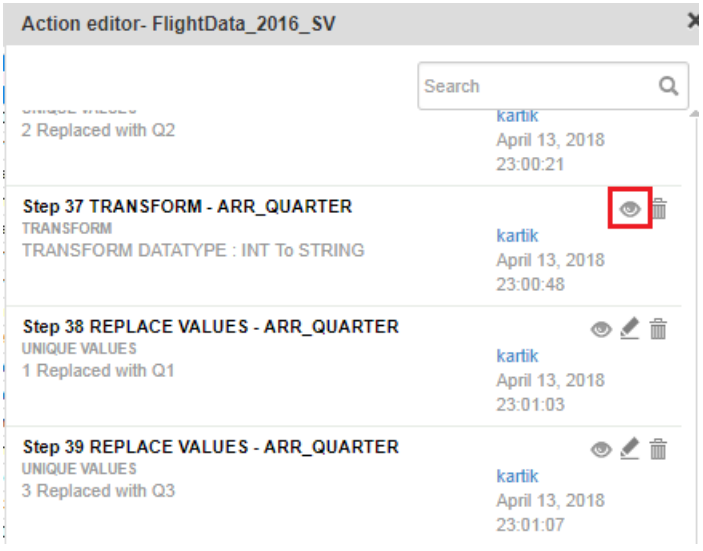
> Note:
> The edit icon is displayed only for some actions, not for all actions.

3. You can select the check box adjacent to the actions you want to delete and click the Delete icon to delete multiple actions.



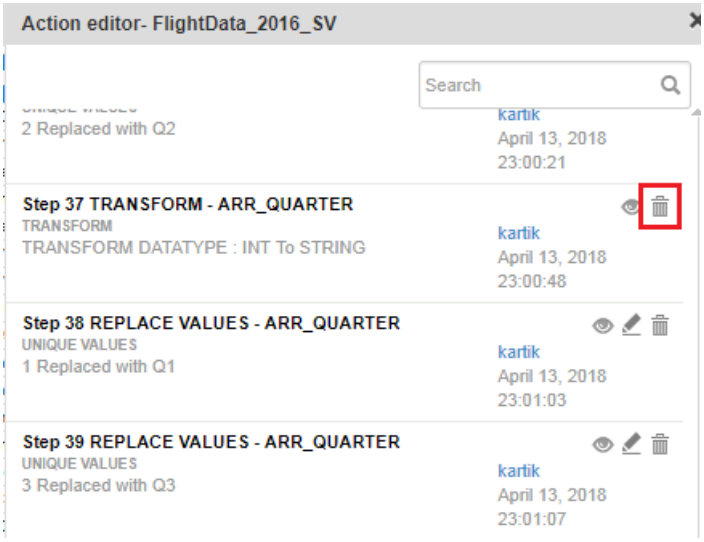
ACTION EDITOR – DELETE MULTIPLE ACTIONS

4. You can click the inactivate icon to deactivate the action.



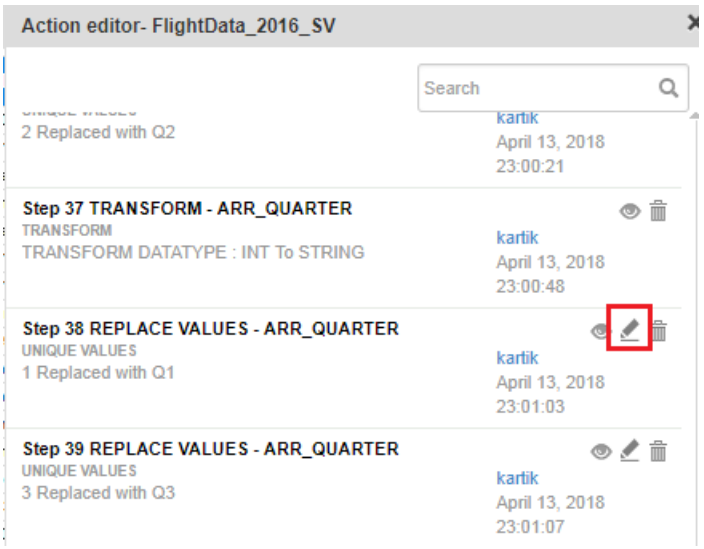
ACTION EDITOR—DEACTIVATE AN ACTION

5. You can click the delete icon to delete the action.



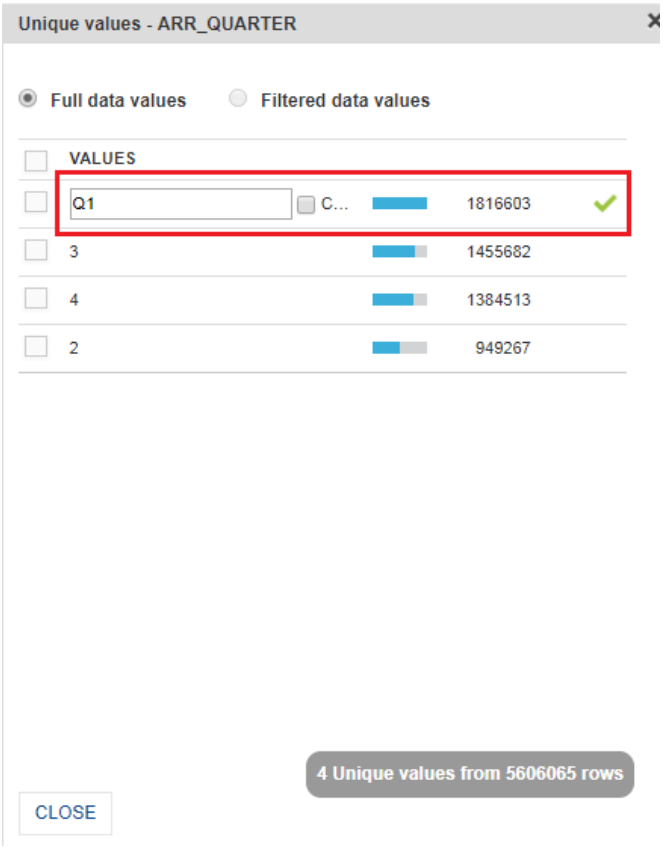
ACTION EDITOR—DELETE AN ACTION

6. You can click the edit icon adjacent to the action you want to edit to modify operations for that action.



ACTION EDITOR—MODIFY AN ACTION

The system displays the operation being performed in the action.

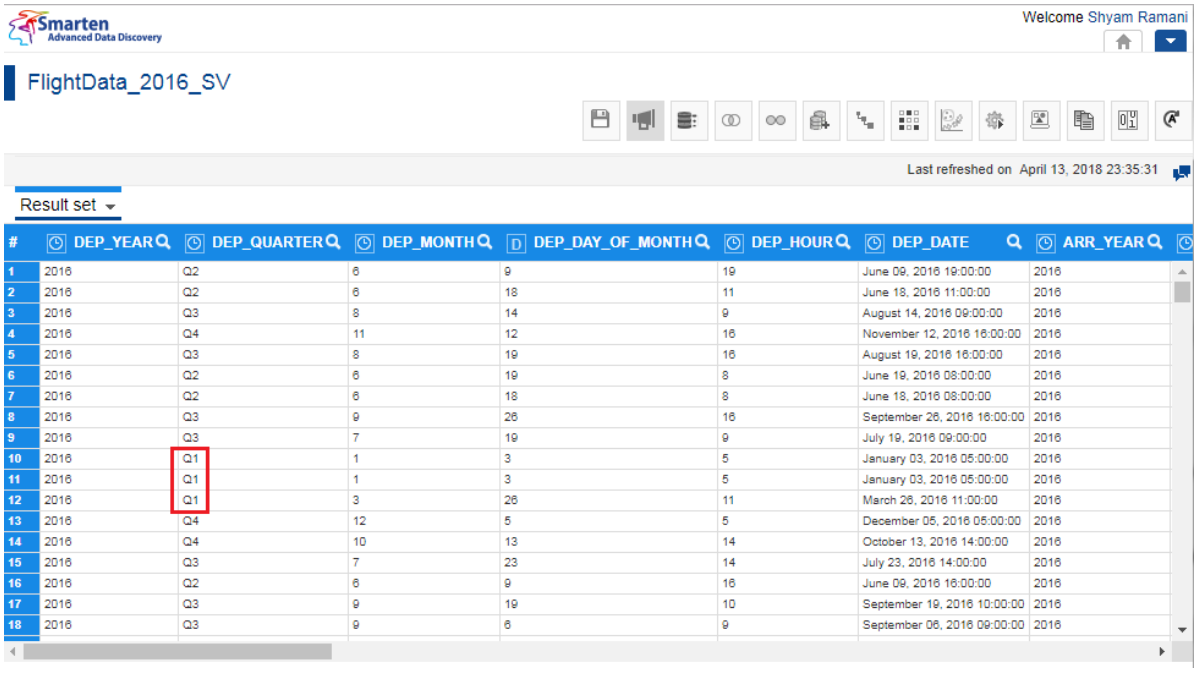


ACTION EDITOR—MODIFYING AN ACTION

In the image above, the action replaces the unique value “1” with “Q1.” You can modify this action to replace the unique value “1” with any other value.

7. Click the OK icon to save the updated value for the action.

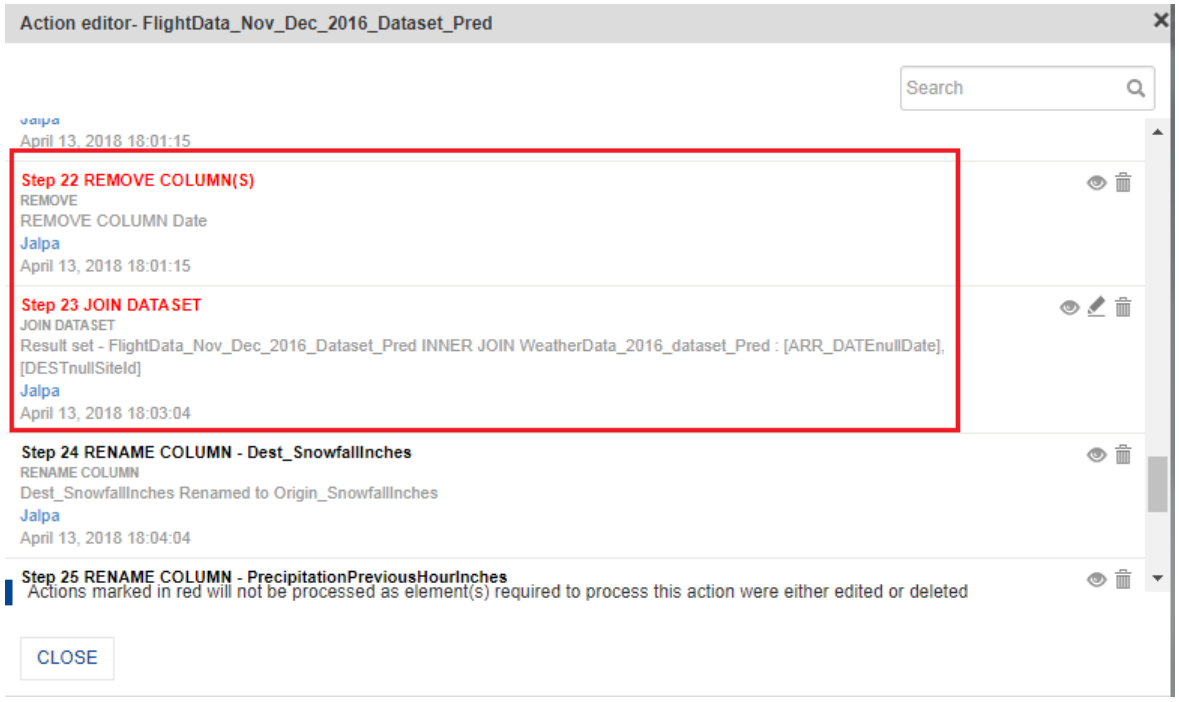
   The system now replaces the unique value “1” in the example above with the updated value.



#	DEP_YEAR	DEP_QUARTER	DEP_MONTH	DEP_DAY_OF_MONTH	DEP_HOUR	DEP_DATE	ARR_YEAR
1	2016	Q2	6	9	19	June 09, 2016 19:00:00	2016
2	2016	Q2	6	18	11	June 18, 2016 11:00:00	2016
3	2016	Q3	8	14	9	August 14, 2016 09:00:00	2016
4	2016	Q4	11	12	16	November 12, 2016 16:00:00	2016
5	2016	Q3	8	19	16	August 19, 2016 16:00:00	2016
6	2016	Q2	6	19	8	June 19, 2016 08:00:00	2016
7	2016	Q2	6	18	8	June 18, 2016 08:00:00	2016
8	2016	Q3	9	26	16	September 26, 2016 16:00:00	2016
9	2016	Q3	7	19	9	July 19, 2016 09:00:00	2016
10	2016	Q1	1	3	5	January 03, 2016 05:00:00	2016
11	2016	Q1	1	3	5	January 03, 2016 05:00:00	2016
12	2016	Q1	3	26	11	March 26, 2016 11:00:00	2016
13	2016	Q4	12	5	5	December 05, 2016 05:00:00	2016
14	2016	Q4	10	13	14	October 13, 2016 14:00:00	2016
15	2016	Q3	7	23	14	July 23, 2016 14:00:00	2016
16	2016	Q2	6	9	16	June 09, 2016 16:00:00	2016
17	2016	Q3	9	19	10	September 19, 2016 10:00:00	2016
18	2016	Q3	9	6	9	September 06, 2016 09:00:00	2016

ACTION EDITOR—REPLACING A VALUE

The system also highlights actions in red that have encountered an error.



ACTION EDITOR—ACTIONS WITH ERROR

7.2.27 Processing the Outlier Values

Outliers are the observations lying outside the overall pattern of distribution.

This function allows users to identify outlier values in the data and replace or remove them from the dataset. Users can also download records detected as outliers in the form of a CSV file.

> Note:
> Outliers are only applicable for numeric data type columns.

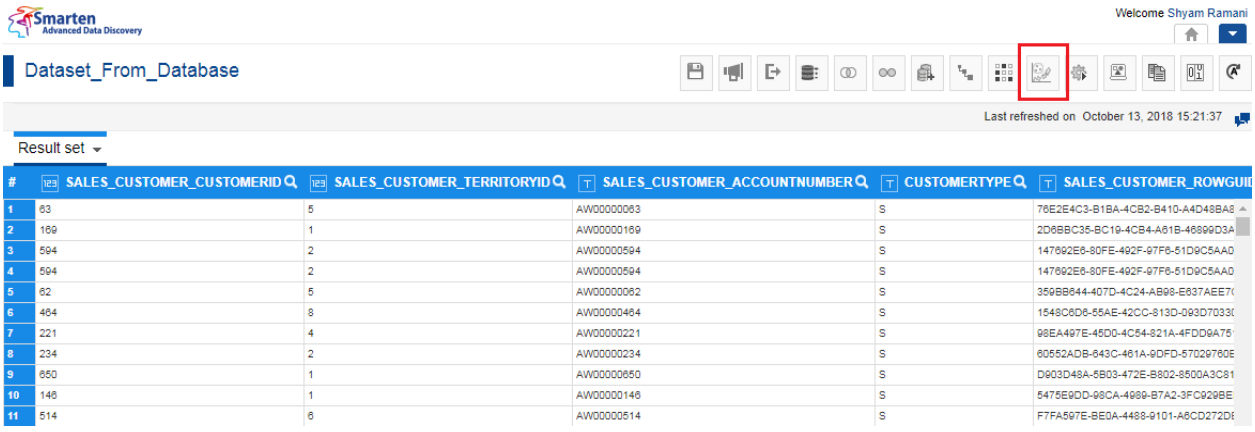
Reference: **Concept Manual > Clean Data > Outliers**

About this task

Use this task to view the outlier values.

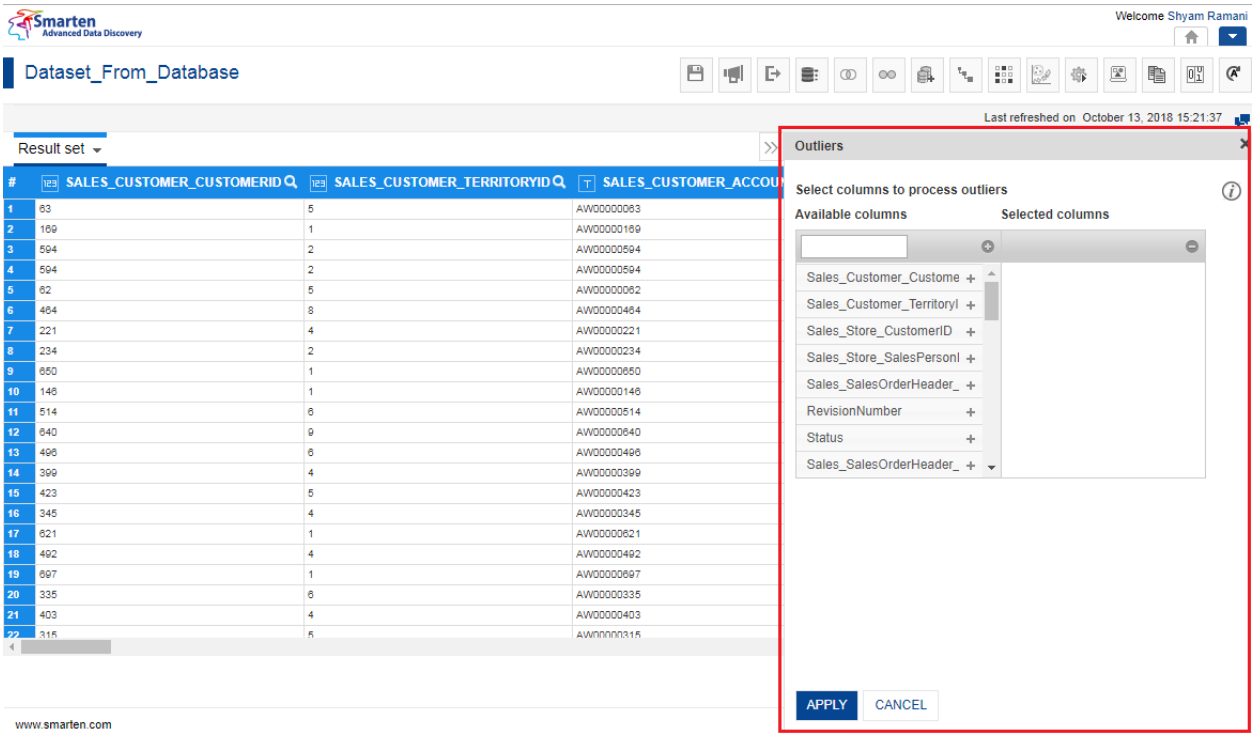
Procedure

1. Open the dataset for which you want to view the outlier values.
2. Click the outliers icon on the toolbar.



OUTLIER VALUES—OPENING THE OUTLIERS DIALOG BOX

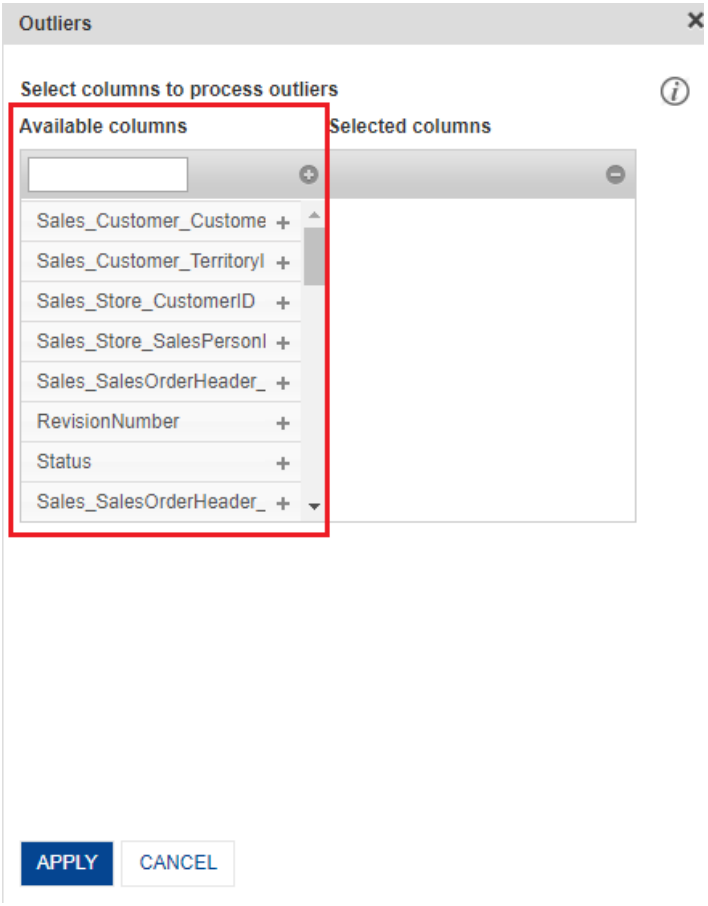
The system displays the **Outliers** dialog box. All the columns with numeric data type available in the dataset are listed within the **Available columns**.



OUTLIER VALUES—THE OUTLIERS DIALOG BOX

3. Click the plus sign adjacent to the columns for which you want to view the outlier values from the **Available columns** section. You can select multiple columns.

> Note:
> If you have selected multiple columns, the system calculates outlier values for all columns separately.

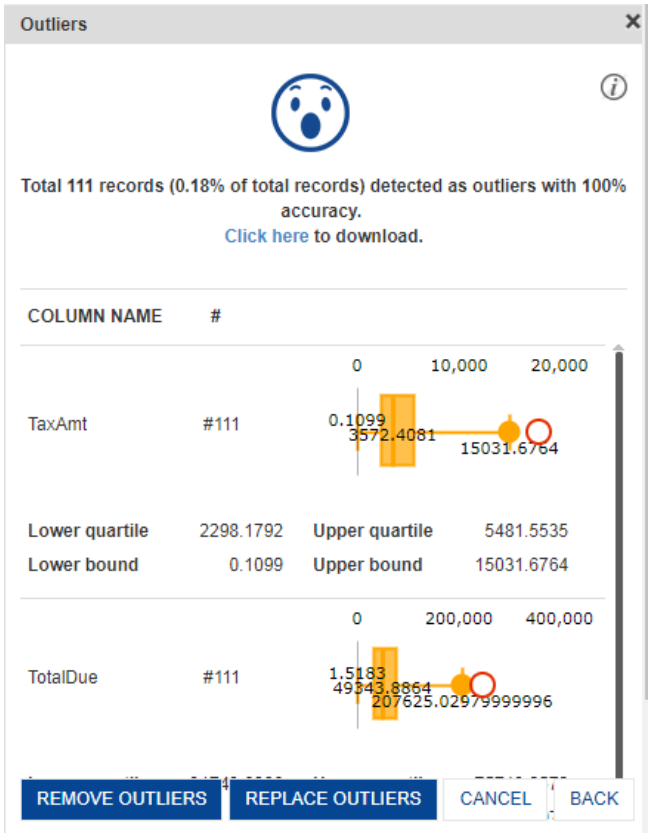


OUTLIER VALUES—AVAILABLE COLUMNS IN WHICH OUTLIER VALUES CAN BE FOUND

4. Click **APPLY**.

   The system displays the number of records that have outlier values.

   You can click the **Click here** option to download the records containing outlier values in a CSV file.



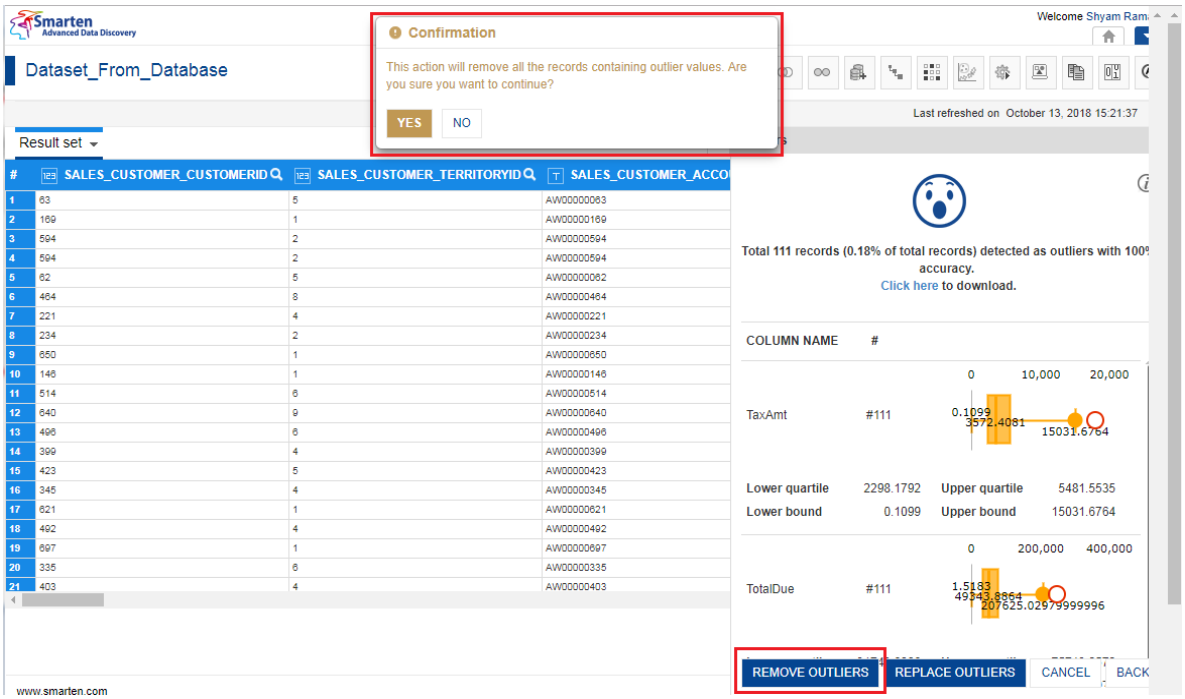
OUTLIER VALUES—OUTLIER VALUES DISPLAYED FOR THE SELECTED COLUMNS

Box plot is displayed with quartiles information to analyze the outlier values and patterns.

OUTLIER VALUES—OUTLIER VALUE INFORMATION

5. You can click **REMOVE OUTLIERS** to remove all records that contain outlier values from the dataset.

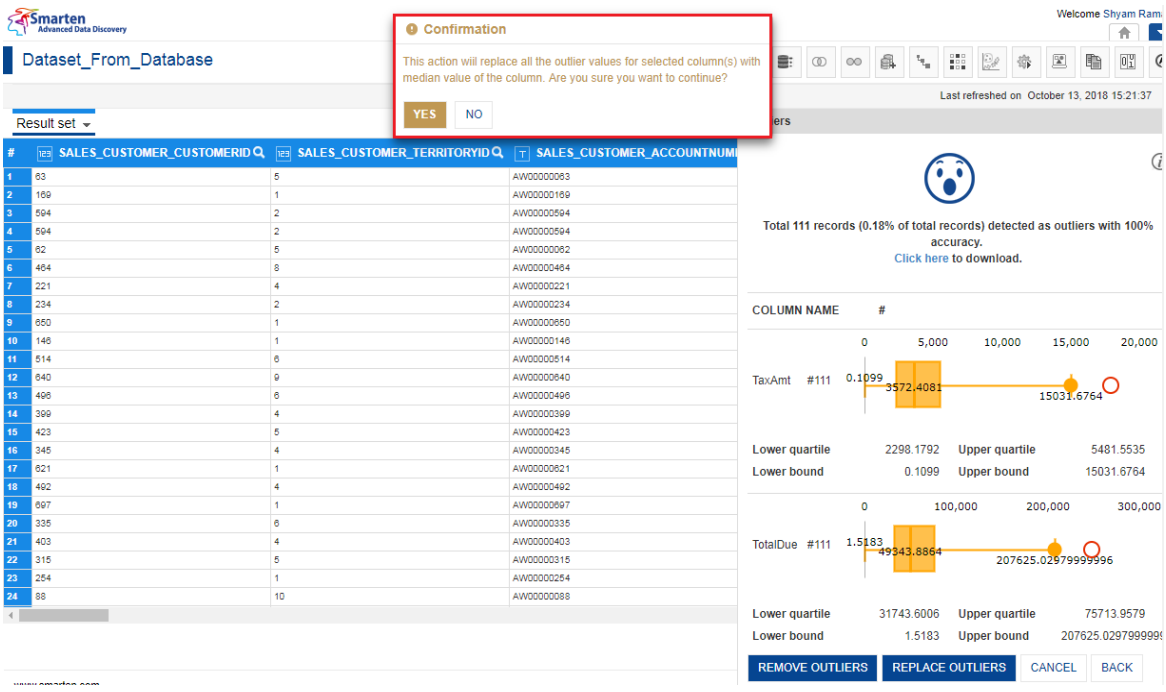
   The system removes all records that contain outlier values after confirmation.



OUTLIER VALUES—REMOVE RECORDS WITH OUTLIER VALUES

6. You can click REPLACE OUTLIERS to replace the outlier values with the median values of all the records for the selected columns.

   The system replaces all outlier values after confirmation.



OUTLIER VALUES—REPLACE OUTLIER VALUES

7. You can click the close icon in the **Outliers** dialog box to close the dialog box.

7.2.28 Sampling the Data

Sampling is a statistical procedure that is concerned with the selection of a subset (a statistical sample) from within a statistical population. By studying the sample, we may fairly generalize our results back to the population from which they were chosen. This not only keeps the cost low but also allows analyzing the sample faster as compared with the entire population.

Smarten SSDP allows users to take a sample from a dataset using two types of sampling methods.

Smarten SSDP recommends the size of the sampling data, which can be changed by the users, who also have a choice to enter the number of records of a dataset to be considered as the sample size.

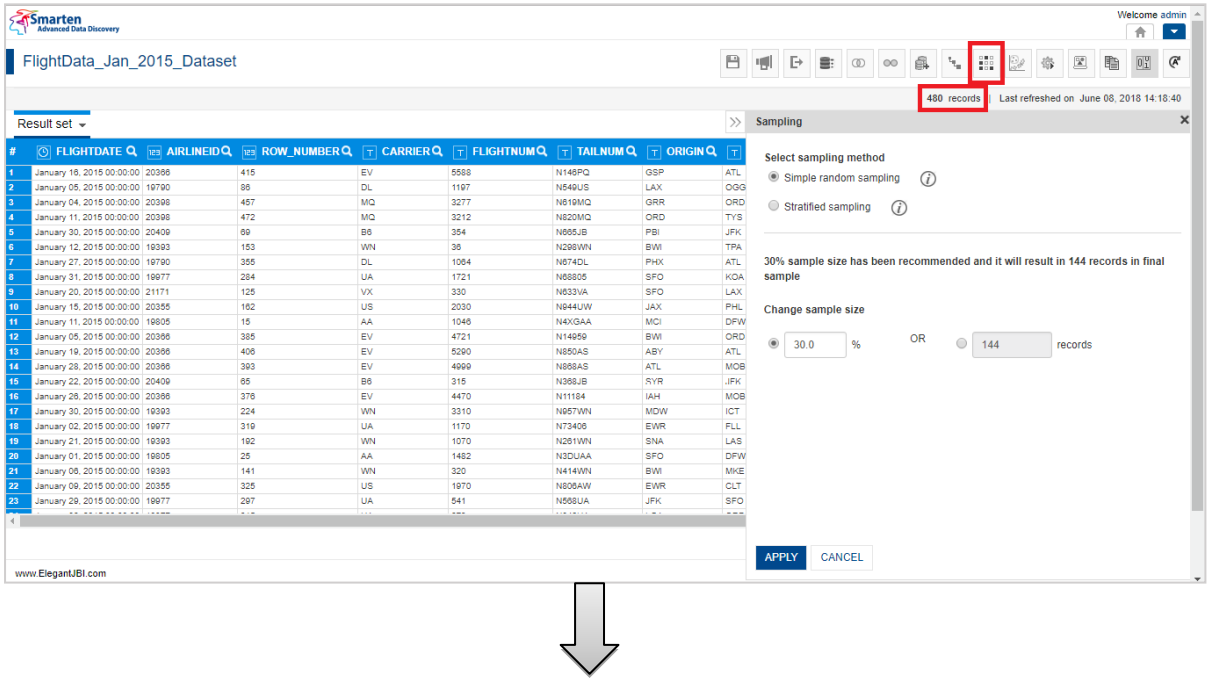
Reference: **Concept Manual > Dataset > Sampling**

7.2.28.1 Sampling—Simple Random Sampling

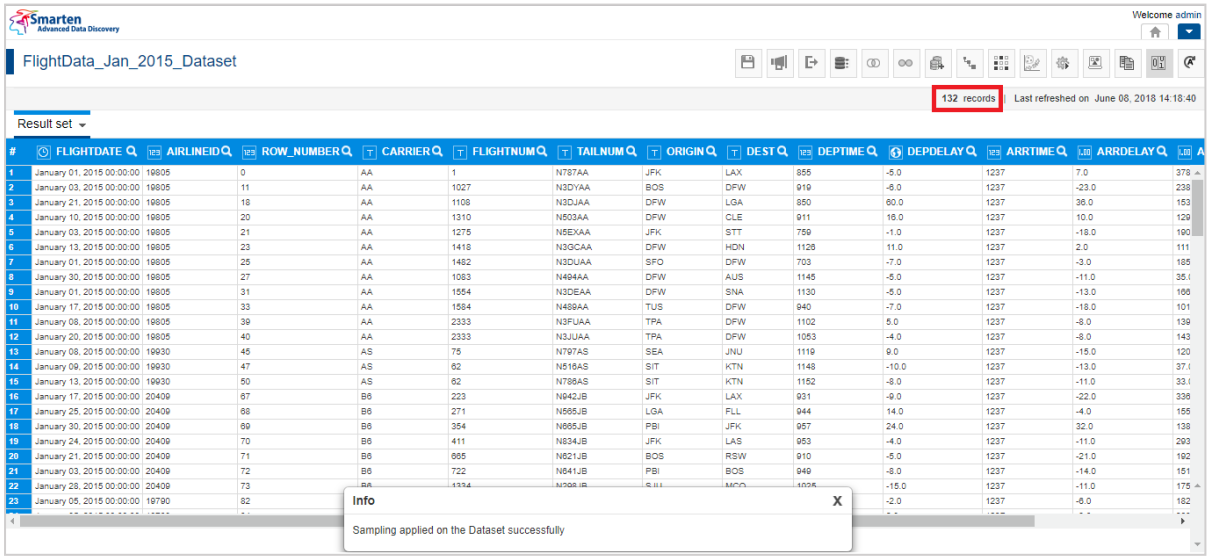
Simple random sampling is a method of sampling in which the selection is determined purely by chance, and every item has an equal chance of being selected. For example, in a lottery system, each member of the population is assigned a number after which the numbers are selected at random.

Shown below is the before and after scenario of “Sampling—Simple random sampling” for a dataset:

Before:



After:



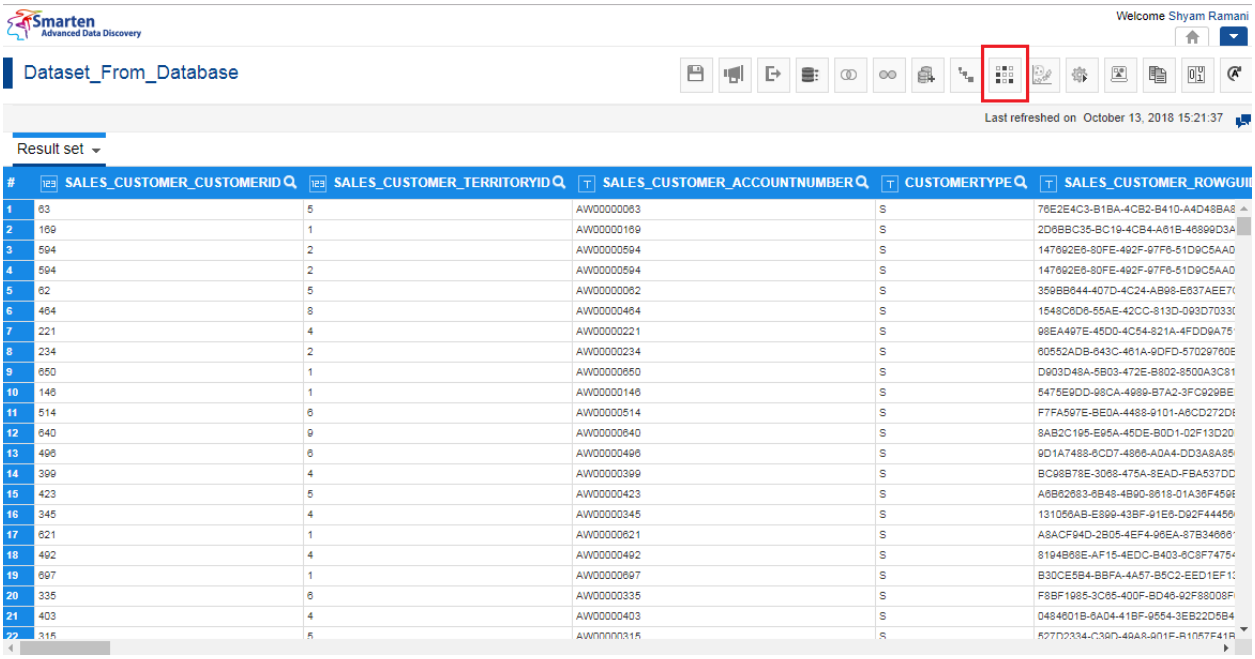
SAMPLING—SIMPLE RANDOM SAMPLING

About this task

Use this task to perform random sampling on the data.

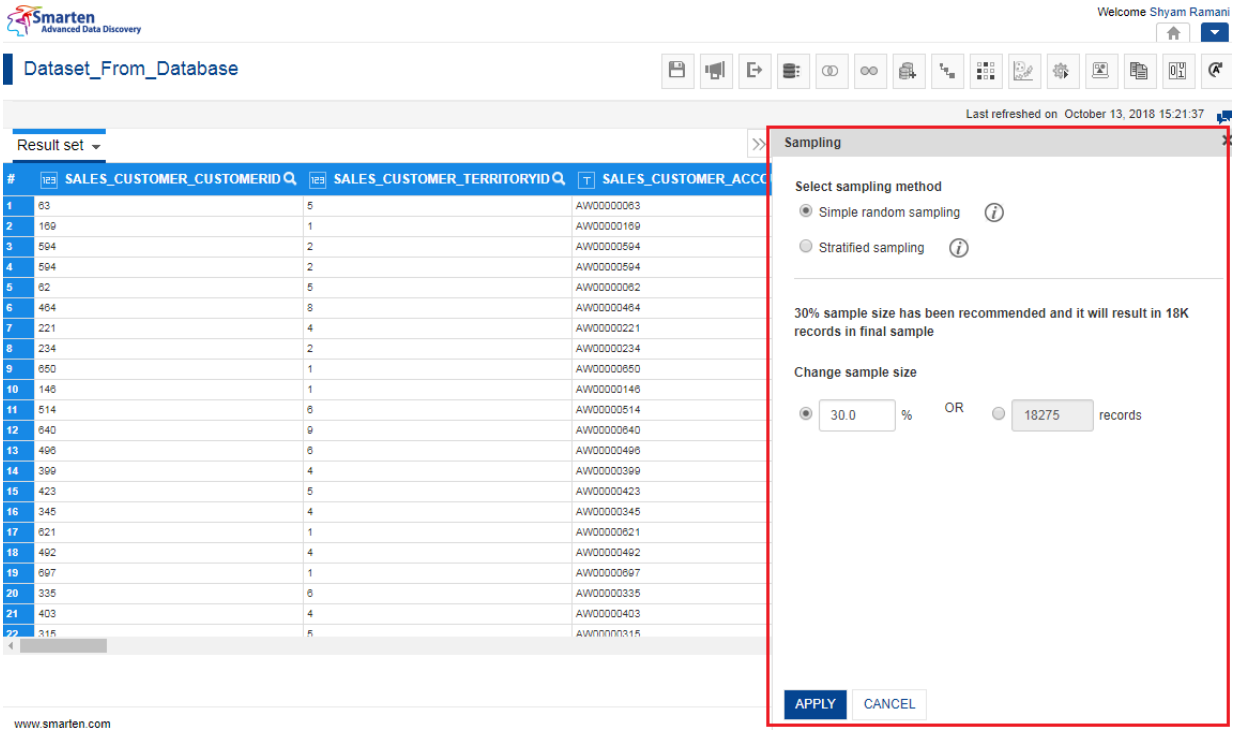
Procedure

1. Open the dataset on which you want to perform random sampling.
2. Click the sampling icon on the toolbar.



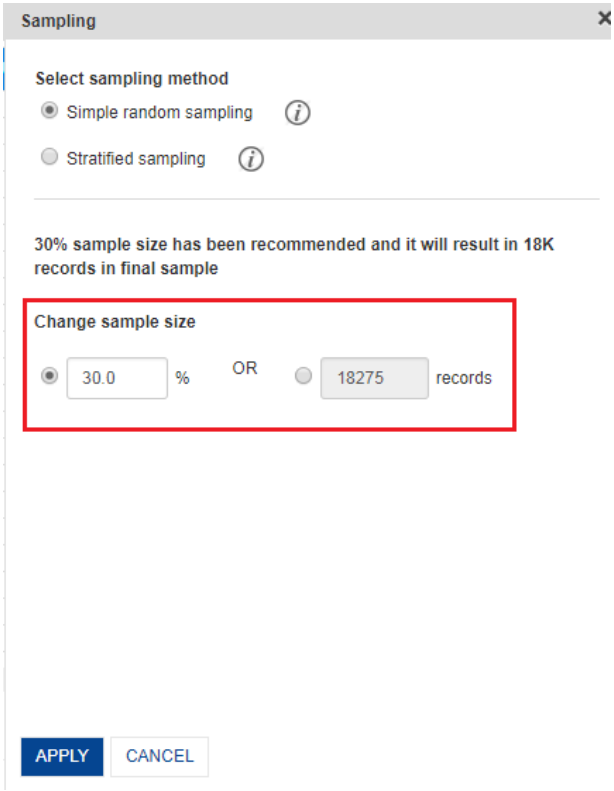
SAMPLING—CLICKING THE SAMPLING ICON

The system displays the **Sampling** dialog box.



SAMPLING—SAMPLING DIALOG BOX

3. Click the **Simple random sampling** option to perform random sampling on the data.
4. Select an option to specify the sample size as a percentage or number of records.



SAMPLING—SPECIFYING SAMPLE SIZE

5. Click **APPLY**.

7.2.28.2 Sampling—Stratified Sampling

Stratified random sampling is a method of sampling that involves the division of a population into smaller groups known as strata. In stratified random sampling, or stratification, the strata are formed based on members' shared attributes or characteristics. For example, subgroups of customers can be formed on the basis of any of their demographics, such as income group, region, and gender. A random sample from each of these subgroups is taken in proportion to the subgroup size relative to the population size, and these samples are then added to form a final stratified random sample. So, if the original dataset had a 1:3 ratio of males and females, the stratified random sample based on gender will also have a 1:3 ratio of males and females.

For example, let us take "CARRIER" as the class to create a sample from a dataset, and there are five subgroups within carriers, which are WN, DL, EV, AA, and UA. Smarten SSDP will take a random sample from each of these subgroups in proportion to the subgroup size relative to the dataset size.

Original data

CARRIER	No of records
WN	101
DL	71
EV	62
AA	50
UA	32

Total number of records in the dataset: 480

30% Sampling

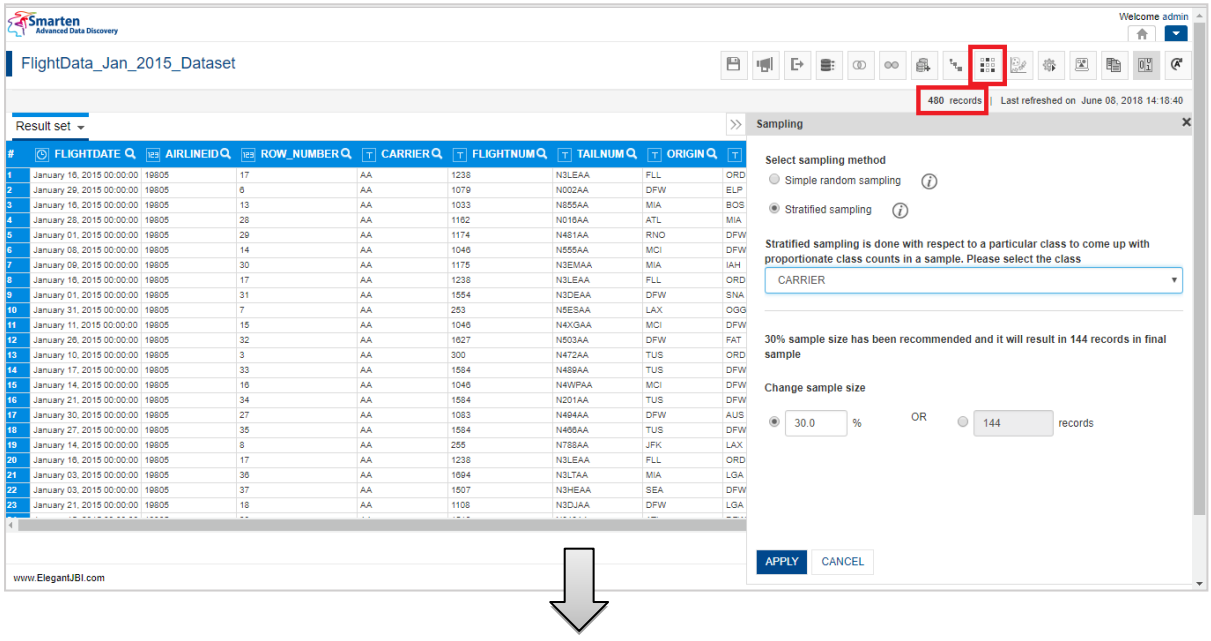
Sample data

CARRIER	No of records (approx)
WN	34
DL	19
EV	17
AA	16
UA	5

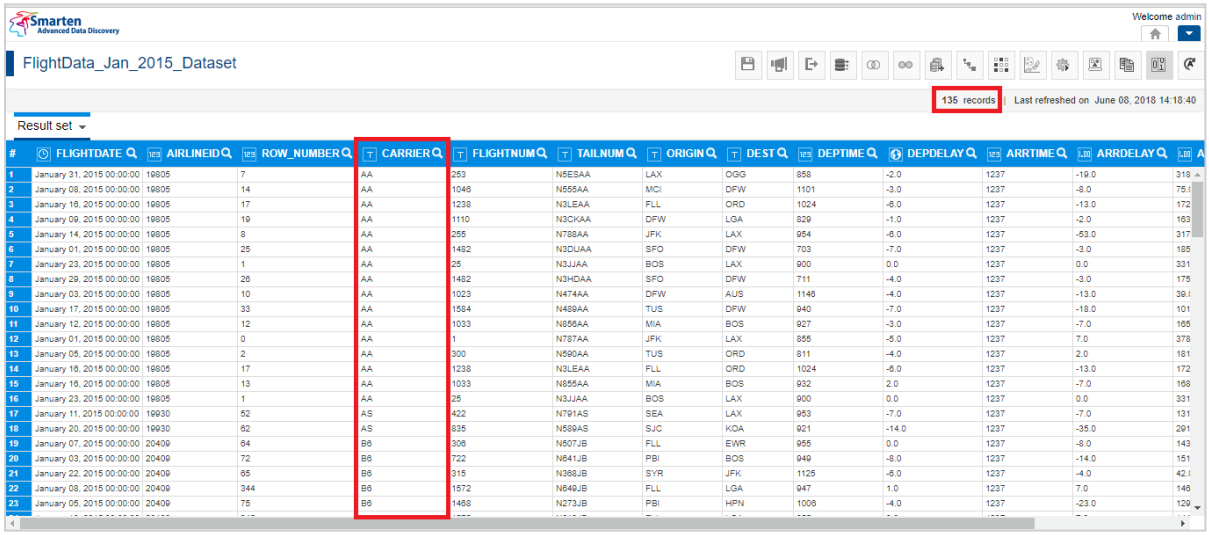
Total number of records in the sample (approx): 135

Shown below is the before and after scenario of "Sampling—Stratified sampling" for a dataset using CARRIER as the subgroup:

Before:



After:



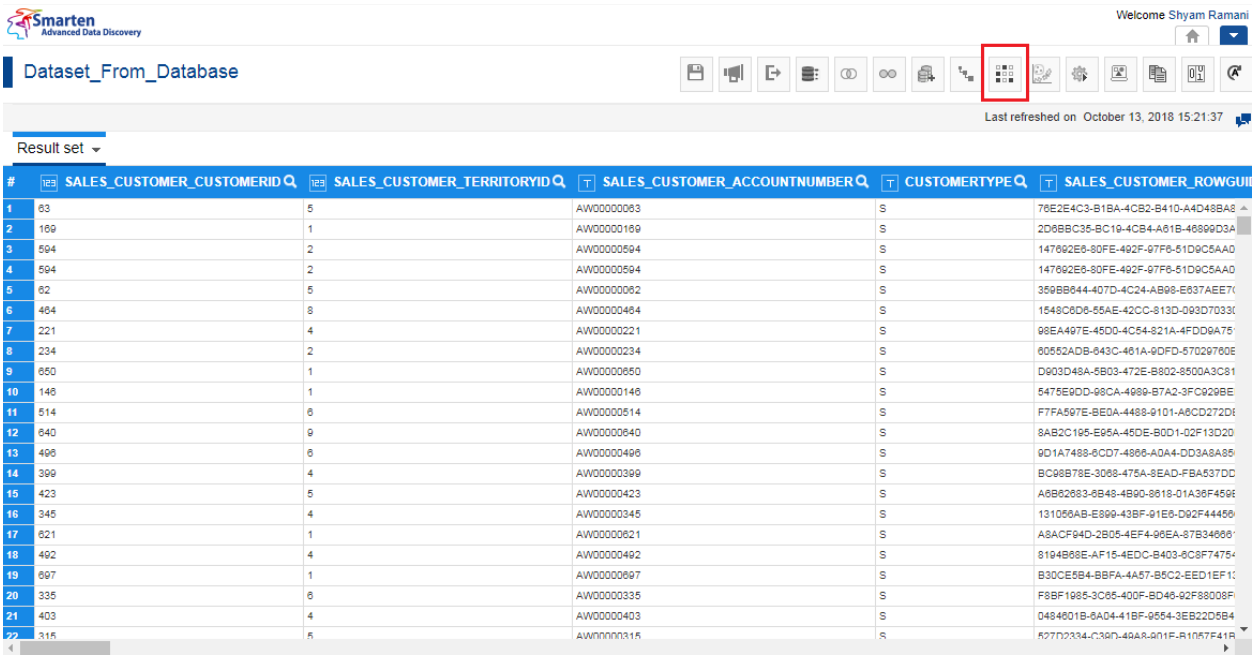
SAMPLING—STRATIFIED SAMPLING

About this task

Use this task to perform stratified sampling on the data.

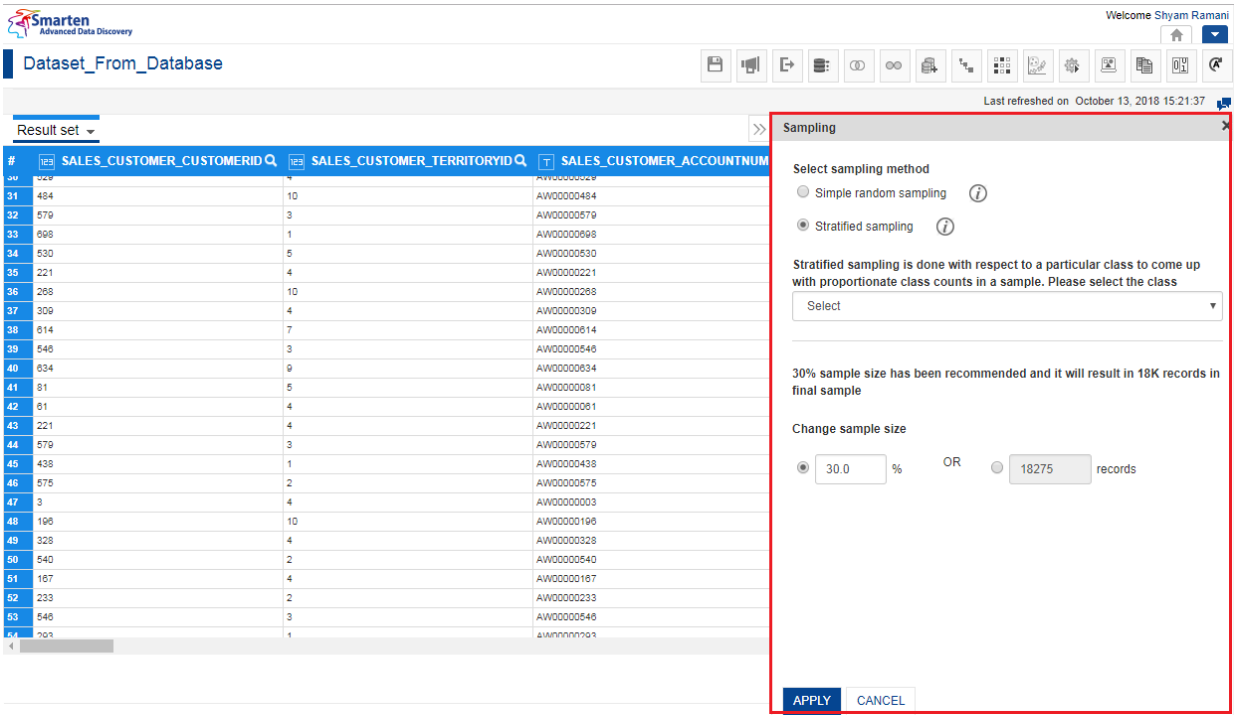
Procedure

1. Open the dataset on which you want to perform stratified sampling.
2. Click the sampling icon on the toolbar.



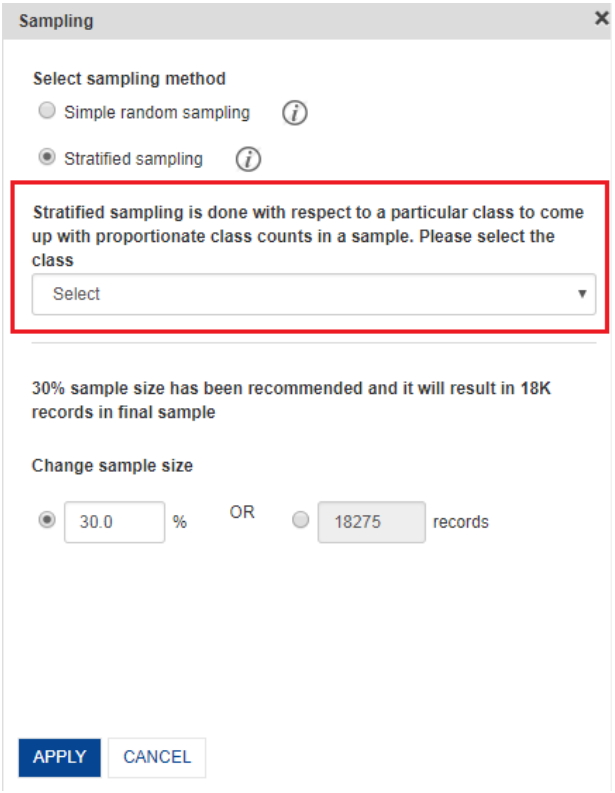
SAMPLING—CLICKING THE SAMPLING ICON

The system displays the **Sampling** dialog box.



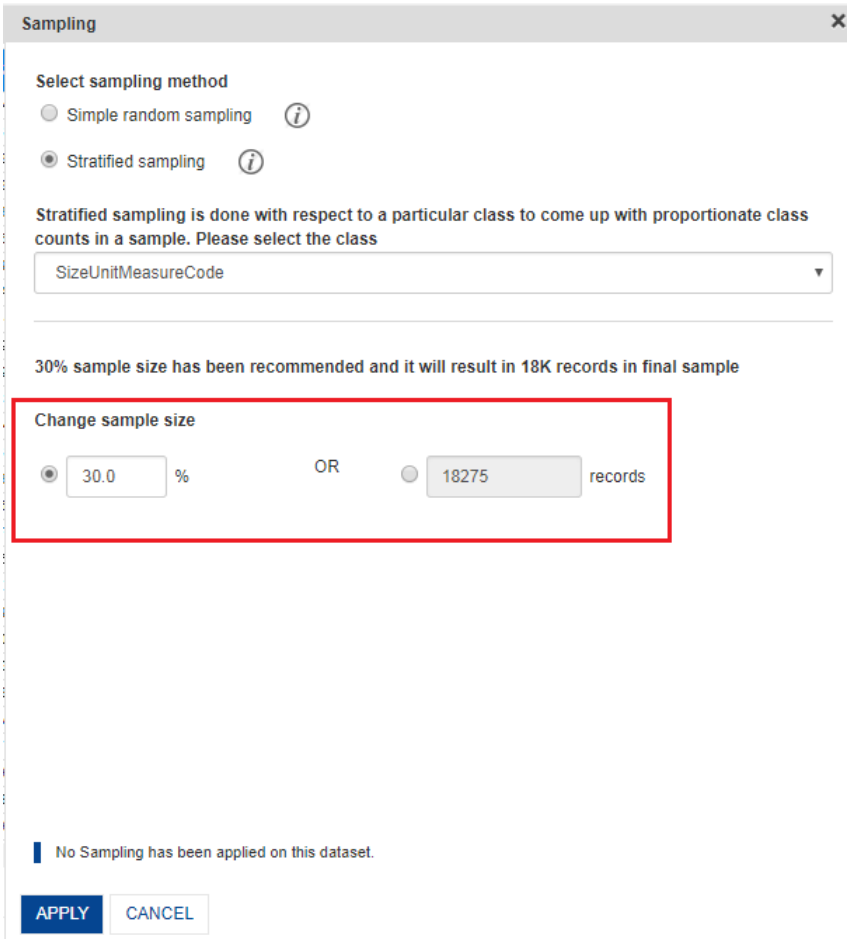
SAMPLING—SAMPLING DIALOG BOX

3. Click the **Stratified sampling** option to perform stratified sampling on the data.
   The system displays the class list.
4. Select an option from the class list to specify the class based on which the sampling must be performed.



SAMPLING—SELECTING THE CLASS

5. Select an option to specify the sample size as a percentage or number of records.



SAMPLING—SPECIFYING SAMPLE SIZE

6. Click **APPLY**.

7.2.29 Adding the Dimension Map

The dimension hierarchy refers to the hierarchical levels of data within a dimension map. Dimension maps can be defined at the cube or dataset level and enable automatic drill down and drill up in front-end objects for users.

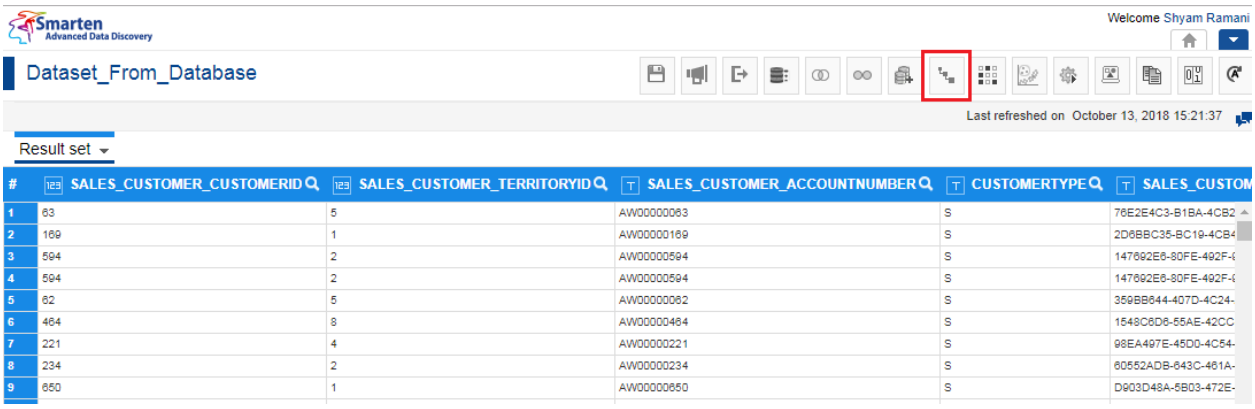
> Reference: **Smarten Concept Manual > Designing the Data Model > Cube Meta Data > Dimension Hierarchy**

About this task

Use this task to add a dimension map.

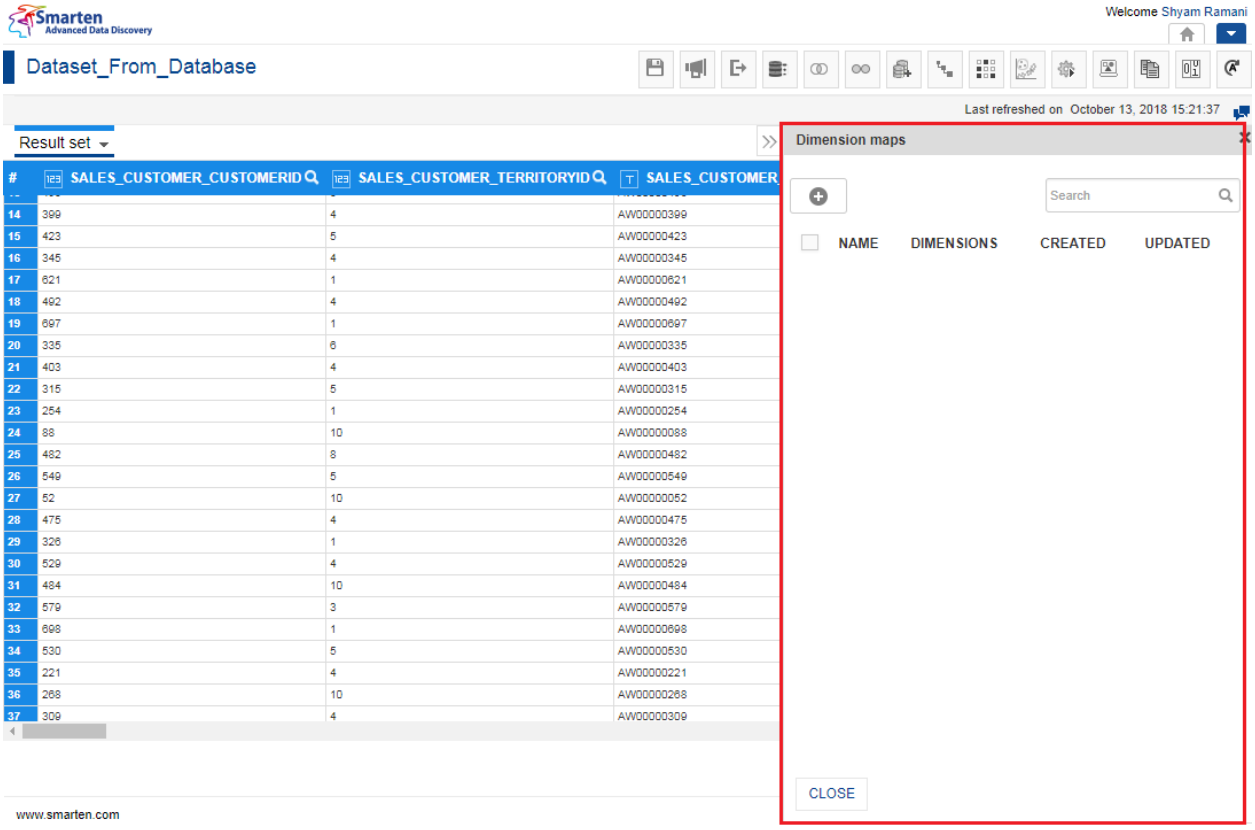
Procedure

1. Open the dataset for which you want to add a dimension map.
2. Click the Dimension map icon on the toolbar.



DIMENSION MAP—OPENING THE DIMENSION MAPS DIALOG BOX

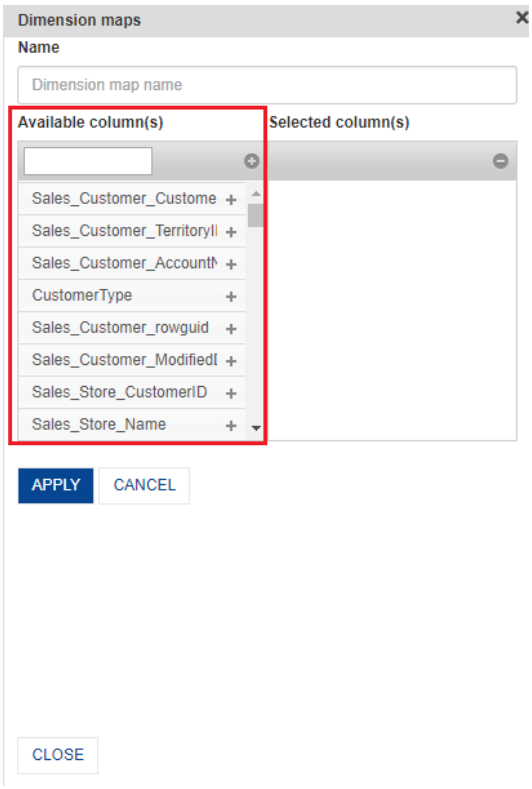
The system displays the **Dimension maps** dialog box.



DIMENSION MAP—THE DIMENSION MAPS DIALOG BOX

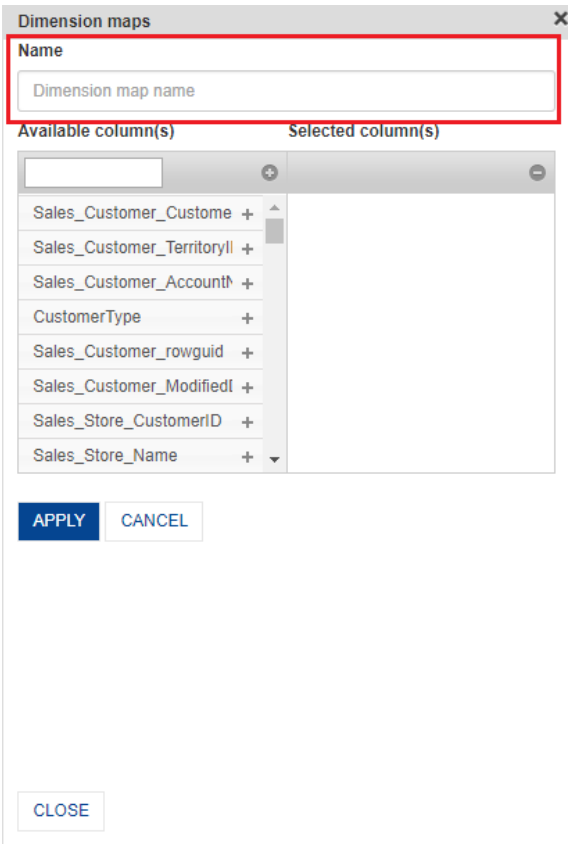
3. Click the Add icon to add a dimension map.

   The system displays all the columns for which a dimension map can be added in the **Available column(s)** section.



DIMENSION MAP—AVAILABLE COLUMNS FOR WHICH A DIMENSION MAP CAN BE ADDED

4. Specify a name for the dimension map in the **Name** box.



DIMENSION MAP—SPECIFY A NAME FOR THE DIMENSION MAP

5. Click the plus sign adjacent to the columns for which you want to add a dimension map from the **Available column(s)** section, and then click **APPLY**.

7.2.29.1 Editing a Dimension Map

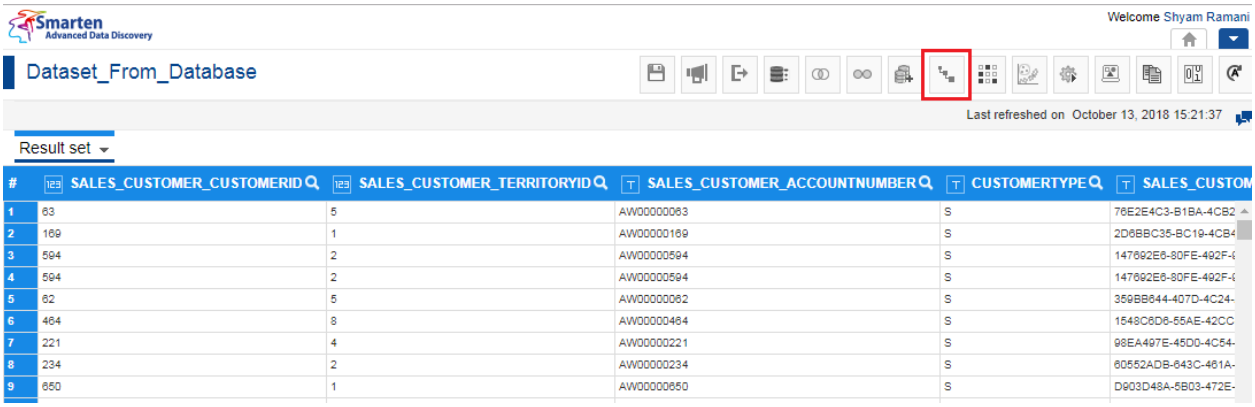
Smarten allows you to edit an existing dimension map. You can add or remove the columns that are used to create a dimension map.

About this task

Use this task to edit a dimension map.

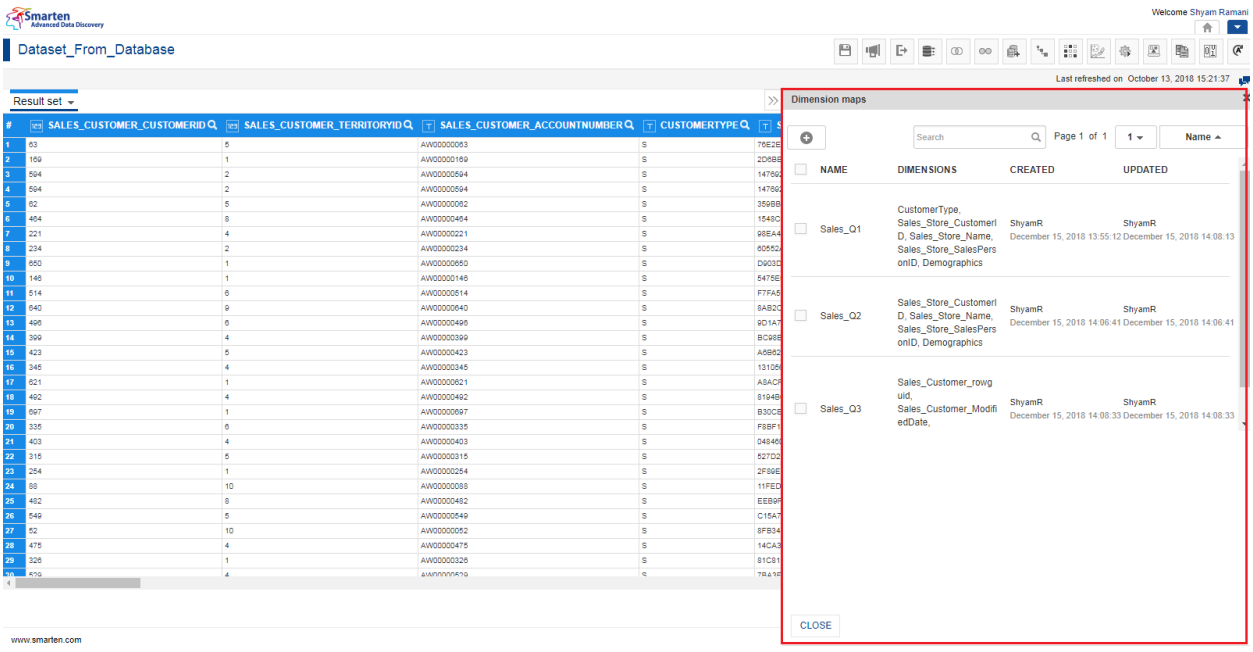
Procedure

1. Open the dataset for which you want to edit a dimension map.
2. Click the Dimension map icon on the toolbar.



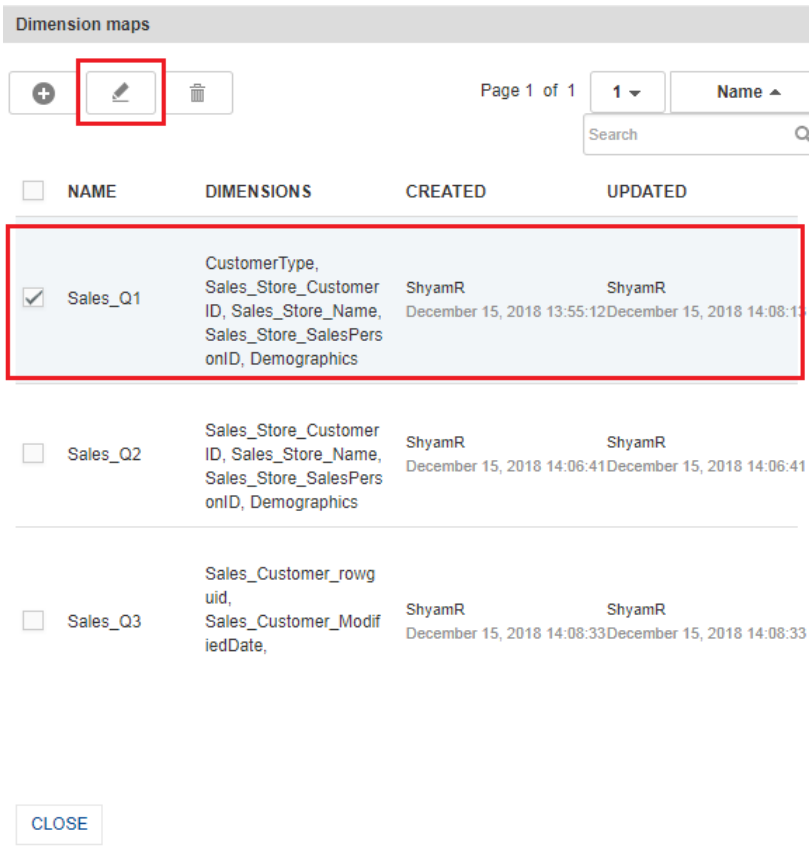
DIMENSION MAP—OPENING THE DIMENSION MAPS DIALOG BOX

The system displays the **Dimension maps** dialog box.



DIMENSION MAP—THE DIMENSION MAPS DIALOG BOX.

3. Select the dimension map you want to edit, and then click the Edit icon.



DIMENSION MAP—EDITING A DIMENSION MAP

4. Click **CLOSE**.

7.2.29.2 Deleting a Dimension Map

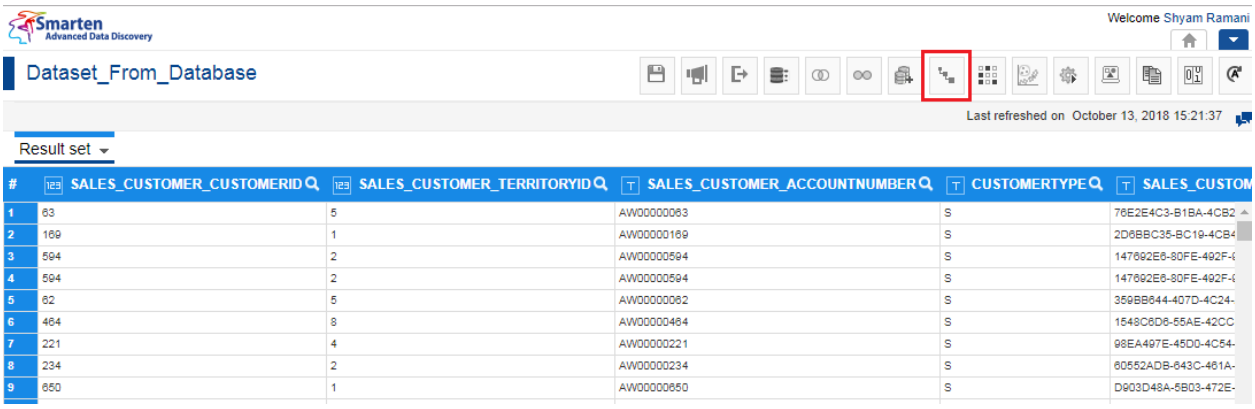
You can delete an existing dimension map.

About this task

Use this task to delete an existing dimension map.

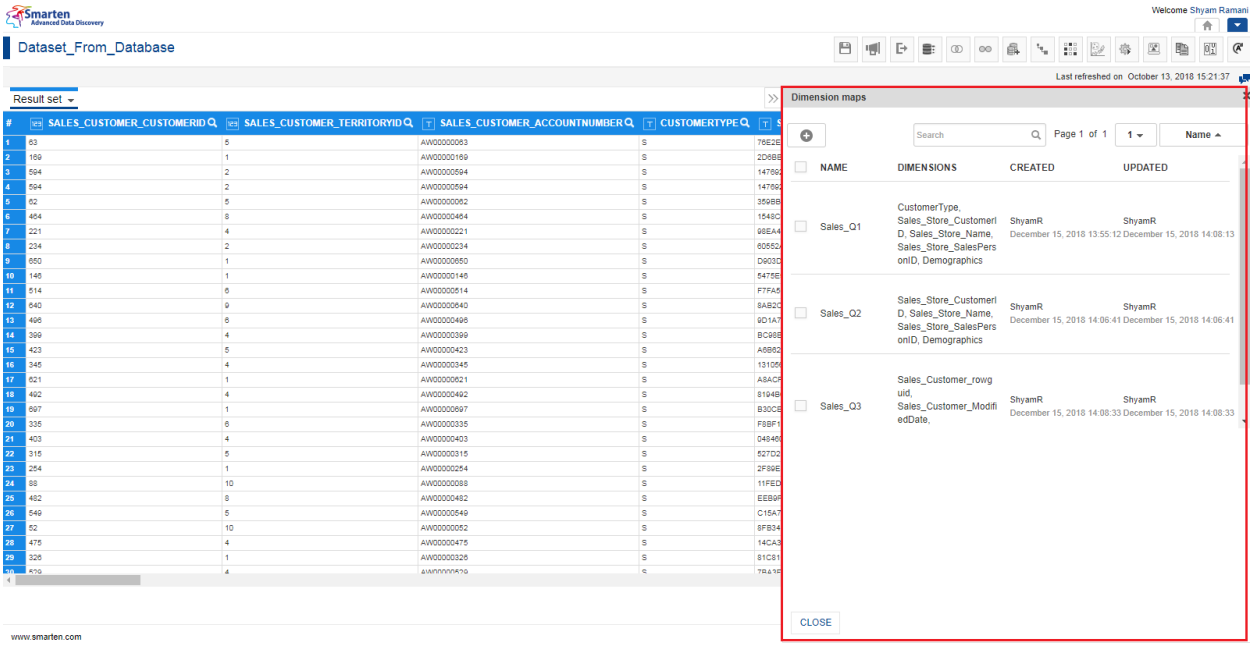
Procedure

1. Open the dataset for which you want to edit a dimension map.
2. Click the Dimension map icon on the toolbar.



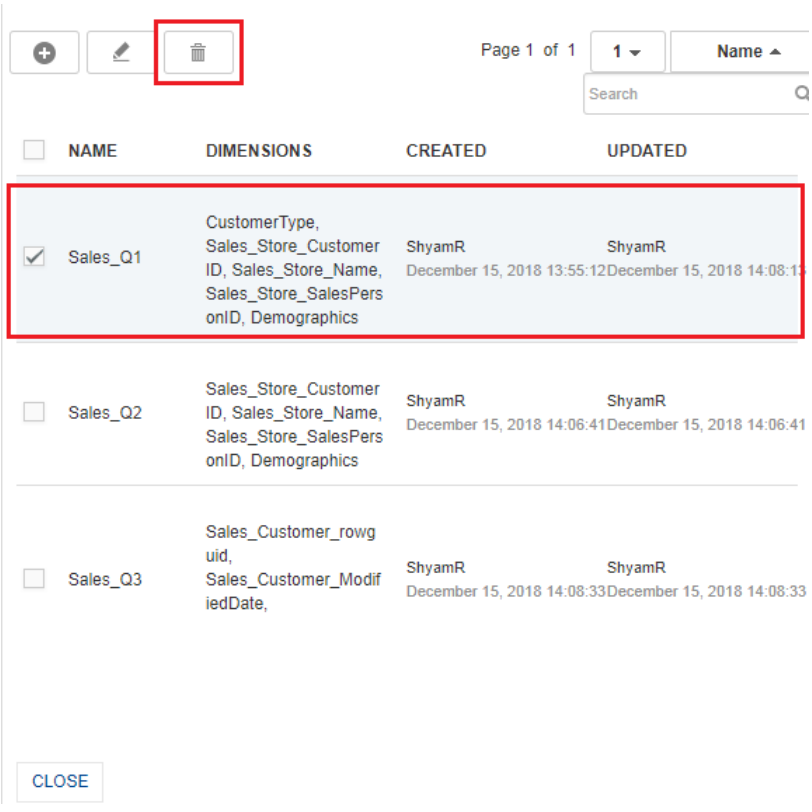
DIMENSION MAP—OPENING THE DIMENSION MAPS DIALOG BOX

The system displays the **Dimension maps** dialog box.



DIMENSION MAP—THE DIMENSION MAPS DIALOG BOX.

3. Select the dimension map you want to delete, and then click the Delete icon.



DIMENSION MAP—DELETING A DIMENSION MAP

4. Click **CLOSE**.

7.2.30 Adding a Dataset

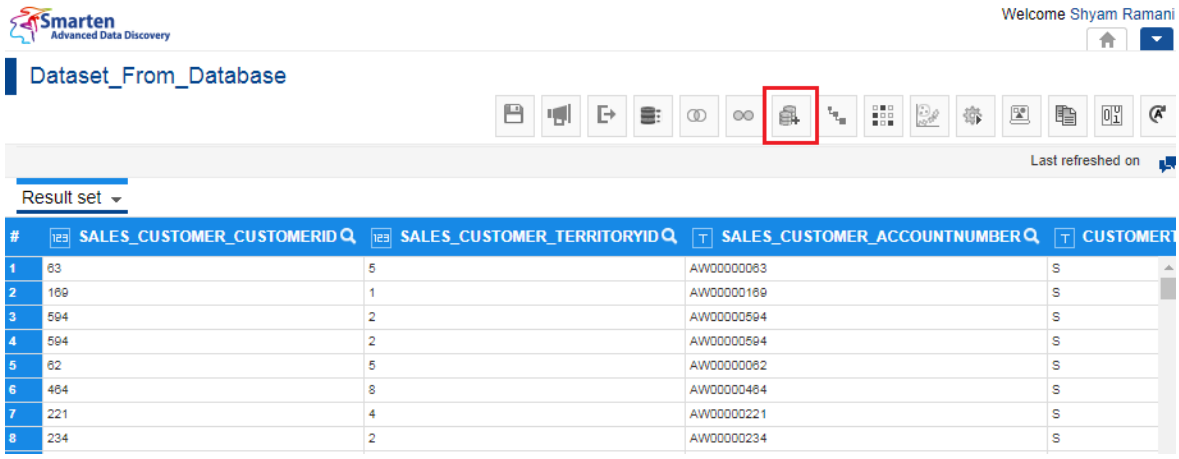
You can add other datasets to the current dataset view. Users can use these datasets to perform blend operations, such as Append and JOIN. These datasets are in read-only mode, and users can view or explore data of these datasets.

About this task

Use this task to add a dataset.

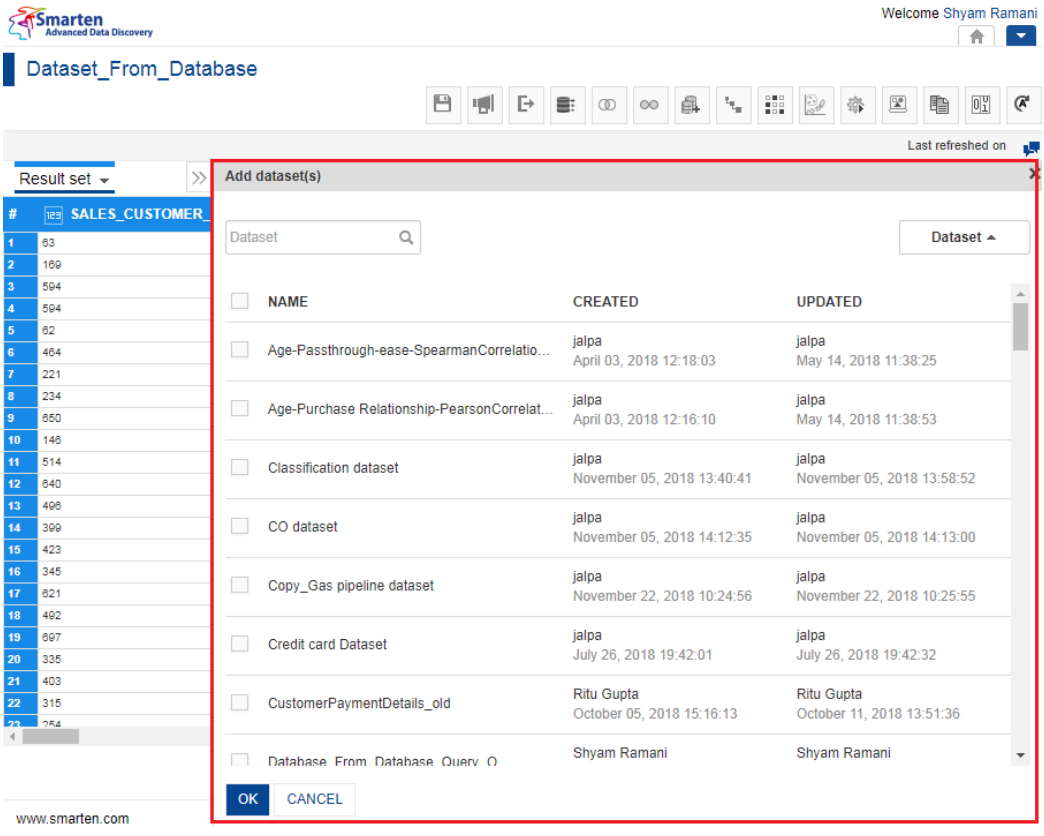
Procedure

1. Open the dataset for which you want to add a dataset.
2. Click the Add dataset icon on the toolbar.



ADDING A DATASET—CLICKING THE ADD DATASET ICON

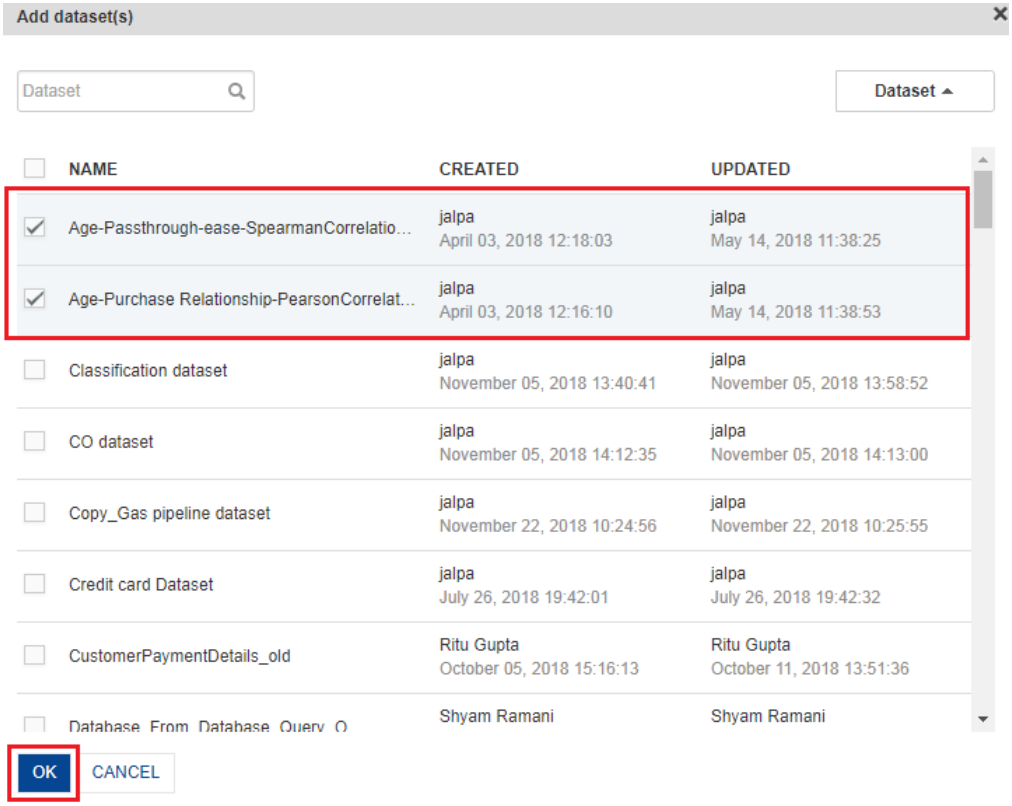
The system displays the **Add dataset(s)** dialog box.



ADD A DATASET—ADD DATASET DIALOG BOX

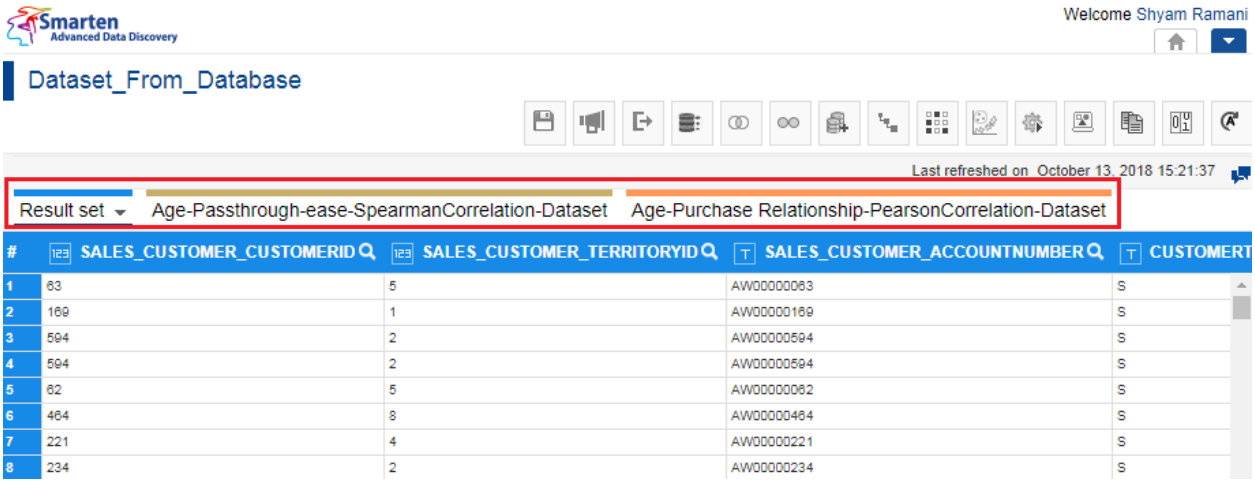
3. Select the datasets you want to add, and then click **OK**.

   You can add one or more datasets to the current dataset.



ADD A DATASET—ADDING DATASETS

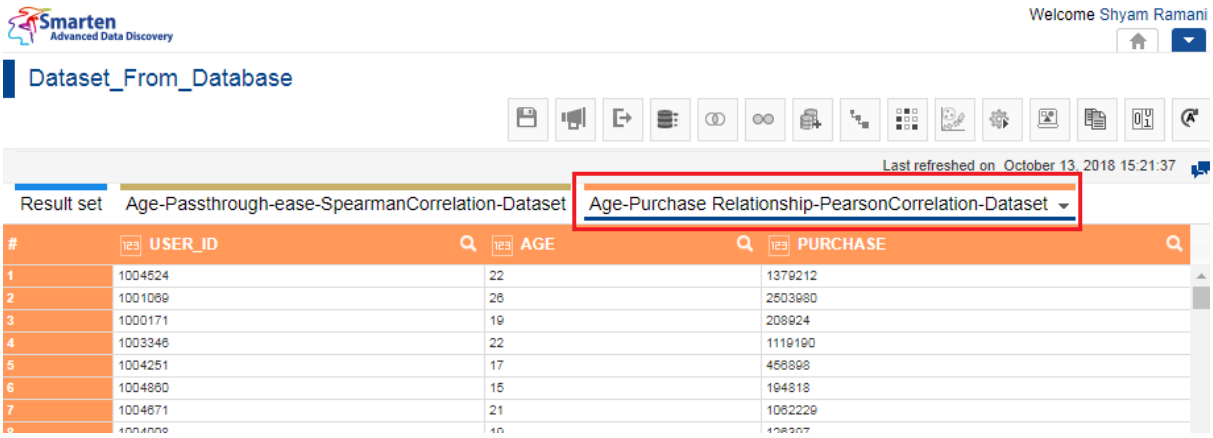
4. The added dataset is displayed adjacent to the current dataset.



ADD A DATASET—ADDED DATASETS

   The system highlights the added datasets in a different color from the current dataset.

5. You can click the dataset name to view data of that dataset.



ADD A DATASET—VIEWING DATA OF AN ADDED DATASET

7.2.30.1 Removing an Added Dataset

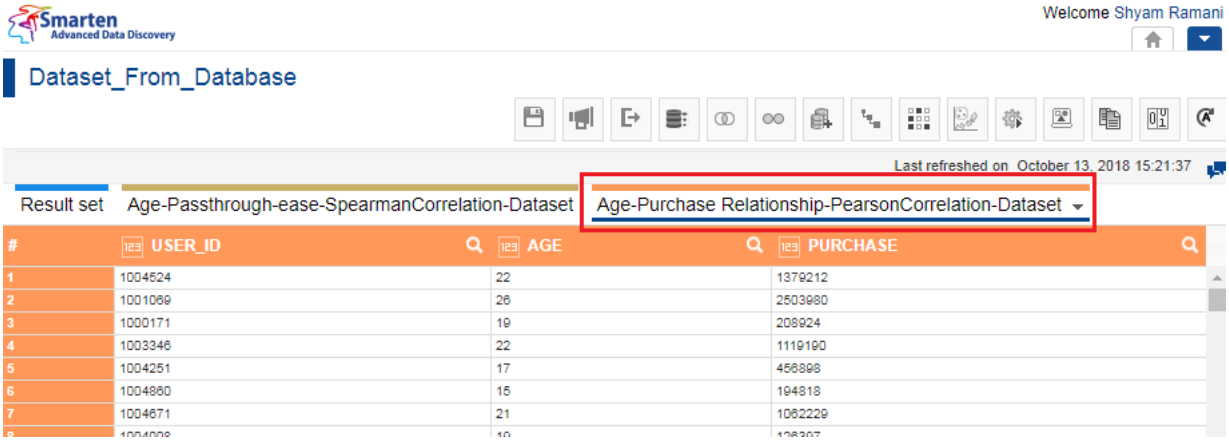
You can remove a dataset added to another dataset.

About this task

Use this task to remove a dataset added to another dataset.

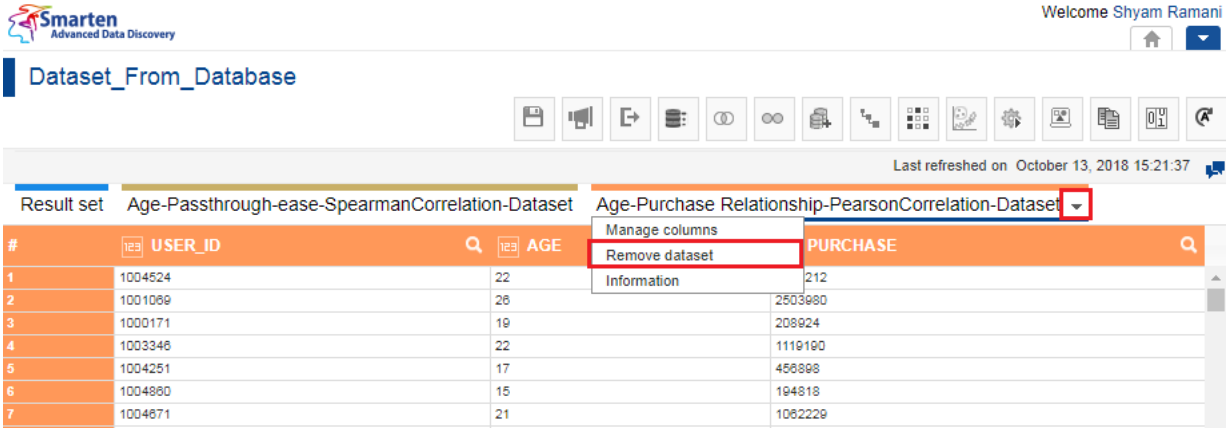
Procedure

1. Open the dataset from which you want to remove an added dataset.
2. Click the dataset you want to remove.



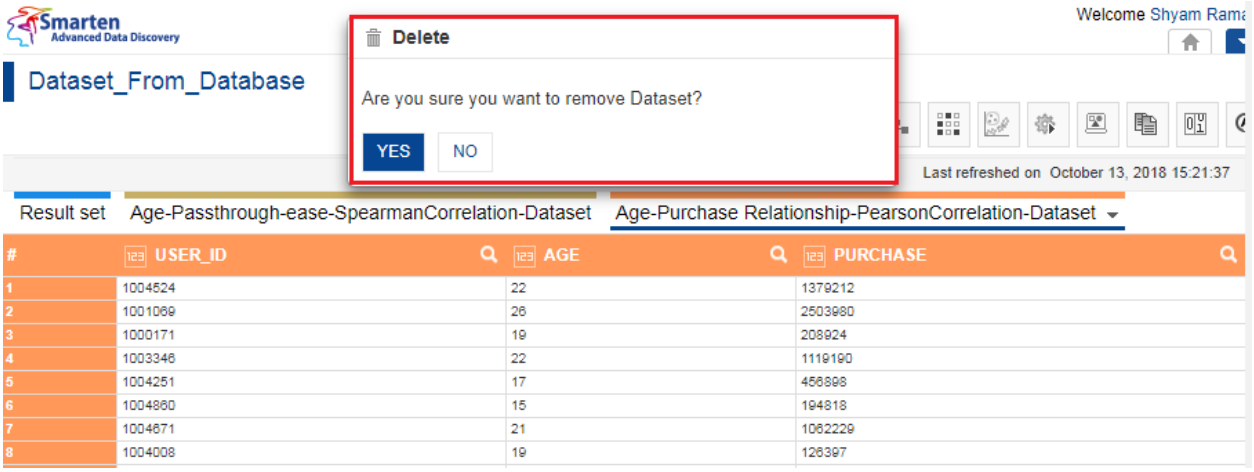
REMOVE A DATASET—SELECTING THE DATASET TO BE REMOVED

3. Click the Menu icon for the selected dataset, and then click Remove dataset.



REMOVE A DATASET —REMOVING A DATASET

The system removed the dataset after confirmation.



REMOVE A DATASET—CONFIRMATION TO REMOVE A DATASET

7.2.31 Blending Data

Users can blend data from two or more datasets based on common fields between them. Assisted by the auto-suggestions and recommendations by Smarten SSDP, users are able to blend the data quickly, efficiently, and without any formal knowledge of SQL or scripting.

7.2.31.1 Blending Data—Append Operation

Using this functionality, users can append datasets with the help of auto-match column suggestions provided by Smarten SSDP. Users are also able to create a copy of resultset before applying the blend operation.

Reference:	**Concept Manual > Blend Data > Append**

About this task

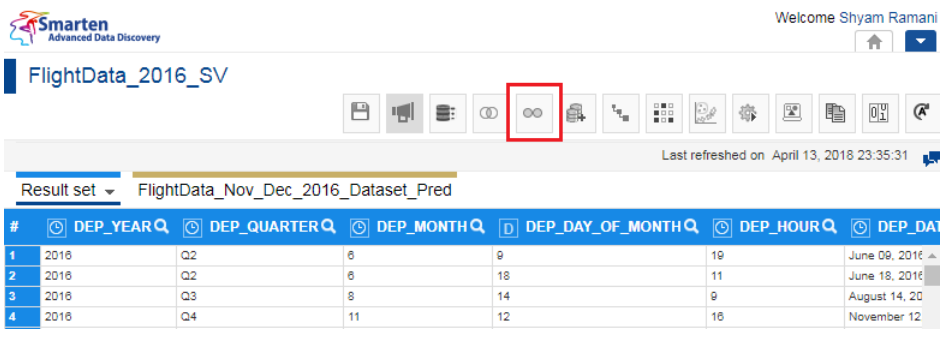
Use this task to append a dataset.

Procedure

1. Open the dataset you want to append with another dataset.

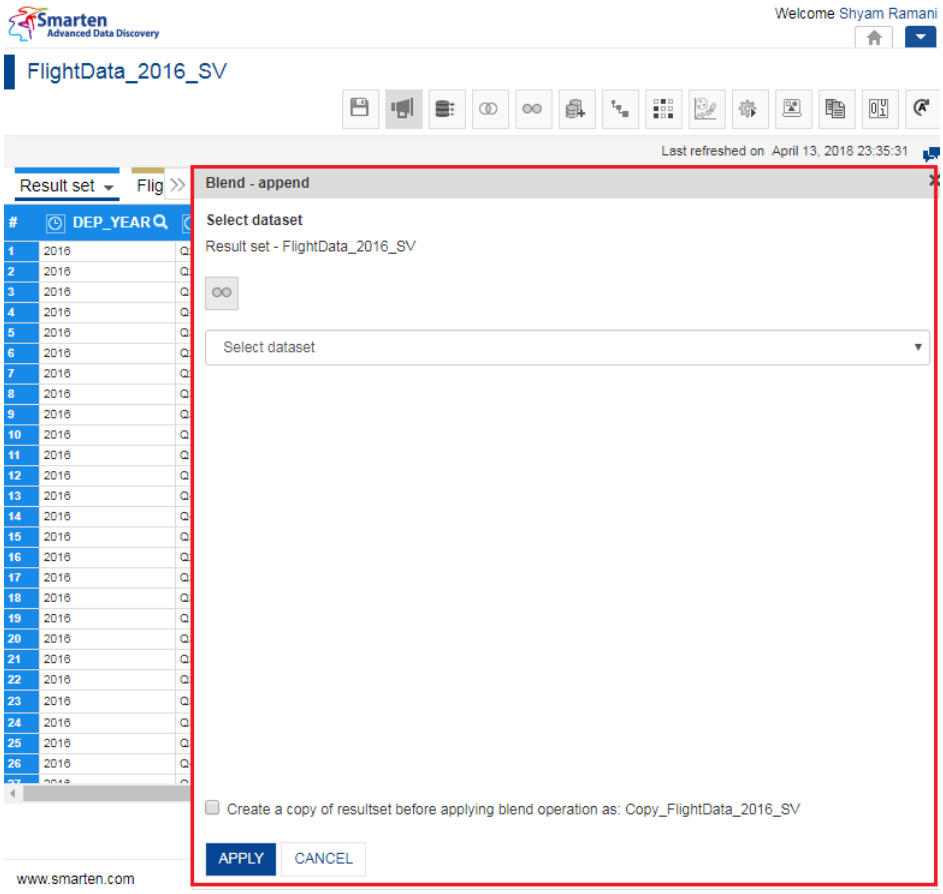
> Note:
> You must add datasets that you want to use before performing the blend operation.

2. Click the Blend - append icon on the toolbar.



BLEND APPEND—OPENING THE BLEND - APPEND DIALOG BOX

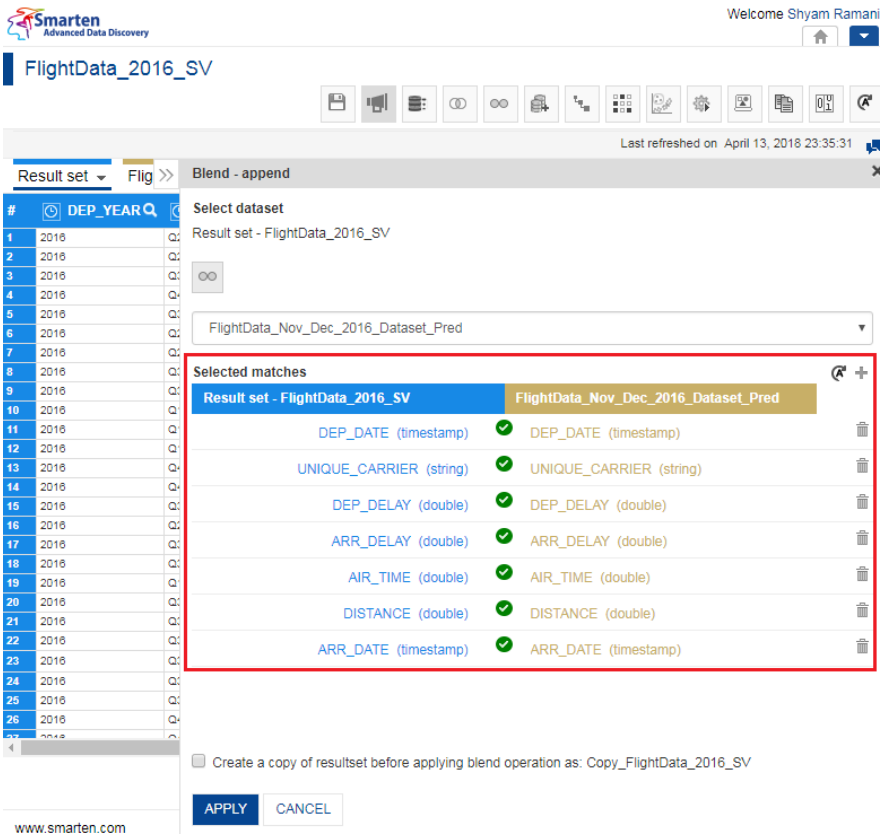
The system displays the **Blend - append** dialog box.



BLEND APPEND—THE BLEND - APPEND DIALOG BOX

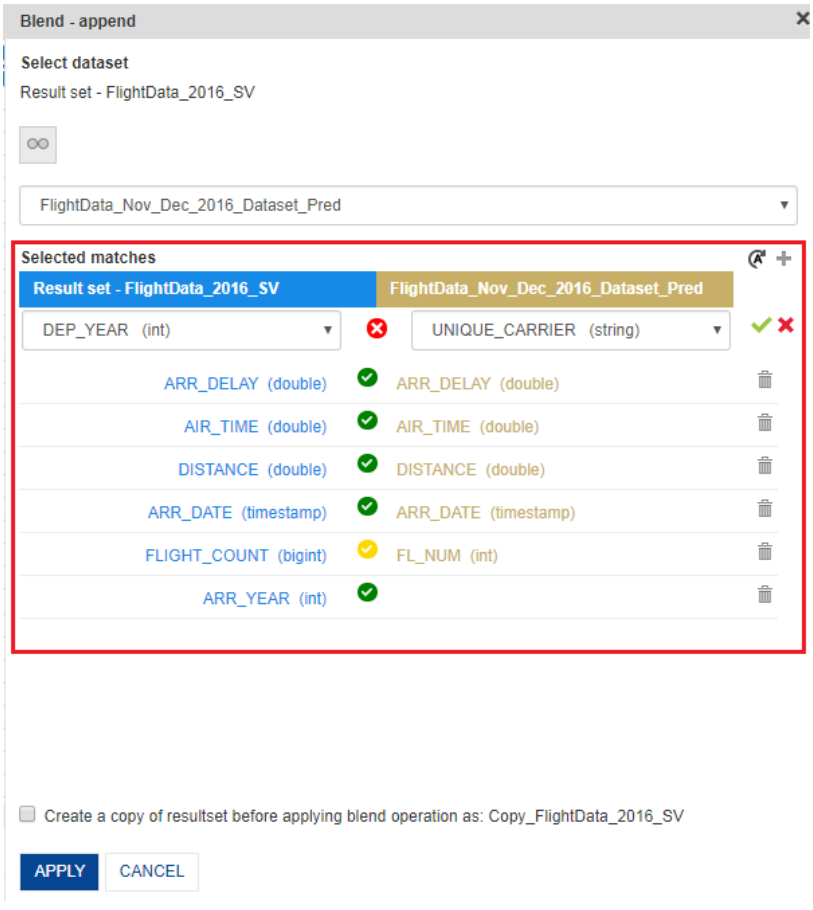
3. Select the dataset you want to append to the current dataset from the **Select dataset** list.

> Note:
> The datasets available in the list are the datasets you have added to the current dataset. If you have not added any datasets to the current dataset, no option will be available in the list.



APPEND A DATASET—SELECT A DATASET THAT WILL BE APPENDED TO THE CURRENT DATASET

The system automatically displays a list of columns with matching criteria.



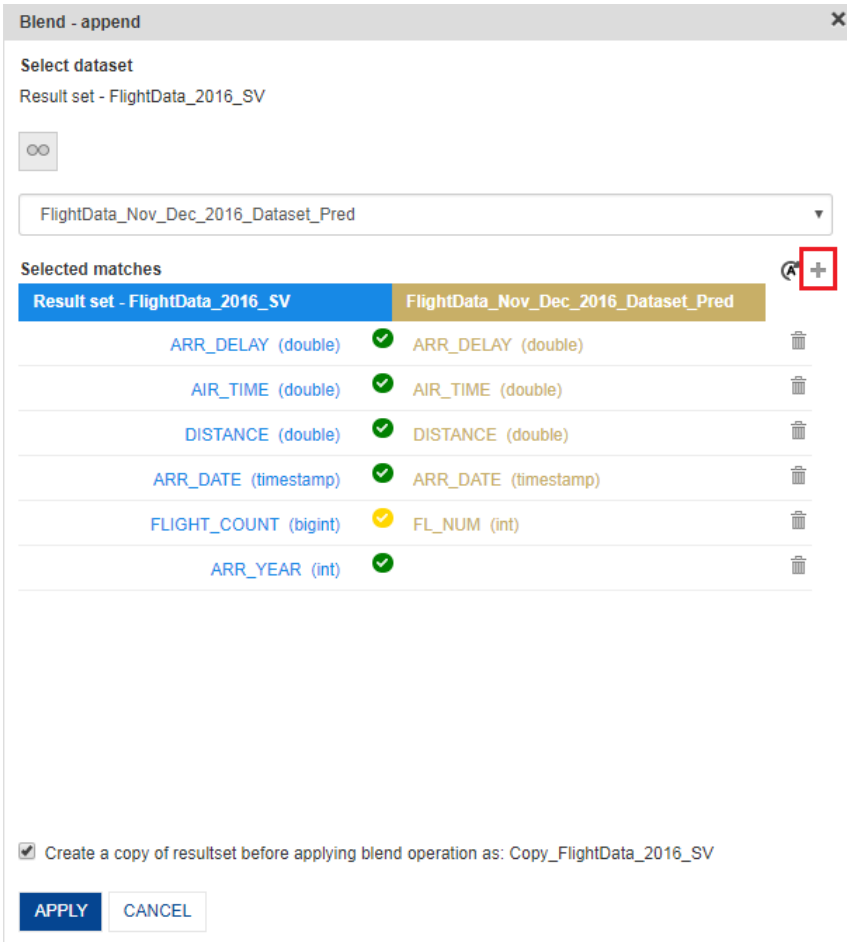
APPEND A DATASET—LIST OF COLUMNS WITH MATCHING VALUES

The system displays a check mark highlighted with the following colors:

- **Green**: If the data type of the selected columns is the same. For example, in the above image, the AIR_TIME column from both datasets is of the double data type.
- **Yellow**: If the data type of the selected columns is similar. Similar data type refers to the data types that are not exactly same, but the data in both columns are similar. For example, number data type may be of Integer, Big Integer, Float, or Double. For example, in the image above, FLIGHT_COUNT and FL_NUM are of big integer and integer data type respectively and can be considered as similar data types.
  When the append operation is applied on similar data types, the system applies the upper data types casting among the data type of the selected columns on the appended columns. In the example above, the system will apply data type of FLIGHT_COUNT columns as big integer is upper data type than the integer data type.
- **Red**: If the data type of the selected columns do not match.

4. Apart from the automatically suggested column matches, the system allows you to add other matching columns having similar data type.

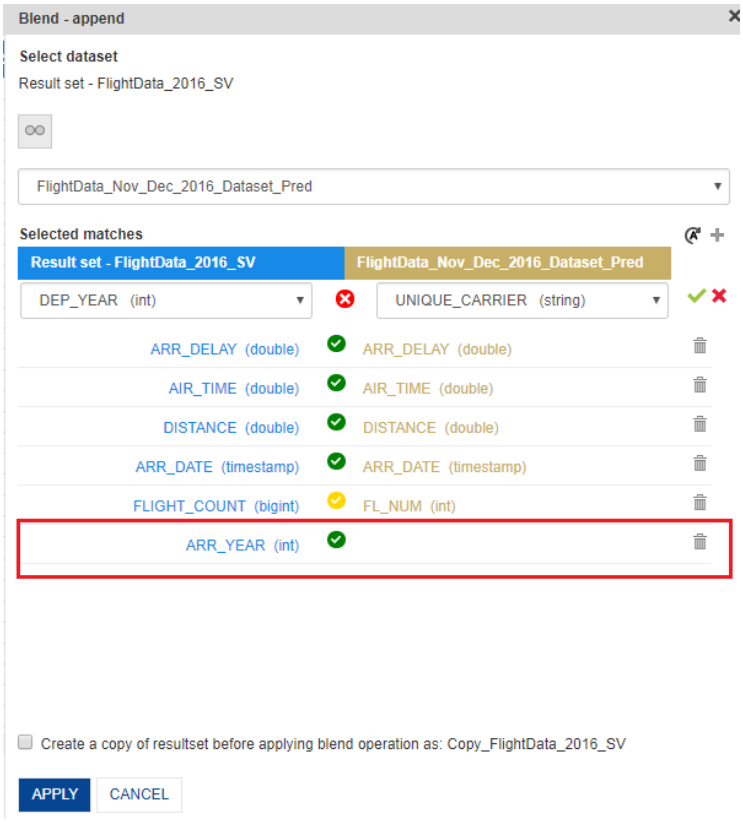
   You can click the Add icon for column match from both the datasets to create another match criteria.



APPEND A DATASET—SELECT COLUMNS FOR MATCH CRITERIA

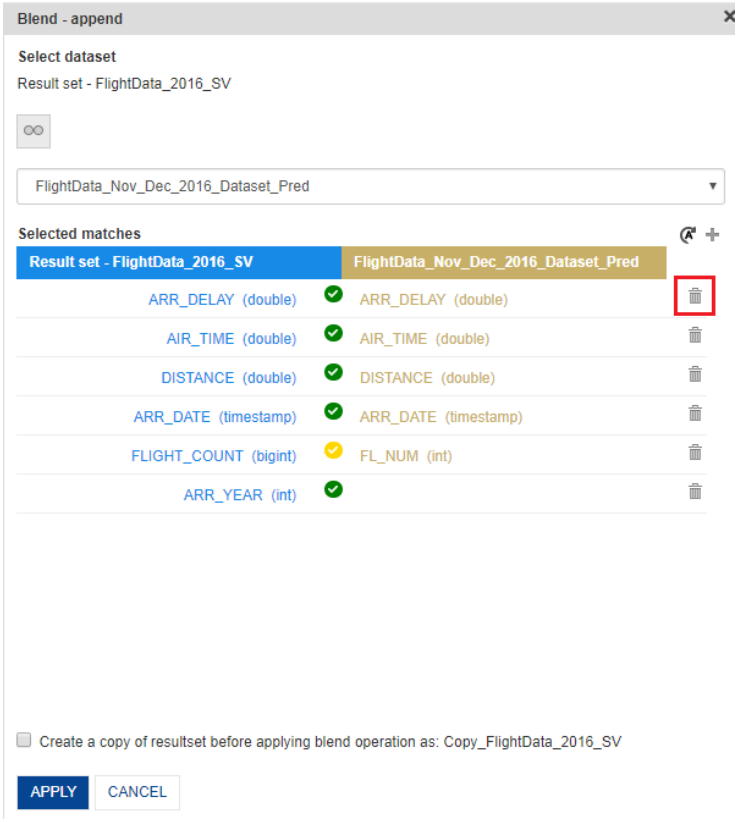
You can add column match criteria in which you select a column from only one dataset and do not select any column from another dataset. For such column match criteria, the system appends data of the selected column with null values. For example, in the image below,

ARR_YEAR is selected for FlightData_2016_SV dataset, but no column is selected for FlightData_Nov_Dec_2016_Dataset_Pred dataset. In this example, the system will append data of the ARR_YEAR column with null values.



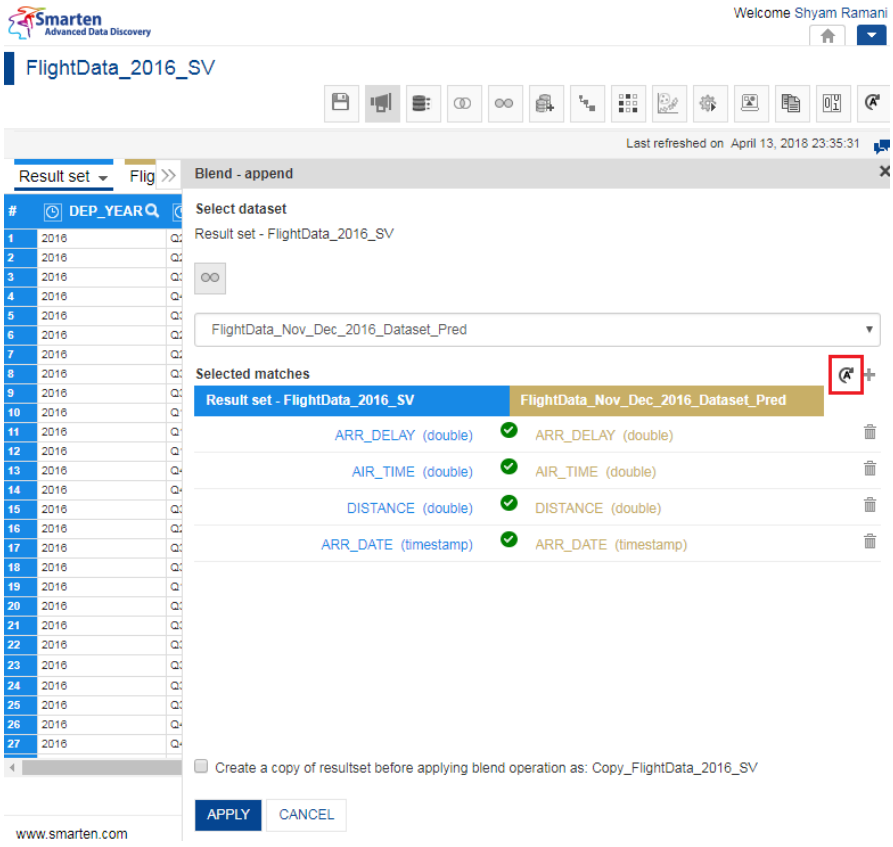
APPEND A DATASET—CRITERIA WITH A SINGLE COLUMN SELECTED

5. You can click the Delete icon adjacent to match criteria to delete that criteria.



APPEND A DATASET—DELETING MATCH CRITERIA

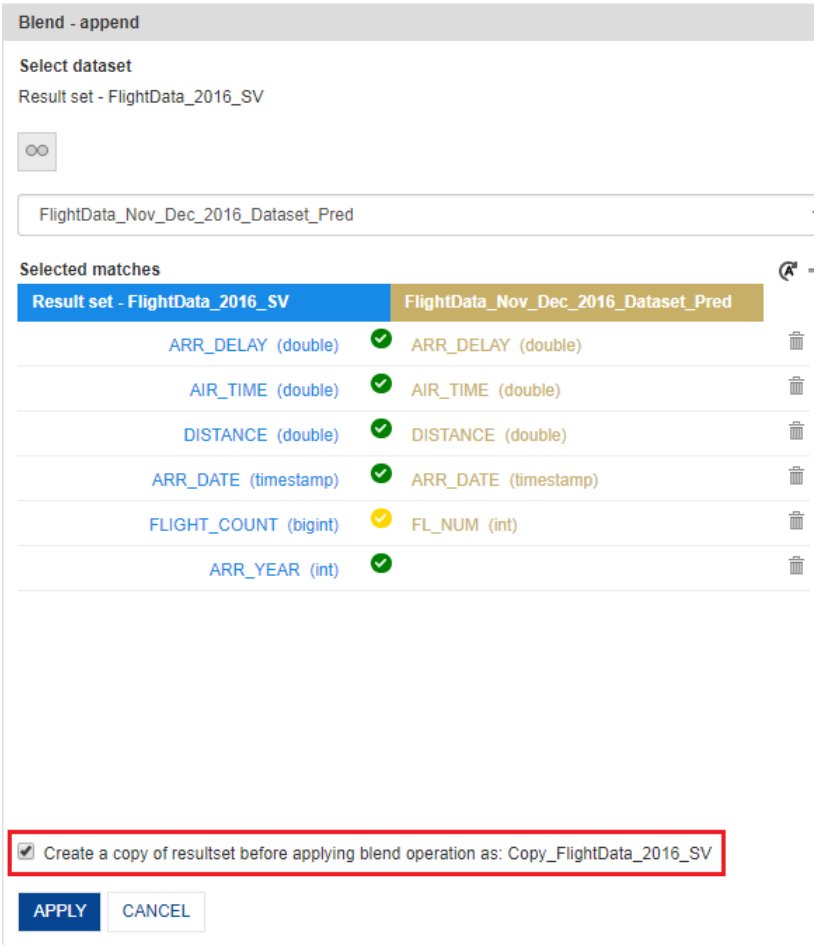
6. You can click the Auto icon to restore the auto-suggested matches.
   The system retains the new match criteria you added.



APPEND A DATASET—RESTORING AUTO-SUGGESTED MATCHES

7. You can select the **Create a copy of the resultset before applying the blend operation** as an option to create a copy of the resultset before the append operation is performed.

   The system creates a copy of the current dataset with the name displayed along the option. For example, in the image above, the system will create a copy of the current resultset with the name “Copy_Dataset_From_Database.”



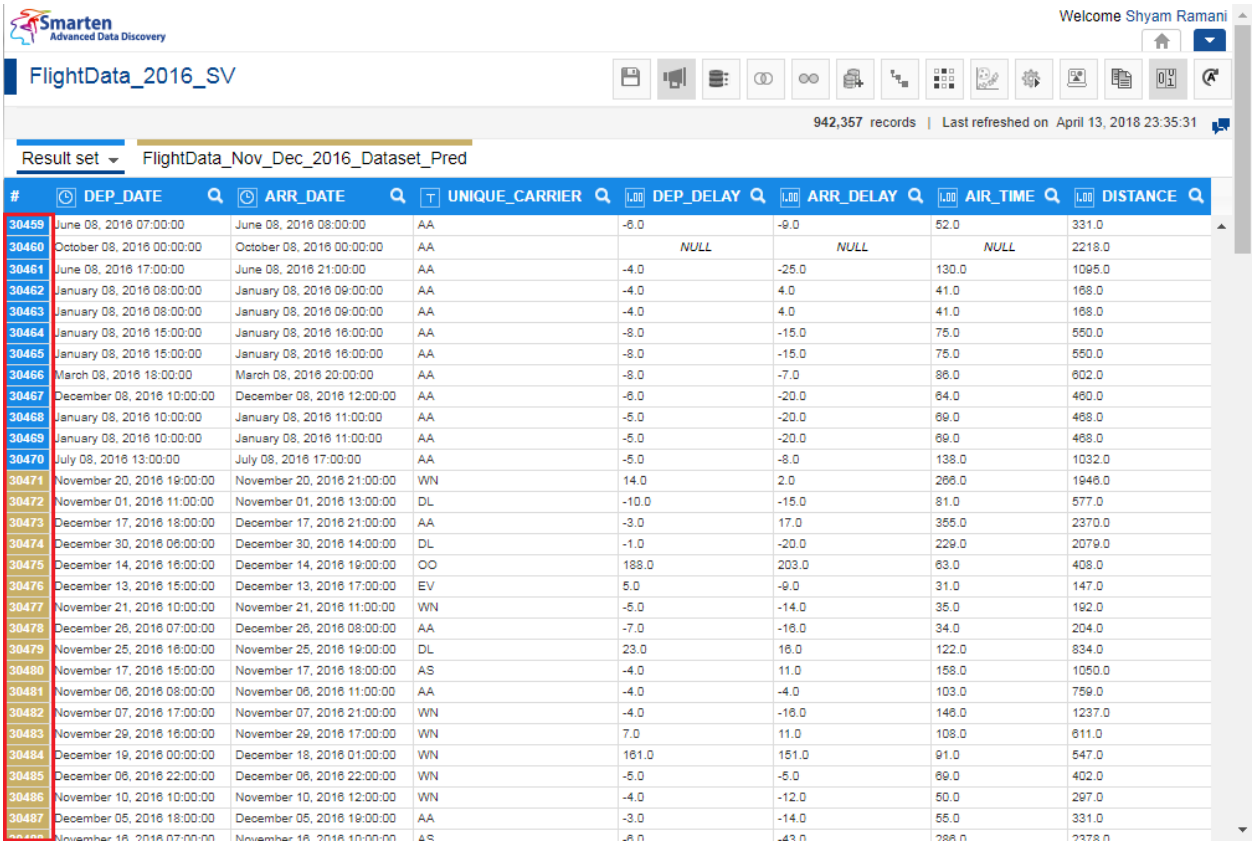
APPEND A DATASET—OPTION TO CREATE A COPY OF RESULTSET

If the Create a copy of resultset before applying the blend operation option is not selected, the system does not create a copy of the resultset and will append selected dataset data into the resultset.

8. Click **APPLY**.

The system updates the current dataset with the columns selected in the append criteria. The total number of records in the updated dataset is the combination of the rows available in the current dataset and the rows available in the added dataset. For example, if the Flight_Data_Jan dataset with 2,000 records is appended with the Flight_Data_Feb dataset, which has 2,500 records, the Flight_Data_Jan dataset after the append operation will have 4,500 records.

The system highlights the appended data with the color of the dataset it belongs to. In the example above, the system will highlight the 2,500 records in the resultset of Flight_Data_Jan with the color of the Flight_Data_Feb dataset.



#	DEP_DATE	ARR_DATE	UNIQUE_CARRIER	DEP_DELAY	ARR_DELAY	AIR_TIME	DISTANCE
30459	June 08, 2016 07:00:00	June 08, 2016 08:00:00	AA	-6.0	-9.0	52.0	331.0
30460	October 08, 2016 00:00:00	October 08, 2016 00:00:00	AA	NULL	NULL	NULL	2218.0
30461	June 08, 2016 17:00:00	June 08, 2016 21:00:00	AA	-4.0	-25.0	130.0	1095.0
30462	January 08, 2016 08:00:00	January 08, 2016 09:00:00	AA	-4.0	4.0	41.0	168.0
30463	January 08, 2016 08:00:00	January 08, 2016 09:00:00	AA	-4.0	4.0	41.0	168.0
30464	January 08, 2016 15:00:00	January 08, 2016 16:00:00	AA	-8.0	-15.0	75.0	550.0
30465	January 08, 2016 15:00:00	January 08, 2016 16:00:00	AA	-8.0	-15.0	75.0	550.0
30466	March 08, 2016 18:00:00	March 08, 2016 20:00:00	AA	-8.0	-7.0	86.0	602.0
30467	December 08, 2016 10:00:00	December 08, 2016 12:00:00	AA	-6.0	-20.0	64.0	460.0
30468	January 08, 2016 10:00:00	January 08, 2016 11:00:00	AA	-5.0	-20.0	69.0	468.0
30469	January 08, 2016 10:00:00	January 08, 2016 11:00:00	AA	-5.0	-20.0	69.0	468.0
30470	July 08, 2016 13:00:00	July 08, 2016 17:00:00	AA	-5.0	-8.0	138.0	1032.0
30471	November 20, 2016 19:00:00	November 20, 2016 21:00:00	WN	14.0	2.0	266.0	1946.0
30472	November 01, 2016 11:00:00	November 01, 2016 13:00:00	DL	-10.0	-15.0	81.0	577.0
30473	December 17, 2016 18:00:00	December 17, 2016 21:00:00	AA	-3.0	17.0	355.0	2370.0
30474	December 30, 2016 06:00:00	December 30, 2016 14:00:00	DL	-1.0	-20.0	229.0	2079.0
30475	December 14, 2016 16:00:00	December 14, 2016 19:00:00	OO	188.0	203.0	63.0	408.0
30476	December 13, 2016 15:00:00	December 13, 2016 17:00:00	EV	5.0	-9.0	31.0	147.0
30477	November 21, 2016 10:00:00	November 21, 2016 11:00:00	WN	-5.0	-14.0	35.0	192.0
30478	December 26, 2016 07:00:00	December 26, 2016 08:00:00	AA	-7.0	-16.0	34.0	204.0
30479	November 25, 2016 16:00:00	November 25, 2016 19:00:00	DL	23.0	16.0	122.0	834.0
30480	November 17, 2016 15:00:00	November 17, 2016 18:00:00	AS	-4.0	11.0	158.0	1050.0
30481	November 06, 2016 08:00:00	November 06, 2016 11:00:00	AA	-4.0	-4.0	103.0	759.0
30482	November 07, 2016 17:00:00	November 07, 2016 21:00:00	WN	-4.0	-16.0	146.0	1237.0
30483	November 29, 2016 16:00:00	November 29, 2016 17:00:00	WN	7.0	11.0	108.0	611.0
30484	December 19, 2016 00:00:00	December 18, 2016 01:00:00	WN	161.0	151.0	91.0	547.0
30485	December 06, 2016 22:00:00	December 06, 2016 22:00:00	WN	-5.0	-5.0	69.0	402.0
30486	November 10, 2016 10:00:00	November 10, 2016 12:00:00	WN	-4.0	-12.0	50.0	297.0
30487	December 05, 2016 18:00:00	December 05, 2016 19:00:00	AA	-3.0	-14.0	55.0	331.0

APPEND A DATASET—COLOR HIGHLIGHT INDICATING WHICH DATASET THE RECORDS BELONG TO

> Note:
> You cannot use the same dataset to append a dataset more than once.

7.2.31.2 Blending Data—Join Operation

Using this functionality, users can combine (JOIN) more than two datasets with the help of auto-suggestion to identify possible JOINs and their relative value and strength. Users can also create a copy of the resultset before applying the blend operation. Different types of Joins are available, such as Left, Inner, Right, and Outer.

Reference:	**Concept Manual > Blend Data > Join**

About this task

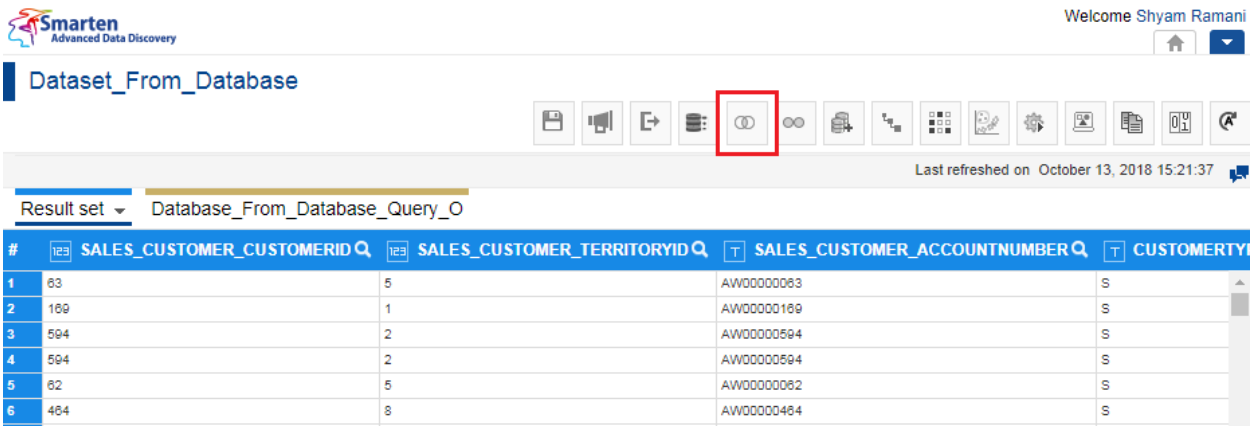
Use this task to join a dataset.

Procedure

1. Open the dataset you want to join with another dataset.

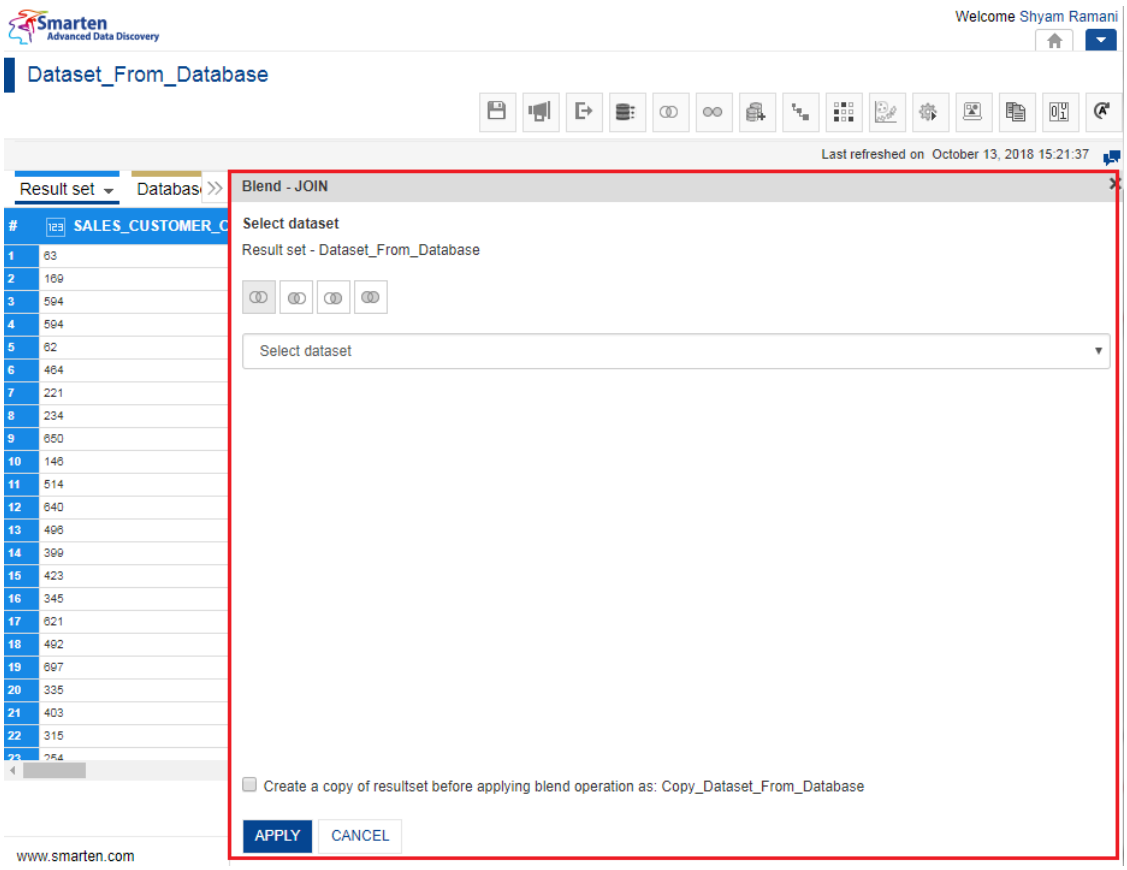
> Note:
> You must add datasets that you want to use before performing the blend operation.

2. Click the Blend - join icon on the toolbar.



BLEND JOIN—OPENING THE BLEND - JOIN DIALOG BOX

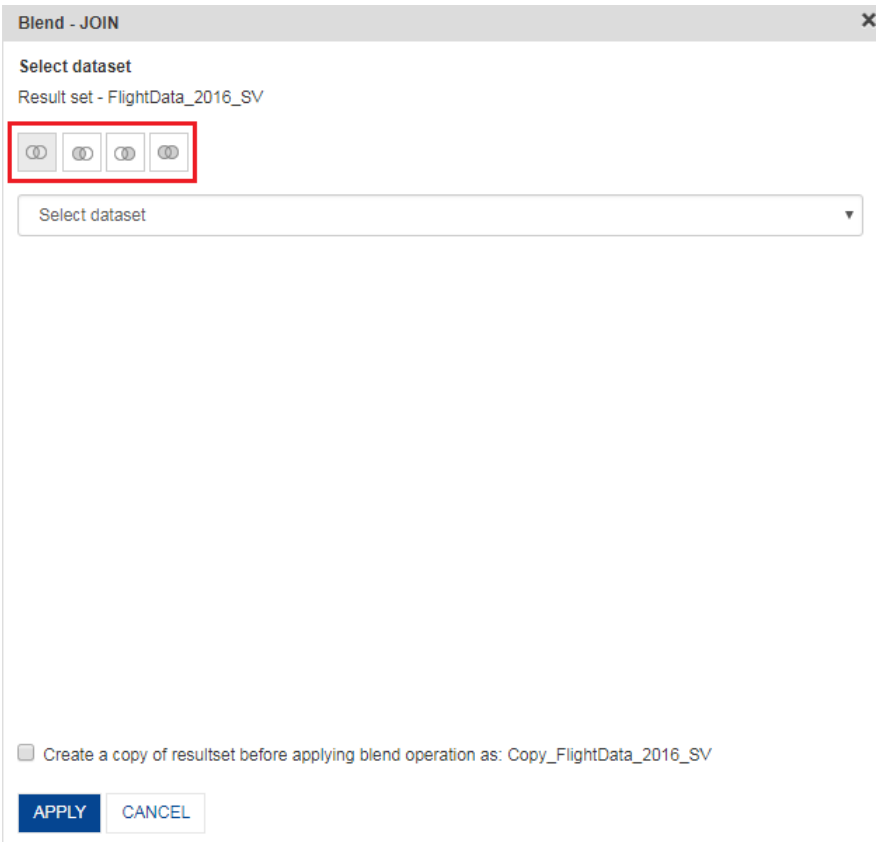
The system displays the **Blend - JOIN** dialog box.



BLEND JOIN—THE BLEND - JOIN DIALOG BOX

3. Select the type of join you want to apply. The following options are available:

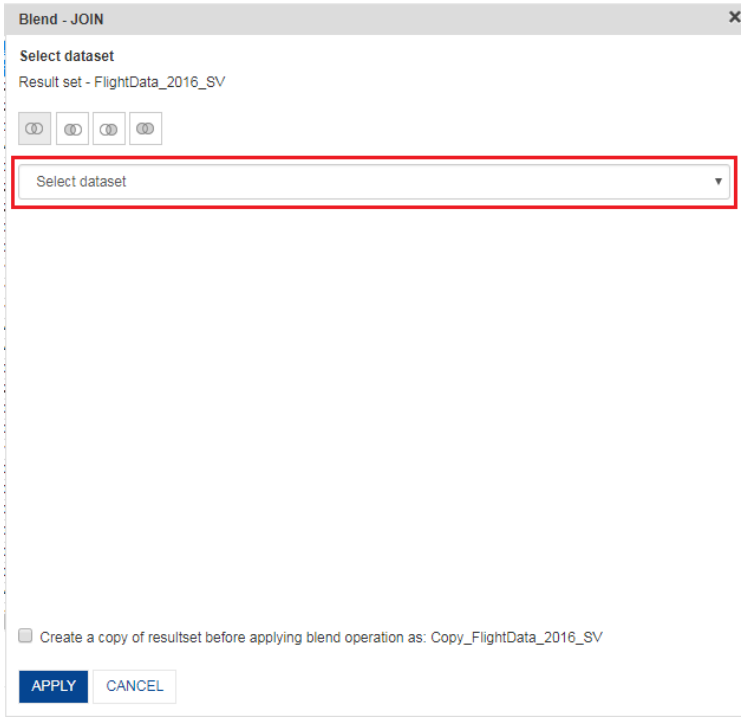
- **Equi**: Select this option if you want to apply equijoin between the two datasets.
- **Left**: Select this option if you want to apply left join between the two datasets.
- **Right**: Select this option if you want to apply right join between the two datasets.
- **Outer**: Select this option if you want to apply outer join between the two datasets.



JOIN A DATASET—SELECTING TYPE OF JOIN TO BE APPLIED

4. Select the dataset you want to join to the current dataset from the **Select dataset** list.

> Note:
> The datasets available in the list are the datasets you have added to the current dataset. If you have not added any datasets to the current dataset, no option will be available in the list.



JOIN A DATASET—SELECT A DATASET THAT WILL BE JOINED TO THE CURRENT DATASET

The system automatically displays a list of columns with matching criteria.

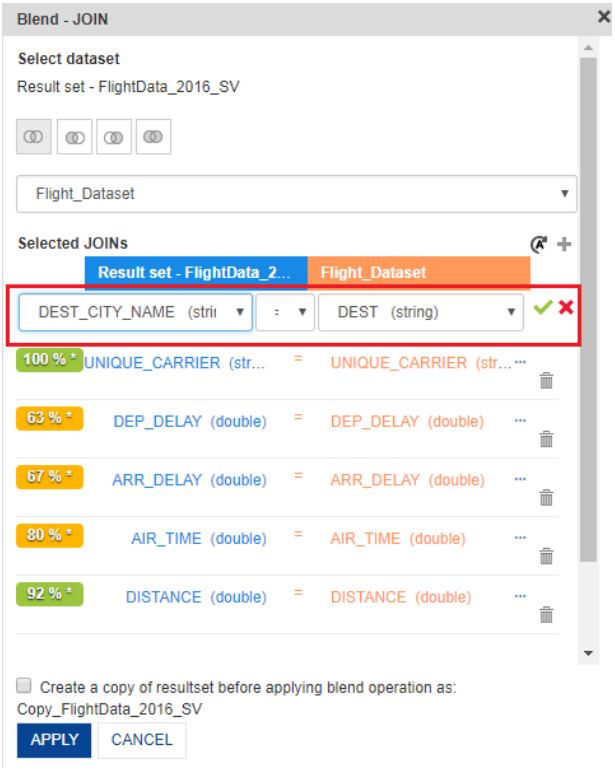
JOIN A DATASET—LIST OF COLUMNS WITH MATCHING VALUES

5. Apart from the automatically suggested column matches, the system allows you to add other matching columns.

   You can click the Add icon for manually adding another JOIN criteria.

   The following join conditions are available:

   - **=** : Select this option if you want to apply equal to operation for the join condition.
   - **!=** : Select this option if you want to apply not equal to operation for the join condition.
   - **<** : Select this option if you want to apply less than operation for the join condition.
   - **<=** : Select this option if you want to apply less than and equal to operation for the join condition.
   - **>** : Select this option if you want to apply greater than operation for the join condition.
   - **>=** : Select this option if you want to apply greater than and equal to operation for the join condition.



JOIN A DATASET—SELECT COLUMNS FOR MATCH CRITERIA

The system displays the JOIN strength as a percentage. The background of the percentage value is displayed in the following colors:

- **Green**: Indicates that more than 80% of records are identical between the two datasets.
- **Orange:** Indicates that more than 50% and less 80% of records are identical between the two datasets.
- **Red:** Indicates that less than 50% of records are identical between the two datasets.

JOIN A DATASET—MATCHING RECORDS AS A PERCENTAGE

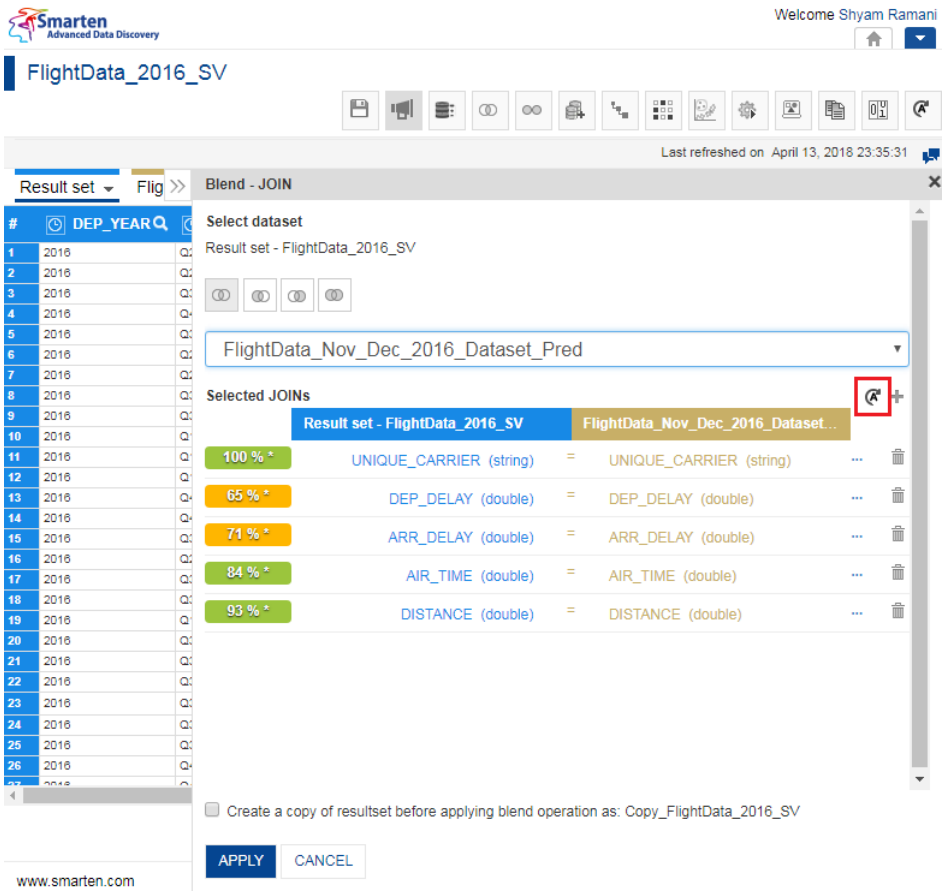
> Note:
> The "*" in the percentage value indicates that the value is calculated based on the sample data. Click the percentage value to calculate the actual percentage match for all the data.

6. You can click the Delete icon adjacent to a join condition to delete that condition.

JOIN A DATASET—DELETING A JOIN CONDITION

7. You can click the Auto icon to restore the auto-suggested matches the system has provided.

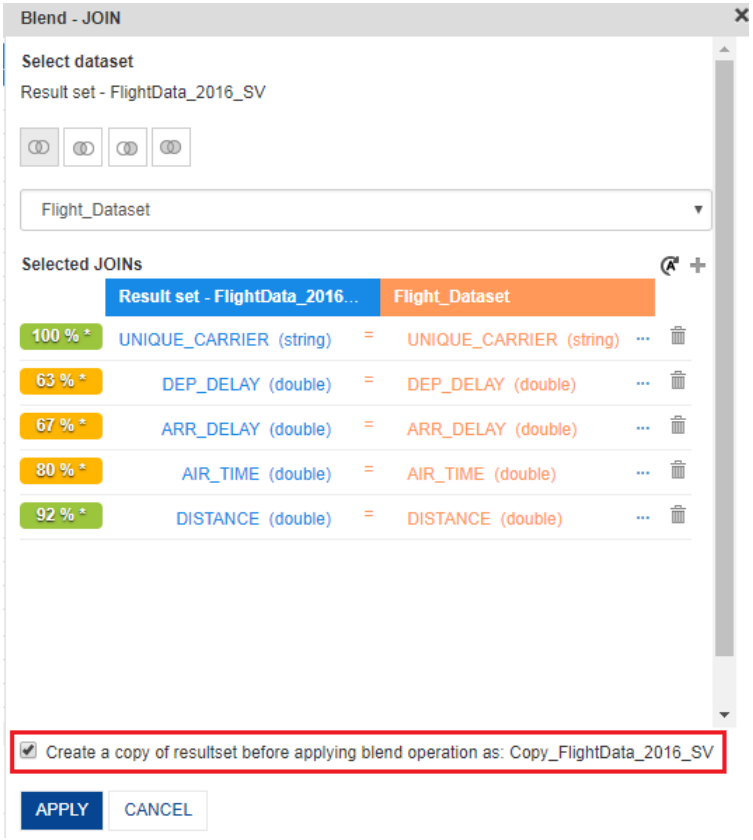
   The system restores the auto-suggested matches while retaining the new match criteria you added.



APPEND A DATASET—RESTORING AUTO-SUGGESTED MATCHES

8. Select the **Create a copy of the resultset before applying blend operation** as checkbox to create a copy of the resultset.

   The system creates a copy of the current dataset with the name displayed along the option. For example, in the image above, the system will create a copy of the current resultset with the name “Copy_Dataset_From_Database.”

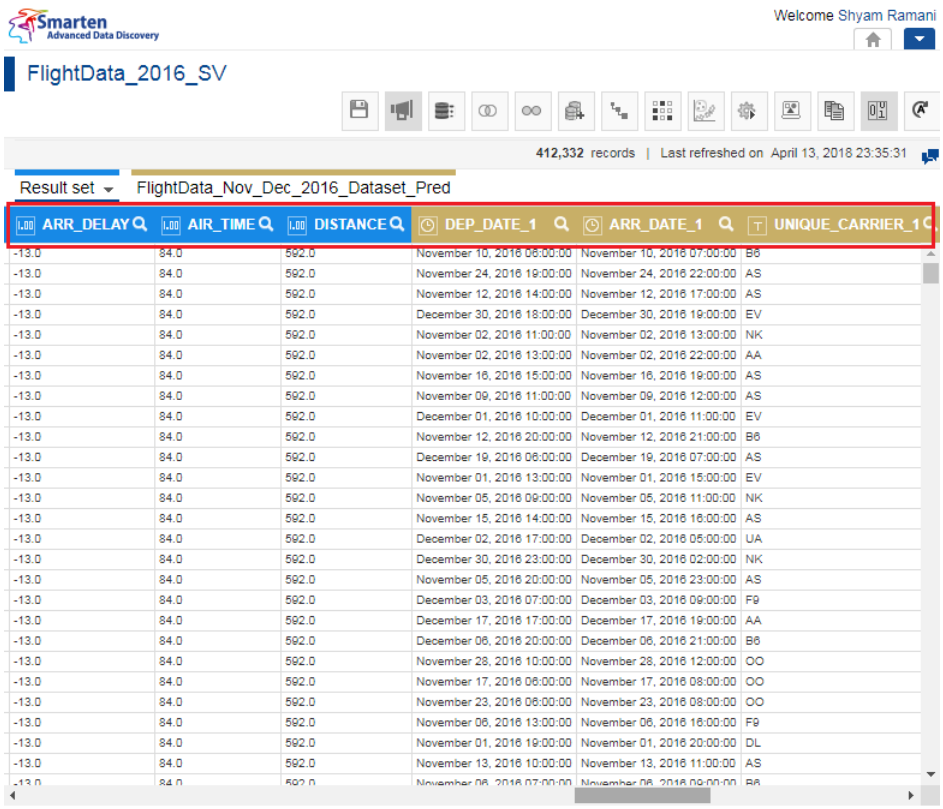


JOIN A DATASET—OPTION TO CREATE A COPY OF RESULTSET

If the **Create a copy of resultset before applying blend operation** option is not selected, the system does not create a copy of the resultset and will join the selected dataset data into the resultset.

9. Click **APPLY**.

The system joins the current dataset with the data of the selected dataset based on the join condition. The system highlights the data of the joined columns with the color of the dataset they belong to. For example, for the AIR_TIME column in the image above, the data from FlightData_2016 will be highlighted in blue, whereas the data from Flight_Dataset will be highlighted.



JOIN A DATASET—COLOR HIGHLIGHT INDICATING WHICH DATASET THE RECORD BELONGS TO

7.2.32 Aggregating Data of a Dataset

Aggregation allows you to aggregate and reduce the dataset by using various data operations on measure and date type columns.

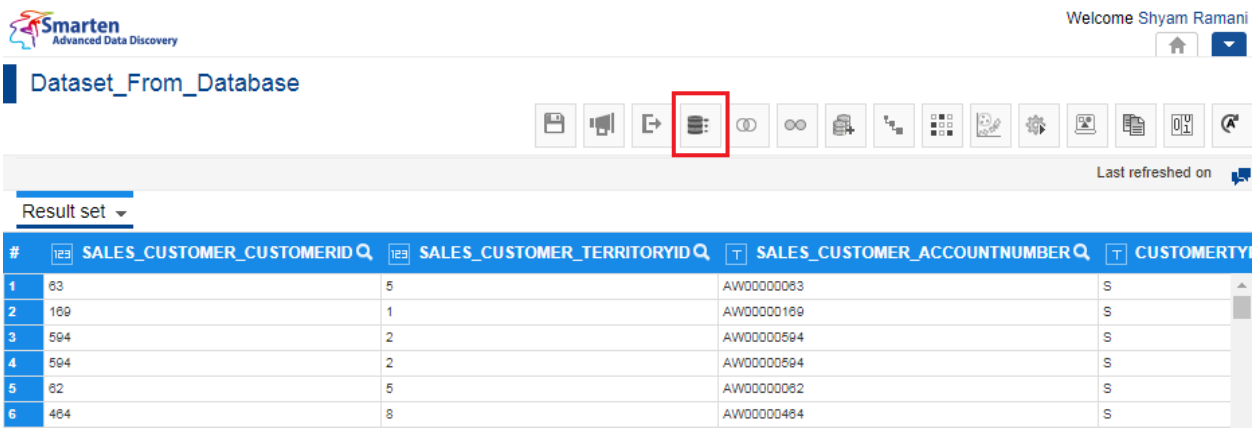
Reference:	**Concept Manual > Transform Data > Aggregate**

About this task

Use this task to aggregate and reduce data in a dataset.

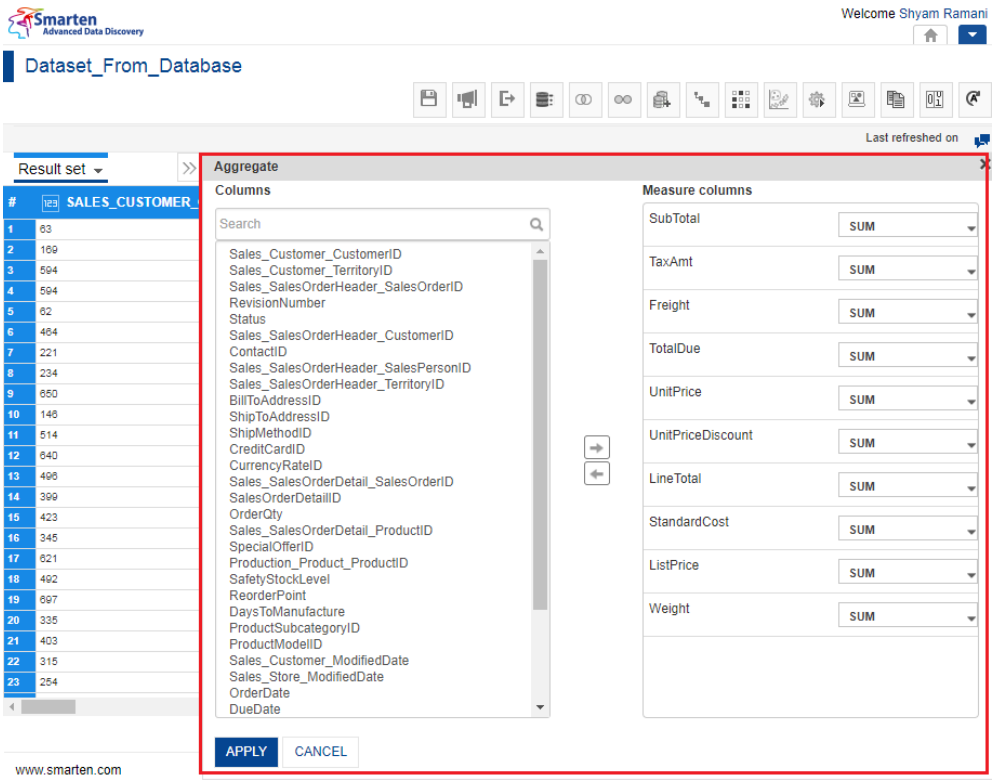
Procedure

1. Open the dataset for which you want to aggregate and reduce data.
2. Click the Aggregate icon on the toolbar.



AGGREGATE—OPENING THE AGGREGATE DIALOG BOX

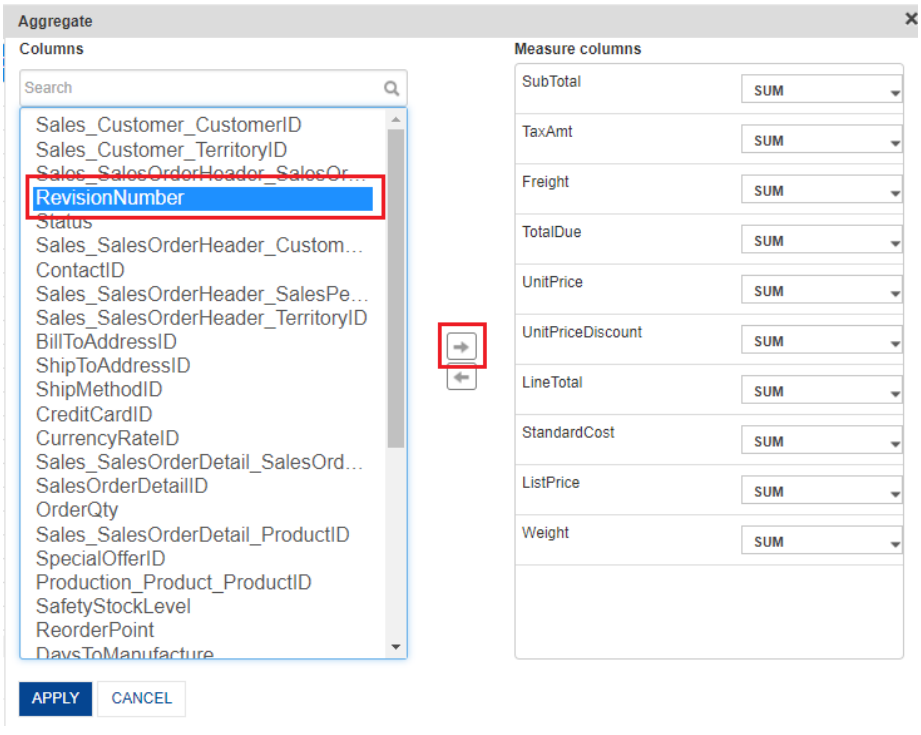
The system displays the **Aggregate** dialog box.



AGGREGATE—THE AGGREGATE DIALOG BOX

The system displays all the measure columns and date type columns for which you can aggregate the data.

3. You can select a column from the Column section and then click the right arrow to add that column to the Measure columns section.



AGGREGATE—ADDING COLUMNS TO THE MEASURE COLUMNS SECTION

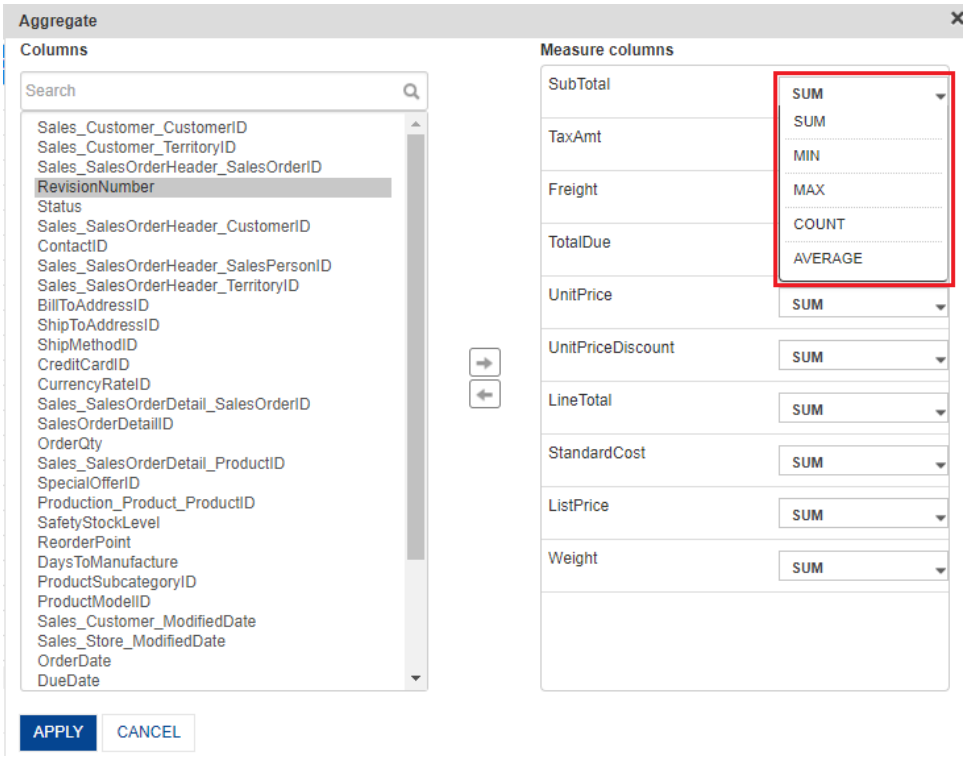
4. Select an operation to be performed for the columns available in the **Measure column** section.

Reference: **Concept Manual > Transform Data > Aggregate> Sum, Min, Max, Count, Average**

The following options are available:

- **SUM**: This function returns the sum of the selected Measure columns for the aggregated row.
- **MIN**: This function returns the minimum value of the selected Measure column for the aggregated row.
- **MAX**: This function returns the maximum value of the selected Measure column for the aggregated row.
- **COUNT**: This function returns the count of the selected Measure column for the aggregated row.
- **AVERAGE**: This function returns the average value of the selected Measure column for the aggregated row.

> Note:
> The options available for columns in the Measure column section depend on the data type of the column. For example, only MIN, MAX, and COUNT options are available for the columns with date data type.



AGGREGATE—OPERATIONS TO AGGREGATE DATA

5. Click **APPLY**.

The system aggregates the data based on the options you have selected. The resultant dataset contains reduced data based on the granularity of the data available in the dataset.

7.2.33 Exporting a Dataset Data

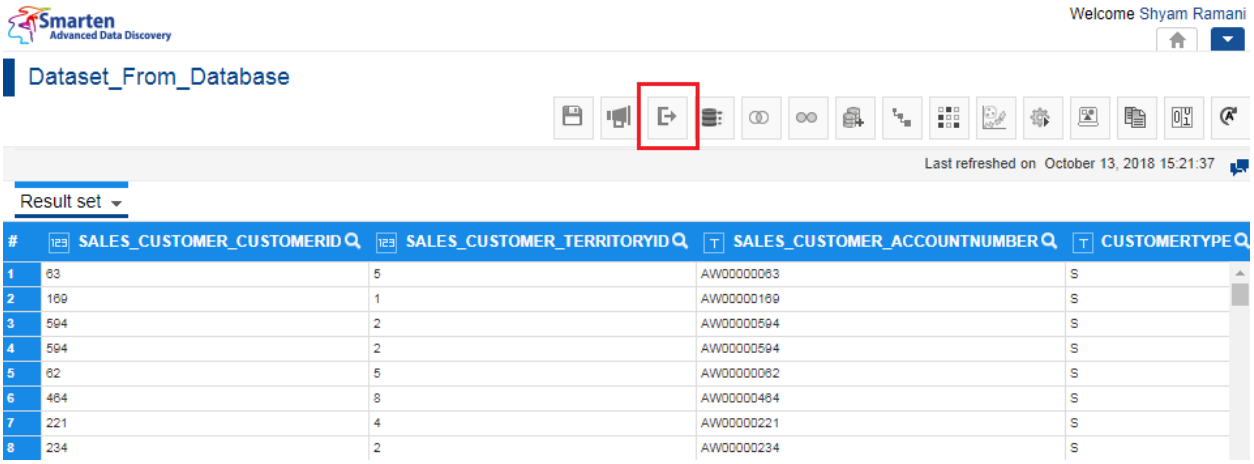
You can export data of a dataset in CSV format. The exported CSV file contains entire or limited data based on the option selected while exporting the data.

About this task

Use this task to export a dataset.

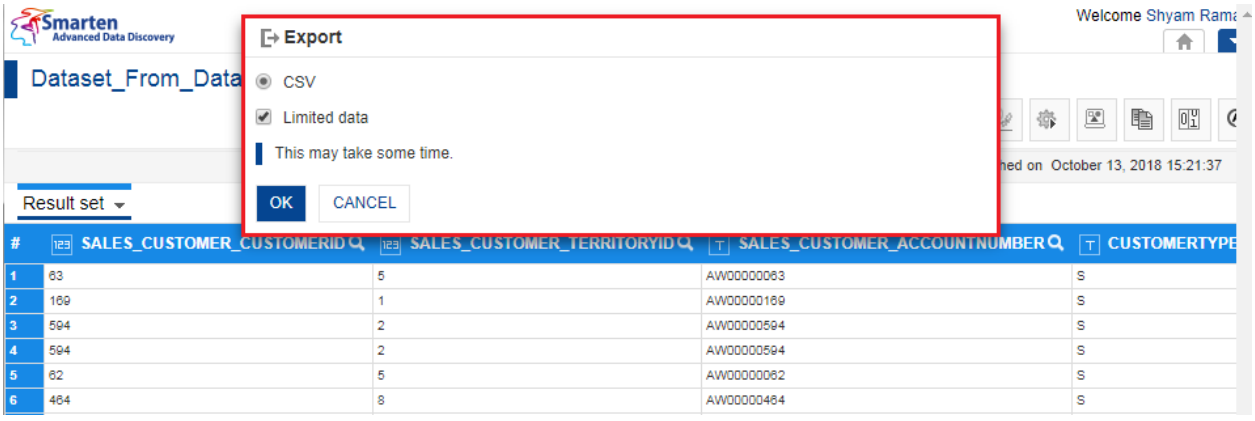
Procedure

1. Open the dataset you want to export.
2. Click the Export icon on the toolbar.



EXPORT A DATASET—OPENING THE EXPORT DIALOG BOX

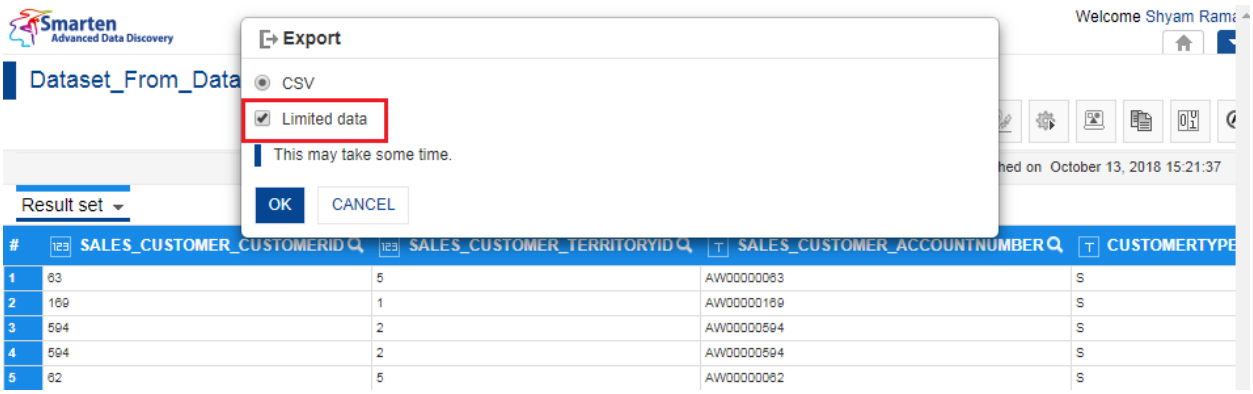
The system displays the **Export** dialog box.



EXPORT A DATASET—THE EXPORT DIALOG BOX

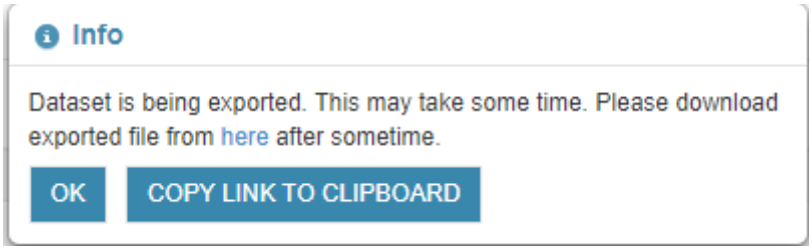
3. You can select the **Limited data** option to export limited data, and then click **OK**.

   If you have not selected this option, the system exports all the data available in the dataset.



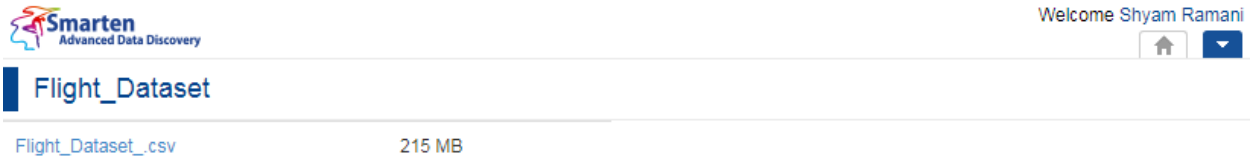
EXPORT A DATASET—OPTION TO EXPORT LIMITED DATA FROM THE DATASET

> Note:
> If the data to be exported is huge, the system displays a message that the data is being exported, and you can download it after it is exported using the link provided.



EXPORT A DATASET—MESSAGE PROVIDING A LINK TO DOWNLOAD THE EXPORTED FILE

4. You click **COPY LINK TO CLIPBOARD** to copy the link to download the exported file. Click **OK** on the **Info** dialog box.



EXPORT A DATASET—DOWNLOADED CSV FILE

7.2.34 Publishing a Dataset

Publishing a dataset makes a dataset available for collaboration and reuse by other Smarten users, who have access to a full suite of Smart Data Visualization, Plug n' Play Predictive Analysis, Dashboards, and Reports to analyze, present, and share results from the published datasets. By default, a dataset is unpublished, and only the user who created the dataset can access that dataset. Once a dataset is published, other users who have been given access rights by the creator of the dataset can access the dataset as per the permissions given to them.

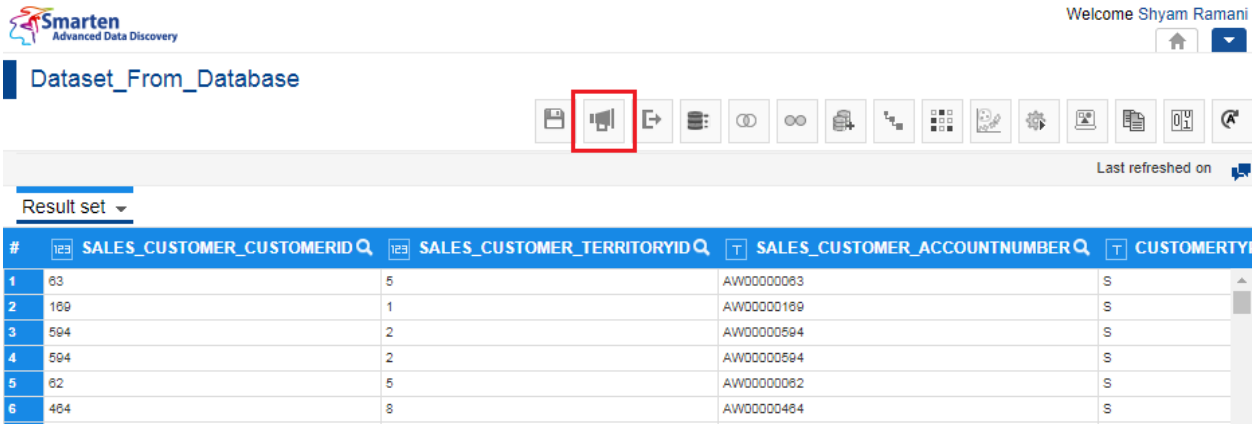
Reference: **Concept Manual > Publish and Refresh Data > Publish Dataset**

About this task

Use this task to publish a dataset.

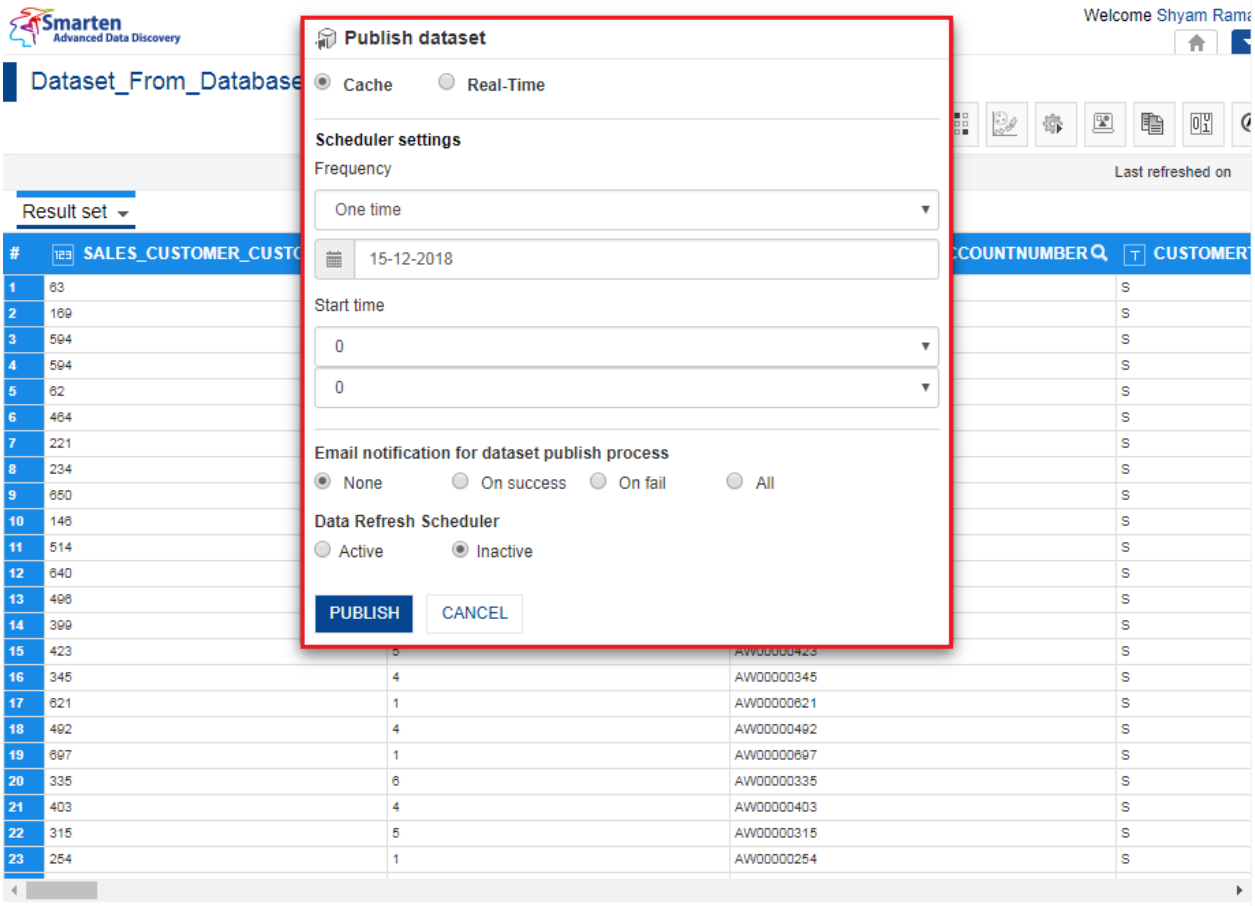
Procedure

1. Open the dataset you want to publish.
2. Click the Publish icon on the toolbar.



PUBLISH A DATASET—OPENING THE PUBLISH DATASET DIALOG BOX

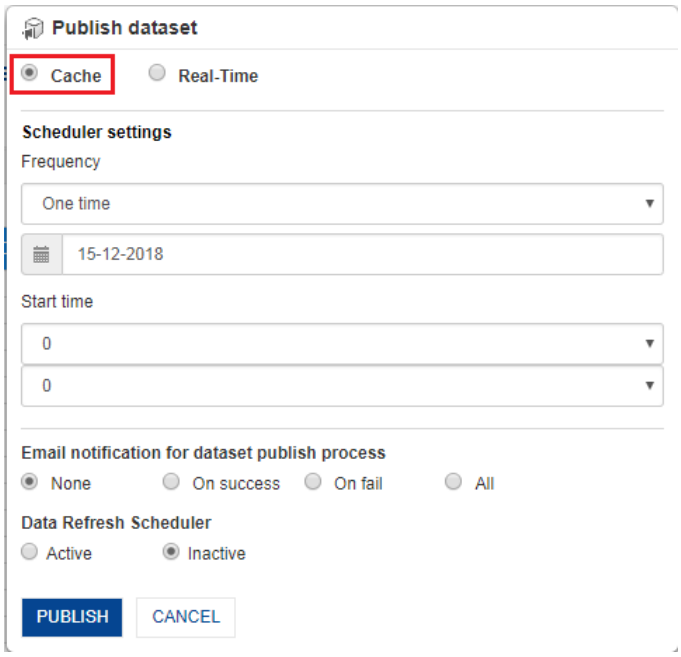
The system displays the **Publish dataset** dialog box.



PUBLISH A DATASET—THE PUBLISH DATASET DIALOG BOX

3. Select the **Cache** option to publish the dataset as a cached dataset.

   Datasets published as cache datasets store data in a columnar data structure and need to be updated periodically from the data sources with the help of a scheduler.



PUBLISH A DATASET—OPTION TO PUBLISH A DATASET AS A CACHE DATASET

   a) Select an option from the Frequency list to specify how often you want to refresh the data of the dataset.

   > Note:
   > Based on the option you select from the Frequency list, the system displays relevant options.

   b) Select an option to specify events for which an email notification must be sent.

   The following options are available:

   - **None**: No email notification should be sent.
   - **On success**: Email notification to be sent only if the dataset is published successfully.
   - **On fail**: Email notification to be sent only if the dataset is not published successfully.
   - **All**: Email notification should be sent for all events.

   c) Select an option to specify whether the data refresh scheduler should be active or inactive.

4. Select the **Real-time** option to publish the dataset as a real-time dataset.
   Datasets published as real-time datasets do not store data in a columnar data structure. They extract the latest data from Data Sources as and when required.

5. Click **Publish**.

7.2.35 Saving a Dataset

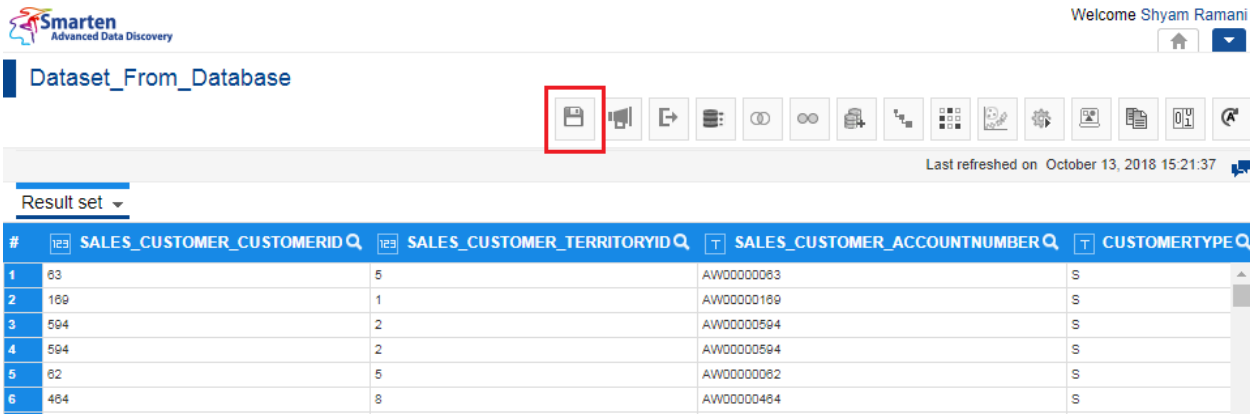
You must save a dataset every time you perform an operation on a dataset. If you do not save a dataset, all the operations you have applied on the dataset are lost.

About this task

Use this task to save a dataset.

Procedure

1. Open the dataset that you want to save.
2. Click the Save icon on the toolbar after you have applied all the operations.



SAVING A DATASET—CLICKING THE SAVE ICON

The system saves the dataset along with all the operations applied.

7.2.36 Dataset Rebuild methods

Rebuild dataset enables you to update dataset with the latest data from the data source. You can update the dataset with entire data - **From scratch** or append only the new data available in the data source - **Incremental**.

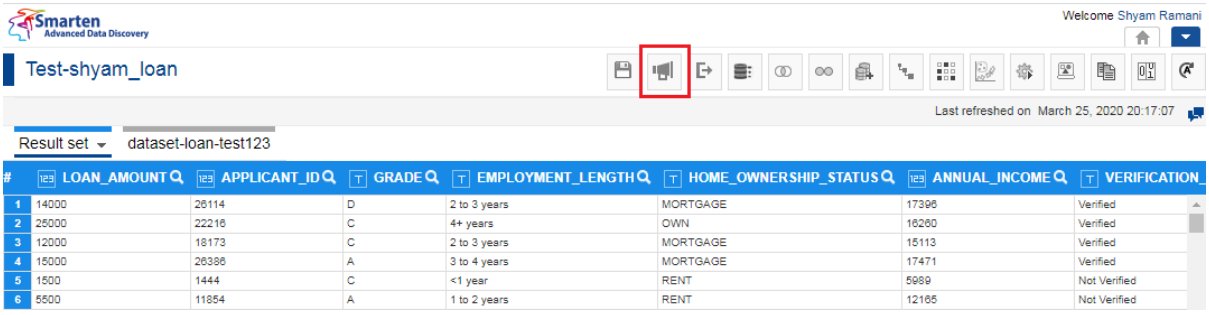
Reference: **Concept Manual > Publish and Refresh Data > Refresh Dataset**

About this task

Use this task to update a dataset.

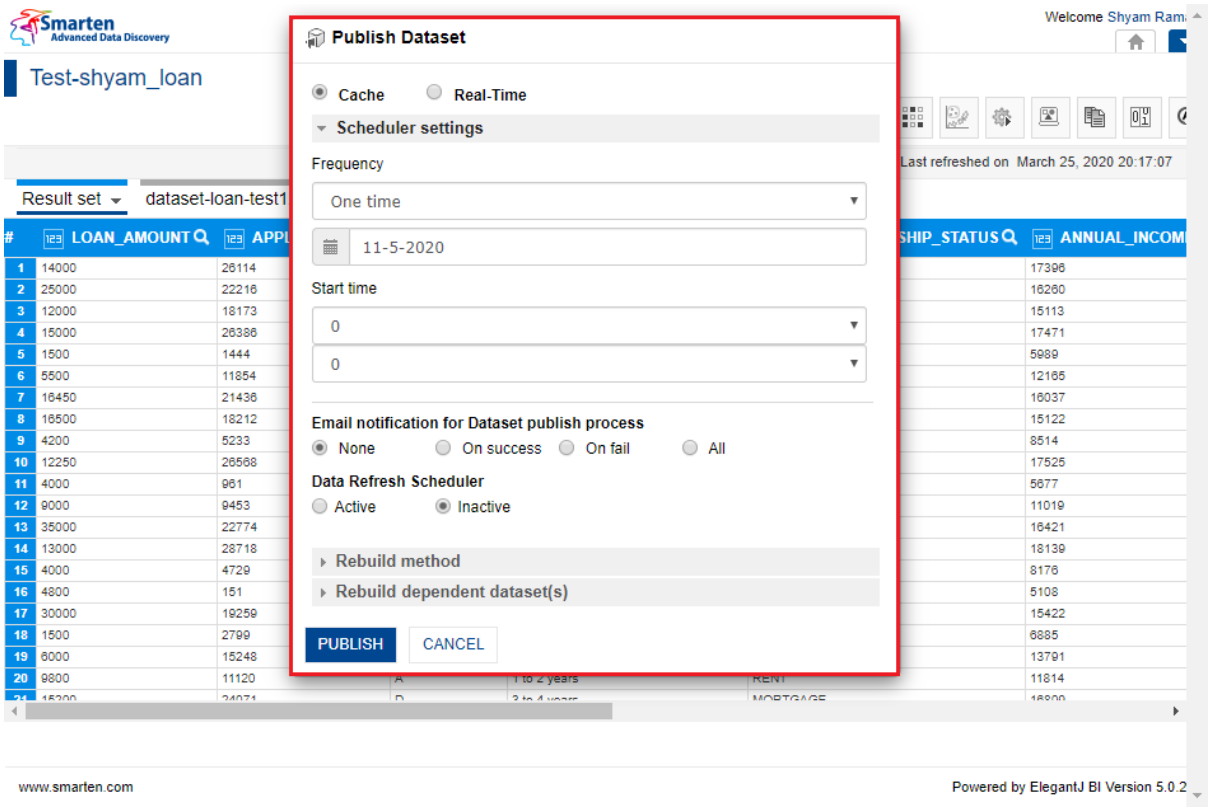
Procedure

1. Open the dataset you want to update.
2. Click the Publish icon on the toolbar.



UPDATE DATASET—OPENING THE PUBLISH DATASET DIALOG BOX

The system displays the **Publish dataset** dialog box.



UPDATE DATASET—THE PUBLISH DATASET DIALOG BOX

7.2.36.1 From scratch

You can update dataset from scratch. In this method, the system retrieves the latest data from the datasource and overwrites the data in the dataset.

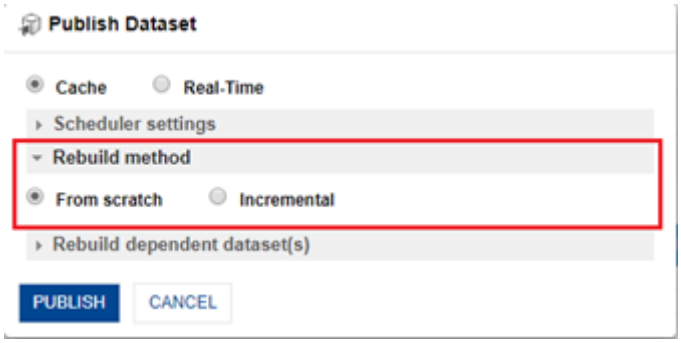
Reference: **Concept Manual > Refresh Data > Refresh methods > From scratch**

About this task

Use this task to update a dataset from scratch.

Procedure

1. Click **Rebuild method** to specify the option to specify how you want to update the dataset.



REBUILD METHOD – FROM SCRATCH

2. Select the **From scratch** option.
   In this option, the system overwrites the data in the dataset with the new data available in the datasource.
3. Click **PUBLISH**.
   The system publishes dataset with new data from the data source.

7.2.36.2 Incremental

You can update dataset with incremental option. In incremental option, system retrieves data from data source and appends only new data into the dataset. Smarten supports two options for incremental update, one is the **append all rows retrieved from data source** and another is the **append new rows identified based on unique ID column**.

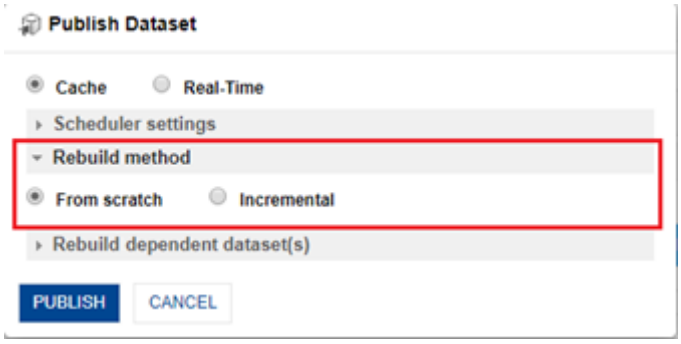
Reference: **Concept Manual > Refresh Data > Refresh methods > Incremental**

About this task

Use this task to update a dataset incrementally.

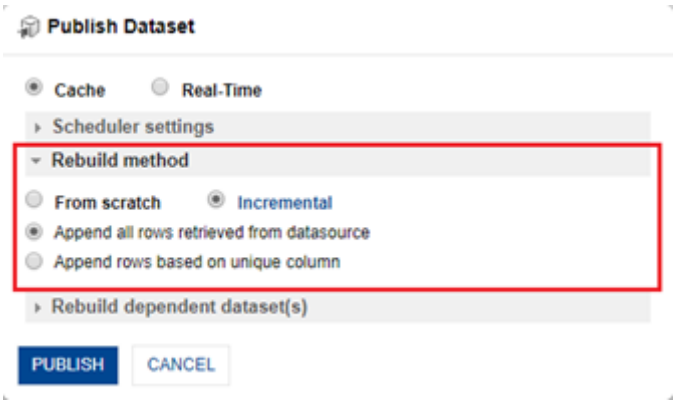
Procedure

1. Click **Rebuild method** to specify the option to specify how you want to update the dataset.



REBUILD METHOD - INCREMENTAL

2. Select Incremental option to append the new data available in the datasource to the dataset every time the dataset is updated.

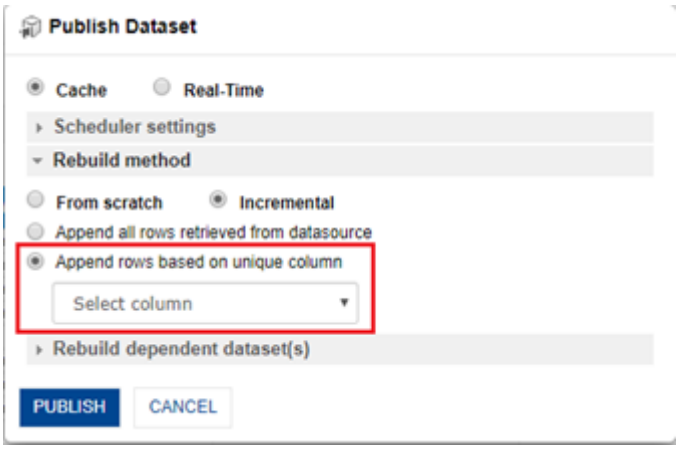


UPDATEDATASET – INCREMENTAL REBUILD OPTIONS

3. You can select the **Append all rows retrieved from datasource** option if you want to append the all records from the data source to the dataset.
4. You can select the **Append rows based on unique column** option if you want to append the new data based on the selected unique column. Select the column based on which you want to perform incremental update of the data in the dataset.

   For example, if you have selected the 'ID' column as a unique column from a dataset and the highest value in that column is '250' in the dataset. When you update the dataset, the system retrieves only those records that have value greater than '250' in the 'ID' column and appends that data to the dataset.

> Note:
> You can only select numeric unique ID column or Date data type column as unique column.



UPDATE DATASET- INCREMENTAL UPDATE BASED ON A COLUMN

5. Click **PUBLISH**.
   The system publishes the dataset with Incremental rebuild method.

7.2.37 Update Dependent Datasets

You must save a dataset every time you perform an operation on a dataset. If you do not save a You can update a dependent datasets while updating the current dataset. A dependent dataset is a dataset that are used in blend operations -JOIN, APPEND or a parent dataset. This feature provides facility to rebuild dependant datasets first and then rebuild current dataset. So, current dataset contains latest data from dependant datasets as well. It helps avoid overlapping of different rebuild tasks which generally causes reflection of older data.

For example, we have Orders dataset. For customer details, it is joined with Customers dataset. Now if we setup different schedulers to rebuild Customers and Orders dataset and it is configured in a way that Customers dataset gets rebuilt first, and then Orders dataset gets rebuilt, then there are chances that, scenario, Customers dataset rebuild is in progress and Orders rebuild process gets started. In such a scenario, Orders dataset will not have latest Customers data.

In such cases, if we use Rebuild dependant dataset feature, while rebuilding Orders dataset, system will first rebuild Customers dataset first, once Customers rebuild data is complete, then it will rebuild Orders dataset, and Orders dataset will always have latest Customers dataset reference records and values.

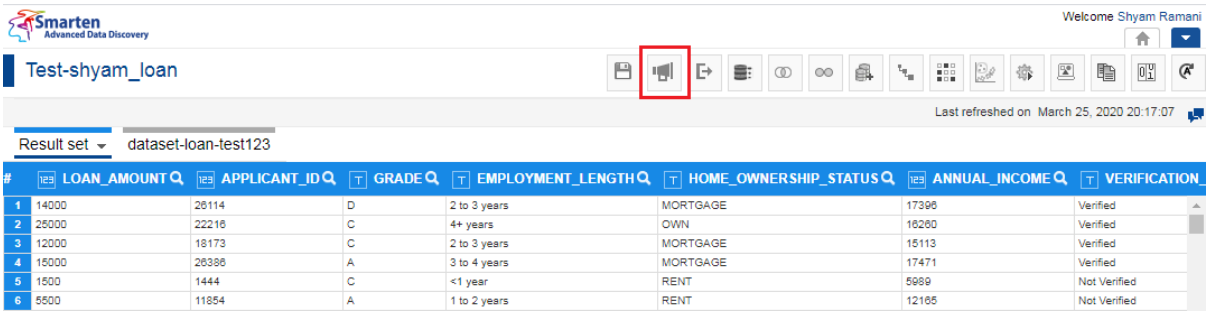
> Note
> Dependant datasets will only be rebuilt when current dataset is rebuilt through scheduler using option explained above. Refresh from source option will override dependent dataset processing configuration, and only rebuild current dataset.

About this task

Use this task to updatea dependent dataset.

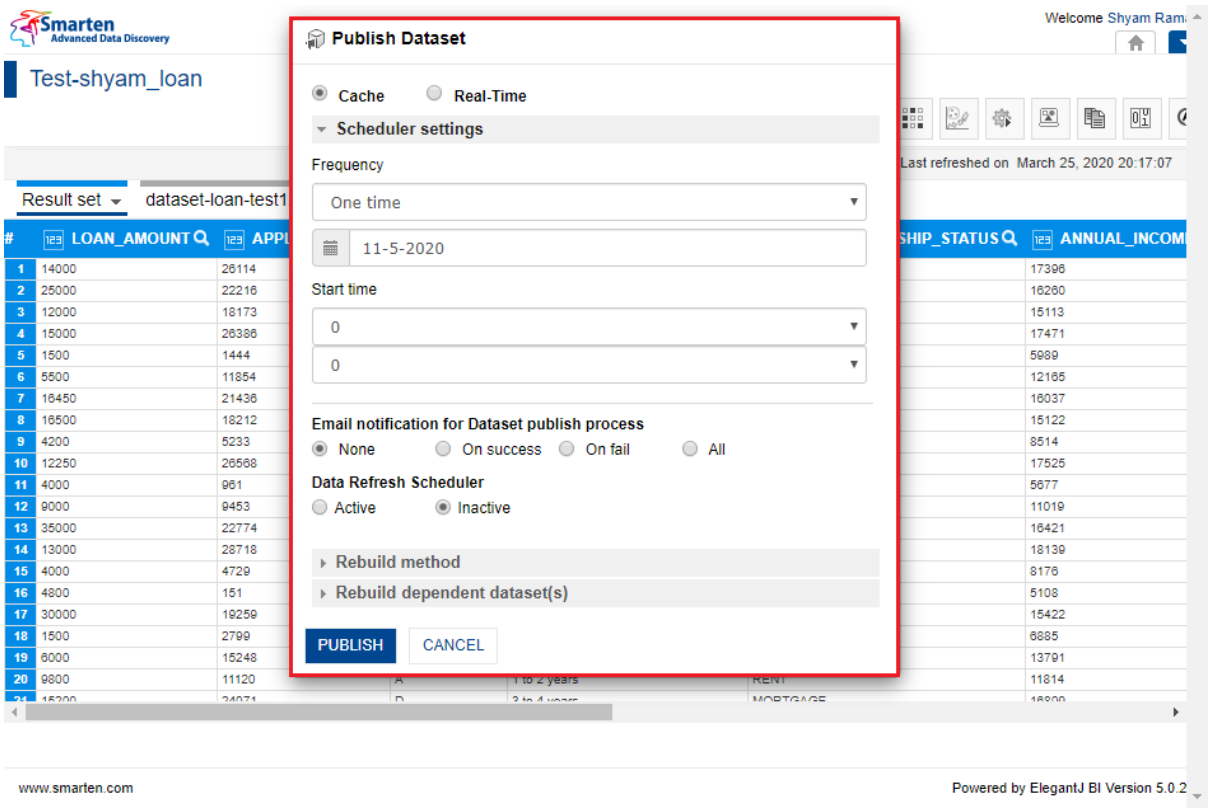
Procedure

1. Open the dataset you want to update.
2. Click the Publish icon on the toolbar.



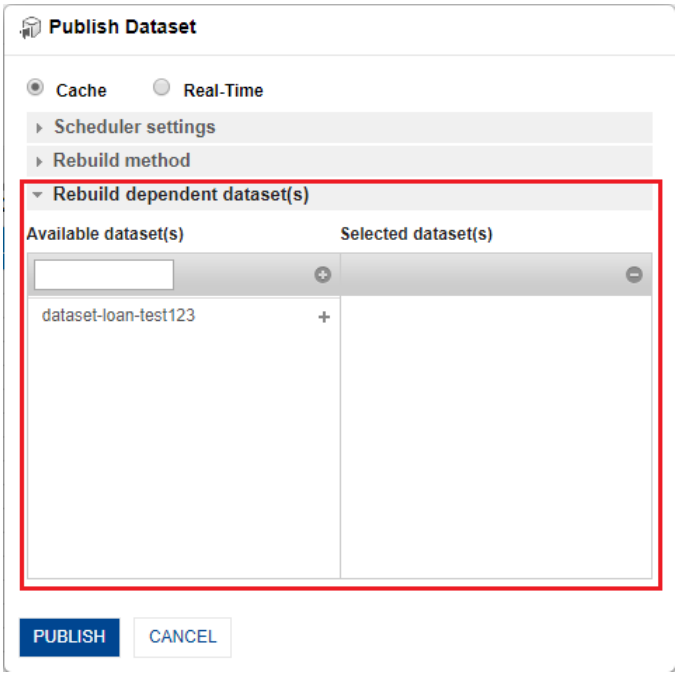
UPDATE DATASET—OPENING THE PUBLISH DATASET DIALOG BOX

The system displays the **Publish Dataset** dialog box.



UPDATEDATASET—THE PUBLISH DATASET DIALOG BOX

3. Click the **Rebuild dependent dataset(s)** option to specify the dependent dataset that you want to build when the current dataset is rebuilt.
4. Click the plus sign adjacent to the dataset that you want to update with the current dataset.



UPDATEDATASET – SELECT DEPENDENT DATASET

5. Click **Publish**.
   The system publishes the dataset with rebuild dependant datasets option.

8 Product and Support Information

Find more information about Smarten and its features at www.smarten.com
Support: support@smarten.com
Sales: sales@smarten.com
Feedback & Suggestions: support@smarten.com
Support & Knowledgebase Portal: support.smarten.com