Smarten Working with Clickless Analytics

Document Information	
Document ID	Smarten-Working-with-Clickless-Analytics
Document Version	1.0
Product Version	5.2
Date	14-June-2021
Recipient	NA
Author	EMTPL

Statement of Confidentiality, Disclaimer and Copyright

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.

Smarten

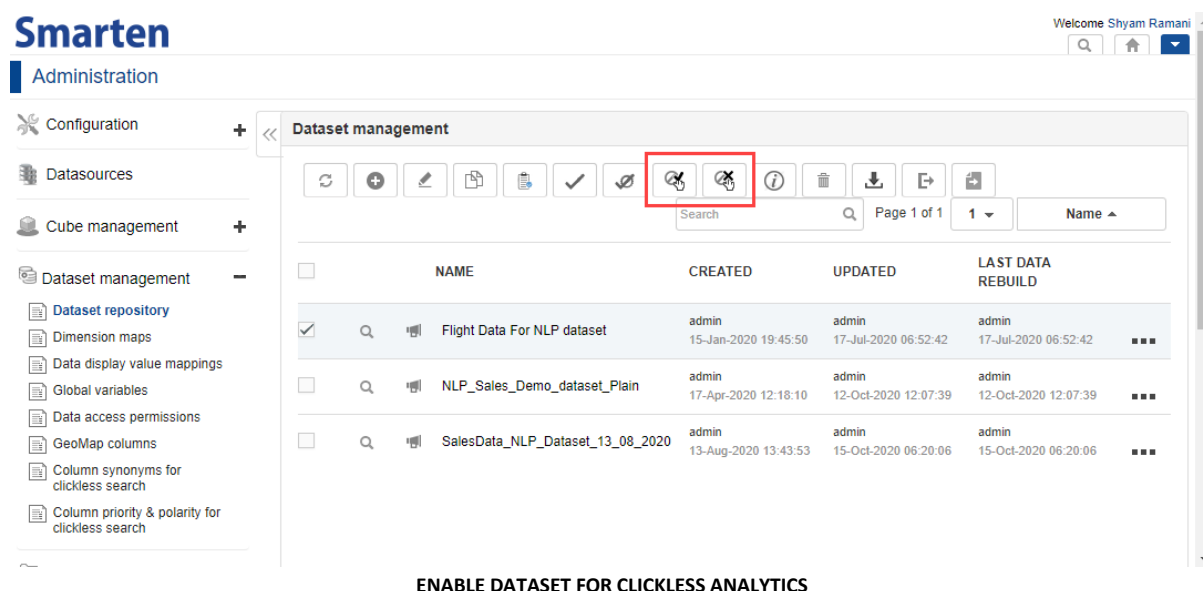## Contents

1   Introducing Clickless Analytics

Clickless Analytics allows business users to ask questions in basic English and get answers. Clickless Analytics provides a seamless augmented analytics experience, using machine learning and natural language processing (NLP). Users can simply enter a search query in basic English, such as, "What was the sales of Bakery in 2019," and the system translates the query and returns the results in an appropriate form, such as visualization, tables, numbers, or descriptions. It reduces skill-set requirements or training required to query data and helps organizations democratize the use of analytics.

2   Clickless Analytics Process Overview

The process of Clickless Analytics starts with identifying the dataset that you want to query and get answers. You can work with default settings and auto recommendations generated for a dataset, but to improve the quality of the search and results, you need to configure synonyms, column priority, default aggregation, measure polarity, phonetics, column types, and abbreviations. You might also need to update the predefined system definitions that are available in the various files.

2.1   Enable Dataset for Clickless Analytics

Clickless Analytics can be enabled for the published dataset within Smarten. By default, Clickless Analytics is disabled, and you need to enable it for the required datasets that will be used for the Clickless Analytics Search query. An Administrator can enable or disable Clickless Analytics for a dataset.

ENABLE DATASET FOR CLICKLESS ANALYTICS

2.2   Define Dataset specific NLP Definitions

Once you have enabled a dataset for Clickless Analytics, you need to specify certain definitions for the dataset. These definitions help improve the quality of the search and semantics.

2.2.1   Synonyms

As for synonyms, as the name suggests, you need to provide alternate terms for a column of the dataset. These are the possible terms that a business user can use in the query to search for the data

from the datasets. For example, if there is an "Employee Name" column available in a dataset, the synonyms that can be added are, e.g., sales representative, salesperson, salesman, and sales rep.

You can provide synonyms manually or use the **Auto Synonyms** option, which uses a predefined library of synonyms for various values in dataset columns.

You have the option to use manually defined synonyms, auto synonyms, or both.

2.2.2 Column Priority

When the system processes the query and finds multiple columns for a specific token, it is necessary for the system to identify one appropriate column from them. Setting priority for columns helps the system select appropriate columns from the identified columns and provide accurate results. For example, there are two columns in the dataset "SalesAmount" and "SalesQty," and the user has entered a query, such as "Sales of 2020." For "Sales" token, the system identified two columns: SalesAmount and SalesQty. If the column priority of these columns is set as 1 for Salesamount and 2 for Salesqty respectively, the system will select the Salesamount column for "Sales" token and show Salesamount in the result.

If you have not set the priority for the columns in a dataset, the system automatically selects the column with higher density (higher number of unique values).

2.2.3 Default Aggregation on Columns

All measure columns are aggregated with the sum operation by default. However, aggregating some measure columns with the sum operation may not be meaningful. For such measure columns, you can specify the default aggregation so that when that measure column is used in the query result the system uses the aggregation you have specified.

For example, sum data operation is suitable for the "Salesamount" measure. For the "Unitprice" column, sum data operation is not meaningful. So, you can specify "average" data operation for this measure as default aggregation. So, the system will apply average data operation when it is used in the query result.

2.2.4 Measure Polarity

Polarity is used to determine if a high value or a low value is better for a column. For example, Profit measure has high polarity, as higher value is better. Discount measure has low polarity, as lower value is better. You can specify polarity of measure based on its nature. The system uses measure polarity when a user queries ranking—top- or bottom-related queries on a measure.

2.2.5 Phonetics Search

Clickless Analytics allows you to enable a phonetic search for a column. When phonetics search is enabled, the system uses machine learning to search for terms that sound like the one that is used in the query. For example, phonix, phoenix, sells, and sales.

Phonetics search enables you to provide accurate results in the scenarios where a business user might have used the wrong term or spelling or dialect (that sounds similar) in the query. For example, if a business user wants to know the employee with the highest sales in Phoenix, AZ, and the query that the user used is "who has the highest sales in phonix," the user has misspelled Phoenix as phonixin in the query. However, if phonetics search is enabled for the relevant column,

the system will show correct results, i.e., information about the employee with the highest sales in Phoenix, AZ.

2.2.6   Identify Column Types

Clickless Analytics uses machine learning to identify column types, such as person name, geo column, or time dimension, based on data in the column and column type marking on the dataset. Identified columns are used to resolve queries with WH questions.

For example, for the question, "Who sold most in Arizona," the system identifies the "SalesPerson" column as the person name and provides results based on this column.

For the question, "Where did we sell the most bakery items in 2020," the system will identify the "State" geo column and show results accordingly.

2.2.7   Abbreviations

You need to specify the abbreviations used in the data or the ones that a business user can use in the query. For example, if you have used AZ for Arizona in the data, then you need to specify the abbreviation in the system to get appropriate results. The system uses the prebuild library that contains the standard abbreviations that can be used in the query or in the data. You can update the files in the library to include the abbreviations as per your requirements.

If you have used the query "Which is the highest selling product in AZ," the system retrieves information about the highest-selling product in Arizona. In the query, we have used "AZ," which is the abbreviation of the state of Arizona. However, in the data, we have used the name of the state, i.e., Arizona, and not the abbreviations. The system was able to retrieve correct data because of the abbreviations we have specified.

2.3   Query using Clickless Analytics

Clickless Analytics enables you to easily query for the required information from a dataset. It provides a similar experience, such as Google search bar, to query the data. All you need to do is type the query for the data you want to search in plain English, and the Clickless Analytics will return the relevant result.

For example, if you want to know which employee sold the maximum number of bakery items last month, you need to select the dataset that you want to use and type the "Who sold the highest number of bakery items last month" query or any query with similar semantics in the Clickless Analytics query bar. The system processes the query you have provided and retrieves the relevant result from the selected dataset.

The Clickless Analytics query bar is shown in the image below. You can select the dataset for which you want to query the data. The dataset selector displays datasets for which Clickless Analytics is enabled by the Administrator.

CLICKLESS ANALYTICS: CLICKLESS QUERY BAR
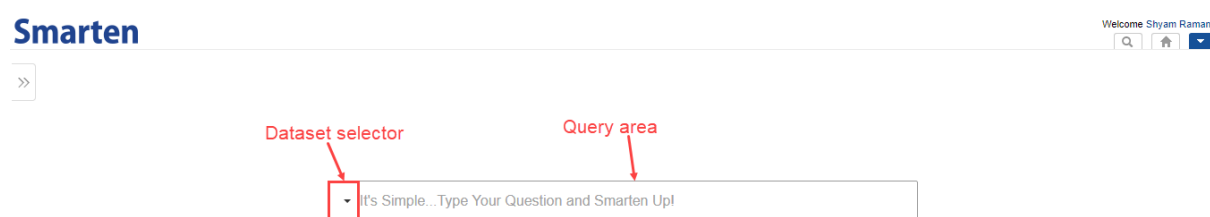
After selecting the dataset, you can type the query you want to use. When you start typing a query, the system suggests relevant tokens and queries. This will help the user to build the query.
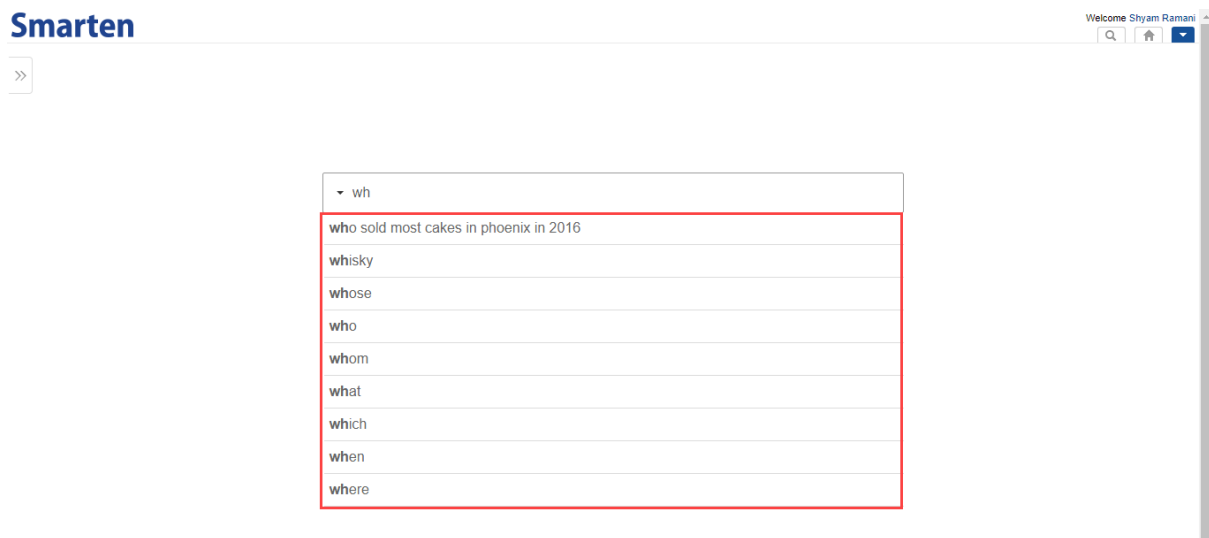
CLICKLESS ANALYTICS: AUTO-SUGGESTION & TYPE-AHEAD FOR THE QUERY
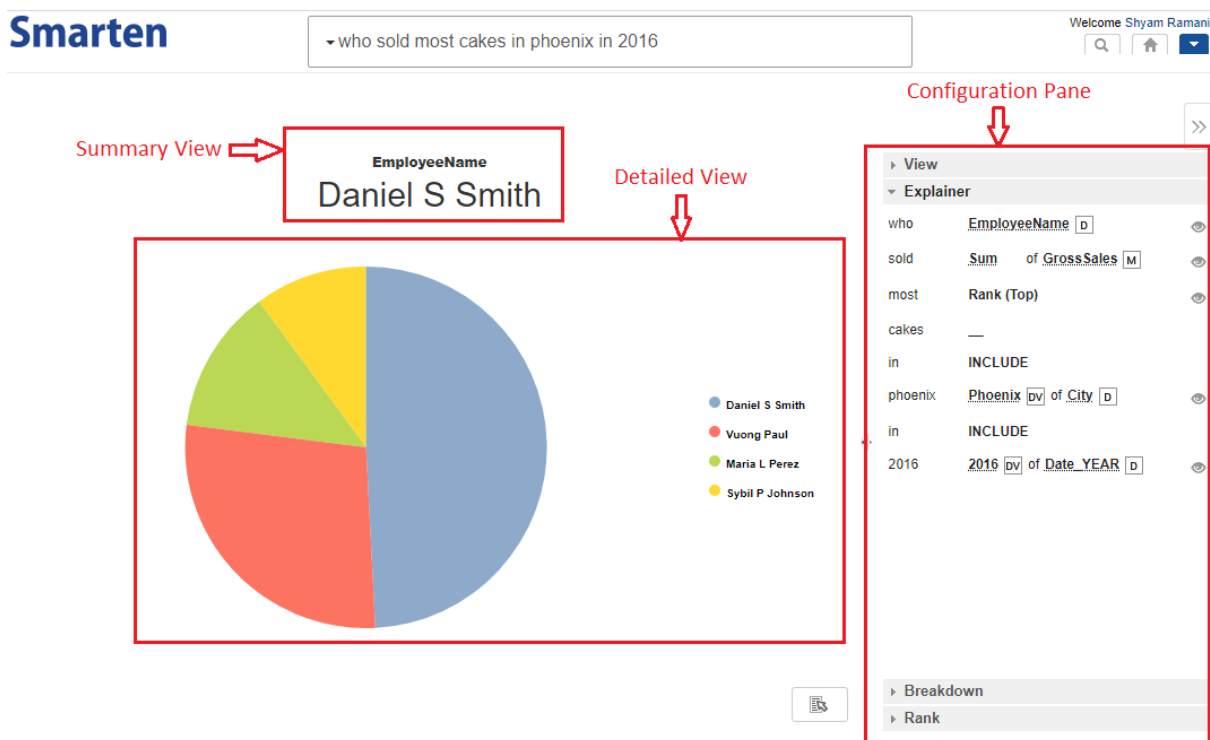
2.4 Analyze and use Clickless Analytics Results

After you enter a query, the system retrieves the relevant result. The system creates an appropriate visualization based on the result retrieved and displays the visualization.

The user can explore the results by configuring visualisations, filters, breakdown columns, and ranks.

2.4.1 Result View

The result view displays the resultant data for the query. The following image depicts the result displayed for the query:
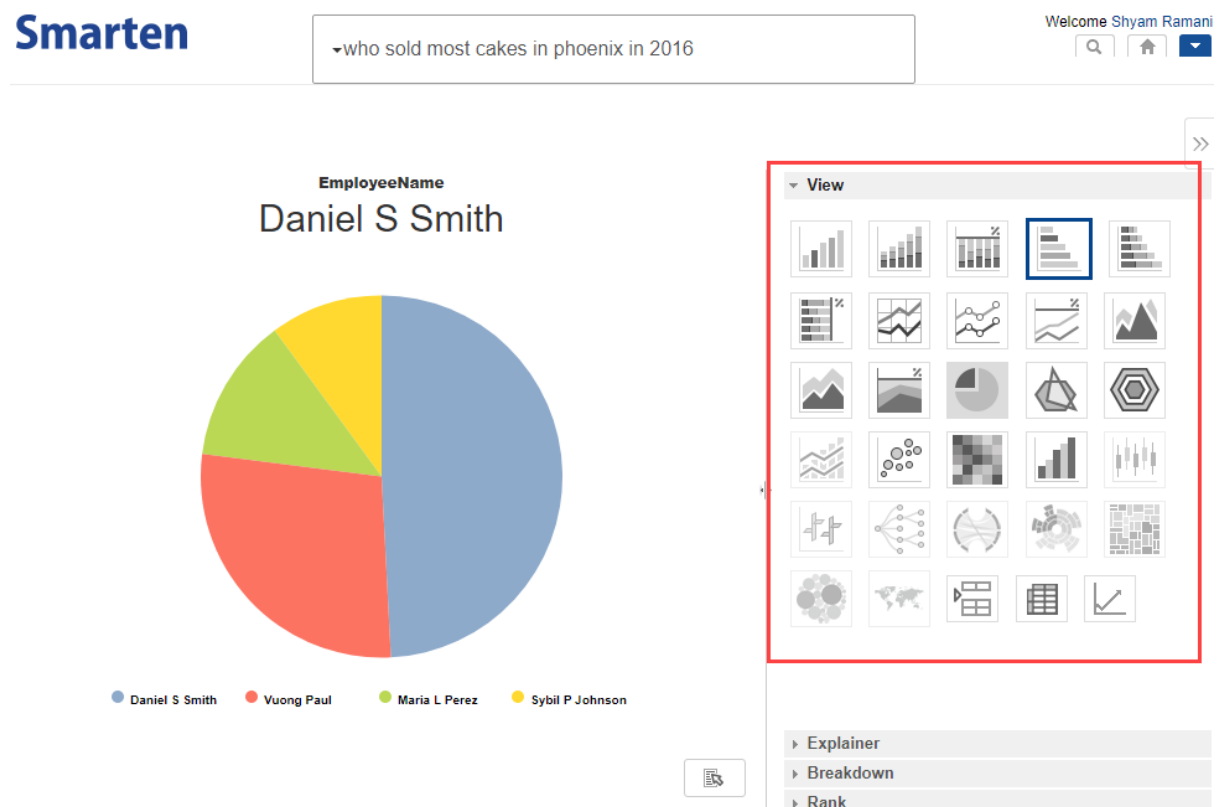

CLICKLESS ANALYTICS: RESULT VIEW

The result view contains the following sections:

- **Summary View**: This section provides the summary of the result. For example, in the above image, the query is to find the employee who sold the maximum number of cakes in Phoenix in 2016, and the summary view displays the name of the employee. Some query results may not display a summary view.

- **Detailed View**: This section provides visualisation for the query result. The information in the detailed view is displayed in the form of a SmartenView, crosstab, tabular, or KPI object. For example, in the above image, the result is displayed as a pie chart.

- **Configuration Pane**: This section allows you to change the view and the result. You can change the view, add filters, add breakdown columns, and set ranking to further explore the results. It also provides Explainer view to explain translations of natural query tokens performed by the system.

2.4.1.1 Visualization

The system generates the resultant object using the most appropriate visualisation for the query. Users can change the views to explore different visualisations supported by the system. Visualisations that are not supported by the schema of the resultset are disabled.

CLICKLESS ANALYTICS: THE VIEW PANE

2.4.1.2 Explainer

Explainer provides explanation of the query and how natural query tokens are interpreted by the system. It shows each token of the query and how it is mapped with dataset columns and values or aggregation operations or other keywords.

For example, in the following image, we have used the query "Who sold the most cakes in Phoenix in 2016," and the Explainer section shows which column and values the system has used for each token of the query:

CLICKLESS ANALYTICS: EXPLAINER

In the above image, the 'D' and 'M' icons next to the column name indicates whether that column is mapped as a dimension or a measure. You can modify the result by activating or deactivating a particular token, change the column mapping, and change the aggregation used for the measure columns.
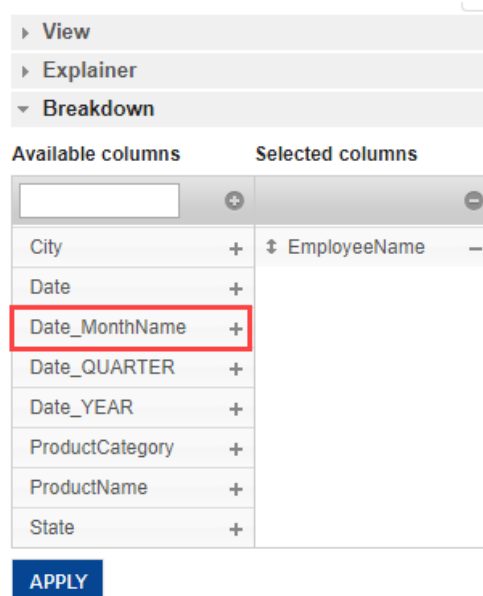
The following image shows the change after we deactivated the Phoenix token from the query and the change in the resultset:
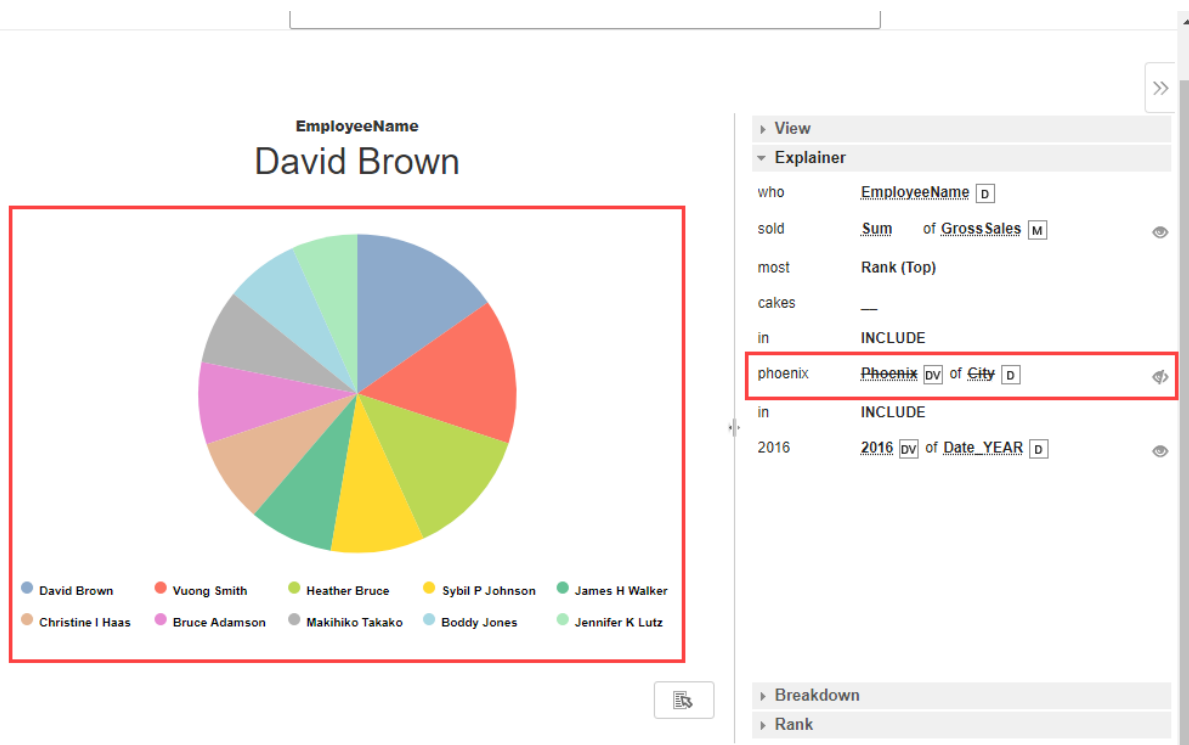
**CLICKLESS ANALYTICS: DEACTIVATING A TOKEN FROM THE QUERY**

2.4.1.3 Breakdown

The Breakdown section lists the columns from the selected dataset that you can use to further break down the result of the query. For example, if the query is to find the employee who sold the most cakes in 2016, you can use columns in the Breakdown section to narrow down the results, e.g., by month, quarter, and city. You can add or remove dimension columns to modify the resultant object based on your requirement. For example, in the following image, you can add the "Date_MonthName" column to modify the resultant object to display the sale of cakes for each month:

CLICKLESS ANALYTICS: BREAKDOWN SECTION

The following image shows the change after we add the "Date_MonthName" column and the change in the resultset:
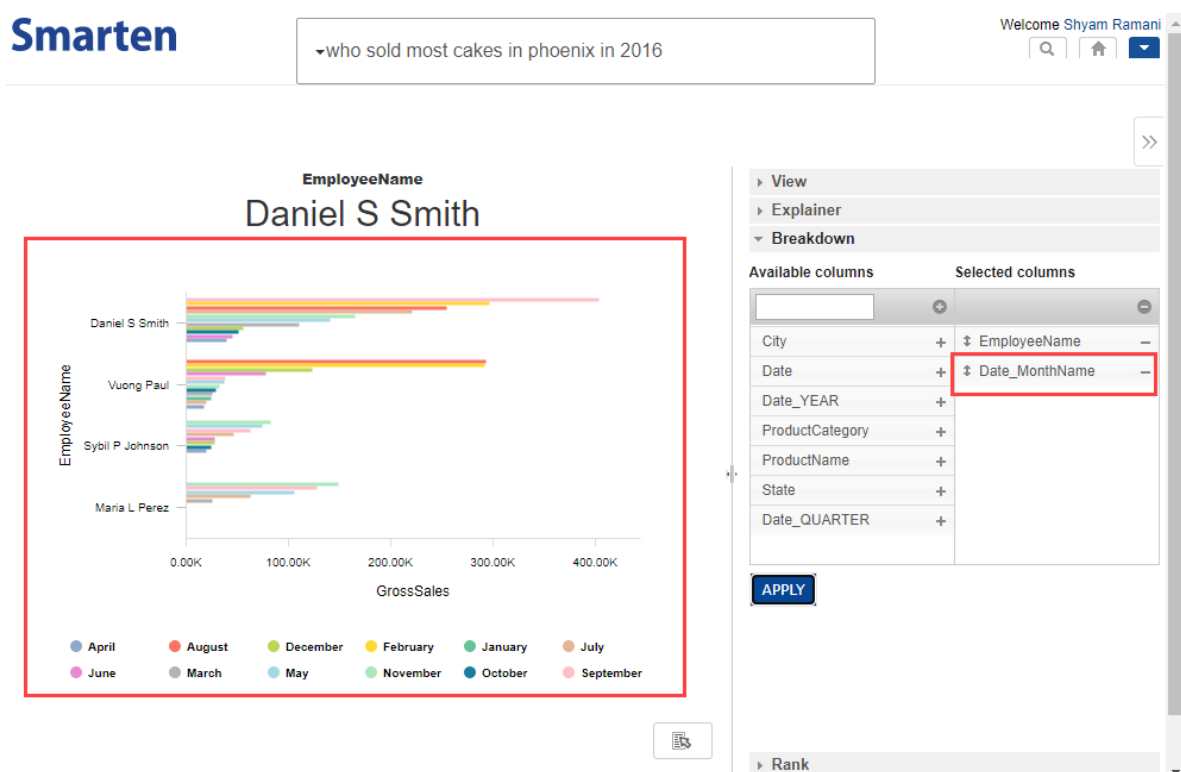
CLICKLESS ANALYTICS: ADDING AN AVAILABLE COLUMN

2.4.1.4 Rank

The Rank section is shown when a query uses rank tokens, such as most, highest, and lowest. The Rank section provides information about the dimension columns and measure columns on which the rank is applied. You can change the columns on which the rank is applied, sorting from "Top" to "Bottom" and vice versa, and update the number of values to be displayed.
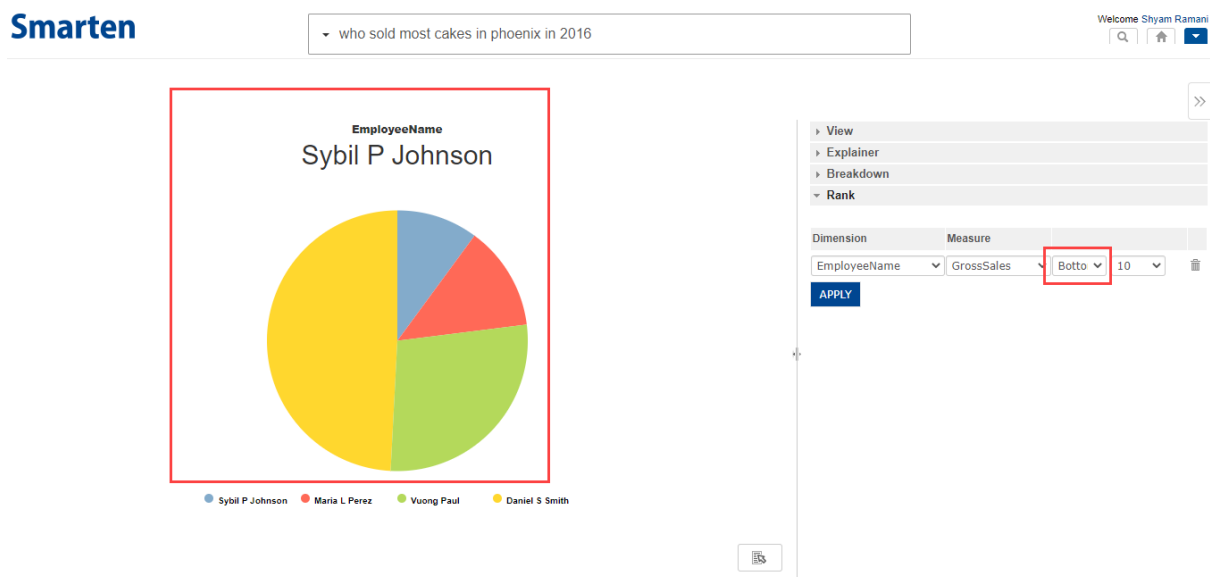

CLICKLESS ANALYTICS: RANK SECTION

The following image shows the change after we change the sorting to "Bottom" and the change in the resultset:

CLICKLESS ANALYTICS: UPDATING THE RANK

2.4.1.5 View Dataset Columns

The dataset columns section provides details about the available columns in the selected dataset. This can help the user look at dataset metadata and available columns for drafting queries.

CLICKLESS ANALYTICS: DATASET COLUMNS

2.4.2 Explore Result in Detail

The **Explore in detail** option allows you to open the resultant object in the respective Smarten object detailed view, and you can explore that object in detail and also save the object for future use.
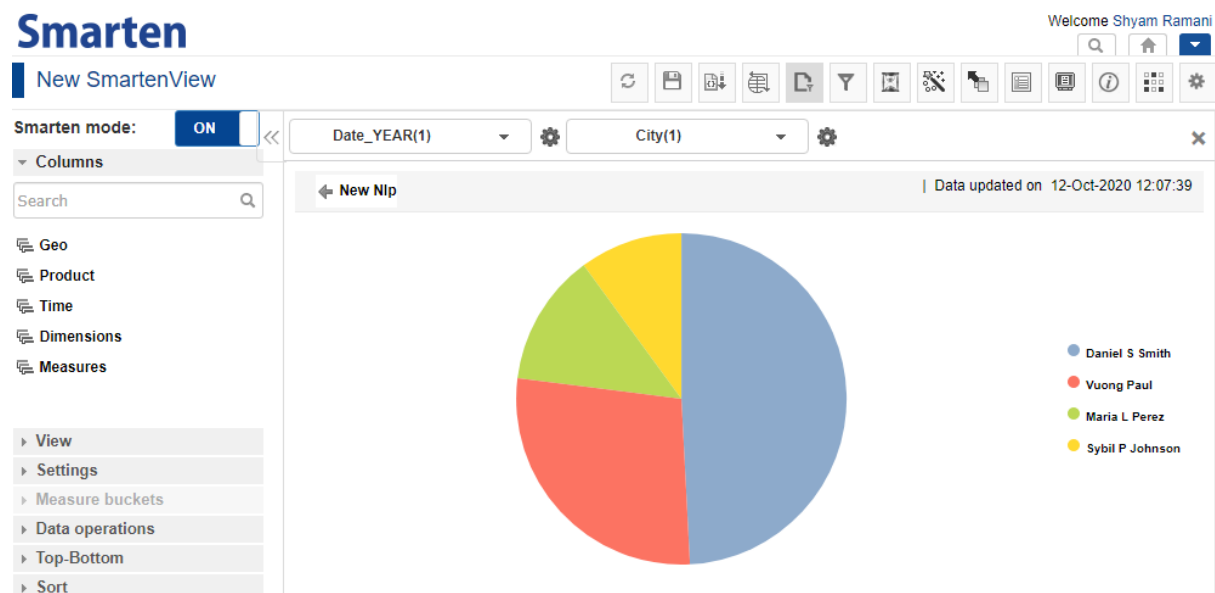
CLICKLESS ANALYTICS: EXPLORE IN DETAIL

3 Predefined System Definitions

Clickless Analytics uses predefined definitions that are specified in various files while processing a query. These definitions are used to accurately interpret the query you have provided. The Administrator can update these files to add or modify any definitions as per business requirements and evolving vocabularies.

The system definition files contain such information as synonyms, abbreviations, dates of festivals, and date format. These are the CSV files that contain relevant definitions that are used by Clickless Analytics to interpret the query and to get appropriate results.

The following system definition files are located at the *<Smarten_data_folder>/resources/NLP_PRE_REQUISITES* folder in your Smarten deployment:

- SMARTEN_AUTO_SYNONYMS

- SMARTEN_NLP_COMPARISION

- SMARTEN_NLP_CURRENCY

- SMARTEN_NLP_WORD_WEIGHTAGE

- SMARTEN_NLP_ABBREVIATIONS

- SMARTEN_NLP_DATE_PERIOD_DEFINITIONS

- SMARTEN_NLP_OPERATORS

3.1 System Files Data Structure and Examples

The system definition files are already populated with the relevant data. You can add or update the values in any of these files based on your requirements.
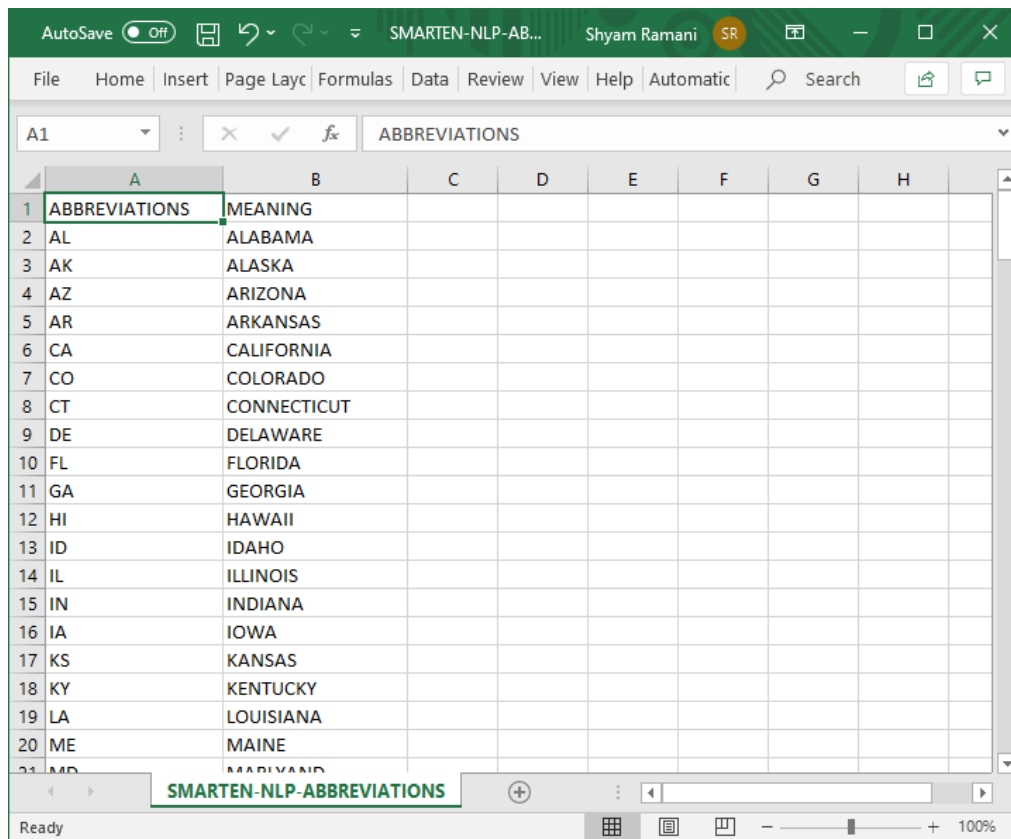
3.1.1 Abbreviations

The SMARTEN_NLP_ABBREVIATIONS system definition file contains the abbreviations that may be used in the data available in the datasets or the query. Values in this file are used to map the abbreviations used in the datasets with the full name in the query. For example, if you have used Arizona in the dataset and AZ in the query, as this abbreviation is available in the system file, the system will be able to map both values and retrieve accurate results.

The structure of the SMARTEN_NLP_ABBREVIATIONS system definition file is as below:

Column Name	Description
Abbreviation	Contains the abbreviated values.
Meaning	Contains the meaning of those abbreviations.

The following image shows the file layout with some of the default values available in the SMARTEN_NLP_ABBREVIATIONS system file:
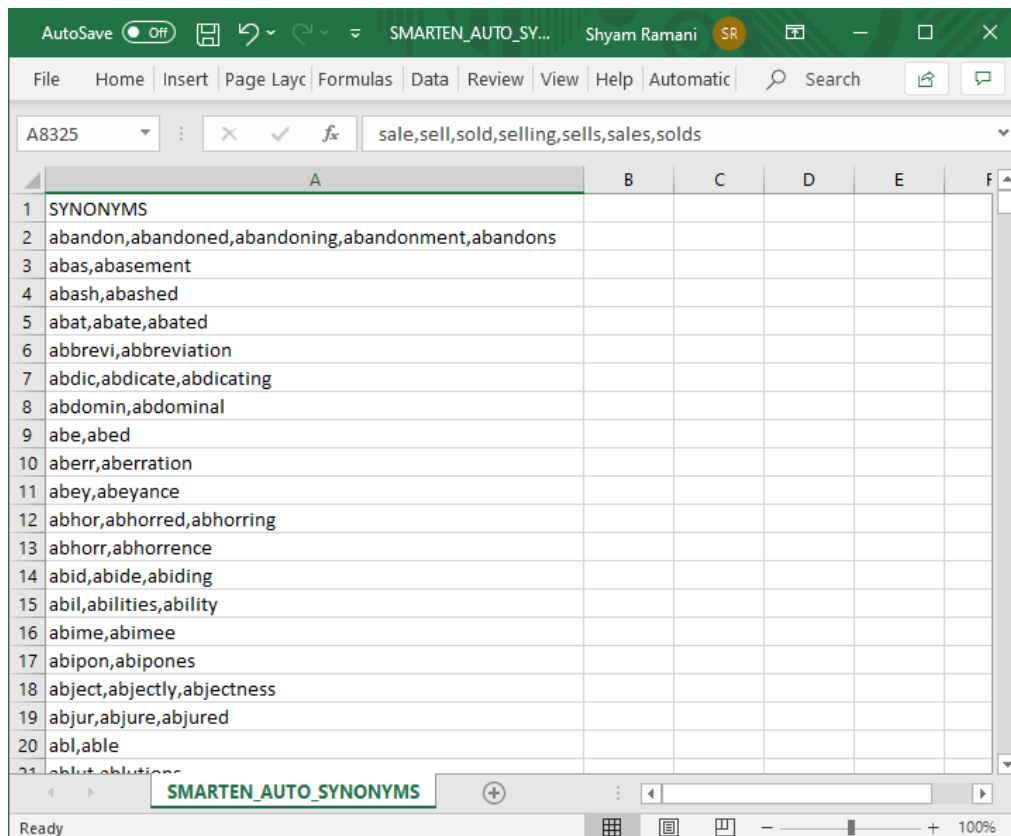
CLICKLESS ANALYTICS: ABBREVIATIONS DEFINITION FILE

3.1.2 Synonyms

The SMARTEN_AUTO_SYNONYMS system definition file contains multiple synonyms of a single word. For example, sell, sold, selling, sells, and sales are defined as synonyms in the SMARTEN_AUTO_SYNONYMS definition file. Values in this file are used to map the other synonyms used in the datasets or in the query. For example, if you have used "Sale" in the dataset and "Sold" in the query, the system will retrieve the accurate results.

The structure of the SMARTEN_AUTO_SYNONYMS system definition file is as below:

Column Name	Description
Synonyms	Contains the comma separated synonyms of the same word.

The following image shows some of the default values available in the SMARTEN_AUTO_SYNONYMS system file:
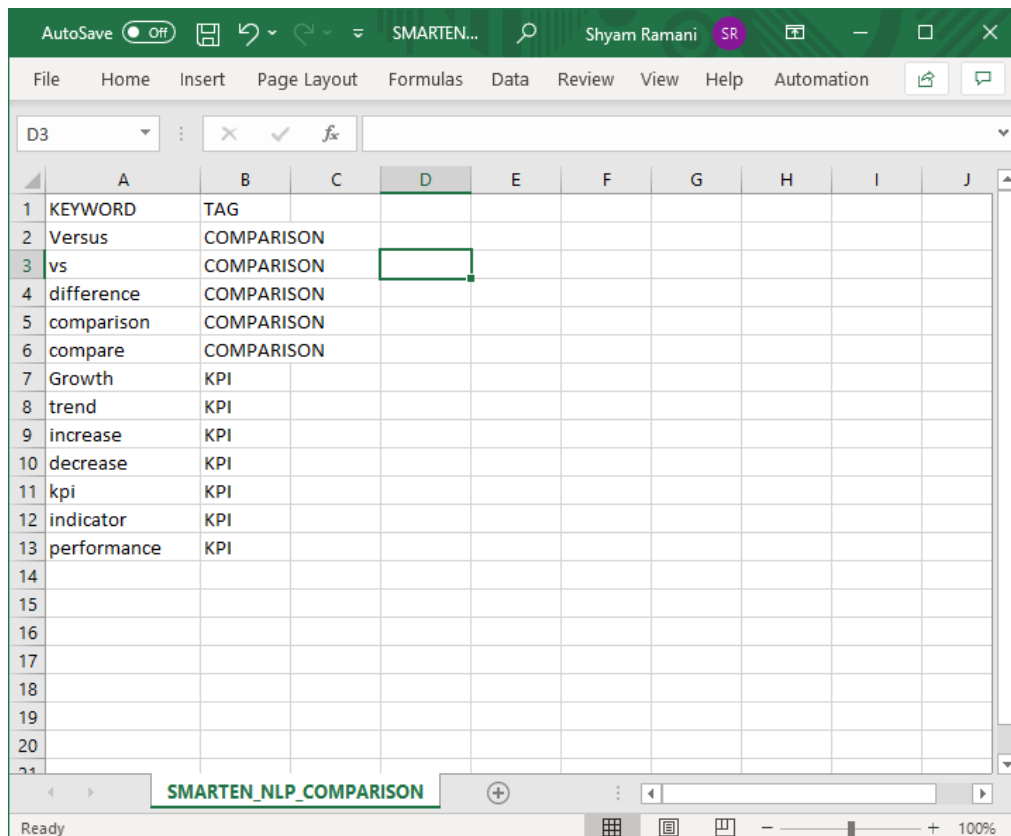
CLICKLESS ANALYTICS: SYNONYMS DEFINITION FILE

3.1.3 Comparison

The SMARTEN_NLP_COMPARISON system definition file contains information about the keywords that Clickless Analytics uses to identify whether that keyword is used to make a comparison or retrieve a KPI object. You can add keywords for comparison or a KPI category as per your requirement.

The structure of the SMARTEN_NLP_COMPARISON system definition file is shown below:

Column Name	Description
Keyword	Contains the keywords.
Tag	Specifies the keyword category and whether the system must use that keyword as COMPARISION or KPI.

The following image shows some of the default values available in the SMARTEN_NLP_COMPARISION system file:

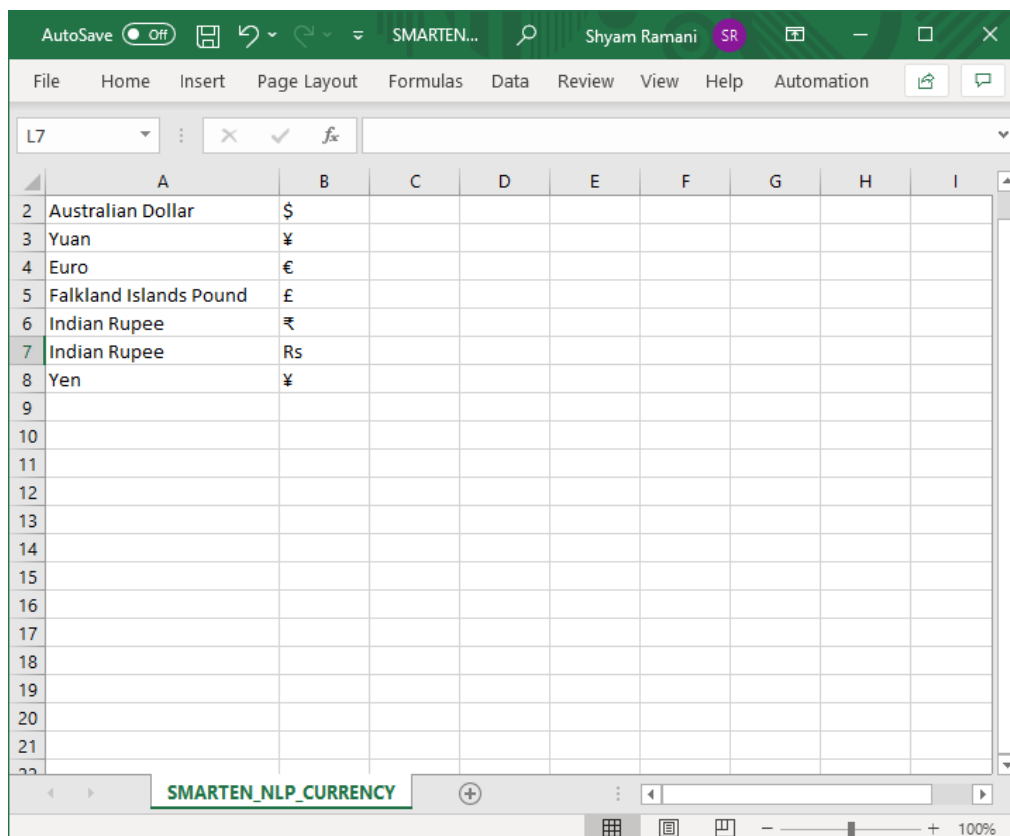**CLICKLESS ANALYTICS: COMPARISON DEFINITION FILE**

3.1.4 Currency

The SMARTEN_NLP_CURRENCY system definition file contains information about the various currencies and their symbols. You can add the name of the currency and its symbol as per your requirement.

The structure of the SMARTEN_NLP_CURRENCY system definition file is as below:

Column Name	Description
Name	Contains the name of the currency.
Symbol	Contains the symbol of the currency.

The following image shows some of the default values available in the SMARTEN_NLP_CURRENCY system file:

CLICKLESS ANALYTICS: CURRENCY FILE

3.1.5   Word Weightage

The SMARTEN_NLP_WORD_WEIGHTAGE system definition file contains information about whether a keyword must be used as a word weightage or a linker. Word weightage is used to interpret whether a keyword is positive or negative. For example, the "maximum" keyword is considered positive, and the "minimum" keyword is considered negative. Similarly, the linker is used to interpret whether a keyword represents inclusion or exclusion of value. For example, when "with," "within," and "from" are used in the query, they mean inclusion, and "barring," "except," and "anti" mean exclusion. You can add keywords for word weightage or linker based on your requirements.

> Note: If you are adding a keyword for the Word Weight category, you can only specify the "Positive" or "Negative" tag for the keyword. Similarly, if you are adding a keyword for the Linker category, you can only specify the "Include" and "Exclude" tag for the keyword.

The structure of the SMARTEN_NLP_WORD_WEIGHTAGE system definition file is as below:

Column Name	Description
Keyword	Contains the keywords.
Tag	Specifies the keyword tag value. For Word Weightage, the possible value can be POSITIVE or NEGATIVE. For the Linker keyword, the possible value can be INCLUDE or EXCLUDE.
Category	Specifies the keyword category whether it is WordWeightage or Linker.

The following image shows some of the default values available in the SMARTEN_NLP_WORD_WEIGHTAGE system file:

CLICKLESS ANALYTICS: WORD WEIGHTAGE DEFINITION FILE

3.1.6 Operators

The SMARTEN_NLP_OPERATORS system definition file contains information about the various operators that can be used in the query. You cannot add or remove an entry in this file. You can only update the "Alternate names" column of an operator. For example, for the "+" operator, you can specify "plus," "addition," "add," and "sum" as its alternate names. The system will interpret these keywords in the query and apply the appropriate operation to get the result.

The structure of the SMARTEN_REPLACE_OPERATORS system definition file is as below:

Column Name	Description
Operator	Contains the operators.
Alternate names	Contains the alternate names for the operators.
Type	Contains information about the type of operators, such as Math operator, Operator, and Aggregator.
Linker	Contains information about Linker text.
Display_Value	Contains the display value of the operator that will be displayed for the operators in the Explainer.
Operator_Column_Type	Contains information about the type of the column for which this operator can be used, e.g., Measure or Dimension.

The following image shows some of the default values available in the SMARTEN_NLP_REPLACE_OPERATORS system file:

CLICKLESS ANALYTICS: REPLACE OPERATORS DEFINITION FILE

4 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