

Analytics Canvas Tutorial:

Data exploration



Overview

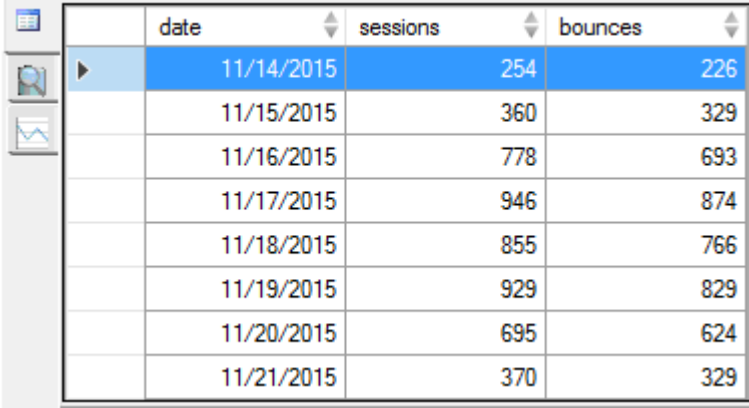
Welcome to Analytics Canvas's tutorial. This is one of a number of detailed tutorials in which we explain how each feature within Analytics Canvas works, with additional context to help you understand it better, and to use it to best suit your own needs.

Analytics Canvas offers the following data profiling features:

- Data Viewer: Data preview of the raw data
- Data Profiler: Data structure and selection of statistics
- Data Chart: Graphical representation of the time series data.
- Data Profiler Block: Deeper exploration of the data.

This tutorial will introduce the main functionalities of each data profiling feature.

Data Viewer



	date	sessions	bounces
▶	11/14/2015	254	226
	11/15/2015	360	329
	11/16/2015	778	693
	11/17/2015	946	874
	11/18/2015	855	766
	11/19/2015	929	829
	11/20/2015	695	624
	11/21/2015	370	329

Figure 1 Data Viewer

At the bottom of the screen, you will be able to see a preview of the data that you loaded in. You can also open Data Viewer by clicking on the input or output stub of a block, or on a connector line between blocks.

There are many things available to you in Data Viewer. If you want to sort the data in a column, in ascending or descending order, you can click on the little triangles in the right corner next to the column names. You can explore the data by right-clicking and selecting "Show Data Type by Color". Dates will be highlighted in light blue and numbers in brown.



Data Profiler

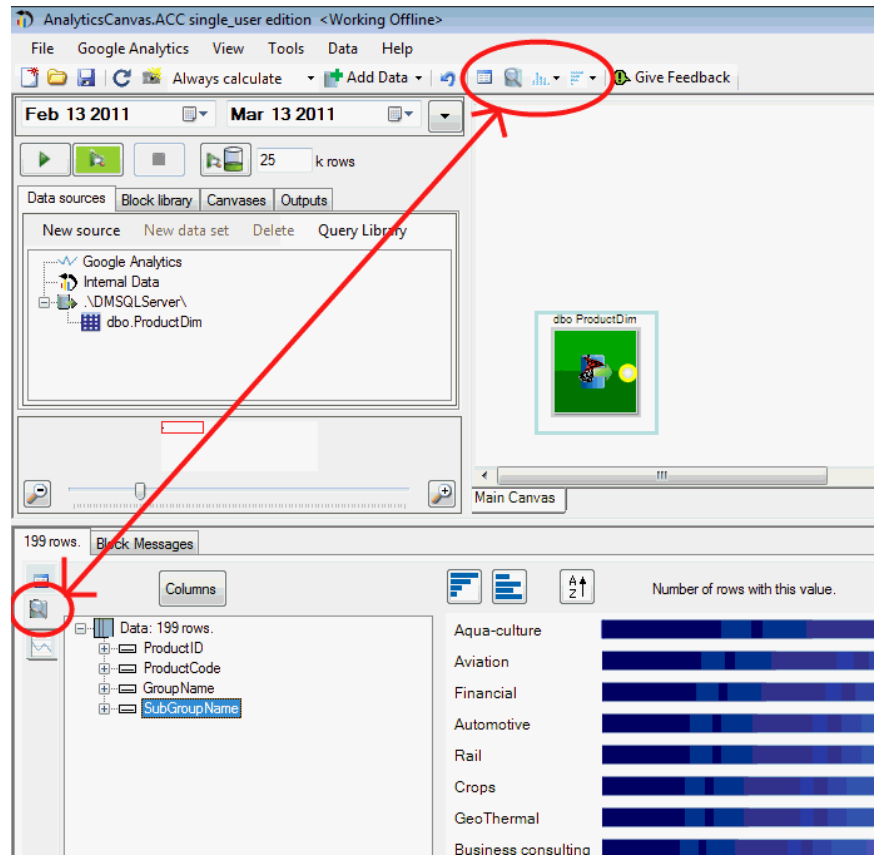


Figure 2 Accessing the Data Profiler

Data Profiler allows you to quickly evaluate the quality of your data. To access the Data Profiler, click on the input or output stub of a block, or on a connector line between blocks. Then select the “Data Profiler” tab at the lower left of the screen, or in the toolbar buttons above the Main Canvas.

Column Name	Data Type	Null Rows	Missing Rows	Populated Rows	Completeness	Cardinality	Uniqueness	Min	Max	Average
date	DateTime	0	0	31	100%	31	100%	11/1...	12/1...	-
sessions	Number	0	0	31	100%	31	100%	254	946	654.387...
bounces	Number	0	0	31	100%	30	96%	226	874	589.161...

Figure 3 Using Data Profiler

If your dataset is too large, you may be prompted to select which columns you wish to profile; otherwise all columns are provided.

For each column that is profiled, the following metrics are calculated:

- **Column Name** - the name of the column being profiled.



- **Data Type** - the declared data type for the column.
- **Null rows** - the count of rows that contain the Null value.
- **Missing rows** - for string types the count of rows containing the empty string.
- **Populated rows** - the count of rows that are not Null or missing.
- **Completeness** - the percentage of rows that are Populated (Populated/Total Rows).
- **Cardinality** - the number of unique values contained in the row.
- **Uniqueness** - the cardinality divided by the number of populated rows.
- **Minimum** - the minimum value within the column for numeric columns, the earliest date for date columns.
- **Maximum** - the maximum value within the column for numeric columns, the latest date for date columns.
- **Average** - the average of all values for numeric columns.

You can drill down into the rows by clicking on the bar in the column category count graphs.

Data Chart

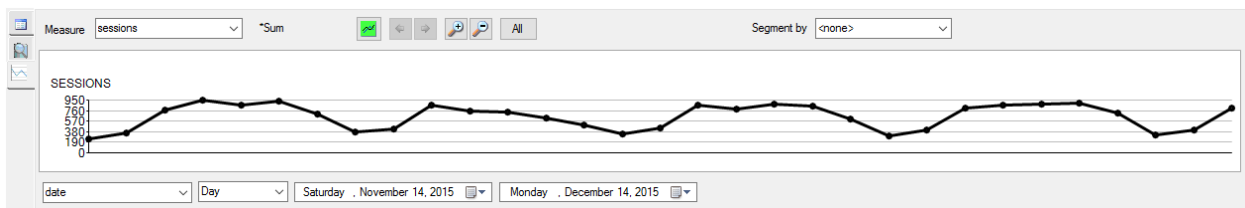


Figure 4 Data Chart

The Data Chart tab is used to visualize the dataset. The Data Chart displays variables as a time series chart, and it can be used to examine changes in data over time. You can zoom in and out by clicking on the data and selecting the time frame that you are interested in.

Conclusion

Thank you for reading this tutorial. We invite you to continue with the tutorial training to learn more about using Analytics Canvas.