Analytics Canvas Tutorial:

Data Profiler Block

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.
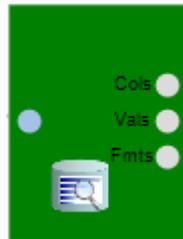
Data Profiler Block

Figure 1 Data Profiler Block

The Data Profiler Block can be found in the Block Library. Simply drag and drop it onto the Main Canvas, and connect it to the input data to start the analysis.

By using the Data Profiler Block, data on your Canvas can be sampled and profiled, and the results used just as in any other block. Use this block to produce information about your data, and to monitor data profile trends over time.

The top output of the Data Profiling Block contains the following:

- **Profile Set Name** - the name of the profiler block; many blocks can be used within one canvas.

- **Capture Date** - the date/time stamp of the profiling run.

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

- **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.

- **Maximum** - the maximum value within the column for numeric columns.

- **Average** - the average of all values for numeric columns.

- **Earliest Date** - the earliest date for date columns.

- **Latest Date** - the latest date for date columns.

		Profile Set Name	Capture Date	Column Name	Null Rows	Missing Rows	Populated Rows	Completeness	Cardinality	Uniqueness	Minimum	Maximum	Average	Earliest Date	Latest Date
	▶	Data profiler blo...	12/15/2015 ...	date	0	0	31	100	31	100	0	0	0	11/14/2015	12/14/2015
		Data profiler blo...	12/15/2015 ...	sessions	0	0	31	100	31	100	254	946	654.387...	_null_	_null_
		Data profiler blo...	12/15/2015 ...	bounces	0	0	31	100	30	96	226	874	589.161...	_null_	_null_

Figure 2 Data Profiler Block Top Output

The Data Profiler Block is a great tool with which to explore your data and drill deeper into data quality issues. Custom rules can be created to address specific issues with data. For example, we can explore the format of the column "Transaction N" to see if it contains letters. We set the rules, and click the "Run" or "Refresh" buttons.

Figure 3 Data Profiler Block Rules

As you can see, there are no letters in the column "Transaction N".

Conclusion

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