

Viewing Data

Viewing Data

Viewing data as it flows from tool to tool can help you evaluate and validate important decisions, such as whether a process affected data as expected, as you progress through your analysis. Designer provides many windows into seeing and understanding the data in our workflow, providing the opportunity to quickly assess the condition of data values or thoroughly investigate the values in a column of data.

Workflow Background

Three input datasets contain data on trees surveyed in New York City. The workflow also contains a manually created a look-up table of borough codes and names. At this point in the workflow's development, some changes to column names, column order, and data types inputs have been applied. However, before moving ahead with further analysis, it is important to explore the data to plan the next steps in the workflow.

Exploring the Results Window

First, view the input containing data from trees surveyed that were planted in 2016. Click the Input Data tool.

Clicking the Input Data tool on the canvas populates the Results window with the data for the selected tool. Values from the ten (10) columns and 3000 rows in this input are visible.

Investigate another input: Tree Survey_2018, which is an Alteryx database. Click the Input Data tool to view the data it contains.

The Results window populates with the contents of this input. However, this time, the Results window displays an important message: only a fraction of the 200,000 rows that this dataset contains are being displayed in this view. Large datasets that exceed the memory storage limit of a tool anchor require a different approach to view its contents: a Browse tool.

Drag a Browse tool onto the Canvas and connect it to the Input Data tool. Click "Run" to populate the Browse tool and click the Browse tool on the Canvas to open its Results window.

The Browse Tool's Results Window

By using a Browse tool, all 200,000 rows that this input contains are now displayed in the Results window. In addition to seeing the entire contents of a dataset, the Browse tool offers an interactive experience of data exploration. Data profiling in the Browse tool's Configuration window allows you to assess the quality, distribution and attributes of data in a single column.

Browsing by Data Type

Choose a column to explore with the Browse tool's data profiling.

Date values are plotted as a histogram to display the number of rows containing a date value in a specific month. The values in the column [Planting Date] are a date data type, which is set to a size, or length, of ten (10) characters. Of the 200,000 rows of data, 345 unique dates are represented and 37 cells in the column contain a null value. The distribution of planting data span from the minimum, or earliest, date to the maximum, or latest, date.

The measured height of each tree is a numeric data type, specifically Int 16. When profiling numeric data types, information about the spread of this data such as the minimum and maximum values, the average, standard deviation, variance and percentile values, is available. As a visual demonstration, Data Profiling also includes a frequency histogram of the unique data values and a plot of the spread.

One of the most important uses of the Browse tool's Data Profiling is assessing the quality of data in a column. Data Profiling categorizes data as OK, Empty, Null, or Not OK, providing important insight into additional data cleansing that may be required for downstream analysis to be carried out successfully. For string data types, data profiling returns information that is more relevant to working with these values, such as the distributions of the values' length, alphanumeric range, and the presence of whitespace.

The values in the column [Centroid] are spatial objects and, instead of being viewed in the Browse tool's Data Profiling, must be viewed with a map.

Viewing Data without the Browse Tool

While Browse tools provide in-depth views into the quality and condition of your data, they are memory-intensive and can increase the time and processing power needed to run a workflow. As you build out a workflow, investigate data frequently and quickly using the data stored in a tool's input and output anchors.

Click the Select tool.

By default, a tool's Results window displays the results, or output, from that tool. View the data before a change to the data stream was applied in the Select tool. In the Results window, click the icon for the input anchor.

Prior to changes applied in the Select tool, the incoming data contained columns that were unnecessary for analysis, named unclearly, and not placed in a desired order. Confirm that the changes to re-order, rename and remove columns were applied by viewing the Select tool's output.

Click the icon for the output anchor.

The views from the tool anchors allows you to quickly assess that a tool's configuration has correctly altered your data, allowing you to confidently progress in your workflow development.