

Data QA & Profiling Tools <Doc 000027X-Day 00006Y BI>

Monday, July 31, 2023 4:08 PM

07.10.2023

Profiling Tools like column quality, column distribution and column profile allow you to explore the quality, composition and distribution of your data before loading it into the Power BI Front-End

Column Quality

Shows the percentages of Valid, Errors or Empty Columns
You can Hover over the Column Quality Box to see the Number of Records in each category
Click the ... (Options Menu) to remove duplicates, errors or empty values
Profiling tools are a way to find all of this in one place

Column Distribution

Provides a sample distribution of the data in a column
Hover over the column quality box to see the number of distinct & unique records
Click the ... (Options Menu) to remove duplicates, errors or empty values

Column Profile

Provides a more holistic view of the Data in a column, including sample distribution and profiling Statistics
Column Statistics Provide more detailed profiling metrics
You can Hover over the Value Distribution for Suggested Transformations

Column Quality Steps

At the Top of the Power Query Editor Inside the View Option Click the Column quality Checkbox so that it is Currently Checked to see how much of the data has Errors

The Percentages appear in the first row of the table

By Default Column quality only checks the first 1,000 rows

We need to update the Column Profiling from the first 1,000 rows to the entire data set

At the Bottom of the Power Query Editor Click the Column profiling based on top 1000 rows Option

Click the Column profiling based on entire data set Option

Hover over the Column Header Percentages to view Power Query's Suggestions

Click the Remove Errors Option

Inside Power Query's Applied Steps Section the Removed Errors Step will be added

The Percentages will update

Hover over the Percentages to view Power Query's Suggestions

Click the Remove Empty Option

Inside Power Query's Applied Steps Section the Filtered Rows Step will be added

→ If you want to view only the Errors then...

Hover over the Column Header Percentages Inside after Checking the Column quality Checkbox

Right Click the Percentages

Click the Keep Errors Option

Inside Power Query's Applied Steps Section the Kept Errors Step will be added

A New Table of just the errors will be generated

Click the Error Link inside any of the cells

That specific Error will appear on the Screen Inside the Power Query Editor

Column Distribution Steps

At the Top of the Power Query Editor Inside the View Option Click the Column distribution Checkbox so that it is Currently Checked to see a Sample Distribution of values within each

column

Hover over the Column Header Distribution to see the Distinct and Unique Percentages

The Remove Duplicates Option is available to us, however, in our example we're looking at names so duplicates are necessary

Column Profile Steps

At the Top of the Power Query Editor Inside the View Option Click the Column profile Checkbox so that it is Currently Checked to see Column Statistics and Value Distribution

Sometimes a Column might not provide a lot of value in its Column Statistics and Value Distribution, you can select another column to potentially get a more clear picture

★ Data Prep and QA is often very time consuming and Data Profiling Tools can speed up the Process