ASSIGNMENT: TABLE TRANSFORMATIONS

NEW MESSAGE

From: **Ethan T. Langer** *(Analytics Manager)*

Subject: **Welcome aboard!**

Hello, and welcome to the team!

We're excited that you'll be helping us develop our new internal reports in Power BI. Looks like you've already gotten started, but we have some new data to add to the model.

Could you please create two new queries to connect to the **Product Category Lookup** and **Product Subcategory Lookup** files attached, and help with a few modifications to the product table?

Thanks!
-ETL

Product Category Lookup
Product Subcategory Lookup

Reply Forward

Key Objectives

1. Create queries to connect to the two new .csv files
2. Name your queries **Product Category Lookup** and **Product Subcategory Lookup**
3. Confirm that column headers have been promoted and that all data types are correct
4. Add a new column to extract all characters before the dash ("-") in the **Product SKU** column, and name it "**SKU Type**"
5. Update the **SKU Type** calculation above to return all characters before *second* dash, instead of the first
6. Replace zeros (**0**) in the **Product Style** column with "**NA**"
7. Close and load to your data model

ASSIGNMENT: TEXT TOOLS

NEW MESSAGE

From: **Ethan T. Langer** *(Analytics Manager)*

Subject: **Customer domains**

Hi!

We're looking to better understand where our customers may be coming from, based on their email domains.

Could you please create a new column in the customer table that will allow us do this?

Thanks!
-ETL

Reply Forward

Key Objectives

1. Duplicate the email address column and name it **"Domain Name"**
2. In the new column, remove all text/characters except for the domain name
3. Use transformation steps to clean up and capitalize the domain names (i.e. "**Adventure Works**")
4. Save & Apply changes

PRO TIP: STORAGE & CONNECTION MODES

Power BI Desktop supports several types of **storage** and **connection modes**:

- **Import**: Tables are stored in-memory within Power BI and queries are fulfilled by cached data *(default)*
- **DirectQuery**: Tables are connected directly to the source and queries are executed on-demand at the data source
- **Composite Model (Dual)**: Tables come from a mix of Import and DirectQuery modes, or integrate multiple DirectQuery tables
- **Live Connection**: Connect to pre-published Power BI datasets in Power BI Service or Azure Analysis Services

Import

- ✓ Dataset is less than 1GB (after compression) & fast performance
- ✓ Source data does not change frequently
- ✓ No restrictions on Power Query, data modeling, and DAX functions

DirectQuery

- ✓ Dataset is too large to be stored in-memory
- ✓ Source data changes frequently and reports must reflect changes
- ✓ Company policy states that data can only be accessed from the original source

Composite Model

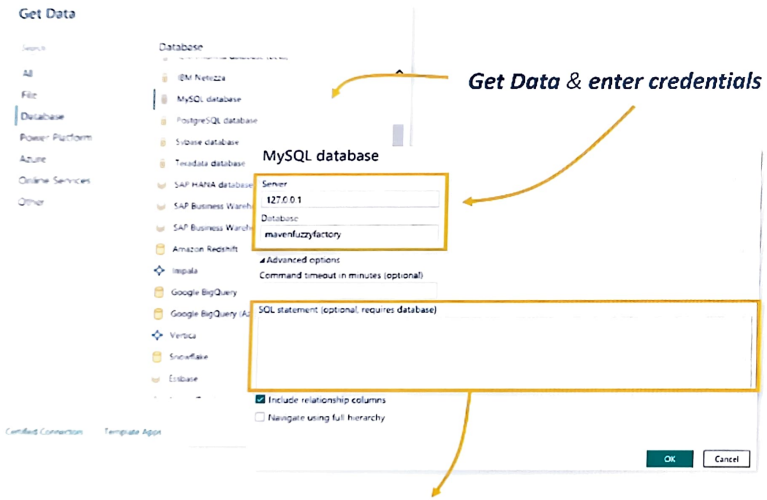
- ✓ Boost performance by setting appropriate storage for each table
- ✓ Combine a DirectQuery model with additional imported data
- ✓ Create a single model from two or more DirectQuery models

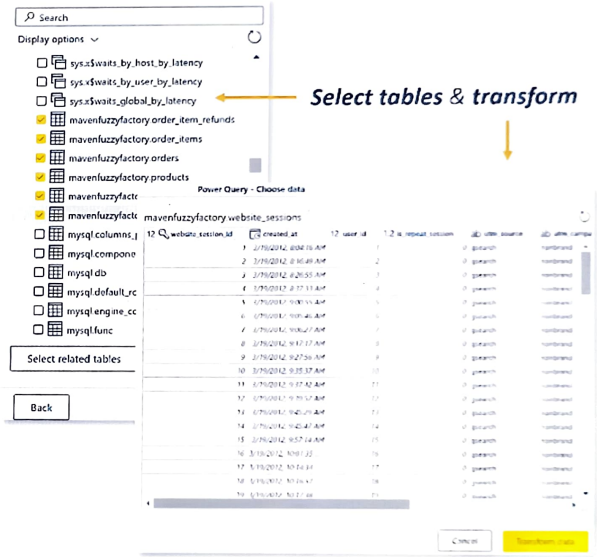
Live Connection

- ✓ Create one dataset that serves as a central source of truth
- ✓ Analyst teams can create different reports from the same source
- ✓ Multi-developer teams where one user builds the model and another works on visualization

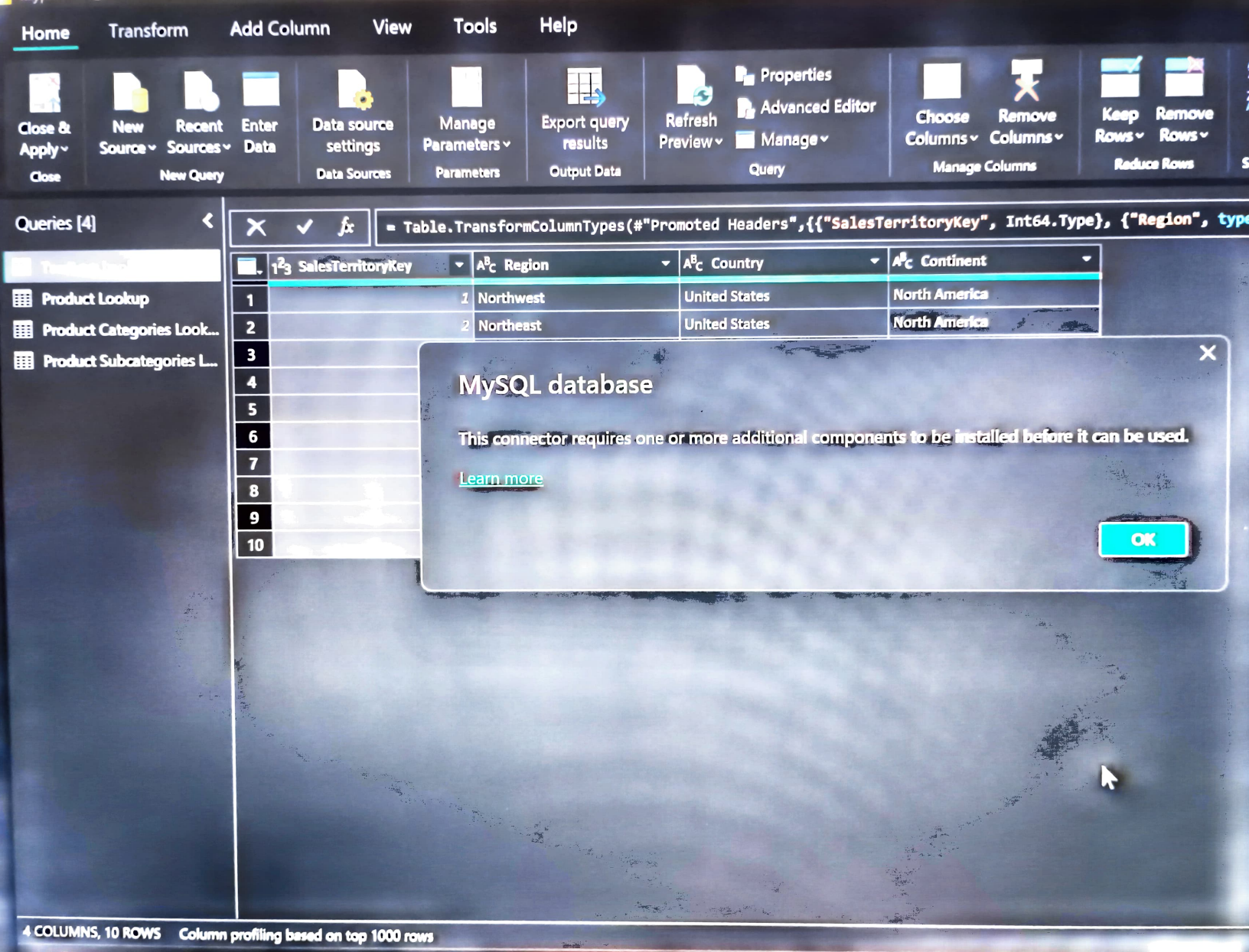
CONNECTING TO A DATABASE

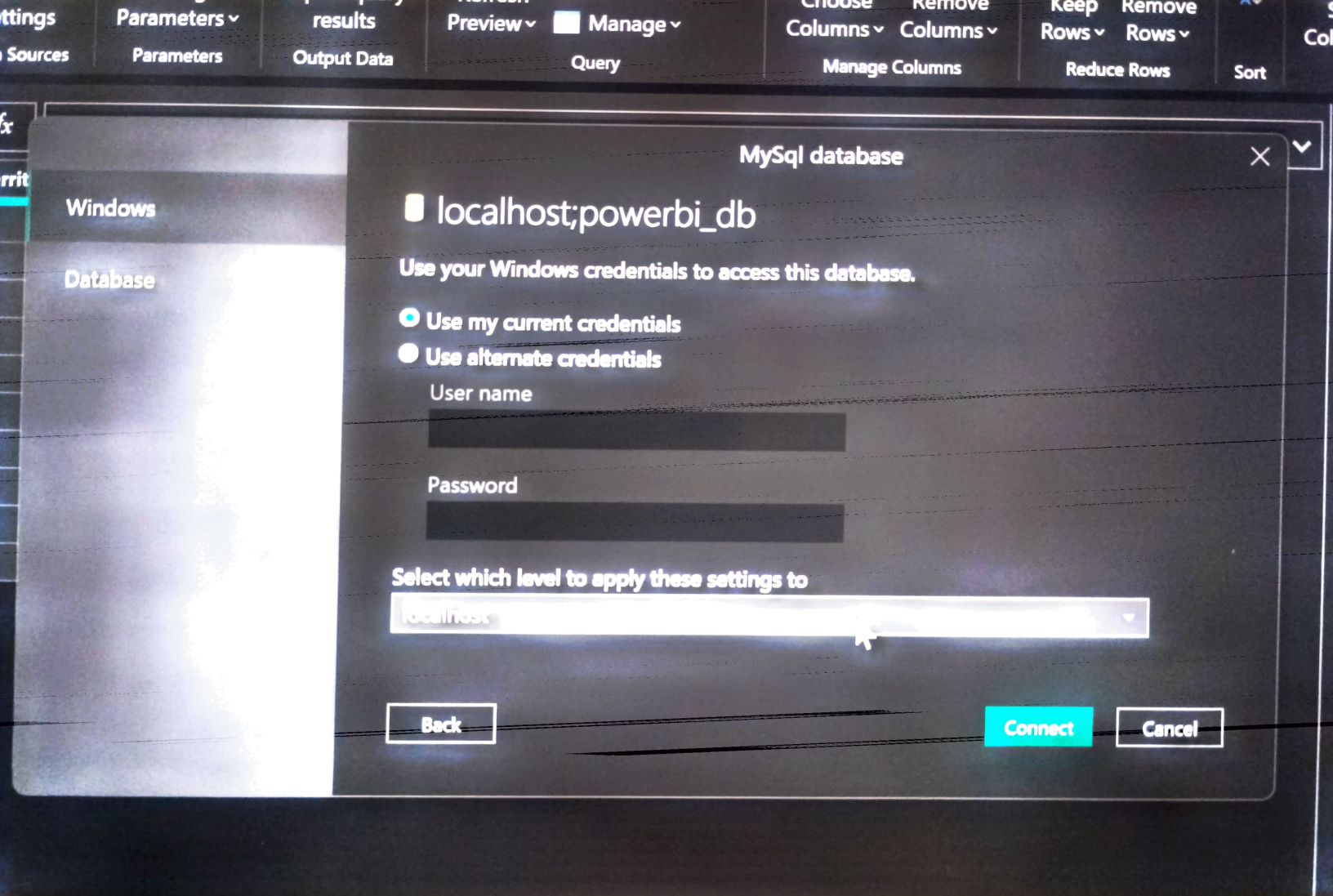
Power Query can connect to data from various **database sources** including SQL Server, MS Access, MySQL, PostgreSQL, Oracle, SAP, and more





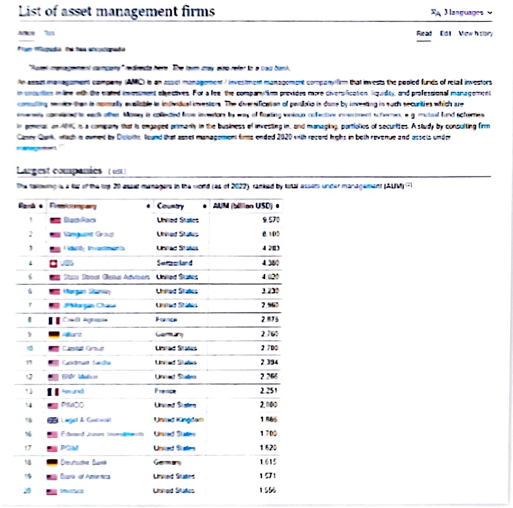
Write custom or advanced queries *with SQL statements (optional)*


Home
Transform
Add Column
View
Tools
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Export query results
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Close
New Query
Data Sources
Parameters
Output Data
Query
Manage Columns
Reduce Rows
Queries [4]
Product Lookup
Product Categories Look...
Product Subcategories L...
= Table.TransformColumnTypes(#"Promoted Headers",{{"SalesTerritoryKey", Int64.Type}, {"Region",
SalesTerritoryKey
Region
Country
Continent
1 Northwest United States North America
2 Northeast United States North America
MySQL database
This connector requires one or more additional components to be installed before it can be used.
Learn more
OK
4 COLUMNS, 10 ROWS
Column profiling based on top 1000 rows


Parameters
Parameters
results
Output Data
Preview
Manage
Query
Choose Columns
Remove Columns
Manage Columns
Keep Rows
Remove Rows
Reduce Rows
Sort
MySql database
Windows
Database
localhost;powerbi_db
Use your Windows credentials to access this database.
Use my current credentials
Use alternate credentials
User name
Password
Select which level to apply these settings to
Back
Connect
Cancel

EXTRACTING DATA FROM THE WEB

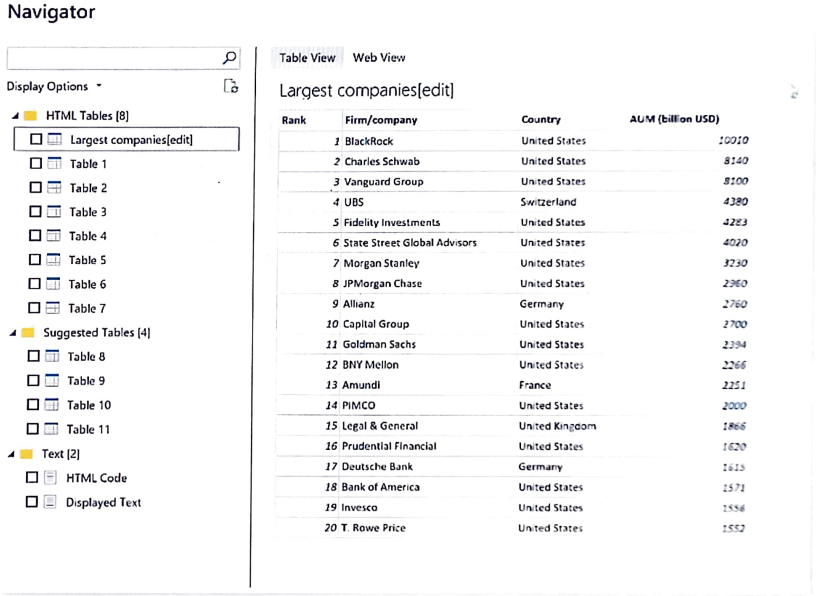
Power Query includes a native **Web connector** for importing web-hosted files (csv, xlsx, etc.) or scraping URLs for anything that Power Query can identify as a structured table


List of asset management firms

Largest companies (edit)

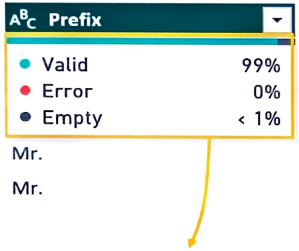
Rank	Firm/company	Country	AUM (billion USD)
1	BlackRock	United States	9 570
2	Vanguard Group	United States	8 100
3	Fidelity Investments	United States	4 283
4	UBS	Switzerland	4 380
5	State Street Global Advisors	United States	4 020
6	Morgan Stanley	United States	3 230
7	JPMorgan Chase	United States	2 960
8	Crédit Agricole	France	2 875
9	Allianz	Germany	2 760
10	Capital Group	United States	2 700
11	Goldman Sachs	United States	2 394
12	BNP Paribas	United States	2 266
13	Amundi	France	2 251
14	PIMCO	United States	2 000
15	Legal & General	United Kingdom	1 866
16	Edward Jones Investments	United States	1 780
17	PGIM	United States	1 620
18	Deutsche Bank	Germany	1 615
19	Bank of America	United States	1 571
20	Invesco	United States	1 556

https://en.wikipedia.org/wiki/List_of_asset_management_firms

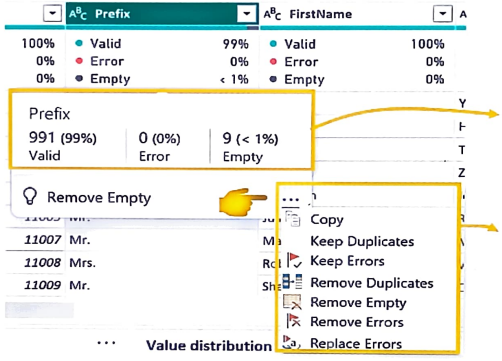


DATA PROFILING: COLUMN QUALITY

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 percentage of values within a column that are* ***valid****, contain* ***errors****, or are* ***empty***



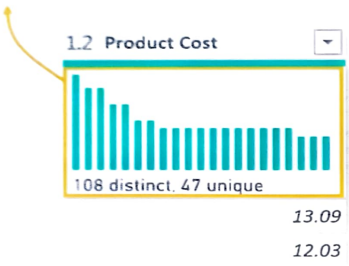
Hover over the column quality box to see the ***number of records*** *in each category*

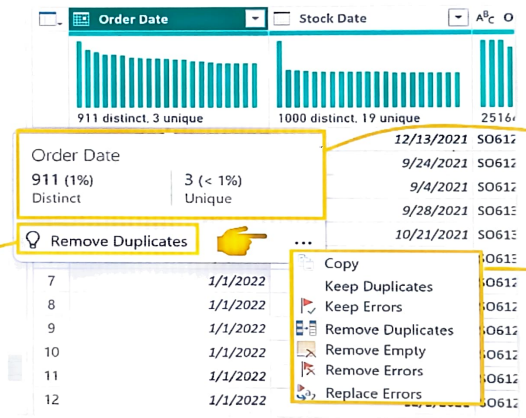
Click the ***options menu*** *to remove duplicates, errors or empty values*

PRO TIP: Profiling tools are a great way to **quickly find and address common data quality issues in one place**, instead of having to manually apply multiple tools or filters

DATA PROFILING: COLUMN DISTRIBUTION

Column distribution *provides a sample distribution of the data in a column*





Suggested action *based on column distribution results*

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*

DATA PROFILING: COLUMN PROFILE

Column profile *provides a more holistic view of the data in a column, including a sample distribution and profiling statistics*

Column statistics *provide more detailed profiling metrics, including:*

Count = 293

(total number of values in column)

Distinct Count = 119

(total number of distinct values, whether they appear once or multiple times)

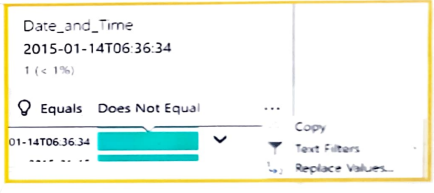
Unique = 82

(total number of values that appear exactly once)

Min & Max

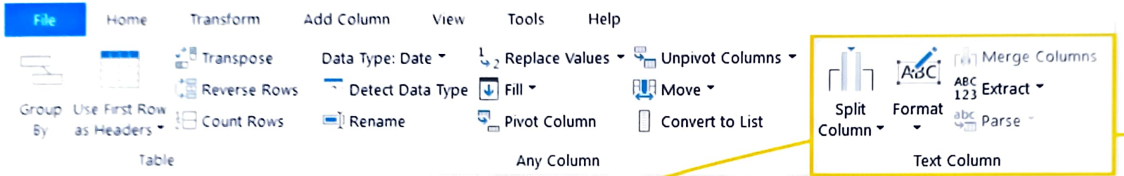
(lowest and highest observed values)

***Note**: Typically only useful for numerical values*



*Hover over the value distribution bar for **suggested transformations** and additional options*

TEXT TOOLS



Split a text column based on a specific delimiter, number of characters, or other attributes

- By Delimiter
- By Number of Characters
- By Positions
- By Lowercase to Uppercase
- By Uppercase to Lowercase
- By Digit to Non-Digit
- By Non-Digit to Digit

- Length
- First Characters
- Last Characters
- Range
- Text Before Delimiter
- Text After Delimiter
- Text Between Delimiters

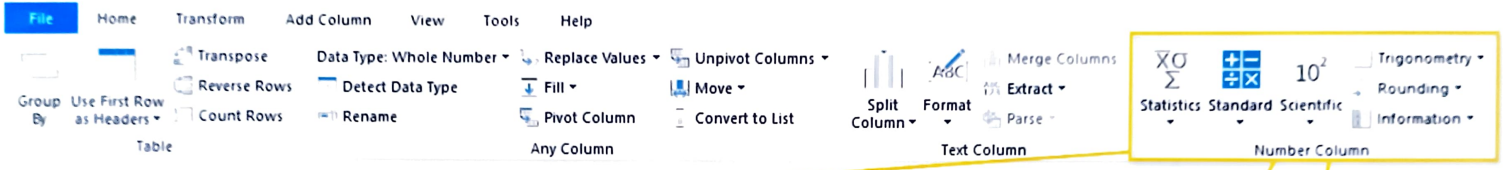
Extract characters from text based on fixed lengths, first/last characters, ranges or delimiters

- lowercase
- UPPERCASE
- Capitalize Each Word
- Trim
- Clean
- Add Prefix
- Add Suffix

Format a text column to upper, lower or proper case, or add a prefix or suffix

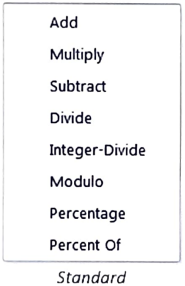
Tip: *Use "Trim" to eliminate leading & trailing spaces, or "Clean" to remove non-printable characters*

NUMERICAL TOOLS

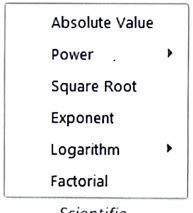


Statistics functions allow you to evaluate basic stats for a selected column (sum, min/max, average, count, count distinct, etc.)

Note: *These tools return a SINGLE value, and are commonly used to explore a table rather than prepare it for loading*



Standard

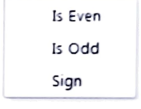


Scientific

Trigonometry

Standard, **Scientific** and **Trigonometry** tools allow you to apply standard operations (addition, multiplication, division, etc.) or more advanced calculations (power, logarithm, sine, tangent, etc.) to each value in a column

Note: *Unlike the Statistics tools, these are applied to each row in the table*



Information tools allow you to define binary flags (*1/0 or TRUE/FALSE*) to mark rows as even, odd, positive or negative

ASSIGNMENT: NUMERICAL TOOLS

NEW MESSAGE

From: **Ethan T. Langer** *(Analytics Manager)*

Subject: **Need some stats for leadership**

Hi again,

Leadership is asking us to validate some high-level stats about our products and customers. Can you please help me answer the following questions?

We don't really need to store these values anywhere, so make sure to restore the tables back to their original state once you're done pulling the stats.

Thank you!
-ETL

Reply Forward

Key Objectives

1. What is our average product cost?
2. How many colors do we sell our products in?
3. How many distinct customers do we have?
4. What is the maximum annual customer income?
5. Return the tables to their original state

SOLUTION: TEXT TOOLS

NEW MESSAGE

From: **Ethan T. Langer** *(Analytics Manager)*

Subject: **Customer domains**

Hi!

We're looking to better understand where our customers may be coming from, based on their email domains.

Could you please create a new column in the customer table that will allow us do this?

Thanks!
-ETL

Reply Forward

Solution Preview