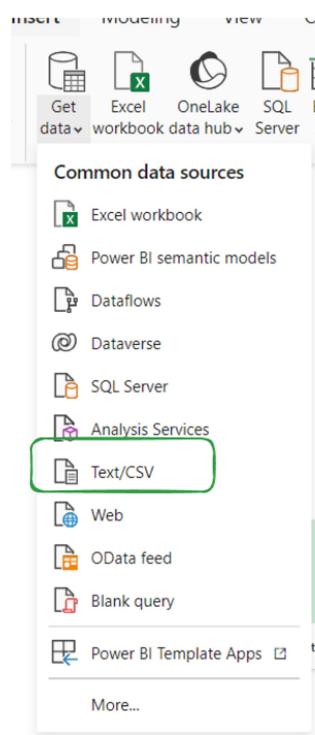
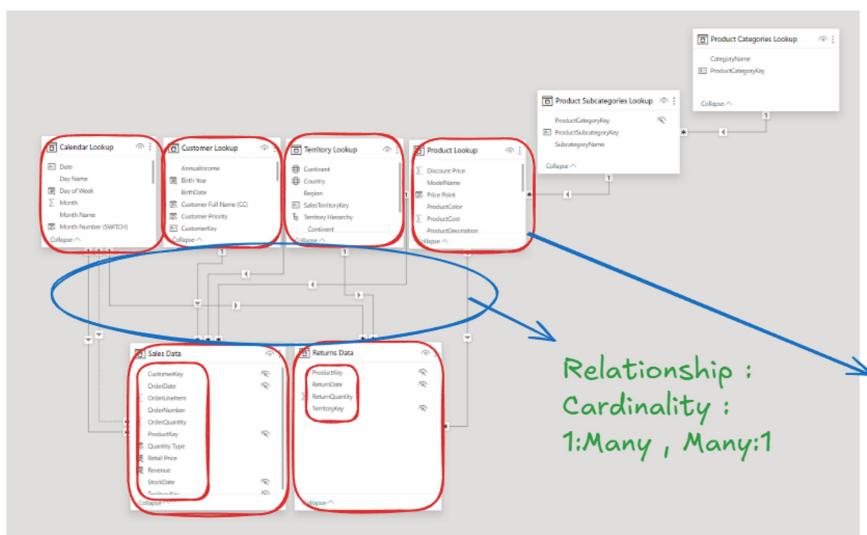
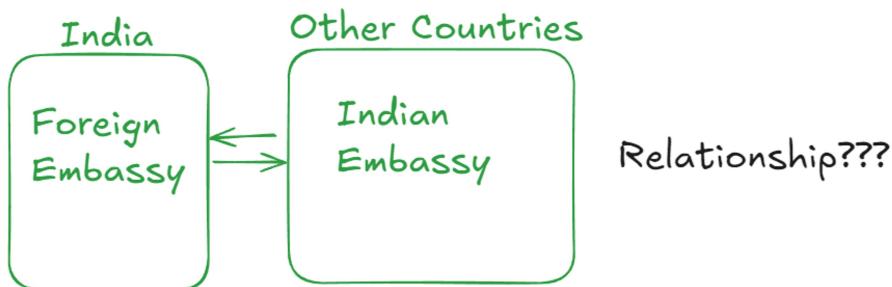
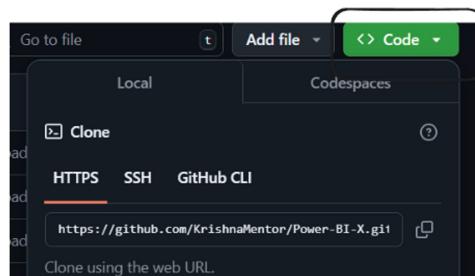


Power BI - Data Transformation - Lecture 2

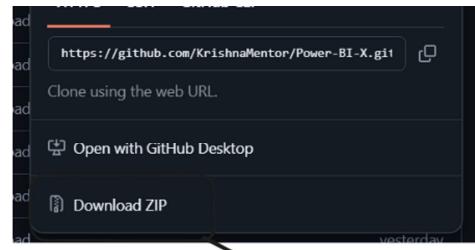
Power BI - Data Transformation - Lecture 2.



Step1 : Load the data from your local System.
[Download it from Github]



More...



Folder.

Step 2 : We all are ready with the dataset.

CONNECTING & SHAPING DATA

In this section we'll connect to source files and cover some of the most common techniques for extracting, cleaning, and shaping data to prepare it for modeling and analysis.

TOPICS WE'LL COVER:

1. Intro to Power Query

2. The Query Editor

3. Data QA & Profiling

4. Calendar Tools

5. Data Connectors

6. Connection Modes

7. Table Transformations

8. Combining Queries

GOALS FOR THIS SECTION:

- Explore Power BI's query editor and understand the role that Power Query plays in the larger BI workflow
- Introduce different types of connectors and connectivity modes available for getting data into Power BI
- Review tools for checking data quality and key profiling metrics like column distribution, empty values, errors and outliers
- Transform tables using text, numerical and date/time tools, pivot and group records, and create new conditional columns
- Practice combining, modifying and refreshing queries

FRONT-END VS. BACK-END



- Power BI Desktop essentially has two distinct environments: a front-end and a back-end

- The front-end includes the Data, Model & Report views, where most of the modeling, analysis and visualization takes place
- The back-end includes the Power Query Editor, where raw data is extracted, transformed, and loaded to the front-end (ETL)

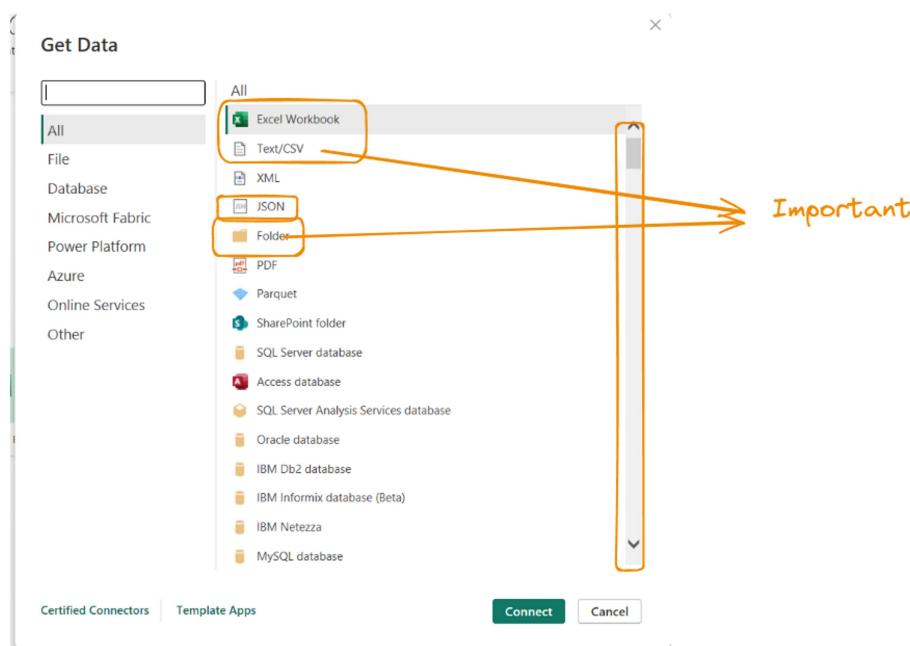
BACK-END

- Connect & extract data using pre-built connectors
- Profile & QA the data to explore, clean and prepare it for modeling and analysis
- Transform & shape tables to add new features, modify values, group records, or sort and filter columns
- Merge or append queries to join and combine them prior to loading to the front-end

FRONT-END

- Build data models by creating table relationships between primary and foreign keys
- Add calculated measures & columns using Data Analysis Expressions (DAX)
- Design reports to visualize the data and create interactive, dynamic dashboards
- Publish & share your Power BI workbooks using Power BI Service (cloud application)

TYPES OF DATA CONNECTORS



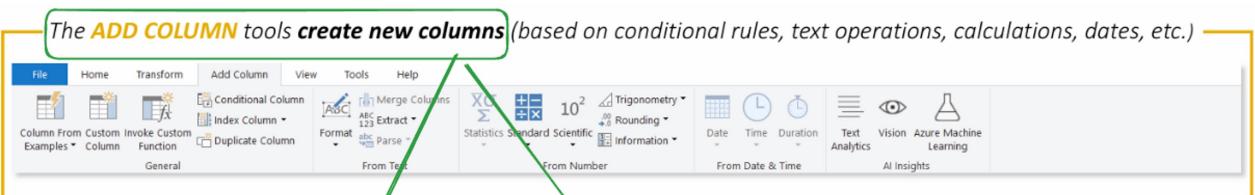
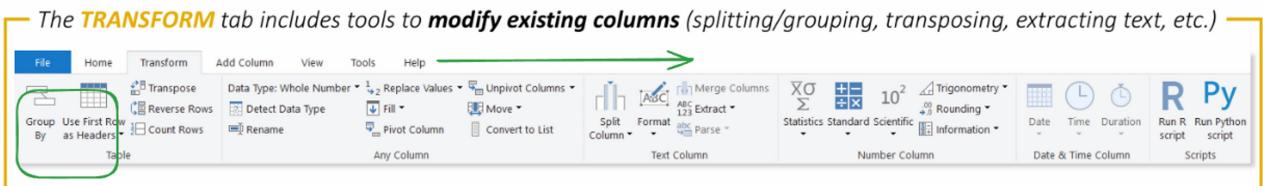
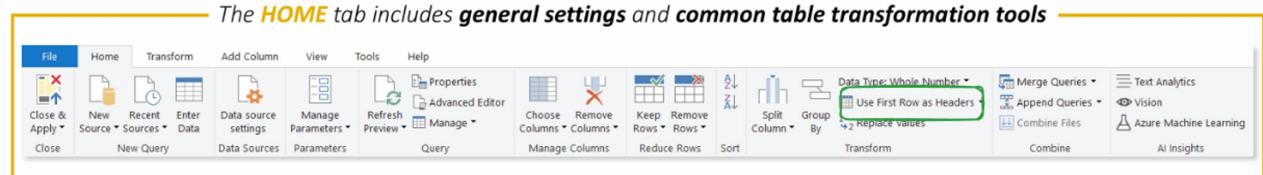
POWER QUERY EDITOR

The screenshot shows the Power Query Editor interface. At the top, there's a ribbon with tabs like File, Home, Insert, Modeling, View, Optimize, Help, and External Tools. The 'External Tools' tab is highlighted with a yellow box. Below the ribbon is a toolbar with icons for file operations, transformation tools, and help. The main area is divided into several sections:

- Queries Pane**: A list of all queries, shown on the left side of the editor.
- Table Preview**: A preview of the data being transformed, located at the bottom center.
- Query Editing Tools**: A panel on the right containing various icons for table transformations, calculated columns, etc.
- Formula Bar**: A bar at the top right labeled '(this is "M" code)' where users can enter M code.
- Table Name & Properties**: A panel showing the current table's name ('Customer Lookup') and properties.
- Applied Steps**: A panel listing the steps taken in the query, such as 'Source', 'Transformed Headers', 'Changed Type', etc.

Arrows point from the labels to their respective parts in the interface.

QUERY EDITING TOOLS



isMarried	No. of children	Is Parent?
Yes/No	0, 1, 2, 3	YES
M/S/U		NO

Extra Spaces.
[Row Context]

New Column - Adding Column by Power Query Editor.
Calculated Column - DAX [Data Analysis Expression].

1. CountIf, CountIfs, SUMIFS, SUMIF, AVERAGE, AVERAGEIFS, etc.
2. Formulae : Where you are using operator,

Row Context ...

Adding a new Columns : Function calling:
IF(No. Of Children > 0, IsParent ? Yes : No)

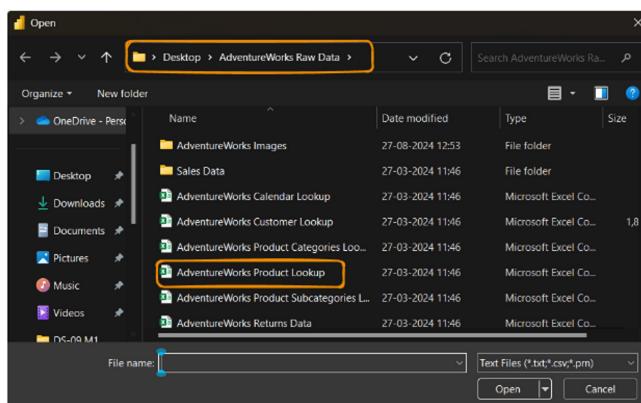
AdventureWorks Product Lookup.csv

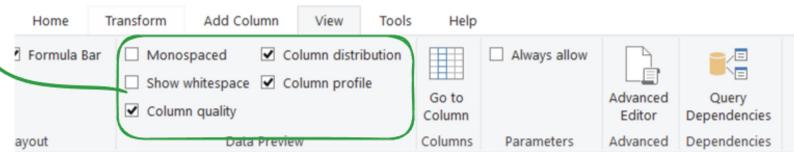
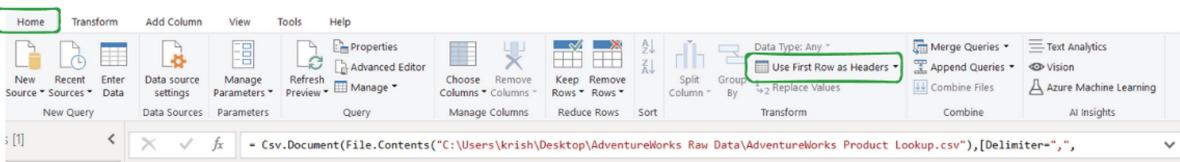
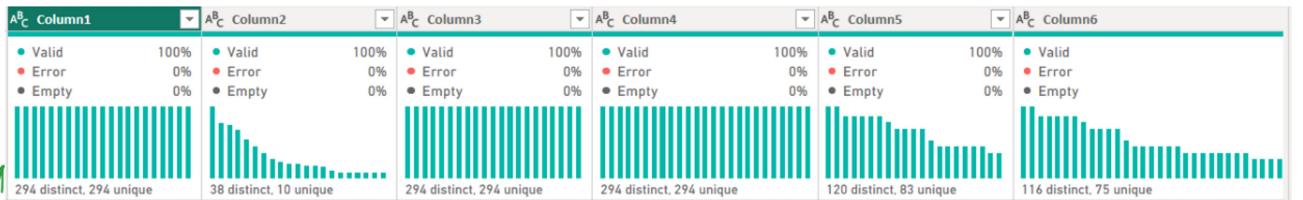
We are now trying to transform Product Lookup

File Origin Delimiter Data Type Detection
1252: Western European (Windows) Comma Based on first 200 rows

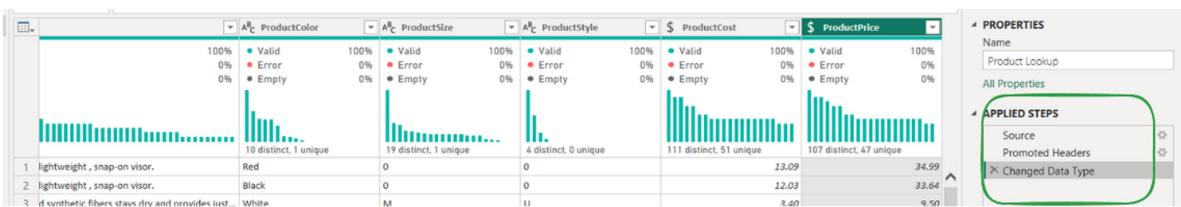
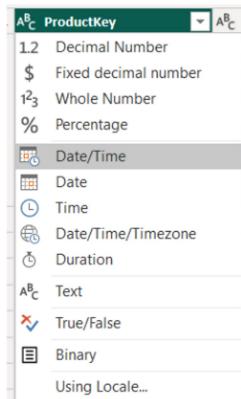
Column1	Column2	Column3	Column4	Column5	Column6
ProductKey	ProductSubcategoryKey	ProductSKU	ProductName	ModelName	ProductDescription
214	31	HL-US09-R	Sport-100 Helmet, Red	Sport-100	Universal fit, well-vented, lightweight , snap-on viso
215	31	HL-US09	Sport-100 Helmet, Black	Sport-100	Universal fit, well-vented, lightweight , snap-on viso
218	23	SO-B909-M	Mountain Bike Socks, M	Mountain Bike Socks	Combination of natural and synthetic fibers stays dry
219	23	SO-B909-L	Mountain Bike Socks, L	Mountain Bike Socks	Combination of natural and synthetic fibers stays dry
220	31	HL-US09-B	Sport-100 Helmet, Blue	Sport-100	Universal fit, well-vented, lightweight , snap-on viso
223	19	CA-1094	AWC Logo Cap	Cycling Cap	Traditional style with a flip-up brim; one-size fits all.
226	21	LI-0192-S	Long-Sleeve Logo Jersey, S	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling jersey
229	21	LI-0192-M	Long-Sleeve Logo Jersey, M	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling jersey
232	21	LI-0192-L	Long-Sleeve Logo Jersey, L	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling jersey
235	21	LI-0192-X	Long-Sleeve Logo Jersey, XL	Long-Sleeve Logo Jersey	Unisex long-sleeve AWC logo microfiber cycling jersey
238	14	FR-R92R-62	HL Road Frame - Red, 62	HL Road Frame	Our lightest and best quality aluminum frame made
241	14	FR-R92R-44	HL Road Frame - Red, 44	HL Road Frame	Our lightest and best quality aluminum frame made
244	14	FR-R92R-48	HL Road Frame - Red, 48	HL Road Frame	Our lightest and best quality aluminum frame made
247	14	FR-R92R-52	HL Road Frame - Red, 52	HL Road Frame	Our lightest and best quality aluminum frame made
250	14	FR-R92R-56	HL Road Frame - Red, 56	HL Road Frame	Our lightest and best quality aluminum frame made
253	14	FR-R38B-58	LL Road Frame - Black, 58	LL Road Frame	The LL Frame provides a safe comfortable ride, while
256	14	FR-R38B-60	LL Road Frame - Black, 60	LL Road Frame	The LL Frame provides a safe comfortable ride, while
259	14	FR-R38B-62	LL Road Frame - Black, 62	LL Road Frame	The LL Frame provides a safe comfortable ride, while
262	14	FR-R38B-44	LL Road Frame - Red, 44	LL Road Frame	The LL Frame provides a safe comfortable ride, while

Extract Table Using Examples Load Transform Data Cancel





"111.11" * 2 = "111.11*2"
111 * 2 = 222



Assignment : Transforming Data

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

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Added SKU Result
- Renamed Columns
- Replaced Value

Text Before Delimiter

Enter the delimiter that marks the end of what you would like to extract.

Delimiter: -

Advanced options

Scan for the delimiter: From the start of the input

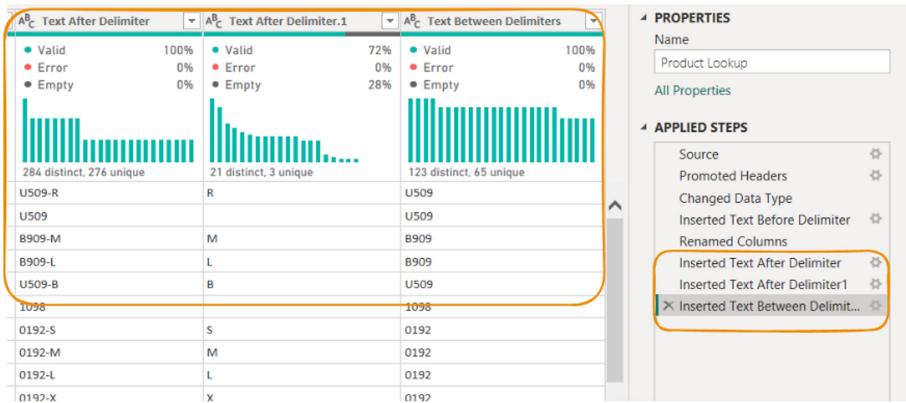
Number of delimiters to skip: 1

skipping the first occurrence of "-"

OK Cancel

delimiter - separator

[HL]-[U509]-[R]



DL @ (3C)- 1246

Text Between Delimiters

Enter the delimiters that mark the beginning and end of the text.

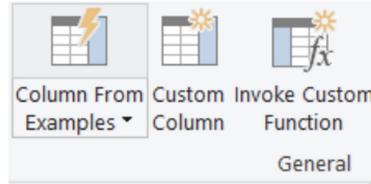
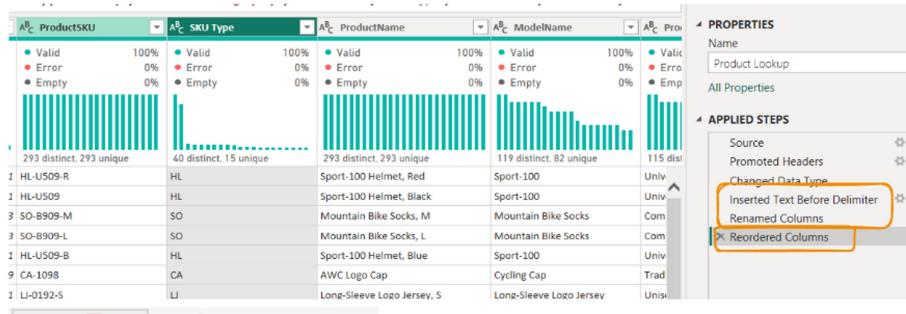
Start delimiter

@

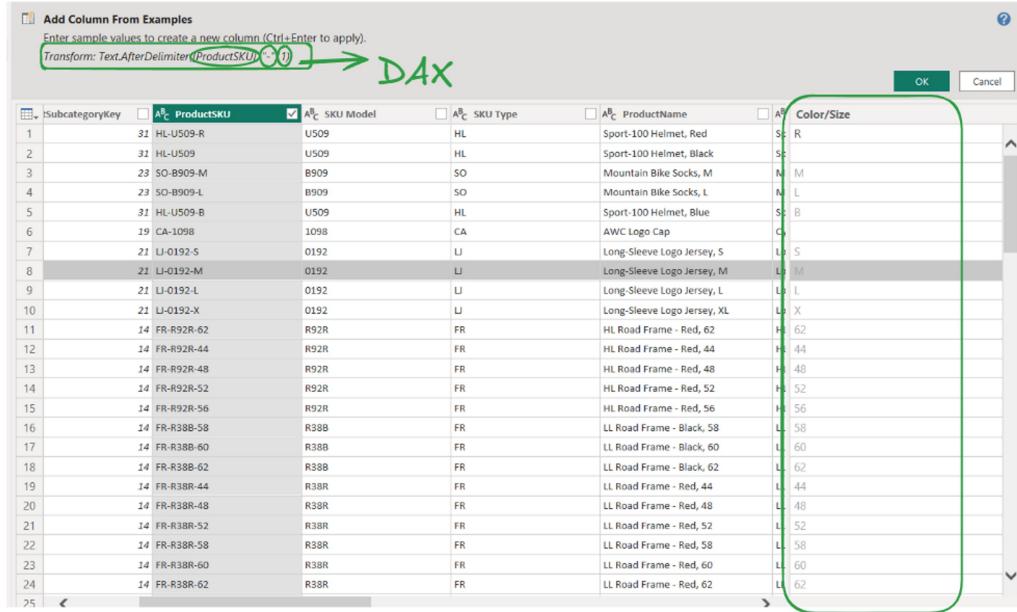
End delimiter

-

Advanced options



20 Hour Rule



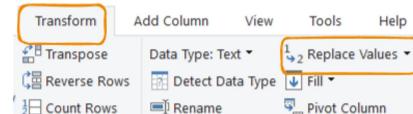
Replace Values

Replace one value with another in the selected columns.

Value To Find
0

Replace With
NA

[Advanced options](#)



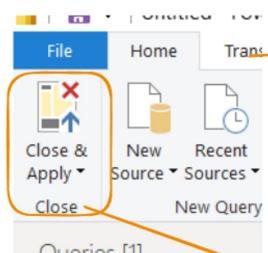
[OK](#) [Cancel](#)

ABC ProductStyle	\$ ProductCost	\$ ProductPrice
Valid 100%	Valid 100%	Valid 100%
Error 0%	Error 0%	Error 0%
Empty 0%	Empty 0%	Empty 0%
NA	13.09	107 distinct, 47 unique
NA	12.03	
U	3.40	
U	3.40	
NA	12.03	
U	5.71	

4 distinct, 0 unique

Properties pane:

- Name: Product Lookup
- All Properties
- Applied Steps:
 - Source
 - Promoted Headers
 - Changed Data Type
 - Inserted Text Before Delimiter
 - Renamed Columns
 - Reordered Columns
 - Replaced Value



→ Power Query Editor : Cleaning Process/ Transformation

Queries [1]

Data

... >

Search

> Product Lookup

Updates the clean data to front End.

Front End Changes : Once it is loaded from BackEnd [P.Q.E]

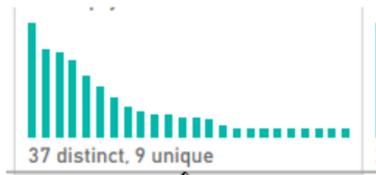
ProductKey	ProductSubcategoryKey	ProductSKU	SKU Type	ProductName	ModelName	ProductDescription
310	2	BK-R93R-62	BK	Road-150 Red, 62	Road-150	This bike is ridden by race winners. Developed with the Adventure Works Cycles professional race team. It
311	2	BK-R93R-44	BK	Road-150 Red, 44	Road-150	This bike is ridden by race winners. Developed with the Adventure Works Cycles professional race team. It
312	2	BK-R93R-48	BK	Road-150 Red, 48	Road-150	This bike is ridden by race winners. Developed with the Adventure Works Cycles professional race team. It
313	2	BK-R93R-52	BK	Road-150 Red, 52	Road-150	This bike is ridden by race winners. Developed with the Adventure Works Cycles professional race team. It
314	2	BK-R93R-56	BK	Road-150 Red, 56	Road-150	This bike is ridden by race winners. Developed with the Adventure Works Cycles professional race team. It
315	2	BK-R68R-58	BK	Road-450 Red, 58	Road-450	A true multi-sport bike that offers streamlined riding and a revolutionary design. Aerodynamic design lets
316	2	BK-R68R-60	BK	Road-450 Red, 60	Road-450	A true multi-sport bike that offers streamlined riding and a revolutionary design. Aerodynamic design lets
317	2	BK-R68R-44	BK	Road-450 Red, 44	Road-450	A true multi-sport bike that offers streamlined riding and a revolutionary design. Aerodynamic design lets
318	2	BK-R68R-48	BK	Road-450 Red, 48	Road-450	A true multi-sport bike that offers streamlined riding and a revolutionary design. Aerodynamic design lets

The screenshot shows the 'Product Lookup' table in Power BI Desktop. The table has columns: ModelName, ProductColor, ProductCost, ProductDescription, ProductKey, ProductName, ProductPrice, ProductSize, and ProductSKU. A 'Collapse ^' button is at the bottom left. Below the table, there's a navigation bar with 'All tables' and a '+' icon.

Don't forget to save the file

Title

214		31	HL-U509-R	HL	S
215		31	HL-U509	HL	S
218		23	SO-B909-M	SO	M
219		23	SO-B909-L	SO	M
220		31	HL-U509-B	HL	S
223		19	CA-1098	CA	A
226		21	LJ-0192-S	LJ	L
229		21	LJ-0192-M	LJ	L
232		21	LJ-0192-L	LJ	L
235		21	LJ-0192-X	LJ	L
238		14	FR-R92R-62	FR	F
241		14	FR-R92R-44	FR	F
244		14	FR-R92R-48	FR	F
247		14	FR-R92R-52	FR	F



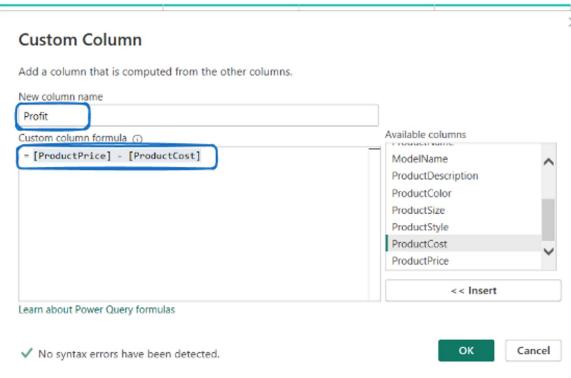
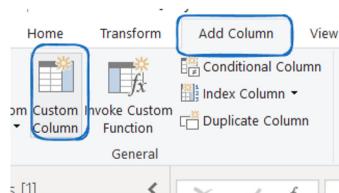
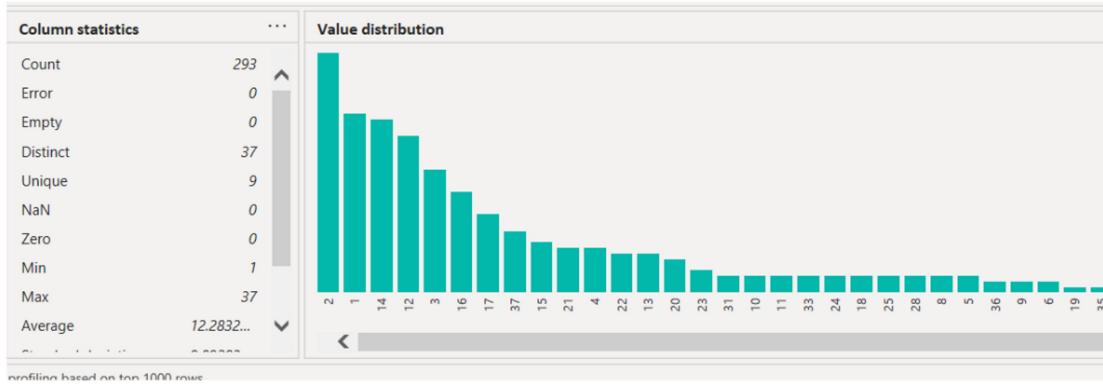
unique : 9 products are uniquely present : 1 time available only

distinct : 37 items we counted as one. but they can be in multiples.

121K - 5
231S - 3
342W - 17
T432 - 25

distinct : if it is present 2 or more . Counting it as 1. but dropping it into distinct category.

Column Profile : Providing the stats of the selected column.



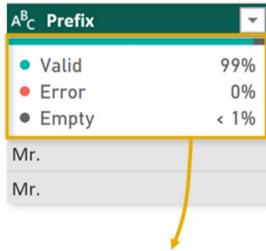
Numerical Data.

Column From Example is more useful : when we are dealing with textual data.

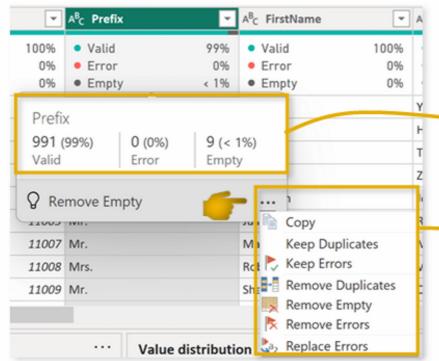
ProductColor	ProductSize	ProductStyle	Random Column
Red	O	NA	Red - O
Black	O	NA	Black - O
White	M	U	White - M
White	L	U	White - L
Blue	O	NA	Blue - O
Multi	O	U	Multi - O
Multi	S	U	Multi - S
Multi	M	U	Multi - M
Multi	L	U	Multi - L
Multi	XL	U	Multi - XL

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



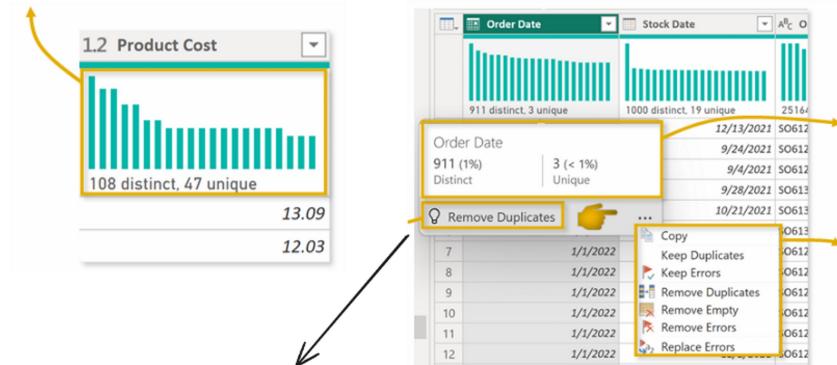
Hover over the column quality box to see the number of records in each category

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

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

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

COLUMN PROFILE

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

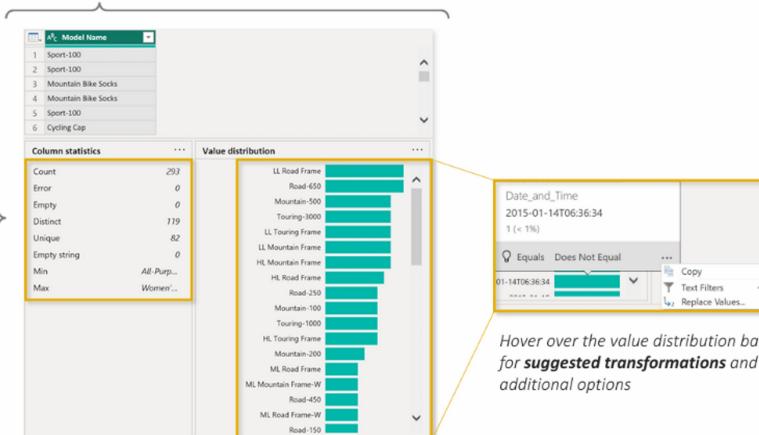
Column statistics provide more detailed profiling metrics, including:

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