

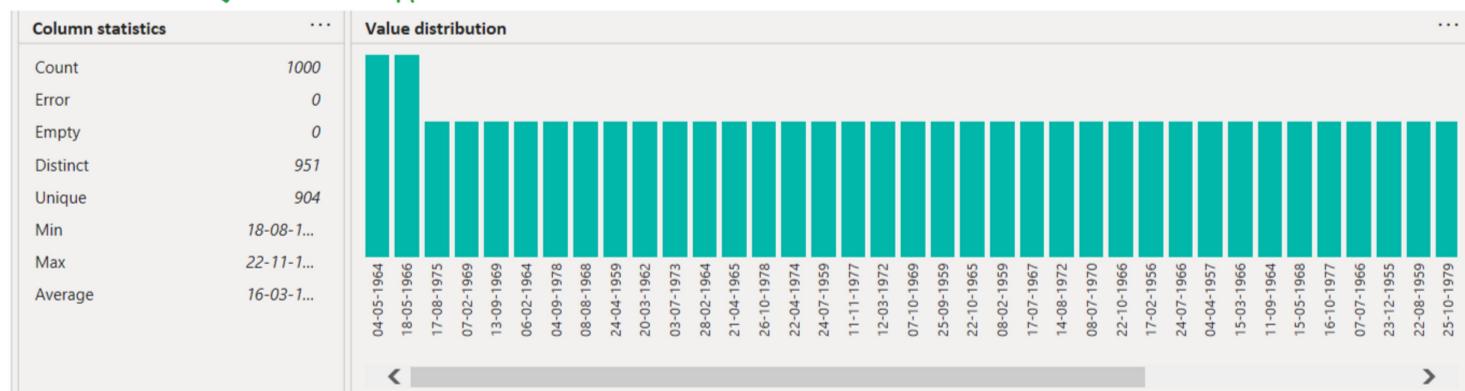
## Power BI - Data Transformation - p1 - Lecture 2

### Power BI - Data Transformation - p1 - Lecture 2

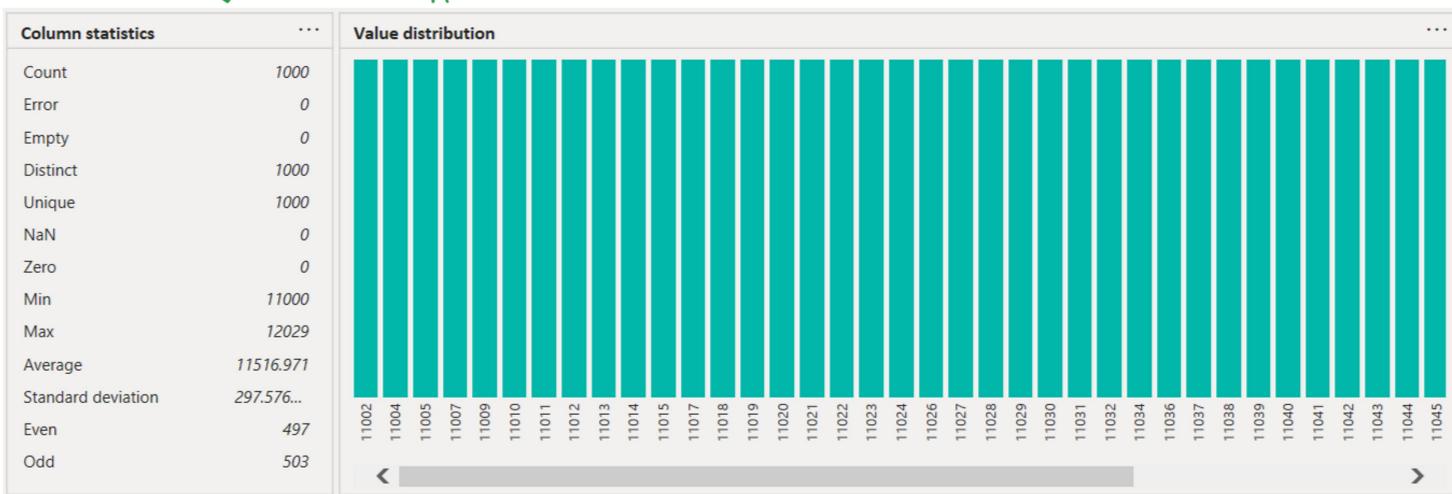
#### Column Profiling - "String Type"



#### Column Profiling - "Date Type"



#### Column Profiling - "Numeric Type"





Distinct - 1 or more Value

Unique - exactly 1 value

3 distinct Values

1 unique

[Volleyball, Badminton,  
Cricket]

[Badminton]

Volleyball - 2

Badminton - 1

Cricket - 2

## 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)

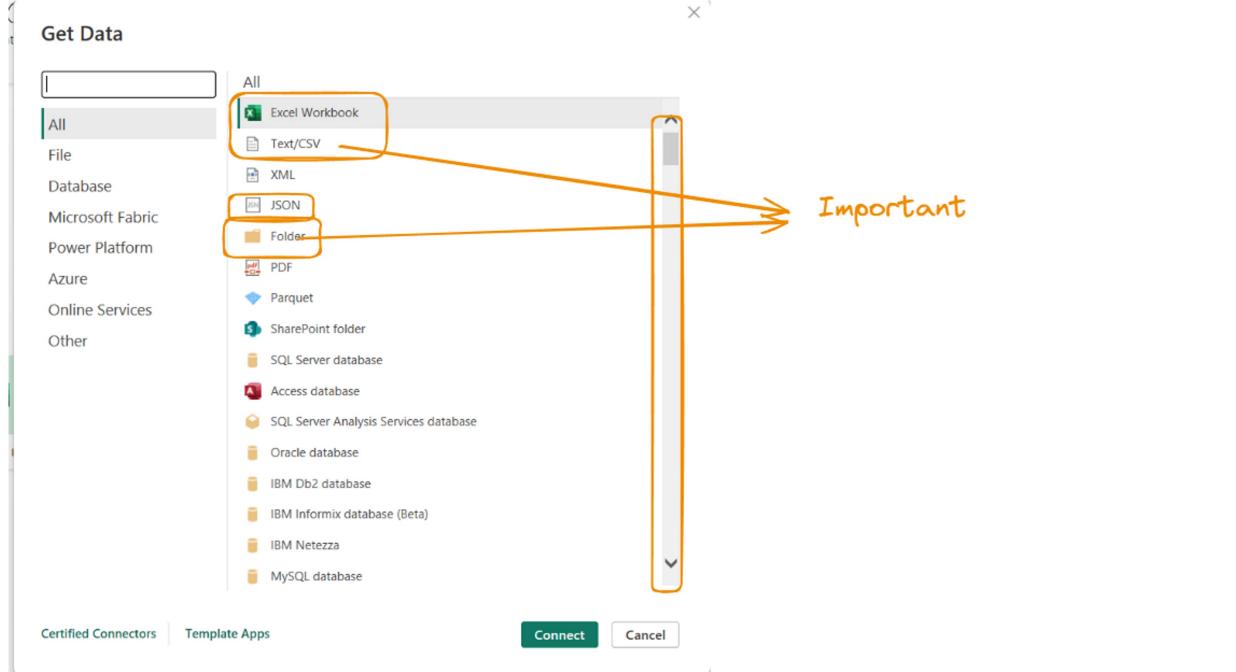
### BACK-END

- 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



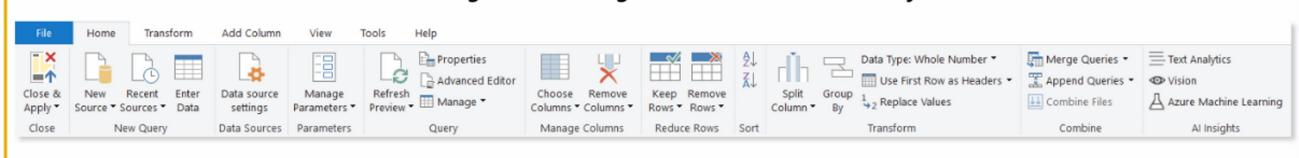
## POWER QUERY EDITOR

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. Below the ribbon is a toolbar with icons for Paste, Copy, Format painter, and others. The main area is divided into several sections:

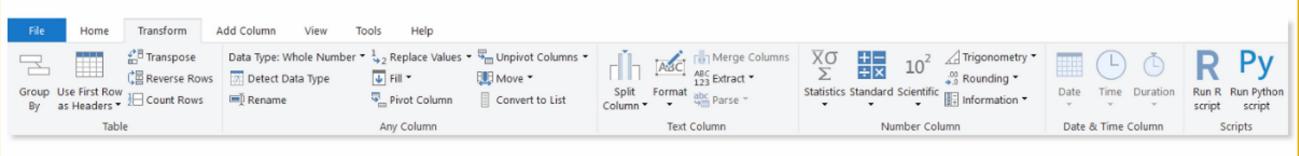
- Queries Pane (list of all queries):** Located on the left, it shows a tree view of queries: Transform file from Sales Data, Other Queries (0), Calendar Lookup, Customer Lookup, and so on.
- Table Preview:** Shows a preview of the data with columns: Customer Key, Prefix, First Name, Last Name, Birth Date, Marital Status, and Gender.
- Formula Bar (this is "M" code):** Displays the formula: =Table1|RemoveColumns({"Customer"}).
- Query Editing Tools (Table transformations, calculated columns, etc.):** A ribbon-like bar above the preview with various transformation icons.
- Query Settings:** A floating pane on the right containing:
  - Properties:** Source, Primary Key, Column Type, Column Rename, Capitalized Each Word, Inserted Text Before Column, Inserted Text Before Delimiter, Renamed Columns, Inserted Text Between Delimiter, Renamed Column1, Renamed Column2, Capitalized Each Word1, Inserted Year, Renamed Column3, Added Conditional Column, Shared Rows.
  - Applied Steps:** A list of steps applied to the query, such as Remove Columns, Rename, and Capitalize Each Word.
- Table Name & Properties:** A section pointing to the properties in the Query Settings pane.
- Applied Steps (like a macro):** A section pointing to the 'Applied Steps' list in the Query Settings pane.

## QUERY EDITING TOOLS

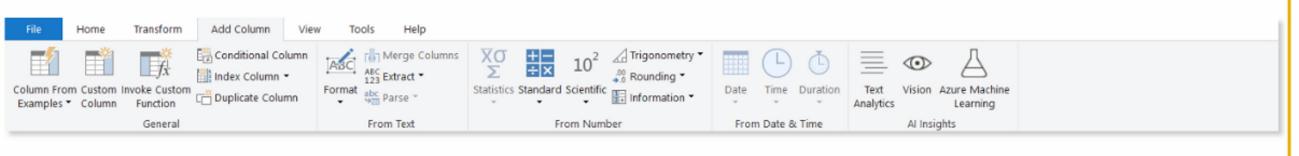
The **HOME** tab includes **general settings** and **common table transformation tools**



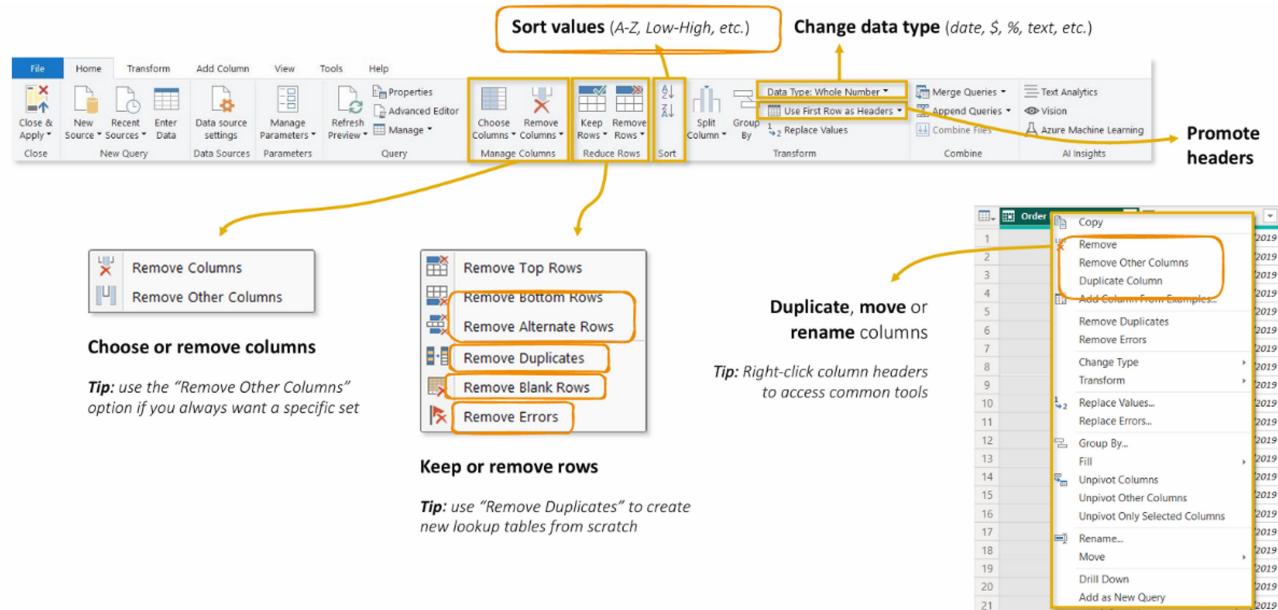
The **TRANSFORM** tab includes tools to **modify existing columns** (splitting/grouping, transposing, extracting text, etc.)



The **ADD COLUMN** tools **create new columns** (based on conditional rules, text operations, calculations, dates, etc.)



## BASIC TABLE TRANSFORMATIONS



## 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

The screenshot shows the Power Query Editor interface. On the right, a sidebar titled 'APPLIED STEPS' lists the following steps:

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

Below the sidebar, there is a preview of a SKU string: "HL-U509-R". A green box highlights the part "U509".

The main area is titled "Text Before Delimiter". It contains the following fields:

- "Delimiter": A text input field containing a hyphen (-).
- "Advanced options": A section with "Scan for the delimiter" (set to "From the start of the input") and "Number of delimiters to skip" (set to 1). A green arrow points from the "Number of delimiters to skip" field to the text "skipping the first occurrence of '-'".

At the bottom right are "OK" and "Cancel" buttons.

column 1

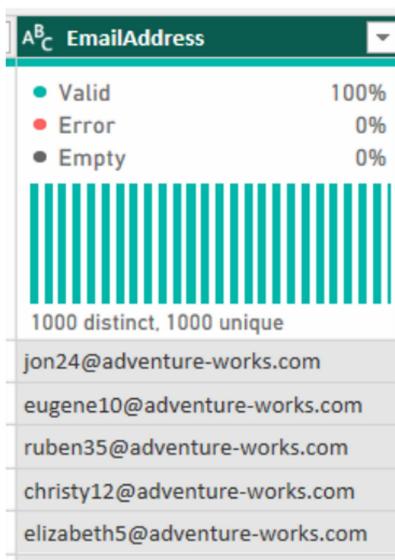
Delhi, India
Berlin, Germany
Mumbai, India
Toronto, Canada

City

Delhi
Berlin
Mumbai
Toronto

Country

India
Germany
India
Canada



Username

jon24
-------

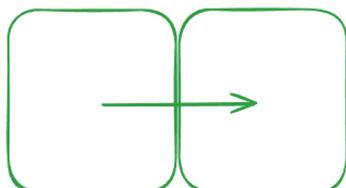
Domain Name

adventure-works
-----------------

"@" - Delimeter/Separator

Merging Queries

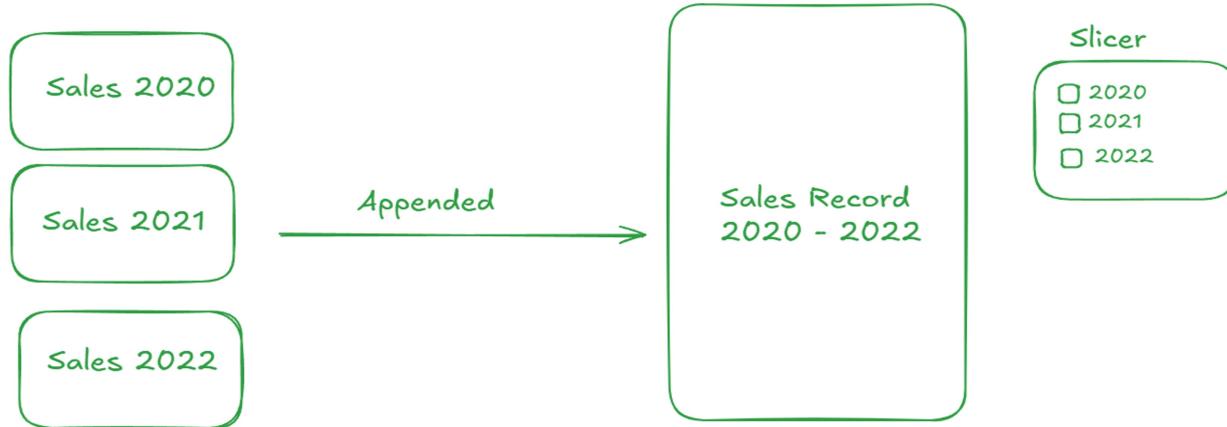
→ Joins in SQL



Horizontal Direction

[This can only be possible if there should be having some relationship exists.]

**Appending Queries** → Joining different **identical** Table in Vertical Direction.



## 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.

## 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

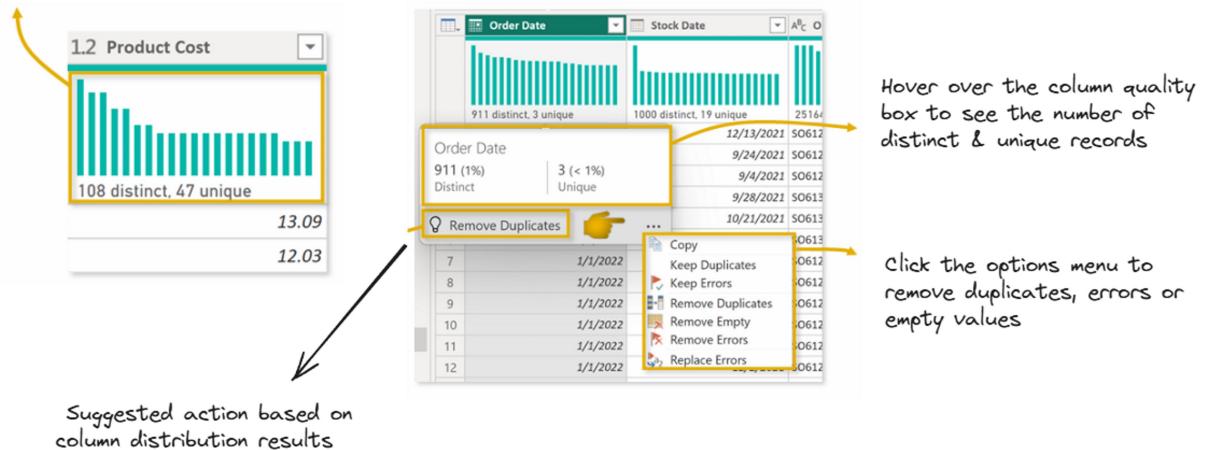
Hover over the column quality box to see the number of records in each category

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



## 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

