SUMMARY: INTRODUCTION TO POWER QUERY

SESSION OVERVIEW:

By the end of this session, the students will be able to:
- Understand the importance of Power Query in Excel.
- Understand different ways by which data can be imported.
- Understand the interface of the Power Query Editor in Excel.
- Understand different types of transformation using Power Query.

KEY TOPICS AND EXAMPLES:

NOTE: You can explore different Get Data features in Excel. The three types mentioned below are used frequently in the industry if we are working with Excel.

Understanding the importance of Power Query in Excel:

1. **What is Power Query:**

    Power Query is a data transformation and manipulation tool developed by Microsoft, available within Microsoft Excel, Power BI, and other Microsoft products. Its primary function is to help users import, transform, and cleanse data from various sources to prepare it for analysis and reporting.

    NOTE:
    In Power Query, the base data—meaning the original data source itself—does not change. Power Query works on the principle of loading data into its environment and then allowing you to perform various transformations on this data within Power Query. These transformations include cleaning, shaping, merging, appending, and more, but they all occur within the Power Query editor and do not affect the original data source.

2. **Importance of Power Query:**

    a. **Data Integration:** Power Query allows users to easily connect to and import data from a wide range of sources, including databases, Excel files, text files, websites, and more. This capability enables analysts to integrate data from disparate sources into a single cohesive dataset for analysis.

    b. **Data Cleansing and Transformation:** Power Query provides a user-friendly interface for cleaning and transforming raw data into a usable format. Users can perform a variety of data cleansing operations, such as removing duplicates, filtering

rows, splitting columns, and changing data types. This ensures data quality and consistency, which are essential for accurate analysis.

    c.   **Data Shaping and Modeling:** With Power Query, users can reshape and model data to meet their specific analysis requirements. This includes tasks such as pivoting and unpivoting data, aggregating data, and merging or appending multiple datasets together. These capabilities empower analysts to structure data in a way that best supports their analytical objectives.

    d.   **Data Exploration and Discovery:** By facilitating data exploration and discovery, Power Query helps users uncover insights and patterns hidden within their data. Analysts can easily manipulate and visualize data to identify trends, outliers, and correlations, leading to more informed decision-making.

Understanding different ways by which data can be imported:

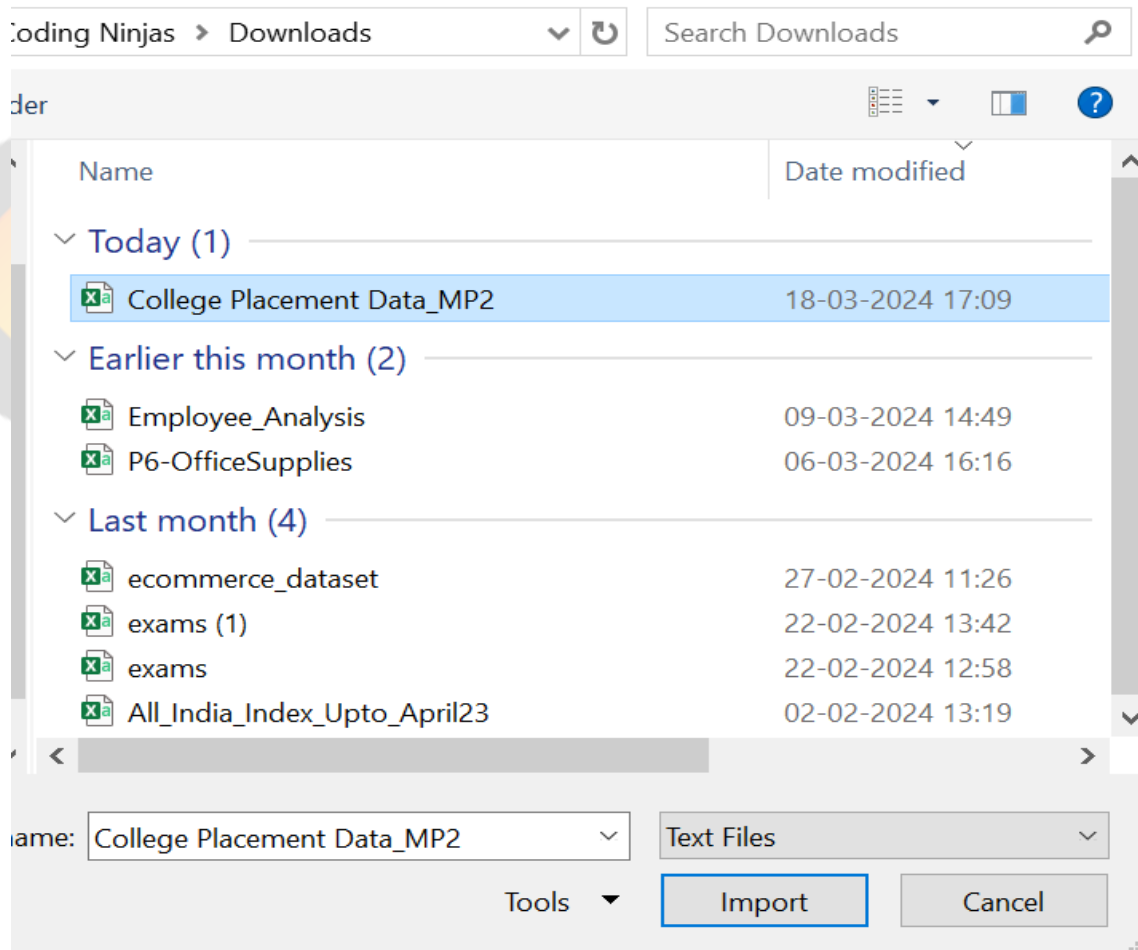   **1.**   **Get Data from TXT/CSV:**

- Go to Data Tab
- Get & Transform Data Group
- Get Data
- Text/CSV file
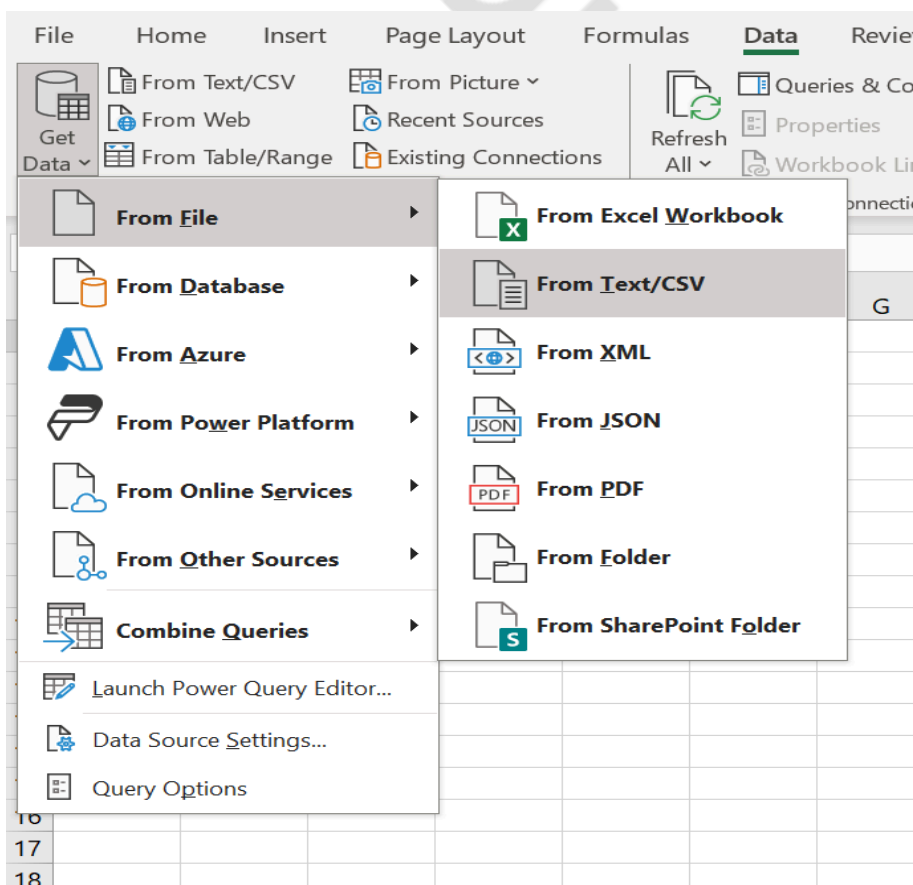
- Select file from the downloaded files that you have.

(**NOTE**: This is just a sample example, you can try to import whatever downloaded file you have in your system. It is not compulsory to use the college placement data in your system.)

Name	Date modified
∨ **Today (1)**	
College Placement Data_MP2	18-03-2024 17:09
∨ **Earlier this month (2)**	
Employee_Analysis	09-03-2024 14:49
P6-OfficeSupplies	06-03-2024 16:16
∨ **Last month (4)**	
ecommerce_dataset	27-02-2024 11:26
exams (1)	22-02-2024 13:42
exams	22-02-2024 12:58
All_India_Index_Upto_April23	02-02-2024 13:19

Name: College Placement Data_MP2   Text Files

Tools ▾   Import   Cancel

- Click Import.

College Placement Data_MP2.csv

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

s_id	Name	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	specialisation	mba_p	wo
1	Rick Hansen	M	67	Others	91	Others	Commerce	58	Sci&Tech	Mkt&HR	58.8	No
2	Justin Ritter	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Mkt&Fin	66.28	Yes
3	Craig Reiter	M	65	Central	68	Central	Arts	64	Comm&Mgmt	Mkt&Fin	57.8	No
4	Katherine Murray	M	56	Central	52	Central	Science	52	Sci&Tech	Mkt&HR	59.43	No
5	Rick Hansen	M	85.8	Central	73.6	Central	Commerce	73.3	Comm&Mgmt	Mkt&Fin	55.5	No
6	Jim Mitchum	M	55	Others	49.8	Others	Science	67.25	Sci&Tech	Mkt&Fin	51.58	Yes
7	Toby Swindell	F	46	Others	49.2	Others	Commerce	79	Comm&Mgmt	Mkt&Fin	53.29	No
8	Mick Brown	M	82	Central	64	Central	Science	66	Sci&Tech	Mkt&Fin	62.14	Yes
9	Jane Waco	M	73	Central	79	Central	Commerce	72	Comm&Mgmt	Mkt&Fin	61.29	No
10	Joseph Holt	M	58	Central	70	Central	Commerce	61	Comm&Mgmt	Mkt&Fin	52.21	No
11	Greg Maxwell	M	58	Central	61	Central	Commerce	60	Comm&Mgmt	Mkt&HR	60.85	Yes
12	Anthony Jacobs	M	69.6	Central	68.4	Central	Commerce	78.3	Comm&Mgmt	Mkt&Fin	63.7	Yes
13	Magdelene Morse	F	47	Central	55	Others	Science	65	Comm&Mgmt	Mkt&HR	65.04	No
14	Vicky Freymann	F	77	Central	87	Central	Commerce	59	Comm&Mgmt	Mkt&Fin	68.63	No
15	Peter Fuller	M	62	Central	47	Central	Commerce	50	Comm&Mgmt	Mkt&HR	54.96	No
16	Ben Peterman	F	65	Central	75	Central	Commerce	69	Comm&Mgmt	Mkt&Fin	64.66	Yes
17	Thomas Boland	M	63	Central	66.2	Central	Commerce	65.6	Comm&Mgmt	Mkt&Fin	62.54	Yes
18	Patrick Jones	F	55	Central	67	Central	Commerce	64	Comm&Mgmt	Mkt&Fin	67.28	No
19	Jim Sink	F	63	Central	66	Central	Commerce	64	Comm&Mgmt	Mkt&HR	64.08	No
20	Ritsa Hightower	M	60	Others	67	Others	Arts	70	Comm&Mgmt	Mkt&Fin	77.89	Yes

Load ▾ | **Transform Data** | **Cancel**

- After it gets imported we have to click load and your desired file will get loaded into your Power Query Editor as well as in your Excel.

2. **Get Data from Web:**
   - Open the web page which needs to be imported into Excel.
   - Copy the URL
     (**NOTE**: You can use any of your desired URLs available on the internet to fetch the data. It is not compulsory to use the URL visible in the image below.)

← → C ⌂ 🔒 contextures.com/xlsampledata01.html#winter

⚡ StackBlitz | Instant...   🐙 Your Repositories   🔵 SQL Search   🔵 C++ Search   🔵 Search | Sales Navi...
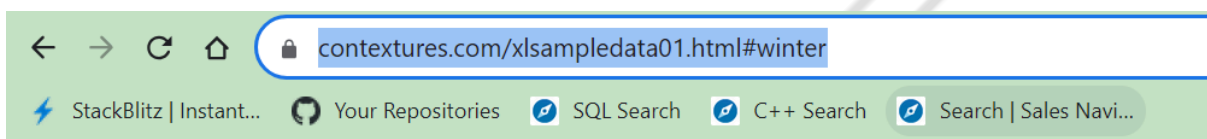
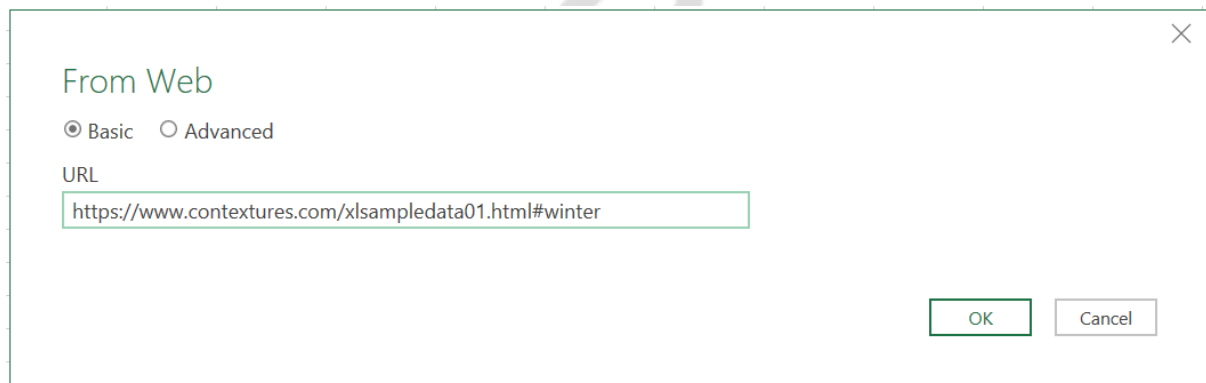## 5) Hockey Players Data

This sample file has data from the 2018 Olympic Hockey teams, fro the USA.

- There are 15 columns of data, including 4 columns with calculatio
- There are 96 rows of data in the hockey player table.

- Open Excel ->Data Tab->Get Data-> From other Sources-> From Web
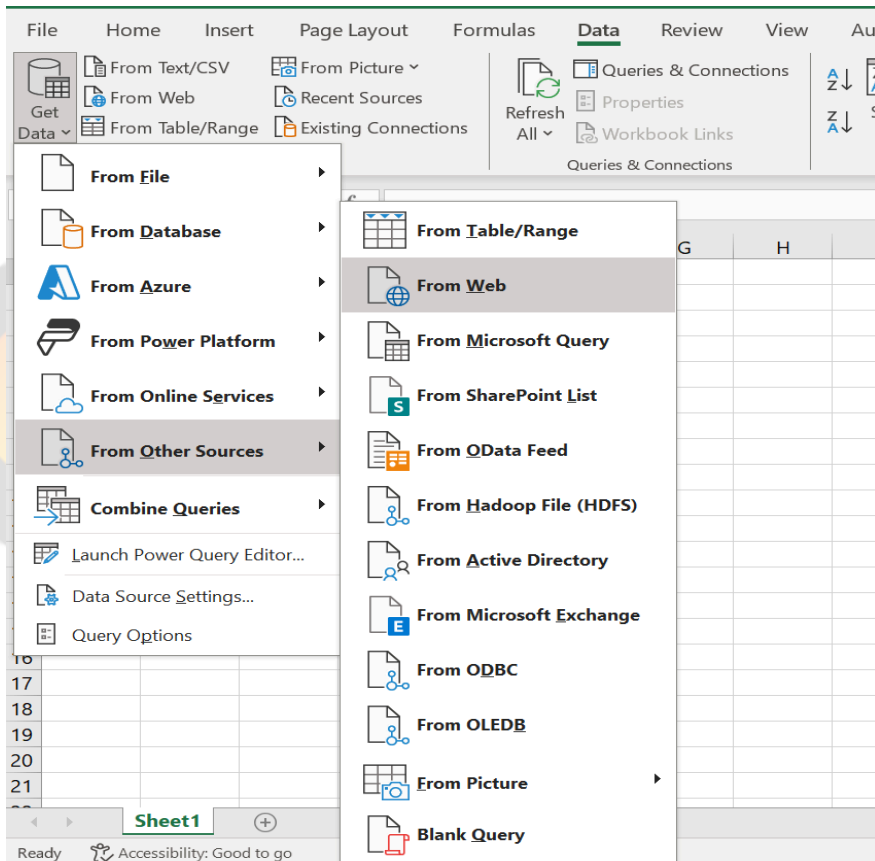
- Paste URL->Ok

- Excel will make a connection with the webpage.
- After the connection established, one will get a new window (A list of the tables will be available, by clicking on any table, one can see the preview of the table's data on the right side)

Navigator

Table View   Web View

Sample Data

OrderDate	Region	Rep	Item	Units	UnitCost	T
1/6/2020	East	Jones	Pencil	95	1.9	
1/23/2020	Central	Kivell	Binder	50	19.9	
2/9/2020	Central	Jardine	Pencil	36	4.9	
2/26/2020	Central	Gill	Pen	27	19.9	
3/15/2020	West	Sorvino	Pencil	56	2.9	
4/1/2020	East	Jones	Binder	60	4.9	
4/18/2020	Central	Andrews	Pencil	75	1.9	
5/5/2020	Central	Jardine	Pencil	90	4.9	
5/22/2020	West	Thompson	Pencil	32	1.9	
6/8/2020	East	Jones	Binder	60	8.9	
6/25/2020	Central	Morgan	Pencil	90	4.9	
7/12/2020	East	Howard	Binder	29	1.9	
7/29/2020	East	Parent	Binder	81	19.9	
8/15/2020	East	Jones	Pencil	35	4.9	
9/1/2020	Central	Smith	Desk	2	12	
9/18/2020	East	Jones	Pen Set	16	15.9	
10/5/2020	Central	Morgan	Binder	28	8.9	
10/22/2020	East	Jones	Pen	64	8.9	
11/8/2020	East	Parent	Pen	15	19.9	
11/25/2020	Central	Kivell	Pen Set	96	4.9	

Display Options in left panel:
- https://www.contextures.com/xlsampledata01....
  - Document
  - **Sample Data**

Load | Transform Data | Cancel

- Select table ->close & Load. After clicking load a dialogue box will appear which will guide you to choose the view that you want in your workbook. Select the location where you want to put the data and click Ok.

Import Data

Select how you want to view this data in your workbook.
- ● Table
- ○ PivotTable Report
- ○ PivotChart
- ○ Only Create Connection

Where do you want to put the data?
- ○ Existing worksheet:
  - =A1
- ● New worksheet

☐ Add this data to the Data Model

Properties... | OK | Cancel

- Once you click Ok, your desired table will be created a new worksheet/Existing worksheet.

OrderDate	Region	Rep	Item	Units	UnitCost	Total
1/6/2020	East	Jones	Pencil	95	1.99	189.05
1/23/2020	Central	Kivell	Binder	50	19.99	999.5
2/9/2020	Central	Jardine	Pencil	36	4.99	179.64
2/26/2020	Central	Gill	Pen	27	19.99	539.73
3/15/2020	West	Sorvino	Pencil	56	2.99	167.44
4/1/2020	East	Jones	Binder	60	4.99	299.4
4/18/2020	Central	Andrews	Pencil	75	1.99	149.25
5/5/2020	Central	Jardine	Pencil	90	4.99	449.1
5/22/2020	West	Thompson	Pencil	32	1.99	63.68
6/8/2020	East	Jones	Binder	60	8.99	539.4
6/25/2020	Central	Morgan	Pencil	90	4.99	449.1
7/12/2020	East	Howard	Binder	29	1.99	57.71
7/29/2020	East	Parent	Binder	81	19.99	1619.19
8/15/2020	East	Jones	Pencil	35	4.99	174.65
9/1/2020	Central	Smith	Desk	2	125	250
9/18/2020	East	Jones	Pen Set	16	15.99	255.84
10/5/2020	Central	Morgan	Binder	28	8.99	251.72
10/22/2020	East	Jones	Pen	64	8.99	575.36

3. **Get data from table/range**:
   - Select the range/table from where you want to get the data.
   - Go to the data tab and click get data.
   - Go to From other sources and click From Table/Range.
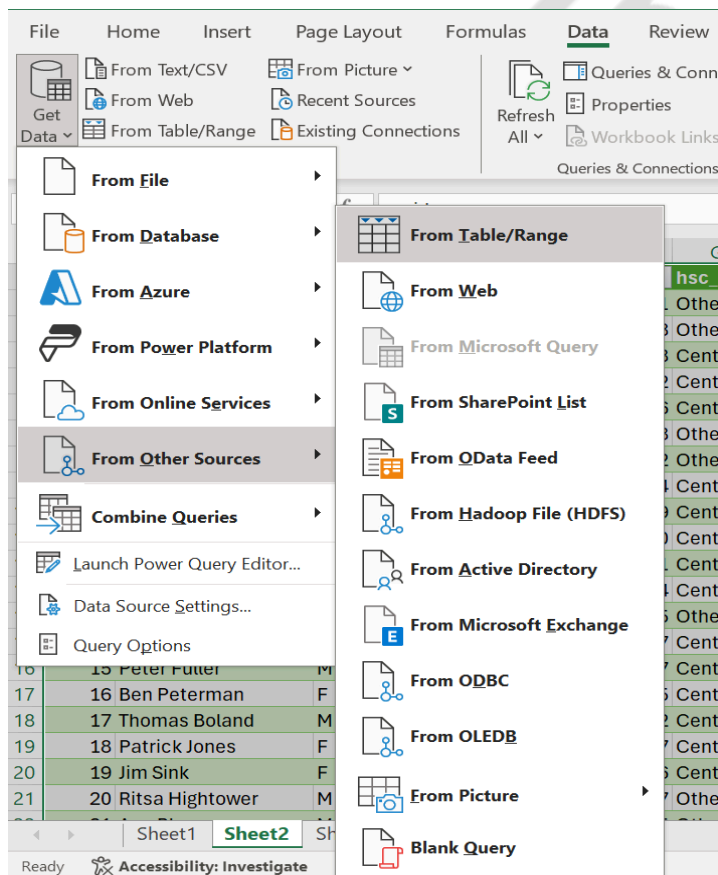
- After clicking From Table/Range, the data will automatically open in the Power Query Editor.
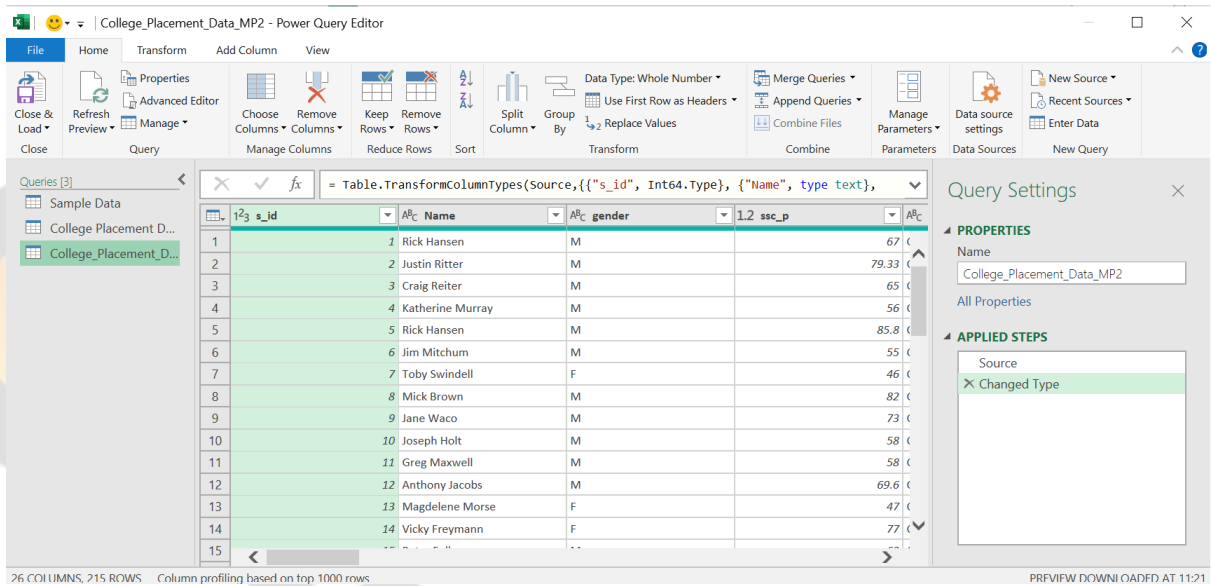

Understanding the interface of the Power Query Editor:

1. **Query Editor Window:** This is the primary workspace where you'll perform data transformation tasks. It consists of several tabs and ribbons, each containing a set of tools and functionalities for working with your data.
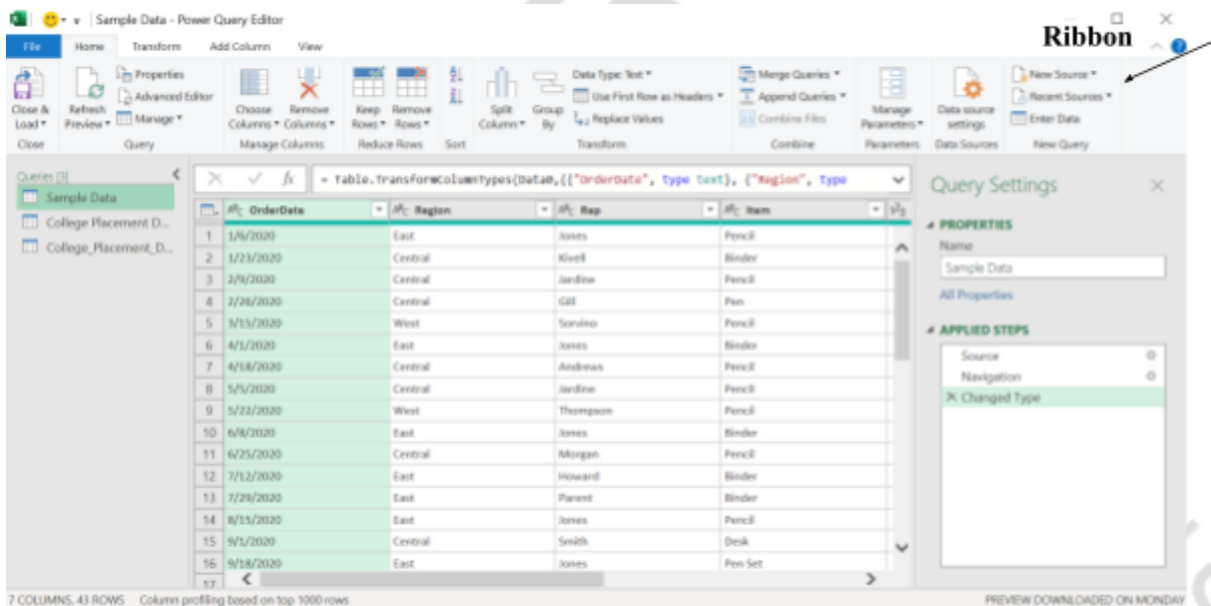
This is how the Power Query Editor looks, which is more or less similar to Excel. The arrow indicates the ribbon above which there are different tabs like Home, Transform, Add Column and view.

2. **Query Settings Pane:** Located on the right side of the Query Editor window, the Query Settings pane provides options and settings specific to the selected query. It includes sections such as Applied Steps, Query Properties, and Query Dependencies.
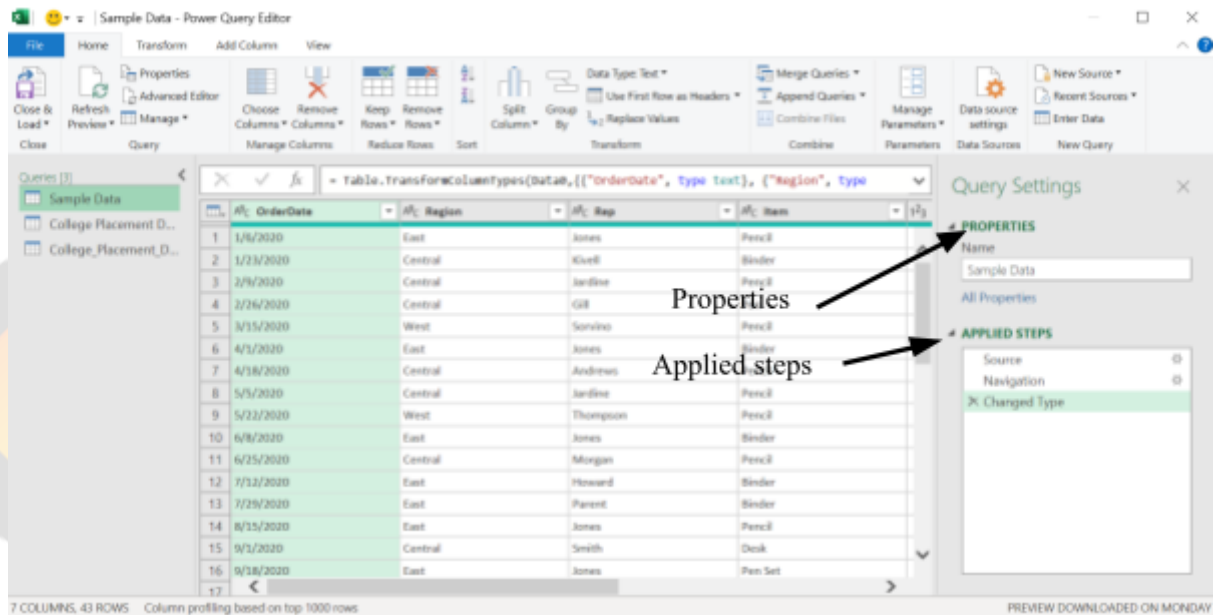
The properties settings help us to change the name of the file in Power Query.

Query properties: This section displays metadata about the selected query, such as its name, description, and source. You can use this section to edit the query's properties or add descriptive information to help document your analysis process.

Applied steps: This section lists all the transformation steps that have been applied to the data. Each step represents a specific action performed on the data, such as renaming columns, filtering rows, or transforming data types. You can review, edit, or delete individual steps to modify the data transformation process.

3. **Formula bar:** At the top of the Query Editor window, you'll find the formula bar.

Understand different types of transformation using Power Query:

Reference dataset

1. **Close and load:**
   - When you click "Close & Load," all the transformations, edits, and data cleaning operations you performed in the Power Query Editor are finalized. This step applies all the changes to your original data set.
   - You also have the option to "Close & Load To…" which allows you to choose not only to load the data into a worksheet or the data model but also to create a connection only without immediately loading the data into the Excel interface.

2. **Manage columns:(Dataset)**

   a. **Choose columns:**

   - Look for the "Manage Columns" group, then click on the "Choose Columns" button. This action will open the "Choose Columns" dialog box.
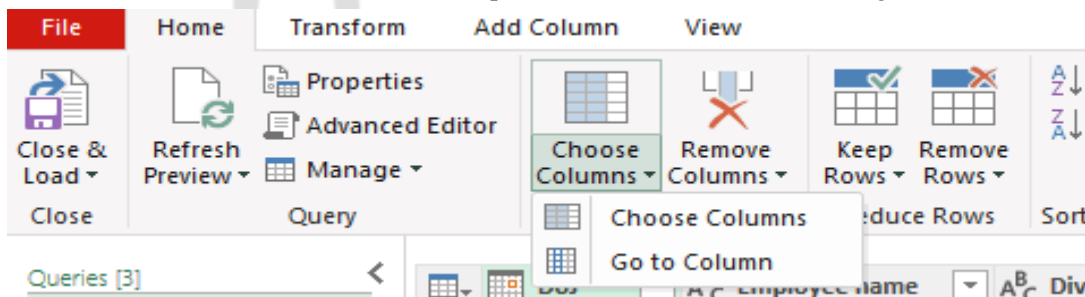

   - In the "Choose Columns" dialog box, you'll see a list of all the columns in your dataset.
     You can then check the boxes next to the names of the columns you want to keep in your final dataset. Unchecked columns will be removed.
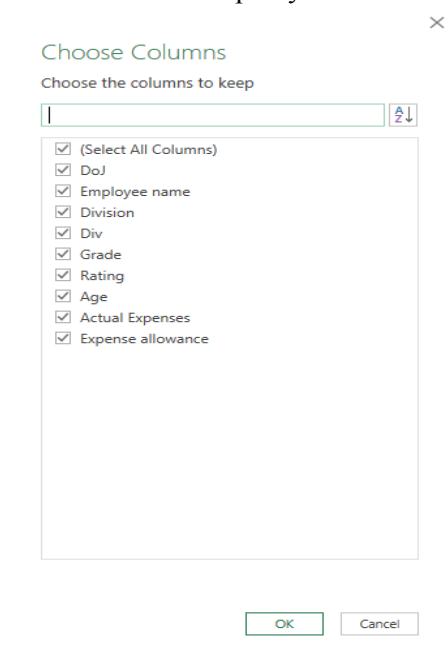
b. **Remove columns:**
  - In the Power Query Editor, click the column or hold down Ctrl to select multiple columns you wish to remove.
  - Right-click one of the selected columns and choose "Remove Columns" from the context menu. Alternatively, after selecting the columns, you can go to the "Home" tab and use the "Remove Columns" button directly.
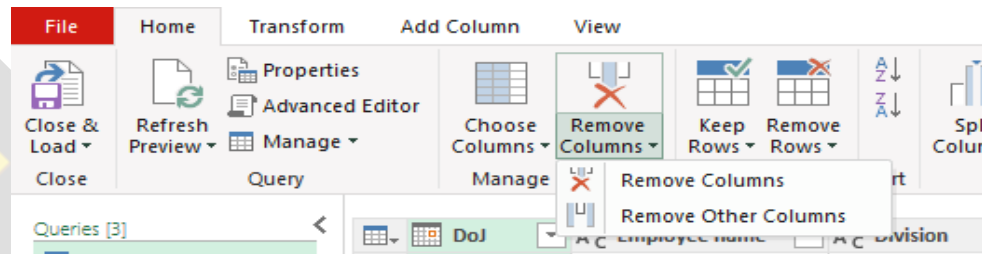
c. **Renaming columns:**

  - In the Query Editor window, click on the header of the column you want to rename to select it.
  - Right-click on the selected column header.
  - From the context menu, choose "Rename".
  - Alternatively, you can also double-click on the column header to enter edit mode and directly type the new name for the column.

3. **Reduce rows: (Dataset)**

  a. **Keep rows:**

  Significance:
  - By keeping only relevant rows, you reduce the dataset size, making subsequent transformations faster and more efficient.
  - It helps focus the analysis on the data that matters, eliminating noise from irrelevant entries.
  - It's an essential part of data cleaning, enabling the removal of outliers or erroneous data that could skew analysis results.

  Steps:
  - **Keep Top Rows:** To keep a specified number of top rows, go to the "Home" tab, click "Keep Rows" in the "Reduce Rows" dropdown, and select "Keep Top Rows." Enter the number of rows you want to keep.
  - **Keep Bottom Rows:** Similar to keeping top rows, select "Keep Bottom Rows" from the "Keep Rows" dropdown and specify the number of rows from the bottom you wish to keep.
  - **Keep Range of Rows:** To keep a specific range of rows, select "Keep Range" from the "Keep Rows" dropdown. You'll need to specify the starting row and the number of rows to keep.
  - **Keep Duplicates:** If you want to keep only the rows that have duplicate values in certain columns, go to "Keep Rows" and select "Keep Duplicates."

- **Keep Errors:** To keep only the rows that contain errors, select "Keep Errors" from the "Keep Rows" dropdown.
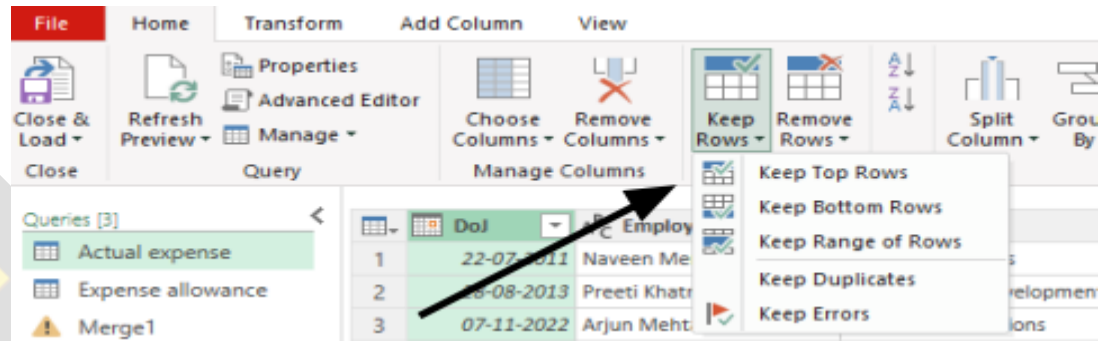

b. **Remove rows:**
- **Remove Top Rows**: If you need to remove a set number of rows starting from the top, you can use this option by specifying the number of rows to remove.
- **Remove Bottom Rows:** Similar to removing top rows, this option allows you to specify the number of rows to remove from the bottom of your dataset.
- **Remove Blank Rows**: To clean your dataset by removing rows where all cells are blank, select this option. It's useful for datasets with incomplete records.
- **Remove Duplicate Rows**: This feature allows you to remove rows that are exact duplicates of another row, helping in deduplicating your data. You can choose specific columns to check for duplicates, or select all columns.
- **Remove Errors:** If some rows contain errors in one or more cells, this option can remove those entire rows to ensure data quality.
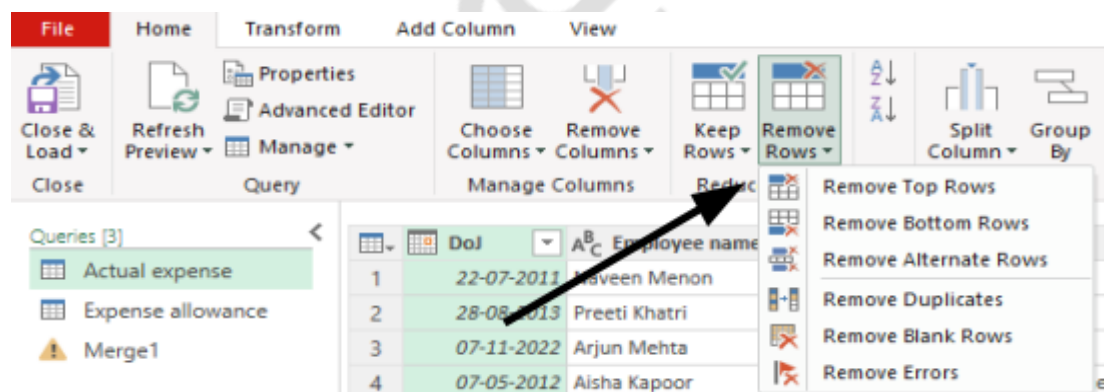
4. **Sorting of data:**
- In Power Query, you can sort your data by one or more columns in ascending or descending order.
- To sort data, select the column(s) you want to sort by, then go to the Home tab and click on "Sort Ascending" or "Sort Descending" in the Transform group.
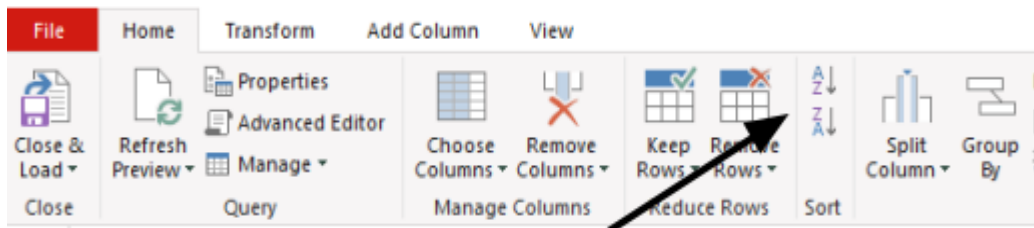
5. Filtering of data:
- Filtering allows you to include or exclude rows from your dataset based on specific criteria.
- To filter data, click on the drop-down arrow next to the column header you want to filter, then uncheck/select the values you want to include/exclude.

6. Changing data type:
- You can change the data type of a column to a different data type, such as text, number, date, etc.
- To change the data type, right-click on the column header, select "Change Type", and choose the desired data type from the list.
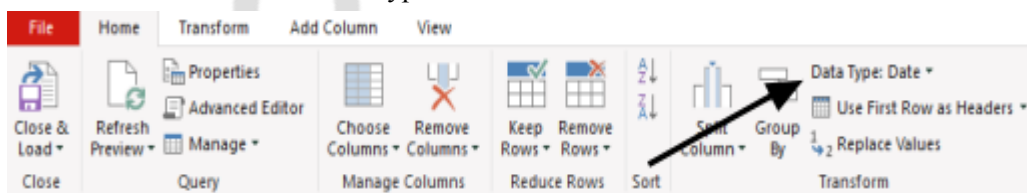

7. Split text to Columns:

- If you have text data that needs to be split into separate columns, you can use the "Split Column" feature.
- Select the column you want to split, go to the Home tab, click on "Split Column", and choose the delimiter (e.g., comma, space) to split the text.

8. How to handle missing values in Power Query:
- **a. Using replace option:**
  In Power Query, the "Replace" feature is used to replace specific values in your dataset with other values. It's a powerful tool for data cleaning and transformation.

  - **Correcting typos and inconsistencies**: If your dataset contains typos or inconsistencies in values, you can use the "Replace" feature to standardize them. For example, you can replace variations of "USA" with "United States" or "UK" with "United Kingdom".

  - **Handling missing or null values**: You can use the "Replace" feature to replace missing or null values with a specific value or placeholder. This ensures consistency in your dataset and prevents errors in downstream analysis.
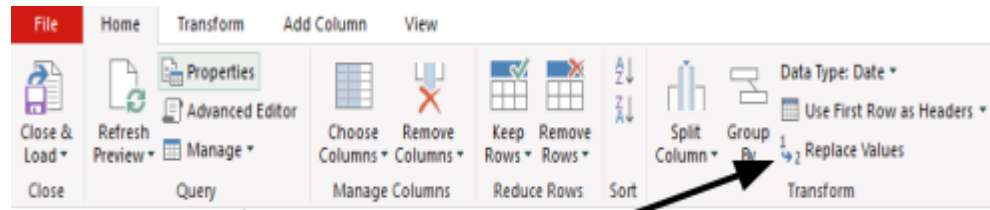
9. Add new data in the table and refresh:

You can add the new data in the original table and click refresh from the ribbon above. Once you click refresh all the related data will get updated and the data in the Power Query Editor will also get updated automatically.

10. Use Group By in Power Query:

(Dataset)

In Power Query for Excel, the "Group By" feature is used to aggregate data based on one or more columns. This feature is particularly useful when you need to summarize data, perform aggregations, or combine data from multiple records into single values based on a grouping column.

- Once your data is loaded from Get Data it will open in the Power Query Editor. This is where you can apply various transformations to your data.
- In the Power Query Editor, select the column you want to group by. You can do this by clicking on the column header.
- With the column selected, go to the "Home" tab on the ribbon. Click on the "Group By" button in the "Transform" group. This will open the "Group By" window.
- In the "Group By" window, you will see options to select the column you want to group by and to define the aggregation.
- You can choose the column(s) from the dropdown menu under "Group By".
- Then, you can add aggregation by selecting a new column name, choosing an aggregation function (like sum, count, average, min, max), and selecting the column to aggregate.

Examples of Aggregations:
- Sum: Calculate the total for numerical columns.
- Average: Compute the average for numerical columns.
- Count: Count the number of rows in each group.
- Min/Max: Find the minimum or maximum value in each group.

Group By

Specify the columns to group by and one or more outputs.

○ Basic   ● Advanced

Division ▼

Add grouping

New column name	Operation	Column
Count	Count Rows ▼	▼

Add aggregation

OK   Cancel