Data Science and Analytics (IT3080)

B.Sc (Hons) in IT (IT Specialization)Year 3

SLIIT
Discover Your Future

Lab Sheet 06

Accessing and Preparing Data Using Power BI Desktop

Introduction

By the end of this lab, you will have learned:

- How to load data from Comma-Separated Values (CSV) sources
- How to manipulate the data to prepare it for reporting
- How to prepare the tables in Power Query and load them into the model

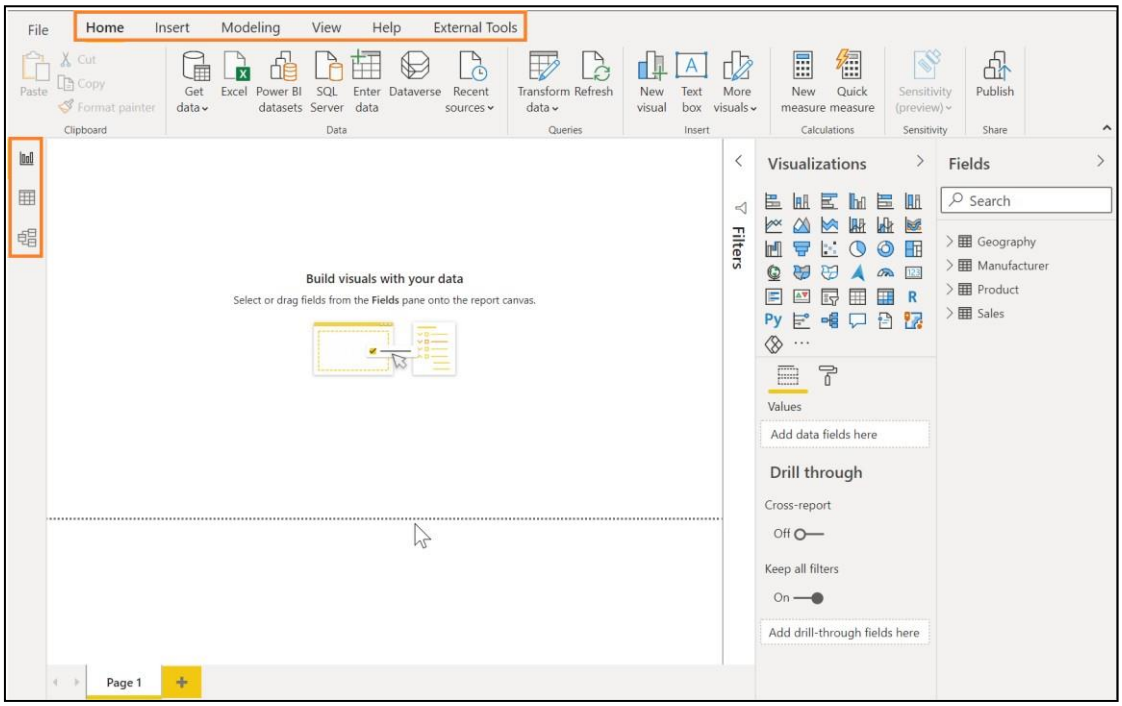
Power BI has the capability of ETL.

Extract – CSV, EXCEL, MS Access, SQL, NO SQL, Web API, Web Scraping, Cloud
Transform – Imputation (Dealing with missing values)
Add or remove columns/ rows
Merging of columns
Rounding floating values to decimal
Splitting columns etc.
Load

Power BI Desktop

Power BI Desktop - Layout

Let's start with the main **Power BI Desktop** window and become familiar with the distinct sections available.



1. On the top of the window, you see the **Home** tab where the most common operations you perform are available.

2. The **Insert** tab in the ribbon allows you to insert shapes, a text box or new visuals

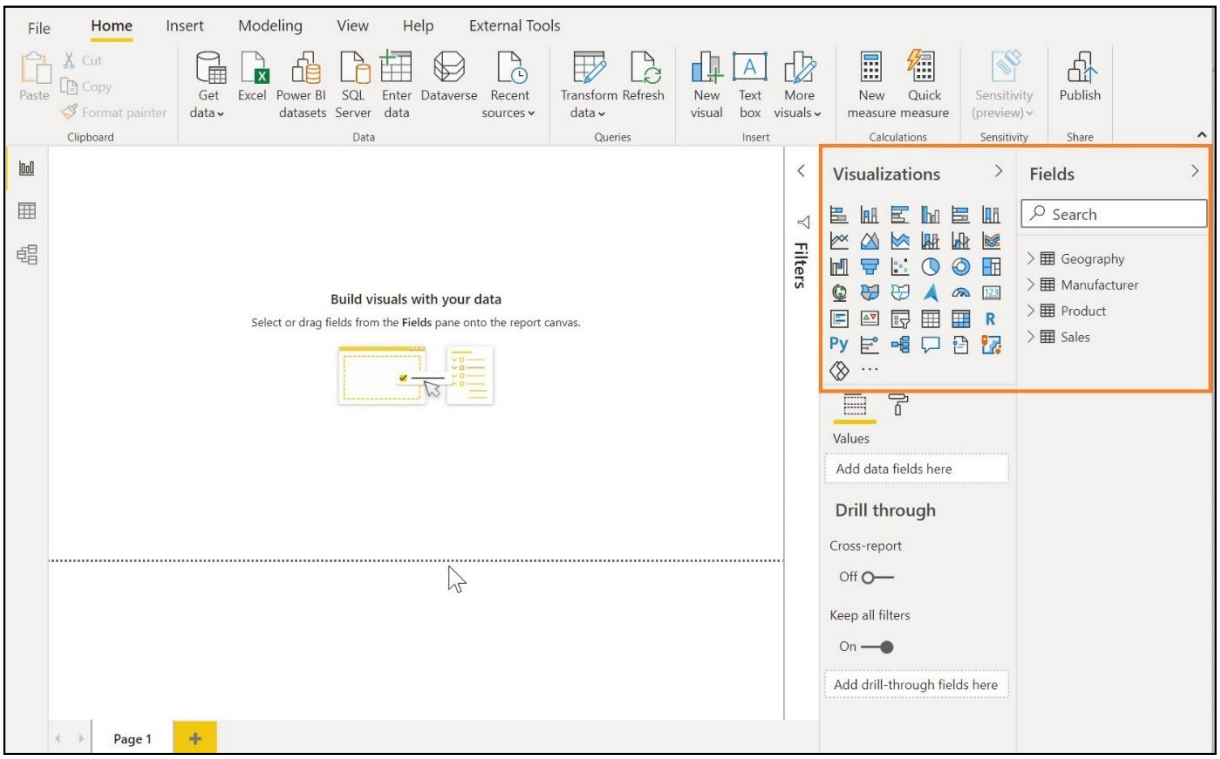
3. The **Modeling** tab in the ribbon enables additional data modelling capabilities like adding custom columns and calculating measures.

4. The **View** tab has options to format the page layout.

5. The **Help** tab provides self-help options like guided learning, training videos and links to online communities, partner showcase and consulting services.

6. On the left side of the window, you have three icons, **Report, Data and Model**. If you hover over the icons, you can see the tooltips. Switching between these allows you to see the data and the relationships between the tables.

7. The center **white space** is the canvas where you will be creating visuals.



8. The **Visualizations** panel on the right allows you to select visualizations, add values to the visuals, andadd columns to the axis or filters.
9. The **Fields** window on the right panel is where you see the list of tables which were generated from the queries. Click the ▼ icon (downward facing triangle) next to a table name to expand the field list forthat table.

Power BI Desktop – Accessing Data

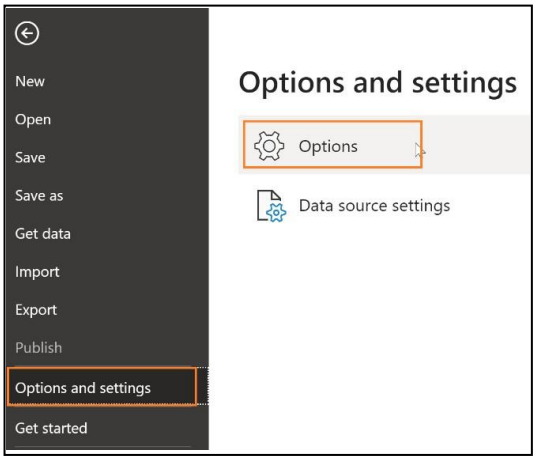
Power BI Desktop - Get Data

Let's start by looking at the data files. The dataset contains Calendar, Customers, Products and Sales data in CSV files. Sales data in 3 different files which represents the years 2015,2016 and 2017.

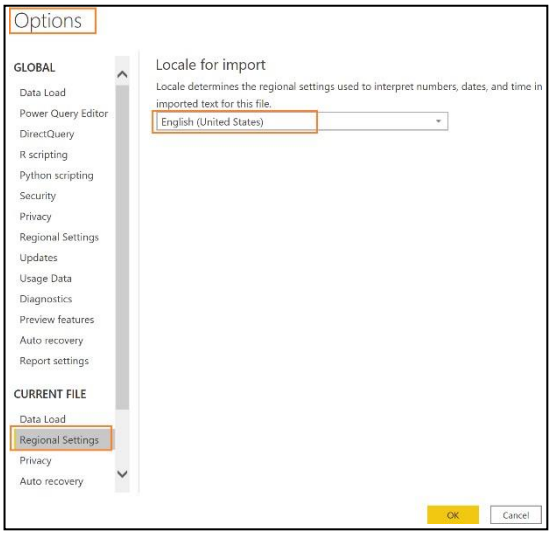
1. If you don't have the **Power BI Desktop** open, launch it now.
2. Click **Already have a Power BI Account? Sign in** option.
3. **Sign in** using your Power BI credentials.
4. You will see the startup screen opens. Click on the **X** on the top right corner of the dialog box to close it.

Let's set the **Locale** to US English to make it convenient in the rest of this lab.

5. From the ribbon, click **File**, then click **Options and settings**, then click **Options**.

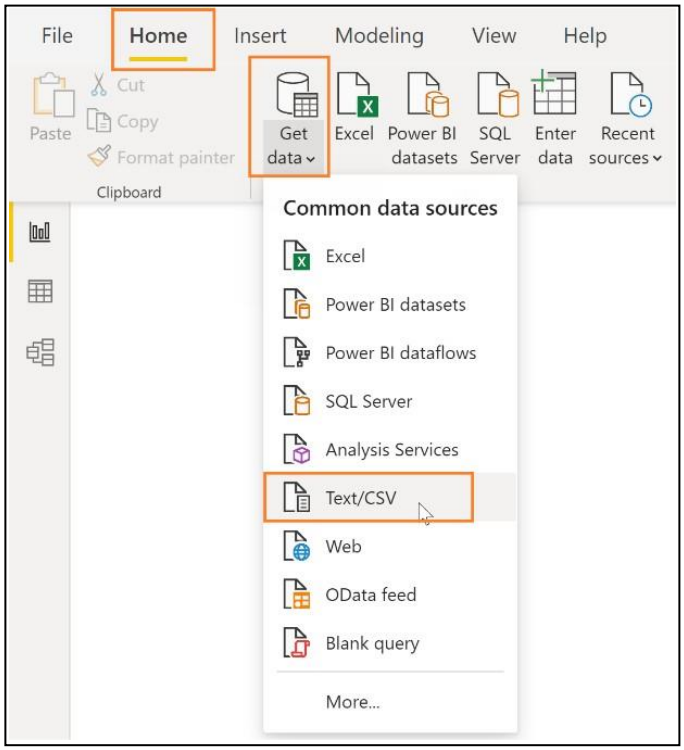


6. In the left panel of **Options** dialog box, click **Regional Settings** under **Current File**.
7. From the **Locale** drop-down, click **English (United States).**
8. Click **OK** to close the dialog box.



The next step is to load data to Power BI Desktop. We will load Calendar data which is in CSV file.

9. From the ribbon, click **Home** and then click the **Get Data** drop-down arrow**.**
10. Click **Text/CSV.**

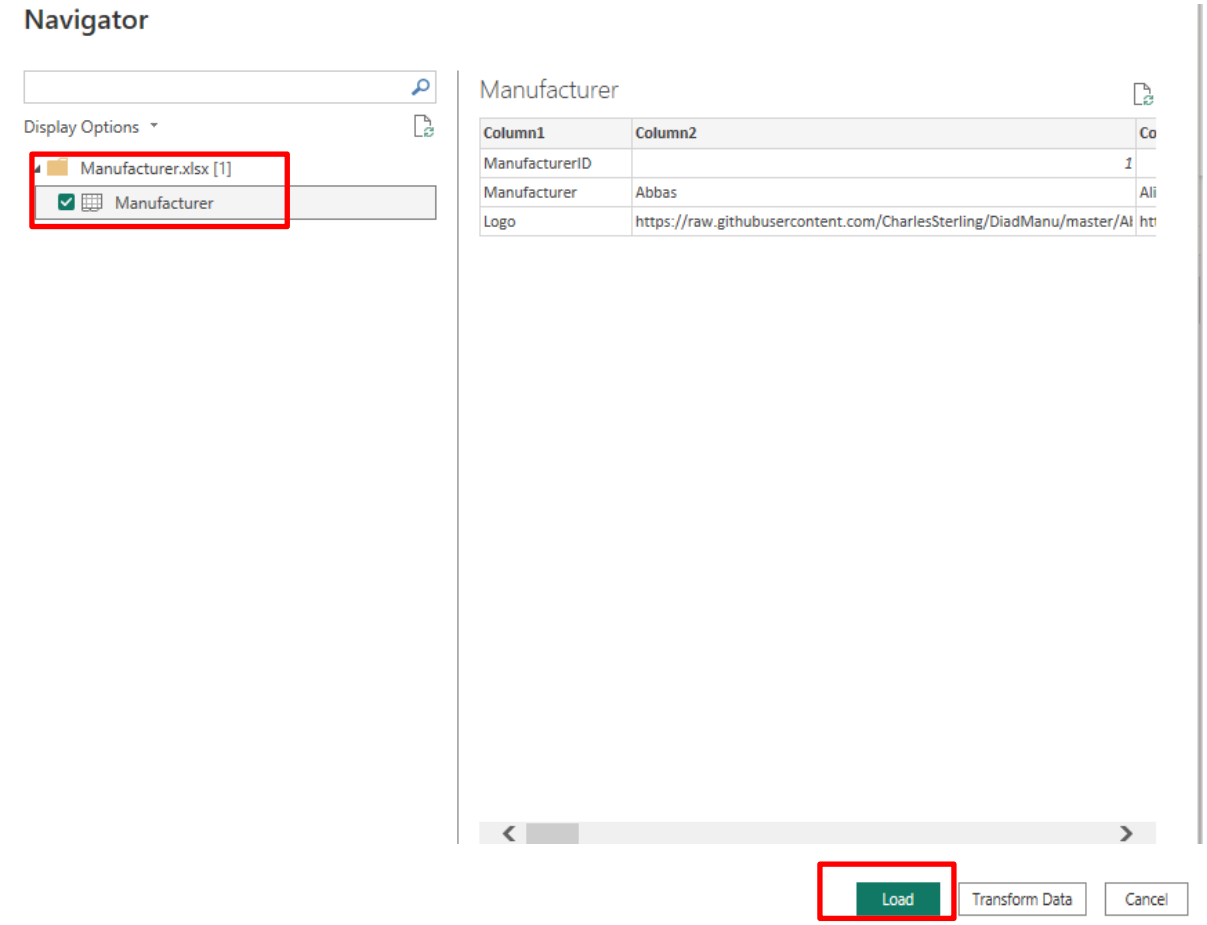


11. Browse to the **Calendar.csv** file location, and then click **Calendar.csv.**
12. Click the **Open** button.

Power BI detects the data type within each column. Load the data in **Customers.csv** and **Products.csv** files as above.

Now let's get the data that is in Excel source file.

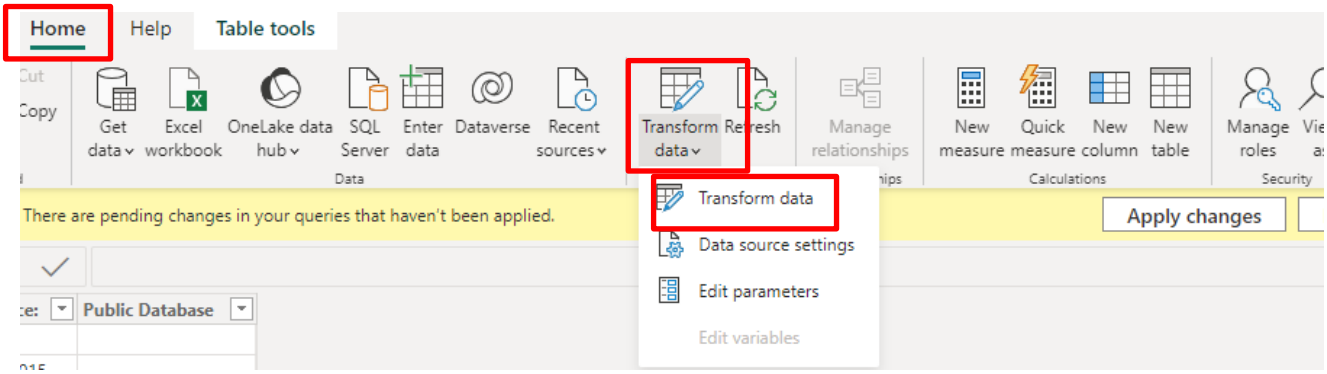
13. From the ribbon, click **Home** and then click the **Get Data** drop-down arrow**.**
14. Click **Excel.**
15. Browse to the **Manufacturer.xls** file location, and then click **Manufacturer.xls.**
16.Click the **Open** button.
17. **Navigator** Dialog box opens.
18. Select the **Manufacturer** and click **Load**.



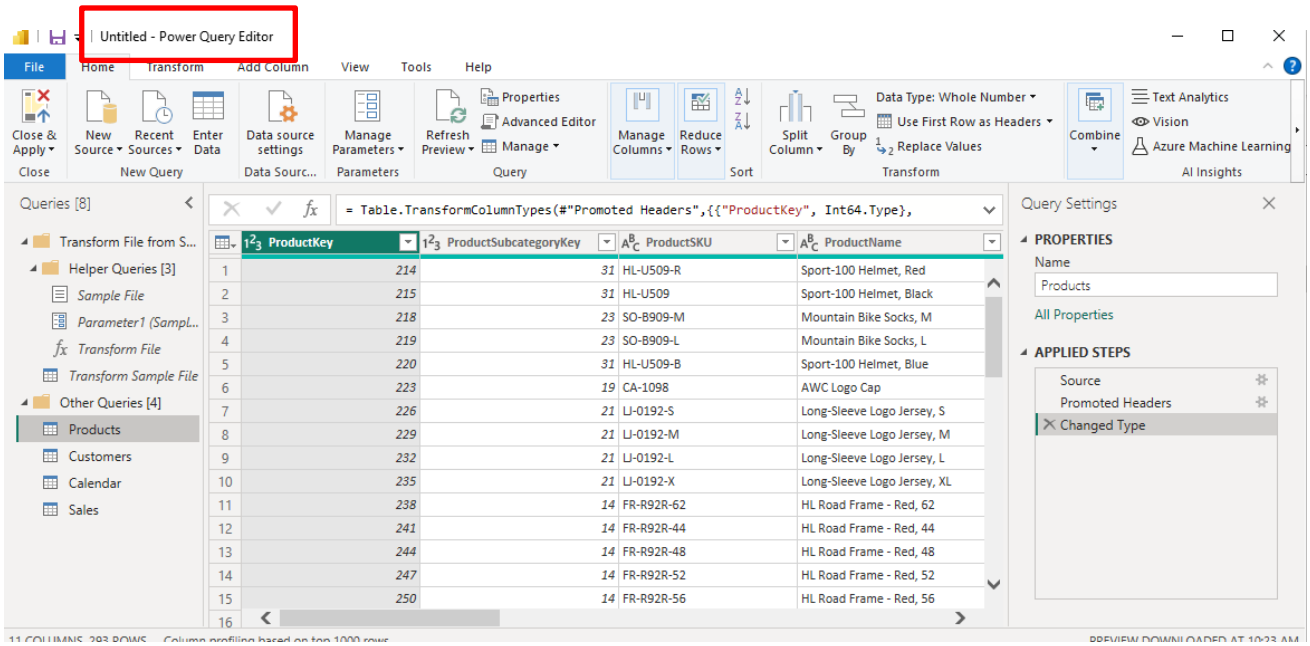
After completing your selection, you have three options – Load, Edit or Cancel.

- **Load** adds the data from the source into Power BI Desktop for you to start creating reports.
- **Transform Data** allows you to perform data shaping operations such as merging columns, adding additional columns, changing data types of columns as well as bringing in additional data.
- **Cancel** gets you back to the main canvas.

16. Go to the **Home** tab and click **Transform Data** as shown in the screenshot. A new window opens.



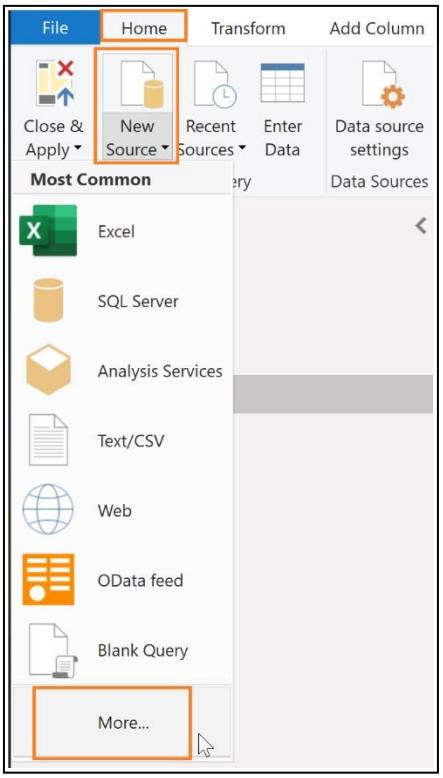
You should be in the Query Editor window as shown in the screenshot below. The Query Editor is used to perform data shaping operations.



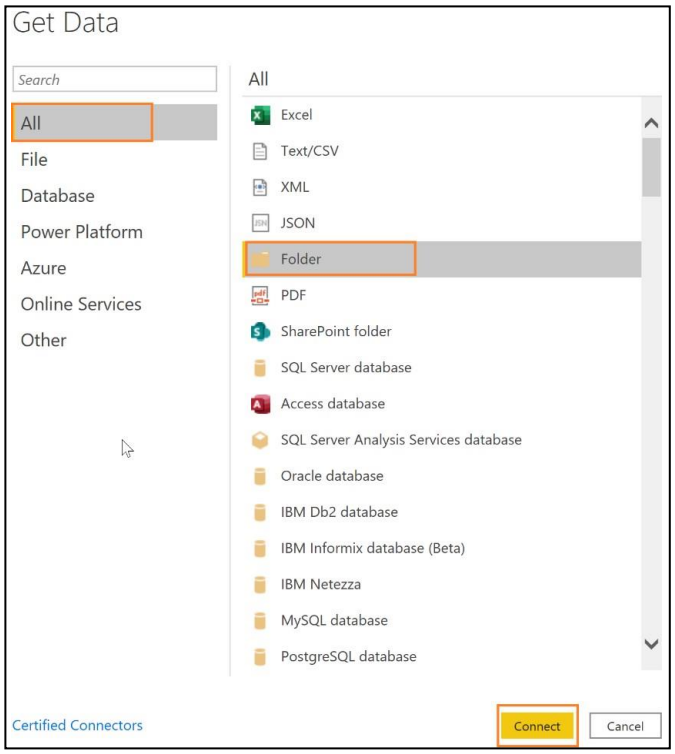
Power BI Desktop - Adding additional data

All the sales data which are belonging to years 2015, 2016 and 2017 are in three different files. Power BI provides an easier way to load all the files in a folder together.

17. On the **Home** tab of the Query Editor, click on the **New Source** drop-down menu.
18. Click **More…** as shown in the figure.



19. The Get Data dialog box opens.
20. In the **Get Data** dialog box, click **Folder** as shown in the diagram.
21. Click **Connect** and the **Folder** dialog box will open.

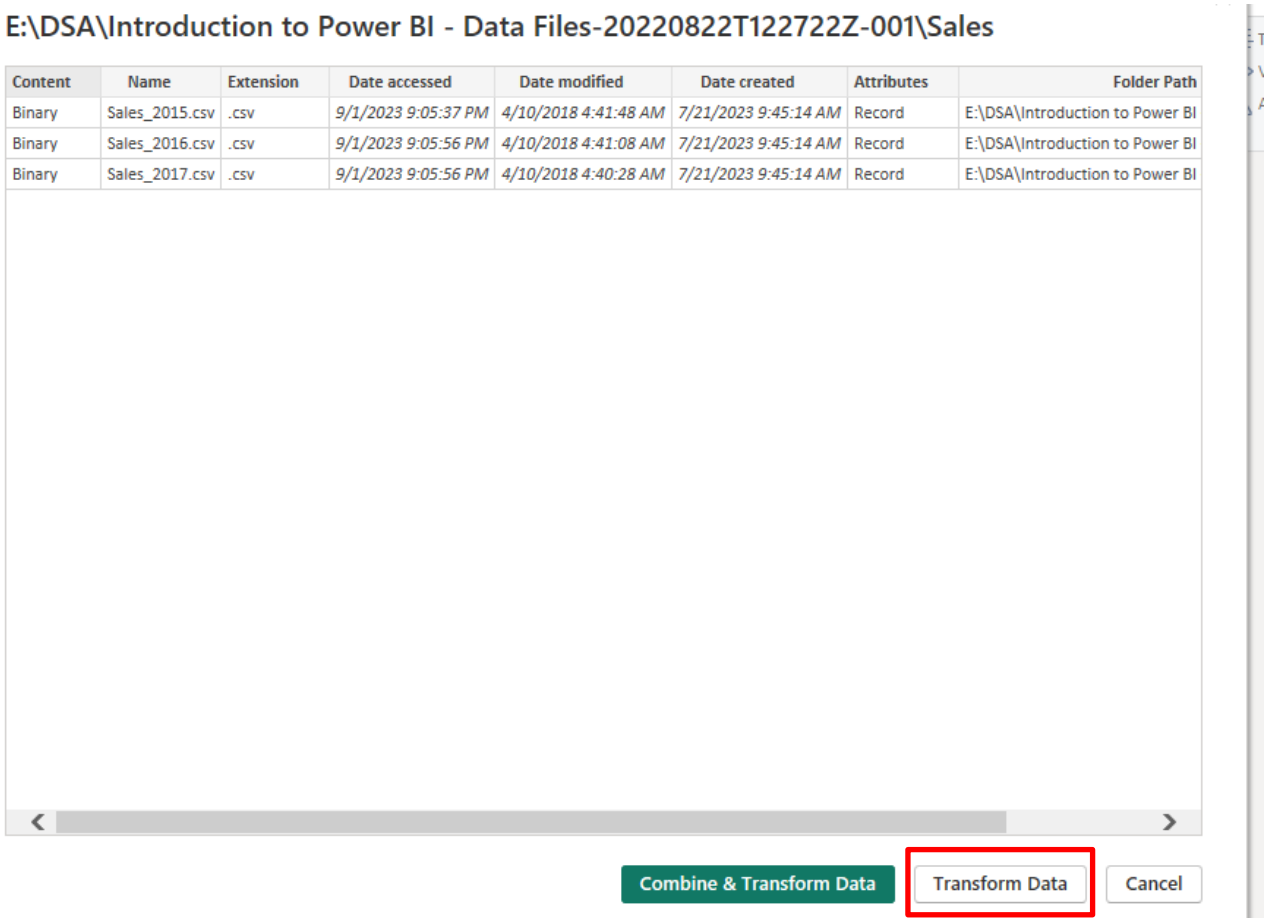


22. Click the **Browse…** button.
23. In the **Browse for Folder** dialog box, navigate to the folder location where you saved the **Sales_2015, Sales_2016** and **Sales_2017** CSV files. (All the three files should be in a different folder)
24. Open and Click the folder.
25. Click **OK** (to close the **Browse for Folder** dialog box).
26. Click **OK** (to close the **Folder** dialog box).

Note: This approach will load all the files located in the folder. This is useful when you have a group that puts files on an FTP site each month and you are not always sure of the names of the files or the number of files. All the files must be of the same file type with columns in the same order.

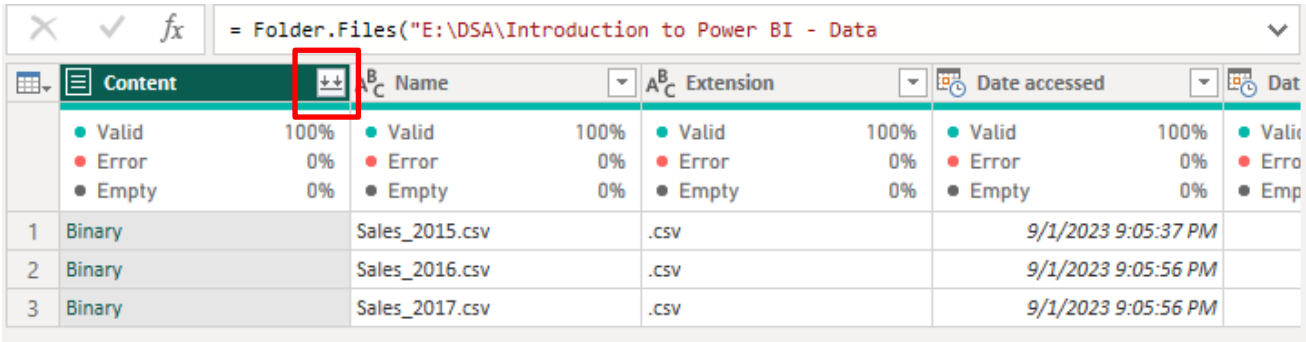
The dialog box will display the list of files in the folder.

27. Click **Transform Data**.



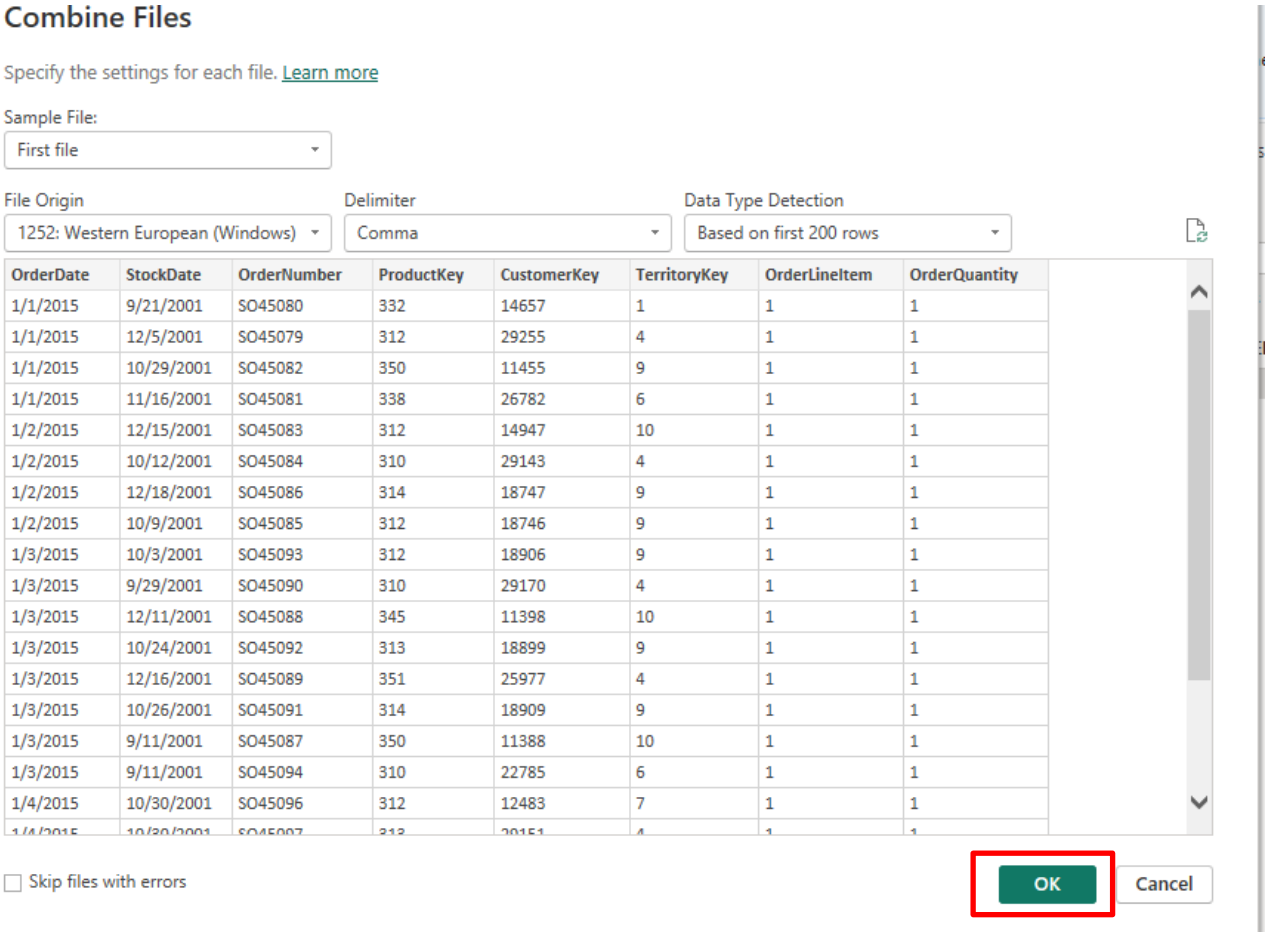
Note: The data in your file for **Date accessed**, **Date modified**, and **Date created** might be different than the dates displayed in the screenshot.

Click the two down arrows before the **Content**.



The **Combine Files** dialog box will open. By default, Power BI will again detect the data type based on the first 200 rows. Notice there is an option to select various file Delimiters. The file we are working with is Comma delimited, so let's leave the Delimiter option as Comma.

28. Click **OK**.



Also notice that on the right panel under **APPLIED STEPS** you will see the list of transformations and steps that have been applied. You can navigate through each change made to the data by clicking on the step. Steps can also be deleted by clicking on the **X** that appears to the left of the step. The properties of each step can be reviewed by clicking on the **gear** to the right of the step.

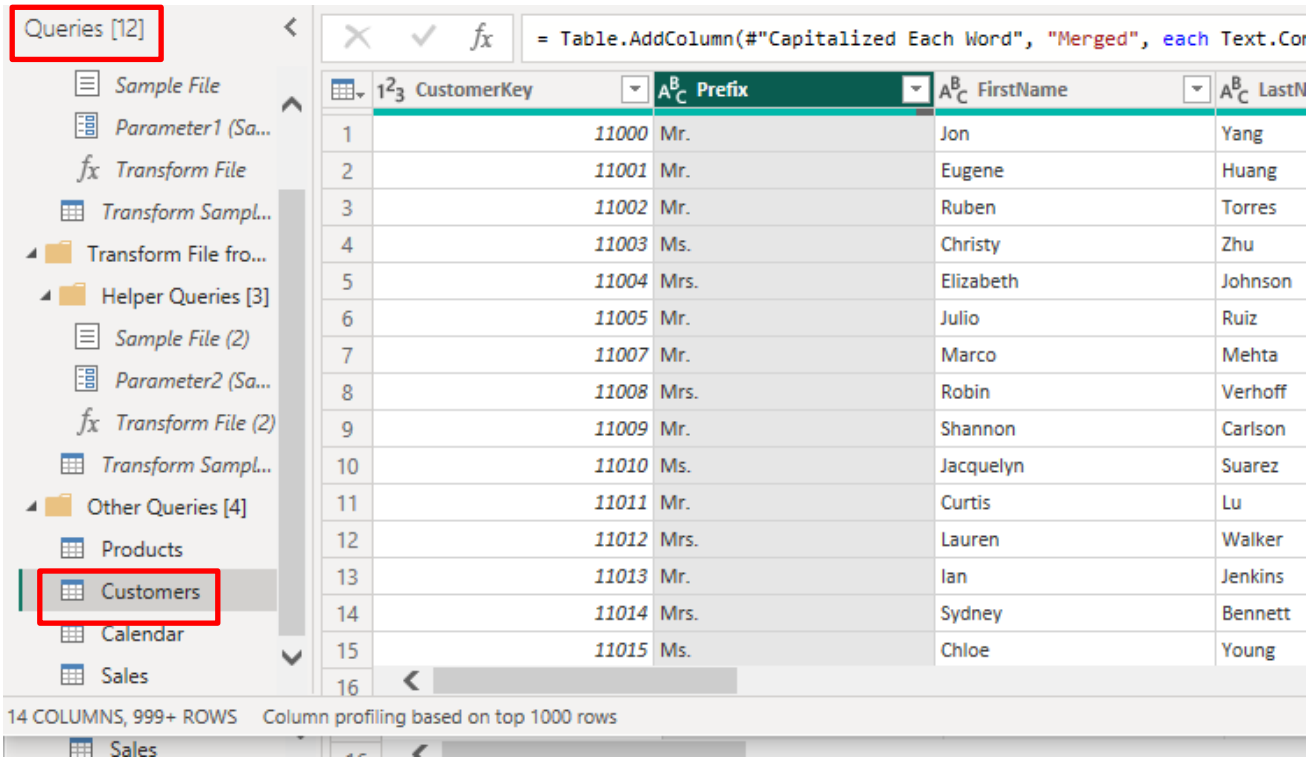
Power BI Desktop – Data Preparation

In this section, we will explore methods to transform data in the data model. Transforming the data by renaming tables, updating data types, and appending tables together ensures that the data is ready to be used for reporting. In some instances, this means cleaning the data up so that similar sets of data can be combined. In other instances, groups of data are renamed so that they are more easily recognized by end users and report writing is simplified.

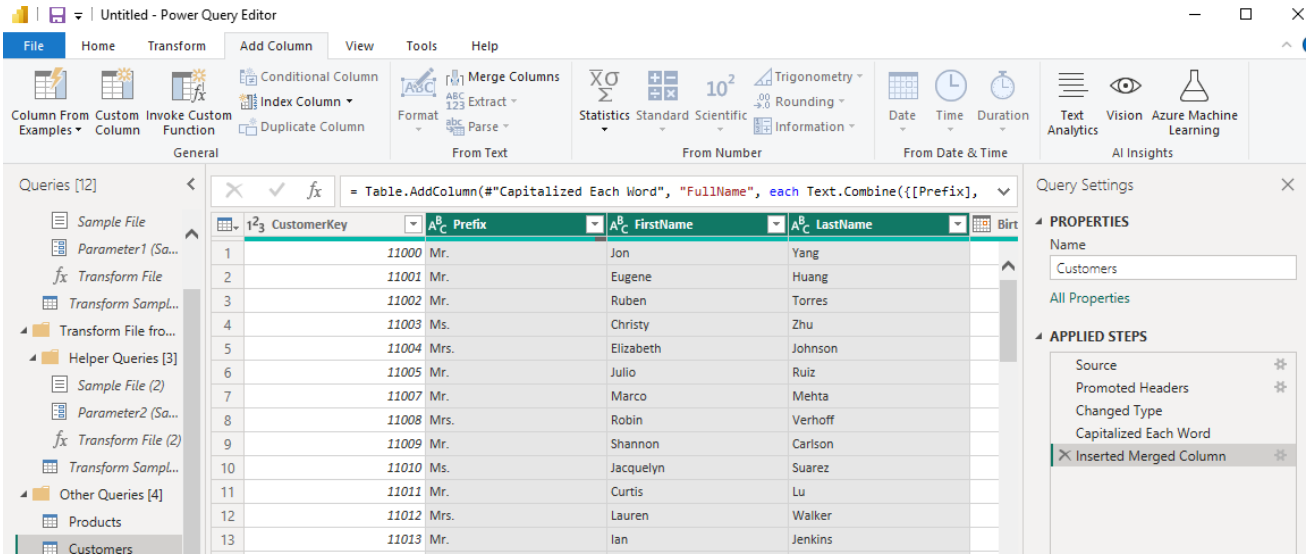
Power BI Desktop – Merge and Rename Columns

The Query Editor window should appear as shown below.
If formula bar is disabled, you can turn on the formula bar from the View ribbon.

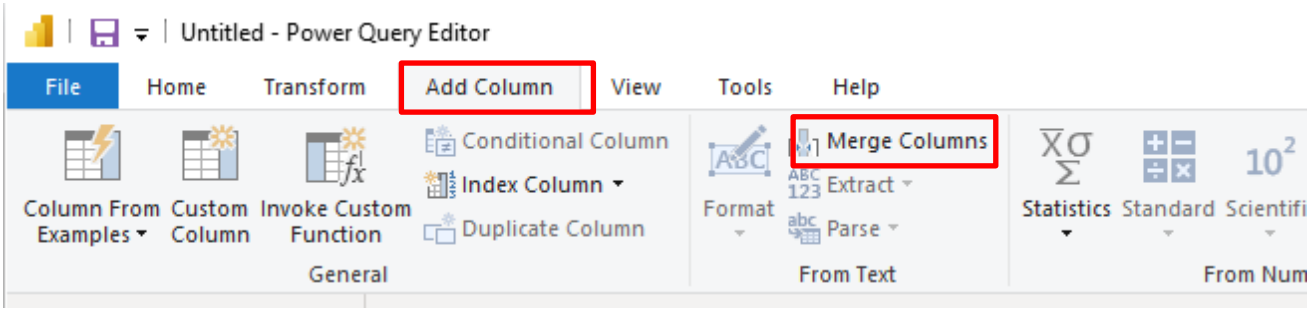
1. Select the **Customers** table from the **Queries** pane in the left hand side.



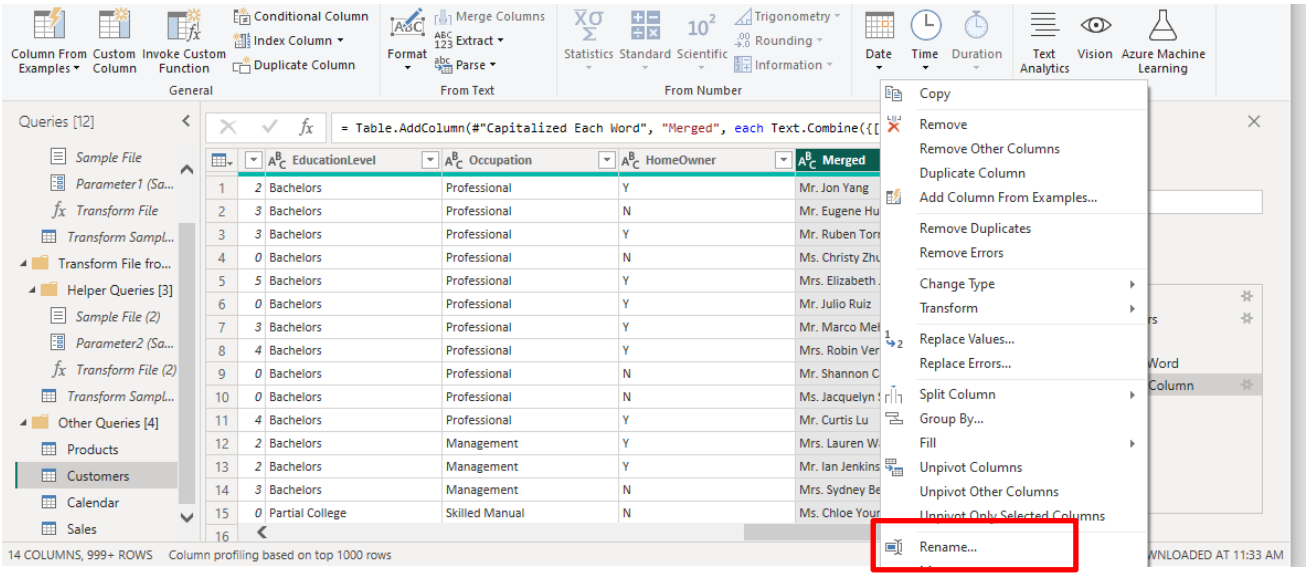
2. Select the columns you want to merge using the ‘Ctrl’ key.



3. Under **Add Column** tab select **Merge Columns**. The merged column will appear as the last column in the **Customers** table.



4. Column name will appear as **Merged**. Right click the **Merged** column and select **Rename.**

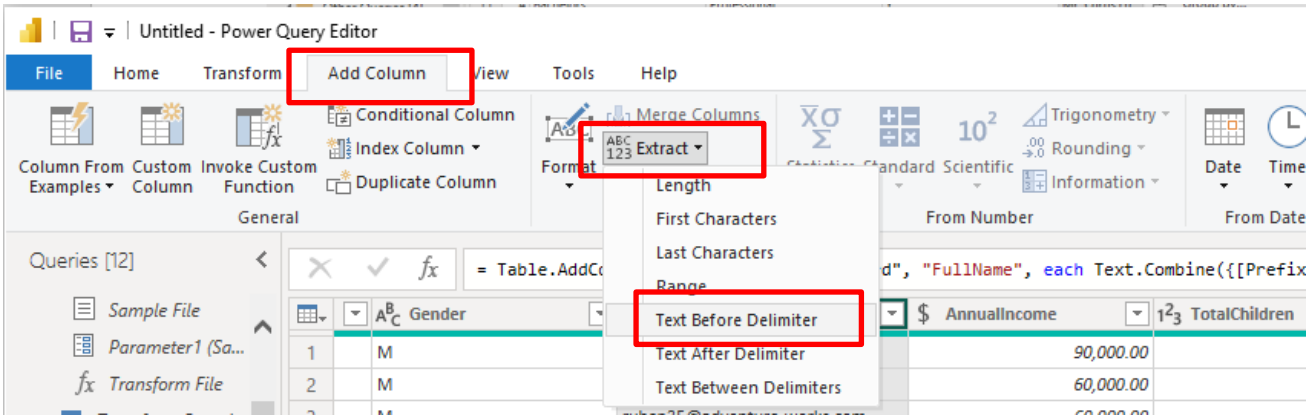


5. Type the new name as **FullName**.

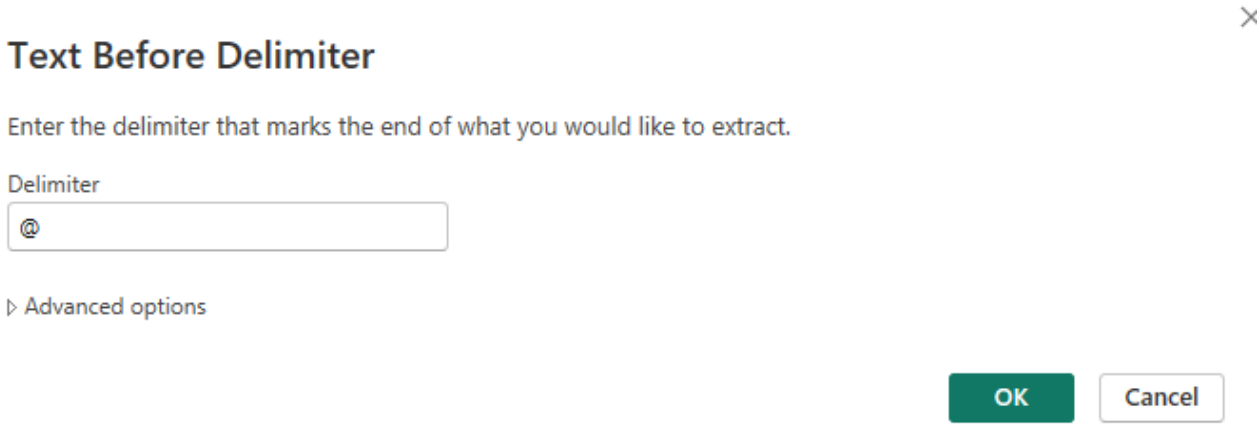
Power BI Desktop – Extract useful data from columns

In the Customers query, notice the **EmailAddress** column. Let's extract the username from the email and save it in a new column.

1. From the ribbon, click **Add Column**, click **Extract**, and then click **Text Before Delimiter**. The **Text Before Delimiter dialog** box opens.



2. Enter @ symbol as shown in the screenshot.
3. Click **OK**.

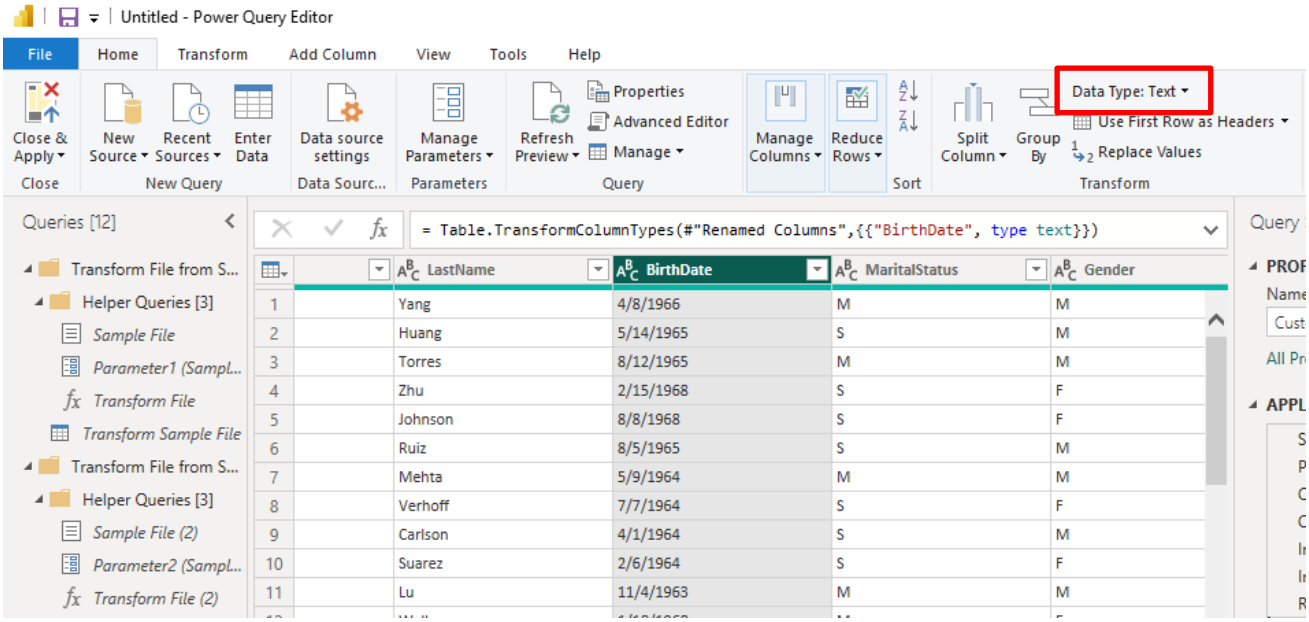


4. The new column will add at the end of the **Customers** table. Rename the column to **Username**.

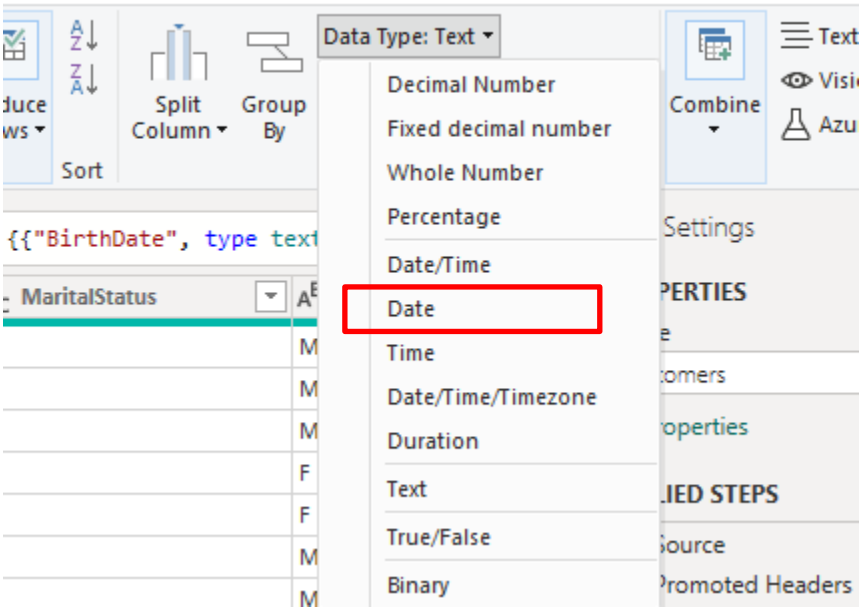
   Note: If the currency values are with the currency symbol, before the analysis it is needed to extract the currency value without the currency symbol.

Power BI Desktop – Change the data type of a column

Select each column and check the data type. **BirthDate** column has the data type **Text**.



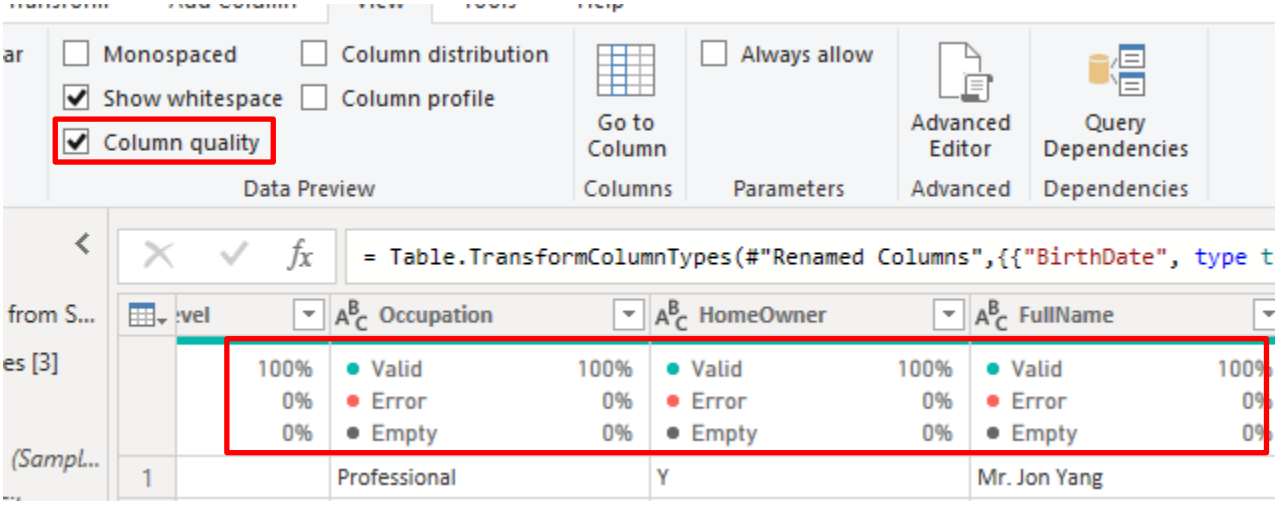
Click on the down arrow and select the data type **Date** from the drop down list.



Power BI Desktop – Check the quality of the columns

Power BI has the capability of checking valid, erroneous and empty cells automatically. To check the quality of the columns,

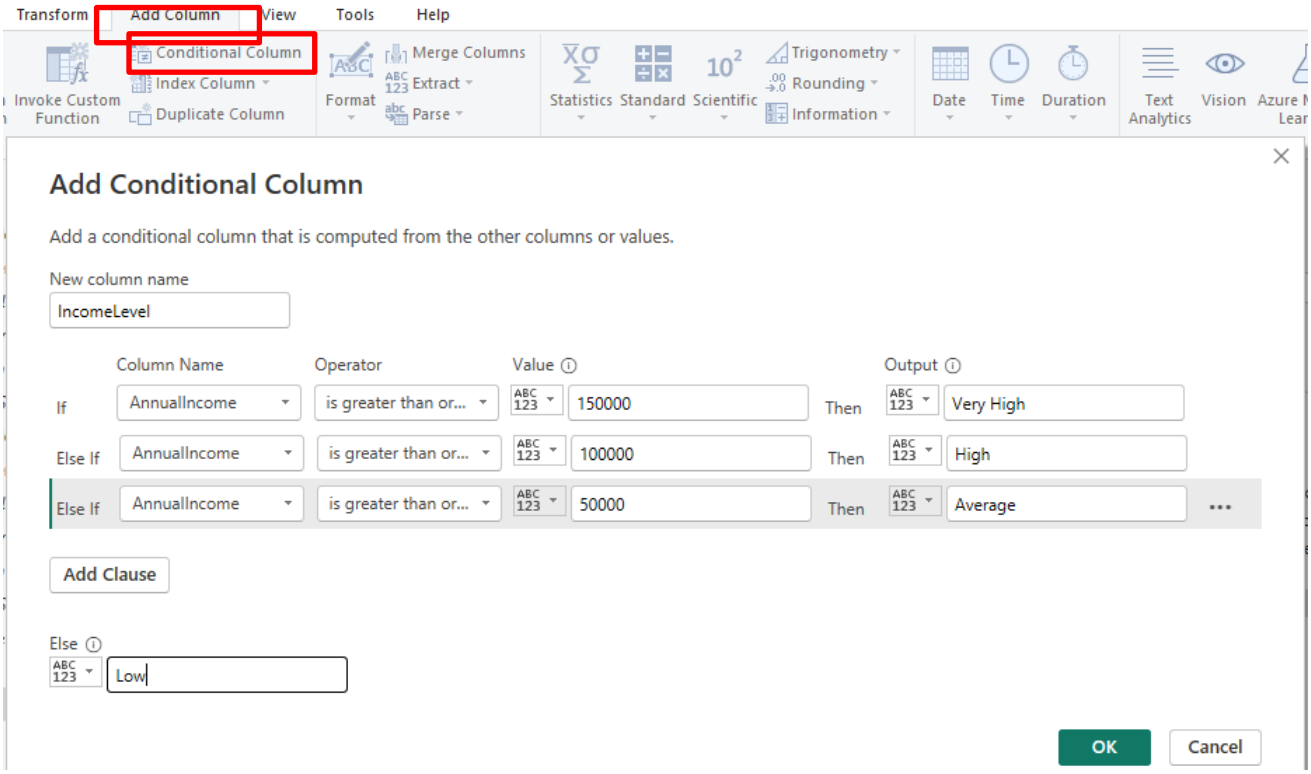
1. Click **View** tab on the ribbon.
2. Then check the box before **Column quality**.
3. Percentage values of valid, error and empty cells will display below the column names.



Power BI Desktop – Convert numerical data to categorical

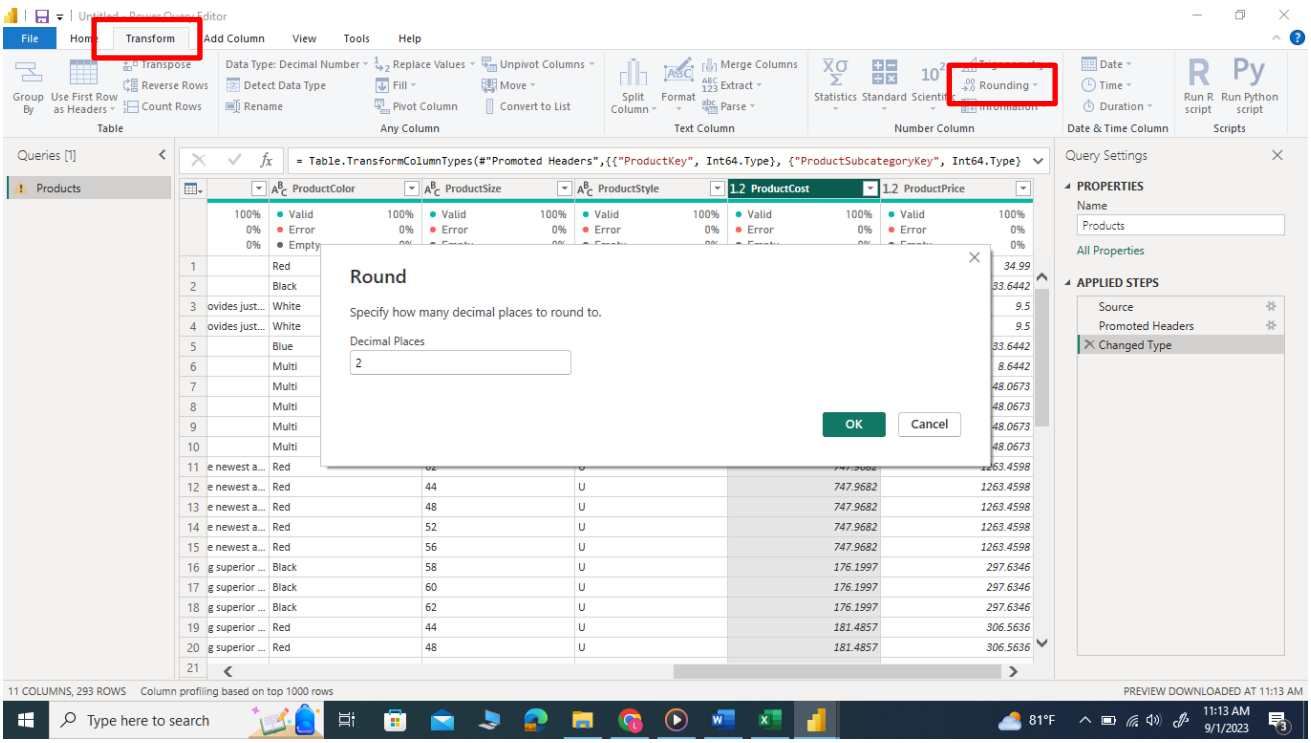
AnnualIncome column has numerical data. Let's convert it into categorical.

1. From the ribbon, click **Add Column** and then click **Conditional Column**. The **Add Conditional Column** dialog box opens.
2. Type the New column name as **IncomeLevel.**
3. Enter the categories as following and then click **OK**.
   (Click the **Add Clause** button to enter conditions for Else If)



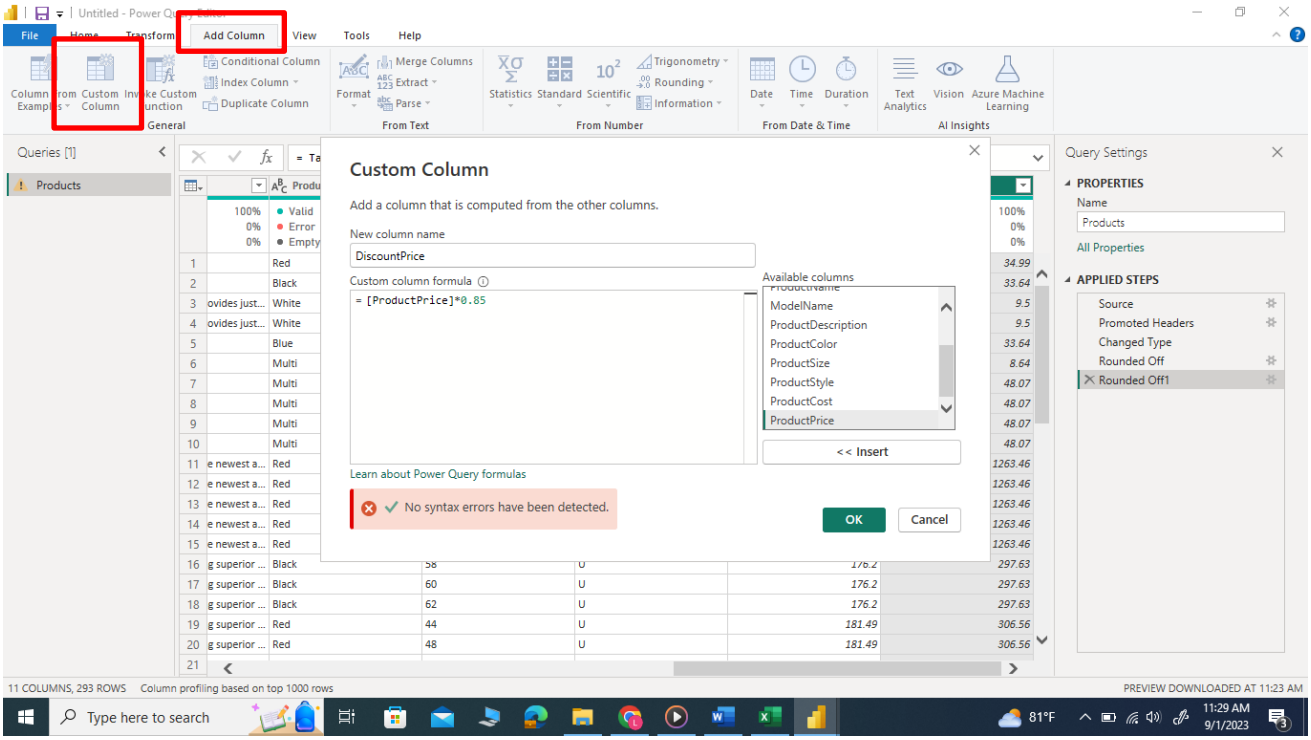
Power BI Desktop – Rounding floating values to decimal places

1. From the ribbon, click **Transform** and then click **Rounding**. The **Round** dialog box opens.
2. Enter the decimal places to round to**.**
3. Click **OK**.



Power BI Desktop – Adding a calculated column

1. From the ribbon, click **Add Coulmn** and then click **Custom Column**. The **Custom Column** dialog box opens.
2. Enter the **New column name** as **DiscountPrice.**
3. Enter the **Custom Column formula** as follows and Click **OK**.
4. Change the data type into **Decimal Number**.
5. Round the decimal places into two.

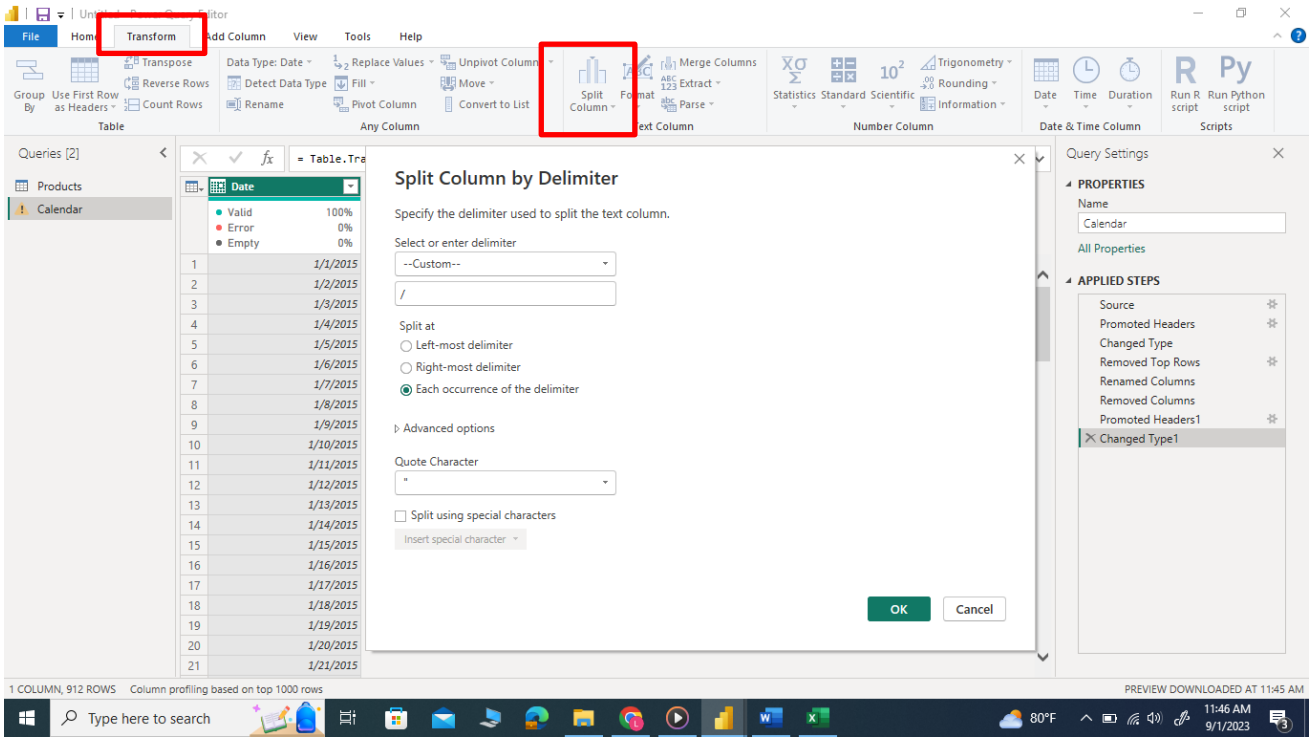


Power BI Desktop – Splitting columns

Power BI detects the date in month/date/year format. If the date is not in that format, split the date and merge the columns in the order month/date/year. Let's split the date into three columns.

1. From the left panel, click the **Calendar** Query.
2. Click the **Date** column.
3. From the ribbon, click **Transform**, click **Split Column**, and then click **By Delimiter**. The Split **Column byDelimiter dialog** box opens.
4. In the dialog box, make sure that **Custom** is selected in the **Select or enter delimiter** drop-down menu.
   Note: The **Select or enter delimiter** drop-down menu has some of the standard delimiters like comma, colon, and so on.
5. Notice that in the text area, there is a slash (/). Power BI assumes we want to split by (/). Make sure it has selected.

6. Click **OK**.



Note: If the delimiter occurs multiple times, the **Split at** section provides the option to split only once (either left most or right most) or the option to split the column on each occurrence of the delimiter.

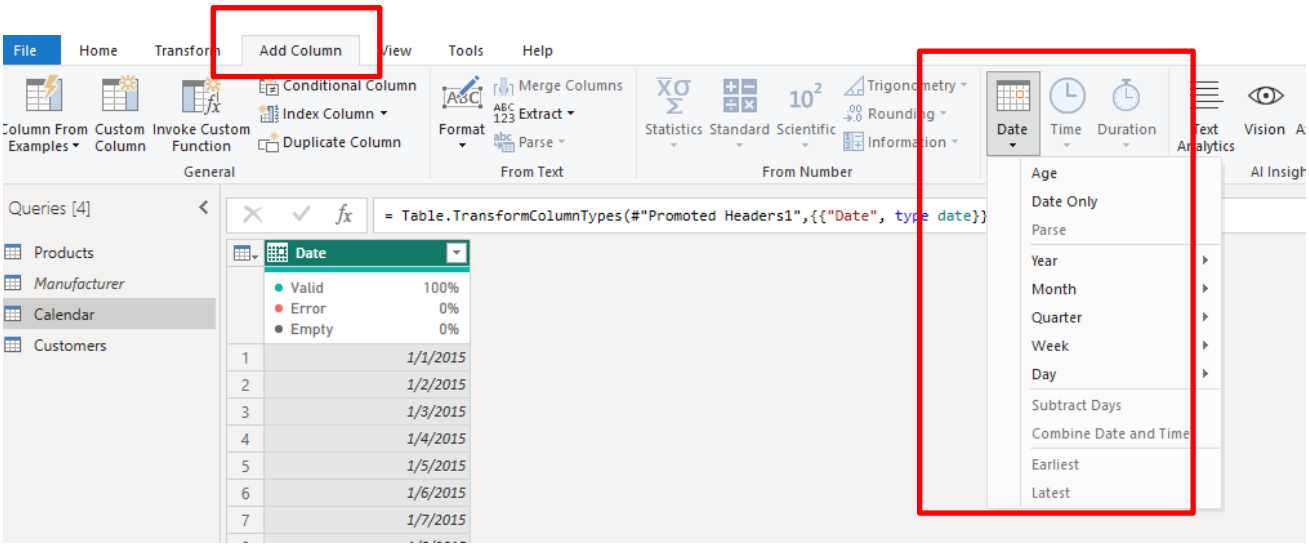
In this scenario, the delimiter (/) occurs twice, therefore the Date column is split into three columns.

7. Then merge the columns considering the order date/month/year.

Power BI Desktop – Adding New Columns

1. From the left panel, click the **Calendar** Query.
2. Click the **Date** column.
3. From the ribbon, click **Add Column**, and then click the down arrow of **Date**. Select **Year, Month, Name of the Month, Start of the Month, Start of the Week** and **Name of the Day** one by one from the drop down list.

   Note: Select the Date column every time before add a new column.



Power BI Desktop – Filling empty values

In our scenario, some of the data is not in the right format. Power BI provides extensive transformation capabilities to clean and prepare data to meet your needs. Let's start with the **Products** query.

Notice that the **ProductStyle** column has a lot of null values. Hover over the green/gray bar (known as the quality bar) below the column header. This allows you to easily identify errors and empty values in yourdata previews.

	ProductColor		ProductSize		ProductStyle		1.2 Product
100%	Valid	100%	Valid	100%	Valid	71%	Valid
0%	Error	0%	Error	0%	Error	0%	Error
0%	Empty	0%	Empty	0%	Empty	28%	Empty
	Red		0				
	Black		0				
ovides just...	White		M		U		
ovides just...	White		L		U		
	Blue		0				
	Multi		0		U		
	Multi		S		U		
	Multi		M		U		
	Multi		L		U		
	Multi		XL		U		
e newest a...	Red		62		U		
e newest a...	Red		44		U		
e newest a...	Red		48		U		

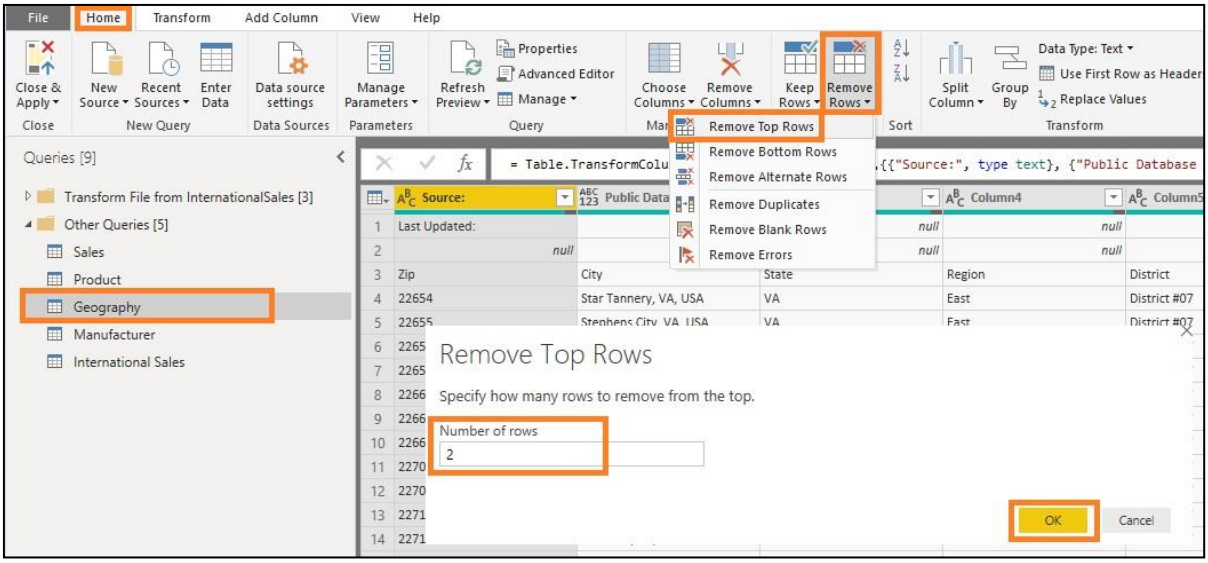
1. In the left panel, click the **Products** Query.
2. Click the **ProductStyle** column.
3. From the ribbon, click **Transform**, click **Fill**, and then click **Down**.

Notice how all the null values are filled with the appropriate Product Style values.

Power BI Desktop – Removing unwanted rows

In the **Calendar** query, notice that the first two rows are informational. They are not part of the data.

1. In the left panel, click the **Calendar** query.
2. From the ribbon, click **Home**, click **Remove Rows**, and then click **Remove Top Rows**.
3. The **Remove Top Rows** dialog box opens. Enter **2** in the text box since we want to remove the topinformational data row and the blank second row.
4. Click **OK**.



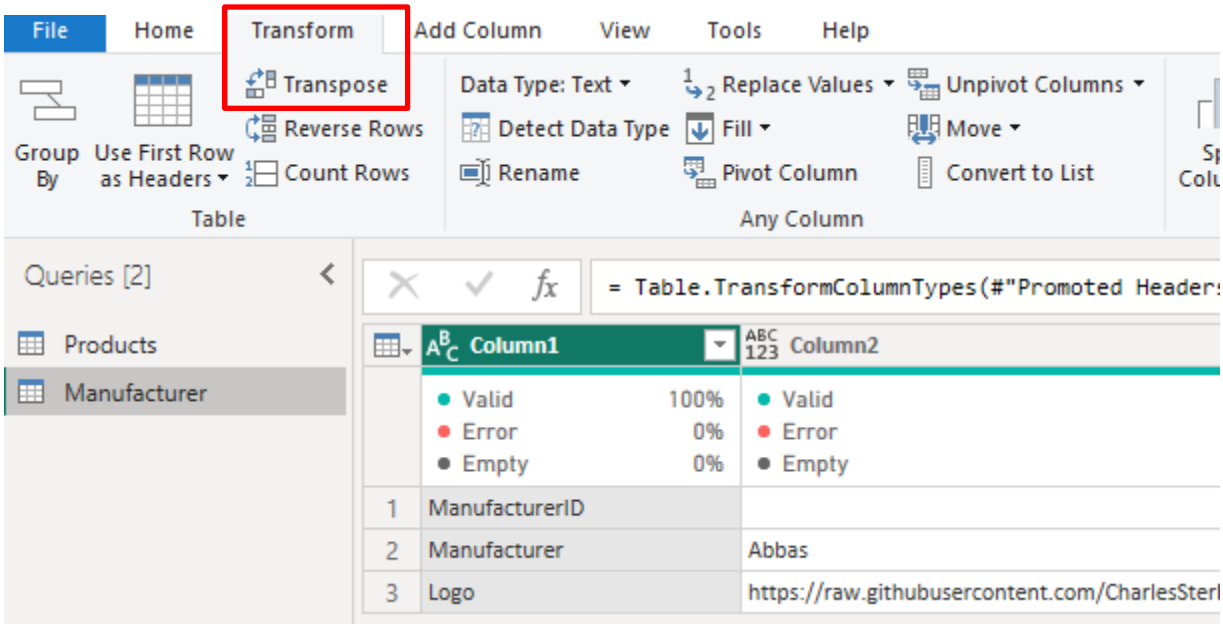
Notice the first row in the **Calendar** query is now the column header. Let's make it a header.

5. With **Calendar** query selected in the left panel, from the ribbon click **Home**, and then click **Use First Row as Headers.**

With that step, Power BI will predict the data type of each field again.

Power BI Desktop – Transposing data

1. From the left panel, click the **Manufacturer** Query. Notice that the **ManufacturerID**, **Manufacturer**, and **Logo** data is laid across in rows. Also notice that the header is not useful. We need to transpose the table to meet our needs.
2. From the ribbon click **Transform** and then click **Transpose**.



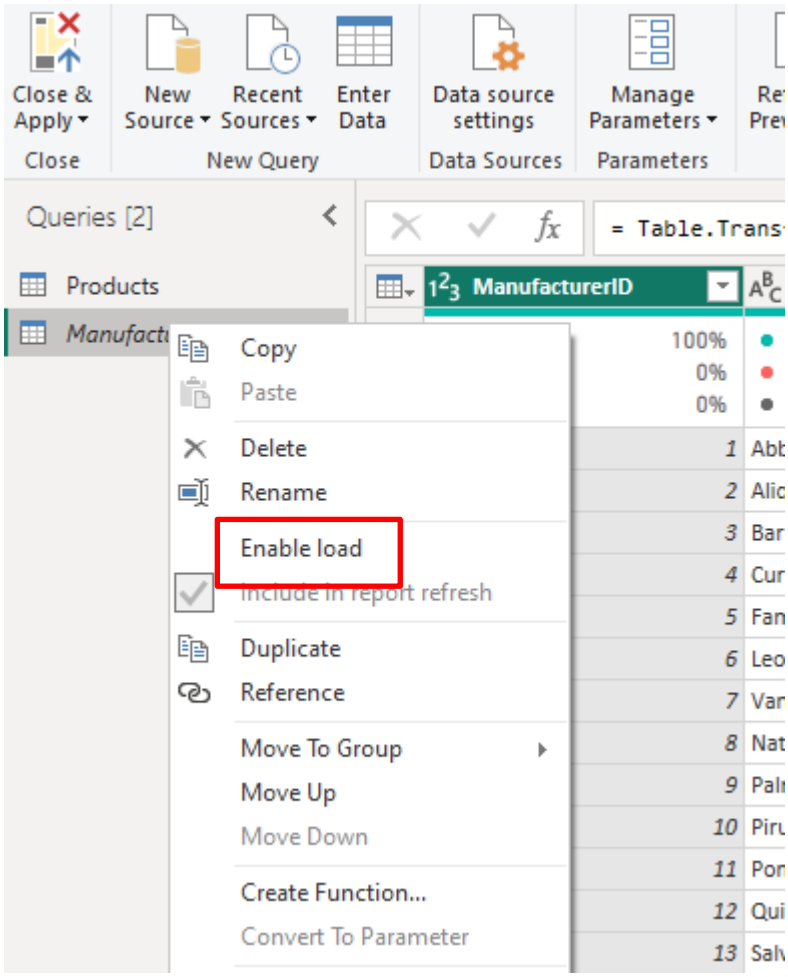
Notice that this transposes the data into columns. Now we need the first row to be the header.

3. From the ribbon click **Home** and then click **Use First Row as Headers**.

Notice that now the **Manufacturer** table is laid out the way we need it with a header and values along columns.

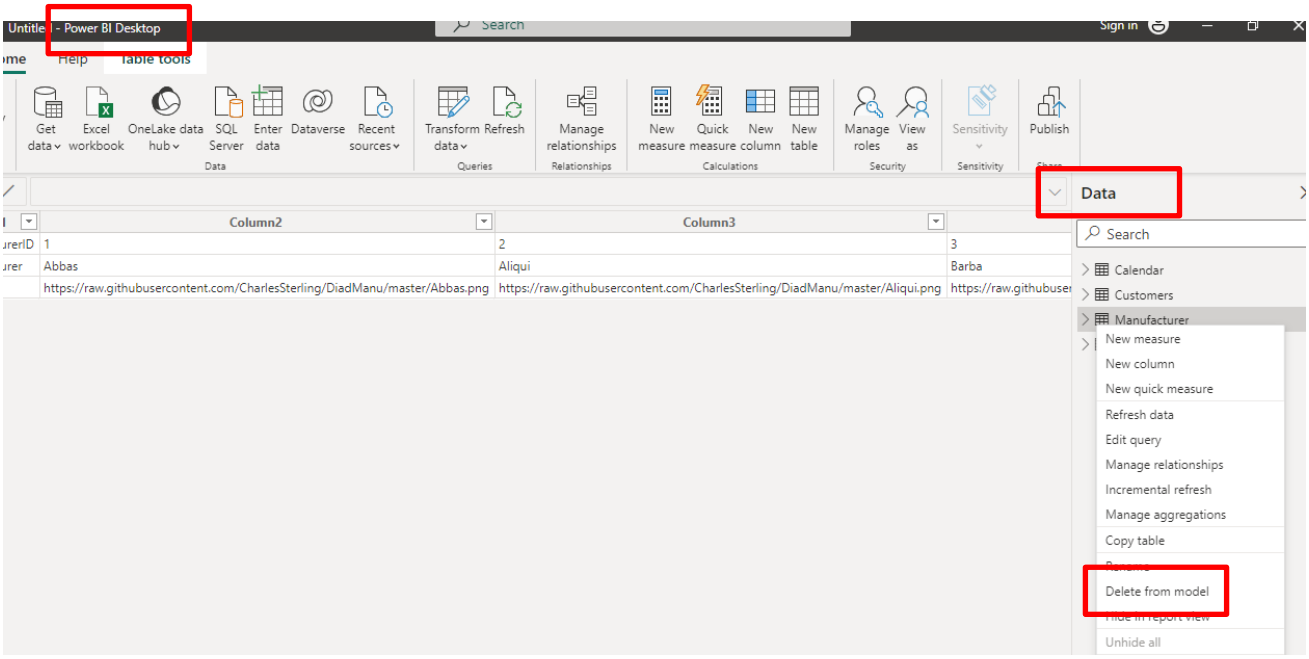
Manufacturer table has no connection with the other tables. So we don't need the **Manufacturer** table to load into the data model. Let's prevent the **Manufacturer** table from loading into the data model.

4. From the Queries panel on the left, click the **Manufacturer** query.
5. Right-click and then click **Enable Load**. This will disable loading **Manufacturer** data.



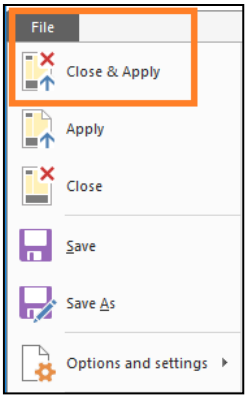
You can remove the tables which are not need for the analysis.

1. Go to the **Power BI Desktop** view.
2. Right click on the table name which needs to be deleted from the right hand side **Data** bar.
3. Select the **Delete from model**.

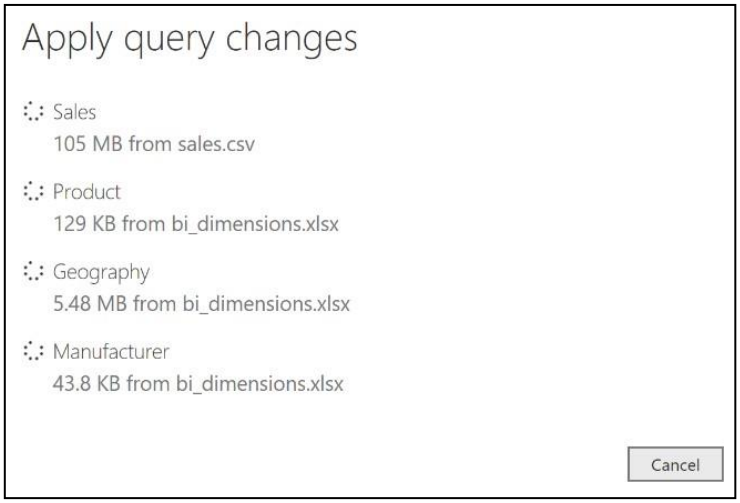


You have now successfully completed import and data shaping operations and are ready to load the data into the Power BI Desktop data model to visualize the data.

4. Click **File** and then click **Close & Apply**. This will close out the power query window and apply all changes



All the data will be loaded in memory in the Power BI Desktop. You will see the progress dialog box with the number of rows being loaded in each table as shown in the above Figure.



Note: It may take several minutes to load all the tables.

5. Click **File** and then click **Save** to save the file after the data loading is complete. Name the file as "**MyFirstPowerBIModel**".

Activity 01

Use the given data and do the following activities

1. Load the L1-Transaction data into PowerBI.
2. Extract the date from the transaction column and rename
3. Extract the time from the transaction column and rename
4. Round off long. And lat. Values to nearest 4 decimal places
5. Convert the country name into Uppercase
6. Add a new column as product category and categorize the products in to A, B and C (A-Product1, B- Product2 and C- Product3).

Activity 02

1. Download FoodMart_Transactions_1997 and FoodMart_Transactions_1998 csv files.
2. Load the data into PowerBI.
3. Load Product_Lookup and Store_Lookup datasets to PowerBI.
4. Remove the source column.
5. Make a separate column for transaction year and display the results.

6. Load Customer_Lookup, data file to PowerBI.
7. Make one column for the customer's name with full name and last name together.
8. If the member card is golden, add a new column and name them as Premium member.
9. Add a new column the categorize the customers and High, Average and Low according to their income.