Power BI Desktop

Power BI Desktop - Accessing Data

In this section, you will import VanArsdel and its competitors USA sales data. Then you import and merging sales data from other countries.

Power BI Desktop - Get Data

Let's start with looking at the data files. The dataset contains sales data of VanArsdel and other competitors. We have 7 years of transaction data by day, product and zip code for each manufacturer. We are going to analyze data from 7 countries.

USA sales data is in a csv file located in /Data/USSales folder.

Sales of all other countries is in /Data/InternationalSales folder. Each countries sales data is in a csv file in this folder.

Product, Geography and Manufacturer information in an excel file in /Data/USSales/bi_dimensions.xlsx.

1. Open **/Data/USSales/bi_dimensions.xlsx.** Notice the first sheet has **Product** information. The sheet has a header and product data is in a named table. Also notice Category column has a bunch of empty cells.

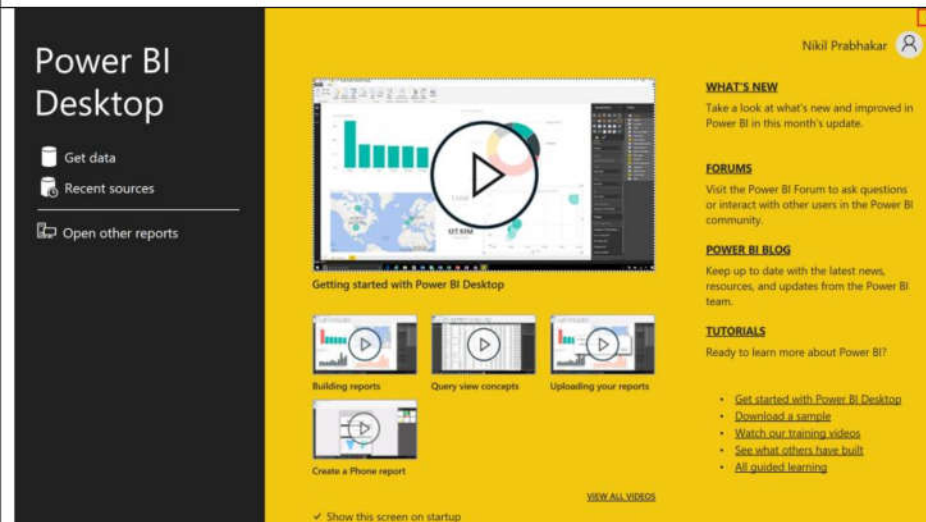|	A	B	C	D	E	F
1	**Source:**	Public Database				
2	**Last Upda**	Monday, February 1, 2016				
3						
4	**Zip**	**City**	**State**	**Region**	**District**	**Country**
5	22654	Star Tannery, VA, USA	VA	East	District #07	USA
6	22655	Stephens City, VA, USA	VA	East	District #07	USA
7	22656	Stephenson, VA, USA	VA	East	District #07	USA
8	22657	Strasburg, VA, USA	VA	East	District #07	USA
9	22660	Toms Brook, VA, USA	VA	East	District #07	USA
10	22663	White Post, VA, USA	VA	East	District #07	USA
11	22664	Woodstock, VA, USA	VA	East	District #07	USA
12	22701	Culpeper, VA, USA	VA	East	District #07	USA
13	22709	Aroda, VA, USA	VA	East	District #07	USA
14	22711	Banco, VA, USA	VA	East	District #07	USA
15	22712	Bealeton, VA, USA	VA	East	District #07	USA

product | manufacturer | **geo**

|	
Manufacturer sheet has data laid out across the sheet and with no column headers and it has a couple of blank rows and a note in row 7. **Geo sheet** has geography information. The first couple of rows has data details. Actual data starts from row 4. We will start by connecting to data from these different files and perform data cleaning and transformation operations.	
2. If you don't have the **Power BI Desktop** open, launch it now. 3. Select **Already have a Power BI Account? Sign in** option. 4. **Sign in** using your Power BI credentials. 5. Startup screen opens. Click on **X** on the top right corner of the dialog to close it.	**Power BI Desktop** Get data Recent sources Open other reports Nikil Prabhakar **WHAT'S NEW** Take a look at what's new and improved in Power BI in this month's update. **FORUMS** Visit the Power BI Forum to ask questions or interact with other users in the Power BI community. **POWER BI BLOG** Keep up to date with the latest news, resources, and updates from the Power BI team. **TUTORIALS** Ready to learn more about Power BI? • Get started with Power BI Desktop • Download a sample • Watch our training videos • See what others have built • All guided learning Getting started with Power BI Desktop Building reports Query view concepts Uploading your reports Create a Phone report VIEW ALL VIDEOS ✓ Show this screen on startup

Let's set up the locale to US English, to make it convenient to go through the rest of this lab.

6. From the ribbon, select **File -> Options and settings -> Options**.
7. In the left panel of Options dialog, select **Regional Settings**.
8. From the **Locale** drop down select **English (United States).**
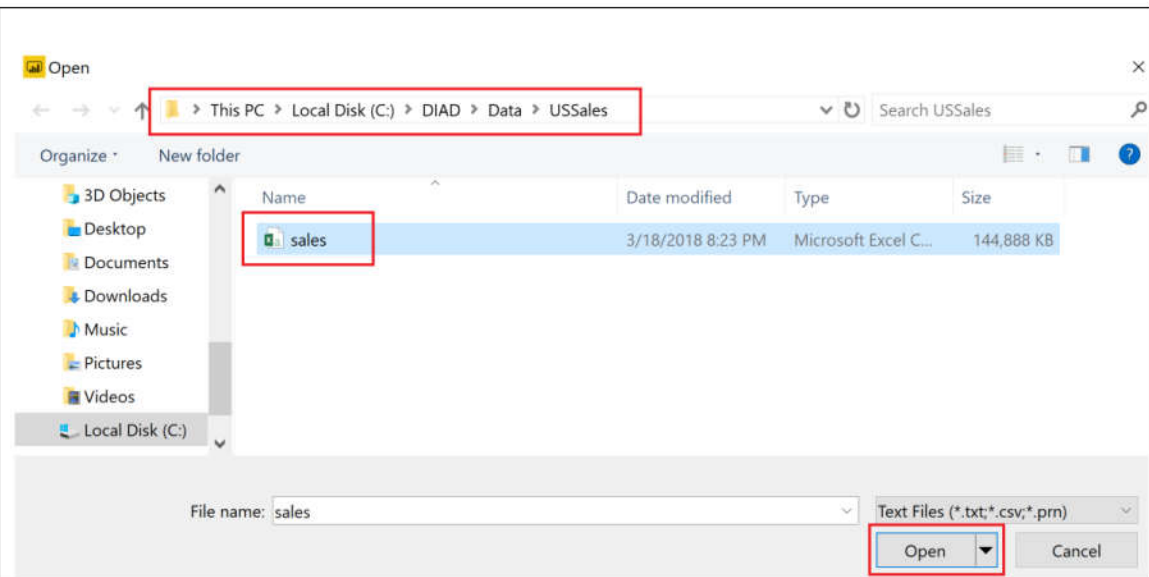9. Select **OK** to close the dialog.
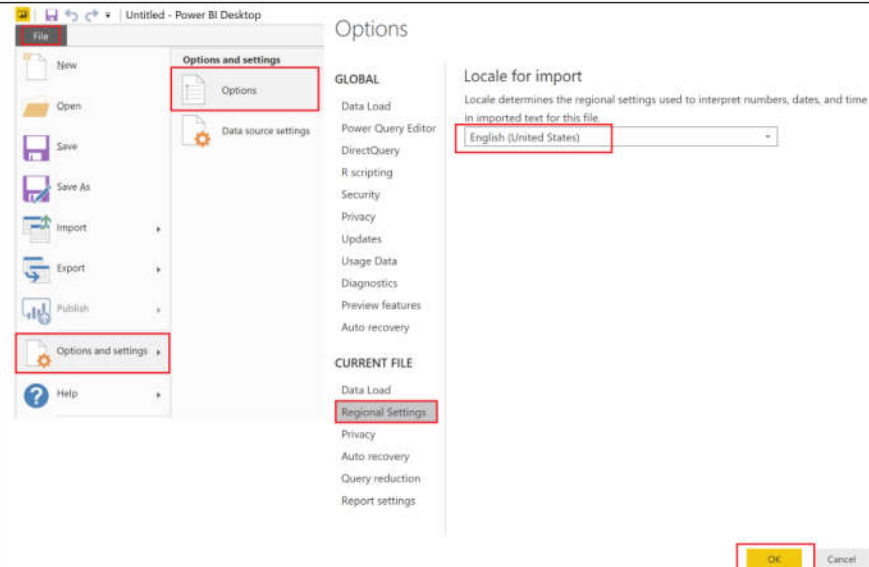
First step is to load data to Power BI Desktop. We will load USA Sales data which is in comma separated value(CSV) files.

10. From the ribbon, select **Home -> Get Data.**
11. Select **Text/CSV.**

Note: Power BI Desktop has the capability to connect to 70+ data sources. We are using csv and excel data files in this lab for simplicity.
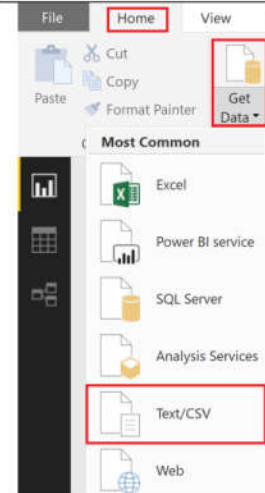
12. Browse to **DIAD\Data\USSales** folder and select **sales.csv**. 13. Click **Open**.	
Power BI detects the data type of each column. There are options to detect the data type based on the first 200 rows or based on the entire dataset or not detect it. Since our dataset is large and it will take time and resources to scan the complete data set, let's leave the default option of selecting dataset based on the first 200 rows. After completing your selection, you have three options – Load, Edit or Cancel. • **Load**, loads the data from the source into Power BI Desktop for you to start creating reports. • **Edit** 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.	
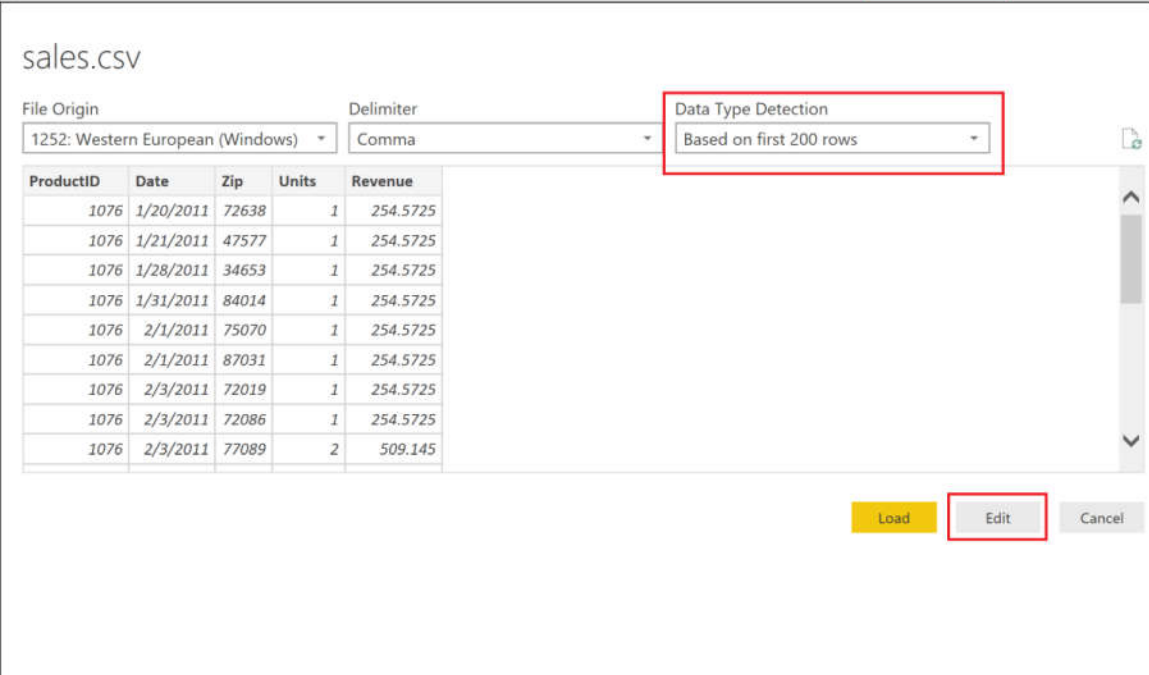
|	
• **Cancel** gets you back to the main canvas. 14. Click **Edit** as shown in the screenshot. A new window opens.	
You should be in the Query Editor window as shown in the screenshot to the right. Query Editor is used to perform data shaping operations. Notice the sales file you connected to shows as a query in the left panel. You see a preview of the data in the center panel. Power BI predicts data type of each field (based on the first 200 rows) which is indicated next to the column header. In the right panel, steps that Query Editor performs are recorded. **Note**: You will be bringing in sales data from other countries as well as performing certain data shaping operations. 15. Notice Power BI has set Zip field to data type Whole Number. To ensure that Zip codes which start with zero don't lose the leading zero, we will format them as text. Highlight the **Zip column**. From the ribbon, select **Home -> Data Type** and update it to **Text**. 16. **Change Column Type** dialog opens. Select **Replace Current** button which overwrites Power BI's predicted datatype.	
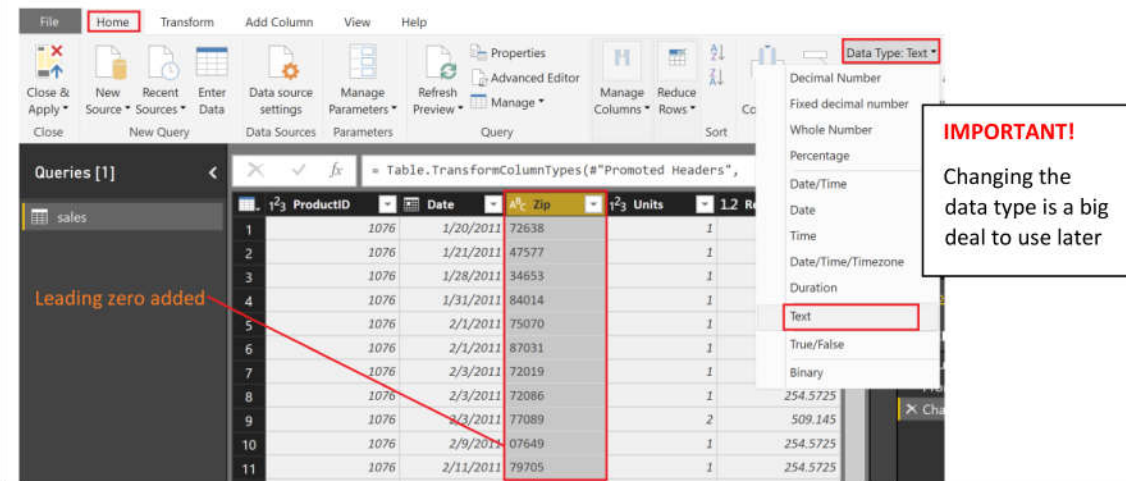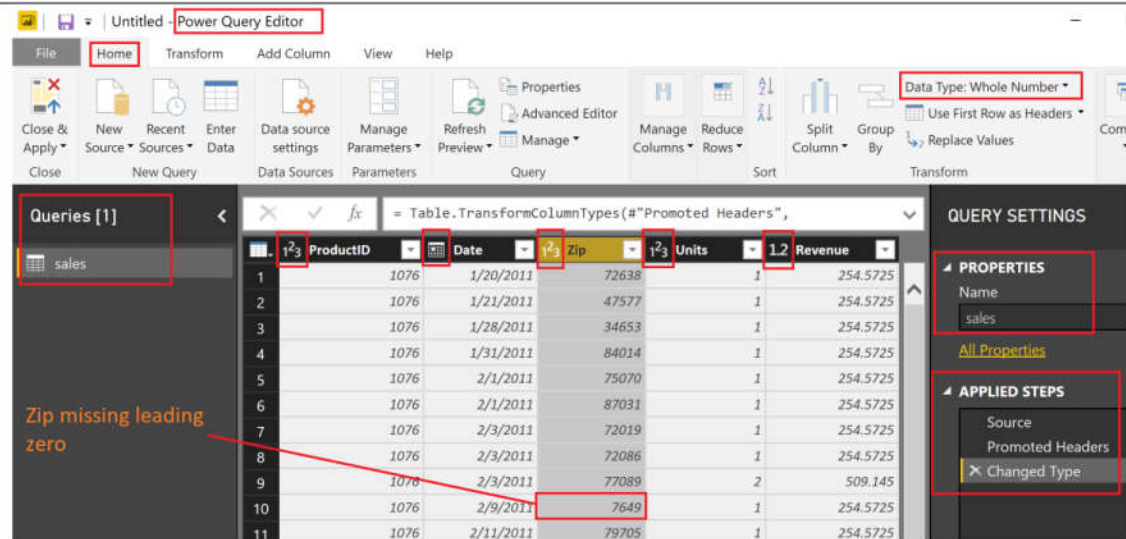
Now let's get the data that is in excel source file. 17. From the ribbon, select **Home -> New Source -> Excel.**	
18. Browse to **DIAD\Data\USSales** folder and select **bi_dimensions.xlsx**. Navigator dialog opens.	
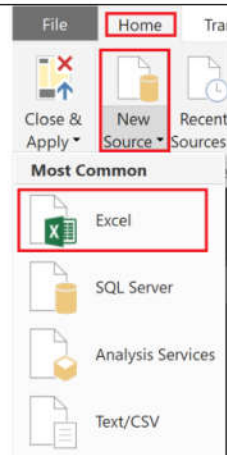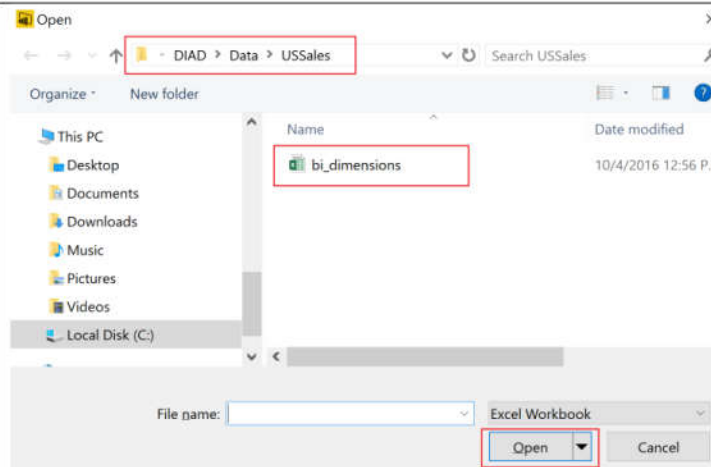
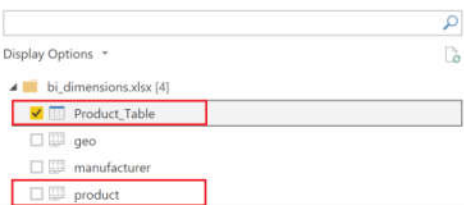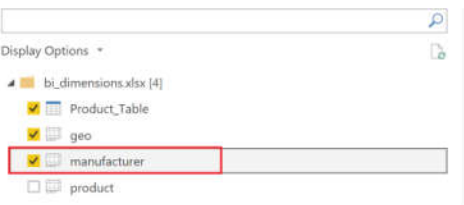|	
19. Navigator dialog lists 3 sheets that are in the excel workbook. It also lists the Product named table. **Select product** from the left panel and in preview panel notice the first row is the header. This is not part of the data. 20. **Unselect product** from the left panel. **Select Product_Table**. Notice this has only the contents of the named table. This is the data we need. **Note**: Table names are differentiated from Worksheet names by using different icons.	Navigator product 	Column1	Column2	Column3	Column4	Column5
---	---	---	---	---
Product Details		null	null	null	null
ProductID	Product	Category	ManufacturerID	Price
	1 Abbas MA-01|All Season	Mix		1 USD 412.13
	2 Abbas MA-02|All Season	null		1 USD 329.78
	3 Abbas MA-03|All Season	null		1 USD 963.38
	4 Abbas MA-04|All Season	null		1 USD 828.98 | Navigator Product_Table | ProductID	Product	Category	ManufacturerID	Price
---	---	---	---	---
1	Abbas MA-01|All Season	Mix	1	USD 412.13
2	Abbas MA-02|All Season	null	1	USD 329.78
3	Abbas MA-03|All Season	null	1	USD 963.38
4	Abbas MA-04|All Season	null	1	USD 828.98
5	Abbas MA-05|All Season	null	1	USD 745.5
7	Abbas MA-07|All Season	null	1	USD 451.45	
21. From the left panel, **select geo**. In the preview panel notice the first couple of rows are headers that are not part of the data. We will remove them shortly. 22. From the left panel, **select manufacturer**. In the preview panel notice the last couple of rows are footers that are not part of the data. We will remove them shortly. 23. Select **OK**. (Make sure Product_Table, geo and manufacturer are selected in the left panel) Notice all 3 sheets are added as queries in the Query Editor.	Navigator geo 	Column1	Column2	Column3	Column4	Column5	Column6
---	---	---	---	---	---
Source:	Public Database	null	null	null	nu
Last Updated:	2/1/2016	null	null	null	nu
null	null	null	null	null	nu
Zip	City	State	Region	District	Country
22654	Star Tannery, VA, USA	VA	East	District #07	USA
22655	Stephens City, VA, USA	VA	East	District #07	USA | Navigator manufacturer | Column1	Column2	Co
---	---	---
ManufacturerID		1
Manufacturer	Abbas	Ali
Logo	https://raw.githubusercontent.com/CharlesSterling/DiadManu/master/AI	hti
	null	null
	null	null
List of Suppliers and Manufacturers		null	

Power BI Desktop - Adding additional data

International subsidiaries have agreed to provide their sales data so that the company's sales can be analyzed together. You've created a folder where they will each put their data.

Power BI Desktop – Using Fill feature

Some of the data provided is not in the right format. Power BI provides extensive transformation capabilities to clean and prepare the data to meet our needs. Let's start with Product query. Notice that Category column has a lot of null values. Looks like there is are values in Category column only when the value changes. We need to fill it down to have values in each row.

| 4. From the left panel, select **Product** Query.
5. Select **Category** column.
6. From the ribbon select **Transform -> Fill -> Down**.

Notice now all the null values are filled with the appropriate Category values. | |

Version: 04.10.2018   Copyright 2018 Microsoft   **20** | P a g e

Maintained by: Microsoft Corporation

In Product query, notice that the Product column. Looks like two fields are concatenated into one field with a pipe (|) separator. Let's split them into two columns. This will be useful when we build visuals, so we can analyze based on both fields.

7. From the left panel, select **Product** Query.

8. Select **Product** column.

9. From the ribbon select **Home -> Split Column -> By Delimiter**. Split Column by Delimiter dialog opens.

10. In the dialog, make sure **Custom** is selected in the **Select or enter delimiter** dropdown.

Note: Select or enter delimiter dropdown has some of the standard delimiters like comma, colon, etc.

11. Notice in the text area, there is a hyphen (-). Power BI assumes we want to split by hypen. **Remove hyphen** symbol and **enter pipe symbol (|)** as shown in the screenshot.

12. Select **OK**.

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

In this scenario delimiter occurs only once, hence Product column is split into 2 columns.
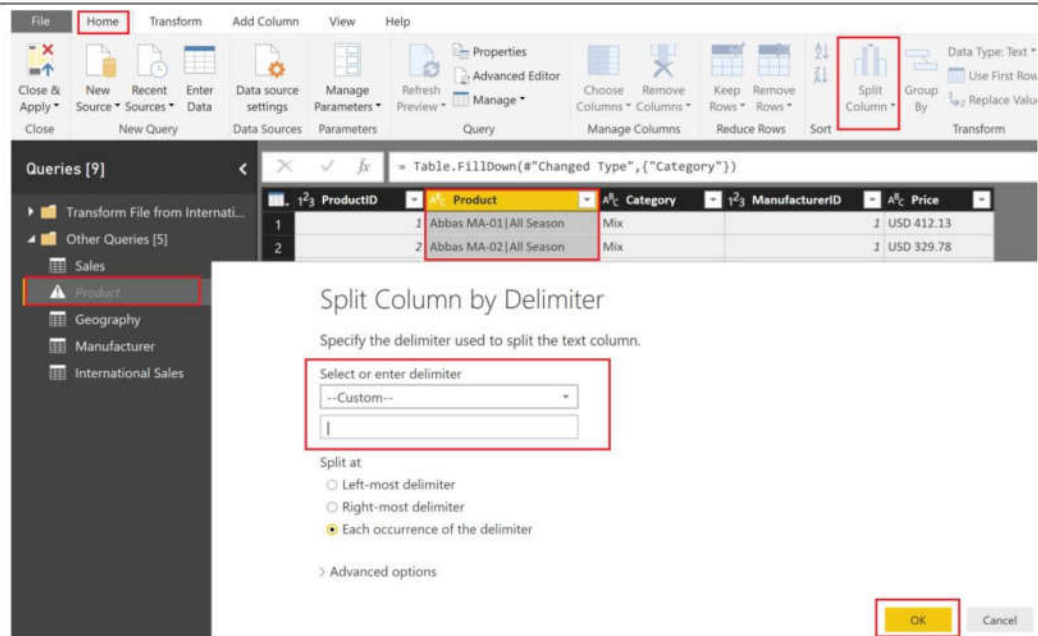
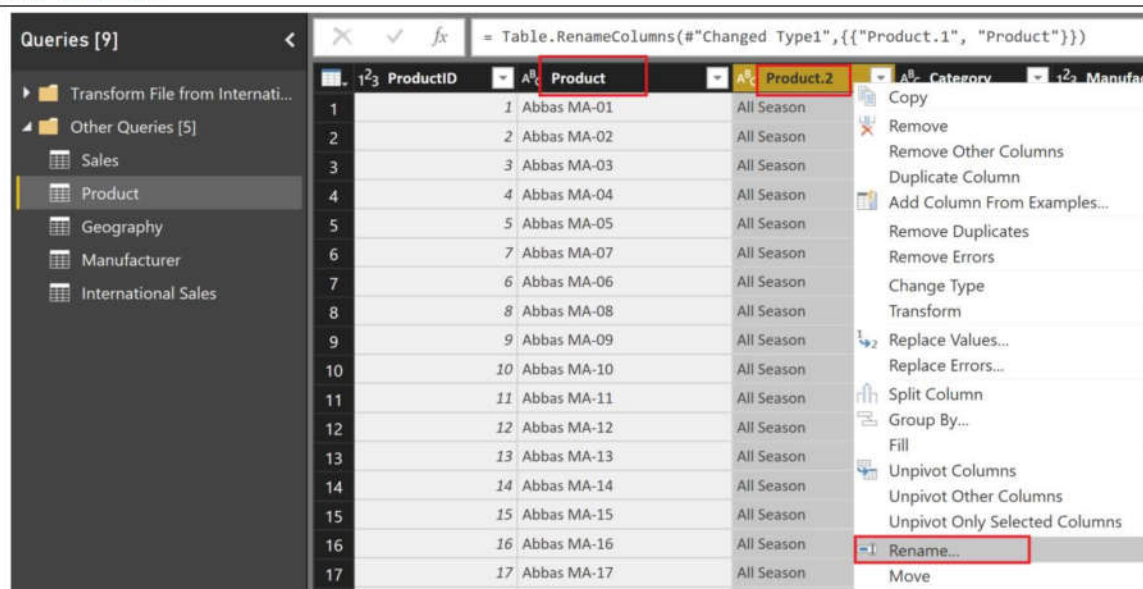## Power BI Desktop – Using Rename Column feature

Let's rename the columns.

13. Select **Product.1** column. **Right click** next to the column name.

14. Select **Rename** from the selection dialog.

15. **Rename** the field to **Product**.

16. Similarly rename **Product.2** to **Segment**.

## Power BI Desktop – Using Column From Examples feature

In Product query, notice that the Price column. You will see price and currency concatenated into one field. To do any calculations we just need the numeric value. It will be good to split this field into two columns. We can use the split feature like earlier or we can use Column From Examples. Column From Examples is handy in scenarios where the pattern is more complex than a delimiter.

17. From the left panel, select **Product** Query.

18. From the ribbon, select **Add Column -> Column From Examples**.

19. In the **first row of Column1** enter the first Price value which is **412.13** and click enter Notice as you enter, Power BI knows that you want to split Price column. The formula it uses is displayed as well.

20. **Double click** column header **Text After Delimiter** to rename it.

21. **Rename** the column to **MSRP**.

22. Click **OK** to apply the changes.
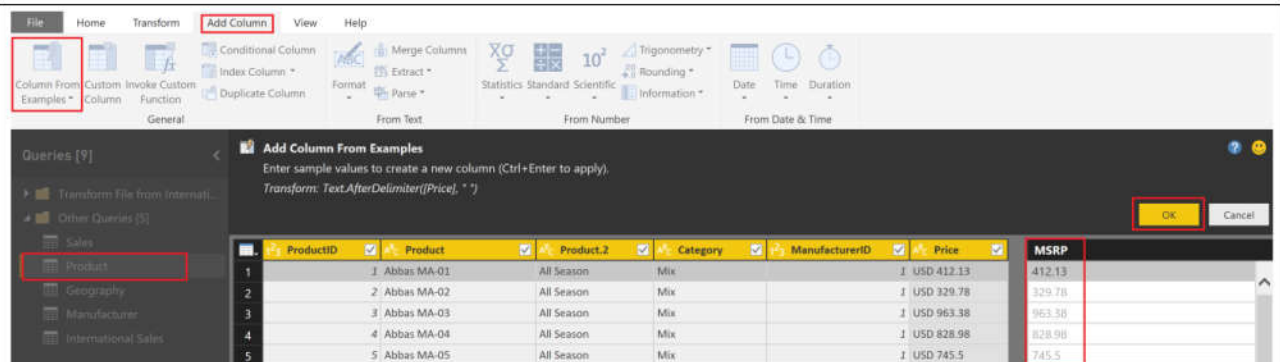
Notice MSRP field is of data type text. It must be a decimal. Let's change it.

23. Select **ABC** in **MSRP** column.

24. From the selection dialog, select **Decimal Number**.

Notice all the steps we performed on the Product query are being recorded under **APPLIED STEPS** in the right panel.
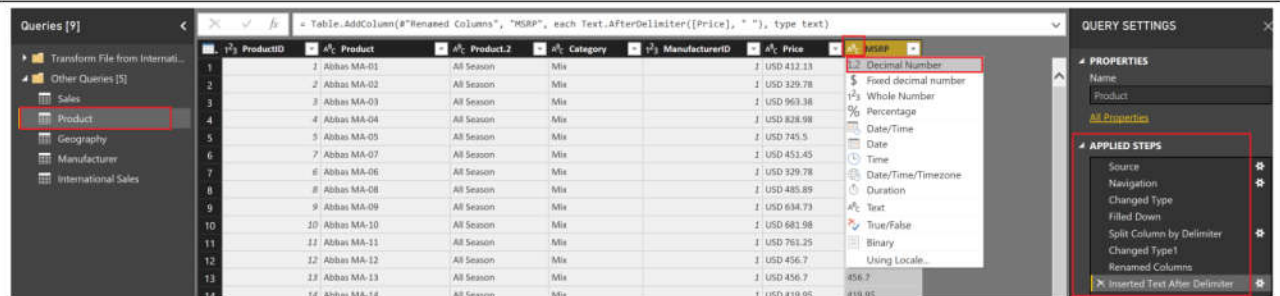
Similarly, let's create a currency column.

25. From the left panel, select **Product** Query.

26. From the ribbon, select **Add Column ->
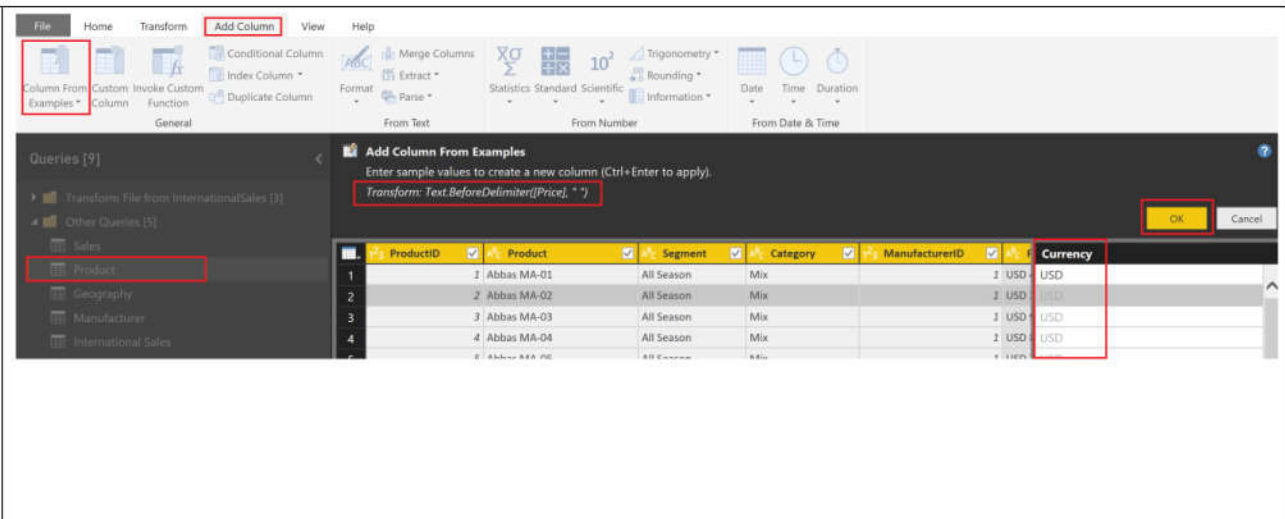Column From Examples**.

27. In the **first row of Column1** enter the first
Currency value as **USD** and click enter
Notice as you enter, Power BI knows that you
want to split Price column. The formula it uses
is displayed as well.

28. **Double click** column header **Text After
Delimiter** to rename it.

29. **Rename** the column to **Currency**.

30. Click **OK** to apply the changes.

Now that we have split Price into MSRP and
Currency columns, we don't need Price
column. Let's remove it.

31. From the left panel, select **Product** Query.

32. **Right click** next to **Price** column.

33. Select **Remove**.

Power BI Desktop – Using Add/Remove Rows feature

In Geography query, notice that first two rows are informational. It is not part of the data. Similarly, in Manufacturer query the last couple of
rows are not part of the data. Let's remove them so we have a clean dataset.
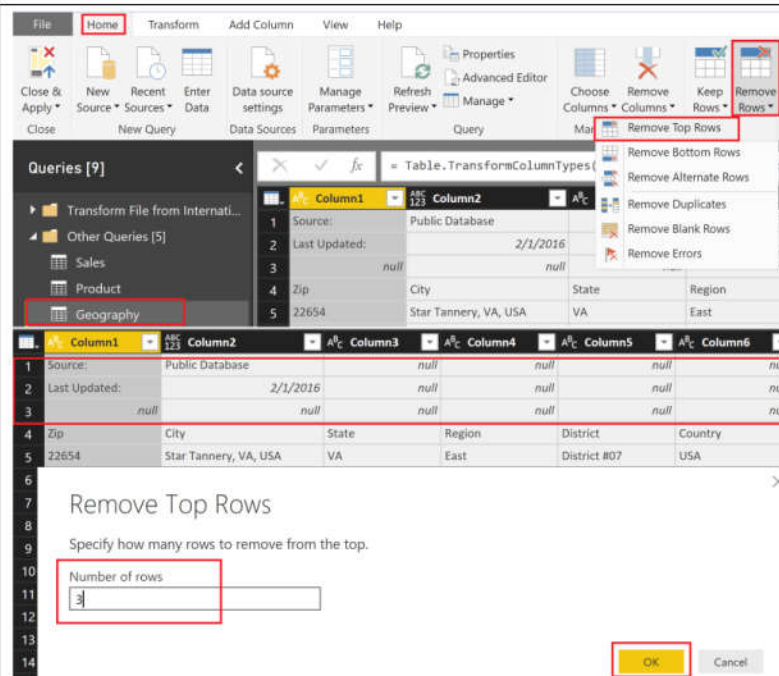
34. In the left panel, select **Geography** query.

35. From the ribbon, select **Home -> Remove Rows -> Remove Top Rows**.

36. Remove Top Rows dialog opens. Enter **3** in the text box, since we want to remove the top 2 informational data rows and the blank 3rd row.

37. Select **OK**.

Notice the first row in Geography query now is the column header. So let's make it a header

38. With **Geography** query selected in the left panel, from the ribbon select **Home -> Use First Row as Headers**

Notice column Zip is of data type number. Let's change it to text as we did earlier. If we don't we will see errors when we load the data.

39. Select **123** next to Zip Column. From the dialog, select **Text**.

40. Select **Replace Current** in the **Change Column Type** dialog.
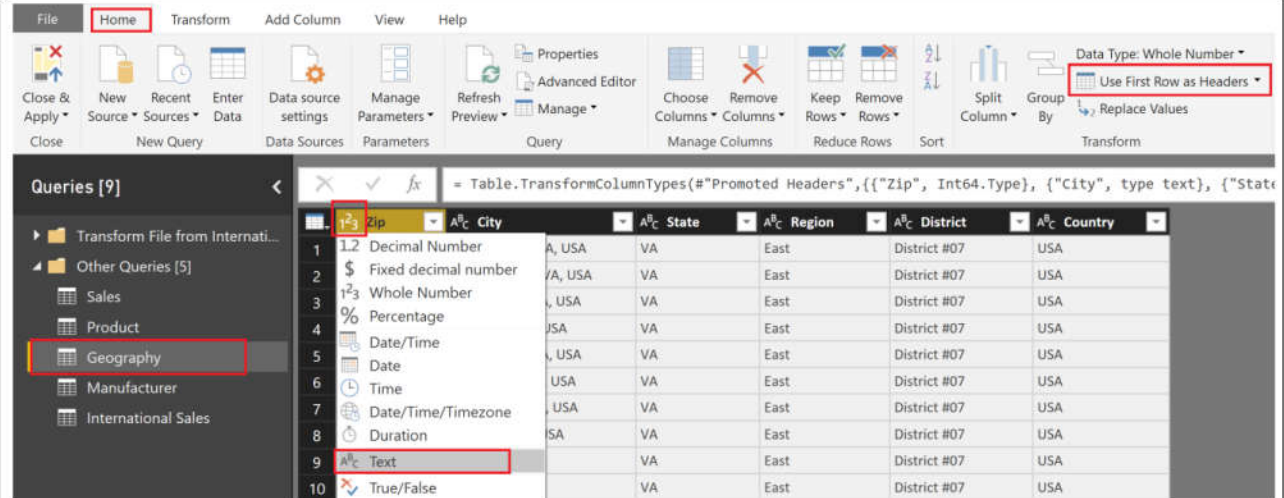
41. From the left panel, select **Manufacturer** query. Notice the bottom 3 rows are not part of the data. Let's remove it

42. From the ribbon, select **Home -> Remove Rows -> Remove Bottom Rows**

43. Remove Bottom Rows dialog opens. Enter **3** in **Number of rows text box**.

44. Select **OK**.
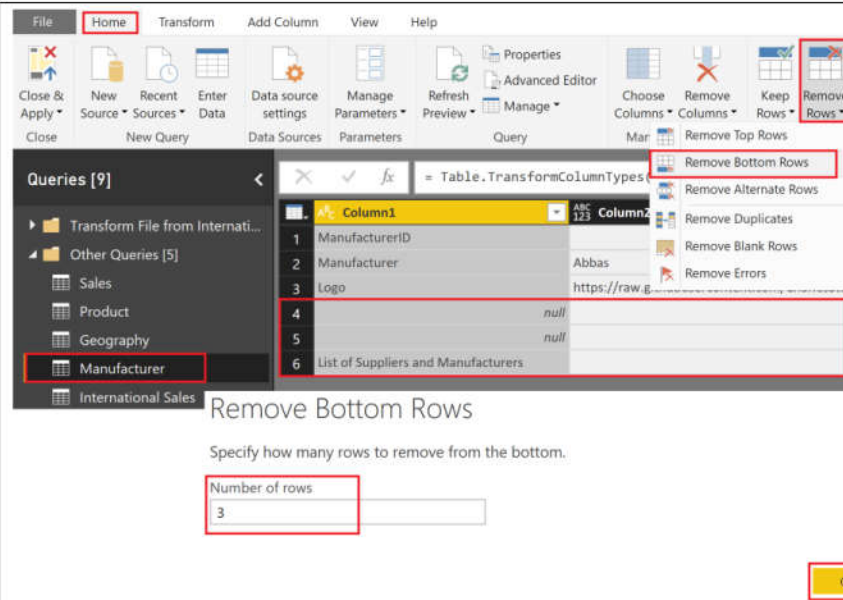
Power BI Desktop – Using Transpose feature

45. From the left panel, select **Manufacturer** Query. Notice ManufacturerID, Manufacturer and Logo data is laid across in rows. And the header is not useful. We need to transpose the table to meet our needs.

46. From the ribbon select **Transform -> Transpose**.

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