Mastering Data Analytics in Excel

Get Data into Excel

Get Data into Excel

Lab story

This lab is designed to introduce you to Power Query in Excel application and how to connect to data and how to use data preview techniques to understand the characteristics and quality of the source data.

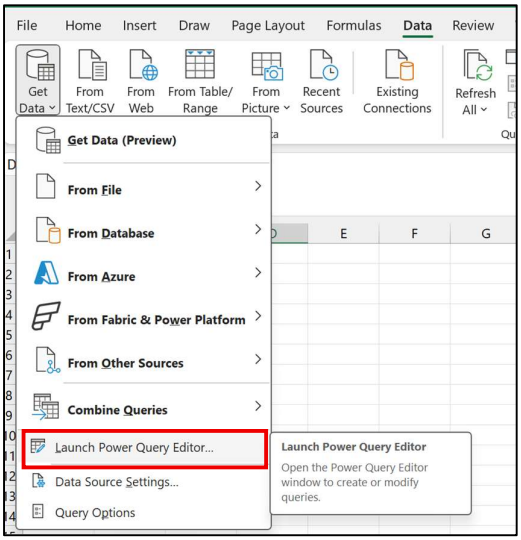
Objectives:

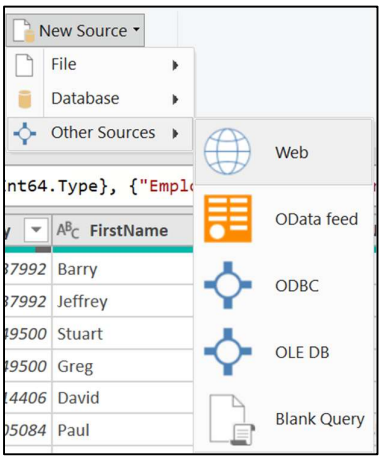
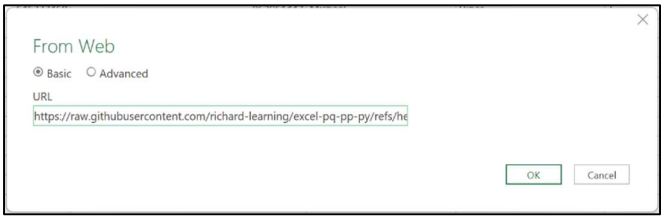
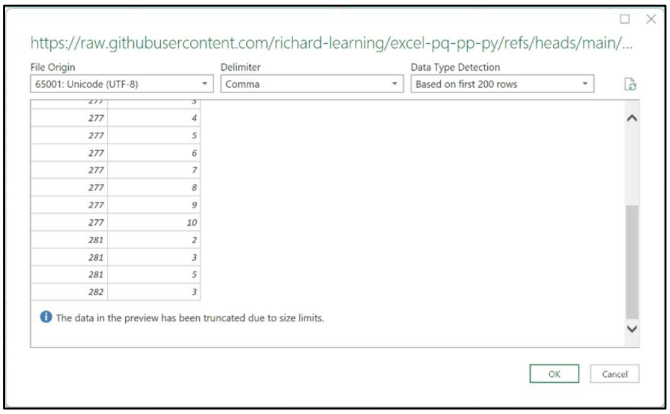
In this lab, you learn how to:

- Open Power Query in Excel.
- Connect to different data sources.
- Preview source data with Power Query.
- Use data profiling features in Power Query.

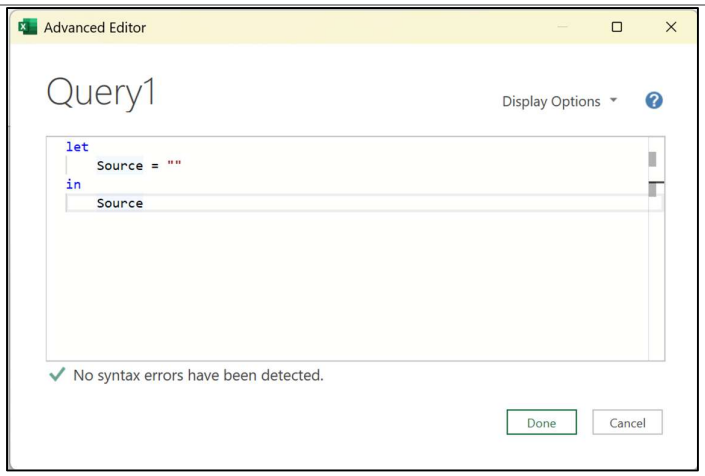
This lab should take approximately 30 minutes.

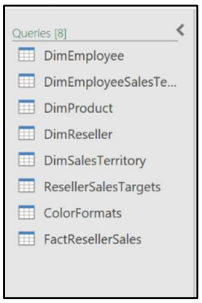
Get started with Power Query

1	To complete this exercise, first open a web browser and enter the following URL to download the zip folder: `https://github.com/richard-learning/excel-pq-pp-py/raw/refs/heads/main/01-get-data-in-excel/01-get-data-in-excel.zip`
2	Extract the folder to the Download folder.
3	Open the **01-Starter-Sales Analysis.xlsx** file.
4	Launch the Power Query editor. 
5	With the Power Query editor open you will notice the **DimEmployee** already queried.
6	Explore the Query Settings; this is where you can track all transformation done.

Get data from Web	
1	This task teaches you how to get Data from the web and import tables, which create queries in Power Query.
2	On the **Home** ribbon tab, from inside the **New Query** group, select **New Source**. 
3	A dialog box will ask for a URL. Enter the following and click **OK**: https://raw.githubusercontent.com/richard-learning/excel-pq-pp-py/refs/heads/main/00-data/DimEmployeeSalesTerritory.csv 
4	A preview of the data to be import will be shown. Click **OK**. 
5	Repeat steps 2 to 4 for the following tables. DimProduct https://raw.githubusercontent.com/richard-learning/excel-pq-pp-py/refs/heads/main/00-data/DimProduct.csv

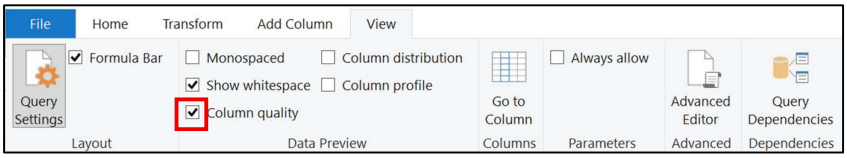
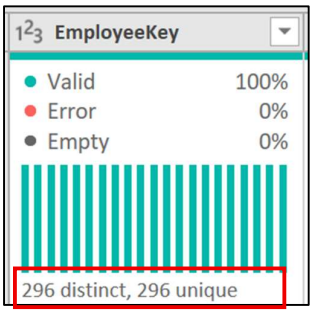
	DimReseller https://raw.githubusercontent.com/richard-learning/excel-pq-pp-py/refs/heads/main/00-data/DimReseller.csv DimSalesTerritory https://raw.githubusercontent.com/richard-learning/excel-pq-pp-py/refs/heads/main/00-data/DimSalesTerritory.csv
6	You've now connected to four tables from a Web source.
Get data from a CSV file	
1	In this task, you'll create a new query based on CSV files.
2	To add a new query, in the **Power Query Editor** window, on the **Home** ribbon tab, from inside the **New Query** group, select the **New Source** down-arrow, and then select **File>Text/CSV**.
3	Navigate to the **Downloads > 01-get-data-in-excel** folder you extracted earlier and select the **ResellerSalesTargets.csv** file. Select **Open**
4	In the **ResellerSalesTargets.csv** window, review the preview data. Select **OK**
5	In the **Queries** pane, notice the addition of the **ResellerSalesTargets** query
6	Repeat the steps to create a query based on the **ColorFormats.csv** file.
7	You should now have two new queries, **ResellerSalesTargets** and **ColorFormats**.
Get Data using M code	
1	In this task, you'll create a new query using M code. M code is the language that is used in Power Query for loading and transformation
2	To add a new query, in the **Power Query Editor** window, on the **Home** ribbon tab, from inside the **New Query** group, select the **New Source** down-arrow, and then select **File>Other Source/Blank Query**.
3	In the Properties section on the right side, change the Name to **FactResellerSales**.
4	In the **Power Query Editor** window, on the **Home** ribbon tab, from inside the **Query** group, select **Advance Editor**.

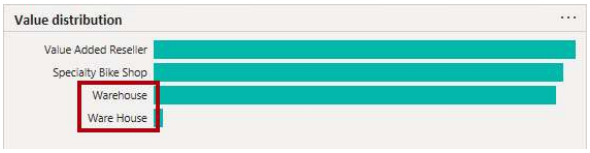


5	Navigate to the **Downloads > 01-get-data-in-excel** folder you extracted earlier and open the **snippet.txt** file
6	Copy the text under the header **FactResellerSales M Code** into the advance editor and click **Done**.
7	You should now have all eight queries loaded. 

Preview Data in Power Query Editor

1	This task introduces the Power Query Editor and allows you to review and profile the data. This helps you determine how to clean and transform the data later. You'll also review both dimension tables prefixed with "Dim" and fact tables prefixed with "Fact".
2	In the **Power Query Editor** window, at the left, notice the **Queries** pane. The **Queries** pane contains one query for each table you checked
3	Select the **DimEmployee** query
4	At the bottom left corner of the status bar, some table statistics are provided—the table has 27 columns, and 296 rows. 

5	To assess column quality, on the **View** ribbon tab, from inside the **Data Preview** group, check **Column Quality**. The column quality feature allows you to easily determine the percentage of valid, error, or empty values found in columns. 
6	Notice that the **Position** column has 94% empty (null) rows. 
7	To assess column distribution, on the **View** ribbon tab, from inside the **Data Preview** group, check **Column Distribution**
8	Review the **Position** column again, and notice that there are four distinct values, and one unique value
9	Review the column distribution for the **EmployeeKey** column—there are 296 distinct values, and 296 unique values 
Note: When the distinct and unique counts are the same, it means the column contains unique values. When modelling, it's important that some model tables have unique columns. These unique columns can be used to create one-to-many relationships	
10	In the **Queries** pane, select the **DimReseller** query
11	To view column values, on the **View** ribbon tab, from inside the **Data Preview** group, check **Column Profile**

12	Select the **BusinessType** column header, and notice the new pane beneath the data preview pane. Review the column statistics and value distribution in the data preview pane 
13	Hover the cursor over the **Ware House** bar, and notice that there are five rows with this value
14	In the **Queries** pane, select the **FactResellerSales** query
15	Review the column quality for the **TotalProductCost** column, and notice that 8% of the rows are empty
Lab complete	
1	You may choose to save your excel file, though it's not necessary for this lab. In the next exercise, you'll work with a pre-made starter file.