Extracting Data from SharePoint Online using a Function Query

Lab Time: 45-60 minutes

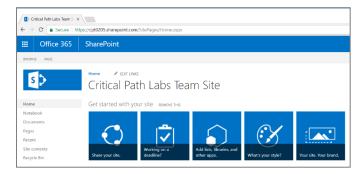
Lab Folder: C:\Student\Modules\02_Queries\Lab\

Lab Overview: In this lab you will begin by designing a query to extract expense data from an unstructured text file in a SharePoint Online document library. You will then convert your query into a function query to extract data from multiple files into a single table.

Exercise 1: Build a Query to Extract Data from an Unstructured Text File

In the following exercise, you will use the Query Editor window to design an advanced query using a query function.

- 1. Navigate to the SharePoint team site at the root of your SharePoint tenancy.
 - a) The SharePoint site should have a URL in the form of https://[Your tenant name].sharepoint.com.

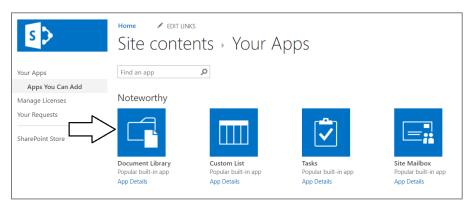


You can use any SharePoint Online team site for this exercise as long as you have permissions in the site to create new lists.

- 2. Create a new document library named **Data** to store the files with expense data.
 - a) Drop down the Site Actions menu and select Add an app.



b) Select **Document Library** as the type of list to create.



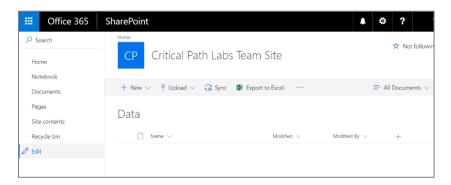
c) In the Adding Document Library dialog, add a Name of Data and click Create.



d) Once the **Data** document library has been created, navigate to its default view.



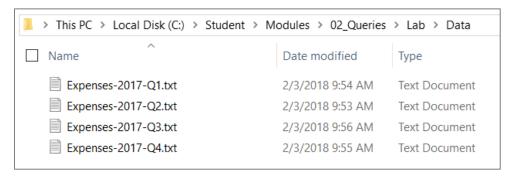
e) You should now be at the default view for the Data document library.



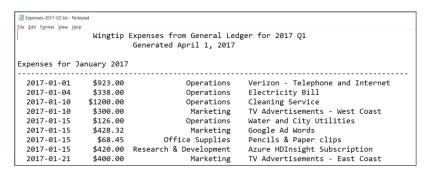
- 3. Upload data files to the Data document library.
 - a) Using Windows Explorer, examine the data files into the follow folder.

C:\Student\Modules\02_Queries\Lab\Data

b) You should see the following four files as shown in the following screenshot.



c) Double-click on the file named Expenses-2017-Q1.txt. to open it in Windows Notepad and inspect its contents.

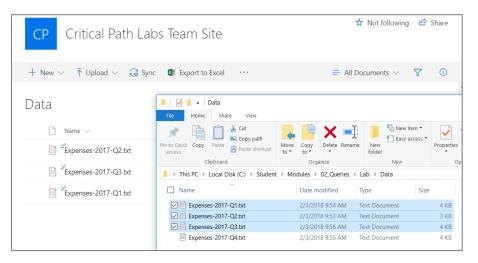


As you can see, this file is an unstructured text file with fixed-width lines which contain expense data. The other three files in the Data folder have expense data for different time periods, but the format of their contents is the same.

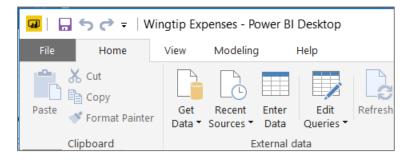
- d) Upload the following three files to the **Data** document library.
 - i) Expenses-2017-Q1.txt
 - ii) Expenses-2017-Q2.txt
 - iii) Expenses-2017-Q3.txt

Note that you should NOT upload the fourth file named Expenses-2017-Q4.txt. You will upload the last file later in this lab.

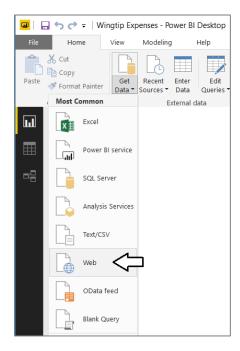
e) You should be able to verify that those three files have been upload to the Data document library.



- 4. Create a new Power BI Desktop project named Wingtip Expenses.pbix.
 - a) Launch Power BI Desktop.
 - b) Begin by saving the new project and give it a name of **Wingtip Expenses.pbix**.



- 5. Create a new query to import data from the file in the Data document library named Expenses-2017-Q1.txt.
 - a) Drop down the Get Data menu and select the Web command.



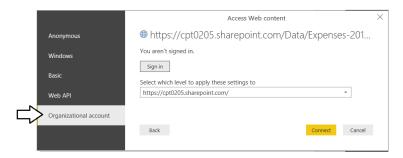
- b) In the **From Web** dialog, add the path to the file named **Expenses-2017-Q1.txt**. The path should include the base URL of your SharePoint site along with the relative file path which is **/Data/Expenses-2017-Q1.txt**.
- c) Your path should look something like the following URL

https://cpt0205.sharepoint.com/Data/Expenses-2017-Q1.txt

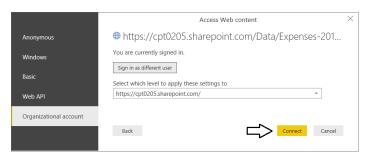
d) Once you have added the file path to the **From Web** dialog, click **OK**.



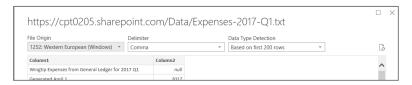
e) If you are prompted to login with the Access Web content dialog, select Organizational account and click Sign in.



f) Once you have signed in on the Access Web content dialog, click the Connect button



g) You should now be prompted with the dialog shown in the following screenshot.



h) Set the Delimiter dropdown menu option to --Fixed Width--.



i) In the textbox under the **Delimiter** dropdown menu, add the following column positions.

0,12,24,48

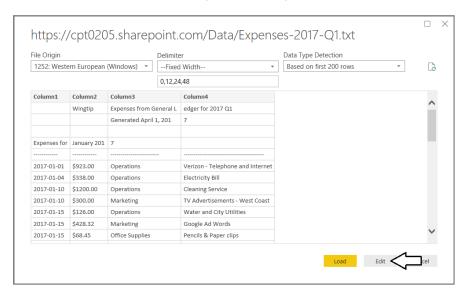
j) Make sure your fixed-width column positions match the following screenshot.



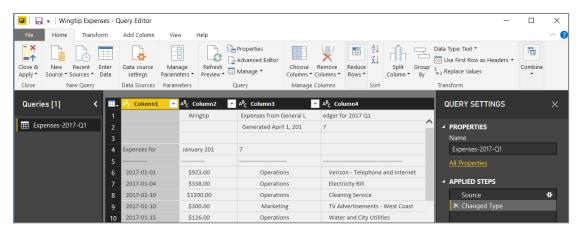
k) At this point, you should be able to see data in some of the rows conforming to the fixed-width column scheme.



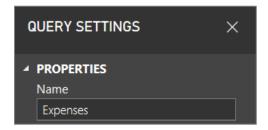
) Click the Edit button at the bottom right of the dialog to open the new query in the Query Editor window.



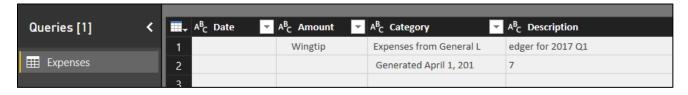
m) Your new query should appear in the Query Editor window.



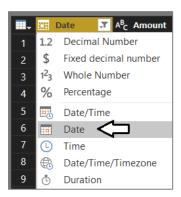
n) In the QUERY SETTING pane in the upper right, modify the name of the query to Expenses.



o) Update the 4 column names in the query to Date, Amount, Category and Description as shown in the following screenshot.



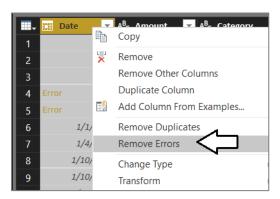
p) Use the Data type dropdown menu at the top left of the Date column to change its type to Date.



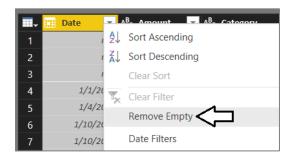
q) You should observe that rows that contain valid data are able to convert their **Date** column values to actual dates while the rows that do not contain valid data show either errors or null values in the **Date** column.



r) Right-click on the **Date** column header and select the **Remove Errors** command.



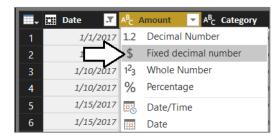
s) Click the dropdown menu on the right-side of the **Date** column and select the **Remove Empty** command.



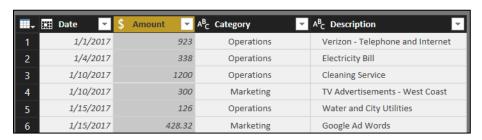
t) The query output should now only contain rows with a valid date value in the **Date** column.



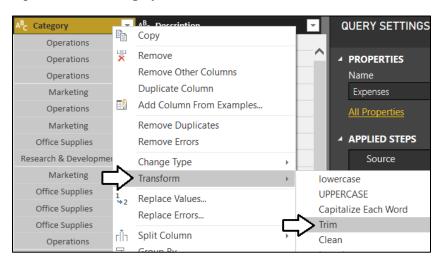
u) Modify the data type of the **Amount** column to be **Fixed decimal number**.



v) The data type for the Amount column should now show a dollar sign to indicate its type is Fixed decimal number.



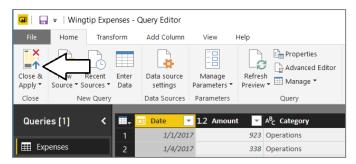
w) Right click on the Category column header and select the Transform > Trim command.



- x) Right click on the **Description** column header and select the **Transform > Trim** command.
- y) The Category and the Description columns should no longer have any extra whitespace.



z) You are done designing this query. Click the **Close and Apply** button to run the query and close the Query Editor window.



This would be a good time to save the work you have completed so far.

Save your work in the Wingtip Expenses project by clicking the Save icon in the upper left of the Power BI Desktop window.



- 7. Create a report to display expense data.
 - a) In the Power BI Desktop application window, navigate to report design view.



b) Add a new Clustered column chart visual.



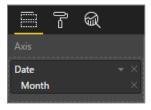
c) Using the mouse, reposition the column chart visual so it takes up the entire page in the report.



d) Click on the checkbox for the **Date** column in the **FIELDS** list on the right. When you do this, you should see that a date hierarchy is automatically added to the **Axis** well.



e) Remove Year, Quarter and Day from the date hierarchy so that only Month remains as shown in the following screenshot.



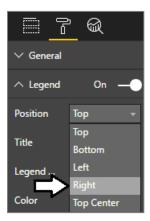
f) Add the Category field to the Legend well and add the Amount field to the Value well.



g) You report should now match the following screenshot.



h) Modify the **Position** property in the **Legend** section of the **Format** pane so that column chart displays its legend on the right.

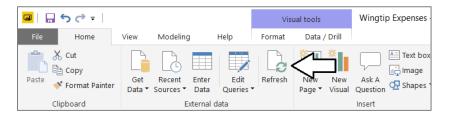


i) In the lower left corner of the report page, update the page title to **Expenses**.



Now that you have created the query and the report, you should be able to refresh the data at any time. In the next step, you will test this by refreshing the report. Note that nothing will change because the data will remain the same. However, the key point is that refreshing the report will import the most recent data into your project without having to make any updates to your query or report.

- 8. Refresh the data in the Wingtip Expenses project.
 - a) Click on the Refresh button in the Home tab of the ribbon.



b) Wait while Power BI Desktop imports the current expenses data.

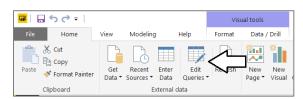


9. Save your work clicking the Save icon in the upper, left-hand side of the Power BI Desktop application window.

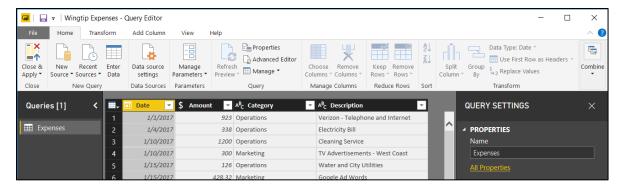
Exercise 2: Designing a Function Query to Extract Data from Multiple Files

In the following exercise, you will continue to work on the Power BI Desktop project named **Wingtip Expenses.pbix** to design a function query to implement an advanced query design where the data from multiple files can be imported into a single table.

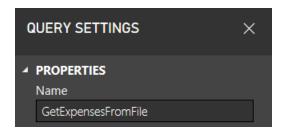
- 1. Convert the **Expenses** query into a function query.
 - a) In the Power BI Desktop application window, click the Edit Queries button to navigate to the Query Editor window.



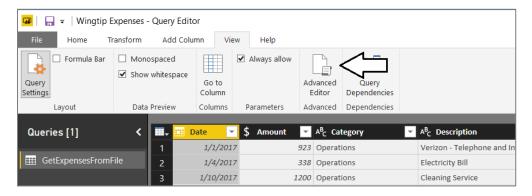
b) At this point, your project should contain a single query named **Expenses**.



c) Rename the query from **Expenses** to **GetExpensesFromFile**.



d) Click the Advanced Editor button from the View tab to display the Advanced Editor dialog.



e) You should see an editable view of the M code from the GetExpensesFromFile query in the Advanced Editor window.



f) You should also be able to see a function call to Web.Contents with a path to the file named Expenses-2017-Q1.txt.

- g) Place your cursor at the very beginning of the M code in front of the let statement.
- h) Add the following M code to the top of the query.

```
(FilePath as text) =>
```

The way that you convert a query into a function query is by adding parentheses and the arrow operator. In this scenario, you are defining your query function to accept a single text parameter named **FilePath**.

i) At this point, your Advanced Editor dialog should match the following screenshot.

```
GetExpensesFromF

(FilePath as text) =>

let

Source = Csv.Document(Web.Content
#"Changed Type" = Table.Transform
#"Renamed Columns" = Table.Rename
```

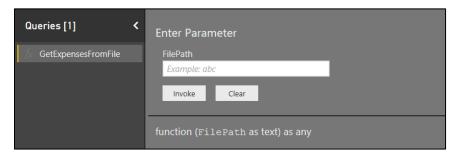
- j) Move down in the M code to the line which assign a value to **Source**.
- k) Update the call to Web.Contents to replace the hard-coded file path to the parameter named FilePath.

```
Source = Csv.Document(Web.Contents(FilePath), 4, {0,12,24,48}, null, 1252),
```

Your Advanced Editor dialog should match the following screenshot.



- m) Click the **Done** button in the Advanced Editor dialog to close it and save the changes to your M code.
- n) Note that the Query Editor window now displays the GetExpensesFromFile query differently.



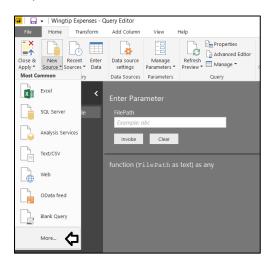
Once you convert a query into a function query, you can no longer edit it in the standard fashion. Instead, you must make any additional changes to this query by modifying its M code in the Advanced Editor window. This should not be a problem in this scenario because you did all the work to design the query logic before you converted it into a function query.

o) If you inspect the query in the Queries list on the left, you can see it now has an fx icon indicating the query returns a function.

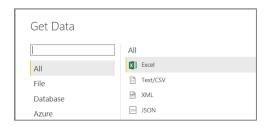


p) Save your work clicking the Save icon in the upper, left-hand side of the Query Editor window.

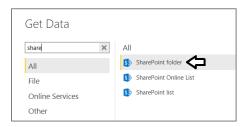
- Create a new query named Expenses to call the GetExpensesFromFile function query.
 - a) Drop down the **New Source** menu and select **More...**.



b) When the **Get Data** dialog appears, place your cursor in the search textbox.



- c) Type in "share" to see the available SharePoint datasources.
- d) Select the SharePoint folder datasource.

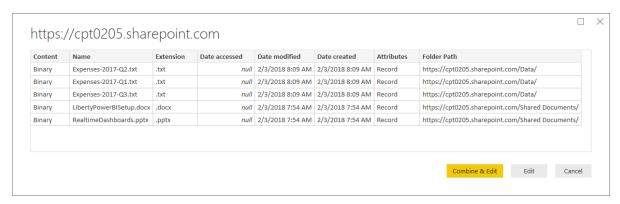


- e) Click the Connect button at the bottom right of the Get Data dialog.
- f) When promoted by the SharePoint folder dialog, enter the base URL to your SharePoint site and click OK.



It's a bit counter-intuitive with the SharePoint folder datasource. But, yes you pass the base URL to the site not the path to the library.

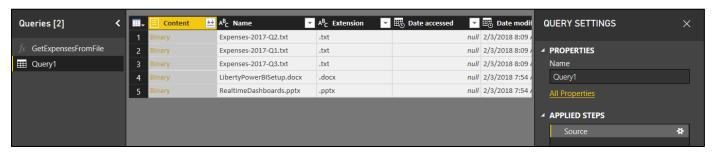
g) You should now see a dialog with a list of files from all document libraries in the site.



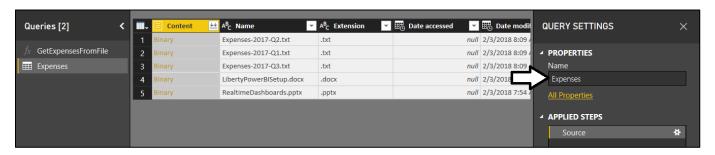
h) Click the **Edit** button to open the new query in the Query Editor window.



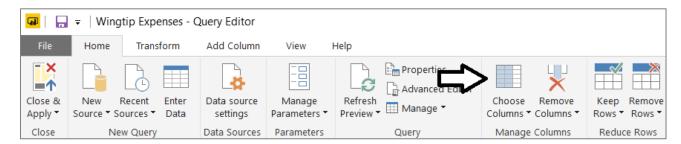
i) When the Query Editor window opens, should see a query named Query1 as shown in the following screenshot.



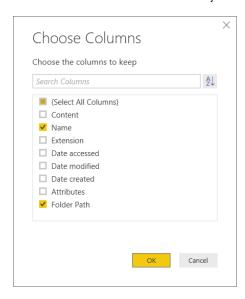
i) Rename the guery from Query1 to Expenses.



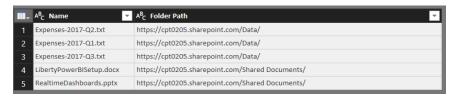
k) Click the Choose Columns button in the Home tab of the Query Editor window.



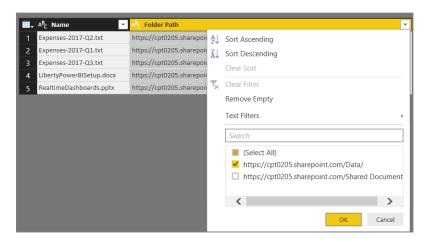
Unselect all columns and then select just the Name column and the Folder Path column then click OK.



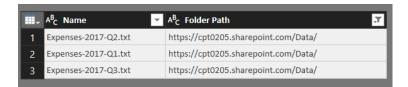
m) At this point, your query output should match the following screenshot.



n) Using the dropdown menu on the right side of the Folder Path column, select just the folder path that ends with /Data/.

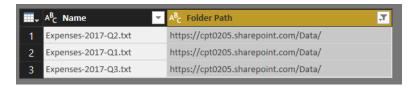


o) Now your query results should only display the three data files from the **Data** document library.



Over the next few steps you will combine the two columns into a single column named **FilePath** by adding a **Merge Column** step. Note that you must select the column on the right first and then the column on the left second to merge the columns correctly.

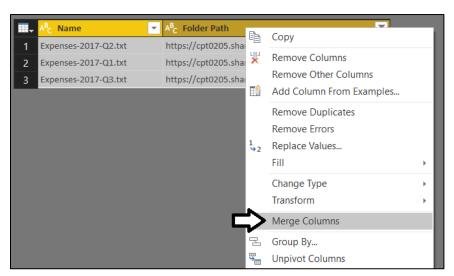
- 3. Add a Merge Columns step to create a new column named FilePath.
 - a) Using the mouse, select the **Folder Path** column by clicking its column header.



b) With the Folder Path column selected, hold down the Ctrl key and click the Name column so both columns are selected.



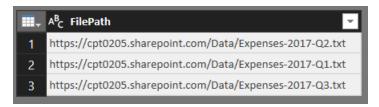
c) Right-click the Folder Path column header and select the Merge Columns command.



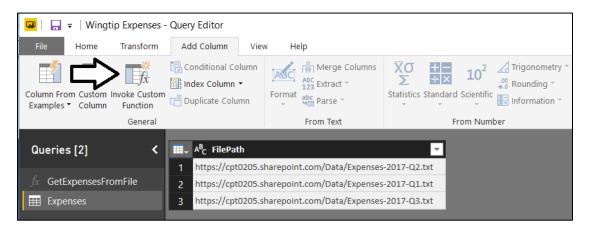
- d) In the Merge Columns dialog...
 - i) Leave the **Separator** set to **--None--**.
 - ii) Add a New column name of FilePath.
 - iii) Click OK.



e) The query results should now show a single column named FilePath as shown in the following screenshot.



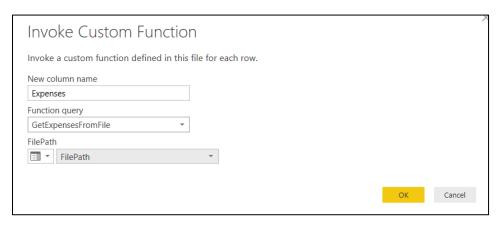
- 4. Modify the Expenses query to call the function query named GetExpensesFromFile.
 - a) In the Query Editor window, make sure the Expenses query is the selected in the Queries list on the left.
 - b) Navigate to the Add Columns tab.
 - c) Click the Invoke Custom Function button in the ribbon.



- d) In the Invoke Custom Function dialog...
 - i) Add a New column name of Expenses.
 - ii) Drop down the Function query menu and select GetExpensesFromFile.
 - iii) Use the dropdown menu under the FilePath parameter and select Column Name.



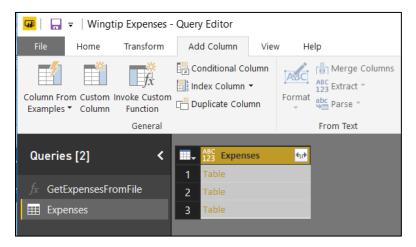
iv) Use the dropdown to configure the FilePath parameter with the FilePath column and then click OK.



e) You should see a new column named Expenses whose values contains Table objects as shown in the following screenshot.



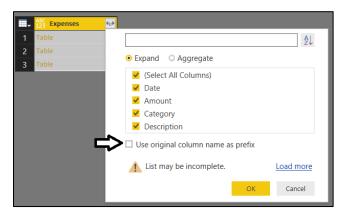
f) Remove the column named FilePath from the query so the query results only show the new column named Expenses.



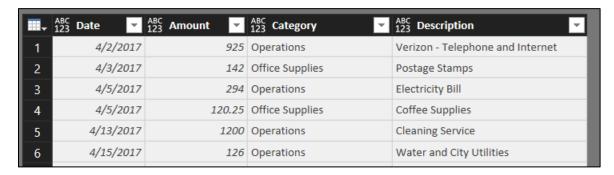
- 5. Expand the Table objects in the Expenses column into rows.
 - a) Click the **Expand** button on the right side of the **Expenses** column header.



- b) Select the columns named Date, Amount, Category and Description.
- c) Make sure the **Use original column name as prefix** checkbox is not selected.
- d) Click **OK** to expand the rows for each **Table** object.

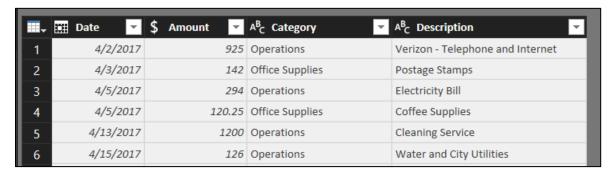


e) You query should now return a separate row for each expense.

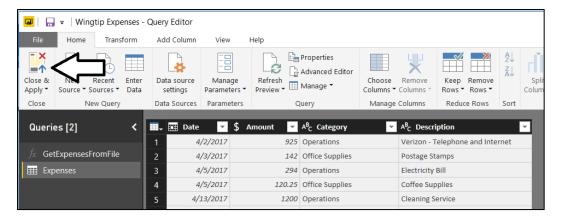


You will notice the datatype for the columns in the table are not set to the correct types. Your next task is to fix that.

- 6. Update the datatypes for the columns in the **Expenses** query.
 - a) Update the datatype of the **Date** column to **Date**.
 - b) Update the datatype of the **Amount** column to **Fixed Decimal number**.
 - c) Update the datatype of the Category column to Text.
 - d) Update the datatype of the **Description** column to **Text**.

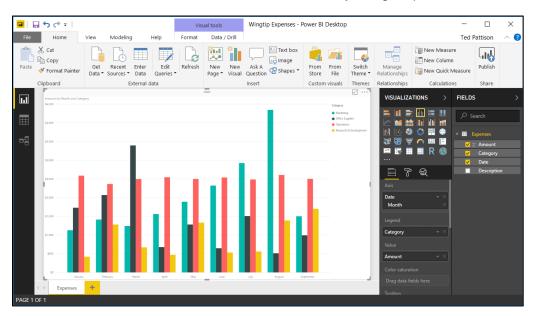


- 7. Test out the query to make sure it works properly.
 - a) Click the Close and Apply button in the ribbon to close the Query Editor window and to execute the Expenses query.

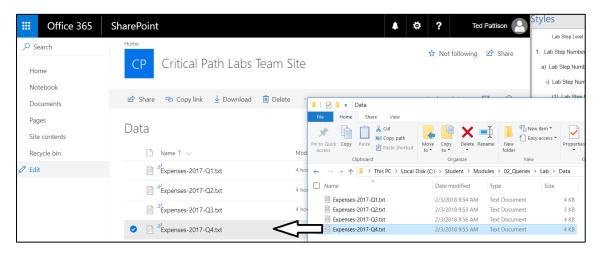


When the **Expenses** query executes, it should call the **GetExpensesFromFile** passing the name of each of the three data files you uploaded to the **Data** document library in SharePoint Online. The query should also append the rows of data from each of these three data files into a single table named **Expenses**. The report with the clustered column chart should still display correctly, but now it should show data combined from three different files to include expenses for **Q1**, **Q2** and **Q3**.

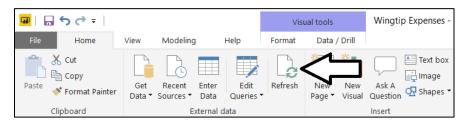
- 8. Inspect the **Expenses** page of the report you created in the previous exercise.
 - a) The clustered column chart should now show data from January through September.



- 9. Add another data file to the Data document library in SharePoint Online.
 - a) Using the browser, return to the **Data** document library in your SharePoint site.
 - b) Upload the one remaining file named Expenses-2017-Q4.txt.



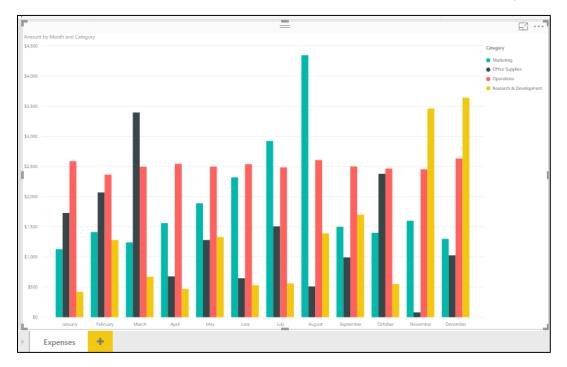
- 10. Refresh the data in the Power BI Desktop project named Wingtip expenses.pbix.
 - a) Return to the Wingtip expenses.pbix project in Power BI Desktop.
 - b) Click the **Refresh** button on the **Home** tab to re-execute the **Expenses** query.



c) When the Expenses query runs, the Refresh dialog indicates that it is also executing GetExpensesFromFile.



d) Once the query has completed, you should see that the column chart now displays data through December.



11. Save your work clicking the Save icon in the upper, left-hand side of the Power BI Desktop application window.

Congratulations. You have now designed a complex query using a function query. Keep in mind that you will be continuing to work on the Power BI Desktop project named **Wingtip Expenses.pbix** in the next lab on data modeling.