Data Access & Preparation

Read Me: This lab focuses on getting Power BI Project up and running, working with Power Query features to transform data coming from a SQL Azure database. Labs that follow will build of this lab.

Lab Time: 45 minutes

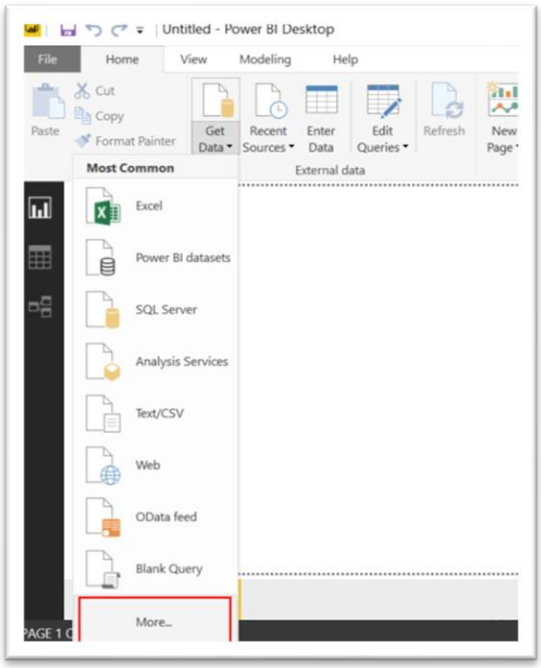
Exercise 1: Importing Data from Azure SQL Database

In this exercise you will create a Power BI Project that imports data from SQL Azure Database

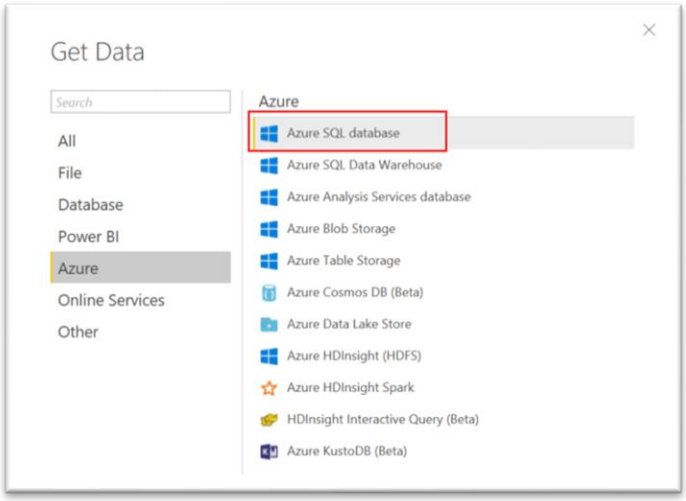
1. Start a new Power BI project by launching Power BI Desktop
2. Save the project as **product_sales.pbix** in the **Class_Project** folder

   C:\armely\Class_Projects\product_sales.pbix

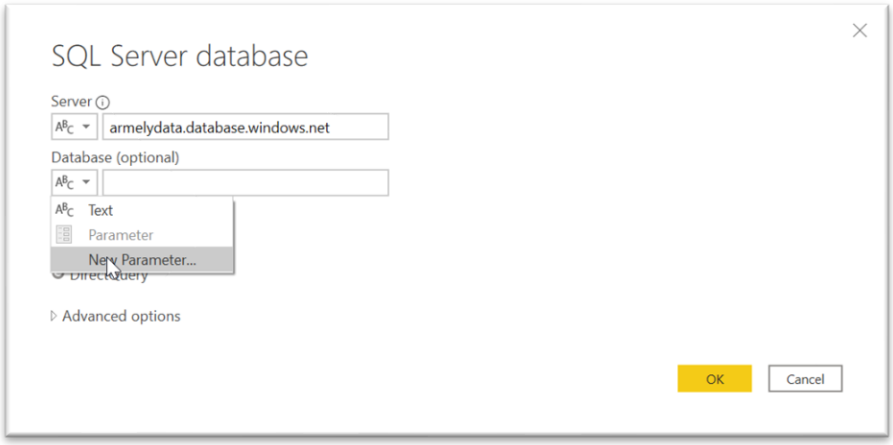
3. Click on Get Data menu button and select More at the bottom



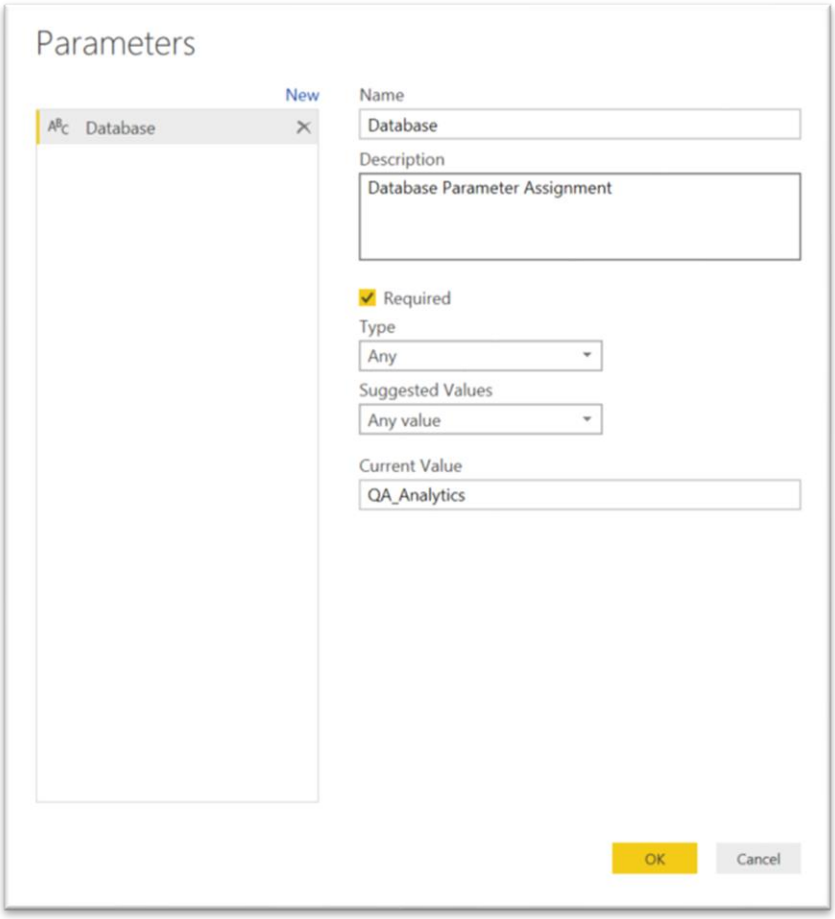
4. Select **Azure SQL Database** from **Azure** subsection



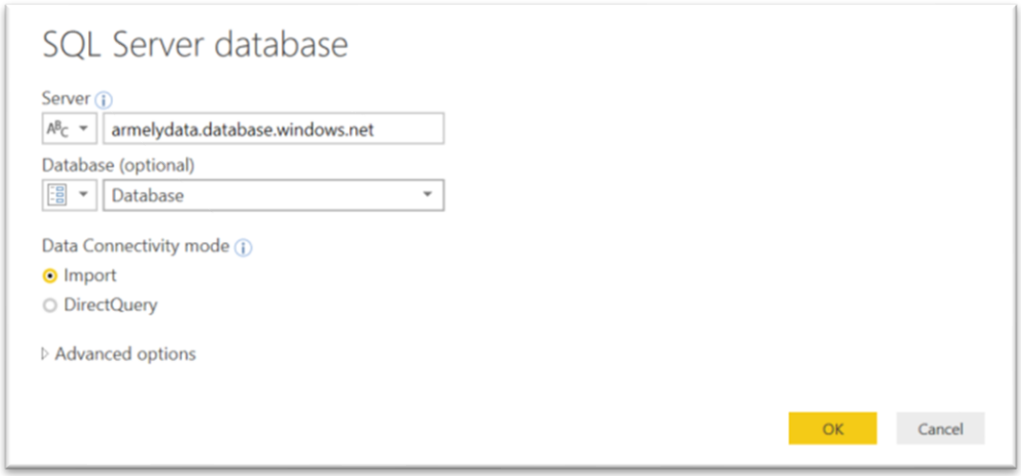
5. When prompted for SQL Server database information, provide the following
    a. **Important** Enter Server name value as **armelydata.database.windows.net**
    b. **For Database (Optional)**
        i. **Click the**  **Icon and select New Parameter**



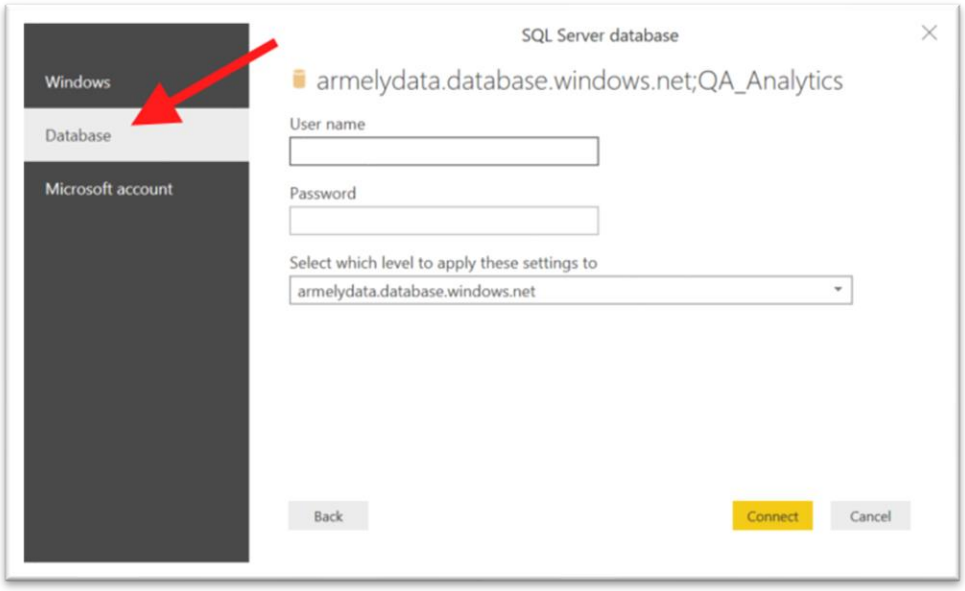
        ii. In the Parameter screen, set the following values
            1. Name: **Database**
            2. Description: Database Parameter assignment
            3. Required: Checked
            4. Type: Any
            5. Suggested Values: Any value
            6. Current Value: **QA_Analytics**
        iii. Your values should appear as follows, click OK after completion.



iv. Your connection should appear as follows, click OK to continue.

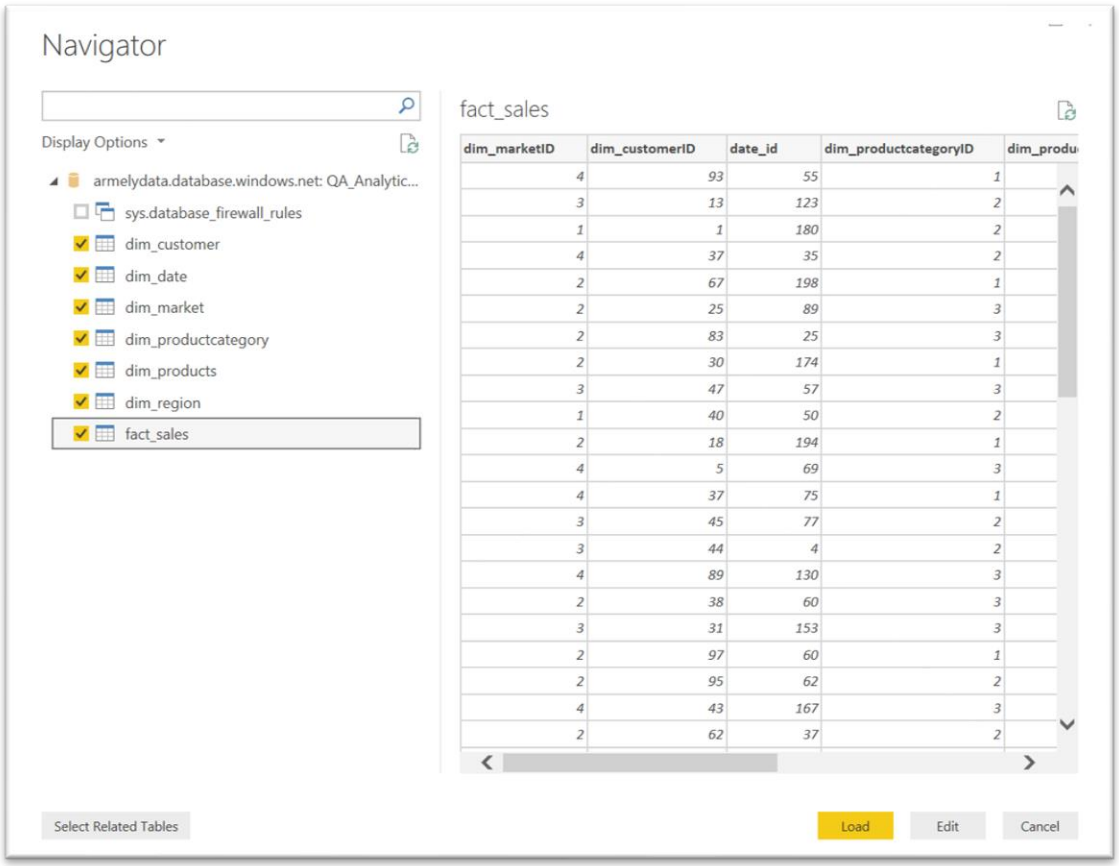


6. The connection will require credentials to authenticate. (Since it's a cloud-based database, we will use **Database Connection**)
    a. Windows – for network-based connections
    b. Database – for cloud-based connection
    c. Microsoft Account –Azure IAM

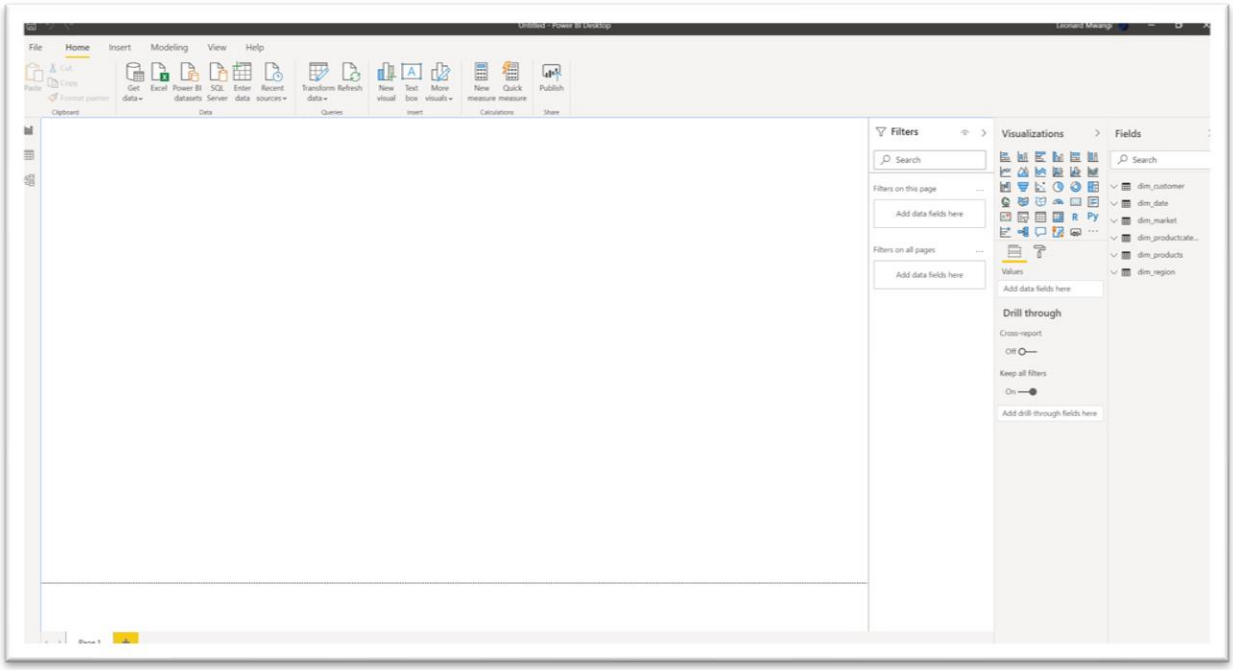


7. Enter the credentials for SQL user account as provided below.
    a. Username: **training**
    b. Password: **Th!st@Train**

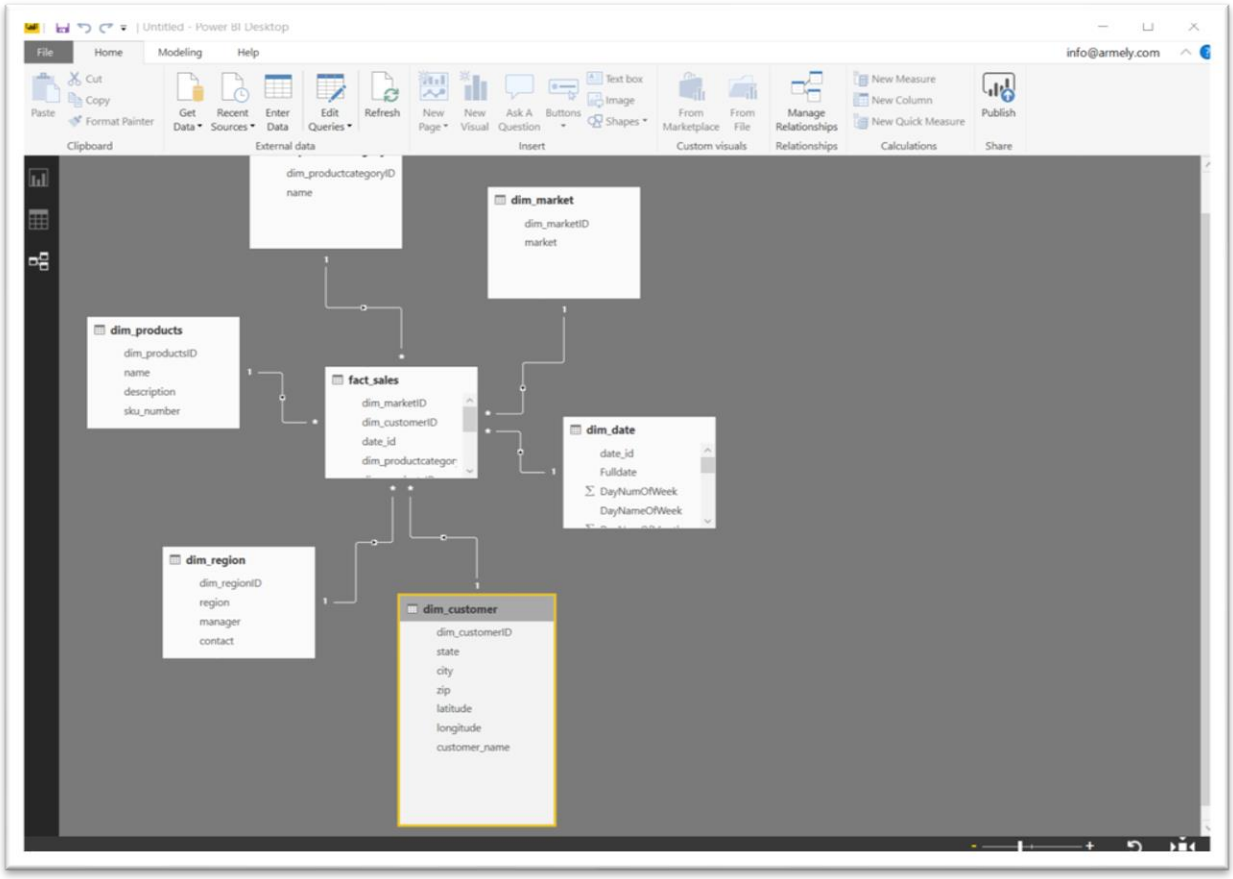
8. The connection gives you access to database tables **Check** the following tables and **Click Load**



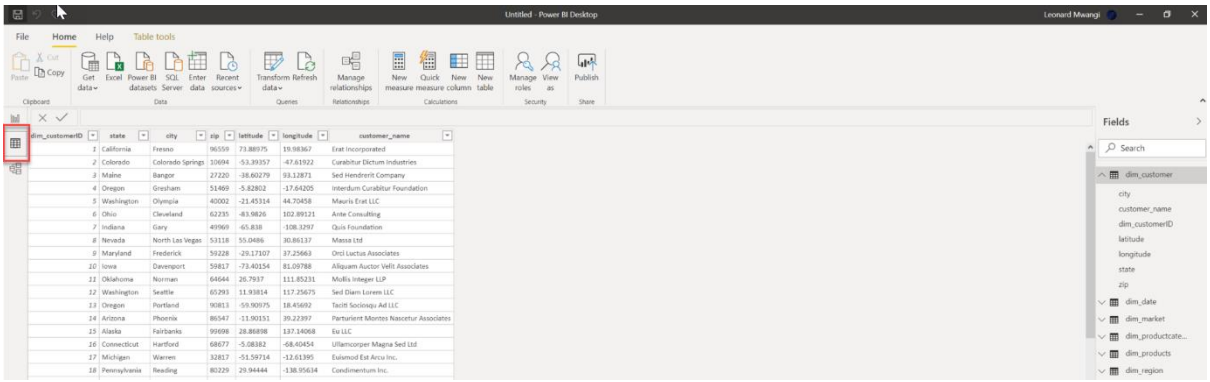
9. After loading your Power BI Desktop should show as follows



10. Click on Relationships, **Power BI has autodetected relationships based on column names.**



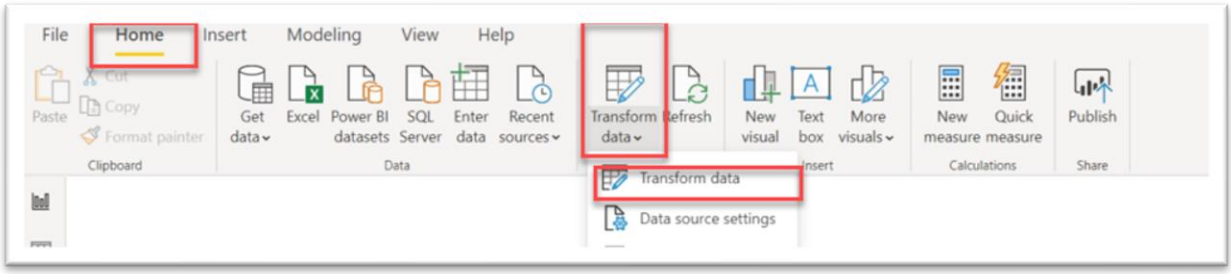
11. Review the data in **Data View**



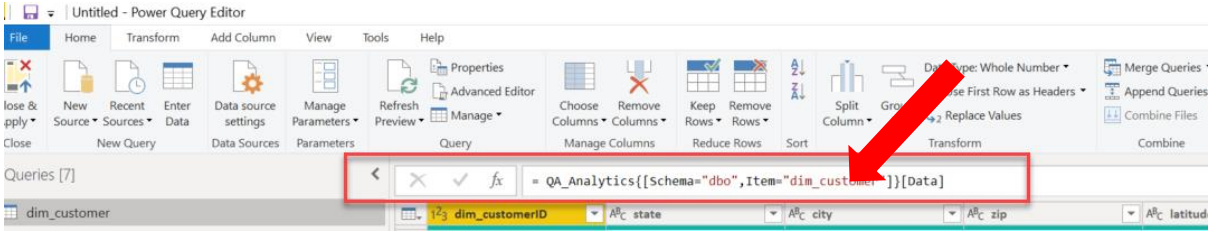
Exercise 2: Reshaping Customer Data

In this exercise you will use **Query Editor** to modify **Customers** query and perform transformation

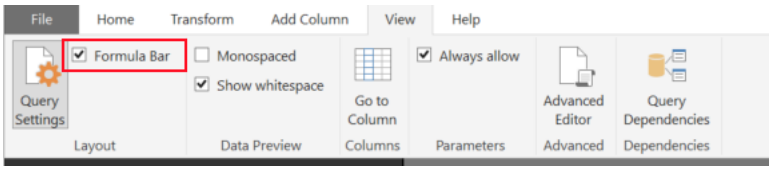
1. Make sure your **product_sales.pbix** project is open.
2. Click on the drop-down next to **Transform Data** and select **Transform data**



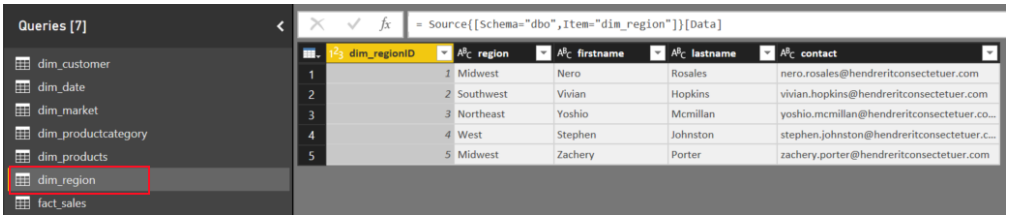
3. Ensure that the formula bar is visible



   a. **If the formular bar is not visible**, check the formula bar from the **view menu**



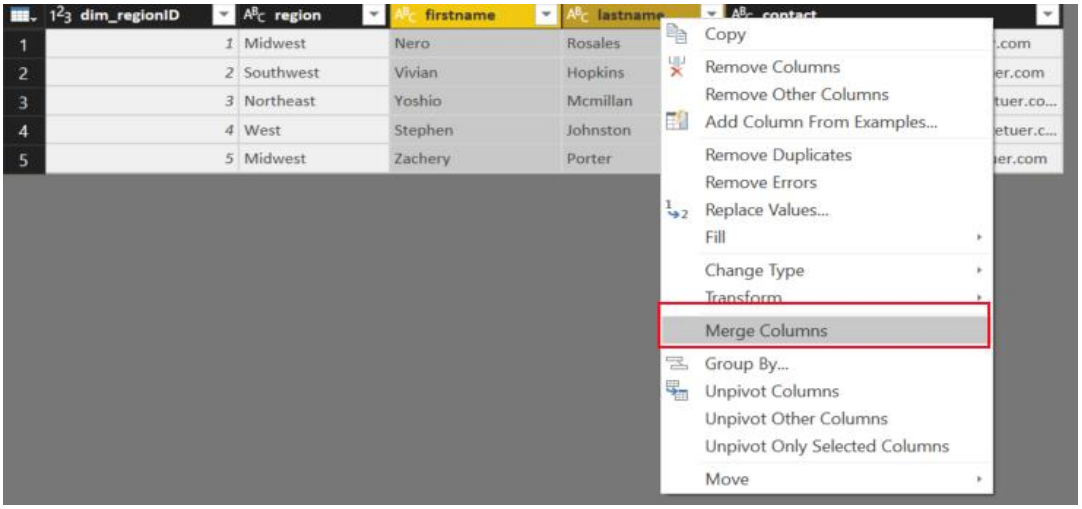
4. In this step, we will merge **firstname** and **lastname** to create manager column
   a. Click on dim_region from queries



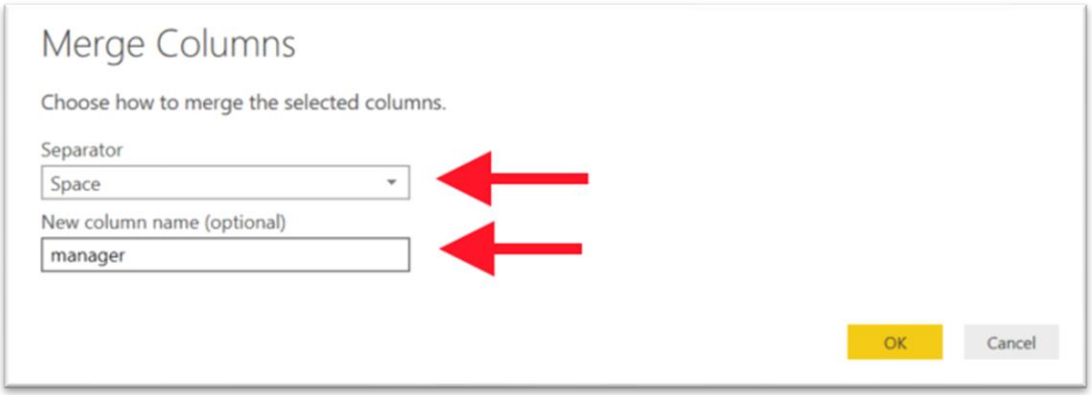
b. Click on **firstname** column header

c. Hold down the **SHIFT** key and click on **lastname** column header

d. **Important** Let the SHIFT key off and right click on selected columns and select **Merge Column** menu command



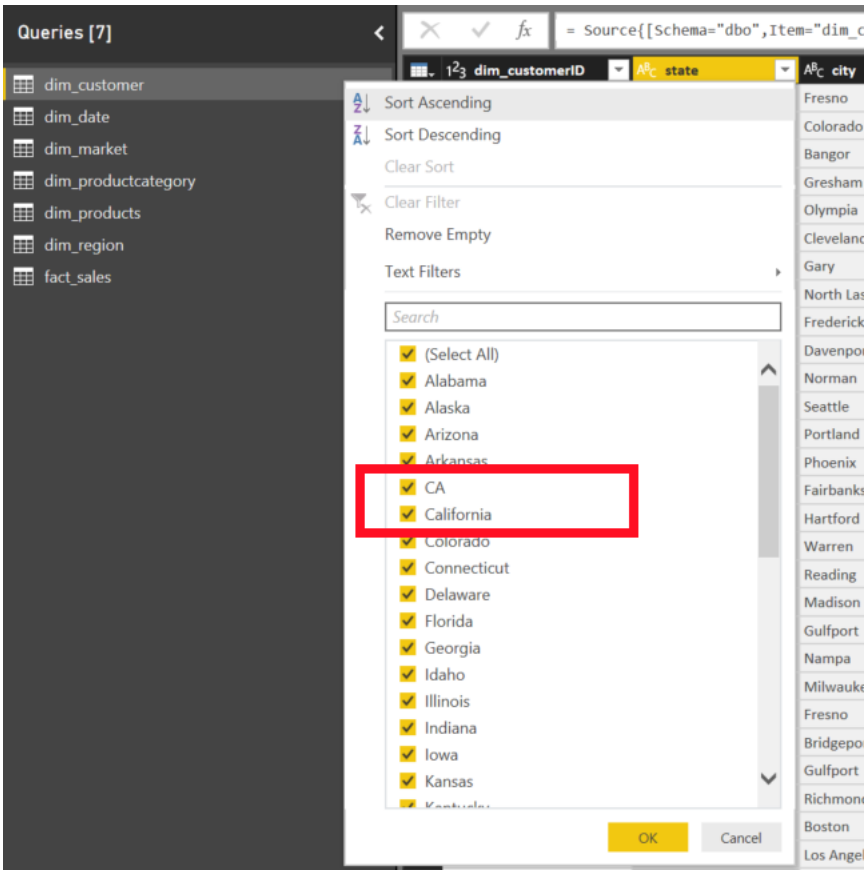
e. **Important** in the Merge Column dialog, drop down the Separator control and select Space. On New Column name, use **Manager** as value and click OK button.



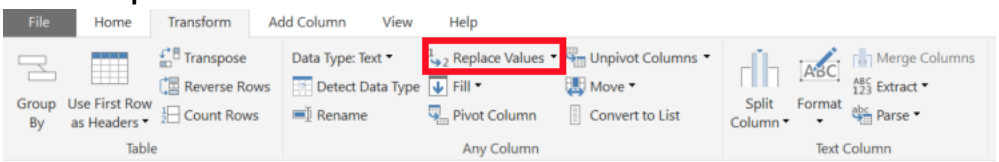
f. The outcome should look as follows

	dim_regionID	region	manager	contact
1	1	Midwest	Rosales Nero	nero.rosales@hendreritconsectetuer.com
2	2	Southwest	Hopkins Vivian	vivian.hopkins@hendreritconsectetuer.com
3	3	Northeast	Mcmillan Yoshio	yoshio.mcmillan@hendreritconsectetuer.co...
4	4	West	Johnston Stephen	stephen.johnston@hendreritconsectetuer.c...
5	5	Midwest	Porter Zachery	zachery.porter@hendreritconsectetuer.com

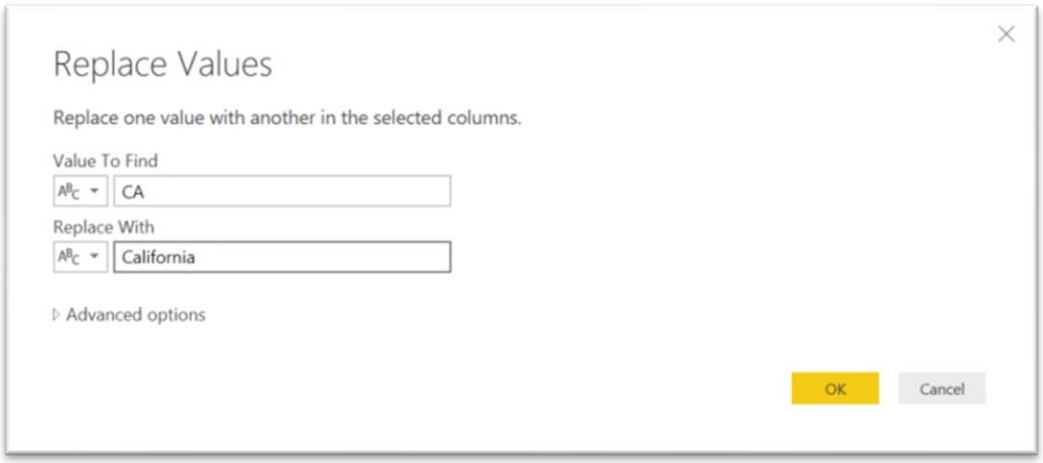
5. **Important Fixing Data** issues, in this step, we will fix data issues so that there can be consistency
   a. Click on the **dim_customer** table to select it. Inspect the **state** data, you will notice **CA** and **California** which would show as two different records.



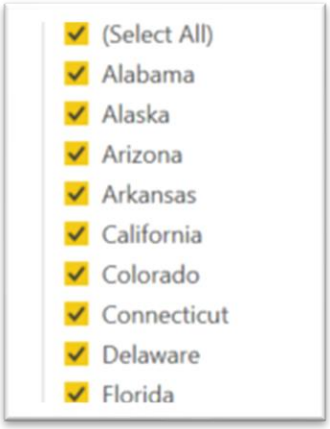
   b. Click on **Transform** tab on the ribbon, make sure the **state column** is highlighted and click on **Replace Values**



c. In the **Replace Value dialog, enter CA** in the value to find and **California** in **Replace with field**

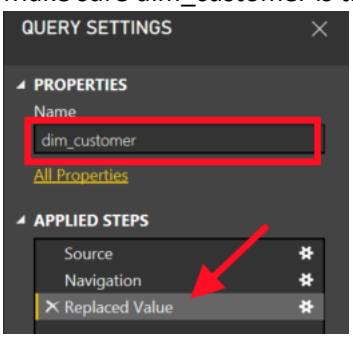


d. The results should not contain CA

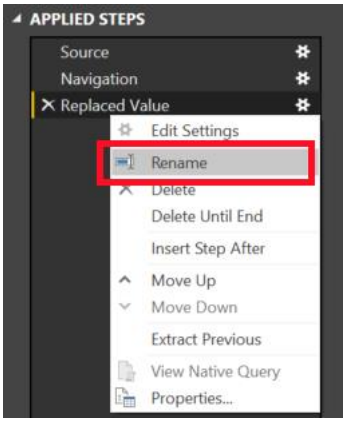


6. It's a good practice to rename the **Applied Steps** to help trace your path in the future. On the right in the **Query Setting**,

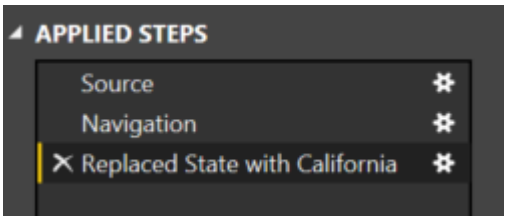
a. Make sure dim_customer is the table you are working with



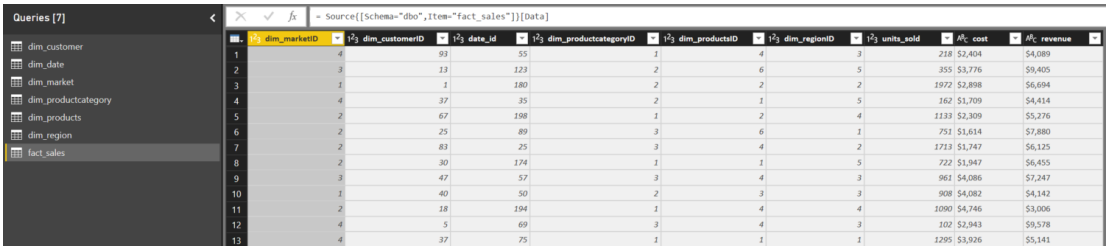
b. Right click on the step and select **Rename**



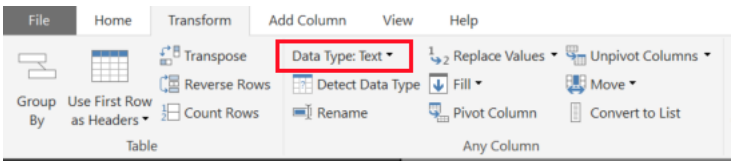
c. End result



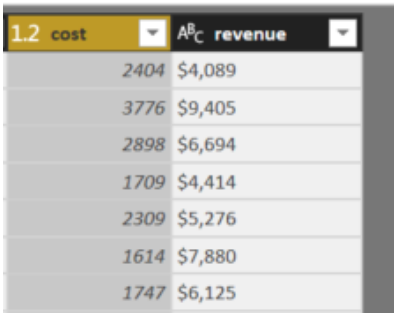
7. In this step, we will fix the **Cost** in **Fact_Sales** table from text to Decimal so that it can accommodate currency
   a. Select the **fact_sales** Query on the left



   b. Select the header for **cost** column, click on **Transform** and change the **Data Type** from Text to **Decimal**

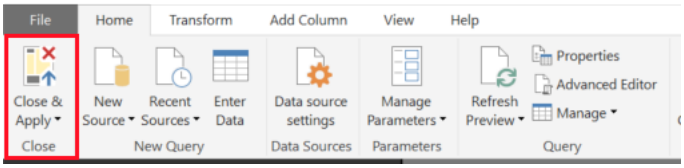


   c. Results

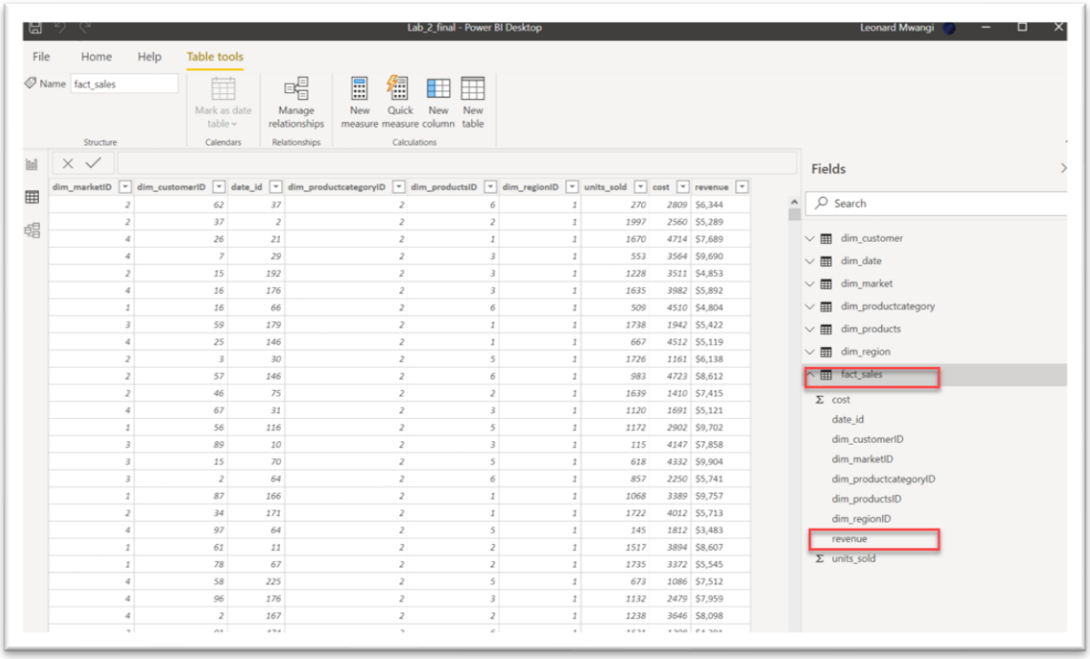


1.2 cost	revenue
2404	$4,089
3776	$9,405
2898	$6,694
1709	$4,414
2309	$5,276
1614	$7,880
1747	$6,125

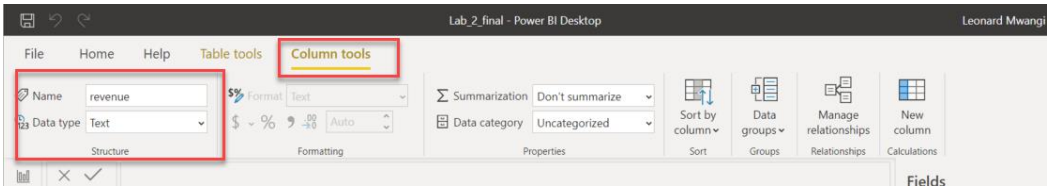
8. Click on Close & Apply button to apply the changes



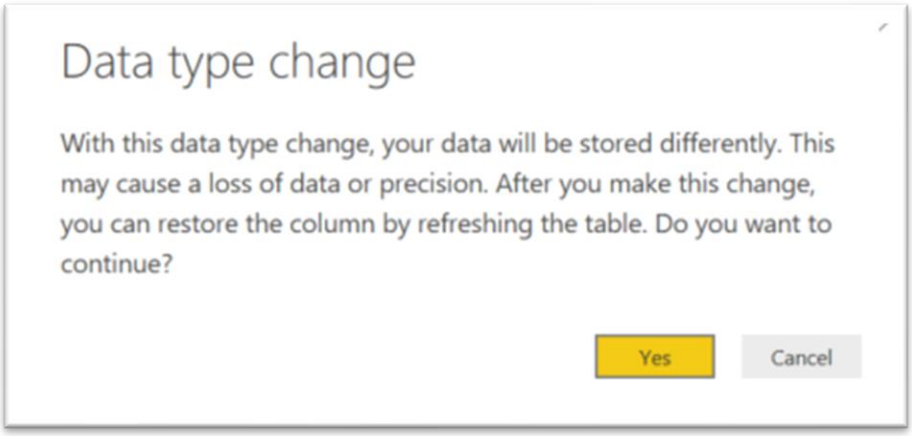
9. Changing the Data Type from Data View
   a. Click on **revenue** from left to highlight it



   b. On the **Column tools,** tab change the **Data Type** to **Decimal**

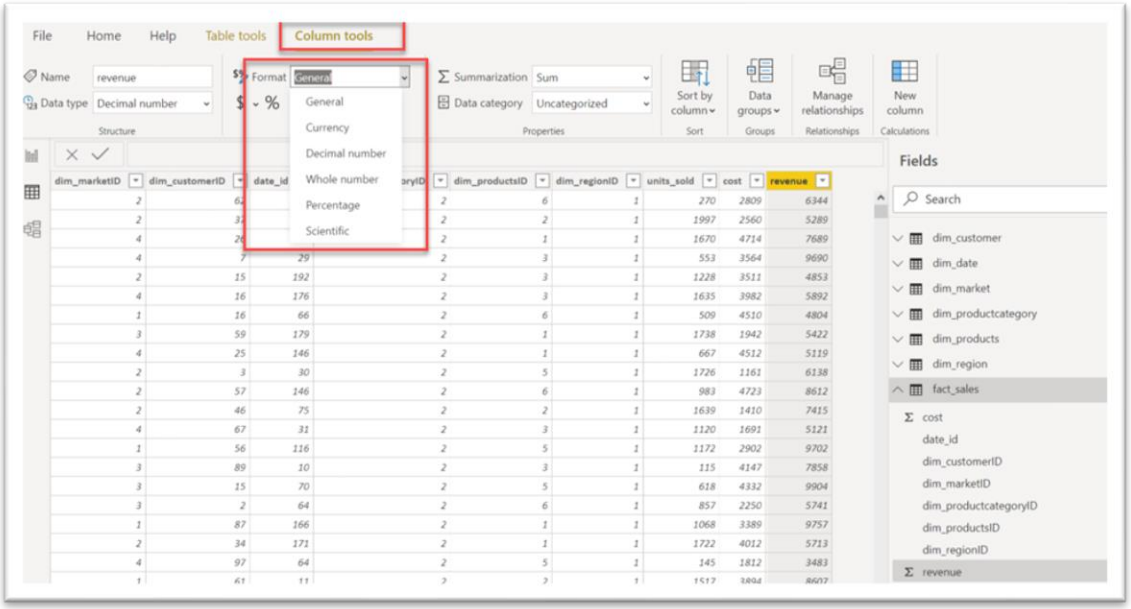


c. You will get a warning on data change



10. Important Change the **Cost** and **Revenue** to Currency

a. Highlight **Cost** column in the Data view, click on Modeling and select **Format** and change it to Currency as shown below.



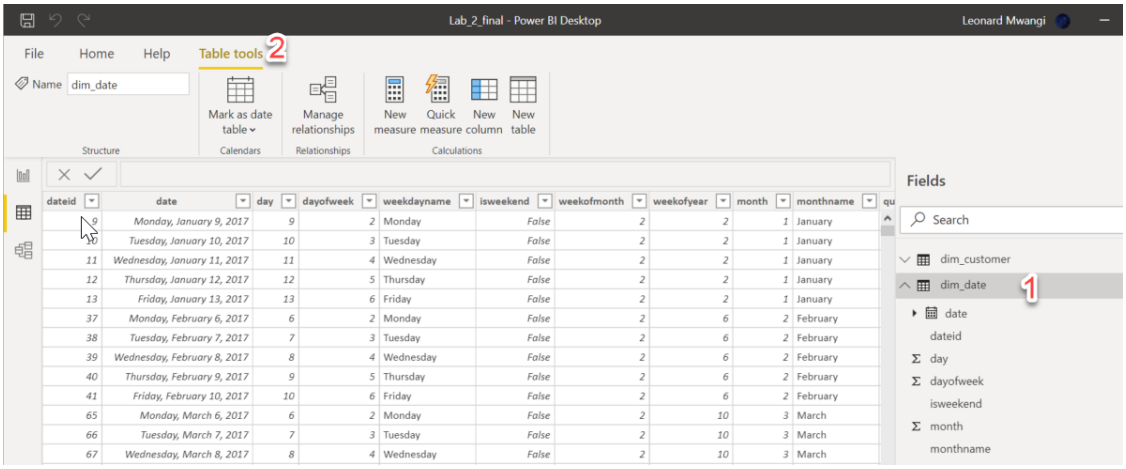
b. The results should be

cost	revenue
$1,614	$7,880.00
$1,747	$6,125.00
$4,086	$7,247.00
$2,943	$9,578.00
$4,235	$9,125.00

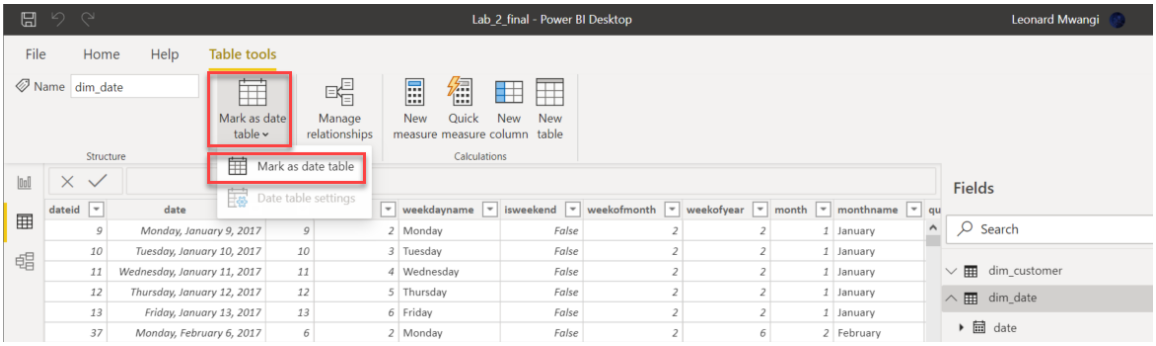
Exercise 3: Enable Time Intelligence

In this exercise you will configure Dates Table to mark it as date table so that it can allow date-based functions.

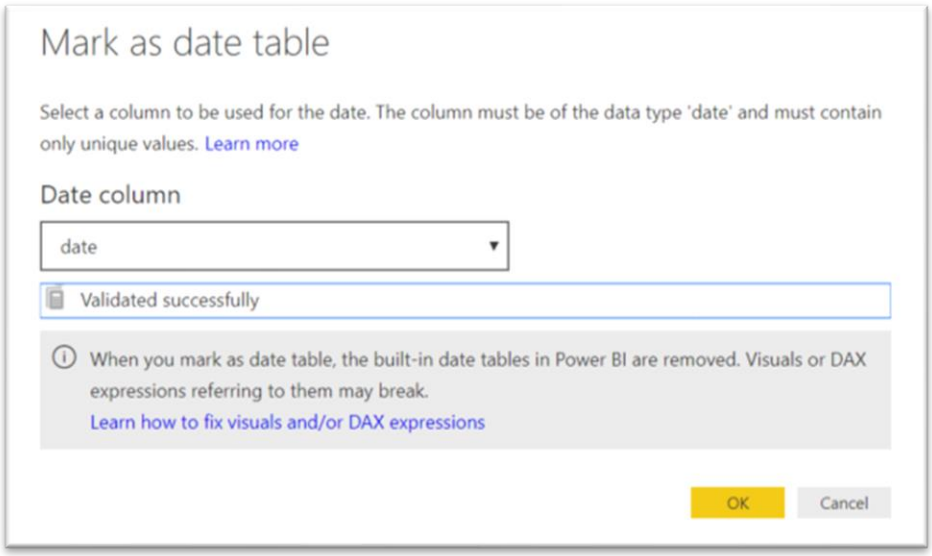
1. In this step, you will mark the dim_date table as a time dimension.
    a. Select **dim_date** table from the **Data View** then click on **Table tools**



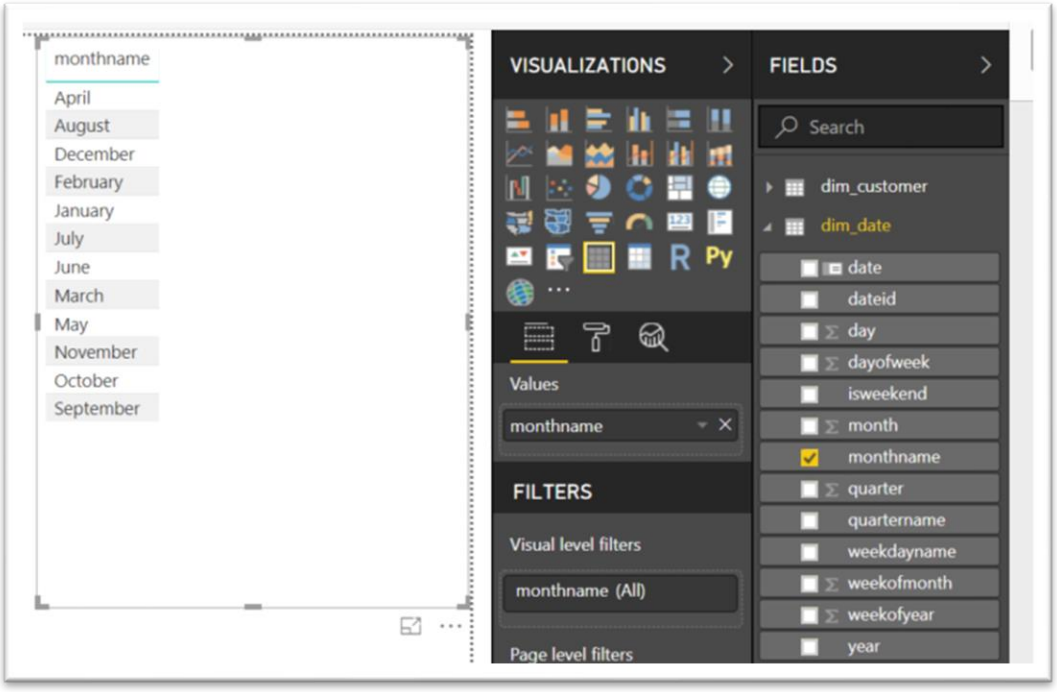
    b. On the ribbon select **Mark Table** as **Date Table**



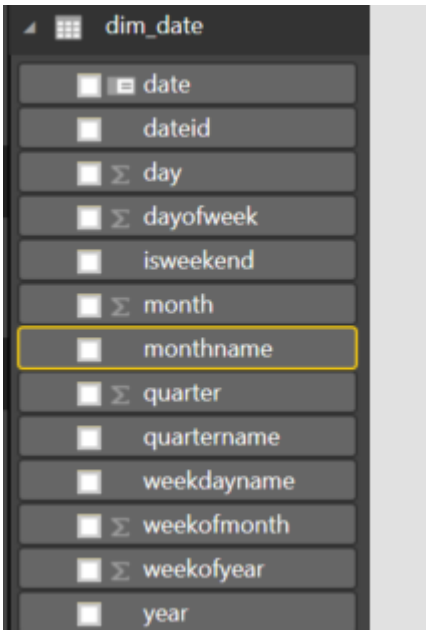
    c. Select date as the Date Column and click OK.



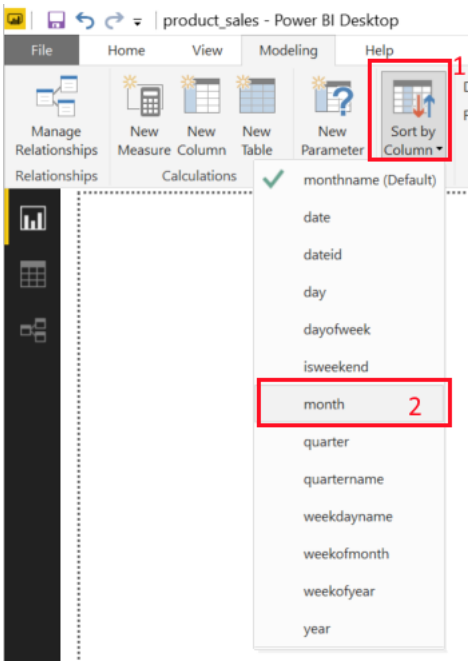
2. In this step, you will fix month sorting from alphabetic to proper sorting.
    a. Click on **Report View** and put a check mark on **monthname**, you will notice monthname is sorted alphabetically



    b. Highlight the monthname column (to show with a yellow marking around it)



c. On Modeling, tab click on Sort by Column and set the sort to Month



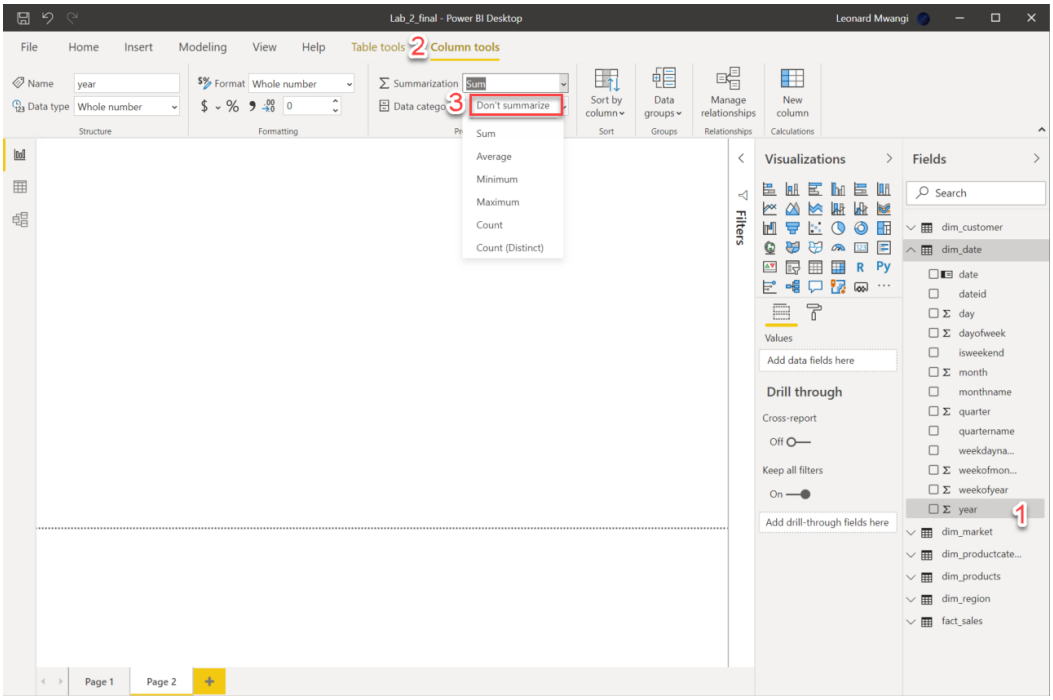
d. The sort order now should be by month number.

Exercise 4: Change Column Properties

Power BI by default sets any value that's a number to be summable, not all numbers are supposed to be summable. Numbers like years, months, dates should not be summed. In this exercise you will configure some date table columns to not summarize.

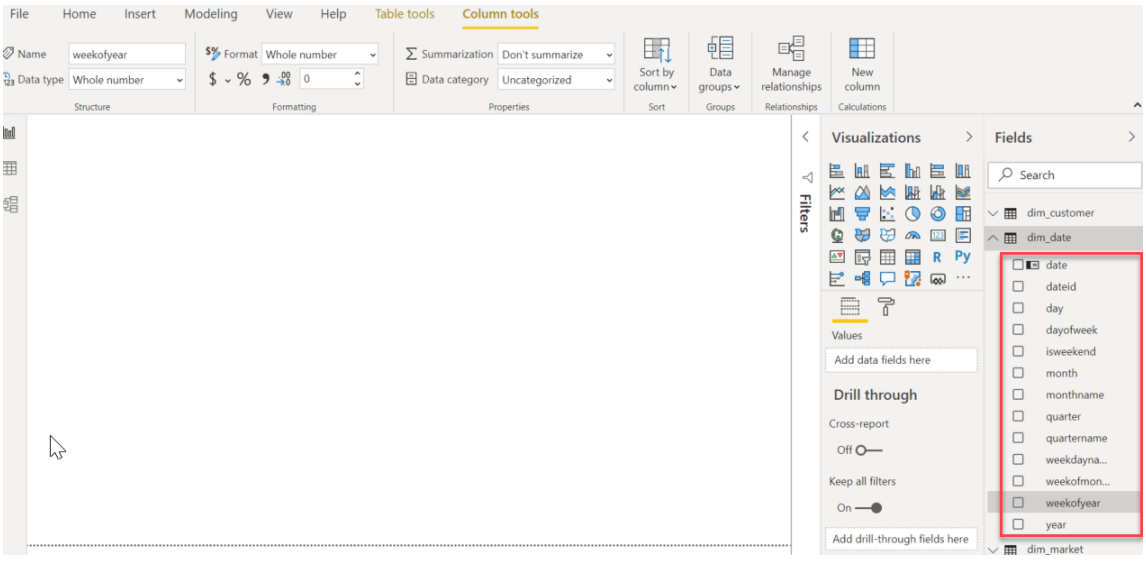
1. In this step, you will change the column properties for year in dim_date table from summable to not summarize column. Summable columns are marked with **Σ** before the column name.

a. **Important** Click on **Year** column from Fields section and select **Column tools,** change Summarization to "**Don't Summarize**"



b. Repeat this process for the following columns
   i. Day
   ii. Dayofweek
   iii. Month
   iv. Quarter
   v. WeekofMonth
   vi. WeekofYear

c. Result should be as follows



2. Save your work by clicking the save button in the upper left corner of the Power BI Desktop Window.

Congratulation, you have completed this lab for modifying queries to transform data as its being loaded into the data model of your project. You will continue to work on this project in subsequent labs.