

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

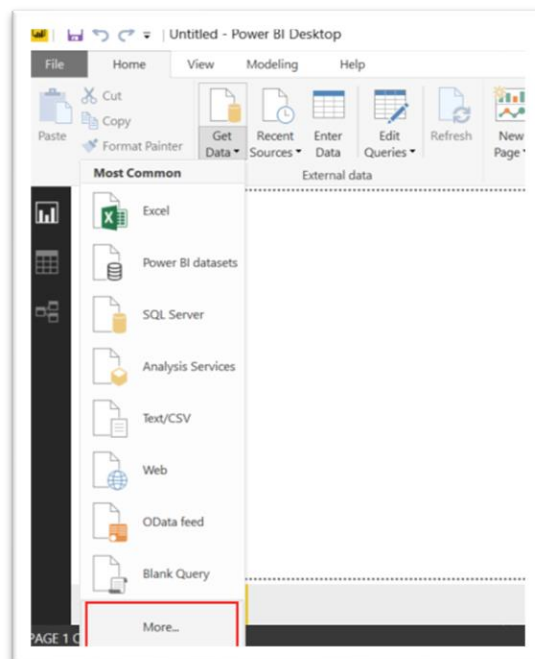
Exercise 1: Importing Data from Azure SQL Database

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

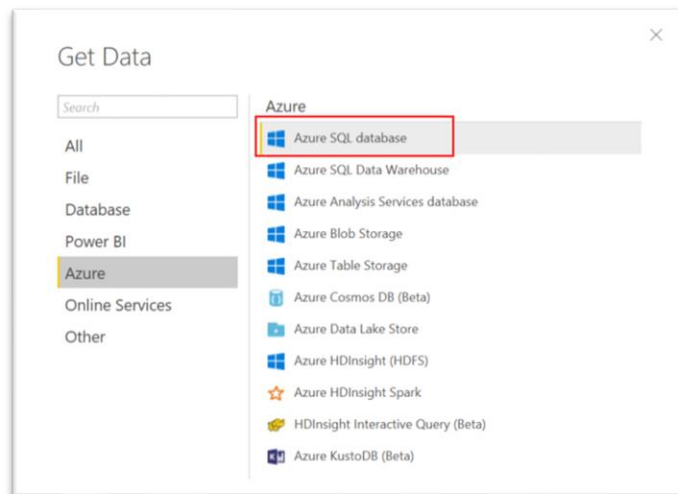
1. Start a new Power BI project by launching Power BI Desktop
2. Save the project as **product_sales.pbix** in the **Class_Project** folder

C:\armely\Class_Projects\product_sales.pbix

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

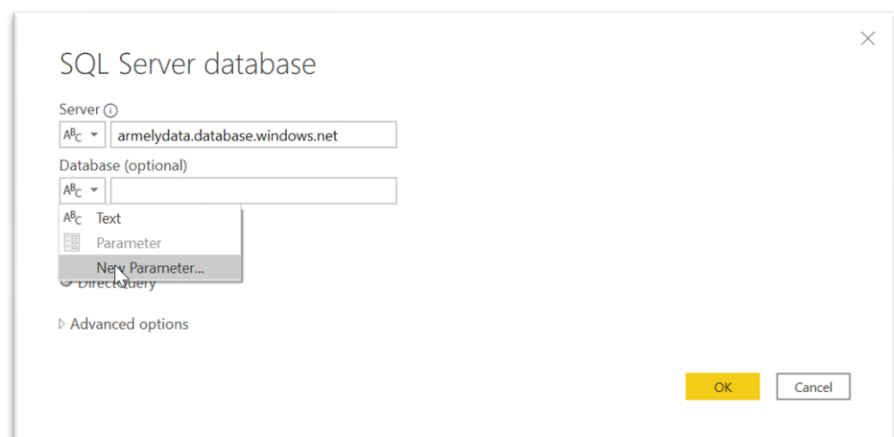


4. Select **Azure SQL Database** from **Azure** subsection



5. When prompted for SQL Server database information, provide the following
- Important** Enter Server name value as **armelydata.database.windows.net**
 - For Database (Optional)**

- i. Click the  icon and select **New Parameter**



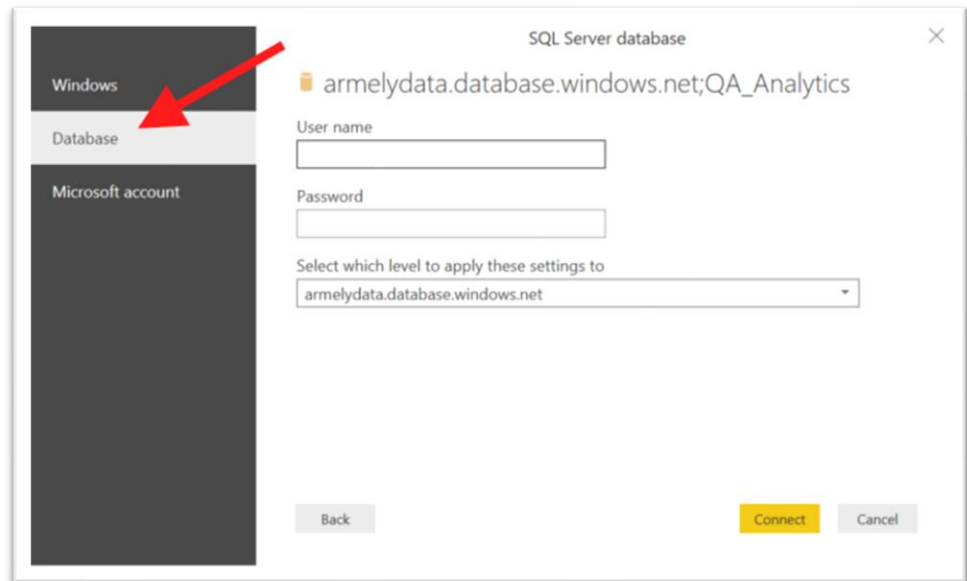
- ii. In the Parameter screen, set the following values
- Name: **Database**
 - Description: Database Parameter assignment
 - Required: Checked
 - Type: Any
 - Suggested Values: Any value
 - Current Value: **QA_Analytics**
- iii. Your values should appear as follows, click OK after completion.

The 'Parameters' dialog box in Power BI is shown. It has a title bar 'Parameters' and a 'New' button. On the left, there is a list box with 'Database' selected. On the right, there are several fields: 'Name' with the value 'Database', 'Description' with the value 'Database Parameter Assignment', a checked 'Required' checkbox, 'Type' with a dropdown menu showing 'Any', 'Suggested Values' with a dropdown menu showing 'Any value', and 'Current Value' with the value 'QA_Analytics'. At the bottom right, there are 'OK' and 'Cancel' buttons.

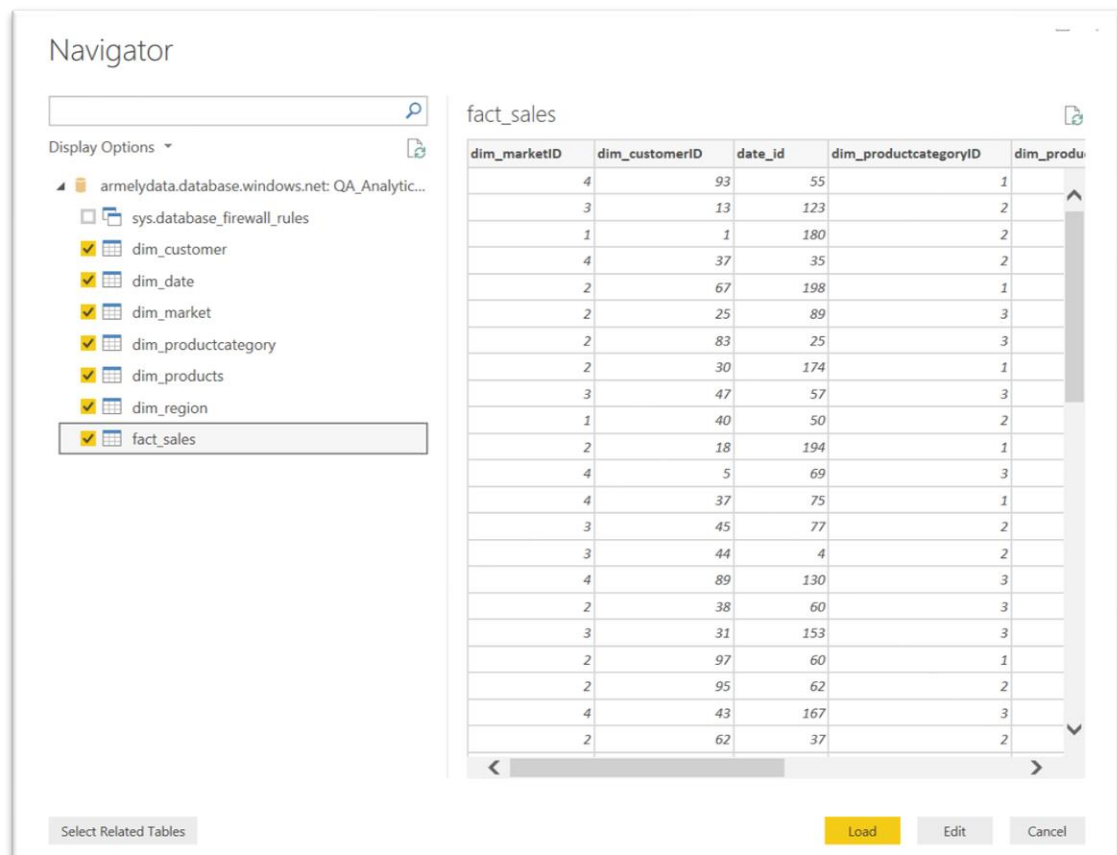
iv. Your connection should appear as follows, click OK to continue.

The 'SQL Server database' connection dialog box is shown. It has a title bar 'SQL Server database'. The 'Server' field is set to 'armelydata.database.windows.net'. The 'Database (optional)' field is set to 'Database'. The 'Data Connectivity mode' section has 'Import' selected. There is an 'Advanced options' link at the bottom. At the bottom right, there are 'OK' and 'Cancel' buttons.

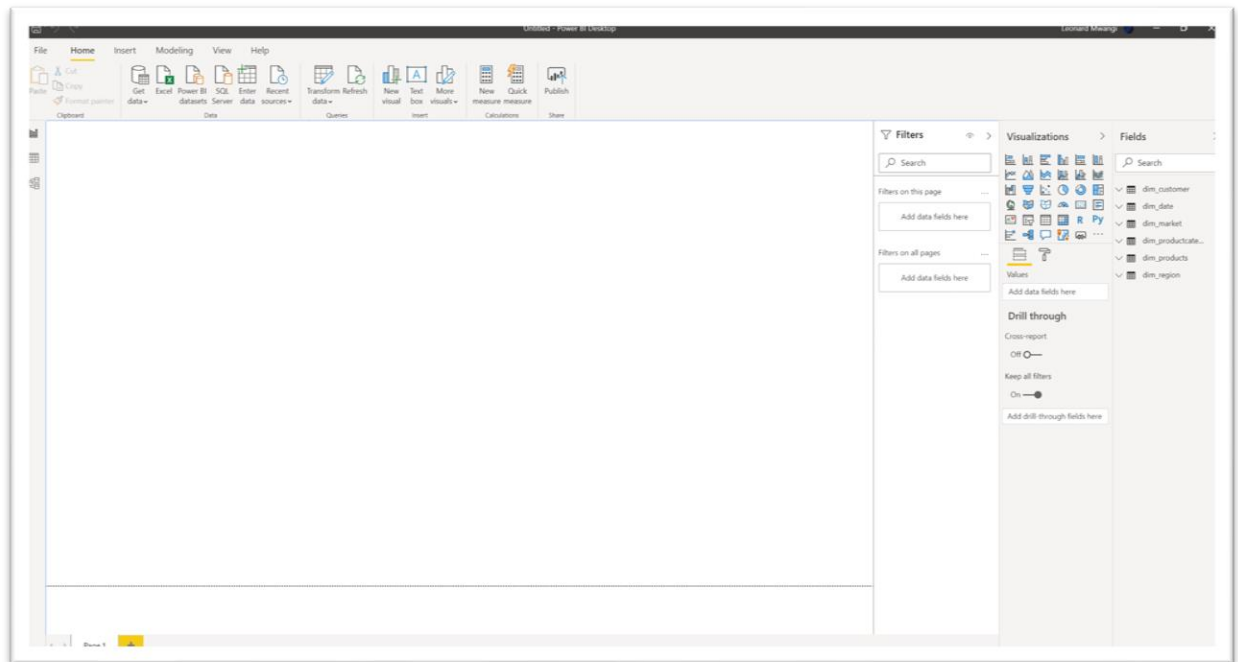
6. The connection will require credentials to authenticate. (Since it's a cloud-based database, we will use **Database Connection**)
 - a. Windows – for network-based connections
 - b. Database – for cloud-based connection**
 - c. Microsoft Account –Azure IAM



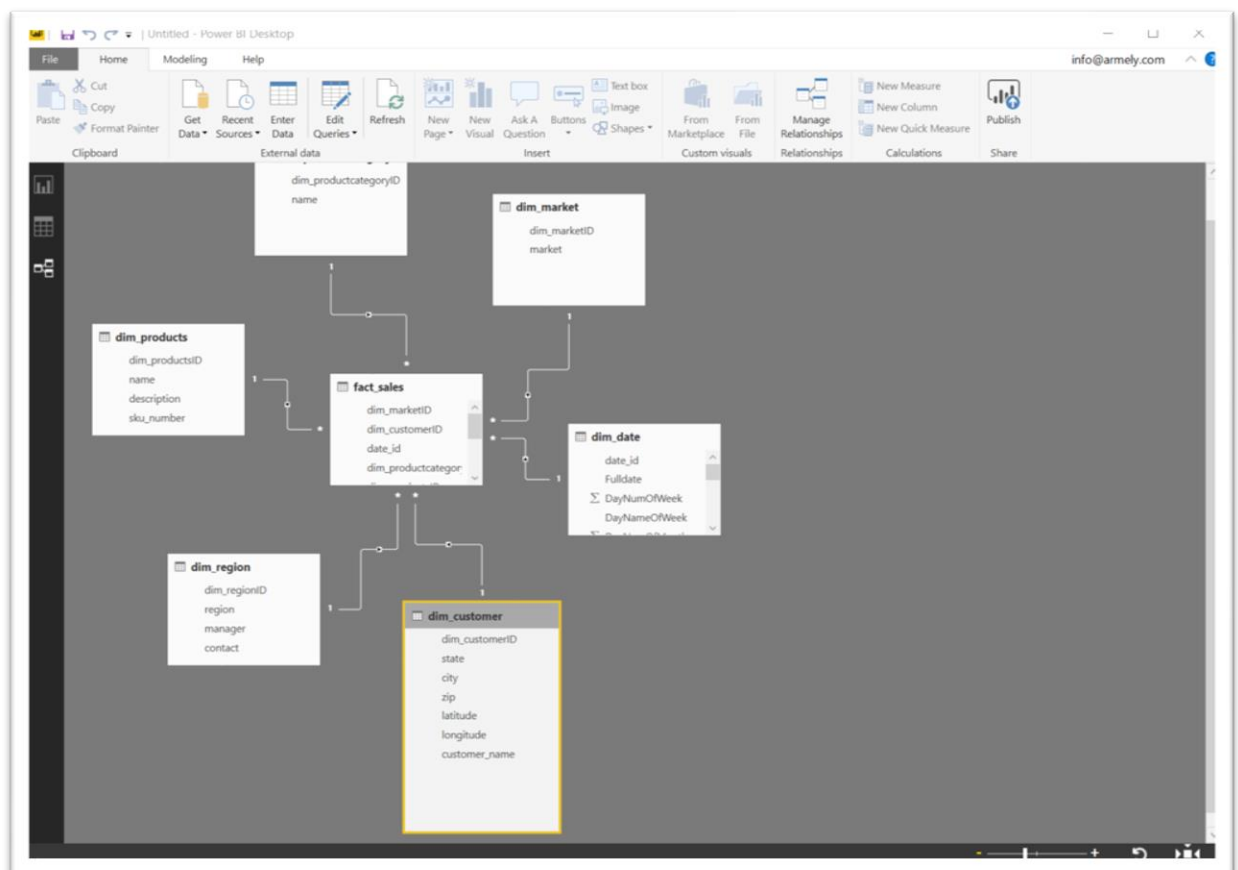
7. Enter the credentials for SQL user account as provided below.
 - a. Username: **training**
 - b. Password: **Th!st@Train**
8. The connection gives you access to database tables **Check** the following tables and **Click Load**



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



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



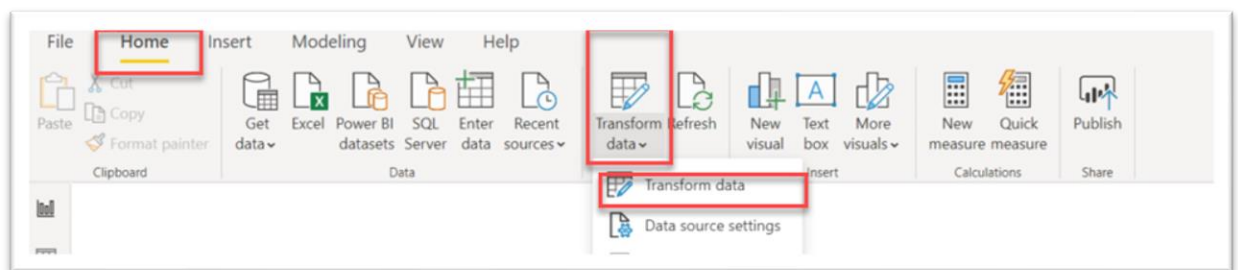
11. Review the data in Data View

	state	city	zip	latitude	longitude	customer_name
1	California	Fresno	96559	73.88975	19.98367	Erat Incorporated
2	Colorado	Colorado Springs	10694	-53.39357	-47.63922	Caraburtt Dicum Industries
3	Maine	Bangor	27220	-38.60279	93.32871	Sed Henderitt Company
4	Oregon	Gresham	51489	-5.62862	-17.64205	Interdum Curabitur Foundation
5	Washington	Olympia	42002	-13.45324	84.29058	Maurit East LLC
6	Ohio	Cleveland	62335	-83.9826	82.89121	Anne Consulting
7	Indiana	Gary	49669	-65.838	-108.3297	Quis Foundation
8	Nevada	North Las Vegas	53118	15.0486	30.86137	Massa Ltd
9	Maryland	Frederick	59228	-29.17207	37.25963	Orti Luctus Associates
10	Iowa	Des Moines	59817	-73.48254	81.09788	Aliquam Acitor Nelli Associates
11	Oklahoma	Norman	64644	26.7937	111.85331	Mulla Integer LLP
12	Washington	Seattle	65293	13.93814	117.25675	Sed Diam Lorem LLC
13	Oregon	Portland	90813	59.90975	18.45692	Tacit Sociosque Ad LLC
14	Arizona	Phoenix	80547	-11.90151	39.22397	Parturient Montes Nascetur Associates
15	Alaska	Fairbanks	99586	-18.96986	157.14668	Eu LLC
16	Connecticut	Hartford	68677	-5.98382	-68.40454	Ullamcorper Magna Sed Ltd
17	Michigan	Warren	32857	-51.59714	-12.61395	Eutemod Est Acca Inc
18	Pennsylvania	Reading	80229	29.94444	-138.95634	Condimentum Inc

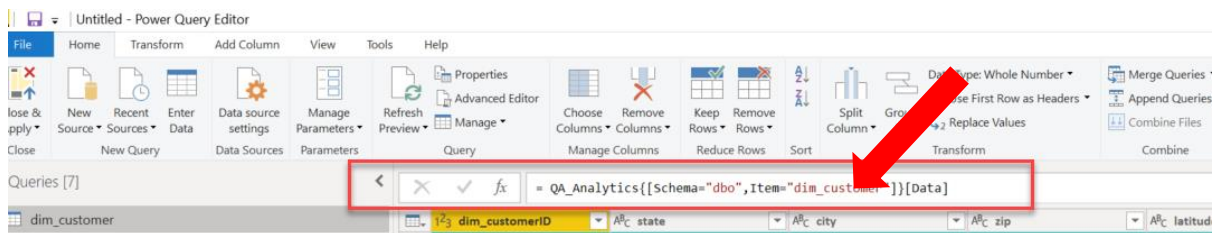
Exercise 2: Reshaping Customer Data

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

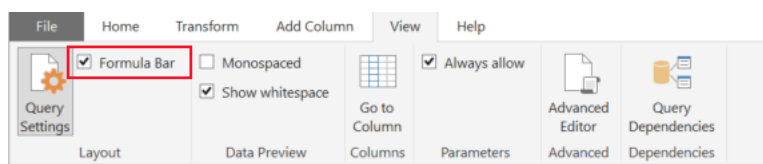
1. Make sure your **product_sales.pbix** project is open.
2. Click on the drop-down next to **Transform Data** and select **Transform data**



3. Ensure that the formula bar is visible



- a. If the formula 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

Queries [7]

- dim_customer
- dim_date
- dim_market
- dim_productcategory
- dim_products
- dim_region**
- fact_sales

fx = Source([Schema="dbo",Item="dim_region"])[Data]

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

- Click on **firstname** column header
- Hold down the **SHIFT** key and click on **lastname** column header

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

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

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

- Copy
- Remove Columns
- Remove Other Columns
- Add Column From Examples...
- Remove Duplicates
- Remove Errors
- Replace Values...
- Fill
- Change Type
- Transform
- Merge Columns**
- Group By...
- Unpivot Columns
- Unpivot Other Columns
- Unpivot Only Selected Columns
- Move

- 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.

Merge Columns

Choose how to merge the selected columns.

Separator
Space

New column name (optional)
manager

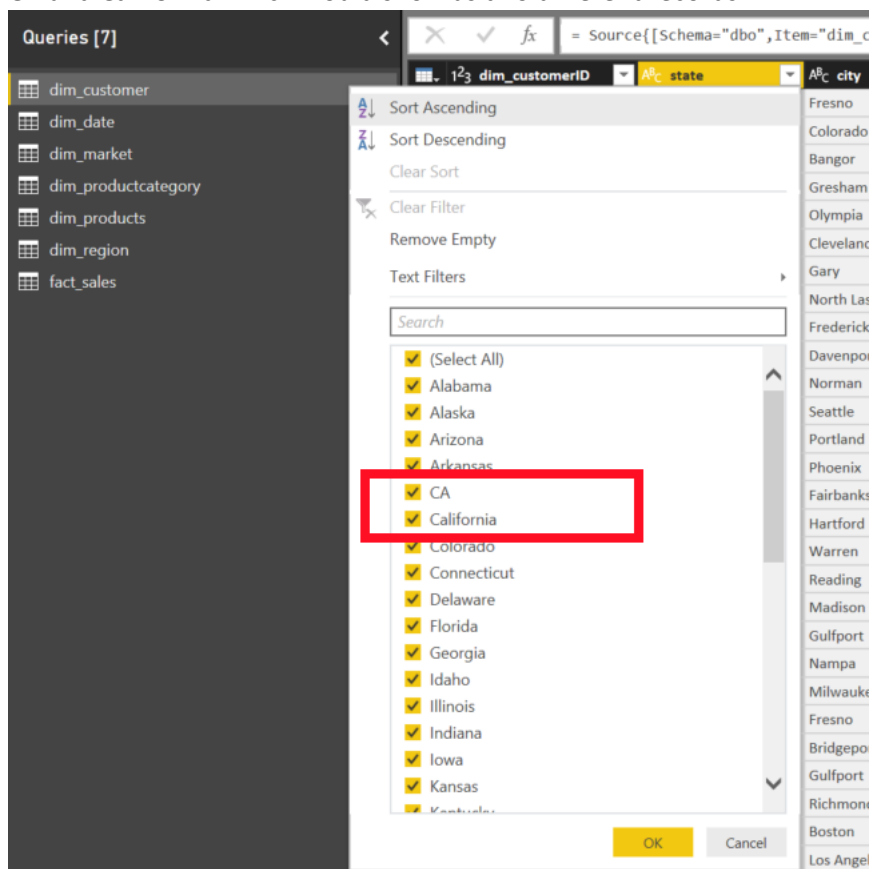
OK Cancel

- f. The outcome should look as follows

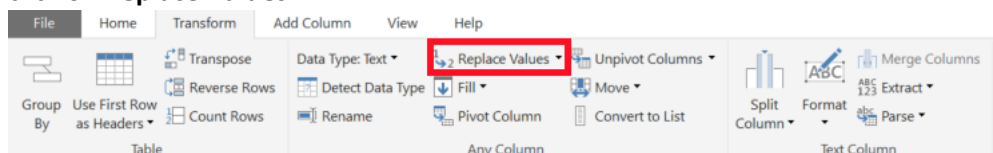
	123 dim_regionID	ABC region	ABC manager	ABC contact
1	1	Midwest	Rosales Nero	nero.rosales@hendreritconsectetuer.com
2	2	Southwest	Hopkins Vivian	vivian.hopkins@hendreritconsectetuer.com
3	3	Northeast	Mcmillan Yoshio	yoshio.mcmillan@hendreritconsectetuer.co...
4	4	West	Johnston Stephen	stephen.johnston@hendreritconsectetuer.c...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

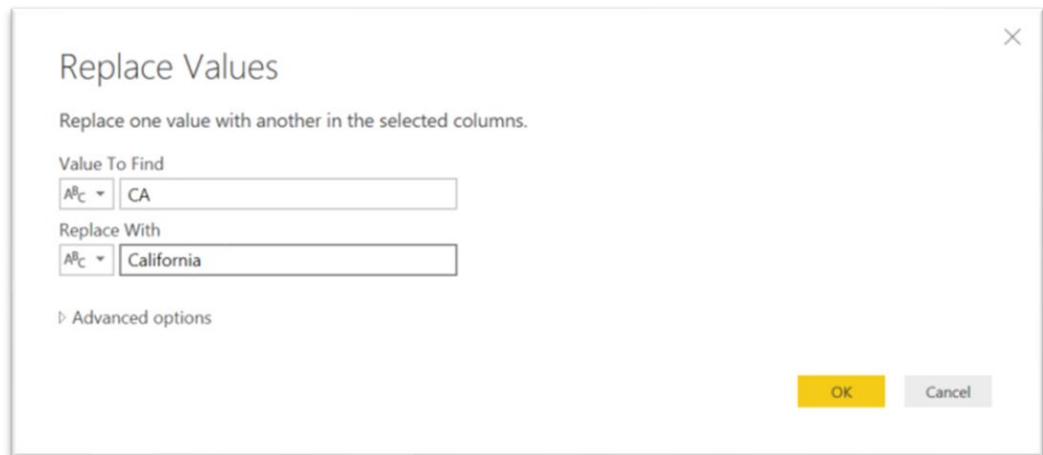
- a. Click on the **dim_customer** table to select it. Inspect the **state** data, you will notice **CA** and **California** which would show as two different records.



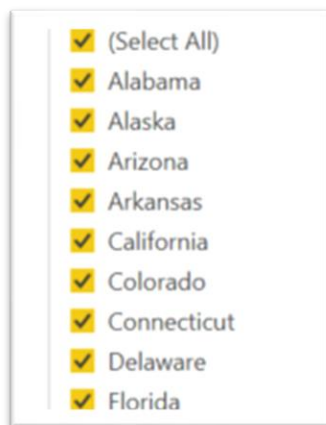
- b. Click on **Transform** tab on the ribbon, make sure the **state** column is highlighted and click on **Replace Values**



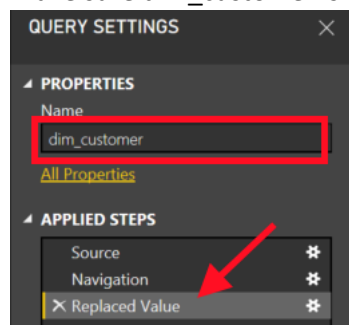
- c. In the **Replace Value dialog**, enter **CA** in the value to find and **California** in **Replace with field**



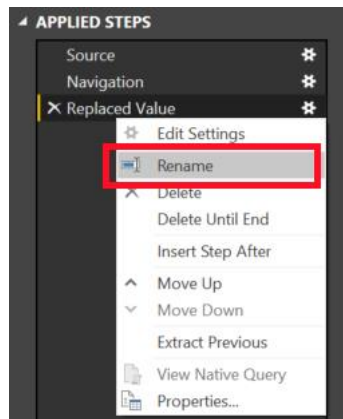
- d. The results should not contain CA



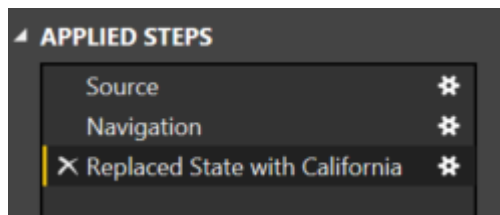
6. It's a good practice to rename the **Applied Steps** to help trace your path in the future. On the right in the **Query Setting**,
- a. Make sure **dim_customer** is the table you are working with



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



c. End result

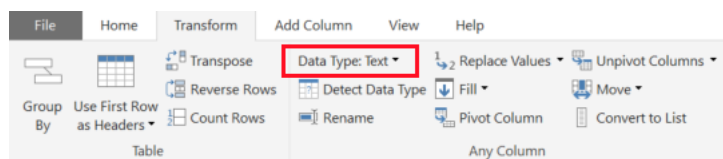


7. In this step, we will fix the **Cost** in **Fact_Sales** table from text to Decimal so that it can accommodate currency

a. Select the **fact_sales** Query on the left

	dim_marketID	dim_customerID	date_id	dim_productcategoryID	dim_productID	dim_regionID	units_sold	cost	revenue
1	4	91	35	1	4	3	219	\$2,404	\$4,089
2	3	13	123	2	6	5	355	\$3,776	\$9,405
3	1	1	180	2	2	2	1972	\$2,898	\$6,694
4	4	37	35	2	1	5	162	\$1,709	\$4,414
5	2	67	198	1	2	4	1133	\$2,309	\$5,276
6	2	25	89	3	6	1	751	\$1,614	\$7,880
7	2	83	25	3	4	2	1713	\$1,747	\$6,125
8	2	30	174	1	1	5	722	\$1,947	\$6,455
9	3	47	57	3	4	3	961	\$4,086	\$7,247
10	1	40	50	2	3	3	908	\$4,082	\$4,142
11	2	18	194	1	4	4	1090	\$4,746	\$3,006
12	4	5	69	3	4	3	102	\$2,943	\$9,578
13	4	37	75	1	1	1	1295	\$3,926	\$5,141

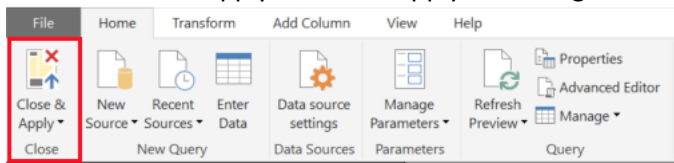
b. Select the header for **cost** column, click on **Transform** and change the **Data Type** from Text to **Decimal**



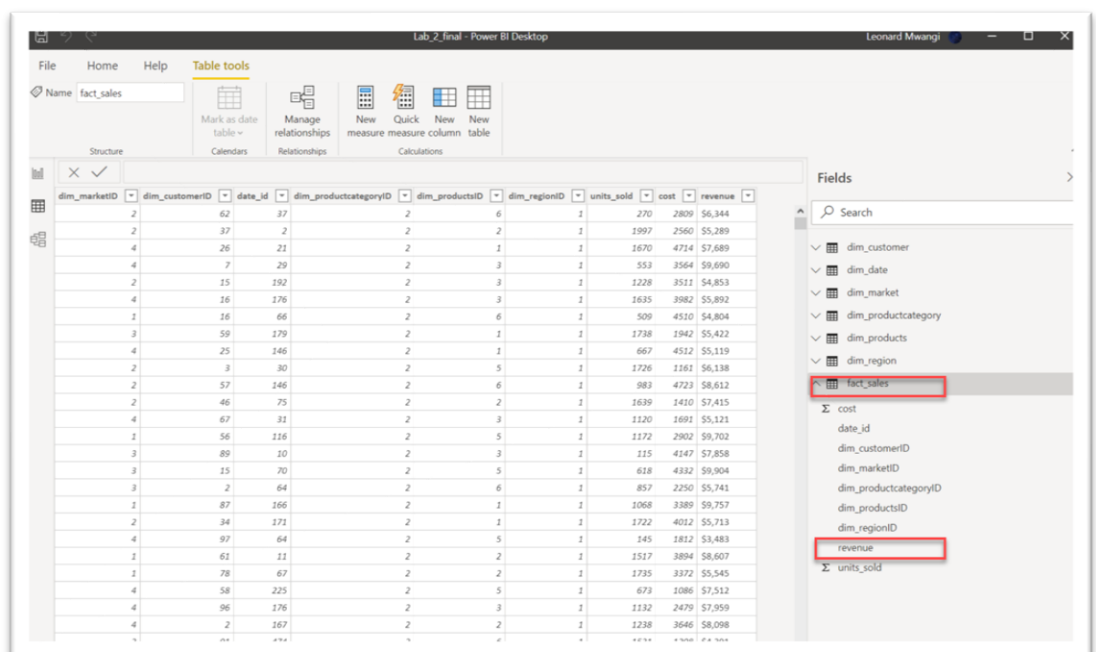
c. Results

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

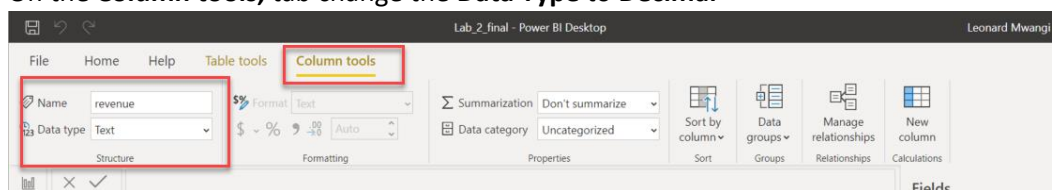
8. Click on Close & Apply button to apply the changes



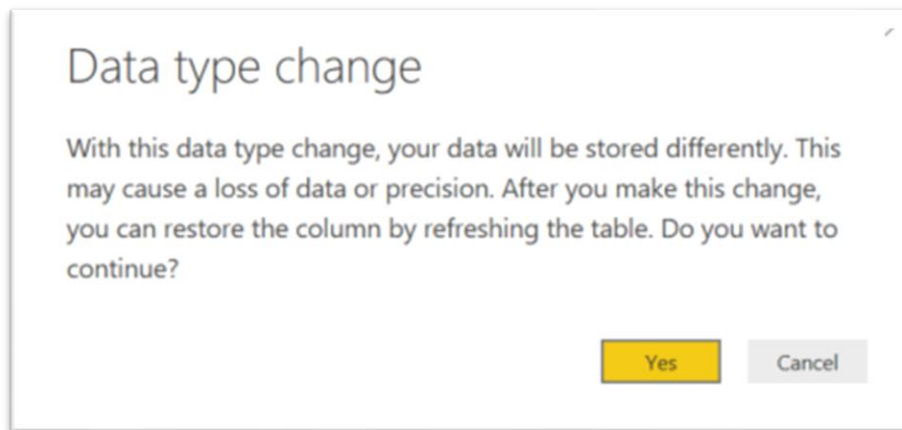
9. Changing the Data Type from Data View
a. Click on **revenue** from left to highlight it



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

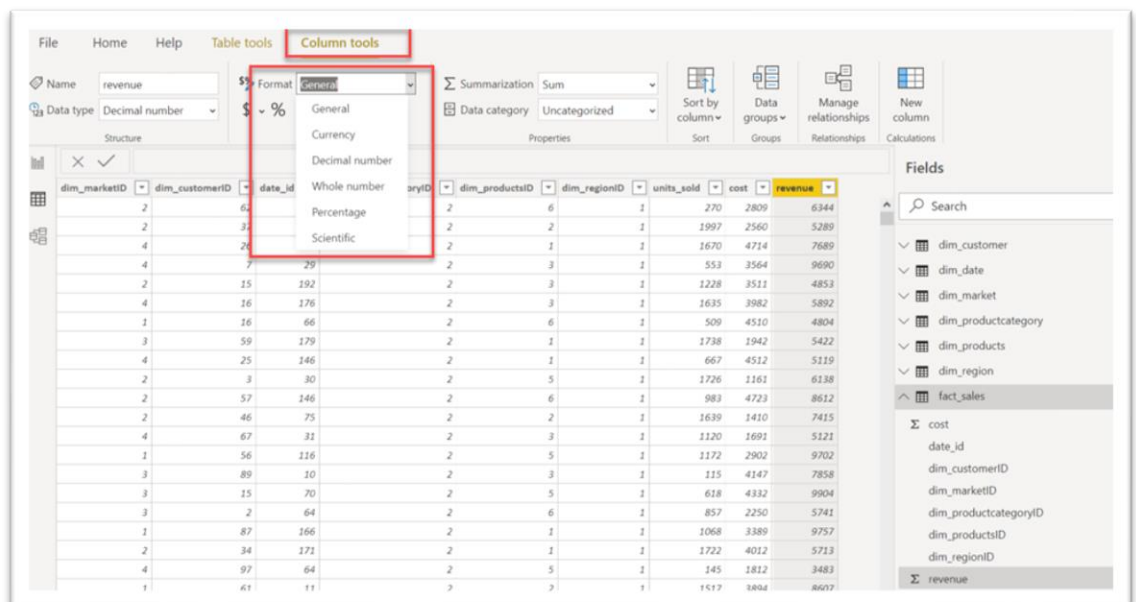


- c. You will get a warning on data change



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.



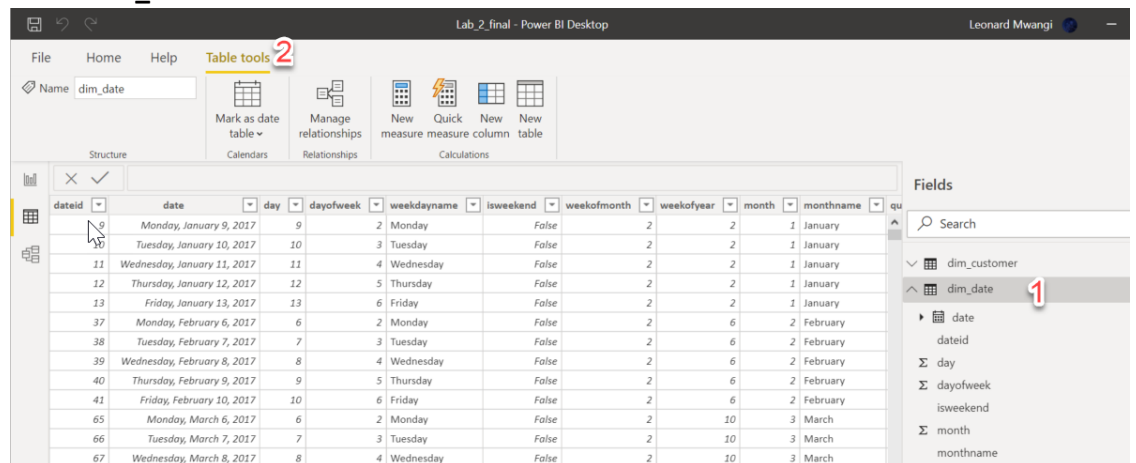
- b. The results should be

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

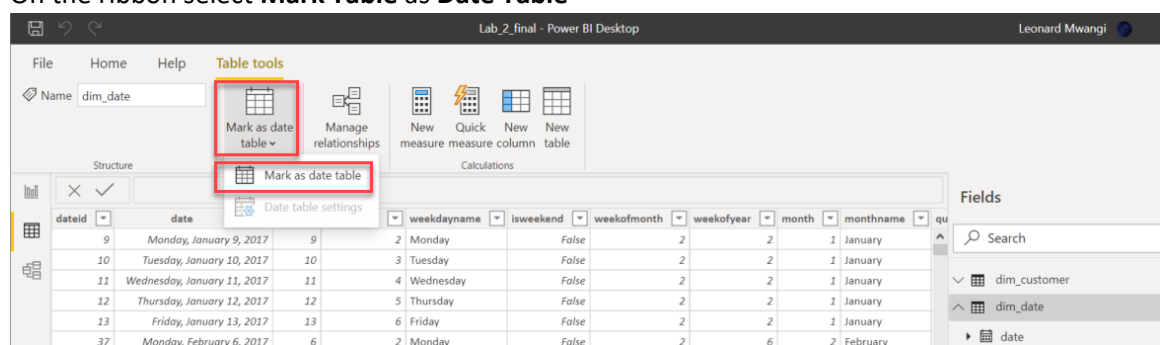
Exercise 3: Enable Time Intelligence

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

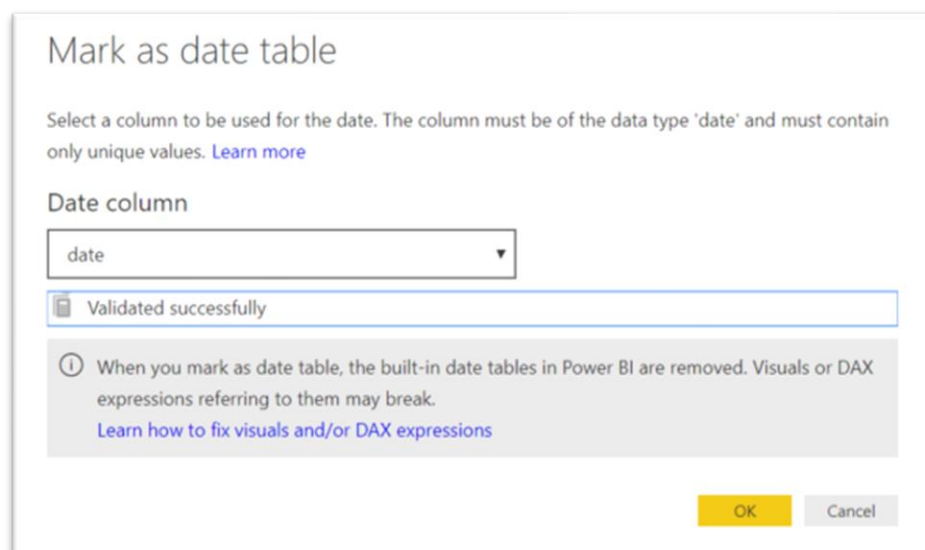
1. In this step, you will mark the dim_date table as a time dimension.
 - a. Select **dim_date** table from the **Data View** then click on **Table tools**



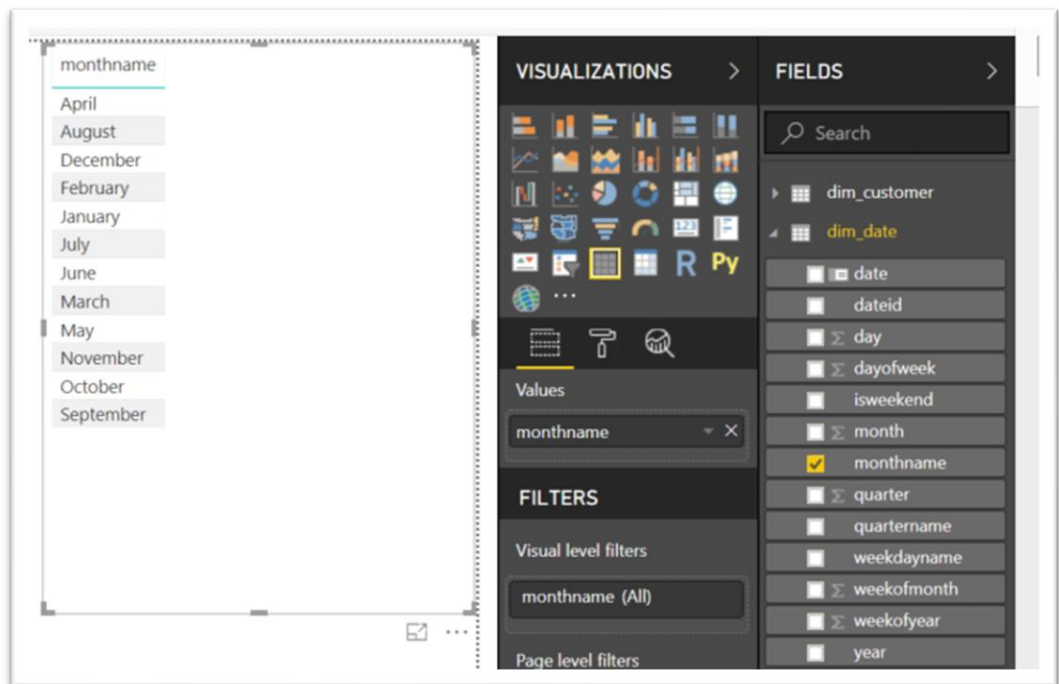
- b. On the ribbon select **Mark Table as Date Table**



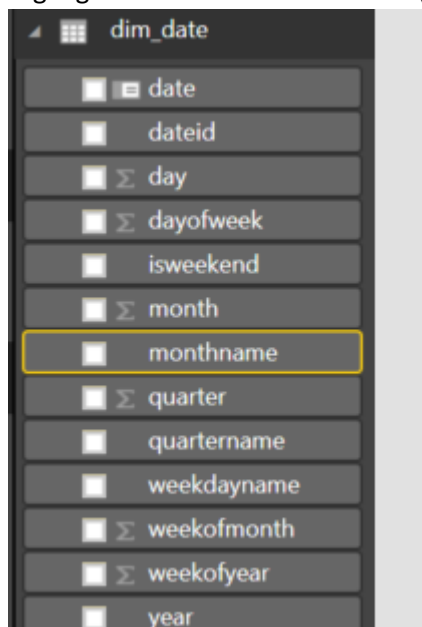
- c. Select date as the Date Column and click OK.



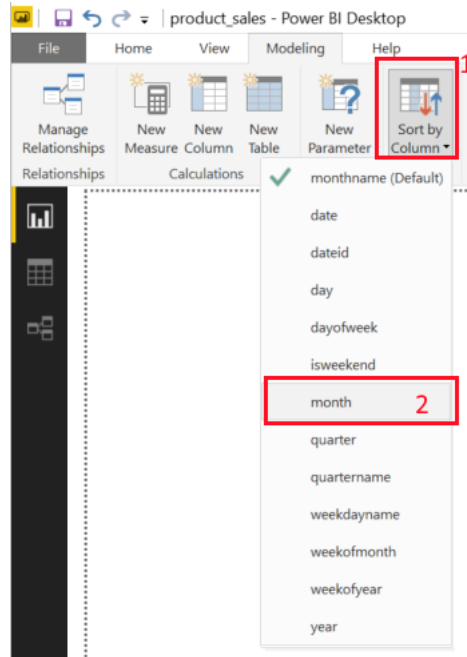
2. In this step, you will fix month sorting from alphabetic to proper sorting.
 - a. Click on **Report View** and put a check mark on **monthname**, you will notice monthname is sorted alphabetically



- 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.

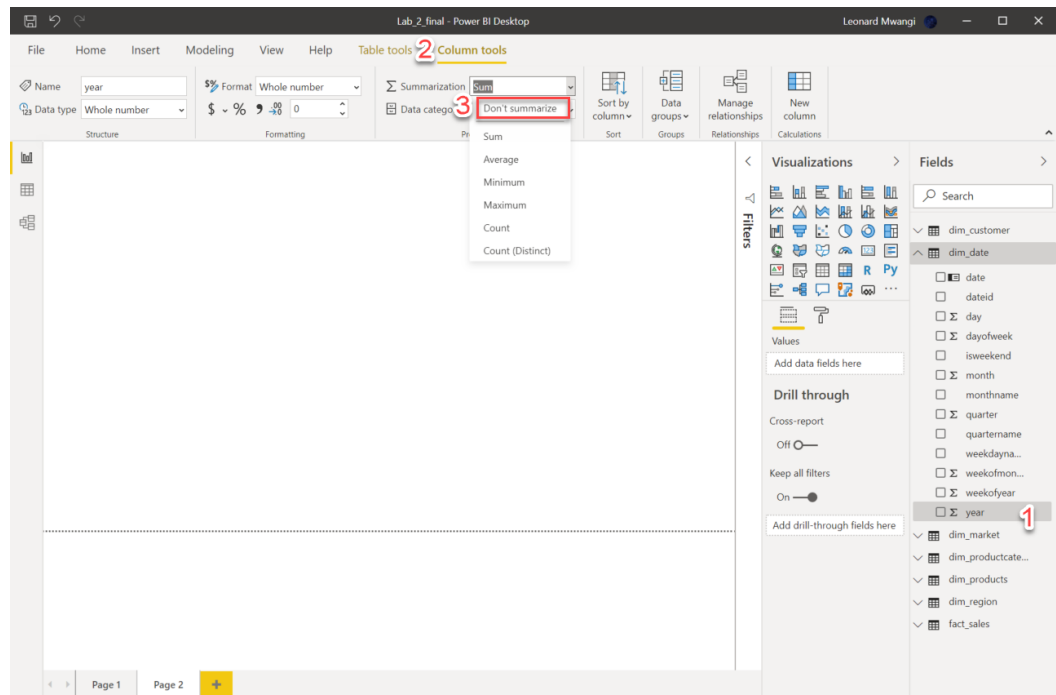
monthname
January
February
March
April
May
June
July
August
September
October
November
December

Exercise 4: Change Column Properties

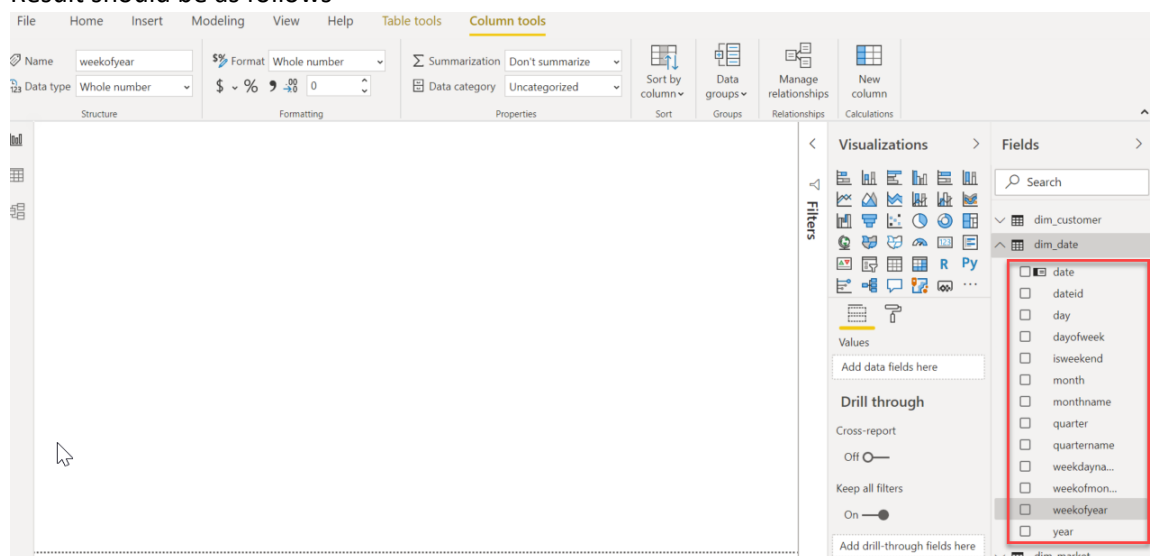
Power BI by default sets any value that's a number to be summable, not all numbers are supposed to be summable. Numbers like years, months, dates should not be summed. In this exercise you will configure some date table columns to not summarize.

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
- Day
 - Dayofweek
 - Month
 - Quarter
 - WeekofMonth
 - 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.