

CSCE 5320: Scientific Data Visualization

Activity 4

Submission Guidelines:

1. Assignment submission is individual.
2. Take screenshots of the entire screen including the date and time.
3. Mark every figure using a number and refer that figure while explaining the figure understanding using 200 - 300 words.
4. The students are required to submit both tutorial and tasks along with the PowerBI Worksheets file.
5. The similarity score for your document should be less than 15%.
6. Please don't submit your work in a zip file – separate file in the same submission.
7. Submission after the deadline is considered as late submission.

Task 1

Tutorial 1: Power Query(Data Transformation and Group By)

Group By feature allows us to aggregate the data based on specific criteria. We will be performing Aggregation functions like sum, average, count, etc. using group by which is like SQL.

Step 1:- Load the 1st Dataset Named Transactions.csv and click on Transform Data.

The screenshot shows the Power BI Desktop interface with the 'transactions.csv' data source loaded. The 'File Origin' is set to '1252: Western European (Windows)', the 'Delimiter' is 'Comma', and 'Data Type Detection' is 'Based on first 200 rows'. A preview of the data is shown in a table format.

date	store_nbr	transactions
1/1/2013	25	770
1/2/2013	1	2111
1/2/2013	2	2358
1/2/2013	3	3487
1/2/2013	4	1922
1/2/2013	5	1903
1/2/2013	6	2143
1/2/2013	7	1874
1/2/2013	8	3250
1/2/2013	9	2940
1/2/2013	10	1293
1/2/2013	11	3547
1/2/2013	12	1362
1/2/2013	13	1102
1/2/2013	14	2002
1/2/2013	15	1622
1/2/2013	16	1167
1/2/2013	17	1580
1/2/2013	18	1635
1/2/2013	19	1369

The data in the preview has been truncated due to size limits.

Buttons: Extract Table Using Examples, Load, Transform Data, Cancel.

Page 1 of 1

Step 2:- Once you click on Transform Data it takes you to the Power Query Editor where you can make changes to the dataset.

Now, Click on New Source Located at the Top and load the rest of the Datasets.

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply Close New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Manage Columns Reduce Rows Sort Split Column Group By Data Type: Date Use First Row as Headers Replace Values Combine

Queries [1] transactions

Table.TransformColumnTypes(#"Promoted Headers",{{"date",

Query Settings

PROPERTIES

Name

transactions

All Properties

APPLIED STEPS

Source

Promoted Headers

Changed Type

3 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 10:50 PM

Step 3:- I Have loaded 3 Datasets or Tables in total where there is no common attributes except the Date Column in the DailyDelhiClimateTest.csv

The screenshot displays the Microsoft Power Query Editor window. The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The Transform tab is active, showing options like Manage Columns, Reduce Rows, Sort, Split Column, Group By, and Transform. The main area shows a data table with the following columns: permitdate, store_nbr, and transactions. The table contains 23 rows of data, with dates ranging from 1/1/2013 to 1/2/2013. The right-hand pane shows the Query Settings for the 'transactions' query, including the Name, All Properties, and Applied Steps (Source, Promoted Headers, Changed Type). A watermark 'Activate Windows' is visible in the bottom right corner.

1	2	3
permitdate	store_nbr	transactions
1/1/2013	25	770
1/2/2013	1	2111
1/2/2013	2	2358
1/2/2013	3	3487
1/2/2013	4	1922
1/2/2013	5	1903
1/2/2013	6	2143
1/2/2013	7	1874
1/2/2013	8	3250
1/2/2013	9	2940
1/2/2013	10	1293
1/2/2013	11	3547
1/2/2013	12	1362
1/2/2013	13	1102
1/2/2013	14	2002
1/2/2013	15	1622
1/2/2013	16	1167
1/2/2013	17	1580
1/2/2013	18	1635
1/2/2013	19	1369
1/2/2013	23	1381
1/2/2013	24	2605
1/2/2013	25	1038

3 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 10:50 PM

Step 4:- Now I want to get the columns min and max values. You can do that by using a query. Right click and click on New Query -> Blank Query. I have renamed the Date to permitdate. You can also rename the attributes name just by clicking on the attribute and renaming it with the other name.

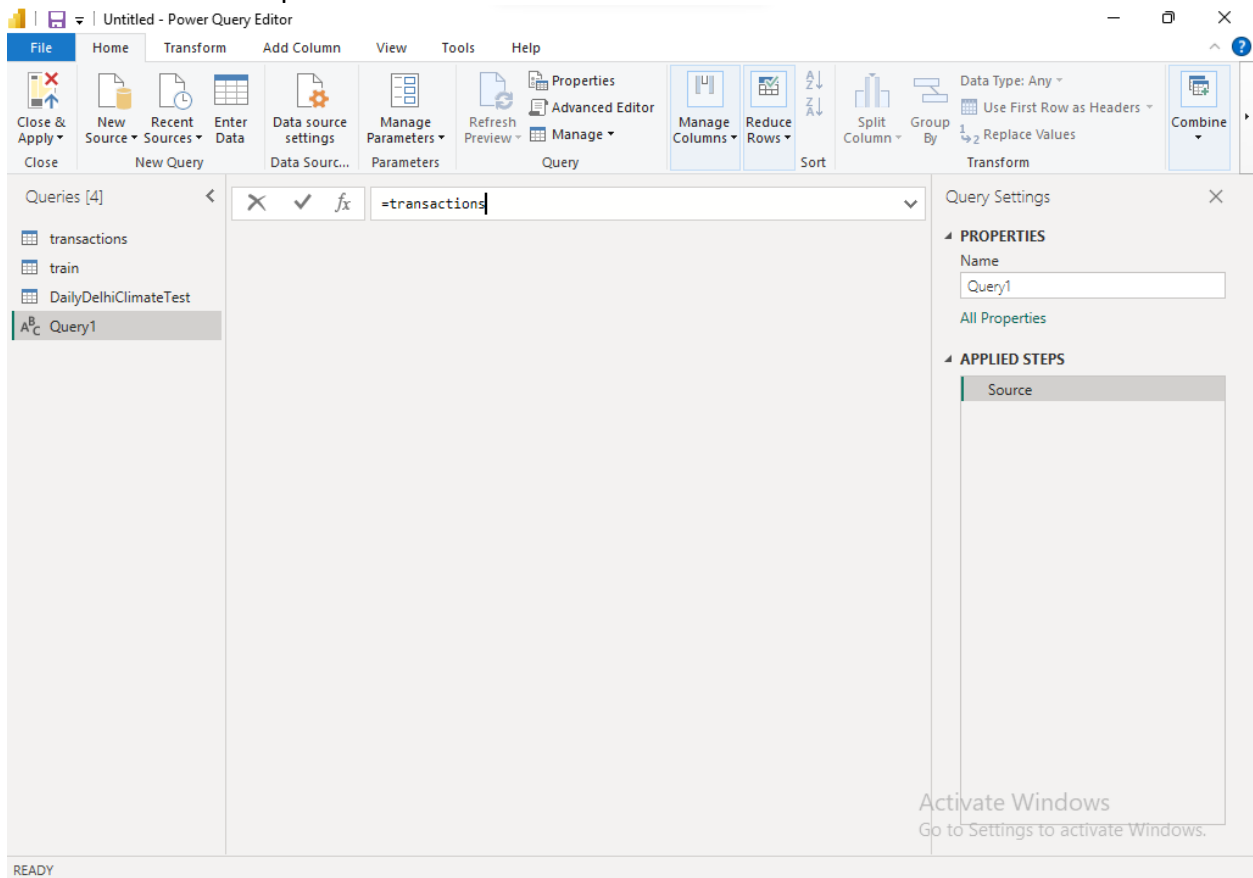
The screenshot displays the Power Query Editor window. The main area shows a table with three columns: 'permitdate', 'store_nbr', and 'transactions'. The 'permitdate' column is highlighted, and a context menu is open, showing options such as 'New Query', 'Blank Query', and 'Recent Sources'. The 'Query Settings' pane on the right shows the 'Renamed Columns' step. The status bar at the bottom indicates '3 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

	permitdate	store_nbr	transactions
1	1/1/2013	25	770
2	1/2/2013	1	2111
3	1/2/2013	2	2358
4	1/2/2013	3	3487
5	1/2/2013	4	1922
6	1/2/2013	5	1903
7	1/2/2013	6	2143
8	1/2/2013	7	1874
9	1/2/2013	8	3250
10	1/2/2013	9	2940
11	1/2/2013	10	1293
12	1/2/2013	11	3547
13	1/2/2013	12	1362
14	1/2/2013	13	1102
15	1/2/2013	14	2002
16	1/2/2013	15	1622
17	1/2/2013	16	1167
18	1/2/2013	17	1580
19	1/2/2013	18	1635
20	1/2/2013	19	1369
21	1/2/2013	23	1381
22	1/2/2013	24	2605
23	1/2/2013	25	1038

3 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 10:50 PM

Step 5:- You can observe a new Query1 added to the Queries tab where I have selected the transactions table to perform modifications.



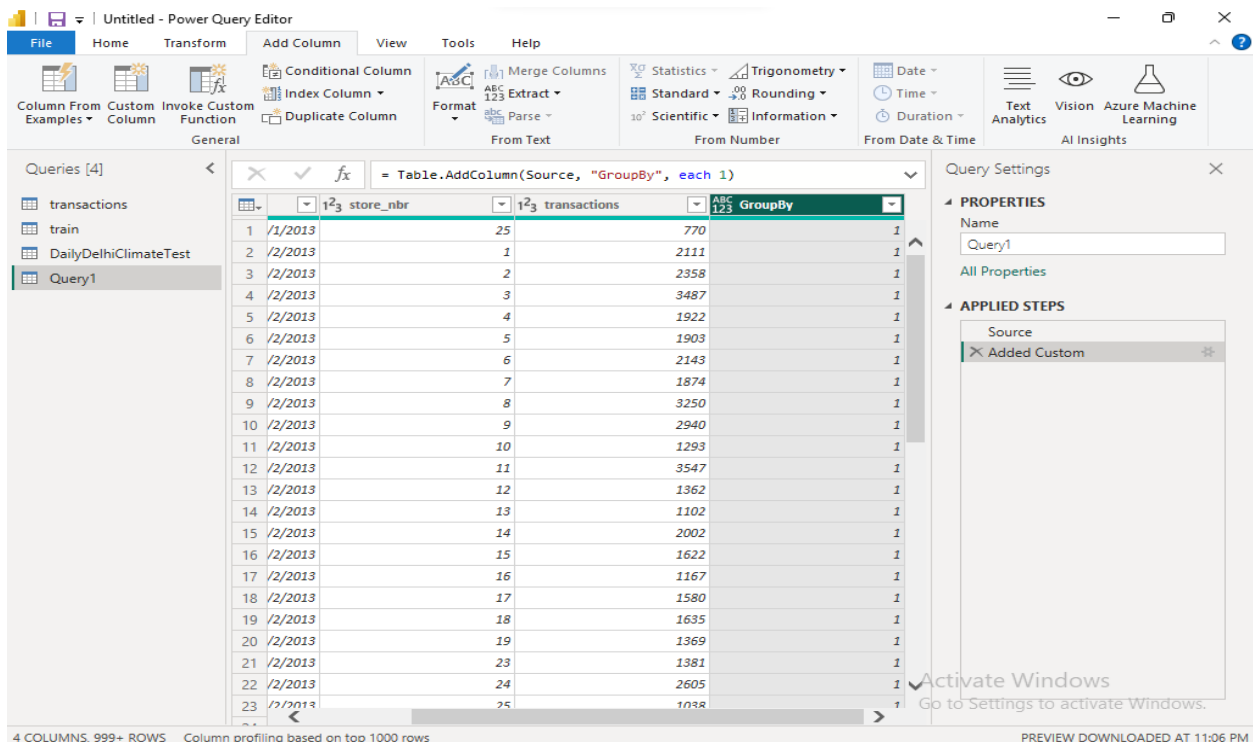
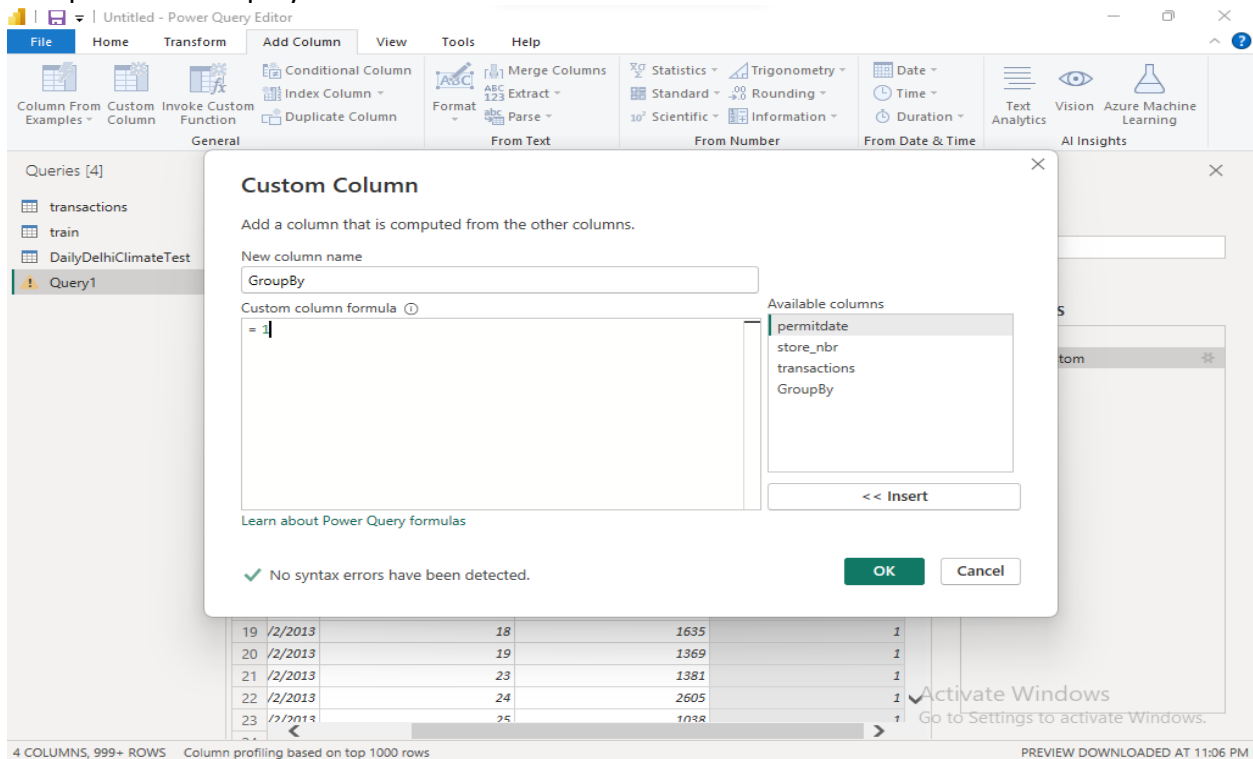
Step 6:- The “=transactions” display the Transactions table. You will be able to find the Applied steps at the right side under Query settings where the steps performed in that table will be displayed. You can click on the X mark on that step if you want to remove that step from the changes made.

The screenshot displays the Power Query Editor window titled "Untitled - Power Query Editor". The ribbon includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The "Add Column" tab is active, showing options like Conditional Column, Index Column, Duplicate Column, Merge Columns, Extract, Format, Parse, Statistics, Trigonometry, Standard, Rounding, Scientific, and Information. The main area shows a table with the following data:

	permitdate	store_nbr	transactions
1	1/1/2013	25	770
2	1/2/2013	1	2111
3	1/2/2013	2	2358
4	1/2/2013	3	3487
5	1/2/2013	4	1922
6	1/2/2013	5	1903
7	1/2/2013	6	2143
8	1/2/2013	7	1874
9	1/2/2013	8	3250
10	1/2/2013	9	2940
11	1/2/2013	10	1293
12	1/2/2013	11	3547
13	1/2/2013	12	1362
14	1/2/2013	13	1102
15	1/2/2013	14	2002
16	1/2/2013	15	1622
17	1/2/2013	16	1167
18	1/2/2013	17	1580
19	1/2/2013	18	1635
20	1/2/2013	19	1369
21	1/2/2013	23	1381
22	1/2/2013	24	2605
23	1/2/2013	25	1038

The right sidebar shows the "Query Settings" pane. Under "PROPERTIES", the Name is "Query1". Under "APPLIED STEPS", the step "Source" is listed. An "Activate Windows" watermark is visible at the bottom right.

Step 7:- In add column, Click on Custom column to create a new column. But our main purpose is to perform Group by. The cells will be inserted with a 1.



Step 8:- Right click on the attribute and click on group By.

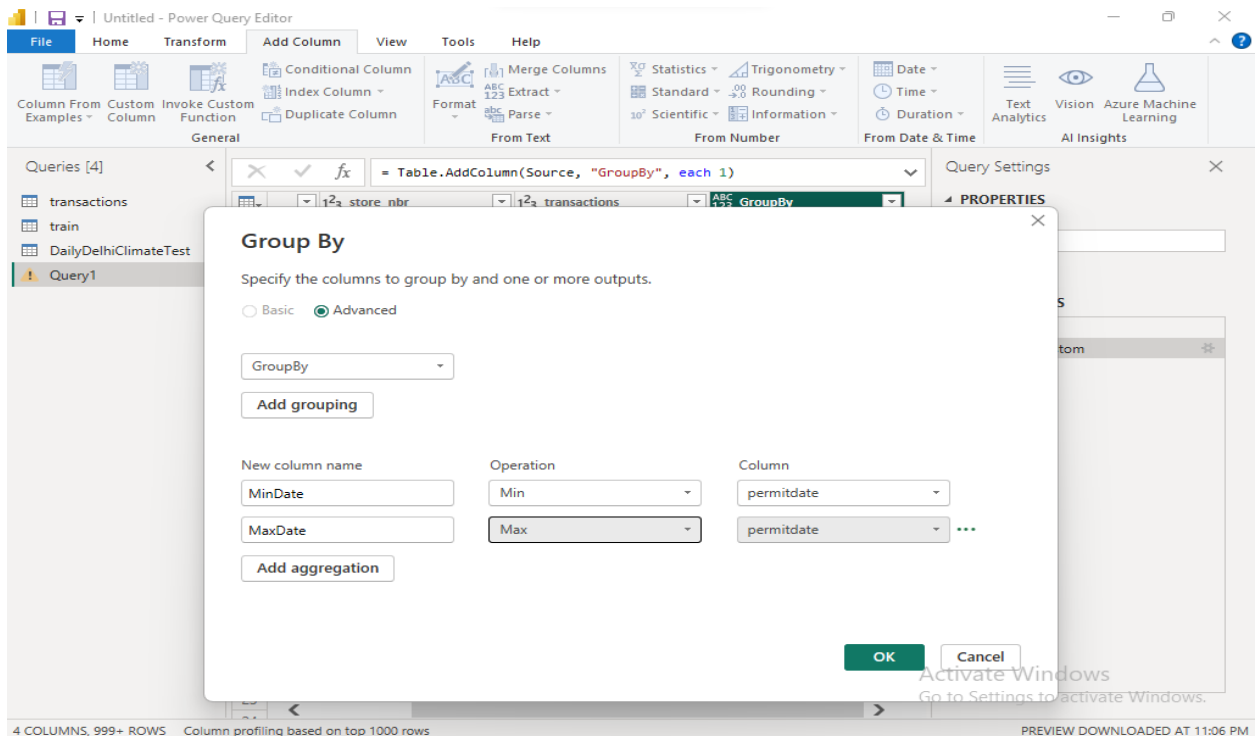
The screenshot shows the Power Query Editor interface. The 'Queries' pane on the left lists 'transactions', 'train', 'DailyDelhiClimateTest', and 'Query1'. The main area displays a table with the following data:

	date	store_nbr	transactions	GroupBy
1	/1/2013	25	770	
2	/2/2013	1	2111	
3	/2/2013	2	2358	
4	/2/2013	3	3487	
5	/2/2013	4	1922	
6	/2/2013	5	1903	
7	/2/2013	6	2143	
8	/2/2013	7	1874	
9	/2/2013	8	3250	
10	/2/2013	9	2940	
11	/2/2013	10	1293	
12	/2/2013	11	3547	
13	/2/2013	12	1362	
14	/2/2013	13	1102	
15	/2/2013	14	2002	
16	/2/2013	15	1622	
17	/2/2013	16	1167	
18	/2/2013	17	1580	
19	/2/2013	18	1635	
20	/2/2013	19	1369	
21	/2/2013	23	1381	
22	/2/2013	24	2605	
23	/2/2013	25	1038	

A right-click context menu is open over the 'GroupBy' column. The menu options include: Copy, Remove, Remove Other Columns, Duplicate Column, Add Column From Examples..., Remove Duplicates, Remove Errors, Change Type, Transform, Replace Values..., Replace Errors..., **Group By...** (highlighted), Fill, Unpivot Columns, Unpivot Other Columns, Unpivot Only Selected Columns, Rename..., Move, Drill Down, and Add as New Query.

4 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 11:06 PM

Step 9:- Go the advanced options and type the same min and max dates as shown in the fig below and select the attribute for which you want the group by function to be applied.



Step 10:- Now the min date and the max date are displayed. From 1/1/2013 and 8/15/2017. You can even round the dates to the nearest day, month, and year. For ex:- 8/15/2017 can be rounded until 8/31/2017.

As my dates are already in order I don't want to sort them again based on the years starting from 2013 to 2017.

The screenshot displays the Power Query Editor interface. The main area shows a table with the following data:

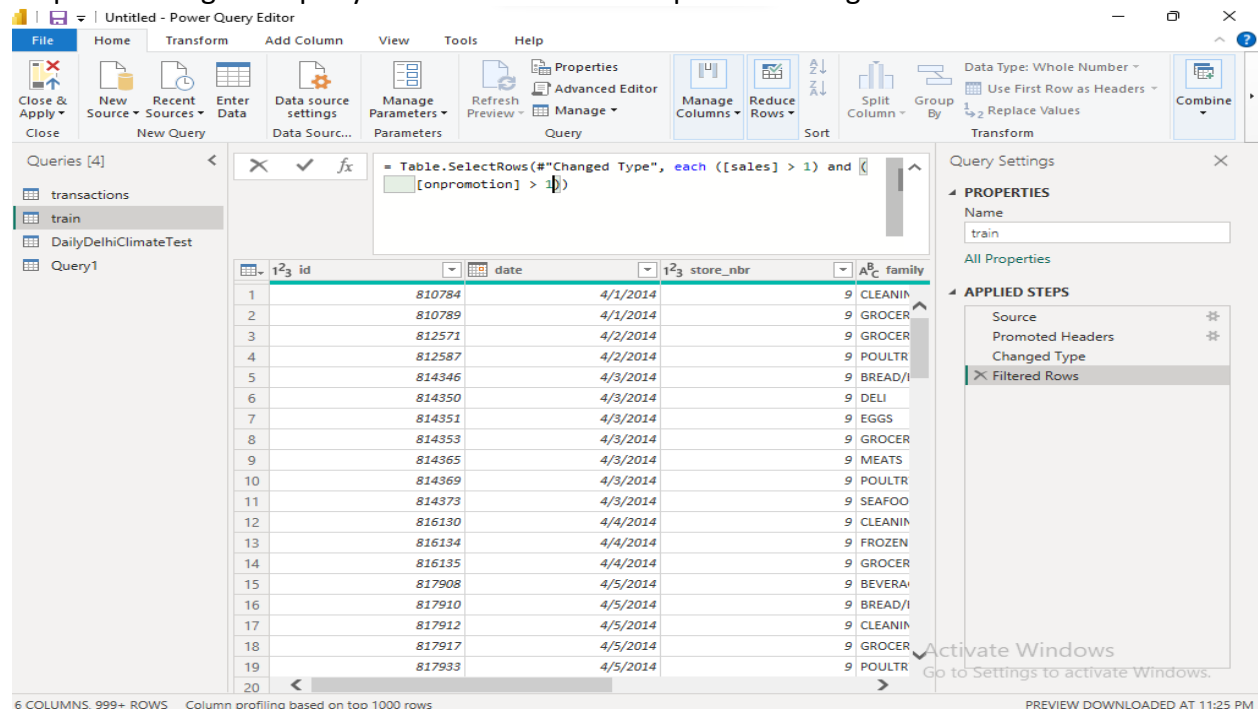
GroupBy	MinDate	MaxDate
1	1/1/2013	8/15/2017

The formula bar at the top indicates the query is defined as: `= Table.Group("#Added Custom", {"GroupBy"}, {{"MinDate", each`

The right-hand pane shows the 'Query Settings' for 'Query1', including the 'APPLIED STEPS' list which contains 'Source', 'Added Custom', and 'Grouped Rows'.

At the bottom of the window, a status bar indicates '3 COLUMNS, 1 ROW' and 'Column profiling based on top 1000 rows'. A watermark at the bottom right reads 'PREVIEW DOWNLOADED AT 11:10 PM'.

Step 11:- Change the query where the sales and onpromotion is greater than 1.



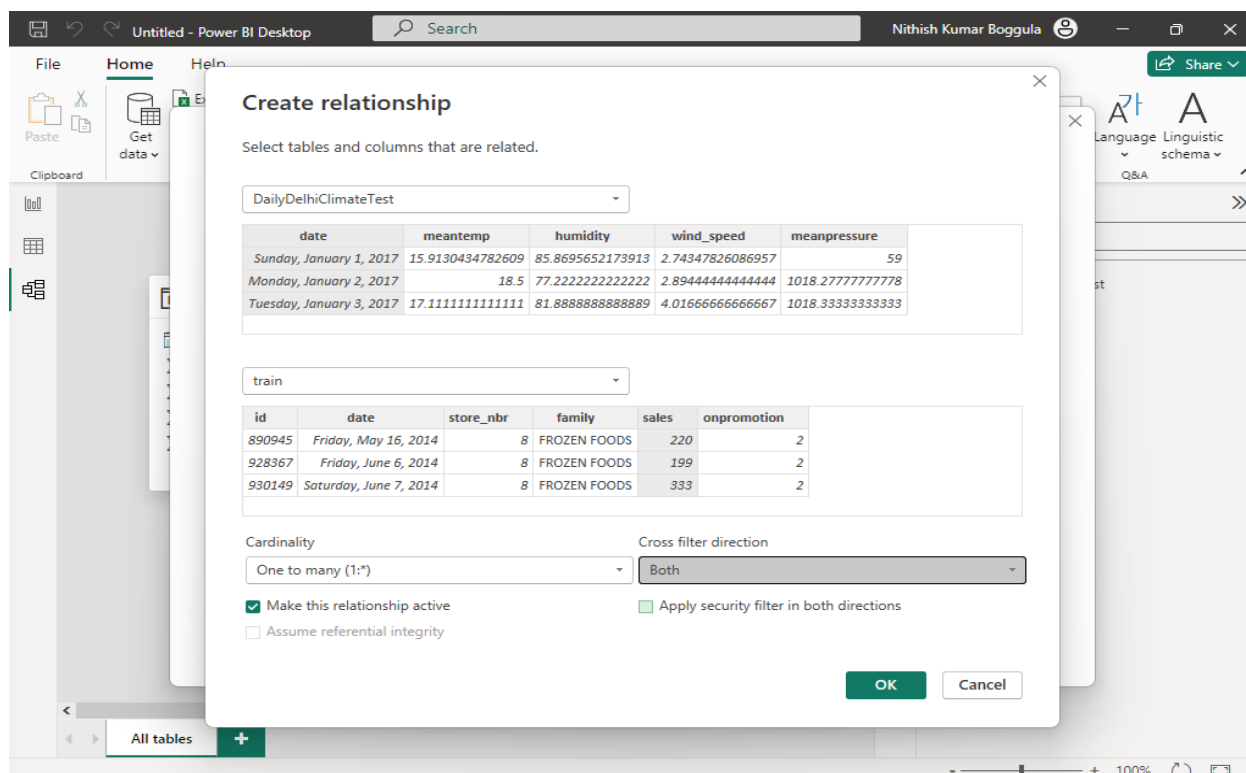
Question 1	Points
<p>Follow the Tutorial 1 and complete the below task for the given attributes.</p> <p>Rename the store_nbr to your desired name to all the tables.</p> <p>Apply group by on any table and perform the same steps mentioned in the tutorial for any other table except train.</p> <p>Use the Query tab as Step 11 and just display the ID's below 700000.</p> <p>Explain any difficulties while performing these operations in detail in about 100 – 200 words.</p>	<p>Visualization – 5 Marks (each)</p> <p>Explanation – 5 Marks</p>
<p>Please answer the following question</p> <p>A. Filter the Dates in Transaction in descending order and explain how you did that along with the screenshot.</p>	<p>Explanation with screenshot – 10 Marks</p>
	<p>Total 20 Marks</p>

Task 2

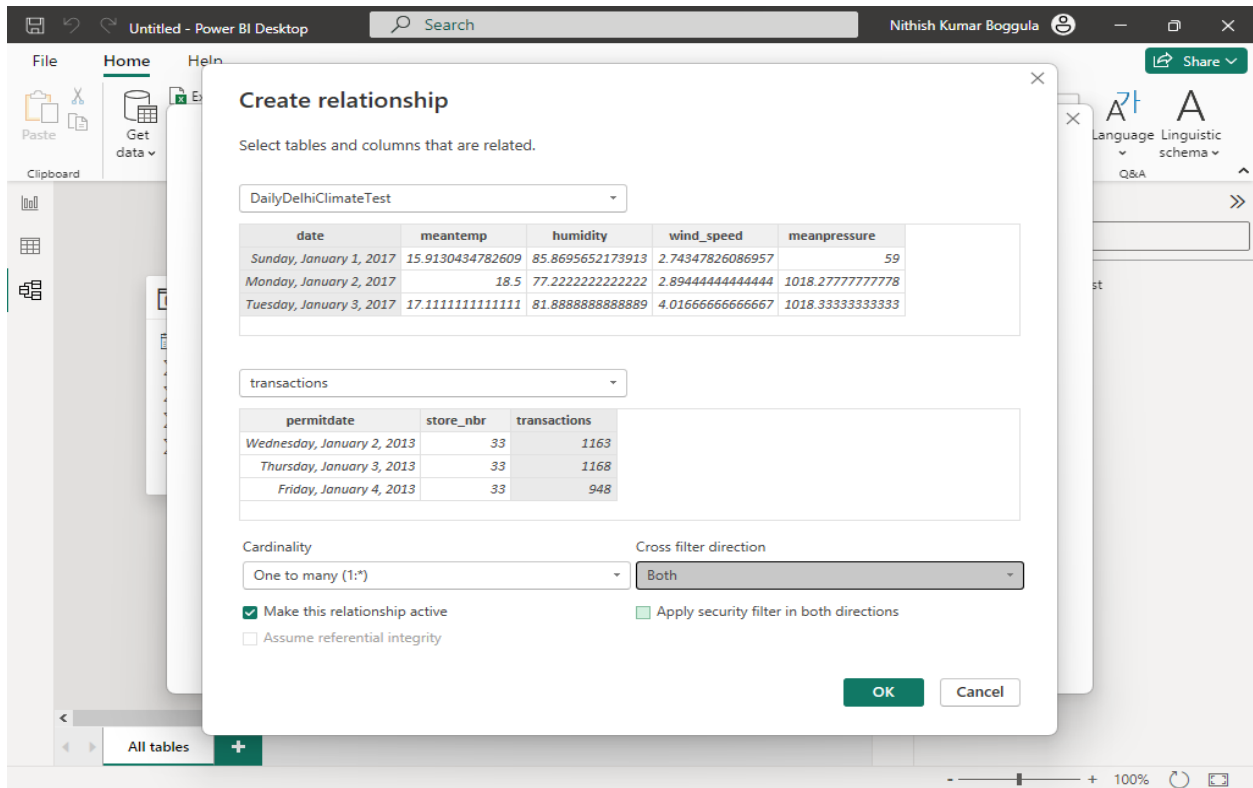
Tutorial 2: Creating Relationships

We will be creating relationship between the attributes and display the cardinality of the selected attributes.

Step 1:- Go to the model view and right click on any table -> Create Relationship. Create all the possible and necessary relations which can be one to one, one to many, etc.....
Follow the below relationships and create relationships between the tables and attributes.



Step 2:- I am creating a connection between DailyDelhiClimateTest table and transactions table.



Step 3:- We are creating a relationship between train and transaction table. I have selected 1 attribute from each table to link the tables. The cardinality is automatically populated based on the data and attributes selected.

Untitled - Power BI Desktop

Search

Nithish Kumar Boggula

Share

Create relationship

Select tables and columns that are related.

train

id	date	store_nbr	family	sales	onpromotion
890945	Friday, May 16, 2014	8	FROZEN FOODS	220	2
928367	Friday, June 6, 2014	8	FROZEN FOODS	199	2
930149	Saturday, June 7, 2014	8	FROZEN FOODS	333	2

transactions

permitdate	store_nbr	transactions
Wednesday, January 2, 2013	33	1163
Thursday, January 3, 2013	33	1168
Friday, January 4, 2013	33	948

Cardinality: Many to many (*)

Cross filter direction: Both

☐ Make this relationship active

☐ Assume referential integrity

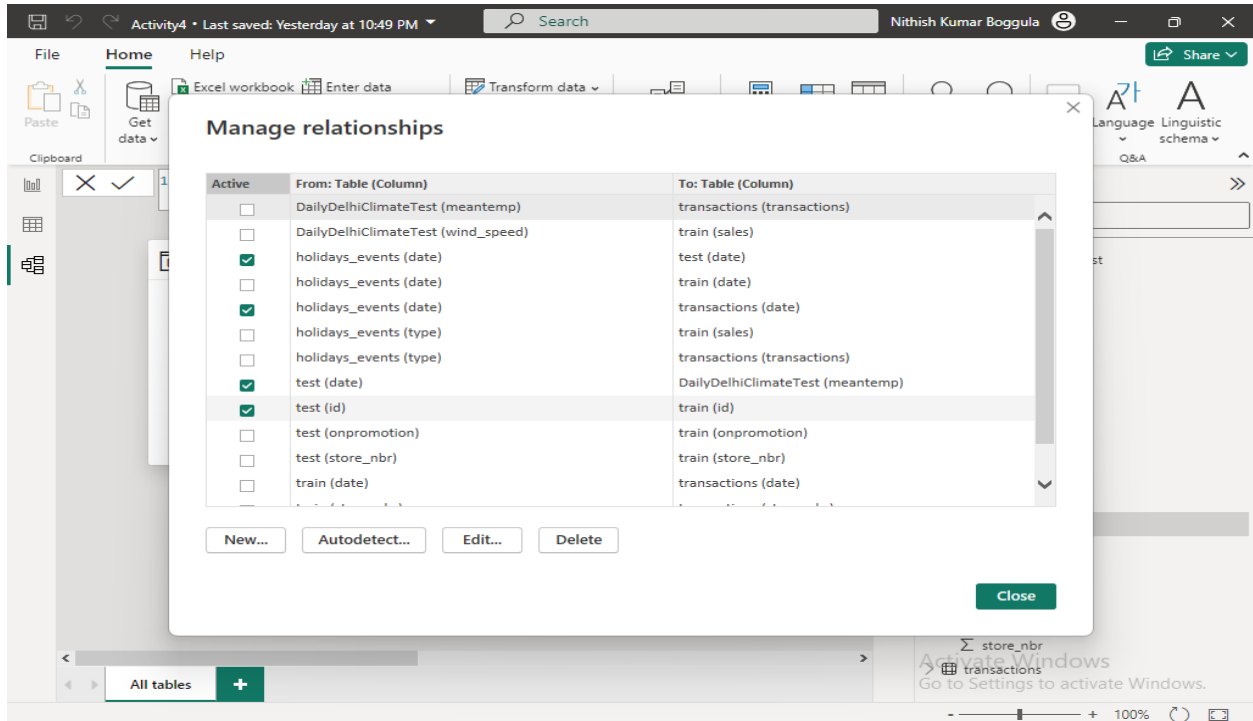
☐ Apply security filter in both directions

! This relationship has cardinality Many-Many. This should only be used if it is expected that neither column (sales and transactions) contains unique values, and that the significantly different behavior of Many-many relationships is understood. [Learn more](#)

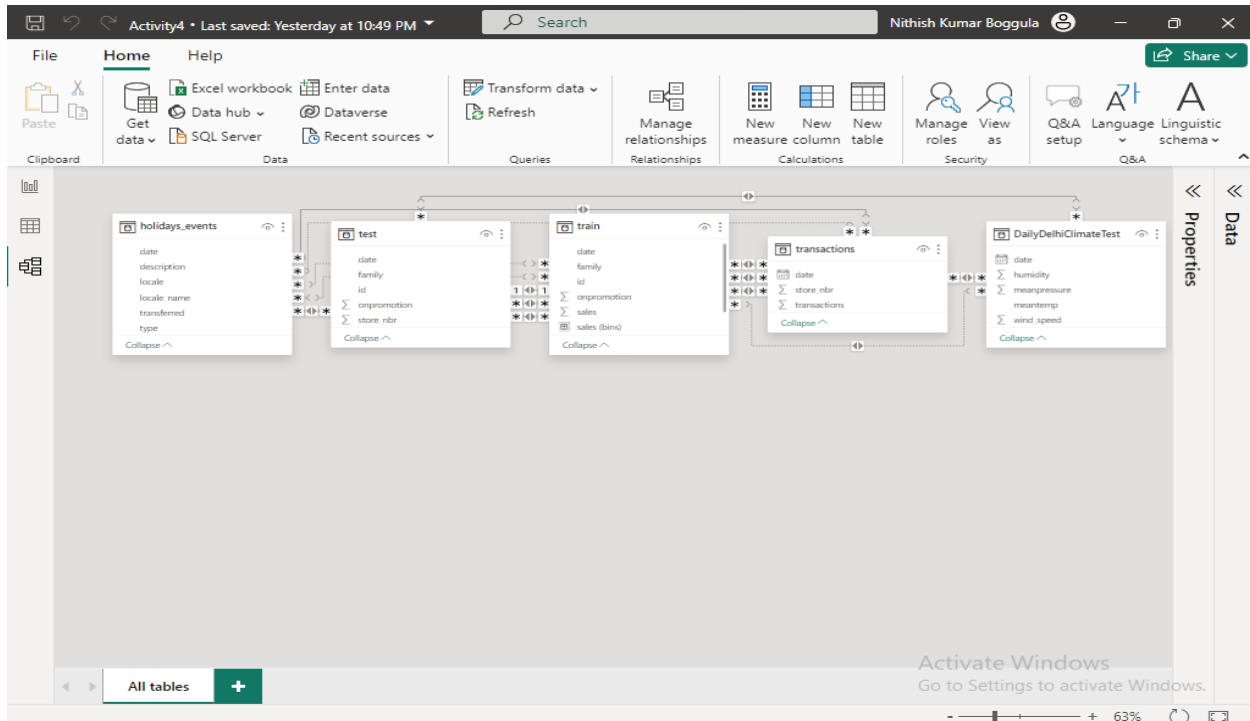
OK Cancel

Step 4:- The created relationships can be viewed using the manage relationships where the active relationships give us that a connection between a table is active. Only a single connection between tables can be active.

For ex:- you can't create 2 active relationships between the tables train and transactions and vice versa.



Step 5:- The model view of the tables can be viewed after the relationships between the tables are established.

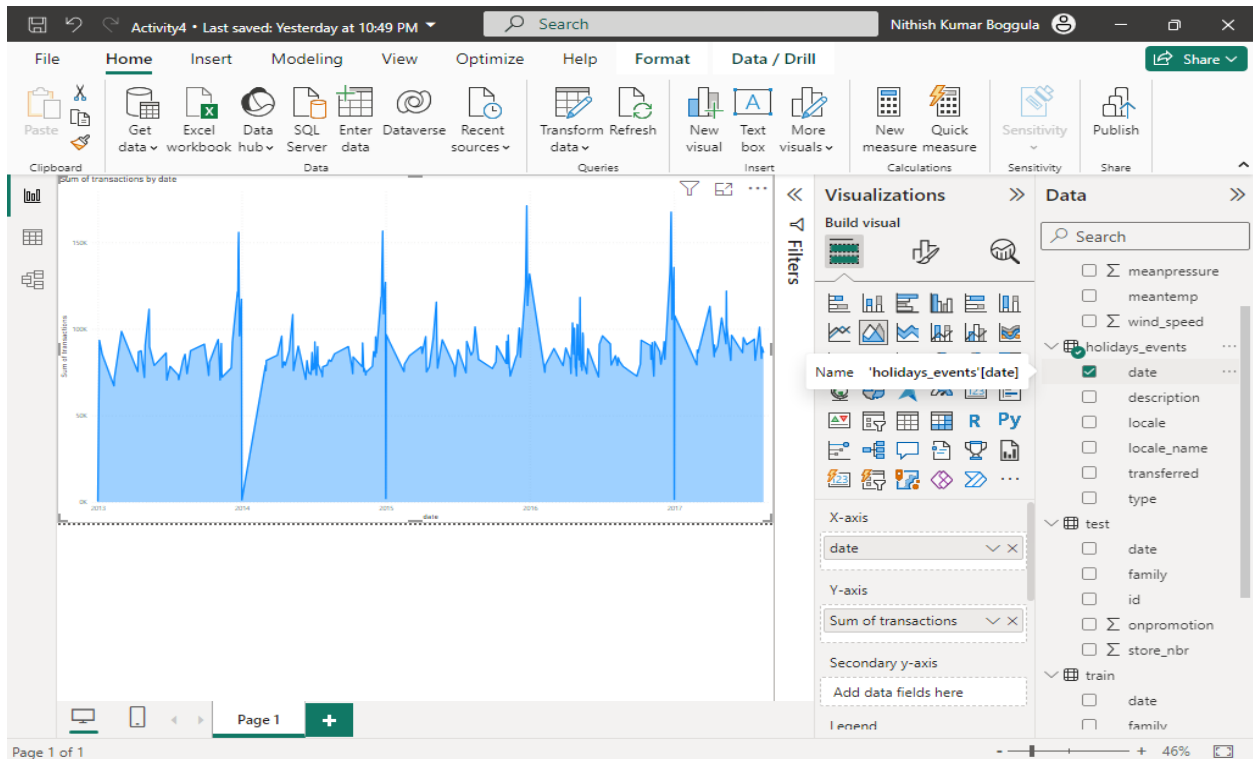


Question 2	Points
<p>Follow the above tutorial and try to find all the possible relationships for the tables.</p> <p>Find all the one to one and many to one relationship among the attributes.</p> <p>Will cardinality of the relationship really affect the relationship and the way the attributes interact with the other?</p>	<p>Explanation with Screenshot – 10 Marks</p>
<p>Please answer the following Questions</p> <p>A. What happens if you select cross filter direction as single and filter based on the attributes?</p> <p>B. Can we use multiple relations in the table and set them as active? Please explain your opinion.</p> <p>(Screenshot Mandatory)</p>	<p>Explanation with Screenshot – 10 Marks</p>
	<p>Total – 20 Marks</p>

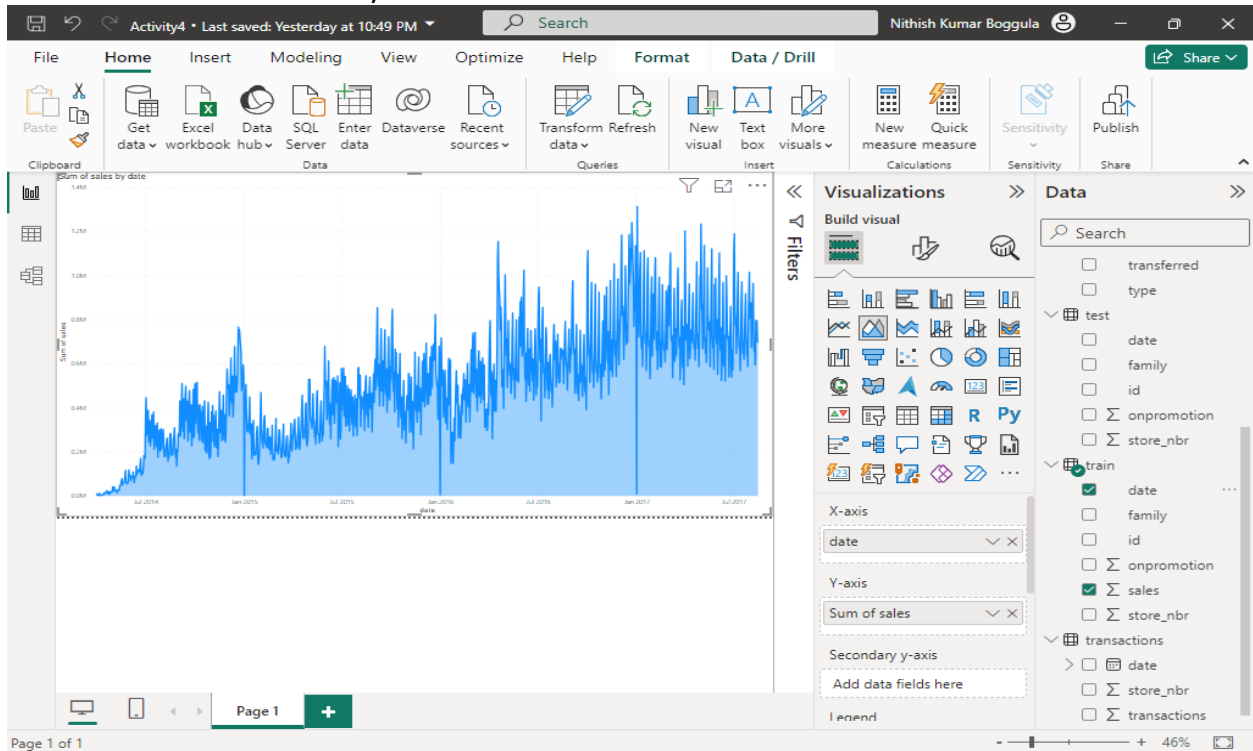
Task 3

Tutorial 3: Visualizing using different tables

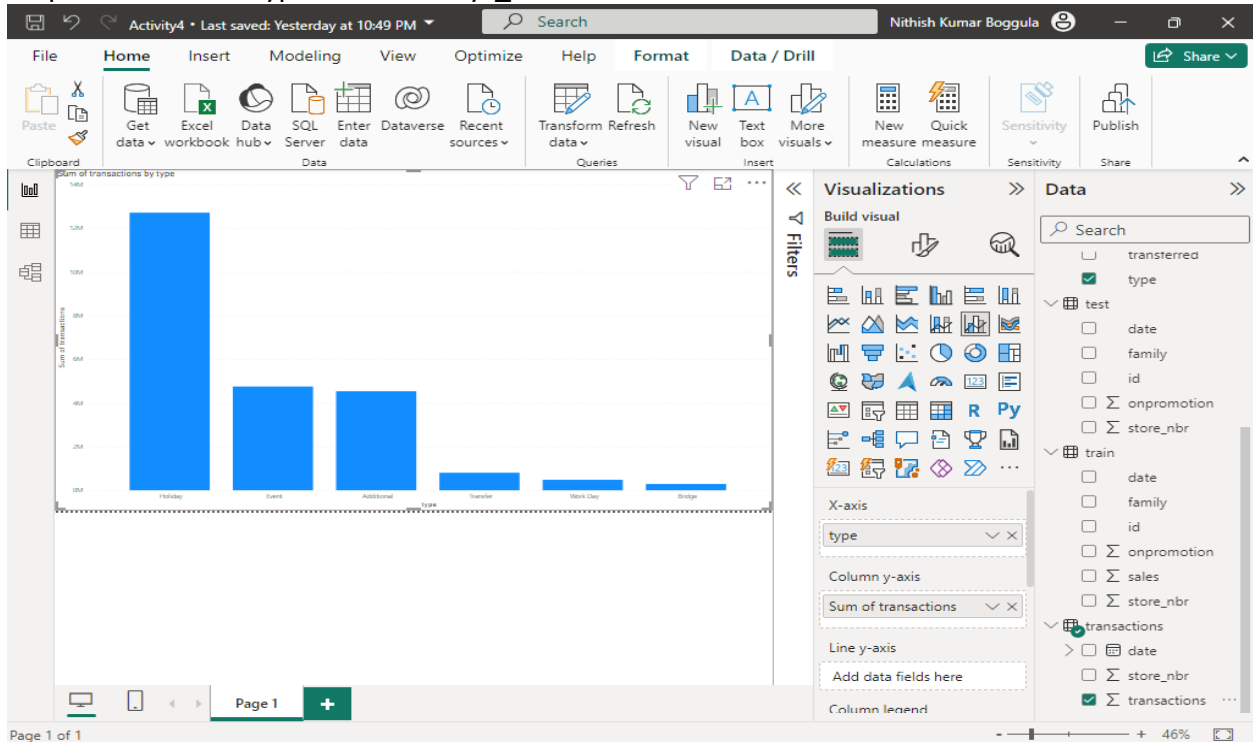
Step 1:- Select the date from the holidays_events table and transactions from the transaction table and click on the Area chart to visualize the sum of transactions with respect to years.



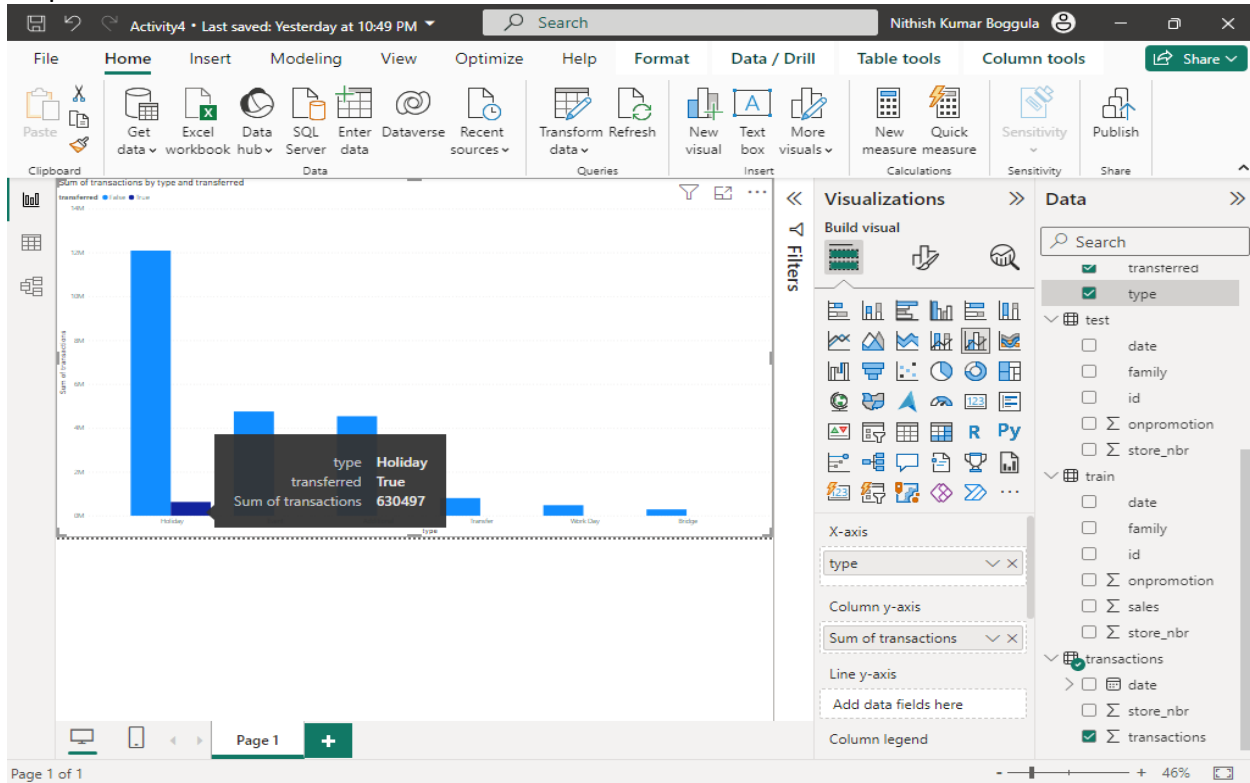
Step 2:- Now I am selecting the sales and date from the same train table and use the area chart to visualize the sales across years.



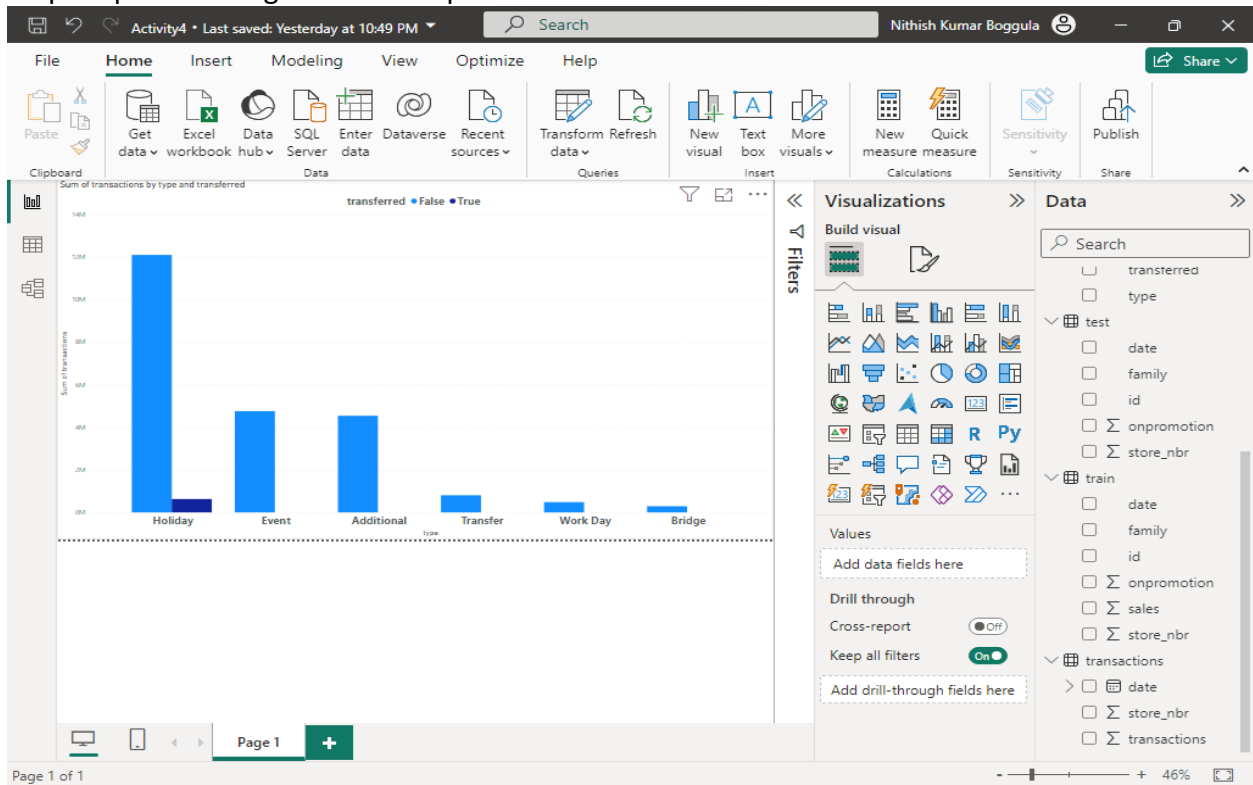
Step 3:- Select the Type from holidays_events and transactions and select the visualization.



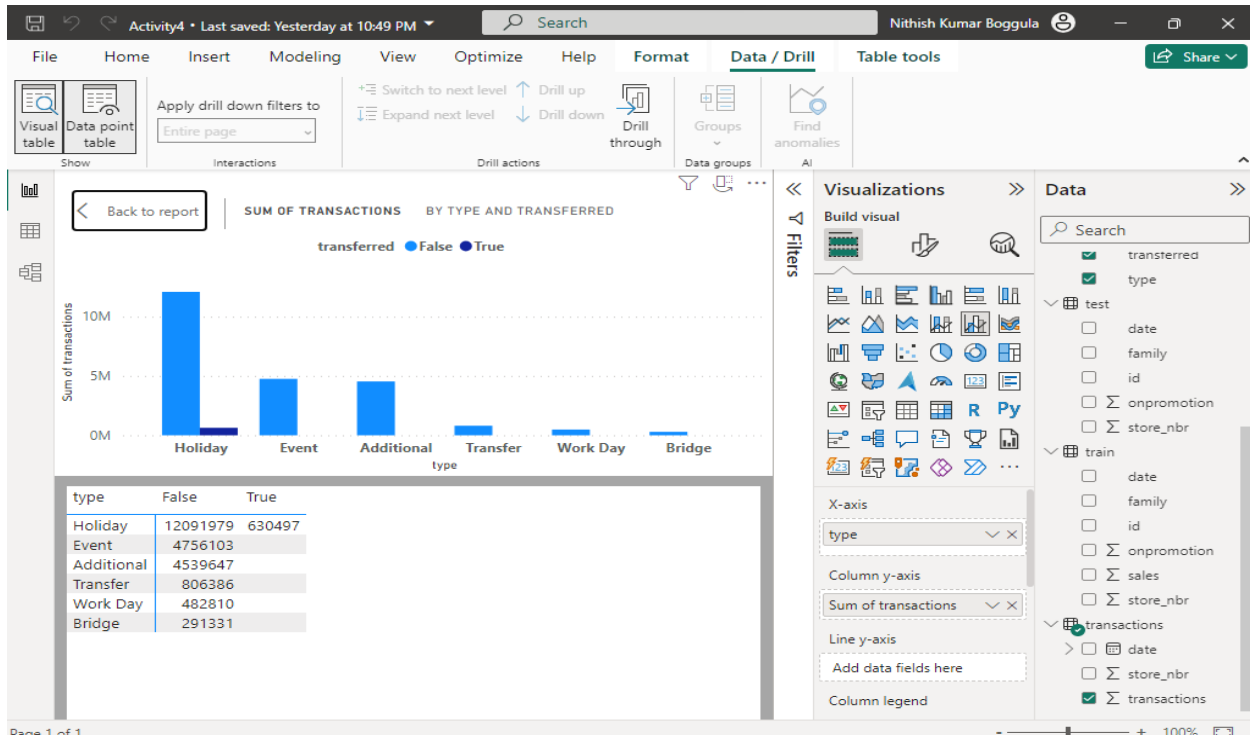
Step 4:- Now add another attribute i.e. Transferred



Step 5:- place the legend at the top center



Step 6:- Click on visual table in Data/Drill to get the information about the graph in tabular format.



Question 3	Points
<p>Follow the Tutorial 3 and complete the below task for the given attributes.</p> <p>Select the attributes Transferred, Type and Date from Holidays_events and transactions from transactions table.</p> <p>Take another page and select date, sales, and store_nbr from train and use a clustered column chart and explain both the attributes roles in the visualization and explain your findings.</p> <p>Use at least 3 visualizations and provide a detailed understanding. (Note:- Visualizations Mandatory)</p>	<p>Explanation – 5 Marks</p> <p>Visualization – 5 Marks</p>
<p>Display the Visual Table for the above attributes in separate pages.</p> <p>Answer the following questions</p> <p>A. What is the Highest sales in 2016 and How much compared to the 2nd Highest?</p> <p>B. Why does the attributes work differently for different visualizations?(Explain the question based on the attributes given in the above question)</p> <p>(Visualizations and screenshots mandatory to support your answer)</p>	<p>Explanation with screenshots – 15 Marks</p>
	<p>Total – 25 Marks</p>

Tutorial 4

Tutorial 4: Creating an attribute to group the attributes

In this tutorial we will be creating a group for the selected attribute by using the below query. So, use the default bin settings and complete the tutorial.

Step 1:- Here we will be doing forecasting on various attributes.

Click on Custom Column and give the below query to get the mean. If you want to insert the attributes then just click on the attributes and click on insert to use them in the query.

Activity4

File Home Transform Add Column View Tools Help

Column From Custom Invoke Custom
Examples Column Function

General

Conditional Column
Index Column
Duplicate Column

Format
Merge Columns
Extract
Parse

Statistics
Standard
Scientific

Trigonometry
Rounding
Information

Date
Time
Duration

Text
Analytics

Vision
Azure Machine
Learning

AI Insights

Queries [5]

test
train
transactions
holidays_events
DailyDelhiClimateTest

Custom Column

Add a column that is computed from the other columns.

New column name
Mean

Custom column formula
= AVERAGE(DailyDelhiClimateTest[meantemp])

Available columns
date
meantemp
humidity
wind_speed
meanpressure

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

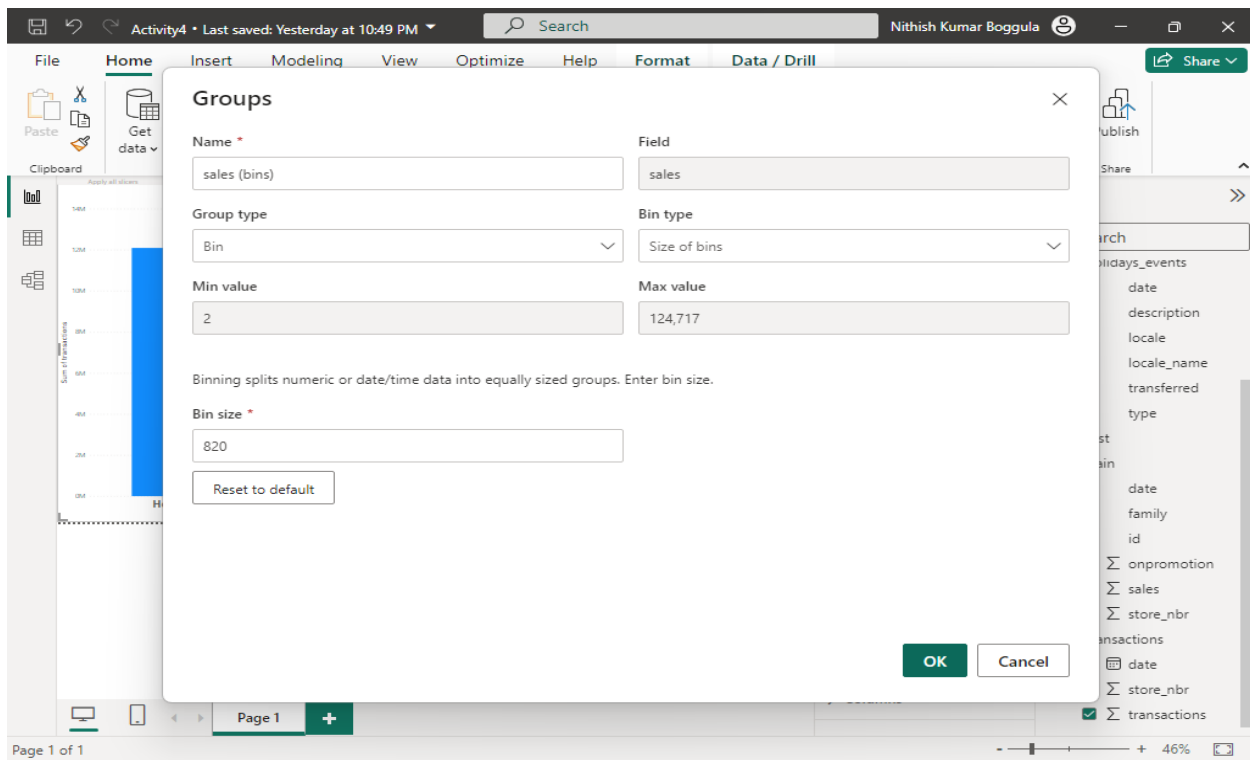
OK Cancel

19	1/19/2017	14.61904762	75.14285714
20	1/20/2017	15.26315789	66.47368421
21	1/21/2017	15.39130435	70.86956522
22	1/22/2017	18.44	76.24
23	1/23/2017	18.11764706	76

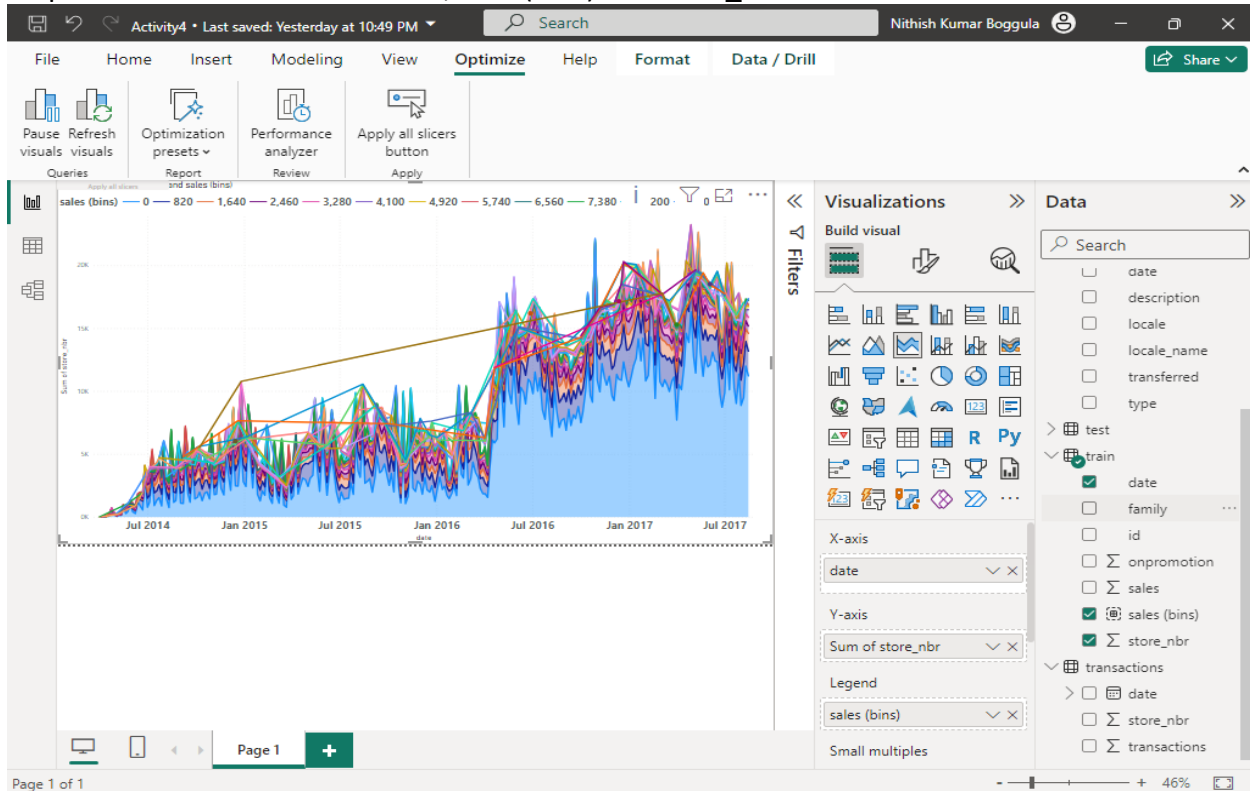
5 COLUMNS, 114 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 2:36 AM

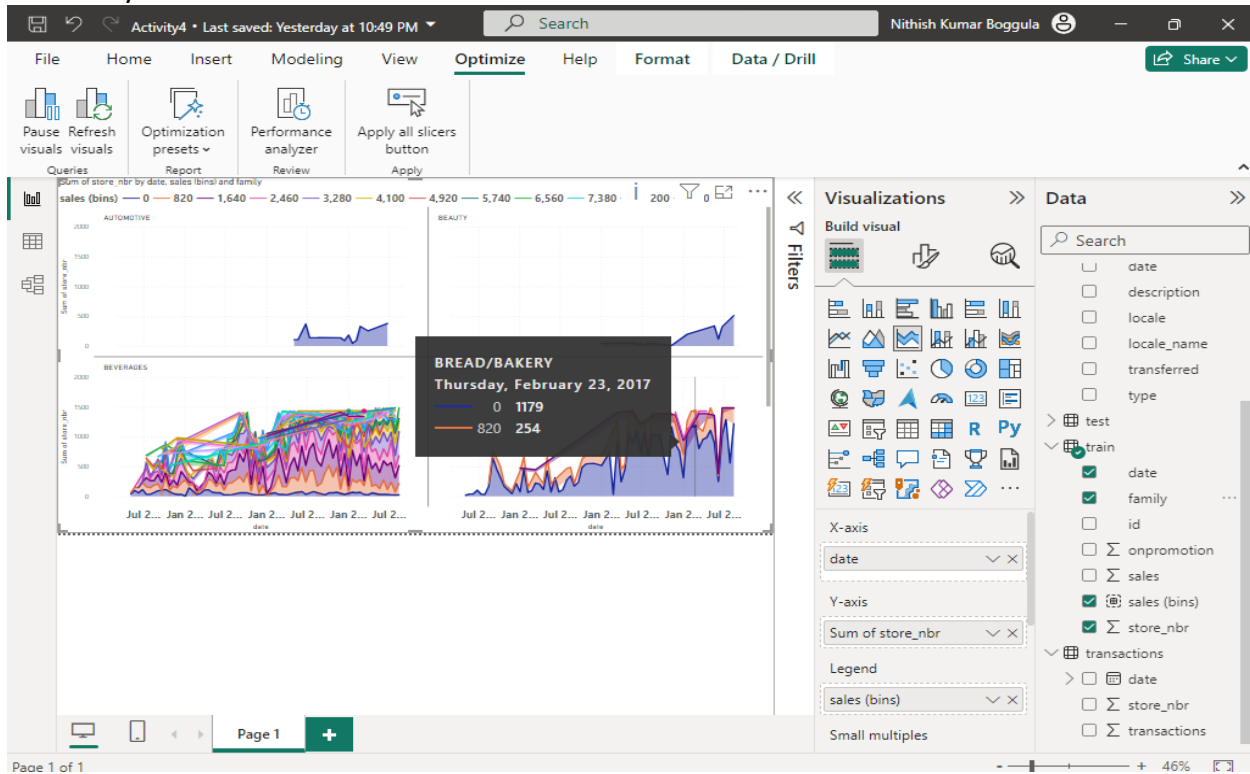
Step 2:- Grouping based on a sales attribute.
Use the default settings and click OK.



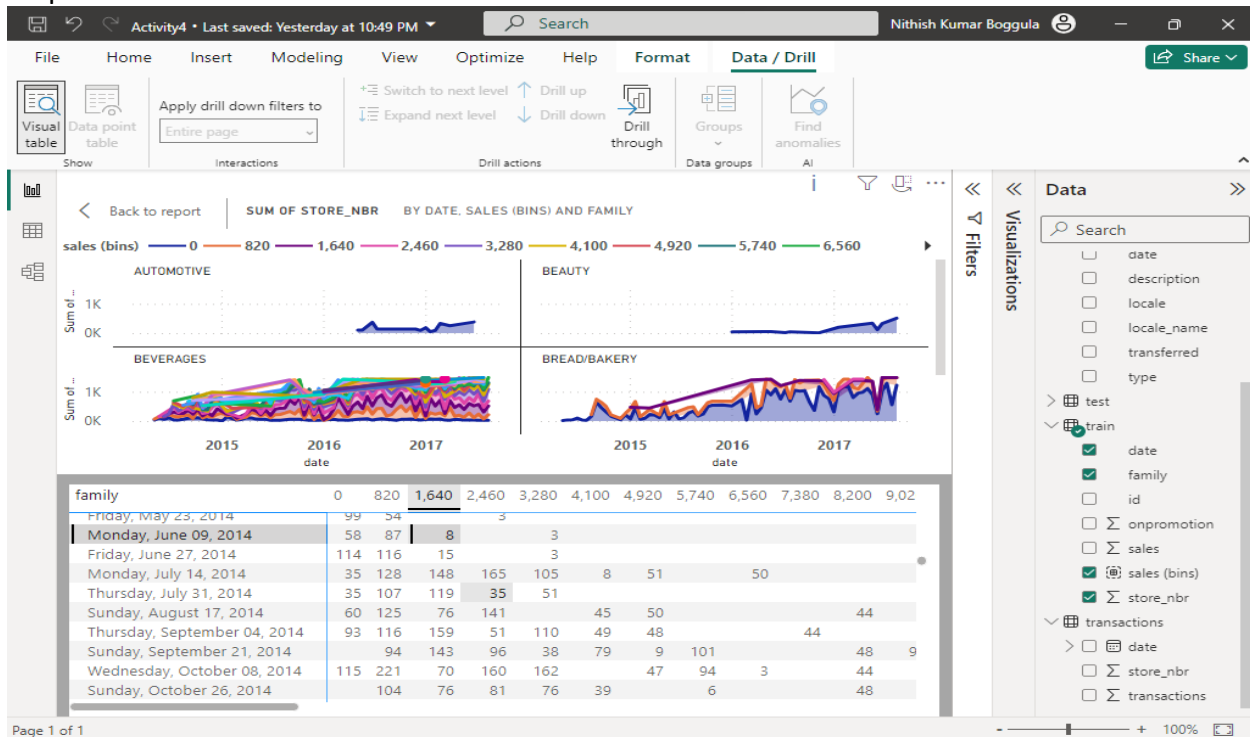
Step 3:- Select the attributes Date, sales(bins) and store_nbr and use the stacked area chart.



Step 4:- Select attribute Family and we can observe that the sales have been divided based on the family.



Step 5:- Get the visual table for the above Stacked area chart.



Question 4	Points
<p>Follow the above tutorial complete the following tasks.</p> <p>Use transactions to use grouping to the attribute instead of Sales.</p> <p>Select all the attributes in Transactions table to visualize and remove quarters from the date in the X axis.</p> <p>Note the changes and compare the above attributes using a line graph and Explain your understanding in detail in about 100 – 200 Words. (2-3 visualizations Mandatory)</p>	<p>Visualization – 5 Marks(each) Explanation – 5 Marks</p>
A. Just click on sum of transactions and change it to count and explain why it decreases.	Explanation with visualization – 5 Marks
	Total – 20 Marks

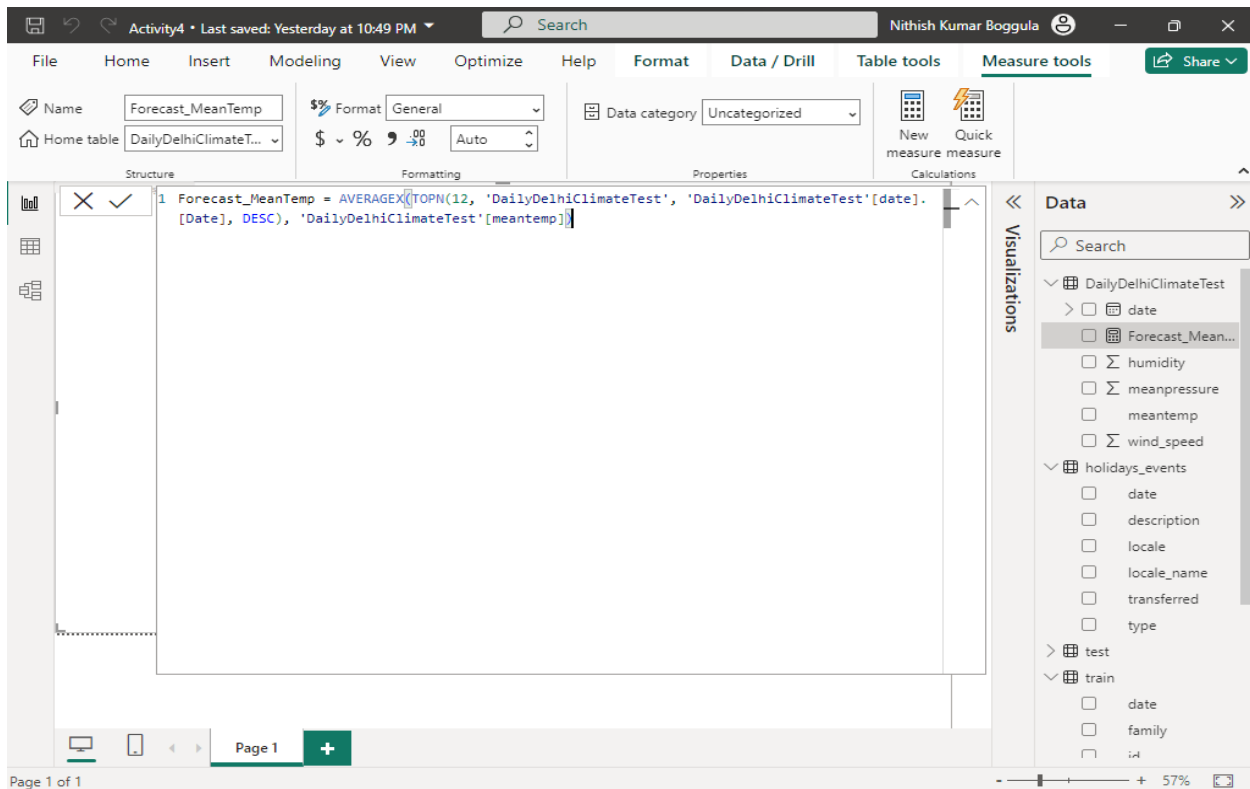
Task 5

Tutorial 5: Forecasting

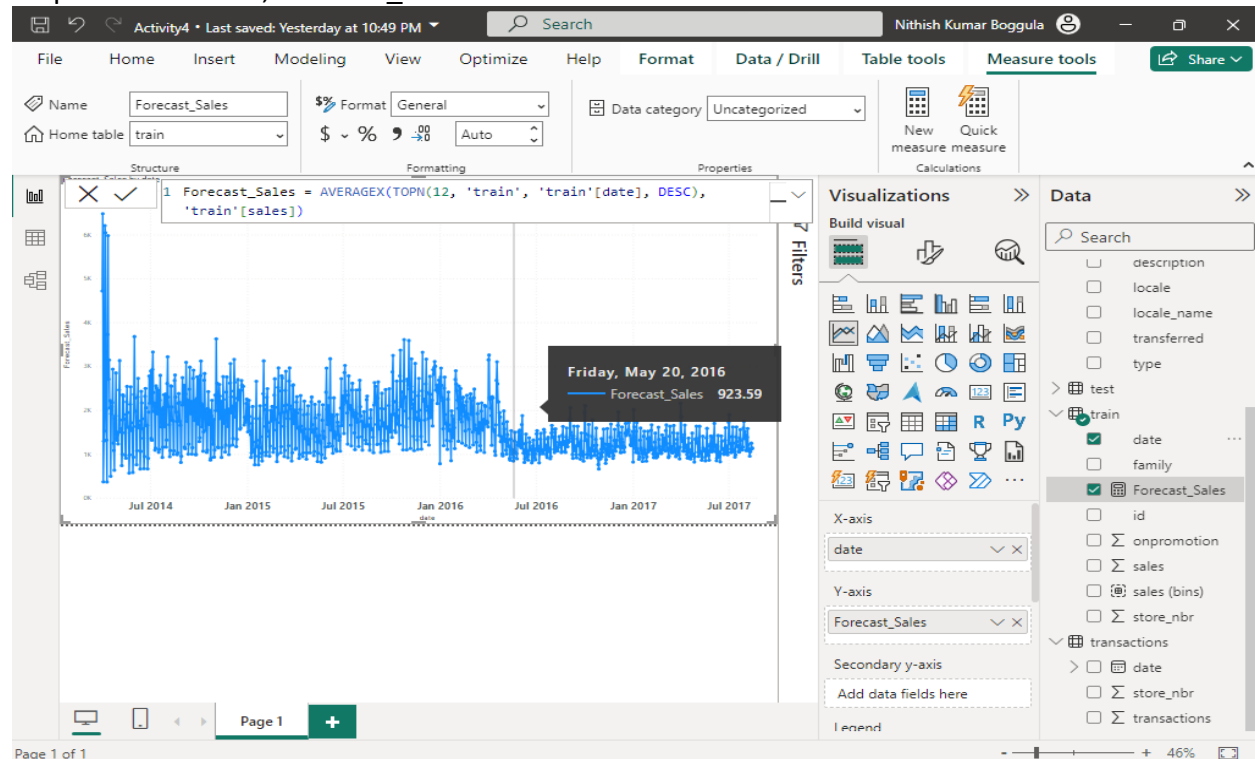
In this task we will forecast the data along with time i.e. year. You will need to use the below query to create a new attribute and select that attribute to forecast the sales.

Step 1:- Using the New measure by naming it Forecast_MeanTemp and choosing the home table as DailyDelhiClimateTest.

Use the below Query and you can see that an attribute named by Forecast_MeanTemp will be created under DailyDelhiClimateTest.



Step 2:- Select date, Forecast_Sales attributes and select the Line chart.



<p>Question 5</p> <p>Follow the above Tutorial and complete the below tasks.</p> <p>Use Transactions attribute and create a new measure by taking the query as a reference and build a query on your own to forecast the transactions data along with the date.</p> <p>Any additional attributes along with the above are encouraged. (2 visualizations Mandatory)</p>	<p>Points</p> <p>Visualization – 5 Marks</p> <p>Explanation – 5 Marks</p>
<p>Answer the Following Questions</p> <p>A. Explain how Time series forecasting is important and why?</p> <p>B. Explain your Complete Understanding of the Activity in 250- 300 words.</p>	<p>Explanation – 5 Marks</p>
	<p>Total – 15 Marks</p>

