CSCE 5320: Scientific Data Visualization
Activity 4

Submission Guidelines:

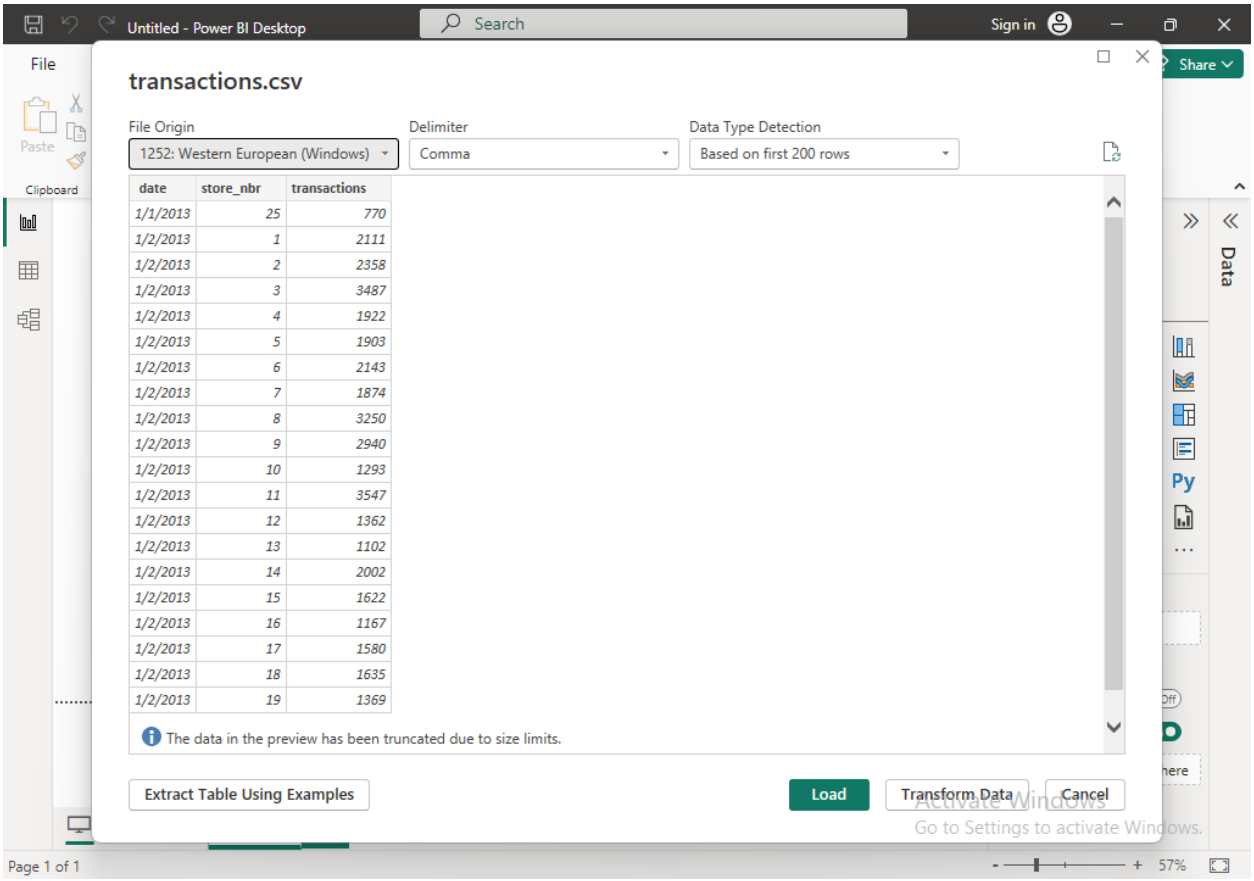
1. Assignment submission is individual.
2. Take screenshots of the entire screen including the date and time.
3. Mark every figure using a number and refer that figure while explaining the figure understanding using 200 - 300 words.
4. The students are required to submit both tutorial and tasks along 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

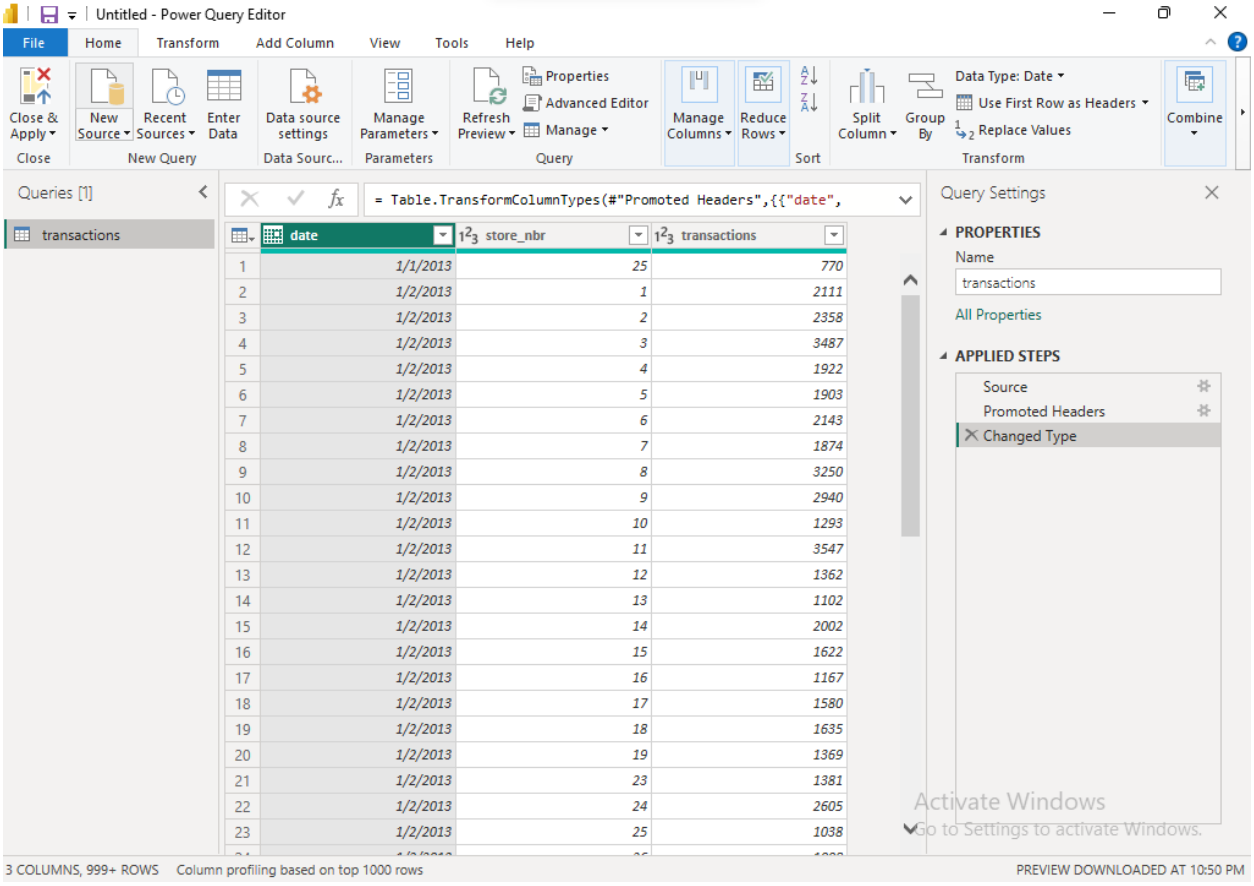
Tutorial 1: Power Query(Data Transformation and Group By)

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.

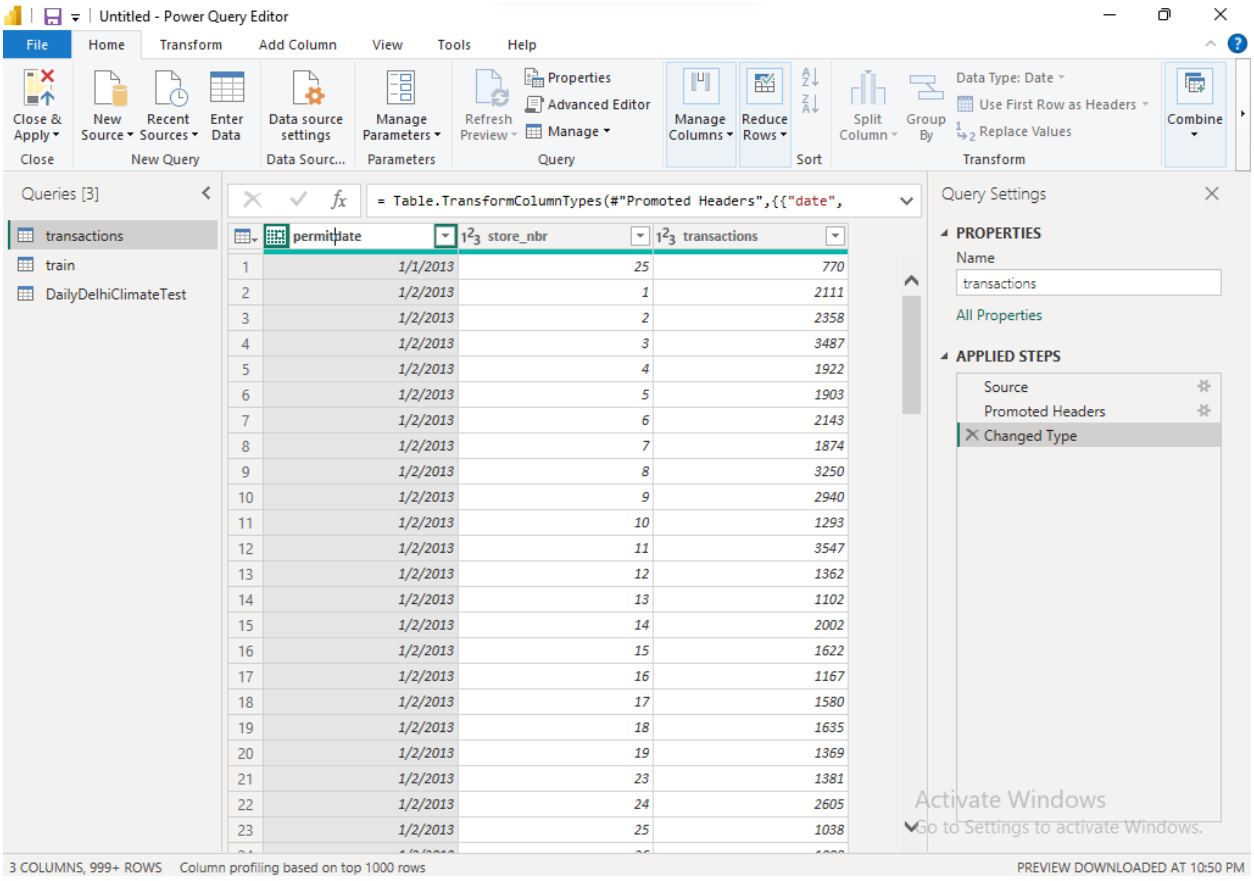
Step 1:- Load the 1st Dataset Named Transactions.csv and click on Transform Data.



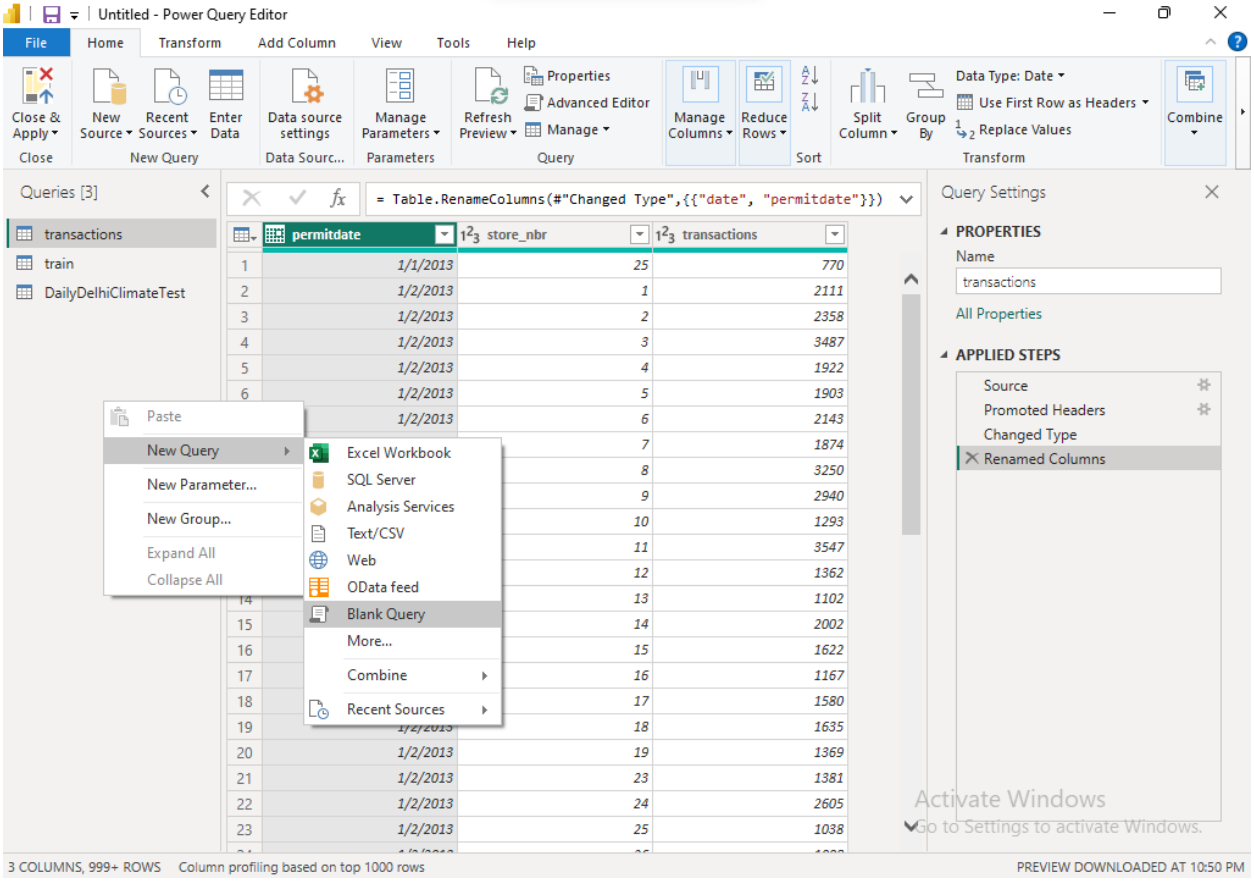
Step 2:- Once you click on Transform Data it takes you to the Power Query Editor where you can make changes to the dataset.
Now, Click on New Source Located at the Top and load the rest of the Datasets.



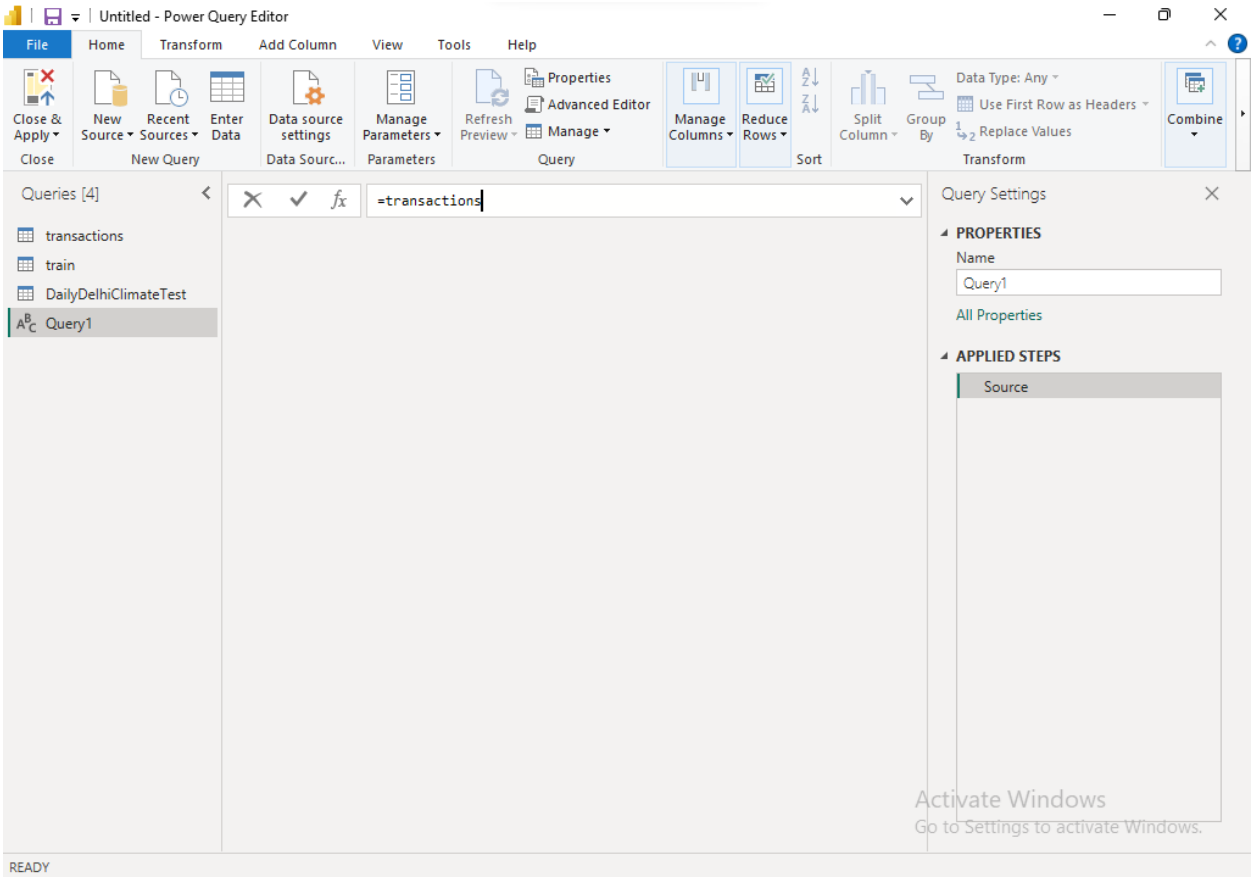
Step 3:- I Have loaded 3 Datasets or Tables in total where there is no common attributes except the Date Column in the DailyDelhiClimateTest.csv



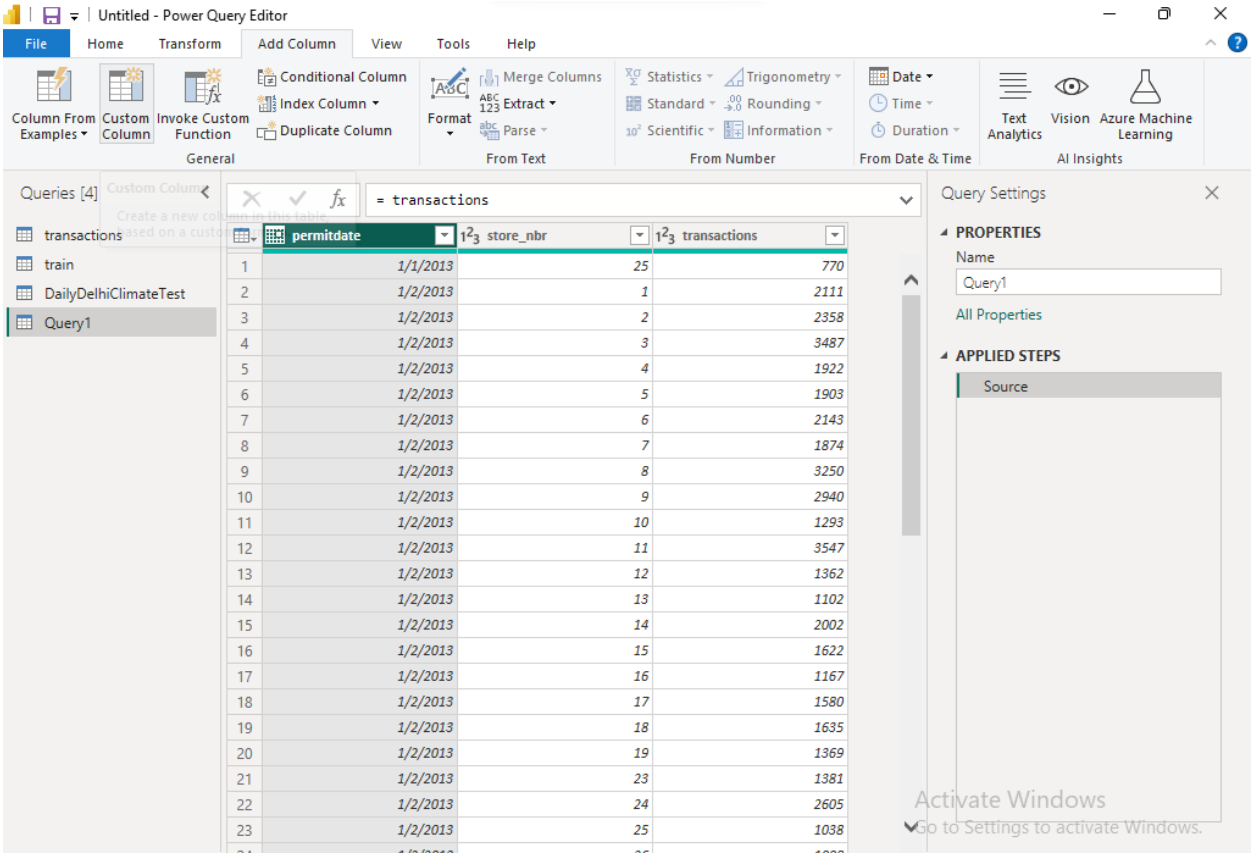
Step 4:- Now I want to get the columns min and max values. You can do that by using a query. Right click and click on New Query -> Blank Query.
I have renamed the Date to permitdate. You can also rename the attributes name just by clicking on the attribute and renaming it with the other name.



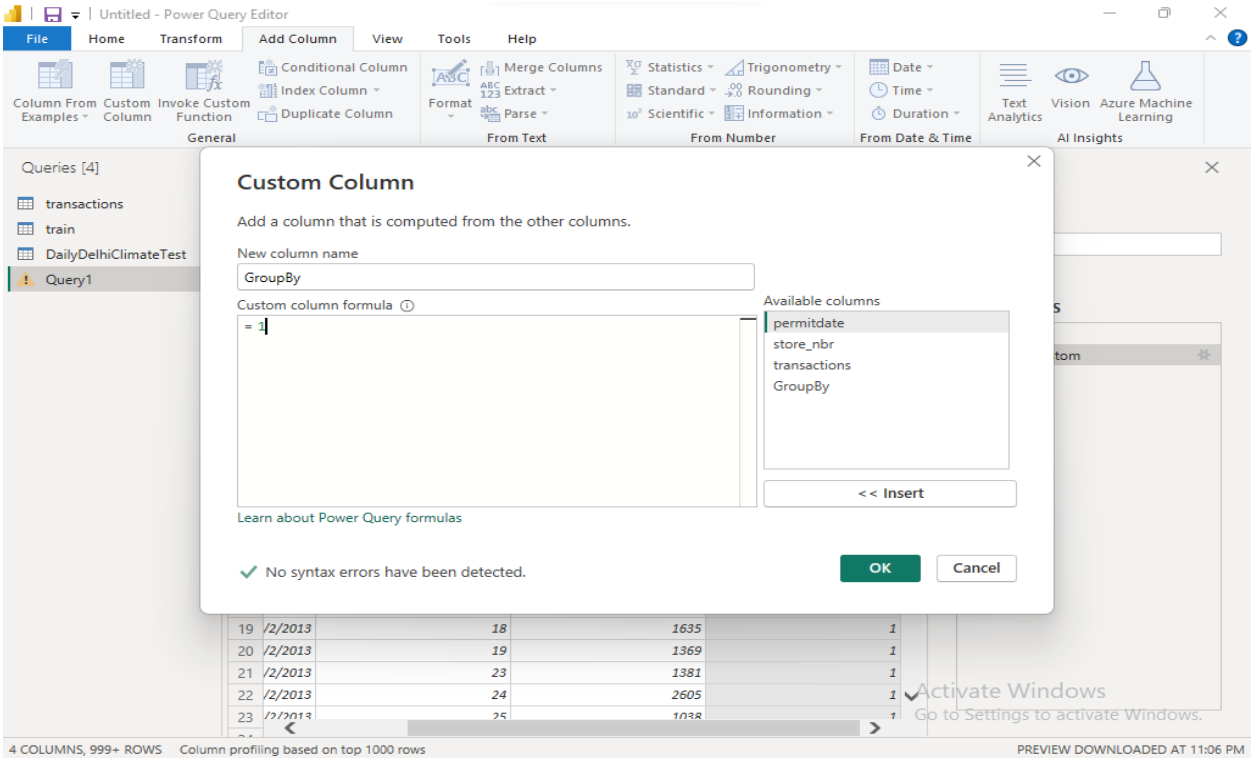
Step 5:- You can observe a new Query1 added to the Queries tab where I have selected the transactions table to perform modifications.

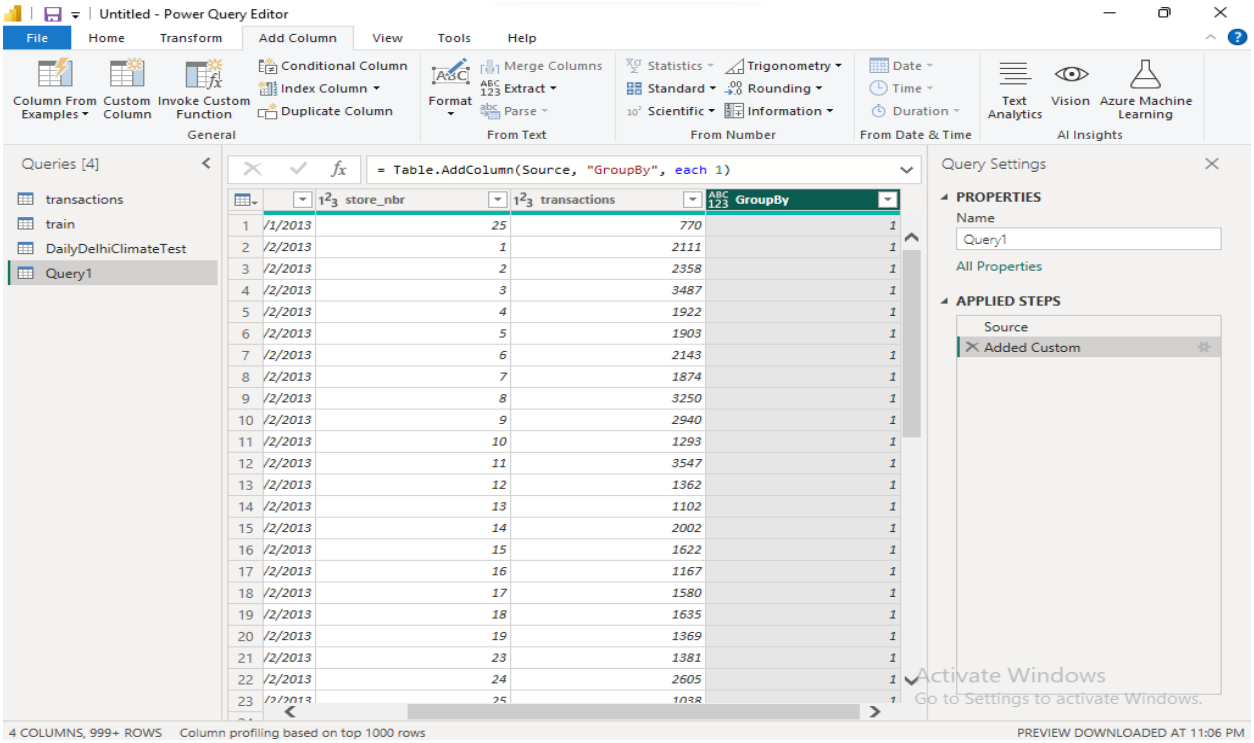


Step 6:- The “=transactions” display the Transactions table. You will be able to find the Applied steps at the right side under Query settings where the steps performed in that table will be displayed. You can click on the X mark on that step if you want to remove that step from the changes made.

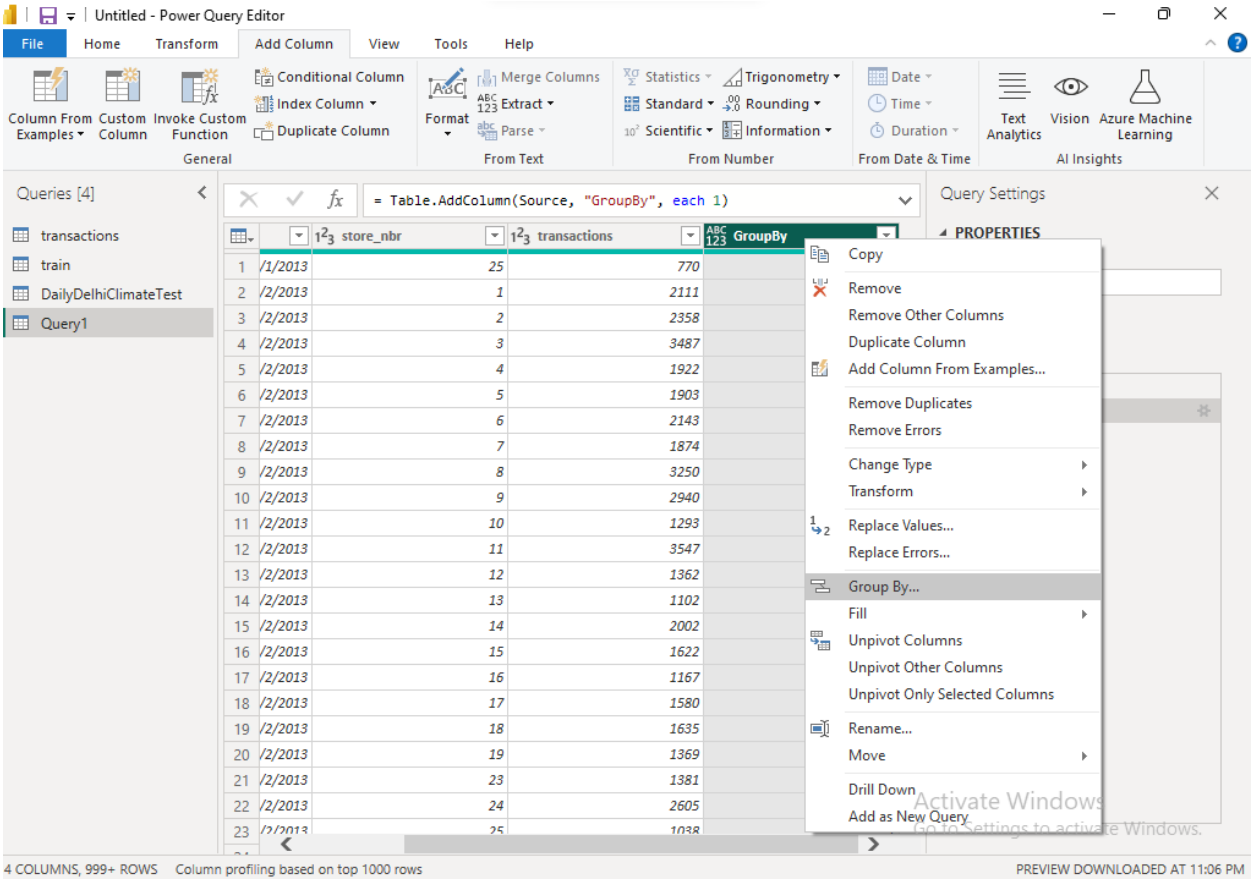


Step 7:- In add column, Click on Custom column to create a new column. But our main purpose is to perform Group by. The cells will be inserted with a 1.

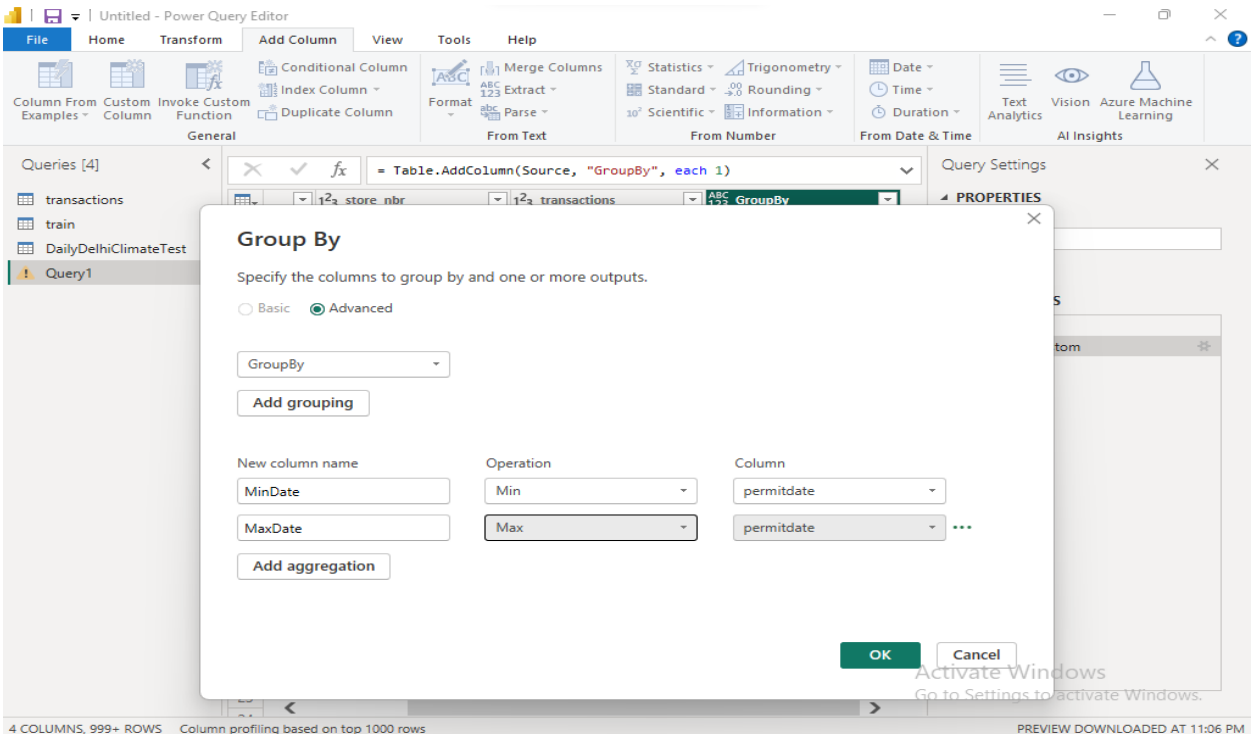




Step 8:- Right click on the attribute and click on group By.

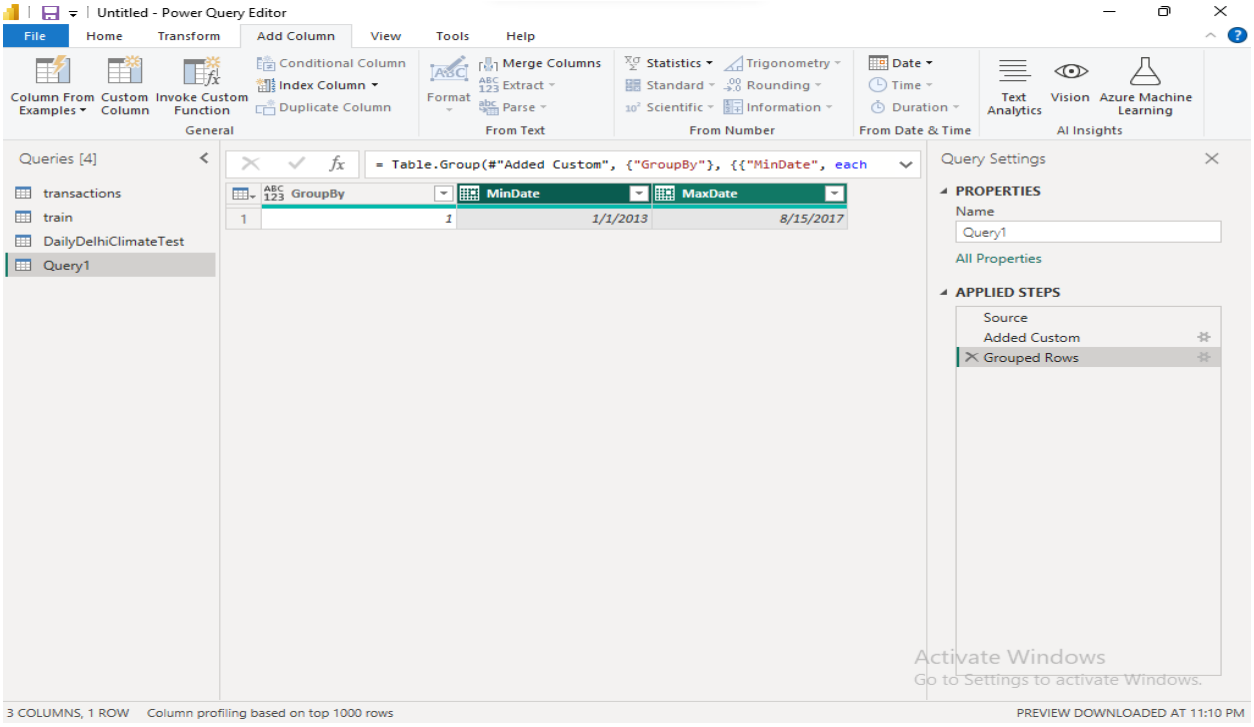


Step 9:- Go the advanced options and type the same min and max dates as shown in the fig below and select the attribute for which you want the group by function to be applied.

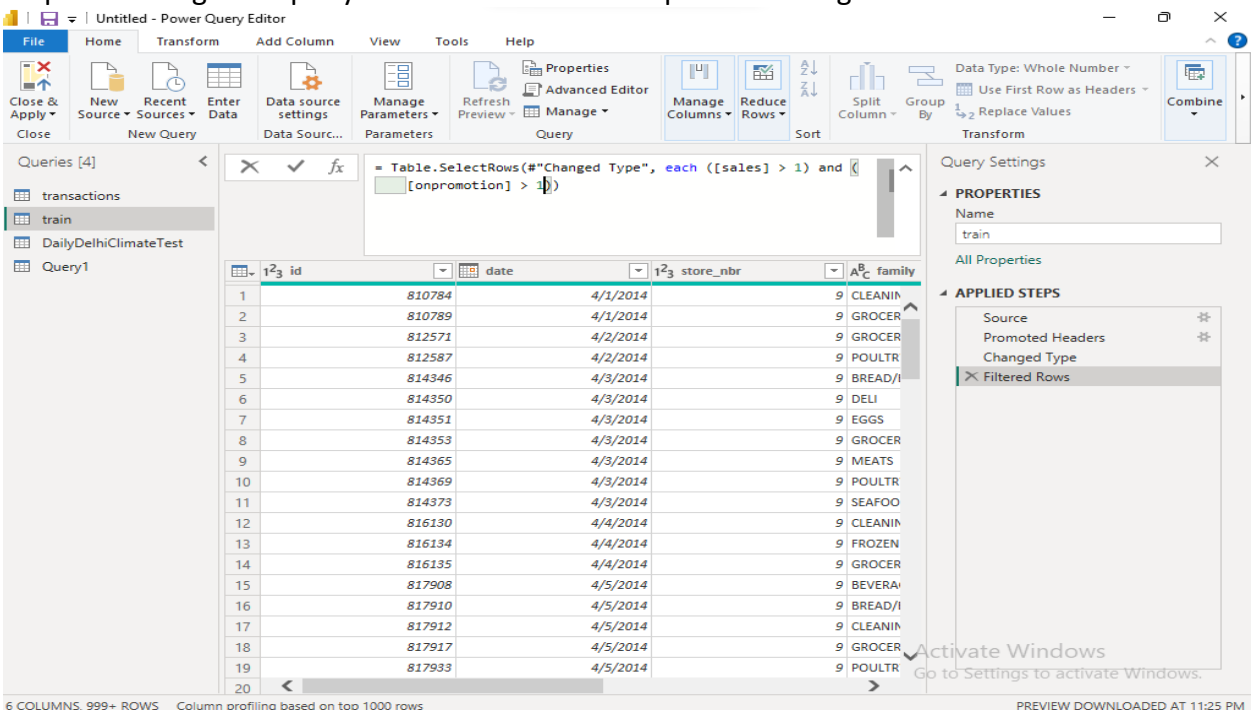

Untitled - Power Query Editor
File
Home
Transform
Add Column
View
Tools
Help
Column From Examples
Custom Column
Invoke Custom Function
Conditional Column
Index Column
Duplicate Column
General
Format
Merge Columns
Extract
Parse
From Text
Statistics
Standard
Scientific
Trigonometry
Rounding
Information
From Number
Date
Time
Duration
From Date & Time
Text Analytics
Vision
Azure Machine Learning
AI Insights
Queries [4]
transactions
train
DailyDelhiClimateTest
Query1
= Table.AddColumn(Source, "GroupBy", each 1)
store_nbr
transactions
GroupBy
Query Settings
PROPERTIES
Group By
Specify the columns to group by and one or more outputs.
Basic
Advanced
GroupBy
Add grouping
New column name
Operation
Column
MinDate
Min
permitdate
MaxDate
Max
permitdate
Add aggregation
OK
Cancel
Activate Windows
Go to Settings to activate Windows.
4 COLUMNS, 999+ ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 11:06 PM

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.



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



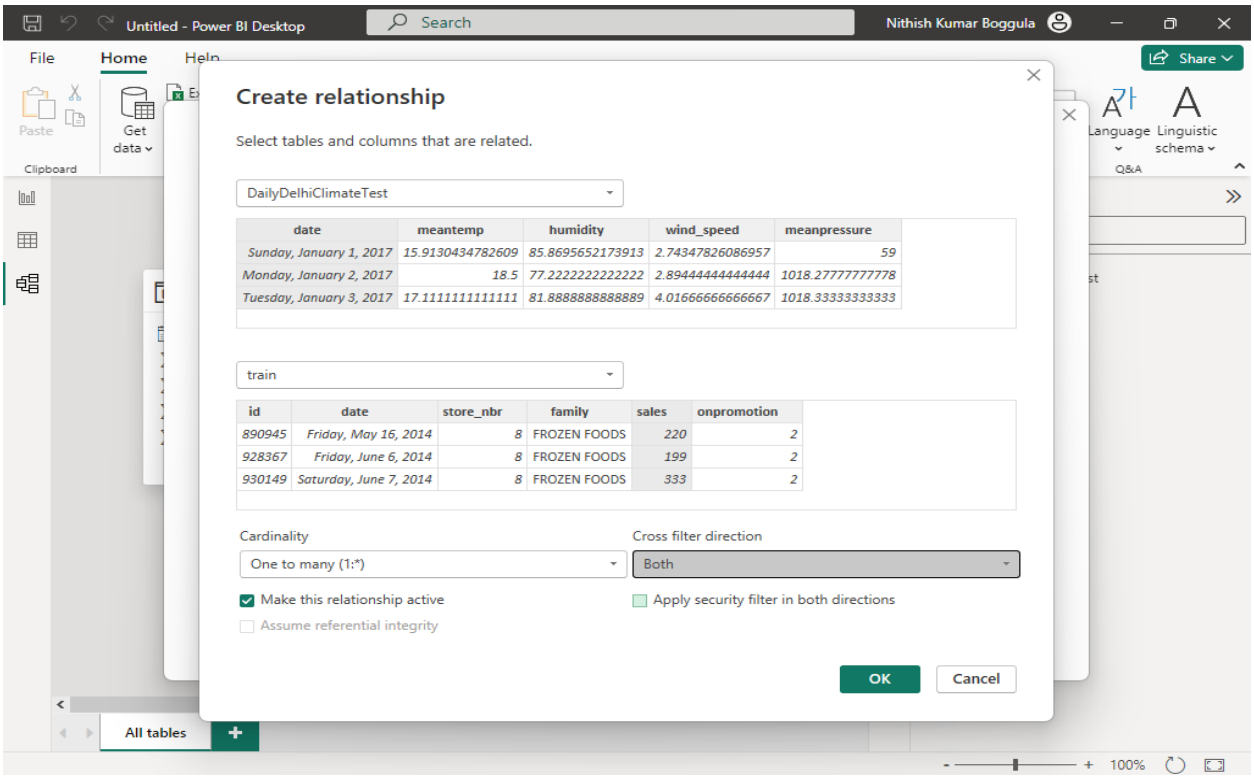
Question 1	Points
Follow the Tutorial 1 and complete the below task for the given attributes. Rename the store_nbr to your desired name to all the tables. Apply group by on any table and perform the same steps mentioned in the tutorial for any other table except train. Use the Query tab as Step 11 and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

Task 2

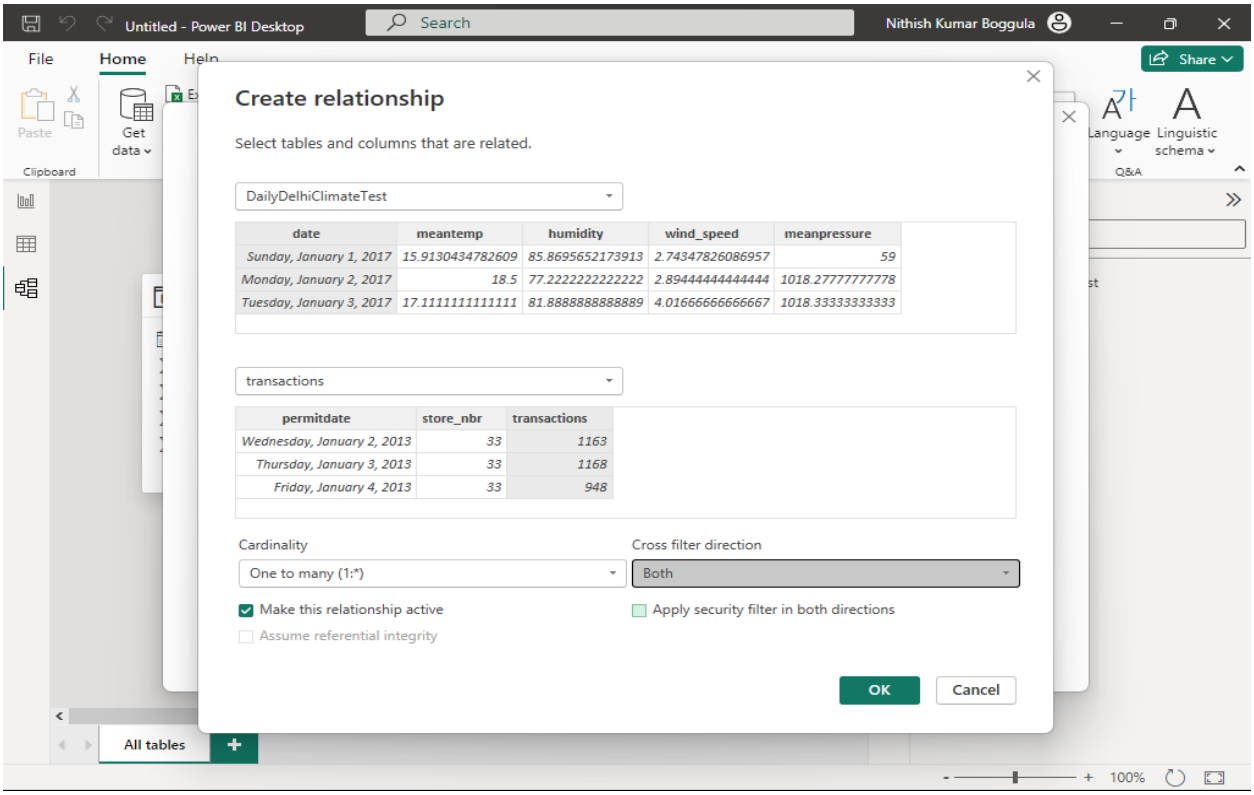
Tutorial 2: Creating Relationships

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

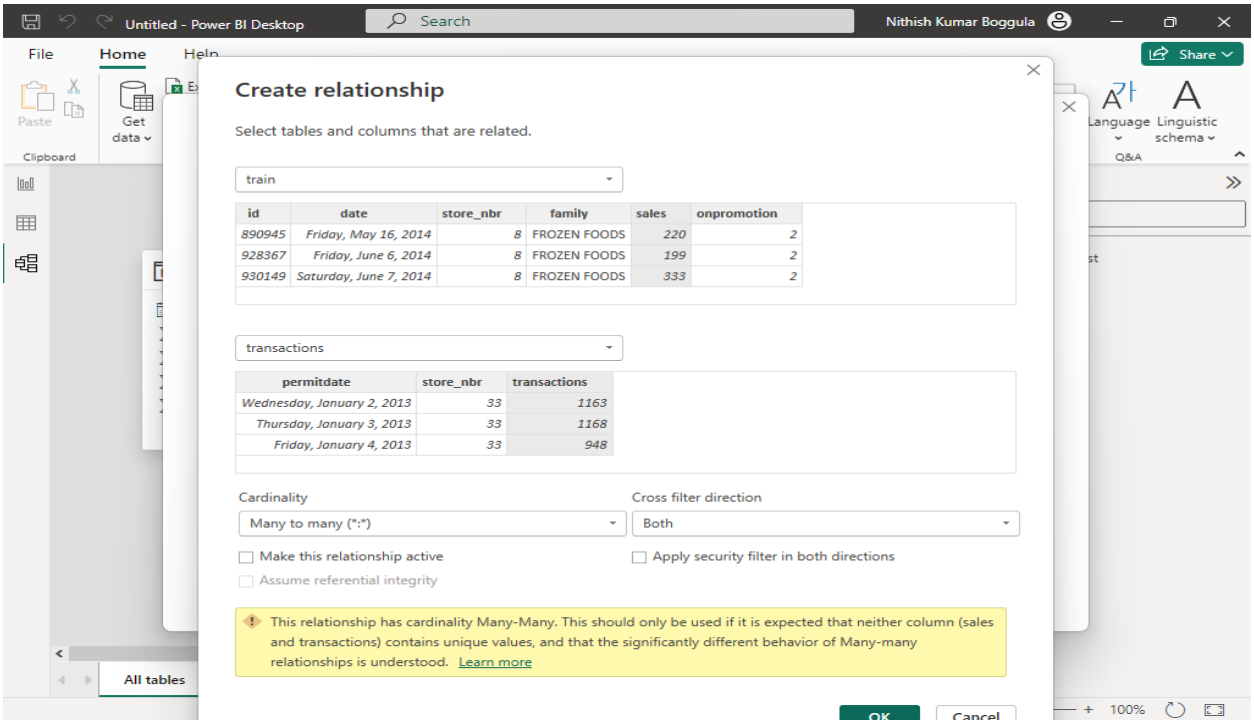
Step 1:- Go to the model view and right click on any table -> Create Relationship. Create all the possible and necessary relations which can be one to one, one to many, etc.....
Follow the below relationships and create relationships between the tables and attributes.



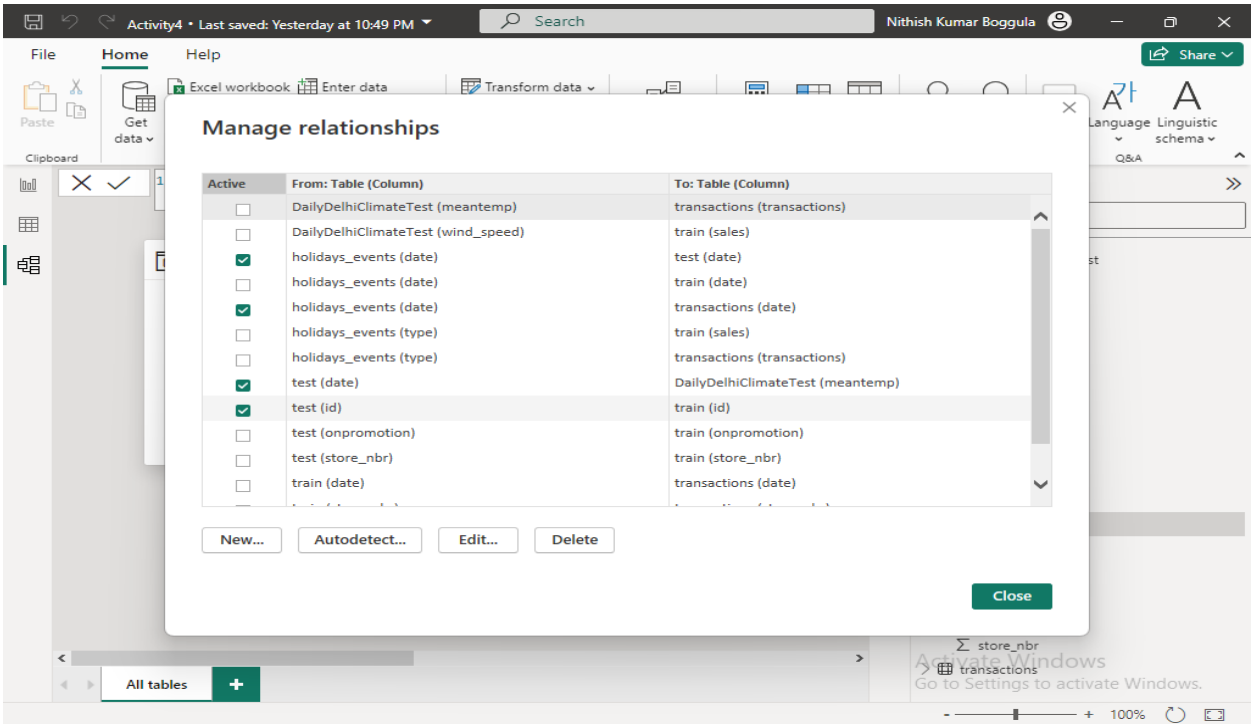
Step 2:- I am creating a connection between DailyDelhiClimateTest table and transactions table.



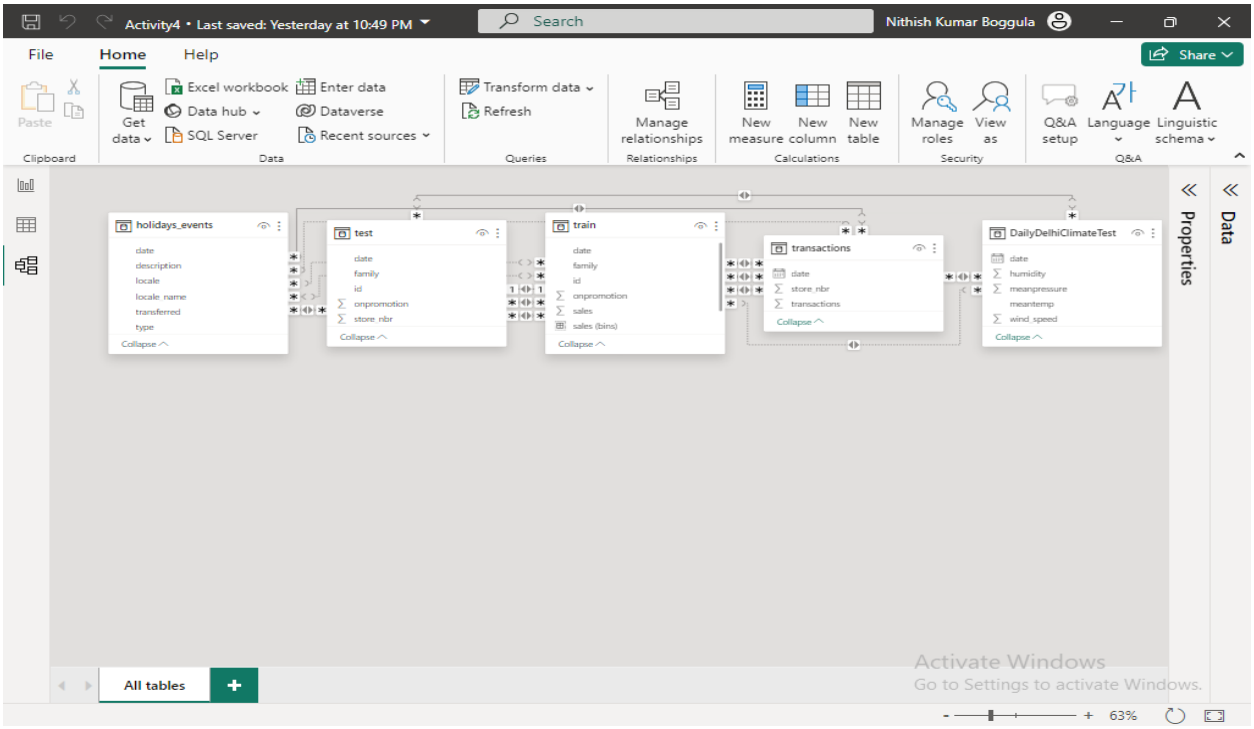
Step 3:- We are creating a relationship between train and transaction table. I have selected 1 attribute from each table to link the tables. The cardinality is automatically populated based on the data and attributes selected.



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.

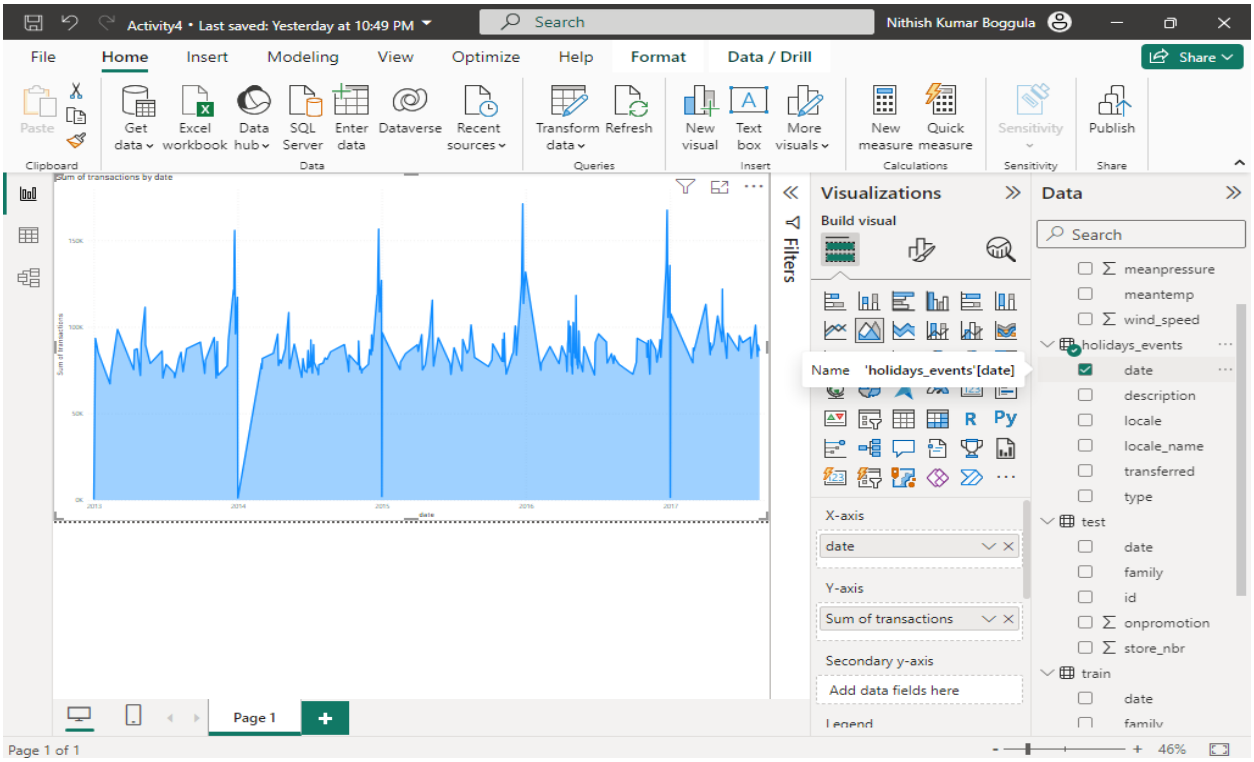


Question 2	Points
Follow the above tutorial and try to find all the possible relationships for the tables. Find all the one to one and many to one relationship among the attributes. Will cardinality of the relationship really affect the relationship and the way the attributes interact with the other?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

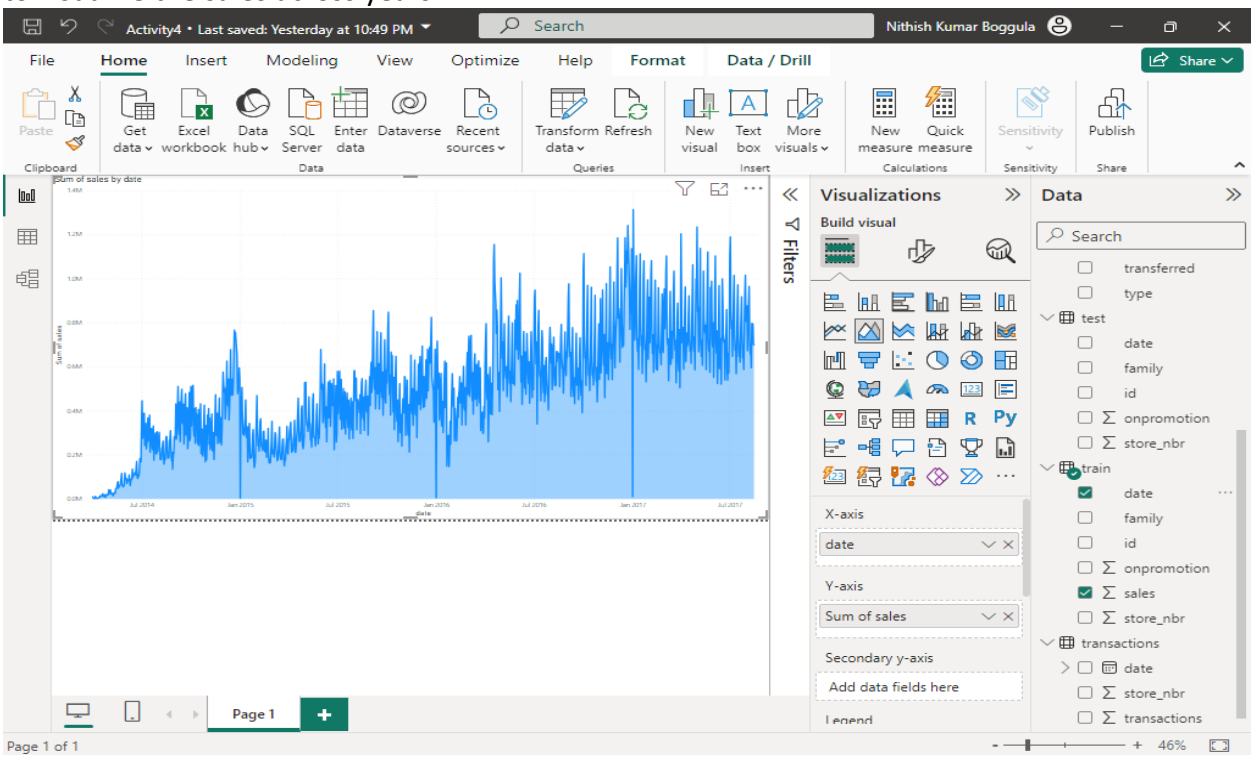
Task 3

Tutorial 3: Visualizing using different tables

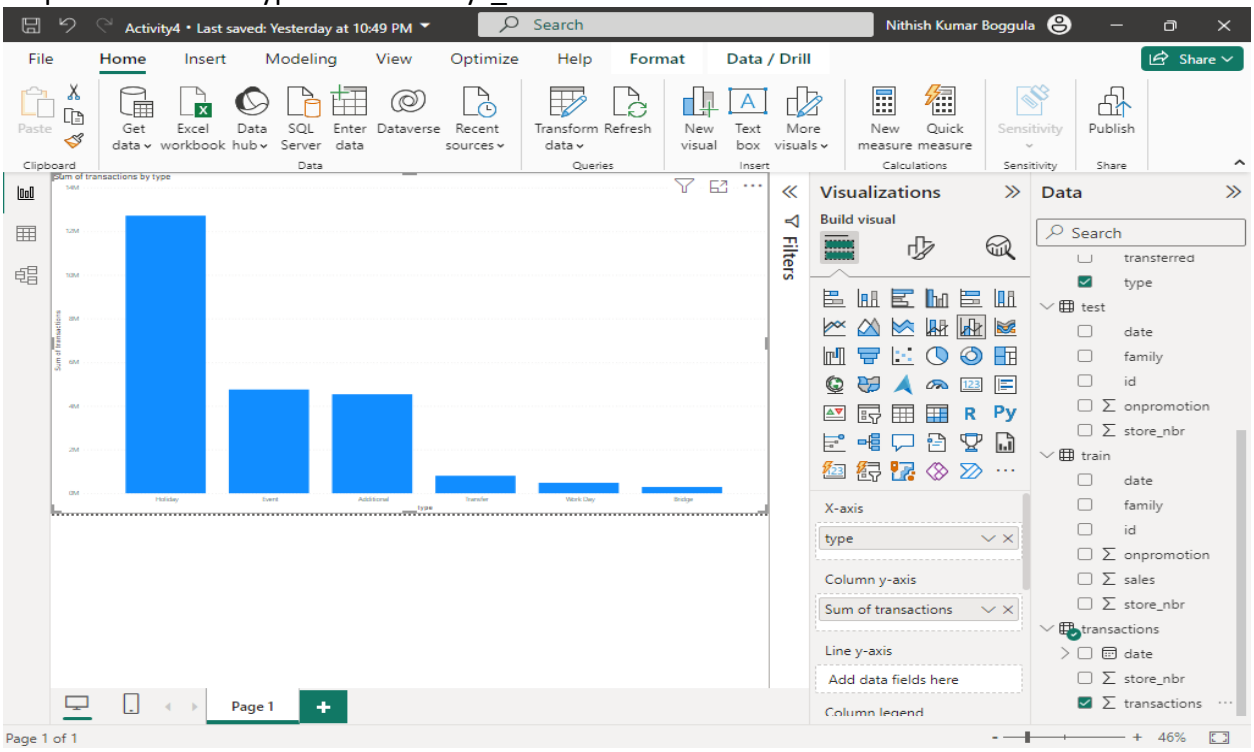
Step 1:- Select the date from the holidays_events table and transactions from the transaction table and click on the Area chart to visualize the sum of transactions with respect to years.



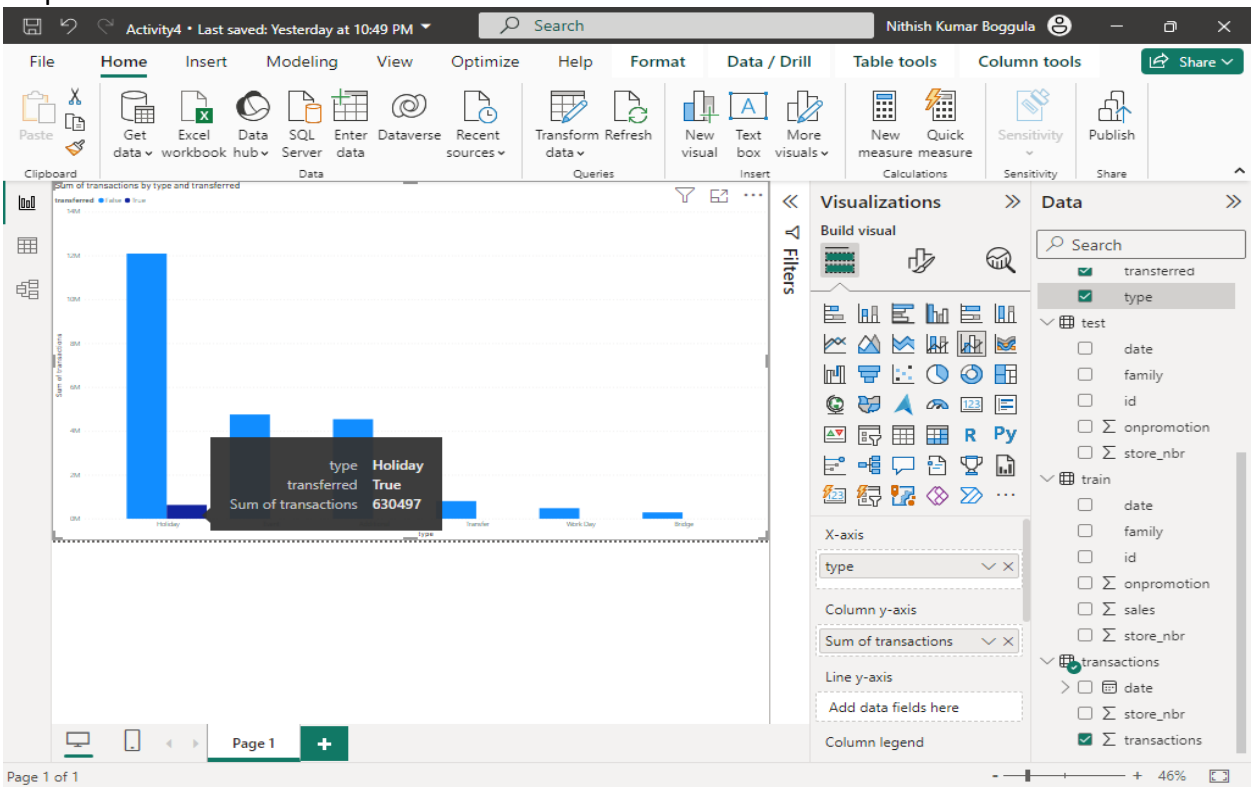
Step 2:- Now I am selecting the sales and date from the same train table and use the area chart to visualize the sales across years.



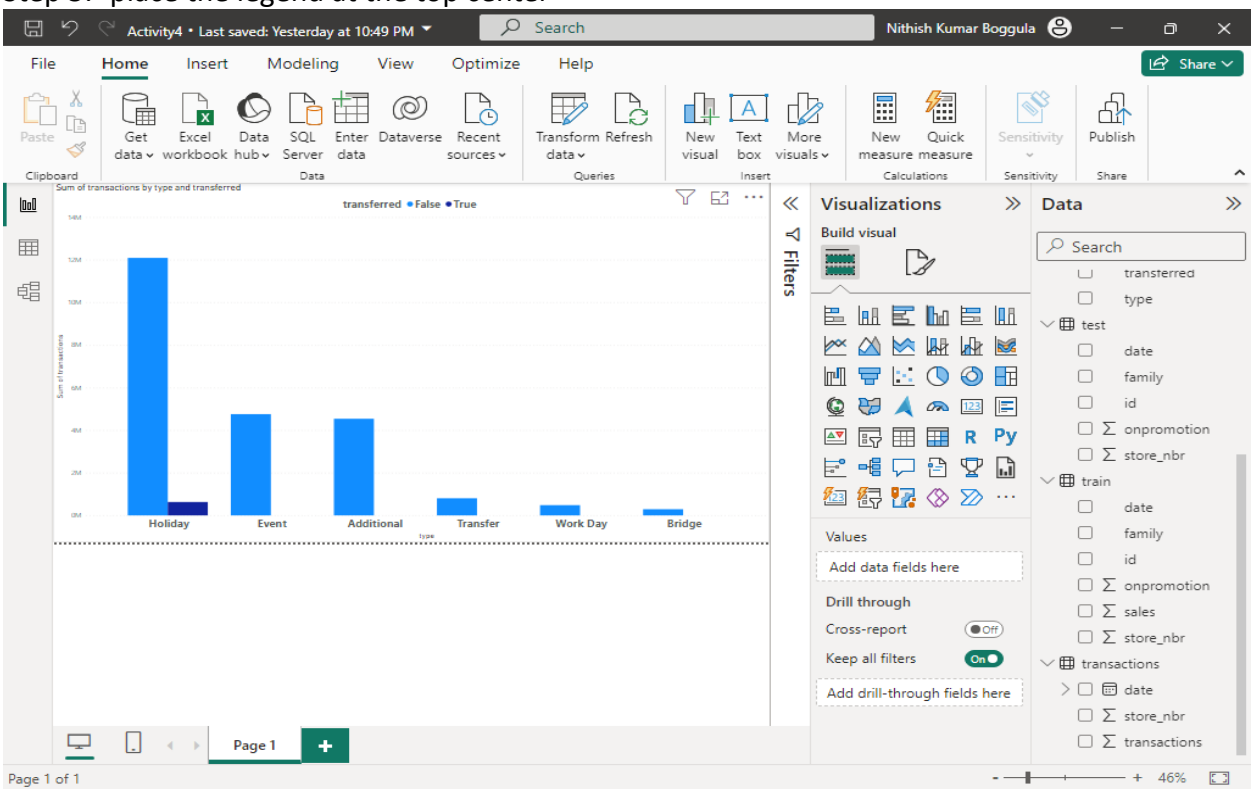
Step 3:- Select the Type from holidays_events and transactions and select the visualization.



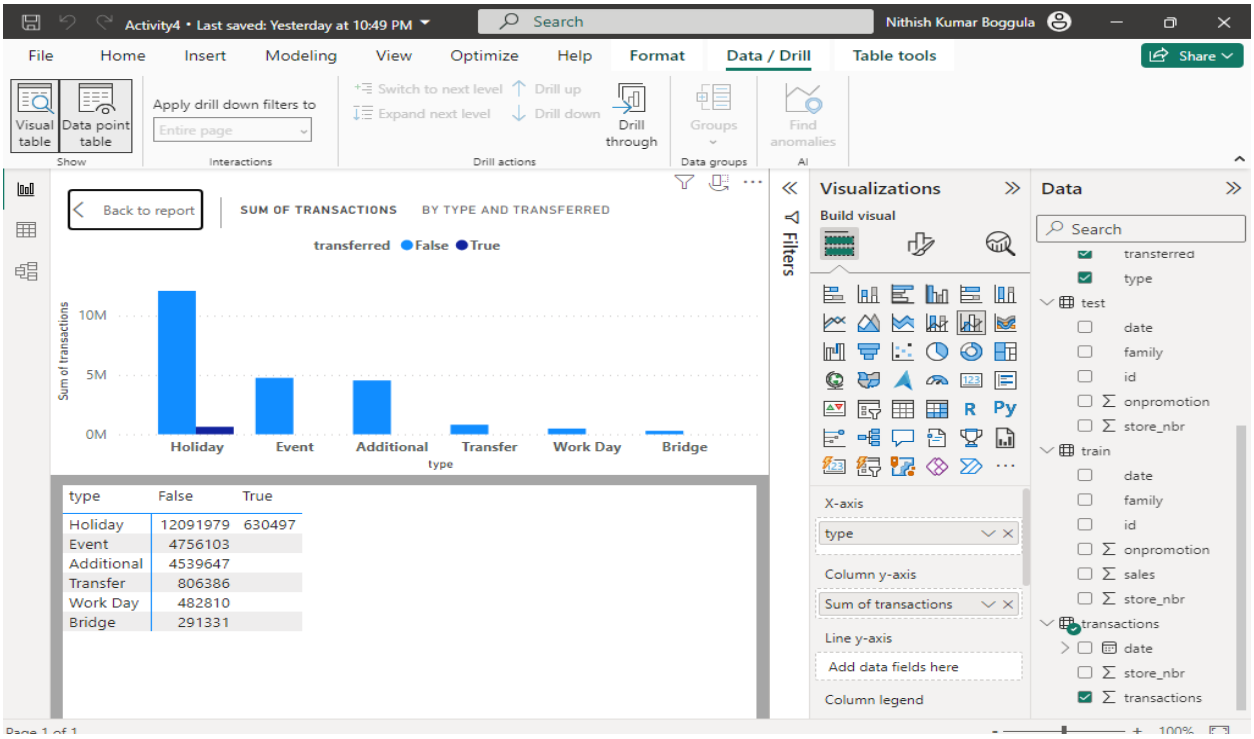
Step 4:- Now add another attribute i.e. Transferred



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.



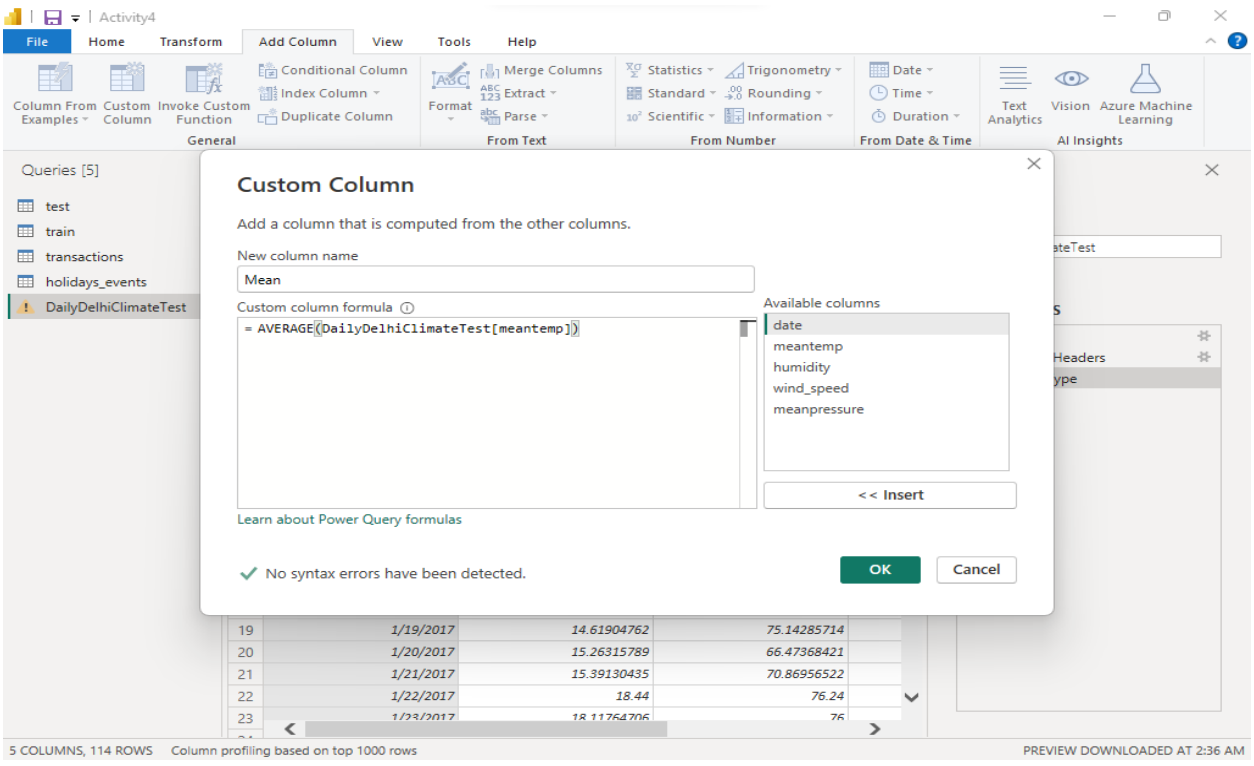
Question 3	Points
Follow the Tutorial 3 and complete the below task for the given attributes. Select the attributes Transferred, Type and Date from Holidays_events and transactions from transactions table. Take another page and select date, sales, and store_nbr from train and use a clustered column chart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

Tutorial 4

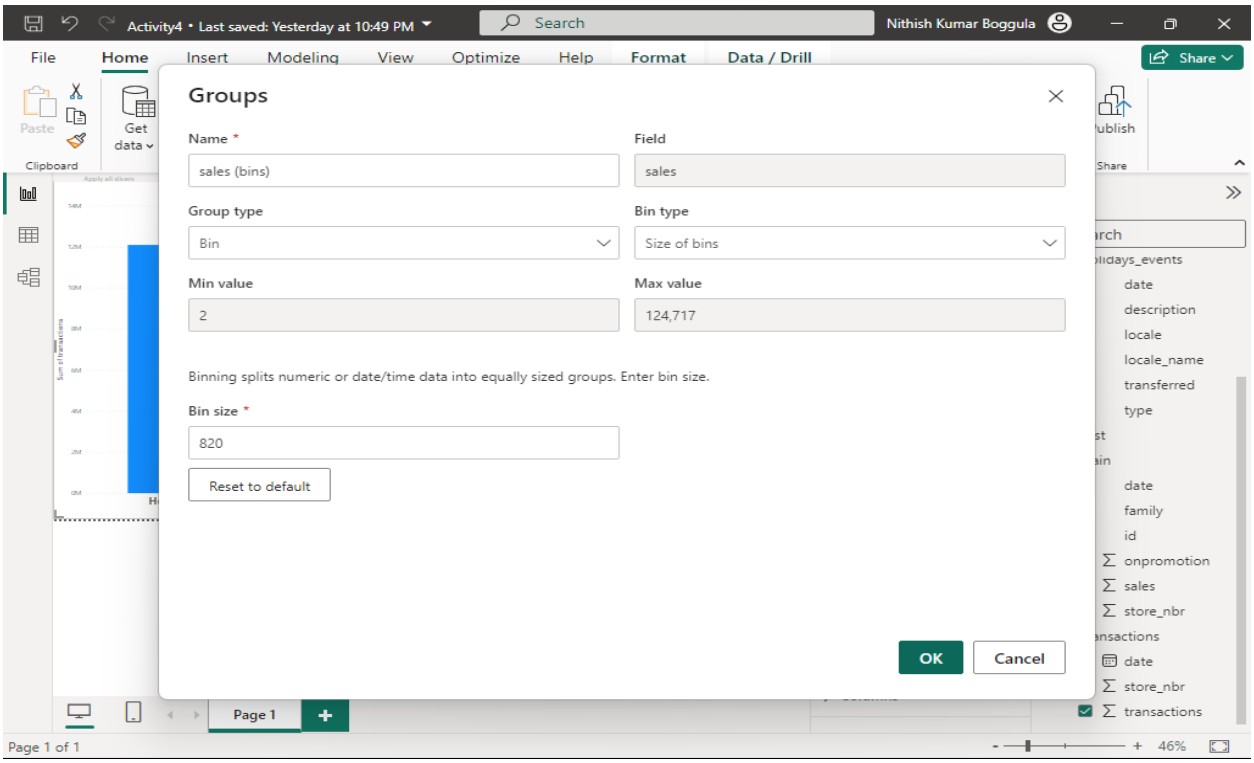
Tutorial 4: Creating an attribute to group the attributes

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.

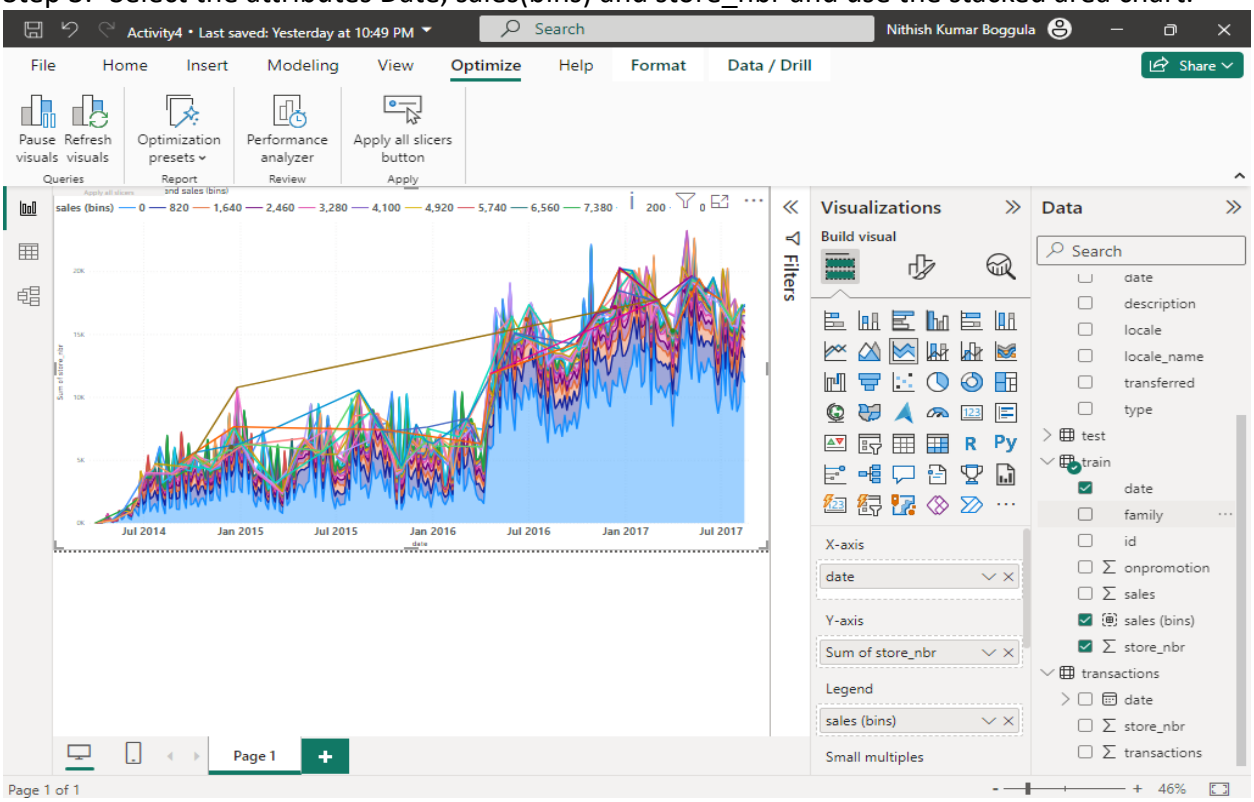
Step 1:- Here we will be doing forecasting on various attributes.
Click on Custom Column and give the below query to get the mean. If you want to insert the attributes then just click on the attributes and click on insert to use them in the query.



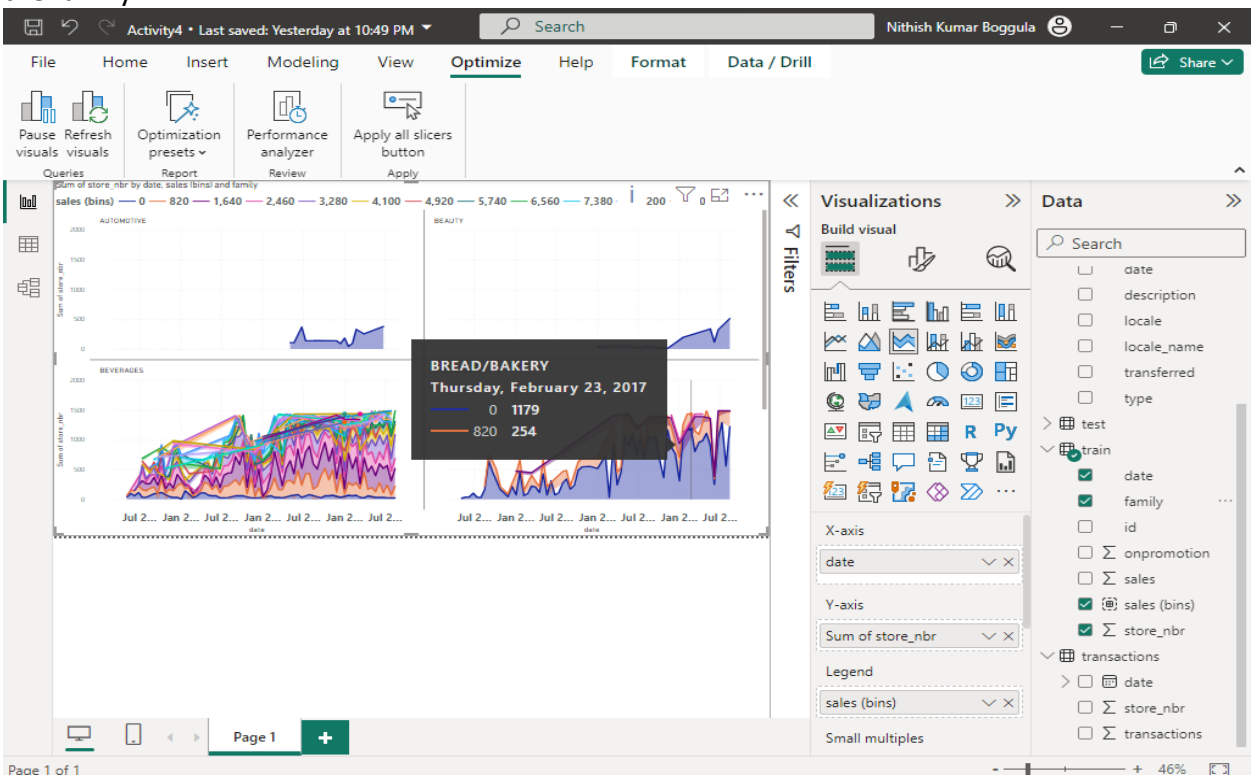
Step 2:- Grouping based on a sales attribute.
Use the default settings and click OK.



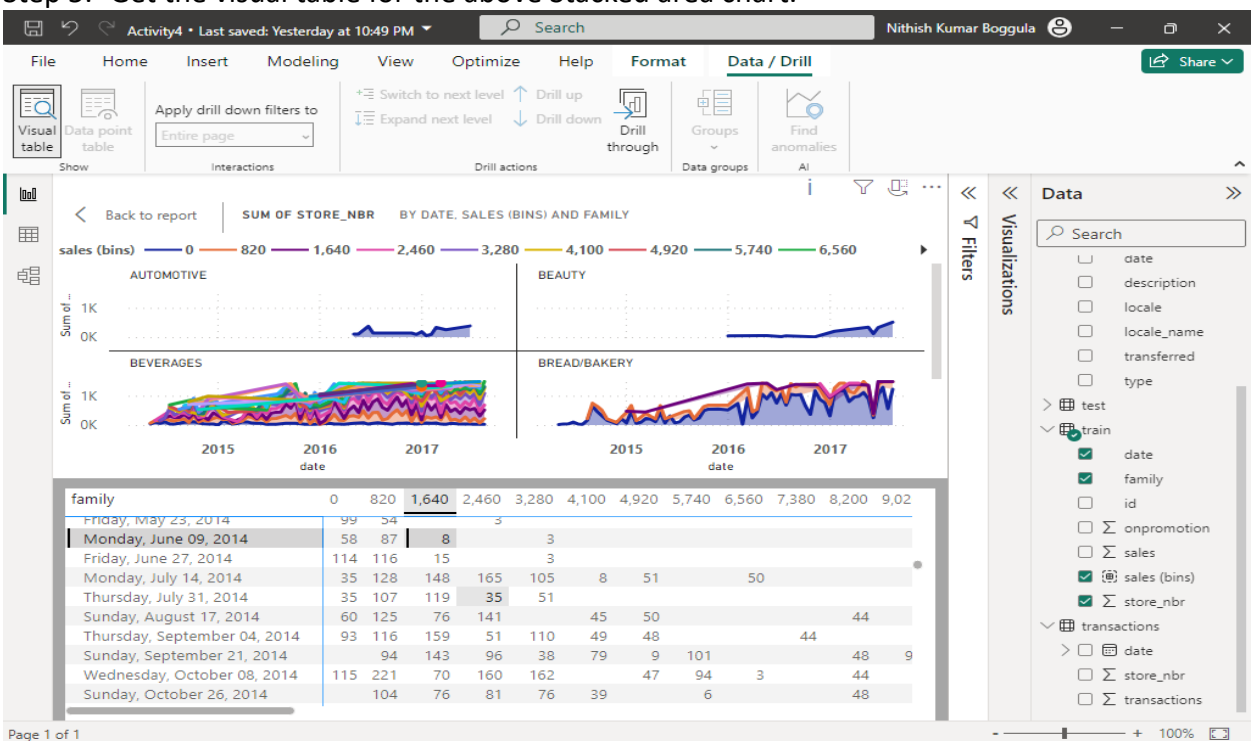
Step 3:- Select the attributes Date, sales(bins) and store_nbr and use the stacked area chart.



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.



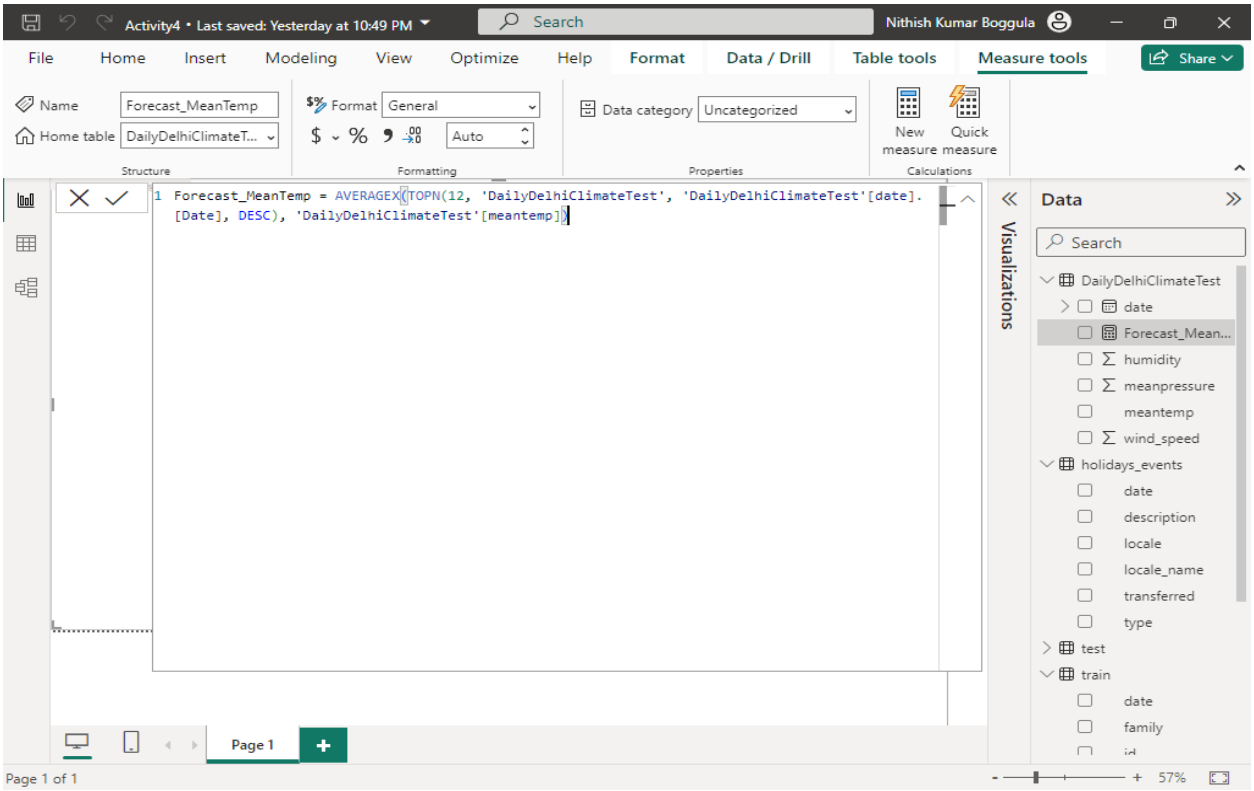
Question 4	Points
Follow the above tutorial complete the following tasks. Use transactions to use grouping to the attribute instead of Sales. Select all the attributes in Transactions table to visualize and remove quarters from the date in the X axis. Note the changes and compare the above attributes using a line graph and Explain your understanding in detail in about 100 – 200 Words. (2-3 visualizations Mandatory)	Visualization – 5 Marks(each) Explanation – 5 Marks
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

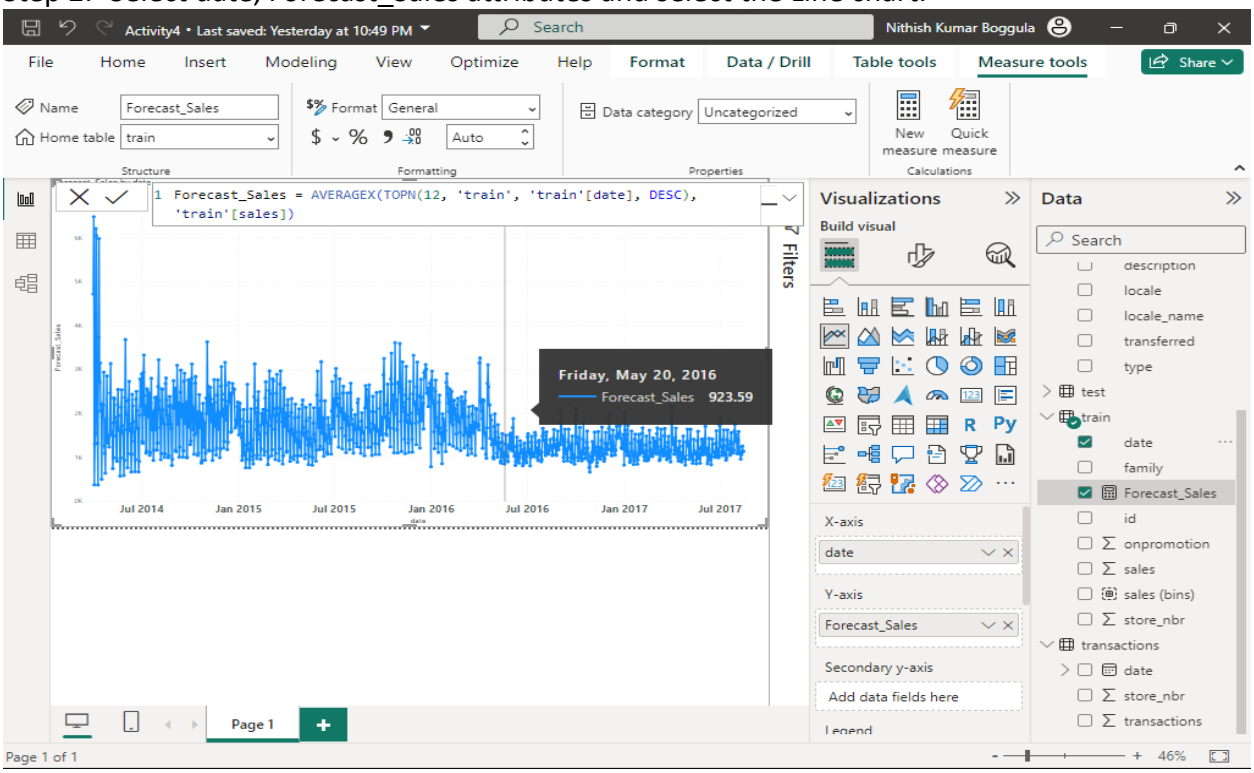
Tutorial 5: Forecasting

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.



Question 5	Points
Follow the above Tutorial and complete the below tasks. 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. Any additional attributes along with the above are encouraged. (2 visualizations Mandatory)	Visualization – 5 Marks Explanation – 5 Marks
Answer the Following Questions A. Explain how Time series forecasting is important and why? B. Explain your Complete Understanding of the Activity in 250- 300 words.	Explanation – 5 Marks
	Total – 15 Marks