

CSCE 5320: SCIENTIFIC DATA VISUALIZATION ACTIVITY-4 TUTORIAL

R. Nikhitha

Date: 09/23/2024

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

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

The screenshot shows the Microsoft Power BI Data Editor interface. At the top, there is a file browser window titled 'Open' showing files in 'SDV > ACT4'. The file 'transactions' is selected and highlighted in blue. Below the browser is a 'File name:' dropdown set to 'transactions' and an 'Open' button. To the right of the browser is a ribbon bar with tabs like 'Organize', 'New folder', 'Insert', 'Calculations', 'Sensitivity', 'Publish', 'View recovered files', 'Visualizations', and 'Fields'. The 'Fields' section on the right indicates 'You haven't loaded any data yet. Get data'. In the center, a preview pane displays the 'transactions.csv' file with columns 'date', 'store_nbr', and 'transactions'. The data shows daily sales figures from January 1, 2013, to January 20, 2013. At the bottom of the preview pane, there are buttons for 'Extract Table Using Examples', 'Load', 'Transform Data', and 'Cancel'. The status bar at the bottom of the screen shows the date and time as 9/19/2024 6:42 PM.

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 screenshot shows the Microsoft Power Query Editor interface. The main area displays a table with columns 'date' and 'store_nbr'. The 'Applied Steps' pane on the right shows a single step named 'Changed Type' applied to the 'store_nbr' column. The status bar at the bottom indicates the date as 9/19/2024 and the time as 6:42 PM.

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.

The screenshot shows the Microsoft Power Query Editor interface. A file browser window is open in the background, showing a folder path of 'SDV > ACT4' containing several CSV files. The main area of the editor shows the 'transactions' dataset with columns 'date' and 'store_nbr'. The 'Applied Steps' pane on the right shows a single step named 'Changed Type' applied to the 'store_nbr' column. The status bar at the bottom indicates the date as 9/19/2024 and the time as 6:45 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

Queries [3]

	date	meantemp	humidity	wind_speed	meanpressure
1	1/1/2017	15.91304348	85.86956522	2.743478261	1018.277
2	1/2/2017	18.5	77.22222222	2.894444444	1018.333
3	1/3/2017	17.11111111	81.88888889	4.016666667	1018.333
4	1/4/2017	18.7	70.05	4.545	1018.333
5	1/5/2017	18.38888889	74.94444444	3.3	1014.333
6	1/6/2017	19.31818182	79.31818182	8.681818182	1011.772
7	1/7/2017	14.70833333	95.83333333	10.041666667	1011.772
8	1/8/2017	15.68421053	83.52631579	1.95	1015
9	1/9/2017	14.57424857	80.80952381	6.542857143	1015.952
10	1/10/2017	12.11111111	71.94444444	9.361111111	1016.888
11	1/11/2017	11	72.11111111	9.772222222	1016.777
12	1/12/2017	11.78947368	74.57894737	6.626315789	1016.368
13	1/13/2017	13.23529412	67.05882353	6.435294118	1017.529
14	1/14/2017	13.2	74.28	5.276	1018
15	1/15/2017	16.43478261	72.56521739	3.630434783	1018.130
16	1/16/2017	14.65	78.45	10.38	1017
17	1/17/2017	11.72222222	84.44444444	8.038888889	1018.388
18	1/18/2017	13.04166667	78.33333333	6.029166667	1021.958
19	1/19/2017	14.61904762	75.14285714	10.33809524	1022.809
20	1/20/2017	15.26315789	66.47368421	11.22631579	1021.789
21					

5 COLUMNS, 114 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 6:44 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.

Queries [3]

	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		10	1293
12		11	3547
13		12	1362
14		13	1102
15		14	2002
16		15	1622
17		16	1167
18		17	1580
19		18	1635
20		19	1369
21		20	1381

3 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 7:25 PM

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

The screenshot shows the Power Query Editor interface. The top menu bar includes File, Home, Transform, Add Column, View, Tools, and Help. The ribbon below the menu has sections for Close & Apply, Source, Recent Sources, Enter Data, Data Source Settings, Manage Parameters, Refresh Preview, Advanced Editor, Properties, Choose Columns, Remove Columns, Keep Rows, Remove Rows, Sort, Split Column, Group By, Replace Values, Data Type: Any, Merge Queries, Append Queries, Text Analytics, Vision, Combine Files, Azure Machine Learning, and AI Insights. On the left, the 'Queries [4]' list shows 'transactions', 'train-2', 'DailyDelhiClimateTest', and 'Query1' (which is selected). The main area displays the 'transactions' table with columns: permitdate, store_nbr, and transactions. The bottom right corner shows the system status bar with the date 9/21/2024 and time 7:29 PM.

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.

This screenshot shows the same Power Query Editor interface as above, but the 'transactions' table is now fully visible in the main preview area. The 'Applied Steps' pane on the right lists a single step named 'Source'. The bottom status bar indicates 'PREVIEW DOWNLOADED AT 7:29 PM' and the date '9/21/2024'.

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

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.

Custom Column

Add a column that is computed from the other columns.

New column name:

Custom column formula:

Available columns:

- permitdate
- store_nbr
- transactions

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel

3 COLUMNS, 199+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 7:29 PM

= Table.AddColumn(Source, "GroupBy", each 1)

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

4 COLUMNS, 199+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 7:33 PM

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

The screenshot shows the Power Query Editor interface. A table is displayed with three columns: 'permidate', 'store_nbr', and 'transactions'. The 'GroupBy' column is highlighted. A context menu is open at the top of the 'GroupBy' column, with 'Group By...' selected. The 'APPLIED STEPS' pane on the right shows a single step named 'Added Custom'. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED AT 7:33 PM' and the date '9/21/2024'.

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.

The screenshot shows the Power Query Editor with the 'Group By' dialog box open. The dialog box allows specifying columns to group by and one or more outputs. Under 'Advanced' settings, two aggregations are defined: 'minDate' with operation 'Min' and 'permidate' column, and 'maxDate' with operation 'Max' and 'permidate' column. The 'OK' button is visible at the bottom right of the dialog box. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED AT 7:33 PM' and the date '9/21/2024'.

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 shows the Power Query Editor interface with the following details:

- File** tab is selected.
- Transform** ribbon tab is active.
- Queries [4]** pane on the left lists: transactions, train-2, DailyDelhiClimateTest, and Query1 (selected).
- Applied Steps** pane on the right shows a single step: "Grouped Rows".
- Preview** pane displays a table with three columns: "minDate" (1/1/2013), "maxDate" (8/15/2017), and a third column containing the number 1.
- Query Settings** pane shows the query name as "Query1".
- Properties** pane shows the query name as "Query1".
- Bottom status bar: 3 COLUMNS, 1 ROW, PREVIEW DOWNLOADED AT 7:36 PM, 7:37 PM, 9/21/2024.

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

The screenshot shows the Power Query Editor interface with the following details:

- File** tab is selected.
- Transform** ribbon tab is active.
- Queries [4]** pane on the left lists: transactions, train-2, DailyDelhiClimateTest, and Query1 (selected).
- Applied Steps** pane on the right shows a single step: "Filtered Rows".
- Preview** pane displays a table with columns: id, date, store_nbr, family, and sales. The data shows various grocery items and their sales figures.
- Query Settings** pane shows the query name as "train-2".
- Properties** pane shows the query name as "train-2".
- Bottom status bar: 6 COLUMNS, 199+ ROWS, PREVIEW DOWNLOADED AT 10:38 PM, 10:47 PM, 9/21/2024.

Tutorial 2: Creating Relationships

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.....

The screenshot shows the Power BI Model view. On the left, the 'Tables' pane lists 'DailyDelhiClimateTest' and 'train-2'. The 'Model' pane on the right shows a list of relationships. A 'New relationship' dialog box is open, showing 'DailyDelhiClimateTest' as the 'From table' and 'train-2' as the 'To table'. Under 'Cardinality', 'One to many (1:*)' is selected. Under 'Cross-filter direction', 'Single' is selected. A checkbox 'Make this relationship active' is checked. The 'Save' button is highlighted.

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

The screenshot shows the Power BI Model view. On the left, the 'Tables' pane lists 'DailyDelhiClimateTest' and 'transactions'. The 'Model' pane on the right shows a list of relationships. A 'New relationship' dialog box is open, showing 'DailyDelhiClimateTest' as the 'From table' and 'transactions' as the 'To table'. Under 'Cardinality', 'One to many (1:*)' is selected. Under 'Cross-filter direction', 'Single' is selected. A checkbox 'Make this relationship active' is checked. The 'Save' button is highlighted.

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.

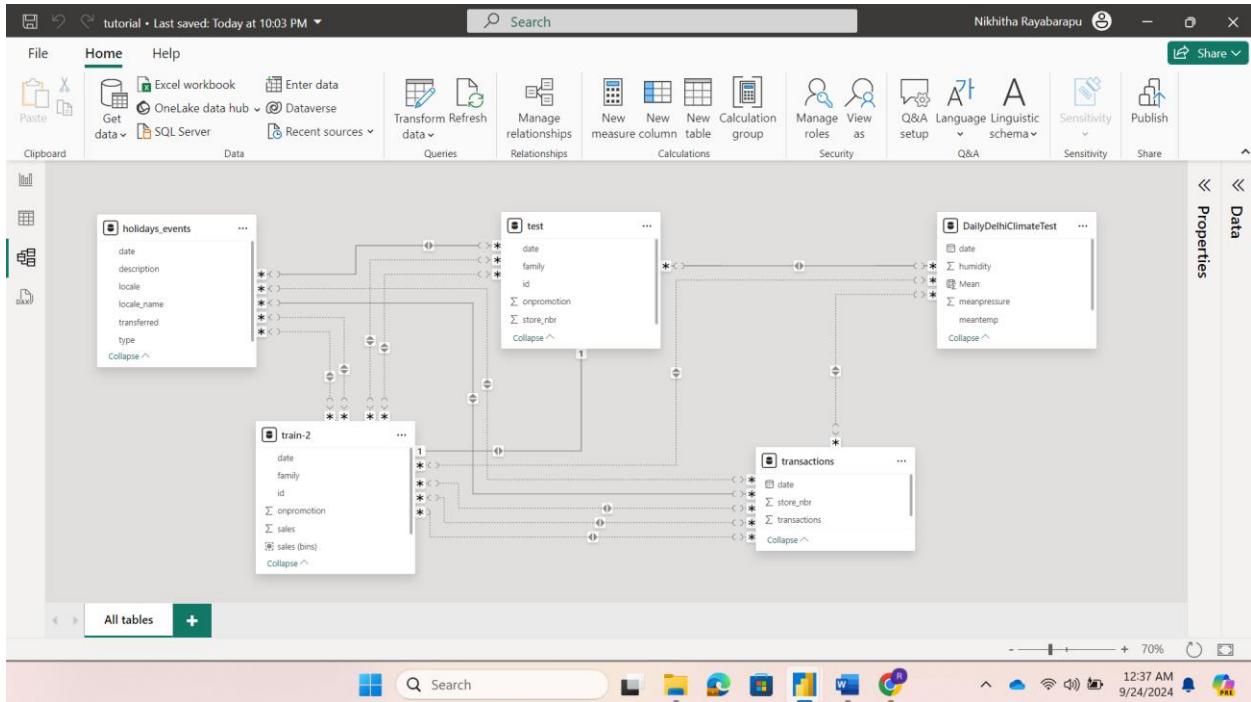
The screenshot shows the Power BI Data Model interface. A 'New relationship' dialog box is open, prompting the user to select related tables and columns. The 'From table' dropdown is set to 'train-2', and the 'To table' dropdown is set to 'transactions'. Both tables are displayed as data grids. The 'sales' column in the 'train-2' grid and the 'transactions' column in the 'transactions' grid are highlighted with green boxes, indicating they are selected for the relationship. Below the grids, the 'Cardinality' dropdown is set to 'Many to many (*.*)' and the 'Cross-filter direction' dropdown is set to 'Both'. There are three checkboxes at the bottom: 'Make this relationship active' (unchecked), 'Apply security filter in both directions' (unchecked), and 'Assume referential integrity' (unchecked). The 'Save' button is visible at the bottom right of the dialog.

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.

The screenshot shows the 'Manage relationships' dialog box. It lists several relationships between different tables in the model. The columns in the table are 'From: table (column)', 'Relationship', 'To: table (column)', and 'Status'. The 'Status' column indicates whether each relationship is 'Active' or 'Inactive'. The 'Close' button is visible at the bottom right of the dialog.

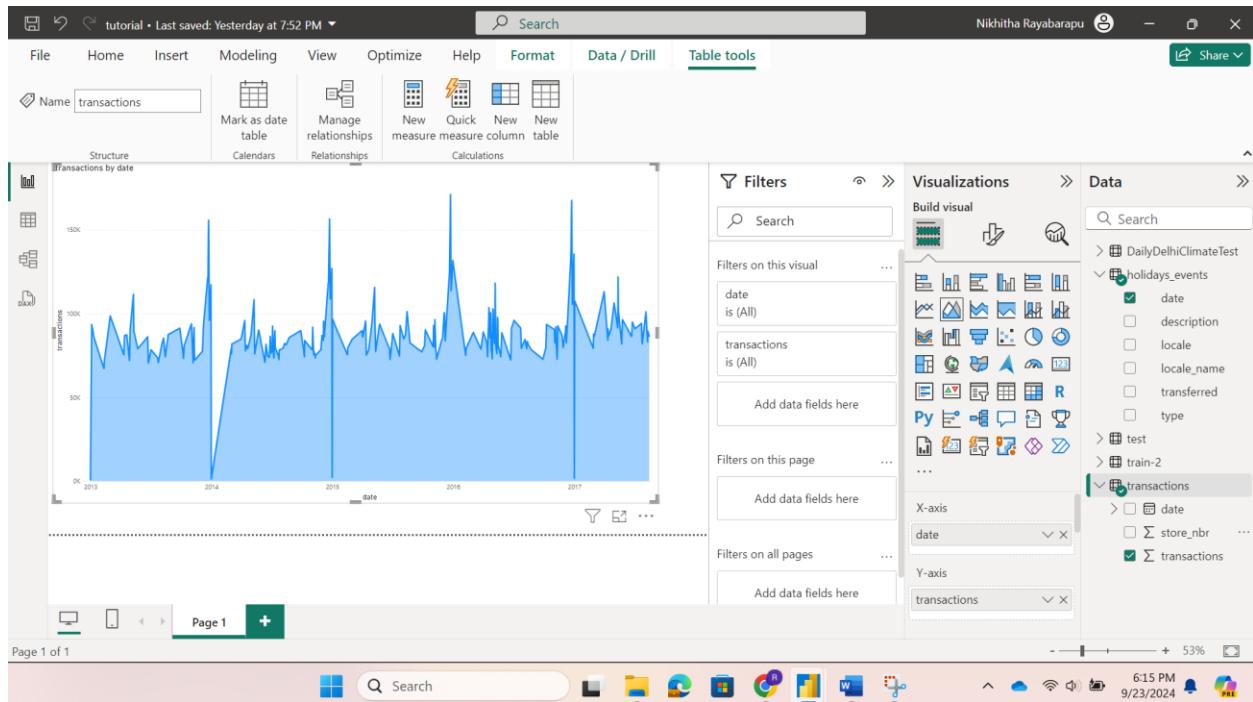
From: table (column)	Relationship	To: table (column)	Status
DailyDelhiClimateTest (meantemp)		transactions (transactions)	Inactive
DailyDelhiClimateTest (wind_speed)		train-2 (sales)	Inactive
holidays_events (date)		test (date)	Active
holidays_events (date)		train-2 (date)	Inactive
holidays_events (date)		transactions (date)	Active
holidays_events (type)		train-2 (sales)	Inactive
holidays_events (type)		transactions (transactions)	Inactive
test (date)		DailyDelhiClimateTest (meantemp)	Active
test (onpromotion)		train-2 (onpromotion)	Inactive
test (store_nbr)		train-2 (store_nbr)	Inactive

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

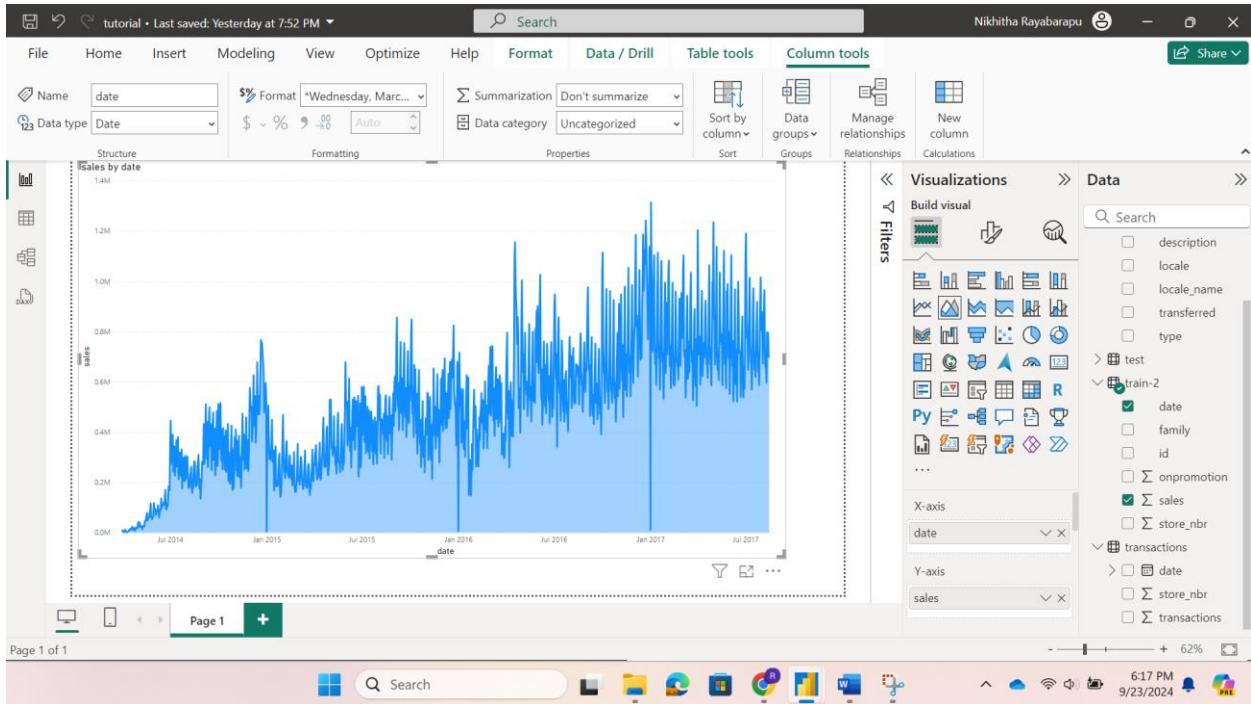


Tutorial 3: Visualizing using different tables

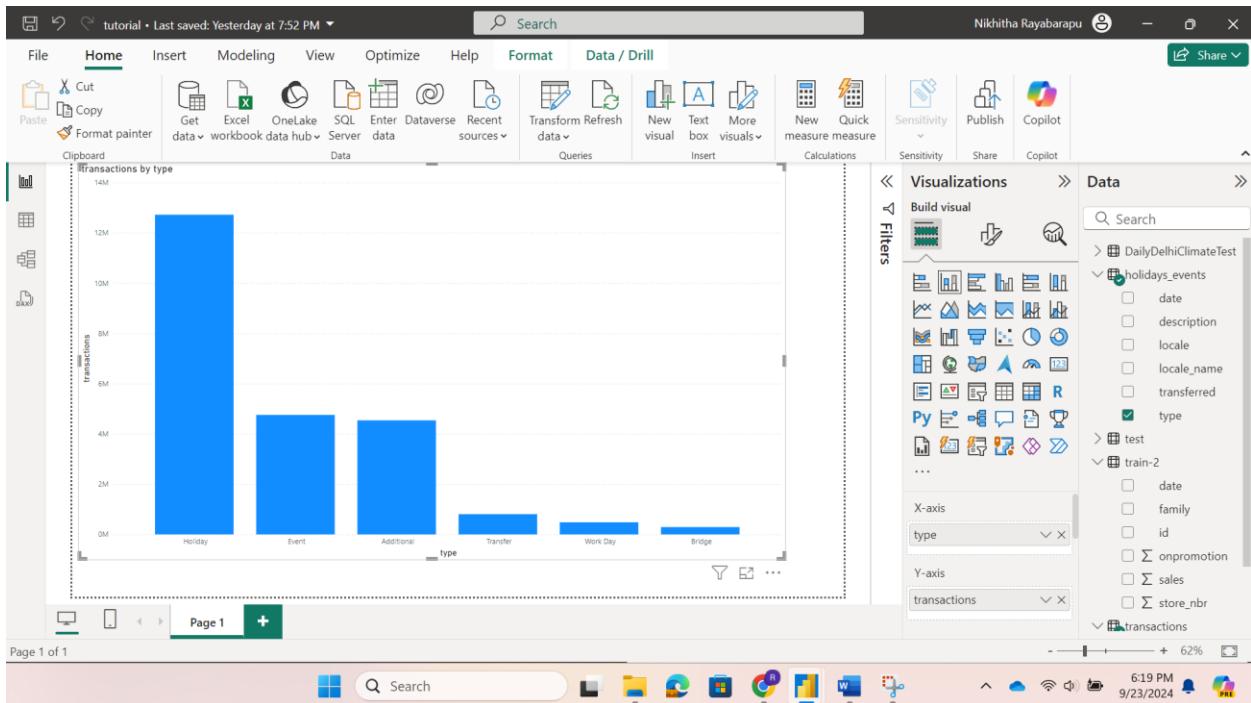
Step 1:- Select the date from the holidays_events table and transactions from the transaction table and click on the Area chart to visualize the sum of transactions with respect to years.



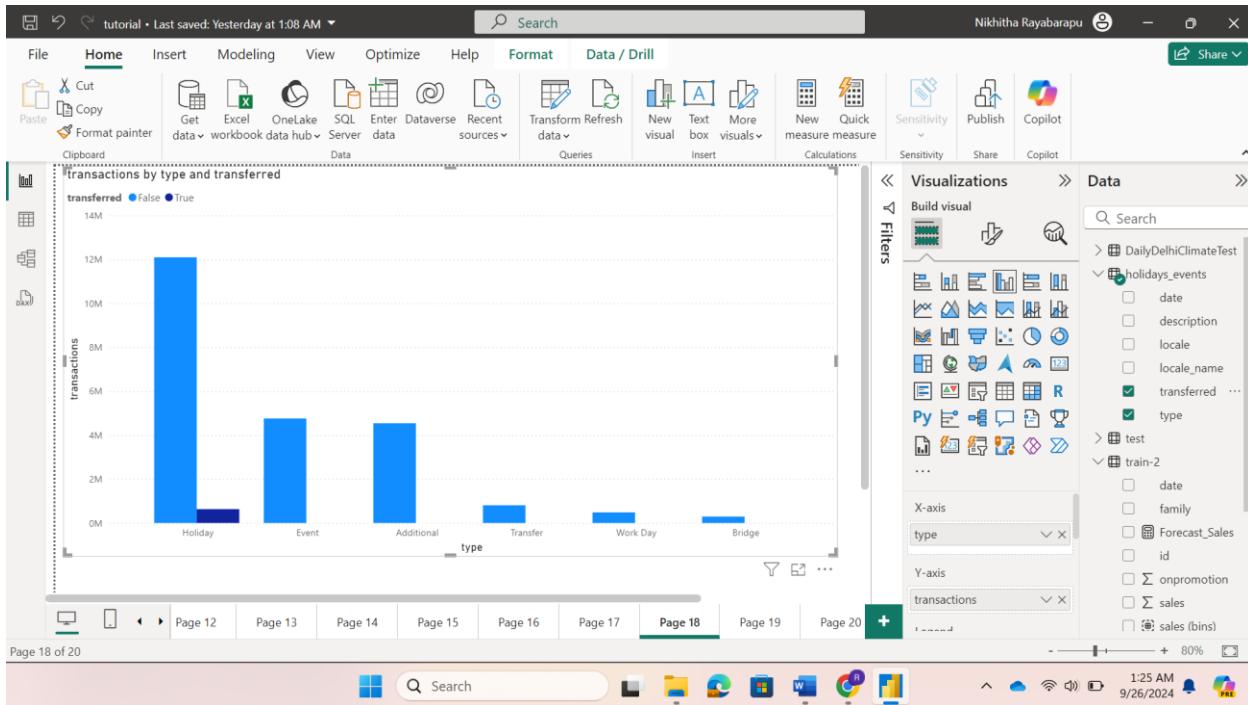
Step 2:- Now I am selecting the sales and date from the same train table and use the area chart to visualize the sales across years.



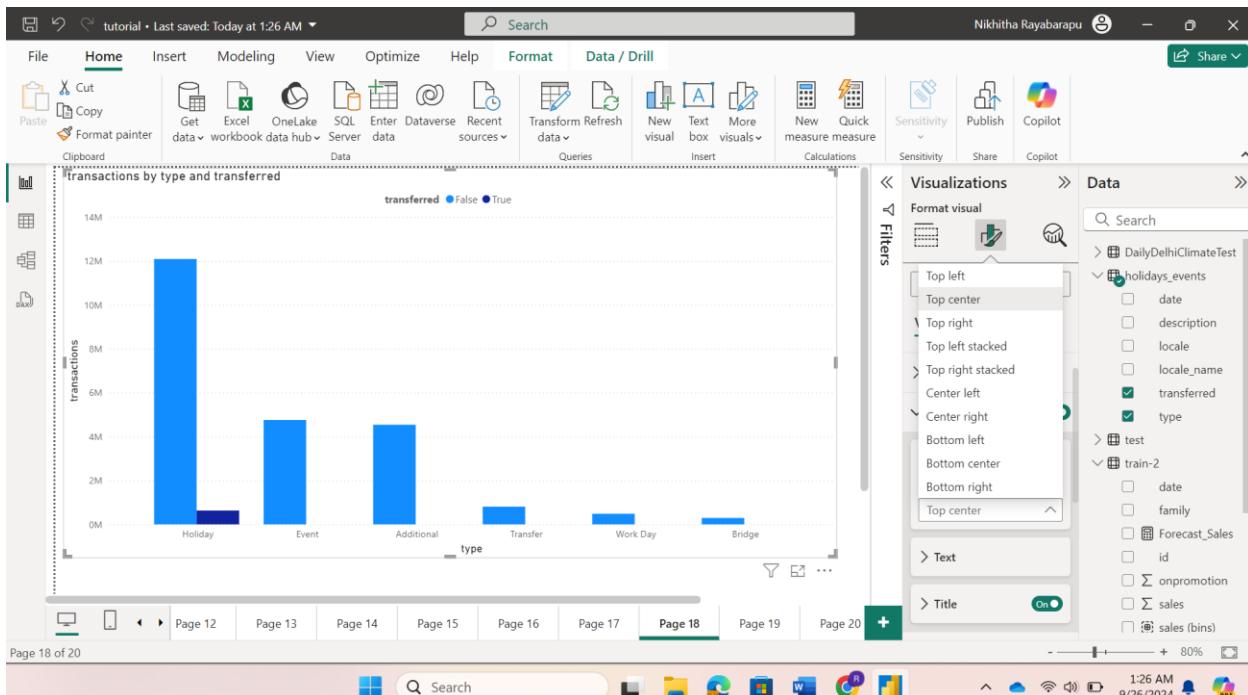
Step 3:- Select the Type from holidays_events and transactions and select the visualization.



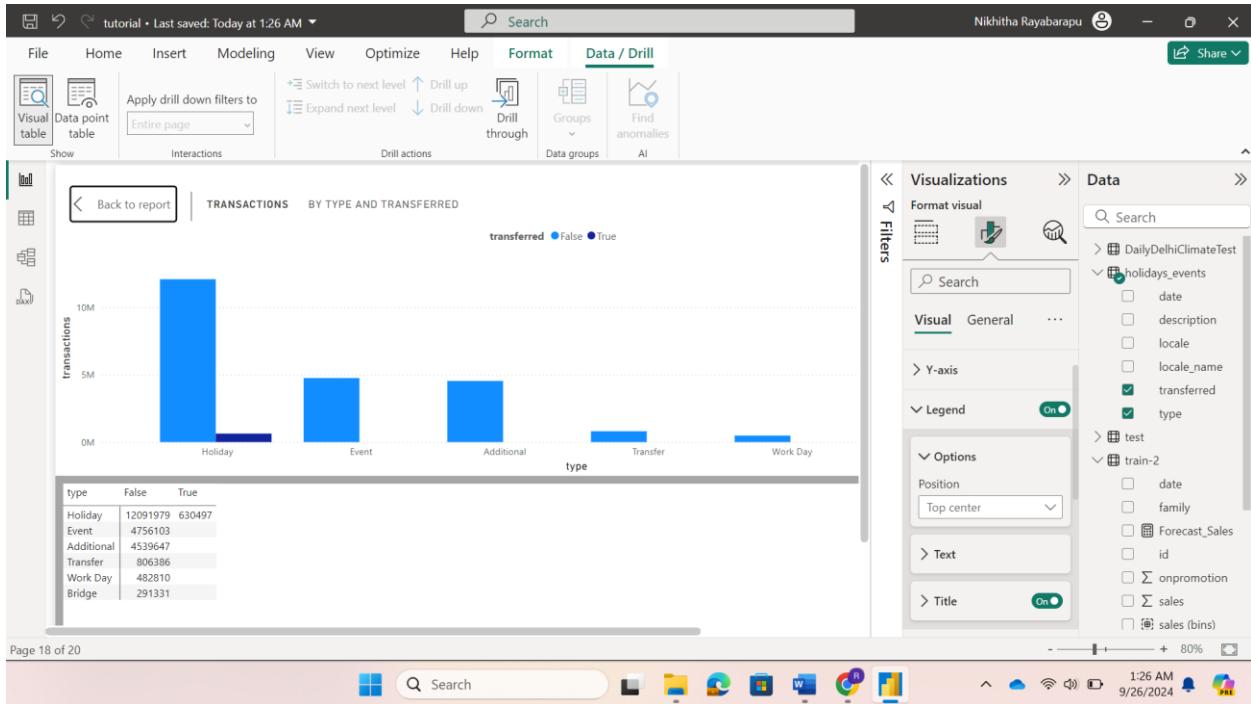
Step 4:- Now add another attribute i.e. Transferred



Step 5:- place the legend at the top center

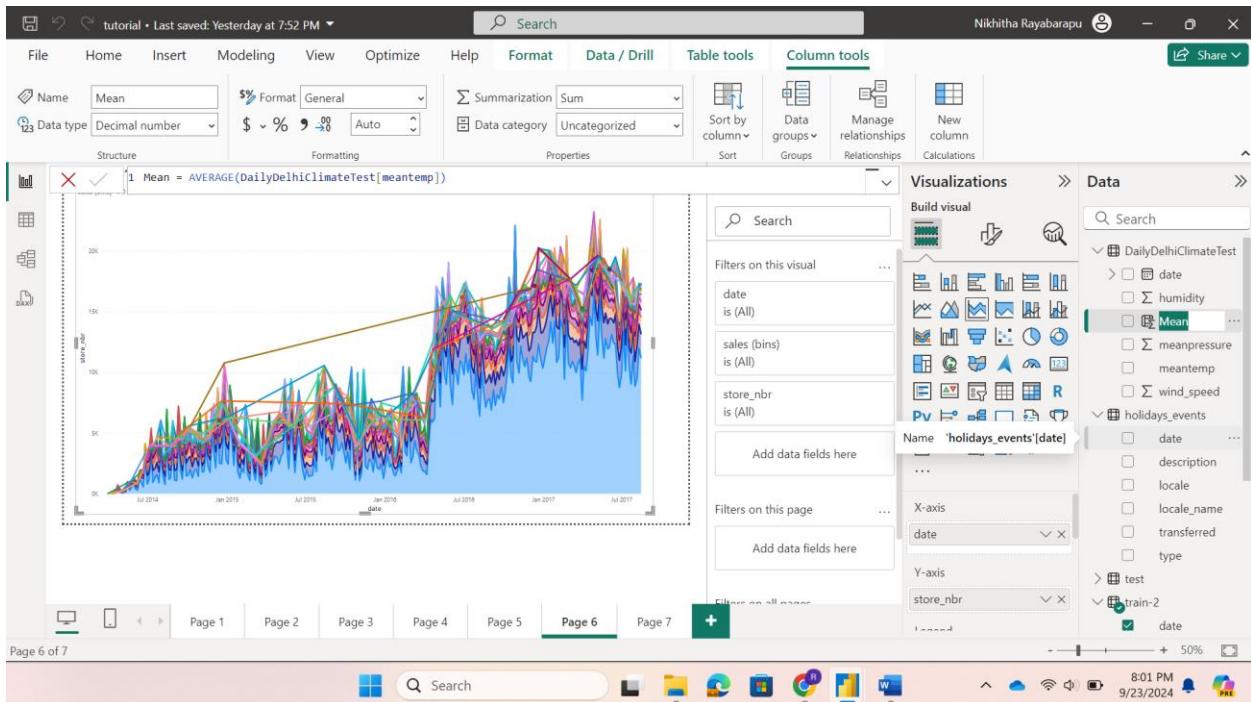


Step 6:- Click on visual table in Data/Drill to get the information about the graph in tabular format.

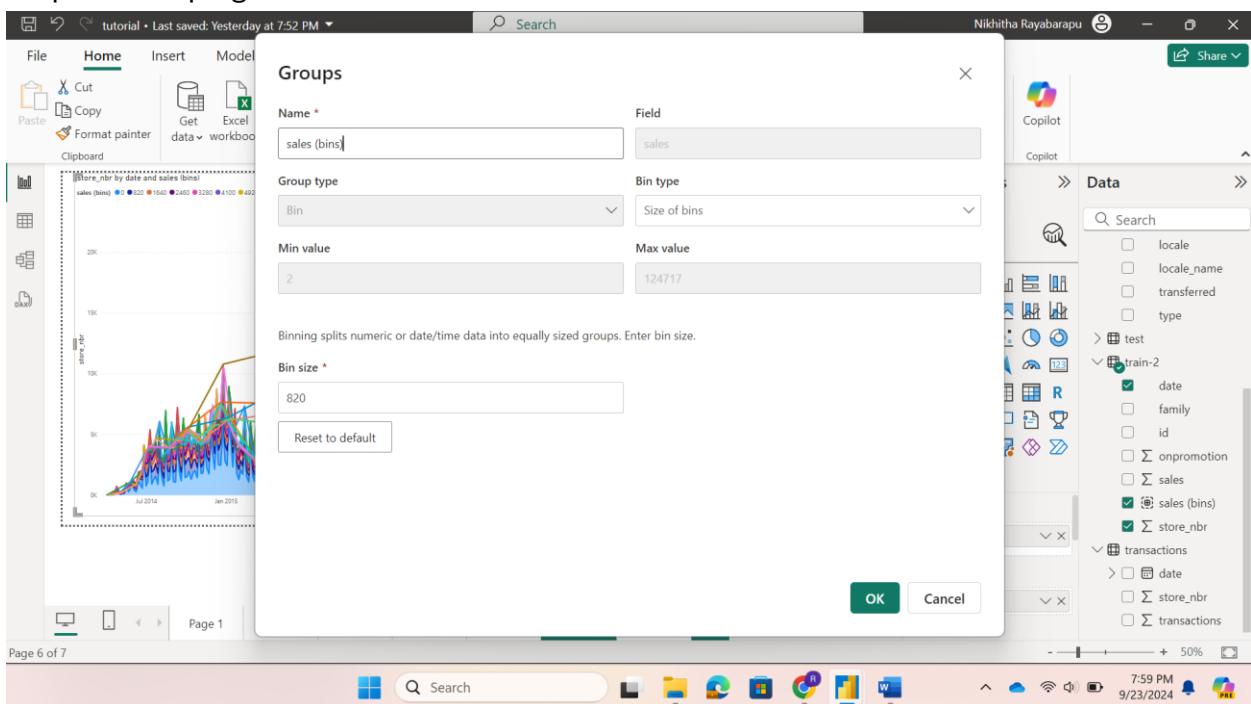


Tutorial 4: Creating an attribute to group the attributes

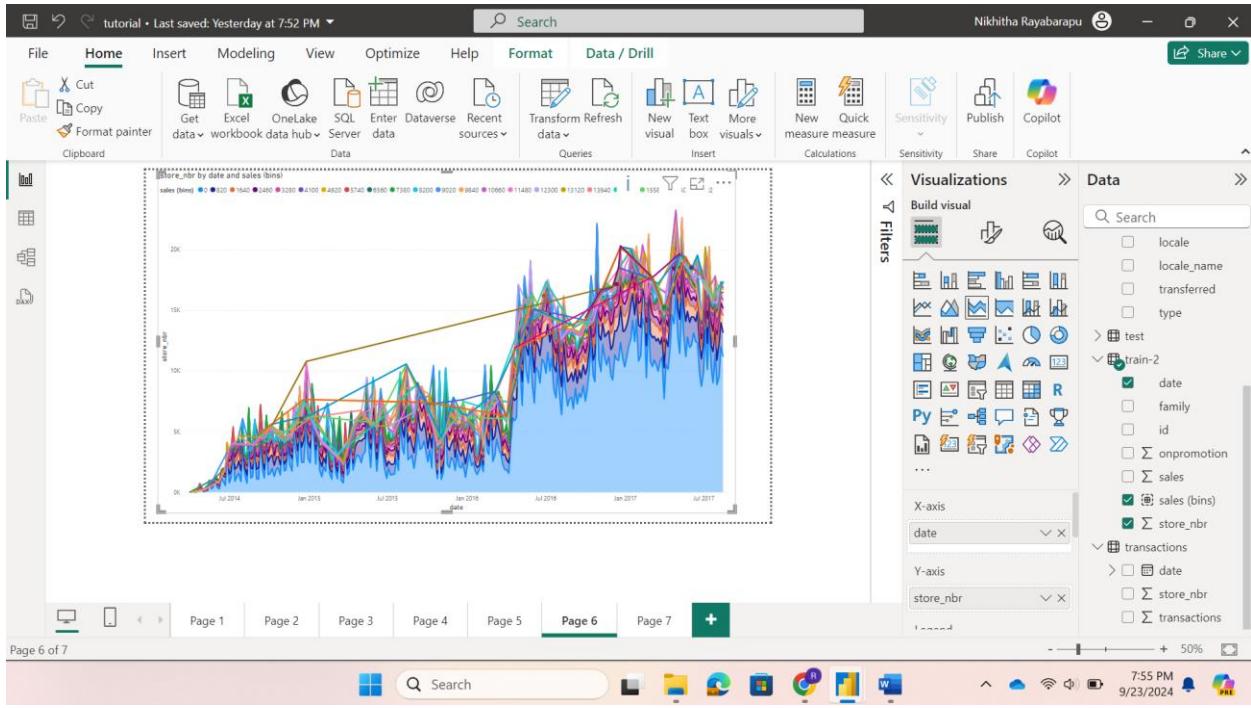
Step 1: Click on Custom Column and give the below query to get the mean. If you want to insert the attributes then just click on the attributes and click on insert to use them in the query.



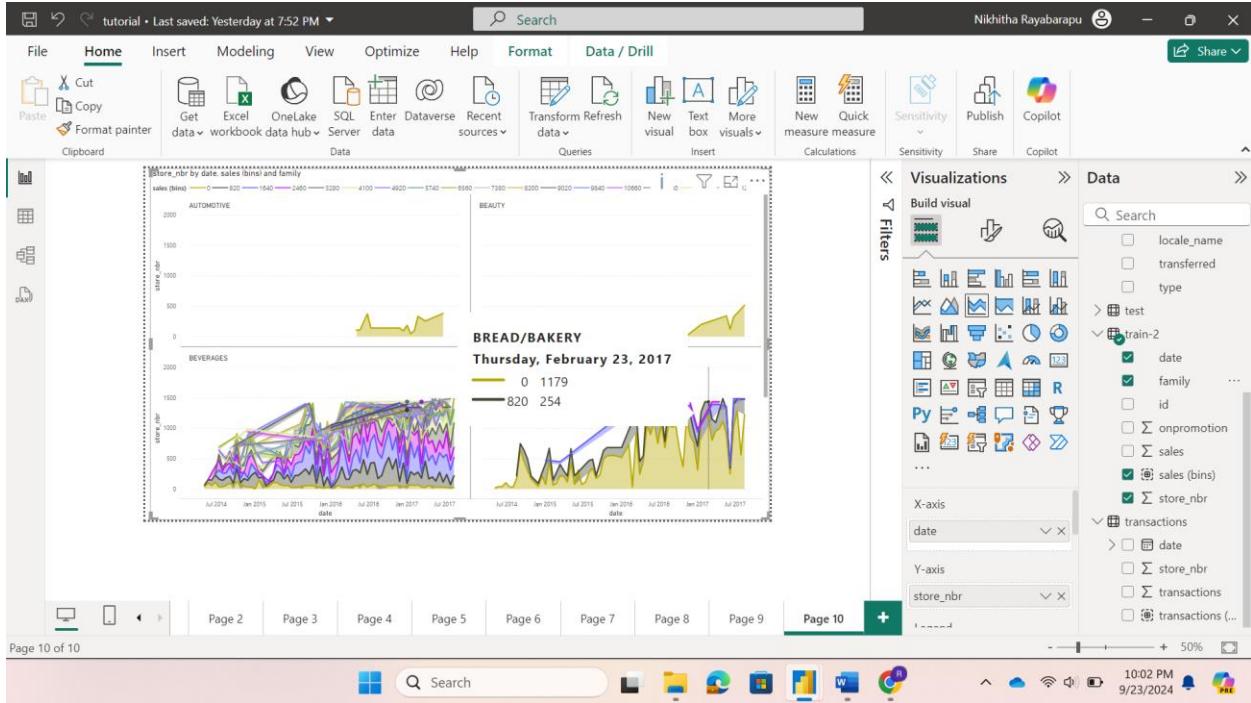
Step 2:- Grouping based on a sales attribute.



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.

Screenshot of Microsoft Power BI interface showing a stacked area chart and its corresponding visual table.

Visual Area:

- Title:** STORE_NBR BY DATE, SALES (BINS) AND FAMILY
- Legend:** sales (bins) (0, 820, 1640, 2460, 3280, 4100, 4920, 5740, 6560, 7380, 8200, 9020, 9840, 10660, 11480, 13300, 15120)
- Stacked Area Chart:** Displays sales over time for four categories: AUTOMOTIVE, BEAUTY, BEVERAGES, and BREAD/BAKERY.
- Data Table:** A detailed table showing daily sales figures for each category across various dates.

Data View:

- Visualizations:** A list of available visual types including charts, tables, and maps.
- Data:** A list of data sources and their fields:
 - test:** date, family, id, onpromotion, sales, store_nbr, transactions
 - train-2:** date, family, id, sales (bins), store_nbr, transactions
- Filters:** A section for applying filters to the data.

Page 10 of 10

10:05 PM
9/23/2024

Tutorial 5: Forecasting

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.

The screenshot shows the Power BI desktop interface. The ribbon is at the top with tabs like File, Home, Insert, Modeling, View, Optimize, Help, Table tools, and Measure tools. The Measure tools tab is selected. In the center, there's a code editor window with the following DAX query:

```
1 Forecast_MeanTemp = AVERAGEX(TOPN(12, 'DailyDelhiClimateTest', 'DailyDelhiClimateTest'[date].[Date], DESC), 'DailyDelhiClimateTest'[meantemp])
```

Below the code editor, there's a message: "Build visuals with your data. Select or drag fields from the Data pane onto the report canvas." To the right, the Data pane shows the structure of the "DailyDelhiClimateTest" table, including fields like date, meantemp, humidity, meanpressure, and wind_speed. The status bar at the bottom indicates "Page 12 of 12".

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

The screenshot shows the Power BI desktop interface again. The ribbon has the Measure tools tab selected. The code editor contains the following DAX query:

```
1 Forecast_Sales = AVERAGEX(TOPN(12, 'train-2', 'train-2'[date], DESC), 'train-2'[sales])
```

The main area displays a line chart titled "Sunday, July 23, 2017" with a single data series labeled "Forecast_Sales" showing values around 1,558.38. The X-axis is "date" and the Y-axis is "Forecast_Sales". The Data pane on the right shows the "train-2" table structure, including fields like date, family, id, onpromotion, sales, sales(bins), and store_nbr. The status bar at the bottom indicates "Page 12 of 12".