CSCE 5320 Section(s) 003,600 (Spring 2024 1)

Scientific Data Visualization

Name: Reddybathuni Jyothirmai
Student ID: 11658930

Tutorial-1 Explanation:
In the tutorial, I first loaded the Transactions.csv dataset into the Power BI application to perform data transformations (see Fig. 1.). Later, using the Power Query Editor, I made some changes to the data like renaming the column and also loading new datasets(see Fig. 2.). I used the Custom Column option to add a new column to the existing dataset(Fig. 4.). I then performed the Group By operation in Fig 6. In Fig 7, we can see the result from the Group By operation. The min date and the max date are displayed from the dataset. In Fig 8, I performed a query operation using the Power Query editor; the query results in the rows which have sales greater than 1 and on promotion greater than 1.

Question-1 Explanation:
I first renamed the store_nbr column name to store_id in all the tables as shown in Fig 9. Using the tutorial as a reference, I added another custom column to perform the Group By operation in Fig 10. I then added the group by function with the minimum and maximum dates from the date attribute in the test dataset. In Fig 12, we can see the result of the Group By function. I used the Power Query Editor to perform the query, which resulted in rows with a store_id of less than 700000. The query and the output of the query can be seen in Fig 13. When compared to Tableau, Power BI was so complex and rigid. The user interface of Tableau is smoother and easier to understand Power BI. Initially, I needed to understand how the queries work and the Power Query Editor.

Question 1. A
I filtered the dates in the transactions dataset by clicking the down arrow which is present beside the name of the column. I found the option to sort the values in descending order(see Fig.14). The values of the date column are sorted in a descending manner in Fig 15.

Tutorial-2 Explanation:
In Fig 16, I created a relationship between the date attribute in the DailyDelhiClimateTest table and the sales attribute in the train table. It is a one-to-many relationship. I chose cross-filter detection for both sides and made the relationship active. In Fig 17, I made another relationship between attributes like the previous one. Fig 18 shows all the relationships that can be made active or all the relationships that are present in the model view. Fig 19 is the model view, a graphical representation of the tables and the relationships between them.

Question-2 Explanation:
I could find all the one-to-one and many-to-one relationships between the attributes by manually checking each table and all the attributes in them. I could find 2 one-to-one relationships and 10 Many-to-one relationships.

Cardinality can affect how attributes interact with each other.

- In a one-to-one relationship, every instance of one table relates to exactly one instance of another, and vice versa. For example, in Fig 22, we can see that the date attribute in the DailyDelhiClimateTest table is exactly related to the id attribute in the train-2 table.
- In a one-to-many relationship, each instance of one table relates to one or more instances of another. Still, each instance of the second table can be related to only one instance in the first table.
- In a many-to-many relationship, each instance of one table is related to one or more instances of another, and vice versa.
- In a many-to-one relationship, each instance of one table relates to one instance of another. Still, each instance of the second table can be related to one or more instances in the first table.

Question 2A :

When we select cross filter direction as single, that means that if there are any changes in the table, the changes are visible in all the tables that it is linked to but not in the original table. If we use Fig 31 as a reference, the cross-filter direction is selected as both, therefore if there is a change in either DailyDelhiClimateTest or in the transactions table, it will be affected in both of them. From Fig 32, we can see that I selected the cross-filter detection as single. If there is a change in the transaction table, the changes are visible only in the transaction table and not in the DailyDelhiClimateTest table. It works the same way in the opposite direction. Changes in DailyDelhiClimateTest will not affect the transactions table.

Question 2B:

No, we can not have multiple relations in one table and set them as active. In Fig 33, I tried to make two relationships that already have an indirect relationship active. PowerBi gave a pop-up that said we could not make it active.

Tutorial 3 Explanation:

In Fig 34, I used the date attribute from the holiday_events table and the transactions attribute from the transaction table, which is visualized using the Area chart. In the next Fig 35, I added the sales attribute and date attribute to the train-2 table. From this visualization, we can identify the trend of sales changing each year. In Fig 36, I used a bar chart to visualize the total number of transactions by each type. Fig 37 is an extension of the previous visualization. I just added an attribute from the holiday_events called transferred. In Fig 38, I placed the legend in the top center using the format visual option. In Fig 39, I used the visual table to describe the data visual table.

Question 3 Explanation:

In Fig. 40, I selected the attributes mentioned in the tutorial, which are Transferred, Type, and Date from the holiday_events and transactions from the transactions table. I visualized it in a Table(see Fig 40). Fig 41 is a stacked bar chart that shows the sum of transactions by type, date, and transactions. In Fig 42, I used a waterfall chart to understand the changes in the

attributes. We can see that the attributes are kept the same, but the visualization and the insights we can derive from the visualization change. The waterfall chart is mostly used to understand the changes in a particular value and how it changes over time. In Fig 43, we can see the sum of sales for each year using a clustered column chart. In Fig 44, I made a visualization using a stacked bar chart with the same attributes. We can see the changes in the result that the attributes are swapped with row and column. The stacked bar chart made it easier to interpret the data. Fig 45 is the waterfall chart for the date, sales, and store_id from the train-2 dataset.

Question 3A

Fig 40 and Fig 46 are the visual tables for the attributes on separate pages.
The highest sale in 2016 was 124717, and the 2nd highest was 87439, which is 37278 lower than the highest sale in 2016.

Question 3B

Attributes work differently for different visualizations because of various reasons, in Fig 42,43 we can observe that because of a change in the type of visualization, the attributes change. Therefore, the major reason for the change is that the attributes automatically adjust according to the visualization and how the data is used in the particular task. The relationships created in Task 2 are also a factor for the change in attributes.

Tutorial 4 explanation:

In Fig 48, I created a custom column with the formula given. The formula is used to identify the mean of the values in the attribute. Then, by right-clicking on the attribute in the Data pane, I could group the values using the default settings given in Power BI(see Fig 49). I created a stacked area chart using the attributes created in the previous steps(see Fig 50). From the visualizations, we can identify the sum of sales in a specific month and year. In the next Fig 51, I added the family attribute to the visualization, so now I can clearly identify the sales for each category. In Fig 52 I also added the table dataset for a clearer understanding. The table and the stacked area chart are interlinked, and if we click on a particular row in the table, we can see the visualization only regarding that row.

Question 4 explanation:

I worked with a transaction table that had various attributes. To understand the data better, I grouped the transactions by all the available attributes and removed the Quarter from the date hierarchy(see Fig 53). I then proceeded to create two visualizations. For the first visualization, I used a stacked column chart(see Fig 54), and for the second one, I opted for a scatter plot(see Fig 55). However, I noticed that the attributes for the visualizations were different. While the X and Y axis remained constant for the column chart and line graph, they were different for the scatter plot. I tried to make a scatter plot with the same axes, but the results were messy.(Fig 56).

Question 4A

From fig 57 and 58, we can see that the sum of transactions is greater than the count of transactions because the sum is a measure that adds up the values in the transaction whereas count is used to find the number of values in the transaction attribute.

Tutorial 5 explanation:
I used the new measure option to create a new attribute, which is the average of the 12 rows from the original dataset for the meanTemp(see Fig 59). I also used the same query to find the average of sales attributes and performed visualizations with those attributes(Fig 60). In Fig 61, we can see the visualizations for the measures created.

Question 5 Explanation:
In Fig 62, I created a new measure with a query implemented on the transactions table which averages the transaction. I used the same measure to build visualisation which can be seen in Fig 63 and Fig 64. Fig 63 is a Clustered column chart that shows the result of the query I created and the transactions. Fig 64 is a Funnel chart that shows the same result as the previous visualization.

Question 5A:
Time series forecasting enables firms to predict future trends and patterns using historical data. This predictive information enables businesses to make better decisions, plan resources more wisely, and forecast alterations in the market. Time series forecasting allows companies in real-life situations to make decisions based on analytics rather than human decisions. This can help companies generate profit and revenue.

Question 5B:
I learned how to transform data and perform queries in Power BI. I found it really complicated initially, but with the help of internet resources, I was able to complete the task. In this activity, I added relationships between different datasets, calculated new measures, added new custom columns, created visualizations for various attributes, and understood the changes in attributes.