**Tableau Introduction-Assignment 1**

1. “twbx” is a bundled workbook for Tableau. The original .twb file is bundled with the data source in this package. It can be compared to a compressed file. It contains all the information and instructions required to operate in Tableau. Since the data is contained within the .twbx file itself, one can still access and use it without a network or Internet connection. The .twb file and the data source can be separated from the .twbx file by unpacking it. Pick up any dataset of your choice, create a simple bar chart using the fields of the dataset and save the visualization created in .twbx format. Analyze the properties of the newly created twbx file and segregate the .twbx file into .twb and data source.
2. Briefly explain the utility of the Tableau bookmark feature and create a simple bookmark file. Observe the format of the bookmark file and mention the location in which it is saved.
3. Using the “Sample-Superstore.xls” file, create a scatter and a bubble plot between different measures in the dataset and observe the type of correlation (negative or positive or no correlation) between them. Draw a comparison between the bubble chart and the scatter plot.
4. Consider that you are an HR representative for a multinational company. The staff database is under your control. There are certain details regarding employees that you must never divulge. However, there are many bits of information about employee abilities and skills that may be shared. Using the data extract option in tableau, build a packaged worksheet and use the option "Hide All Unused Fields" in the data extract feature to hide all the fields- dimensions and measures which you haven’t used in the visualization and do not wish to share with employees. Feel free to use any HR dataset or you may even create a dummy data for illustration purpose.
5. Discuss the differences between the “Measure Names” and “Measure Values” pre-defined features in Tableau. Using the “PowerStore\_USA” dataset available in your iNeuron resources, create a visualization using “Measure Names” and “Measure Values” and mention the fields that fall into each category- “Measure Names” and “Measure Values”.

[*PFA links for the datasets used.*](https://drive.google.com/drive/folders/123UyMRbrReCjyn1K4g_FhsjKx6cP4zLH?usp=sharing)

Ans 1)

i) The main difference between .twb file and .twbx file is .twb files contain information about the workbook structure, including worksheets, dashboards, calculated fields, etc where as twbx files contain the complete workbook structure ( .twb) along with the data source and any custom images.

ii) Another major difference is the difference in the file size the .twb file is smaller in size whereas .twbx is bigger in size.

iii) In .twbx files the data source is directly within the file and .twb files do not contain the data source.

Ans 2)

Utility of Bookmark Feature :

1. Bookmark feature captures the current state of the dashboard or worksheet, filters, parameters, etc. you can also create multiple bookmarks for different scenarios which enables comparison between various data points or time frames without having to manually adjust filters and parameters each time.
2. Bookmark feature enables for the storytelling of the tableau dashboard which you can explain while presenting the dashboard.
3. It helps while presenting for creating interactive presentations in tableau also by saving views user can quickly return to specific analysis points.

Path:

3) Ans

i) In the Scatter plot of sum of sales vs sum of profit we can see that there is significant positive correlation

vs ii) In the bubble plot of sum of sales vs sum of profit we can see that there is no sign of any correlation between the 2 quantities here the size of the bubble is the sum of sales of the particular quantity and color of the bubble corresponds to the profit from that particular quantity. Hence there is significant difference between the bubble plot and the Scatter plot.

1. Done simply “ Hide all Unused Fields ” from the Extract option.

1. As I was not able to find the dataset “PowerStore\_USA” I am answering the question based on the dataset named “Sample-Superstore”

Difference between measure names and measure values: Measure names fields contains the names of all the measure fields in the dataset. Whereas Measure values is the field that contains the values of all the measure fields in the dataset

Measure Names in the dataset are:

Order ID

Postal Code

Product ID

Product Name

Region

Ship Mode

Ship Date

Segment

Measure Values in the dataset are:

Discount

Profit

Quantity

Sales

Latitude

Longitude

Orders