designed by freepik.com

DATA VISUALIZATION IN TABLEAU

GEOVISUALIZATION AND ADVANCED CARTOGRAPHY

PRATICHHYA SHARMA (12031082)
DEPARTMENT OF GEOINFORMATION
University of Salzburg

Table of Contents

Table of Figures

Objective

To get familiar with Tableau for interactive visualization and exploration of data.

Tool used

Tableau Public 2020.3

Data Used

Blue bikes trip records of Feburary 2020 (csv) in Boston, USA

Overview

Data Visualization is the representation of data in a diagrammatic form. It sets a bridge for communication of data with simple visualization and provides the user with a visual summary of the information making it easier to identify pattern rather than looking through each data. To improve our skill in data visualization as a cartography student we were guided through a tool (Tableau here) by Prof. Barabara Hofer that would help an individual to present a piece of data as information.

In this task, I have used location-based data of February 2020 in Bostan provided by the Bluebikes System to know the use of the bike in the area just before the lockdown due to Corona. It was downloaded as an excel file containing data on Trip Duration in second; Start and Stop Date and Time; Start and End Station Coordinates, Name and ID, User-type, Birthyear and Gender. Through visual interpretation with the data, I was able to discover the answers to my several questions which I will be discussing in this document.

Questions

> **Question: Is there any Station having same End Station? If yes, what is the total trip duration from these common stations?**

Column Used: Start Station Lognitude, Start Station Lognitude, Start Station Name, End Station Name, Tripduration

Observation: Yes, there were few Start Station in different locations having the same End Station with varying trip duration. The observation was done in the form of a map as shown in Fig. 1. For better visualization, only the Start Station whose Total Trip Duration is a minimum of 2 mins are displayed. The higher the trip duration the bigger is the icon whereas colour is given based on End Station. So the Start Station having same End Station are given similar colour. Here, the reason behind the selection of OpenStreetMap as a base-map is to give a better visualization to the user who wish to see the station with respect to local routes. (Note: Color Value were randomly given.)
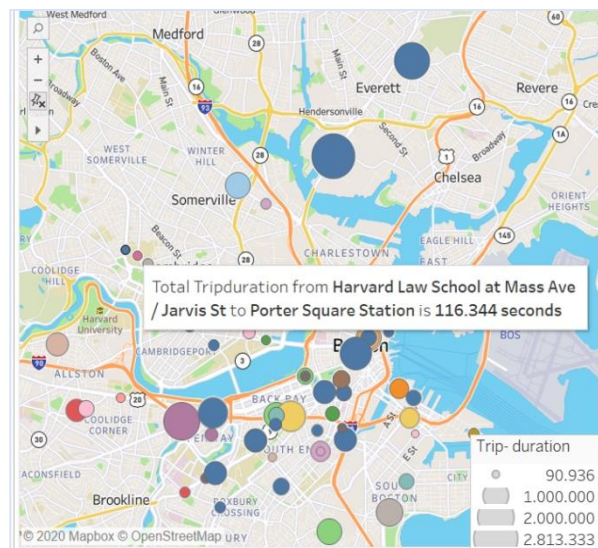

Fig 1: Map view with Start Station having same End Station

Question: At what time of the day, highest trip duration is observed?

Column Used: Starttime and Trip Duration

Observation: Total trip duration for each hour in a month was plotted in a line discrete chart against the weekdays and was observed that there is a peak trip in between 11 am to 6 pm in a week. It was identified that on Monday at 6 pm, Tuesday at 3 pm, Wednesday at 4 pm, Thursday at 5 pm, Friday & Saturday at 11 pm and on Sunday at 4 pm bike trip rises. Among all the weekday, it was a bit unusual but on Wednesday there was the highest bike trip with a total of 1.5 hour around 4 pm in February.
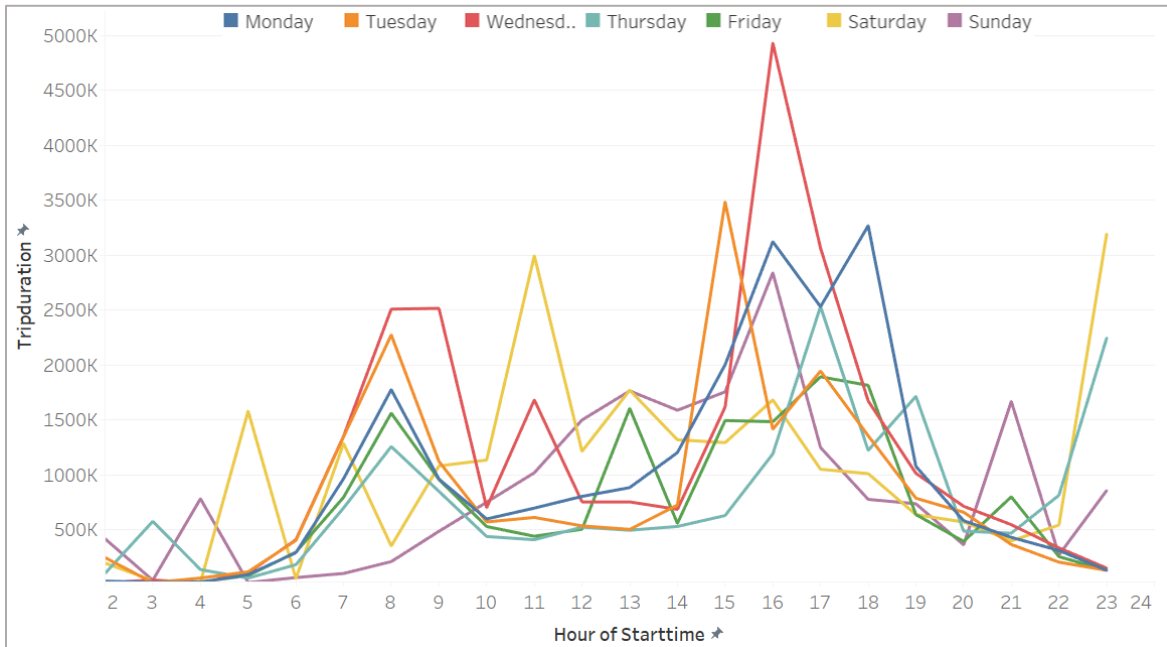

Fig 2: Line Chart for Total Trip Duration every hour in a week for a month

Question: What is the Average, Maximum and Minimum age of the people taking part in this survey?

Column Used: Gender

Observation: The data had over 132,825 records for participants of different age with different user-type travelled in bike around 41 hours in February. The data were grouped into several categories such as a participant's gender was included male, female and undefined (who didn't define their gender in the survey form) and each gender could be grouped as either customer (single trip) or subscriber (monthly and annual user). According to me, I believed that Stacked Chart could be the best representation at the moment. Based on used data it a simple chart as shown in Fig. 3 was obtained with the value as mentioned in the table.

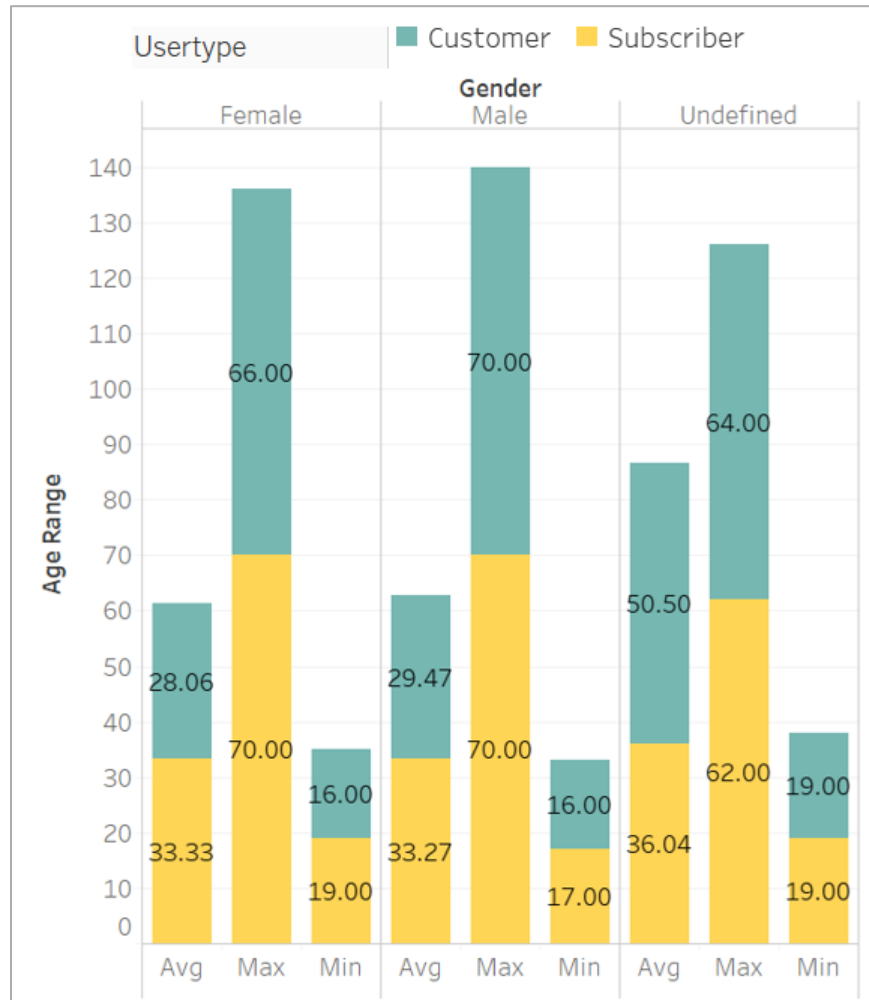|	Customer			Subscriber		
	Female	Male	Undefined	Female	Male	Undefined
Avg. Age	28	30	51	33	33	36
Max. Age	66	70	64	70	70	62
Min. Age	16	16	19	19	17	19

Fig 3: Participants Age Variation

Question: What is the total trip duration travel by each gender group? Can I also differentiate them based on both gender and user type?

Column Used: Gender, Tripduration and Calculated field based on Usertype for each Gender Value.

Observation: Yes, when thinking logical I can get a numerical figure for the total trip duration for each gender taking into consideration their user-type by using trip duration column instead of age in the previously generated stacked chart. Thus, I could successfully get an answer to my question that for the female customer total trip duration was 6,455.341 seconds whereas for subscriber was 21,646.035 seconds. Similarly, for customer and subscriber male, it was 14,141.298 seconds and 68,163.593 seconds respectively. Also, 35,355.462 and 790 seconds was recorded for undefined customer and subscriber respectively as shown in Fig. 4.

Though I was able to answer the question readily, when thinking out of the box I tried to represent the same data in the different form. Since I had the information for the total number of participant and total trip duration, I chose to go with a pie chart which may not be the best solution in such situation but could be a good solution. When preparing a pie-chart, I divided it based on gender which was sub categorised as Female Subscriber, Female Customer,

Fig 4: Total Trip Duration

Male Subscriber, Male Customer, Undefined Customer and Undefined Subscriber. The same value was obtained as from that of the previous method when computation was carried out. From the pie chart, I could say that among 6,825 Undefined, 96,964 Male and 29,036 Female, 91,074 Subscriber Male had the highest contribution with 68,163.593 seconds trip duration.
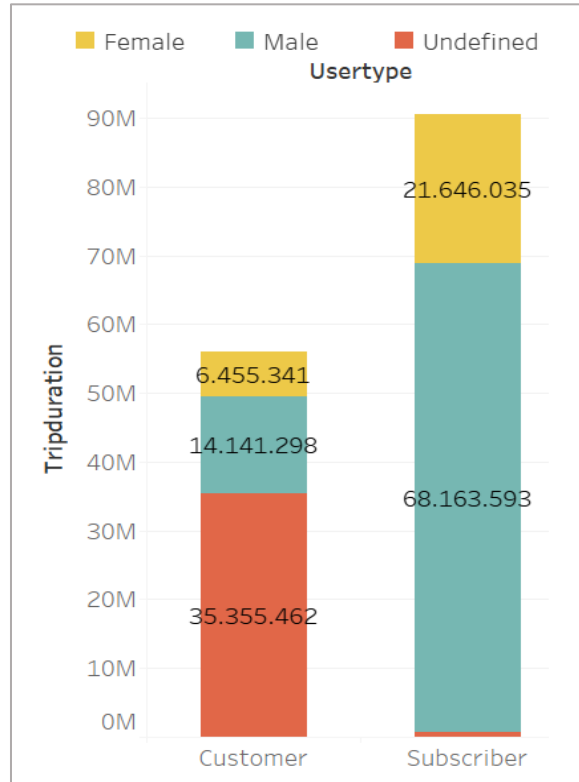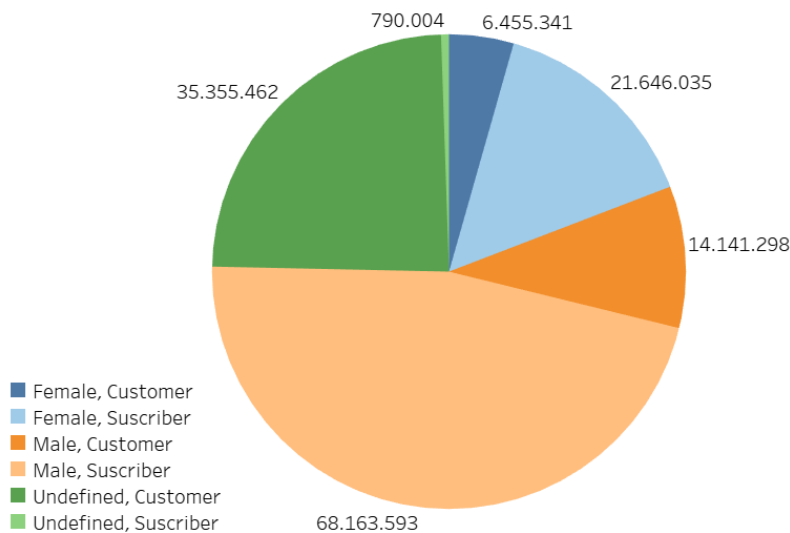

Fig 5: Pie-Chart for Total Trip Duration

Question: Which is the most popular start station with a minimum of 1000 records?

Column Used: Data Count and Start Station Name

Observation: Among 30 stations filtered out as having more than 1000 participants recorded, MIT at Mass Ave/ Amherst Station had the maximum with 3,739 participants making it the most popular station in Boston for February. The color is represented as white to white and size of box as small to big based on the information as minumum to maximum.
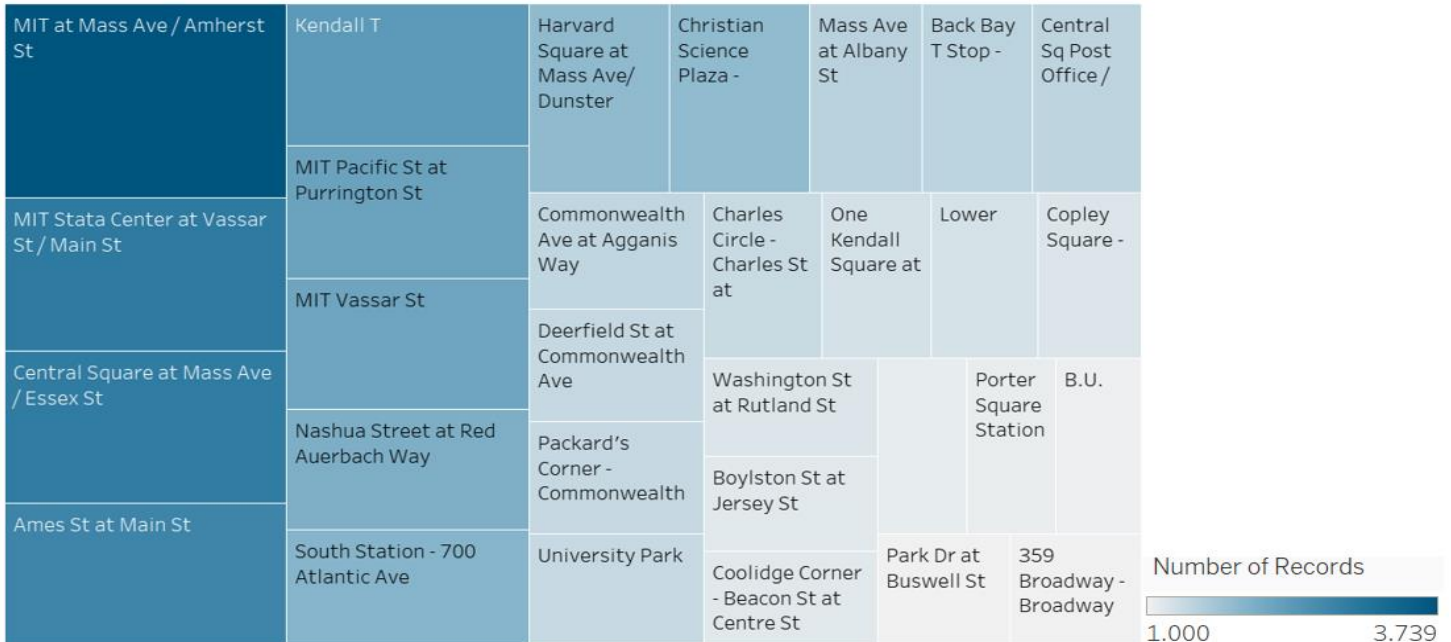

Fig 6: Most Popular Stations

Result

Tableau allowed me to combine all my sheet into a single page as a simple dashboard. I was able to create an interactive dashboard that summarized the answer to all my question as shown in Fig.7. In Addition to it, I was able to integrate an action functionality to the dashboard that allowed me to inspect on detail for each location reflecting in all analysis except the most Popular Station as shown in Fig.8.

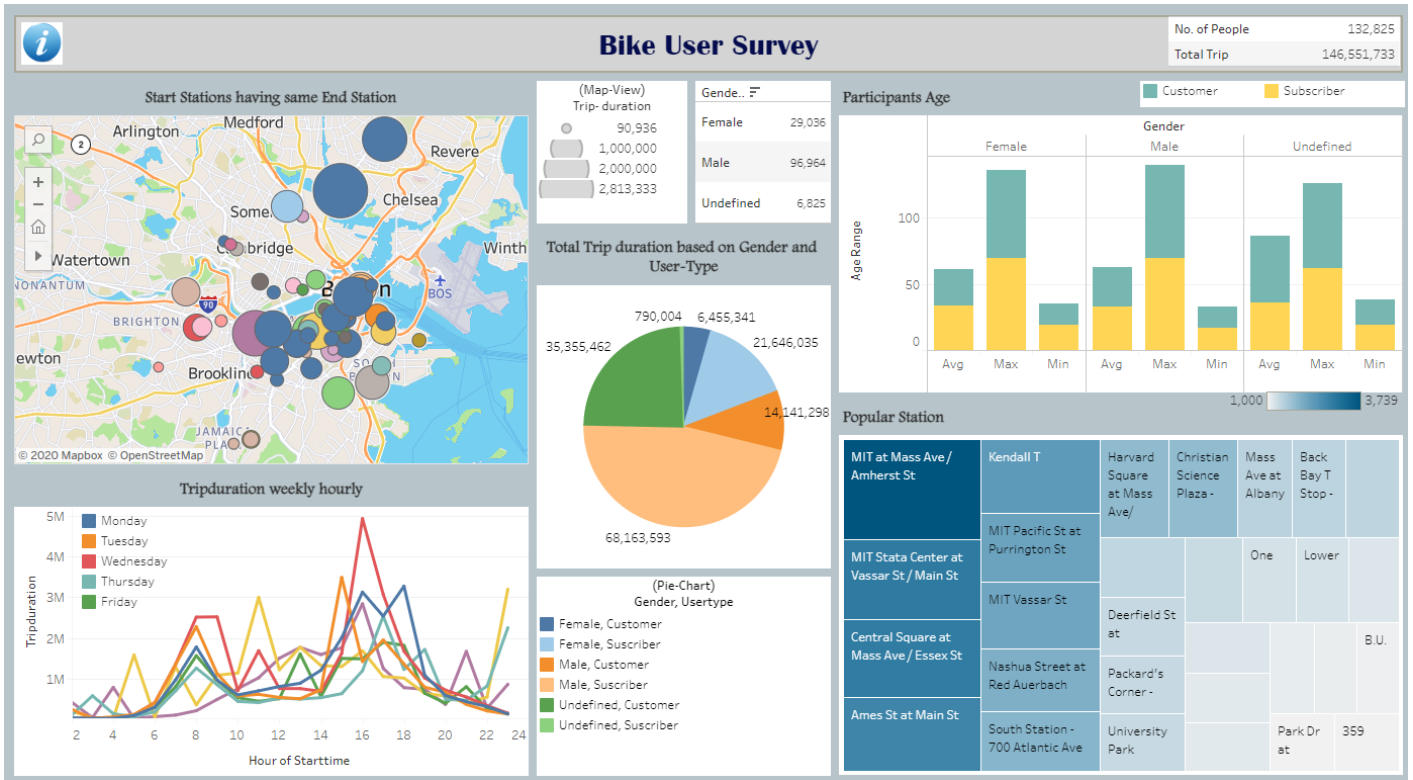Link: https://public.tableau.com/profile/pratichhya.sharma5543#!/vizhome/geoviz_v3/Dashboard1
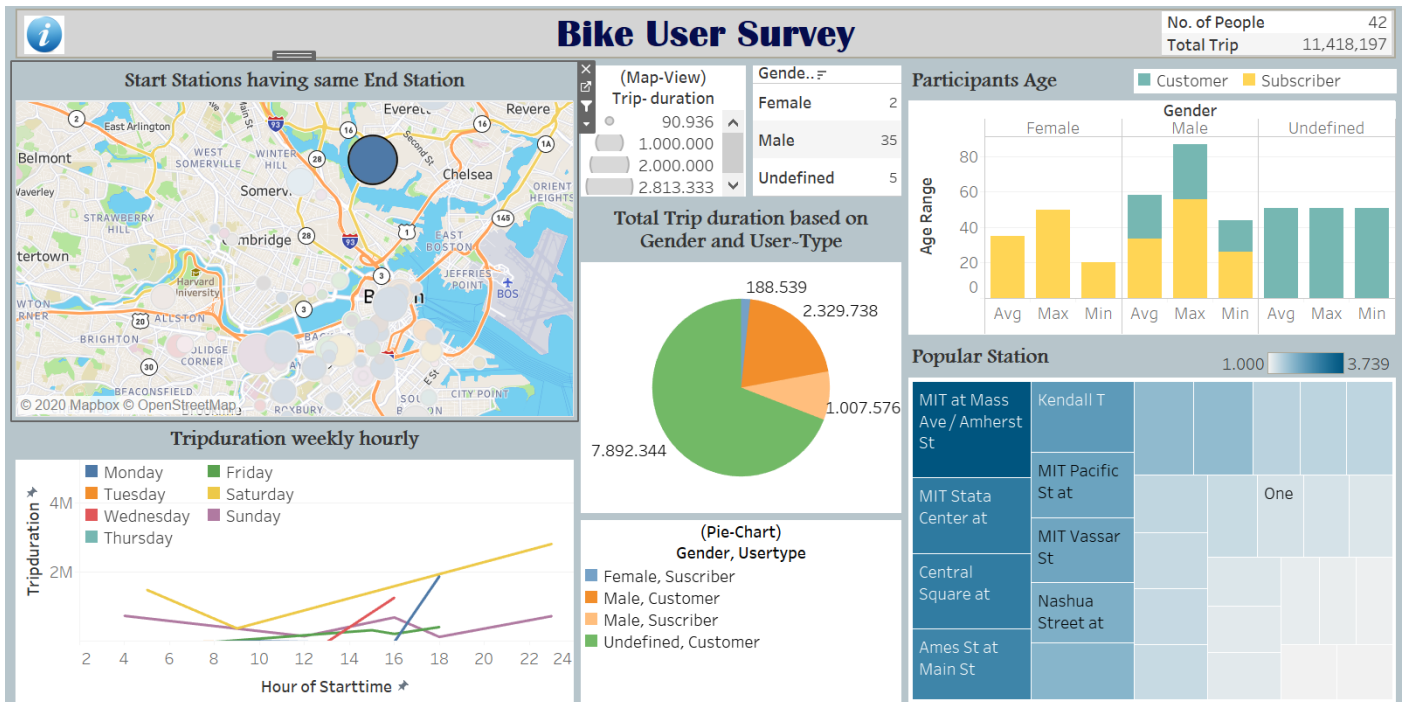
Fig 7: Bike Survey Dashboard

Fig 8: Action Functionality in the Bike Survey Dashboard

Discussion

Tableau was a very interesting and helpful tool for better and quicker visualization of a data but still, there were some challenges faced while using the software. They have been summarized under the following table as the strength and weakness of it:

Strength	Weakness
Faster and Better visualization of Interactive diagrams	Complex platform as many things are available but are mixed up.
Doesnot require advance knowledge on Programming Languages	Maybe not a good solution when there is continously updating data dashboard needed
Easy to backup	Legend Customization Issue when trying to display single sentence by breaking it into lines
Can handle large amount of data	The layout was always disturbed due to my screen resolution.