Reflective Report

P2630030

2022

**SAS Studio – Citi Bike**

**Reflective Report:**

In this assignment, we had to analyse CitiBike, a privately owned bicycle sharing system in New York City. We were given an xlsx file to analyse the data according to the tasks given to us. Looking at citibike data, it was found that there were null entries in the dataset as some of the users were unable to return/dock the bikes to the station.

The task was to delete the rows with null entries. Hence, I used the SQL procedure using the delete from and where statement. A total of 359 rows were deleted from the dataset and a library was created for the temp base to work on. In the excel file, we were given ‘started\_at’ and ’ended\_at’ in DateTime format and we had to format it in another column to analyse the data with ease. In the next procedure, I used the date part and time part functions to define their ‘DATETIME16’ format into ‘mmddyy10’ and ‘TIME8’.

After importing the Citi bike data we deleted the empty or incorrectly entered data from end\_station\_id and end\_station\_name as some bikes were not returned by using the delete function and where statement. After deleting the null entries, we split the date and time and reformatted the date and time by using datepart and time part respectively

To understand the frequency and most common station in comparison to others, I opted for freqplot to define the most common start station ID. The diagram below defines the most common station, JC005, which was frequently opted start station by the customers. The second diagram defines the most most frequent opted end station by the customers

Chart, bar chart

Description automatically generated

Figure 1/Start Station ID

Chart, bar chart

Description automatically generated

Figure 2/End Station ID

To understand the usage of customers by their membership. I decided to use the bar graph here because it is clarifying deep details better than tables and estimates key values at a glance (Geographyfieldwork.com, 2022).

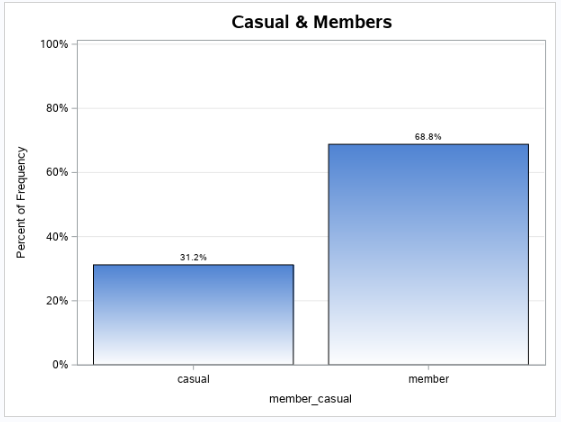


Figure 3/Casual & Members Customer

‌I was able to get the duration of the bike by calculating the time customer spends on their journey and during the analysis, I found out that the most common time spent on the bike was 5 min.

After getting the duration of the most common journey I analyzed to find the start station name for both members and casual customers. After analysis, I found out that the most popular starting station for the member is Grove ST Path and for the casual customer is Hoboken Terminal. For this analysis, I have used a histogram because of its simplicity, versatility, and its insightful look etc. (Six Sigma Daily, 2020).

After completing that analysis, I did a further analysis of members' return times during the evening, morning, and afternoon. During the analysis, I found out that most of the bikes were returned in the afternoon and the least of them were returned during the evening. For this analysis, I used the pie chart because it is simpler than other graphs and can summarize large data (Geographyfieldwork.com, 2022).

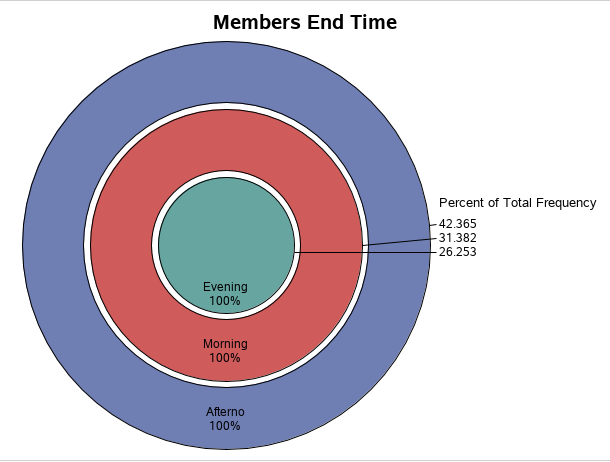


Figure 4/Members End time

**Dashboard:**

I began by comparing member and casual customers in the dashboard. To understand business ideology, it’s critical for any company to understand ground information about its customers. My dashboard shows the most popular start station by ID compared to other Stations, making it a key point for their revenue. Following that, to understand customers and meet their requirements and come up with a better decision, I defined the most common duration. The most popular stations and organisations among casual and member customers are then displayed.

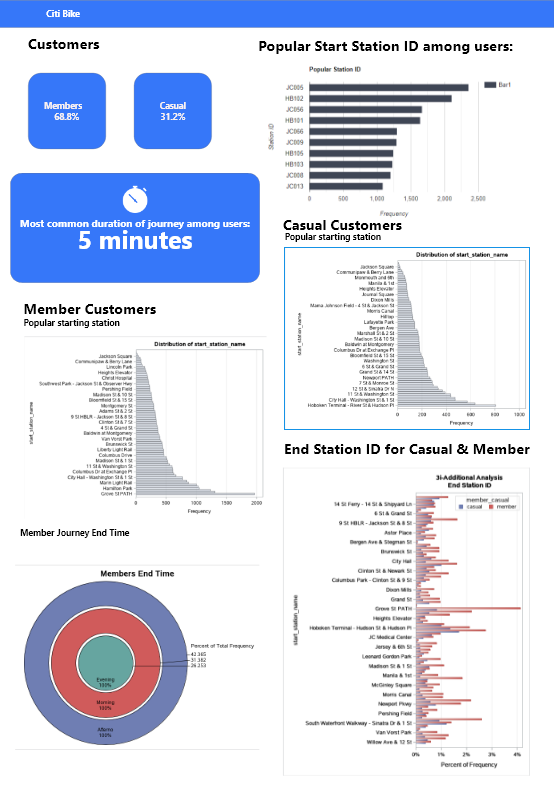


Figure 5/CitiBike Dashboard

## SAS Studio Code:

Figure 6/ Initial Query



Figure 7/Main query



Figure 8/Main query

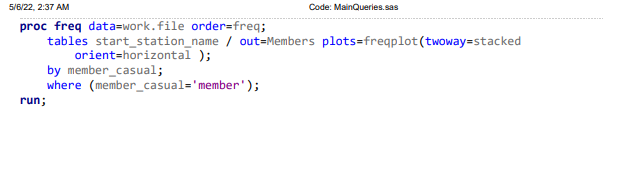


Figure 9/Main Query



Figure 10/Final Query

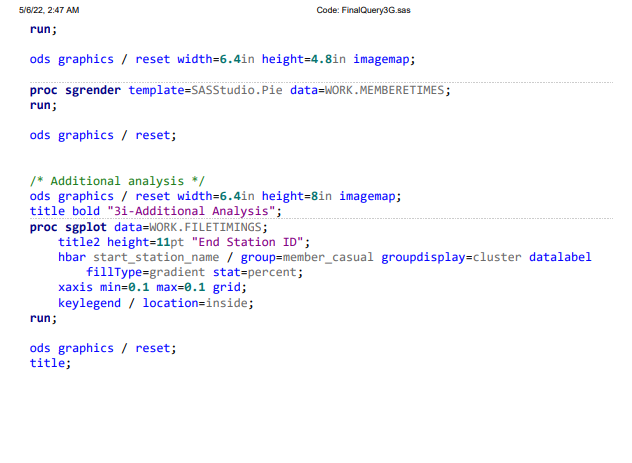


Figure 11/Final Query

## SAS Studio Graph

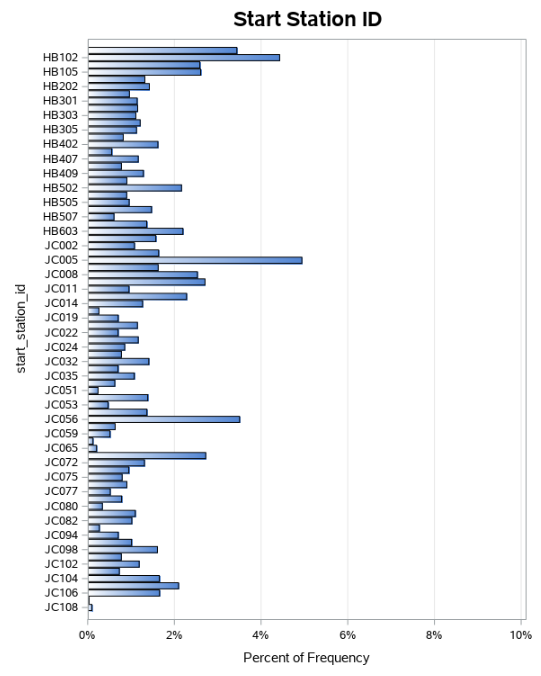


Figure 12/Start Station ID

Chart, bar chart

Description automatically generated

Figure 13/End Station ID

Chart

Description automatically generated

Figure 14/Casual & Members

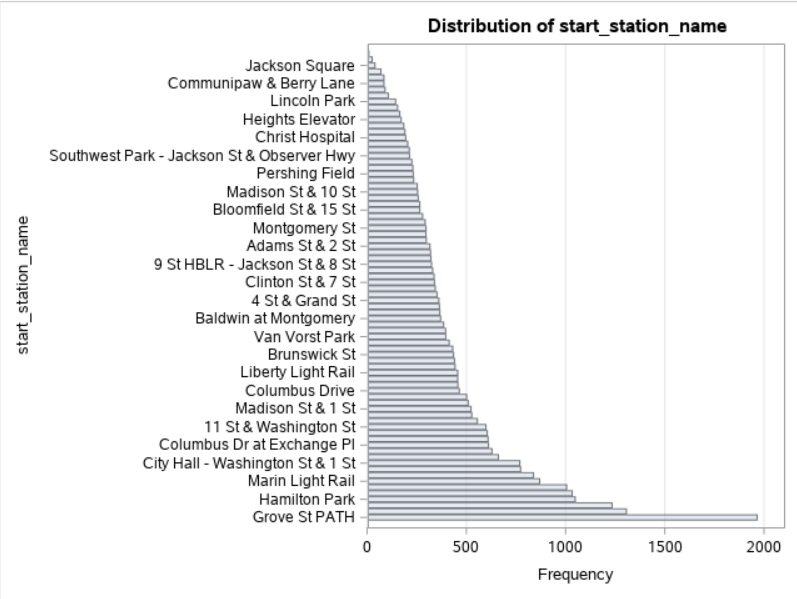


Figure 15/ Members Start Station Name

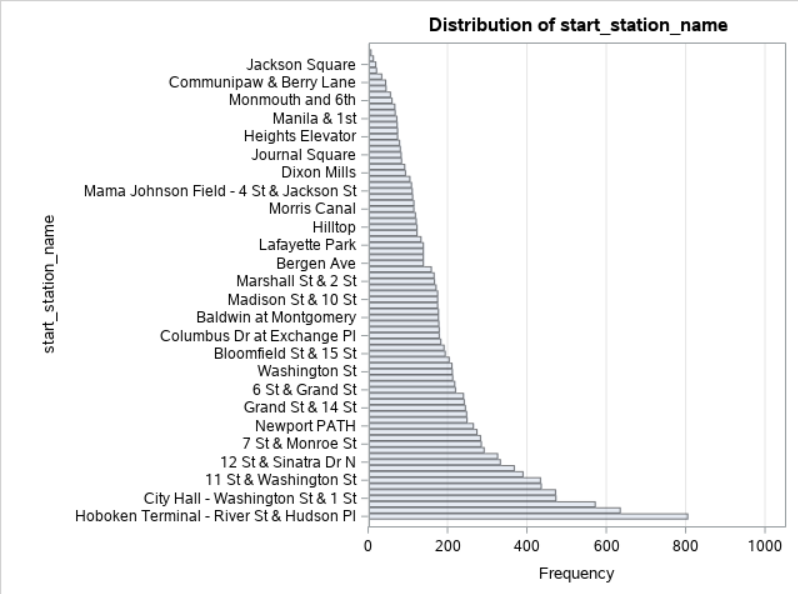


Figure 16/Casual Start Station name

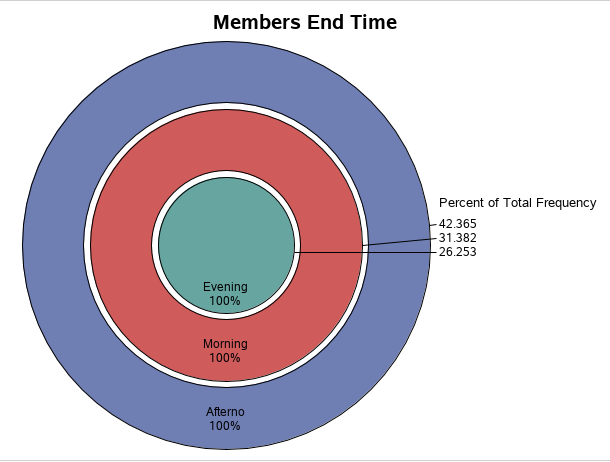


Figure 17/Member End time

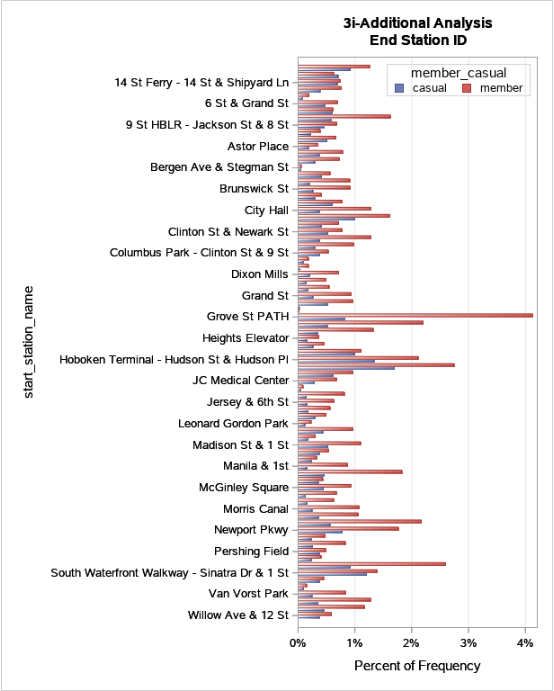


Figure 18/ Members & Casual Users End Station

## References

DataFlair. (2018). *Advantages of SAS | Disadvantages of SAS Programming - DataFlair*. [online] Available at: https://data-flair.training/blogs/disadvantages-and-advantages-of-sas/ [Accessed 6 May 2022].

Geographyfieldwork.com. (2022). *Data Presentation: Bar Charts*. [online] Available at: https://geographyfieldwork.com/DataPresentationBarCharts.htm [Accessed 6 May 2022].

Six Sigma Daily. (2020). *Six Sigma Daily*. [online] Available at: https://www.sixsigmadaily.com/six-sigma-tools-histogram/#:~:text=The%20main%20advantages%20of%20a,pricing%20plans%20and%20marketing%20campaigns. [Accessed 6 May 2022].

https://www.linkedin.com/learning/ (2022). *IMAT2214 - Designing a Dashboard*. [online] Available at: https://www.linkedin.com/learning/interaction-design-dashboards-and-visualization?contextUrn=urn%3Ali%3AlearningCollection%3A6782598689094438912&u=68130314 [Accessed 6 May 2022].

‌

‌