Bike Sharing Data Analysis – Demographic and Station Analysis

Part 1: Data Preparation

The data cleaning process of two datasets from the year 2017 for Q1 and Q2, from the Divvy bike sharing program had several steps which improved the integrity of data for analysis. The first step was to read and concatenate the two datasets 'Divvy_Trips_2017_Q1' and 'Divvy_Trips_2017_Q2', which resulted in a new data frame 'df3' followed by dealing with missing and duplicate data. The 'gender' column contained a lot of missing values which were filled with 'Prefer not to disclose'. The 'birthyear' column was dropped, as it was deemed irrelevant beyond age verification. When checked for duplicate entries in df3, eleven entries were found and dropped. The data type of 'start_time' and 'end_time' was changed to datetime. The cleaned dataset was then saved and named as "trips_2017_Q1Q2".

Similarly, the dataset of stations for Q1 and Q2 was stored in data frame df4 and was examined for duplicate entries and null values. The only changes that df4 encountered were changing the datatype of 'online_date' column from object to datetime and a minor change in the city label 'Chicago' by removing extra space. At last, we merged df3 and df4 using an inner join which resulted in creating df5 and was then saved as "trips_and_stations_2017_Q1Q2.csv". This consolidated dataset will result in easier analysis of data

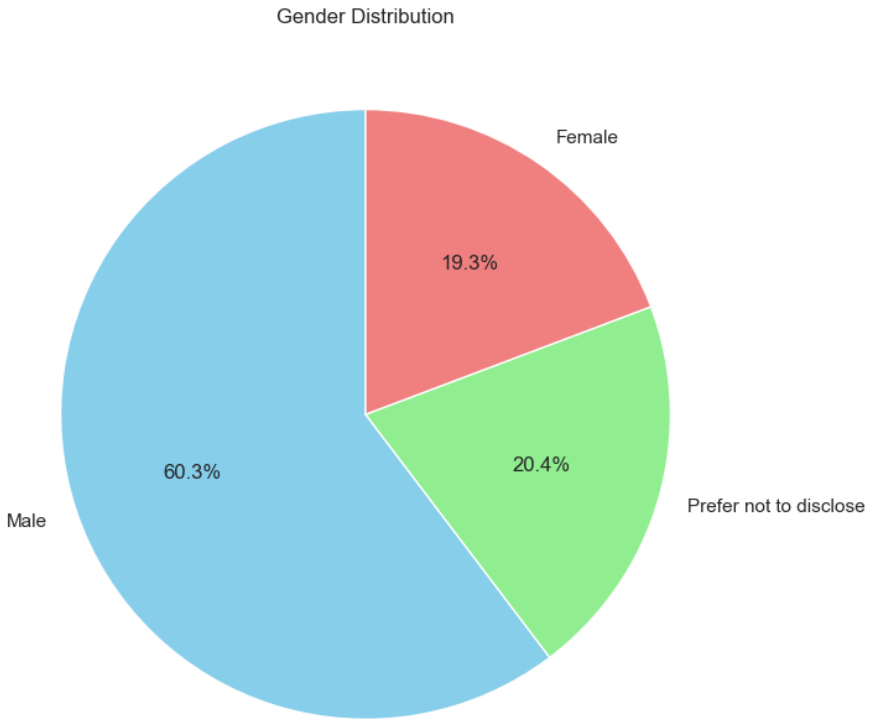
Part 2: Data Exploration and Analysis

This report sheds light on a comprehensive analysis of bike-sharing data, which is focused towards gender distribution, user types, ride patterns, and station usage. It presents a wide range of information, with valuable insights on user behaviour and preferences.

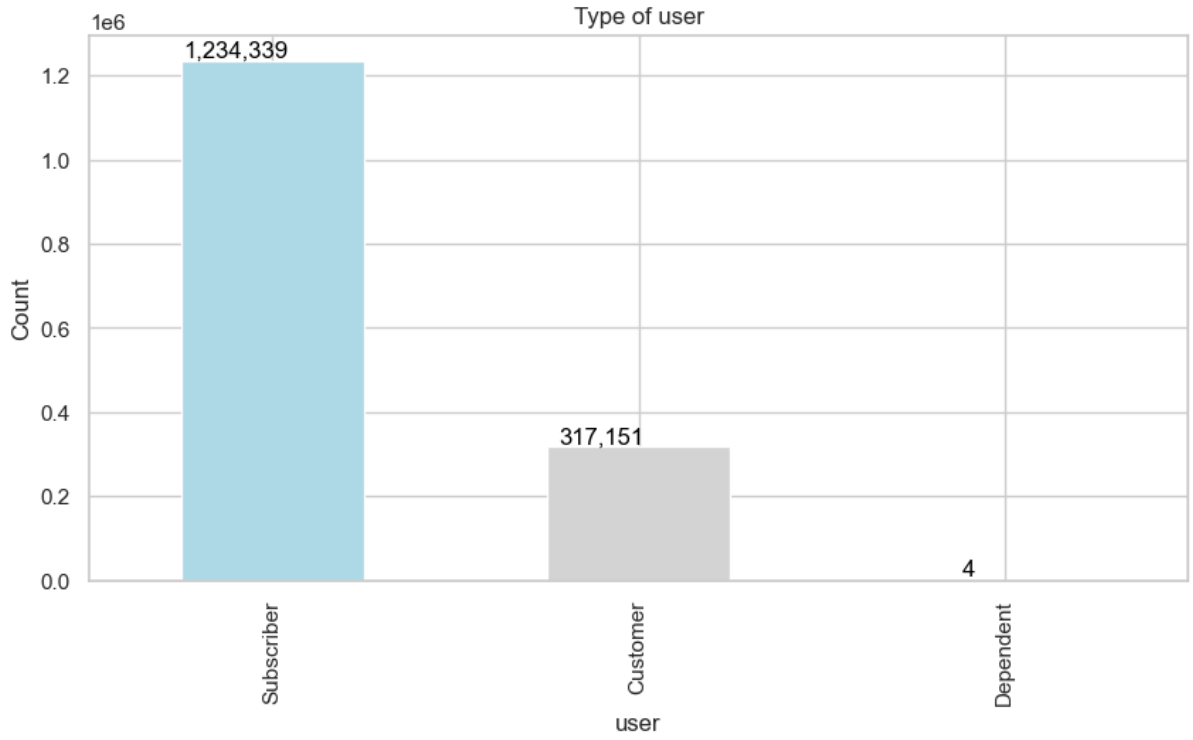
Ride Analysis

Gender Distribution

There's a significant gap among number of male and female bike riders with 935,854 males, whereas only 298,784 female riders. Moreover, we do not know about the gender of customers, thus it has been filled by "Prefer not to disclose". While comparing total number of rides taken by male and female, it was discovered that men are 3.13x times more likely to ride bike than females. Though data for customer user type is not given, but I believe, if it could be recovered somehow, I expect distribution would be roughly the same.

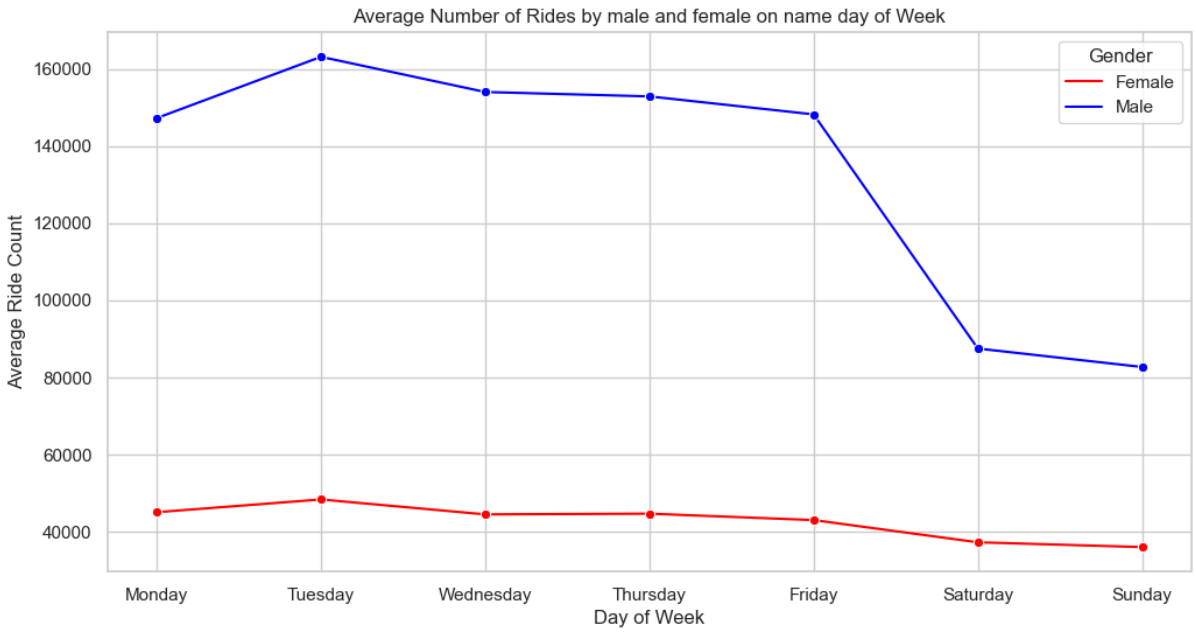


Usertype Analysis



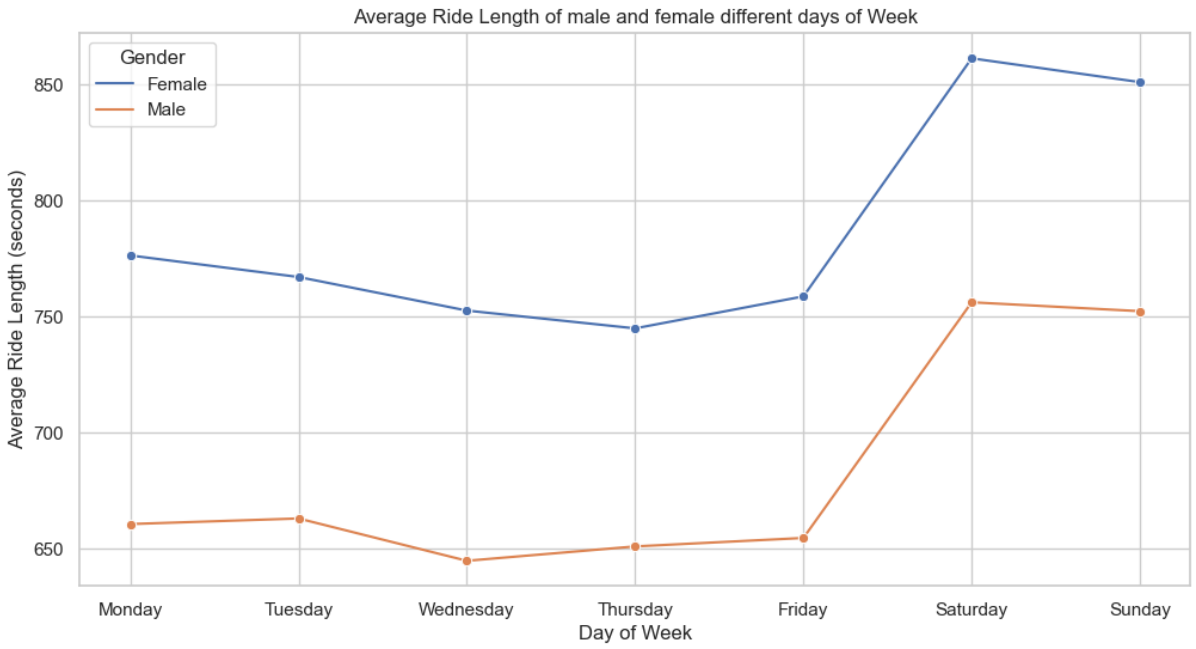
Subscribers tend to be the major user followed by customers. I believe, the company should focus on converting their customers into subscribers, which will eventually lead to revenue growth.

Q1. What is the average number of rides taken by male and females on a daily basis?



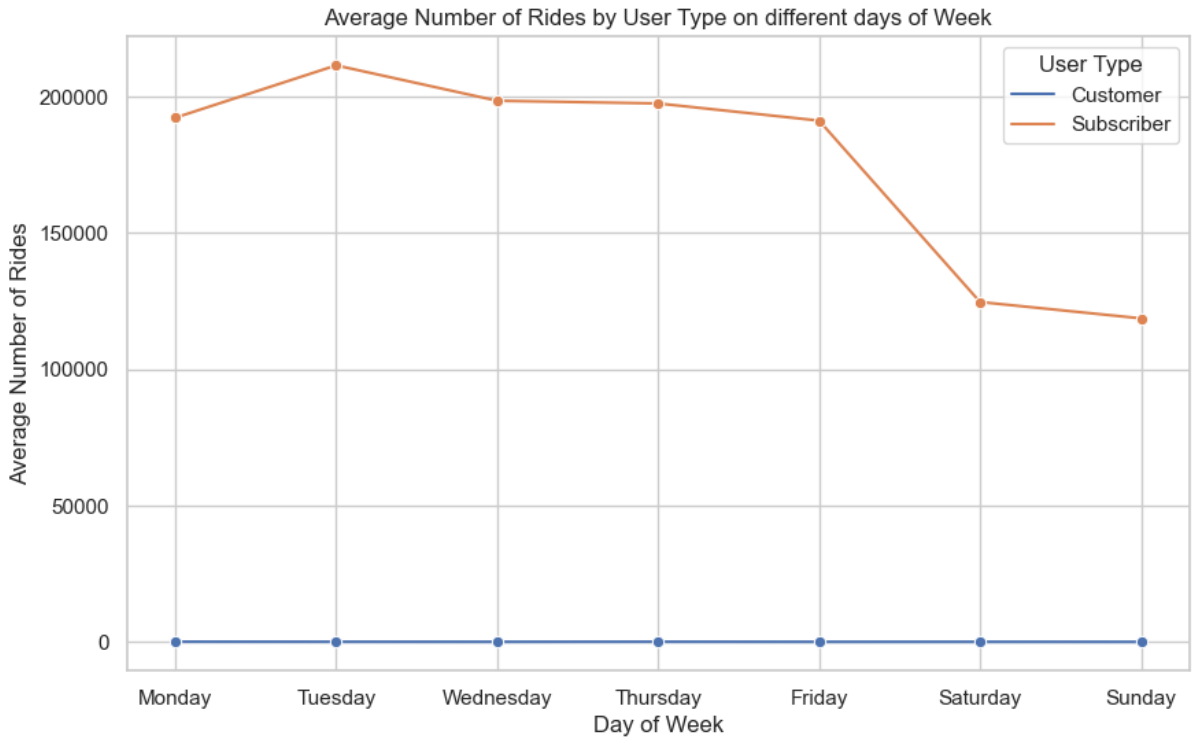
There is a wide gap in between number of rides taken by a male and female on weekday. Males have a good intensity of number of rides taken on Monday and peaking on Tuesday, while it dropped massively on Saturday and Sunday especially for males and it could be a potential area for promotion to increase number of rides on weekends. Females have very a smaller number of rides on weekday as well as weekend.

Q2. What is the weekly average ride length of a male or female?



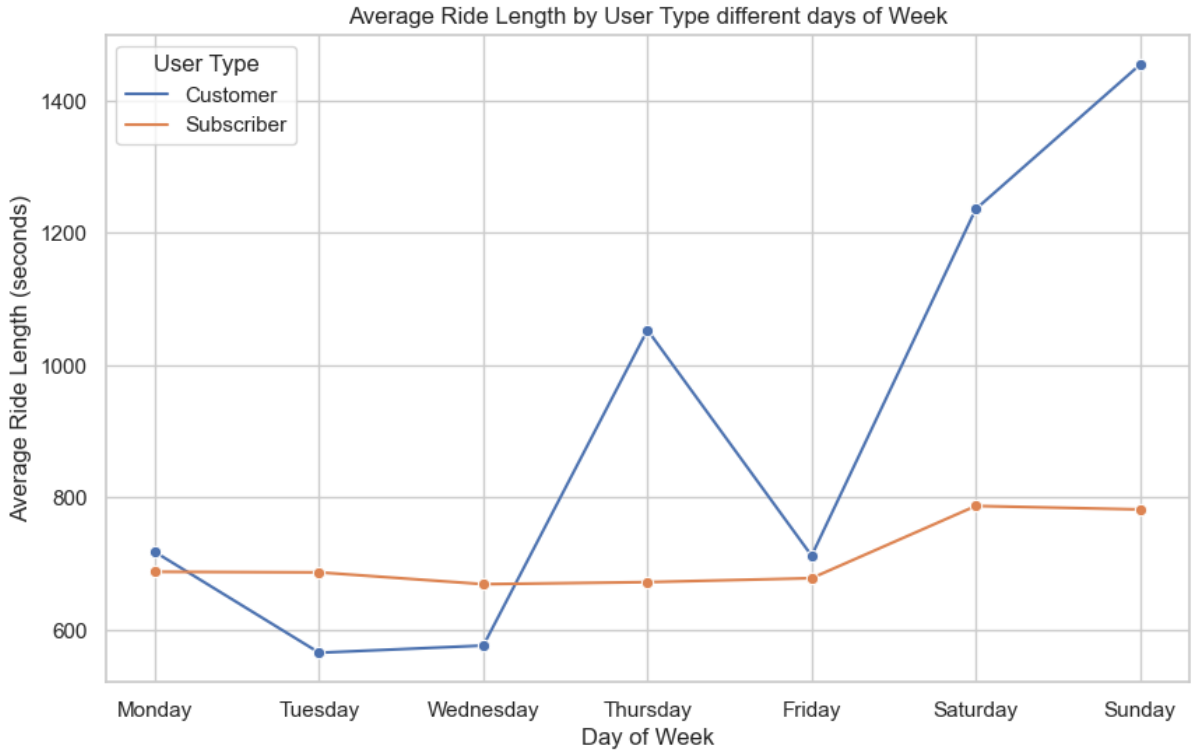
Weekly average ride length of female surpasses male. Males have shorter ride duration in terms of females. The trend remains same on weekdays as well as weekends. Though there's a slight change which should be noticed that average ride length for a male as well as female has increased massively over the weekend. It could be due to plenty of leisure time available during the weekend.

Q3. What is the average number of rides taken by subscriber and customer on different days of week?



On an average, subscribers have a greater number of rides on Tuesday and Wednesday. Whereas, customers have maintained a low ride count compared to subscriber. Though customers have a consistent pattern to ride bike but subscribers are more inclined to ride bike on weekdays.

Q4. What is the average ride length of a customer and subscriber on each name day of week?

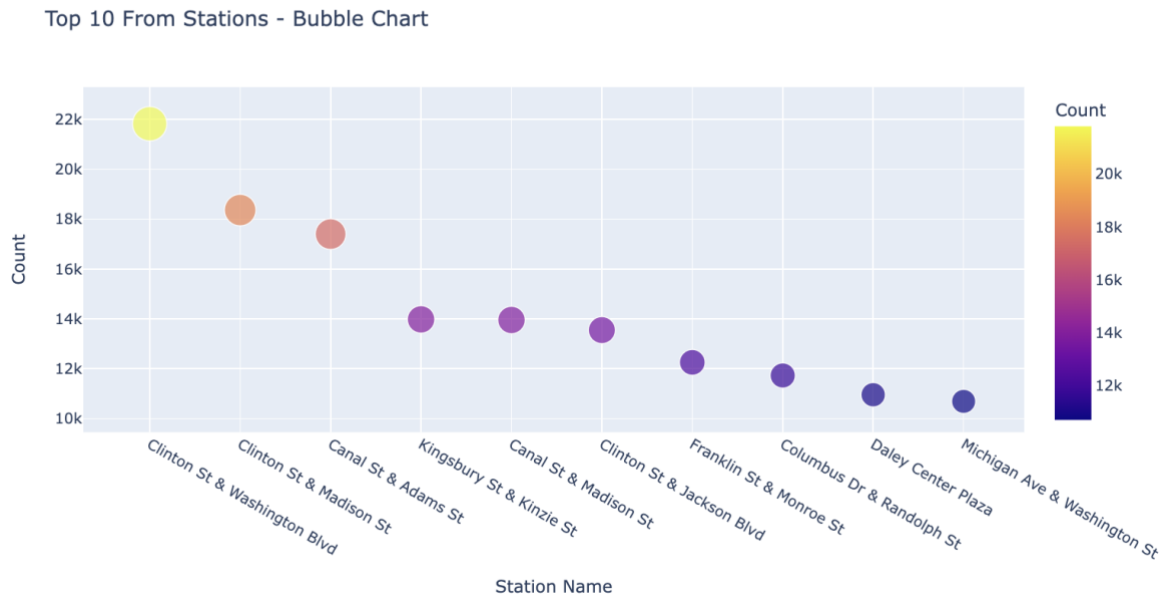


Customers have a longer ride duration on weekends especially on Sunday, whereas subscribers have short and consistent duration over the week. This shows that customers might be using these bikes for leisure, whereas subscribers opt for routine and shorter trips. The peak for average ride length of a customer was on Sunday whereas, average ride length of a subscriber was at its peak on Saturday.

Station Analysis

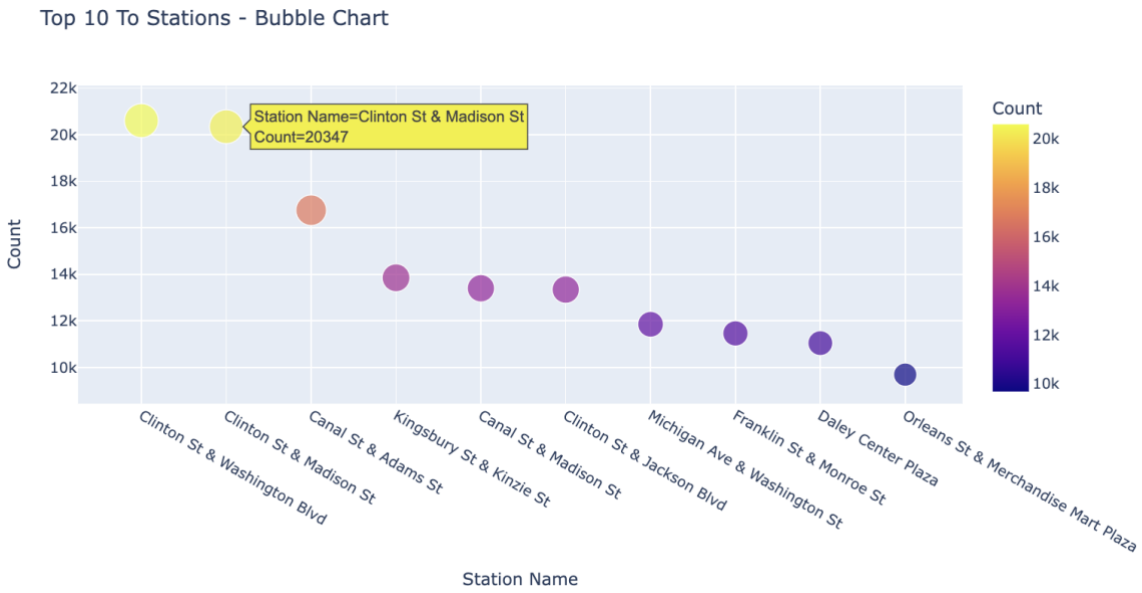
In total there are 582 unique stations where riders can start and end their rides. Every station can store bikes.

Q1. What are the top 10 stations where riders start their ride?



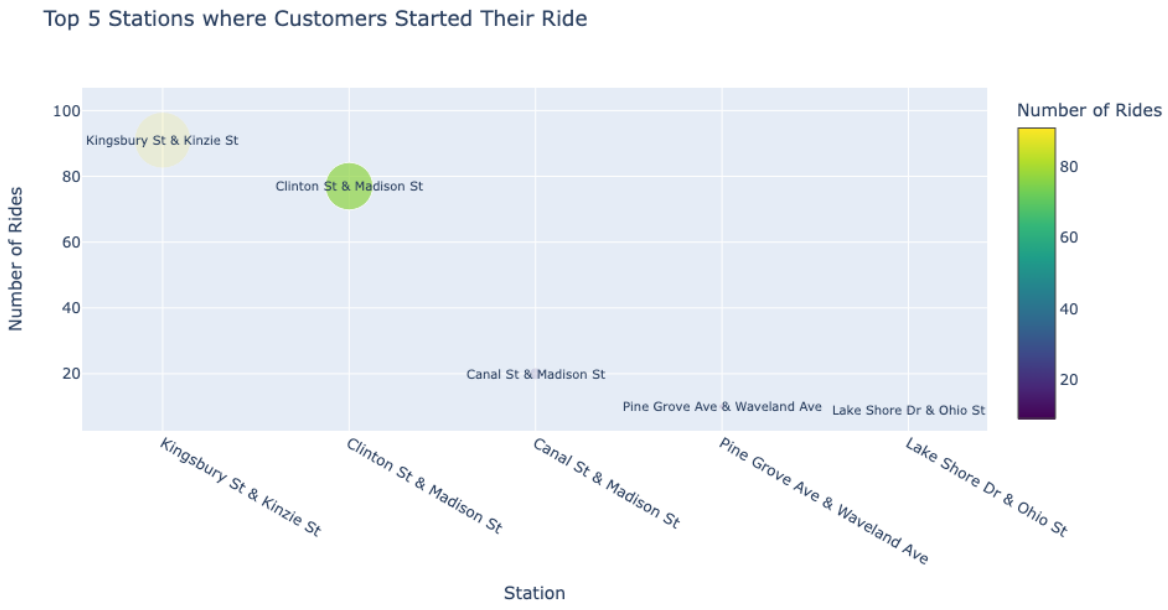
The top 10 stations account for 11.72% of all the traffic start points on our bike network. These top stations are located in and around the heart of Chicago to the heart of the golden west.

Q2. Which are the top 10 stations where riders terminate their ride?

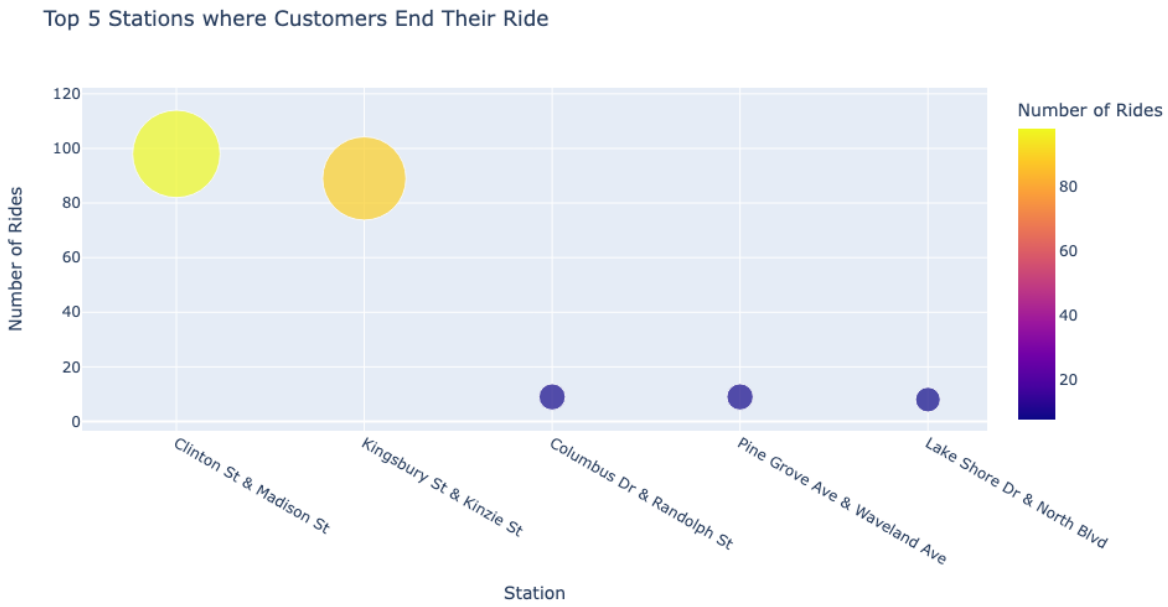


Similarly, the top ten stations where riders end their ride are located in and around the central areas. The top ten end stations account for ~ 11.52% of all of the traffic end points on our bike network.

Q3. What are the top 5 stations where customers start their ride?

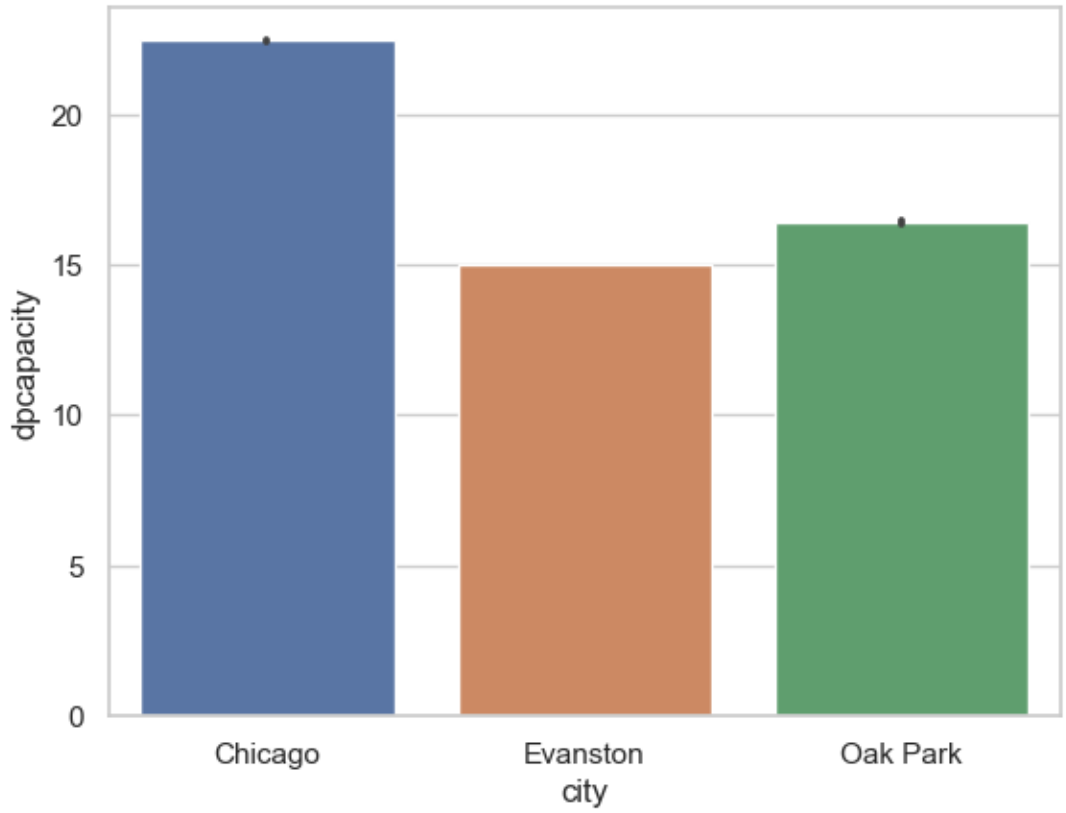


Q4. What are the top 5 stations where customers end their ride?



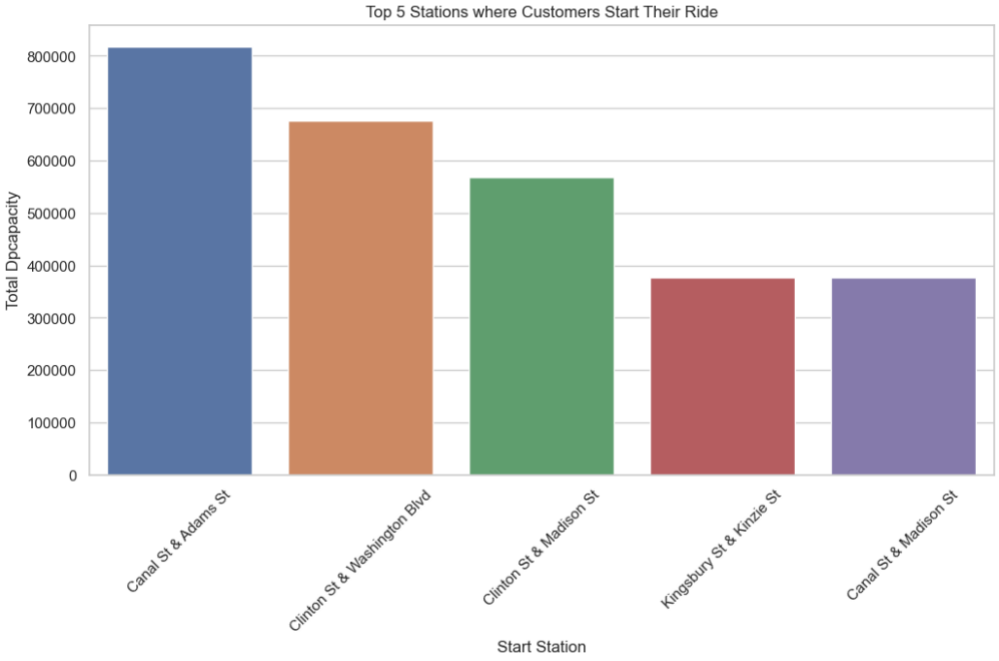
Kingsbury St & Kinzie St and Clinton St & Madison St being the most popular station among customers to start and end their ride. The company should put out survey forms or feedback around the area to know what they can do to convert these customers into subscribers.

Q5. What is the city wise dpcapacity?



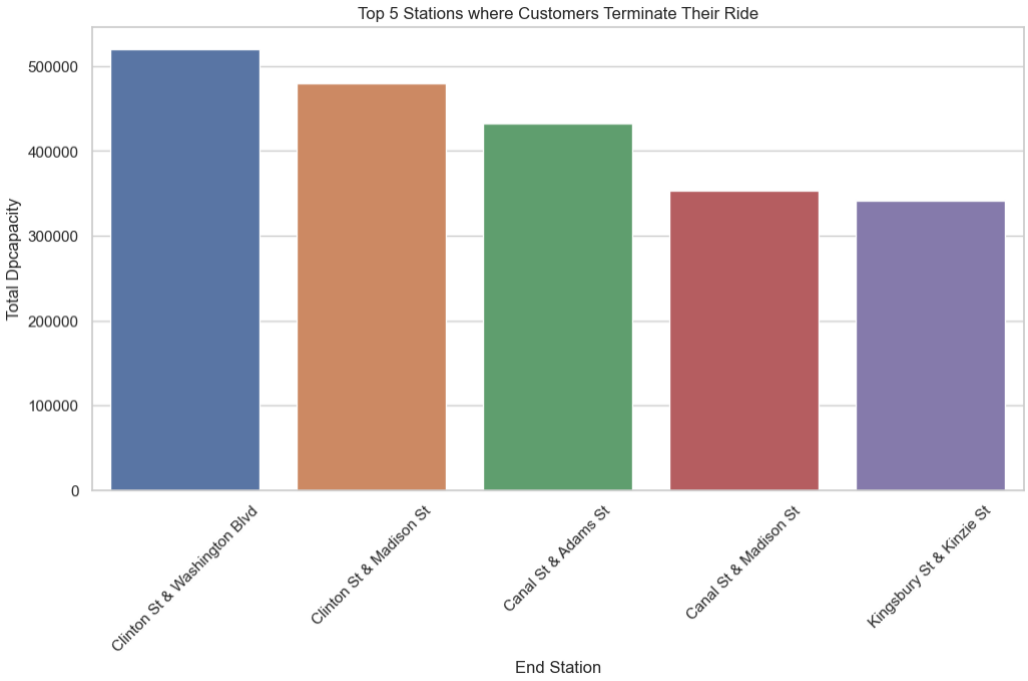
Chicago being at the top in terms of dpcapacity, followed by Oak Park and Evanston City.

Q6. What is the dpcapacity of the top 5 stations where customers start their ride?



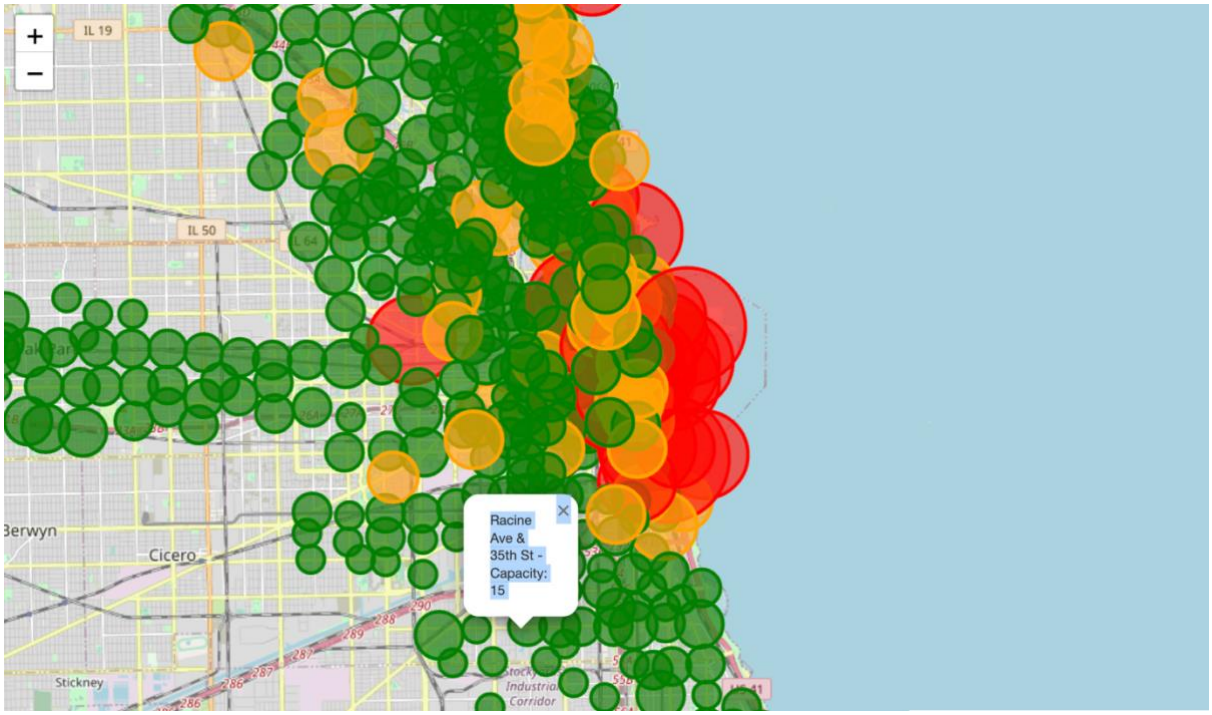
These stations have enough capacity to provide vehicles if a new customer comes in. These stations are also some of the top stations from where members start their ride.

Q7. What is the dpcapacity of the top 5 stations where customers end their ride?



Canal St & Adams St is quite for customers to end their ride followed by Clinton St & Washington Blvd. This information is essential for the company to meet demand effectively at these popular stations.

Q8. What is the station wise dpcapacity?



This is an interactive map generated using the folium library in python Each circle marker tells about the station and its properties. As can be seen station marked red are held with highest dpcapacity greater than 30, stations reflecting orange have a dpcapacity between 20 and 30 and green less than 20. With help of AI and other technologies this data can be tracked on real time basis, which will eventually help riders figure out their nearest stations with parking capacity.

Suggestions

1. I would suggest company should focus on increasing number of female riders as it will help in getting more subscribers and customers as the women riders have comparatively more ride length or time compared to men, so it will eventually increase the revenue of firm.
2. I believe they should start organizing events focused to gather female riders and educate them about the safety flexibility and convenience of using Divvy bikes.
3. Can offer gender specific discounts or promotions for bringing in signups
4. I believe the company should create marketing strategies focused towards customer using top start stations to convert customers into subscribers
5. The overall rider strength is from Chicago. I believe, Divvy bike company should expand around Oak Park and then Evanston city.

6. It is evident that weekends have high ride time, company should create a campaign around tourist destination during leisure time, which in turn will help in marketing of the brand as well.
7. Should employ their sales staff at few of their popular stations with a tablet to make it easier for customers to sign up on the spot and guide them about the benefits of being a subscriber such as cost saving or exclusive perks.

Conclusions

This data analysis report provides insights into the bike sharing services, addressing gender disparities, user type, and station activities. The gender gap in ridership with higher number of males compared to females suggests company to create strategies and focus on participation of female riders. Companies can organize events focused on female, or promotional offer for females, and could guide them about convenience and feasibility, which could attract a greater number of female riders, which will lead to increase in revenue.

Additionally, strategies focused towards popular start stations and promotions in locations such as Oak Park and Evanston to attract more customers and subscribers' conversion. Marketing initiatives for weekend tours around tourist hotspots can lead to increase in brand exposure. I believe following the suggestions cited above company can bring in more customer and convert customers into subscribers. Following these practices will lead to increase in revenue.

References

1. https://chicagology.com/chicagostreets/madisonstreet/#:~:text=Madison%20street%20runs%20straight%20from,is%20the%20Golden%20Gate%20street.
2. https://towardsdatascience.com/creating-a-simple-map-with-folium-and-python-4c083abfff94
3. https://plotly.com/python/bubble-charts/
4. https://python-graph-gallery.com/line-chart/
5. https://matplotlib.org/