“Cyclistic” Case Study – Google Data Analytics Capstone

This is the report that deals with the findings on the data analytics done on the “Cyclistic” data set as part of Google Data Analytics course Capstone. The report is broken down into 6 parts which discusses different phases of the data analytics involved in coming up with the findings of the case study.

The data used was the cyclistic data case study problem statement. I used data from Nov, 2020 to Oct, 2021 (12 months) to so my data analysis.

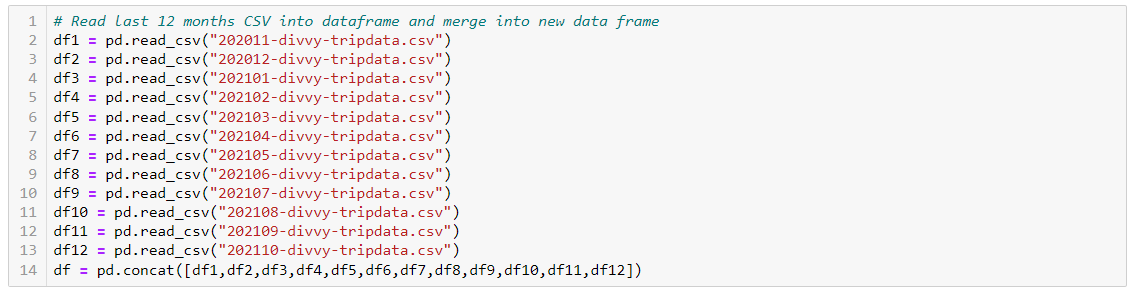
Ask

We would want to answer these three questions through our data analysis

* How do annual members and casual riders use Cyclistic bikes diﬀerently?
* Why would casual riders buy Cyclistic annual memberships?
* How can Cyclistic use digital media to inﬂuence casual riders to become members?

Prepare

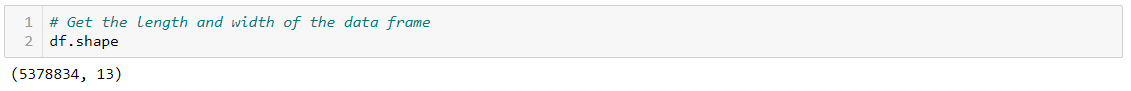
As previously mentioned, data from Nov, 2020 to Oct, 2021 we taken and data analysis was performed on that. A large data frame is created with all the data of 12 months.



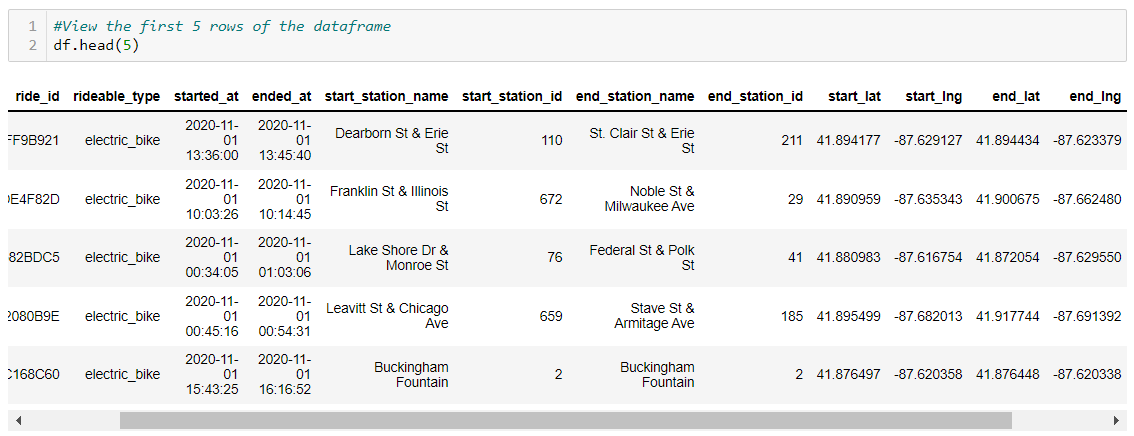
Process

In this stage we try to understand the data and clean/remove erroneous and missing data.

We first check the size of the data



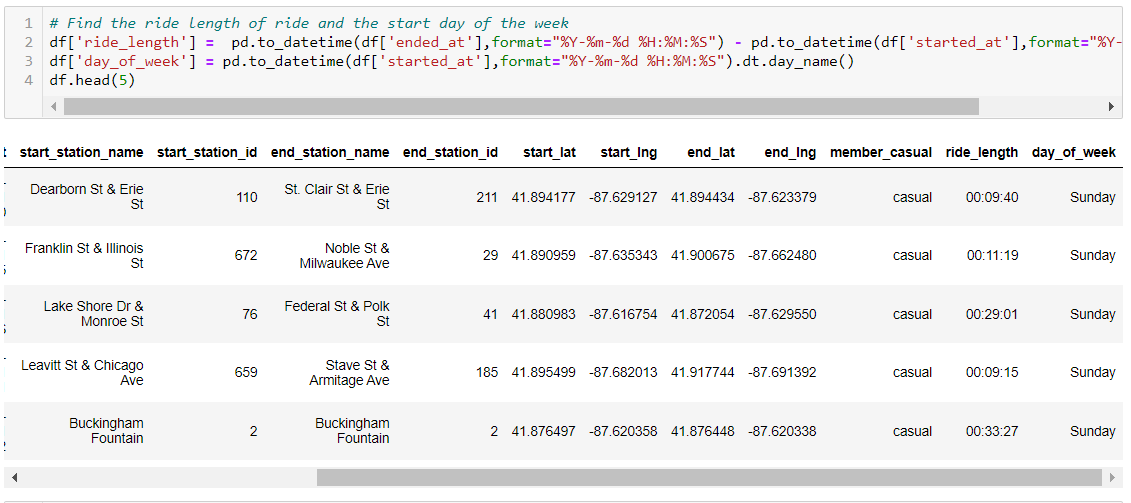
Now, we see first few rows of the data



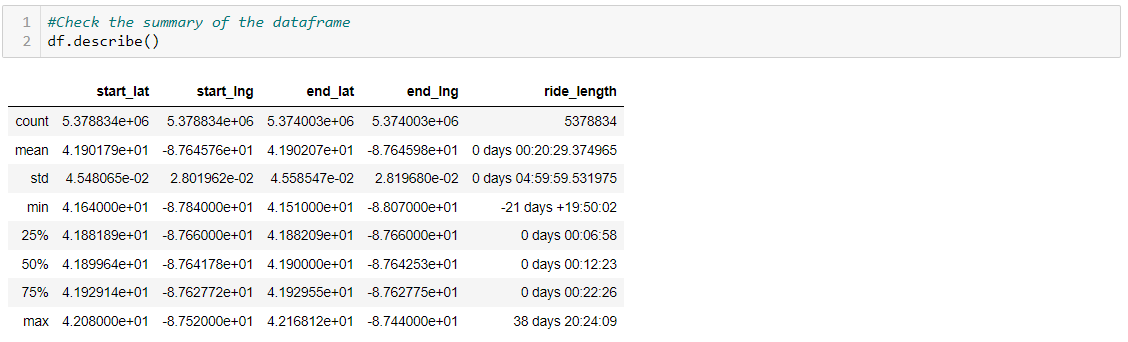
Now we try to remove duplicates, but we see the numbers of rows are still same, which means that data has no duplicates.



Next, we add two new columns ride length and day of the week, which shows the length of the ride for each ride and the day of the week on which that ride was taken.



Now when we find the summary, we see that ride lengths are negative.



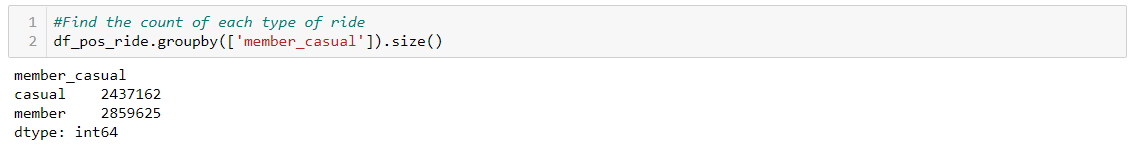
So, we remove, those records,



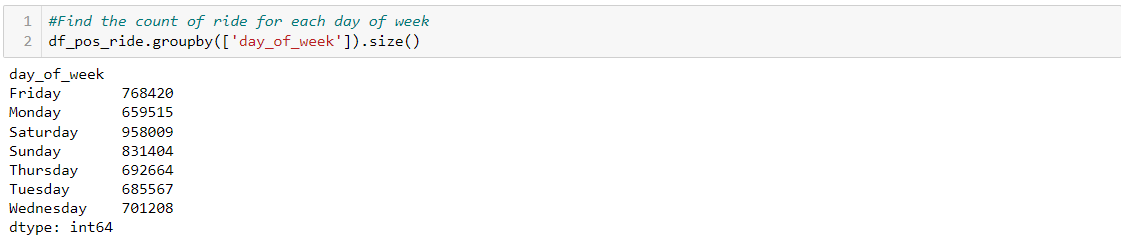
Analyse

In the analysis stage we try to summarise different aspect of data to make observations.

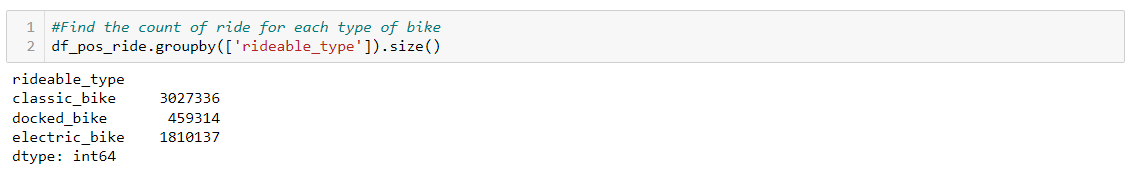
We find our dataset contribution, we see that our data has more data for member than for casual rider, so it might be good idea to get more casual ride data.



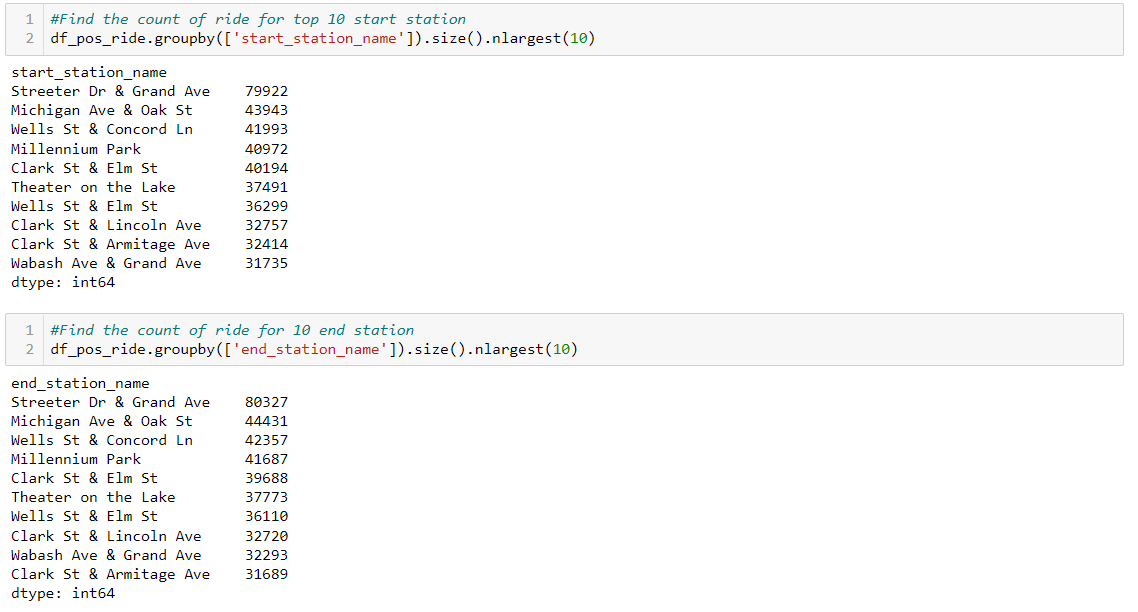
Now we check ride count for each day and see that “Saturday” is the busiest day.



We also check the count to find the type of bike favourite among the riders. It is not clear that classic bike is the most favourite bike or if the count is more since classic bike has more availability.



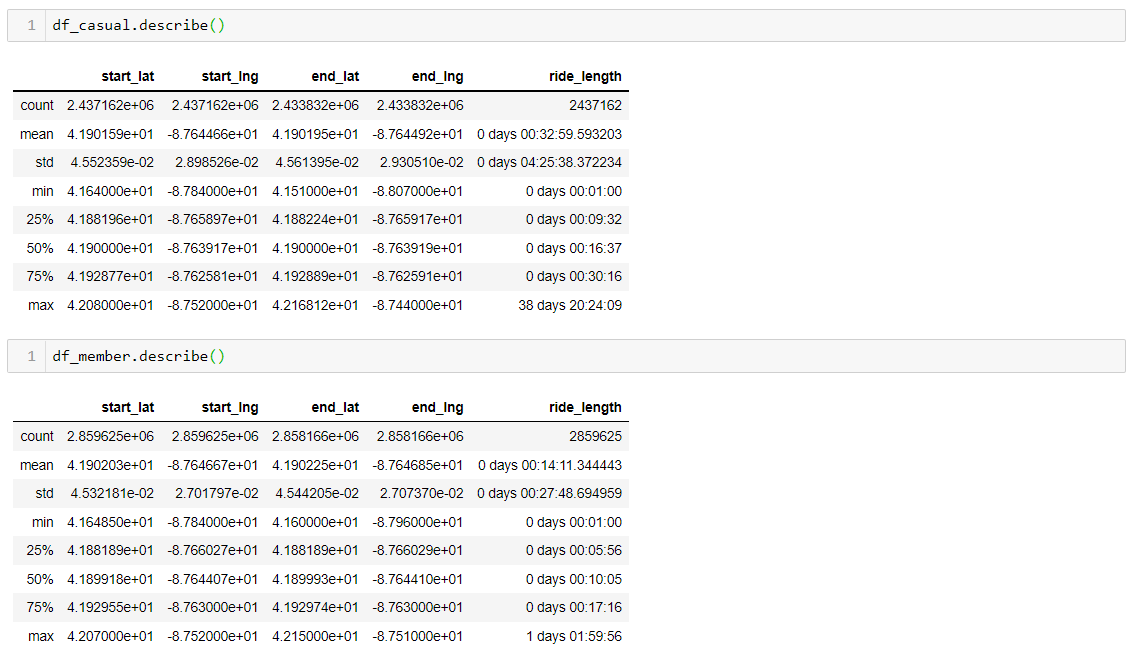
Next, we find the top 10 stations from which riders start and end the trip.



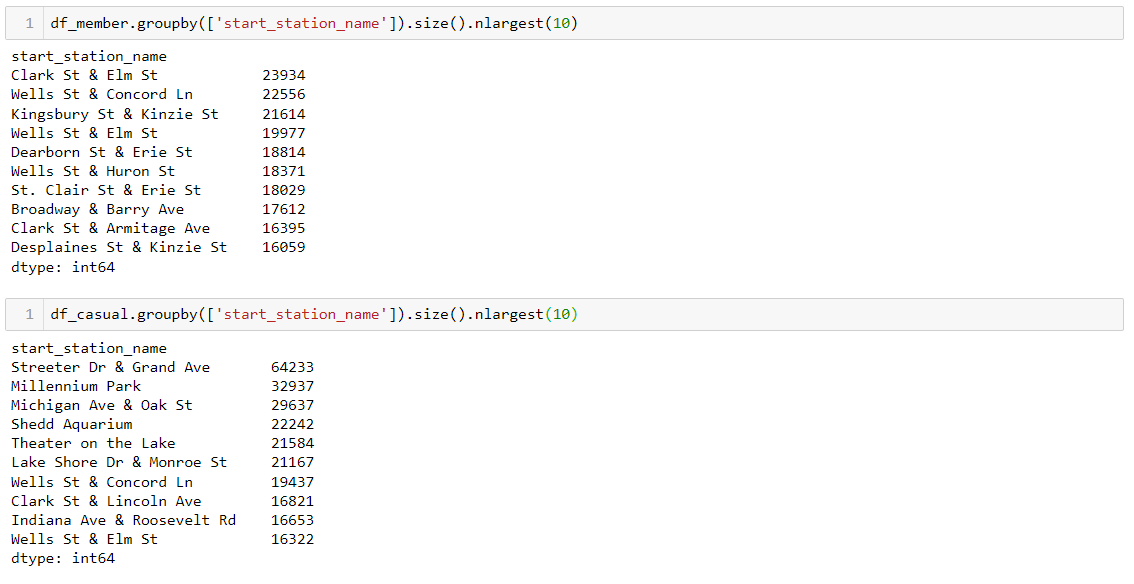
Next, we summarise data based on rider type. We plot two graphs which show the ride count for each type of rider for each day of the week.

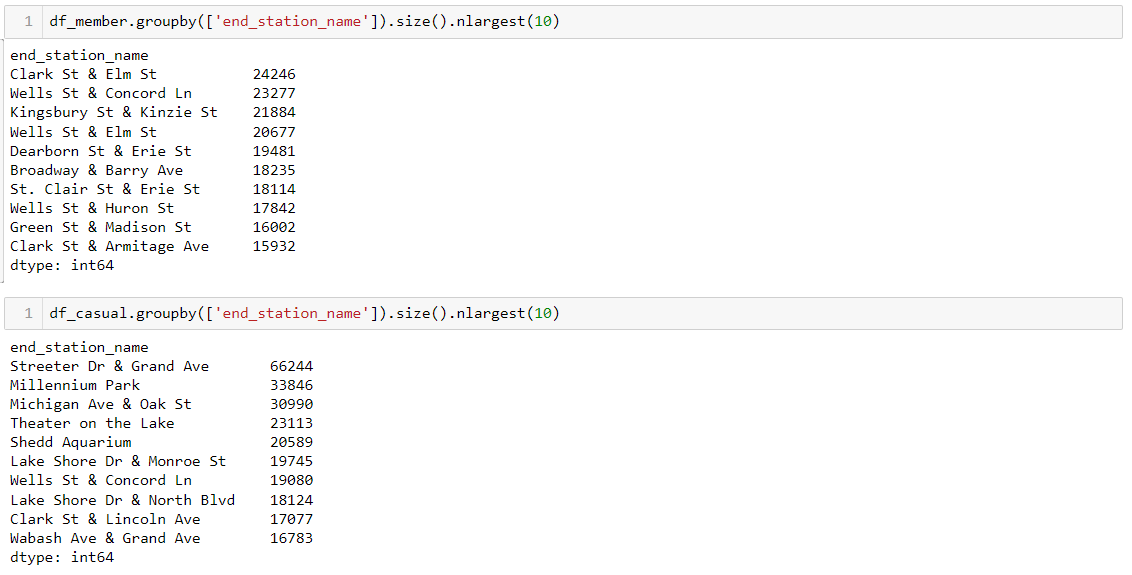


Now, we summarise ride length data for each type of rider.

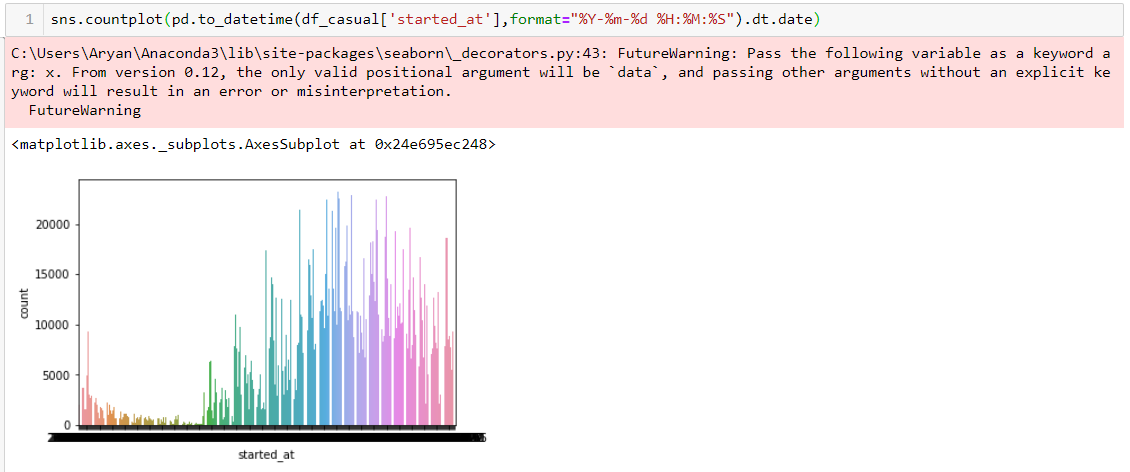


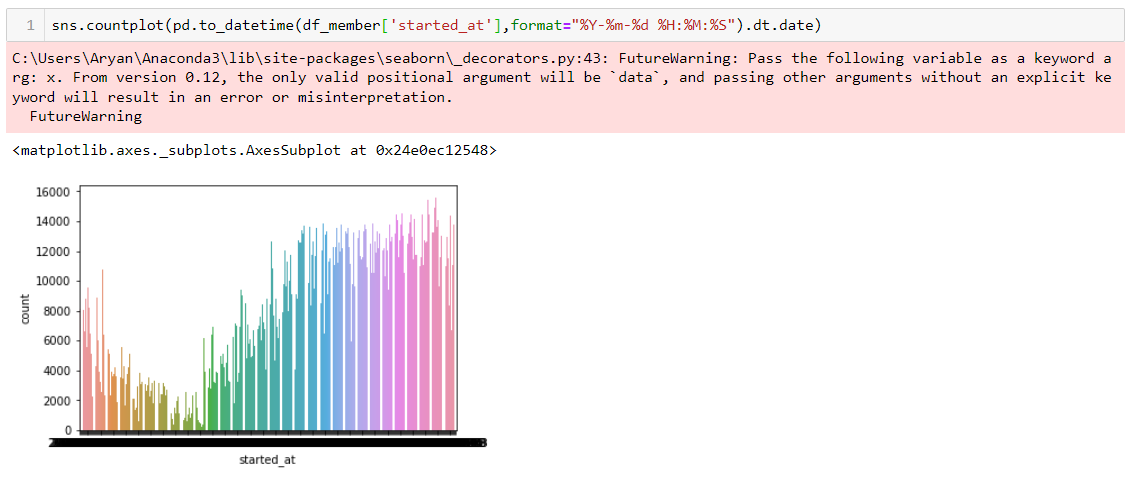
Now we find the top stations where the riders start and end trip.





Next when we check the usage of bike across the year for each rider type and see that the usage for casual riders is more seasonal.





Share

From the above analysis though we can get initial indication towards the question asked above. But to come up with concrete recommendation, more data about the rider would be required.

Following are the observations based on above analysis on understanding how casual riders are different from member riders:

* Member riders ride bike usage is almost same through out the week except on Sunday, while the bike usage for casual rider the usage is more on weekends.
* Casual bikers bike usage is most on Saturday.
* The average duration of ride file casual rider is ~32 minutes and for member rider is ~14 minutes.
* We can see a seasonal usage of bike. This is more evident in case of casual riders than member rides, who avoid riding bike only when the weather is very cold.

Act

Below are my recommendations on how to influence casual riders to become members?

* Since the casual rider’s bike usage is significantly higher on weekends, introduce a weekend membership plan. Later show that annual membership would be cheaper than taking weekend membership.
* We know the top stations from which a casual biker starts or ends a trip. We can put advertisement of annual membership on those location.
* We can give a discount on annual membership during the summer months, so that we have higher probability to getting more riders signed into the membership as it’s the peak season.
* Keep the daily ride fare low for short trips and increase it significantly once the it crosses 30 minutes mark to make the rider buy week membership and slowly entice them to buy annual membership.
* Since most of the member riders seem to be office goers as observed from the data, we can target the casual riders who go to office by other medium, if such data about causal rider is available.