Google Data Analytics Capstone: Cyclistic Case study

Analysis by- Anurag Bhagat

Introduction-

Introduction Welcome to the Cyclistic bike-share analysis case study! In this case study, you work for a fictional company, Cyclistic, along with some key team members. In order to answer the business questions, follow the steps of the dataanalysis process: Ask, Prepare, Process, Analyze, Share, and Act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path

Scenario:-

You are a junior data analyst working on the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Stakeholders :-

1. **Lily Moreno**: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels. ●

2. Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic's mission and business goals—as well as how you, as a junior data analyst, can help Cyclistic achieve them. ●

3. Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

The business task:-

Analyse the Cyclistic data set for the year 2020 to understand how annual members and casual riders use Cyclistic bikes differently.

Deliverables:-

1. A description of all data sources used
2. Documentation of any cleaning or manipulation of data
3. A summary of the analysis
4. Supporting visualisations and key findings
5. Top three to four recommendations based on the analysis

1.A description of all data sources used:-

The data-set which will be used for the analysis has been provided by Cyclistic management. It consists of CSV files, It contains file of 2019-on quarterly basis. I have used tables only from year 2019, for one year analysis . It contain four table , each of one quarter. I have also included two file related to station, one is start station and other is end station which will be used to find the location of the table including their longitude and latitude , which further be used for the distance coverd calculation.

*This is public data that has been made available by Motivate International Inc. under this license.

2- Cleaning and manipulating the data:-

Data Transformation Steps:-

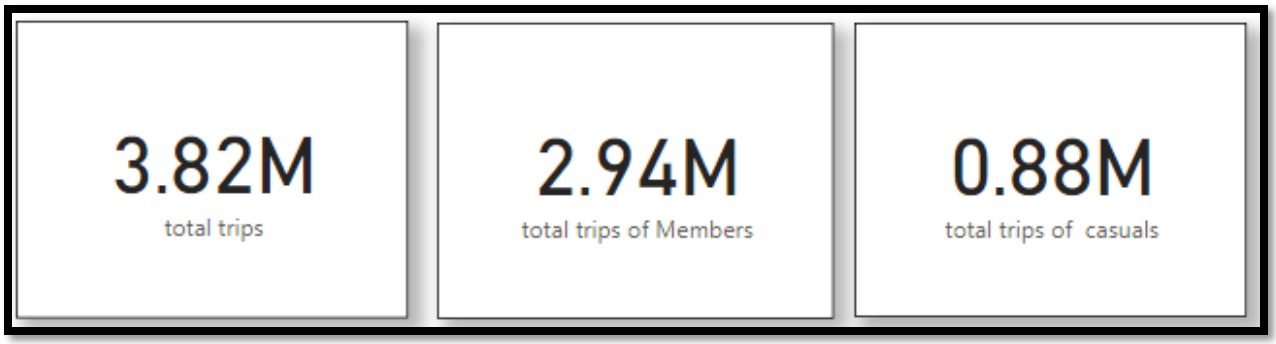
1. **Data Cleaning—** **Handling Missing Values:** Replaced missing values in the Birthday Year column with "null". Replaced missing values in the Gender column with "null".
2. **Feature Engineering:**-- **New Columns:-**Created a new column called 'ride_length' to calculate the duration of each trip. Created a new column called 'day_of_week' to indicate the day of the week for each trip.also created a column "distance" to calculate total distance traveled in each tirp.
3. **Data Validation:**--Ensured that missing values were properly handled to maintain data integrity,Verified the accuracy of the calculated trip duration values.
4. **Standardizing the table:-** all the tables had differently column name ,I made them same it will help in further merging the table for better understanding and visualization .
5. **Check for the continuity of the data:-** I try to find out is the data is continuous or not . my findings says that data is continous.
6. **Data type error**:- I checked the data type of all the columns and corrected the wrong data typed.
7. **Making the table similar:**- There were four table for 2019, each one is for one quarter , but the second quarter table was named differently so I changed the name of column in such a way that all tabe become similar , Now all the table can be append .

3-A summary of the analysis:- For the analysis purpose I used power bi Desktop , I used for ETL , analysis and visualization. Before I started analysis I loaded all the related file in the Power BI desktop. I did my deatialed ETL there . Finally I started my visualization which are as follows.

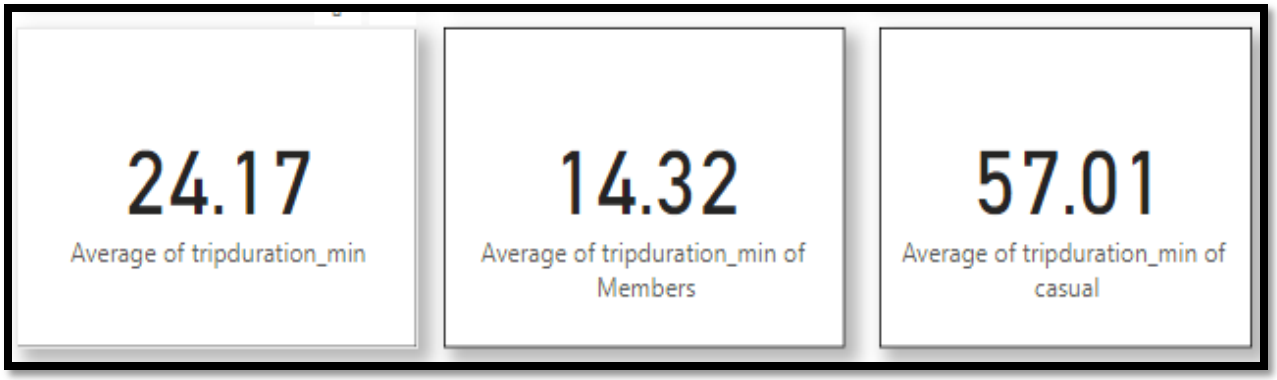
4-Supporting visualisations and key findings:-

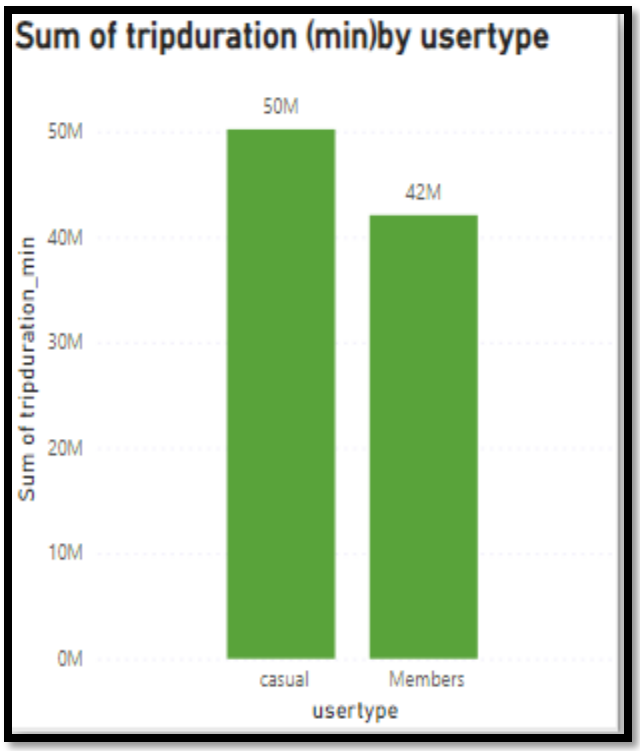
Total no of trips-

First of all we will be visualizing the numbers of trips , which could be infer as number of rides.Here we can see the total numbers of trips is 3.82 million. In which trips done by members is 2.94 millions and trips done by casuals is 0.88 millions.

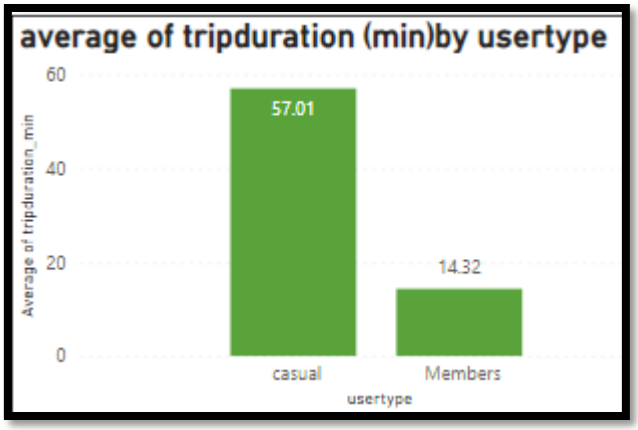


Average trip duration –

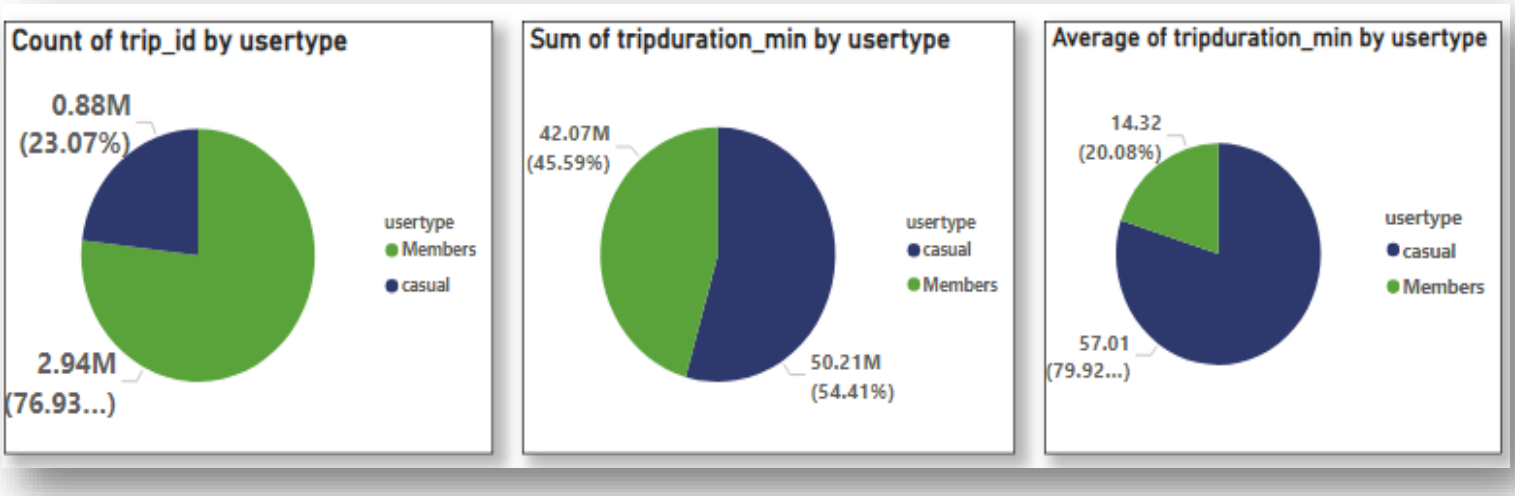




Here It is clearly seen that average trip duration in minute is greater for Members than casual , what does it mean , It mean that casuals were doing longer trip than Members .I will show evidence for this .In this Bar char we can easily see that sum of total trip duration for casuals is more than Member



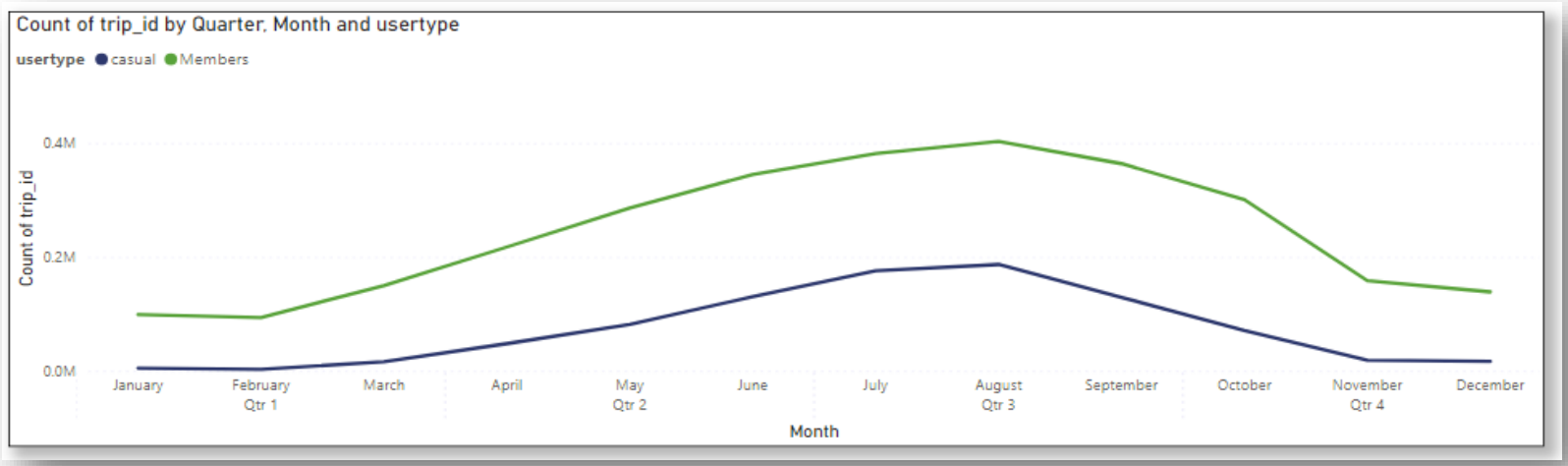
This is a big finding despite the larger number of members , cauals are having more traveling , they are spending more time on bike, this can be further displayed as

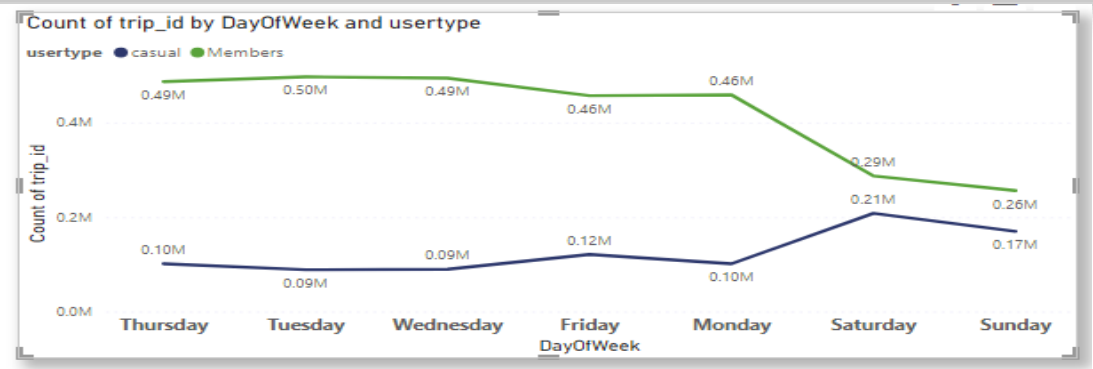


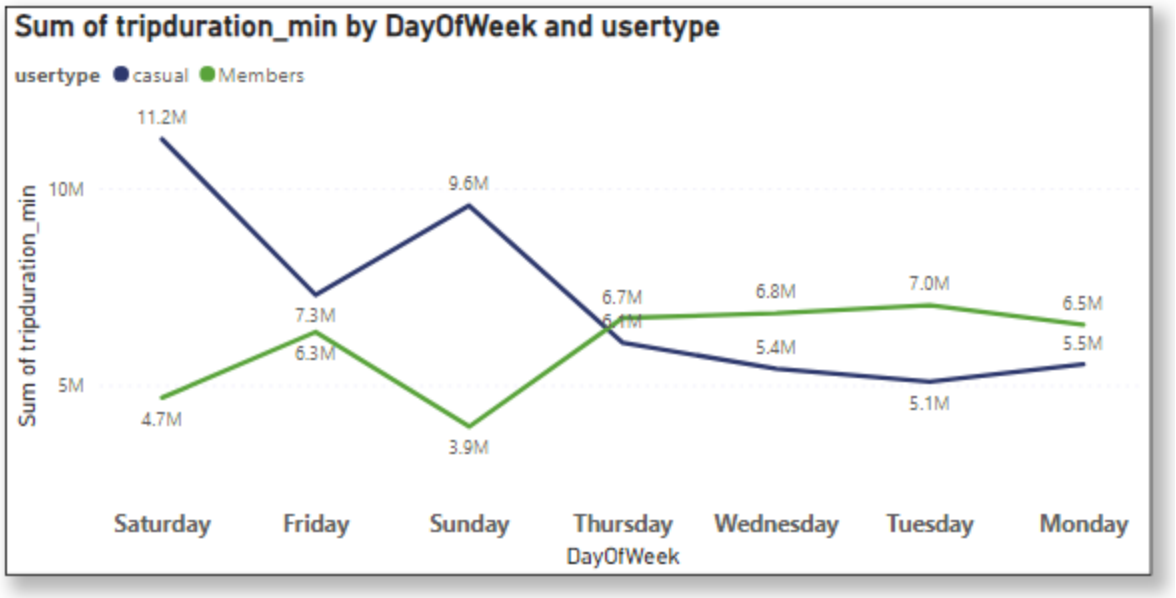
follows.

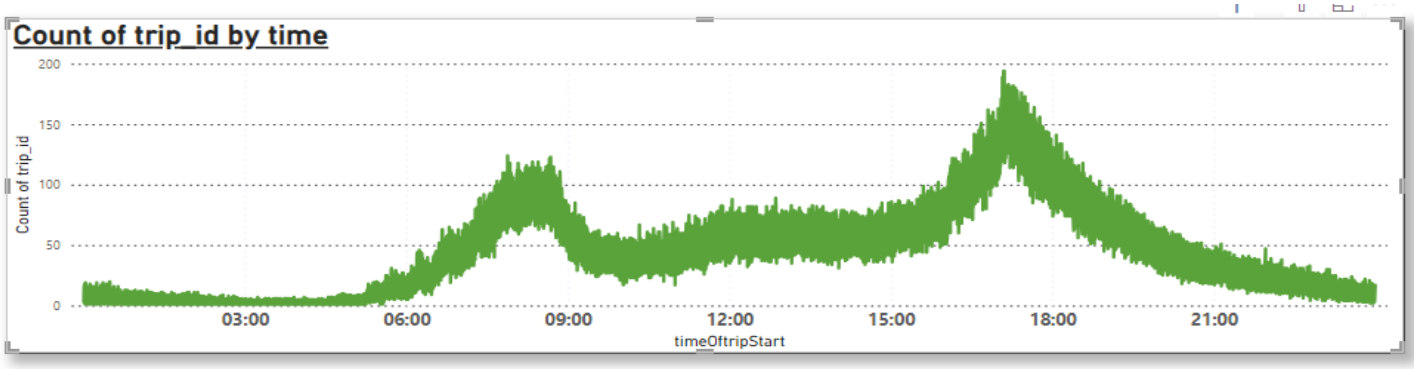
For Lily Moreno: The director of marketing and manager:-

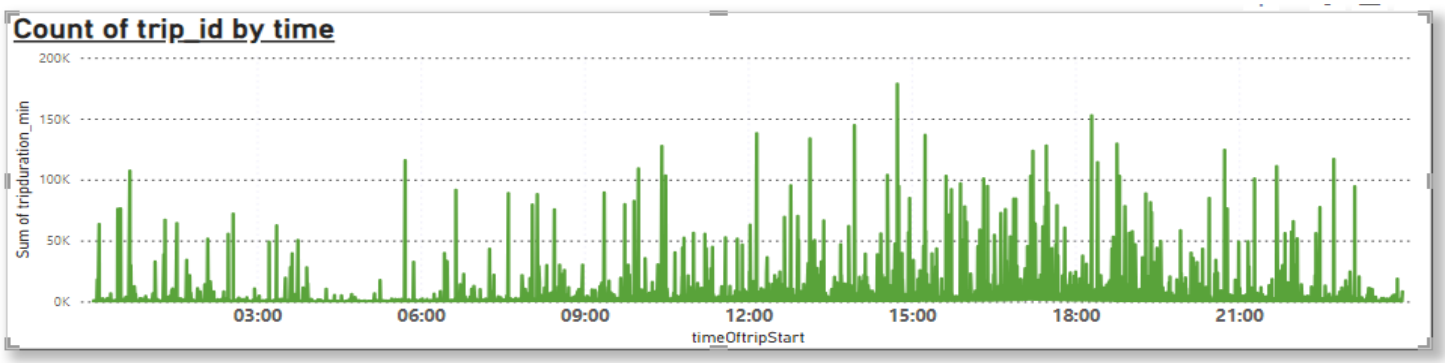
If marketing team is looking for trends for marketing and campaign here it is –

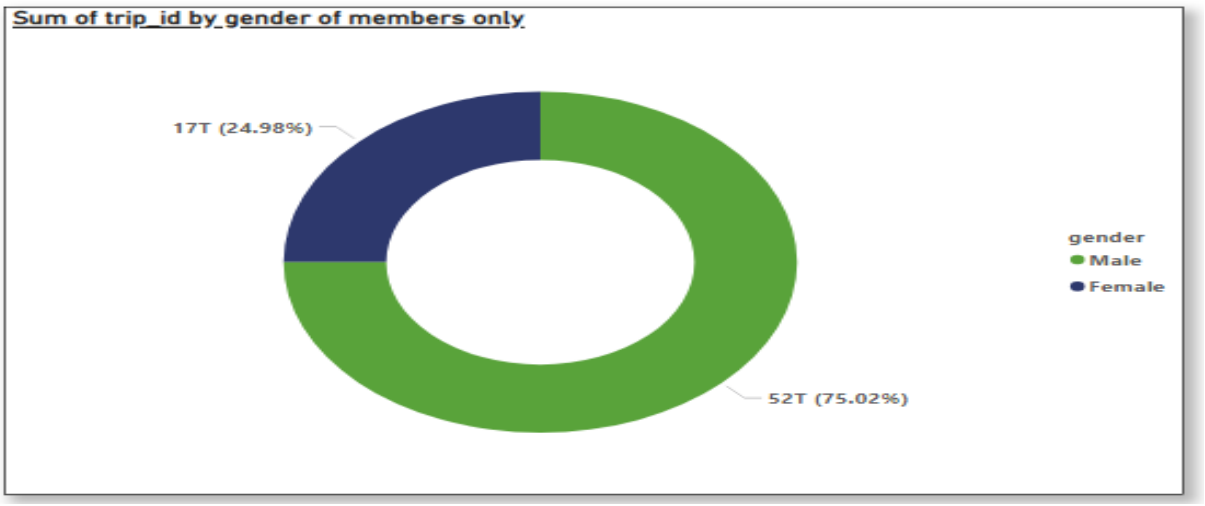


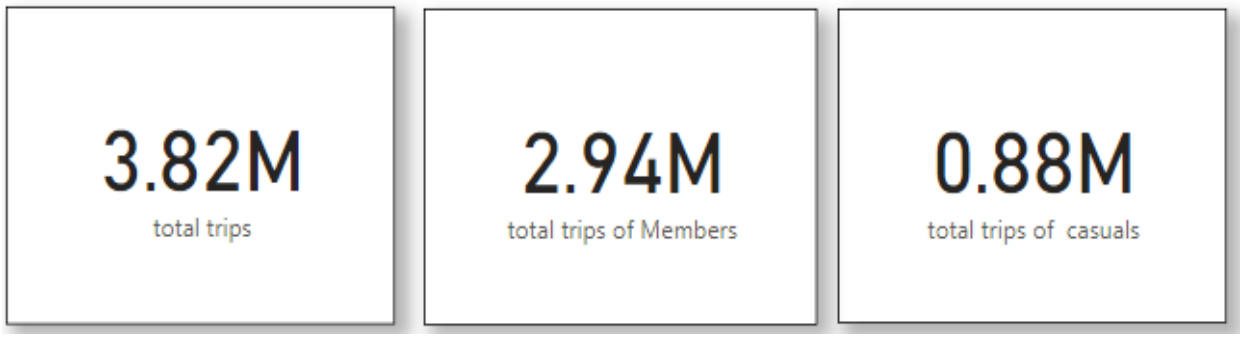



Sum of tripduration_min by DayOfWeek and usertype
usertype casual Members
Sum of tripduration_min
10M
5M
11.2M
9.6M
7.3M
6.3M
6.7M
6.1M
6.8M
7.0M
6.5M
5.4M
5.5M
5.1M
4.7M
3.9M
Saturday
Friday
Sunday
Thursday
Wednesday
Tuesday
Monday
DayOfWeek

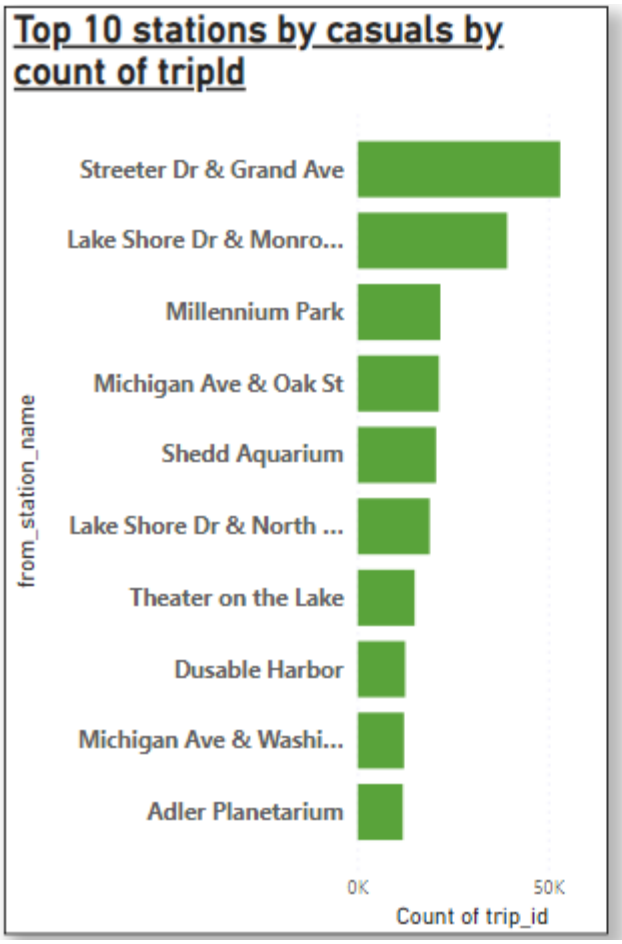

Count of trip_id by time
Count of trip_id
200
150
100
50
0
03:00
06:00
09:00
12:00
15:00
18:00
21:00
timeOftripStart

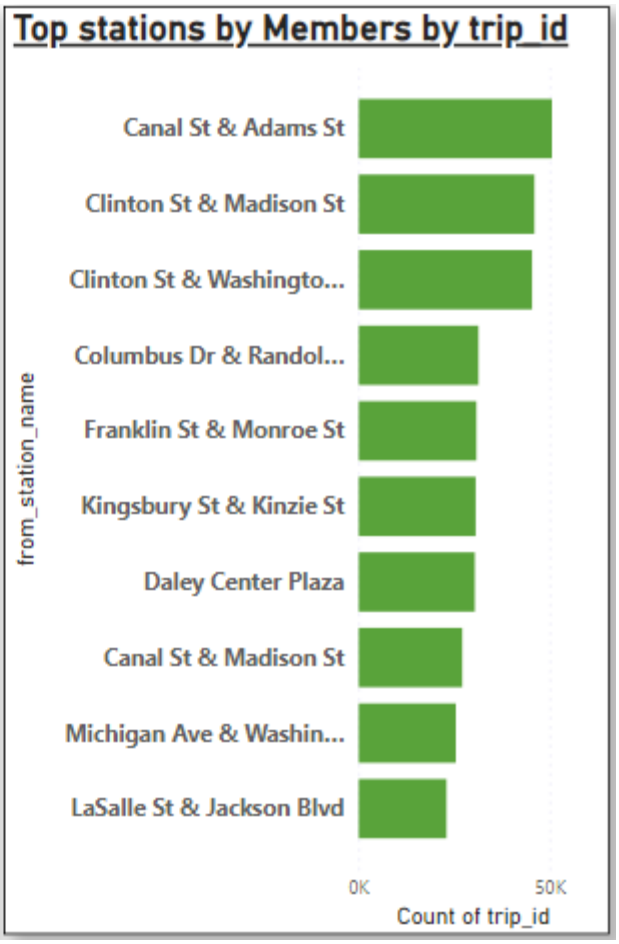

Count of trip_id by time
Sum of tripduration_min
200K
150K
100K
50K
0K
03:00
06:00
09:00
12:00
15:00
18:00
21:00
timeOftripStart


Sum of trip_id by gender of members only
17T (24.98%)
52T (75.02%)
gender
Male
Female

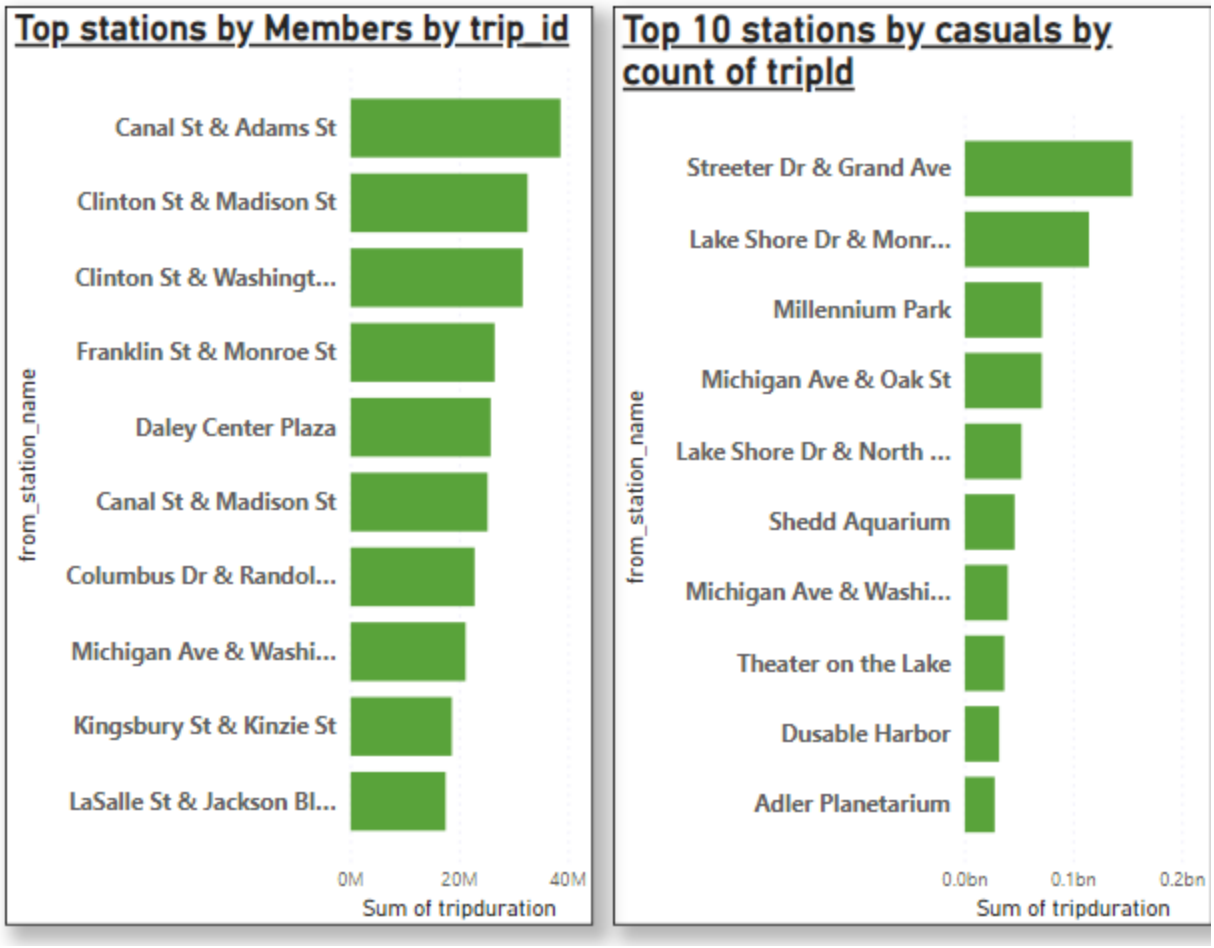


Here I created list of top ten station based on count of trip Id which would be beneficial for campaign team for marketing -





-



6. Your top three recommendations based on your analysis-

1. Marketing team should make a campaign which should run on high traffic months like during third quarter that is July to September.
2. Company should insure good facilities during weekend because on these days casuals people are more which can be converted into members by just giving more attention and running some good campaign.
3. Between 13:00 to 18:00 numbers of both type of customers is more marketing team utilize these time as opportunity and just little more good service during these time can be beneficial. During these hours company can increase the number of worker.
4. There are certain station which has high no of both type of customers, some Special look after these stations can be very beneficial.