

Google Data Analytics

Data Analysis

Capstone Case Study

Ashwini Nayak R

<https://www.coursera.org/user/73d06da80dc2aa250a13bb42d7287439>

December 2023



Google-Data-Analytics

Capstone-Case-Study-1

CYCLISTIC-BIKE-SHARE



Introduction

Welcome to the Cyclistic bike-share analysis case study. In this case study we will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act which we have learned in the 8 courses of professional track in order to answer the key business questions. It gives me a pleasure to share my findings on data given using various tool I studied so far. The following tools are used for analysis

About the company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Scenario

I am a junior data analyst working in 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, my 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 our recommendations, so they must be backed up with compelling data insights and professional data visualizations.



Points to be taken into account

- Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders.
- The director of marketing believes that maximizing the number of annual members will be key to future growth.
- Rather than creating a marketing campaign that targets all-new customers, there is a very good chance to convert casual riders into members
- casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Phase 1: ASK

Ask the right question

The following three questions will guide the future marketing program:

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

In the case study we are focusing on the first question

Identify the business task

We consider two business task here

1. Investigate on variation and difference between casual riders and annual riders
2. Focusing on process of converting casual riders into annual riders

Identify key stakeholders.

- **The director of marketing:** who is responsible for the development of campaigns and initiatives to promote the bike-share program.
- **The Cyclistic executive team:** who is detail-oriented and will decide whether to approve the recommended marketing program.
- **The Cyclistic marketing analytics team:** which is a team of data analysts who are responsible for collecting, analysing, and reporting data that helps in building marketing strategies.

Phase 2: PREPARE

The preparation of the data is the second step in the data analysis process. During this stage, I must comprehend the many metrics needed for analysis and ascertain whether the data is internal or external by locating it. When handling the data, it's also crucial to take any required security precautions into account.

Data Location: I am using Cyclistic's historical data for this project, which is open to the public and kept on their cloud servers. You can access the data [[Here](#)]. I will be concentrating on data from 2023, which is set up into 11 distinct CSV files, each of which corresponds to a month of the year.

Data Organization

1. Divvy's public historical trip data is organized by month and year and saved as a zip file.
2. Each csv file is structured utilizing rows and columns.

Data Credibility ROCCC?

This data ROCCCs! Taken from Divvy's public historical trip data, this data is.

1. Reliable: this data is accurate, complete, and unbiased.
2. Original: this data is validated by the original source
3. Comprehensive: this data contains all the critical information to answer the questions
4. Current: this data is useful and up to the date
5. Cited: this data is credible and cited from this source [[Here](#)]

Data License

1. The data is maintained and made available by Motivate International Inc. under this [license](#)

2. that data-privacy issues prohibit me from using riders' personally identifiable information.
3. I won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

Phase 3: Data Processing

Data cleaning and Data Manipulation

1. Merged datasets of October and November month in Excel
2. Removed unwanted 4 columns Source.Name, start_lat, start_lng, end_lat, end_lng so that we have dataset consists of 9 columns and 8,99,631 rows.
3. The **rideable_type** column has two categories: **classic-bike** and **electric-bike**(by using filter).
4. There are 4,44,520 classic bike and 4,55,111 electric bikes
5. Found that more than **15.5%** of the rows have blank (using **IFBLANK** function) entries in some columns like (**start_station_name**) and (**end_station_name**).

COLUMN NAME	NUMBER OF BLANKS
ride_id	0
Rideable type	0
Started at	0
Start time	0
Ended at	0
End time	0
Ride length	0
Start station name	139415
Start atation id	139415
End station name	147152
End_station_id	147152
Member_casual	0

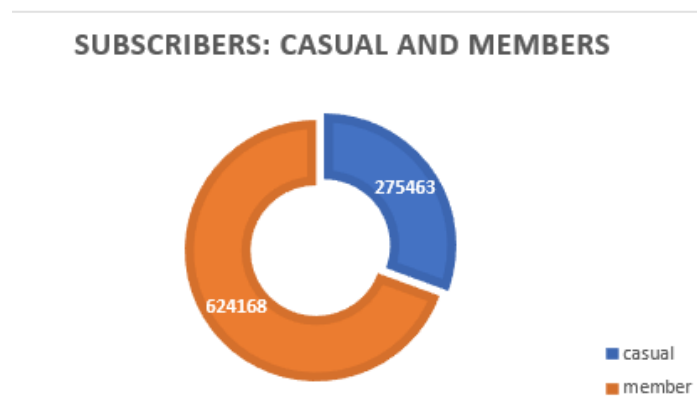
6. Deleting these rows would significantly reduce the amount of data available for analysis and could potentially affect the results. Since there is valuable data in other columns, I decided to treat these blank entries as N/A "Not Available" using **IF** function instead of deleting them.
7. It is observed that the ride_ids are not proper by using filter after checking number of characters in that column using **LEN** function. Since it doesn't play any significant role in our analysis, it will be kept as it is.
8. Separated the columns for time and date. Created column for day and month separately using **TEXT** function. This will help analysing about riders preference to time and day, upon which the marketing strategies can be built.
9. Created a new column called ride_length by taking the difference of the columns end_time and the start_time. Since the excel could not take the difference of time such

- as 23:03:43 from 00:22:34, I used a **MOD** function. Checked for correct values using filter. When applied **AVERAGE** function to this column, it came out to be 00:13:32
- Now the dataset consists of 14 columns and 8,99,631 rows
 - Rechecked for blanks in every columns using filters.
 - The headers are made bold to stand out from the rest and this row is filled with gray color to distinguish it from the remaining rows. The first row is then frozen by using 'freeze' command from view menu. This helps us to identify the column clearly even while scrolling. The table now looks as follows:

ride_id	rideable_type	started_at	start_day	month	start_time	ended_at	end_day	end_time	ride_length	start_station	start_station_name	end_station	end_station_name	member_casual
4449097279F88BE7	classic_bil	08-10-2023 00:00	Sunday	October	10:36:26	08-10-2023 00:00	Sunday	10:49:19	00:12:53	Orleans St	620 Sheffield	TA1309000	member	
9CF060543CA7B439	electric_b	11-10-2023 00:00	Wednesd	October	17:23:59	11-10-2023 00:00	Wednesd	17:36:08	00:12:09	Desplaine	TA1306000	Sheffield	TA1309000	member
667F21F4D6BDE69C	electric_b	12-10-2023 00:00	Thursday	October	07:02:33	12-10-2023 00:00	Thursday	07:06:53	00:04:20	Orleans St	620 Franklin S	TA1307000	member	
F92714CC68019B96	classic_bil	24-10-2023 00:00	Tuesday	October	19:13:03	24-10-2023 00:00	Tuesday	19:18:29	00:05:26	Desplaine	TA1306000	Franklin S	TA1307000	member
5E34BA5DE945A9CC	classic_bil	09-10-2023 00:00	Monday	October	18:19:26	09-10-2023 00:00	Monday	18:30:56	00:11:30	Desplaine	TA1306000	Franklin S	TA1307000	member
F7D7420AFAC53CD9	electric_b	04-10-2023 00:00	Wednesd	October	17:10:59	04-10-2023 00:00	Wednesd	17:25:21	00:14:22	Orleans St	620 Sheffield	TA1309000	member	
870B2D4CD112D7B7	electric_b	31-10-2023 00:00	Tuesday	October	17:32:20	31-10-2023 00:00	Tuesday	17:44:20	00:12:00	Orleans St	620 Sheffield	TA1309000	member	
D9179D36E32D456C	classic_bil	02-10-2023 00:00	Monday	October	18:51:51	02-10-2023 00:00	Monday	18:57:09	00:05:18	Desplaine	TA1306000	Franklin S	TA1307000	casual
F8E131281F722FEF	classic_bil	17-10-2023 00:00	Tuesday	October	08:28:18	17-10-2023 00:00	Tuesday	08:50:03	00:21:45	Calumet A	13102 Morgan St	TA1307000	member	
91938871748FA05	classic_bil	17-10-2023 00:00	Tuesday	October	19:17:38	17-10-2023 00:00	Tuesday	19:32:23	00:14:45	Wolcott A	TA1309000	Morgan St	TA1307000	member
1918FA255C1820FC	classic_bil	04-10-2023 00:00	Wednesd	October	15:24:00	04-10-2023 00:00	Wednesd	15:32:30	00:08:30	Wolcott A	TA1309000	Morgan St	TA1307000	member
52E0BC8010C7DFE9	classic_bil	17-10-2023 00:00	Tuesday	October	11:49:58	17-10-2023 00:00	Tuesday	12:02:57	00:12:59	Larrabee S	TA1306000	Franklin S	TA1307000	member
0552E9F51D63509C	classic_bil	18-10-2023 00:00	Wednesd	October	16:28:33	18-10-2023 00:00	Wednesd	16:53:27	00:24:54	Wolcott A	TA1309000	Morgan St	TA1307000	member
3A06FDA7CB096C8D	electric_b	04-10-2023 00:00	Wednesd	October	15:47:31	04-10-2023 00:00	Wednesd	15:53:01	00:05:30	Western A	13068 Logan Blvr	TA1308000	member	
0A577547F72D98C9	classic_bil	25-10-2023 00:00	Wednesd	October	16:24:45	25-10-2023 00:00	Wednesd	16:32:00	00:07:15	Wolcott A	TA1309000	Morgan St	TA1307000	member
F914F9932D04F1C6	electric_b	11-10-2023 00:00	Wednesd	October	15:44:11	11-10-2023 00:00	Wednesd	15:56:50	00:12:39	Western A	13068 Sheffield	TA1309000	member	
2E536F84DAC106A9	classic_bil	22-10-2023 00:00	Sunday	October	14:12:30	22-10-2023 00:00	Sunday	14:22:48	00:10:18	Larrabee S	TA1306000	Sheffield	TA1309000	casual
E03FD93521EEA202	electric_b	13-10-2023 00:00	Friday	October	03:01:38	13-10-2023 00:00	Friday	03:14:29	00:12:51	Western A	13068 Public Rac	866	casual	
DA0803D1C2EB4F62	electric_b	29-10-2023 00:00	Sunday	October	13:52:13	29-10-2023 00:00	Sunday	14:00:51	00:08:38	Wolcott A	TA1309000	Morgan St	TA1307000	casual
517756C65FC62A95	classic_bil	05-10-2023 00:00	Thursday	October	16:09:03	05-10-2023 00:00	Thursday	16:26:42	00:17:39	Western A	13068 Central Pa	15644	casual	

Phase 5 & 6 : Analysing ,Visualizing and Sharing Data

- Since we are into converting casual riders to members, we first count number of casual riders and member riders in October and November month **COUNTIF** function can also be used to get the result.



- The mean of ride length is 00:13:32 while maximum ride length being 23:59:59 by a casual rider on Sunday 01-10-2023

- After creating the pivot table, it was found that the University Ave & 57th St is the start station preferred by riders in maximum number of 6851 times (using **LARGE** function). This station is also minimum in number as end station. The following is the pivot table reference.

Row Labels	Count of start_station_name
University Ave & 57th St	6851
Grand Total	6851

start_station_name

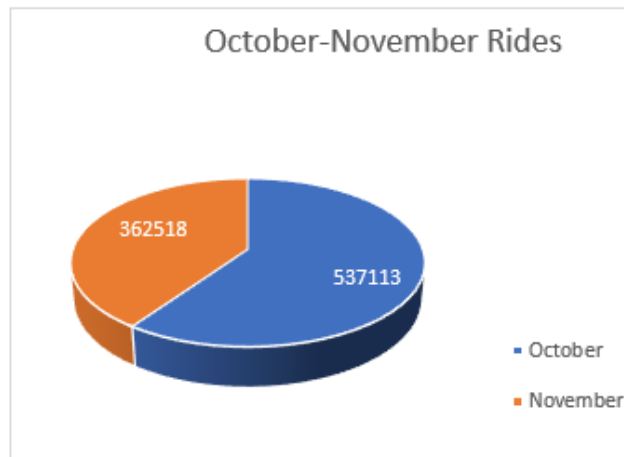
- Troy Ave & 27th St
- Troy St & Elston Ave
- Troy St & Grace St
- Troy St & Jackson Bl...
- Troy St & North Ave
- Union Ave & Root St
- University Ave & 57...**
- University Ave & 59...

Row Labels	Count of end_station_name
University Ave & 57th St	6728
Grand Total	6728

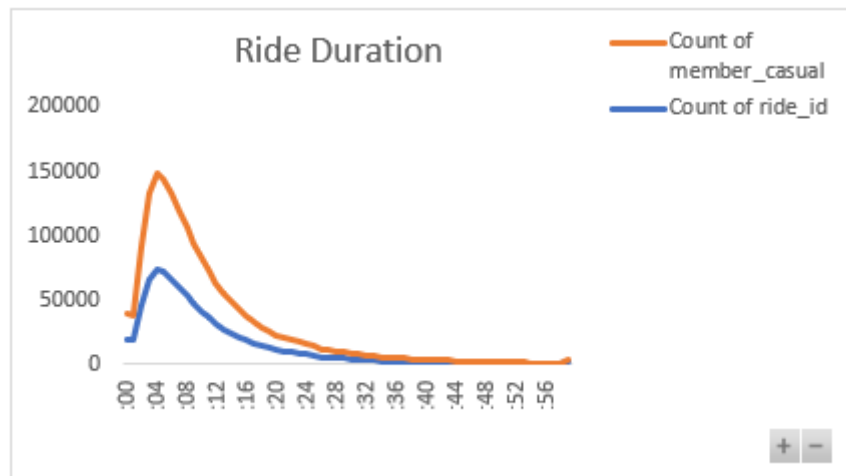
end_station_name

- Troy St & Elston Ave
- Troy St & Grace St
- Troy St & Jackson Bl...
- Troy St & North Ave
- Union Ave & Root St
- University Ave & 57...**
- University Ave & 59...
- University Library (...)

- As we are interested in knowing about subscribers, the count of casual riders and member riders in the month of October and November are noted down

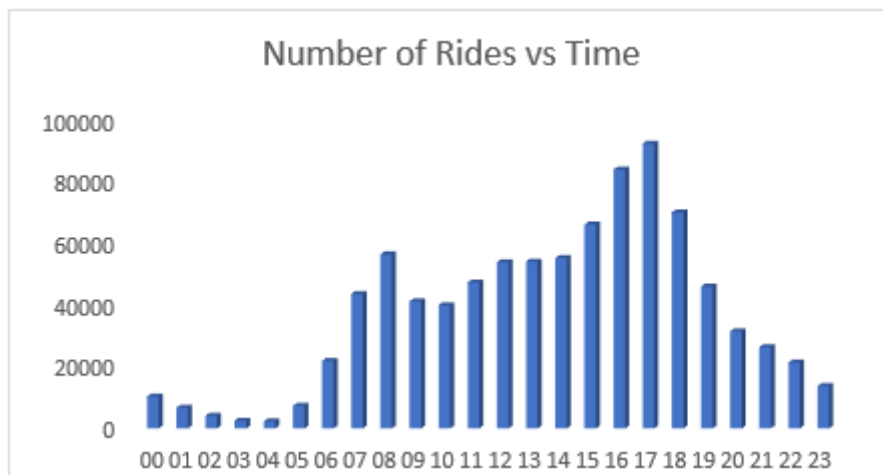


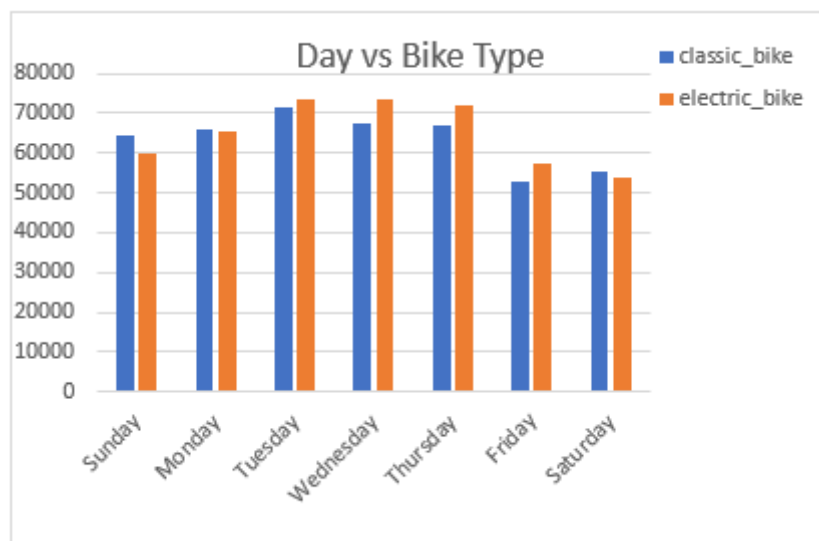
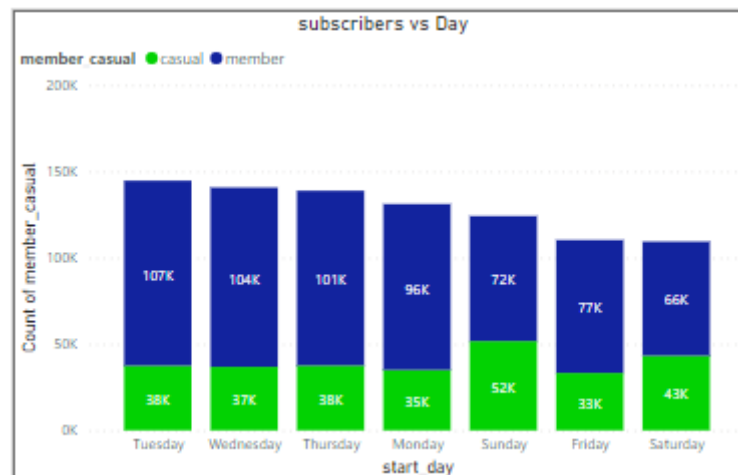
- The average of ride duration for both type of subscribers came out to be 00:13:32 which is obtained using **AVERAGE** function.



But it can be seen that the curve is not normally distributed, data is concentrated more towards one side of the scale, while the long tail on the right side. Hence it is positively skewed curve.

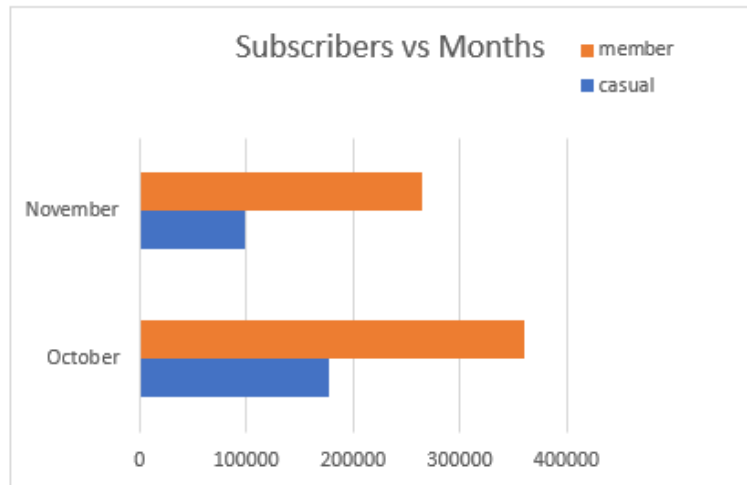
- Analysis on bike type, subscriber type with time and day play important role to draw thoughtful insight. Use of Excel and Power BI was made to observe these relations.



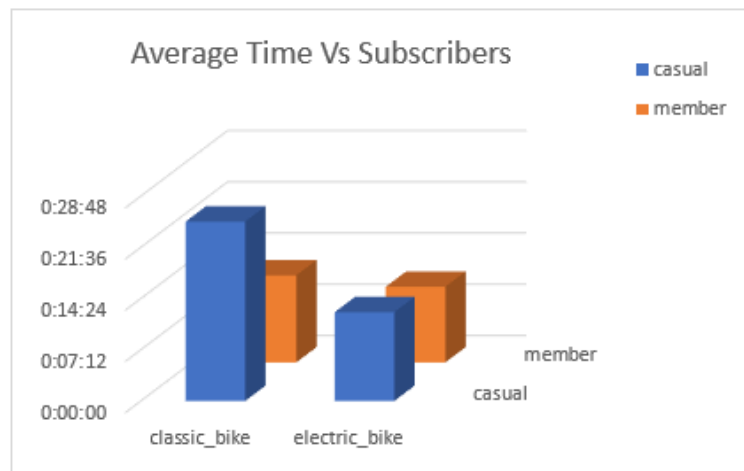


It can be seen in the first chart that there 2 peak hours for bike riders 8 A.M and 7P.M. And in the next two charts it can be observed that in the mid-week Tuesday, Wednesday and Thursday, number of bike riders are more compared to other days. It is evident from the second chart that number of casual riders are more on weekends. The least number of bike rides can be seen on Saturdays.

- Since we have the data regarding two months, it is possible to take look on surface level about the subscribers in these two months.

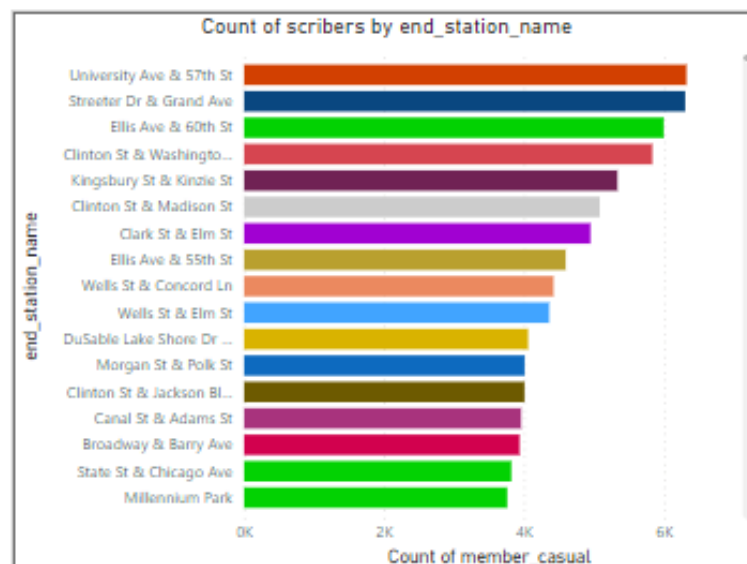
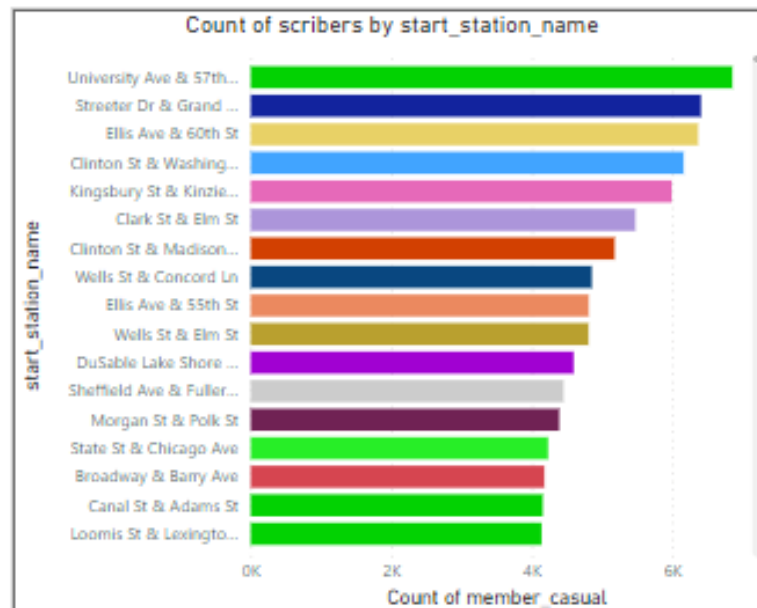


It can be seen that there is a decrease in number of both casual and member riders from October to November.



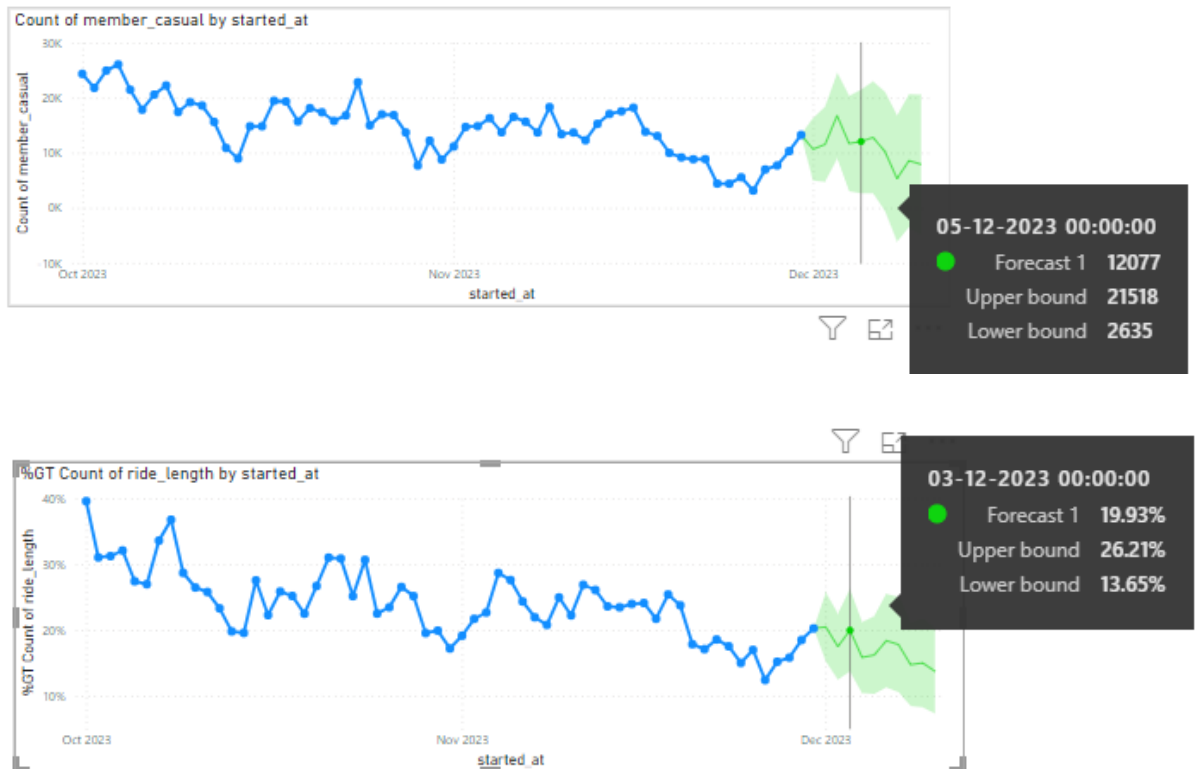
It is evident from the chart that number of casual riders are more and riders prefer classic bike for the ride.

- Here is reference to some of the most preferred start stations and end stations.

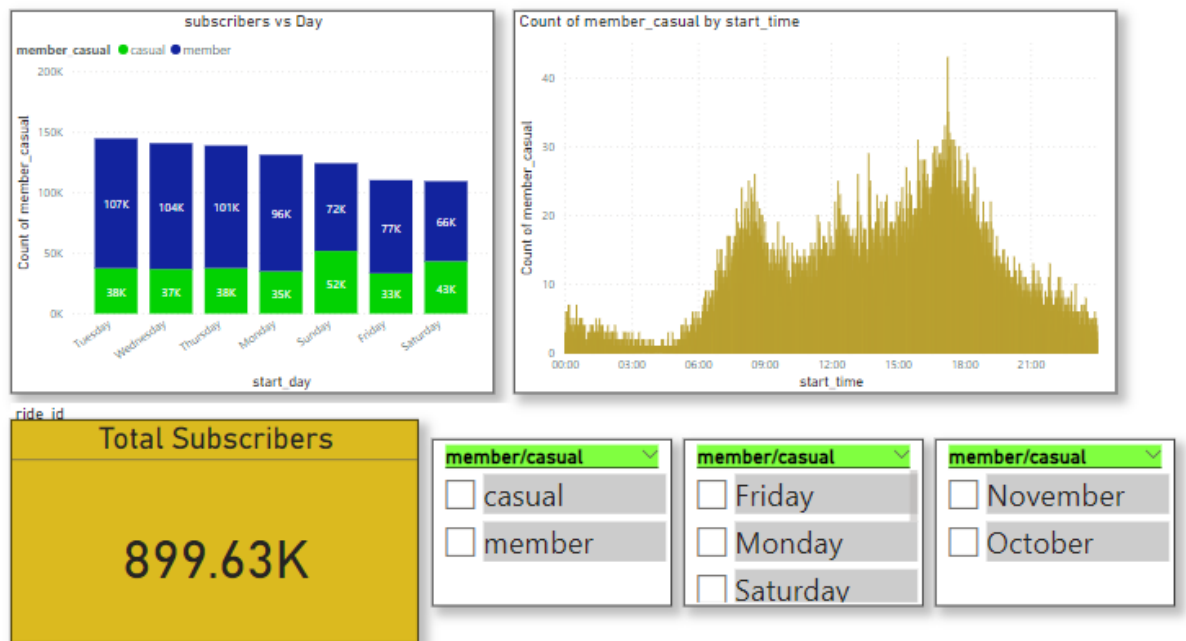


As mentioned earlier it can be seen that the most preferred start and end station is University Ave & 57th St.

- Now based on the historical trends it is possible to predict the future results using power BI. This information can be valuable for understanding the factors influencing business metrics.



- A simple dashboard for basic result is as follows



Phase 6: Act

Recommendations and Conclusions

- University Ave and 57th St, Streeter Dr & Grand, Ellis Ave & 60th St, Cinton St & Washington Blvd are some of highly utilized stations for initiating rides, maintaining systematic and hygienic environment is crucial to ensure the continued attraction of riders to these station. Posting some eye-catching offers about annual membership may encourage casual riders to subscribe for such packages in these stations.
- Providing customized app for annual subscribers and general app for casual riders for booking purpose may help the costumers for booking their rides. Through provision of online booking on bike rides the offers can be made to claim several benefits for both riders and service providers. Offering online booking enhances the overall convenience for riders, making it easier for them to access and use the bike-sharing service.
- The member riders use the bikes more frequently than casual riders in terms of total usage. Casual riders have longer ride times compared to member riders. So it is essential to know the information about the reasons behind riders to be casual and acting upon that to convert them into members accordingly
- If the trends in last two months are followed, it can be observed that the number of customers decreased from October to may by which we may predict that in the next month there might be further decrease in number of customers. The same was seen while forecasting in power BI. This gives the hint that in the winter there might be a loss of customers
- Both the casual and members prefer classic bikes hence availability of more number of classic bikes will facilitate the business objective.
- Offers on electric bike annual membership such as free service on charging stations may encourage casual riders to get the annual membership.
- since casual riders use bikes the most on weekends, it may be a good idea to focus marketing efforts on Saturdays and Sundays. This could help increase the number of casual riders who become members.
- By making customers aware about health benefits of riding bike, many programs such as fun activities, bike marathon for all the annual riders will encourage casual riders to take annual subscription.