

A

Presentation

by

Lilian Nwafor

A Junior Data Analyst

Maximizing the Annual membership of the Cyclistic company

A bike-share company in Chicago

The aim is to analyze the productivity level between casual and member status in the Cyclistic organization so as to decide if converting casual to member or vice versa will improve the profit margin.

The key stakeholders or factors to consider are the status of the bikers, the type of bike used, the ride length and the time it takes to complete their daily tasks.

Problem Statement

- How has the rideable type usage compared to the time duration of daily ride affected the productivity of each cyclist?
- How to determine the activities of the status (member and casual) of the cyclist using their daily ride length

Data Collection

Dataset was collected from [Google Data Analytics Case Study | Kaggle](https://www.kaggle.com/leetdum/google-data-analytics-case-study)
[https://www.kaggle.com › leetdum › google-data-analyt...](https://www.kaggle.com/leetdum/google-data-analytics-case-study)

These files were gathered from an Amazon Web Services webpage titled **Index of bucket "divvy-tripdata"** <https://divvy-tripdata.s3.amazonaws.com/index.html>.

12 months datasets were selected from January to December 2021

From the datasets, variables like Start time and End_Time, Start_Lng and End_Lng, helped in answering the business questions because getting the difference of both variables will help to better understand the activities of Cyclistic company.

Data Cleaning and Wrangling Approach

Summary: Data cleaning and manipulation was done with R statistical software and merging of files was done using Jupiter notebook in python. There was no duplicates. Columns were dropped because they contained large number of rows with missing data and most variables are irrelevant to this analysis. New columns were added in order to get the difference of Started & ended time, start lat & end lat, start lng & end lng. I examined the summary of the test statistics. The mean of ride length for each month was checked. Getting the time difference and ride length difference helped me to decide the riders to be converted to a membership status because the longer the time spent or ride length, the more profit to the organization.

Details Below:

Columns Dropped:

Start_Station_Name
Start_Station_ID
End_Station_Name
End_Station_ID
X

The above listed columns were dropped because some contained large number of rows with missing data and most variables are irrelevant to this analysis.



Columns Added:

Time_Diff

Lat

Ride_Lng

New columns were added in order to get the difference of Started & ended time, start lat & end lat, start lng & end lng.

Column Renamed:

Member_casual = Status

Column was changed to status for clearer understanding.

The difference of the following column was derived:

Started_Time - Ended_Time = Time_Diff

Start_lng - End_lng = Lag

Start_Lng - End_Lng = Ride_Lng

Extracted Columns For Analysis

Ride_ID

Rideable_Type

Status

Ride_Lng

Time_Diff

Lat

Table Merging and Filtering

Parameters filtered were:

The ride length ≥ 0.1

Time difference > 400

All months of each filtered parameters was merged into one year. The merging was done using Jupiter Notebook.

To view the data cleaning and wrangling details, check link below:

<https://github.com/Li-Ndibe/Bike-Analysis-Project.git>

Dataset Description

The detailed of the data statistics are listed below:

Month	No of Columns	No. of Rows
January	6	131573
February	6	96,834
March	6	49,622
April	6	214,272
May	6	337,230
June	6	337,230
July	6	531,633
August	6	729,595
September	6	822,410
October	6	804,352
November	6	519,380
December	6	631,226

ANALYSIS

The following slides show the bar plots of the percentage of each activities of the rideable type and status.

Slide 9: Shows that Electric bike is mostly used compared to Docked and Classic bikes. Electric bike was used more in February unlike other months.

Slide 10: Compares the clients status activities. Casual bikes are the highest clients of the Cyclistic compared to the members, this may be due to high cost of registration membership status.

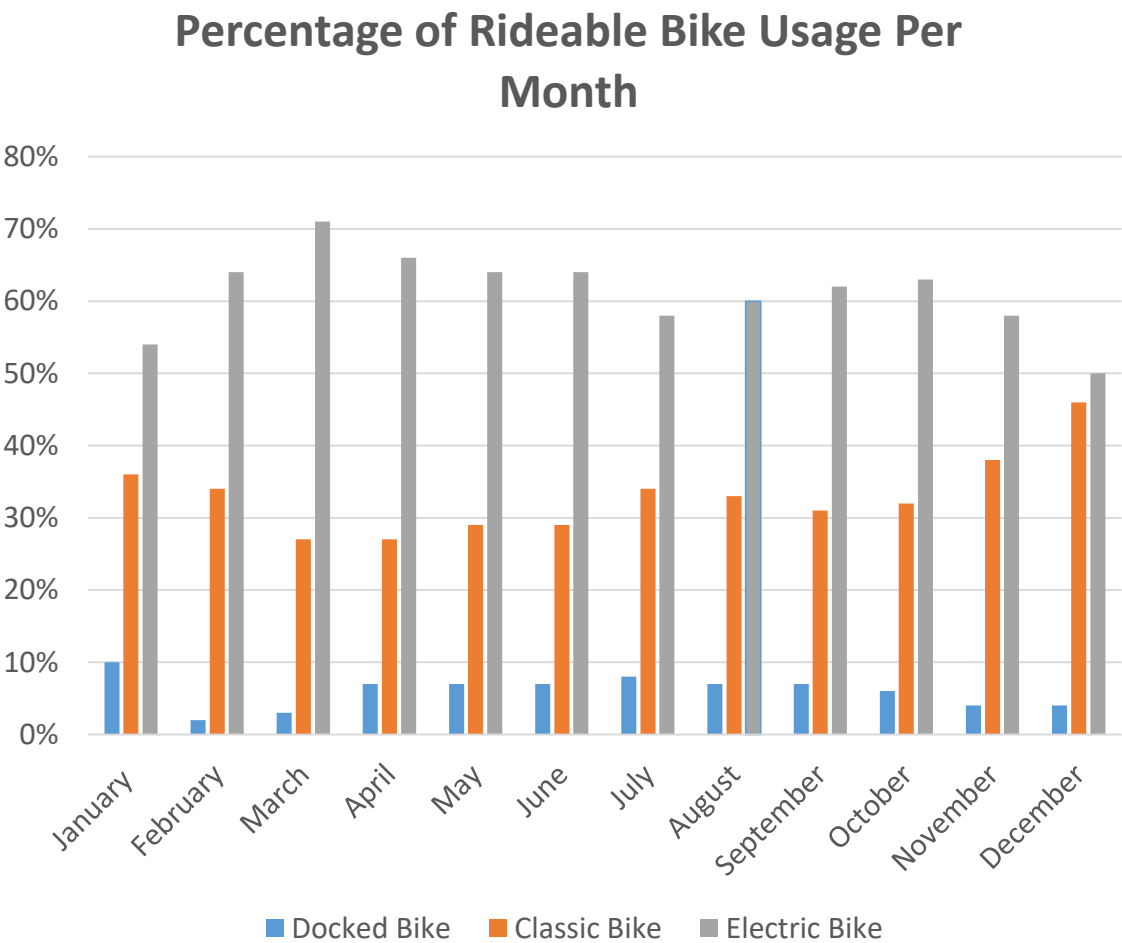
Other metrics visualized and analyzed are as follows:

- a. Average Time Difference of the Clients For Each Month
- b. Maximum and Minimum Ride Length
- c. Mean of Ride Length of Each Bike Type for Each Month
- d. Average Time Difference of Each Rideable Bike For Each Month
- e. Mean of Time Difference greater or equal to 400 with Ride Length
- f. Average Mean of Ride Length for Each Month

The results can be found in the Tableau page <https://public.tableau.com/app/profile/lindibe>

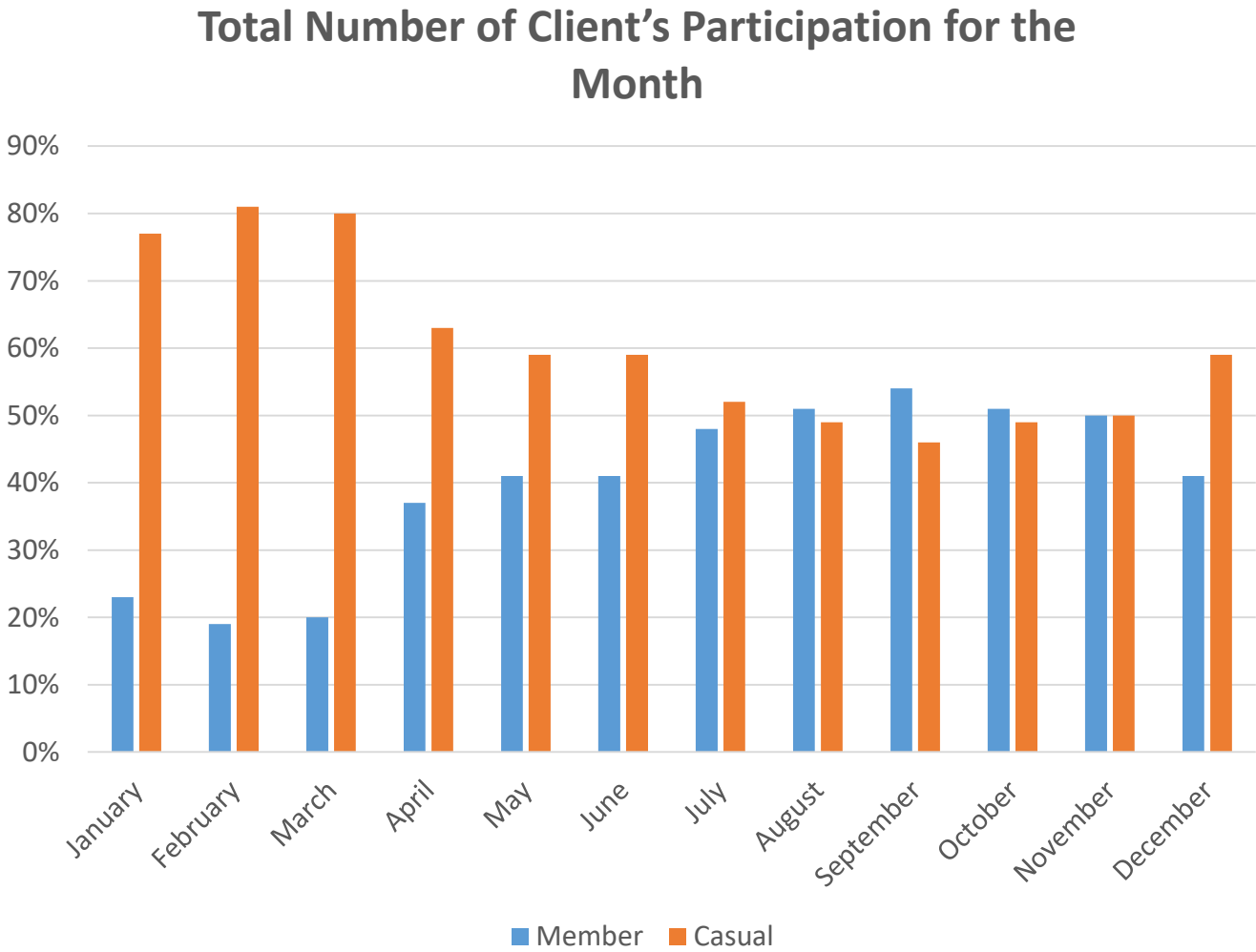
Below is the monthly analysis of types of bike used. This is to determine factors affecting the bike usage and to understand the reason behind usage preferences. Generally, Electric bike is more used by clients which could be due to its flexible and easy operating nature. Electric bike was more used in the month of March compared to the other months. Factors such as cost, weather, promotional activities may be the reasons for it's preference.

Month/Bike	Docked Bike	Classic Bike	Electric Bike
January	10%	36%	54%
February	2%	34%	64%
March	3%	27%	71%
April	7%	27%	66%
May	7%	29%	64%
June	7%	29%	64%
July	8%	34%	58%
August	7%	33%	60%
September	7%	31%	62%
October	6%	32%	63%
November	4%	38%	58%
December	4%	46%	50%



The Chart below shows the total number of client’s participation for each month. Casual clients participate more than the member clients except for the months of August, September and October. There is need to look into the factors for the sudden drop of Casual clients participation and rise in member interest in Bike-share. Overall Casual members are more actively involved in the year of 2021.

Month/Status	Member	Casual
January	23%	77%
February	19%	81%
March	20%	80%
April	37%	63%
May	41%	59%
June	41%	59%
July	48%	52%
August	51%	49%
September	54%	46%
October	51%	49%
November	50%	50%
December	41%	59%



OBSERVATION

From the visualization, the following were observed:

- a. The highest maximum ride length 0.3532 in the month of April, while the minimum ride Length was 0.21263 in the month of October.
 - b. The mean of ride length of different group of clients for each month were visualized. Average mean of casual clients were relatively lower than the member clients except for the month of March than recorded .000555. Member clients in the month of May and June recorded the mean of 0.0003130 respectively.
 - c. Though the member clients spent more time in the Cyclistic company but the casual clients are more in number, therefore their ride length is higher than the member clients.
-

Challenge

The major challenge encountered was the large number of cells with missing values which obviously affected the accuracy of the analysis. I believe if there was no or minimal missing values, a nearly accurate result would have gotten.

Summary

In conclusion, the casual members participate more in the Cyclistic company considering the time spent daily in the organization. Converting Status of the clients from casual to member would not change much, focus should be on the activities of the clients using duration of seconds spent, length of the ride and type of bike used.

Providing more of the electric rideable bikes should be prioritized because it is more in demand by the clients unlike the docked and classic bikes, profit of the company relies on it.

I believe that using a benchmark of time difference of 400 seconds and above which is approximately 7 minutes daily should be put into consideration when making a business decision because the longer a rider rides, the more money is being generated.

THANK YOU