Cyclistic company customers' analysis

Michael Meng, 29 Sep 2021

1 Project Background and working environment setting

1.1 Project background introduction

Cyclistic bike-share analysis is a frictional case study organized by the Google Data Analytics capstone project team. In the project hypothesis, the marketing director of Cyclistic company, Lily Moreno, wants to design a marketing strategy for converting casual riders to annual members to reinforce the company's future. The marketing strategy should answer the following questions.

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?

This section analysis tries to answer the first question, finding behavior differences between customers of two types. And any patterns, trends will be helpful for future project advances.

1.2 Data resources and ROCCC inspect

- Cyclistic is a fictional company for the case study.
- The data has been made available by Motivate International Inc. under this license.
- The data was downloaded from amazonaws website link.
- As case study statement required, downloaded previous 12 months data from Sep 2020 to Aug 2021.

1.3 Install required packages and other environments setting

The analysis will need OS to upload data, Numpy and Pandas help interact with data, and Pyplot and Seaborn libraries to plot.

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

2 Processing data: uploading, inspecting, cleaning, and compiling

2.1 Downloading and Uploading data:

- Download data from Amazonaws website (Link provided in section 1.2), and store 12 monthly data in the same folder for convenience of following processes.
- Upload monthly data

```
import os
os.chdir('D:/resume_portfolio/cyclistic_google_case1/divvy_tripdata')
m8_2021 = pd.read_csv('202108-divvy-tripdata.csv')
m8_2021.info
```

```
m6_2021 = pd.read_csv('202106-divvy-tripdata.csv')
m5_2021 = pd.read_csv('202105-divvy-tripdata.csv')
m4_2021 = pd.read_csv('202104-divvy-tripdata.csv')
m3_2021 = pd.read_csv('202103-divvy-tripdata.csv')
m2_2021 = pd.read_csv('202102-divvy-tripdata.csv')
m1_2021 = pd.read_csv('202101-divvy-tripdata.csv')
```

2.2 Inspect monthly data and stack the full year dataset.

```
m8_2021.columns
```

```
Index(['ride_id', 'rideable_type', 'started_at', 'ended_at',
       'start_station_name', 'start_station_id', 'end_station_name',
       'end_station_id', 'start_lat', 'start_lng', 'end_lat', 'end_lng',
       'member_casual'],
      dtype='object')
```

- Other useful functions or features: info, shape, dtypes, type(column_name), head()
- Stack yearly data.

data = pd.concat([m8_2021, m7_2021, m6_2021, m5_2021, m4_2021,m3_2021, m2_2021, m1_2021, m12_2020,m11_2020,m10_2020,m09_2020])

2.3 Inspect full year data

2.3.1 Missing value

data.isnull().sum() to get an instinct view.

2.3.2 Numeric features

data.describe().T: Found some wrong data need deleted

2.3.3 Categorical features

Unique values and number

data.member_casual.value_counts()

2.4 Feature engineering

- Need increase duration, weekday and month columns for future analysis
  - data['duration'] = data['ended_at'] - data['started_at']
  - data['month'] = data["started_at"].dt.strftime("%Y-%m")
  - data['weekday']=data['started_at'].dt.day_name()
- column datatype converting
  - data['duration']=data['duration'].astype('timedelta64[s]')

2.5 Data cleaning

Delete duration that is less or equal to 0, that's impossible or data operator assigned a negative number representing bikes in maintenance.

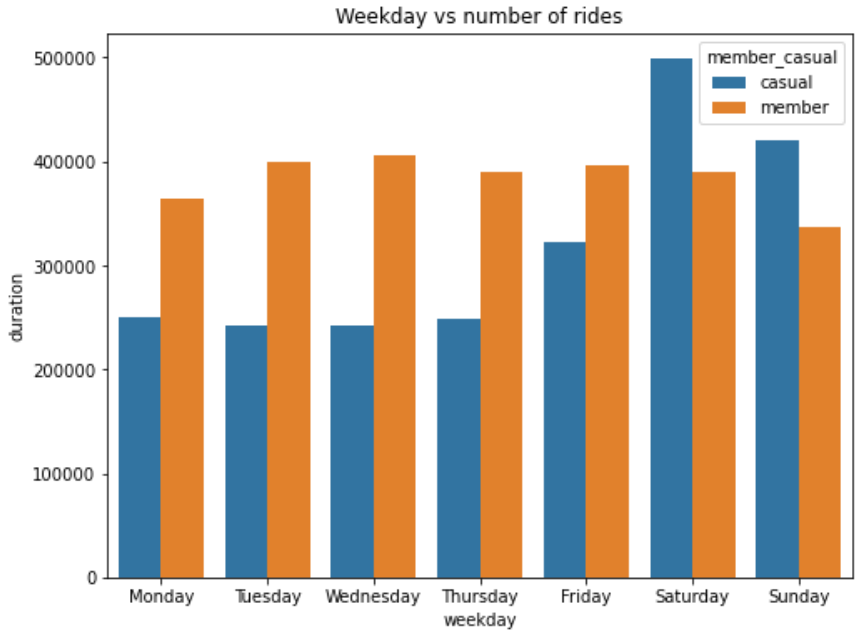
- data = data.drop(data[data.duration<=0].index)

3 Data Analysis

3.1 Weekly trends analysis

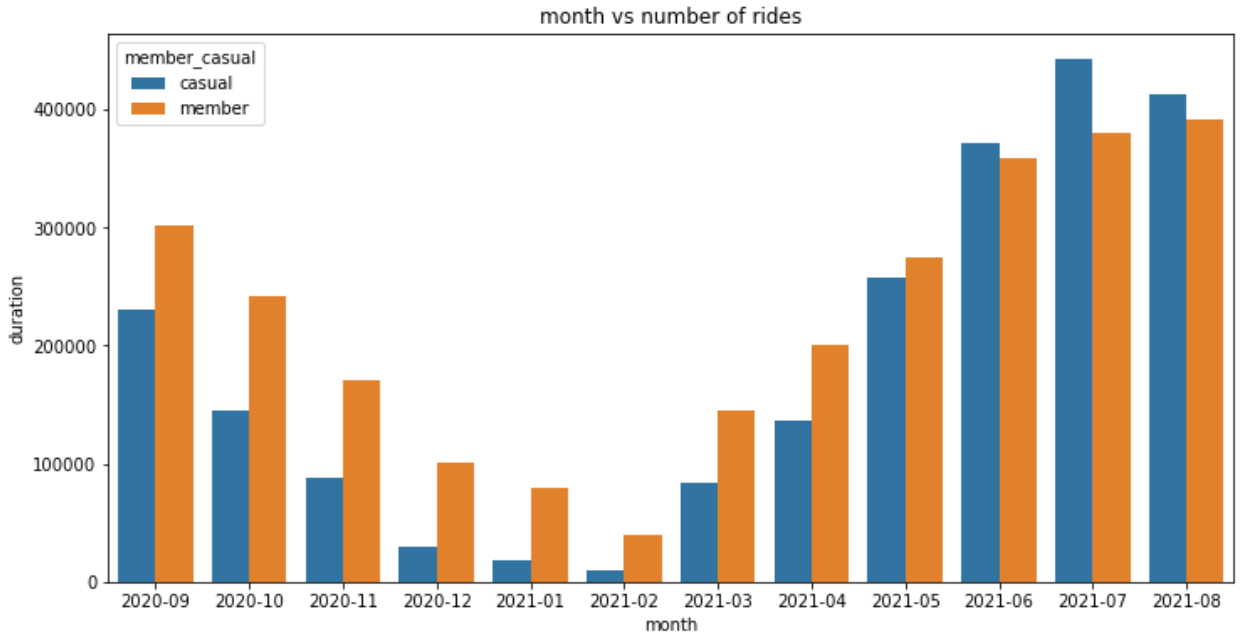
Weekend rides is 1.26 times of workdays.

Casual customer like weekend

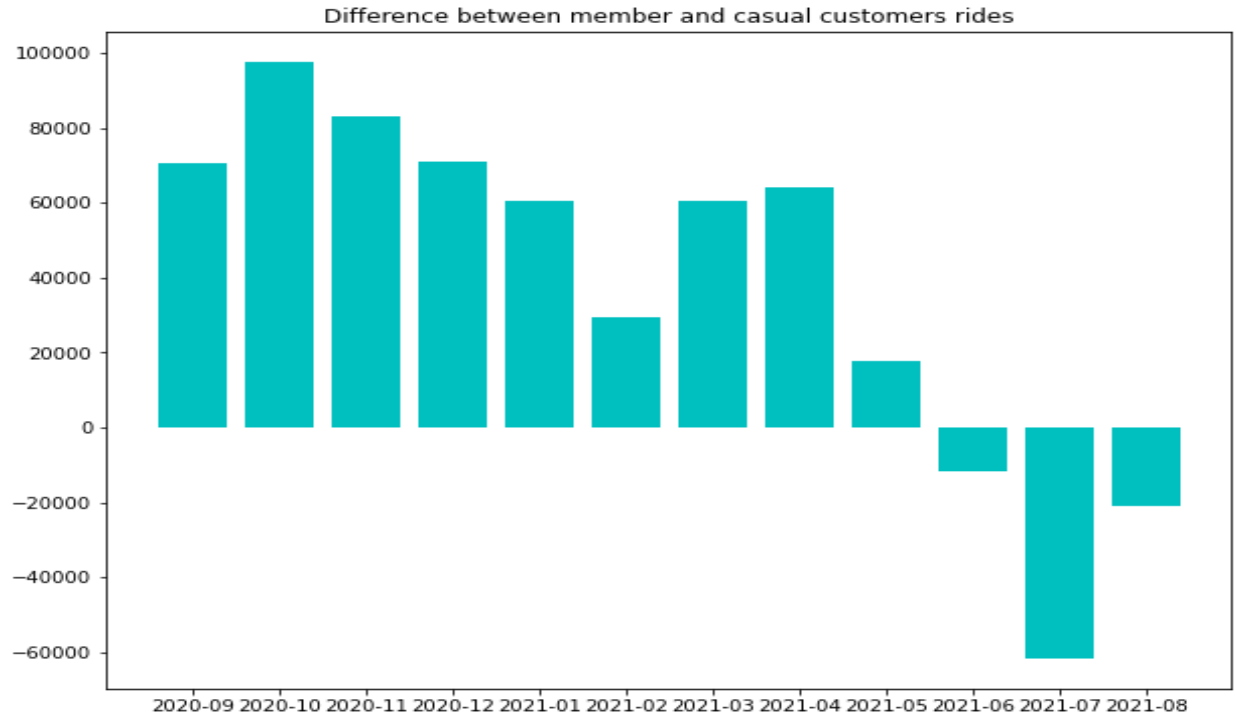


3.2 Analysis rides against month

The business whole trend is increasing, although there is bad performance during the first half year.

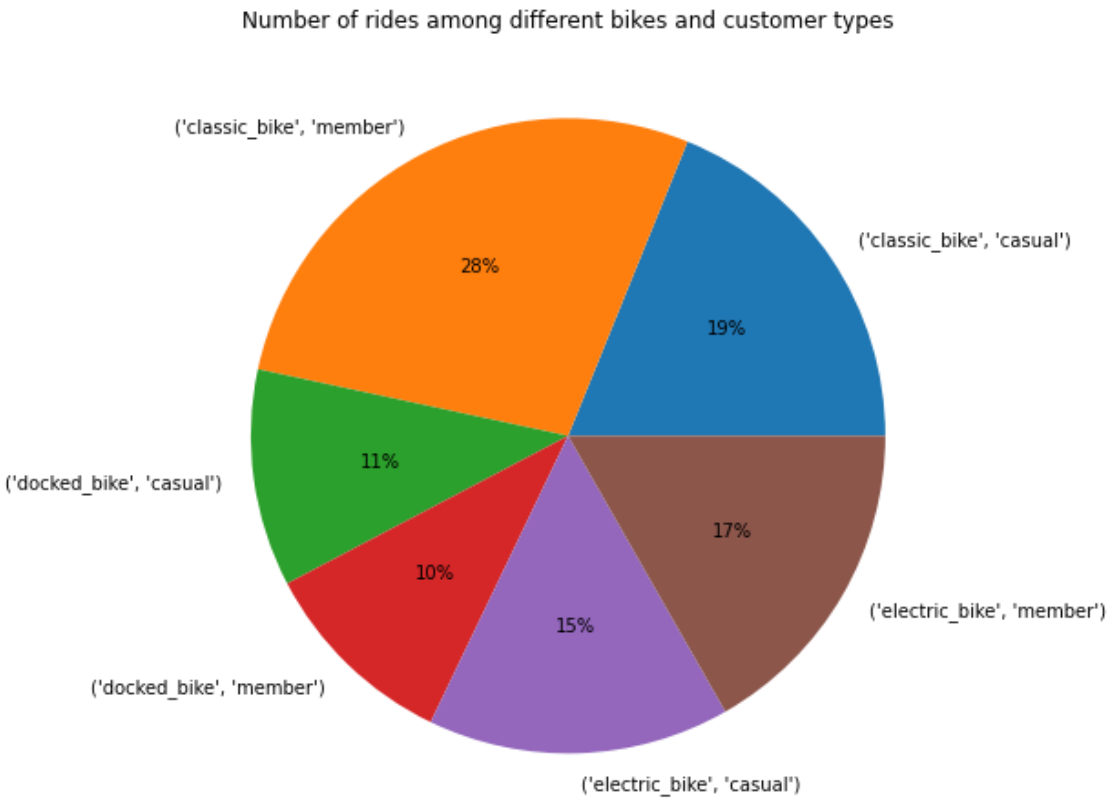


But rides contributed by member customer can't increase as fast as casual ones

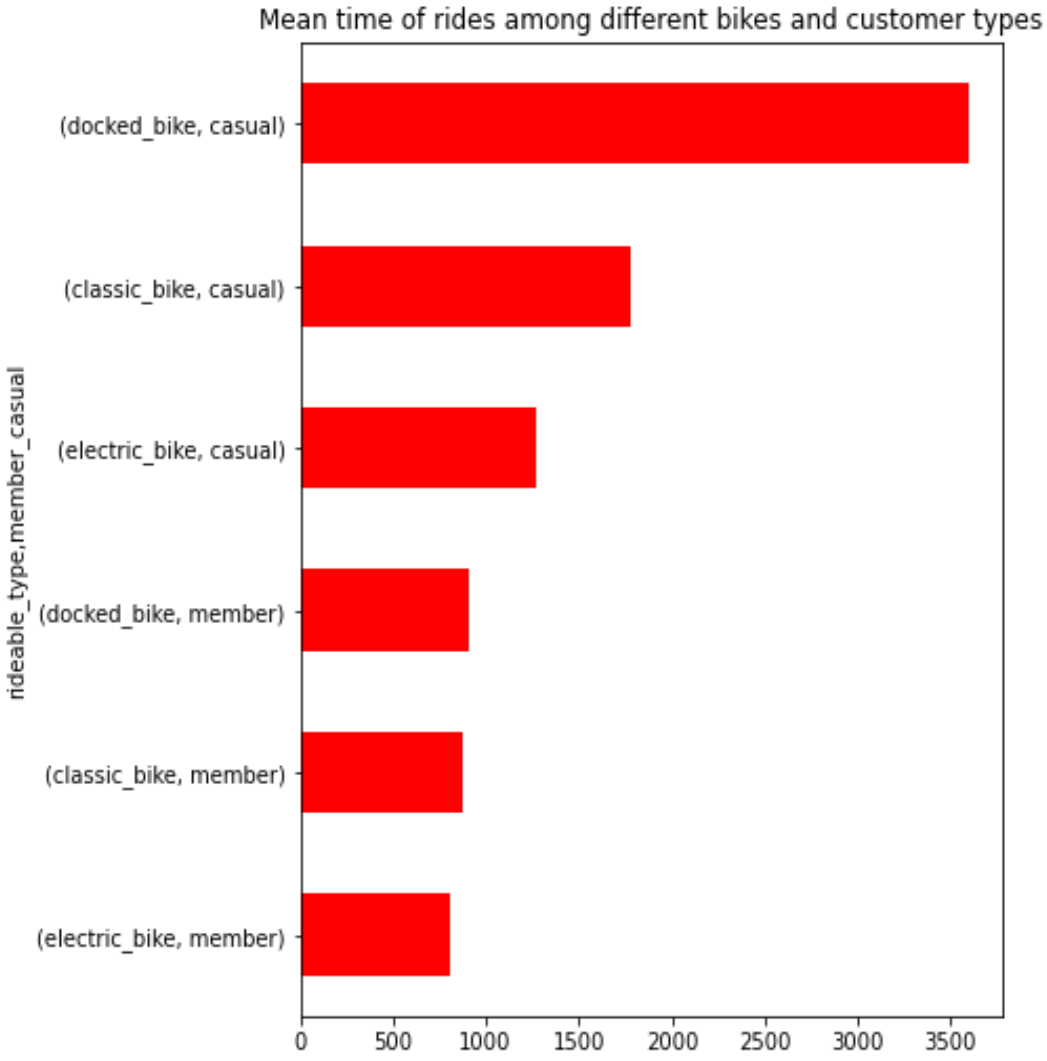


3.3 Analysis among bike and customer types

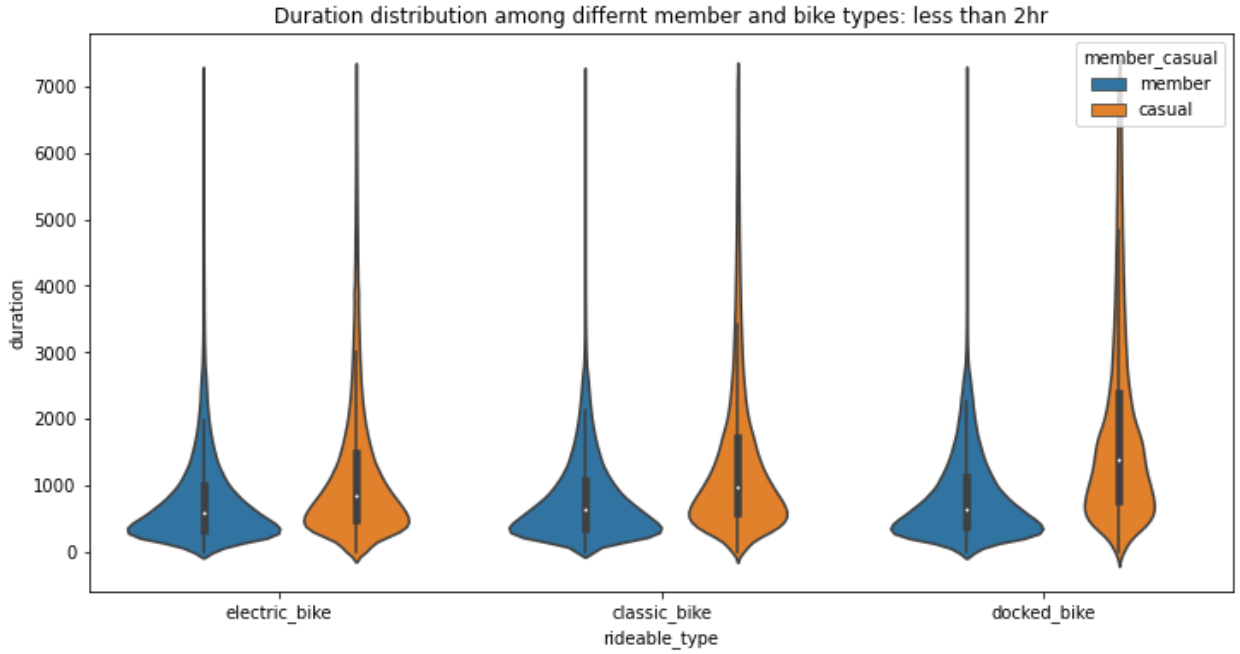
Classic bikes contribute main parts of business volume, no matter membership owners or casual customers.



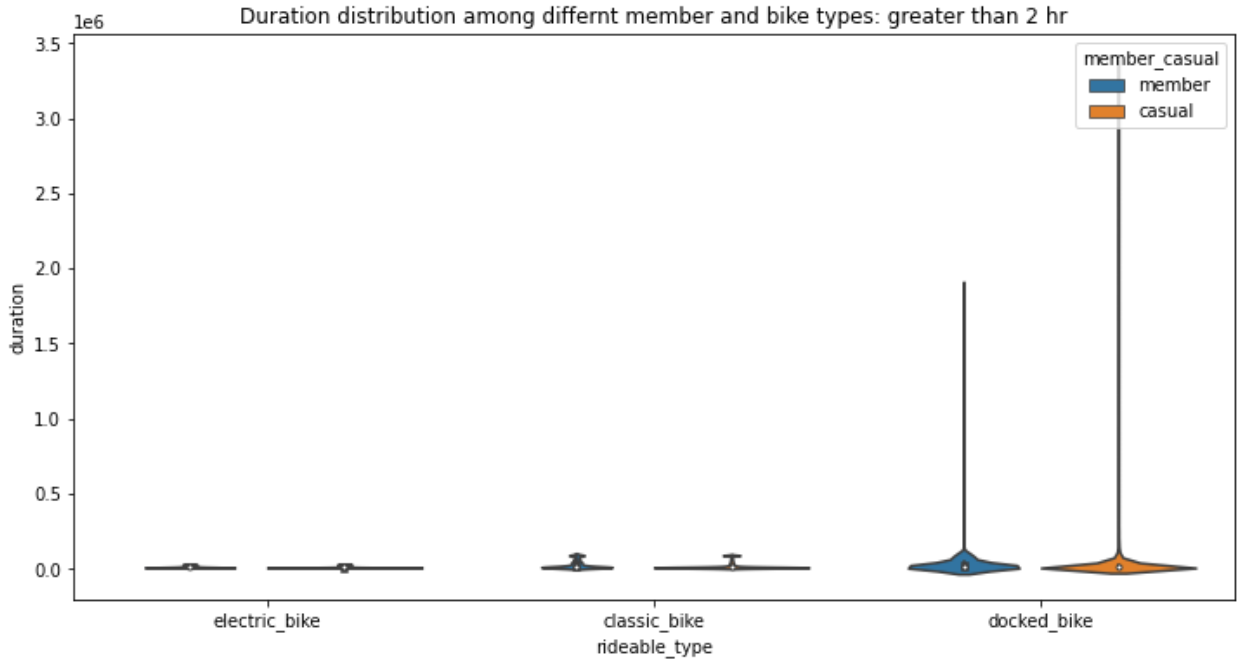
Casual customers usually take more time during one ride.
Docked bikes usually take more time than other two types.



More clear distribution of time of rides. Member owners usually presented more reasonable ride activity.



Docked bikes dominated long time rides(more than 2 hour), especially for casual customers.



4 Conclusions and future suggestions.

4.1 From the analysis above we can get the insights

4.1.1 The increasing ratio of rides contributed by member type customers is slower than the other one. We need to focus on converting casual customers to membership.

4.1.2 Casual customer's duration usually takes more than 2 times than membership, which may affect the company's financial performance by decreasing bikes' turn-over rate.

4.1.3 Casual customers like use our service on weekend, that's reasonable and can be used as a marketing strategy.

4.1.4 Docked bikes need to pay attention to long time occupation.

4.2 Future suggestions

4.2.1 Financial influence during different rides and customer types.

4.2.2 The unusual reason for the long time occupation of docked bikes.