Quarterly Analysis Report

Background

This report aims to help our shareholders to uncover two phenomena and make some recommendations. The data was sourced from https://www.citibikenyc.com/system-data., and its timeframe ranging from 06.22 to 09.22 was a summer season in America.

Pre-process

The data structure has been changed for few times when I investigated the above link. The following part is how the data has been recorded.

2021.02:
- Add 'Ride id', 'Ride Type'
- Remove 'Bike ID ', 'Duration', 'Birth', 'Gender'
- Time: change to mm/dd/yyyy hh:mm
- Station Id: xxxx > xxxx.xx (More stations)
- User Type=>member_casual

2021.06:
- Add 3 rideable type ('Classic', 'Electric', 'Docked')

To clean and transform the data for correctly analyse the data, I execute the following steps and then combine 4 files into 1 file by using Python3.
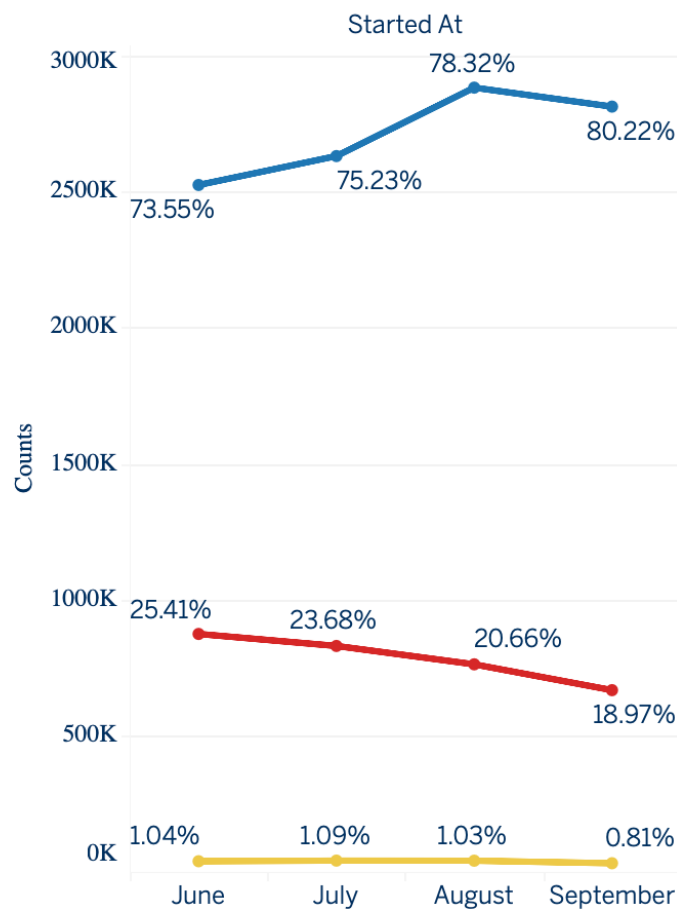
1. Insert "Trip Time" column: The difference between start time and end time.
2. Insert "Status" column:
    a. Valid: Casual users:2~30 mins + Member: 2~45 mins.
    b. Overtime: Casual users:30 mins~1day + Member: 45 mins~1day.
    c. Timeout Error: 'Trip Time' less than 0 min.
    d. Unsuccessful Renting: 'Trip Time' between 0~2 mins.
    e. Investigate: 'Trip Time' longer than 1 day. (We need a new policy)
3. Insert "Error Type" column:
    a. Completed: No missing data
    b. GPS Error: Record has latitude and longitude but no end station's id and name.
    c. Data Loading Error: Record doesn't have station's id, name, latitude and longitude.
4. Adjust each stations' id, latitude, and longitude to make sure each station has only one id, one latitude, and one longitude.

First Phenomena: Clients' Preference

To understand our clients' consumption pattern, I am going to analyse the data from several aspects. The purpose of discovering our clients' preference could help us to understand what our customers want and how we allocate our resources more efficiently.

Bike Type aspect

Bike Type Preference

Started At

The summer season in America is from June to August. From the plot, we have found three trends. First, users preferred to rent classic bikes instead of electric bikes or docked bikes. Secondly, even though the total number of rented classic bikes dropped from 2.88 million in July 2022 to 2.81 million in August 2022, the ratio of choosing class bikes still increased from 73.55% in June 2022 to 80.22% in August 2022. Lastly, on average, most users' renting time is 12.72 minutes. It means people would choose classic bikes rather than paying extra money to use electric bikes.

Average Using Time

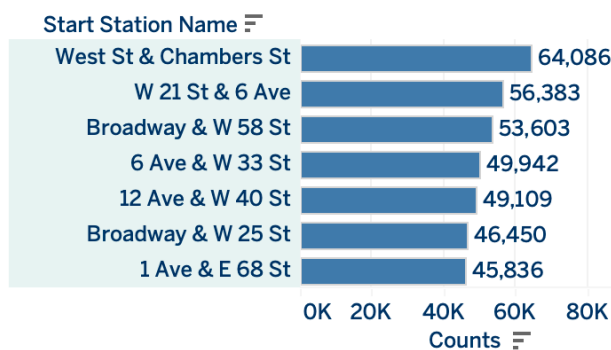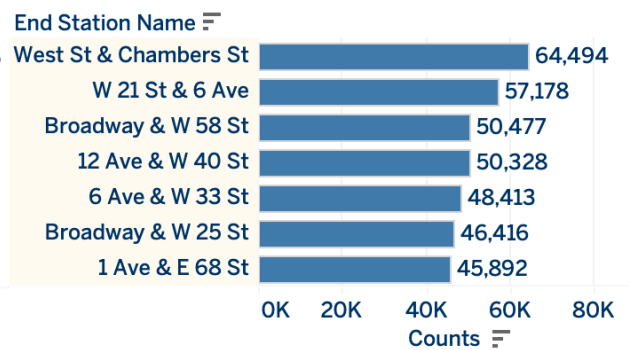| Status	
Valid	12.72
Over Time	72.38

Renting Time aspect

Time Preference
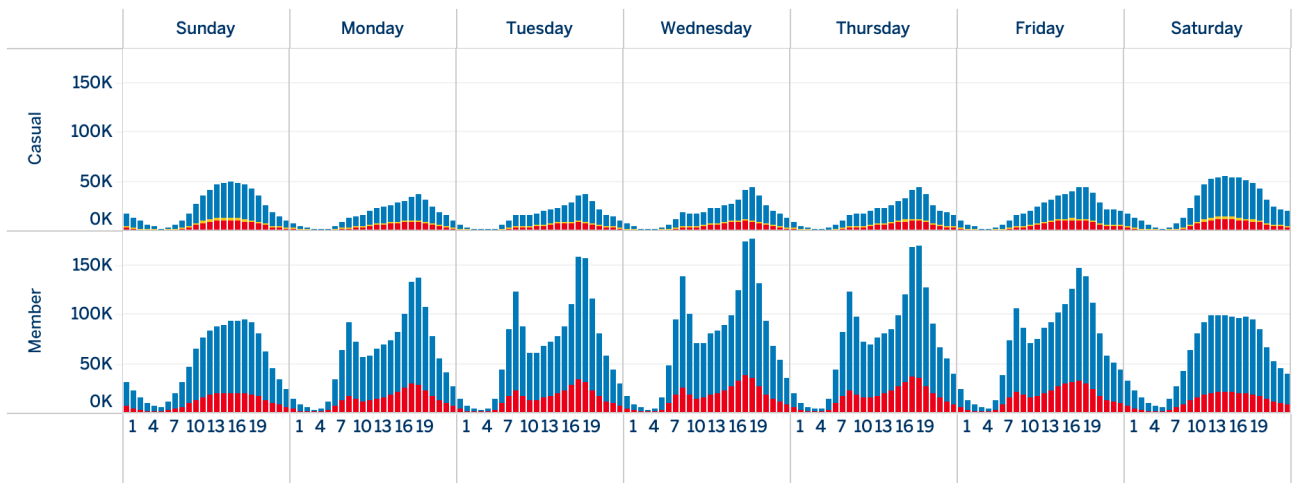
According to the plot above, the casual user group's usage pattern was entirely different from the member group. Casual customers tended to rent bikes on Saturday and Sunday and members tended to do so on weekdays. During weekdays, there are two clear peak usage times at 8am and at 5pm in the member group. On the other hand, most casual customers rented bikes in the afternoon. We could assume that most of our members have stable working time. During the weekend, the distribution of usage time is a normal distribution shape. It may indicate that people living in the New York City area have two-days-off living style.

Bike Station aspect

Top7 Start Station

Start Station Name	Counts
West St & Chambers St	64,086
W 21 St & 6 Ave	56,383
Broadway & W 58 St	53,603
6 Ave & W 33 St	49,942
12 Ave & W 40 St	49,109
Broadway & W 25 St	46,450
1 Ave & E 68 St	45,836

Top7 End Station

| End Station Name	Counts
West St & Chambers St	64,494
W 21 St & 6 Ave	57,178
Broadway & W 58 St	50,477
12 Ave & W 40 St	50,328
6 Ave & W 33 St	48,413
Broadway & W 25 St	46,416
1 Ave & E 68 St	45,892

These two bar charts show the top 7 bike stations where our clients liked to start and end their journeys. We can see that the station list from both was the same and even the visit times were similar. As these stations had high usage volume, arranging regular maintenance for these stations' software and hardware equipment must be prioritized.

Second Phenomena: Error Analysis

In the Pre-process section, I mentioned that I adjusted each stations' id, latitude, and longitude to make sure each station has only one id, one latitude, and one longitude. The reason I did the step is that I found out one bike station has multiple ids, latitudes, and longitudes. It may cause trouble if we want to investigate a specific station or do further research in some areas by using station id, latitudes, and longitudes. Therefore, I present the error part I found in the data set as another important issue.

Top 10 counts of errors by station
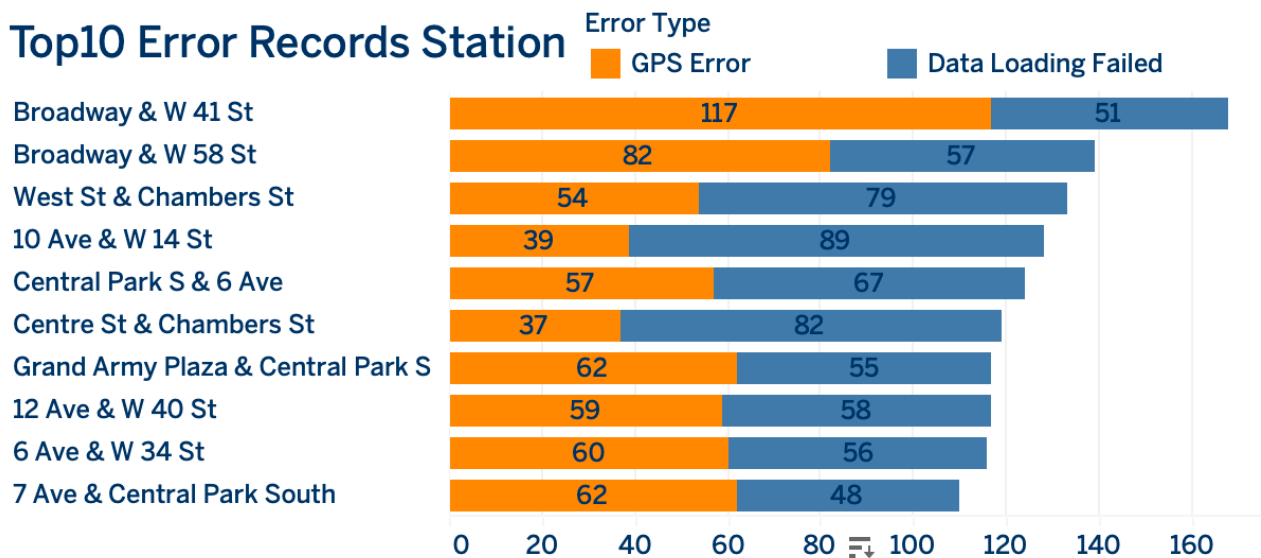
Compared to the previous slide, only three out of seven top- usage-volume stations were in the list. It means we need to launch a procedure to find out what happened on these stations. Do we need to upgrade the stations' equipment?
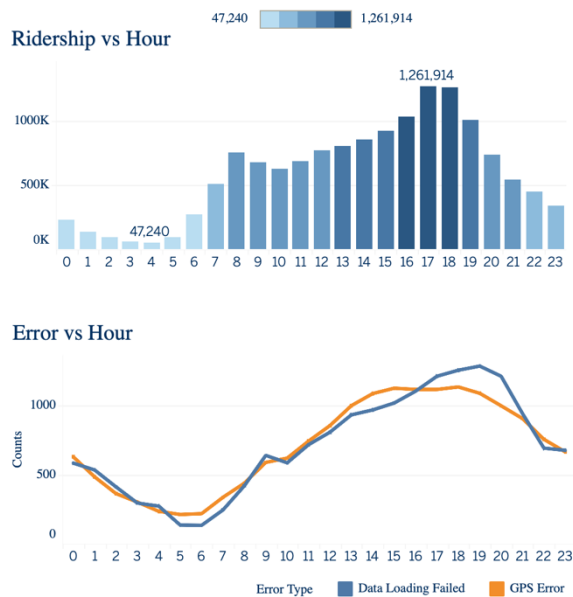
Error vs Hour

Error vs Hour

These two charts help us to understand how data errors came out. The bar chart shows in these 4 months' time frame, how many numbers of our clients used our services in each hour. According to the chart, there were around 126 million users at 5 pm and 6pm respectively.

From the line chart, based on the same hour-distribution, we have found out errors often triggered between 6pm to 7pm and it happened an hour later when the number of users peaked.

Based on two analyses from two paragraphs above, we should upgrade our systems and equipment as our system now cannot process efficiently that high amount of input and output data during which around 78 thousand (126 million users in 4 months) users flow into bike stations to return their bikes in an hour.

Status Analysis

Although the number of 'Investigate', 'Timeout Error', or 'Unsuccessful Renting' is not that many compared to the total number of records, we still need to find a way to decrease the number as many as we can.

| Status	
Investigate	17,879
Timeout Error	436
Unsuccessful Renting	770,619

In the 'Investigate' section, the record of the trip time is longer than 24 hours would be listed in it. CityBike has a clear rule that restricting users should not use more than 30 minutes per trip (45 minutes for our member). Renting a bike for more than one day is an intolerable behaviour as it will be a burden to our business's operation and impact our other customers who abide by our rules.

In the 'Timeout Error' section, the record of the trip time is less than 0 would be listed in it. It is great that we only had 436 records as our system processes huge data every day. However, we might be able to achieve the '0' target by improving the signal strength and the connections between each station.

In the 'Unsuccessful Renting' section, the record of trip time between 0 to 2 minutes would be listed in it. There are many reasons that may cause this situation. My recommendation is recruiting more technicians to do more regular service on dock machines and bikes. Therefore, people might not need to redock a bike and choose another bike which has better conditions or other factors force our customers to dock in and out in 2 minutes.

Popularity Of Station For Starting A Journey
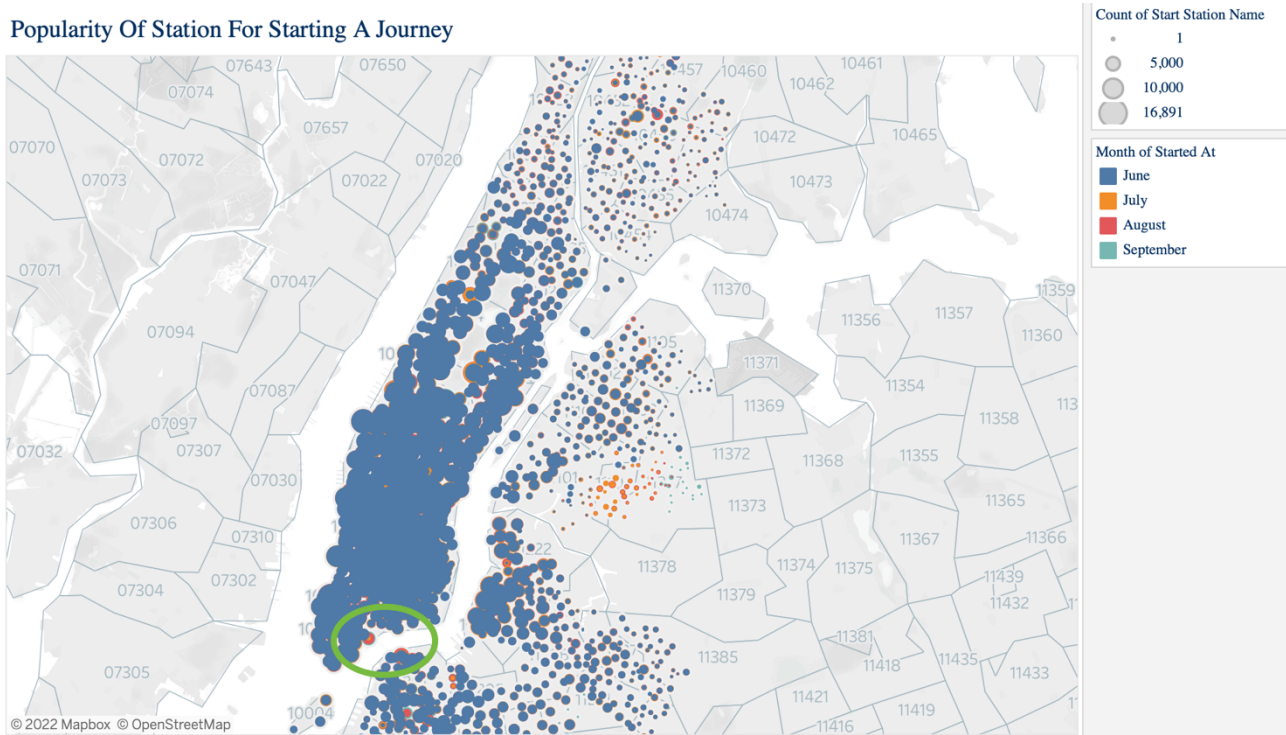
The most popular origins are stations near large transit and tourist hot spots, including the stations near Central Park, Penn Station (10001) and Union Square (10003), and the area around Union Square in Manhattan. After using colours to group our data by the months, we can see that our service area is spreading out in The Bronx, Queens, and Brooklyn.

Highlights:

1. The number of users in Sunnyside (11104) and Woodside (11377) increased significantly and gradually grew in The Bronx.

2. The bike station id (5096.12) in Peck Slip (10038) and station id (4936.14) in Brooklyn (11201) were growing month by month. The reason might be the combination of ferry and bike provides our users a commute with money-saving and time-saving between Manhattan and Brooklyn.

Popularity Of Station For Ending A Journey
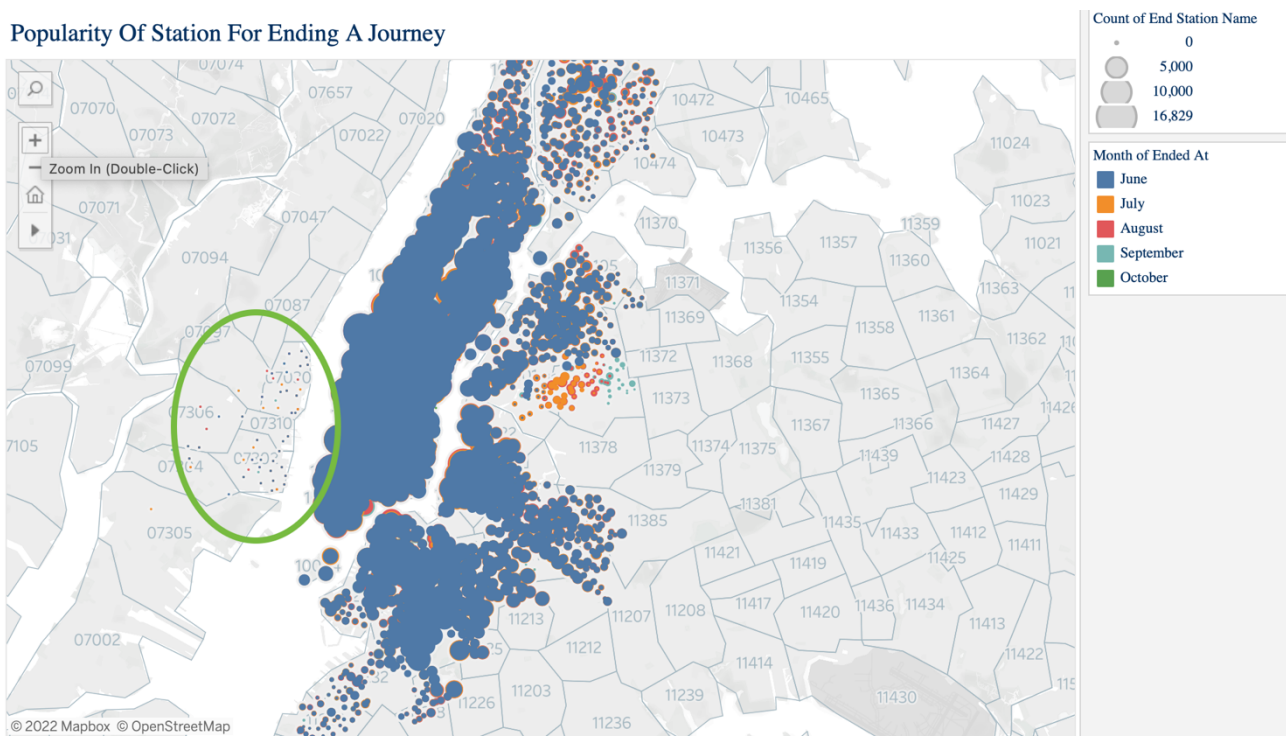
As I mentioned in the slide "Station Preference", basically, the popularity of stations for users to end their journeys is pretty much the same. And compared to the previous slide, the destination part also inherited the same business's growing pattern.

However, there is one thing worth diving into. There are small numbers of users who would return their bikes in the Jersey City area. At the moment, Citybike has two bike-sharing projects, one in the New York city area and another one in the Jersey city area. It may indicate there is an unknown consumption pattern which these data sets cannot support an explanation.

Conclusion

Overall, the business is gradually growing, the number of bike stations is increasing and the area of using our services is expanding. It is a good sign for our company. However, we still need to plan ahead to deal with unpredictable situations that happen when the business scales up.

I have made three recommendations. Firstly, comprehensively upgrade our network systems and hardware equipment to provide our customer better service quality. Secondly, modifying our existed renting rule to stop people rent our bikes more than 1 day a trip, which could help us cost down efficiently. Lastly, we could build up a relationship with the ferry company or bus companies to strengthen the importance of our bike business in the traffic network.