JSP

Jeff Pinegar

jeffpinegar1@gmail.com

717-982-0516

Assignment: Tableau – Citi Bike Challenge

Due Date: January 13, 2023

Tableau – Citi Bike Challenge

Executive Summary

I conducted an exploratory analysis of the ridership of Citi Bike in NYC. This consisted of Building a data frame of 2022 rides and then exploring that data to see if there were patterns of phenomena of interest. Two items of interest eventually revealed themselves (1) patterns of ride duration and (2) the sub-category of round-trip rides.

The patterns and duration of rides

The patterns and duration of rides reveal the best time of day to perform quick preventive maintenance on bikes at the stations and the best days to take bikes out of circulation for more elaborate preventative maintenance.

Analysis of Round-trip rides

The analysis of round-trip rides revealed possibilities for improvements to our billing for casual users and perhaps the creation of a "health membership class."

Presentation of findings

Two dashboards were created, one for each "item of interest." Following are the details behind the creation of these dashboards and the findings.

Table Contents

Data and Data Preparation

Data Selection

For this exercise, I am analyzing the data from 2022. While trends over time can help anticipate the future, at this time, I felt it was better to understand the present. Then, if necessary, I would return to the past to see if it could foretell the future.

Furthermore, this business, at this time, has very little inertia. Therefore, I don't need to predict very far into the future to keep the company on track. Moving kiosks and adding bikes in response to changing demands is relatively easy. The most effective way to shrink the pool of bikes is to wait and take them out of circulation when they need repair.

Data Source: https://citibikenyc.com/system-data

Data Preparation

The 2022 data set includes a monthly file, each with more than a million rides each—way more data than can quickly and efficiently be analyzed. More importantly, way more data than is needed for practical analysis.

I randomly selected 1% of the data from each month using Python pandas. Then, the selected data was put into a single data frame, cleaned, and transformed the data frame in the following ways.

- Step 1: Basic collection and screening
    - Dropped incomplete records (this was done before the selection of 1%)
    - Converted the string representation of start and end times to datetimes
    - Calculated the duration of a ride in Seconds and added that to the data frame
    - Determined the Day of Week rides started and added it to the data frame
    - Determined if the start location and end location using the were the same using the station ID and added a Round-Trip (T/F) column to the data frame
    - Rides of negative duration were eliminated. We are not focused on the time-traveling public but will monitor this for future opportunities.
    - Based on the station ID and the start and end location data calculated, a single location for each station. I merged this data into the data frame.
- Step 2: Establish Station location. Records contain where bikes started and stop and station IDs. I used this to calculate Station location. Station locations were calculated by averaging longitude and latitude data from every ride that started on ended at the station.
    - Calculated the station location information
    - Used Excel to add this to each record as x_station_lat and x_station_lng where x is either start or end.
- Rides of less than 1 min and longer than 24 hours duration were eliminated as outlines or errors

Once cleaned and transformed, this data was exported as a CVS file to be ingested by Tableau.

Patterns and Duration of Rides

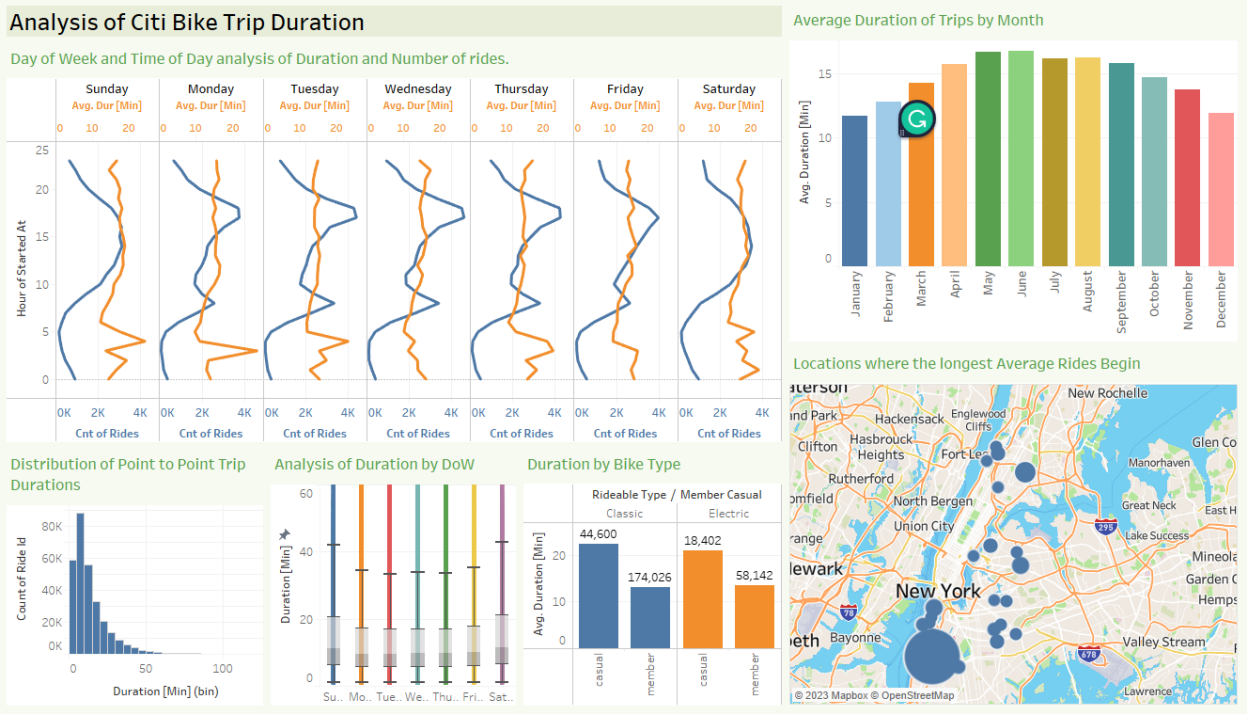

Figure 1: Analysis of Citi Bike Trip Durations

The dashboard created is shown in Figure 1. The link to this dashboard is:
https://public.tableau.com/app/profile/jpinegar2023/viz/City_Bike_Challenge_Posted_02/DurrationDashBd?publish=yes

Patterns through the Week

Weekday and weekend rides have very distinct by different patterns; see Figure 2. Weekday rides have very shape peak ridership around 8 AM with another extended peak between 4:30 and 7:00 PM. These peaks follow a traditional workday schedule. The afternoon peak is extended because, in addition to commuters, it includes others who are up and active in the afternoon but not in the morning.

The weekend profile is almost a mirror image of the weekday. The activity starts picking up gradually around 9 AM, builds to about 4 PM, then gently declines.

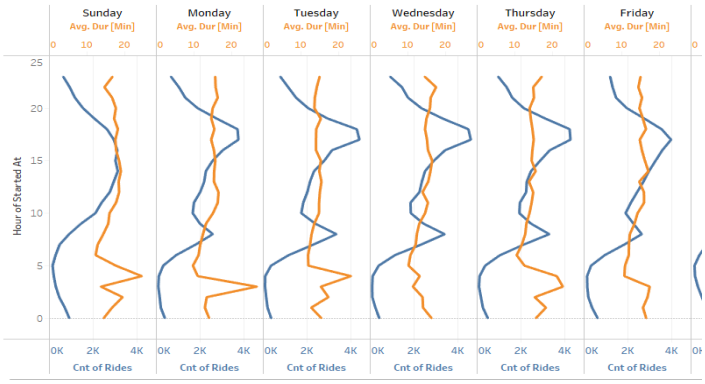

Figure 2: Daily Patterns

As for the duration of travel, the average trip is between 10 and 20 minutes. The only significant variation in this data is between midnight and 6 AM. The underlying cause for this variation still needs to be investigated. The working theory is that a measurable number of bikes are taken out at this time and abandoned.

Frequency of rentals by time-of-day and day-of-week

While not directly tied to the duration of travel, this usage heat map presents a clearer view of the count of riders and peak activity, as shown in Figure 2. This heatmap makes it much more apparent that 3:30 PM to 8:00 PM represents the highest rate of bike utilization.

Count of trips by Time of Day

Hour of Started At	Sunday	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday
0	904	386	357	411	484	613	977
1	701	203	194	236	260	363	670
2	447	152	132	96	172	192	496
3	290	89	80	88	84	145	312
4	199	113	87	109	111	136	180
5	144	311	353	386	364	316	159
6	282	901	1,151	1,121	1,041	939	380
7	463	1,778	2,275	2,312	2,113	1,819	710
8	896	2,569	3,328	3,387	3,280	2,624	1,061
9	1,440	1,984	2,364	2,424	2,448	2,244	1,811
10	2,094	1,655	1,729	1,847	1,920	1,886	2,456
11	2,374	1,711	1,821	1,832	1,946	2,130	2,795
12	2,747	1,994	2,005	2,240	2,231	2,412	3,200
13	2,928	2,178	2,128	2,349	2,257	2,702	3,357
14	3,114	2,254	2,307	2,495	2,450	2,950	3,470
15	3,006	2,561	2,778	2,898	2,866	3,253	3,369
16	3,092	3,026	3,128	3,409	3,308	3,560	3,309
17	2,932	3,734	4,375	4,578	4,245	3,958	3,129
18	2,620	3,696	4,256	4,468	4,210	3,529	2,854
19	2,053	2,757	2,946	3,314	3,115	2,754	2,476
20	1,530	1,881	2,044	2,329	2,122	1,945	1,849
21	1,162	1,317	1,491	1,720	1,596	1,454	1,364
22	924	1,029	1,135	1,445	1,343	1,297	1,280
23	632	668	778	911	960	1,187	1,184

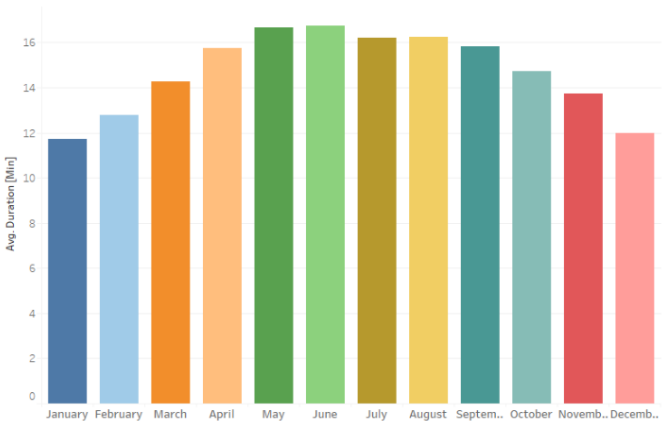

Figure 4: Seasonal Variation

Figure 3: TOD, DOW heatmap

Seasonal Variations

While it is intuitive to assume that more rides would be made during the summer months, it also turns out that the rides are either further (greater distance) or done at a more leisurely pace. The average duration ranges from a low of less than 12 min in January to more than 16 min from May to August, see Figure 4.

Distribution of Duration

Drilling further into the duration data, we see that the vast number of trips are 10 min or less; see Figure 5 below. This would suggest two possible pricing strategies one would maximize revenue, and the other would maximize ridership (getting people off the roads a subway).

To maximize revenue, price with a relatively high flat fee for a ride and a low fee for the duration ($4 plus $0.20 per minute). To maximize ridership, have no flat fee and change for the usage ($0.30 / min). In all cases, I would recommend a time-based fee to encourage people to return the bike to circulation.

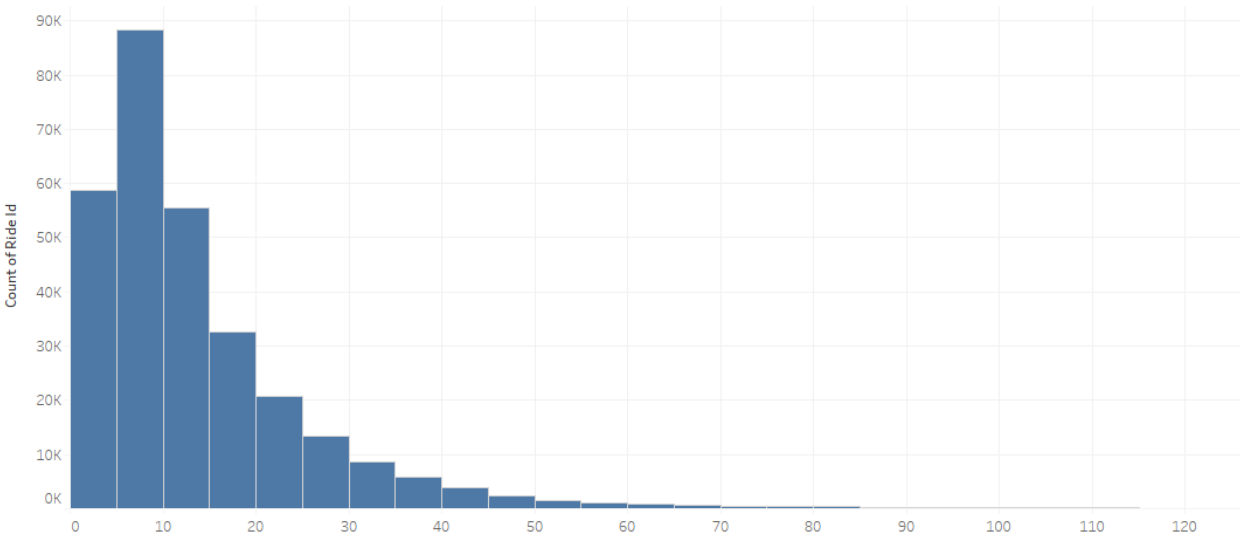

Figure 5: Distribution of Duration

Duration variation by DOW

Examining the duration of trips even closer, we see the duration of trips are almost identical Monday through Friday (Figure 6), with longer trips on Saturday and Sunday. While the average and median duration of trips are relatively stable and short, every day of the week has many outliers with much longer duration.

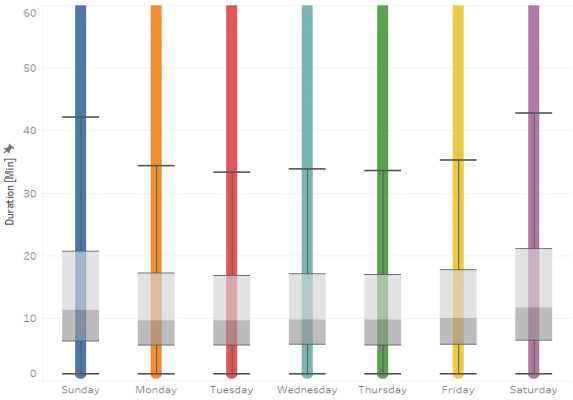

Figure 6: Duration Variation by DOW

Members vs. Casual Riders

Examining membership and bike type, we see two interesting things. First, (Figure 7), notice that average members take shorter trips. We attribute this high usage short duration to commuting to work. While members indeed use the bike for other activities, when we look at the average, it is overwhelmed by short-duration trips to work (or similar routine trips).

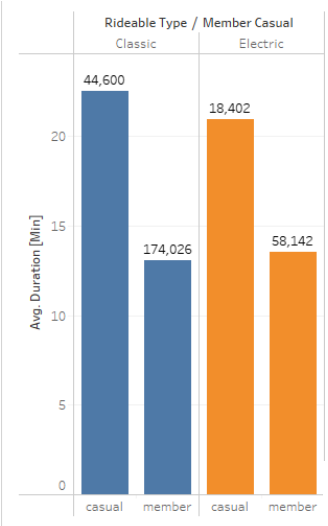

Figure 7: Member vs Casual Usage

The second and surprising finding is that the selection of bike type (classic vs. electric) has very little to do with the duration of the trip! This needs further investigation. It may simply reflect the availability of bikes or the condition of the available bikes.

Analysis of Round-trip Rides

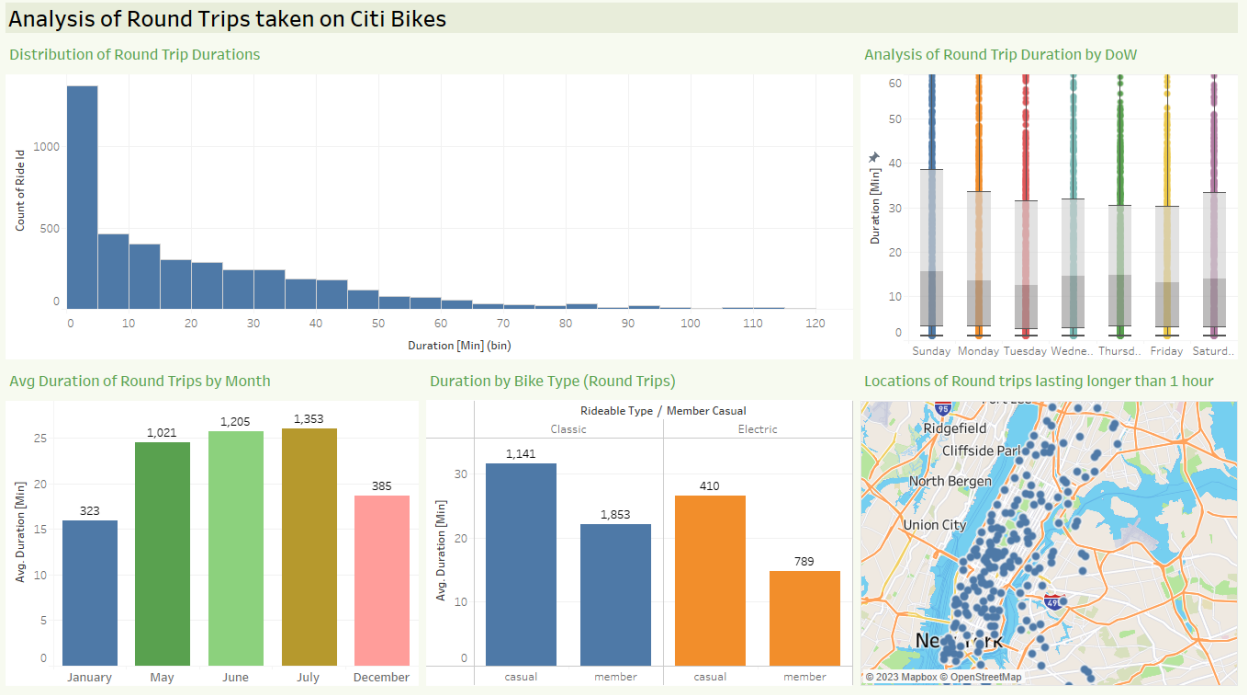

Figure 8: Analysis of Round-trip Rides

The dashboard created is shown in Figure 8. The link to this dashboard is:
https://public.tableau.com/app/profile/jpinegar2023/viz/City_Bike_Challenge_Posted_02/DurrationDashBd?publish=yes

Distribution of Round-trip Durations

The distribution of round-trip durations points to the most significant factor in round-trips … "I forgot something." The vast majority of the trips are less than 5 minutes, even less than 2 minutes; see top left of Figure 8. This would suggest examining the pricing policy for trips less than 2 min to determine the best strategy for dealing with these customer errors.

Alternative Views:

- Rides less than 2 minutes are free. This will avoid customer frustration and anger. Anger can take the form of bike vandalism or feelings of "what a rip-off, I am never using Citi Bike again."
- Free rides of less than 2 minutes will lead to frivolous rides, resulting in less availability for those who want or need the bike for a longer distance.

Seasonality of Round-trip rides

The vast majority of round-trip rides take place in the summer months (Figure 9). This seasonality also suggests these trips are discretionary (exercise, pleasure, etc.) and not mission-driven (going to work or running an errand.

Discretionary also means that they can be influenced by pricing. This suggests we could increase ridership by offering low-cost round-trip rides from August through April. Furthermore, we could ship the demand in time by offering discounts for start times between 10 and noon on the weekdays and before 9 AM on Saturdays and Sundays.

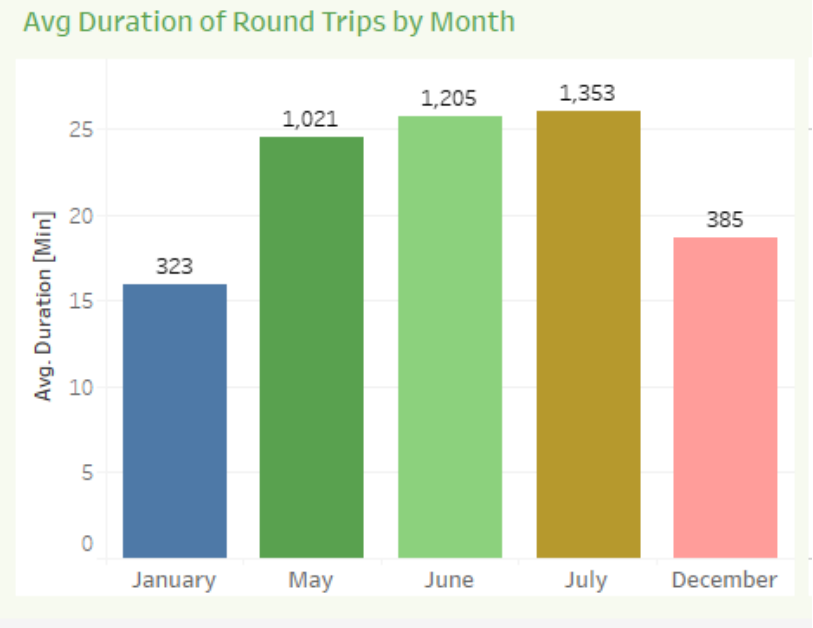

Figure 9: Round Trip Rides

Bike Type for Round Trips

Looking closer at round-trip riders and the bikes they choose (Figure 10) reinforces the hypothesis that there are two purposes for these trips (1) exercise and (2) pleasure. From the count of members vs. casual, we can assume the members have a lower threshold for usage. For example, members may only have 20 minutes, but since it does not cost them more, they take a ride. While the casual user decides there is not enough value for that 20 min ride.

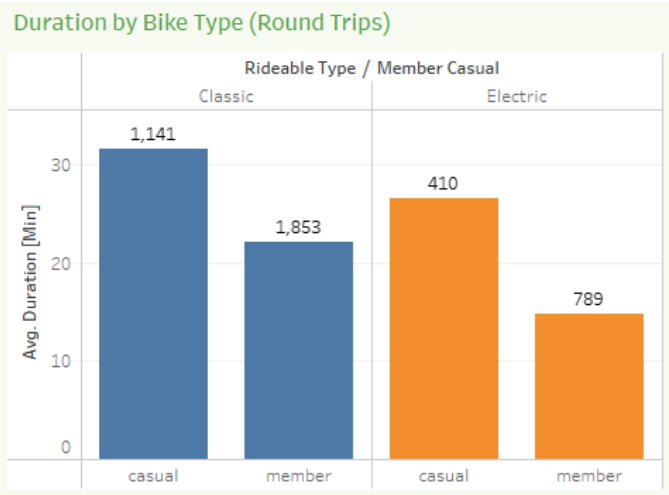

Figure 10: Bike Type for Round trips

Conclusions and Action Items

Based on this investigation of the duration and timing of rides in general and the round-trip riders, I believe several opportunities for growth and efficiency should be investigated.

- Develop a strategy for increasing the round-trip ridership. The elements of this program should consider.
    - A round-trip-only membership
    - Structures to encourage riding outside of the peak usage hours
    - Perhaps an awards program to promote longer rides. This is good for social health and could grow the business by eating into the market share from health cubs.

- Optimize the bike maintenance plans based on the natural variation in usage by time, day, and season. For example:
    - Frequent light on-site maintenance and inspection could be performed between the mid-day peaks
        - Oiling the chains
        - Check tire pressure
        - Collecting damaged bike
        - Relocating bike to where surges will occur
    - Periodic maintenance in-shop maintenance during the overnight hours from 9 PM to 6 AM.
        - Replacement of tires
        - Realignment of bent wheels
        - Checking brakes
        - Replacing worn/damaged seats or hand grips
    - Inventory optimization (major repair or replacements) during late fall, October to November, and early spring, February to March.
        - Repainting
        - Repairing damaged frames
        - Catching up on the bikes that were removed for periodic maintenance, which turned out to be more involved
        - Removing bikes from circulation that may be frail (decreasing demands on light maintenance and inspection.)