**Citi Bike Analytics**

The data period is between July 2019 and December 2019.

From the monthly trip data, we can see the number of trips reached the highest by the end of summer (September) and went down since temperature dropped down. This indicates that people prefer to ride bikes during summer.

Most of the Citi Bike users are subscribers (annual members) instead of customers who hold 24-hour pass or 3-day pass.

After analyzing peak hours during summer (July, August, September) and winter (November, December), we can conclude that peak hours are 8 am and 5~6 pm, matching with regular traffic rush hours.

By analyzing the total number of trips, the most popular start stations are:

1. Grove St PATH

2. Hamilton Park

3. Harborside

4. Sip Ave

5. Newport PATH

6. Columbus Dr

7. Marin Light Rail

8. Newport Pkwy

9. Warren St

10. Newark Ave

The least popular start stations are:

1. JCBS Depot

2. Jackson Square

3. Union St

4. Dey st

5. Leonard Gordon Park

6. Communipaw & Berry Lane

7. Christ Hospital

8. Riverview Park

9. Glenwood Ave

10. Hilltop

The most popular end stations are:

1. Grove St PATH

2. Hamilton Park

3. Columbus Dr

4. Harborside

5. Newport PATH

6. Sip Ave

7. Marin Light Rail

8. Newport Pkwy

9. Newark Ave

10. Warren St

The least popular end stations are:

1. 6 Ave & Canal St

2. 6 Ave & W 33 St

3. Barrow St & Hudson St

4. Bayard St & Baxter St

5. Broadway & Battery Pl

6. Broadway & E 14 St

7. Broadway & W 56 St

8. E 16 St & 5 Ave

9. E 53 St & Madison Ave

10. E 56 St & 3 Ave

According to the gender data, male users take almost three times Citi Bike trips to female users. The time trend of each gender matches the monthly conclusion that highest usage is in September and then goes down as winter comes.

After calculating the female trips percentage by using total number of trips taken by female users divided by total number of trips, we can see female users prefer to ride during summer and autumn. The percentage reaches the highest in October.

Since the data has not recorded the distance of each trip, I make an assumption:

The distance of each trip is the distance between the start station and the end station. Since only coordinates are recorded, I used the formula:

distance = 3963 \* arccos[(sin(lat1) \* sin(lat2)) + cos(lat1) \* cos(lat2) \* cos(long2 – long1)]

lat = latitude / 57.29577951

long = longitude / 57.29577951

The conclusion is: The more times a bike is used, the higher total trips distances will be.

But the assumption is not perfect, there must be some trips like riding a bike to go for grocery and returning it to the same or nearby station. So, the analysis regarding distance might not be as robust as the gender analysis and station analysis.