

Introduction

The purpose of this report is to visualize historic rental bike data in New York City, New York in order to search for trends that can lead to actionable insights. In a large city such as New York a diverse array of transportation options will always be necessary to serve the population and keep traffic congestion to a minimum. Bike rentals are one option of many and should be encouraged because of the availability of bike lanes and the benefit of being an environmentally friendly selection.

The Tableau story can be viewed at the link below (A login to Tableau public may be required):
<https://prod-useast-a.online.tableau.com/t/bethanylindberg/views/Citibike/CitiBikeStory>

Description of the Data

The citibike data was downloaded from <https://www.citibikenyc.com/system-data>. For this analysis the time period chosen is January 2018 to June 2019. Parameters are listed below.

Trip Data:

Tripduration: integer representing total length of trip in minutes

Starttime: Start of trip date and time e.g. 1/24/2018 7:31:33 PM

Stoptime: Stop of trip date and time e.g. Stoptime1/24/2018 7:41:08 PM

Start Station Id

Start Station Name

Start Station Latitude

Start Station Longitude

End Station Id

End Station Name

End Station Latitude

End Station Longitude

Bikeid

Usertype: User type for this trip (Customer or Subscriber)

Birth Year: User birth year if known

Gender: Integer code for gender

Date: Date of trip

Gender Text: Gender string

Daily Ridership:

Date

Trips over the past 24-hours (midnight to 11:59pm)

Miles traveled today (midnight to 11:59 pm)

Total Annual Members (All Time)

New Members: Difference between prior day numbers and current entry

24-Hour Passes Purchased (midnight to 11:59 pm)

3-Day Passes Purchased (midnight to 11:59 pm)

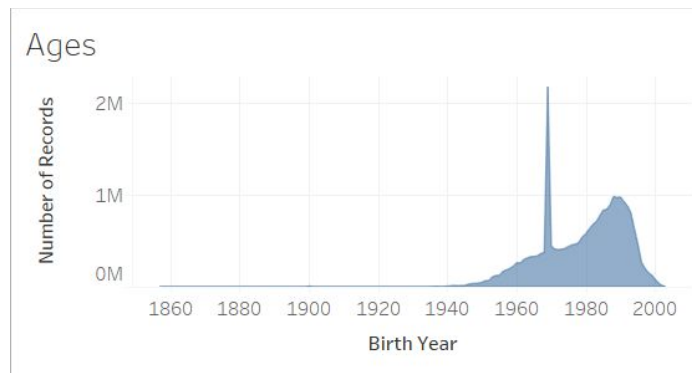
Total Passes Purchased: Sum of 24-hour and 3-day passes purchase

Precipitation and temperature data was downloaded from here:

<http://www.cnyweather.com/wxraindetail.php?year=2019>. Parameters collected included daily accumulated precipitation in inches and daily high and low temperatures in Fahrenheit.

Limitations of the data

As shown in the below area graph, there is an unusual spike of birth years listed at 1969 and several records for every year between 1857 and 1939. These years were filtered out of the demographic analysis.



Methodology

The below aggregation was done in Jupyter Notebooks with Python and Pandas to estimate how many rides are taken for business purposes.

Weekdays comparison:

We can assume that tourists' usage of Citibike stays consistent during the week. So we can calculate average number of rides for Saturday and Sunday to get Recreational usage. We can also calculate average number of rides Monday-Friday. The difference between 2 would be roughly equal to usage for Business purposes.

```
df = pd.read_csv('tripdata.csv')
# Add column for weekdays
df["Weekday"] = df["starttime"].map(lambda x: dt.datetime.strptime(x, '%Y-%m-%d %H:%M:%S.%f') if pd.notnull(x) else np.nan)\
    .map(lambda x: calendar.day_name[x.weekday()] if pd.notnull(x) else np.nan)
df
```

```
# Calculate Number of rides by Weekend and Weekday
weekend_days = df.loc[df['Weekday'].isin(['Saturday','Sunday'])]['date'].nunique()
weekdays = df.loc[~df['Weekday'].isin(['Saturday','Sunday'])]['date'].nunique()

weekend_rides = df.loc[df['Weekday'].isin(['Saturday','Sunday'])]['Weekday'].count()
weekday_rides = df.loc[~df['Weekday'].isin(['Saturday','Sunday'])]['Weekday'].count()
```

```
weekday_daily_average = round(weekday_rides/weekdays,0)
weekend_daily_average = round(weekend_rides/weekend_days,0)

approximate_daily_business = weekday_daily_average - weekend_daily_average

ride_summary = pd.DataFrame({"Avg Daily Sat-Sun":int(weekend_daily_average),\
    "Avg Weekday":int(weekday_daily_average),\
    "Approx. Daily Business Rides":int(approximate_daily_business)},index=[""])

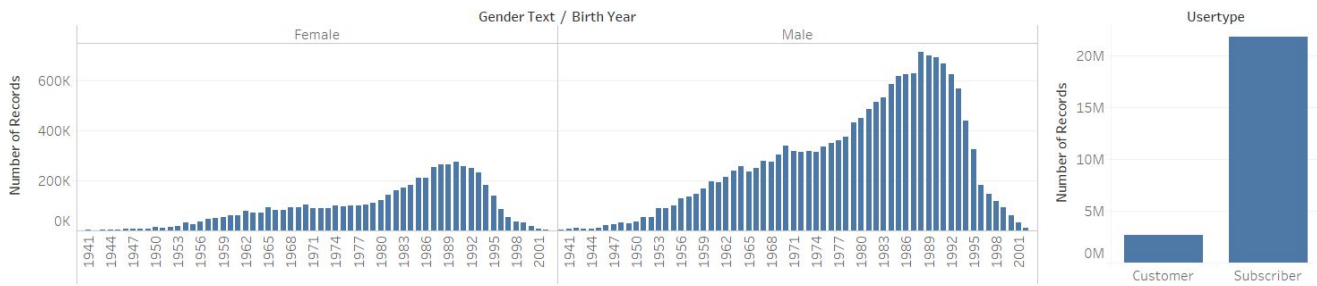
ride_summary
```

Avg Daily Sat-Sun	Avg Weekday	Approx. Daily Business Rides
40116	50327	10211

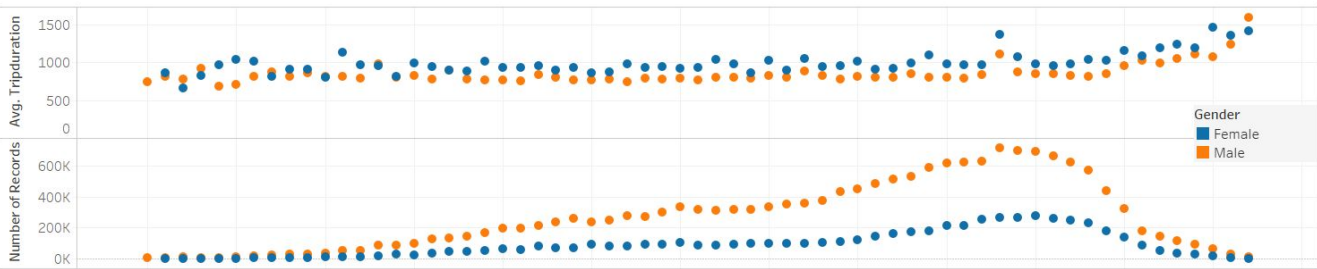
For explanatory analysis the below graphs were produced in Tableau.

Demographics

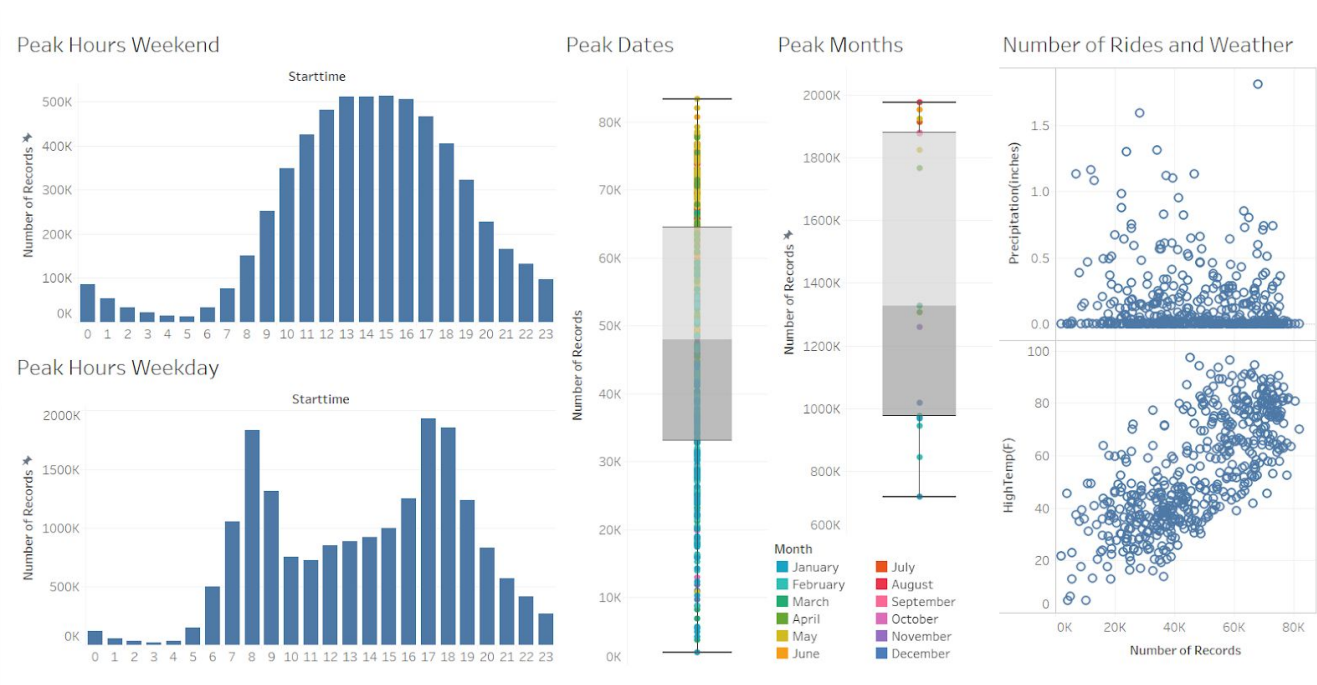
Demographics



Trips by Age and Gender

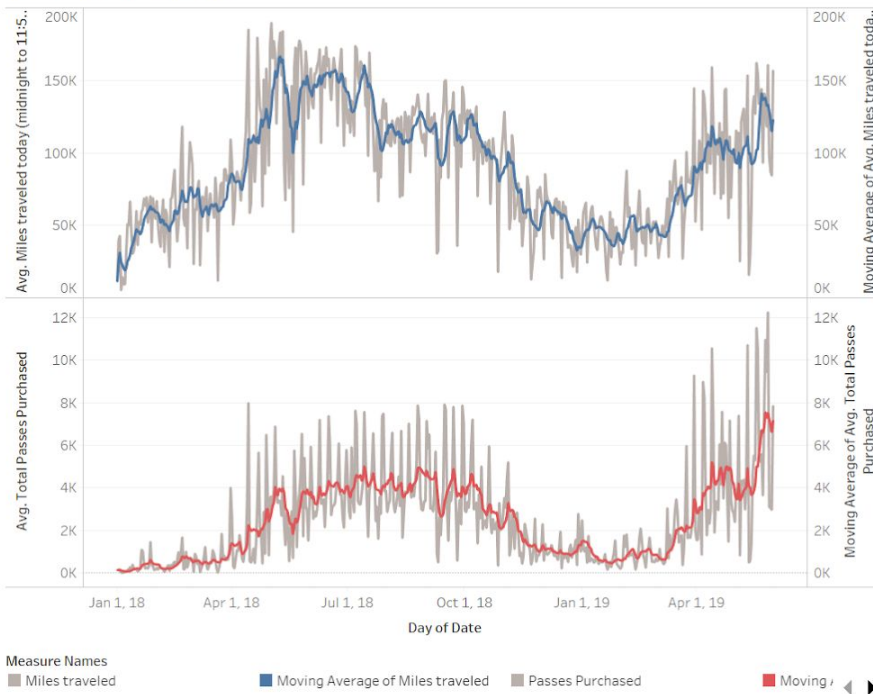


Peak

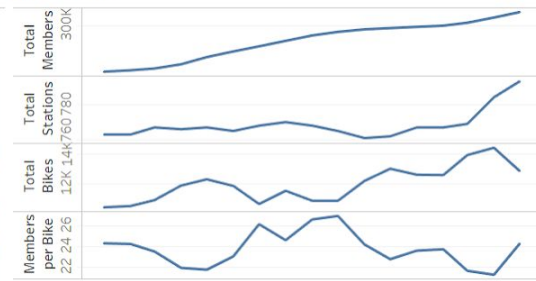


Trends

Passes & Miles Trend



Fleet Lines



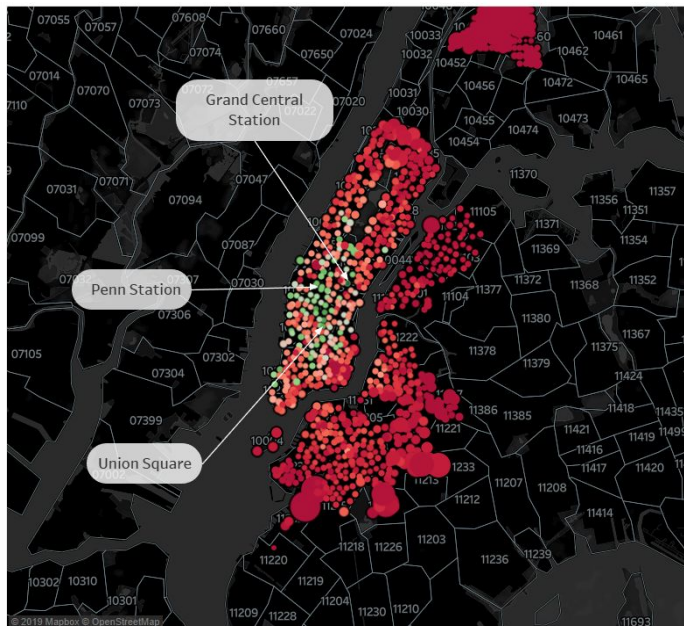
Fleet

Year of Date	Month of Date	Bike Count	Station Count	Total Members	Members per Bike
2018	January	10,449	763	254,539	24
	February	10,534	763	255,944	24
	March	10,935	767	257,701	24
	April	11,913	766	261,874	22
	May	12,335	767	268,998	22
	June	11,889	765	274,628	23
	July	10,672	768	279,702	26
	August	11,561	770	285,087	25
	September	10,892	768	290,467	27
	October	10,887	765	294,007	27
	November	12,228	761	296,497	24
	December	13,046	762	297,748	23
2019	January	12,642	767	298,975	24
	February	12,620	767	300,157	24
	March	13,966	769	303,060	22
	April	14,457	784	308,192	21
	May	12,896	793	313,606	24

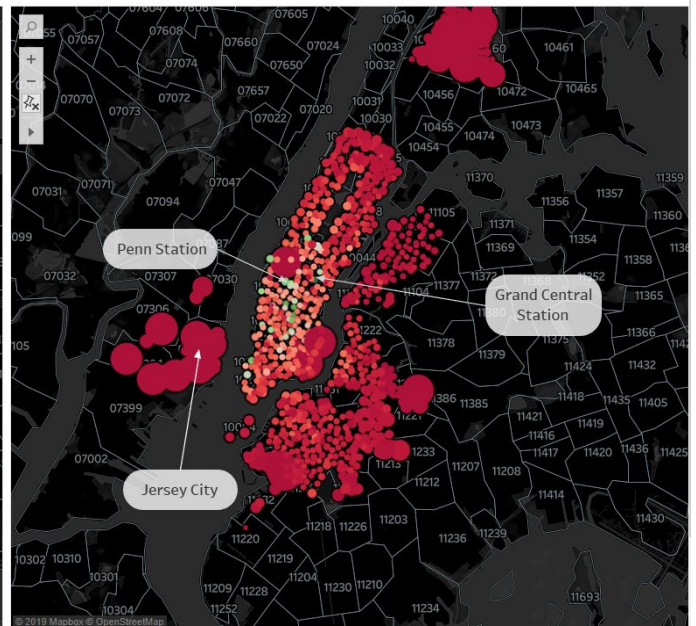
Tableau was also used to produce two sets of maps.

Static map annotated map showing trips over the entire dataset

Static Map - Trip Start



Static Map - Trip End



Dynamic map showing each day of data set (this is a gif)

Dynamic Map - Trip Start - January 1, 2018, Monday Dynamic Map - Trip End - January 1, 2018, Monday



Results of Analysis

Demographic analysis revealed that most of the rides are taken by subscribers with less than 12% of total rides taken by an unsubscribed customer. Distribution of rides by gender and year reveals that there is a similar left skewed trend for both genders with peak birth years between 1980 and 1993. Reviewing trends by gender and age of number of rides and average trip duration shows that male users take more rides in every age group while females take longer trips at almost every birth year.

A review of trips when reviewed by time of day and week reveals that peak hours differ between weekend and weekdays. During weekdays, most trips are taken during morning and evening rush hours, while most trips are taken in the early afternoon on weekend. Overall, more trips being taken during the weekdays than Saturday and Sunday. An analysis of peak months and dates shows that there are no outliers, all dates and months falling with a normal distribution and peak months being late Spring and early Summer while the months with the lowest number or records falling in the winter Months.

Examination of relationships between number or records and weather shows that changes in precipitation does not appear to have any association with changes in number of rides. Temperature does have a weak positive linear relationship with number of rides, but with an r^2 of .54 is not a useful metric for predictive analytics.

The static map confirms what many would expect of the dataset with most trips being centered around Manhattan hot spots and the longest trips ending farther from the city center. Based on the analysis 3 of the 5 most popular Citibike stations are near major transportations hubs:

- Grand Central (major train station)
- Union Square (major subway station where multiple lines

connect)

-Penn Station (major train station)

A review of the purchases and miles traveled over time shows a clear seasonality previously observed in the plot of peak dates. An overview shows up upward trend in number of bikes, total members and total stations and a downward trend in bikes per member.

Conclusions

The recent downward trend in bikes available per member may result in insufficient supply during the busiest months of the year, and thus this report suggests a few proactive measures should be taken to avoid this. Bike fleet should be systematically taken out of service and all preventative maintenance done during the colder months and new bikes purchased periodically based on current fleet numbers.

Currently there is an opportunity to increase membership by recruiting more heavily among female residents and tourists. With a targeted marketing campaign and more features that would appeal to this demographic there is potential to tap into a new market segment and increase not only overall rides but also average trip duration.