

Operating Report

City of Chicago, 2017

Claire Lubash

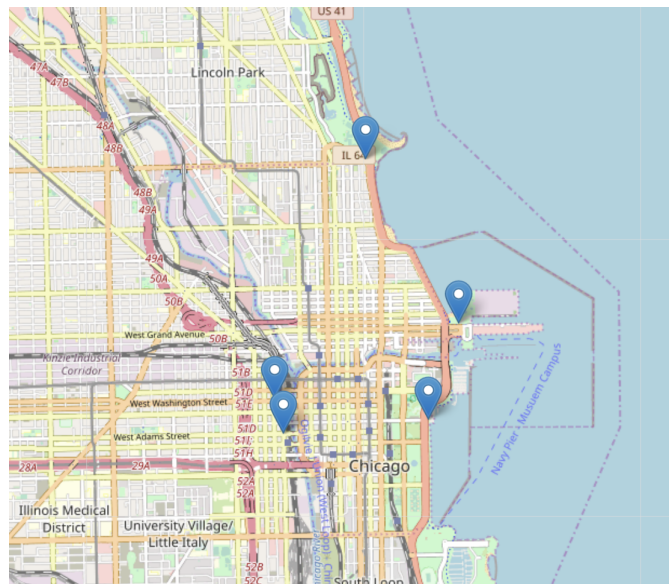
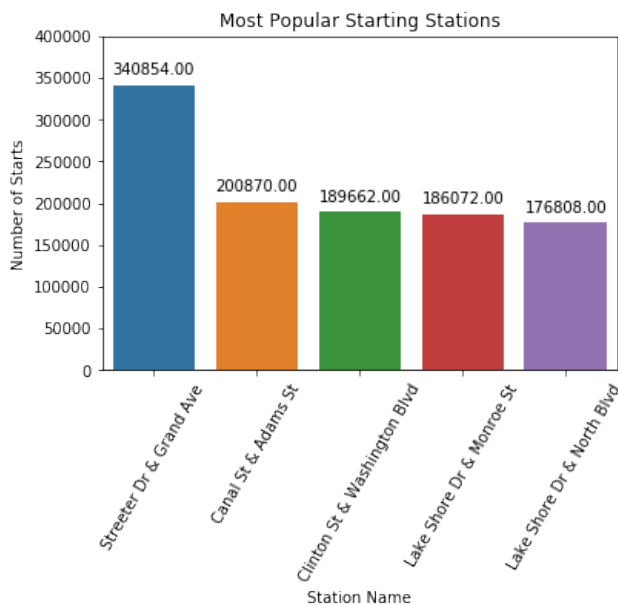
1 Introduction

The data sourced from all Divvy Bikes trips and stations in Chicago in 2017 was used to analyze and visualize various aspects and relationships within the data. External data depicting traffic spikes during rush hour and weather patterns were also added. Ultimately, a model was created in order to predict the time a bike trip would take given a starting and ending station. The model predicted with a low accuracy, due to an influx of data points within the set categorized by a short distance traveled and a long trip duration. Additionally, data was retrieved from Yelp, which described various restaurant reviews in Chicago in 2017. Through various textual analysis techniques, the most popular cuisines were determined, as well as the most frequent words used to review them and the most popular topics amongst all reviews. Finally, a correlation was established between star rating and the sentiment of a given review.

2 Data Visualization

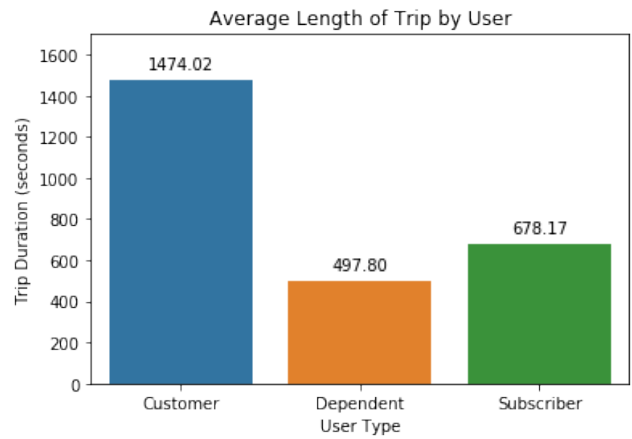
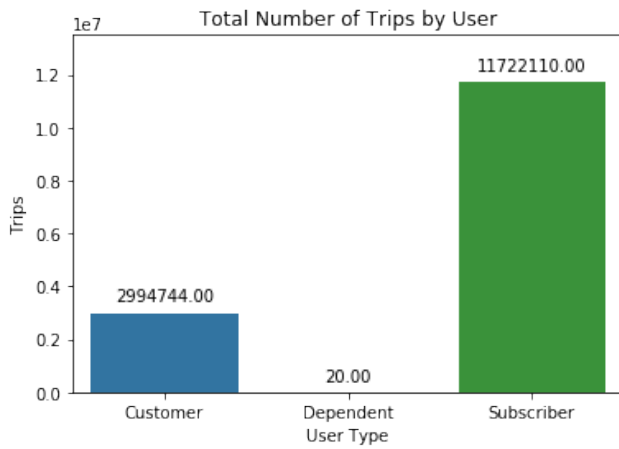
2.1 Most Popular Starting Stations

The most popular starting station is Streeter Drive & Grand Avenue, with over 340,000 starts in 2017. The top 5 most popular starting stations are shown geographically over a map of Chicago.



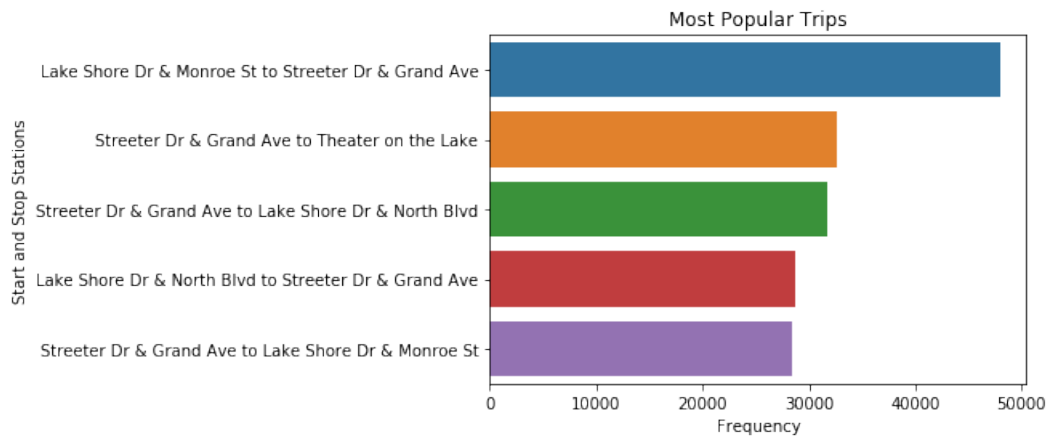
2.2 Trip Duration by User Type

During 2017, subscribers took a significantly larger amount of trips than customers. However, on average, customers had longer trips than subscribers.



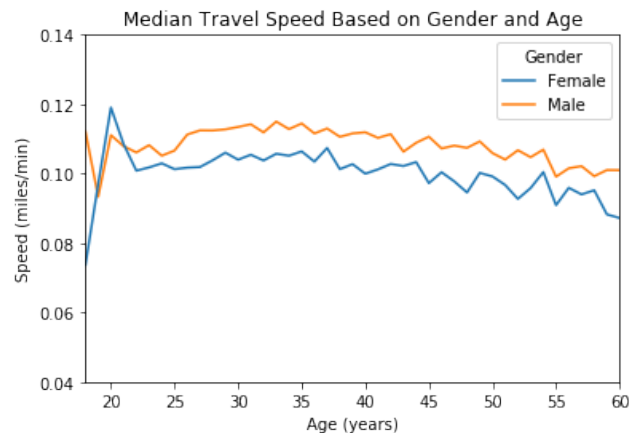
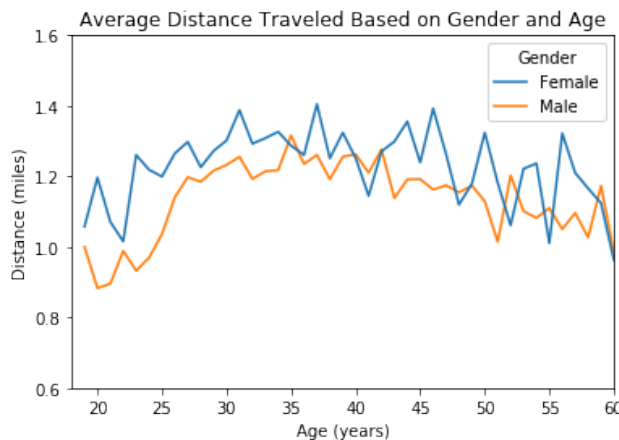
2.3 Most Popular Trips

The most popular trip taken was from Lake Shore Drive & Monroe Street to Streeter Drive & Grand Avenue.



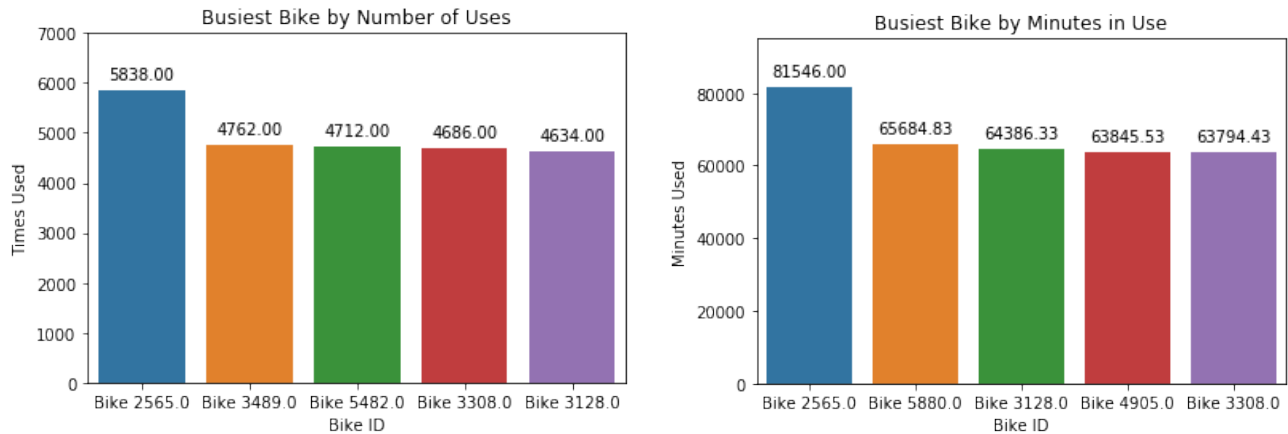
2.4 Rider Performance by Gender & Age

In general, female riders are traveling further distances, but male riders are traveling at a faster speed.



2.5 Busiest Bikes

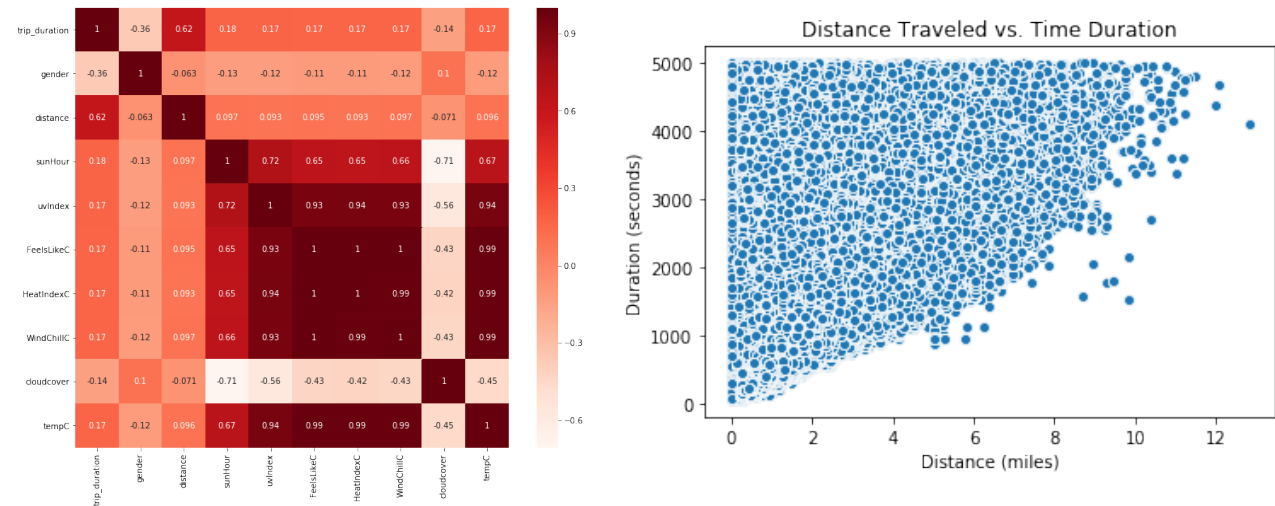
Based on both number of uses and minutes, Bike 2565 was the busiest bike in Chicago in 2017.



3 Data Analysis

3.1 Feature Importance

Unsurprisingly, the heat map provides evidence that the most impactful feature on trip duration is distance. Therefore, we will take a closer look into the relationship between the two variables. We notice a large portion of our data consists of trips with a low distance and a high duration of time.



4 Model

A predictive model was created utilizing the highest correlated features with trip duration, as determined by the heat map above. However, due to the spread of data as shown in the scatter plot above, the model only predicts with 59% accuracy. We can see that a kiosk feature like this is possible. In order to allow for a potentially more accurate model, a recommendation would be to take a random subset of the data and use more sophisticated algorithms such as a random forest classifier. This could allow for a more accurate prediction, even though the data has such skewed points.

5 Yelp Reviews

5.1 Popular Words Based on Cuisine

From the data sourced from the Yelp website, we find that the top most reviewed restaurant cuisines are:

- Mexican
- American
- Korean
- Indian
- Chinese

Below are word clouds that were generated from the most frequently used words in reviews for each of those cuisines.



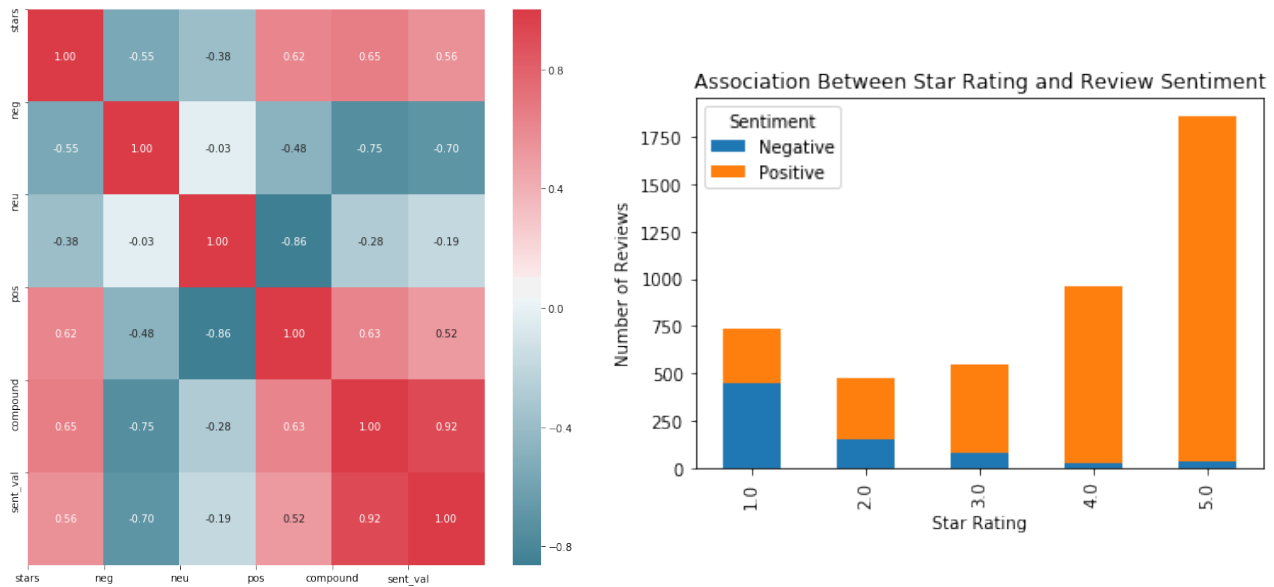
5.2 Topics Extracted From Reviews

Through textual analysis, we inferred the following list to be the most popular topics throughout all the Yelp restaurant reviews in 2017.



5.3 Correlation Between Star Rating and Sentiment of Review

The heat map shows a correlation of 0.56 between star rating and sentiment. This confirms that there is a positive correlation between the two variables. However, the bar chart shows that almost all of the five star restaurants had positive reviews, while only about two-thirds of one star restaurants had negative reviews.



6 Conclusion

From our analysis from both the Divvy Bikes and Yelp review data sets, we conclude the following points:

- Top 5 Stations With Most Starts
 1. Streeter Dr & Grand Ave (340,854)
 2. Canal St & Adams St (200,870)
 3. Clinton St & Washington Blvd (189,662)
 4. Lake Shore Dr & Monroe St (186,072)
 5. Lake Shore Dr & North Blvd (176,808)
- Subscribers take more trips, but customers have longer trips.
- Most Popular Trips
 1. Lake Shore Dr & Monroe St to Streeter Dr & Grand Ave (48,040)
 2. Streeter Dr & Grand Ave to Theater on the Lake (32,612)
 3. Streeter Dr & Grand Ave to Lake Shore Dr & North Blvd (31,748)
 4. Lake Shore Dr & North Blvd to Streeter Dr & Grand Ave (28,700)
 5. Streeter Dr & Grand Ave to Lake Shore Dr & Monroe St (28,440)
- Female riders travel further, but male rider travel faster.
- The busiest bike is Bike #2565. (5,838 uses & 81,546 minutes in use)
- A model can be used to predict the time of a trip given a starting and ending station with 42% accuracy.
- The most reviewed cuisines are Mexican, American, Korean, Indian, and Chinese.
- The most frequent topics found within restaurant reviews are:

1. Wait Time
 2. Sushi
 3. BBQ
 4. Chinese Food
 5. Friendly Staff
 6. American Food
 7. Mexican Food
- There is a positive correlation of 0.56 between star rating and sentiment of review.
 - Five star restaurants are more likely to receive positive reviews than one star restaurants receiving negative reviews.