**Citi Bike Trip Analysis**

698 Project

Chunmei Zhu

1. Introduction

I am interesting in the life style of people in New York. Citi Bike Ride has been an essential behavior for some New Yorkers. Where do Citi Bikers ride? When do they ride? How far do they go? Which stations are most popular? What days of the week are most rides taken on?

1.1 Background

Citi Bike is a largest bike share system in the nation, launched in May 2013 in New York City. It is available daily for use 24 hours and riders have access to thousands of bikes at hundreds of stations across Manhattan, Brooklyn, Queens and Jersey City. The bikes can be unlocked from one station and returned to any other station for one-way trips throughout the city.

Citi Bike is operated by Motivate, the global leader in bike share. Motivate works to re-envision how people experience and move around cities. In 2018 Citi Bike has over 143,000 members and grew to 12,000 bikes. Riders take their 50 million trips to 750 stations a year. (<https://www.citibikenyc.com/about> )

It was designed for quick trips with convenience in mind, and it’s a fun and affordable way to get around town. Single ride costs $3 up to 30 minutes and Day Pass costs $12 unlimited 30-minute rides in a 24-hour period. It also provides $14.95 monthly commitment and $169 annual membership for unlimited 45-minute rides. (<https://www.citibikenyc.com/pricing> )

1.2 Problem Statement

In 2017 the mayor De Blasio needs a better understanding of Citi Bike ridership and asked for an operating report of Citi Bike. There are five main points in the mayor’s request as following.

* What are the top 5 stations with the most starts with number of start?
* How long duration per trip based on different type of users?
* Where are most popular trips based on start station and stop station?
* How riders performance by gender and age based on average trip distance (station to station), median speed (trip duration / distance traveled)?
* What is the busiest bike in NYC during the year? How many times was it used? How many minutes was it in use?

Additionally, the Mayor wants to pitch to Citi Bike and needs to improve its feasibility. He would like Citi Bike to add a new feature to their kiosks: “Enter a destination and we’ll tell you how long the trip will take”. He wants to build a model that can predict how long a trip will take given a starting point and destination. (<https://towardsdatascience.com/citi-bike-2017-analysis-efd298e6c22c>)

1. Data

2.1 Data Source (more or less)

Citi Bike Rides data is an open public data source provided by Citi Bike System. The data is available from launch to now and has been updated monthly in the official website (<https://www.citibikenyc.com/system-data>).

For further analysis and modeling, I also combine some import factors, like NYC Weather Data by hourly/ daily; MTA fair price update, Uber rides, ect.

2.2 Attributes

In Citi Bike trip data, there are 11 attributes for each trip, which includes Trip Duration (seconds), Start Time and Date, Stop Time and Date, Start Station Name, End Station Name, Station ID, Station Lat/Long, Bike ID, User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member), Gender (Zero=unknown; 1=male; 2=female), and Year of Birth.

1. Challenge

This is a big data analysis. Dealing with missing data and model adjusting from time series data are challenged in prediction. I will use R/Python to data mining and data visualization in a build-in dashboard.

1. Expected Results

I will answer the mayor’s questions for the basic analysis. Then I will build a number of rides perdition model base on temperature change, weather status and price change of competitors effected on rider’s behaviors. One model for riders is how many trips to make a fair value day, monthly, annual membership. And I will try one model for Operator profit evaluation (need program maintaining cost).