Analysis on Citi-Bike Trip Durations

hx517[1]

[1]New York University (NYU)

November 9, 2017

Abstract

Abstract content goes here

Abstract

Citi-Bike is becoming more and more popular in NYC daily public transportation. This project focuses on the distribution of trip duration and see whether most of Citi-Bike trips are lower than 20 minutes, in order to see whether Citibike will be as a solution to travel a short distance, or as a routine commuter program. The study found that the fraction of long(>20min) trips over total number of trips on weekdays is less than the fraction of short(<20min) trips over total number of trips on weekdays.

Introduction

Citi Bike is a privately owned public bicycle sharing system serving New York City and Jersey City, New Jersey. Named after lead sponsor Citigroup, it is operated by Motivate (formerly Alta Bicycle Share), with former Metropolitan Transportation Authority CEO Jay Walder as chief executive. The system's bikes and stations use BIXI-branded technology from PBSC Urban Solutions.This research tries to answer the question of whether there are more long trips of Citi Bike then short trips of it on weekdays.

Data

The data used in this research is retrieved from Citi Bike's open data. The analysis is based on all the trips in January 2014. The research try to figure out the riding behavior on weekdays. Thus I transfer the column "start time", which is in string format, to column "date" with function "to date time". Meanwhile, since only trip duration and weekdays are required for this research, other features are all dropped.

Methodology

Chi-square test is chosen for this research. It is because the research works on two unpaired group - long trips and short trips. At the same time, although the trip duration are in numeric types, I put them into two categories - long trips (>20 minutes) and short trips (<20 minutes).

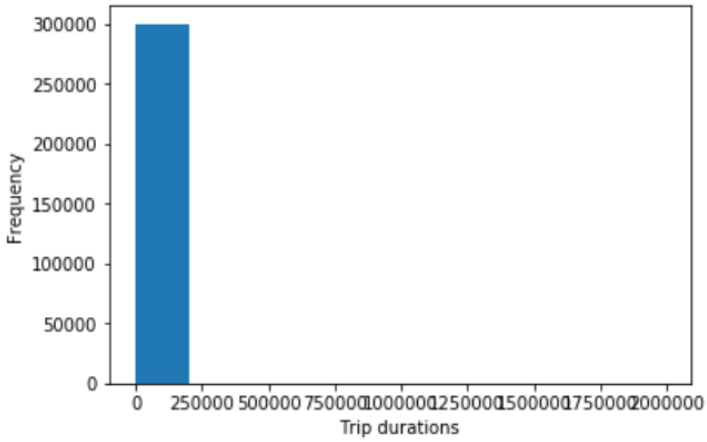

Figure 1: Frequency of Trip durations in January 2014 before removing outliers.

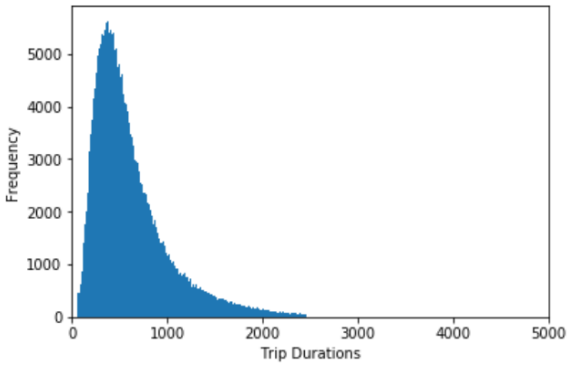

Figure 2: Frequency of Trip durations in January 2014 after removing outliers with 3-sigma rules.

Figure 1: This is a caption

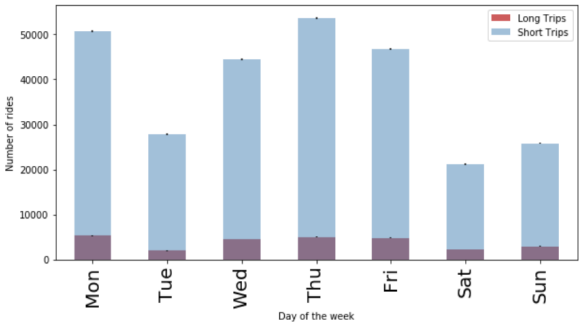

Figure 1b: Distribution of Citibike bikers by ride time in January 2014, absolute counts, with statistical errors

Figure 2: This is a caption

Conclusions

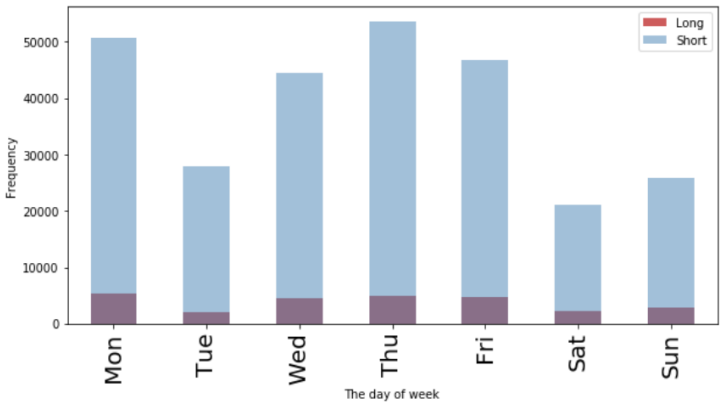

Figure 3a: Distribution of Citibike bikers by trip time in January 2014, absolute counts

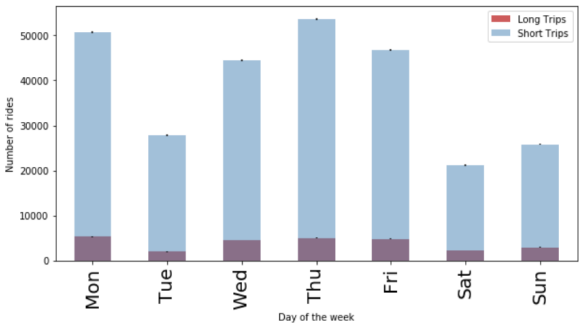

Figure 1b: Distribution of Citibike bikers by ride time in January 2014, absolute counts, with statistical errors

Figure 3: This is a caption

Fig.3 show the distribution of Citi Bike trips on riding durations.

Obviously, the number of short trips is about triple the number of long trips when we separate the trips by the day of week. Chi-square test also gives the same result that the null hypothesis is rejected on significance level 0.05. The chi-square statistics is 183.5 , which is far higher than the threshold of 3.84 (one-tail, alpha = 0.05, degree of freedom = 1). The result matches the assumption of the "last mile" theory that most of people regard Citi Bike as a commuter for short distance/time. However, there are some weaknesses in the analysis.

The data I used is the trips in January2014. It's very likely that people chose not to ride for too long time in cold winter. The better way is to choose the data in each season in order to ensure that there is no bias in data selection.