



A Citi Bike Analysis of Trip Durations Between Weekday and Weekend

Yuwen Chang¹

ABSTRACT

Citi Bike is an important component of NYC’s public transportation. Much research has been done to analyze user and trip characteristics in order to provide better services. An important question is whether there exist differences in rider behaviors between weekday and weekend. If biking is more likely to be for leisure on weekends and more of a commuting tool on weekdays, it may call for different bike allocation strategies. However, my research shows that there is no significant difference between the two regarding trip durations that took place in January 2015.

INTRODUCTION

Citi Bike is NYC’s bike sharing system. With more than 300 stations and 34,000 rides each day, Citi Bike serves as the last mile solution for public transportation network. However, balancing bikes at stations has been a great challenge. Bike stations may be emptied or overwhelmed in certain peak time periods. If weekday trips are more of a commuting tool while weekend biking is more likely to be for leisure, different bike allocation and balancing strategies may apply due to different rider behaviors (e.g., longer commute, different origin-destination clustering, etc.). This research thus aims to answer the sub-question of whether or not weekend trips last longer in average in comparison with weekday rides.

DATA

The data used in this analysis is retrieved from Citi Bike’s open trip data. The time range was specified to all trips that took place in January 2015 for this analysis. Since the original data only contains the most basic format of start and end of trip timestamps, I process the data in order to extract an indicator for weekends that allows me to group all trip data by weekdays versus weekends. The preliminary exploratory analysis of data, as shown in Fig. 1, shows that most trip durations are within half an hour.

METHODOLOGY

With the normality assumption for the population distribution of Citi Bike trip durations, I am testing a single variable (i.e., trip duration) between two categories within an unpaired data set that has 285,552 observations. According to these criteria,

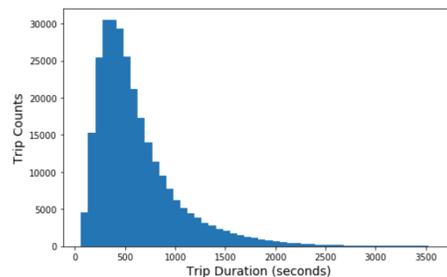


Figure 1. Histogram of NYC Citi Bike trip durations in January 2015

I chose to perform Welch’s t-test, which compares the difference in trip duration and test if the difference is significant at the designated significance level of 0.05.

My peer Srikanth also suggests conducting either a z-test or Welch t-test depending on the homogeneity of variance and under the assumption that the data is parametric. I adopt the t-test with the consideration that there may be differences between the variances of two groups.

I thus conduct a one-tail test under the null hypothesis that the average trip duration on weekends is the same as or shorter than that of weekdays, on which hasty commuting takes a greater proportion.

CONCLUSIONS

Both Fig 2 and Fig 3 show that there is seemingly little difference among each day of week and between weekdays and weekends regarding the average trip duration, which is around 11 minutes. The result of the t-test also shows no statistical significance for the difference. The t statistic for the test is 0.65, which is far smaller than the threshold of 1.646 (one-tail, alpha = 0.05). I can not reject the null hypothesis nor conclude that weekend trips are longer than those of weekdays in average.

Some possible explanations for this outcome include that: (a) people don’t use Citi Bike for leisure; (b) the 30- and 45-minute constraints on each ride before charging extra money limit the duration of each trip, in other words, people could be riding longer trips by switching bikes; (c) people don’t ride bikes for leisure on cold snowy days. It turns out the null

hypothesis is successfully rejected in the data set from July 2017.

The good news is that this matches my intuition from the visualization. However, some weakness is to be considered. First, the distribution of trip duration, as seen in Fig 1, is more identical to a Poisson or chi-square distribution. I may need to revisit the assumption of the distribution being Gaussian. The second improvement to be made is to include data from different months. Using data from winter, in which people use bikes less, may fail to capture the trip pattern in other seasons.

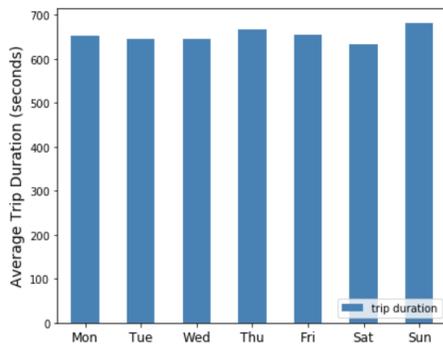


Figure 2. Histogram of average bike trip duration by day of week

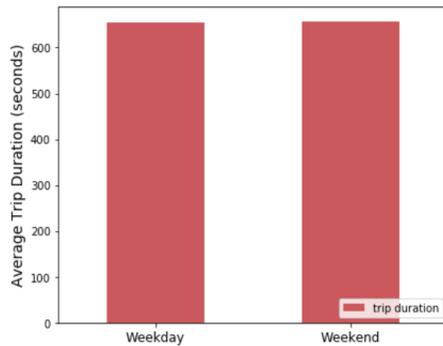


Figure 3. Comparison of average trip duration between weekday and weekend