# **Citi Bike Ridership Mini-Research:** Young People Are More Likely to Use Citi Bikes on Weekends

ZIMAN ZHOU, DONGJIE FAN
*New York University, Center for Urban Science and Progress(CUSP)*
October 16, 2016

**Abstract**

This study aims to find out whether or not young people ride bikes on weekends more often than that of middle-aged people. The analysis performs a hypothesis test (Z-test) to compare the ratio of the number of young people using citi bikes on weekends over weekdays to that of mid-age people. The result shows that the under 5% significance level, the ratio of the number of young people biking on weekends over week days (7 days) is greater than the counterpart of middle-aged people.

*Keywords:* Citi Bike, Hypothesis Test, Z-test , Age

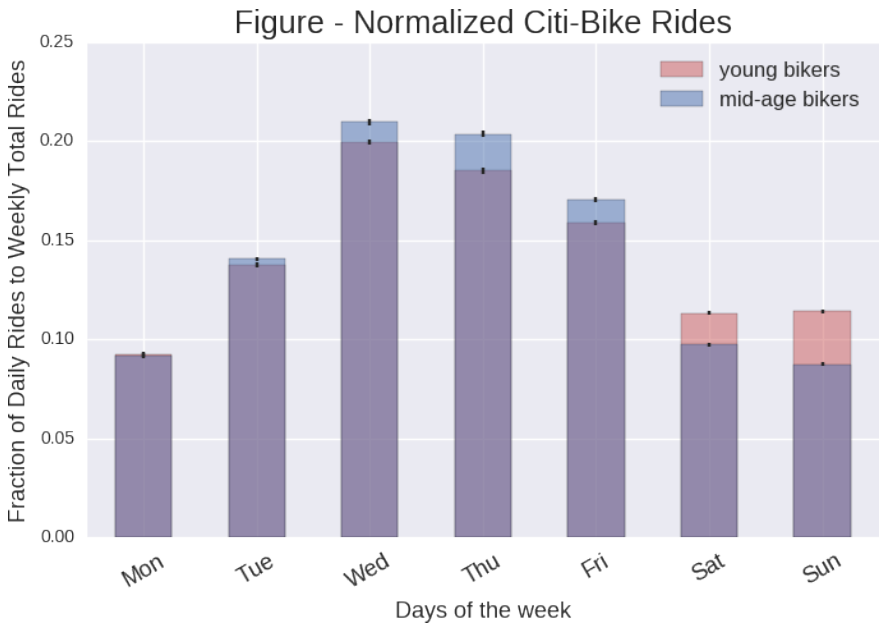
## Introduction

In a fast pace modern city like New York, citi bike has become not only one of the most popular alternatives for commuting, but also a crucial component of a city's gradually formed network system of both transportation and social activities. As a source from which quite comprehensive datasets can be acquired, citi bike is a great subject for researchers to study citizen's behavior through patterns in its ridership. This study aims to find out whether or not young people ride bikes on weekends more often than that of middle-age people, with the assumption that the bikers' usage of citi bikes fully reflects their personal preferences — biking only for general use rather than heavily commuting purpose.

## Data Availability and Processing

All processed data used to perform the statistical test is from: $https://s3.amazonaws.com/tripdata$ which is documented on a monthly basis. The data wrangling process follows the idea of reproducibility and includes the following stages:

1. Enable checking and downloading data to a pointed directory each time when searching for data of a specific month, so the existed data becomes retrievable. We choose **February 2015** citi bike data for our research.
2. Read the data with Pandas Dataframe; select and modify the attributes as needed(i.e. create a binary "age group" by calculating the ages using "birth year"). Label each row with $18 \leqslant age < 40$ and $40 \leqslant age < 60$ as **young** and **middle-aged** respectively.
3. Plots histograms to visualize the normalized fraction of young and middle-aged bikers' average biking trip counts as well as each individual group on each day of the week.
4. Considers the errors of average daily riding counts on weekdays and weekends for both biker groups.



## Analysis

Since the number of population is large (>30) and the standard deviation of population is known, we choose Z-test to do hypothesis test[1][2]. According to the question we focus on, we set

$$H_0 : \frac{\#\,of\ young\ on\ weekends}{\#\,of\ young\ on\ week\ days} \leqslant \frac{\#\,of\ middle-aged\ on\ weekends}{\#\,of\ middle-aged\ on\ week\ days}$$

$$H_a : \frac{\#\,of\ young\ on\ weekends}{\#\,of\ young\ on\ week\ days} > \frac{\#\,of\ middle-aged\ on\ weekends}{\#\,of\ middle-aged\ on\ week\ days}$$

According to the formula[3], we can get $z - score = 24.4665$ and corresponding p value is lower than 0.05. Thus we choose to reject $H_0$, and choose $H_a$ which is the ratio of the number of young people biking on weekends over week days (7 days) is greater than the counterpart of middle-aged people.

## Conclusion

Based on the result of Z-test, we can conclude that under the 0.05 statistical significance, young riders is much more likely to ride citibike than middele-aged people. The reasons behind the result might be diverse, for example, there are more social activities on weekends for young people which make them ride citi bike much more. Furthur research might foucs on extracting data from different months in avoid of seasonal influence.

## References


1. Difference between z-test and t-test. `http://www.differencebetween.net/miscellaneous/difference-between-z-test-and-t-test/`.
2. Z-test vs t-test. `https://www.packtpub.com/mapt/book/Big%20Data%20&%20Business%20Intelligence/9781784390150/2/ch02lvl1sec20/Z-test+vs+T-test`.
3. Hypothesis test: Difference between proportions. `http://stattrek.com/hypothesis-test/difference-in-proportions.aspx?Tutorial=AP`.