Citibike Ridership

pa1303[1]

[1]Affiliation not available

November 7, 2017

Abstract:

Citibike, a company that provides public bikes across NYC has made their data publicly available. One question this paper tries to answer with this data is whether there is a difference in the length of the trip duration of male and female riders. A t-test is used to compare the two groups and the results show that there is no difference between how long rides are based on gender.

Introduction:

Citibike, is a private company that provides a public bike share service in parts of Manhattan, Queens, Brooklyn, and New Jersey. with hundreds of docking stations around the areas. Every month a dataset is released that contains information about citibike usage. One question to be answered from this dataset is if there is a difference between how long men and women use citi -ike. This result would be significant because it could identify if the use cases based on gender are different.

Data:

The dataset used was the most recent data available for August 2017. Each row represents a trip taken on a citibike. One column contains the gender and another contains the tripduration. There were a few trips that were extremely long and outliers in the data. In order to view the results of the most common trips the data was trimmed by 10%. Based on gender Figure 1 shows the distribution of the tripduration. It can be observed that both distributions look very similar for male and female.

Methodology:

In order to test the hypothesis a t test for two means with equal variance was used with a significance level $\alpha = .05$. The hypothesis is this:

$H0: \ \ average\ female\ trip\ duration\ =\ average\ male\ trip\ duration$

$H1: \ average\ female\ trip\ duration \neq average\ male\ trip\ duration$

The reasoning for this test is that is has high robustness against non-normality. Figure 1 shows that the distributions are not normal and have a long right tail. An alternative to the t test was a z-test however the biggest issue was non-normality with the dataset.

Conclusion:

The value of the t test is is 0. The t test shows that we cannot reject the null hypothesis that there is no difference between the trip durations of males and females.

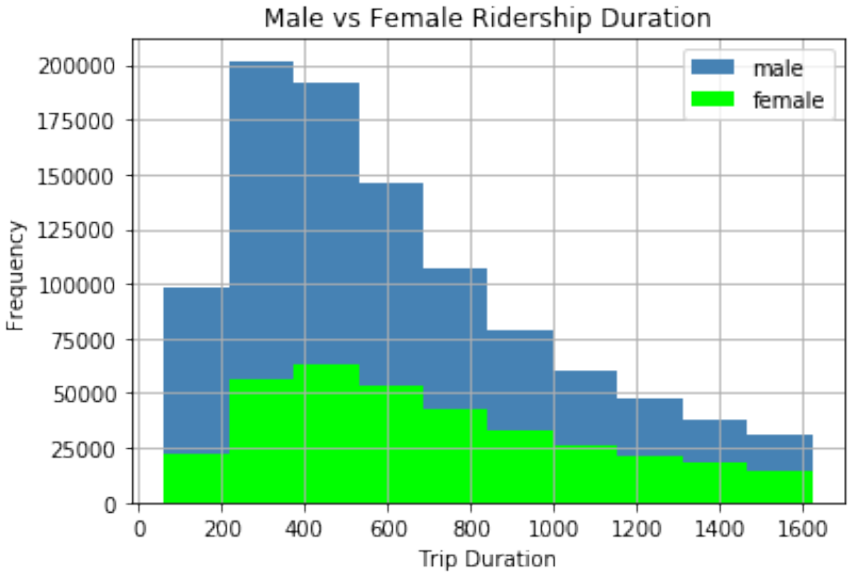

Figure 1: Frequency of trip durations based on Male or Female