

Citibike Review

Rebecca Scheidegger¹ and Emily Padvorac¹

¹NYU Center for Urban Science & Progress

November 6, 2017

Abstract

An analysis of Citibike data from July 2017 considered if the average trip duration by gender was statistically significant. The null hypothesis that the average trip duration for males is shorter than or equal to the average trip duration for females was rejected based on a two-sample T-test and a Z-test.

Introduction

Citibike is a popular bike sharing program in the NY metro area, including Manhattan, Queens, Brooklyn, Staten Island, and New Jersey. Launched in 2013, Citibike collects a wealth of data on each ride and rider, including gender and trip duration. An interesting analysis is to consider whether male and female riders exhibit different behaviors. We sought to understand if average trip duration for males is shorter than or equal to the average trip duration of females.

Data

An archive of Citibike data is easily accessible through Citibike. Data, in a CSV format, was obtained directly from the Citibike website (<https://www.citibikenyc.com/system-data>) in October 2017. The data was read using a Pandas data frame. Instances where gender was “Unknown” were not taken into account.

| | Average Female Trip Duration | Average Male Trip Duration |
|------------|------------------------------|----------------------------|
| starttime | | |
| 2017-07-01 | 21.782166 | 18.111268 |
| 2017-07-02 | 24.100578 | 20.141165 |
| 2017-07-03 | 19.129823 | 18.276041 |
| 2017-07-04 | 19.975015 | 22.380926 |
| 2017-07-05 | 18.953602 | 15.635412 |
| 2017-07-06 | 16.465340 | 15.280398 |
| 2017-07-07 | 14.734951 | 13.839599 |
| 2017-07-08 | 17.335252 | 18.146678 |
| 2017-07-09 | 19.346256 | 16.705310 |
| 2017-07-10 | 16.771729 | 15.003669 |

Figure 1: Fig. 1: A sample of the first ten days of July 2017 shows the average trip duration in minutes for males and females.

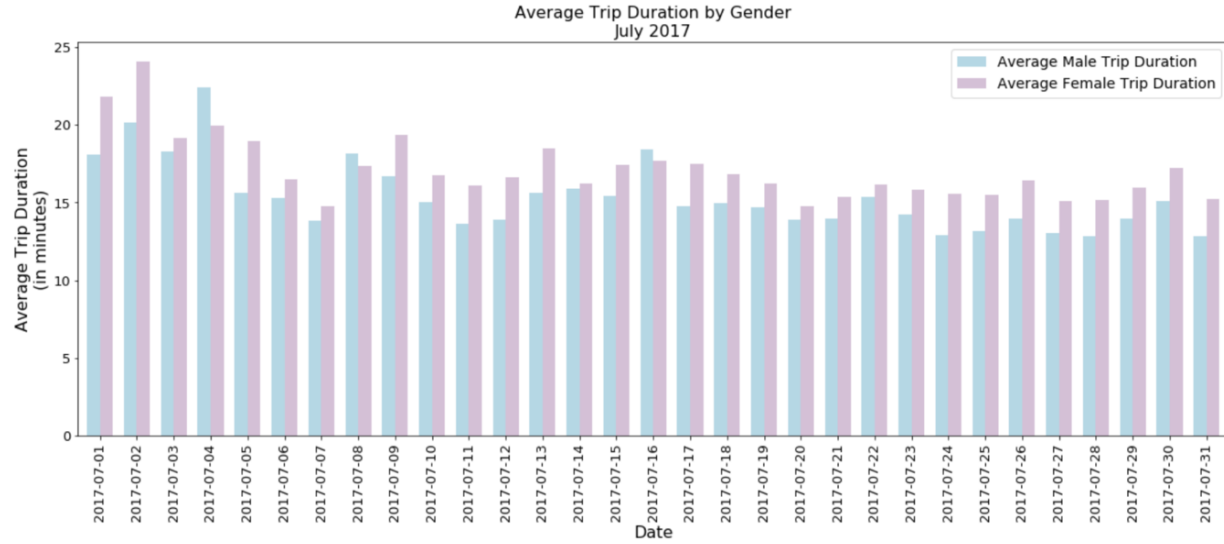


Figure 2: Fig. 2: A visual representation of average trip duration per day by gender in July 2017 indicates that except for three days, women on average took longer trips. Additionally, this indicates the average trip duration for either gender falls into about a ten minute range, from approximately 14 minutes to 24 minutes

Methodology

To test the analysis, two tests were employed: a two sample T Statistic and a Z test. The significance level was set at 0.05. It is generally suggested T-tests are used when $n < 30$ and Z-tests are used when $n > 30$. In this analysis, $n=31$. Because n was so close to the demarcation, we opted to run both tests.

A two sample T Statistic is a good choice because the sample size is small ($n=31$). The results of the two-sample T Statistic were statistic=3.12864137 and pvalue =0.00271001.

A two sample Z test was also done. The result of the two-sample Z test was 4.257. This result is more than 4 standard deviations away from the mean.

With our Z-test being 4.257 and T statistic p-value being 0.0027, we can reject the null hypothesis.

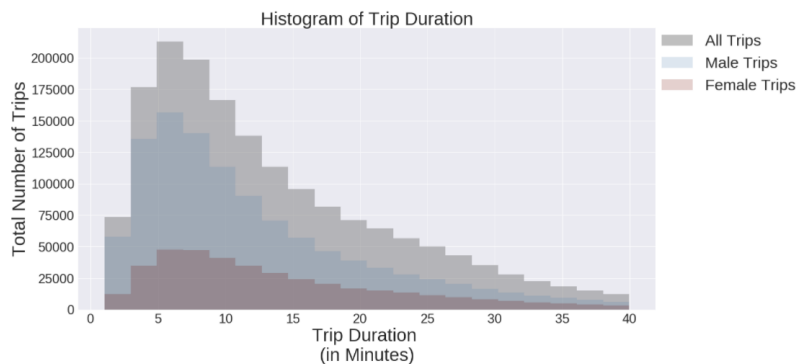


Figure 3: Figure 3: A histogram showing the trip duration for the total amount of trips taken during July 2017, total number of male trips in July 2017, and total number of female trips in July 2017.

Conclusions

One feasible way to strengthen this analysis would be to select a larger time frame, as this is only considering one month. Given more data, it would also be interesting to explore behavior by gender based on the purpose of trips (commuting vs leisure), but Citibike does not currently track this. A proxy for this could be looking at day of week and time of day. For instance, a trip beginning at 8:30am on a Monday would likely be a commuting trip, versus a trip at 3:00pm on Saturday.