

Seniors Utilize Citi Bikes for Similar Ride Durations Compared to Younger Users

Colin Andrus¹

¹New York University Shanghai

November 7, 2017

Abstract

Citi Bike is a bike sharing service operating in the New York City metro area that rents bikes for rides between its stations. People of all ages use Citi Bike, including senior citizens. Using publicly available Citi Bike data, I tested whether senior citizens took rides of shorter times than younger users. My findings show that seniors did not take trips of significantly shorter ride durations.

Introduction

Citi Bike is a bike sharing service operating in the New York City metro area. Bikes can be rented from and returned to a collection of stations spread throughout Manhattan, Brooklyn, Queens and Jersey City. For each unique Citi Bike trip, data is collected for the trip and the rider. I sought to determine if seniors (>64 years old) take shorter rides on average than other users of Citi Bike. I hypothesized that seniors would take shorter rides than younger people, possibly due to tiring quicker or not needing to commute between neighborhoods to work since many seniors are retired. Knowing if seniors ride Citi Bikes for shorter periods of times could provide valuable insights for spacing of bike docking stations and marketing strategies aimed at seniors, a rapidly growing consumer group.

Data

Data from Citi Bike trips was taken in January and July of 2016 was collected. For my analysis the 2 features of interest that I kept were age and trip duration (seconds). Both features contained outliers that did not represent real bike ride durations or highly improbable ages.

Trips with riders born before the year 1927 were removed. I believe these birth years were incorrect since most 90 year olds are relatively frail and would use other methods for transportation than a bike share. Also all rides without a birth year were removed rather than provided with average age to avoid classifying them all as younger riders.

Outlier rides with a trip duration above 3 hours were also removed. These rides likely were recorded as being very long due to an incorrectly docked bike since it is quite expensive to rent a bike for that much time and it is doubtful many people would do it intentionally.

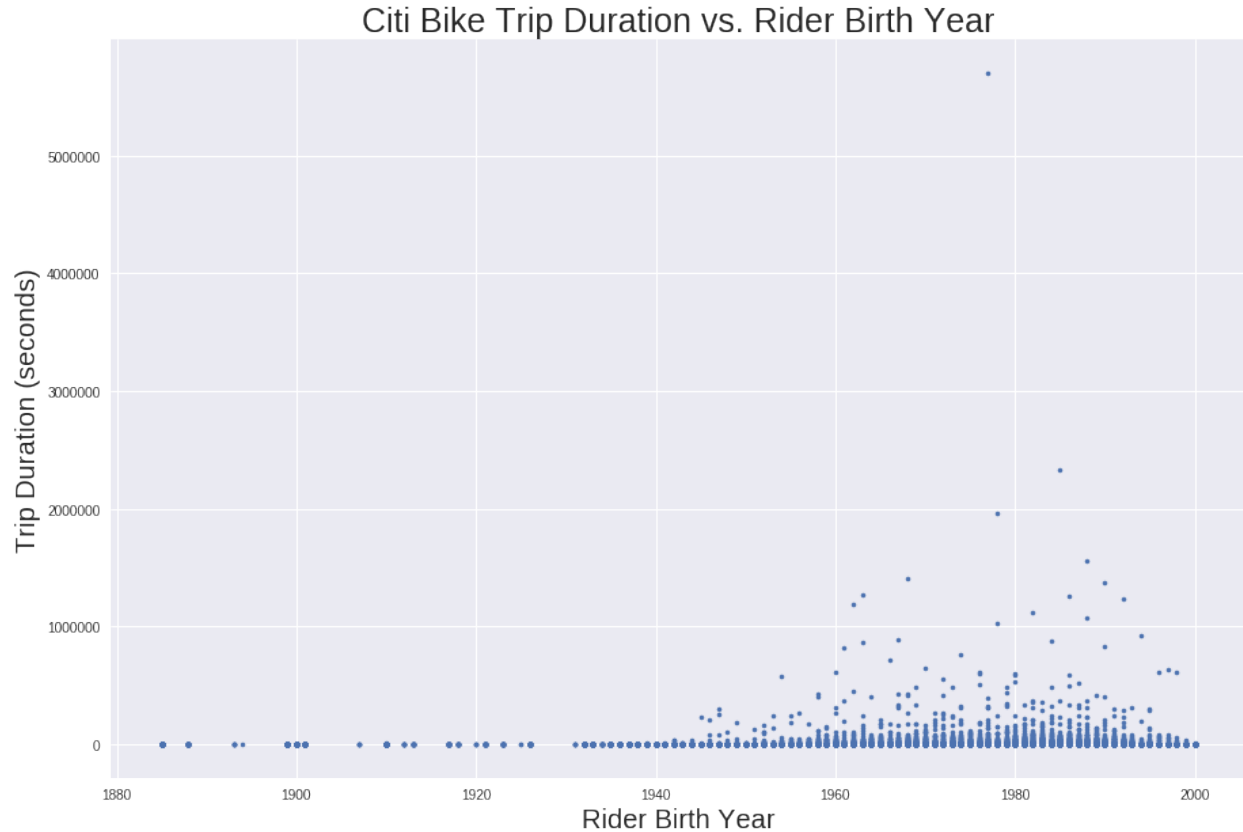


Figure 1: Scatter plot of trip duration vs. birth year shows few rider claiming to be extremely elderly, likely do to errors or intentionally incorrect inputs.

Methodology

To test whether seniors rode Citi Bikes for shorter rides than younger people, I split rides into a senior group and a non-senior group. A cut-off of being born before 1953 (older than 64) was used to classify seniors. Then data was log transformed to fit obtain a more normal distribution.

In order to test my hypothesis that the mean ride time for seniors is significantly less than the mean ride time for non-seniors I decided to use a two sample z-test. This make sense for my data since I will be comparing normally distributed data with my log transformed ride duration values. One other method that I considered, but decided against was splitting rides into groups of long and short and then using a chi-square test for proportions. I decided against this because I would have to supply a cutoff for short and long trips, and I could bias the results depending on what I chose.

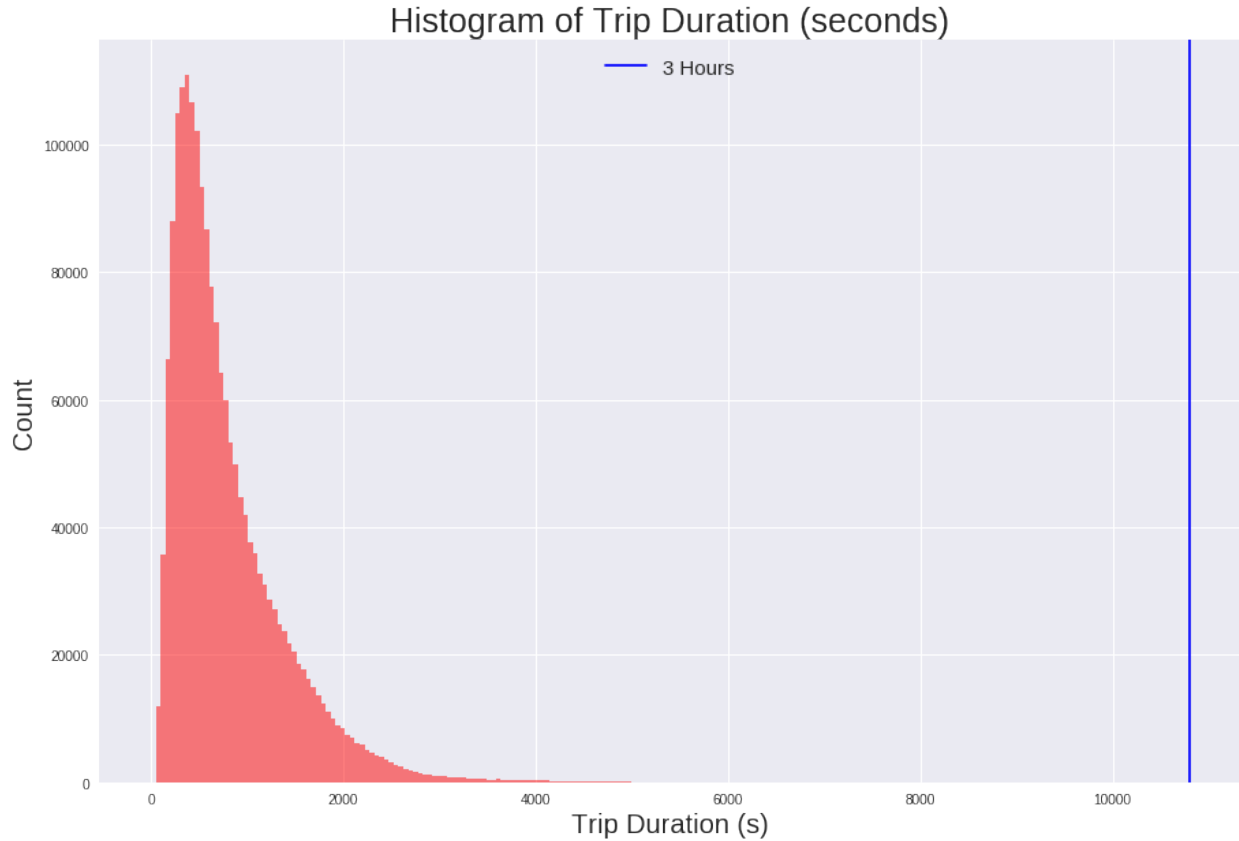


Figure 2: Histogram of trip duration shows a lognormal distribution of trip times. Also shown is a bar for a 3 hour trip which I will use as a cutoff to remove trips that did not have bikes docked correctly at their destination. As a subscriber to Citi Bike, a 3 hour trip would cost an addition \$22.50 on top of your yearly subscription. Doing that on purpose seems unlikely.

Conclusions

Using a 2 sample z-test with a null hypothesis of the mean of senior's trip durations being greater than the mean trip duration of younger people's rides I calculated a z-test statistics of 13.796 with an associated p-value of 1. This means we cannot reject out null hypothesis that senior's bike ride durations are greater than or equal to those of younger people. In fact when looking at a box plot of ride durations seniors have higher ride durations.

There are some possible issues this method of analysis. One of the biggest is that riders can input their birth year incorrectly. Possibly there are many riders that I have counted as seniors even after removing outliers that are actually younger or seniors that put in a younger age. This analysis would much more useful if the ages were certainly correct.

One strength of this analysis is that it clearly shows that there is not a big shift in ride duration for people at age 65 or above, assuming most people tell the truth about their age, it answers the original question I had about senior's ride duration.

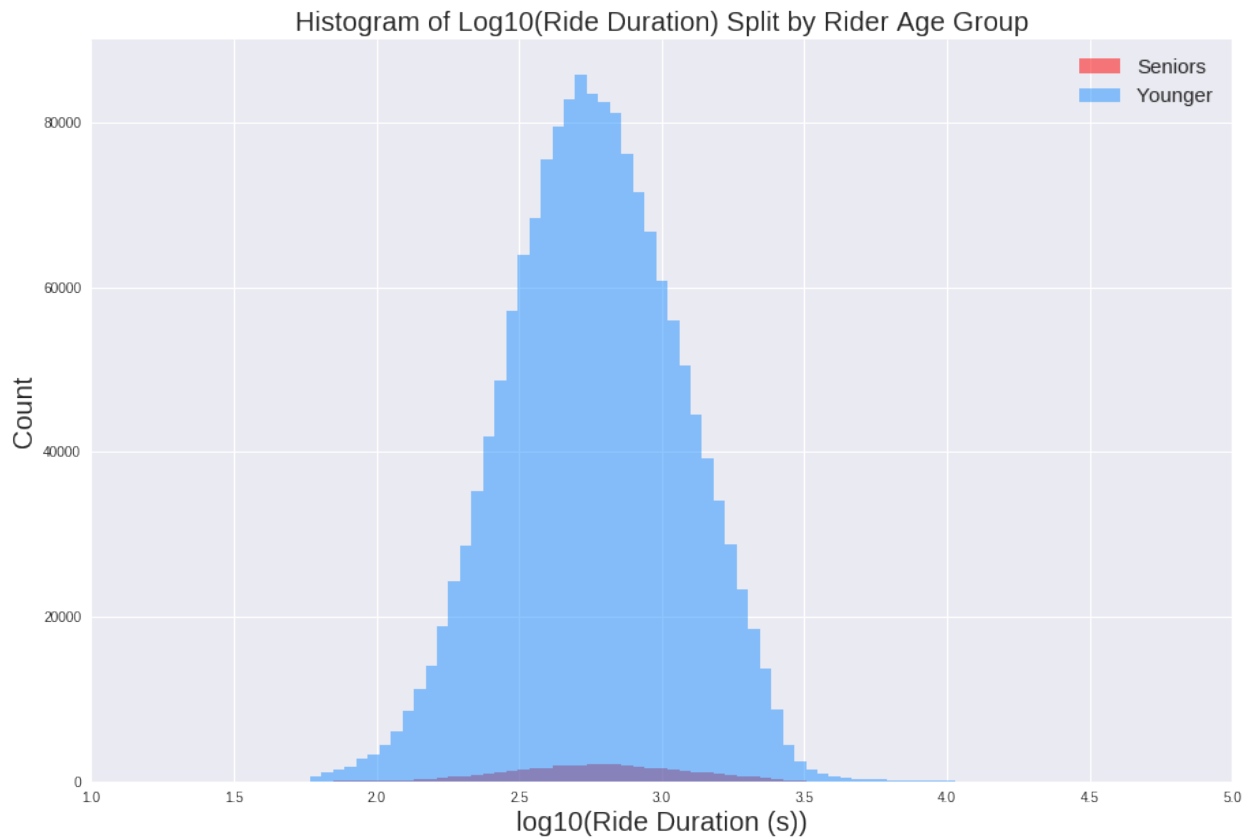


Figure 3: Histogram of $\log_{10}(\text{Trip Durations})$ shows data is relatively close to normal distribution.

Jupyter Notebook

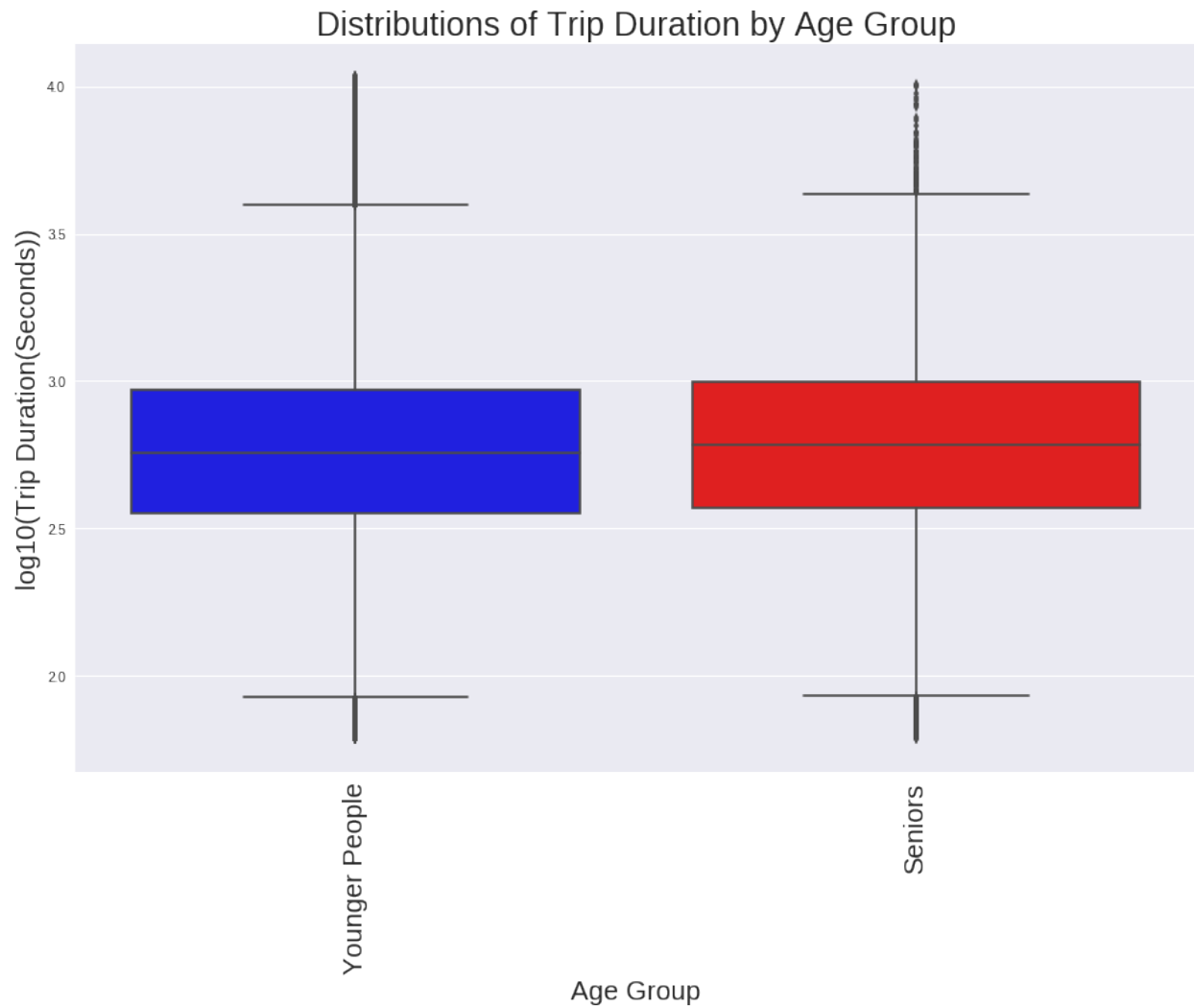


Figure 4: This box plot of log10 transformed trip durations for seniors and younger people shows that there is clearly not a significant difference in means between senior trip durations and younger people's trip durations



Figure 5: Click on “Code” to run the associated .ipynb