Inferential Stats Applied to Capstone 1

To apply inferential stats to my capstone, I first defined my goals and the variables I wanted to look at. For my first test, I wanted to see what the confidence interval was for the average duration of rides in minutes. I first defined my function that takes as input a pandas series, and took a function argument which is by default the mean, repetition count of default 10,000 times, and calculates the 95% confidence interval.

The above mentioned function returns a histogram with the bootstrap resampled test statistics and prints the confidence interval. This allowed me to see that the test statistic of average minute duration for rides was well within the 95% confidence interval. The null hypothesis was thus that the avg. duration of rides is different from 15.47 min. In fact, the confidence interval was 14.68 to 16.36 min., and so we cannot reject the null hypothesis.

I performed the same above examination for the age and cost of ride variables and found the same result: a test statistic well in the range.

I next wanted to look at the difference between two test statistics for two different populations. For this I chose to split by user type: subscribers and customers. I looked difference in the standard deviations of the cost of rides for these two populations. What my bootstrap resampling of standard deviation differences showed me was that the test statistic was within the range again and so I cannot say there is a difference between the two populations.