Citibike: Average Trip Duration Based on Payment Class

Davey Ives[1], Srikanth[2], and Imran Khan[3]

[1]New York University Shanghai
[2]Affiliation not available
[3]New York University (NYU)

November 9, 2017

Abstract:

Those who pay for Citibike rides on a pay-as-you-go basis (called 'Customers' in our study) will, on average, take longer trips than those who have Citibike memberships (called 'Subscribers' in our study).

Introduction:

Citibike is a bike-sharing system that was initiated in New York City in 2013. Users may ride the bikes for a fee, which can be paid out either through a membership plan, or a pay-as-you-go plan. Our group chose to investigate the difference in average trip duration between the two payment classes (Subscriber and Customer) outlined in the project abstract.

Understanding differences in trip duration with respect to payment class could have significant impacts for future policies at Citibike. Citibike could reevaluate the cost of these two payment classes based on how bikes are being used. There are other more practical and operational considerations that could address how bikes are marketed, maintained, and priced throughout the year.

Data:

Our group used data from two months worth of Citibike rides (January 2017 and July 2017) in order to see if there was a seasonal difference in the length of rides between the two ridership classes. We found the average Subscriber ride time to be shorter than the average Customer ride time in both months. Figures 1 and 2 below illustrate this trend.

Average trip duration in July 2017 is greater across both payment classes compared to average trip duration in January 2017. For January however, the difference in trip duration across the two payment classes is much smaller, which results in us not being able to reject our null hypothesis.

Methodology:

To test the null hypothesis given the type of data analyzed and the question asked, we prefer to utilize the t-test. We noted that the data has means for two groups for where both the t-test and z-test could be applied. The t-test was the better selection in our opinion as we had two groups with no standard deviation valuation of the population. The t-test has a better application than the z-test because the uncertainty in the sample variance is not accounted for as the z-test is incorporated.

Conclusions:

Based on the outcome of the t-test, we could not reject the null hypothesis for the month of January, but we could reject the null hypothesis for the month of July.
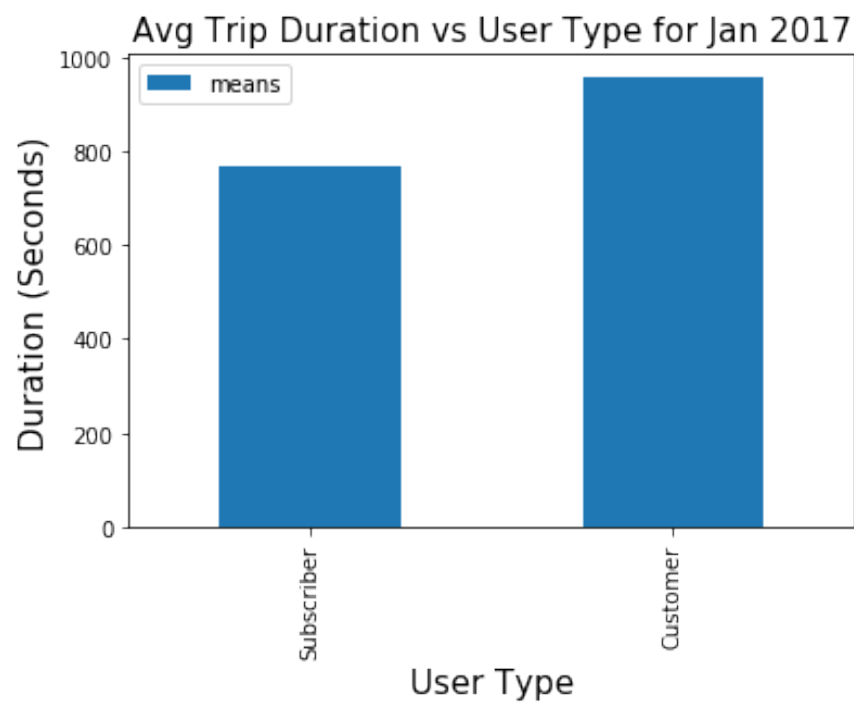
Figure 1: January 2017 Avg. Trip Duration by Payment Class. Note how close these two means are.
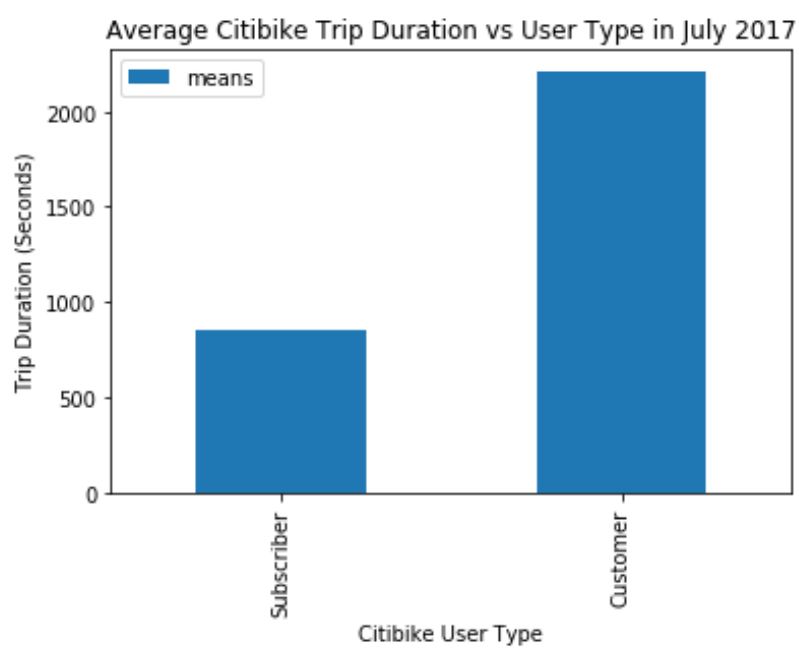

Figure 2: July 2017 Avg. Trip Duration by Payment Class. Note how much further apart these means are.

This conclusion was supported by the results of the t-test, which showed a significance level of 0.17 for the month of January, and a much lower 2.08 x 10^-244 for the month of July. In plain figures, we could see this illustrated by the difference in the means for the two months (see figures 3 and 4 below).

	means
Subscriber	769.908443
Customer	960.451832

Figure 3: Mean trip duration (in seconds) for Subscriber and Customer Classes during January 2017. Note how close the means are to one another.

	means
Subscriber	854.975088
Customer	2215.592281

Figure 4: Mean trip duration (in seconds) for Subscriber and Customer Classes during July 2017. Note how much further these means are compared to the same data in January 2017.

Stronger relationships between trip duration and payment class for July, but not for January, could provide management at Citibike with a host of operational conclusions that could allow them to better optimize their service.