Explanatory Analysis of Ford GoBike System Data

Cathy Moy
July 2019

Effects of Age Group, Gender, and User Type to Average Trip Duration in Minutes

Investigation Overview

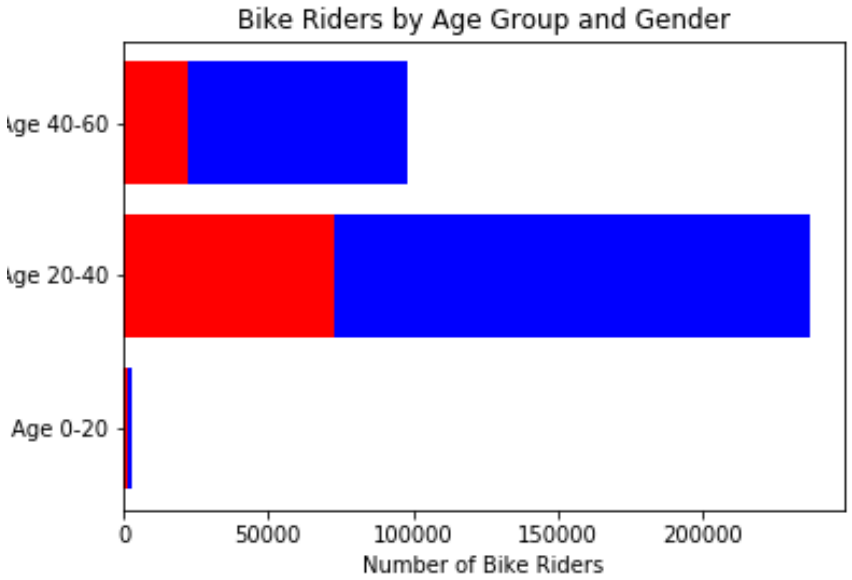
In this investigation, I wanted to look at the characteristics of bike riders that could be used to predict their average trip duration for bike rides with the Ford GoBike System. The main focus was the effect that user type, gender, and age group may have on average trip duration.

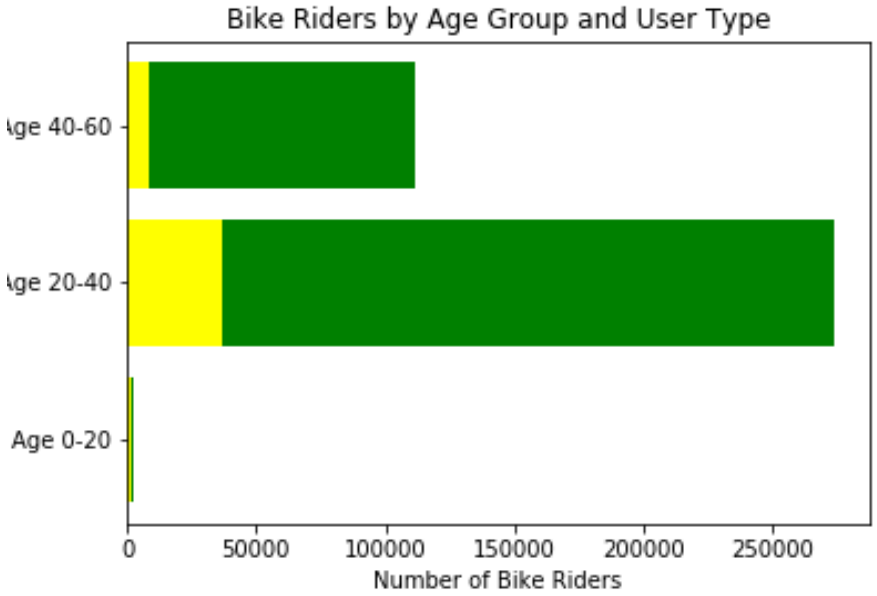
Dataset Overview

The data consisted of attributes, time, and location of bike rides and bike riders. The attributes included user type, gender, birth year as well as additional measurements such as bike id, start and end time, and information about bike stations. There were 519,700 entries of data to start with and there was a need for some data wrangling. So then, 84,602 entries were removed from the analysis due to inconsistencies or missing information.

Distribution of Bike Riders

There was some uneven distribution of Bike Riders among all characteristics of focus. There were significantly more males than females represented in this dataset for gender. Similarly, there were significantly more subscribers than customers represented in this dataset for user type. Additionally, there was very little representation of age group 0-20 than to age groups 20-40 and 40-60. From discovering the uneven distribution of bike riders in this dataset, this will be taken account when finalizing findings. The uneven distribution can be visually seen in the plots on the next page.





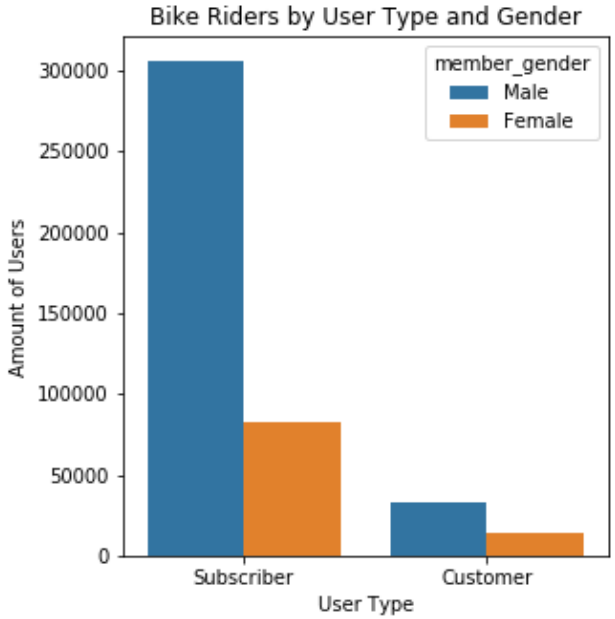
Color Coding of Stacked Bar Plots

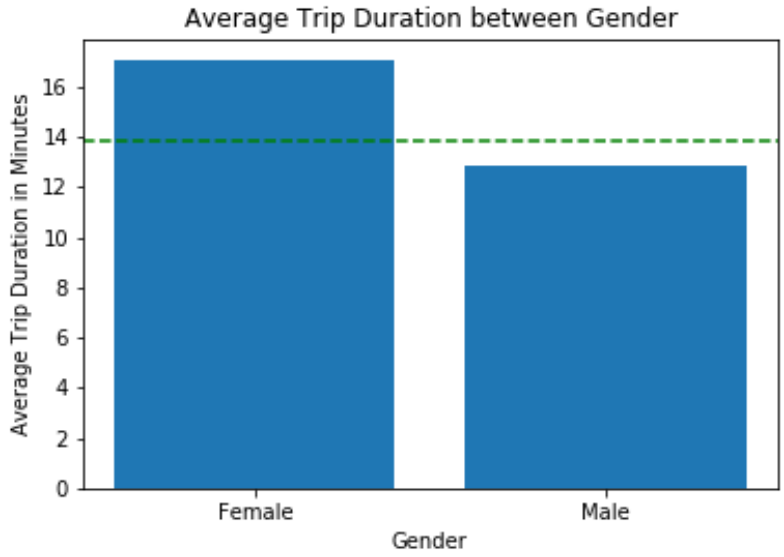
Subscribers: Green

Customers: Yellow

Females: Red

Males: Blue

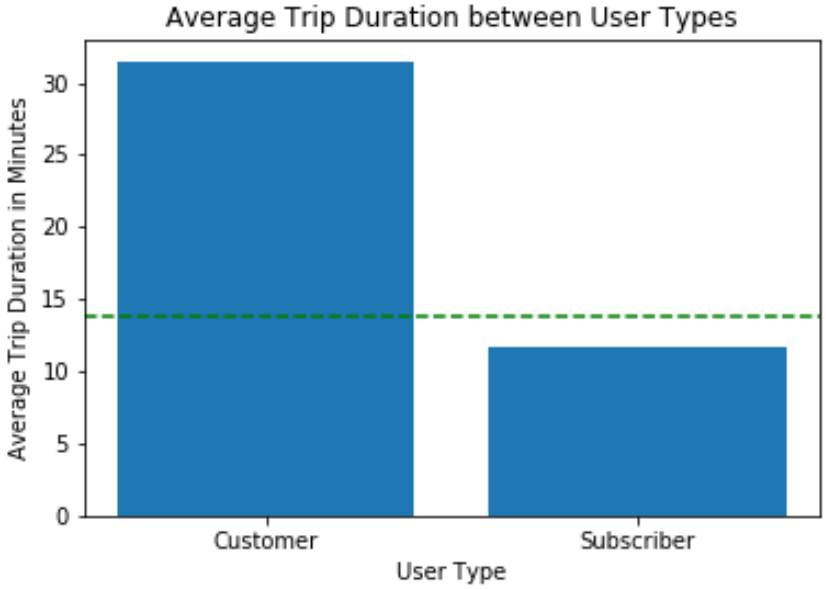




Average Trip Duration vs. Gender

Looking at Average Trip Duration to Gender, Females were above average while Males were below average. Out of the 435,098 entries, there were 338,224 males and 96,874 females. There is nearly 3.5 times as much male representation than is there for females in this dataset. This uneven distribution may have skewed the average trip duration between gender for females.

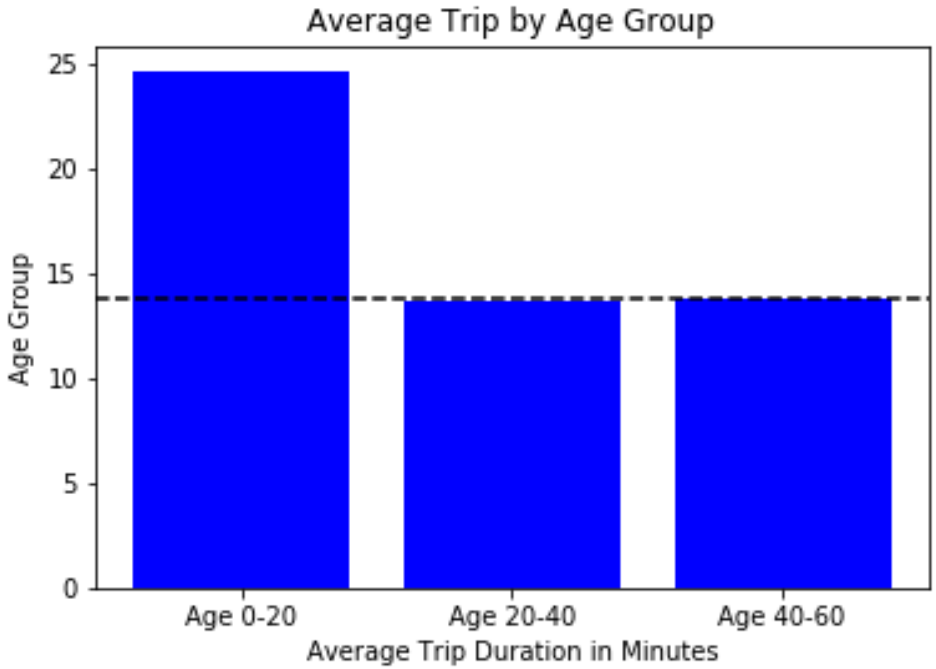
In the general sense, with overall average of 13.82 minutes as represented with the dotted line, female average was 17.06 minutes while males average was 12.89 minutes.



Average Trip Duration vs. User Type

Looking at Average Trip Duration between User Types, Customers were significantly above average while Subscribers were below average. Out of the 435,098 entries, there were 387,763 subscribers and 47,335 customers. Similar to gender, this uneven distribution may skew the average trip duration between user type for customers.

In the general sense, with overall average of 13.82 minutes represented by the dotted line, subscribers average was 11.67 minutes while customers average was 31.47 minutes. Customers average is almost twice as much as overall average. This suggest that customers may take longer bike rides as they are making one-time purchase to make use of their purchase.



Average Trip Duration vs. Age Group

Looking at Average Trip Duration between Age Groups, Age Group of 0-20 were significantly above average while both Age Groups of 20-40 and 40-60 were just in the range of overall average. Out of the 435,098 entries, there were 4,713 for age 0-20, 310,204 for age 20-40, and 120,721 for age 40-60. Similar to gender and user type, this uneven distribution may skew the average trip duration between these age groups, especially for 0-20.

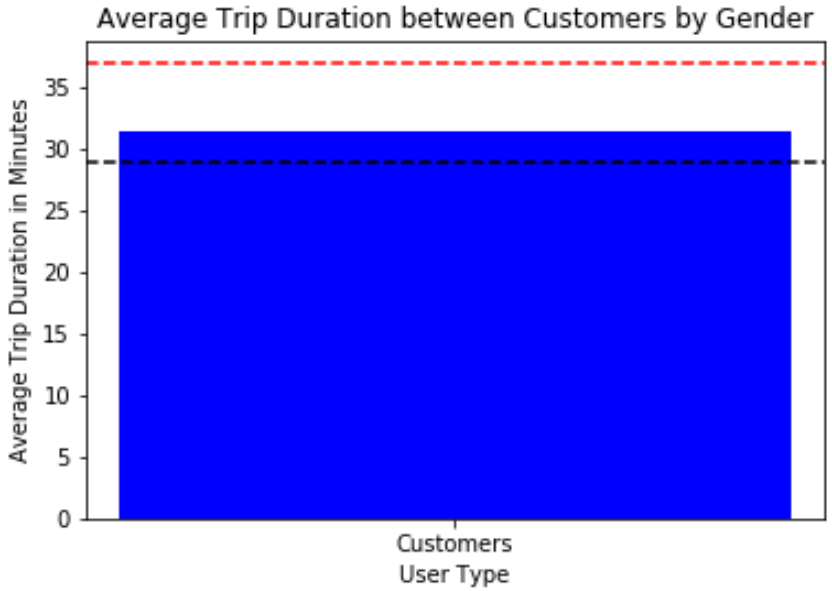
In the general sense, with overall average of 13.82 minutes represented by the dotted line, age 0-20 had an average of 24.61 minutes, age 20-40 with 13.66 minutes, and age 40-60 with 13.85 minutes. Age 20-40 and Age 40-60 had very close averages and to the overall. With consideration to the very low representation of age group 0-20, I can suggest that age may have a small factor to average trip duration in which those in the age group of 0-20 may be likely to have a longer bike ride than those of other age groups.

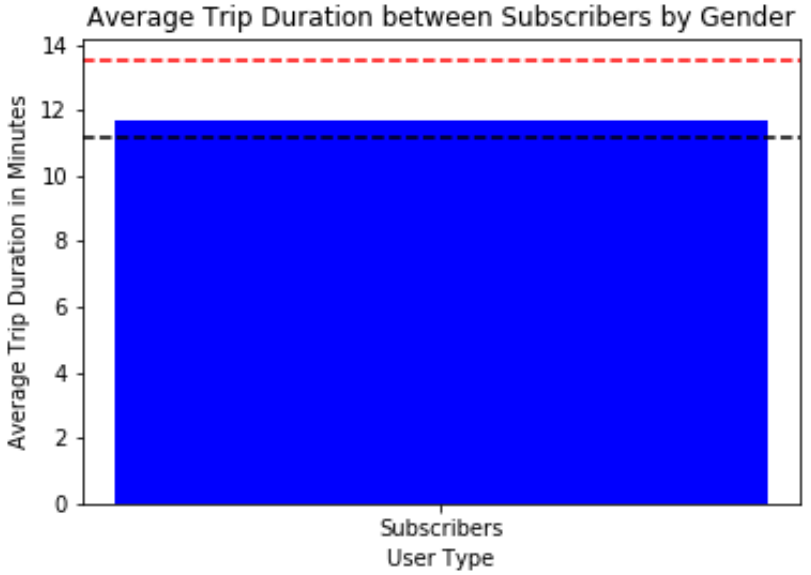
Average Trip Duration by User Type vs. Gender

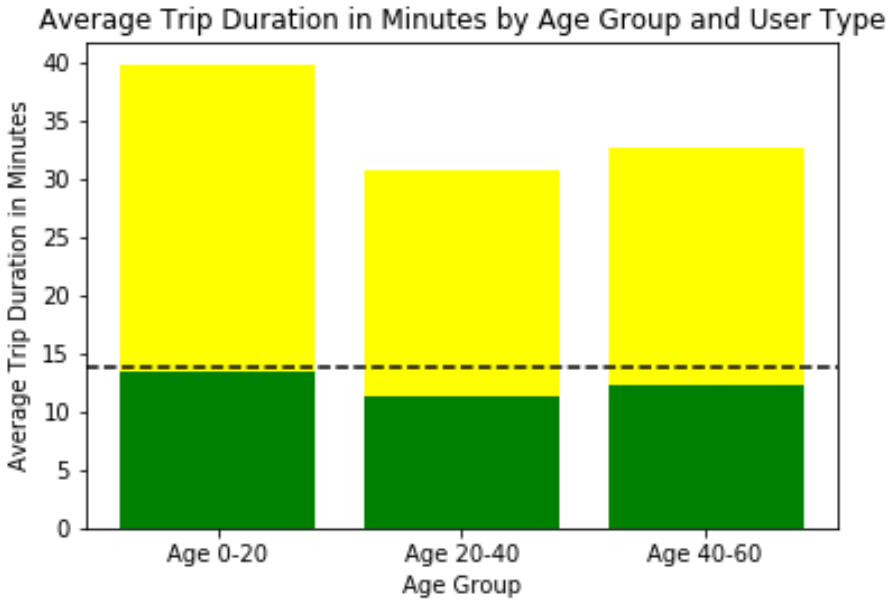
From these two plots, the red dotted line represents the average trip duration for females and the black dotted line represents the the average trip duration for males. The gender is being compared against designated user type found on the x-axis. The bars represent the overall averages of those user types.

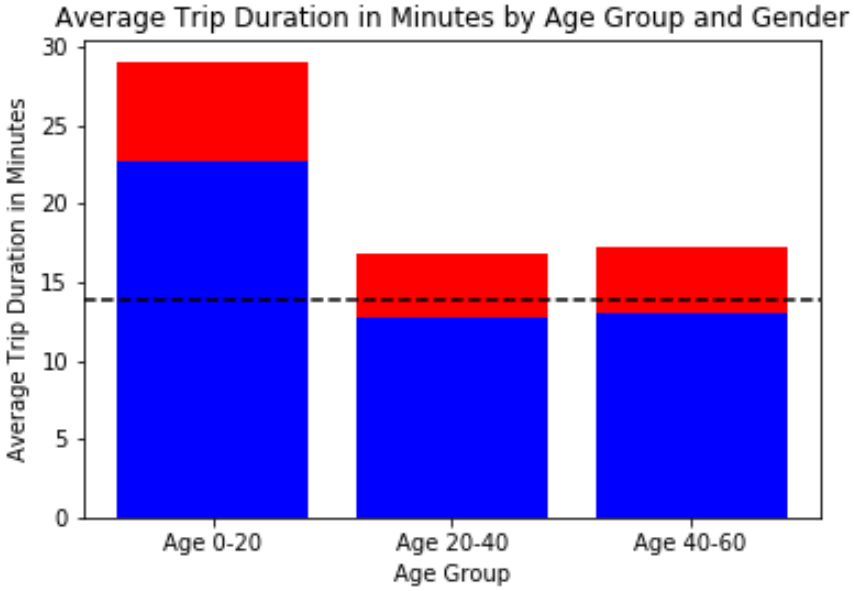
In connection to gender to user type with the average trip duration, there seems to be some connection. Given these two visuals, they nearly look the same in which females have a higher average than males even with disregard to the scale on the y-axis.

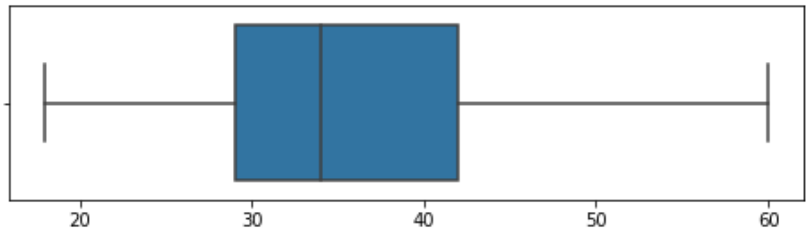
With customers, females have a higher average than males and this goes for subscribers as well. This observation supports the finding that females generally have a higher average than males and that customers have a higher average than subscribers. Therefore, it can be suggested that average trip duration may depend on gender and user type. However, the fact that there was an uneven distribution from the start, it should be taken account that data findings may be skewed to those with lower representation such as females and customers.











Average Trip Duration by Age Group by User Type & Gender

The top bar plot focuses on averages of age group and user type to overall average. The bottom bar plot focuses on averages of age group and gender to overall average. The heights of the colored bars represent their averages, the color coding is *subscribers*: green, *customers*: yellow, *females*: red, and *males*: blue. The dotted black line represents the overall average of 13. 82 minutes in both plots. The boxplot is also given to show the breakdown of ages of the dataset in which most are under the ages of 30 to 45.

Given these visuals, it could be suggestive to say age has some role as seen with the age group of 0-20 and how it is shown significantly higher than the other two age groups. Age 20-40 and 40-60 seem to be similar with their averages among user type and gender. Once again, customers and females have higher averages in each age group, but this could be from its low representation in the dataset. Generally, average trip duration seems to be dependent on user type, age group, and gender. It was found that longer bike rides were taken by customers, females, and those in age group 0-20.