Dataset

This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area. I'm most interested in figuring out what features are best for predicting the business revenue and asset(bike) depreciation by referring to the duration time and distance. In addition, I am also interested in finding marketing points that are likely to improve efficiency and effectiveness in the business. I expect the type of users will have the most potent effect on duration and distance because subscribers are more likely to use the bike than customer's along long distance. I also think this will be variable depending on the day of the week.

Summary of Findings

Most of the stations are located in three densely populated zones around San Francisco Bay. I assume that the primary purpose of using the bike is commuting from home to the office or neighborhood shopping, which is not just for exercise. Duration forms almost standard distribution, where the average time is about 12.1 minutes. This could possibly mean that my initial assumption on the purpose of use is correct. Because the average use time is less than 10 minutes, proportionally, the average distance is 1.69 kilometers. Most users park the bike where 1 to 2 kilometers away from the initial station. About 2.1% of checked-in users did not use the bike. Interestingly, non-subscribers used longer distances than subscribers, and so do those who do not use the bike-sharing services during the entire trip. Subscribers are almost eight times as many as non-subscribers, and those are mostly male who is not sharing the bike during their entire trip. The duration per member's birth year forms a normal distribution to the 8.6-minute point; however, the distance plot is skewed to the left. This possibly explains that the age group between mid-'20s and mid-'30s are remarkable active users.

Key Insights

The subscribers less used the bike than non-subscribers, which is probably a vital marketing point because we can increase the profit margin, revenue minus cost, with the membership contract. Although we figured out there are fewer users during the weekends, but they used more time. With these findings, we possibly recommend last-In bikes(high-tier inventory) to first-Out(serve) to weekend users with premium memberships.

List of resources during the creation of the project

Stackoverflow and GitHub repositories