

ANALYZING BAY WHEELS BIKE SHARING (2018 QUARTER IV))

by Ayoade Olayiwola

Dataset & Data Overview

In this investigation of the Bay Wheels Bike Sharing System (formerly Ford GoBike System), I set out to examine the characteristics of riders as it relates to trips taking and their use of the bike sharing service.

The dataset used contains monthly individual trip data from October 2018 to December 2018 in CSV format covering the greater San Francisco Bay area. Though data exists for every month, I chose to work with data for the last quarter of 2018. When combined this dataset contains 446,956 rows and 16 columns, each row representing an individual's trip records on 16 variables. The dataset contains information about the trips, the station and the members. The dataset hosted on the site (<https://www.lyft.com/bikes/bay-wheels/system-data>) has excluded some variable which are important to my investigation. I was however able to locate a dataset with the missing variable. The project leaves room to source your own dataset.

I performed an exploratory data analysis on the dataset using Python data science and data visualization libraries to explore the dataset's variables and understand the data's structure, patterns and relationships. This was followed with exploratory analysis from simple univariate relationships up through multivariate relationships.

Summary of Findings

There was a lot more subscribers using the bike sharing system than (casual) customers overall

Both categories use the service the most in October of the quarter examined.

The predominant usage of the service is for work commute; this is validated on many grounds. Different usage patterns and riding habits are observed between the two types of riders. More usage is seen generally on weekdays than on weekends;

Subscribers tend to have much shorter/quicker trips compared to customers which makes subscriber usage more efficient.

Users of the service on Monday through to Friday are slightly older than those who ride on weekends. Usage of the service is seen to decline with age

Key Insights for Presentation

Monthly usage of subscribers and customers which shows a lot more subscribers than customers who use the service

Distribution of Trip Durations which shows subscribers make shorter/quicker trips than customers

Usage by hours shows peak usage at work commute hours

Distribution of User Age: In the case of age, the distribution is more concentrated between 25 to about 40 years old.

Resources Used and websites visited

<https://stackoverflow.com/questions>

<https://www.kaggle.com/residentmario/plotting-with-seaborn>

<https://www.drawingfromdata.com/setting-figure-size-using-seaborn-and-matplotlib>

<https://github.com/nbalotaibi>

<https://seaborn.pydata.org/generated/seaborn.FacetGrid.html>

<https://www.programcreek.com/python/example/96204/seaborn.FacetGrid>

<https://github.com/asajm>

<https://seaborn.pydata.org/generated/seaborn.boxplot.html>

<https://www.earthdatascience.org/courses/intro-to-earth-data-science/file-formats/use-text-files/format-text-with-markdown-jupyter-notebook/>