Dataset

The project is based on the Baywheels data set, available here: <https://www.lyft.com/bikes/bay-wheels/system-data>

The data details e-bike rides taken in the San Francisco Bay Area every month. Feel free to look at <https://www.visualization.bike/baywheels/od/2020/> to get familiar with the data set. There are several Github projects looking at the data as well.

R, Python or SQL can be used to perform the analysis. Please attach your answer with programming code.

Candidate can submit answers in a public (or privately shared) Github repo.

1. Write code for the following and show the tabular output (B, C, D)
   1. Find the station id with the most ride minutes by month
   2. Find the origin station with the most originations (count).
   3. For each week of the year, derive the total and average ride durations for casual and members each.
   4. What are some of the most popular trips (start station/end station pairs) that most members took?
2. Using a BI tool such as Tableau Public, visualize the following: (B, C, python notebook for D ok)
   1. On the San Francisco Bay Area map, show the intensity of rides (borrows) at every station. A drop down should enable selection of the month.
   2. On the SF Bay Area map, show the top 5 stations by rides initiated. Allow selection by month, subscription status
   3. Visualize week over week/month over month change in rides. Allow selection by station, member status.
3. Exploratory data analysis (D)
   1. Perform some statistical tests for the categories (member status, bike type)
   2. What is the distribution of categorical and discrete variables like user type, bike type?
   3. What is the optimal number of bikes in the system?
   4. Any obvious patterns and correlations? (Time of ride, casual vs member )
   5. Histogram of time (hour) of ride start and station
   6. Distribution of minutes per ride
   7. Should there be more bikes at certain stations at certain times of the day? (Assume that the number of bike slots at a station is the peak number of rides that day)