

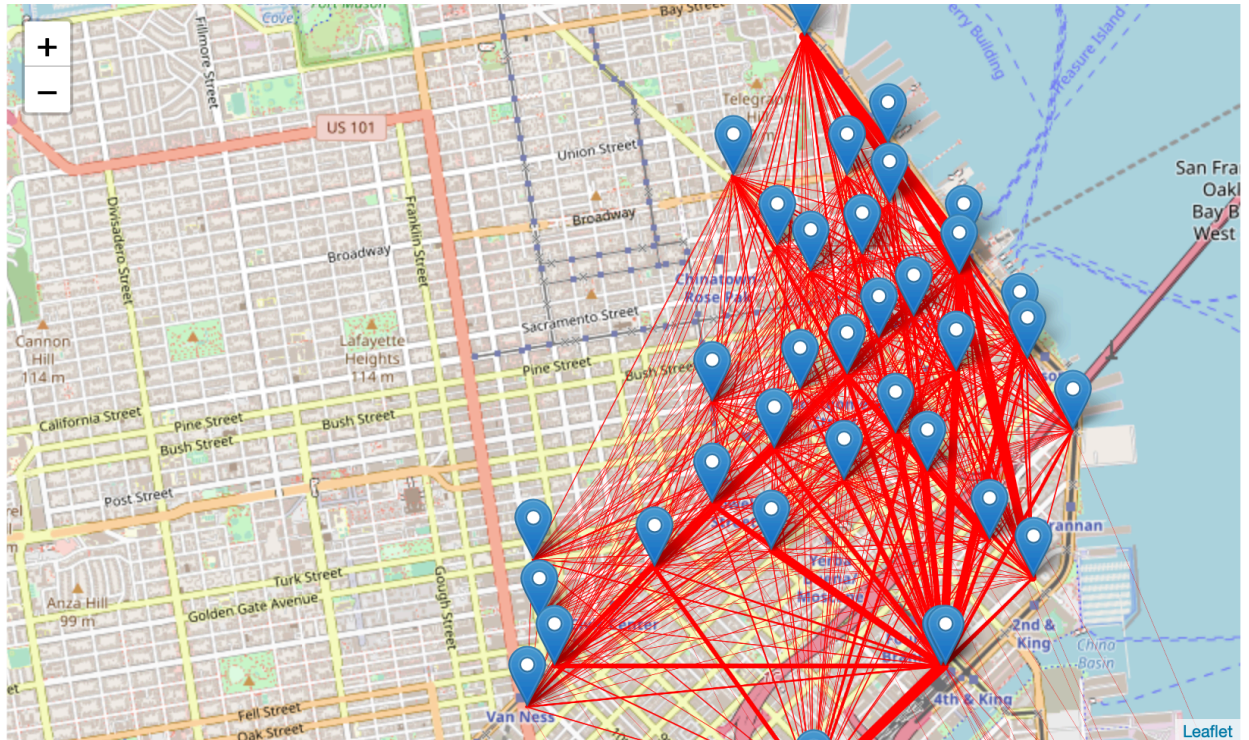
**Final Project**  
**STAT 5329/6329**

**Please do the following using the Python programming language.**

1. Download the four datasets (stations.csv, status.csv, trips.csv, weather.csv) from the Kaggle website <https://www.kaggle.com/datasets/benhamner/sf-bay-area-bike-share>.
  - a. From trips.csv data, which records each and individual trips, construct an aggregated data that includes daily rides for subscriber and daily rides for customers for each day and each station id. For other numerical variables such as duration, please compute their daily averages for each day and station id. Please name this data as “dailyrides”.  
**[hint: you can use group by commands]**
  - b. Convert the status.csv to daily data by creating new variables “avg\_no\_of\_bikes\_available” (daily averages of # bikes\_available for each station\_id) and “avg\_no\_of\_docks\_available” (daily averages of # docks\_available for each station\_id). Please name this data as “dailystatus”
  - c. Merge “dailyrides” with “dailystatus” and with the other datasets and create a consolidated single dataset.
  - d. Create new variables “weekday”, “month”. From the variable “installation\_date”, please create a variable “timesinceinstall” which is the time difference (in days) from the time of installation to current date.
  - e. Using the longitude and latitude, please plot the location of the stations on the map.
  - f. Please plot the total number of rides (in log(x+1) scale) for different visibility levels for both subscribers and customers. Interpret the plot.
  - g. Please plot the total number of rides (in log(x+1) scale) for different weather events for both subscribers and customers. Interpret the plot.
  - h. Plot the daily rides (as a time series plot) for both subscribers and customers. Interpret the plot.

**2. Please do the following using R programming language**

- a. From trips.csv data, for the city San Francisco, please draw a network diagram to show the busiest bike riding routes on the map. For example, please see the below picture. Can you identify the top 2 busiest routes?



**b.** Please repeat step (a) for the remaining cities.