

# Data Science Intensive Capstone Project Proposal

Chang-Heng Wang

- **Topic: Bay area bike share program analysis**

- **Introduction:**

The bay area bike share program offers a fleet of 700 bikes and 70 stations spanning the San Francisco and San Jose region, allowing users to freely rent bikes from and return to any station. With the official bike trip dataset, we would like to explore the behavior of bike share program users and gain insights toward operations that could improve the user experiences.

In particular, the objectives of this projects are as follows:

- Understand how usage profile relates with times of day, weather conditions, purpose (commute/leisure) etc.
- Identify stations that constantly involves bike or dock availability issue and predict the occurrences
- Understand how pricing plans (annual or short-term) affect the user behavior
- Suggest sites for expansion based on popularity of the neighborhood

- **Intended clients:**

We believe that the bike share system should be a synergy between the system operator and the users. Therefore, the target clients of this project should also be both the system operator and the users. For the system operator, we expect them to use the result of this project to make efficient operational decisions toward bike/dock availability issue, and to determine sites for system expansion. For the users, we would deliver suggestions of which stations to pick up and return a bike based on their location and the predicted bike/dock availability. We expect to enhance their experience with more accurate prediction (at the time they arrive at the station) instead of mere current bike/dock availability.

- **Reference data sources:**

- Dataset on Google BigQuery:  
<https://cloud.google.com/bigquery/public-data/bay-bike-share>
- Dataset from bike share program: <http://www.bayareabikeshare.com/open-data>
- Google maps API: <https://developers.google.com/maps/>
- Google places API: <https://developers.google.com/places/web-service/intro>

- **Solution approaches:**

The intended solution approaches would include the followings:

- From combining the trip data and the weather data, explore how the time of day, weather, etc. affect the usage of bike share system
- Find the occurrences of bike/dock absence issue, and determine the main factors that lead to the occurrences
- Use better machine learning model to predict the bike/dock availability for each station
- Build web app or mobile app to show the recommendation of start and end station (based on user query), and use the collected user queries to further refine the performance of prediction

- **Deliverables:**

The deliverables would include the followings:

- Report paper for the project
- Slides that promote the project
- Codes for data analysis and machine learning
- Web app or mobile app that shows real time recommendation for users