

Bay Area Bike Share.

The Bay Area Bike Share scheme is a network of 70 docking stations containing 700 bikes available to rent to the public. Anyone can pick up a bike in one station and drop it off at another. Distributed throughout San Francisco, Mountain View, Palo Alto and San Jose the bikes are intended to provide a cheap, fun and healthy form of transport to residents and tourists.

Each docking station contains a limited number of bikes and docking bays. *Motivate*, the company which operates the Bay Area Bike Share currently operate a two-to-one policy of bikes to bike-docking places which they hope ensures that docking stations always contain available bikes and available docking bays 365 days a year and 24 hours a day.

In 2015, *Motivate* announced that the system will undergo a ten-fold system expansion from 700 to 7000 bikes. They plan to roll this expansion out over a two year period after public consultation and technical siting analysis.

We will analyse usage statistics available at <http://www.bayareabikeshare.com/open-data> which contains bike number, trip start/end date and time, trip start/end station, whether the rider is an annual or 3-day-pass subscriber, zip-code of the annual members, weather information and bike and dock availability. Using this data we hope to understand:

- How is the network currently used? Which are the most popular routes and docking stations? When is it most busy?
- How do residents and tourists use the system? Are there any differences?
- Does the two-to-one policy work? Are there ever full or empty docking stations?

We hope to be able to suggest to Motivate:

- Where to build new stations in order to accommodate the expansion.
- What times are best to perform necessary maintenance on the bikes and stations.
- Liaising with the local authorities; where could cycle paths be built in order to make the system safer.

We will use R packages `dplyr` and `tidyr` to clean, sort and initially analyse this data. R's `ggplot` package will assist in visualising and presenting the statistical analyses. Using graph analysis, rank algorithms and clustering algorithms on the data set will help us find the busiest docking stations.

We will deliver our code used during the research and our findings and suggestions in the form of a report and slides to be presented.