**Capstone 1 proposal** by Eunice Kim

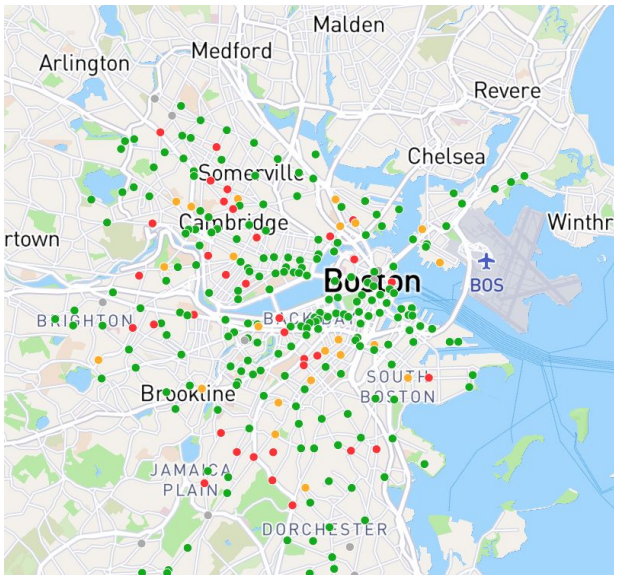
Boston Bike share system:

demand forecast and recommendation of locations of new bike stations.

## 0. Background

Blue Bike is a bike share system in the Boston metropolitan area. As of May 2019, the system deploys over 280 stations with more than 2000 bikes. However, the system only covers Boston city proper and a few surrounding cities--Cambridge, Somerville and Brookline--located in the northwest and southwest of Boston.

It would be beneficial for the community if Blue Bike expands to more neighborhoods. We turn our attention to the north and northeast of Boston, especially Malden, Everett (not visible from the map, but it is located right below Malden), and Revere. These cities are located near the existing Orange and Blue subway lines of the city. This is a good indicator that there will be great demands for Blue Bike as there are enough population and traffic flow in the areas already.





On the other hand, as Boston's old nickname, the Athens of America, suggests, Boston and its surrounding cities are home to numerous universities and colleges. According to Boston Planning and Development Agency, Boston city proper has 29 colleges and universities with the total student enrollment close to 140,000. This is approximately 20% of the Boston population, 685,000, and it is not a negligible number. Including the nearby cities, the numbers grow even higher.

Naturally, the number of residents in the Boston metropolitan area fluctuates based on academic seasons. Many students leave in mid-May and return in mid-August. We hypothesize that this seasonal migration affects the number of Blue Bike users.

## 1. Problems

Problem 1: Combining the current bike docking station and the subway station geo data, we will find the relationship between the two. From the trip history data, we will examine if bikes are used to commute to and from subway stations. We will then propose suitable locations for future bike docking stations near the Orange and Blue subway lines.

Problem 2: We will develop a model explaining the usage pattern. In particular, with the weather and academic seasons taken into account, we will predict when and which docking stations are prone to be low on inventory.

## 2. Target clients

Our clients will be Blue Bike.

Blue Bike is currently owned by Boston, Somerville, Cambridge, and Brookline. Blue Bike may use our results to expand their business in the north and northeast of Boston.

The company will be able to better predict the customer demands. Understanding varying demands based on the weather and seasons can help them rebalance their bikes more efficiently.

## 3. Data Collection

Most of the data is readily available in CSV, txt, pdf files or through API. Here are the links to the data.

1. Monthly trip history (2015-present), geolocation of docking stations as of 2017 can be obtained from:

https://www.bluebikes.com/system-data

2. Weather information:
https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00014739/detail

3. A list of Boston Subway stations with geolocation and other information:

https://www.mbta.com/developers

4. Number of enrollments in universities and colleges in Boston:

http://www.bostonplans.org/getattachment/3e8bfacf-27c1-4b55-adee-29c5d79f4a38

## 4. Tentative approach

1. Gather weather data from https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00014739/detail and align it with bike trip history data.
2. Classify each day with the weather information (Temperature, no Rain/snow, light rain/snow, heavy rain/snow)
3. Gather academic season information (Off: mid-May to mid-August & mid-Dec to mid-Jan, On: other time)
4. Classify each day as a working day or offday (Use federal/Mass holidays)
5. Find the usage patterns in Pandas
6. Create map based visualisations on (such as cartodb.com ) to plot the location of new stations.

## 5. Deliverables

We will present our results and visualizations in slides. We will also attach the codes in Jupyter notebook.