Alana King

U06232647

BU CS591 L1 Project Report

I. Introduction

My project focused on transportation within the city, and its usability. I began by looking at time efficiency compared to cost, and found that biking is one of the most cost efficient modes of transportation. This encouraged me to look into Hubway, a public bike share program with over 185 station in the greater Boston area. Like all transportation, however, it does have downfalls. One major issue with Hubway is the limitations of the stations. Each station has a set amount of docks for the bikes. With limited bikes and limited docks at each station, it is possible to arrive at a station to find that there is no room to dock the bike you had rented from another station, or arrive at a station to find that there are no bikes at all. This leaves users with no choice but to walk to another station if they want to use or return a bike. In some cases, the next closest stop could be over a mile away, defeating the purpose and efficiency of using one of the bikes. My project uses data collected by Hubway to help minimize the chances of one of the previously described situations from happening.

II. Data Sets

Hubway provides a lot of data in the form of CSV files, so the majority of my project used this data

a. Hubway Stations:

   http://datamechanics.io/data/aking17/hubway_stations.csv

   This data set contained information on all the Hubway stations in Boston. The data included a station ID number, station name, the municipality in which it is located, latitude and longitude of the station, and whether or not it is still in use.

b. Station Status

   Stationstatus.csv

This data set contained over 30 million entries on 49 popular stations within Boston. The data set logged updates on the number of bikes available, and the number of empty docks available at every hour of the day. With millions of hours of data, this set was one of the most useful in determining which stations were problem stations.

c. Station Availability

Station_availability.csv

This data set uses the information from the station status file. I parsed the data so that each station had a clear log of the amount of bikes at the station, and the capacity of the station, for every hour of the week. This made it easier to map the station availability to the physical location of each data point. Each station has a column containing a tuple in the form of (number of bikes available, number of docks available) for every recorded hour.

d. Hubway Trips

Hubway_trips.csv

This file contains data from trips that specific users took. The data contains information such as the user ID, duration of the trip, the date, the start and end time, the start and end station, the bike ID number, user subscription type, user zipcode, user birthdate, and gender.

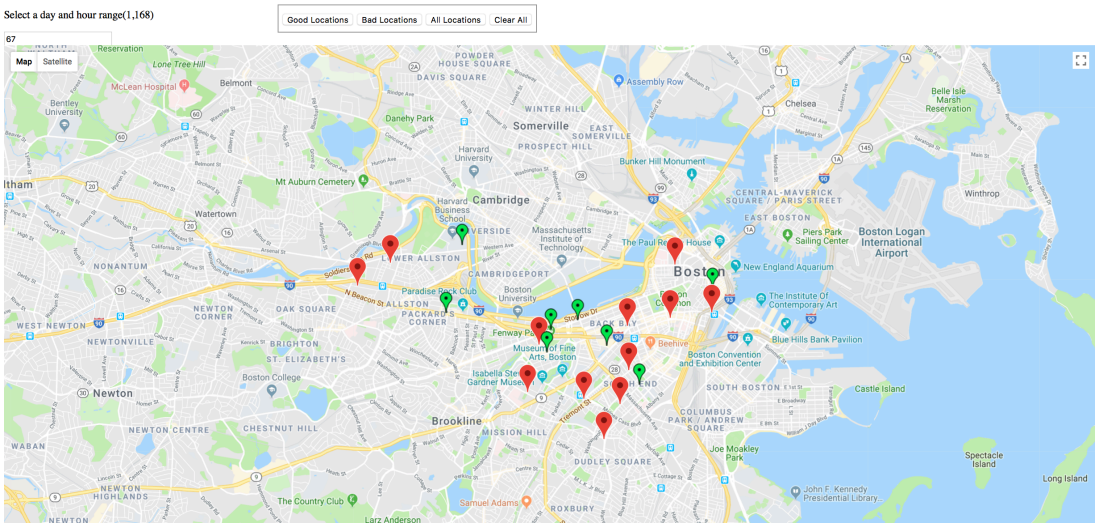
e. Good Bad Trips

Goodbad.csv

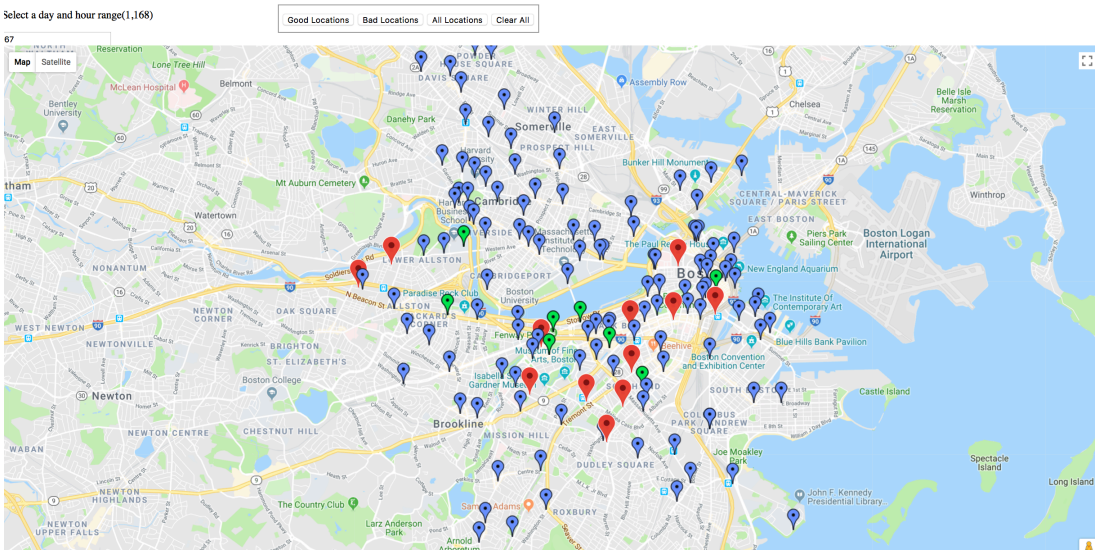
This is a file of the parsed stationstatus data. It contains two columns in the form of a list of the “good” stations and a list of the “bad” stations at every hour. A good station is where 75% of the docks have bikes and a bad station is where only 25% of the docks have bikes.

III. Plotting the Data

I used the google maps API to create a map with the capabilities of plotting all of the stations longitude and latitude on a map in order to have a better visualization of the most popular stations. Then I parsed the data to find the stations that had 75% or more of the docks full and marked those as stations that you were likely to get a bike, but may be hard to return a bike to within the next hour. Those stations are marked in green on the station map. I also located the stations that only had 25% or less of the docs full, making them an unreliable place to get a bike within the next hour, but easy to return a bike to, and marked them in red. A map of all the stations are plotted in blue. These maps can be generated for any hour of the week.



Map of Boston with the good and bad stations displayed for the given time.



Map of Boston with all the stations displayed.

To create this web interface, I stored my data in a mongodb, and used flask to connect it to the web interface. Each button calls on information from the database and uses the results to set a marker for the proper location. By setting the hour in the top left corner of the screen, the website can then call the proper information for the good, and the bad stations and create a visualization of their locations.

IV. Minimizing Travel

From this information, it is not hard to make an informed decision about where to pick up or drop off a bike. I parsed the data and found the next closest location for each station so that a user can easily find a place to dock or pick up a bike. The furthest a user would have to go to get to the next station is approximately 1.2 miles, and on average the user would only need to travel 0.27 miles.

V. Conclusions

Hubway's extensive amount of data makes it easy to track trip and station statistics. This allows the company to determine which aspects of their service need to be improved in order to improve user experience and popularity of their services. With over 185 stations, over 1800 shared bikes, and plans to expand, it is important that they are constantly interpreting their data so they can better serve the growing demand, and the complications that come with it.

The space and bike limitation problem is an open ended problem because as the company's user base expands, this problem will continue to occur. Since there are limitations on space and resources, there is no definitive way to solve this problem. Their stations rely on users constantly redistributing bikes. However, some locations are more popular to travel too. In addition, the bikes are used for commuting. Many users want to use the bikes to commute into the city in the morning, and back to the more residential areas in the evening, putting strain on those stations at the given times.

The station use also varies due to weather, as it is more appealing to ride a bike in nicer weather. If I had more time, I would use a weather API to determine how much the weather effects the bike usage. As Hubway continues to extend their stations hours into the colder and rainier seasons, this will become a larger issue. Again this would be an open ended problem, because Boston has unpredictable weather.