Taking Gender Equity to the Streets

CS504: Data Mechanics

Kerin Grewal and Stephanie Shin

Introduction

This project, founded through Spark! and the The Mayor's Office of Women's Advancement at the City of Boston, was aimed to create more female representation within Boston through the various street names around the city. Boston's current demographics include 52% women, yet street names do not reflect this equal divide. We wanted to see what the best candidates were for renaming streets within the 23 different neighborhoods of Boston. By working on this project, we hope to help evolve and create a more inclusive and representative city to live in. We used many different developer tools to create this project including PyCharm, Python 3.7, Google APIs, Flask, MongoDB, and more.

Datasets and Data Transformations

The data sets we utilized are as follows:

- Boston Street Names (boston_street_names.json - datamechanics.io)
    - This data set was collected from last semester's project as a CSV and was uploaded to datamechanics.io as a .json file. It is a list of all the street names in Boston, their genders, their zip codes, and their ranks (a number that describes the length of the street (based on how many zipcodes it covers). This was the base of our data as it is a list of our street names.
    - This data was uploaded to our repo database as kgrewal_shin2.street_names

	full_name	street-name	name-length	gender	gender2	zipcodes	rank
1	A Street,	A	2	unknown	F	2136	1
2	A Street,	A	2	unknown	F	02210-02127	2
3	Abbot Street,	Abbot	6	male	M	2124	2
4	Abbotsford Street,	Abbotsford	11	unknown	M	2121	1
5	Abby Road,	Abby	5	female	M	2135	1
6	Aberdeen Street,	Aberdeen	9	female	M	2215	1
7	Acacia Road,	Acacia	7	female	F	2132	1
8	Academic Way,	Academic Way	12	female	M	2134	1
9	Academy Court,	Academy Court	13	unknown	F	2119	1
10	Academy Road,	Academy	8	male	F	2119	1
11	Academy Terrace,	Academy Terrace	15	unknown	F	2119	1
12	Academy Hill Road,	Academy Hill	13	unknown	F	2135	3
13	Acadia Street,	Acadia	7	female	F	2127	1
14	Achorn Circle,	Achorn Circle	13	unknown	M	2130	1
15	Ackley Place,	Ackley	7	male	F	2130	1
16	Acorn Street,	Acorn	6	unknown	M	2108	1
17	Acton Street,	Acton	6	male	M	2136	1
18	Ada Street,	Ada	4	female	F	2131	1
19	Adair Road,	Adair	6	unknown	M	2135	1
20	Adams Park,	Adams Park	10	male	M		1
21	Adams Place,	Adams	6	male	M	2127	1

- Boston Landmarks (boston_landmarks.json - datamechanics.io)
  - This data was collected from the Boston Landmarks Commission (BLC) at http://bostonopendata-boston.opendata.arcgis.com/datasets/7a7aca614ad740e99b060e0ee787a228_3 as a CSV and was transformed and uploaded to datamechanics.io as a .json file. This gives us the names of all the landmarks in Boston, the address (which includes street name), and the corresponding neighborhood.
  - This data was uploaded to our repo database as kgrewal_shin2.landmarks
- Boston Neighborhoods (.geojson)
  - This data was collected from https://raw.githubusercontent.com/codeforamerica/click_that_hood/master/public/data/boston.geojson. It includes the geographic boundaries of each neighborhood in Boston.
  - This data was uploaded to our repo database as kgrewal_shin2.neighborhoods
- Boston Public Schools (.geojson)
  - This data was collected from https://data.boston.gov/dataset/public-schools/resource/6c48e501-3dba-44f3-912f-8a5f309d5df4. It includes the locations of all the public schools in Boston, their addresses, city, and zip code.

    - This data was uploaded to our repo database as kgrewal_shin2.pub_schools
- Boston Uber data (boston_common_ubers.json - datamechanics.io)
    - This data was collected from Uber's statistics website at [https://movement.uber.com/explore/boston/travel-times/query?lang=en-US&si=1116&ti=&ag=censustracts&dt[tpb]=ALL_DAY&dt[wd;]=1,2,3,4,5,6,7&dt[dr][sd]=2018-01-01&dt[dr][ed]=2018-01-31&cd=](https://movement.uber.com/explore/boston/travel-times/query?lang=en-US&si=1116&ti=&ag=censustracts&dt[tpb]=ALL_DAY&dt[wd;]=1,2,3,4,5,6,7&dt[dr][sd]=2018-01-01&dt[dr][ed]=2018-01-31&cd=). It was downloaded as a .csv and transformed and uploaded to datamechanics.io as a .json file. It includes information of Uber rides originating at the Boston Common, their destination location, mean travel time, and range of travel time. We chose to use Uber data originating at the Boston Common due to its central and popular location.
    - This data was uploaded to our repo database as kgrewal_shin2.ubers

Origin Movei	Origin Display Name	Destination Movement ID	Destination Display Name	Date Range	Mean Travel	Range - Low	Range - Upper Bound Travel Time (Seconc
290	Boston Common, 139 Tremont St, Boston,	3	0 Flaherty Way, D Street / West B	1/1/2018 - 2/28/	619	440	870
290	Boston Common, 139 Tremont St, Boston,	5	0 Queensberry Street, Fenway/Ker	1/1/2018 - 2/28/	510	343	757
290	Boston Common, 139 Tremont St, Boston,	13	200 Liberty Street, Braintree	1/1/2018 - 2/28/	1483	1088	2019
290	Boston Common, 139 Tremont St, Boston,	14	400 Cushing Street, Hingham	1/1/2018 - 2/28/	1753	1345	2284
290	Boston Common, 139 Tremont St, Boston,	16	100 Kenrick Street, Oak Square, Bc	1/1/2018 - 2/28/	1345	1026	1762
290	Boston Common, 139 Tremont St, Boston,	17	0 Maggi Road, Revere	1/1/2018 - 2/28/	1288	922	1798
290	Boston Common, 139 Tremont St, Boston,	18	0 Lake Street, Billerica	1/1/2018 - 2/28/	2197	1672	2886
290	Boston Common, 139 Tremont St, Boston,	19	200 Ocean Street, Lynn	1/1/2018 - 2/28/	1962	1492	2579
290	Boston Common, 139 Tremont St, Boston,	20	0 Florence Street, Everett	1/1/2018 - 2/28/	1338	947	1890
290	Boston Common, 139 Tremont St, Boston,	21	0 Canal Lane, Somerville	1/1/2018 - 2/28/	751	487	1157
290	Boston Common, 139 Tremont St, Boston,	27	100 Inner Belt Road, Inner Belt, Sc	1/1/2018 - 2/28/	654	406	1051
290	Boston Common, 139 Tremont St, Boston,	28	0 Lincoln Road, Brookline	1/1/2018 - 2/28/	1081	791	1476
290	Boston Common, 139 Tremont St, Boston,	29	100 Crosby Street, Arlington	1/1/2018 - 2/28/	1358	977	1885
290	Boston Common, 139 Tremont St, Boston,	30	0 Stagecoach Way, Hopkinton	1/1/2018 - 2/28/	2275	1809	2860
290	Boston Common, 139 Tremont St, Boston,	32	0 Gainsborough Street, Fenway/Ke	1/1/2018 - 2/28/	645	437	951
290	Boston Common, 139 Tremont St, Boston	33	200 Prospect Avenue, Revere	1/1/2018 - 2/28/	981	668	1439

- Boston Neighborhood Zip Codes (neighborhood-zipcodes.json - datamechanics.io)
    - This data was taken from a table on http://archive.boston.com/news/local/articles/2007/04/15/sixfigurezipcodes_city/. We took the data and transformed it into a CSV and then transformed and uploaded it as a .json file to datamechanics.io.
    - This data was uploaded to our repo database as kgrewal_shin2.neigh_zip

origin	neighborhood	zero	zipcode
2101	Downtown Boston	0	2101
2108	Beacon Hill	0	2108
2109	Markets / Inner Harbor	0	2109
2110	Financial District / Wharves	0	2110
2111	Chinatown	0	2111
2112	Downtown Boston	0	2112
2113	North End	0	2113
2114	West End	0	2114
2115	Fenway	0	2215
2116	Back Bay	0	2116
2117	Downtown Boston	0	2117
2118	South End	0	2118
2119	Roxbury	0	2119
2120	Roxbury Crossing	0	2120
2121	Roxbury	0	2121

- Boston Zip Codes and their Geographic Locations (MA_zip_codes.json - datamechanics.io)
  - We got this dataset from https://gist.github.com/erichurst/7882666/ and it contains all US zipcodes and their corresponding Lat and Lon. We only kept the codes from to Boston.
  - This data was uploaded to our repo database as kgrewal_shin2.ma_zip_loc

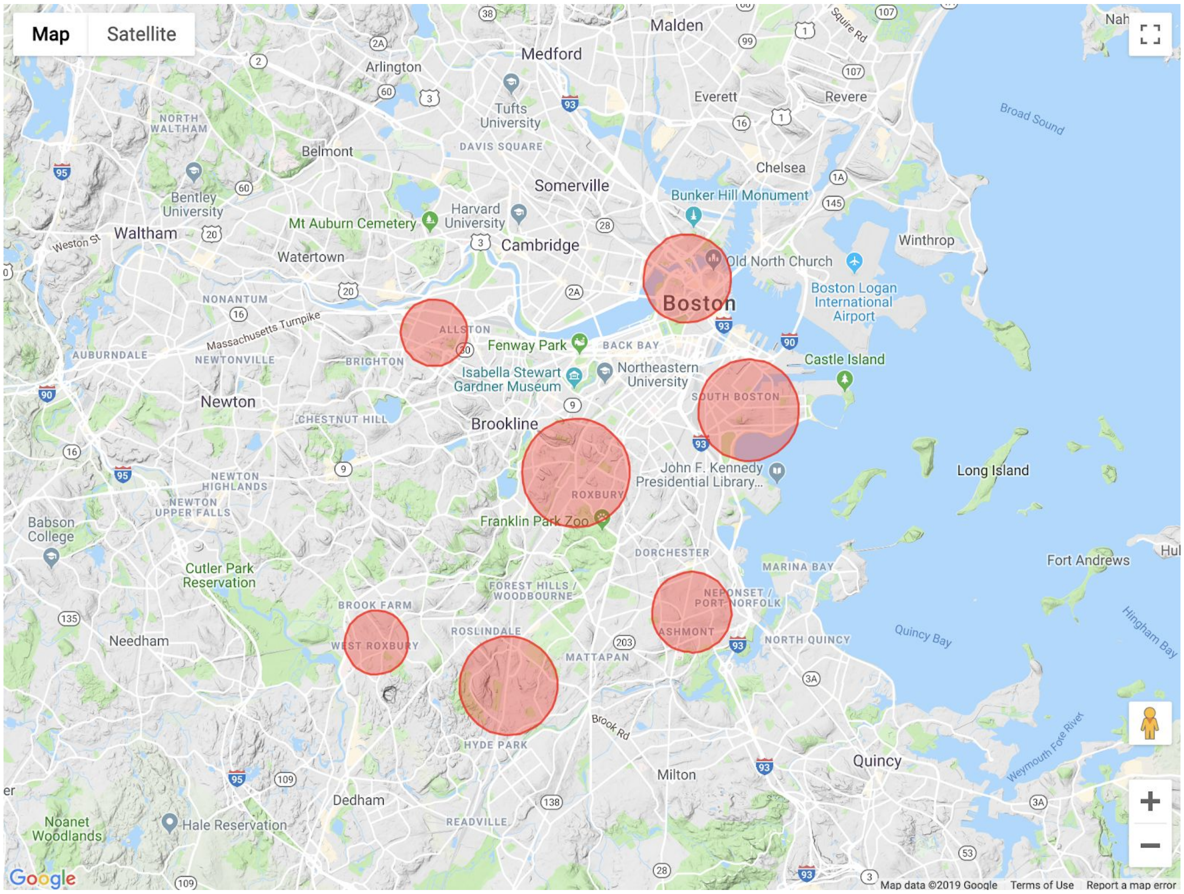
The data transformations we performed/data sets we assembled were:

- Streets without Landmarks
  - We combined street_names and landmarks in order to find streets that are not where landmarks are and selecting those that are only non-female names (unknown gender or male street names).
  - This data was uploaded to our repo database as kgrewal_shin2.streets_without_landmarks.
- Streets without Public Schools
  - We combined street_names and pub_schools in order to find streets that are not where public schools are and selecting those that are only non-female names (unknown or male street names).
  - This data was uploaded to our repo database as kgrewal_shin2.streets_without_schools.
- Uber Aggregates

    - We took the ubers dataset and performed an aggregate sum in order to locate the areas of Boston where ubers travel most frequently around the different areas of Boston from the Boston Common.
    - This data was uploaded to our repo database as kgrewal_shin2.neighborhood_ubers.
- Unclaimed streets
    - We took the streets without landmarks and streets without public schools data in order to find a list of “unclaimed streets”- streets that are not female gendered, at any Boston landmarks, or any Boston public schools. This data became our base for the streets that are the best candidates for being renamed.
    - This data was uploaded to our repo database as kgrewal_shin2.unclaimed_streets.
- Streets and their Neighborhoods
    - We combined the unclaimed_streets data and the neighborhood and their zip codes data in order to assign every “unclaimed street” to a neighborhood based off of their zip code.
    - This data was uploaded to our repo database as kgrewal_shin2.neigh_streets.

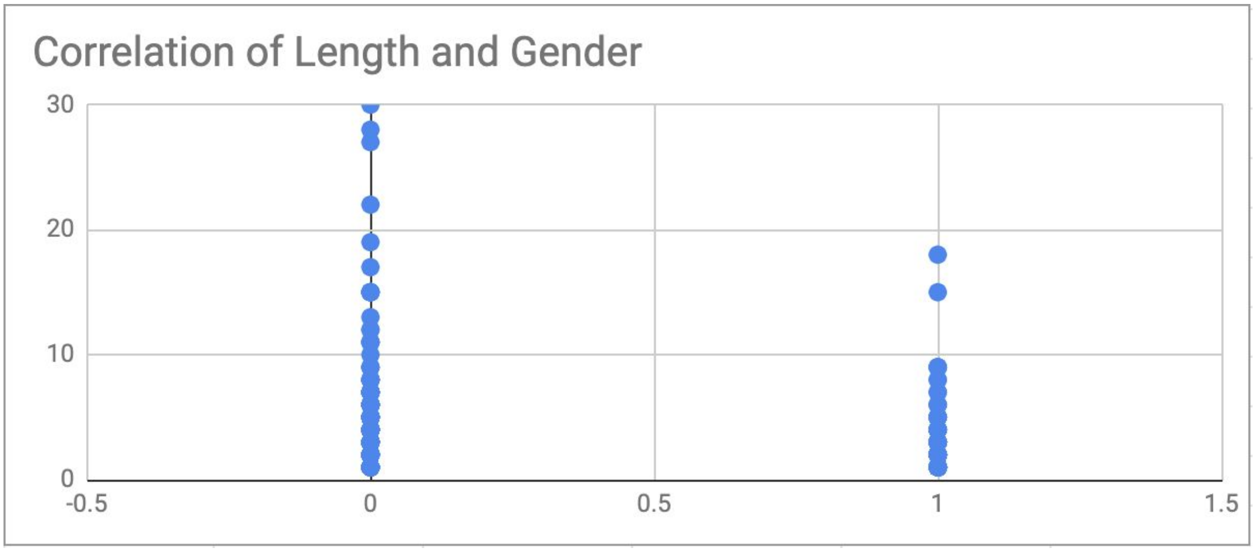
Optimization and Statistical Analysis

We wanted to find where the most common areas of “unclaimed streets” were in order to find the best areas for renaming streets. Unclaimed streets represent streets which, based on our analysis, do not have landmarks or schools on them. To do this, we implemented a non-trivial optimization technique- k-means. We performed k-means clustering on the unclaimed_streets data, then did a count to find how many streets existed in each cluster. We found that our unclaimed streets data was well spread throughout the city. The clusters were all of significant sizes, meaning that no particular area was lacking applicable streets. We mapped this data out on our final user



interactive web interface in order to more easily visualize where street names could most likely be renamed. The clusters are saved within our repo database as kgrewal_shin2.street_kmeans. One of the difficulties we had with using k means is that we can not keep the count in each cluster easily. In order to get a count of how many streets exist in each cluster we used a distance equation and added each street to the count of the cluster with the closest mean.

Furthermore, we wanted to discover if there was a relationship between street length and gender. In order to do this, we performed a statistical analysis of finding the bivariate correlation coefficient between the lengths of our unclaimed streets and the genders (setting male/unknown to 0 and female to 1). We hypothesized that longer roads were more likely to be “major roads” (assuming that major roads are longer roads) and also male gendered roads, and we believed that we would be potentially renaming more major roads instead of less-significant roads.



We found a correlation coefficient of -.0203697 and a p-value of .148325. Therefore, we found no correlation between length and gender. This means that the streets with the most potentially to be renamed will be of multiple lengths and with plenty of significant suggestions. This correlation coefficient and p-value were uploaded into our repo database as kgrewal_shin2.street_length_vs_gender.

Interesting Issues

Most of our issues stemmed from the datasets as we believe the most important part of solving this problem is cleaning and finding the correct data. For example, the public schools data does not necessarily cover universities in Boston (which there are many and plenty of).

Furthermore, we found interesting issues when looking into traffic data for Boston. After various research we found it hard to find a centralized location with an abundance of traffic information. Going to Ubers data, we found that we could only download data from a certain period of time with a starting destination and end destination or just a starting destination (which we utilized). However, this obviously does not show all the possible Uber data we would like to collect, as we only ended up using data with rides from the Boston Common. An even greater issue regarding Uber data is realizing that many different demographics and areas of Boston may not have access to Uber, a smart phone, the necessary financial situation, and many other factors that are required when calling an Uber. This data is most definitely biased towards the wealthier areas of Boston, and this is something we have to acknowledge and would try to offset in the future.

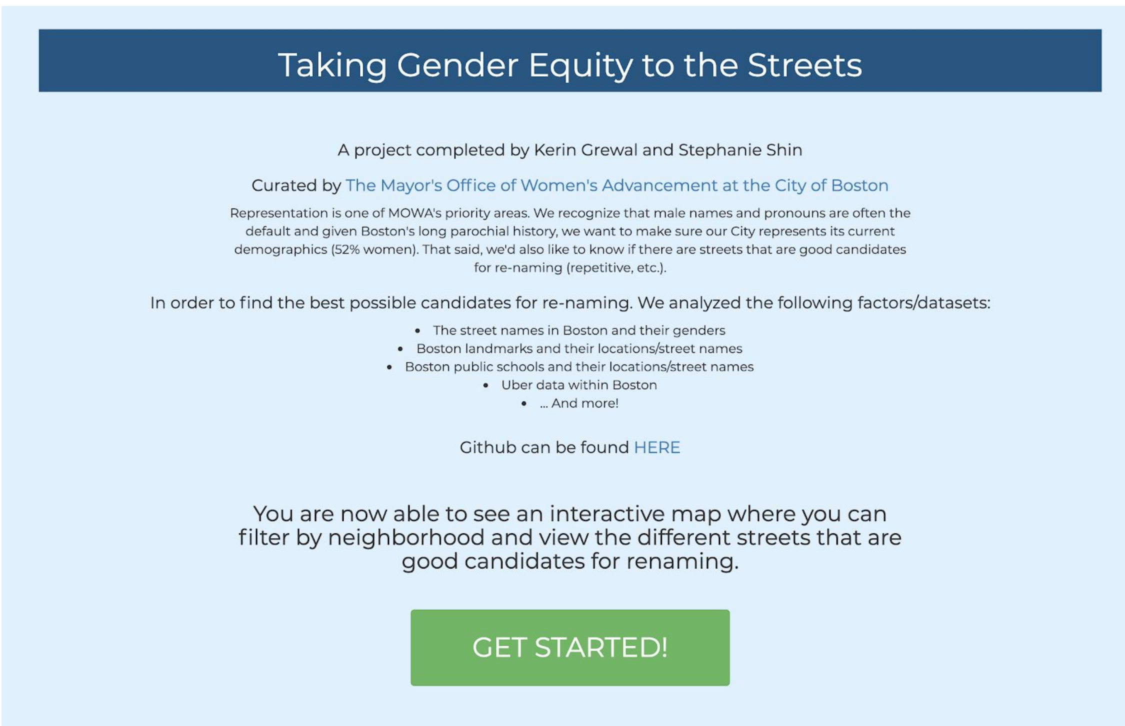
Regarding the usability of finding a correlation coefficient, this technique wasn't necessarily the most appropriate or relevant technique. This was more in order to find answers to curious questions, instead of furthering our main problem of finding streets to be renamed. Generally, with this specific problem, we had trouble finding specific statistical analyses to perform as it was mostly narrowing and selecting the data we originally found.

Generally, we found many issues within the limitations of the available data and the formats within them. For example, we found a dataset that included major roads in Boston and their rankings, but the data set was too large for us to handle (at the moment). With more resources and time, we would have been able to utilize another relevant data set. Most of the issues we developed were in response to being able to have much more data to influence the possible streets to be renamed and these would most definitely be addressed if we were to continue to attempt to solve this problem.

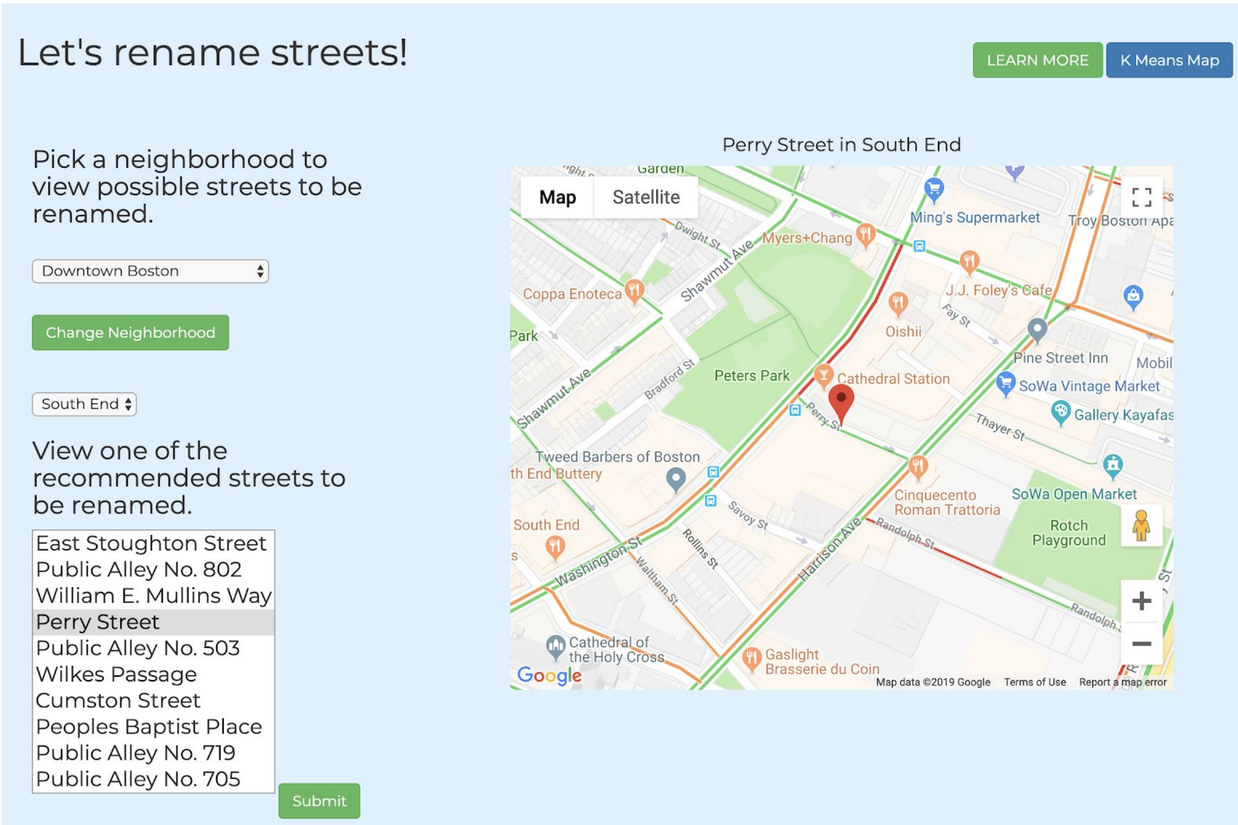
Summary

Our end result is an interactive web application where you are able to select neighborhoods and see a list of the streets that are good candidates for renaming and then able to locate them and view them on the map to visualize and assess the possibility of this street being renamed. We found that within the less central neighborhoods of Boston (such as Roxbury, Mattapan, Dorchester) there were many possible streets that could be renamed whereas in Downtown Boston, we had no possible options based on our datasets. However, this is not a definitive problem as this completely depends on the type of data used to come to these conclusions.

Here is the opening page for our web-based visualization:

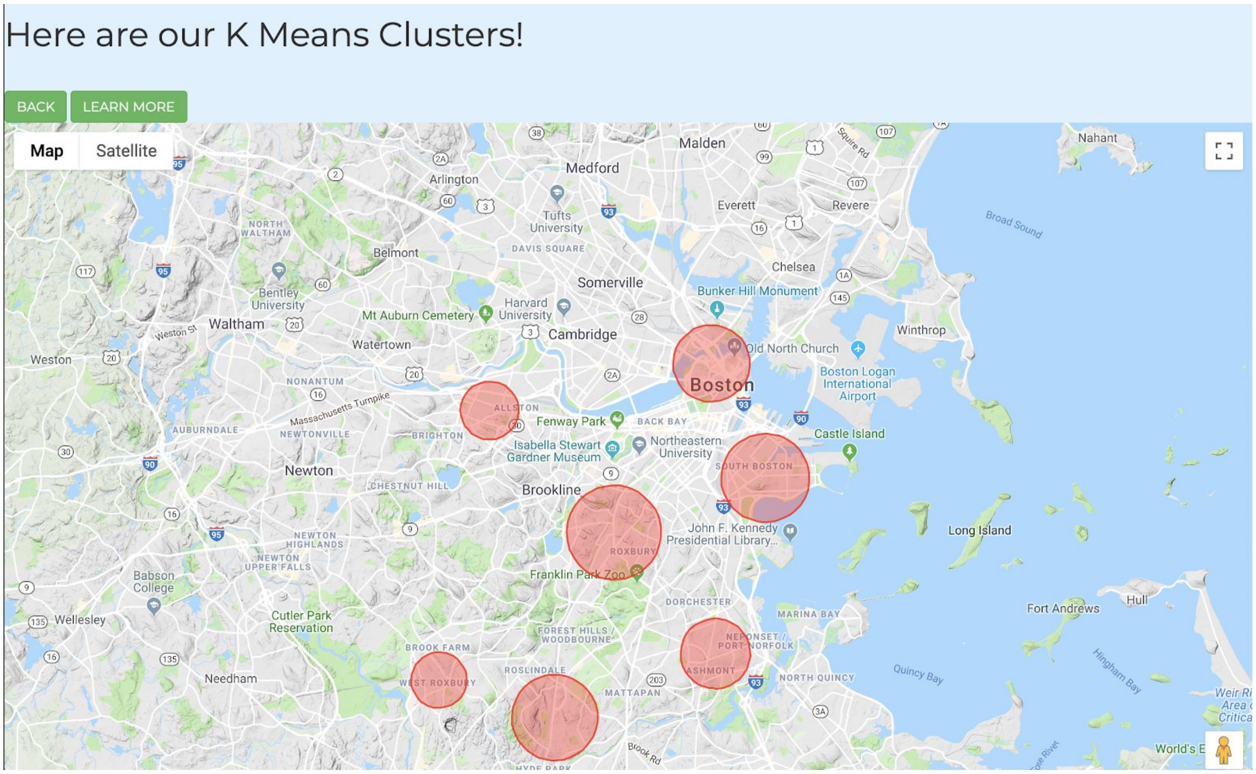


And the interactive part of the visualization:



This example shows Perry Street in the South End. The user views this after first selecting South End as a neighborhood and then Perry Street from the list of recommended streets to be renamed.

Here is the k-means map on our web-based interactive site:



Moving Forward

Regarding data sets, if we were to continue this project we definitely would have wanted to look into further traffic data (the T, car traffic in general in less central neighborhoods, walking data if possible, etc). We also would have loved to look into when streets were renamed and disregard the ones that were recently renamed. We found that the uber data did not give us enough information about traffic around the city because of the way in which it is provided. We were forced to choose a location where the ubers were either going or coming from as our base for the dataset.

Additionally, roads frequently traveled on like Storrow Drive will not be a frequent drop off or pick up location for an uber and will therefore provide insignificant data. If we could find an api that gives us traffic data for a given street, it would be more useful.

Furthermore, it would be interesting to make sure race plays a greater role in choosing possible streets to be named, more specifically, not renaming any streets that are named after a person of color. As there are limited methods and resources to decide whether a name itself is from a person of color or not, this would be a thin line to walk moving forward. This may be something that would have to be looked at by an individual instead of using an API or a data transformation, but those resources may not be available to do so.

Within the web application, we would also find it interesting to open it to the public in order to create some sort of voting or polling in order to see if other people would be interested in renaming certain streets (as this would affect different peoples lives throughout each neighborhoods community). Before we can do this, we must narrow down our list of recommendations even further so that each neighborhood in boston has around the same number of suggestions. If we had more time, we would pull in more information about famous athletes, colleges, the names of different politicians or public figures, and more that may be the basis for some of the street names we did not filter out in this project.