**CS 591 L1 Data Mechanics**

**Improving Public Schools in Boston**

**by Wenjing Lyu, Zetian Wu, Guangxing Ren**

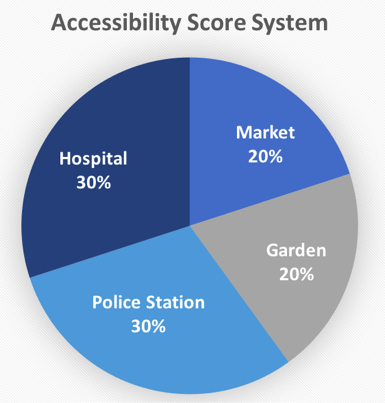
**Motivation**

Education is always the priority in Boston since high-quality education is in the heritage and spirit of the city. However, Public Schools students and parents are always complaining about the neighbourhoods and accessible amenities around schools areas. Therefore, we wish to provide improvement suggestions for public schools in different regions by evaluating the accessibility of each public schools in Boston. The reflection given by the result of our project might also be helpful for Boston government to decide potential good location to build new public schools.

**The Data Sets We Used**

Our Data Sets are from two sources: City of Boston and Boston Open Data. With Boston Public Schools dataset, we retrieve Location of Boston Public Schools for school year 2012-2013. Four accessibilities data sets are Location of Police Stations, Gardens, Markets, and Hospitals.

**Transformation 1: K-Means and Scoring Mechanism**

 We used K-means clustering algorithm to divide Boston public school into 5 regions by calculating 5 clustered central points and calculating which is the nearest central point for each school.To calculate the accessibility score, we only consider the accessibilities within 2 km range of each school. The accessibilities scores of each school are generated by the following equation:

Score = #Garden + #Market + 1.5 #Police Station +1.5 #Hospital.

We originally try to divide Boston public school into 6 regions, but 2 of them overlap. Then, we calculate the average score of each region and apply 90 percent confidence interval to test whether the score of certain region is significant below the average.

If a region has a score that significant below the average, the schools in that cluster should make contact with government to demand an increase of certain accessibilities around their neighbourhood.

**Analytical Results**

School Sample Size=133

Region Size=5

School Average score=56.03

School Std = 38.80

Region 1: School Number: 17 Average:40.2 T score:0.82 P value:0.44

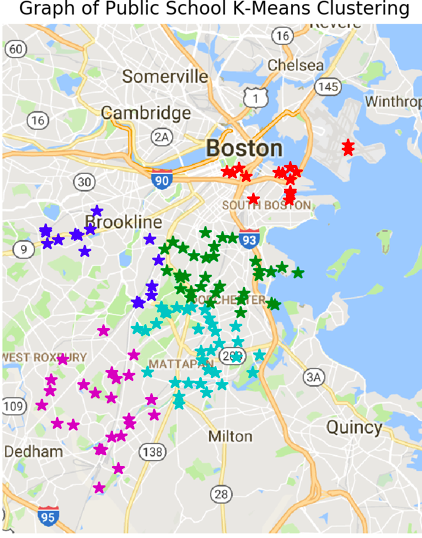
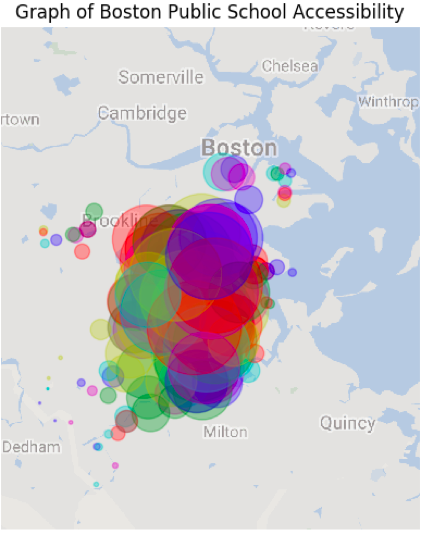
Region 2: School Number:37 Average:81.36 T score: -1.44 P value:0.22

Region 3: School Number:38 Average:78.65 T score: -1.26 P value: 0.27

Region 4: School Number: 15 Average: 29.06 T score:1.55 P value:0.19

Region 5: School Number:26 Average:18 T score:2.18 P value:0.094

Conclusion: Only Region 5 is significantly below the average, thus we suggest the schools in that region make contact with government to make improvement.

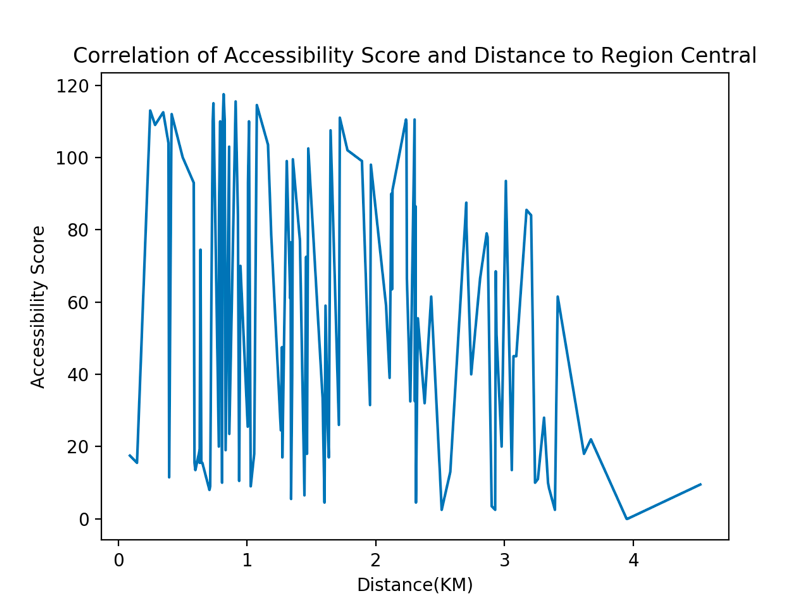


**Transformation 2: Correlation**

We calculate the correlations between the Distances( Distance between school and its cluster centra) and Accessibilities Score(the accessibilities score of individual school)of every public school in Boston to understand whether the the school closer to clustered center usually has higher accessibility. A high correlation coefficients(>0.5) means schools that close to its cluster central have more accessibility.

The results show that the correlation coefficient between distance and accessibility is -0.27 and P value is 1.0 which means that there is no correlation between the Distances and Accessibilities of Boston public schools.

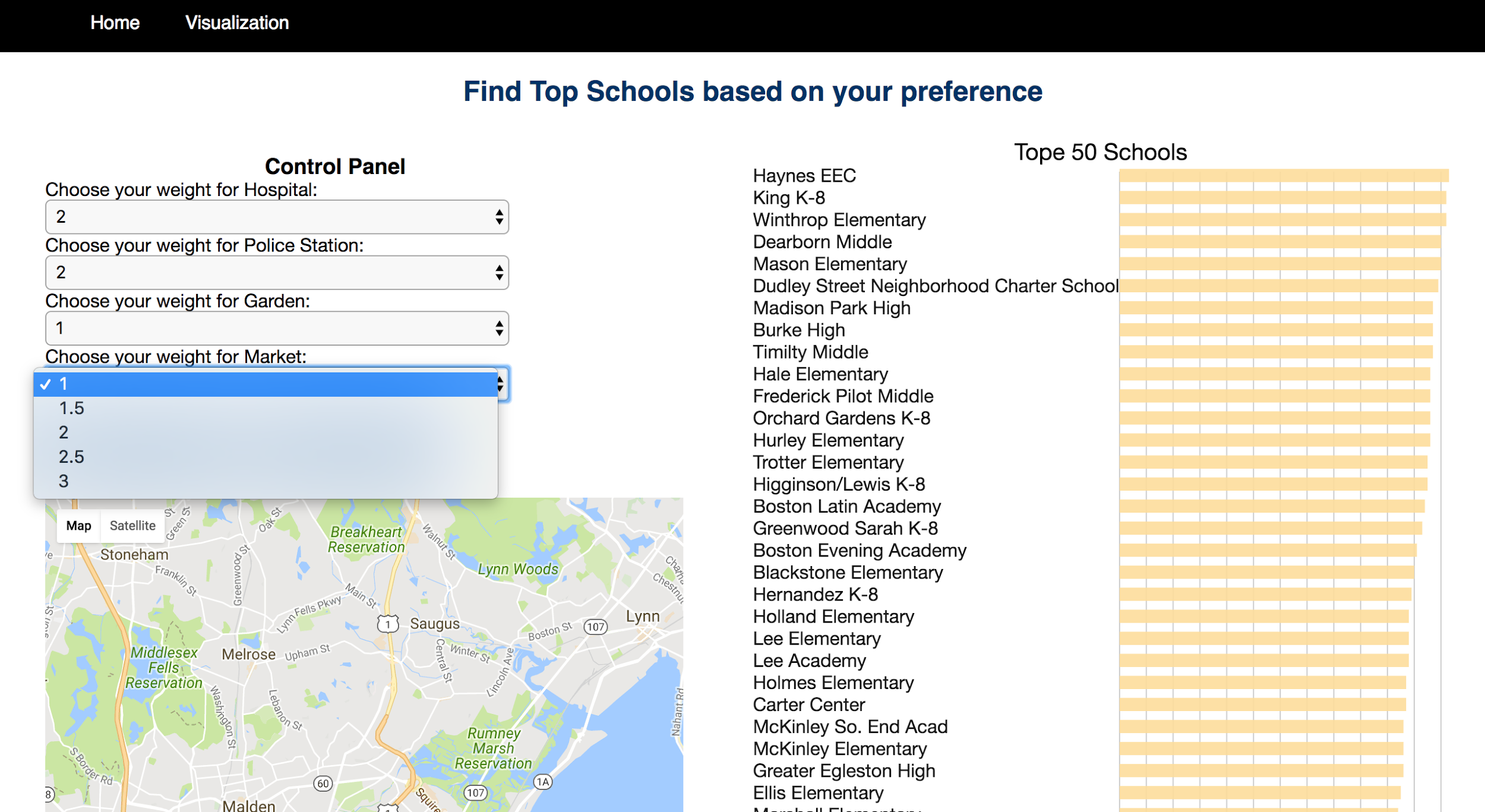
Consequently, the result infers that public schools are not clustered based on accessibility. Thus it is necessary to look into the accessibility score of each score and optimize their accessibility needs.

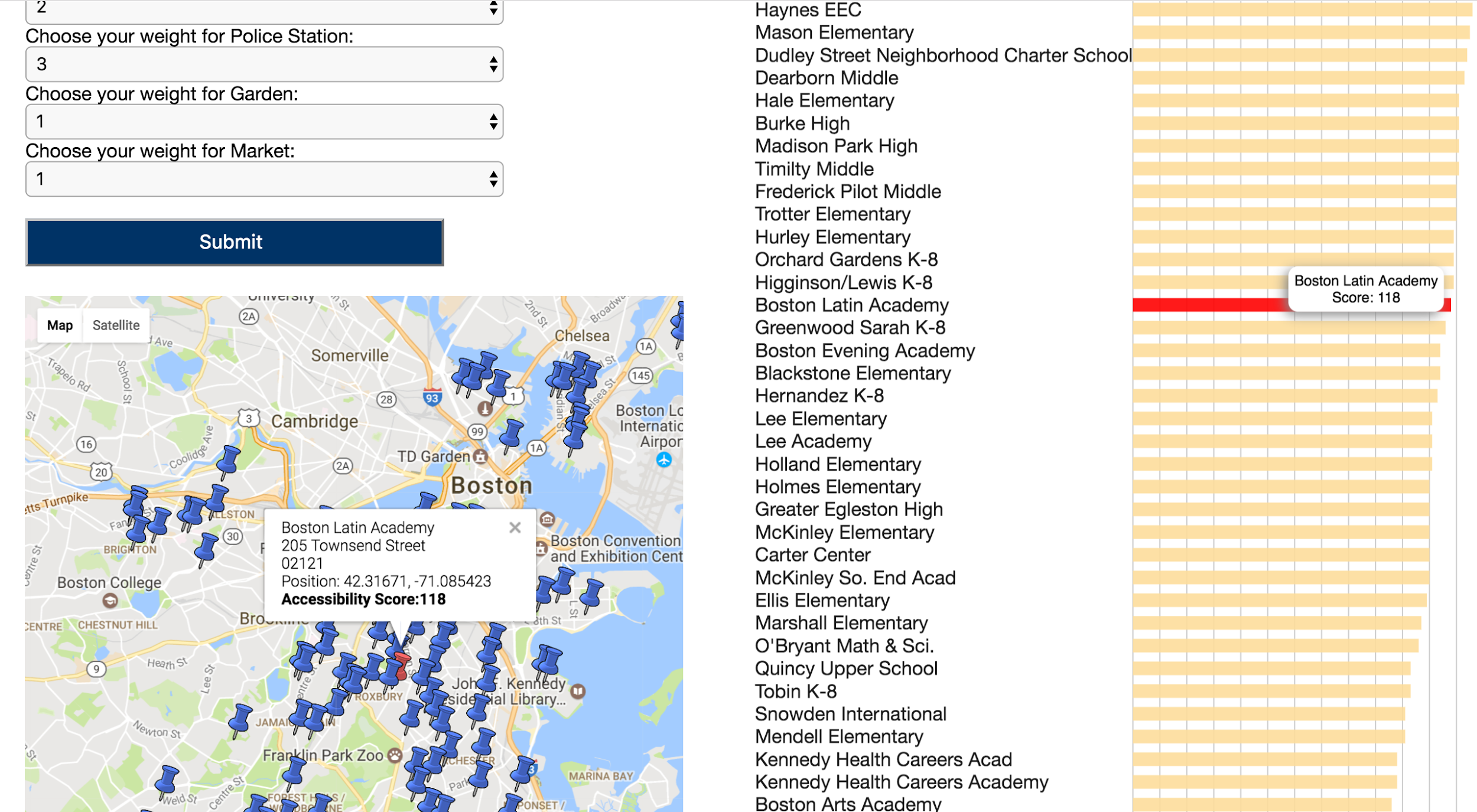


Conclusion: Correlation Diagram indicates that there are zero relationship between Accessibility Scores and Distance to Region Central

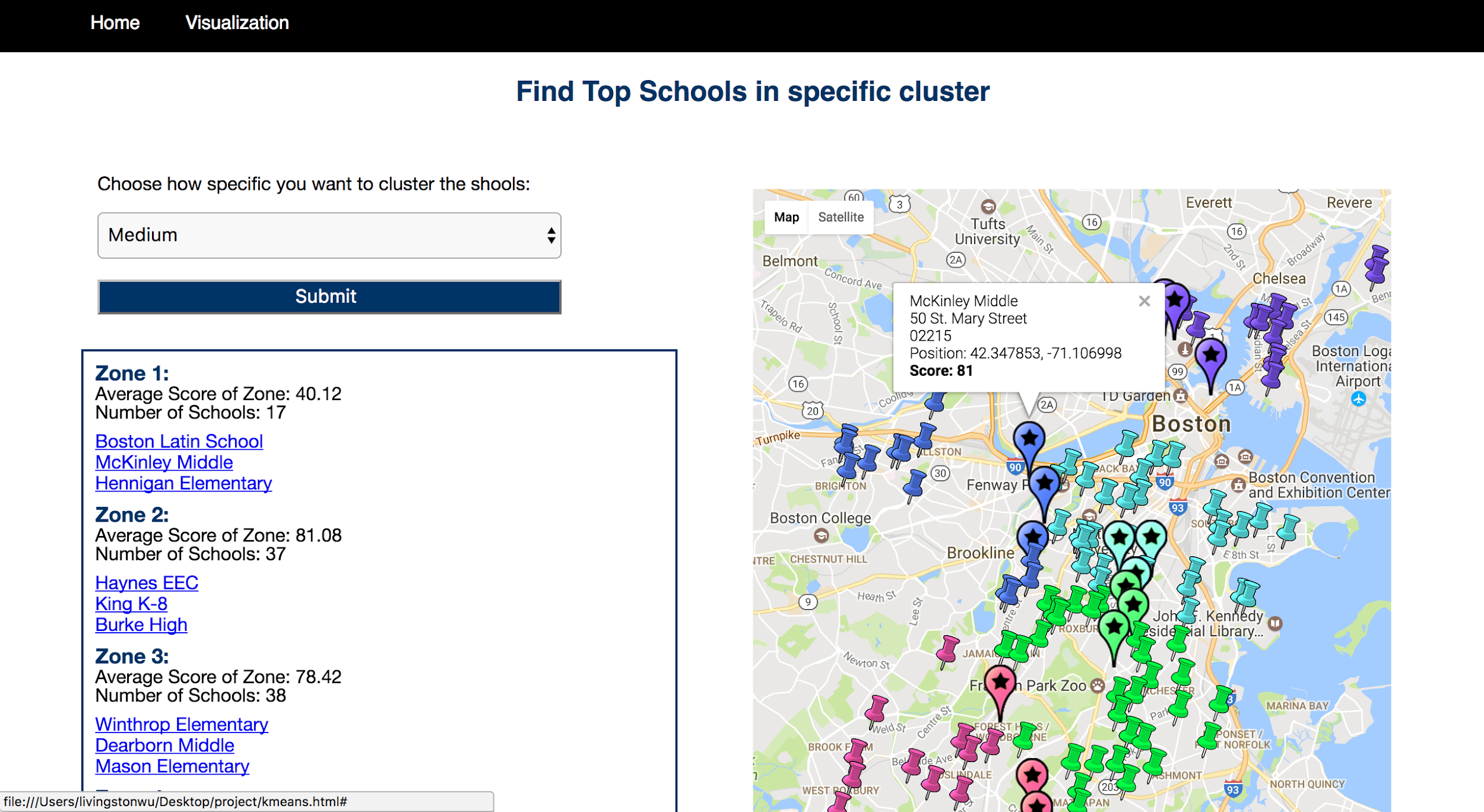
**Web Service**

We build two interactive features, the first one is to customize the scoring algorithm system to find top schools based on one’s own preference. We have four components for the final score for each school, and they are hospital, police station, garden, and market. For each factor, you can scroll down and choose the weight of 1, 1.5, 2, 2.5 or 3 based on your preference. Then click the submit button to see the ranking of all public schools based on your customized formula. If you hover on any one of the school over the ranking, you will see the score information and location of the school on the map.





The second visualization we did is to find top schools based on different cluster levels. You can choose general, medium or specific clustering based on your preference. Different cluster levels will divide all public schools into different number of area denoted by different colors. Similar to our first interactive feature, you can hover over the map or the column on the left to find the detailed information of any specific school.





**Future Work:**

Regarding our future work, we have two following ideas of improvement.

First of all, our Data Sets will not be limited by accessibilities, we will include datas like crime reports and car accident reports to make our analysis comprehensive.

Secondly, we will focus on individual school rather than large regions. Therefore, we can provide more detail suggestions for individual school.

**The Data Sets Involved**

Police:<https://data.cityofboston.gov/Public-Safety/Boston-Police-District-Stations/23yb-cufe>

API: <https://data.cityofboston.gov/resource/pyxn-r3i2.json>

Public\_school:<https://data.cityofboston.gov/dataset/Boston-Public-Schools-School-Year-2012-2013-/e29s-ympv>

API:<https://data.cityofboston.gov/resource/492y-i77g.json>

Hospital:<https://data.cityofboston.gov/Public-Health/Hospital-Locations/46f7-2snz> API:<https://data.cityofboston.gov/resource/u6fv-m8v4.json>

Market:<https://data.mass.gov/Health/Farmers-Markets/3v56-znm2> API:<https://data.mass.gov/resource/66t5-f563.json>

Garden:[Http://data.cityofboston.gov/Health/Xommunity-Garden/cr3i-jj7v](http://data.cityofboston.gov/Health/Xommunity-Garden/cr3i-jj7v)

API:<https://data.cityofboston.gov/resource/rdqf-ter7.json>