CS 591 L1 Data Mechanics

Improving Public Schools in Boston

by Wenjing Lyu Zetian Wu Guangxing Ren

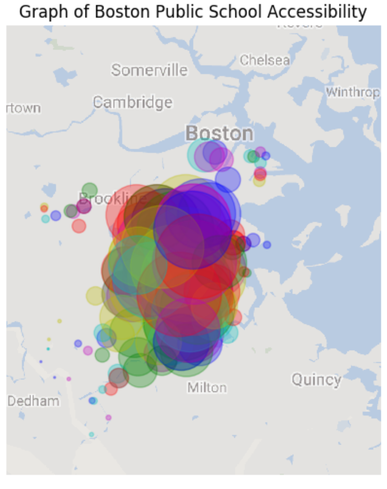
Motivation

- Education is always the priority in Boston.
- Public Schools students and parents always complain about the neighbourhoods around schools areas.
- Goals: We wish to provide improvement suggestions for public schools in different regions by evaluating the accessibility of each public schools in Boston

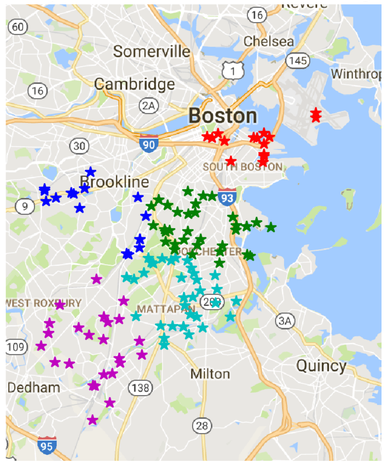
The Data Sets Involved

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

Graph of Boston Public School Accessibility



Graph of Public School K-Means Clustering



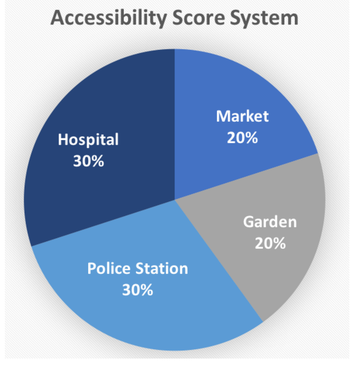
Transformation 1

K-Means and Scoring Mechanism

- We use 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
- 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.
- Results: School Sample Size:133 Region Size:5 School Average score= 56.03 School Std = 38.80

	School Number	*Average*	*T Score*	*P-Value*
Region 1	***17***	***40.2***	***0.82***	***0.44***
Region 2	***37***	***81.36***	***-1.44***	***0.22***
Region 3	***38***	***78.65***	***-1.26***	***0.27***
Region 4	***15***	***29.06***	***1.55***	***0.19***
Region 5	***26***	***18***	***2.18***	***0.094***

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



Future Work

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

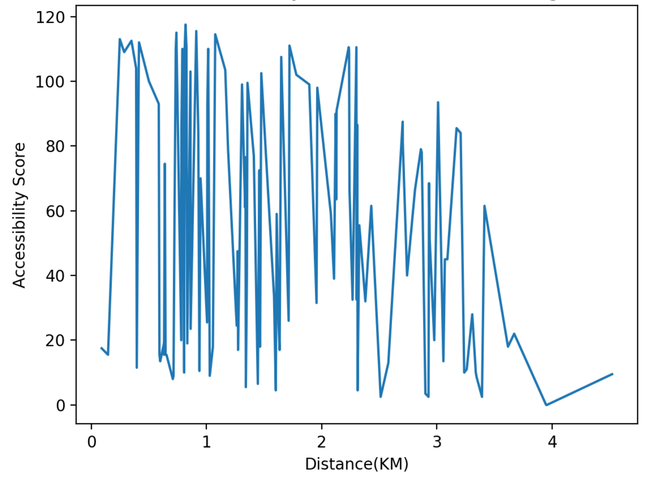
1. Our Data Sets will not be limited by accessibilities, we will include datas like crime reports and car accident reports to make our analysis comprehensive.
2. We will focus on individual school rather than large regions.

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.

Correlation of Accessibility Score and Distance to Region Central



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