Francis Zamora
Javiar Titus
Data Mechanics

Algorithms and Analyses to Advance Boston's Progress Towards Carbon Neutrality

Introduction & Motivation

Carbon emissions are at an all time high, and are posing a looming threat to the well being of planet earth; however, major cities are taking steps to curb their carbon emission by pledging to be zero emission cities. The City of Boston has pledged to be a zero emission city by 2050. We are looking at Boston's data on the eco-friendly services and facilities it currently offers: electric charging stations, hubway stations, points in an open spaces, and points in a bike networks. We define points in a bike network to be a pair of latitude and longitude coordinates which are part of any given bike network, and we define points in an open space to be a pair of latitude and longitude coordinates which are part of any given open space. Specifically, this project focuses on three different objectives: calculating a green score for each neighborhood, finding correlations between eco-friendly facilities/services, and simulating the improvement of green scores through constraint satisfaction algorithms. The foundation of our algorithms and analyses we performed took into account positioning of these eco-friendly facilities/services within neighborhoods in Boston.

Data Sets

Bike Network Data from Boston Maps Open Data

Neighborhood Data from Analyze Boston

Hubway Station Data from Analyze Boston

Charging Station Data from Boston Maps Open Data

Open_Space.csv from BostonMaps: Open Data;

Calculating a Green Score

Several comparisons and aggregations were performed in order to acquire the grouping of eco-friendly services/facilities within neighborhoods in Boston; we leveraged the use of the techniques from the first section of the course in order to transform the data. With this data we took the averages of the placement of each of the individual eco-friendly facilities and services across neighborhoods, such as the average of the electric charging stations across neighborhoods , average of the hubway stations across neighborhoods, etc. We acquired four separate averages and combined them in order to create one average that represents the average of the averages we acquired. We call this new average the limit. Based on the tuple of the form (Neighborhood Name, Number of Charing Stations, Number of Hubway Stations, Number of

Points from Bike Network, Number of Points from Open Space) we then iterated through every element, excluding the neighborhood name. The score calculation has two variations and is represented through this model:

Objective: Calculate a green score for each of Boston's 26 neighborhoods

$$\text{Score} = \begin{cases} \dfrac{\sum_{i=1}^{n} \frac{F_i}{4} + (\text{Item}_i - \frac{F_i}{4}) \times (.2)}{10} \\[2ex] \dfrac{\sum_{i=1}^{n} (\text{Item}_i)}{10} \end{cases}$$

*we define items as charging stations, hubway stations, bike network points, and open space points

where F_i is the average number of each item per neighborhood

Note: Fi/4 is the acquired of average of averages, which we refer to in this paper as the limit.

Correlation Between Green Developments

Using this data we performed a statistical analysis using the Pearson product-moment correlation: ρ(x, y)=cov(x, y) / (σ(x) · σ(y)) in an attempt to find out if there exists any correlation between subset entities of our data and if these correlations corresponded to the number of placements of select entities in each neighborhood. To do this, we iterated through all the possible subsets of two entities within neighborhoods and calculated correlations with the help of SciPy library. Our findings are presented in an easy to read table:

X	Y	Correlation Coefficient	P Value
Charging Stations	Hubway Stations	0.4066	0.0392
Charging Stations	Bike Networks	-0.2419	.2336
Charging Stations	Open Spaces	-0.2611	0.1974
Hubway Stations	Bike Networks	-0.1785	0.3827
Hubway Stations	Open Spaces	0.0503	0.8071
Bike Networks	Open Spaces	0.6459	0.0003

Constraint Satisfaction Algorithms

We took the green scores we computed initially and set up a constraint satisfaction problem where we attempted to optimize the green score for each neighborhood given a range of budgets from $200,000 to $2,400,000. The constraints we added attempted to make the computation realistic, meaning that solutions of 0 or less than 0 were not acceptable, in an effort to deplete the budget. We used the z3 solver to help us solve the constraint satisfaction problems for all the ranges of the budget. We also randomized the maximum and minimum number of specific entities that could be built in neighborhoods in an effort to create a unique solution for each neighborhood, and this data was placed in a MongoDB collection which our visualization makes use of during its execution.

Visualization and Web Service

We took our data that we acquired from our constraint satisfaction algorithms and made a web application that features a map of Boston, an interactive slider that allows you to pick the budget across each neighborhood, and allows the user to hover over each neighborhood and see the corresponding green score with the given budget that they select, as well as the corresponding addition of eco-friendly facilities that would be added in accordance with the budget. We made use of a web application written in python in conjunction with the flask library in order to retrieve the data from the collection. The interactive visualization lets the user pick through a range of budgets and watch the number of green facilities/services change as well as the score with each budget. The html document was able to retrieve these results using the python web application, and the visualization was made using LeafletJS, and D3.js.

Conclusion

In our findings, we implemented a robust algorithm to calculate a green score for each neighborhood, we discovered certain correlations between the placement of eco-friendly facilities and services across neighborhoods, and readjusted green scores of each neighborhood through a series of constraint satisfaction algorithms that make an unbiased attempt at satisfying constraints for each neighborhood with a given budget. Furthermore, we created a visual tool that allows the user to see the changes in the green score that would occur given a new budget within our range. The algorithms and analyzes we developed are taking the city of Boston and its neighborhoods in the right direction towards its Carbon Neutrality goal.

Future Work

These basic implementations require more tweaking to produce results of a higher degree of significance; however, with the consideration of more dimensions into each neighborhoods green score we would be able to tweak our green score algorithm, as well as correcting certain biases. Moreover, we would increase the range of the budget for the constraint satisfaction algorithms, and explore more correlations between the extra dimensions of eco-friendly facilities and services which we will add in the future. One problem we faced was that the number of green facilities would go up with the budget, but not necessarily the score which means that the score algorithm needs to be tweaked accordingly. We hope to expand our green score algorithm, correlation analysis, and budget visualizations to include neighborhoods in Cambridge, and the Greater Boston area. Furthermore, we want to implement a real time platform for the City of Boston which includes the use of the green score, constraint satisfaction algorithms, and budget visualization, as well as machine learning algorithms.