BOSTON UNIVERSITY

Computer Science Department

City of Boston services: Police Patrol allocation & Inspection analysis

Stephanie Alibrandi, Javier Arguello

Abstract

The City of Boston is performing a significant effort on collecting data. The diversity of these publicly available datasets allows us to combine them in order to improve city services. Particularly, this project focuses on three different studies:

- Improving Police Patrol allocation.
- Find correlation between inspections and reviews.
- Ranking zip codes.

Datasets

City of Boston:

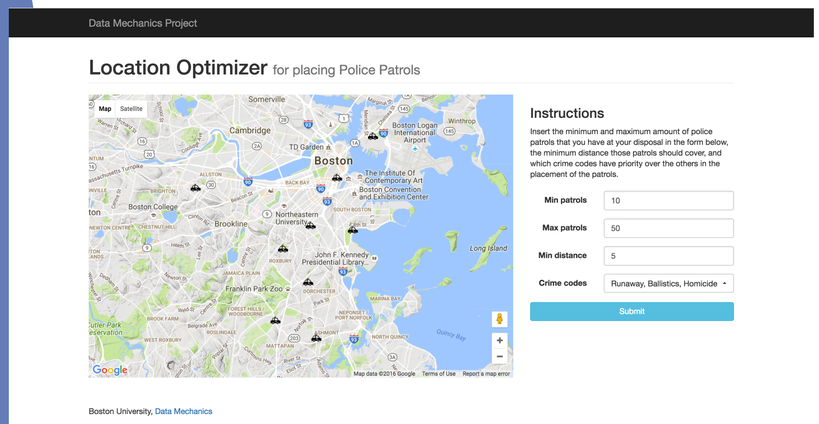
- Crime Incident Reports
- School Grounds
- Hospital Locations
- Food Establishment Inspections
- 311, Open Service Requests

Yelp Academic Dataset:

- Yelp business reviews
- Yelp business

Improving Police Patrol allocation

Goal: Find coordinates where Police Patrols should be allocated to minimize the distance to crime areas.

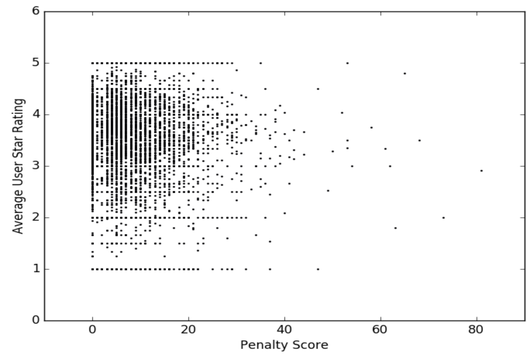


- Used **k-means** to find optimal patrol locations.
- Customizable input.
- **RESTful API** using python package FlaskAPI.
- U**ser-friendly interface** to visualize results on Google Maps using D3, Bootstrap, HTML and CSS.

Correlation between inspections and reviews

Goal: determine if food inspections and user reviews are correlated.

	Correlation Coefficient	P-value
Minor	-0.036	0.006
Major	-0.038	0.003
Severe	-0.03	0.02
Penalty score	-0.042	0.001
# violations	-0.041	0.001

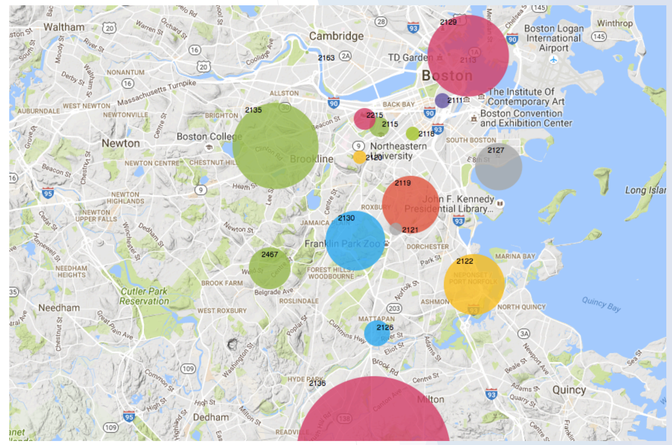


- $penaltyScore_r = minor_r + major_r + severe_r$
- Lack of unique identifier between the datasets made the **merging process complicated**.
- Negative correlation with **high significance**.
- **Reviews assigned by date** to a particular inspection.
- Using **Pearson Correlation Coefficient**.

Ranking zip codes

Goal: rank zip codes considering several attributes

$(\#crimes, \#311\ reports, \#passed\ inspections, \#schools, \#hospitals)$



- Populated missing zip codes.
- Used **multi-objective query** (skyline query).
- **Equal importance** to all attributes.
- Result set is *Pareto optimal*.

Future Work

- Take **residential** data and perform a **ranking** to determine the best neighborhood to live in.
- Build a model to **predict inspection failure** based on user reviews.

Conclusions

- Built a tool for dispatching police patrols to maximize coverage area of specific historic crimes.
- Correlation between inspection result and user reviews exists.
- Modeled and ranked the zip codes with a multi-objective query.