Find the best living area in Boston

Minteng Xie Yue Lei Zhi Dou

BOSTON UNIVERSITY

Introduction

Problems:

- Finding the most livable area in Boston with specific preference and constraints are complex
- The potential development of a certain area is critical for people to make the decision

As a result, we implement website for users including:

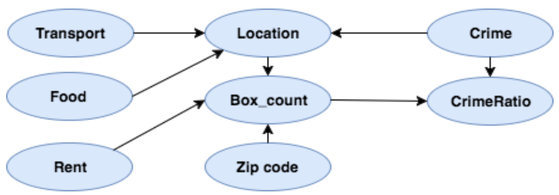
- Optimization model quantizes the degree of livability of different areas and gets the most suitable one for certain requirement
- Statistical analysis gets internal relationship of different factors and crime variances in a certain area by correlation analysis.

Dataset

Four core datasets:

- Average **Rent** for Boston Cities
- **Transport** data (MBTA Stops)
- Active **Food** Establishment Licenses
- **Crime** Incidents In Boston(2012-2017)

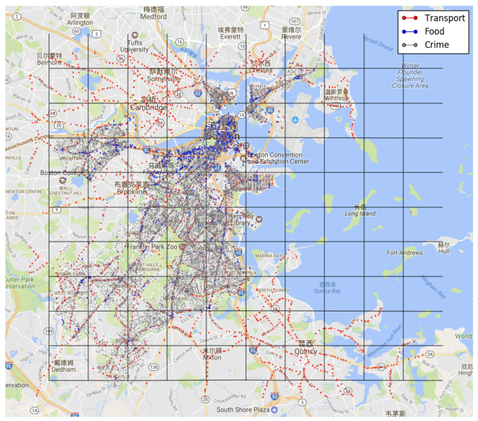
Transformation



Algorithm

1. Optimization

We build a big rectangle using the left bottom/ right top coordinates of Boston. Then we separate this rectangle into 10 x 10 grids so that we can solve the optimization problem by: maximize **#restaurant**, **#MBTA stops** and minimize **#crime incidents** and **rent price**.



Algorithm

We map the four numbers/counts into **Ratings**, then users could select the **Input** ratings for 4 attributes,

If we find **Ratings** all are above the **Input**, we implement:

$$argmax_{R \epsilon Ratings} \sum_{i=1}^{4} (R_i - Input_i)$$

The result is the location with maximal sum of four attributes.

Else, we implement:

$$argmin_{R \epsilon Ratings} \sum_{i=1}^{4} (R_i - Input_i)^2$$

The result is the location with minimal distance from the **Input** ratings.

2. Statistical Analysis

Correlation coefficient model of studying the linear relationship between different random variable:

$$\rho(x,y) = cov(x,y)/(\sigma(x) \cdot \sigma(y))$$

To verify the correlation coefficient, hypothesis testing is applied and set

$$H_0: The\ correlation\ is\ an\ artifact$$

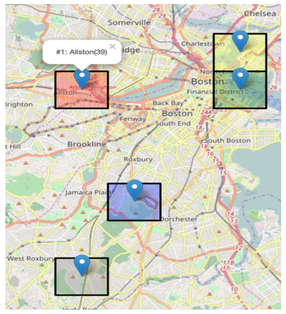
Results

1. Optimization

Input of requirements are [5, 4, 2, 2]

Top fitted areas:

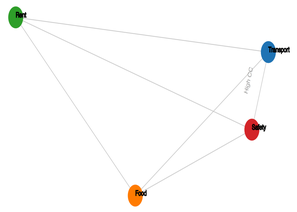
Area	Zipcode	Food	Transport	Safety	Rent
Allston	02134	★★★★★	★★★★	★★	★
Roxbury	02121	★★★★★	★★★★	★	★★★
Downtown	02110	★★★★★	★★★	★★	★
East Boston	02128	★★★★	★★★★	★	★★
Roslindale	02131	★★★★	★★★	★★	★★★



2. Statistical Analysis

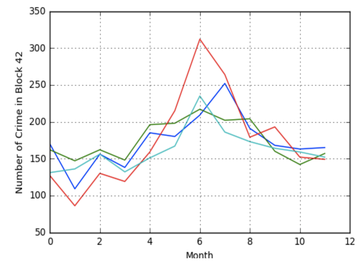
To study the relationship between these four attributes (f, t, s, r), correlation coefficient and p-value were calculated

Value Pair	Correlation Coefficient	p-value
Transport - Food	0.18443133126957592	0.24835610469426553
Transport - Rent	-0.063998564292522625	0.6909797629197301 8
Transport - Safety	-0.76594432658403122	0.00000000537058223953
Food - Rent	-0.22292437540342935	0.16122291338006584
Food - Safety	-0.35851557322730288	0.021350190751979622
Rent - Safety	0.00863790426302407 48	0.9572538186779973

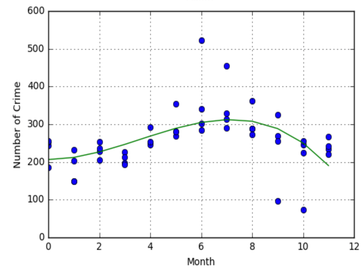


To study the consistency of crime in every block. Let X_{ij} be the number of crimes happened in block *i* in the year *j*.

Results



year	correlation coefficient	P-value
2013-2014	0.782569291953	0.0026219859084
2014-2015	0.796445834967	0.0019334569301
2015-2016	0.928294156304	0.0000132251510



block 42 – Downtown has perfect consistency, assumption, which means this consistency will be kept in this year.

Fitting method to predict of block 42

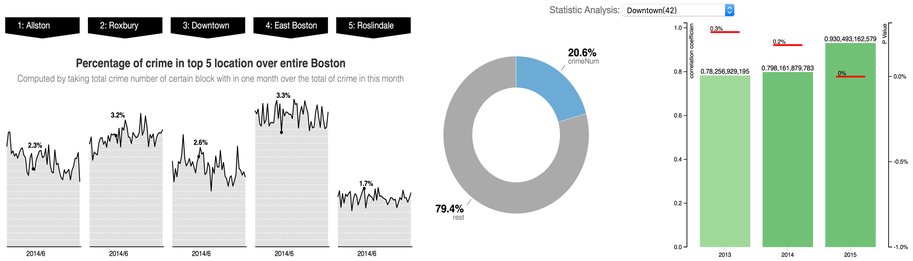
$$A = Y \backslash X$$

Web Pages

Optimization

Statistical Analysis

Project Link



Future work

- Traffic condition will be added as a supplement for parts of transportation.
- More influential factors such as entertainment, public facility will be counted.
- Add factor analysis to figure out which of these factors are the most important.
- Time series analysis will be applied to present the changes and trend of different factors.