



IBM Capstone Project

Data Analysis on Choosing District
for New Residential Project

Author : Tse Chun Ho



Table of Content

1. Background
2. Data Source
3. Data Visualization
4. K-means Clustering
5. Results and Conclusion



Background

- Equity Residential(EQR) is going to develop a new residential project in Boston
- I am one of the data Scientists in Equity Residential
- EQR got a huge success on an apartment project in Downtown
- The new location of the new residential project have two requirement:
 - The chosen district need to have low crime rate
 - The chosen district have the similar venues with Downtown



Data Source

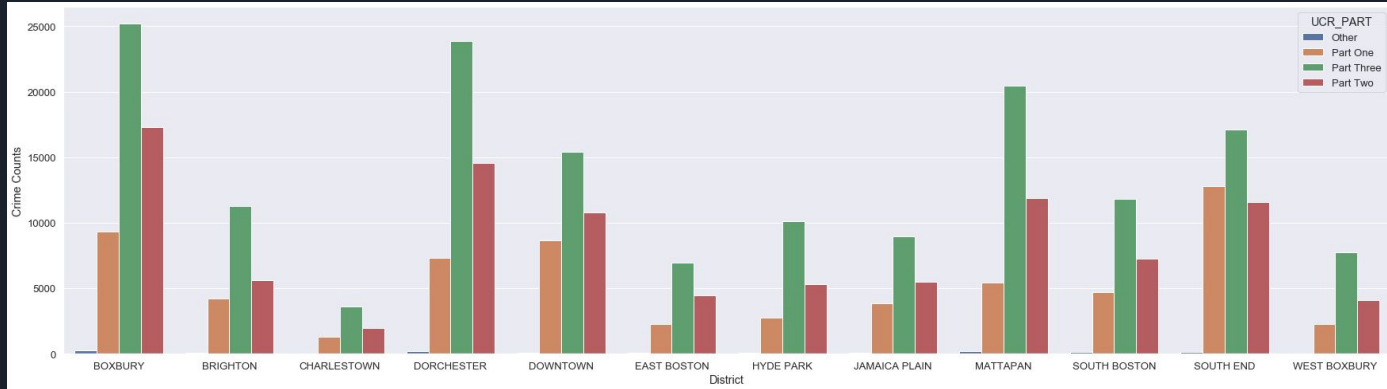
1. Crime Record from 2015 to 2018
 - Downloaded from Official Boston Police Website
2. Foursquare API
 - Purpose : getting nearby venues of a district

Data Visualization - Crime Counts



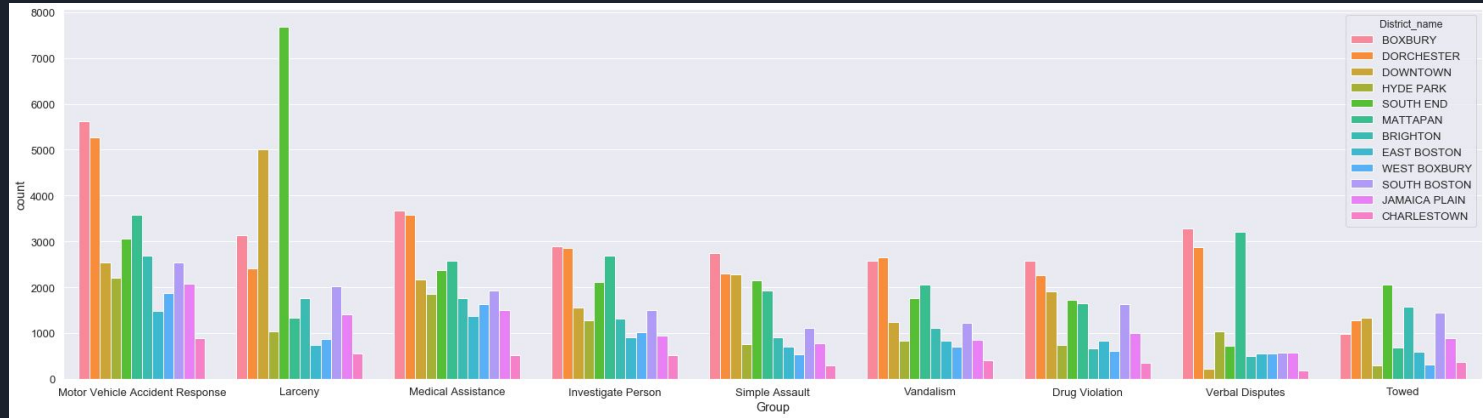
- Roxbury, Dorchester and South End have the highest crime rate in Boston
- High crime rate group : Roxbury, Dorchester, South End and Mattapan
- Middle crime rate group : Downtown, South Boston, Brighton and Jamaica Plain
- Low crime rate group : Hyde Park, West Roxbury, East Boston and Chlestown

Data Visualization - UCR Part



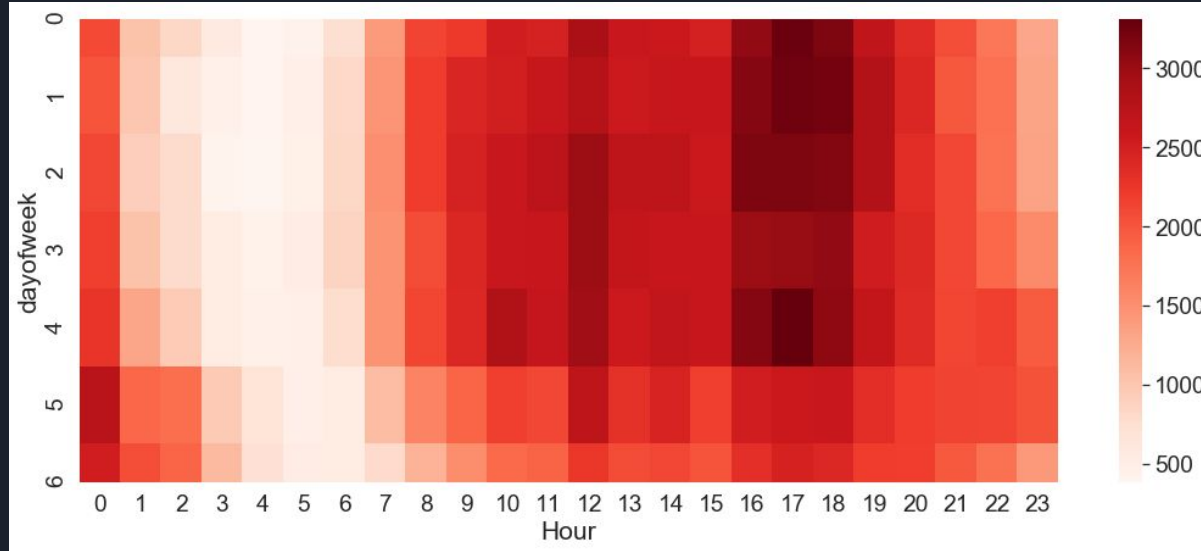
- UCR part one includes serious crime
- Highest UCR part one district : South End, Boxbury and Downtown
- Highest UCR part two district : Boxbury, Dorchester and Mattapan
- Decent district : Charlestown and East Boston

Data Visualization - Crime Category



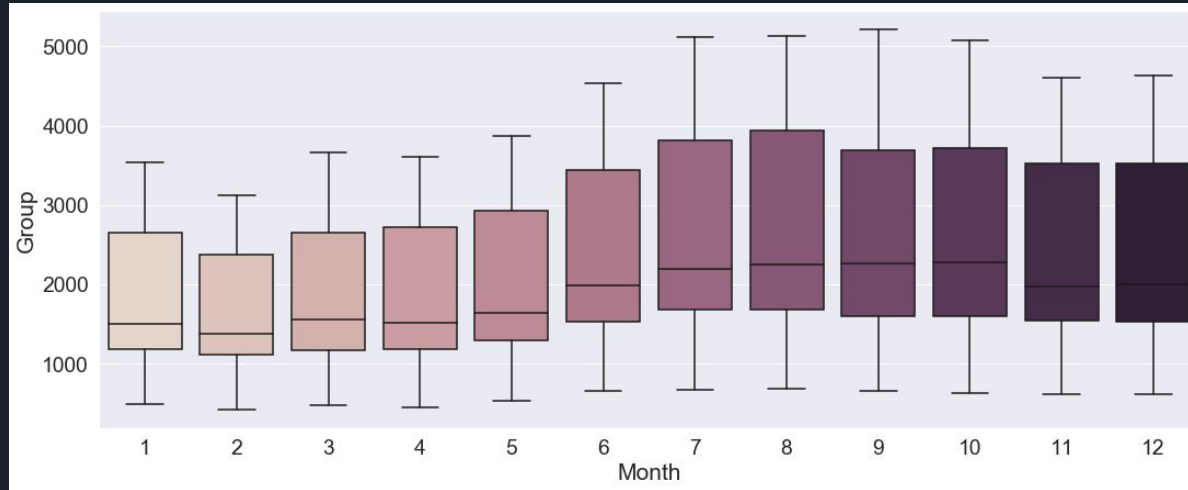
- Larceny, Simple Assault, Vandalism and drug violation will highly affect the living quality
- High Larceny District : Boxbury, South End and Downtown
- High Simple Assault District : Boxbury, Dorchester and Downtown
- High Vandalism District : Boxbury, Dorchester and Mattapan
- High Drug Violation District : Boxbury, Dorchester and Downtown

Data Visualiztion - Crime occured time



- Lots of crime happened at 16:00 to 18:00
- Lots of crime happened on weekdays

Data Visualiztion - Crime occured time



- Crime happened in July to October are more than other months
- August have a higher crime counts than others
 - Upper quartile and lower quartile are higher



K-means Clustering

Similarity with Downtown(%)	
district	
SOUTH BOSTON	90.0
SOUTH END	70.0
MATTAPAN	70.0
JAMAICA PLAIN	70.0
DORCHESTER	60.0
EAST BOSTON	50.0
BOXBURY	30.0
BRIGHTON	30.0
CHARLESTOWN	20.0
HYDE PARK	10.0
WEST BOXBURY	0.0

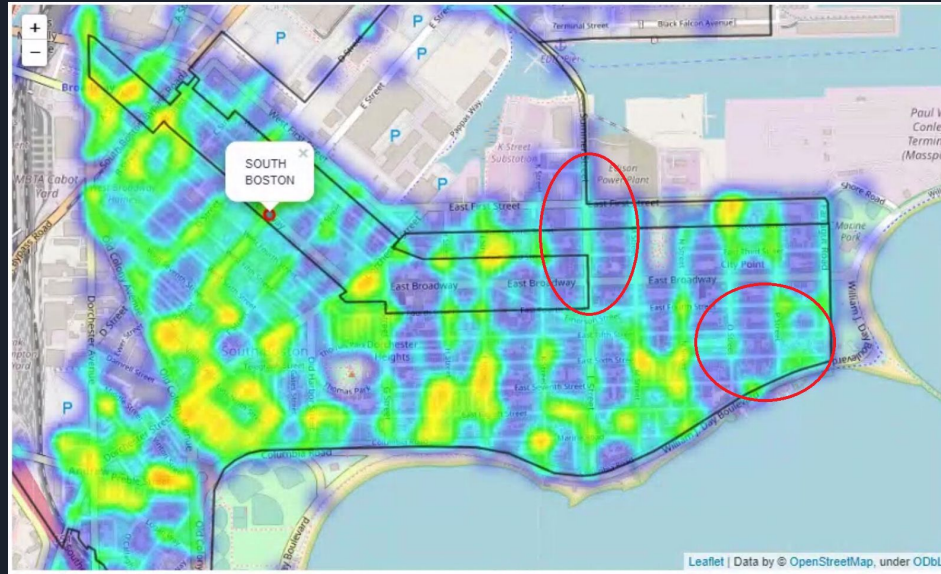
- South Boston have the most similar venues environment to Downtown
- South End, Mattapan and Jamaica Plain are the second similar venues environment to Downtown



Result and Conclusion

- South Boston is the chosen district for the new residential project
 - The venues and environment are the most similar to Downtown
 - Crime rate is in middle crime rate group
 - The crime counts in larceny, simple assault, vandalism and drug violation are in average

Result and Conclusion



- Heat map shows these two circles have a lower crime rate
- these two areas can be a good location of the new residential project