THE BATTLE OF THE NEIGHBORHOODS

LEVERAGING LOCATION DATA FOR PROBLEM SOLVING

NOT BEING FAMILIAR WITH A LOCATION WHEN RELOCATING CAN BE TROUBLESOME … SO START LOOKING IN THE RIGHT PLACE

- Focus on basic needs/requirements
  - 1 or more people
- Gather and evaluate data on requirements of the greater area
- Predicting neighborhoods as a starting point for further investigation saves time
  - Work smarter…not harder

BACKGROUND

- Individual relocating to Boston with family
-

Basic Needs	Family Wants
Security	To live in a safe environment
Close to working locations	Must have loads of restaurants in the area, Close to Museum (place of work)
Public Elementary school	Neighborhood close to school

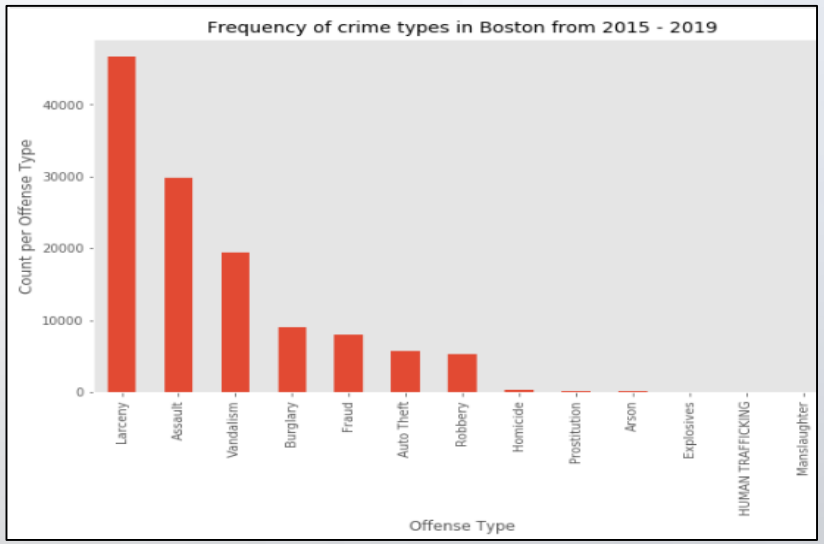
DATA ACQUISITION AND CLEANING

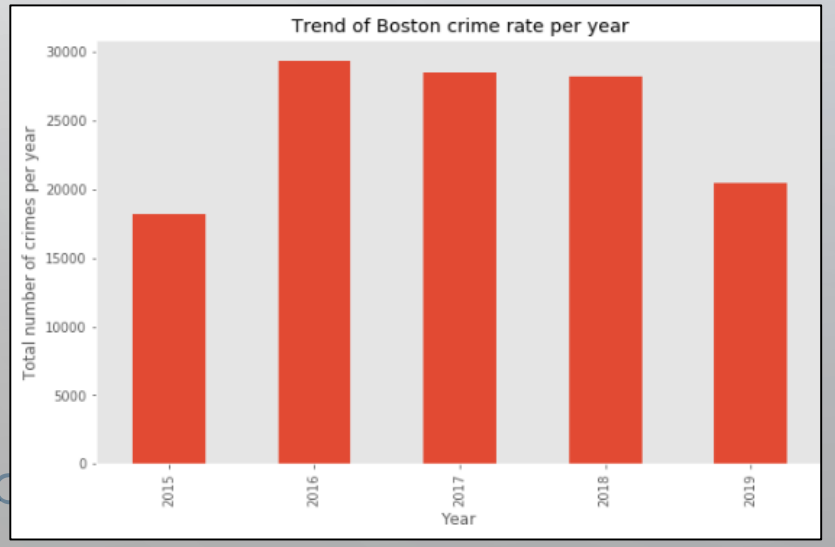
- Crime data – Analyze Boston (https://data.boston.gov)
  - Records for crime incident reports as provided by the Boston Police Department for 2015 – 2019 .
  - Scraped the location, type and timeframe
  - ~125 000 data entries in cleaned dataset

- School data – Analyze Boston (https://data.boston.gov)
  - School building investment data containing location information as well as type of school
  - Extracted preferred school type with location
  - ~48 data entries in cleaned dataset

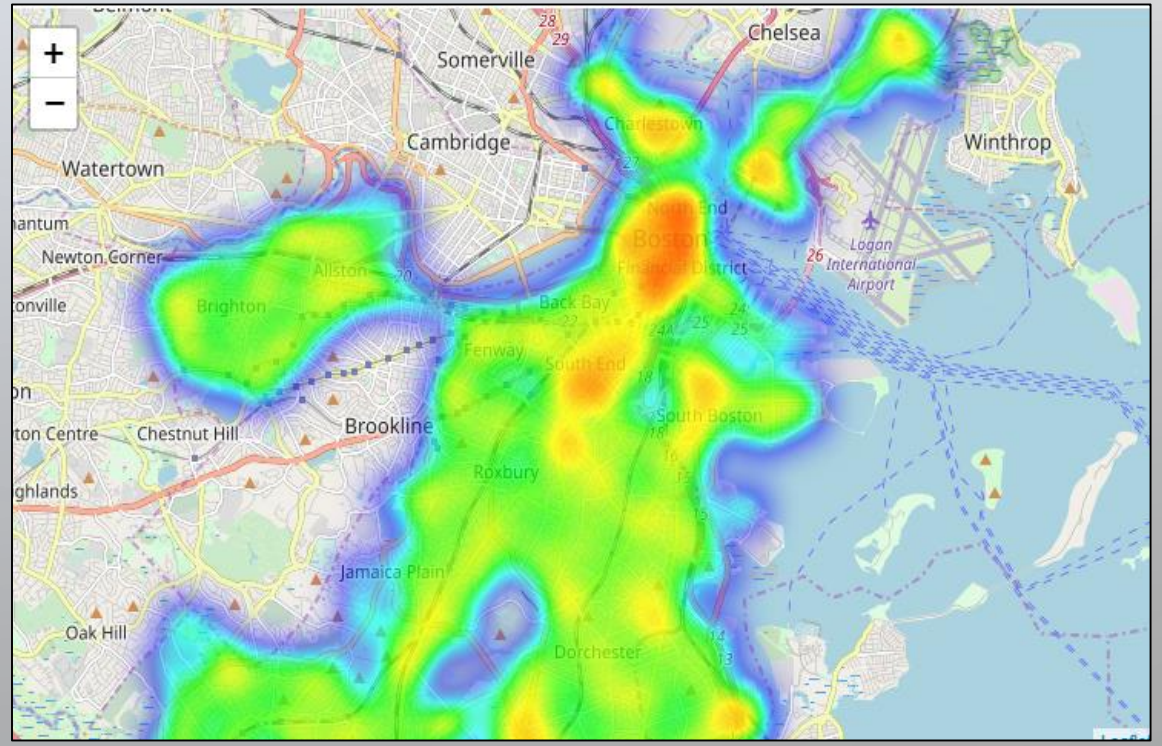
- FourSquare data – https://foursquare.com
  - Work location and - The Museum of Science coordinates
  - Restaurants – Scraped location data around the proximity of the museum including type of restaurant
  - ~27 data entries in cleaned dataset

CRIME DATA



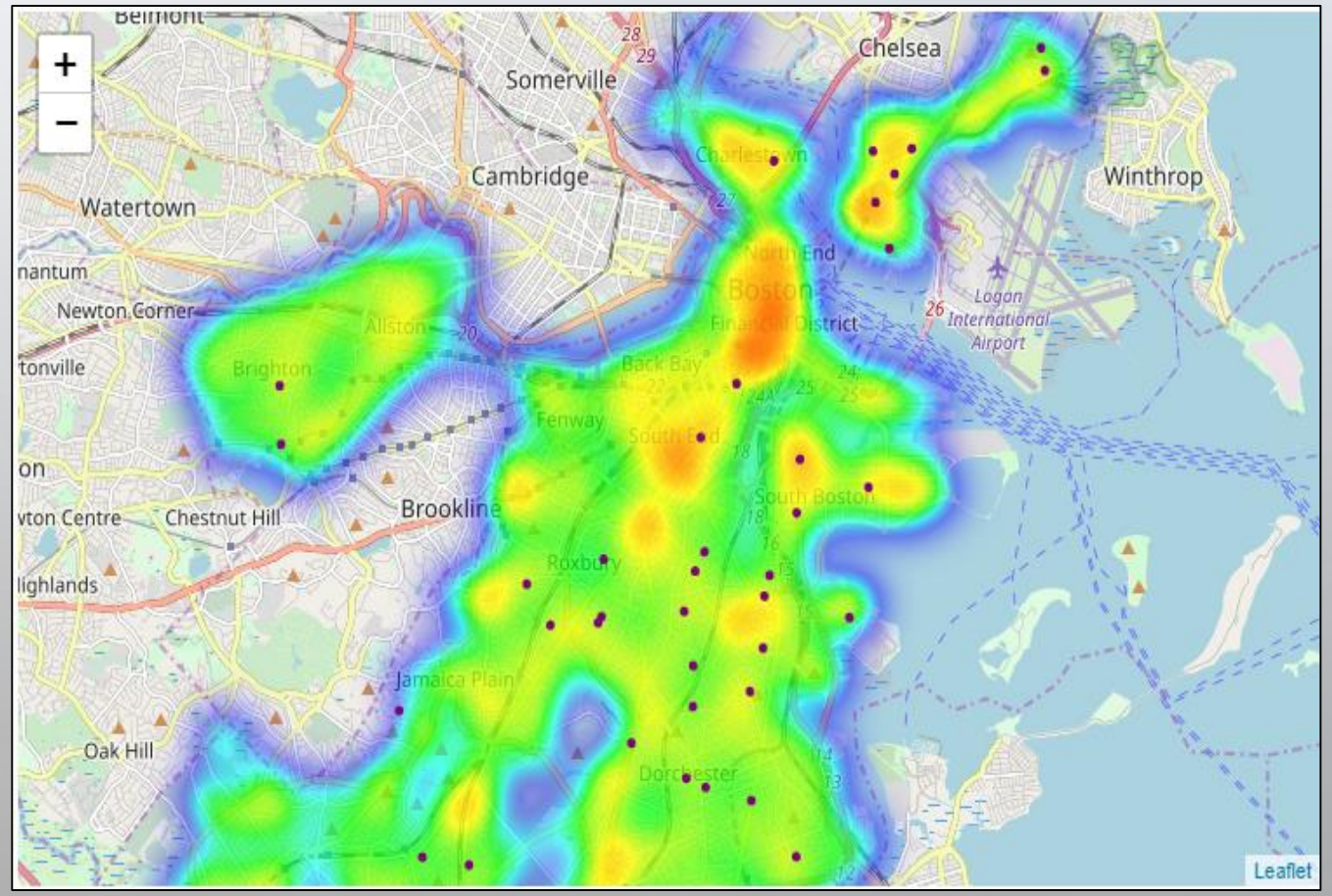


- Top 3 Crimes:: Larceny, Assault & Vandalism
- Crime trend indicates a decrease
- High crime location in center of Boston

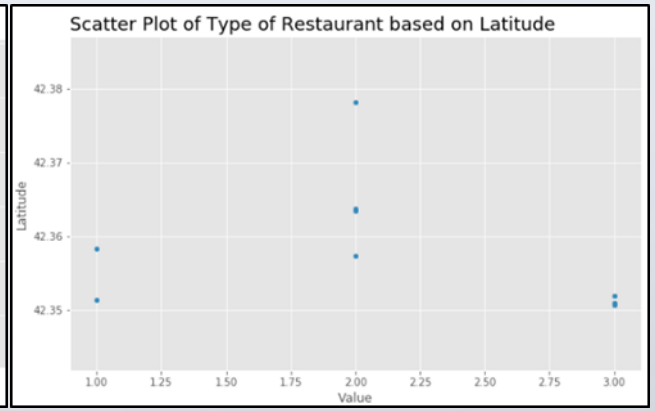


SCHOOL DATA

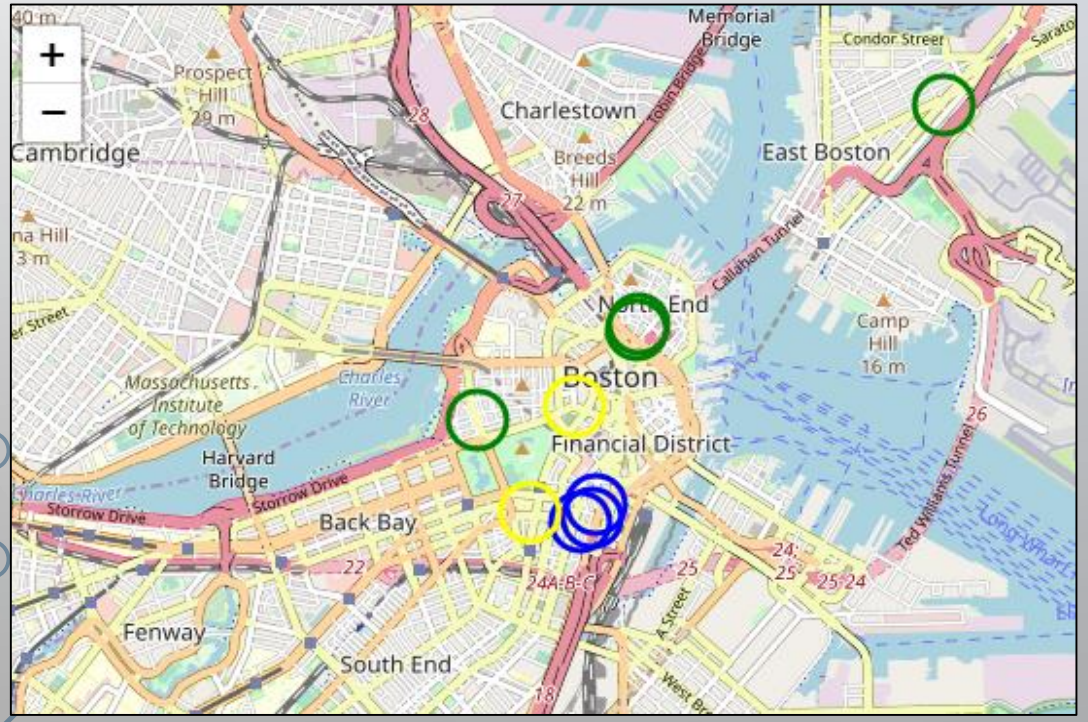
- There are various options for schools, but data on it's own the data proves inconclusive

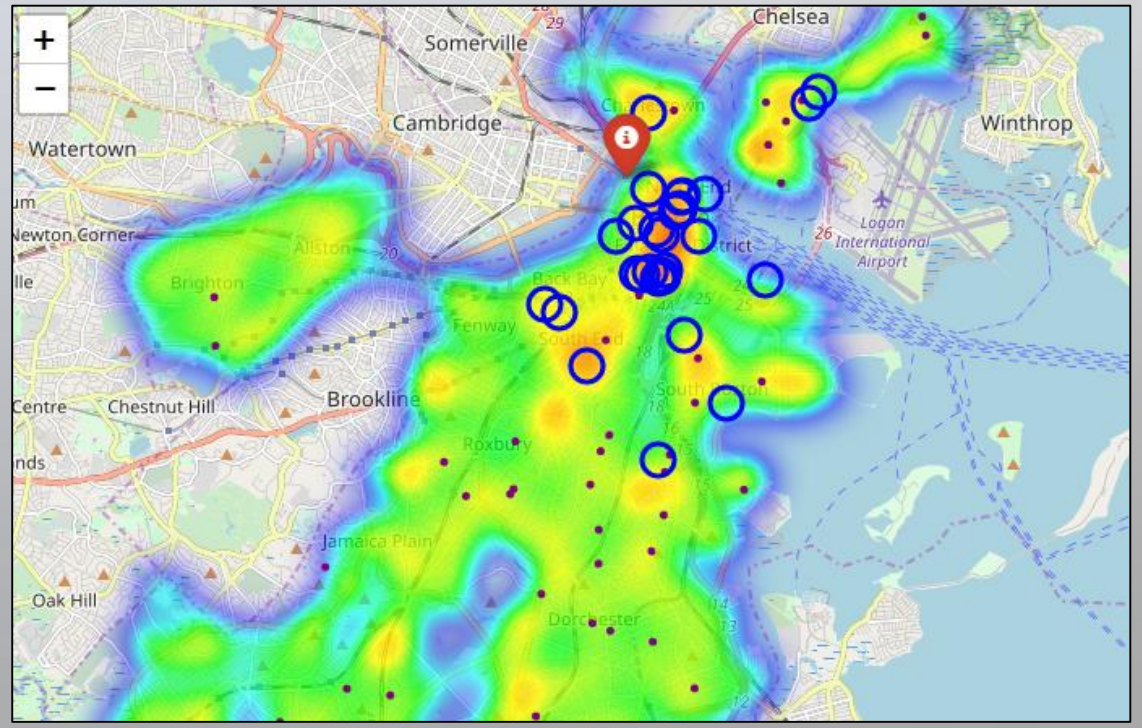


FOURSQUARE DATA



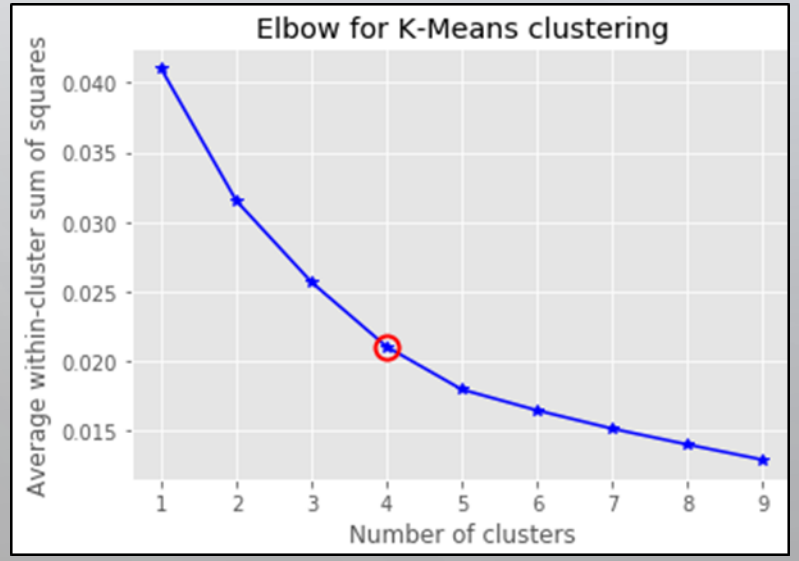
- Only Chinese restaurant types are clustered in Chinatown
- There are various neighborhood options based on all restaurant locations

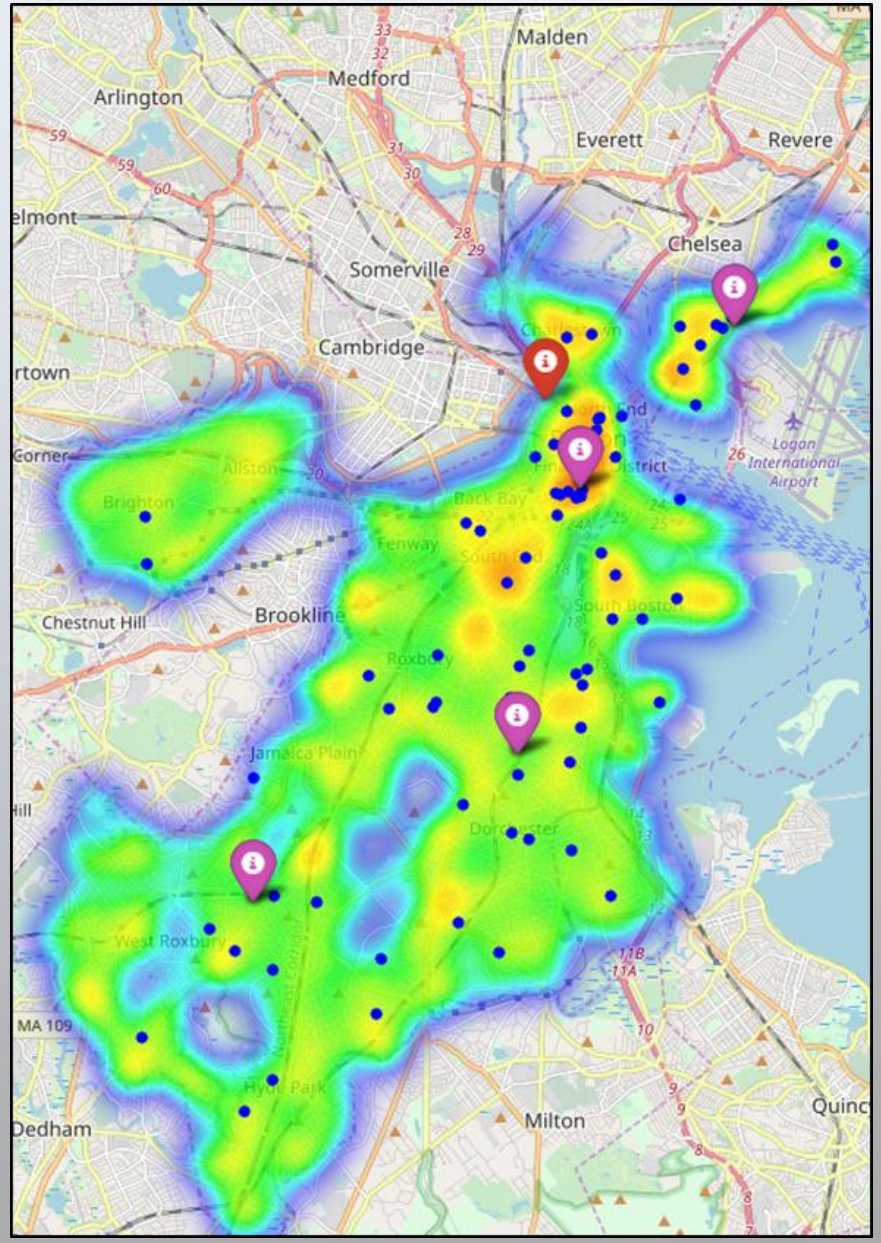




K-MEANS CLUSTERING

- Good K to use in the model is 4
- Data fitted to model to form clusters
- Cluster centroids are marked on map
- Clusters are ideal starting location for further investigation





RESULTS

- Various datasets on it's own can have inconclusive results
  - Crime data – Located all over Boston
  - School data – Scattered across neighborhoods with various locations rom Museum
  - FourSquare Restaurant data – Few datapoints in a small region

- Four ideal neighborhoods identified to further investigate, but only three conformed to basic needs. Neighborhoods to consider are:
  - East Boston, Roxbury and Roslindale

CONCLUSION

- K-means clustering prediction model indicates clearer clusters and when superimposed on a crime heat map gives better direction towards ideal neighborhoods to live in
- Recommended neighborhoods to be considered as a starting point for further extensive analysis
- Room for improvement on model accuracy
- Can include more preferences and can include more detailed results