CLUSTER BOSTON CITY WITH CRIME DATA

AUTHOR : MANASA CHAKKA

Introduction

- Boston city is center of excellence in many areas
- There are many establishments like schools, colleges, universities and technological companies in the city .
- They attract around 400,000 people to the city every year.
- Unfortunately, like with all other major cities crime rate is also high in Boston.
- This develops a need for the relocating people to find a safe home in Boston.

Data Acquisition and Cleaning

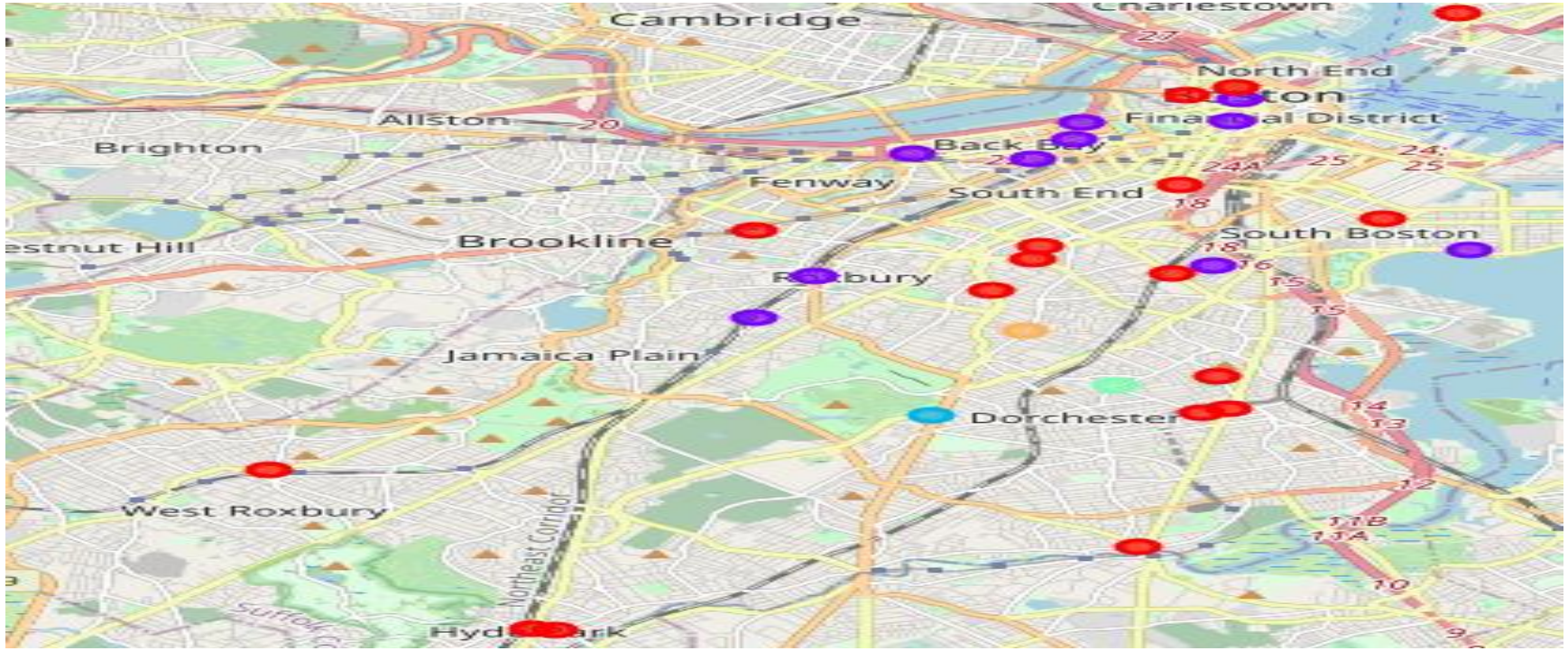
- We require crime data of the city digitalized by Boston Police Department that is readily available in Kaggle.
- We also require venue data around the area of crime occurrence using Foursquare API
- The columns having more NaN values are dropped from crime data.
- Foursquare API returns a json file with a lot of information unnecessary. We extract only name, location and category.

Methodology

- We select the features offense code group from crime data and venue category from venue data to apply our model on.
- But both are categorical variables. So we use one-hot encoding technique to convert them to numerical variables.
- The dataset is very large and we cannot afford to make so many calls to the API.
- So we select top 30 crime areas, extract rows from crime data belonging only to those areas.
- The algorithm we use to cluster is k-means cluster and fit data with it.

Results

- The data we fit is divided into five clusters. The clusters are formed as shown below.



Discussion

- By observing the clusters formed it is clear that motor vehicle accident is the most common offense recorded in all the columns.
- The cluster labelled 0 has seventeen streets recorded under it with larceny and drug violation as next common crimes.
- The cluster labelled 1 has ten streets under it with vandalism and assault as most common crimes.
- The areas under remaining 3 clusters are outliers of whole remaining data.

Conclusion

- In this study I have used crime and venue data of Boston to cluster the city based on crime rates.
- The features used are offense code group and venue category.
- The crime irrespective of cluster is high at central and north corner of Boston .
- There are not much educational institutions in this area. So the newcomer would do good to avoid this areas.