Predicting the crime rate based on the venues categories prevailing in London boroughs

MAY 2020

Introduction

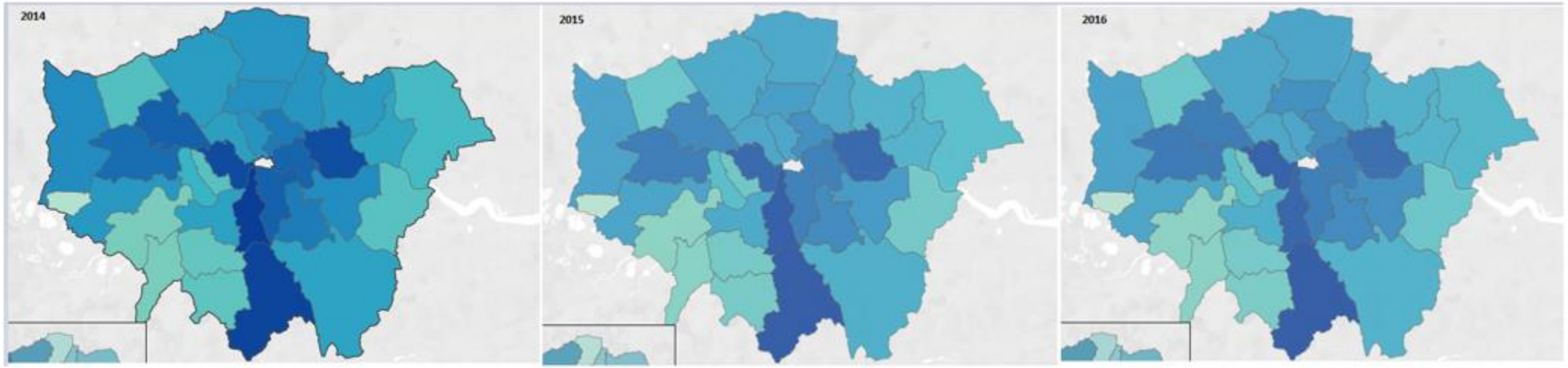


Figure 1. London boroughs crime rate for 2014 /2015/ 2016 years

- Some areas of London remain safer than others persistently in years
- The London areas (Boroughs->Wards) have different *character*
- Character can be defined by prevalence of venues' categories
- Can we explore dependency between **character** and **crime rate**?

Interest

- Local authorities can impact on the crime rate by restricting and increasing prevalence of some areas
- Property Investors can explore areas with a good potential growth of real estate valuation

Data acquisition

- The following information was loaded from official London Government Data Store Portal for 625 London wards for 2019 year:
  - Population denstity
  - Recorded crime feed (for crime type ***Violence against the person)***
- Geo Location data was scraped from MapIt (mySociety)portal for wards
- 15000+ venues were exported by using Foursquare API and allocated to corresponding wards

Clustering by crime rate

- All wards were processed with K-Means and split into 3 categories

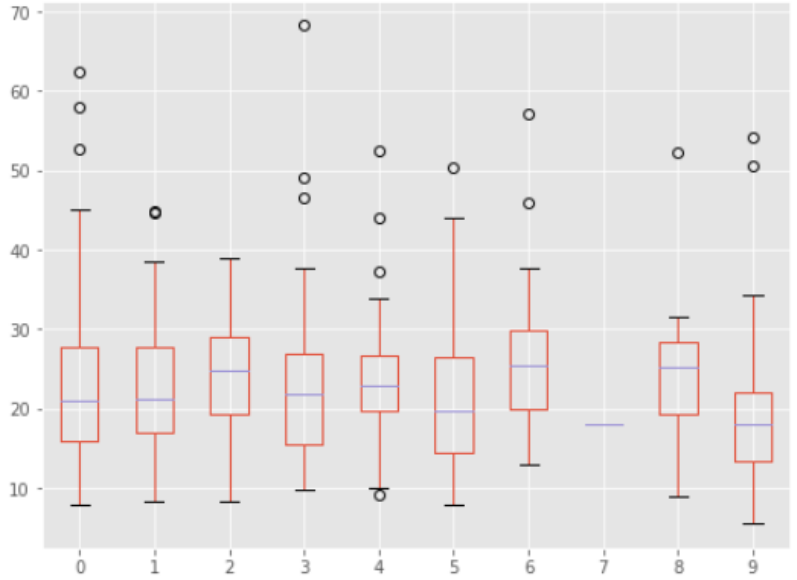
Category	Min rate	Max rate	Number of wards
Low	5.61	21.26	272
Medium	21.36	35.35	242
High	35.77	68.20	145

Table 2. Final clustering by crime rate

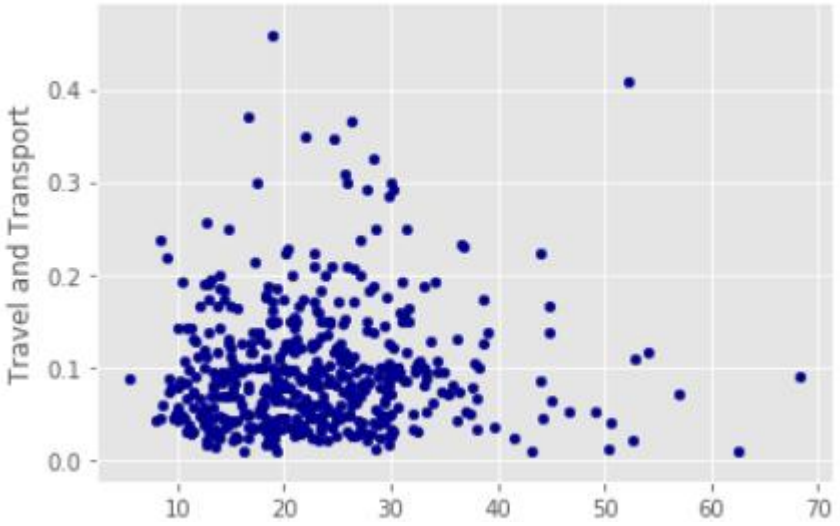
- As results of the clustering 2 wards were excluded for being too touristic and therefore not usable for normal cases

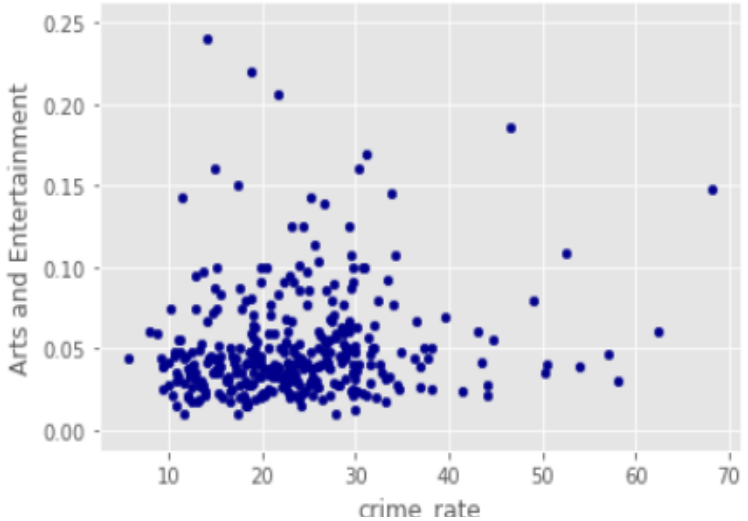
Clustering by prevailing categories

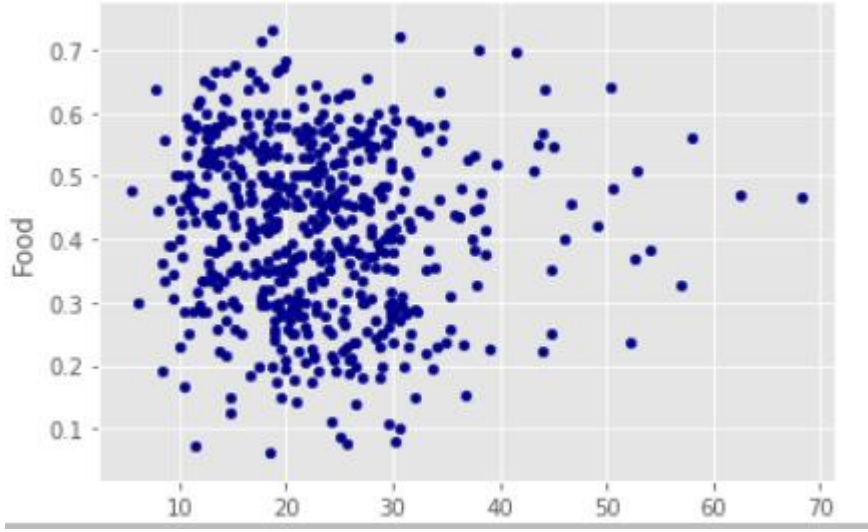
- K-Means with 10 clusters was used
- Cluster “7” was excluded as being as another exception (Theme Park ward)
- The crime rate varies quite randomly with each cluster (see figure)

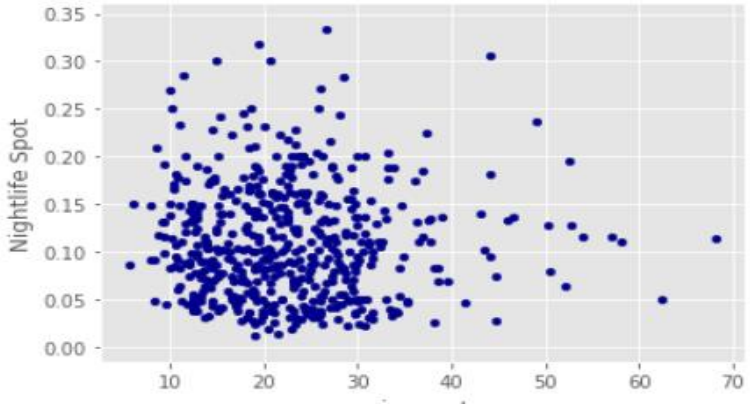


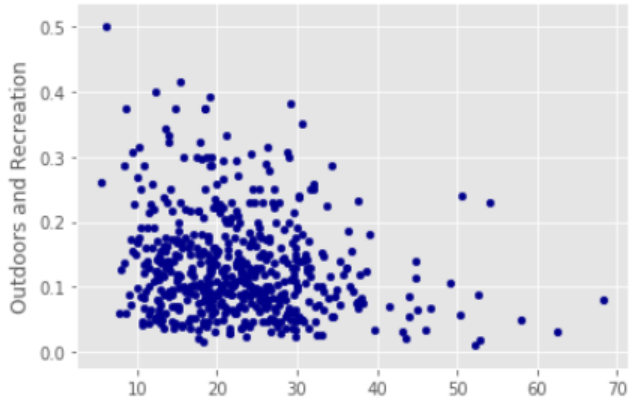
Dependency between the categories and crime rate











Predicting models

- Decision tree model was built with 0.4910 accuracy
- SVM model was built with F1 score of 0.42117

Conclusion and future directions

- Shown that crime rate is a complex measure which doesn't depend to venues' categories in a simplistic way
- This research can be performed for different type of crimes (Theft, Car crimes) to check if accuracy can improve
- This research can be applied to other large municipalities in the UK (Birmingham, Manchester) or even outside of the country