Clustering of London's Crime Areas

Alom Miah

24 April 2021

Contents

Introduction 3
The Data 4
The Crime Data 4
Foursquare Data 5
Methodology 6
Results and discussion 7
Crime Data 7
Venue Data 8
Conclusion 9

Introduction

The purpose of this project is to segment the different neighbourhoods of London by the different types of crime which occur. This would be of use to:

1) Residents of London - So they can protect themselves of such crimes.
2) Tourists - As above and for an awareness of what crimes to expect in different areas.
3) Businesses - As above and so they can take out appropriate insurance.
4) Insurance companies - Can set insurance premiums taking into consideration the prevalence of crime in a certain neighbourhood.

London will also be segmented by Venues (from the Foursquare API) to see if there is a link to crime.

The Data

The Crime Data

The Crime data was obtained using the UK Police's API/Crime Data, filtering the data from crimes in 2019 (to exclude impacts of Covid-19) by the Metropolitan Police Service which is London's police force. The crime data comes as monthly csv files which will be combined for the year of 2019.

Documentation on the API can be found on: https://data.police.uk/docs/ and https://police-api-client-python.readthedocs.io/en/latest/ and Crime Data from https://data.police.uk/data/

Crimes for which location data exists are presented by neighbourhood and latitude and longitude of the crime location. The latitude and longitude data of the neighbourhood of the crime will be used to cluster London's neighbourhood.

The latitude & longitude coordinates for centroids for neighbourhoods were created from the crime data creating a crime centroid for each neighbourhood. The "LSOA name" is the neighbourhood's name and "Crime type" is the other important heading in the crime data.

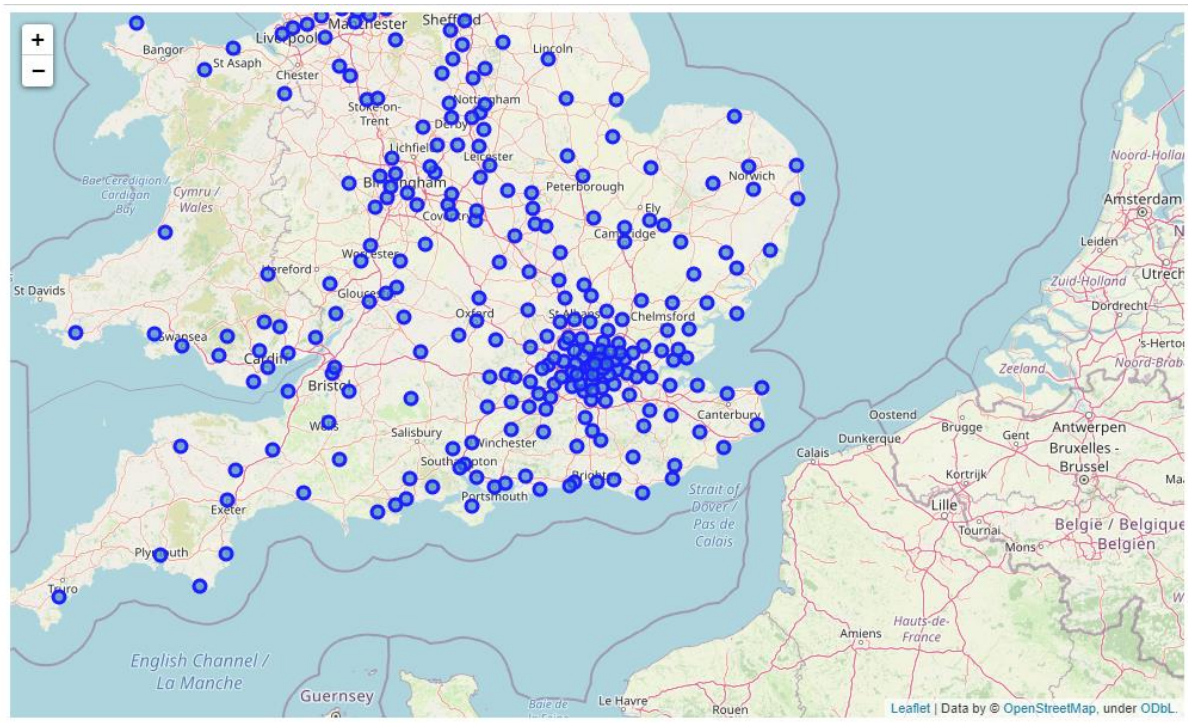

Figure 1 - Crime Neighbourhoods Before Cleaning Data

The crime data had to be cleaned as there were 1,108,042 reported crime incidents to the Metropolitan Police in 2019 but only 1,093,359 reported crime incidents actually took place in London which covers a significant portion of the 1,108,042 reported crime.

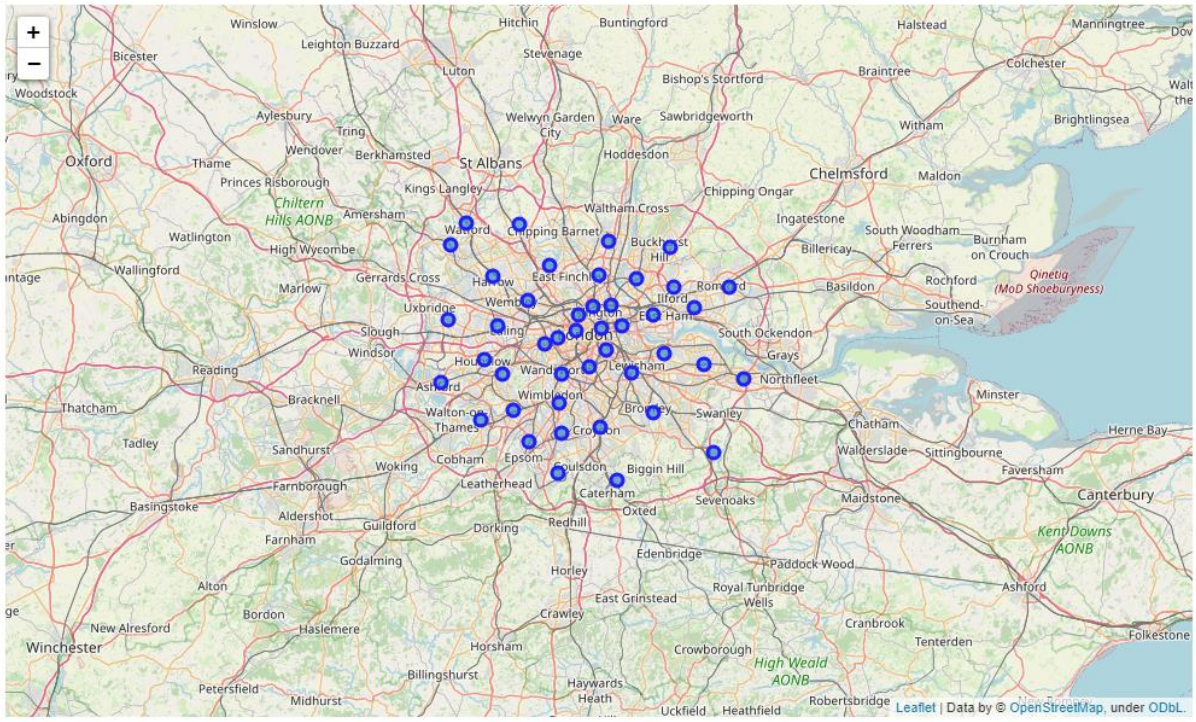

Figure 2 - Crime Neighbourhoods After Cleaning Data

Foursquare Data

Using the crime Neighbourhood data we can use the Foursquare API will be used to get info on venues in each neighbourhood.

The number of venues were found within a radius of 2km from each neighbourhood crime centroid. As we are looking at the crime centroid of each neighbourhood we want to see if nearby venues have an association with the crimes committed.

Methodology

For this project we looked to segment the different neighbourhoods of London by the most prevalent crime in that area.

In the first step we collected the crime data from the Metropolitan police and cleaned this data for crime that only occurred in London.

In the second step, we clustered the neighbourhoods by the most prevalent crimes in the area.

In the third step, we obtained venue data from the Foursquare API for venues near each neighbourhood centroid.

In the fourth step, we clustered the neighbourhood by the most prevalent venues.

In the fifth and final step, we compared the crime and venue clustered neighbourhoods to see if there is any similarity.

Results and discussion

Crime Data

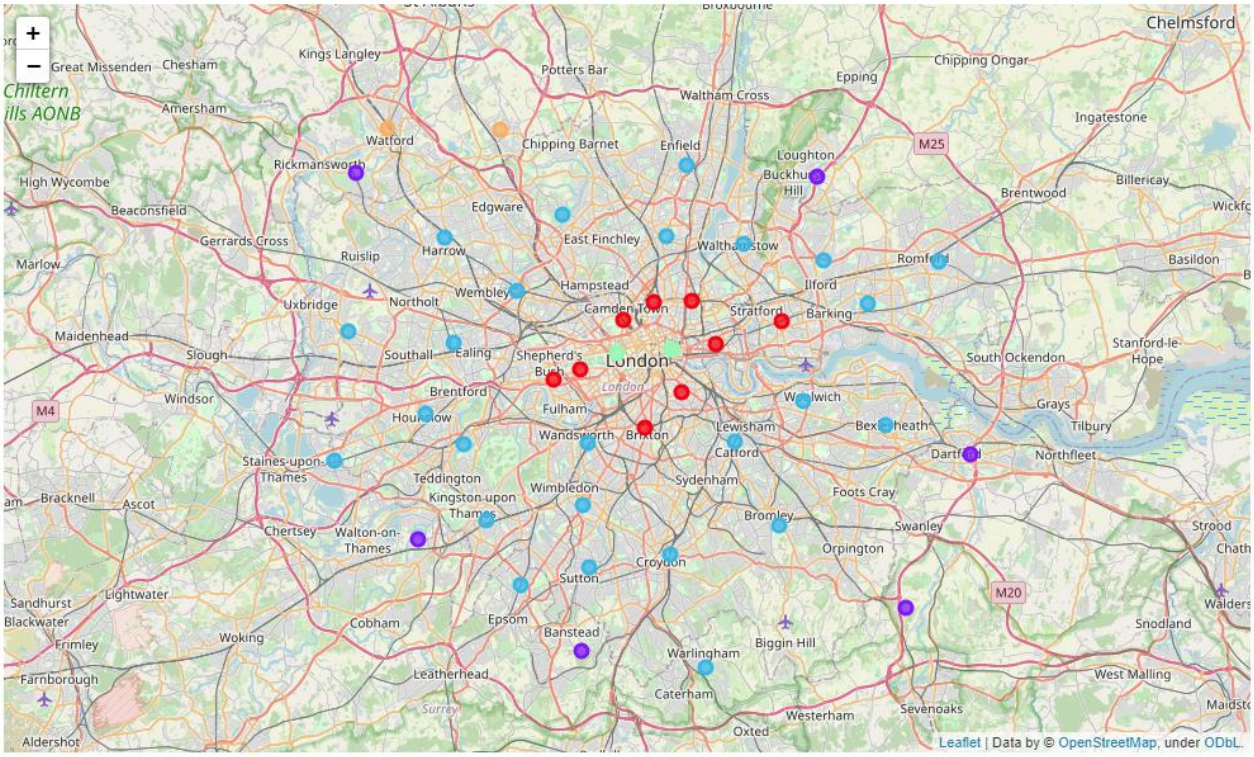

Figure 3 – Clustered crime categories in London

It looks like the clusters for crime in London vary radially from the centre of London. Delving deeper into the clusters (above) it appears that:

Cluster 3 (Green) - Central London: The crimes here are primarily theft related which makes sense as this is the centre of London with many tourist attractions and so there is ample opportunity for theft.

Cluster 0 (Red) - Inner City: These represent an area that surrounds the centre of London and is very populous with many residents and families here living close to the centre of London for work. The majority of crimes here are related to Anti-social behaviour and violence or theft. This makes sense due to these being a very densely populated area where people are less likely to get along.

Cluster 2 (Blue) - Inner Suburbs: This cluster can be described as the inner suburbs which is slightly less densely populated than the Inner city. The majority of crimes being committed here are a combination of violent crimes, anti-social behaviour and vehicle crime. This makes sense as we move away from the centre of London, the police force have more ground to cover which present more opportunity for criminals to be violent and not get caught immediately, also moving away from the centre of London there is more reason for people to have cars and therefore more opportunity to steal cars. However, it is still densely populated enough for anti-social behaviour still to be a dominant crime.

Cluster 1 (Purple) - Outer suburbs: This cluster can be described as the outer suburbs which is much less densely populated than the Inner city. The majority of crimes here are violent crimes and vehicle theft. Which makes sense as per the logic in cluster to as we move further away from the centre of London, the police force have more ground to cover which present more opportunity

for criminals to be violent and not get caught immediately, also moving away from the centre of London there is more reason for people to have cars and therefore more opportunity to steal cars.

Cluster 4 (Orange) - Outer towns: These locations can be considered as Towns within Greater London which have their own identities and issues. With crime types likely to be unique to each area. That being said the majority of crime is violent which makes sense as these areas are less densely populated than other areas in London.

Venue Data

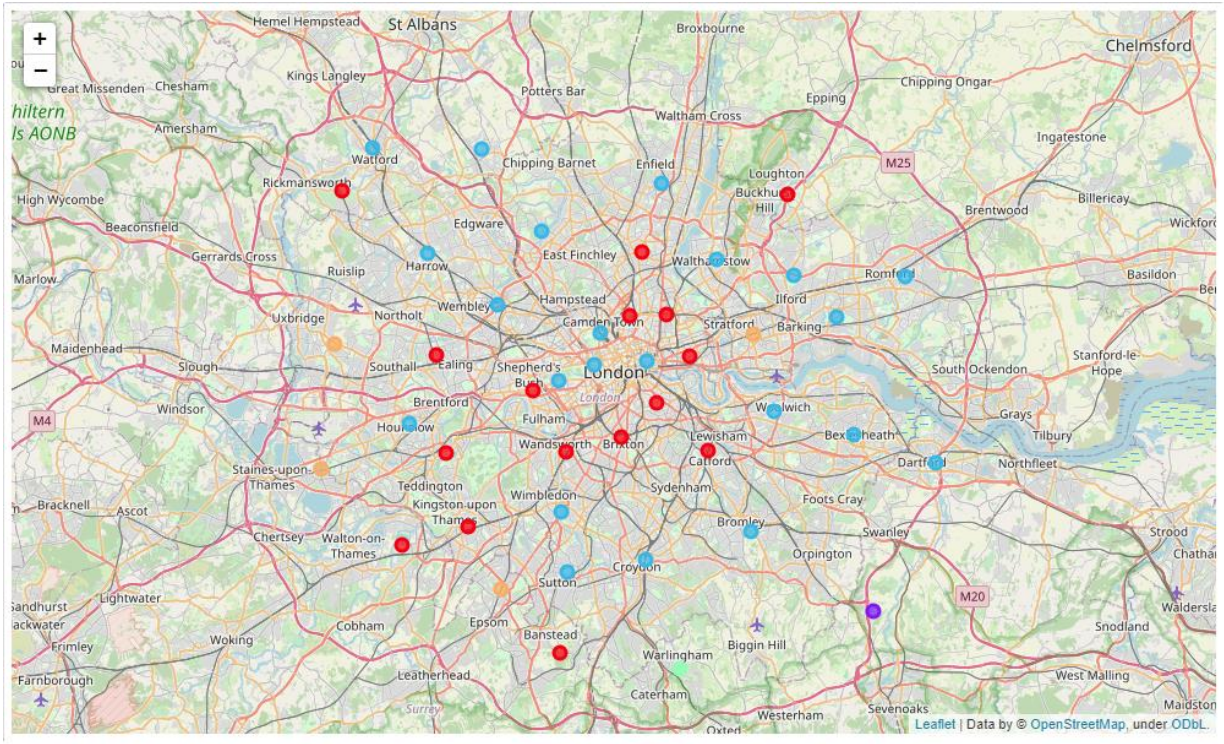

Figure 4 – Clustered venue categories near crime centroids in London's neighbourhoods

It looks like there is no clear geographic link between neighbourhood location in London and the clustering of venues. Regardless we will delve deeper into the clusters (see above) to see what the venue clusters represent:

Cluster 0 (Red) - This cluster appears to represent more diverse and complex neighbourhoods given the varying locations present at each location (similar to cluster 2). But it is clear the most crime in these locations occur near pubs.

Cluster 1 (Purple) - This cluster appears to similar to cluster 3 and contains only one neighbourhood on the outskirts of London with venues near crime hotspots mainly being the pub or train station.

Cluster 2 (Blue) - This cluster appears to represent more diverse and complex neighbourhoods given the varying locations present at each location (similar to cluster 0). It appears the most common crime location for this cluster is either a pub, cafe or grocery store.

Cluster 3 (Green) - This cluster appears to be similar to cluster 1 and contains only one neighbourhood on the outskirts of London with venues near crime hotspots mainly being the train station or park.

Cluster 4 (Orange) - This cluster appears to contain mainly residential neighbourhoods without much complexity with venues near crime hotspots mainly being grocery stores and pubs which makes sense as these are areas where many people interact with each other.

Conclusion

The purpose of this project was to segment the different neighbourhoods of London by the different types of crime which occur and the venues near crime locations. This was so that stakeholders could be aware of common crimes in a neighbourhood and any linkages to crime.

From the crime data (from the Metropolitan Police) we found that crime in London varies radially from the centre of London. With Central London being the centre of thefts and as you move outwards you will find anti-social behaviour morphing into more violent crime.

Clustering and looking at venues near the crime centroids of each neighbourhood (using the Foursquare API) there was no clear linkage between the crime location clusters and venue clusters and so it is unlikely that the clustering of venues gives rise to crime but more the opportunity at a geographic location and the people who live and work there. The venue data did however typically indicate pubs being the closest venue to the crime centroid of most neighbourhoods which makes sense as drinking does not lead to the best decisions.

Therefore, stakeholders can conclude that crime will most likely increase around locations containing pubs and in order to compensate for this local government should ensure establishments have the correct licences in place in order to compensate for potentially higher insurance costs and the need for more police presence.