Coursera Capstone

IBM Data Science Professional Certificate

The Greater London Borough Analysis

By

Enyinna Echetanna Egejuru

December 2020

The Greater London Borough Analysis

1. Introduction

Greater London is categorised as a cosmopolitan city being the home to roughly 8.9 million people (Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2019) from diverse backgrounds, London is also one of the top financial capitals in Europe. This city is made up of 33 boroughs each having its own distinct characteristics. Choosing a location within Greater London boroughs to setup shop or to reside can be extremely difficult objective to achieve, however with the help of population data, crime data and location data from Foursquare, we will aid you in making a well-informed decision.

Business Problem

In Greater London, which borough would be a better place to start up a business or live in? By using data science methodology and machine learning techniques like K-Means Clustering the objective of this capstone project is to analyse gathered data and select the best borough in the Greater London which aims to provide an answer to the above question highlighted.

Who would this report interest (Target Audience)

Entrepreneurs set out to either start a business or expand their existing business in the Greater London

Persons planning to move to Greater London and looking for the perfect location that will meet their expectations

Upcoming data scientists that would like to have an insight to some of the most used techniques to obtain data, analyse it and finally, tell a story out of it.

2. Data Description

For this project, the following data is required

- Population data of Greater London
- List of boroughs in Greater London
- Crime data for Greater London
- Co-ordinate data (Longitude and Latitude) to get the venue data
- Venue data

Data Sources:

1. Population / Location Data
   We get population data for Greater London, this data shows the 2019 estimates from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland/mid2

019april2020localauthoritydistrictcodes It contains population information for the United Kingdom broken down by age, this data is filtered to show the London 33 boroughs.
For geolocation data we use the geopy package in python to derive the co-ordinate data for the listed 33 boroughs.

2. Venues Data – leveraging on the Foursquare API
   Foursquare API provides point of interest information and geolocation based on coordinates provided, we will use this API to get relevant venue location information for Greater London.

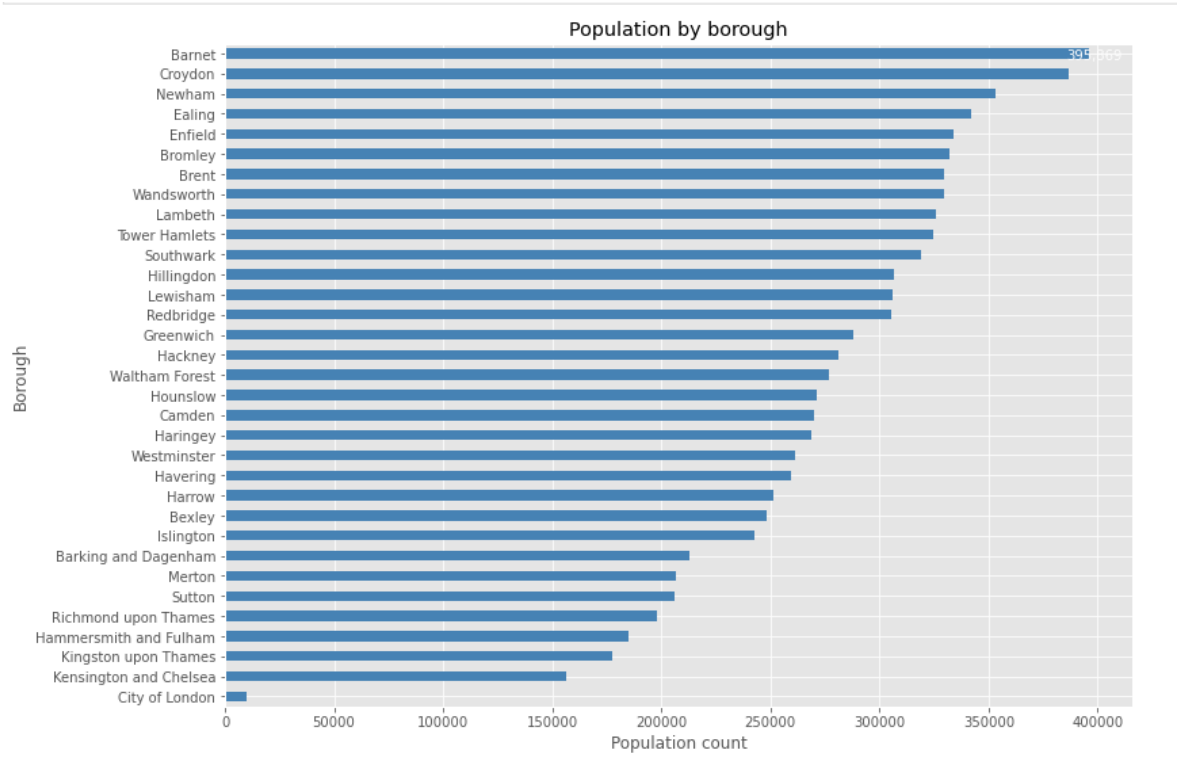
3. Crime Data
   Criminal activity for each borough will be analysed using Recorded Crime: Geographic Breakdown for Greater London - MPS Borough Level Crime (most recent 24 months) from https://data.london.gov.uk/dataset/recorded_crime_summary. It contains information about the number of crimes at borough levels of London per month, according to crime type from the Metropolitan Police Service.

3. Methodology

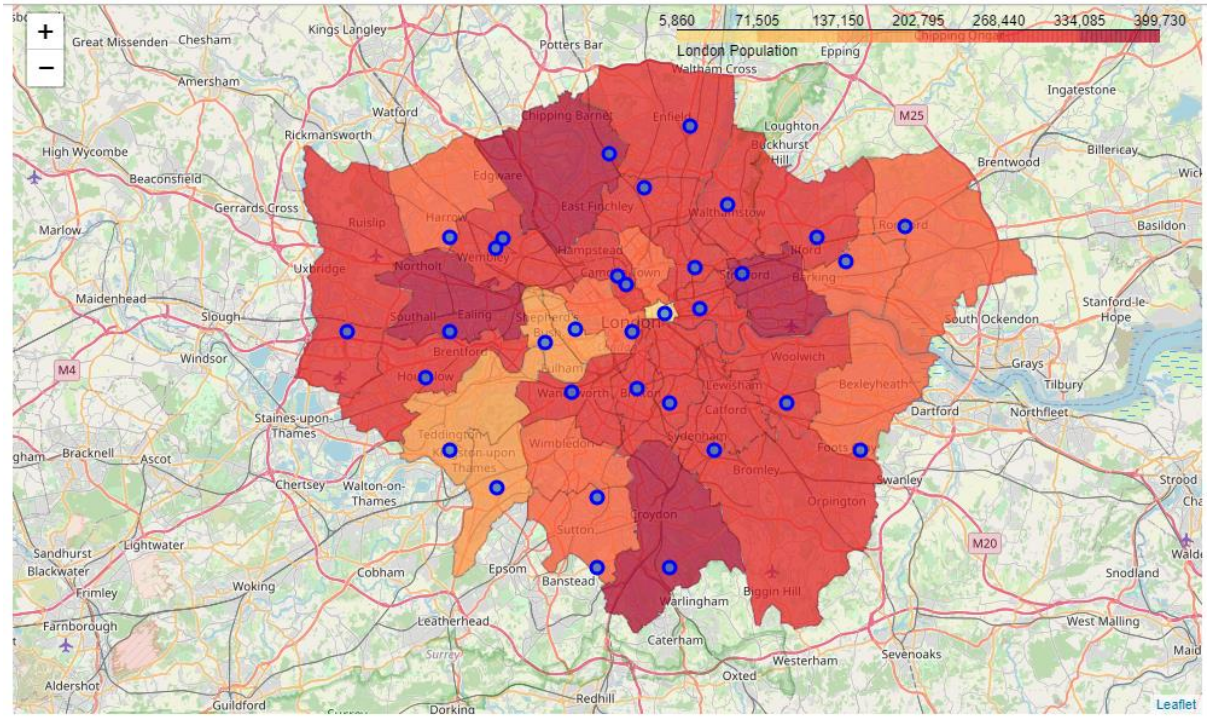
In this section we will explore the data gathered using visualisations to understand the data better.

Location and Population Data

Examining the population data, we see that the borough with the highest population is Barnet while the least populated borough is the City of London

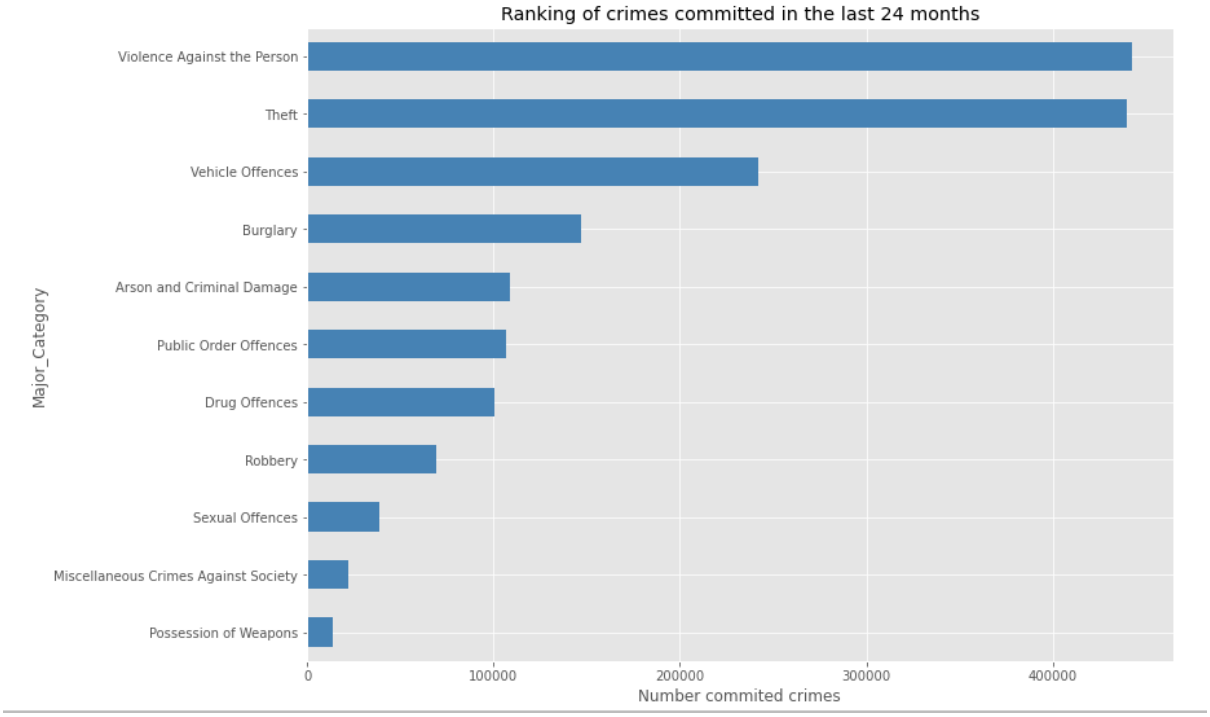


Using Geopy library we were also able to get the co-ordinates for the boroughs and plot them on a map using the Folium Geospatial visualization library.

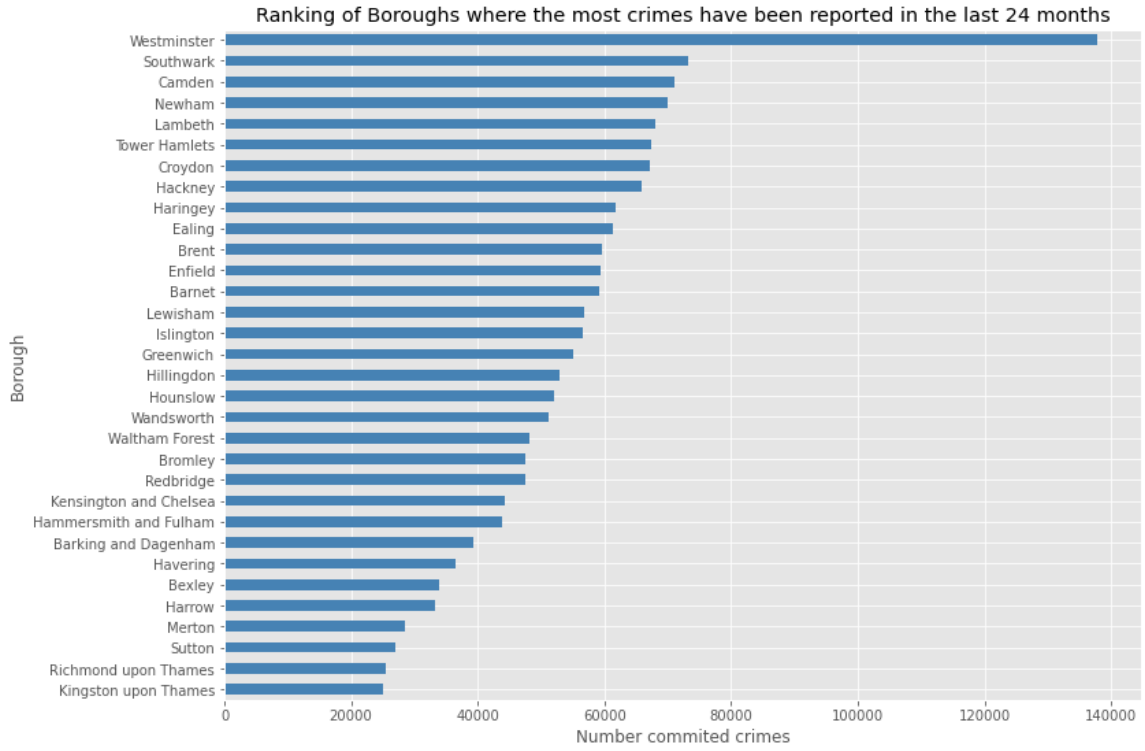


Crime Data

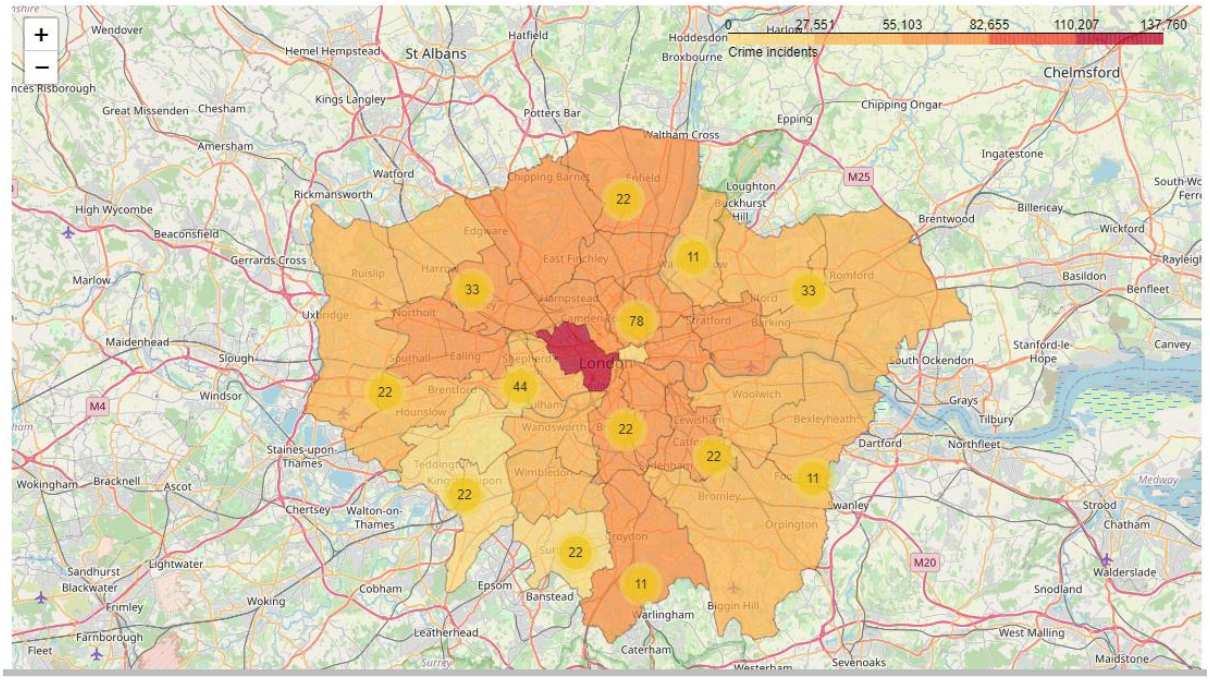
Examining the crime data, we can see the most committed crime is Violence against the person followed by theft and vehicle offences and the least committed crime is possession of weapons



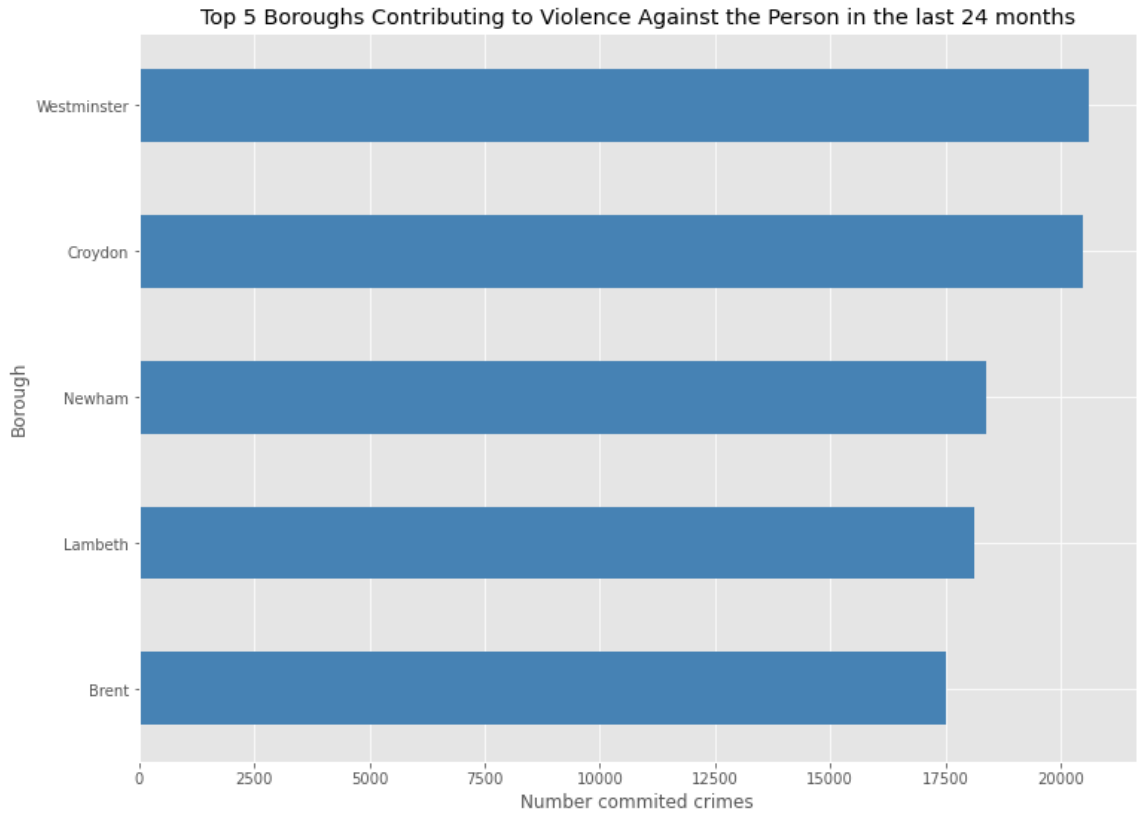
Looking at where these crimes where committed we can see that Westminster and Southwark are the boroughs ranking the highest committed crimes while Kingston upon Thames has the least reported cases.



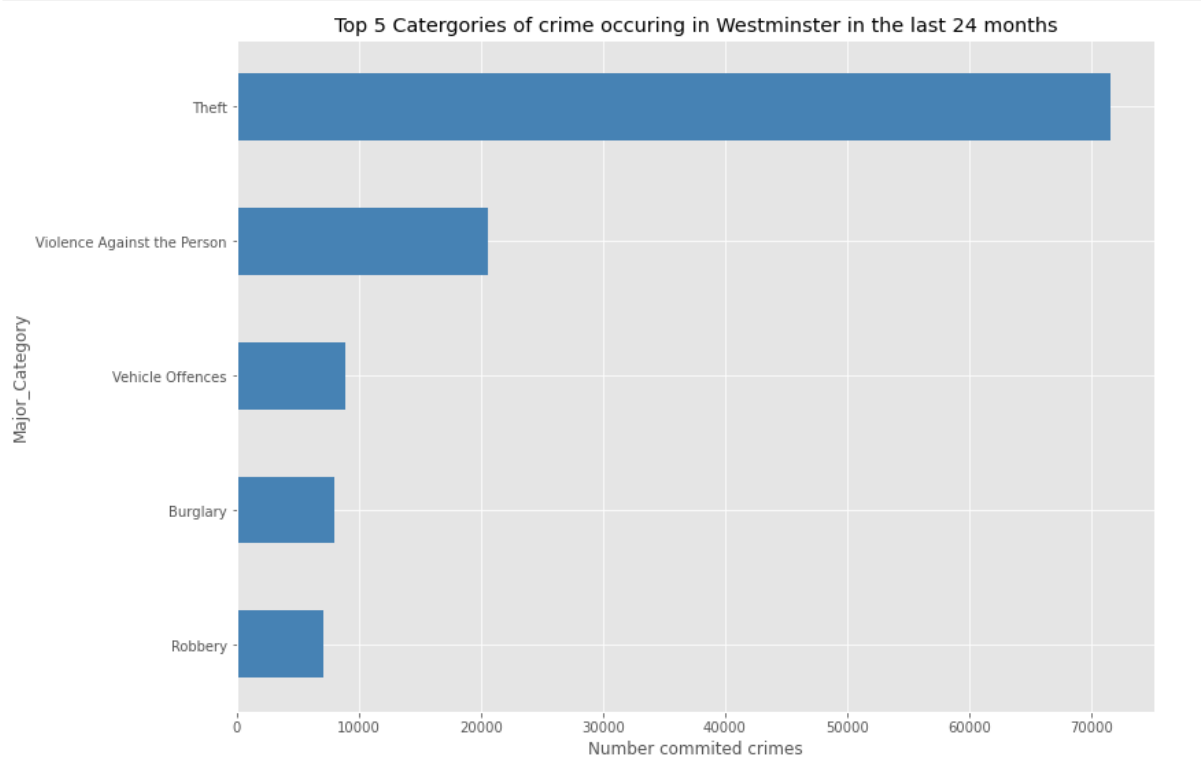
Plotting the above information on a map we have the following choropleth visualization below which shows the crimes committed by borough overlayed with the category of crime committed.



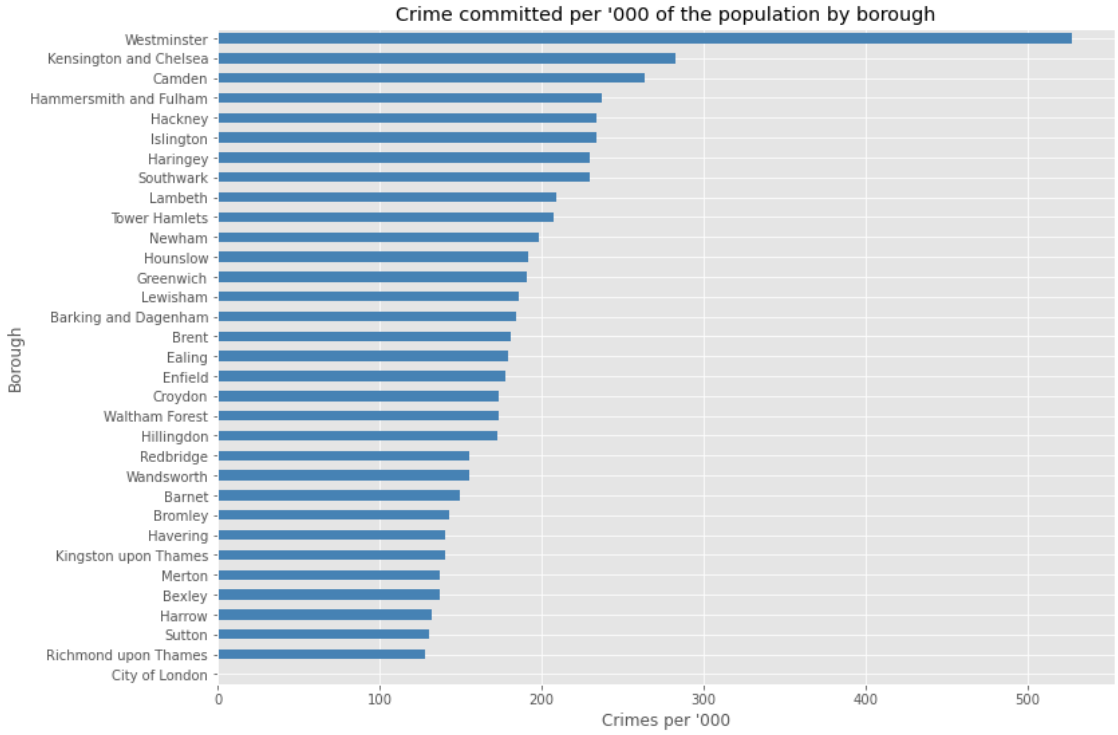
Delving deeper we breakdown the highest reported category of crime – Violence Against the Person and we see that Westminster is the borough where such case is being reported the most followed by Croydon then Newham.



On the other hand, when looking at the borough of Westminster we see that the most reported crime is Theft, followed by Violence Against the Person with the least case as robbery.

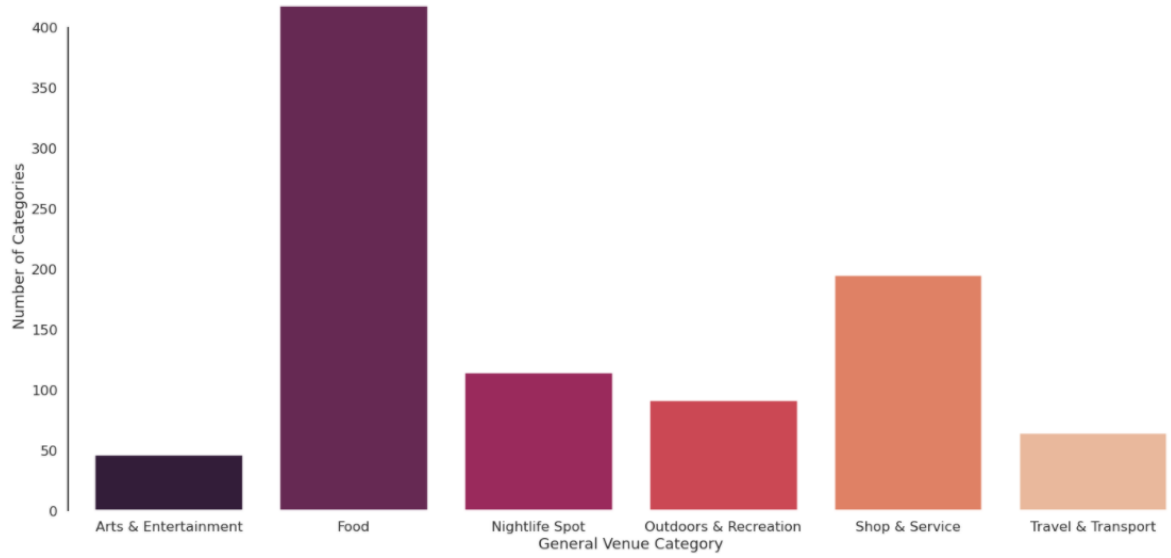


Nominalising the crimes committed by the population of each individual borough we see that Westminster also ranks the highest followed by Kensington and Chelsea and the least is Richmond upon Thames.



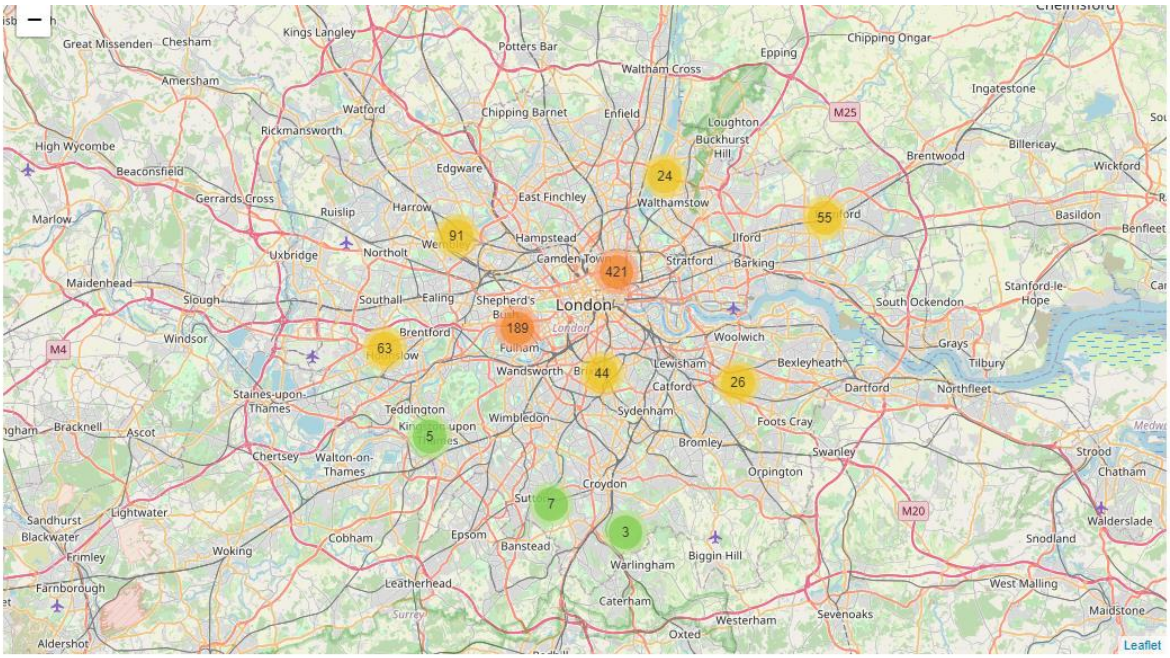
Venue Data

In deriving the Venue data, we take the borough information from the location data and gather the data of venues using Foursquare API and loaded it into a Pandas data frame, of which we had 928 venues, the venues were classified in to six broad categories – Nightlife spot, Food, Shop & Service, Outdoor & Recreation, Travel & Transport and Arts & Entertainment. Plotting the count of each category below

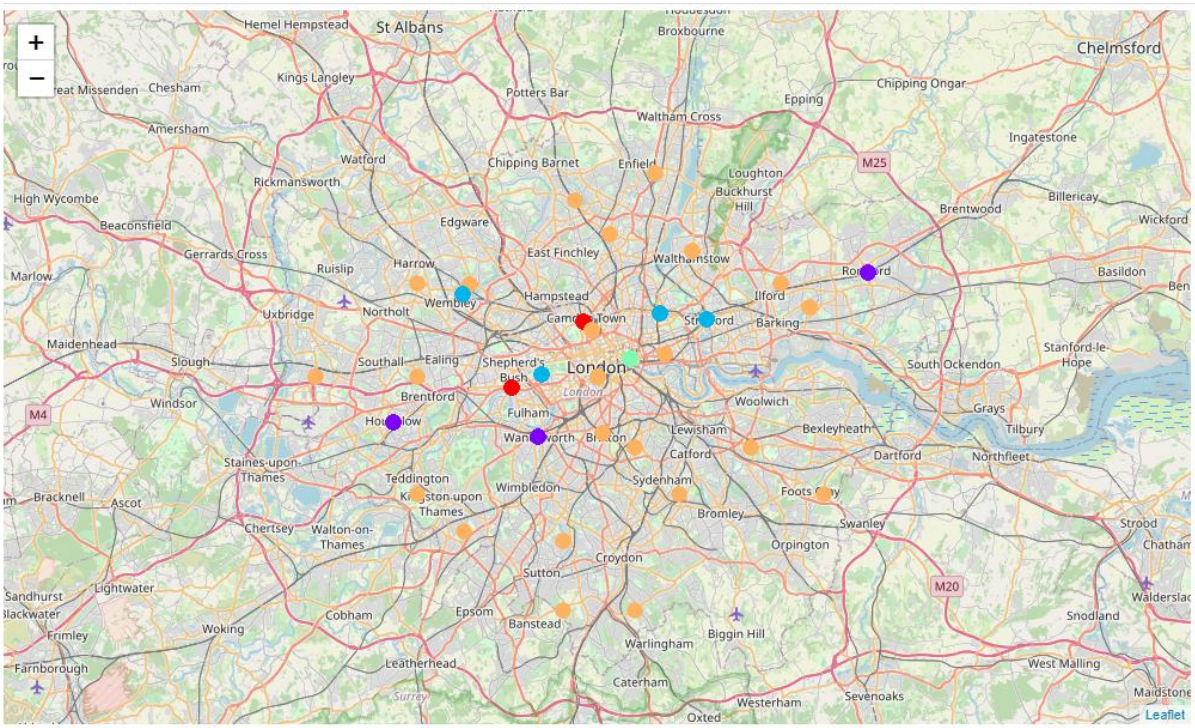


From the chart above we see that ‘Food’ is the most popular General Venue category among business owners followed by ‘Shop & Service’

Imposing the data into a cluster map we can how the different types of General Venue Category are spread across the city.

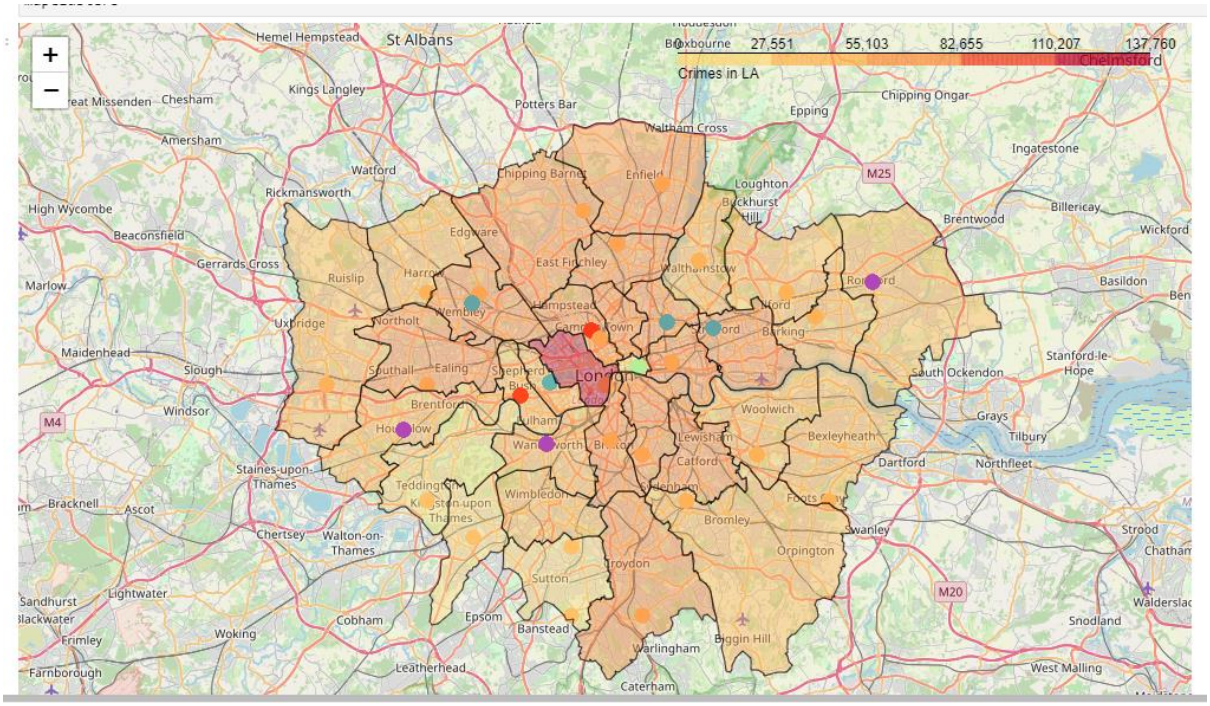


Clustering these venues into five clusters using the K-means algorithm and visualizing them on the map we have the following figure below



We can see that the cluster with the most venues is Cluster 4 and it is evenly distributed across the city.

Layering the clusters on the crime data we have the following map below



4. Results and Discussions

- All the venues can be grouped into 5 clusters
- Of all the clusters, Cluster 4 has least number of boroughs (1) and "Outdoor & Recreation" venue category is the most popular among boroughs in Cluster-5
- "Food" venue category is the most popular among boroughs in clusters 1,2 & 3
- Among all the venue categories, "Food" is the most popular category.
- "Outdoor & Recreation" seems to be the second popular venue category followed by "Shop & Service"
- Westminster and Southwark have the highest number of crimes recorded, however when compared to population Westminster and Kensington and Chelsea have the highest per capita ('000)

5. Conclusion

- In this Capstone project, we analysed the boroughs in Greater London, this data which contained the population density per borough was sourced from the Office of National Statistics, using Geopy and the Folium library, we were able to retrieve co-ordinate data and visualise this data on a map.

- Following this, we analysed the Metropolitan Police Service crime data for the past 24 months to determine which category and borough had the highest number of crimes and using a Choropleth map we visualised this data.
- Finally using the Foursquare API, we were able to retrieve details of venues in neighbourhood and divided them into 5 clusters, also plotting them on a map layering with the crime data.