



IBM Capstone Project Report



1 Introduction/Business Problem

Problem: ***Deciding where to build a new police station in London***

My project will use the python skills which I have learnt so far to help decide where to build a new police station in London. I have chosen to address this problem because I have a genuine interest in what the data may reveal. Areas with a high crime rates may benefit from police station of close proximity to improve response time to crimes. In addition to crime data, the report will also use demographic data of each London Borough and statistical analysis to help drive decision making. This analysis may be of use to the govt/police to assist them in a decision to construct a new police station based on crime hotspots.

2 Data

The data used will be used is:

Crime Data:

- The crime data is sourced from the London Datastore (<https://data.london.gov.uk/dataset/london-borough-profiles>)
- It is a CSV file which contains a range of information for each borough in London.
- The London Datastore describe says the data “helps paint a general picture of an area by presenting a range of headline indicator data in spreadsheet form to help show statistics covering demographic, economic, social and environmental datasets for each borough, alongside relevant comparator areas.”
- The CSV file will be downloaded and then the crime data for each borough will be extracted.

Location Data:

- The location data is sourced from CARTO (https://joshuaboyd1.carto.com/tables/london_boroughs_proper/public)
- It is a GEOJson file which contains the polygon coordinates for each London Borough.
- This will be used to help develop a choropleth map for crime rates across each London Borough.

Venue Data:

- The venue data about police stations is from Foursquare API call.
- Multiple data calls will be used from locations in each London borough to produce a map of all the current police station locations across London.
- This will be superimposed as markers with the Folium package.
- Geopandas can then be used to determine current number of police stations in each London borough.

3 Methodology

The first step was to produce a visual representation of the data on choropleth map. This involved:

- Getting data from Foursquare API about police station locations in London (`url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)`).
- This was then filtered and transformed into a dataframe.
- These were then plotted as blue circles on a map using the Folium package.
- Location data was uploaded to IBM Watson Studio from the GEOJson file.
- Crime data was uploaded to IBM Watson Studio as a CSV file and imported as a dataframe.
- The location data was combined with crime data to produce a choropleth map which was superimposed on the other map. Boroughs which were red had high crime rates and boroughs which were yellow had low crime rates.

The second step involved data analysis to see if there was a relationship between crime rates and number of police stations:

- The number of police stations in each borough was calculated and then stored in dataframe.
- A scatter graph was plotted with the number of police stations as x-axis and the total crime in 2014 as the y-axis.
- A linear regression was conducted using matplotlib to determine if there was a relationship between the number of police stations and total crime per year. This was superimposed on the scatter graph.
- R squared was then used as an inferential statistical technique to determine how well the linear regression fitted the data.

The third step was to determine which of the boroughs had the most number of crimes per police station. Ultimately, this would help drive the decision about which borough needs the police station the most.

4 Results

Figure 1: Choropleth map of crime rate for each London Borough with blue circles representing every police station.

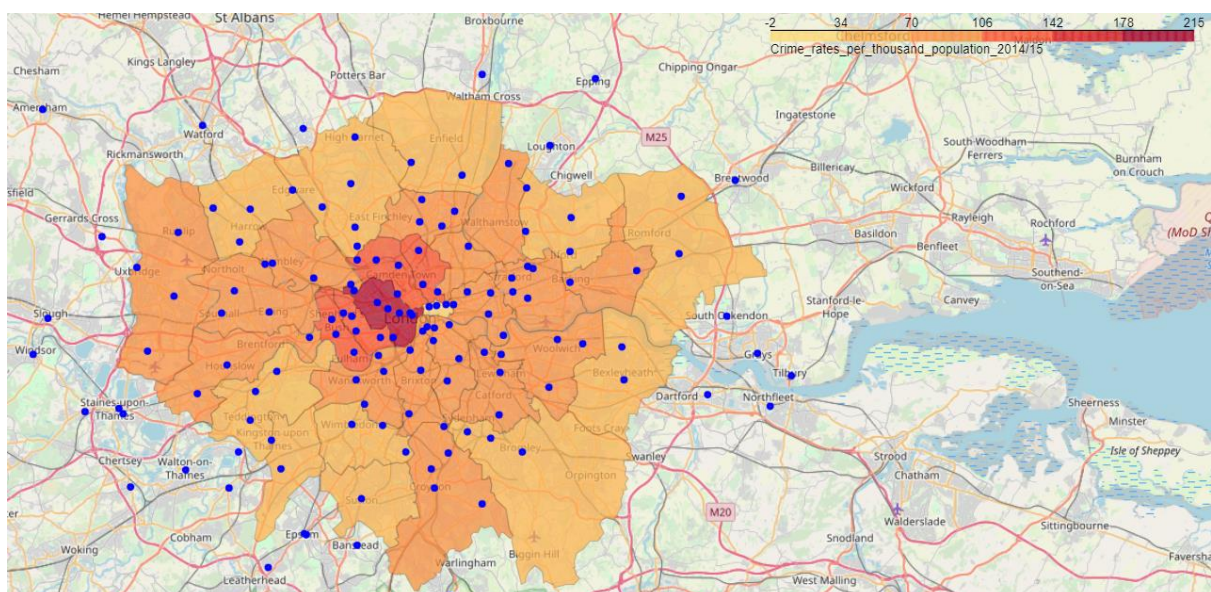
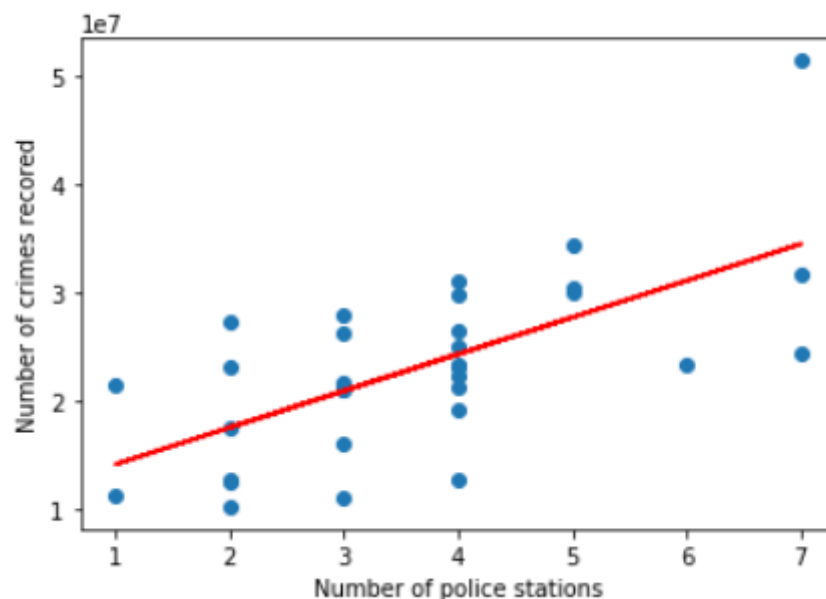


Figure 2: Scatter plot of relationship between number of police station and total crime for each borough with a linear regression.



- Results show a positive linear relationship between number of police stations and total crimes recorded. This shows that more police stations are put in areas with larger numbers of recorded crime.
- The R squared value was: **0.418**. This shows a low-to-moderate accuracy for the line of best fit.

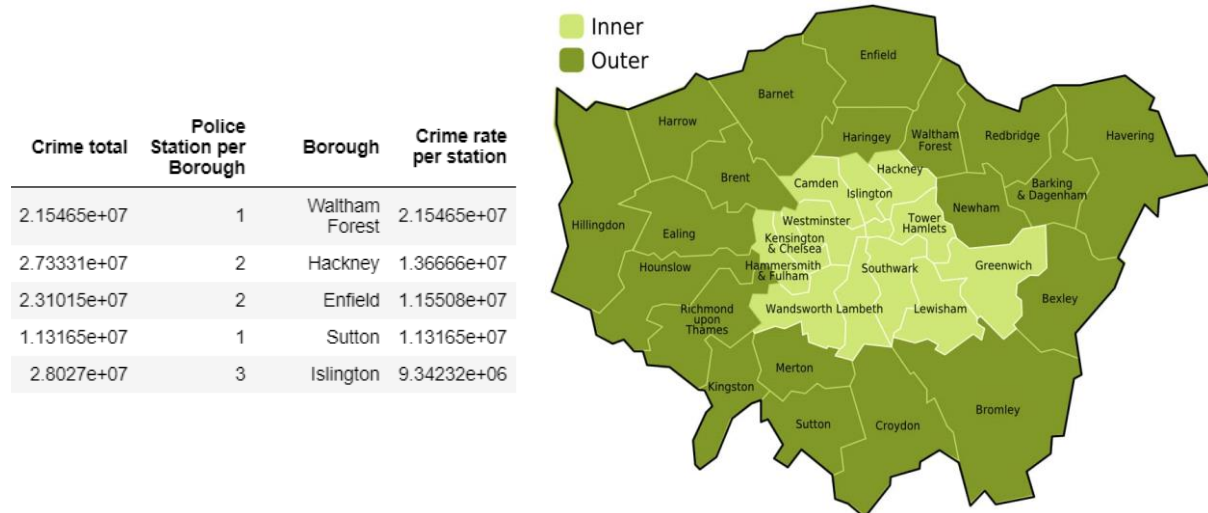
Figure 3: Table show the top 5 boroughs ranked by the total crime rate per police station.

Crime total	Police Station per Borough	Borough	Crime rate per station
2.15465e+07	1	Waltham Forest	2.15465e+07
2.73331e+07	2	Hackney	1.36666e+07
2.31015e+07	2	Enfield	1.15508e+07
1.13165e+07	1	Sutton	1.13165e+07
2.8027e+07	3	Islington	9.34232e+06

5 Discussion

The chorpleth map provided a very nice visualisation of the data. However, due to the number of diffenet police stations, it was difficult to make any inferences/insights into the data. Therefore, a linear regression was used to provide analysis over whether there was a pattern between the total number of crimes in a borough and the number of police stations in a borough. The results showed that there was a correlation, albeit weak. This indicated that boroughs with higher crime volumes normally had a higher number of police stations. This

analysis was then used to determine which boroughs had the highest crime rate per station. The results are showed in the Table 3 above (and copied below):



This can be used to determine where new police station should be built. It can be seen that the the two boroughs with highest crime rate per police station are Waltham Forest and Hackney. Interestingly, they are also very close geographically. Therefore, **I suggest that the new police station be built in Waltham Forest** in the south of the borough so that it can serve both boroughs. Additionally, Figure 1 shows that currently, the single police station in Waltham Forest is located towards the north of the borough. Therefore, it would definitely make sense to try and construct a new police station in the south of the borough if possible.

I think it is important to discuss the limitations of the analysis. The analysis made no consideration of:

- The size of each police station (e.g. number of cars, officers etc.).
- The area of each borough which needed to be covered.
- Proximity of police stations in a different borough which could respond to the same crime.
- Only a single year of crime data was considered.

A more in-depth study could be completed to consider these factors and provide a more accurate answer.

6 Conclusion

The project used data science techniques to help answer a real-world problem. Location data form Foursquare was combine with socioeconomic data to help determine where to build a new police station in London. The final result concluded that a new police station should be built in Waltham Forest. This decision was made using a range of different inputs. Future studies could use even more indicators and deeper analysis to provide an even more accurate answer. However, I think the results of this report will be useful for the UK Govt and for London planning authority.