Investigation into the relationship between local businesses and venues and crime in London

IBM Data Science Professional Certificate Capstone Project

Author: Brenda Cundy
Date: 03 July 2021

Table of Contents

Table of Figures

1. Disclaimer

This is not a formal analysis. It has been commissioned as a capstone project for IBM Data Science Professional Certificate. No reliance should be placed on conclusions and analysis within. This report has been produced under a number of constraints, including time and requirements of the capstone project and should not be considered fully representative of the quality of work the author can produce under professional conditions.

2. Executive Summary

This report looks at the link between types of business in London and crime in the surrounding area.

Crime, such as burglary estimated to cost businesses £1.6 billion a year, and this report has found that increased business units in an area does indeed lead to an increase of burglary, though it is not clear if this is simply due to the increase in targets. Other crimes are also increased, and tentative findings of note include Finance and Insurance causing a disproportionate increase in violent disorder – though it is not discounted that this not due to such businesses tending to be located in hotspots for such crimes.

Construction and motor trades were tentatively associated with a decrease in crime.

An in depth study of venues with a 24 hour alcohol licence was conducted. The evidence was inconclusive, although suggestive that they were indeed linked to an increase in crime, particularly of harassment. Of note there was no evidence large supermarkets with a 24 hour licence increased crime, indeed the evidence suggested they decreased it, but caution is recommended due to the small variation in numbers.

Due to limitations on availability of data this study was conducted at London borough level, which aggregated a diverse set of neighbourhoods and produces a small sample size, therefore it is recommended further analysis conducted to verify these conclusions, suggestions are outlined at the end of this report.

It is recommended that the implications on the crime rate are considered when granting 24 hour alcohol licences and the impact of other businesses is weighed against their benefits.

3. Introduction

London is the capital city of the United Kingdom, with an estimated population of 9.4 million[1]. Modern London is a diverse city, with varied neighbourhoods. Crime is a concern to residents, workers and business owners alike, in May 2021 a YouGov survey revealed that crime and housing were the biggest concerns to Londoners[2].

The costs of crime can be direct i.e. for the victim of theft, or indirect, i.e. loss of business revenue, increased insurance. In the 2018 report *The economic and social impact of crime 2nd Edition*[3] the Home office included estimated costs of UK crime in 2015/16 to be £50 billion against individuals and £9 billion against businesses.

3.1 Problem Statement

This report will address the following question:

- Within the context of London, is there evidence of a relationship between the types of facilities/industries in a locality and crime?

3.2 Scope

In scope:

- Identification of overall trends
- Consideration of different categories of crime
- Deep dive into the impact of 24 hour alcohol licensed premises
- Recommendations for follow-up analysis

Out of scope:

- Understanding the nature of a relationship i.e. if increased presence of certain businesses is in response to crime rate or a cause of it

3.3 Stakeholders

Then intended stakeholders for this report are:

Stakeholder	**Interest**	**Example**
Local planners	for consideration in approving planning applications, and deciding planning strategies	If 24 hour licensed premises are shown to be associated with increased crime licenses applications for such licenses may be rejected
Local residents and business owners	For understanding potential impact of local facilities and businesses on crime rate	if certain venues are shown to be associated with increased crime they may wish to petition against such venues opening.
Prospective residents and business owners	For understanding potential crime profile on an area based on the local facilities and businesses	if certain venues are shown to be associated with increased crime, they may wish to avoid properties near to such such premises

1 https://worldpopulationreview.com/world-cities/london-population
2 https://yougov.co.uk/topics/politics/articles-reports/2021/05/06/housing-and-crime-most-pressing-issues-facing-new-
3 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/954485/the-economic-and-social-costs-of-crime-horr99.pdf

4. Background

Geographic Considerations

Insight into London's history provides context for the formation of neighbourhoods and political boundaries against which data is collected. London is a diverse and a historic city, first referenced as Londinium in the 2nd Century AD. It is the largest city in the UK, and currently, by population the 32nd largest in the world (source: worldometer). London was subject to rapid growth associated with industrialisation the 19th Century AD, when the population grow fivefold, and immigration in the late 20th century following mid century decline. On the whole, the growth of London has been organic rather than planned. Many neighbourhoods feature 19th century street layouts, interspersed with 1940s bomb damage and 1960s slum clearance and redevelopment. Redevelopment in old and new neighbourhoods continues to this day.

In addition to population growth the geographical area considered London has grown to incorporate what were once separate settlement boundaries. Notably in 1965, London was redefined, and what were parts of neighbouring counties became included as London boroughs This history has resulted in a diverse geography, with neighbourhoods not necessarily fitting within the political boundaries by which London is governed. Many London neighbourhoods represent historic parishes, once separate to London., others have developed around transport links, for example the London underground stations.

Today London is governed by division into 32 local government districts referred to boroughs each governed by a London borough council..These make up the ceremonial county of Greater London. The "City of London", the historic centre of London, is a separate ceremonial county. However, the two counties together comprise the region of Greater London, all of which is also governed by the Greater London Authority. This is the definition of London used in this report.

Each London borough is subdivided into wards. These wards represent areas of population and seats in the borough council, they are not distinct communities and can be residential with commercial centres and amenities in nearby wards.

Crime data Collection

Greater London is covered by three separate police forces, which collate crime data

Police Force:	**Responsible for policing of:**
Metropolitan Police	The vast majority of London
City of London Police	City of London
British Transport Police	The national rail network and the London Underground

Crime Categorisation

Different stakeholders will primarily be concerned with different categories of crime. *The economic and social impact of crime 2nd Edition* captured particular concern for businesses: *Thefts from businesses make up almost 90% of business crime but account for approximately half of the total estimated costs of crime against businesses (£4.2bn), as each crime has a low impact on society. In contrast, robberies and burglaries against businesses – estimated to cost £2bn and £1.6bn respectively – make up over 40% of the costs of crime, but account for only 5% of all crimes against businesses.*

5. Data

Two primary sources of data will be used for this report:

1. **Foursquare**: an independent location data platform, which harvest details and recommendation of venues as input by users of it's mobile application "Foursquare City Guide"
2. **The London Datastore:** a free and open data-sharing portal, provided by the London Assembly and Mayor of London.

This reports utilises three groupings of data:
a. Crime
b. Local businesses/venues
c. Ancillary data to support analysis

5.1 London Crime Data

Data source: The London Datastore. *Recorded Crime Summary: Geographic Breakdown*. Provided by the Metropolitan Police https://data.london.gov.uk/dataset/recorded_crime_summary

Summary of data:

MPS Borough level crime (most recent 24 months):	Number of offences by month, by category (both major and minor), by borough for 24 months (the most recent that data is available for)
MPS Ward level crime (most recent 24 months):	Number of offences by month, by category (both major and minor), by ward for 24 months (the most recent that data is available for)
Borough level crime (historic):	Number of offences by month, by category (both major and minor), by borough for 2008 to 2018

	MajorText	MinorText	LookUp_BoroughName	201905	201906	201907	201908	201909	201910	201911	...	202007	202008	202009	202010	202011	202012	202101	202102	202103	202104
0	Arson and Criminal Damage	Arson	Barking and Dagenham	11	3	5	3	6	9	8	...	4	6	2	7	4	2	4	6	4	6
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	140	113	134	118	109	109	97	...	122	114	116	120	100	109	100	104	80	100
2	Burglary	Burglary - Business and Community	Barking and Dagenham	21	27	31	35	37	30	30	...	28	23	32	21	18	24	20	18	14	12
3	Burglary	Burglary - Residential	Barking and Dagenham	114	96	71	67	80	97	114	...	72	63	54	68	90	91	69	90	71	75
4	Drug Offences	Drug Trafficking	Barking and Dagenham	9	6	11	8	7	9	14	...	21	9	12	13	17	13	12	9	7	6

Figure 1: MPS Borough Level Crime (most recent 24 month). First 5 records

See appendix for details of crime categories

Data cleansing: None required, data is complete and good quality.

Data limitations: Data used is crime statistics published by the Metropolitan police so will exclude the City of London, and crime recorded on the London underground and national rail network.

5.2 Local businesses/venues data

This report will consider three different sources of data reflecting the facilities and venues. The London datastore contains additional relevant data which could be used for follow-up research.

5.2.1 Foursquare

Data source:
GET https://api.foursquare.com/v2/venues/explore
Foursquare API “explore”[4] endpoint returns up to 50 recommended venues. These venues are those from a geographic area either specified by a

- geocodable location name, or
- longitude/latitude coordinates provided with a search radius.

The underlying Foursquare places database is crowd sourced.

The most meaningful approach to harvesting Foursquare venue data for this analysis will be finalised post data exploration. Options include by name or coordinates and by neighbourhood or borough

Summary of data:
Foursquare API returns a JSON response, details of which can be found on the developer reference guide.

For this exercise the

- venue name
- venue location (latitude, longitude)
- venue category

will be considered.

The category provides a description of the type of venue and allows grouping of similar venues for further analysis

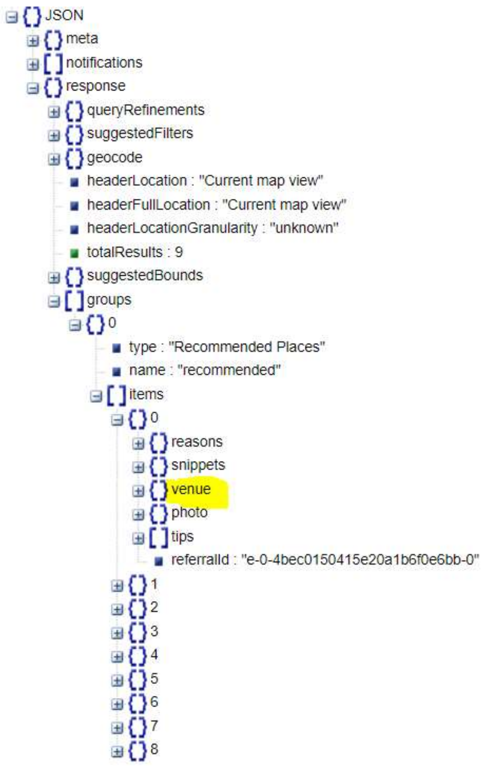


Figure 2: Example Foursquare response

4 https://developer.foursquare.com/docs/api-reference/venues/explore/

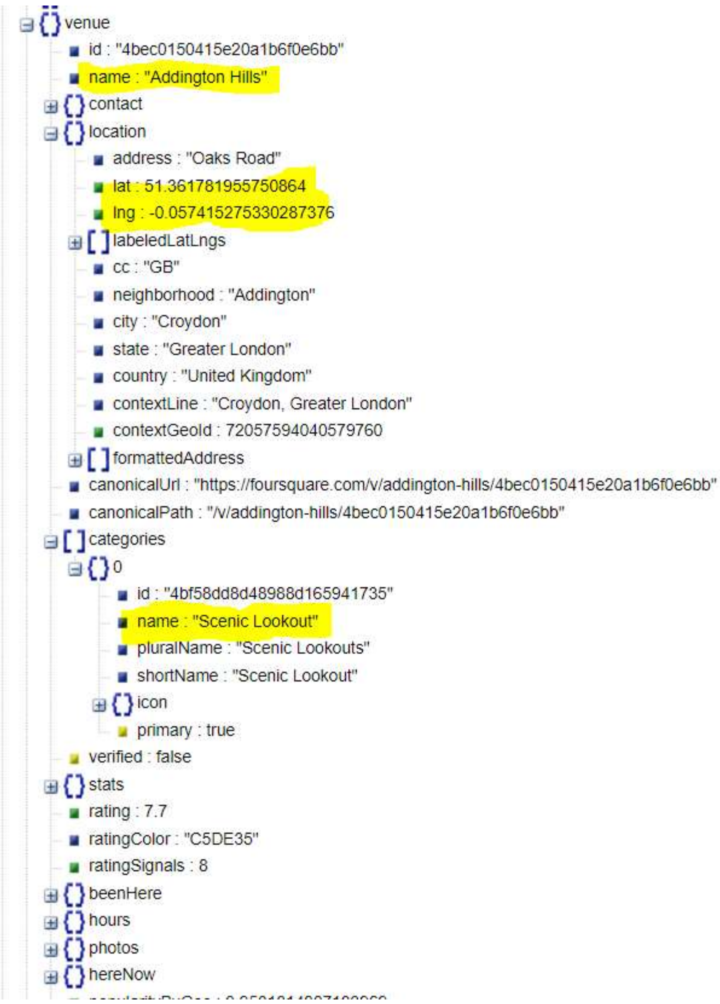


Figure 3: Example Foursquare response - venue node expanded

Data cleansing: Not required

Data limitations: There are a number of limitations with foursquare data

1. Crowd sourced data is biased by the demographic which uses foursquare
2. The geographic area data is obtained for may not match, or be representative of the geographic area crime data is available for
3. The limit of responses may prevent the returned venues being representative of the area searched. This limit is 50 as per documentation, but experience suggests is actually 100.
4. The allocation of category can be subjective, for example coffee shop vs cafe, pub vs gastropub, fried chicken joint vs fast food restaurant.

5.2.2 Venues with 24 Hour Licence

This data allows detailed analysis into a particular venue type.

Data source: The London Data Store *Alcohol and late-night refreshment licensing statistics – licensed premises 24 hour*. Provided by the Home Office.
https://data.london.gov.uk/dataset/alcohol-and-late-night-refreshment-licensing-statistics

Summary of data:

Data is provided on an annual basis, with the most recent data being from 2018.

Per year the number of licensed premises is provided for each borough (licensing authority), broken down by premises type with totals and subtotals across categories of premises.

Propose to use 2018 data with 2017 data used to fill in gaps

	Premises with 24-hour alcohol licences											
			Supermarkets and stores				Hotel bars					
Licensing authority	Total	Pubs, bars and nightclubs	Total	Large supermarkets	Other convenience stores	Supermarket and store type not reported	Total	Open 24 hours to residents and general public	Open 24 hours to residents and their guests only	Hotel bar type not reported	Other premises types	Premises type not reported
Barking and Dagenham	:	:	3	1	2	0	:	0	:	:	0	:
Barnet	:	:	:	:	:	:	:	:	:	:	:	:
Bexley	4	0	4	1	3	0	0	0	0	0	0	0
Brent	46	2	28	4	24	0	3	3	0	0	13	0
Bromley	14	0	11	4	7	0	3	0	3	0	0	0
Camden	:	:	:	:	:	:	:	:	:	:	:	:

Figure 4: Example data on licensed premises (24 hour license)

Data cleansing:
Data is cleansed to enable analysis:

- There are a number of gaps in the data (shown as :), where that data is not available. Use the previous years' value if available.
- Ensure totals and subtotals reflect any data cleansing
- If any data is missing for 2 consecutive years exclude
- Exclude any features (columns) where more than eight records are 0, as there will be insufficient data for meaningful analysis

Borough	Total	Total Supermarkets and stores	Large supermarkets	Other convenience stores	Total Hotel bars	Open 24 hours to residents and their guests only
Barking and Dagenham	3	3	1	2	0	0
Barnet	0	0	0	0	0	0
Bexley	4	4	1	3	0	0
Brent	46	28	4	24	3	0
Bromley	14	11	4	7	3	3

Figure 5: Example cleansed data showing feature reduction

5.2.3 Local Units by industry

Data source:
The London datastore: *Local unity by broad industry group, borough.* Provided by the Office of National Statistics.
https://data.london.gov.uk/dataset/local-units-broad-industry-group-borough

Summary of data:

Data is available for each year 2003 to 2020. The number of businesses (local units such as a factory or a shop) by Broad Industry Groups, per borough is provided per year.

		UK SIC 2007																	
Code	Area	SIC07: 01-03 : Agriculture , forestry & fishing	SIC07: 05-39 : Production	SIC07: 41-43 : Constructio n	SIC07: 45 : Motor trades	SIC07: 46 : Wholesale	SIC07: 47 : Retail	SIC07: 49-53 : Transport & Storage (inc. postal)	SIC07: 55-56 : Accommodatio n & food services	SIC07: 58-63 : Information & communicatio n	SIC07: 64-66 : Finance & insurance	SIC07: 68 : Property	SIC07: 69-75 : Professional , scientific & technical	SIC07: 77-82 : Business administratio n & support services	SIC07: 84 : Public administratio n & defence	SIC07: 85 : Educatio n	SIC07: 86-88 : Health	SIC07: 90-99 : Arts, entertainment , recreation & other services	SIC07: Total
E09000001	City of London	20	730	725	25	605	995	285	1,280	2,185	3,700	1,055	8,005	6,105	55	215	330	1,050	27,365
E09000002	Barking and Dagenham	10	330	1,460	225	410	645	530	410	610	90	145	810	780	170	200	605	360	7,790
E09000003	Barnet	30	625	3,200	350	1,155	2,165	600	1,000	2,570	530	2,055	4,910	2,205	55	535	1,250	1,585	24,820
E09000004	Bexley	10	445	1,905	260	335	805	380	570	1,065	180	225	1,510	840	25	225	555	560	9,895
E09000005	Brent	5	585	2,255	410	955	1,600	670	915	1,870	270	755	2,630	1,355	60	320	790	1,015	16,460
E09000006	Bromley	60	495	2,355	290	525	1,665	310	855	2,080	390	510	3,465	1,570	25	380	900	1,195	17,070
E09000007	Camden	30	1,005	1,515	150	1,240	2,450	465	2,140	4,610	855	1,525	10,385	3,530	115	680	1,205	2,845	34,745
E09000008	Croydon	15	495	2,285	390	540	1,540	500	1,020	2,030	335	525	2,930	1,445	50	420	1,095	1,000	16,615

Figure 6: Example local units data as available from London Data Store

Data cleansing:
Data is cleansed to enable analysis:

- Additional UK regional data is included in the dataset, records removed as only London Borough data is required
- No gaps identified in data

Data limitations:
Number of units provides a single snapshot, in reality businesses are not static, opening and closing through the year.

5.3 Ancillary Data

The following data is utilised to support analysis

5.3.1 London Population Data

Population data enables the primary data to be population adjusted – e.g. crime rate per 10,000 capita to be calculated.

Data source:
London Datastore: *London Borough Profiles*. Provided by the Greater London Authority
https://data.london.gov.uk/dataset/london-borough-profiles
London Datastore: *London Ward Profiles*: Provided by the Greater London Authority
https://data.london.gov.uk/dataset/ward-profiles-and-atlas

Summary of data:
Extensive indicators are provided about each London ward and borough. Of interest to this analysis is population.
The population figure provided is:
2017 Estimate : Borough level data
2015 Estimate: Ward level data

	Code	Area_name	Inner/_Outer_London	GLA_Population_Estimate_2017	GLA_Household_Estimate_2017	Inland_Area_(Hectares)	Population_density_(per_hectare)_2017	Average_Age,_2017	Proportion_of_population_ag 15,_
0	E09000001	City of London	Inner London	8800	5326	290	30.3	43.2	
1	E09000002	Barking and Dagenham	Outer London	209000	78188	3,611	57.9	32.9	
2	E09000003	Barnet	Outer London	389600	151423	8,675	44.9	37.3	
3	E09000004	Bexley	Outer London	244300	97736	6,058	40.3	39.0	
4	E09000005	Brent	Outer London	332100	121048	4,323	76.8	35.6	

Figure 7: Header of Borough profile data

Data limitations:
It is assumed that population growth has been consistent across boroughs/wards since the date of the population estimates, therefore the population number is a reasonable proxy to use for current population.

5.3.2 Geocoding

Nominatim API is used to obtained coordinates of locations, based on a place name (i.e. London boroughs and neighbourhoods.
https://wiki.openstreetmap.org/wiki/Nominatim
This is used for data visualisations, and is investigated for utilising the Foursquare API to obtain venue data for a specific geography.

5.3.3 London maps

The python Folium library is used for visualisations http://python-visualization.github.io/folium/

London boroughs are overlaid using ***london_boroughs_proper.geojson*** sourced from:
https://joshuaboyd1.carto.com/tables/london_boroughs_proper/public/map
London wards are overlaid using ***london-wards-2014.geojson*** sourced from:
https://github.com/ft-interactive/geo-data

5.3.4 List of London Neighbourhoods

Wikipedia provide a list of London neighbourhoods is scrapped from
'https://en.wikipedia.org/wiki/List_of_areas_of_London'

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge[9]	ILFORD	IG2	020	TQ455895
Aldgate	City[10]	LONDON	EC3	020	TQ334813
Aldwych	Westminster[10]	LONDON	WC2	020	TQ307810
Alperton	Brent[11]	WEMBLEY	HA0	020	TQ185835
Anerley	Bromley[11]	LONDON	SE20	020	TQ345695
Angel	Islington[8]	LONDON	EC1, N1	020	TQ345665
Aperfield	Bromley[11]	WESTERHAM	TN16	01959	TQ425585

Figure 8: List of London neighbourhoods as shown on Wikipedia

5.3.5 List of London borough locations

The Wikipedia list of London boroughs provides coordinates for each
https://en.wikipedia.org/wiki/List_of_London_boroughs

Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2019 est)[1]	Co-ordinates	Nr. in map
Barking and Dagenham[note 1]			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	212,906	51.5607°N 0.1557°E	25
Barnet			Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	395,896	51.6252°N 0.1517°W	31
Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	248,287	51.4549°N 0.1505°E	23
Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	329,771	51.5588°N 0.2817°W	12
Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	332,336	51.4039°N 0.0198°E	20
Camden	✓		Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	270,029	51.5290°N 0.1255°W	11

Figure 9: List of London boroughs (in part) as shown on Wikipedia

6. Data Exploration

6.1 London Crime Data

This exploration has been conducted on the recent Borough Level crime data.

6.1.1 Crime trends over most recent 24 months

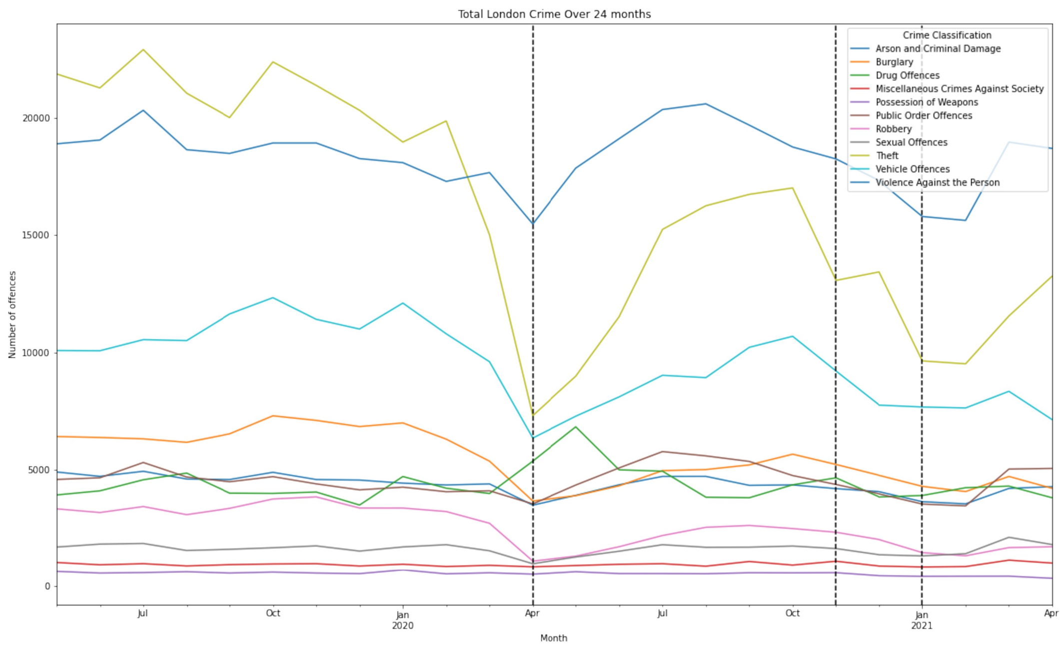


Figure 10: Crime, split to major category, over 24 months

This shows different major categories of crime over 24 months. This 24 month period includes the corona virus pandemic, and the UK lockdowns are approximately marked. The graph indicates that:

a) Crime trends have been influenced by the Covid 19 pandemic

b) Not all crimes follow the same trend.

For example “possession of weapon” has been consistent throughout the time frame, “drugs offences” increased with the initial lockdown whereas most crimes dropped

6.1.2 Crime trends across minor crime category

The major crime category “Drug Offences” is divided into minor crime categories as follows:

Major Category	Minor Category
Drug Offences	Drug Trafficking
	Possession of Drug

This is examined as an example to determine whether there is value in data analysis at the minor categorisation level.

The 24 month time series of crime data shows the two subcategories follow different patterns.

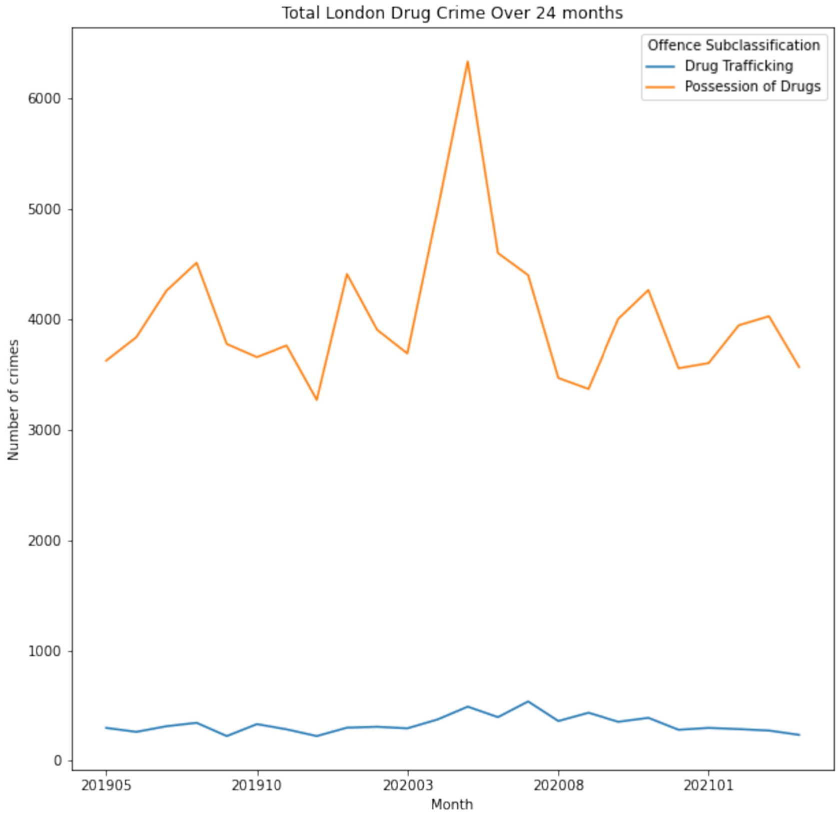


Figure 11: Drug offences split to minor categories over 24 months

The distribution of drug offences across the boroughs shows that the two sub categories of drug offences follow different trends to the major category.

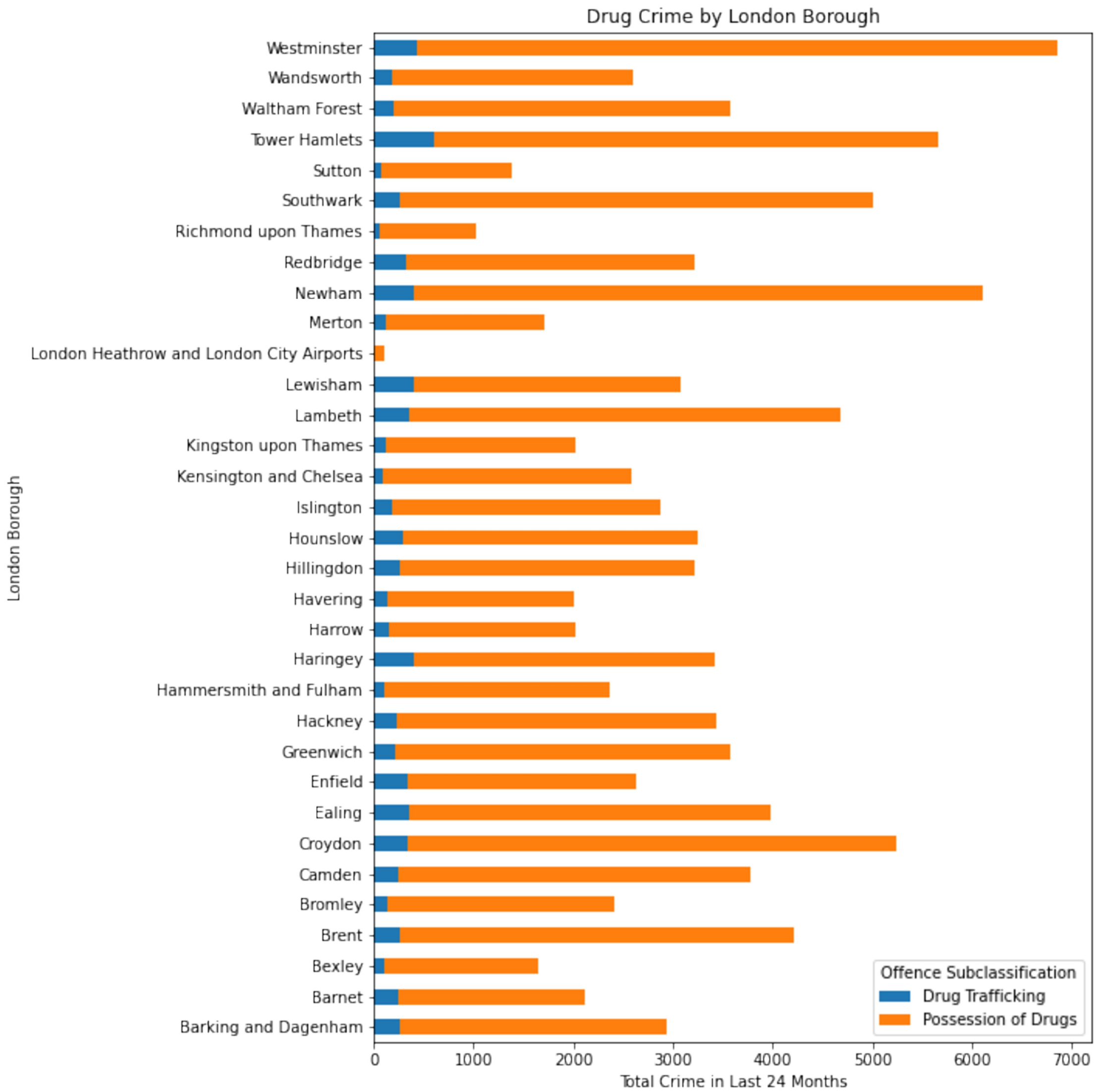
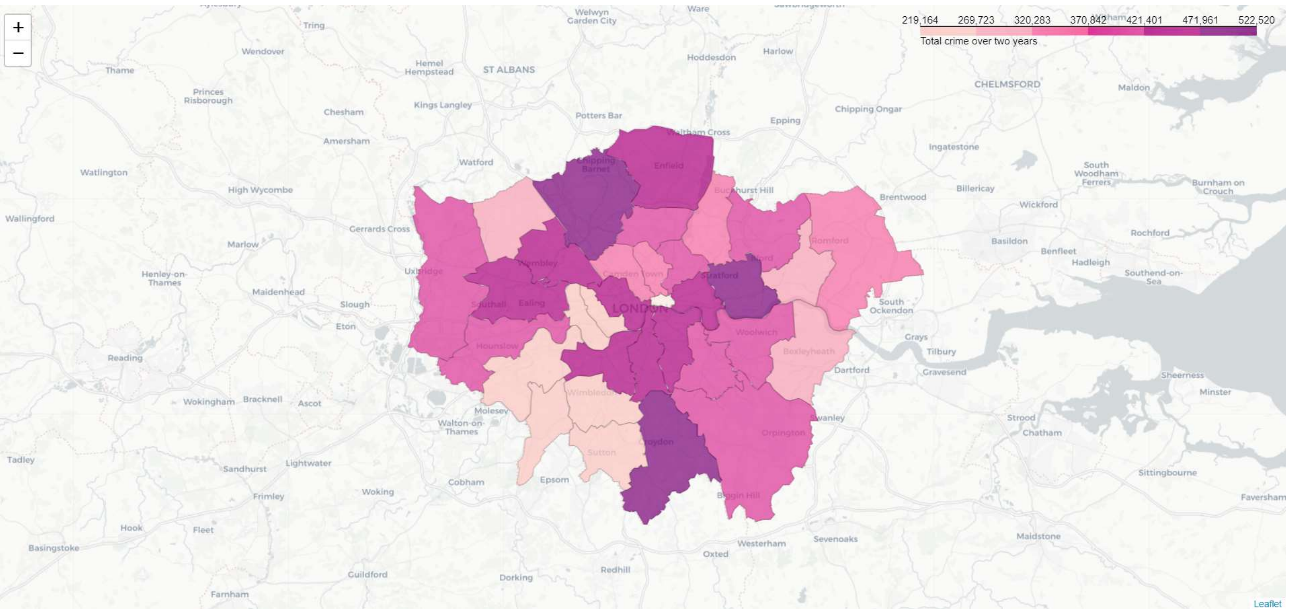


Figure 12: Total drug offences over 24 months split to sub category per Borough

6.1.3 *Crime trends across geography*

The distribution of crime across boroughs confirm crime is not evenly distributed.



***Fi*gure 13: Total crime in 24 months across London Boroughs**

The different categorisations of crimes are distributed differently. “

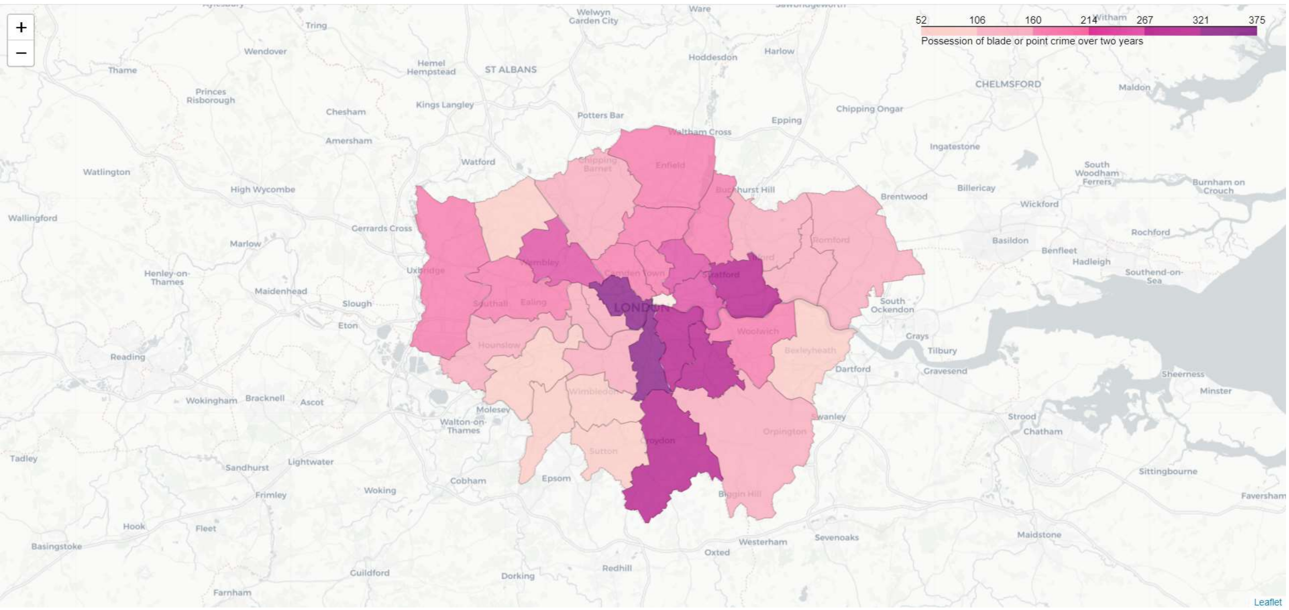
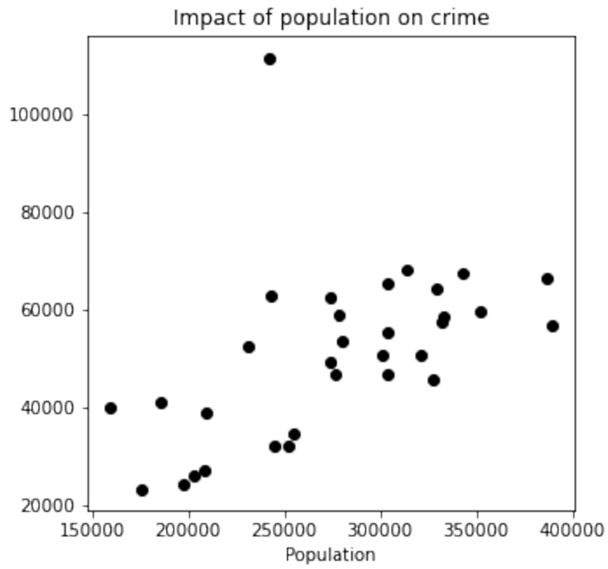


Figure 14: Knife possession in 24 months across London Boroughs

6.1.4 *Impact of population size*

The distribution of crime across the boroughs could simply be a consequence of population size.



There is evidence of a correlation between population and total crime.

To see the impact this has geographically, consider the same visualisations above, using population adjusted crime number i.e. number of offences per 10,000 capital.

Figure 15: Population vs number of crimes

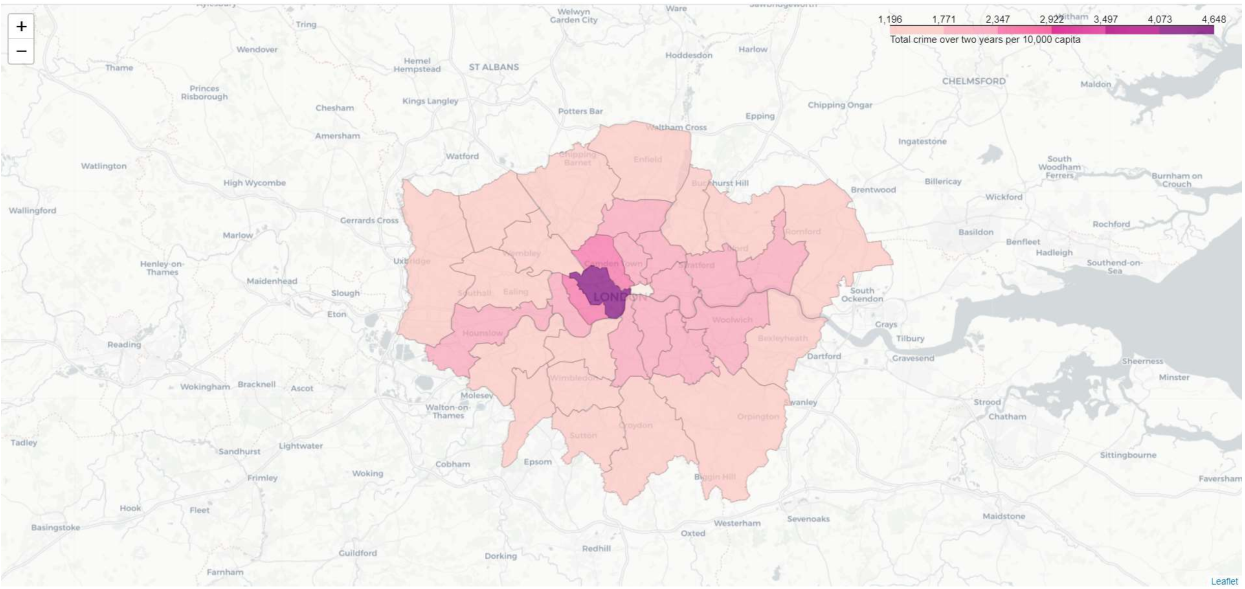


Figure 16: Total crime per 10,000 capita over 24 months

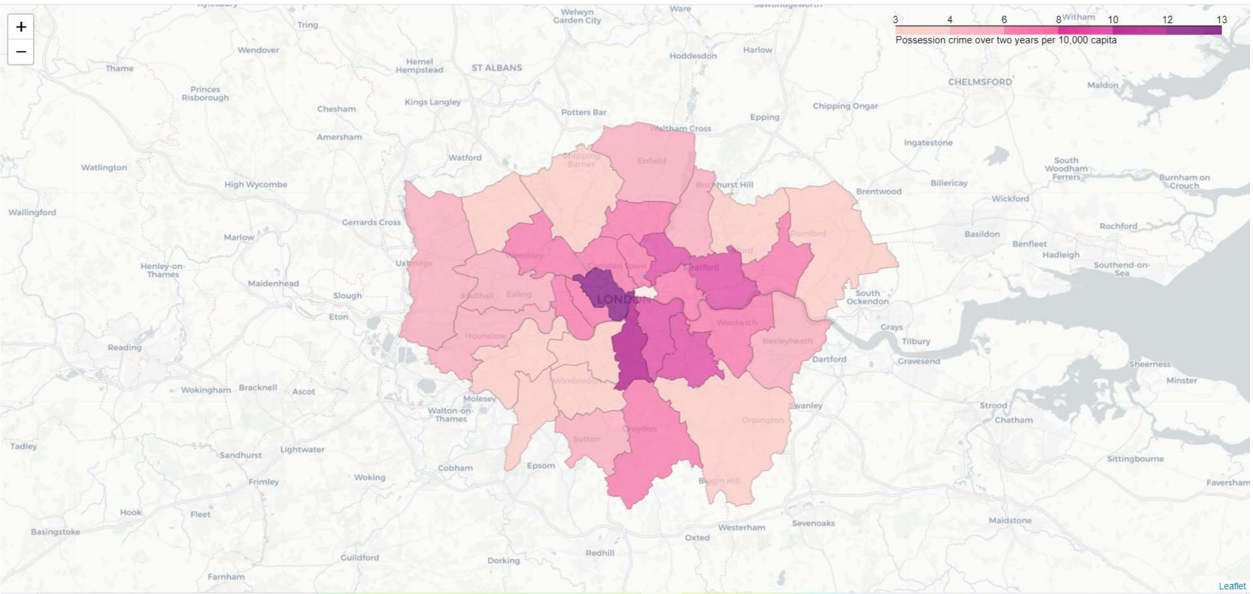


Figure 17: Knife possession per 10,000 capita over 24 months

This adjustment shows a different distribution, and confirms that crime is not uniform across the London boroughs. The highest concentration of Crime is in the London borough of the City of Westminster.

6.1.5 Ward level breakdown

So far, the crime data explored has been at borough level. Data also exists at ward level. This can be visualised to inspect a more granular crime distribution***.***

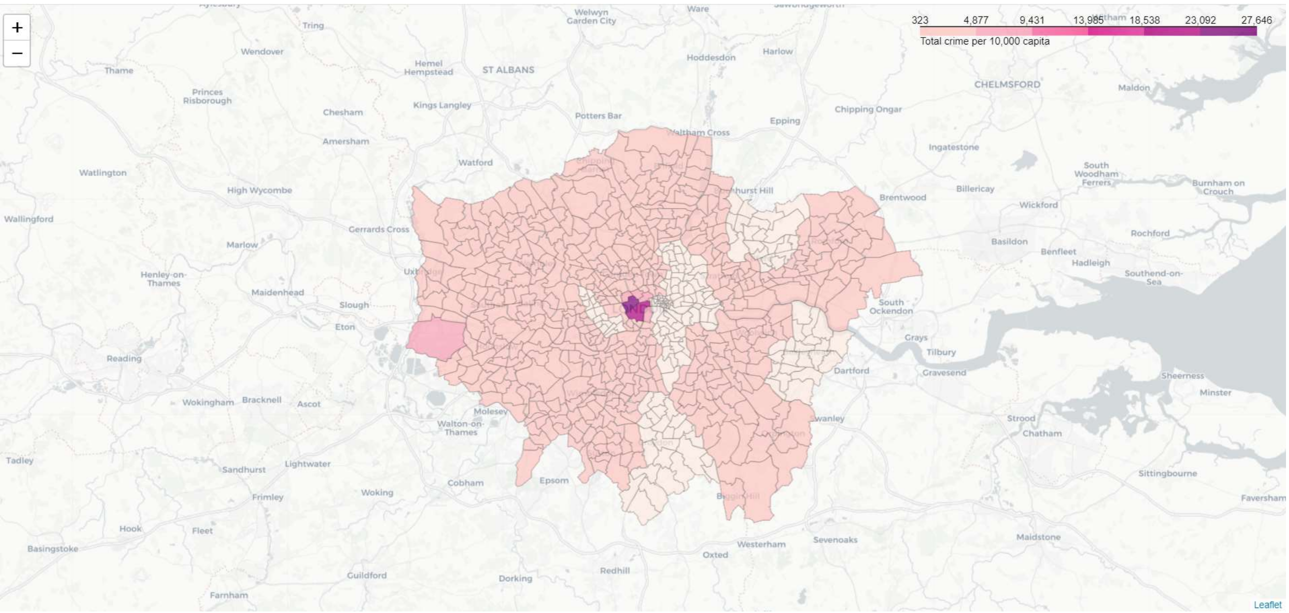


Figure 18: Total crime per 10,000 capita across London wards

In the above, no crime is displayed in a number of wards, this is due to a mismatch in the map overlay which is based on the 2014 ward boundaries, and the crime data which is based on the current boundaries.

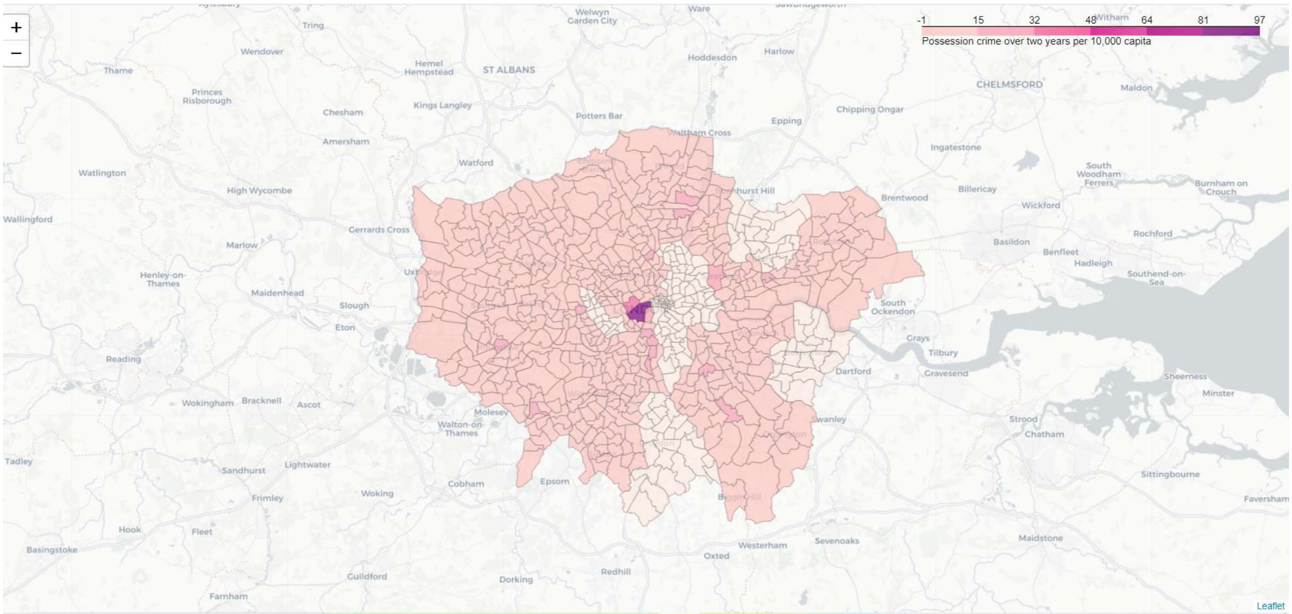


Figure 19: Knife possession per 10,000 capita across London wards

Two wards have significantly higher crime that the rest of London. These wards can be identified as "West End" and "St James's" by sorting the underlying data. These wards are in the London borough of the City of Westminster.

Ward	WardName	Total Crime
E05000649	West End	27377.777778
E05000644	St James's	20419.917012
E05000129	Bloomsbury	7796.581197
E05000138	Holborn and Covent Garden	6796.350365
E05000641	Marylebone High Street	5978.475336
...	...	...
E05000298	Pinner South	695.774648
E05000311	Hacton	655.555556
E05000126	Shortlands	608.866995
E05000407	Coombe Vale	594.029851
E05000307	Cranham	591.304348

Figure 20: Wards sorted by total crime, top and bottom 5

Removing these wards from the data and replotting shows a more nuanced distribution of crime, albeit one limited by the wards which can be plotted.

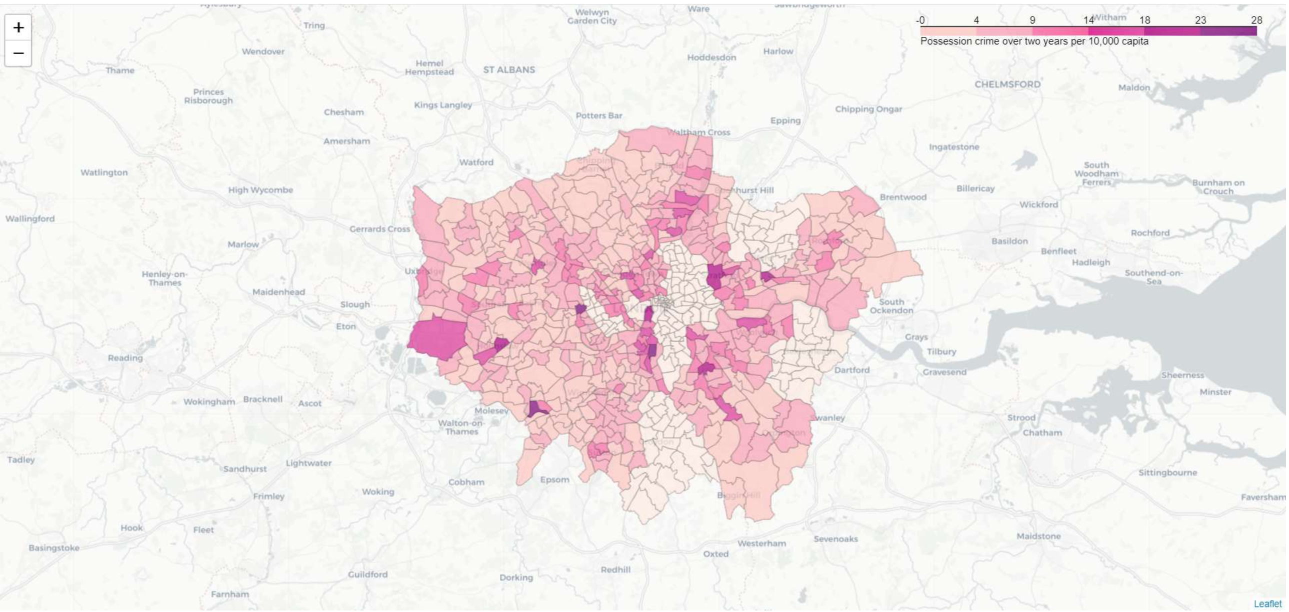


Figure 21: Population adjusted crime by ward, excluding St James and West End

6.1.6 Crime data exploration conclusions

From this exploration the following conclusions have been made:
1. It is appropriate to consider total crime of a period (24 months for current data), to reduce the impact of anomalous events
2. Crime of different categories (both major categories and minor sub-categories) follows different patterns, so this granularity should be considered
3. Crime distribution varies across London
4. There are certain hotspots of crime relative to population density (St James and the West End within the borough of the City of Westminster).

6.1.7 Data transformation:

Data aggregated and pivoted to show total crime for each borough

MinorText / Borough	Absconding from Lawful Custody	Aggravated Vehicle Taking	Aiding Suicide	Arson	Bail Offences	Bicycle Theft	Bigamy	Burglary - Business and Community	Burglary - Residential	Criminal Damage	...	Soliciting for Prostitution	Theft from Person	Theft from a Motor Vehicle	Theft or Taking of a Motor Vehicle	Threat or Possession With Intent to Commit Crimina	Violence with Injury	Violence without Injury	Violent Disorder	Wildlife Crime	Total Crime
Barking and Dagenham	0.0	62.0	0.0	114.0	1.0	350.0	3.0	593.0	1979.0	2588.0	...	0.0	1014.0	2145.0	2140.0	141.0	3988.0	8897.0	5.0	2.0	38797.0
Barnet	0.0	74.0	0.0	114.0	0.0	549.0	0.0	1155.0	4742.0	3745.0	...	0.0	1239.0	7167.0	2644.0	158.0	4472.0	10602.0	4.0	0.0	56945.0
Bexley	0.0	61.0	0.0	123.0	0.0	185.0	0.0	518.0	1734.0	2835.0	...	0.0	279.0	2993.0	1467.0	171.0	3402.0	7120.0	2.0	1.0	31950.0
Brent	7.0	81.0	0.0	134.0	1.0	726.0	5.0	868.0	3527.0	3944.0	...	0.0	1467.0	4931.0	2281.0	169.0	5613.0	12085.0	9.0	0.0	57417.0
Bromley	1.0	62.0	0.0	207.0	0.0	302.0	0.0	892.0	3027.0	3498.0	...	0.0	694.0	4541.0	1925.0	205.0	4108.0	8825.0	2.0	0.0	45711.0

Figure 22: Minor categorisation by borough: Example recent crime data post transformation

MajorText / Borough	Arson and Criminal Damage	Burglary	Drug Offences	Miscellaneous Crimes Against Society	Possession of Weapons	Public Order Offences	Robbery	Sexual Offences	Theft	Vehicle Offences	Violence Against the Person	Total Crime
Barking and Dagenham	2702	2572	2974	662	332	2408	1569	1281	6539	4870	12888	38797
Barnet	3859	5897	2153	743	304	3756	1966	1160	10722	11304	15081	56945
Bexley	2958	2252	1656	497	232	2479	609	766	4664	5310	10527	31950
Brent	4078	4395	4228	779	521	4097	2074	1226	9920	8381	17718	57417
Bromley	3705	3919	2436	677	325	3291	925	1029	8697	7771	12936	45711

Figure 23: Major categorisation by borough: Example recent crime data post transformation

6.2 Industry/Venue Data

6.2.1 Foursquare

As per requirements of the IBM Data Science Professional certificate capstone project, Foursquare API has been used to obtain data about London. Before utilising data, the most reliable method of extracting data for London is explored.

6.2.1.1 Data Obtained by Borough

6.2.1.1.1 Using Central Points

Central coordinates for each borough can be obtained from the geopy library (Nominatim). These locations could be used for obtaining venue data (i.e. by searching for venues within a fixed radius)

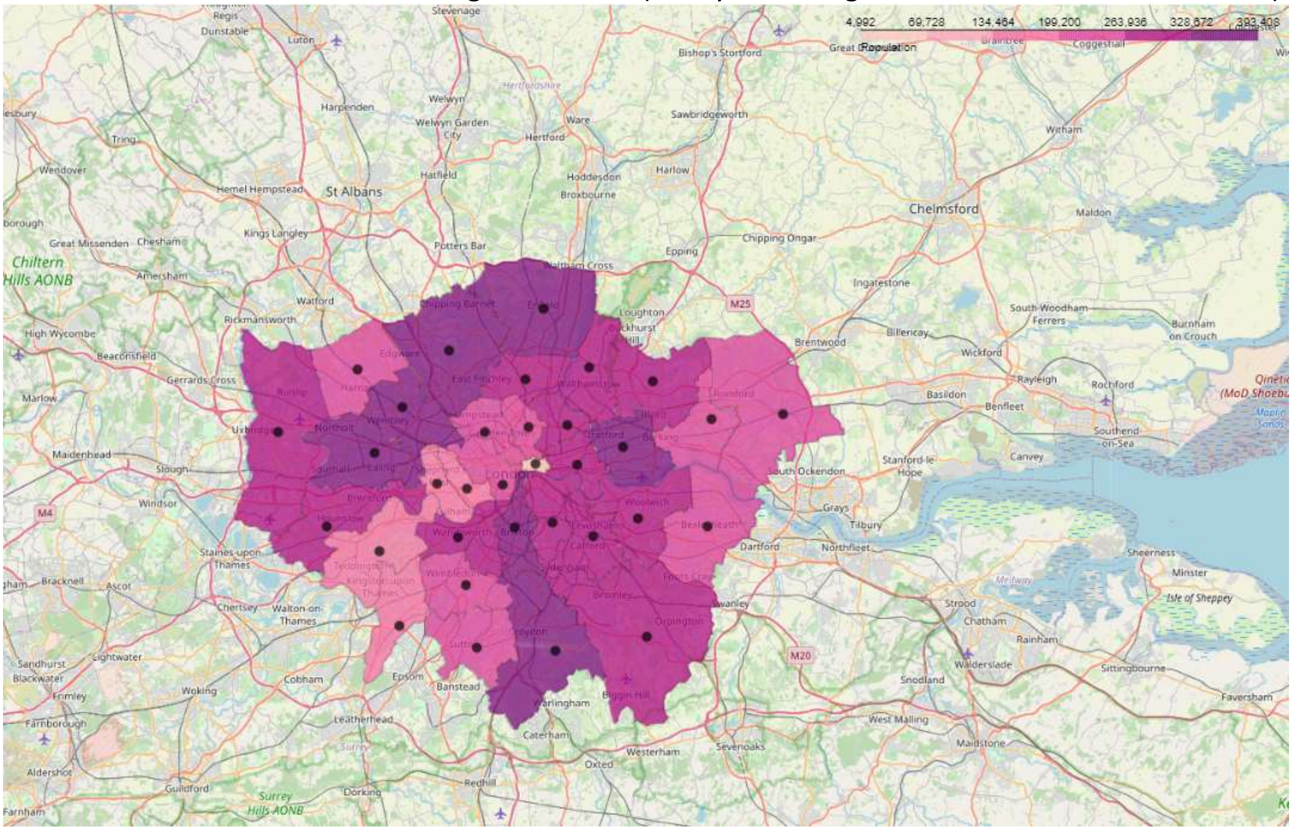


Figure 24: Central points of London boroughs obtained via Nominatim shown against a choropleth of borough population

Inspection of the shape of London boroughs and the location of central points shows the following challenges:

a) The centre of the borough may or may not be close to where venues are located, this is particularly relevant in larger outer London boroughs.
b) The slice of a borough taken by searching a radius around a central point may or may not be representative of the borough.
c) Due to the irregular shapes of London boroughs the centre points can be very close to borough boundaries, so venues returned are likely to be from other boroughs.

d) An appropriate radius of inner London boroughs (which are smaller, and may have venues more densely located) may not be appropriate for outer London boroughs

Conclusion: Given these limitations, the robustness of any analysis on the areas will be compromised, therefore a better approach is required

6.2.1.1.2 Using Borough within Foursquare API

Venue data was retrieved for London boroughs by reference to the borough name.

Visualising the venues retrieved shows a poor match to the well defined borough. Therefore this is not considered a viable approach.

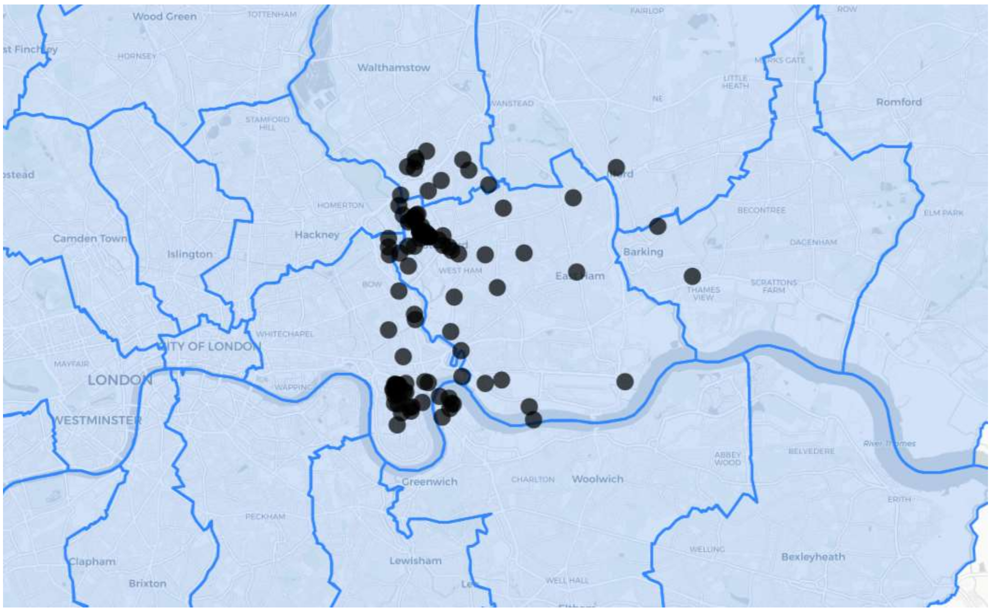


Figure 24: Venues retrieved from Foursquare for "Newham, Greater London, United Kingdom" as location name

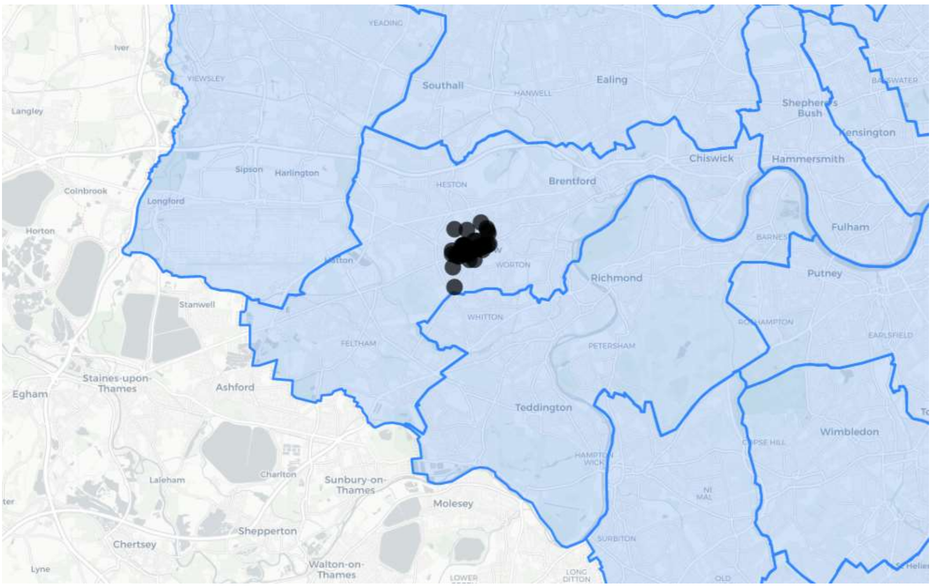


Figure 25: Venues retrieved from Foursquare using "Hounslow, Greater London, United Kingdom" as venue name

6.2.1.2 *Data Obtained by Neighbourhood*

Neighbourhoods in London often refer to subsumed towns and villages, as well as newer developments around geographic features. No formal definition or list of neighbourhoods has been found, but the list taken from Wikipedia is geocodable and coordinates have been obtained using Nominatim.

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref	latitude	longitude
526	Woolwich	Greenwich	LONDON	SE18	020	TQ435795	51.482670	0.062334
527	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655	51.378503	-0.241660
528	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815	51.521393	-0.240554
529	Yeading	Hillingdon	HAYES	UB4	020	TQ115825	51.527239	-0.399270
530	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804	51.512866	-0.474152

Figure 26: Example of Neighbourhoods scrapped from Wikipedia with longitude and latitude added

Not all neighbourhoods are recognised as geocodable by Foursquare:

```
Neighborhood Geocodable
False    188
True     330
```

For neighbourhoods which are not recognised, search by radius can be used, given the differing population density of inner and outer London propose to use a search radius of 600m for inner London (taken as where post town is London) and 1200m for outer London (taken as where post town is not London). It is anticipated this approach will obtain representative venue data for all neighbourhoods. Note: This approach was used for data harvesting but in final analysis only the geocodable neighbourhoods were used.

To sanity check the venue data received, the venues for a neighbourhood with which the author has some familiarity has been inspected and visualised.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Neighborhood Geocodable
15991	Leyton	51.569673	-0.015681	Marmelo Kitchen	51.563913	-0.005926	Restaurant	True
15992	Leyton	51.569673	-0.015681	Yardarm	51.564402	-0.006960	Wine Bar	True
15993	Leyton	51.569673	-0.015681	Leyton Orient Supporters Club	51.559931	-0.013497	Sports Bar	True
15994	Leyton	51.569673	-0.015681	Michael's Fish Bar	51.569407	-0.003578	Fish & Chips Shop	True
15995	Leyton	51.569673	-0.015681	Deeney's Café	51.562088	-0.010269	Café	True
...	...	...	...	...	...	...	...	...
16056	Leyton	51.569673	-0.015681	TK Maxx	51.556548	-0.007183	Clothing Store	True
16057	Leyton	51.569673	-0.015681	Costa Coffee	51.555784	-0.008333	Coffee Shop	True
16058	Leyton	51.569673	-0.015681	Leyton Mills Retail Park	51.556145	-0.008739	Shopping Plaza	True
16059	Leyton	51.569673	-0.015681	Asda	51.555901	-0.009369	Supermarket	True
16060	Leyton	51.569673	-0.015681	Burger King	51.555484	-0.008305	Fast Food Restaurant	True

Figure 27: Venue data obtained for Leyton, Greater London, United Kingdom

Visualisation of Leyton

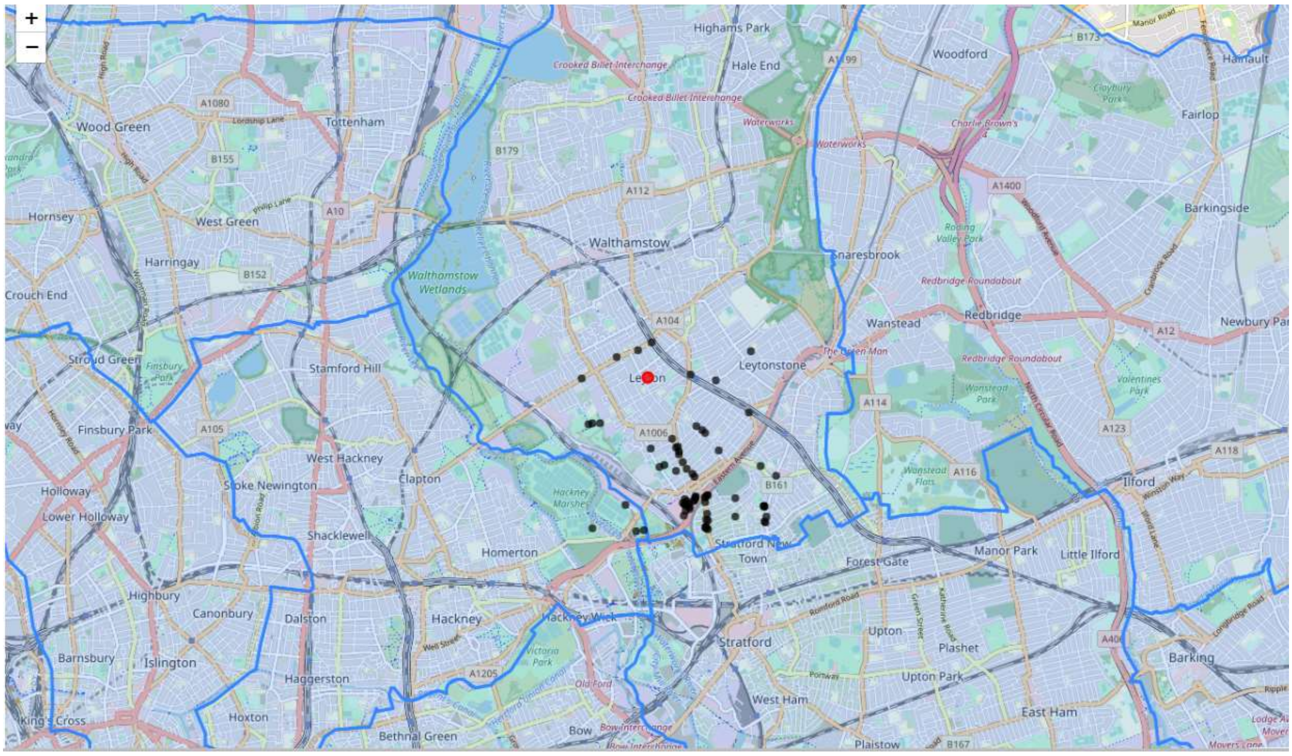


Figure 28: Visualisation of venues retrieved for Leyton

Applying existing knowledge, it is confirmed that this location is roughly Leyton, though a scatting of venues referenced would normally be considered to belong to neighbouring neighbourhoods (Leytonstone and Homerton). The coordinates obtained from Nominatim are marked in red, and it is noted that searching in a radius from this point would have obtained less accurate results.

6.2.1.3 Data Exploration Conclusion

The most meaningful data obtained from foursquare is at London neighbourhood level when searching on neighbourhood name, this data set is however not comprehensive.

6.2.1.4 Data Transformation

Data for each neighbourhood is pivoted and normalised, to show what percentage of a neighbourhoods venues is of each category.

	Neighbourhood	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Windmill	Wine Bar	Wine Shop	Winery	Wings Joint	Women's Store	Xinjiang Restaurant	Yoga Studio	Zoo	Zoo Exhibit
0	Abbey Wood	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	Acton	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
2	Addington	0.0	0.0	0.000000	0.027778	0.027778	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	Aldgate	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.020000	0.020000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	Aldwych	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.016667	...	0.0	0.016667	0.016667	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
5	Anerley	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
6	Angel	0.0	0.0	0.015385	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.015385	0.0	0.0	0.0	0.0	0.015385	0.0	0.0
7	Arkley	0.0	0.0	0.025641	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
8	Balham	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.024390	0.0	0.0
9	Barbican	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

Figure 29: Example transformed neighbourhood venue data as obtained from Foursquare

6.2.2 24 Hour Licenced Venues

Data is provided for each borough broken down by venue type. For this analysis only venue types with sufficient data for meaningful analysis have been considered. Venues types with 0 occurences in 8 or more boroughs were excluded. The remaining venue types are as follows:

Total	~~Pubs, bars and nightclubs~~	
	Total Supermarkets and stores	Large supermarkets
		Other convenience stores
		~~Supermarket and store type not reported~~
	Total Hotel bars	~~Open 24 hours to residents and general public~~
		Open 24 hours to residents and their guests only
		~~Hotel bar type not reported~~
	~~Other premises types~~	

The number of total venues with a 24 hour alcohol license has been plotted for each borough.

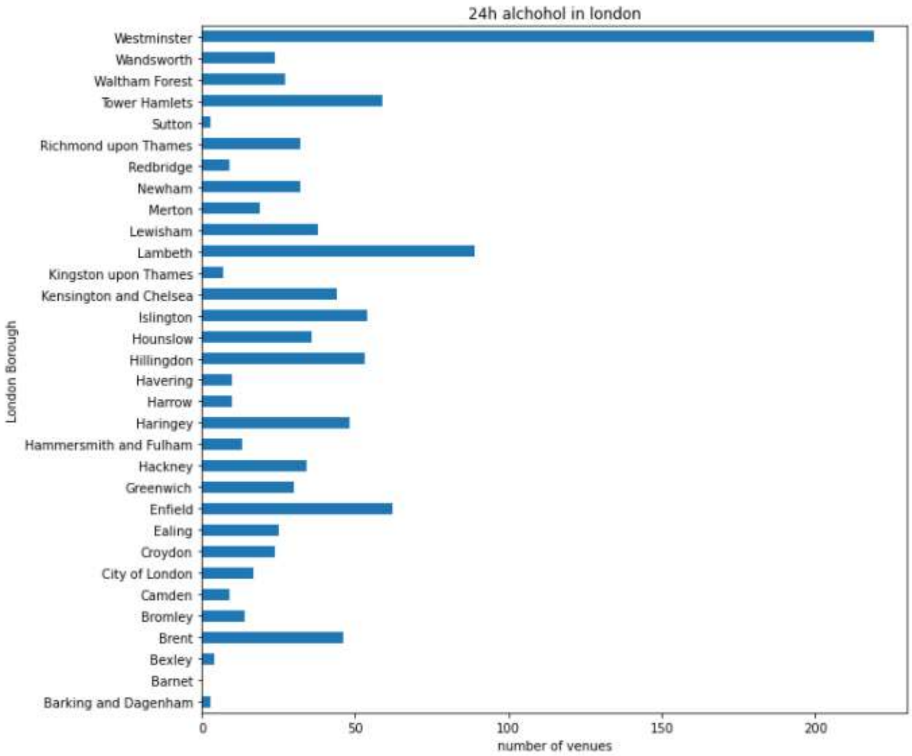


Figure 30: Total number of 24 hour licensed venues per Borough

This shows Westminster is an outlier, which fits with knowledge this is a hospitality centre. It is also noted that some other London boroughs such as Sutton, Bexley, Barnet and Barking and Dagenham have via limited numbers.

Next consider the venues, population adjusted i.e. number of venues per 10,000 capita.

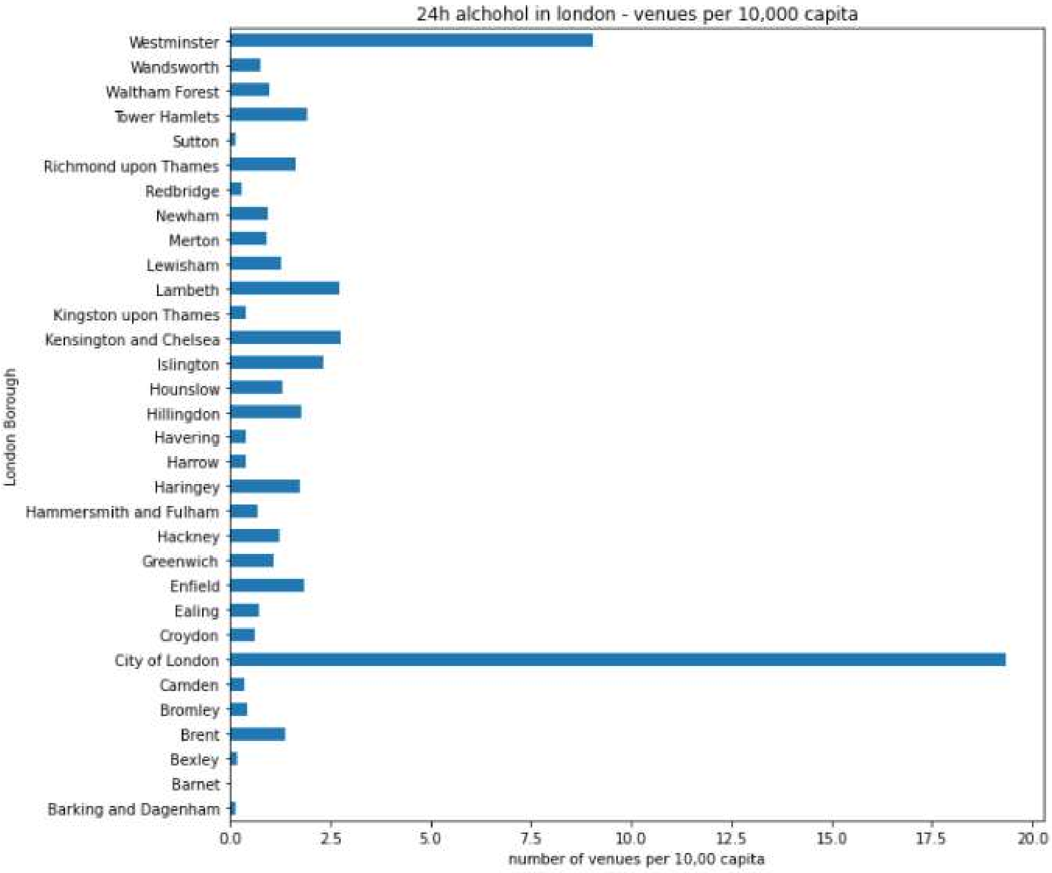


Figure 31: Total number of 24 hour venues per Borough population adjusted

This shows the commercial and less residential boroughs of Westminster and city of London are outliers. It is noted that analysis may be more meaningful with these excluded.

6.2.2.1 Data transformation:

In line with crime analysis consider the total number of venues per 10,000 capita.

Borough	Total	Total Supermarkets and stores	Large supermarkets	Other convenience stores	Total Hotel bars	Open 24 hours to residents and their guests only
Barking and Dagenham	0.143541	0.143541	0.047847	0.095694	0.000000	0.000000
Barnet	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Bexley	0.163733	0.163733	0.040933	0.122800	0.000000	0.000000
Brent	1.385125	0.843120	0.120446	0.722674	0.090334	0.000000
Bromley	0.426959	0.335468	0.121988	0.213480	0.091491	0.091491

Figure 32: Example transformed venues with 24 hour licence data

6.2.3 Local Units by Industry

Consider a simple visualisation of where a single industry is distributed across London, in this case "Finance and insurance"

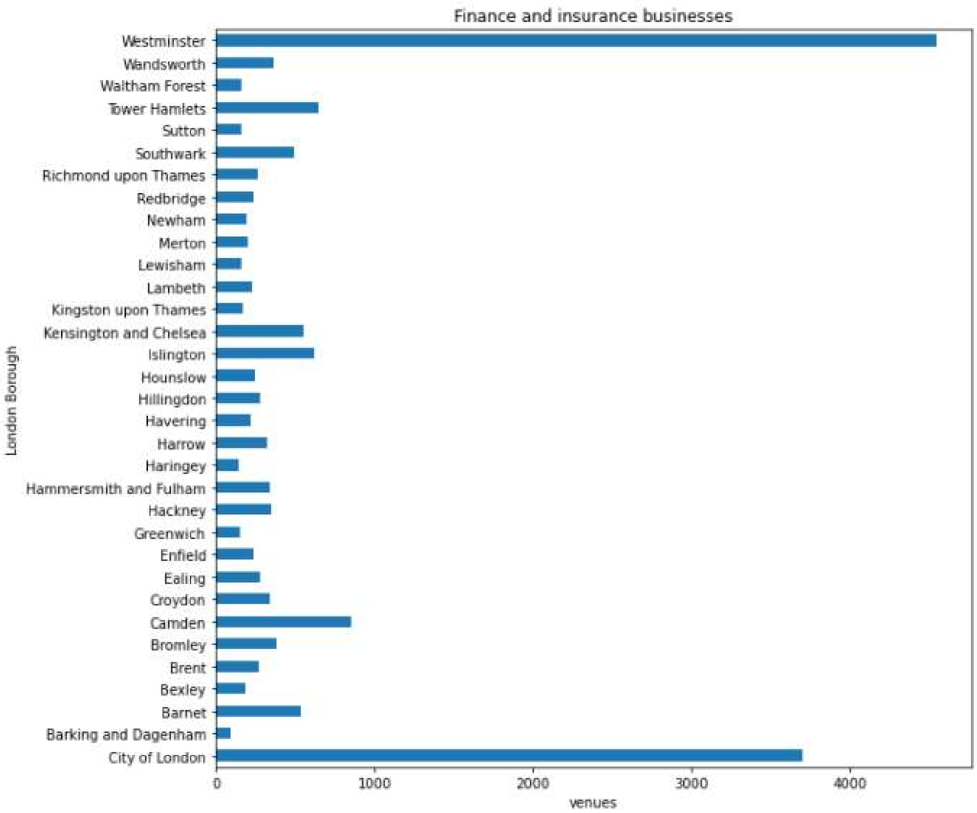


Figure 33: Distribution of financial services

This shows most finance and insurance is in the city of London or Westminster.

Next look at the total number of business units in each borough. This has been spit by industry type to give a sense of the diversity of local business units.

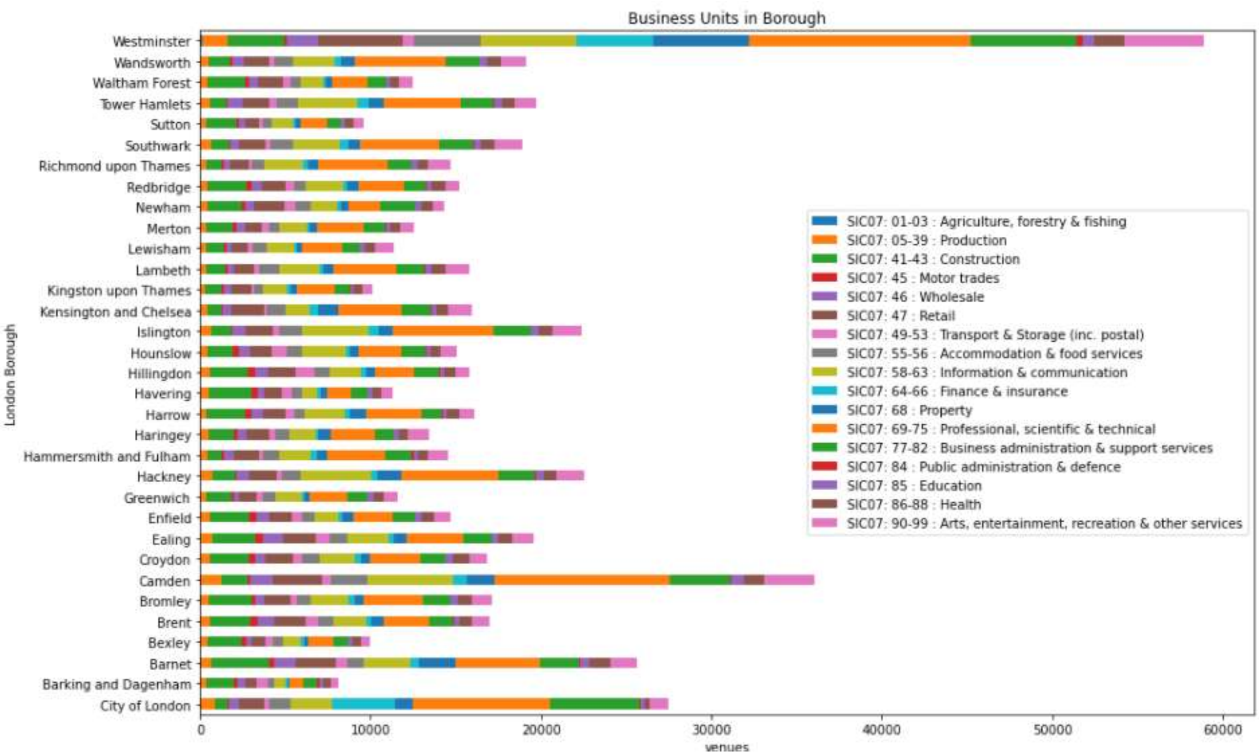


Figure 34: Industry totals per London Borough

Population adjusted this show's the city of London to be an outlier for number of businesses.

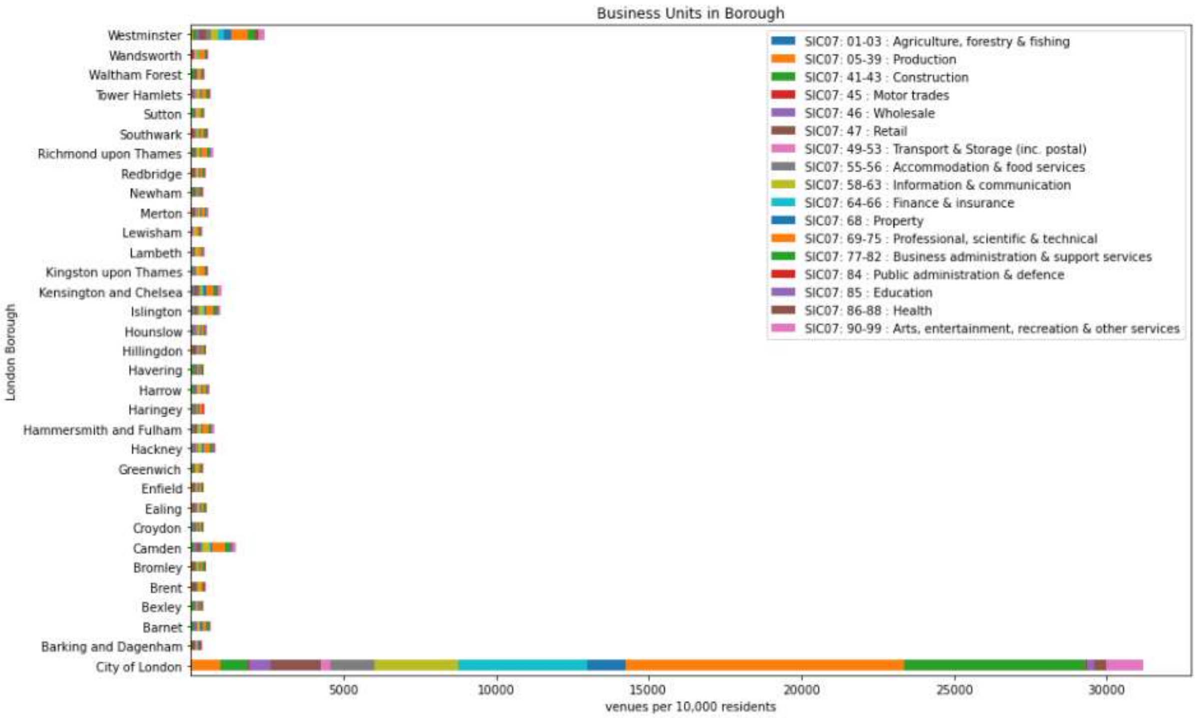


Figure 35: Business units per 10,000 capita across London boroughs

Excluding the city of London provides more meaningful visualisation for the remaining boroughs

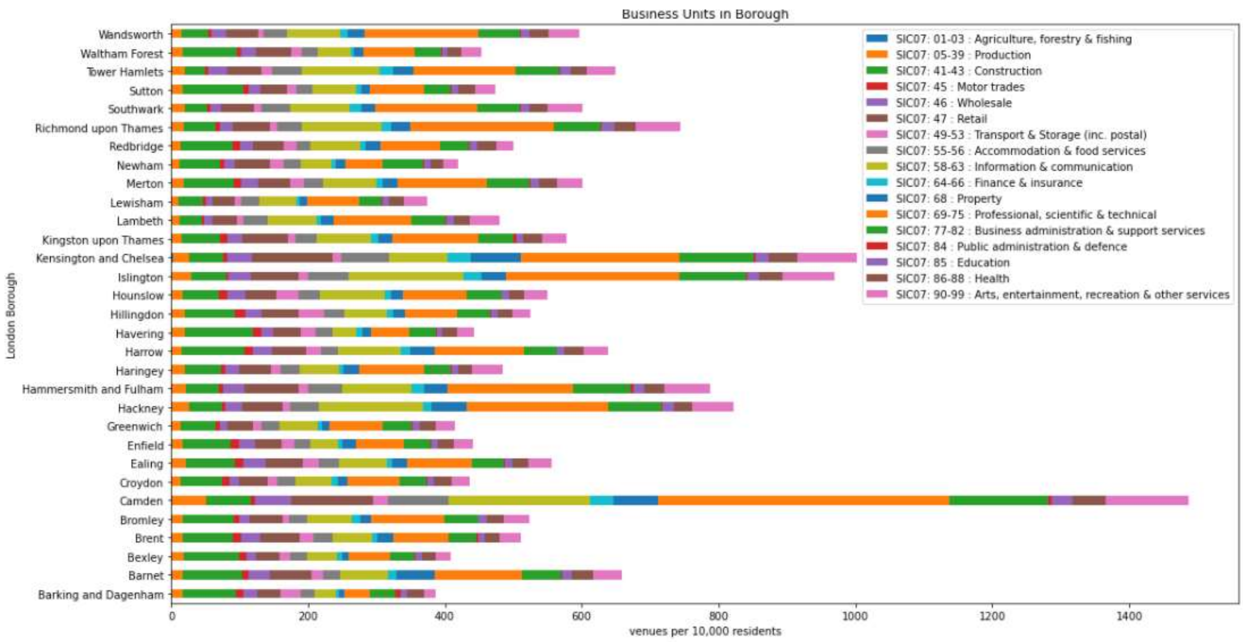


Figure 36: Business units per 10,00 capita across Greater London (excluding City)

To understand the make up of each borough, consider the percentage of total industry each industry type represents. The variation in size of different colour bands shows the variation in proportion of different industries within the boroughs. For example Havering has more construction businesses and less finance than Wandsworth.

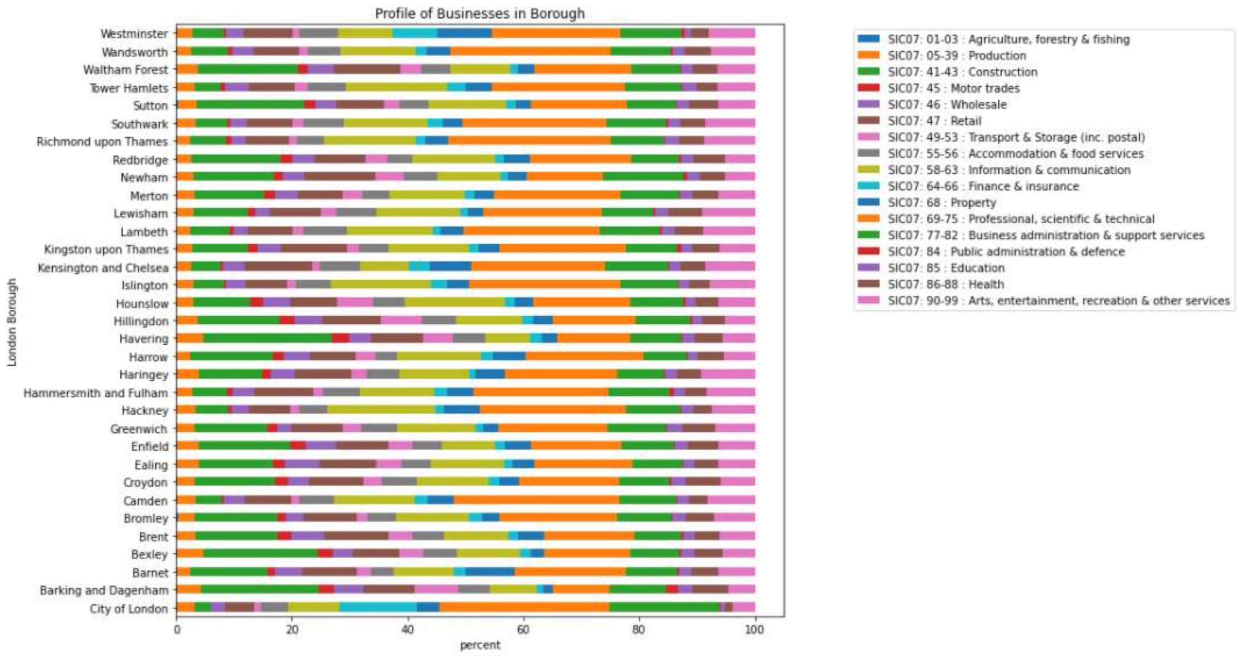


Figure 37: Industry profile of each London Borough

In conclusion there is variation in number and proportion of industry type of business units across the London boroughs.

6.2.4 Data transformation:

Create population adjusted data

Borough	SIC07: 01-03 : Agriculture, forestry & fishing	SIC07: 05-39 : Production	SIC07: 41-43 : Construction	SIC07: 45 : Motor trades	SIC07: 46 : Wholesale	SIC07: 47 : Retail	SIC07: 49-53 : Transport & Storage (inc. postal)	SIC07: 55-56 : Accommodation & food services	SIC07: 58-63 : Information & communication	SIC07: 64-66 : Finance & insurance	SIC07: 68 : Property	SIC07: 69-75 : Professional, scientific & technical	SIC07: 77-82 : Business administration & support services	SIC07: 84 : Public administration & defence	SIC07: 85 : Education	SIC07: 86-88 : Health	SIC07: 90-99 : Arts, entertainment, recreation & other services	SIC07: Total
City of London	28.409091	948.863636	892.045455	34.090909	727.272727	1619.318182	329.545455	1454.545455	2744.318182	4204.545455	1244.318182	9136.363636	5937.500000	62.500000	244.318182	380.681818	1215.909091	31204.545455
Barking and Dagenham	0.478469	16.028708	78.708134	10.765550	19.377990	33.971292	29.665072	20.574163	31.818182	4.306220	6.698565	37.799043	37.799043	8.133971	9.330144	23.684211	18.181818	387.320574
Barnet	0.898357	15.657084	86.883984	9.240246	30.929158	61.601643	16.298768	25.667351	69.045175	13.732033	55.056468	127.181725	58.264887	1.411704	13.860370	31.570842	41.324435	658.624230
Bexley	0.409333	18.419975	81.047892	11.051985	13.507982	33.360622	16.577978	23.741302	44.617274	7.572657	9.209988	60.990585	34.997953	1.023332	9.414654	20.057307	22.717970	408.718788
Brent	0.150557	17.314062	72.267389	12.195122	28.756399	57.061126	20.776874	28.154170	57.211683	8.130081	23.336344	79.795242	41.403192	1.957242	9.485095	22.433002	31.165312	511.592894

Figure 38: Example population adjusted local industry units data

7. Methodology

Following data exploration (see section 6) the following approaches will be taken for the different data types:

7.1 24 Hour Licenced Venues

- As this data is from 2018, the 2018 crime data will be used.
- Crimes over 2018 will be aggregated
- Crimes and venues will be population adjusted
- Only crimes and venue types with adequate data will be considered
- Pearson correlation coefficients will be calculated to identify potential pairwise correlation between venue types and crimes.
- Scatter plots, with lines of best fit if appropriate will be plotted to explore relationships between 24 hour licenced venues and crime

7.2 Local Units by Industry

- Crimes over the most recent 24 months will be aggregated
- Crimes and industry units will be population adjusted
- Only crimes with adequate data will be considered
- Pearson correlation coefficients will be calculated to identify potential pairwise correlation between venue types and crimes.
- Scatter plots, with lines of best fit if appropriate will be plotted to explore relationships between 24 hour licenced venues and crime

7.3 Foursquare recommended venues

- Foursquare data will be used to form clusters of neighbourhoods based on similar venue profiles. Venue profiles will be determined using normalisation of number of venues of given category in a geography.
- K-means clustering will be used to identify clusters of similar venue profiles.
- Map based visualisations will be used to enable manual inspection for correlations. i.e. cluster labels superimposed on a choropleth map of crime data.

8. Results

8.1 24 Hour Licenced Venues

Initial inspection indicated a number of crimes had correlation with number of 24 hour venues. To look for broad linear relationships Pearson pairwise coefficients were calculated for all combinations of crime categorisation and venue

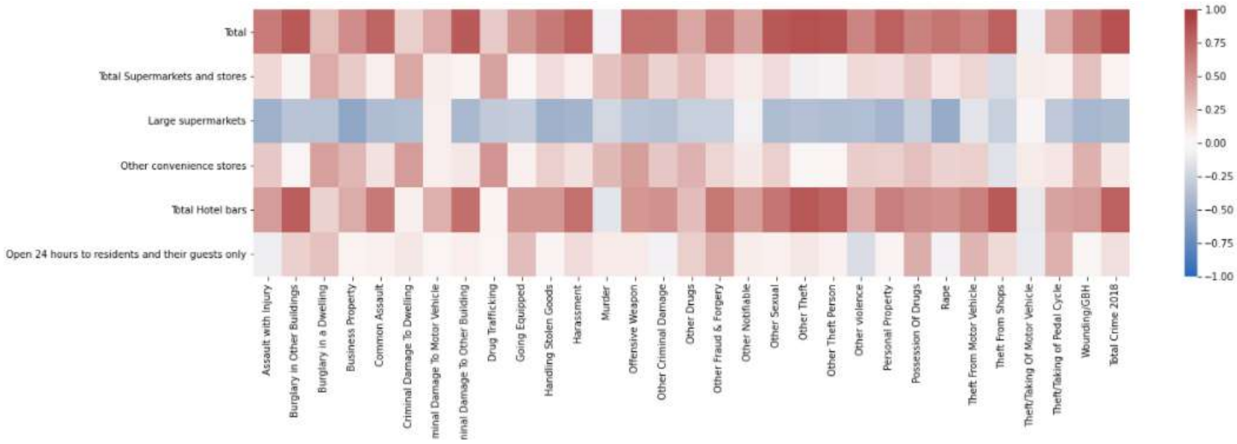
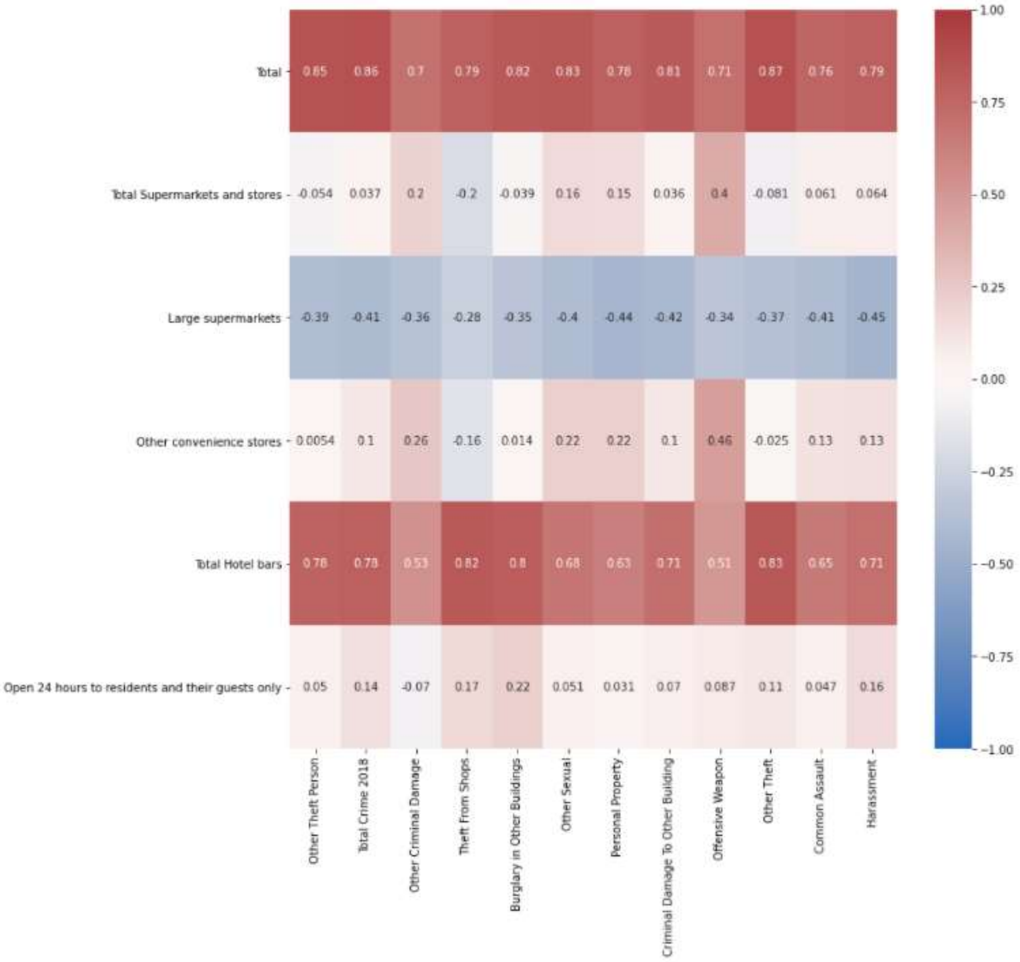


Figure 39: Correlation matrix of minor crime categorisations and 24 licensed venues

This indicates that some but not all crimes have a linear correlation with certain types of licenced venues. Of particular note there is no suggestion 24 hour licenced supermarkets increase crime, and indeed some suggestion they reduced crime.



The crimes with strong evidence of correlation (absolute value Pearson coefficient of 0.7 or more for one or more venue type) were considered for more in depth analysis

Figure 40: Heat-map showing Pearson correlation coefficients of crimes with highest linear correlation to 24 hour alcohol venues

To dig a bit deeper, some of these have been plotted.

These show positive

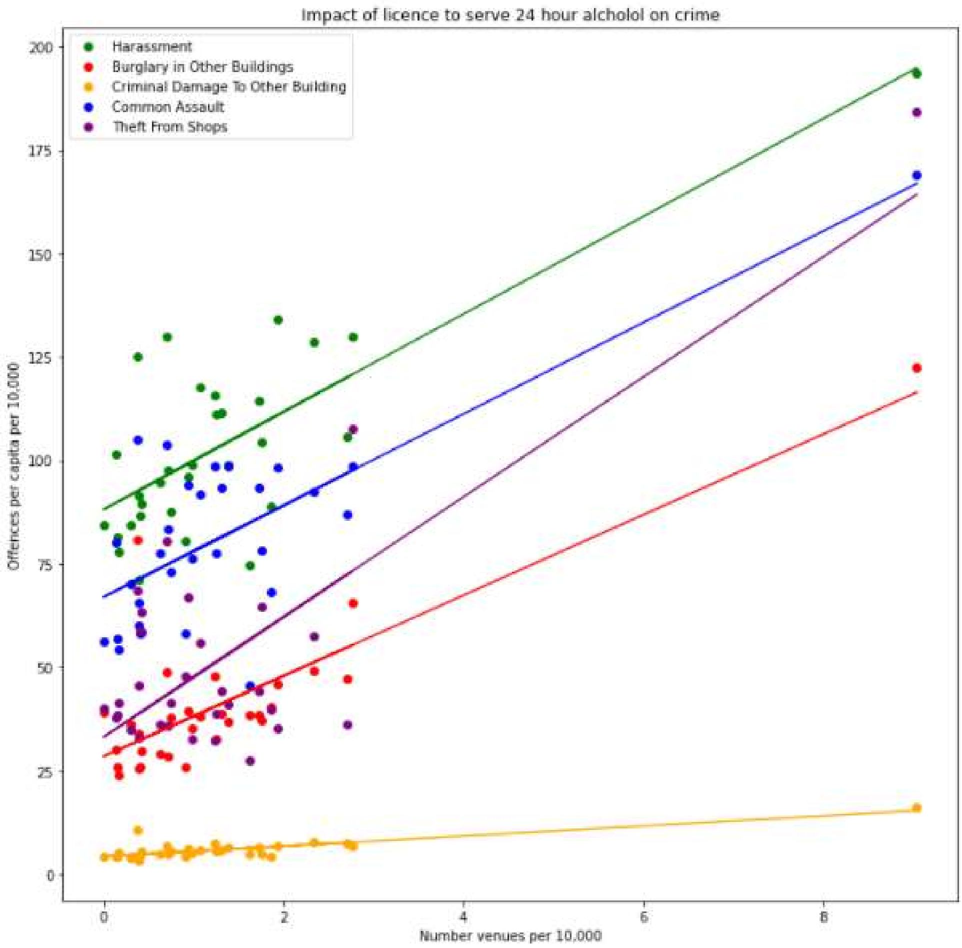


Figure 41: Impact of number of 24 hour licensed venues on crime

correlation, with each crime having main cluster and a single extreme outlier. However removing the outlier of Westminster the correlation looks less clear.

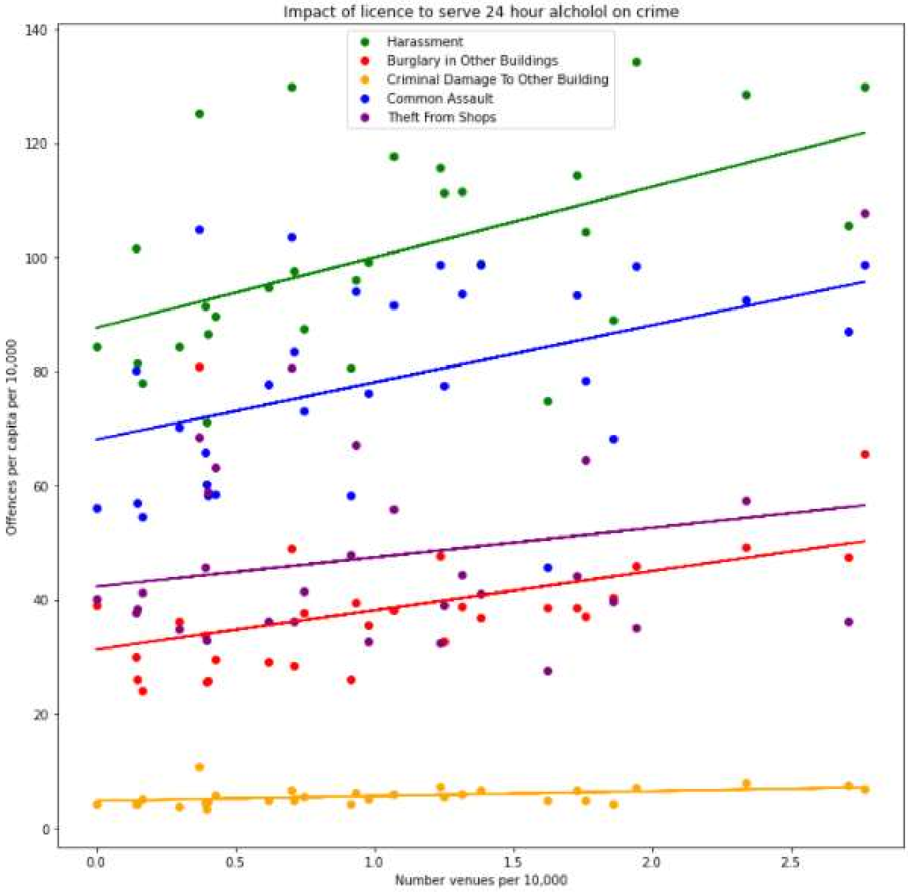


Figure 42: Impact of number of 24 hour licensed venues on crime excluding Westminster

Indeed repeating the correlation analysis, excluding Westminster there is less strong indication of correlation, particularly considering the small sample size.

It is noted that large supermarkets have been consistently indicated to have a negative correlation to crime. An example of this is visualised, with the more positive correlation for hotel bars included for comparison.

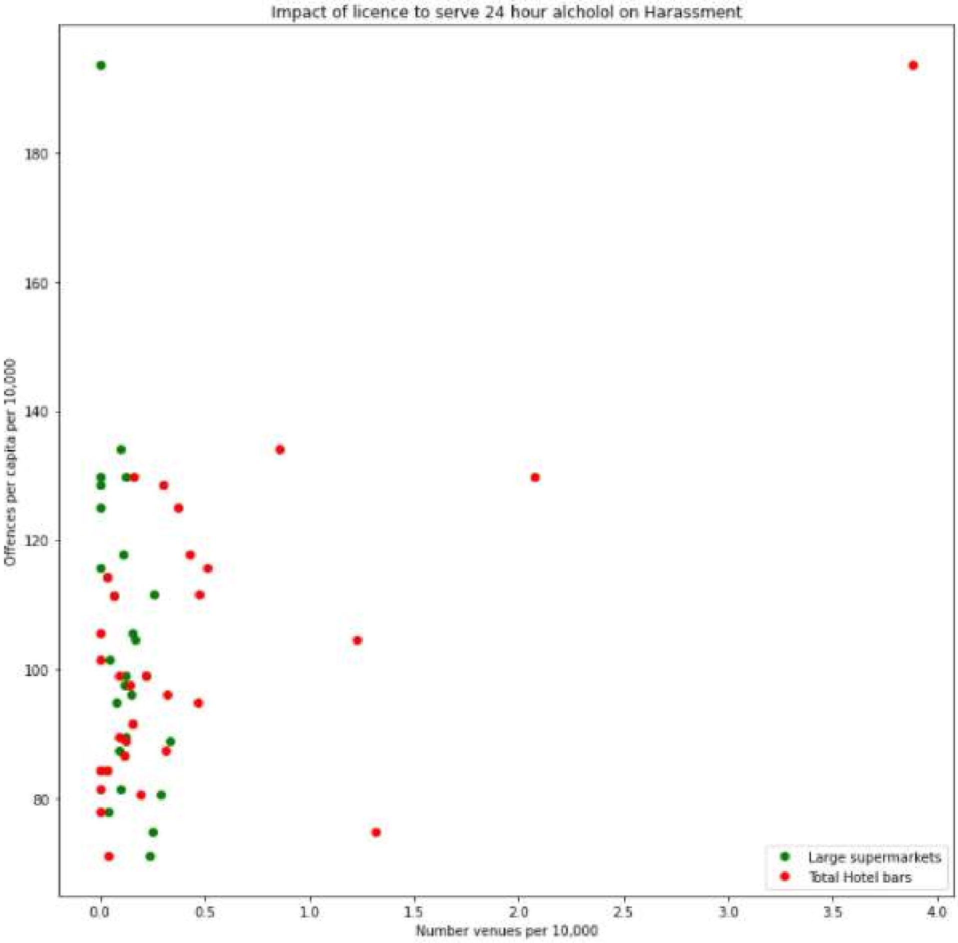


Figure 43: Negative trend harassment vs large supermarkets compared with hotel bars

Although this indicates a slight negative trend, it is noted that there are not large numbers of 24 hour licenced venues and no great variation across the boroughs making the results hard to interpret.

8.2 Local Units by Industry

Evidence was found for correlation between some types of crime and some types of industries.

1. The correlation between crime and number of units, was less pronounced for most industry types when data was population adjusted.

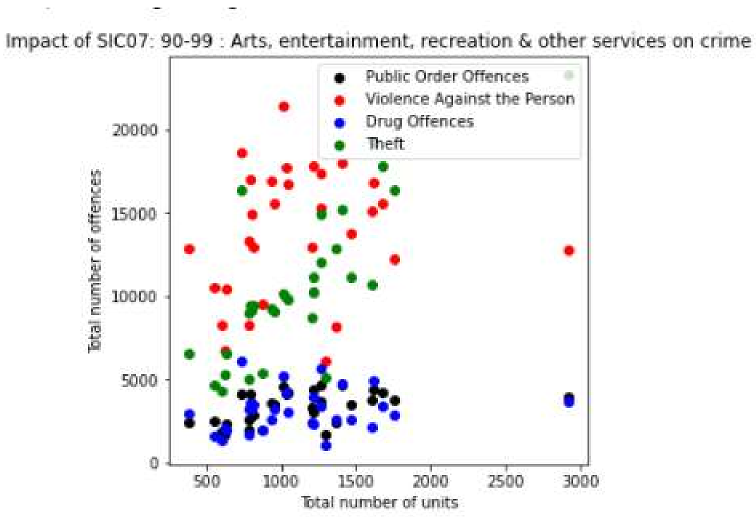


Figure 44: Impact of unit of SIC07 90-99 on four major crime categories

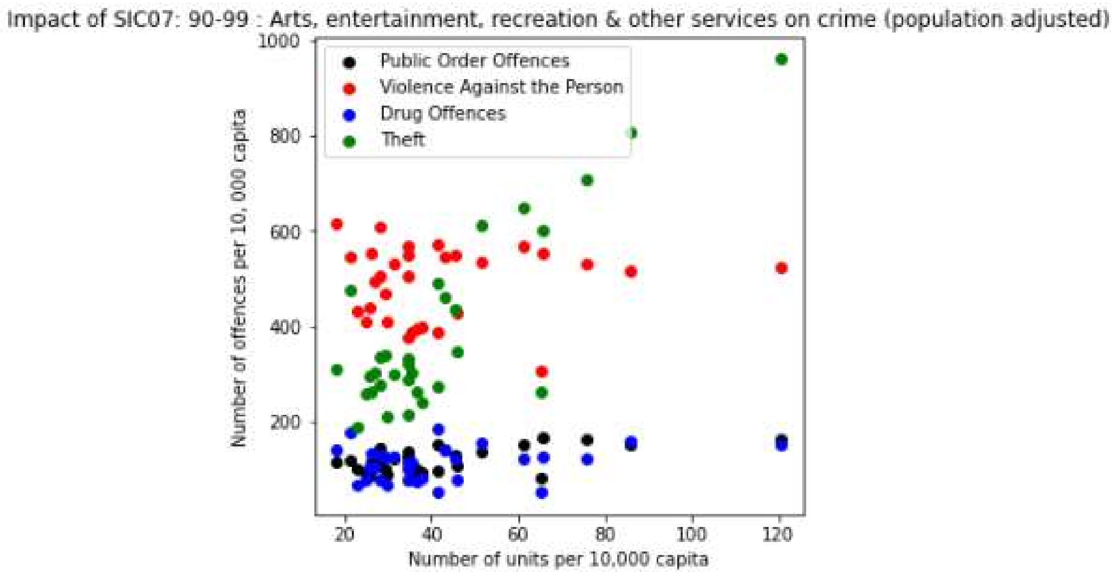


Figure 45: Impact of unit of SIC07 90-99 on four major crime categories (population adjusted)

2. Westminster was an outlier in the data and when removed, trends changed.

Consider the lines of best fit for “obscene publications” across different industries.

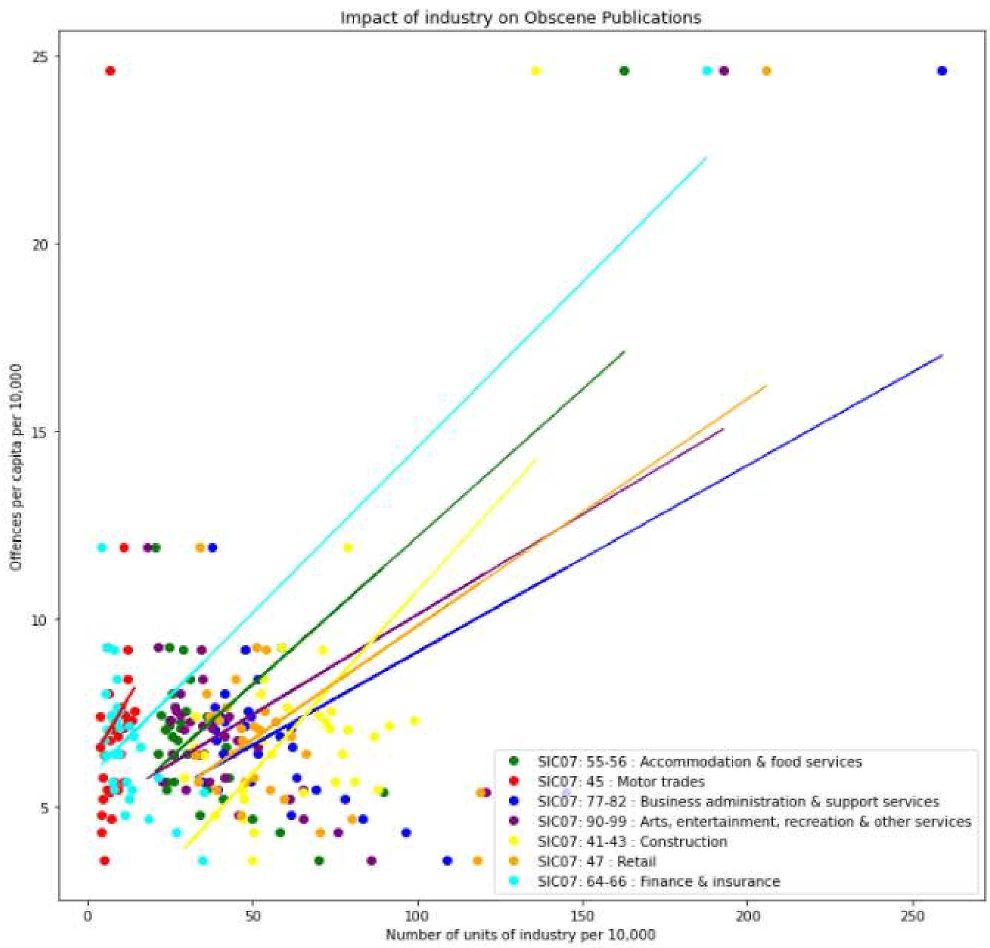
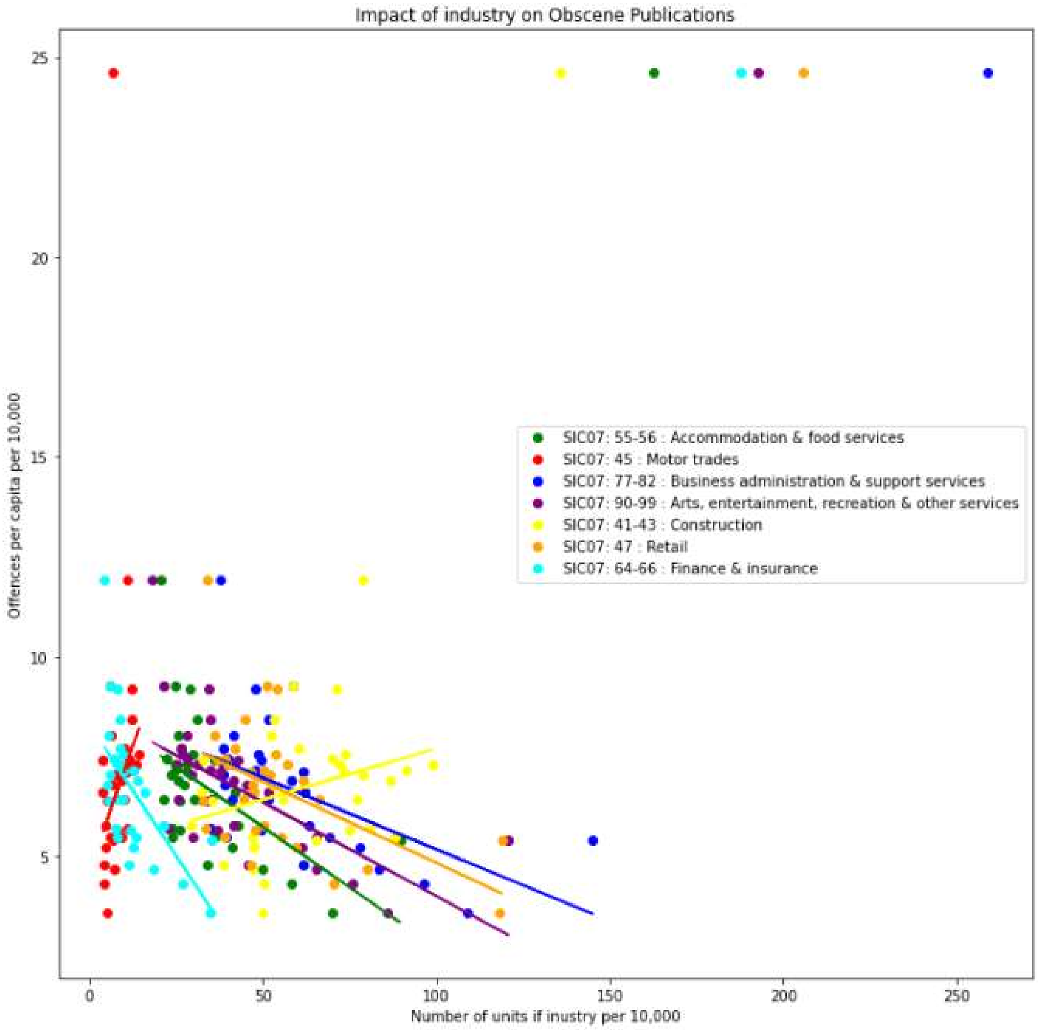


Figure 46: Best fit of obscene publication across multiple industries

This shows positive correlations, though “best fit” is clearly not a “good fit”. The following shows the same data but the outlier (which is the city of Westminster) excluded from the best fit calculation.



Here "best fit" is a

***F*igure 47: Best fit for obscene publication across multiple industries excluding Westminster data**

negative correlation for all industries plotted except construction and motor trades.

3 The minor crimes with the strongest evidence of correlation are:

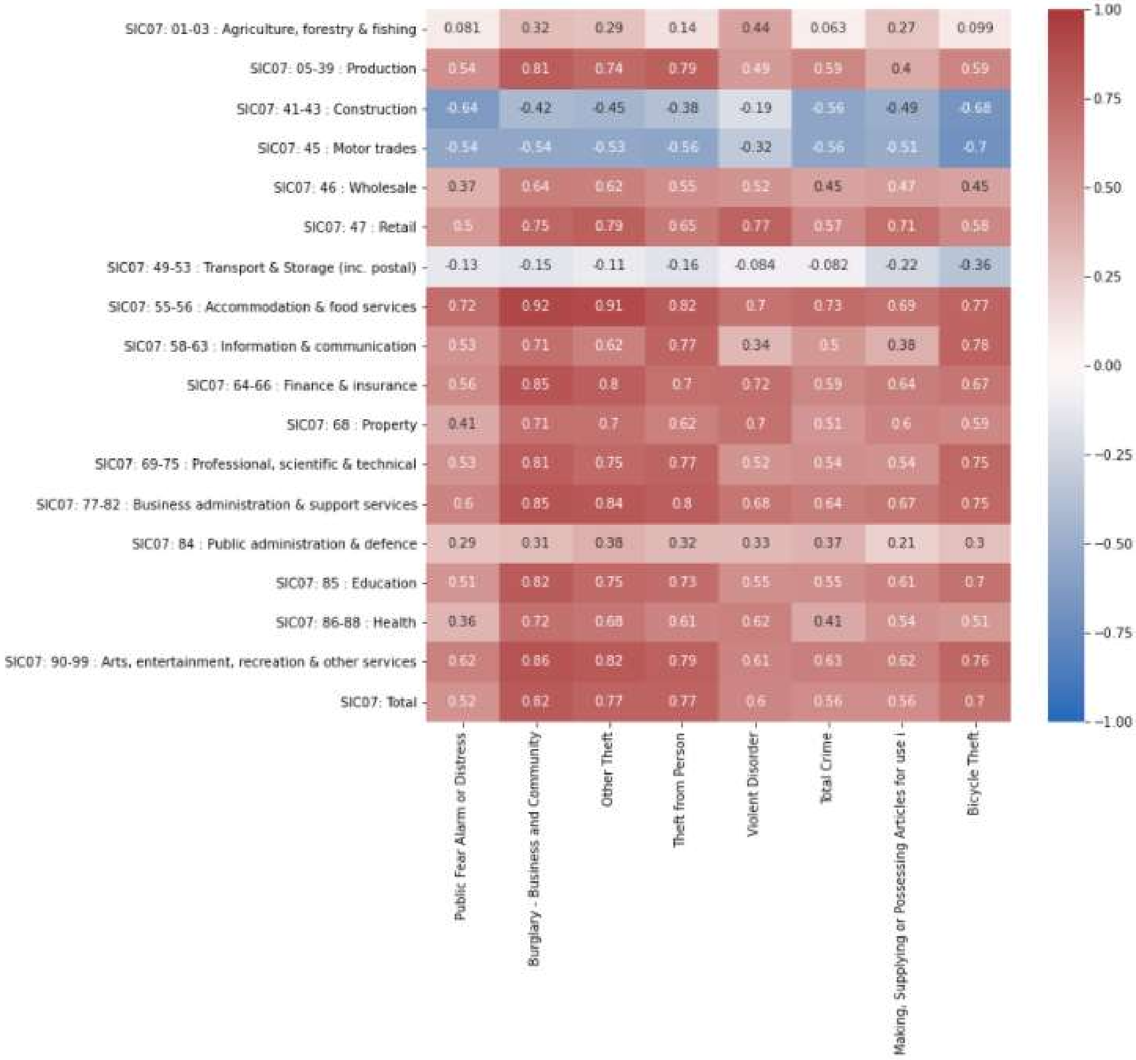


Figure 48: Crimes with most evidence of correlation to industry units (Westminster excluded)

A selection of these relationships have been plotted to enable further examination.

Bicycle theft has the strongest evidence of negative correlation. To examine this consider the following plot:

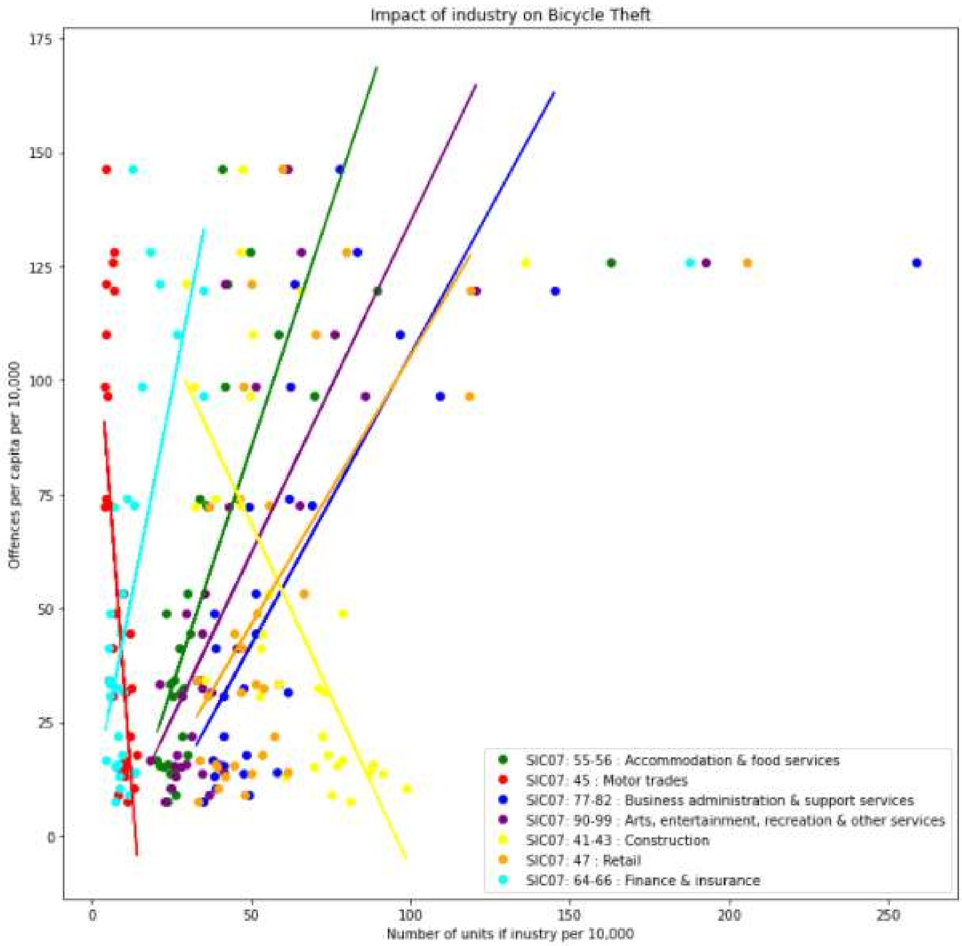


Figure 49: Best fit of bicycle theft across various industries

The negative correlation is hard to see, so considered the following reduced industry plot

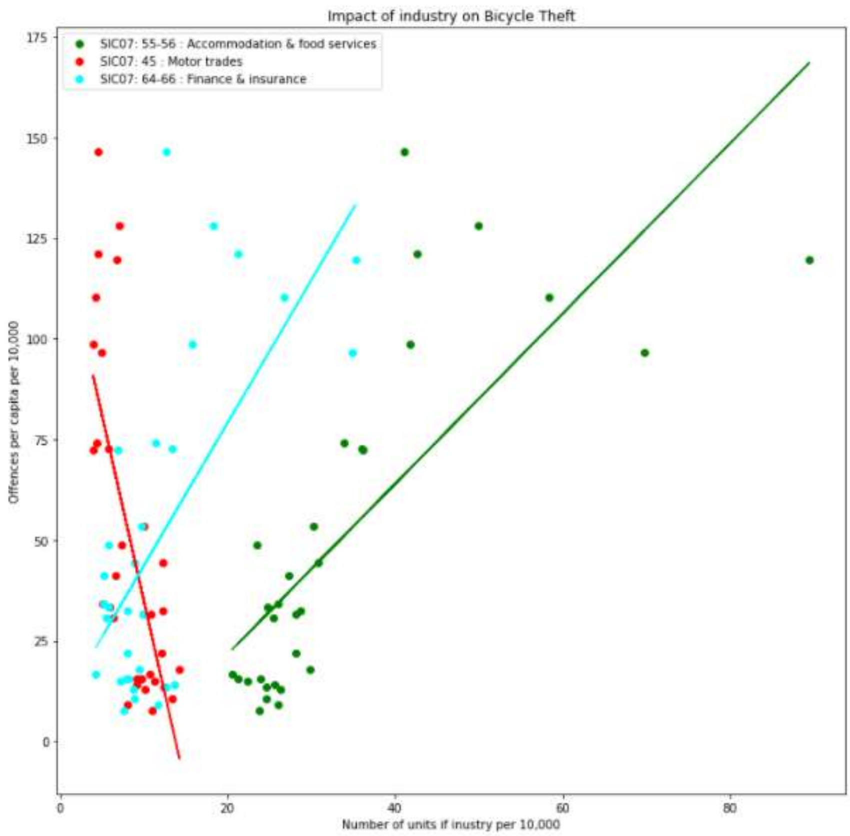


Figure 50: Best fit of bicycle theft across three industries

Business and Community Burglary is of particular interest to the stakeholders of this report, due to the huge cost of business burglary each year.

This suggests

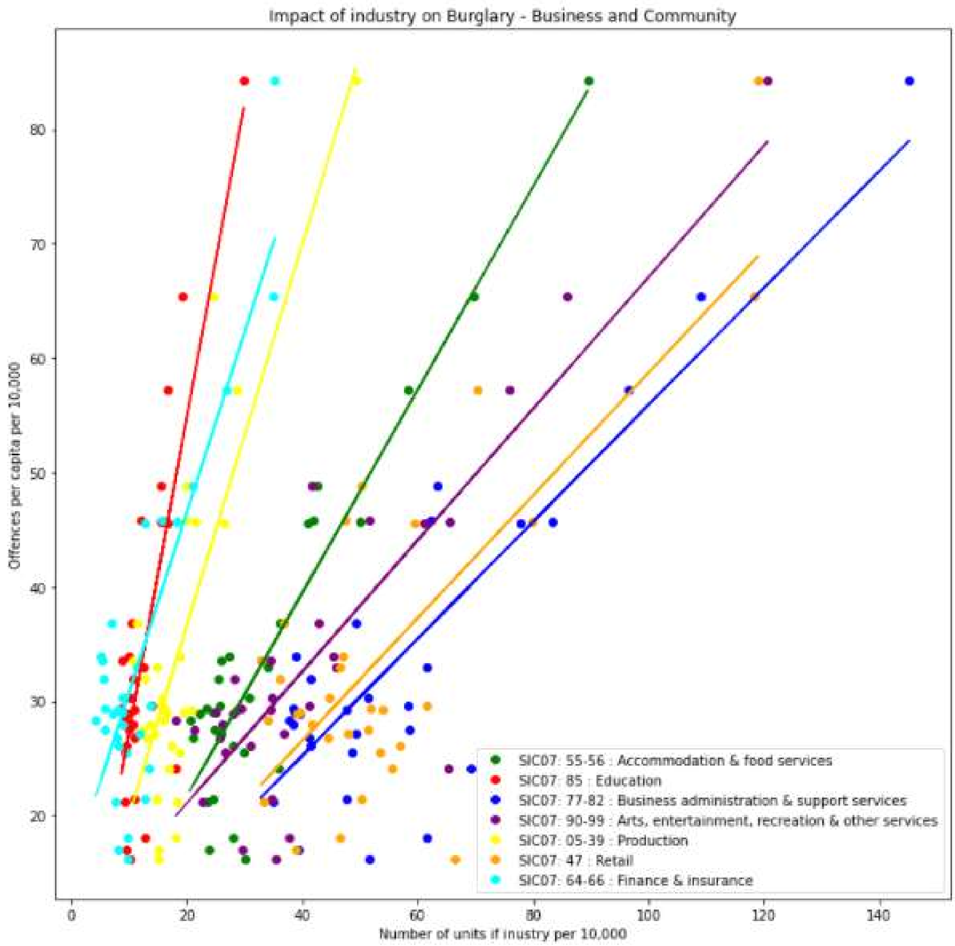


Figure 51: Best fit of Burglary business or community

Education and Production, having the steepest gradients have the most impact on Burglary – Business and Community.

Next consider “violent disorder”

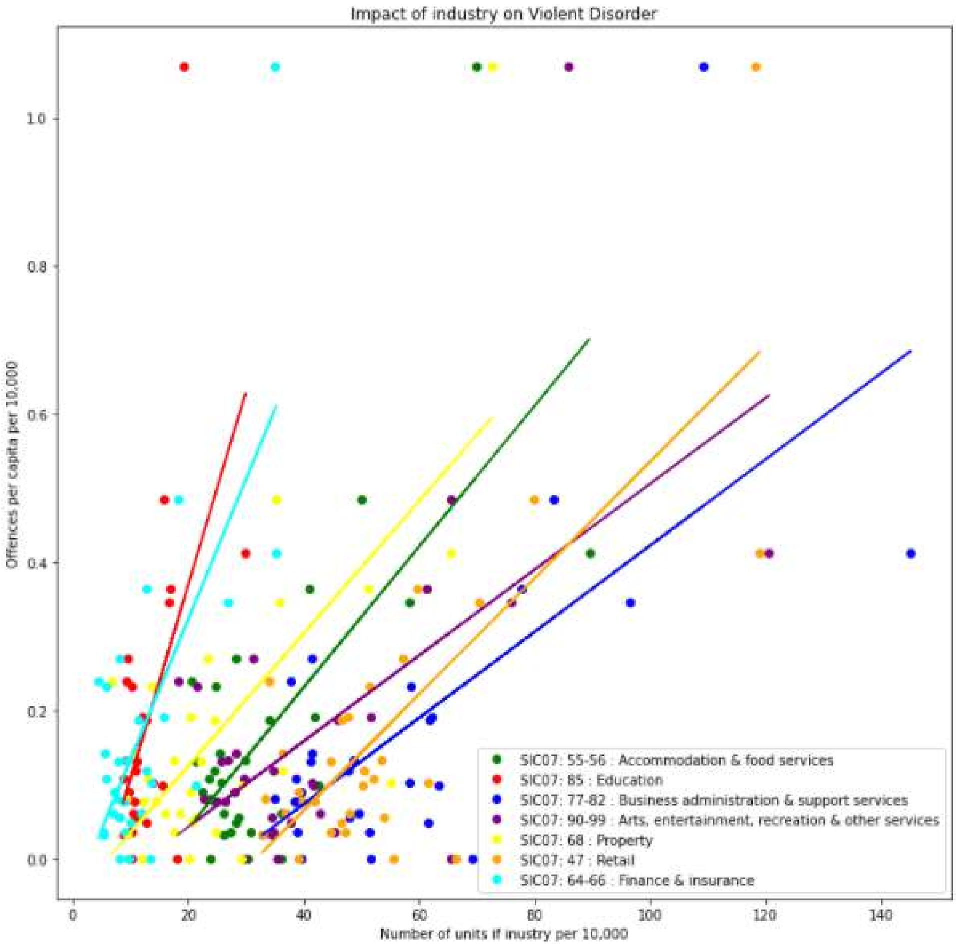


Figure 52: Best fit of violent disorder

Education and Finance& Insurance have strongest link to violent disorder – though note correlation is weak for education.

8.3 Foursquare recommended venues

The recommended venue data from Foursquares, was grouped by venue category, then normalised to create a profile of each neighbourhood's venues.

	Neighbourhood	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Windmill	Wine Bar	Wine Shop	Winery	Wings Joint	Women's Store	Xinjiang Restaurant	Yoga Studio	Zoo	Zoo Exhibit
0	Abbey Wood	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	Acton	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
2	Addington	0.0	0.0	0.000000	0.027778	0.027778	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	Aldgate	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.020000	0.020000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	Aldwych	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.016667	...	0.0	0.016667	0.016667	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
5	Anerley	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
6	Angel	0.0	0.0	0.015385	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.015385	0.0	0.0	0.0	0.0	0.015385	0.0	0.0
7	Arkley	0.0	0.0	0.025641	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
8	Balham	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.024390	0.0	0.0
9	Barbican	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

Figure 53: Example venue profiles

This was further manipulated to show what the 10 most common categories of venues recommended in each neighbourhood were.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbey Wood	Supermarket	Warehouse Store	Gym / Fitness Center	Train Station	Grocery Store	Historic Site	Clothing Store	Mobile Phone Shop	Fast Food Restaurant	Flea Market
1	Acton	Fast Food Restaurant	Brewery	Train Station	Hookah Bar	Food & Drink Shop	Bakery	Middle Eastern Restaurant	Bar	Chinese Restaurant	Pastry Shop
2	Addington	Gaming Cafe	Forest	Gym / Fitness Center	Rugby Pitch	Mediterranean Restaurant	Tram Station	Train Station	Golf Course	Pharmacy	Sandwich Place
3	Aldgate	Garden	Speakeasy	Bakery	Mediterranean Restaurant	Hotel Bar	Hotel	Office	French Restaurant	Beer Bar	Middle Eastern Restaurant
4	Aldwych	Restaurant	Burger Joint	Gelato Shop	Opera House	Noodle House	Gym	Spanish Restaurant	Gym / Fitness Center	Steakhouse	Museum
...	...	...	...	...	...	...	...	...	...	...	...
320	Woodside Park	Grocery Store	Gym / Fitness Center	Gift Shop	Newsagent	Golf Course	Café	Gourmet Shop	Greek Restaurant	Burger Joint	Beer Bar
321	Woolwich	Burger Joint	Platform	Boat or Ferry	Hotel	Pier	Tunnel	Steakhouse	Sandwich Place	Bakery	Pharmacy
322	Worcester Park	Pizza Place	Steakhouse	English Restaurant	Pharmacy	Grocery Store	Supermarket	Coffee Shop	BBQ Joint	Train Station	Bus Stop
323	Yeading	Hotel	Café	Sandwich Place	Park	Coffee Shop	Pharmacy	Supermarket	Clothing Store	Mexican Restaurant	Military Base
324	Yiewsley	Burger Joint	Gym	Coffee Shop	Bar	Fast Food Restaurant	Theater	Fish & Chips Shop	Chinese Restaurant	Gym / Fitness Center	Bookstore

Figure 54: Example top 10 venue categories for London neighbourhoods

This demonstrated a wide variety of popular venue categories, including less obvious choices, given these are recommended venues, including gas station, bus stop, train platform.

K-means clustering was used to identify 8 clusters of venues. Geocodable neighbourhoods only were considered. These clusters have been plotted on a map, which shows that although location not an input to the clustering algorithm, the geographic distribution does not appear random.

The cluster map was then overlaid with crime per 10,000 capita at borough level. Initial focus was with the crime categories which previous analysis showed were most likely to be impacted by industry units in an area ie Theft, Burglary from business or community and Violent disorder.

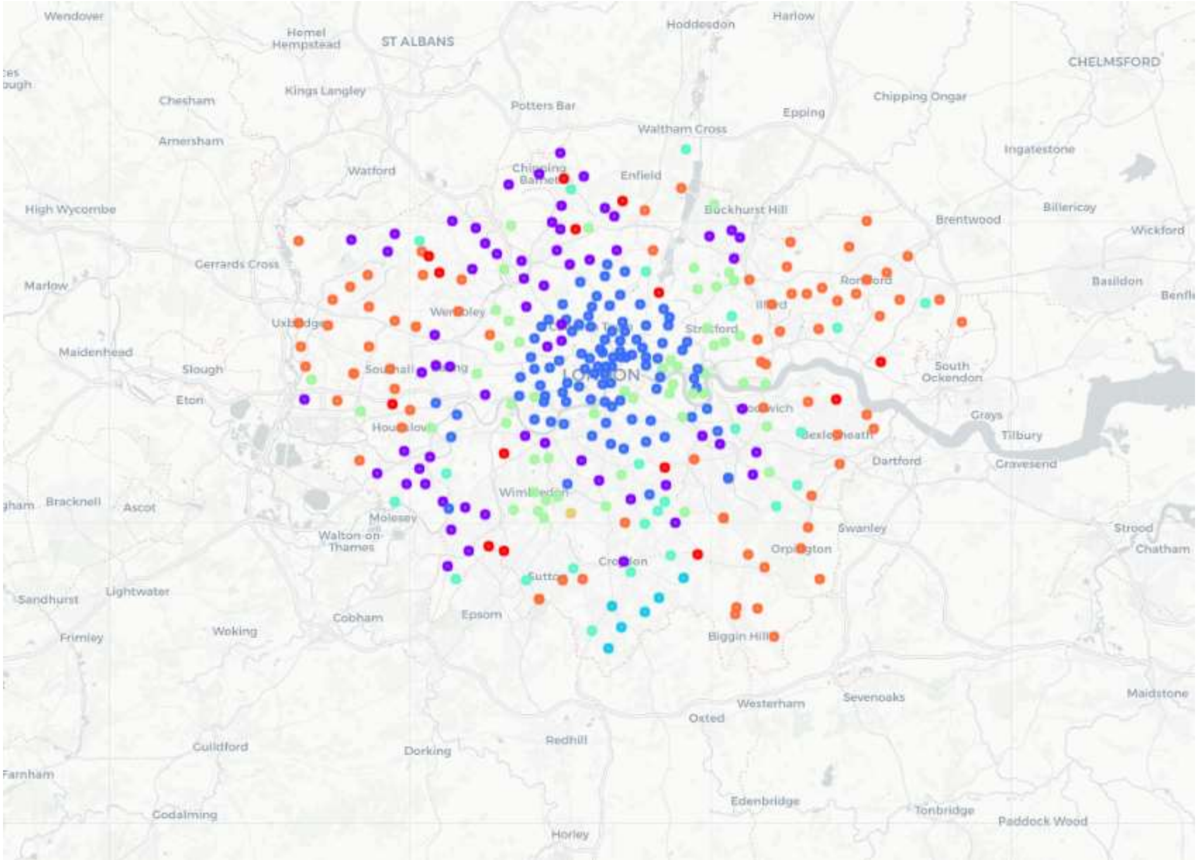


Figure 55: Eight clusters of London neighbourhoods based on venue profile

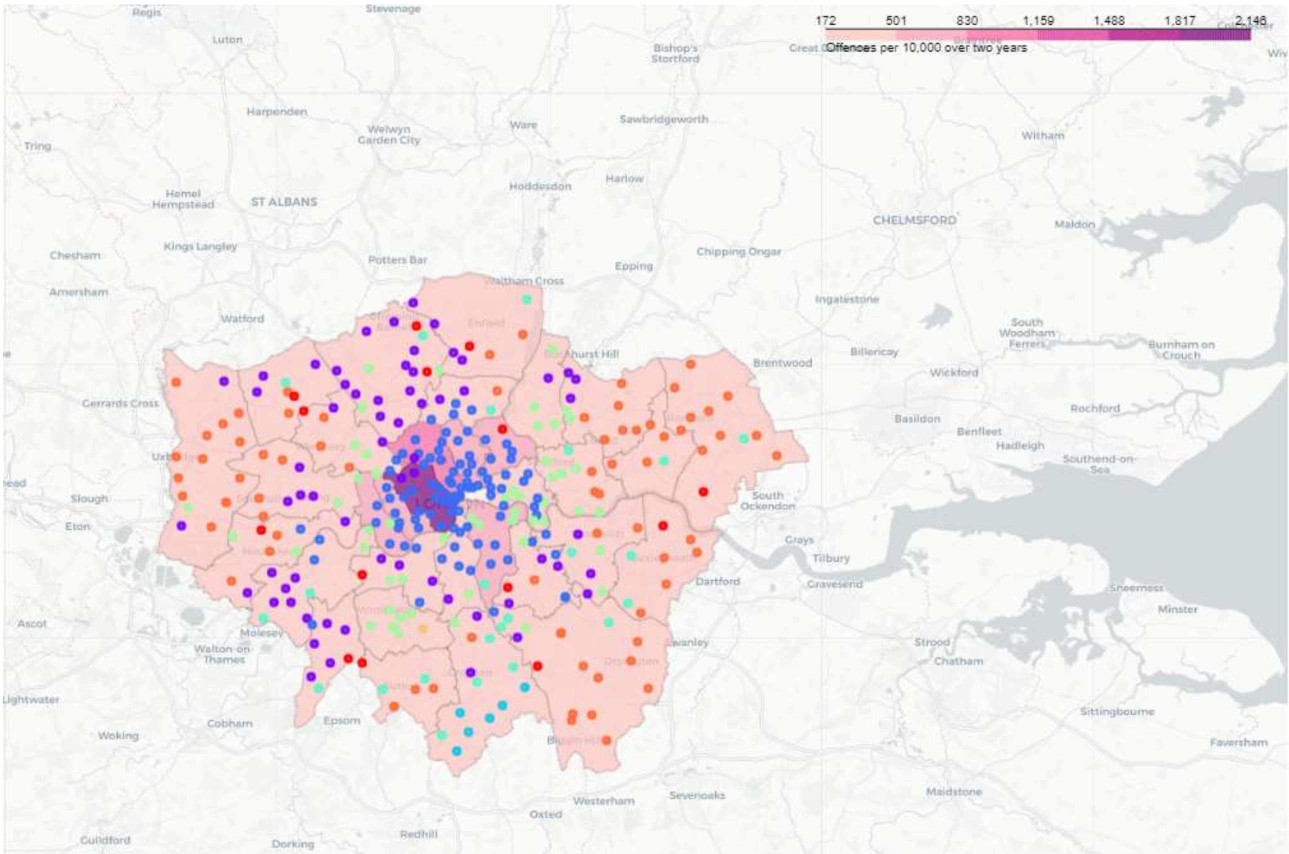


Figure 56: Theft shown against neighbourhood clusters

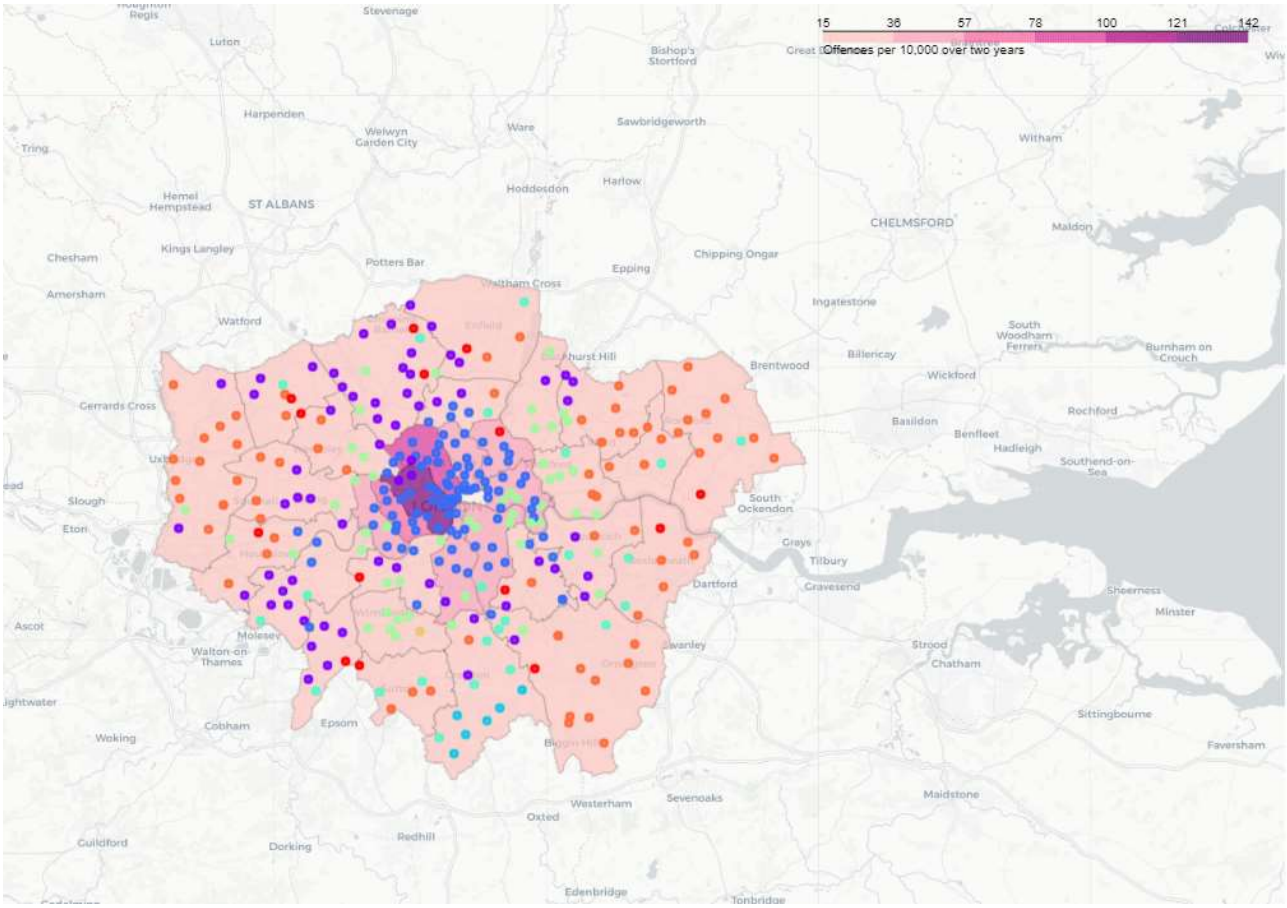


Figure 57: Burglary - Business and Community shown against neighbourhood clusters

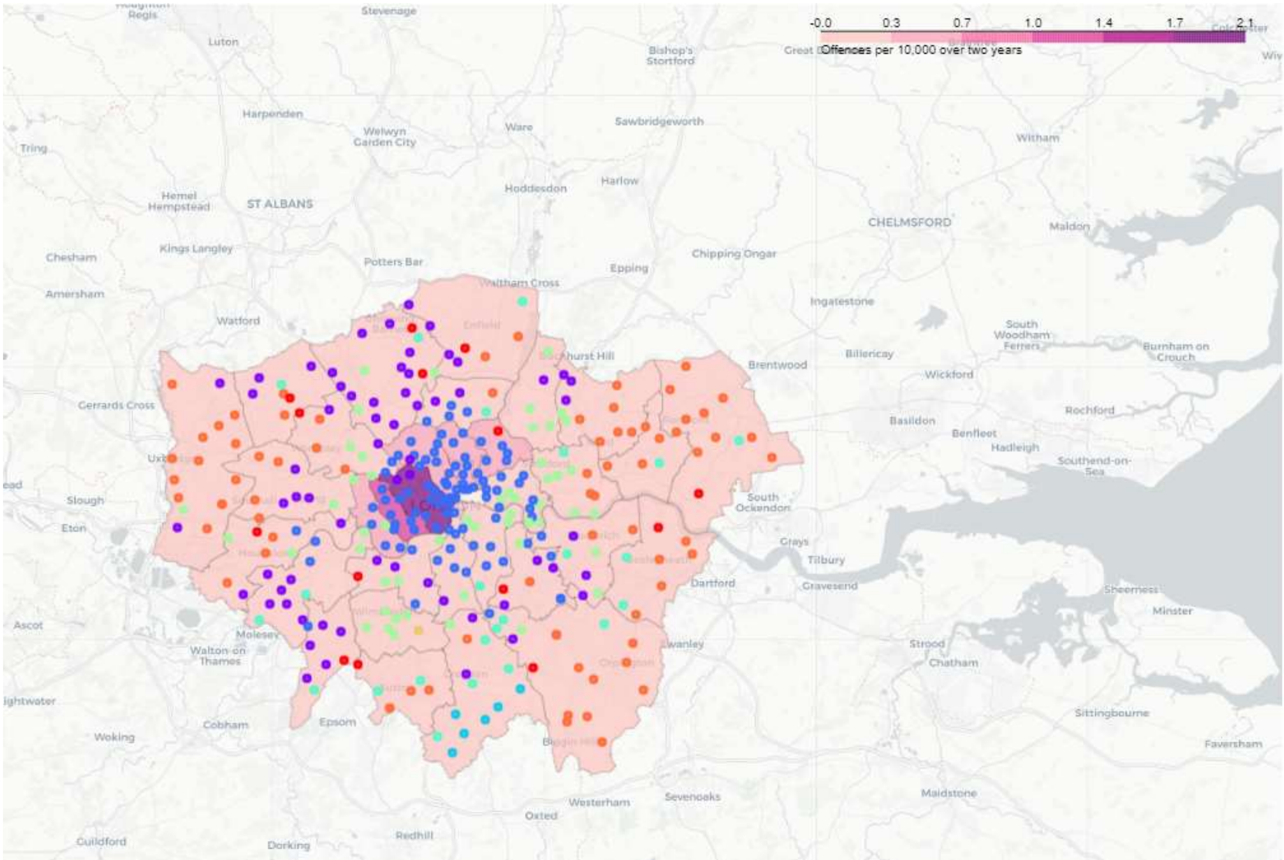


Figure 58: Violent disorder shown against neighbourhood clusters

These crimes(theft, burglary – business and community) and Violent disorder, follow the same distribution pattern with crime focussed on central London.

Drug offences was considered as it has a more diverse geographic spread. However no relation with neighbourhood clusters was discernable.

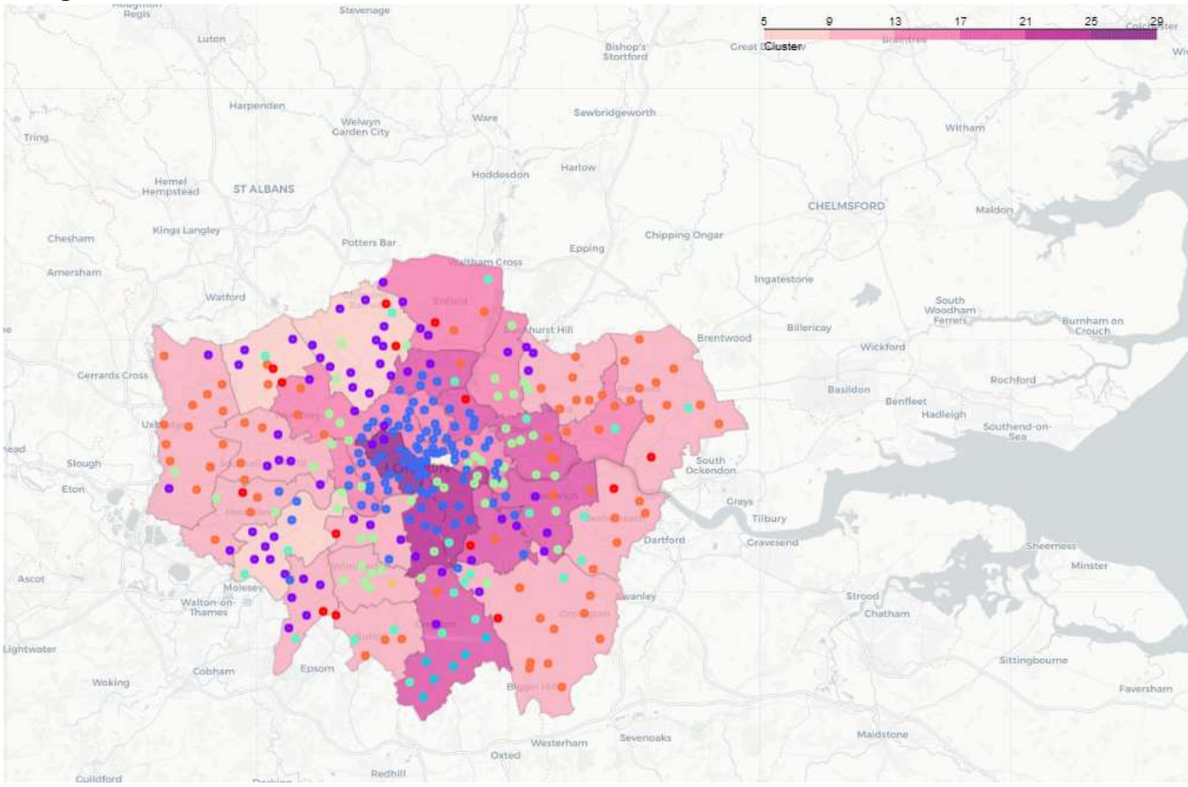


Figure 59: Drug offences shown against neighbourhood clusters

This analysis was repeated with other crime categories, not displayed here, and alternative numbers of clusters, however no pattern was noted.

9. Discussion

This report is to address the question:

- **Within the context of London, is there evidence of a relationship between the types of facilities/industries in a locality and crime?**

9.1 24 Hour Licenced Venues

There is some indication that crimes such as harassment, non residential burglary, theft from shops and common assault increases with the number of 24 hour licensed venues, though it is noted that all venue types are equal and the presence of 24 hour licenced large supermarkets does not appear to be problematic..

9.2 Local Units by Industry

There is some indication that, in general increased units, ie increase number of local businesses increased crime. The results indicated, that the impact was different across different industries and different crimes. Of note. Education and Production had the most impact on the rate Burglary – Business and Community. The scope of the analysis did not cover who the victim was, so it is not clear if this correlation was simply that these industries provide things to steal. Financial Services and Insurance appeared to have the strongest correlation with violent disorder. Interesting construction and motor trades, had no evidence of link to increased crime rate, indeed the evidence tentatively indicated a decrease.

9.3 Foursquare recommended venues

Nothing discernable was indicated from Foursquares recommended venue data, further analysis could be done for example meaningful grouping of venue categories, however given limitations in how data is sourced it seems unlikely.

9.4 Next steps

Although evidence has been found, the findings need to be treated with caution and are not conclusive as the sample size is small. It is therefore recommended that further analysis be conducted, for example:

- seeing if identified trends persisted in other years.
- seeing if impact of changes across years confirms findings, i.e. number of units increased, did crime?
- obtaining and completing analysis on more granular data, i.e. if neighbourhood level data could be obtained.

10. Conclusion

The businesses in an area appear to be linked to the crime profile, however correlation does not equal causation, and the presence of confounding factors is likely. It must be noted however that this study was limited by a small sample size of large geographic areas (London boroughs).

This report is focussed on a downside risk of business units in the area, however it is anticipated the stakeholders would also balance the upsides brought by additional business.

Appendix

List of crime classifications

In March 2019, the Metropolitan Police Service started to provide offences grouped by the updated Home Office crime classifications . This currently only covers the most recent 24 months of data, but historic data using the previous categories is available separately back to January 2008.

Below is a list of the crime types covered under the new HO categories (*not available at LSOA level):
Major Category: Minor Category
Arson and Criminal Damage - Arson / Criminal Damage
Burglary: Burglary – Business and Community / Burglary – Residential**
Drug Offences: Drug Trafficking / Possession of Drugs
Miscellaneous Crimes Against Society: Absconding from Lawful Custody / Bail Offences / Bigamy / Concealing an Infant Death Close to Birth / Dangerous Driving / Disclosure, Obstruction, False or Misleading State / Exploitation of Prostitution / Forgery or Use of Drug Prescription / Fraud or Forgery Associated with Driver Records / Going Equipped for Stealing / Handling Stolen Goods / Making, Supplying or Possessing Articles for use i / Obscene Publications / Offender Management Act / Other Forgery / Other Notifiable Offences / Perjury / Perverting Course of Justice / Possession of False Documents / Profiting From or Concealing Proceeds of Crime / Soliciting for Prostitution / Threat or Possession With Intent to Commit Criminal / Wildlife Crime
Possession of Weapons: Other Firearm Offences / Possession of Firearm with Intent / Possession of Firearms Offences / Possession of Other Weapon / Possession of Article with Blade or Point
Public Order Offences: Other Offences Against the State, or Public Order / Public Fear Alarm or Distress / Racially or Religiously Aggravated Public Fear / Violent Disorder
Robbery: Robbery of Business Property / Robbery of Personal Property
Sexual Offences*: Other Sexual Offences / Rape
Theft: Bicycle Theft / Other Theft / Shoplifting / Theft from Person
Vehicle Offences: Aggravated Vehicle Taking / Interfering with a Motor Vehicle / Theft from a Motor Vehicle / Theft or Taking of a Motor Vehicle
Violence Against the Person: Homicide / Violence with Injury / Violence without Injury
(source: https://data.london.gov.uk/dataset/recorded_crime_summary)

Key python libraries used

This analysis will be conducted in python, libraries used will include:

BeautifulSoup	for web scraping
numpy	to handle data in a vectorized manner
pandas	for data analysis
geopy.geocoders Nominatim	to convert an address into latitude and longitude values
sklearn.cluster	for k-means clustering
folium	for map rendering
matplotlib	for visualisations
seaborn	for visualisations