

IBM APPLIED DATA SCIENCE CAPSTONE

EXPLORING NEIGHBOURHOODS IN LONDON

by

IBRAHEEM KEKERE-EKUN

30th January, 2021

Table of Contents

1. Introduction

2. Data

3. Methodology

4. Results

5. Discussion

6. Conclusion

7. References

Figure 1: Wikipedia page showing list of areas in London

2.3.2. Location data

The Wikipedia page lacked information about the coordinates of the neighbourhoods. Therefore, Geocoder library was used to fetch the coordinates for each neighbourhood.

	Neighbourhood	Borough	Area	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	SE2	51.482490	0.119194
1	Acton	Ealing, Hammersmith and Fulham	W3, W4	51.495053	-0.220839
2	Addington	Croydon	CR0	51.344980	-0.012265
3	Addiscombe	Croydon	CR0	51.344980	-0.012265
4	Albany Park	Bexley	DA5, DA14	51.457075	0.184164

Figure 2: Geographical coordinates of neighbourhoods in London

The coordinates of each borough was gotten from a geojson file. This data was useful in visualizing choropleth maps ^[2].

2.3.3. Crime rate data

The data containing crime rates in London boroughs was gotten from finder.com ^[3]. The webpage contained information of the crime rates per 1000 people and the total number of crimes reported for each borough in 2020.

Number of crimes per borough 2019/2020			Table
Search for a borough			
Borough	Crime count	Crime rate	
Barking and Dagenham	21,169	99.5 per 1,000	
Barnet	31,911	80.6 per 1,000	
Bexley	18,323	73.8 per 1,000	
Brent	31,725	96.2 per 1,000	

Figure 3: webpage containing crime rates in London

2.3.4. House price data

The average housing price, as well as price per square feet in each London neighbourhood was gotten from propertydata.co.uk ^[4].

Property data for London

Prices, rents, yields, growth & more in London

[<< Back to all cities](#)

London is a major global city and world cultural capital, with strengths in the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism, and transportation. A major settlement for 2,000 years, London is the world's largest financial centre and the 5th or 6th richest city in the world.

The property investment capital of the UK, London has a diverse array of housing and communities across a giant urban area. You can find different types of long-let investment property in various areas to suit most budgets, and being the world's most-visited city, as measured by international arrivals, means there is no shortage of short-let opportunities too. Having the largest concentration of higher education institutes in Europe guarantees that London has an active student property rental market as well.

There are 217 postcode districts with significant property markets in and around London, and here we break down key property data for each.

Make informed property decisions - [sign up today](#).

Area	Avg yield	Avg price	£/sqft	5yr +/-	Explore data
BR1	3.4%	£445,600	£455	+15%	Explore data
BR2	3.3%	£484,100	£475	+14%	Explore data
BR3	3.4%	£445,400	£490	+14%	Explore data
BR5	3.0%	£477,700	£426	+18%	Explore data
BR6	3.0%	£538,200	£477	+14%	Explore data
BR7	2.8%	£605,254	£501	+15%	Explore data

Figure 4: webpage showing average house prices per square feet [£] in each London borough

2.3.5. Venue data

Foursquare api ^[5] was used to get data of all the venues and their categories for each London neighbourhood. Foursquare is an American location data provider.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Wood	51.482490	0.119194	Clam Field Recreation Area	51.479830	0.121023	Athletics & Sports
1	Abbey Wood	51.482490	0.119194	Abbey Wood Caravan Club	51.485502	0.120014	Campground
2	Abbey Wood	51.482490	0.119194	Bostall Woods	51.480026	0.115305	Forest
3	Abbey Wood	51.482490	0.119194	99 Bus to WW	51.481398	0.125636	Bus Station
4	Acton	51.495053	-0.220839	The Queen's Head	51.495975	-0.220711	Pub

Figure 5: dataframe containing venues, venue categories and venue coordinates in London

2.4. Data cleaning

Before any modelling was done, it was important to do clean the data.

In the datasets, only required columns were retained. Unwanted characters such as £ signs, brackets, commas, and strings were removed from the datasets. This was important to make calculations and plots easy as Python doesn't read numbers with symbols.

There are 32 boroughs in London. The London neighbourhood dataset from Wikipedia had the number of unique boroughs in London to be 60. This was because some boroughs were merged together to form another borough, thereby increasing the number of boroughs. The data was merged with the crime rate and housing data and the number of boroughs was reduced to 32.

The columns were renamed and those that had the wrong data types were corrected. The picture below shows the refined data set.

	Neighbourhood	Borough	Area	Latitude	Longitude	Crime count	Crime rate	Avg price	£/sqft
0	Addington	Croydon	CR0	51.344980	-0.012265	36790	95.1	343779.0	440.0
1	Addiscombe	Croydon	CR0	51.344980	-0.012265	36790	95.1	343779.0	440.0
2	Coombe	Croydon	CR0	51.344980	-0.012265	36790	95.1	343779.0	440.0
3	Coulsdon	Croydon	CR5	51.300165	-0.123698	36790	95.1	524906.0	446.0
4	Croydon	Croydon	CR0	51.344980	-0.012265	36790	95.1	343779.0	440.0

Figure 6: The dataset after cleansing

3. Methodology

3.1. Exploratory data analysis

This step was very important in the project. It helped to summarize the characteristics of the data and gain more insight into the data. The following observations were made:

- There were a total of 416 rows or neighbourhoods in the dataset
- There were 32 unique boroughs in the dataset.
- There were no Nan values in the dataset

3.2. Data preprocessing

Foursquare api was used to return the venues in London. The data contained each venue's category and coordinates. There were 226 unique venue categories in the data. The data was converted into numerical variables because K Means algorithm does a better job at prediction when taking in numbers

as inputs. This technique is known as One-hot encoding. One-hot encoding simply creates additional features based on the number of unique values in the categorical feature. Every unique value in the category will be added as a feature.

3.3.Algorithm

The first step in K Means is to randomly select k centroids, where k is equal to the number of clusters you choose. Centroids are data points representing the center of a cluster. The main element of the algorithm works by a two-step process called expectation-maximization. The expectation step assigns each data point to its nearest centroid. Then, the maximization step computes the mean of all the points for each cluster and sets the new centroid.

In this project, a model called KElbowVisualizer was used to show the optimum value for k .

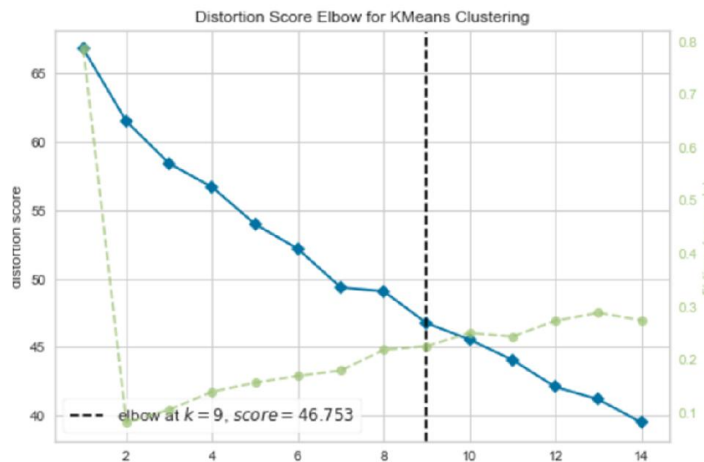


Figure 7: Predicting the number of clusters

The optimum value for k was predicted to be 8 with a distortion score of 46.753.

The data was fitted into the K Means algorithm and 9 clusters were produced.

The top 10 most common venues in London were retrieved and stored in a new dataframe with their corresponding neighbourhood, borough and cluster.

[06]:

	Neighbourhood	Borough	Area	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
2	Addington	Croydon	CR0	51.344980	-0.012265	7	Convenience Store	Construction & Landscaping	Grocery Store	Tr Stat
3	Addiscombe	Croydon	CR0	51.344980	-0.012265	7	Convenience Store	Construction & Landscaping	Grocery Store	Tr Stat
5	Aldborough Hatch	Redbridge	IG2	51.577600	0.070668	7	Pizza Place	Grocery Store	Bus Stop	Bak
7	Aldwych	Westminster	WC2	51.512847	-0.121873	8	Steakhouse	Fast Food Restaurant	Fabric Shop	Playgrou

Figure 9: The top 10 most common venues in each neighbourhood

4. Results

Folium library was used to visualize the clusters on the map of London. The following table shows the colours representing each of the clusters.

Cluster0	Green
Cluster1	Blue
Cluster2	Purple
Cluster3	Red
Cluster4	Orange
Cluster5	Brown
Cluster6	Pink
Cluster7	Yellow

The picture below shows the map of London with each marker representing a neighbourhood.

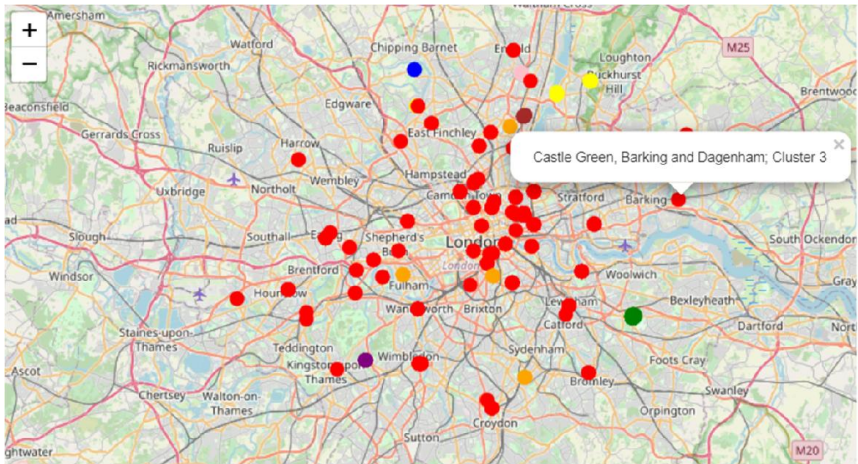


Figure 8: Each marker represents a neighbourhood in London and different colors represent different clusters.

4.1. Cluster analysis

After clustering, each cluster was analyzed and the number of each of the most common venue category was retrieved. The following results were obtained:

4.1.1. Cluster0

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Cor \
0	Barnehurst	Bexley	DA7	18323	73.8	445730.0	406.0	51.451113	0.054593	Low £/sqft	...	Mediterranean Restaurant	Dessert Shop	Pizza Place	
1	Barnes Cray	Bexley	DA1	18323	73.8	307564.0	363.0	51.450560	0.054320	Low £/sqft	...	Mediterranean Restaurant	Dessert Shop	Supermarket	Pizza
2	Belvedere	Bexley	DA17	18323	73.8	313878.0	351.0	51.451470	0.055490	Medium £/sqft	...	Dessert Shop	Pub	Park	Rest
3	Bexley (also Old Bexley, Bexley Village)	Bexley	DA5	18323	73.8	506967.0	426.0	51.451097	0.054533	Medium £/sqft	...	Mediterranean Restaurant	Dessert Shop	Pizza Place	
4	Blackfen	Bexley	DA15	18323	73.8	427473.0	436.0	51.451360	0.055090	Low £/sqft	...	Italian Restaurant	Pub	Park	Mediterr
5	Colliers	Bexley	DA8	18323	73.8	316433.0	344.0	51.450810	0.054550	Low £/sqft	...	Mediterranean Restaurant	Dessert Shop	Pizza Place	
6	Crayford	Bexley	DA1	18323	73.8	307564.0	363.0	51.450560	0.054320	Medium £/sqft	...	Mediterranean Restaurant	Dessert Shop	Supermarket	Pizza

Figure 10: Cluster0

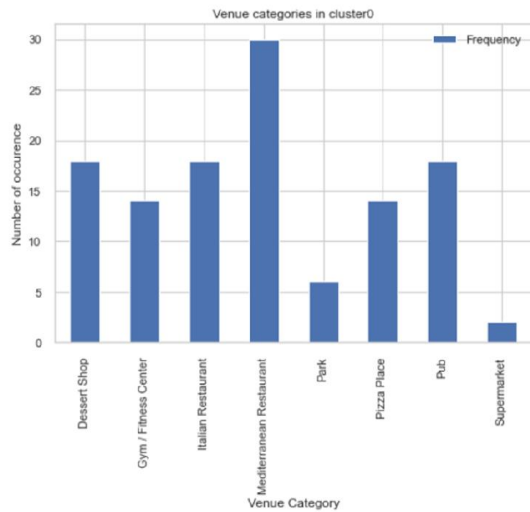


Figure 11: The most common venue categories in cluster0

Cluster0 was in the Bexley borough. The neighbourhoods in this cluster were Barnehurst, Barnes Cray, Belvedere, Bexley (also Old Bexley, Bexley Village), Blackfen, Colyers, Crayford, Crook Log, East Wickham, Foots Cray, Lamorbey, Lessness Heath, North Cray, North End, Northumberland Heath, Slade Green, Upton, and Welling. The most common venue categories were Mediterranean restaurants, pubs and Italian restaurants.

4.1.2. Cluster1

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Childs Hill	Barnet	NW2	31911	80.6	530540.0	629.0	51.63568	-0.19847	Low £/sqft	...	Bus Stop	Yoga Studio	Escape Room	Food & Drink Shop	F
1	East Finchley	Barnet	N2	31911	80.6	615603.0	550.0	51.63568	-0.19847	Low £/sqft	...	Bus Stop	Yoga Studio	Escape Room	Food & Drink Shop	F
2	Hampstead Garden Suburb	Barnet	N2	31911	80.6	615603.0	550.0	51.63568	-0.19847	Low £/sqft	...	Bus Stop	Yoga Studio	Escape Room	Food & Drink Shop	F

3 rows × 21 columns

Figure 12: Cluster1

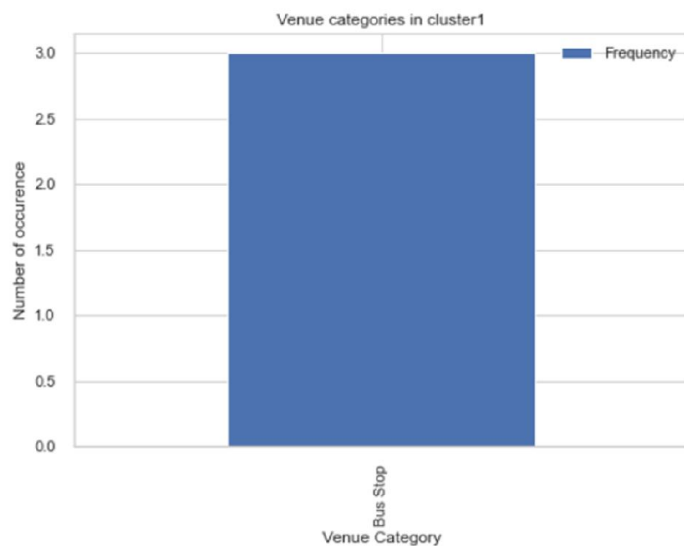


Figure 13: Venues in cluster1

Cluster1 was in the Barnet borough. The neighbourhoods in this cluster were Childs Hill, East Finchley, and Hampstead Garden Suburb. The most common venue categories were Bus stops.

4.1.3. Cluster2

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	...
0	Canning Town	Newham	E16	37107	105.1	483098.0	555.0	51.41781	-0.25775	Low £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
1	Custom House	Newham	E16	37107	105.1	483098.0	555.0	51.41781	-0.25775	Low £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
2	East Ham	Newham	E6	37107	105.1	381311.0	417.0	51.41781	-0.25775	Low £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
3	Forest Gate	Newham	E7	37107	105.1	471401.0	471.0	51.41781	-0.25775	Low £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
4	Little Ilford	Newham	E12	37107	105.1	417461.0	410.0	51.41781	-0.25775	Low £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
5	Manor Park	Newham	E12	37107	105.1	417461.0	410.0	51.41781	-0.25775	Medium £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
6	Maryland	Newham	E15	37107	105.1	433181.0	543.0	51.41781	-0.25775	Medium £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...
7	North Woolwich	Newham	E16	37107	105.1	483098.0	555.0	51.41781	-0.25775	Medium £/sqft	...	Nutritionist	Bakery	Golf Course	Yoga Studio	...

Figure 14: Cluster2

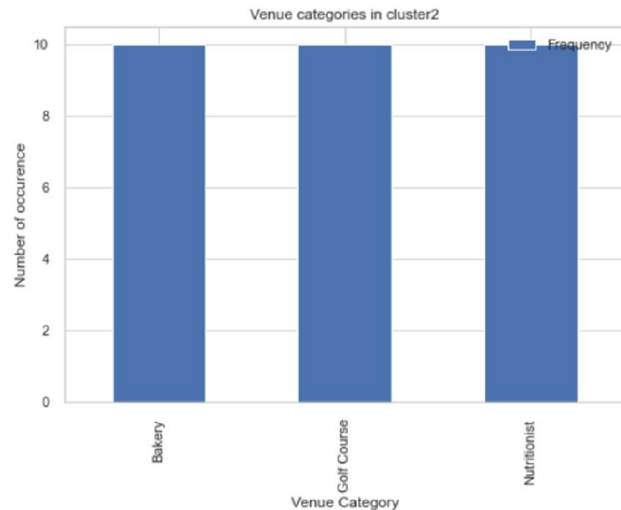


Figure 15: Venue categories in cluster2

Cluster2 was in the Newham borough. The neighbourhoods in this cluster were Canning Town, Custom House, East Ham, Forest Gate, Little Ilford, Manor Park, Maryland, North Woolwich, Silvertown, and Stratford. The most common venue categories were bakeries, golf courses and nutritionists.

4.1.4. Cluster3

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	...
0	Addington	Croydon	CR0	36790	95.1	343090.0	439.0	51.381893	-0.106196	Low £/sqft	...	Platform	Pizza Place	Dessert Shop	Bakery	...
1	Addiscombe	Croydon	CR0	36790	95.1	343090.0	439.0	51.381893	-0.106196	Low £/sqft	...	Platform	Pizza Place	Dessert Shop	Bakery	...
2	Coombe	Croydon	CR0	36790	95.1	343090.0	439.0	51.381893	-0.106196	Low £/sqft	...	Pub	Café	Pizza Place	Italian Restaurant	...
3	Croydon	Croydon	CR0	36790	95.1	343090.0	439.0	51.381893	-0.106196	Low £/sqft	...	Platform	Pizza Place	Dessert Shop	Bakery	...
4	Forestdale	Croydon	CR0	36790	95.1	343090.0	439.0	51.381893	-0.106196	Low £/sqft	...	Platform	Pizza Place	Dessert Shop	Bakery	...
...
289	Shoreditch	Hackney	N1	34510	122.8	712517.0	907.0	51.545050	-0.055320	Low £/sqft	...	Pub	Coffee Shop	Café	Cocktail Bar	...

Figure 16: Cluster3

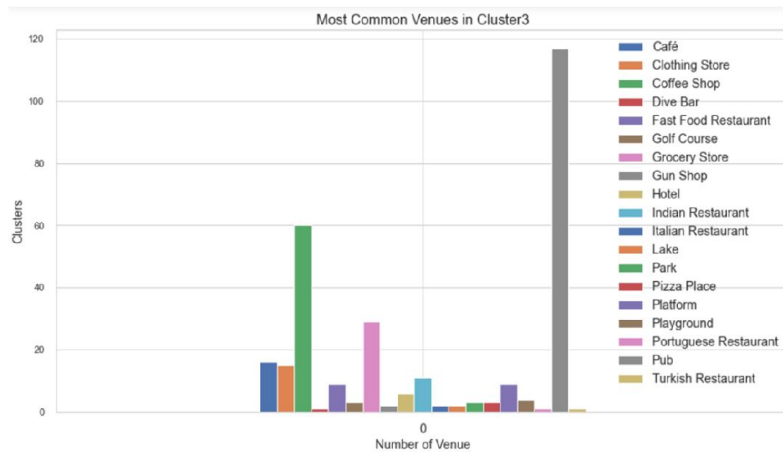


Figure 16: Most common venue categories in cluster3

Cluster3 was in the Croydon, Bexley, Redbridge, Westminster, Brent, Bromley, Islington, Havering, Barnet, Enfield, Wandsworth, Southwark, Barking and Dagenham, Richmond upon Thames, Newham, Sutton, Ealing, Lewisham, Harrow, Camden, Kingston upon Thames, Tower Hamlets, Greenwich, Haringey, Hounslow, Lambeth, Kensington and Chelsea, Merton, and Hackney boroughs. Therefore, cluster3 had the most number of neighbourhoods. The most common venue categories were pubs, coffee shops and grocery stores.

4.1.5. Cluster4

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Co
0	Coulsdon	Croydon	CR5	36790	95.1	530180.0	437.0	51.405360	-0.055540	Low £/sqft	...	Grocery Store	Café	Park	Ent Rest
1	Kenley	Croydon	CR8	36790	95.1	443532.0	415.0	51.405260	-0.055720	Low £/sqft	...	Grocery Store	Café	Park	Ent Rest
2	Old Coulsdon	Croydon	CR5	36790	95.1	530180.0	437.0	51.405360	-0.055540	Low £/sqft	...	Grocery Store	Café	Park	Ent Rest
3	Purley	Croydon	CR8	36790	95.1	443532.0	415.0	51.405260	-0.055720	Low £/sqft	...	Grocery Store	Café	Park	Ent Rest
4	Riddlesdown	Croydon	CR8	36790	95.1	443532.0	415.0	51.405260	-0.055720	Low £/sqft	...	Grocery Store	Café	Park	Ent Rest
5	Senderstead	Croydon	CR2	36790	95.1	380293.0	442.0	51.405460	-0.055350	Low £/sqft	...	Grocery Store	Café	Park	Ent Rest
6	Seahurst	Croydon	SE25	36790	95.1	351493.0	445.0	51.593480	-0.083420	Low £/sqft	...	Playground	Fast Food Restaurant	Chinese Restaurant	

Figure 17: Cluster4

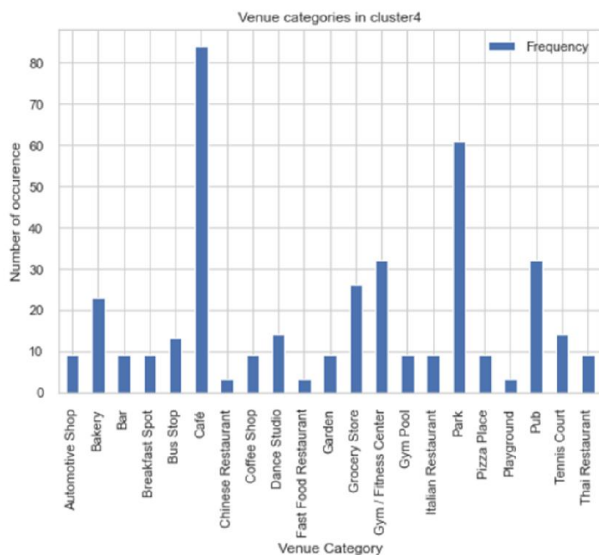


Figure 18: Venue categories in cluster4

The following table shows the boroughs and neighbourhoods in cluster4

Borough	Neighbourhood
Croydon	Coulsdon, Kenley, Old Coulsdon, Purley, Riddlesdown, Sanderstead, Selhurst, Selsdon, South Croydon, South Norwood and Upper Norwood.
Hammersmith and Fulham	Fulham, Hammersmith, Old Oak Common, Parsons Green, Sands End, Shepherd's Bush, West Kensington, White City and Wormwood Scrubs.
Hillingdon	Cowley, Harefield, Harlington, Harmondsworth, Ickenham, Longford, Northwood, Ruislip, Sipson, South Ruislip, Uxbridge, West Drayton, Yeadon, and Yiewsley.

The most common venue categories were parks, cafés, pubs and gym/fitness centres.

4.1.6. Cluster5

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Anerley	Bromley	SE20	25802	77.6	384045.0	520.0	51.601511	-0.066365	Low £/sqft	...	Sporting Goods Shop	Bar	Hostel	Pub
1	Aperfield	Bromley	TN16	25802	77.6	444540.0	411.0	51.601520	-0.066750	Low £/sqft	...	Sporting Goods Shop	Bar	Brewery	Mediterranean Restaurant
2	Bickley	Bromley	BR3	25802	77.6	448269.0	503.0	51.601380	-0.066960	High £/sqft	...	Sporting Goods Shop	Bar	Hostel	Pub
3	Biggin Hill	Bromley	TN16	25802	77.6	444540.0	411.0	51.601520	-0.066750	Low £/sqft	...	Sporting Goods Shop	Bar	Brewery	Mediterranean Restaurant
4	Bromley Common	Bromley	BR3	25802	77.6	448269.0	503.0	51.601380	-0.066960	Low £/sqft	...	Sporting Goods Shop	Bar	Hostel	Pub
5	Chelsfield	Bromley	BR6	25802	77.6	544046.0	474.0	51.601380	-0.066750	Medium £/sqft	...	Sporting Goods Shop	Bar	Hostel	Pub
6	Chislehurst	Bromley	BR3	25802	77.6	544046.0	474.0	51.601380	-0.066750	Medium £/sqft	...	Sporting Goods Shop	Bar	Hostel	Pub

Figure 19: Cluster5

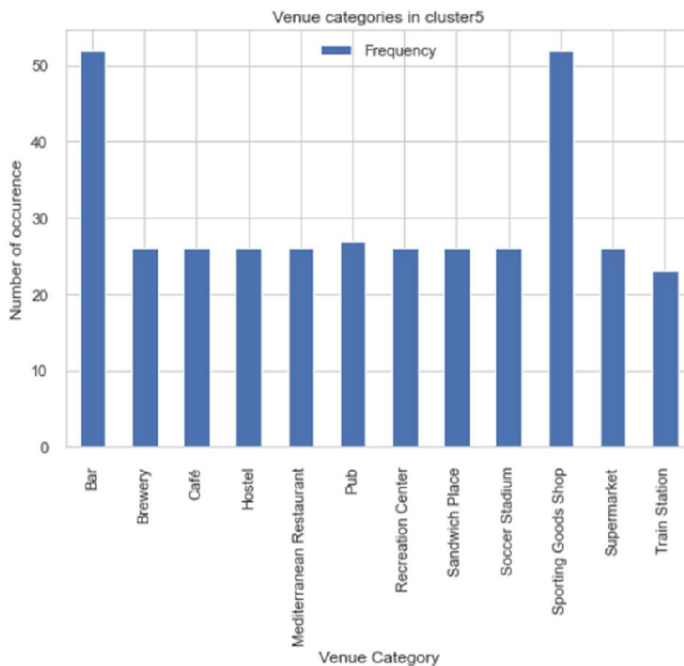


Figure 20: Venue categories in cluster5

Cluster5 was in the Bromley borough. The neighbourhoods in this cluster were Anerley, Aperfield, Bickley, Biggin Hill, Bromley Common, Chelsfield, Chislehurst, Cudham, Derry Downs, Downe, Eden Park, Elmers End, Elmstead, Goddington, Hayes, Hazelwood, Keston, Leaves Green, Locksbottom,

Mottingham, Penge, Petts Wood, Pratt's Bottom, Southborough, St Mary Cray and St Paul's Cray. The most common venue categories were bars, sporting goods shops and pubs.

4.1.7. Cluster6

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Harrow Weald	Harrow	HA3	17909	71.3	491667.0	478.0	51.63327	-0.07023	Low £/sqft	...	Park	Bakery	Museum	Film Studio	Electronics Store
1	Hatch End	Harrow	HA5	17909	71.3	576077.0	561.0	51.63325	-0.07012	Medium £/sqft	...	Park	Bakery	Museum	Film Studio	Electronics Store
2	Pinner	Harrow	HA5	17909	71.3	576077.0	561.0	51.63325	-0.07012	Low £/sqft	...	Park	Bakery	Museum	Film Studio	Electronics Store
3	Rayners Lane	Harrow	HA5	17909	71.3	576077.0	561.0	51.63325	-0.07012	Low £/sqft	...	Park	Bakery	Museum	Film Studio	Electronics Store
4	South Harrow	Harrow	HA2	17909	71.3	481380.0	493.0	51.63303	-0.07063	Low £/sqft	...	Park	Bakery	Museum	Home Service	Electronics Store
5	Wealdstone	Harrow	HA3	17909	71.3	491667.0	478.0	51.63327	-0.07023	Low £/sqft	...	Park	Bakery	Museum	Film Studio	Electronics Store

Figure 21: Cluster6

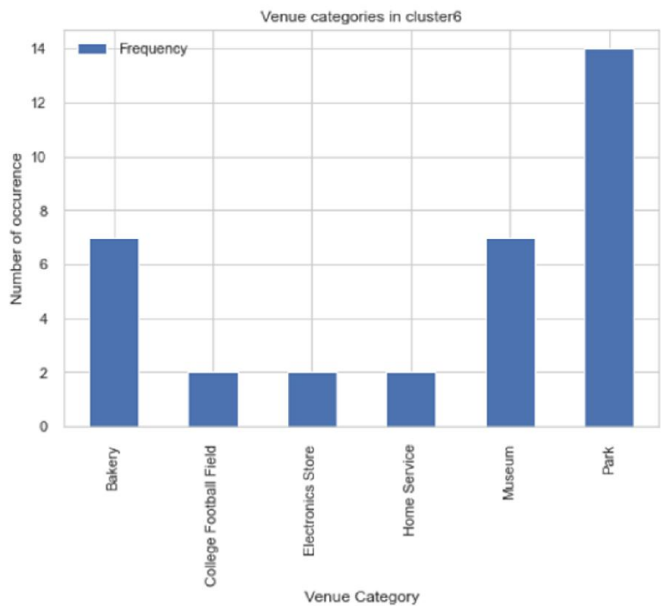


Figure 22: Bar plot showing venue categories in cluster6

Cluster6 was in the Harrow borough. The neighbourhoods in this cluster were Harrow Weald, Hatch End, Pinner, Rayners Lane, South Harrow, Wealdstone and West Harrow. The most common venue categories were parks, bakeries and museums.

4.1.8. Cluster7

	Neighbourhood	Borough	Area	Crime count	Crime rate	Avg price	£/sqft	Latitude	Longitude	price_binned	...	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Aldwych	Westminster	WC2	73603	281.7	1652502.0	1017.0	51.62773	0.01322	Low £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
1	Baywater	Westminster	W2	73603	281.7	1035301.0	1226.0	51.62773	0.01322	Low £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
2	Charing Cross	Westminster	WC2	73603	281.7	1652502.0	1017.0	51.62773	0.01322	Medium £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
3	Chinatown	Westminster	W1	73603	281.7	2542697.0	1177.0	51.62774	0.01319	Medium £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
4	Covent Garden	Westminster	WC2	73603	281.7	1652502.0	1017.0	51.62773	0.01322	Medium £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
5	Lisson Grove	Westminster	NW8	73603	281.7	1097470.0	1088.0	51.62772	0.01306	High £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
6	Maida Vale	Westminster	W9	73603	281.7	749526.0	950.0	51.62773	0.01288	High £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
7	Marylebone (also St Marylebone)	Westminster	W1	73603	281.7	2542697.0	1177.0	51.62774	0.01319	High £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store
8	Mayfair	Westminster	W1	73603	281.7	2542697.0	1177.0	51.62774	0.01319	High £/sqft	...	Park	Pub	Athletics & Sports	Yoga Studio	Electronics Store

Figure 23: Cluster7

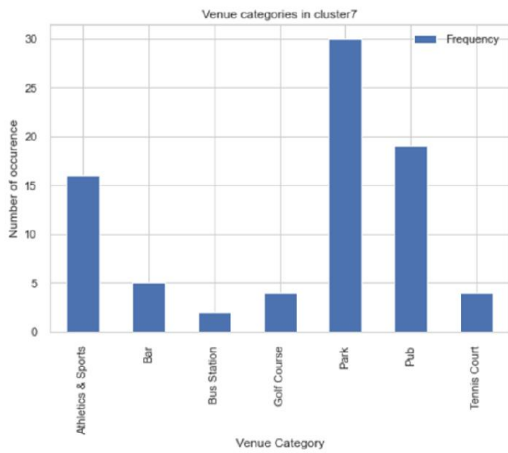


Figure 24: Bar plot showing venue categories in cluster7

The following table shows the boroughs and neighbourhoods in cluster7

Borough	Neighbourhood
Brent	Kingsbury, Preston, Tokyngto and Wembley Park
Waltham Forest	Cann Hall, Chingford, Highams Park, Leytonstone, Upper Walthamstow, Walthamstow and Walthamstow Village
Westminster	Aldwych, Bayswater, Charing Cross, Chinatown, Covent Garden, Lisson Grove, Maida Vale, Marylebone (also St Marylebone), Mayfair, Paddington, Soho and St John's Wood.

The most common venue categories were parks, pubs and athletics and sports.

At the end of the cluster analysis, the top three most common venues in each neighbourhood and their numbers were obtained.

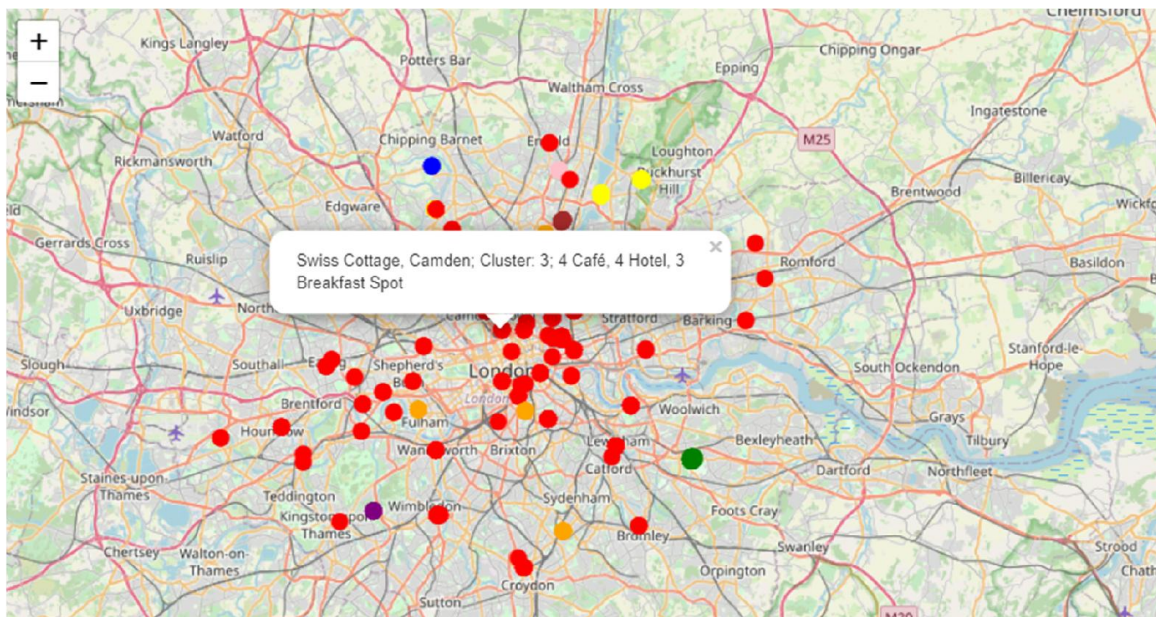


Figure 25: Map of London showing clusters and number of top three venues.

Cluster3 had the most number of venue categories and neighbourhoods. The picture below shows the map of London with only neighbourhoods in cluster3. From the map, we see that cluster3 is scattered around London.

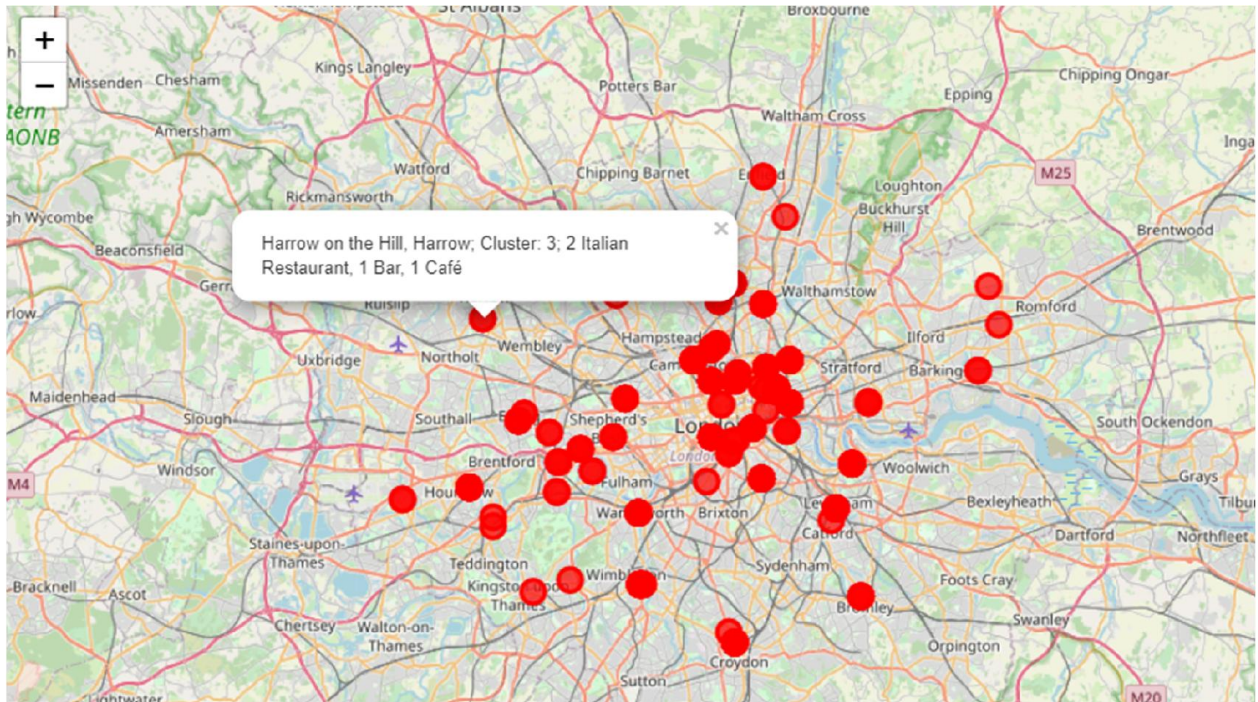


Figure 26: Map showing neighbourhoods in cluster3

Clusters on crime rate choropleth map

Using Folium's choropleth feature, we can visualize the crime rates in London. The darker the colour, the higher the crime rate.

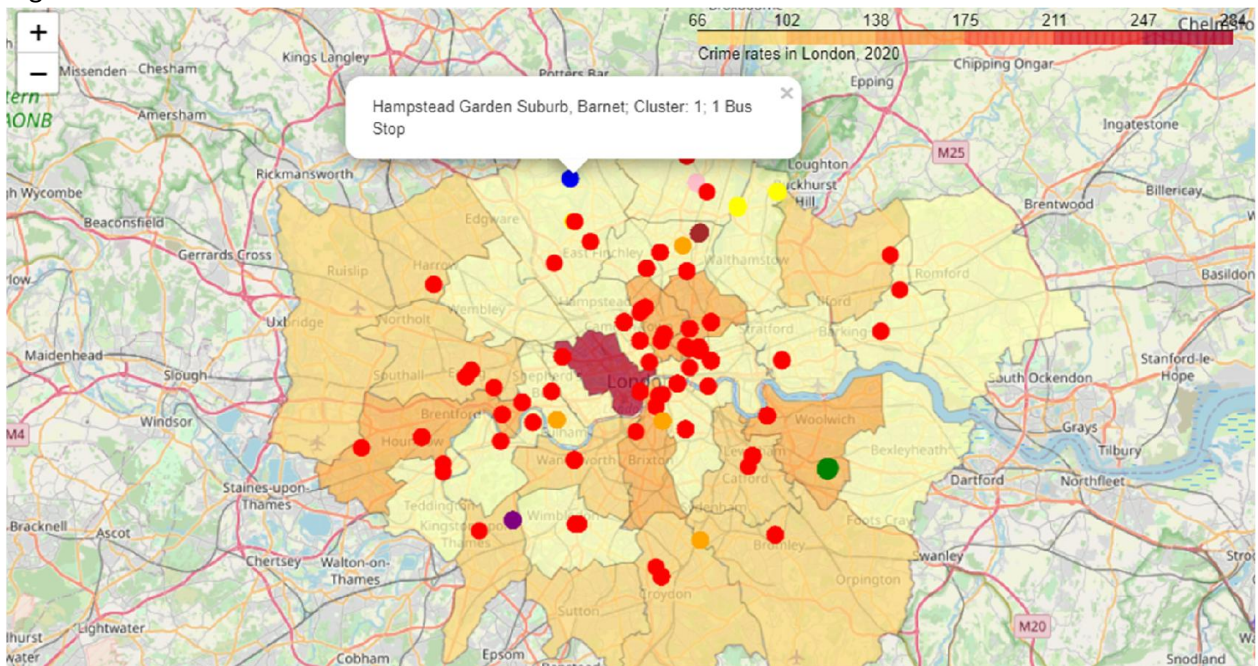


Figure 27: Crime rate choropleth map showing clusters

Clusters on average house sale price choropleth map

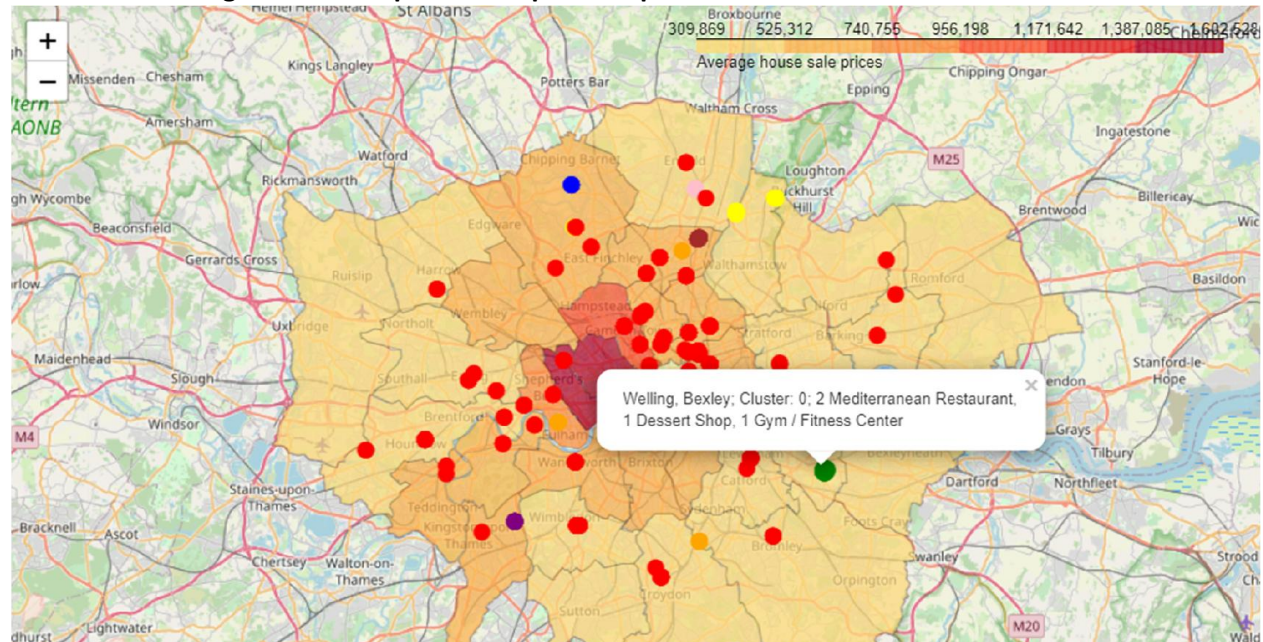


Figure 28: Average house sale price choropleth map showing clusters

Clusters on average house price per square feet choropleth map

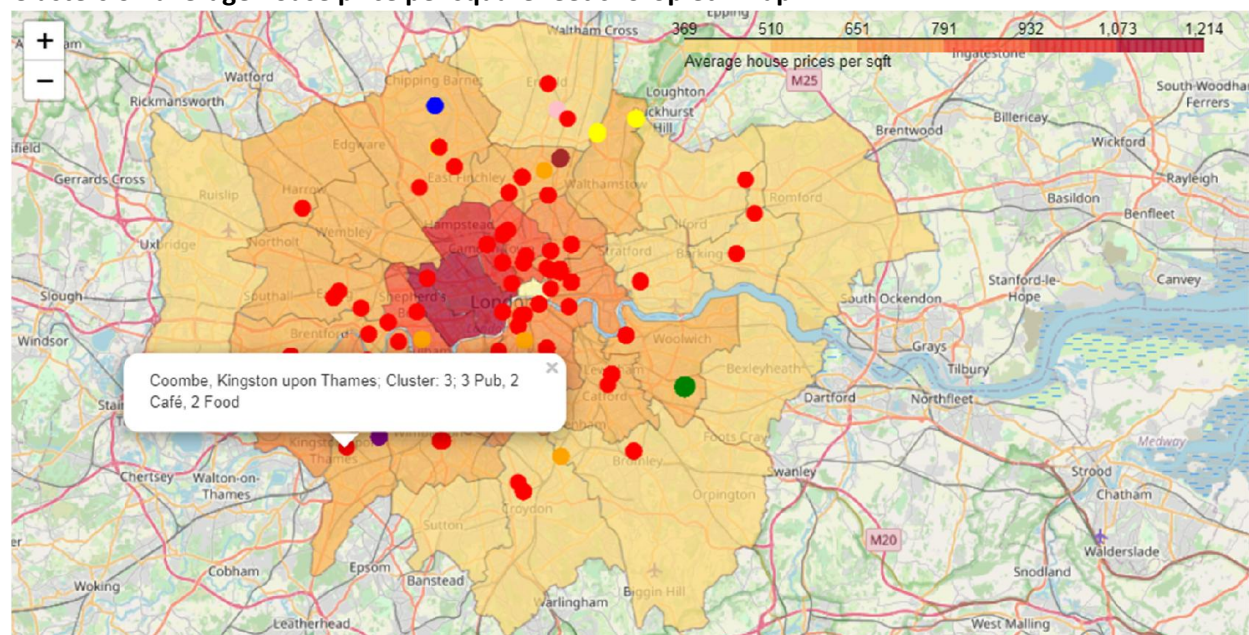


Figure 29: Average house price per square feet choropleth map showing clusters

From the maps above, it can be seen that the more expensive houses are located in the middle of London.

4.2. Boroughs with low average house price per sqft and below the median crime rate

4.2.1. Finding the neighbourhoods below the median crime rate

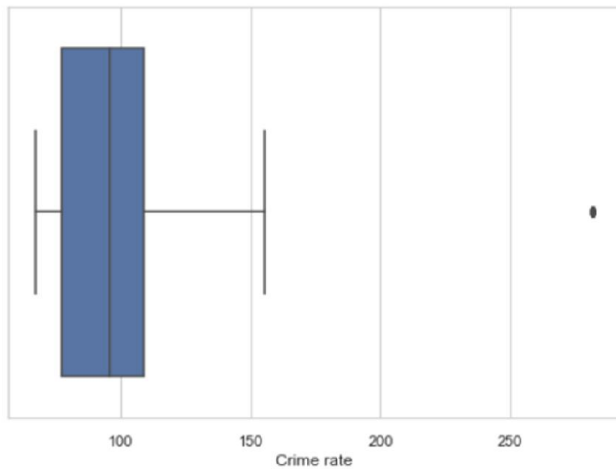


Figure 30: Box plot showing the crime rates

Barnet, Bexley, Bromley, Croydon, Ealing, Enfield, Harrow, Havering, Hillingdon, Kingston upon Thames, Merton, Redbridge, Richmond upon Thames, Sutton, Waltham Forest and Wandsworth' have crime rates below the median crime rate of 96.1.

Boroughs with average house sale price per square feet in the low range

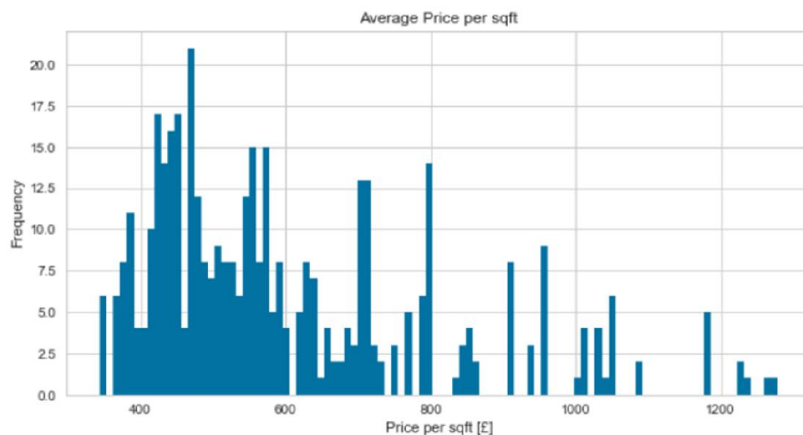


Figure 31: Bar plot showing the average house price per sqft

Croydon, Bexley, Redbridge, Westminster, Brent, Bromley, Islington, Havering, Barnet, Enfield, Wandsworth, Southwark, Barking and Dagenham, Richmond upon Thames, Newham, Sutton, Ealing, Lewisham, Harrow, Camden, Kingston upon Thames, Tower Hamlets, Greenwich, Haringey, Hounslow, Waltham Forest, Kensington and Chelsea, Merton, Hillingdon, Hackney and Hammersmith and Fulham have the average housing price per sqft in the low range.

Boroughs with low average price per sqft and below the median crime rate

Sixteen boroughs were found to have these characteristics. They are; Barnet, Bexley, Bromley, Croydon, Ealing, Enfield, Harrow, Havering, Hillingdon, Kingston upon Thames, Merton Redbridge, Richmond upon Thames, Sutton, Waltham Forest and Wandsworth.

5. Discussion

The aim of this project was to make it easy for anyone interested in the project easily understand the results of the project. We were able to cluster over 400 neighbourhoods in London containing 266 venue categories using K Means clustering algorithm. Nine clusters were obtained. We analyzed each cluster and showed the number of each of the most common venue category. Pubs, parks and cafés were among the most common venues overall. Cluster3 had the most number of venue categories. We retrieved the top three most common venues for each neighbourhood. The top three were selected because some neighbourhoods had very few venues.

We were able to visualize the crime rates and average house sale prices on the map of London. Finally, we retrieved the boroughs with low average house sale price and crime rates.

6. Conclusion

This project mainly focused on crime rates, house prices and common venues as criteria for choosing a neighbourhood to move to. There are other factors that people put into consideration before they move, but these are some of the most important ones. Therefore, the results of this project can be a good place to start for anyone that wants to move to London.

7. References

- [1] [Neighbourhood data from Wikipedia](#)
- [2] [Borough coordinates from github](#)
- [3] [Crime rates data from finder.com](#)
- [4] [Average house sale price](#)
- [5] [Venue data from Foursquare.com](#)