Dog Friendly London Neighborhoods

1. Introduction :

Trying to make a move to London with pets will require good planning and some research. London is a sprawling metropolis and whilst all the main tourist attractions are fairly close to each other, the surrounding areas make for a large and, to those who're not familiar with the city, somewhat confusing location.

Many rental properties do not accept pets; pet-friendly rentals in London are limited and tend to get snapped up pretty quickly. In order to succeed in your search you need to focus on certain neighbourhoods.

2. Objective:

This project aims to identify best neighbourhoods for small families with a dog. Main criterias are:

01 | Crime rate

02 | Commute time to Bank station

03 | 2 bedroom rental price

04 | Nearby parks

3. Data acquisition and cleaning

01 | Neighborhoods	**02 | Crime rate**	**03 | Commute time**
04 | Rental prices	**05 | Nearby Parks**	

01 | Neighborhoods

The first thing to understand about London is how the different areas are described. There are 33 **boroughs** of London, each with their own infrastructure, local government, council and highly individual feel and identity. These boroughs are then further broken down into what are known as '**postcodes**.'

Each Postcode is made up of the following elements: PO1 3AX

- PO the area. There are 124 postcode areas in the UK
- 1 the district. There are approximately 20 **Postcode districts** in an area
- 3 the sector. There are approximately 3000 addresses in a sector.
- AX the Unit. There are approximately 15 addresses per unit.

A Lower Layer Super Output Area (LSOA) is a geographic area designed to improve the reporting of small area statistics in England and Wales.

There are now **4,836** lower layer super output areas (LSOA) and **190** postcode districts in London. In this project, the results will be displayed in LSOA level.

Source : London data store - Greater London Authority
Source link :
1.https://data.london.gov.uk/dataset/postcode-directory-for-london
2.https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london
Data details: Postcode, Latitude, Longitude, LSOA

LSOAs in London

Source: Office for National Statistics, 2019

	objectid	lsoa11nm	lsoa11nmw	st_areashape	st_lengthshape	geometry	Postcode	LSOA2	Latitude	Longitude	Postcode District
LSOA											
E01000722	5	Bromley 023A	Bromley 023A	569162.040237	3524.308527	POLYGON ((0.1146063324029615 51.39254569837126...	BR5 2EE	E00003519	51.389834	0.107552	BR5
E01033603	31	Westminster 009I	Westminster 009I	26843.034373	769.684192	POLYGON ((-0.1651940481248741 51.5223349961171...	NW1 5DF	E00023578	51.521158	-0.167708	NW1
E01003626	37	Newham 015B	Newham 015B	84191.765717	1413.516140	POLYGON ((0.05359712608222828 51.5390032900314...	E6 2AG	E00018285	51.540082	0.057697	E6
E01003989	38	Southwark 018C	Southwark 018C	242134.008988	2505.869788	POLYGON ((-0.05785501055406365 51.484316385276...	SE151RZ	E00020073	51.480043	-0.064614	SE1
E01004197	41	Tower Hamlets 002A	Tower Hamlets 002A	204102.320562	1785.985330	POLYGON ((-0.0525324737766623 51.5348034485191...	E2 9HJ	E00021170	51.533201	-0.053391	E2

02 | Crime Rate

Recorded crime rates in the UK include threats and cases where no physical violence was used so are hard to compare with international statistics. In this project all reported crime categories are included in calculations.

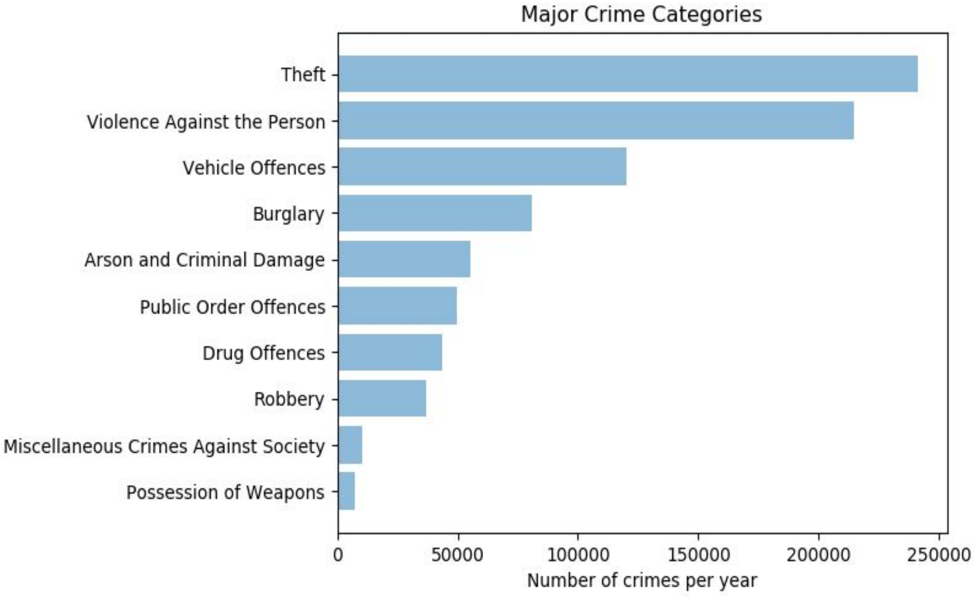
> **Source :** London data store , Greater London Authority ; Metropolitan Police Service
> **Source links :**
> **1.** https://data.london.gov.uk/dataset/recorded_crime_summary
> **2.**https://data.london.gov.uk/dataset/lsoa-atlas
> **Data details:**
> 1. LSOA Level number of crimes for the last 24 months. major crime category, minor category, LSOA, borough
> 2. LSOA level population

	LSOA	Borough	Major Category	Minor Category	201810	201811	201812	201901	201902	201903	201904	201905	201906	201907	201908	201909	Sum_Crime
0	E01000006	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	1	0	0	0	0	0	0	1
1	E01000007	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	0	0	0	0	0
2	E01000008	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	0	0	0	0	0
3	E01000009	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	0	0	0	0	0
4	E01000010	Barking and Dagenham	Arson and Criminal Damage	Arson	1	0	0	1	0	0	0	1	0	0	0	0	3
5	E01000012	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	0	0	0	0	0
6	E01000013	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	0	0	0	0	0
7	E01000015	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	0	0	0	0	0
8	E01000016	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	1	1	0	0	0	0	0	2
9	E01000021	Barking and Dagenham	Arson and Criminal Damage	Arson	0	0	0	0	0	0	0	0	1	0	0	0	1

The most frequent Minor Crime Categories:

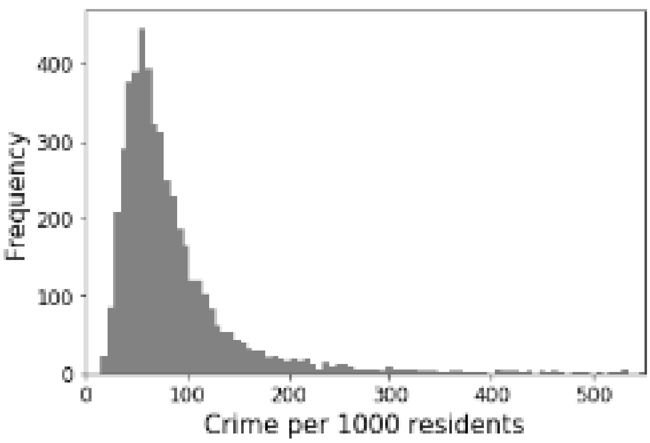
	Minor Category	Sum_Crime
46	Burglary - Residential	60107
47	Theft from a Motor Vehicle	73203
48	Violence with Injury	76480
49	Other Theft	126827
50	Violence without Injury	137972

Majority of the crimes are in three categories: theft, violence against person, vehicle offences.

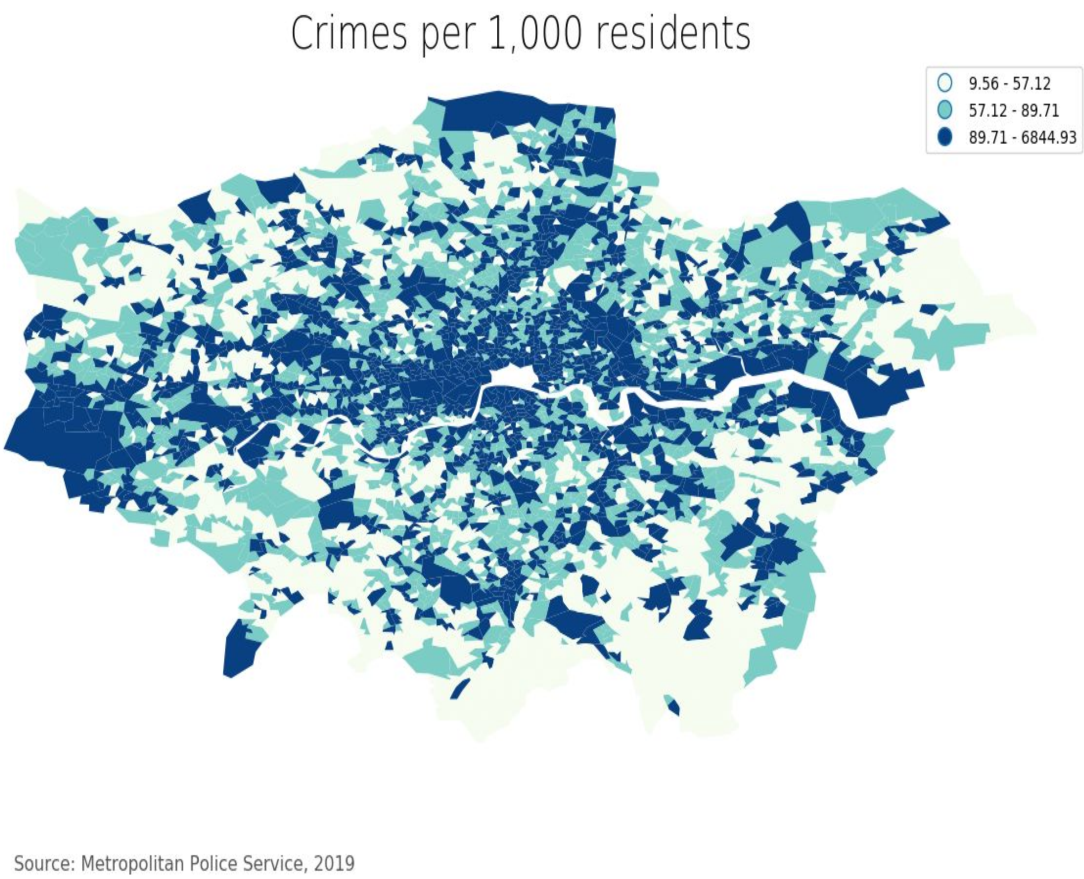


Most maps using count of reports of crime end up being maps of the population density of cities. In this project, crime rate is calculated by normalizing crime count by population

$$\left[\frac{\text{Total number of crimes}}{\text{Total Population} \div 1{,}000} \right]$$



UK crime figures are available nationally at borough and LSOA level broken down by category. Crime index is calculated by using last 12 months crime figures between Oct 2018 and Sep 2019.

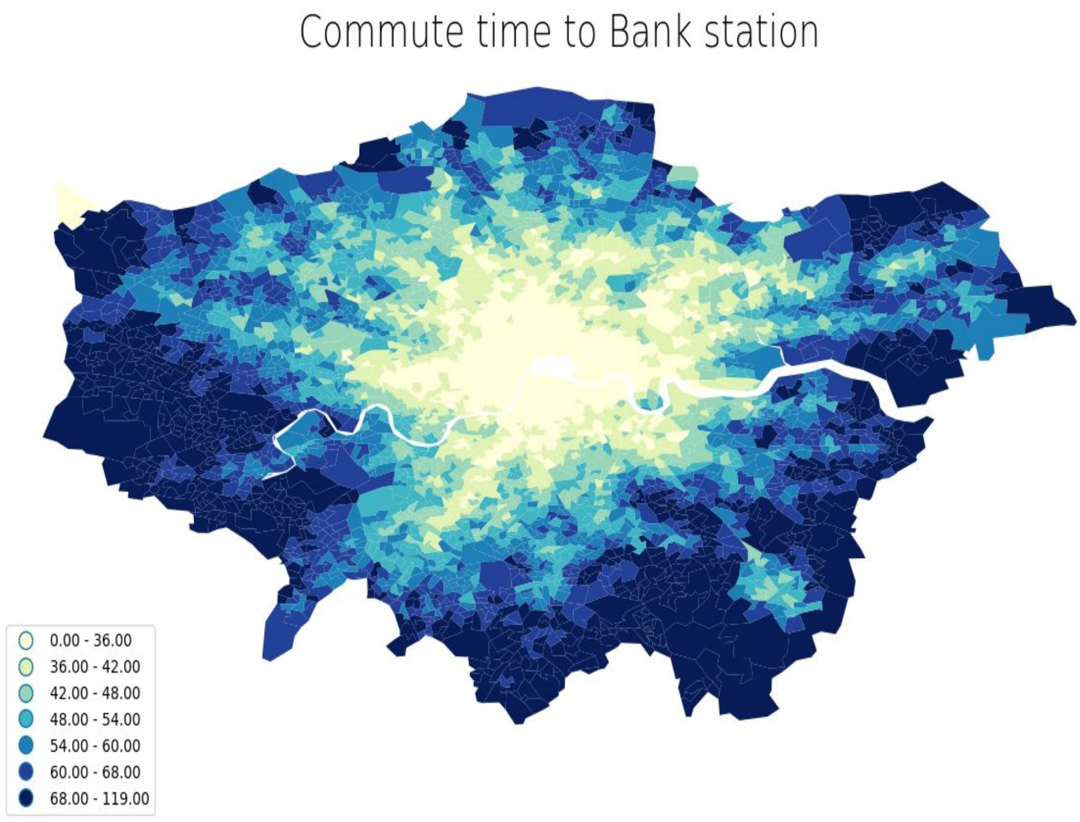


03 | Commute Time

London has one of the largest transportation networks in the world, covering 1,569 km^2. When choosing a location to live, it's important to use the transport network to your advantage.

Source : London data store , Transport for London
Source link :
https://data.london.gov.uk/download/mylondon/c2e9ebc1-935b-460c-9361-293398d84fe5/MyLondon_traveltime_to_Bank_station_OA.csv
Data details: Commuting duration to Bank station per each LSOA

Bank station is located in the centre of the city and one of the busiest underground stations located in central London.



Source: Transport for London, 2019

04 | Rental Prices

Market conditions can fluctuate greatly across London's many localised and specialist markets. Rents can of course vary depending on lots of factors, including the location and number of bedrooms. Take for example the average rents in Croydon and Westminster – at opposite ends of the spectrum – of £1,136 and £2,526 respectively

Source : London data store , Valuation Office Agency

Source link :
https://www.gov.uk/government/publications/private-rental-market-in-london-october-2018-to-september-2019
Data details: Data has price categories based on room number: studio,1,2,3,4+.

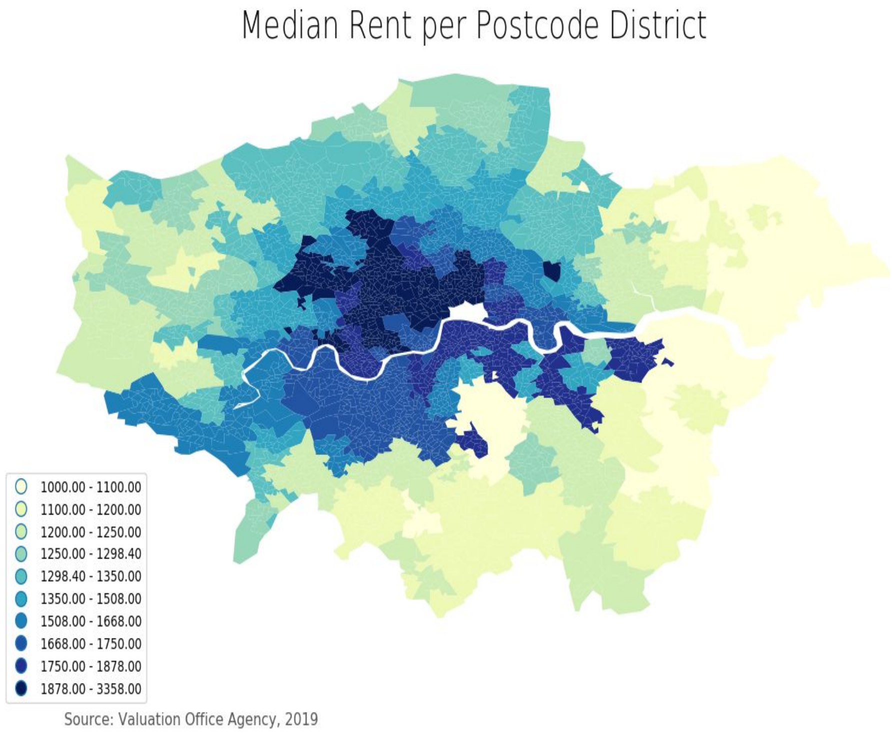
These are the postcodes with the highest average monthly rent prices in London for two bedroom properties.

	Postcode District	Median_Rent
185	W1U	3055
186	W8	3142
187	W1H	3142
188	SW3	3250
189	W1G	3358

These are the postcodes with the lowest average monthly rent prices in London for two bedroom properties.

	Postcode District	Median_Rent
0	DA8	1000
1	DA7	1000
2	DA1	1000
3	RM5	1023
4	SE2	1050

The data on average two bedroom private property rents is given at postcode district level (SW19 or E7, for example), and is based on a sample covering the last 12 months.



05 | Nearby Parks

I have used Foursquare places API to calculate the number of parks for each Postal district.

Source : Foursquare API
Foursquare Category Id:
National Park: 52e81612bcbc57f1066b7a21
State / Provincial Park: 5bae9231bedf3950379f89d0
Park: 4bf58dd8d48988d163941735
Data details: Number of parks within 3000 meters for each Postcode district

4. Methodology

For this project I am using the K-means Clustering algorithm to cluster the London neighborhoods. This is one of the simplest algorithms which is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from the unlabeled data.

The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

- The process begins with k centroids initialised at random.
- These centroids are used to assign points to its nearest cluster.
- The mean of all points within the cluster is then used to update the position of the centroids.
- The above steps are repeated until the values of the centroids stabilise.

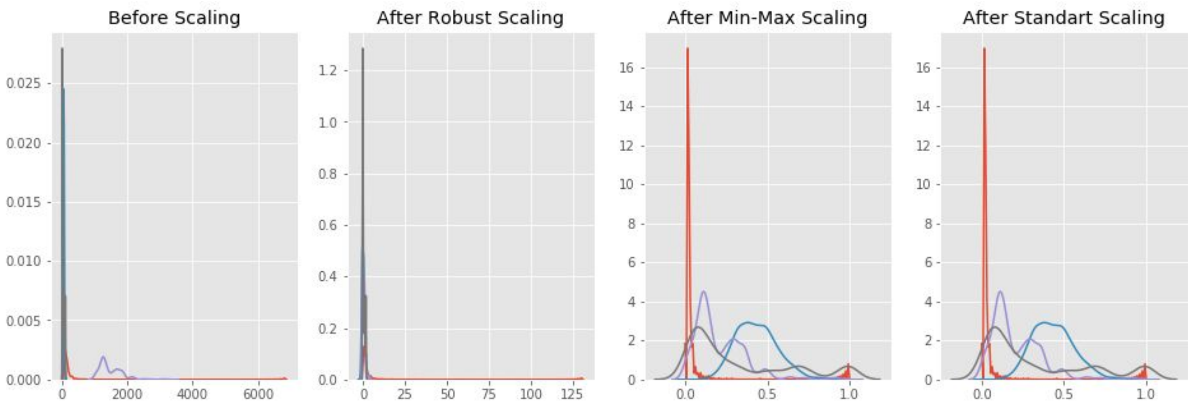
4.1 Missing value Handling

For k-means clustering any observation even with one missing dimension must be specially handled. While prepping data I have excluded missing observations

Since there are only few observations with missing values, I have excluded these observations from clustering

4.1 Normalizing data

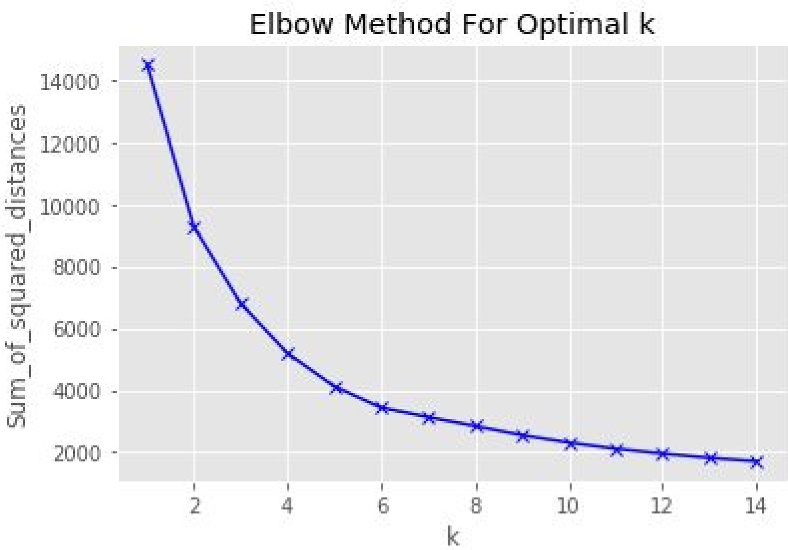
K-Means is sensitive to outliers, so as a first step I have used standard scaling to normalize my data.



4.2 Deciding on K

The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all clusters.

As k increases, the sum of squared distance tends to zero. Below is a plot of sum of squared distances for k in the range specified above. If the plot looks like an arm, then the elbow on the arm is optimal k,



The elbow chart for the dataset does not have a clear elbow. Instead, we see a fairly smooth curve, and it's unclear what is the best value of k to choose.

Before reevaluating whether clustering is the right thing to do on my data, I have tried a different method for determining the optimal k :silhouette score.

The range of the Silhouette value is between +1 and -1. A high value is desirable and indicates that the point is placed in the correct cluster. If many points have a negative Silhouette value, it may indicate that we have created too many or too few clusters.

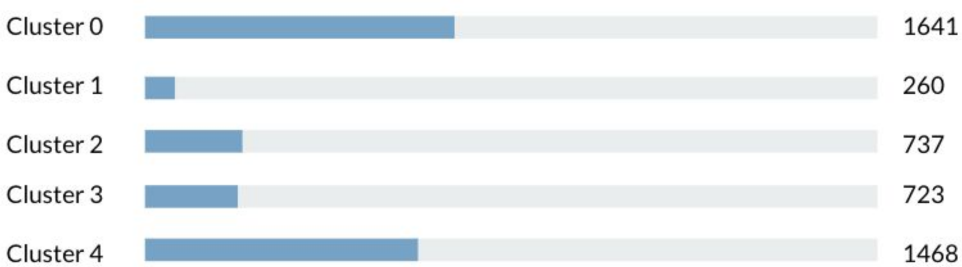
The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b).

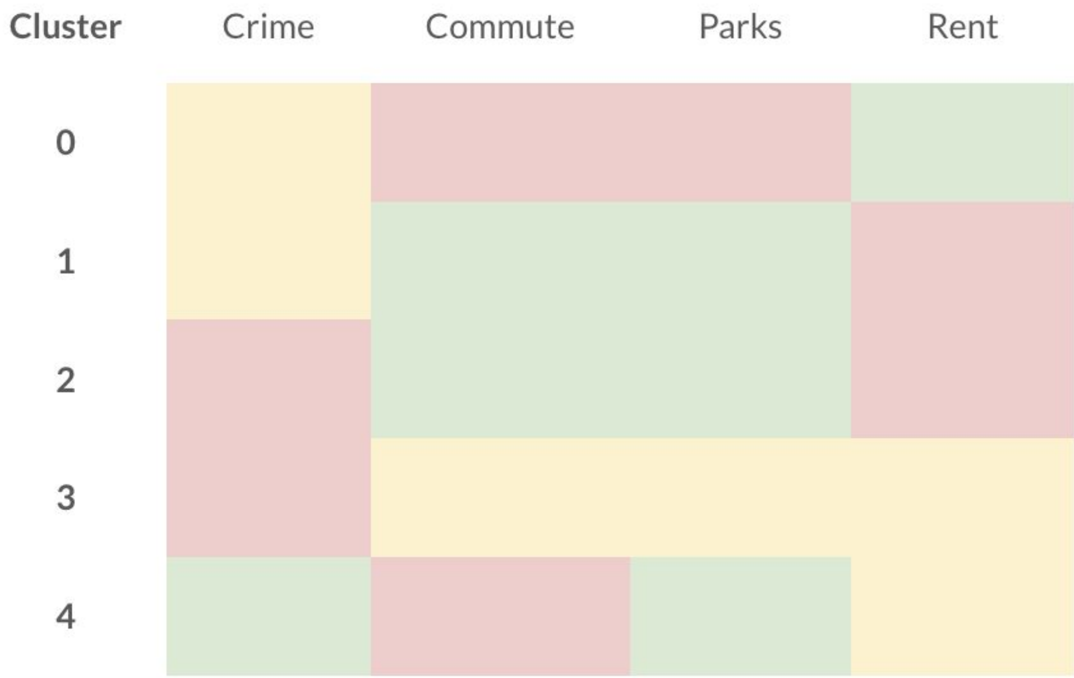
```
For n_clusters=2, The Silhouette Coefficient is 0.3488809848407165
For n_clusters=3, The Silhouette Coefficient is 0.36647887082724245
For n_clusters=4, The Silhouette Coefficient is 0.36097861107087464
For n_clusters=5, The Silhouette Coefficient is 0.3784644277627227
For n_clusters=6, The Silhouette Coefficient is 0.3507493220093311
For n_clusters=7, The Silhouette Coefficient is 0.3443922200983591
For n_clusters=8, The Silhouette Coefficient is 0.3195952305628103
For n_clusters=9, The Silhouette Coefficient is 0.32055082062108536
For n_clusters=10, The Silhouette Coefficient is 0.31860128259604914
For n_clusters=11, The Silhouette Coefficient is 0.3316256614361113
For n_clusters=12, The Silhouette Coefficient is 0.34261855779640193
For n_clusters=13, The Silhouette Coefficient is 0.3235371936639249
For n_clusters=14, The Silhouette Coefficient is 0.33632478293116086
```

5.Results

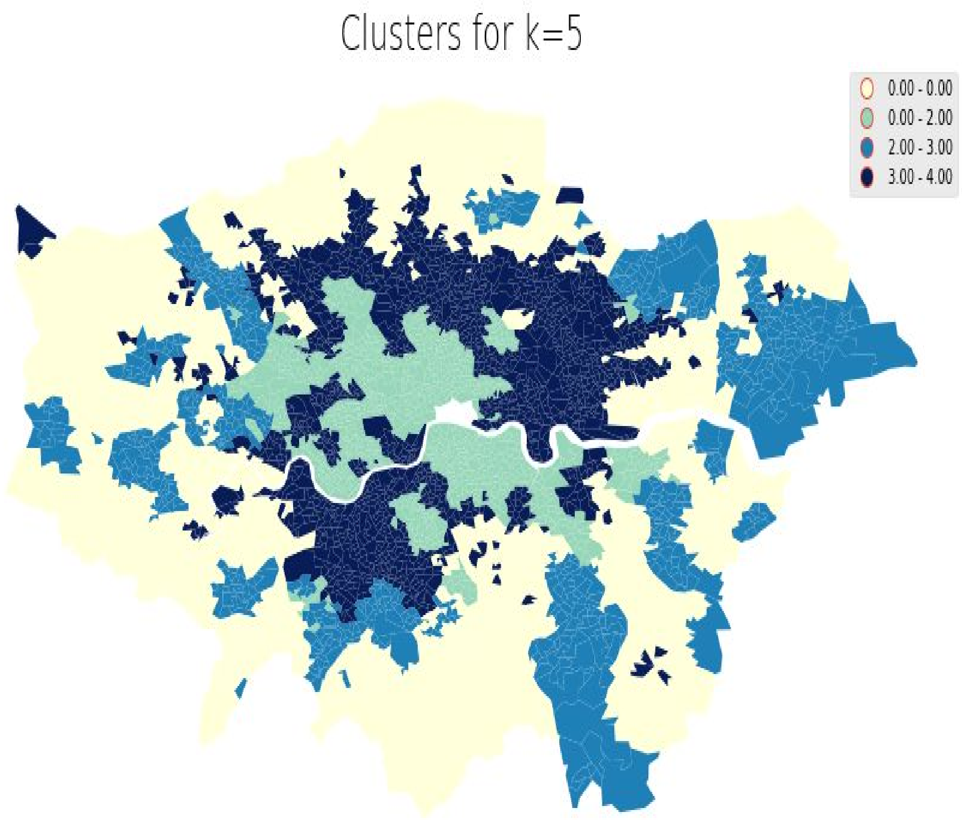
As expected there are no neighbourhoods in London that ticks all the boxes. We can eliminate cluster 1 due to low concentration of parks within the area. For Cluster 4, proximity to parks should be considered, because area does not have enough parks to cover walking distances.

The remaining clusters 2-3-5 is dog friendly, but in terms of commute duration, rent and crime they differentiate.





Before choosing a neighbourhood users will need to decide on their priorities and budget. For a short commute time you will need to pay more and deal with higher crime rates.



Best neighbourhoods to live for the below listed criteria:

- Dog friendly
- Commute time to Bank station < 45
- Crime low
- 2 bed rent price <2000

	Postcode	Cluster_Labels	public_transport_time_mins	crime_per_1000	Median_Rent	Total_venues
0	IG2 6QN	3	42.0	48.686739	1263.0	60
1	IG2 6EJ	3	41.0	46.562386	1263.0	60
2	HA9 8PN	2	44.0	32.540676	1400.0	100
3	SE8 4HU	2	43.0	31.683168	1500.0	100
4	SE8 3AN	2	41.0	50.779286	1500.0	100
5	SE5 8UU	2	43.0	45.278137	1500.0	80
6	E5 9TU	2	39.0	43.619792	1650.0	56
7	E5 8JR	2	36.0	47.890536	1650.0	56
8	N7 8GF	2	29.0	20.995334	1712.0	71
9	N7 0JP	2	33.0	48.531290	1712.0	71

Top 10 worst places to live with a dog due to high crime rate and rent prices

	Postcode	Cluster_Labels	public_transport_time_mins	crime_per_1000	Median_Rent	Total_venues
0	IG2 6QN	3	42.0	48.686739	1263.0	60
1	IG2 6EJ	3	41.0	46.562386	1263.0	60
2	HA9 8PN	2	44.0	32.540676	1400.0	100
3	SE8 4HU	2	43.0	31.683168	1500.0	100
4	SE8 3AN	2	41.0	50.779286	1500.0	100
5	SE5 8UU	2	43.0	45.278137	1500.0	80
6	E5 9TU	2	39.0	43.619792	1650.0	56
7	E5 8JR	2	36.0	47.890536	1650.0	56
8	N7 8GF	2	29.0	20.995334	1712.0	71
9	N7 0JP	2	33.0	48.531290	1712.0	71

6.Further Analysis & Improvement Ideas

A. Dog friendly venues for analysis.

Dogs are welcome London's parks and green spaces, public transportation and pubs but your chances of enjoying a restaurant with your dog is quite low. As a further step, dog friendly restaurants , veterinarian, and doggy day care details can be added.

- Veterinarian: 4d954af4a243a5684765b473
- Pet Service: 5032897c91d4c4b30a586d69
- Pet Store: 4bf58dd8d48988d100951735
- Pet Café: 56aa371be4b08b9a8d573508
- Dog Run: 4bf58dd8d48988d1e5941735
- Veterinarian: 4d954af4a243a5684765b473

B. Calculating Nearby parks in LSOA level:

In this project due to daily limits for Foursquare API calls, I have calculated the number of parks within 3000 m and on Postcode district level. Ideally, to consider walking distances query should be limited to 1000 meters for each LSOA.

C. Add tube line category:

After a few months in London I realised the importance of underground lines when choosing your neighbourhood. Not all underground lines are equal. Some lines are hotter than others, some lines have older trains than others, and some lines are busier than others. Also tube availability/disruptions data would be interesting to work on.

D. Depth analysis in crime data:

Downsides of my crime calculation is counting each crime without considering crime type. For further analysis data can be remodelled according to crime types e.g. violent crimes. Also normalizing data using daily visitors , especially for touristic areas like central London would improve clustering results.