Dog Friendly London Neighbourhoods

Buse Uluask

Oct 2019

TOC

Overview

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.

Target audience

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

Data Sources

01 | Neighborhoods

London data store - Greater London Authority

Office for National Statistics (ONS)

02 | Crime rate

Metropolitan Police Service

London data store - Greater London Authority

03 | Commute time to Bank station

Transport for London

04 | 2 bedroom rental prices

Valuation Office Agency

05 | Nearby Parks

Foursquare Places API

Neighborhoods

01

There are 33 boroughs of London. These boroughs are then further broken down into what are known as ‘postcodes.’

Each postcode begins with one or two letters, followed by a one or two digit number. This part is also called ‘postcode district’ and identifies a particular area within a borough.

Neighborhoods

01

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.

LSOAs in London

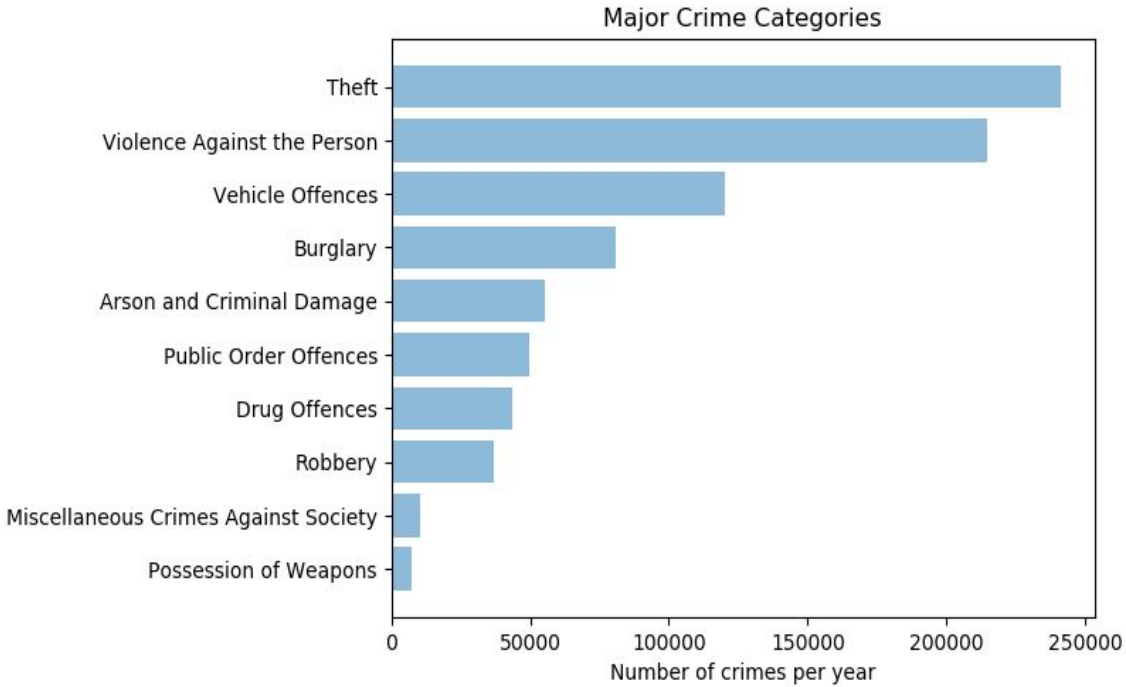
Source: Office for National Statistics, 2019

Crime Rate

02

Recorded crime rates in the UK include threats and cases where no physical violence was used so are hard to compare with international statistics.

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



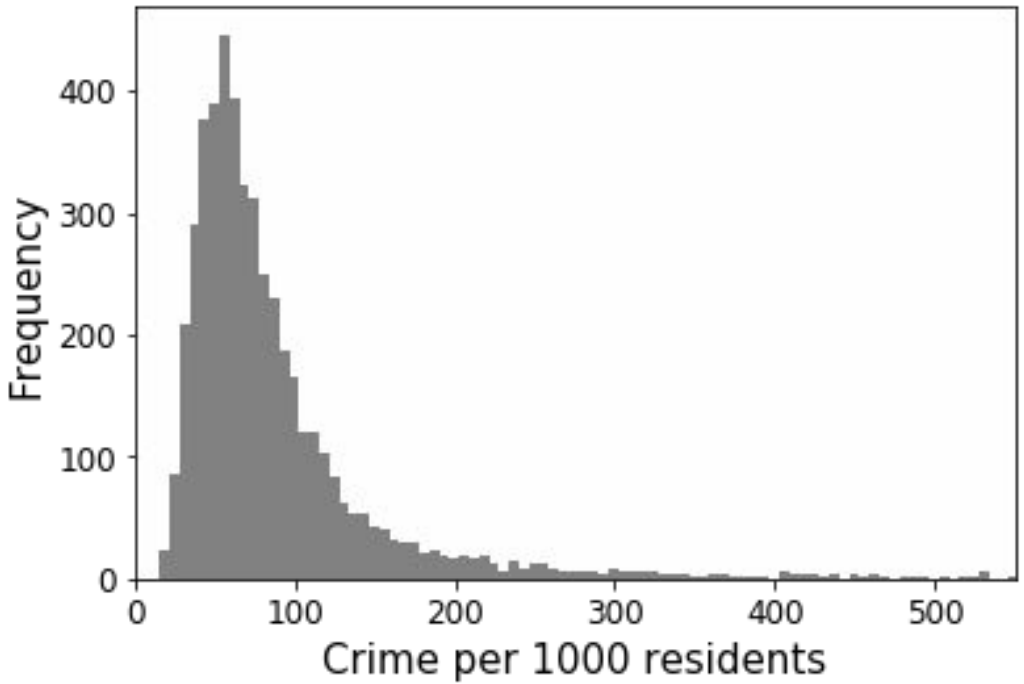
Crime Rate

02

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

$$\frac{\text{Total number of crimes}}{\text{Total Population} \div 1{,}000}$$

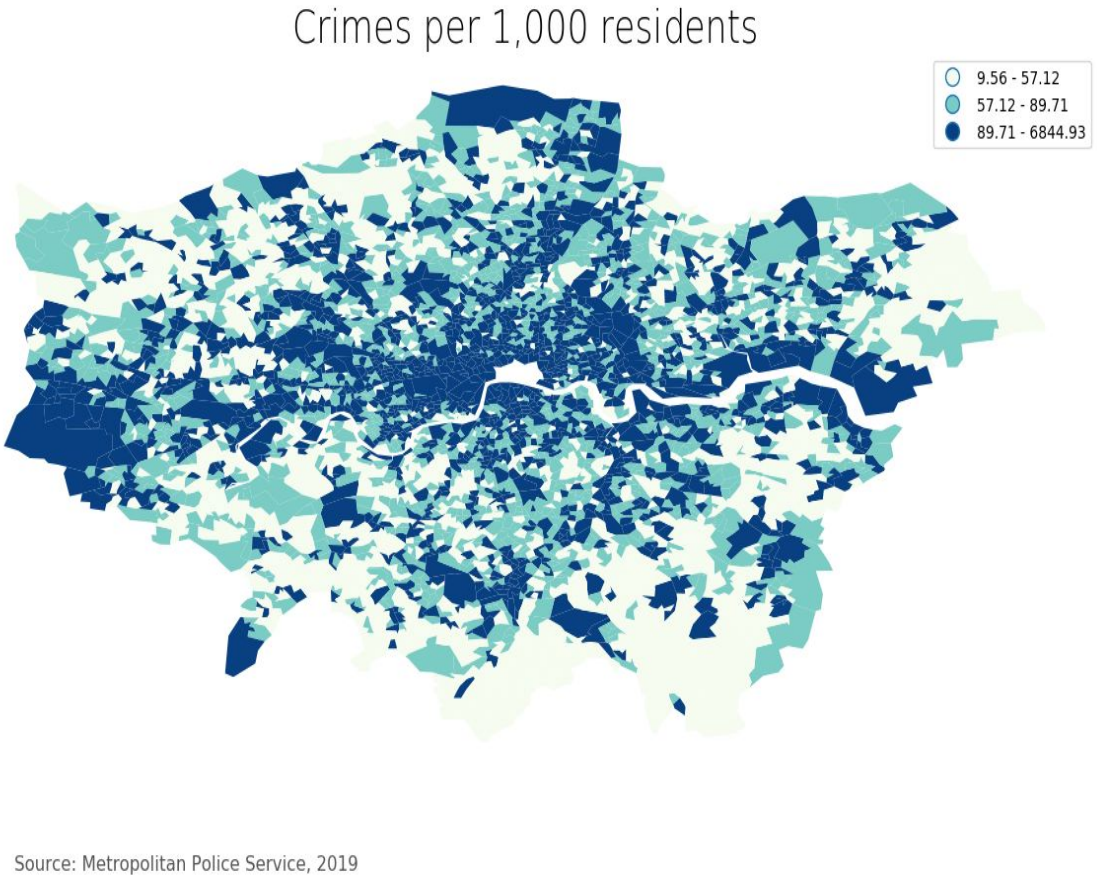


Crime Rate

02

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.



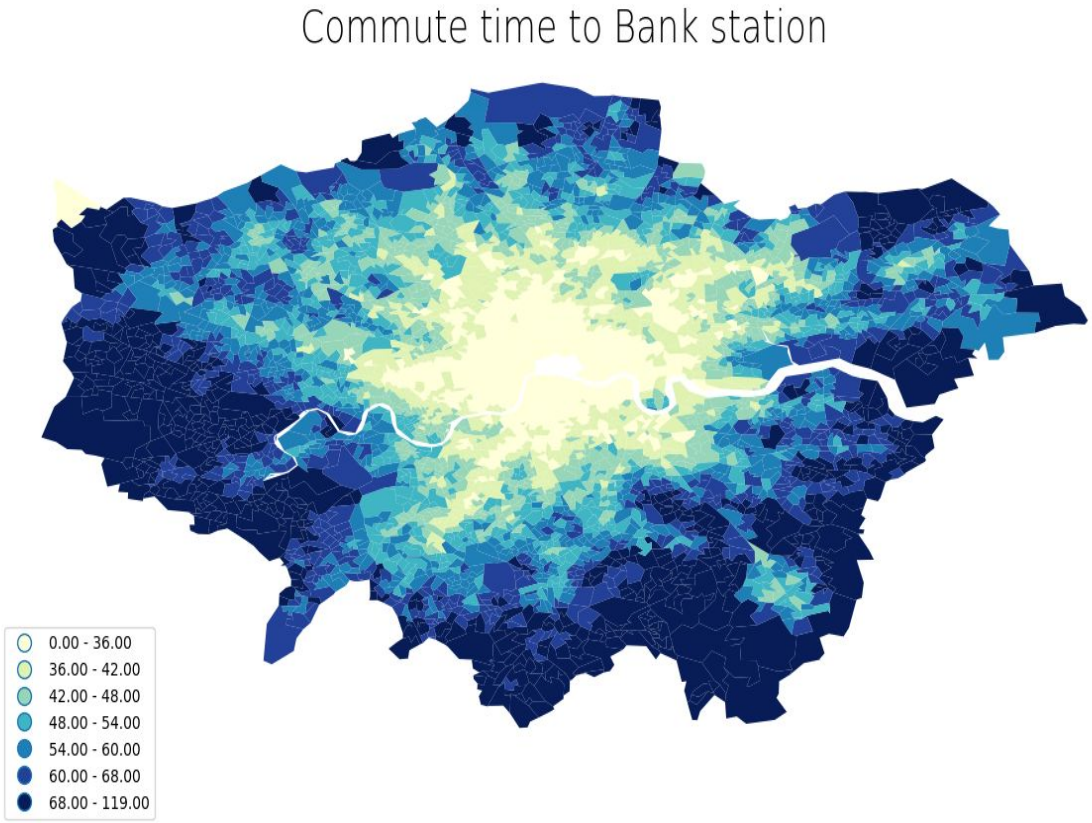
Source: Metropolitan Police Service, 2019

Commute

03

London has one of the largest transport networks in the world, covering 1,569 km². When choosing a location to live, it's important to use the transport network to your advantage.

Bank station is one of the busiest underground station located in central London.



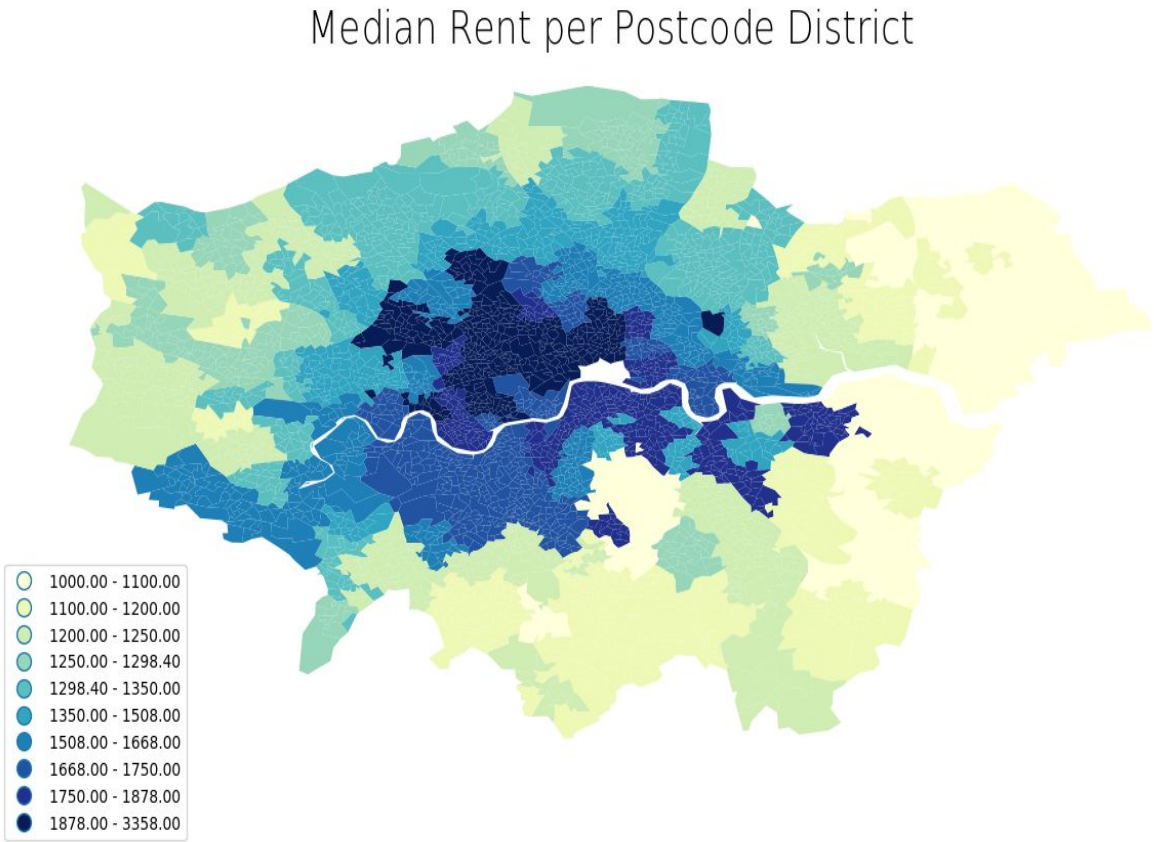
Source: Transport for London, 2019

Rental prices

04

Two bedroom properties have a varied audience and price.

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.

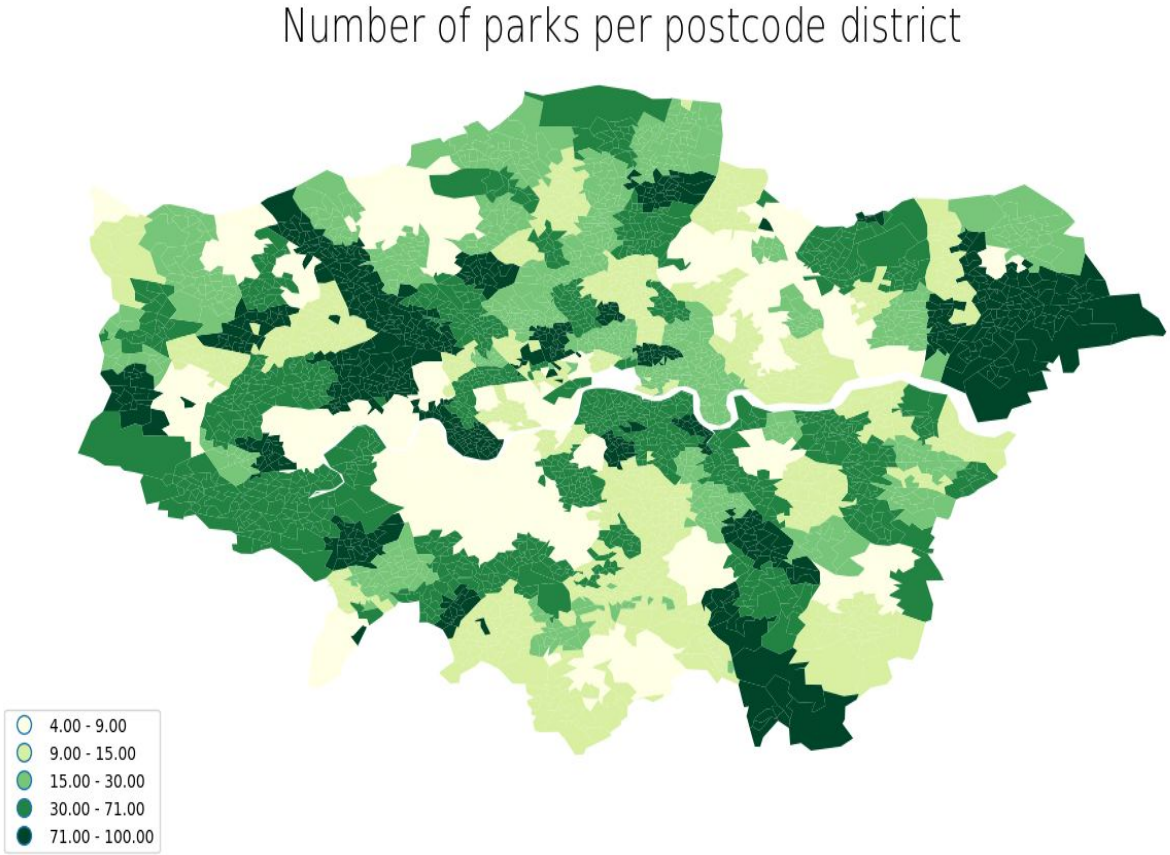


Source: Valuation Office Agency, 2019

Parks

05

Despite London is made of 40% public green space, including 3,000 parks and totaling 35,000 acres. Some postcode districts has too little green space compared to London average.



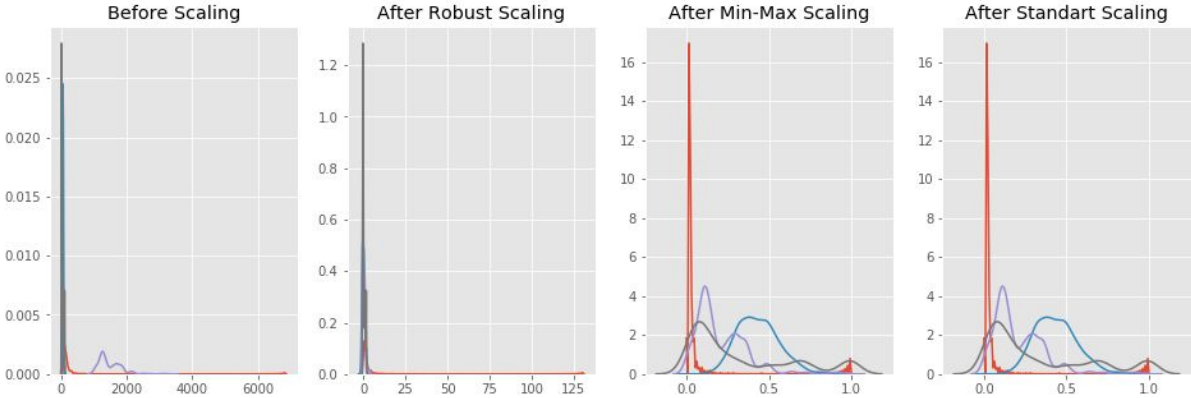
Source: Foursquare API, 2019

Methodology

01

For this project I am using the K-means Clustering algorithm to cluster the London neighborhoods. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K.

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



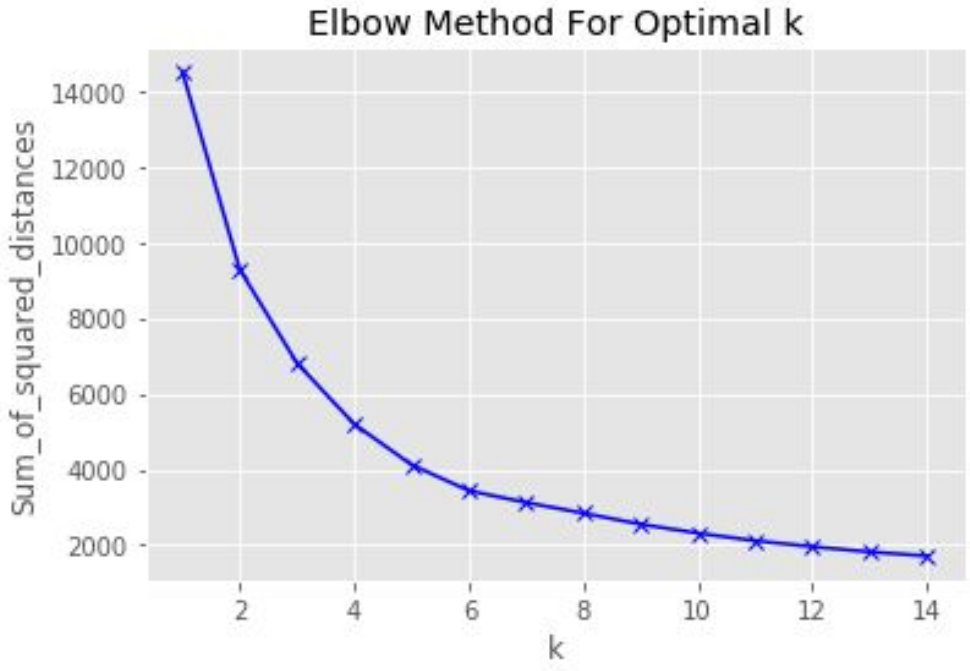
Methodology

02

After normalizing the data, elbow method is used to identify the best value of k.

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

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.



Methodology

03

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.

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
```

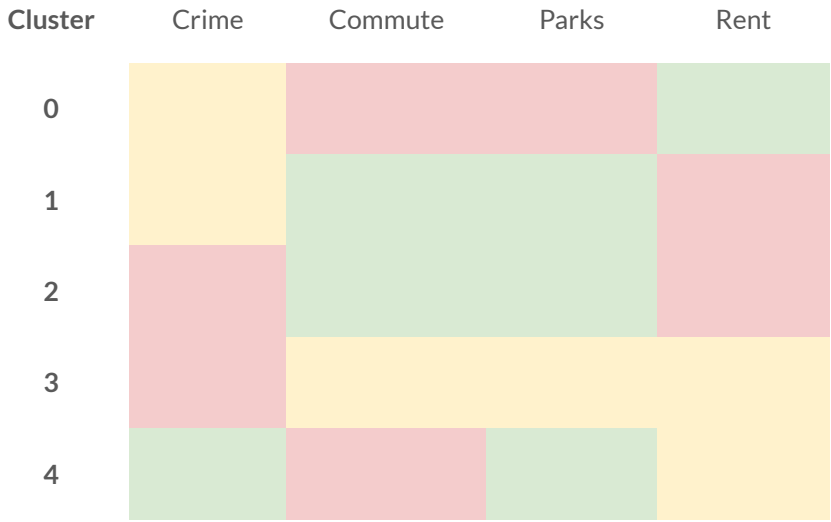
Results

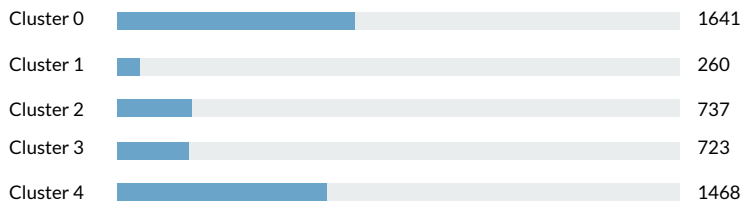
01

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.



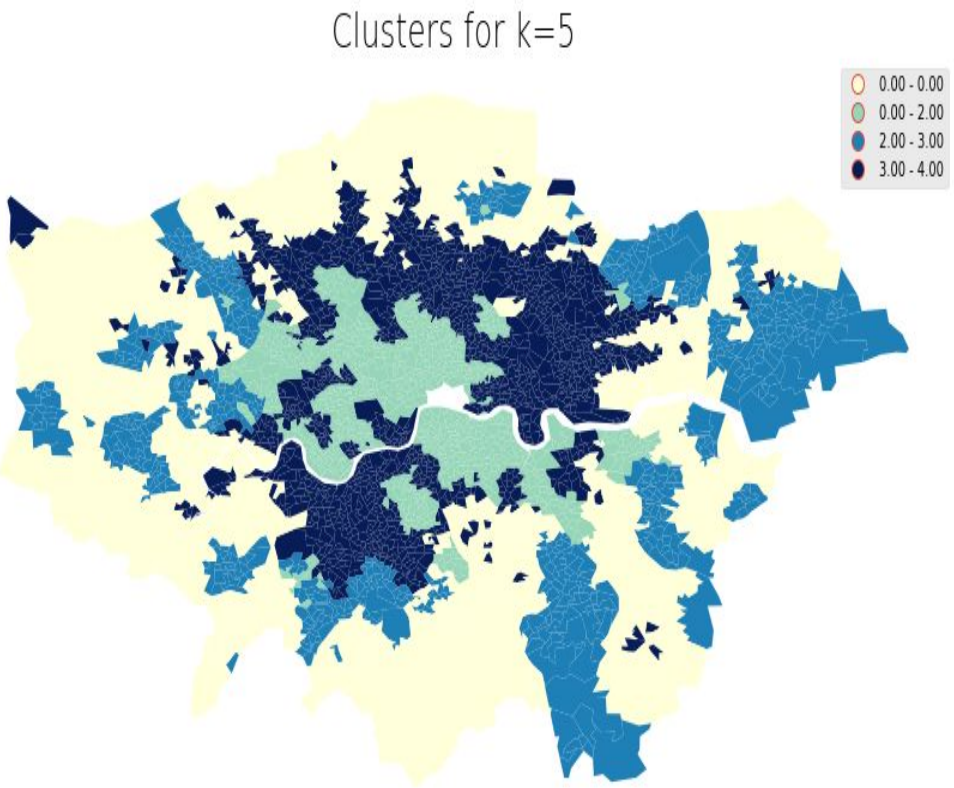


Results

02

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.



Results

03

Best neighbourhoods for the below listed criterias:

- 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

Results

03

Top 10 worst places to live with a dog:

- High crime rate
- High rent prices
- Not dog friendly

	Postcode	Cluster_Labels	public_transport_time_mins	crime_per_1000	Median_Rent	Total_venues
0	W1R 5LG	1	26.0	6844.928751	2536.0	13
1	WC2N4HY	1	19.0	5032.723239	2536.0	31
2	W1A 0ZA	1	25.0	3524.840764	2536.0	10
3	WC2N5HX	1	18.0	3128.979144	2536.0	31
4	WC2E7QZ	1	16.0	1935.302391	2536.0	31
5	W1F 0ET	1	27.0	1714.285714	2536.0	79
6	EC2A3SL	1	20.0	1409.948542	1950.0	40
7	W1H 0EN	1	31.0	1364.250338	3142.0	10
8	W1P 0JH	1	21.0	1182.452063	2536.0	22
9	W1X 7LF	1	28.0	974.876847	2536.0	8

Improvement Ideas

01 | Adding more dog friendly venues to 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.

02 | 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.

03 | 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.

04 | 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

Thank you.