

London Neighborhood for Expatriates

Coursera Applied Data Science Capstone Project

1 Introduction

This project aims to help expatriates decide which neighborhood to stay when they arrive in London.

1.1 Background

London is a global city and one of the world's largest financial centers. More than one-fifth of the United Kingdom's biggest companies are based in Central London and this provides several job opportunities. The capital city also has a flourishing art scene. It is the world's third largest film production center and one of four fashion capitals. As a cosmopolitan city, London has a diverse group of inhabitants and cultures. This makes it an attractive destination for expatriates.

1.2 Problem

One of the most important yet stressful aspects of moving to another country is deciding on accommodation. It can be difficult to start looking for a house in a new country, especially when one has not been there before. This project thus aims to guide expatriates on where they can start looking for a suitable accommodation.

1.3 Interest

This project will be of interest to young expatriates who are looking to move to London. It will be applicable to groups who prefer a quieter lifestyle and those who prefer a livelier environment.

2 Data

The data used in this project were extracted from multiple sources. This includes the following:

2.1 Geographical Coordinates of London

The data was taken from geographical coordinates of each London borough were taken from https://en.wikipedia.org/wiki/List_of_London_boroughs. This was used to map the location of each borough and create a visual representation of how far each neighbourhood was from the city centre.

2.2 Foursquare Data

Foursquare was used to extract the most common venues in each neighborhood. The top 10 venues in each neighborhood were then compared. This will give an overview of the type of amenities one can expect to have when staying there.

For example, the data could show an area with more hotels in the neighborhood. This might not be an ideal location for someone who wants to settle down in London. However, an area with more grocery stores and recreational sites could provide a better living environment. This would make it a more comfortable neighborhood to live in.

2.3 London Borough Profiles

There are 33 boroughs in London and each borough has information on the crime rates per 1,000 population, median house prices, and average public transport accessibility score. The data was taken from the London Database (<https://data.london.gov.uk/>). Where the crime rates were unavailable on the London Database, they were taken from <https://www.statista.com/statistics/380963/london-crime-rate/>.

The cost of a house is a primary consideration when house-hunting. Therefore, the median house price of each neighbourhood would give a good indication of affordability. As the safety and accessibility of an area are also key factors in choosing a neighbourhood, the information would help to narrow down an ideal area to live in.

3 Methodology

The machine learning technique of k-means clustering was used on the data, and the silhouette score was used to find the optimum value of k.

3.1 Exploratory data analysis

The neighbourhoods were plotted on the map of London for a visual representation of where they were located. This might give a spatial indication of any patterns in clusters that would emerge later.

Location of London Neighbourhoods



The venues within a 1000m of each neighbourhood was then explored using Foursquare and grouped into categories. This would give an indication of the type of amenities one can expect and the general environment of the neighbourhood.

Snapshot of the Top 10 Most Common Venues in Neighbourhoods

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
Barking and Dagenham	Bus Stop	Pool	Gym / Fitness Center	Gas Station	Soccer Field	Golf Course	Supermarket	Turkish Restaurant	Diner	Grocery Store
Barnet	Park	Pub	Bus Stop	Gym	Café	Fast Food Restaurant	Falafel Restaurant	Farm	Farmers Market	Yoga Studio
Bexley	Pub	Clothing Store	Fast Food Restaurant	Supermarket	Coffee Shop	Hotel	American Restaurant	Pharmacy	Italian Restaurant	Chinese Restaurant
Brent	Hotel	Coffee Shop	Bar	Sporting Goods Shop	Clothing Store	Grocery Store	Sandwich Place	Indian Restaurant	Warehouse Store	Pizza Place
Bromley	Pub	Clothing Store	Coffee Shop	Supermarket	Indian Restaurant	Gym / Fitness Center	Café	Park	Pizza Place	Bar
Camden	Coffee Shop	Hotel	Pub	Café	Sandwich Place	Burger Joint	Pizza Place	Breakfast Spot	Bookstore	Plaza
City of London	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Scenic Lookout	Hotel	Plaza	Boxing Gym	Seafood Restaurant	Steakhouse	French Restaurant
Croydon	Coffee Shop	Pub	Clothing Store	Hotel	Indian Restaurant	Italian Restaurant	Mediterranean Restaurant	Bookstore	Portuguese Restaurant	Sushi Restaurant

3.2 K-Means Clustering

K-means clustering aggregates values together based on similarities between the data points. The value of k determines the number of centroids to be defined in the dataset, and thus the number of resultant clusters.

This machine learning technique was used to group the neighbourhoods and show the distinct characteristics of each cluster. For example, clusters with neighbourhoods that have pubs as one of the most common venues could have a livelier nightlife. On the other hand, a cluster with neighbourhoods that have parks as one of the most common venues could be more relaxed. Choosing a cluster based on such indicators would then help to narrow down a neighbourhood which is more suited for one's lifestyle.

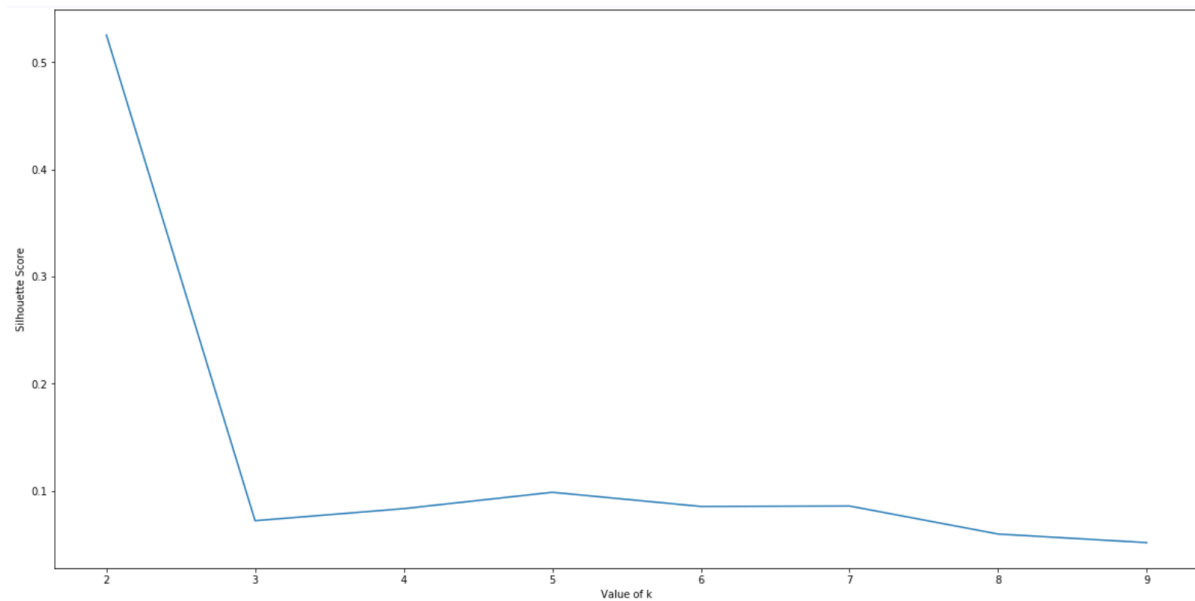
K-means clustering was an ideal technique to use in this situation as the characteristics of the groups were not defined. Unsupervised learning on unlabelled data thus had to be done and k-means clustering was able to do so.

3.3 Silhouette Score

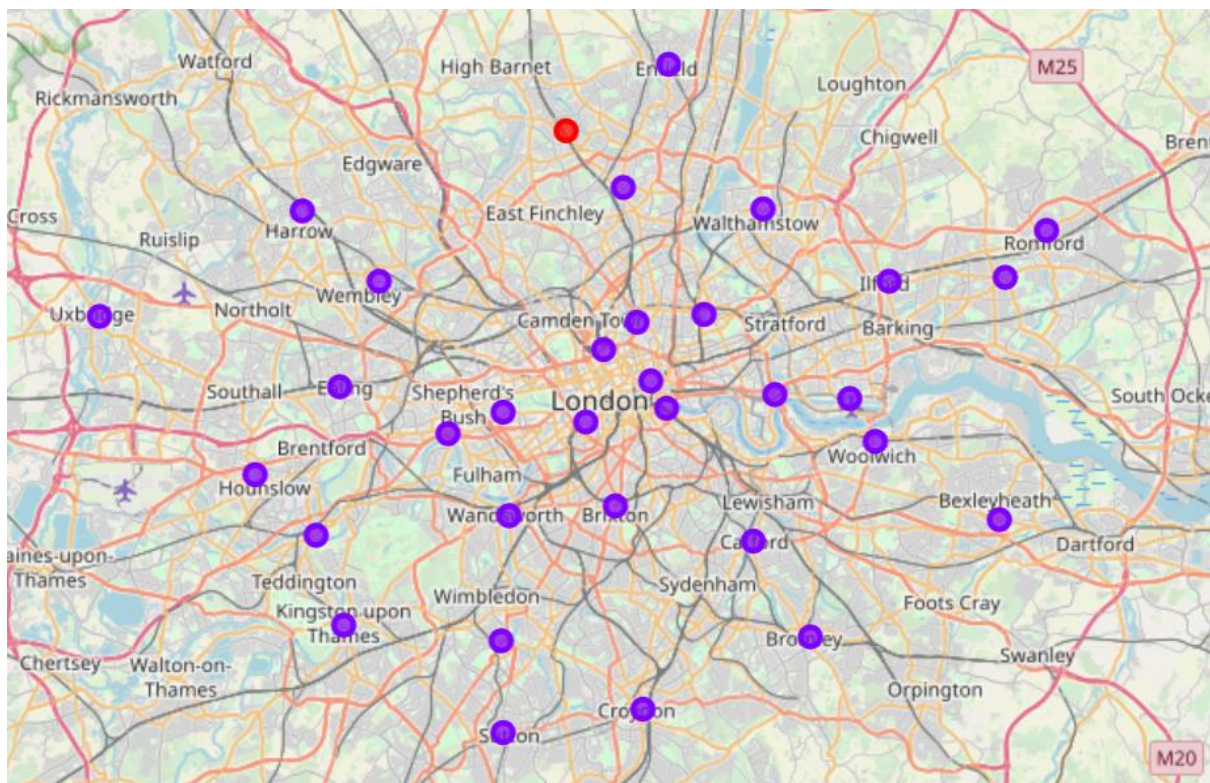
The silhouette score was used to determine the optimum value of k to use for k -means clustering. This measures the mean intra-cluster distance to determine how similar values are in the cluster, and the mean inter-cluster distance to determine how different a cluster is from another.

Based on the silhouette score, the most optimum value of k was 2. The neighbourhoods were thus split into 2 clusters.

Graph of silhouette score against value of k



London Neighbourhoods grouped into 2 clusters



3.4 Sorting by crime rates, house price and transport accessibility

The map above shows that one cluster has only one neighbourhood. For the other cluster, the next step is to choose a specific neighbourhood to live. To do so, more information about each neighbourhood should be considered. This includes crimes rates which would indicate how safe a neighbourhood was. Neighbourhoods were sorted from lowest to highest crime rate. Transport accessibility would indicate how connected the neighbourhood was and the relative ease of commuting. Neighbourhoods were sorted from highest to lowest accessibility score. House prices would determine whether one could afford to live in the neighbourhood. Neighbourhoods were sorted from lowest to highest price.

In addition to these indicators, neighbourhoods with either hotels or pubs as one of the top two most common venues were removed. This was to ensure that the neighbourhoods would not be too noisy for potential residents.

4 Results

The following are the cluster of neighbourhoods obtained after k-means clustering with k=2. There was only one neighbourhood in Cluster 1. Cluster 2 had been sorted according to the crime rate, house price and public transport accessibility scores. This will be further discussed in the next section.

Results for Cluster 1

	Neighborhood	Crime rates per thousand population	Median House Price	Average Public Transport Accessibility score	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
2	Barnet	62.738426	445000	2.996701	Park	Pub	Bus Stop	Gym	Café	Fast Food Restaurant	Falafel Restaurant	Farm	Farmers Market	Yoga Studio

Top 5 results for Cluster 2

	Neighborhood	Crime rates per thousand population	Median House Price	Average Public Transport Accessibility score	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
14	Harrow	50.396101	396150	2.870196	Coffee Shop	Indian Restaurant	Sandwich Place	Clothing Store	Gym / Fitness Center	Fast Food Restaurant	Pharmacy	Park	Afghan Restaurant	Thai Restaurant
28	Sutton	55.856141	320000	2.854086	Clothing Store	Coffee Shop	Pub	Pizza Place	Café	Sandwich Place	Supermarket	Bar	Bakery	Italian Restaurant
20	Kingston upon Thames	58.455691	410000	2.928251	Coffee Shop	Café	Pub	Clothing Store	Italian Restaurant	Burger Joint	Thai Restaurant	Park	Department Store	Bookstore
23	Merton	59.747174	415000	3.405588	Supermarket	Café	Tram Station	Train Station	Park	Italian Restaurant	Fast Food Restaurant	Pizza Place	Bookstore	Sandwich Place
25	Redbridge	69.687656	345000	2.982147	Grocery Store	Clothing Store	Fast Food Restaurant	Supermarket	Coffee Shop	Bakery	Pub	Sandwich Place	Gym / Fitness Center	Bookstore

5 Discussion

An ideal neighbourhood will be selected for two types of expatriates: one who prefers a quieter lifestyle, and one who prefers a livelier environment.

5.1 Quiet Lifestyle

Analysing the two clusters, cluster 1 is more suited for an expatriate who prefers a quiet lifestyle. In fact, Barnet is the only neighbourhood in the cluster. The most common venue in Barnet are parks and there are also plenty of farms. Its location around green spaces make it ideal for those who want a more relaxed environment or slower pace of life. It is also easy to get food with many cafes and restaurants around. There are also farmers markets to get groceries from.

In terms of transport accessibility, Barnet has a relatively lower score compared to other neighbourhoods. However, bus stops are the third most common venue and should be able to mitigate the low score. The crime rate in the neighbourhood is low and the cost of housing is mid-range. This makes it an ideal neighbourhood for those who prefer a quiet lifestyle.

5.2 Lively Environment

Cluster 2 has more entertainment spaces and retail outlets, including pubs and clothing stores. This can be seen from the table below. This makes it a preferred group of neighbourhoods to live in for those who are looking for a lively environment to live in.

Before neighbourhoods with hotels or pubs as the first 2 most common venues were removed

	Neighborhood	Crime rates per thousand population	Median House Price	Average Public Transport Accessibility score	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	City of London	85.460000	799999	7.862302	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Scenic Lookout	Hotel	Plaza	Boxing Gym	Seafood Restaurant	Steakhouse	French Restaurant
1	Barking and Dagenham	83.359060	243500	2.970626	Bus Stop	Pool	Gym / Fitness Center	Gas Station	Soccer Field	Golf Course	Supermarket	Turkish Restaurant	Diner	Grocery Store
3	Bexley	51.827942	275000	2.552134	Pub	Clothing Store	Fast Food Restaurant	Supermarket	Coffee Shop	Hotel	American Restaurant	Pharmacy	Italian Restaurant	Chinese Restaurant
4	Brent	78.801881	407250	3.653713	Hotel	Coffee Shop	Bar	Sporting Goods Shop	Clothing Store	Grocery Store	Sandwich Place	Indian Restaurant	Warehouse Store	Pizza Place
5	Bromley	64.133577	374975	2.779859	Pub	Clothing Store	Coffee Shop	Supermarket	Indian Restaurant	Gym / Fitness Center	Café	Park	Pizza Place	Bar

Harrow and Sutton emerged as ideal choices in this cluster. Both had many clothing stores for shopping. There were also plenty of coffee shops to chill alone or hang out with friends. For day-to-day living, there were many eateries and supermarkets.

Sutton also had many pubs and bars, but these venues were not the most common in the neighbourhood. Neighbourhoods that had hotels and pubs as the top two most common venues were already removed from consideration to ensure the area was not be too busy or disruptive.

Both neighbourhoods had low crime rates and house prices, making it safe and affordable to live in.

6 Conclusion

To conclude, this project guides expatriates in choosing where to stay in London as they would be unfamiliar with the city. K-means clustering was used to group the neighbourhoods and it was found that one cluster was more suited for those who preferred a quiet lifestyle while the other suited those who preferred a livelier environment.

Barnet was ideal for those who preferred a quiet lifestyle due to the many green spaces in the area. On the other hand, Harrow and Sutton was ideal for those who preferred a livelier environment. These

neighbourhoods had many retail and entertainment spaces yet had the necessary amenities for everyday living. This included supermarkets and eateries.