Identifying Rental Opportunities in London using Machine Learning

Joel Strickland

1. Introduction

1.1. Background

London is one of the most expensive cities for renters to live in the whole of Europe. Rents can vary depending on a variety of factors, including location, number of bedrooms, and local services. Single rooms are often popular with young professionals moving to the capital, two and three bedroom properties have a varied audience, including families to sharers.

1.2. Problem

When searching for the perfect locket of London to call home, looking at rental prices and identifying the ideal location can seem very intimidating. It is common knowledge that renting a home privately in the capital can be pricey, but knowing which borough is suitable for your own personal budget can really help keep costs down.

The purpose of this project is to identify ideal rental locations in London for house sharers, a family, or young professionals. To identify the ideal rental borough and property the data (crime, rental, venue type) will be filtered using the following criteria:

- Narrowing London borough and ward locations based on a < £1200 per person rental budget,
- Narrowing London borough and ward locations by identifying neighbourhoods with relatively low crime (< 35,000 crimes in a 24 month period),
- Filtering identified borough properties using K Means clustering using rental price and access to local amenities (restaurants, pubs etc.) as the features of interest.

1.3. Interest

Who might be interested in this project?

- House sharers and families looking for a safe place to live with access to local services.
- Landlords looking to target specific demographics.
- Local services trying to better understand their target market.

2. Data Acquisition and Cleaning

2.1. Data Sources

The London borough rental statistics can be found from the gov.uk website:

- https://data.london.gov.uk/dataset/average-private-rents-borough

The London ward rental statistics can be found from the Rightmove website:

- https://www.rightmove.co.uk/

The London borough and ward crime statistics can be found from the gov.uk website:

- https://data.london.gov.uk/dataset/recorded_crime_summary

Local popular amenity data can be found by leveraging the Foursquare API:

- https://developer.foursquare.com/

2.2. Data Wrangling

In this work, the bulk of the data cleaning was performed on the wrangling of the crime and rental data tables with the purpose of identifying low-crime and low-rent boroughs of London. This was accomplished by identifying boroughs whose lower rent quartile was below £1200 and the total number of crimes committed in a 24 month period lower than 35,000. From this filtering, 5 out of 32 London boroughs were identified that met these requirements (Merton, Harrow, Kingston upon Thames, Bexley, Sutton). Following this, the location of each borough from the city centre was calculated using geo coordinate data. Bexley was found to have the best combination of low-crime, low-rent, and proximity to London city centre; hence, it was chosen for further analysis. Bexley rental data was then extracted from the Rightmove website and 163 properties were identified. The rental data was filtered so only the rental price, latitude, and longitude were extracted. If the reader would like more information in regard to the data wrangling stage of this project, then please the GitHub code:

https://github.com/joel-strickland/Capstone_Project_Battle_of_the_neighbourhoods/blob/main/Identifying%20Rental%20Opportunities%20in%20London%20using%20Machine%20Learning.ipynb

3. Exploratory Data Analysis

As demonstrated in Figure 1 below, Kingston upon Thames is the safest London borough and Bexley has the cheapest rent. This analysis does not consider the population size of each borough, which may influence the relative crime statistics. Considering the difference in crime between boroughs is small, the saving from renting in a cheaper borough may outweigh the downside of living in an area with slightly more reported crime. However, before any conclusions are made about the best rental location it is important to understand the borough distance from London city centre, since this is typically where the majority of the most well-paying jobs are.

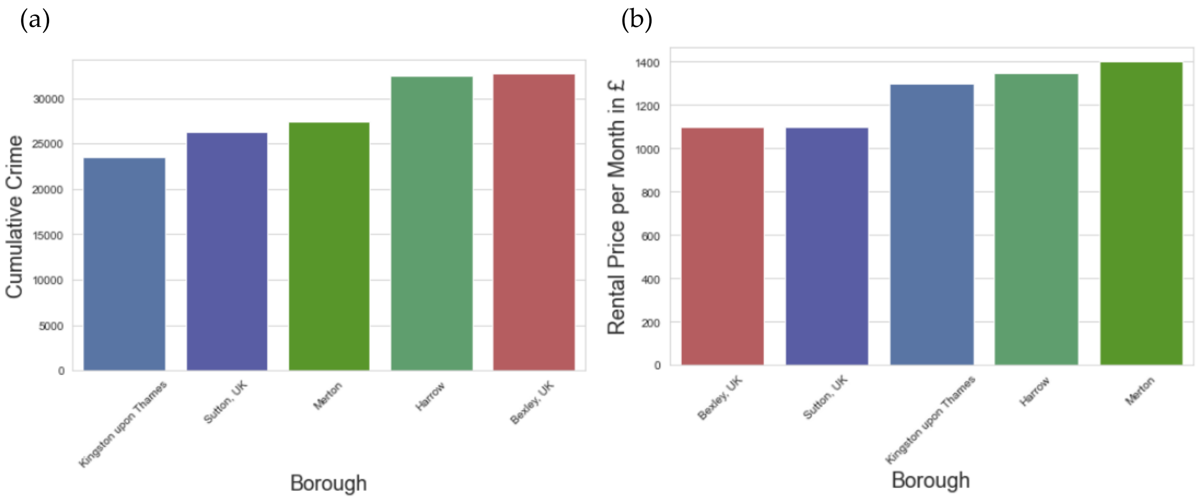

Figure 1 – **(a)** Crime level over a 24 month period for the top 5 safest low rent boroughs. **(b)** Rental price for the top 5 safest low rent boroughs.

The locations of the safest and cheapest rent boroughs within London are plotted in Figure 2(a). the boroughs are all relatively far away from the city centre ($> 11\ km - \boldsymbol{Figure}\ \mathbf{2}(c)$). There appears to a population of safe boroughs with low-rents on the south-west side of London (Sutton, Merton, Kingston upon Thames). As can be seen from Figure 2(b), there is a decreasing trend of price with increasing distance from the city centre. However, Bexley appears to be an outlier since its average rental rates are £200 below Kingston upon Thames which is the same distance from the city centre. Therefore, Bexley appears to be an ideal rental location with close proximity to London, low-crime, and low-rental rates. The location of all rental properties for Bexley (found on the Rightmove website – 10/05/2021) are indicated on Figure 2(c).

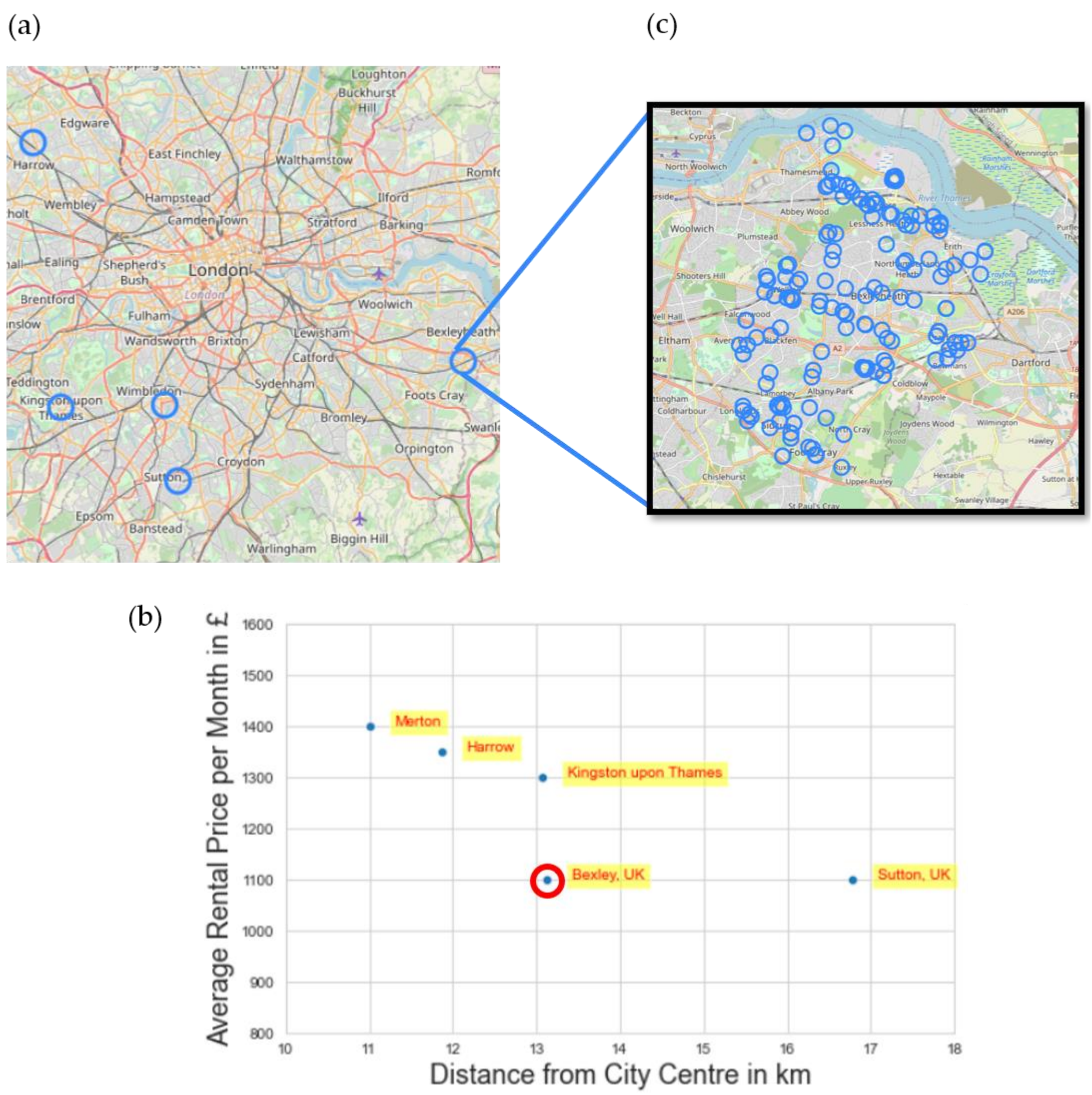

Figure 2 – (a) Identifying the locations of the low-rent low-crime borough in London. **(b)** A plot of how rental price varies with distance from the city centre, where Bexley is identified as a price outlier. **(c)** A map of rental property locations in Bexley.

4. K-Means Clustering

4.1. Cluster by Price

The rental properties in Bexley were filtered against their rental price using K-Means clustering. A more detailed filtering could have been performed including data such as, bedroom and bathroom number, however, this was outside the scope of the present article. After the first price filtering, three clusters were identified. A cheap cluster (mean = £755, σ = £184) with 55 properties, a moderately priced cluster (mean = £1297, σ = £187) with 99 properties, and an expensive cluster (mean = £2697, σ = £469) with 9 properties. The cheap rental cluster was taken forward for further analysis since all of the properties were <£1000.

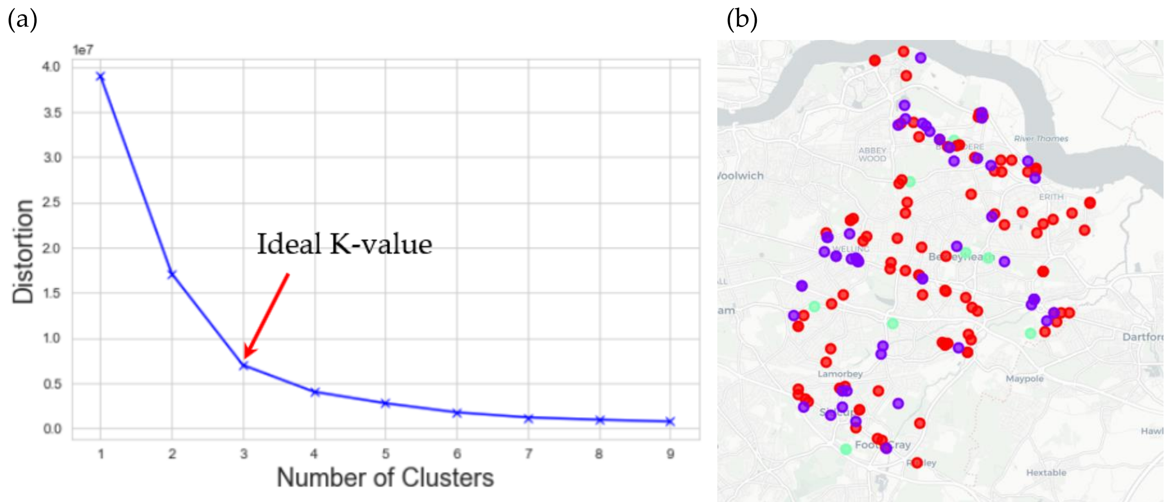

Figure 3 - (a) The elbow method showing the optimal k-value (k = 3) for the variation in observed rental prices. **(b)** Plotting the three identified clusters. Purple is cheap (mean = £755); red is moderately priced (mean = £1297); green is seriously expensive (mean = £2697).

After extracting the properties that resided in the cheap rental property cluster (Figure 3(b) – purple), a further price clustering was performed. Two sub-price clusters were identified, a really cheap cluster (mean = £581, σ = £85) with 26 properties, and a moderately cheap cluster (mean = £910, σ = £75) with 29 properties (Figure 4(b)). The 26 properties identified in the really cheap cluster were then used for further clustering analysis, however, this time the properties were clustered based on their proximity to local amenities rather than their price.

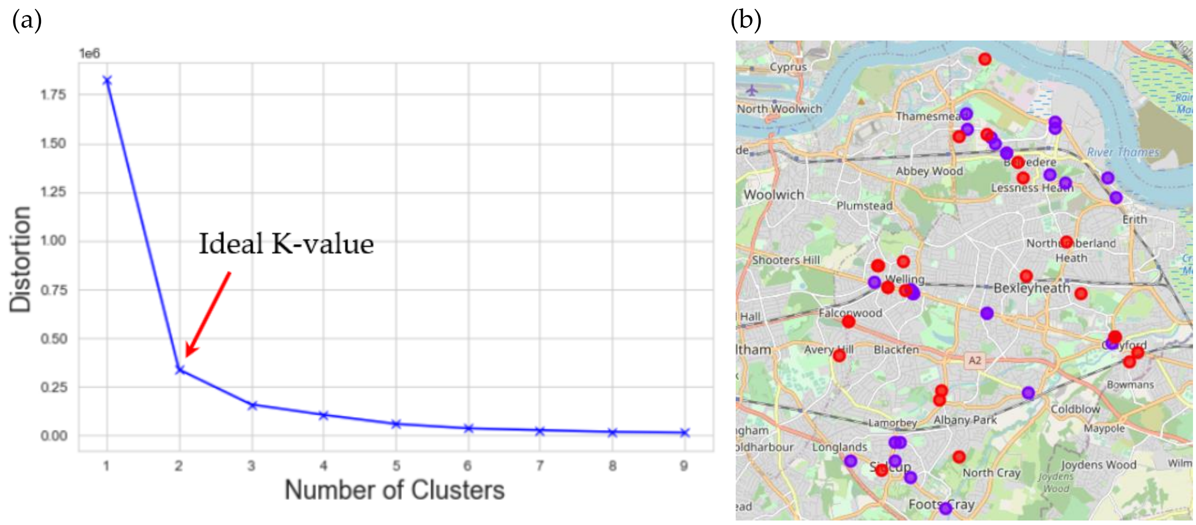

Figure 4 - (a) The elbow method showing the optimal k-value (k = 2) for the variation in observed rental prices from the cheap cluster (purple in Figure 3(b)). **(b)** Plotting the two identified sub-price clusters. Purple is moderately cheap (mean = £910); red is really cheap (mean = £581).

4.2. Cluster by Local Amenities

The Foursquare API was used to identify venues within 200m of each property found within the really cheap cluster (red circles in Figure 4(b)). Only 14 out of the 26 properties in the cluster were within a

200m distance from any local amenities on the Foursquare database. The location data indicated that 66 venues were close to the remaining 14 properties, of which there were 27 unique categories. To further split the remaining 14 properties into clusters, the venue category was chosen as the feature of interest and k-means performed on the data. A K-value of six was chosen (Figure 5(a)), however, the ideal K-value could not be determined with certainty from the data, since only 14 properties and 66 venues resided within the dataset. From post-mortem venue analysis, a K-value of six appeared to be the most appropriate since this grouped cluster zero (red) into pubs and cluster 1 (purple) into retail venues. The other clusters appeared to be more or less determined from the location of one or two unique nearby venues, such as a golf course or a nature trail.

(a) (b)

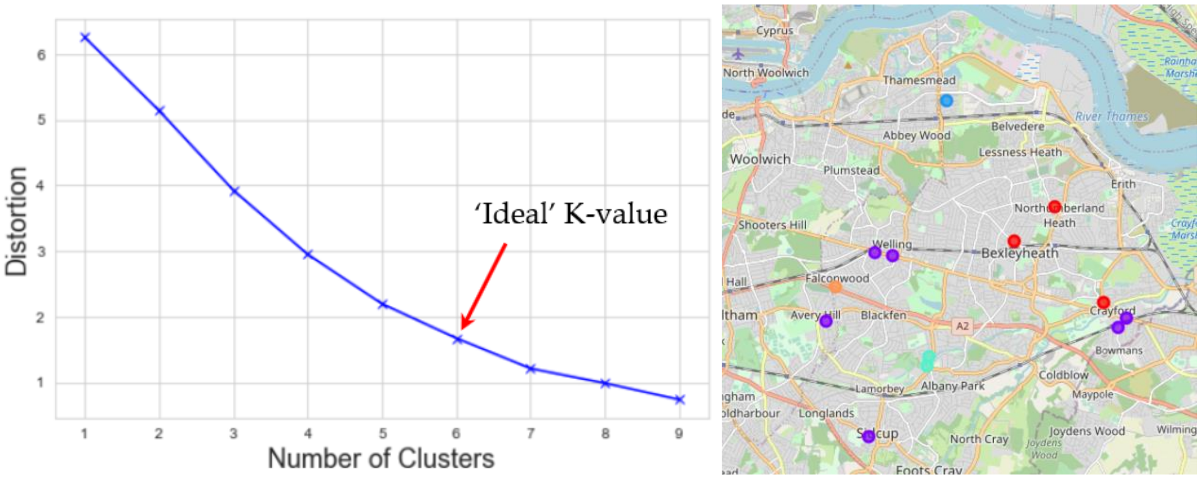

Figure 5 –(a) The elbow method showing the optimal k-value ($k \sim 6$) for the variation in venue types near the really cheap properties (red in Figure 4(b)). **(b)** Plotting the six identified venue type clusters. Purple is Highstreet venues; red is pubs; all others are clustered based on unique local venues.

Assuming shared tenancy, family, or young professional, the ideal location for a property rental would therefore be in either the pub cluster (red) or the Highstreet cluster (purple). It is pertinent to note, that quick access by train into London city centre is usually very important for a Londoner, therefore, properties around train stations (such as the Crayford region - Figure 5(b)), may make ideal rental locations for some tenants. However, all identified properties in Figure 5(b) are in a low-crime borough with low-rent, near local amenities, and close to a train station for quick access to the city centre.

5. Conclusions

The following project demonstrates the usefulness of applying data science principles to identify trends within structured data. Bexley was identified as the ideal London borough for rental properties when filtering based on rental price, crime level, and location London city centre. K-means clustering was further employed to provide unique insights into the distribution of rental price in Bexley. Moreover, Foursquare API location data has been leveraged to cluster properties based on their proximity to local amenities. Nine low-crime, low-rent properties with access to local amenities have been identified that would be suitable for shared tenancy, family, or young professional.