

The Battle of the Neighborhoods

Xiayi Cheng

Abstract:

In this report, I summarise my work on analysing London boroughs to identify the most suitable area for a unique foreign business to expand. The report also includes correlation analysis between rent price movement and the borough venue categories.

1. Introduction:

London stands out as one of the most iconic cities of Europe. Developed from the British colonialism, the city became a signature for multi-cultural society. Given the enormous size of London, the community has many things to offer, which provide career, social and leisure opportunities for a well-balanced daily life. However, London is also well-known because of its high living cost, in particular the housing price. Only the few wealthiest are able to afford the properties here. Industrial renters, on the other hand, rely on office space rentals. For many industrial organisations, the rental expense is one of the largest expense in their balance sheet. A wisely selected occupation space could reduce rental expense and bring logistical advantages over competitors, which play a crucial role in the success of their business. Estate agencies provides property recommendation services to industrial renters. Understanding the underlying driving factors for the rental price movement is particularly important for providing industrial client-based services. Using this information, the estate agencies would be able to tailor down a property selection, which could enhance the growth of the client's business and avoid unnecessary rental cost for the clients in the future.

1.1 Business problem

A unique foreign chain restaurant (“XYZ's”) has decided to expand their business in London for the next few years. They aim to establish various cosy restaurants **across the city to maximise their exposure** in a new culture. To **minimise the cost**, they wish to have a stable rental expense. Individual stores are self-supporting regarding groceries. Hence, they **rely on the close-by supermarkets** to avoid delivery costs. “XYZ's” has seek consultancy from the estate agency in founding the potential areas in London, which align with their situation.

2. Data acquisition and cleansing

To address previously mentioned challenges in the “XYZ's” business, following data are useful in this initial stage:

- Rent prices statistics from London Datastore of MAYOR OF LONDON:

This data set is provided by the government and documents the average, median, lower and higher quantile rent prices for each borough. The data are acquired every quarter from 2011 to 2019.

- London borough's area sizes, population distribution and postal codes:

Borough sizes and population are easily obtained from Wikipedia. I use pandas' web-scraping to read the table for further data analysis. Over 300 000 postal codes of London were extracted. These will be used for venue exploration in Foursquare.

- Overall venues statistics in each London borough

Due to limited number of available call in Foursquare, I only use every 100 postal code in Foursquare to explore the nearby venues. To minimise the missing areas, the exploration radius is enlarge to 3km.

In addition, the venue categories provided by Foursquare are often very detailed, which are over 400 different venue categories. To reduce the number of categories, I have introduced some major categories, such as Pub/bar, Transport, Restaurants, Fastfood, Tourism, Nature...

- Supermarket statistics for each London borough

As our clients specifically required good shopping opportunities, I have extracted the number of supermarkets in each borough using the postal codes mentioned earlier.

According to a BBC survey, Tesco, Sainsbury's, M&S, Aldi, Asda, Waitrose, Co-op are the most common stores in UK, which also define the scope of my data acquisition.

3. Methodology

To quantify the rent price movement, analysis will be carried out on the derived quantities about the rent prices: the mean rent price and annual rent increment in percentage. As part of the data exploration and to understand the available data, I use pairwise correlation and multilinear regression to analyse the rent dependence on the venue categories in each borough.

Extrapolation on the rent prices are made for the next 3, 5 and 10 years. These are compared with

the population density, supermarket density to deliver a selection of boroughs. This selection is then separately confirmed by the technique: K-mean clustering.

4. Data exploration and analysis

4.1 Average rent price and annual percentage rent price increment

To quantify the rent price movement, we use the mean price and percentage annual rent increase, where the latter is extracted by using linear regression. Figure 1 shows these plotted in choropleth maps.

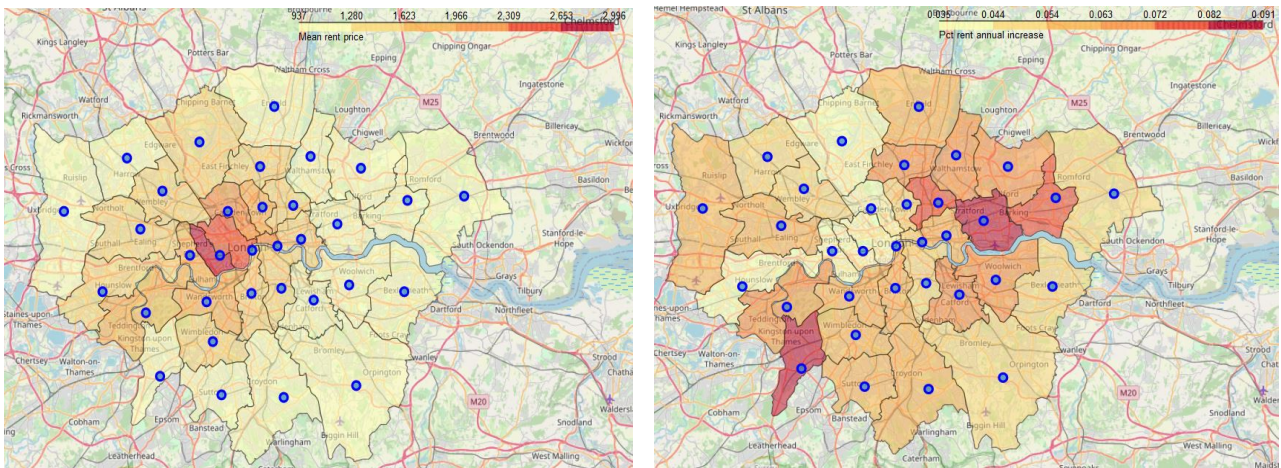


Fig. 1: The mean rent price in London Boroughs (left). Red indicates higher price and yellow cheaper price. The annual price increment is shown in the right, where strong increments are observed in the east and south-west.

The averaged rent price appears to be the highest in Kensington and Chelsea, and slowly decays towards the outside of the city centre. The price increments can be used as an indicator for developing borough. Fig. 1 (right) reveals two areas, where the more development is in process.

In the bar plot below (Fig. 2), it is clear to see that the borough with high rent prices has reached a saturation, which results in the lowest price increment.

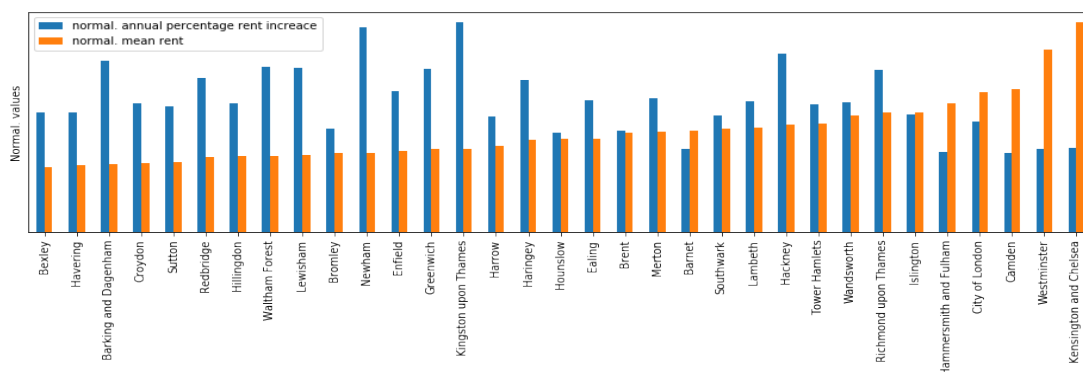


Fig. 2: The mean rent price and price increments are normalised to illustrate the relative trends. Kensington and Chelsea, Westminster and Camden with the highest rent also have the lowest rent increments, indicating saturation.

If the client is interested in renting in these boroughs, the overall rental expenses for the next 3, 5 and 10 years are extrapolated and illustrated In Fig. 3.

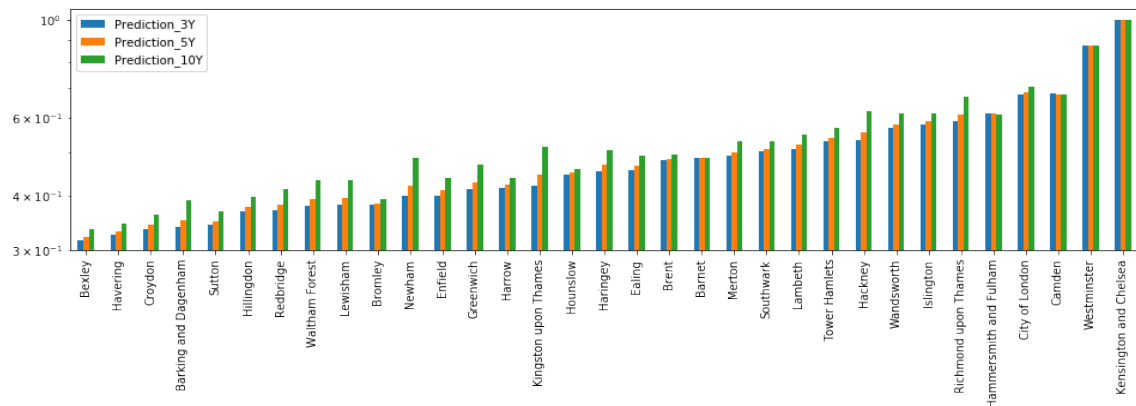


Fig. 3: Extrapolated relative rental expenses over the next 3, 5 and 10 years.

Boroughs with the larger rent increments only have significant price impact on the overall rental expenses over the next 10 year. For short rent, up to 5 years, the expenses will be mainly driven by the mean rent price.

4.2 City features

Table 1 shows that top 10 venue categories across London and their overall contribution, which are acquired using Foursquare.

Generally, most of the venues are concentrated around the city centre. Population density and supermarket density are exhibited in Fig. 4.

Table. 1: Venues distribution

Category	Contribution [%]
Restaurants	14
Pub/Bar/Clubs	6.2
Stores	6.1
Cafe	5.3
Fastfood	3.4
Grocery	1.9
Hotel	1.8
Transport	1.3
Gym/Fitness	< 1
Nature	< 1

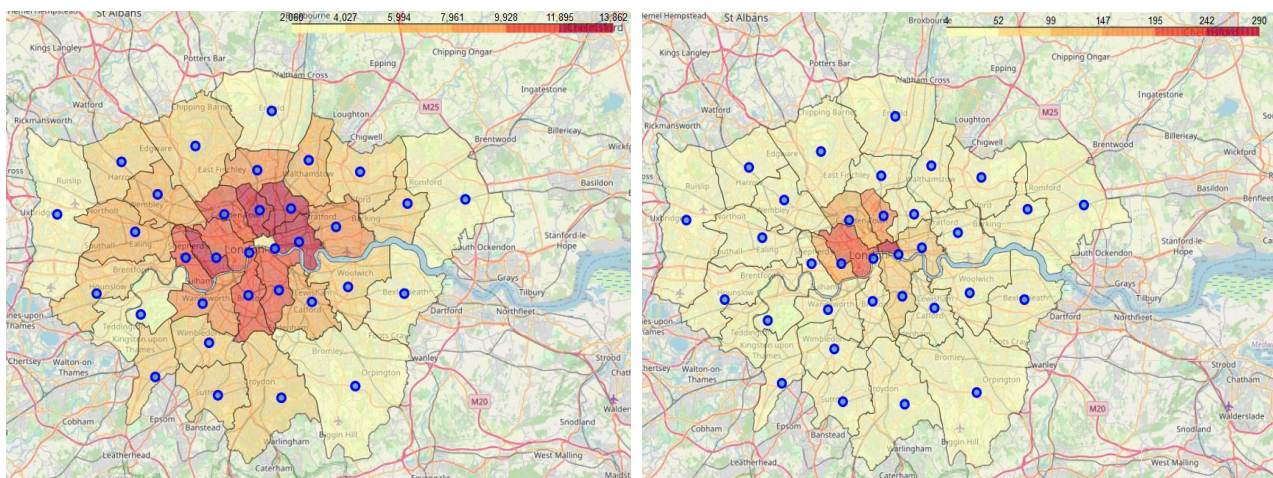


Fig. 4: Population density (left) and supermarket density (right).

The boroughs with highest population density are located around the City of London, where a

slightly higher density towards the north east London is observed. It is important to note, that City of London is mainly occupied by offices and lack of local residence, which gave a false impression of low population. This is reflected in the density of supermarkets across London. Fig. 4 shows that the density of supermarket dominates in Westminster and City of London, as these are boroughs with highest flow of people.

4.3 Correlations with the rent price

Using pairwise correlation, I have analysed the relation between City features and the rent prices.

Table 2: Correlation statistics on mean rent price.

	Correlation	p-value
meanRent	1.000000	1.543639e-239
Gym / Fitness Center	0.767801	1.840014e-07
Cafe	0.738031	9.506965e-07
Restaurant	0.715363	2.887128e-06
Fastfood	0.714782	2.966305e-06
Pub/Bar	0.710950	3.540658e-06
Nature	0.709658	3.756168e-06
SupermarketDensity	0.709161	3.842026e-06
Hotel	0.676878	1.523213e-05
Grocery	0.669854	2.010398e-05
Store	0.626253	9.679650e-05
PopulationDensity	0.569546	5.412393e-04

Table 2 show the pairwise correlation between the top 10 venue categories and the mean rent price. I used p-value as a preliminary measure for the correctness of the fit. This suggests that the rent price is main driven by the quality of living (variety of entertainment, food and shopping opportunities). Interestingly, the rent price is less affected by the transport opportunities.

With multi-dimensional linear regression, we could extract the contribution of the venue categories to the rent price movement. Fig. 5. show the distribution for mean rent price (orange) and percentage annual rent increments (blue).

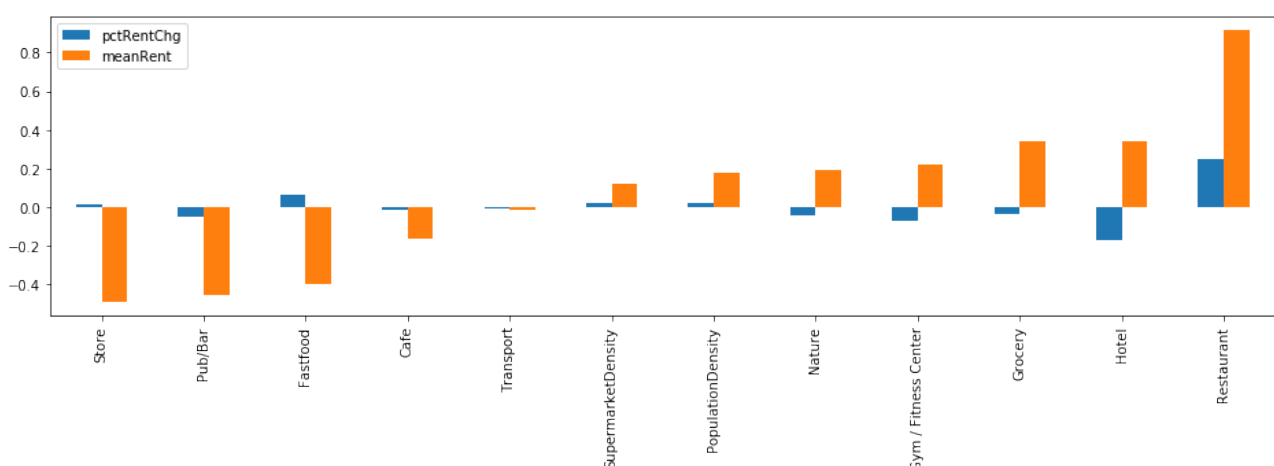


Fig. 5: Contribution of venue categories to the rent price movement

The mean rent price appears to be mainly driven by the number of restaurants, Hotels and grocery opportunities. Surprisingly, multi-dimensional linear regression suggests that the rent price is negative affected by the pub/bars and stores in the borough. On the other hand, the rental

increments are only driven by the existing restaurants, in the absence of other categories of venues.

Both methods agree on the fact that transport opportunities don't contribute to the rent price movement.

4.4 Comparing Population density, Supermarket density and the projected rental expenses

To address the need of our client, the three key parameters are popularity, shopping opportunities and reduced rental expenses. Hence, we focus on finding the boroughs with a suitable population and supermarket density and a low projected overall rental expenses. As rents are mostly a major part of the overall expense, we show the latter three quantities ranked after boroughs with increasing rent prices for 3 years in Fig. 6.

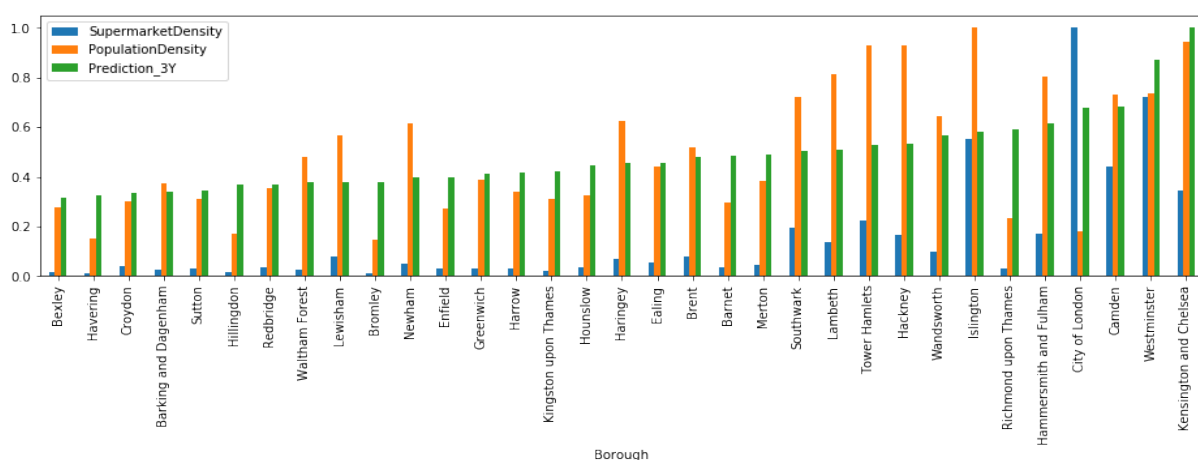
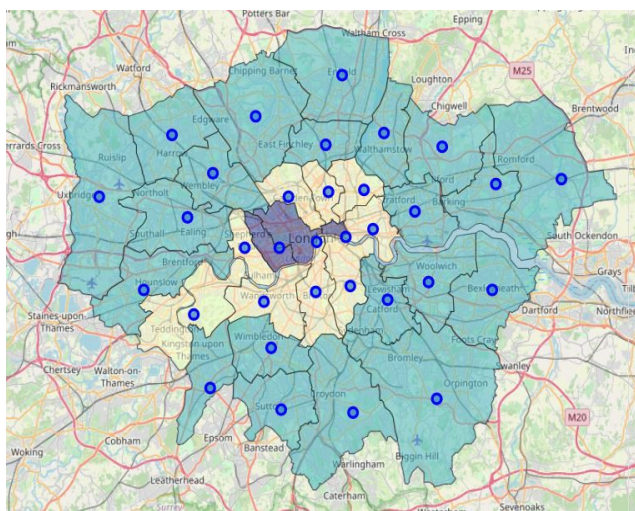


Fig. 6: Supermarket density compared with population density and rental expenses.

The first three boroughs with the most expensive rental expenses stand out with well balance popularity and shopping opportunities.

Southwark, Lambeth, Tower Hamlets, Hackney, Wandsworth, Hammersmith and City of London appear by this analysis to be ideal for opening a business due to a range of moderate rent price and decent amount of exposure and local shopping opportunities. As previously mentioned, City of London has low local residents, but is actually high in popularity.

K-mean clustering has also confirmed our finding as shown in Fig. 7 (right), where boroughs with same suitability are clustered into the same group shown in different shadowed areas. It is worth noting that the Richmond upon Thames has been falsely identified as a potential borough.



5. Result and conclusion

Our analysis have shown that the most expensive borough is the Kensington and Chelsea, which is followed by Westminster. Boroughs located around these two appear to fall into the next rent price class. However, rent cost seems to be in general higher in the west than in the east. In contrast, east(north) and south(west) London appear to experience a larger annual rent increase. This trend indicates an higher demand in housing and show potential future development. Particular heavily developing boroughs are Kingston, Hackney, Newham and Barking.

Using multi-dimensional linear regression, we found out that the mean rent price is mostly driven by the number of hotels, Grocery, Gym, Nature, which are all indications of balanced living options. Transport and supermarket don't really affect the rental price. After narrowing down our selection scheme, we recommend:

- Camden, Hackney, Hammersmith and Fulham, Islington, Lambeth, Southwark, Tower Hamlets, Wandsworth and City of London

as suitable borough to expand the business. However, our client has the option to go for more expensive boroughs to trade for higher population density.

6. Outlook

This was the initial step of identifying suitable boroughs to grow. The next step would identify the most suitable location within a borough for further optimisation. In addition, an exclusive analysis on the local competitors would contribute greatly to the decision making process of our client.