Open a restaurant in London

an overview of London boroughs

Mark Long

2020-03-22

1 Introduction

1.1 Background

London is considered to be one of the world's most important global cities and has been called the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment-friendly, and most-popular-for-work city.

London attract lots of new inhabitants, new investors each year because above reasons.

1.2 Business Problem

London is a huge city, not everyone knows London well before they go there. It is risky to move to a place you barely known.

This report addressed this problem and provide a investment guide to people who want to open a restaurant in London.

It supposed to help those new commers to have a general business view of London. So, they can better decide where they are going.

1.3 Assumptions

Usually businesses are grouped together, people like to go to a place with multiple restaurants to have a wider range of choice. If you want to open a restaurant, It is better to go to those places which already have restaurants to minimize the risk. However, having same type of restaurants at the same place will be disaster due to competition, We must carefully select correct type of restaurant to increase diversity and decrease competition.

Besides that, richer neighborhoods usually have higher spending power, you would like to setup your business there to maximize the revenue. In this case, It is assumed higher house sales price will indicate higher spending power.

So, this report is based on two assumptions:

1. Multiple but not the same type of restaurants at the same place indicates lesser risk.
2. Higher average house sales price indicates higher revenue.

2 Data

2.1 Data Source

- List of boroughs can be obtained on Wikipedia. https://en.wikipedia.org/wiki/List_of_London_boroughs
  - Borough names and geo coordinates can be found here; geo coordinates will be used in foursquare API to find recommended venues.
- By using foursquare API, list of recommended venues can be fetched base on above coordinates. https://api.foursquare.com/v2/venues/explore
- List of top venues of each borough can be found here to decide the functionality of a borough
  - House sale price data on Land registry of UK, I will use data from last two years (2018, 2019 123945 transactions in total). http://prod1.publicdata.landregistry.gov.uk.s3-website-eu-west-1.amazonaws.com/pp-2019.csv
  - District(borough) to house sale price mapping can be extracted here.

2.2 Data Cleaning

Data downloaded from Land registry of UK contains house sales data from All UK. To have more focus, I select all records for Greater London area.

3 Methodology

3.1 House price analyze

The average house price in London is very High, 835457 pounds in average.

BRENT	683713.4982
BROMLEY	513846.0275
CAMDEN	2057129.3731
CITY OF LONDON	8117333.4499
CITY OF WESTMINSTER	3187531.6141
CROYDON	499684.5322
EALING	648575.8090
ENFIELD	530180.7377
GREENWICH	513601.2540
HACKNEY	828780.8992

Figure 1. Mean house price of London boroughs

By look at the boxplot, we can see there are quite a lot of outliners.

Figure 2. Boxplot of house price of London boroughs

Depends on the type of the restaurant, we can choose to serve high-end customers (outliners) or average customers. In this report I'll focus on average customers. So, I must reduce the effect of outliners. I will use median price to represent the house price of a borough.

BRENT	470000.0000
BROMLEY	430000.0000
CAMDEN	770000.0000
CITY OF LONDON	923080.0000
CITY OF WESTMINSTER	970000.0000
CROYDON	372500.0000
EALING	460000.0000
ENFIELD	385550.0000
GREENWICH	430000.0000
HACKNEY	557087.5000

Figure 3. Median house price of London boroughs

Then the median house price of London is 455000 pounds now. By look at more detailed data, it looks more normal now.

3.2 Borough geo data analyzes

Data from " List_of_London_boroughs" contains geographic information of different boroughs. Each borough has very different size.

Figure 4. Map of London boroughs

However foursquare API returns data based on a radius of center coordinate. To make foursquare returns more representative data of each borough, I defined a unique radius for each borough.

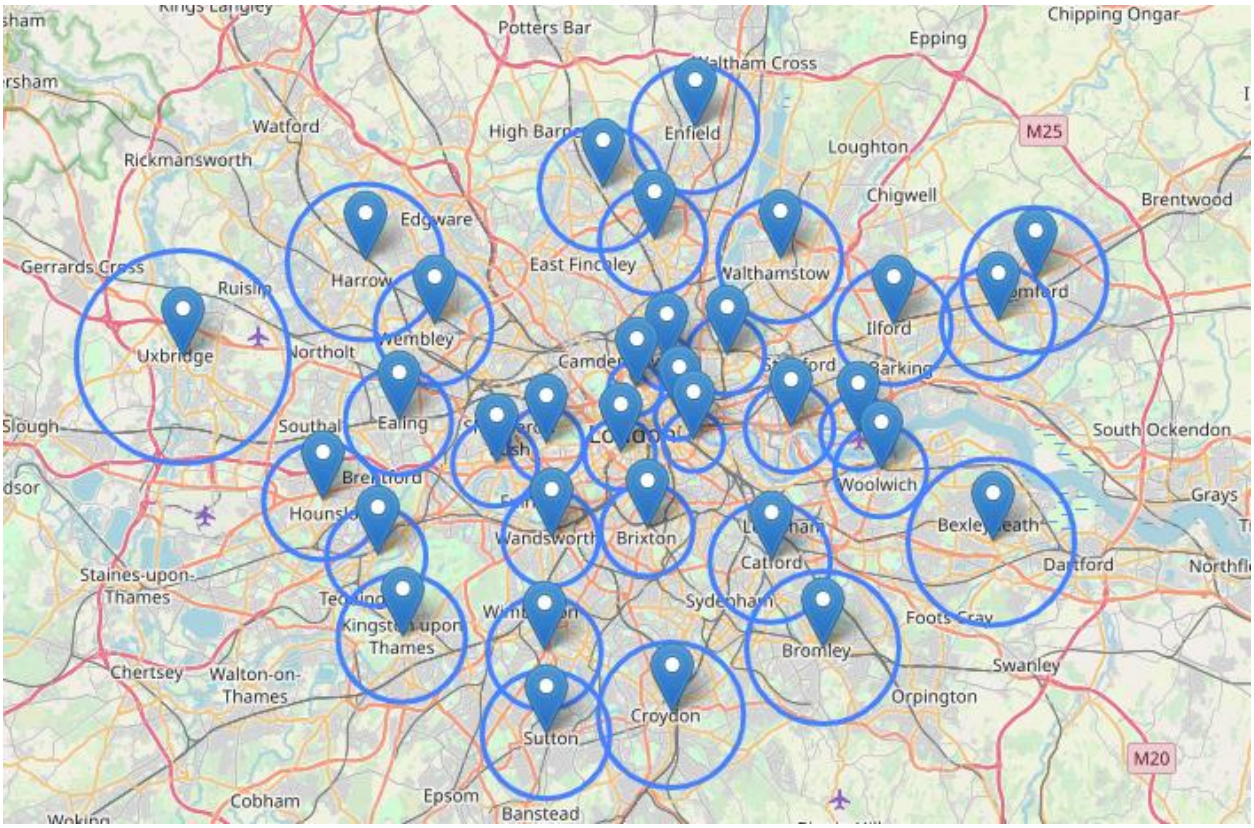

Figure 5. Foursquare

The radius is calculated by use the average distance of nearest 3 coordinates.

3.3 Unsupervised Clustering by K-Mean

I use foursquare API to return top 100 venues of each borough. 100 is a limit of foursquare API.

Category of a venue is used to indicate the functionality of a venue, the more category appears in the venue list of a borough, the more significant this category for this venue.

By using one hot method, I encode the category of venues into integer features. And group the data by borough then take the mean value, I'm prepared to feed the data to K-Mean model.

3.3.1 Find best K for K-Mean model

Before I start to use K mean model to cluster the data, I must decide what is the best number for K.

By use Elbow method, I ran the evaluation 20 times.

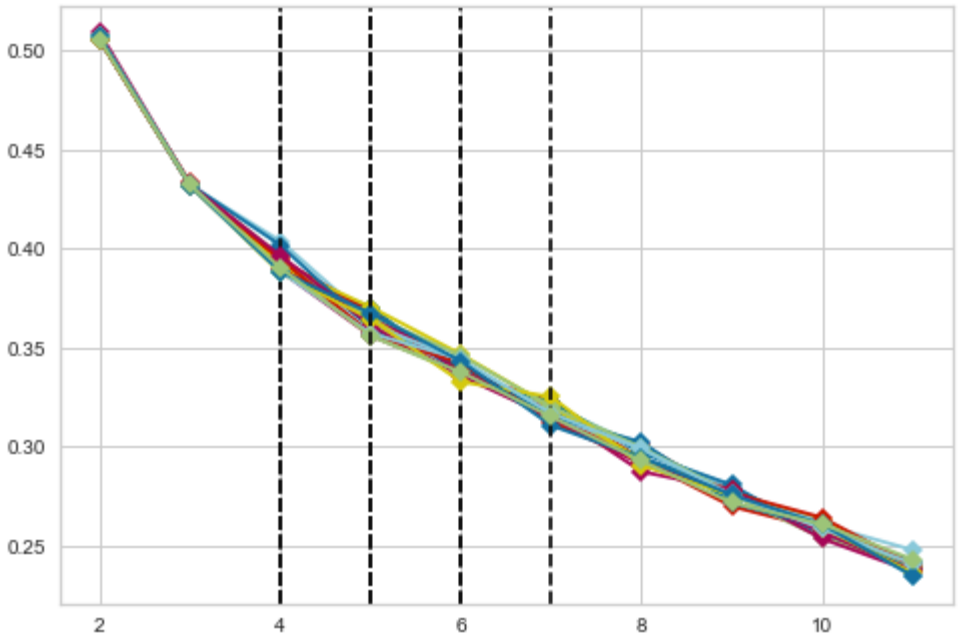

Figure 6. Graphic for elbow method

The result is [6, 4, 5, 5, 4, 5, 6, 5, 4, 4, 4, 4, 5, 5, 5, 7, 5]

It successfully detected 17 times of elbow with average value of 5. So I will use 5 as the optimal K for this report.

3.3.2 Cluster data with K-Mean model

With K decided, I feed the data to K-Mean model to cluster similar boroughs.

After that, I sort the result by house price inside the cluster so We can give suggestion easily based on the descending median house price of the borough.

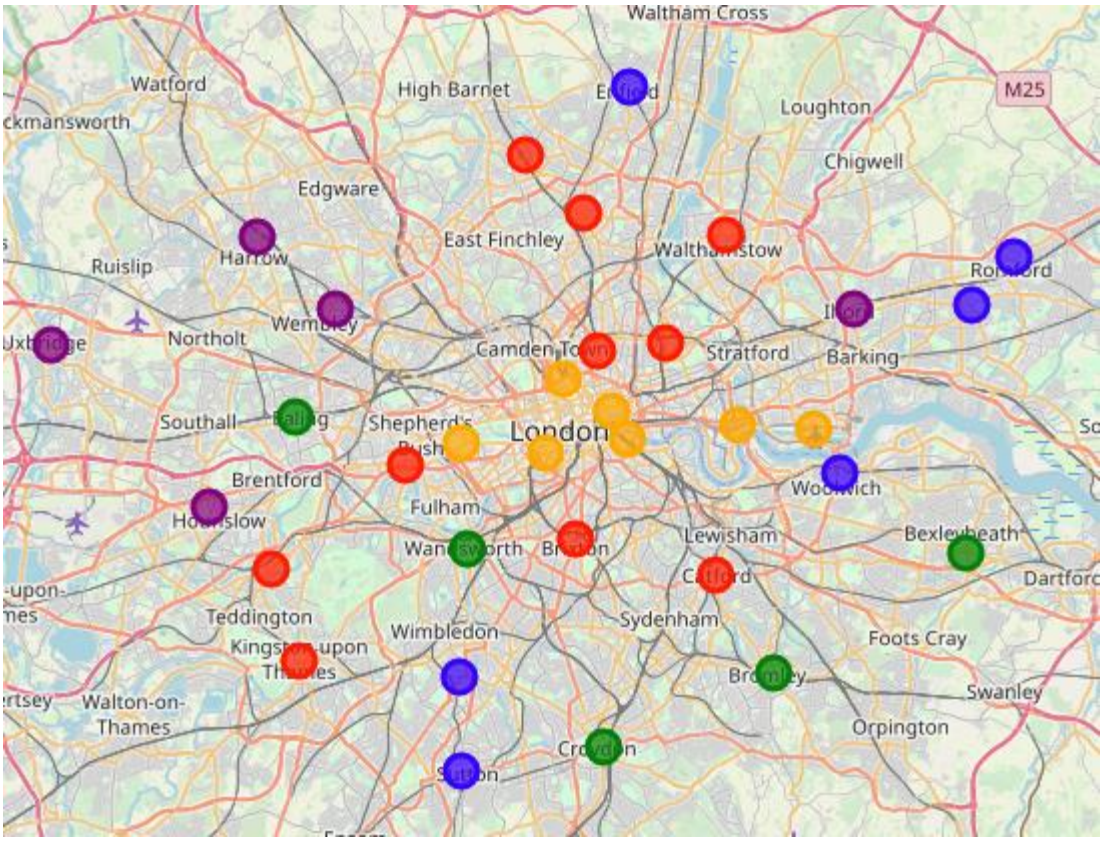

Figure 7. Clustered boroughs

4 Results

4.1 Cluster 0

Those areas have good night lives because of those pubs, Cafe, and coffee shops.

If you want to open a restaurant, go to HAMMERSMITH AND FULHAM, and try to avoid pizza, French restaurant, Japanese restaurant and Turkish restaurant.

	borough	latitude	longitude	price	radius	cluster	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 M Comm Ve
11	HAMMERSMITH AND FULHAM	51.4927	-0.2339	740000.0000	2263	0	Pub	Café	Coffee Shop	Gastropub	Pizza Place	Park	French Restaurant	H
25	RICHMOND UPON THAMES	51.4479	-0.3260	623945.0000	2682	0	Pub	Café	Coffee Shop	Park	Rugby Stadium	Italian Restaurant	Garden	H
17	ISLINGTON	51.5416	-0.1022	587000.0000	1400	0	Pub	Coffee Shop	Bakery	Gastropub	Cocktail Bar	Park	Café	Mediterran Restau
10	HACKNEY	51.5450	-0.0553	557087.5000	2062	0	Pub	Coffee Shop	Bakery	Café	Brewery	Bookstore	Pizza Place	Yoga St
20	LAMBETH	51.4607	-0.1163	520000.0000	2484	0	Pub	Coffee Shop	Market	Café	Restaurant	Brewery	Japanese Restaurant	Pizza Pl
1	BARNET	51.6252	-0.1517	500000.0000	3291	0	Café	Pub	Coffee Shop	Park	Turkish Restaurant	Italian Restaurant	Supermarket	Grocery S
12	HARINGEY	51.6000	-0.1119	500000.0000	2807	0	Café	Turkish Restaurant	Pub	Coffee Shop	Park	Bakery	Greek Restaurant	Mediterran Restau
19	KINGSTON UPON THAMES	51.4085	-0.3064	480000.0000	3472	0	Pub	Café	Garden	Park	Hotel	Gastropub	Coffee Shop	Gym / Fitn Ce
29	WALTHAM FOREST	51.5908	-0.0134	440000.0000	3298	0	Pub	Café	Coffee Shop	Pizza Place	Park	Restaurant	Turkish Restaurant	Brev
21	LEWISHAM	51.4452	-0.0209	409995.0000	3253	0	Pub	Park	Café	Coffee Shop	Gastropub	Italian Restaurant	Indian Restaurant	Gym / Fitn Ce

Figure 8. Cluster 0 of London boroughs

4.2 Cluster 1

Areas for livings because of the supermarkets and grocery stores. With convenient transportations the difference in average house price are not so big.

If you want to open a restaurant, try to go to MERTON, and try to avoid Sushi, Indian, and Italian restaurants.

	borough	latitude	longitude	price	radius	cluster	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 Cor V
22	MERTON	51.4014	-0.1958	464100.0000	3028	1	Park	Supermarket	Pub	Sushi Restaurant	Indian Restaurant	Grocery Store	Fast Food Restaurant	Resta
9	GREENWICH	51.4892	0.0648	430000.0000	2483	1	Pub	Park	Grocery Store	Hotel	Gym / Fitness Center	Café	Coffee Shop	E
8	ENFIELD	51.6538	-0.0799	385550.0000	3445	1	Supermarket	Pub	Turkish Restaurant	Coffee Shop	Grocery Store	Train Station	Gym / Fitness Center	
27	SUTTON	51.3618	-0.1945	380000.0000	3432	1	Pub	Grocery Store	Supermarket	Coffee Shop	Park	Hotel	Italian Restaurant	F C
14	HAVERING	51.5812	0.1837	360000.0000	3918	1	Grocery Store	Coffee Shop	Supermarket	Pub	Bakery	Italian Restaurant	Fast Food Restaurant	
0	BARKING AND DAGENHAM	51.5607	0.1557	309000.0000	3039	1	Grocery Store	Coffee Shop	Park	Supermarket	Pub	Fast Food Restaurant	Shopping Mall	Furn

Figure 9. Cluster 1 of London boroughs

4.3 Cluster 2

Places mixed with nature and inhabitants.

Open a restaurant in WANDSWORTH, and avoid pizza, burger, French and Thai restaurant.

	borough	latitude	longitude	price	radius	cluster	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 Mos Commo Venu
30	WANDSWORTH	51.4567	-0.1910	610000.0000	2541	2	Coffee Shop	Pub	Park	Pizza Place	Café	Supermarket	Bakery	Frenc Restaurar
7	EALING	51.5130	-0.3089	460000.0000	2839	2	Park	Pub	Coffee Shop	Grocery Store	Hotel	Café	Burger Joint	Pizz Plac
4	BROMLEY	51.4039	0.0198	430000.0000	4055	2	Pub	Coffee Shop	Park	Gym / Fitness Center	Clothing Store	Grocery Store	Pizza Place	Indiar Restaurar
6	CROYDON	51.3714	-0.0977	372500.0000	3894	2	Coffee Shop	Pub	Grocery Store	Pizza Place	Café	Park	Indian Restaurant	Bookstor
2	BEXLEY	51.4549	0.1505	348620.0000	4513	2	Pub	Grocery Store	Supermarket	Italian Restaurant	Furniture / Home Store	Park	Coffee Shop	Clothin Stor

Figure 10. Cluster 2 of London boroughs

4.4 Cluster 3

Those areas have lots of Indian restaurant. The house price is quite average here.

Go to BRENT for slightly higher price. Avoid open another Indian restaurant in those areas.

	borough	latitude	longitude	price	radius	cluster	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 I Com V
3	BRENT	51.5588	-0.2817	470000.0000	3089	3	Coffee Shop	Indian Restaurant	Sandwich Place	Hotel	Clothing Store	Gym / Fitness Center	Grocery Store	Spo Goods
13	HARROW	51.5898	-0.3346	454500.0000	4134	3	Indian Restaurant	Coffee Shop	Pub	Gym / Fitness Center	Supermarket	Café	Park	It Resta
24	REDBRIDGE	51.5590	0.0741	415000.0000	3116	3	Indian Restaurant	Park	Supermarket	Fast Food Restaurant	Clothing Store	Coffee Shop	Portuguese Restaurant	Sanc R
15	HILLINGDON	51.5441	-0.4760	407000.0000	5652	3	Pub	Indian Restaurant	Coffee Shop	Supermarket	Gym / Fitness Center	Golf Course	Park	
16	HOUNSLOW	51.4746	-0.3680	400000.0000	3099	3	Pub	Coffee Shop	Indian Restaurant	Clothing Store	Grocery Store	Hotel	Bakery	Superm

Figure 11. Cluster 3 of London boroughs

4.5 Cluster 4

Areas with highest house price, has lots of hotels and coffee shop, good places for tourism and business.

Go to KENSINGTON AND CHELSEA if possible and avoid opening Italian and Indian restaurant.

	borough	latitude	longitude	price	radius	cluster	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 Mos Commo Venu
18	KENSINGTON AND CHELSEA	51.5020	-0.1947	1270000.0000	1997	4	Restaurant	Hotel	Gym / Fitness Center	Italian Restaurant	Bakery	Indian Restaurant	Pub	Clothin Stor
31	CITY OF WESTMINSTER	51.4973	-0.1372	970000.0000	1891	4	Hotel	Theater	Plaza	Art Museum	Garden	Cocktail Bar	Japanese Restaurant	India Restaura
32	CITY OF LONDON	51.5155	-0.0922	923080.0000	1218	4	Gym / Fitness Center	Coffee Shop	Hotel	Steakhouse	Garden	Cocktail Bar	Vietnamese Restaurant	Sala Plac
5	CAMDEN	51.5290	-0.1255	770000.0000	1419	4	Coffee Shop	Café	Hotel	Pizza Place	Bookstore	History Museum	Beer Bar	Burge Joi
26	SOUTHWARK	51.5035	-0.0804	550000.0000	1632	4	Coffee Shop	Hotel	Cocktail Bar	Italian Restaurant	Bakery	Garden	Scenic Lookout	Restaura
28	TOWER HAMLETS	51.5099	-0.0059	475000.0000	2348	4	Coffee Shop	Hotel	Burger Joint	Gym / Fitness Center	Pub	Bar	Park	Plaz
23	NEWHAM	51.5077	0.0469	402500.0000	2014	4	Hotel	Coffee Shop	Café	Gym / Fitness Center	Sandwich Place	Pub	Supermarket	Groce Stor

Figure 12. Cluster 4 of London boroughs

5 Discussion

The data from Land register UK does not contains size information. This might lead some bias of average purchase power of that borough. We might need to find another data source to obtain more detailed data, hence improve the accuracy of the model.

London is very big; each borough of London is big as well. Foursquare API can only return top 100 venues of a borough. Those 100 venues might not have a good representation of that borough. Our result is heavily depending on foursquare ’s database.

6 Conclusion

In this report, I categorize London’s boroughs by its functionality. It shows the most common venues in a borough and sorted the similar boroughs with house price information.

It can help a newcomer to get familiar with London, help them to decide which borough they want to go. And if they want to open a restaurant, it also shows which kind of restaurant should be avoided.

The same model and principles can be applied to other cities as well, just by replace corresponding data.