Open a restaurant

in London

an overview of London boroughs

Mark Long

2020-03-22

# Introduction

## 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.

## 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.

## 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.

# Data

## 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.

## 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.

# Methodology

## House price analyze

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

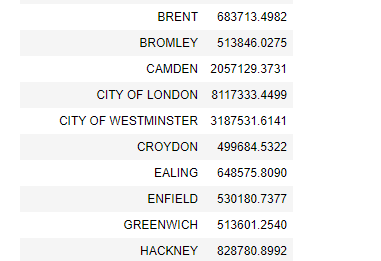


Figure 1. Mean house price of London boroughs

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.

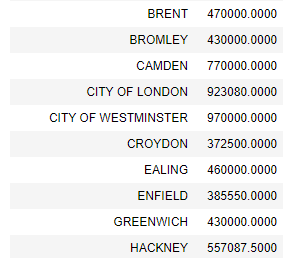


Figure 3. Median house price of London boroughs

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

## Borough geo data analyzes

Data from “ List\_of\_London\_boroughs” contains geographic information of different boroughs. Each borough has very different size.

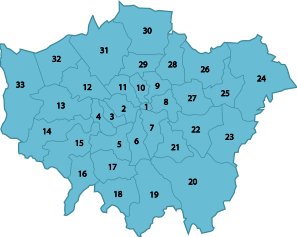


Figure 4. Map of London boroughs

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.

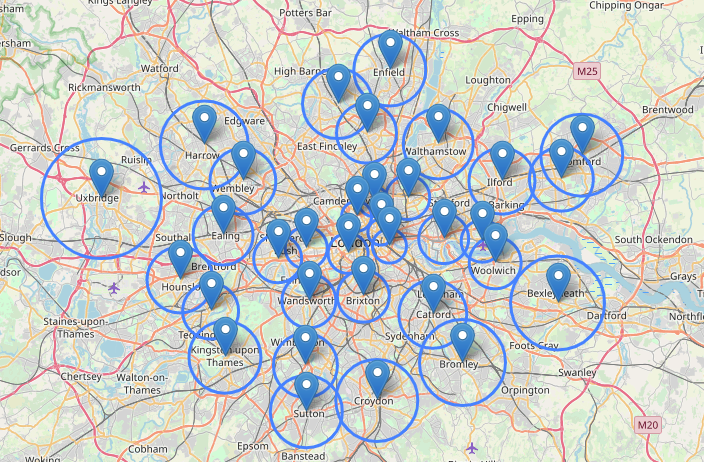


Figure 5. Foursquare

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

## 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.

### 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.

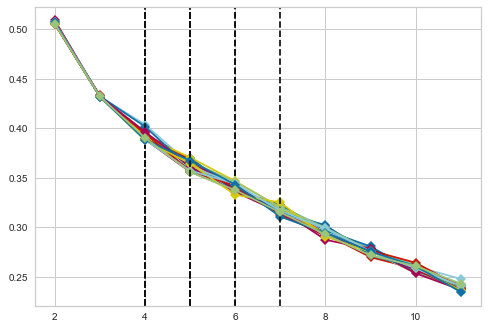


Figure 6. Graphic for elbow method

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.

### 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.

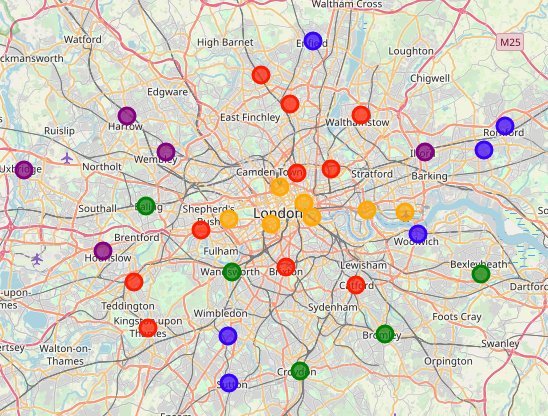


Figure 7. Clustered boroughs

# Results

## 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.

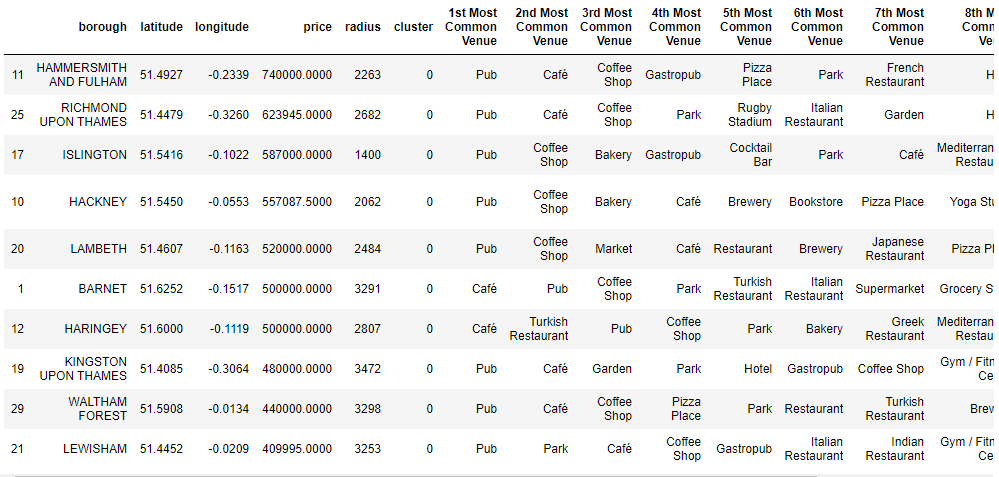


Figure 8. Cluster 0 of London boroughs

## 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.



Figure 9. Cluster 1 of London boroughs

## Cluster 2

Places mixed with nature and inhabitants.

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



Figure 10. Cluster 2 of London boroughs

## 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.



Figure 11. Cluster 3 of London boroughs

## 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.

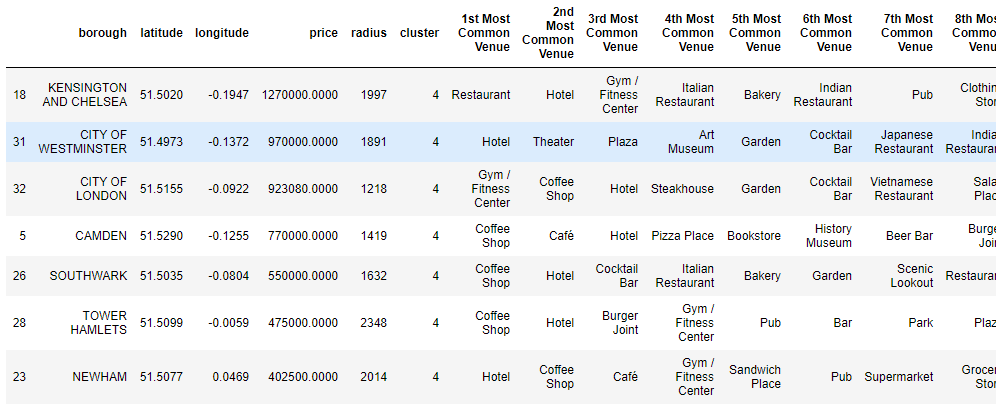


Figure 12. Cluster 4 of London boroughs

# 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.

# 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.