The Battle of Neighborhood

Jun Wang
April 30, 2020

1. Introduction

Background
London is the economic and political centre of UK, and its land prices have always been high. Before Brexit earlier in this year, someone warned that if Brexit, the value houses in UK may fall and this will have an impact on the British market. Recently, the impact of coronavirus has also shocked the UK housing market, according to Global Property Consultancy, UK house prices would probably fall by 3% in 2020, and expects the prices to recover in 2021. In the current scenario, it could be a new opportunity for business operators and investors.

Problem
The land price directly affects the operating conditions of some commercial operators, that could be reflected through leasing or purchasing the land to operate, especially have more impacts on the catering industry. Usually, the rent costs account for a large portion of the expenditure. therefore, the choice of location and whether the operator could balance the revenue and the cost is of vital importance.
To explore this issue, we will explore the components of venues in the Greater London area, the distribution of regional land prices, what venues are popular with the public, and recommend different business models based on the operator's situation.

Interested Audience
People who are moving to large cities as well as new immigrants that want to start their businesses, catering in this case, are more likely interested in this case.

2. Data

Data that will be collected:
- List of postcode and coordinates of Greater London area
- List of the average land value and value change for the area covered under certain postcodes
- List of venues information around given postcodes (venue latitude, longitude and category)

- Most common and popular venues in the Greater London area

The location that includes the postcode and coordinates of the Greater London area are extracted from www.doogal.co.uk, which collects and sort out UK postcodes as well as data from Land Registry data from GOV.UK, and this website has more clear and organized data provided to the public. These data are used to search land value information through Zoopla, nearby venues information through FourSquare, as well as be plotted on maps.

The land value information comes from www.zoopla.co.uk is a property website, which is more reliable because it updates housing market information on a monthly basis. This website provides API to obtain data, however, the free version has a lot of restrictions, so the web crawler is used to get land value information. The land value data are used to identify how the highest and lowest land value area distributed, and it will be helpful to indicate different business models and strategies. The value change could be used to decide future trend to make better decisions.

To explore a given location, such as London in this case, the FourSquare API is used to get the most common venue categories in each postcode area, which are used to group postcode into clusters and build the visualization.

3. Data Wrangling

Dataset downloaded from Doogal contains much information of the Greater London area that we are not going to use in this case, so we just import some necessary columns such as Postcode, Latitude, Longitude and In Use?. The first three columns are quite intuitive, and the last one is to check whether the certain postcode is in use or not. If it is not in use, then we do not need that piece of information and should drop it. What is more, all rows that contain NaN values should also be dropped. After cleaning the dataset, we get a clean table that contains more than 170,000 pieces of information as follow:

	Postcode	Latitude	Longitude
0	BR1 1AA	51.401546	0.015415
1	BR1 1AB	51.406333	0.015208
2	BR1 1AE	51.404543	0.014195
3	BR1 1AF	51.401392	0.014948
4	BR1 1AG	51.401392	0.014948

The next step is to obtain land value information. As we do not find any suitable dataset that could be better used in this case, we decide to use web crawler technology to search from Zoopla, which is a website that provides very updated housing value information. Our dataset now has a very large number of data, and it will take a very long time to search information for each piece of postcode, we simply divide the postcode into two parts, and only select the first part as our new postcode, which could represent a small area that will not have a large difference in land value.

	Postcode	Latitude	Longitude	Average Value	Value Change%
0	BR1	51.412007	0.020858	530834	5.52
1	BR2	51.387282	0.022403	570400	8.37
2	BR3	51.404585	-0.030556	592643	7.03
3	BR4	51.375123	-0.008384	595940	7.20
4	BR5	51.391689	0.103339	453718	7.02

4. Exploratory Data Analysis

As we are exploring the land value in the Greater London area, we sort the data in ascending order according to the average value and divide data in to three parts according the average land value, the top 50 valued areas, the middle 50 valued areas and the least 50 valued areas to observe their distribution.
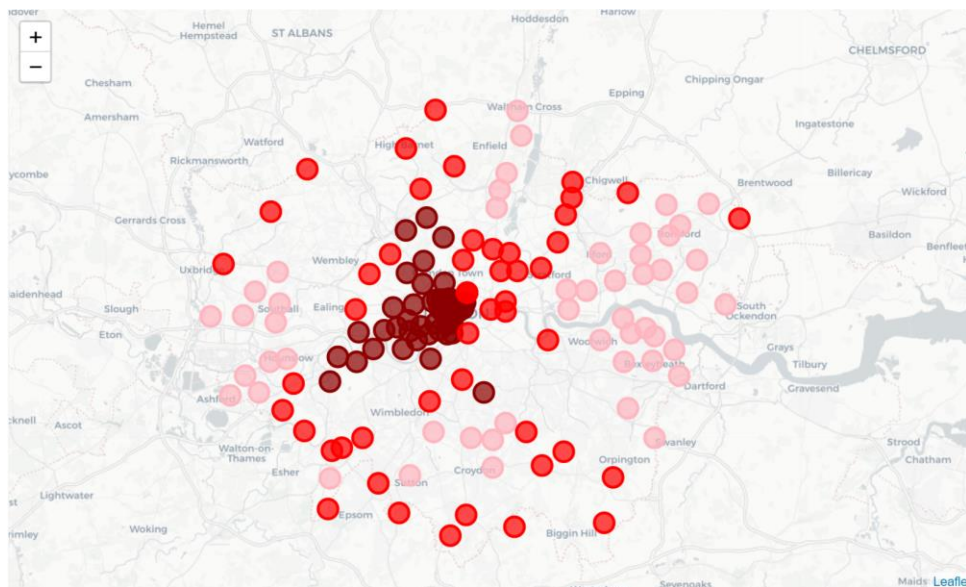
It is can been seen that the dark red circles represent the top 50 valued areas, the red circles represent the middle 50 valued areas, and the pink circles represent the least 50 valued

areas. By scattering these points on GLA map, we could find that the most valued areas are distributed in the centre of London and the least valued areas are distributed outer of London. However, we mainly focus on the top and least 50 areas in this case.

Now, we connect with FourSquare API to explore these postcodes further to obtain the name, category and coordinates of the venues around each postcode area. We set dummy variable of category that if a postcode area has this type of category, the value will be 1, otherwise 0, and count the mean to get its frequency. Then, we create the new data frames to display the top 10 common venues for each postcode location, we could see that for different locations, the most common are not the same, which means that the popularity of venue type for different locations are also different. The table of common venues for top 50 is shown as follow:

	Postcode	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
0	N2	Park	Pub	Café	Bakery	Coffee_Shop	Turkish_Restaurant	Grocery_Store	Forest
1	N6	Pub	Park	Coffee_Shop	Gastropub	Bakery	Café	Pizza_Place	Movie_Theater
2	NW1	Hotel	Pizza_Place	Pub	Garden	Movie_Theater	Zoo_Exhibit	Bakery	Bookstore
3	NW11	Park	Grocery_Store	Pub	Bakery	Turkish_Restaurant	Italian_Restaurant	Coffee_Shop	Café
4	NW3	Park	Pub	Bakery	Coffee_Shop	Zoo_Exhibit	History_Museum	Middle_Eastern_Restaurant	Café

To make it more clear, we build data visualization to find most common venues for top/least 50 land value areas. For the top 50 areas, the most common ones are Park, Hotel and Café, while for the least 50 areas, the most common ones are Coffee Shop, Supermarket, Grocery Store and Pub. The size of font is based on the occurrence frequency of the venue category.


Most Common Venues in Top 50 Land Value in London   Most Common Venues in Least 50 Land Value in London

Furthermore, we explore the most common venues in the whole Greater London area by restricting the boundaries of the search area. After processing data in a similar way we did above, we find that, overall, people are more inclined to outdoor activities when comparing with the result that we find by dividing GLA into smaller areas which gives us some information to make decisions.

Top 10 Common Venues in Top 50 Land Value in London

Top 10 Common Venues in Least 50 Land Value in London

Top 20 Common Venues in Greater London
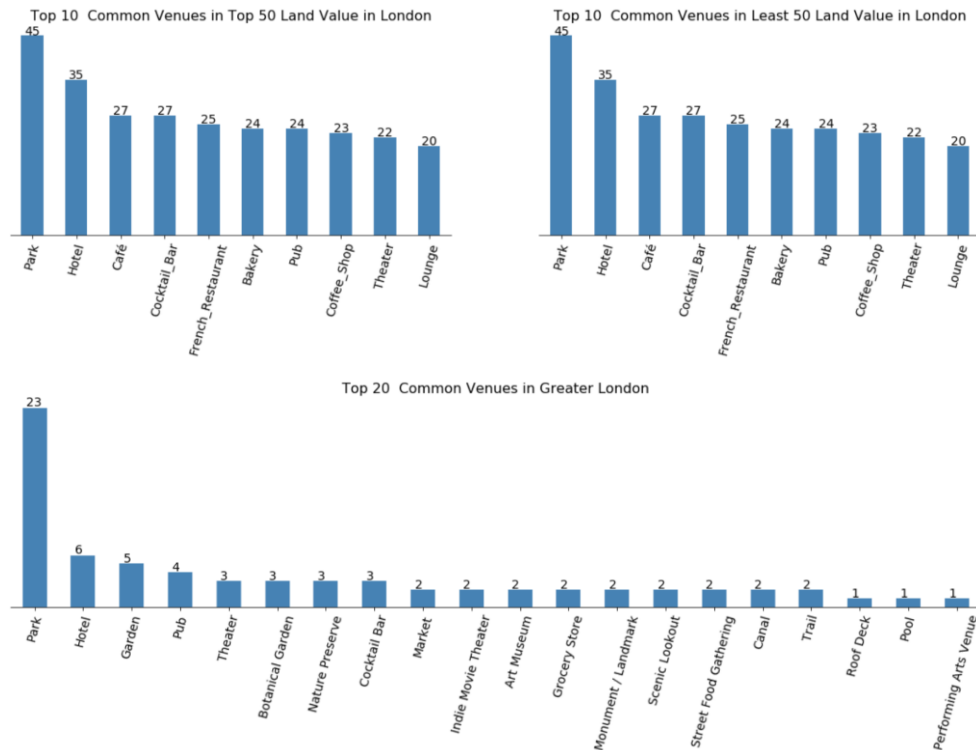
By scattering the points on map, we could see how these venues distributed around the postcode areas. The black points represent the top venues in the whole GLA area.
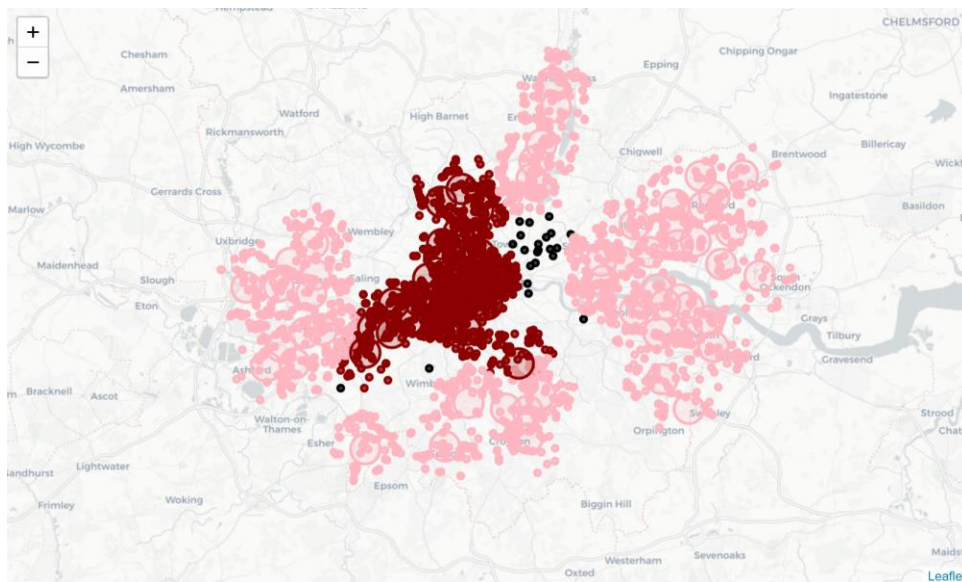

The Brexit effect did not break down London's housing market, and we can clearly see that although there is not a huge increase, the hosing price during past 12 months still has an upward trend. This could be good news for people are going to start their own businesses. London, as the world's second largest economic and financial centre, it is hard to believe

that house price will fall sharply, therefore, the decrease in the rise could be an opportunity to invest.
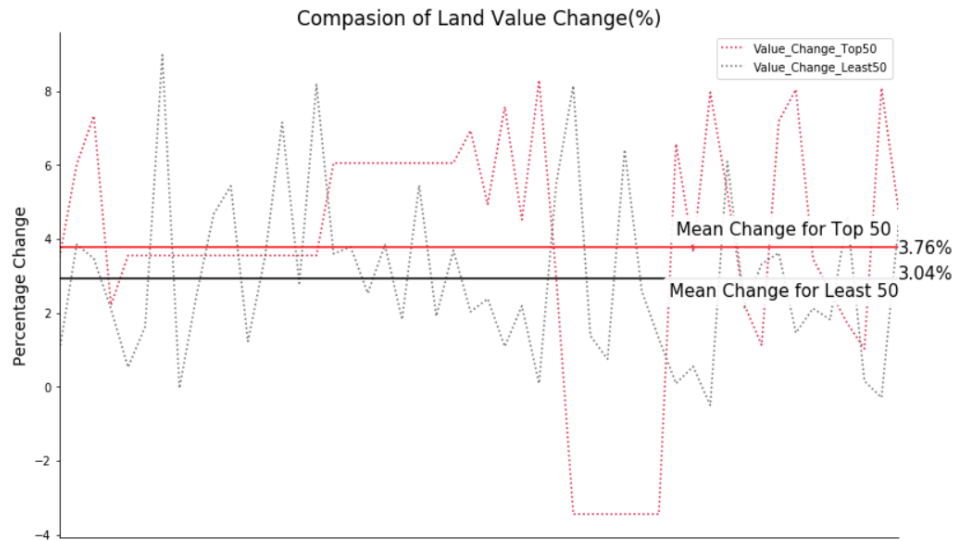
After filtering the data, we are also interested in how the venues are clustered. Therefore, we run K-Means algorithm to figure this out. The principle is that dividing the data into different categories, and the distance between the data points in the same category should be very close, that is, the closer the data should be, the similar the data is. In another word, the similarity between data is inversely proportional to the Euclidean distance between them.

The two graphs below are about how the data is clustered, and we choose five as the number of clusters. The left hand side graph is the clusters regarding the top valued areas, and the right hand side graph is the clusters regarding the least valued areas.
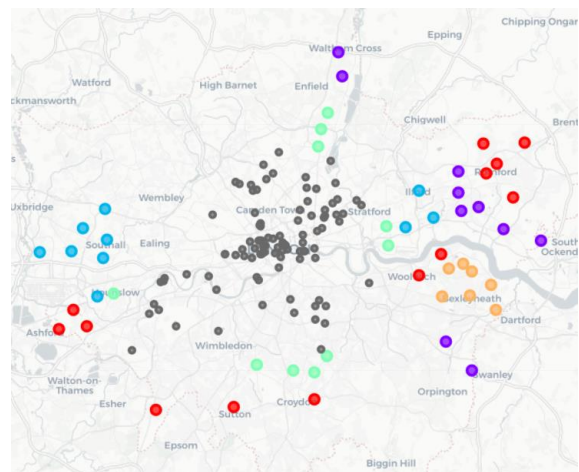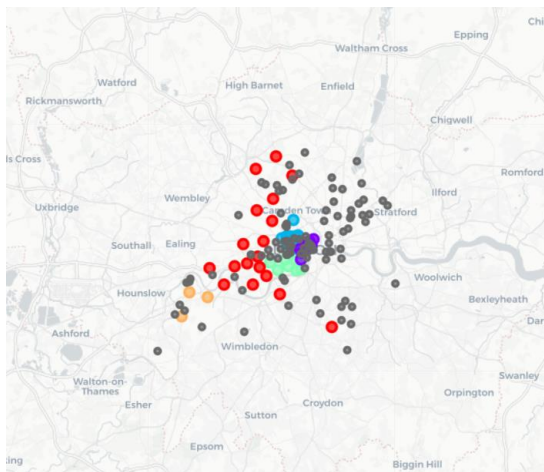
5. Discussion and Conclusion

From the statistical table and visualization tool, we could find that the land value in Greater London area is a circle that takes the city centre as the centre of the circle, and gradually decreases toward its periphery. The closer to the centre, the more expensive the housing price, and the price difference can be more than five times, which is huge. We mainly focus on the most valued and the least valued areas in GLA. For these different regions, the characteristic of data we find from FourSquare has a big difference. In the central area, the most common venues are Park, Hotel, French Restaurant and Café etc., which are relatively high consumption place, and most of them are for people to socialize. However, around the fringe area, the most common venues are Coffee Shop, Supermarket, Grocery Store and Pub etc. Most are places to meet people's daily needs. When we expand the perspective again, to observe the whole situation in GLA, Park, Botanical Garden and Theater seem to be very popular, which are mainly distributed in the northeast in GLA between the two regions.

For a restaurant operator who wants to find a suitable place to open a business, it's important to choose a location based on his or her situation. If he or she has a sufficient funding background, it is obvious that it can choose any place, including the city centre. However, what we are discussing here is that most people would consider the cost of opening a new business. For a person who has very little start-up capital, the mobile dining car is a very good choice. As the most popular in GLA are the Parks or Botanical Gardens which have open space and high traffic of people, more revenue will be generated, and it will not require much capital for land rental. For a person who wants to find a place for opening a restaurant, he or she will need to consider the needs of people living in different regions as well as his or her foresight for the market, which could be the balance between revenue and cost, as well as the perspective on the housing market after the rising pressure from the Brexit, also, how much the COVID-19 will influence the housing price in GLA, it is still difficult to predict due to the rapidly changing market and policy from the government.

The gap in land value does not mean that there is a gap in wealth within people, because this is not persuasive, and further investigation and research are needed, but it can somewhat explain the difference in people's lifestyles. After all these analyses, I hope this analysis can be helpful for restaurant operators.