Battle of the Neighbourhoods
Applied Data Science Capstone

Introduction

London has a population of roughly eight million people and finding the best spot to open a hospitality business is difficult. I aim to make this process easier by analysing different locations in London. I will use the example of finding the optimal location to open a Thai Restaurant.

The ideal location will be found from a combination of factors such as the area average annual house price increase, proximity to central London, proximity to other restaurants and more specifically, proximity to other Thai Restaurants.

Ultimately I will suggest a few neighbourhoods and identify what are the key aspects that make them a good place to open a business.

Data

The population per London Borough is from 150,000 to 300,000. Instead of boroughs, I will look at smaller subdivisions of London in order to split London into more manageable sizes. To do this lets look at the Parliament Constituenties which are roughly 80,000 in population. Lets get the names by scraping the data from wikipedia using **beautiful soup**. https://en.wikipedia.org/wiki/London_boroughs.

I will get the latitude and longitude coordinates of the centre of London and each of the Boroughs by using the **google maps geocoding API**.

Data relating to parliamentry constituencies is well documented. In order to find the most up and coming areas I will look at which areas had the largest increase in house prices over the past year. The data will be scraped from an excel document from the UK's Office for National Statistics Website. https://www.ons.gov.uk/peoplepopulationandcommunity/housing/bulletins/housepricestatisticsforsmallareas/yearendingseptember2019

Using **foursquare API** I can get information on the type of venues in the areas to see where there are fewest restaurants.

Methodology

There were 73 London Constituencies which were scrapped from the wikipedia page website.

Using google maps Geocoding API we could then get the latitude and longitude coordinates which we could then use to plot the centre points of the constituencies on a map using folium.
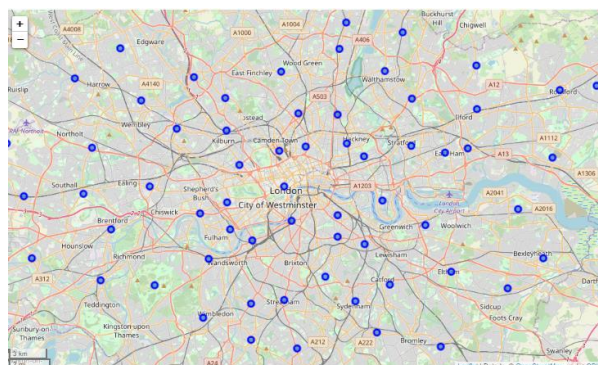
Then I calculated the distance to the centre of London and found the median house price and annual percentage change from the government office for national statistics.

	Constituency	Latitude	Longitude	Median House Price GBP	18/19 % Change	Distance to centre
0	Greenwich and Woolwich	51.483000	0.028000	515000.0	8.421053	11151.190300
1	Bethnal Green and Bow	51.530858	-0.040193	536500.0	7.300000	6616.858593
2	Feltham and Heston	51.460000	-0.412000	359000.0	6.845238	20435.641039
3	Dulwich and West Norwood	51.447000	-0.084000	550000.0	6.692532	7370.658052
4	Hackney South and Shoreditch	51.540000	-0.060000	595000.0	6.250000	5942.274411

I was able to use foursquare API to get the locations of all the restaurants in London. There were found to be 3282 restaurants in total and only 64 Thai restaurants. Plotted on a map we can see restaurants in blue and Thai restaurants in red. As there were so few Thai restaurants, I decided not to account for this when assessing future locations to open a restaurant.
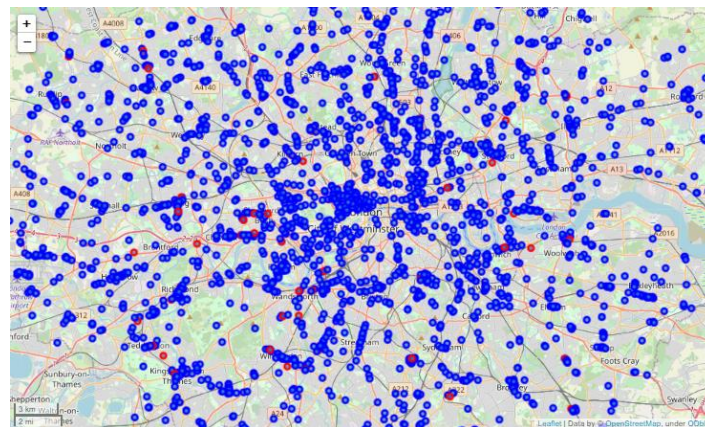
Results

To find acceptable locations I created a grid structure with latitude and longitude coordinates and then filtered through to find only the locations that did not have a restaurant in 400 m. Plotted on a heat map we can see the locations where there is a high density of restaurants and the acceptable locations which are outside of the high density areas.
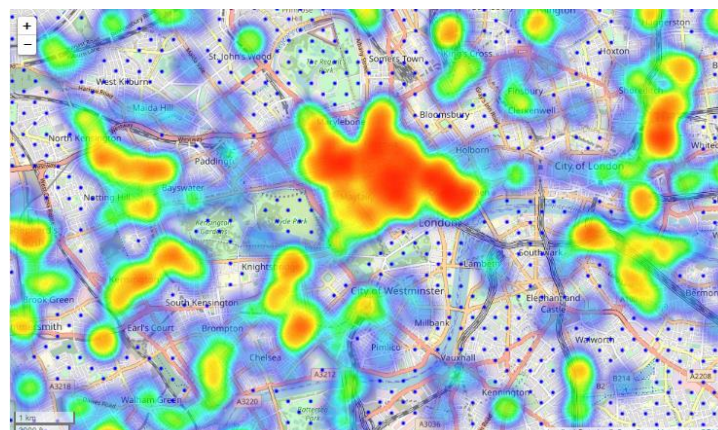
I then conducted Kmeans clustering to group these into neighbourhoods of there own. I used 15 clusters. It was found that there was a very large cluster south of the river so this is a potential place to open a restaurant.

```
1  from sklearn.cluster import KMeans
2
3  number_of_clusters = 15
4
5  kmeans = KMeans(n_clusters=number_of_clusters, random_state=0).fit(Acceptable_Locations[['Location_Lat','Location_Lng']].val
```

Discussion

Our analysis shows a number of potential locations to open a restaurant. Particularly in the borough of Camberwell and Peckham which is south of the river. There are a number of promising locations in Bethnal Green and Bow which showed the second largest increase in average house price and is relatively close to the centre. Hence, by the methodology above, this is the best place to open a restaurant in London.

Conclusion

I have found a number of locations to open a restaurant that do not have another restaurant in a 400m radius. I then grouped these neighbourhoods by using kmeans clustering. Further work would find the best of these locations based on their proximity to the centre of London and by how much the house price had increased in the last year.