Predicting the best Borough in outer London for a food/drink business

1. Introduction

1.1 Background

London is the capital and largest city of England and the United Kingdom. The city stands on the River Thames in the south-east of England, at the head of its 50-mile (80 km) estuary leading to the North Sea. London has been a major settlement for two millennia. London is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. This means that the market is highly competitive meaning that the cost of doing business is one of the highest. But what about those who do not have enough money, but still want to open a restaurant/drink business outside inner city limits? Where will it be cheapest and will have enough people living around to be popular? Where the competition is not too overwhelming? Therefore, any new business setting must be analyzed thoroughly to have the most profit of it. Statistical analysis of the proper data is mandatory because it will certainly reduce the risk of failure.

1.2 Problem

Data that might contribute to determining the best outer borough in London might include the Boroughs outside the city of London and their population, this type of businesses around them, and the lowest rent possible. This project aims to predict which outer Borough is the most suitable for opening of such type of business.

1.2 Interest

Obviously, everyone from an individual to a small or big company would be extremely interested in an accurate prediction of the best venue to make a profit of such business.

2. Data acquisition and cleaning

2.1 Data sources

Most of the preferable data such as Boroughs of outer London with their coordinates, rent data, and venues data can be found online in Wikipedia (https://en.wikipedia.org/wiki/List_of_London_boroughs') and in the 4Square API.

2.2 Data cleaning

Data were downloaded and scraped in one table. However, there was a problem because Wikipedia provide us with some information that were not needed for our analysis such as Borough council, political situation, and the inner Boroughs of London. These data were deleted, and we only kept those that mattered to us such as Name, Area, Population, Coordinates, and Rent.
Through the explore function in 4SQuare we get a dataset of venues, from which we request the specific venues we take interest in.

2.3 Feature selection

From the dataset we created, we transform it to show the top 5 places we can better work with. Then, we merge the dataframes we created through our analysis into one dataframe, that contains all data values that will best help us with our analysis.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	Supermarket	Grocery Store	Park	Coffee Shop	Pub
1	Bexley	Pizza Place	Coffee Shop	Ice Cream Shop	Discount Store	Bakery
2	Bromley	Pub	Grocery Store	Coffee Shop	Park	Pizza Place
3	Enfield	Pub	Coffee Shop	Turkish Restaurant	Greek Restaurant	Garden Center
4	Haringey	Café	Pub	Park	Coffee Shop	Turkish Restaurant
5	Havering	Hotel	Park	Coffee Shop	Garden	Bakery
6	Merton	Pub	Park	Coffee Shop	Café	Bar
7	Redbridge	Pub	Park	Coffee Shop	Restaurant	Italian Restaurant

Figure 1.Dataframe of all Boroughs divided in 3 clusters with most common venues.

3. Exploratory Data analysis

First, we got the necessary information on London Boroughs, dropping the extras, that would not be needed for the analysis. Then we renamed the columns, making the dataset better on the eyes. Because of extra notes in the Wiki page, we only renamed some of the Boroughs. Due to the staggering difference in rent price, as well as the number of venues in London, we filtered the data to have only the Outer boroughs going forward. We edited the coordinates, found the Boroughs with the lowest rent, counted the venues and found the most popular Boroughs.

3.1 Clustering¶

Once the Boroughs are selected, we cluster them to analyze their similarities and differences and we can find the advantages or disadvantages of each Borough selected.

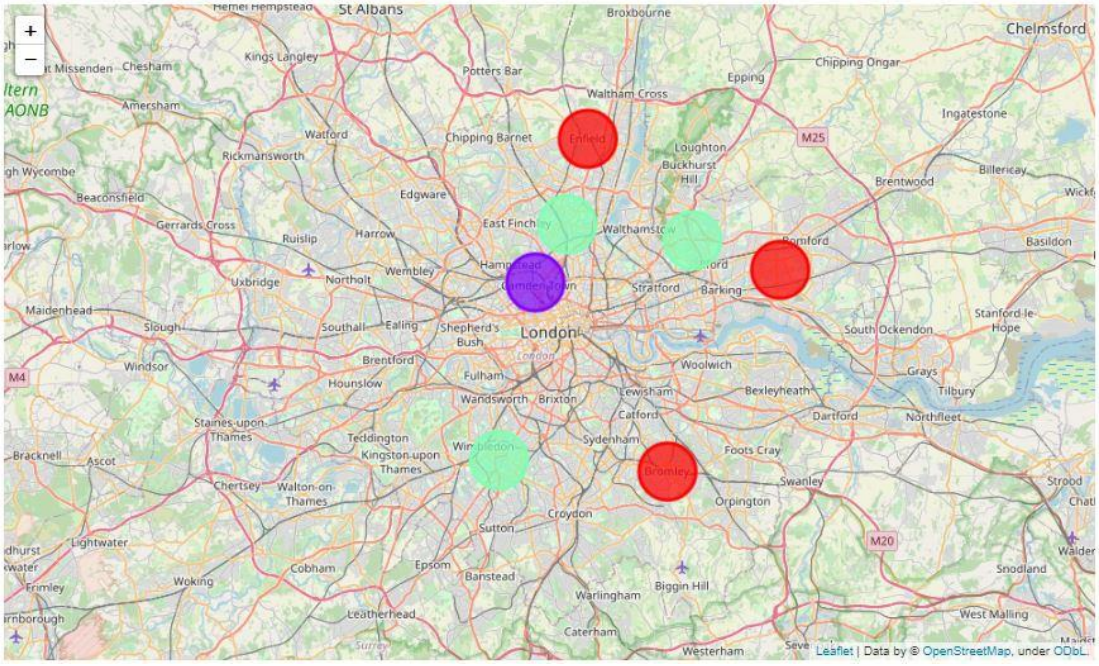

Figure 2.The clusters of the Boroughs of the outer London.

4. Results

In the **first** cluster which includes *Barking and Dagenham, Bromley,* and *Enfield* we see that the highest rent is in *Bromley*, while the other two Boroughs share the same price of rent. The lowest in popularity type of venue is **Pizza Place** for *Bromley*, **Garden Center** for *Enfield* (which is not similar with the type of business we are interested in), and **Pub** for *Barking and Dagenham*.

In the **second** cluster which includes *Bexley*, and *Havering* we see that the highest rent is in *Bexley*. The lowest in popularity type of venue is **Bakery** for both Boroughs.

In the **third** cluster which includes *Haringey, Merton*, and *Redbridge* we see that the highest rent is in *Merton*, while the lowest is in *Haringey*. The lowest in popularity type of venue is **Turkish restaurant** for *Haringey*, **Bar** for *Merton*, and **Italian Restaurant** for *Redbridge*.

5. Discussion

Taking into consideration the population, the max rent, and the least common venue of these 3 clusters, we came into the conclusion that *Enfield, Havering*, and *Haringey* are the best places to open a food/drink business. Although a vast variety of information are lacking

or cannot be retrieved at that moment, our data analysis provided a slight insight for a more profitable business move.

6. Conclusion

In conclusion, in this analysis I used some of the most common libraries to clear and manipulate data, and the 4square API to extract info about different types of venues in the outer London. By clustering the desirable Boroughs, I managed to conclude about the most suitable areas to open a food/drink business. Of course, due to lack of certain data, further future analysis is needed.