IBM Data Science Professional Certificate
The best area to build a Brazilian Restaurant in London

Larissa R. Teixeira

April 2020

1 Introduction

According to GLAECONOMICS -Tourism in London, London is one of the most attractive tourism cities in Europe and the most visited city in 2010. In 2009 there were 151,000 Brazilian visitors to the UK and 2010 there were 177,000 Brazilian visitors. Numbers of Brazilian visitors are predicted to increase by 32 per cent. Brazilian visitors tend overwhelmingly to be on holiday, young and making their first visit to the UK and spend approximately 118 million euros (2010 Census Data). For this reason, it is possible to think about opening a new business in London for Brazilians and other visitors. Obviously, any business decision must be carefully analyzed so as not to risk bankruptcy.

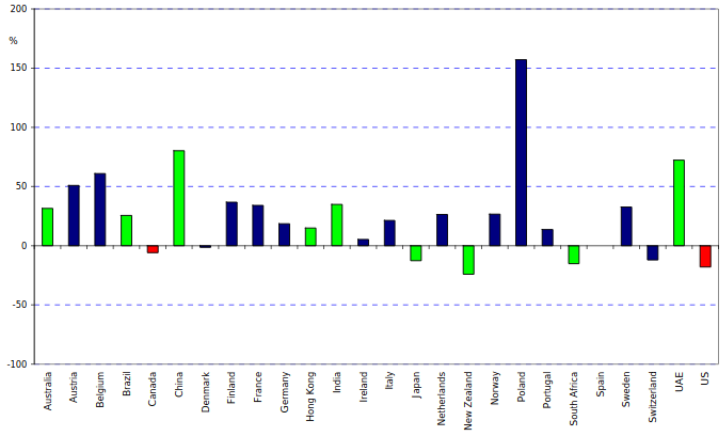


Figure 1: Growth in spend by country (percentage change 2007-2010). Source: ONS – International Passenger Survey detailed data / GLA Economics

1.1 Business Problem

Considering that London attracts tourists from all over the world, mainly Brazilians and that tourists always want to try new types of cuisine, it is easy to use a machine learning algorithm to analyze which is the best place to open a new Brazilian restaurant in London based on the number of businesses in London's neighborhoods and the population of each. The objective of this project is to analyse and select the best area to open a new Brazilian Restaurant in London. For this analysis, the concepts of data science and machine learning algorithm are used as clustering to answer the business question: What is the best area to build a Brazilian restaurant in London, if a developer is looking for a new business to open?

1.2 Target Audience

This project is for people who would like to investing or open a new Brazilian Restaurant in London. London is bursting with bustling Brazilian restaurants that give expats a taste of home and the rest of London a chance to experience Brazil's culture and delicious cuisine. For this reason, investors can consider London as a good city for this type of business.

2 Data

The dataset used for this problem were extracted from London datastore, where contains the neighborhoods, borough and the population of each neighborhood.

- For each neighborhood of London, we will get the latitude and longitude through the library GeoPy.
- To explore and recommend the best places to open a Brazilian restaurant, we will access the data through the FourSquare API interface, where we explore each neighborhood to find out which are the most frequent venues category in each one.
- To achieve the answer for this problem, we need clean e filter the dataset based on some criteria explained later.

	Code	Borough	Neighborhood	Year	Population	Hectares	Square_Kilometres
0	E05000026	Barking and Dagenham	Abbey	2020	16246	127.9	1.279
1	E05000455	Merton	Abbey	2020	10933	139.3	1.393
2	E05000630	Westminster	Abbey Road	2020	13331	109.5	1.095
3	E05000214	Greenwich	Abbey Wood	2020	16376	265.3	2.653
4	E05000382	Kensington and Chelsea	Addington	2020	11203	62.9	0.629

Figure 2: First five rows of the dataset

3 Methodology

After obtaining the dataset mentioned above, the Pandas python's library was used to read data about the city of London, where the following columns needed for this problem are selected which are: Boroughs, neighborhoods, neighborhood population, square kilometers of neighborhood.

However, this is just a list of address, composed of the union of the neighborhood and the Borough, of London. It is necessary to get the geographical coordinates such as latitude and longitude of each address in order to be able to use Foursquare API to get the information about the venues category of each neighborhood. For this, it was used the GeoPy python's library where through the address it is possible to get the latitude and longitude of the respective location. Next, the Foursquare API was used to the the top 200 venues category of each neighborhood through the latitude and longitude that are within a radius of 500 meters of London centroid. After that, the dataset was explored to understanding how many neighborhoods exists in London, the London's most crowded boroughs and which neighborhood contains Brazilian Restaurants.

After the visualization analysis, only the neighborhoods where exists Brazilian Restaurants was taken to clustering. For last, to get the number of cluster need for this problem, it was used the Elbow method.The Elbow method looks at the total within-cluster sum of square (WCSS) as a function of the number of clusters. For this problem, the number of cluster chosen was 3, where the change in WCSS begins to level off is 3.

The neighborhoods that contain Brazilian restaurants also contain other types of restaurants. So an average population per square kilometer was made for each neighborhood, compared to the number of restaurants in each one.

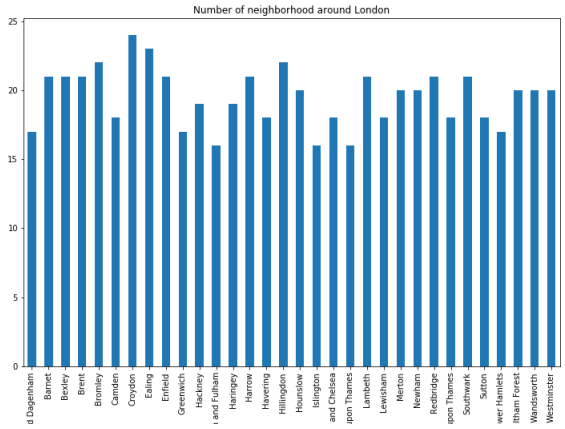


Figure 3: Number of neighborhoods for each borough

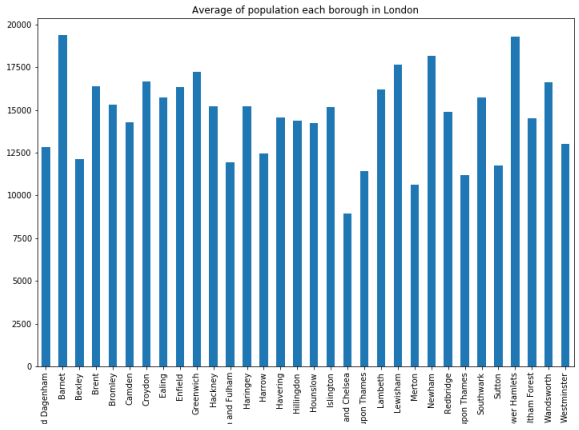


Figure 4: Average of population per borough

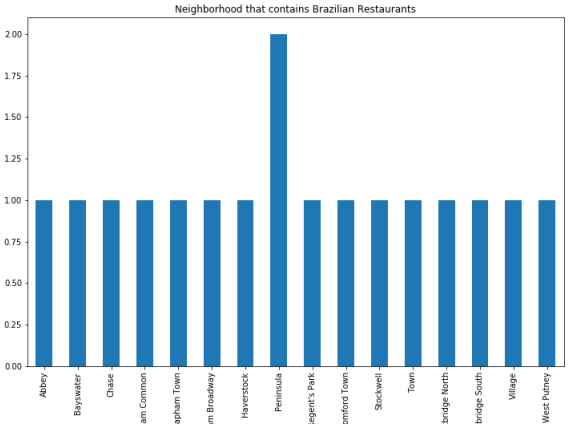


Figure 5: Neighborhood which contains Braizilian Restaurants

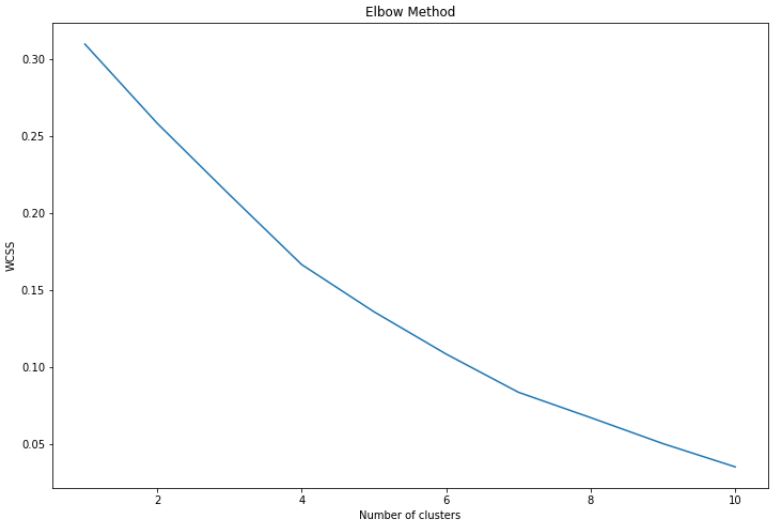


Figure 6: WCSS based on number of cluster

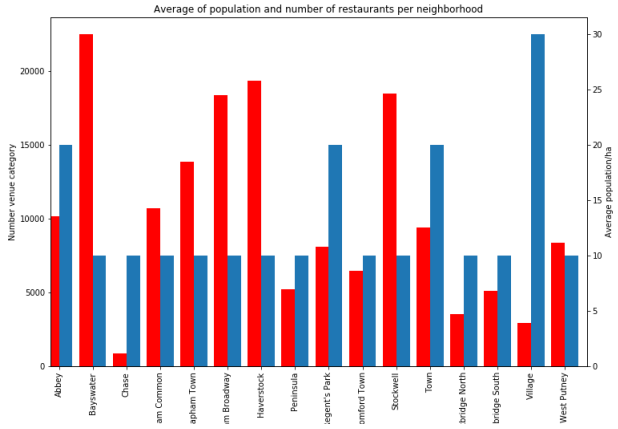
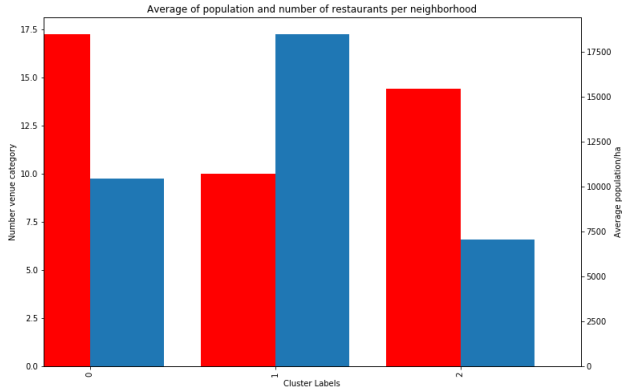


Figure 7: Population and number of restaurants per neighborhood

4 Results

The results from k-means categorize the neighborhoods into 3 clusters based on the number of restaurants existing in each cluster and the respective population:

- Cluster 0: Neighborhoods with the largest population of the three clusters and average restaurants equal to ten. This cluster contains one neighborhood with Brazilian restaurants.
- Cluster 1: Neighborhoods with the highest average of restaurants and the lowest population of the three clusters.
- Cluster 2: Neighborhoods with a large population and an average of fewer than 10 restaurants. This cluster contains two neighborhoods with Brazilian restaurants.



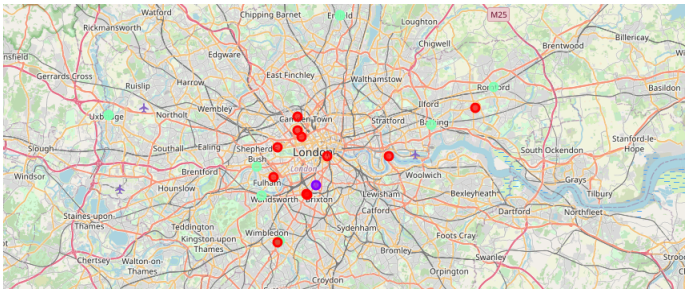


Figure 8: Cluster 0: Red, Cluster 1: Purple, Cluster 2: Green

5 Discussion

After all the analysis, it is possible to realize that the number of Brazilian Restaurants around London is too low. Many of the neighborhoods do not have Brazilian restaurants, even though the population is high. So, the neighborhoods where it is recommended to open a Brazilian Restaurant belongs to cluster 0 and 2.

	Neighborhood	Population_per_hectares	Venue Category	Latitude	Longitude	Cluster Labels
0	Abbey	10171.781437	20	51.474371	-0.053908	2
1	Bayswater	22509.652510	10	51.512276	-0.188385	0
2	Chase	865.380076	10	51.652085	-0.081017	2
3	Clapham Common	10704.393150	10	51.462075	-0.137359	0
4	Clapham Town	13845.535714	10	51.462292	-0.138856	0
5	Fulham Broadway	18371.814093	10	51.480883	-0.194349	0
6	Haverstock	19370.218579	10	51.544868	-0.153213	0
7	Peninsula	5228.469433	10	51.502931	0.003197	0
8	Regent's Park	8097.615086	20	51.526930	-0.150284	0
9	Romford Town	6473.831124	10	51.576046	0.182265	2
10	Town	9381.200139	20	51.572105	-0.152579	2
11	Uxbridge North	3509.605008	10	51.546655	-0.477511	2
12	Uxbridge South	5110.683349	10	51.546655	-0.477511	2
13	Village	2945.899874	30	51.489281	-0.047017	0
14	West Putney	8371.614301	10	51.460934	-0.215348	2

Figure 9: Neighborhoods recommended to open a Brazilian restaurant

One detail that was not taken into account for this analysis was the data from the neighborhoods most visited by tourists. This data can be added in future works to have a more accurate recommendation.

6 Conclusion

In this project, it was necessary identifying a business problem to try to solve. To do so, it was required a dataset extraction and prepare the data to use a machine learning algorithm, in this case K-means. The algorithm divide the data into three clusters based on the neighborhood similarities. Based K means result, the recommendation to open a new Brazilian restaurant in London is in the neighborhoods in cluster 0 and 2 because they have the highest average population per square kilometers and an average of fewer than 10 restaurants. Also, exist just three Brazilian restaurants around this neighborhoods. This project is important to avoid investors and people that would like to open a new business overcrowded areas in their decisions to open a new Brazilian Restaurant and understanding the best location for starts this new business.