IBM Applied data science capstone

North vs South: A Portuguese (vege)Tale

Final Assignment Report

Maria Ana Batalha - 7 de Maio de 2019

Introduction

Tourism in Portugal serves millions of international and domestic tourists. Tourists visit to see cities, historic landmarks, enjoy beaches, or religious sites. In 2017, Portugal had 23 million visitors. The most popular destinations were Lisbon, Porto, Fatima, the Portuguese Riviera and the Algarve. This project aims to explore the differences between Lisboa and Porto, the two biggest Portuguese cities. In 2018 these two cities had more tourists than London or Barcelona and have had an incredible growth over the past years. During this project I expect to explore the different neighborhoods in both cities, compare and contrast them and see why they are so popular among tourists.

This project can also be the first step to a new business to decide where to step foot. In which city there is a wider market to explore. The **aim** of this project is to answer the question: Where is the best city to open a new Vegetarian Restaurant, Lisboa or Porto? If we want to open it in Lisboa, what is the best neighborhood for it.

For this project Foursquare data will be used for both cities and in the end a cluster analysis will be made to compare different neighborhoods in Lisbon. *k*-means clustering algorithm will be used to complete this task. Finally, I will use the Folium library to visualize the neighborhoods in Lisbon and their emerging clusters. An exploratory analysis of the type of restaurants in both cities will be made, to explore where is best to open a new vegetarian restaurant.

Data

Based on this problem, the factors that will impact my decision are as follows:
(Same analysis will be done for Lisboa and Porto)

- number of existing restaurants in the city (any type of restaurant)
- number of Vegetarian restaurants in the city, if any.

<u>Focusing on Lisbon's Neighborhoods:</u>

Data sources needed to extract the information to be used in the analysis:

- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- The Lisbon's neighborhoods will be extracted from a dataset available in *Kaggle* from airbnb dataset. I've downloaded that dataset and I'll use it to extract Lisbon's neighborhoods and their geocoordinates.

Methodology

For this project the first step will be to investigate how many vegetarian restaurants there are in both Lisboa and Porto. For this **Foursquare data** will be used and the results will be counted.

Then, focusing only on Lisboa and using the airbnb dataset to obtain the neighborhoods names and coordinates the same analysis will be done, now for each neighborhood. In the end a ***k*-means clustering algorithm** will be used to compare and aggregate the neighborhoods into clusters based on the type of restaurants that already exist. The optimum *k* will be chosen based on the elbow method, which iterates the algorithm through a range of different *k*'s. When plotted, the sum of squared differences is minimum for the best *k*. The decision where to open a vegetarian restaurant will result from the analysis of each cluster.
Also a plot of the Lisbon's map will be provided with the different clusters.

Results

Lisboa and Porto: Restaurants

The first step was to obtain both Lisboa and Porto geo coordinates through **geolocator.**

- The geographical coordinates of **Lisboa** are **38.7077507, -9.1365919.**
- The geographical coordinates of **Porto** are **41.1494512, -8.6107884.**

Using Foursquare and the previous coordinates it was possible to obtain all the restaurants of Lisboa and Porto. The url used and the inputs were as follows:

- search_query = 'Restaurant'
- radius = 6000
- LIMIT = 1000
- url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude_por, longitude_por, VERSION, search_query, radius, LIMIT)

Number and type of restaurants in Lisboa:

```
counts_lis = df_lis.groupby('categories')['name'].nunique()
counts_lis
```

```
categories
African Restaurant          2
Asian Restaurant            2
BBQ Joint                   1
Breakfast Spot              1
Cocktail Bar                1
Himalayan Restaurant        2
Hotel                       1
Indian Restaurant           2
Italian Restaurant          1
Japanese Restaurant         1
Mexican Restaurant          1
Pakistani Restaurant        1
Portuguese Restaurant      20
Restaurant                 11
Seafood Restaurant          1
South Indian Restaurant     1
Steakhouse                  1
```

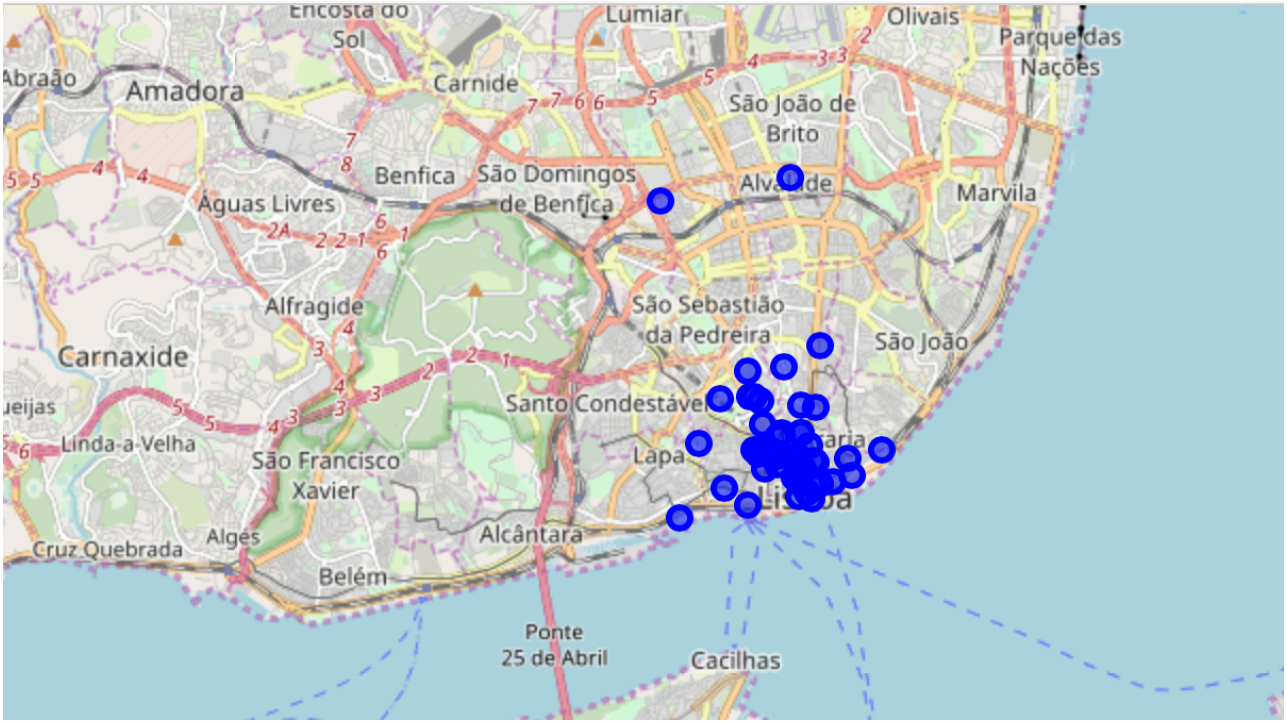
Number and type of restaurants in Porto:

```
counts_por = df_por.groupby('categories')['name'].nunique()
counts_por
```

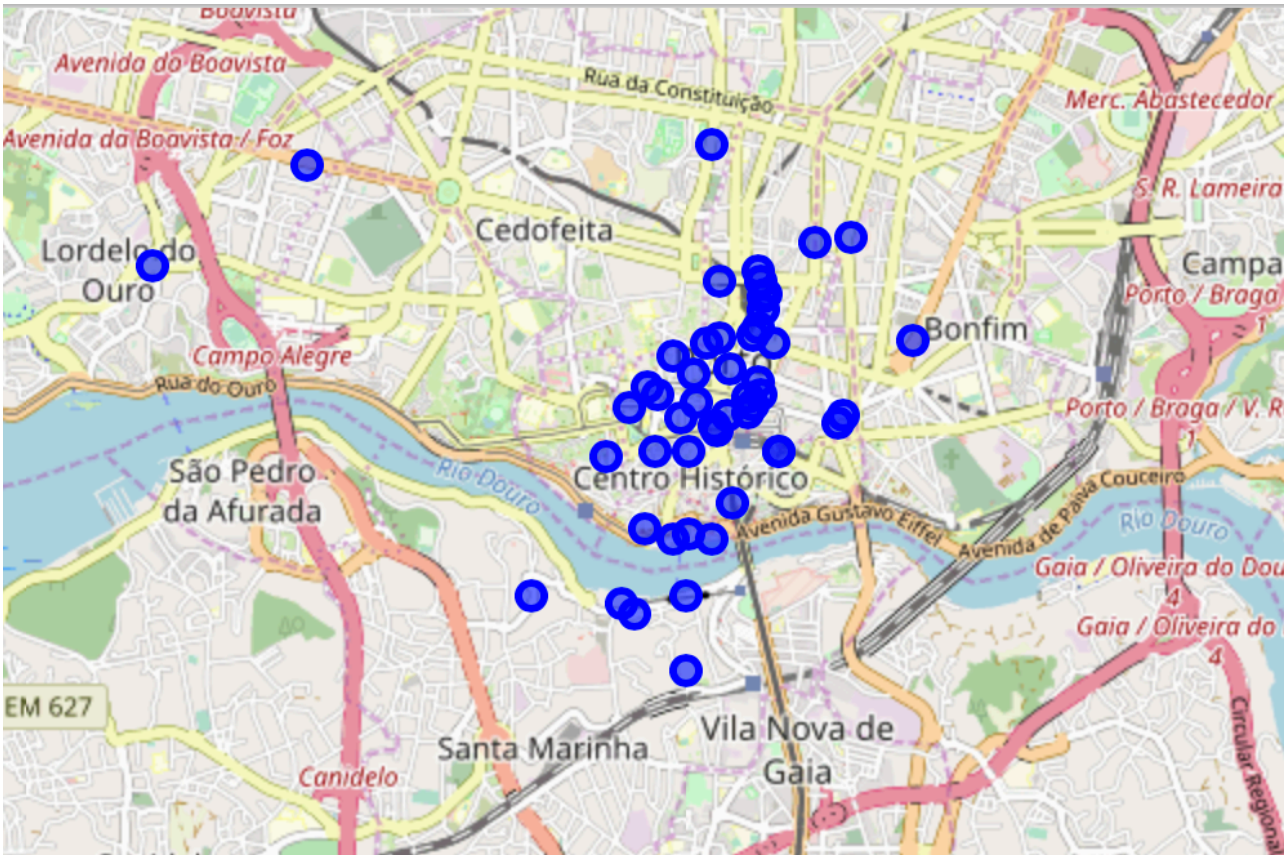
```
categories
Afghan Restaurant            1
Asian Restaurant             1
Bar                          1
Chinese Restaurant           1
Gastropub                    1
Italian Restaurant           1
Mediterranean Restaurant     1
Portuguese Restaurant       28
Restaurant                  10
Seafood Restaurant           1
Tapas Restaurant             2
Turkish Restaurant           2
```

Using folium, is possible to visualize all these in the map:

- Lisboa



- Porto



Lisboa Neighborhoods Cluster Analysis

After the comparison between Lisboa and Porto, let's focus only on Lisboa and its neighborhoods and check where is the best to open a vegetarian restaurant. First it was necessary to explore the airbnb dataset to obtain the names and the geo coordinates of each neighborhood in Lisboa.

Then, the Foursquare database was used again to obtain all restaurants from each neighborhood. The same query was used: "Restaurant".

```
Misericórdia
Belém
Santa Maria Maior
Avenidas Novas
Areeiro
São Vicente
Estrela
Alcântara
Alvalade
Campo de Ourique
Penha de França
Santo António
Olivais
Arroios
Lumiar
Parque das Nações
São Domingos de Benfica
Beato
Benfica
Campolide
Ajuda
Carnide
Marvila
Santa Clara
```

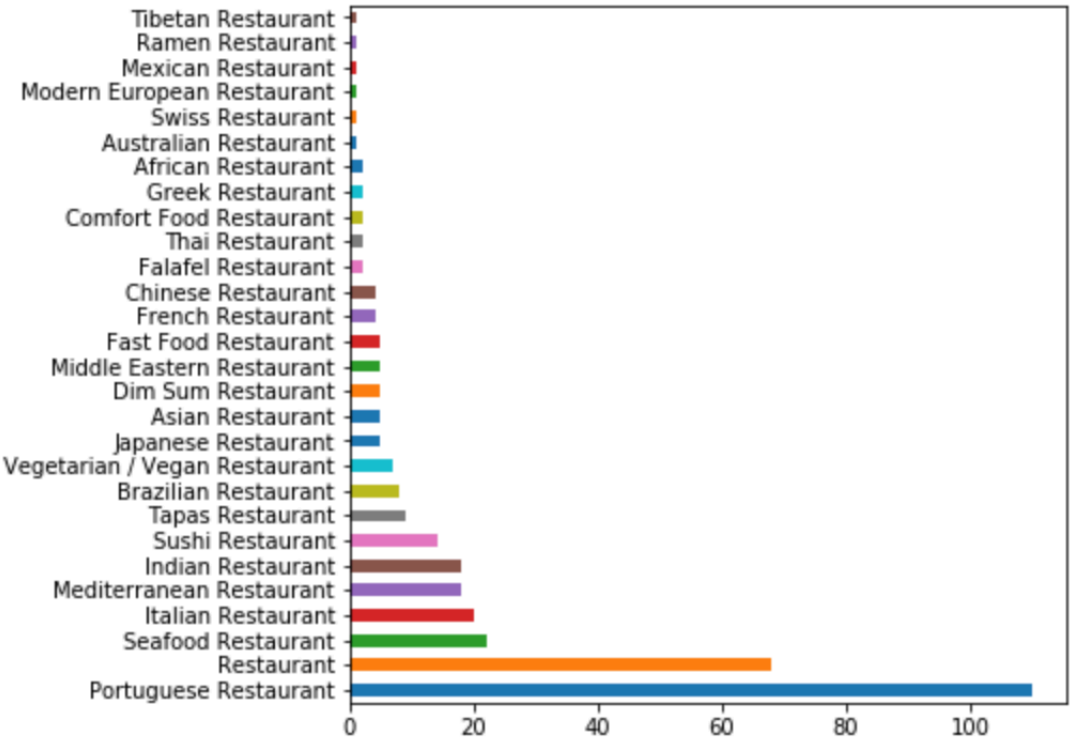


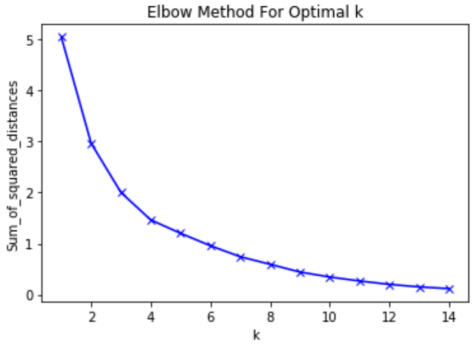
Fig 1: Count of the different types of Restaurants in Lisboa.

According to these counts there are only 7 vegetarian restaurants in Lisboa, the next step will be to check in which neighborhoods they're located and then perform a cluster analysis on all neighborhoods and group them in terms of types of restaurants.

Cluster Analysis

For each neighborhood, the frequency of each type of restaurant was calculated and the most common restaurants per location were obtained.

To perform a *k*-means cluster analysis first we need to find the optimum *k*. For this the elbow method was used in which the sum of squared distances is calculated for a range of *k*, the value that minimizes the sum of squared distances should be chosen as the number of clusters to form.



In this case is not clear where is the end of the elbow but the k=7 was chosen. After the cluster algorithm, the plot of the Lisboa's map with the different clusters formed was plotted.

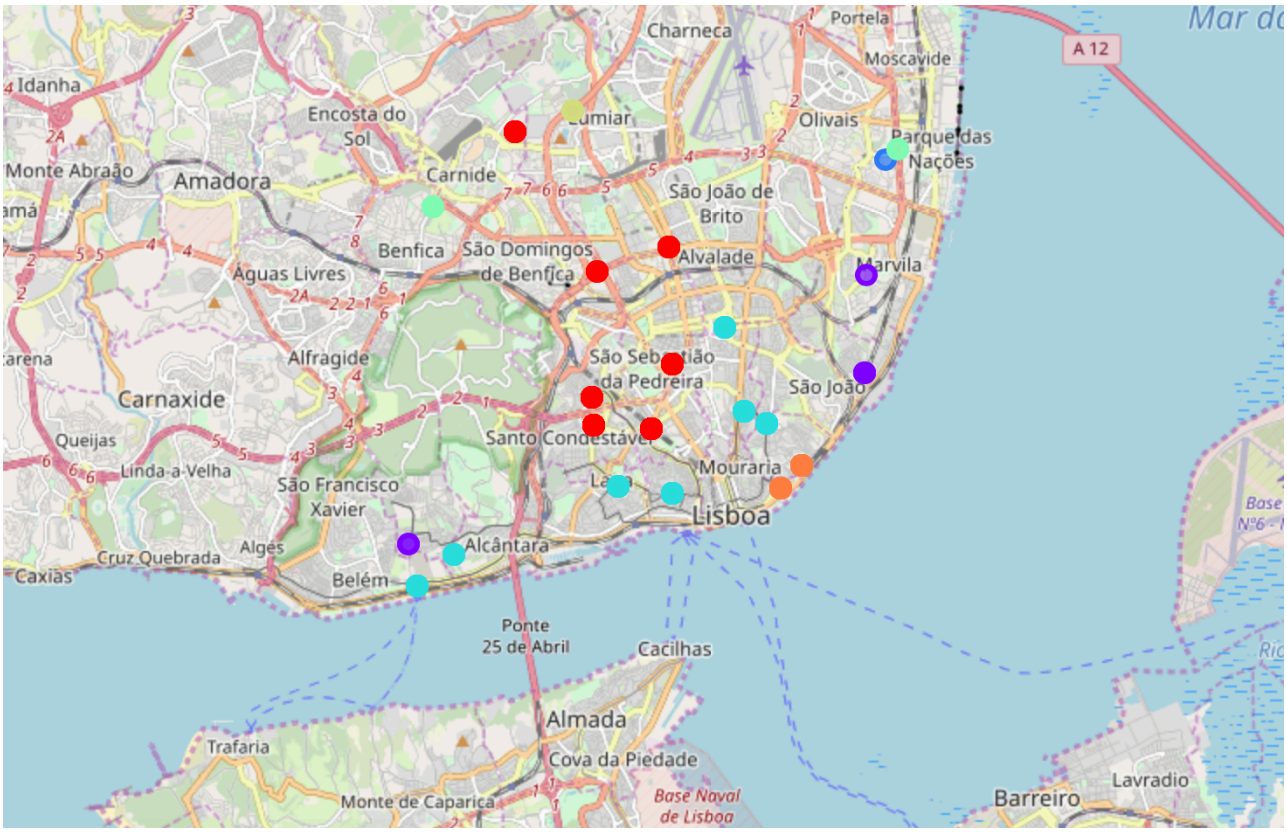


Fig 2: k-means cluster Analysis of Lisboa's neighborhoods's type of restaurants. (k=7)

From the analysis of each cluster, the ones in which there are vegetarian restaurants are cluster 0, 3 and 6.

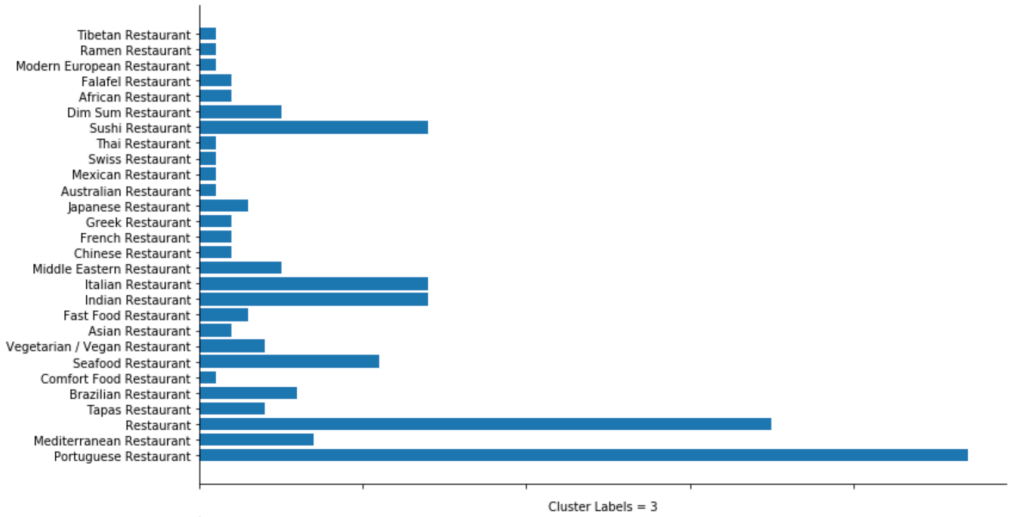


Fig 3: Cluster 3 in detail, counts of each type of restaurant (Neighborhoods: Arroios, Areeiro, Penha de França, Alcantara, Belém, Estrela, Misericórdia)

Discussion and Conclusions

From the first comparison of the two cities it is clear that there are not much vegetarian restaurants in Lisboa nor in Porto. Both cities would benefit from a new vegetarian restaurant, possibly near other venues related with a healthier lifestyle, like gyms or beauty centers. A possible second step would be to look for “health” venues to choose the best place for the restaurant.

From the second analysis, on one hand it is better to open a vegetarian restaurant if there isn’t any restaurant of the same type in the surroundings, on the other hand, the fact that there is a vegetarian restaurant in some neighborhood may indicate that there are costumers that want that type of food. Maybe if looking for health related venues on the same neighborhoods would be the best way to check if there will be clients for it. In Lisboa, specially in those typical neighborhoods with older population the most popular restaurants would be the ones with typical Portuguese food, instead of vegetarian. Perhaps would be best to look for recent neighborhoods with younger people which are more prone to this type of kitchen.

There are still a lot of next studies that could be carried out to narrow the search for the ideal place for a vegetarian restaurant such as costumer evaluation, the venues surrounding or if it is a more tourist place or not.