

# Clustering Restaurants in Lisbon Neighbourhoods

Applied Data Science Capstone by IBM/Coursera

Gustavo Bakker

February 14, 2020

## 1.

### Introduction

**1.1 Background** Lisbon is the capital and the largest city of Portugal, with its population of about 500 thousand people living in the main area and about 2.8 million people in the metropolitan area, is the 10th most populous urban area of the European union. The city has not only native portuguese people but a lot of emigrant people coming from around the world, being a multicultural city that offers different kinds of events, restaurants, shops, concerts, museums, etc.

**1.2 Problem** There are tons of restaurants in Lisbon and they are of different kinds and from different cultures. In this project we will see how certain types of restaurant and its food will vary in different neighbourhoods, with the goal of giving understanding of how this can vary with the variance of the location in a certain city. We will use a clustering technique to separate restaurants by categories in each area.

**1.3 Interest/Stakeholders** This will contribute to the city of Lisbon by providing a valuable information about the kind of restaurants that exist in a certain neighbourhood and also helping entrepreneurs who want to start a new business in town or even future applications that could use this information to recommend different types of food to people.

## 2. Data acquisition and cleaning

### 2.1 Data sources:

We will use the following resources to get our data:

### **Lisbon City Neighbourhoods Dataset**

Source: [https://pt.wikipedia.org/wiki/Lista\\_de\\_freguesias\\_de\\_Lisboa](https://pt.wikipedia.org/wiki/Lista_de_freguesias_de_Lisboa).

This is a table from Wikipedia (portuguese version), that has all the 24 neighbourhoods of Lisbon city, containing information about the population and area for each one.

### **Foursquare API**

Source: <https://developer.foursquare.com/docs>

This is a location data provider consisting of a RESTful API that allows to retrieve information of venues of a certain neighbourhood using coordinates, giving us a rich JSON file with details about restaurants and their locations, provided by the neighbourhoods dataset of Lisbon.

## **2.2 Data cleaning:**

Data collected from the dataset provided by Wikipedia has to be cleaned because it has some information including unnecessary columns and some columns with not so clear names.

There was not so much data cleaning in this study since we were using a geopy and Foursquare API that gave us most of the information we wanted for our simple table, providing the necessary coordinates for the neighbourhoods and also a JSON file containing all the details about venues for our study.

## **3. Methodology**

We are filtering all venues by showing only the restaurant categories so we can see all the restaurants for each neighbourhood.

Using **geopy** library we could retrieve a coordinates like latitude and longitude for each of the neighbourhoods, so then we can use the **Foursquare API** to use those coordinates to get the locals of interest for that area of the map, in a *JSON* file that contains all the details.

Then it is created a dataframe with Neighborhood, Latitude and Longitude details of Lisbon's neighborhoods.

We used a technique of filtering this dataframe by the venue categories that contain the word 'Restaurant', so instead of generalization for food we're only filtering the restaurants.

With a clustering algorithm we can organize restaurants by similarities in different neighbourhoods, and this will be shown in the map provided.

Those clusters were added to a new dataframe in which we associate each cluster number for each neighbourhood along with the coordinates of in the map and the venue category.

## 4. Results

Using some great Python libraries like **seaborn**, it was possible to print out some interesting graphics to show us the most common type of restaurants for each neighbourhood.

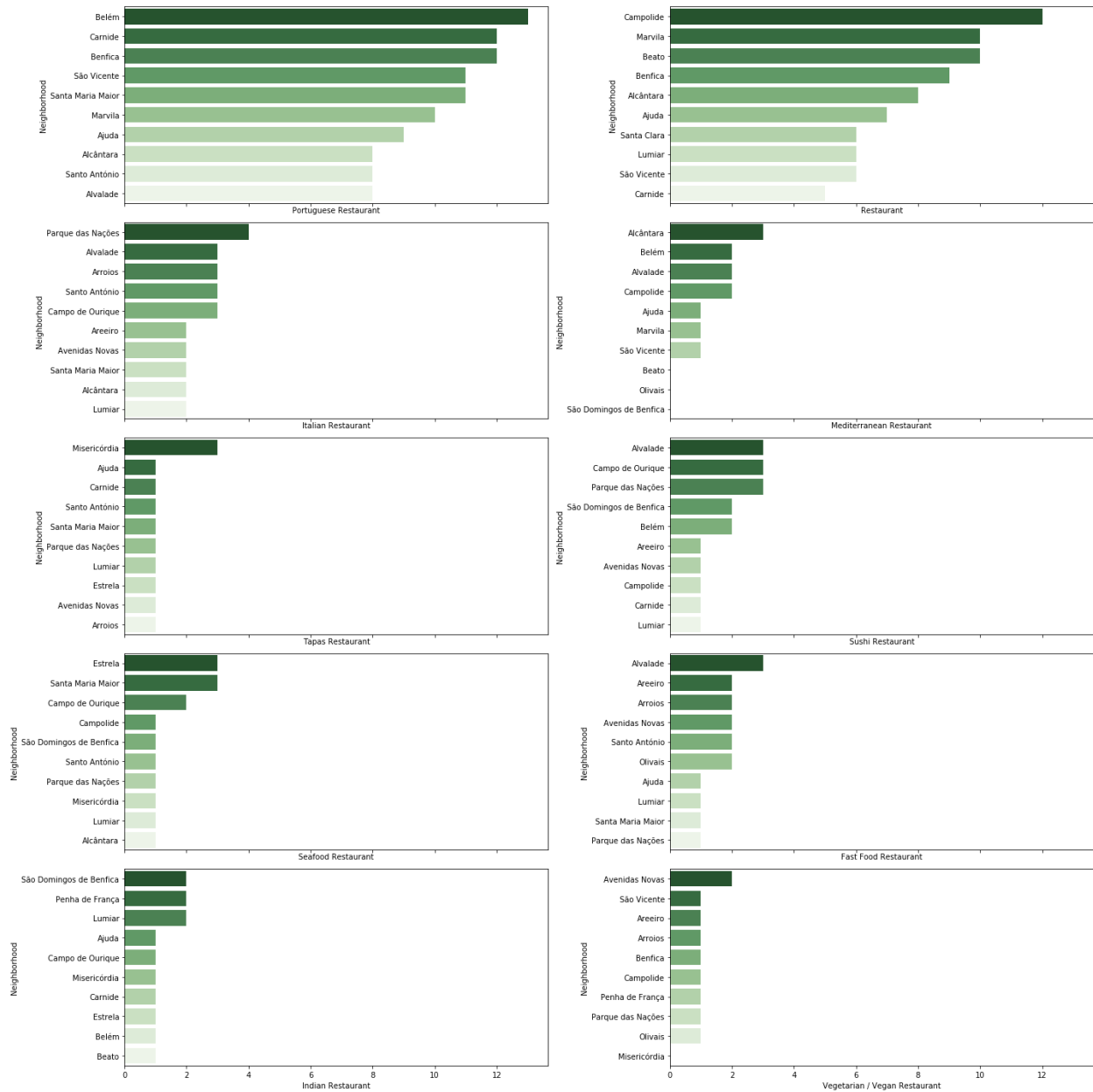


Fig. 1 - EDA showing common restaurants per neighbourhood in Lisbon.

## Clustering - Machine Learning using K-means

For this kind of data in which we do not have such a big data to train and we simply want to find patterns in the map with venues, K-means is a good option since it is a unsupervised learning algorithm that will cluster the different neighbourhoods that have similarities in terms of restaurants.

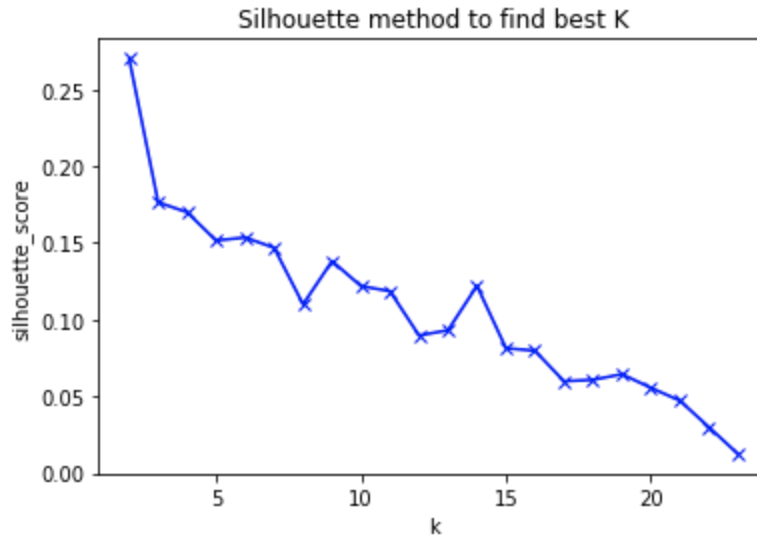


Fig. 2 - Silhouette method to find best K for k-means

To find the best value for the number of clusters K that were needed, we used the Silhouette method, since is a great option and a more reliable one compared to other ones like Elbow method.

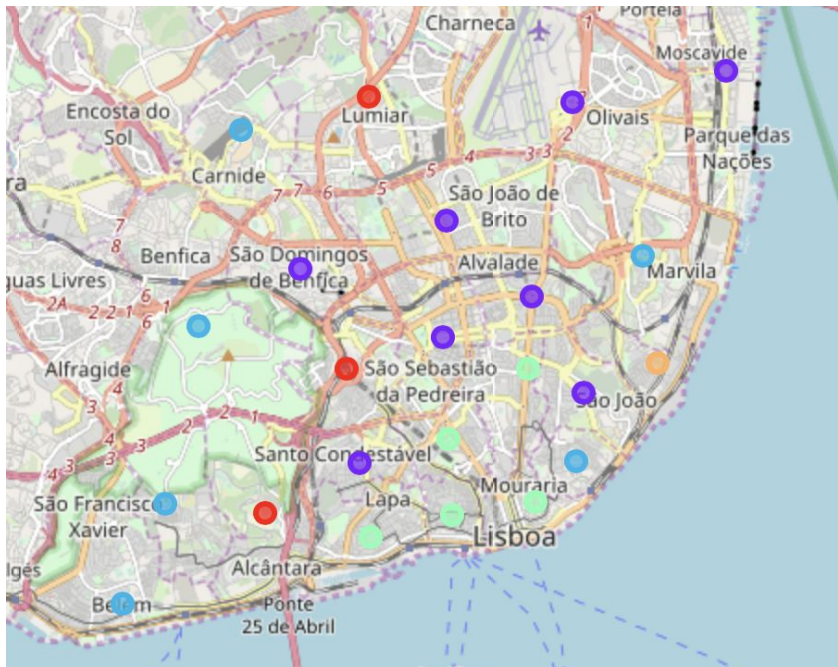


Fig. 3 - Foursquare map showing the 5 clusters for the 24 neighbourhoods in Lisbon

This map above show us the 5 different clusters that we could obtain for the 24 neighbourhoods in Lisbon city area, representing different patterns of restaurants in those areas.

## 5. Discussion

It is interesting to see that the neighbourhood **Belém** holds the first position in the ranking for neighbourhood with the most typical **Portuguese Restaurants**, while, for example, **Parque das Nações** is the one with more **Italian Restaurants**.

**Alcântara** is the best zone for **Mediterranean** foods.

**Misericórdia** for **Tapas**.

**Alvalade, Campo de Ourique and Parque das Nações** are great choices for **Sushi** restaurants.

**Estrela** has a lot of **Seafood** restaurants.

**Alvalade** also has a lot of **Fast Food** restaurants.

**Indian** restaurants can be found in **São Domingos de Benfica**.

**Avenidas Novas** is where **Vegetarian and Vegans** should go.

We can now understand the distribution of restaurants in Lisbon city, and see that most of them are typical Portuguese ones or generic ones. A great conclusion that we had was that after the typical Portuguese restaurants, the Italian ones are the most common ones along with the Fast Food. Next to them, come the Seafood and Sushi restaurants which are also very common in the city.

To have a clear idea of what are the Top 10 most common restaurants in town we can see the table below:

Portuguese Restaurant	182
Restaurant	115
Italian Restaurant	32
Fast Food Restaurant	24
Seafood Restaurant	20
Sushi Restaurant	18
Chinese Restaurant	17
Tapas Restaurant	15
Indian Restaurant	14
Mediterranean Restaurant	12

After this idea of what kind of restaurants we could have in Lisbon city area, let's see what are the important things we found in this study:

The best value for the number of clusters was 5, meaning that we could have 5 different types of restaurant cultures if we put all them together in these 5 categories.

For example, in **Cluster 4** we can see that next to portuguese restaurants, there are a lot of **Seafood restaurants** and also south europe restaurants like **Italian and Spanish (tapas)**.

#### Cluster 4

```
lisbon_merged.loc[lisbon_merged['Cluster Labels'] == 3, lisbon_merged.columns[[1] + list(range(5, lisbon_merged.shap
```

	Population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	31 653	3	Portuguese Restaurant	Italian Restaurant	Chinese Restaurant	Fast Food Restaurant	Indian Restaurant
20	20 128	3	Portuguese Restaurant	Seafood Restaurant	Restaurant	Mexican Restaurant	Middle Eastern Restaurant
28	13 044	3	Portuguese Restaurant	Tapas Restaurant	Italian Restaurant	African Restaurant	Asian Restaurant
46	12 822	3	Portuguese Restaurant	Seafood Restaurant	African Restaurant	Italian Restaurant	Restaurant
47	11 836	3	Portuguese Restaurant	Italian Restaurant	Restaurant	Fast Food Restaurant	Himalayan Restaurant

Fig. 4 - Example of a cluster and its common restaurants

Nevertheless, typical Portuguese restaurants superpass by a great margin, the number of the other restaurants, which seems to make sense.

## 6. Conclusion

With this study our goal is to provide better knowledge of what we could eat in different parts of Lisbon, and make this a starting point for future studies that we could make using other cities. This can help for example, tourists or emigrants to make better choices in terms of food options and give a more personalized experience to people in general by providing recommendations of food by location. This project can be used by mobile apps or websites to recommend and suggest restaurants for each neighbourhood.