Coursera Capstone

IBM Applied Data Science Capstone

Opening a bar in Lisbon, Portugal

Sara Nunes
April 2020

1. Introduction

Portugal has seen a huge increase in tourists in the past few years, especially in Lisbon, the biggest city in the country that has been competing with Porto as the number one destination in Portugal. Lisbon is being seek by tourists for many reasons and one of them is the people searching for a city with a good nightlife. Due to this being a big part of the city´s attractions, an investor would like to open a bar in Lisbon. This investor would like to open a pre-drinking bar which location would be closed to nightclubs and other bars.

2. Business Problem

The objective of this project is to identify and outline the best neighborhoods in the city of Lisbon for nightlife. Using data analysis tools and machine learning algorithms like clustering and data from the city´s venues and districts it is possible to identify and group those neighborhoods into categories.

3. Data

The data needed to solve this problem is:

- List of neighborhoods in Lisbon. We get the list on https://en.wikipedia.org/wiki/Category:Parishes_of_Lisbon and create our neighborhoods data frame.
- Coordinates of the neighborhoods. These are obtained with geopy library based on district names and added to the data frame.
- Information about the nightlife spots in the different neighborhoods. The Foursquare API is used to collect this information.

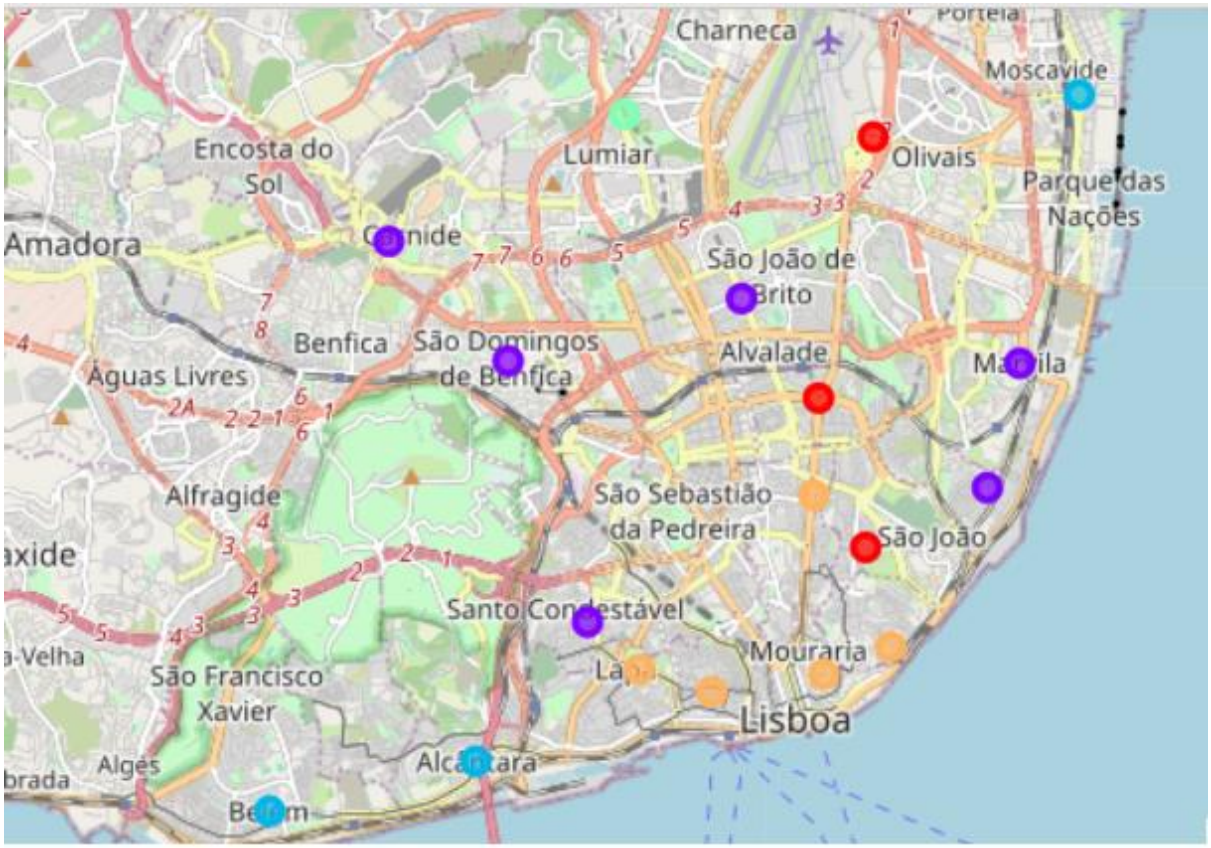
4. Methodology

The goal of this project is to group the different neighborhoods in Lisbon by their nightlife spots and compare them.

For this, it is necessary to analyze the most five most common venues in each district. After K-means clustering is used to put the districts that have similar most common venues together. After, each cluster is analyzed to try to determine which one is more attractive to open the bar.

5. Results

The clusters are located across Lisbon in the following way:



Cluster 0: In these districts we will mostly find nightclubs and different kinds of bars. This would be a good location to open a bar

Cluster 1: In these districts we find mostly bars, tea rooms and breweries. These are places with bars already but no nightclubs so not the best place to open a bar.

Cluster 2: This district has a karaoke bar and some bars and breweries. Not the best place to open a bar

Cluster 3: These districts have bars and some nightclubs and cafés. These would be good locations to open a bar too.

Cluster 4: These districts have lounges, hotel bars and pubs. Not the best place to open a bar since it has to nightclubs

6. Discussion

My analysis shows that districts in Lisbon have different nightlife spots, and therefore it's possible to cluster them in various categories.

I learned that the most common in Lisbon are bars, nightclubs and cafés.

Bars are prevailing in the city center, while restaurants are prevailing outside the city centre. Interestingly, only two clusters showed a good number of nightclubs.

Finally, I'd like to note that the analysis was done on the places returned by Foursquare API with 'nightlife spot' category. Analyzing nightlife spots for other purposed may give a different analysis, but the goal of this project was to focus on places to open a pre-drinking bar, so it was important that it was located near nightclubs.

7. Conclusion

This project help nightlife investors to better understand the city districts from a nightlife perspective. I grouped similar districts into clusters and showed their differences. There's no better cluster, everyone can decide what's best for them, because each cluster has many nightlife spots. The project can be used as part of a bigger study on Lisbon districts to find different investment opportunities using another category, not just bars but many venue categories.