**Analyzing the Neighborhoods in Porto to Start a new Restaurant**

**IBM Applied Data Science Capstone Project**

**19/07/2021**

Carlos Fortes

Conteúdo

[Introduction 3](#_Toc78836170)

[Data 3](#_Toc78836171)

[Discussion 5](#_Toc78836172)

[Conclusion 5](#_Toc78836173)

# Introduction

While opening a restaurant can be a very lucrative business, a lack of demand causes many restaurants to close within the first year of opening. There are many different factors that can account for a restaurant’s success such as location, competition and quality of the food. This is an important question that every business owner must face when choosing whether to open a restaurant or not, as well as location of the business.

To demonstrate the process of picking a location for a client opening a business, the project will focus on answering the following question: “If the client wanted to open an Mexican Restaurant in Porto, what areas are the best options to open the restaurant?” For an Mexican Restaurant, the location and competition are both determined by where the restaurant is opened. If there are too many Mexican Restaurants in the local vicinity, the profitability of the restaurant will be severely decreased.

# Data

To answer the business problem, the following factors must be extracted from various data sources:

• Neighborhoods in Porto and coordinates.

• Number of Restaurants in Each Neighborhood (Foursquare API)

• Number of Mexican Restaurants in Each Neighborhood (Foursquare API)

Methodology

Uma imagem com mesa

Descrição gerada automaticamenteThe first step of the project was to combine the Porto dataset, containing the postal code, borough, neighborhood name, latitude and longitude for each neighborhood in Porto, and the census dataset.

The next step was to visualize the location of the various neighborhoods within Porto to obtain a general understanding the location. As seen on the map, the neighborhoods are densely clustered near downtown Porto and spread out as the distance from downtown increases. This is important because while some postal codes might not have many restaurants, if the area is located near downtown, adjacent regions can heavily impact the profitability of the restaurant. Uma imagem com mapa

Descrição gerada automaticamente

Now that the region has been clearly visualized, the Foursquare API was used to explore each neighborhood and return the top 100 venues within 1500 meters of the longitude and latitude for each neighborhood. The extracted venue categories were encoded using one-hot encoding and the total restaurants and Mexican restaurants in each region were calculated.

Uma imagem com mesa

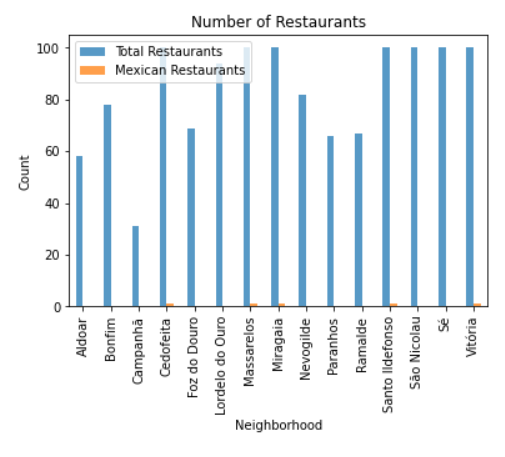
Descrição gerada automaticamenteWith the resulting data Borough name, Latitude, Longitude and Density columns of each region were dropped from the dataFrame. Total number of restaurants and coordinates of restaurants were used to train a kMeans clustering algorithm with 5 clusters (Figure 3). The characteristics of the resulting clusters can be found in Table 1.

|  |  |
| --- | --- |
| Cluster 0 = Red | Cluster 0 Outer city type |
| Cluster 1 = Purple | Cluster 1 Downtown |
| Cluster 2 = Blue | Cluster 2 Inner city Type |

Uma imagem com mapa

Descrição gerada automaticamente

# Discussion

 From the results of the clustering algorithm, it was determined that neighborhoods corresponding to cluster 1 or 2 were the best choice for opening a Mexican restaurant based on the normalized area for restaurants. This narrowed down possible locations into two clusters. Cedofeita, Massarelos and Miragaia region were eliminated due to the large number of restaurants in the area and already having a Mexican restaurant in the area. From the 6 remaining regions, I would recommend that the client open his/her restaurant in either Bonfim or Lordelo do Ouro, as they are further away from downtown making rent cheaper and having less competition nearby.

# Conclusion

Opening a restaurant is a complex task that can lead to a large monetary loss if not done properly. Thus, extensive research about the area would greatly increase the likelihood of the restaurant succeeding. From the project above, I demonstrated the workflow necessary for a client to determine what area the restaurant should open. For specifically, I determined that the optimal location to open an Indian restaurant in Porto should be in the cluster 2 region.