

Applied Data Science Capstone Project

The Battle of Neighborhoods

June 10th, 2020

1. Introduction

The main goal of this project is to gather information in order to recommend a new contractor the best area in the city of Porto to open a hotel.

Porto or Oporto, is the second-largest city in Portugal, one of the Iberian Peninsula's major urban areas. Oporto city has a population of 237,559 and a metropolitan area with 2.4 million people (2019) in an area of 2,395 km² (925 sq mi), making it the second-largest urban area in Portugal. It is recognized as a gamma-level global city by the Globalization and World Cities (GaWC) Study Group, the only Portuguese city besides Lisbon to be recognised as a global city.

Located along the Douro River estuary in northern Portugal, Oporto is one of the oldest European centres, and its core was proclaimed a World Heritage Site by UNESCO in 1996, as "Historic Centre of Oporto, Luiz I Bridge and Monastery of Serra do Pilar". The historic area is also a National Monument of Portugal. The western part of its urban area extends to the coastline of the Atlantic Ocean. Its settlement dates back many centuries, when it was an outpost of the Roman Empire. Its combined Celtic-Latin name, Portus Cale, has been referred to as the origin of the name Portugal, based on transliteration and oral evolution from Latin. In Portuguese, the name of the city includes a definite article: o Porto ("the port"), which is where its English name "Oporto" comes from.

Port wine, one of Portugal's most famous exports, is named after Oporto, since the metropolitan area, and in particular the cellars of Vila Nova de Gaia, were responsible for the packaging, transport, and export of fortified wine. In 2014 and 2017, Oporto was elected The Best European Destination by the Best European Destinations Agency. Oporto is on the Portuguese Way path of the Camino de Santiago. Oporto is also the home of one of the world's most successful football clubs, FC Porto.

Tourism in Porto has grown widely in recent years. The city has to offer not only a spectacular gastronomic but cultural experience. Vast kilometers of beaches and lots of activities that allowed tourists to have fun and relax during their holidays.

a. Business Problem

With all this, the demand and competition in business has also increased. Finding the best are to place a new hotel in town by answering the following:

- Areas with the most hotels;
- Areas with more restaurants;
- Areas with more places of interest to visit.

2.Data Section

a. Data sources and data cleaning

The dataset utilized for this project contained information about the different Boroughs and Counties in Portugal, as well as their geographical coordinates (Latitude and Longitude). The datasets can be found in the following repository: <https://simplemaps.com/data/pt-cities>. Due to the size of the country and their Boroughs it was decided to analyze data in a parish level and not neighborhoods, as the size of a neighborhood in Portugal wouldn't show any significant results.

In order to obtain data about the Porto's parishes coordinates the library Geopy and the data junction was done in excel file and imported to IBM as a dataset.

The venues information in Porto were obtained through the Foursquare API and merged with the master dataset.

3.Methodology

- Several dependencies and libraries were utilized in this project, such as: Numpy, Pandas, Geopy, Matplotlib, Seaborn, Sklearn, Scipy and Folium.
- Gather information about Porto district, which has 7 parishes. The master data included the following parameters: Borough, Parish, Latitude and Longitude.

- The Foursquare API and explore Porto's venues (limit of 100 venues radius=1000) - Top 5 venues per Parish.
- Clustering the Parishes - use of folium maps to present different clusters.
- Parish with more venues - two parishes of interest were further investigated.
- Hotels, food places and places to visit per parish of interest.

4. Results

a. Dataframe with the master data

The master was transformed in a dataframe and contained information about the seven parishes in Porto Borough as we can see in Table 1.

Table 1 - Master dataset containing the geospatial information about the parishes in Porto.

	Borough	Parish	Latitude	Longitude
0	Porto	Bonfim	41.151070	-8.593957
1	Porto	Campanhã	41.159674	-8.576985
2	Porto	Paranhos	41.173513	-8.605900
3	Porto	Ramalde	41.169962	-8.643317
4	Porto	União das freguesias de Aldoar, Foz do Douro e...	41.161152	-8.675639
5	Porto	União das freguesias de Cedofeita, Santo Ildef...	41.140242	-8.615174
6	Porto	União das freguesias de Lordelo do Ouro e Mass...	41.151706	-8.641936

b. Geographic details of Porto and its parishes

The Folium library for Python was used in order to obtain the geographical details about Porto and its parishes by using latitude and longitude of each one of them. (Figure 1).

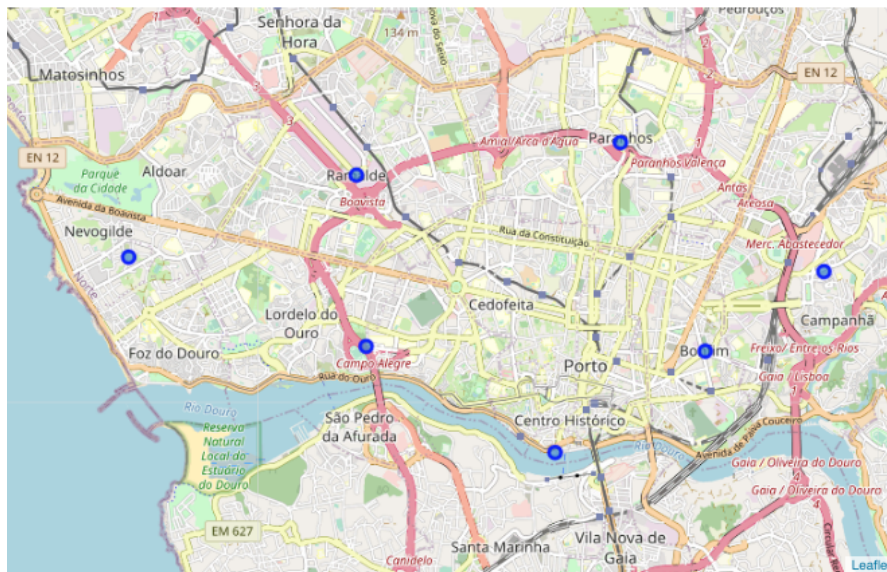


Figure 1 - Geographic distribution of parishes in Porto borough.

c. Porto venues with Foursquare API

The parishes were explored by using the Foursquare API with a limit of 100 venues per parish and the radius of 1000 meters. This information was merged with the initial master dataset as we can see in Table 2.

Table 2 - Venues information per parish.

	Parish	Parish Latitude	Parish Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bonfim	41.15107	-8.593957	Mesa 325	41.148224	-8.597471	Coffee Shop
1	Bonfim	41.15107	-8.593957	Suribachi Centro Macrobiotico	41.149043	-8.597248	Vegetarian / Vegan Restaurant
2	Bonfim	41.15107	-8.593957	A Cozinha do Manel	41.146196	-8.592846	Portuguese Restaurant
3	Bonfim	41.15107	-8.593957	Eurostars do Heroísmo	41.146280	-8.593437	Hotel
4	Bonfim	41.15107	-8.593957	Manuel Alves	41.156655	-8.594456	Restaurant

This information was organized so we could observe the most common venues categories per parish (Table 3).

Table 3 - Top 5 venue category per parish.

	Parish	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bonfim	Restaurant	Café	Portuguese Restaurant	Bakery	Hotel
1	Campanhã	Clothing Store	Coffee Shop	Café	Sandwich Place	Restaurant
2	Paranhos	Supermarket	Café	Bar	Bakery	Portuguese Restaurant
3	Ramalde	Café	Supermarket	Soccer Field	Portuguese Restaurant	Bakery
4	União das freguesias de Aldoar, Foz do Douro e...	Beach	Restaurant	Café	Bakery	Bar

From further analyzes it was observed 136 unique categories were retrieved by Foursquare API in the given conditions.

d. Clustering parishes

As we could see there are some common venue categories in each parish, for this reason the K-means algorithm was used in order to cluster the parishes.

The Elbow Method was used in order to obtain the optimal k (number of clusters), but unfortunately no major conclusions were taken from it, most probably because K-means is highly sensitive to preprocessing. If one attribute is on a much larger scale than the others, it will dominate the output (Chart 1).

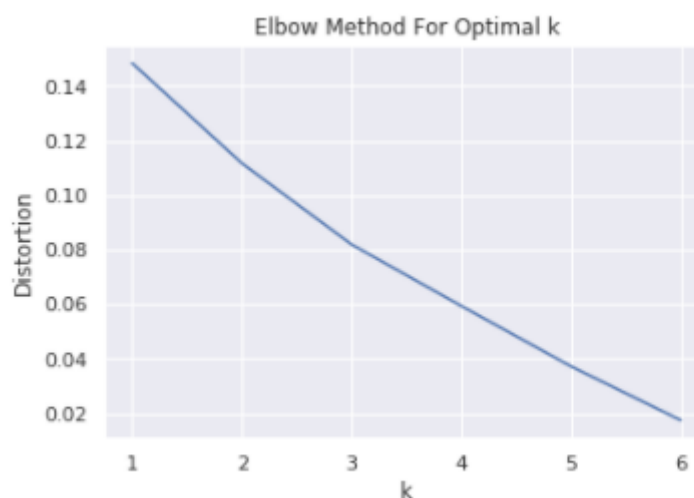


Chart 1 - Elbow method for optimal k.

Four clusters were considered and by observing the number of 1st Common Venue in each cluster we could categorize them (Chart 2) and then were inserted with the parish coordinates so we could see their location (Figure 2).

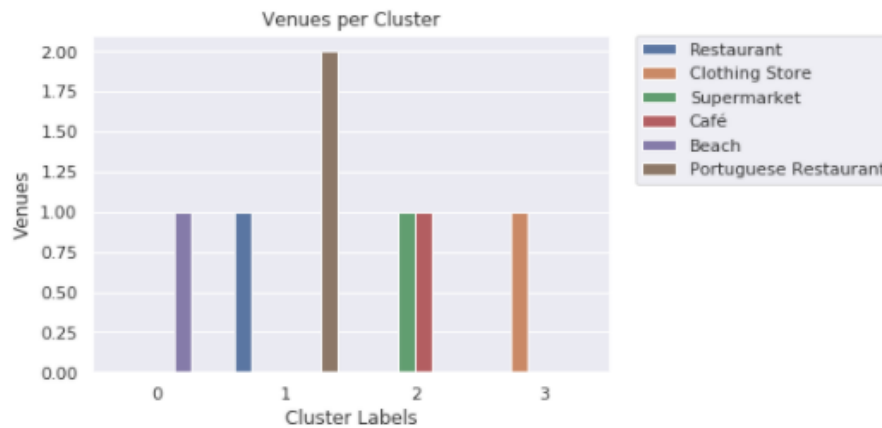


Chart 2 - Venues category per cluster.

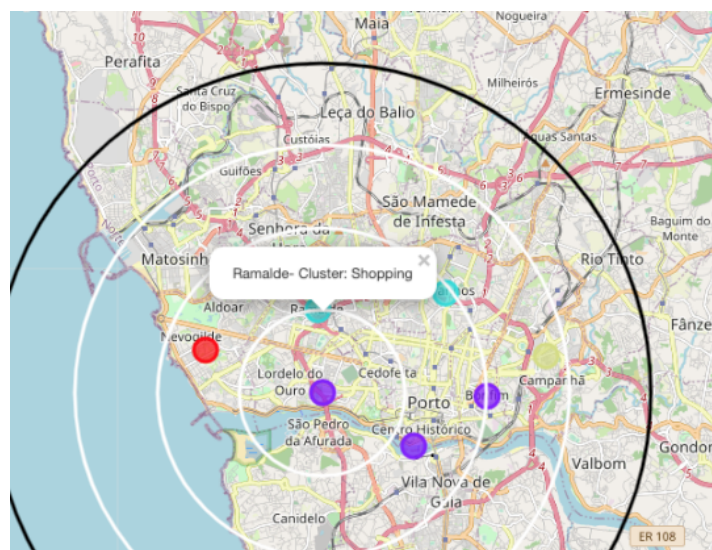


Figure 2 - Cluster's categories.

e. Venues per parish

From the Chart 3 we can see that **União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória** and **Bonfim** are the parishes with the most venues. The two parishes will be considered as the parishes of interest and will be studied in more detail.

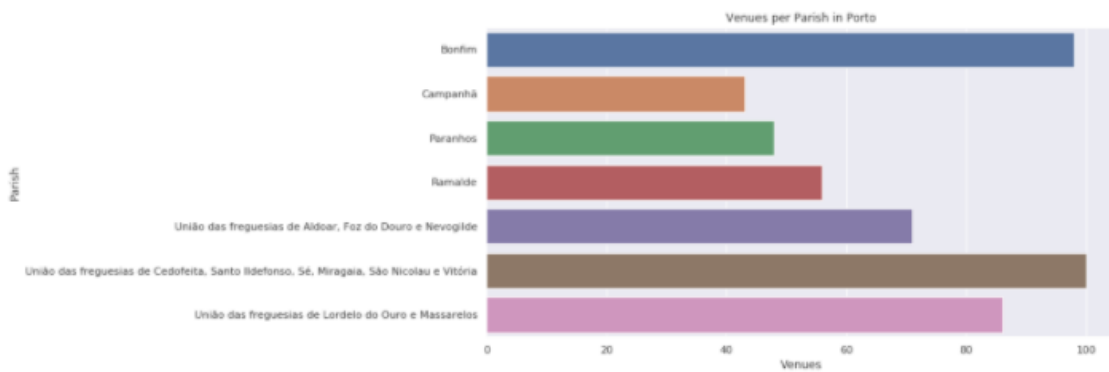


Chart 3 - Venues per parish.

By using Folium Heatmap (Figure 3) we could observe the main focus of the venues in the two parishes of interest : **União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória** and **Bonfim**.



Figure 3 - Heat Map of venues in União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória and Bonfim.

União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória:

In the below chat we can see that this parish is:

- Strong variety of restaurants, more prominent with Portuguese cuisine;
- Places of interest to visit such as: churches, Street Market, Historic sites, Parks, Monuments, train station, event spaces and wine shops;
- Accommodations options such as: hotel and hostels.

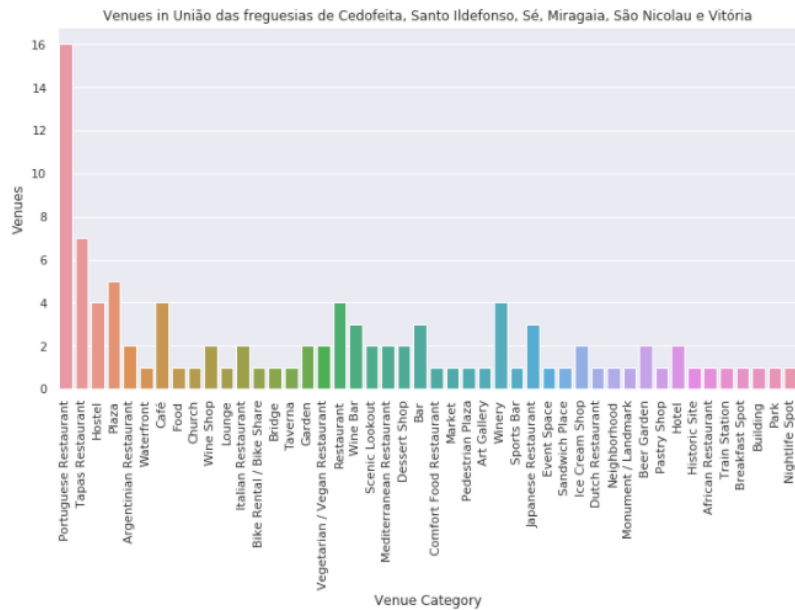


Chart 4 - Venues in União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória.

Bonfim:

In the below chat we can see that this parish is:

- Strong variety of restaurants, also prominent portuguese cuisine;
- Rich in Bakeries and coffee shops;
- Places of interest to visit such as: churches, Street Market, Theater, Dive Bar;
- Accommodations options such as: hotel, hostels and motels.

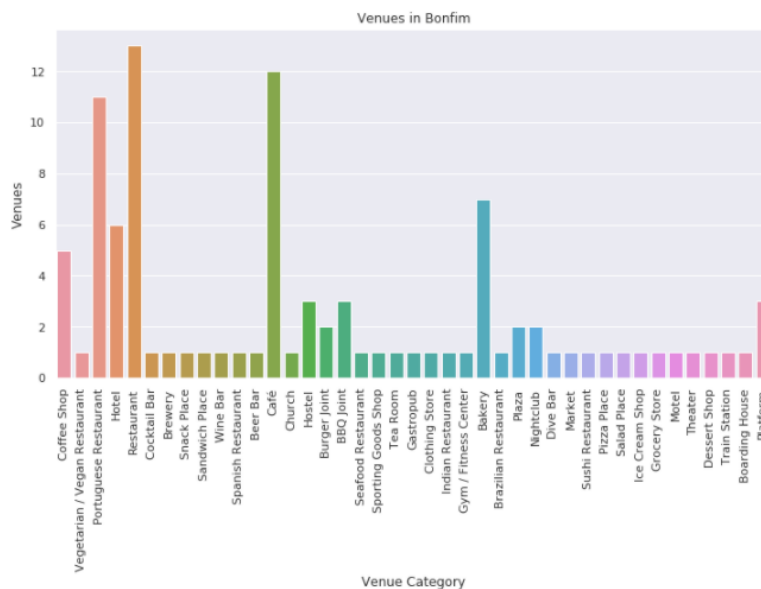


Chart 5 - Venues in Bonfim.

f. Venues of interest

Accommodation per parish:

From the barchart and the heatmap below we can see that Bonfim is the parish with the focus of accommodation options, such as hotels, hostels and motels.

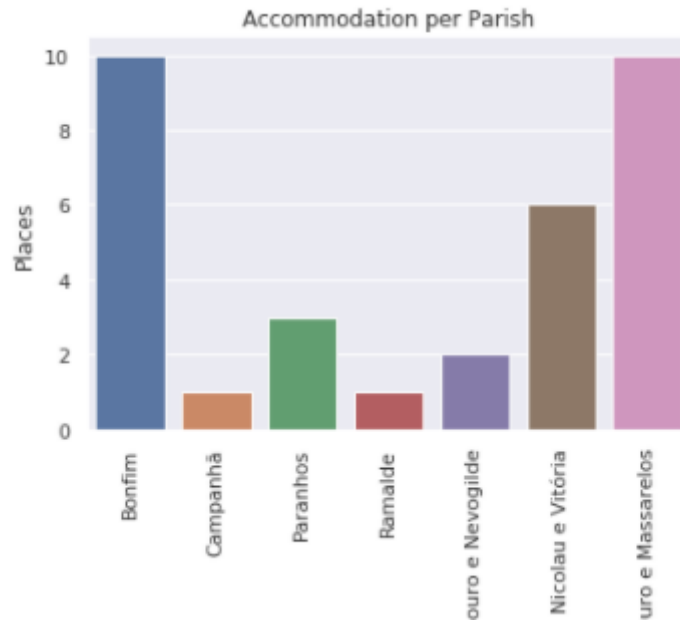


Chart 6 - Accommodation per parish.

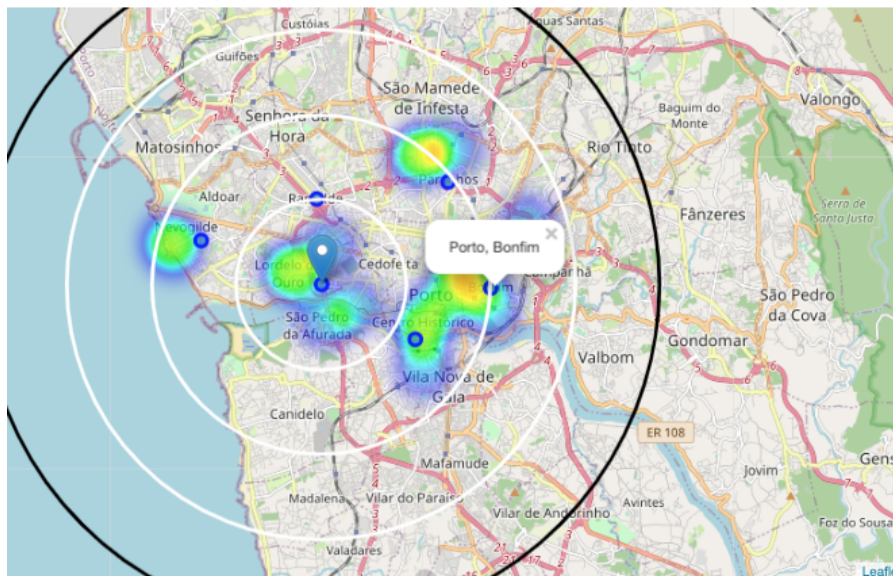


Figure 3 - Heatmap with accommodations distribution in Porto.

Food places per parish:

From the below chart we can see that União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória and Bonfim are the focus areas for food places.

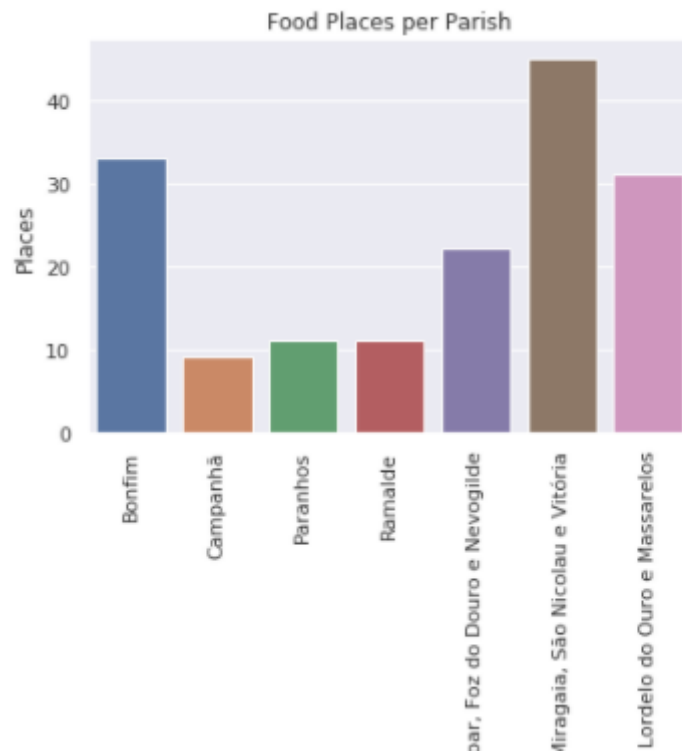


Chart 7 - Food places per parish.

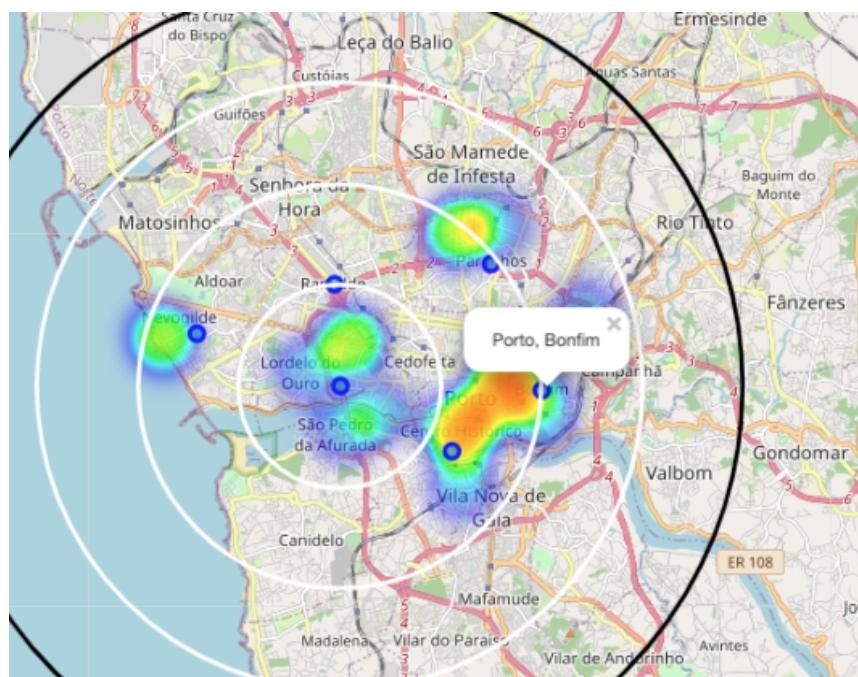


Figure 4 - Heatmap with food places distribution in Porto.

Places to visit:

From the barchart and heatmap below, we can see that União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória is the focus of interest places to visit.

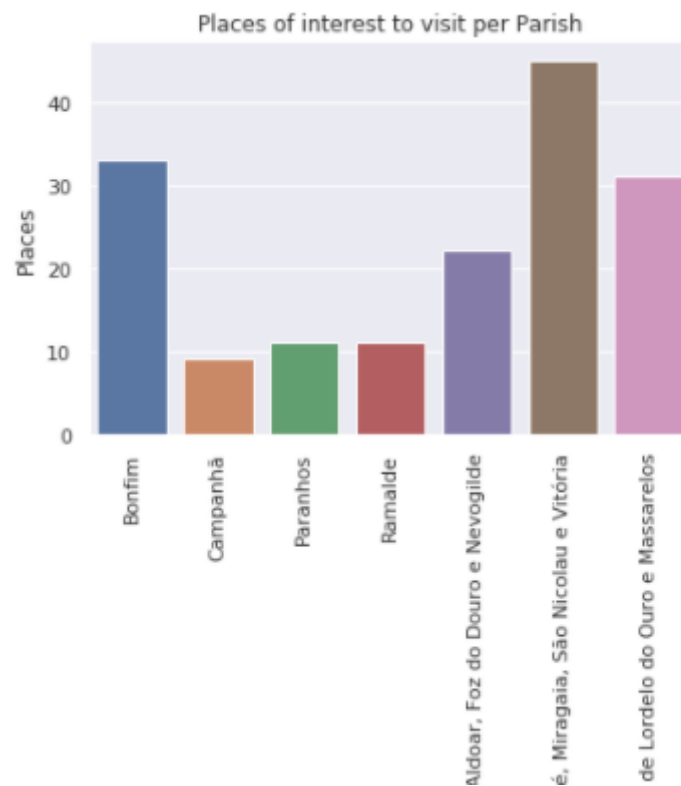


Chart 8 - Places of interest to visit per parish.



Figure 4 - Heatmap with places of interest distribution in Porto.

5. Discussion and Conclusions:

The areas taken into consideration were Parishes and not neighborhoods due to the country size and the distribution in general.

Porto is a Borough and a County, so another valid option would have been to check another borough from Porto County, such as Vila Nova de Gaia.

The two Parishes analyzed in detail, both shown to be good options to place a new Hotel, but of them have a huge amount of different restaurants when compared with other parishes. União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória area would be most probably the better area because competition would be lower.