

---

# IBM Applied Data Science Capstone

— The best parish in Porto to place —  
a new hotel.

---

# Introduction

The main goal of this project is to gather information in order to recommend a newly contractor the best area in the city of Porto to open an 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 km<sup>2</sup> (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.

# Business Problem:

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.

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

1. Areas with the most hotels
3. Areas with most restaurants
4. Areas with the most museums/attractions to visit

# Data Section:

For this project we will be using the following data:

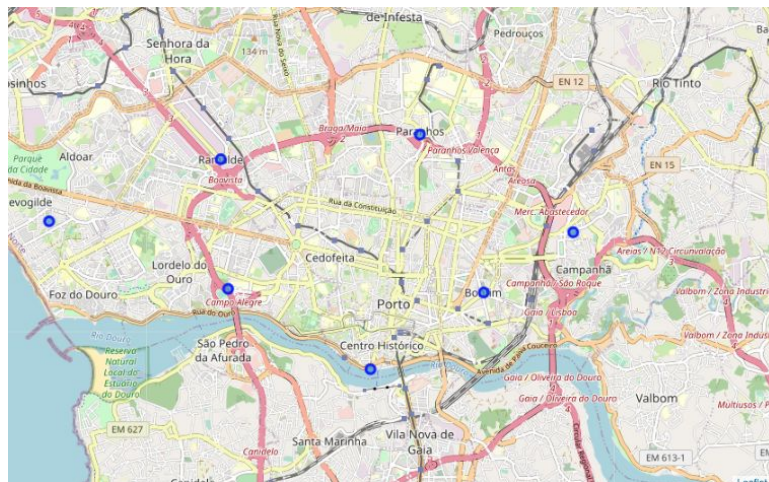
- Portugal's data with the different Boroughs, latitude and longitude.
  - Data source with Portuguese boroughs and geospatial information:  
<https://simplemaps.com/data/pt-cities>
- Venues information in each neighborhood of Porto:
- Data source: Foursquare API

# Methodology:

1. Gather information about Porto district, which has 7 parishes. My master data included the following parameters: *Borough, Parish, Latitude and Longitude*;
2. Call Foursquare API and explore Porto's venues (limit of 100 venues radius=1000) - Top 5 venues per Parish;
3. Clustering the Parishes - use of folium maps to present different clusters;
4. Parish with more venues;
5. Hotels per Parish of interest;
6. Food places per Parish of interest;
7. Places to visit per Parish of interest.

# Master data - Porto Parishes:

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

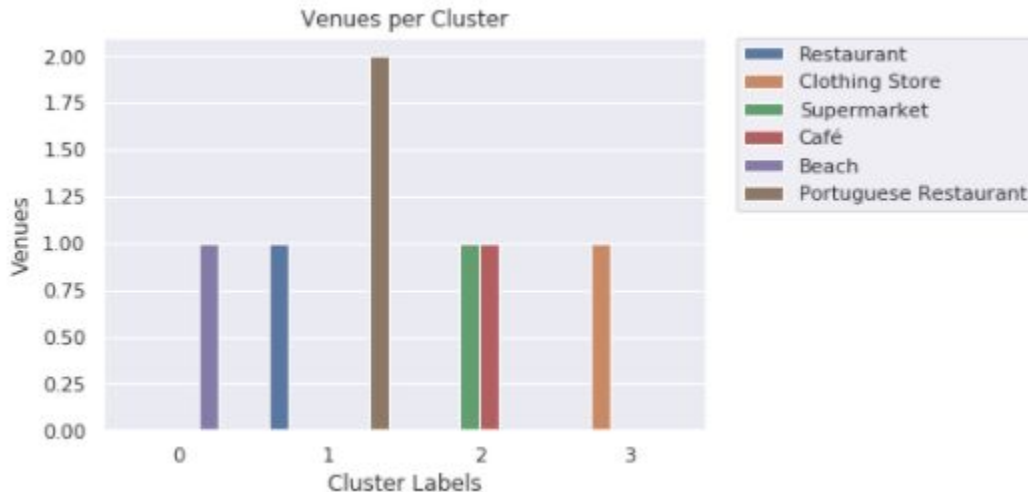


# Top 5 venues per Parish:

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

# Clustering the parishes:

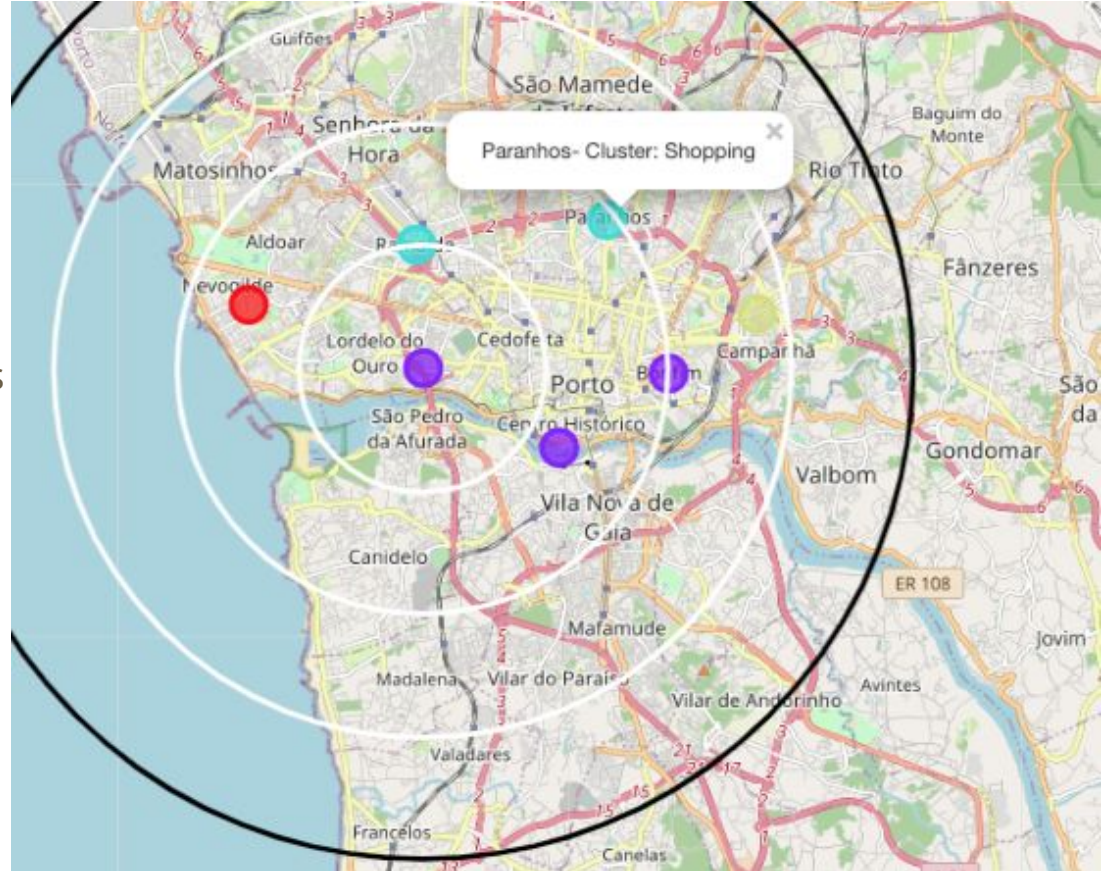
- The data was divided in 4 clusters and it was further analyzed in order to understand the similarities between the venues in each cluster.
- The cluster were classified considering their venues.





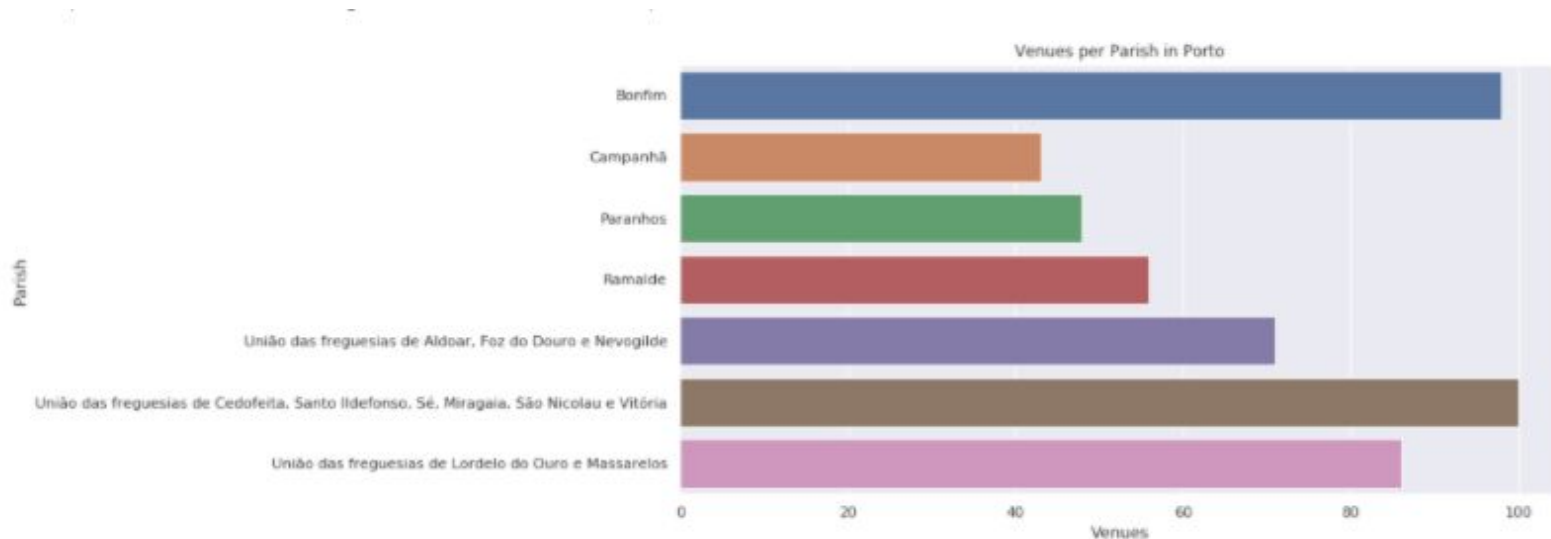
# Cluster categories:

1. Cluster 0 - Beach
2. Cluster 1 - Restaurants
3. Cluster 2 - Coffee Shops and Groceries
4. Cluster 3 - Shopping



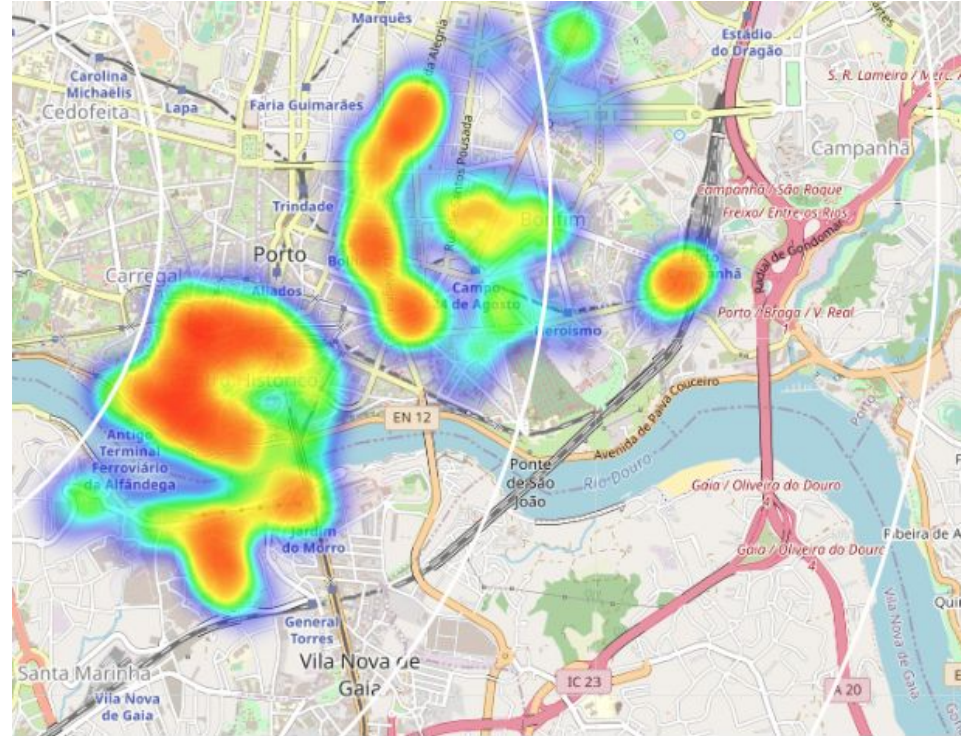
# Parishes with more venues:

In order to know the most interesting area, the distribution of venues in each parish was checked. From previous chart 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.



# Parishes with more venues:

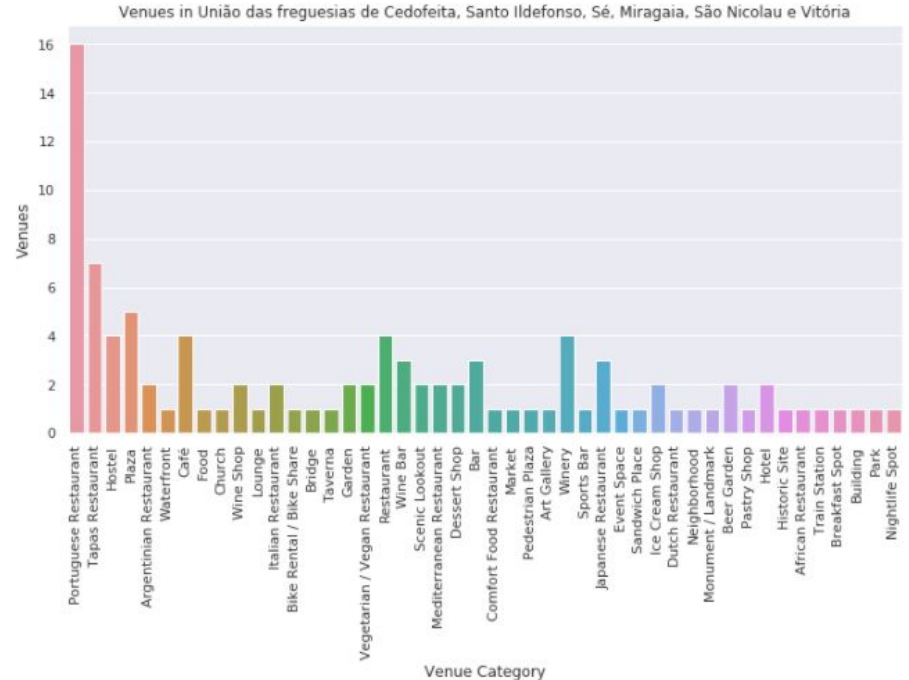
Here we can see a Heatmap of the venues in these two Parishes. It is possible to see that both them are really close to the Historic centre of Porto, so both could be potential areas to place a new Hotel.



# Parish with more venues:

From the chart of venues in **União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória** we can see that:

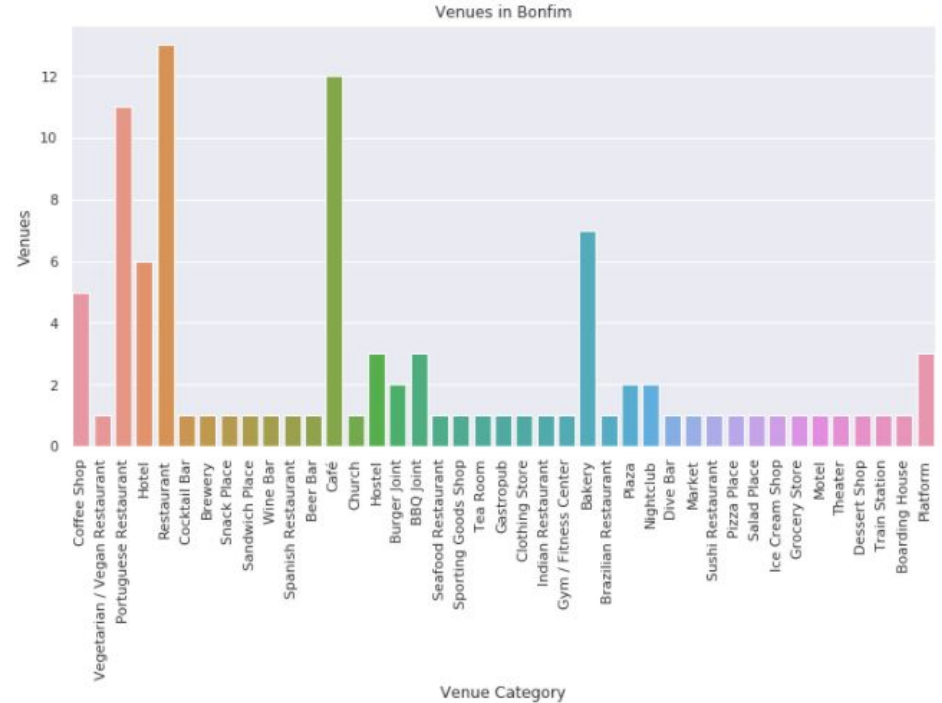
- Strong variety of restaurants, more prominent with Portuguese cuisine;
- Places of interest to visit such as: churches, Street Market, Historic sites, Parks, Monuments, train station, event spaces and wine shops;
- Accommodations options such as: hotel and hostels.



# Parish with more venues:

From the chart of venues in **Bonfim** we can see that:

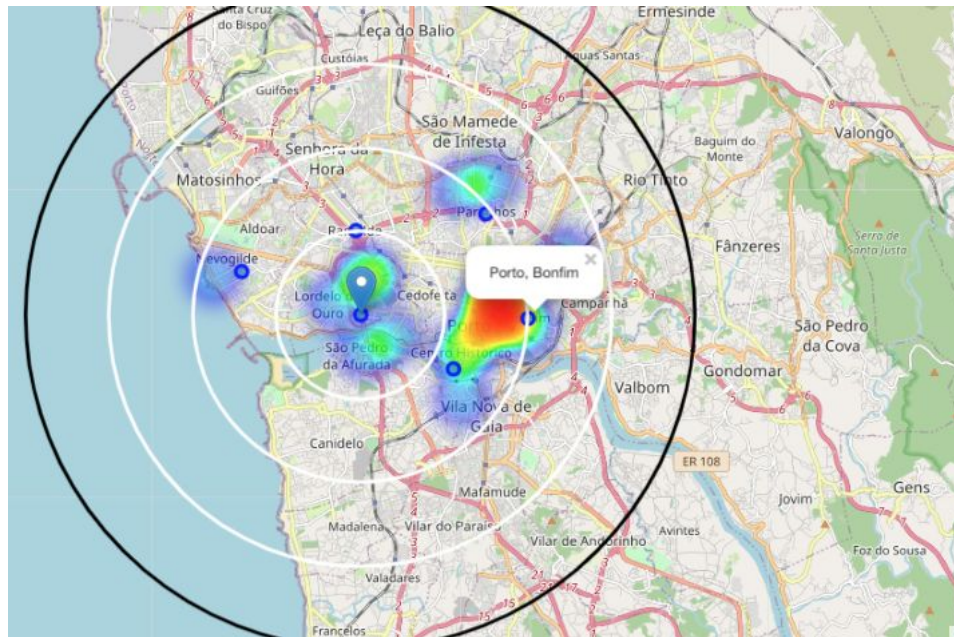
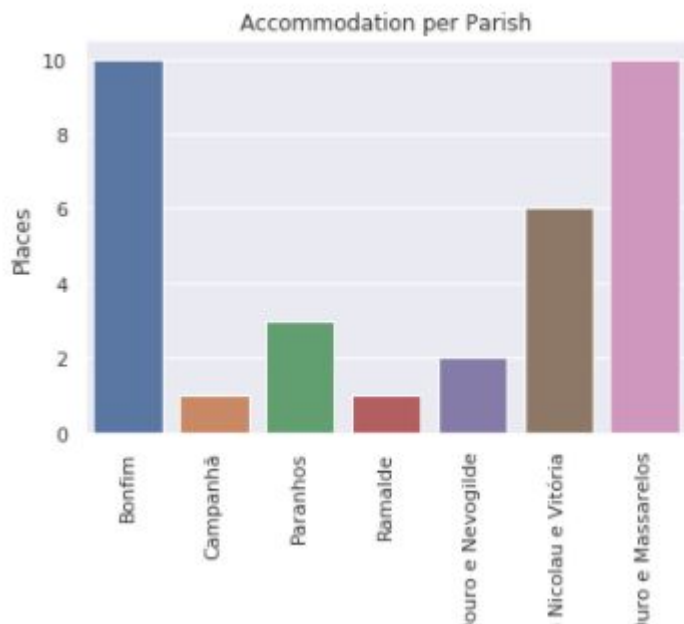
- Strong variety of restaurants, also prominent portuguese cuisine;
- Rich in Bakeries and coffee shops;
- Places of interest to visit such as: churches, Street Market, Theater, Dive Bar;
- Accommodations options such as: hotel, hostels and motels.





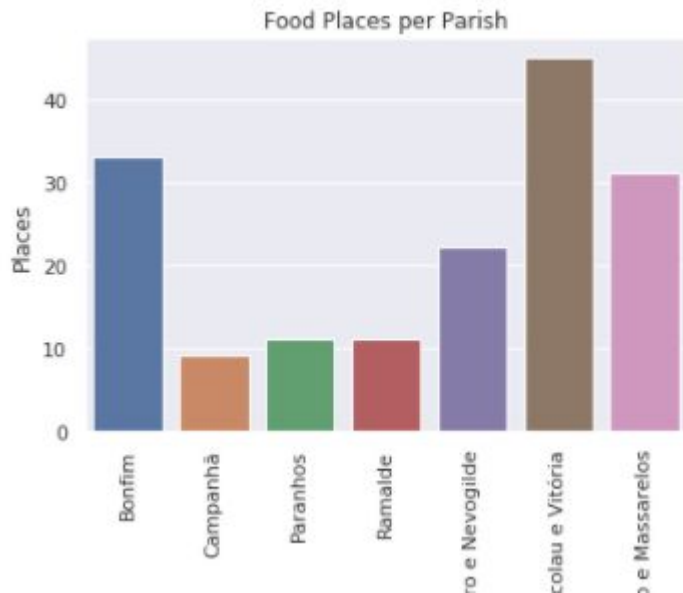
# Parishes with more venues - Accommodation:

From the charts below we can see that **Bonfim** is the focus of the accommodations areas in Porto:



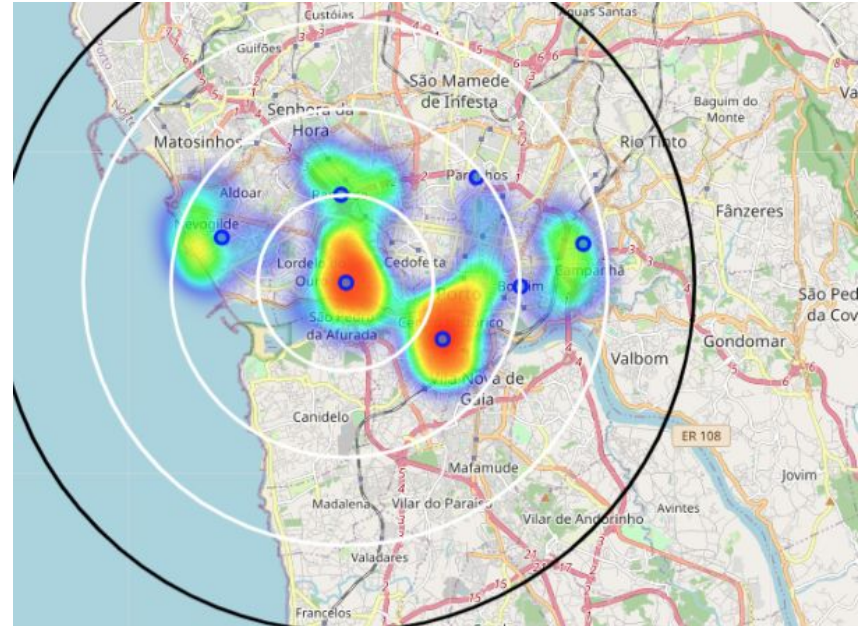
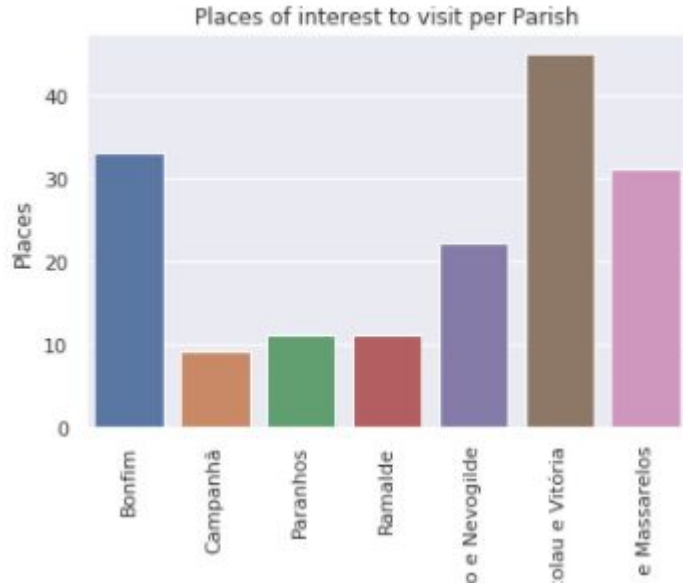
## Parish with more venues - Food places:

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.



# Parish with more venues - Places to visit:

União das freguesias de Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória is the focus of interest places to visit.





# Discussion:

- The Elbow Method was performed in order to find the optimal K, but unfortunately, I was unable to take any conclusions 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. For that I have tested different K and check which would make more sense.
- Two parishes of interest were selected and further analyzed because of the number of venues, hotels and proximity with the city center.

# 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 boroughs from Porto County.
- 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.