Opening a Sushi Restaurant in the city of Oporto

IBM Data Science Professional Certificate Project Capstone

João Miguel Freitas da Silva

April 2021

1. Introduction

The hypothetical business problem of this project is making a reasoned decision about where to open a sushi restaurant in the city of Oporto in Portugal. The city of Oporto is very trendy at the moment and is a touristic boom. On top of that there is a youthful culture regarding new experiences inherent to the city, which can lead to a great curiosity and openness towards experience non-portuguese food. The intent of this project is finding an optimal location area for the restaurant, taking into account accessibility, concurrence and distance from hotels, schools and universities. To better understand the characteristics of the possible restaurant location I'm going to subdivide the city of Oporto into its official districts and then perform unsupervised machine learning to create clusters of districts based on the features described above. The description of the different Oporto's areas can be helpful to guide the line of thought of potential investors in this particular city.

2. Data Acquisition

The data needed for this project is:

1. List of districts of Oporto
2. Coordinates (Latitude and Longitude) of the districts
3. List of several kinds of venues for each district
4. Most popular venues for each district

The list of districts of Oporto was obtained exploring the city council website. The coordinates of each district were obtained with the help of Google Maps. Finally, the list

of different kinds of venues and the most popular venues for each district was obtained through a Foursquare API.

3. Methodology

3.1 Oporto Districts

Based on the Oporto's city council, the city has seven districts:

1. Union of districts of Aldoar/Foz do Douro/Nevogilde
2. Union of districts of Cedofeita/Santo Ildefonso/Sé/Miragaia/São Nicolau /Vitória
3. Union of districts of Lordelo do Ouro/Massarelos
4. Bonfim
5. Campanhã
6. Paranhos
7. Ramalde

Note: For now on I refer to the second district as city center.

The figure below is a representation of the districts named above,

Figure 1: Oporto's Districts

In order to represent each circle, I created circles of radius proportional to the size of each district in terms of area. The coordinates of each district were obtained using Google Maps. The resultant *dataframe* is shown in the figure 2, followed by a visualization of the districts of Oporto, figure 3.

	Districts	Latitude	Longitude	Radius
0	Aldoar/Foz/Nevogilde	41.161290	-8.667521	1500
1	Cedofeita/Ildefonso/Sé/Miragaia/Nicolau/Vitória	41.151764	-8.613740	1300
2	Lordelo/Massarelos	41.156975	-8.639555	1300
3	Bonfim	41.152498	-8.588829	1100
4	Campanhã	41.160127	-8.568764	1300
5	Paranhos	41.174884	-8.607466	1600
6	Ramalde	41.177788	-8.638421	1500

Figure 2: *Dataframe* of the districts

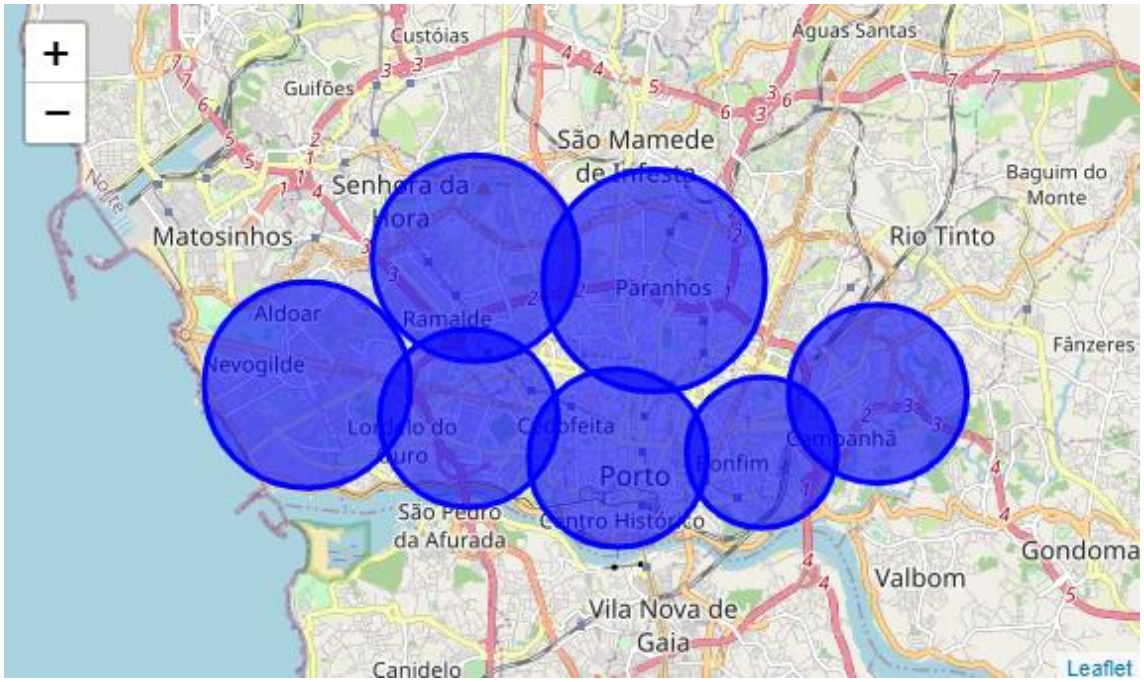

Figure 3: Visualization of the districts

3.2 Exploring the districts

In order to decide about the optimal district to open a sushi restaurant I explored different types of venues that I found relevant for each district, namely,

1. Sushi Restaurants
2. Hotels
3. Schools and Universities
4. Means of Transport

5. Popular Venues

The number of sushi restaurants in each district is relevant to analyze he concurrence. The number of hotels is useful has a measure of how touristic a district is. Furthermore, people who are hosted in a hotel are likely to look for lunch or dinner in a restaurant outside the hotel. Schools and Universities are a good indicator for potential clients, since younger people is generally more open to experience new types of cuisine and to follow trends and, sushi is certainly a food trend right now. On top of that, schools and universities can lead to group lunches, and sushi is a great option for this particular case. The accessibility of the sushi restaurant location is also very important, that's why I decided to explore the means of transport of each district. Finally, the popular venues are a good indicator to measure the most popular venues, which can be a crucial indicator to decide where to open the restaurant. In addition, more than just search for the popular venues, I also search for popular sushi related venues for each district.

For each venue type I calculated the density per district. The density per district is simply the total number of a certain venue in a certain district divided by the radius of the district, which is an indicator of its size. This parameter is important to normalize the venues analysis since the districts have different areas.

3.3 Clustering the districts

After the districts exploration I performed unsupervised machine learning using a K-Means Algorithm, in order to create groups of districts. The features used to cluster the districts were the density of sushi restaurants, density of hotels, density of Schools and Universities and Density of Transports. To estimate the best number of clusters I used the elbow method. Applying this method, one chooses the value of *k* (number of clusters) associated with a great decrease in slope for the distortion score. In other words, for values of *k* higher than the one selected with the elbow method, the decrease in the clustering error for higher values of *k* is small.

4. Results

4.1 District's venues exploration

In this section I'm going to present the results regarding the venues of each district. The presentation is on the form of bar plots with the number and density of each particular venue.

4.1.1 Sushi Restaurants

The district with higher number and density of sushi restaurants is the city center. The district of Bonfim has an extremely low density for this venue. The city of Oporto has a total of 54 sushi restaurants.

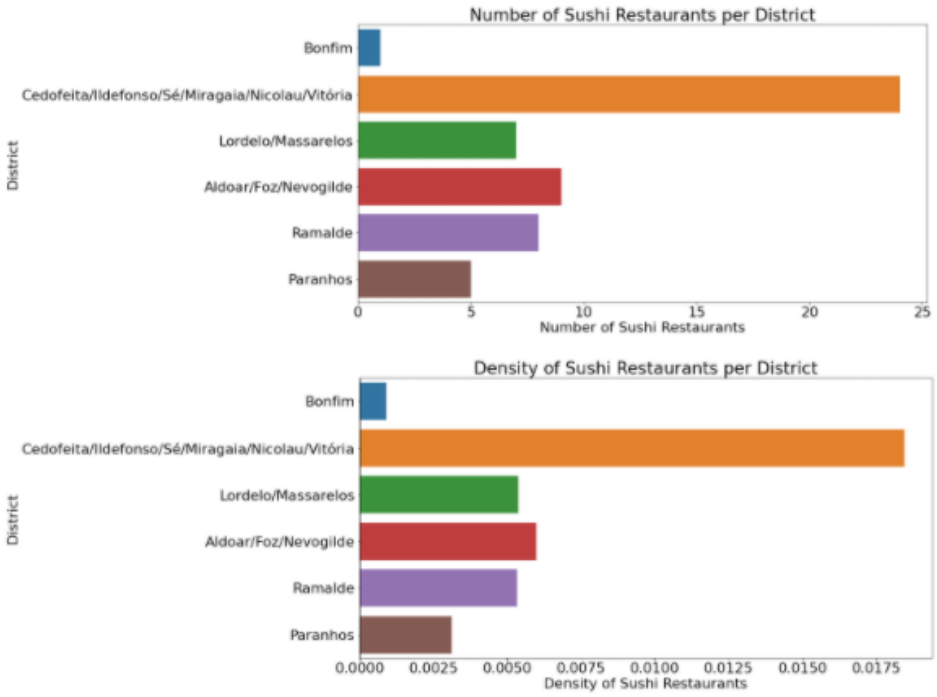

Figure 4: Sushi Restaurants for each district

4.1.2 Hotels

There is a total of 141 hotels in the city. The majority of them are located in the city center, as expected. Following the city center, the two districts with higher hotel density, with a huge gap comparing to the remaining districts, are Bonfim and Lordelo/Massarelos. The count of the number and the calculation of the density of the hotels is represented in figure 5.

4.1.3 Schools & Universities

There are 214 schools or universities in the city of Oporto. Once again the city center has the higher density of schools and universities, followed by Paranhos, which is a district with a lot of academic facilities, and Lordelo/Massarelos. The visualization of this venue is shown in figure 6.

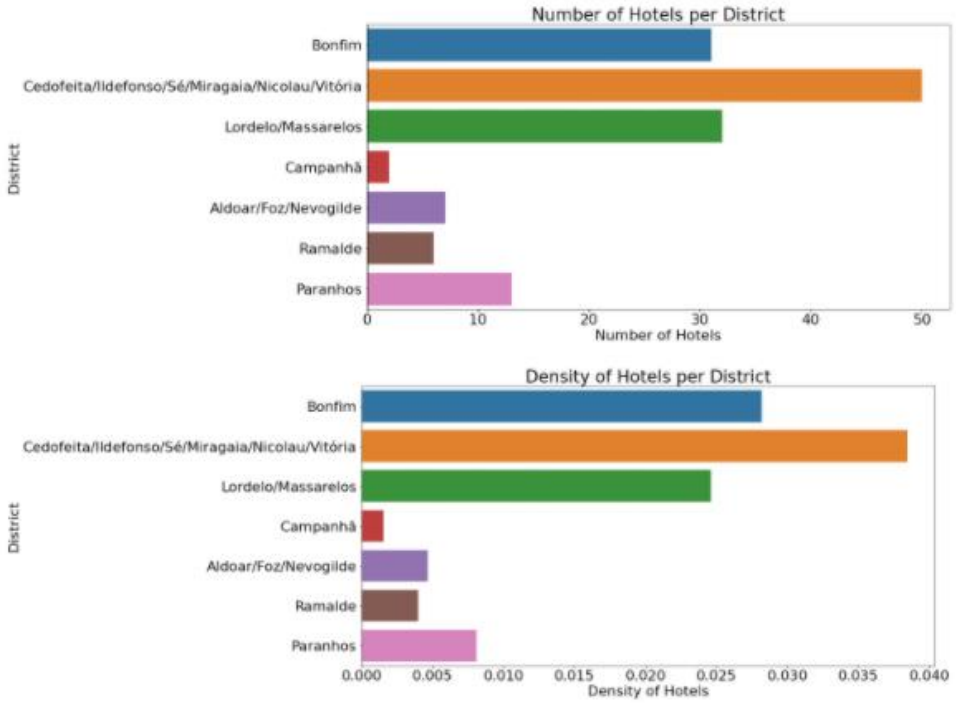

Figure 5: Hotels for each district

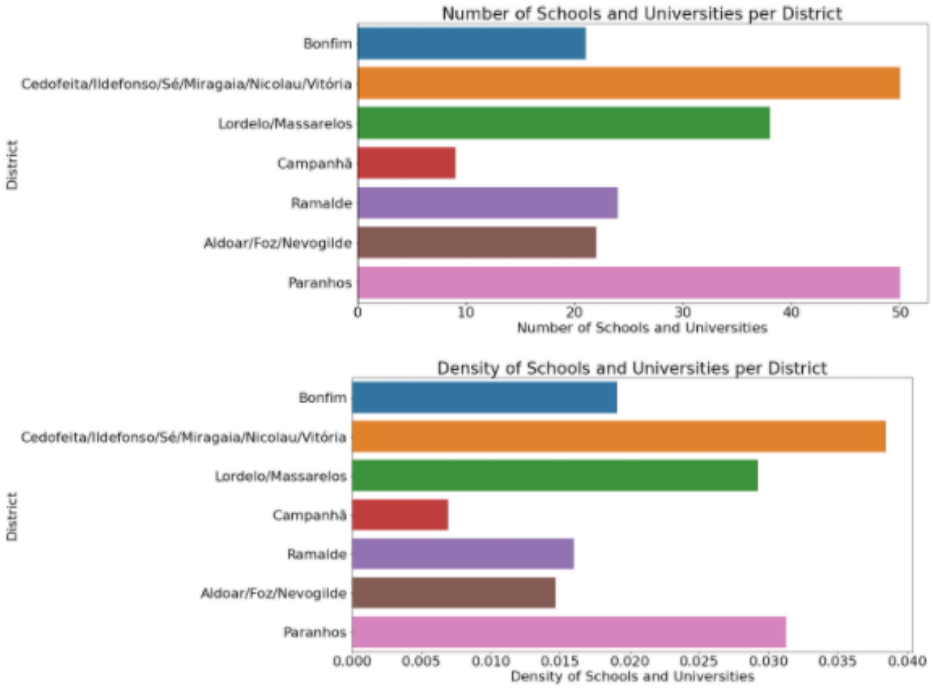

Figure 6: Schools and Universities for each district

4.1.4 Means of Transportation

As expected, there is a very significant gap between the density of means of transport comparing the city center and the other districts. The transports density in Aldoar/Foz/Nevogilde is extremely low in comparison to the other districts.

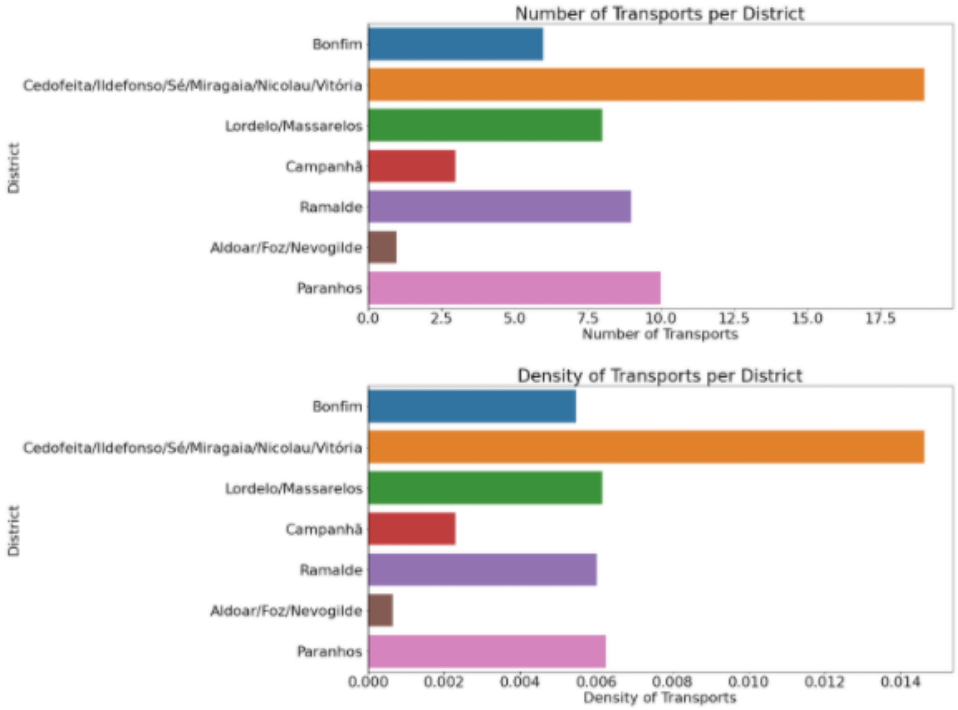

Figure 7: Means of Transport for each district

4.2 Popular Venues of each district

In this section I present the top 5 most popular venues for each district, as well as the number of sushi venues that are popular at the moment. This exploration is important to see if the trendy venues of each district are somehow favorable or a good indicator for the project purpose.

4.2.1 Bonfim

There is a clear trend for restaurants in the district of Bonfim. This can be a good indicator that this district is a good candidate for the sushi restaurant. However, the only sushi restaurant of this district is not trendy. Figure 8 shows the Top 5 Most Popular Venues.

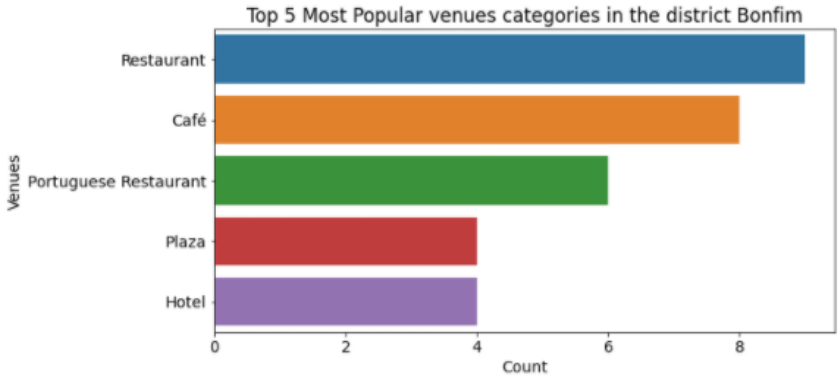

Figure 8: Top 5 Most Popular Venues in Bonfim

4.2.2 Cedofeita/Santo Ildefonso/Sé/Miragaia/São Nicolau/Vitória (City Center)

There is a clear trend for tourism related venues in the city center, like Hostels, Plazas and Restaurants. Although, only 3 out of the 24 sushi restaurants in this district are trendy at the moment.

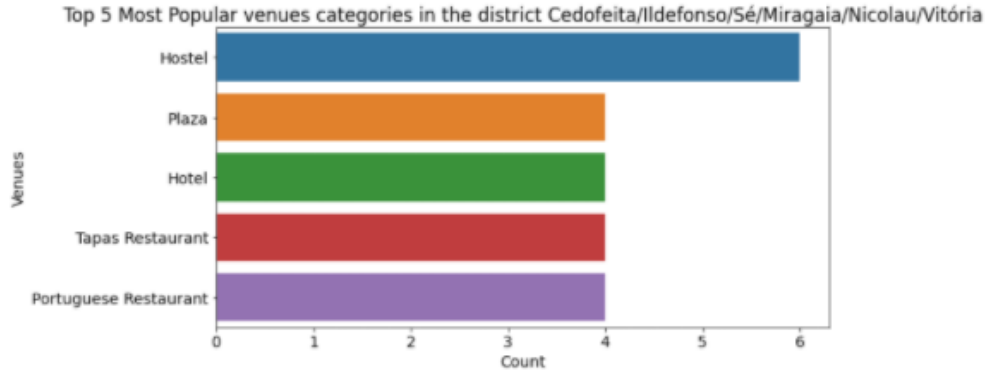

Figure 9: Top 5 Most Popular Venues in the city center

4.2.3 Lordelo/Massarelos

There is a clear trend for hotels and restaurants in the district of Lordelo/Massarelos. However, only 1 out of 7 sushi restaurants in this district are popular.

4.2.4 Campanhã

There is no particularly trendy category in this district. The most popular ones only have two instances. Furthermore, this district has no sushi restaurant yet.

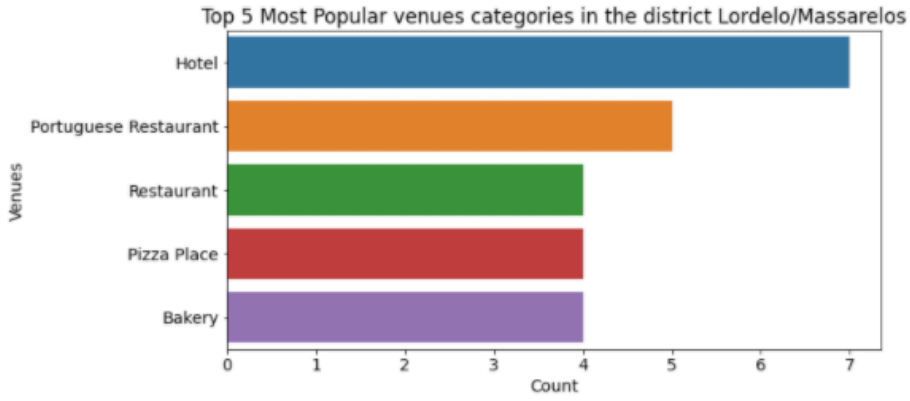

Figure 10: Top 5 Most Popular Venues in Lordelo/Massarelos

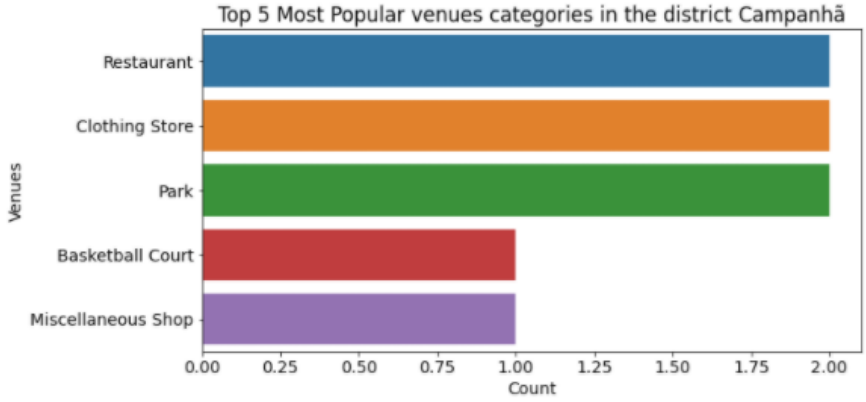

Figure 11: Top 5 Most Popular Venues in Campanhã

4.2.5 Aldoar/Foz/Nevogilde

The most trendy places in this district are generally food related. Surprisingly, 8 out of 9 sushi restaurants in this district are trendy at the moment. This indicator alone can be significant to choose this district as the location of the restaurant.

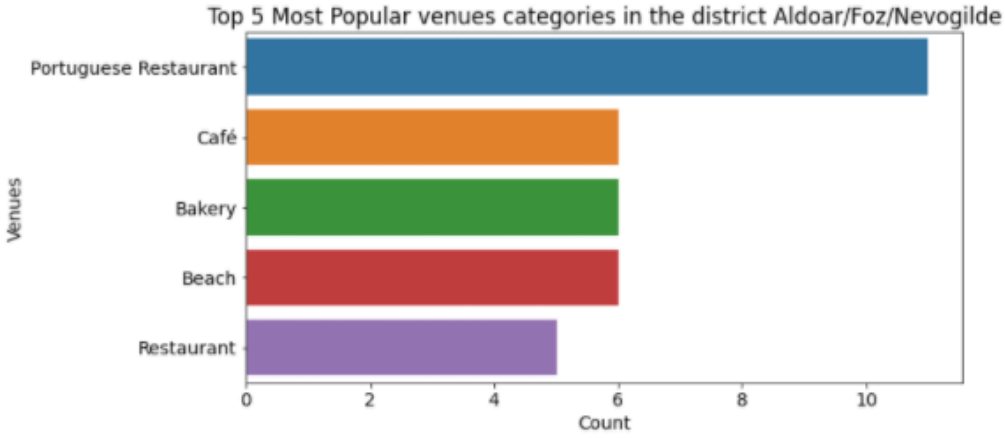

Figure 12: Top 5 Most Popular Venues in Aldoar/Foz/Nevogilde

4.2.6 Ramalde

The majority of the trendy places in this district are food related. However, only one out of the 8 sushi restaurants is trendy.

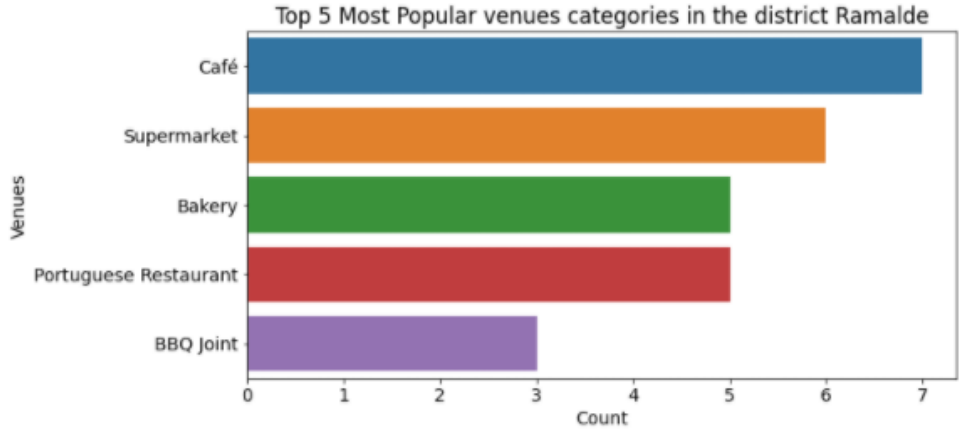

Figure 13: Top 5 Most Popular Venues in Ramalde

4.2.7 Paranhos

Once again, the majority of the trendy places are food related but only a small percentage of the sushi restaurants are trendy, in this district only 1 out of 5.

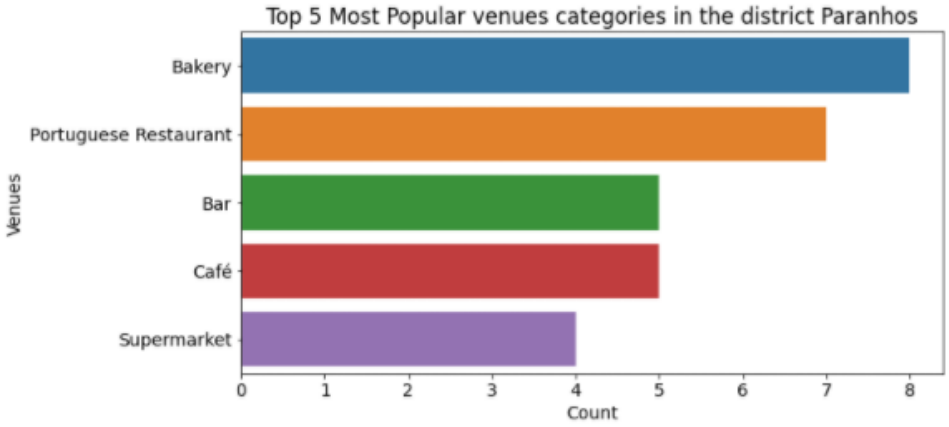

Figure 14: Top 5 Most Populat Venues in Paranhos

4.3 Clustering the districts

Based on the elbow method shown in figure 15 I decided to use 3 seeds for clustering (k = 3). Each cluster has a different color and for visualization purposes each district is represented by 4 circles, whose radius represent the density of the venues analyzed.

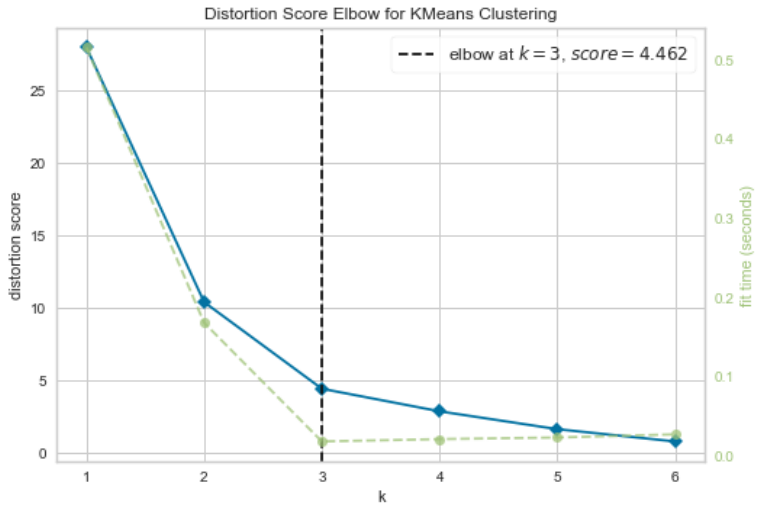

Figure 15: Elbow Method Representation

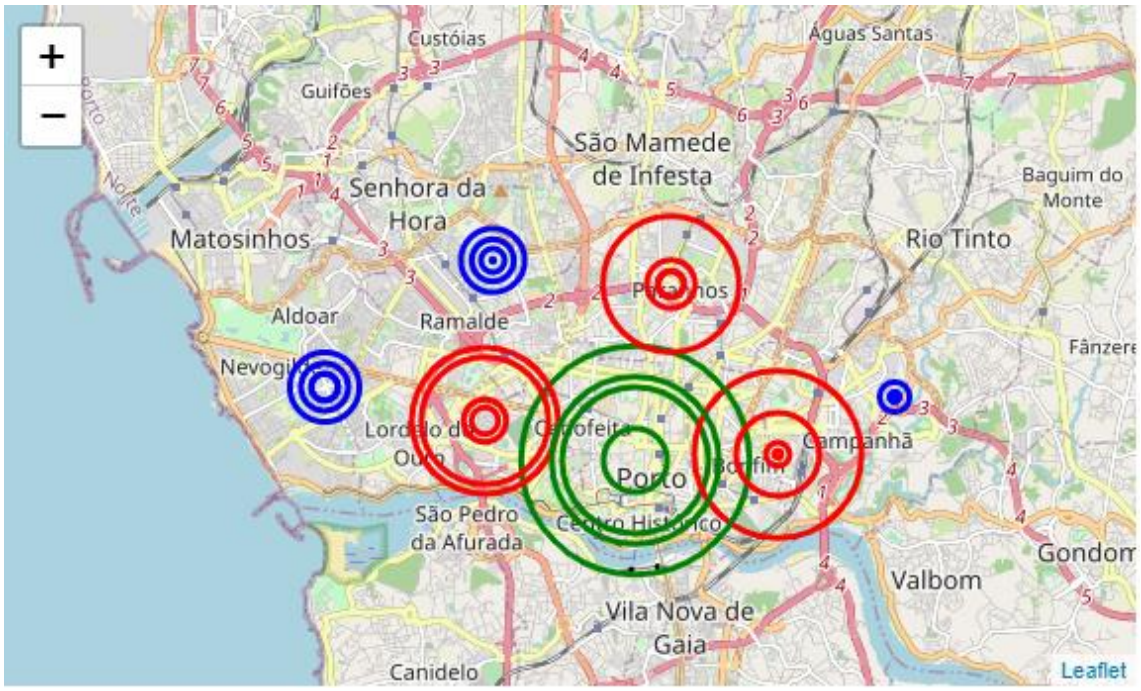

Figure 16: Clusters Visualization

Clearly, the blue districts are the ones with a small amount of venues considered critical for this analysis. There is a significant difference between the number of venues of the green (city center) and red districts.

5 Discussion

Based on the results I would recommend the districts of Bonfim or Aldoar/Foz/Nevogilde for the location of the sushi restaurant.

Although the district of Bonfim (Red) has not so many critical venues comparing to the green district, it has only one Sushi Restaurant (comparing to the green district). In addition, the top 3 most popular venues of Bonfim are restaurants and café's, which represent a tendency to food related establishments to perform well in this district. The other red districts are Lordelo/Massarelos and Paranhos, but, comparing to Bonfim, the densities of sushi restaurants are 6 and 3.4 times higher respectively, which represent a much higher concurrence in this districts comparing to the district of Bonfim.

The fact that 8/9 of the sushi restaurants are trendy in the district of Aldor/Foz/Nevogilde is a sufficient indicator, in my opinion, to consider this district as a good choice for the investment. A possible explanation why sushi is really trendy in this district is that people who live in there are known for being above average in terms of wealthiness, which can lead to a higher affluence of sushi restaurants, which are typically expensive if the sushi is of good quality. Bellow there's a visualization of the sushi venues in this district area (represented by a yellow circle) and the sushi venues with a radius of 300 meters (blue circles). There is clearly a lot of space to create a new sushi venue. Once can clearly see that, even extremely close sushi restaurants have the ability to be trendy in this district.

Figure 17: Sushi venues in the district of Aldoar/Foz/Nevogilde

6 Conclusion

The results obtained are a strong indicator of where to invest in a sushi restaurant in the city of Oporto. Furthermore, I based my final decision in two different criteria: The number of crucial venues for each district and how trendy sushi restaurants are in a specific district.

There are several important criteria to have in consideration in order to make a final decision that were not explored in this project, for instance, the rent per meter square in each district.