Capstone Project: Lighthouse Classification

Iván Martínez Suárez

1 Introduction

1.1 Background

Galicia is a spanish region with many kilometers of coastline. In order to avoid maritime accidents, along the coastline there are plenty of lighthouses to guide the ships. These places are usually isolated and are far from touristic areas.

The Galician government has a plan to boost tourism in the villages near these lighthouses: they will subsidize newly opened catering businesses located less than 3 kilometers from a lighthouse. A key fact is that the amount of money given to a new business will be proportional to the age of the lighthouse, since this feature will increase the tourist attraction of that place.

1.2 Business problem

A company is interested in creating a new catering business, fulfilling the requirements to receive the grant. The problem is knowing which is the best lighthouse to place the food business, because within a radius of 3 km from many of the Galician lighthouses there are already restaurants and bars.

The purpose of the analysis will be finding the best place to install the catering business, according to the restaurant density in that area and the age of the lighthouse.

2 Data

2.1 Source, clean up and selection

We can find a list of the main spanish lighthouses at Wikipedia (click here). This table contains, apart from the name and coordinates of each lighthouse, other features such as height, light range or year of construction. Some of these features will be used during the data analysis, in particular coordinates and built year.

It is necessary to make an important effort to clean up this data table, so that the final format is readable for us. Then, specific data for Galician lighthouses is extracted, because these will be the geographical extent of our analysis.

We have to drop some columns and rows, extract the adequate information from every feature and many other issues. Then, we center our analysis in Galician lighthouses, as we have data for all spanish ones. All the process is be showed in detail in the project notebook.

Faro	Imagen	Localización & Coordenadas	Provincia / Comunidad	Autoridad Portuaria	Cuerpo de agua	Año de construcción	Altura del soporte	Altura del plano focal	Rango
Adra[6]		Adra 36.74796, -3.03098	Almería Andalucía	Almería	Mar Mediterráneo	1889	26 metros (85 pies)	49 metros (161 pies)	16 millas náuticas (30 km)
Águilas[7]		Águilas 37.40169, -1.57802	Murcia	Cartagena	Mar Mediterráneo	1860	23 metros (75 pies)	30 metros (98 pies)	13 millas náuticas (24 km)
Ahorcados[8] [9]		Isla de los Ahorcados 38.81451, 1.41178	Islas Baleares	Baleares	Mar Mediterráneo	1856	17 metros (56 pies)	27 metros (89 pies)	10 millas náuticas (19 km)
Albir[10] [11]		Alfaz del Pi 38.56363, -0.05007	Alicante Comunidad Valenciana	Alicante	Mar Mediterráneo	1863	8 metros (26 pies)	112 metros (367 pies)	15 millas náuticas (28 km)

Figure 1: Some of the Galician lighthouses and their features before cleaning (Wikipedia).

	Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range
0	Cabo Corrubedo	42.57627	-9.09009	Galicia	1854.0	14.0	32.0	15.0
1	Cabo Finisterre	42.88236	-9.27196	Galicia	1853.0	17.0	143.0	23.0
2	Cabo Ortegal	43.77107	-7.87018	Galicia	1954.0	10.0	124.0	18.0
3	Cabo Prior	43.56761	-8.31453	Galicia	1853.0	7.0	107.0	22.0
4	Cabo Prioriño	43.45879	-8.34033	Galicia	1854.0	5.0	36.0	23.0

Figure 2: Some of the Galician lighthouses and their features after cleaning.

2.2 Foursquare data: most common venues

Once we have the final data, we will use Foursquare API to find out the nearest restaurants and bars from every lighthouse. For that, we will only look for *food* venues, as we can see in detail in the notebook. Moreover, we will limit our search to 10 venues per lighthouse within a radius of 3 km.

We prepare and mix our previous data with data from Foursquare, and then we proceed to the analysis stage, in which we will carry out machine learning clustering. Namely, we will cluster the Galician lighthouses according to the most popular food venues in their vicinities, using the *kMeans* algorithm.

Lighthouse Name	Bakery	Brazilian Restaurant	Breakfast Spot	Burger Joint	Cafeteria	Café	Diner	Food	Food Court	Gastropub
Cabo Corrubedo	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.00
Cabo Finisterre	0.0	0.0	0.0	0.0	0.00	0.0	0.1	0.1	0.0	0.00
Cabo Ortegal	0.0	0.0	0.0	0.0	0.00	0.0	0.5	0.0	0.0	0.00
Cabo Prioriño	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.00
Cabo Silleiro	0.0	0.0	0.0	0.0	0.25	0.0	0.0	0.0	0.0	0.00

Figure 3: Frequency of venues for some lighthouses (from Foursquare).

3 Methodology

As we said before, once we have prepared our dataset (with Wikipedia and Foursquare inputs), it is time to machine learning, in particular clustering.

Before we start clustering, let's display all the lighthouses in the map of Galicia:

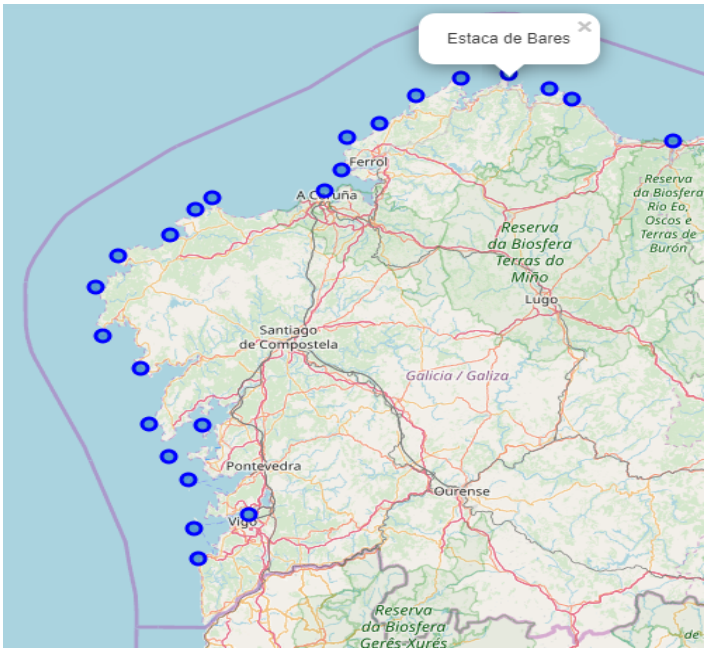

Figure 4: Map of Galician lighthouses.

3.1 KMeans Clustering

As we did in the previous labs, kMeans algorithm is used. This time we set up the algorithm with 6 clusters. We expect that after the clustering, one of the resulting groups fits our requirements to install a new food business in a place near one of the lighthouses in the group.

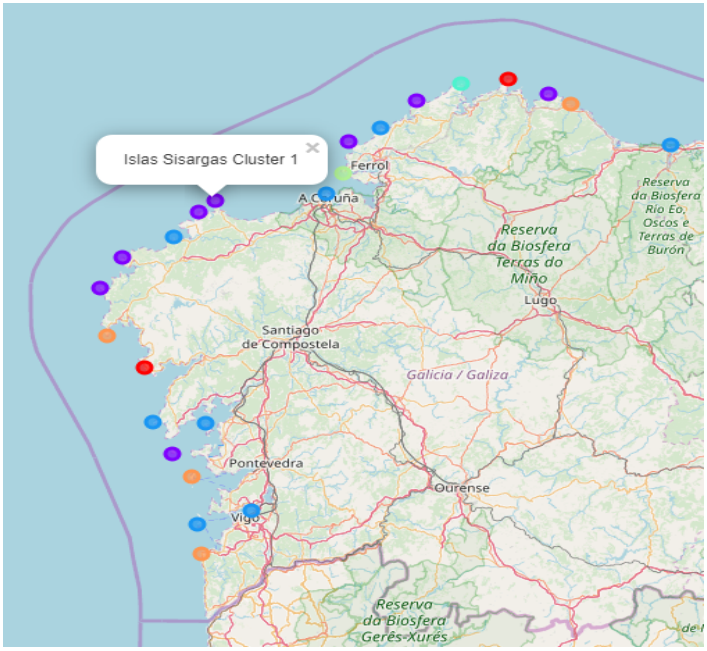

Figure 5: Clustering of Galician lighthouses.

At the following lines we briefly describe each resulting cluster:

3.1.1 Cluster 0 (red)

There are only 2 lighthouses in Cluster 0, characterized by its most common venue: seafood restaurant. This kind of food is very typical of Galicia.

	Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
7	Estaca de Bares	43.78623	-7.68430	Galicia	1850.0	10.0	101.0	25.0	0	Seafood Restaurant	Spanish Restaurant	None
18	Punta Insua	42.77249	-9.12503	Galicia	1921.0	14.0	27.0	15.0	0	Seafood Restaurant	None	None

Figure 6: Cluster 0.

3.1.2 Cluster 1 (purple)

The members of Cluster 1 are characterized by the total absence of venues in its vicinity. They are our candidates to install the new catering business and we will come back to analyze this cluster later.

	Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
3	Cabo Prior	43.56761	-8.31453	Galicia	1853.0	7.0	107.0	22.0	1	None	None	None
6	Cabo Villano	43.16041	-9.21093	Galicia	1896.0	25.0	104.0	28.0	1	None	None	None
8	Isla de Sálvora	42.46586	-9.01308	Galicia	1852.0	16.0	40.0	21.0	1	None	None	None
11	Islas Sisargas	43.35991	-8.84457	Galicia	1853.0	11.0	110.0	23.0	1	None	None	None
16	Punta Candelaria	43.71083	-8.04709	Galicia	1954.0	9.0	89.0	21.0	1	None	None	None
20	Punta Nariga	43.32068	-8.91010	Galicia	1998.0	39.0	55.0	22.0	1	None	None	None
21	Punta Roncadoira	43.73576	-7.52530	Galicia	1984.0	14.0	94.0	21.0	1	None	None	None
22	Touriñán (Toriñana)	43.05336	-9.29814	Galicia	1898.0	14.0	65.0	24.0	1	None	None	None

Figure 7: Cluster 1.

3.1.3 Cluster 2 (blue)

Cluster 2 members have at least 3 kinds of venues, mainly Spanish and Tapas Restaurants.

Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Cabo Corrubedo	42.57627	-9.09009	Galicia	1854.0	14.0	32.0	15.0	2	Tapas Restaurant	Spanish Restaurant	Restaurant
Isla Pancha	43.55654	-7.04204	Galicia	1859.0	13.0	28.0	21.0	2	Spanish Restaurant	Seafood Restaurant	Restaurant
La Guía	42.25953	-8.70213	Galicia	1844.0	21.0	37.0	15.0	2	Spanish Restaurant	Seafood Restaurant	Restaurant
Monte del Faro	42.21412	-8.91571	Galicia	1853.0	10.0	187.0	22.0	2	Spanish Restaurant	Snack Place	Seafood Restaurant
Punta Cabalo	42.57242	-8.88399	Galicia	1853.0	5.0	13.0	10.0	2	Spanish Restaurant	Seafood Restaurant	Gastropub
Punta Frouxeira	43.61806	-8.18836	Galicia	1992.0	30.0	75.0	20.0	2	Spanish Restaurant	Café	Breakfast Spot
Punta Laxe	43.23214	-9.01121	Galicia	1995.0	11.0	66.0	20.0	2	Tapas Restaurant	Spanish Restaurant	Snack Place
Torre de Hércules	43.38594	-8.40648	Galicia	150.0	49.0	106.0	23.0	2	Spanish Restaurant	Tapas Restaurant	Restaurant

Figure 8: Cluster 2.

3.1.4 Cluster 3 (turquoise)

Cluster 3 has only one member. Its most common venue is seafood restaurant and the second is diner.

	Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	Cabo Ortegal	43.77107	-7.87018	Galicia	1954.0	10.0	124.0	18.0	3	Seafood Restaurant	Diner	None

Figure 9: Cluster 3.

3.1.5 Cluster 4 (green)

Cluster 4 has only one member due to its singular most common venue. It is not usual to find a turkish restaurant nearby a lighthouse in Galicia.

	Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
4	Cabo Prioriño	43.45879	-8.34033	Galicia	1854.0	5.0	36.0	23.0	4	Turkish Restaurant	None	None

Figure 10: Cluster 4.

3.1.6 Cluster 5 (orange)

Cluster 5 is the most heterogeneous one. All the lighthouses have several venues and they are diverse.

	Lighthouse Name	Latitude	Longitude	Community	YearBuilt	TowerHeight	FocalHeight	Range	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Cabo Finisterre	42.88236	-9.27196	Galicia	1853.0	17.0	143.0	23.0	5	Seafood Restaurant	Spanish Restaurant	Sandwich Place
5	Cabo Silleiro	42.10436	-8.89655	Galicia	1862.0	30.0	85.0	24.0	5	Restaurant	Seafood Restaurant	Cafeteria
9	Isla de Ons	42.38244	-8.93639	Galicia	1865.0	12.0	127.0	25.0	5	Food	Tapas Restaurant	Seafood Restaurant
14	Punta Atalaya (o San Ciprián)	43.70050	-7.43680	Galicia	1860.0	14.0	41.0	15.0	5	Seafood Restaurant	Restaurant	Pizza Place

Figure 11: Cluster 5.

4 Final Results

According to the previous analysis, the best lighthouses to create a new food business are the members of Cluster 1, because of their lack of any other food venues nearby them.

At this point, it is important to remember that the government grant will be proportional to the age of the lighthouse. So, the best places for our business in Galicia are, in descending order:

Lighthouse Name	YearBuilt
Isla de Sálvora	1852.0
Cabo Prior	1853.0
Islas Sisargas	1853.0
Cabo Villano	1896.0
Touriñán (Toriñana)	1898.0
Punta Candelaria	1954.0
Punta Roncadoira	1984.0
Punta Nariga	1998.0

Figure 12: List of the best Galician lighthouses to place a new food business.

There are 8 candidate lighthouses to place our business, sorted by its year of construction. The first lighthouse was built more than a century and a half before the last one, so the amount of money received from the Galicia government could be very different in each case.

5 Discussion

Despite we had reasonable results for our purpose, it is important to keep in mind that there are many other limitations to build up a new restaurant, bar or any other food business.

For instance, according to Table 12, the preferred lighthouse is *Isla de Sálvora*. But this lighthouse is located on an inhabited island, whose name is Sálvora! It may not be a good a idea to place your food business where there are no people.

6 Conclusions and Future Work

In order to avoid this issues, in the future it would be necessary to input more information to the clustering algorithm or even before this stage.

Other important question is the chose of the number of clusters, which was quite successful this time, according to the results, but could be improved.