

# Goiko grill restaurant analysis

## 1. Introduction

The Spanish fast-food chain Goiko Grill has many restaurants throughout Spain, right now they want to continue expanding, but before constructing more establishments they want to analyse the current locals.

To do so they have decided to classify their restaurants based on the competition near them, they hope to see if there are any patterns among the restaurants.

Once the classification is done, they will use some private data of their own (like profitability of each local) to decide with type of cluster is better. Later they will use this information to decide the locations of future restaurants

## 2. Data description

In order to solve this problem several different types of data are required and have been leveraged using a wide array of apis and code packages. Namely the data used is the following

- Goiko grill web page: Data regarding the location of the restaurants has been scrapped from the Goiko grill's web page using beautiful soup <https://www.goiko.com/todos-nuestros-restaurantes/>
- Latitude and longitude data: it has been collected manually since no working (and free) options were found
- Neighbouring venues using foursquare: data from type of venues around the restaurants has been leveraged using the foursquare api.

Finally, using together all the data, a clustering model was built, which clustered the restaurants depending on which type of neighbourhood they are located according to the venues in their surroundings.

Since I had to find the geo coordinates for the restaurants manually, I decided to only take one restaurant from each city. Which means that we end up with only 30 restaurants instead of 100+

## 3. Methodology

- Data pre-processing:

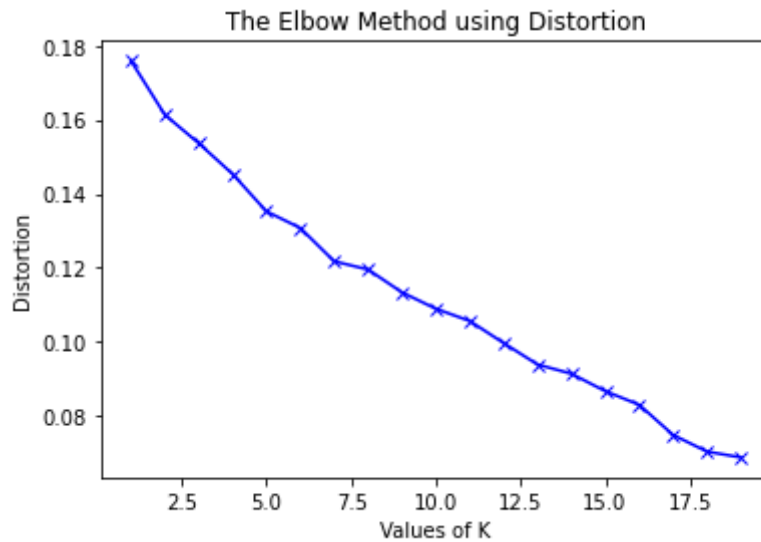
In order to conduct the data preprocessing I eliminated the location column and used instead the city column, additionally I cleaned and ordered the data I acquired with beautiful soup.

The data I obtained from foursquare I grouped and clustered

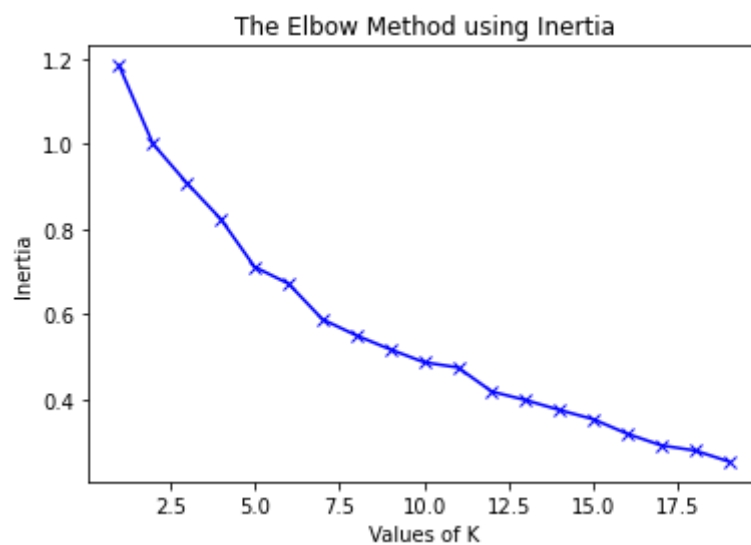
- Statistical testing:

In order to find the most accurate number of clusters to divide the data into I performed the elbow test with both inertia and distortion, the results are the following

Result using elbow method with distortion



Results using elbow method with inertia



The values of both inertia and distortion are low, yet, we cannot find a clear number of clusters, since the data frame is not that big I decided to divide the data into four different clusters

- Algorithms:

I decided to use the K means clustering algorithm, the reason why I decided this algorithm is that I wanted to cluster the data based on neighbourhoods' similarities, furthermore, it is an efficient algorithm. Additionally, using the elbow method, you can accurately choose the appropriate k for the clustering.

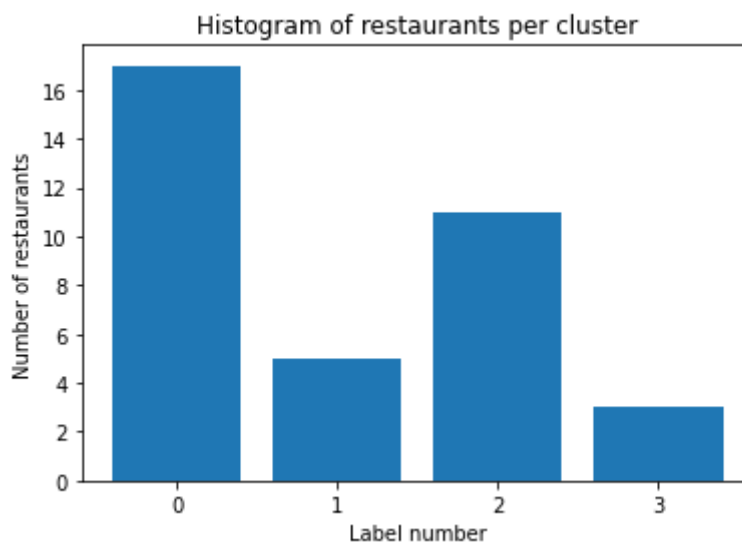
	City	Location	Latitude	Longitude	Cluster Labels
0	Madrid	María de Molina 20, Madrid	40.437680	-3.686744	1
1	Valencia	Martínez Cubells 4, Valencia	39.468176	-0.374816	1
2	Zaragoza	San Miguel 7, Zaragoza	41.651058	-0.880645	0
3	Alicante	Pablo Iglesias 5, Alicante	38.347451	-0.488203	2
4	Barcelona	Sant Joan 58, Barcelona	41.396366	2.174007	1

## 4. Results

After running our model, following the elbow method, we were able to divide the dataset into four different clusters, accordingly this clusters were named in the following fashion

- Label 0: Leisure areas, restaurants of this cluster are located in city centres in areas where there are leisure activities and grab and go food options ('tapas')
- Label 1: Cosmopolitan cities, restaurants of this clusters are located in some of the most cosmopolitans' cities of Spain, and therefore are surrounded by modern venues
- Label 2: City based shopping areas, restaurants of this cluster are located on city centres in areas were there is a predominance of restaurants and shops
- Label 3: Out of city shopping malls, restaurants of this cluster are located on shopping malls that are not located in the outskirts of the city.

This is a graphical representation of the number of restaurants per cluster:



## 5. Discussion

Our analysis has revealed the existence of 4 different types of restaurants based on the type of neighbourhood they are located in. From this point on, to continue the analysis proprietary data would

be necessary, such as restaurant profitability, land costs... with this type of information we could analyse the profitability of the different types of restaurants. Once we know which type of restaurant (if any), is significantly better than the others, Goiko grill would be able to build new restaurants in those areas which resemble the most profitable cluster.

## 6. Conclusion

This analysis has yielded significant information that has value of its own worth, but in order to fully leverage the analysis conducted we would need access to more data, which only the pertaining company has access to.

Nonetheless. The analysis has revealed how are divided the restaurants, additionally, it reveals where in Spain are located the restaurants and which type of restaurant is available in each city.