Healthy Food Store Chain opening in Barcelona (Spain)

Lucas Murciano

April 2020

1. Introduction

1.1 Background

During the last years, people has become more interested to the topics related with health and nutrition. This is in good part related to the big influence of social media, where everyday we can see athletes, models, and other people with healthy lifestyle and tons of followers sharing their daily routines, what of course include where do they go to eat and where they obtain the products to prepare food at home.

So, now more tan ever, people is willing to increase the spending in nutrition because of the direct relation assumed to the health. this regard, big existing food and supermarket chains, and also new creation ones, have no lost their time in order to offer a healthy branch of products and restaurants.

1.2 Problem

An existing supermarket Company based in Barcelona (Spain) has decided to open 3 new concept stores related to healthy food and lifestyle. In this regard, the Company needs the analysis of the diffferent commerce models of neighbourhoods of Barcelona in order to decide the most suitable locations for the new stores. Renting cost for each area will be considered.

1.3 Interest

As the study is going to cover an analysis of the commerce models in Barcelona, insights obtained are interesting to every company that wants to open new stores or moving existing ones. Also can be interesting to people in general who want to decide a neighbourhood to live based on the relation between nearby commerce and renting costs.

2. Data Aquisition and Cleaning

2.1 Data Sources

In first place, geographical location features of each neighbourhood have been obtained. The most reliable source for this info has been the [official metropolitan transports of Barcelona Open Data page](#), where geographical points for each metro station are included.

For commerce data, exploring function of the Foursquare APIs has been used for each neighbourhood centered on its metro station and with a limit of 100 commerces.

For renting data, detailed and relaiable infographic of the avg cost (€/m2) for each neighbourhood based on it's metro station can be found in the [Real State services web Idealista](#).

2.2 Data Cleaning

Geolocation data for each metro station is obtained in xlsx format, so it can be directly transformed into a Dataframe by Pandas. Information regarding name, metro line and geographical coordinates will be kept.

As the renting costs are in infographic format, data for each metro station will be added manually to the xlsx obtained for the metro stations, so it can be included in the same dataframe.

For the commerce data, JSON format with 100 venues for each neighbourhood is obtained and can be casted directly into a Dataframe. Data will be grouped by neighbourhoods in order to obtain the rate of different commerce categories by location so k-means clustering can be performed to split the areas of Barcelona into similar groups.

2.3 Features

Key Features for the analysis will be: Neighborhood Cluster, Neighborhood existing healthy related commerces, renting cost in the area.

With this 3 figures for each area, analysis of which clusters include the most current healthy stores will be performed so, other areas within the same cluster with less competence rate and competitive renting costs can be found to set the 3 new stores.

3. Exploratory Data Analysis

3.1 Data Properties

Once the location and price Data has been cleaned we obtain a **definitive list of 135 neighborhoods** in the city of Barcelona with the following information:

CODI_GRUP_ESTACIO	NOM_ESTACIO	PICTO	Long	Lat	PREU (€/M2)
6660111	Hospital de Bellvitge	L1	2.107242	41.344677	12.78
6660112	Bellvitge	L1	2.110918	41.350974	12.85
6660113	Av. Carrilet	L1	2.102645	41.358522	13.11
6660114	Rambla Just Oliveras	L1	2.099749	41.364090	13.23
6660115	Can Serra	L1	2.102755	41.367693	13.33

Table 1. Example of Data Cleaned by Neighbourhood

Using the values of Latitude and Longitude of each neighbourhood in the Foursquare API, with a **radius of 500m** and a **venue list limit of 100** the list of Barcelona is obtained.

Grouping the venues listed by neighbourhood we obtain the Histogram below:
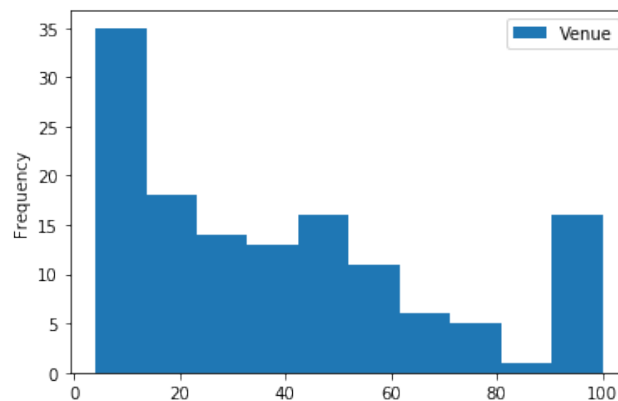

Figure 1. Frequency of Neighbourhoods by Total venues

It can be seen that:

- 35% aprox of Neighbourhoods have 20 or less venues
- 65% have 20 or more venues

By converting the venue categories into dummies, it is possible to obtain the most common venues by neighbourhood:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Aeroport T1	Airport Lounge	Clothing Store	Spanish Restaurant	Airport Service	Coffee Shop	Bookstore	Café	Sporting Goods Shop	Rental Car Location	Duty-free Shop

Table 2. Example of Most Common Venues by Neighbourhood

3.2 Neighbourhood K Means Clustering

With the proportion of each of the **309 venue categories** by neighbourhood, it has been built a KMeans model in order to cluster Barcelona Areas into similar ones.

Firstly the Elbow method is performed in order to define which k value is the most suitable for the clustering:
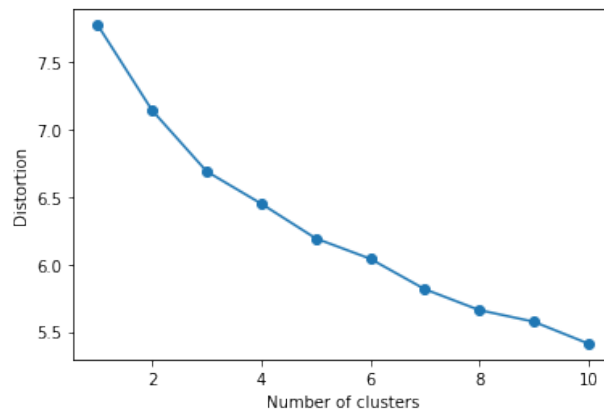

Figure 2. K Means dispersion by k for Neighbourhood Clustering by Venues

As it can be seen, there is no clear elbow in the graph where the slope changes dramatically, so **k=5** is taken as is a reasonable number of initial clusters.

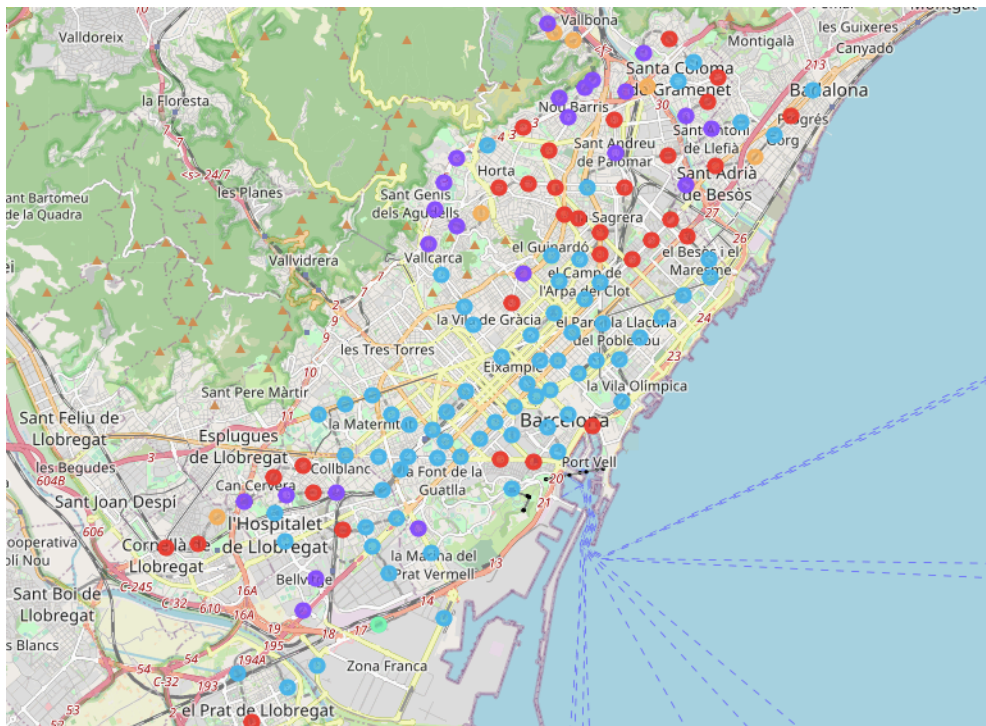Performing KMeans, we obtain the following classification:

Figure 3. Barcelona Clusters by Venue Categories

- **Cluster 0: Residential Areas in Downtown** with Tapas restaurants and supermarkets
- **Cluster 1: Residential Areas in the Outksirts** with Parks and Sport Centers
- **Cluster 2: Turistic Areas in Downtown** with Hotels and Restaurants
- **Cluster 3: Barcelona Harbor**
- **Cluster 4: Industrial Areas in the Outskirts**

3.3 Health Food Stores by Cluster

By grouping venue data into clusters, we obtain the distribution of the competence presence by area in Barcelona
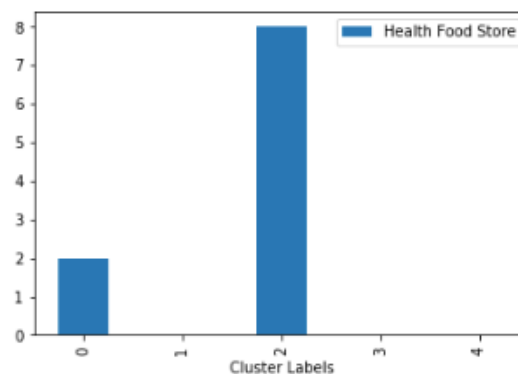

Figure 4. Health Food Stores by Cluster

As shown, all the competitors are present in the Downtown areas, where the major part of the Barcelona social life occurs, so **suitable locations for the store openings have to be seek within Cluster 2.**

4. Model Development

4.1 Competitors vs Prices

Once the area with most potential customers is found, the next part is to define which are the most suitable locations. For this we have taken into consideration 2 main parameters: **Presence of competitors and renting prices**.

We first display a scatter plot of this two parameters within the Cluster 2 neighbourhoods in order to check any possible correlation.
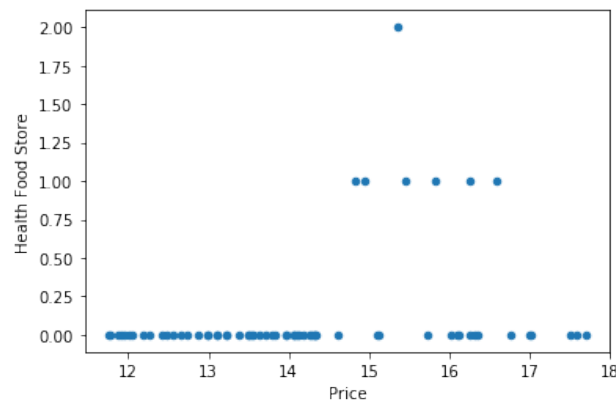
Figure 5. Competitors presence vs Renting Price on Cluster 2

As seen above, competitors presence is higher in areas with higher prices of renting, what could indicate that higher profits can be obtained in downtown areas.

4.2 K Means Clustering: Competitors vs Prices

In order to clear out insights obtained in the previous section, we perform K means clustering over this group of areas.

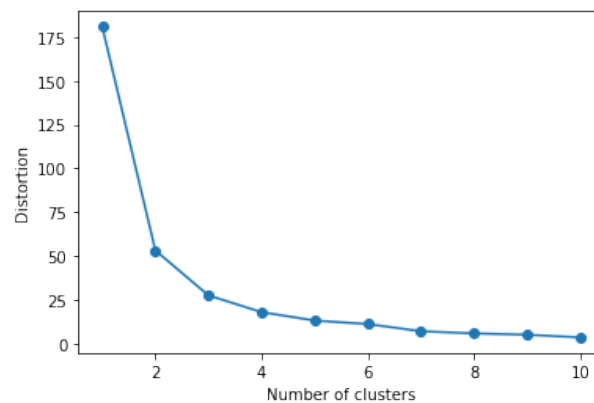We first perform Elbow method to find most suitable k:

Figure 6. K Means dispersion by k for Neighbourhood Clustering by Competitors and Rent Price

So, with **k=3** we perform k means clustering on price and competitors presence over Cluster 2 areas in order to obtain the following sub-division (See map below):

- **Sub Cluster 0: Downtown neighbourhoods** with **strong presence of the competence** and **medium prices**
- **Sub Cluster 1: Residential Areas** near the Outskirts with **no competence and low prices**
- **Sub Cluster 2: Turistic Downtown** with **weak competence and high prices**
- **Sub Cluster 3: Downtown neighbourhoods** with **no presence of competence and medium prices** *(Optimum Area)*
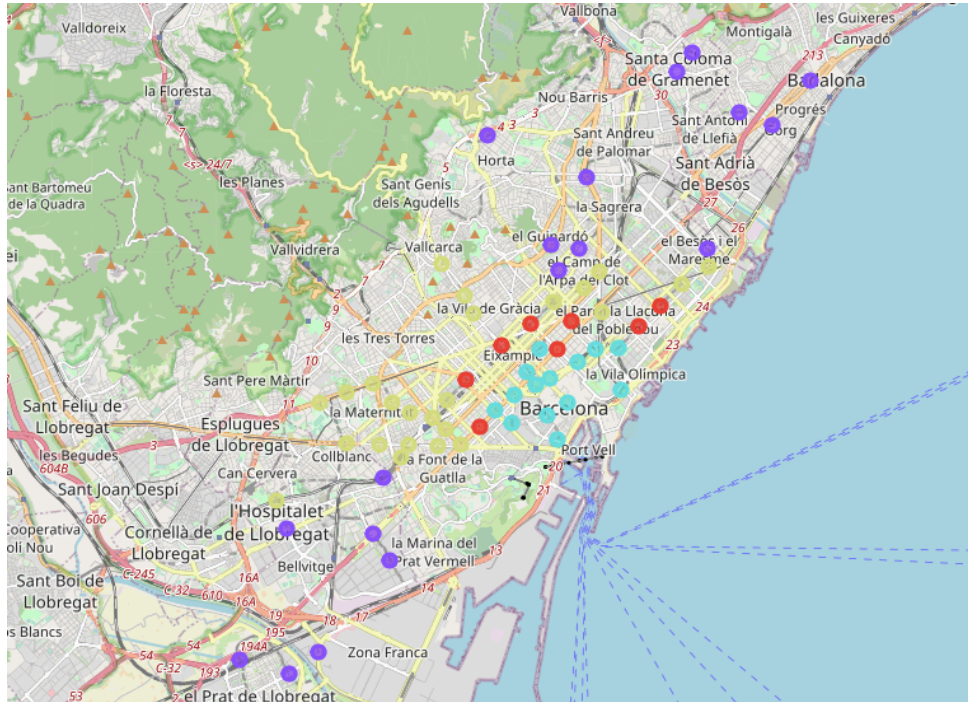
Figure 7. Barcelona Downtown Clusters by Competitors and Rent Price

The 24 marks in green show the areas listed below in the Cluster 2 – Sub Cluster 3 that are considered the **most suitable for placing the new stores**.

Neighborhood	Price	Subcluster	Health Food Store
Rambla Just Oliveras	13.23	3	0
Mercat Nou	13.22	3	0
Plaça de Sants	13.38	3	0
Hostafrancs	13.72	3	0
Espanya	14.18	3	0
Glòries	14.31	3	0
Clot	13.50	3	0
Sagrada Família	14.06	3	0
Encants	13.56	3	0
Zona Universitària	13.96	3	0
Palau Reial	14.29	3	0
Maria Cristina	14.61	3	0
Les Corts	14.12	3	0
Plaça del Centre	13.80	3	0
Sants Estació	13.96	3	0
Tarragona	14.11	3	0
Fontana	14.31	3	0
Lesseps	13.83	3	0
Vallcarca	14.27	3	0
El Maresme	Fòrum	13.53	3	0
Selva de Mar	14.06	3	0
Collblanc	13.64	3	0
Badal	13.50	3	0
Entença	14.34	3	0

Table 3. List of Most Suitable Locations and its Rent Prices

5. Conclusions

Data analysis of Barcelona venues and renting prices has been performed in order to determine best possible locations for Health Food Stores opening.

Two consecutive K Means clustering have been performed with the following purposes:

- 1st: Categorize neighbourhoods by their main venues
- 2nd: Find the most suitable location within the most likely cluster found in the first step

By performing this analysis, a **final list of 24 out of the initial 135 neighbourhoods has been obtained** as the best possible locations for the new stores.

6. Future Directions

With insights obtained from this study, a knowledge base of Barcelona venue distribution has been set, so further analysis can be performed taking into consideration aspects like income per neighbourhood, number of passers-by, closeness to related venues (Fitness centers), etc.

Furthermore, this base can also be used to studies on other venue categories of interest.