

Healthy Food Store & Restaurant 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 different 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 reliable 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.