IBM DATA SCIENCE

COURSERA CAPSTONE FINAL PROJECT

SPORT NUTRITION STORE @ GRANADA, SPAIN

Enrique Benítez López

February 2020

Introduction

Nowadays, there can be no doubt that diet and sport practice play an important role in our lives. They have been scientifically proven to be the major factors in the promotion, restoration and maintenance of good health throughout the entire life course. There is currently a huge amount of information accessible to everybody as well as an increase on sport facilities. People do more exercise and are more aware of the importance of a good diet in his purpose to live inside a healthier body.

The problem

The objective of this capstone project is to analyze and select the best location in the city of Granada, Spain to open a new sports nutrition store. Using data science methodology and machine learning techniques like clustering, this project aims to fin what is the best location in Granada to start a sports nutrition business.

The audience

As a healthy lifestyle is becoming more and more popular in our current society, lots of investors are deciding to put their money in projects related to wellbeing, such as gyms, sport infrastructure or specific nutrition stores.

About the city…

Granada is the capital city of the province of Granada, in the autonomous community of Andalusia, Spain. Granada is located at the foot of the Sierra Nevada mountains, at the confluence of four rivers, the Darro, the Genil, the Monachil and the Beiro. It sits at an average elevation of 738 m (2,421 ft) above sea level, yet is only one hour by car from the Mediterranean coast, the Costa Tropical. Nearby is the Sierra Nevada Ski Station.

In the 2005 national census, the population of the city of Granada proper was 236,982, and the population of the entire urban area was estimated to be 472,638, ranking as the 13th-largest urban area of Spain. Its nearest airport is Federico García Lorca Airport.

The Alhambra, an Arab citadel and palace, is located in Granada. It is the most renowned building of the Islamic historical legacy with its many cultural attractions that make Granada a popular destination among the tourist cities of Spain. The Almohad influence on architecture is also preserved in the Granada neighborhood called the Albayzín with its fine examples of Moorish and Morisco construction. Granada is also well-known within Spain for the University of Granada which has an estimated 82,000 students spread over five different campuses in the city.

Data

In order to solve the problem, three different types of data will be needed:

- **List of neighborhoods and municipalities in Granada**

  Neighborhoods of Granada are available in this link of Wikipedia whereas the XLS file with the list of municipalities has to be downloaded from the SIMA platform of the Junta de Andalucía, which is the organization that manages all official data of Granada.

- **Latitude and longitude coordinates of those locations**

  We will get the geographical coordinates of the neighborhoods using python Geocoder, paying attention to it in order to get the right geographical coordinates of our neighborhoods of interest.

- **Venue data, particularly data related to Gyms**

  We will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API Will provide many categories of the venue data, we are particularly interested in the **Gym** category in order to help us to solve the business problem put forward.

  This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.