

Capstone Project – Optimal location for a Colombian restaurant in Madrid

Introduction

Madrid is the capital and most populous city of Spain. The city has almost 3.3 million inhabitants and a metropolitan area population of approximately 6.5 million. It is the second-largest city in the European Union (EU), surpassed only by Berlin, and its monocentric metropolitan area is the second-largest in the EU, surpassed only by Paris. The municipality covers 604.3 km².

As the capital city of Spain, the city has attracted many immigrants from around the world, with most of the immigrants coming from Latin American countries. In 2020, around 76 % of the registered population was Spain-born, while, regarding the foreign-born population (24 %), the bulk of it relates to the Americas (around 16 % of the total population), and a lesser fraction of the population is born in other European, Asian and African countries. With its diverse culture, comes diverse food items. There are many restaurants in Madrid, each belonging to different categories like Spanish, Argentinian, Peruvian, Ecuadorian, Chinese.

Problem description

The main task is to find the best possible location or the most optimal, for a Colombian restaurant in Madrid. To accomplish this task, an analytical approach will be used, based on advanced Machine Learning techniques and Data Analysis, mainly Clustering and some Data Visualization techniques. During the process of analysis, several data transformations will be performed to find the best possible data format for the Machine Learning model. Once the data is set up and prepared, a modeling process will be carried out, and this statistical analysis will provide the best possible places to locate the Colombian restaurant.

Data Presentation

The data that will be used to develop this project is based on two sites:

- The Foursquare API, this data will be accessed via Python, and used to obtain the most common venues per neighborhood in the city of Madrid. It is possible to have a taste of how the city's venues are distributed, what the most common places are for leisure, and for the most part, it will provide an idea about the preferences of the inhabitants.
- The Madrid City Hall's Web Portal, this site provides several data sources of great utility to solve this problem. The files are provided in Excel format and are built over a statistical exploitation. The data contains updated information about the immigrant population per nationality. This data will be analyzed in such a way that one could determine the best location of a new venue/restaurant/other based on nationality. For make things simpler, it will be assumed for this exercise that people preferences vary according to their nationality, and people from one specific country will be more attracted to place that matches the environment and culture of their own countries, rather than the ones from foreign countries.