IBM Applied Data Science Capstone

Latino Restaurant opportunity in Manhattan, New York, NY

By: Rodrigo Martins

May 2020

Introduction

For this Capstone project, I am creating a hypothetical scenario for an entrepreneur who wants to explore opening a Latino Restaurant in Manhattan, New York, NY, such as Brazilian, Peruvian, etc. The idea behind this project is that there may not be enough Latino Restaurants in Manhattan, and it might be a great opportunity for this entrepreneur since there are a lot of Hispanic / Latinos living in this city. It would be better if the entrepreneur opens this kind of restaurant in locations where Latino food is popular. Finding the best location to open a restaurant is one of the most important decisions for an entrepreneur, and because of that this project has an objective of help him to find the most suitable location.

Business Problem

The objective of this capstone project is to analyze and select the best location in Manhattan, New York, NY to open a Latino Restaurant. This project will use data science methodology and machine learning algorithms like clustering to help an entrepreneur solve the following question: Where would be the best locations to open a Latino Restaurant in Manhattan, New York?

Target Audience

This project is particularly useful to entrepreneurs looking to open or invest in a new Latino Restaurant in Manhattan, New York since there is a big community of Latinos living there. According to the data released on Wikipedia, Manhattan, New York is populated by 27.5% of Hispanic or Latino, or approximately 2,287,905 people from this ethnicity, which give a good opportunity to entrepreneurs who want to open a Latino Restaurant and explore the lack of Latino cuisine in the city.

Data

To solve the problem, we will need the following data:

    I.   List of neighborhoods in Manhattan, New York. This defines the scope of this project.

II. Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and to get the venue data.

III. Venue data, particularly data related to restaurants. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This webpage (https://geo.nyu.edu/catalog/nyu_2451_34572) contains a list of neighborhoods in Manhattan, New York. We will download the data in a json file format, and use Python packages such as Pandas, to clean the data. The geographical coordinates of the neighborhoods will be obtained by cleaning the json file, which already contains the latitude and longitude coordinates of the neighborhoods.

Moreover, we will use the Foursquare API to get the venue data for all neighborhoods. Then, the Foursquare API will provide the categories of the venue data or cuisines that exist in Manhattan. The focus of this project is Latino Cuisine, such as Brazilian and Peruvian, in order to help us to solve the business problem.

This is a project that will make use of many data science skills, from working with Python libraries, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).