Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Coffee Shop in Sumaré, Brazil.

For coffee lovers, it's always welcome a nice place with a good location, specially when you want invite somebody to talk and share a good coffee.

Connection it's what we want between people, for some people it is practically synonymous of "Good morning."

Others prefer to drink it in the afternoon and many even confess that they drink several cups throughout the day, the fact is that, regardless of preferences, coffee is part of the lives of the majority of the population.

As the result, there a lot of coffee shop in the city of Sumaré and places being built, opening a new coffee shop there are a lot of considerations and the most important decision is the place that will be.

Business Problem

The target of this capstone project is identify and select the best locations in the city of Sumaré, Brazil to open a new coffee shop. Using tools of data science methodology and machine learning techniques like clustering, this project will provide results to check the best place in the city of Sumaré, Brazil.

Target Audience

This research is particular useful to new investors looking for open a new coffee shop in Sumaré, Brazil. According to Exame Magazine from brazil, the Coffee sector will grow in the next 10 years 25%.

And will be needed stores to sell coffee products and promote their brands to actual public and new ones.

Data

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

1. List of neighborhoods in Sumaré, Brazil. This defines the scope of this project which is confined to the city of Sumaré in the state of São Paulo in Brazil.
2. Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
3. Venue data, particularly data related to Coffee shops. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This Wikipedia page (https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Sumaré) contains a list of neighborhoods in Sumaré, with a total of 188 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.
After that, we will use Foursquare API to get the venue data for those neighborhoods.
Foursquare API will provide many categories of the venue data, we are particularly interested in the Coffee Shop category in order to help us to solve the business problem.
This is a project that will make use of many data science skills, web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).