Coursera Capstone

IMB Applied Data Science Capstone

Opening a Steakhouse in Florence, Italy

Solano Campos

July 2019

Introduction

For many tourists, enjoy the real culture of the country that you are visiting is a goal. We all know that Italy is very famous for the pizza and the pasta, but, essentially in Toscana Region, a region where the tradition of the agriculture and pecuary have maintained along of the years, the creation of high quality sheep's have been develop to a free system of native alimentation, which corresponds to the highest quality of the meat. Italians love meat, being it from cows, sheep's, pigs, birds or deer's, but most part of the meats are used in sophisticated plates or as a side dish, this way, just the better cuts from the meat are utilized, and the others are sold for banana's price. In Uruguay and South of Brazil, the climate and the creation of the sheep are very similar to the Toscana model, in the other hand, the meat is the main dish of the lunch or dinner, ate in a barbecue without any side dish. This has happened in Italy when the hungry have hitted Europe, but, as the conditions were getting better, the European stopped to innovate and maintain the tradition and started to eat just noble parts of meat. Regain this culture, in the hearth of Toscana, would be not just lucrative, but a very important step towards perpetuating the roots of a people with a lot of history.

Business Problem

The objective of this Capstone Project is to analyze and select the best location in Florence to open a new Steakhouse. Building all the Project from zero, passing thought the entire methodology, data architecture and machine learning techniques to improve the algorithms along time. Providing the best solution that answers the question: At Florence, if a business man is looking to open a new restaurant, what kind and where you will recommend that he open it?

Target Audience of this project

This project is useful for any entrepreneur with will to create history, any investor that wants to enter in one of the most competitive markets from Italy, that is the Food Industry, with a different idea that will connect people through past to future, perpetuating a culture that would be forbidden. Sheep Breeders are also interested in this project regarding their production and the will that they have to serve their own country instead of keep exporting most part of the meat produced in their fields, as the historical quote says: From the earth, to the earth.

Data

- List of neighborhoods in Florence. This defines the scope of this project, which is restricted to Florence city in Toscana District at Italy.
- Latitude and Longitude coordinates of those neighborhoods, it's required in order to plot maps, discover venues, explore the city and develop the clusters.
- Venue Data, the most important data of the project, will be used as the main variable to perform machine learning clustering and all exploration of the region.

Sources of Data and Methods to extract them

The Wikipedia page(https://commons.wikimedia.org/wiki/Category:Neighborhoods_in_Florence) contains a list of neighborhoods in Florence, with a total of 33. 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 the latitude and longitude coordinates of the neighborhoods.

After that, will be used 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 various categories of the venue data, we are particularly interested in the food industries categories 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, will be presented the Methodology section where will be discussed the steps taken in this project, the data analysis that we did and the machine learning technique that was used.