IBM Applied Data Science Capstone

Opening a new Chinese Restaurant in Milan, Italy

Nunzio Logallo

August 2020

Introduction

Nowadays Traditional Cuisine Restaurant are an essential part of Multicultural Cities like Milan. In each big city we have a lot of different restaurant like Japanese, Korean, Mexican, Turkish and so on. Anyway, if we look at the competition part of opening a new restaurant in big cities there is a difference between restaurant of different culture, for example, it would be easier to open a Libanese restaurant than a Sushi Restaurant because of the number of opened Libanese restaurants and the consequent low competition. Opening a Chinese Restaurant is even more difficult than Sushi Restaurant in Milan because of the high number of them and the presence of a specific neighborhood called “Chinatown”, anyway, there are busy zones without any Chinese Restaurant.

Business Problem

The main goal of this project is to analyze all the neighborhoods in Milan and select the best locations to open a new Chinese Restaurant. Using Data Science Methodology and Machine Learning this project aims to answer a question: “If an entrepreneur is looking to open a new Chinese Restaurant in Milan, where would you recommend to open it?”.

Data

Obviously, Data and their sources are very important in this project and more specifically we need these data:

- List of neighborhoods in the city of Milan where we should be careful to select only the main city and not the metropolitan area;
- Latitude and Longitude of the neighborhoods, very important to plot a map and get venues data;
- Venues data, related to Chinese Restaurant, we will use it to perform clustering on neighborhoods.

Sources and extraction

Respectively, the sources and the methods to extract the data above are:

- Wikipedia (https://en.wikipedia.org/wiki/Category:Districts_of_Milan), where there is a list of 77 neighborhoods. We can web scraping technique using the Python libraries “beautifulsoup” and “requests” to extract the names of the neighborhoods and put them inside a Python list;
- Python Geocoder package, to extract the latitude and longitude of each neighborhood in the list, being careful about searching the neighborhoods inside the City of Milan instead of the metropolitan area;

- Foursquare API to extract the venue data of each neighborhood. Foursquare is one of the best sources to extract this type of data, having an extremely large database of 105+ million places and used by 150000 developers. With Foursquare we can filter the 900 available categories of venues and get only one we need: Chinese Restaurant.