Capstone Project of IBM Data Science Specialisation

By Andrea Di Sarra

Select optimum location to start up a new wine cellar shop in the City of Milan

1. Introduction

1.1 Introduction

The city of Milan is one of the most interesting Business Case in Italy for its proximity to other major European capitals and for an undoubtable ecosystem which attract investors, therefore creating an hub whereas most of the Italians, especially the southern, are willing to move.

Milan is also famous all over the world for the fashion industry, the city centre through the famous "Quadrilatero della Moda" has a variety of fashion shops of the most important brands (i.e. Prada, Gucci etc.) which is home for travellers and fashion bloggers.

The city is hosting the headquarters of many companies, included but not limited to those of the energy or engineering sectors, the financial / stock market, legal firms etc. Milan is also home of top ranked universities which enhance the variety of the population which is established in the city.

It is also well-known that **tourists** from abroad are attracted from **Italian cuisine** and **wines**, and those three components are key pillars of the Italian economy.

1.2 Business Problem

The objective of this capstone is to analyse, study and select the most attractive place where to establish a new wine cellar shop where to showcase Italians' wines.

The business question behind this objective is therefore the following: if a private investor in the field of the wine production is interested to open up a new shop in the city of Milan, where do you would recommend to start up the activity?

1.3 Target Audience

There are so many excellent wine producers in Italy, as well as company which are active in the distribution of such products. Any of those players, the owner of a selected brand, re-seller or private investor who want to start up a new business, might be willing to understand optimum location where to establish their shop in one of the most diverse city of Italy.

2. Data Description & Sources

- The nine boroughs of Milan

  https://en.wikipedia.org/wiki/Municipalities_of_Milan

- The postal codes of Milan

  http://dati.comune.milano.it/dataset/ds634-numeri-civici-coordinate

- The Latitude and Longitude of boroughs and districts as required is taken from geocoder

- Venue data, paying particular attention to the restaurants & wine shops to cluster the city of Milan accordingly.

3. Methods and Application of Data Science toolkit

The data as resulted from web scarping, are imported in Python.

Data wrangling and cleaning is performed through the help of beautiful soap package and data transformation is done to couple the boroughs of Milan with respective postal code to enable afterward the utilisation of geocoder to get latitudes and longitudes coordinates.

The Foursquare API played major role to find out venues information for the comprehensive dataframe we have been building and ultimately allowed to find out 10 most common venues for each location.

Clustering methods through application of K-means algorithm has been used to characterise the city and its districts, therefore through ranking of main activities to understand the nature of existing businesses.

Foilum library is applied to facilitate the data visualization to make faster and most informed decisions.

4. Results & Discussions

Milan neighbourhoods are plotted on the map to have more visual insights on the locations with respect to the city center (Fig.1).
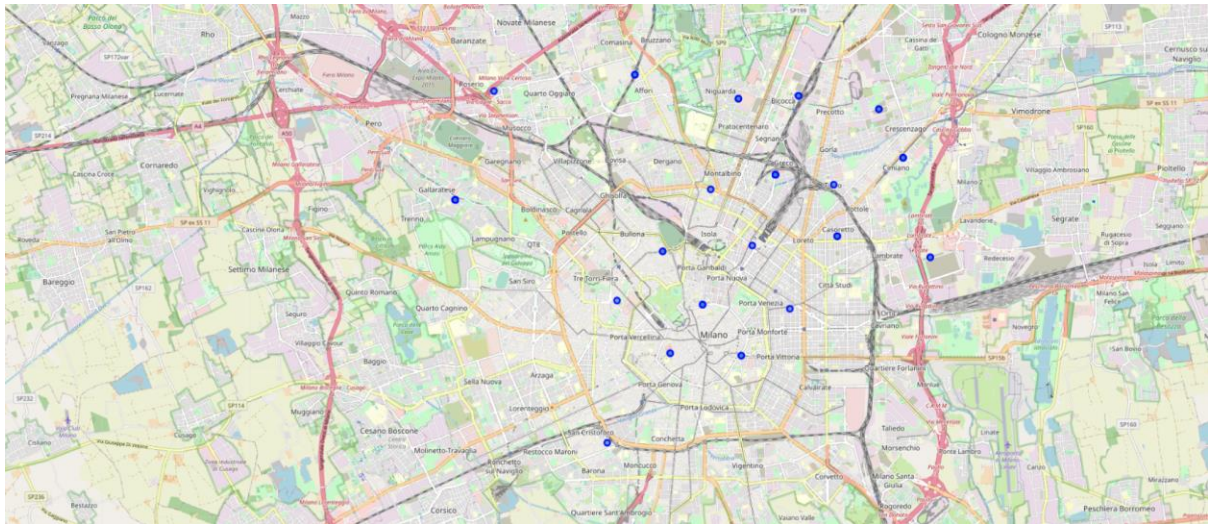
Fig. 1 – Milan neighbourhoods maps.

K-means clustering application will allow to perform clustering of subject data and once more the mapping will enable quick overlook and understating.
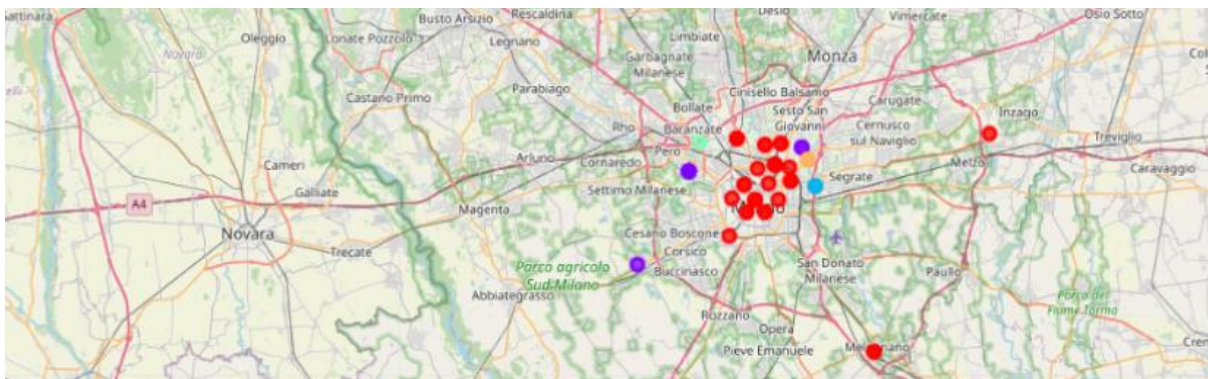

Fig. 2 – Milan neighbourhoods clusters.

As main outcomes, the most interesting clusters for the objective of such project are as follows:

- **Cluster 0** has most food & entertainment businesses and many cocktail bars which suggest an intense nigh life.

- **Cluster 1** has also similar businesses however clothing stores, bookstores are also available.

5. Conclusion section

Wine shop conceived as a re-seller of wine product shall be treated differently from wine or cocktail bar and are recommended in area like those identified in Cluster 1.

Adriano neighbourhood which is under Cluster 1 is a good candidate as, from available data showcase also high Density population/inhabitans/km2.

It's located in the north east of the city, out from city center which also suggest in principle affordable rental prices. Frequency of Wine shops and wine bars, according to Foursquare are null, therefore it's definitely the recommended location to start up wine shop business.