DATA SCIENCE CAPSTONE PROJECT

BY ANDREA DI SARRA

SELECT OPTIMUM LOCATION TO START UP WINE CELLAR SHOP IN THE CITY OF MILAN

The city of Milan is one of the most interesting Business Case in Italy for the following:

- Its Proximity to other major European capitals;

- The fashion Capital through the famous "Quadrilatero della Moda" host major brands

- Home for travellers and fashion bloggers

- Major holding already chosen Milan as the optimum hub for their headquarters.

Milan city is an undoubtable ecosystem which attract investors.

BUSINESS IDEA

Tourism, Italian cuisine and **wines** are key pillars of the Italian economy.

Milan might be the successful Business Case where to open-up a wine shop.

Idea for who?

Any private investor or wine producer/re-seller willing to leverage on this idea could find

the solution through application of machine learning algorithm.

DATA REQUIREMENTS

- The nine boroughs of Milan city : https://en.wikipedia.org/wiki/Municipalities_of_Milan

- The postal codes of Milan city : http://dati.comune.milano.it/dataset/ds634-numeri-civici-coordinate

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

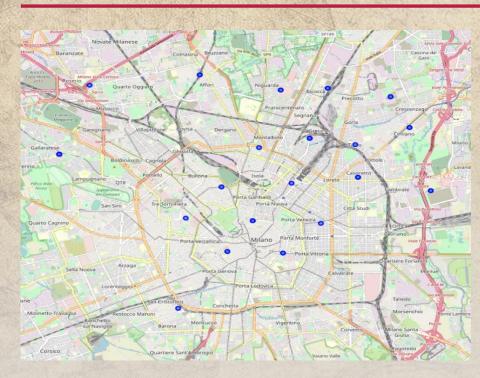- Venue data, paying particular attention to restaurants & wine shops to cluster the city of Milan accordingly.
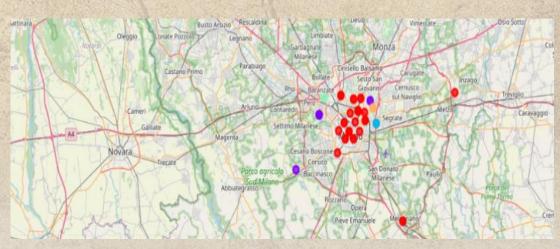
METHODS AND APPLICATION OF DATA SCIENCE TOOLKIT

1. Web Scraping and data importing to Python

2. Data wrangling and cleansing via beautiful soap package

3. Data merging to build up dataframe of interest

4. Foursquare API call to find out venues information for selected neighbourhoods

5. K-means algorithm application to cluster such data

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

DATA VISUALIZATION

K-means clustering application allows to perform clustering of venues data as obtained trough Foursquare API calls.

Quick look on the mapping is granted by Folium.

Milan neighbourhoods are plotted on the map to have more visual insights on the locations with respect to the city center.

CONCLUSIONS & RECOMMENDATIONS

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

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

- Wine shop conceived as a re-seller of wine product are recommended in area like those identified in **Cluster 1.**

- **Conclusions:** 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 the recommended location to start up new wine shop business.