Project Report

Applied Data Science Capstone

The Battle of Neighborhoods

Business Problem:

I live in Italy and Italy is famous throughout the world for its cuisine. You can find different cuisines from different regions of Italy in the cities (specially the major cities), and if you are looking for food from other countries, there are many options as well.

Fast foods are quite popular nowadays, specially among students and workers, who just want to grab a quick bite and get back to their studies or work. Among many fast food options, one that is quite popular in Italy (and many other places) is Turkish Kebab.

For a hypothetical client who wants to open a Turkish Kebab restaurant in Milan, I will look into and analyse the spread of the Turkish Kebab restaurants that are already there in the city, and come up with a suggestion on where would be better to open such a restaurant.

Data:

I chose the city of Milan to perform this analysis on, as it has more and better data (accuracy) compared to other cities in Italy. To perform this analysis I will use 2 data sources: Wikipedia, Foursquare.

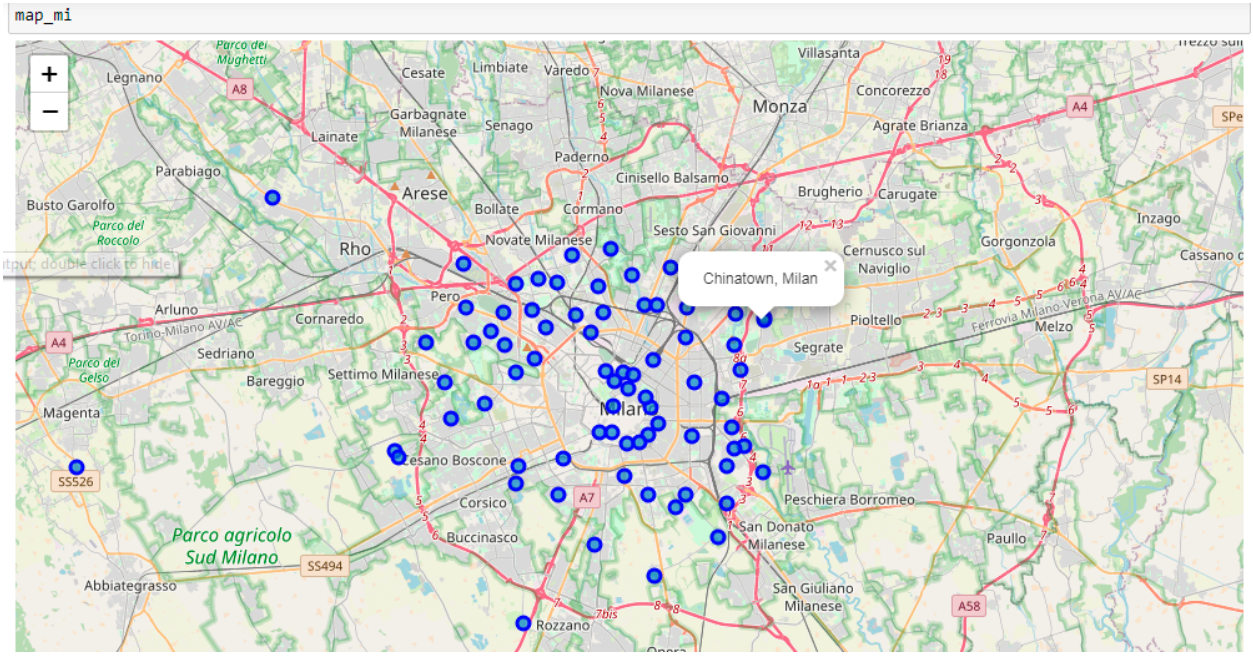
Wikipedia
I will use this Wikipedia page to get to know the neighborhoods of Milan by scraping the page using BeatifulSoup.

Foursquare
I will use the Foursquare API to explore and fetch the venues that are present in a neighborhood. After entering my personal API credentials such as client_id, client_secret, and version, we can access the API and can retrieve the top 200 venues that are within a radius of 2000 meters.

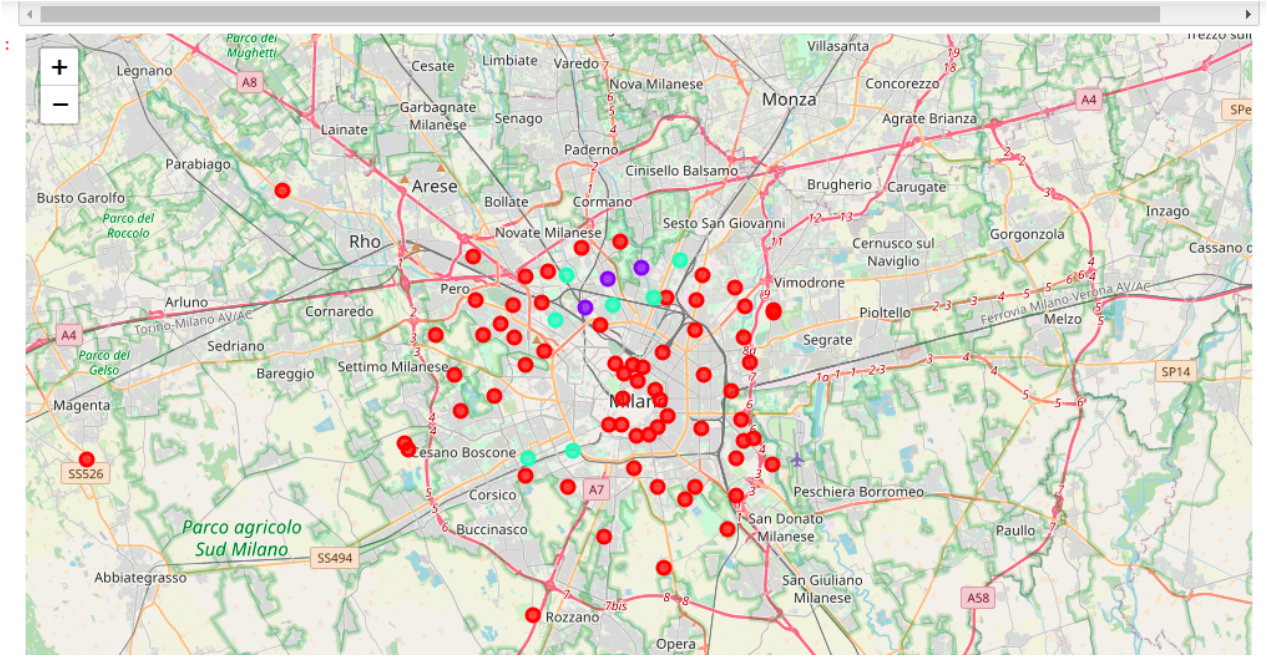
Methodology:

The methodology used in this project consisted of two components; the identification of neighborhoods in Milan, finding their geographical coordinates and visualizing them on the map. Then the venues were retrieved in those neighborhoods using the Foursqaure API. Finally the KMeans algorithm was used to cluster those venues, with number of clusters set to 3.

Here is the map of Milan with its neighborhoods:



And here are the clusters:



Results:

As seen from the output tables and clustered map, the best areas to open up a Turkish Kebab restaurant are in cluster 0, followed by cluster 2. Meanwhile cluster 1 has the highest number of Turkish Kebab restaurants, so it is not recommended to open up a restaurant in those areas.

Discussion:

Most of the Turkish Kebab restaurants are concentrated in areas that are slightly away from the center of the city of Milan, with the highest rate in cluster 1 and moderate rate in cluster 2.

This represents a great opportunity and high potential to open new Turkish Kebab restaurants as there is little or no competition from existing kebab restaurants in the neighborhoods in cluster 0. Meanwhile, kebab restaurants in the neighborhoods in cluster 1 face quite high competition due to the excess supply and high kebab restaurant density. Therefore, based on this analysis, I would suggest to the clients to open a new kebab restaurant in the neighborhoods of cluster 0. Clients can also open new kebab restaurants in the neighborhoods in cluster 2 with moderate competition. However, I would strongly advise the clients to avoid neighborhoods in cluster 1, which is already a highly concentrated kebab restaurant area.

Conclusion:

This project has demonstrated the potential, capability and practicality of Data Science and Machine Learning concepts. Even with access to limited data, I was able to perform a quite comprehensive analysis and draw very interesting and useful results.