IBM APPLIED DATA SCIENCE CAPSTONE

Recommendation System on the opportunity to open a Wine Bar in a Neighborhood of Toronto

Massimo D'Alessandro June 19th 2019

1 INTRODUCTION

ABSTRACT IDEA:

A new investor has contacted us looking to open a Wine Bar in Toronto. The investor operates in the field of bar and restaurant and has many successful wine bars in the major cities of Italy, mainly in Milan. The investor wants to expand the wine bar activity in Canada starting from Toronto.
The investor is looking for guidance to understand if Toronto could be a successful location, proving if there is a similarity with Milan, and which is the best neighborhood where to locate the Wine Bar.

BUSINESS PROBLEM:

To prove the similarity between Toronto and Milan, a neighborhood segmenting and clustering analysis will be conducted grouping together neighborhoods in Milan and Toronto. The similarity between neighborhoods is measured by comparing the venues in each neighborhood. The type of venues used in this study includes shop, restaurant, gym, museum etc.. After proving a similarity between the two cities a new segmenting and clustering analysis will be conducted to understand which Neighborhoods of Toronto is similar to the neighborhood of Milan where the wine bar is located. Target Neighborhoods of Toronto are those with a lack of Wine bars and similar to Milan neighborhoods where the successful wine bar is located.

This study uses Foursquare API and other data sources to acquire location data.

TARGET AUDIENCE:

The specific target audience is the "sales department" of the company, that is in charge to investigate where to locate new shop around the world.

2 DATA COLLECTION

FOURSQUARE PLACES API:

FOURSQUARE is a social location service that allows users to explore the world around them.
Using a free account, it is possible to download up to 100 venues for request.
THE FOURSQUARE API allows application developers to interact with the Foursquare platform.
The API itself is a RESTful set of addresses to which you can send requests and get responses.
The API allows querying places and users, exploring popular places, and checking out reviews

and photographs for these places.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: Web site where to scrape postal code of Toronto Canada

https://worldpostalcode.com/italy/lombardia/milano': Web site where to scrape postal code of Milan Italy

geolocator.geocode: API to extract the latitude and longitude of postal code.

3 Data Cleaning

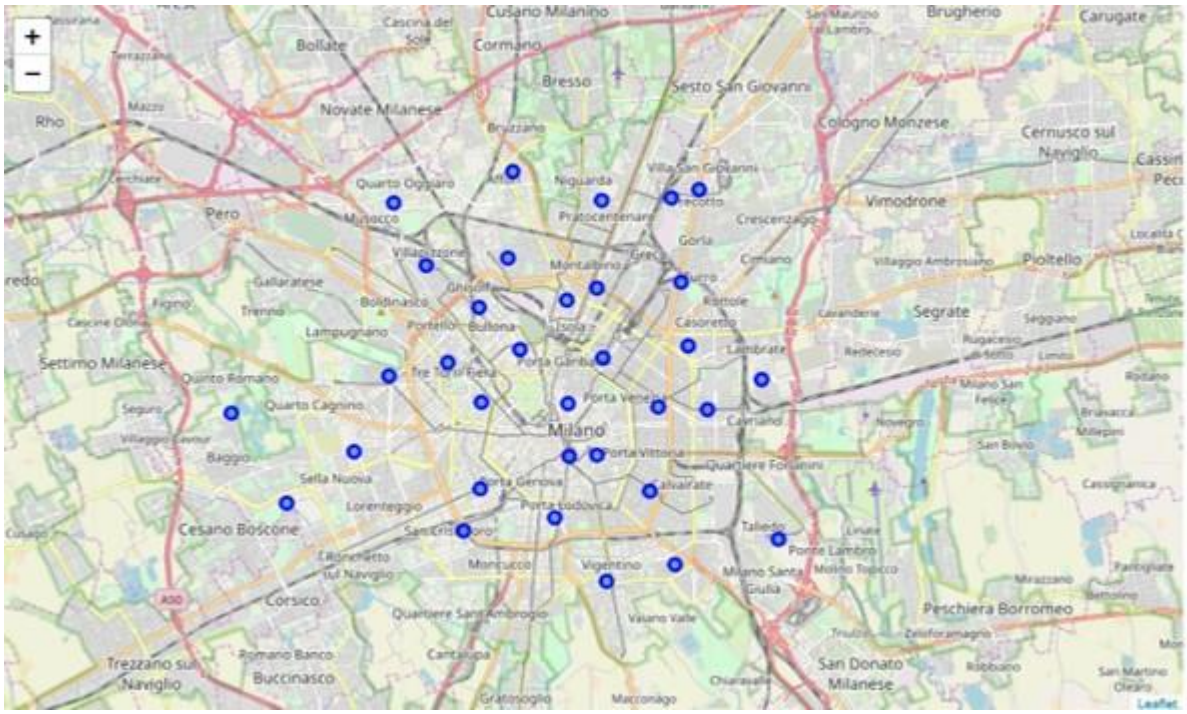
With WEB scraping, the Postal codes Of Milan and Toronto have been extracted; after that, has been extract the Longitude and Latitude of each postal code. Geolocator, for some reason, added a wrong latitude/longitude to 4 Postal code in Milan, for this reason these Postal code have been eliminated.

Venues extracted from “Foursquare” hadn’t wrong value, everything was fine.

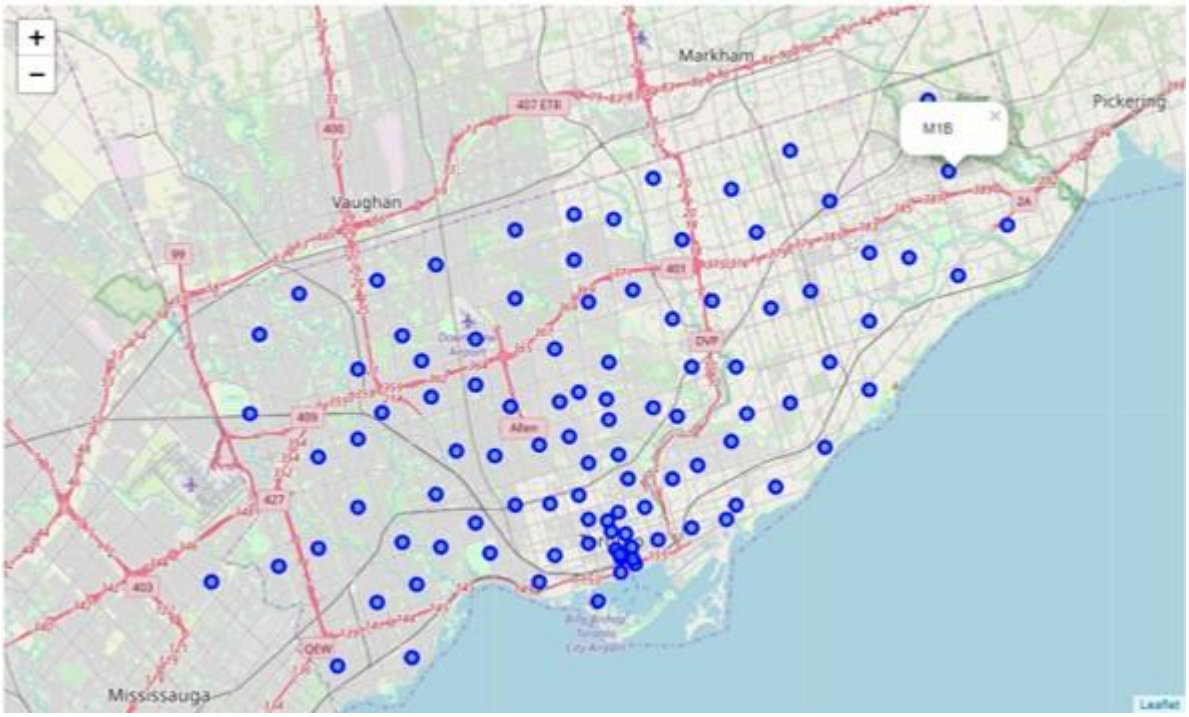
4 Exploratory Data Analysis

- Represented on a map all neighborhoods of Milan and Toronto.
- Evaluated the number of Wine bar per Neighborhood.
- Evaluated the number of venues “wine bar” like per Neighborhood.
- The wine bar of the investor (That’s Wine) is located in Milan in the postal code area 20122; in the same area there are two more Wine bars.
- Represented on the map the location of Wine bars in Milan and Toronto.

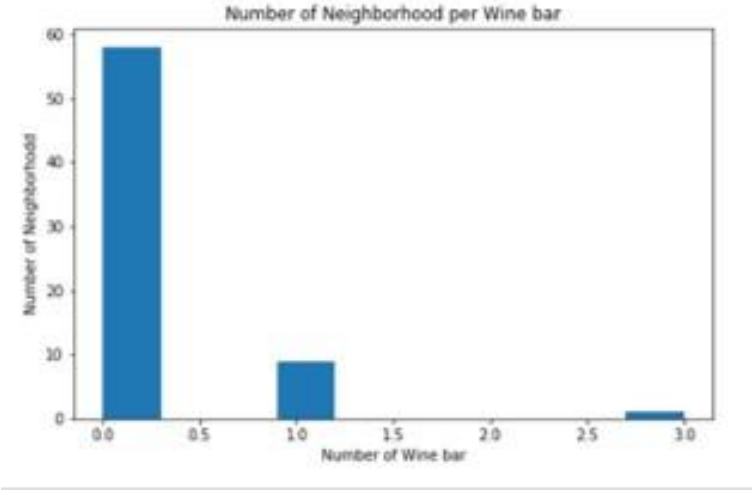
Milano Map with all the Postal codes area



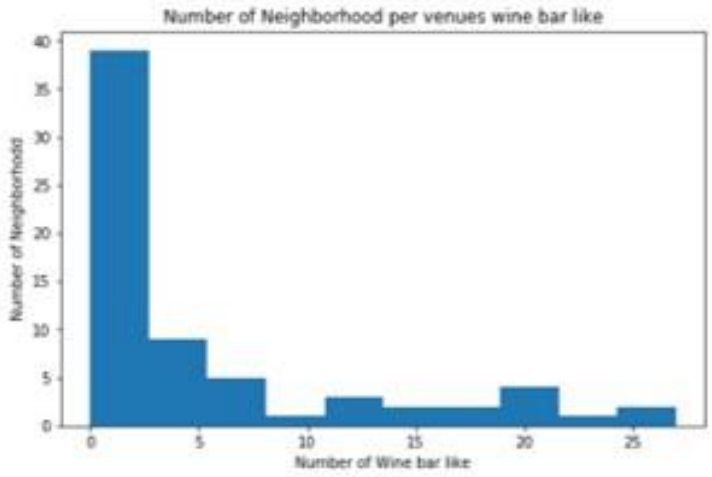
Toronto Map with all the Postal codes area



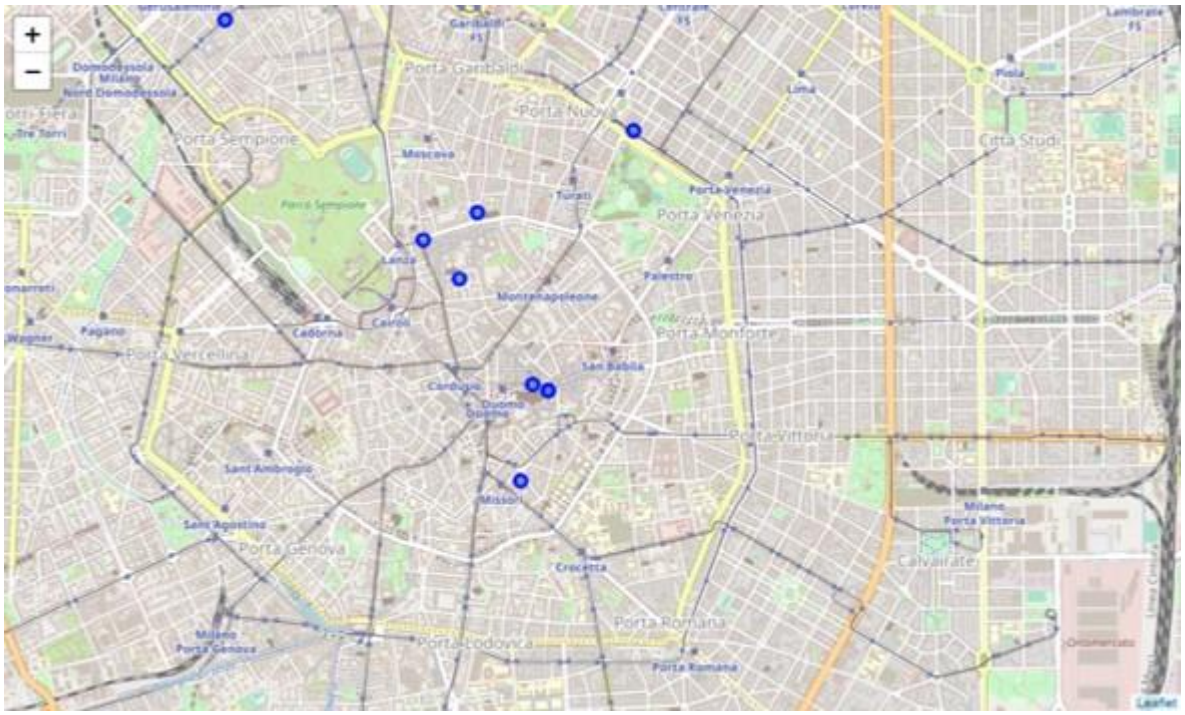
Wine Bar distribution per area: few areas have 1 Wine Bar, only two areas have 3 Wine bars;



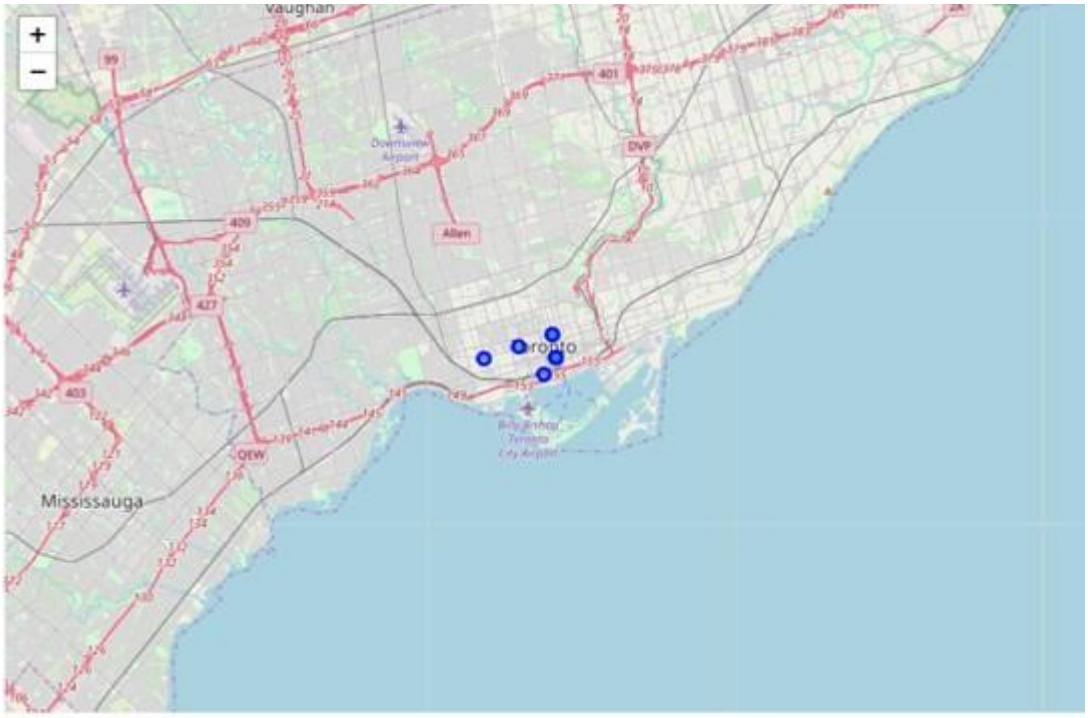
“Wine Bar” like venues distribution per area: in some area there is a high concentration of wine bar like venues; 25 venues in 3 area.



“Wine Bar” per area in Milan:



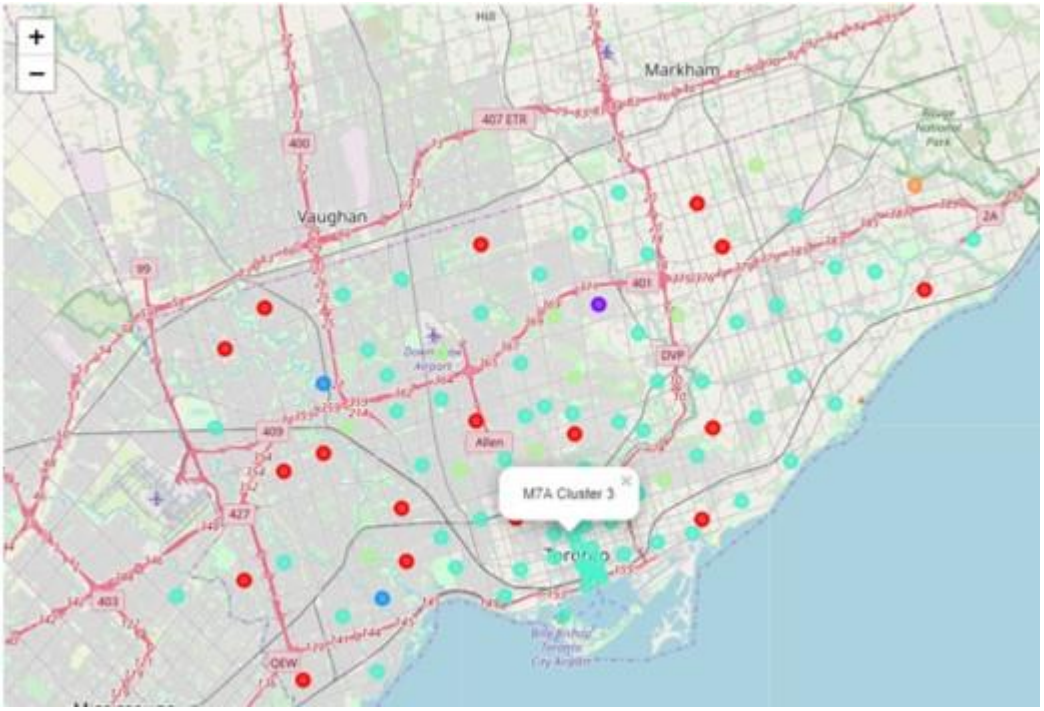
“Wine Bar” per area in Toronto:



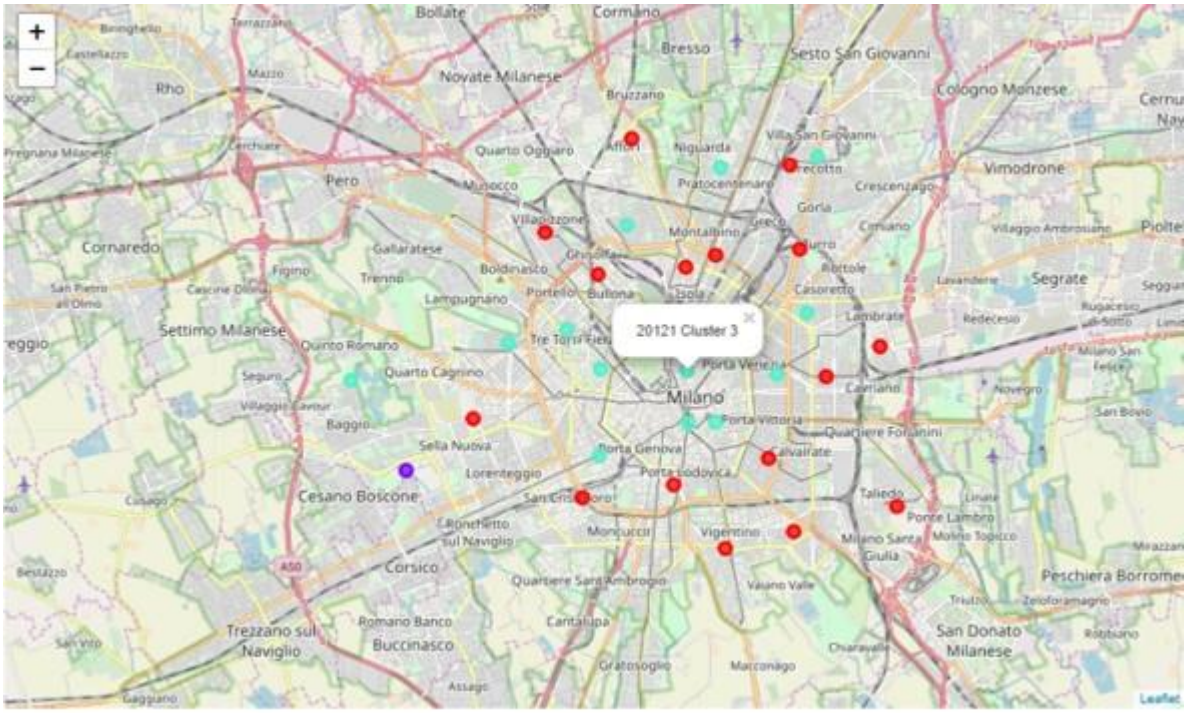
4 Predictive Modeling

The clustering model K-means has been applied to identify similar neighborhoods in Milan and Toronto. This clustering has been applied putting together the neighborhoods of Toronto and Milan which features have been represented by the venues extracted from Foursquare.
A k of 6 has been used.
From the map below it is clear that there is a good similarity between Toronto and Milan, especially in the center area where the Wine bar in Milan is located.

Neighborhood Clusters in Toronto:



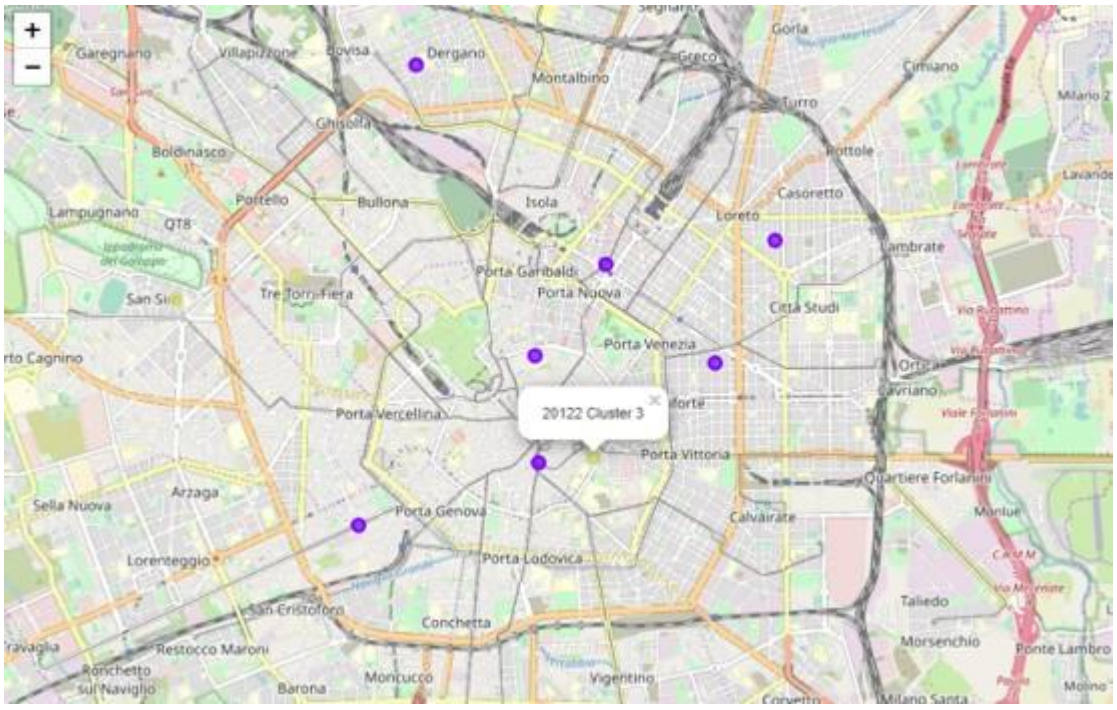
Neighborhood Clusters in Milan:



The cluster number 3 has been identified as the cluster with more similarity between Milano and Toronto and where the Wine Bar in Milan is located.
At this point the cluster 3 has been isolated, both for Milan and Toronto, and a new clustering model has been applied to identify the neighborhoods more like the one where the Wine Bar in Milan is located.
Below the map of Toronto and Milan re-clustering the cluster number 3:

Toronto:

Milan:

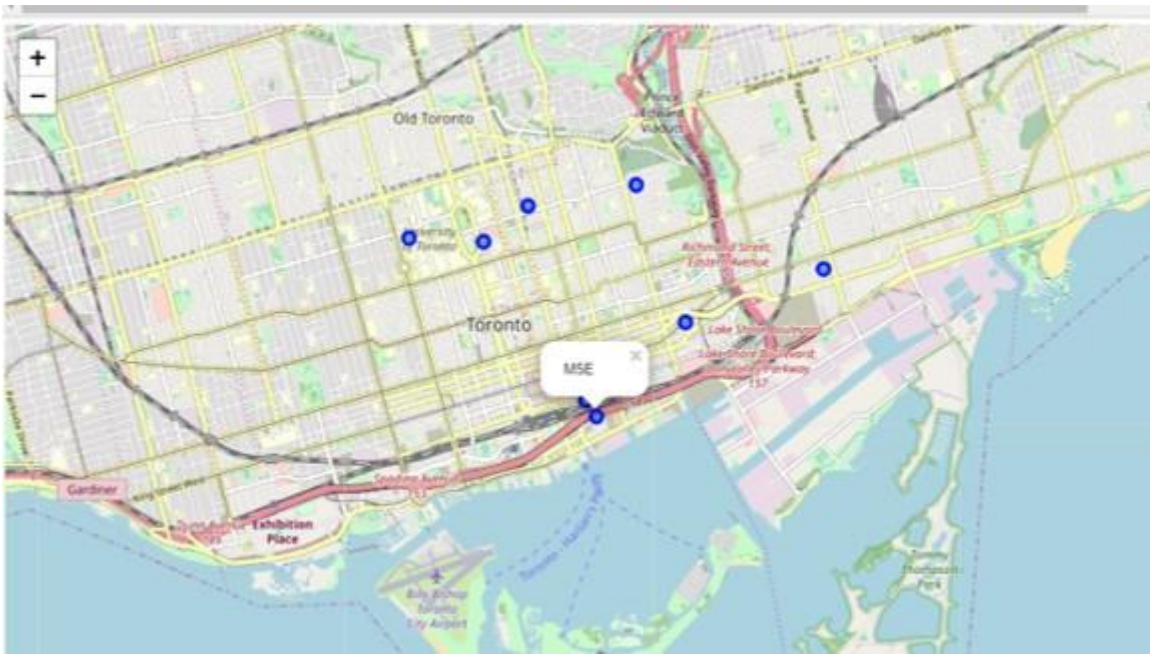


After clustering, it results that the neighborhood in Milan, where the wine bar is located, has numerous similar neighborhoods in Toronto.
We can deduct that the wine bar in Toronto has to be located in the neighborhoods of the cluster 3.
To refine the analysis we can define other conditions to choose the right place.

Condition could be:

- The neighborhood of the cluster has no Wine Bar.
- The neighborhood of the cluster has more than 5 venues like Wine Bar among Wine Shop, Bar, Beer Bar, Café, Cocktail Bar, Coffee Shop.
- The neighborhood is not far from the center, less than 4 KMs.

The following map represent the 9 neighborhoods where the wine bar has to be located, they are the most similar to the neighborhood of Milan where the successful wine bar is located:



5 Conclusion

In this study, I analyzed the relationship between neighborhoods in two cities to find if there is a similarity between them. After the similarity was proved a new clustering model was applied to group together similar neighborhoods in the two cities; similar neighborhoods can have similar activities. In this way it was possible to identify the right neighborhoods in Toronto where to locate a successful Wine Bar already located in Milan. The same exercise can be done for any other activity.