MACHINE LEARNING LOCATION EVALUATION TO START AN ITALIAN RESTAURANT IN TORONTO

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE ON COURSERA

MARCH 2021

ANGELO ANTONIO BIANCAMANO

1. INTRODUCTION

Background

- For this Capstone project, the hypothetical scenario is represented by an Italian who wants to explore opening opportunity of an authentic Italian restaurant in Toronto area.
- The idea behind this project is that there may not be enough Italian restaurants in Toronto. Let's say Italian food is similar to other European cuisines: this Italian person might think to open his restaurant in locations where European food is popular (in other words, having many European restaurants in the neighborhood).
- With this aim, find the right location to open such a restaurant is one of the most important decisions for this Italian person, and with this project I would like to help him find the most suitable location.

1. INTRODUCTION

Business Problem

- The objective of this capstone project is to find the most suitable location for the Italian person to open a new Italian restaurant in Toronto, Canada. The question data science methods and machine learning methods such as clustering can help us to have a solution is: in Toronto, if anyone wants to open an Italian restaurant, where should he consider opening it?

1. INTRODUCTION

Target Audience

- The Italian person who wants to find the best location to open authentic Italian restaurant.

2. DATA

To solve this problem, we need the below list of data:

- List of neighborhoods in Toronto, Canada.
- Latitude and Longitude of these neighborhoods.
- Venues data related to European restaurants: this will help us find the neighborhoods that are most suitable to open an Italian restaurant.

3. EXTRACTING DATA

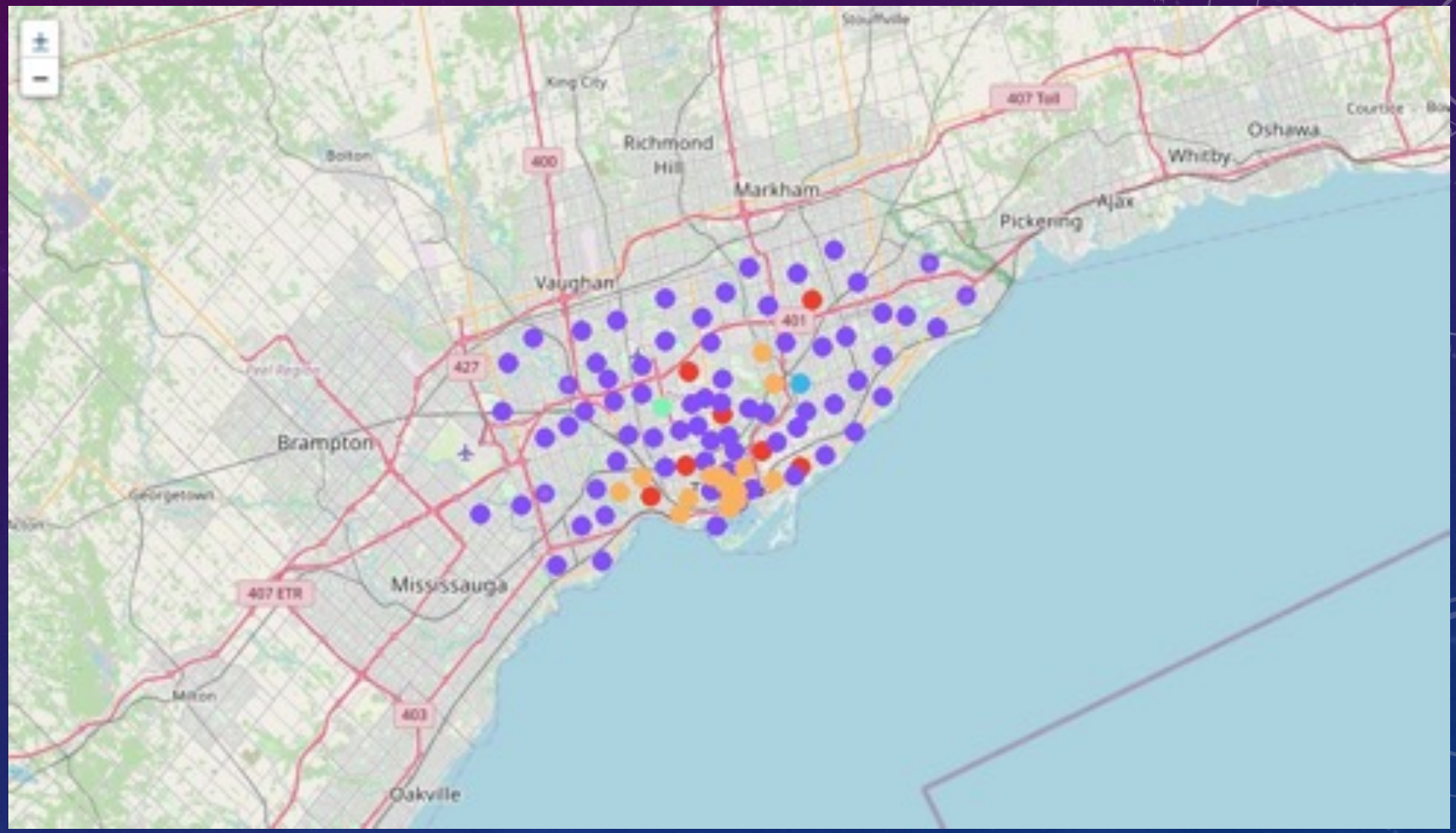
- Scraping of Toronto neighborhoods via Wikipedia.
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package.
- Using Foursquare API to get venue data related to these neighborhoods.

4. METHODOLOGY

- Import all the necessary libraries.
- Extract list of neighborhoods from Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- Pandas html table web page scraping method.
- Build Toronto map with Foursquare, Geocoder and Folium
- Foursquare API to pull the list of top 100 venues within 500 meters radius.
- From Foursquare, I was able to get the names, categories, latitude and longitude of the venues from Foursquare.
- Neighborhood analysis by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category.
- Later, I did some queries to find out typical European Restaurants.
- Clustering by using K-means clustering.

5. RESULTS

- Clusters:



6. RECOMMENDATIONS

- From analysis, it came out that better clusters to start an Italian Restaurants are 0, 2, 3.

7. LIMITATIONS AND SUGGESTS

- It could be useful to extend analysis starting also from other European Restaurants.

8. CONCLUSIONS

- In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

THANK YOU!!!!