Battle of the Neighborhoods

Attique Ur Rehman

# Introduction

As part of Applied Data Science Capstone, this project encompasses the core concepts learned in the Coursera-IBM Data Science Professional Certificate Course. In the given context, this project employed techniques like data acquisition, cleaning, visualization, and analysis. Modeling based on machine learning algorithms is also a key component of this project.

**Business Problem**

Toronto, the capital of the province of Ontario, is one of the major Canadian cities. Being a dynamic metropolis, it attracts people from multi-ethnic groups, consequently it is not only the most densely populated region of the Canada but also provides numerous growth opportunities. In this context, the food business is one of the promising sectors to invest, as being a multicultural city, Toronto offers wide range of cuisines.

To invest in the food business, this project aims to identify the best venue to open a restaurant that offers continental food, not only to better serve the community but also provide an edge to flourish the business.

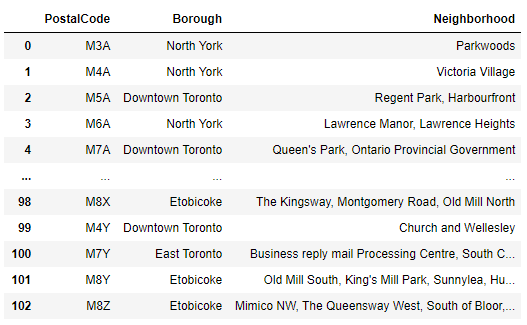
**Target Audience**

The targeting audiences of this project are entrepreneurs and community members. As from the best location/venue of the continental restaurant, the entrepreneurs can benefit in terms of more profit from their business. On the other side, the best location of the continental restaurant will also facilitate the community members.

# Data Description

# As Toronto being the intended city to open the continental-food restaurant, the postal code data of Toronto is acquired from Wikipedia (the web-link to the acquired data is provided as, <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>). The acquired data is transformed to *pandas* dataframe for further processing. The extracted dataframe consist of 3 columns namely, Postal Codes, Borough, and Neighborhood, which are further cleaned for further visualization and analysis. The geographical coordinates of each postal code are obtained from <https://cocl.us/Geospatial_data>. Moreover, *Geocoder* and *Foursquare* are also employed in this project for visualizing and pertaining the venue data for geographical location suitability.

Some of the results snapshots of the acquired data is shown in Figure 1-2



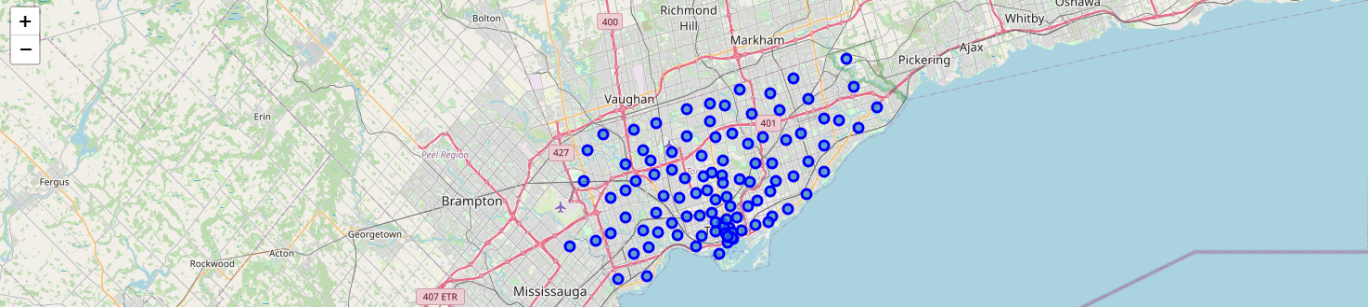
**Figure 1. Acquired Data from** [**https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)



**Figure 2. Data merged with the geographical coordinates acquired from** [**https://cocl.us/Geospatial\_data**](https://cocl.us/Geospatial_data)

# Results

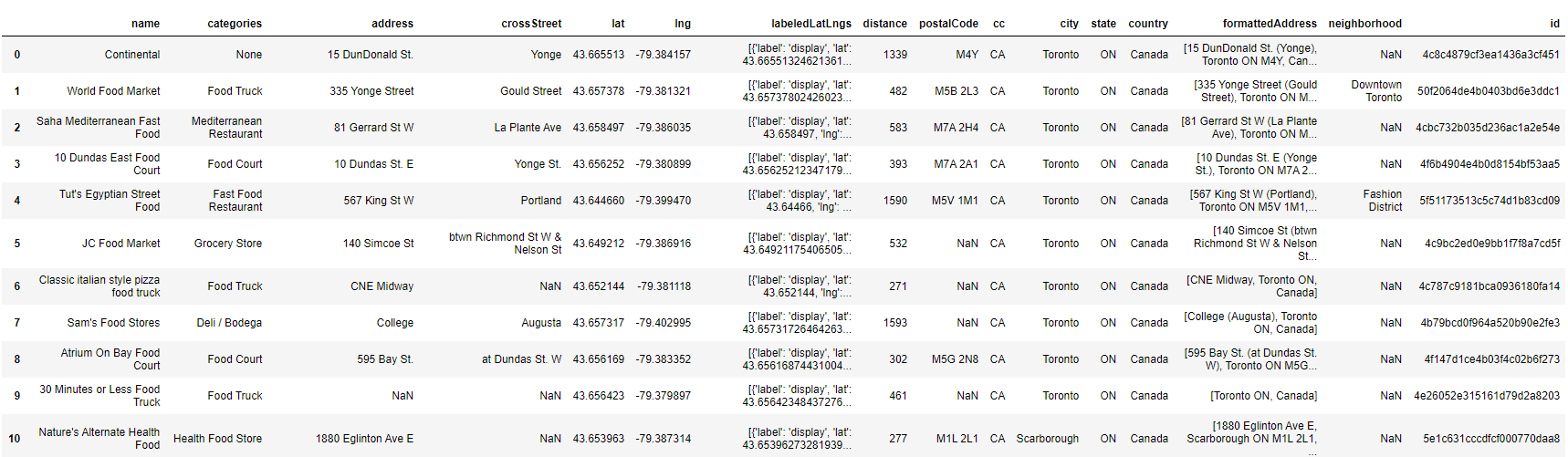
The acquired data is cleaned and geospatially analysed in the context of the neighborhoods in Toronto. In the given context, Figure 3 shows the results.

****

**Figure 3. Neighborhoods in Toronto** (represented by blue circles)

Further, Foursquare API is employed to search for existing continental food restaurant and the existing restaurant are mapped on the geographical map to extract their locations. However, for said purposes the extracted data related to continental food is thoroughly cleaned for further processing. Table 1 present the cleaned data in the given context, where Table 2 highlights the names of restaurants offering continental food.

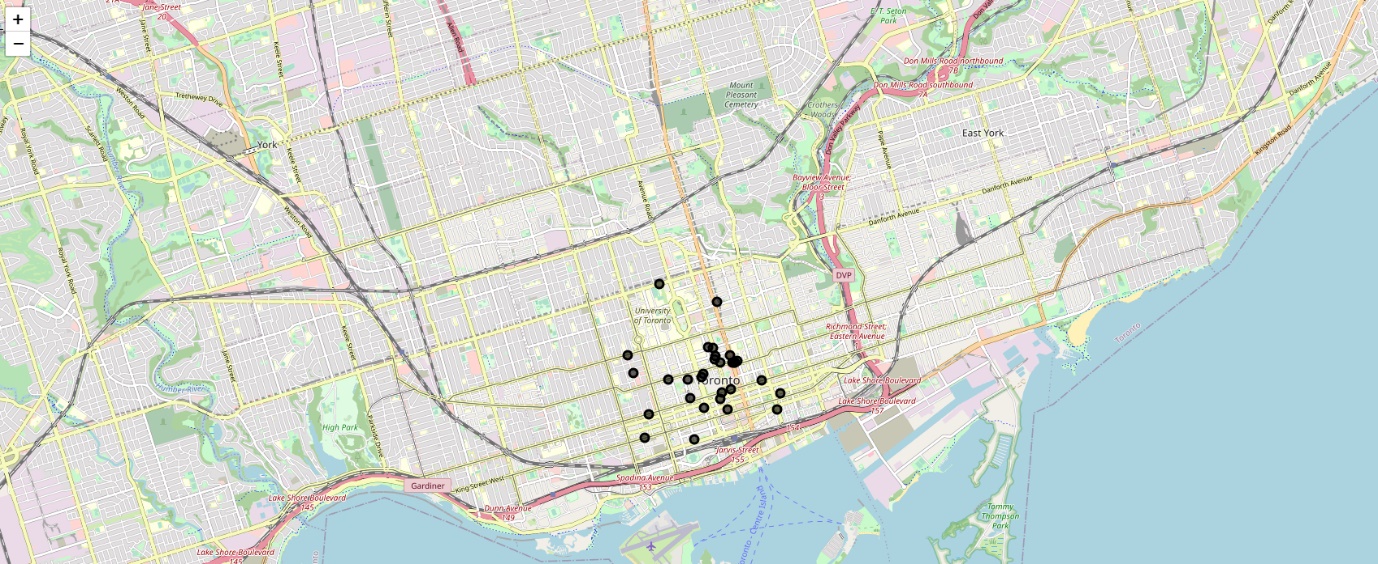
**Table 1. Information of restaurants offering continental food**



**Table 2. Information of restaurants offering continental food**

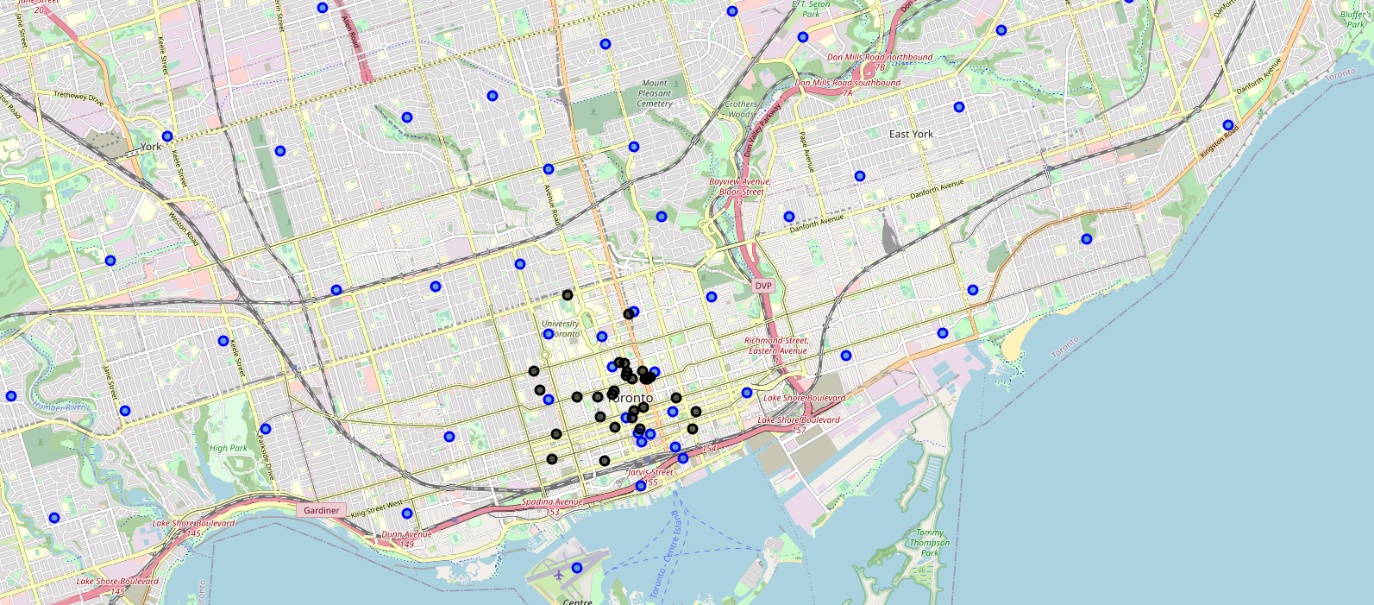


In the given context, Figure 4 shows the geographical locations of the existing restaurants offering continental food in the Toronto region.



**Figure 4. Restaurants offering Continental food in Toronto** (represented by black circles)

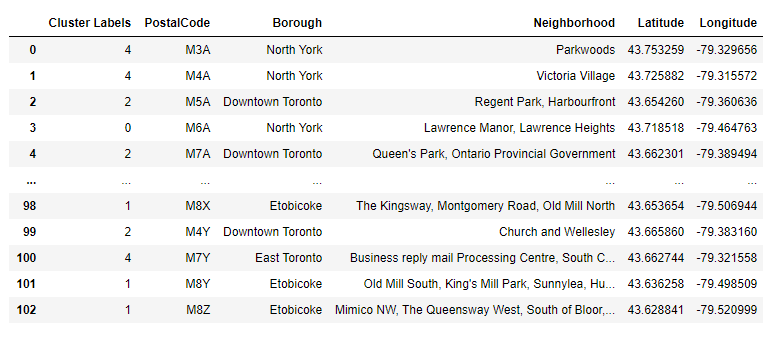
The existing restaurants are also mapped along with the postal codes to better visualize the existing of the restaurants against different locations. Figure 5 shows the corresponding results.

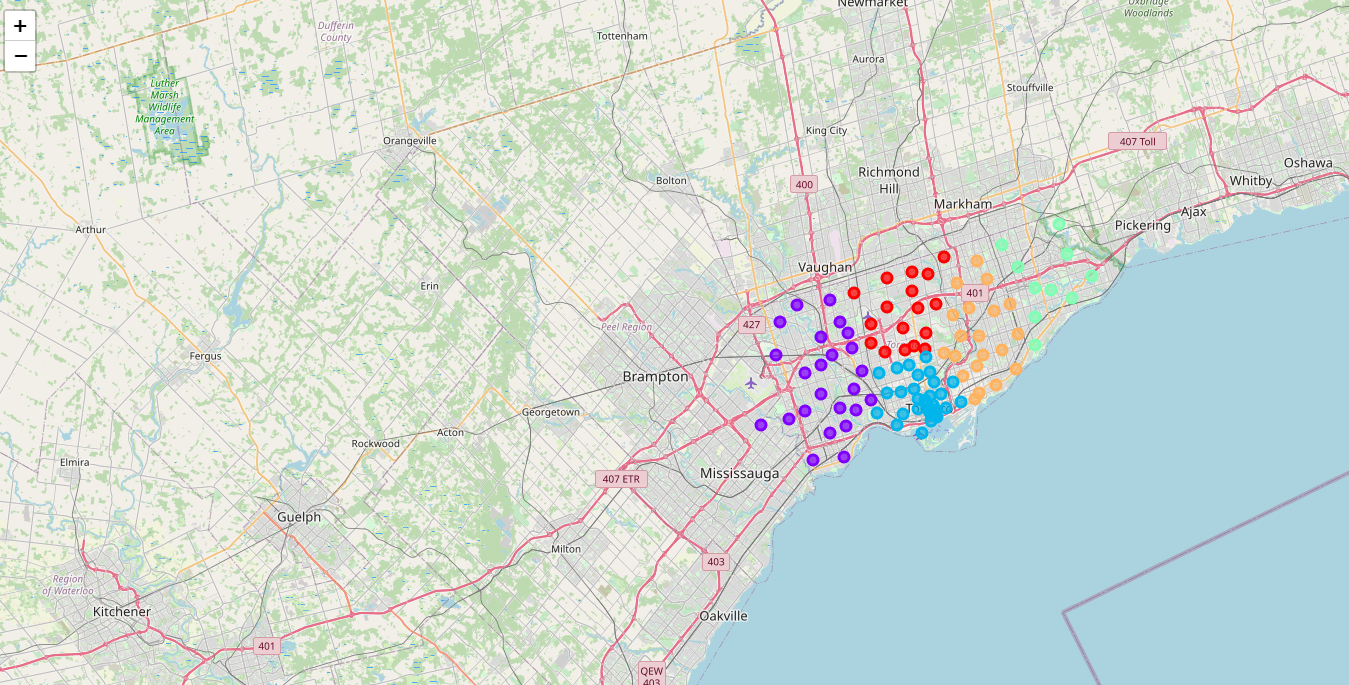


**Figure 5. Restaurants vs Postal Codes** (in black and blue, respectively)

The neighborhoods are also clustered using k-means algorithm, the corresponding results are presented in Table 3 and Figure 6.

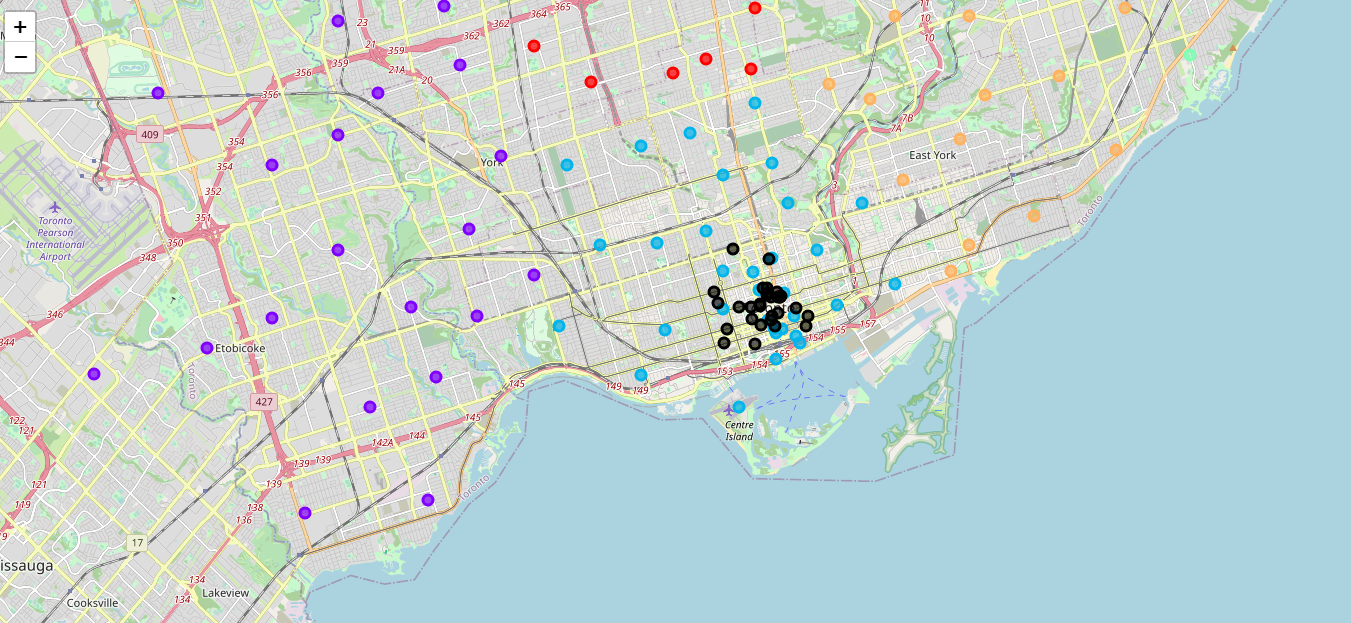
**Table 3. Clustering**





**Figure 6. Clustering**

The clustering information is also mapped against the location of the restaurants offering continental food in Toronto. Figure 7 presents the corresponding results.



**Figure 7. Clustering and Restaurants Locations**

**Discussion and Conclusion**

It is observed from the extracted results, under given conditions, that all the continental food restaurants are in cluster 2 (presented in light blue color), where clusters 0, 1, 3, and 4 does not have any restaurant that offers continental food. Hence for opening a new restaurant that offers continental food, it is better to invest in the regions that lies in clusters 0, 1, 3, and 4. Moreover, even within cluster 2 there are locations (as seen in Figure 7) that possess potential for opening a restaurant offering continental food.