**Strategic Analysis for Establishing a Restaurant**

**A. Introduction**

**A.1 Background**

As part of a Data Science Analysis, I am taking an example of Toronto city to explore it for establishing a Restaurant. For this analysis, I have considered using an Indian Restaurant, but this analysis can be used for any type of restaurant.

Toronto is the provincial capital. It is one of the most multicultural and multiracial cities in the world. In 2016, 51.5% of the residents of the city proper belonged to a visible minority group, compared with 49.1% in 2011, and 13.6% in 1981. Toronto also has established ethnic neighbourhoods such as the multiple Chinatowns, Little Italy, Little India and Roncesvalles, which celebrate the city's multiculturalism. With around 11% Indian population, it will be interesting to establish an Indian restaurant but just need to find the best location for it.

**A.2 Business Problem**

To establish an Indian restaurant in Toronto, the problem is to locate the strategic location which would be the optimal option in order to maximize business profits.

The objective of this project is will mainly focus on this. It will use Foursquare API to get various location details and will use relevant data from Toronto Open Data Portal.

**A.3 Interest**

Obviously, investors will be very interested in looking for the best location for its restaurant as location plays a dominant role for its success. Others who wants to explore the neighbourhoods may also be interested.

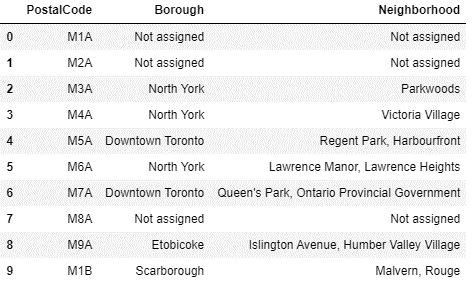
**B. Data Description**

To consider the problem we can list the data as below:

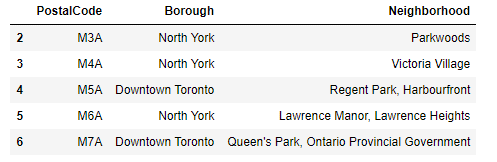
1. Information about Toronto list of Postal Codes in Canada can be obtained from Wikipedia page <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>. It has required details like Postal code, Borough and Neighbourhood. Data is required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas data frame so that it is in a structured format dataset.
2. Geospatial data is used to get the latitude and longitude of the location <http://cocl.us/Geospatial_data/Geospatial_data.csv>.
3. Toronto Open Data Portal <https://open.toronto.ca/> is used to get Details of City of Toronto Neighbourhood Profile <https://open.toronto.ca/dataset/neighbourhood-profiles/> . It is used to get the population detail of all areas.
4. Boundaries details of city of Toronto Neighbourhoods is taken from Toronto Open Data Portal <https://open.toronto.ca/dataset/neighbourhoods/>
5. Foursquare is a technology company that built a massive dataset of location data <https://developer.foursquare.com/>. Foursquare API can be used to get information regarding different venues in Toronto, along with its latitude and longitude.

**C. Methodology**

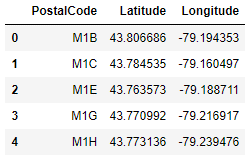
The postal codes in Canada, which begins with M are located within the city of Toronto. I scrapped the data from Wikipedia page using BeautifulSoup python library. This gave me the information about all Postal Codes, Boroughs and its Neighbourhood. After scrapping, the raw data looks like:



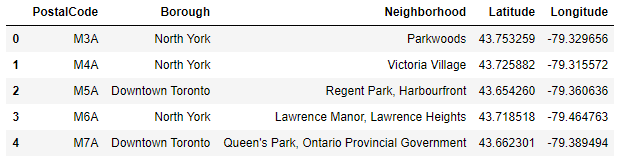
Data is cleaned by removing Postal codes where Borough is “Not assigned”. After cleaning, my data frame looks like:

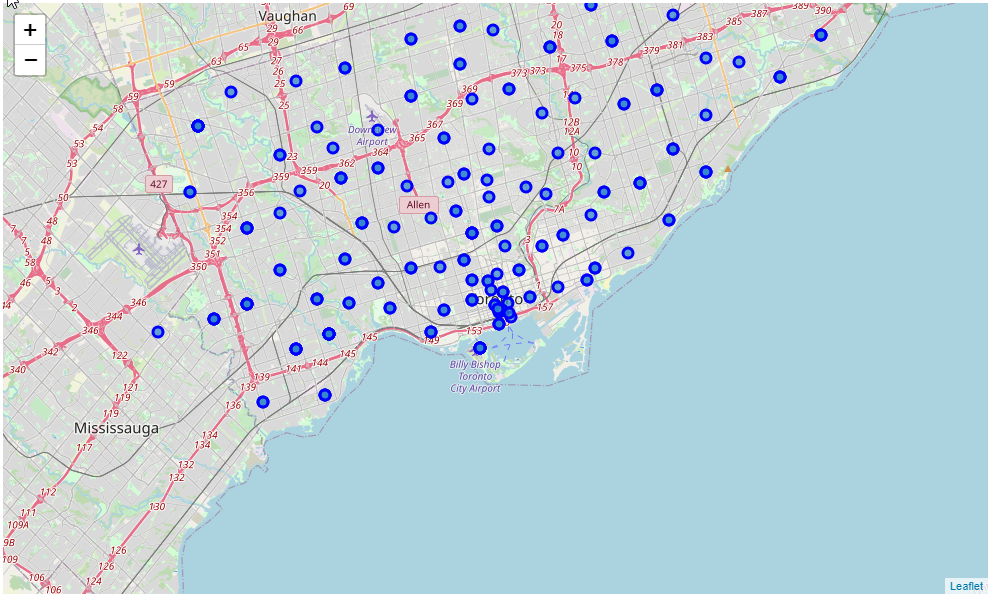


Using Geospatial data provides the latitude and longitude of the Postal Codes. The raw Geospatial data looks like:

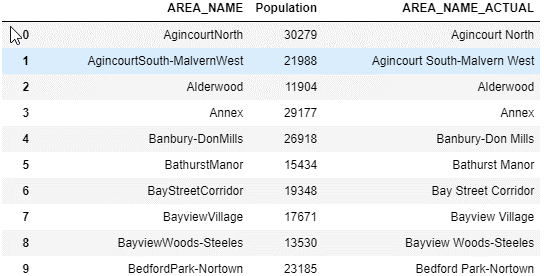


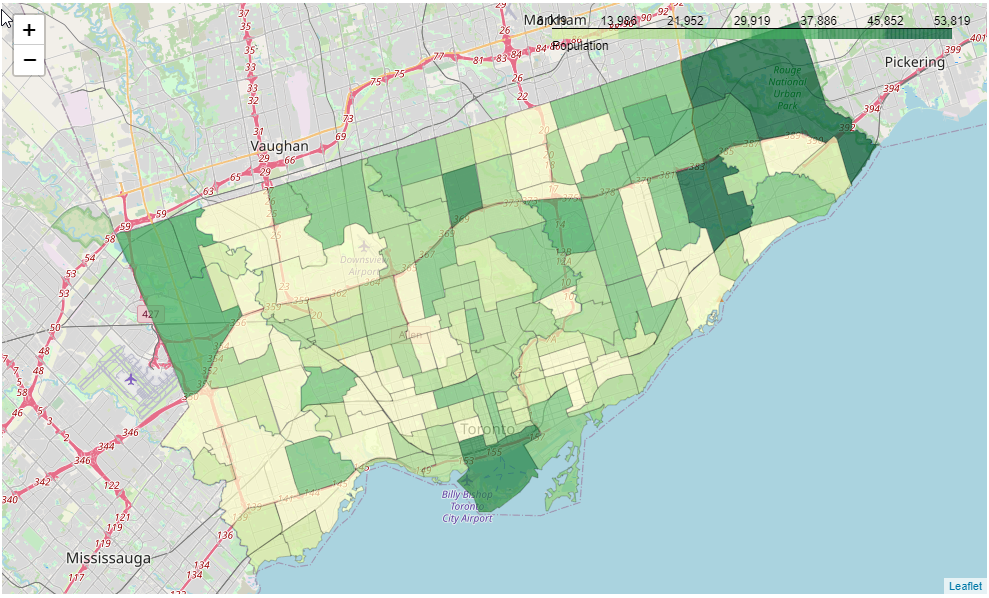
I used python folium library to visualize geographic details of Toronto and its boroughs and I created a map of Toronto with boroughs superimposed on top. I used latitude and longitude to get the visual as below:



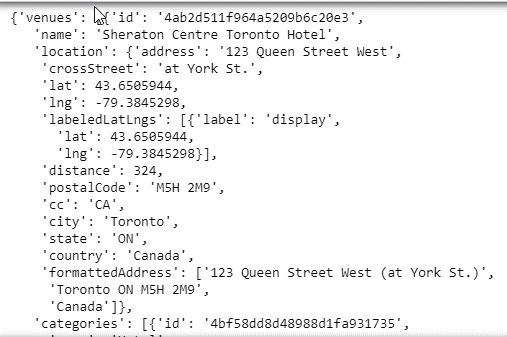


From Toronto Open Data, I retrieved the population details of all Neighbourhoods and prepared a folium map to represent it.

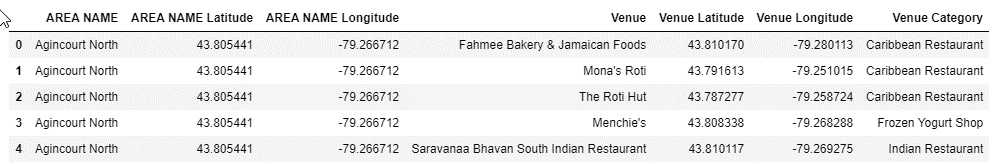


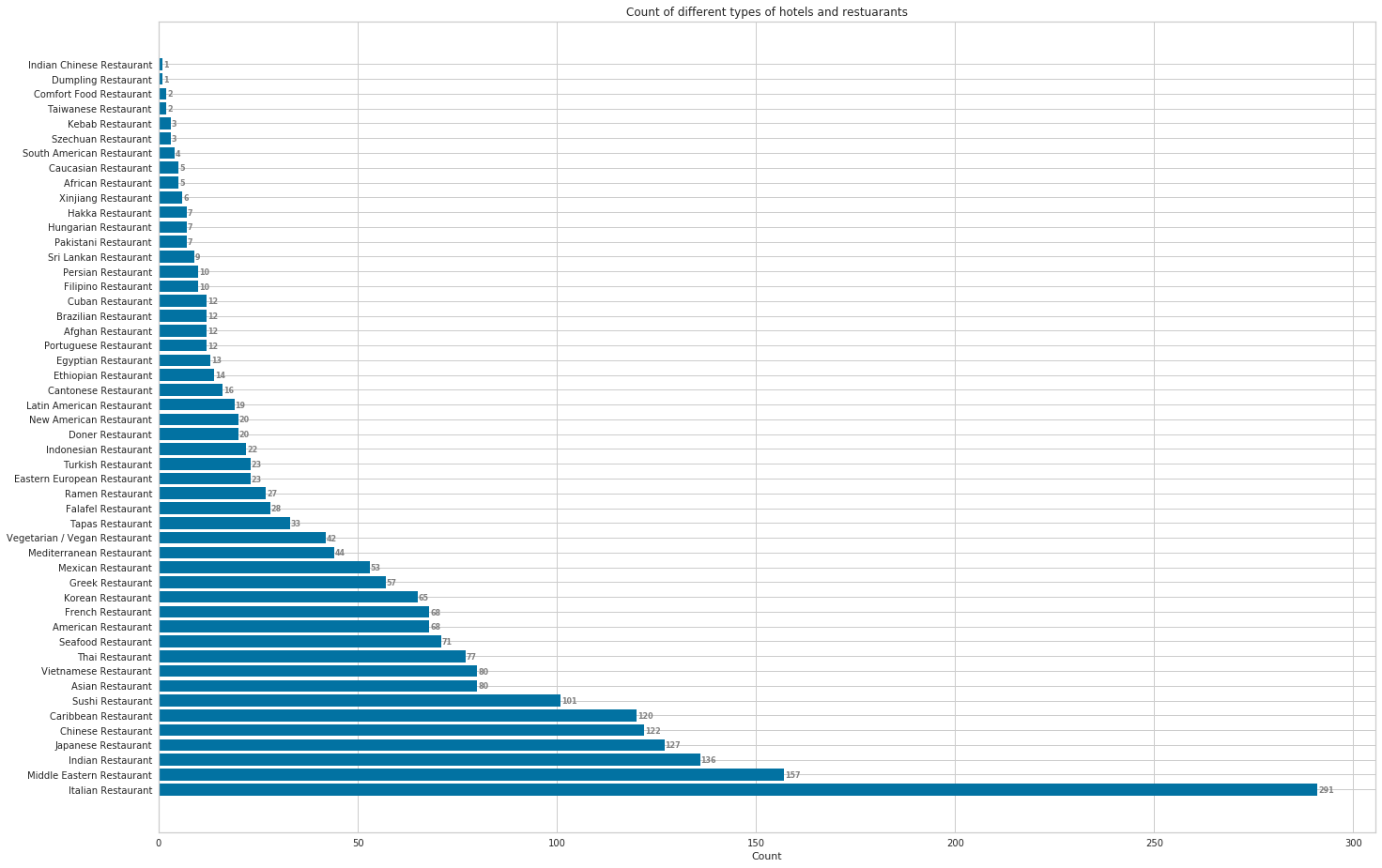


I utilized the foursquare API to get all hotels and restaurants at Toronto. I set the LIMIT to 1500 and the radius as 5000 meters. Sample raw data:

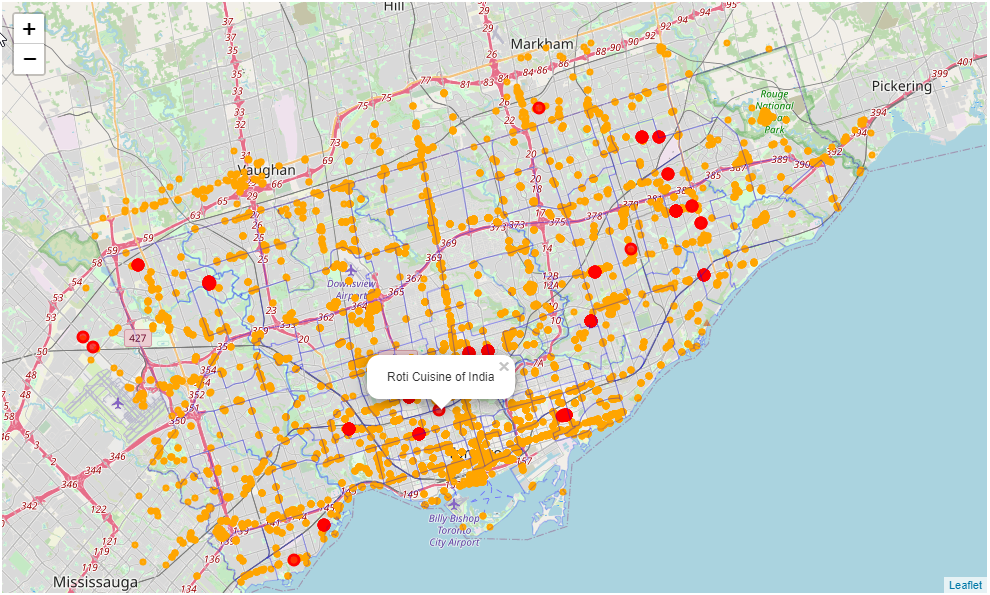


All hotels and restaurants data are wrangled, cleaned and formatted in a structured format as below with required, name, categories, location, id, latitude and longitude details:

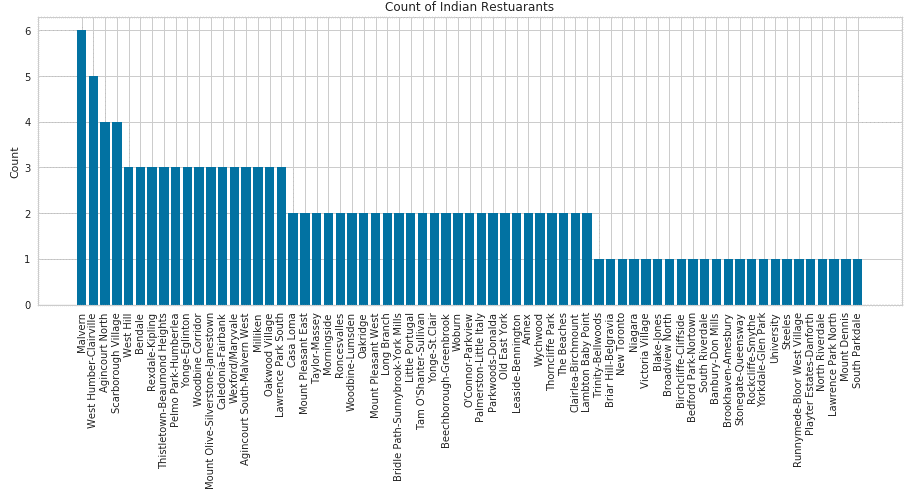


Using Matplotlib, a horizontal bar graph is prepared to plot count of different kind of Restaurants against each category:

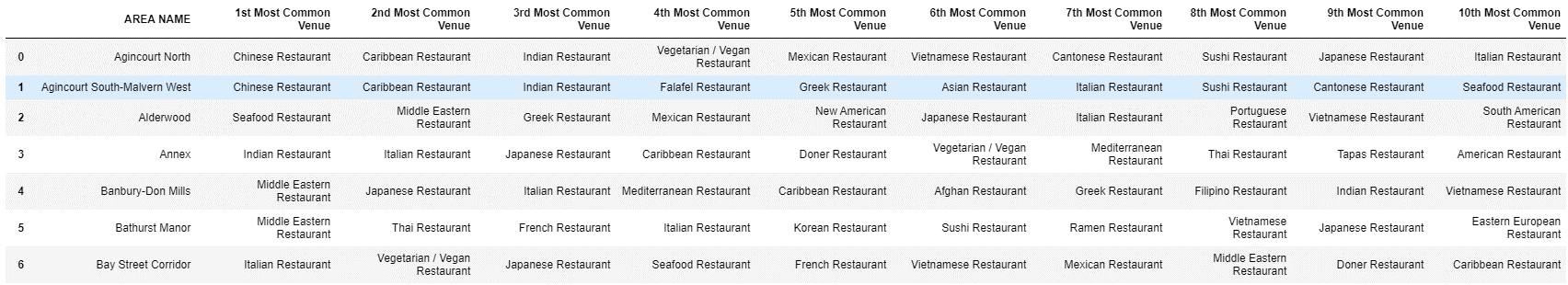
For visualizing, I plotted all venues on the Toronto map and superimposed “Indian Restaurant” on top with red dot.



Count of Indian Restaurant in Areas in different Area is as below:

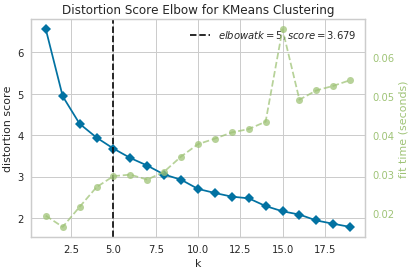


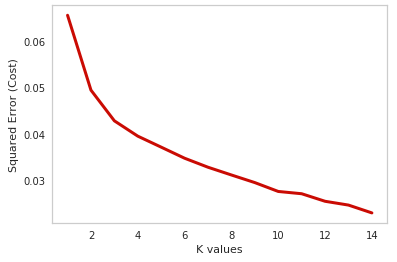
I created a table to list the top 10 venue category for each area as in below:



As K-Means Algorithm is one of the most commonly used cluster mechanisms. I used it to cluster the Area with some common venue.

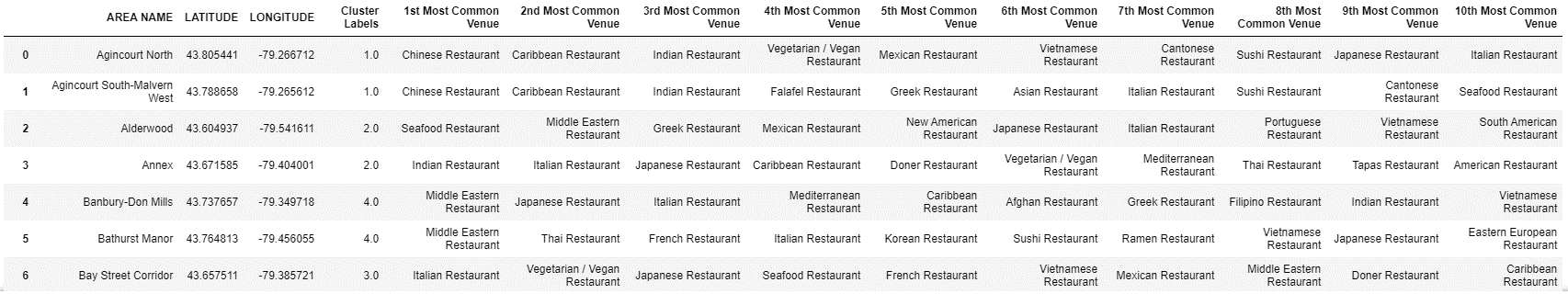
With Elbow method, I determined the that 5 will be the optimum k of the K-Means and thus clustered the Area into 5 clusters.





**D. Results**

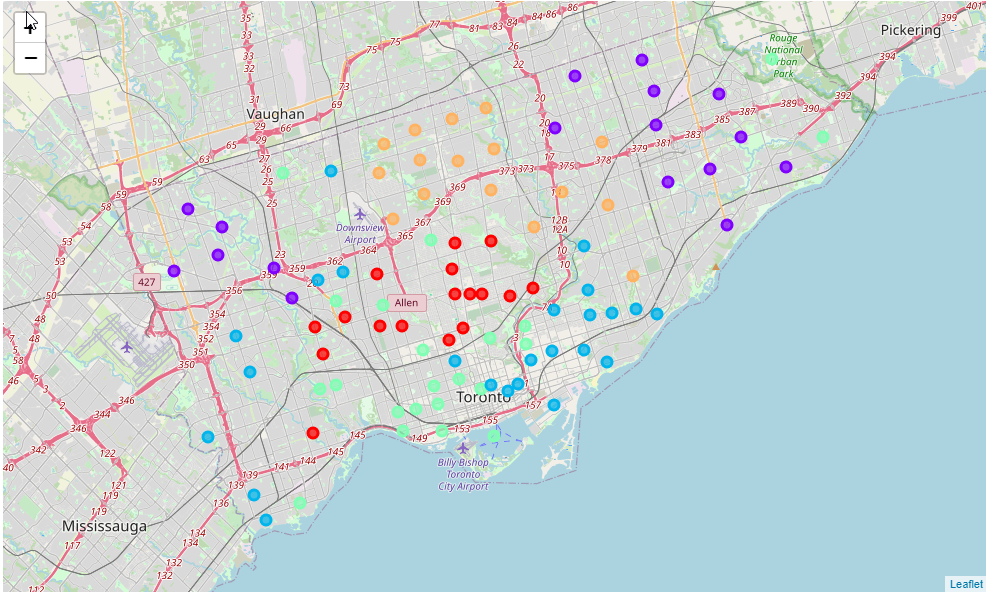
After applying clustering, the data would look like:



I labelled each cluster as follow based on most visited Cuisine in the Area:

* Cluster 0 – Italian/Indian Cuisine
* Cluster 1 – Chinese Cuisine
* Cluster 2 – Indian/Japanese Cuisine
* Cluster 3 – Italian/French Cuisine
* Cluster 4 – Middle Eastern Cuisine

Clusters can be visualized on map as:



**E. Discussion**

With a recorded population of 2,731,571 in 2016, Toronto is the most populous city in Canada and the fourth most populous city in North America. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, aerospace, environmental innovation, food services, and tourism.

I performed data analysis on population variation in different areas and visualized it.

I also mapped the different venues and spotted out Indian Restaurant so that visually it can be analysed.

I used the K-mean algorithm to cluster different areas based on its most visited venues. I have analysed venues with only 50 types of different restaurants with 9769 Restaurant details. The analysis can be made more accurate by expanding types of venues.

**F. Conclusion**

As a result of such diversity in cuisine, people can use this analyse to check for the most densely populated and mark the suitable type of restaurants.