Introduction:

Toronto is the provincial capital of Ontario and the most populous city in Canada. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group,[26] and over 200 distinct ethnic origins are represented among its inhabitants. In 2011, the most commonly reported religion in Toronto was Christianity, adhered to by 54.1 per cent of the population. Other religions significantly practised in the city are Islam (8.2 per cent), Hinduism (5.6 per cent), Judaism (3.8 per cent), Buddhism (2.7 per cent), and Sikhism (0.8 per cent). The cuisine of Toronto reflects Toronto's size and multicultural diversity. Different ethnic neighbourhoods throughout the city focus on specific cuisines, such as authentic Chinese and Vietnamese found in the city's Chinatowns, Korean in Koreatown, Greek on The Danforth, Italian cuisine in Little Italy and Corso Italia, and Indian in Little India.

My client, an immigrant in Canada from India with a cooking expertise. He wants to open a food-chain focusing on Indian cuisine in Toronto. For the starting of the business, he needs some suggestion regarding the neighborhood of Toronto. In this regard, he has two options: being more competitive or being for innovative. If he wants to go competitive then he will want to locate his first restaurant to be in the center where there are a lot of highly rated Indian restaurant. Else he will want to locate his restaurant somewhere nearby where Indian cuisines are not available or not that much highly rated. He needs both kind of suggestion in terms of location.

Data:

We will be using Foursquare API for the geographical data and the venues nearby each neighborhood. And also use user ratings to find where Indian cuisine is highly rated. We can have two approaches in this case, use the rating of the places to grab the most customers in those already popular places or open in a different place where Indian cuisines are not highly rated. And for the latitude and longitude we will be using the file provided in http://cocl.us/Geospatial_data.

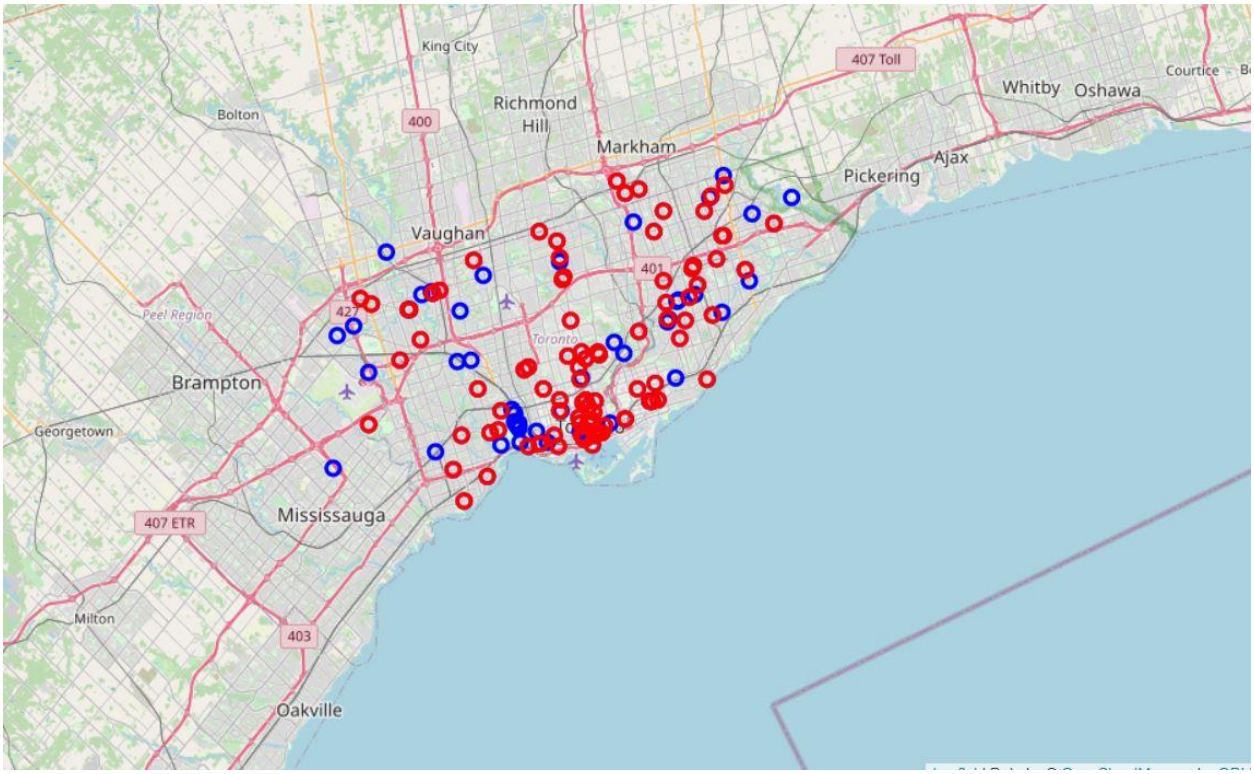
Methodology:

From the data provided in http://cocl.us/Geospatial_data we had the postal codes of each borough and the longitude and latitude of each postal. We then use Foursquare API to get any Indian places nearby that postal code. We want to have all the venues associated with Indian culture as our primary customers will be centered in that area. Some of these places may be restaurants as well. We will be using Foursquare API again for getting any category of those places. A deep inspection into the dataset reveals that not all of those indian restaurants are not mentioned as “restaurants”. There are some inconsistency in the naming of the category of each places from Foursquare API. We performed some exploratory data analysis to have all those keywords regarding all those restaurants. We again used Foursquare API to get the rating of each restaurant. But unfortunately, most of the places were unrated. So, we had to drop that idea.

For the clustering, we used k means algorithm. For choosing the optimum number of clusters an iteration is performed for different number of clusters ranging from two to number of points. A threshold of 3km is taken into account. Whenever the distance between the cluster center and the point is greater than 3 km our algorithm is iterated to the next number. For the distance calculations, we will be using “Haversine Formula”, which determines great circle distance between two points on a sphere given their longitudes and latitudes. We perform the same clustering action on the restaurant data as well. After achieving the optimum number of clusters in both cases, we will get an idea about where our Indian crowd will most likely be.

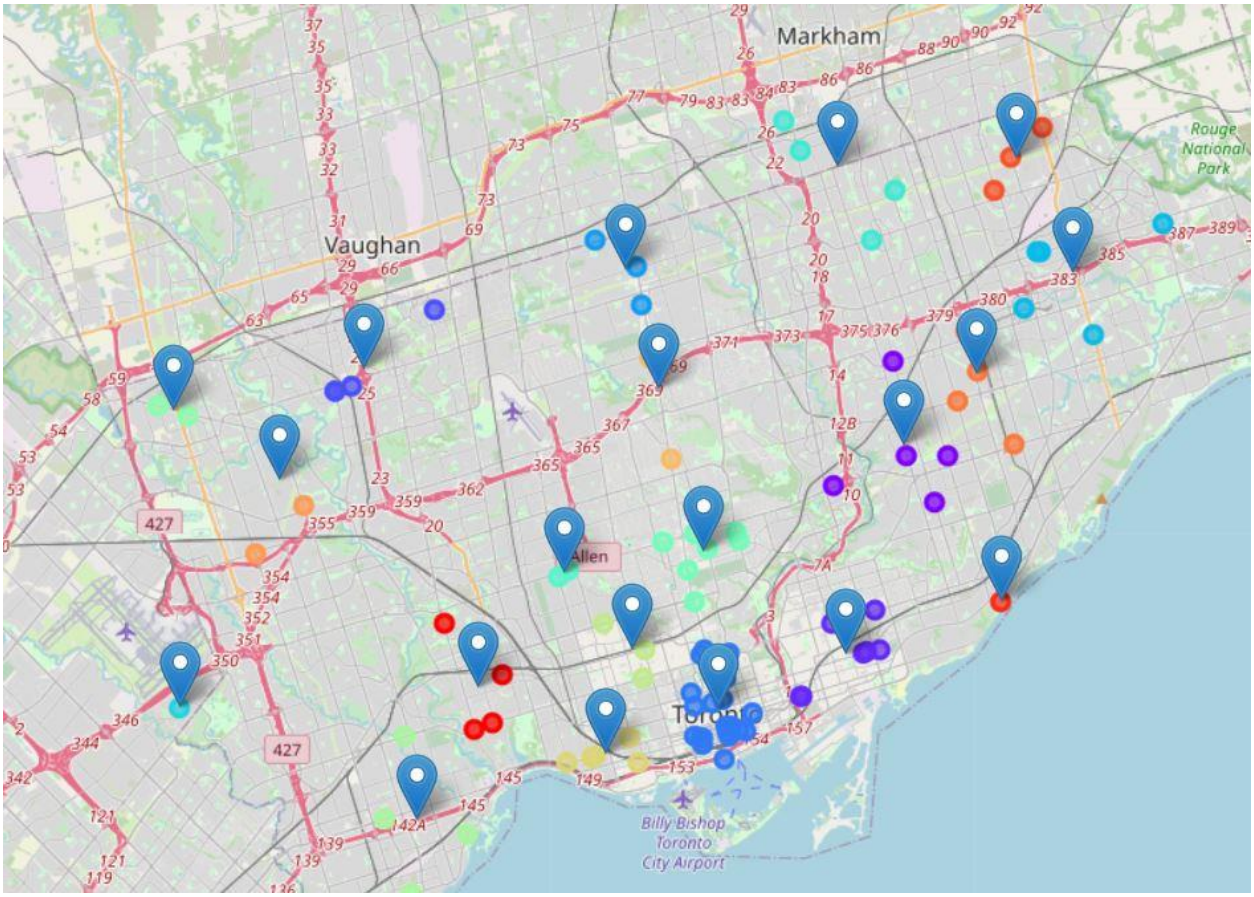
Results:

Here is a view of all the ‘Indian’ places in blue circle and ‘Indian’ restaurants in red circle.



We can see the some highly densed restaurants in the shore area and mostly the places labeled as ‘Indian’ are restaurants. To get a further insights, we will perform the clustering.

After clustering we get 26 clusters for all the Indian places e.g. restaurants, grocery store, religious places etc. And after clustering the restaurants we get 20 clusters of Indian restaurants. By this inspection we can say approximately there are four other places where our restaurant can be set up.



Discussion:

From comparing the clusters we get four location data where a restaurant can be set up. But exploring a new area for a restaurant focusing on a foreign cuisine can be little risky as well. So, the answer comes with the client's preference, whether he will be competitive or innovative. In the first case, he can go for the heavily densed shore area where we already have a tons of Indian restaurants, and there will be less trouble for making crowd. But there will be competition from others. And in the latter case, he will explore a new area where he will have to draw attention of the crowd from the start but there will be no competition from others.

Conclusion:

We can run the same process after we have enough data in the foursquare API with the user rating of those 'Indian' place when we will be able to suggest more accurate and demanding area for the restaurant. If our client wants to upgrade to a multi-cuisine restaurant in the future then this process can be re-run for all the restaurants with same cuisine. By this method, we can offer our client with locations suitable to their taste and maximize the profit.