Coursera Capstone Project

Where to Open a South Asian Restaurant in Toronto and Mississauga

Introduction

Immigration to Canada is a life changing decision. It has far reaching consequences for an individual and his/her immediate family members .The immediate challenge to a new emigree is to find a suitable location to live and work in Canada. With the current economic conditions in the corona impacted world, finding a suitable Job is an uphill task at best. Thus starting a new life in a new world requires an out of the box approach.

For me, food has been a passion all my life. I consider myself a connoisseur of south Asian food. Now, in a tough economic situation with well paying jobs hard to comeby, i will transform my passion into a business.

Problem Description

My aim is to immigrate to Canada and open a South Asian Restaurant in the Toronto/Mississauga area .

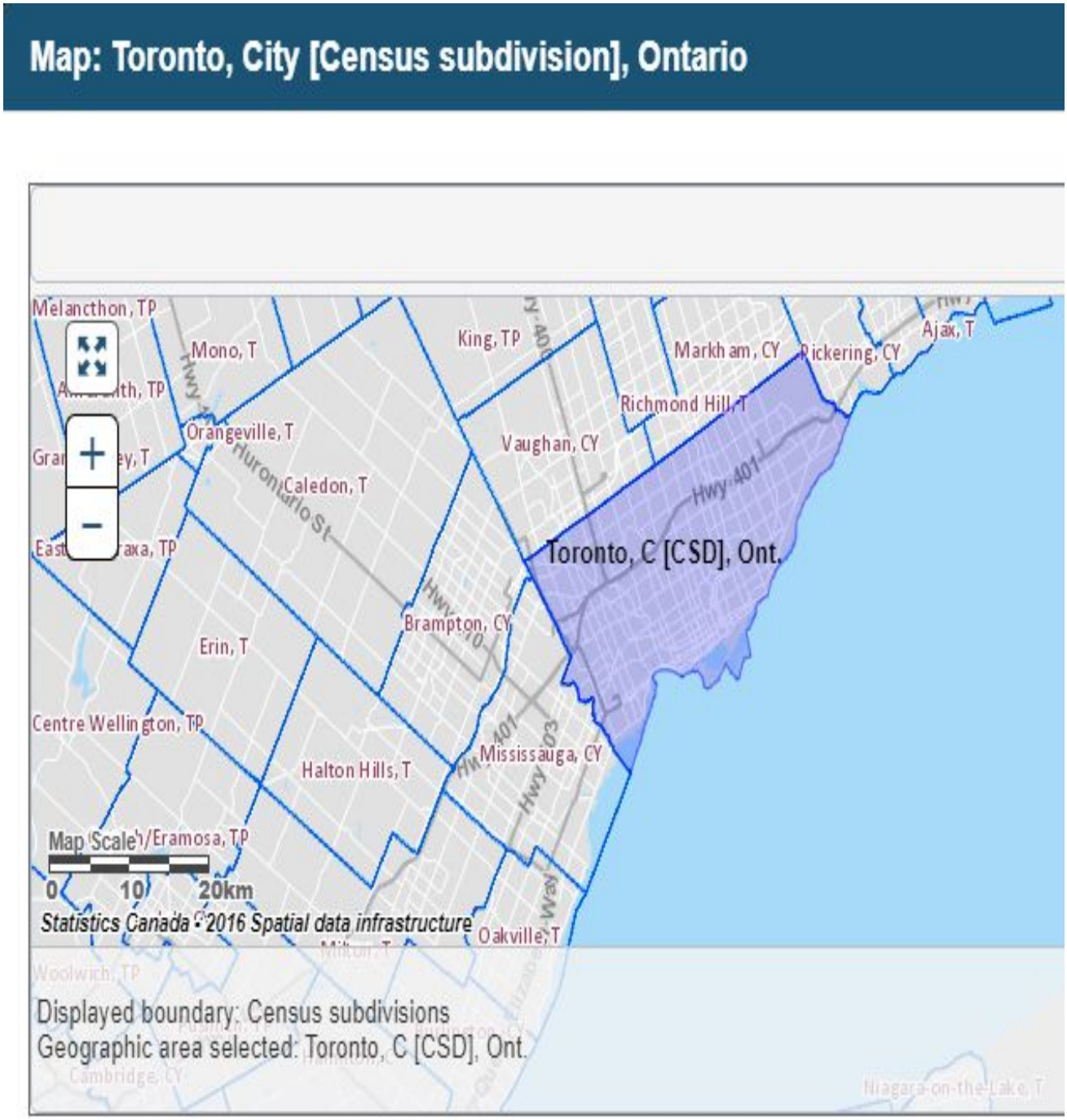
I will utilize the knowledge gained from the courses included in this IBM Certification to do an in depth analysis of my target market using location data provided by Four Square and Demographic Data Provided by Statistics Canada, which is the national statistical office for the Government of Canada.

Due to proximity to Toronto and Mississauga I have combined the neighborhoods of Toronto and Mississauga in my analysis so that I have a holistic view of the Toronto Mississauga restaurant market targeted especially at the South Asian community in these areas.

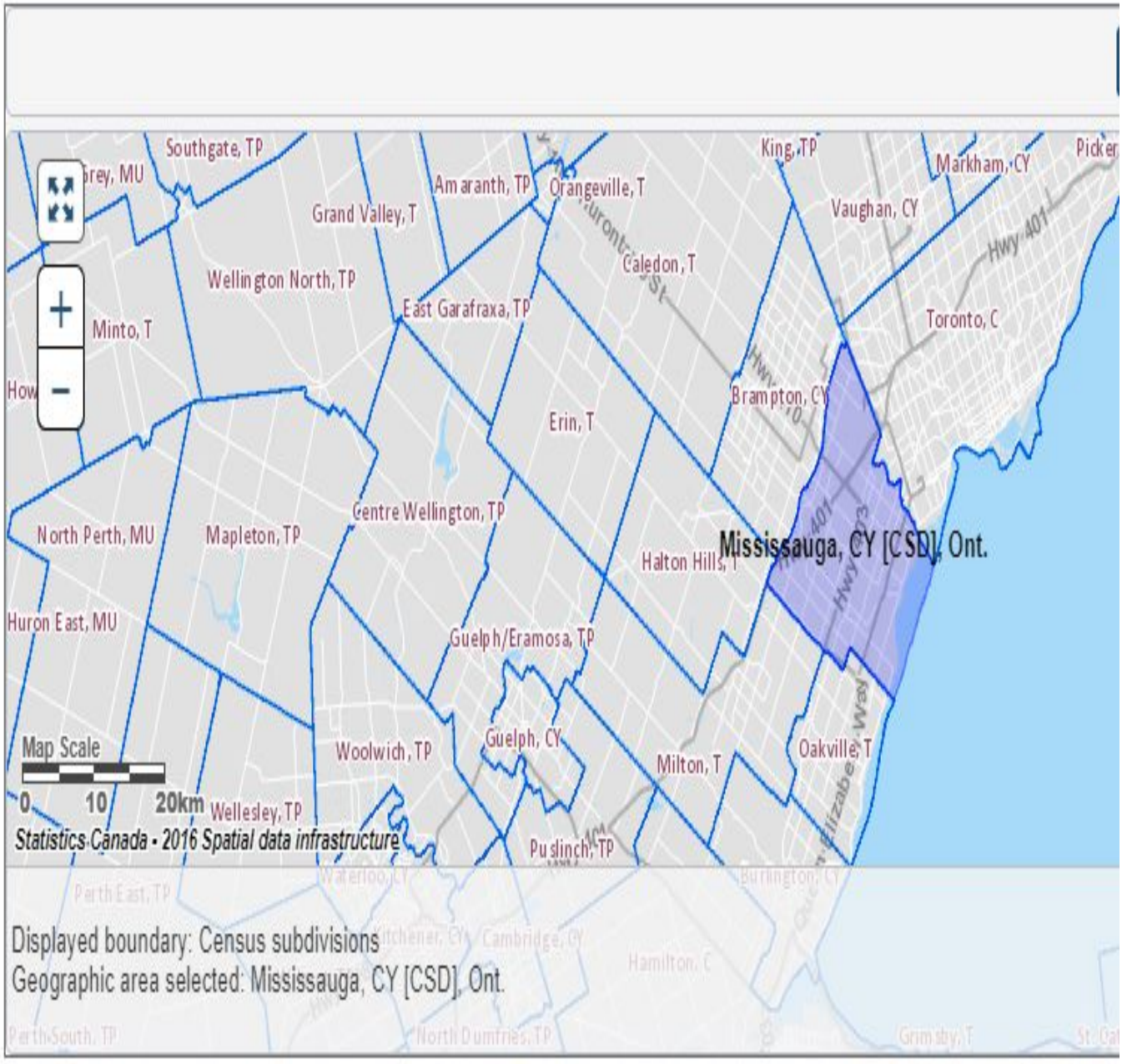
The in depth analysis of the target market using the sources mentioned above using the data analysis skills learned in this specialization will help me answer following key questions before i open my South Asian food restaurant business in Toronto/Mississauga Market

1. Total number of Restaurants Serving the Toronto/Mississauga Market
2. Location of Restaurants that serve the same target Population of the South Asian Community in Toronto/Mississauga that i am targeting.
3. Potential Coverage area of the restaurants that serve the South Asian community that i plan to serve by my restaurant.
4. Location of recreation facilities in the Toronto/Mississauga market like parks, shopping malls , museums and sporting arenas that may attract visitors from my target south Asian community.
5. And Finally Location for my proposed restaurant based on analysis of aforementioned factors using data analysis techniques learned in this specialization.

Background


Map: Toronto, City [Census subdivision], Ontario
Melancthon, TP
Mono, T
King, TP
Markham, CY
Pickering, CY
Ajax, T
Richmond Hill, T
Orangeville, T
Vaughan, CY
Caledon, T
Hwy 401
Toronto, C [CSD], Ont.
Brampton, CY
Erin, T
Centre Wellington, TP
Halton Hills, T
Mississauga, CY
Map Scale
0 10 20km
Statistics Canada - 2016 Spatial data infrastructure
Oakville, T
Displayed boundary: Census subdivisions
Geographic area selected: Toronto, C [CSD], Ont.

Map: Mississauga, City [Census subdivision], Ontario



Demographics of Toronto and Mississauga

Total population for Toronto is 2,731,571 while Mississauga has a population of 7,21,599 as recorded in the Census of 2016.

Our target market is the South Asian Immigrant Population residing in Toronto and Mississauga. The size of this target market can be measured using the Mother Tongue demographics as measured Canadian Census of 2016. This community speaks Indo Aryan languages with an addition of Pushto which is part of North Westren Pakistan.

Total Indo Aryan or Pushto Speakers in Toronto are 138,625 and 4355 respectively. That makes a total target population of 142,980.

Total Indo Aryan or Pushto Speakers in Mississauga are 85,525 and 1165 respectively. That makes a total target population of 86,690.

Based on these statistics , the combined target population of Indo Aryan and Pushto Speakers in Toronto and Mississauga is 229,670.

Sources of Data

Demographic Data Source

The demographics data has been collected from Statistics Canada Which is the national statistical office for Canada. This agency ensures, collects and distributes the key information on Canada's economy, society and environment that they require to function effectively as citizens and decision makers.

https://www.statcan.gc.ca/eng/start

Toronto :

https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3520005&Geo2=PR&Code2=35&SearchText=Toronto&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=3520005&TABID=1&type=0

Mississauga:

https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3521005&Geo2=PR&Code2=35&SearchText=Mississauga&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=3521005&TABID=1&type=0

Location Data Source

The Location Data has been collected from Four Square API using methods explained in the assignments.

https://developer.foursquare.com/docs/api-reference/venues/search/

Description of Data and How it will enable me to Solve the Problem

Description of Data Required for the Project :

1. Name and Location of the Neighborhoods for Toronto
2. Name and Location of the Neighborhoods for MIssissauga
3. Location, Names and Categories of All Venues in a radius of 6 Km for each neighborhood.
4. Location and Names of all Restaurants that can be termed as Competitors serving the same South Asian community in Toronto Mississauga
5. Criteria for a suitable location for a South Asian restaurant
6. Display of Location of Competing Restaurants on Folium Maps to better ascertain suitable location of the proposed restaurant based on criteria already set.
7. Final Proposed Location.

How Data will solve the Problem :

1. Name and Location of Toronto and Mississauga Neighborhoods will be combined into one combined collection.

2. The location of each Neighborhood will serve as the centre of the circle with a radius of 6 km.
3. Using the FourSquare API, I will get the location, names and categories of all the locations in the defined radius.
4. The radius has been kept wide enough to overlap with other neighborhoods so that coverage area of a particular venue is ascertained. However I will take care to only unique venues when doing the competitor analysis.
5. Thus a venue at a certain distance within 6 Km radius of multiple neighborhoods will appear to serve all these neighborhoods but will be considered as one unique location in our competitor analysis.
6. After collecting all the venue data , I will collect the names and location for a particular category of restaurants that will serve the same south asian community which is our target market.
7. After collection of all venues and competitor locations in the target market, I will set the criteria of the location for our restaurant based on information gained for competitors.
8. All venues and competitor locations as set in criteria will be displayed in a Folium based map and reviewed.
9. Finally a location will be chosen based on the set criteria and marked on the map.
10. Once the location is finalized , I can start working on project management aspects of the South Asian restaurant business.

Methodology

Three key concepts form the foundation of methodology that I have used to find out the optimum location for my proposed restaurant.

1. ***Overlapping Boundaries with 6km radius from each neighborhood***

A restaurant located in a particular neighborhood is not restricted to serve customers from other neighborhoods. So coverage area of each competitor restaurant must be calculated using a specific distance yardstick.

For example, for a casual late night dinner at a local burger joint, customers may travel about 5-6 kms from their neighborhood . Similarly office workers may look for a good lunch out of office within their short lunch break at a place near their work location.

Based on these observations, I have set the local coverage area for a restaurant at a radius of 6km from its given location in a particular neighborhood. The FourSquare API provides the details of Venues , location, category and name among other information for neighborhoods in a given radius.

The key point to note is that once data has been extracted form Foursquare API , a single restaurant may appear for multiple neighborhoods . This may seem an issue at first but a close observation reveals that this is in fact what I wanted. The name of Restaurant appearing against each neighborhood is infact its coverage area. However I can easily find the unique number of restaurants using the .unique() method.

2. ***Proximity to a major a location with high footfall***

An important consideration is identifying the locations which have a high foot fall. These locations include but not limited to Parks, Movie theatres , shopping malls etc etc. The identification of these locations is necessary because any proposed location for a restaurant must have a high foot fall venue nearby to attract secondary traffic from customers , who after visiting the primary venue, take the opportunity to visit the nearby restaurant for a meal before heading home. This case is especially applicable to Shopping malls and movie theatres.

3. Identifying the Size of Target Market

The target market has been identified using the Canadian census data [the link to this data has been mentioned in the Source section presented earlier in this report]. The target market for my South Asian restaurant speaks Indo Arayan languages which includes among others Urdu, Hindi, Gujrati, Punjabi languages. In addition Pashto speakers which is not part of the indo Aryan language but is spoken in north westren Pakistan and is considered a part of indo iranian languages has been added to the target market .

These numbers have been extracted separately for Toronto and Mississauga but have been combined in the numbers presented earlier in this report [Section: Demographics of Toronto and Mississauga]

4. Quality of Data

The preprocessing of Data has confirmed The FourSquare API and Canadian Census Data as excellent sources with no missing values nor any duplication.

The data extracted from Four API had 258 unique types of venues and seven types of data extracted for each particular venue. This included

Neighborhood Name ,

Neighborhood Latitude

Neighborhood Longitude,

Venue name

Venue Latitude

Venue Longitude.

5. Who is the Competitor ?

An interesting question arose during the project on how to define a competitor. Should it be restricted to specialty restaurants that specifically mention themselves as Indian or Pakistani restaurants as per classification available in FourSquare or I must consider all the

restaurants that attract south asian communities . Although work was done on multiple restaurants that offer middle eastren, turkish and arabic cuisine , however in the final part , I have displayed only Indian and Pakistani restaurants as the primary competitors for the proposed restaurant.

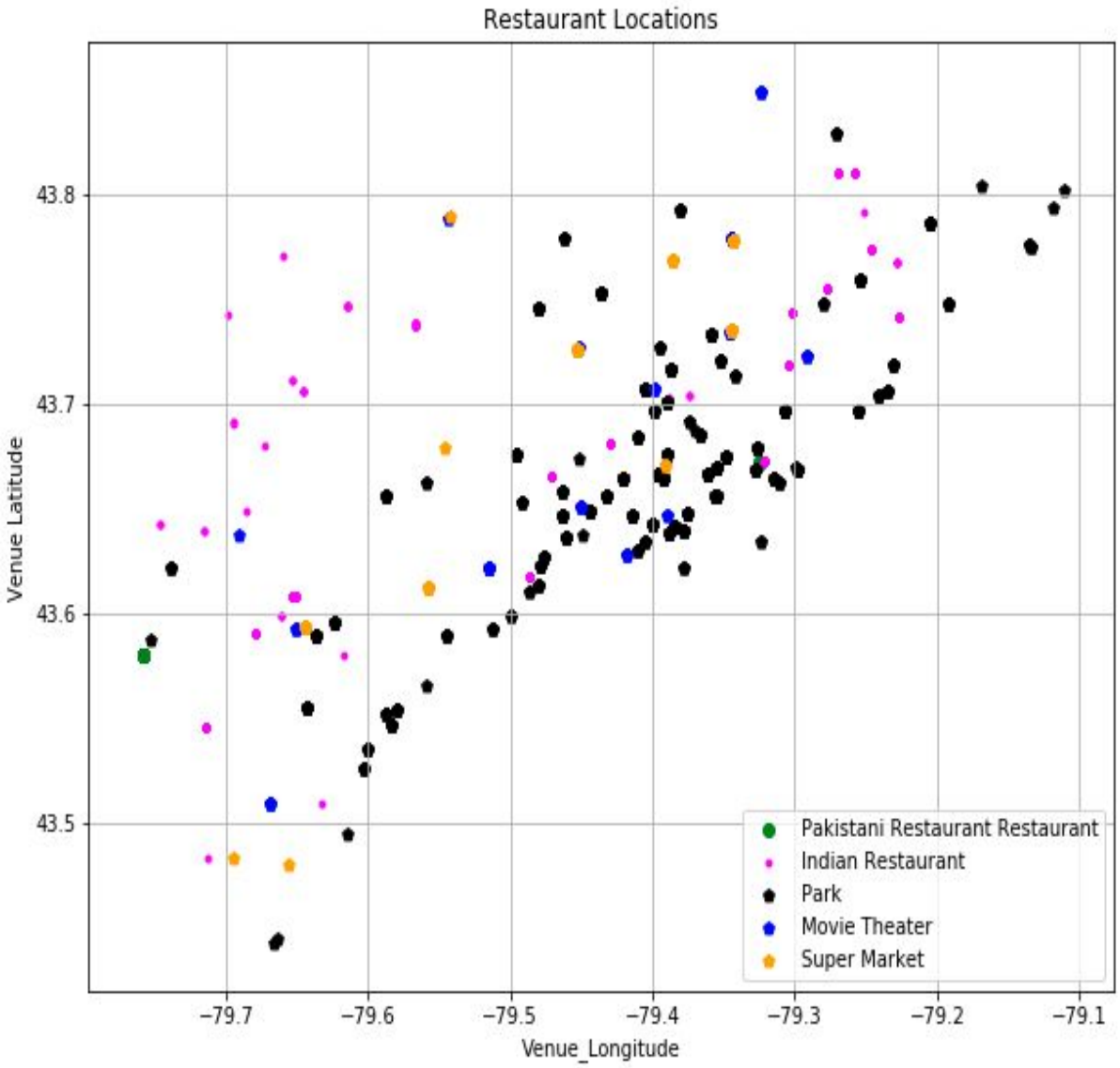
6. **Using K Means Clustering was found to be inadequate**

During the early part of the project for Week 3, I used Kmeans clustering for all venues in central Toronto. It was observed that the Kmeans clustering resulted in a generalized clusters of venues (depending on number of clusters chosen from 2 to 5. I found that the results were more understandable with 3 clusters where results showed that city had three types of venues i.e Parks , Gyms/sporting arena and general Venues like restaurants and movie theatres.

I found that this generalized clustering was inadequate for our project and more specific analysis required using data analysis techniques provided by python.

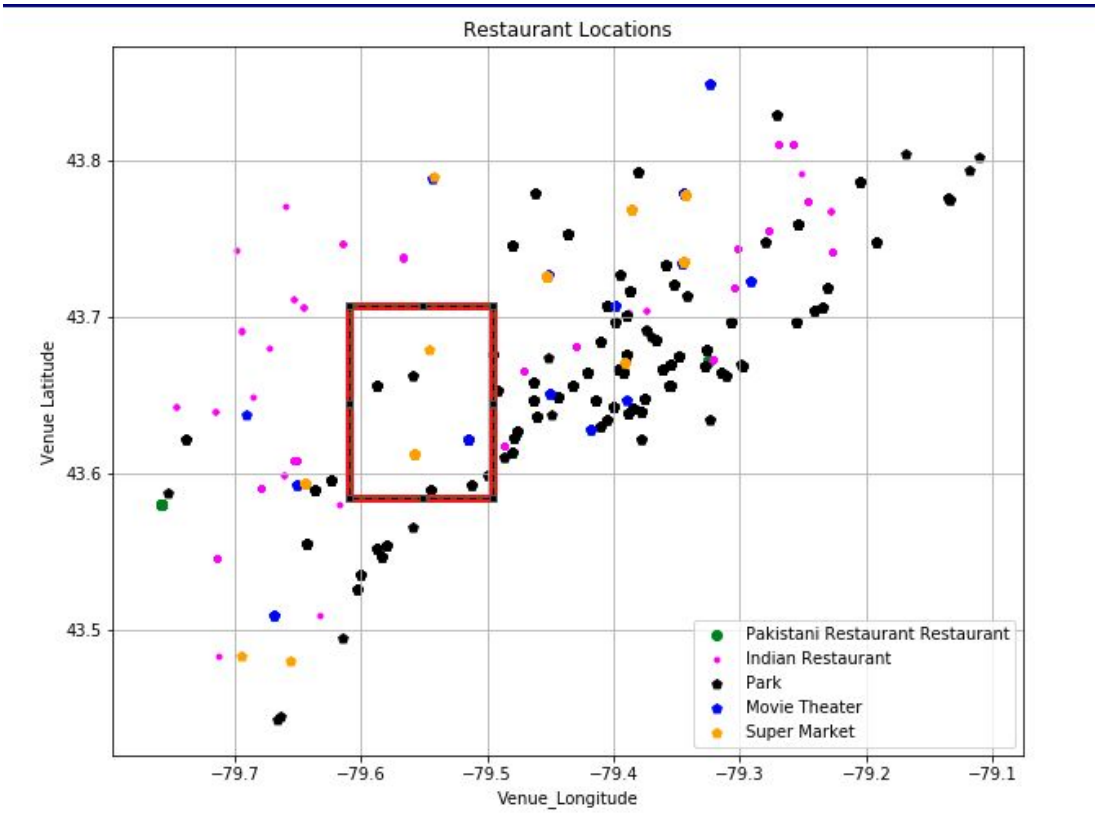
Results

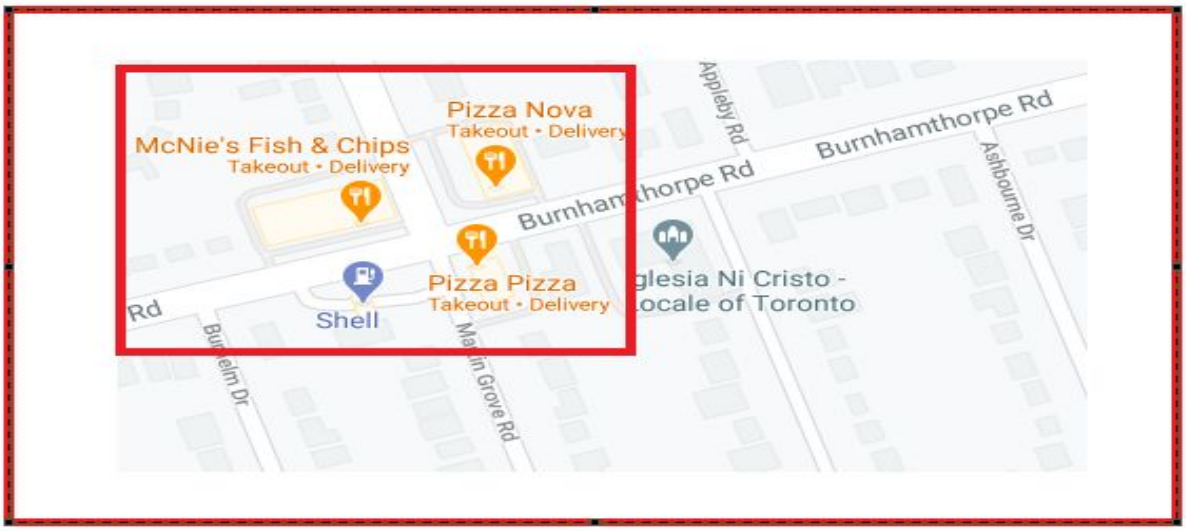
Total Competitors are 39. [37 Indian restaurants and 2 Pakistani Restaurants]



Proposed Location

Proposed location has been selected using the criteria explained above and nearest available commercial space for opening a Business.





Discussion

Following Important points need to be considered

1. Source of Data must be authenticated and well reputed. My Choice of Statistics Canada was a good choice considering data available had no issues like missing values and it was well formatted available in various formats. This saved alot of time that i might have spent in preprocessing of data and fixing issues with the available data .
2. Use of python based Data analysis techniques in combination with Location data providers like Foursquare provide an excellent opportunity to explore the massive data now available free . It has a transformational impact on whatever project you opt to start.
3. The use of Machine learning techniques must be done carefully as they may provide results that may be too generic for your particular project.

Conclusion

Finally , this project was a great learning opportunity for me as i had to utilize my learnings from the complete certification and complete a project that at first glance seemed difficult.

Use of FourSquare with Python Analysis techniques enabled me to analyse the location data of Toronto and Mississauga and helped me make an informed decision on the location of my proposed South Asian restaurant.

Thanks to IBM and Coursera for a such a great learning opportunity.