Introduction: Business Problem

In this project we will try to find an **optimal location for a Bar**. Specifically, this report will be targeted to stakeholders interested in **opening a Bar in Toronto, Canada**.

Canada is a country which allows migrants from other countries, as the number of immigrants are getting permanent residence in the Canada, most of them try to open a new business for the survival. As, Toronto is the city with high population, high facilities immigrants try to find places to open a new business in Toronto but struggles which business can be open at which location of the city. To overcome this, we will analyse the venues at different Borough by taking the example of Bar business.

Since there are lots of Bar in Toronto, we will try to detect locations that are not already crowded with Bar. We are also particularly interested in areas with no Bar is in vicinity. We would also prefer locations as close to city centre as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighbourhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing Bar in the neighbourhood
- number of and distance to Bars in the neighbourhood, if any
- distance of neighbourhood from city centre

We decided to use regularly spaced grid of locations, cantered around city center, to define our neighbourhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of restaurants and their type and location in every neighbourhood will be obtained using **Foursquare API**
- coordinate of Toronto center will be obtained using **Google Maps API geocoding** of well-known Toronto location

we will us the Wikipedia page for extracting the data

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

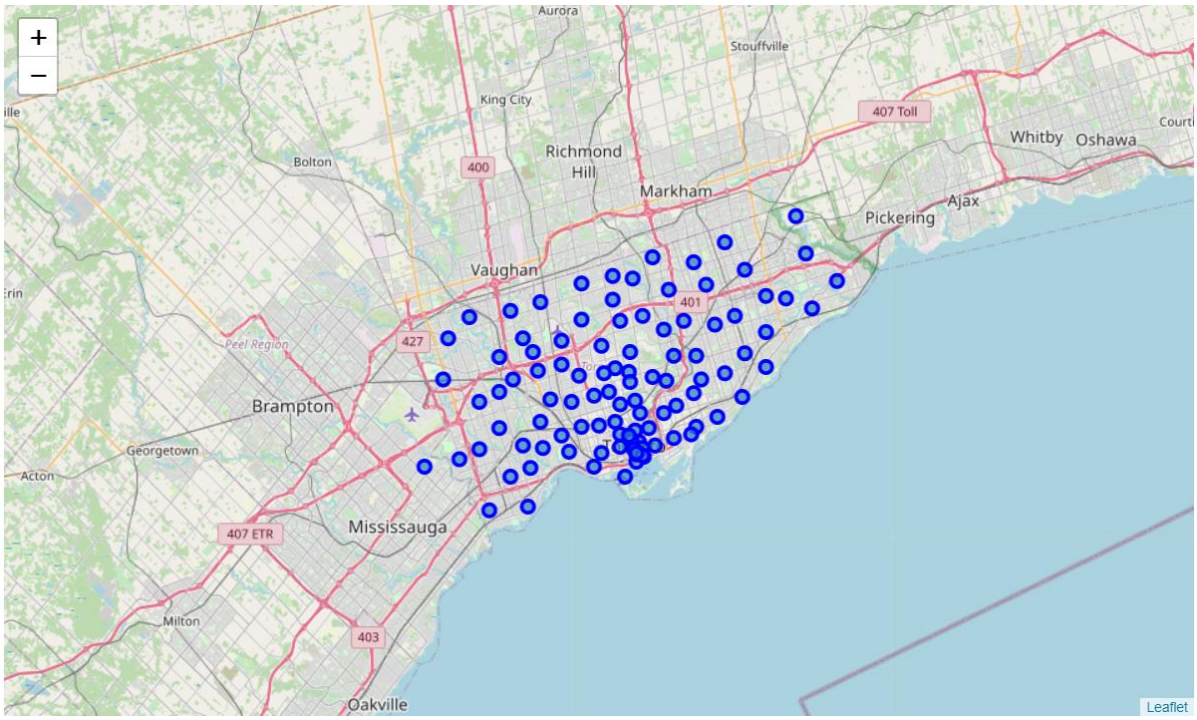
The following shows the dataframe we obtained after wrangling -

	Postalcode	Bourough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...

We obtained the longitude and latitude of all the venues so as to identify the exact location.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	RIGHT WAY TO GOLF	43.785177	-79.161108	Golf Course
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
5	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
6	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
7	Guildwood, Morningside, West Hill	43.763573	-79.188711	Enterprise Rent-A-Car	43.764076	-79.193406	Rental Car Location
8	Guildwood, Morningside, West Hill	43.763573	-79.188711	Woburn Medical Centre	43.766631	-79.192286	Medical Center
9	Guildwood, Morningside, West Hill	43.763573	-79.188711	Lawrence Ave E & Kingston Rd	43.767704	-79.189490	Intersection

The map shows all the neighbourhoods locations as per there longitude and latitudes-



Methodology

In this project we will direct our efforts on detecting the 5 most common venues in the Neighbourhoods along with its frequency.

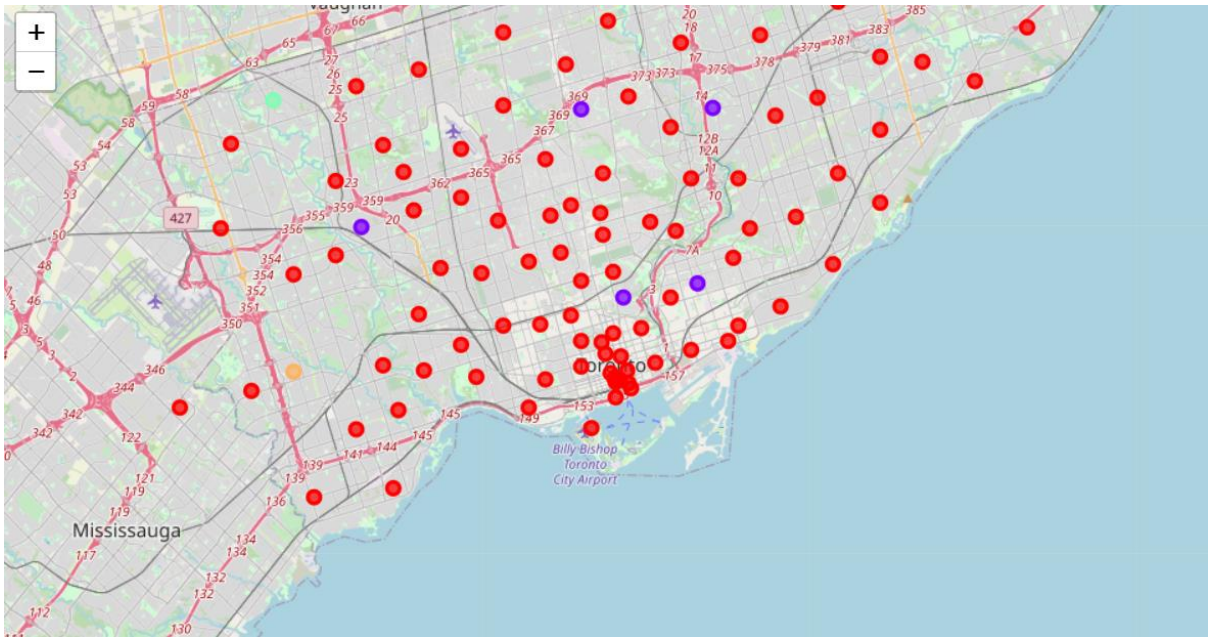
```
----Agincourt----
                       venue  freq
0                      Lounge   0.2
1           Breakfast Spot   0.2
2  Latin American Restaurant   0.2
3               Skating Rink   0.2
4           Clothing Store   0.2


----Alderwood, Long Branch----
            venue  freq
0     Pizza Place   0.2
1        Pharmacy   0.1
2             Gym   0.1
3    Skating Rink   0.1
4  Sandwich Place   0.1
```

Second step in our analysis will be looking at the top 10 most common venues in each Neighbourhood and storing the data in new data frame for further analysis after **clustering. We used KMeans for clustering.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Latin American Restaurant	Skating Rink	Clothing Store	Breakfast Spot	Dog Run	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
1	Alderwood, Long Branch	Pizza Place	Gym	Pharmacy	Sandwich Place	Athletics & Sports	Pool	Pub	Skating Rink	Coffee Shop	Curling Ice
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Pizza Place	Pharmacy	Sushi Restaurant	Middle Eastern Restaurant	Shopping Mall	Deli / Bodega	Restaurant	Fried Chicken Joint
3	Bayview Village	Café	Bank	Chinese Restaurant	Japanese Restaurant	Yoga Studio	Department Store	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
4	Bedford Park, Lawrence Manor East	Coffee Shop	Restaurant	Sandwich Place	Italian Restaurant	Sushi Restaurant	Comfort Food Restaurant	Pharmacy	Pizza Place	Café	Pub

Following map shows clustering with K=5



In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with less no of Bars. We will present map of all such locations but also create clusters (using k-means clustering) of those locations. We will observe and find out the locations where we can set-up Bars for better Business.

Below is the example of how we examined each cluster for the different venues.

Cluster 3 - No Bars

```
In [41]: df_c3=toronto_merged.loc[toronto_merged['Cluster_Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.sha
         df_c3
```

Out[41]:

	Bourough	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Scarborough	2	Fast Food Restaurant	Deli / Bodega	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Doner Restaurant	Dog Run

Cluster 2 - Bars are present only in North York

In [40]:
```
df_c2=toronto_merged.loc[toronto_merged['Cluster_Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.sha
df_c2
```

Out[40]:

	Bourough	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	Scarborough	1	Park	Playground	Distribution Center	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
23	North York	1	Park	Bar	Bank	Convenience Store	Eastern European Restaurant	Drugstore	Donut Shop	Doner Restaurant	Dog Run	Deli / Bodega
25	North York	1	Park	Convenience Store	Food & Drink Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Deli / Bodega
40	East York	1	Park	Convenience Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Yoga Studio	Deli / Bodega
50	Downtown Toronto	1	Park	Trail	Playground	Dance Studio	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
98	York	1	Park	Convenience Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Yoga Studio	Deli / Bodega

Results

The following are the outcomes of analysis of 5 clusters-

- In cluster 1 - Bars are present in Scarborough and Etobicoke Borough.
- In cluster 2 - Bars are present in only North York.
- In cluster 3, 4 and 5 - Bars are not present.

Discussion

Hence, for people who are interested in opening the Bar business can easily opt for any on cluster 3 or 4 or 5.Ba

In case the stakeholders wants the business to be in cluster 1 or 2, they can choose Borough other than Scarborough and Etobicoke for cluster-1 and borough except North York in cluster 2.

Conclusion

Purpose of this project was to identify Toronto areas close to center with low number of Bars in order to aid stakeholders in narrowing down the search for optimal location for a business. By calculating Bar density(frequency) distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby bars. Clustering of those locations was then performed in order to create major zones of interest and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal Bar location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.