Predicting the suitable location for Gambling Centre

Mahesh Warghane

1. Introduction

1.1 Background

Gambling is the practice or act of playing games of chance for a stake. In most cases, the stake is money. However, if the gambler has run out of money, the stake could include any possession. Gamble means the same as betting or wagering. We call somebody who bets a gambler. Gamblers bet on something that results either in a gain for them, or a loss.

Toronto is one of the most populous cities in North America. And it is home to many migrants from all over the world. Immigration from past few decades has changed the demography of Toronto completely. Now there are many restaurants and shopping complexes all over the city. Also due to this Toronto also gets an excellent Tourist attention every year independent of the quarters of the year. The city is a major Centre for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the Canada.

All these factors contribute towards making the market of Toronto highly competitive. So any person or any organization which is looking to invest here needs to analyze the locality of Toronto completely. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

1.2 Problem

Data that might contribute to determining a suitable and profitable location for a client. Gathering the complete information of localities of Toronto city. This project aims to predict the most suitable location for a Gambling Centre in Toronto, Canada.

1.3 Interest

This data can play a vital role for those organizations or investors who are looking for opening their ventures in the locality of Toronto city. Since this data will be used to identify the top localities of Toronto so peoples from different domains can get data which is useful for them e.g. cafés, shops, etc.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

We get the data of neighbourhoods of Toronto from here. Since data was not readily available so we have to do lots of work on that. Now, to get the required data we have used the Web-Scrapping technique. This technique is useful for those datasets which are meant to be acquired from web. Some insights of the data acquired from the web.

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0. However, the postal code M0R 8T0 is assigned to an Amazo
volume addresses.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills

Figure 1 data before Web-Scrapping

2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values from earlier seasons, because of lack of record keeping. For this project is concerned I'll only use the data which will give the information about the neighborhoods of Toronto city. Also I have excluded those boroughs which had no values or more precisely "Not Assigned". Take a look at this dataset which was obtained after this process.

```
In [5]: df=df[df['Borough']!='Not assigned']

In [10]: df.head(11)
Out[10]:
```

	Postalcode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
8	M9A	Etobicoke	Islington Avenue, Humber Valley Village
9	M1B	Scarborough	Malvern, Rouge
11	M3B	North York	Don Mills
12	M4B	East York	Parkview Hill, Woodbine Gardens
13	M5B	Downtown Toronto	Garden District, Ryerson
14	M6B	North York	Glencairn

Figure 2 data after Web-Scrapping

After eliminating the null boroughs my next job was to merge the Neighbourhoods of the similar postal codes. E.g. from if two rows have the same postal code and two different Neighbourhoods, so we need to merge them. Also the indexing of the table not quite good, so we also need to rest the index of the table.

```
In [17]: df_merge.head()
```

Out[17]:

	Postalcode	Borough	Neighborhood
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

```
In [18]: df_merge.shape

Out[18]: (103, 3)
```

Figure 3 After merging neighbourhood of same Postal Code

Now there's a possibility that the above information on the web may be false. So we cannot blindly trust that. Hence to avoid this confusion we'll use this .csv file. After this we'll add two columns i.e. latitude and longitude accompanied with the postal code.

Out[21]:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Figure 4 table with latitude and longitude

After this we will make a single data frame consisting of a Postal code, boroughs, neighbourhoods, latitude and longitude. This data frame shows us all the postal codes of Canada starting with 'M'. And assigning latitude and longitude to each and every postal code.

Out[24]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure 5 Postal codes of Canada with location co-ordinates

Now we are ready to get our required dataset which is the Postal codes of Toronto city. Let's take a look at it…

Out[25]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Figure 6 final data set

3. Methodology

Here we will briefly discuss about the methodology. Since the required data set was not readily available to us so we have to do it lots of data cleaning. We obtained our data set from the web and extracted that data set through the process of web-scrapping.

After getting acquainted by the desired data set our next job is to use that data set and obtain the required information about the localities of Toronto city. And for this purpose, we required an API, there are many APIs who offers this service but in our project we have used the *FOURSQUARE* API.

To use the *FOURSQUARE* API in our project we have to define some credentials such as Foursquare ID, Foursquare Secret and Foursquare API version.

After this we need to create the API request URL and also we will use a variable to get the response of the request. Also, we need to make sure that we should only get the relevant information about the venues. Now, after doing all these processes we got the names of the venues near by the Toronto City.

We needed this information so as to know which are the venues near by the city that are often visited by the citizens. We obtain the list of top 100 venues of the city. We use this method because our job is to locate the place for our client where they can open their business. So much the area is popular and visited by public there are more chances of getting profit out of that.

```
In [52]: toronto_venues = getNearbyVenues(names=toronto_data['Neighborhood'],
                                   latitudes=toronto_data['Latitude'],
                                   longitudes=toronto_data['Longitude']
                                  )
```

```
The Beaches
The Danforth West, Riverdale
India Bazaar, The Beaches West
Studio District
Lawrence Park
Davisville North
North Toronto West,  Lawrence Park
Davisville
Moore Park, Summerhill East
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
Rosedale
St. James Town, Cabbagetown
Church and Wellesley
Regent Park, Harbourfront
Garden District, Ryerson
St. James Town
Berczy Park
Central Bay Street
Richmond, Adelaide, King
Harbourfront East, Union Station, Toronto Islands
Toronto Dominion Centre, Design Exchange
Commerce Court, Victoria Hotel
Roselawn
Forest Hill North & West, Forest Hill Road Park
The Annex, North Midtown, Yorkville
University of Toronto, Harbord
Kensington Market, Chinatown, Grange Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport
Stn A PO Boxes
First Canadian Place, Underground city
Christie
Dufferin, Dovercourt Village
Little Portugal, Trinity
```

Figure 7 NearbyVenues around Toronto City

To make it more easy to understand we have converted the above list to a data frame. Because it is always a good approach to convert the tables or lists into data frames.

```
In [53]: toronto_venues.head()
```

Out[53]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop

Figure 8 DataFrame consisting of Venues

After this we have grouped the neighbourhoods with respect to the venues. We are doing this because it will help us in the analysing section.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	58	58	58	58	58	58
Brockton, Parkdale Village, Exhibition Place	23	23	23	23	23	23
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	15	15	15	15	15	15
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	17	17	17	17	17	17
Central Bay Street	61	61	61	61	61	61
Christie	16	16	16	16	16	16
Church and Wellesley	76	76	76	76	76	76
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	34	34	34	34	34	34
Davisville North	9	9	9	9	9	9
Dufferin, Dovercourt Village	16	16	16	16	16	16
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West, Forest Hill Road Park	5	5	5	5	5	5

Figure 9 Grouping neighborhoods by venue counts

4. Analysing Section

After applying all the methodology we obtain the dataset for all the neighbourhoods of Toronto City. Now we will analyse each neighbourhood. Since the neighbourhood column has object as a datatype. So we'll use one-hot encoding technique to convert each neighbourhood to float type.

Out[56]:

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store
0	Berczy Park	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.017241	0.0
1	Brockton, Parkdale Village, Exhibition Place	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0
2	Business reply mail Processing Centre, South C...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0
3	CN Tower, King and Spadina, Railway Lands, Har...	0.0	0.058824	0.058824	0.058824	0.117647	0.117647	0.117647	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0
4	Central Bay Street	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.016393	0.0

Figure 10 One-hot encoding

Now further we have analysed the top 5 locations of each neighbourhood. Getting complete information about all these places makes our idea much clear for predicting the right location for our client.

```
----Berczy Park----
               venue  freq
0        Coffee Shop  0.10
1         Restaurant  0.03
2        Cocktail Bar 0.03
3  Seafood Restaurant 0.03
4      Farmers Market 0.03

----Central Bay Street----
                venue  freq
0         Coffee Shop  0.18
1                Café  0.05
2  Italian Restaurant  0.05
3      Sandwich Place  0.05
4      Bubble Tea Shop 0.03

----Brockton, Parkdale Village, Exhibition Place----
               venue  freq
0               Café  0.13
1        Coffee Shop  0.09
2     Breakfast Spot  0.09
3                Gym  0.04
4 Italian Restaurant  0.04

----Christie----
          venue  freq
0 Grocery Store  0.25
1          Café  0.19
2          Park  0.12
3    Baby Store  0.06
4   Coffee Shop  0.06
```

Figure 11 Top_5 venues of some neighbourhoods

From the above figure we can easily see top 5 venues of some of the few neighbourhoods. Similarly we can do this for other neighbourhoods too. After this we have put all these into a data frame.

Out[65]:

	Neighborhood	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	Berczy Park	Coffee Shop	Café	Cheese Shop	Farmers Market	Bakery	Seafood Restaurant	Restaurant	Beer Bar	Cocktail Bar	Lounge
1	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Coffee Shop	Grocery Store	Bakery	Stadium	Burrito Place	Restaurant	Climbing Gym	Pet Store
2	Business reply mail Processing Centre, South C...	Light Rail Station	Skate Park	Garden Center	Brewery	Auto Workshop	Fast Food Restaurant	Farmers Market	Burrito Place	Restaurant	Garden
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Airport Terminal	Boutique	Harbor / Marina	Boat or Ferry	Rental Car Location	Bar	Plane	Coffee Shop
4	Central Bay Street	Coffee Shop	Café	Sandwich Place	Italian Restaurant	Department Store	Salad Place	Japanese Restaurant	Bubble Tea Shop	Burger Joint	Juice Bar

Figure 12 DataFrame consisting top venues of different neighbourhoods

5. Discussion and Observation

After analysing is done we discuss, what we got from the data set. It is always prefer to view the observations rather discussing it. So we have used k-means clustering approach.

K-means clustering is a good technique as it can be used to make different clusters on the map with different colours. To make it clearer and more precise we have create a new data frame that includes the cluster as well as the top 10 venues for each neighborhood.

	Postalcode	Borough	Neighborhood	Latitude	Longitude	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
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Yoga Studio	Department Store	Electronics Store	Eastern European Restaurant
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	2	Greek Restaurant	Italian Restaurant	Coffee Shop	Restaurant	Furniture / Home Store	Ice Cream Shop	Pub
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	1	Park	Sandwich Place	Pizza Place	Pub	Liquor Store	Burrito Place	Italian Restaurant
43	M4M	East Toronto	Studio District	43.659526	-79.340923	2	Café	Coffee Shop	Brewery	Gastropub	Bakery	American Restaurant	Yoga Studio
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	1	Park	Bus Line	Swim School	Dim Sum Restaurant	Falafel Restaurant	Ethiopian Restaurant	Electronics Store

Figure 13 DataFrame consisting of clusters of various neighbourhoods

Clusters on the map can be seen as :-

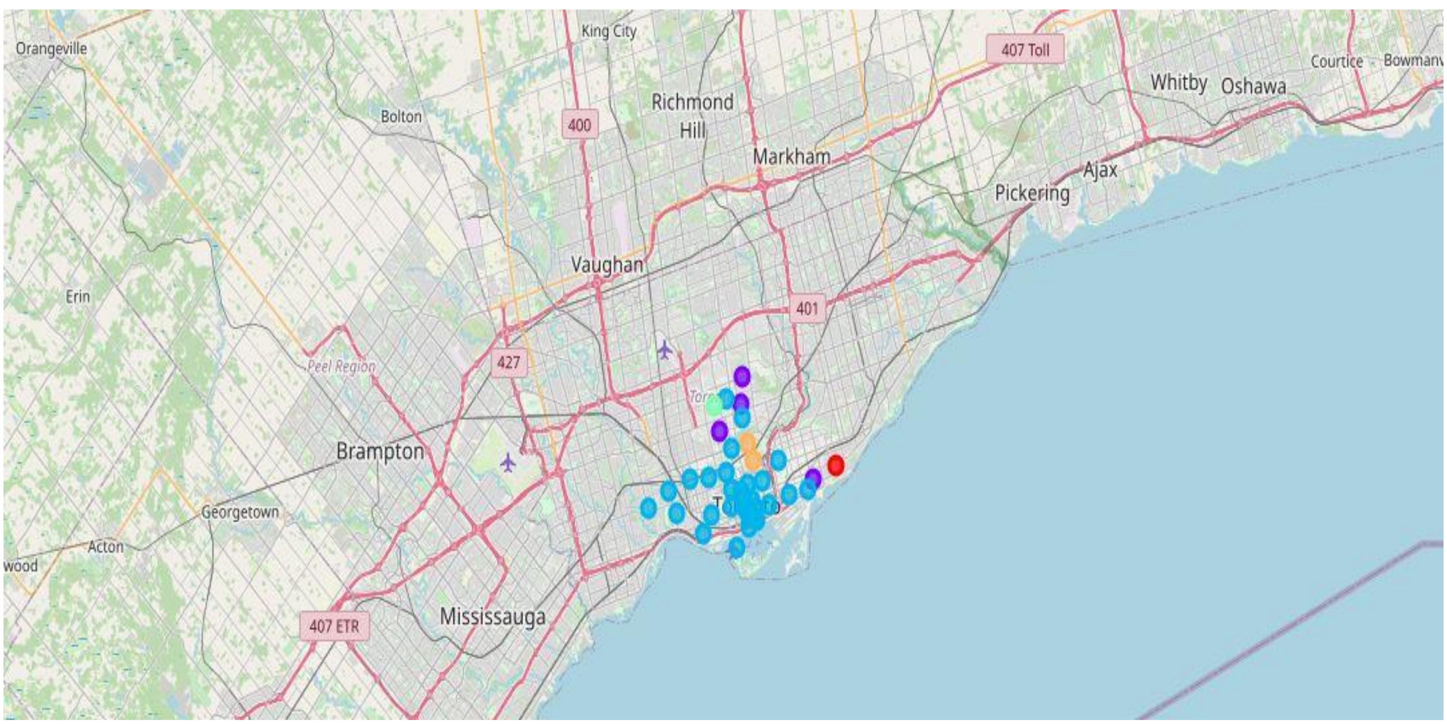

Figure 14 Clusters showing neighbourhoods of Toronto

As we can observe from the above map that there are places which are mostly populated all the time. So there's a great chance of getting customers for our client's Gambling Centre. The denser the cluster the higher the probability of a good locality for our client's project.

6. Result Section

After clustering on map is done we have short-listed the localities having denser clusters since probability of flourishing is much higher in these localities for our client. We have also looked at the competitors in the locality so as to make it much more clear for our client. Since out of the data we noticed that there are not much competitors for our client but there are some vendors who offer trading and betting as a service. So, based on our results and observations we were able to predict the location for a Gambling Centre of our client.

7. Conclusion

In this project we have short-listed the top localities of Toronto city, now this data can also play a vital role for vendors who wants to open their café in other parts of Toronto too. We have done a lot of work for getting the accurate data for the neighbourhoods of Toronto. Also, to get the top venues of the city we used the Foursquare API which has played a vital role in clustering the map.