

Capstone Project - The Battle of Neighborhoods (Week 1)

July 9, 2019

1 Battle of Neighborhoods (Week 1)

1.1 1. Introduction Section

1.1.1 Description of the Problem

A restaurant is a business which prepares and serves food and drink to customers. It is no surprise to learn that owning a restaurant comes with its fair share of responsibilities that require you to be on your feet. To be successful in the long run, a restaurant needs an accessible location, a theme or style that appeals to a broad range of customers, and a solid menu. Restaurants must stand out from the crowd and finalizing a space within the budget.

So it is evident that to survive in such competitive market it is very important to strategically plan. Often there's an issue of time constraint that keeps a restaurant owner preoccupied with intricacies. Whether it is sourcing the right ingredients, obtaining too many licences, finalizing a space within your budget to interviewing the right staff — restaurateurs have to consider everything beginning from a pin to piano before taking the right foot forward.

First move is very important, thereby choice of location is very important. So, HSB India (P) Ltd., plan to choose the correct location to start its first restaurant in Toronto, Canada.

1.1.2 Discussion of the Background - Prospects of a Indian Restaurant in Toronto, Canada.

Toronto, Canada's largest city is consistently rated among the world's top most liveable cities. Toronto boasts a multicultural population, diverse neighbourhoods, and world class cultural attractions, entertainment, dining and shopping.

Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. Its varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities, attract over 25 million tourists each year.

People have travelled through and inhabited the Toronto area and the diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. According to the 2011 National Household Survey (NHS), 1,264,395 non-Whites, or 20.2% of Canada's visible minority population, live in the city of Toronto; of this, approximately 70 per cent are of Asian ancestry.

These indicates that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. 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.

Target Audience

1. Business personnel who wants to invest or open a ethnic restaurant. This analysis will be a comprehensive guide to start or expand restaurants targeting the ethnic diversity in Toronto.
2. Any Freelancer who wants to start a new restaurant in Toronto.
3. New graduates, to find reasonable lunch/breakfast place close to office.
4. Budding Data Scientists, who wants to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it.

Success Criteria: The success criteria of the project will be a good recommendation of Borough/Neighborhood choice to HSB India (P) Ltd., based on lack of such restaurants in that location and nearest suppliers of ingredients.

1.2 2. Data Section

Neighborhood has a total of 11 boroughs and 103 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 11 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Postal codes beginning with M are located within the City of Toronto in the province of Ontario. I will use the BeautifulSoup website scraping library to extract the content on the Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) as XML and transform data that is in the table of postal codes into a pandas dataframe.

I will read the csv file that has the geographical coordinates of each postal code from http://cocl.us/Geospatial_data into a pandas dataframe. Perform a merge the Latitude and Longitude from the Geospatial dataframe with the Toronto neighbourhoods using inner join.

Next will slice the dataframe to have only the boroughs that contain the word Toronto. I will use the Foursquare API to explore neighborhoods in Toronto City. Toronto City geographical coordinates data used as input for the Foursquare API, to explore venues information in each neighborhood.

1.2.1 2.1 Preparation for Data

First, let's download all the dependencies that we will need.

```
[2]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#!conda install -c conda-forge geopy --yes # uncomment this line to install geopy library
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
```

```

from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

# import BeautifulSoup for webscraping
from bs4 import BeautifulSoup

#!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line to install Folium library
import folium # map rendering library

print('Libraries imported.')

```

Libraries imported.

Download and Scrape the Data I used the BeautifulSoup website scraping library to extract the content on the Wikipedia page as XML.

```

[3]: url = requests.get('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M').text
LoP_Canada = BeautifulSoup(url,'xml')

```

Next, obtain the data that is in the table of postal codes and transform the data into a pandas dataframe.

```

[4]: table_LoP = LoP_Canada.find('table')
LoP_values = table_LoP.find_all('td')

postcode = []
borough = []
neighbourhood = []

for i in range(0, len(LoP_values), 3):
    postcode.append(LoP_values[i].text.strip())
    borough.append(LoP_values[i+1].text.strip())
    neighbourhood.append(LoP_values[i+2].text.strip())

df_LoP = pd.DataFrame(data=[postcode, borough, neighbourhood]).transpose()

# dataframe consist of three columns: PostalCode, Borough, and Neighborhood
df_LoP.columns = ['PostalCode', 'Borough', 'Neighbourhood']

```

Let's take a quick look at the data.

```

[5]: df_LoP.head()

```

```

[5]: PostalCode      Borough      Neighbourhood
0      M1A      Not assigned      Not assigned

```

1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

Cleanup the Dataset

```
[6]: # Drop rows with Borough not assigned
df_LoP = df_LoP[~df_LoP['Borough'].str.contains("Not assigned") == True]

# More than one neighborhood for a Borough are combined into one row with the
# neighborhoods separated with a comma
df_gp_LoP = df_LoP.groupby(['PostalCode', 'Borough'])['Neighbourhood'].apply(', '.join).
    reset_index()
df_gp_LoP.columns = ['PostalCode', 'Borough', 'Neighbourhood']

# If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the
# same as the borough.
df_gp_LoP['Neighbourhood'].replace('Not assigned', "Queen's Park", inplace=True)
df_gp_LoP
```

```
[6]:
```

	PostalCode	Borough \
0	M1B	Scarborough
1	M1C	Scarborough
2	M1E	Scarborough
3	M1G	Scarborough
4	M1H	Scarborough
5	M1J	Scarborough
6	M1K	Scarborough
7	M1L	Scarborough
8	M1M	Scarborough
9	M1N	Scarborough
10	M1P	Scarborough
11	M1R	Scarborough
12	M1S	Scarborough
13	M1T	Scarborough
14	M1V	Scarborough
15	M1W	Scarborough
16	M1X	Scarborough
17	M2H	North York
18	M2J	North York
19	M2K	North York
20	M2L	North York
21	M2M	North York
22	M2N	North York
23	M2P	North York
24	M2R	North York

25	M3A	North York
26	M3B	North York
27	M3C	North York
28	M3H	North York
29	M3J	North York
30	M3K	North York
31	M3L	North York
32	M3M	North York
33	M3N	North York
34	M4A	North York
35	M4B	East York
36	M4C	East York
37	M4E	East Toronto
38	M4G	East York
39	M4H	East York
40	M4J	East York
41	M4K	East Toronto
42	M4L	East Toronto
43	M4M	East Toronto
44	M4N	Central Toronto
45	M4P	Central Toronto
46	M4R	Central Toronto
47	M4S	Central Toronto
48	M4T	Central Toronto
49	M4V	Central Toronto
50	M4W	Downtown Toronto
51	M4X	Downtown Toronto
52	M4Y	Downtown Toronto
53	M5A	Downtown Toronto
54	M5B	Downtown Toronto
55	M5C	Downtown Toronto
56	M5E	Downtown Toronto
57	M5G	Downtown Toronto
58	M5H	Downtown Toronto
59	M5J	Downtown Toronto
60	M5K	Downtown Toronto
61	M5L	Downtown Toronto
62	M5M	North York
63	M5N	Central Toronto
64	M5P	Central Toronto
65	M5R	Central Toronto
66	M5S	Downtown Toronto
67	M5T	Downtown Toronto
68	M5V	Downtown Toronto
69	M5W	Downtown Toronto
70	M5X	Downtown Toronto
71	M6A	North York

72	M6B	North York
73	M6C	York
74	M6E	York
75	M6G	Downtown Toronto
76	M6H	West Toronto
77	M6J	West Toronto
78	M6K	West Toronto
79	M6L	North York
80	M6M	York
81	M6N	York
82	M6P	West Toronto
83	M6R	West Toronto
84	M6S	West Toronto
85	M7A	Queen's Park
86	M7R	Mississauga
87	M7Y	East Toronto
88	M8V	Etobicoke
89	M8W	Etobicoke
90	M8X	Etobicoke
91	M8Y	Etobicoke
92	M8Z	Etobicoke
93	M9A	Etobicoke
94	M9B	Etobicoke
95	M9C	Etobicoke
96	M9L	North York
97	M9M	North York
98	M9N	York
99	M9P	Etobicoke
100	M9R	Etobicoke
101	M9V	Etobicoke
102	M9W	Etobicoke

	Neighbourhood
0	Rouge, Malvern
1	Highland Creek, Rouge Hill, Port Union
2	Guildwood, Morningside, West Hill
3	Woburn
4	Cedarbrae
5	Scarborough Village
6	East Birchmount Park, Ionview, Kennedy Park
7	Clairlea, Golden Mile, Oakridge
8	Cliffcrest, Cliffside, Scarborough Village West
9	Birch Cliff, Cliffside West
10	Dorset Park, Scarborough Town Centre, Wexford . . .
11	Maryvale, Wexford
12	Agincourt
13	Clarks Corners, Sullivan, Tam O'Shanter

14	Agincourt North, L'Amoreaux East, Milliken, St. . .
15	L'Amoreaux West
16	Upper Rouge
17	Hillcrest Village
18	Fairview, Henry Farm, Oriole
19	Bayview Village
20	Silver Hills, York Mills
21	Newtonbrook, Willowdale
22	Willowdale South
23	York Mills West
24	Willowdale West
25	Parkwoods
26	Don Mills North
27	Flemingdon Park, Don Mills South
28	Bathurst Manor, Downsview North, Wilson Heights
29	Northwood Park, York University
30	CFB Toronto, Downsview East
31	Downsview West
32	Downsview Central
33	Downsview Northwest
34	Victoria Village
35	Woodbine Gardens, Parkview Hill
36	Woodbine Heights
37	The Beaches
38	Leaside
39	Thorncliffe Park
40	East Toronto
41	The Danforth West, Riverdale
42	The Beaches West, India Bazaar
43	Studio District
44	Lawrence Park
45	Davisville North
46	North Toronto West
47	Davisville
48	Moore Park, Summerhill East
49	Deer Park, Forest Hill SE, Rathnelly, South Hi. . .
50	Rosedale
51	Cabbagetown, St. James Town
52	Church and Wellesley
53	Harbourfront, Regent Park
54	Ryerson, Garden District
55	St. James Town
56	Berczy Park
57	Central Bay Street
58	Adelaide, King, Richmond
59	Harbourfront East, Toronto Islands, Union Station
60	Design Exchange, Toronto Dominion Centre

61 Commerce Court, Victoria Hotel
 62 Bedford Park, Lawrence Manor East
 63 Roselawn
 64 Forest Hill North, Forest Hill West
 65 The Annex, North Midtown, Yorkville
 66 Harbord, University of Toronto
 67 Chinatown, Grange Park, Kensington Market
 68 CN Tower, Bathurst Quay, Island airport, Harbo...
 69 Stn A PO Boxes 25 The Esplanade
 70 First Canadian Place, Underground city
 71 Lawrence Heights, Lawrence Manor
 72 Glencairn
 73 Humewood-Cedarvale
 74 Caledonia-Fairbanks
 75 Christie
 76 Dovercourt Village, Dufferin
 77 Little Portugal, Trinity
 78 Brockton, Exhibition Place, Parkdale Village
 79 Downsview, North Park, Upwood Park
 80 Del Ray, Keelesdale, Mount Dennis, Silverthorn
 81 The Junction North, Runnymede
 82 High Park, The Junction South
 83 Parkdale, Roncesvalles
 84 Runnymede, Swansea
 85 Queen's Park
 86 Canada Post Gateway Processing Centre
 87 Business Reply Mail Processing Centre 969 Eastern
 88 Humber Bay Shores, Mimico South, New Toronto
 89 Alderwood, Long Branch
 90 The Kingsway, Montgomery Road, Old Mill North
 91 Humber Bay, King's Mill Park, Kingsway Park So...
 92 Kingsway Park South West, Mimico NW, The Queen...
 93 Islington Avenue
 94 Cloverdale, Islington, Martin Grove, Princess ...
 95 Bloordeale Gardens, Eringate, Markland Wood, Ol...
 96 Humber Summit
 97 Emery, Humberlea
 98 Weston
 99 Westmount
 100 Kingsview Village, Martin Grove Gardens, Richv...
 101 Albion Gardens, Beaumond Heights, Humbergate, ...
 102 Northwest

Read the csv file that has the geographical coordinates of each postal code from http://cocl.us/Geospatial_data into a pandas dataframe.

```
[7]: df_geo = pd.read_csv('http://cocl.us/Geospatial_data')
df_geo.columns = ['PostalCode', 'Latitude', 'Longitude']
```


Perform a merge the Latitude and Longitude from the Geospatial dataframe with the Toronto neighbourhoods using inner join. Print the top 5 from the merged dataframe.

```
[8]: df_pos = pd.merge(df_gp_LoP, df_geo, on=['PostalCode'], how='inner')

df_tor_neigh = df_pos[['PostalCode', 'Borough', 'Neighbourhood', 'Latitude', 'Longitude']].
    ↳copy()

df_tor_neigh.head()
```

```
[8]: PostalCode    Borough    Neighbourhood  Latitude \
0      M1B  Scarborough    Rouge, Malvern  43.806686
1      M1C  Scarborough  Highland Creek, Rouge Hill, Port Union  43.784535
2      M1E  Scarborough    Guildwood, Morningside, West Hill  43.763573
3      M1G  Scarborough                Woburn  43.770992
4      M1H  Scarborough    Cedarbrae  43.773136

Longitude
0 -79.194353
1 -79.160497
2 -79.188711
3 -79.216917
4 -79.239476
```

Print the number of rows of the dataframe.

```
[9]: df_tor_neigh.shape
```

```
[9]: (103, 5)
```

Let's simplify the dataframe and cluster only the neighborhoods in Toronto. So let's slice the original dataframe and just take portion of dataframe where Boroughs contain word Toronto.

```
[10]: toronto_data = df_tor_neigh[df_tor_neigh['Borough'].str.contains("Toronto")].
    ↳reset_index(drop=True)
toronto_data.head()
```

```
[10]: PostalCode    Borough    Neighbourhood  Latitude \
0      M4E  East Toronto    The Beaches  43.676357
1      M4K  East Toronto  The Danforth West, Riverdale  43.679557
2      M4L  East Toronto  The Beaches West, India Bazaar  43.668999
3      M4M  East Toronto    Studio District  43.659526
4      M4N  Central Toronto    Lawrence Park  43.728020

Longitude
0 -79.293031
1 -79.352188
2 -79.315572
3 -79.340923
4 -79.388790
```

1.3 Conclusion

1.3.1 Week 1: Description of Problem, Discussion of Background and Data

We get the initial Dataframe with Postcode, Borough and corresponding Neighbourhoods in each Borough and the coordinates of those major Neighbourhoods.

So as the next step I will use [Foursquare](#) data and obtain information on restaurants. With these, we can start with our battle of neighborhoods for opening a restaurant in Toronto.

1.3.2 Thank You !