

# **CAPSTONE PROJECT ASSIGNMENT**

## **RESTAURANT BEST LOCATION ANALYSIS**

**2020**

Submitted by  
ANAND R NAIR

## **TABLE OF CONTENTS**

- 1. INTRODUCTION/PROBLEM STATEMENT**
- 2. DATA DESCRIPTION**
- 3. SOLUTION APPROACH**
- 4. METHODOLOGY**
- 5. RESULTS**
- 6. DISCUSIION**
- 7. CONCLUSION**

## 1. INTRODUCTION / BUSINESS PROBLEM STATEMENT

Find the best locations for starting a new Restaurant in Toronto City

A global restaurant chain company wants to start a restaurant shop in one of the best neighbourhoods in Toronto city. They want to find the best neighbourhood locations where there are no, or minimal restaurants currently exists in the area. They are looking for less competitors in the area but a more popular place to start with.

Since they are a multicuisine restaurant chains, they want to consider all types of restaurants that provide different country dishes.

The problem is to find and filter such locations and apply the best statistical approaches to find the best location(s) and visualize with statistical proofs to confirm the analysis.

## 2. DATA DESCRIPTION

The primary data that will be needed is the popular locations within Toronto city and a reliable source of data.

We have the Canadian postal data from the wiki which we can leverage to find the popular boroughs and the neighbourhoods that exists in each borough.

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

The data primarily consists of the below details in tabular format

1. Postal Code
2. Borough Name
3. Neighbourhood(s)

<b>M4A</b> North York (Victoria Village)	<b>M5A</b> Downtown Toronto (Regent Park / Harbourfront)	<b>M6A</b> North York (Lawrence Manor / Lawrence Heights)
<b>M4B</b> East York (Parkview Hill / Woodbine Gardens)	<b>M5B</b> Downtown Toronto (Garden District, Ryerson)	<b>M6B</b> North York (Glencairn)
<b>M4C</b> East York (Woodbine Heights)	<b>M5C</b> Downtown Toronto (St. James Town)	<b>M6C</b> York (Humewood-Cedarvale)

### 3. SOLUTION APPROACH

The existing data is tabular format which we cannot input directly to any data analysis tools. Below is a step by step solution approach to leverage the table data to find suitable locations.

1. Scrap the Canada Postal data from the web page into processable input format
2. Get the location data for each neighbourhood/borough and update the input data
3. Visualize the neighbourhoods in Canada
4. Get the popular neighbourhoods within a defined geographic circle of Toronto
5. Explore the neighbourhoods in Toronto
6. Analyse each neighbourhood based on restaurant data
7. Cluster each neighbourhood based on restaurant data
8. Examine and selected the best clusters with that satisfies the restaurant location requirements

#### Detailed Solution Approach

1. *Scrap the Canada Postal data from the web page into processable input format*

The webpage data in the tabular form needs to be read and converted into a suitable data structure that can be used as an input to any analysing tools. The web scraping libraries such as Scrapy, beautiful soup or python requests to convert the data into usable format. In this assignment we'll use the python requests. The expectation is to create a new tabular data structure in below format.

Postal Code	Borough	Neighbourhood(s)
-------------	---------	------------------

2. *Get the location data for each borough and update the input data*

The primary data must be updated with additional useful data, such as geographic location data with latitudes and longitudes to get more information about the places. We can leverage geopy library or third-party data sets to do this. In this assignment we'll use one of the third-party data sets to combine the location data to the input data set of the Canadian neighbourhoods.

The above-mentioned in 1. table will be updated to below format.

Postal Code	Borough	Neighbourhood(s)	Latitude	Longitude
-------------	---------	------------------	----------	-----------

3. *Visualize the Canadian neighbourhoods*

This is an optional step. However, it is useful to get a good high-level visual view of the initial input data set that is going to be analysed and refined in the further stages. We can leverage the existing visualization libraries such as Folium to get this view.

4. *Get the popular neighbourhoods within a defined geographic circle*

In this step we'll get the popular neighbourhoods within a more defined geographical circle in and around Toronto. This requires two steps.

- a. First, the above table must be filtered for records of Toronto.
- b. Pass the required input data to a location data API such as the Foursquare API to get more details on the Toronto neighbourhoods which can be input to the next stages in the solution approach.

The results from the location data API such as FourSquare can be used to get a lot of information about the explored locations such as its popularity and the reasons for its popularity. This can be

filtered to identify the suitable locations that can be taken as candidates for the problem statement. A sample response data is shown below for another location.

*Below is a Sample Four Square Data of a Diner in Tibett Ave, US*

```
{'reasons': {'count': 0,
  'items': [{'summary': 'This spot is popular',
    'type': 'general',
    'reasonName': 'globalInteractionReason'}]},
'venue': {'id': '4b79cc46f964a520c5122fe3',
  'name': 'Tibbett Diner',
  'location': {'address': '3033 Tibbett Ave',
    'crossStreet': 'btwn 230th & 231st',
    'lat': 40.8804044222466,
    'lng': -73.90893738006402,
    'labeledLatLngs': [{'label': 'display',
      'lat': 40.8804044222466,
      'lng': -73.90893738006402}]},
  'distance': 452,
  'postalCode': '10463',
  'cc': 'US',
  'city': 'Bronx',
  'state': 'NY',
  'country': 'United States',
  'formattedAddress': ['3033 Tibbett Ave (btwn 230th & 231st)',
    'Bronx, NY 10463',
    'United States']},
'categories': [{'id': '4bf58dd8d48988d147941735',
  'name': 'Diner',
  'pluralName': 'Diners',
  'shortName': 'Diner',
  'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/diner_',
    'suffix': '.png'},
  'primary': True}],
'photos': {'count': 0, 'groups': []},
'referralId': 'e-0-4b79cc46f964a520c5122fe3-2'},
```

5. *Explore the neighbourhoods in Toronto*

The above step 4 will give you the neighbourhood data in Toronto with additional features and details. This data can be explored to find the suitable locations to identify where all restaurants exists.

6. *Analyse each neighbourhood based on restaurant data*

We can analyse each neighbourhood based on additional data such as if the same venue has a more popularly visited additional place such as a hotels or parks. This increases the business chances for a restaurant.

7. *Cluster each neighbourhood based on restaurant data*

In this step we'll cluster the results based on the above business criteria and select the best locations for listing in a tabular and visual format.

8. *Examine and selected the best clusters with that satisfies restaurant location requirements*

Examine the best clusters and list them as suitable candidate places for starting new restaurant and locations.

#### 4. METHODOLOGY

Technology : Python, Jupyter Notebook, Anaconda/ FourSquare API

Python Libraries : Python requests / Folium/ Geopy

#### Download and data set and Wrangle it

1. The address data obtained from the Wiki was scrapped using the Python requests into input data frame format of python.

Out[57]:

	PostalCode	Borough	Neighborhood
0	M9A	Etobicoke	Islington Avenue
1	M9B	Etobicoke	West Deane Park , Princess Gardens , Martin Grove , Islington , Cloverdale
2	M9C	Etobicoke	Eringate , Bloordale Gardens , Old Burnhamthorpe , Markland Wood
3	M6E	York	Caledonia-Fairbanks
4	M6G	Downtown Toronto	Christie
5	M6H	West Toronto	Dufferin , Dovercourt Village
6	M6J	West Toronto	Little Portugal , Trinity
7	M6K	West Toronto	Brockton , Parkdale Village , Exhibition Place
8	M9L	North York	Humber Summit
9	M9M	North York	Humberlea , Emery
10	M9N	York	Weston
11	M9P	Etobicoke	Westmount
12	M9R	Etobicoke	Kingsview Village , St. Phillips , Martin Grove Gardens , Richview Gardens
13	M6S	West Toronto	Runnymede , Swansea
14	M5T	Downtown Toronto	Kensington Market , Chinatown , Grange Park
15	M9V	Etobicoke	South Steeles , Silverstone , Humbergate , Jamestown , Mount Olive , Beaumont Heights , Thistletown , Albion Gardens
16	M9W	EtobicokeNorthwest	Clairville , Humberwood , Woodbine Downs , West Humber , Kipling Heights , Rexdale , Elms , Tandridge , Old Rexdale
17	M8X	Etobicoke	The Kingsway , Montgomery Road , Old Mill North
18	M8Y	Etobicoke	Old Mill South , King's Mill Park , Sunnylea , Humber Bay , Mimico NE , The Queensway East , Royal York South East , Kingsway Park South East
19	M8Z	Etobicoke	Mimico NW , The Queensway West , South of Bloor , Kingsway Park South West , Royal York South West

2. Filter the data to exclude the areas which does not have a borough. Without a borough we cannot get a more accurate neighbourhood.
3. Load the external geocode data and then assign it to each matching postal code. Add additional columns latitude and longitude to the postal area data frame.
4. Filter the Toronto neighbourhoods from the result set.

## Capstone Project – Battle of Neighbourhoods

Out [58] :

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M9A	Etobicoke	Islington Avenue	43.6679	-79.5322
1	M9B	Etobicoke	West Deane Park , Princess Gardens , Martin Grove , Islington , Cloverdale	43.6509	-79.5547
2	M9C	Etobicoke	Eringate , Bloordale Gardens , Old Burnhamthorpe , Markland Wood	43.6435	-79.5772
3	M6E	York	Caledonia-Fairbanks	43.689	-79.4535
4	M6G	Downtown Toronto	Christie	43.6695	-79.4226
5	M6H	West Toronto	Dufferin , Dovercourt Village	43.669	-79.4423
6	M6J	West Toronto	Little Portugal , Trinity	43.6479	-79.4197
7	M6K	West Toronto	Brockton , Parkdale Village , Exhibition Place	43.6368	-79.4282
8	M9L	North York	Humber Summit	43.7563	-79.566
9	M9M	North York	Humberlea , Emery	43.7248	-79.5322
10	M9N	York	Weston	43.7069	-79.5182
11	M9P	Etobicoke	Westmount	43.6963	-79.5322
12	M9R	Etobicoke	Kingsview Village , St. Phillips , Martin Grove Gardens , Richview Gardens	43.6889	-79.5547
13	M6S	West Toronto	Runnymede , Swansea	43.6516	-79.4844
14	M5T	Downtown Toronto	Kensington Market , Chinatown , Grange Park	43.6532	-79.4
15	M9V	Etobicoke	South Steeles , Silverstone , Humbergate , Jamestown , Mount Olive , Beaumont Heights , Thistletown , Albion Gardens	43.7394	-79.5884
16	M9W	EtobicokeNorthwest	Clairville , Humberwood , Woodbine Downs , West Humber , Kipling Heights , Rexdale , Elms , Tandridge , Old Rexdale	43.7067	-79.5941
17	M8X	Etobicoke	The Kingsway , Montgomery Road , Old Mill North	43.6537	-79.5069
18	M8Y	Etobicoke	Old Mill South , King's Mill Park , Sunnylea , Humber Bay , Mimico NE , The Queensway East , Royal York South East , Kingsway Park South East	43.6363	-79.4985
19	M8Z	Etobicoke	Mimico NW , The Queensway West , South of Bloor , Kingsway Park South West , Royal York South West	43.6288	-79.521

### Explore and Analyse the data set

1. Get the neighbourhood data and all its nearby venues using the Foursquared API.
2. Looked for the neighbourhoods that show a high mean value for the frequency for the Restaurants and other Diner segments.

Out [63] :

	Neighborhood	Average Restaurant Frequency
0	Brockton , Parkdale Village , Exhibition Place	0.004348
1	Christie	0.003704
2	Dufferin , Dovercourt Village	0.004167
3	Kensington Market , Chinatown , Grange Park	0.011111
4	Little Portugal , Trinity	0.010303
5	Runnymede , Swansea	0.008333

### Cluster the data

1. Since there are multiple categories that fall into each Venue, we used the K-Means clustering to group the Neighbourhoods into different clusters.
2. In this case the number of clusters chosen were 5 and the data was clustered into an increasing order of most common avenues.
3. Based on the results, the top 5 clusters were chosen to see which clusters had a restaurant in the most common initial venues.

- Also, other types of diners were considered to determine the best locations to start a new restaurant.
- The mean value for each type of restaurant can be grouped for each Neighbourhood as below.

Out[27]:

	Neighborhood	Asian Restaurant	Belgian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Doner Restaurant	Dumpling Restaurant	Empanada Restaurant	Falafel Restaurant	Fast Food Restaurant	Re
0	Brockton , Parkdale Village , Exhibition Place	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.0000	0
1	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.0000	0
2	Dufferin , Dovercourt Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.0625	0
3	Kensington Market , Chinatown , Grange Park	0.000000	0.011494	0.011494	0.022989	0.022989	0.000000	0.011494	0.034483	0.011494	0.000	0.0000	0

## 5. RESULTS

Based on the analysis, we can see that the below locations have higher mean values for restaurants.

- Among these Kensington Market, China Town has a higher mean frequency of restaurants and **Christie neighbourhood in Downtown Toronto has a lower value of mean frequency of restaurants.**

Out[45]:

	Neighborhood	Average Restaurant Frequency
0	Brockton , Parkdale Village , Exhibition Place	0.004348
1	Christie	0.003704
2	Dufferin , Dovercourt Village	0.004167
3	Kensington Market , Chinatown , Grange Park	0.011111
4	Little Portugal , Trinity	0.010303
5	Runnymede , Swansea	0.008333

- Based on the most common venue ordering also, **Christie shows a restaurant in the 10<sup>th</sup> most common venue only, which makes it a most suitable location for the restaurant company chain to start its first initiative in Toronto.**

Also, Christie is the second most popular neighbourhood also.

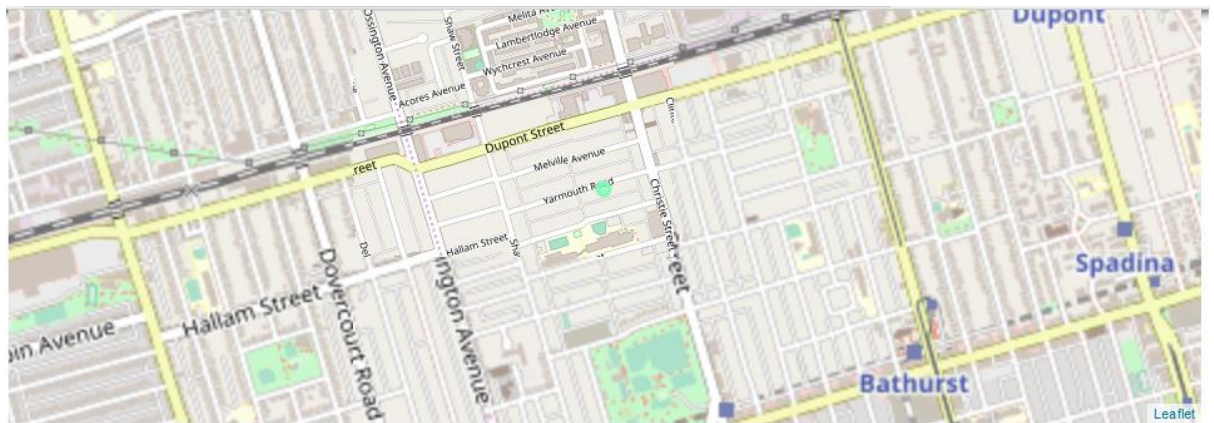


Out[13]:

	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	Brockton , Parkdale Village , Exhibition Place	Coffee Shop	Café	Breakfast Spot	Furniture / Home Store	Restaurant	Climbing Gym	Japanese Restaurant	Italian Restaurant	Nightclub	Intersection
1	Christie	Grocery Store	Café	Park	Nightclub	Candy Store	Coffee Shop	Athletics & Sports	Baby Store	Gas Station	Italian Restaurant
2	Dufferin , Dovercourt Village	Bakery	Pharmacy	Brewery	Supermarket	Pool	Café	Park	Fast Food Restaurant	Gym / Fitness Center	Bar
3	Kensington Market , Chinatown , Grange Park	Bar	Vietnamese Restaurant	Café	Bakery	Vegetarian / Vegan Restaurant	Coffee Shop	Mexican Restaurant	Dumpling Restaurant	Noodle House	Burger Joint
4	Little Portugal , Trinity	Bar	Coffee Shop	Restaurant	Asian Restaurant	Pizza Place	Café	Men's Store	Vietnamese Restaurant	Wine Bar	Mac & Cheese Joint

The second-best choice would be Dufferin, Dovercourt Village which has the next higher restaurant frequency and has a restaurant as its most common venue in 8<sup>th</sup> position only.

#### MAP SHOWING CHRISTIE STREET AS THE BEST LOCATION



## 6. DISCUSSION

In addition to above, other types of Diners also can be considered if needed to get a more results on the location. However, it may enlarge the scope of search and reduce accuracy. Also the neighbourhoods available under Toronto is comparatively less. Hence a more rich data set would definitely improve the results.

## 7. CONCLUSION

To summarize, the analysis of around 5 neighbourhoods in Toronto revealed two best locations to start a new restaurant with one looks most promising.

The Restaurant location study report can be generalized, expanded and applied to any city around the globe to find solutions to similar problems.