

This capstone project 'Battle of Neighborhoods' is part of Applied Data Science capstone course, Coursera, which is a sub-course of 'IBM Data Science Professional Certificate'

---

Ankush Sharma

## Contents

---

Background   Problem statement.....	3
Description of the background .....	3
Data   Implementation (Solution to problem) .....	4
Methodology .....	6
Result - Analysis outcome .....	11
Discussions and Recommendations .....	12
Conclusion .....	12

## Background | Problem statement

---

The most trending, essential and leading business in most of the western countries and other is the *food industry*. The business case considered focusses on identifying suitable location for *starting a restaurant in the neighborhoods of Toronto city, Ontario, Canada*.

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 6,139,404 in 2019 and has been predicted to reach 6,491,286 in 2025 at the growth rate of 0.93%. Hence we will be focussing in Toronto city and its borough for this project.

### Key Benefitters | Target-audience:

Business ventures, Entrepreneurs, Investors, Startups companies, Vendors etc.

## Description of the background

---

With the help of Data Science platform, necessary evaluations will be done for each neighborhood based on certain criteria, to find the most popular and suitable business location helping the key benefitters to choose on their own. Thus, saving the time and efforts spent for data analysis, and preventing the 60% to 80% of failing of restaurants!

According to the statistics provided by CNBC report CNBC report, around 60% of new restaurant fail within the first year and nearly 80% shutter before fifth anniversary. The most common and top reason is the location of the restaurant and the lack of self-awareness about the business's appropriate location. To help identifying the right business location among many towns in Toronto city for starting the restaurant, this research project provides a recommendation considering various factors including but not limited to analyzing venues in and around the neighborhoods of Toronto city, comparing and providing the best and suitable place nearby Toronto city, Canada to start and establish the restaurant successfully!

## Data | Implementation (Solution to problem)

---

This research has utilized and implemented following,

### Dataset

The dataset used for the project is extracted from Wikipedia link

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

---

- Postcode
- Borough
- Neighborhood
- Latitude
- Longitude

### Details of Neighborhood

- Venues
- Borough
- Latitude
- Longitude
- Venue category

### Population

- Name of the neighborhood
- Population rate

Geographical details are queried from **Foursquare** platform, performing **API call traversal** and processing the resulting **JSON file to pandas dataframe**.

### Libraries used

- Pandas
- Matplotlib
- Numpy
- Folium
- Scikit-learn

## Algorithms • K-means clustering

It is one of the most common and simplest unsupervised machine learning algorithms. This is used for clustering the similar group of datapoints together which is not similar to other set of clusters of datapoints.

This is done by randomly selecting centroids, which is the beginning point of every clusters and performing iterative calculation to optimize the positions of the centroids.

The identification and optimization of clusters stops when there is no change in their values as the clustering is found to be successful.

## Purpose of K-means in project

- This clustering algorithm is used for identification of most common venues in the neighborhoods / towns of Toronto city, Canada to identify the best and most suitable location for intended business amongst them all. This will help the key benefiter to precisely choose location suitable for their restaurant business.

## End-result

- The outcome will suggest the suitable business location automatically reducing the cost of analysis, time and efforts for the same.
- Henceforth, **preventing the occurrence of 60% and 80% shut-off** level of restaurants as per the reports of CNBC (**60%** of new restaurant fail within the *first year* and nearly **80%** shutter before *fifth anniversary*)

## Factors considered

For the analysis of considered problem statement, following are the factors addressed,

- Popular neighborhoods
- Statistics on population in neighborhoods
- Most interested venues **Data extraction**
- Primary data is extracted from, Wikipedia, <https://www.cnn.com/2016/01/20/heres-the-real-reasonwhy-most-restaurants-fail.html>
- Details of neighborhood is extracted using the Foursquare platform
- Population details are fetched from, [https://en.wikipedia.org/wiki/Demographics\\_of\\_Toronto\\_neighbourhoods](https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods)
- Geospatial data is utilized from [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

This project is published in my Github repo [https://github.com/eshrnks/Coursera\\_Capstone](https://github.com/eshrnks/Coursera_Capstone)

## Methodology

This project objective is to recommend a suitable business location for starting a restaurant in neighborhoods of Toronto city, based on certain criteria. Analysis will be performed on the neighborhoods within 5km from the center of city.

### First stage

- **Required data** - Postcodes | Borough | Neighborhood | Latitude | Longitude

#### Basic Info table

	Postcode	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

#### Primary table

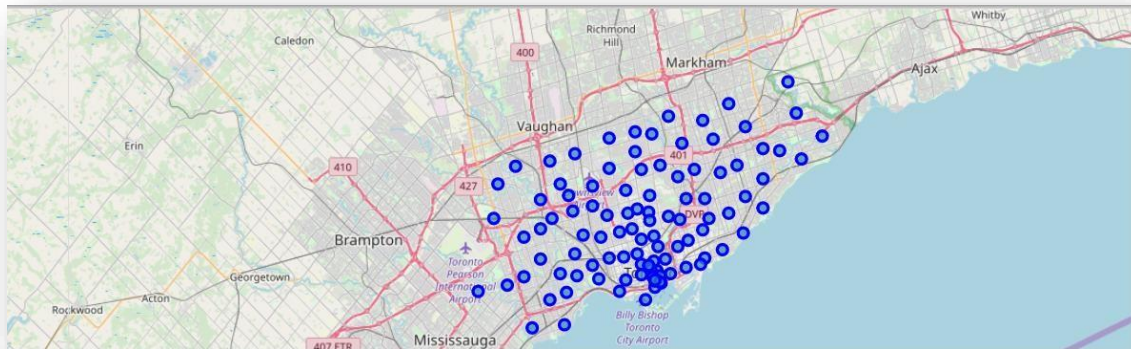
	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H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

- **Necessary processing and refining of primary data** - where the data is refined to 38 Neighborhoods for 4 Borough of Toronto city

## Second stage

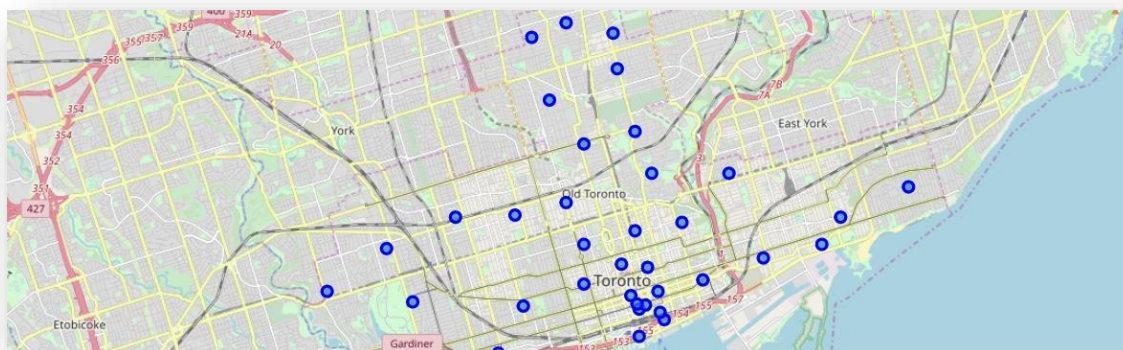
- **Exploration and identification of Venues** - Venues in nearby neighborhood for the selected Borough will be identified using the Foursquare API service. It is plotted as map superimposed in the map of Toronto city. This is done using folium library. Later, identifying the top 10 most common and resident's preferred venues in order to identify which neighborhood prefers and are more interested in food and restaurants.

### All Borough of Toronto



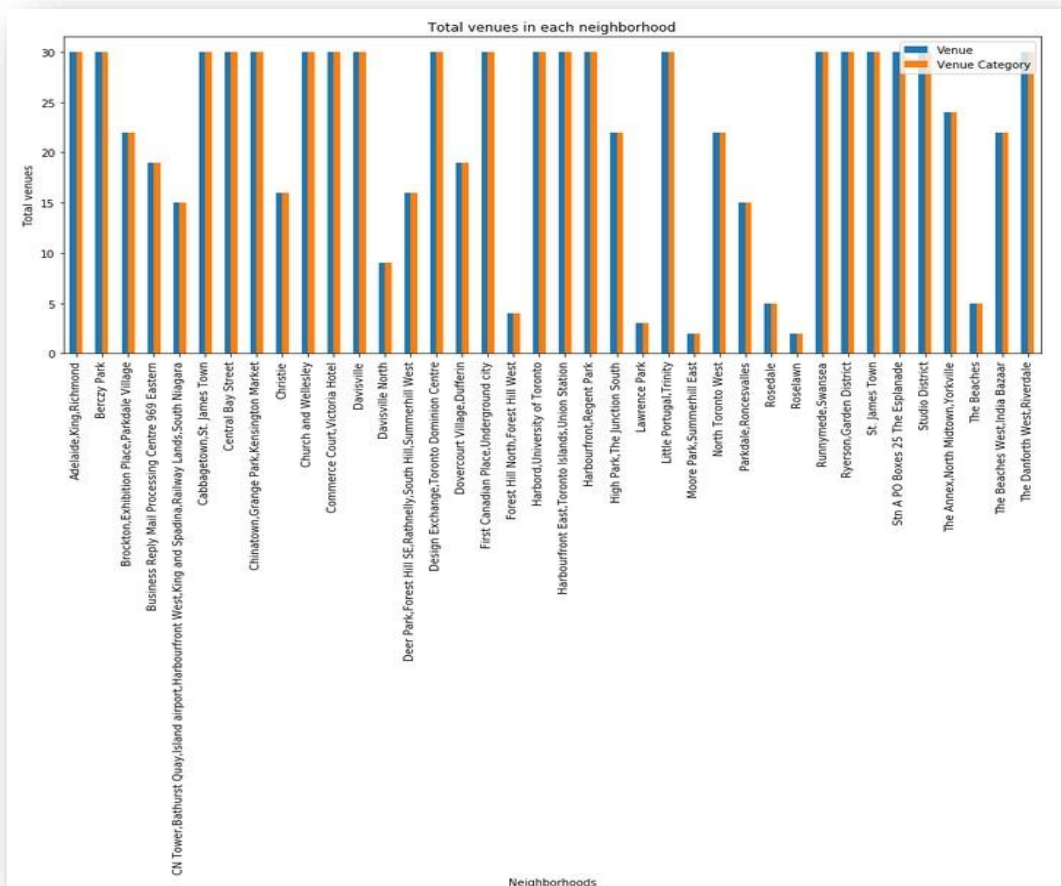
The above map is generated with all borough of Canada, using the Folium API. This has total coordinates of 103 being superimposed in the map.

### Selected Borough having Toronto neighbourhood



The above map is the plotting of the Borough restricted to Toronto as our purpose focusses on Toronto city. This map has been refined with 88 coordinates and has been done using Folium API. Each borough is found to have unique 38 neighbourhoods.

## Venues



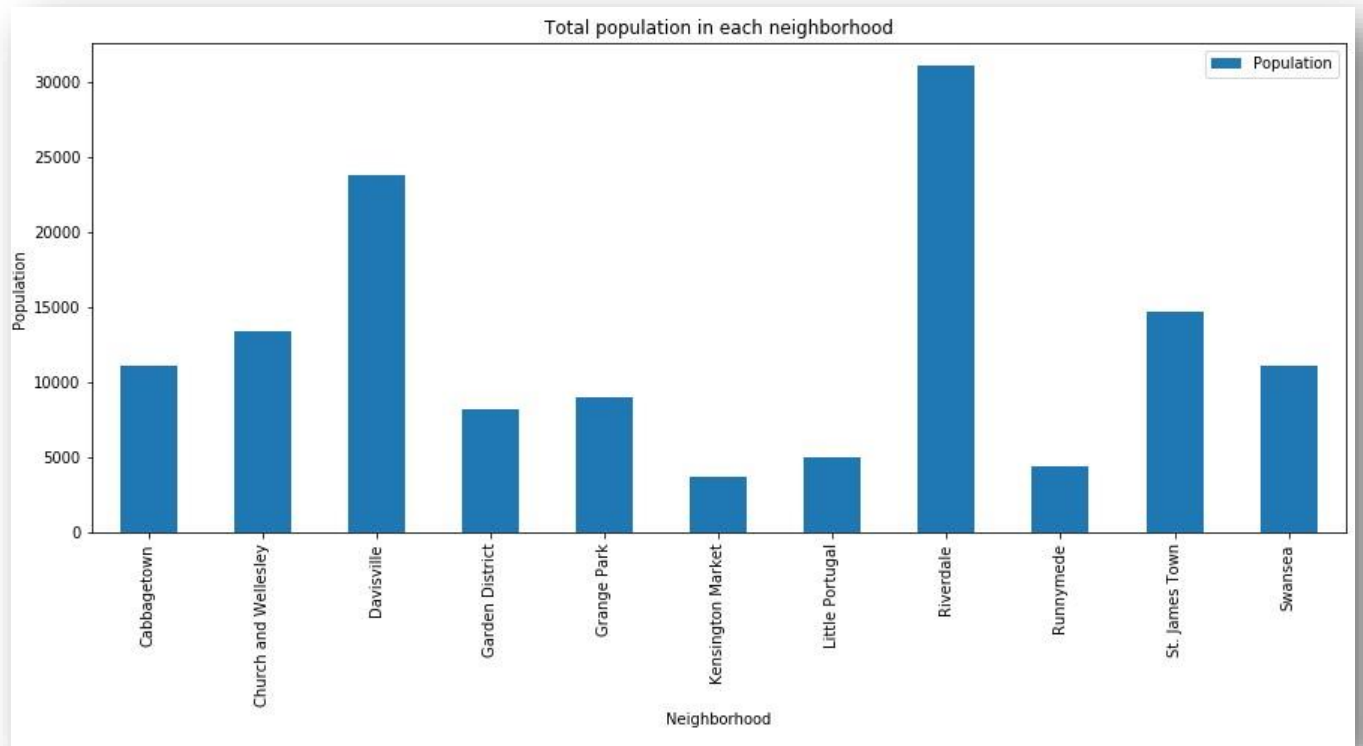
This bar chart was created to project the total venues obtained for each neighbourhood. This was generated with the help of matplotlib library.

This chart helped to identify the neighbourhoods having top venues. Below provided is the table displaying the names of the neighbourhoods with top venues, the rest of the neighbourhoods are removed from the list for better visualization of the data.



	Venue	Venue Category
Neighborhood		
Adelaide,King,Richmond	30	30
Berczy Park	30	30
Cabbagetown,St. James Town	30	30
Central Bay Street	30	30
Chinatown,Grange Park,Kensington Market	30	30
Church and Wellesley	30	30
Commerce Court,Victoria Hotel	30	30
Davisville	30	30
Design Exchange,Toronto Dominion Centre	30	30
First Canadian Place,Underground city	30	30
Harbord,University of Toronto	30	30
Harbourfront East,Toronto Islands,Union Station	30	30
Harbourfront,Regent Park	30	30
Little Portugal,Trinity	30	30
Runnymede,Swansea	30	30
Ryerson,Garden District	30	30
St. James Town	30	30
Stn A PO Boxes 25 The Esplanade	30	30
Studio District	30	30
The Danforth West,Riverdale	30	30

- **Identifying Population rate for selected neighbourhood** - Here, we will be scraping the data from web and converting to dataframe using Pandas library. This will help us compare and analyze the data with our previous findings, in identification of suitable business location



From the above chart, it clearly depicts the fact the **Riverdale** has more population conveying the potential of revenue generating opportunity and capability.

## Final stage

Algorithm to perform clustering- K-Means clustering is performed to accurate our analysis and save the 60% to 80% of shut-off rate of restaurants due to location problem

Postcode	Borough	Neighbourhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Other Great Outdoors	Trail	Health Food Store	Pub	Wine Shop	Creperie	Eastern European Restaurant	Dog Run	Discount Store	Diner
41	M4K	East Toronto	The Danforth West/Riverdale	43.679557	-79.352188	0	Greek Restaurant	Ice Cream Shop	Italian Restaurant	Yoga Studio	Pizza Place	Cosmetics Shop	Pub	Restaurant	Dessert Shop	Diner
42	M4L	East Toronto	The Beaches West/India Bazaar	43.668999	-79.315572	0	Park	Sandwich Place	Gym	Sushi Restaurant	Pet Store	Pizza Place	Movie Theater	Pub	Burrito Place	Burger Joint
43	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	American Restaurant	Italian Restaurant	Bakery	Cheese Shop	Coworking Space	Chinese Restaurant	Fish Market	Middle Eastern Restaurant
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	3	Bus Line	Park	Swim School	Creperie	Eastern European Restaurant	Dog Run	Discount Store	Diner	Dessert Shop	Deli / Bodega

There were more common venues in boroughs in the dataset. Hence, I used KMeans clustering, an unsupervised algorithm to cluster the boroughs having common venues. Here, the boroughs were clustered into 5 and the resulting table is displayed above which has top ten most common venues among the boroughs.

Each 5 clusters are evaluated separately.

## Result - Analysis outcome

---

### Initial outcome

From the above analysis, following are the observations

- Total records received - 103 data
- Post refining - 38 unique Neighborhoods | 4 Borough of Toronto city | 187 unique categories
- Final refining - 20 Neighborhoods has more venues which was reduced from 38 previously

Following are the neighborhood suggested for starting the restaurant business as these neighborhoods have the more venues which may be challenging but yet it clearly shows the high opportunity rate for starting the business

1. Adelaide,King,Richmond
2. Berczy Park
3. Cabbagetown,St. James Town
4. Central Bay Street
5. Chinatown,Grange Park,Kensington Market
6. Church and Wellesley
7. Commerce Court,Victoria Hotel
8. Davisville
9. Design Exchange,Toronto Dominion Centre
10. First Canadian Place,Underground city
11. Harbord,University of Toronto
12. Harbourfront East,Toronto Islands,Union Station
13. Harbourfront,Regent Park
14. Little Portugal,Trinity
15. Runnymede,Swansea
16. Ryerson,Garden District
17. St. James Town
18. Stn A PO Boxes 25 The Esplanade

19. Studio District

20. The Danforth West, Riverdale

## Discussions and Recommendations

---

After further research, following are the result obtained,

The most populated neighbourhood is **Riverdale**. Also, if we retrieve the most common venue in Riverdale it is found that people are more interested in 'restaurants' comparing to other neighbourhoods. Hence this 'Riverdale' might be a good choice to start a restaurant from the analysis performed based on the following criteria

- Venue interest of people residing in the Riverdale
- Population rate
- Clustering

Result from overall observation, analysis we will be able to provide a suggestion on suitable business location to key benefiteres or target audiences to start their restaurant business and establish successfully.

As per the outcome of this research, 'Riverdale' is most suitable for our business purpose. However, this research can be taken forward by considering the proximity of location with tourist's spots, educational centres and further more criteria. Also, if the target audience are opting to purchase on property then we can consider including the property values in each neighbourhood for further analysis.

## Conclusion

---

The purpose of this project is to recommend a suitable business location for starting a restaurant in the neighborhoods of Toronto city, Ontario, Canada. All the required data were collection through web scraping method and Foursquare API for venue informations. The analysis was based on the three criteria, population of neighborhood (which is more essential to determine and evaluate the earning capability and opportunity available in that neighborhood, most common or resident's preferred venues (this is to understand the interest domain of the people living in each neighborhood, this surely helps us to understand the potential impact it may create after the establishment), clustering (for more accurate results and further options in neighborhoods enabling the target audience to choose from the same).

As discussed earlier, as per the outcome of this research, 'Riverdale' is most suitable for our business purpose. However, this research can be taken forward by considering the proximity of location with tourist's spots, educational centres and further more criteria. Also, if the target audience are opting to purchase on property then we can consider including the property values in each neighbourhood for further analysis.