

INDIAN RESTAURANTS PROBE

Exploring Toronto Neighborhoods - to open an Indian Restaurant
Web-Scraping, Foursquare API, Folium Map & Machine Learning

PREPARED BY-
KRUNAL PANCHAL (DATA ANALYST)

Introduction

Problem Statement: Prospects of opening an Indian Restaurant in Toronto, Canada.

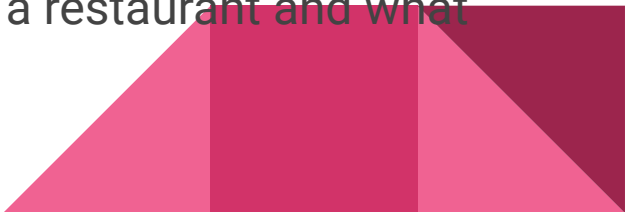
Toronto, the capital of the province of Ontario, is the most populous Canadian city. Its diversity is reflected in Toronto's ethnic neighbourhoods such as Chinatown, Corso Italia, Greektown, Kensington Market, Koreatown, Little India, Little Italy, Little Jamaica, Little Portugal & Roncesvalles. One of the most immigrant-friendly cities in North America with more than half of the entire Indian Canadian population residing in Toronto it is one of the best places to start an Indian restaurant.



In this project, we will proceed with the step by step analysis of the neighbourhoods of Toronto in order to find insights for the best suitable place to open an Indian Restaurant. Toronto shelter a greater number of Indians than any other city in Canada, so it is a good idea to start the restaurant here, but we just need to make sure whether it is a profitable idea or not. If so, where we can place it, so it yields more profit to the owner.

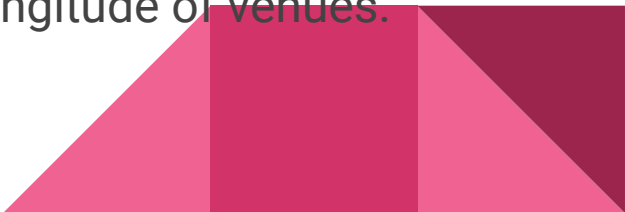


Target Audience

- Business personnel who wants to invest or open an Indian restaurant in Toronto. This analysis will be a comprehensive guide to start or expand restaurants targeting the Indian crowd.
 - Indian crowd who wants to find neighbourhoods with lots of option for Indian restaurants.
 - Data scientist or Business analyst who wants to explore and enhance their skills by finding insights from the given data in order to gain their storytelling skills as a part of the project.
 - Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.
- 


Data Overview

The data required for the analysis of the neighbourhoods of Toronto is a combination of different files collected from various sources. following are the lists of sources:

- the lists of postal codes of Canada are collected from Wikipedia "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M" which contains information regarding boroughs and neighbourhoods.
 - using geospatial data "https://cocl.us/Geospatial_data" for the geographical coordinates of the neighbourhoods.
 - using foursquare API to get the information about different venues in Toronto. The information contains name, category, latitude, longitude of venues.
- 

Methodology

Collecting data from different sources:

- Wikipedia-“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M” for information regarding boroughs and neighbourhoods of Canada and creating a data frame after performing EDA.
 - Geospatial data-“https://cocl.us/Geospatial_data” for the geographical coordinates of the neighbourhoods and merging it with the data frame for further analysis.
 - Foursquare API-to retrieve venues of Toronto which contains name, category, latitude and longitude information that is used to analyse the number of Indian Restaurants in each neighbourhood.
- 

	Borough	PostalCode	Neighborhood	Latitude	Longitude
0	Central Toronto	M4N	Lawrence Park	43.728020	-79.388790
1	Central Toronto	M4P	Davisville North	43.712751	-79.390197
2	Central Toronto	M4R	North Toronto West	43.715383	-79.405678
3	Central Toronto	M4S	Davisville	43.704324	-79.388790
4	Central Toronto	M4T	Moore Park, Summerhill East	43.689574	-79.383160

Fig.1



Fig.2

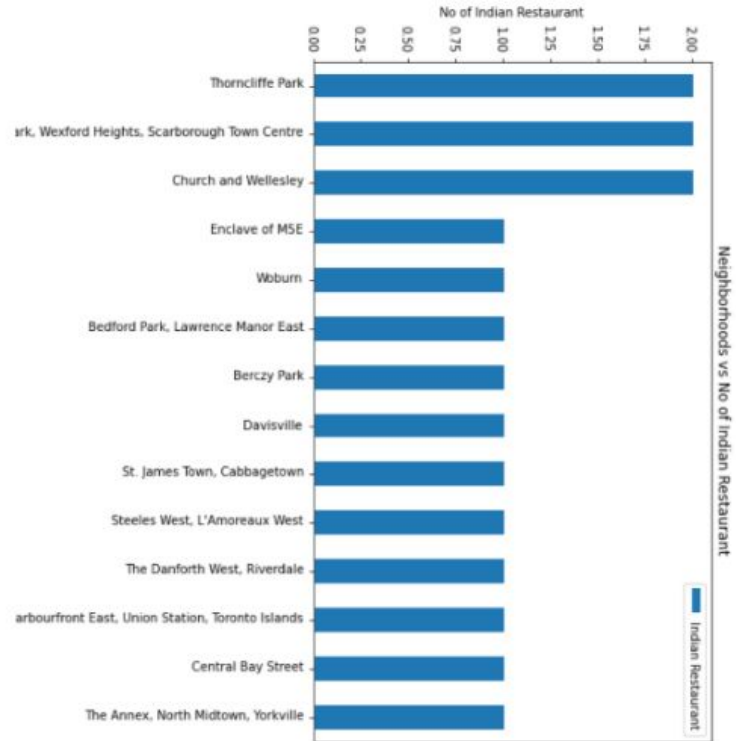
map containing all neighbourhood of Toronto

Analysis

After all the relevant data was collected, the Grouping of data was done on the venue dataframe to know the number of places present in each neighborhood and stored in a dataframe.

	Neighborhood	Indian Restaurant
0	Agincourt	0.00
1	Alderwood, Long Branch	0.00
2	Bathurst Manor, Wilson Heights, Downsview North	0.00
3	Bayview Village	0.00
4	Bedford Park, Lawrence Manor East	0.04
...
94	Willowdale West	0.00
95	Willowdale, Newtonbrook	0.00
96	Woburn	0.25
97	Woodbine Heights	0.00
98	York Mills West	0.00

99 rows × 2 columns



Barplot: neighbourhood vs number of India Restaurant

Examine clusters

Cluster Map consisting each neighbourhood in its respective cluster



Cluster 0 (red color)-

only 1 neighbourhood is present.

	Borough	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Indian Restaurant
34	East York	M4H	Thorncliffe Park	43.705369	-79.349372	0.0	0.1

Cluster 1 (purple color) -

no restaurant are present in this cluster of neighbourhood.

Borough	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Indian Restaurant
Central Toronto	M4N	Lawrence Park	43.728020	-79.388790	1.0	0.0
Central Toronto	M4P	Davisville North	43.712751	-79.390197	1.0	0.0
Central Toronto	M4R	North Toronto West	43.715383	-79.405678	1.0	0.0
Central Toronto	M4V	Summerhill West, Rathnelly, South Hill, Forest...	43.686412	-79.400049	1.0	0.0
Central Toronto	M5N	Roselawn	43.711695	-79.416936	1.0	0.0
...
York	M6C	Humewood-Cedarvale	43.693781	-79.428191	1.0	0.0
York	M6E	Caledonia-Fairbanks	43.689026	-79.453512	1.0	0.0
York	M6M	Del Ray, Mount Dennis, Keelsdale and Silverthorn	43.691116	-79.476013	1.0	0.0
York	M6N	Runnymede, The Junction North	43.673185	-79.487262	1.0	0.0
York	M9N	Weston	43.706876	-79.518188	1.0	0.0

rows x 7 columns

Cluster 2 (blue color)-

3 neighbourhood with greater number of restaurant.

	Borough	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Indian Restaurant
82	Scarborough	M1P	Dorset Park, Wexford Heights, Scarborough Town...	43.75741	-79.273304	2.0	0.333333

Cluster 3 (skyblue color) -
one neighbourhood with good number of restaurants.

	Borough	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Indian Restaurant
75	Scarborough	M1G	Woburn	43.770992	-79.216917	3.0	0.25



Cluster 4(green color) -
neighbourhood with medium populated restaurants.

	Borough	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Indian Restaurant
3	Central Toronto	M4S	Davisville	43.704324	-79.388790	4.0	0.040000
7	Central Toronto	M5R	The Annex, North Midtown, Yorkville	43.672710	-79.405678	4.0	0.055556
65	North York	M5M	Bedford Park, Lawrence Manor East	43.733283	-79.419750	4.0	0.040000
87	Scarborough	M1W	Steeles West, L'Amoreaux West	43.799525	-79.318389	4.0	0.062500

Cluster 5 (orange color)-
neighbourhood with 1-3 number of restaurants.

	Borough	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Indian Restaurant
9	Downtown Toronto	M4X	St. James Town, Cabbagetown	43.667967	-79.367675	5.0	0.026316
10	Downtown Toronto	M4Y	Church and Wellesley	43.665860	-79.383160	5.0	0.028986
14	Downtown Toronto	M5E	Berczy Park	43.644771	-79.373306	5.0	0.021739
15	Downtown Toronto	M5G	Central Bay Street	43.657952	-79.387383	5.0	0.016667
25	Downtown Toronto Stn A	M5W	Enclave of M5E	43.646435	-79.374846	5.0	0.011765
27	East Toronto	M4K	The Danforth West, Riverdale	43.679557	-79.352188	5.0	0.023810

Results

After a complete analysis of the neighbourhood of Toronto for the opening of Indian Restaurants as a part of business problem using various data and applying algorithms to find the best suitable place or neighbourhood for restaurant, the following are the results :

- from the above plot, we can see out of all the neighbourhoods only Thorncliffe Park, Dorset Park, Wexford Heights, Scarborough, Church and Wellesley has the highest number of Indian Restaurants with higher population density.
- after observing the data, opening a restaurant in Scarborough would be an appropriate place as seen from the above clusters.



Conclusion

To conclude this project, the insights for the given business problem was concluded. Many python libraries were used to fetch, manipulate the contents and analyze and visualize the datasets. The Foursquare API helps to get the venues in the neighbourhoods of Toronto. Machine learning Algorithms and techniques were used to predict the output of given data and for visualising folium library is used. A data scientist can use a similar procedure to predict some different outputs like the opening of malls in nearby areas. Similarly, we can analyze the data for different purposes such as the opening of cuisine or gym. Finally, this project was completed as a real-life problem with useful outputs and challenges using data science

