

1 DATA

Based on the problem statement. The factors which will influence the final decision are Indian restaurant in neighbourhood and south Indian population in the area.

1.1 DATA SOURCE

Following data source will be needed to extract the required information

1. Neighbourhood of Toronto
 - a. List of Postal code of Canada: M"
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
2. Geographical coordinates of the neighbourhoods
 - a. https://cocl.us/Geospatial_data
3. Demographics of Toronto
 - a. Source: https://en.wikipedia.org/wiki/Demographics_of_Toronto
4. Number of existing Indian restaurants in the neighbourhood
 - a. Source: Foursquare API

1.2 DATA CLEANING

1. Scraping Toronto Neighbourhoods data from Wikipedia -
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Data frame will have 3 columns

Postcode Borough Neighbourhood

- Borough having value will be process. "Not assigned" Borough will be deleted from data frame.
- If there is no Neighbourhood for the Borough, then Neighbourhood will be same as Borough

	Postcode	Borough	Neighbourhood
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

2. Adding geographical coordinate to the neighbourhood

	Postcode	Borough	Neighbourhood		Latitude	Longitude
0	M3A		North York	Parkwoods	43.753259	-79.329656
1	M4A		North York	Victoria Village	43.725882	-79.315572
2	M5A		Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636
3	M6A		North York	Lawrence Manor , Lawrence Heights	43.718518	-79.464763

3. Scraping demographics of Toronto in order to obtain the data of visible minorities.

Riding		EthnicOrigin	EthnicPop	Minority	MinorityPop
0	Parkdale-High Park	English	26.2	Black	5.3
1	Etobicoke Centre	Italian	27.1	South Asian	5.9
2	Etobicoke-Lakeshore	English	27.7	South Asian	5.5
3	Toronto-St. Paul's	English	28.3	Black	5.1
4	Eglinton-Lawrence	Canadian	31.8	Filipino	10.7
5	Davenport	Portuguese	31.9	Black	6.4
6	University-Rosedale	English	32.7	Chinese	14
7	Toronto-Danforth	English	32.8	Chinese	12.3
8	Beaches-East York	English	33.6	South Asian	10.9
9	Don Valley West	English	41.5	South Asian	13.3
10	Spadina-Fort York	English	42.8	Chinese	14.8
11	York Centre	Filipino	46.5	Filipino	16.5
12	Toronto Centre	English	49.7	South Asian	11.8
13	York South-Weston	Portuguese	54.9	Black	23.2
14	Scarborough Southwest	Canadian	56.5	South Asian	21.6
15	Don Valley East	East Indian	58.4	South Asian	17.1

4. Get venue data for neighbourhood using Foursquare API

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant