Introduction:

As one of the most culturally diverse cities on the planet, Toronto evades categorization. While restaurants in other cosmopolitan cities like Paris or Tokyo have iconic, immediately identifiable cuisines, Canada's largest city manages to remain a moving target.

Indian restaurants have become so popular in Toronto now it seems that there is one on every corner. Starting an Indian restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

Business Problem :

Our client wants to open his business in Toronto area, so I focus on that borough during my analysis. We define potential neighborhood based on the number of Indian restaurants which are operating right in each neighborhood. Toronto has full potential but also is a very challenging place to open a business because of high competition. New restaurants should be open in an area that inadequate neighborhood in this way the bar can attract more customers. Therefore, this analysis necessary to ensure that we have enough Indian community and that we are not so close to other Indian restaurants.

Data Collection and cleaning:

Data Sources:

- The first step in the analysis was to find all the boroughs in Toronto. This data could be obtained from a wiki page,

("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")

- The geographical coordinate for the boroughs of Toronto are found from the following link

('https://cocl.us/Geospatial_data')

- Neighbourhoods details are found using Foursquare API .
- Using this API we will get all the venues in each neighbourhood. We can filter these venues to get only Indian restaurants.

('https://api.foursquare.com/v2/venues/explore?)

Data Cleaning

All borough names are extracted from Wiki using web scraping by utilizing pandas HTML table scraping method as it is easier to pull tabular data directly from a web page into the data frame

	Postal code	Borough	Neighborhood
0	M1B	Scarborough	Malvern / Rouge
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek
2	M1E	Scarborough	Guildwood / Morningside / West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

- Tried getting geographical coordinate for the boroughs of Toronto using Geocoder Package but it was not working so used the CSV file provided by IBM team to match the coordinates of Toronto neighbourhoods

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
1	M2H	North York	Hillcrest Village	43.803762	-79.363452
2	M4B	East York	Parkview Hill / Woodbine Gardens	43.706397	-79.309937
3	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
4	M4G	East York	Leaside	43.709060	-79.363452
5	M4M	East Toronto	Studio District	43.659526	-79.340923

Finally Using Foursquare Location Data, the 100 most popular venues in a radius of 500m for each district is Toronto is obtained. The data obtained is a JSON _le, and we need to turn that into a data-frame. This final dataset will contain:

	PostalCode	Borough	Neighborhood	BoroughLatitude	BoroughLongitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	Downtown Toronto	43.653232	-79.385296	Neighborhood
1	M4E	East Toronto	The Beaches	43.676357	-79.293031	Textile Museum of Canada	43.654396	-79.386500	Art Museum
2	M4E	East Toronto	The Beaches	43.676357	-79.293031	Cafe Plenty	43.654571	-79.389450	Café
3	M4E	East Toronto	The Beaches	43.676357	-79.293031	Sansotei Ramen 三草亭	43.655157	-79.386501	Ramen Restaurant
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	Japango	43.655268	-79.385165	Sushi Restaurant

Methodology:

Exploratory Data Analysis

In this section, we will analysis the data frames built in the above section. And come up with the analysis on the best neighborhood for Indian restaurants

Analyze the venues in Toronto

Check how many venues for each neighborhood

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	55	55	55	55	55	55
Brockton / Parkdale Village / Exhibition Place	24	24	24	24	24	24
Business reply mail Processing CentrE	16	16	16	16	16	16
CN Tower / King and Spadina / Railway Lands / Harbourfront West / Bathurst Quay / South Niagara / Island airport	16	16	16	16	16	16
Central Bay Street	60	60	60	60	60	60
Christie	16	16	16	16	16	16

Group rows by neighborhood and mean of venue

	Neighborhoods	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	...
0	Berczy Park	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	...
1	Brockton / Parkdale Village / Exhibition Place	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	...
2	Business reply mail Processing CentrE	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	...
3	CN Tower / King and Spadina / Railway Lands / ...	0.0625	0.0625	0.0625	0.125	0.1875	0.125	0.000000	0.000000	0.00	...
4	Central Bay Street	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	...
5	Christie	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	...

Identify the Top 10 neighborhoods with Indian restaurant

	Neighborhoods	Indian Restaurant
0	Berczy Park	0.000000
1	Brockton / Parkdale Village / Exhibition Place	0.000000
2	Business reply mail Processing CentrE	0.000000
3	CN Tower / King and Spadina / Railway Lands / ...	0.000000
4	Central Bay Street	0.016667
5	Christie	0.000000
6	Church and Wellesley	0.012987

With this data, I analyzed each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. I made a justification to explicitly look for "Indian restaurants". finally, I performed the clustering method by using k-means clustering. K-means clustering algorithm categorizes k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible.