

INTRODUCTION

THIS REPORT IS INTENDED TO PROVIDE AN INSIGHT ON "HOW DATA SCIENCE CAN BE USED TO SOLVE A BUSINESS PROBLEM?"



BUSINESS SCENARIO BACKGROUND

This report will try to address the following question for any entrepreneur who is planning to invest in opening a new Indian restaurant in Toronto city.

- How many Indian restaurants are already present in Toronto City?
- Which neighbourhoods in Toronto city has an Indian restaurant?
- Which neighbourhoods in Toronto city will be ideal to start a new India Restaurant?
- What are the descriptive highlights of these ideal neighbourhoods?



MSA North York Parkosocia MAA North York Victora Village MSA Oomtown Toront Regent Park, Harbourhront MSA North York Lawrence Manor, Lawrence Heights MTA Domitown Toronto Cusenip Park, Ortanio Provincial Government MSA Not assigned Not assigned MSA Elobicicke Islington Avenue, Hamber Valley Village MSB Scarborough Malvern, Rouge MSB Not assigned Not assigned MSB Not assigned MSB Not seigned MSB Not seigned Not

MILA

Not assigned

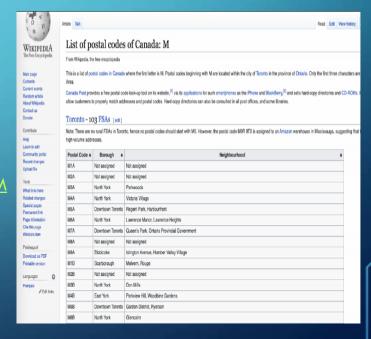
DATA DESCRIPTION AND ACQUISITION

1. Postal Codes of Canada – Web Scraping

A list of all the postal codes with assigned Boroughs and neighbourhoods is available on a wiki page.

https://en.wikipedia.org/wiki/List of postal codes of Canada: M

Web Scraping - Using pandas.read html()



DATA DESCRIPTION AND ACQUISITION

- 2. Geographical coordinates of the postal codes Web Scraping
- This is the main input for acquiring other venue related data.
 - This data can be extracted using
 - Geocoder Python package
 - > Using a direct link to a csv file http://cod.us/Geospatial date



Boroughs of Toronto plotted on Map using Folium library.

- 3. Venues data of Toronto City Foursquare API
- The venue data details all types of venues, their category and geocoordinates for any desired location.
- For this study, the venue data was extracted using the Foursquare API.
- The venue data consists of the following as shown in figure below.
 - Venue Name
 - Venue Category
 - Venue Latitude
 - Venue Longitude

	name	categories	lat	Ing
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008
3	Impact Kitchen	Restaurant	43.656369	-79.356980
4	Body Blitz Spa East	Spa	43.654735	-79,359874

DATA DESCRIPTION AND ACQUISITION

- 4. Final dataset for analysis
- The datasets 1-3 were merged together to create one data frame.
- the data frame had to be transformed to represent the venues in terms of the neighbourhoods.
- the rows with same neighbourhood names were grouped together.
- One hot encoding was used for the same.

	Neighbourhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Berczy Park	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0	0.018182	0.000000	0.0	0.0
1	Brockton, Parkdale Village, Exhibition Place	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
2	Business reply mail Processing Centre, South C	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
3	CN Tower, King and Spadina, Railway Lands, Har	0.0	0.0625	0.0625	0.0625	0.125	0.125	0.0625	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
4	Central Bay Street	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0	0.000000	0.014706	0.0	0.0

METHODOLOGY

A. EXPLORATORY DATA ANALYSIS

For the exploratory data analysis the following was done.

A1. TOTAL VENUE CATEGORIES.

The number of 'to-do' options available to individual living in Toronto.

A2. TOP 10 VENUES

The top 10 venues by their occurrence in the neighbourhood.

are any Indian restaurant in top 10 category in any neighbourhood?

A3. INDIAN RESTAURANTS IN TORONTO.

The total number of Indian restaurants and their frequency of occurrence in each of the neighbourhoods.

	Neighbaurhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighbourhood						
Barczy Park	11	. 10	. 15	. 55	. 15	55
Breckton, Parkdale Village, Exhibition Place	23	29	. 25		29	23
Business reply mail Processing Centre, South Central Letter Processing Plant Security	16	76	16	36	16.	16
CN Tower, King and Speditss, Ballway Lands, Harbourboot West, Softwart Quest, South Niegars, Inland sirpart	16	16	16	10	16	16
Central Bay Street	58	66	. 68	68	66	64

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	Srd Most Common Venue	4th Meet Common Versus	Sth Most Common Venue	Oth Most Common Venue	7th Most Common Venue	Oth Mest Common Worse	Oth Mest Common Venue	10th Most Common Venue
0	Securitus.	Caffee Ship	Sedicid Redsylett	Cockled Bar	Farmers Market	ber far	Restaurant	Cheese Shop	towery	Sandwich Place	Separtment Store
1	Brockton, Parkitele Village, Exhibition Place	CWN	Seekfast Seek	Nagericle	Coffee Shop	Cinting Syn	Burnto Place	Restaurant	Restaurant	Intersection.	Ber
,	Business repty mail: Processing Centre, South C.	Part	Ficta Place	Light Rad Station	Skete Flori	Barto Pice	famers Market	Fact Food Restaurant	Butcher	Receiver	Recording Studio
1	CN Town, King and Spetina, Reliver Lands, Mar.	Alport Lounge	Airport Service	States	Harbor / Marina	Boot or Yeary	Rental Car Location	No.	Place	Catter Shop	Soulphure Garden
4	Central Bay Street	Cartes Shop	Carlo	Total Restoraci	Sentwich Place	Saled Place	Subble Tea Shop	Department Store	Burger Josep	Japanese Reiduzant	Thai Restaurant

	Neighbourhood	Venue
o	Central Bay Street	1
1	Church and Wellesley	7
2	Davisville	
33	Harbourfront East, Union Station, Toronto Islands	1
4	St. James Town, Cabbagetown	3
5	The Annex, North Midtown, Yarkville	1
6	The Danforth West, Riverdale	- 1

METHODOLOGY

B. INFERENTIAL STATISTICS - MACHINE LEARNING

I have used K-mean clustering method for clustering the neighbourhoods.

Number of clusters - 4

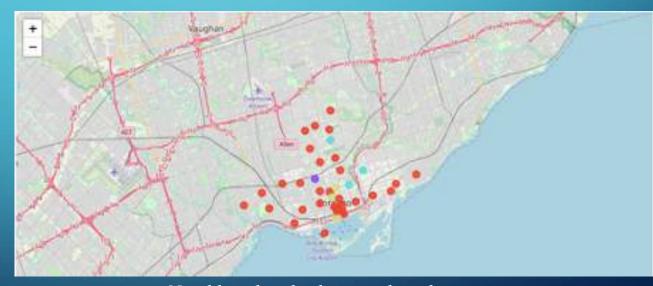
The following clusters were created.

1.Cluster 1 – Red

2.Cluster 2 – Purple

3. Cluster 3 – Aquamarine

4.Cluster 4 – Dark khaki'

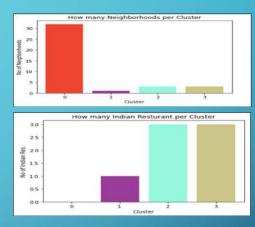


Neighbourhoods clusters plotted on a map

RESULTS

Following analysis were done to identify the best neighbourhoods for opening an Indian Restaurant.

- 1. Neighbourhood in clusters
- Cluster 0 has highest number ie 32 neighbourhood in it. Followed by cluster 2 and 3 with 3 neighbourhoods each in them and finally cluster 1 with just one neighbourhood in it.
- 2. Indian restaurants across clusters
 There are no Indian restaurants present in cluster 0. Also
 there are 3 Indian restaurants present in cluster 2 and 3 each
 and 1 in cluster 1.



Cluster 0 - Red	Cluster 1 - Purple
 Has no Indian restaurants in it. Spread across all 4 Borough 32 Neighbourhoods. Ideal cluster 	 Present in only 1 borough Only 2 neighbourhoods in this cluster 1 Indian Restaurant already present in it Not an ideal choice
Cluster 2 - Aquamarine	Cluster 3 - Dark kaki'
 Present in only 3 borough Spread across 5 neighbourhoods 3 Indian Restaurant already present. Not an ideal choice 	 Present in only 3 borough Spread across 5 neighbourhoods 3 Indian Restaurant already present. Not an ideal choice

DISCUSSION

- Clearly cluster 1, 2, 3 are not our preferred choices to start the Indian restaurant
- Cluster 0 seems to be suitable.
- Within cluster 0 there are few neighbourhoods which have very common venues like Parks and bars.
- This would give us a good sense of which neighbourhood would be a better choice.

	Borough	n Neighbourhood	1st Most Common Venue	n Common	Common	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	Common	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
15	East Toronto	구	Park	k Gym	Sushi Restaurant	Sandwich Place	Liquor Store	Burrito Place	Restaurant	Italian Restaurant	Fast Food Restaurant	Fish & Chips Shop
18	Central Toronto	Lawrence Park	Park	k Swim School		Escape Room	Electronics Store	Eastern European Restaurant	Dumpling	Donut Shop	Doner Restaurant	Dog Run
20	Central Toronto		Park	k Hotel	Breakfast Spot	Sandwich Place	Food & Drink Shop	Dog Run	Dance Studio	Department Store	Gym / Fitness Center	Cosmetics Shop
21	Central Toronto		Park	Sushi Restaurant		Trail	Escape Room	Electronics Store	HURODOAD	Dumpling Restaurant	Donut Shop	Doner Restaurant
33	Downtown Toronto	Rosedale	Park	k Trail	Playground	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling	Donut Shop	Doner Restaurant	Dog Run
	Borough	Neighbourhood Co			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
11	West Toronto	Little Portugal, Trinity	Bar	= = 100	ietnamese Restaurant	Restaurant	Café	Asian N Restaurant	Men's Store	Cuban Restaurant	Record Shop	Pizza Place

CONCLUSION

Only cluster 0 spread across 32 neighbourhoods and with no existing Indian Restaurant present in it is the ideal choice to start a new Indian Restaurant.

Potential neighbourhoods with lots of parks and bars around

- 1. India Bazaar
- 2. The Beaches West
- 3. Little Portugal
- 4. Trinity

This Analysis can be made more robust by adding demographic data. This will create clusters with more detailed features and will help us making more informed decisions.

THANK YOU

Find me on LinkedIn and the codes for this project on GitHub