

Exploring Toronto Neighbourhood

_IBM Capstone report

- By Vasu Khanna

Introduction:

A investors life is full of risks . Too lessen the risks in his life and make investments a bit more risk free in Toronto. This notebooks aims to do that. This may also help people who set up advertisement, political or other campaigns. This tells about what people of all neighbourhoods visit most often in a particular neighbourhood.

Problems being solved:

- Tries to make life of investor better
- Tells more about a particular neighbourhood
- Helps people to decide where to buy a house
- Helps people set up effective advertising campaigns

DATA

Data source:

web scraping data from wikipedia about toronto neighbourhood. Also using Foresquare api to get most visited places in a particular neighbourhood

Data cleaning:

Using beautiful soup to scrape data from wikipedia, then making a dataframe of it, adding latitude and longitude to dataframe. Using Foresquare api to get information of top 10 trending places in the neighbourhood. Putting in data frame and removing things that are not required

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Malvern	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
3	Rouge	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
4	Rouge Hill	43.784535	-79.160497	Chris Effects Painting	43.784343	-79.163742	Construction & Landscaping

Feature selection:

	Neighborhood	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	St
0	Adelaide	0.0	0.000000	0.0000	0.0000	0.000	0.000	0.000	0.033333	0.0	...	0.00
1	Agincourt	0.0	0.000000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.0	...	0.00
2	Agincourt North	0.0	0.000000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.0	...	0.00
3	Albion Gardens	0.0	0.000000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.0	...	0.00
4	Alderwood	0.0	0.000000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.0	...	0.00
5	Bathurst Manor	0.0	0.000000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.0	...	0.00
6	Bathurst Quay	0.0	0.062500	0.0625	0.0625	0.125	0.125	0.125	0.000000	0.0	...	0.00

We use one hot encoding technique to convert the venue category to numbers. We then see the frequency of a category in each neighbourhood, in a dataframe.

Exploring data:

Make a dataframe with top 10 most visited categories of places in each neighbourhood

	Neighborhood	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
0	Adelaide	Steakhouse	Café	Pizza Place	Asian Restaurant	Hotel	Concert Hall	Monument / Landmark	Plaza	Food Court	Lounge
1	Agincourt	Lounge	Breakfast Spot	Chinese Restaurant	Sandwich Place	Dance Studio	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run
2	Agincourt North	Park	Playground	Asian Restaurant	Curling Ice	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Discount Store
3	Albion Gardens	Grocery Store	Liquor Store	Pharmacy	Fast Food Restaurant	Fried Chicken Joint	Sandwich Place	Pizza Place	Beer Store	Video Store	Coffee Shop
4	Alderwood	Pizza Place	Gym	Skating Rink	Pharmacy	Pool	Pub	Dance Studio	Sandwich Place	Coffee Shop	Fruit & Vegetable Store

Predictive Modelling:

Using k-means clustering technique , clustering neighbourhoods of toronto based on the categories of different venues as features. Then plotting it on a map using folium.

Conclusion:

The top categories can clearly be seen in the neighbourhood and the map highlights it too. Based on the investors interest of business or place he can choose a neighbourhood.

