



Toronto city explorer

Project report

saksham kulshrestha

Introduction/ business problem

A lot of potential thoughts and considerations are given to the locations whenever we visit a city. What if we could look at specific important/busy areas in the city or place of interest of a city we are visiting. In that case what if we could cluster neighbourhoods on the basis of the potential tourism interests. It would help travellers, tourism agencies and real estate investors too.

Explanation

Toronto is a fast growing financial hub. With increasing city centres and sparkling residential hubs. This makes tourism and real estate one of many growing business ventures Toronto has had. The city is well connected and well planned and is financially flourishing and that aspect brings in a lot of travellers. Hence determining interests and hotspots is definitely beneficial and aids new tourists and investors

Who are our stakeholders?

Anyone who's interested in the region, it can be individuals, tourists or even the residents who would want to know if they are in the right neighbourhood

Data

Data used in this project is retrieved from wiki data and foursquare API data Processed under a valid foursquare developer account, geospatial data was also retrieved for the purpose of mapping and geo-visualization.

Methodology

We explore and analyse data using pandas (python) in a jupyter terminal the insights were processed and found through a clustering analysis and the findings were depicted using folium in a geospatial mapped environment

Data/exploration

The data corresponding to the city was

	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	[Rouge, Malvern]	43.806686	-79.194353
1	M1C	Scarborough	[Highland Creek, Rouge Hill, Port Union]	43.784535	-79.160497
2	M1E	Scarborough	[Guildwood, Morningside, West Hill]	43.763573	-79.188711
3	M1G	Scarborough	[Woburn]	43.770992	-79.216917
4	M1H	Scarborough	[Cedarbrae]	43.773136	-79.239476

The venue data by foursquare is as follows

[Rouge, Malvern]
[Highland Creek, Rouge Hill, Port Union]
[Guildwood, Morningside, West Hill]
[Woburn]
[Cedarbrae]
[Scarborough Village]
[East Birchmount Park, Ionview, Kennedy Park]
[Clairlea, Golden Mile, Oakridge]
[Cliffcrest, Cliffside, Scarborough Village West]
[Birch Cliff, Cliffside West]
[Dorset Park, Scarborough Town Centre, Wexford Heights]
[Maryvale, Wexford]
[Agincourt]
[Clarks Corners, Sullivan, Tam O'Shanter]
[Agincourt North, L'Amoreaux East, Milliken, Steeles East]
[L'Amoreaux West]
[Upper Rouge]
[Hillcrest Village]
[Fairview, Henry Farm, Oriole]
[Bayview Village]
[Silver Hills, York Mills]
[Newtonbrook, Willowdale]
[Willowdale South]
[York Mills West]
[Willowdale West]
[Parkwoods]
[Don Mills North]
[Flemingdon Park, Don Mills South]

We processed all the above data and explored various cumulative datasets

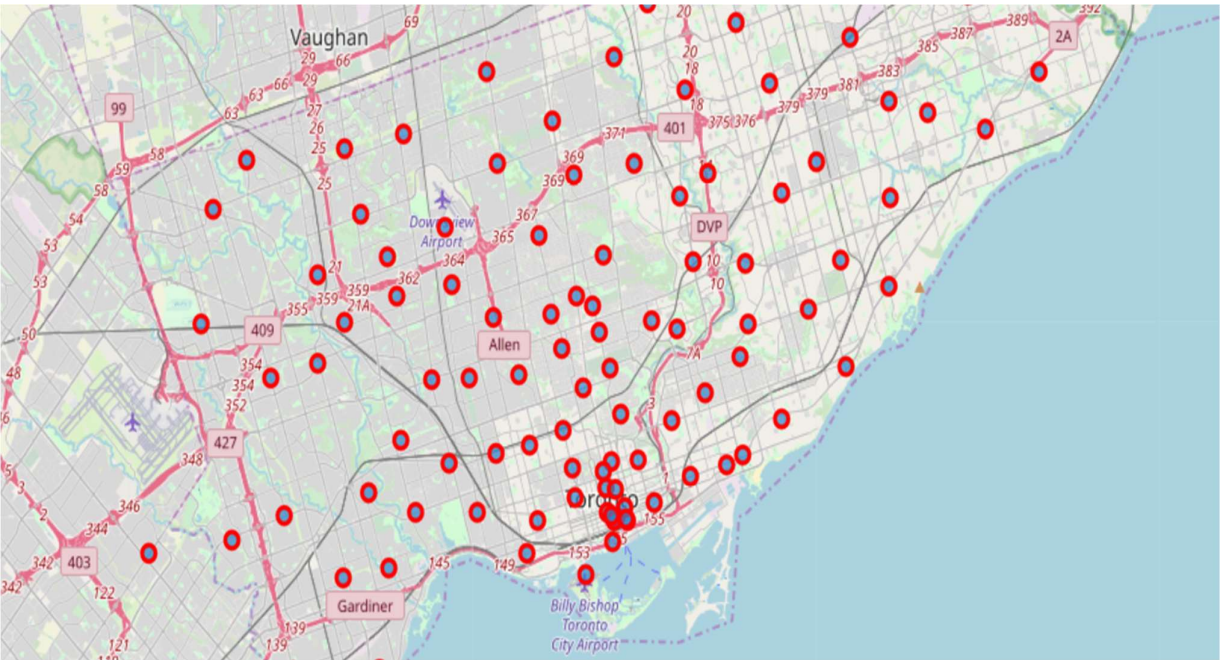
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	[Rouge, Malvern]	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	[Rouge, Malvern]	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	[Highland Creek, Rouge Hill, Port Union]	43.784535	-79.160497	Chris Effects Painting	43.784343	-79.163742	Construction & Landscaping
3	[Highland Creek, Rouge Hill, Port Union]	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
4	[Guildwood, Morningside, West Hill]	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
5	[Guildwood, Morningside, West Hill]	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
6	[Guildwood, Morningside, West Hill]	43.763573	-79.188711	Marina Spa	43.766000	-79.191000	Spa
7	[Guildwood, Morningside, West Hill]	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
8	[Guildwood, Morningside, West Hill]	43.763573	-79.188711	chatr Mobile	43.765917	-79.191672	Tech Startup

We also assorted the same data with respect to our needs

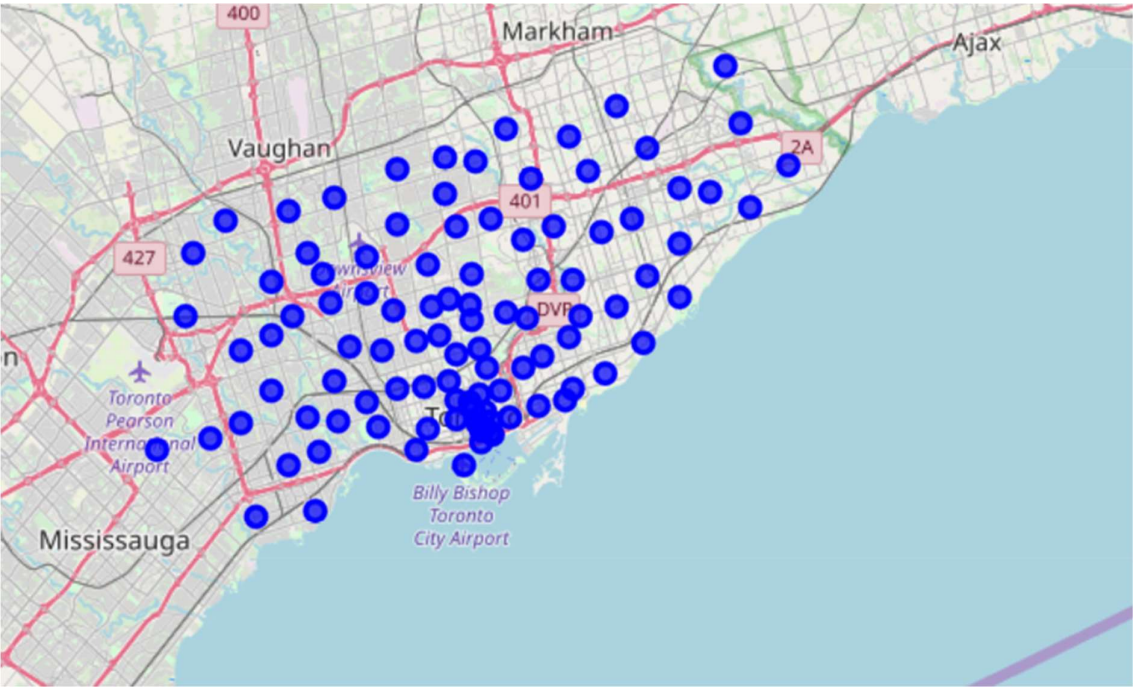
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	[Adelaide, King, Richmond]	Café	Steakhouse	Asian Restaurant	Pizza Place	Hotel	Greek Restaurant	Opera House	Monument / Landmark	Concert Hall	Plaza
1	[Agincourt North, L'Amoreaux East, Milliken, S...]	Park	Asian Restaurant	Playground	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Discount Store	Diner	Dim Sum Restaurant
2	[Agincourt]	Lounge	Chinese Restaurant	Breakfast Spot	Sandwich Place	Curling Ice	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Dog Run	Discount Store
3	[Albion Gardens, Beaumont Heights, Humbertgate,...]	Grocery Store	Coffee Shop	Fast Food Restaurant	Beer Store	Video Store	Liquor Store	Pharmacy	Pizza Place	Sandwich Place	Fried Chicken Joint
4	[Aldenwood, Long Branch]	Pizza Place	Pool	Coffee Shop	Skating Rink	Gym	Pharmacy	Pub	Dance Studio	Sandwich Place	Women's Store

Results

And we found the following results when we visualized the above data by folium



Clustered the venues by their corresponding neighbourhoods



Discussion

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into cells.

Conclusion

The data related to problem was analysed and provided insights were followed and observed using required visualisation techniques