BATTLE OF NEIGHBORHOOD – TORRONTO

1. Introduction:

Someone who loves to research on places before visiting them, the idea of finding best of Toronto is exciting. As a new visitor it is difficult to find out best places to visit in the restricted time period and also to find something which really suits their liking. Someone might just be happy looking at a sunset, while someone else might seek adventure. There are people who find calm in books while others want to see castles from their fairy tales. Foodies to drinkers, meditators to explorers everyone is a stakeholder to this problem. Every person who visits a new location faces this problem to find suitable top rated locations of their liking.

Most of the websites provide a list of top 10 or top 20 places to visit in a specific city, which is difficult to conclude if these lists would fulfil a user's need to find places which the locals like the most and would also suit his/her taste. Hence as part of this Capstone project I tried solving this problem by finding various venue options available to a tourist based on his or her liking. And further, suggested which neighbourhood holds most of the attractions which users might find fascinating and enable them book a hotel in, thus saving travel expenses and yet be able to visit a wide group of locations.

Using this Project I tried to segregate different neighbourhoods to find the one which has everything for a particular type of traveller.

2. Data

To solve the problem faced by travelers, I have used a sample data from a website with neighborhood details of Toronto. This data contains postal code, borough names and neighborhoods in Toronto and added all this data to a data frame using Beautifulsoap API.

	Postal_Code	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood\n, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park\n, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West\n
9	M1N	Scarborough	Birch Cliff, Cliffside West\n
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...

After creating data frame, I gathered the Latitude and Longitude details of the neighborhoods.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Downtown Toronto	Harbour front	43.640080	-79.380150
1	M5A	Downtown Toronto	Regent Park	43.660706	-79.360457
2	M5B	Downtown Toronto	Ryerson	43.621573	-79.559130
3	M5B	Downtown Toronto	Garden District	43.656502	-79.377128
4	M5C	Downtown Toronto	St. James Town	43.669403	-79.372704
5	M4E	East Toronto	The Beaches	43.671024	-79.296712
6	M5E	Downtown Toronto	Berczy Park	43.648001	-79.375385
7	M6G	Downtown Toronto	Christie	43.664111	-79.418405
8	M5H	Downtown Toronto	Adelaide	43.650298	-79.380477
9	M5H	Downtown Toronto	King	43.648949	-79.377754
10	M5H	Downtown Toronto	Richmond	43.655443	-79.355780

Further, I have used foursquare API to find out the popular venues around Toronto within 500m of the neighborhoods. The list of venues and categories of these venues are mapped in each category to a higher level split the traveler might be interested in.

Identified categories:

Food	Art	Nature	Adventure	History	Outdoors

To divide into these categories I found the unique categories and manually found the keywords which we usually find in categories of these venues which helped me divide the venues.

NEIGHBORHOOD			VENUE				
Place	Latitude	Longitude	Venue	Latitude	Longitude	Category	Summary
Harbour front	43.64008	-79.38015	Lake Ontario	43.639398	-79.379589	Lake	This spot is popular
Harbour front	43.64008	-79.38015	Harbour Square Park	43.639253	-79.378395	Park	This spot is popular
Harbour front	43.64008	-79.38015	Harbour front	43.639526	-79.380688	Neighbour-hood	This spot is popular
Harbour front	43.64008	-79.38015	Beaver Tails	43.639899	-79.380197	Bakery	This spot is popular
Harbour Front	43.64008	-79.38015	Sharetea-Waterfront	43.640176	-79.379606	Bubble Tea Shop	This spot is popular

A major chunk of venues fall into category – food. Hence it cannot be considered a dominant factor while forming clusters. We can probably look into neighbourhoods which have a higher number of various attractions and suggest user based on his interest which neighbourhood can be a better option for him / her to get a hotel.

Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Summary	Venue ID	My Categories
Harbour front	43.64008	-79.38015	Lake Ontario	43.639398	-79.379589	Lake	This spot is popular	4d07f8041657a35d19272ae7	Nature
Harbour front	43.64008	-79.38015	Harbour Square Park	43.639253	-79.378395	Park	This spot is popular	4e49413e81dc766f3e3d6312	Nature
Harbour front	43.64008	-79.38015	Harbourfront	43.639526	-79.380688	Hike	This spot is popular	4bfaa3494a67c928d08528cf	Adventure
Harbour front	43.64008	-79.38015	BeaverTails	43.639899	-79.380197	Bakery	This spot is popular	55a19437498eeea53fa58b54	Food
Harbour front	43.64008	-79.38015	Sharetea-Waterfront	43.640176	-79.379606	Bubble Tea Shop	This spot is popular	57b22076498efeeeb93cfc47	Food

Final Dataset:

Postal Code	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
M5A	Downtown Toronto	Harbourfront	43.640080	-79.380150	2	art	nature	adventure	outdoor	history
M5A	Downtown Toronto	Regent Park	43.660706	-79.360457	1	art	outdoor	nature	history	adventure
M5B	Downtown Toronto	Ryerson	43.621573	-79.559130	2	art	adventure	outdoor	nature	history
M5B	Downtown Toronto	Garden District	43.656502	-79.377128	1	art	adventure	outdoor	nature	history
M5C	Downtown Toronto	St. James Town	43.669403	-79.372704	1	outdoor	nature	history	art	adventure

3. Methodology

To explore the data, the counts and proportions of the category types were checked. The most striking point is food not being a prominent type. Hence, removed it from the main data and then formed clusters using K Means.

4. Conclusions:

I was able to form 4 clusters which suggest user specific neighborhood to stay during their travel.

- Nature explorers and Adventurers
- Neighborhoods seem best for Art and Nature lovers.
- Neighborhoods seem best for someone who loves to stay outdoors be it Nature or Adventure
- Best for art Nature and Adventure.

The choice remains with the Traveler to choose what fits the best to his choice. But this clustering would give idea to him / her to where to start from.

5. Discussions:

I have been able to identify something clusters of neighbourhoods with details of top venues. This can further be taken up to suggest an itinerary, which could be a next part of our next quest.