Coursera Capstone - The Battles of Neighborhoods

Data Description

Data Description:

As we need to explore, segment, and cluster the neighborhoods in the city of Toronto, the Toronto neighborhoods data is key for this project. Unfortunately, the data is for the Toronto neighborhood data is not widely available on the Internet in the structured format, hence we need to scrap it through an existing Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. The data should contain the coordinates for each of the neighborhood in Toronto that will help us to further obtaining more information critical for this project. We will also like to obtain the key information like below; such as number of residences information for each neighborhood which is one of the key factors for the neighborhood of choice in the final recommendation report. The data needs to be clean up and eventually in a structured format.

1. Neighborhood Name
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Number of residences in each neighborhood.

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.8067	-79.1944
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.7845	-79.1605
2	M1E	Scarborough	Guildwood]], Morningside, West Hill	43.7636	-79.1887
3	M1G	Scarborough	Woburn	43.771	-79.2169
4	M1H	Scarborough	Cedarbrae	43.7731	-79.2395
5	M1J	Scarborough	Scarborough Village	43.7447	-79.2395
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.7279	-79.262
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.6915	-79.2866
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.7163	-79.2395
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.6927	-79.2648

Data Features:

We will be leveraging on features in a reliable location information provider such as the Foursquare.com to explore the various types of venues and its categories available in each neighborhood. We will also need to understand the trending of these venues in the respective neighborhood. The information obtained per neighborhood will be as such:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude

Conclusion

We will also use clustering techniques such as the K-Clustering to segment and cluster these neighborhoods so that we can group them together to understand their similarities and what best we can do for these types of neighborhoods.

With all these features, techniques and data, we will then be able to come up with a best recommendation to the management of ABC Online Grocery to their problem which is where is the best neighborhood for them to first start off to offer their services. For an example, we will not want to enter a neighborhood whereby there is already a high concentration of grocery stores available or there is a high trending of such stores upcoming in the neighborhood; we will like to recommend a neighborhood where we know that there will be a higher demand of such delivery service due to the lack of supply in that area.