Location Data

Location in examination

Toronto . The neighborhood data of Toronto will be obtained via the scraping method that was introduced in week 3's assignment .

Explanation & discussion

For my case , instead of just narrowing down the neighborhood to those which has 'Toronto' in it , I will be utilizing the whole data set and not perform any filtering . This is to reflect the real life scenario where no suburbs will be missed out and the data user could have a detailed view of all neighborhood clusters in Toronto.

In order to effectively mark the location of all neighborhood , I will also be appending the scrapped dataset with their respective latitude and longitude , example as below :

	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

K-means clustering will then be applied onto the data sets where each suburbs will be marked with specific clusters on the map of Toronto.

The neighborhood will be clustered according to the venues that are in them. To obtain extra details regarding the venues , I will be using the EXPLORE endpoint as part of foursquare location API , where the result would look like this :

	Neighborhod	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aq
0	The Beaches	0	0	0	0	0	0	0	0	0	0	0
1	The Beaches	0	0	0	0	0	0	0	0	0	0	0
2	The Beaches	0	0	0	0	0	0	0	0	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	0	0
4	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	0	0