Exploring the cities on New York and Toronto

Raju Kalidindi

Background and Description of the problem: I am a software engineer working for a major bank in Minneapolis, Minnesota USA. I am currently looking for Data Engineer positions and have attended a few interviews from clients at New York City and Toronto. I currently live close to Minneapolis downtown and enjoy many amenities and venues in Minneapolis. As a regular practitioner of yoga and meditation, it is important for me to relocate to neighborhoods that have these facilities to help me to continue with my interests. I would like to relocate to a city, which has similar or better facilities than Minneapolis.

The purpose of this project is to help people who are planning to relocate to new cities, to use data science to compare venues in different cities or neighborhoods. By using the latitude and longitude coordinates and the Foursquare application we can map the venues available in different neighborhoods. Selecting the appropriate neighborhood which has venues of interest, of the person who is relocating, will help them to meet their desired lifestyle requirements.

Since I am close to getting offers from clients located in New York and Toronto, I would like to use the skills gained in this course to compare the different venues and amenities in New York and Toronto cities.

Description of the data that will be used to solve the problem: Data for New York city will be downloaded from https://cocl.us/new_york_dataset. Data for Toronto postal codes will be downloaded from https://data.mongabay.com/igapo/toronto_zip_codes.htm. For the venues and amenities in Toronto, we will utilize Foursquare application. We will use data wrangling and k-Means clustering to explore the data for the two cities.

Manhattan and Toronto datasets used for analysis: Sample of the datasets used for solving the business problem are shown below.

Manhattan Dataset

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210

Toronto Dataset:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806696	-79.194353

1	M1C	Scarborough	Highlan Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.7770992	-79.216917

The above datasets have the names of the neighborhoods of the above cities and the latitude and longitude coordinates.

We can use the above geographical coordinates and the Foursquare application to get data on different venues at the above locations. Using data science tools, we can get the top or most common venues for each of the neighborhoods. We can next apply the k-means classification algorithm to classify the venues into clusters. The algorithm can then be used to assign labels to each of the clusters. The dataset can also be used to list neighborhoods which are closest to a venue which is of interest to the person who is planning to relocate to this city.