**DT Applied Data Science Capstone Wk 4 Peer Graded Assignment**

Have you ever been faced with a move to a new and unfamiliar city? How did you choose where you would focus your search for housing? This project will allow a user to find the best match between neighborhoods comparing the current neighborhood/city by its available amenities to the gaining city and finding the closest matching neighborhood.

Approach

This project will use venue data from Foursquare paired with web data used to identify the parameters of neighborhoods within a city. Using the venues surrounding the current neighborhood as a profile, the nearest matching neighborhood in the gaining city is determined similar to the way NetFlix recommends movies based on what you’ve watched (your Netflix profile).

Source Data

Neighborhood data will be harvested from online data sets to identify neighborhoods by name and lat/long coordinates. Example:

|  |  |  |
| --- | --- | --- |
| **Neighborhood** | **Latitude** | **Longitude** |
| The Beaches | 43.676357 | -79.293031 |
| Studio District | 43.659526 | -79.340923 |
| Lawrence Park | 43.728020 | -79.388790 |
| Davisville North | 43.712751 | -79.390197 |

Popular venue data will be harvested from Foursquare and normalized to determine the “profiles” of each neighborhood. Example:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Neighborhood** | **Art Gallery** | **BBQ Joint** | **Bagel Shop** | **…** | **Thai Restaurant** | **Vegetarian / Vegan Restaurant** |
| Berczy Park | 0.017857 | 0.017857 | 0.017857 | … | 0.017857 | 0.017857 |

K-Nearest-Neighbor will be used to determine the closest match between the current neighborhood and all of the neighborhoods in the gaining city.

Sources

Foursquare venue data <https://foursquare.com/>

New York City neighborhood data <https://cocl.us/new_york_dataset>

Toronto neighborhood data <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>