

IBM Coursera Capstone Project

A Place Like Home

1. Introduction

This project emphasizes the need to compare neighborhoods and find similarity between them. One main use case for this in the real world is that, if someone is moving from one city to another, they may want to settle in a place that's similar to their old neighborhood. But it's not feasible for the person to visit each neighborhood in the new city to observe and find out the one that they need. As Foursquare API contains humungous amounts of location data, we can leverage that to explore and find out similar neighborhoods in a matter of minutes. In this project, I have compared one neighborhood from Toronto city to all neighborhoods in New York city to find out the ones that are similar. This way a person who's moving into from Toronto to New York, can easily find the place that they like to live.

2. Data

The Toronto neighborhood data is obtained from a CSV file which includes Borough, Neighborhood, Latitude and Longitude columns. For convenience to identify the neighborhood in the later section of the problem, a city column has been added with a singular value "Toronto". The resulting data-frame looks like below.

	City	Borough	Neighborhood	Latitude	Longitude
0	Toronto	North York	Parkwoods	43.753259	-79.329656
1	Toronto	North York	Victoria Village	43.725882	-79.315572
2	Toronto	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	Toronto	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	Toronto	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

The New York Neighborhood data is obtained from a JSON file. The columns are similar to Toronto data and like above, a city column is added with value "New York". The resulting data-frame looks like below.

	City	Borough	Neighborhood	Latitude	Longitude
0	New York	Bronx	Wakefield	40.894705	-73.847201
1	New York	Bronx	Co-op City	40.874294	-73.829939
2	New York	Bronx	Eastchester	40.887556	-73.827806
3	New York	Bronx	Fieldston	40.895437	-73.905643
4	New York	Bronx	Riverdale	40.890834	-73.912585

The Foursquare API is used to get the venues in each neighborhood to compare them and identify which are similar to the neighborhood selected from the Toronto data.