

FINAL REPORT

Introduction:

In this project, we'll be using foursquare API to get the data regarding two cities, New York City and The city of Toronto. Then we'll find their neighborhoods and cluster them according to their venue categories. As a result, we can find out which portion of each city is similar to the other and which are not. If anyone wants to examine a portion of a particular city, say, someone wants to check out a restaurant in a particular city then what is the best possibility to find that kind of restaurant to the other city and in which portion. Neighborhoods that are in the same cluster will be similar to their categories.

Data:

Firstly we'll have the data with the borough, neighborhood, latitude, and longitude of each city. Then we'll merge those two datasets into one and by using foursquare API we'll collect the venues for each neighborhood. Each venue consists of four items. Unique id, name, categories, and location. Our main focus will be on venue categories. We'll extract different venue categories along with the venues and using them we'll cluster the full dataset.