Predicting likeness of Neighborhoods in cities

Ibrahim Ramadan
March 8, 2021

1 INTRODUCTION

1.1 BACKGROUND

Let's, say you are traveling to a city you have never and want to know where is the best places to go. Some cities such as NYC and Toronto are just so big that there is no way you have the time to go everywhere in those large cities. Outside of spending alot of valualbe time on google there is really know way to know where to go. Let's say you've been to Toronto and you liked a certain Neighborhoods because of the venues that were present. I am going to use the Foursquare API with ML algorithms to suggest neighborhoods in one city based on their similarity in venue categories.

1.2 PROBLEM

Whenever I visit a city that I've never been to I always run into the issue of what are the areas to go. Unless you know someone who has been to every neighborhood in a city like NYC. Now lets say you live in Toronto and you love a particular Neighborhood in Toronto if only there was a way to find a similar neighborhood in NYC that has similar venues. My application attempts to match neighborhoods in cities based on the types of venues in their respective neighborhoods.

1.2.1 INTEREST

Anybody who travels travels would be interested. Also I could see travel sites such as expedia which could suggest neighborhoods to stay in based on previous places you've stayed or type of venues that you like.

2 Data acquistion and Cleaning

2.1 Data Sources and Cleaning

For the NYC data, I will be using the newyork$_data.jsondataset$.

For the Toronto data I will be using scraping off the Wikipedia site of canadian postal codes and neighborhoods. To scrape the data I will be using Beautiful Soup. I will be getting the geospatial data for Toronto using https://cocl.us/Geospatial$_atadataset.Iwilljoinbothdatasetsusingthepo$

To get venue information I will be using the Foursquare API. I will be using get calls such as search and query to acquire this data from the API. I will be using venue categories as way to cluster the data.

3 Exploratory Data Analysis

NYC neighborhoods are a lot more dense than Toronto neighborhoods when it comes to venues.

Overall NYC and Toronto are not similar when you look at the overall cities.

4 Results

Used Kmeans clustering with a k of 5. Clusters 1, 3 and 4 gave the best clusters. There were many similar neighborhoods in those clusters.

5 Observations

Cluster 5 had a good amount of neighborhoods in NYC, but none in Toronto. When you look at the top venues in this cluster it looks like beach popped up a lot. If you look through the toronto venues there really were no beaches.

6 Conclusion

Clusters 1, 3 and 4 gave were the most helpful clusters for finding like neighborhoods. Especially cluster 3 which had 34 like neighborhoods in Toronto and 119 in NYC. If I continued clustering using venues I would lower to k to 3 clusters. Also I could cluster using other features:

- Peak time of business(If looking for nightlife)
- Popularity of venues
- Age group of customers
- Venue ratings