## The Battle of Neighborhoods - Search for Similar Cities for Graduating University Students

##### Dexin Wang for Capstone Project Assignment

Feb 28, 2020

### Introduction

A survey shows some graduating students from a New York university are searching for information to support their decision of coming relocation. Assuming that the US job market is prospering, the students want to find US cities or city neighborhoods much like New York where they enjoy the nearby venues or amenities. Though there may be many different criteria, numbers and variety of the venues as well as population density are among the most important factors to support the relocation decision. Now the question is, which cities or city neighborhoods are similar to New York the students could choose from? The students prefer to have multiple choices to leverage with job opportunities and other considerations.

### Project Objective

To help the students make decision of relocation, the project objective is to collect and analyze city or neighborhood venue data, from which draw insights about which cities are more similar to the current city they live. In addition to the lists of similar cities, the most popular venues are also presented for further information along with grouped city visualization.

### Analytic Approach

Based on the needs of the customers– in this case the students, it is appropriate to find city neighborhood similarity by clustering the similar cities as groups.

### Data Description

Apparently, we need data of as many cities as possible across the country. So we started with the 200 largest cities by population in 2020 at <http://worldpopulationreview.com/us-cities/#cities>. We can read a table from the website, which has city names, states, population, population density, location coordinates, etc. For the neighborhood venues, we use the data from the Foursquare <https://foursquare.com> by exploring nearby venues defined by a radius of the given city neighborhoods. Let us assume that the city neighborhood centered at the location coordinates is the neighborhood the students might be interested to relocate to. In fact, there are multiple city neighborhoods included in the city list for some metropolitan areas.

In summary, our data is composed of city population densities and nearby venues around the locations for 200 largest US cities. The data has the information as the most important factors mentioned earlier to support the students’ relocation decision.

(The sections below to be completed for week 5 assignment)

### Feature Extraction

### Clustering

### Clusters Validation

### Results Interpretation

### Conclusion