Introduction

Two of the largest areas in the Great Lakes region of the United States are Detroit and Chicago. Many of the area's larger businesses have ties to both cities and large portions of the population in Chicago originate from Michigan, specifically the metro Detroit region.

Having lived in Chicago for over 10 years, I'm relocating to the metro Detroit area to be closer to family and friends. I'd like to research neighborhoods in the metro Detroit area that are comparable to those of Chicago to give insights as to where I'd like to relocate.

This information may also be used to help tourism in the Detroit area or to attract younger talent and recent college graduates to the area. Additionally, it may be useful for anyone else potentially interested in relocating from Chicago to Detroit or vice versa.

Data

For this study, I'll use Foursquare data to analyze what types of venues are similar for areas in Detroit and Chicago. More specifically, I plan to use a data set of combined Detroit and Chicago data and cluster them. Hopefully patterns will emerge and Detroit and Chicago areas will cluster together. Based on that information and my first hand knowledge of neighborhoods in Chicago, I'll be able to make an informed decision on areas I may be interested in residing in the Detroit area.

Since Chicago is substantially larger by population than Detroit and most of the areas of economic growth in recent years have been in Oakland County, MI, this analysis will include neighborhoods from both Wayne (including Detroit) and Oakland counties. Additionally, since many areas in Wayne and Oakland counties are suburban, I'll include all of Cook County for Chicago. Furthermore, Ann Arbor is relatively close to Detroit and is worth including due to its proximity and economic diversity.

In addition to data from Foursquare, this study will require a list of neighborhoods in Chicago and Detroit, a list of community areas in Wayne, Oakland, and Cook counties, and a list of neighborhoods in Ann Arbor. Geographic coordinates for each of these areas will also be used to map the clusters and create an easily readable visualization. Most of this data will be retrieved from the Internet using web scraping techniques.

As a last measure, it would be interesting to cluster the data set of these communities using a walk score index from walkscore.com. However, this may skew the data clustering since Chicago's population is substantially more dense and many areas in Chicago would likely not cluster with Detroit areas due to this factor. However, walkability could be a big factor for many people looking to relocate to the Detroit area from Chicago and some areas should cluster.