Braden Smolko

Introduction

My girlfriend and I graduated from our undergraduate programs this past spring and have been trying to decide where we want to potentially move in the future. With all of the changing situations surrounding COVID-19 we had decided to stay with a family member in Frederick, MD. We have grown to really love living in downtown Frederick, so I thought that it would be a fun and informative project to leverage my knowledge from this certificate to create an analytical approach to finding a similar place to live.

We have been considering a couple of living options for our next steps, but we don't know which option that we would enjoy the most. We do know that we likely want to be near places with family and a place where we can walk to different things like a park or a creek. This means that most of our options are in the Baltimore-Washington Metropolitan area. I have worked with my girlfriend to compile a list of about 30 different options. I will use the K-means algorithm to cluster the different cities together and see which group contains Frederick.

This should help guide us and maybe other people in our situation that would like to see other options in the area that are similar to their own.

Data

I will be using the location and venue data from the Foursquare API to obtain features of these different cities. I will obtain coordinates for each city or center point that we have on our list and then obtain information about the most popular 100 venues within a 5km radius. From these API calls I will extract the venues and their categories for each city. I will one-hot encode this data and then use the mean values collapsed across the city to form the feature set for each city. The cities will then be clustered using the K-means algorithm and the clusters will then be further examined to find the matching group to Frederick.