Capstone Capstone Project

IBM Data Science Professional Certificate

Introduction

Background

Texas and California are considered to be significantly different states. Especially in terms of political approach, California is a liberal-leaning state while Texas is a conservative-leaning state. The cost of living is also considerably higher in California than in Texas. Texas is known for its unpredictable weather. These, among other things, make moving between these states is a unique experience for movers. I attempted to find rudimentary comparability between cities in these two sates using K-mean clustering.

Business Problem

If a person wants to move between two polarizing states of California and Texas, they would like to see what are the cities that have similar characteristics in terms of amenities. It would give them peace of mind knowing that they could have similar access to the businesses and entertainment options once they moved to the new city. So, I tried to answer the problem: What are the similar cities between Texas and California?

Target Audience

The primary audience for this project would be the potential movers between states of California and Texas.

Data

I will use the list of cities of California and Texas from the following sources.

Texas: https://www.texas-demographics.com/cities_by_population

California: https://www.texas-demographics.com/cities_by_population

Data will contain the list of cities and the respective population. I will remove big cities from the analysis as those would have an abundance of amenities and can easily be compared to big cities in the other state. Therefore, not suitable for this type of basic comparison.

I will follow the following steps for obtaining and preparing data.

1. Use web scraping techniques to obtain the list of cities from above publicly available pages. Python requests and beautifulsoup packages.
2. Obtain the latitude and longitude coordinates for the cities using Python Geocoder package or Mapquest Geocoding API.
3. Obtain the list of venues (amenities) for those cities using Foursqaure API