IBM Data Science Capstone Project

Christopher Lopez-Simons

November 18, 2020

Problem/Background

An avid European traveler wants to visit one of the big cities in the United States. However, he also loves the outdoors and wants to be able to visit various outdoor venues so he can explore not only the big city on some days but also the natural habitat that the United States offers. Back in Europe, he spends a lot of his free time either going to parks or beaches to just enjoy the sun and trees around him. On other days he loves to go on hiking whether it be on trails or at natural parks. I have been tasked by him to find the top ideal locations for him to travel to in the United States so that he can then pick his final travel destination from those selections. He is not restricted by finances or time, so those variables can be ignored for this task.

Data Required

The first task that needs to be done to complete this task is to scrape the most populous cities in the United States. From there that data will need to be inserted into a dataframe then cleaned to ensure consistency of each city and state element in the dataframe. A longitude and latitude must be found for each state and inputted into the dataframe. With the longitude and latitude data, an explore api using Foursquare can be done on each city to find the top venues within a large radius of the city to ensure nearby outdoor activities can be found. A pandas dummy dataframe can then be created to assign each city with a mean of each type of venue so that a KMeans algorithm can be run to cluster similar cities. A folium map will then be created to showing which city belongs to which cluster. The clusters will be analyzed for their top venue types to see which includes the most outdoor activities. A new folium map will be created with only the cities belonging to the cluster that exhibits the most outdoor activities. Along with that new folium map, a new data frame will be presented with the top venues for each city in that cluster to the customer so he can decide which city to travel to.