Battle of the Neighborhoods: Where should an entrepreneur open a new yoga studio in Washington, DC?

Introduction

A young business entrepreneur is interested in opening a yoga studio in Washington, DC. They are new to the area and aren't sure which neighborhood would be the best fit to open their yoga studio. They want to make sure the yoga studio is located in a neighborhood that is hip/trendy enough that the yoga studio would have plenty of customers, but they also don't want to open it in a neighborhood that already has a lot of yoga studios, since the competition here might be too high for a new yoga studio to succeed.

This project will use Foursquare API data to explore and cluster neighborhoods in Washington DC based on the top 10 venue types located in each neighborhood. The results of this exploration and clustering analysis will then be used to select the top three neighborhoods for the entrepreneur to consider opening their new yoga studio.

Data Sources

To conduct a neighborhood analysis of Washington, DC, I will use two main sources of data. The first is a list of zip codes in Washington, DC with their corresponding neighborhoods, located on this page (https://www.cccarto.com/dc/index.html). I was not able to find a structured table of this information online, but this webpage contains the data in list format. For example, the first line of the list is as follows, "Adams Morgan Neighborhood 20009 Zip Code." I will attempt to use BeautifulSoup to scrape these data from the page and format them into a pandas dataframe. If I am not able to scrape the data successfully, then I will copy it into an excel file and work from there.

Next, I will use Foursquare API location data to extract and geocode venues in each of the neighborhoods in Washington, DC. Foursquare API location data includes many different characteristics of venues, including venue type, venue name, venue address, and latitude and longitude of the venues. I can then use these features to overlay the venues onto my map of Washington, DC to perform the neighborhood analyses.