

Introduction

The client is interested in opening a yoga studio in the Sacramento County, California. The population of the county, according to 2010 census, is 1,418,788 distributed in seven cities whose total area is 2570 km². The population of the county is diverse, consisting of neighborhoods of varying mean incomes, housing prices, infrastructures and so forth. To be profitable, the studio should be opened in an area with few other yoga studios and consisting of large population with relatively high income. Choosing a location to open a studio in such a large area consisting of varied demographic is a daunting task if done manually. To this end, various data analysis techniques will be employed to choose the best neighborhood for the studio to thrive.

Data

Various kinds of information is needed to suggest the best neighborhood to open the studio. The information about the various neighborhoods in Sacramento County will be scrapped from:

<https://geo.nyu.edu/catalog/stanford-cm480wn0393>

The website will be scrapped using BeautifulSoup module to get zip codes of all the neighborhoods of the Sacramento County. The website also has information about the the name of the neighborhood for each zip code. Same tool will be used to scrap the data of median incomes and housing prices per zip codes of Sacramento County from:

<https://www.bestplaces.net/find/zip.aspx?st=CA&county=06067>

The name of the venues in different neighborhoods will be obtained from the Foursquare location database. Using K-means clustering, the neighborhoods will be divided into several categories. Combined information from clustering, median income and population will be used to choose the best possible location to open a yoga studio in the Sacramento County.