Where to Retire in San Diego

IBM Data Science Capstone Project

Dapeng Wang

2020/02/08

Abstract

San Diego is well known for its good weather and beautiful city, which make it a top choice for retirement. There are more than 100 neighborhoods in San Diego alone, not including the surrounding cities. Which neighborhood to choose if one decides to retire in this city? It all depends on what each neighborhood can offer and each individual's preference.

In this study, data were scraped and analyzed on each neighborhood, they were clustered into different groups by what venues is available from each neighborhood. A recommendation model can give top 5 neighborhoods based on each individual's preference.

Introduction

I have lived in San Diego for many years and really enjoy the life here. Good weather, warm winter, and cool summer, beautiful coastal line, and variety of food, just to name a few. But there are over 100 neighborhoods in San Diego, I do not have the luxury to experience all of them.

Many people, like John, often ask me:"If I want to retire in San Diego, which area should I live?" That is really a big question to answer without diving in to see what each area can offer.

Lucky, after a few hours of research, I found a complete list of neighborhoods in San Diego. Armed with newly acquired skills on web scraping, Foursquare API, and machine learning, I can quickly make some comparison on the neighborhoods and give a recommendation based on John's preference, whether he likes to eat restaurant or fast food, whether he likes to enjoy museum or parks.

This is only a preliminary study on limited data. Many other factors, like the average house price or crime rate in the area, are not considered in the study. These can be included in the future for a improved model.

Business Problem

The objective of this study is to analyze and select the best locations in San Diego for a person to retire. Using data scraping and machine learning techniques like clustering and recommendation, this project will answer the question: If a person wants to retire in San Diego, which neighborhood can you recommend?

Data

The following data is collected in this study:

1. A complete list of neighborhoods in San Diego from Wikipedia. The link is:
   https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San_Diego

   I used Python requests and beautifulsoup packages to scrape the San Diego neighborhood list from Wikipedia page

2. Latitude and Longitude data on each neighborhood. These are retrieved with the geocoder package, data is saved along with the neighborhood names

3. Venue data in each neighborhood from Foursquare API

4. Use Folium package to display the neighborhood overlaid on map of San Diego.