Coursera IBM Data Science Certificate

Capstone Project

The Best Neighborhood for a New Resident of Miami

Courtney T Green

June 2020

Introduction

For many people throughout the United States and beyond, Miami is a dream vacation spot filled with exciting nightlife, beautiful beaches, intriguing art galleries and posh hotels. However, Miami is more than just a hot holiday getaway. Being the cultural, economic and financial center of southern Florida makes Miami an attractive place for young professionals to relocate. Not only is the city of Miami ranked third-richest in the United States, it is seventh in terms of "business activity, human capital, information exchange, cultural experience, and political engagement". But Miami is also the sixth most densely populated major city in the States, having a population of over 460,000 people living in the city and 6.1 million in the metropolitan area. For this reason, a new resident has a daunting task of choosing where exactly in the city they should live.

Business Problem

The objective of this project is to analyze and cluster the twenty-five neighborhoods of Miami based on venues and location to give a new or prospective resident a clear idea of which neighborhoods they may be interested in living. This project seeks to answer the question: If I am moving to Miami and prefer living near x,y, and z, which neighborhoods should I consider?

Data

To give adequate solutions to the problem, I will need the following data:

- List of major neighborhoods in Miami, FL, as currently defined by the city of Miami
- Latitude and Longitude coordinates of these neighborhoods
- Venue data for each neighborhood

Sources of data and methods of extraction:

The Wikipedia page "List of neighborhoods in Miami" (https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami) contains a table listing the major neighborhoods in Miami, totaling twenty-five. It also gives the geographical coordinates of these neighborhoods. I will use web scraping to extract this table from Wikipedia into a Jupyter notebook and convert it to a Pandas dataframe using Python. The table also provides 2010 population and sub-neighborhood information, but that data is not necessary for the scope of this project.

I will gather the venues data for the neighborhoods using Foursquare API. With a considerable database of over 105 million places, Foursquare will be able to give me a large number of venues present in each neighborhood, which I will then use to determine a general characteristic of these neighborhoods.