Battle of Neighborhoods - Miami

Introduction: Business Problem

Miami is one of the largest city in Southeast United States and the Miami Metropolitan area is one of the most populated in South Florida. Also known as the Greater Miami Area it comprises of three counties namely Miami-Dade, Broward and Palm Beach. I have been living in Miami for over 10 years now and I still haven't explored the Miami area completely. Also over these years I have helped many of my colleagues who moved to Miami area. When people move to a new city they research a lot about the new place, and factors like safety, schools, good neighborhood with most amenities close by etc. play a very important role. An application that can do this exploratory analysis and provide a comparison between neighborhoods would be very beneficial.

The purpose of this project is to explore the Miami-Dade and the Miami Metropolitan areas, look at the crime rate, population density and also list the 10 most common venues in some of the biggest cities in the Greater Miami area. This project will help Users save a lot of time by providing a lot of initial analysis and recommendations for which they would have to spend hours doing it manually. Also this project uses K-Means clustering unsupervised machine learning algorithm to cluster the venues based on the various categories such as beaches, restaurants, parks etc. This will not only help the user to compare the neighborhoods but also give them more insights to choose one neighborhood over the other.

Datasets and APIs:

Based on definition of the problem, we would need

- The list of cities in Miami-Dade county and the Miami Metropolitan Area
- The most common venues in each of the neighborhood
- Crime database from last year for the area

We will be using the following data sources and APIs:

Datasets:-

- List of communities in Miami-Dade County, Florida Wikipedia page to get the list of neighborhoods/cities in Miami-Dade County
    - Once we have the list of the cities in Miami-Dade county we will get their respective coordinates.
    - Using those coordinates we will cluster the cities and use a Folium visualization to represent them on the map.
    - Then using the Foursquare API we will get the list of venues nearby to each of the cities and rank them based on the frequency.

- Miami metropolitan area Wikipedia page to get the data about largest cities in South Florida
    - From this we will get the list of major neighborhoods/cities in South Florida.

    - This table also gives the population of the major cities which we will use in a visualization to show the most populated neighborhoods.

- Florida Crimes database for 2019 from http://www.fdle.state.fl.us/FSAC/Documents/Excel/2019/FL_Index_Crime_by_Jurisdiction_2019.aspx
    - We will get Total Crime Index and Crime rate per 100,000 population from this table.
    - We will utilize this data to determine the safest neighborhoods.

APIs:-

- Foursquare API:
    - This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. We will use this to get the list of venues nearby to the neighborhoods that we are exploring

- Folium - Python Visualization Library
    - This Python visualization library would be used to visualize the neighborhoods cluster distribution of Miami city over an interactive leaflet map.

- Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods.