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 haven't explored the Miami area completely yet. 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 county area, look at the crime rate, population density, list the 10 most common venues across the cities. Additionally we will explore South Florida also known as the Miami Metropolitan area and get a list of India restaurants in this area, which by the way is the most asked question by my colleagues. 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

https://en.wikipedia.org/wiki/List_of_communities_in_Miami-Dade_County,_Florida

  - 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

https://en.wikipedia.org/wiki/Miami_metropolitan_area

- 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
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.

Methodology:

• First we collect the Cities data from the Wikipedia pages mentioned in the Data section by Web scraping and data wrangling using the Beautiful Soap library.
• Using the geocoded library get the coordinates for the cities
• Then using the Foursquare API we get the a list of 100 venues in the cities in Miami Dade, County.
• We then narrowed down the list to the top 10 most common venues in each city
• We utilized elbow curve method to find the optimum number of clusters.
• Next we used K-mean Clustering method to cluster the cities together and did the exploratory analysis on that data.
• Using the Folium map we put the clusters on the map (Fig.1)

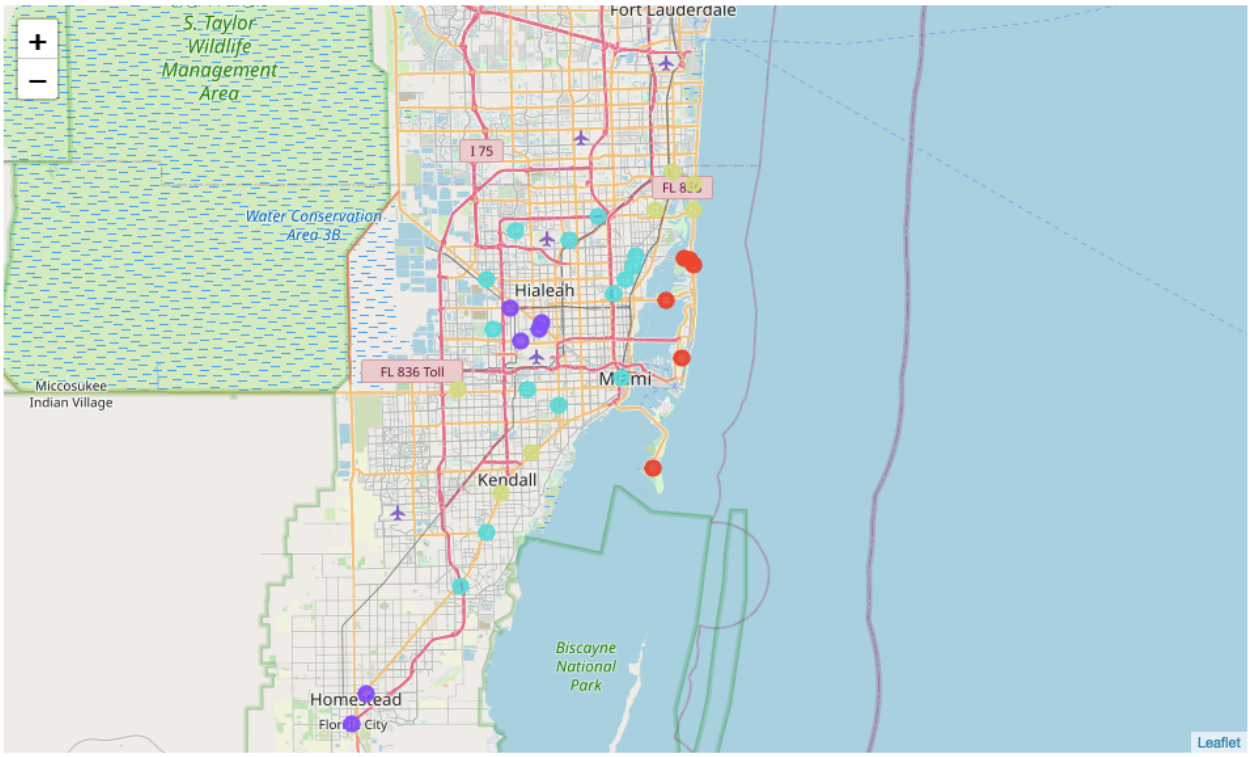

Fig 1. Miami-Dade City Clusters

- We also used the Crime data to determine the safest cities in Miami-Dade County. Fig. 2

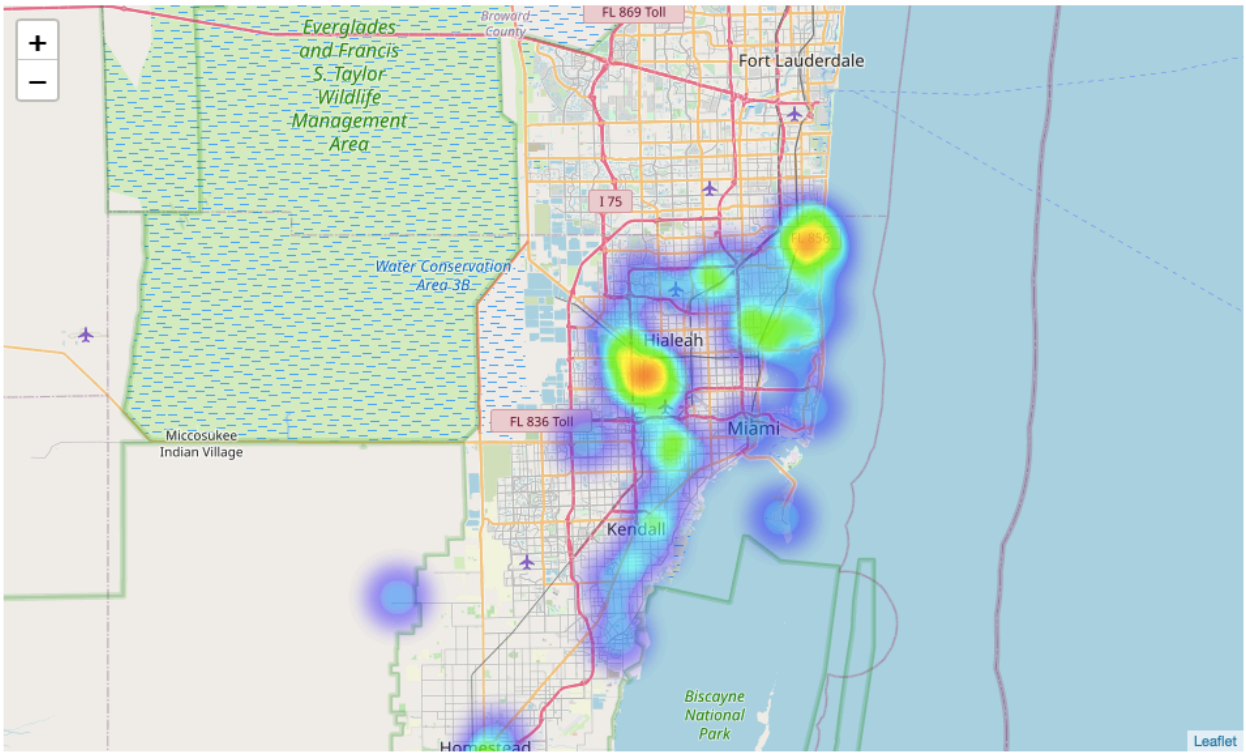

Fig 2. Miami-Dade Cities Crime Rate Heat Map

- Additionally we explored the cities in Miami Metropolitan area to get the list of Indian restaurants. (Fig.3)

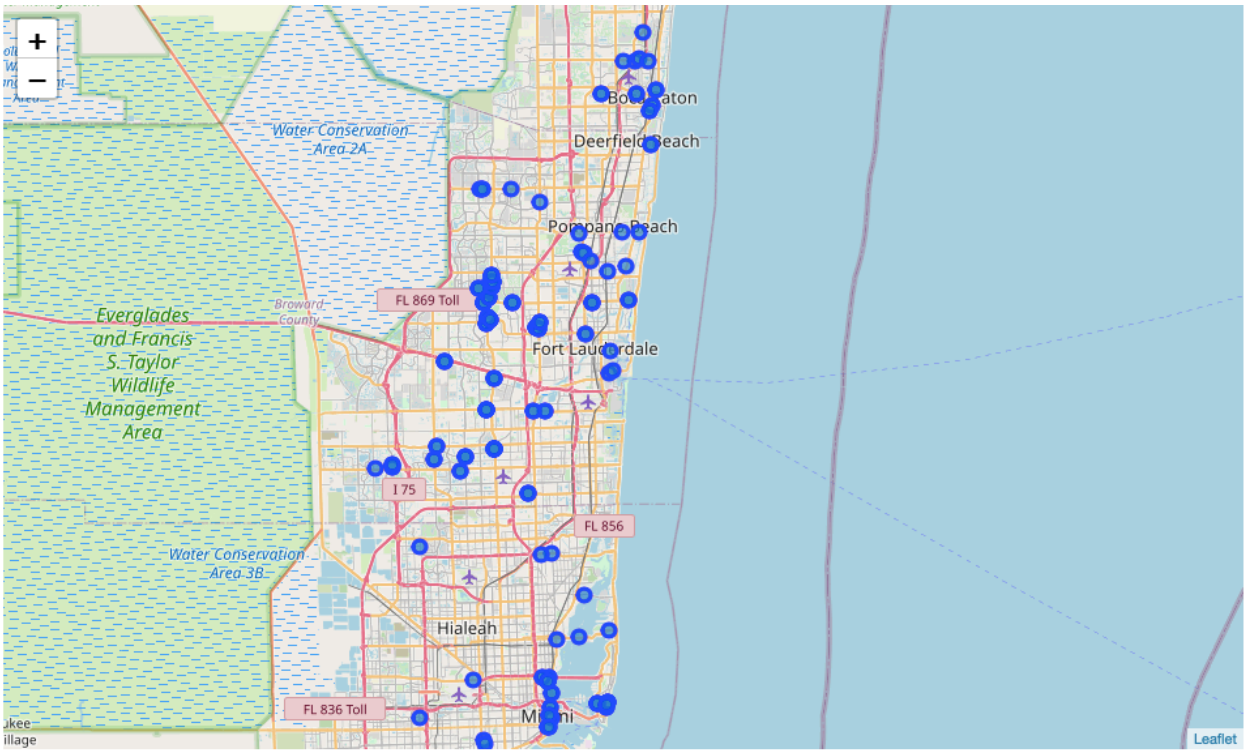

Fig. 3 Indian Restaurants in South Florida

Results:

Our exploratory analysis for Miami Dade county and its cities shows that it has a good mix of different types of Venues catering to varied interests of people. Being a tourist destination also may be a reason for this. We were exploring the Miami Dade cities with respect to various amenities, so we divided the cities into clusters. Based on the clusters data we can see that Cluster 0 the most popular venues are the beaches and restaurants. So people who love the beaches and love eating out can choose to live in these cities. Cluster 2 has a good mix of Parks and restaurants. Cluster 3 has lot of shopping options. Additionally we also used the Crime data in these cities to determine the safest cities to live.Results clearly show that cities that are most populated also have high crime rates. The heat maps and the bar plot help to visualize the data.

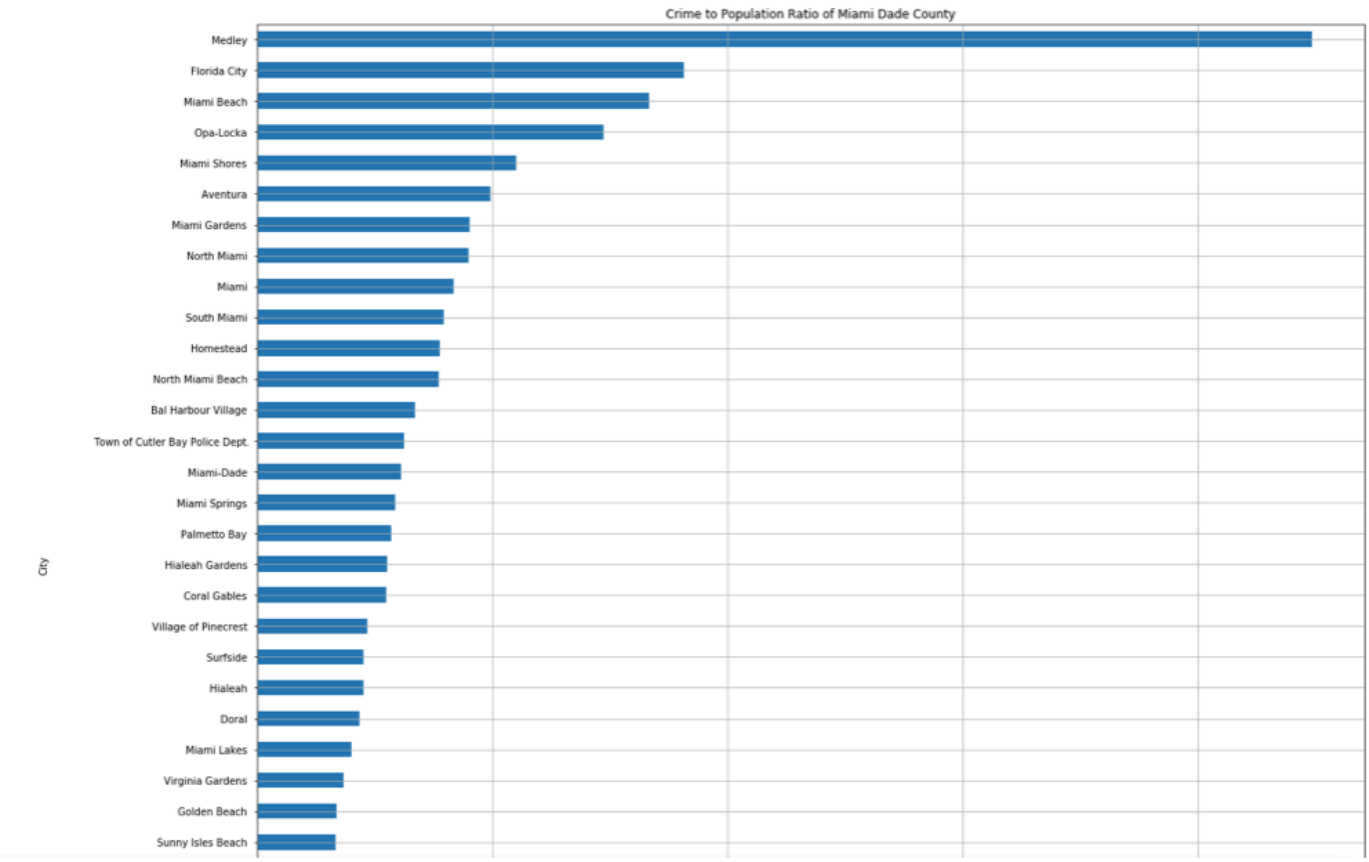

Fig 4. Crime Rate per 100,100 population in Miami Date Cities

The second part of our analysis was to explore the South Florida Area also known as the Miami Metropolitan Area. We looked at the Population across these cities.(Fig. 5)

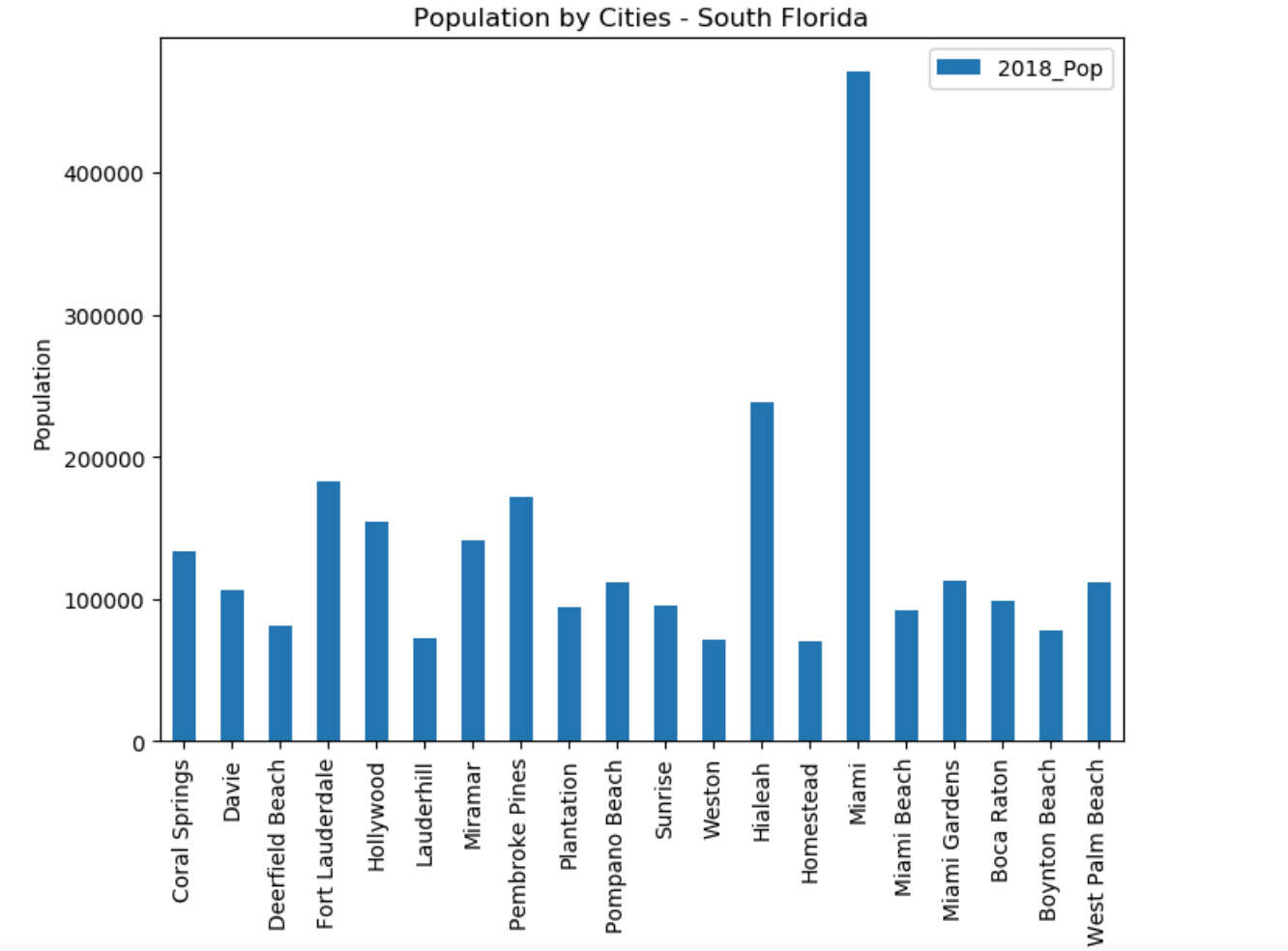

Fig.5 Population by Cities - South Florida

The analysis mainly focussed on Indian restaurants in the largest cities in South Florida. Based on the results we can infer that there are lot of Indian restaurants to choose from in a 5 miles radius from the major cities(Fig.6)

Fig 6. Number of Indian Restaurants by Cities in South Florida

The above analysis will help anyone who wishes to move to the Miami area in their decision to choose a city of their liking.

Conclusion:

The purpose of the project was to explore the Miami Dade county and the Miami Metropolitan area. In order to aid the users with the exploring we clustered the cities in Miami Dade county to come up with recommendation of cities with better facilities and lesser crime rates.

First we gathered the data of Cities in Miami Dade County from the Wikepedia page and the largest cities data in Miami Metropolitan area from Wikipedia page. We also got the crimes data from the webpage: http://www.fdle.state.fl.us/FSAC/Documents/Excel/2019/FL_Index_Crime_by_Jurisdiction_2019.aspx . Then as part of Methodology section we did the following :-
a. Collect and Inspect Data
b. Explore and Understand Data
c. Data preparation and preprocessing
d. Modeling.

In particular, in the modeling section, we used the k-means clustering technique to cluster the various cities.Then using the Foursquare API first we got the list of 100 venues and then narrowed it down to the top 10 most common venues. Then using the

K-means clustering we clustered the neighborhoods to get the most popular venues by the clusters.

Our analysis shows that there are lot of options in each of cities around the Miami area. We can infer from the analysis that cities like Doral and Miami Lakes with plenty of amenities and relatively low crime rate are better options for living if one doesnt want to live in a quite neighborhood. Whereas cities like Sunny isles and Miami Beach are popular among people who want to live near the beach.

The last part of out analysis helped us to get a list of Indian restaurants in the Miami Metropolitan area. The conclusion we came to is that there are plenty of options available.

With this we can conclude that the analysis will help users with the required information.