The Battle of Neighborhoods

Determining the Best Neighborhood to Open an Italian Restaurant in Miami

Daphne Cheung

*November 2020*

**Introduction**

Matilda is a restaurateur from New York City, who has established two successful Italian restaurants on the east coast. She plans on expanding her restaurant business to Miami and would like to open an Italian restaurant in a neighborhood with an existing clientele for such cuisine. With that in mind, I will be helping Matilda in determining the best neighborhood to open up her new restaurant in Miami using data from the Foursquare API.

**Problem Statement**

The objective of this project is to determine the best Miami neighborhood for Matilda to establish her new Italian restaurant. However, she does not have any knowledge on where Italian restaurants are currently located in Miami, consequently, where there may also be an existing clientele for her restaurant. Using machine learning and data from the Foursquare API, I will be able to answer the question of: which neighborhoods should a restaurateur open an Italian restaurant in Miami?

**Target Audience**

The target audience for this solution is a restaurateur who seeks to open an Italian restaurant in Miami but does not have the knowledge to know where they should do so.

**Data**

1. **Data Requirements**

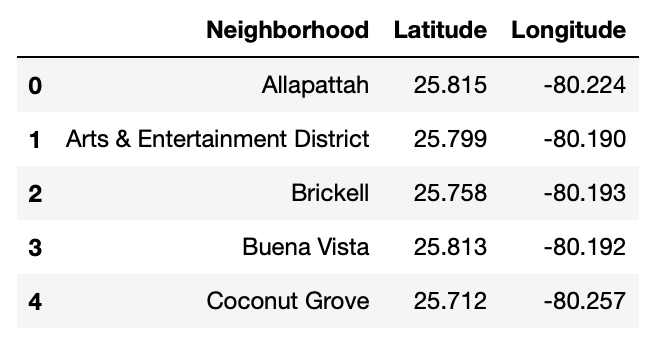
* Neighborhoods in Miami
* Latitude and Longitude of Neighborhoods in Miami
* Italian Restaurant Venues in Miami

1. **Data Sources**

* Scrape Wikipedia for all neighborhoods in Miami
* Obtain latitude and longitude of those neighborhoods using the Geocoder package
* Use the Foursquare API to retrieve Italian restaurant venues in Miami

1. **Data Transformation & Cleaning**

While the data had been scraped from Wikipedia, it needed to be transformed from an HTML table into a Pandas DataFrame. Once read into a Dataframe, columns consisted of: Neighborhood, Demonym, Population2010, Population/KM², Sub-Neighborhoods, and Coordinates. Neighborhoods with missing coordinates were removed from the DataFrame. As only the Neighborhood and Coordinates data were necessary, all other columns were removed. Both latitude and longitude were in the same column. I separated the two values so that latitude and longitude sat in different columns.

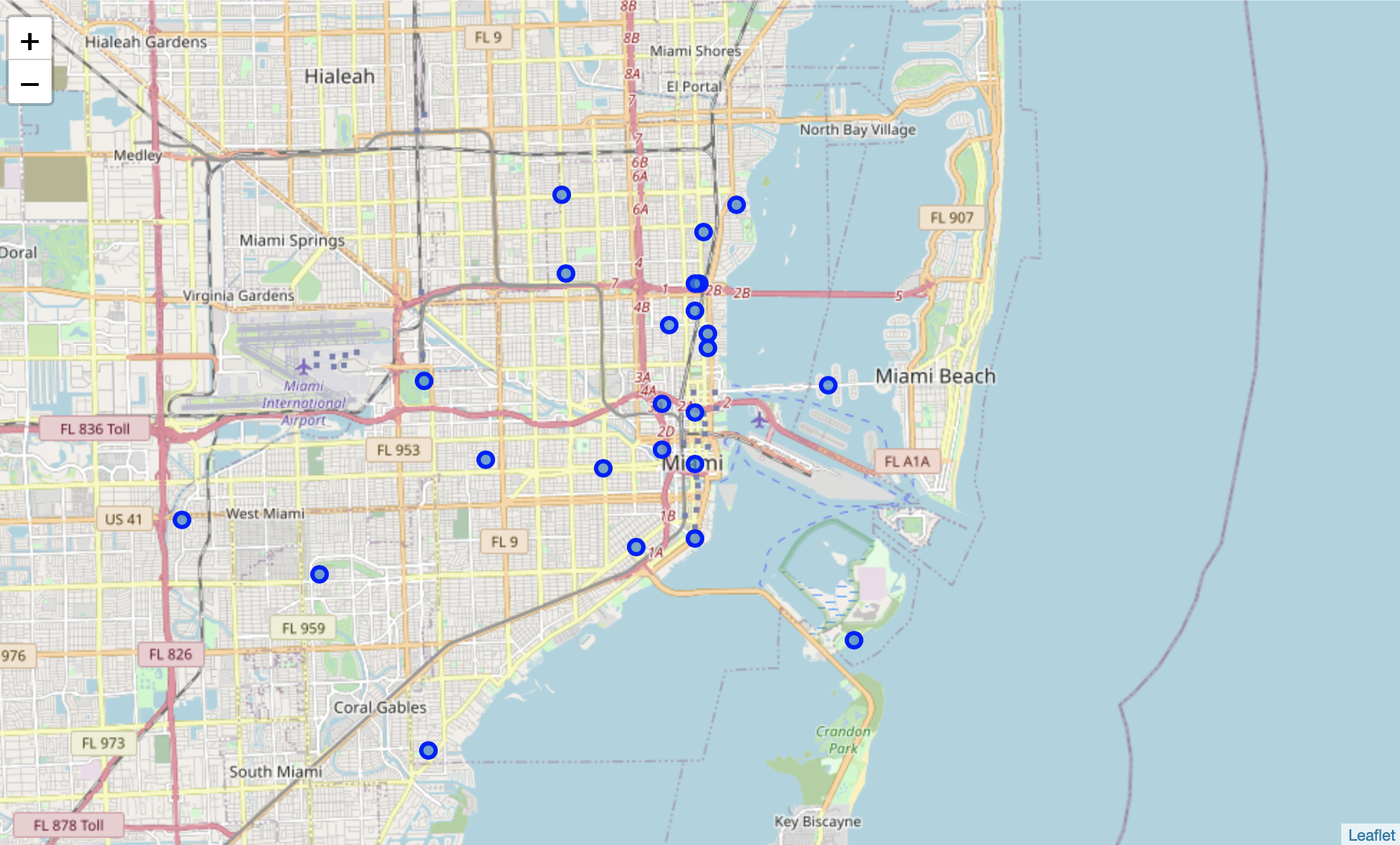


I was left with a DataFrame with 25 neighborhoods in Miami, along with each of their latitude and longitude coordinates.

**Exploratory Analysis**

1. **Mapping Out Miami and its Neighborhoods**

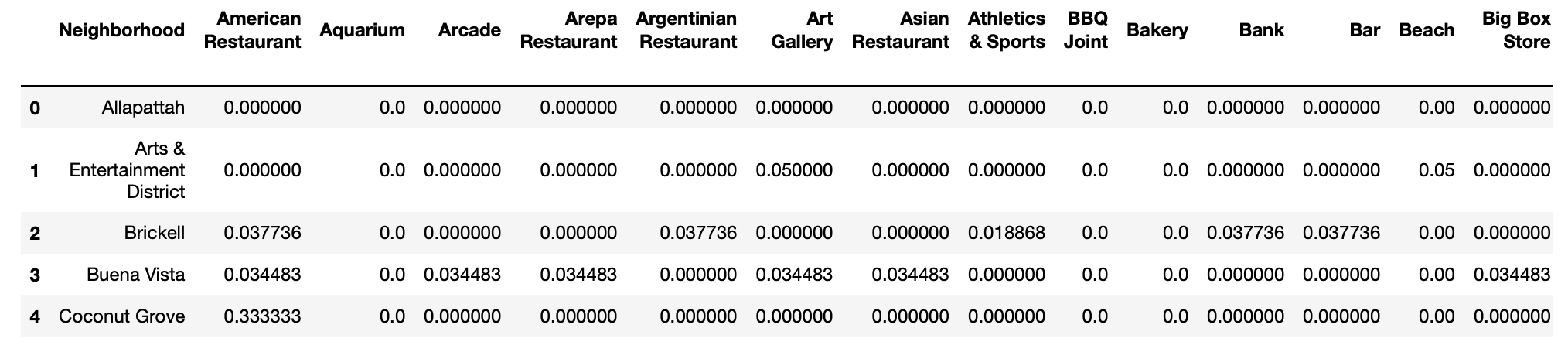
To visualize Miami and its neighborhoods, I used the Folium library to map out Miami and its neighborhoods using the coordinates documented above.



1. **Extracting Top Venues of Each Neighborhood**

Using the latitudes and longitudes of each neighborhood, I began to extract the top venues of each neighborhood for exploration. Extraction of venues came with the venue name, as well as venue latitude, venue longitude, and venue category. I then counted the number of top venues for each neighborhood. This extraction resulted in 137 unique venue categories in Miami, one of which was “Italian Restaurant.” Each venue was transformed into a dummy variable for future grouping and analysis of neighborhoods.

Venues were then grouped by Neighborhood, where frequency of each category of venue was averaged.



Based on those averages, I extracted the top five venue categories for each of the Miami neighborhoods, nine of which contained “Italian Restaurant” in their top five venue categories.

**Methodology**

To identify the best neighborhood for Matilda to open an Italian restaurant, I utilized the K-Means clustering algorithm in scikit-learn on neighborhoods that contained Italian restaurants.

I instructed the K-Means algorithm to partition the neighborhoods into three different clusters based on their similarities. The scikit-learn algorithm then fit the K-Means model on Miami neighborhoods containing Italian restaurants and produced cluster labels for each neighborhood. I merged the dataset of Miami neighborhoods with Italian restaurants with the results from the K-Means algorithm for further analysis. The merged dataset was then merged with additional neighborhood location data to be visualized using the Folium package.

**Results**

Based on the results produced by the K-Means clustering algorithm, the best neighborhood in Miami to open an Italian restaurant would be in cluster 2, in the neighborhoods of The Roads and Upper Eastside. In analyzing the individual cluster, its neighborhoods had the greatest concentration of Italian restaurants, averaging around 14% in frequency. Cluster 0 consisted of neighborhoods with an average of 6% in frequency of Italian restaurants. And cluster 1 consisted of neighborhoods with an average of .08% in frequency of Italian restaurants.

**Discussion**

From the results produced by the K-Means algorithm, Matilda should open her first Italian restaurant in Miami in either The Roads or the Upper Eastside. There will likely be an existing clientele for her to take advantage of. However, it may also be beneficial for her to explore neighborhoods in cluster 1. While there are few Italian restaurants in those neighborhoods, she has a first-mover advantage. Matilda would have fewer direct competitors in those neighborhoods.

**Conclusion**

All in all, I have explored neighborhoods in Miami that have Italian restaurants already. I have visualized Miami neighborhoods and collected venue data from the Foursquare API to gain an understanding of the neighborhoods’ existing businesses. I then clustered those neighborhoods to find the neighborhoods best suited for Matilda to open her Italian restaurant. Ultimately, results concluded that the best neighborhoods to open an Italian restaurant would either be The Roads or the Upper Eastside.