ANALYSIS OF MIAMI NEIGHBORHOODS: HEALTH/FITNESS VENUES

IBM Applied Data Science Capstone Report

Joshua D. Hudson

May, 2020

TABLE OF CONTENTS

INTRODUCTION

Background

Among its many attractions, Miami, Florida is known for its great weather, beaches, nightlife, food, and multi-cultural population. Generally speaking, the residents of Miami are very active and interested in maintaining physical fitness and healthy dieting. This is exhibited when strolling through local parks and beaches, as fitness and sport amenities are almost always being utilized, and at some times, there is even a line of people standing and waiting to be the next to do as many pull-ups or push-ups they can manage in front of gathering crowds and passers-by. Yoga studios are almost always booked full-time, and local gyms and fitness centers are operating at full capacity during early evening hours when the work force finishes the day. These trends and behaviors of the population can lead to very interesting investment opportunities, provided the location is easily accessible, and the right population is served.

Problem

A local business investor has approached the company to analyze the main neighborhoods within the City of Miami, Florida (not including other incorporated municipalities of Miami-Dade County). The investor's vision is to open a new health club/gym that is adjacent to a new health food market. The purpose is to have both venues side-by-side to serve the health club patrons with the convenience of shopping for healthy food options.

The project analysis is to consider neighborhood population density and the venue preferences of the general population in the area. Multiple neighborhoods can be clustered together if their characteristics are similar, and a project location can be determined to best serve nearby patrons and shoppers.

Stakeholders and Interest

The business investor who has developed this project may also have additional investors who would be interested if the analysis proves to be favorable for the project.

DATA ANALYSIS

Data for this analysis was collected from various sources. The following is a list of sources used.

- List of Miami neighborhoods with approximate GPS coordinate locations: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami
- Demographic and land area data for Midtown Miami: https://www.point2homes.com/US/Neighborhood/FL/Midtown-Edgewater-Demographics.html and https://www.cpexecutive.com/post/midtown-opportunities-l-l-c-acquires-22-acres-of-land-in-midtown-miami/
- Supplemental latitude/longitude coordinates were found using Google Maps where initial data was missing. www.googlemaps.com
- FourSquare database of venues to evaluate neighborhood characteristics and business competition (connect via API): https://api.foursquare.com/v2/venues/...

As data was imported into the analysis (via Jupyter Notebook), it was reviewed, cleaned, and formatted appropriately. In order to accomplish this, some assumptions had to be made.

- It was noted that during the time of this analysis and reporting, significant developments have taken place with respect to the COVID-19 pandemic. Currently gyms and fitness clubs are ordered to be closed in the state of Florida. This analysis assumes gyms and fitness centers will be permitted to re-open and return to normal business in the relatively near future.
- The data sets evaluated and conclusions made are based on the venue and location data available at the time of evaluation along with common knowledge of the areas. Factors such as real estate market values, property taxes, local development regulations, etc. were not considered in this evaluation.
- In the originally imported neighborhood database from Wikipedia.com, the Health District neighborhood had no latitude and longitude coordinates. Googlemaps.com was used to estimate the approximate center of this area. The Civic Center train station was assumed to be this approximate location.
- The Midtown neighborhood (most of which is formerly known as Edgewater) was missing population and land area data, as this area was not considered a significantly populated area until redevelopment efforts in 2010 made it a thriving neighborhood.

Data from the two Web sites referenced above were assumed to be appropriate for this analysis.

- Population and land area data for Venetian Islands was not provided in the Wikepedia.com table, and other various sources were extremely inconsistent. Because of this, Venetian Islands neighborhood was excluded from this analysis.

Once the above-listed assumptions were implemented, the data was checked for formatting and converted appropriately for use. After the Miami Neighborhood data was cleaned and formatted, FourSquare credentials were set up, and an API request was utilized to review all available types of venues within 1,000 meters of each of the respective neighborhoods' GPS coordinates. Below are examples of the data output at these steps in the analysis.

Out[11]:

	Neighborhood	Population2010	Population/Km²	Sub-neighborhoods	Latitude	Longitude
0	Allapattah	54289	4401	NaN	25.815	-80.224
1	Arts & Entertainment District	11033	7948	NaN	25.799	-80.19
2	Brickell	31759	14541	West Brickell	25.758	-80.193
3	Buena Vista	9058	3540	Buena Vista East Historic District and Design ...	25.813	-80.192
4	Coconut Grove	20076	3091	Center Grove, Northeast Coconut Grove, Southwe...	25.712	-80.257
5	Coral Way	35062	4496	Coral Gate, Golden Pines, Shenandoah, Historic...	25.75	-80.283
6	Design District	3573	3623	NaN	25.813	-80.193
7	Downtown	71000	10613	Brickell, Central Business District (CBD), Dow...	25.774	-80.193
8	Edgewater	15005	6675	NaN	25.802	-80.19
9	Flagami	50834	5665	Alameda, Grapeland Heights, and Fairlawn	25.762	-80.316
10	Grapeland Heights	14004	4130	NaN	25.792	-80.258
11	Health District	2705	2148	NaN	25.790	-80.215
12	Liberty City	19725	3733	NaN	25.832	-80.225
13	Little Haiti	29760	3840	Lemon City (aka Little River)	25.824	-80.191
14	Little Havana	76163	8423	Riverside and South River Drive Historic District	25.773	-80.215
15	Lummus Park	3027	3680	NaN	25.777	-80.201
16	Midtown	3162	13747.8	Edgewater and Wynwood	25.807	-80.193
17	Overtown	6736	3405	Spring Garden	25.787	-80.201
18	Park West	4655	3635	NaN	25.785	-80.193
19	The Roads	7327	4899	NaN	25.756	-80.207
20	Upper Eastside	12525	2513	Bay Point Estates, Bayside District, Belle Mea...	25.83	-80.183
21	Virginia Key	14	4.01146	NaN	25.736	-80.155
22	West Flagler	31407	4428	NaN	25.775	-80.243
23	Wynwood	7277	2983	Wynwood Art District and Wynwood Fashion District	25.804	-80.199

Table 1: General Miami Neighborhood Data from Data Sources

For visual reference, a map was created using the folium package to review the approximate locations of each of the evaluated neighborhoods.

Out[17]:

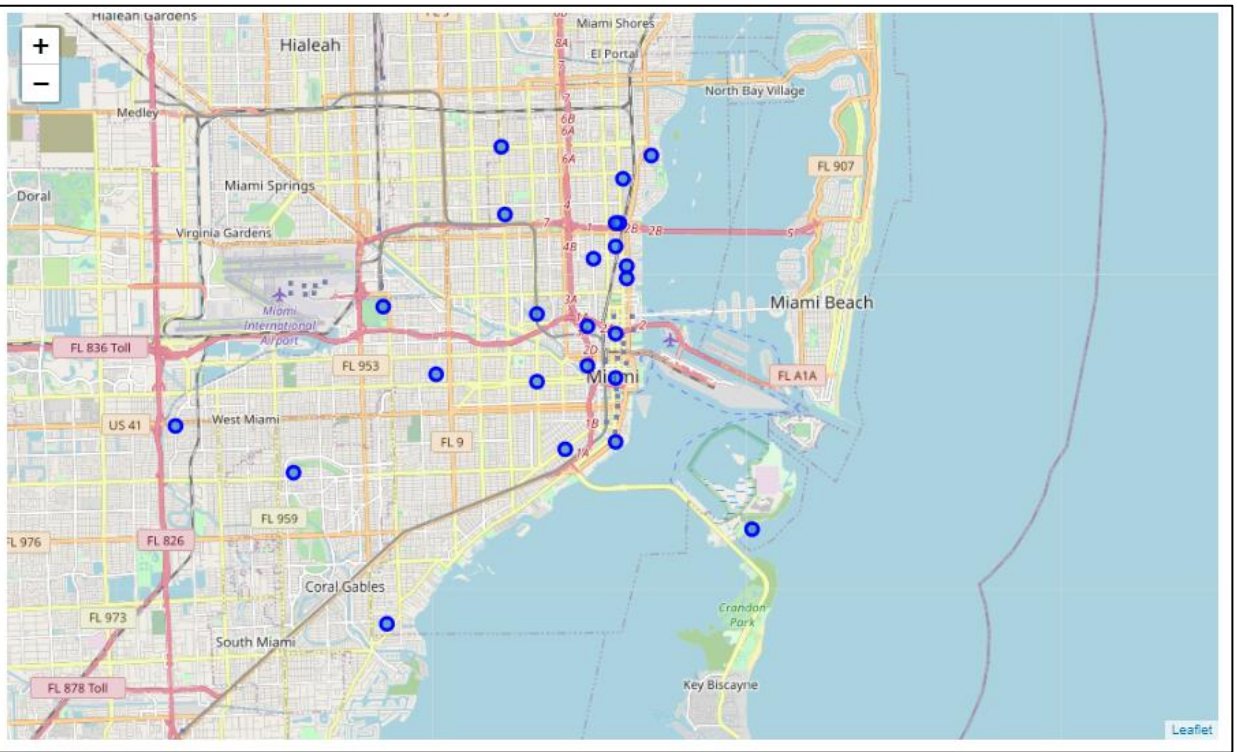

Figure 1: Map of Miami Neighborhoods

In [20]:
```
print(mia_venues.shape)
mia_venues.head(15)
```

```
(1418, 7)
```

Out[20]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allapattah	25.815	-80.224	Club Tipico Dominicano	25.809557	-80.218593	Nightclub
1	Allapattah	25.815	-80.224	Family Dollar	25.807208	-80.223503	Discount Store
2	Allapattah	25.815	-80.224	Redbox	25.808122	-80.224456	Video Store
3	Allapattah	25.815	-80.224	Charles Hadley Pool	25.819565	-80.216753	Park
4	Allapattah	25.815	-80.224	Winn-Dixie	25.808179	-80.224911	Grocery Store
5	Allapattah	25.815	-80.224	El Presidente Supermarket	25.809744	-80.231959	Food & Drink Shop
6	Allapattah	25.815	-80.224	Fritura Dominicana	25.809588	-80.223622	Food Truck
7	Allapattah	25.815	-80.224	MDT Metrorail - Earlington Heights Station	25.812449	-80.229974	Light Rail Station
8	Allapattah	25.815	-80.224	Chimi Churri Los Primos	25.809275	-80.226073	Food Truck
9	Allapattah	25.815	-80.224	Cafeteria Amarilis Dominican Food	25.808877	-80.223655	Spanish Restaurant
10	Allapattah	25.815	-80.224	amarillis	25.808877	-80.223655	Latin American Restaurant
11	Allapattah	25.815	-80.224	Navarro Discount Pharmacy	25.819749	-80.230086	Pharmacy
12	Allapattah	25.815	-80.224	Joe's Meat Market	25.816633	-80.232063	Grocery Store
13	Allapattah	25.815	-80.224	Charles Hadley Park	25.820589	-80.218300	Park
14	Allapattah	25.815	-80.224	classys	25.815200	-80.232600	Food & Drink Shop

Table 2: Sample Venue Output for Miami Neighborhoods (first 15 lines)

From this, venue data was then transformed to a one-hot encoding format to evaluate the frequency occurrence of each category. Then each of the venue categories were sorted for each neighborhood, and the top ten most common venues were displayed. A sample of this output is shown below in Table 3.

Out[28]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allapattah	Park	Food & Drink Shop	Food Truck	Grocery Store	Cosmetics Shop	Spanish Restaurant	Light Rail Station	Latin American Restaurant	Discount Store	Nightclub
1	Arts & Entertainment District	Art Gallery	Restaurant	Ice Cream Shop	Bar	Juice Bar	Gym	Food Truck	Peruvian Restaurant	Coffee Shop	Gym / Fitness Center
2	Brickell	Hotel	Pizza Place	Italian Restaurant	Yoga Studio	Middle Eastern Restaurant	Residential Building (Apartment / Condo)	Coffee Shop	Argentinian Restaurant	Gym	Latin American Restaurant
3	Buena Vista	Art Gallery	Italian Restaurant	Pizza Place	Coffee Shop	Gym	Café	Park	Furniture / Home Store	Jewelry Store	Business Service
4	Coconut Grove	Park	Playground	Trail	Garden	Boat or Ferry	Plaza	American Restaurant	Cosmetics Shop	Cuban Restaurant	Film Studio
5	Coral Way	Pizza Place	Café	Playground	Sports Bar	History Museum	Gym	Grocery Store	Golf Course	Intersection	Dive Bar
6	Design District	Italian Restaurant	Art Gallery	Coffee Shop	Pizza Place	Gym	Café	Park	Furniture / Home Store	Business Service	Pharmacy
7	Downtown	Hotel	Seafood Restaurant	Italian Restaurant	Peruvian Restaurant	Residential Building (Apartment / Condo)	Spa	Gym	Café	Cosmetics Shop	Brazilian Restaurant
8	Edgewater	Art Gallery	Ice Cream Shop	Restaurant	Coffee Shop	Pizza Place	Mexican Restaurant	Bar	Italian Restaurant	Food Truck	Gym / Fitness Center
9	Flagami	Bakery	Liquor Store	Seafood Restaurant	Latin American Restaurant	Hobby Shop	Gas Station	Department Store	Chinese Restaurant	Record Shop	Discount Store
10	Grapeland Heights	Rental Car Location	Hotel	Bus Station	Gym / Fitness Center	Airport Service	Gym	Train Station	Hotel Pool	Gas Station	Latin American Restaurant
11	Health District	Sandwich Place	Café	Convenience Store	Bakery	Fast Food Restaurant	Coffee Shop	Light Rail Station	Gourmet Shop	Mexican Restaurant	Cuban Restaurant

Table 3: Sample of Top 10 Venues Output for Each Miami Neighborhood

From the data shown in Table 3, one can see that health-oriented venues such as gyms/fitness centers, juice bars, trails, gardens, parks, etc. appear with relative frequency among the neighborhoods.

METHODOLOGY OF ANALYSIS

The next step in the analysis was to review clusters of neighborhoods to find any similar characteristics among small groups. If multiple neighborhoods share similar venue rankings, then it may be advantageous to locate a new fitness center/health food store in a central location to potentially service more than one neighborhood. In order to do this, k-means clustering methods were utilized. The goal was to sort the neighborhoods into five clusters based on their most common venue rankings. Once the clusters were created, a new neighborhood map was created to review the approximate geographies of each.

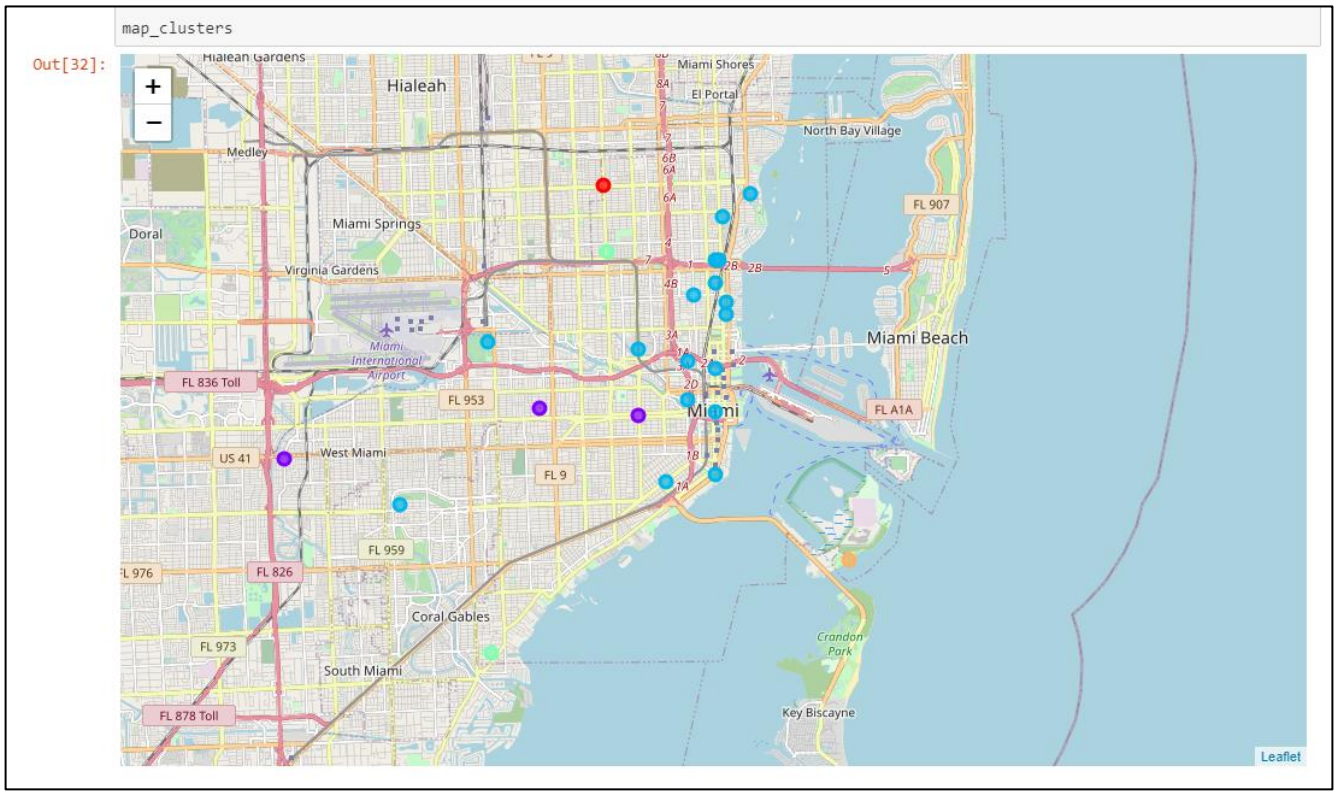

Figure 2: Map of Miami Neighborhood Clusters

From this, each of the clusters were then evaluated individually to try to identify distinguishing characteristics, if any. Below is an example of one neighborhood cluster data set.

Cluster 2

```
In [36]: Cluster2=mia_merged.loc[mia_merged['Cluster Labels'] == 1, mia_merged.columns[[0] + list(range(1, mia_merged.shape[1]))]]
         Cluster2
```

Out[36]:

	Neighborhood	Population2010	Population/Km²	Sub-neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	
9	Flagami	50834	5665	Alameda, Grapeland Heights, and Fairlawn	25.762	-80.316	1	Liquor Store	Bakery	Seafood Restaurant	Cuban Restaurant	Re St
14	Little Havana	76163	8423	Riverside and South River Drive Historic District	25.773	-80.215	1	Cuban Restaurant	Smoke Shop	Latin American Restaurant	Pharmacy	M Re
22	West Flagler	31407	4428	NaN	25.775	-80.243	1	Latin American Restaurant	Pharmacy	Pizza Place	Asian Restaurant	Ba

Table 4: Sample of Miami Neighborhood Cluster Data

Each of the clusters were then evaluated for their approximate geographic centroid, population, and population density. The approximate geographic centroid may be used as a location reference for the new gym/health food venue. For the purposes of this evaluation, this was defined as the median latitude and longitude of the Cluster Neighborhoods. The population and population density are evaluated to ensure the new venue will service an adequate population quantity and also have sufficient customers within a close proximity. Population is considered to be the total population of the Cluster Neighborhoods. For the purposes of this study, the population density of a Cluster is calculated to be the mean population density of the Cluster Neighborhoods.

Out[38]:

	Cluster 2
Latitude	25.773
Longitude	-80.243
Population2010	158404.000
Population/Km²	6172.000

Table 5: Sample of Cluster Location and Population Summary

RESULTS

Results from this analysis were noted to change over time, and it should be expected the results will continue to change as time goes on. It should not be expected that the data set evaluated in this report would be the same as if it were collected and evaluated at any time in the future. For example, it was observed that if the analysis notebook tool were executed on a particular day, and edits were made to code, formatting, comments, etc., and the data were later re-run on a subsequent day, the newly executed data workbook would yield different data clusters. This is likely due to the "live" nature of the data utilized from Foursquare and the API connection. As Foursquare users input data, rate venues, etc. with time, data intended for this analysis could potentially be affected, as was observed. Data analyzed in this report was generated May 7, 2020.

Additionally, by utilizing k-means clustering methods to create the neighborhood clusters, there is an introduction of the iterative process in determining the clusters that is not perfect. The k-means clustering algorithm uses a random initial placement of cluster centroids and converges when its algorithm does not change the centroid location. This method is widely used in the computer science field but does not guarantee to find the optimum.

For the five neighborhood clusters, below is a summary table of the cluster characteristics evaluated for this analysis.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
No. of Neighborhoods	1	3	16	2	1
Latitude	25.832	25.773	25.792	25.764	25.736
Longitude	-80.225	-80.243	-80.193	-80.241	-80.155
Population	19,725	158,404	267,668	74,365	14
Population/km^2	3,733	6,172	5,671	3,746	4
Approx. Area (km^2)	5.3	25.7	47.2	19.9	3.5

Table 6: Cluster Summary

For reference and illustrative purposes, below is a map of the approximate centroids of the clusters.

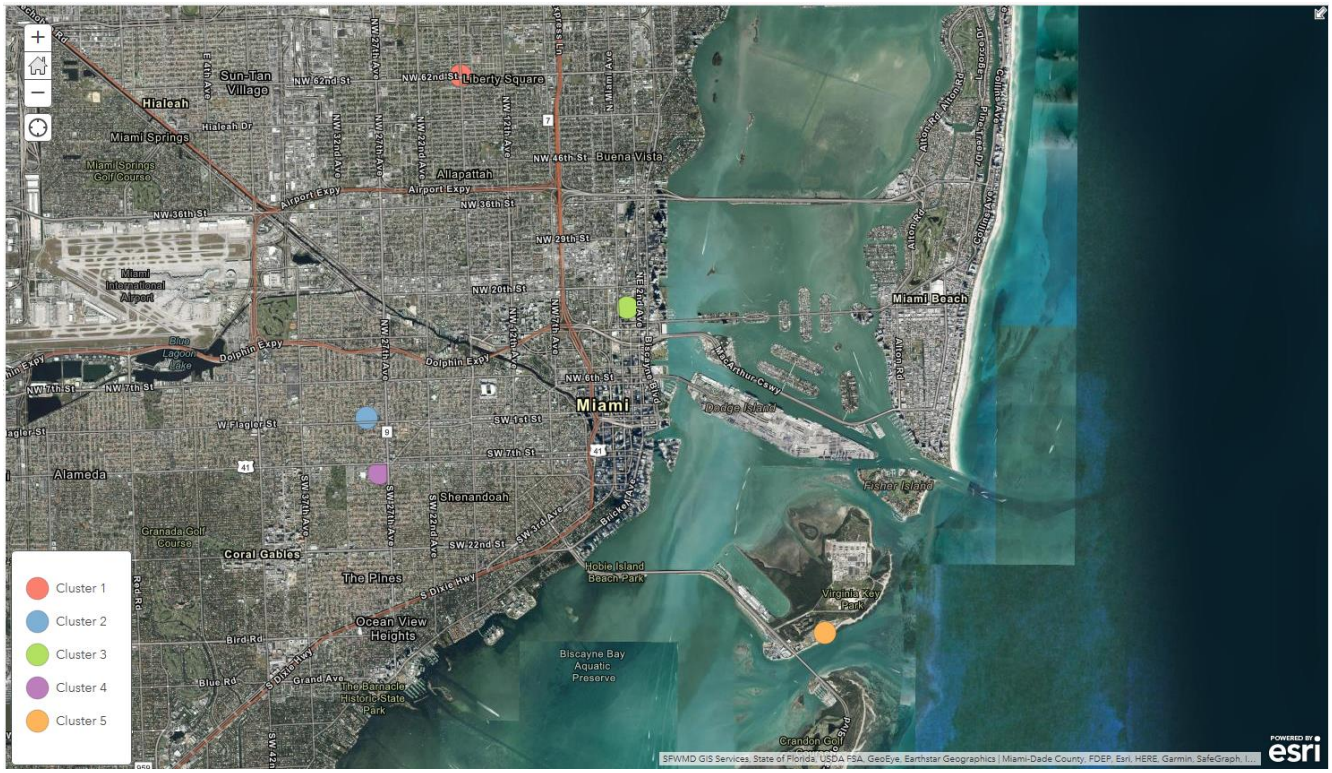

Figure 3: Map of Neighborhood Cluster Centroids

DISCUSSION OF RESULTS

Based on the previously presented data and results, each of the neighborhood clusters will be discussed to evaluate each as a potential location for the proposed gym and health food venue.

Cluster 1 Results:

In this analysis, Cluster 1 has only one neighborhood, Liberty City. Unfortunately, Liberty City is known for having high crime rates and relatively unsafe conditions; however, it is noted that the area is experiencing a redevelopment effort to improve infrastructure, crime prevention, transportation, and overall appearance. Despite having gym/fitness center in the top 10 venues, it is unlikely the Liberty City area will be selected as a potential location until redevelopment efforts have made significant process and can be verified.

Cluster 2 Results:

Cluster 2 is made of three neighborhoods: Flagami, Little Havana, and West Flagler. Upon review of the Top 10 Venues for each of the neighborhoods in the Cluster, it does not appear gyms, fitness centers, or health food stores rank very highly. These three neighborhoods appear to favor restaurants very heavily. Given this data and general knowledge of the area, it is unlikely this Cluster will be the selected location of the new venue, despite having the largest population density of the Clusters.

Cluster 3 Results:

Cluster 3 is by far the largest of the clusters in land area and overall population. It is made of 16 neighborhoods and is home to nearly 268,000 residents. Its land area is more than 47 km^2 and is spread throughout various areas of the city. Some of the neighborhoods in this cluster would certainly be great locations for the new venue. A neighborhood such as Midtown is a great area of Miami that has a relatively large population density and a young, vibrant atmosphere. Individually, the neighborhood would certainly be considered; however, Cluster 3 as a whole covers too much area to be able to evaluate a particular location.

Cluster 4 Results:

Allapattah and Coconut Grove make up Cluster 4 in this evaluation. Of the two neighborhoods, Coconut Grove appears to have a more health-oriented venue preference with a park, garden, trail, playground, and fish market all in its Top 10 venues. Coconut Grove also has a fantastic

view of the water, a marina, and a very vibrant and active population. There is a nice mix of commercial and residential development in Coconut Grove, which makes it a very attractive option for the location of the new gym and health food venue. Conversely, Allapattah is not a very desirable location for a business in the city, and its geographic location in relation to Coconut Grove gives some confusion as to why these neighborhoods were clustered together.

Cluster 5 Results:

Cluster 5 is only the Virginia Key neighborhood. This area is known most for its parks, and there is only a very small population on the island to serve. Virginia Key will not be considered as a potential location of the proposed venue.

CONCLUSION

From the data gathered, presented, and evaluated, it is very difficult to determine the best neighborhood cluster for the proposed venue. Evaluated in their entireties, the clusters offer some valuable data, but they do not present convincing data to lead one to determine where is the best location for the new gym and health food venue.

Please note that the conclusions presented herein are based on the venue and location data available at the time of evaluation along with common knowledge of the areas. Factors such as real estate market values, property taxes, local development regulations, etc. were not considered in this evaluation.

On the other hand, a few individual neighborhoods appeared to be very attractive, namely Midtown and Coconut Grove. Although very different from one another, both areas have very active and vibrant atmospheres with shops, restaurants, and a healthy mix of commercial and residential developments. Midtown is a more urban-style neighborhood, while Coconut Grove is more eccentric with older buildings and a view of Biscayne Bay.

Given that Coconut Grove has a much larger population and more health-focused venues and amenities than those of Midtown, it is concluded that the Coconut Grove neighborhood should be the most strongly considered as the location of the proposed gym and health food venue, solely based on the data evaluated and common knowledge of the area.