

Machine Learning Capstone

Where in Miami to put a new gay bar?

By

Carilda A. Thomas

1. Business Problem

I want to open a gay bar in Miami, Florida. I selected Miami because I used to live there.



Figure 1: KitKat Lounge and Supper Club

2. Location Requirements

- It must be in an area of other gay businesses. Modeling after the gay blocks on North Halstead Street in Chicago (Boys' Town), an area of gay businesses, especially entertainment industry businesses like bars, benefit from proximity so patrons can easily go from one to another in an evening's "bar crawl".
- Distance from schools. Different neighborhoods have differing requirements on how far a bar should be from any school. Some are 500 feet, the majority are 300 feet and one neighborhood mandates 1000 feet. I've arbitrarily selected 300 feet.
- Select district considering local crime rate, paying especial attention to hate crimes.



Figure 2: Miami Skyline

3. *Stakeholders*

- Myself
- Investors who help finance the undertaking
- The Miami gay community

4. Data

- **Miami Neighborhood Data**

Neighborhood data will be scraped from https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami and saved in a pandas dataframe.

Only neighborhood names and coordinates will be extracted from this dataframe.

Here is a partial screenshot:

Neighborhoods in Miami ^[2]					
Neighborhood ↕	Demonym ↕	Population 2010 ↕	Population/ Km² ↕	Sub-neighborhoods	Coordinates ↕
Allapattah		54,289	4,401		25.815,-80.224
Arts & Entertainment District		11,033	7,948		25.799,-80.190
Brickell	Brickellite	31,759	14,541	West Brickell	25.758,-80.193
Buena Vista		9,058	3,540	Buena Vista East Historic District and Design District	25.813,-80.192
Coconut Grove	Grovite	20,076	3,091	Center Grove, Northeast Coconut Grove, Southwest Coconut Grove, and the West Grove (Black Grove)	25.712,-80.257
Coral Way		35,062	4,496	Coral Gate, Golden Pines, Shenandoah, and Silver Bluff	25.750,-80.283

Here is an excerpt of the neighborhoods dataframe after extracting only neighborhood and splitting the coordinates column into latitude and longitude.

	Neighborhood	Latitude	Longitude
0	Allapattah	25.815	-80.224
1	Arts & Entertainment District	25.799	-80.190
2	Brickell	25.758	-80.193
3	Buena Vista	25.813	-80.192
4	Coconut Grove	25.712	-80.257

Figure 3: Reduced neighborhoods dataframe

- **Boys Town data for reference**

Neighborhood name as 'Boys Town' and the latitude and longitude of the KitKat club (a prominent Boys Town establishment) will be appended to the dataframe.

The objective is to see similarity between Boys Town and neighborhoods in Miami.

- **Foursquare**

Foursquare is used to ascertain the top venues for each neighborhood.

The venue information is processed to return the top 10 venues by count for each neighborhood. This is saved in a dataframe and the neighborhoods are then clustered using kmeans clustering.

Here is an excerpt of the venues table before clustering:

Foursquare found 216 unique venue categories.

5. Methodology

1. Scrape the webpage https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami to get a list of Miami neighborhoods and their coordinates using BeautifulSoup and save the results into a pandas dataframe.
2. The coordinates have to be split into latitude and longitude columns and these columns must be converted to float for further processing.
3. Get location data for Boys Town in Chicago using the address of the KitKat club, a well-known gay bar and restaurant in Chicago, with a python geocoder. Append this to the Miami neighborhoods dataframe. This will be used as the “gay standard” when clustering.
4. Define a function to get “nearby” venues from Foursquare, given a maximum distance and a count limit.
5. Run the above function on all the neighborhoods, including the Boys Town neighborhood. This creates a dataframe where each row is a unique combination of neighborhood and venue category.
6. Use pandas get_dummies to create a “one-hot” dataframe where there is a column for each venue category and either a one or a zero in the cell depending

on whether or not this venue appears in this neighborhood.

7. Group the dataframe by neighborhood. Select the top ten venues for each neighborhood. This is an excerpt of the result:

	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
0	Allapattah	Park	Food & Drink Shop	Pizza Place	Cosmetics Shop	South American Restaurant	Discount Store	Video Store	Bakery	Grocery Store
1	Arts & Entertainment District	Art Gallery	Restaurant	Ice Cream Shop	Bar	New American Restaurant	Juice Bar	Beer Garden	Pizza Place	Peruvian Restaurant
2	Brickell	Hotel	Italian Restaurant	Pizza Place	Bar	Restaurant	Yoga Studio	Latin American Restaurant	Argentinian Restaurant	Japanese Restaurant
3	Buena Vista	Art Gallery	Coffee Shop	Italian Restaurant	Furniture / Home Store	Café	Pizza Place	Grocery Store	South American Restaurant	Cosmetic Shop
4	Coconut Grove	Park	Boat or Ferry	Trail	Playground	Cosmetics Shop	Garden	Plaza	American Restaurant	Harbor Marina

Figure 5: Most common venues for some neighborhoods

8. Apply K-means clustering with 5 clusters and add the resultant labels to the dataframe.
9. For each cluster label value, select out the neighborhoods that have this value.

When this is complete, find the cluster containing Boys Town and assume this is the ideal cluster for a gay bar.

6. Results

Kmeans was run with several different numbers of clusters but the KitKat cluster always has 20 neighborhoods while the others are distributed randomly.

7. Discussion

I varied the number of top venues regarded, as well as the number of clusters but the largest cluster is always the one that contains KitKat.

8. Conclusions

The results of the clustering suggest that Miami is very homogeneous, so it probably doesn't matter in which neighborhood I choose to locate my new gay bar.

9. Future Directions

The analysis could be done over including Miami Beach neighborhoods, or with only Miami Beach neighborhoods. The neighborhoods could be mapped against schools to eliminate neighborhoods that would violate legal strictures about proximity of gay bars to schools.

I didn't do any of this because I really don't care much for this analysis. I don't think it adds anything to the sum knowledge in the universe.