Best Locations for a Night Club in New York

1. Introduction

1.1. Background

Assume an entertainment company is choosing a location for a night club in New York City. This report will suggest the top neighborhood for the night club by analyzing data obtained using the Four Square API and data for New York City's neighborhoods.

1.2. Problem

Neighborhoods will be mapped and clustered based on their night life. A neighborhood which already has a lively amount of night life would be a candidate. It is assumed that the interested entertainment company wants to choose an optimal location for their night club which will attract a lot of patrons. Locating the club in a popular, trendy area of the city is important. Neighborhoods with other clubs within walking distance are a good bet for oftentimes club patrons will hop from club to club, especially, if there are other genres of music close by.

1.3. Interest

The bar and night club industry has steady grown since the mid 1990's reaching 23.15 billion U.S. dollars in 2015 [1]. Running a night club can cost several million dollars just for a yearly lease depending on location. A well-placed dance club can gross revenues anywhere from $5,000 to $35,000 nightly [2]. As Harold Samuel, the real estate magnate, once said "Location, Location, Location". Location is of major influence on a brick-and-mortar business and how successful it will become. This analysis will help determine an optimal location for a night club.

2. Data Description

2.1. Data Sources

As mentioned above, the data used will be venue data obtained using the Four Square API and New York neighborhood data obtained from https://geo.nyu.edu/catalog/nyu_2451_34572 . This is the same data that was used in a previous assignment for this course. This data is used to identify the location of each neighborhood in New York using longitude and latitude. The Foursquare API is used to access night club venue information.

There was a third source of data that defined the boundaries of each neighborhood. It was downloaded from github at this location: https://github.com/veltman/snd3/blob/master/data/nyc-

neighborhoods.geo.json . This data set was used to assign clubs into neighborhoods using the actual boundaries of the neighborhoods rather than their distance from some center of a neighborhood. The idea is that neighborhoods are not circles but irregular polygons, and that there is more similarity of a business within a neighborhood rather than across neighborhoods. That is, two streets could be one block over from each, exist in different neighborhoods and be significantly different economically even though they are relatively close in distance.

2.2. Data Cleaning

The data downloaded from FourSquare was reformatted from JSON into a Pandas data frame. Only a few of the downloaded fields were used from the Four Square data set. They were:

https://geo.nyu.edu/catalog/nyu_2451_34572

This data set was used to get the geographical centers of the neighborhoods. It was used with the FourSquare API to get all of the clubs within 3000 meters of the neighborhood center. This data set was checked for duplicates, invalid neighborhood names and bad latitude and longitude.

Data Label	Description
Borough	Name of the borough the neighborhood is within
Neighborhood	Name of the neighborhood
Latitude	The latitude of the neighborhoods
Longitude	The longitude of the neighborhoods

Table 1 Coursera Data Set

FourSquare API

The four square API was used to pull down the follow information. A search radius of 3000 meters was used to find all clubs within 3000 meters of the neighborhood centers defined above.

Data Label	Description
lat	Latitude of the club
lng	Longitude of the club
distance	Distance the club was from the requested search location
categories	The type of venue
address	The address of the venue
labeledLatLngs	Representation of the latitude and longitude as a label
postalCode	The zip code of the club
cc	Country Code
city	The city the club is within
state	The state the club is within
country	The country the club is within
formattedAddress	The formatted address
crossStreet	The cross street
id	The FourSquare id

Table 2 Data from FourSquare API

The id field was was important for the analysis for it was used to look up more information about the venue. It was determined that many of the venues that had a NaN id were also closed temporarily due to

COVID-19. This was done by choosing clubs in the candidate neighborhoods, which are the ones with the most clubs, and then checking for ratings. The clubs in the candidate neighborhoods without ratings were double checked by doing some google searches, and it was discovered that these clubs, without ids, were indeed closed. Considering the circumstances and business closing to control the spread of COVID-19 this made sense.

https://github.com/veltman/snd3/blob/master/data/nyc-neighborhoods.geo.json

This data set was used to get the geographical boundaries of each neighborhood. Using the neighborhood center data was convenient to use to query the FourSquare API; however, once the clubs were obtained using the FourSquare API it was convenient to group the clubs by neighborhood using the boundaries of each neighborhood. This data was converted to Polygons and the club locations were checked to see which Polygon contained the neighborhood. The data was plotted on a map to visual the neighborhoods and ensure that they were correct.

2.3. Feature Selection

The name of the club, the neighborhood of the club in which it is contained, the neighborhood center that the club is closest to, the longitude of the club, the latitude of the club and the id of the club were all selected. The id was used to obtain the rating of the club. For any club that did not have a rating, a negative rating of -1.0 was assigned to penalize the neighborhood. A neighborhood with a significant number of closed clubs would be heavily penalized. This was designed to avoid neighborhoods with a significant number of closed clubs. The negative was chosen to ensure that the neighborhood was sufficiently penalized. A more negative rating could also have been used but -1.0 was chosen.

3. Methodology

3.1. Exploratory Data Analysis

The clubs were plotted on a map of New York along with the neighborhood centers. An example of such a map is shown below. The centers of the neighborhoods are show in red. The club locations are shown in blue:

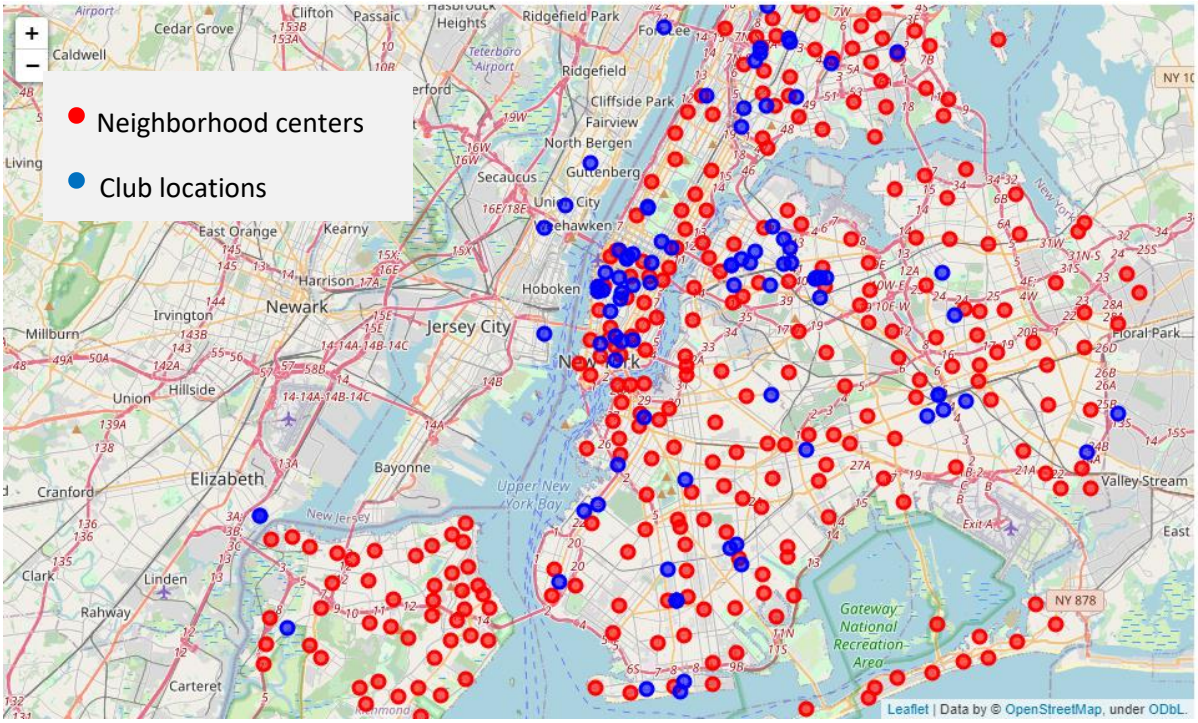

Figure 1 New York Neighborhood Centers and Club Locations

The number of clubs per neighborhood were counted and plotted on the choropleth map. The neighborhoods with the most clubs are shown in red in the figure below:

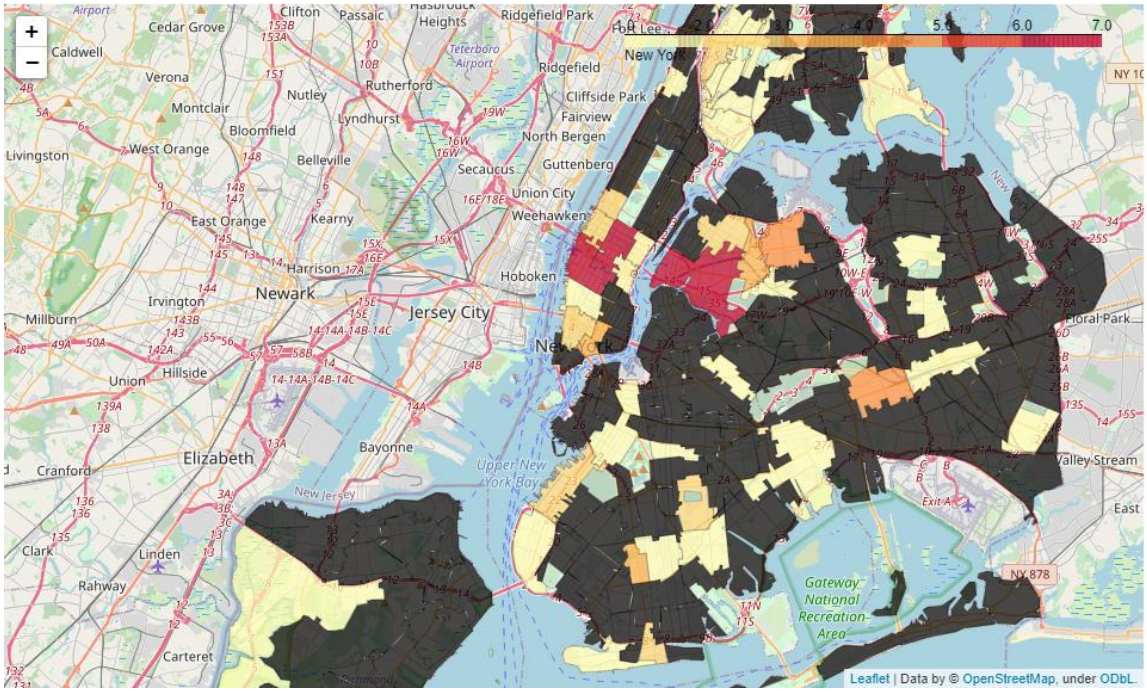

Figure 2 Choropleth Mapping showing Neighborhoods with Most Clubs

The neighborhoods can be sorted in descending order of the number of clubs within their boundaries. The plot below shows the top neighborhoods with clubs:

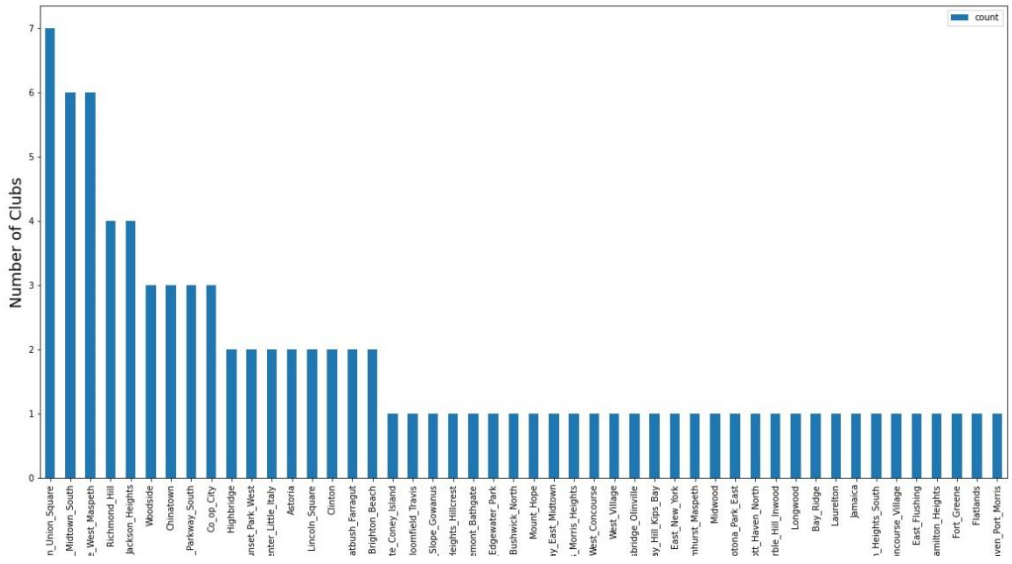

Figure 3 Neighborhoods sorted in order of Number of Clubs

The top portion of the above chart is shown zoomed in below so that one can get a better idea of the neighborhoods with the most clubs.

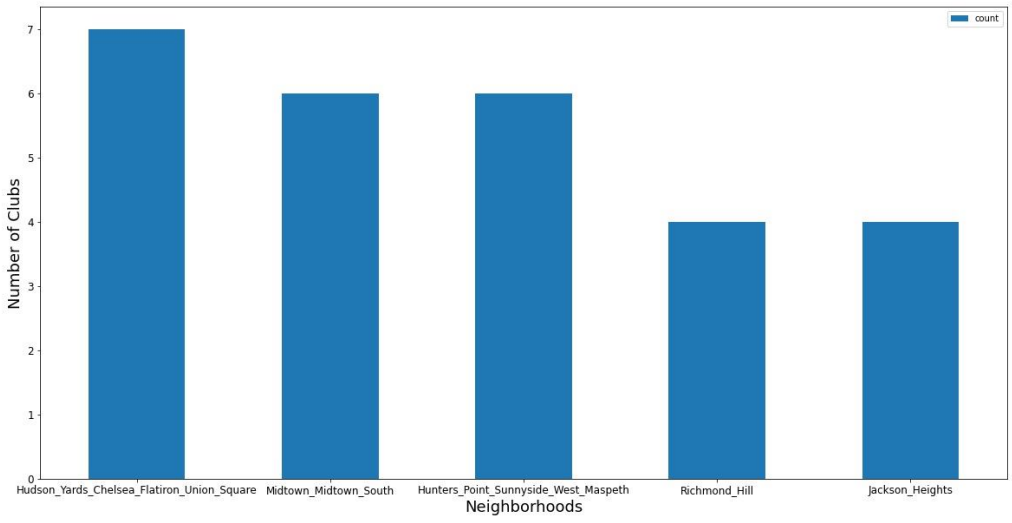

Figure 4 Top Neighborhoods with Clubs

One can see that the top neighborhood groupings above match to the choropleth map.

3.2. Machine learning

K-means clustering was also used to cluster the clubs based on locations and ratings. The ratings were obtained using the FourSquare API. Three centroids were used to group the clubs.

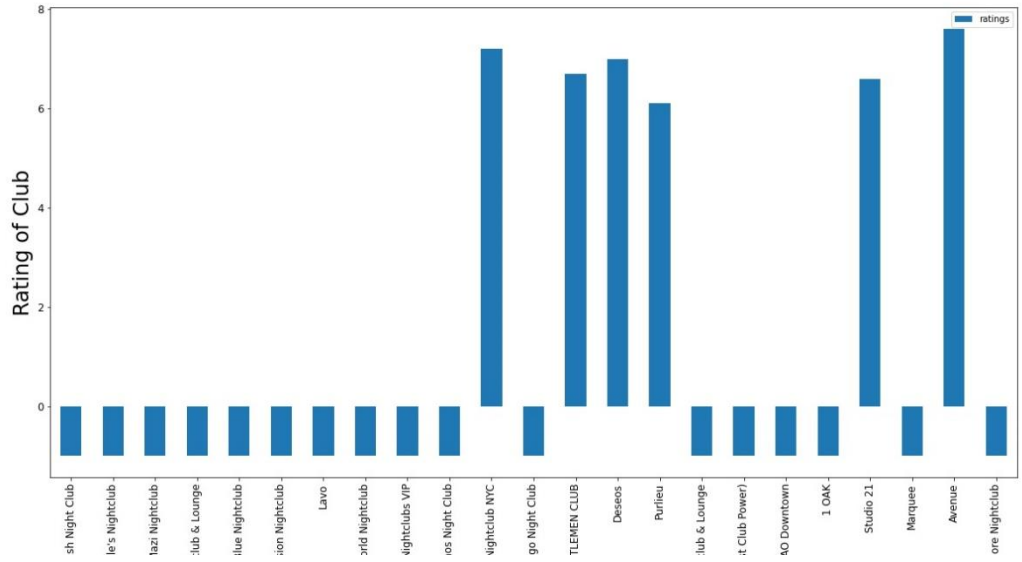

Figure 5 Club Ratings

The clubs were sorted in order of ratings and the top ratings are shown below.

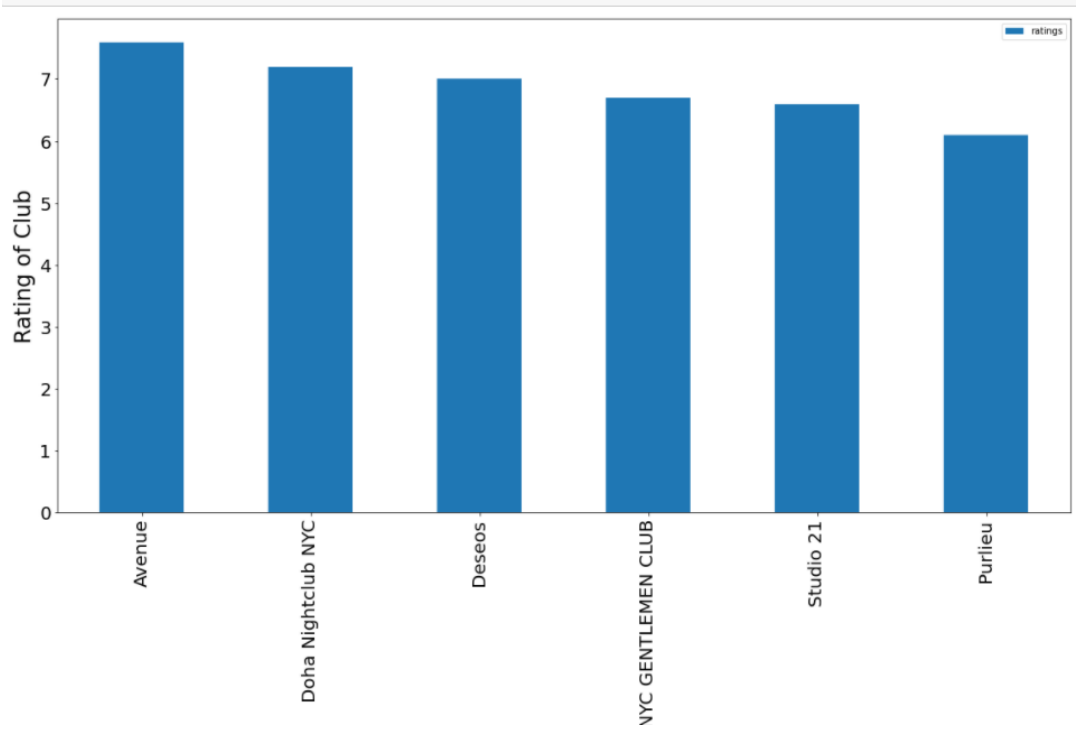

Figure 6 Top Six Club Ratings

The centroids of the clusters were created and plotted below in Figure 8. The centroid with the highest rating is shown below in green. It had an average rating of 6.866. The remaining centroids had overall negative ratings and are shown in black.

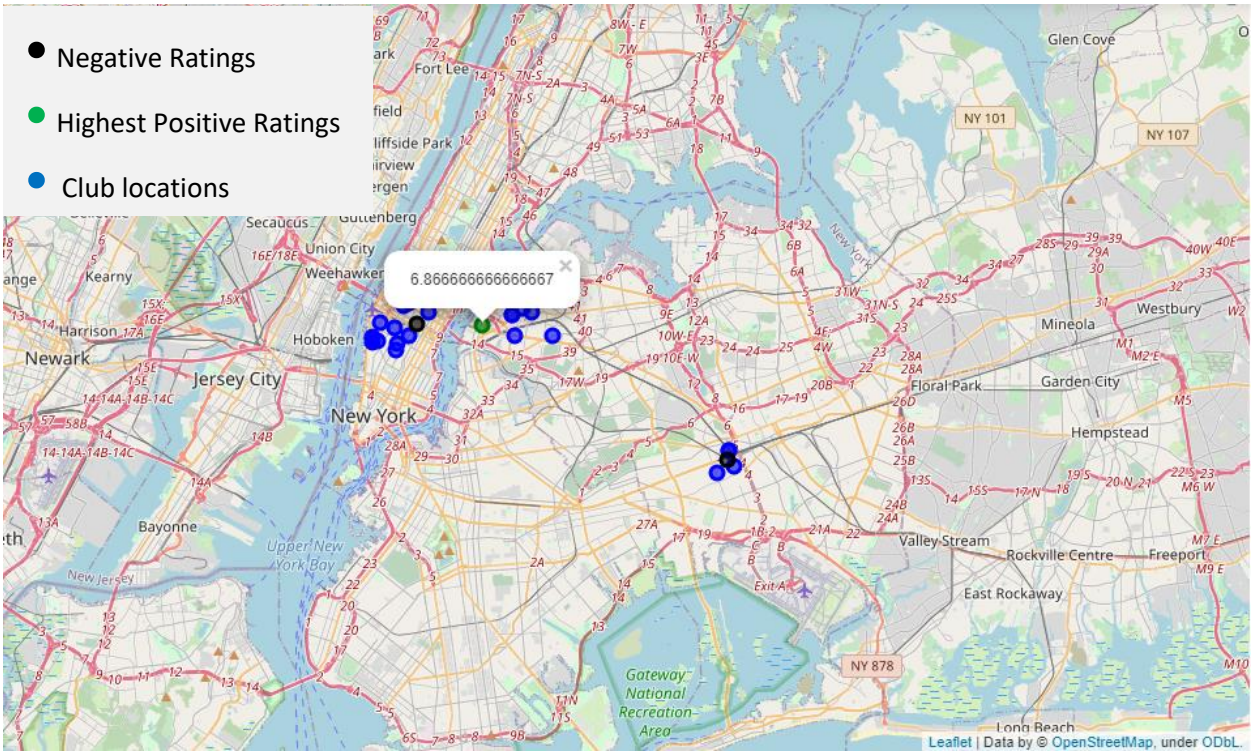

Figure 7 Centroids of the Clubs with Highest Ratings

4. Results

The centroid of the cluster with the highest average rating is located within the bounds of 'Hunters_Point_Sunnyside_West_Maspeth' boundary. The closest neighborhood center to the cluster centroid is "Hunters Point". These methods are in agreement so the neighborhood with clubs that have the highest average rating is "Hunters Point". The other areas seem to have a significant number of closed clubs leading to an overall negative rating for the additional centroids. It is interesting that K-Means clustering computed the same result that the initial exploratory analysis was trending towards.

Notice how the centroid falls within one of the areas with the highest number of clubs. This area also turned out to be one of the areas with the most clubs still open:

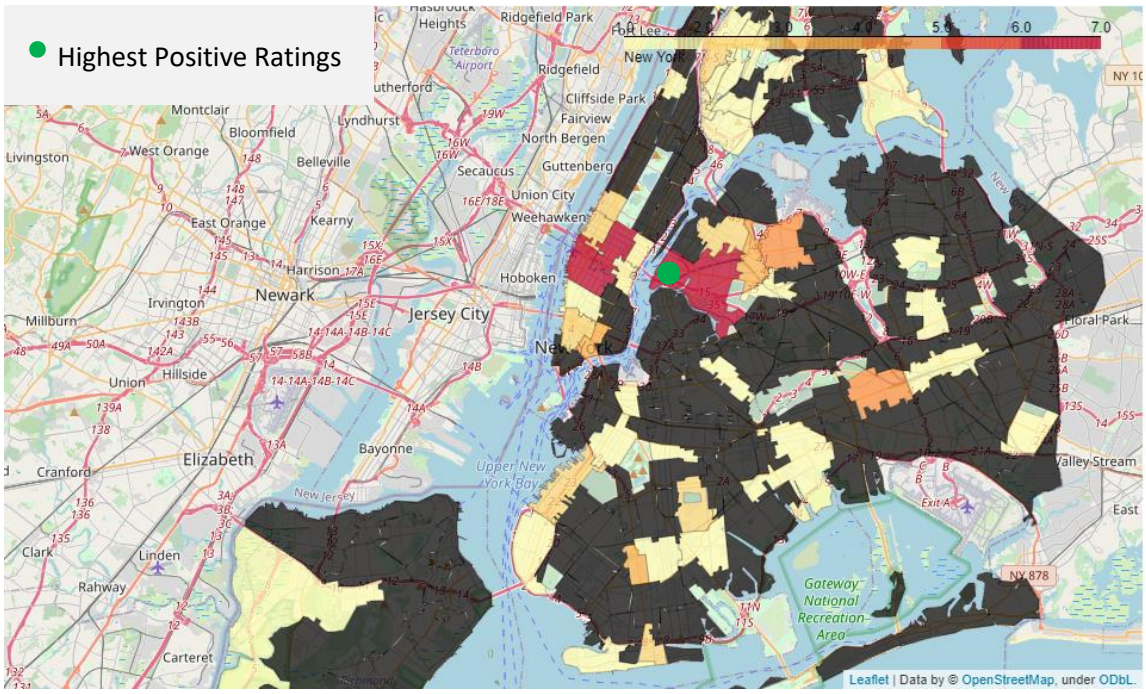

Figure 8 Agreement with Exploratory Analysis

5. Discussion

The analysis could be run again after the COVID-19 crisis ends. Many clubs are currently closed, but this neighborhood seems to have the highest resiliency with clubs still open. The number of closed clubs was in retrospect not unexpected considering the current lock downs and circumstances surrounding COVID-19. The previous historical data may have limited ability to predict the performance of a club because of the unique nature of the COVID-19 crisis and any lasting effects that it might have.

6. Conclusions

The neighborhood determined is one of the few with clubs still open; therefore, one can posit that this location may lead the recovery of the night club scene. Additional information would be helpful such as the amount of patrons per weekend, or the amount of money spent at each club. This would help the entertainment company determine how much profit that they could expect from a certain location. Analysis of the real estate prices would also be helpful in any future analysis to help maximize profit by trading a less expensive location with one that makes less profit. Obtaining such datasets that would provide additional insights would be helpful in any future analysis. To get accurate up-to-date data a paid source rather than an open source may have to be found.

References

[1] https://www.statista.com/topics/1752/bars-and-nightclubs

[2] https://www.referenceforbusiness.com/business-plans/Business-Plans-Volume-07/Nightclub.html