Battle of the neighbourhoods

Opening a new Jazz Club in Manhattan

Author: Mark Chinnock
Date: June 2020

Content

1. Introduction

I am a successful Jazz musician wanting to move to New York City and open a new venue of my own in Manhattan. I want to locate my venue away from other clubs but in a similar neighborhood. I have a limited knowledge of the neighborhoods.

1.1 Background

Where is the best location to open a new Jazz Club in Manhattan? Determine similar neighborhoods to where existing Jazz venues are and find similar ones without a local Jazz Club.

1.2 Problem

The business understanding is to identify which neighborhood(s) are a best match for a location, based on:

- Finding where existing Jazz Clubs are currently located
- Discovering what type of neighborhood category they are in
- Find a similar neighborhood that doesn't currently have a Jazz Venue

1.3 Audience

The target audience for this would be anyone considering moving to Manhattan without detailed knowledge of the area. This report is a specific study on Jazz Club categories but this approach and the search criteria could easily be amended to repeat the exercise for a different category of venue.

2. Data

To investigate this problem the following datasets of information were required:

- List of neighborhoods in Manhattan
- Population density
- Location of existing Jazz Clubs and other venues
- Population density

Using the neighborhood data from the dataset previously provided on this course (https://cocl.us/new_york_dataset) we can obtain the geo coordinates of the Manhattan neighborhoods and store the following information:

```
Borough     Neighborhood     Latitude     Longitude
```

6 Manhattan Marble Hill 40.876551 -73.910660
100 Manhattan Chinatown 40.715618 -73.994279
101 Manhattan Washington Heights 40.851903 -73.936900

This data provides:

1. Borough name - eg Manhattan
2. Neighborhood - eg marble hill, highlights the area within Manhattan
3. Latitude - latitude coordinate for mapping
4. Longitude - longitude coordinate for mapping

Additionally, we can retrieve Population information for each neighborhood from the NYC open source: https://opendata.cityofnewyork.us/

This data can can read in using geopandas and then we are able to perform various mapping functions against the dataframe which may help with the investigation.

The neighborhood data will be compared using the venue data retrieved from foursquare to produce clusters of similar neighborhoods. The Jazz Club venues will then be placed in those clusters to see whether there is a pattern of cluster for a venue such as a jazz Club to help form a recommendation.

3. Methodology

In this project we directed our efforts on finding existing Jazz Clubs in the Manhattan borough of New York City, identify the neighborhoods where they reside, and then looking for neighborhoods with similar characteristics based on other venue types and population density, but with fewer Jazz Clubs (or hopefully none!)

In first step I collected the required data: location and type (category) of every Jazz Club within 10km from the centre of Manhattan. We have also identified all the venues within Manhattan currently held in the Foursquare database, which returned data as shown in the following sample.

	Borough	Neighborhood	Latitude	Longitude
6	Manhattan	Marble Hill	40.876551	-73.910660
100	Manhattan	Chinatown	40.715618	-73.994279
101	Manhattan	Washington Heights	40.851903	-73.936900
102	Manhattan	Inwood	40.867684	-73.921210
103	Manhattan	Hamilton Heights	40.823604	-73.949688

Table 1: sample data collected from foursquare

I used python folium library to visualize geographic details of Manhattan and its neighborhoods. G**eodata** and **choropleth maps** were used to identify population density (neighborhood area / population count recorded) in each neighborhood, and comparing which neighborhoods are similar but without Jazz Clubs.

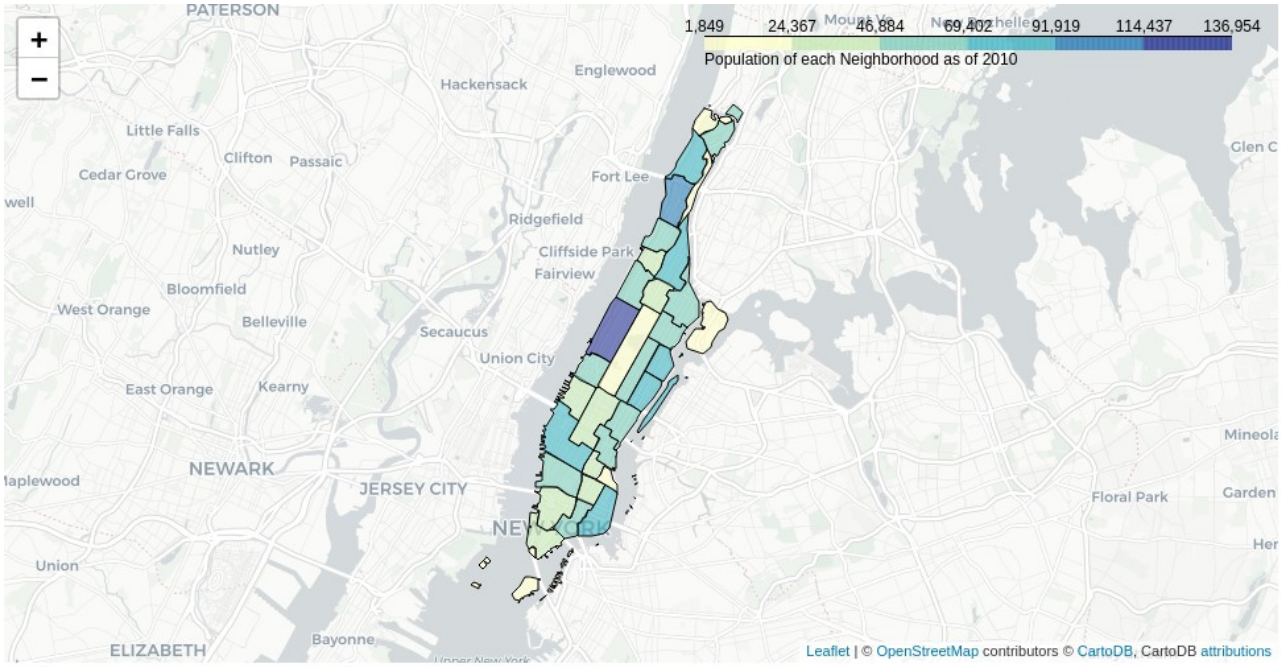

Illustration 1: Population density for each of the neighborhoods

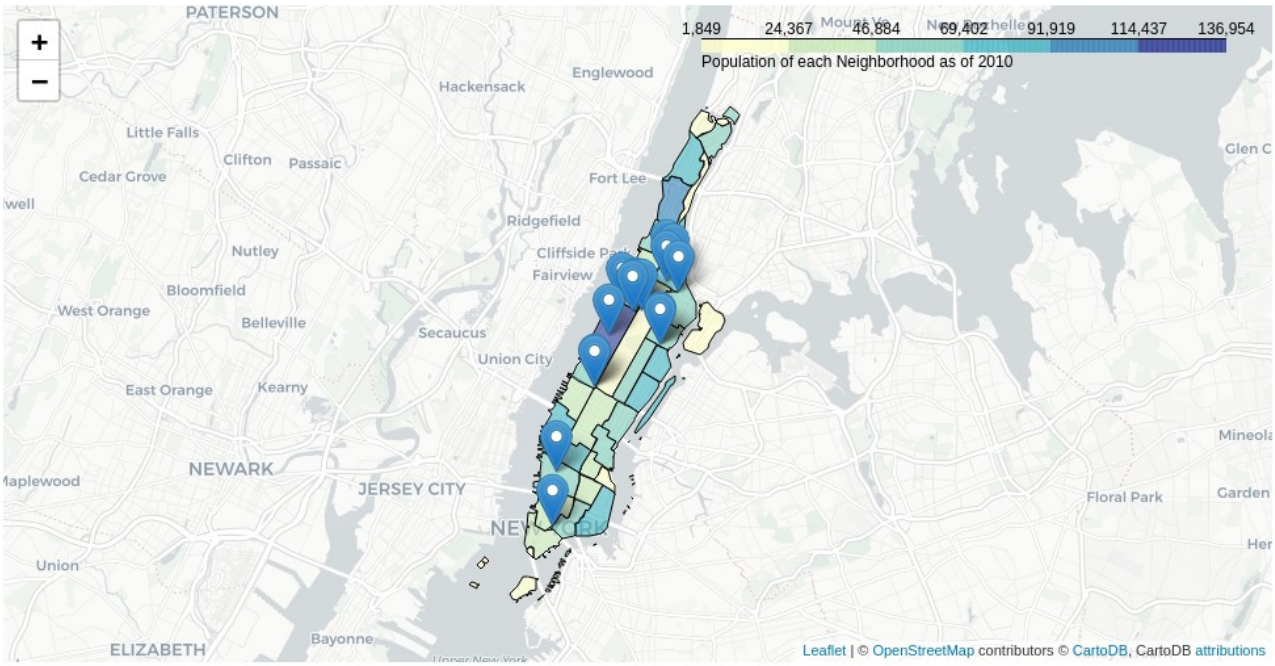

Illustration 2: Population density with Jazz Club location markers overlaid

The Jazz clubs are located in neighborhoods with very similar population density. The most densely populated neighborhood has the most Jazz clubs, so that would indicate this is an important feature that should not be overlooked.

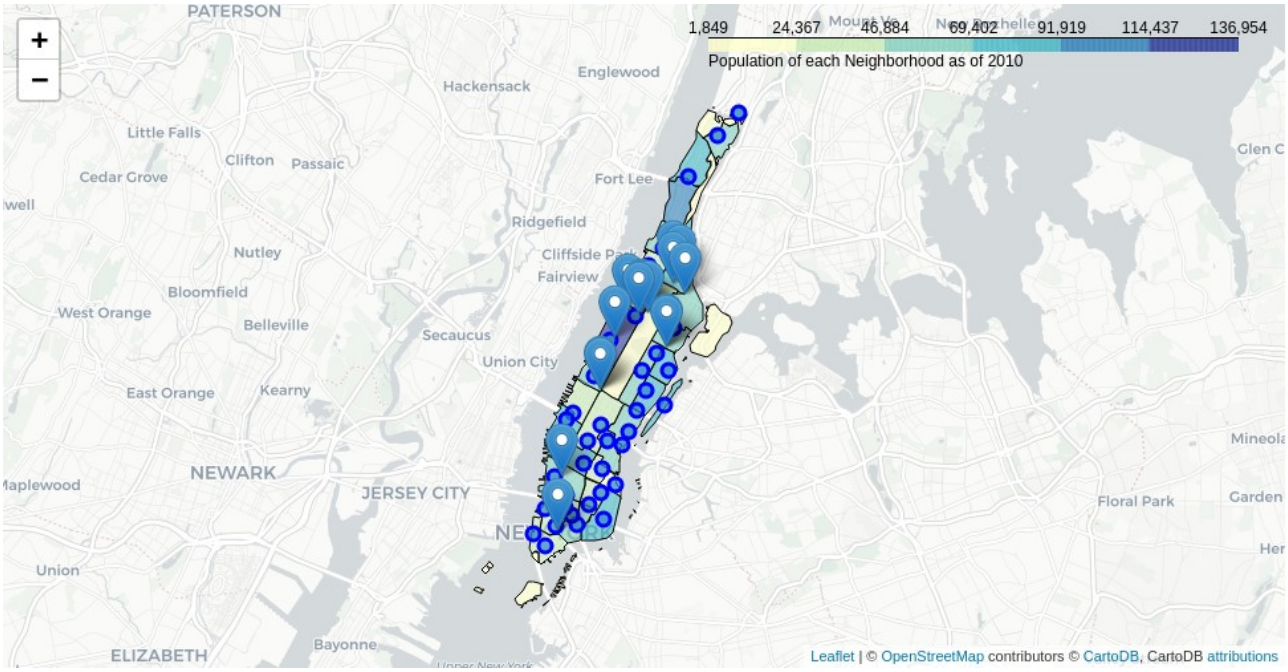

Illustration 3: neighborhood central points before clustering was applied

	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	Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site	Playground	Gourmet Shop	Food Court	Mexican Restaurant	Shopping Mall
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Bookstore	Gym / Fitness Center	Gym	Italian Restaurant	Pizza Place	Wine Shop	Vietnamese Restaurant
2	Central Harlem	African Restaurant	Chinese Restaurant	Seafood Restaurant	Bar	French Restaurant	Gym / Fitness Center	American Restaurant	Park	Cafeteria	Library
3	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	Café	Bakery	American Restaurant	Cocktail Bar	Theater	Italian Restaurant	Bar
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Spa	Bar	Ice Cream Shop	Coffee Shop	American Restaurant	Optical Shop

Table 2: table showing a sample of venue data collected for all the neighborhoods

Once all the venue data for all categories of venue was collected from foursquare API, this data was stored in a dataframe (above). We have some common venue categories in the neighborhoods so unsupervised learning K-means algorithm was chosen to cluster the data.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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
6	Manhattan	Marble Hill	40.876551	-73.910660	4	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Pharmacy	Supplement Shop	Steakhouse	Seafood Restaurant
100	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Spa	Bar	Ice Cream Shop	Coffee Shop
101	Manhattan	Washington Heights	40.851903	-73.936900	3	Café	Bakery	Mobile Phone Shop	Mexican Restaurant	Donut Shop	Latin American Restaurant	Supermarket	Tapas Restaurant
102	Manhattan	Inwood	40.867684	-73.921210	3	Lounge	Mexican Restaurant	Restaurant	Bakery	Café	Frozen Yogurt Shop	Spanish Restaurant	Caribbean Restaurant
103	Manhattan	Hamilton Heights	40.823604	-73.949688	3	Pizza Place	Coffee Shop	Deli / Bodega	Café	Mexican Restaurant	Sandwich Place	Sushi Restaurant	Cocktail Bar

Table 3: the data after cluster labelling has been applied

When we take a look at the specific neighborhoods containing Jazz venues we see they are all catergorized as cluster 0

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	9t Co
105	Manhattan	Central Harlem	40.815976	-73.943211	0	African Restaurant	Chinese Restaurant	Seafood Restaurant	Bar	French Restaurant	Gym / Fitness Center	American Restaurant	Park	C
107	Manhattan	Upper East Side	40.775639	-73.960508	0	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Bakery	French Restaurant	Spa	Juice Bar	Yoga Studio	Ar Res
117	Manhattan	Greenwich Village	40.726933	-73.999914	0	Italian Restaurant	Café	Sushi Restaurant	Bar	Dessert Shop	Seafood Restaurant	Caribbean Restaurant	Sandwich Place	
118	Manhattan	East Village	40.727847	-73.982226	0	Bar	Mexican Restaurant	Cocktail Bar	Coffee Shop	Pizza Place	Speakeasy	Wine Bar	Juice Bar	Ice
123	Manhattan	West Village	40.734434	-74.006180	0	Italian Restaurant	Wine Bar	American Restaurant	Pizza Place	Park	Jazz Club	Cocktail Bar	New American Restaurant	

Table 4: each neighborhood with an existing jazz club is cluster 0

The clustering label can be visually represented on a standard map with colours applied for each cluster:

Cluster 0 = red

Cluster 1 = purple

Cluster 2 = blue

Cluster 3 = green

Cluster 4= orange

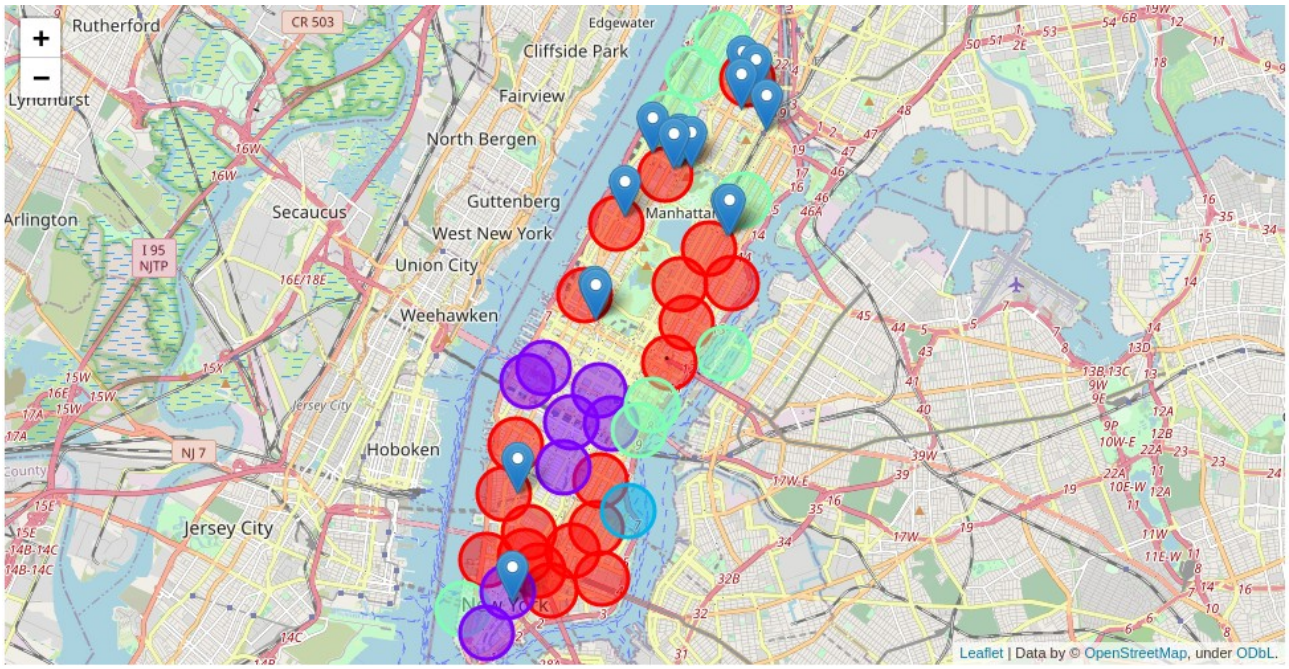

Illustration 4: neighborhoods after clustering has grouped and coloured similar neighborhoods

The map shows how all the existing Jazz venues are within cluter 0 (red) .

4. Results

The analysis shows that there are approximately 30 Jazz Club venues spread all around Manhattan with more centred around Manhattan Valley, near West 106th Street, and 3 near Central Harlem, around West 131-133rd Street. These are both towards the northern end of Manhattan.

All of the neighborhoods with existing Jazz Clubs were put into the cluster 0.

There are several cluster 0 neighborhoods with similar population density without an existing Jazz Club towards the east and south east of Manhattan (namely Lower East Side, Lenox Hill) with similar population density to the existing Jazz neighborhoods but are distanced from existing clubs.

5. Discussion

From the analysis, cluster 0 neighborhoods appear to be good candidates to locate a Jazz Club. However, there could be other factors as to why they haven't been chosen as venues before. Also, perhaps it would be better to locate a specialist music club such as a Jazz venue nearer to other existing clubs where it may benefit from an existing ambiance and clientele.

6. Conclusion

The purpose of this project was to identify a potential Jazz Club location in Manhattan in a neighhorhood where ther wasn't currently a Jazz Club. Additional information of characteristics of neighborhoods, population density and clustering of similar venue types together were also used to provide a scientific approach to the research.

The final decision on where to location a new Jazz Club would still require additional information and research, taking into account the general atmosphere of a location, closeness to major roads, rental prices, whether tourism numbers has an effect on such a venue, etc.