

A New Concept in Coffee, a New Concept in Jazz

**Analysis of New York City Neighborhoods in support of
A new business proposal**

Prepared by Charles Baldwin

September 2019

Table of Contents

1. Introduction
 - a. Background on the business proposal
 - b. Business problem
2. Data
 - a. Neighborhood data
 - b. FourSquare data
3. Methodology
 - a. Coffee Shop Density
 - b. Musical Venue Analysis
4. Results
5. Discussion
 - a. Observations
 - b. Recommendations
6. Conclusion

1. Introduction

Great coffee and great jazz are enjoyed around the world by people of all nationalities, and in New York City we are lucky to have some of the best places on earth to experience both.

Our clients, a group of jazz musicians and producers, understand this and have begun planning for a new kind of coffee experience in New York City. They have engaged us to execute neighborhood analysis within all five boroughs of the city to identify the most favorable areas for this venture.

a. Background on the business proposal

During the day the venue will operate as a regular gourmet coffee shop but with a jazz theme, with vintage memorabilia such as autographed photos and records on display, and jazz playing through a sound system. This isn't just another coffee shop however: although coffee will be the primary daytime business the venue will include a stage area in the back for jazz performances at night, when alcohol will be available in addition to coffee.

Not only will formal shows be staged, several times a week the venue will host "open mic" sessions, with a house band ready to serve as a backing group to play standards with customers who want to perform. My clients plan to have professional-grade audio/visual equipment to record these sessions so that customers can purchase a recording of their performance.

The belief is that by having a coffee shop as the main daytime purpose of the business, with a love of jazz as the focus, the shop/club will attract jazz-loving customers for the coffee and music and help build an audience for the evening shows.

b. Business Problem

While there are a number of professional jazz clubs in New York City there are few, if any, which include "open mic" opportunities and none that also focus on great coffee. The problem we seek to solve is which neighborhoods in New York would be the best candidates for such a unique venue.

The target demographic is jazz lovers, amateur jazz musicians, and jazz professionals from the New York metropolitan area and visiting from around the world. Our approach will be to identify neighborhoods with a higher concentration of music-related activity, but which would also be good locations for a business selling coffee.

2. Data

Since our goal is to identify neighborhoods where music is popular we will use FourSquare data on New York City. The underlying neighborhood data is a list of boroughs, neighborhoods and associated coordinates publically available from New York University.

By using FourSquare to highlight neighborhoods with music venues and other music-related businesses (e.g. record stores), then clustering them accordingly, we can identify areas where music is popular and musicians like to spend time. Once a target list is identified we will execute analysis of coffee shop density to identify an area within that list where there is a lower concentration of existing shops.

a. Neighborhood Data

The data set used to identify New York City neighborhoods is available here: NYC Neighborhood Data (https://geo.nyu.edu/catalog/nyu_2451_34572)

Within the New York City neighborhood data we will be using three primary fields. Here are the field names, associated data, and an example of each:

Field Name	Associated Data	Example
borough	Borough name	<i>Manhattan</i>
name	Neighborhood name	<i>Upper East Side</i>
coordinates	Neighborhood geographical coordinates	<i>[-73.96050763135, 40.775638573301805]</i>

Coordinates will be split into Latitude and Longitude; as such the initial neighborhood data set includes 306 neighborhoods and 4 columns.

b. FourSquare data

From the returned FourSquare data we will use the following fields for our project:

Field Name	Associated Data	Example
venue : location : lat	Venue latitude	<i>40.8769755336728</i>
venue : location : lng	Venue longitude	<i>-73.90675193198494</i>
categories : name	Venue Category	<i>Ice Cream Shop</i>

For our analysis we will focus on the following FourSquare categories:

- Record Shop
- Performing Arts Venue
- Jazz Club
- Concert Hall
- Music School
- Music Venue
- Nightclub
- Opera House
- Coffee Shop (to eliminate neighborhoods with high coffee shop density)

By merging the neighborhood and FourSquare data we can identify venue categories of interest within a certain radius of the neighborhood's central coordinates - we will set this radius as 800 meters to ensure we have a robust set of data to work with.

Once we have data on every neighborhood in New York City we will use k-clustering analysis to group the neighborhoods to help understand the types of nearby venues. This will allow identification of neighborhoods that have a concentration of music-related venues, and then within that list we will seek to identify those with fewer coffee shops.

3. Methodology

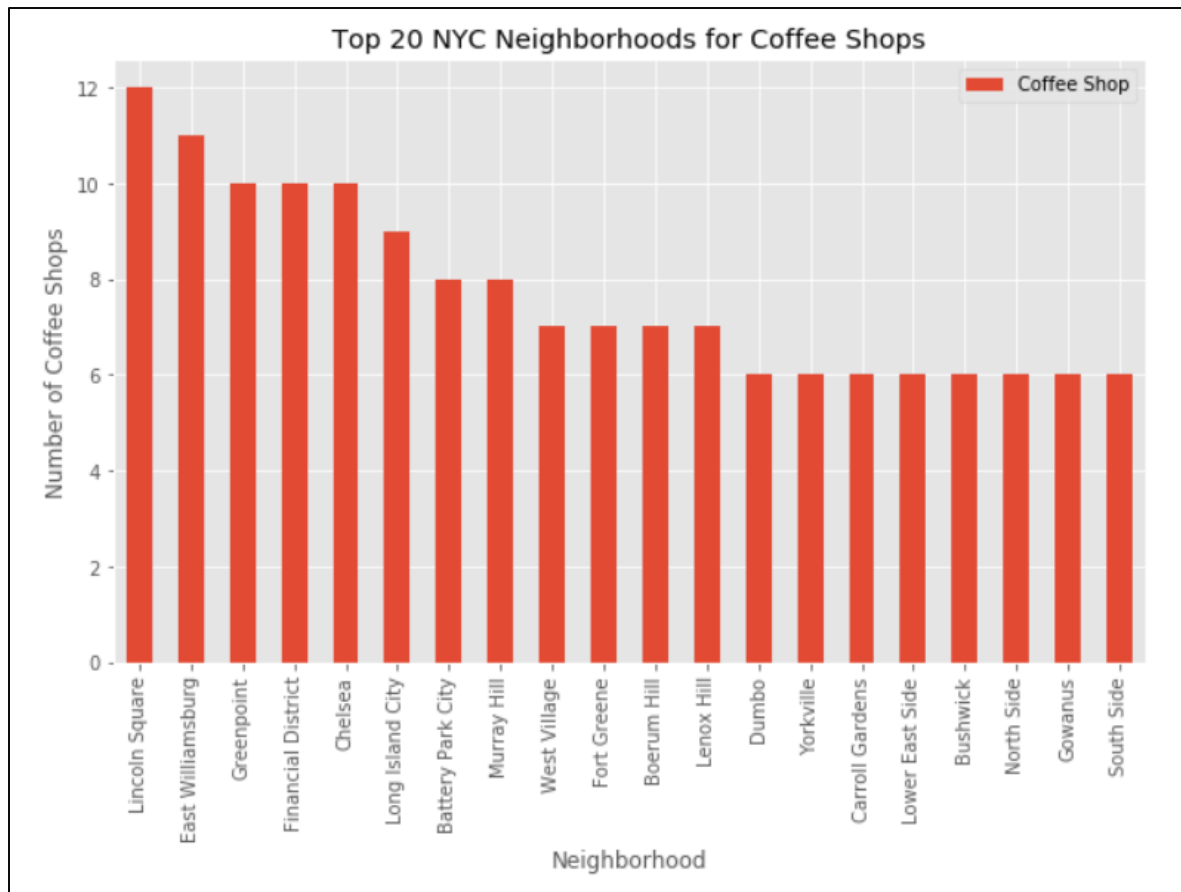
First I created a data frame for New York with all nearby venues, then created a new dataframe with only the categories of interest as noted above: 8 relating to music, 1 for coffee shops.

Next after normalizing the data using onehot encoding I split the data into two separate data frames, one just for music venues and another just for coffee shops to facilitate density analysis.

a. Coffee Shop Density

There are 384 coffee shops in New York City as returned by FourSquare using an 800 meter radius of neighborhood central coordinates. This data will be merged back with the top neighborhoods returned by the musical venue analysis.

Fig 1: Top 20 neighborhoods by coffee shop density



b. Musical Venue Analysis

Since there are not a large number of music-related venues in our data set I did some exploratory analysis. By varying the radius of the FourSquare search I was able to build a larger set of data for analysis without introducing overlap, finally settling on 800 meters.

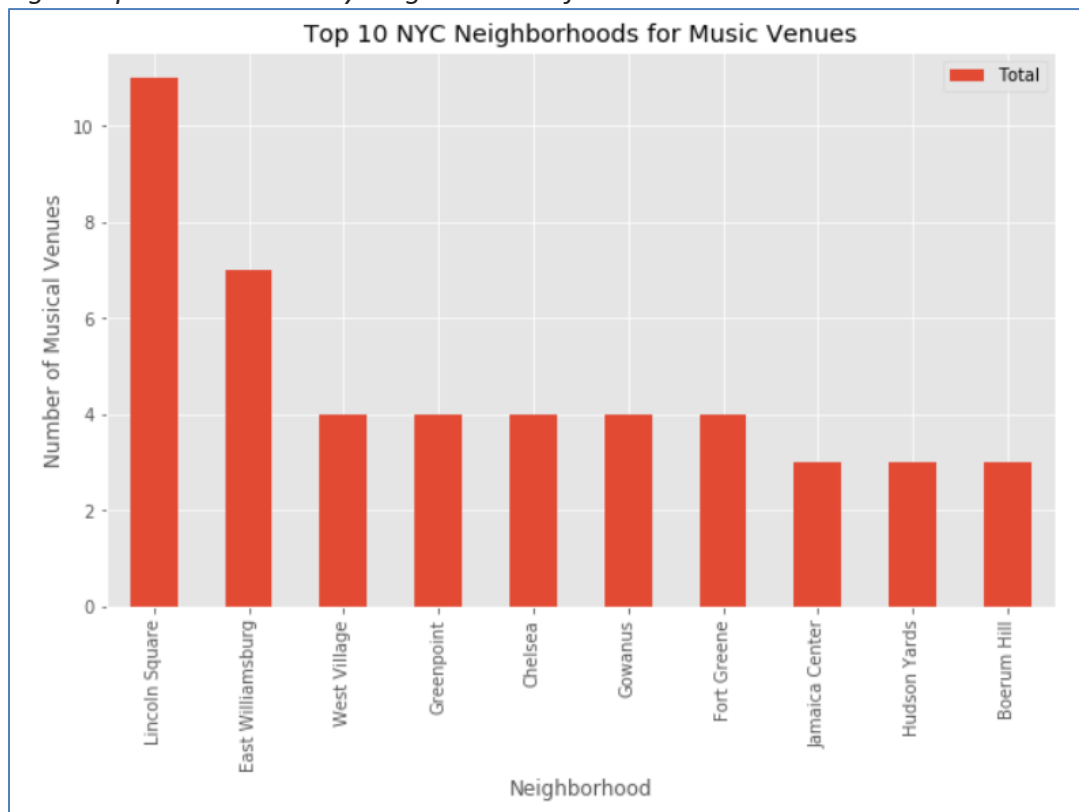
I then removed neighborhoods that have zero relevant data since they were no longer candidates for our search. This brought the total number of target neighborhoods down to 58.

The final New York City in-scope data set has:

- 58 neighborhoods
- 8 musical venue categories
- 108 total musical venues

Which neighborhoods have the highest number of musical venues? Let's check that data.

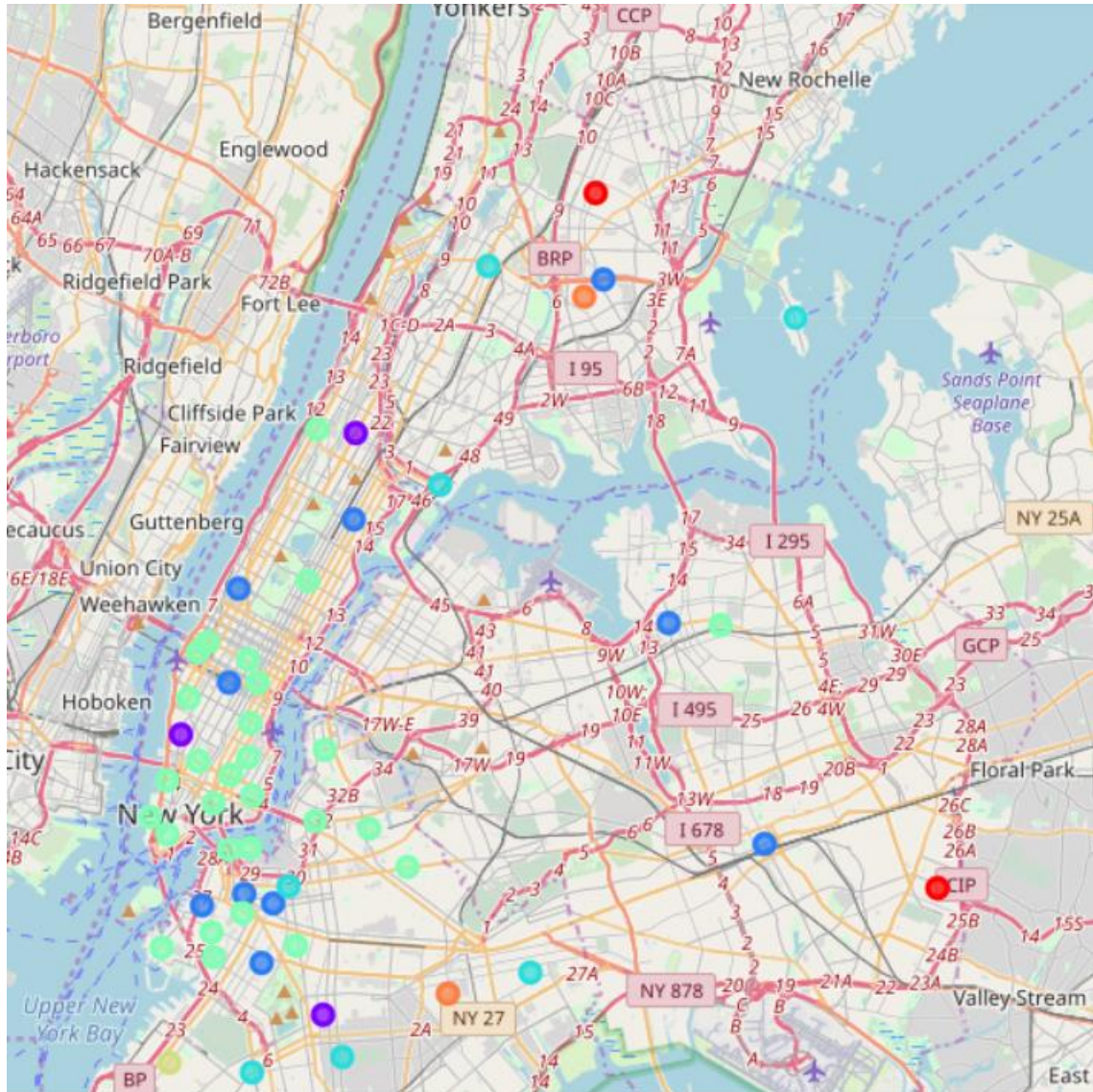
Fig 2: Top 10 New York City neighborhoods for Music Venues



With this final data set I used SSE analysis (Sum of Squared Errors) partnered with Silhouette scores to identify the ideal K for clustering. This analysis indicated that seven clusters would result in the best combination of low error and high silhouette score.

With that I ran the clustering using $k=7$ and mapped the results as shown below.

Fig 3: Map of New York City showing clusters



It's clear that most musical venues are in either Manhattan or Brooklyn.

I then examined the data for each cluster to identify neighborhoods that are good candidates for a jazz venue. The method used was to select neighborhoods where:

- Jazz Club or Night Club is the top venue
- Other relevant musical venues (such as Music Venue or Record Shop) are second or third most common venue.

This resulted in **13 target neighborhoods**.

I merged this target list with the number of coffee shops in each of the neighborhoods, then sorted to identify those with the lowest coffee shop density.

4. Results

The final number of target neighborhoods totaled thirteen: 7 in Manhattan, 4 in Brooklyn, 1 in Queens, and 1 in Staten Island. These represented a range of music venues and neighborhoods, any of which would be good candidates for a jazz venue. By including the number of coffee shops in each neighborhood we were able to rank the full list of targets:

Borough	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Number of Coffee Shops
Brooklyn	Manhattan Terrace	Jazz Club	Record Shop	Performing Arts Venue	2
Manhattan	Central Harlem	Music Venue	Jazz Club	Record Shop	2
Brooklyn	Prospect Lefferts Gardens	Record Shop	Music Venue	Jazz Club	3
Manhattan	Greenwich Village	Record Shop	Jazz Club	Performing Arts Venue	4
Manhattan	Upper East Side	Jazz Club	Record Shop	Performing Arts Venue	5
Manhattan	Hudson Yards	Nightclub	Music School	Concert Hall	5
Brooklyn	South Side	Nightclub	Music Venue	Record Shop	6
Brooklyn	Bushwick	Nightclub	Music Venue	Record Shop	7
Manhattan	Murray Hill	Jazz Club	Record Shop	Performing Arts Venue	8
Manhattan	West Village	Jazz Club	Performing Arts Venue	Record Shop	8
Queens	Murray Hill	Jazz Club	Record Shop	Performing Arts Venue	8
Manhattan	Chelsea	Nightclub	Record Shop	Performing Arts Venue	10
Staten Island	Chelsea	Nightclub	Record Shop	Performing Arts Venue	10

5. Discussion

a. Observations

There are only four neighborhoods with fewer than 5 local coffee shops, so to ensure enough variety, and to allow for my clients some room for personal preference or expertise, I selected the top 10: Six in Manhattan and four in Brooklyn.

Borough	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Number of Coffee Shops
Brooklyn	Manhattan Terrace	Jazz Club	Record Shop	Performing Arts Venue	2
Manhattan	Central Harlem	Music Venue	Jazz Club	Record Shop	2
Brooklyn	Prospect Lefferts Gardens	Record Shop	Music Venue	Jazz Club	3
Manhattan	Greenwich Village	Record Shop	Jazz Club	Performing Arts Venue	4
Manhattan	Upper East Side	Jazz Club	Record Shop	Performing Arts Venue	5
Manhattan	Hudson Yards	Nightclub	Music School	Concert Hall	5
Brooklyn	South Side	Nightclub	Music Venue	Record Shop	6
Brooklyn	Bushwick	Nightclub	Music Venue	Record Shop	7
Manhattan	Murray Hill	Jazz Club	Record Shop	Performing Arts Venue	8
Manhattan	West Village	Jazz Club	Performing Arts Venue	Record Shop	8

From this target list we can observe the following:

- In 4 neighborhoods a jazz club is the most common musical venue
- In 3 others a jazz club is 2nd or 3rd most common
- The remaining three neighborhoods have night clubs as the 1st most common venue, plus music venues or record shops as 2nd or 3rd.

The seven neighborhoods with an existing jazz venue would be great starting points, as there is already an indication that musicians and music lovers come to those neighborhoods. As the search expands the other neighborhoods would also be strong candidates.

b. Recommendation

My recommendation is to begin with the seven neighborhoods that have a local jazz club, then through internal discussion agree on the best 3-4 for a deeper dive. This should include a review of real estate availability, rent prices, accessibility to public transportation, and any personal preferences or requirements.

If a suitable candidate is not identified in the first round of analysis I would recommend that the next 3-4 candidates undergo the same kind of deep dive, and so on until a narrow list of candidate neighborhoods is identified. Once done my clients can engage local real estate professionals to begin identification of available locations within each neighborhood and actually start the process of setting up the business.

6. Conclusion

Considering the unique nature of this business proposal we had to take into consideration two distinct types of businesses and target audiences: Jazz lovers and coffee lovers. Although it's safe to say that coffee lovers can be found in any neighborhood, finding the right balance of low coffee shop density and relatively high density of jazz or music venues presented an interesting challenge.

The results of our analysis made it clear that although there are many fewer musical venues than coffee shops in New York City, there are enough musically-oriented neighborhoods to allow deeper review. Although we started with 306 New York City neighborhoods we were able to reduce that to 58 candidate neighborhoods with at least one musical venue. By executing KNN clustering we were able to review 7 distinct clusters and select 13 strong candidates for our clients. Selecting the top 10 based on coffee shop density gives them an excellent starting point for further analysis.

Additional data that could be considered for this analysis includes demographics within New York City (incomes, age groups, etc), average rent costs and real estate turnover data, plus detailed information on coffee consumption by neighborhood in New York City.