Tuning-In To NYC's Most Musical Neighborhoods

Michael Cascio

2019/04/02

DRAFT
DRAFT
DRAFT
DRAFT
DRAFT

Table of Contents

Abstract

Machine learning allows for the creation of computational models capable of identifying patterns in multidimensional datasets. This project aims to leverage location data and machine learning algorithms to identify the most musical neighborhoods in New York City.

Introduction

Background

Music is a form of art that has, and probably always will be, deeply embedded within the cultural activity of cities, communities, and groups of people more generally. Music is a means of communication, expression, and sometimes even protest with the power to peacefully bring together large amounts of like-minded people, influence popular culture, and hypnotize you with a memorable lyric that you end up singing in the shower subliminally for weeks on end even after consciously being disappointed in yourself for doing so…..I digress....

Problem

Cities are, in part, composed of musical entities such as record shops, instrument vendors, concert halls, amphitheaters, and more, that not only provide to the music needs of local citizens but also to tourists from around the world. For bigger cities, music entities can be spread apart, resulting in an ecosystem of hip niche neighborhoods that evolve and change over time. This ecosystem is often learned by humans looking for a cool music scene through either natural life experience (wandering/flaneur) or recommendations in the form of internet reviews, comments, and conversations with people in-real-life.

This project aims to quantify and monitor the state of neighborhoods in a major metropolitan city, New York City, to identify clusters of similar music scenes.

Interested Stakeholders

Different parties may be interested in a model that is able to quantify neighborhood similarity based on music venue frequency and thus group neighborhoods of similar music profiles. Such a model would be able to inform renters and home buyers who prefer to live where the music is happening that they're next home is properly located. Future music venue start-ups can utilize the model to identify neighborhoods lacking live music venues and ensure they are investing in an area that is not saturated. Future music retail vendors, sellers of things like records and instruments, can similarly utilize the model to ensure they are launching a business where competition is in their favor.

Methodology

Data Sources

NYU Spatial Data Repository

I am using the '2014 New York City Neighborhood Names' dataset hosted by NYU's Spatial Data Repository as the basis for the neighborhood names and associated location centroids [0]. The image to the right shows a sample of this information:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Foursquare - 'Places API'

I will be using Foursquare's 'Places API' to acquire data related to 'venues' (as defined by Foursquare) categorized to be somehow associated with music [1]. It is important to note that Foursquare defines a 'venue' as a place that one can go to, or check-in to, and that a 'venue' is not necessarily a music venue but can be any establishment such as a restaurant or type of retail shop. Each Foursquare 'venue' is assigned a 'Category' and each 'Category' is associated with a particular 'CategoryID'. The image to the right shows the 'CategoryID' values provided by Foursquare that will be used to acquire music related venues within New York City:

Foursquare Music-Related Venue CategoryIDs

```
Music Venue = '4bf58dd8d48988d1e5931735'

Jazz Club = '4bf58dd8d48988d1e7931735'

Piano Bar = '4bf58dd8d48988d1e8931735'

Rock Club = '4bf58dd8d48988d1e9931735'

Concert Hall = '5032792091d4c4b30a586d5c'

Amphitheater = '56aa371be4b08b9a8d5734db'

Music Festival = '5267e4d9e4b0ec79466e48d1'

Nightlife Spot = '4d4b7105d754a06376d81259'

Music Store = '4bf58dd8d48988d1fe941735'

Record Shop = '4bf58dd8d48988d10d951735'

Recording Studio = '52f2ab2ebcbc57f1066b8b37'
```

Data Cleaning

Feature Selection

Exploratory Data Analysis

Results

Predictive Modeling

Discussion

Conclusion

References

[0] - [2014 New York City Neighborhood Names - NYU Spatial Data Repository](#)

[1] - ['Places API' Documentation - Foursquare](#)