

Dan Weber
Capstone Data Science
7/8/2020

Introduction/Business Problem:

When researching and compiling information to use for direction in where to place a venue, location data and existing venue data is paramount to your success. The audience for this data analysis will be useful for prospective business owners and investors looking to establish venues in Buffalo, NY, namely concert halls.

The business problem introduced here is that Buffalo, NY has many established concert halls already, so location of these is needed, as well as how close they are to one another, which neighborhoods contain multiple venues, and of course which areas contain none (potential for establishing a prominent venue in such a location).

By using available and wrangled data for Buffalo NY, we will parse the data and explore it in order to draw conclusions on ideal locations for our new venue. By doing this, we create a useful guide to address the issues faced by our audience. Where is the best location? How many venues of our kind already exist? Is there a market for us here? These questions and others will be summarized in this data analysis project.

Data Section:

Using data about Buffalo NY which has been wrangled and cleaned, we will explore the neighborhoods of this city and use Foursquare data and location data to pinpoint the neighborhoods as well as the venues nearby.

By creating a pandas dataframe from the data wrangled from "https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Buffalo,_New_York", the list of neighborhoods in Buffalo NY can be organized and put into columns showing name, latitude, longitude, and therefore can be used with Foursquare venue data in order to compare and contrast these neighborhoods statistics for our purposes.

Using these sources of data, we will create visual aids such as maps with Folium, as well as cluster the neighborhood data, giving us the best representation of the available data. By doing so, we can address our business problem fully and effectively.

Methodology:

(Choosing city)

The data analysis process began with the choosing of our target location (city) for a new music venue project, the city in this case being

Buffalo, NY. Once the city is located, the neighborhoods which exist and make up the land included in the city need to be identified and mapped.

(Identifying neighborhoods)

The next step taken is to identify and find the latitude and longitude coordinates for the neighborhoods in Buffalo. A list of neighborhoods and their coordinates was found (source in data section) and this data was requested.

(Getting data)

Using BeautifulSoup, a request was sent to retrieve the data and parse it. Specifically, just the table containing our neighborhood names.

(Data into dataframes)

Using Pandas, the data retrieved using BeautifulSoup now gives us the ability to display the neighborhoods in columns with their corresponding coordinates (lat, long)

(Venue data)

By making calls to the Foursquare API, information on venues located in the neighborhoods of Buffalo can be retrieved for use in our data analysis. This will give us a lot of data to work with, but what's needed most is their coordinates (lat, long) and their type, to narrow down the data to music venues and concert halls.

(New dataframes)

Now that there is added data to what is being used, it needs to be imported into a useable and simple dataframe. The new dataframe will include neighborhood list, their lat/long coords, venues in each neighborhood, and those venues lat/long as well.

(Sort into specific useable data)

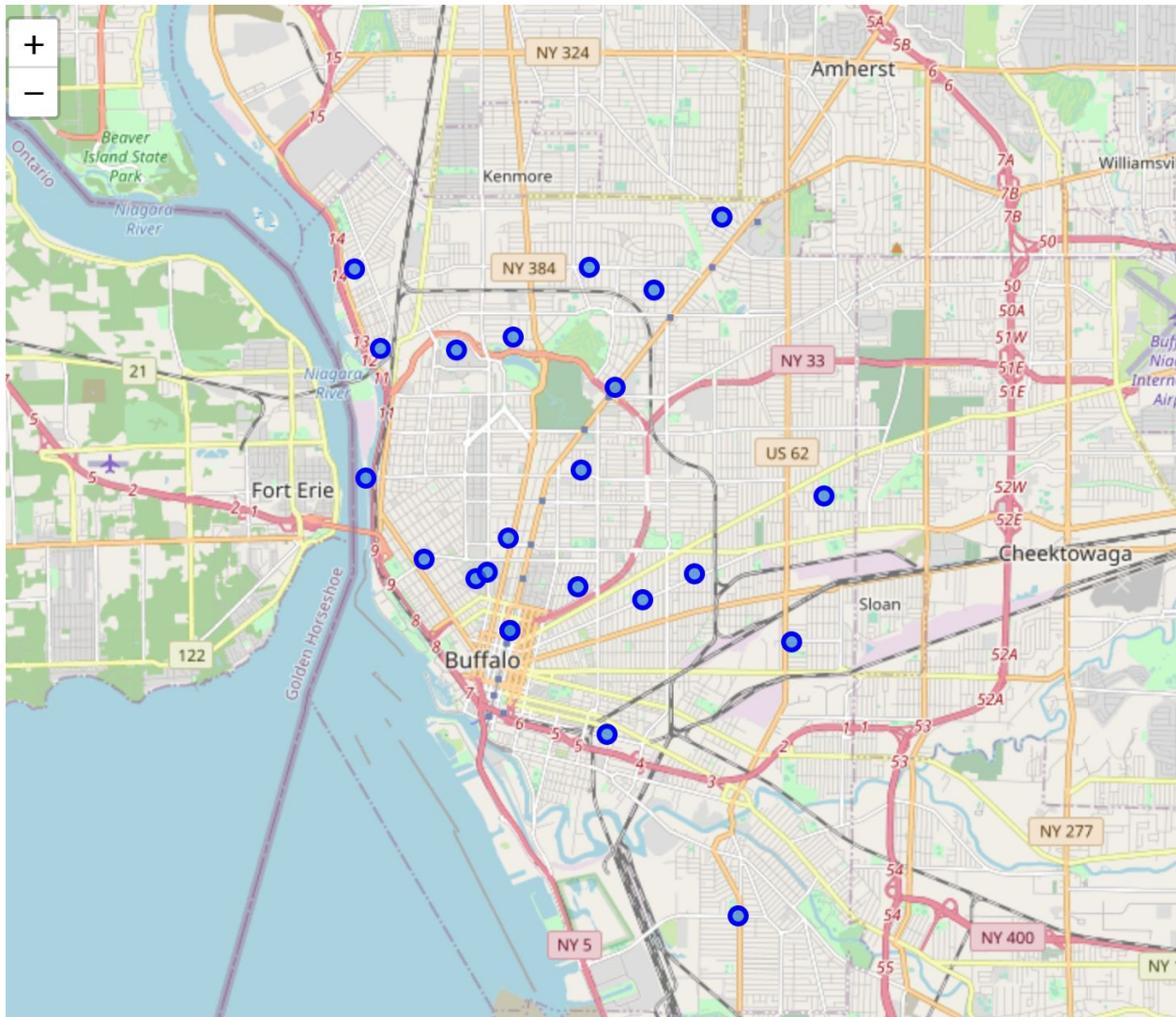
The data needed to decide on a location for our music venue requires a sorting by venue type to narrow it down to concert halls, performing arts centers, and music venues. Using results from these three types show all the data needed for analysis.

(Create visuals and maps)

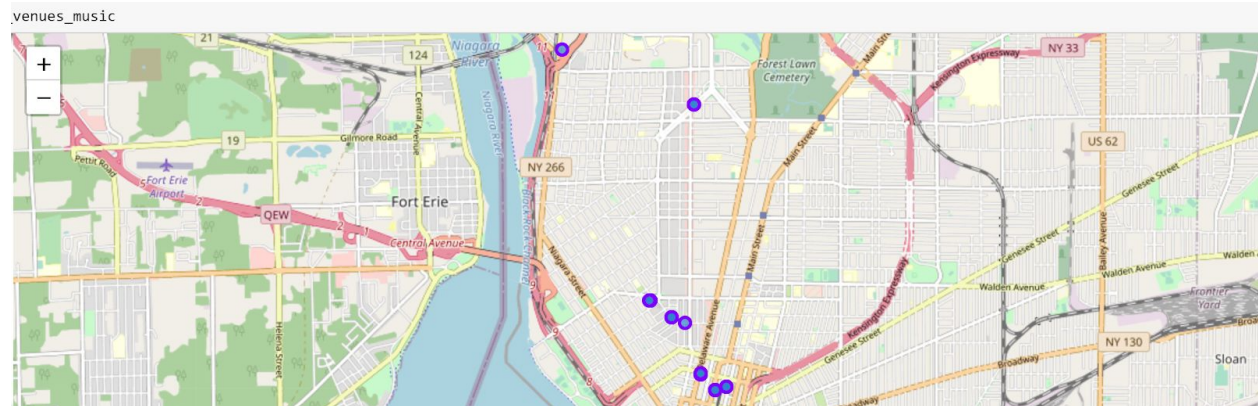
To better assess the data, create multiple visuals.

(Display data easily and make easily readable)

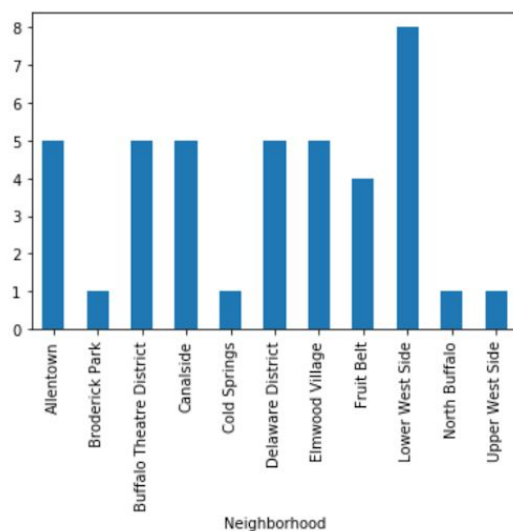
Results: (map, graph, data retrieved)



Map created using Folium, displaying Buffalo's neighborhoods used in data analysis



A second map, once the specific venues were narrowed down to display music venue locations.



From this we can spot that Upper West Side, Cold Springs, North Buffalo, and Broderick Park have the least of our venue type

Graph showing the neighborhoods, sorted by music venue location proximity.

Discussion:

The results from this data analysis have revealed valuable information for potential business owners and investors alike when considering one of Buffalo's neighborhoods for a new music venue to be opened.

By obtaining data about each music venue and each neighborhood, it was then possible to draw several conclusions, the first being which areas had the least or most venue occurrences and proximity to neighborhoods. In this case, Upper West Side, Cold Springs, North Buffalo, and Broderick Park have 1 or less venues within their respective neighborhood's land area. Now these four neighborhoods can be further researched and analyzed themselves using another data set or another data analysis, this time using population data, census data, and more to take the next step in planning.

Conclusion:

There are several neighborhoods in Buffalo which fit the description for the ideal place to open a music venue. They are lacking in music venue locations already, as well as including music venues with lower ratings or limited range of music acts (such as a philharmonic or opera house, club, etc). Therefore, a music venue which is designed to be multi-purpose, housing touring bands as well as solo singers, djs, and the majority of popular artists, would have a high appeal to those living or near its location.

Notes

-Some music venues carried mislabeled “venuetypes” during data analysis. This was addressed in my coding, but potential for loss of a small amount of venue data possible-