Finding Similar Neighborhoods Based on Most Common Type of Venue

Allison Wu

January 19, 2021

1. Introduction

1.1. Background

Location data can be very useful for people moving to a new city to find an area that is very similar to their old neighborhood in their old city. They can know before they move that their new area has resources and venues that will interest them and help them adjust to their new environment more quickly because it is actually familiar and comfortable rather than new and foreign. This is of personal interest to this data scientist who is looking to relocate to a new city soon.

For someone who currently lives in one city, can we find the most similar neighborhood in another city based on the kinds of venues that are most common in that neighborhood in order to ease the relocation transition?

1.2. Problem

In this project, we will explore this question - specifically for someone who currently lives in Charleston, South Carolina, is moving to Pittsburgh, Pennsylvania, and is interested in neighborhood similarity based on the number of craft breweries in that neighborhood. Additionally, we will look at median home values in those neighborhoods to better inform our decision about where to relocate with a home-buying budget of $300,000.

These two cities, on the surface, are very different. Charleston is a coastal city located in the Southeastern United States. Peak summer temperatures average 91°F (32.78°C), while average low temperatures in the winter rarely reach below 38°F (3.33°C) (US Climate Data, 2021). According to the US Census Bureau, the estimated population in July 2019 was 137,566 with an estimated 55,889 households (US Census Bureau, 2019). Charleston's demographics were estimated to be 74.1% White, 21.7% Black/African American, 1.9% Asian, and 3.2% Hispanic/Latino (US Census Bureau, 2019). There were also 2.3 craft breweries per capita (Conway, J., 2020). The median home value in the Charleston metropolitan area is $348,052 (Zillow, 2020).

On the other hand, Pittsburgh is a land-locked city located in the Northeastern United States. Peak summer temperatures average 83°F (28.33°C), while average low temperatures in the winter can plunge as cold as 21°F (-6.11°C) (US Climate Data, 2021). According to the US Census Bureau, the estimated population in July 2019 was 300,286 with an estimated 138,058 households (US Census Bureau, 2019). Pittsburgh's demographics were estimated to be 66.8% White, 23.0% Black/African American, 5.8% Asian, and 3.2% Hispanic/Latino (US Census Bureau, 2019). There were also 4.1 craft breweries per capita (Conway, J., 2020). The median home value in the Pittsburgh metropolitan area is $195,195 (Zillow, 2020).

Though they seem very different, our goal is to find similarity between individual neighborhoods in each city. In this example, we already know the neighborhoods of Charleston very well, as well as which neighborhoods have the highest density of craft breweries of interest. Our goal is to compare unknown Pittsburgh neighborhoods to known Charleston neighborhoods to find an area that one is more accustomed to, which will make adjusting to the new city easier.

2. Data

2.1. Data Sources & Acquisition

Based on our business problem and specific criteria to be looked at, we needed data for the following factors:

- Neighborhoods in Charleston, SC
- Neighborhoods in Pittsburgh, PA
- The number of breweries per neighborhood in Charleston, SC
- The number of breweries per neighborhood in Pittsburgh, PA
- Median home values per neighborhood in Charleston, SC
- Median home values per neighborhood in Pittsburgh, PA

First, we obtained median home values for specific neighborhoods in each city from Zillow's publicly available database (Zillow Home Value Index, 2020). Under Housing Data, we exported the data base for ZHVI Single-Family Homes Time Series ($) based on Neighborhood. It was exported & automatically downloaded in .csv format.

Second, we obtained a JSON list of 100 venues within 1 mile (1609.34 meters) of each neighborhood center's geospatial coordinates from the Foursquare API. We looked at venues categorized as breweries and calculated the frequency of occurrence of this kind of venue in each neighborhood. This is how we decided which Pittsburgh neighborhood was most similar to which Charleston neighborhood.

2.2. Data Cleaning & Processing

As mentioned above, data extracted from Zillow's research data was downloaded as a .csv file. This file was imported to IBM Data Cloud and converted to a Pandas dataframe, then the data was cleaned to only include the relevant neighborhoods in either Charleston, SC or Pittsburgh, PA. We found 26 neighborhoods in Charleston (Figure 1a) and 77 neighborhoods in Pittsburgh (Figure 1b). Using the Geopy package Nominatim, latitude and longitude were found for each neighborhood's center. In the chs_zhvi dataframe, we initially found two neighborhoods that were misspelled, so they did not return any geospatial data. We also found one neighborhood in the pit_zhvi dataframe that was misspelled. We corrected these misspellings and found the latitude and longitude for these neighborhoods.

In order to check the accuracy of Nominatim's coordinate assignments, we used Folium to map and visualize each neighborhood within each city. Based on these visualizations, we found that 13 neighborhoods in Charleston (Figure 2a) and 14 neighborhoods in Pittsburgh (Figure 2b) were assigned incorrect coordinates. This may be because the neighborhoods are too small to be added to Nominatim or the neighborhood names are too specific to the city and are named for another, more well-known region. For example, the French Quarter, Charleston, SC was assigned the coordinates to the French Quarter, New Orleans, LA. We found the northeast and southwest bounds of each city using the Geopy package GoogleV3, then compared the assigned coordinates of each neighborhood to determine which neighborhoods were assigned out-of-bounds coordinates. For these neighborhoods, we manually researched and updated the coordinates to be within bounds of each respective city, then visualized the maps again (Figure 3a & Figure 3b).

Once the appropriate coordinates were assigned to each neighborhood, we were able to extract JSON data for each neighborhood from the Foursquare API to find the closest 100 venues within a one-mile (1609.34 meter) radius. From this JSON data, we were able to further extract venue name, venue category, venue latitude, venue longitude into another dataframe, grouped by neighborhood.

3. Methodology

After ensuring all geographic coordinates were correctly assigned, we used those coordinates to anchor each neighborhood. We defined a function to pull out the venue name, category, latitude, & longitude from the Foursquare API for the first 100 venues within a one-mile radius of the neighborhood's coordinates. We found 1,711 venues in the Charleston neighborhoods and 4,451 venues in the Pittsburgh neighborhoods. We then took each venue category and created dummy-codes for each possible category to label each venue. For Charleston venues, we found 188 different categories. For Pittsburgh venues, we found 285 different categories.

Next, we grouped all venues by neighborhood and calculated the frequency of each venue category by finding the mean of that venue category occurrence per neighborhood. Since our venue category of interest was brewery, we pulled out the column with brewery frequency, as well as neighborhood name into a new dataframe. This new dataframe was then merged with the Zillow Home Value Index dataframe for each city. Each merged dataframe was then sorted by brewery frequency in descending order, so that the neighborhood with the highest frequency of breweries within a one-mile radius of neighborhood center was at the top of the dataframe. From there, we visually compared brewery venue frequency to match neighborhoods across Charleston and Pittsburgh to find the most similar neighborhoods based on this venue criteria.

4. Results

Our analysis shows that the Charleston neighborhood with the highest frequency of brewery as a venue within a one-mile radius of city center is Silver Hill-Magnolia, with a frequency of 0.24. This means that almost a quarter of the venues in this neighborhood are categorized as a brewery. At the end of 2020, this neighborhood had a median home value of $205,635.00 (Figure 4a).

Our analysis also shows that there are two Pittsburgh neighborhoods with the highest frequency of brewery as a venue within a one-mile radius of city center with a frequency of 0.125 - Spring Garden and Upper Lawrenceville. At the end of 2020, these neighborhoods had median home values of $129,360.00 and $251,403.00, respectively (Figure 4b).

However, the intention of this project was to find the neighborhoods which are most similar between Charleston and Pittsburgh based on the frequency of breweries as a venue. Based on this criteria, Silver Hill-Magnolia is not very similar to Spring Garden or Upper Lawrenceville. We can take another look at Charleston neighborhoods and find that East Central has a frequency of breweries as a venue of 0.113 (Figure 4a). This is still not a perfect match but is more similar to the frequency of 0.125 found in Spring Garden & Upper Lawrenceville. Although we do not need the median home value in East Central to inform our home-buying decision in Pittsburgh, we can see that the median home value at the end of 2020 in East Central was $343,867.00.

5. Discussion

For this project's purpose, it would be most helpful to see what current (January 2021) median home values are in the identified neighborhoods to be as precise in informing our relocation decision as possible, but that data is not yet available from Zillow. There was also a bit of manual data cleaning that was involved in our process because home neighborhoods are more specific and narrowly bounded than neighborhoods for venues, based on Foursquare's API data. In order to compare as many neighborhoods as possible and visualize breweries all across each city, the decision was made to manually enter the city center coordinates for each neighborhood that was not defined by the Foursquare API. This process may be easier for bigger/more populous/more tourist-centric cities because the data may be more robust.

For future projects and to better inform others that are looking at relocating to an unfamiliar city, additional analysis on neighborhoods can be done with other factors, such as school districts, proximity to grocery stores, and/or proximity to a known workplace (if relocation is due to job change), to help narrow down the ideal neighborhood(s) even more. This project was very narrowly focused based on this data scientist's personal interests and hobbies, so the similarity criteria was very specific. To suit a broader stakeholder base, other relevant data can be sourced and used in a similar manner to compare neighborhoods. If multiple factors are being used to inform the decision, a statistical test, such as ANOVA, may be of interest to statistically and objectively analyze similarity rather than rough comparisons.

6. Conclusion

The purpose of this project was to find the most similar neighborhood(s) between Charleston, SC and Pittsburgh, PA based on the number of breweries in each neighborhood. A secondary purpose of this project was to evaluate median home value in each of those neighborhoods to inform and narrow down our search for a new home while relocating from Charleston to Pittsburgh. Our analysis consisted of taking neighborhoods for which we have home value data and examining the venues within those neighborhoods to identify the frequency of breweries within each neighborhood. Once we found which neighborhoods had the most breweries, we could easily compare the neighborhoods from Charleston and Pittsburgh to find which ones were most similar/had the closest frequency of breweries per neighborhood.

Based on our analysis, we can say that the Spring Garden and Upper Lawrenceville neighborhoods in Pittsburgh are most similar to the Silver Hill-Magnolia neighborhood in Charleston. We can also easily compare median home values in these neighborhoods. We can see that median home values in Spring Garden ($129,360.00) and Upper Lawrenceville ($251,403.00) fall well below our home-buying budget of $300,000.

Thus, we can conclude that moving from Charleston to either Spring Garden or Upper Lawrenceville will help a craft beer enthusiast feel at home in their new surroundings more quickly

because of the greater frequency of a venue of interest. We can also conclude that moving to either of these neighborhoods will be affordable for this relocator since neither neighborhood's median home value exceeds the predetermined home-buying budget.

References

Conway, J. (2020, June). *Craft beer breweries per capita in the U.S. by state, 2019.* https://www.statista.com/statistics/319978/craft-beer-breweries-per-capita-in-the-us-by-state/

United States Census Bureau (2019). *U.S. Census Bureau QuickFacts: Charleston city, South Carolina.* https://www.census.gov/quickfacts/fact/table/charlestoncitysouthcarolina/PST045219

United States Census Bureau (2019). *U.S. Census Bureau QuickFacts: Pittsburgh city, Pennsylvania.* https://www.census.gov/quickfacts/fact/table/pittsburghcitypennsylvania/PST045219

U.S. Climate Data (2021). *Climate Charleston AFB – South Carolina and Weather Averages Charleston.* https://www.usclimatedata.com/climate/charleston-afb/south-carolina/united-states/ussc0052

U.S. Climate Data (2021). *Climate Pittsburgh – Pennsylvania and Weather Averages Pittsburgh.* https://www.usclimatedata.com/climate/pittsburgh/pennsylvania/united-states/uspa3601

Zillow, Inc. (2021). *Charleston SC Home Prices & Home Values.* https://www.zillow.com/charleston-sc/home-values/

Zillow, Inc. (2021). *Housing Data – Zillow Research*. https://www.zillow.com/research/data/

Zillow, Inc. (2021). *Pittsburgh PA Home Prices & Home Values.* https://www.zillow.com/pittsburgh-pa/home-values/

Appendix

Figure 1a. The first five entries in dataframe chs_zhvi, which shows neighborhood names and median home values.

In [4]:
```
chs_zhvi=zillow_zhvi[zillow_zhvi['City']=='Charleston'].reset_index(drop=True)
chs_zhvi=chs_zhvi[chs_zhvi['State']=='SC'].reset_index(drop=True)
chs_zhvi=chs_zhvi.loc[:,['RegionName','State','City','Metro','2020-11-30']]
chs_zhvi.rename(columns={'2020-11-30':'Median Home Value 2020-11'},inplace=True)
chs_zhvi['neighborhood']=chs_zhvi[['RegionName', 'State']].agg(', '.join, axis=1)
print('there are',chs_zhvi.shape[0],"neighborhoods in Zillow's Home Index database for Charleston, SC")
chs_zhvi.head()
```

```
there are 26 neighborhoods in Zillow's Home Index database for Charleston, SC
```

Out[4]:

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood
0	Harleston Village	SC	Charleston	Charleston-North Charleston	892637.0	Harleston Village, SC
1	Daniel Island	SC	Charleston	Charleston-North Charleston	876967.0	Daniel Island, SC
2	Cannonborough-Elliottbororugh	SC	Charleston	Charleston-North Charleston	568652.0	Cannonborough-Elliottbororugh, SC
3	Shadowmoss	SC	Charleston	Charleston-North Charleston	314499.0	Shadowmoss, SC
4	Wagener Terrace	SC	Charleston	Charleston-North Charleston	564916.0	Wagener Terrace, SC

We extracted information for 26 neighborhoods in Charleston.

Figure 1b. The first five entries in dataframe pit_zhvi, which shows neighborhood names and median home values.

In [5]:
```
pit_zhvi=zillow_zhvi[zillow_zhvi['City']=='Pittsburgh'].reset_index(drop=True)
pit_zhvi=pit_zhvi[pit_zhvi['State']=='PA'].reset_index(drop=True)
pit_zhvi=pit_zhvi.loc[:,['RegionName','State','City','Metro','2020-11-30']]
pit_zhvi.rename(columns={'2020-11-30':'Median Home Value 2020-11'},inplace=True)
pit_zhvi['neighborhood']=pit_zhvi[['RegionName', 'State']].agg(', '.join, axis=1)
print('there are',pit_zhvi.shape[0],"neighborhoods in Zillow's Home Index database for Pittsburgh, PA")
pit_zhvi.head()
```

```
there are 77 neighborhoods in Zillow's Home Index database for Pittsburgh, PA
```

Out[5]:

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood
0	Mount Lebanon	PA	Pittsburgh	Pittsburgh	354231.0	Mount Lebanon, PA
1	Squirrel Hill South	PA	Pittsburgh	Pittsburgh	442777.0	Squirrel Hill South, PA
2	Shadyside	PA	Pittsburgh	Pittsburgh	612067.0	Shadyside, PA
3	Brookline	PA	Pittsburgh	Pittsburgh	163387.0	Brookline, PA
4	Squirrel Hill North	PA	Pittsburgh	Pittsburgh	669792.0	Squirrel Hill North, PA

We extracted information for 77 neighborhoods in Pittsburgh.

Figure 2a. First map visualization of Charleston neighborhoods which shows some neighborhood outside of city bounds.

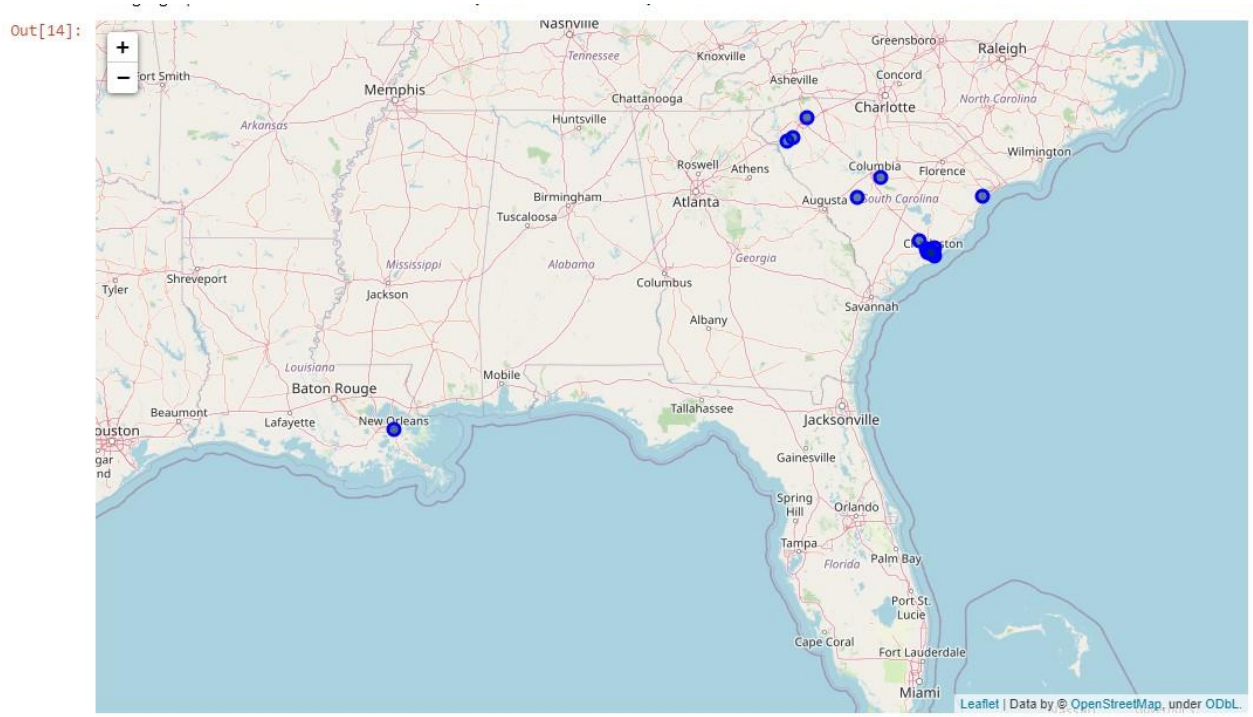

Figure 2b. First map visualization of Pittsburgh neighborhoods which shows some neighborhood outside of city bounds.

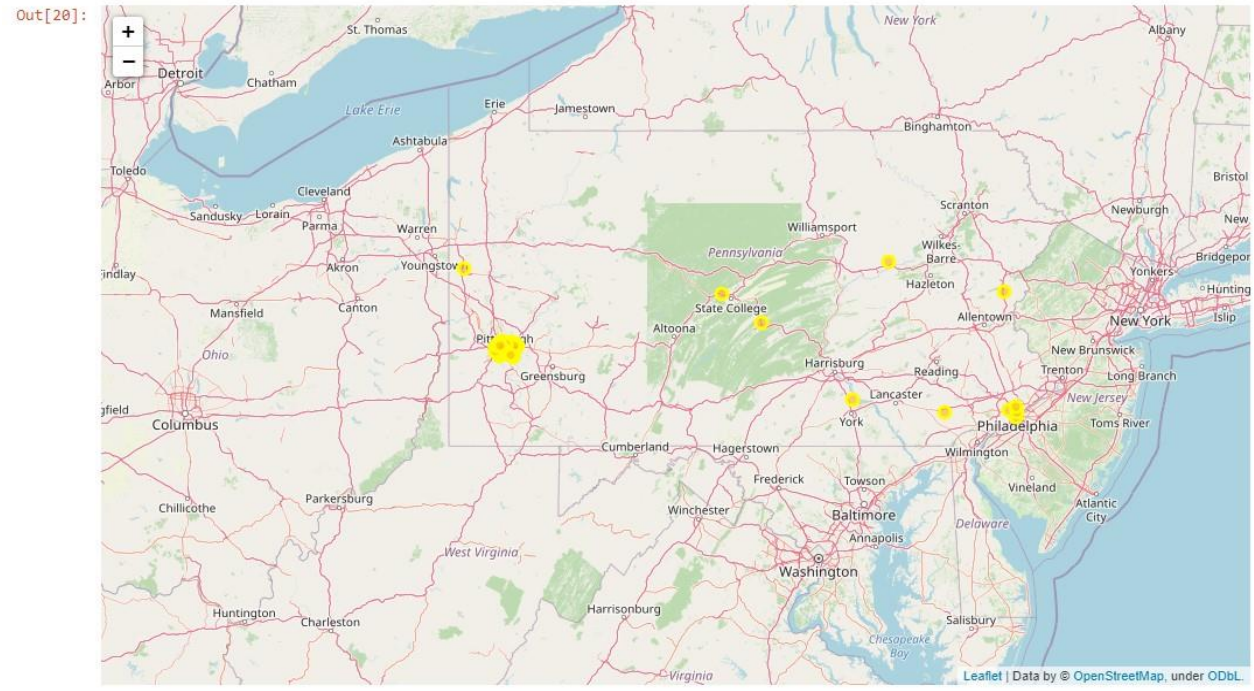

Figure 3a. Second map visualization of Charleston neighborhoods which shows all neighborhoods within city bounds.

Out[19]:

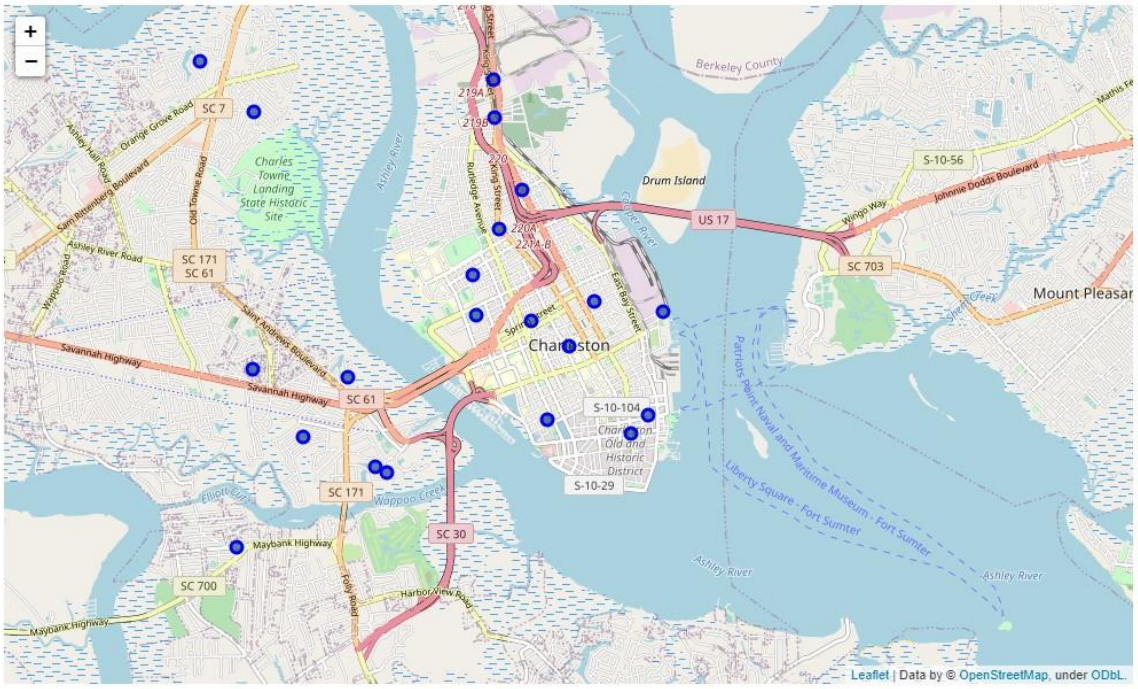

Figure 3b. Second map visualization of Pittsburgh neighborhoods which shows all neighborhoods within city bounds.

Out[25]:

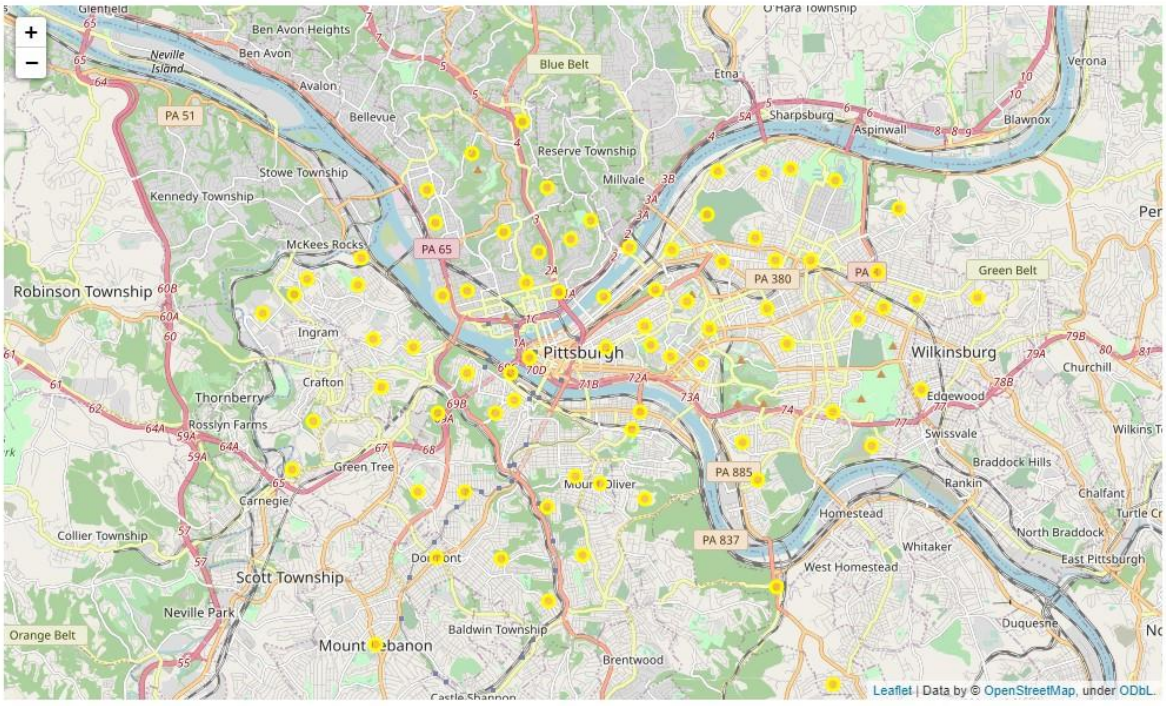

Figure 4a. Final dataframe with all neighborhood names, coordinates, and frequency of brewery venues in Charleston. First 10 neighborhoods shown.

Out[35]:

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood	lat	long	Brewery
13	Silver Hill-Magnolia	SC	Charleston	Charleston-North Charleston	205635.0	Silver Hill-Magnolia, SC	32.820900	-79.951600	0.240000
4	Wagener Terrace	SC	Charleston	Charleston-North Charleston	564916.0	Wagener Terrace, SC	32.816200	-79.951400	0.162162
14	East Central	SC	Charleston	Charleston-North Charleston	343867.0	East Central, SC	32.807000	-79.947300	0.112903
6	North Central	SC	Charleston	Charleston-North Charleston	455721.0	North Central, SC	32.802100	-79.950700	0.063291
20	Sandhurst	SC	Charleston	Charleston-North Charleston	439524.0	Sandhurst, SC	32.823232	-79.995645	0.034483
19	Hampton Park Terrace	SC	Charleston	Charleston-North Charleston	696325.0	Hampton Park Terrace, SC	32.796300	-79.954700	0.034091
17	Northbridge Terrace	SC	Charleston	Charleston-North Charleston	405576.0	Northbridge Terrace, SC	32.816843	-79.987590	0.027778
2	Cannonborough-Elliotborough	SC	Charleston	Charleston-North Charleston	568652.0	Cannonborough-Elliotborough, SC	32.790600	-79.945900	0.020000
15	South Windermere	SC	Charleston	Charleston-North Charleston	627225.0	South Windermere, SC	32.776010	-79.980090	0.012195
24	Moreland	SC	Charleston	Charleston-North Charleston	511431.0	Moreland, SC	32.783510	-79.973423	0.011628

Figure 4b. Final dataframe with all neighborhood names, coordinates, and frequency of brewery venues in Pittsburgh. First 10 neighborhoods shown.

Out[36]:

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood	lat	long	Brewery
63	Spring Garden	PA	Pittsburgh	Pittsburgh	129360.0	Spring Garden, PA	40.471000	-79.988000	0.125000
41	Upper Lawrenceville	PA	Pittsburgh	Pittsburgh	251403.0	Upper Lawrenceville, PA	40.482037	-79.951025	0.125000
39	Spring Hill - City View	PA	Pittsburgh	Pittsburgh	132676.0	Spring Hill - City View, PA	40.466730	-79.993929	0.086957
21	Lincoln-Lemington-Belmar	PA	Pittsburgh	Pittsburgh	57826.0	Lincoln-Lemington-Belmar, PA	40.473523	-79.898046	0.076923
22	Central Lawrenceville	PA	Pittsburgh	Pittsburgh	291035.0	Central Lawrenceville, PA	40.472349	-79.953866	0.070000
20	Stanton Heights	PA	Pittsburgh	Pittsburgh	214786.0	Stanton Heights, PA	40.481407	-79.937572	0.066667
45	Lower Lawrenceville	PA	Pittsburgh	Pittsburgh	267422.0	Lower Lawrenceville, PA	40.464514	-79.964399	0.060000
71	Herrs Island	PA	Pittsburgh	Pittsburgh	650633.0	Herrs Island, PA	40.465062	-79.976818	0.060000
35	Morningside	PA	Pittsburgh	Pittsburgh	232856.0	Morningside, PA	40.482434	-79.929518	0.053571
69	Chartiers City	PA	Pittsburgh	Pittsburgh	118351.0	Chartiers City, PA	40.458012	-80.070694	0.043478