



An Analysis of Pittsburgh Neighborhoods: A Guide for Incoming Homeowners

By Ryan Mooney

Introduction

Once a city of dreary smog covered skies and droning steel mining, Pittsburgh has made quite the change in the last couple decades to become a booming city of healthcare, technology, and innovation. Attracting new tech company offices from Facebook, Uber, and Google, Pittsburgh has been heading in the direction of becoming a hotspot for millennials seeking new and upcoming jobs in booming industries. Pittsburgh has already been a powerhouse for healthcare and engineering with its US Steel, UPMC Health Network, and various pharmaceutical industrial complexes, so it can really hold its own in its diverse portfolio of desirable jobs.

Not only does it have the jobs, but it has the attractions to keep the populace entertained. Long known for its music, the Carnegie Theater still boasts some of the greatest performances in the nation. The sports arenas are some of the most beautiful around, sitting on the edge of the 3 rivers the city is known for, they have been the home to many championships, rightly giving the city one of its nicknames as the "City of Champions". It has even started to attract more large budget productions, such as The Dark Knight Rises, Jack Reacher, and The Perks of Being a Wallflower.

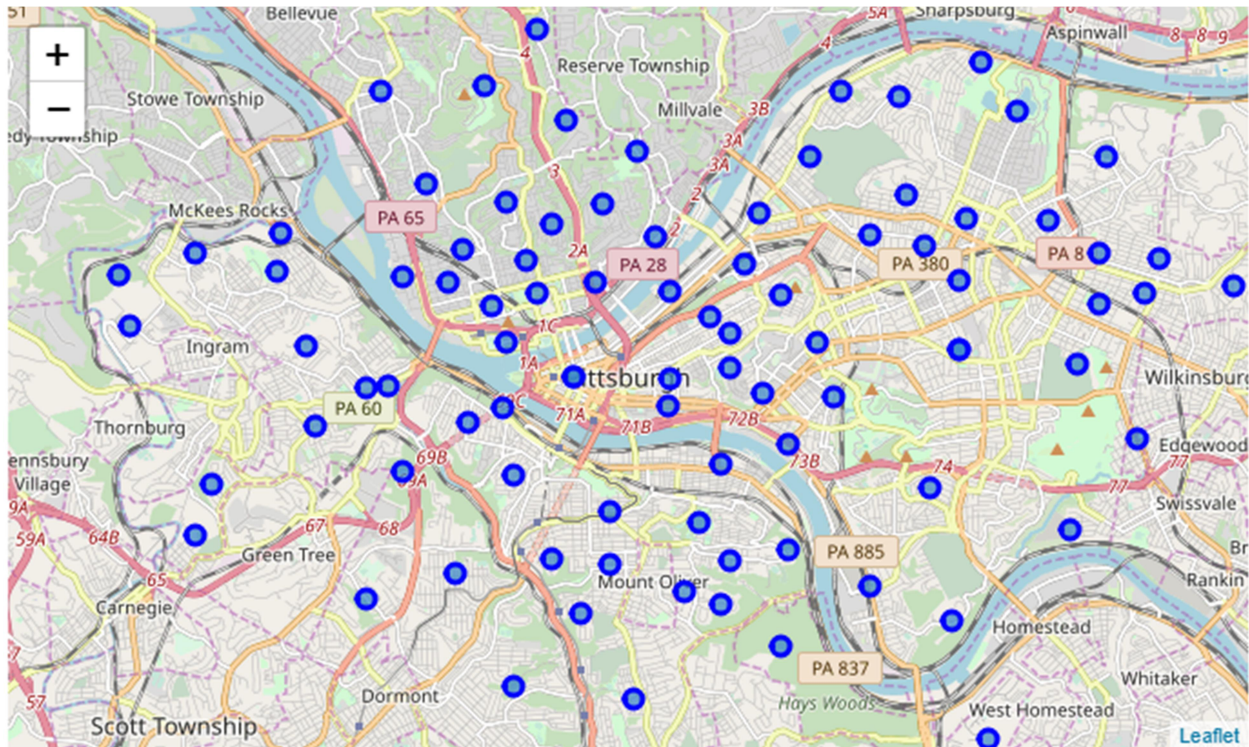
With all these enticing options, it is no wonder that people are flocking to Pittsburgh and its surrounding areas. While I grew up there, life took me across the country, but now as I am planning to head back, I have been wondering how the various neighborhoods stack up against each other. So with this project I aim to compare the neighborhoods against each other, looking at aspects of their local attractions and housing costs to see how they rank up against each other. Hopefully this analysis will help others looking at making the move as well!

Data and Data Sources

In order to analyze the city, I first needed to collect the neighborhoods names and their coordinates. By using the Wikipedia page for the Neighborhoods of Pittsburgh, I was able to find the neighborhoods names and aliases. I organized them into a CSV file and then used the Google Maps API in order to get the coordinates for each neighborhood. I iterated through the CSV and put the neighborhoods address into a call to Google's geocoding API function. The response it sent back included various information, but I was only interested in the latitude and longitude. All this information was organized into a pandas dataframe.

	Neighborhood	Latitude	Longitude
0	Allegheny Center	40.4538	-80.0074
1	Allegheny West	40.452	-80.0158
2	Allentown	40.4223	-79.9934
3	Arlington	40.4153	-79.971
4	Arlington Heights	40.4169	-79.9598

From there it was easy to use Folium's mapping feature to quickly make a map of Pittsburgh to ensure that my data looked correct. Once I was confident there was no issues with my data frame, I could then move on to collecting venue data using Foursquare's API.

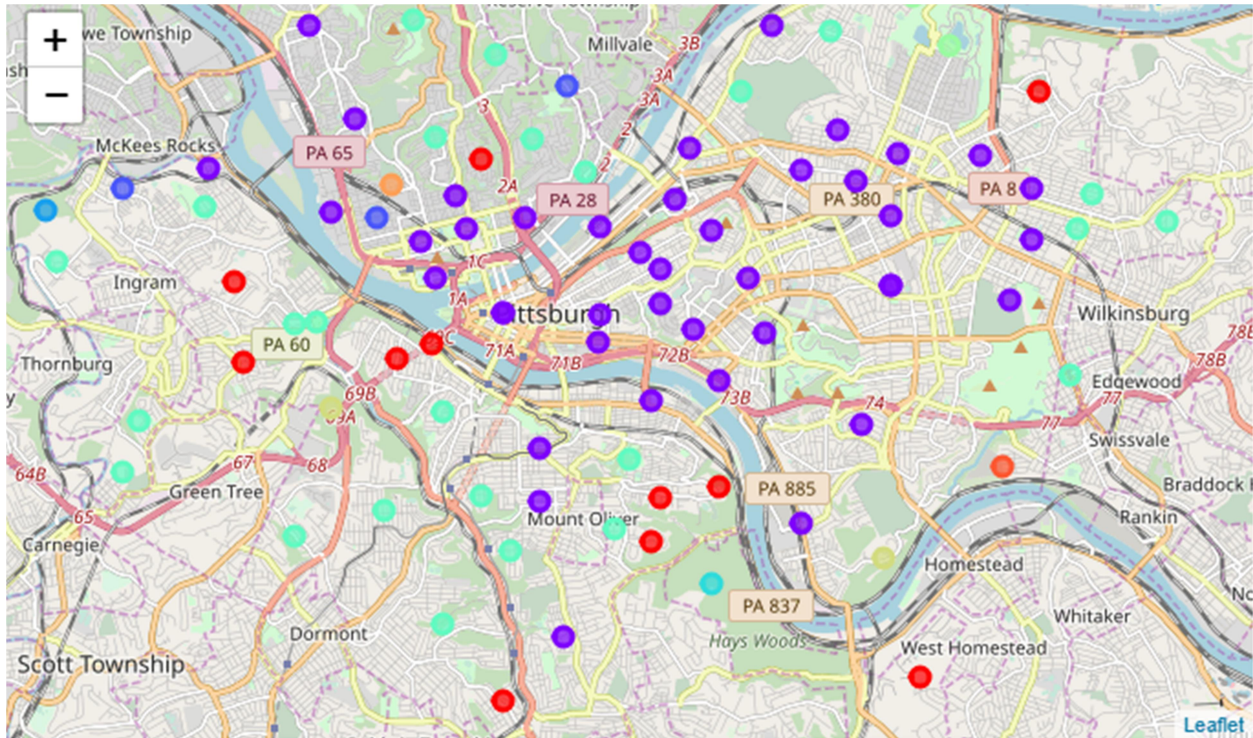


Methodology

Using the dataframe that included the neighborhood names and coordinates of Pittsburgh, we could explore each neighborhood with Foursquare’s API in order to see what sort of establishments each had. This could then be used to categorize and organize the neighborhoods into clusters. We started off by making calls to the Foursquare API asking for the nearby venues using each neighborhoods coordinates. I had to tweak the radius it would look in to ensure that most neighborhoods got enough results to analyze. Once I tested out the function on one neighborhood, I implemented a function that got the nearby venues for all of the neighborhoods by making an API call for each set of coordinates.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allegheny Center	40.453786	-80.007377	Children's Museum of Pittsburgh	40.452793	-80.006569	Museum
1	Allegheny Center	40.453786	-80.007377	National Aviary	40.453154	-80.010049	Zoo
2	Allegheny Center	40.453786	-80.007377	El Burro	40.455860	-80.006689	Mexican Restaurant
3	Allegheny Center	40.453786	-80.007377	Federal Galley	40.451605	-80.006045	Comfort Food Restaurant
4	Allegheny Center	40.453786	-80.007377	Brugge On North	40.455400	-80.007478	Belgian Restaurant

Once we had a pandas dataframe that included all of the venues in the city, it was time to organize them. I used one hot encoding in order to find the frequency of each type of venue in each neighborhood. We grouped this data into a new data frame that listed each neighborhood and the frequency of each type of venue in it. From there, it was easy to organize the frequencies into the 10 most common venues in each neighborhood. We used these common venues in order to cluster the neighborhoods using K-Means clustering. This gave us 10 clusters based on the commonalities of venues that each neighborhood shared. From here we were able to analyze the data based on their common attractions.



Results

By looking at some of the highest frequency and unique venues in each cluster, we were able to get a better sense of what each cluster was like. Here are the list of venues I found for each cluster.

- Cluster 1: American restaurants, Zoos, Other outdoor activities
- Cluster 2: Italian, Pizza, Coffee shops
- Cluster 3: Zoo, Food Court
- Cluster 4: Zoo, Food trucks
- Cluster 5: American Restaurant, Zoo
- Cluster 6: Bars, Parks, Home Service
- Cluster 7: Zoo, American Restaurant
- Cluster 8: Gym, Zoo, Food and drink shop
- Cluster 9: Comedy Club, Bakery, Flower shops
- Cluster 10: Playground, Department stores

Discussion

Through this analysis, we have found a large variety of attractions to bring you to each part of the city. In order to categorize them better, though, we might try and narrow down what makes each cluster unique. To analyze each cluster, we will take a look at the most common venues in each cluster and see if that tells us something about what kind of neighborhoods they are.

Looking at the clusters this way, the first thing I notice is clusters 3 and 4 are quite similar, as are clusters 5 and 7. So we could essentially group these together if we would like. Once that is done, we notice

certain themes among the clusters. For example, if you like American food, perhaps a neighborhood in cluster 1, 5, or 7 might suit you. If you like being close to outdoor attractions, maybe cluster 1 might be for you. While this approach isn't exact in separating the neighborhoods, it does give you a unique look at the city.

There are a few things that could be changed in future iterations of this to possibly give better results. In future attempts, it might be of interest to vary the radius that we look at in order to prevent overlap between the neighborhoods. As of now, a radius of 550 was needed in order to prevent some of the outlying neighborhoods to return with no venues, but this caused the more densely packed venues to return a large number of venues. There might also be some tweaks that could be interesting as to how the venues are categorized. Maybe categorizing the venues differently could give us a glimpse as to which neighborhoods are more "party oriented", which are "foodie heavens", or which are "shoppaholic dreams", for example.

Conclusion

Using the Foursquare API, we can get a unique look at different neighborhoods of a city in order to determine the unique attractions each have to offer. We can utilize it as an interesting tool in determining which neighborhood might be better suited to the needs of an incoming home owner. For me, I found Cluster 2 to be the most enticing, and while looking through I settled on Lower Lawrenceville as my favorite option. Its varied selection of foods and attractions made it quite the choice. There are certainly some shortcomings to this technique, such as overlapping and categorizing, but I think it posed a great new way to look at cities. I believe this approach could certainly help a home owner choose the right place to live!

References

- 1, "List Of Pittsburgh Neighborhoods". *En.Wikipedia.Org*, 2019, https://en.wikipedia.org/wiki/List_of_Pittsburgh_neighborhoods.
2. Google Maps API
3. Foursquare API
3. Folium Library