Identifying Neighborhood Commonality in Two Neighboring Cities Based on Affordability and Trendy Venues

Patrick Reed
Coursera Capstone Project

Problem: Finding Neighborhood Commonality

Goal: Identify Commonalities in Neighborhoods between two Cities,

- Use Home Value index for Affordability
- Use Number of nearby Breweries / Coffee Shops for “trendiness”

Users: Interest in this analysis applies to multiple users interested in identifying aspects of their local or neighboring city.

- House Shoppers - Looking for neighborhoods that fit their lifestyle
- Real Estate Agents - Providing customers with detailed options to meet their needs
- Investors - Searching for Market Opportunities

Data: Acquisition and Cleaning

Data Sources:

- Name and geographic location (Latitude and Longitude) of each of the neighborhood,
  - https://www.city-data.com/ for city information and lists of neighborhoods,
  - Google Earth (https://earth.google.com/), for neighborhood coordinates,
- Neighborhood home value data,
  - ZILLOW Home Value Index (https://www.zillow.com/home-values/)
- Nearby venue identification based neighborhood coordinates,
  - FourSquare website and API (https://foursquare.com/)

Data Cleaning:

- Combine all data sources, grouped by neighborhood
- Remove extraneous values (i.e. cafes, and bars vs coffee shops and breweries)
- Deal with missing data (use city mean home value / replace NAN with 0 for venues)
- Normalize Home value on scale with venue counts

Finished Data

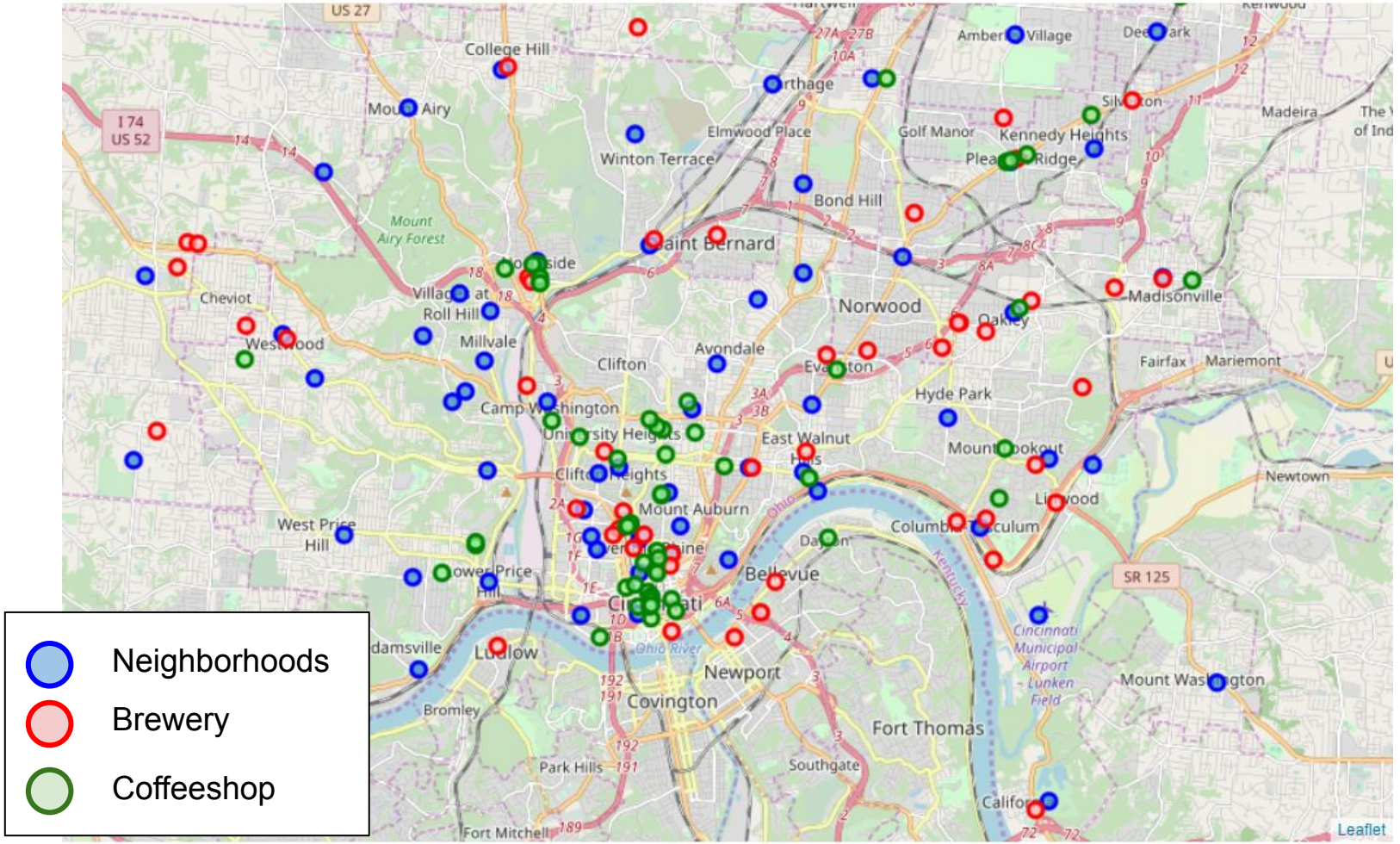
	Neighborhood	Latitude	Longitude	City	State	County	House Value	Brewery	Coffeeshop	HV_Norm
0	Avondale	39.144963	-84.497811	Cincinnati	OH	Hamilton	81719	0	0	0.81719
1	Bond Hill	39.177785	-84.477659	Cincinnati	OH	Hamilton	111614	0	0	1.11614
2	California	39.065338	-84.419893	Cincinnati	OH	Hamilton	128577	1	0	1.28577
3	Camp Washington	39.137950	-84.537609	Cincinnati	OH	Hamilton	58310	2	2	0.58310
4	Carthage	39.195869	-84.485014	Cincinnati	OH	Hamilton	68406	0	0	0.68406
5	Clifton Heights	39.125934	-84.520908	Cincinnati	OH	Hamilton	298535	8	4	2.98535
6	College Hill	39.198536	-84.548428	Cincinnati	OH	Hamilton	130138	1	0	1.30138
7	Columbia-Tusculum	39.115193	-84.436140	Cincinnati	OH	Hamilton	325009	5	2	3.25009
8	Corryville	39.136807	-84.503866	Cincinnati	OH	Hamilton	158269	1	5	1.58269
9	CUF	39.125115	-84.525842	Cincinnati	OH	Hamilton	161286	8	3	1.61286
10	East End	39.099214	-84.422519	Cincinnati	OH	Hamilton	318827	1	0	3.18827
11	East Price Hill	39.106141	-84.569386	Cincinnati	OH	Hamilton	67706	0	0	0.67706
12	East Walnut HIlls	39.125191	-84.477637	Cincinnati	OH	Hamilton	171615	2	3	1.71615
13	East Westwood	39.150129	-84.566766	Cincinnati	OH	Hamilton	57827	0	0	0.57827
14	English Woods	39.139781	-84.556888	Cincinnati	OH	Hamilton	77999	1	0	0.77999

Exploratory Data Analysis

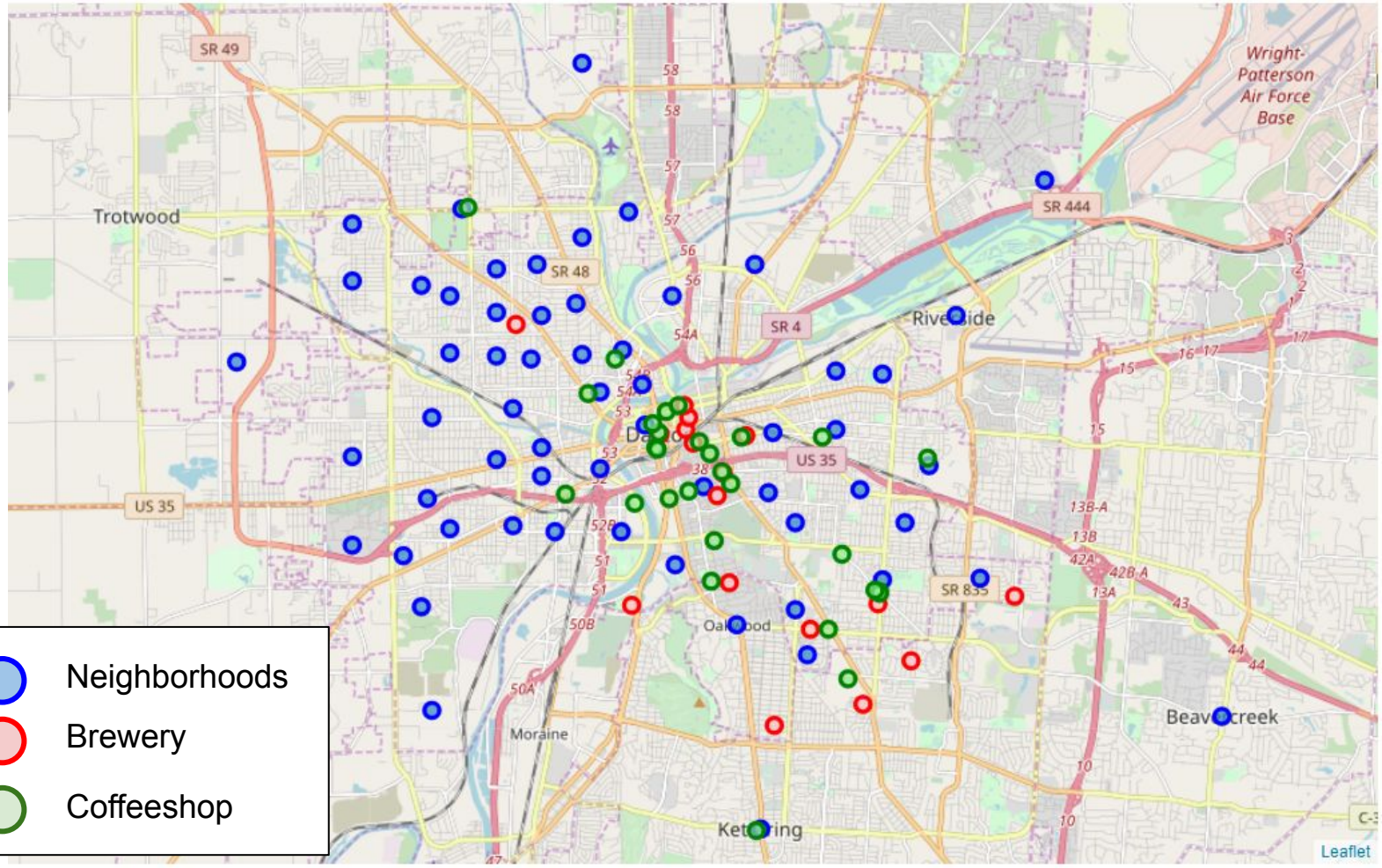
- View the neighborhood and venue data on a map:
  - Initial Observation: Greater Number of Cincinnati Venues

- View City Home Value Histograms
  - Initial Observation: Much greater Home Values in Cincinnati
  - Higher Home Value at largest bin
  - Max Value for Cincinnati nearly double that of Dayton

Cincinnati, Ohio Neighborhoods and Venues



Dayton, Ohio Neighborhoods and Venues

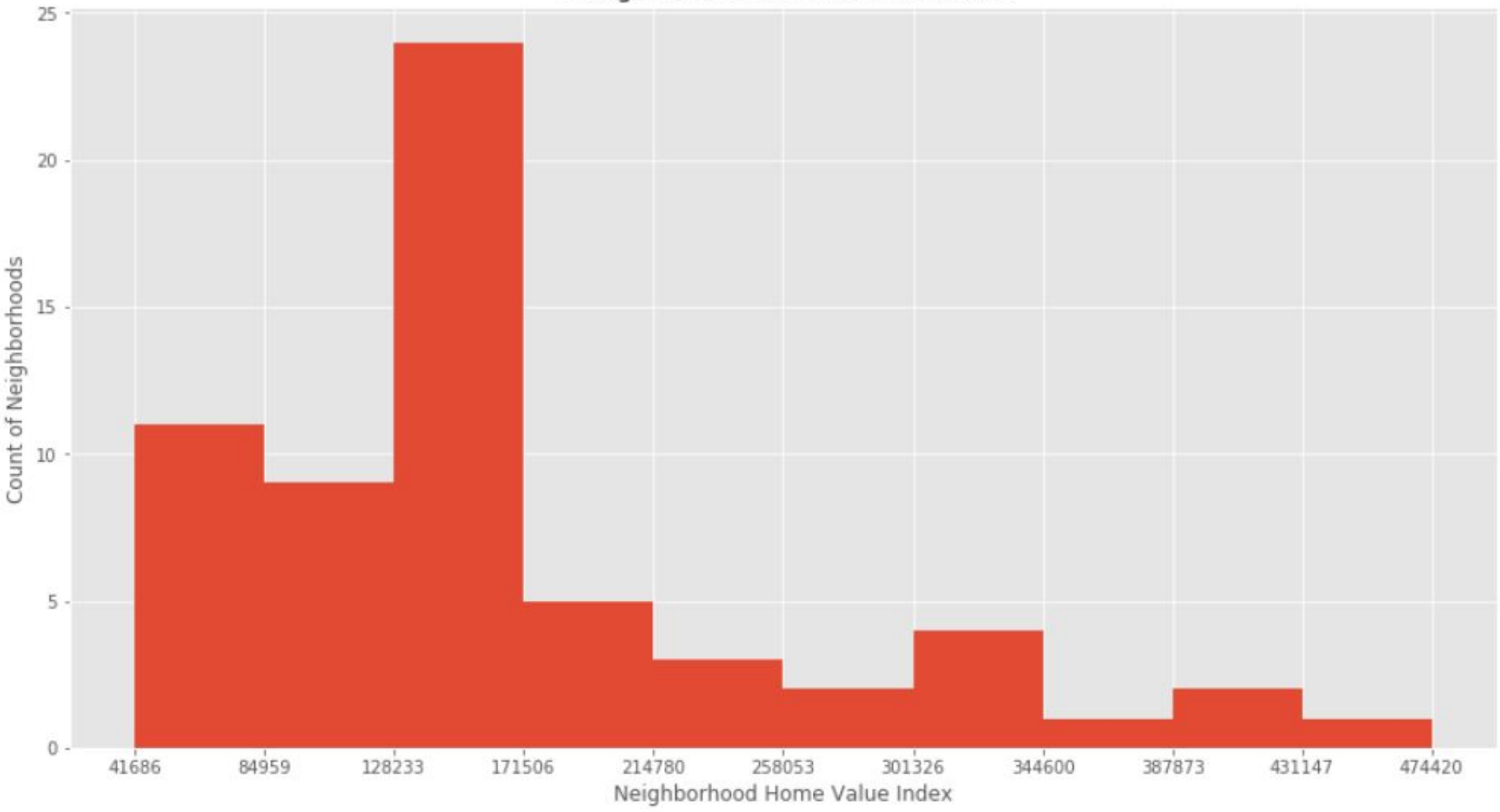


Neighborhoods

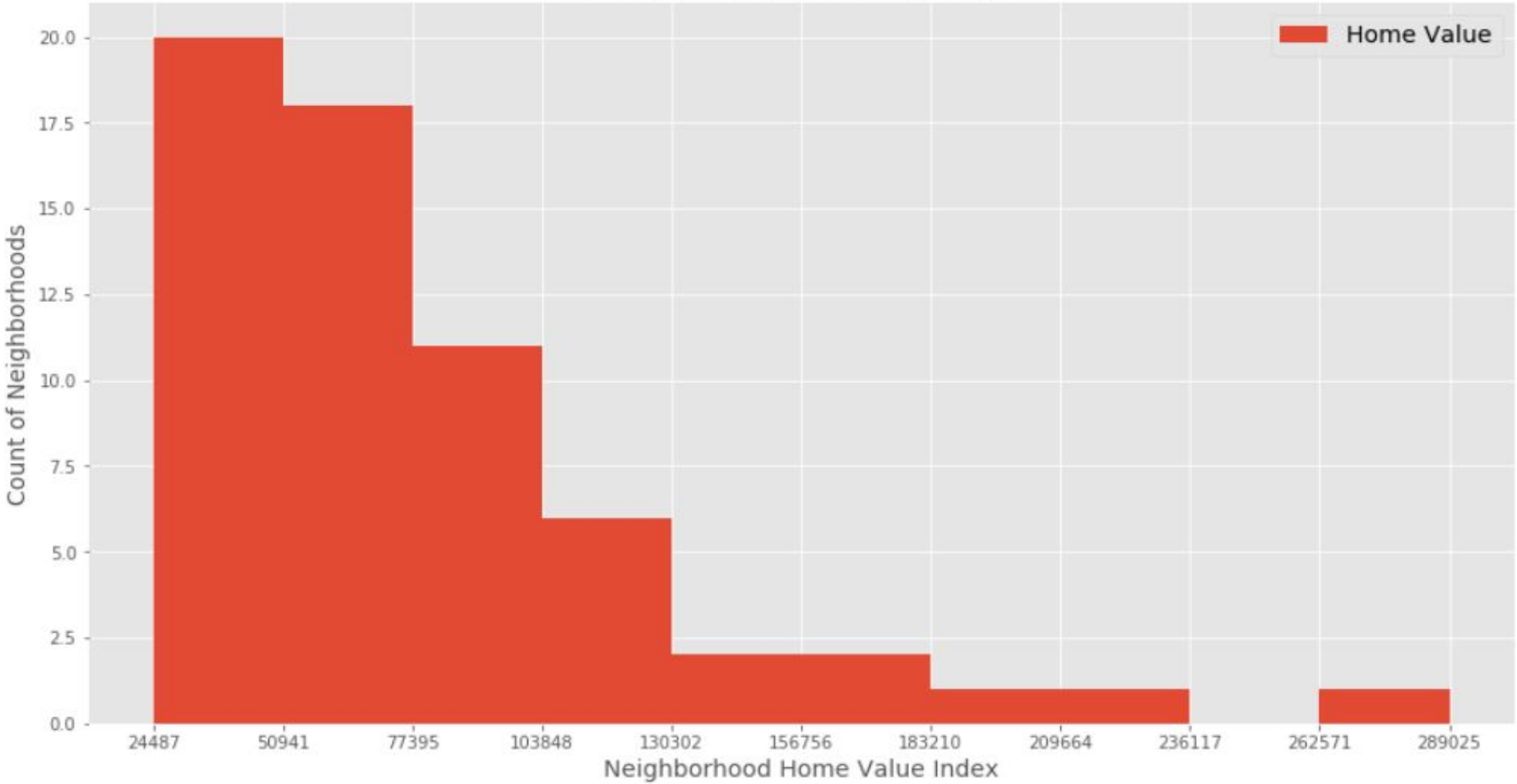
Brewery

Coffeeshop

Histogram of Cincinnati Neighborhood Home Values



Histogram of Dayton Neighborhood Home Values

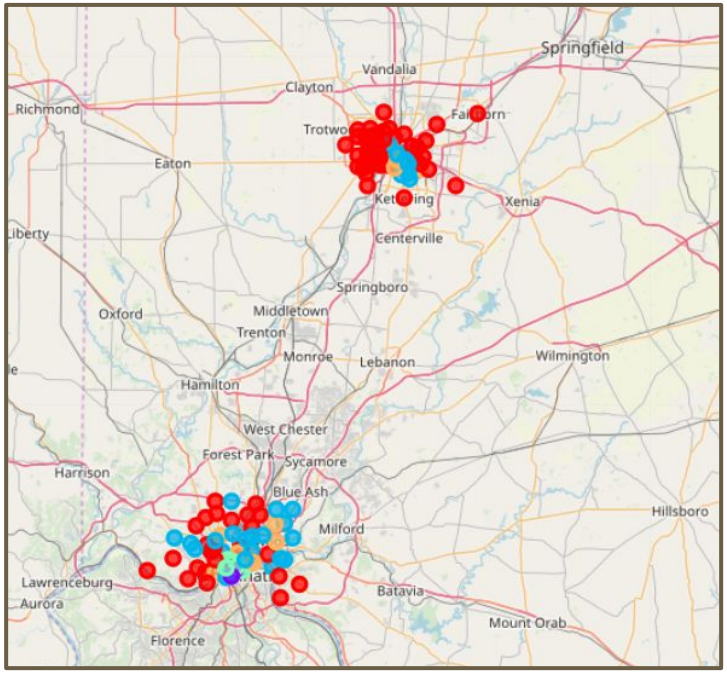


K-Means Clustering

- “Kmeans” Algorithm:
  - One of the most common clustering technique.
  - Iterative, unsupervised learning method.
  - Identifies unique, non-overlapping clusters of data.

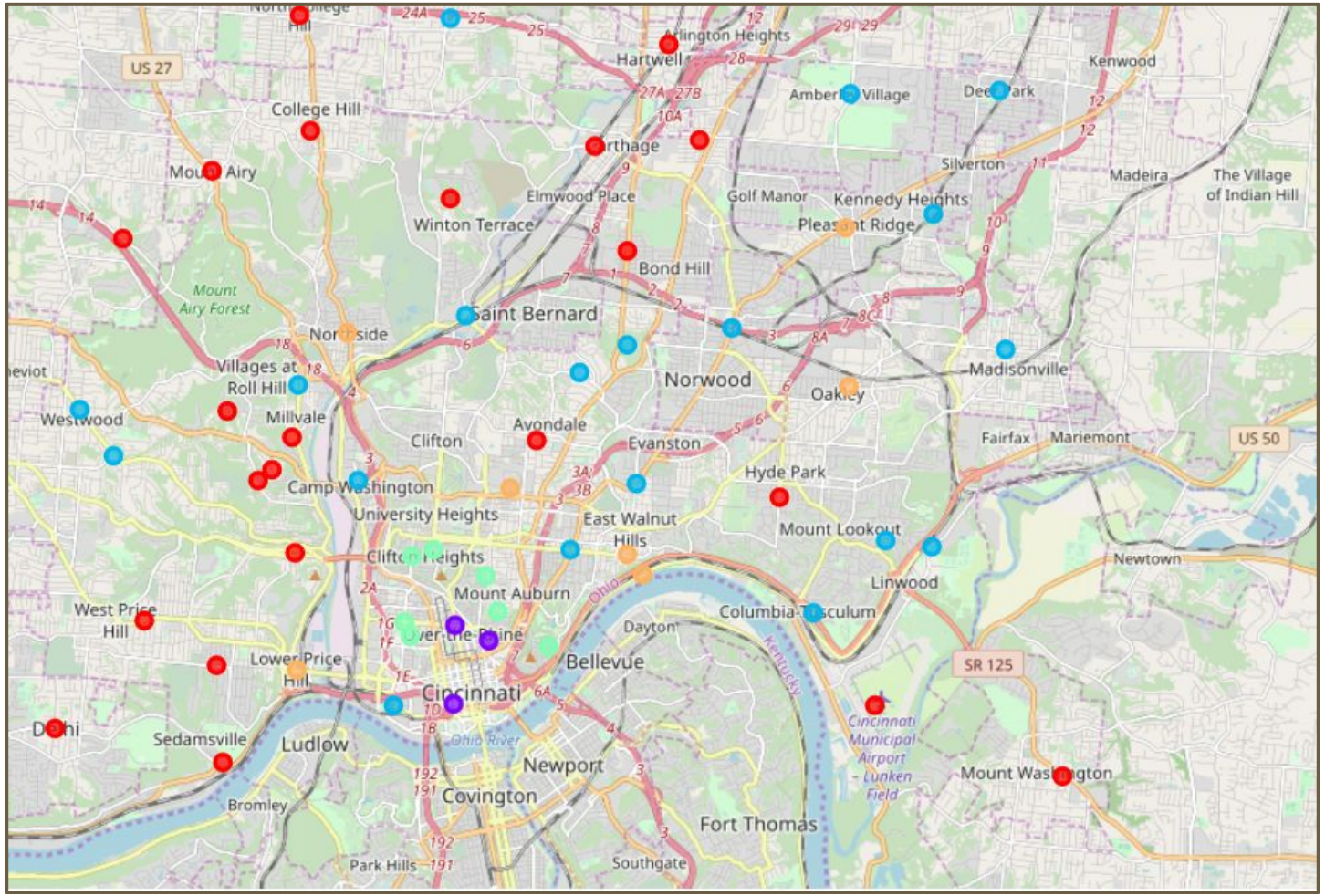
- Methods Utilized:
  - Scikit-learn Python Module
  - Tested 3, 4, and 5 Clusters
  - Determined 5 Clusters to be Ideal

Results

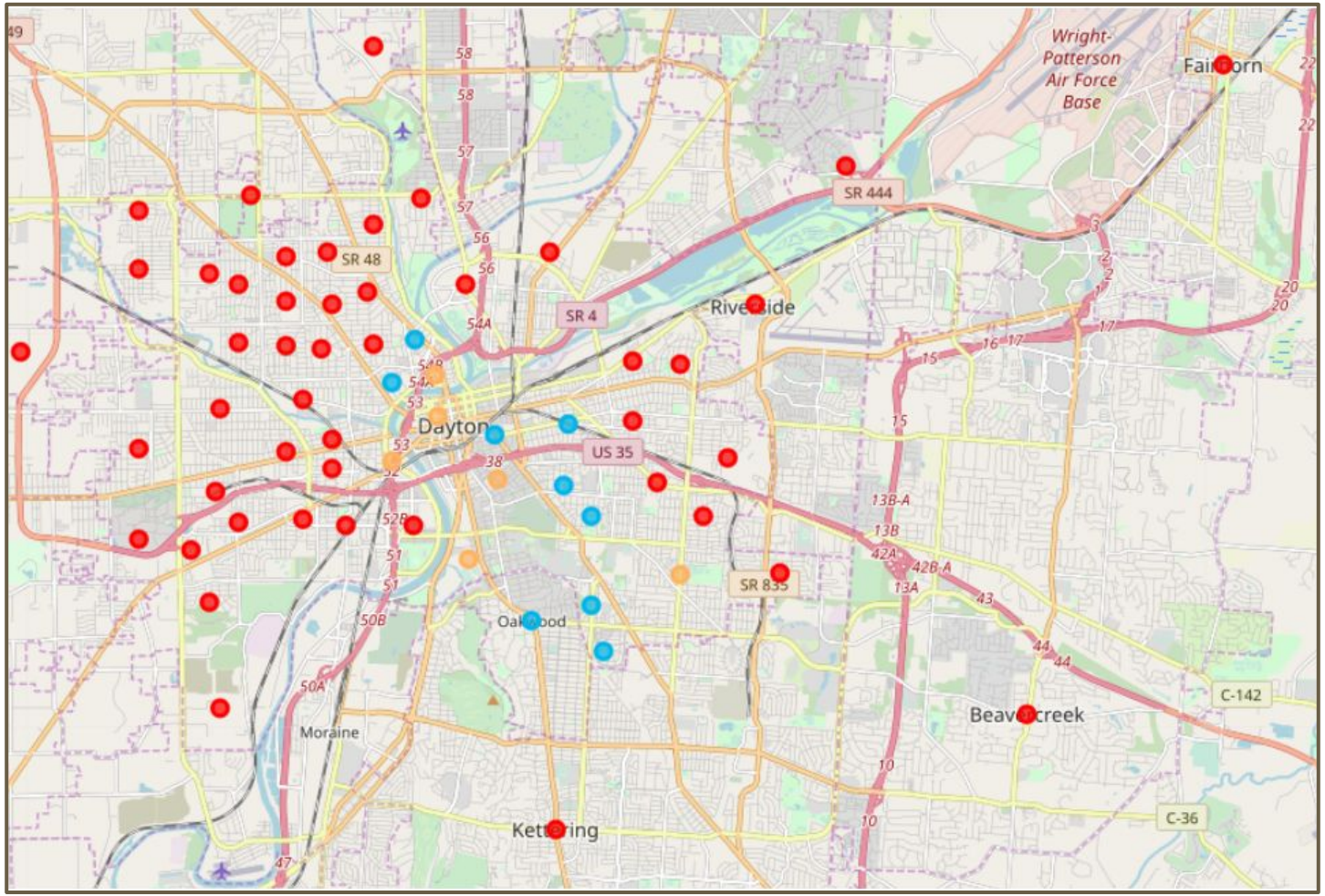


- Cluster 0:
  - 72 Neighborhoods
- Cluster 1:
  - 3 Neighborhoods (Cincinnati Only)
- Cluster 2:
  - 29 Neighborhoods
- Cluster 3:
  - 7 Neighborhoods
- Cluster 4:
  - 13 Neighborhoods

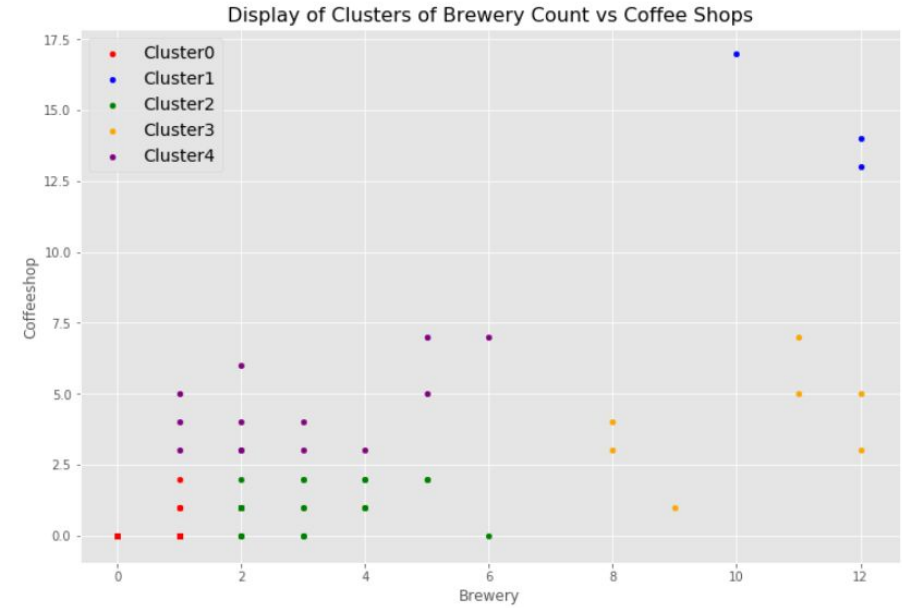
Cincinnati Cluster Results

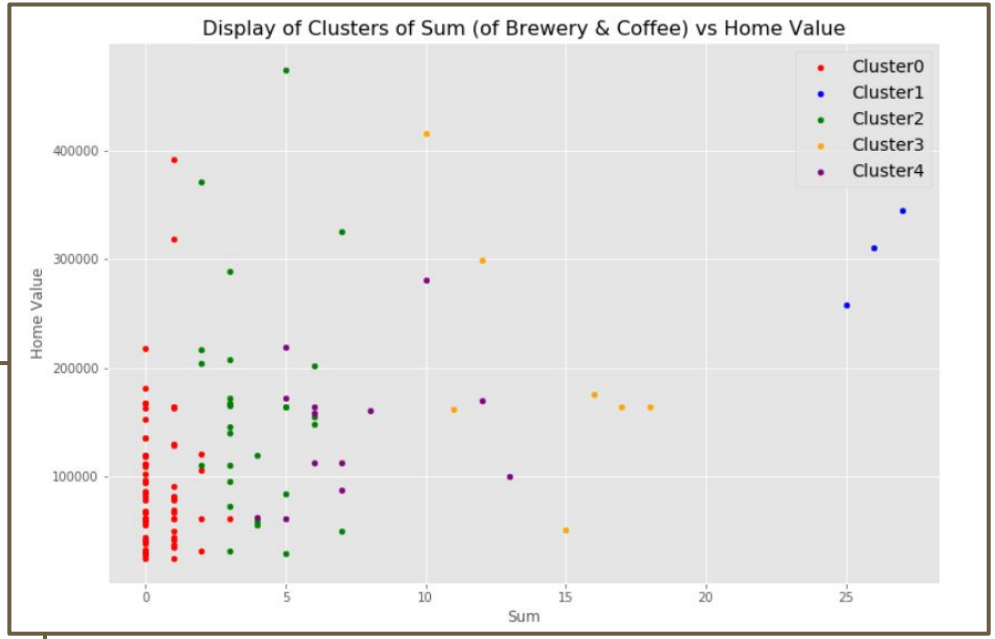


Dayton Cluster Results



Cluster Results





Conclusions

- **Cluster 0:** Low price-range, low-venue count neighborhoods
  - Least desirable neighborhoods for this project.
- **Cluster 1:** High price-range,venue-dense neighborhoods.
  - Very Expensive Real Estate
  - Very dense venue counts
- **Cluster 2:** Modest affordability and Venue density
  - Still on the low-end of both.
- **Cluster 3:** Slightly Over Desired affordability and venue density
  - Still on the high-end of both
- **Cluster 4:** The sweet spot in this analysis.
  - Balanced levels of both affordability and venue density.
  - This cluster also has decent representation in both cities.