Capstone Project – Finding a Seattle Brewery Location

Mike Burghart

December 13, 2020

1. Introduction / Business Problem

People in the Pacific Northwest love beer, with over 150 breweries operating in the Seattle / Tacoma. metropolitan area. The goal of this project is to find a neighborhood in Seattle proper to open up a new brewery / taproom. Because there are already so many breweries already operating in Seattle, we would prefer neighborhoods that are currently not heavily saturated with other breweries. Additionally, we would like to find locations with good population density, has a relatively high household income and do not have a large number of breweries/bars nearby that might compete with our brewery (the taproom plans on having rotating food trucks.). We will use the k-means clustering algorithm to complete this task. Finally, you will use the Folium library to visualize the neighborhoods in Seattle and their emerging clusters.

2. Data Sources

The data I will be using comes from the city of Seattle website https://data-seattlecitygis.opendata.arcgis.com/datasets/a-community-reporting-areas-profile-acs-5-year-2013-2017. The data is available in several forms including csv, shapefile, kml and GeoJSON. There is quite a bit of useful information available for each Community Reporting Area (CRA) in the data set. For each CRA (or neighborhood), a few examples of data available include: size of the neighborhood, population, population density, median household income and average gross rent. In order to make the analysis using Python a bit easier, I first took a shapefile of the Community Reporting Areas (CRA) and created a latitude/longitude centroid for each neighborhood using the QGIS software. This would allow searching for bars / restaurants in each CRA using the Foursquare location data a bit easier. This centroid latitude / longitude information was added to the CRA csv file. Because there are a significant number of fields, I edited my csv file to only contain information that was deemed relevant to the analysis. This csv file (along with the GeoJSON file) were the basis for the analysis.

	NEIGHBORHOOD_ID	NEIGHBORHOOD	DISTRICT	AREA_ACRES	AREA_SQMI	TOTAL_POPULATION	MEDIAN_HH_INC
0	1	Ballard	Ballard	492.9	0.77	8649	79162
1	2	North Beach/Blue Ridge	Ballard	1284.4	2.01	12701	94804
2	3	Montlake/Portage Bay	Northeast	951.1	1.49	9732	132573
3	4	Interbay	Magnolia/Queen Anne	1214.6	1.90	11024	74679
4	5	North Capitol Hill	East	283.7	0.44	4807	96220

Table 1: Example of csv data utilized for the analysis

Venues data of breweries/bars in each CRA were obtained using Foursquare. The number of breweries/bars in each CRA were aggregated. As noted earlier, we are trying to avoid areas that are already saturated with breweries/bars, so this information is useful for our analysis. A bar/brewery was considered to be near a neighborhood if it was within approximately 0.75 miles of the neighborhood centroid.

	NEIGHBORHOOD	BARS_IN_AREA
0	Ballard	27
1	North Beach/Blue Ridge	0
2	Montlake/Portage Bay	0
3	Interbay	2
4	North Capitol Hill	1

Table 2: Number of bars/breweries located near each neighborhood

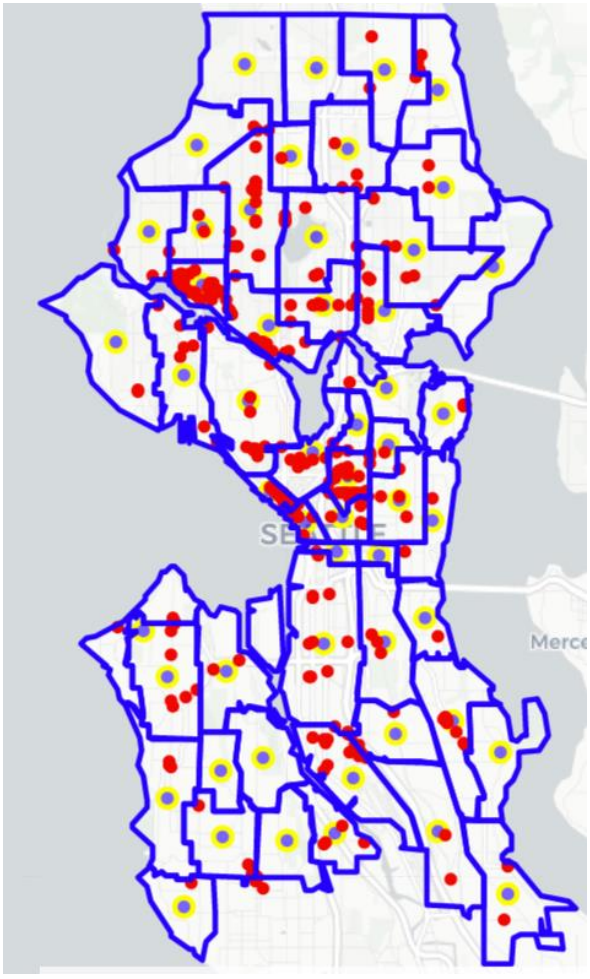


Figure 1: Neighborhoods and bars/breweries. The neighborhoods are outlined in blue, and the centroid is the purplish dot with yellow outline.

3. Methodology

In this project we will direct our efforts on finding neighborhoods that have low brewery/bar density. We will also be looking at neighborhoods with larger populations / population density, higher median income and (hopefully) have reasonable rent.

In first step, we have collected the required data: location and type (category) of every bar/brewery within approximately 0.75 miles from each neighborhood centroid.

In the second step in our analysis will revolve around the exploration of potential neighborhoods for the brewery using k-means clustering. We hope the k-means clustering will effectively group neighborhoods in a manner that will help us easily identify neighborhoods with larger populations / population density, higher median income and (hopefully) have reasonable rent. We will include some mappings to aid with visualizations of bars/breweries throughout the city.

First, we want to try and optimize the number of k-means clusters using the Elbow Method. The graphic results to help are shown below.

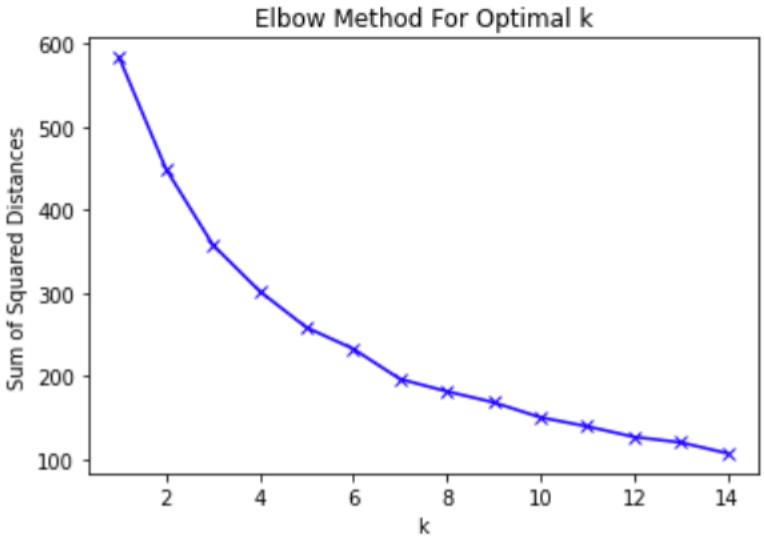


Figure 2: K-means optimization results.

Based on the graph, there wasn't a clear-cut "winner", but it looked as if k=4 or k=5 are good candidates. It was decided to go with k=4 for this part of the analysis.

A visualization of the clustered neighborhoods is shown in a map below.

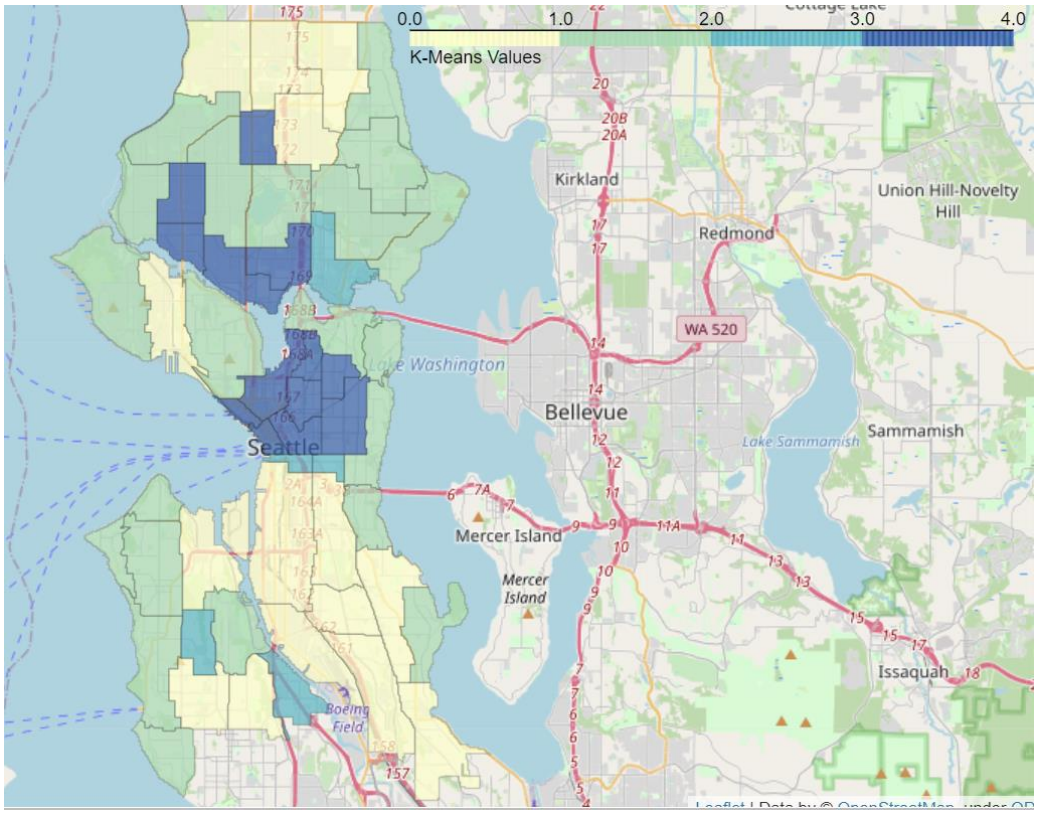


Figure 3: K-means (k=4) clustering of Seattle neighborhoods

We can group each of the k-means clusters (0 – 3) and analyze various demographic categories to see if there are clusters that have profiles that would help find neighborhoods that would be good candidates for breweries.

Cluster	Area (Acres)	Area (Sq. Mi.)	Population	Median Income	Median Rent	% HH Income Below $50,000	% HH Income Between $50,000 and $100,000	% HH Income Above $100,000	Population Density (per acre)	Bars	Bar Density
0	1319	2.06	11646	66859	1254	39.3	30.1	30.5	9.4	7.7	0.195
1	1215	1.90	14809	105784	1500	24.7	24.9	50.5	11.8	6.1	0.038
2	518	0.81	8168	38089	983	59.5	20.8	19.8	17.7	9.4	0.141
3	530	0.83	13573	85226	1516	31.0	27.0	42.0	28.0	27.1	0.260

Some interesting results here. It looks like Cluster 3 may not be a good candidate - there are an average of over 27 bars/breweries in this cluster! The other neighborhoods have much fewer bars on average, although Cluster 0 and 2 have a bit higher bar density (which is the number of bars in a neighborhood divided by the population). Cluster 1 has the highest median income, but more expensive median rent. However, the neighborhoods in Cluster 1 have the largest population, decent population density and low bar density on average. We'll focus further on neighborhoods from Cluster 1.

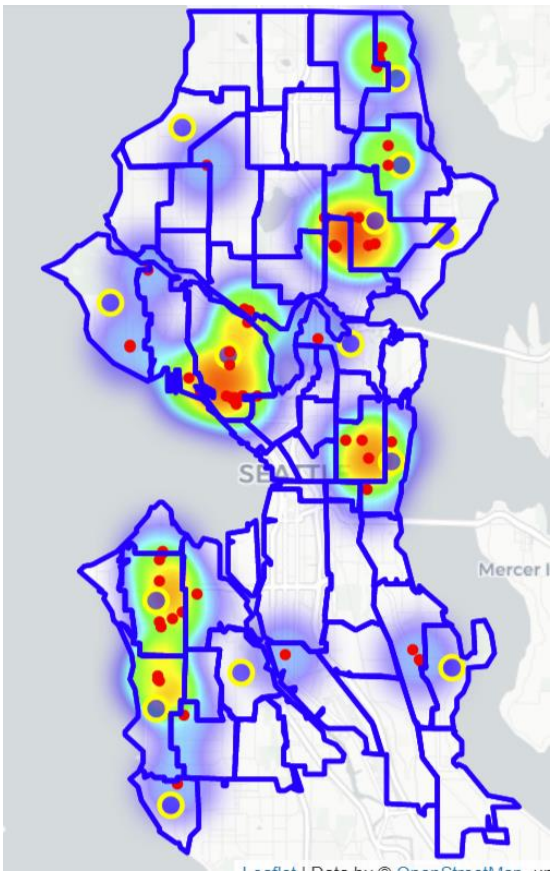
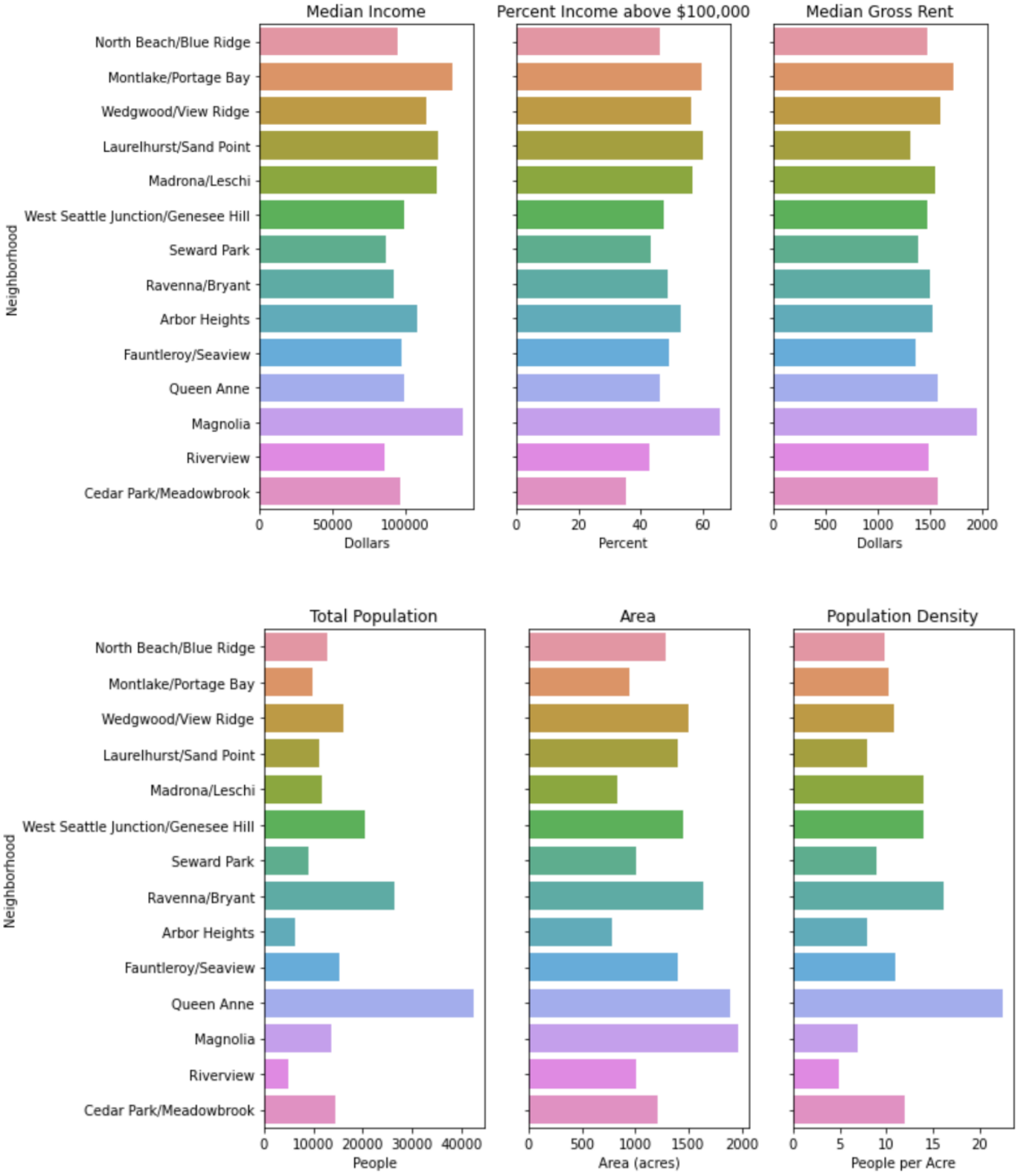


Figure 4: Heat map showing bar/brewery density for our Cluster 1 neighborhoods

Now we'll do some bar plot visualizations for each neighborhood in Cluster 1. We'll look at median income, income over $100,000, median gross rent, total population, area (acres) and population density (people per acre).



Interesting results, again. Magnolia is the most affluent neighborhood, but has the most expensive rent and has a relatively low population / population density. Montlake/Portage Bay, Wedgewood/View Ridge, Laurelhurst/Sandpoint and Madrona/Leschi share similar qualities. Queen Anne has a large population and population density, but is pretty average in terms of income and rent compared to most neighborhoods in this cluster.

Since we don't seem to have a "slam dunk" neighborhood, let's use k-means clustering again on the Cluster 1 data set and see if we can drill down further.

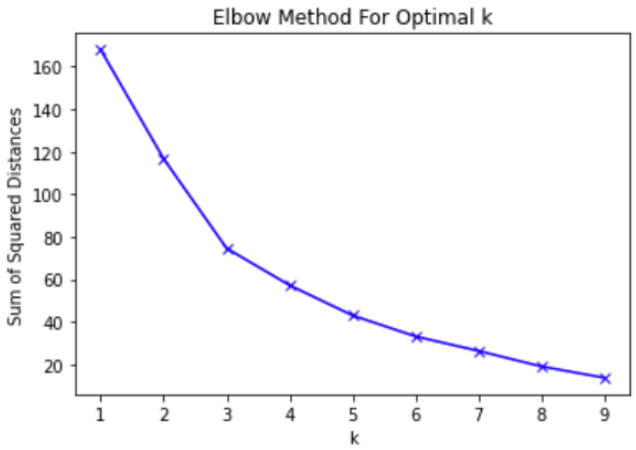


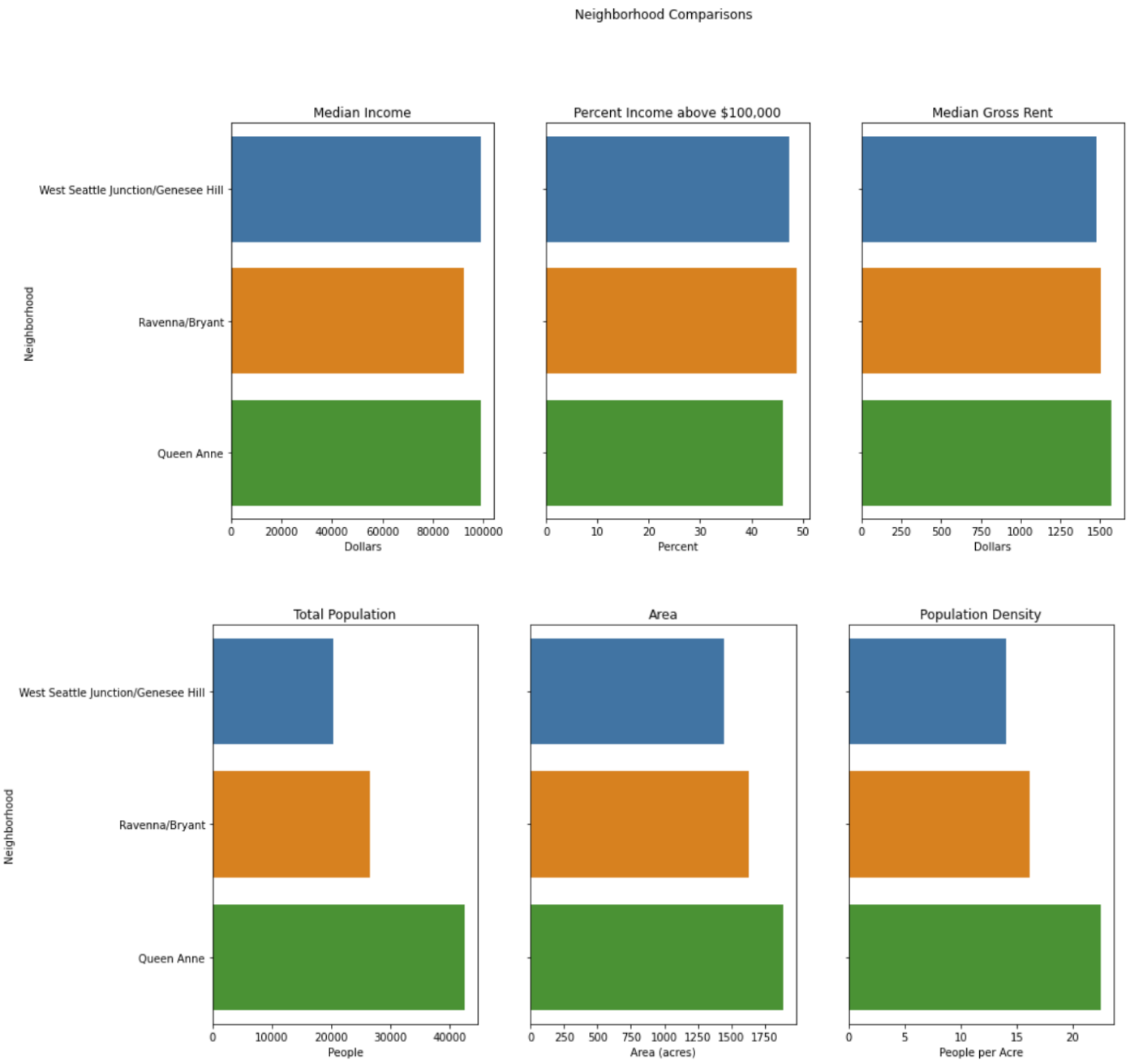
Figure 5: K-means optimization results with Cluster 1 neighborhoods.

Based on the graph, it's pretty clear that k=3 is the best for optimization. We do k-means clustering on the Cluster 1 subset, and can again group by the updated clusters and analyze various demographic categories to see if we can narrow down our search even further.

Cluster	Area (Acres)	Area (Sq. Mi.)	Population	Median Income	Median Rent	% HH Income Below $50,000	% HH Income Between $50,000 and $100,000	% HH Income Above $100,000	Population Density (per acre)	Bars	Bar Density
0	1331	2.08	12467	126170	1627	19.2	21.1	59.7	10.0	2.2	0.018
1	1115	1.74	10454	94787	1468	27.0	28.0	45.0	9.1	2.8	0.026
2	1657	2.59	29800	96880	1517	26.0	26.6	47.4	17.6	12.0	0.041

It looks like each of these Clusters has interesting qualities compared to each other. Cluster 0 has the highest median income, but has the most expensive rent and has low population density. Cluster 1 and Cluster 2 seem to have some similar qualities, although Cluster 2 has a much higher population and population density. There are a higher average number of bars in Cluster 2, but the bar density isn't too bad. Because of the higher population and density, let's take a closer look at neighborhoods from Cluster 2.

Again, we'll create some plots of the neighborhoods from Cluster 2 to help us decide which we think is a better area for out brewery.



4. Results

It appears economically, these three neighborhoods are quite similar, but Queen Anne has much greater population and population density. However, there are other outside considerations that make Queen Anne an attractive location. The Queen Anne neighborhood has historically been a popular area to live, notably because of how close it is to downtown Seattle. Queen Anne is also home to Key Arena, which is a venue that historically was the home of the Seattle SuperSonics pro basketball team before they were stolen away to Oklahoma City in 2008. Key Arena is also home to the Seattle Storm women's pro basketball team, and also hosts a number of concerts and events throughout the year. There were hopes that men's pro basketball would one day return to Key Arena, but the venue needed significant renovating in order to attract any professional sports teams.

In 2018, Seattle was awarded an expansion pro hockey team - the Seattle Kraken. In order to accommodate the new team, Key Arena is currently being renovated and will called "Climate Change Arena", and there will be pro hockey beginning in 2021. This will bring an economic influx to the Queen Anne area, and there are hopes that with a new arena, a men's pro basketball team could potentially relocate to Seattle in the near future, bringing even greater economic stimulus to the neighborhood.

5. Conclusion

There are a number of considerations that one must consider when looking for an ideal location for a brewery (or any other type of business). With a great deal of useful data available through Foursquare and other data sources, one can perform in-depth analyses using tools such as k-means clustering to help drill-down to help meet your business needs.