Samia Islam

August 13, 2020

ANALYSIS OF AMENITIES AVAILABLE TO NEIGHBORHOODS IN LOS ANGELES COUNTY BASED ON MEDIAN INCOME

Introduction

We want to analyze the different types of amenities in neighborhoods in a large metropolitan area and see if there is any correlation of these venues with the median income of the neighborhoods. In particular, we want to see if people living in neighborhoods with lowest median income have access to parks, trails or other facilities dedicated to an overall healthy lifestyle, compared to their counterparts in neighborhoods with highest median income.

Several studies have confirmed that separation from nature is detrimental to human development, health and wellbeing, and that regular contact with nature is required for good mental health. People living more than 1 kilometer away from a green space have nearly 50 percent higher odds of experiencing stress than those living less than 300 meters from a green space.[1] In addition, relaxation has many health benefits and reduces stress and the symptoms of mental health conditions like depression, anxiety and schizophrenia[2]. Spas and gyms can facilitate in such relaxations. We plan to

[1] https://www.nrpa.org/our-work/Three-Pillars/health-wellness/ParksandHealth/fact-sheets/parks-improved-mental-health-quality-life/

[2] https://www.healthdirect.gov.au/relaxation

present through our analysis, a direct comparison of accessibility of facilities related to the health and wellbeing of people between neighborhoods with highest and lowest median income.

The analysis done in this project would be useful for agencies involved in social services, as well as those involved in city planning. It’s true that higher (or highest) income neighborhoods tend to be located in places which are geographically nicer. But what can city planners do to address this issue? Poverty or living in low (including lowest) median income neighborhoods cannot and must not automatically disqualify a person or a family from enjoying nature or having access to a place dedicated to promote relaxation. As is sometimes the case, correlation does not equate to causation and we do not claim any causation through the correlations that we might find in our research. That is beyond the scope of this project and would need further study and analysis.

This project includes gathering median income data of neighborhoods located within a significantly large metropolitan area in the United States of America (from hereon referred to as the US), use Foursquare to find different venues in the neighborhoods of our choice (high income and low income with a minimum population threshold), perform K-Means clustering to cluster the neighborhoods based on their income and population similarities, and then analyze the results. We choose Los Angeles County for our project.

Los Angeles County (from hereon referred to as LA County) is located in the state of California, the largest state by population in the US[3]. It is the most populous county in the US, with more than 10 million people[4] as of 2018. It includes the city of Los Angeles, the largest city in California, with an estimated population of nearly 4 million people[5]. LA County is home to countless Hollywood celebrities and is known around the world for the glitz and glamor that accompany celebrity lifestyles. However, LA County has a poverty rate of 17%, which is more than the national average of 13.1%[6]. Its diverse population is comprised of 9% Black or African American, 15.4% Asian, 48.6% Hispanic or Latino (may include other races) and 26.1% White (not Hispanic or Latino)[7]. A significantly large metropolitan area with varying income brackets, we find that LA County suits the purpose of our research well. An excellent area of future study would be whether this analysis changes based on the racial demography of the neighborhoods, in addition to the median income.

---

3 https://en.wikipedia.org/wiki/California

4 https://en.wikipedia.org/wiki/Los_Angeles_County,_California

5 https://en.wikipedia.org/wiki/Los_Angeles

6 https://datausa.io/profile/geo/los-angeles-county-ca

7 https://www.census.gov/quickfacts/losangelescountycalifornia#qf-headnote-b

Data

To analyze the different values based on the median income of LA County neighborhoods, we will need the following data:

- Neighborhood names
- Median income of each neighborhood
- Population of each neighborhood (this is because we want to include neighborhoods with a population of at least 25,000 only)
- Latitude and longitude location of each neighborhood, because we will use Foursquare to obtain venues in each neighborhood

We found several interesting datasets on the University of Souther California's website on Neighborhood Data for Social Change[8]. They include datasets for LA County and Coachella Valley for demography, income, education, health, real estate, etc.

1. Median Income

We found a dataset with median income information based on LA County neighborhoods for years 2006-2018[9]. First we cleaned up the dataset by dropping columns not needed, like the ones with location and GEOID information. Next, we filtered the year 2018 because that is the latest year that the data is available for and we want to focus on only one year. The dataset had several entries for each neighborhood, based on specific locations at the sub level of each neighborhood. However, to keep our analysis simple, we just want the median income at the neighborhood level. So we took the mean of all the median income entries of each neighborhood and dropped the latitude/longitude information at the sub level. We will obtain that information at the neighbor-

[8] https://usc.data.socrata.com/stories/s/htr6-r22g

[9] https://usc.data.socrata.com/stories/s/u9tw-axc2/

hood level from elsewhere. In the end, we got data for all LA County neighborhoods in the following format:

Median Income	Neighborhood

2. Population

We found a dataset with population information based on LA County neighborhoods for 2006-2018[10]. Similar to the previous dataset on median income, we first cleaned up the dataset by dropping columns not needed, like the ones with location and GEOID information. Next, we filtered the year 2018 because that is the year we will do our analysis on. The dataset had several entries for each neighborhood as in the previous dataset, based on specific locations at the sub level of each neighborhood. However, as we discussed before, we will do our analysis at the neighborhood level and so we aggregated the sum of all entries within each neighborhood to obtain the total population of each neighborhood. We also dropped the latitude/longitude information at the sub level and as discussed previously, we will obtain the location information of each neighborhood from elsewhere. In the end, we got data for all LA County neighborhoods in the following format:

Population	Neighborhood

3. Location:

With a list of the neighborhoods that we had the population and median income data for, we used Nominatim to find the latitude and longitude location for each neighbor-

[10] https://usc.data.socrata.com/stories/s/t8xm-yvpp/

hood. We had to drop a few neighborhoods (<15) from our dataset because their location wasn't available on Nominatim. We added the location to the median income data to get the following for each neighborhoods:

Neighborhood	Median Income	Latitude	Longitude

Finally, we merged the population data with the median income data:

Neighborhood	Median Income	Population	Latitude	Longitude

<u>4. Final data clean-up:</u>

We obtained data for over 250 neighborhoods in LA County. First, we want to focus only on neighborhoods with a considerable population. So we dropped the neighborhoods with less than 30,000 people. This number was picked arbitrarily. Further research needs to be conducted, and is outside the scope of our project, to determine what the minimum population threshold should be. We then filtered our data down to 20 neighborhoods with the lowest median income and 20 neighborhoods with the highest median income. These are the 2 datasets we will be performing our analysis on.

<u>5. How data will be used:</u>

Now that we have our datasets of lowest and highest median income neighborhoods in LA County along with their location information, we will perform K-Means Clustering on the data to cluster the neighborhoods based on their population and median income. Then we will use Foursquare API to find the 20 most common venues in each neighborhood. We will analyze the type of venues that Foursquare returns for each neighborhood and group them according to the type. For example, we will group all restaurant and food related venues together, grocery and convenience stores together, parks, trails and other natural areas together, etc. We will analyze the findings and finally visually present how accessibility to places dedicated to the health and wellbeing of people compares between neighborhoods with the lowest median income and those with the highest median income.

Methodology and Results

After data scrambling and data clean up, the datasets for the lowest and highest median income neighborhoods with a population of over 30,000 in LA County are as follows.

	Neighborhood	Population	Median Income	Latitude	Longitude
0	Watts	45364	31559.11	33.9406	-118.243
1	Pico-Union	42250	33050.08	34.0466	-118.288
2	South Park	38810	34388.12	34.0404	-118.267
3	Historic South-Central	44932	35073.64	34.0162	-118.267
4	Westlake	112237	35582.20	34.0629	-118.273
5	Westmont	33723	35625.43	33.9414	-118.302
6	Vermont-Slauson	34975	36002.88	33.9837	-118.292
7	Exposition Park	34253	36646.86	34.0137	-118.287
8	Florence	53591	37205.23	33.9742	-118.243
9	Central-Alameda	45709	37793.90	34.004	-118.248
10	Vermont Square	46335	38734.00	34.0019	-118.3
11	Boyle Heights	93528	39300.44	34.0437	-118.21
12	Koreatown	104294	39866.16	34.0618	-118.305
13	Florence-Firestone	64162	40171.00	33.9674	-118.243
14	Green Meadows	40766	40592.25	33.941	-118.263
15	East Hollywood	66648	40994.22	34.0904	-118.297
16	Hyde Park	35852	41030.60	33.9854	-118.331
17	Bell Gardens	42641	41041.78	33.9695	-118.15
18	Huntington Park	58694	41869.56	33.9827	-118.212
19	Bell	35270	42517.00	33.9748	-118.187

Fig 1. Neighborhoods with the lowest median income

	Neighborhood	Population	Median Income	Latitude	Longitude
0	Brentwood	33846	164631.20	34.0521	-118.474
1	Manhattan Beach	35573	159236.50	33.8916	-118.395
2	Rancho Palos Verdes	44986	134225.44	33.7483	-118.371
3	Beverly Hills	34362	117823.00	34.0697	-118.396
4	West Hills	39577	111965.67	34.2032	-118.645
5	Redondo Beach	67700	111911.85	33.8456	-118.389
6	Woodland Hills	59012	106937.50	34.1684	-118.606
7	Encino	48544	106607.30	34.1591	-118.502
8	Walnut	30008	106054.25	34.0203	-117.865
9	Studio City	43458	102861.44	34.1484	-118.396
10	Venice	36149	101770.45	33.995	-118.467
11	Cerritos	50172	101168.00	33.8644	-118.054
12	Santa Clarita	190304	99502.98	34.3917	-118.543
13	Santa Monica	92078	98199.32	34.0251	-118.497
14	Granada Hills	58088	98151.00	34.2662	-118.517
15	Chatsworth	52706	98002.62	34.2596	-118.602
16	Altadena	44389	96364.00	34.1863	-118.135
17	Claremont	38195	96302.50	34.0967	-117.72
18	Arcadia	54967	96191.00	34.1362	-118.04
19	Diamond Bar	46346	95933.11	34.0286	-117.81

Fig 2. Neighborhoods with the highest median income

Let's take a look at the population bar graphs of each of the datasets.

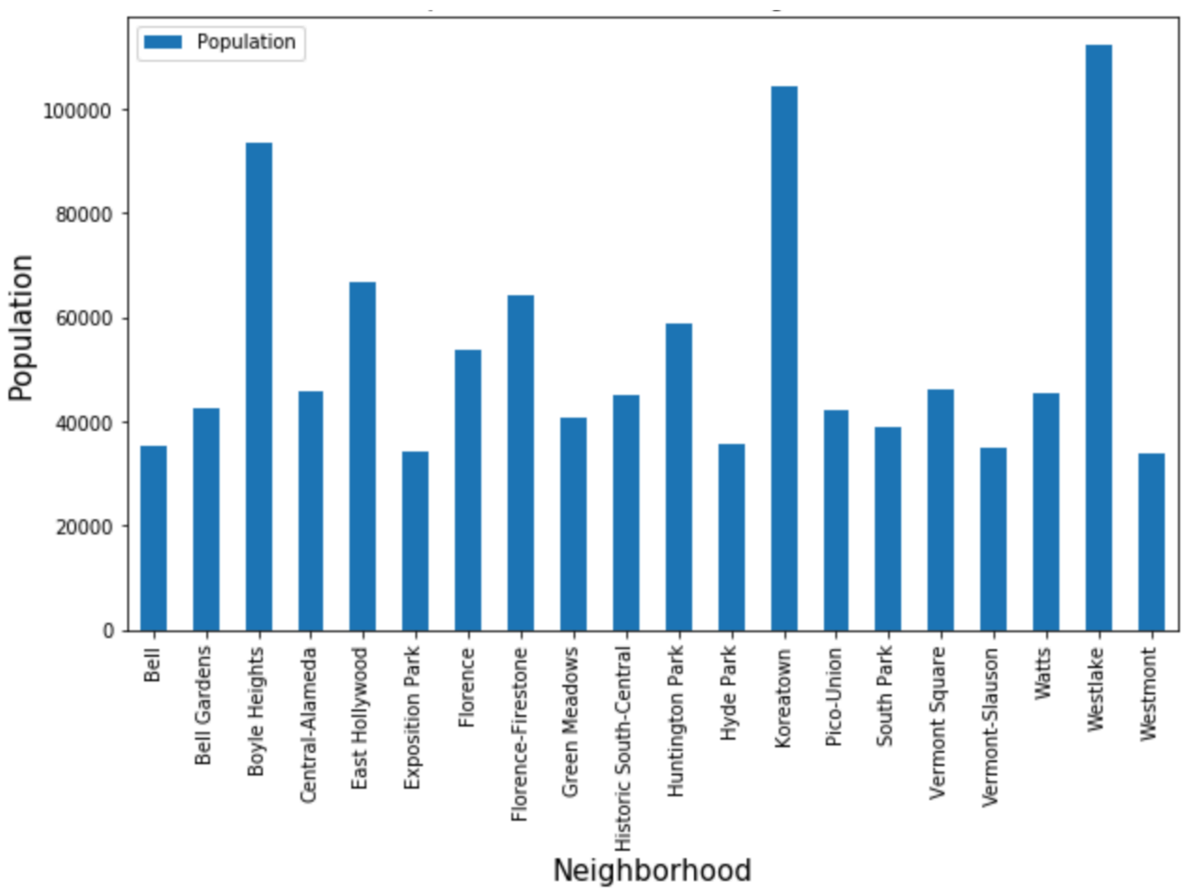

Fig 3. Population for neighborhoods with the lowest median income

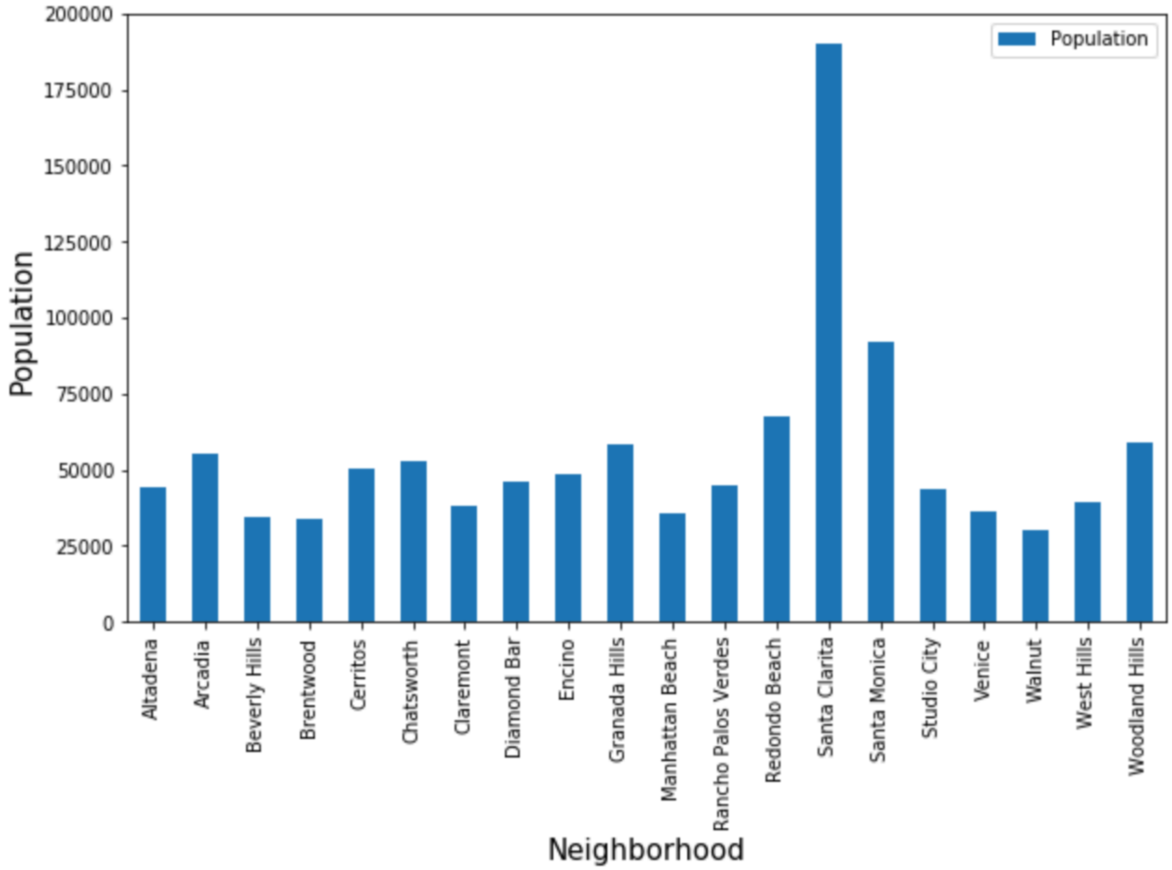

Fig 4. Population for neighborhoods with the lowest median income

We used Folium to visualize the locations of the neighborhoods on the map of LA County.

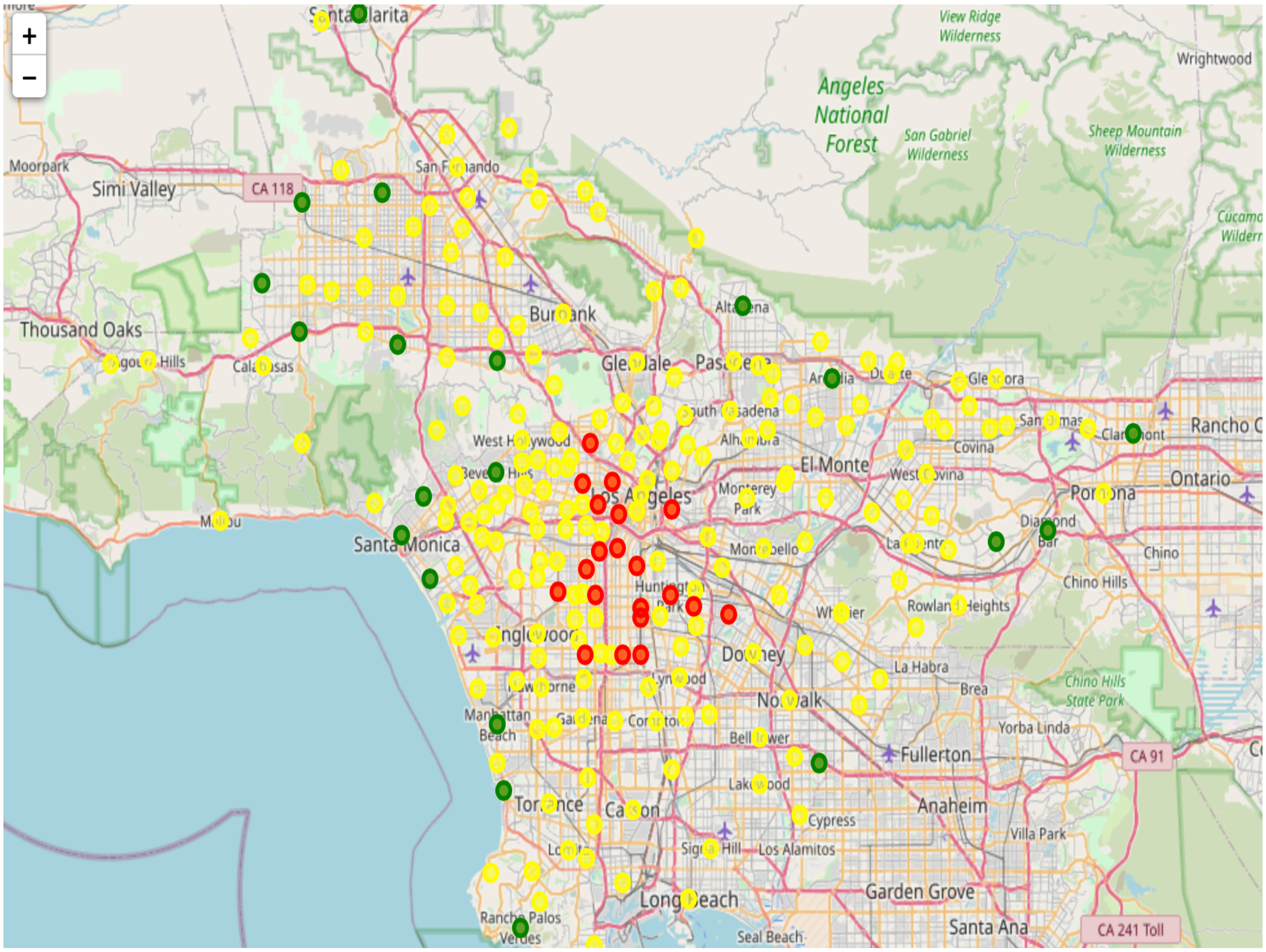

Fig 5. Map of LA County neighborhoods

We see that the neighborhoods with the highest median income (indicated by red dots) are mostly located along the coast and in the more scenic areas. The neighborhoods with the lowest median income (indicated by the green dots) are mostly located within inner city limits.

We want to use K-Means clustering and cluster the neighborhoods in each dataset (lowest and highest median income) based on their population and median income.

The distribution of population and median income are as follows.

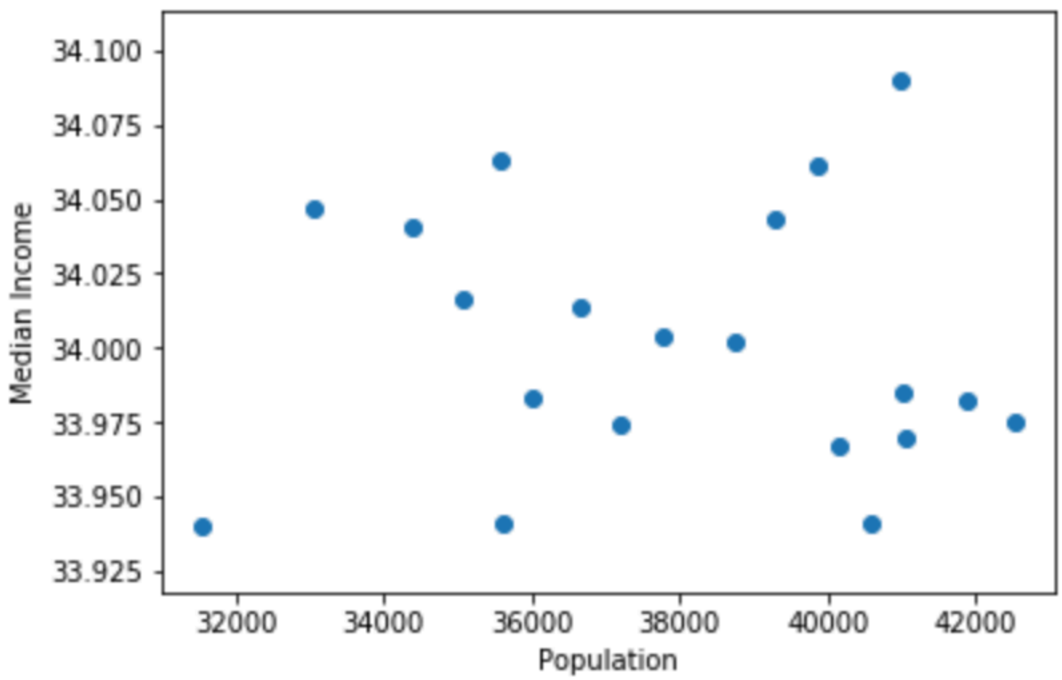

Fig 6. Distribution of Median Income and Population in Lowest Median Income Neighborhoods

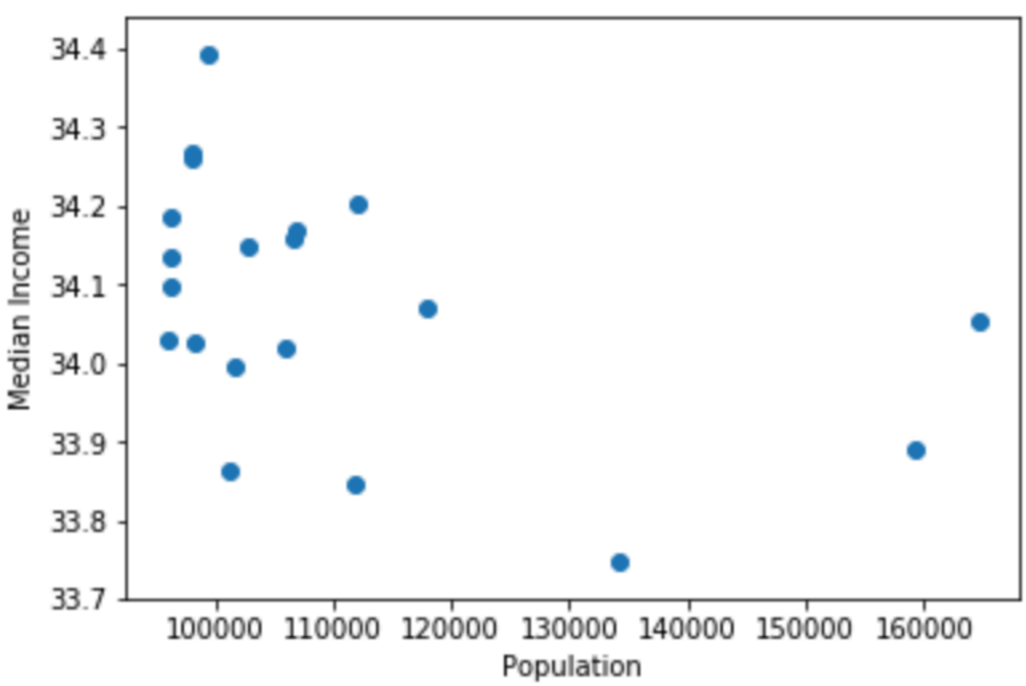

Fig 7. Distribution of Median Income and Population in Highest Median Income Neighborhoods

Before we apply K-Means Clustering on the datasets, we find the optimal values of K for each dataset by using the elbow method.

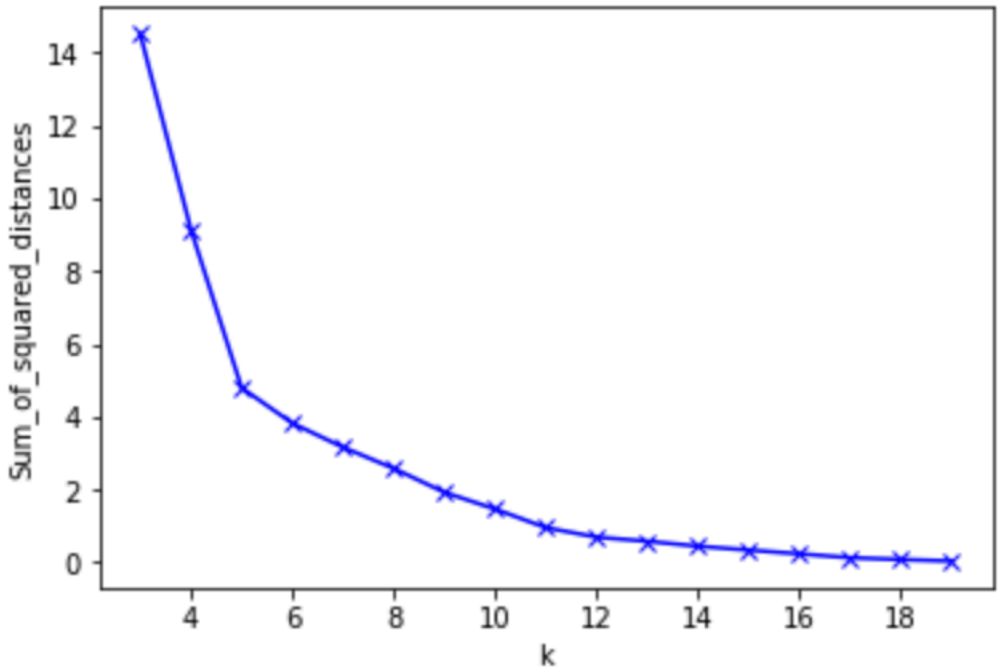

Fig 8. Elbow Method to Find Optimal K for Lowest Median Income Neighborhoods

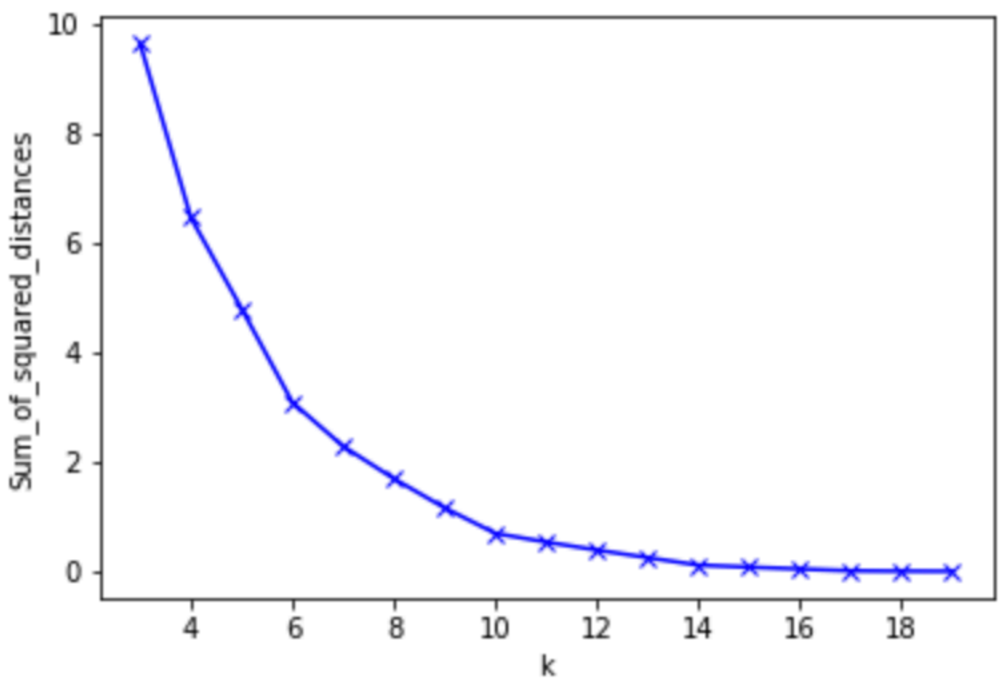

Fig 9. Elbow Method to Find Optimal K for Highest Median Income Neighborhoods

The distribution of the clusters for the datasets are as follows.

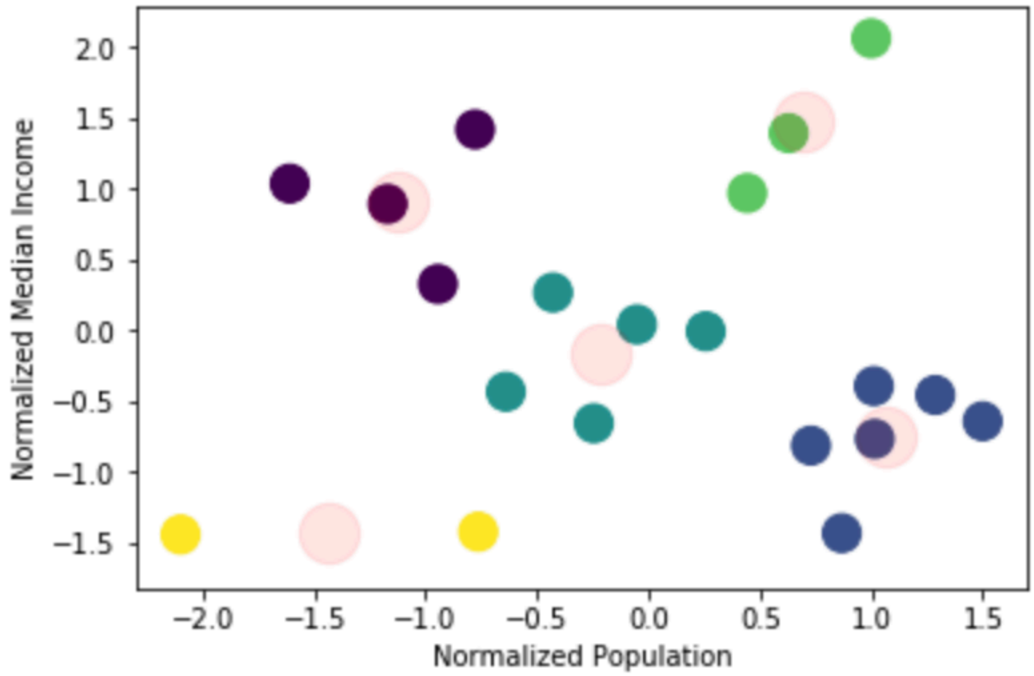

Fig 10. Lowest Median Income Neighborhood Clusters With Respect To Median Income and Population

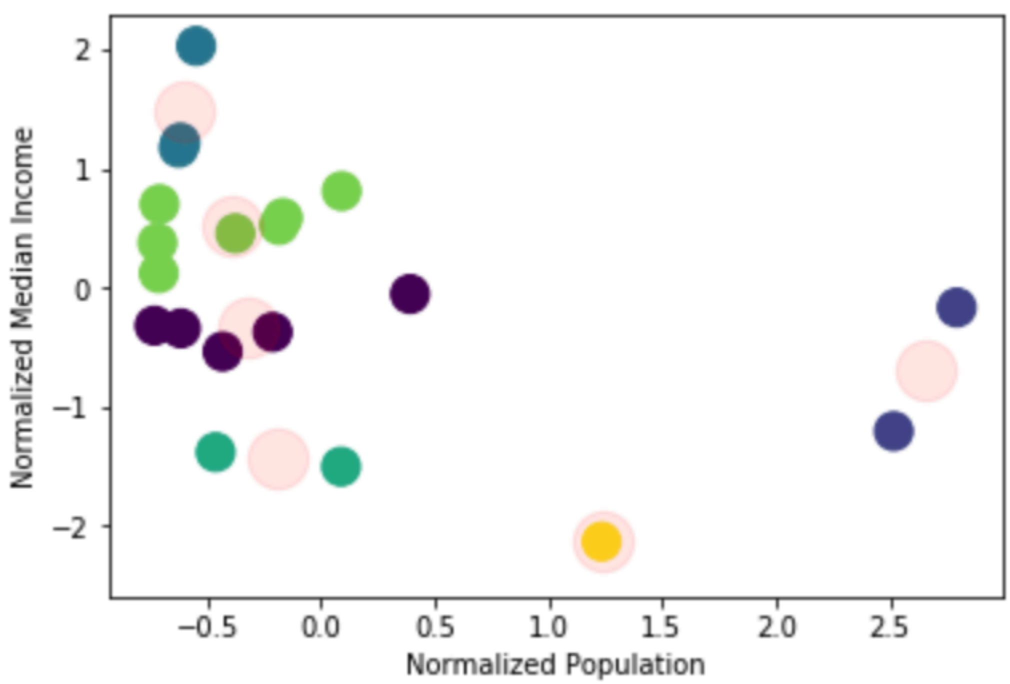

Fig 11. Highest Median Income Neighborhood Clusters With Respect To Median Income and Population

Using Folium, we map the clusters.

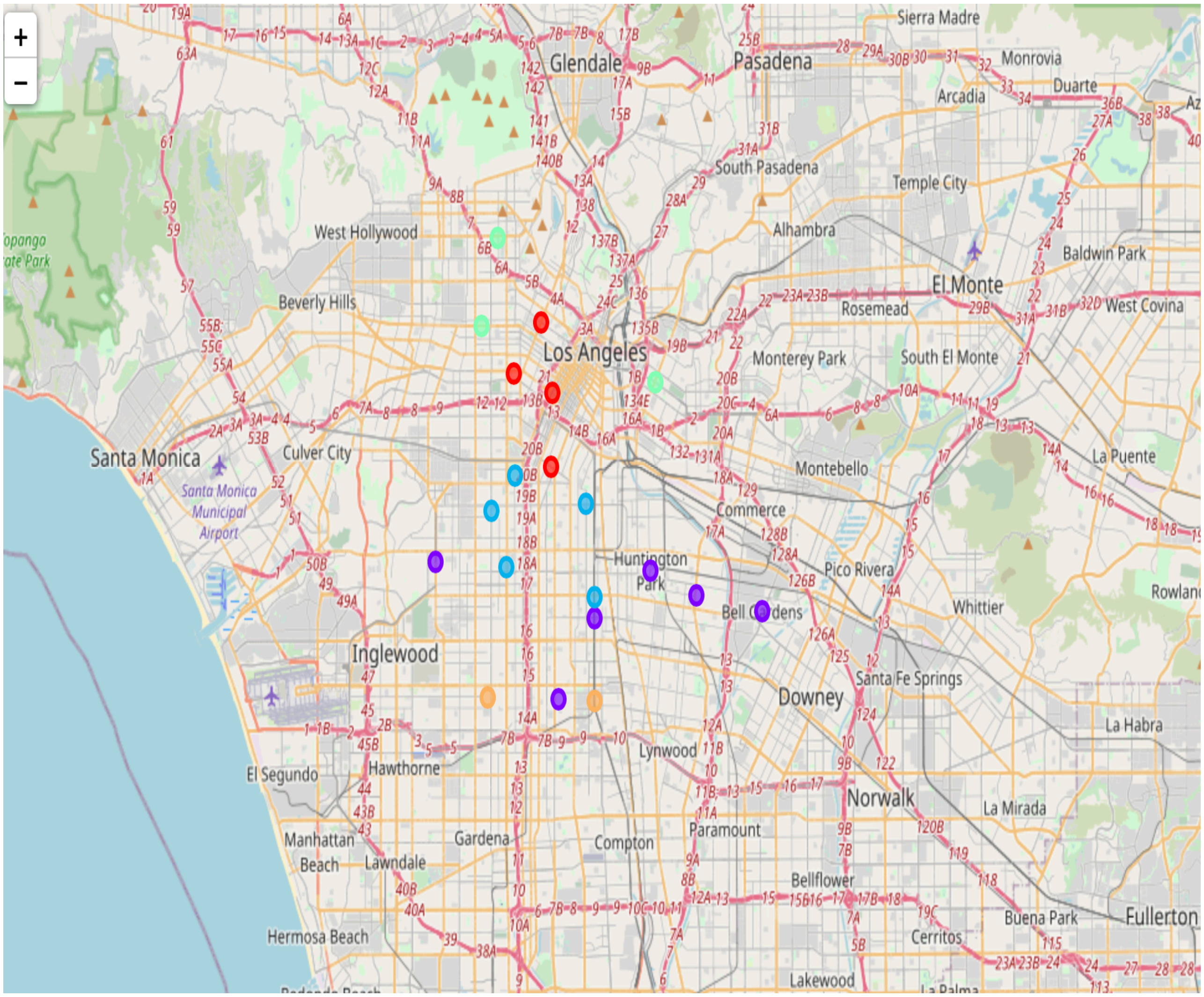

Fig 12. Map of Lowest Median Income Neighborhood Clusters

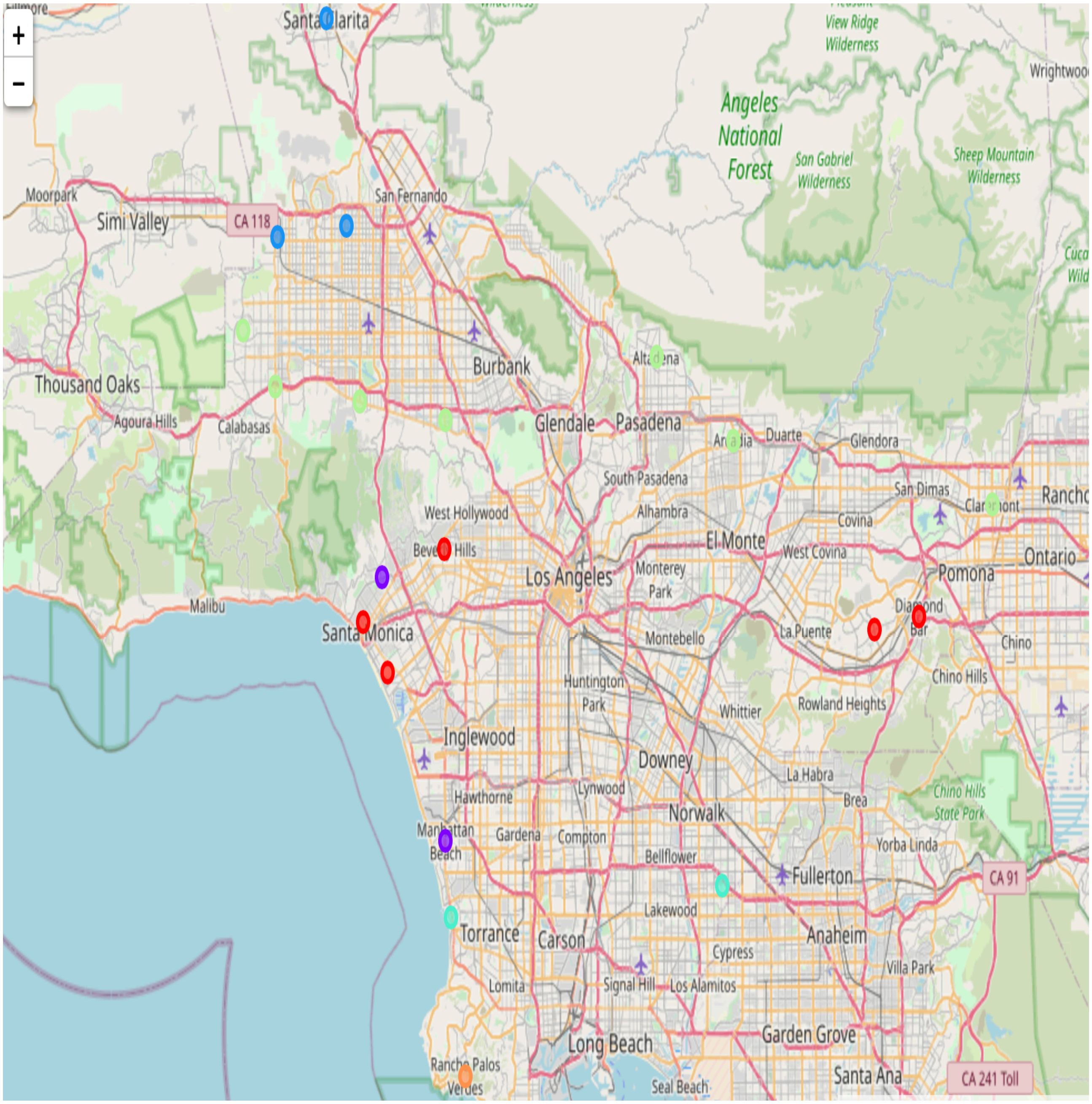

Fig 13. Map of Highest Median Income Neighborhoods Clusters

Using Foursquare API, we find all venues within 500 meters of each neighborhood location. Then we sort the venues to find the top 20 most common venues in each dataset.

	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	...	·
0	0	Mexican Restaurant	Coffee Shop	Food Truck	Fast Food Restaurant	American Restaurant	Clothing Store	Sports Bar	Bar	Grocery Store	...	D
1	1	Mexican Restaurant	Fried Chicken Joint	Burger Joint	Pizza Place	Sandwich Place	Donut Shop	Fast Food Restaurant	Pharmacy	Food	...	Co
2	2	Science Museum	History Museum	Mexican Restaurant	Burger Joint	Museum	Sandwich Place	Park	Food Court	Aquarium	...	
3	3	Korean Restaurant	Café	Bakery	Restaurant	Grocery Store	Ice Cream Shop	Coffee Shop	Burger Joint	Convenience Store	...	F
4	4	Donut Shop	Fast Food Restaurant	Performing Arts Venue	Mobile Phone Shop	Sculpture Garden	Sandwich Place	Café	Bank	Fried Chicken Joint	...	

Fig 14. Most Common Venues in Lowest Median Income Neighborhoods

	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	...
0	0	Mexican Restaurant	Smoke Shop	Convenience Store	Pizza Place	Park	Business Service	Fast Food Restaurant	Food Truck	Home Service	...
1	1	Seafood Restaurant	Bar	Coffee Shop	Juice Bar	Hotel	Café	Burger Joint	Diner	Pharmacy	...
2	2	Coffee Shop	Boutique	Café	Bank	American Restaurant	Clothing Store	Sandwich Place	Pizza Place	Park	...
3	3	Coffee Shop	Italian Restaurant	Bank	Clothing Store	Gym / Fitness Center	Sandwich Place	Hotel	Sports Bar	Pharmacy	...
4	4	Coffee Shop	Mediterranean Restaurant	Ice Cream Shop	Sushi Restaurant	Pizza Place	Bakery	Spa	American Restaurant	Arts & Crafts Store	...
5	5	Sporting Goods Shop	Furniture / Home Store	Trail							...

Fig 15. Most Common Venues in Highest Median Income Neighborhoods

For simplicity, we group the venues as follows:

Groceries - groceries and convenience stores

Food - all food related businesses, except health foods and juice bars

Finance - money/Bank/ATM related businesses

Transport - transportation services like bus stops, train stations, etc.

Shopping - shopping like clothing stores

Parks - parks, gardens and nature trails

Lifestyle - lifestyle stores like spas, gyms, health foods, juice bars, etc.

Health - health related like hospitals, pharmacies, clinics

Entertainment - entertainment like theaters, museums, aquariums

Finally, we count the number of each venue type.

	Cluster Label	Groceries	Food	Finance	Transport	Shopping	Parks	Lifestyle	Health	Entertainment
0	0	1	11	0	0	1	0	0	0	0
1	1	2	11	0	0	1	1	0	1	0
2	2	1	6	0	1	1	1	0	0	5
3	3	2	15	0	0	0	0	0	1	0
4	4	1	6	1	1	0	1	0	1	1

Fig 16. Venue Count by Type in Lowest Median Income Neighborhoods

	Cluster Label	Groceries	Food	Finance	Transport	Shopping	Parks	Lifestyle	Health	Entertainment
0	0	1	8	0	0	0	1	0	0	0
1	1	1	8	0	0	0	1	2	1	0
2	2	0	11	1	0	2	1	2	0	0
3	3	0	5	1	1	1	0	2	1	0
4	4	1	11	1	0	0	1	1	0	0
5	5	0	0	0	0	0	1	0	0	0

Fig 17. Venue Count by Type in Highest Median Income Neighborhoods

We visualize a comparison of the types of venues in each of the two datasets.

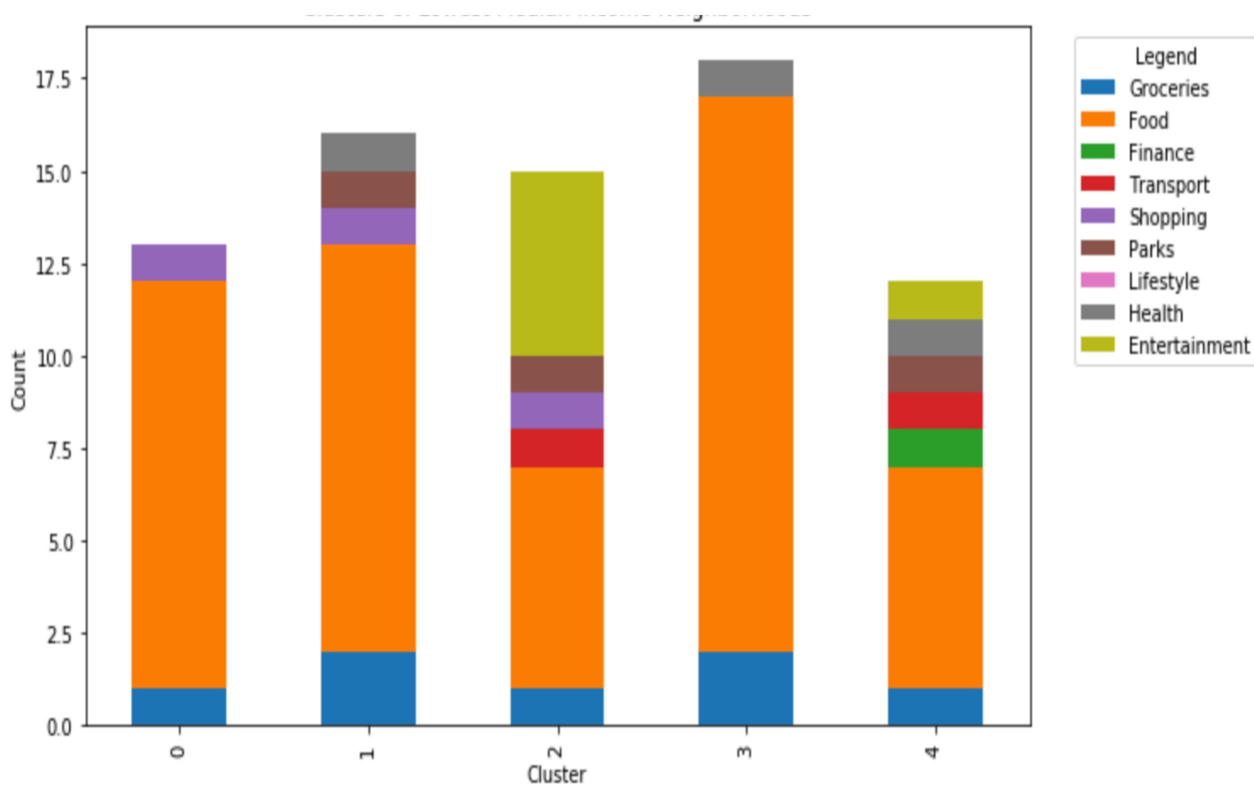

Fig 18. Bar Chart of Venue Count by Type in Lowest Median Income Neighborhoods

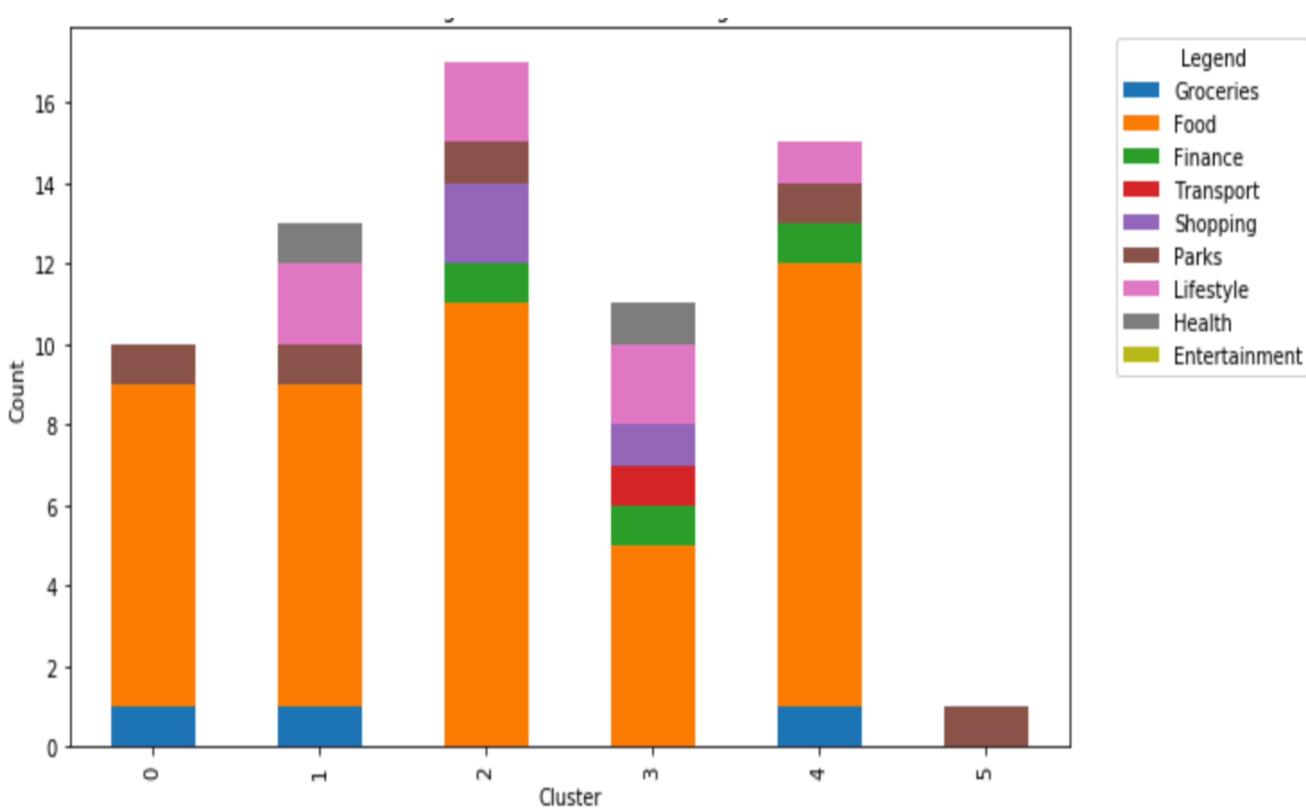

Fig 19. Bar Chart of Venue Count by Type in Highest Median Income Neighborhoods

We combine the Parks and Lifestyle categories to provide a final comparison between the two datasets.

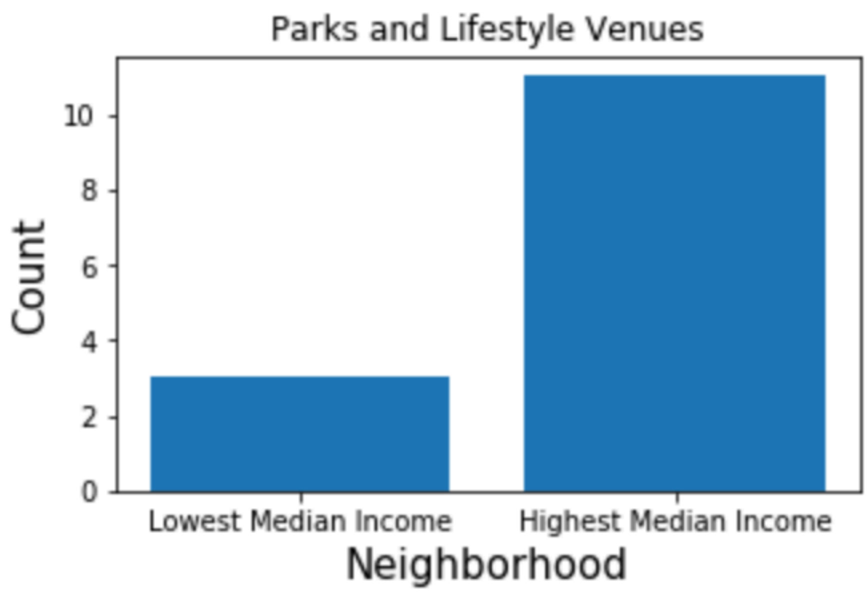

Fig 20. Comparison of Number of Relaxation Venues in Lowest and Highest Median Income Neighborhoods

CONCLUSION

We see from Fig 18 and Fig 19 that the types of venues vary in the lowest median income neighborhood clusters and the highest median income neighborhood clusters. One of the main differences we see is the number of venues that are identified by Foursquare as parks, trails, beaches and other natural areas. They are more common in the neighborhoods with the highest median income, which tend to be located in more naturally nicer areas. Lower income neighborhoods tend to be located within inner city limits and access to parks and trails may be limited. In addition, venues that are related toward healthy lifestyle or overall relaxation, like spas and gyms, are often provided by businesses that require membership and can be relatively expensive.

Therefore, we see in our results that these are also more common in neighborhoods with the highest median income. In fact, none of the five clusters in the lowest median income neighborhoods include spas or gyms.

Finally, in Fig 20 we see a direct comparison between lowest median income and highest median income neighborhoods, with respect to the number of venues related to physical and overall health and wellbeing. We see that neighborhoods with the highest median income have three times more such venues than those with the lowest median income.

In conclusion, this analysis provides city planners and those involved with social welfare an insight into how to improve physical health and wellbeing of those living in neighborhoods with the lowest median income. An income level cannot and must not be a determinant in whether a person has easy access to a place which can provide an opportunity to relax. This is a necessity to a person's wellbeing and must be addressed in strategic planning regarding all neighborhoods.

We believe that our analysis also provides many opportunities for further research. For example, in addition to the median income, does a neighborhood's demography play a role in what types of amenities are accessible to the residents? Also, we can study other large metropolitan areas to see if we get similar results.

Code is available on Github.