Finding your perfect home!!

Problem description

In this project we will try to find the neighborhood in Buffalo,NY where we can buy our home based on our interests and needs. This will be acheived by an analytical approach, programatically go through the data, using maching learning techniques and visualizing using various python packages.

Data will have to collected from different sources, cleaned up and transformed to suit our needs to extend the analysis in acheiving the goal of finding the place where our next home is.

Data Source

We will be using Buffalo neighborhood data file available in the following website. This file contains lot of details which may not be of interest to us for this analysis. We can trim this file to suit our needs. This data set contains the demographics, economic and housing characteristics data for each neighborhood in the Buffalo area. The criteria used for analysis in finding the best neighborhood for buying the home are as follows.

1. Neighborhood with considerable asian population
2. Neighborhood with newer homes
3. Neighborhood with Restauarants https://data.buffalony.gov/stories/s/a235-4wxj (https://data.buffalony.gov/stories/s/a235-4wxj) -- Data Source

The Foursquare Api: This data will be accesed via Python, and used to obtain the most common venues per neighborhood in the city of Buffalo. This will help in identifying where the restaurants and other eateries are situated in the city.

Basic exploration of the data

In [2]:
```
import pandas as pd
```

In [6]:
```
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
client_649a74950f1c4b0db92f74323e7bed40 = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='b7XT5XjhEc_aYdvNOiaI_Ds196f2xEb7J8pvJMFtpPn7',
    ibm_auth_endpoint="https://iam.ng.bluemix.net/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3-api.us-geo.objectstorage.service.networklayer.com')

body = client_649a74950f1c4b0db92f74323e7bed40.get_object(Bucket='capstoneproject-donotdelete-pr-pc7gsdf0moggqt',Key='Neighborhood_Metrics.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

df_data_1 = pd.read_csv(body)
df_data_1.head()
```

Out[6]:

	Neighborhood	Community	ID	Neighborhood Abbreviation	Labor Force Participation Rate	Employment Rate	Private Employment Rate	Government Employment Rate	Self-Employment Rate	Median Income	...	Percent Non-Family Households	Tot Livin Alon
0	Kensington-Bailey	East	16	KB	56.41	89.47	82.11	16.11	1.78	37240	...	41.98	220
1	Central	Central	1	CN	55.17	93.43	80.41	16.76	2.83	54390	...	59.35	76
2	Elmwood Bidwell	West	6	EBW	66.01	94.02	80.12	15.57	4.31	56150	...	76.23	373
3	Riverside	North	11	RS	56.06	95.04	80.33	16.14	3.52	21570	...	42.42	145
4	Fruit Belt	East	25	FB	40.25	89.76	75.15	14.04	10.82	25350	...	38.76	31

5 rows × 124 columns

In [16]:
```
df_data_1[['Neighborhood','Community','Latitude','Longitude','Median Income']]
```

Out[16]:

	Neighborhood	Community	Latitude	Longitude	Median Income
0	Kensington-Bailey	East	42.939776	-78.809881	37240
1	Central	Central	42.875966	-78.877250	54390
2	Elmwood Bidwell	West	42.924511	-78.874531	56150
3	Riverside	North	42.954950	-78.901916	21570
4	Fruit Belt	East	42.899160	-78.860605	25350
5	Pratt-Willert	East	42.887106	-78.860134	26330
6	University Heights	North	42.950995	-78.820645	42130
7	Delavan Grider	East	42.921940	-78.831565	28370
8	Kaisertown	South	42.871164	-78.809378	41280
9	Lower West Side	West	42.892692	-78.885380	18340
10	Elmwood Bryant	West	42.909564	-78.875978	40520
11	Lovejoy	East	42.887584	-78.812094	28780
12	Fillmore-Leroy	East	42.933075	-78.835398	22910
13	Seneca-Cazenovia	South	42.856939	-78.809195	38300
14	Black Rock	North	42.939284	-78.901827	24550
15	Central Park	North	42.948038	-78.837274	61940
16	West Side	West	42.905534	-78.892684	20040
17	Hopkins-Tifft	South	42.846165	-78.841956	36960
18	South Park	South	42.842726	-78.813528	55720
19	Schiller Park	East	42.913340	-78.805546	24020
20	Hamlin Park	East	42.921161	-78.848504	27860
21	Broadway Fillmore	East	42.890301	-78.839558	18940
22	North Park	North	42.951131	-78.860831	53870
23	Parkside	North	42.936658	-78.860027	54000
24	Upper West Side	West	42.922925	-78.893404	31390
25	Genesee-Moselle	East	42.906787	-78.818131	22080
26	Kenfield	East	42.927779	-78.808760	26700
27	First Ward	South	42.865960	-78.858467	27880

	Neighborhood	Community	Latitude	Longitude	Median Income
28	West Hertel	North	42.951184	-78.886673	25980
29	Ellicott	East	42.875295	-78.856977	22190
30	Masten Park	East	42.910959	-78.855598	22520
31	Grant-Amherst	North	42.938881	-78.888609	24385
32	Seneca Babcock	South	42.870176	-78.832377	30290
33	MLK Park	East	42.909308	-78.834789	20480
34	Allentown	West	42.898553	-78.876092	39060

This project can be used by anyone in the lookout for new home and trying the find out which area suits them the most.

Methodology

In this project, I will use folium to visualize the neighborhoods in city of Buffalo. I will use latitude and longitude data available in Buffalo neighborhood dataset along with foursquare API to fetch the different venues available in these neighborhoods.

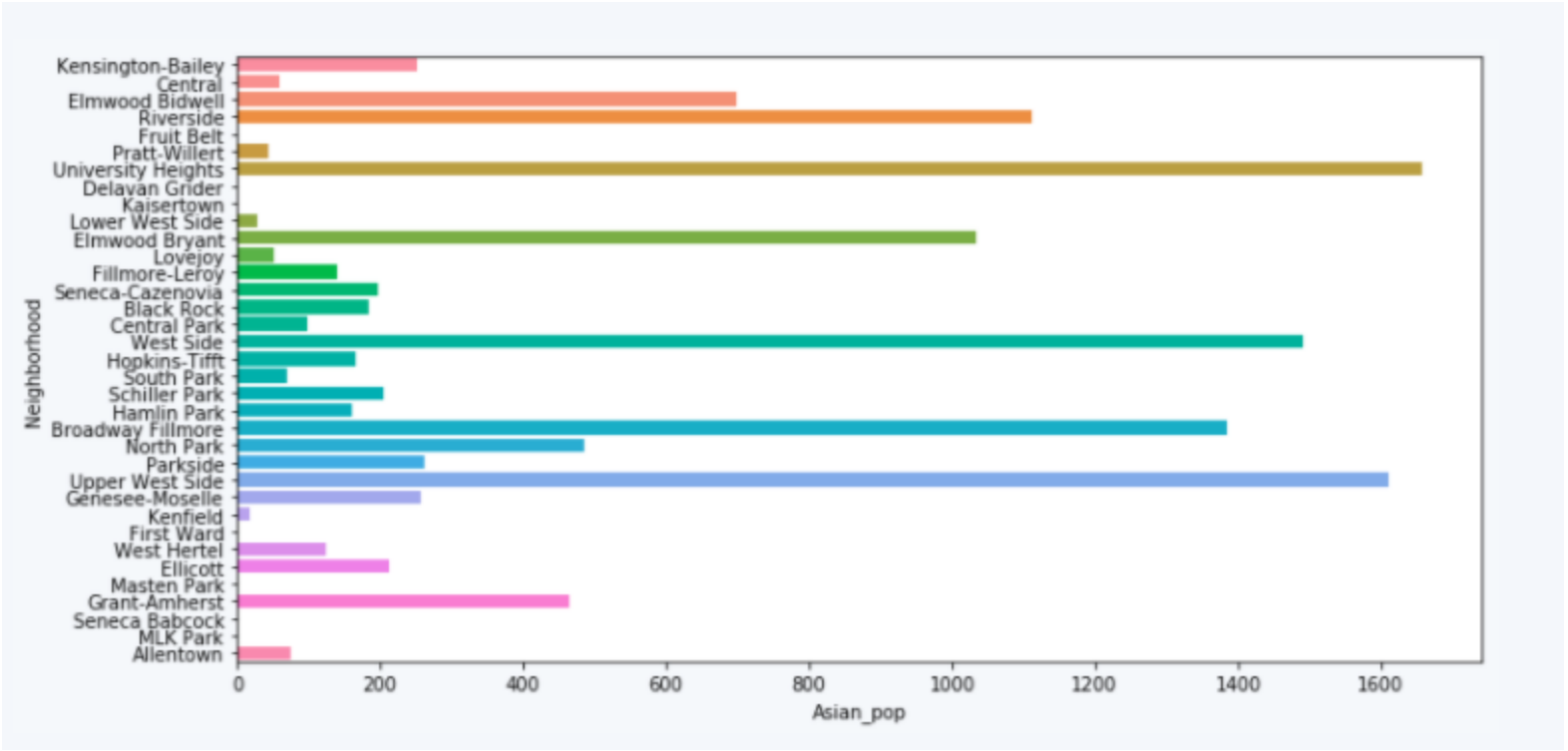
I will be creating a generalized category for these venues. These generalized categories will help us in choosing what's prevalent in which neighborhood.

Using this venue details,I will create One hot coding on Venue categories and Generalized categories. Finally, frequency will be calculated to find out how often a category occurs in particular neighborhood.

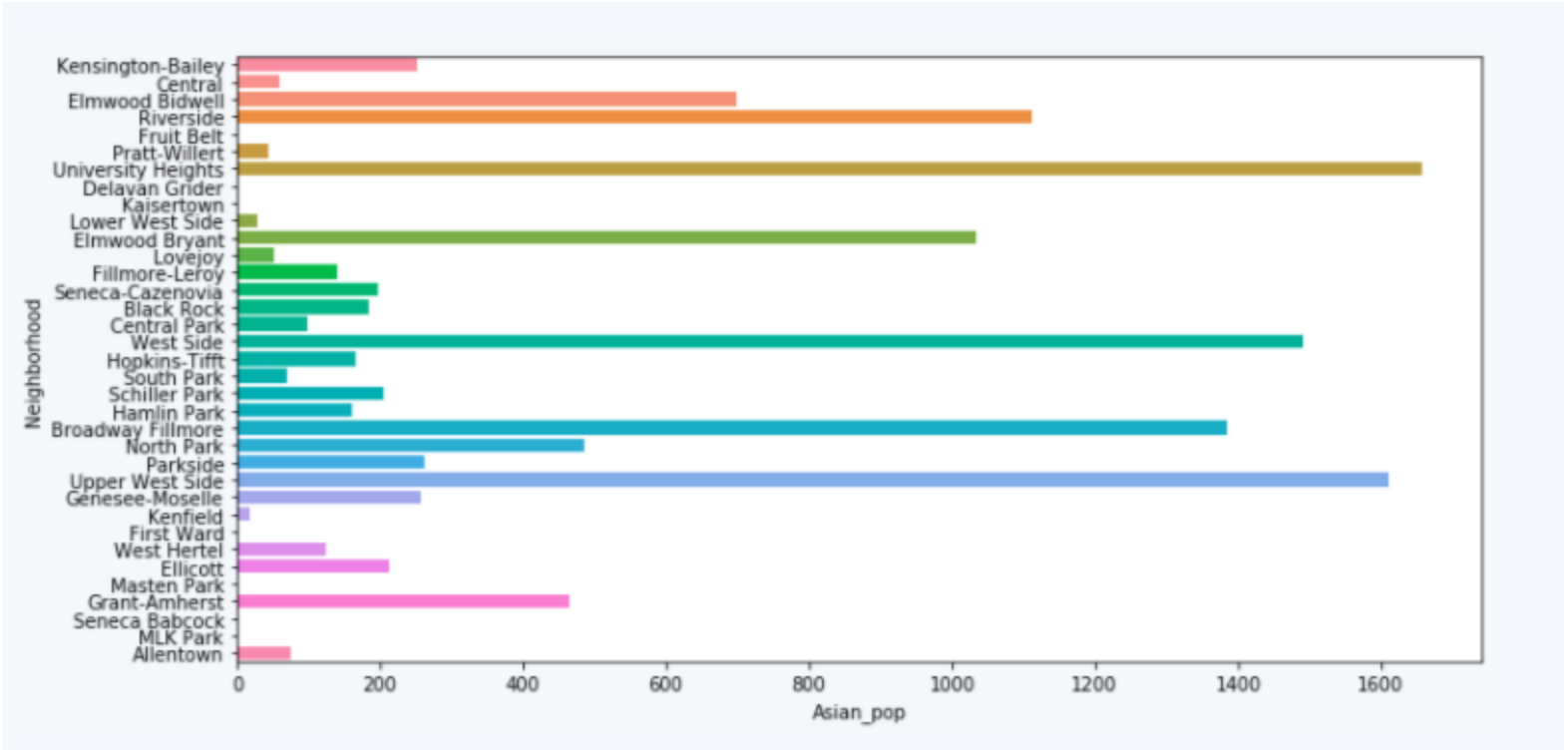
I will be using K-means clustering on this and cluster the neighborhoods to identify similar ones and choose the ones based on our interests. For simplicity I will be using k=5 in this project.

Results

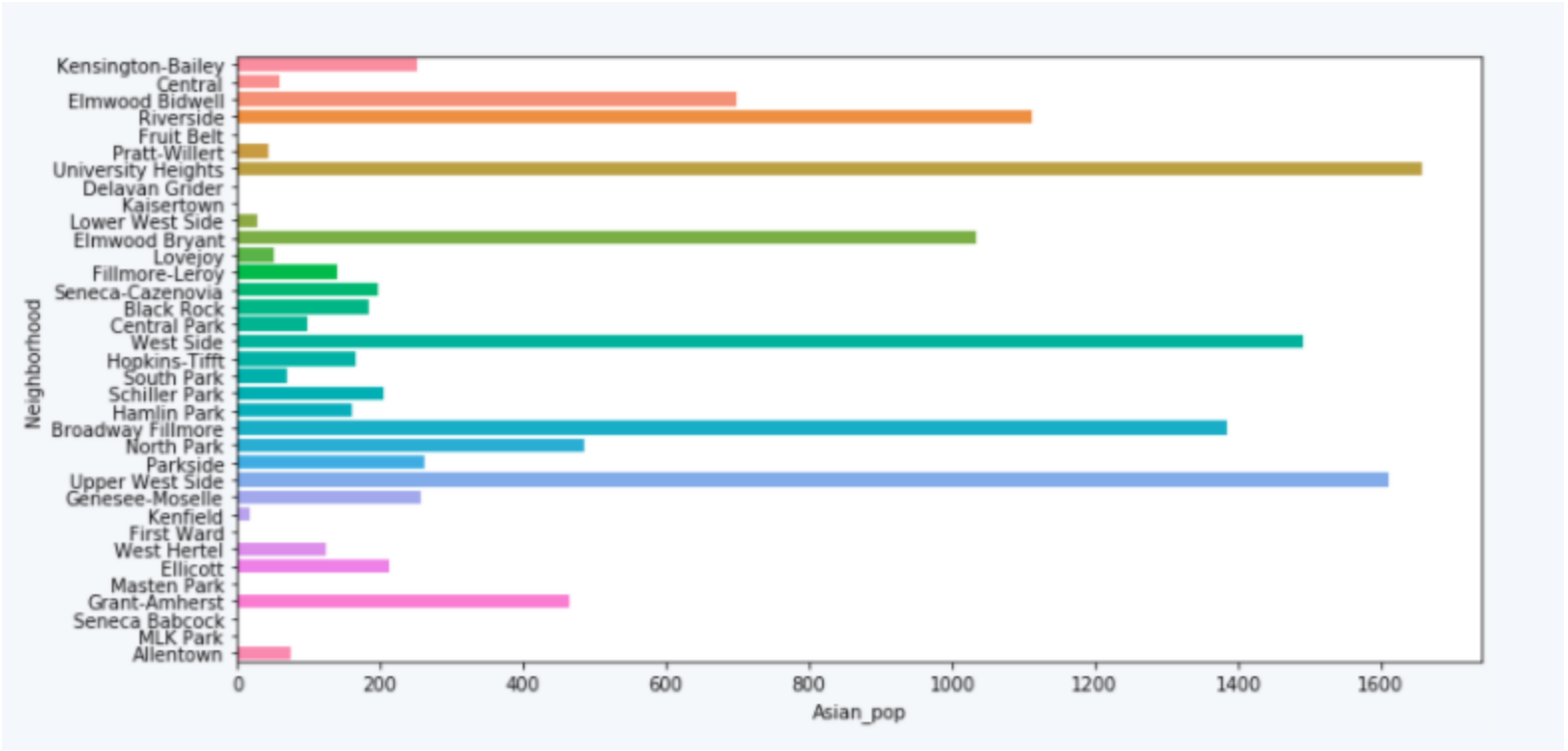
Asian Population Distribution



Units built past 2000



Visualizing Population, Housing, Income data



Top 5 neighborhoods with Asian Population, Median Income and Houses after 2000

Neighborhoods with Asian Population (Top 5)

University Heights

Upper West Side

West Side

Broadway Fillmore

Riverside

Neighborhoods with houses built after 2000 (Top 5)

Pratt-Willert

Masten Park

Lower West Side

Central

Ellicott

Neighborhood median income (Top 5)

Central Park

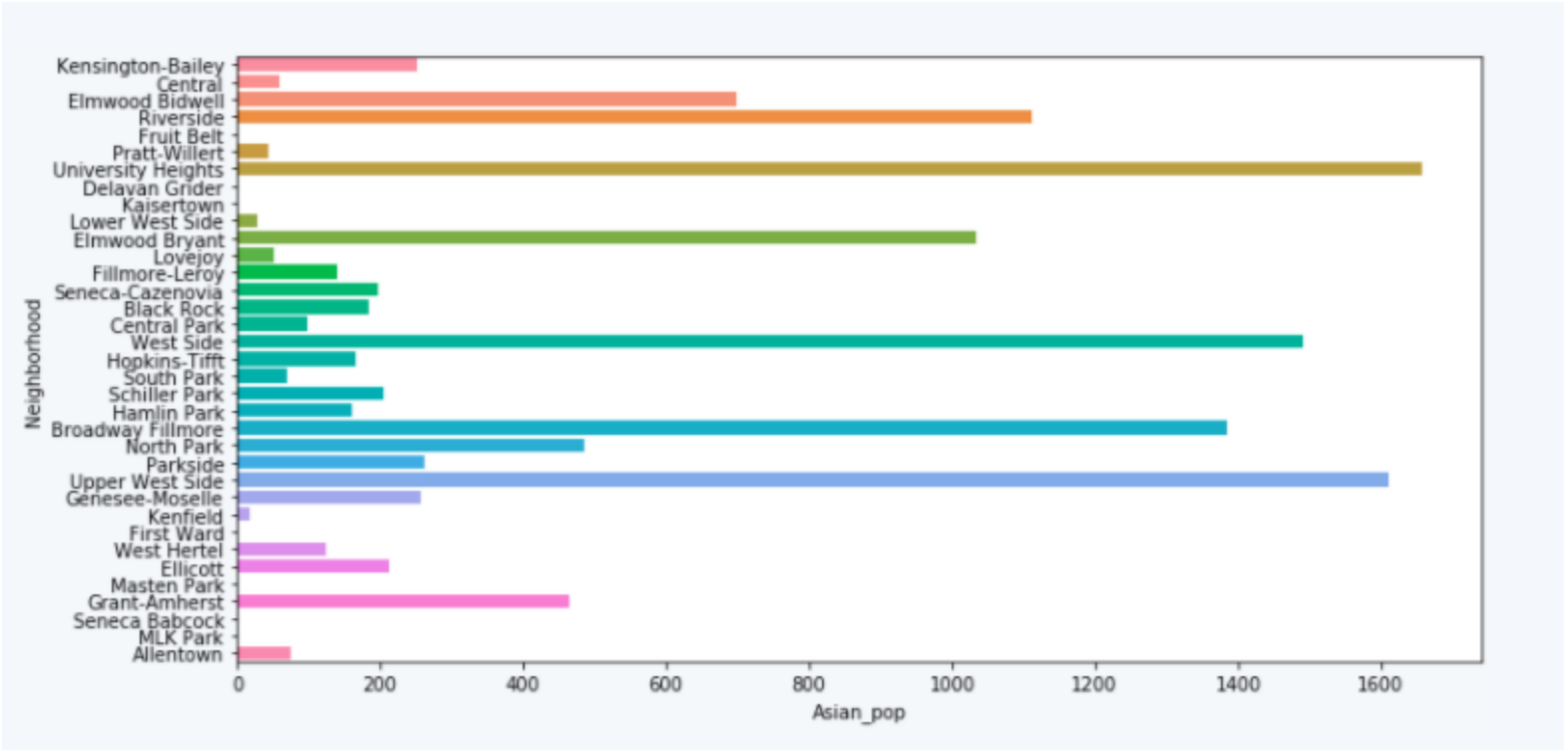
Elmwood Bidwell

South Park

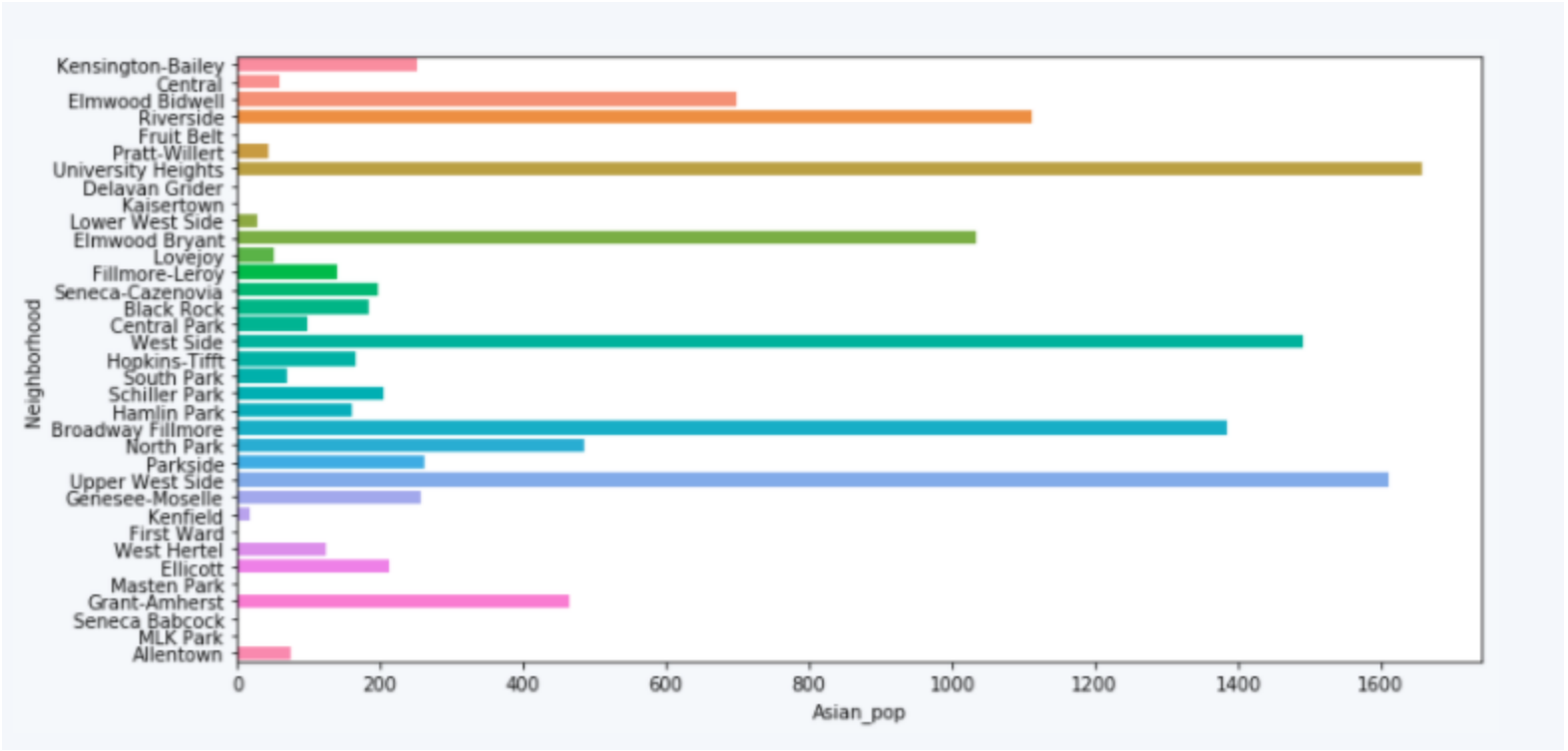
Central

Parkside

Buffalo Neighborhoods in map

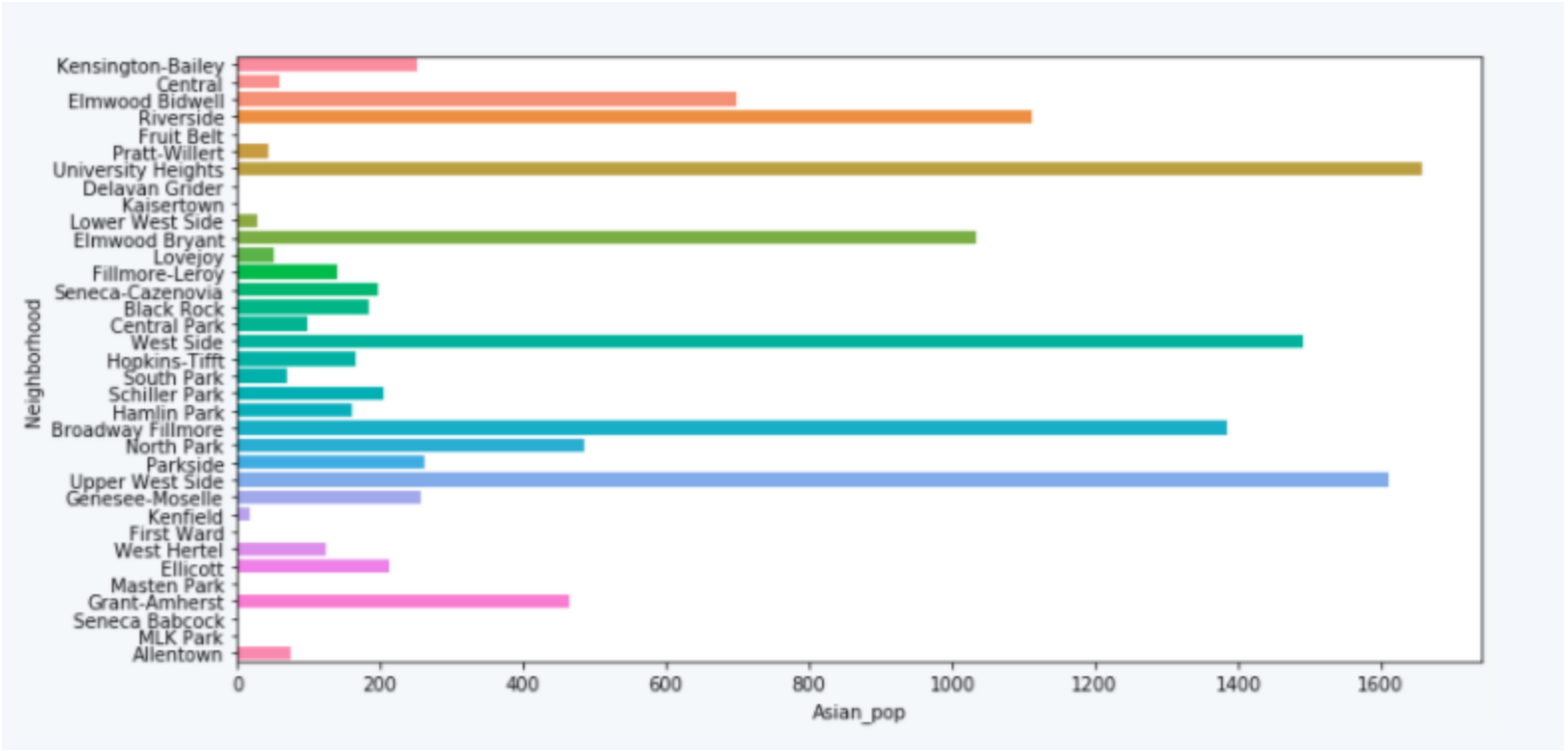


Neighborhood clusters

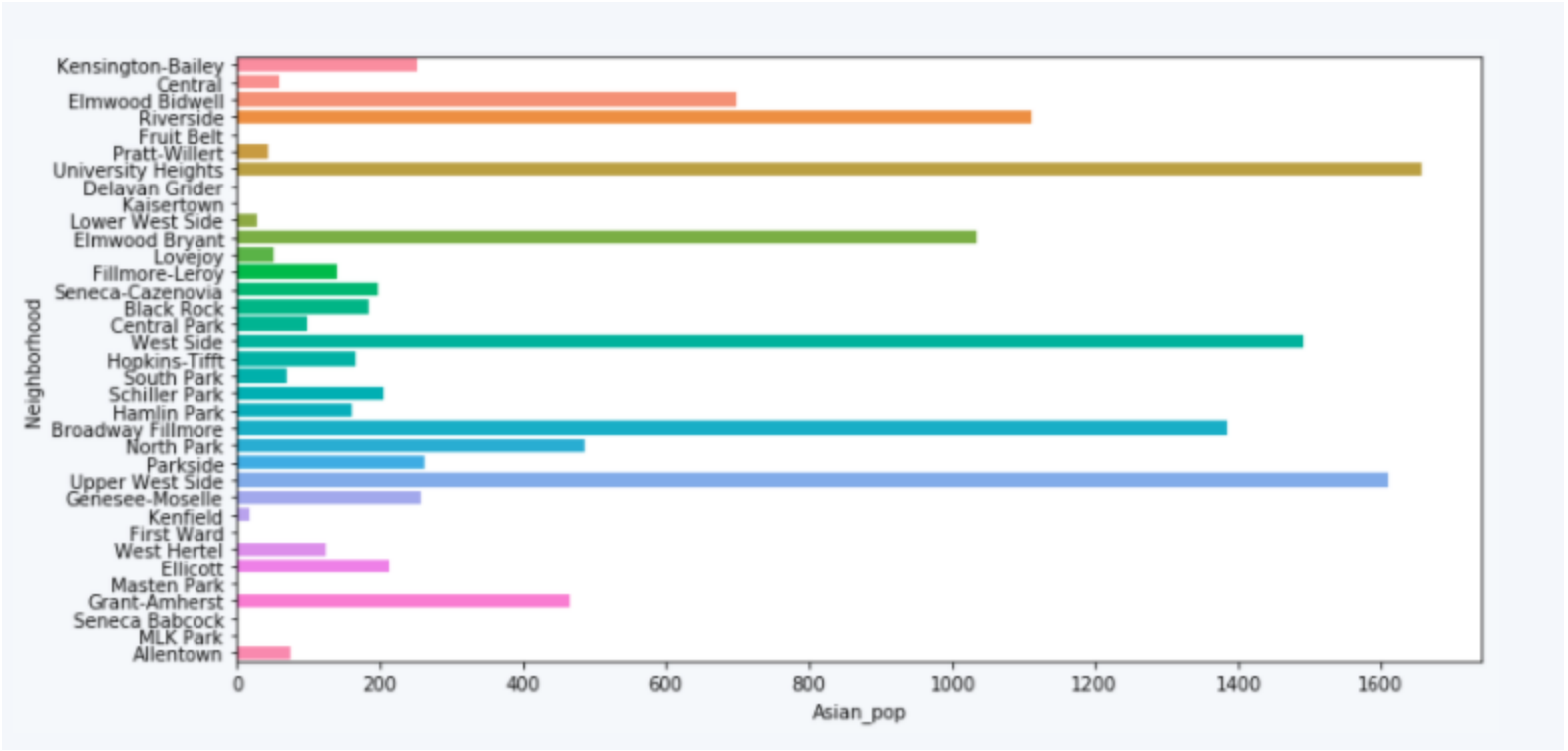


K Means Clustering

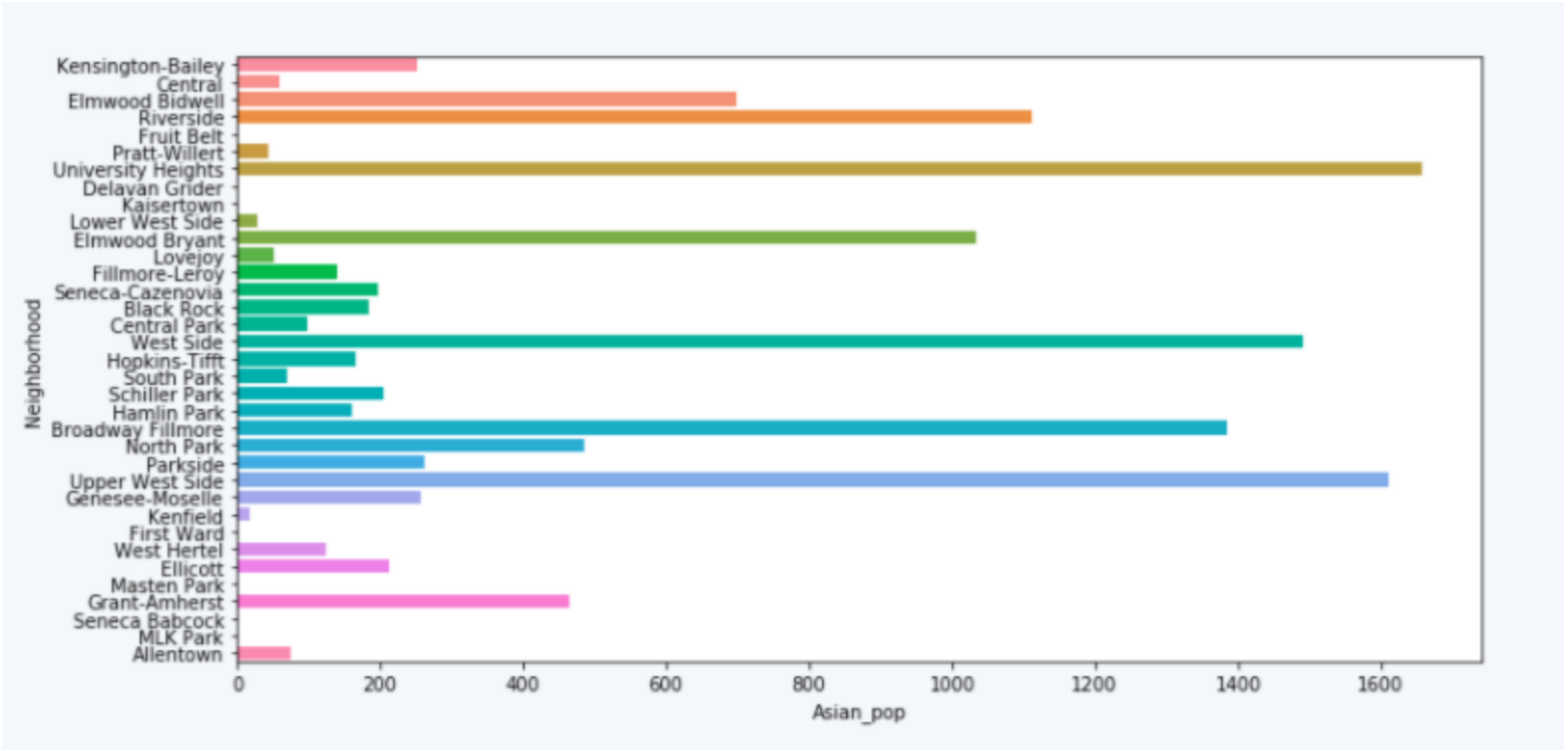
Cluster 0


Kensington-Bailey
Central
Elmwood Bidwell
Riverside
Fruit Belt
Pratt-Willert
University Heights
Delavan Grider
Kaisertown
Lower West Side
Elmwood Bryant
Lovejoy
Fillmore-Leroy
Seneca-Cazenovia
Black Rock
Central Park
West Side
Hopkins-Tifft
South Park
Schiller Park
Hamlin Park
Broadway Fillmore
North Park
Parkside
Upper West Side
Genesee-Moselle
Kenfield
First Ward
West Hertel
Ellicott
Masten Park
Grant-Amherst
Seneca Babcock
MLK Park
Allentown
Neighborhood
0
200
400
600
800
1000
1200
1400
1600
Asian_pop

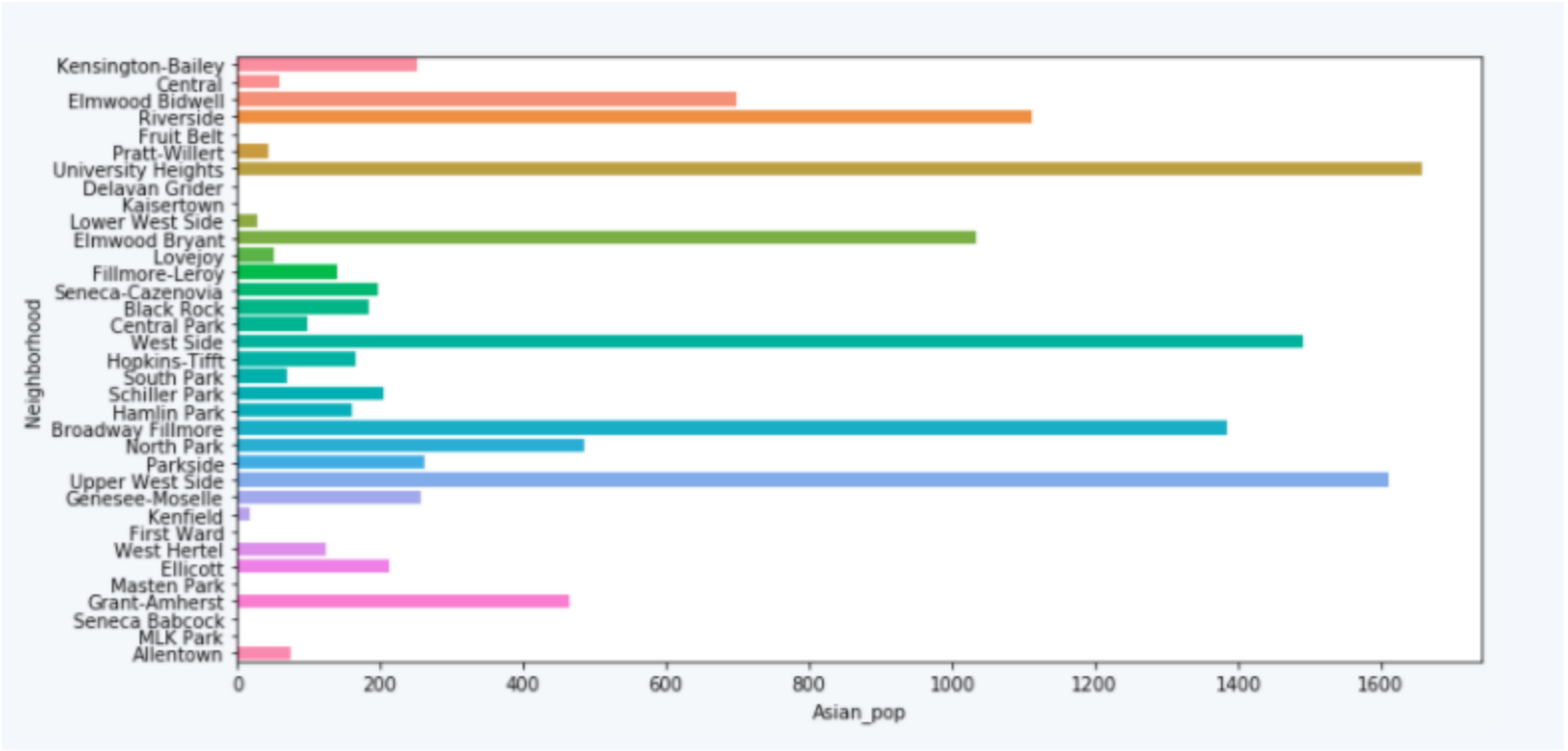
Cluster 1


Kensington-Bailey
Central
Elmwood Bidwell
Riverside
Fruit Belt
Pratt-Willert
University Heights
Delavan Grider
Kaisertown
Lower West Side
Elmwood Bryant
Lovejoy
Fillmore-Leroy
Seneca-Cazenovia
Black Rock
Central Park
West Side
Hopkins-Tifft
South Park
Schiller Park
Hamlin Park
Broadway Fillmore
North Park
Parkside
Upper West Side
Genesee-Moselle
Kenfield
First Ward
West Hertel
Ellicott
Masten Park
Grant-Amherst
Seneca Babcock
MLK Park
Allentown
Neighborhood
0
200
400
600
800
1000
1200
1400
1600
Asian_pop

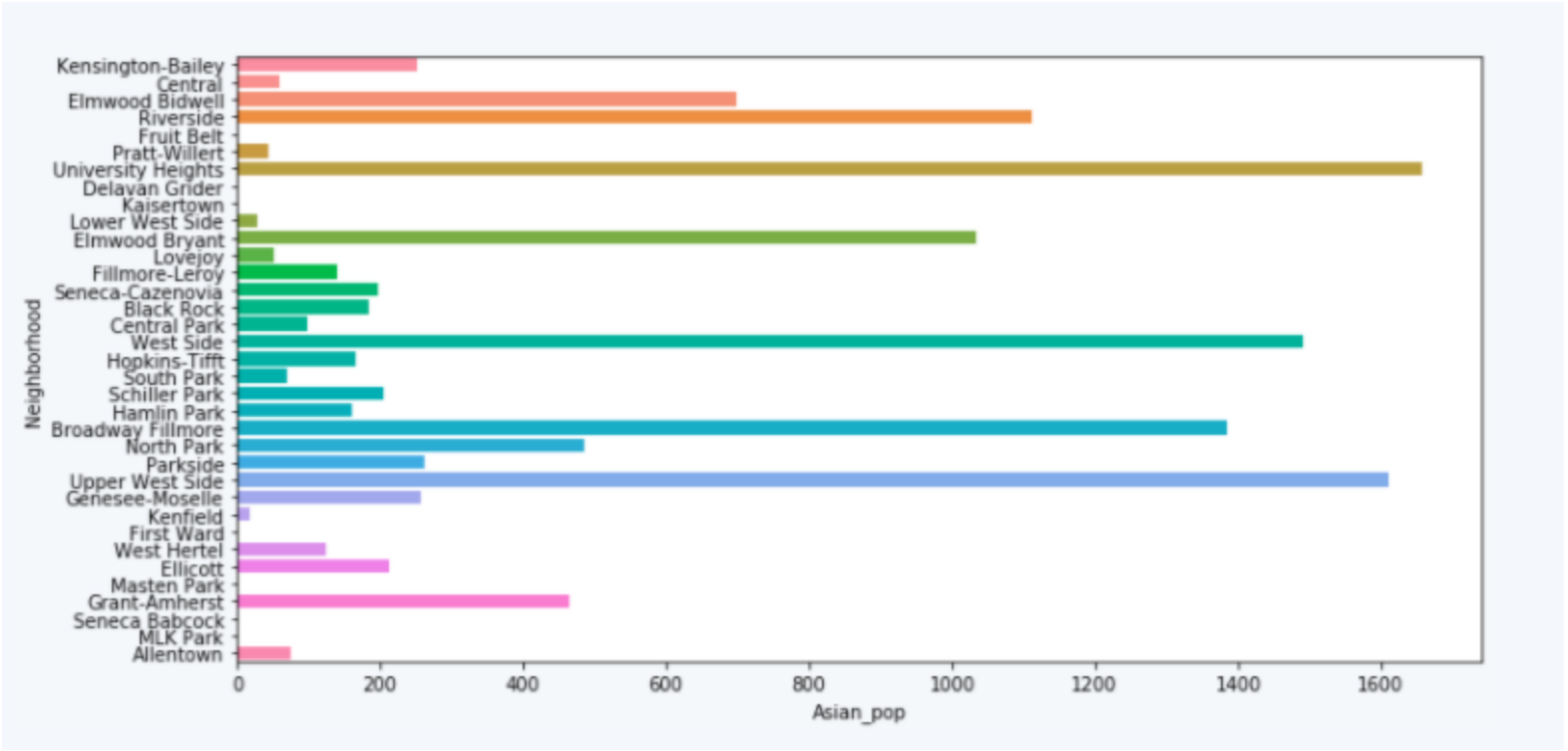
Cluster 2


Kensington-Bailey
Central
Elmwood Bidwell
Riverside
Fruit Belt
Pratt-Willert
University Heights
Delavan Grider
Kaisertown
Lower West Side
Elmwood Bryant
Lovejoy
Fillmore-Leroy
Seneca-Cazenovia
Black Rock
Central Park
West Side
Hopkins-Tifft
South Park
Schiller Park
Hamlin Park
Broadway Fillmore
North Park
Parkside
Upper West Side
Genesee-Moselle
Kenfield
First Ward
West Hertel
Ellicott
Masten Park
Grant-Amherst
Seneca Babcock
MLK Park
Allentown
Neighborhood
0
200
400
600
800
1000
1200
1400
1600
Asian_pop

Cluster 3


Kensington-Bailey
Central
Elmwood Bidwell
Riverside
Fruit Belt
Pratt-Willert
University Heights
Delavan Grider
Kaisertown
Lower West Side
Elmwood Bryant
Lovejoy
Fillmore-Leroy
Seneca-Cazenovia
Black Rock
Central Park
West Side
Hopkins-Tifft
South Park
Schiller Park
Hamlin Park
Broadway Fillmore
North Park
Parkside
Upper West Side
Genesee-Moselle
Kenfield
First Ward
West Hertel
Ellicott
Masten Park
Grant-Amherst
Seneca Babcock
MLK Park
Allentown
Neighborhood
0
200
400
600
800
1000
1200
1400
1600
Asian_pop

Cluster 4


Kensington-Bailey
Central
Elmwood Bidwell
Riverside
Fruit Belt
Pratt-Willert
University Heights
Delavan Grider
Kaisertown
Lower West Side
Elmwood Bryant
Lovejoy
Fillmore-Leroy
Seneca-Cazenovia
Black Rock
Central Park
West Side
Hopkins-Tifft
South Park
Schiller Park
Hamlin Park
Broadway Fillmore
North Park
Parkside
Upper West Side
Genesee-Moselle
Kenfield
First Ward
West Hertel
Ellicott
Masten Park
Grant-Amherst
Seneca Babcock
MLK Park
Allentown
Neighborhood
0
200
400
600
800
1000
1200
1400
1600
Asian_pop

Discussion

On analyzing the data, we could see that there isn't one common neighborhood which has higher density of asian population along with high income people and newer homes.

If we take the criteria of asian population, we can find one neighborhood "Central" with newer homes and high income folks. Central falls under Cluster 1 and it's top venues are Hockey Arena, Restaurants, bars and museum. This will be a perfect neighborhood to move into.

This analysis was limited by only considering couple of variables this can be extended to consider other variables and zero down on a neighborhood that suits one's preferences

Conclusion

The goal of this project is to find a neighborhood which closely matches with our interests. In doing that we were able to find one neighborhood that match our preferences also we could find other qualities of these neighborhoods. There are some neighborhoods which are Restaurant oriented and others are Coffee shop dominants. Here are the few from our analysis below. This analysis can be extended to use more variables to identify the results which closely matches and caters our interests.

Coffee Neighborhoods:

1. Park Side
2. Central Park
3. South Park

Restaurant Neighborhoods:

1. West Side
2. Black Rock
3. Elmwood Bidell

Night life Neighborhoods:

1. Central
2. Fillmore-Leroy
3. Hamlin-Park

References

[1] Buffalo Neighborhoods (https://data.buffalony.gov/stories/s/a235-4wxj)

[2] Foursquare API (https://developer.foursquare.com/)