

NEIGHBORHOODS OF BALTIMORE: PEOPLE, CRIME, AND VENUES

A Data Science Exploration by CJRR



PROBLEM STATEMENT

Jane New Resident is moving to Baltimore. She is:

- A 38 yr old female professional
- Concerned with the crime rate in Baltimore; far above the national average
- Overwhelmed by the number of neighborhoods to consider, >50

To start her house hunt she wants help to:

- Identify potential neighborhoods that are:
 - Low in crime
 - Has shops and restaurants
 - Filled with professionals
- Understand the character of those neighborhoods



ANALYTIC APPROACH

Collect data on each neighborhood:

- Crime rates
- Demographics data:
- Number and types of venues

Cluster the neighborhoods into a manageable number of groups based on those features

Recommend neighborhoods in clusters that align with Jane's criteria

Summarize the character of each neighborhood

DATA SOURCES

Crime and Safety:

- 2015 Crime and Safety data for Baltimore City is available on Data.gov here:
- <https://catalog.data.gov/dataset/crime-and-safety-2015>

Demographics:

- 2015 Census demographics data for Baltimore City is available on Data.gov here:
- <https://catalog.data.gov/dataset/census-demographics-2015>

Venues:

- Foursquare will be used to generate a list of venues near each neighborhood's center.

DATA FEATURES

Crime and Safety:

- Part I Crime Rate per 1,000 Residents
- Violent Crime Rate per 1,000 Residents
- Property Crime Rate per 1,000 Residents

Demographics:

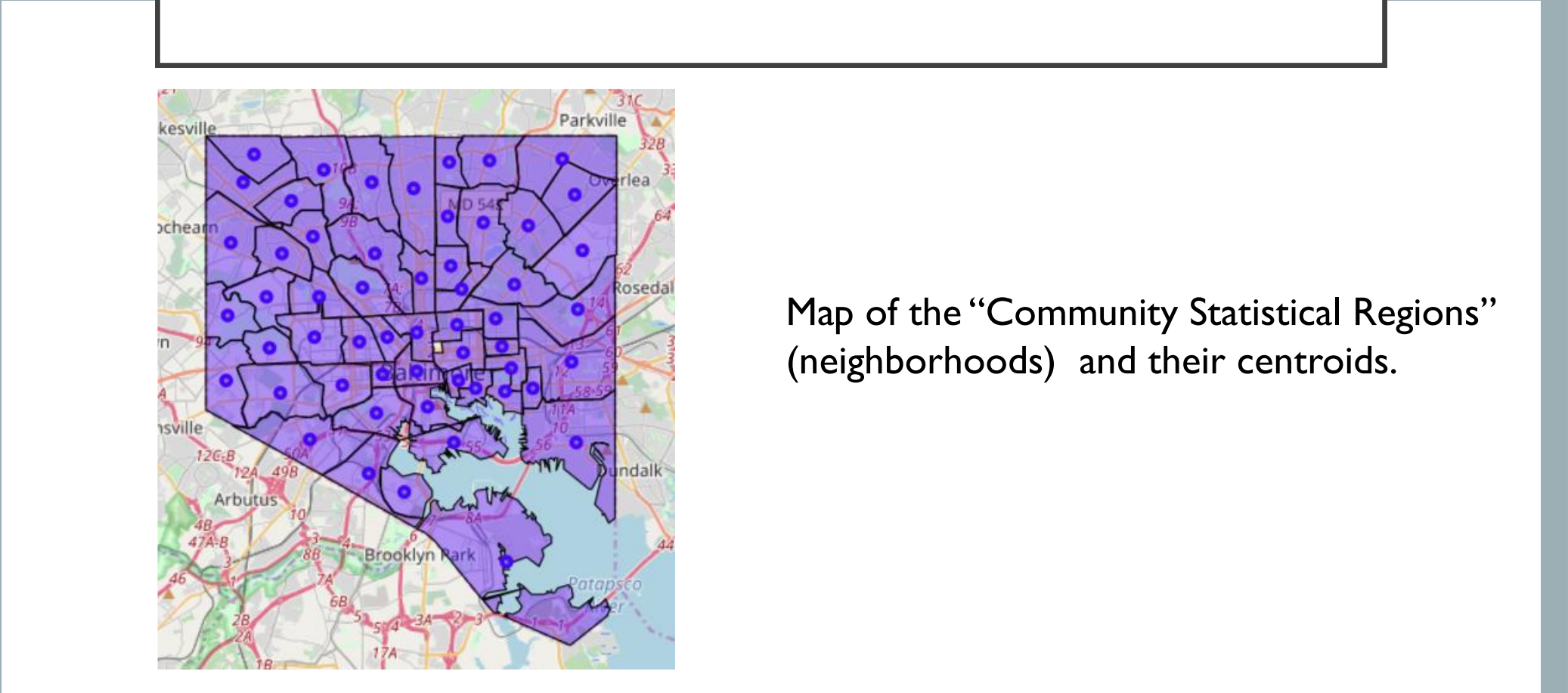
- Percent of Population 25-64 years old
- Average Household Size
- Median Household Income
- Percent of Family Households Living Below the Poverty Line
- Total Number of Households
 - to calculate population density

Venues:

- Number of each type of venues in each neighborhood

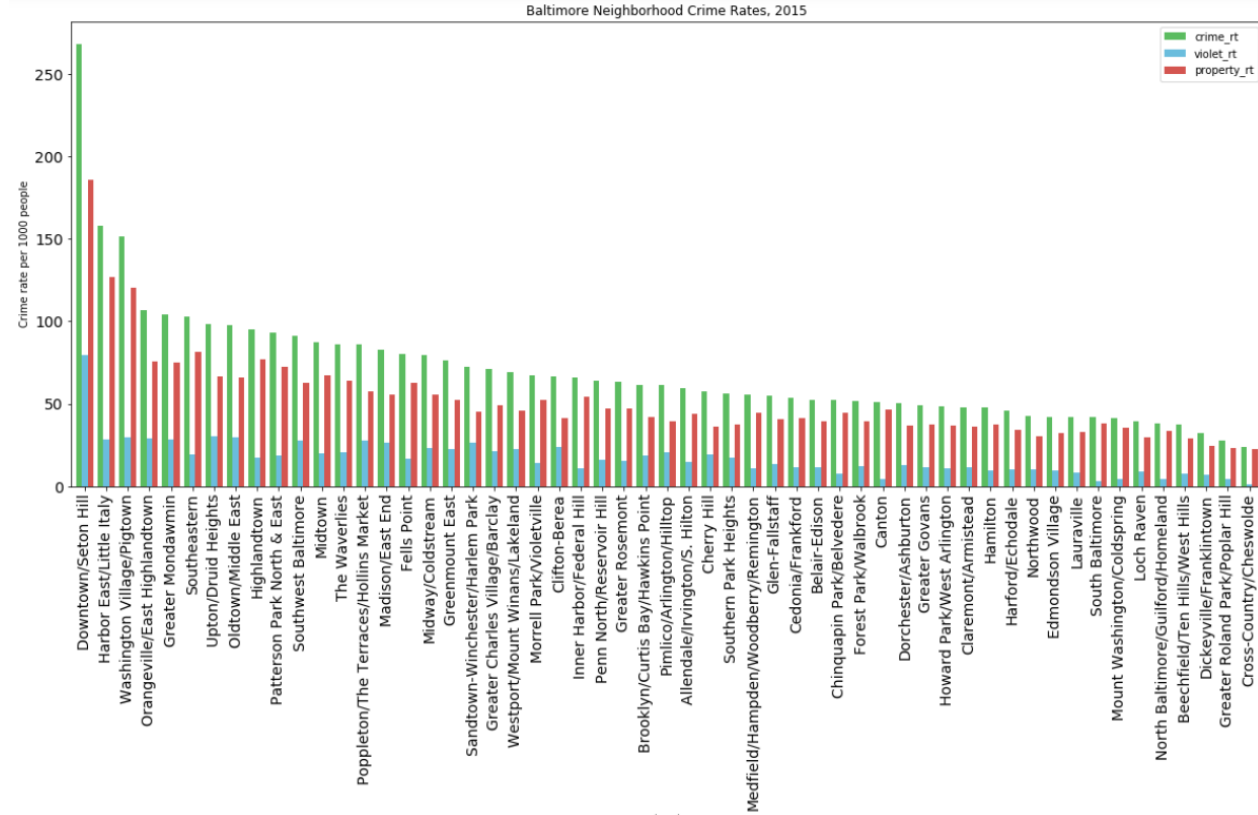
DATA EXPLORATION

DATA EXPLORATION: MAP

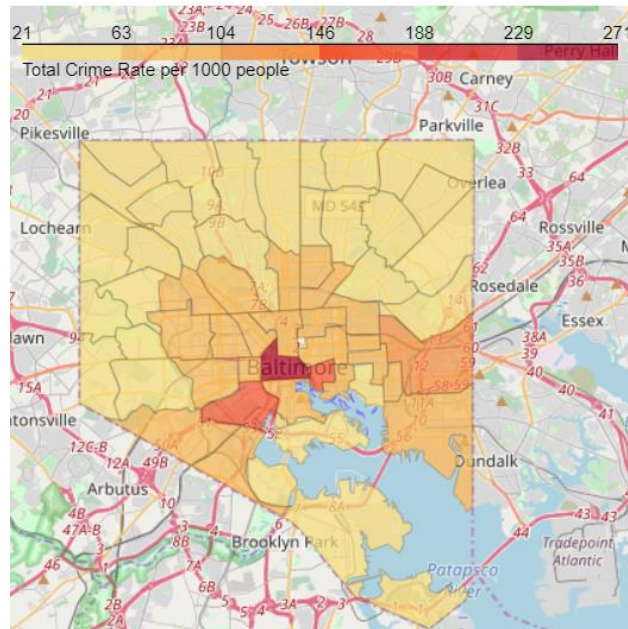


Map of the “Community Statistical Regions”
(neighborhoods) and their centroids.

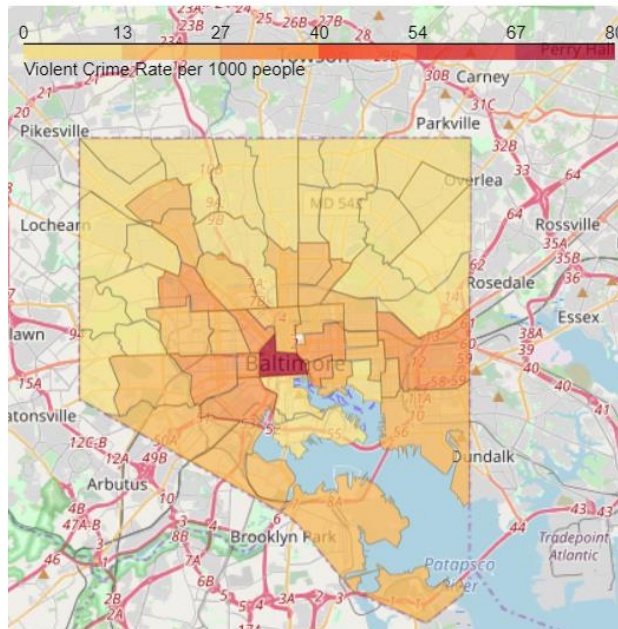
DATA EXPLORATION: CRIME



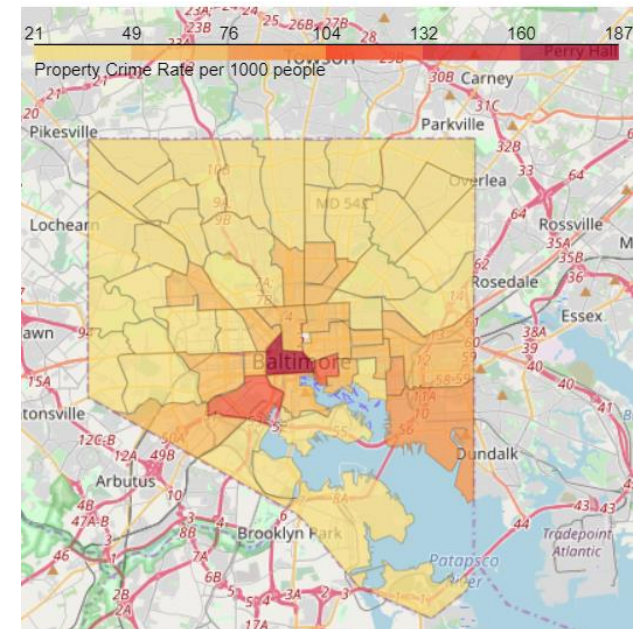
DATA EXPLORATION: CRIME MAPS



Total Crime

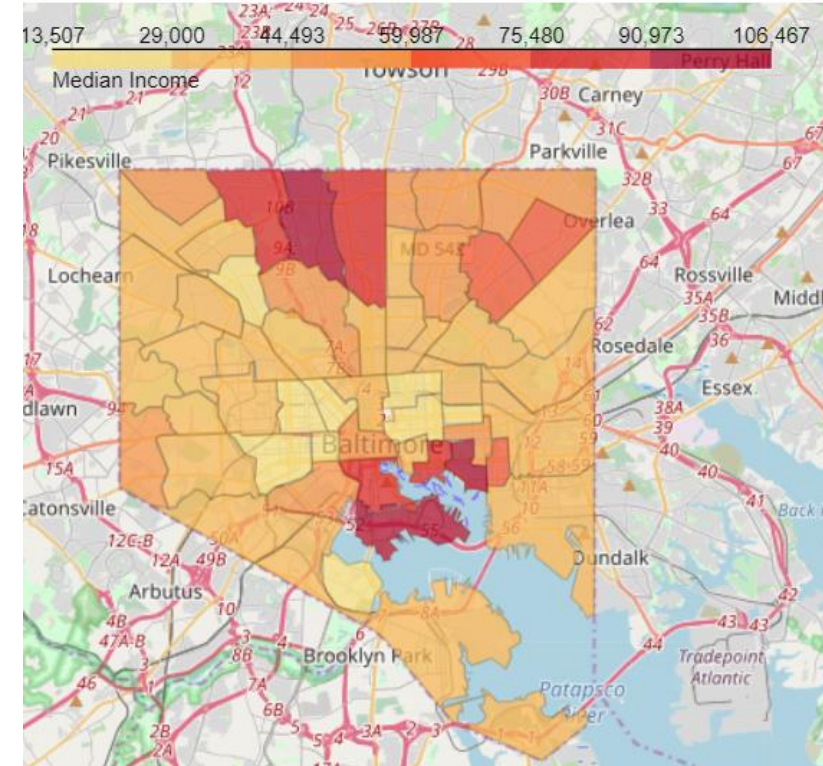
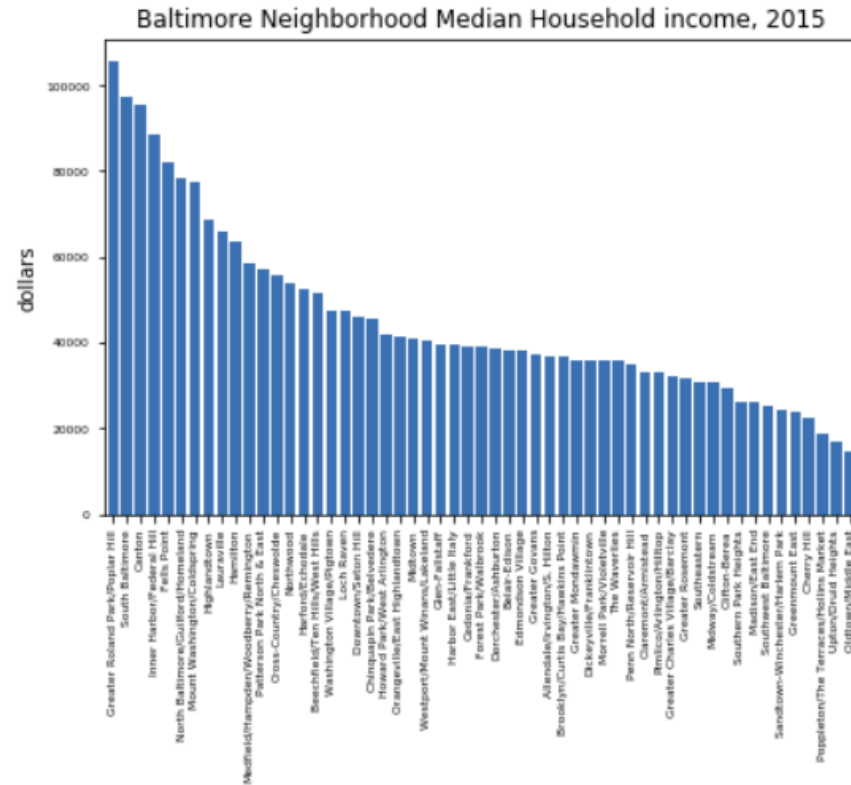


Violent Crime



Property Crime

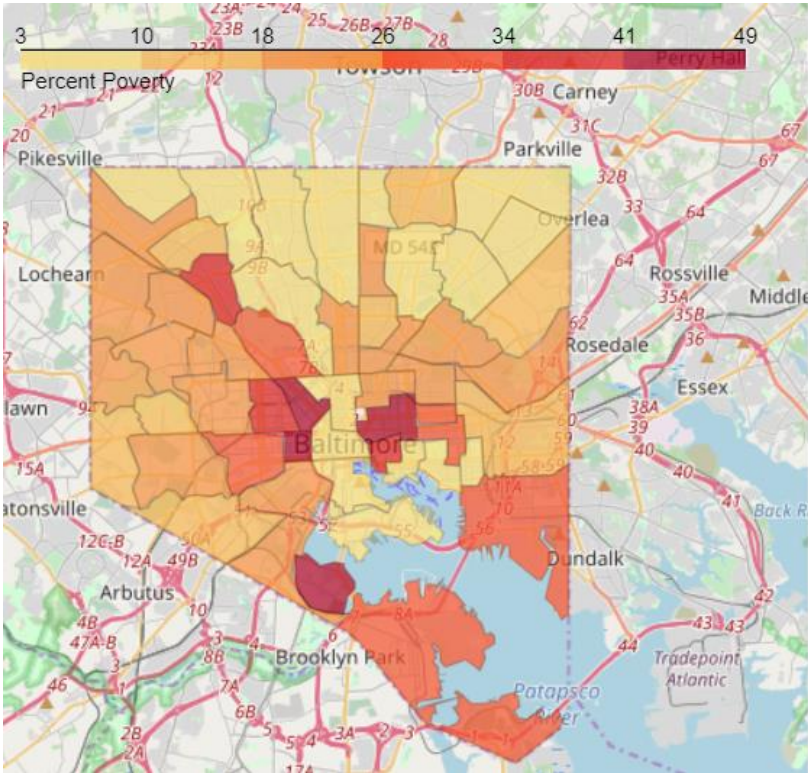
DATA EXPLORATION: POPULATION



MEDIAN INCOME

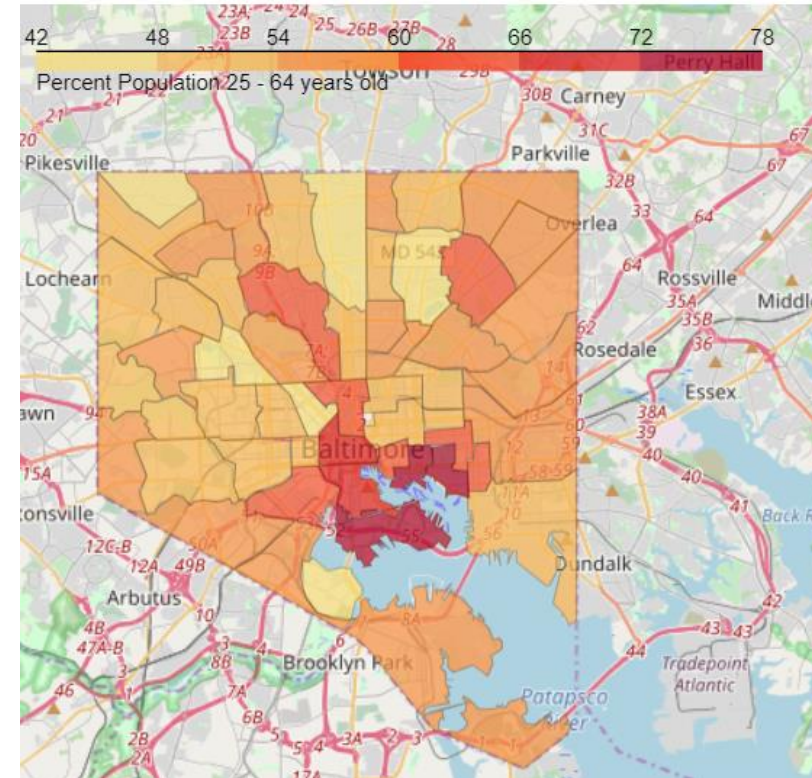
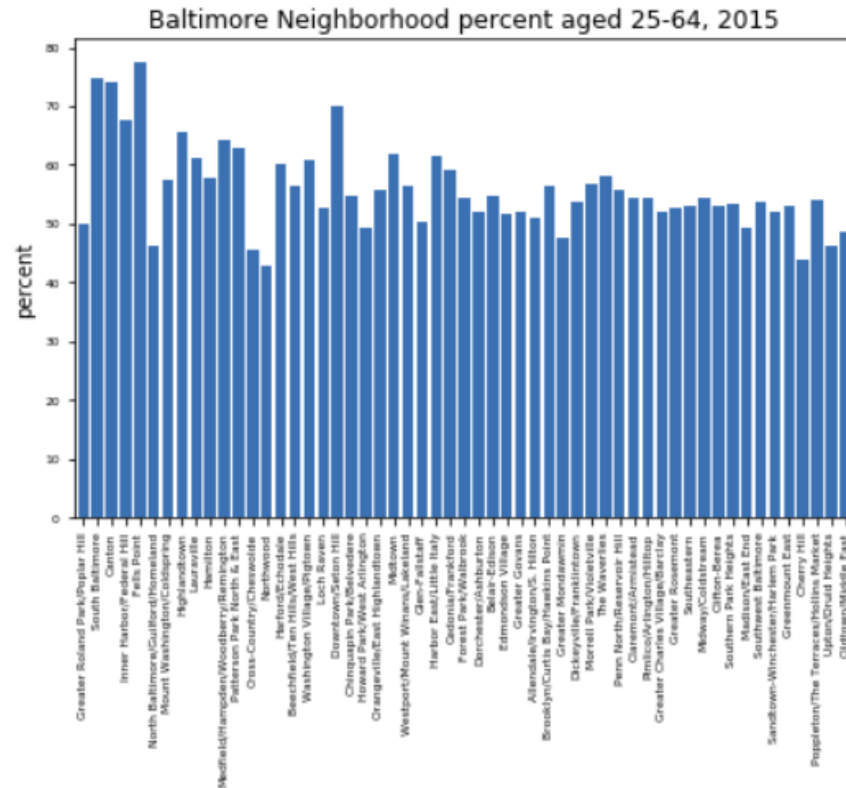
DATA EXPLORATION: POPULATION

Neighborhood	Percent
Greater Roland Park/Venue Hill	10
South Bay	10
Capitol Hill	10
Inner Harbor/Federal Hill	15
Fells Point	15
North Baltimore/Guilford/Homeland	15
Mount Washington/Columbia Heights	15
Highland Park	20
Leawood	20
Hamilton	18
Medfield/Hampden/Woodberry/Barnington	25
Patterson Park North & East	18
Cross Country/Chesapeake	18
Herndon/Leitchwood	18
Beechfield/Ten Hills/West Hills	22
Washington Village/Pigtown	25
Loch Raven	18
Overlook/Seaton Hill	15
Chinatown/Cherry Hill	18
Howard Park/West Arlington	20
Orangeville/East Highlandtown	18
Midtown	25
Westport/Mount Whanna/Lakeview	20
Harbor East/Georgetown Navy Yard	38
Georgetown/Federal Hill	18
Forest Park/Walbrook	20
Dorchester/Ashburton	18
Belair-Edition	18
Edmondson Village	15
Greenbelt	22
Altendale/Springtown/5. Hill	25
Brooklyn/Curtis Bay/Saybrook Point	28
Greater Mondakemin	18
Dickeyville/Franklinview	22
Morrell Park/Violetville	15
Capitol Hill/Walbrook	15
Penn North/Barnaby Hill	20
Greenmont/Homestead	29
Brimco/Arlington/Hilltop	22
Greater Charles Village/Barkley	25
Greater Rossmore	22
Southeastern	25
Northwest	25
Clifton/Berres	24
Southern Park Heights	36
Madison/East End	33
Southeast Baltimore	33
Sandtown-Windward Hill	36
Greenmount Park	25
Cherry Hill	46
Puppleton/The Terrace/Holmes Market	48
Upton/Cruik	46
Oldtown/Middle East	44



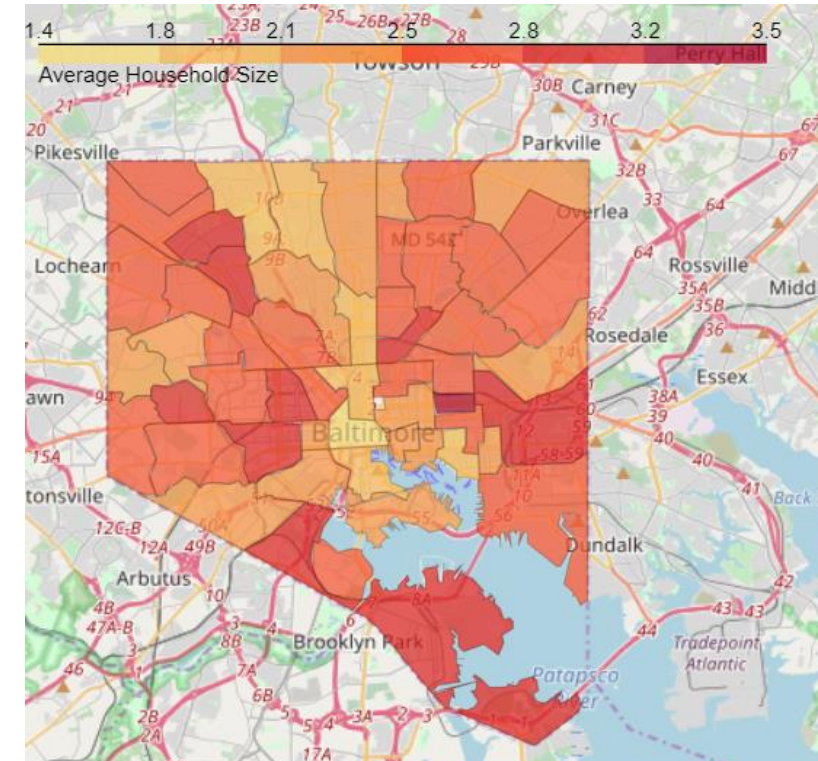
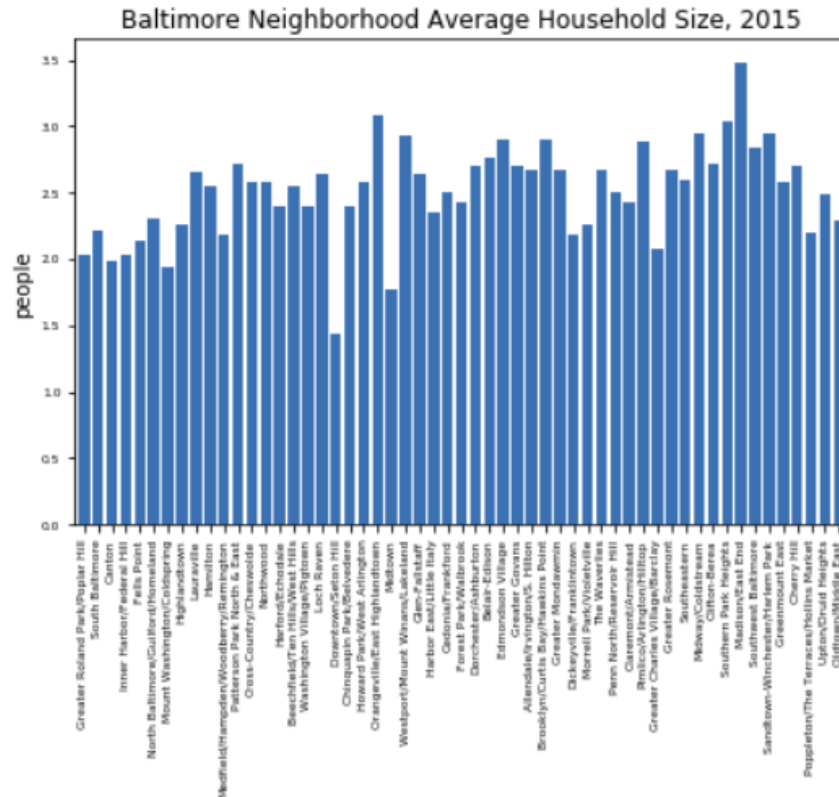
POVERTY

DATA EXPLORATION: POPULATION



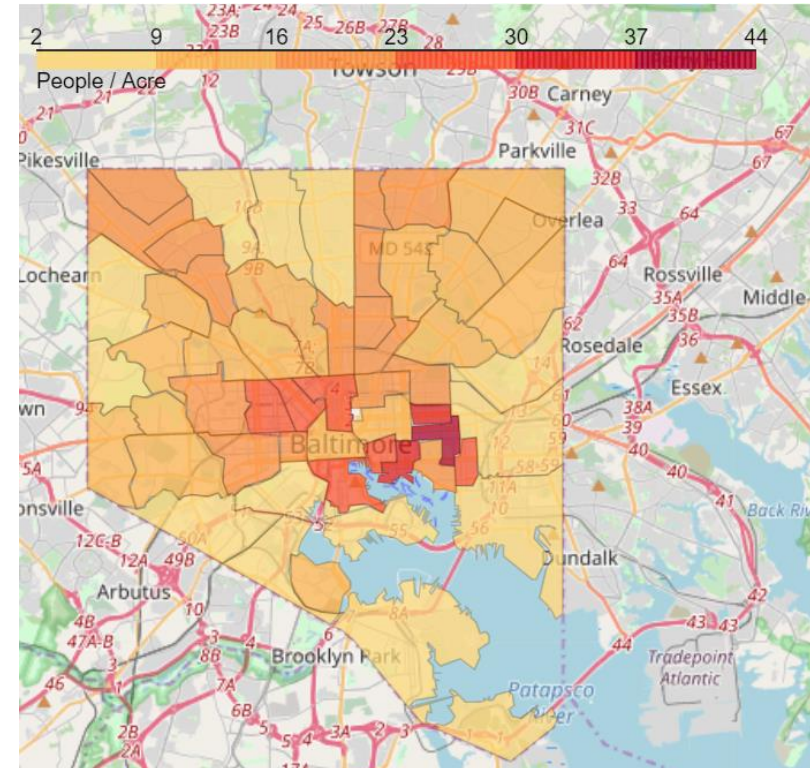
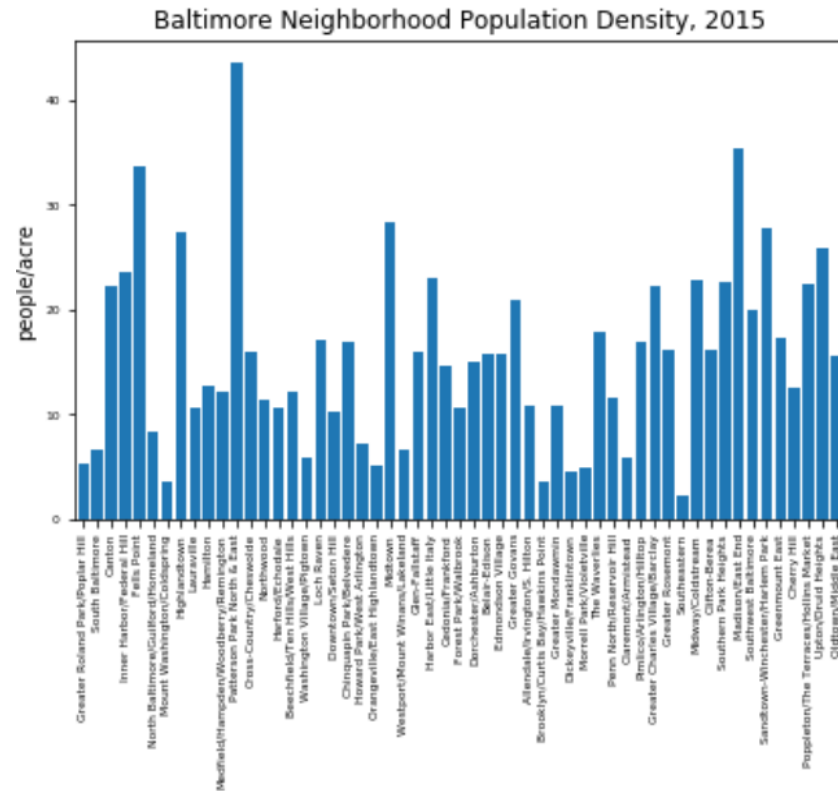
AGED 25-64

DATA EXPLORATION: POPULATION



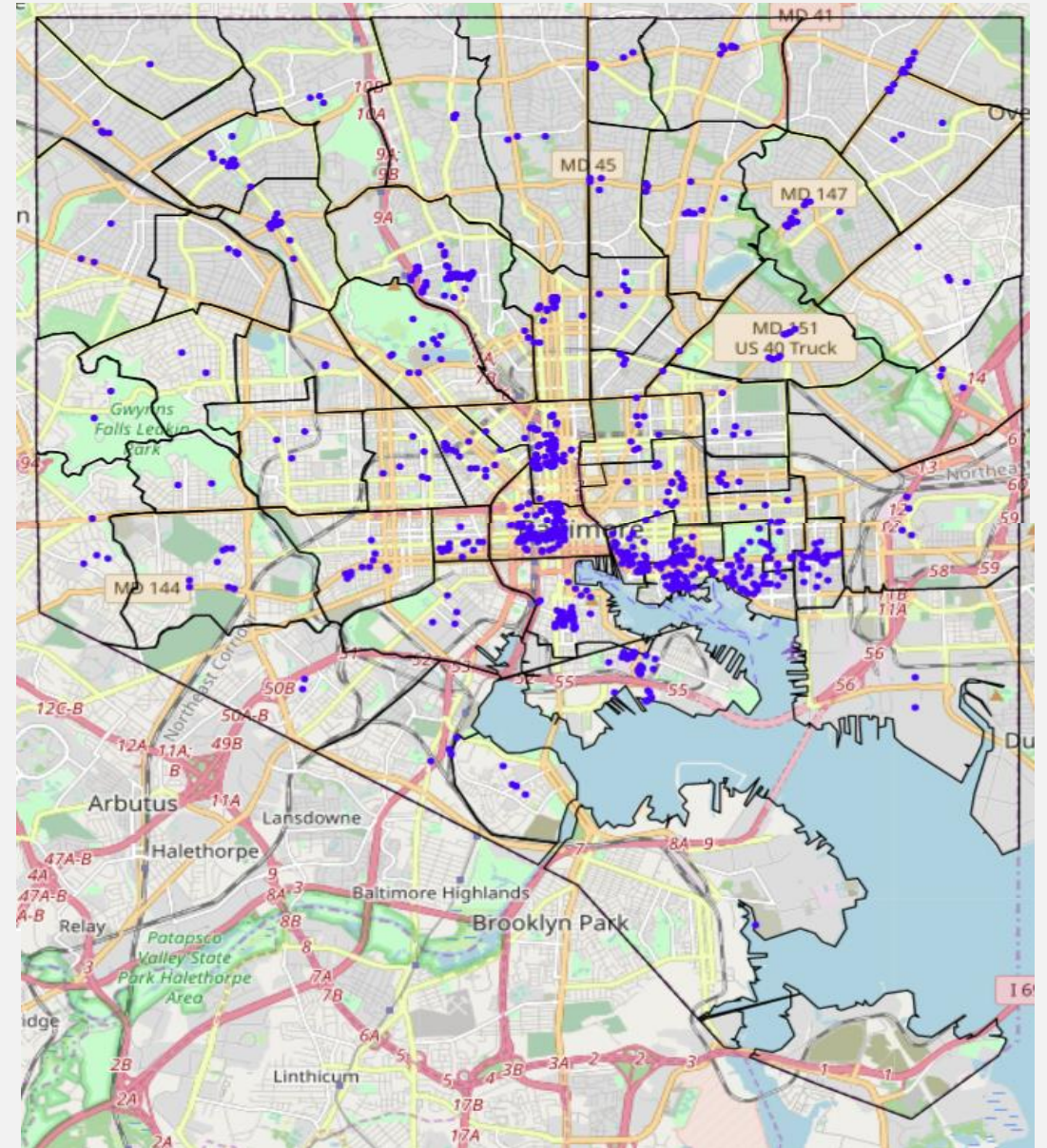
HOUSEHOLD SIZE

DATA EXPLORATION: POPULATION



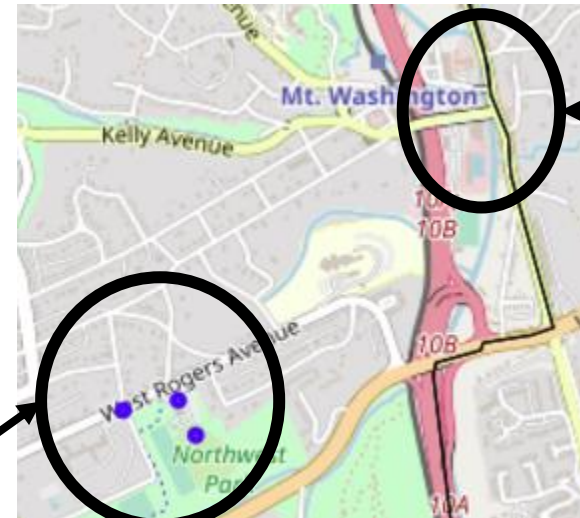
POPULATION DENSITY

889 Venues in 216 Categories from Foursquare



PROBLEM

- Foursquare data is rather sparse for some neighborhoods. It is clearly missing many relevant venues, especially in the more suburban areas.
- Increasing search radius did not resolve this
- Example:



Cute village filled with shops, restaurants and art galleries

Only Foursquares returns

Problem Resolution: NONE – Foursquares use is a requirement

ANALYSIS

CATEGORIZING FEATURE VALUES

Convert numeric values to categories based on statistical distribution:

- Very low ← lowest 10%
- Low ← lowest 10-30%
- Moderate ← lowest 30-70%
- High ← lowest 70-90 %
- Very High ← lowest 90 – 100 % (aka highest 10%)

Except average household size:

- Mostly Singles ← <1.5
- Mostly Couples ← 1.5 – 2.5
- Couples & Small Families ← 2.5 -3
- Families ← >3

EXAMPLE:

	hood	25-64_pct_cat	household_ave_cat	income_median_cat	pct_poverty_cat	density_cat	crime_rt_cat	violet_rt_cat	property_rt_cat
0	Allendale/Irvington/S. Hilton	low	couples & small families	moderate	high	moderate	moderate	moderate	moderate
1	Beechfield/Ten Hills/West Hills	moderate	couples & small families	high	moderate	moderate	very low	low	very low
2	Belair-Edison	moderate	couples & small families	moderate	moderate	moderate	moderate	moderate	moderate
3	Brooklyn/Curtis Bay/Hawkins Point	moderate	couples & small families	moderate	high	very low	moderate	moderate	moderate
4	Canton	very high	mostly couples	very high	very low	high	moderate	very low	moderate
5	Cedonia/Frankford	high	couples & small families	moderate	moderate	moderate	moderate	moderate	moderate

CATEGORIZING FEATURE VALUES

USEFUL TO DESCRIBE NEIGHBORHOODS CHARACTER



MOST COMMON VENUES

Tabulate the most common venues for
each neighborhood

EXAMPLE:

	Neighborhood	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
0	Allendale/Irvington/S. Hilton	Fried Chicken Joint	Karaoke Bar	Discount Store	Cosmetics Shop	Bus Stop	Flower Shop	Thrift / Vintage Store
1	Beechfield/Ten Hills/West Hills	ATM	Residential Building (Apartment / Condo)	Pool	Restaurant	Discount Store	Food Service	Food Court
2	Belair-Edison	Fried Chicken Joint	Discount Store	ATM	Chinese Restaurant	Grocery Store	Food & Drink Shop	Pharmacy
3	Brooklyn/Curtis Bay/Hawkins Point	Food Truck	Yoga Studio	Farmers Market	French Restaurant	Food Service	Food Court	Food & Drink Shop
4	Canton	Bar	American Restaurant	Gym / Fitness Center	Pub	Pizza Place	Mexican Restaurant	Ice Cream Shop
5	Cedonia/Frankford	Garden	Indian Restaurant	Chinese Restaurant	Yoga Studio	Fast Food Restaurant	Food Truck	Food Service

MOST COMMON VENUES

USEFUL TO DESCRIBE NEIGHBORHOODS CHARACTER



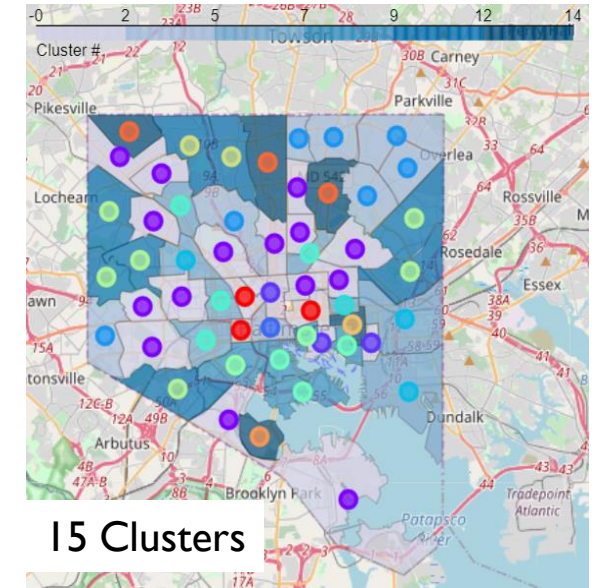
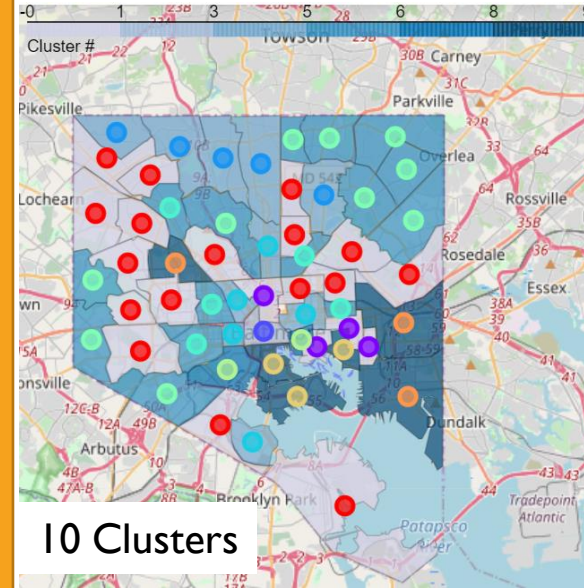
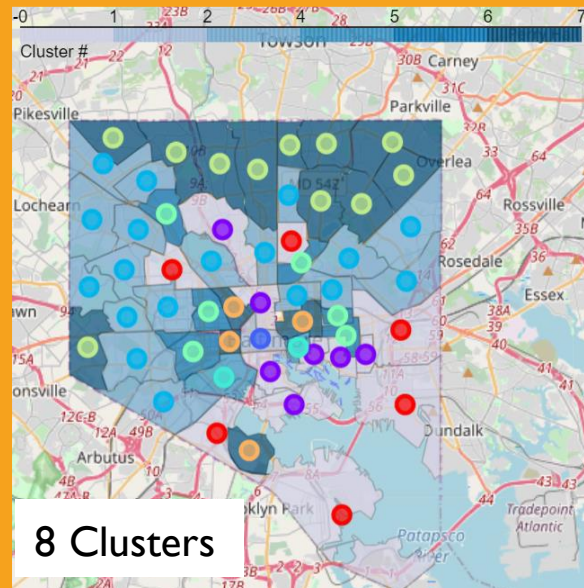
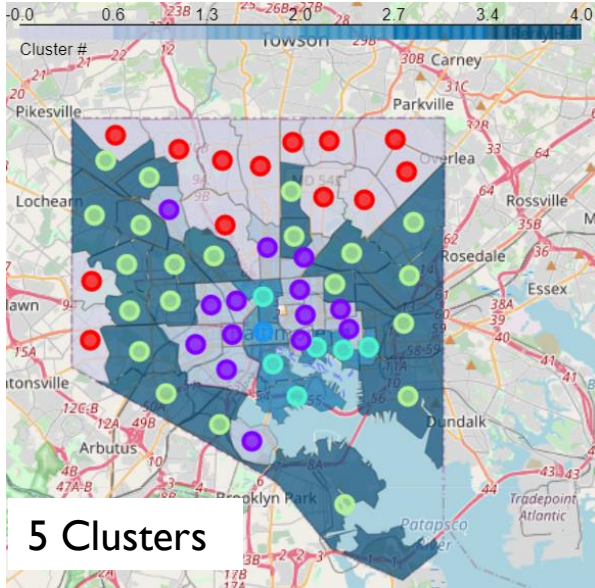
CLUSTER NEIGHBORHOODS

Use K-Means to cluster the neighborhoods based on:

- Crime and Census numeric features
- Crime, Census, and Venue numeric features

Normalize data first and try different numbers of clusters.

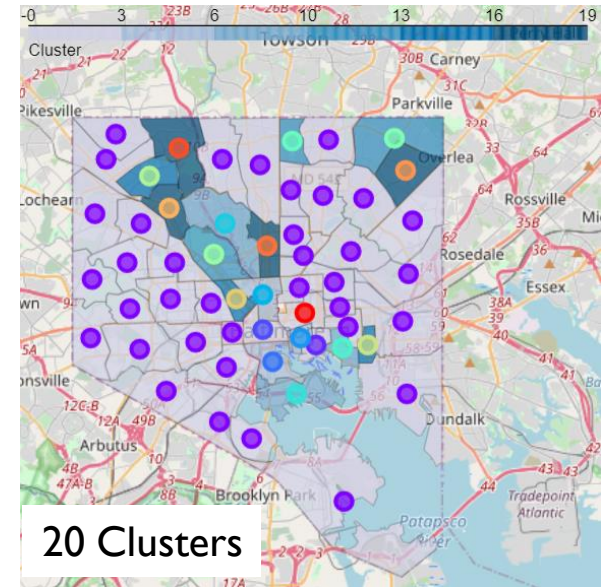
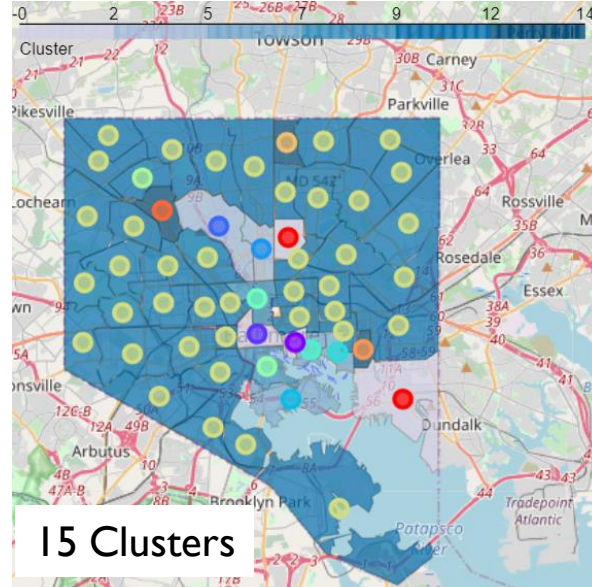
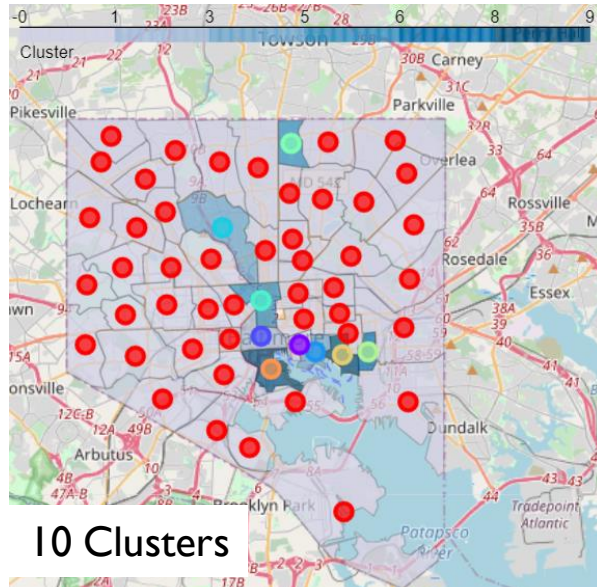
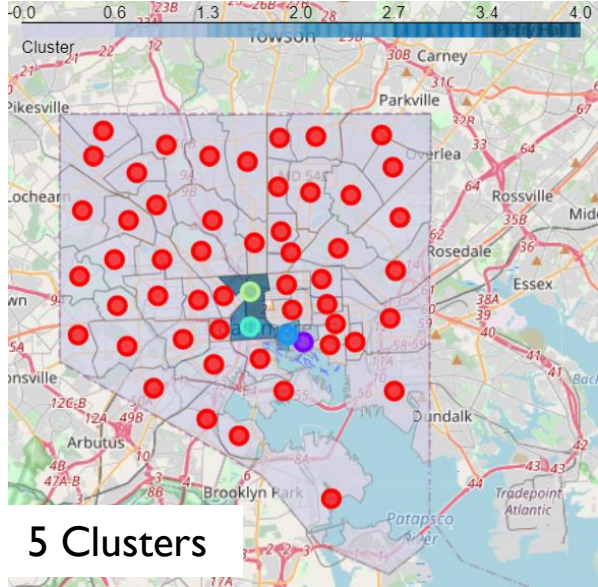
Looks Good!



CLUSTER NEIGHBORHOODS

CRIME AND CENSUS DATA

When including the venue data the results are not useful



CLUSTER NEIGHBORHOODS

CRIME, CENSUS, AND VENUE DATA



CHARACTERIZE THE CLUSTERS

Using the results of K-Means clustering on crime and census data and 8 clusters.

- Find the mean value for each feature
- Categorize the values
- Find the 5 most common venues
- Generate report on each cluster

	25-64_pct	household_ave	income_median	pct_poverty	density	crime_rt	violet_rt	property_rt	ATM	Adult Boutique	...	Waterfall
cluster												
0	54.577119	2.812500	36864.873027	20.662509	7.711712	88.348843	23.099700	64.065607	0.000000	0.000000	...	0.000000
1	69.367368	2.083786	76006.953610	6.380418	22.061153	68.199883	11.886277	55.742072	0.000000	0.142857	...	0.000000
2	69.829190	1.436667	46132.866910	5.454545	10.312222	268.228359	79.429103	185.696556	0.000000	0.000000	...	0.000000
3	53.329947	2.566620	35484.709813	20.650568	13.452264	56.218001	14.502281	40.950887	0.055556	0.000000	...	0.055556
4	61.112011	2.377500	43536.526595	31.422385	14.434481	154.656080	29.049289	123.494436	0.000000	0.000000	...	0.000000
5	54.262141	2.995458	31578.846760	31.609769	28.719827	79.285281	23.323529	54.783713	0.000000	0.000000	...	0.000000
6	53.114143	2.420712	63439.270428	8.204275	11.327016	39.771567	6.968218	32.165119	0.090909	0.000000	...	0.000000
7	48.160009	2.417458	18154.698775	46.321400	19.114445	84.790032	26.778533	56.705256	0.250000	0.000000	...	0.000000

CHARACTERIZE THE CLUSTERS

MEAN FEATURE VALUES

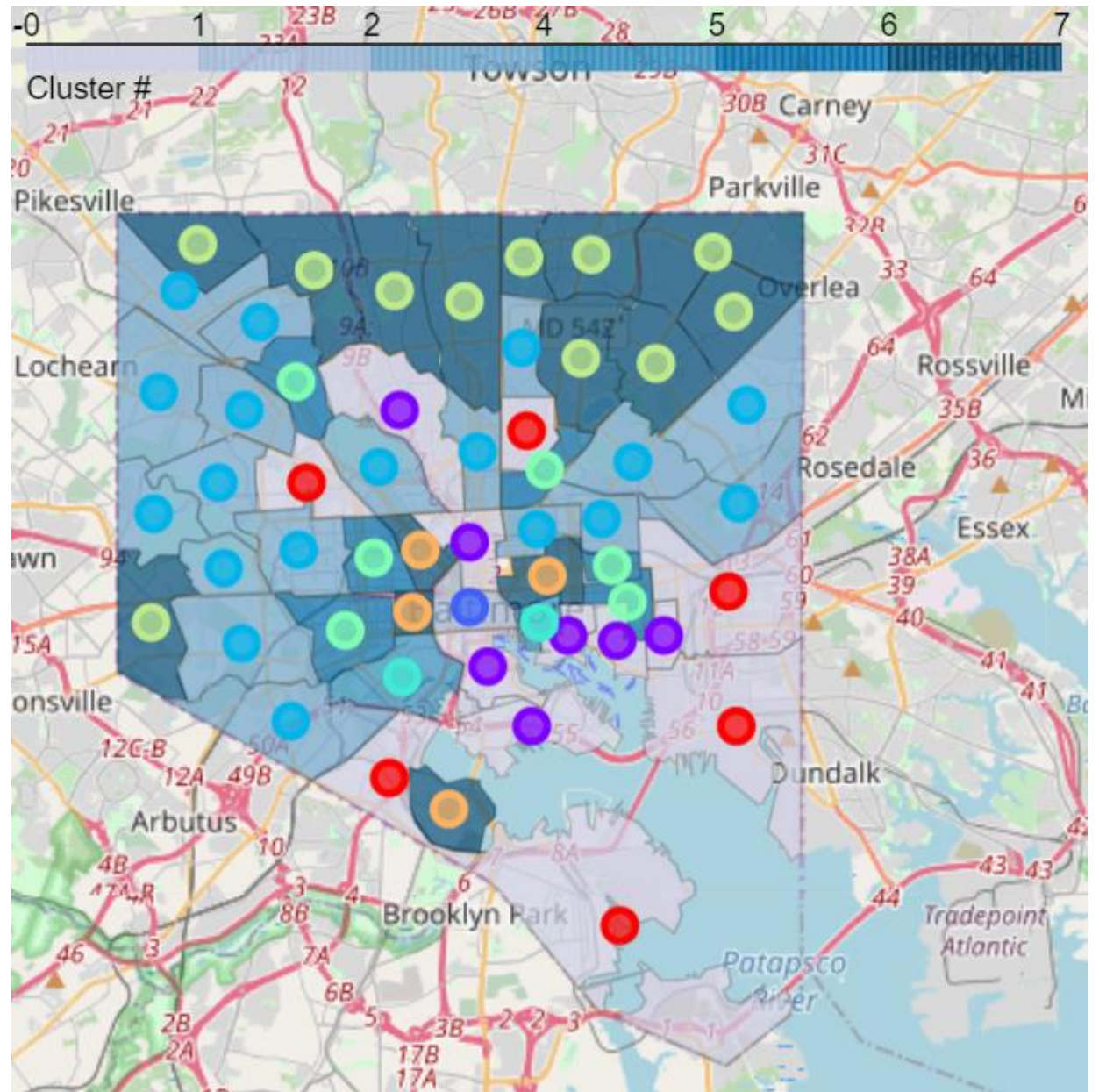
	25-64_pct_cat	household_ave_cat	income_median_cat	pct_poverty_cat	density_cat	crime_rt_cat	violet_rt_cat	property_rt_cat	Top_5_Venues
cluster									
0	moderate	couples & small families	moderate	moderate	low	high	high	high	[Business Service, Coffee Shop, Discount Store,...]
1	very high	mostly couples	high	low	high	moderate	moderate	high	[Bar, Pizza Place, Mexican Restaurant, America...]
2	very high	mostly singles	moderate	very low	low	very high	very high	very high	[Café, Coffee Shop, Sandwich Place, Hotel, Del...]
3	moderate	couples & small families	moderate	moderate	moderate	moderate	moderate	moderate	[Fried Chicken Joint, Park, Convenience Store,...]
4	high	mostly couples	moderate	high	moderate	very high	very high	very high	[Italian Restaurant, Pizza Place, Rental Car L...]
5	moderate	couples & small families	low	high	very high	high	high	moderate	[Fast Food Restaurant, Grocery Store, Deli / B...]
6	moderate	mostly couples	high	low	moderate	very low	low	low	[Pharmacy, Coffee Shop, Flower Shop, Bus Stop,...]
7	low	mostly couples	very low	very high	high	high	high	high	[Pizza Place, Grocery Store, Market, Bus Stop,...]

CHARACTERIZE THE CLUSTERS

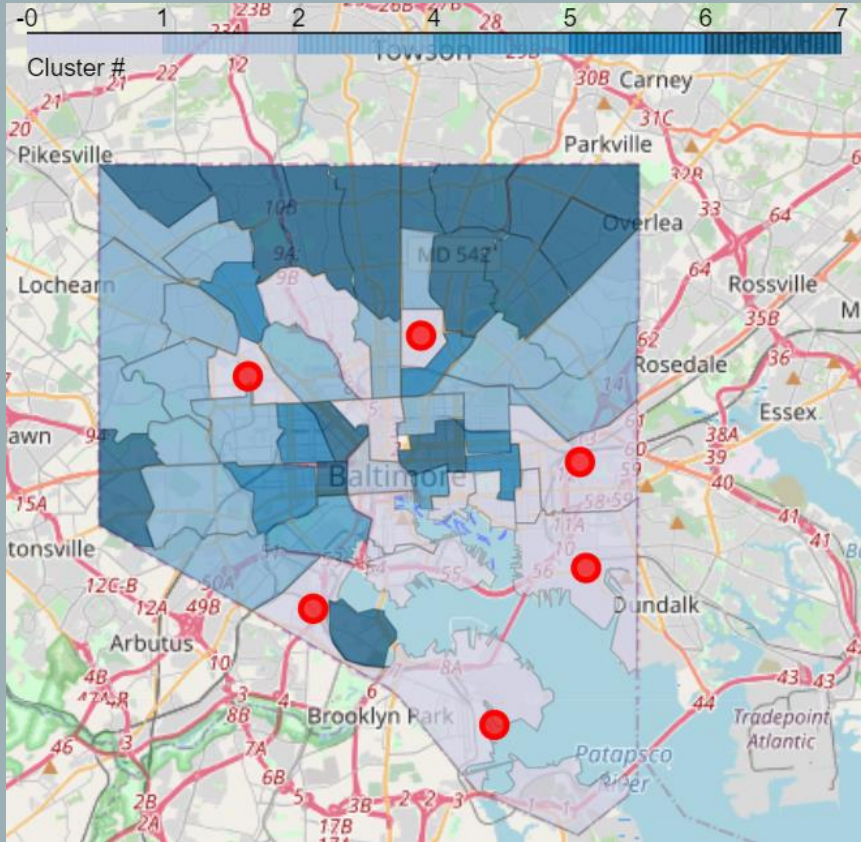
CATEGORIZED FEATURES AND TOP VENUES

THE CLUSTERS

8
CLUSTERS



CLUSTER #0



Summary: High crime, moderate income, moderate poverty, low population density

Recommend: **No.** High crime rate.

Crime:

- Crime: high, average rate is 88.3 per 1000 people
- Violent Crime: high, average rate is 23.1 per 1000 people
- Property Crime: high, average rate is 64.1 per 1000 people

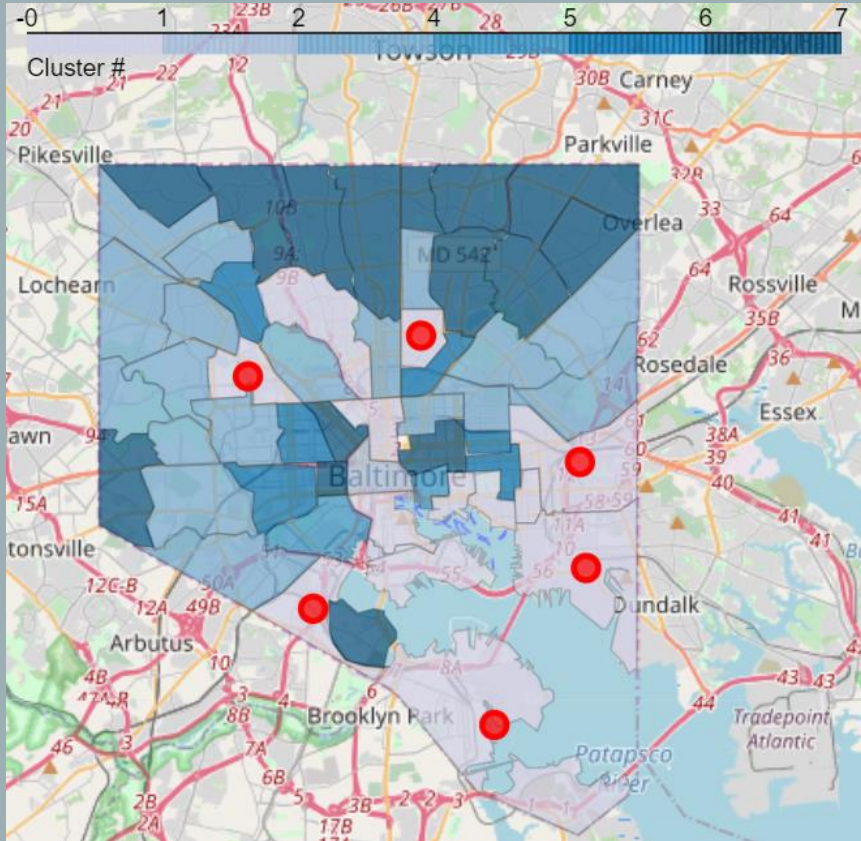
People:

- Median Income: moderate, average is 36864.9 dollars per household
- Poverty: moderate, average percentage of families in poverty is 20.7
- Population Density: low, average population density is 7.7 people per acre
- 25-64 age group: moderate, average percentage of people in the 25-64 years age group is 54.6
- Household Size: couples & small families, average average household size is 2.8 people

CLUSTER #0

Summary: High crime, moderate income, moderate poverty, low population density

Recommend: **No**. High crime rate.



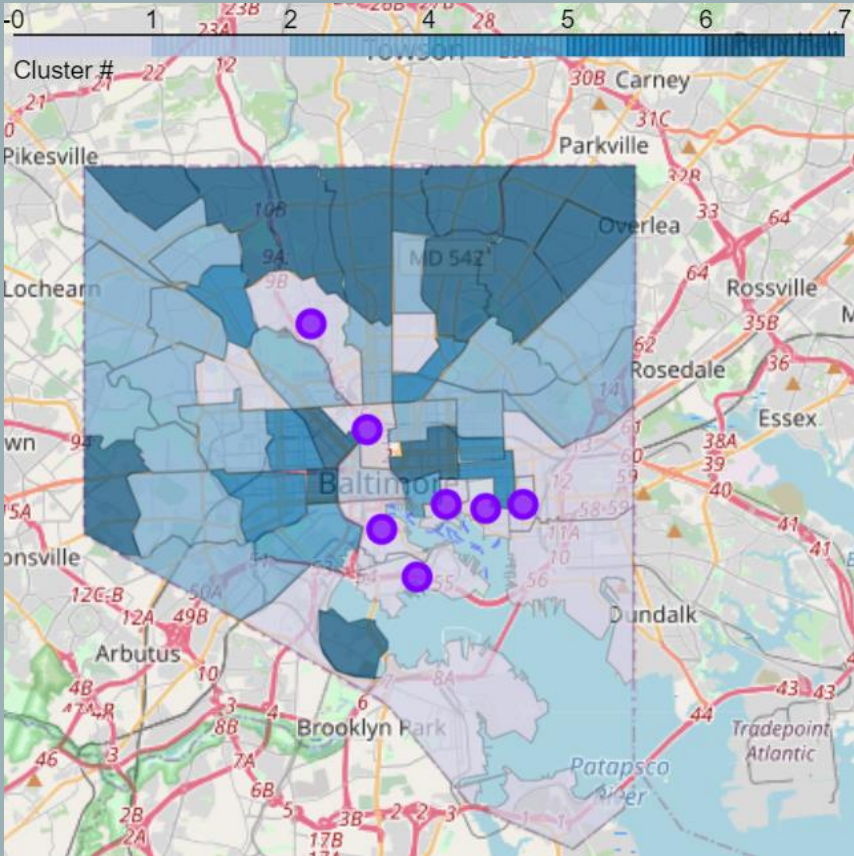
Top Venues

- Business Service
- Coffee Shop
- Discount Store
- Train Station
- Bus Stop

Neighborhoods:

- Brooklyn/Curtis Bay/Hawkins Point
- Greater Mondawmin
- Orangeville/East Highlandtown
- Southeastern
- The Waverlies
- Westport/Mount Winans/Lakeland

CLUSTER #1



Summary: Moderate crime, High income, low poverty, high population density, mostly couples, bars and restaurants

Recommend: **Yes**. Moderate crime, but trendy urban scene with professionals.

Crime:

- Crime: moderate, average rate is 68.2 per 1000 people
- Violent Crime: moderate, average rate is 11.9 per 1000 people
- Property Crime: high, average rate is 55.7 per 1000 people

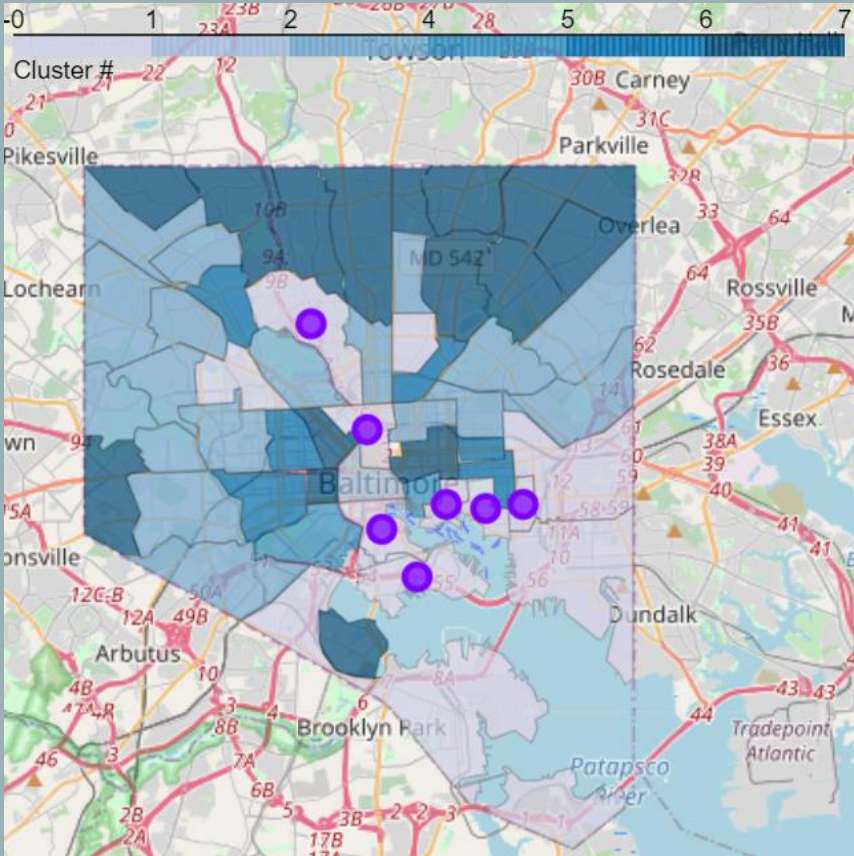
People:

- Median Income: high, average is 76007.0 dollars per household
- Poverty: low, average percentage of families in poverty is 6.4
- Population Density: high, average population density is 22.1 people per acre
- 25-64 age group: very high, average percentage of people in the 25-64 years age group is 69.4
- Household Size: mostly couples, average average household size is 2.1 people

CLUSTER #1

Summary: Moderate crime, High income, low poverty, high population density, mostly couples, bars and restaurants

Recommend: **Yes**. Moderate crime, but trendy urban scene with professionals.



Top Venues

- Bar
- Pizza Place
- Mexican Restaurant
- American Restaurant
- Coffee Shop

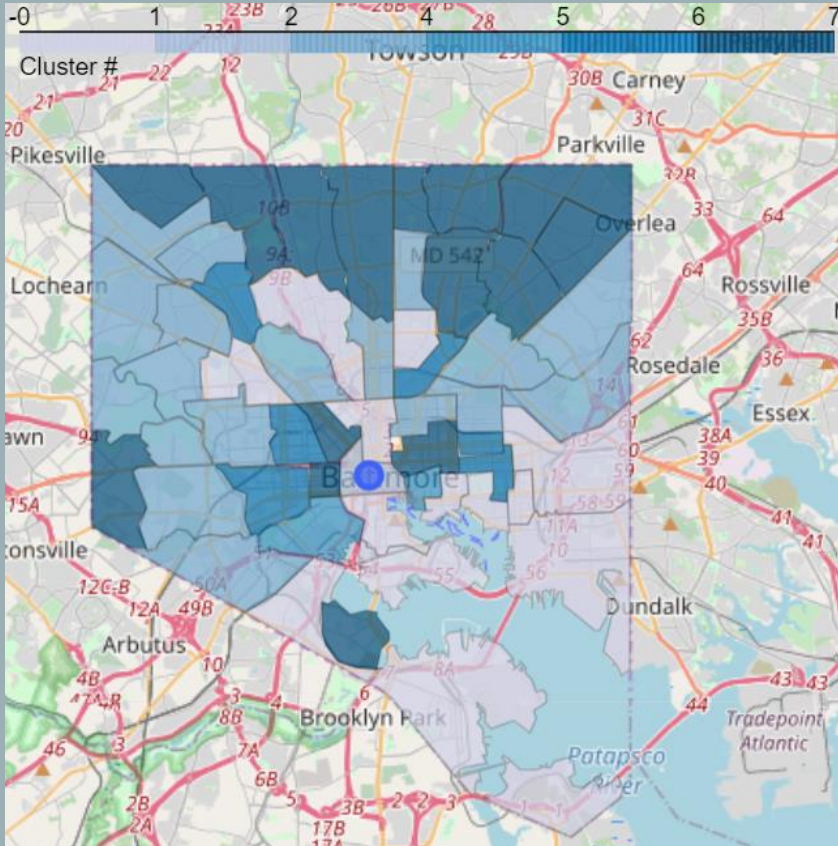
Neighborhoods:

- Canton
- Fells Point
- Highlandtown
- Inner Harbor/Federal Hill
- Medfield/Hampden/Woodberry/Remington
- South Baltimore
- Midtown

CLUSTER #2

Summary: Very high crime

Recommend: **No.** Very high crime.



Crime:

- Crime: very high, average rate is 268.2 per 1000 people
- Violent Crime: very high, average rate is 79.4 per 1000 people
- Property Crime: very high, average rate is 185.7 per 1000 people

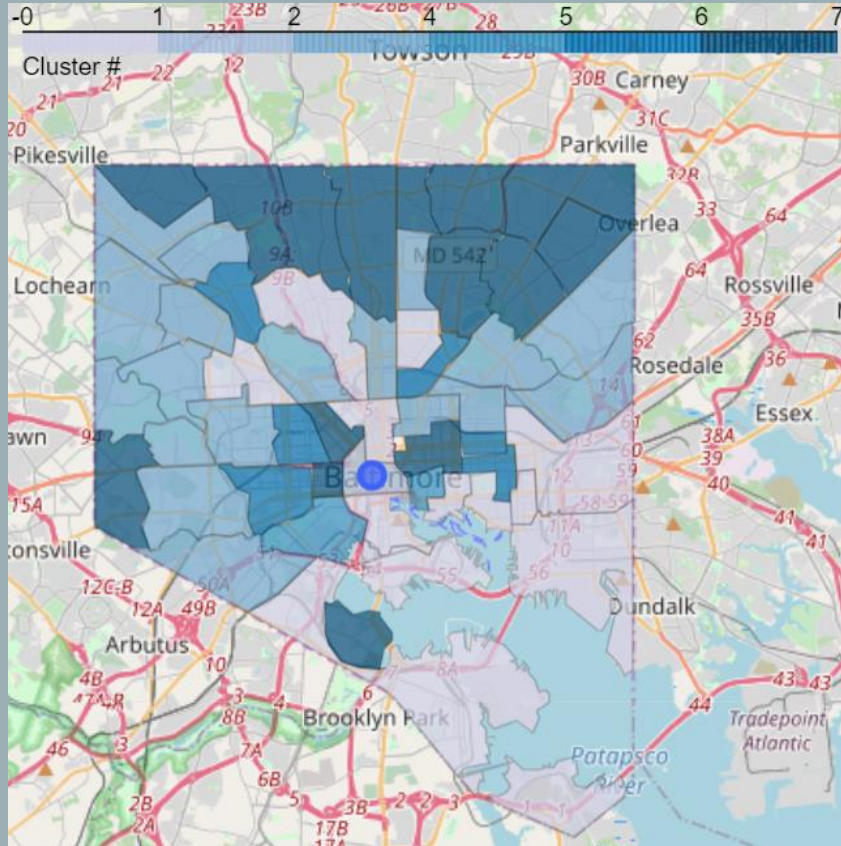
People:

- Median Income: moderate, average is 46132.9 dollars per household
- Poverty: very low, average percentage of families in poverty is 5.5
- Population Density: low, average population density is 10.3 people per acre
- 25-64 age group: very high, average percentage of people in the 25-64 years age group is 69.8
- Household Size: mostly singles, average average household size is 1.4 people

CLUSTER #2

Summary: Very high crime

Recommend: **No.** Very high crime.



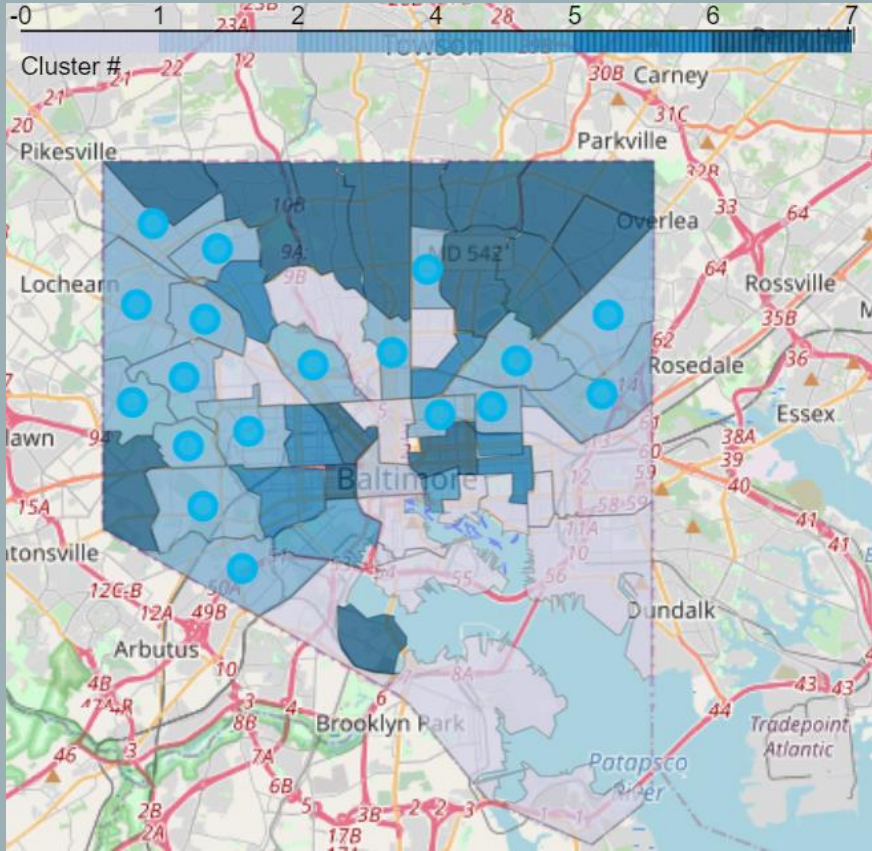
Top Venues

- Café
- Coffee Shop
- Sandwich Place
- Hotel
- Deli / Bodega

Neighborhoods:

- Downtown/Seton Hill

CLUSTER #3



Summary: Moderate everything.

Recommend: **Yes.** Moderate crime, but may be more affordable,. Has a mix of shops and restaurants.

Crime:

- Crime: moderate, average rate is 56.2 per 1000 people
- Violent Crime: moderate, average rate is 14.5 per 1000 people
- Property Crime: moderate, average rate is 41.0 per 1000 people

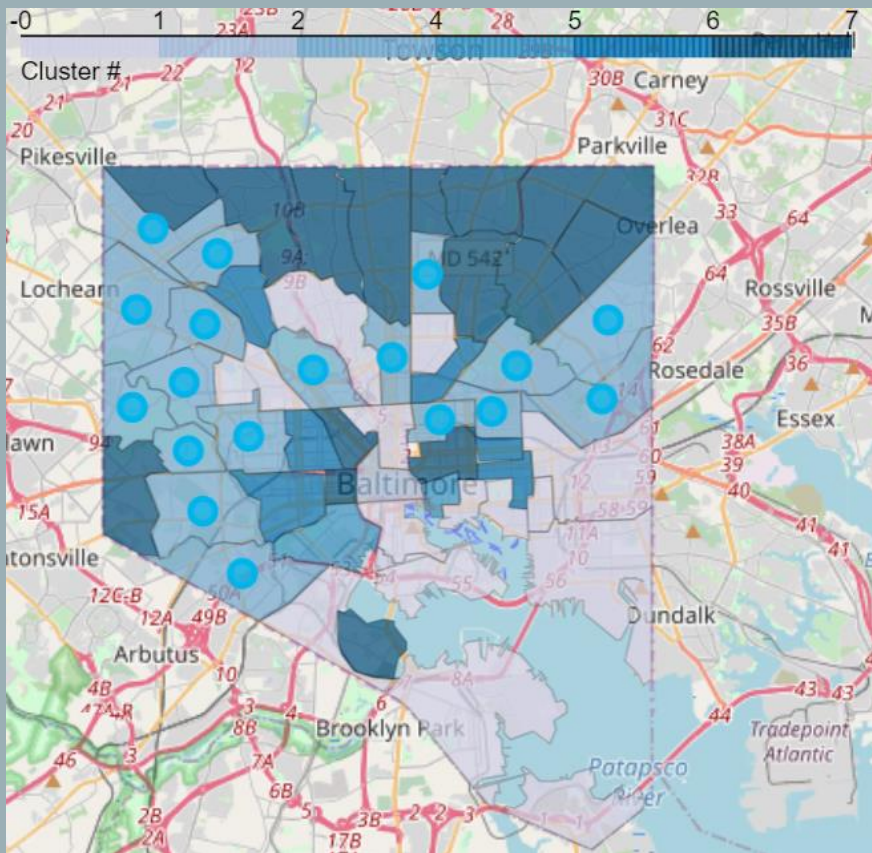
People:

- Median Income: moderate, average is 35484.7 dollars per household
- Poverty: moderate, average percentage of families in poverty is 20.7
- Population Density: moderate, average population density is 13.5 people per acre
- 25-64 age group: moderate, average percentage of people in the 25-64 years age group is 53.3
- Household Size: couples & small families, average average household size is 2.6 people

CLUSTER #3

Summary: Moderate everything.

Recommend: **Yes.** Moderate crime, but may be more affordable,. Has a mix of shops and restaurants.



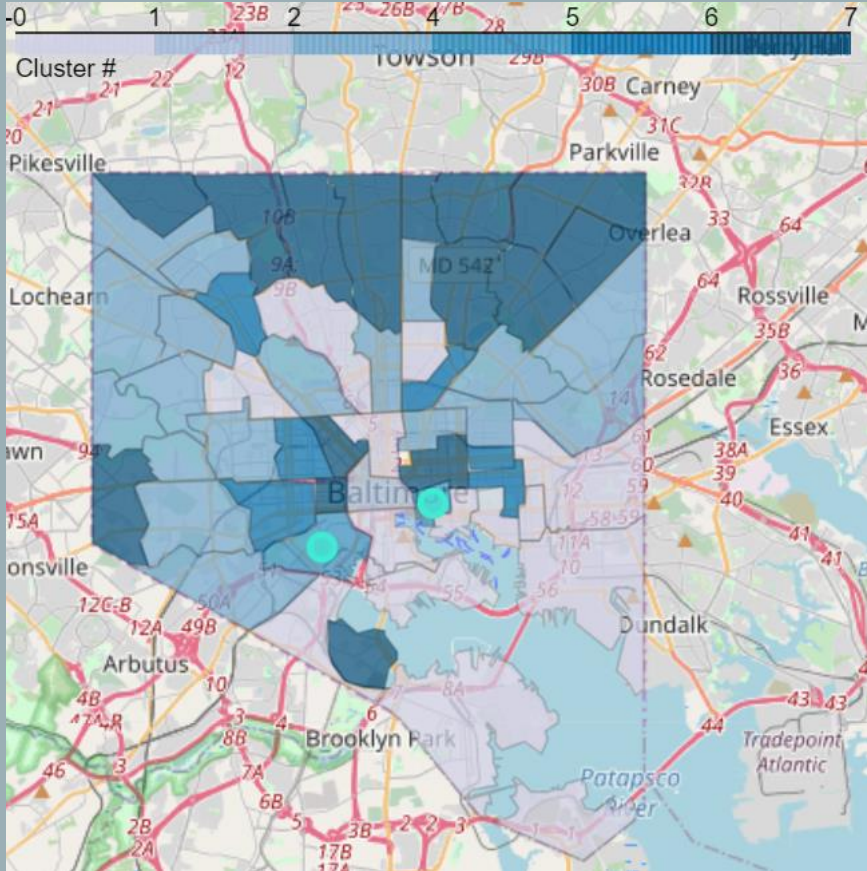
Top Venues

- Fried Chicken Joint
- Park
- Convenience Store
- American Restaurant
- Discount Store

Neighborhoods:

- Allendale/Irvington/S. Hilton
- Belair-Edison
- Cedonia/Frankford
- Claremont/Armistead
- Clifton-Berea
- Dickeyville/Franklintown
- Dorchester/Ashburton
- Edmondson Village
- Forest Park/Walbrook
- Glen-Fallstaff
- Greater Charles Village/Barclay
- Greater Govans
- Greater Rosemont
- Howard Park/West Arlington
- Morrell Park/Violetville
- Penn North/Reservoir Hill
- Pimlico/Arlington/Hilltop
- Greenmount East

CLUSTER #4



Summary: Very high crime, moderate income, high poverty, mostly couples, restaurants

Recommend: **No.** Very high crime, high poverty.

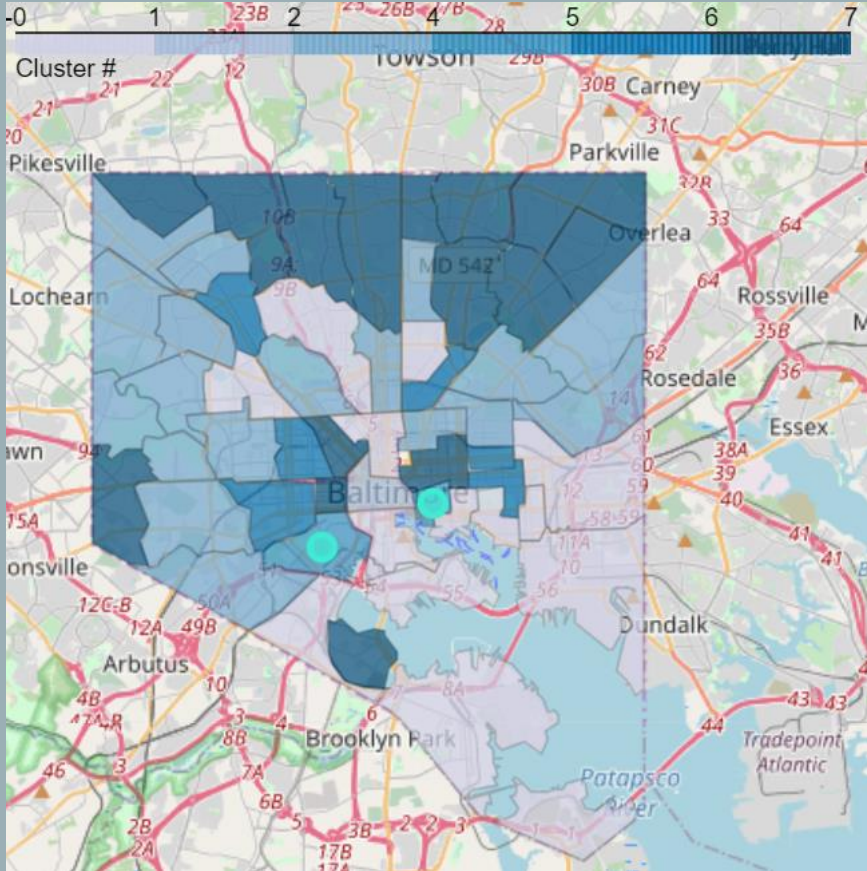
Crime:

- Crime: very high; average rate is 154.7 per 1000 people
- Violent Crime: very high, average rate is 29.0 per 1000 people
- Property Crime: very high, average rate is 123.5 per 1000 people

People:

- Median Income: moderate, average is 43536.5 dollars per household
- Poverty: high, average percentage of families in poverty is 31.4
- Population Density: moderate, average population density is 14.4 people per acre
- 25-64 age group: high, average percentage of people in the 25-64 years age group is 61.1
- Household Size: mostly couples, average average household size is 2.4 people

CLUSTER #4



Summary: Very high crime, moderate income, high poverty, mostly couples, restaurants

Recommend: **No.** Very high crime, high poverty.

Top Venues

- Italian Restaurant
- Pizza Place
- Rental Car Location
- Bakery
- Mobile Phone Shop

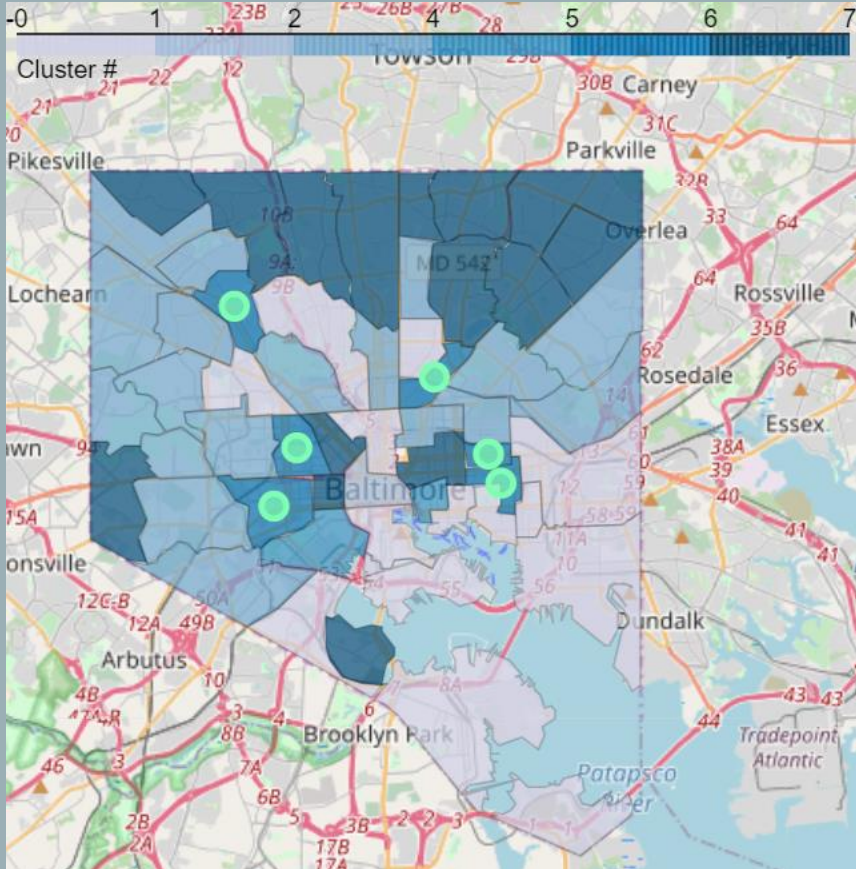
Neighborhoods:

- Washington Village/Pigtown
- Harbor East/Little Italy

CLUSTER #5

Summary: High crime, low income, high population density, high poverty.

Recommend: **No**. This is a poor, high crime urban area.



Crime:

- Crime: high, average rate is 79.3 per 1000 people
- Violent Crime: high, average rate is 23.3 per 1000 people
- Property Crime: moderate, average rate is 54.8 per 1000 people

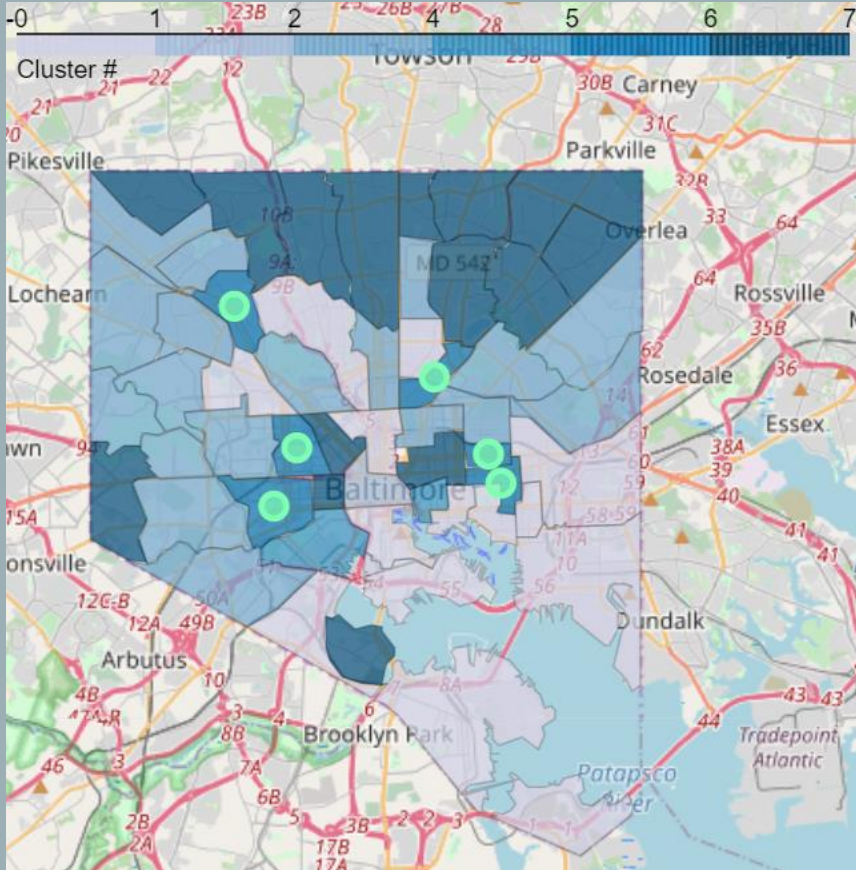
People:

- Median Income: low, average is 31578.8 dollars per household
- Poverty: high, average percentage of families in poverty is 31.6
- Population Density: very high, average population density is 28.7 people per acre
- 25-64 age group: moderate, average percentage of people in the 25-64 years age group is 54.3
- Household Size: couples & small families, average average household size is 3.0 people

CLUSTER #5

Summary: High crime, low income, high population density, high poverty.

Recommend: **No**. This is a poor, high crime urban area.



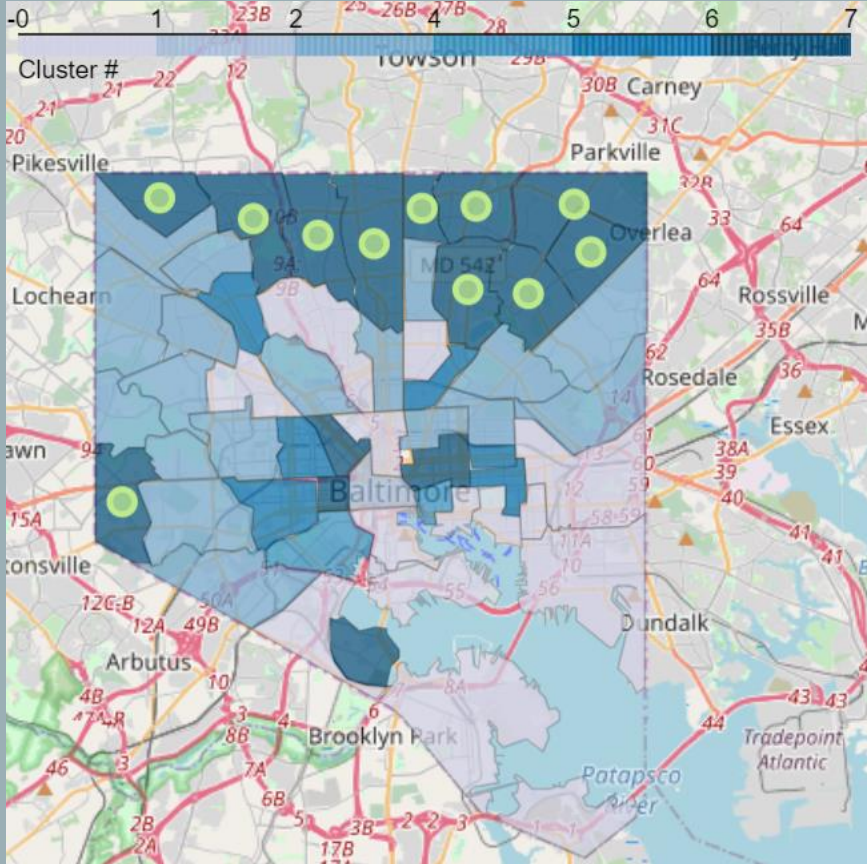
Top Venues

- Fast Food Restaurant
- Grocery Store
- Deli / Bodega
- Seafood Restaurant
- Fried Chicken Joint

Neighborhoods:

- Madison/East End
- Midway/Coldstream
- Patterson Park North & East
- Sandtown-Winchester/Harlem Park
- Southern Park Heights
- Southwest Baltimore

CLUSTER #6



Summary: Very low crime, high income, mostly couples, moderate population density, shops

Recommend: **Yes.** Rather suburban areas.

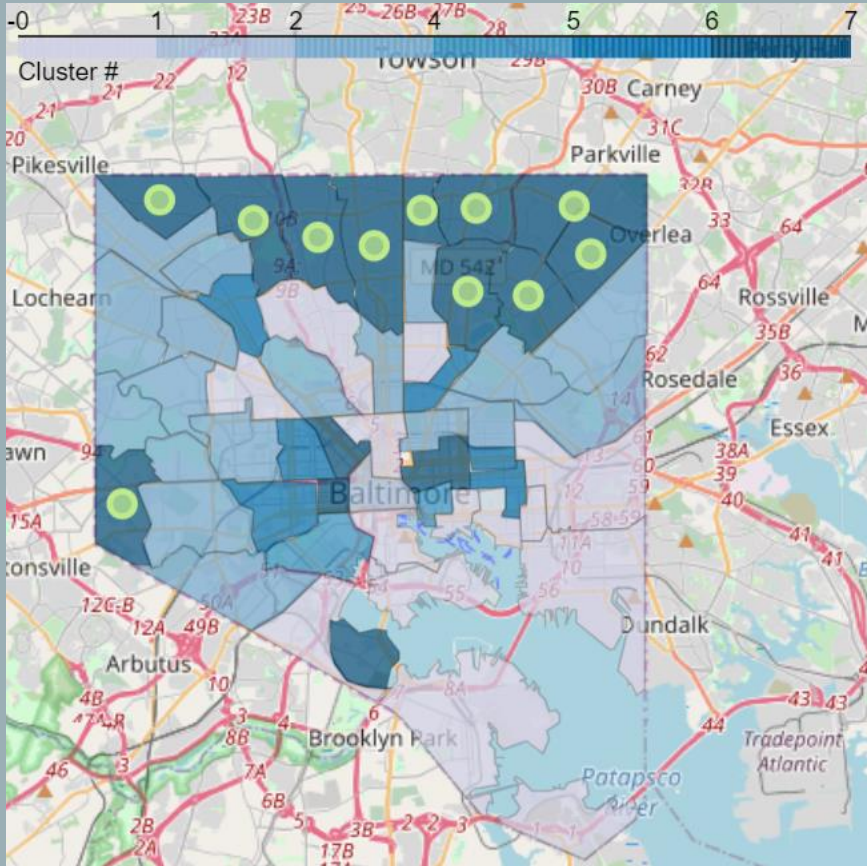
Crime:

- Crime: very low, average rate is 39.8 per 1000 people
- Violent Crime: low, average rate is 7.0 per 1000 people
- Property Crime: low, average rate is 32.2 per 1000 people

People:

- Median Income: high, average is 63439.3 dollars per household
- Poverty: low, average percentage of families in poverty is 8.2
- Population Density: moderate, average population density is 11.3 people per acre
- 25-64 age group: moderate, average percentage of people in the 25-64 years age group is 53.1
- Household Size: mostly couples, average average household size is 2.4 people

CLUSTER #6



Summary: Very low crime, high income, mostly couples, moderate population density, shops

Recommend: **Yes.** Rather suburban areas.

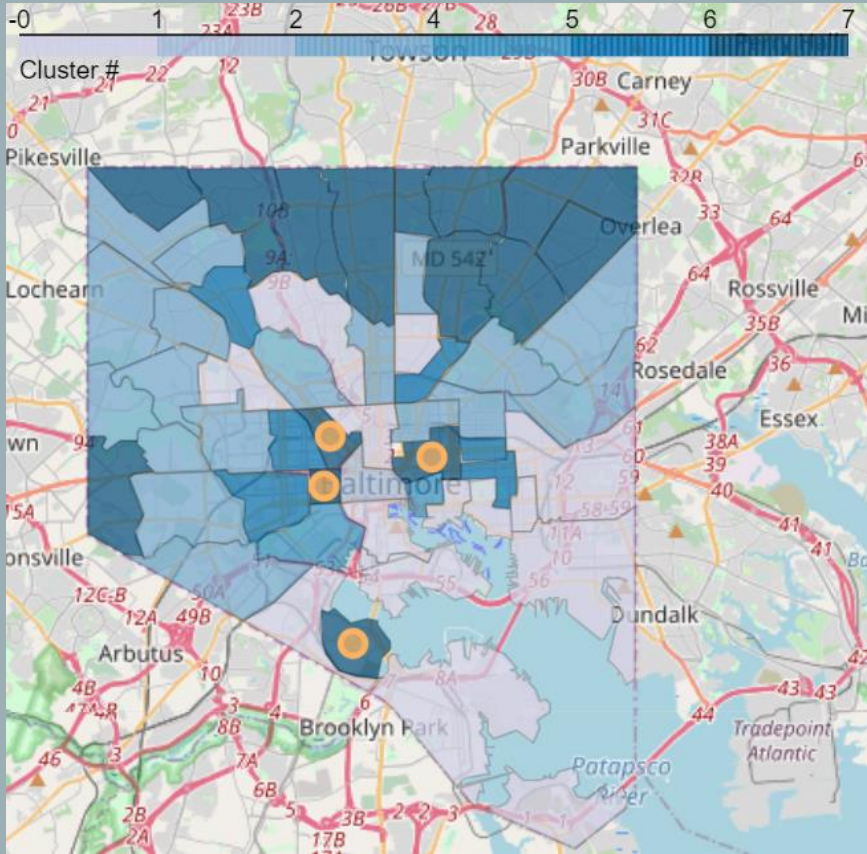
Top Venues

- Pharmacy
- Coffee Shop
- Flower Shop
- Bus Stop
- Grocery Store

Neighborhoods:

- Beechfield/Ten Hills/West Hills
- Chinquapin Park/Belvedere
- Cross-Country/Cheswolde
- Greater Roland Park/Poplar Hill
- Hamilton
- Harford/Echodale
- Lauraville
- Loch Raven
- Mount Washington/Coldspring
- North Baltimore/Guilford/Homeland
- Northwood

CLUSTER #7



Summary: High crime, very low income, very high poverty, high population density, low percentage of 25-64 year-olds

Recommend: **No.** This is high crime, low income, urban area.

Crime:

- Crime: high, average rate is 84.8 per 1000 people
- Violent Crime: high, average rate is 26.8 per 1000 people
- Property Crime: high, average rate is 56.7 per 1000 people

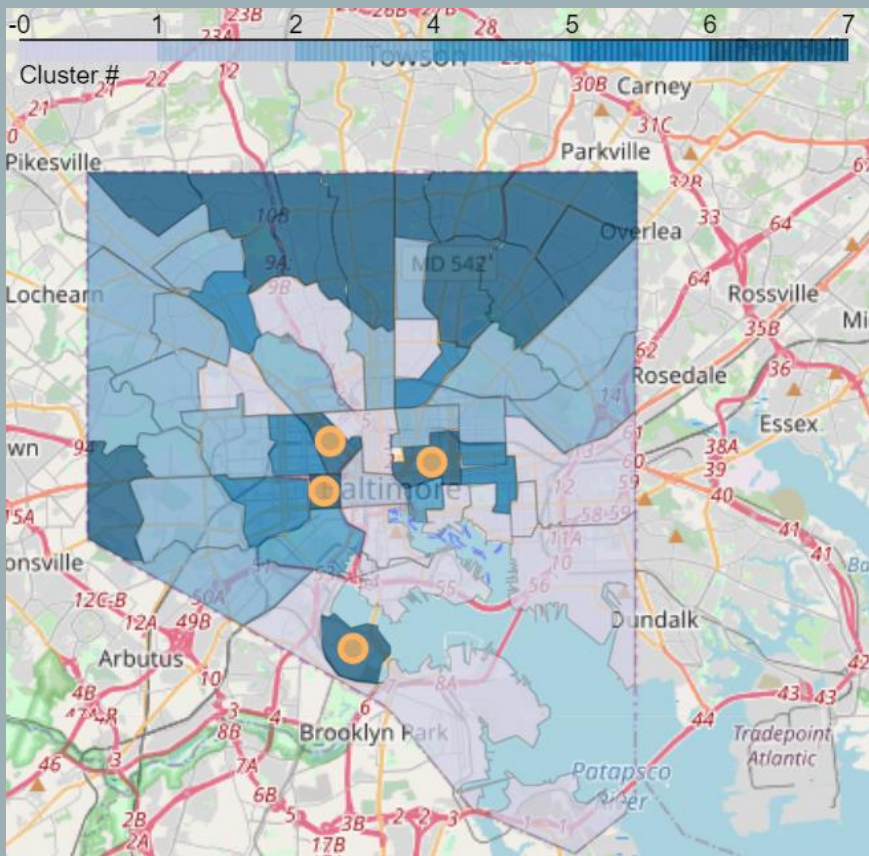
People:

- Median Income: very low, average is 18154.7 dollars per household
- Poverty: very high, average percentage of families in poverty is 46.3
- Population Density: high, average population density is 19.1 people per acre
- 25-64 age group: low, average percentage of people in the 25-64 years age group is 48.2
- Household Size: mostly couples, average average household size is 2.4 people

CLUSTER #7

Summary: High crime, very low income, very high poverty, high population density, low percentage of 25-64 year-olds

Recommend: **No.** This is high crime, low income, urban area.



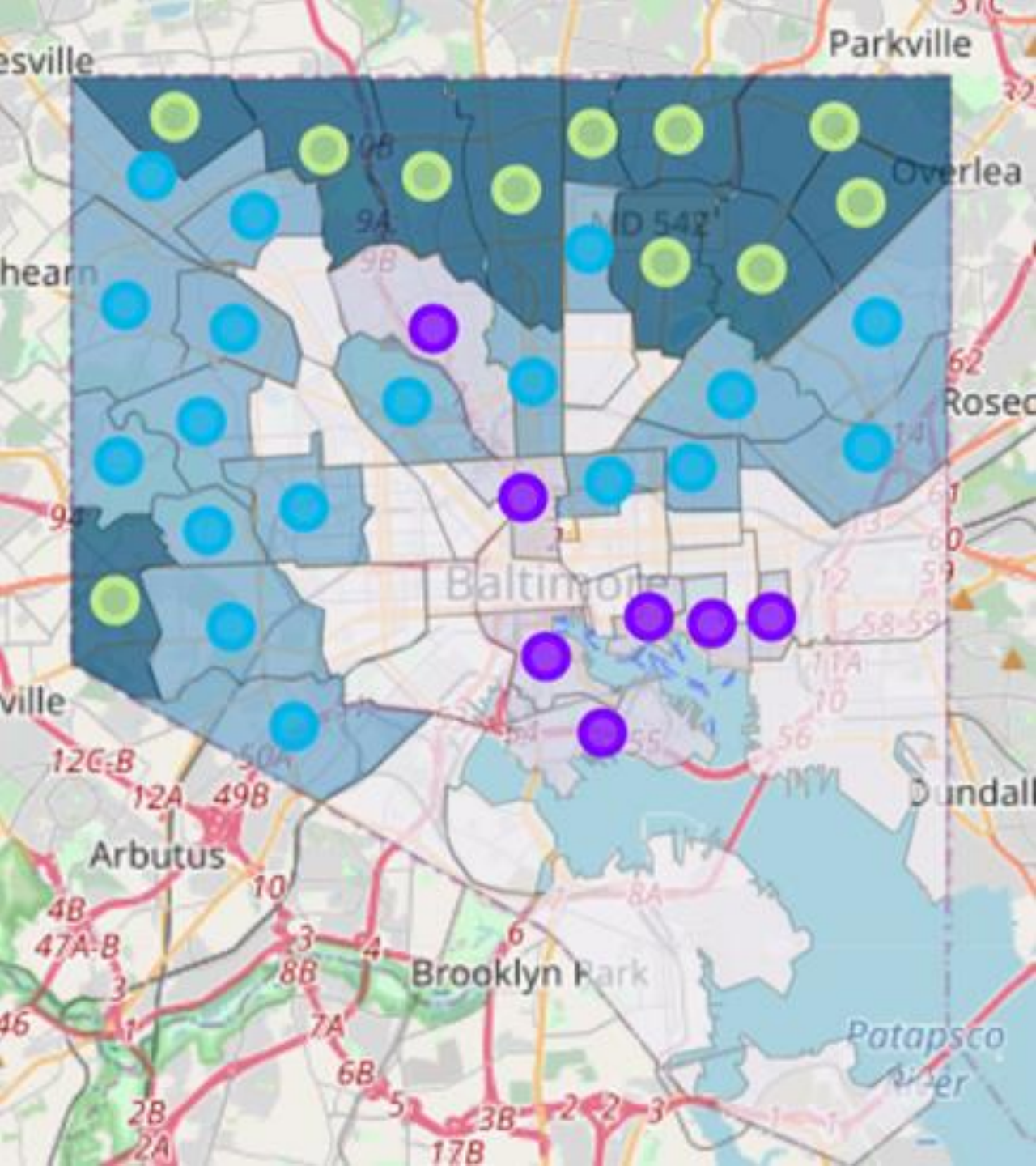
Top Venues

- Pizza Place
- Grocery Store
- Market
- Bus Stop
- Coffee Shop

Neighborhoods:

- Cherry Hill
- Poppleton/The Terraces/Hollins Market
- Oldtown/Middle East
- Upton/Druid Heights

RECOMMENDATIONS



RECOMMENDATION:
CLUSTERS #6, #1, #3

RECOMMENDED NEIGHBORHOODS

Cluster #6:

- Safest, but suburban
- Neighborhoods:
 - Beechfield/Ten Hills/West Hills
 - Chinquapin Park/Belvedere
 - Cross-Country/Cheswolde
 - Greater Roland Park/Poplar Hill
 - Hamilton
 - Harford/Echodale
 - Lauraville
 - Loch Raven
 - Mount Washington/Coldspring
 - North Baltimore/Guilford/Homeland
 - Northwood

Cluster #1:

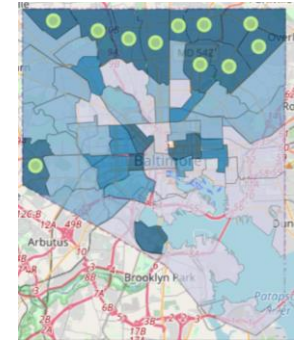
- Trendy, but moderate crime
- Neighborhoods:
 - Canton
 - Fells Point
 - Highlandtown
 - Inner Harbor/Federal Hill
 - Medfield/Hampden/Woodberry/Remington
 - South Baltimore
 - Midtown

Cluster #3:

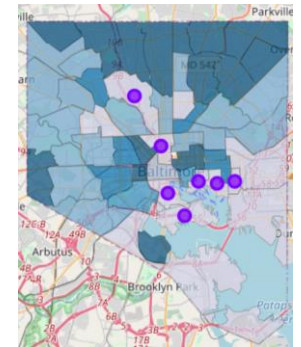
- More affordable, moderate crime
- Neighborhoods:
 - Allendale/Irvington/S. Hilton
 - Belair-Edison
 - Cedonia/Frankford
 - Claremont/Armistead
 - Clifton-Berea
 - Dickeyville/Franklintown
 - Dorchester/Ashburton
 - Edmondson Village
 - Forest Park/Walbrook
 - Glen-Fallstaff
 - Greater Charles Village/Barclay
 - Greater Govans
 - Greater Rosemont
 - Howard Park/West Arlington
 - Morrell Park/Violetville
 - Penn North/Reservoir Hill
 - Pimlico/Arlington/Hilltop
 - Greenmount East

SEE REPORT FOR SUMMARY
OF EACH NEIGHBORHOOD

#6



#1



#3

