Neighborhoods of Baltimore: People, Crime, and Venues

4/09/2020, CJRR

Contents:

1. Introduction: 2
   1.1: Objective: 2
   1.2: Background: 2
   1.3. Analytic Approach: 2
2. Data: 2
   2.1 Data Requirements: 2
   2.2 Data sources: 3
3. Methodology: 6
   3.1: Data Wrangling: 6
   3.2: Data Exploration: 12
   3.3: Analysis: 16
4. Results: 20
   4.1 Cluster #6: 21
   4.2 Cluster #1: 22
   4.3 Cluster #3: 23
5. Discussion 25
   5.1 Recommendation: 25
   5.2: Suggested Improvements: 25
6. Conclusion: 26

Appendix A: Cluster Characteristics: 27

Appendix B: Recommended Neighborhoods 35

1. Introduction:

1.1: Objective:

Characterize Baltimore's neighborhoods in terms of their crime rates, age of the residents, population density, median household income, and venues.

1.2: Background:

Client: The principal client for this project is Jane New Resident. She is a 38-year-old female professional who is moving to Baltimore. To start her house hunt, she is looking to identify low crime neighborhoods, with grocery stores, shops, restaurants and other professionals.

Baltimore: Nicknamed "Charm City", Baltimore is the most populous city in Maryland. It is home to the world-renowned Johns Hopkins University and Hospital and numerous tourist destinations including the Pimlico Race Course, the Inner Harbor, and the Maryland Science Center. However, crime in Baltimore is far above the national average and the city has a practically high murder rate. The city is divided into nine geographic regions and over 50 neighborhoods: each neighborhood having its own character. The city's crime is generally concentrated in the neighborhoods with high rates of poverty. [i]

1.3. Analytic Approach:

To help Jane select neighborhoods for her house hunt, this analysis will collect and generate summary data on the crime, number and type of venues, median income, age, and household size, for each neighborhood. It will then cluster the neighborhoods, summarize each cluster and display the information in an interactive map. This will allow Jane to review the character of each cluster and select a subset of neighborhoods to start searching in.

2. Data:

2.1 Data Requirements:

The analysis described in 1.3 will require the following data:

- List of the neighborhoods in Baltimore
- The geographic boundaries of the neighborhoods
- General crime statistics for each neighborhood
- General demographic statistics for each neighborhood
- Venue information for each neighborhood

The client wants to understand the general character of neighborhoods so she can start house hunting. She is worried about the crime rate. The analysis will need to include an overall picture of the crime rate of each neighborhood.

She is looking to live near other like-professionals – a factor that is not easy to characterize. However demographic information may help identify those neighborhoods – specifically by examining the median income, the average household size, and the percentage of people that are in the 25 to 64-year-old age group. Because the crime rate is related to the high poverty areas, the analysis will include that as well.

She is interested to understand the character of neighborhoods in terms of the available shopping and restaurants. For each neighborhood a listing of the type and number of nearby venues will be used.

There are over 50 neighborhoods in Baltimore. The client wants to consider them in like character clusters so that she does not over-constrain her house-hunting options.

2.2 Data sources:

2.2.1 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

This data set represents the year 2015's values for Crime and Safety from Vital Signs published by the Baltimore Neighborhood Indicators Alliance. For more information, please visit http://www.bniajfi.org/vital_signs.

The crime statistics are organized by neighborhood and include the following features:

- Part 1 Crime Rate per 1,000 Residents
- Violent Crime Rate per 1,000 Residents
- Property Crime Rate per 1,000 Residents
- Juvenile Arrest Rate per 1,000 Juveniles
- Juvenile Arrest Rate for Violent Offenses per 1,000 Juveniles
- Juvenile Arrest Rate for Drug Offenses per 1,000 Juveniles
- Rate of 911 Calls for Service for Shootings per 1,000 Residents
- Rate of Gun Homicides per 10,000 Residents
- Rate of 911 Calls for Service for Common Assaults per 1,000 Residents
- Rate of 911 Calls for Service for Narcotics per 1,000 Residents
- Rate of 911 Calls for Service for Auto Accidents per 1,000 Residents
- Adult Arrest Rate per 1,000 Adults

For this analysis, of the above options, the following statistics will be included:

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

Example:

Community Statistical Areas (CSAs)	Part 1 Crime Rate per 1,000 Residents	Violent Crime Rate per 1,000 Residents	Property Crime Rate per 1,000 Residents
Allendale/Irvington/S. Hilton	59.320466	15.045939	43.842881
Beechfield/Ten Hills/West Hills	37.345075	7.909328	28.946510
Belair-Edison	52.537896	11.770785	39.389068
Brooklyn/Curtis Bay/Hawkins Point	61.644317	18.746051	42.266376
Canton	51.234568	4.567901	46.543210

2.2.2 Demographics:

2015 Census demographics data for Baltimore City is available on Data.gov here:

https://catalog.data.gov/dataset/census-demographics-2015

This data set represents the year 2015's values for Census Demographics from Vital Signs published by the Baltimore Neighborhood Indicators Alliance. For more information, please visit http://www.bniajfi.org/vital_signs.

The demographics statistics are organized by neighborhood and include the following features:

- Percent of Residents - Black/African-American
- Percent of Residents - White/Caucasian Percent of Residents - Asian
- Percent of Residents - Two or More Races
- Percent of Residents - All Other Races
- Percent of Residents - Hispanic Racial Diversity Index
- Percent of Population Under 5 Years Old
- Percent of Population 5-17 years old
- Percent of Population 18-24 years old
- Percent of Population 25-64 years old
- Percent of Population 65 years and over
- Total Number of Households
- Percent of Female-Headed Households with Children Under 18
- Percent of Households with Children Under 18
- Average Household Size
- Median Household Income
- Percent of Households Earning Less than $25,000
- Percent of Households Earning $25,000 to $40,000
- Percent of Households Earning $40,000 to $60,000
- Percent of Households Earning $60,000 to $75,000
- Percent of Households Earning More than $75,000
- Percent of Family Households Living Below the Poverty Line
- Percent of Children Living Below the Poverty Line

For this analysis, of the above options, the following statistics will be included:

- 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

Example:

	Community Statistical Areas (CSAs)	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
0	Allendale/Irvington/S. Hilton	51.001244	2.670000	36701.90674	24.147122	5969
1	Beechfield/Ten Hills/West Hills	56.381067	2.550000	51537.58208	11.169652	5249
2	Belair-Edison	54.652308	2.762500	38173.96825	18.612071	5914
3	Brooklyn/Curtis Bay/Hawkins Point	56.413237	2.906667	36679.05344	28.359564	5012
4	Canton	73.958333	1.986667	95362.40031	2.998605	3958

2.2.3 Neighborhoods and Boundaries:

Both the demographics and crime data sets include a list of neighborhood's names and geographic boundaries. See below for a map of Baltimore Community Statistical Areas, hereafter referred to as neighborhoods, and the centroid of each neighborhood:

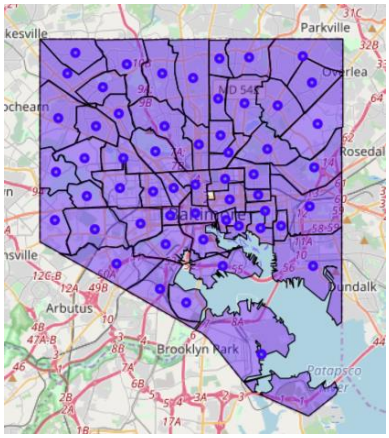

Figure: Baltimore Neighborhoods

2.2.4 Venues types and Counts:

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

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Greater Govans	39.344454	-76.604296	APlus at Sunoco	39.346470	-76.609390	Coffee Shop
Greater Govans	39.344454	-76.604296	Jakhashakxy Professional Services	39.344591	-76.607754	Home Service
Greater Govans	39.344454	-76.604296	Richwood Corner Store	39.346634	-76.607489	Snack Place
Greater Govans	39.344454	-76.604296	League Wellness Center	39.344495	-76.598981	Gym / Fitness Center
Greater Govans	39.344454	-76.604296	Chicken Run	39.345888	-76.609303	Fried Chicken Joint

3. Methodology:

3.1: Data Wrangling:

3.1.1 Formatting and reformatting….

The 2015 crime and statistics and census data files posted on data.gov were a mess. The json files were not in a readable format. The CSV files had formatting inconsistencies. All the data was present, however to make it workable it had to be manually reformatted. When completed there were three files, a CSV containing the crime and safety data, a CSV containing the census data, and a GEOJSON file containing multipolygon shapes for each neighborhood. The reformatted files are posted here:

GEOJSON:

https://github.com/AstroCris/CoureraCourses/blob/master/Baltimore44.geojson

2015 Demographics:

https://github.com/AstroCris/CoureraCourses/blob/master/Census_Demographics__2015__data_clean.csv

2015 Crime and Safety:

https://github.com/AstroCris/CoureraCourses/blob/master/Crime_and_Safety__2015__data_clean.csv

3.1.2 Reading and Slicing:

GEOJSON:

Simply reading the GEOJSON file and using Folium to produce a choropleth map produces the desired result. Each region that corresponds to a neighborhood is colored. The color has no value in this example.

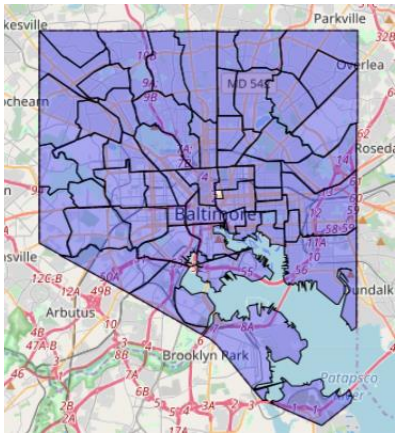

Figure: Neighborhoods

The uncolored region in the center of the city, initially was a concern. However, that region corresponds to the city jail. The data sets do not include any crime or demographics information for this region, nor is it relevant to our house hunting client.

2015 Crime and Safety Data:

Using pandas to read the crime and safety CSV file produces the following dataframe (as viewed with .head())

9]:

	the_geom	OBJECTID	Community Statistical Areas (CSAs)	Part 1 Crime Rate per 1,000 Residents	Violent Crime Rate per 1,000 Residents	Property Crime Rate per 1,000 Residents	Juvenile Arrest Rate per 1,000 Juveniles	Juvenile Arrest Rate for Violent Offenses per 1,000 Juveniles	Juvenile Arrest Rate for Drug Offenses per 1,000 Juveniles	Rate of 911 Calls for Service for Shootings per 1,000 Residents	Rate of Gun Homicides per 10,000 Residents	Rate of 911 Calls for Service for Common Assaults per 1,000 Residents	Rate of 911 Calls for Service for Narcotics per 1,000 Residents	Rate of 911 Calls for Service for Auto Accidents per 1,000 Residents	Adult Arrest Rate per 1,000 Adults	Shape__Area	Shape__Length
0	NaN	1	Allendale/Irvington/S. Hilton	59.320466	15.045939	43.842881	21.177686	8.264463	1.549587	4.501449	0.924955	68.570019	51.674169	66.720109	21.873201	6.377046e+07	38770.16557
1	NaN	2	Beechfield/Ten Hills/West Hills	37.345075	7.909328	28.946510	3.793627	0.758725	0.000000	2.201566	0.407697	42.808219	21.689498	44.357469	9.129512	4.788253e+07	37524.95053
2	NaN	3	Belair-Edison	52.537896	11.770785	39.389068	24.112231	9.206488	0.876808	2.009646	0.516766	65.227377	24.919614	60.231971	20.798866	4.495003e+07	31307.31484
3	NaN	4	Brooklyn/Curtis Bay/Hawkins Point	61.644317	18.746051	42.266376	23.809524	5.602241	7.002801	4.282806	0.702099	132.415924	70.490767	69.367409	55.490196	1.760777e+08	150987.70190
4	MULTIPOLYGON (((-76.571398100367 39.2844148270...	5	Canton	51.234568	4.567901	46.543210	29.629630	0.000000	0.000000	0.370370	0.000000	29.135802	5.432099	57.283951	6.510763	1.540854e+07	23338.61248

The NaN entries in the *the_geom* column are entries where the multipolygon was corrupted in the CSV and were manually removed. Those entries were corrected manually in the GEOJSON file.

Selecting the desired features (as discussed in section 2.2.1) and simplifying the column names produces the following dataframe:

	hood	crime_rt	violet_rt	property_rt
0	Allendale/Irvington/S. Hilton	59.320466	15.045939	43.842881
1	Beechfield/Ten Hills/West Hills	37.345075	7.909328	28.946510
2	Belair-Edison	52.537896	11.770785	39.389068
3	Brooklyn/Curtis Bay/Hawkins Point	61.644317	18.746051	42.266376
4	Canton	51.234568	4.567901	46.543210

2015 Demographics:

Using pandas to read the 2015 Census CSV file produces the following dataframe (as viewed with .head())

	the_geom	OBJECTID	Community Statistical Areas (CSAs)	Percent of Residents - Black/African-American	Percent of Residents - White/Caucasian	Percent of Residents - Asian	Percent of Residents - Two or More Races	Percent of Residents - All Other Races	Percent of Residents - Hispanic	Racial Diversity Index	...	Median Household Income	Percent of Households Earning Less than $25,000	Percent of Households Earning $25,000 to 40,000	Percent of Households Earning $40,000 to 60,000	Percent of Households Earning $60,000 to 75,000	Percent of Households Earning More than $75,000	Percent of Families Living Below the Poverty Line	Percent of Children Living Below the Poverty Line	Shape__Area	Shape__Length
0	NaN	1	Allendale/Irvington/S. Hilton	86.865672	8.955224	0.422886	2.151741	0.043532	1.560945	24.729153	...	36701.90674	36.069693	18.847378	16.786731	12.732451	15.563746	24.147122	38.931487	6.377046e+07	38770.16557
1	NaN	2	Beechfield/Ten Hills/West Hills	78.078191	15.443379	0.838826	2.366687	0.434392	2.838526	39.162063	...	51537.58208	20.384835	17.203277	20.956373	12.269004	29.186512	11.169652	19.416892	4.788253e+07	37524.95053
2	NaN	3	Belair-Edison	88.695385	8.849231	0.547692	0.947692	0.215385	0.744615	21.468088	...	38173.96825	34.950964	16.993575	17.906662	7.592154	22.556645	18.612071	36.882547	4.495003e+07	31307.31484
3	NaN	4	Brooklyn/Curtis Bay/Hawkins Point	37.597429	45.214002	1.374265	3.336524	1.148639	11.329140	69.857510	...	36679.05344	34.078212	19.393456	21.787710	9.517159	15.223464	28.359564	45.008801	1.760777e+08	150987.70190
4	MULTIPOLYGON (((-76.571398100367 39.2844148270...	5	Canton	2.459839	86.847390	3.790161	1.744478	0.238454	4.919679	26.293699	...	95362.40031	8.842850	8.716523	11.748358	7.377463	63.314805	2.998605	5.490849	1.540854e+07	23338.61248

The NaN entries in the *the_geom* column are entries where the multipolygon was corrupted in the csv and were manually removed. Those entries were corrected manually in the GEOJSON file.

Calculate population density per acre by multiplying the total number of households by the average household size and dividing by the area of the neighborhood. Multiplying the result by 43560 converts the result from $1/ft^2$ to 1/acres.

Selecting the desired features (as discussed in section 2.2.2) and simplifying the column names produces the following dataframe:

	hood	25-64_pct	household_ave	income_median	pct_poverty	density
0	Allendale/Irvington/S. Hilton	51.001244	2.670000	36701.90674	24.147122	10.886321
1	Beechfield/Ten Hills/West Hills	56.381067	2.550000	51537.58208	11.169652	12.176643
2	Belair-Edison	54.652308	2.762500	38173.96825	18.612071	15.832208
3	Brooklyn/Curtis Bay/Hawkins Point	56.413237	2.906667	36679.05344	28.359564	3.604041
4	Canton	73.958333	1.986667	95362.40031	2.998605	22.229373

3.1.3 Neighborhood Coordinates

The GEOJSON file contains the multipolygon shape of each neighborhood, expressed as a list of latitude and longitude coordinates. However, the Foursquares API requires a single coordinate point (expressed as latitude and longitude) and a radius (in meters).

The GEOJSON data should be sufficient to calculate reasonable center coordinates for each neighborhood. As a first attempt the average of the latitudes and longitudes for each neighborhood's multipolygon was calculated. However, this produced erroneous results:

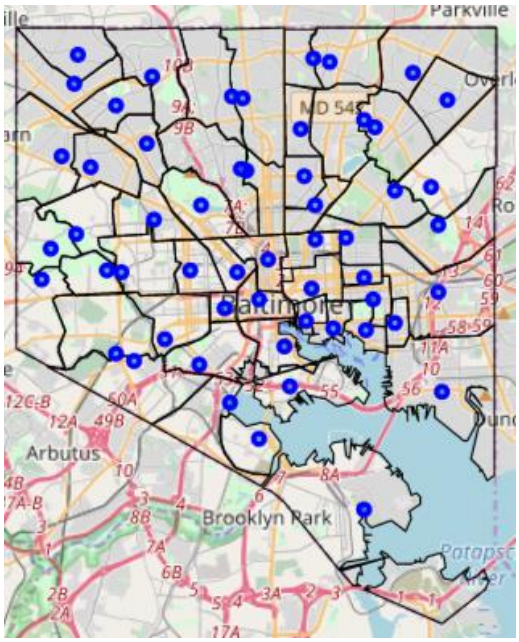

Figure: Neighborhood centers calculated by averaging

This averaging method is biased by the density of points needed to define each edge. A more complicated edge requires more points to define it. This shifts the calculated average coordinates towards that edge.

A more mathematically valid approach is to calculate the centroid of the polygon shape. See centroid for more information. The python library *shapely* has a function to do this. See shapely for more information. Using this function produced superior results:

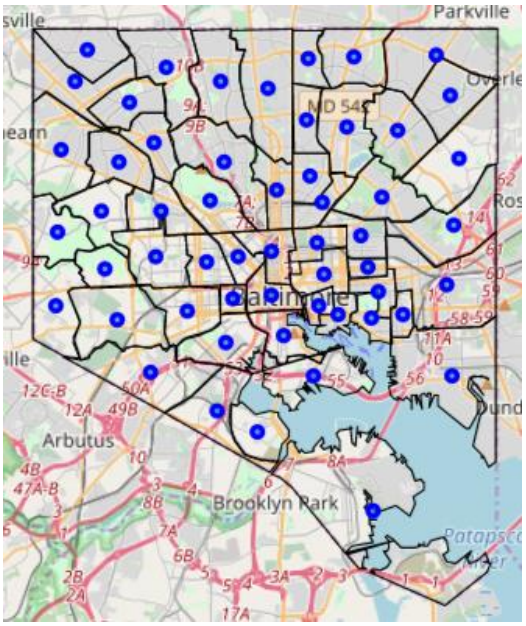

Figure: Neighborhood centroids

Centroid latitude and longitude for each neighborhood was loaded into a dataframe:

index	hood	lat	long
Greater Govans	Greater Govans	39.344454	-76.604296
Hamilton	Hamilton	39.351621	-76.547891
Sandtown-Winchester/Harlem Park	Sandtown-Winchester/Harlem Park	39.301178	-76.643323
Dorchester/Ashburton	Dorchester/Ashburton	39.331750	-76.677971
Howard Park/West Arlington	Howard Park/West Arlington	39.335560	-76.700687

3.1.4 Venue Data:

Foursquare was used to collect a list of venues details (name, location, category) for venues within 1000 meters of each neighborhood's centroid, up to 100 venues for each neighborhood. A total of 889 venues in 216 unique categories were collected. The result loaded into a dataframe has this form:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Greater Govans	39.344454	-76.604296	APlus at Sunoco	39.346470	-76.609390	Coffee Shop
1	Greater Govans	39.344454	-76.604296	Jakhashakxy Professional Services	39.344591	-76.607754	Home Service
2	Greater Govans	39.344454	-76.604296	Richwood Corner Store	39.346634	-76.607489	Snack Place
3	Greater Govans	39.344454	-76.604296	League Wellness Center	39.344495	-76.598981	Gym / Fitness Center
4	Greater Govans	39.344454	-76.604296	Chicken Run	39.345888	-76.609303	Fried Chicken Joint

The geographic distribution of the 889 venues:

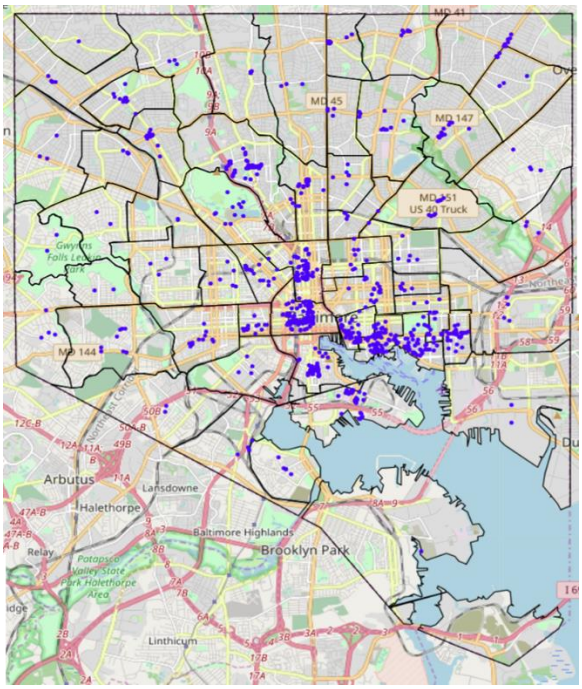

Figure: Foursquare venues within 1000 m of each neighborhood's centroid

On inspection some neighborhoods are sparsely populated with Foursquare venues. For example, Foursquare did not return any venues in Mount Washington Village, despite the presence of shops, restaurants, art galleries, a church, and a school in the village. Increasing the query radius did not resolve this problem. This calls into question the validity of all the results that follow. See Discussion for more detail.

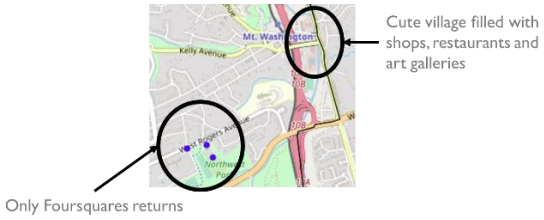

Figure: Where are the Mount Washington Village's venues?

Since the objective is to characterize each neighborhood, the list of venues need to be translated into a summary for each neighborhood. First using one-hot encoding the venue categories are converted into new columns and a '1' is placed to indicate which category that venue is in. Zero is placed otherwise. The result has this form:

	Neighborhood	ATM	Adult Boutique	Afghan Restaurant	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Vietnamese Restaurant	Volleyball Court	Waterfall	Weight Loss Center	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
885	Greater Rosemont	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
886	Greater Rosemont	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
887	North Baltimore/Guilford/Homeland	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
888	North Baltimore/Guilford/Homeland	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
889	North Baltimore/Guilford/Homeland	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	1	0	0

Next the venues are grouped by neighborhood, summing the number of venues for each category:

	Neighborhood	ATM	Adult Boutique	Afghan Restaurant	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Vietnamese Restaurant	Volleyball Court	Waterfall	Weight Loss Center	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	Allendale/Irvington/S. Hilton	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Beechfield/Ten Hills/West Hills	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Belair-Edison	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Brooklyn/Curtis Bay/Hawkins Point	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Canton	0	0	0	3	0	0	0	0	0	...	1	0	0	0	0	0	1	0	0	0

Now the venue data is in a form that is relevant to the analysis objective.

3.1.5 Merging DataFrames

Merging the crime, census, and coordinates dataframes to make a dataframe for summary information:

	hood	25-64_pct	household_ave	income_median	pct_poverty	density	crime_rt	violet_rt	property_rt	lat	long
0	Allendale/Irvington/S. Hilton	51.001244	2.670000	36701.90674	24.147122	10.886321	59.320466	15.045939	43.842881	39.284047	-76.678537
1	Beechfield/Ten Hills/West Hills	56.381067	2.550000	51537.58208	11.169652	12.176643	37.345075	7.909328	28.946510	39.288136	-76.702264
2	Belair-Edison	54.652308	2.762500	38173.96825	18.612071	15.832208	52.537896	11.770785	39.389068	39.321214	-76.574747
3	Brooklyn/Curtis Bay/Hawkins Point	56.413237	2.906667	36679.05344	28.359564	3.604041	61.644317	18.746051	42.266376	39.225856	-76.577810
4	Canton	73.958333	1.986667	95362.40031	2.998605	22.229373	51.234568	4.567901	46.543210	39.284285	-76.578654

Merging the crime, census and venues counts dataframes to make a dataframe that contains the relevant data for clustering:

	hood	25-64_pct	household_ave	income_median	pct_poverty	density	crime_rt	violet_rt	property_rt	ATM	...	Vietnamese Restaurant	Volleyball Court	Waterfall	Weight Loss Center	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	Allendale/Irvington/S. Hilton	51.001244	2.670000	36701.90674	24.147122	10.886321	59.320466	15.045939	43.842881	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Beechfield/Ten Hills/West Hills	56.381067	2.550000	51537.58208	11.169652	12.176643	37.345075	7.909328	28.946510	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Belair-Edison	54.652308	2.762500	38173.96825	18.612071	15.832208	52.537896	11.770785	39.389068	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Brooklyn/Curtis Bay/Hawkins Point	56.413237	2.906667	36679.05344	28.359564	3.604041	61.644317	18.746051	42.266376	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Canton	73.958333	1.986667	95362.40031	2.998605	22.229373	51.234568	4.567901	46.543210	0.0	...	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

3.2: Data Exploration:

3.2.1 Crime Rate:

How do the crime rates vary from neighborhood to neighborhood?

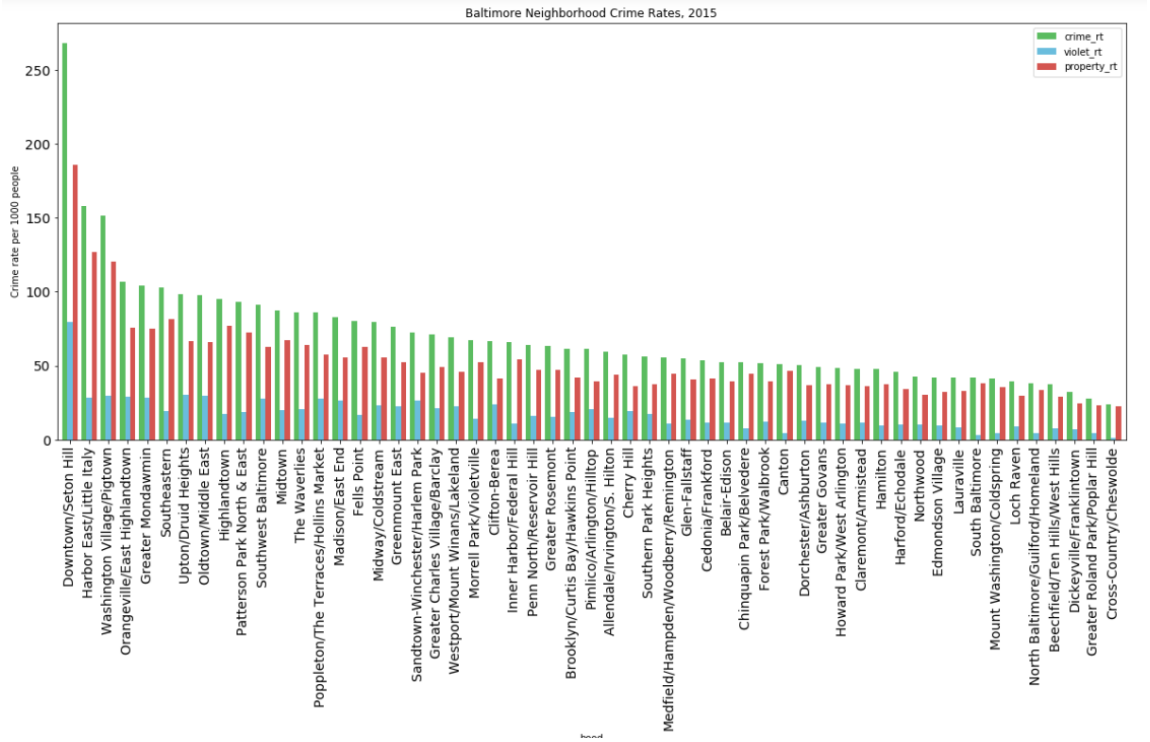

Figure: Total Crime rate (green), violent crime rate (blue), property crime rate (red) in 2015.

Geographically how do the crime rates vary?

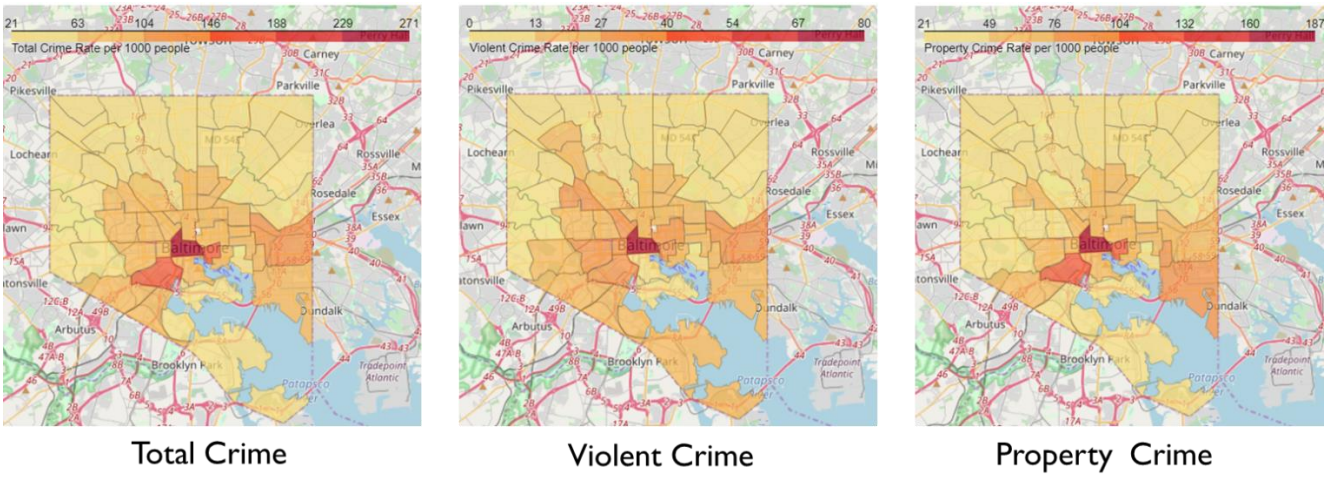

Figure: Geographic distribution of crime rates, per 1000 people, in 2015.

3.2.2 Median Household Income:

How does median household Income vary neighborhood to neighborhood?

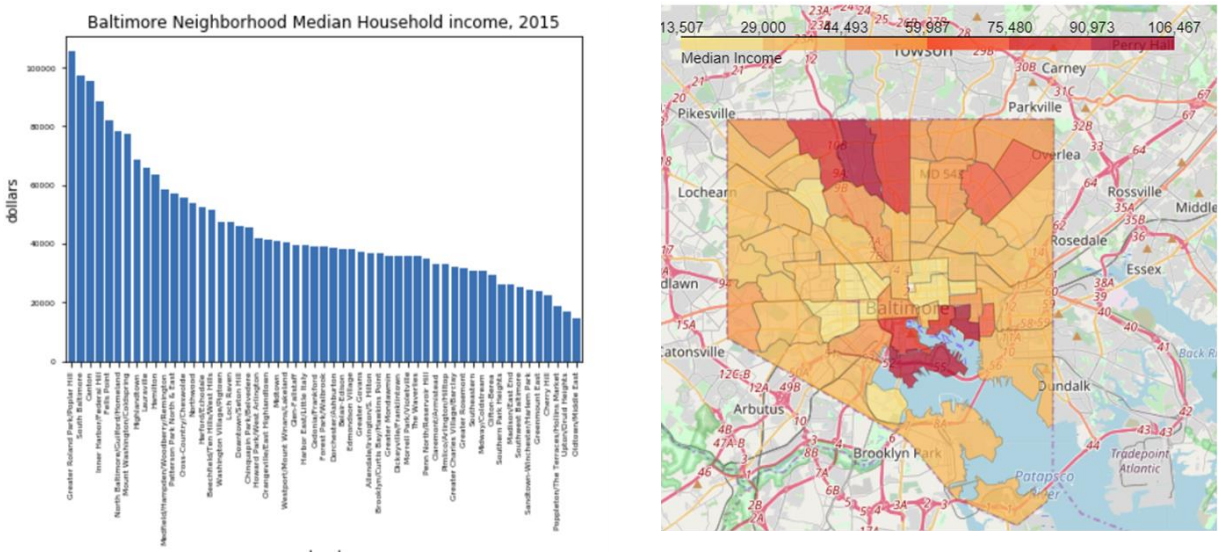

Figure: Median Household Income, 2015

3.2.3 Poverty:

How does the percentage of households in poverty vary neighborhood to neighborhood?

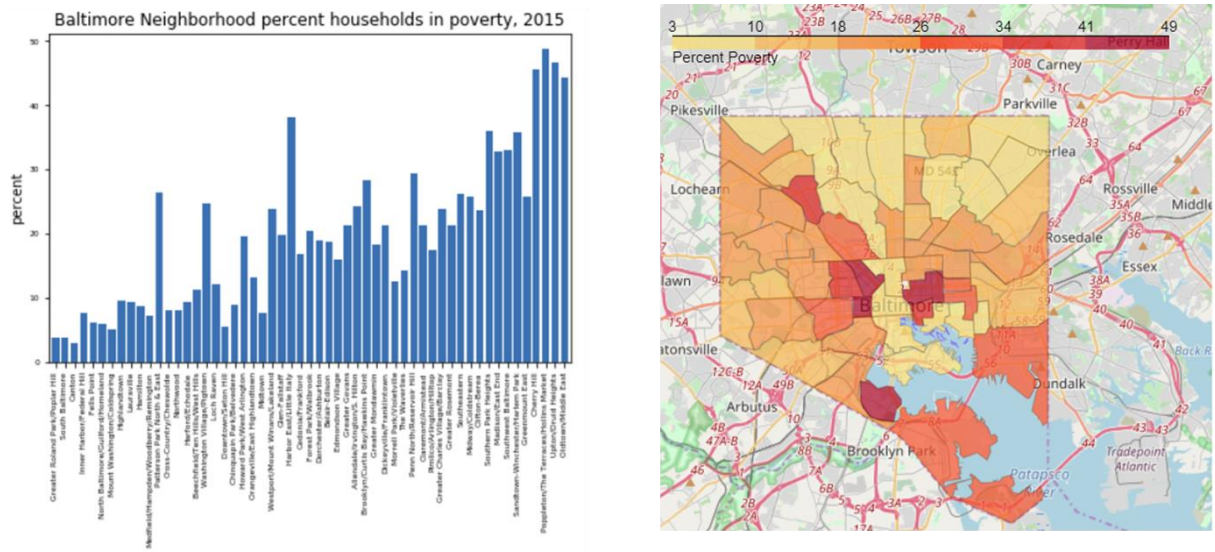

Figure: Percent of household in poverty, 2015

3.2.4 Working age population:

How does the percentage of the population that are 25 -64 years old vary neighborhood to neighborhood?

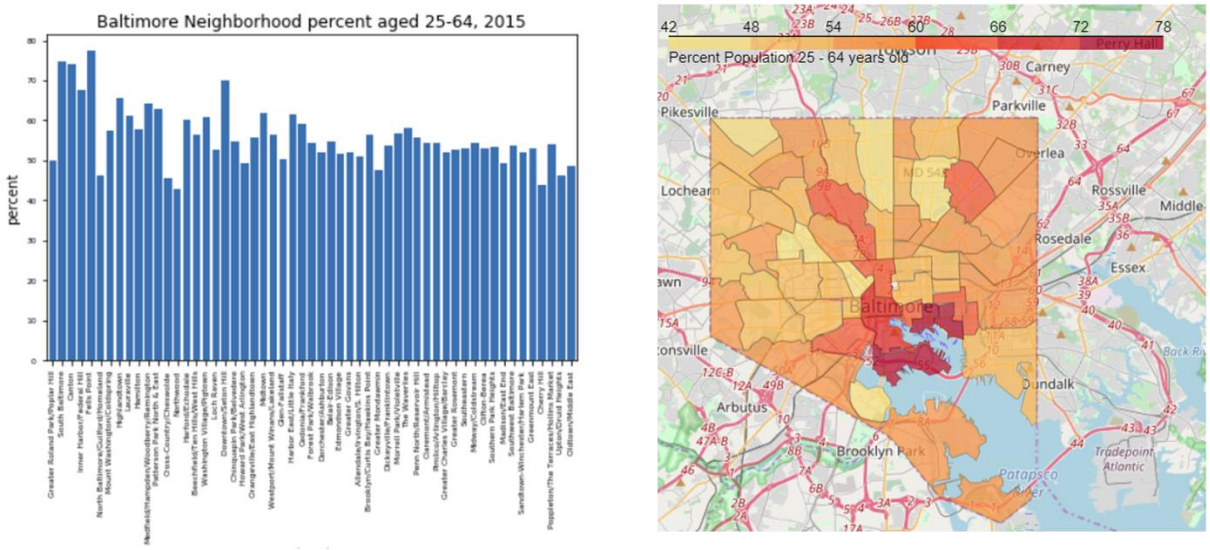

Figure: Percent of the population that is 25 -64 years old, 2015

3.2.5 Household size:

How does the average household size vary neighborhood to neighborhood?

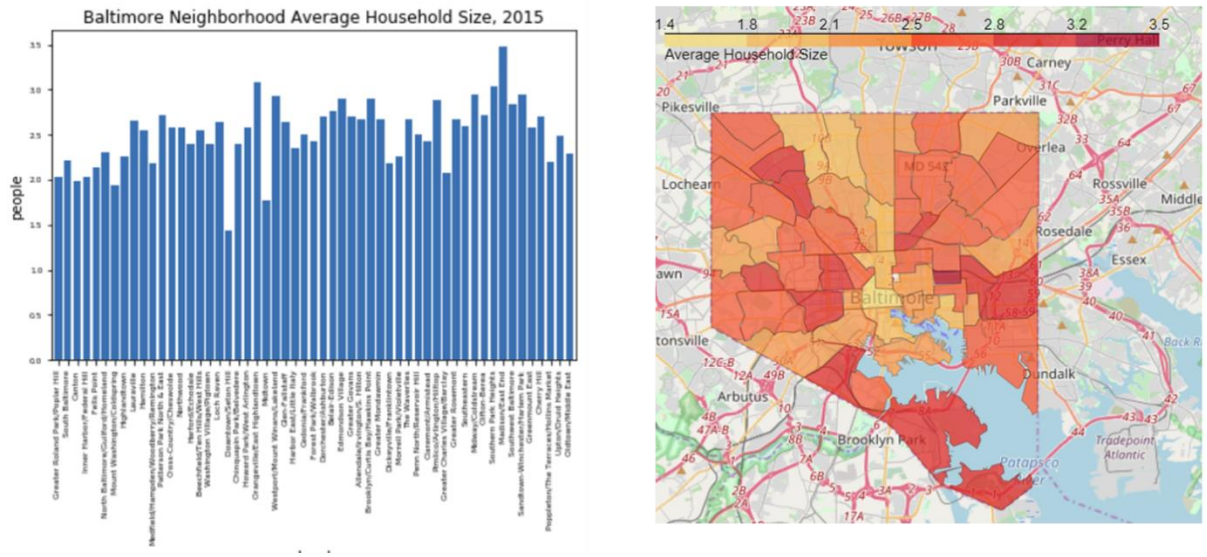

Figure: Average household size, 2015

3.2.6 Population Density:

How does the population density vary neighborhood to neighborhood?

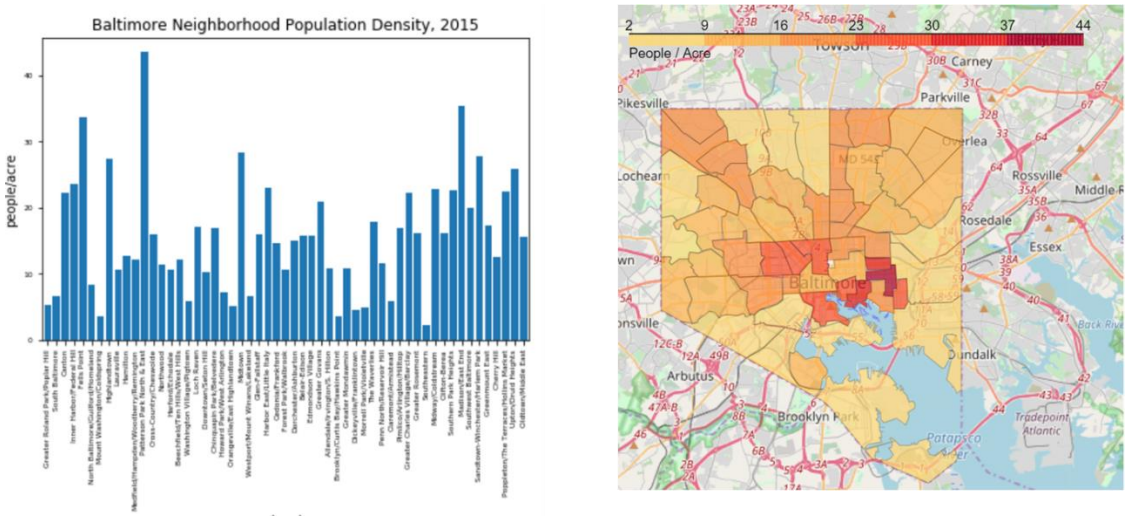

Figure: Neighborhood population density, 2015

3.3: Analysis:

3.3.1 Categorizing Statistical Data:

To help analyze and interpret the crime and census numerical data for each neighborhood, the numerical value for each feature was converted to a categorical value of very low, low, moderate, high, or very high. The categorical value 'very low' was assigned the lowest 10% of the features' values, 'low' was assigned to lowest 10-30% of values, moderate to the lowest 30-70%, high the lowest 70-90% , and very high >90% (equivalently, the highest 10%).

The average household size feature was then corrected and assigned categories that were indicative of the household make up. Average household sizes less than 1.5 were assigned the category 'mostly singles', between 1.5 and 2.5 was assigned 'mostly couples', between 2.5 and 3 'couples & small families', and >3 was assigned the categorical value 'families'.

The resulting values were added to the summary dataframe and has this form:

	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

In the process of translating the numeric features to categorical values based on statistical distribution (<10%, 10-30% etc.) the numerical values of the slicing bins were saved, thus recording numeric values that correspond to the cuts for each feature. These values were saved so that they are available to use later when analyzing the cluster average values for each feature.

3.3.2 Common Venues:

To aid in characterizing each neighborhood, the 7 most common venue types were tabulated. The result has the below form, which was then added to the summary dataframe. These new features will be used to generate summary reports for each recommended neighborhood.

	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

3.3.3 K-Means Clustering:

To group similar neighborhoods for the client's conservation, K-Means clustering was applied to the 'clustering' dataframe (see 3.1.5). This dataframe includes the crime and census data and the one-hot-encoded venue information complied for each neighborhood (see 3.1.4).

Because the venue data was sparse for some neighborhoods, K-Means clustering was first attempted with only the non-venue information. The data was selected from the dataframe and then normalized using StandardScaler.

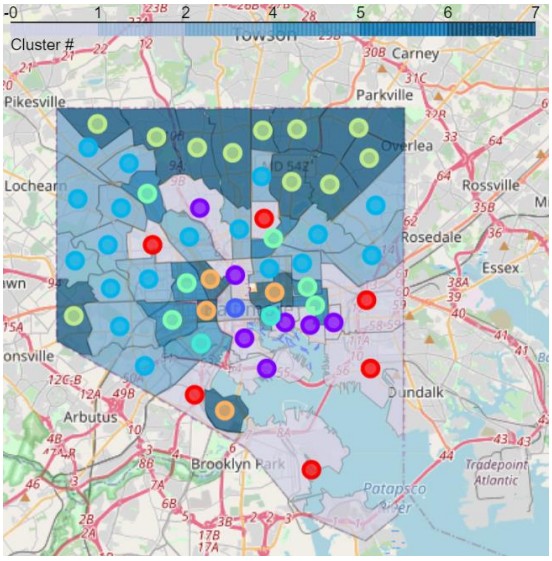

Figure: Clustered neighborhoods, 8 clusters, dot color and shade correspond to cluster number.

The algorithm was run for different numbers of clusters and the results were visually inspected (see below). A domain expert who was familiar with the city, suggested that the 8 cluster result was adequate (above). The 8-cluster result will be used in the rest of this analysis.

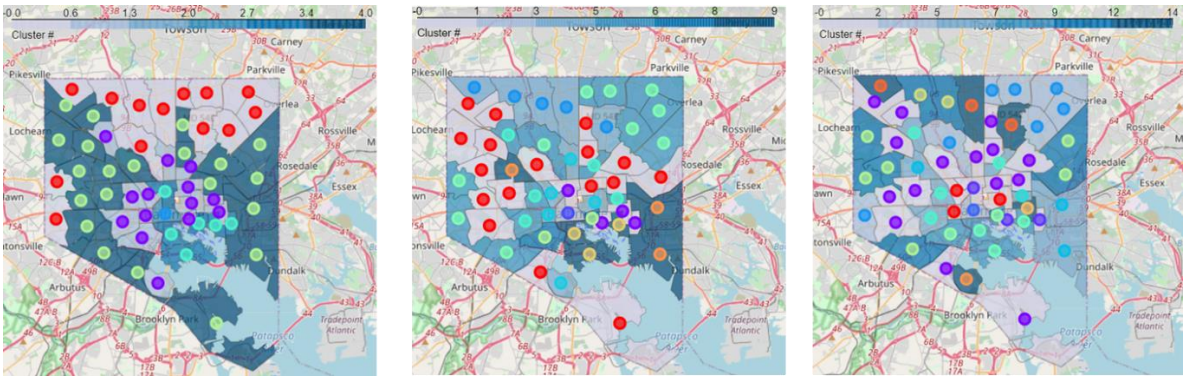

Figure: Clustering results for 5 (left), 10 (center), and 15 (right) clusters.

K-Means clustering was repeated to also include the one-hot encoded venue information. As before the data was first normalized using StandardScaler. The K-Means algorithm was run with different numbers of clusters and the results were visually inspected (see below).

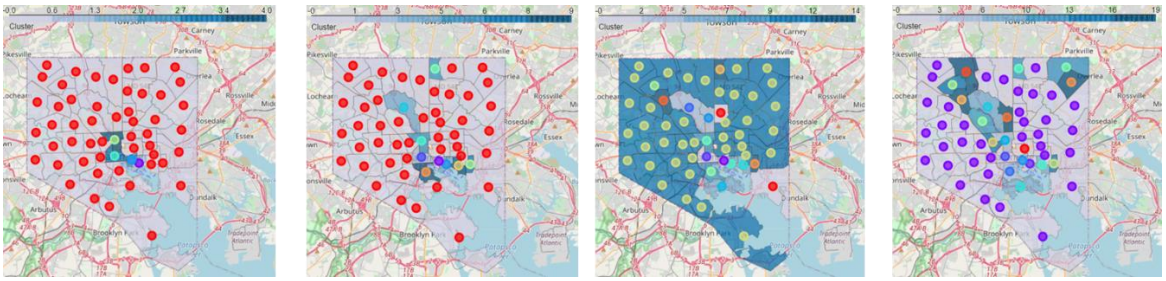

Figure: Clustering results for 5 (left), 10 (center left), 15 (center right) and 20 (right) clusters.

The results of all the clustering attempts, that included the venue information, placed the majority of neighborhoods in a single cluster. Such a result is not useful for the client.

The best clustering result for the current objective is for 8 clusters and without using the venue information.

3.3.4 Summarizing Clusters

To summarize the character of each cluster, the mean value of the features (including the venue counts) for the neighborhoods in that cluster was calculated:

cluster	25-64_pct	household_ave	income_median	pct_poverty	density	crime_rt	violet_rt	property_rt	ATM	Adult Boutique	...	Waterfall
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

Then using the same bins from 3.3.1 (Categorizing Statistical Data) a categorical representation of the cluster average values was added to the dataframe using the same numerical limits as in 3.3.1. This ensures that ‘very high’ means a value that is among the top 10% of the value for all neighborhoods – and not the top 10% for values amongst the cluster means.

The 5 most common venues were tabulated for each cluster as well.

cluster	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	25-64_pct	...	Vietnamese Restaurant	Volleyball Court	Waterfall	Weight Loss Center	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	moderate	couples & small families	moderate	moderate	low	high	high	high	[Business Service, Coffee Shop, Discount Store...	54.577119	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	very high	mostly couples	high	low	high	moderate	moderate	high	[Bar, Pizza Place, Mexican Restaurant, America...	69.367368	...	0.142857	0.142857	0.000000	0.000000	0.285714	0.285714	0.428571	0.000000	0.142857	0.142857
2	very high	mostly singles	moderate	very low	low	very high	very high	very high	[Café, Coffee Shop, Sandwich Place, Hotel, Del...	69.829190	...	3.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	moderate	couples & small families	moderate	moderate	moderate	moderate	moderate	moderate	[Fried Chicken Joint, Park, Convenience Store,...	53.329947	...	0.000000	0.000000	0.055556	0.055556	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000
4	high	mostly couples	moderate	high	moderate	very high	very high	very high	[Italian Restaurant, Pizza Place, Rental Car L...	61.112011	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000

A summary report for each cluster was printed and manually reviewed by a human analyst. See Appendix A and section 4 for the cluster summary report and the analyst's recommendations.

4. Results:

The above analysis clustered the 56 Baltimore neighborhoods into 8 clusters. Appendix A and the map and table below summarize the clusters.

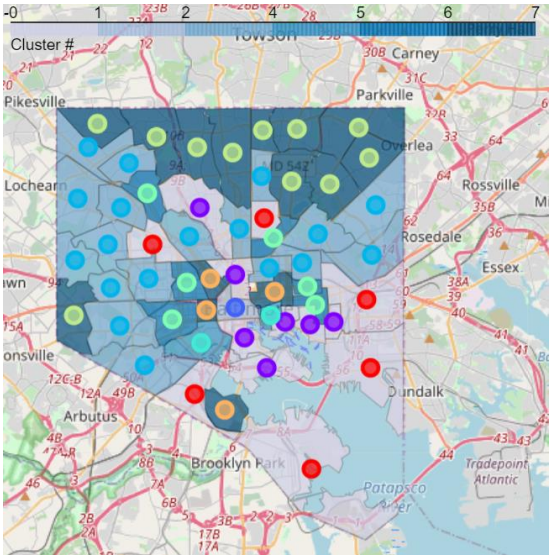

Figure: Neighborhoods of Baltimore Clustered

cluster	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
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,...

A human analyst examined the summary information of each of the clusters. The mean crime rate for clusters #2 and #4 is 'very high'. Similarly, the mean crime rate for clusters #0, #5, and #7 is 'high'. The neighborhoods in these clusters are not suitable for the client.

Remaining clusters are #6, #1 and #3:

cluster	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
1	very high	mostly couples	high	low	high	moderate	moderate	high	[Bar, Pizza Place, Mexican Restaurant, America...
6	moderate	mostly couples	high	low	moderate	very low	low	low	[Pharmacy, Coffee Shop, Flower Shop, Bus Stop,...
3	moderate	couples & small families	moderate	moderate	moderate	moderate	moderate	moderate	[Fried Chicken Joint, Park, Convenience Store,...

4.1 Cluster #6:

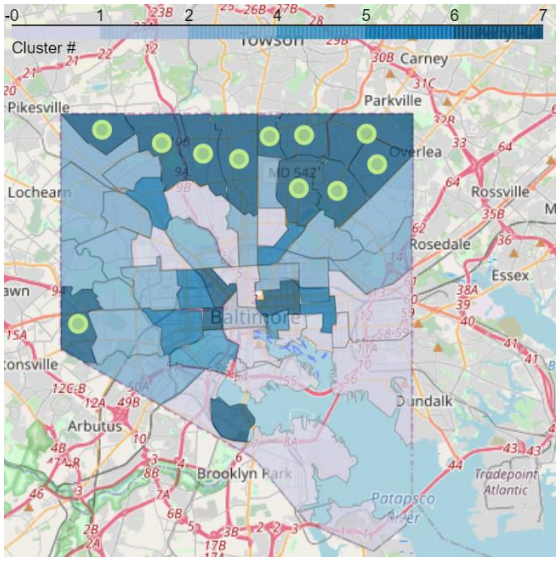

Figure: Cluster #6

Cluster #6 has 'very low' crime, 'high' median income, 'moderate' population density, and the mean average median household size corresponds to 'mostly couples'. A 'moderate' percentage of the population is in the professional age range (25-64). The neighborhoods in this cluster are all near the outer boundary of the city. The most common venues are rather pedestrian and practical: pharmacy, coffee shop, flower shop, bus stop, grocery store.

The location of these neighborhoods, the high median income, moderate population density, and practical venues suggests that these neighborhoods are more suburban in character.

Cluster #6 contains 11 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

See Appendix B for details on each.

4.2 Cluster #1:

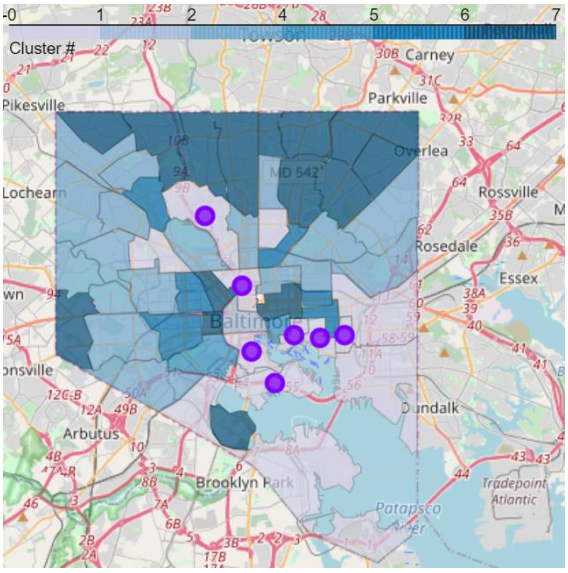

Figure: Cluster #1

Cluster #1 has 'moderate' crime, 'high' median income, 'high' population density, and the mean average median household size corresponds to 'mostly couples'. A 'very high' percentage of the population is in the professional age range (25-64). The neighborhoods in this cluster are all near the center of the city. The most common venues support an active social scene: bar, pizza place, Mexican restaurant, American restaurant, coffee shop.

The location of these neighborhoods, the high median income, high population density, and social venues suggests that this cluster include the trendy urban areas.

Cluster #1 contains 7 neighborhoods:

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

See Appendix B for details on each.

4.3 Cluster #3:

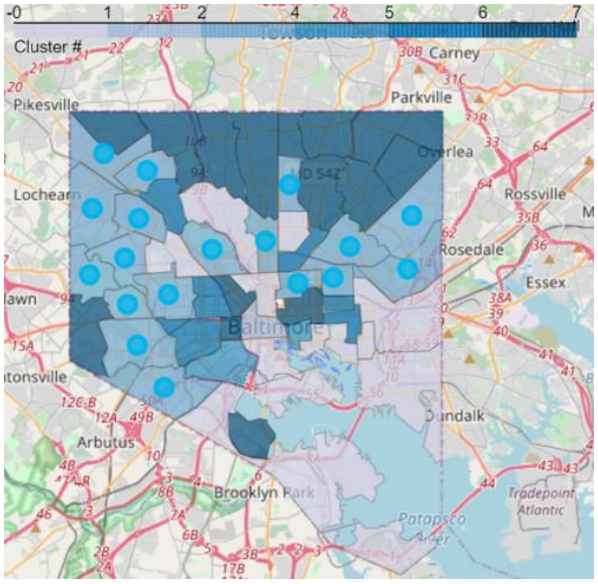

Figure: Cluster #3

Cluster #3 is moderate all around. It has the 'moderate' crime, 'moderate' median income, 'moderate' population density, and the mean average median household size corresponds to 'mostly couples and small families'. A 'moderate' percentage of the population is in the professional age range (25-64). The neighborhoods in this cluster are not as central as cluster #1. Some are closer to the city center than cluster #6. The most common venues are a mix of social and practical: fried chicken joint, park, convenience store, American restaurant, discount store.

The location of these neighborhoods, the moderate median income, moderate population density, and venues suggests that this cluster include is a transition area between more urban and more suburban areas. The moderate median income and the specific common venues (discount store, convenience store) suggest this cluster is possibly less expensive than clusters #1 and #6.

Cluster #3 contains 18 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 Appendix B for details on each.

5. Discussion

5.1 Recommendation:

It is recommended that the client consider the neighborhoods in clusters #6, #1, and #3 as she begins her house hunt. (see Appendix A and B for details.) The analysis suggests that:

- Cluster #6 neighborhoods are more suburban in character with very low crime rate.
- Cluster #1 neighborhoods are upscale urban area with social venues, however a moderate crime rate.
- Cluster #3 neighborhoods are transition regions between the urban and suburban areas. They are likely more affordable than Cluster #1 and Cluster #6 but have moderate crime.

5.2: Suggested Improvements:

This analysis could be improved with more current data, more relevant data, and more accurate data.

More Current Data: The crime and census data used was from 2015 and is 5 years old. More current information would improve the accuracy of the analysis. However, many of the relevant features are unlikely to change dramatically over this time frame.

More Relevant Data: The client was interested in neighborhoods with 'like professionals'. Additional details on the professional make up of each neighborhood could have improved the analysis. Specifically, average years of education and the types of professions for the population.

More Accurate Data: As noted in 3.1.3 the Foursquares venue data collected for this project was rather sparse in some neighborhoods. Consultation with a domain expert confirmed that key venues were missing. More accurate venue information may have resolved the challenges encountered when trying to cluster. Additionally, more accurate venue information would be informative to the client as she reviews the character of the clusters and the neighborhoods.

6. Conclusion:

This analysis successfully characterized Baltimore's neighborhoods in terms of their crime rates, age of the residents, population density, median household income, and venues. The neighborhoods were then clustered based on those factors.

An actionable recommendation was made to the client: the analysis identified three clusters for the client to consider and described the character of each. For each neighborhood in the recommended clusters a detailed report of the character of the neighborhood was provided to the client.

The authors wish Jane New Resident the best of luck in her house hunt!

Appendix A: Cluster Characteristics:

The characteristics of each cluster when clustering on solely the crime and census data follows. Cluster numbers are arbitrary.

Cluster #0:

Summary: High crime, moderate income, moderate poverty, low population density
Recommend: No – high crime.

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, mean average household size is 2.8 people

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, mean average household size is 2.1 people

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, mean average household size is 1.4 people

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 balance 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, mean average household size is 2.6 people

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, mean average household size is 2.4 people

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, mean average household size is 3.0 people

Top Venue:

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, but more 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, mean average household size is 2.4 people

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, mean average household size is 2.4 people

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

Appendix B: Recommended Neighborhoods

A total of 36 neighborhoods are recommended for consideration.

Cluster #6:

Greater Roland Park/Poplar Hill:

Cluster: #6
Crime:

- Crime: very low, 27.9 per 1000 people
- Violent Crime: very low, 4.2 per 1000 people
- Property Crime: very low, 23.5 per 1000 people

People:

- Median Income: very high, 105555.3 dollars per household
- Poverty: very low, percentage of families in poverty is 3.8
- Population Density: low, population density is 5.2 people per acre
- 25-64 age group: low, percentage population is 49.9
- Household Size: mostly couples, average household size is 2.0 people

Top Venues:

- Bank
- Flower Shop
- Pharmacy
- Grocery Store
- Coffee Shop
- Yoga Studio
- Food Truck

Loch Raven:

Cluster: #6
Crime:

- Crime: very low, 39.4 per 1000 people
- Violent Crime: low, 8.9 per 1000 people
- Property Crime: very low, 29.6 per 1000 people

People:

- Median Income: moderate, 47461.9 dollars per household
- Poverty: moderate, percentage of families in poverty is 12.1
- Population Density: moderate, population density is 17.0 people per acre
- 25-64 age group: moderate, percentage population is 52.8
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- Bank
- Dry Cleaner

- Park
- Gym / Fitness Center
- Auto Garage
- Liquor Store
- Bus Stop

Northwood:

Cluster: #6

Crime:

- Crime: low, 42.3 per 1000 people
- Violent Crime: low, 10.5 per 1000 people
- Property Crime: very low, 30.6 per 1000 people

People:

- Median Income: high, 53974.3 dollars per household
- Poverty: low, percentage of families in poverty is 8.0
- Population Density: moderate, population density is 11.3 people per acre
- 25-64 age group: very low, percentage population is 42.7
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- Food Service
- American Restaurant
- Pharmacy
- Pizza Place
- Yoga Studio
- Fast Food Restaurant
- Food Truck

Hamilton:

Cluster: #6

Crime:

- Crime: low, 47.5 per 1000 people
- Violent Crime: low, 9.5 per 1000 people
- Property Crime: low, 37.1 per 1000 people

People:

- Median Income: high, 63600.2 dollars per household
- Poverty: low, percentage of families in poverty is 8.7
- Population Density: moderate, population density is 12.8 people per acre
- 25-64 age group: high, percentage population is 57.8
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- Travel & Transport

- Home Service
- Event Service
- Farm
- Food Truck
- Food Service
- Food Court

Harford/Echodale:

Cluster: #6

Crime:

- Crime: low, 45.7 per 1000 people
- Violent Crime: low, 10.0 per 1000 people
- Property Crime: low, 34.5 per 1000 people

People:

- Median Income: high, 52708.1 dollars per household
- Poverty: low, percentage of families in poverty is 9.3
- Population Density: low, population density is 10.7 people per acre
- 25-64 age group: high, percentage population is 60.1
- Household Size: mostly couples, average household size is 2.4 people

Top Venues:

- Pharmacy
- Italian Restaurant
- Martial Arts Dojo
- BBQ Joint
- Thai Restaurant
- Candy Store
- Gas Station

Lauraville:

Cluster: #6

Crime:

- Crime: low, 41.9 per 1000 people
- Violent Crime: low, 8.1 per 1000 people
- Property Crime: low, 33.2 per 1000 people

People:

- Median Income: high, 66010.9 dollars per household
- Poverty: low, percentage of families in poverty is 9.4
- Population Density: low, population density is 10.7 people per acre
- 25-64 age group: high, percentage population is 61.0
- Household Size: couples & small families, average household size is 2.7 people

Top Venues:

- Coffee Shop
- Farmers Market
- Vegetarian / Vegan Restaurant
- Bus Stop
- Bookstore
- Flower Shop
- Grocery Store

Beechfield/Ten Hills/West Hills:

Cluster: #6

Crime:

- Crime: very low, 37.3 per 1000 people
- Violent Crime: low, 7.9 per 1000 people
- Property Crime: very low, 28.9 per 1000 people

People:

- Median Income: high, 51537.6 dollars per household
- Poverty: moderate, percentage of families in poverty is 11.2
- Population Density: moderate, population density is 12.2 people per acre
- 25-64 age group: moderate, percentage population is 56.4
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- ATM
- Residential Building (Apartment / Condo)
- Pool
- Restaurant
- Discount Store
- Food Service
- Food Court

Cross-Country/Cheswolde:

Cluster: #6

Crime:

- Crime: very low, 23.7 per 1000 people
- Violent Crime: very low, 0.9 per 1000 people
- Property Crime: very low, 22.5 per 1000 people

People:

- Median Income: high, 55964.3 dollars per household
- Poverty: low, percentage of families in poverty is 8.1
- Population Density: moderate, population density is 16.0 people per acre
- 25-64 age group: very low, percentage population is 45.5
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- River
- Yoga Studio
- Farm
- Food Truck
- Food Service
- Food Court
- Food & Drink Shop

Mount Washington/Coldspring:

Cluster: #6

Crime:

- Crime: low, 41.0 per 1000 people
- Violent Crime: very low, 4.6 per 1000 people
- Property Crime: low, 35.6 per 1000 people

People:

- Median Income: high, 77390.2 dollars per household
- Poverty: very low, percentage of families in poverty is 5.1
- Population Density: very low, population density is 3.6 people per acre
- 25-64 age group: high, percentage population is 57.4
- Household Size: mostly couples, average household size is 1.9 people

Top Venues:

- Park
- Café
- Mini Golf
- Yoga Studio
- Fried Chicken Joint
- Food Truck
- Food Service

North Baltimore/Guilford/Homeland:

Cluster: #6

Crime:

- Crime: very low, 38.3 per 1000 people
- Violent Crime: very low, 4.5 per 1000 people
- Property Crime: low, 33.6 per 1000 people

People:

- Median Income: very high, 78174.8 dollars per household
- Poverty: very low, percentage of families in poverty is 5.8
- Population Density: low, population density is 8.4 people per acre
- 25-64 age group: very low, percentage population is 46.1

- Household Size: mostly couples, average household size is 2.3 people

Top Venues:

- Wings Joint
- Social Club
- Gym
- Yoga Studio
- Farmers Market
- Food Truck
- Food Service

Chinquapin Park/Belvedere:

Cluster: #6

Crime:

- Crime: moderate, 52.5 per 1000 people
- Violent Crime: low, 7.6 per 1000 people
- Property Crime: moderate, 44.7 per 1000 people

People:

- Median Income: moderate, 45454.5 dollars per household
- Poverty: low, percentage of families in poverty is 8.9
- Population Density: moderate, population density is 16.9 people per acre
- 25-64 age group: moderate, percentage population is 54.6
- Household Size: mostly couples, average household size is 2.4 people

Top Venues:

- Juice Bar
- Gift Shop
- Dessert Shop
- Ramen Restaurant
- Deli / Bodega
- Sandwich Place
- Eye Doctor

Cluster #3:

Cedonia/Frankford:

Cluster: #3

Crime:

- Crime: moderate, 53.6 per 1000 people
- Violent Crime: moderate, 11.7 per 1000 people
- Property Crime: moderate, 41.5 per 1000 people

People:

- Median Income: moderate, 39300.4 dollars per household

- Poverty: moderate, percentage of families in poverty is 16.9
- Population Density: moderate, population density is 14.5 people per acre
- 25-64 age group: high, percentage population is 59.0
- Household Size: couples & small families, average household size is 2.5 people

Top Venues:

- Garden
- Indian Restaurant
- Chinese Restaurant
- Yoga Studio
- Fast Food Restaurant
- Food Truck
- Food Service

Glen-Fallstaff:

Cluster: #3

Crime:

- Crime: moderate, 55.0 per 1000 people
- Violent Crime: moderate, 13.5 per 1000 people
- Property Crime: moderate, 40.8 per 1000 people

People:

- Median Income: moderate, 39690.8 dollars per household
- Poverty: moderate, percentage of families in poverty is 19.8
- Population Density: moderate, population density is 16.0 people per acre
- 25-64 age group: low, percentage population is 50.2
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- American Restaurant
- Rental Car Location
- Liquor Store
- Yoga Studio
- Farmers Market
- French Restaurant
- Food Truck

Forest Park/Walbrook:

Cluster: #3

Crime:

- Crime: moderate, 51.7 per 1000 people
- Violent Crime: moderate, 12.0 per 1000 people
- Property Crime: moderate, 39.5 per 1000 people

People:

- Median Income: moderate, 39204.8 dollars per household
- Poverty: moderate, percentage of families in poverty is 20.3
- Population Density: low, population density is 10.6 people per acre
- 25-64 age group: moderate, percentage population is 54.5
- Household Size: mostly couples, average household size is 2.4 people

Top Venues:

- Construction & Landscaping
- Yoga Studio
- Farmers Market
- French Restaurant
- Food Truck
- Food Service
- Food Court

Edmondson Village:

Cluster: #3

Crime:

- Crime: low, 42.2 per 1000 people
- Violent Crime: low, 9.6 per 1000 people
- Property Crime: low, 32.0 per 1000 people

People:

- Median Income: moderate, 38042.4 dollars per household
- Poverty: moderate, percentage of families in poverty is 16.0
- Population Density: moderate, population density is 15.7 people per acre
- 25-64 age group: low, percentage population is 51.7
- Household Size: couples & small families, average household size is 2.9 people

Top Venues:

- Convenience Store
- Auto Garage
- Yoga Studio
- Farmers Market
- French Restaurant
- Food Truck
- Food Service

Greenmount East:

Cluster: #3

Crime:

- Crime: moderate, 76.0 per 1000 people
- Violent Crime: high, 22.6 per 1000 people
- Property Crime: moderate, 52.4 per 1000 people

People:

- Median Income: very low, 23625.8 dollars per household
- Poverty: high, percentage of families in poverty is 25.8
- Population Density: moderate, population density is 17.2 people per acre
- 25-64 age group: moderate, percentage population is 53.1
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- Farmers Market
- Drugstore
- Grocery Store
- Community Center
- Flower Shop
- Convenience Store
- Basketball Court

Greater Charles Village/Barclay:

Cluster: #3

Crime:

- Crime: moderate, 71.1 per 1000 people
- Violent Crime: high, 21.4 per 1000 people
- Property Crime: moderate, 49.2 per 1000 people

People:

- Median Income: low, 32241.1 dollars per household
- Poverty: moderate, percentage of families in poverty is 23.8
- Population Density: high, population density is 22.3 people per acre
- 25-64 age group: low, percentage population is 52.1
- Household Size: mostly couples, average household size is 2.1 people

Top Venues:

- Pizza Place
- Café
- Convenience Store
- Brewery
- Chinese Restaurant
- Coffee Shop
- Sculpture Garden

Dickeyville/Franklintown:

Cluster: #3

Crime:

- Crime: very low, 31.9 per 1000 people
- Violent Crime: low, 7.1 per 1000 people

- Property Crime: very low, 24.6 per 1000 people

People:

- Median Income: moderate, 35773.0 dollars per household
- Poverty: moderate, percentage of families in poverty is 21.2
- Population Density: very low, population density is 4.6 people per acre
- 25-64 age group: moderate, percentage population is 53.7
- Household Size: mostly couples, average household size is 2.2 people

Top Venues:

- Garden
- Waterfall
- Home Service
- Yoga Studio
- Farmers Market
- Food Truck
- Food Service

Clifton-Berea:

Cluster: #3

Crime:

- Crime: moderate, 66.5 per 1000 people
- Violent Crime: high, 23.7 per 1000 people
- Property Crime: moderate, 41.2 per 1000 people

People:

- Median Income: low, 29364.2 dollars per household
- Poverty: moderate, percentage of families in poverty is 23.6
- Population Density: moderate, population density is 16.1 people per acre
- 25-64 age group: moderate, percentage population is 53.1
- Household Size: couples & small families, average household size is 2.7 people

Top Venues:

- Wings Joint
- Bus Stop
- Food
- Park
- Public Art
- Yoga Studio
- Farmers Market

Claremont/Armistead:

Cluster: #3

Crime:

- Crime: low, 48.1 per 1000 people

- Violent Crime: moderate, 11.5 per 1000 people
- Property Crime: low, 35.8 per 1000 people

People:

- Median Income: low, 33179.5 dollars per household
- Poverty: moderate, percentage of families in poverty is 21.3
- Population Density: low, population density is 5.9 people per acre
- 25-64 age group: moderate, percentage population is 54.5
- Household Size: mostly couples, average household size is 2.4 people

Top Venues:

- Gas Station
- Hotel
- Fast Food Restaurant
- Yoga Studio
- Farmers Market
- Food Truck
- Food Service

Dorchester/Ashburton:

Cluster: #3

Crime:

- Crime: low, 50.3 per 1000 people
- Violent Crime: moderate, 12.6 per 1000 people
- Property Crime: low, 36.7 per 1000 people

People:

- Median Income: moderate, 38682.9 dollars per household
- Poverty: moderate, percentage of families in poverty is 18.8
- Population Density: moderate, population density is 15.0 people per acre
- 25-64 age group: low, percentage population is 51.9
- Household Size: couples & small families, average household size is 2.7 people

Top Venues:

- Fried Chicken Joint
- Food
- Grocery Store
- Chinese Restaurant
- Frozen Yogurt Shop
- French Restaurant
- Food Truck

Belair-Edison:

Cluster: #3

Crime:

- Crime: moderate, 52.5 per 1000 people
- Violent Crime: moderate, 11.8 per 1000 people
- Property Crime: moderate, 39.4 per 1000 people

People:

- Median Income: moderate, 38174.0 dollars per household
- Poverty: moderate, percentage of families in poverty is 18.6
- Population Density: moderate, population density is 15.8 people per acre
- 25-64 age group: moderate, percentage population is 54.7
- Household Size: couples & small families, average household size is 2.8 people

Top Venues:

- Fried Chicken Joint
- Discount Store
- ATM
- Chinese Restaurant
- Grocery Store
- Food & Drink Shop
- Pharmacy

Greater Govans:

Cluster: #3

Crime:

- Crime: low, 49.0 per 1000 people
- Violent Crime: moderate, 11.3 per 1000 people
- Property Crime: low, 37.3 per 1000 people

People:

- Median Income: moderate, 37457.3 dollars per household
- Poverty: moderate, percentage of families in poverty is 21.3
- Population Density: high, population density is 20.9 people per acre
- 25-64 age group: low, percentage population is 52.1
- Household Size: couples & small families, average household size is 2.7 people

Top Venues:

- Gym / Fitness Center
- Home Service
- Fried Chicken Joint
- Coffee Shop
- Weight Loss Center
- Laundromat
- Snack Place

Greater Rosemont:

Cluster: #3

Crime:

- Crime: moderate, 63.2 per 1000 people
- Violent Crime: moderate, 15.5 per 1000 people
- Property Crime: moderate, 46.8 per 1000 people

People:

- Median Income: low, 31759.1 dollars per household
- Poverty: moderate, percentage of families in poverty is 21.3
- Population Density: moderate, population density is 16.2 people per acre
- 25-64 age group: moderate, percentage population is 52.6
- Household Size: couples & small families, average household size is 2.7 people

Top Venues:

- Intersection
- Lounge
- Deli / Bodega
- Indian Restaurant
- Ice Cream Shop
- Food Truck
- Food Service

Howard Park/West Arlington:

Cluster: #3

Crime:

- Crime: low, 48.5 per 1000 people
- Violent Crime: low, 10.9 per 1000 people
- Property Crime: low, 37.1 per 1000 people

People:

- Median Income: moderate, 41821.9 dollars per household
- Poverty: moderate, percentage of families in poverty is 19.5
- Population Density: low, population density is 7.1 people per acre
- 25-64 age group: low, percentage population is 49.2
- Household Size: couples & small families, average household size is 2.6 people

Top Venues:

- Discount Store
- Bus Stop
- Lawyer
- Yoga Studio
- Fast Food Restaurant
- French Restaurant
- Food Truck

Morrell Park/Violetville:

Cluster: #3

Crime:

- Crime: moderate, 67.0 per 1000 people
- Violent Crime: moderate, 14.1 per 1000 people
- Property Crime: moderate, 52.5 per 1000 people

People:

- Median Income: moderate, 35687.4 dollars per household
- Poverty: moderate, percentage of families in poverty is 12.5
- Population Density: very low, population density is 5.0 people per acre
- 25-64 age group: moderate, percentage population is 56.7
- Household Size: mostly couples, average household size is 2.3 people

Top Venues:

- Construction & Landscaping
- Hobby Shop
- Yoga Studio
- Farmers Market
- French Restaurant
- Food Truck
- Food Service

Penn North/Reservoir Hill:

Cluster: #3

Crime:

- Crime: moderate, 64.2 per 1000 people
- Violent Crime: moderate, 15.8 per 1000 people
- Property Crime: moderate, 47.0 per 1000 people

People:

- Median Income: low, 34879.3 dollars per household
- Poverty: high, percentage of families in poverty is 29.3
- Population Density: moderate, population density is 11.5 people per acre
- 25-64 age group: moderate, percentage population is 55.6
- Household Size: couples & small families, average household size is 2.5 people

Top Venues:

- Park
- Botanical Garden
- Basketball Court
- Farmers Market
- Baseball Field
- Track
- Yoga Studio

Pimlico/Arlington/Hilltop:

Cluster: #3

Crime:

- Crime: moderate, 61.6 per 1000 people
- Violent Crime: high, 20.9 per 1000 people
- Property Crime: moderate, 39.4 per 1000 people

People:

- Median Income: low, 33139.2 dollars per household
- Poverty: moderate, percentage of families in poverty is 17.4
- Population Density: moderate, population density is 16.8 people per acre
- 25-64 age group: moderate, percentage population is 54.3
- Household Size: couples & small families, average household size is 2.9 people

Top Venues:

- Fried Chicken Joint
- Caribbean Restaurant
- Deli / Bodega
- Clothing Store
- Racetrack
- Storage Facility
- Convenience Store

Allendale/Irvington/S. Hilton:

Cluster: #3

Crime:

- Crime: moderate, 59.3 per 1000 people
- Violent Crime: moderate, 15.0 per 1000 people
- Property Crime: moderate, 43.8 per 1000 people

People:

- Median Income: moderate, 36701.9 dollars per household
- Poverty: high, percentage of families in poverty is 24.1
- Population Density: moderate, population density is 10.9 people per acre
- 25-64 age group: low, percentage population is 51.0
- Household Size: couples & small families, average household size is 2.7 people

Top Venues:

- Fried Chicken Joint
- Karaoke Bar
- Discount Store
- Cosmetics Shop
- Bus Stop
- Flower Shop
- Thrift / Vintage Store

Cluster #1:

Inner Harbor/Federal Hill:

Cluster: #1

Crime:

- Crime: moderate, 65.9 per 1000 people
- Violent Crime: moderate, 11.0 per 1000 people
- Property Crime: moderate, 54.5 per 1000 people

People:

- Median Income: very high, 88465.2 dollars per household
- Poverty: low, percentage of families in poverty is 7.5
- Population Density: high, population density is 23.6 people per acre
- 25-64 age group: very high, percentage population is 67.5
- Household Size: mostly couples, average household size is 2.0 people

Top Venues:

- Bar
- Pizza Place
- Mexican Restaurant
- Sushi Restaurant
- Café
- Mediterranean Restaurant
- Pub

Fells Point:

Cluster: #1

Crime:

- Crime: high, 80.2 per 1000 people
- Violent Crime: moderate, 16.6 per 1000 people
- Property Crime: high, 62.6 per 1000 people

People:

- Median Income: very high, 82263.2 dollars per household
- Poverty: low, percentage of families in poverty is 6.2
- Population Density: very high, population density is 33.8 people per acre
- 25-64 age group: very high, percentage population is 77.5
- Household Size: mostly couples, average household size is 2.1 people

Top Venues:

- Bar
- Seafood Restaurant
- Mexican Restaurant
- American Restaurant
- Latin American Restaurant
- Cocktail Bar

- Sushi Restaurant

Highlandtown:

Cluster: #1

Crime:

- Crime: high, 94.9 per 1000 people
- Violent Crime: moderate, 17.4 per 1000 people
- Property Crime: very high, 76.8 per 1000 people

People:

- Median Income: high, 68702.2 dollars per household
- Poverty: low, percentage of families in poverty is 9.6
- Population Density: very high, population density is 27.4 people per acre
- 25-64 age group: very high, percentage population is 65.6
- Household Size: mostly couples, average household size is 2.3 people

Top Venues:

- Dive Bar
- New American Restaurant
- Pizza Place
- Hot Dog Joint
- Mexican Restaurant
- Supermarket
- Pharmacy

Canton:

Cluster: #1

Crime:

- Crime: moderate, 51.2 per 1000 people
- Violent Crime: very low, 4.6 per 1000 people
- Property Crime: moderate, 46.5 per 1000 people

People:

- Median Income: very high, 95362.4 dollars per household
- Poverty: very low, percentage of families in poverty is 3.0
- Population Density: high, population density is 22.2 people per acre
- 25-64 age group: very high, percentage population is 74.0
- Household Size: mostly couples, average household size is 2.0 people

Top Venues:

- Bar
- American Restaurant
- Gym / Fitness Center
- Pub
- Pizza Place

- Mexican Restaurant
- Ice Cream Shop

South Baltimore:

Cluster: #1

Crime:

- Crime: low, 41.8 per 1000 people
- Violent Crime: very low, 3.3 per 1000 people
- Property Crime: moderate, 38.1 per 1000 people

People:

- Median Income: very high, 97441.0 dollars per household
- Poverty: very low, percentage of families in poverty is 3.7
- Population Density: low, population density is 6.7 people per acre
- 25-64 age group: very high, percentage population is 74.7
- Household Size: mostly couples, average household size is 2.2 people

Top Venues:

- Cruise
- Pet Store
- Coffee Shop
- Bar
- Buffet
- Shopping Mall
- Brewery

Midtown:

Cluster: #1

Crime:

- Crime: high, 87.5 per 1000 people
- Violent Crime: moderate, 19.6 per 1000 people
- Property Crime: high, 67.0 per 1000 people

People:

- Median Income: moderate, 41170.5 dollars per household
- Poverty: low, percentage of families in poverty is 7.6
- Population Density: very high, population density is 28.5 people per acre
- 25-64 age group: high, percentage population is 61.9
- Household Size: mostly couples, average household size is 1.8 people

Top Venues:

- Pizza Place
- Deli / Bodega
- Sandwich Place
- Bar

- Arts & Crafts Store
- Sushi Restaurant
- Gay Bar

Medfield/Hampden/Woodberry/Remington:

Cluster: #1

Crime:

- Crime: moderate, 55.8 per 1000 people
- Violent Crime: low, 10.7 per 1000 people
- Property Crime: moderate, 44.7 per 1000 people

People:

- Median Income: high, 58644.2 dollars per household
- Poverty: low, percentage of families in poverty is 7.1
- Population Density: moderate, population density is 12.2 people per acre
- 25-64 age group: high, percentage population is 64.4
- Household Size: mostly couples, average household size is 2.2 people

Top Venues:

- American Restaurant
- Bar
- Pizza Place
- Food Truck
- Brewery
- Italian Restaurant
- Mexican Restaurant

[i] https://en.wikipedia.org/wiki/Baltimore