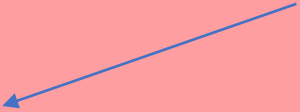
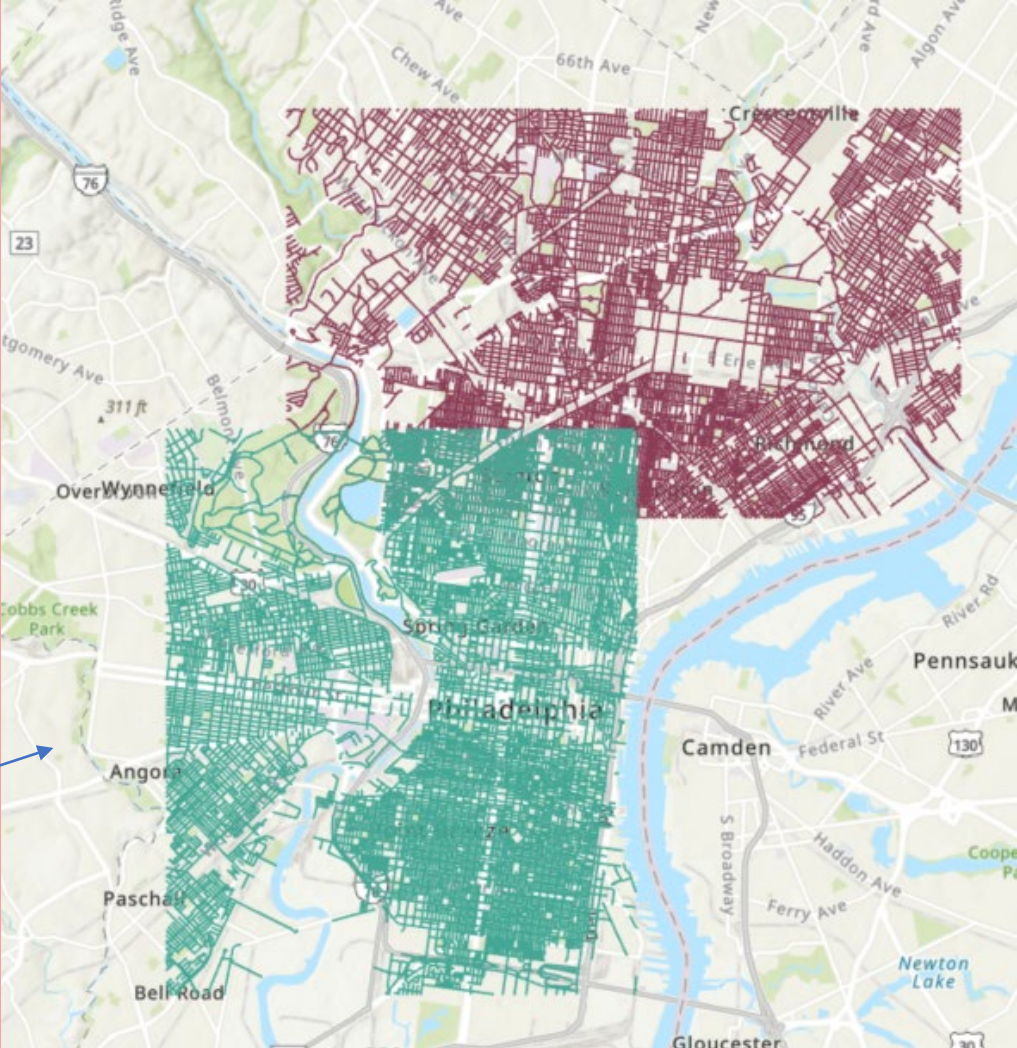


Finding the Likelihood of Gun Incidents in Central Philadelphia



Training on North Philadelphia



Testing on Central Philadelphia

Steps:

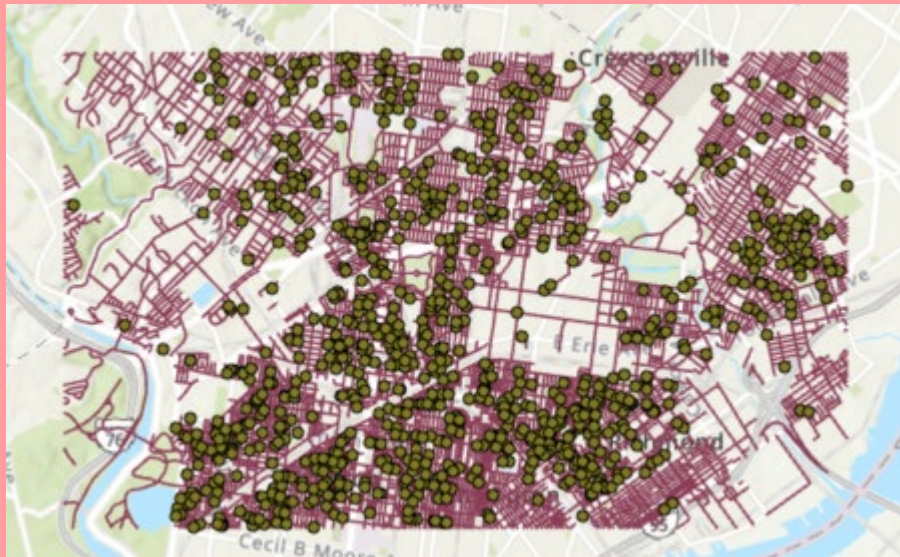
This problem is a predictive modeling problem. It uses unsupervised clustering to train assumptions on one area and then test on another. Thus, we will be using the steps of predictive modeling.

1. Identify the dependent variable. Here that variable is Gun Incidents. We take the *Kernel Density* of Gun Incidents to find the likelihood of shootings occurring across the area of study.
2. Engineer Features, or independent variables, to predict the dependent variable. We use *Kernel Density* for Population variables and *IDW-Interpolation* for Condition variables. These tools enable us to get a smooth range of these values across the region in raster form.
3. Standardize the features using *Slice* so they all have a similar range.
4. Use the *ISO Clustering tool* to input our features and find the model with the combination of features that best captures the gun incidents kernel density layer.
5. Now that we have our trained model, we must test it on the Central Philadelphia area. Repeat Steps 2 and 3 for the same features but in Central Philly.
6. Use *ISO Clustering tool* with the same combination of features in North Philly to predict for Central Philadelphia.
7. Change the symbology layers and colors to reflect the likelihood of gun incidents from Well Below Average to Well Above Average
8. Test if our clustering is accurate by overlaying our cluster layer on top of the intersection of gun incidents in North Philly to see if it captures the hot spots.

Identify the Dependent Variable: Gun Incidents in North Philly

*Before we start make sure to change the environment to North Philly

Use *Kernel Density* to find the density in where clusters of gun incidents are occurring. Set area to Square Feet as we want quite a small area to be covered. Once we have the layer, change Symbology to Stretch and Stretch Type to Standard Deviation.



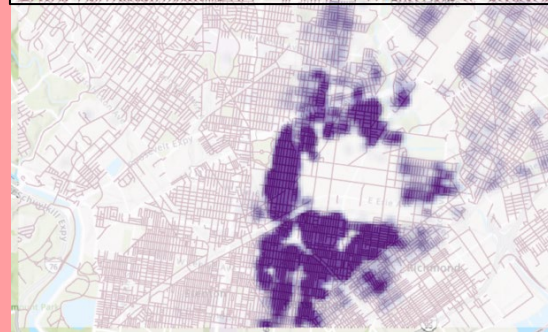
We will be comparing our ISO clustering in North Philly to this dependent variable.

Feature Engineering- Kernel Density

We use Kernel Density for feature engineering the population data because it consists of counts of people and not rates. Additionally, we set the area units to square feet as gun incidents are relatively localized occurrences and we want to be able to track them on a small granular level. We then change the symbology of each map to be in Stretch and the Stretch Type to be Standard Deviation.

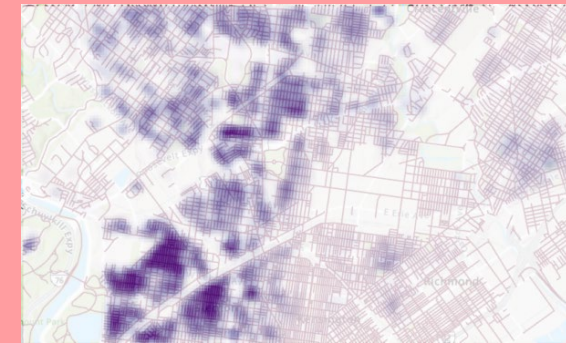
Maps of demographic groups (observationally it seems each occupy a separate space of North Philly)

Hispanic

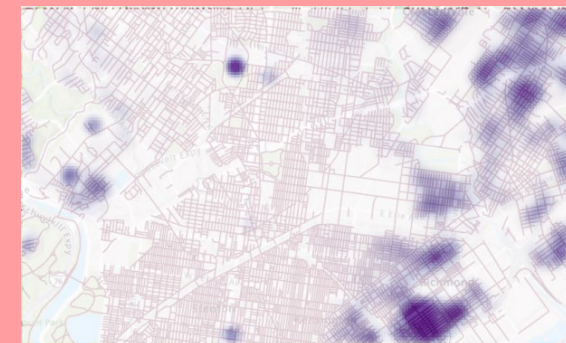
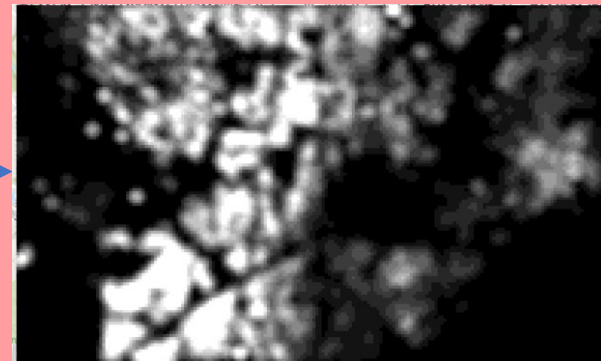


Scaling using Slice

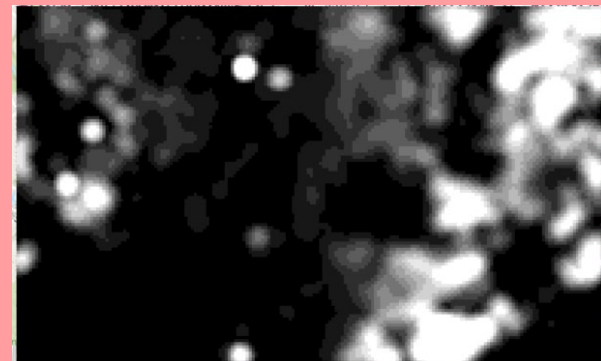
Then we slice the data into 20 classes using Natural Breaks, this helps to standardize the range across demographic groups



Black



White



Geoprocessing Kernel Density

Parameters Environments

Input point or polyline features: North_Populations

Population field: NHBLACK

Output raster: KD_Black

Output cell size: 93.8073887434002

Search radius: []

Area units: Square feet

Output cell values: Densities

Method: Planar

Input barrier features: []

Primary symbology

Stretch: []

Band: Band_1

Color scheme: []

Invert

Value: 0 0.00163279

Label: 0 0.00163279

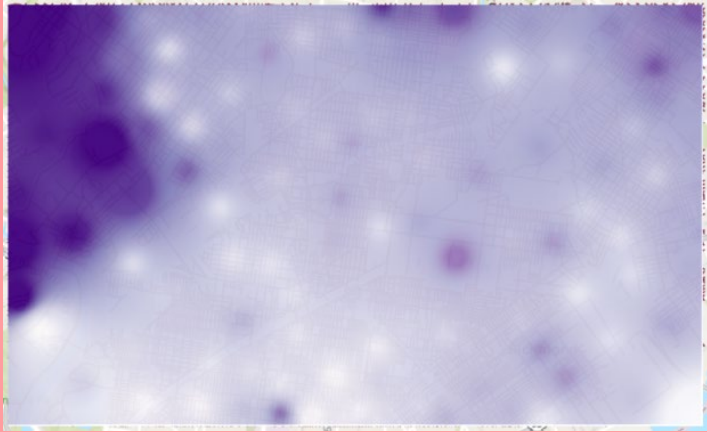
Stretch type: Standard Deviation

Number of standard deviations: 2

Gamma: 1.0

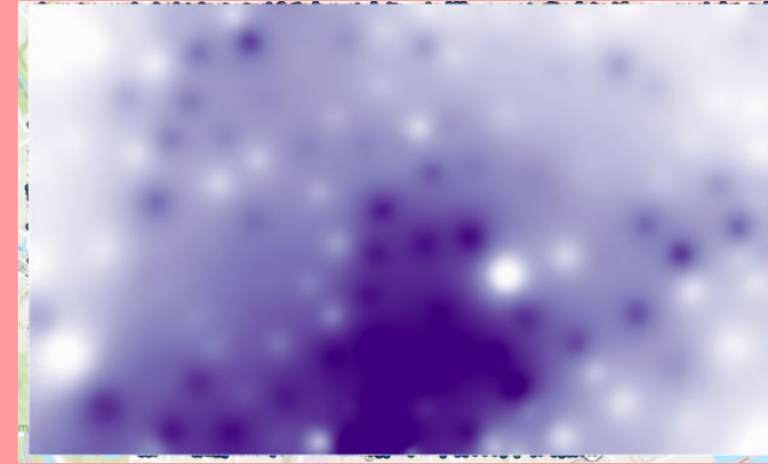
Feature Engineering: IDW- Interpolation

Median Monthly Rent



For our features that are rates or percentages instead of counts, we use interpolation to create a smoothed out range across the region. We use a fixed search radius set to 10000 square feet (roughly 1.8 miles) because again we want our distances to be concentrated on immediate areas. We then change the symbology of each map to be in Stretch and the Stretch Type to be Standard Deviation.

Percent of individuals living below poverty level



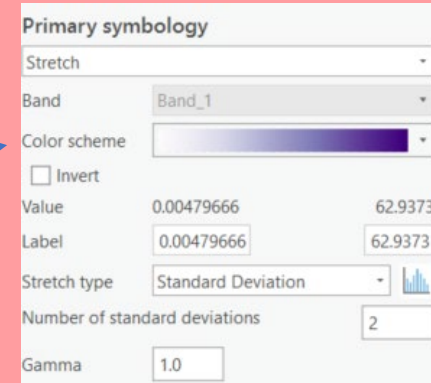
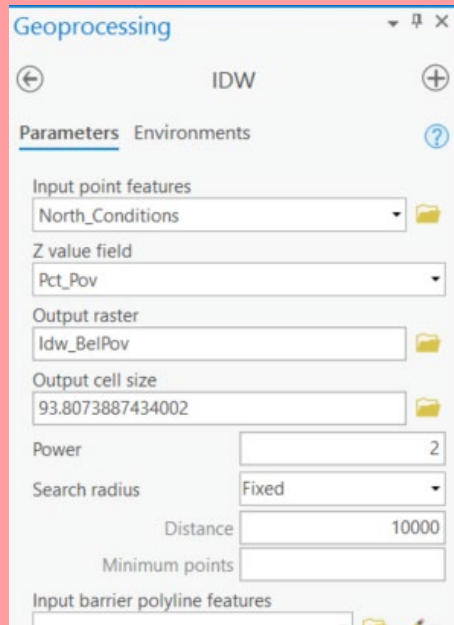
Seems these two features are correlated

Seems these two features are correlated

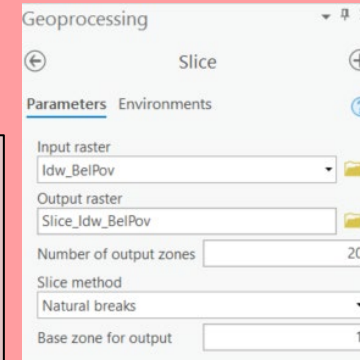
Percentage of people in college for at least two years



Percent of Individuals under the age of 5



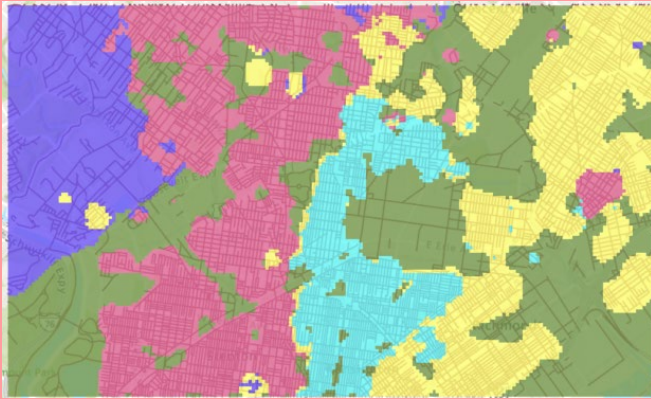
After getting these maps, we also slice them to standardize the range using 20 classes and Natural Breaks, these second maps are not shown here, but the step was taken.



Now we have to build the model!

We used the *ISO Cluster Unsupervised Classification Tool* and added in various combinations of our sliced features. We made sure to set the number of classes to 5 as that is the amount of clusters we want. We tried out three different combinations of features in order to get three different models to work with.

Version 1



Input raster bands

- Slice_College
- Slice_MedRent
- Slice_Hispanic
- Slice_Black
- Slice_White

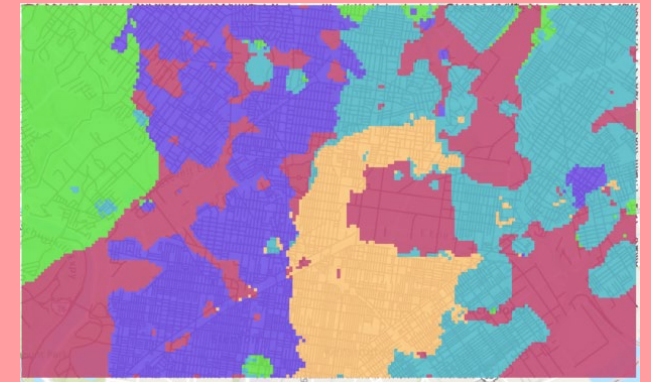
Version 2



Input raster bands

- Slice_Hispanic
- Slice_Black
- Slice_White
- Slice_College

Version 3



Input raster bands

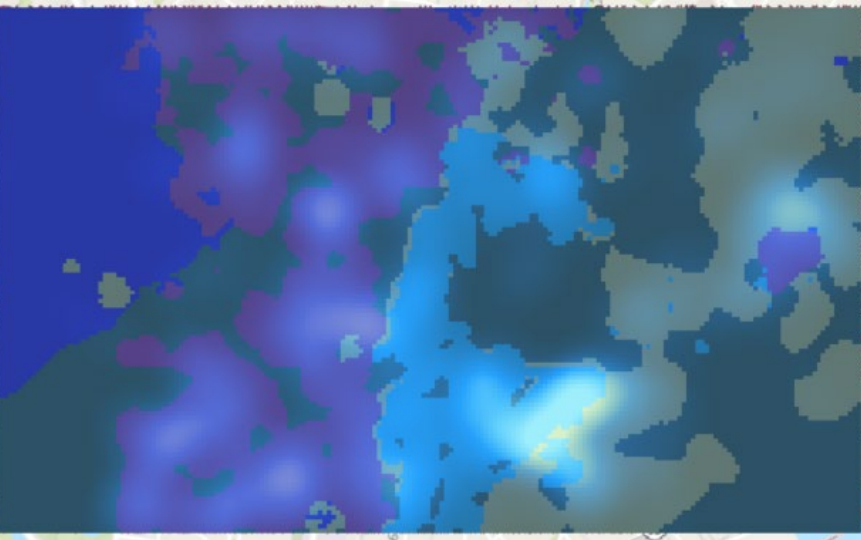
- Slice_Idw_BelSyr
- Slice_Idw_BelPov
- Slice_College
- Slice_MedRent
- Slice_Hispanic
- Slice_Black
- Slice_White

Combinations of Features

Testing the model on the “training data” of Gun incidents in North Philly

Now we test out each cluster version on the heatmap of gun incidents in Philly to see if it has captured the variation in incidents well. We test this by overlaying our cluster models on top of the layer we already made of gun incidents and changing the transparency.

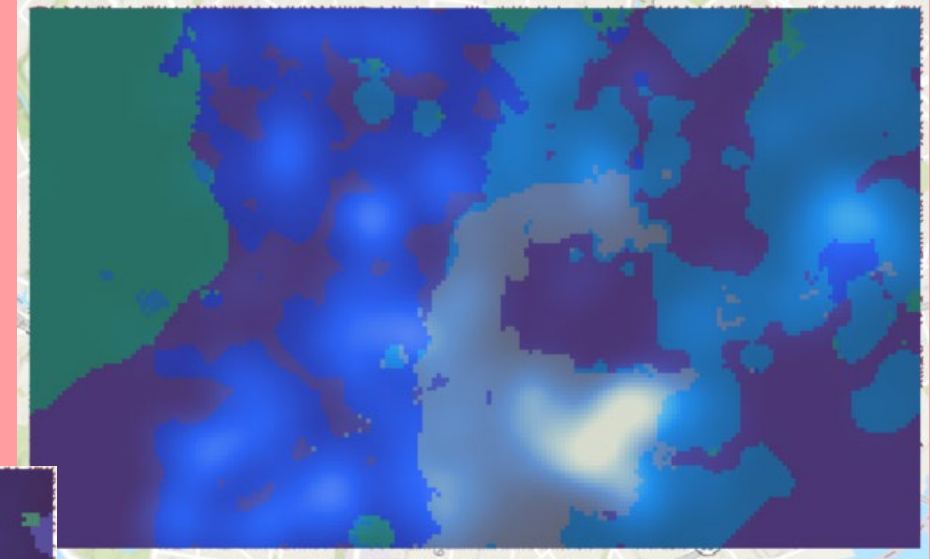
Version 1



Version 2



Version 3



Based on this, it seems version 3 does the best job of capturing the variation in gun incidents. Therefore we will use these same features to predict for gun incidents in Central Philly.

Testing our model on Central Philly

1. We must reset our environment to be on the Central Philly Roads layer and not North Philly.
2. Go through same steps as previously where we use the Kernel Density tool on the population data and the IDW-Interpolation Tool on the Conditions Data
3. Then we slice all 7 layers making 20 classes for each with method of Natural Breaks
4. Change the symbology for each to Stretch with Stretch Type Standard Deviation and inverted Bathymetry #3 color scheme

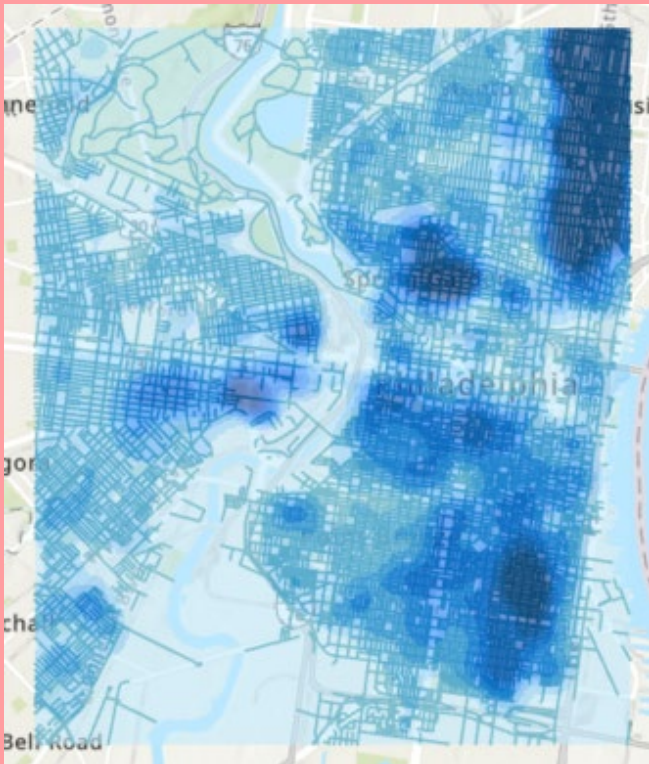
Central Philly Study
Region



Feature Engineering

Sliced Maps of Demographics Features once Kernel Density and Slicing has been completed. We can see here, like in the North Philly data, that the spatial pattern of race in Central Philly occupies quite different areas.

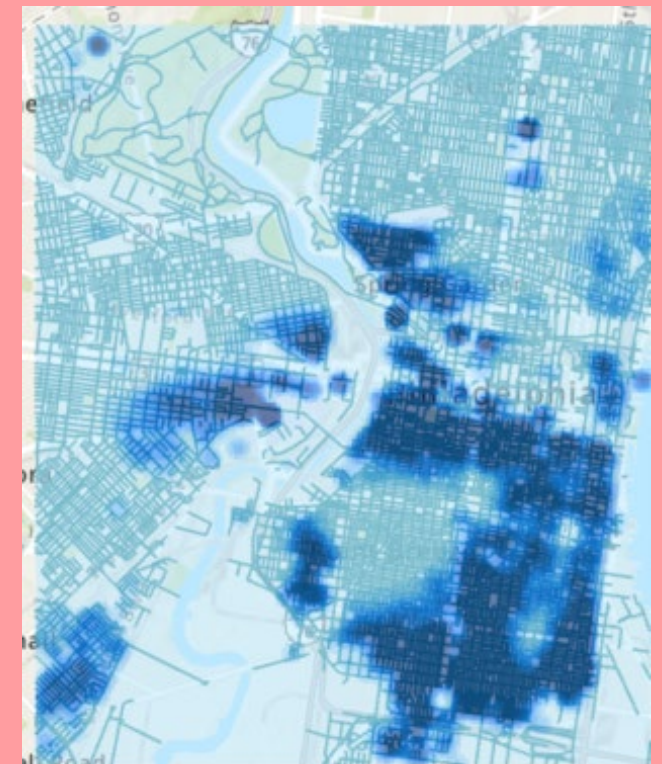
Hispanic



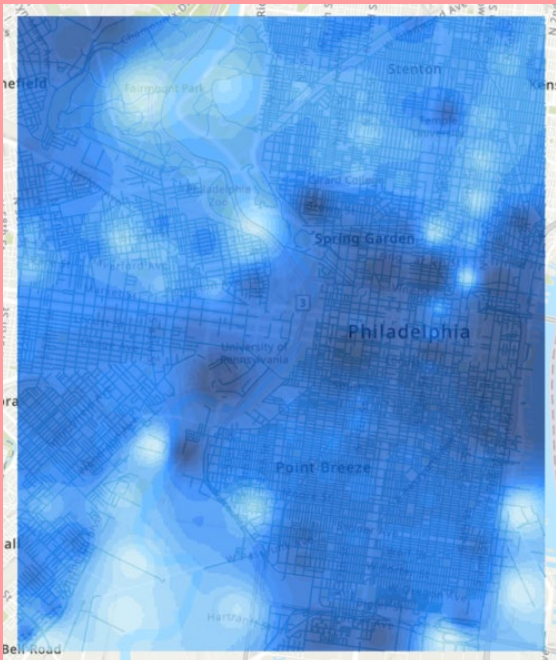
Black



White



Feature Engineering



Median Monthly Rent



Percent of individuals living below poverty level

Sliced Maps of Features Engineered:
Note that similarity in spatial patterns between maps in first column and maps in second. Additionally, note that dark area in the college and median monthly rent map is Center City which is the lighter colored area in the other two maps, showing the patterns are slightly the inverse of each other.

Dark Blue areas are higher density, Light Blue is lower density



Percentage of people in college for at least two years

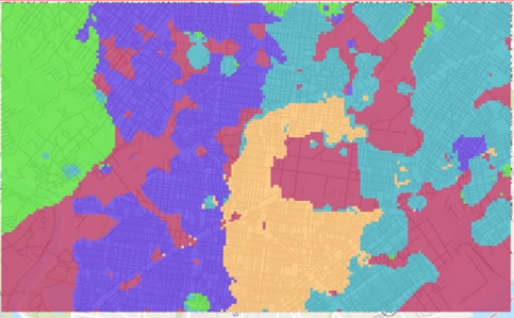
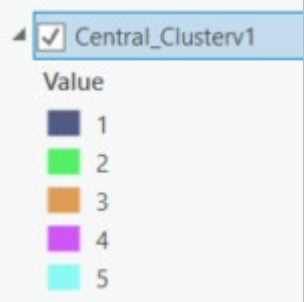
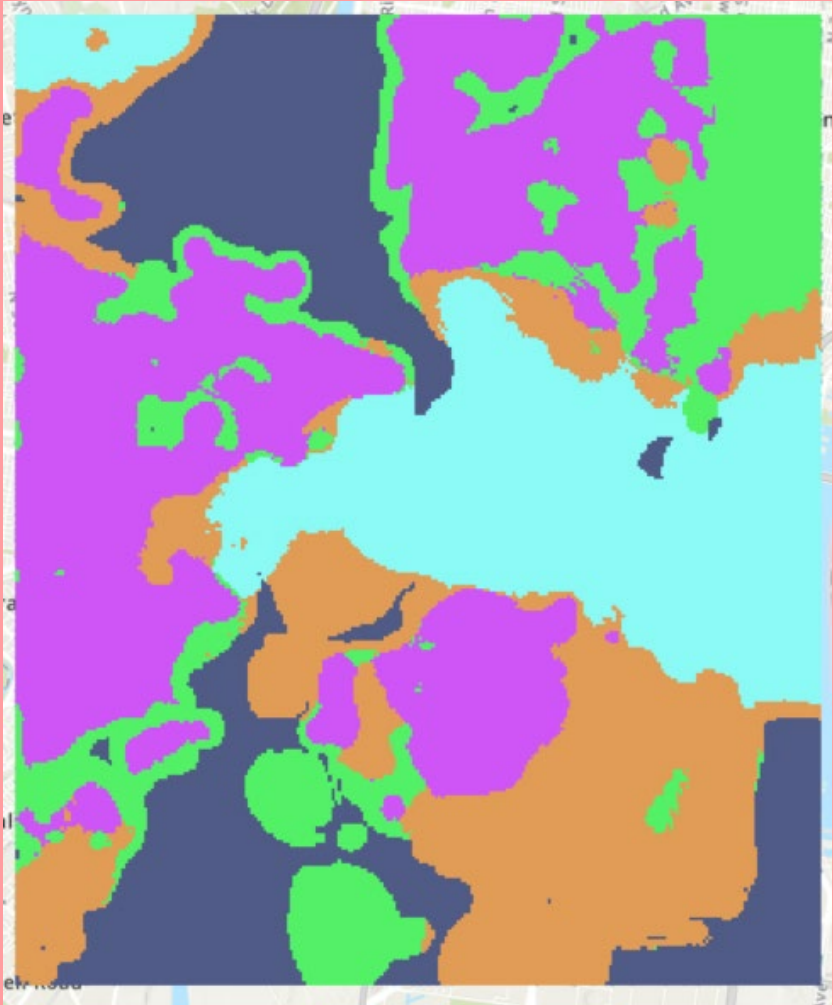
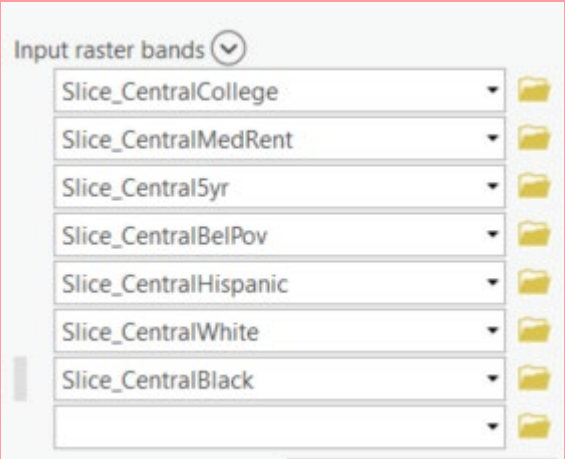


Percent of Individuals under the age of 5

Test the optimal features from North Philly Version 3 Model on Central Philly using ISO Clustering.

Output:

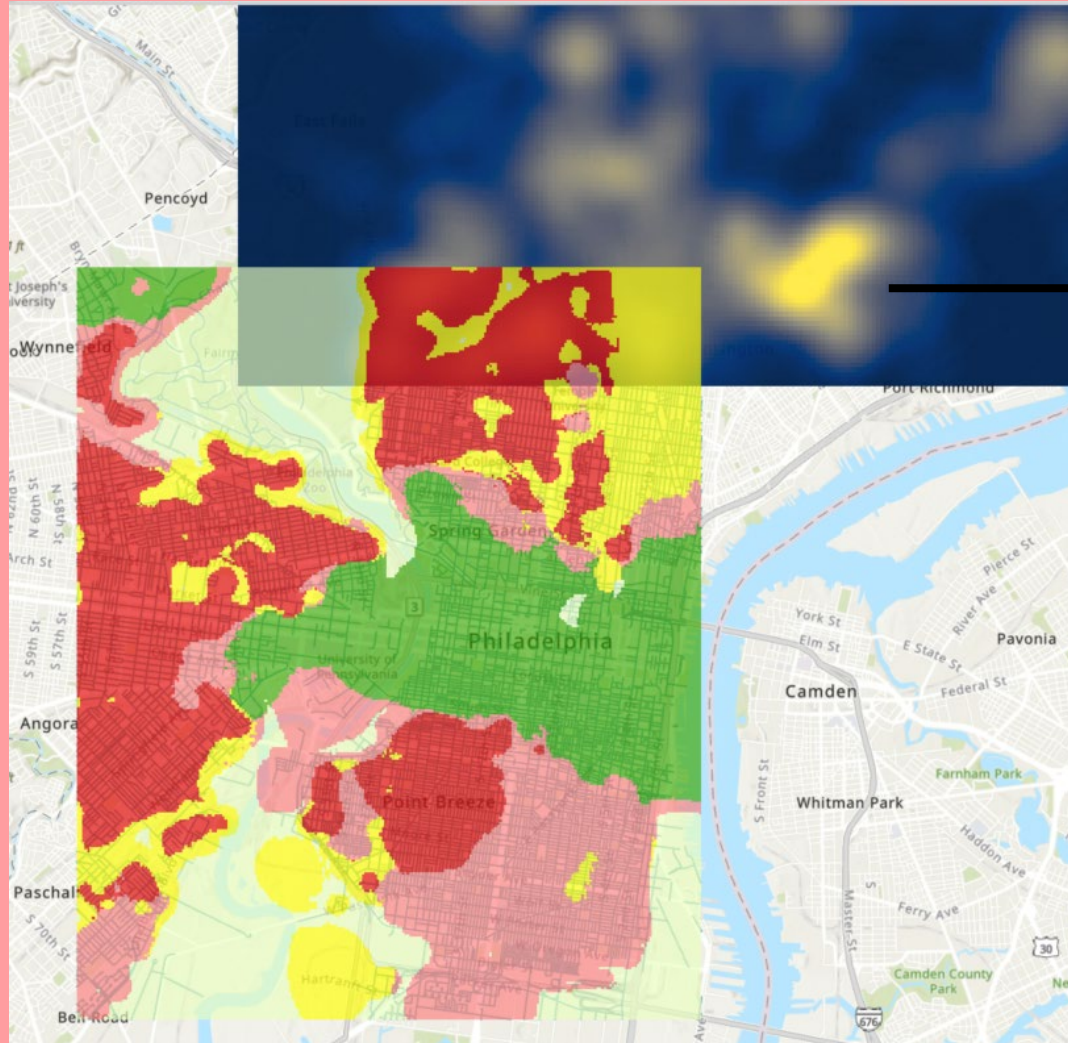
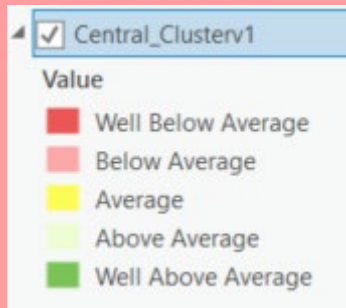
Features:



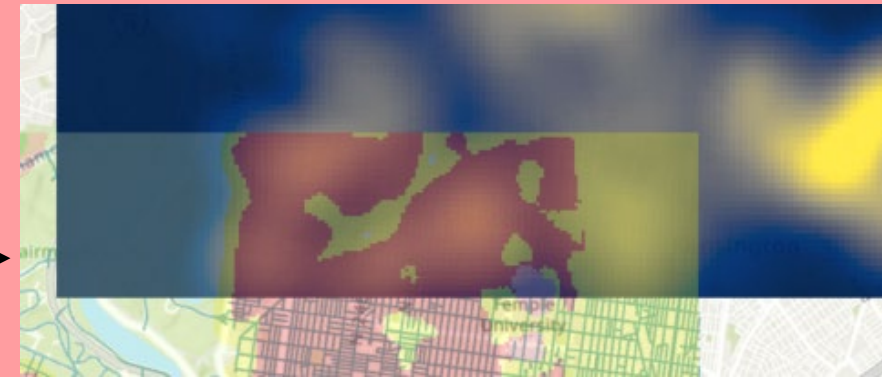
North Philly Version 3 Model for Reference

Validating the Model

Rename the clusters in the Symbology tab to go from Well Below Average to Well Above Average. Because these are clusters labelled arbitrarily from 1-5, we rename them based on the North Philly clustering pattern. We've also changed the colors in the Symbology tab to correspond with the colors that make sense visually to go from "bad" to "good".



Using the North Philly Gun Incidents layer, we can check to see if our clustering holds true. Based on this image, it seems we've captured some gun incidents in two different clusterings. I would say this means we were successful!



Final Map of Central Philly Likelihood of Gun Incidents

