Final Project Report

Jonathan Darius, Ynah Novilla, Ojasvi Godha, Adreyan Distor, and Kushal Tirupathi

Project Description

The goal of this project is to develop a classifier for violent crime rates in communities based on demographic, socioeconomic, and law enforcement-related features. The project utilizes the Random Forest Classifier and K-Means Clustering Algorithm to make predictions regarding the violent crime rate per capita. For this project, we are only considering crimes that are either murder, rape, robbery, or assault. This project aims to contribute valuable insights into the dynamics between community features and crime, fostering informed decision-making for community development and law enforcement initiatives.

Dataset

The dataset consists of 122 predictive, 5 non-predictive, and 1 goal attribute, encompassing a wide range of community characteristics, law enforcement metrics, and socioeconomic factors. This dataset encompasses the US as a whole, so the data will contain diverse data that will take geographic differences into account.

Data Collection and Data Cleaning

Our dataset has plenty of missing values entered in as "?" in various columns that were necessary in creating our model. Dealing with missing values is tricky, since simply turning them into an NA value will only hurt our model accuracy, and replacing each empty value with the mean of their respective features will create inconsistencies within each instance. This is because each instance represents a community, and it is critical that we maintain the relationships between the features and target variables for each community.

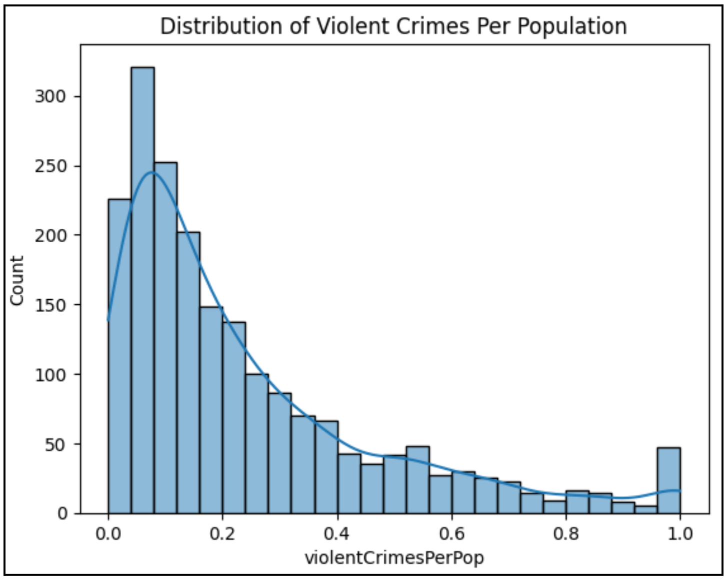
With this in mind, we performed iterative imputation on the missing values using the 'IterativeImputer' from the scikit-learn library. This imputation method creates a linear regression model that considers the relationships between each column, and it is applied iteratively to each column with missing values in the dataset.

Fortunately, all numeric data is already normalized into the decimal range 0.00-1.00. Though one thing we need to be mindful of is the fact that the normalization performed does not preserve the relationships between values between attributes.

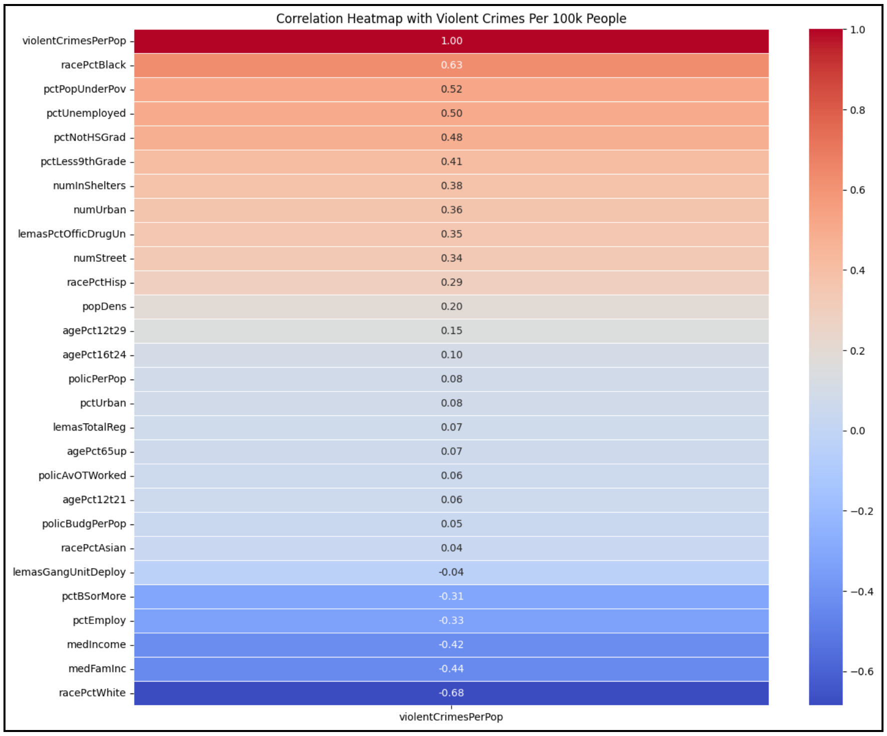
Exploratory Data Analysis

In order to gain a deeper understanding of our dataset, we created multiple visualizations to see how different features correlate with each other.

1. In this distribution of Violent Crimes Per Population in the United States, there are plenty more instances of low crime rate communities in comparison to high crime rate communities. The mean line reflects this as well. More specifically, the most frequent violent crimes per capita is between .0 and .1. It is also important to point out that cities with a high violent crime of 1.0 seem to have a spike of up to ~50 cities. Other than that the frequency starts to converge as the violent crime rate increases.



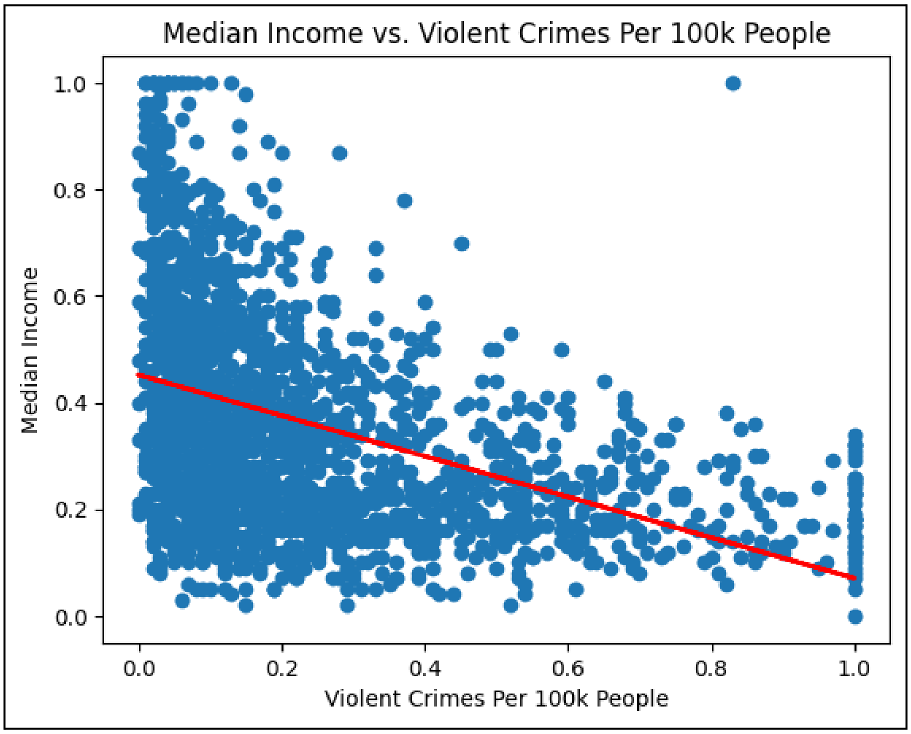
2. The correlation heatmap shows the correlation between a number of factors and violent crime rates. The factors include race, poverty, unemployment, education, income, and housing. The heatmap shows that there is a strong positive correlation between violent crime rates and race, poverty, unemployment, and the percentage of people who have not graduated from high school. There is also a moderate positive correlation between violent crime rates and the percentage of people who are renting their homes and the percentage of people who live in urban areas. There are a few possible explanations for this correlation. One possible explanation is that poverty, unemployment, and low educational attainment can lead to social exclusion and marginalization. Given the lack of resources in underserved communities, we can see that things like poverty, lack of education and opportunity for employment can contribute to a violent crime rate. Most of the communities statistically speaking are predominantly non-white, which indicates a high positive correlation between percentage of African American populace and violent crime rate. On the other hand with strong negative correlation, we have median income, employment, family median income, and percentage of white populace. Statistically speaking, when we look at cities with heavy affluence we can see that there are more thriving businesses, more opportunities for increasing median income, and there is also more of a caucasian predominance in these types of areas, which may lead to the strong negative correlation. In the middle we can see that the features that don't have as much correlation are the age demographics and the police characteristics.



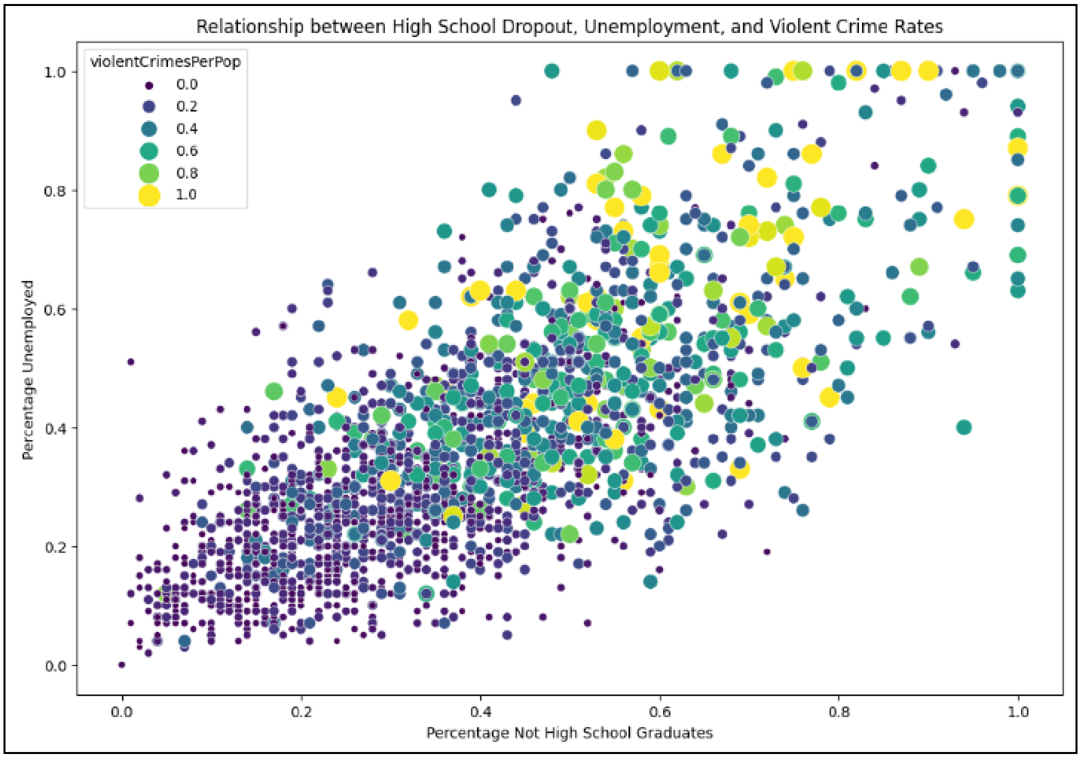
3. This scatter plot of violent crimes per 100,000 people versus median income gives us an idea of these 2 features related to one another. The red line is a negative correlation, which suggests that violent crimes are more likely to occur in areas with lower median income. This is consistent with the heat map that has shown that poverty is a strong predictor of violent crime. There are a number of possible explanations for this relationship. One possibility is that people in poverty are more likely to experience the stressors that can lead to crime, such as unemployment, lack of housing, and exposure to violence. Another possibility is that poverty undermines social cohesion and trust, which can create an environment where crime is more likely to flourish. It is important to note that correlation does not equal causation. Just because there is a relationship between poverty and violent crime does not mean that poverty causes violent crime. However, the research suggests that poverty does play a role in creating the conditions that lead to violent crime. The graph you sent also shows that there is a significant amount of variation in violent crime rates, even among areas with similar median incomes. This suggests that there are other factors, in addition to poverty, that contribute to violent crime. These factors may include:

* The presence of gangs and other criminal organizations
* The availability of guns
* The quality of schools and other community institutions
* The level of trust between the police and the community

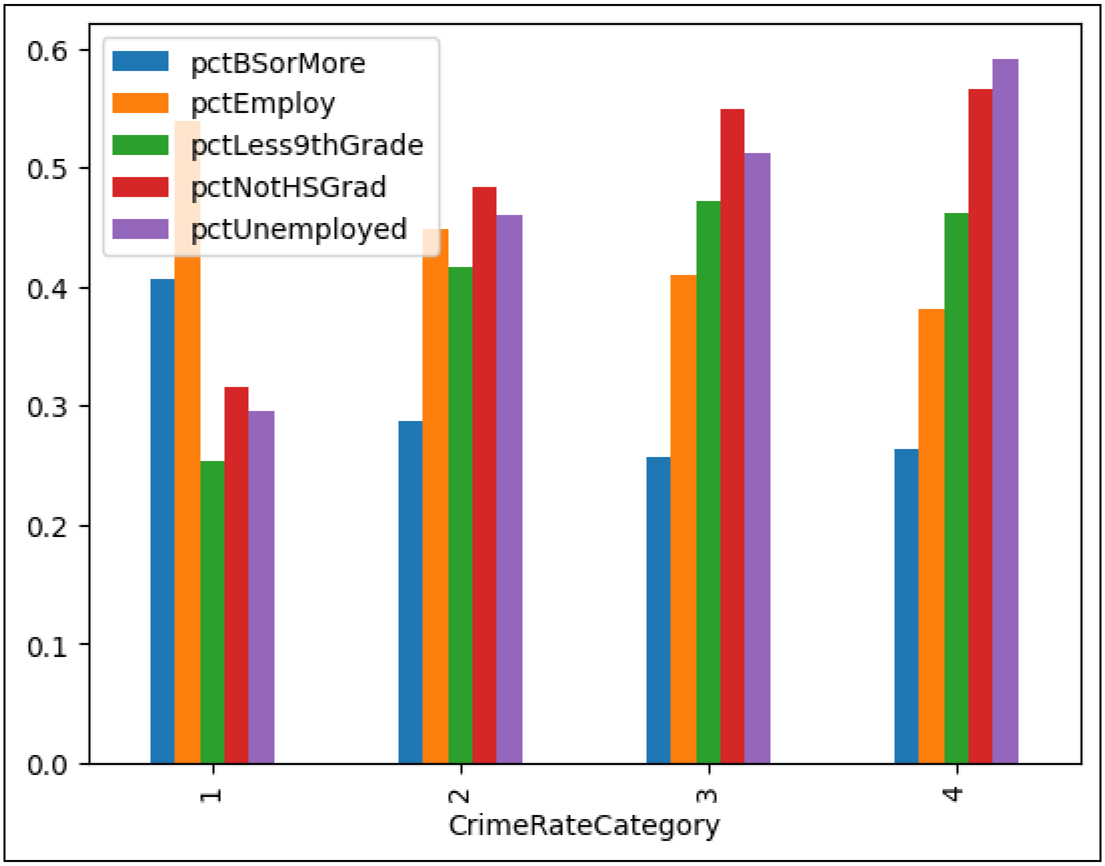
The graph also shows that there are some areas with low median incomes that have relatively low violent crime rates. This suggests that it is possible to reduce violent crime rates in low-income areas, even if it is difficult to completely eliminate poverty.



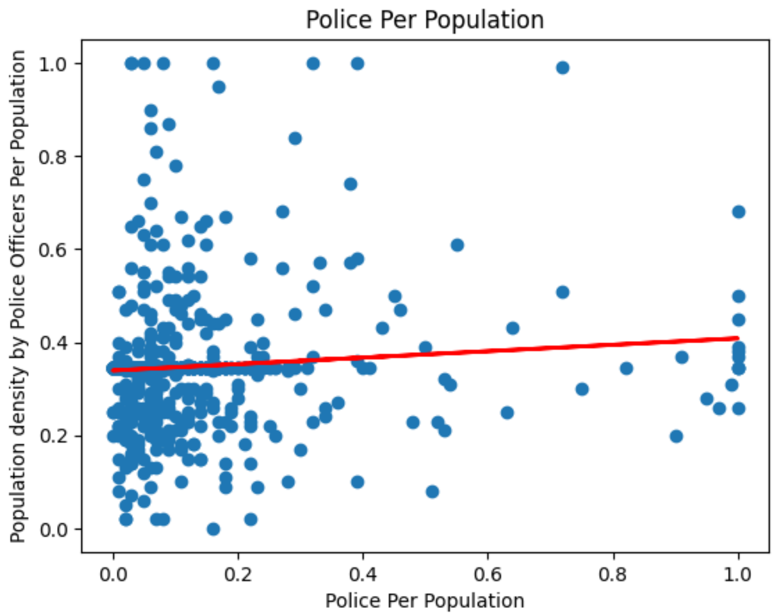
4. We used another scatter plot to show the relationship between the percent of people unemployed vs the percentage of people of non highschool graduates. The trend is a positive correlation, which suggests that violent crimes are more likely to occur in areas with a higher percentage of population with less than a high school diploma. Furthermore, there are more instances of 0, and .2 violent crime. rate near the bottom left of the graph. This implies that as more people are employed, and highschool graduates, the less violent crime there is. Near the middle of the graph there are more instances of crime rates that are .4 and above. To the right of the graph, most instances are equal to or above .6 in violent crime rate. This implies that as more people become unemployed, and don't graduate from highschool, violent crime increases. Overall, this data is a visual representation of the relationship between low educational attainment and violent crime. Additionally, it is a reminder that low educational attainment is a major problem that can have a significant impact on communities.



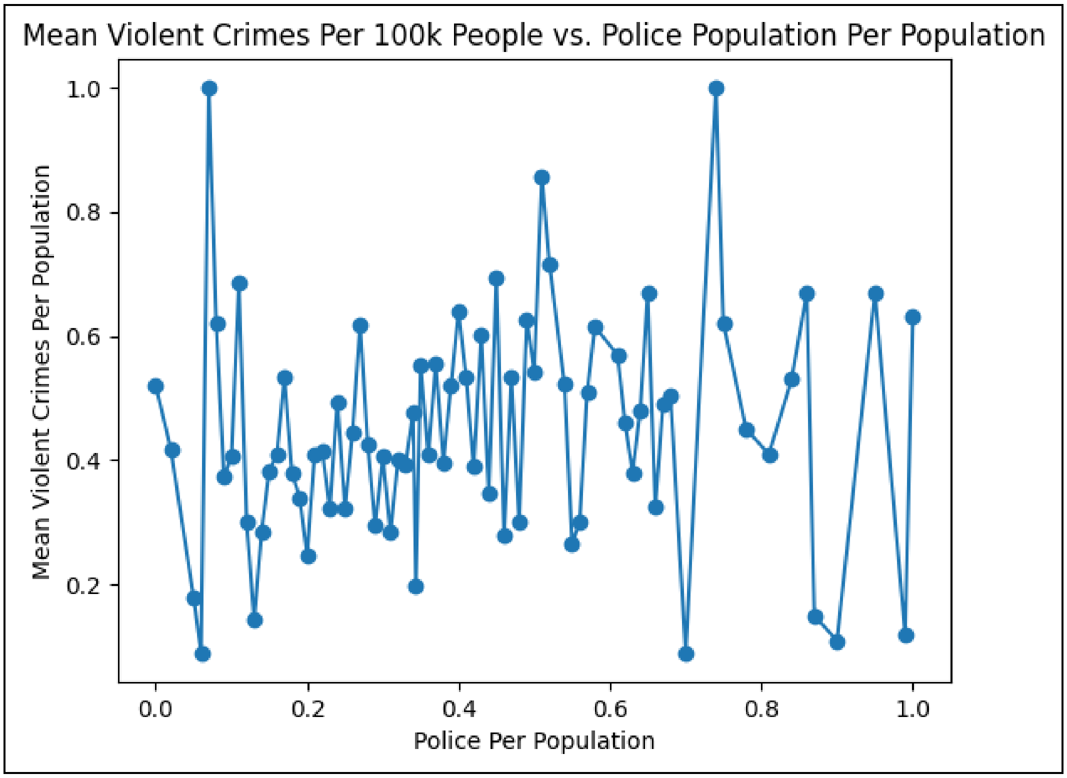
5. We created a bar graph that showed the mean of 5 different variables in each crime rate category. The variables we used are percentage of people 25 and over with less than 9th grade education, percentage of people 25 and over who are high school dropouts, percentage of people 25 and over with college degrees or higher, and the percentages of people 16 and over who are employed and unemployed. This graph shows that as with high crime rates, education seems to become less of a priority. The crime rate category of 1 has the highest average percentage of 16 and over employed while category 4 has the lowest. The average percentage of 25 and over with college degrees or higher also follows this pattern. The other 3 are the complete opposite. The average values of percentage of 16 and over that are unemployed, percentage of 25 and over with less than 9th grade education, and percentage of 25 and over without a high school diploma are the lowest in category 1, small crime rate, and the largest in category 4, large crime rate.



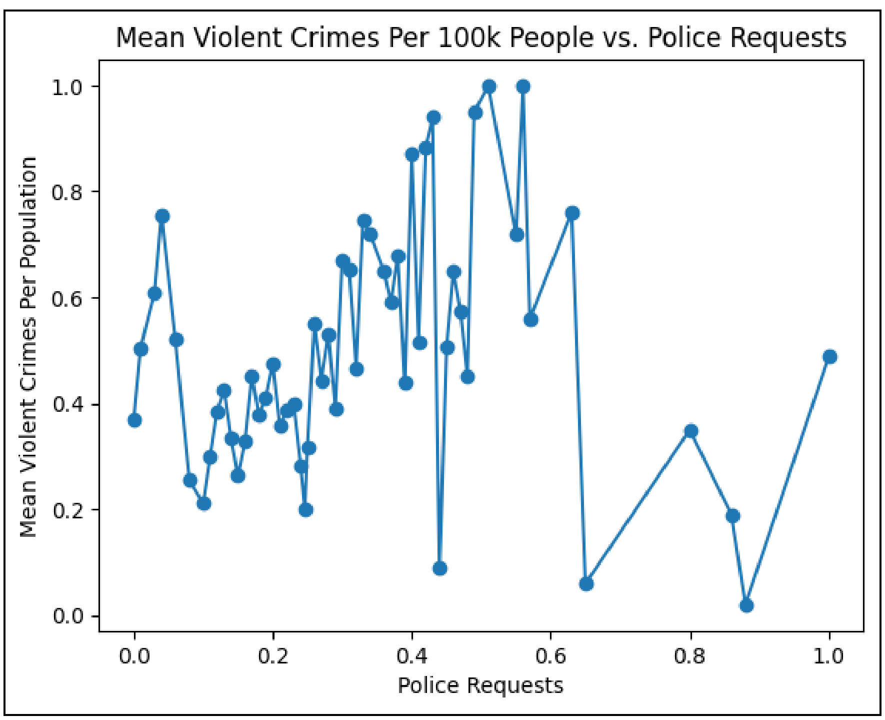
6. We made another scatter plot, this time trying to see if population density and police officers per population were related. In this graph, the data is highly concentrated around low police per population and population density. The highest, and lowest instances of population density are found when police per population is between 0, and .2. As the police per population increases, it does not seem that population consistently decreases or increases. What this graph implies is that the majority of communities have a police per population around .0 and .2. There seems to be no correlation between these variables.



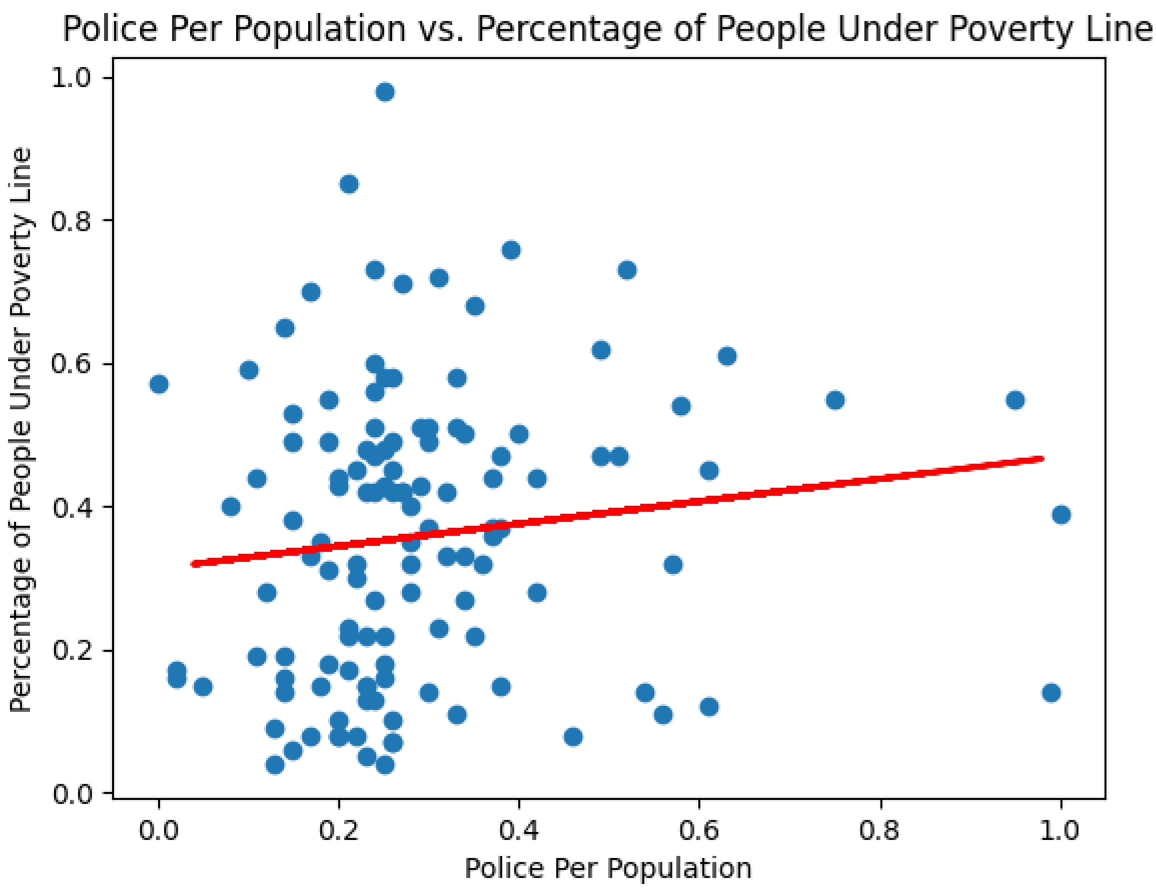
7. We plotted a line graph with the mean of police officers per 100K population versus crime rate. Even though it seems that a higher number of police values would mean lower crime rates, this graph shows that this is not necessarily the case. Even though we can see that there are instances of lower crime rates with higher numbers of police officers, there are plenty of instances otherwise. This makes sense if we remember that the number of police officers in an area is directly related to how dense the population is. The number of officers assigned to an area is most likely based on population size rather than crime rate. The data is scattered enough that we can't draw any sound conclusion of how these two values are related.



8. We plotted another line graph comparing the mean of police requests and crime rate. Similarly to the last graph this data is very scattered, and does not show a trend towards violent crime increasing, or decreasing as police requests increase. Despite this, it's worth noting that at higher police requests after .6, the average violent crime rate is low. We still can't make any conclusion from this though.

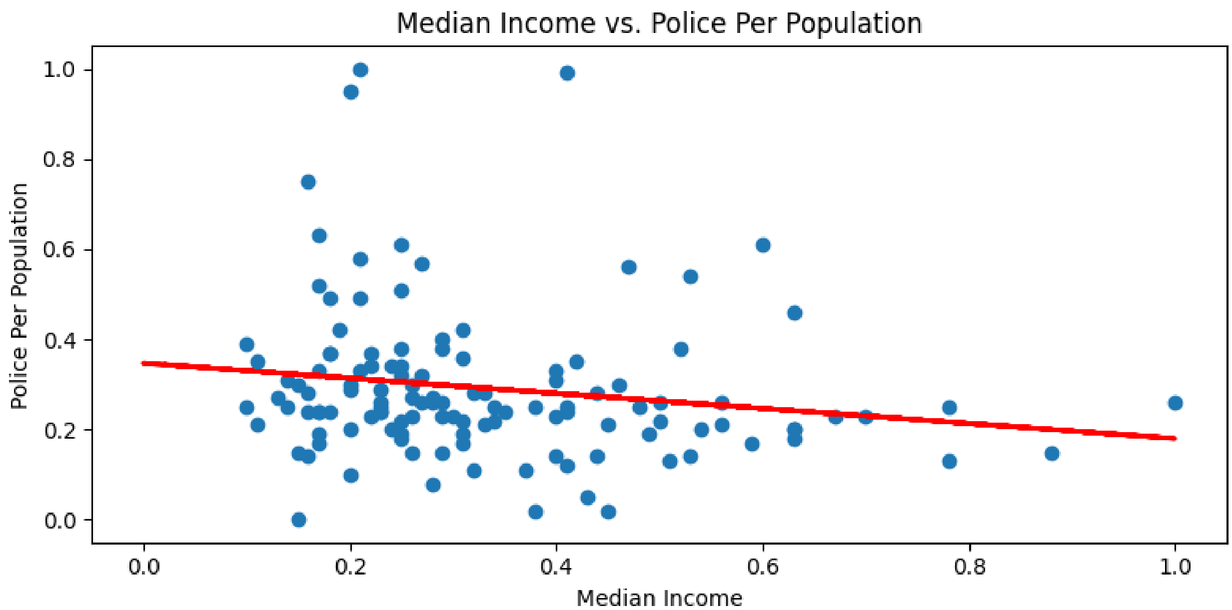


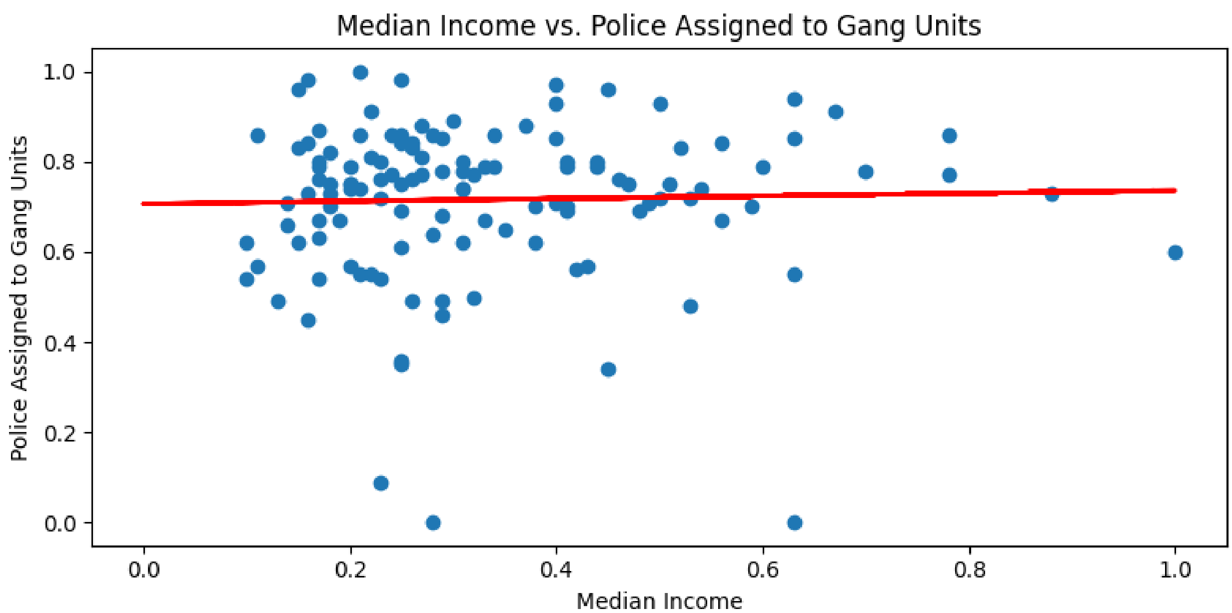
9. Police Per Population v. Percentage of People Under Poverty Line

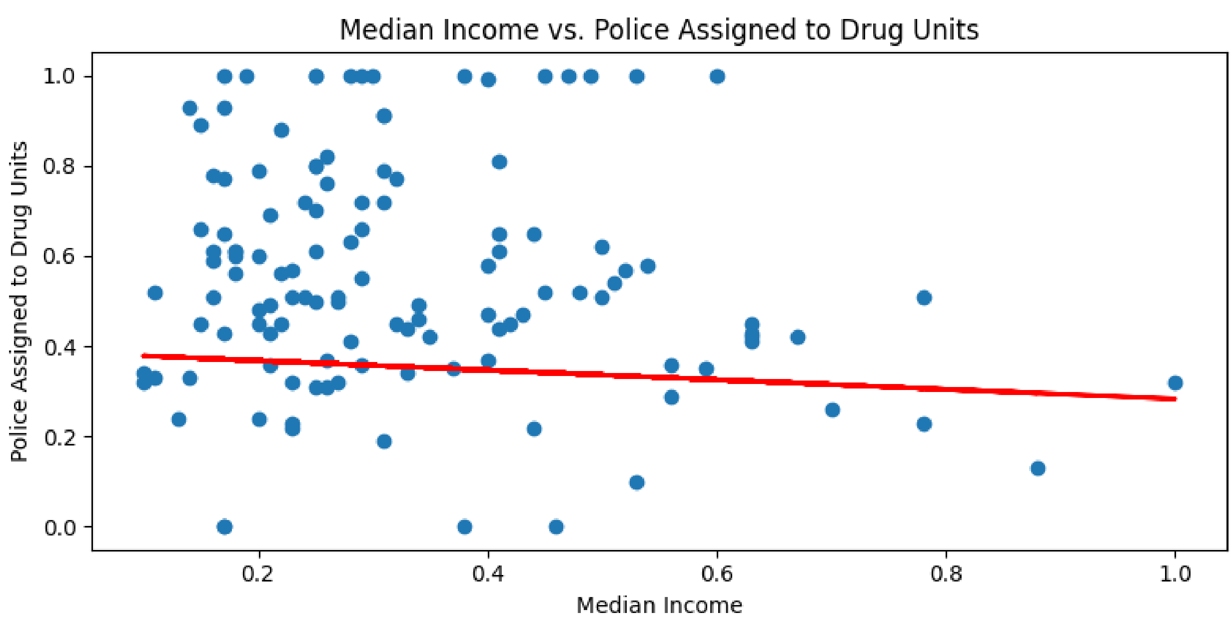


We used a scatter plot to see if there was a relationship between Police Per Population, and Percentage of People Under Poverty Line There is a high concentration of the percentage of people under the poverty line between 0.1, and 0.6 when the police per population is between 0.1, and 0.3. As we go through the graph, there isn't a consistent trend. Even though the linear regression suggests it, there does not seem to be any correlation between police per population and the percentage of people under the poverty line.

10. Median Income vs Police Statistics


Median Income vs. Police Per Population
Police Per Population
Median Income
0.0
0.2
0.4
0.6
0.8
1.0


Median Income vs. Police Assigned to Gang Units
Police Assigned to Gang Units
Median Income
0.0
0.2
0.4
0.6
0.8
1.0


Median Income vs. Police Assigned to Drug Units
Police Assigned to Drug Units
Median Income
0.0
0.2
0.4
0.6
0.8
1.0

When comparing median income and police per population, the police per population is generally less than .4. However, there's an obvious negative trend. This implies that as median income increases, police per population decreases. This relationship could be caused by:

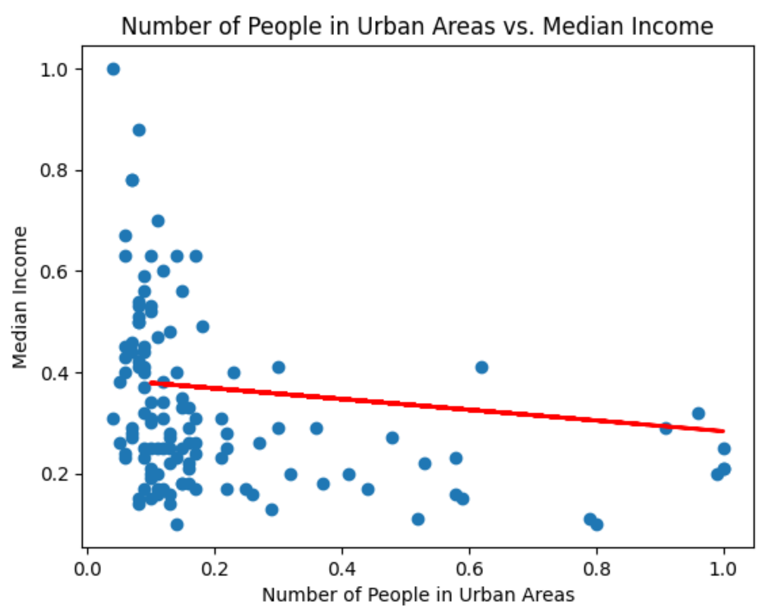
- An increase in population, leading to more people who have low incomes
- Smaller police populations in smaller cities that have higher levels of income

When comparing median income and police assigned to gang units, even though the regression line suggests it, they aren't correlated. Between the median incomes .1 and .4, the data is highly concentrated. As median income increases, the number of police assigned to gang units seems to stay around the same range.

When we compare median income and police assigned to drug units, we can see that there is actually a negative trend. Most instances are clustered between a median income of .1 and .3. Most of these instances have a police assigned to drug units value equal to or above .4. As median income increases, the instances generally go below this range. This could be due to

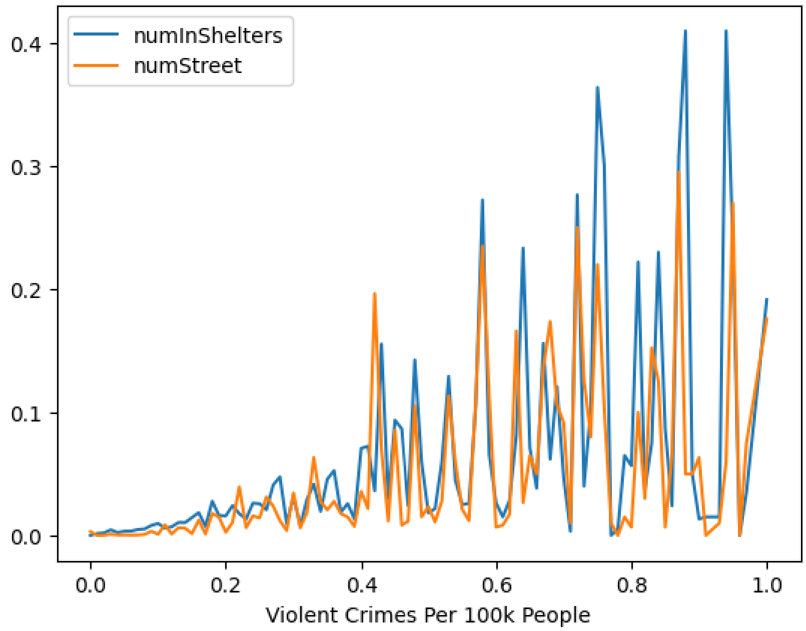
- Gangs being more prevalent in lower income communities due to them lacking support.

11.



In this visualization, we want to see if there's a correlation between the number of people in urban areas and median income. The data is highly concentrated from 0 to .2 in the number of people in urban areas. In this range, we get our highest ranges of median income. As the number of people in urban places, we see median income decrease. This means that there could be a negative correlation between the number of people in urban areas and median income. However, most communities in our data are considered urban, so the more likely reason for this correlation is tied to population density. Higher population density could lead to more varieties of income, some of which could be low.

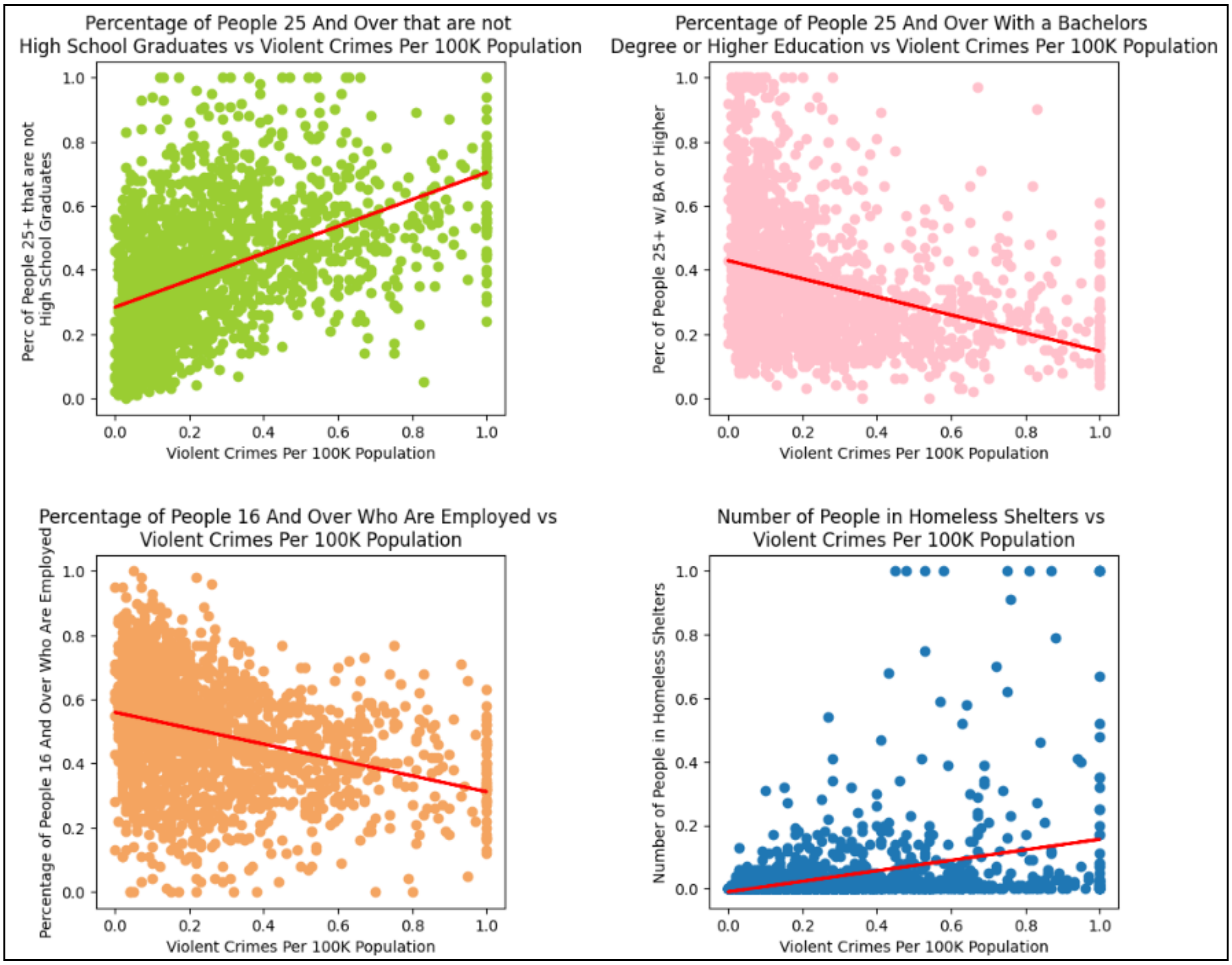
12.



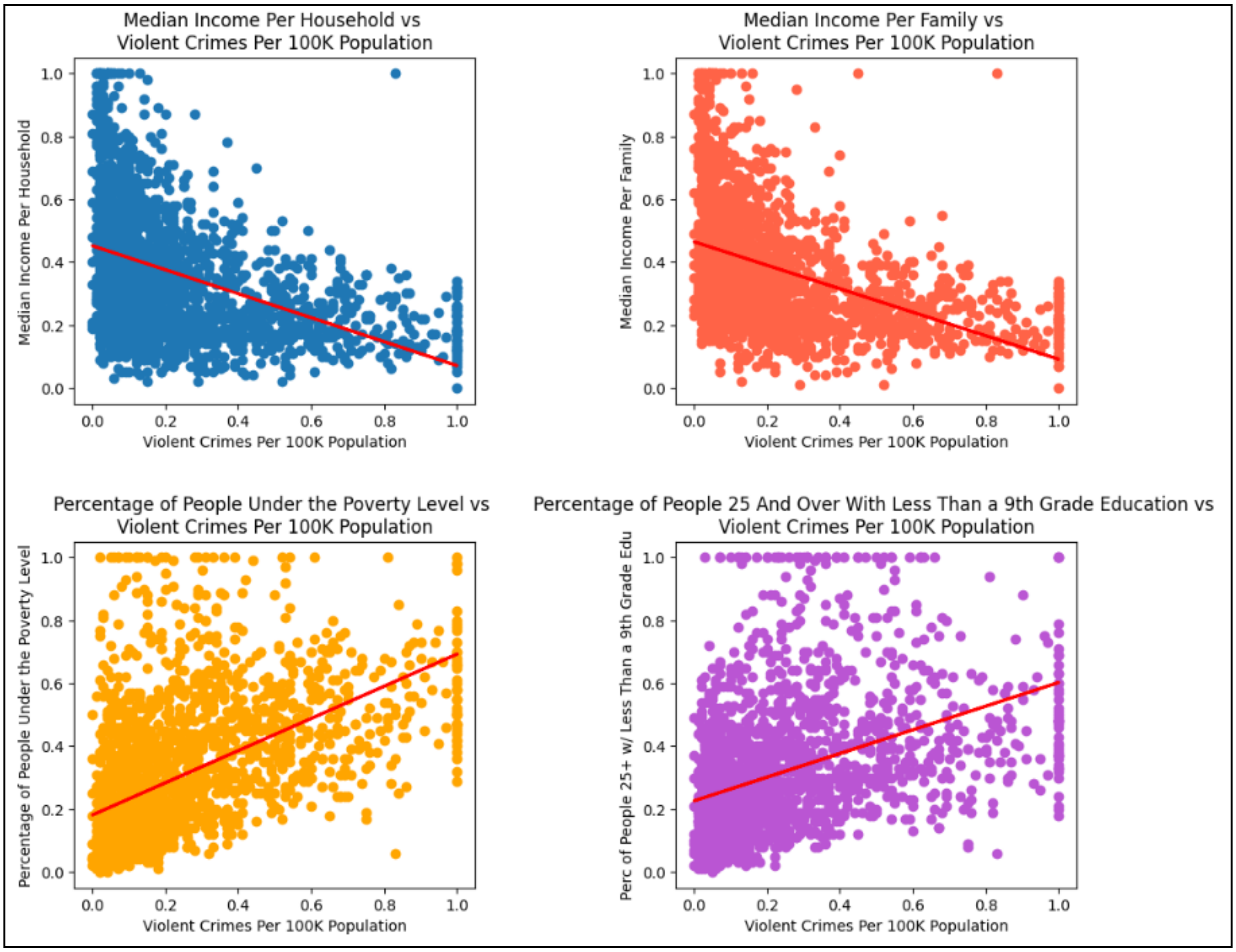
In this line graph we wanted to see if there was a correlation between homelessness and violent crime per 100k people. As we can see, the number of people in Shelter and on the streets have a similar line graph. This makes sense, since people who are homeless often alternate between living on the streets and in shelters. As we can see, as these variables increase, so do violent crimes per 100k people.

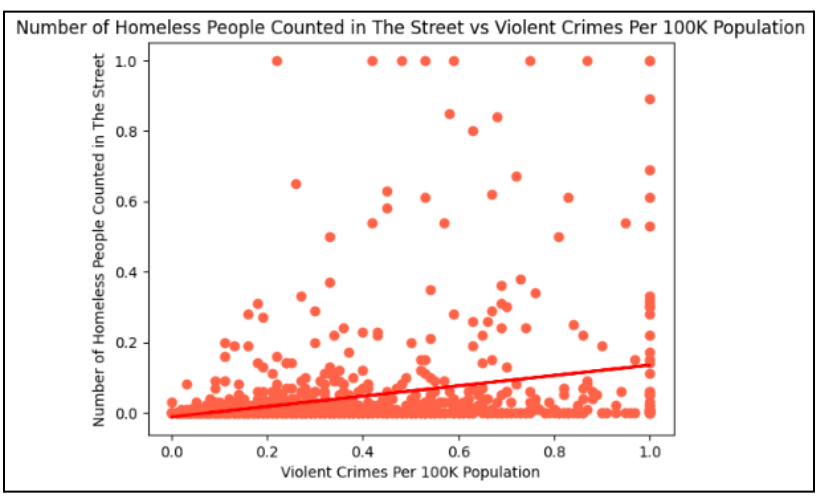
Additionally, we have conducted plenty of hypotheses tests on variables that pertain to demographic, socioeconomic, and law enforcement related features.

1. **Hypothesis 1**: There is a correlation between socioeconomic factors and violent crime rate.
    a. The socioeconomic factors we included are median income per household and per family; percentage of people under the poverty level; percentage of people 25 and over who had less than 9th grade education, didn't graduate high school, and had a bachelor's degree or higher education; percentage of people 16 and older who were employed; the number of people in homeless shelters; and the number of homeless counted in the streets. We used scatter plots and best fit lines to visualize the correlation between each of these features and crime rate.



Percentage of People 25 And Over that are not High School Graduates vs Violent Crimes Per 100K Population
Perc of People 25+ that are not High School Graduates
Violent Crimes Per 100K Population
Percentage of People 25 And Over With a Bachelors Degree or Higher Education vs Violent Crimes Per 100K Population
Perc of People 25+ w/ BA or Higher
Violent Crimes Per 100K Population
Percentage of People 16 And Over Who Are Employed vs Violent Crimes Per 100K Population
Percentage of People 16 And Over Who Are Employed
Violent Crimes Per 100K Population
Number of People in Homeless Shelters vs Violent Crimes Per 100K Population
Number of People in Homeless Shelters
Violent Crimes Per 100K Population

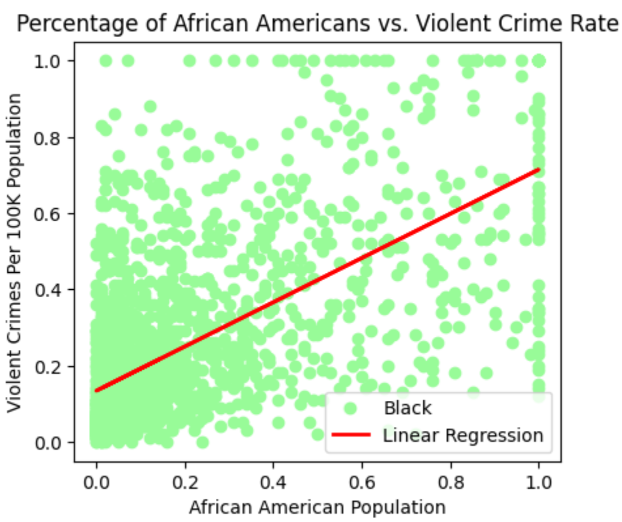


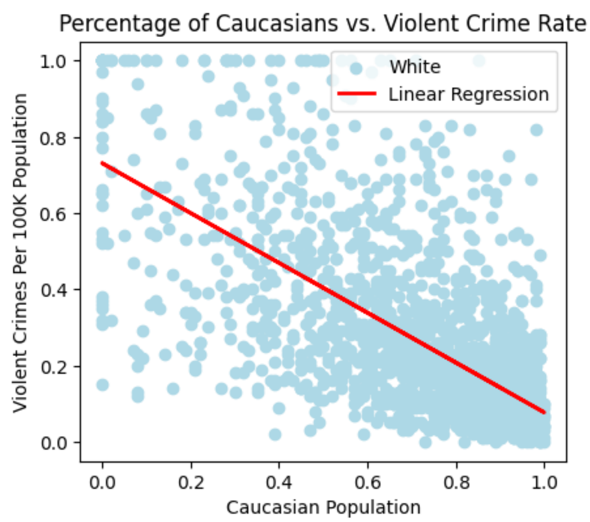


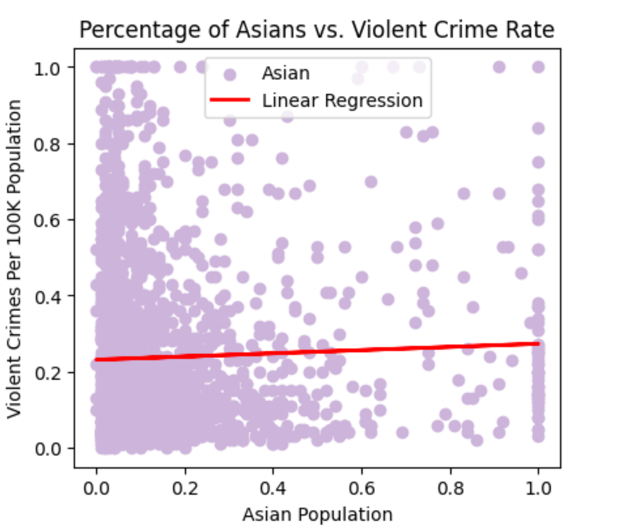
b. In order to see whether the feature was statistically relevant or not, we used the chi squared test for each of them. We concluded that based on all of these tests, not all of these socioeconomic features are statistically relevant for our regression analysis. The median income per household, median income per family, and the percentage of people 25 and over that have a bachelor's degree or higher have no correlation with violent crime per population. The rest have significance, and a correlation: number of homeless in shelters the number of homeless on the streets, the percentage of people under the poverty level, the percentage of people 25 and over with less than 9th grade education, and the percentage of people 25 and over who didn't graduate high school all have a positive correlation with crime rate. The percentage of people 16 and over who are employed is negatively correlated with crime rate. Overall, since the total p-value is under 0.05 we conclude that there is a positive correlation between socioeconomic features and crime rate.

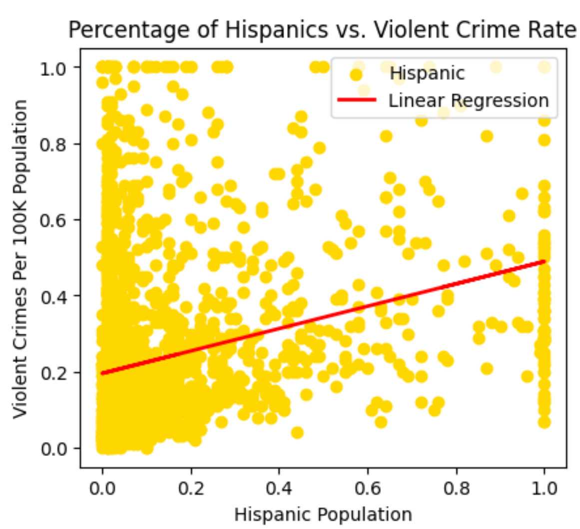
   c. We also used a Pearson correlation test to see whether there was an overall correlation between all of the variables and crime rate. We concluded that there is a positive correlation between socioeconomic features and violent crime rates.

2. **Hypothesis 2**: There is a correlation between demographic and violent crime rate.
   a. *Visualizations:* Within our dataset, demographic is represented by the percentage of African Americans. Caucasians, Asians, and Hisapnics within each community. To visualize the relationships between demographic and crime rate we created scatter plots with a fitted linear regression line on each plot.







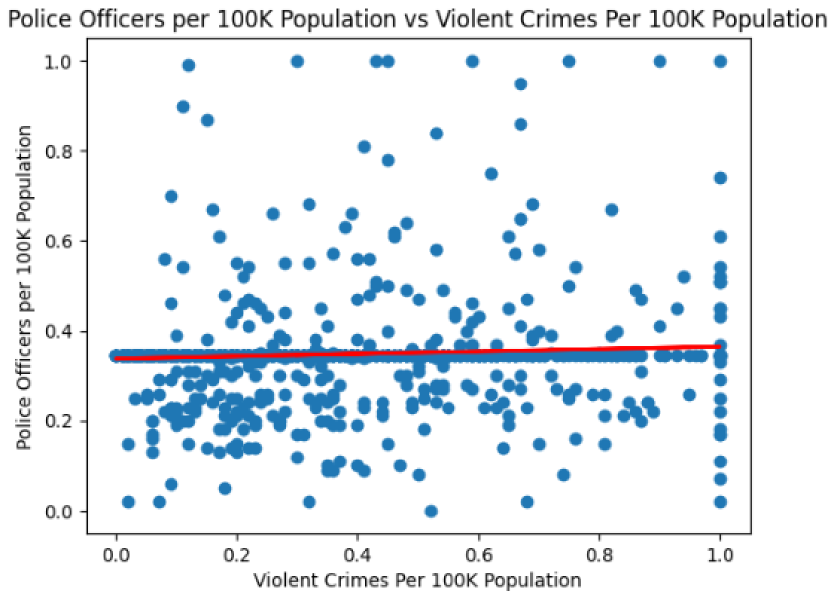


The linear regression line essentially models the relationship between each race and violent crime rate, which gives us a deeper understanding of the effect that each race has. When looking at the graphs, it's clear that the linear regression for African Americans shows a significant opposite trend compared to Caucasians. This difference is due to historical and systemic racism that has negatively affected the prosperity of African Americans. It's important to note that even though these statistical tests show correlation, they don't mean there's a cause-and-effect relationship.

b. *Pearson Coefficients:* Before conducting the statistical test, we first observed the Pearson coefficients for each race to further examine the relationships that exist. These correlations imply that areas with a larger African American and Hispanic population may experience higher levels of violent crime, as they yielded values of 0.63 and 0.29, respectively. The correlation values for Asians is relatively weak at 0.04, though because it is a positive value this suggests a slight tendency for areas with a higher Asian demographic to have a slightly higher rate of violent crimes. The only race that yielded a strong negative value were Caucasians (-0.68), which implies that areas with a larger Caucasian population will tend to have lower levels of violent crime. These coefficients are consistent with the linear regression lines we have plotted.

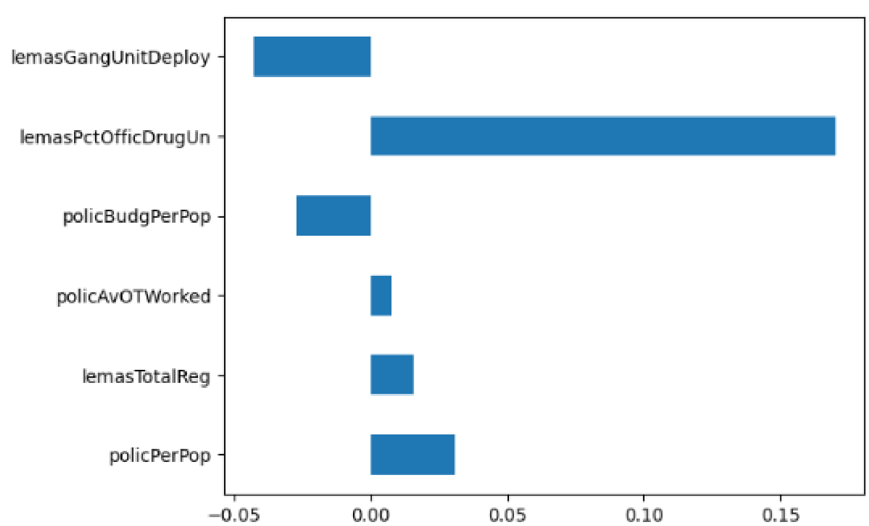
c. *Statistical Test:* To check if there is a statistically significant correlation between each demographic variable and the number of violent crimes per population, we conducted a Pearson correlation coefficient hypothesis test. From the coefficients we found our p-value to be 8.76 e-41, which is less than our alpha value of 0.05. Finally, we can reject the null hypothesis and conclude that there is a significant correlation between demographic and violent crime rates.

3. **Hypothesis 3**: There is a correlation between police-to-population ratio and crime rate.
    a. To visualize this we plotted a scatter plot and a best fit line between the number of police officers per 100K population and crime rate.



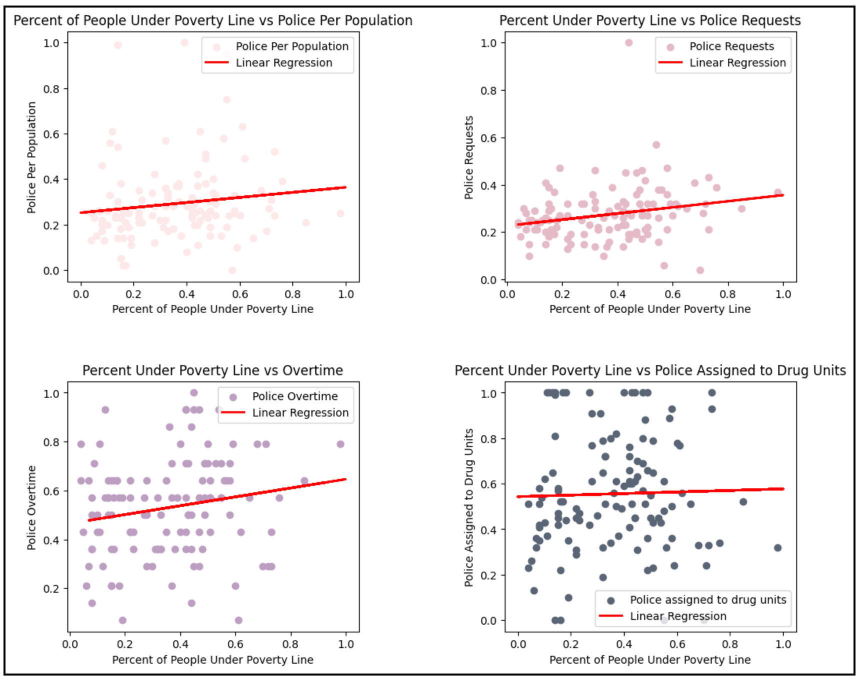
b. We did a chi square test to see if the correlation was significant. We concluded that the chi squared statistic is greater than the critical value and there is a positive correlation. It's statistically significant because the p-value is close to or equal to 0.

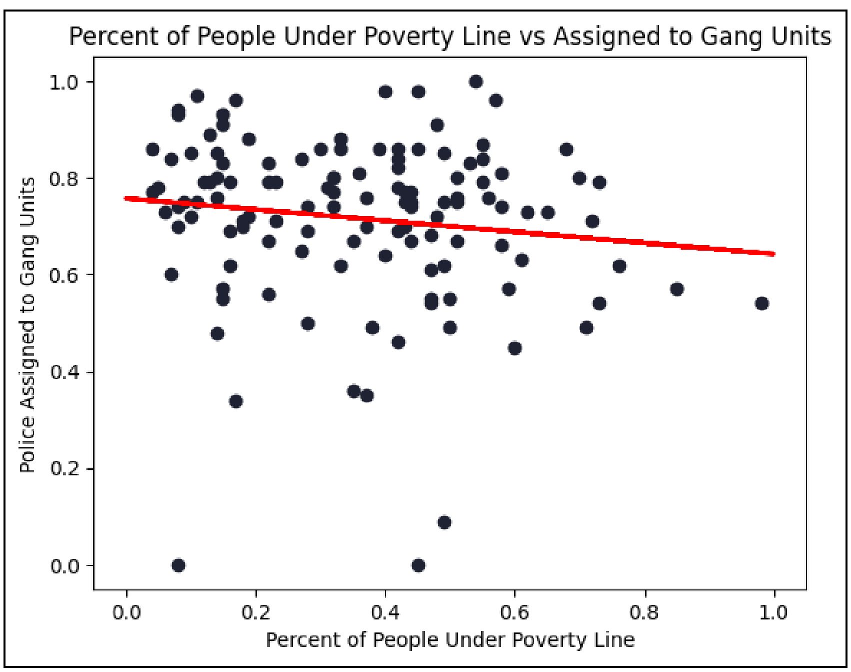
4. **Hypothesis 4**: There is a correlation between the population of those who are below the poverty line and police presence
   a. To analyze this hypothesis we used the number of police officers per 100K population, total requests for police, the average number of police that worked overtime, police operating budget per 100K population, percent of officers assigned to drug units, and whether or not a gang unit was deployed. We first created a bar graph that displays the correlation of each feature with the percentage of people that are under the poverty level.



Then we graphed scatter plots and the best fit lines for each feature against the percentage of people that are under the poverty level.

```
policPerPop              0.030908
lemasTotalReg            0.015731
policAvOTWorked          0.007604
policBudgPerPop         -0.027302
lemasPctOfficDrugUn      0.170169
lemasGangUnitDeploy     -0.042947
```

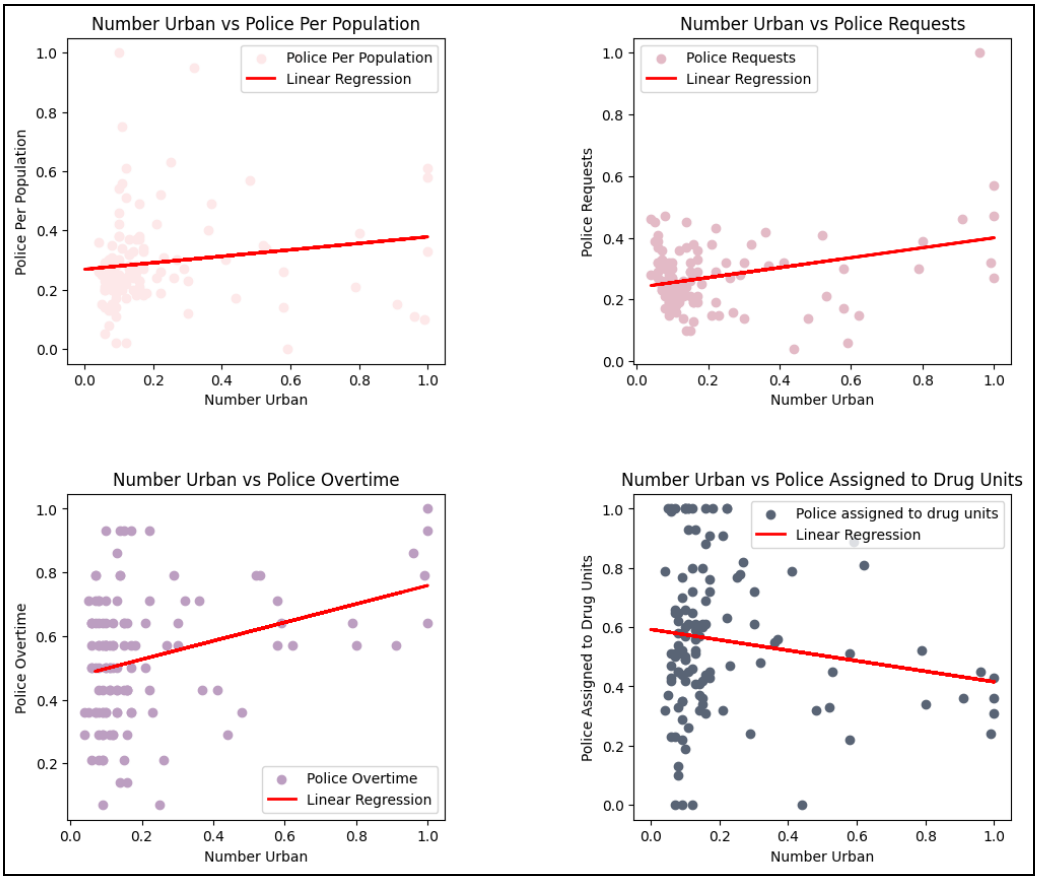


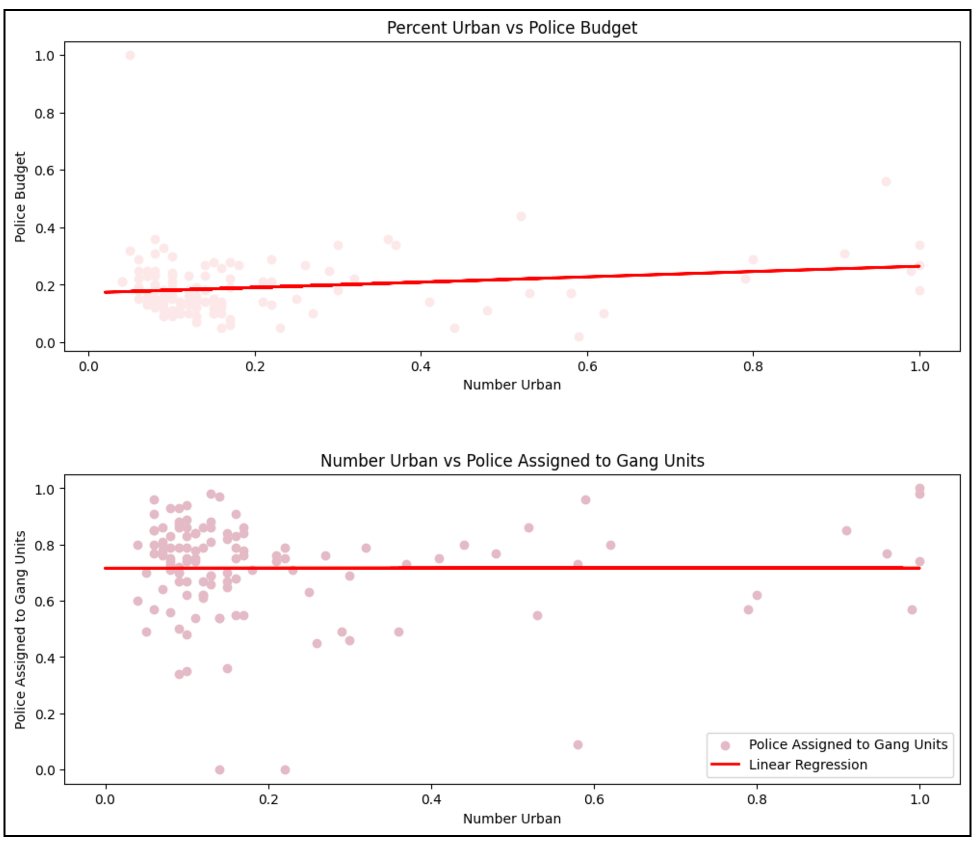


   b. We used chi square tests to analyze each variable. Based on this information and the correlation matrix, we conclude that percent of people under the poverty line is positively correlated with police per 100k people, police requests, police overtime, and police assigned to drug units. Surprisingly, the number of police assigned to gang units is negatively correlated with the percent of people under the poverty line, and police budget per population. This shows police activity is correlated with the percent of people under the poverty line. However, we do not know which is the cause or effect.

5. **Hypothesis 5**: There is a correlation between those who live within an area that is "urban" and police presence
   a. To analyze this hypothesis we used the number of police officers per 100K population, total requests for police, the average number of police that worked overtime, police operating budget per 100K population, percent of officers assigned to drug units, and whether or not a gang unit was deployed. We first created a correlation matrix that displays the correlation of each feature with the number of people living in an area classified as "urban.".

policPerPop	0.134236
lemasTotalReg	-0.070525
policAvOTWorked	0.249489
policBudgPerPop	-0.036114
lemasPctOfficDrugUn	0.478179
lemasGangUnitDeploy	-0.058712


Number Urban vs Police Per Population
Police Per Population
Linear Regression
Police Per Population
Number Urban
Number Urban vs Police Requests
Police Requests
Linear Regression
Police Requests
Number Urban
Number Urban vs Police Overtime
Police Overtime
Linear Regression
Police Overtime
Number Urban
Number Urban vs Police Assigned to Drug Units
Police assigned to drug units
Linear Regression
Police Assigned to Drug Units
Number Urban


Percent Urban vs Police Budget
Police Budget
Number Urban
Number Urban vs Police Assigned to Gang Units
Police Assigned to Gang Units
Police Assigned to Gang Units
Linear Regression
Number Urban

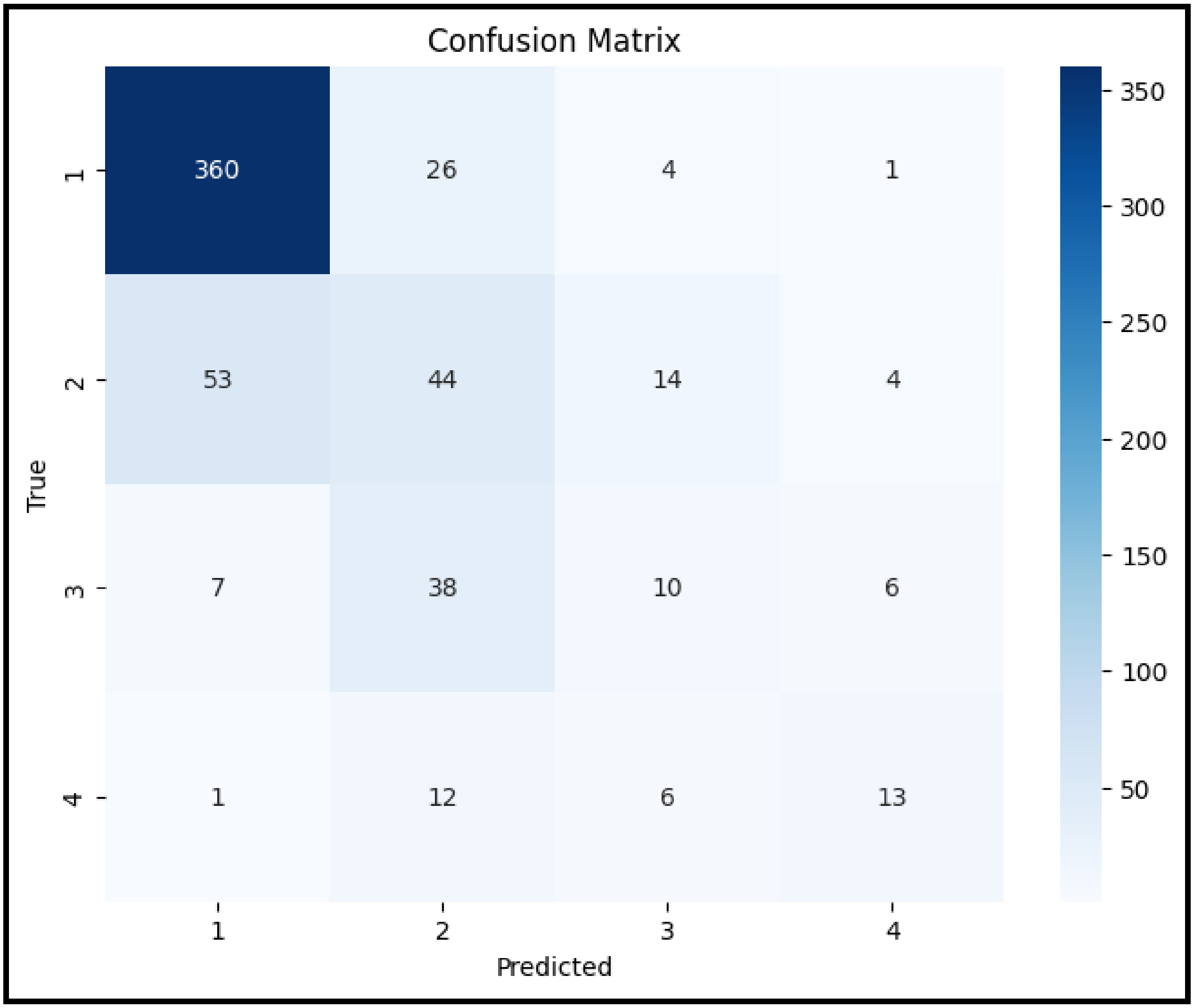
b. We used chi squared analysis tests to see which correlations were statistically relevant. We concluded that the number of people in urban areas is positively correlated with police per population, total police requests, and police overtime. The number of people in urban areas is negatively correlated with police assigned to drug units, and gang units, and police budget..

Supervised Learning

For our Supervised Learning model, we chose to use a Random Forest Classifier algorithm to classify the violent crime rate in 4 different categories. The 4 categories are 1 - Low Crime Rate, 2 - Moderately low Crime Rate, 3 - Moderately High Crime Rate, 4 - High Violent Crime Rate. In terms of what features we chose to train on, the features contain percentages of each race in a given city, their age, percentage of those who live in urban areas, median income, percentage of those who live under the poverty line, and other socioeconomic features. In total, we are training on 25 different features out of the 100+ that our dataset contains. Our training test split is a 70-30 percent split with a random_sampling of 92 to ensure the randomness of our data as it feeds into our model. Before running the model, we also chose to use a hyperparameter optimization technique called “grid-search”.

```
param_grid = {'n_estimators': [10, 30, 50, 100, 150],
              'max_depth': [None, 10, 20],
              'min_samples_split': [2, 5, 10]}
```

The purpose of this is to ensure that we get the most accurate results when running our model, in other words we are maximizing the performance of our model by trying different hyperparameters. After fitting our model we chose to use the best performing random forest tree model and created a confusion matrix to visualize our results of the highest performing model.



After our confusion matrix, we created a classification report to look at the precision, recall, and f1-score for each category.

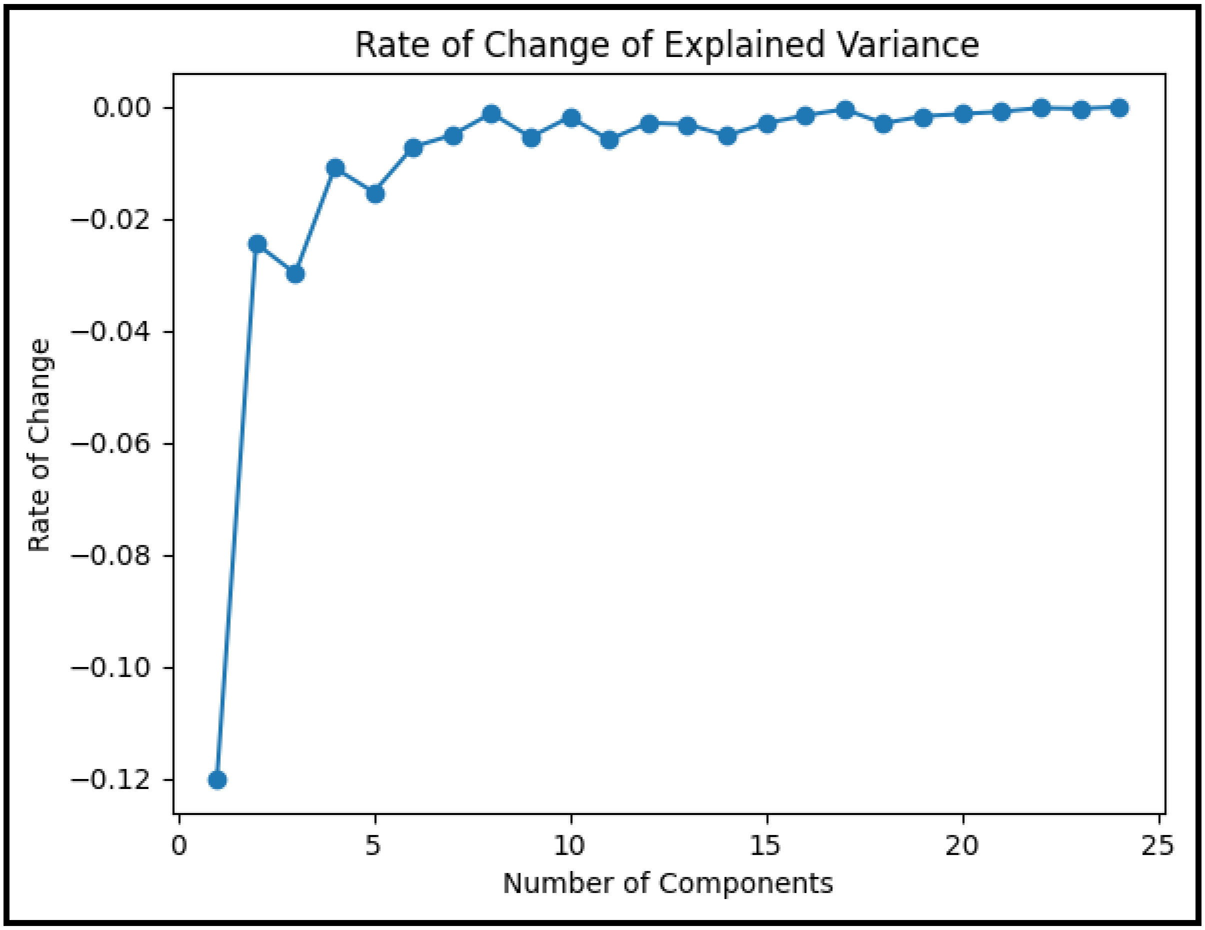
	precision	recall	f1-score	support
1	0.86	0.92	0.89	391
2	0.37	0.38	0.37	115
3	0.29	0.16	0.21	61
4	0.54	0.41	0.46	32
accuracy			0.71	599
macro avg	0.51	0.47	0.48	599
weighted avg	0.69	0.71	0.70	599

Starting with Low Crime Rate, The model is doing a very good job of predicting low crime instances correctly, with a precision of 0.86 and a recall of 0.92. This means that the model is not making many false positives or false negatives for a low crime rate. For moderately low crime rate, the model is doing a less good job of predicting these instances correctly, with a precision of 0.37 and a recall of 0.38. This means that the model is making more false positives and false negatives for moderately low crime. For moderately high crime rate, The model is doing a poor job of predicting these instances correctly, with a precision of 0.29 and a recall of 0.16. This means that the model is making many false positives and false negatives for this instance. Lastly for High Crime rate, The model is doing a fair job of predicting their instances correctly, with a precision of 0.54 and a recall of 0.41. This means that the model is making some false positives and false negatives for High Crime, but not as many as for moderately high. Overall, the model is doing a pretty good job of predicting crime rate, but it could be improved in higher levels if there was more training data for those instances. We believe that an overall accuracy of 71-72% is satisfactory for our classifications. We also conducted cross validation on our model, even though for Random Forest Classifiers it’s not necessary due to the out of the bag features that the classifier has built into it. Here is the validation metric that verifies our results at 5-fold cross validation:

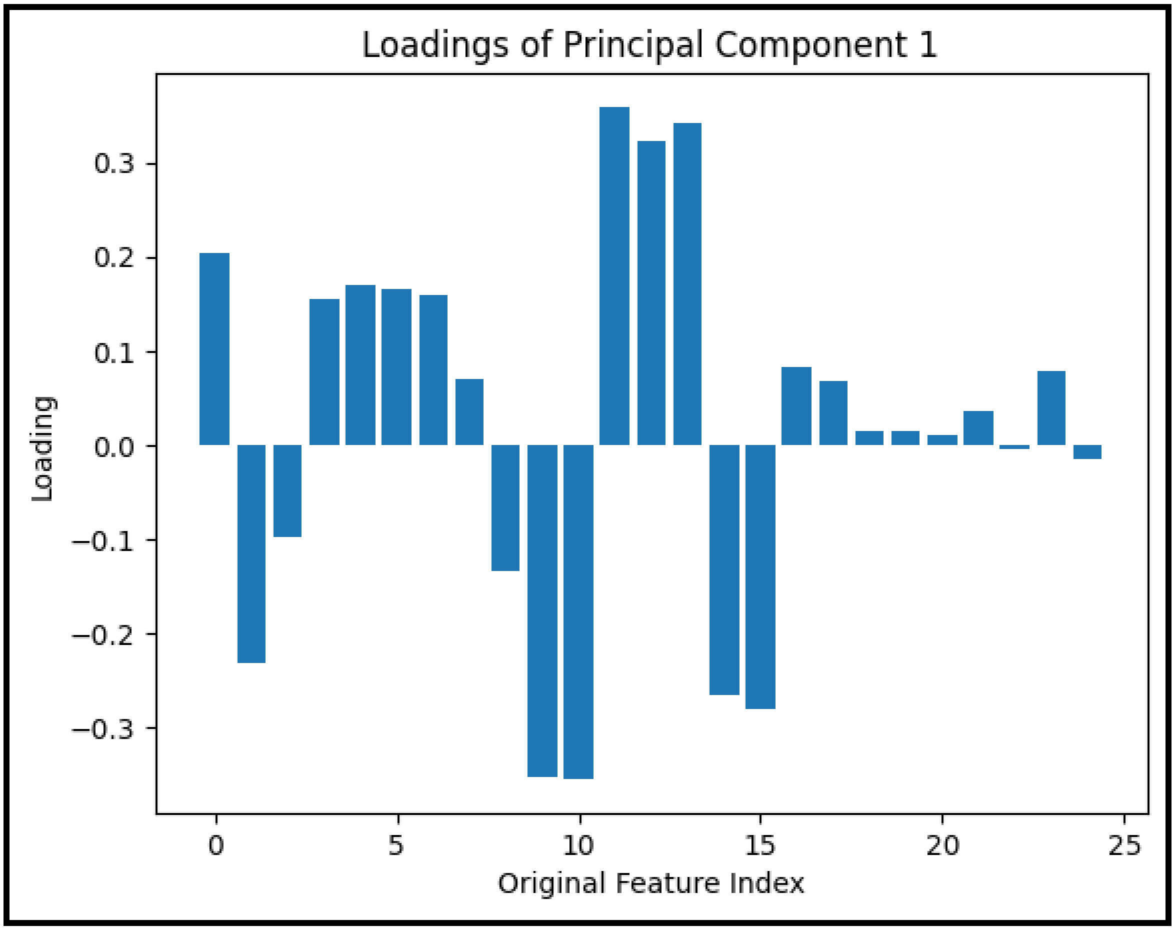
```
Scores for each fold are: [0.71929825 0.73433584 0.73182957 0.73684211 0.71356784]
Average Score: 0.727175
```

Unsupervised Learning

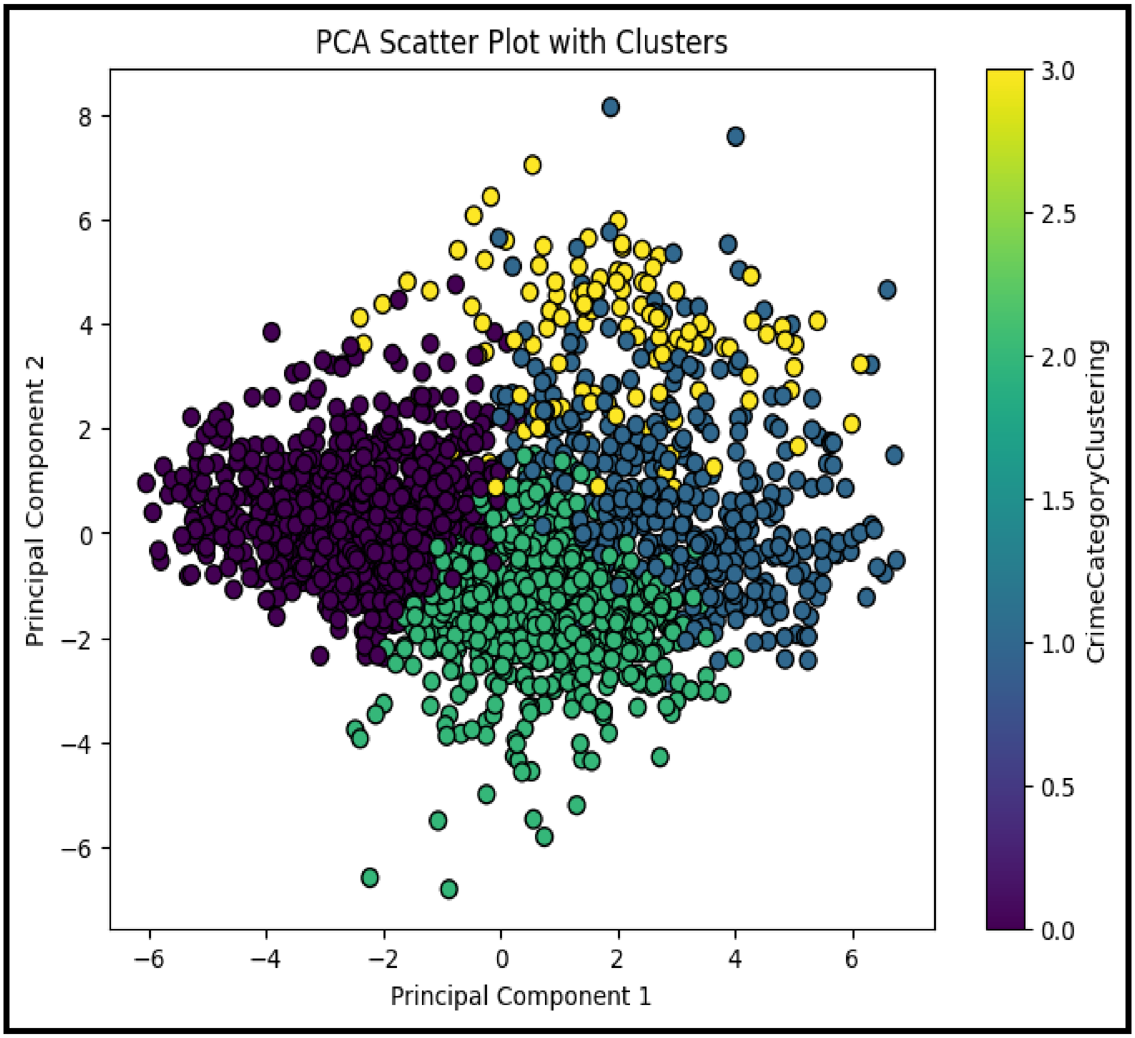
We employed an unsupervised model that predicts violent crime rates in communities based on key demographic, socioeconomic, and law enforcement-related features. The model we decided to use was K-Means Clustering.



First we realized that we are working in high dimensionality, so we wanted to use PCA to scale our feature array so we get the most out of our model. Knowing this, we decided to calculate what n_components we wanted to use for PCA and we created the graph above to determine this. The graph above shows the relationship between rate of change of variance and rate of change in general. Subsequently, it calculates the explained variance and the rate of change of explained variance for different numbers of principal components. The resulting plot helps determine the optimal number of components to retain without creating overly large or overly small clusters. This preprocessing step is valuable for optimizing the representation of high-dimensional data before applying clustering techniques such as k-means. The visualization showed that 15 components were sufficient to retain since the rate of change starts to level off to 0 at that point.



To make our model more effective we performed Principal Component Analysis (PCA) to transform the data, selecting 15 components to ensure higher variance representation. By choosing PCA Component 1 to focus on, we find that there is a positive correlation with positive loading (correlation with variance) in the first original feature index. Given that PCA Component 1 gives the best representation of variance in our dataset, we can use this graph to show which features have the highest positive or negative correlation. This step aimed to capture the most crucial information from the diverse set of community characteristics, law enforcement metrics, and socioeconomic factors present in our dataset. To further validate the effectiveness of our model, we utilized the silhouette score, which measures the cohesion and separation of clusters. This step ensured that the clusters generated by K-Means were well-defined and provided meaningful distinctions between community profiles. The combined application of K-Means clustering and rigorous validation techniques contributes to the reliability and interpretability of our model, shedding light on the complex dynamics between community features and violent crime rates. These findings are crucial for informed decision-making in the realms of community development and law enforcement strategies.



We validated our model through the use of Silhouette score, getting a silhouette score of 0.18. A score of 0.18 shows a low level of separation and cohesion among the clusters generated. This is due to the fact that all the clusters had significant overlap with their influence on violent crime rate. The obtained silhouette score implies that our model's clustering configuration has a moderate degree of appropriateness. The PCA plot is used in hand with our silhouette score of 0.18 to show how much our clusters vary. The above visualization is a 2 dimensional space where each point in the plot corresponds with an observation. The colors are meant to group similar observations based on their features in each PCA component; the clusters on the positive end of the Principal component suggest that those points are similar with respect to the features that contribute positively to the first principal component. This visualization is very important because it tells us that between each category of violent crime rate, it shares many characteristics and variance amongst each of the different categories. The overlap tells us that each category shares similar features with other categories, which lets us know that the split between categories is a lot more blurred than we thought it was.

Conclusion

Our analysis of crime rates and various socio-economic factors has provided valuable insights into how these variables interact. The positive correlation identified between socioeconomic features and violent crime rates underscores the nuanced relationship that exists within communities. Using methods such as Chi Squared Analysis, Pearson correlation coefficient hypothesis tests, and linear regression lines, we have successfully shown the significance of these correlations.

Our investigation into the demographic aspects has revealed a substantial correlation between certain demographic features and violent crime rates. Utilizing Pearson correlation coefficient hypothesis tests and examining linear regression lines for each race against crime rates, we found that for certain ethnic populations that suffered systemic racism among other systemic issues, there was a positive slope for violent crimes on the regression line as the population of those demographics increased, while there was a decrease with more privileged areas..

Our findings were surprising, as it indicated no correlation between the number of police officers per 100K and crime rates. Employing Chi Squared Analysis and exploring linear regression lines, we found that the correlation was 0.3402 which is not significant enough for us to claim correlation.

However, the exploration of police activity has yielded intriguing results. We identified a correlation between police activity and the percentage of people living under the poverty line, shedding light on the socio-economic factors that influence law enforcement engagement. We found this by plotting linear regression lines and found that as the data points increased, so did the slope of the linear regression line.

Furthermore, our investigation into the urban environment and police presence revealed an unexpected result – no discernible correlation between living in an "urban" area and police presence. The linear regression lines did not have any data that fit above or below it, showing that there is significant correlation between police presence and urban environment.

Our project has not only provided statistical evidence but has also raised thought-provoking questions about the influences on crime rates. When comparing the Relationship between Non-high school graduates, unemployment, and violent crimes via a scatter plot, the number of violentCrimesPerPop was at the highest when the percentage of non-high school graduates and unemployment rates were the highest. This research serves as a call to action for communities to be more proactive in keeping their children in school. By doing so, they can positively impact not only the future of individuals but also contribute to the overall well-being and safety of the community.

The insights gained from this study offer valuable guidance for policymakers, educators, and community leaders in formulating strategies that address the root causes of crime and promote sustainable community development. Insights such as the higher the number of police requests coming in directly lead to lower crime rates can be used to help evaluate the Police force and and educate policy makers on what to continue to fund and support and what needs to be changed.

Contributions

Ynah Novilla	- Project Proposal - Data Exploration/Observations - Hypothesis 2: Visualizations and correlations - Final Report - Final Slides
Jonathan Darius	- Project Proposal - Data Cleaning - Filled in missing values using Regression Imputation - Supervised Learning: Visuals, correlation, observations - Unsupervised Learning: Visuals and correlations
Ojasvi Godha	- Data Exploration/Observations - Hypothesis 1: Visualizations and correlations - Hypothesis 3: Visualizations and correlations - Final Report - Final Slides
Adreyan Distor	- Introduction Presentation Slides

	- Data Exploration/Observations - Hypothesis 4: Visualizations and correlations - Hypothesis 5: Visualizations and correlations
Kushal Tirupathi	- Data Cleaning - Filled in missing values using KNN Imputation - Unsupervised Learning: observations - Final Slides - Final Report