



مدينة زويل للعلوم والتكنولوجيا  
Zewail City of Science and Technology

# Final Project - Business Report

CIE 457

Prepared For  
Dr. Mahmoud Abdelaziz

Supervised by  
Eng. Anhar Hassan - Eng. Asmaa Mahmoud

Prepared By

Ahmed Mahmoud	201800683
Ahmed Elghamry	201801254
Ibrahim Hamada	201800739
Hazem Tarek	201800283

## **Table of content :**

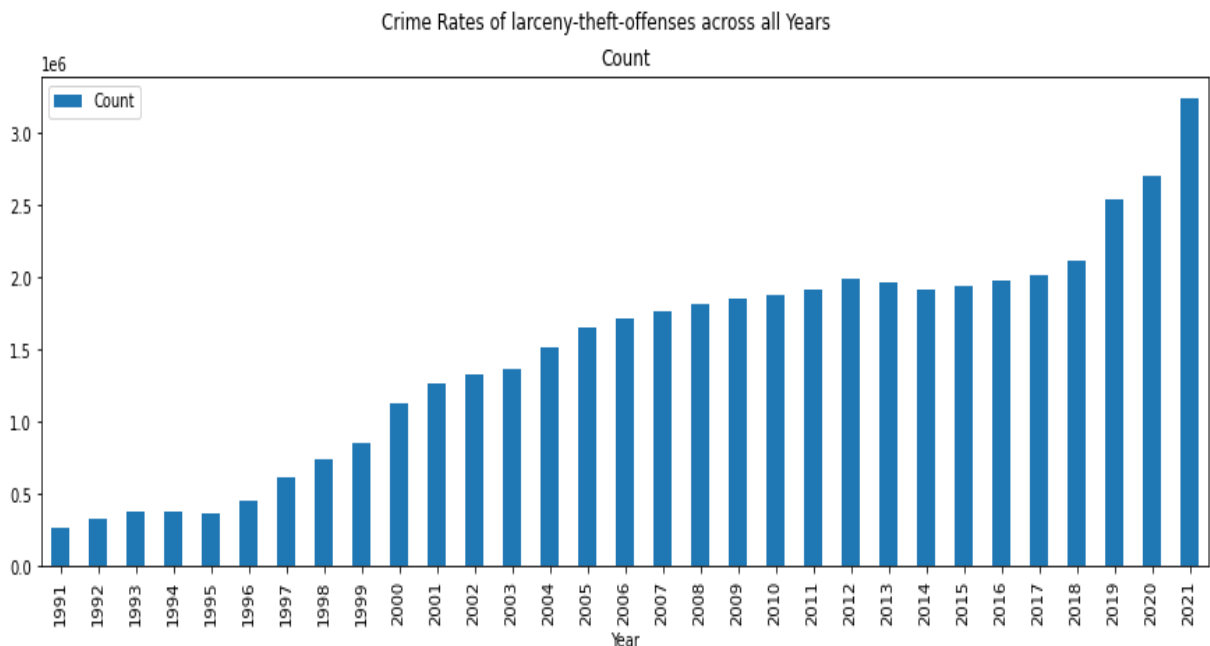
<b>Introduction</b>	<b>3</b>
<b>Exploratory Analysis</b>	<b>3</b>
<b>Question Answering</b>	<b>15</b>
<b>Hypothesis Testing</b>	<b>23</b>
<b>Regression Analysis</b>	<b>25</b>
<b>Bonus Task</b>	<b>31</b>

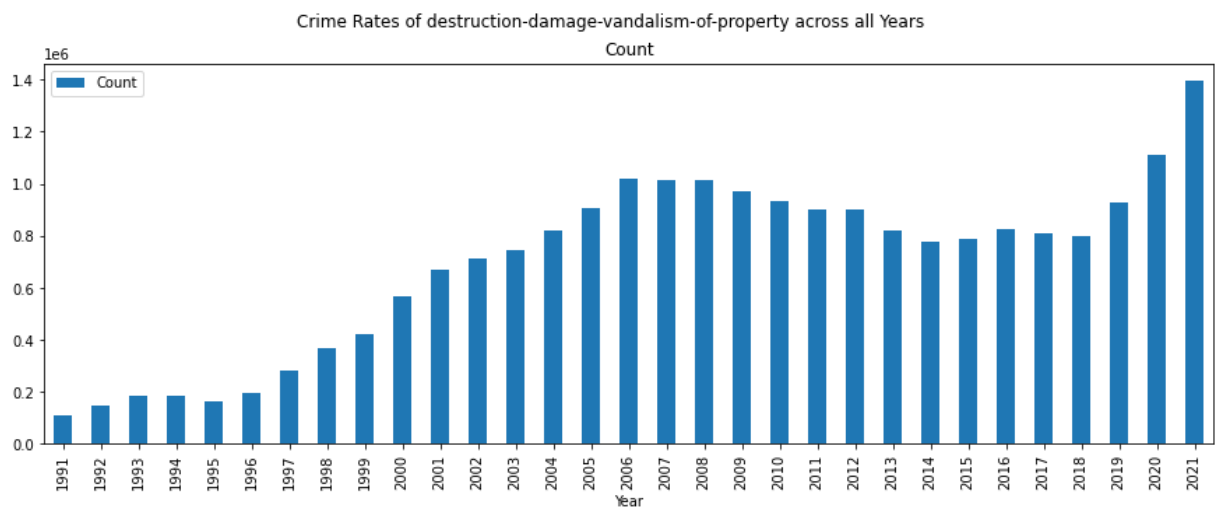
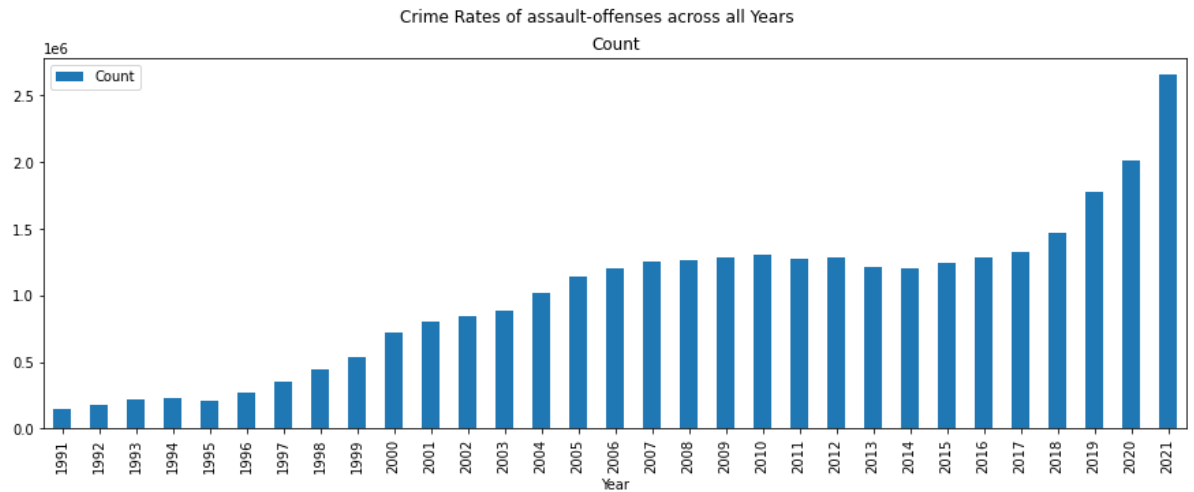
# Introduction

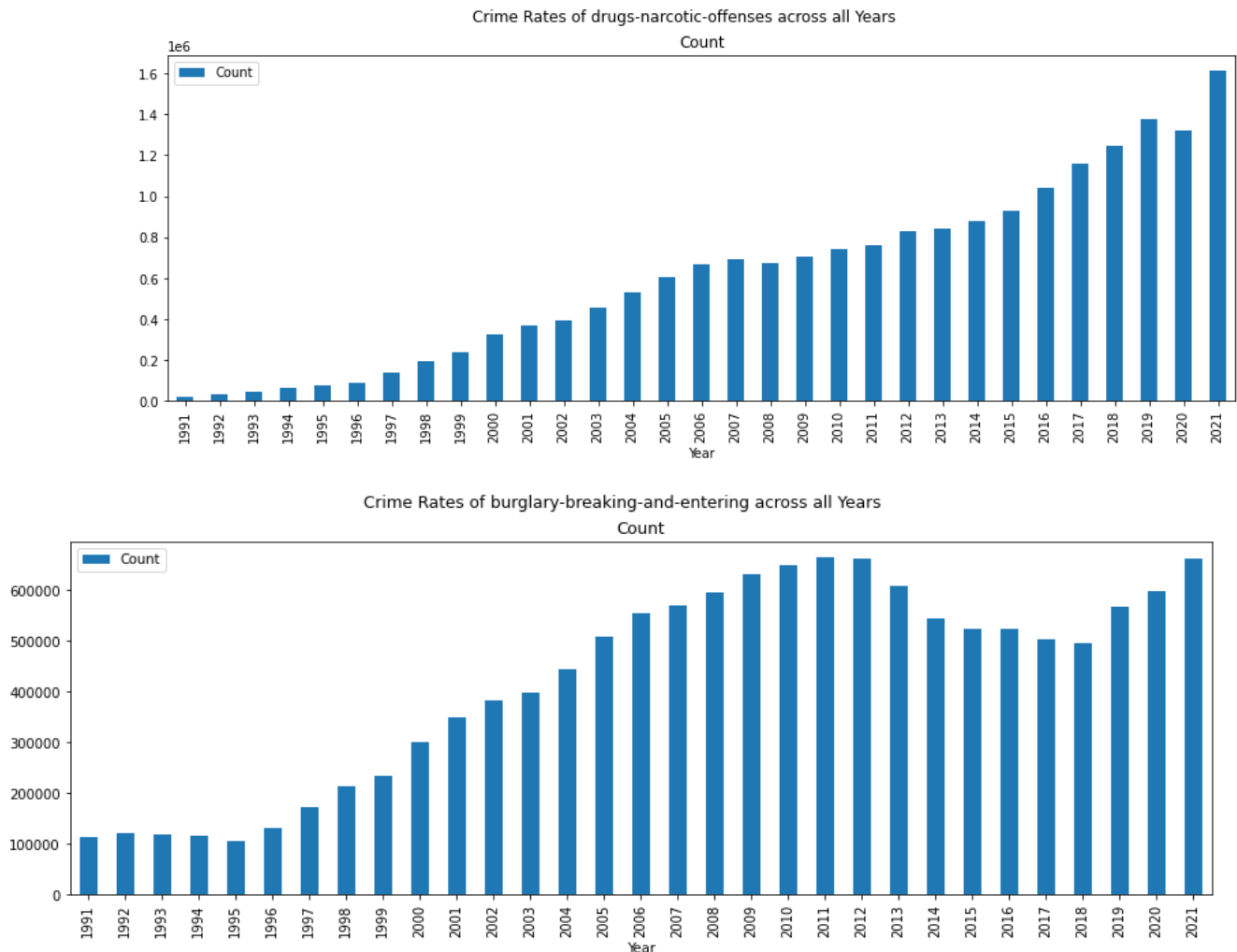
In this project, we will be analyzing four different datasets: The national crime victimization survey (NCVS) data, NIBRS Reported offense count data, Recidivism data for the state of Georgia from 2013 to 2015, and data on firearm laws per state. We will be examining these datasets in order to gain a deeper understanding of crime patterns and trends, as well as the effectiveness of different firearm laws. By the end of this project, we hope to be able to provide valuable insights and recommendations based on our findings.

## Exploratory Analysis

1. National criminal offense rates per year across all available years for the top five most frequent offense categories.

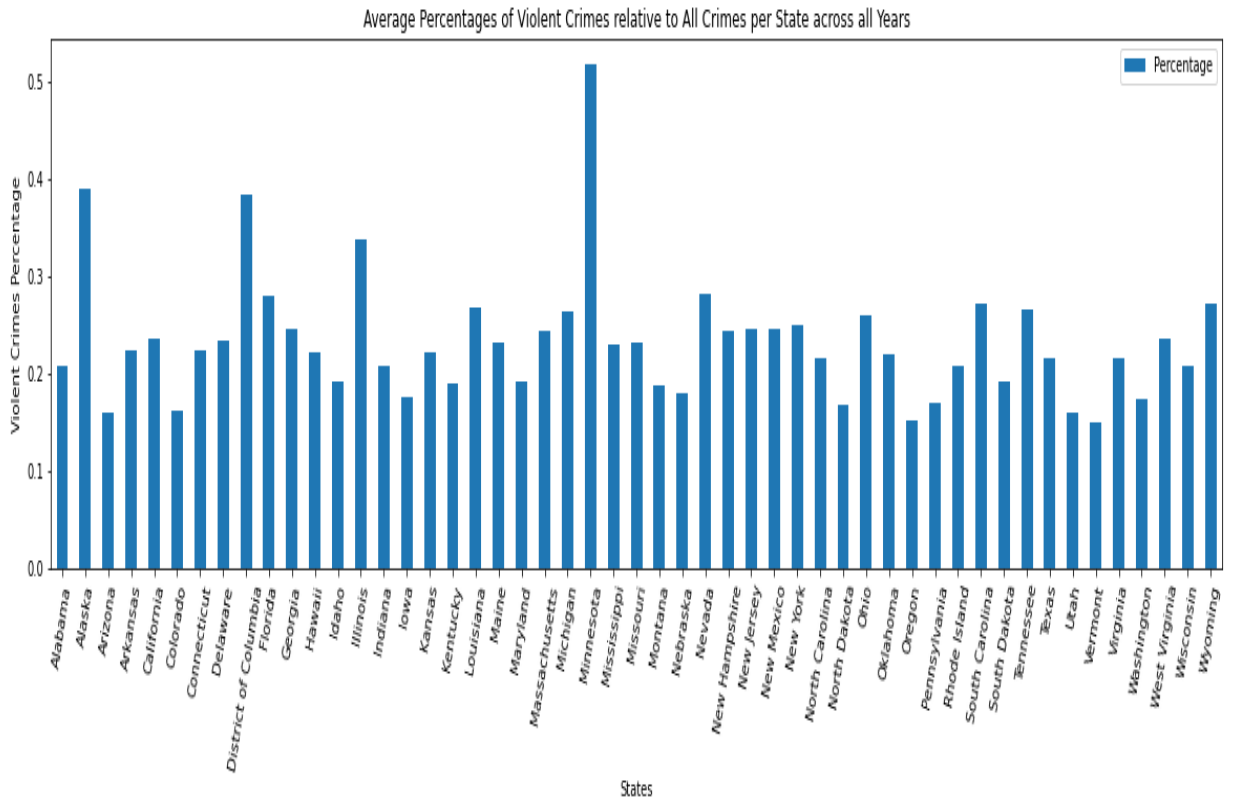






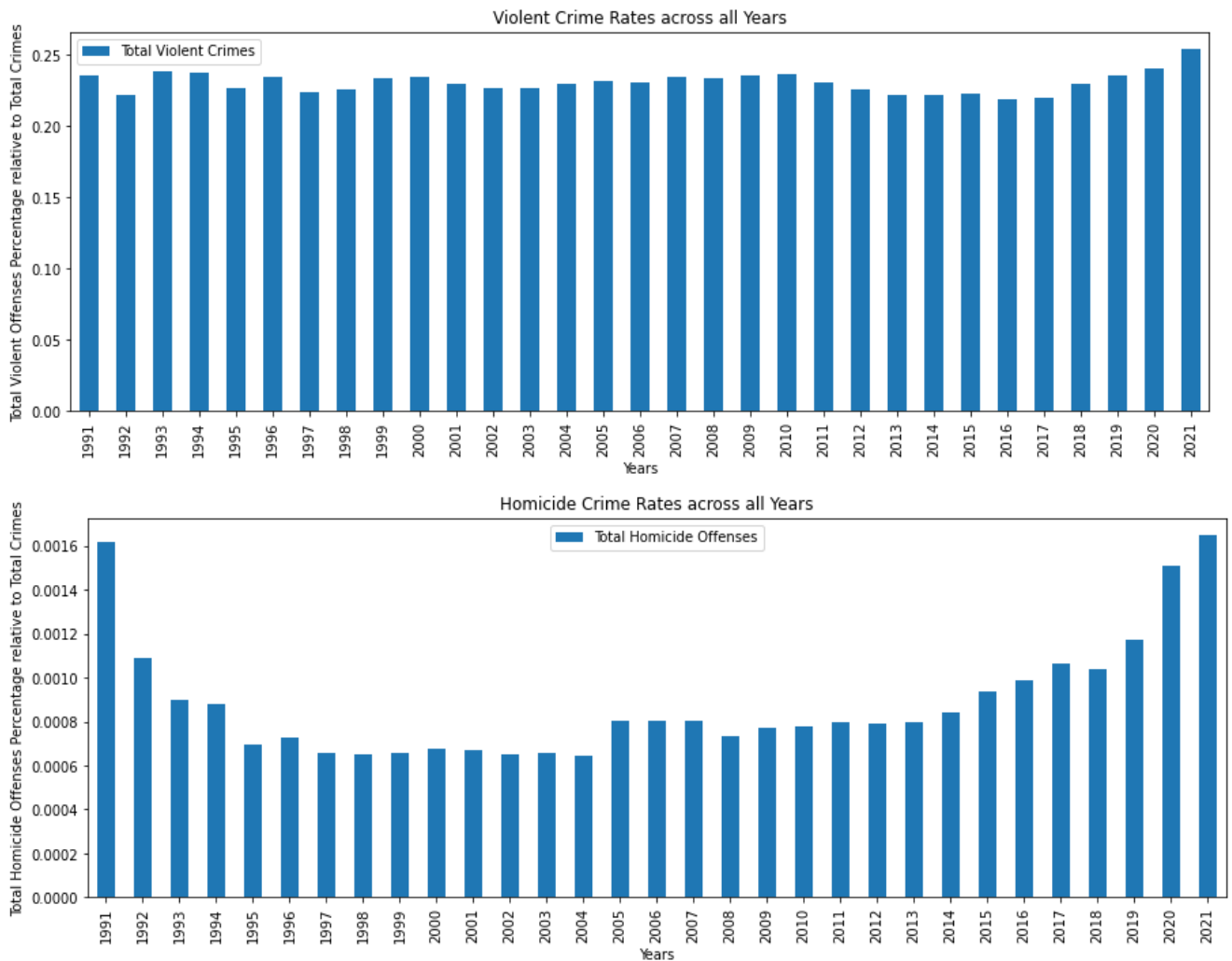
Comment: The crime rates of the top 5 crime categories, which are "larceny-theft-offenses", "assault-offenses", "destruction-damage-vandalism-of-property", "drugs-narcotic-offenses", and "burglary-breaking-and-entering", are increasing along the time. However, some categories have reduced in the middle of 2010s except Drug/Narcotic Offenses. The normalization did not differ in the representation of the crime rates as the population increases approximately linearly.

2. The average percentage of violent crimes relative to total crime per state over all available years.



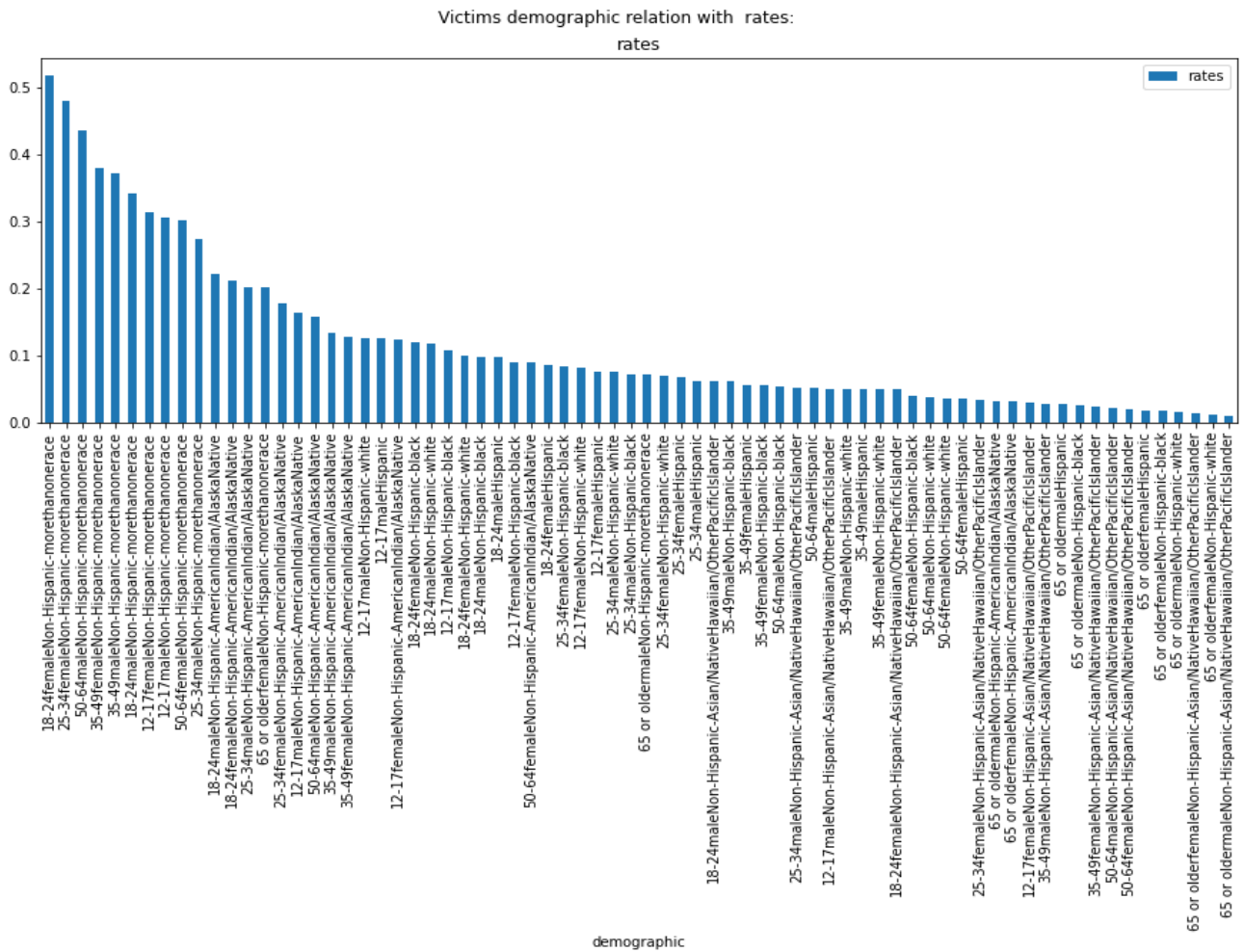
Comment: The average percentage of violent crimes over all years relative to all crime rates shows that the highest state in violent crimes is Minnesota of 50% while the lowest is Vermont.

3. National homicide rates, and total violent crime rates per year over all years.

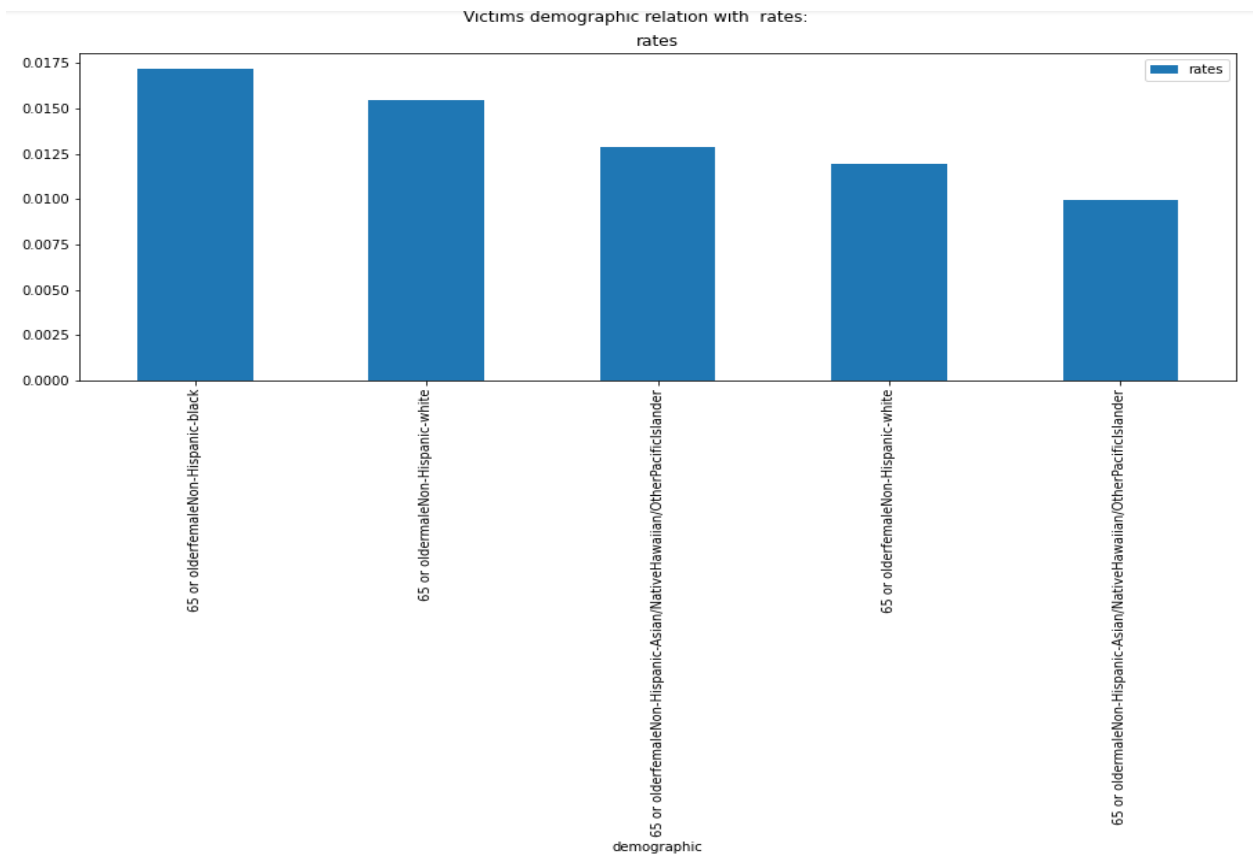
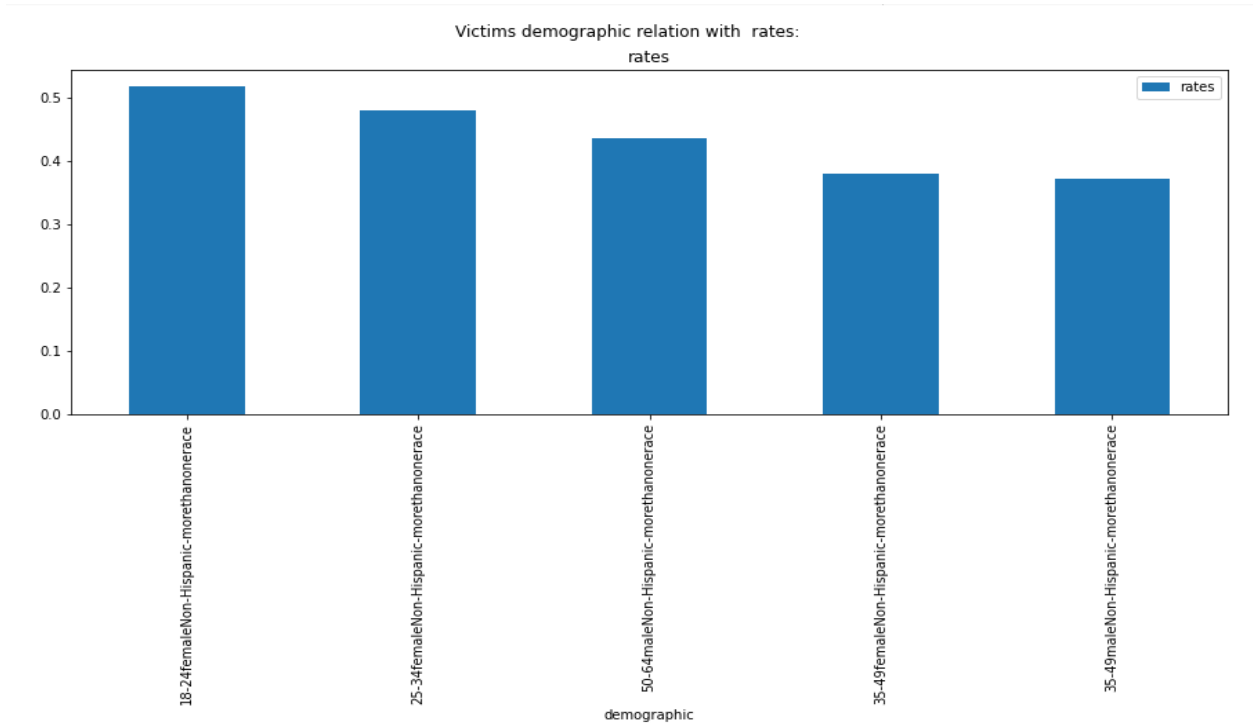


Comment: National homicide rates have been decreasing starting from 1991 till 1997, then it remained approximately constant till 2005. Then, it started to increase till it reached its maximum in 2021. However, the violent crime rates oscillated between 20%-25% over 30 years, which seems to be approximately constant as it has changed a little over the years.

4. The frequency of non-fatal crime incidents in relation to victim demographics.

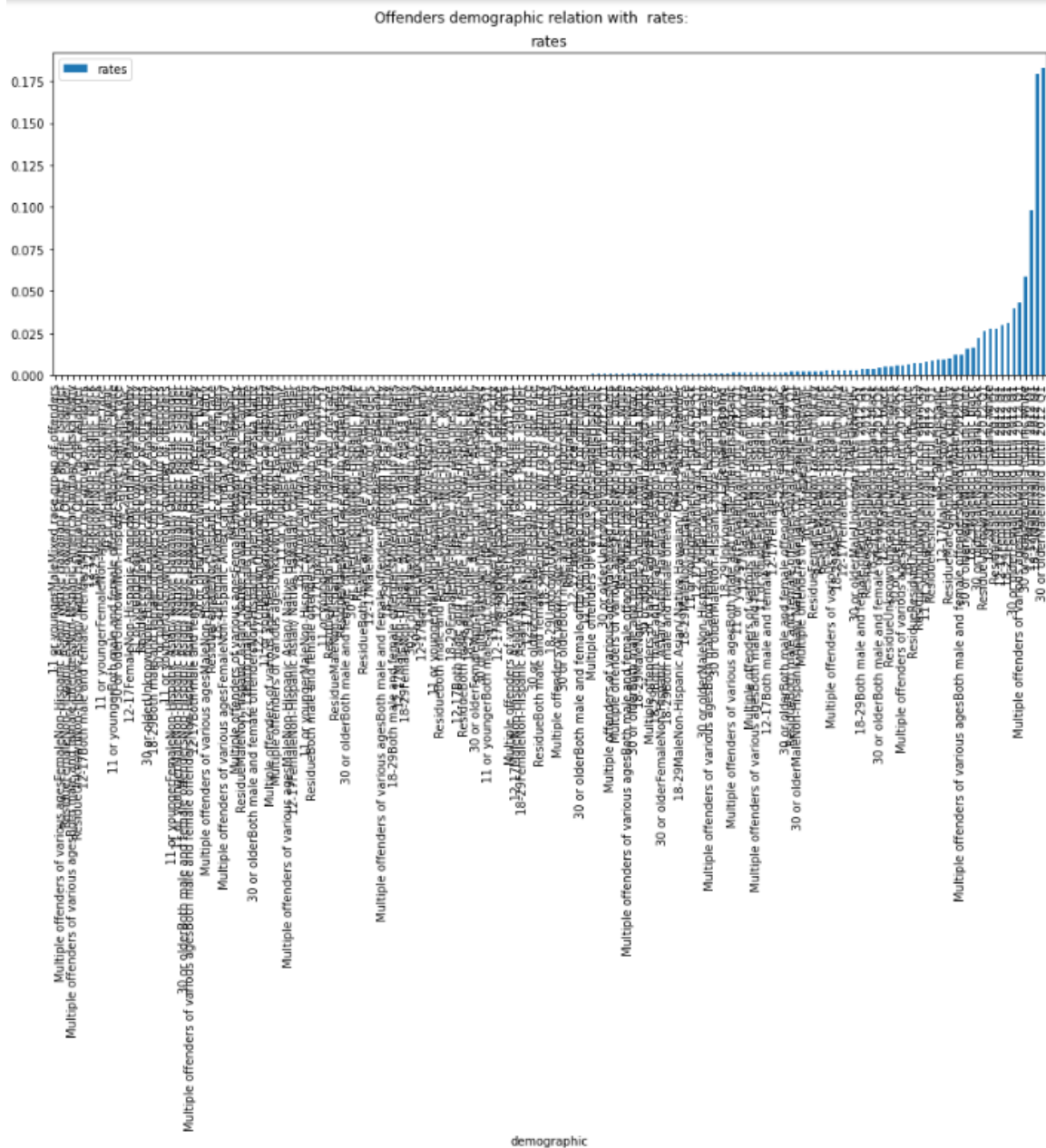


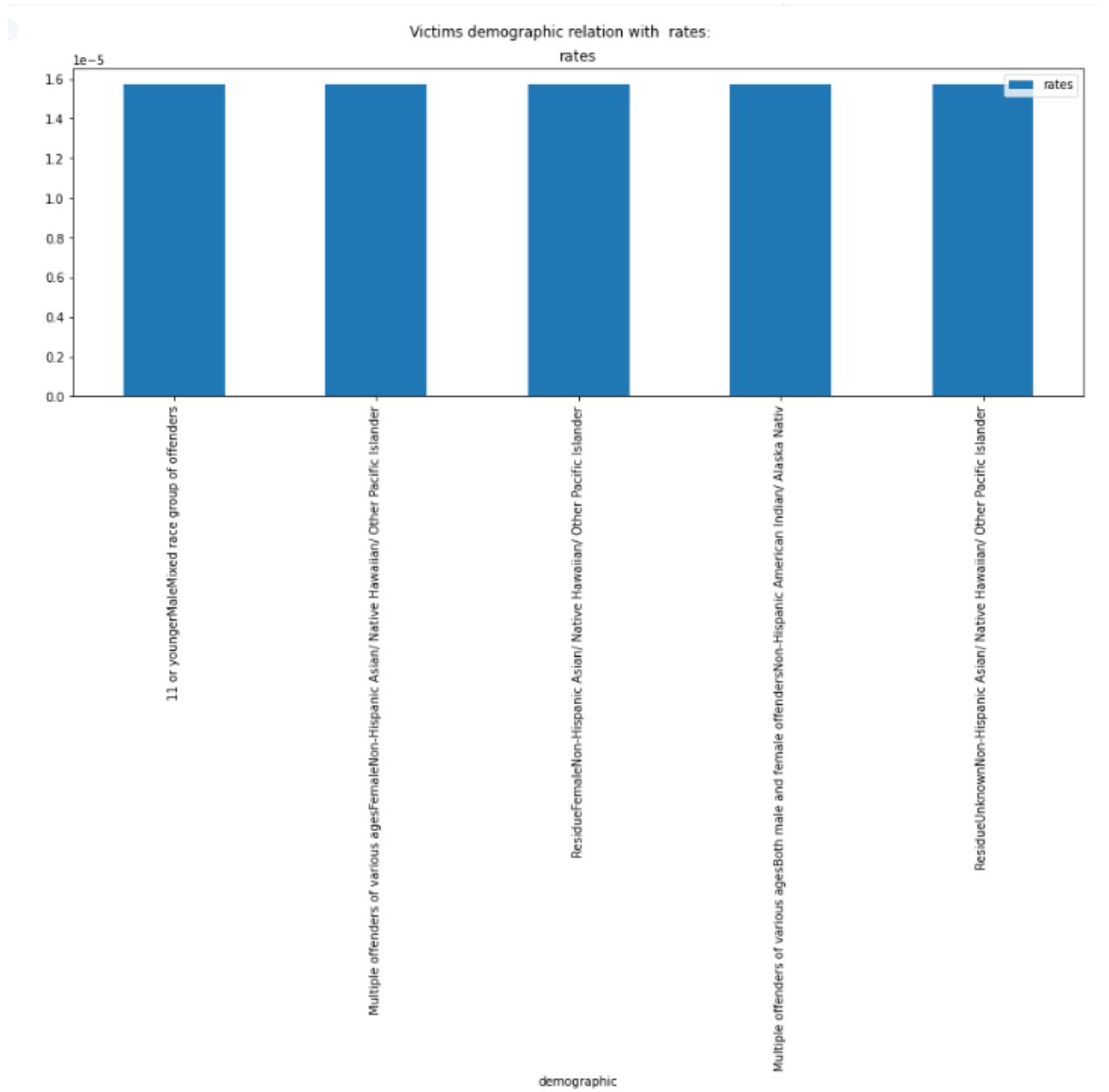


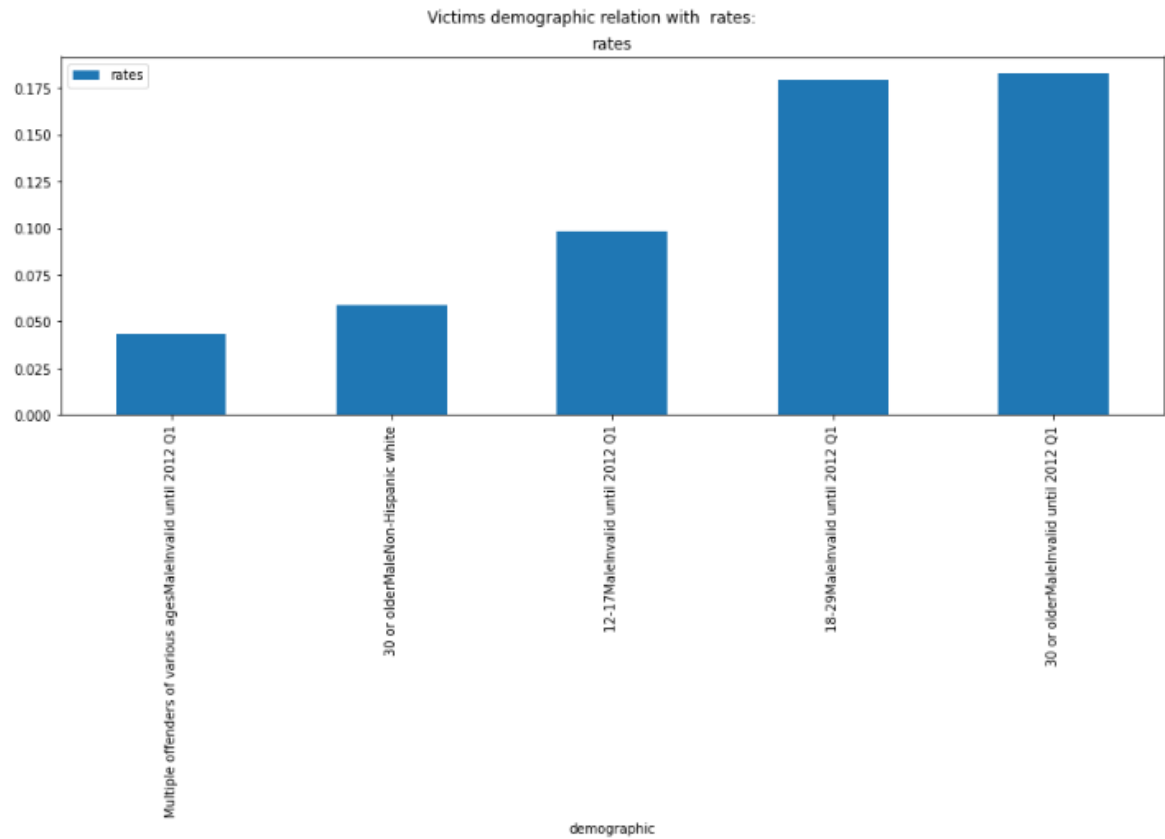


Comment: It is apparent that non Hispanic with more than one race are the most targeted victims (females 18-49 years, and males 25-49) and the lowest targeted victims are older than 65 years non Haspanic males and females of different races (black,white,Asian).

5. The frequency of non-fatal crime incidents in relation to offender demographics.

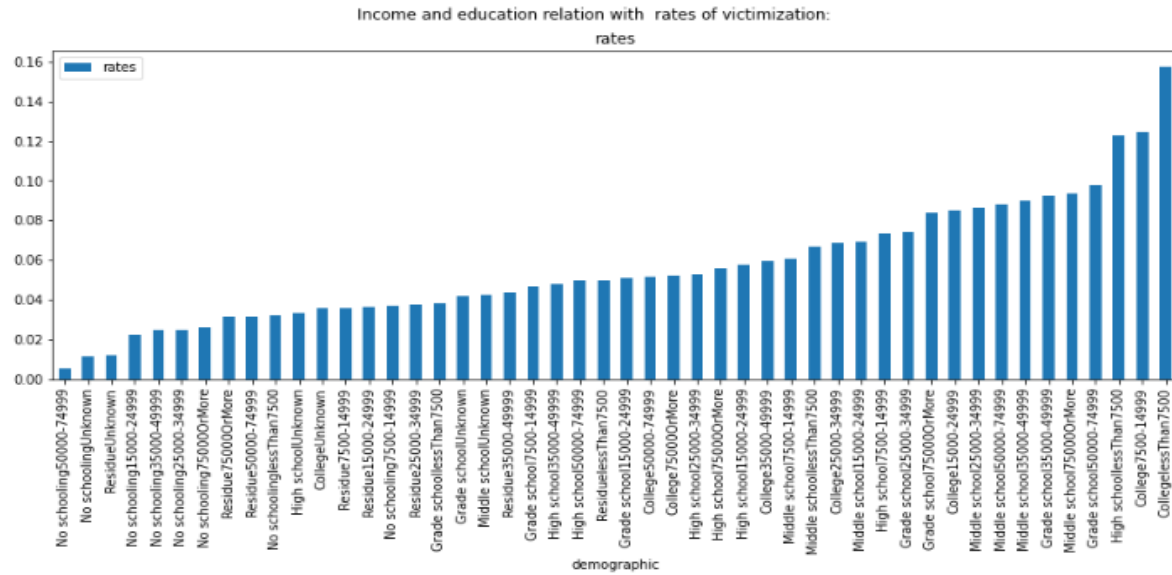


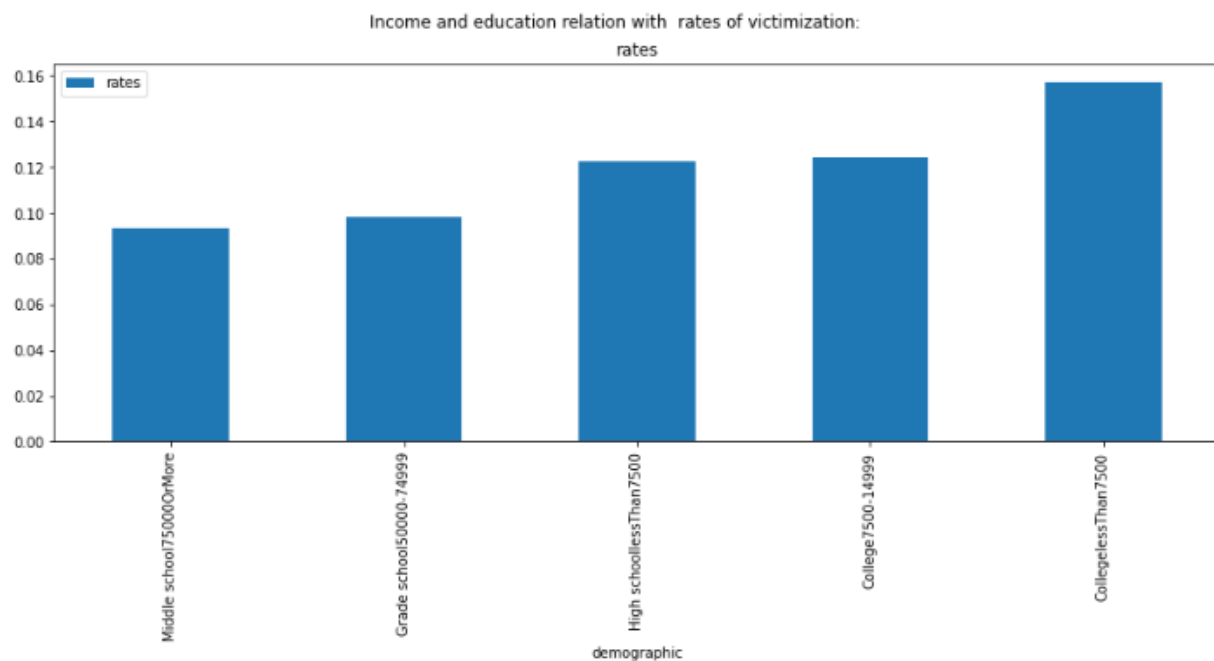
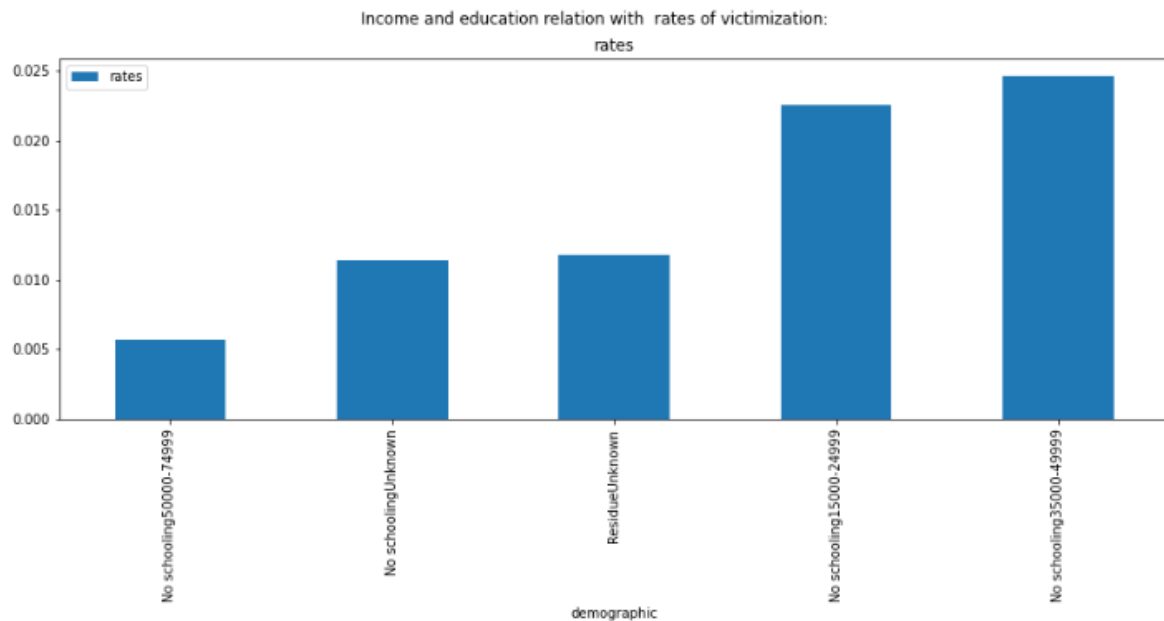




Comment:lowest demographics rates of offender are 11 or younger male group of offenders with more than one race and the highest rates of offenders demographics are in 30 or males

6. The relationship between the victim's education level, their gross household income, and their rate of victimization.

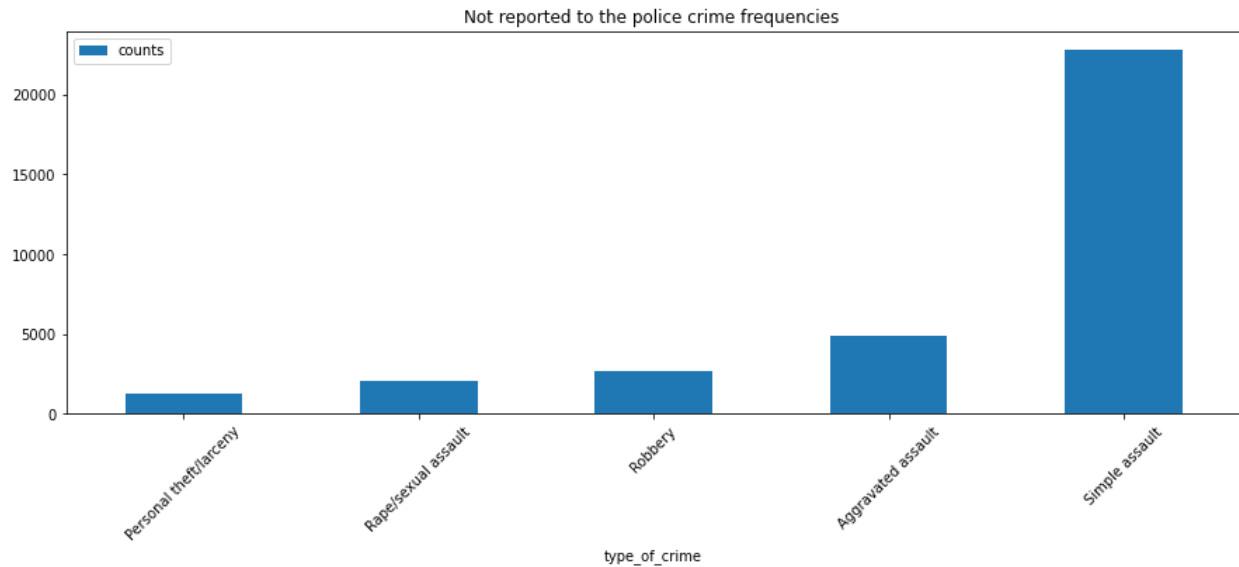




Comment: As shown, being in a college with a moderate income (between 7500-15000) are the most frequent victims, while victims who had no schooling and high income are the least frequent

# Question Answering

1. Which type of non-fatal crime is the most under-reported? Is there an association between the offender-victim relationship and the likelihood of a crime being reported?

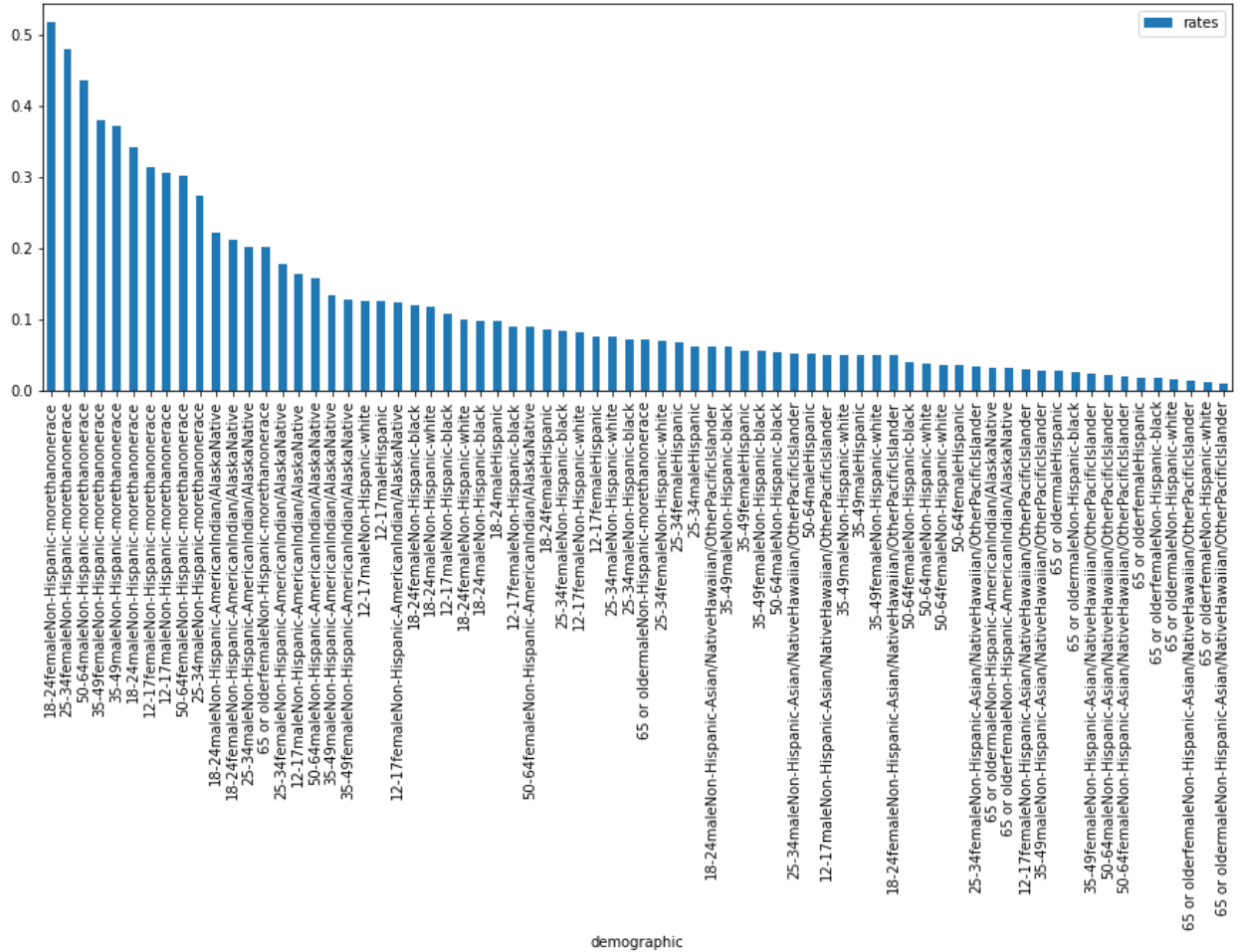


Comment: Simple assault is the most under-reported type of crime.

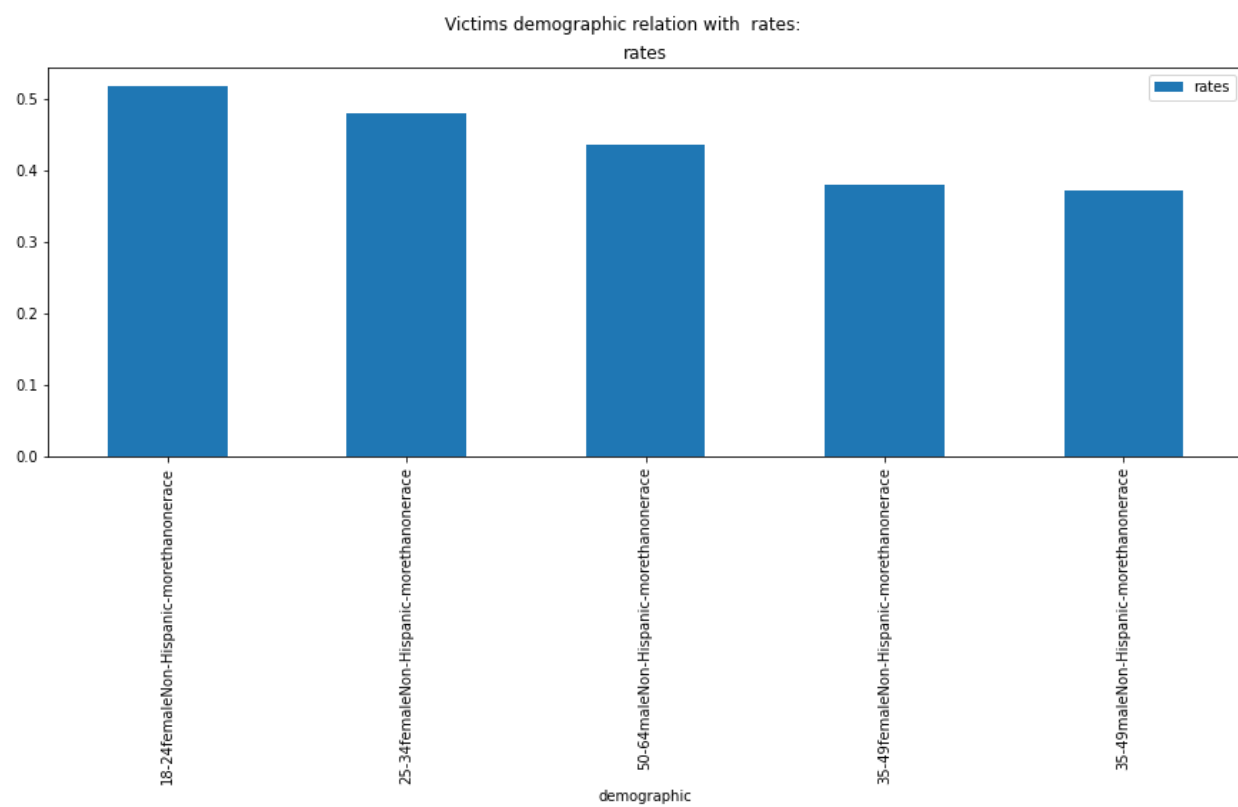
2. Who are the people (the demographic segment) that appear to be most at risk of violent victimization? Who is the least at risk?

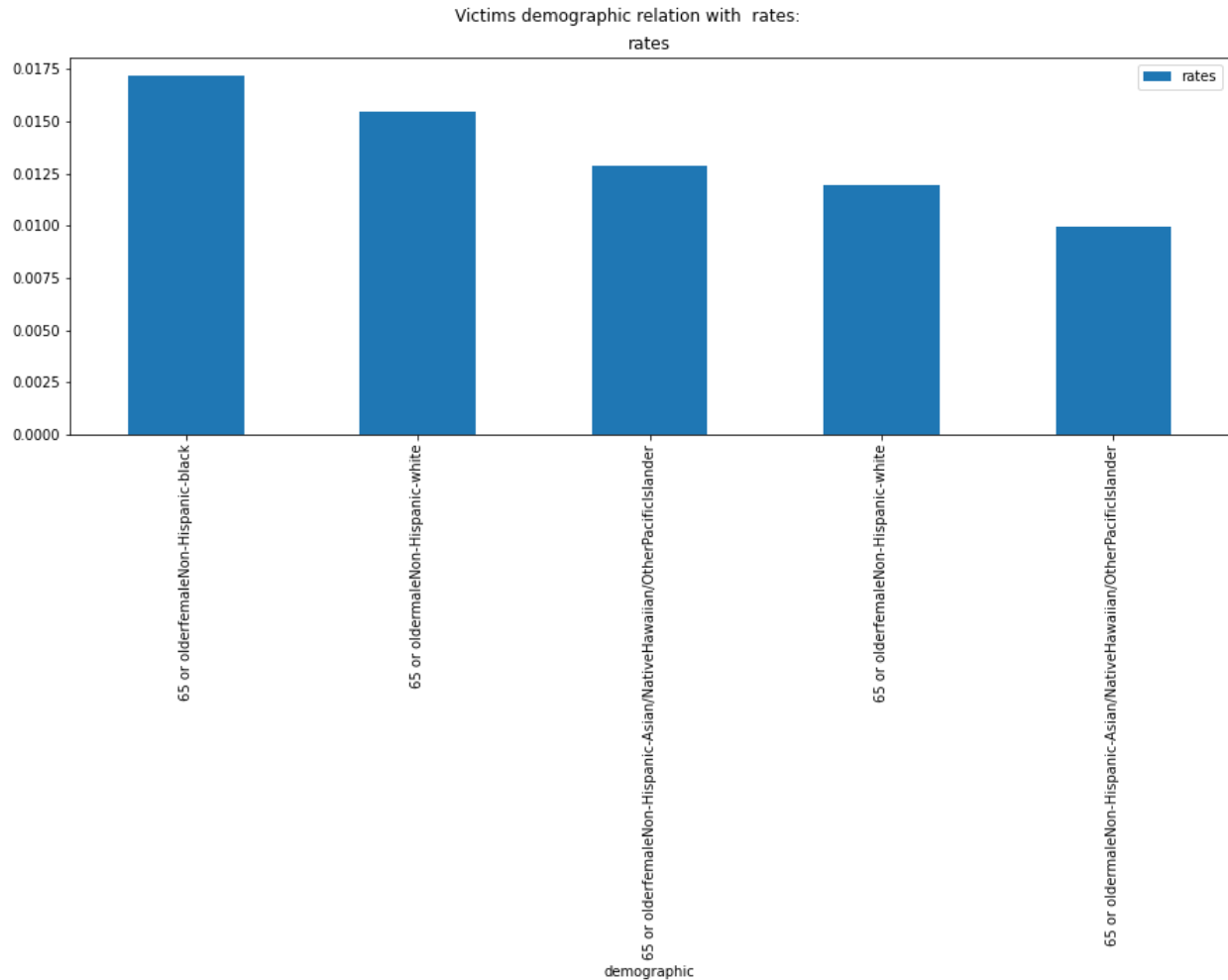
Victims demographic relation with rates:

rates





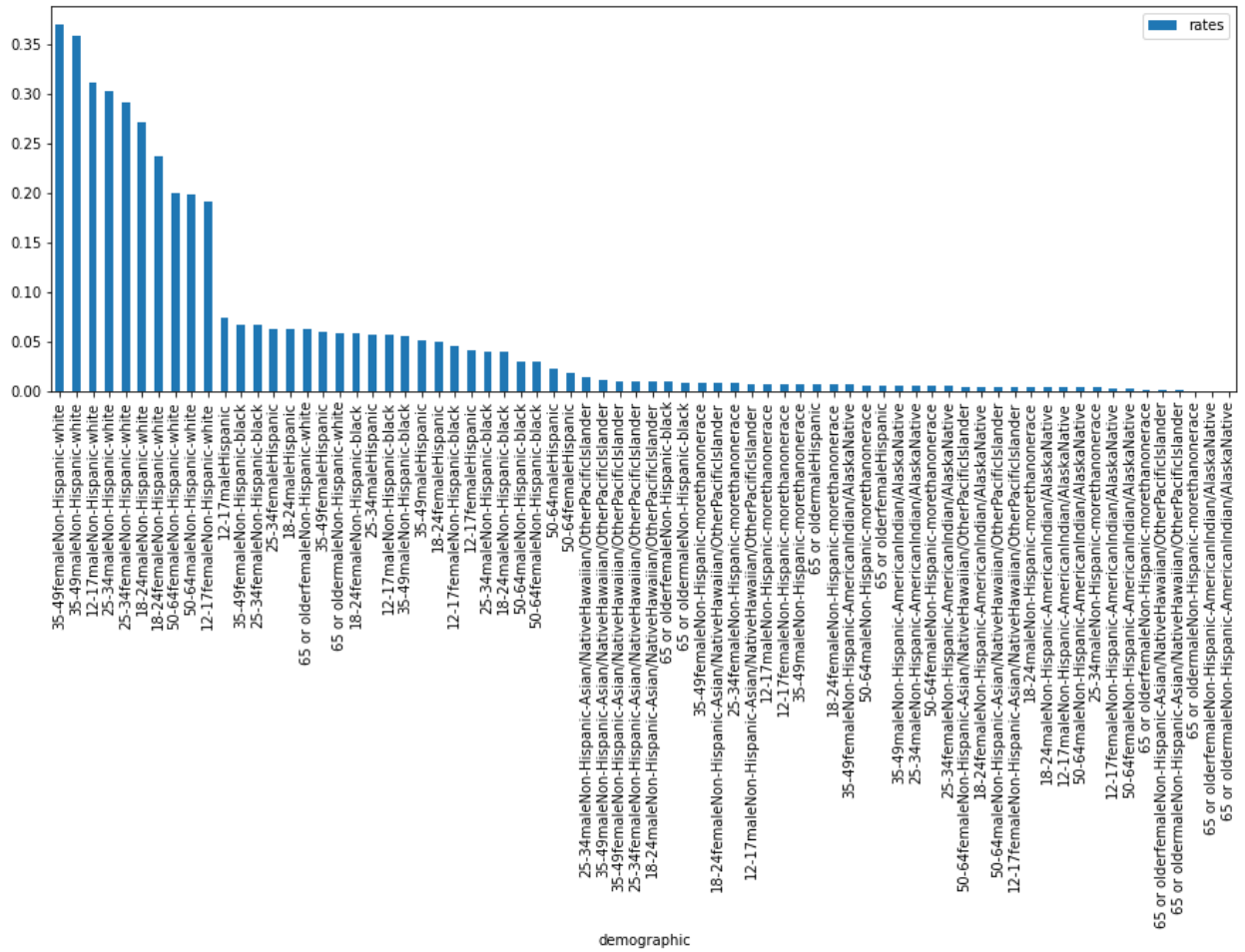


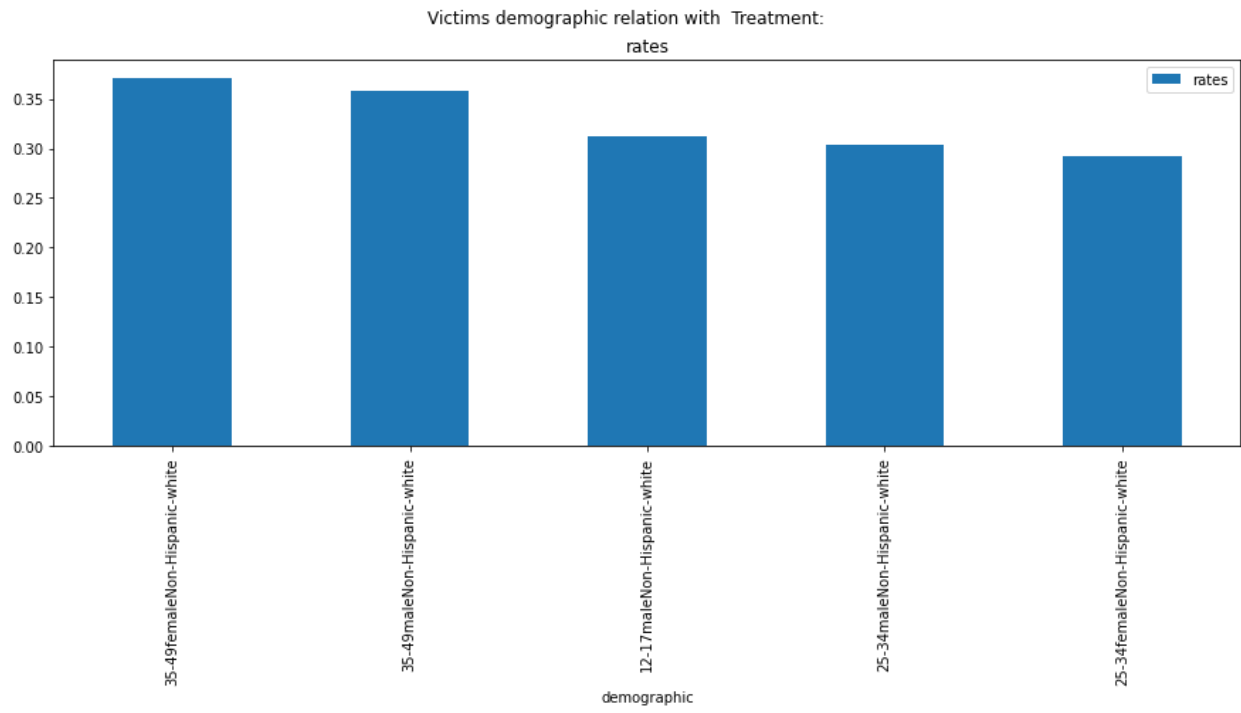


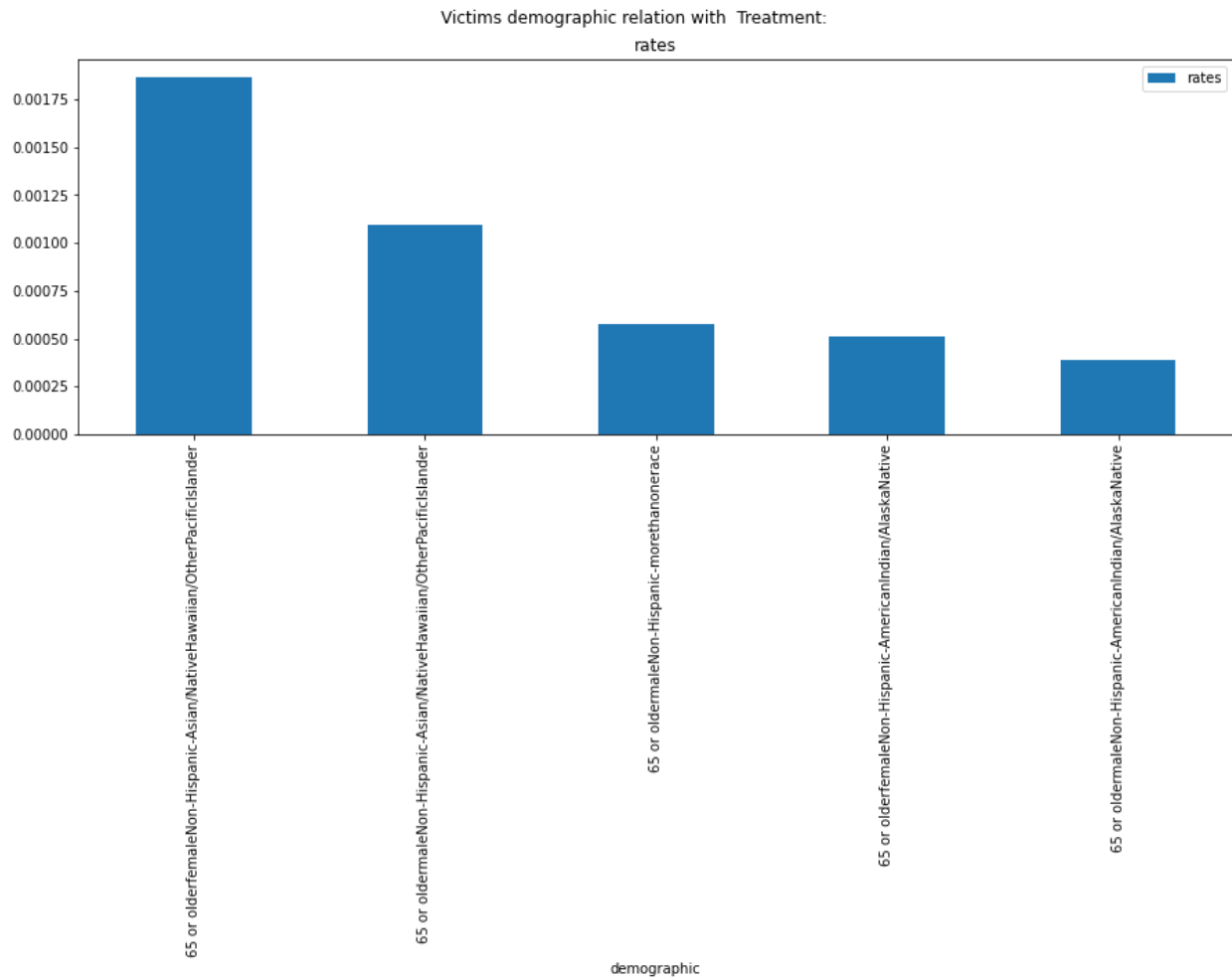
Comment: Non Hispanic females of more than one race of 18-24 years are the most people at risk of violent victimization, while non Hispanic Asian males of 65 years or more are the least at risk of violent victimization.

- Of all victims of non-fatal crimes who suffer an injury, which demographic is the most likely to receive medical attention at the scene? Which is the least likely?

Victims demographic relation with Treatment:  
rates

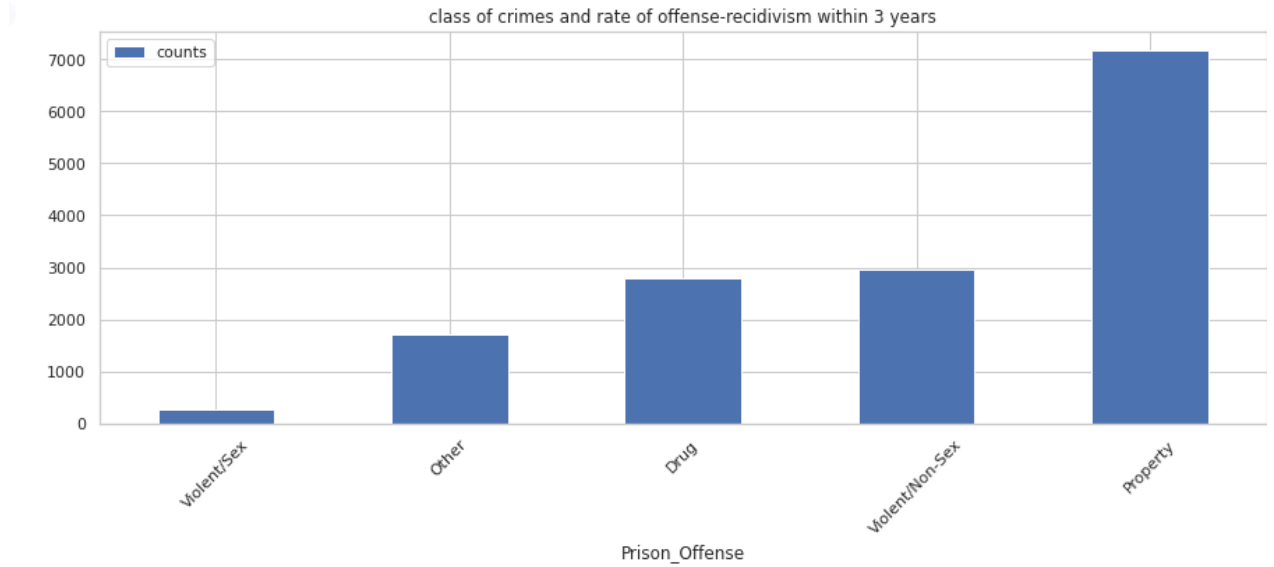






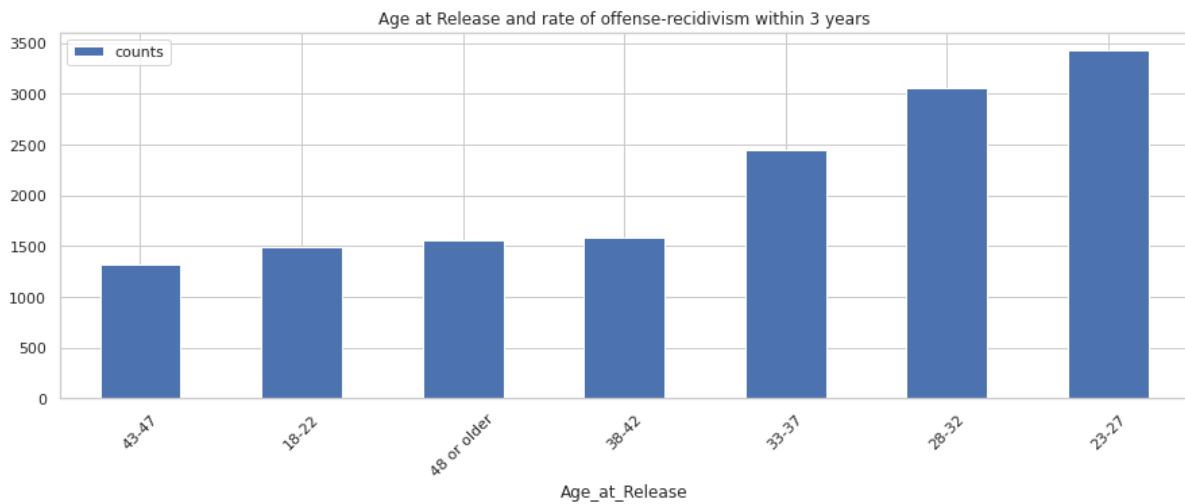
Comment: Non-Hispanic white females of age 35-49 are most likely to be medically attended while suffering from an injury, while the 65-year-old or more non-hispanic American men are least likely to receive medical attention whilst their injury.

#### 4. Class of crimes and rate of offense-recidivism within 3 years



Comment: the plot shows that the property category is associated with the highest rate of Same-offense-recidivism.

#### 5. Age at Release and rate of offense-recidivism within 3 years



Comment: prisoners who are younger at the time of release more or less likely to reoffend than those who are older.

# Hypothesis Testing

## A. Testing Claim 1

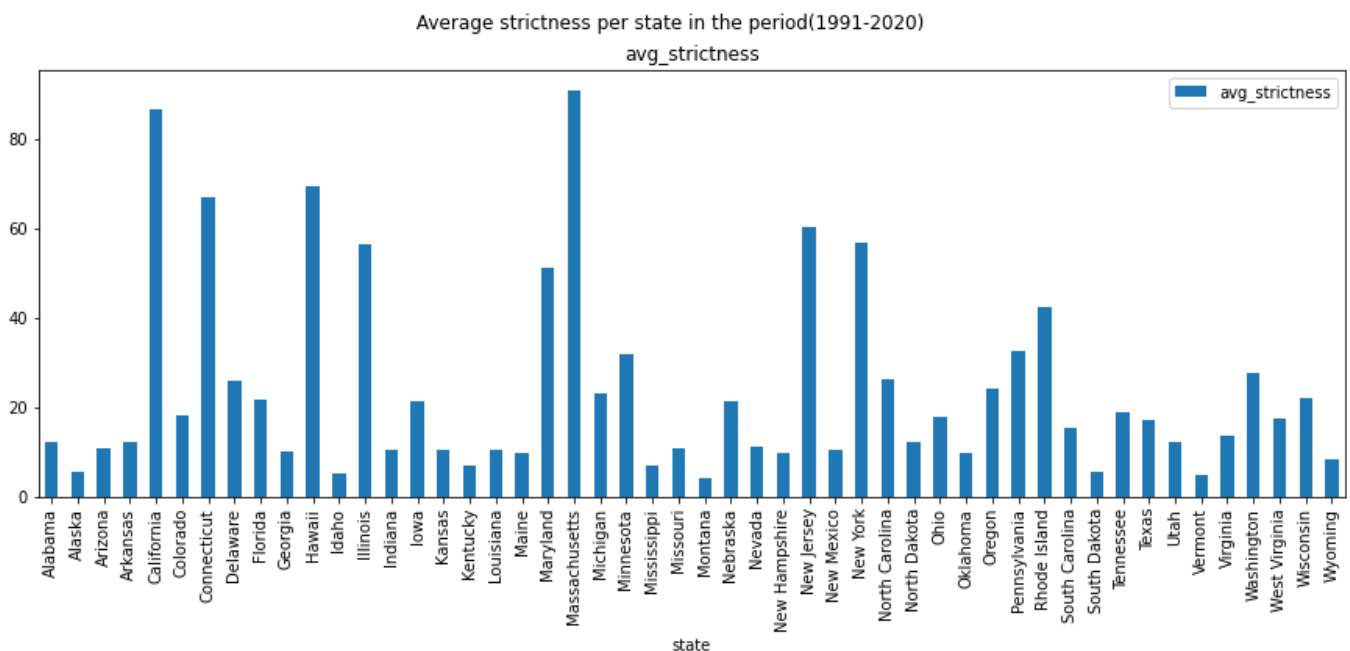
Claim: U.S. states that implement stricter firearm control laws, have lower violent crime rates on average.

Steps: the states are divided into two categories, heavily strict states and less strict states and the threshold that decides will be the mean of the average strictness per state across all years.

Null hypothesis  $H_0$ : there will be no difference in the average crime rate between heavily strict states and the other states.

The test: the test used here is the T-test as we are comparing the means of two different groups.

Test result: the p-value = 0.544, which means that the null hypothesis cannot be rejected. Therefore, we can deduce that American society has violent behavior regardless of laws being strict or not.



Comment: the figure shows the average strictness per state in the period between (1991-2020).

## B. Testing Claim 2

Claim: Black people are assigned a high risk score compared to white people.

Steps: the criminals from the dataset "Recidivism data for the state of Georgia [2013-2015]" will be divided into two categories; Blacks and Whites.

Null hypothesis  $H_0$ : there will be no difference in the Supervision Risk Score between Black and White people.

The test: the test used here is the T-test as we are comparing the means of two different populations.

Test result: the p-value = 0.999, which means that the null hypothesis cannot be rejected. Therefore, we cannot say that the blacks are assigned a high risk score compared to the whites.



# Regression Analysis

## 1. Model's coefficients and p-values

Column Name	Coefficient	P-value
Intercept	8.7584	0.000
Race_BLACK	-0.0360	0.195
Age_at_Release_23-27	-0.4670	0.000
Age_at_Release_28-32	-1.2003	0.000
Age_at_Release_33-37	-1.9663	0.000
Age_at_Release_38-42	-2.5944	0.000
Age_at_Release_43-47	-3.1319	0.000
Age_at_Release_48 or older	-4.0104	0.000
Gang_Affiliated_True	0.4405	0.000
Prior_Conviction_Episodes_Felony_0	-0.0155	0.760
Prior_Conviction_Episodes_Felony_1	-0.1147	0.012
Prior_Conviction_Episodes_Felony_2	-0.1245	0.005
Prior_Conviction_Episodes_Misd_0	0.6543	0.000
Prior_Conviction_Episodes_Misd_1	0.5005	0.000
Prior_Conviction_Episodes_Misd_2	0.4426	0.000
Prior_Conviction_Episodes_Misd_3	0.2322	0.000
Prior_Conviction_Episodes_Viol_True	0.2724	0.000
Prior_Conviction_Episodes_Prop_0	-1.6509	0.000
Prior_Conviction_Episodes_Prop_1	-1.0760	0.000
Prior_Conviction_Episodes_Prop_2	-0.6516	0.000
Prior_Conviction_Episodes_Drug_0	-0.6604	0.000
Prior_Conviction_Episodes_Drug_1	-0.4225	0.000
Prior_Conviction_Episodes_PPViolationCharges_True	0.3391	0.000
Prior_Conviction_Episodes_DomesticViolenceCharges_True	-0.2458	0.000
Prior_Conviction_Episodes_GunCharges_True	0.038	0.000

## 2. Good and Bad predictors

Variables that have p-value  $< 0.05$  are considered as good predictors, while variables that have p-value  $> 0.05$  are bad predictors.

Good Predictors:

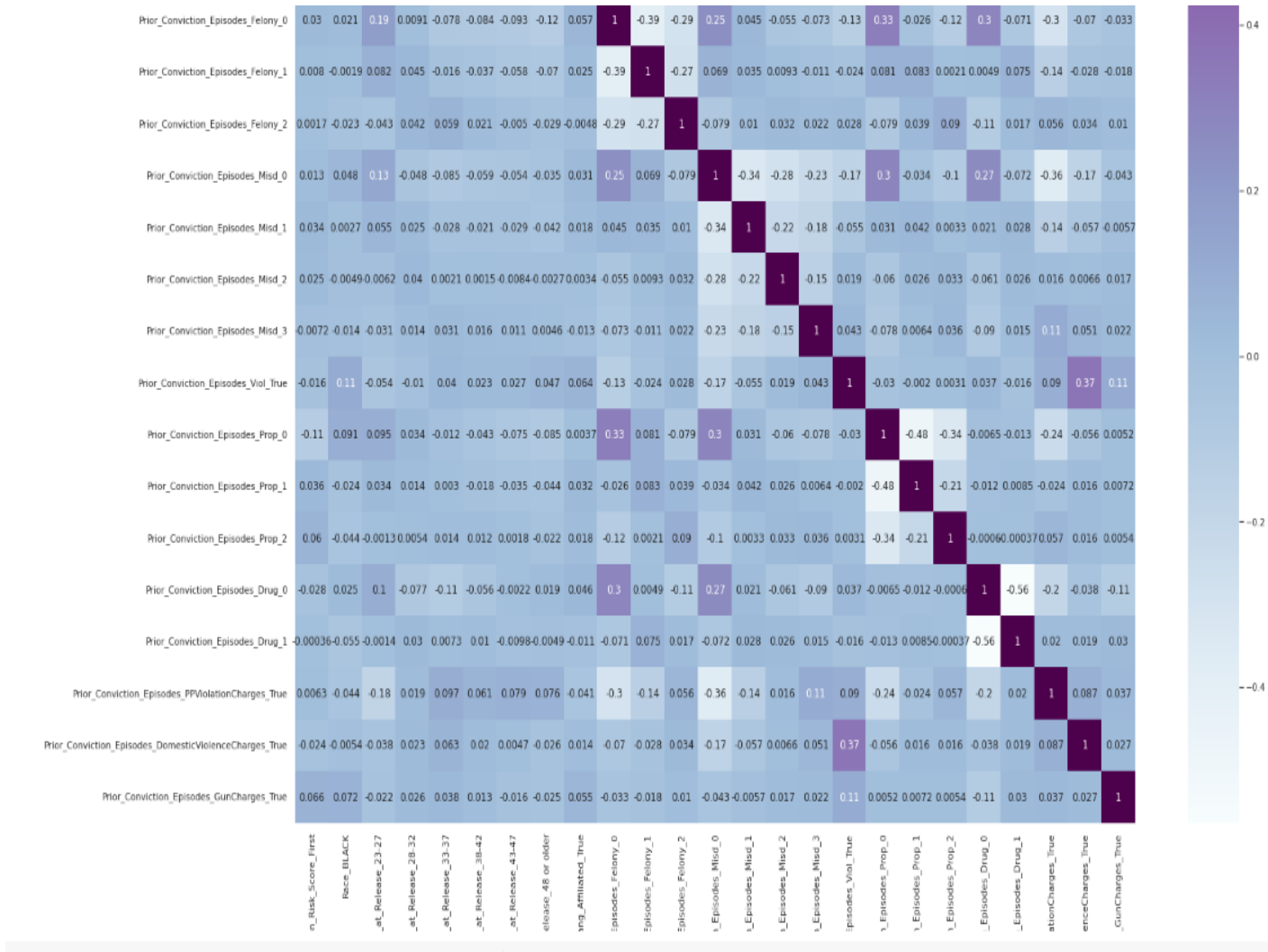
- Intercept
- Race\_BLACK
- Age\_at\_Release\_23-27
- Age\_at\_Release\_28-32
- Age\_at\_Release\_33-37
- Age\_at\_Release\_38-42
- Age\_at\_Release\_43-47
- Age\_at\_Release\_48 or older
- Gang\_Affiliated\_True
- Prior\_Conviction\_Episodes\_Felony\_1
- Prior\_Conviction\_Episodes\_Felony\_2
- Prior\_Conviction\_Episodes\_Misd\_0
- Prior\_Conviction\_Episodes\_Misd\_1
- Prior\_Conviction\_Episodes\_Misd\_2
- Prior\_Conviction\_Episodes\_Misd\_3
- Prior\_Conviction\_Episodes\_Viol\_True
- Prior\_Conviction\_Episodes\_Prop\_0
- Prior\_Conviction\_Episodes\_Prop\_1
- Prior\_Conviction\_Episodes\_Prop\_2
- Prior\_Conviction\_Episodes\_Drug\_0
- Prior\_Conviction\_Episodes\_Drug\_1
- Prior\_Conviction\_Episodes\_PPViolationCharges\_True
- Prior\_Conviction\_Episodes\_DomesticViolenceCharges\_True
- Prior\_Conviction\_Episodes\_GunCharges\_True

Bad Predictors:

- Prior\_Conviction\_Episodes\_Felony\_0

### 3. Correlation between predictors.

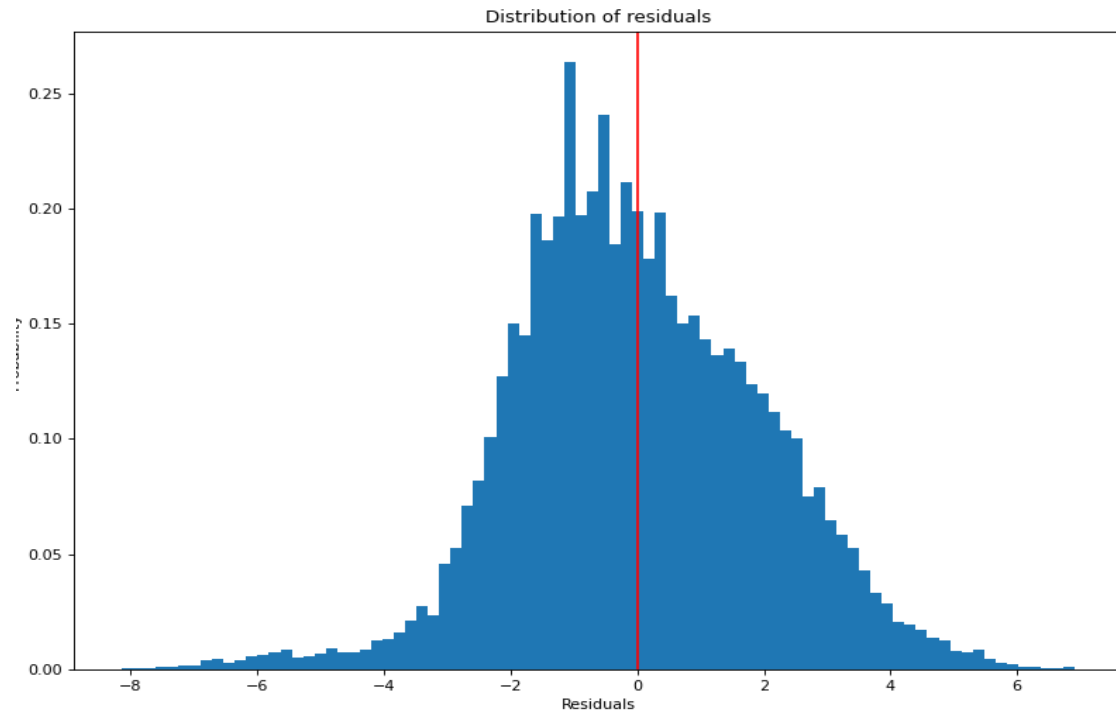
The following graph shows the Correlation matrix.



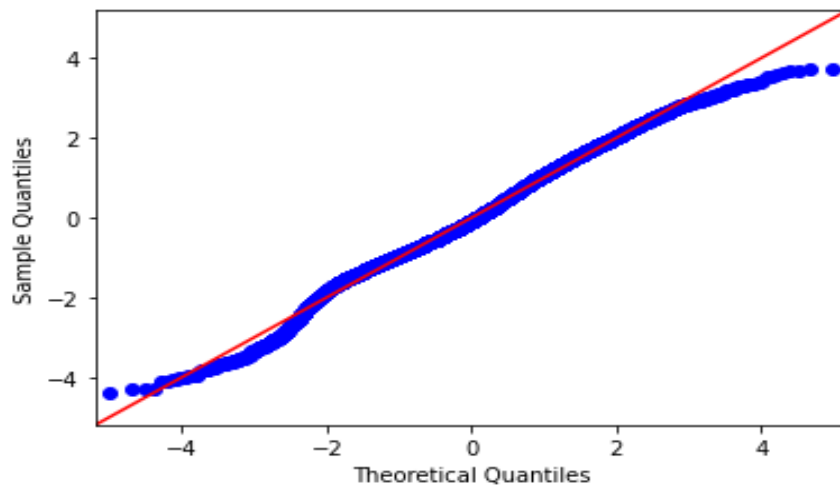
According to the correlation matrix, There are no highly correlated features. However, the features Prior\_Conviction\_Episodes\_Drug\_0 and Prior\_Conviction\_Episodes\_Drug\_1 are slightly correlated.

### 4. Residual Analysis

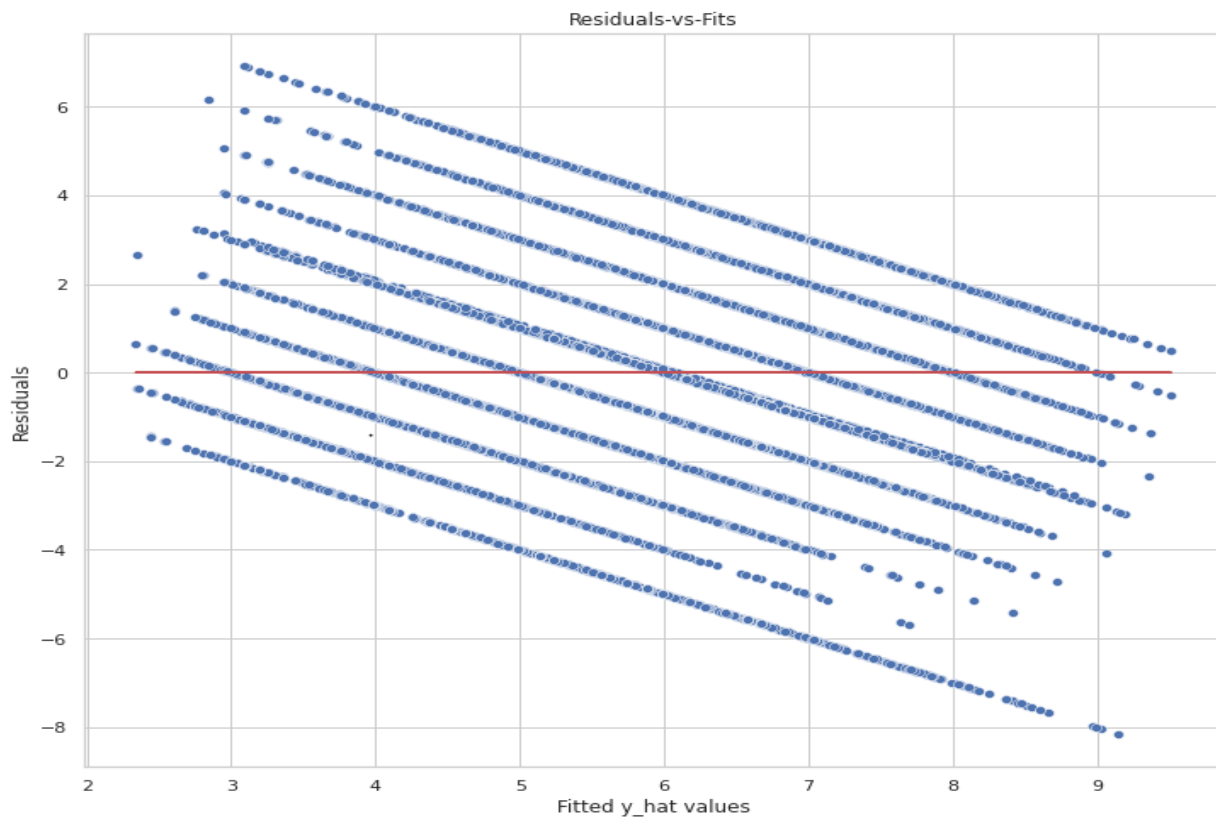
The following graph shows the Distribution of residuals.



The following figure shows the QQ plot where the Residuals terms are normally distributed with nearly 0 mean.

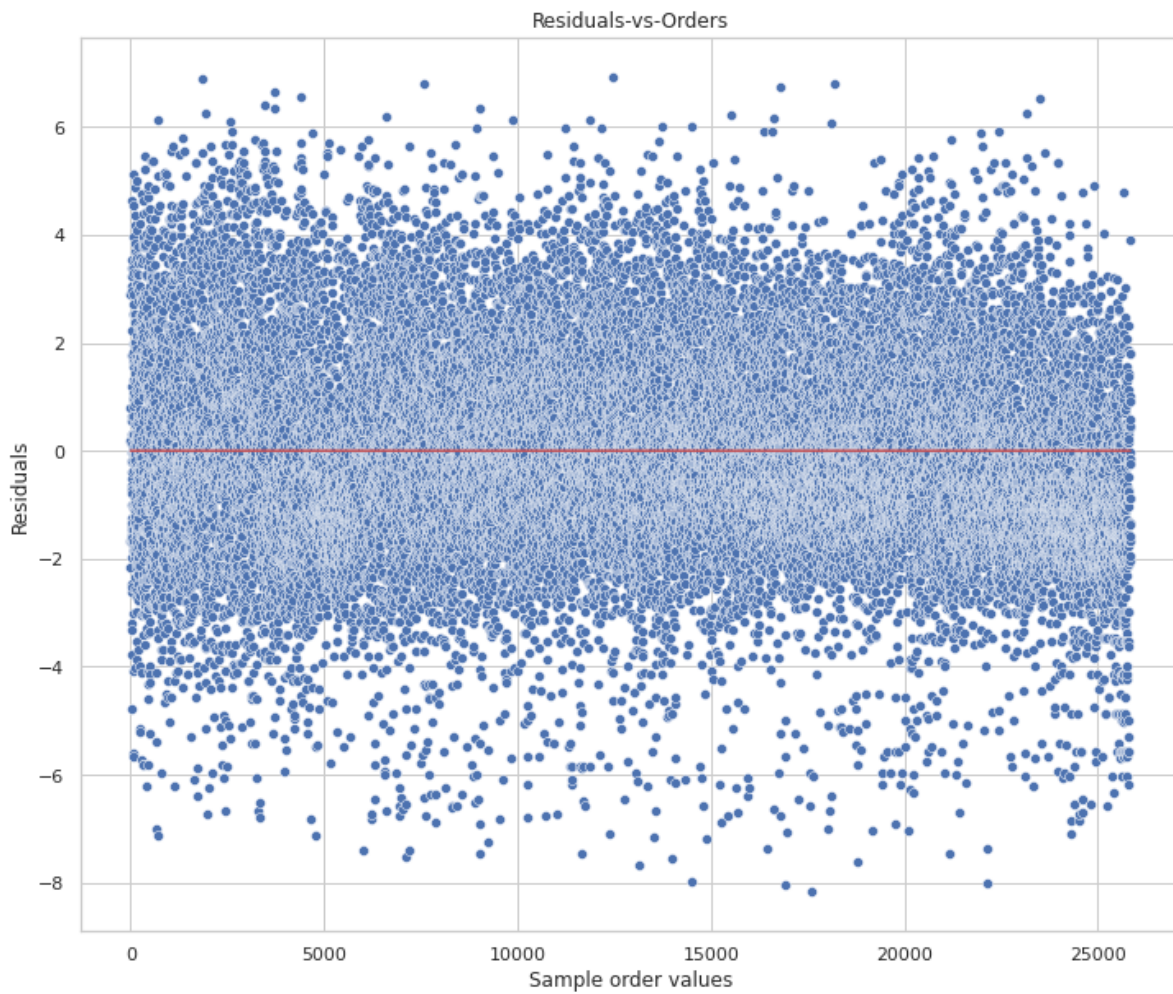


The following figure shows a plot of residuals vs fits.



Comment: It is shown that the residuals depart from 0 in some systematic manner since they are positive for small values and negative for high values. This systematic (non-random) pattern is sufficient to suggest that the regression function is not linear.

The following figure shows the residuals vs orders plot



Comment: There is a uniform random variation pattern about the zero residuals line, then the error terms are independent and not correlated.

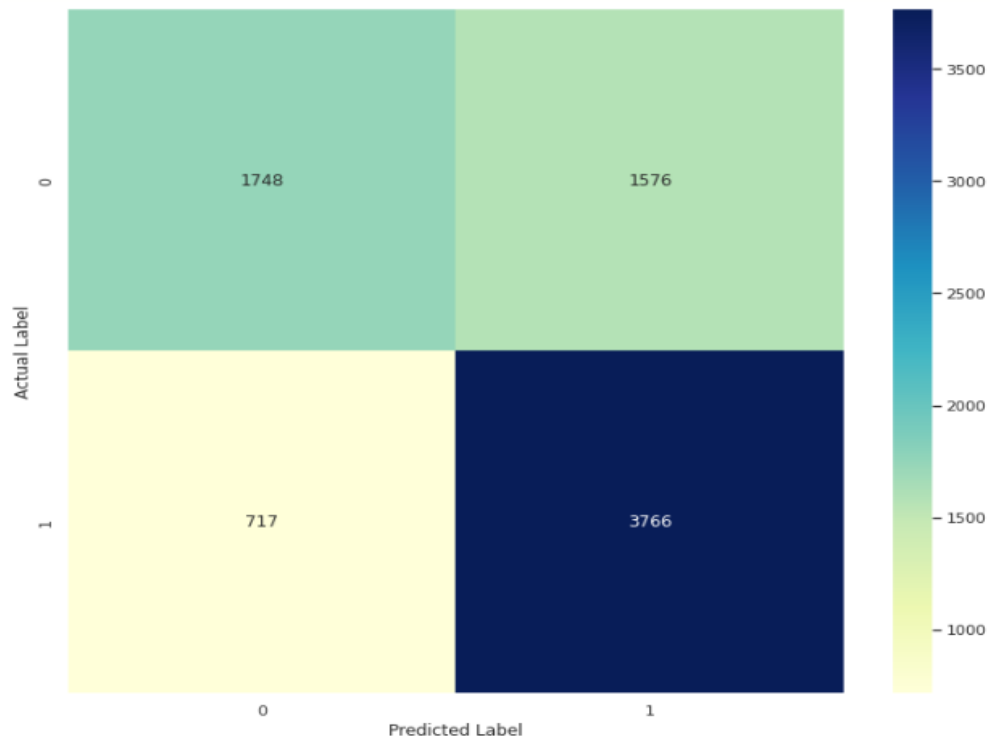
## Bonus Task

We have trained machine learning classifiers to predict the likelihood of recidivism within 3 years of release based on the state of Georgia recidivism records. The following table shows the evaluation metrics for the models.

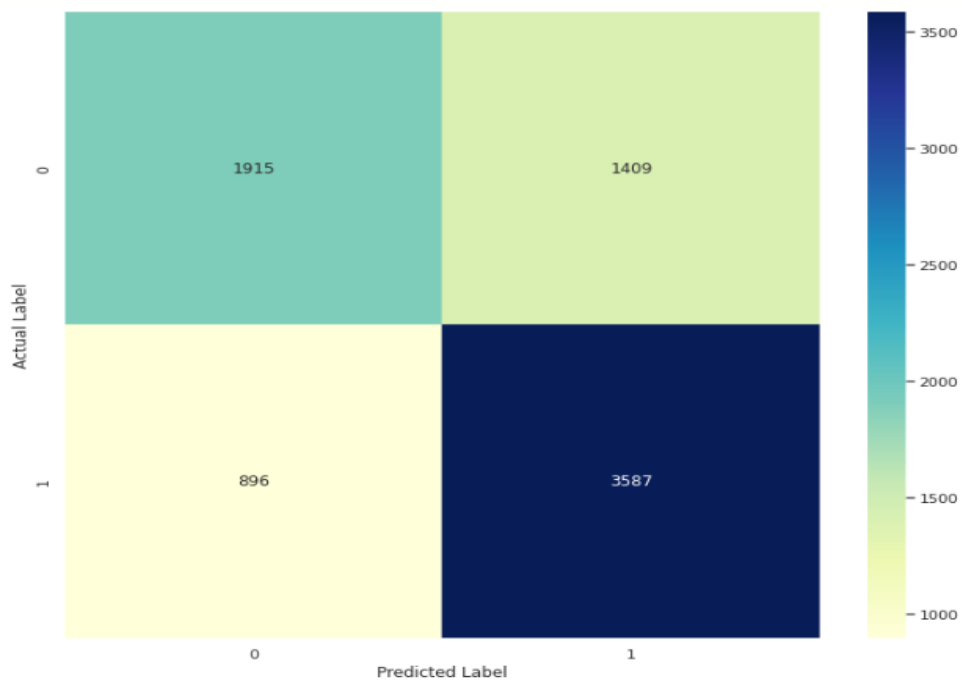
Model	Logistic Regression	Logistic Regression with Feature Selection	Random Forest	AdaBoost	Bagging
Accuracy	0.70629	0.70475	0.73344	0.73344	0.73857
Precision	0.70498	0.71797	0.73726	0.74331	0.74207
Recall	0.84006	0.80013	0.83248	0.81843	0.83493
F1-score	0.76662	0.75683	0.78198	0.77906	0.78577

## Visualizations

### Evaluation of Logistic Regression Model

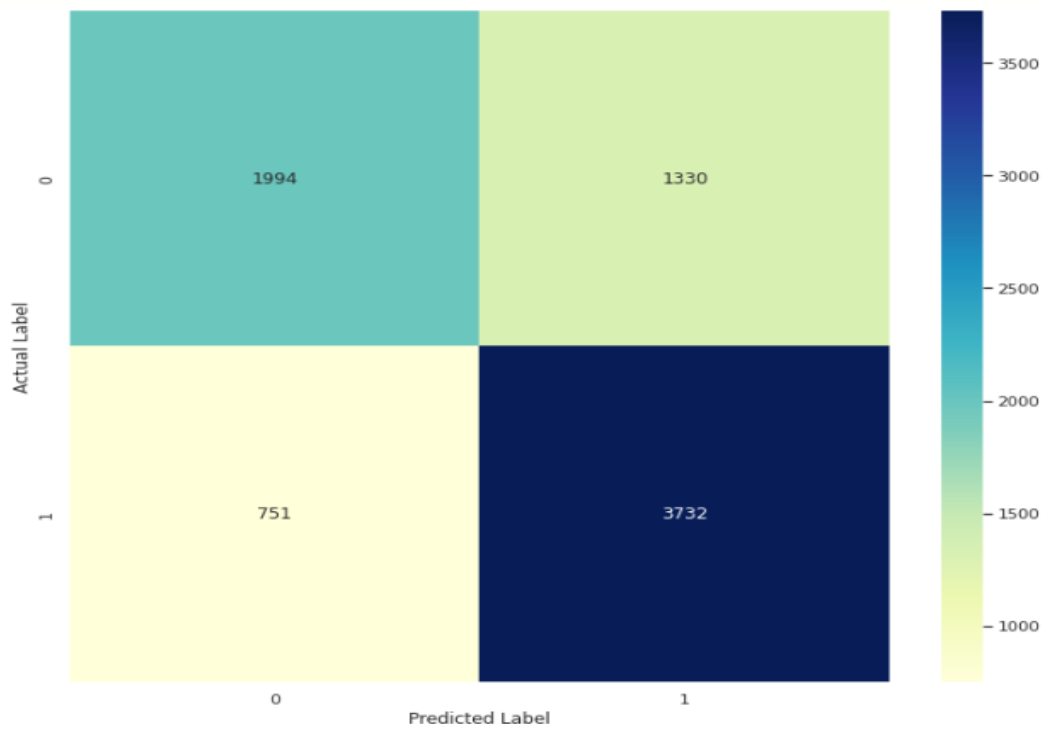


### Evaluation of Logistic Regression with feature selection Model

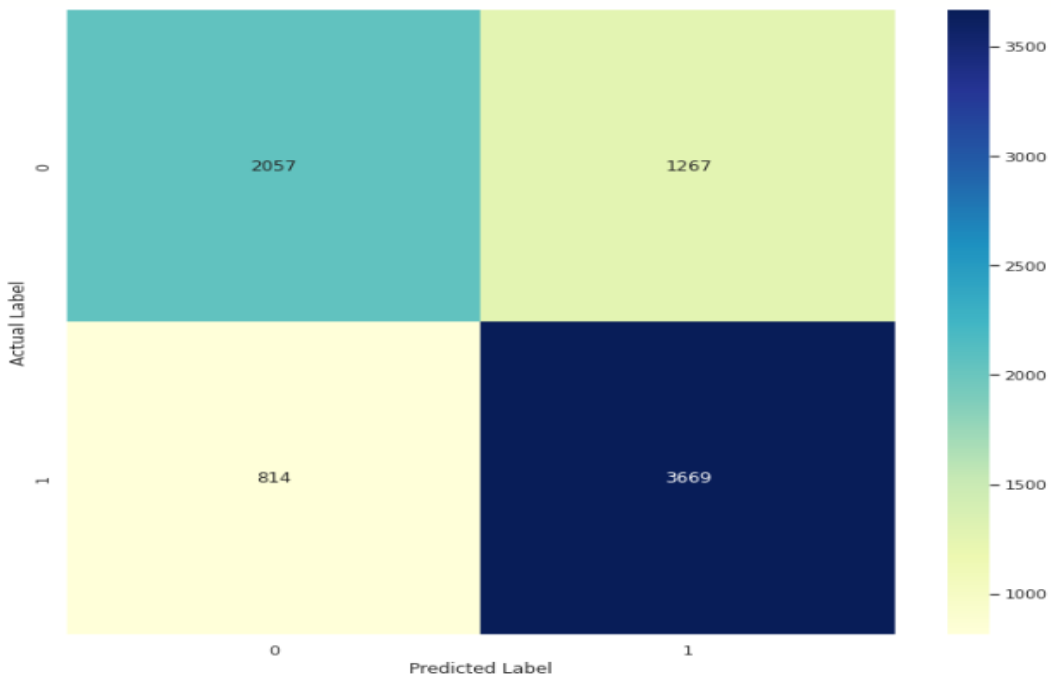




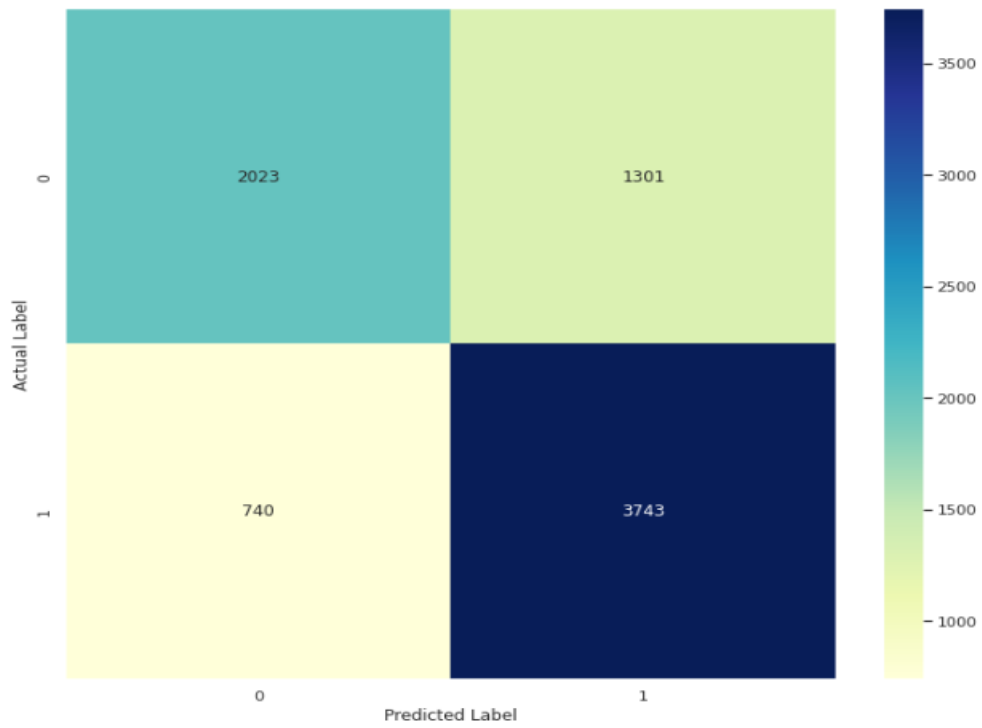
## Evaluation of Random Forest Model



## Evaluation of AdaBoost Model



## Evaluation of Bagging Model



## Conclusion

In conclusion, our data analysis project has provided significant insights into crime patterns and trends, as well as the effectiveness of different firearm laws. By examining The national crime victimization survey (NCVS) data, NIBRS Reported offense count data, Recidivism data for the state of Georgia from 2013 to 2015, and data on firearm laws per state, we have been able to gain a better understanding of these complex issues.

However, it is important to note that there are potential limitations to our findings. For example, the NCVS data is based on self-reported incidents, which may not always be accurate. Additionally, the recidivism data only covers a specific time period and location, and may not be representative of broader trends. Finally, the effectiveness of firearm laws can be difficult to quantify, and there may be other factors at play that contribute to changes in crime rates.

Despite these limitations, we believe that our findings provide valuable insights that can inform policy decisions and help to make our communities safer. Further research and analysis will be necessary to fully understand the complex relationships between crime, firearm laws, and other factors.