# Legal Industry AI-ML Case Study

The Murder Accountability Project is the most complete database of homicides in the United States currently available. This dataset includes murders from the FBI's Supplementary Homicide Report from 1976 to the present and Freedom of Information Act data on more than 22,000 homicides that were not reported to the Justice Department. This dataset includes the age, race, sex, ethnicity of victims and perpetrators, in addition to the relationship between the victim and perpetrator and weapon used.
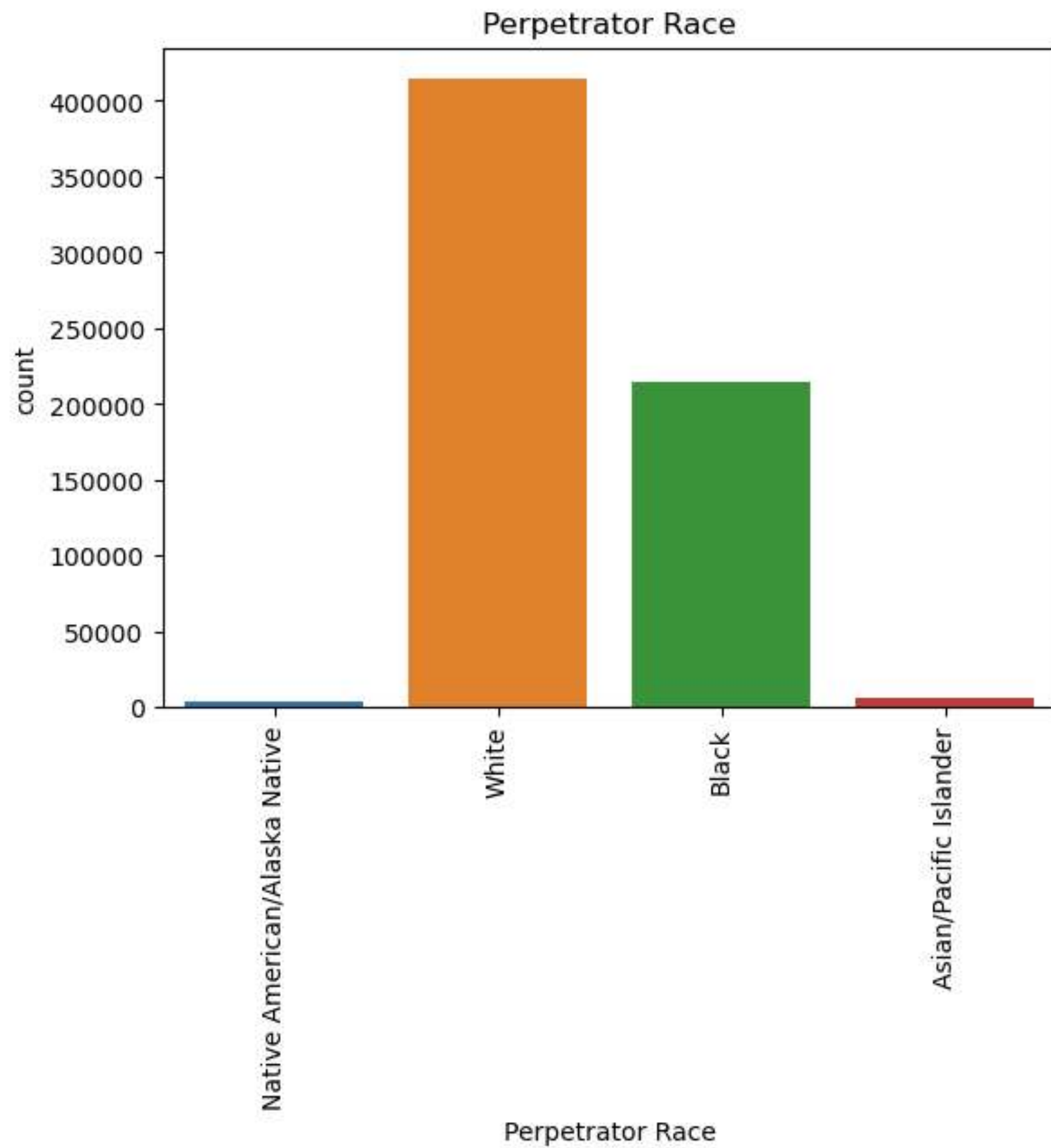
The main aim of the project is to predict whether the crime is solved or not and to identify the factors that determine the crime being solved or not.
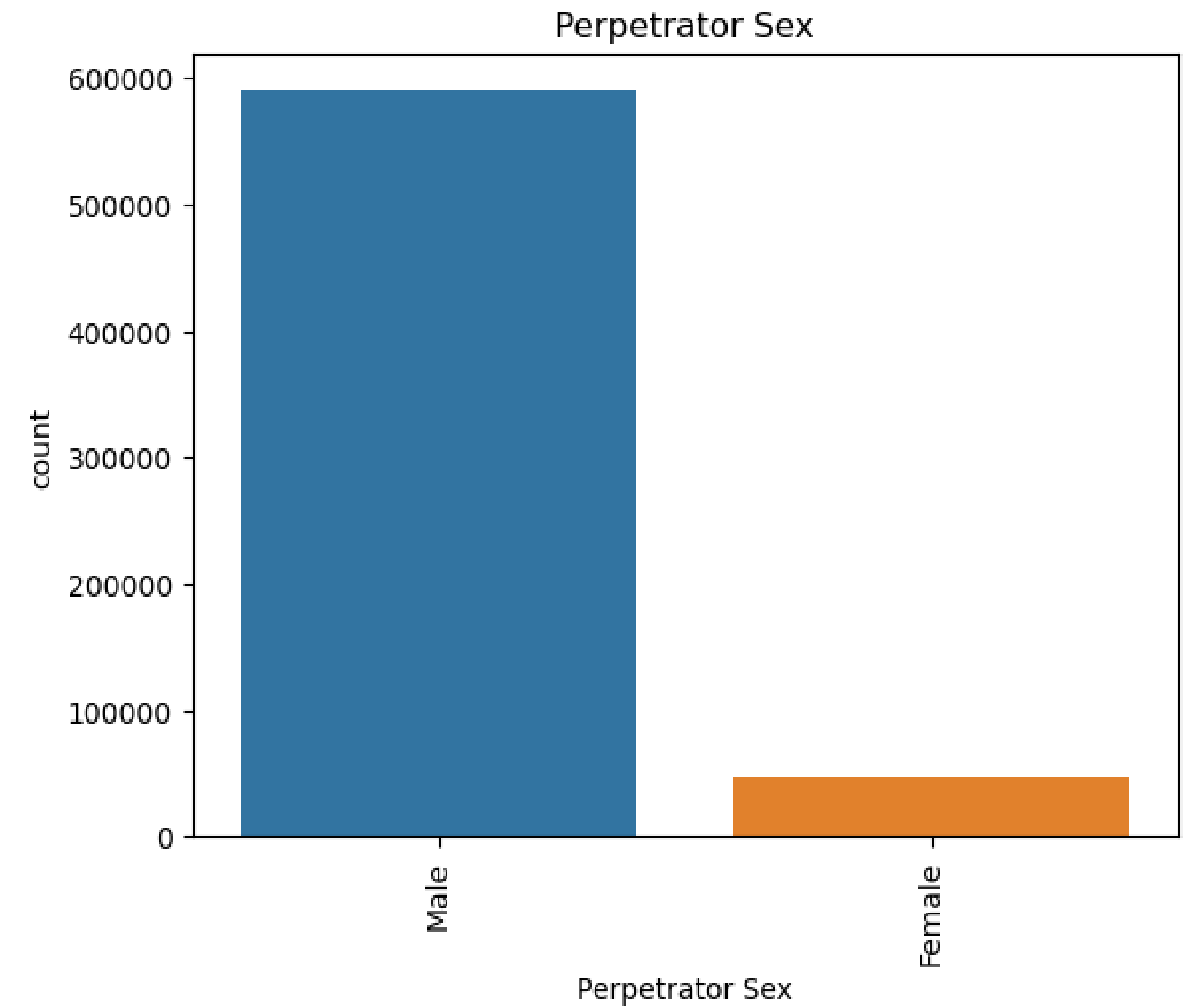
Class Distribution
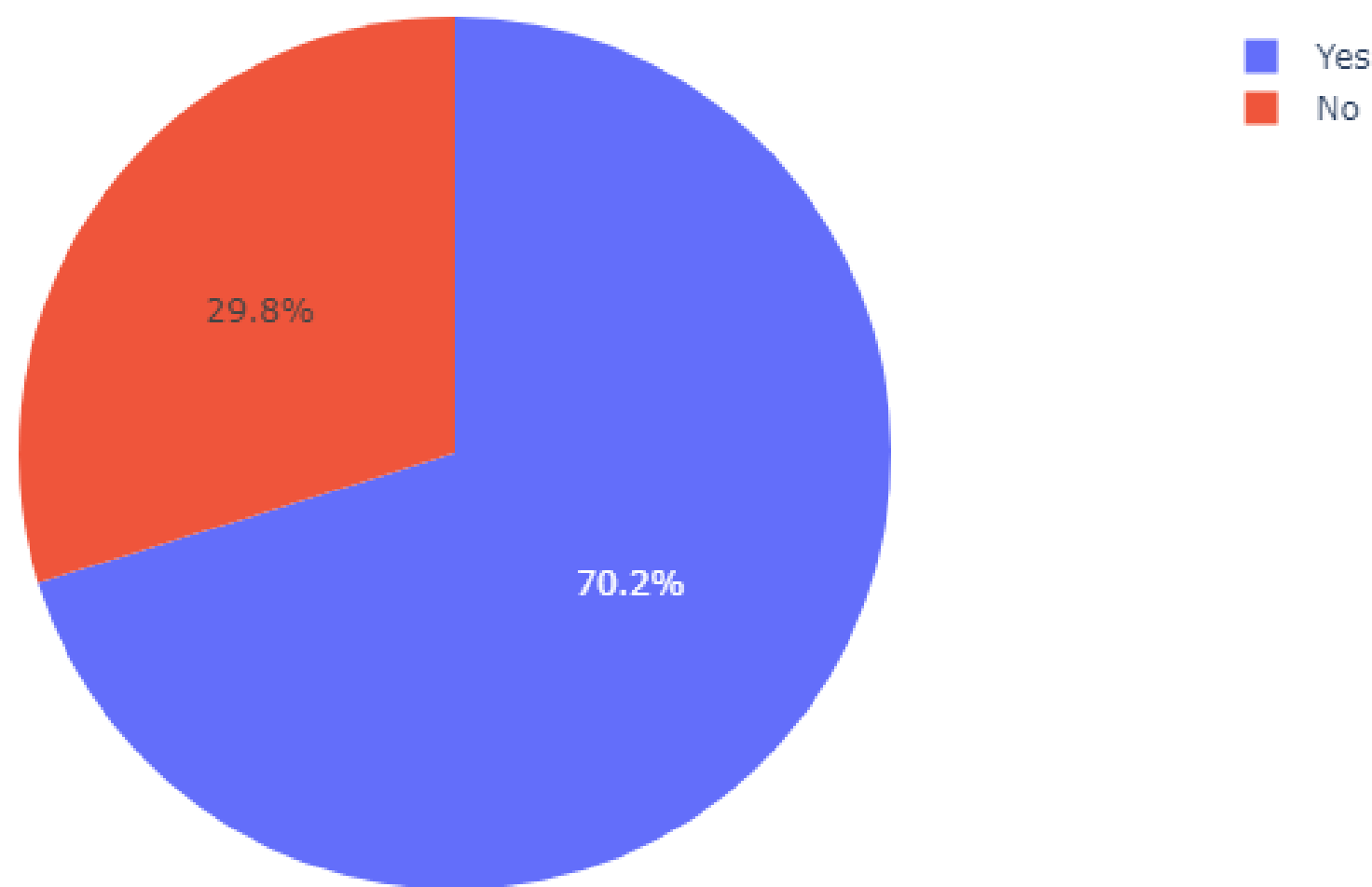
Crime Solved	No. of Homicides
Yes	448172
No	190282
Crime Solved Rate (Yes)	70.2 %
Crime Solved Rate (No)	29.8 %

Features Responsible : Auto-ML

Perpetrator Race

Perpetrator Sex
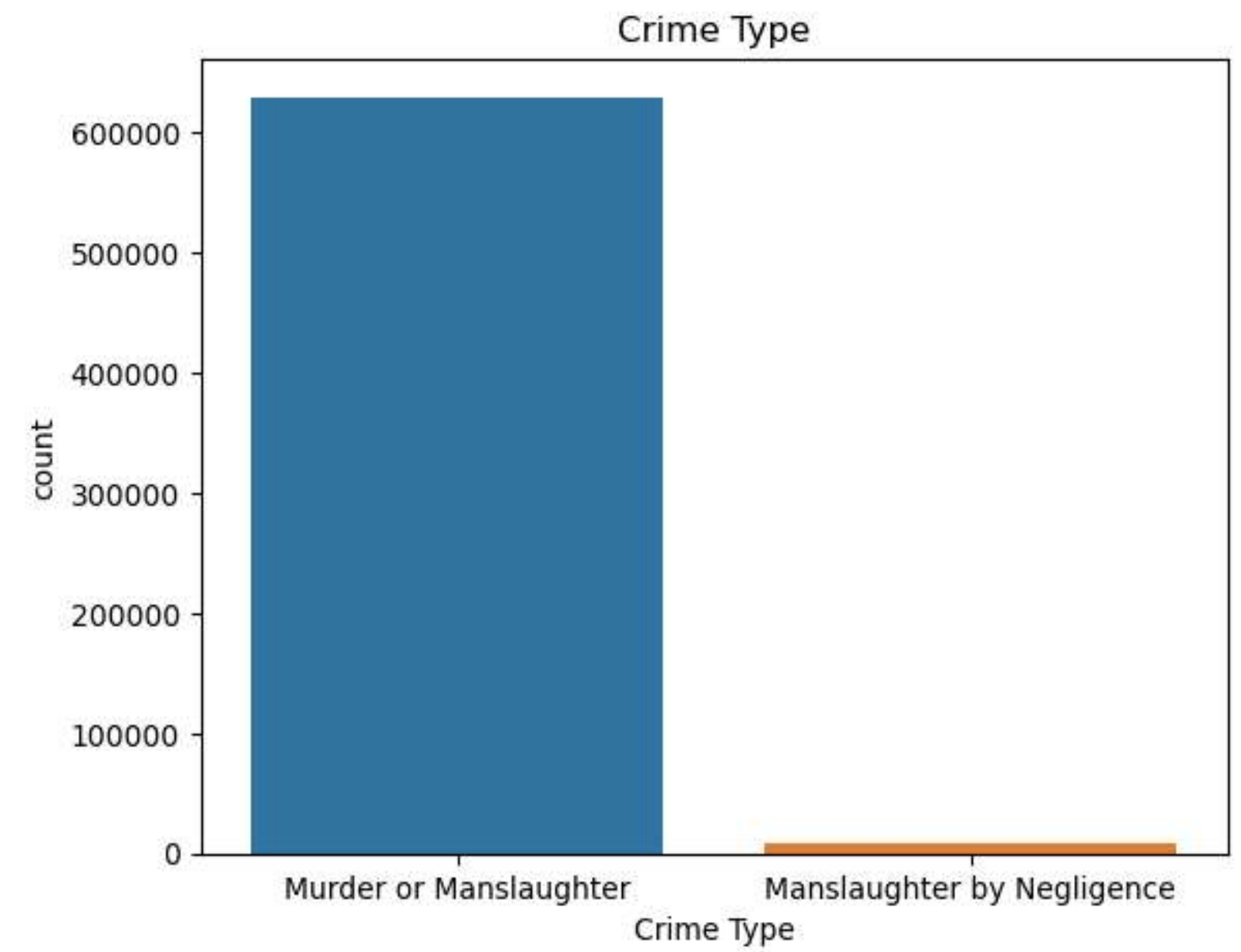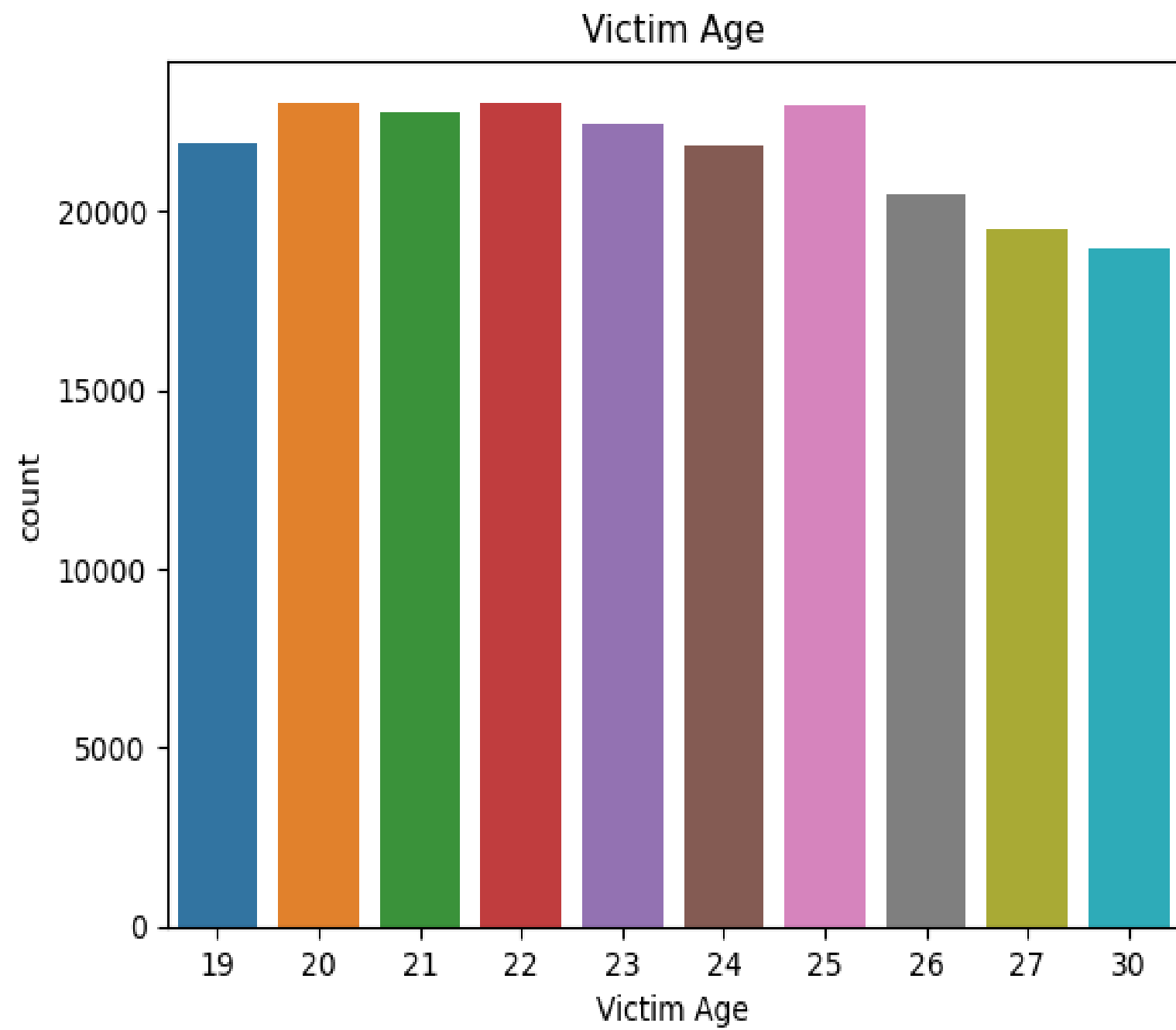
- **Perpetrator Race** = Race of the person commiting the crime
- The white race is shown more prone to commiting crime

- **Perpetrator Sex** = Gender of the perosn commiting the crime
- Males are more prone to commiting the crime

Victim Age

- **Victim Age** = The top 10 age of the victims who endured most crimes
- Most crime victims are aged from 19 to 30

Crime Type

- **Crime Type** = Type of crime cmmited on an individual
- Most type of crime commited is murder or Manslaughter

Auto-ML Methodology Results

Algorithms	Test Accuracy (25 percentile)	Test Accuracy (50 percentile)	Test Accuracy (75 percentile)	Test Accuracy (90 percentile)
Decsion Tree	99.96	99.96	99.93	99.92
Random Forest	99.97	99.97	99.99	99.91
XGBoost	99.98	99.99	99.99	99.99
MLP	99.97	99.98	99.94	99.94
RNN	99.97	99.97	99.8	50
Total Features	11	22	33	40
Avg. Accuracy	99.97	99.974	99.93	89.952

- **Based on our observation from the standard ML algorithms, 75th percentile has the best average accuracy.**

- **XGBoost was the best performing algorithm with 99.9% accuracy in 50, 75 and 90 percentile.**

Conclusion

In additional to solving crime, Auto-ML can be used by police departments to cut costs. With police budgets shrinking around the world, departments everywhere are looking for ways to save and a lot of them are resorting to limiting the number of officers and investigators they have on staff. Auto-ML can use law enforcement agency databases that have all kinds of information on past suspects. The dataset has 638,454 records with 18 Categorical Features and 6 Numerical Features.

For classification, models were created with algorithms using Auto-ML techniques like Decision Tree, Random Forest, XGBoost, Multilayer Perceptron and Recurrent Neural Network. With these models, performance measurement values were obtained for feature sets of 11, 22, 33 and 40. The Auto-ML algorithms were able to predict the crime solved or not with an average accuracy between 89% – 100% and helped to identify factors that determine the crime being solved or not. The major factors include Perpetrator Race, Perpetrator Sex, Victim Age and Crime Type.

Overall, there will be a point where machines will be able to solve a crime beyond a reasonable doubt. This will result in safer communities and eliminate the possibilities of innocent people being accused. Auto-ML can be used in the new generation tools to analyze crime investigations and can have a huge influence on future investigations.