Unraveling the Impact of Age on Crime Victimization

Understanding how age influences crime victimization is critical for developing strategies that prevent crime and support victims. This project explores the relationship between age and the types of crime individuals are likely to experience. By analyzing crime patterns, the study seeks to answer the question: how does age impact the likelihood of being a victim of different crimes? Identifying these patterns can help design targeted interventions to reduce crime and provide better victim support services.

The data for this analysis was sourced from the Charlotte-Mecklenburg Police Department (CMPD) dataset, which spans from 2012 to 2016, and U.S. Census data. The dataset includes various features such as the type of crime, victim age, and location. Initial pre-processing involved cleaning and merging the data to ensure consistency and remove duplicates. Key features such as incident type, victim age, and offense details were used to create a more streamlined dataset for analysis. Age groups were binned into categories (Child, Young Adult, Adult, Mature Adult, Senior) to make the analysis more interpretable. Several pre-processing steps were applied to prepare the data for modeling. First, categorical data was label-encoded to convert into a numerical format, which is necessary for machine learning algorithms. Data cleaning also removed unnecessary duplicates and merged critical features across datasets. This process ensured a reliable foundation for predictive modeling.

Visualizations were employed using boxplots and bar plots to understand the data further. These visuals revealed trends such as younger individuals experiencing more violent crimes while older individuals were more likely to be victims of property crimes. This aligns with prior research but also uncovered new insights, such as the heavy reliance of the model on the "Missing Person" feature. The XGBoost model was chosen for its superior performance in handling multi-class classification. Through hyperparameter tuning, the model achieved a test accuracy of 43.39%. Although other models like Random Forest and Gradient Boosting were tested, XGBoost proved the most effective for this dataset. The project's findings indicate that the "Missing Person" status was highly influential in predicting crime victimization. This suggests that further research could explore why this feature is crucial.

In conclusion, this project demonstrates the significance of age in predicting crime victimization and provides a foundation for more targeted approaches to crime prevention and victim support. The use of machine learning models like XGBoost has shown potential, offering a hopeful outlook for future crime prevention strategies. However, future work should continue to focus on mitigating bias and improving the accuracy of predictions.