Understanding the socioeconomic and housing factors that contribute to violent crime rates is a key objective in crime analysis and public policy. This project explores a subset of the Communities and Crime Dataset, which contains demographic, economic, and crime-related data across various U.S. communities. The target variable for this analysis is ViolentCrimesPerPop, representing the violent crime rate per population.

To enhance the predictive modeling of violent crime rates, we identified and selected 10 key features that are likely to influence crime rates based on socioeconomic and housing conditions. The selected features include:

* PersPerFam - The average number of persons per family.
* PctHousLess3BR - The percentage of houses with less than 3 bedrooms.
* householdsize - The average size of a household.
* NumIlleg - The number of children born out of wedlock.
* state - A numerical identifier for the state.
* PersPerRentOccHous - The number of persons per rent occupied house.
* PctHousOwnOcc - The percentage of houses that are owner-occupied.
* PctWorkMomYoungKids - The percentage of working mothers with young children.
* MedRentPctHousInc - The median rent as a percentage of household income.
* NumUnderPov - The number of individuals living under the poverty line.

To further enhance the predictive power of our dataset, we introduced several engineered features that capture interactions and relationships between key variables. These new features help to account for complex socioeconomic dynamics that may not be fully captured by individual variables. The engineered features include:

* PctHousOwnOcc\_PctHousLess3BR - The ratio of owner-occupied houses to houses with fewer than 3 bedrooms.
* PersPerFam\_householdsize - The ratio of persons per family to the average household size.
* NumUnderPov\_householdsize - The ratio of the number of individuals under poverty to household size.
* NumIlleg\_PctWorkMomYoungKids - The ratio of the number of children born out of wedlock to the percentage of working mothers with young children.
* MedRentPctHousInc\_PctHousOwnOcc - An interaction term between median rent percentage of household income and the percentage of owner-occupied houses.
* PersPerRentOccHous\_PctWorkMomYoungKids - An interaction term between the number of persons per rented occupied house and the percentage of working mothers with young children.

By focusing on these features and their interactions, this project aims to uncover the relationships between housing, economic stressors, and violent crime rates. The dataset provides a foundation for understanding how community-level economic and housing factors contribute to crime and can inform targeted policy interventions.