The code follows a structured approach to analyzing traffic collision data, starting with loading the dataset and identifying categorical variable factors. It then performs descriptive statistics and visualizations, including histograms for numerical variables and count plots for categorical variables, to understand data distribution. Afterward, categorical variables are converted into numerical form, and the dataset is cleaned to handle missing values. The analysis proceeds with exploratory data analysis, where a correlation matrix is generated to identify relationships between variables. Next, three classification models, Logistic Regression, Decision Tree, and Random Forest, are developed to predict fatalities. The top 10 most important features are then selected, and optimized classification models are trained using these selected features, ensuring better model performance and interpretability.