- ***Dataset Overview:***

The dataset provided for fraud detection exhibits a significant class imbalance, with a small number of fraudulent transactions compared to non-fraudulent ones.

Missing values in the dataset were handled, and the numerical features were scaled using StandardScaler. Categorical variables were encoded using label encoding.

- ***Data Visualization:***

The distribution of fraudulent transactions was visualized, highlighting the minimal occurrence of fraud compared to non-fraud cases.

A bar graph was generated to compare the different transaction types for fraud and non-fraud cases, revealing that cash-out transactions were the most common type associated with fraud.

- ***Logistic Regression Model:***

The logistic regression model was trained and evaluated for fraud detection.

The model achieved an accuracy of 90.41%, precision of 91.36%, recall of 89.19%, and F1 score of 90.26%.

Cross-validation demonstrated a mean accuracy of approximately 89.89% with a standard deviation of 0.0258, indicating consistent performance.

- ***Decision Tree Model:***

The decision tree model was trained and evaluated for fraud detection.

The model achieved an accuracy of 99.17%, precision of 98.66%, recall of 99.69%, and F1 score of 99.18%.

Cross-validation showed good accuracy for the decision tree model.

- ***Correlation Analysis:***

A correlation analysis revealed a strong positive correlation between the "Amount" feature and the target variable, indicating its importance in predicting fraud.

- ***Conclusion:***

The code provides a comprehensive analysis of fraud detection using logistic regression and decision tree models.

The decision tree model achieved higher performance metrics, but both models showed promising results in identifying fraudulent transactions.

The insights gained from this code can serve as a foundation for further improvements and fine-tuning to enhance the overall effectiveness of fraud detection systems.

Overall, the code presents an in-depth analysis of fraud detection, highlighting the performance of logistic regression and decision tree models. The correlation analysis and fraud prevention measures further contribute to understanding and mitigating fraud risks.