# Credit Card Fraud Detection: Project Summary

## Data Cleaning and Exploratory Data Analysis

The dataset underwent thorough cleaning and EDA:

* • Duplicate transactions were removed, and missing values were imputed via forward-fill.
* • Outlier analysis was conducted using the IQR method, highlighting anomalies in key features (e.g., V17, V14).
* • Feature engineering added new variables: Time-based (Hour\_of\_Day, Minute\_of\_Hour), amount deciles, and interaction features (e.g., V1\_plus\_V2).
* • Advanced visualization techniques (correlation heatmaps, KDE plots, PCA and t-SNE projections) revealed distinct patterns between legitimate and fraudulent transactions.

## Modeling and Evaluation

A clear modeling strategy was established:

* • Baseline Model: Logistic Regression was chosen for its simplicity and interpretability in classification.
* • Evaluation Metric Identification: ROC-AUC and F1-score were selected as primary metrics.
* • Metric Rationale: ROC-AUC measures the model's ranking ability across thresholds and is robust to class imbalance; F1-score balances precision and recall, critical when false positives and negatives are costly.
* • Interpretation: The baseline logistic model achieved a ROC-AUC of approximately 0.95, indicating strong discrimination between fraud and legitimate transactions.