Interim Report: Fraud Detection Project

BY: Petros Abebe

Date: 20-07-2025

Introduction

Financial fraud poses a significant threat to both e-commerce and banking platforms, leading to substantial financial losses and eroding customer trust. This project aims to develop robust machine learning models to detect fraudulent transactions using both e-commerce and credit card datasets. The solution leverages advanced data analysis, feature engineering, and model explainability techniques to improve fraud detection accuracy and support business decision-making.

Objective

- Analyze and understand the characteristics of fraudulent transactions in e-commerce and credit card datasets.
- Preprocess and engineer features that enhance the predictive power of fraud detection models.
- Build and evaluate machine learning models capable of accurately identifying fraudulent transactions, even in the presence of severe class imbalance.
- Provide interpretable results and actionable business insights using model explainability tools.

Methodology

Data Sources

- **Fraud_Data.csv**: E-commerce transaction data.
- **IpAddress_to_Country.csv**: IP address to country mapping.
- **creditcard.csv**: Bank credit card transaction data.

Exploratory Data Analysis (EDA)

- Assessed data structure, missing values, and class distribution.
- Explored numerical and categorical feature distributions.
- Analyzed time-based patterns and IP-to-country mappings.
- Investigated correlations and feature importance.

Data Preprocessing and Feature Engineering

- Cleaned data by removing duplicates and correcting data types.
- Imputed missing values.
- Merged e-commerce data with IP-to-country mapping for geolocation features.
- Engineered time-based and transaction-based features (e.g., time since signup, transaction frequency).
- Encoded categorical variables and scaled numerical features.
- Addressed class imbalance using SMOTE.

- Split data into training and testing sets.

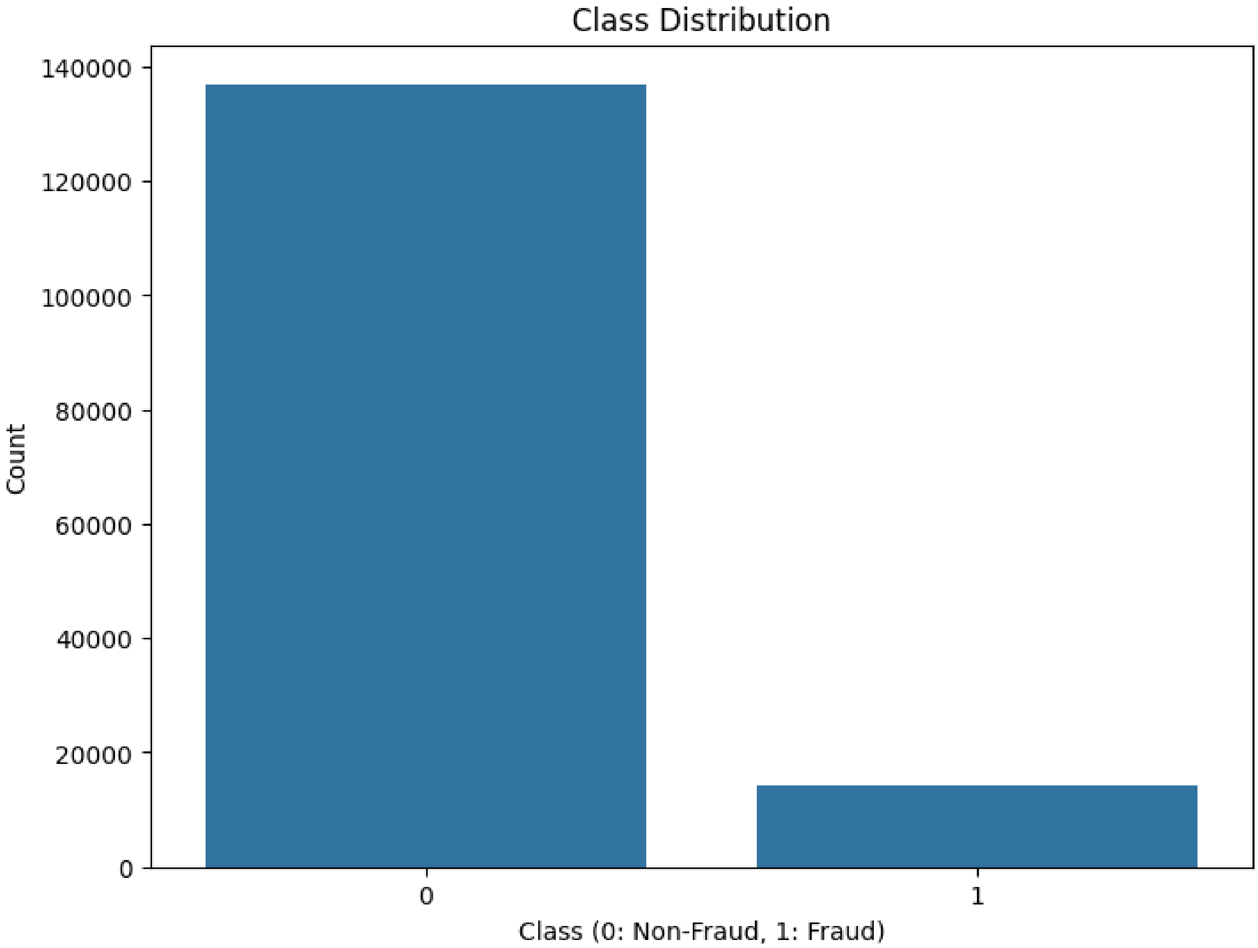
Model Building

- Trained and evaluated Logistic Regression and XGBoost models on both datasets.
- Used metrics such as Accuracy, Precision, Recall, F1 Score, ROC AUC, and PR AUC for evaluation.
- Applied SHAP for model explainability.

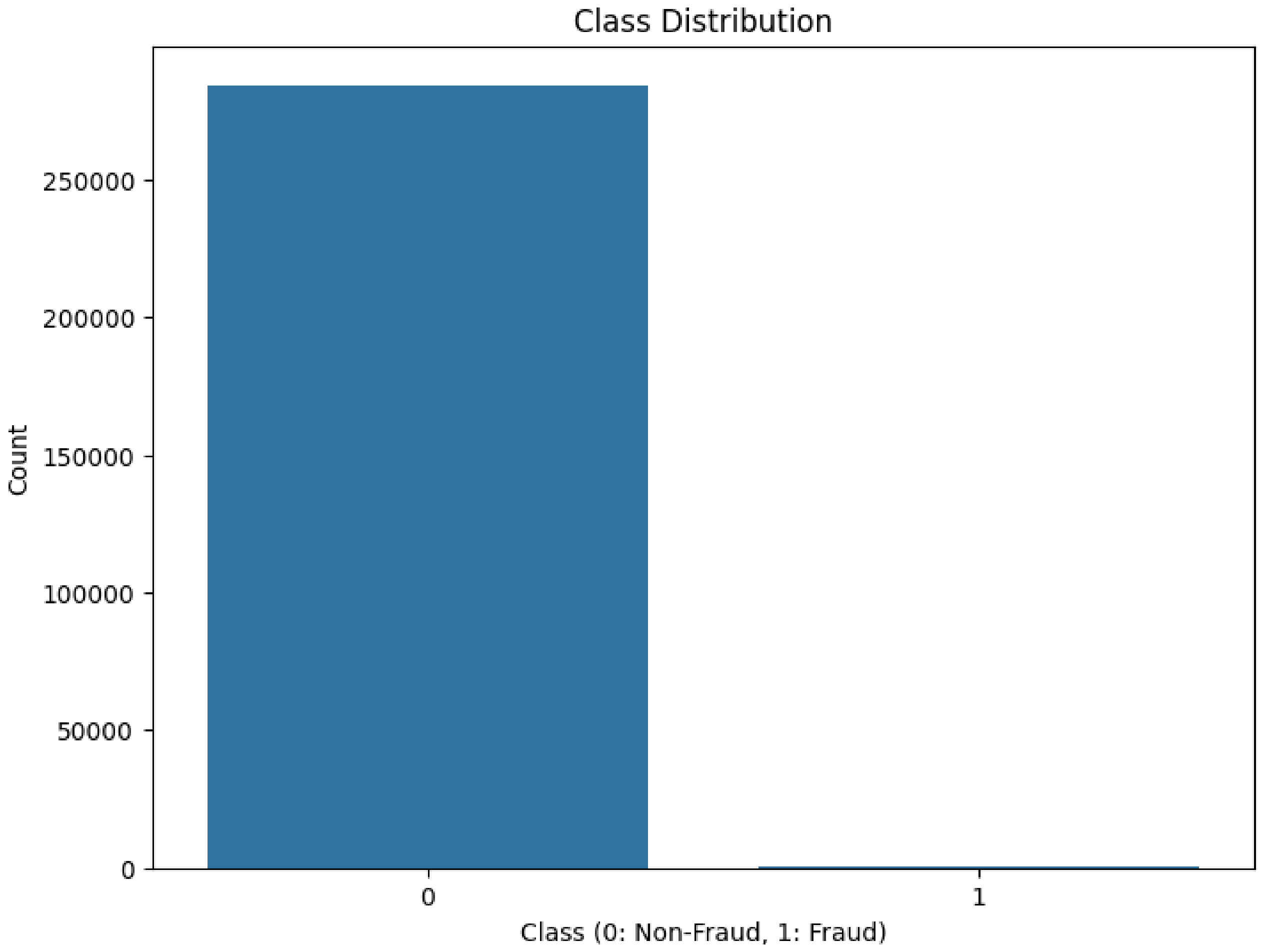
Results and Findings

EDA Insights

- **Class Distribution**

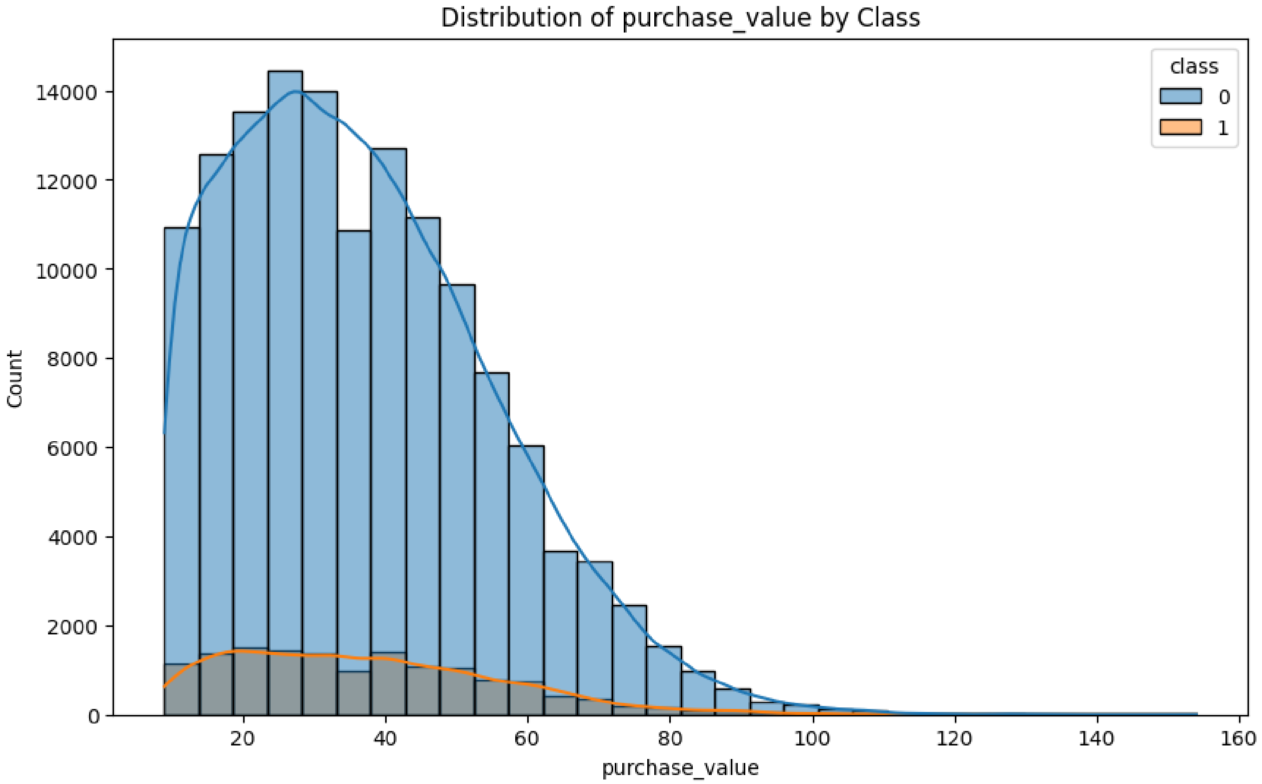


Interpretation: The e-commerce dataset is highly imbalanced, with a small proportion of fraudulent transactions compared to legitimate ones.

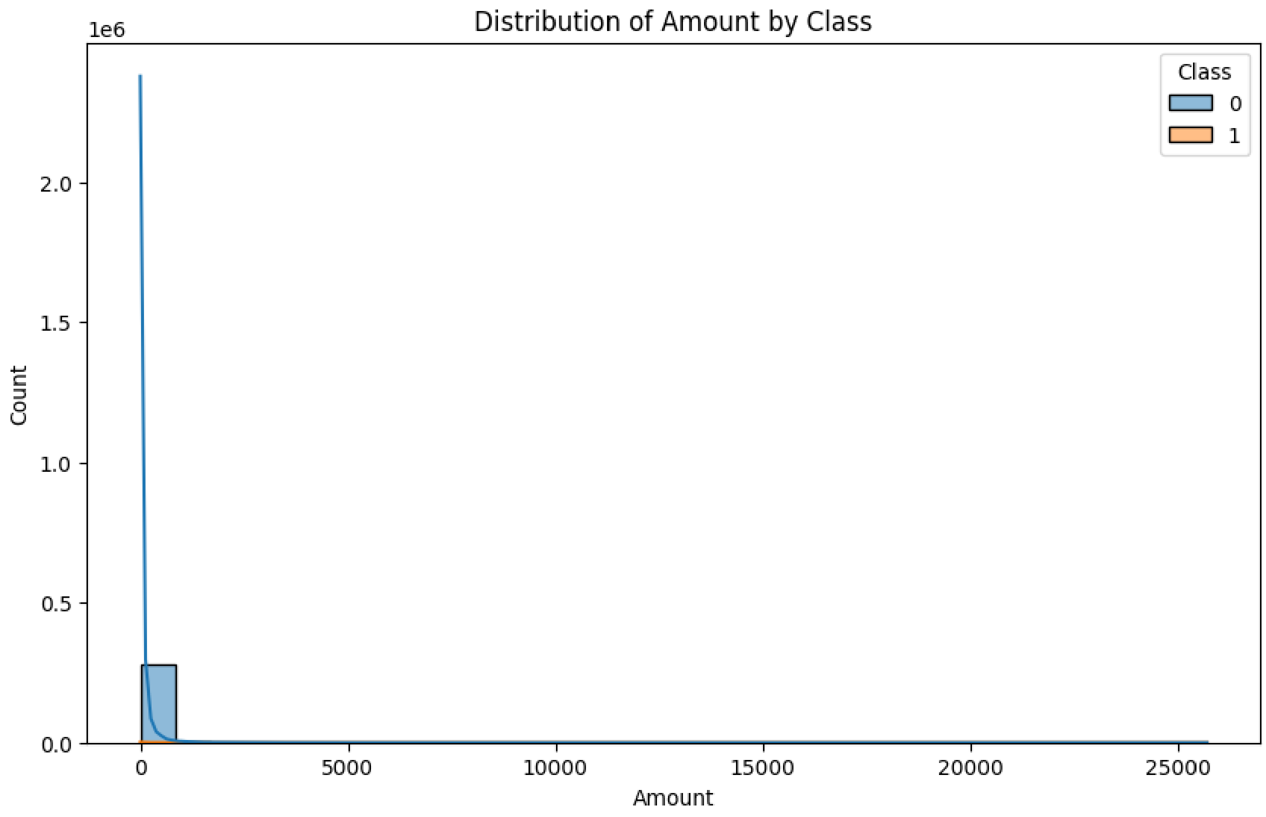


Interpretation: The credit card dataset is even more imbalanced, with fraud cases making up less than 1% of all transactions.

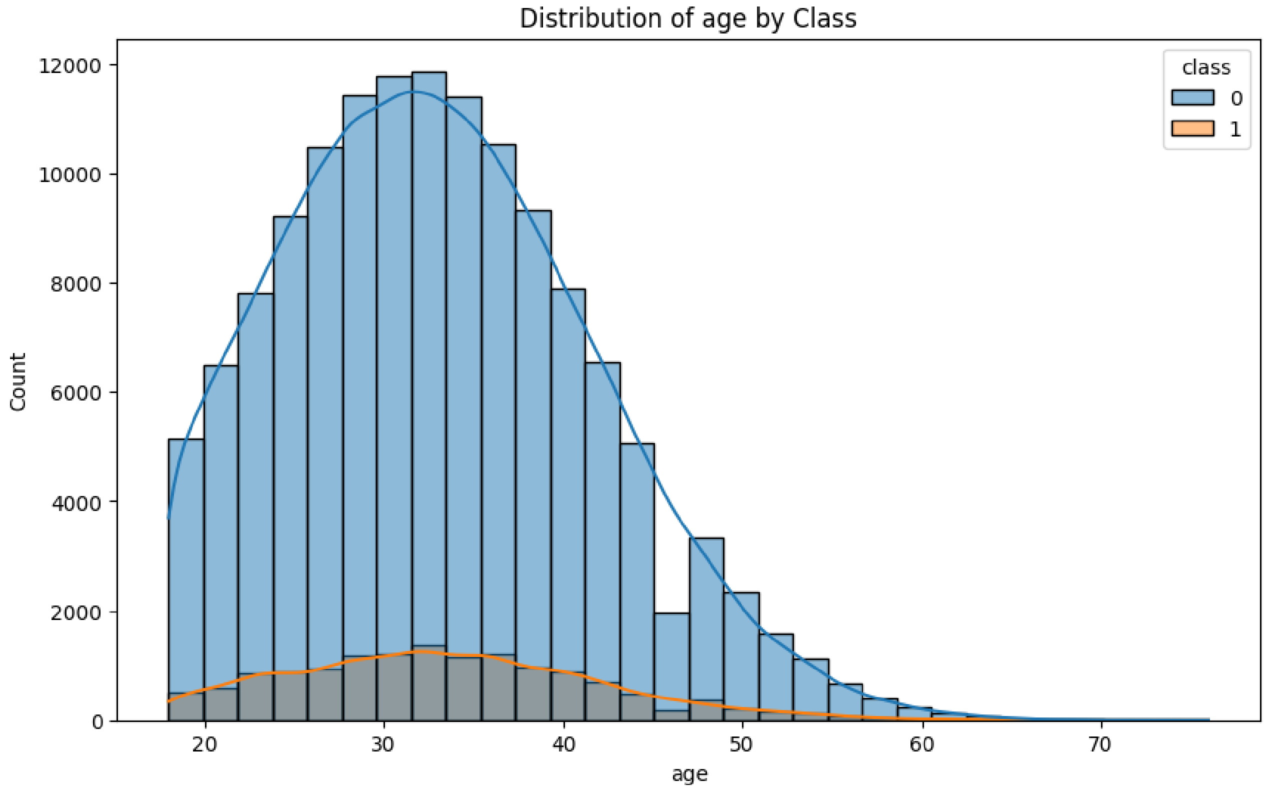
- **Numerical Feature Distributions**



Interpretation: Fraudulent transactions tend to have different purchase value distributions compared to non-fraudulent ones.

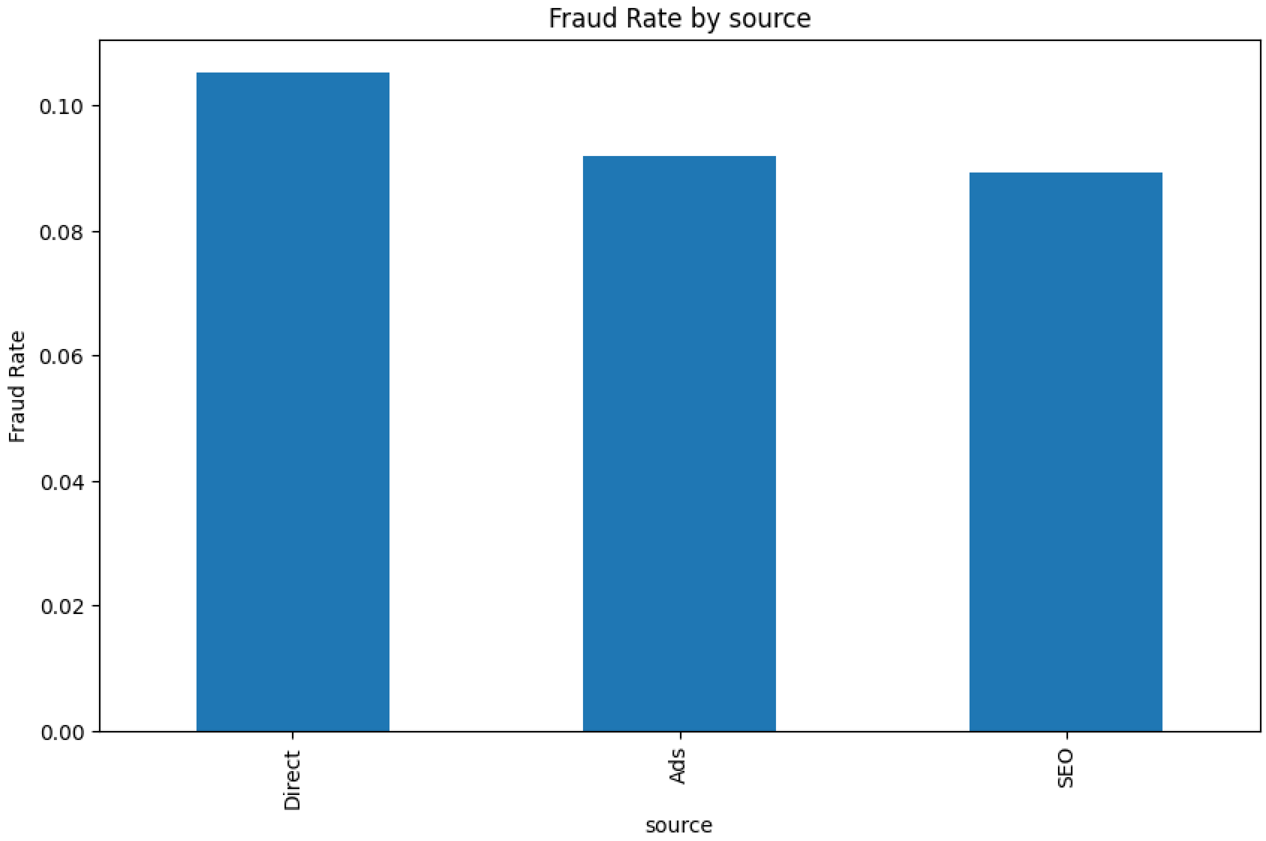


Interpretation: In the credit card data, fraudulent transactions often involve lower amounts, but there is overlap with legitimate transactions.

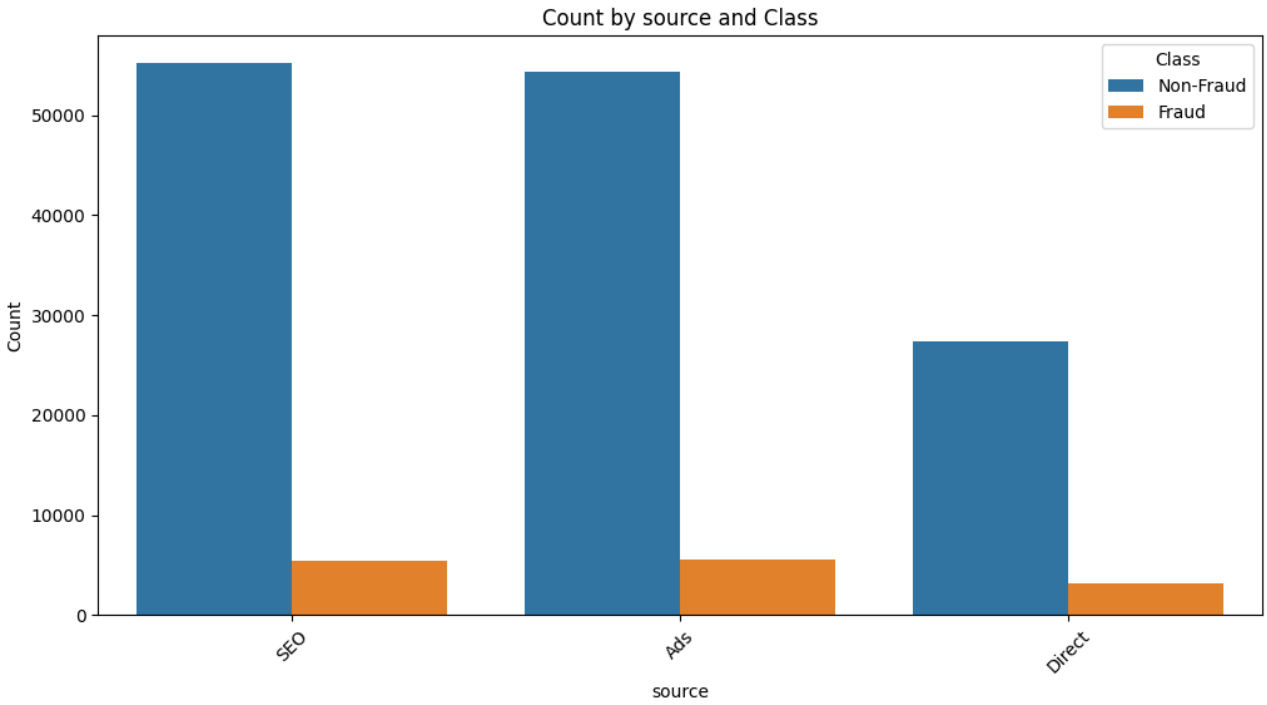


Interpretation: Age distribution shows that certain age groups may be more susceptible to fraud, though both classes span a wide range.

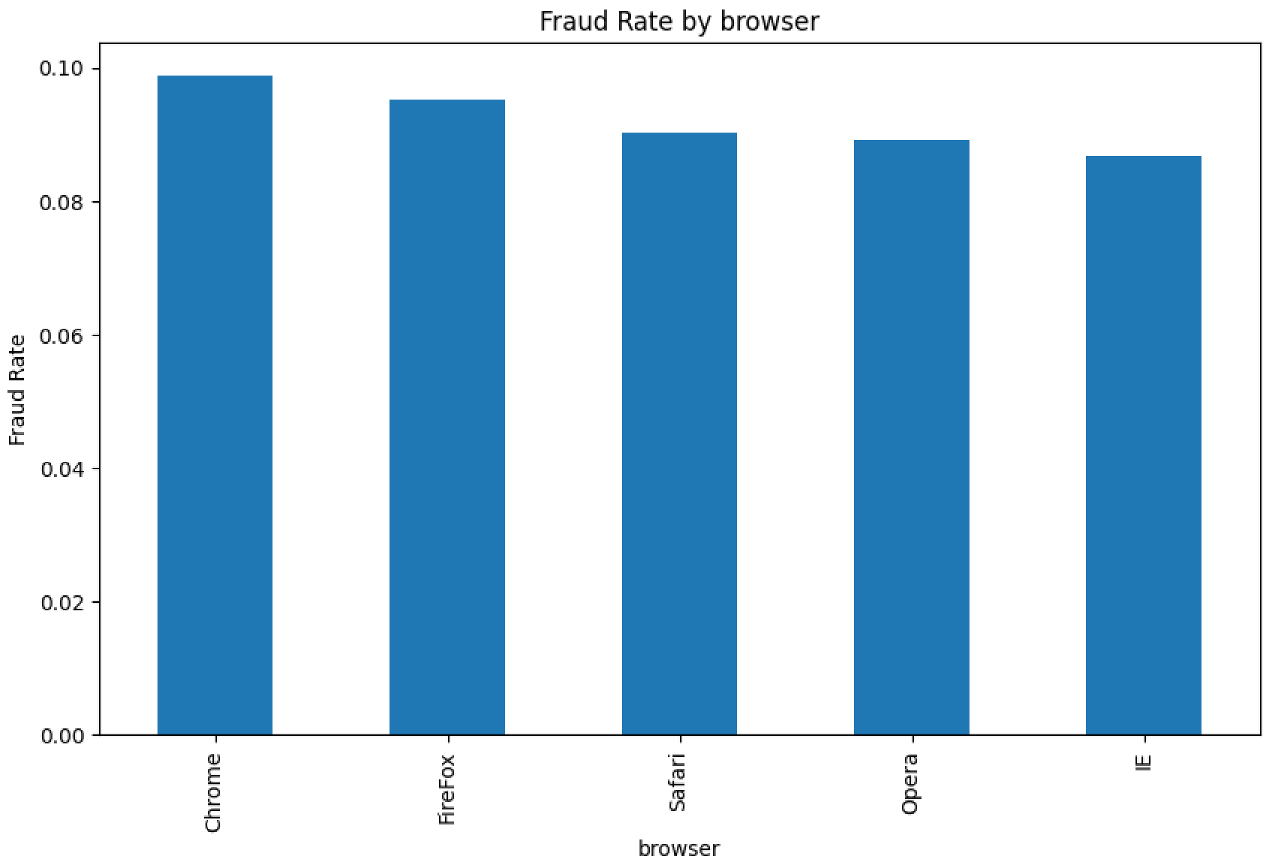
- **Categorical Feature Distributions**



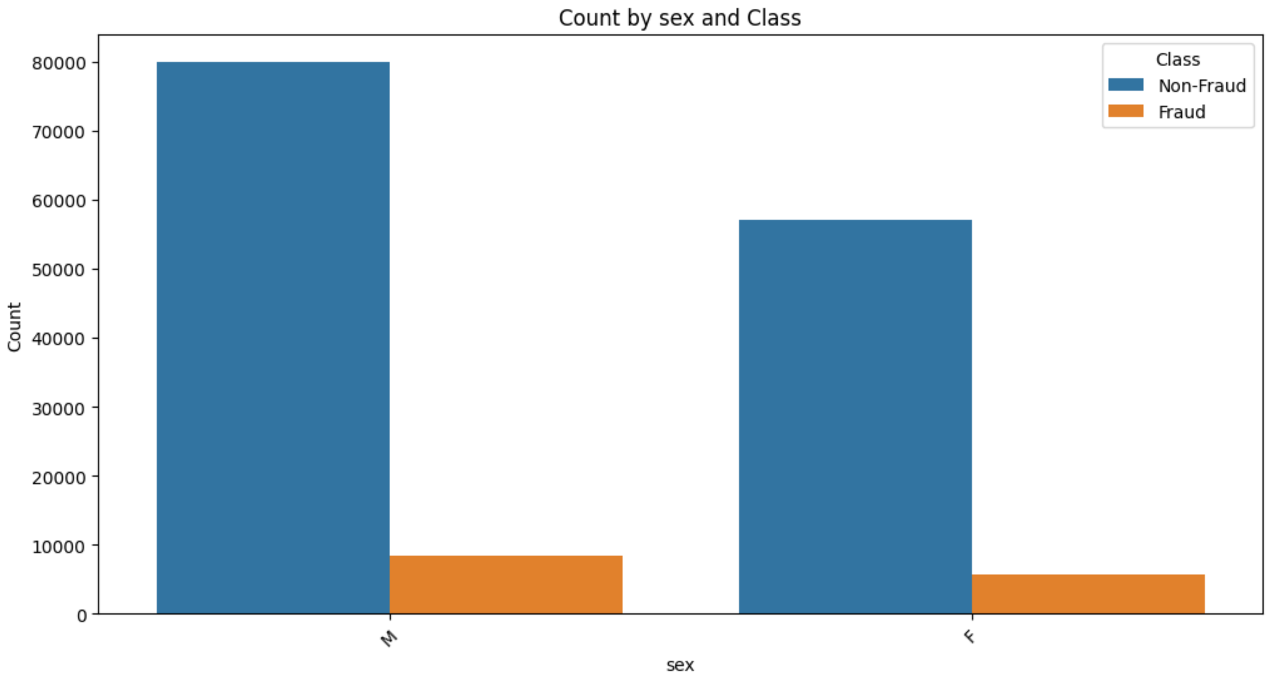
Interpretation: Some sources (e.g., marketing channels) are associated with higher fraud rates than others.



Interpretation: The number of transactions from each source varies, and some sources contribute disproportionately to fraud cases.

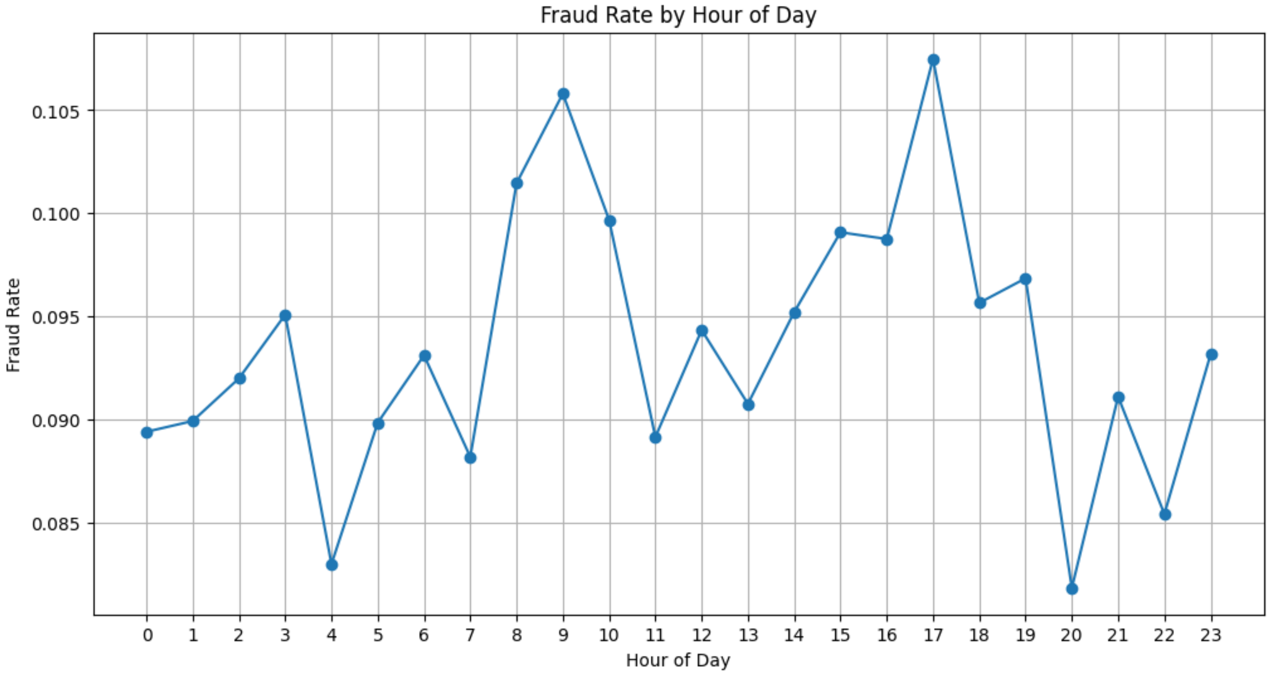


Interpretation: Certain browsers are more frequently used in fraudulent transactions, suggesting possible device or user behavior patterns.

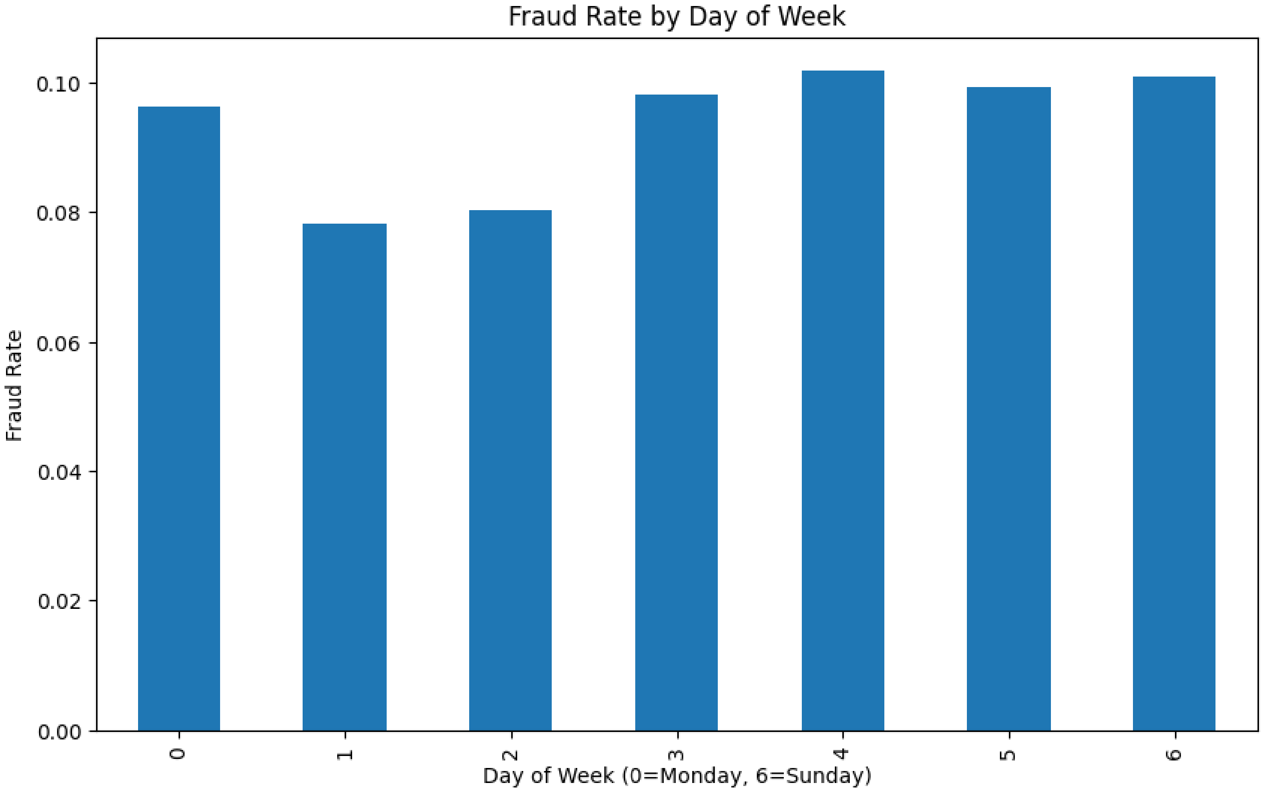


Interpretation: The distribution of fraud by sex is relatively balanced, indicating no strong gender bias in fraudulent activity.

- **Time-based Patterns**

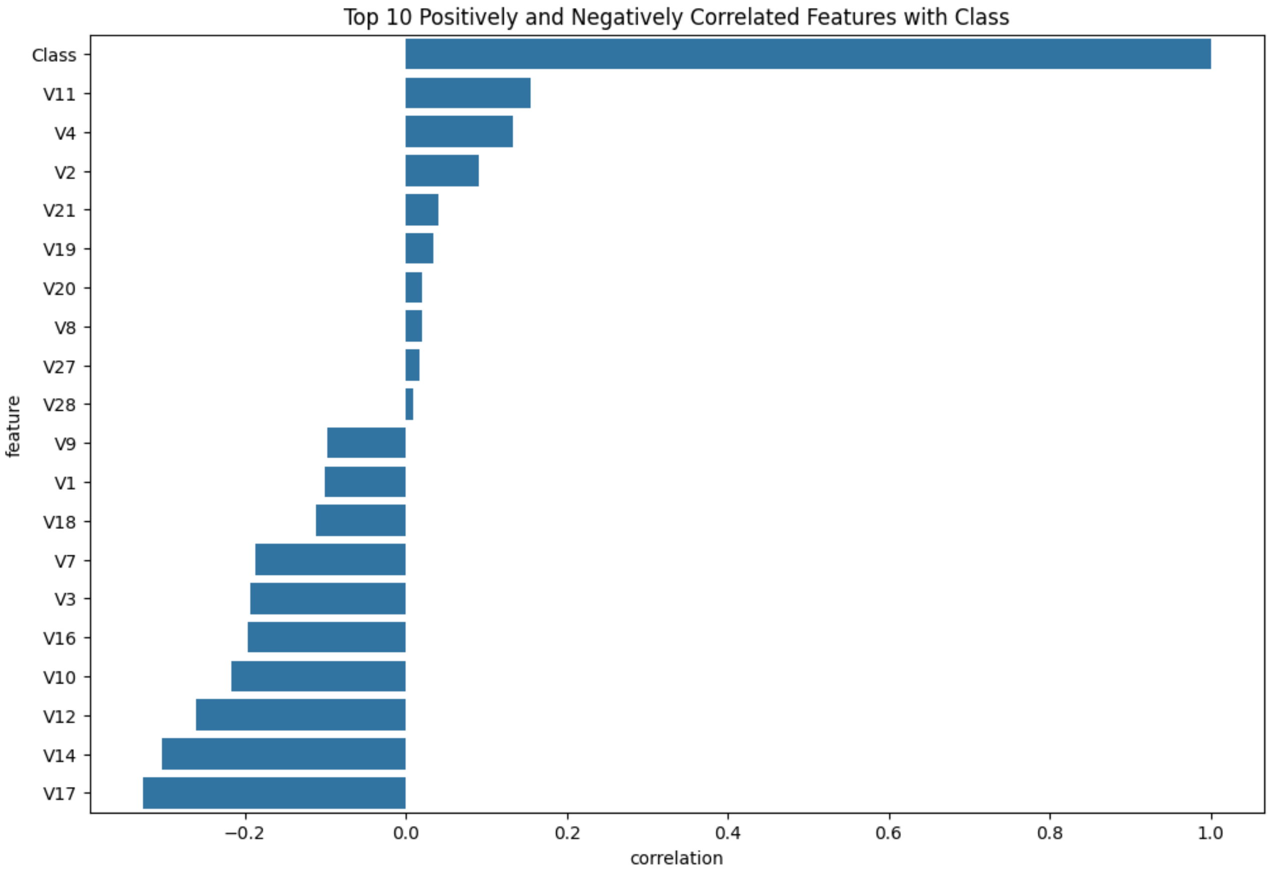


Interpretation: Fraudulent transactions are more likely to occur at specific hours, possibly reflecting attacker behavior or system vulnerabilities.



Interpretation: There are certain days of the week with higher fraud rates, which may inform monitoring and prevention strategies.

- **Correlation and Feature Importance**

  Interpretation: Some features show positive or negative correlation with fraud, but most individual correlations are weak, highlighting the need for complex models.

Data Preprocessing

- All missing values and duplicates were addressed.
- Feature engineering added valuable time, transaction, and geolocation features.
- Categorical variables were encoded, and all features were scaled for modeling.
- Class imbalance was mitigated using SMOTE, improving model training.

Future Work

- **Task 2 - Model Building and Training**
  - **Data Preparation:**
    - Separate features and target variables ([‘Class’ for creditcard, ‘class’ for Fraud_Data]).
    - Perform a train-test split to ensure robust model evaluation.
  - **Model Selection:**
    - Build and compare two models for each dataset:
      - Logistic Regression: As a simple, interpretable baseline.
      - One powerful ensemble model: Random Forest or a Gradient Boosting model (e.g., LightGBM, XGBoost).
  - **Model Training and Evaluation:**
    - Train both models on both datasets.
    - Use appropriate metrics for imbalanced data (AUC-PR, F1-Score, Confusion Matrix).

        - Clearly justify which model is considered "best" and why, based on the evaluation metrics and business context.

- **Task 3 - Model Explainability**

    - Use SHAP (Shapley Additive exPlanations) to interpret the best-performing model.
    - Generate and interpret SHAP plots (e.g., Summary Plot, Force Plot) to understand global and local feature importance.
    - In the final report, explain what these plots reveal about the key drivers of fraud in the data.

- Further optimize model hyperparameters for improved performance.

- Explore additional feature engineering, especially leveraging external data sources.

- Implement real-time fraud detection pipelines.

- Enhance model explainability and integrate business feedback for continuous improvement.

- Deploy the best-performing model in a production environment and monitor its performance.

- **Advanced Model Architectures:** Experiment with deep learning models (e.g., neural networks, autoencoders) and ensemble techniques to further improve fraud detection accuracy.

- **Model Comparison:** Systematically compare the performance of traditional machine learning models (e.g., Logistic Regression, Random Forest, XGBoost) with advanced models on key metrics such as Precision, Recall, F1 Score, ROC AUC, and PR AUC.

- **Threshold Optimization:** Investigate optimal decision thresholds to balance false positives and false negatives according to business needs.

- **Robustness Testing:** Evaluate model robustness to data drift, adversarial attacks, and changes in fraud patterns over time.

- **Cross-Validation:** Employ cross-validation and out-of-sample testing to ensure model generalizability and prevent overfitting.

- **Interpretability:** Continue to use SHAP and other explainability tools to interpret model predictions and validate feature importance.

Conclusion

The interim results demonstrate a strong foundation for effective fraud detection using machine learning. Comprehensive EDA, thoughtful preprocessing, and robust modeling have yielded promising results, particularly with ensemble methods like XGBoost. Continued work will focus on further improving model performance, interpretability, and real-world applicability.