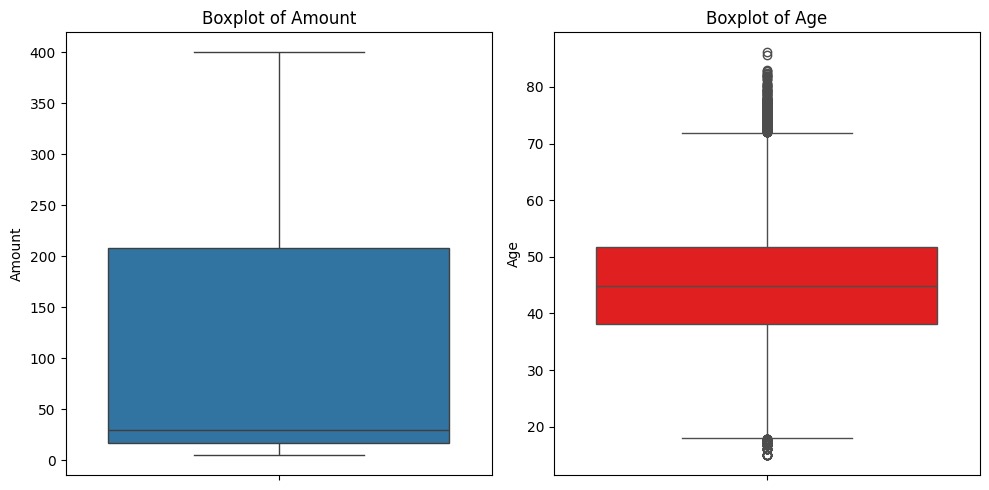
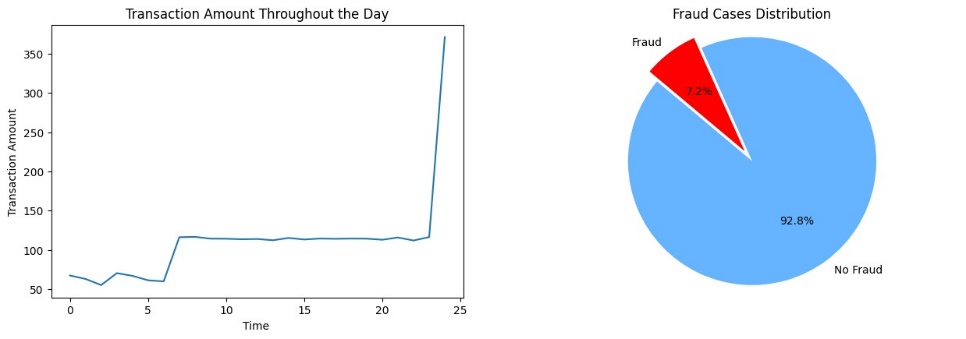
Big Data Analytics in Banking: Transforming Risk Management and Fraud Detection Processes

## Boxplot of Amount and Age

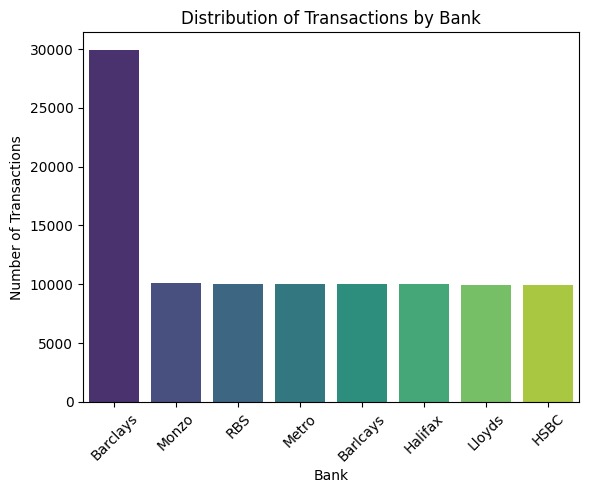


Boxplot of Amount and Age

## Transaction Amount Throughout the Day

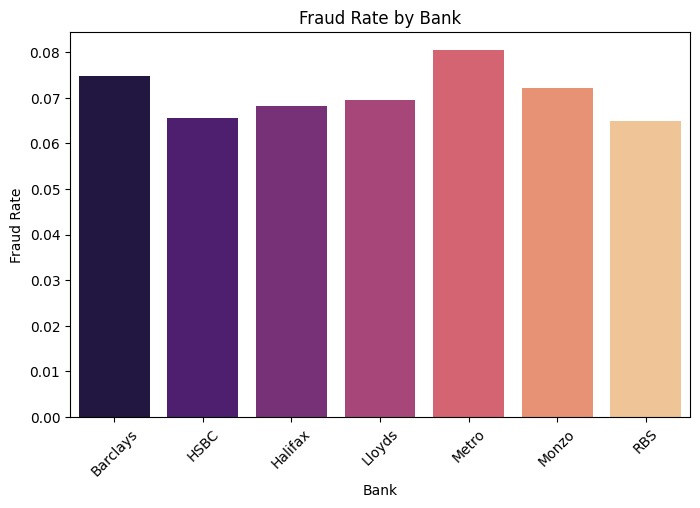


*Line graph of Transaction Amount Throughout the Day and Pie Chart of Fraud cases Distribution*



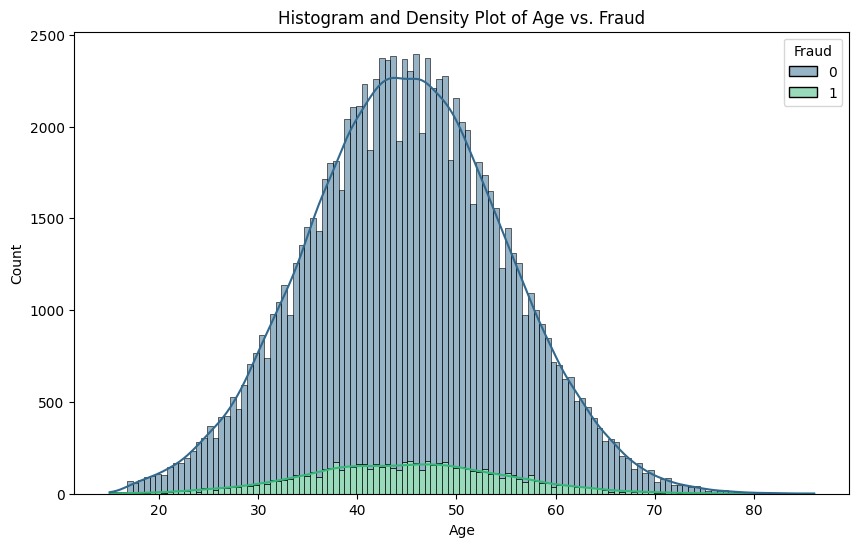
Transaction Amount Throughout the Day

## Fraud Cases Distribution



Bar chart of Fraud Cases Distribution

## Histogram and Density Plot of Age vs. Fraud



Histogram and Density Plot of Age vs. Fraud

## Distribution of Card Types

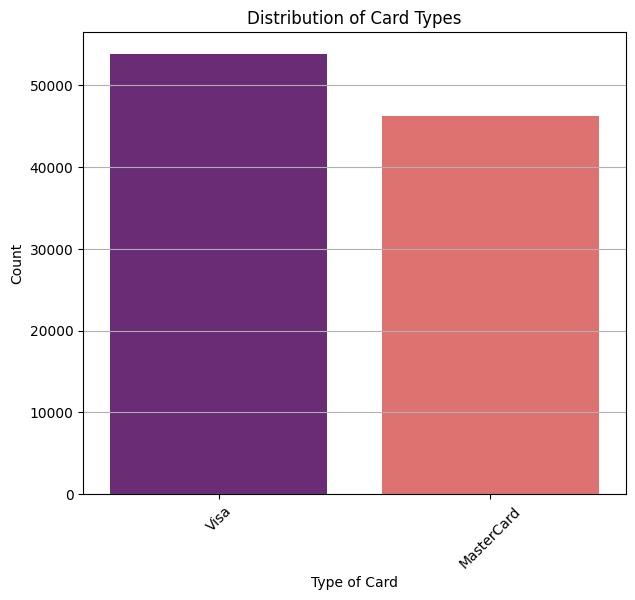
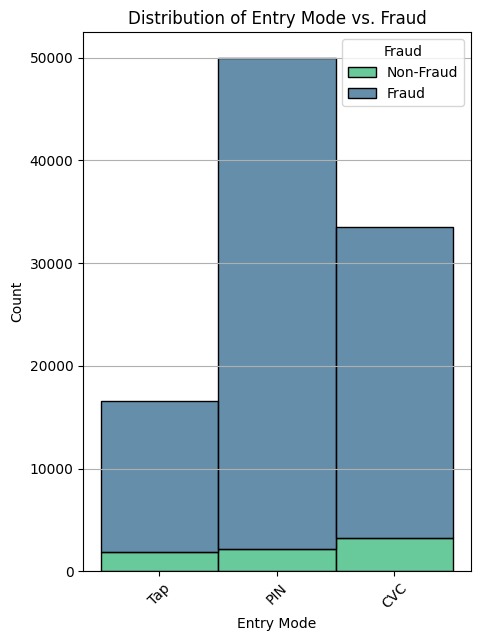


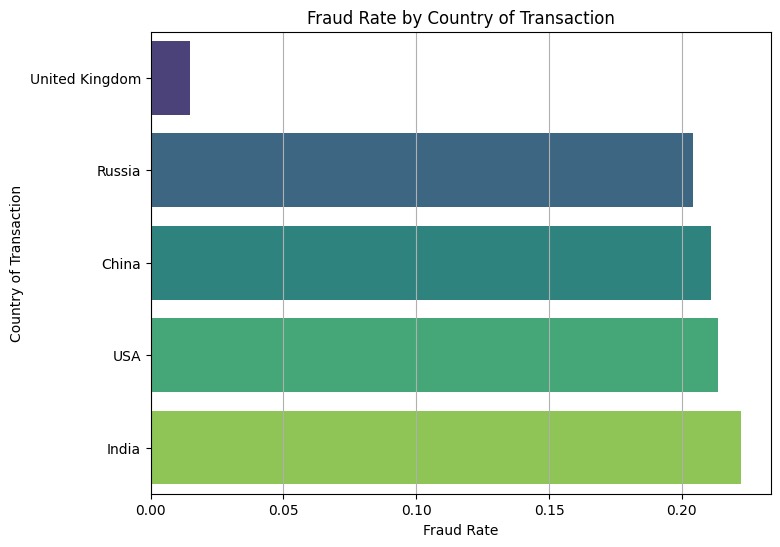
Figure 8 Types of Card distribution

## Distribution of Entry Mode vs. Fraud



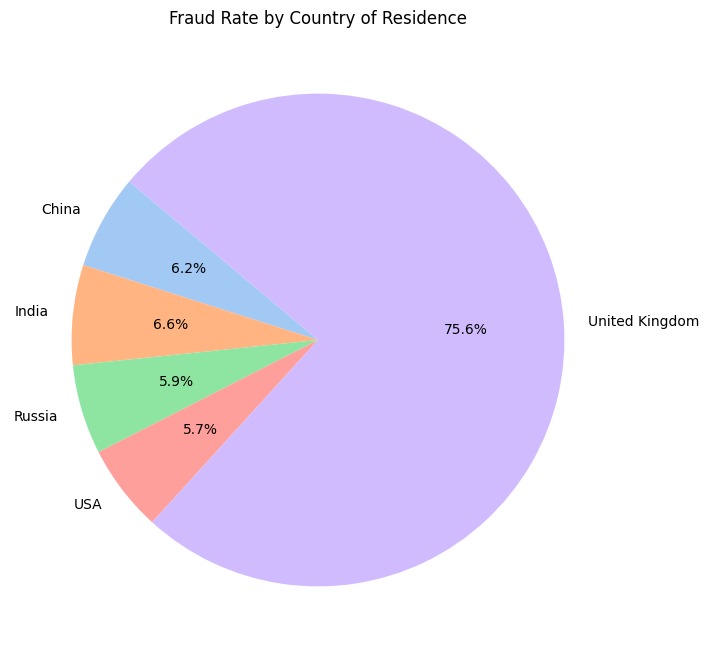
Distribution of Entry Mode vs. Fraud

## 8. Fraud Rate by Country of Transaction



Fraud Rate by Country of Transaction

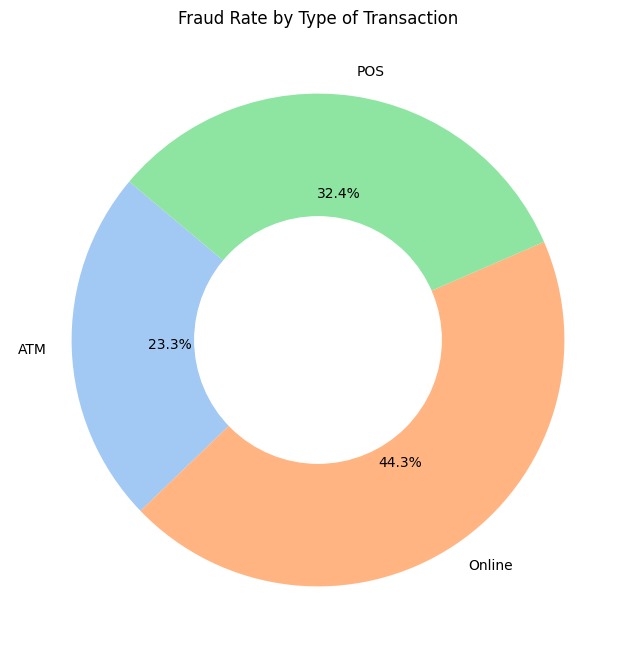
## 9. Fraud Rate by Country of Residence



*Fraud Rate by Country of Residence*

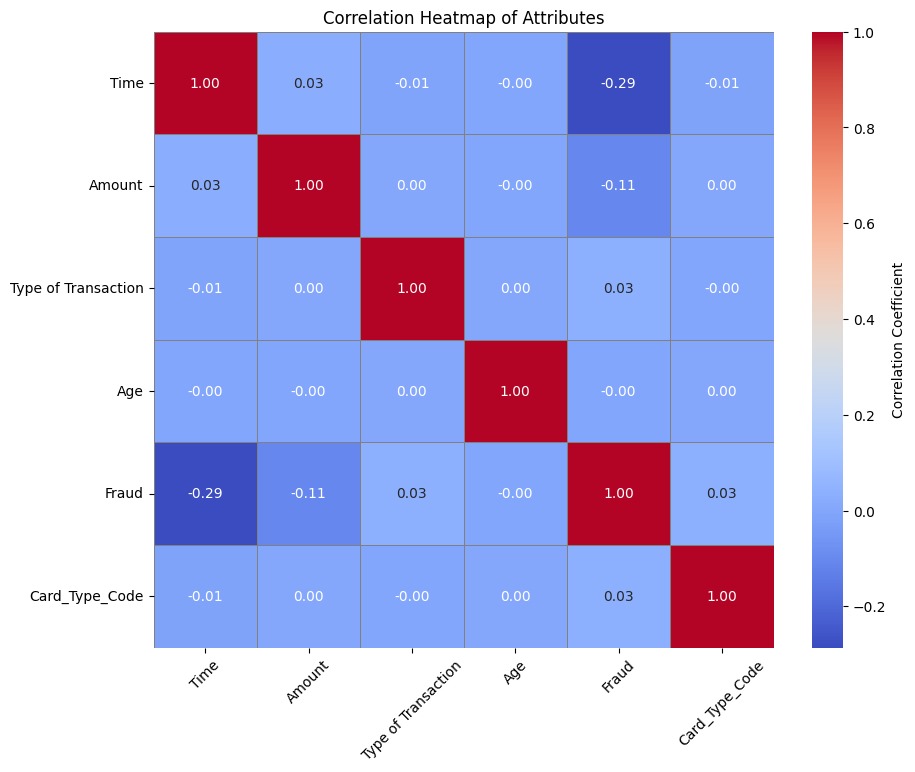
## 

## Fraud Rate by Type of Transaction



Fraud Rate by Type of Transaction

## 11. Correlation Heatmap of Attributes



Correlation heatmap of Attributes

# Logistic Regression

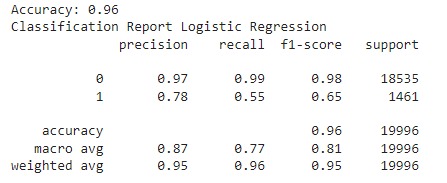


Figure 14. Logistic regression

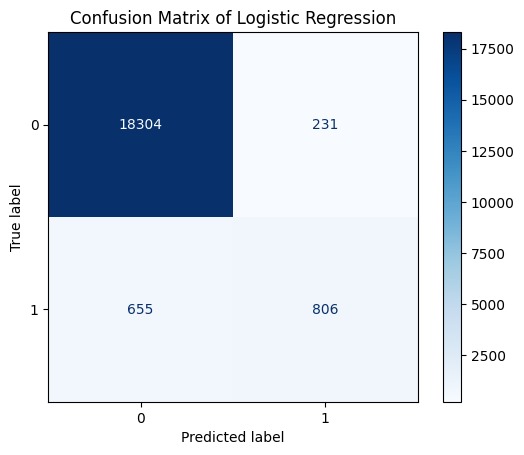


Figure 15. Confusion matrix

## **Decision Tree model**

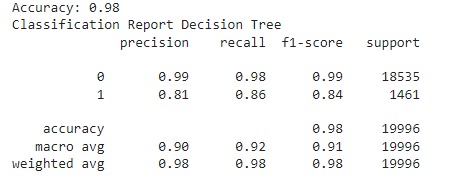
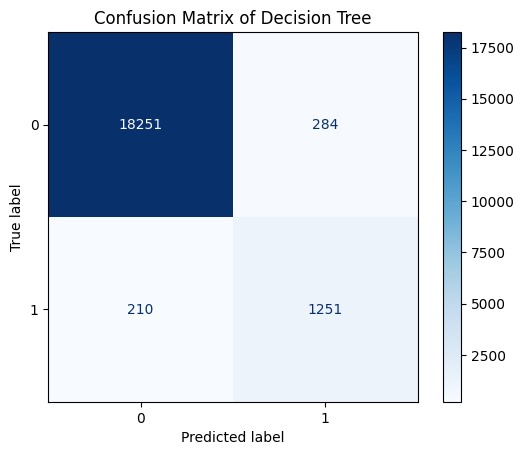


Figure 16 Decision tree classification



*Decision tree confusion matrix*

## **Random Forest model**

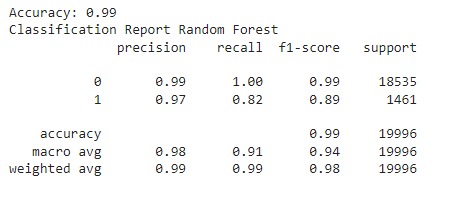
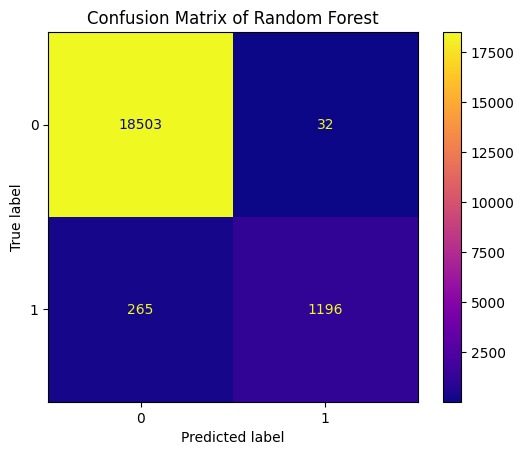


Figure 20. Confusion matrix for Random Forest

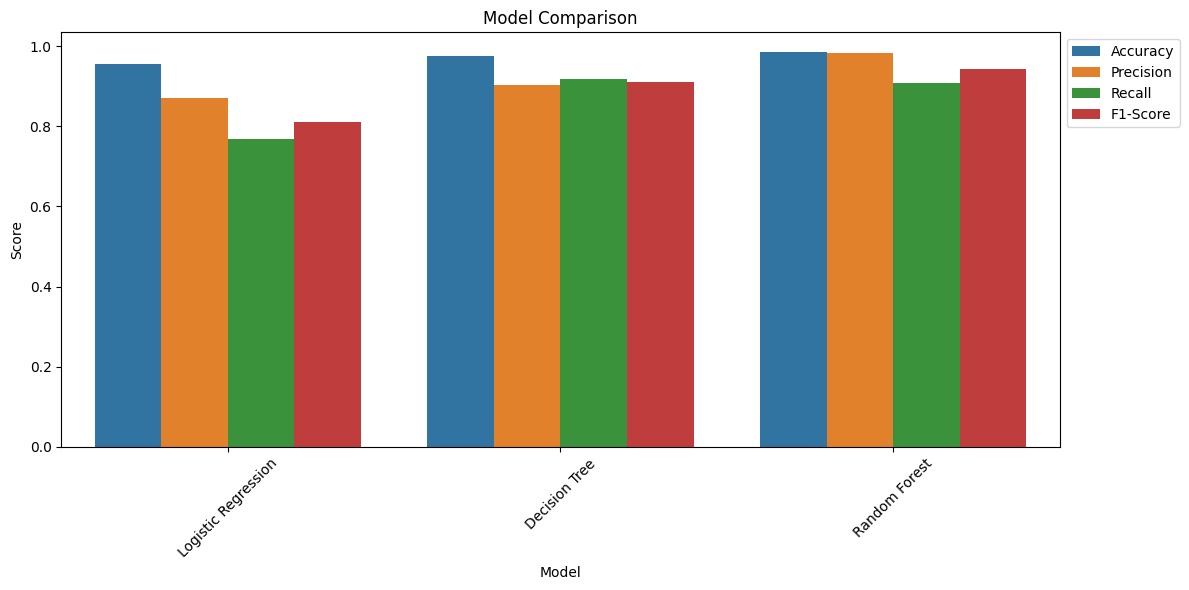


Confusion matrix of random forest

## Model Comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Weighted Precision** | **Weighted Recall** | **Weighted F1-Score** |
| Logistic Regression | 0.96 | 0.95 | 0.96 | 0.95 |
| Decision Tree | 0.98 | 0.98 | 0.98 | 0.98 |
| Random Forest | 0.99 | 0.99 | 0.99 | 0.98 |

Comparison of Models



Comparison of Models