

Building An Intelligent Fraud Detection System

Group 4 Capstone Project





Project Overview

Fraud trends vary by region, with West and Southern Africa facing increased scams, while Asia deals with telecommunication fraud. Commercial banks and health insurers are the most affected, with identity fraud making up 45% of cases in 2023 and expected to reach 50% by year-end. Scam-related frauds have surged by 56% in 2024, surpassing digital payment fraud, with scams now comprising 23% of fraudulent transactions. In Kenya, financial fraud is rising, exemplified by Kiwipay Kenya Limited's Ksh2.3 billion freeze due to suspected debit card fraud. The Central Bank of Kenya links the surge to ICT adoption, low security awareness, and cyber threats, stressing the need for stronger security and public education.


PROJECT OUTLINE

Business Problem
Objectives
Data Source
Visualizations
Modelling
Conclusion

Business Problem

The Central Bank of Kenya (CBK) has identified key risk factors, including the rapid adoption of digital financial services, low consumer awareness of financial security, and evolving cyber threats. Additionally, the rise in scam-related frauds, identity theft, and card skimming has led to significant financial losses, eroding public trust in the financial sector. To address these challenges, we aim to develop a robust fraud detection and prevention model that leverages machine learning, artificial intelligence, and real-time transaction monitoring. This model will enhance the ability of financial institutions in Kenya to proactively detect fraudulent activities, mitigate risks, and strengthen cybersecurity measures, ensuring real-time fraud prevention and minimizing financial losses.

Objectives

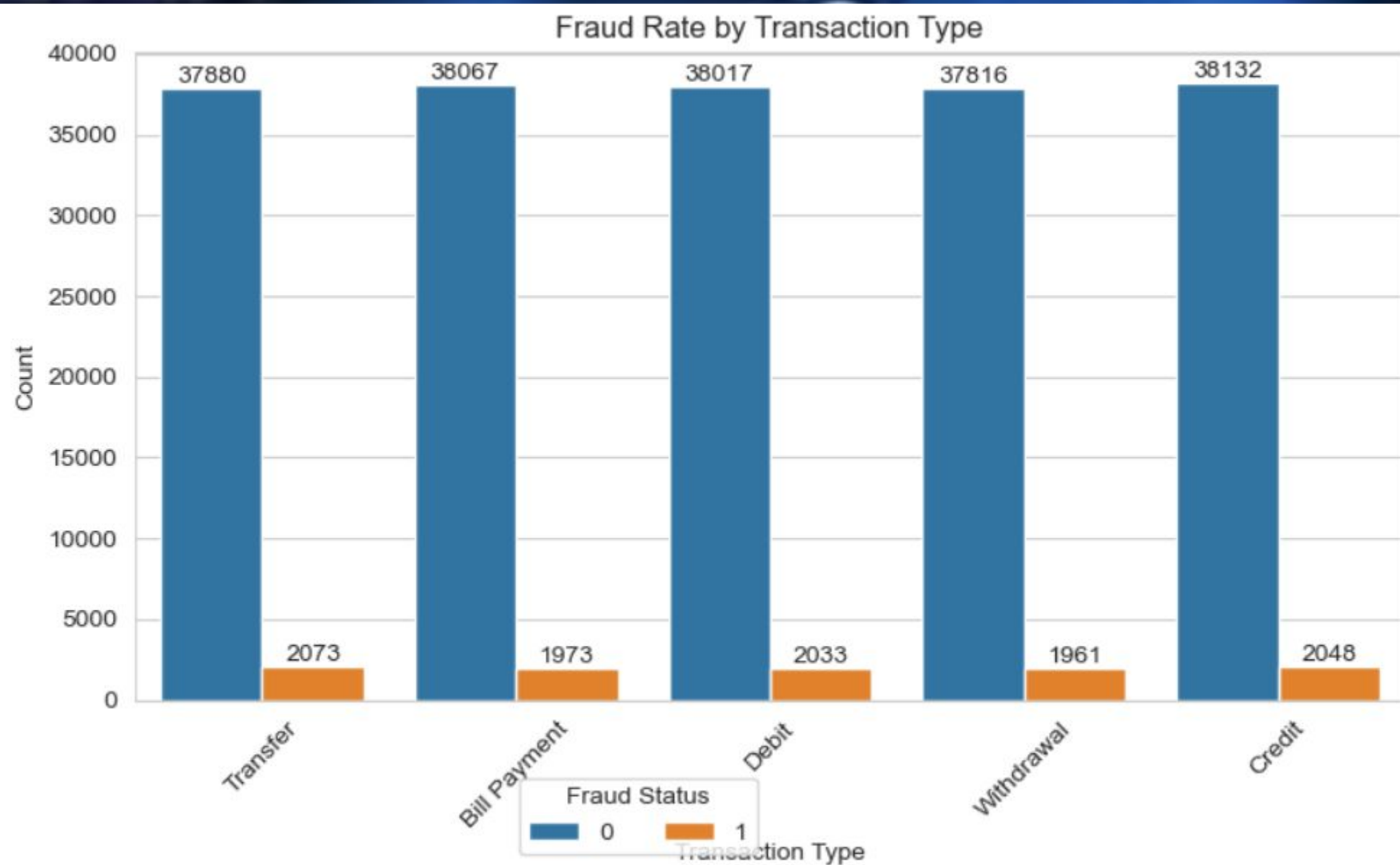
- 
1. Analyze card transaction patterns to detect fraudulent activity.
 2. Develop predictive models to accurately classify transactions as fraudulent or legitimate.
 3. Examine the impact of demographics, such as age and gender, on fraud risks.
 4. Identify peak fraud periods based on transaction dates and times.
 5. Design a real-time fraud detection model for identifying suspicious card transactions.

Data Source

The dataset was obtained from
[<https://www.kaggle.com/datasets/marusagar/bank-transaction-fraud-detection>]

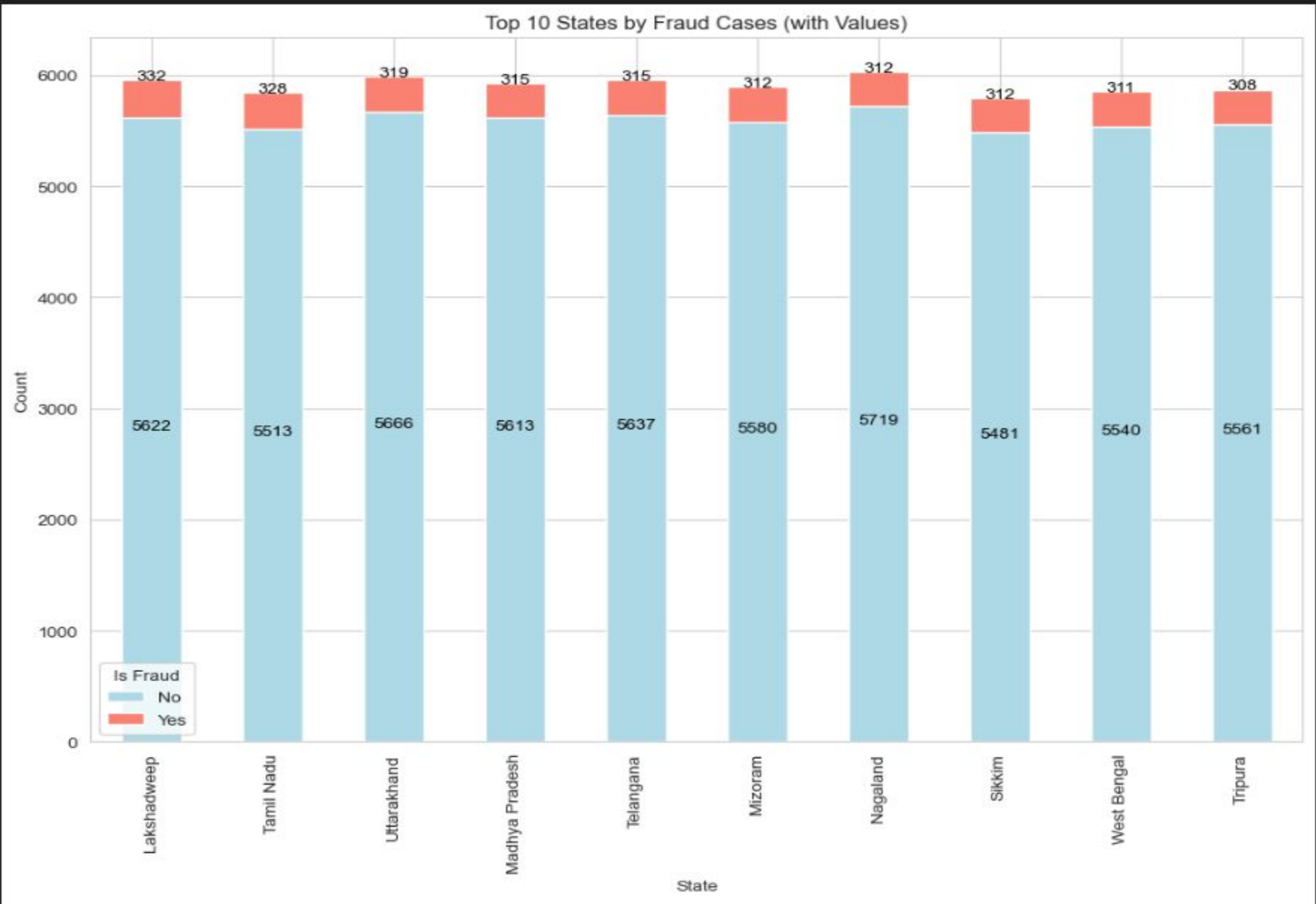
Visualization

Distribution of Fraud Cases By Transaction Type



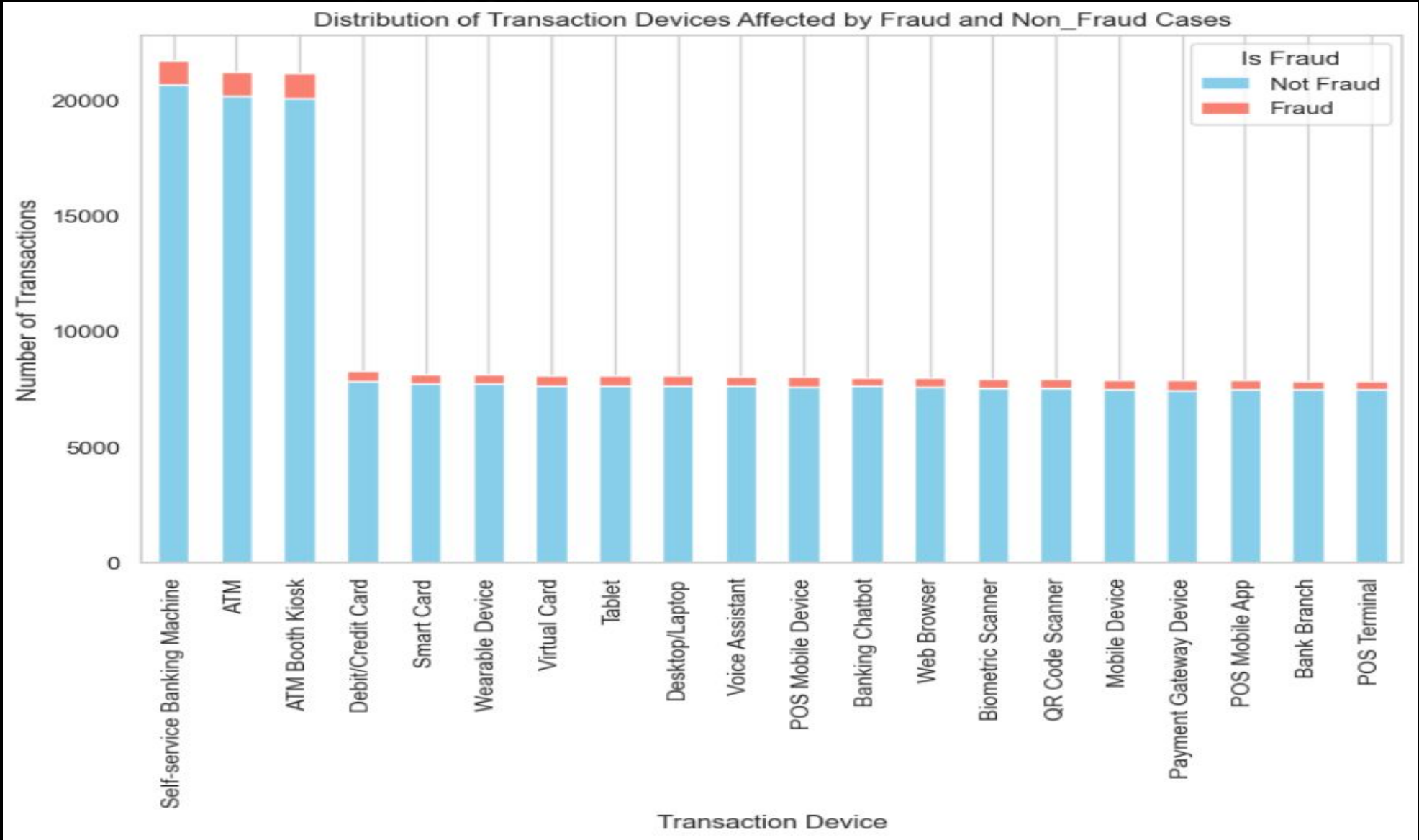
The transaction type with the highest cases of fraud is: Transfer

Distribution of Fraud Cases By State



The state with the highest fraud cases is:
Lakshadweep

Distribution of Fraud Cases By Transaction Devices

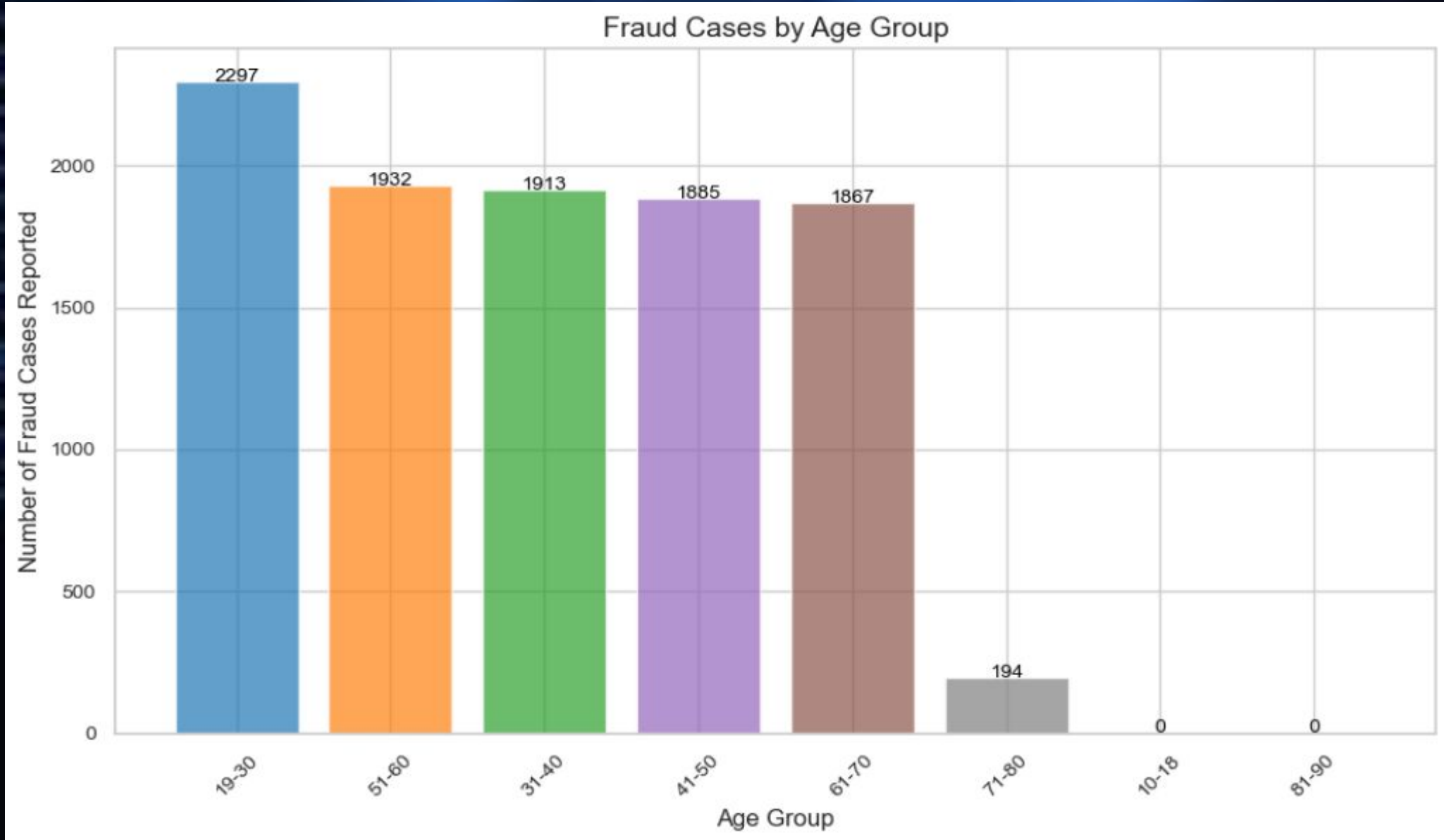


The transaction device with the highest fraud cases is: ATM Booth Kiosk

The transaction device with the lowest fraud cases is: POS Terminal



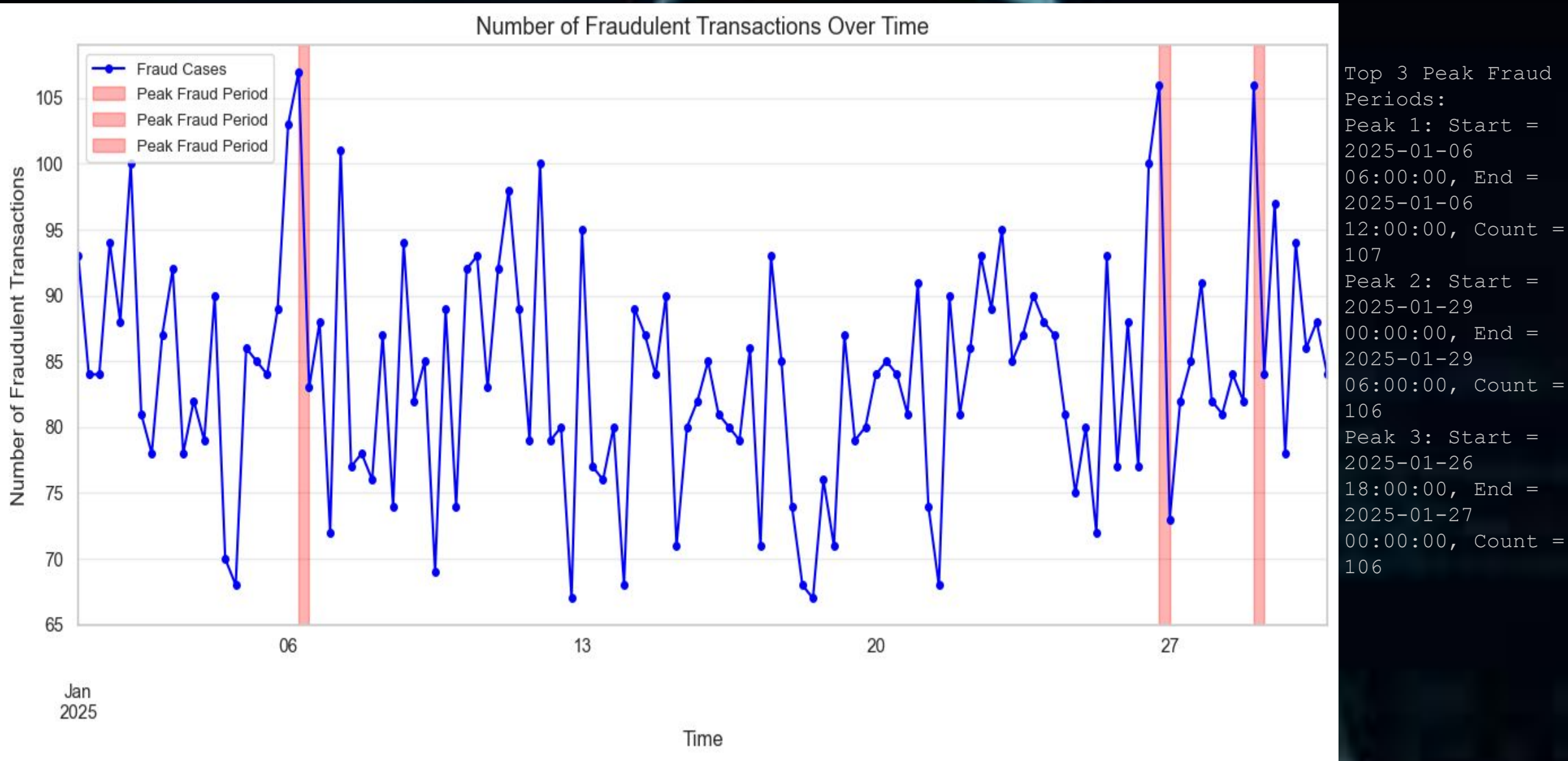
Distribution of Fraud Cases By Age Group



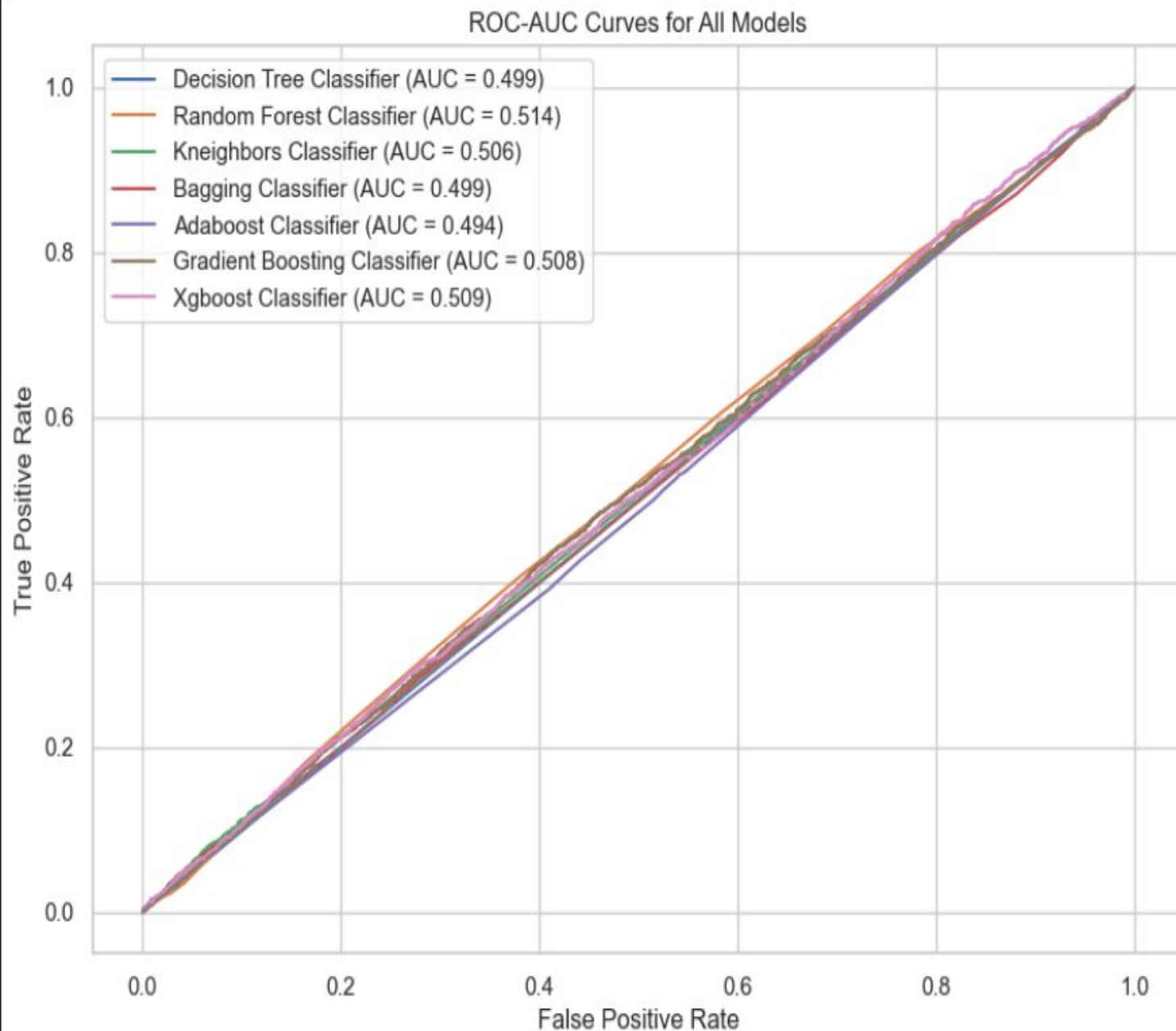
For customers at the age 19-30 and 51 - 60 show significantly higher numbers of fraud cases which further indicates that individuals in the age groups are more vulnerable to fraud.



Distribution of Fraudulent Transactions Over Time



Modelling



Decision Tree Classifier pipeline test accuracy: 0.9026
Random Forest Classifier pipeline test accuracy: 0.9496
Kneighbors Classifier pipeline test accuracy: 0.9485
Bagging Classifier pipeline test accuracy: 0.9496
Adaboost Classifier pipeline test accuracy: 0.4860
Gradient Boosting Classifier pipeline test accuracy: 0.6040
Xgboost Classifier pipeline test accuracy: 0.9496
Classifier with best accuracy: Random Forest Classifier (0.9496)

Random Forest (0.9496), Bagging (0.9496), and XGBoost (0.9496) achieved the highest accuracy & are the best models for this dataset, achieving nearly identical high accuracy.

FINDINGS

1. Customers aged **19-30** and **51-60** show significantly higher instances of fraud, indicating that these age groups are more vulnerable.
2. **Transfers** account for the highest number of fraud cases, with **2,073** reported, followed closely by **credit transactions** with **2,048** cases.
3. The **ATM Booth Kiosk, ATM, and Self-Service Machine** pose the highest fraud risk among transaction devices.
4. The top three peak periods for fraud incidents are during holidays.

Conclusions

1. Financial institutions should analyze fraud trends by age group to assess risk and develop targeted awareness campaigns to reduce fraud incidents.
2. A deeper analysis should be conducted to establish effective controls for mitigating fraud risks in high-frequency areas, particularly transfers and credit transactions.
3. Stronger controls should be implemented for ATM Booth Kiosks, ATMs, and Self-Service Machines, as they report the highest instances of fraud.
4. Financial institutions should enhance fraud monitoring during holidays.



Thank you

**FRAUD
PREVENTION**

A hand in a blue plaid suit jacket points towards a glowing blue circular digital interface. The interface features concentric circles, dashed lines, and small white dots. The words 'FRAUD PREVENTION' are prominently displayed in the center of the circle. The background is dark blue with abstract geometric patterns and light streaks.