FRAUDULENT CLAIM DETECTION

PREDICTIVE MODELING TO PREVENT INSURANCE FRAUD

INTRODUCTION

Insurance fraud leads to significant financial losses and operational inefficiencies. Manual fraud detection methods delay claim processing and increase costs.

PROBLEM STATEMENT

Current fraud investigations rely heavily on manual reviews, leading to delays and errors. Legitimate claims suffer while fraudulent ones may go unnoticed

PROJECT GOAL

To develop a model to classify claims as fraudulent or legitimate using features like claim amount, incident details, customer profile, and approval timelines

ABOUT THE DATASET

- The dataset contains 1000 entries and 39

- columns, including:

- Customer: age, occupation, hobbies

- Policy: coverage, premiums

- Incident: type, severity, location

- Claim: amount, type

- Vehicle: make, model, year

DATA PREPROCESSING & CLEANING

Missing values were addressed (e.g.,property_damage, police_report). Categorical variables were encoded using Label Encoding.Data normalized for modeling
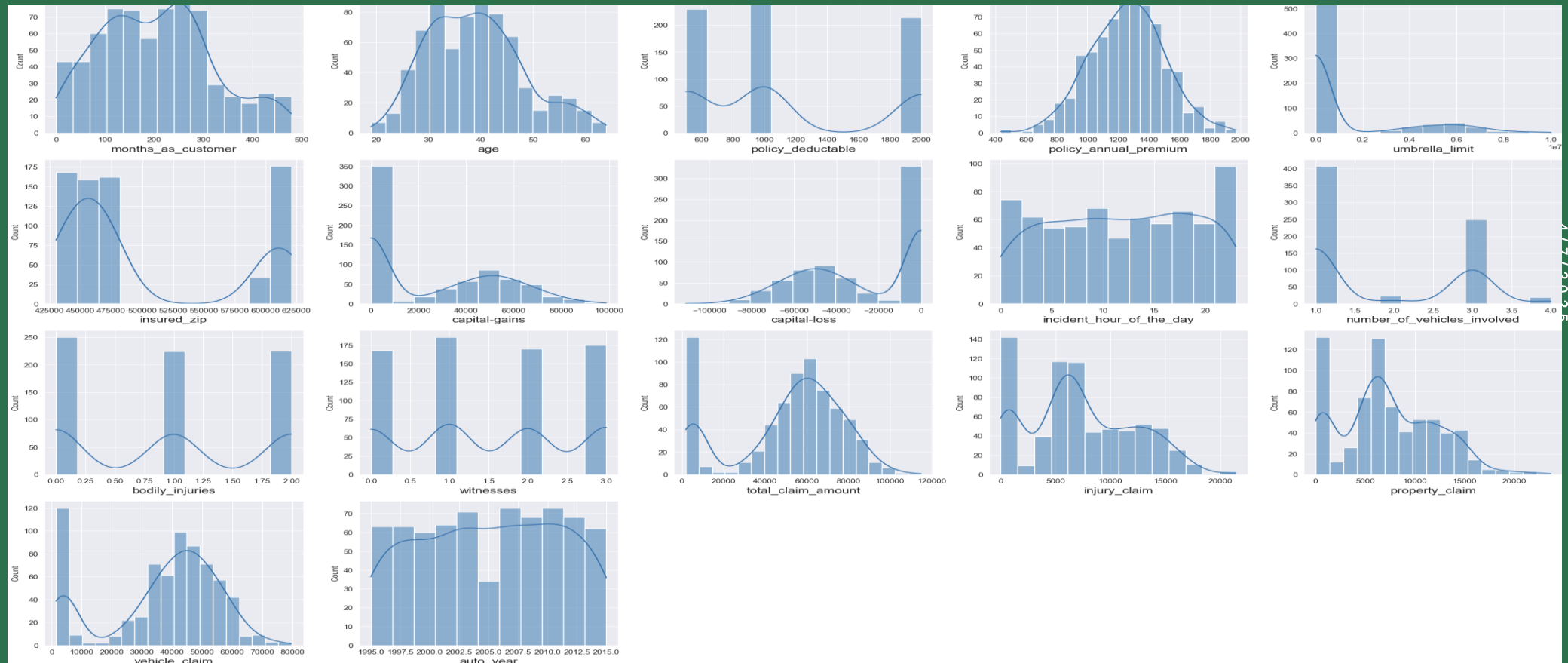
- FRAUDLENT VS LEGITIMATE CLAIMS

Fraudulent

Fraudulent   Legitimate
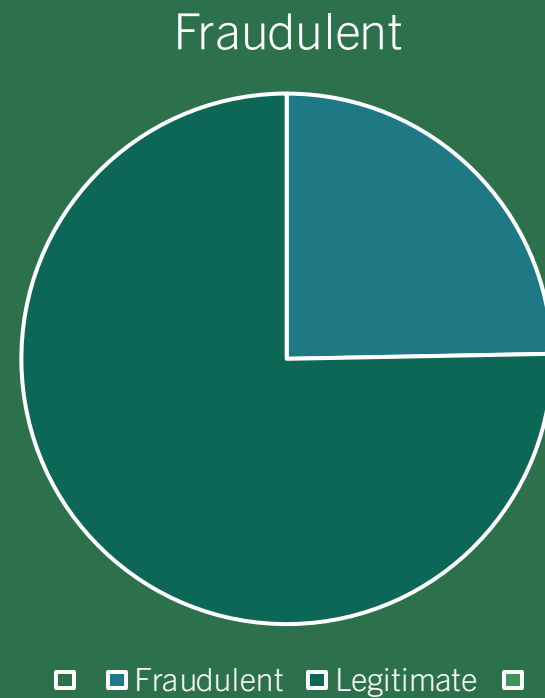
TOP PREDICTIVE FEATURES

Top 10 Important Features for Fraud Detection
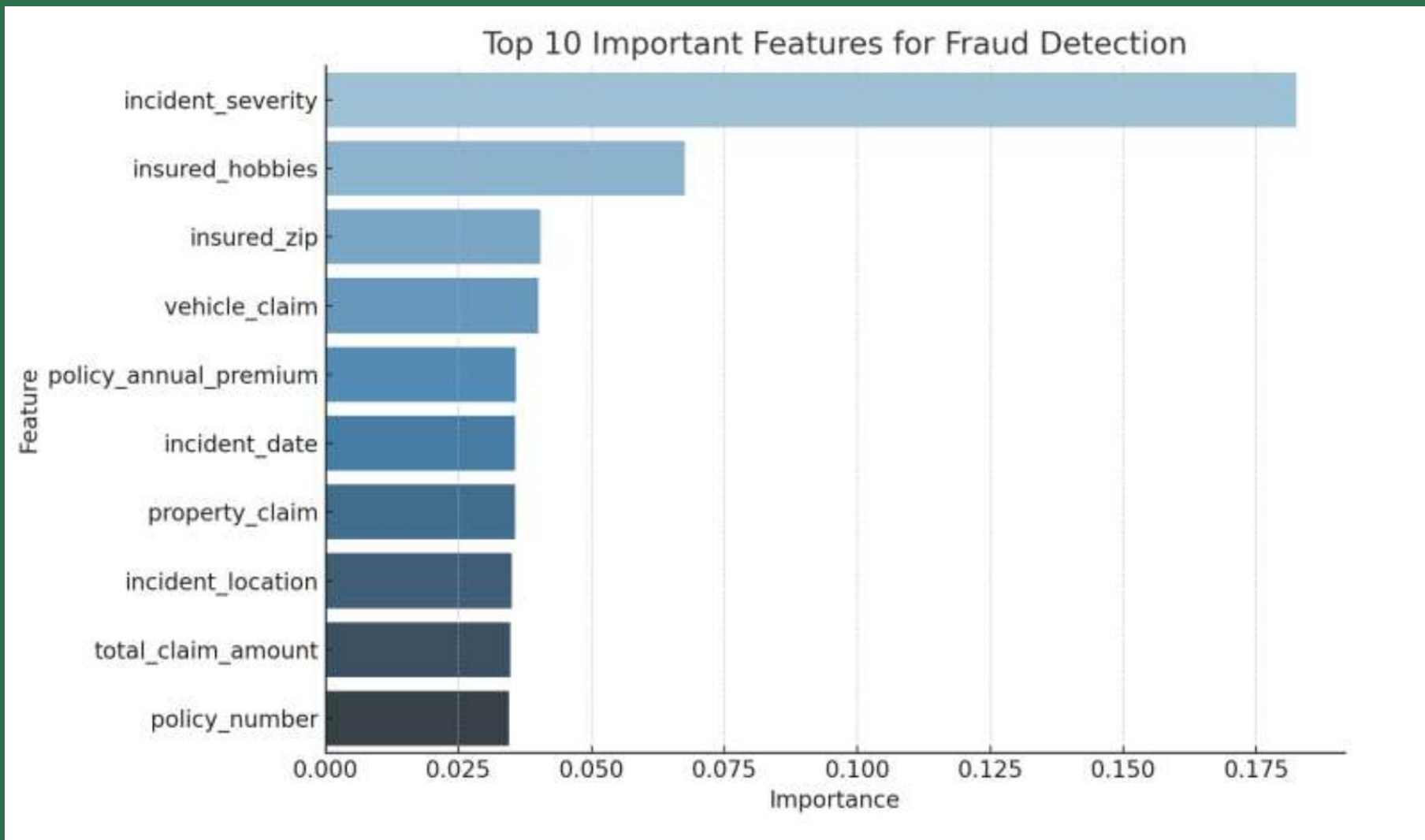
Visualising the distribution of selected numerical features using appropriate plots to understand their characteristics

Investigating the relationships between numerical features to identify potential multicollinearity or dependencies and visualising the correlation structure using an appropriate method to gain insights into feature relationships
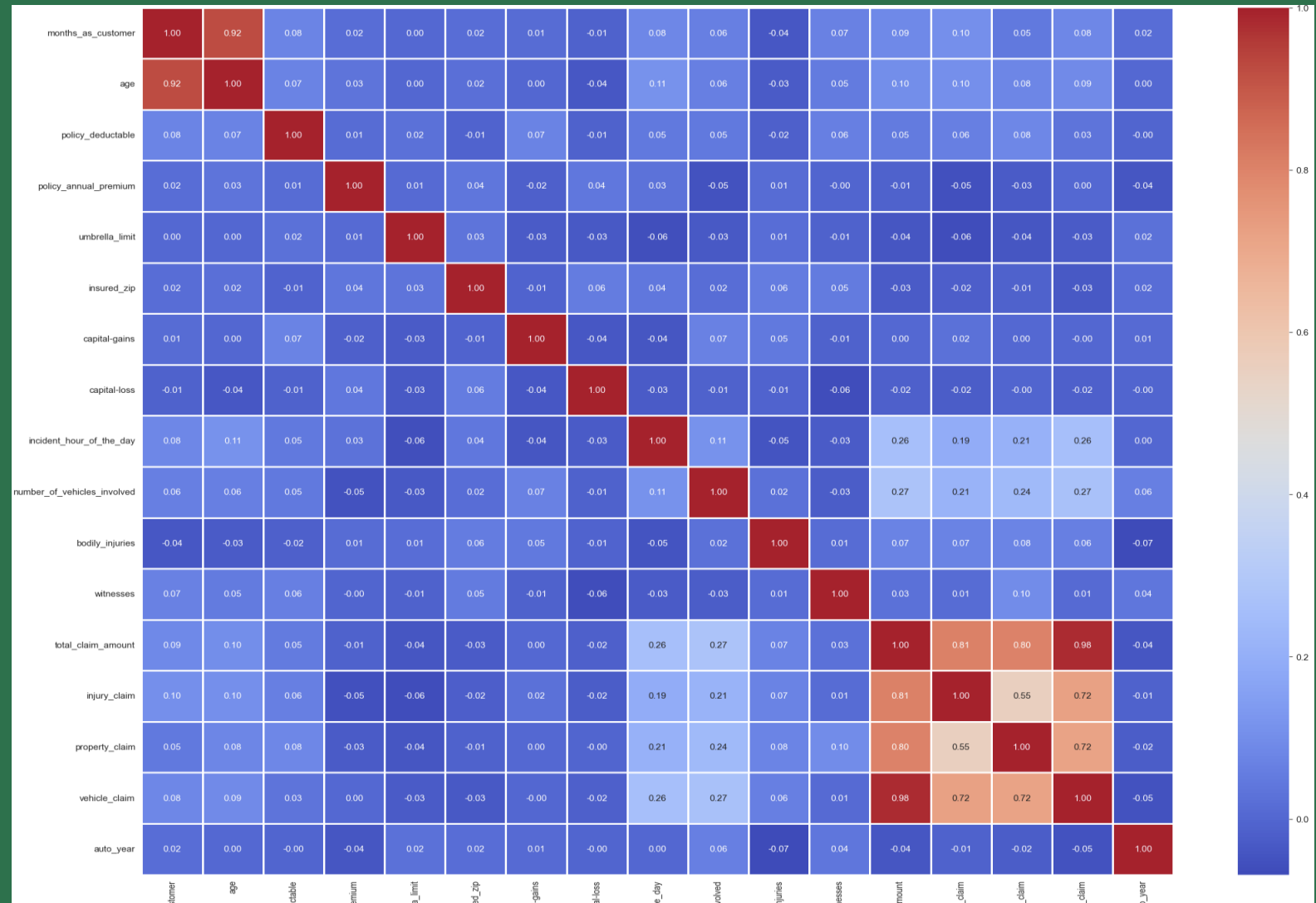
Examining the distribution of the target variable to identify potential class imbalances using visualisation for better understanding
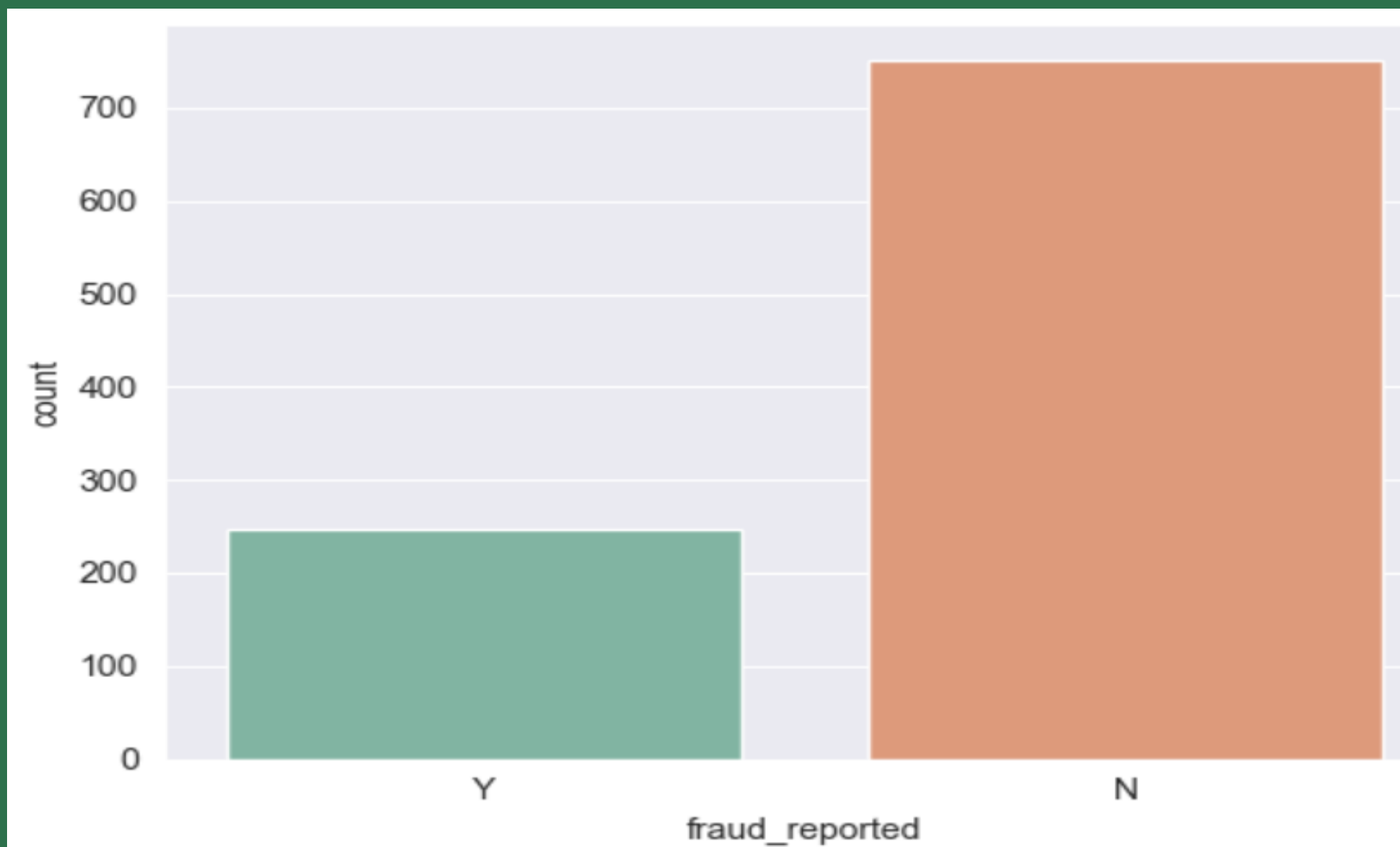
MODELING APPROACH

- We trained a Random Forest classifier using encoded features. Train-test split used to evaluate model performance with cross-validation

MODEL PERFORMANCE

Random Forest showed high accuracy and

strong fraud detection capability. Metrics:

Precision, Recall, F1-Score used to measure

model quality

INSIGHTS FROM THE MODEL

High claim amounts and certain incident types are strong indicators of fraud.

Features like police reports and witness availability also show significance

RECOMMENDATIONS

- Use model scores to flag high-risk claims early.

- Allocate investigator resources based on fraud probability.

- Automate claim triage using predictive output.

BUSINESS IMPLICATIONS

- - Reduce financial losses from fraud.

- - Speed up processing of genuine claims.

- - Improve overall customer satisfaction and operational efficiency.
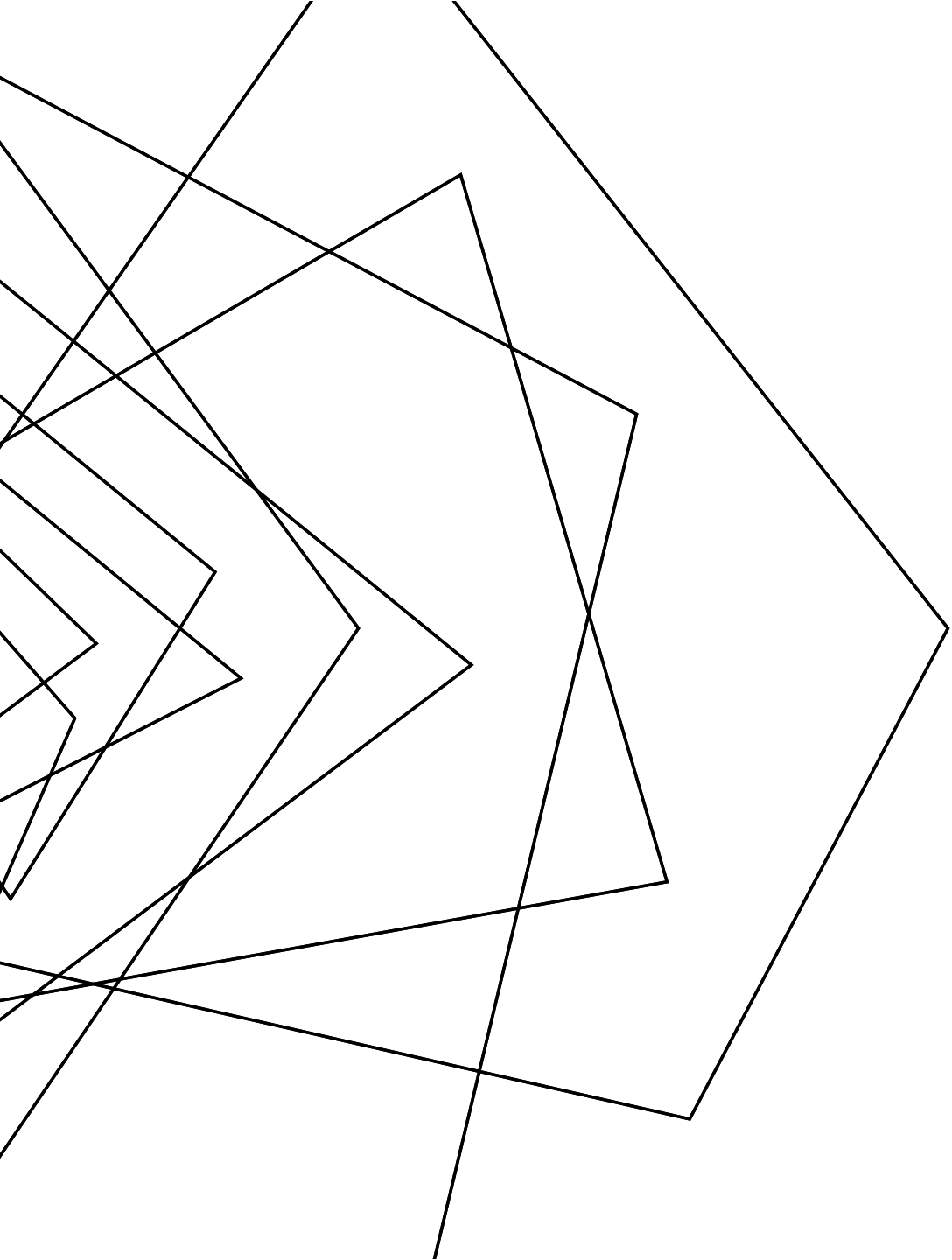
THANK YOU

Brita Tamm
502-555-0152
brita@firstupconsultants.com
www.firstupconsultants.com