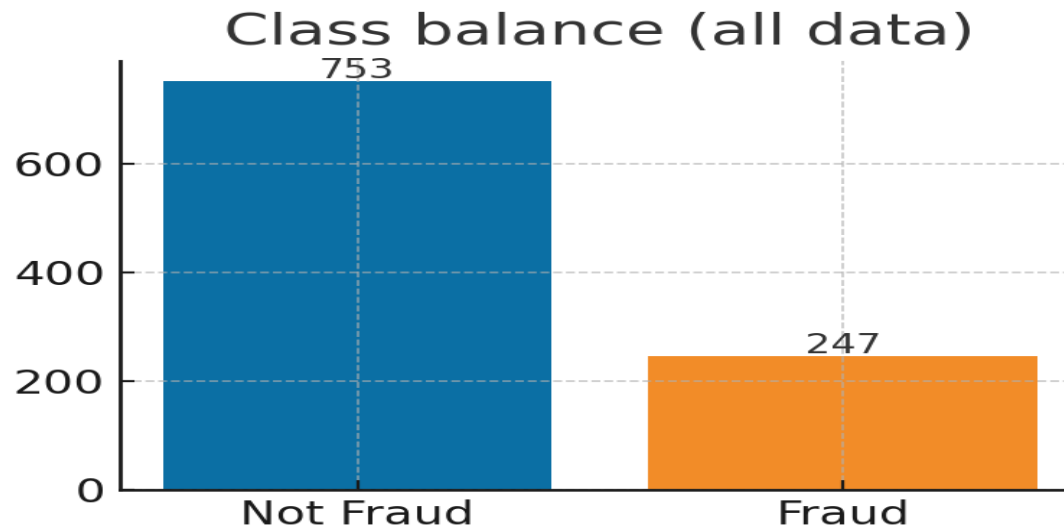


# Fraudulent Claim Detection

Using Machine Learning to prioritize high-risk claims

Prepared by: Naim Shaik



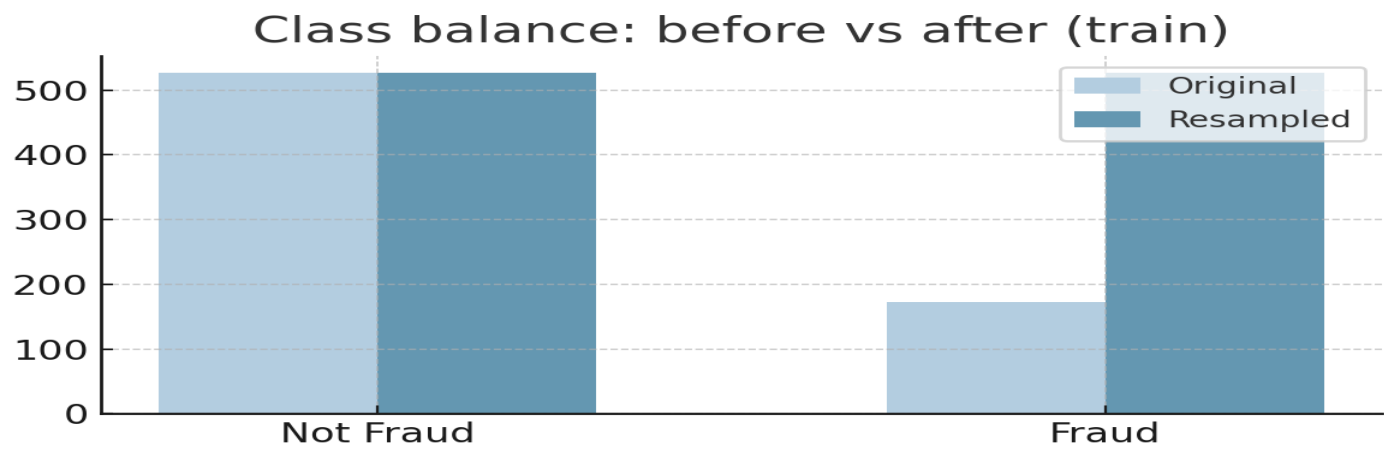
## Problem Statement & Business Value

Insurance fraud results in significant financial loss and operational inefficiency. A predictive model helps to:

- Prioritize claims for investigation
- Reduce manual effort & accelerate legitimate claims
- Minimize financial leakage from fraudulent payouts

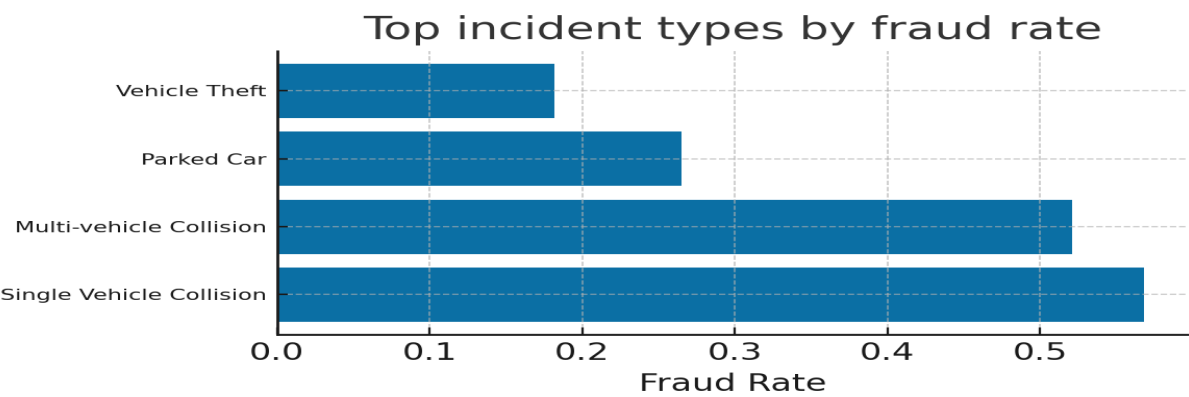
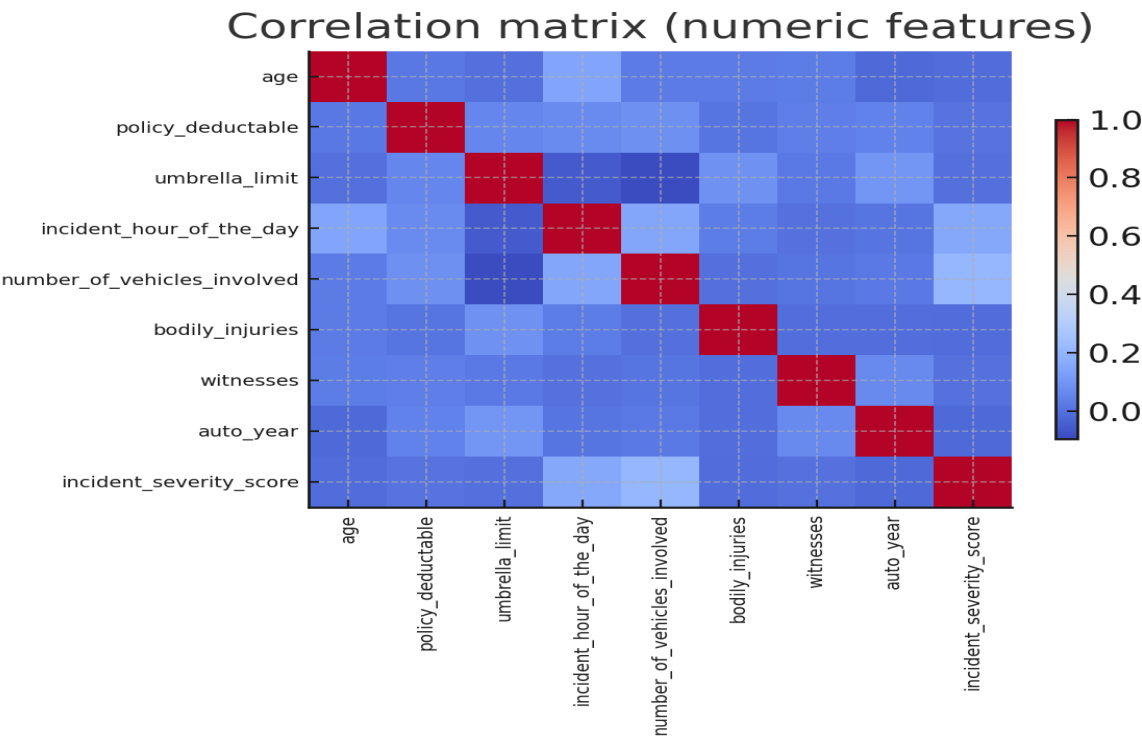
# Dataset Overview

Records: 1000 | Features: 28 | Fraud rate: 24.70%



# EDA Highlights

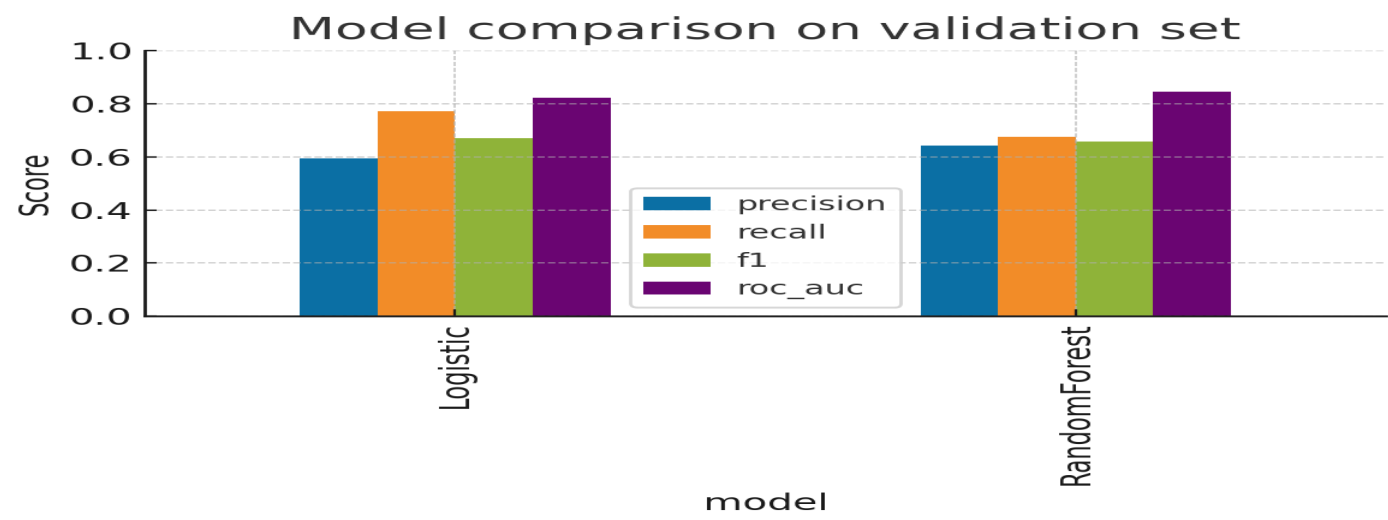
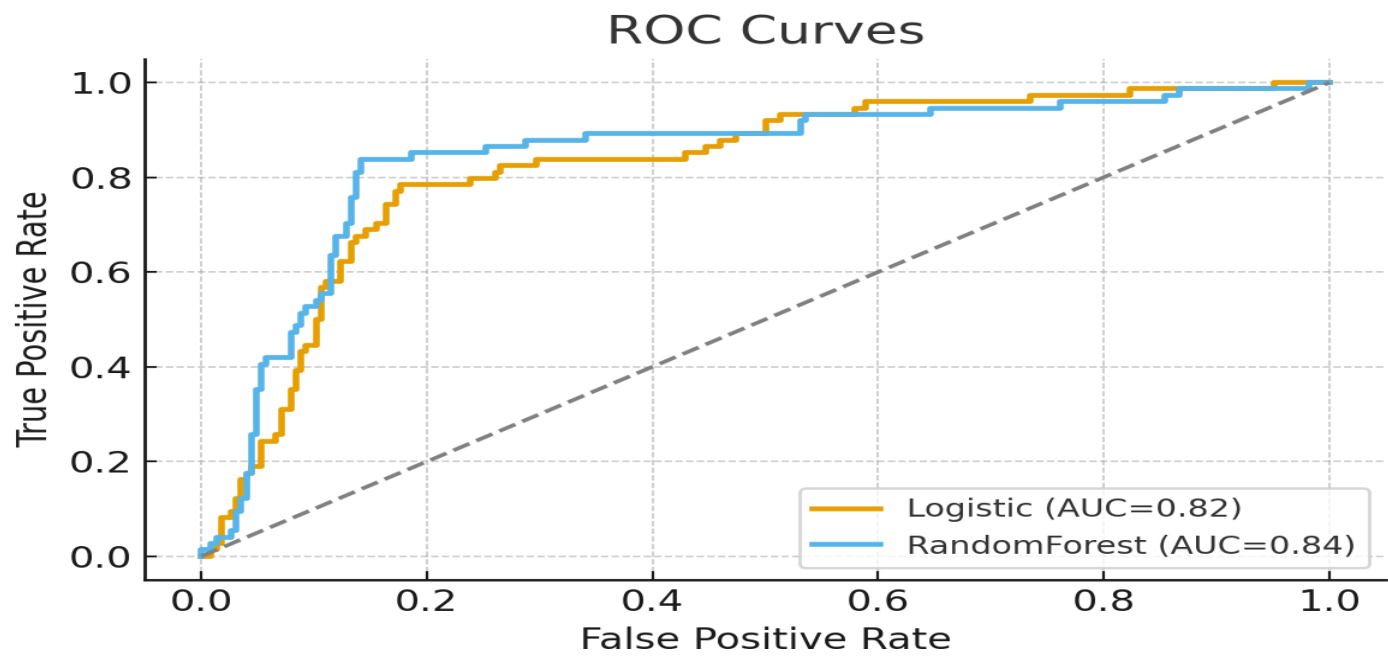
# Feature relationships & Target likelihood



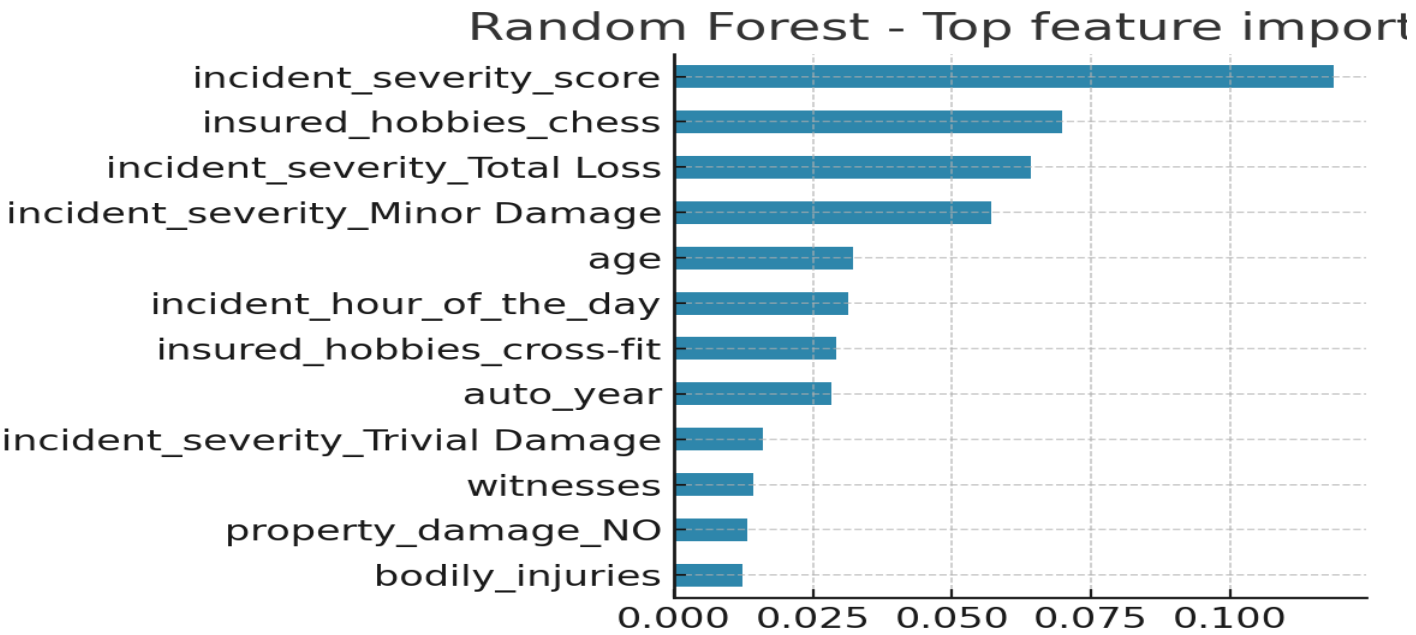
## Modeling Approach

- Resampling (oversampling minority) on training set
- Preprocessing: Standard scaling + One-hot encoding
- Models: Logistic Regression (interpretable) and Random Forest (high performance)

# Model Performance



# Top Predictive Features





## Business Insights from Model

1. High claim amount combined with low premium often indicates elevated fraud risk.
2. Certain incident types have much higher fraud rates — prioritize these for review.
3. Policy-incident gap (claims filed soon after policy start) is a notable signal.

## Recommendations & Next Steps

Short-term:

- Deploy Random Forest model as a triage tool; manually review top-scoring claims.

Mid-term:

- Integrate model with claims processing system; collect more temporal & behavioral features.

Long-term:

- Use SHAP for explainability; build feedback loop for model retraining.

# Thank you

Prepared by: Naim Shaik