By: Kiran K Rao

Case Study: Fraudulent Claim Detection

Objective

Global Insure intends to develop a model that categorizes insurance claims as either fraudulent or legitimate by utilizing historical claim information and customer profiles. The company seeks to forecast which claims may be fraudulent prior to their approval by analyzing features such as claim amounts, customer profiles, and claim types.

In light of this assignment, please address the following questions:

- How can we examine historical claim data to identify patterns that suggest fraudulent claims?

- Which features are the most indicative of fraudulent behavior?

- Is it possible to estimate the probability of fraud for a new claim based on previous data?

- What insights can be derived from the model that could enhance the fraud detection process?

Methodologies

1. Exploratory Data Analysis (EDA)

This process is utilized to comprehend the patterns, anomalies, discrepancies, and relationships within the data points.

Determine the correlations among the features.

Investigate the outliers present in the numerical variables.

Review the categorical variables.

2. Feature Engineering: Feature engineering was conducted on aspects such as the frequency of claims, historical patterns of claim amounts, and customer demographic information including age, income, fraud history, and the claim ratio relative to premiums, among others.

Outcome of EDA
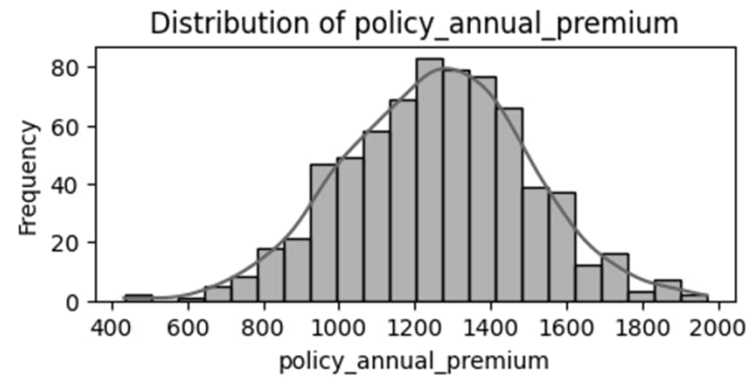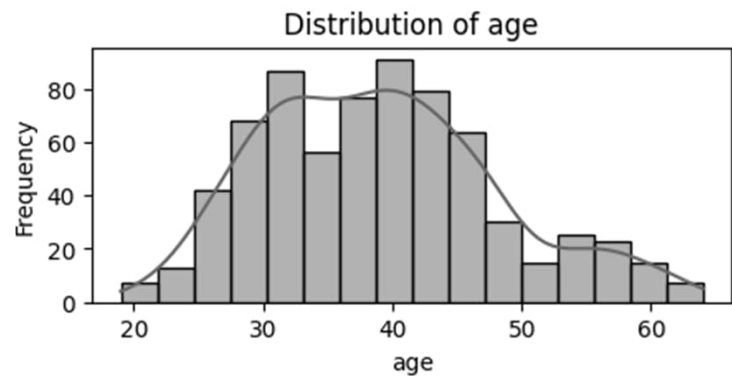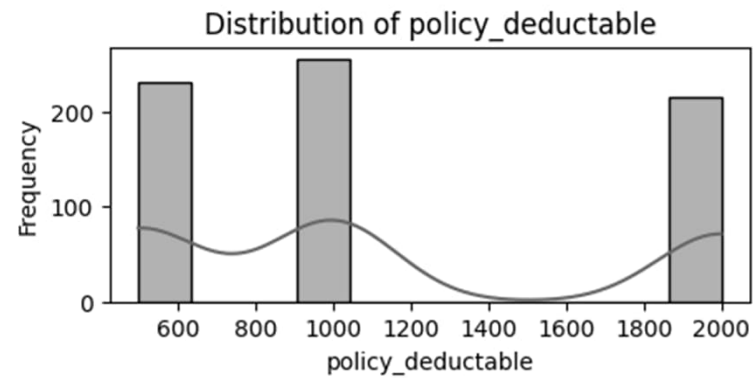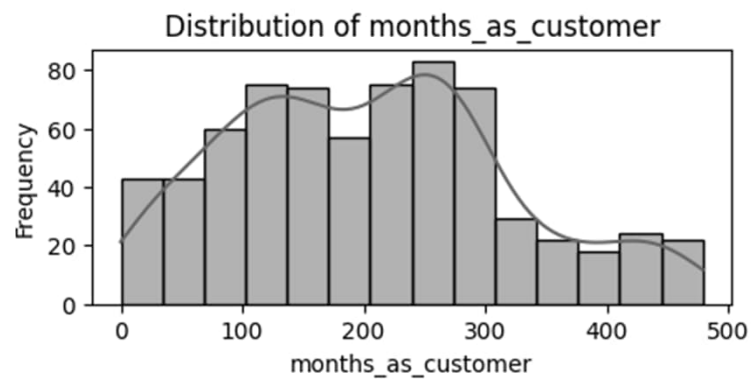
Methodologies..

3. Machine Learning Models:

Supervised Learning-

- Logistic regression, random forest

Unsupervised Learning –

- Isolation forest, autoencoders, clustering methods (k-means etc.) to identify outliers

Hybrid Models-

- Combine supervised and unsupervised methods

4. Evaluation Metrics:

- Focus on high impact features- Claim Amount Ratios, Time-Based patterns, Geographical risk factors
- Handle Data imbalances- Oversampling, under sampling, class weighting
- Clean and standardize the data- Ensure missing values and inconsistencies are handled, standardize categorical variables
- Deployment and real-time monitoring- Early flagging system,

Outliers Handling
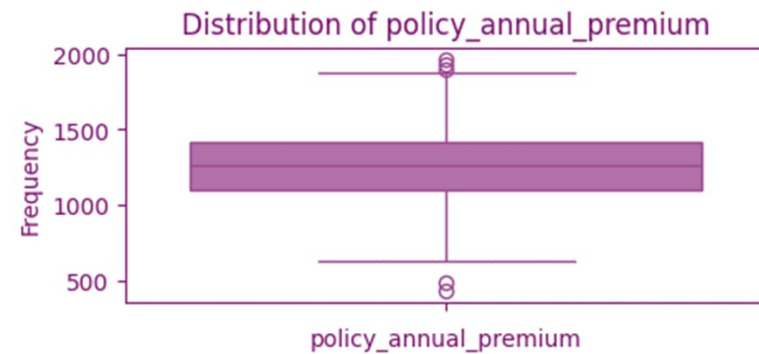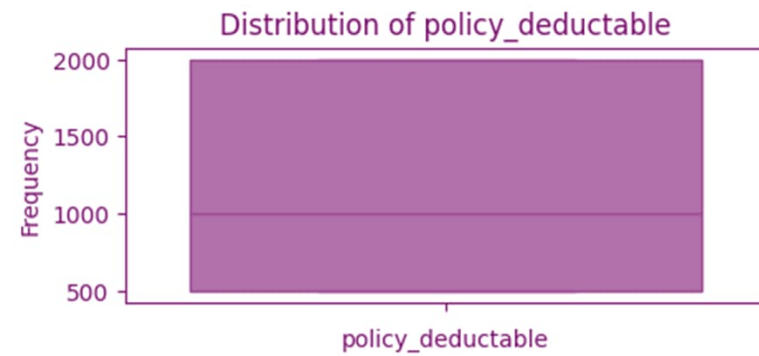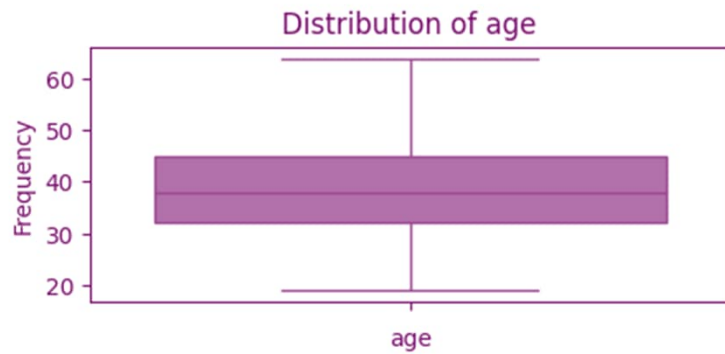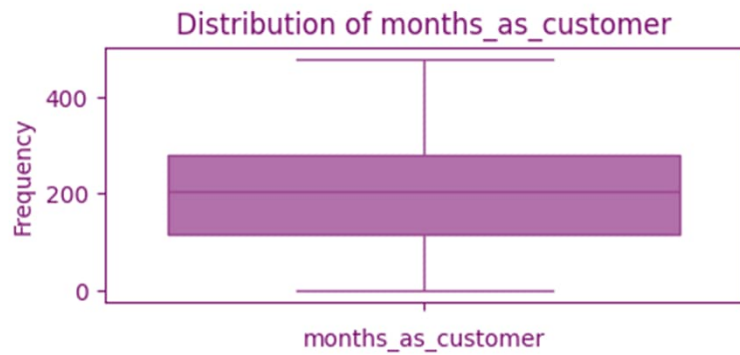
Correlation Analysis

Correlation Matrix of Numerical Features
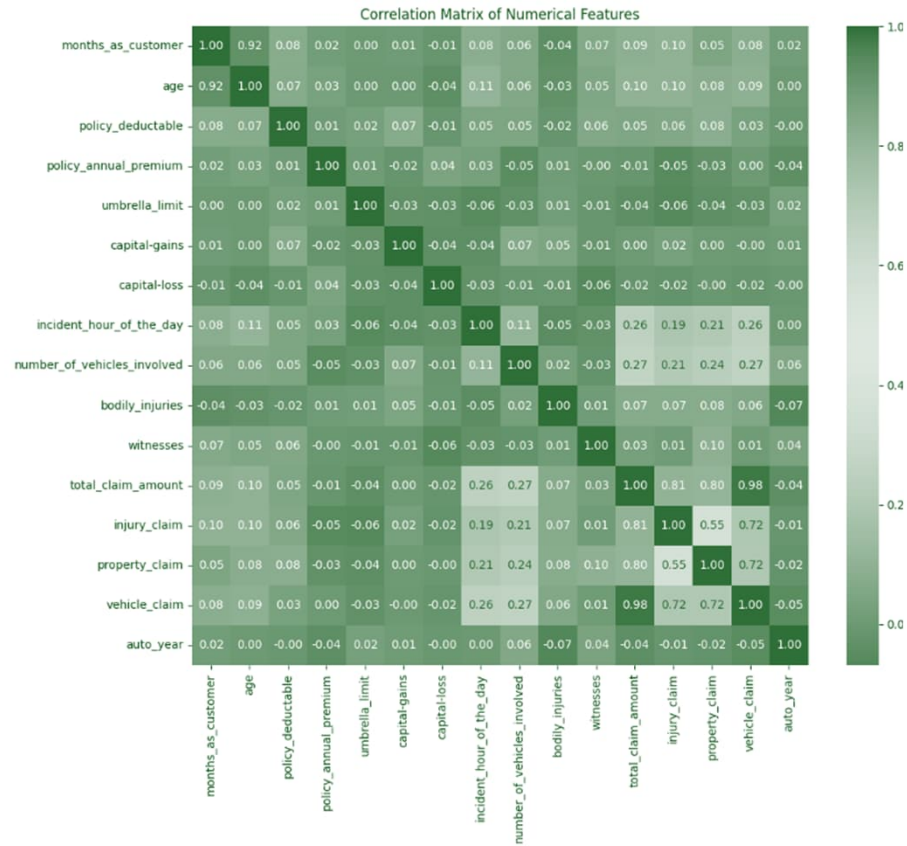
Conclusion

By applying the insights and model created in this project, Global Insurer can greatly enhance its fraud detection abilities. This proactive strategy can substantially reduce financial losses, streamline claim processing, and ultimately improve customer satisfaction. As data continues to expand and change, continuously refining the model through ongoing analysis will guarantee its sustained effectiveness in reducing fraud risk.

Sensitivity: 0.8354430379746836
Specificity: 0.8733031674208145
Precision: 0.7021276595744681
Recall: 0.8354430379746836
F1-score: 0.7630057803468209