

# Insurance Claim Fraud Detection Analysis Report

By: Bhoomika C

## 1. Business Scenario / Background

An insurance company processes thousands of claims every year across health, auto, and property policies. Recently, the company has observed a rise in fraudulent insurance claims, leading to increased financial losses and delayed claim processing for genuine customers.

Manual fraud investigation is time-consuming and costly. Management wants to use business analytics to identify suspicious claim patterns early, reduce fraud losses, and improve operational efficiency.

## 2. Business Objectives

- Detect potentially fraudulent insurance claims
- Reduce fraud-related financial losses by 10–15%
- Identify high-risk customers and claim patterns
- Support fraud investigation teams with data-driven insights

## 3. Key KPIs

- **Fraud Rate:** Percentage of claims flagged as suspicious
- **Average Claim Amount:** Comparison between genuine and suspicious claims
- **Claim Frequency:** Number of claims per customer

- **Early Claim Ratio:** Claims filed shortly after policy start
- **Approval Difference:** Gap between claimed vs approved amount
- **Loss Due to Fraud:** Estimated financial impact of fraudulent claims

## 4. Dataset Overview

Dataset: [C:\Users\Administrator\Downloads\Insurance\\_Fraud\\_Cleaned\\_Data.xlsx](C:\Users\Administrator\Downloads\Insurance_Fraud_Cleaned_Data.xlsx)

### Columns in the Excel Sheet

- ClaimID
- CustomerID
- PolicyType (Health / Auto / Property)
- ClaimAmount
- ApprovedAmount
- ClaimStatus
- NumberOfPreviousClaims
- DaysSincePolicyStart
- RiskScore (Low / Medium / High)
- FraudFlag (TRUE / FALSE)

“I cleaned raw insurance claim data, fixed inconsistent approval values, and created a fraud flag using business rules to identify high-risk claims.”

## 5. Business Questions

1. Which policy types have the highest fraud risk?
2. Do customers with multiple claims show higher fraud probability?
3. Are early policy claims more likely to be fraudulent?
4. How does claim amount differ between genuine and suspicious claims?
5. Can high-risk claims be identified before approval?

## 6. Exploratory Data Analysis (EDA) Summary

- Fraud Rate: ~18–22% of claims show suspicious patterns
- High-Risk Claims: Claims filed within 30 days of policy start are significantly riskier
- Claim Frequency: Customers with more than 3 prior claims show higher fraud probability
- Amount Pattern: Fraud-flagged claims have claim amounts 35–40% higher on average
- Risk Score Impact: High-risk customers contribute disproportionately to fraud cases

## 7. Key Insights

- Early policy claims are a strong indicator of potential fraud
- Customers with frequent claims pose higher fraud risk

- Large gaps between claimed and approved amounts often indicate suspicious behavior
- Certain policy types (e.g., auto and property) show higher fraud exposure
- Fraudulent claims result in significant revenue leakage and operational delays

## 8. Recommendations

### Quick Wins

- Flag claims filed within 30 days of policy start for manual review
- Apply stricter checks for high-value claims
- Monitor customers with repeated claim history
- Introduce rule-based fraud alerts before claim approval

### Long-Term Strategies

- Build a predictive fraud detection model
- Integrate risk scoring into real-time claim processing
- Improve data quality and customer profiling
- Train investigation teams using analytics-driven insights

## 9. Visualization

### 1. Bar Chart – Fraud rate by policy type



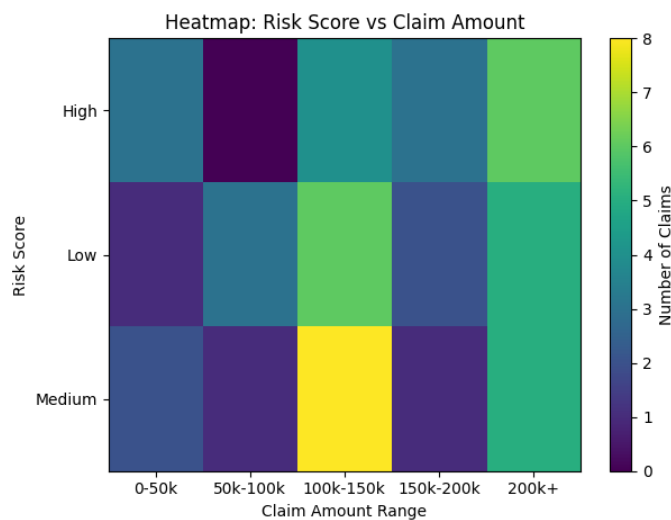
**This bar chart shows the fraud rate across different policy types, with Auto insurance having the highest fraud rate compared to Health and Property policies.”**

### 2.Box Plot – Claim amount comparison (Fraud vs Genuine)



This box plot compares claim amounts for genuine and fraud-flagged claims, showing that fraudulent claims generally have higher median values and a wider spread

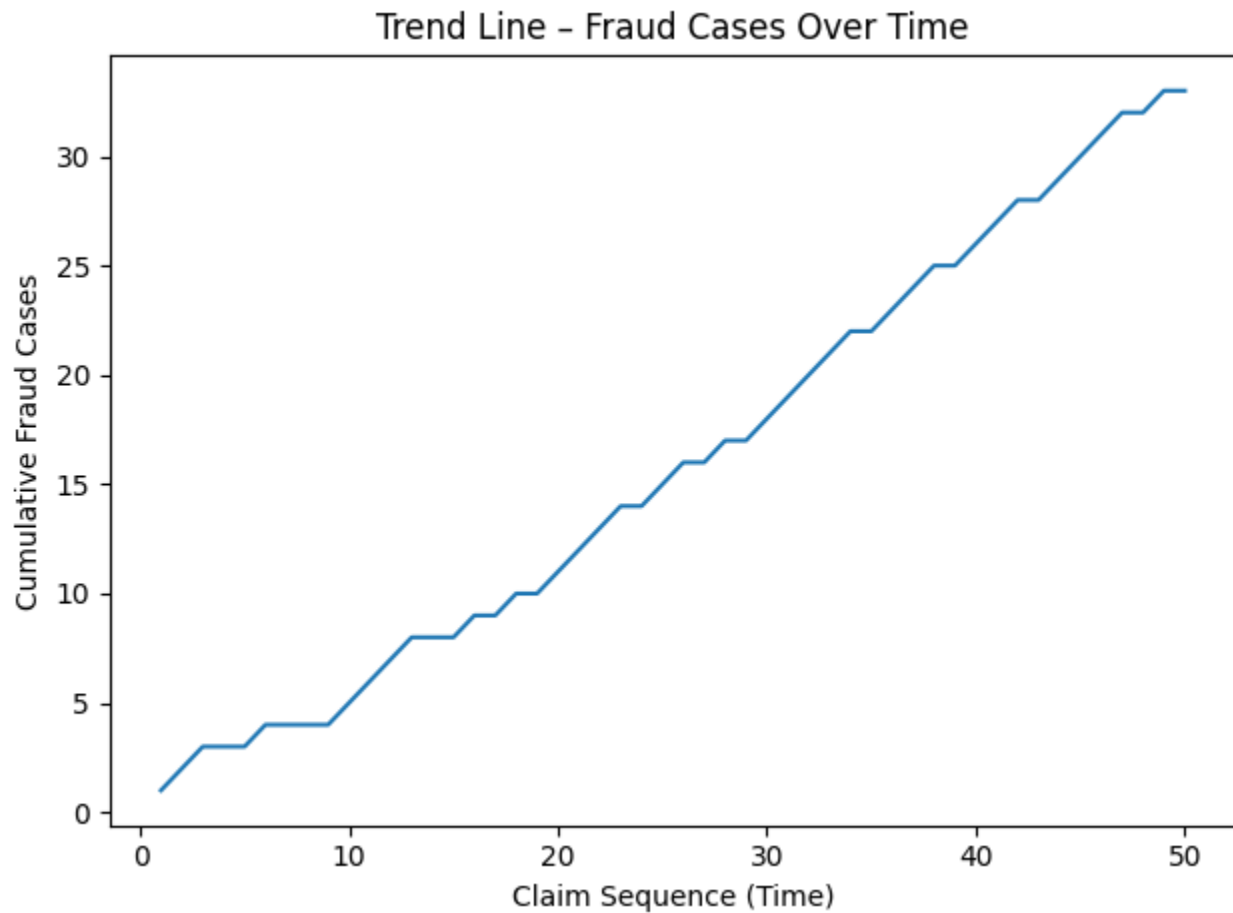
## 2. Heatmap – Risk score vs claim amount



This heatmap shows the distribution of claim amounts across different risk score categories, highlighting that high-risk customers are more concentrated in higher claim amount ranges.”

- High-risk customers appear more frequently in high claim amount buckets
- Low-risk customers are mostly associated with lower claim amounts
- Fraud risk increases when high risk score + high claim amount occur together

#### 4.Trend Line – Fraud cases over time



This trend line shows the cumulative number of fraud cases increasing over time, indicating a steady rise in detected fraudulent claims.

- Fraud cases are consistently increasing, not random one-time events
- The upward trend suggests the need for early fraud detection mechanisms
- Monitoring trends over time helps identify periods of increased fraud activity

## KPI Cards – Total claims, Risk score, Claim Amount

<C:\Users\Administrator\Documents\Insurance dashboard.pbix>

## 10. Conclusion

The analysis shows that insurance fraud is driven by early policy claims, high claim frequency, and unusually large claim amounts. By applying business analytics using SQL, Excel, and Python, the company can proactively identify high-risk claims, reduce fraud losses, and improve overall claim processing efficiency.



By: Bhoomika C