**TITLE:- “AI-Powered Fraud Detection in Auto Insurance: Predictive Modeling for Smarter Claims Management”  
  
(REPORT ON DATA ANALYSIS WITH OUR APPROACH )**

**Objective-**

This project aims to analyze auto insurance claim data to detect and understand fraud patterns using advanced data visualization, feature engineering, and predictive modeling. The goal is to help insurers **identify suspicious claims early**, improve the **efficiency of claim investigations**, and **reduce financial losses** from fraudulent activities.

## ****Business Problem****

## Insurance fraud is a growing concern, leading to significant financial losses each year. Traditional manual claim validation processes are not only time-consuming but also prone to human error. There is a pressing need for **automated, data-driven solutions** that can accurately flag high-risk claims and optimize investigation efforts.

## ****Objectives of the Analysis****

* Explore historical insurance claims to uncover fraud-related patterns.
* Identify high-risk customer segments, vehicles, and geographic zones.
* Visualize and track fraudulent behavior over time and across regions.
* Use machine learning to develop a reliable fraud detection model.
* Support stakeholders with clear, actionable insights and a powerful dashboard.

## ****Dashboard Overview & Visual Insights****

|  |  |
| --- | --- |
| **Visualization** | **Purpose** |
| **Occupation vs. Fraud Rate** | Detect high-risk professions – e.g., Exec-managerial (36.43%) |
| **Map View (State-wise Fraud)** | Identify fraud hotspots – Ohio leads with 3,705 cases |
| **Fraud Over Time (Line Chart)** | Show trends – Week 6 shows a spike in fraud (1,600 cases) |
| **Auto Model vs. Fraud (TreeMap)** | Identify fraud-prone models – RAM (31.01%) and Jetta (29.16%) |
| **Collision Type × Severity** | High fraud in severe + rear/side collisions |
| **Fraud by Gender (Pie Chart)** | Gender distribution – Female: 51.13%, Male: 48.87% |

## ****Key Insights from Data****

* **Occupation Matters:** Fraud rates are highest in “Exec-managerial” and “Farming-fishing” professions.
* **Auto Models Impact Risk:** RAM and Jetta models have the highest fraud incidence.
* **Regional Patterns:** Fraud is concentrated in states like Ohio and Indiana, possibly due to regional fraud rings.
* **Time Factor:** Fraud peaks around **Week 6**, which may suggest tactical filing patterns.
* **Severity and Collision Type Correlation:** Severe accidents combined with rear or side collisions show higher fraud probability.
* **Suspicious Claims:** Many fraudulent claims are **just below ₹20,000**, possibly to evade stricter checks.
* **Data Gaps:** Over 13,000 missing police reports indicate a data quality issue that may be **exploited by fraudsters**.

## ****Key KPIs Summary****

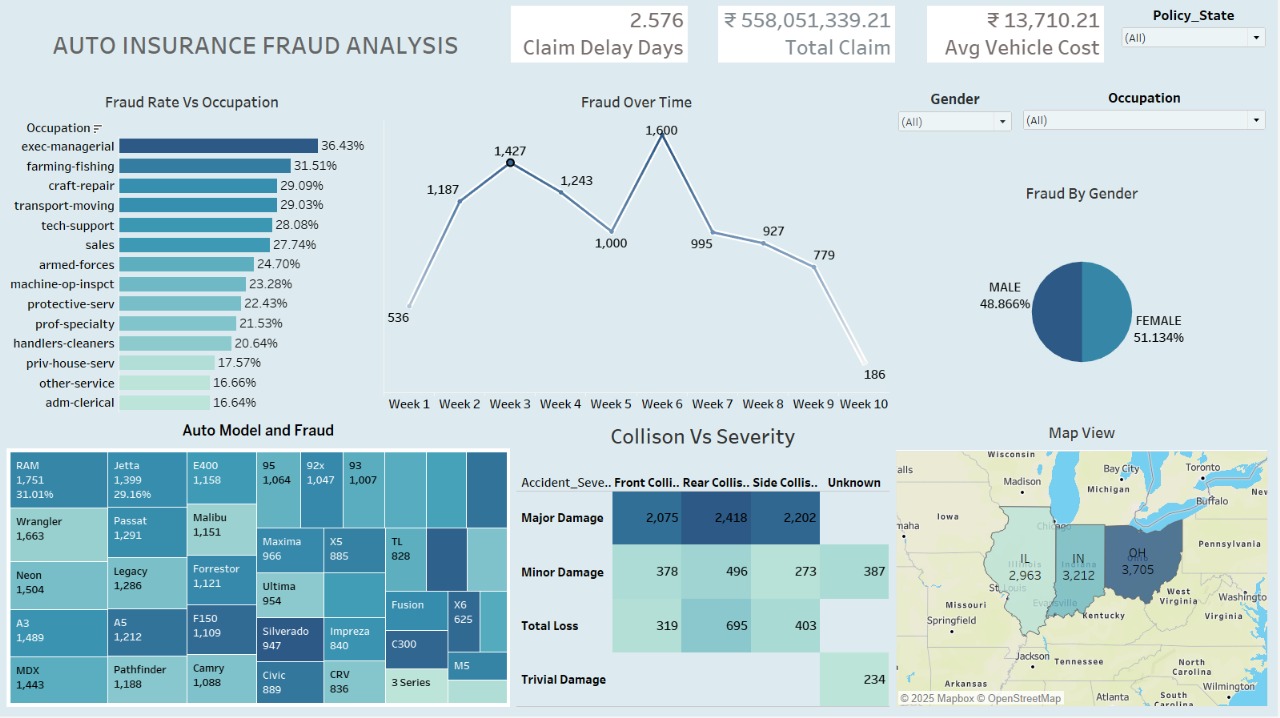
|  |  |  |
| --- | --- | --- |
| **Average Claim Delay** | ~4.5 days | Faster claims may indicate premeditated fraud |
| **Total Claim Volume** | ₹2.4M+ | High value at stake |
| **Avg. Vehicle Cost** | ₹13,000+ | Mid-cost vehicles dominate fraud cases |

## ****Recommendations for Stakeholders****

* Using the insights from this project and dashboard, insurance companies can:
* **Proactively flag suspicious claims** based on factors like model, occupation, and amount.
* **Prioritize investigations** by severity, accident type, or high-risk segments.
* **Refine AI/ML fraud detection models** using engineered features and visualized insights.
* **Enforce policy changes**, e.g., mandatory police report upload, stricter checks on high-risk segments.

## ****Final Outcome****

* The result is a **fully interactive Tableau dashboard** and a **clean, feature-rich machine learning-ready dataset's**, giving insurers:
* Clear visibility into fraud patterns.
* Granular filtering (e.g., by gender, state, occupation).
* Strong foundation for fraud scoring and predictive modeling.
* A powerful tool to **reduce fraud losses and speed up investigations**.



**OUR ANALYSIS**