**Fraud Detection in Auto Insurance Claims – Risk Analysis**

**Name: Atul Sambhaji Shinde**

**Course - Business Analytics (24304G434)**

**Institute: N. L. Dalmia Institute of Management Studies and Research**

**Company Name: Edme Insurance Brokers Ltd.**

**Introduction**

**Problem Statement:**  
Insurance companies are facing financial losses due to rise in fraudulent insurance claims which is affecting both companies and genuine customers.

**Objective: This** analysis aims to

* **Identify fraudulent patterns from previous data.**
* **Detect high-risk segments.**
* **Improve fraud detection strategies.**
* **Support decision-making**

**Tools Used:**

* **Data Cleaning:** Excel, Python (handling missing values, outliers, categorical inconsistencies).
* **Dashboard Development:** Power BI (creating dashboard, visualizing fraudulent trends & insights).

**Unique Claims & Data Overview**

* **Total Claims Analyzed:** 990 (After cleaning)
* **Fraudulent vs. Non-Fraudulent Cases:**
  + **Fraudulent:** *244 cases (24.65% of total claims)*
  + **Non-Fraudulent:** *746 cases (75.35% of total claims*

**Age groups made for convenience:**

* + 18-25**: Young**
  + 26-35**: Adult**
  + 36-45**: Middle Age**
  + 46-55**: Senior**
  + 56-65**: old**

**Fraud by Age & Gender**

* Fraud is the highest among **Middle-aged & Adult categories.**
* **Fraud by Gender:**
  + **53.64**% of fraud cases involve females**.**
  + **46.36**% involve males.

**Segment with Highest Fraud:**

* Education level: *JD*
* Occupation: *Machine-Op-Inspect*
* Hobby: *Reading*

*Missing values:*

* Yes, missing values were found in the column’s **collision type**, **property damage**, **police report available** which were replaced by “unknown” for maintaining data integrity.

*Inconsistencies:*

* Yes, inconsistencies in categorical variables were found in the columns **insured Education Levels, incident type, insured occupation** **which** were standardized for making the data into consistent format.

**Outliers**

Yes, outliers were found in **total claim amount & policy annual premium**  
columns. To prevent data from skewing extreme outliers were removed.

**Insights**

1. **Is There a Relationship Between Policy Deductible & Fraud?**

The fraud rate is mostly consistent with policy deductibles so we can say that it does not influence the fraud likelihood.

1. **Do More Bodily Injuries Correlate with Fraud?**

From the dataset we can say that fraud rate is consistent among bodily injuries hence it will not affect fraud likelihood.

1. **Are Fraud Cases More Likely at Specific Hours?**

Late-night accidents (10 PM - 4 AM) **are more suspicious** for fraud investigations.

1. **Which Incident Types & Severities Have the Most Fraud?**

From the dataset it is clearly inferred that Major Damage & Total Loss cases have more fraud than minor damage cases. Similarly, multi vehicle collisions and single vehicle collisions have more fraud rates than parked /theft car.

Fraudsters may increase damage to cars to maximize their claims amount.

1. **Does Vehicle Year Impact Fraud?**

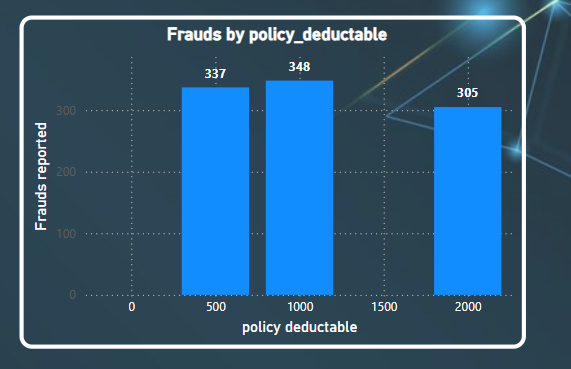
From the dataset we can say that older vehicles (1995-2005 models) report more fraud.Fraudsters targeting older cars for fraudulent claims.

1. **Other Factors: Age, Vehicles, & Claim Amount**

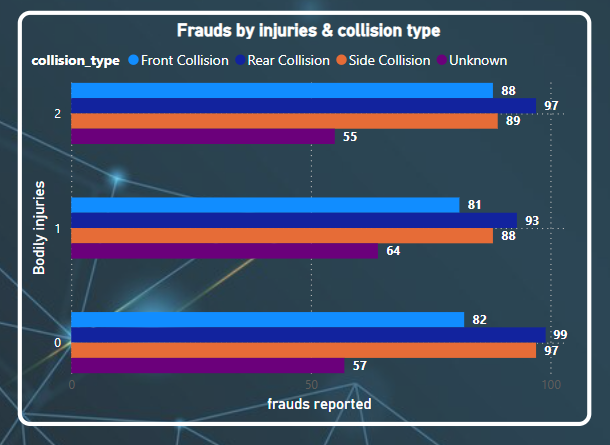
More fraud occurs in cases with multiple vehicles involved.

Higher claim amounts tend to have more fraud cases.

Multi-vehicle claims may be made to increase claim amounts.



The fraud rate is mostly consistent with policy deductibles so we can say that it does not influence the fraud likelihood.



From the dataset, it is clearly inferred that front and rear collisions have higher fraud cases compared to side collisions and unknown collision types. Similarly, cases with bodily injuries (1 or 2) have higher fraud rates than those with no injuries.

A screenshot of a graph

Description automatically generated

From the dataset it is clearly inferred that Major Damage & Total Loss cases have more fraud than minor damage cases. Similarly, multi vehicle collisions and single vehicle collisions have more fraud rates than parked /theft car.

A graph of blue bars

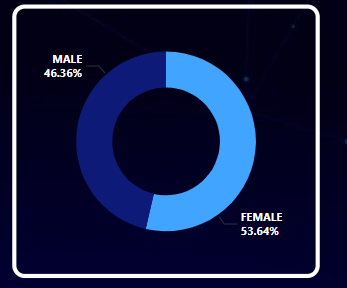
Description automatically generated with medium confidence

From the dataset it is clearly inferred that Middle age & Adults are more involved in fraud cases as compared to senior, young and old ones . Similarly, females are more likely to be involved in fraudulent cases as compared to males. (Age groups are mentioned above)

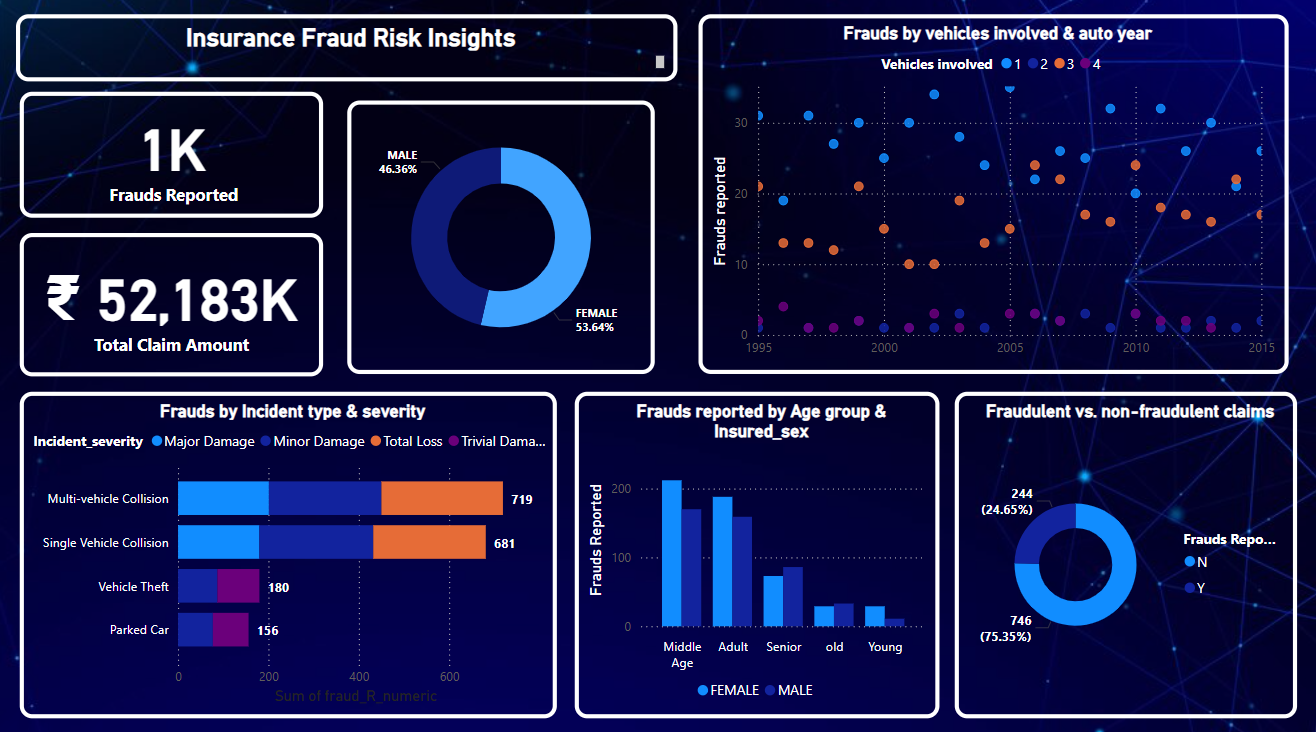
A graph of frauds and auto year

Description automatically generated

From this scatter plot it can be seen that across all years most of the frauds are reported in which single vehicle is involved while fraud cases are lesser in cases where four vehicles are involved .



From this donut chart it is indicated that females (53.64) account for more number of fraud cases as compared to males (46.36%),.Due to which it can be inferred that Insurance companies have to be more specific while dealing with female insurers.



A screenshot of a computer screen

Description automatically generated

