Car Insurance Fraud Analysis

Presented by Noursine Amira, Rahma Haddouchi, Maycem Azaza , Lina Brahmi

1

Business Context

Sector : Property and Casualty (P&C) Insurance.

This project focuses on developing a Business Intelligence (BI) solution to analyze car insurance claims, detect fraud patterns, and understand cost drivers in the Property & Casualty insurance sector.

Project Goals

- Identify fraud patterns and high-risk segments
- Monitor KPIs such as Fraud Rate and Loss Ratio
- Optimize claims approval and settlement time
- Support data-driven decision-making

Business needs

- Undetected insurance fraud increases costs
- Inefficient manual investigation of low-risk claims
- Lack of centralized visibility over claims, drivers, and incidents

Key Research Questions

Understanding Fraud Rates and Influencing Factors

1. What is the overall fraud rate?

2. Does driver age impact fraud probability?

3. Are some vehicle categories more fraud-prone?

4. Do past offenders re-offend?

5. Does witness or police presence reduce fraud?

Process Summary

The project follows an end-to-end BI workflow

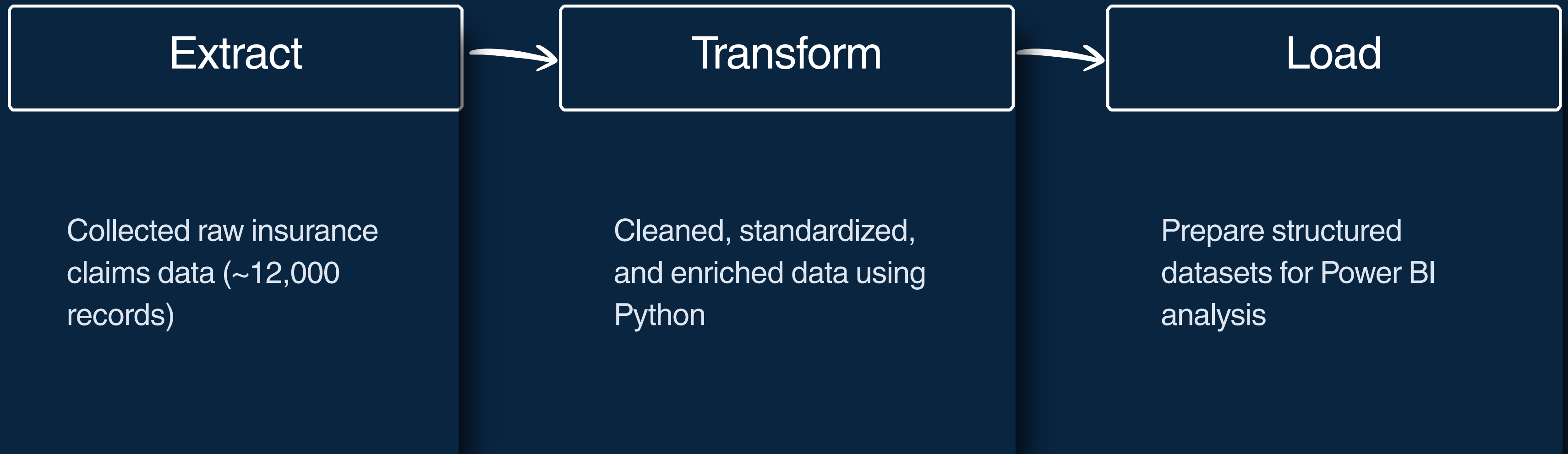Using Python for data preparation and Power BI for dashboards.

Data Gathering

ETL

Data Warehouse

Analysis

Visualization

ETL PROCESS

Extract	Transform	Load
Collected raw insurance claims data (~12,000 records)	Cleaned, standardized, and enriched data using Python	Prepare structured datasets for Power BI analysis

Data Preparation

Handeling missing values

```python
# Replace '?' with NaN and impute
df.replace('?', np.nan, inplace=True)
df['age_of_driver'].fillna(df['age_of_driver'].median(), inplace=True)
df['gender'].fillna(df['gender'].mode()[0], inplace=True)
```

Standardizing categorical fields

```python
# Standardizing Gender text
df['gender'] = df['gender'].replace({'M': 'Male', 'F': 'Female'})

# Converting Fraud Reported to Binary (1/0)
df['fraud_reported'] = df['fraud_reported'].apply(lambda x: 1 if x == '
    Y' else 0)
```
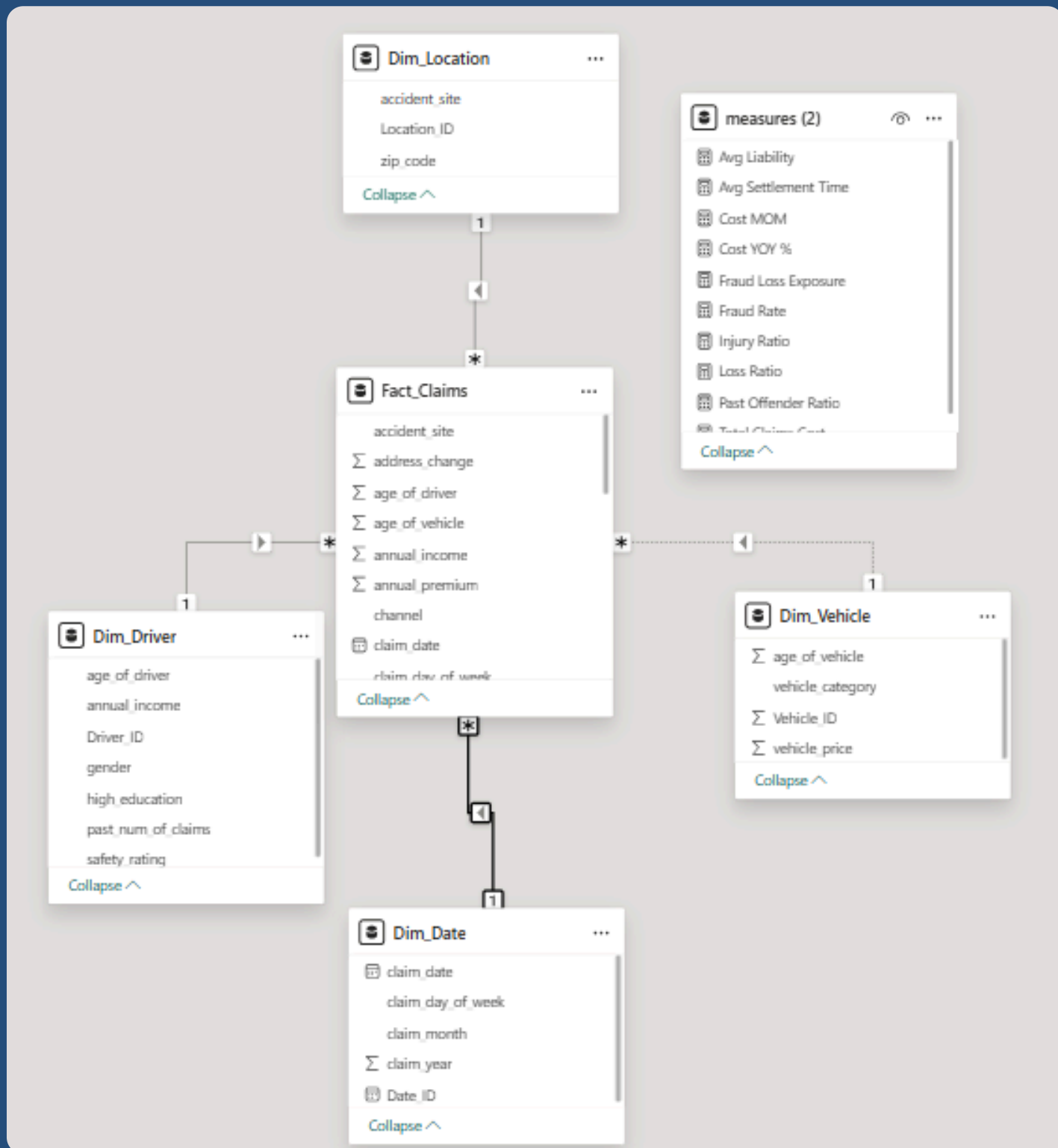
Time Series Preparation

```python
# Extracting Year and Month
df['claim_date'] = pd.to_datetime(df['claim_date'])
df['claim_year'] = df['claim_date'].dt.year
df['claim_month'] = df['claim_date'].dt.month_name()
```

Data Modeling & Storage Design

Star Schema design for analytical efficiency, optimized for Power BI performance

Identification of Fact and Dimension Tables

1.Fact Table:
 The central table of your model, it contains the quantitative data (Measures) such as : Date_ID , Driver_ID, Location_ID ,Vehicle_ID

2.Dimension Tables:
  a. Location : Stores the geographical details of where the accidents occurred or where the policy is based.
  b. Date : A lookup table used for time-intelligence analysis. It allows you to break down claim data by year, month, or day of the week.
 c. Driver: Contains demographic information about the policyholders.
  d. Vehicle: Provides specifics about the insured vehicles

Data Visualization


Executive Summary Dashboard
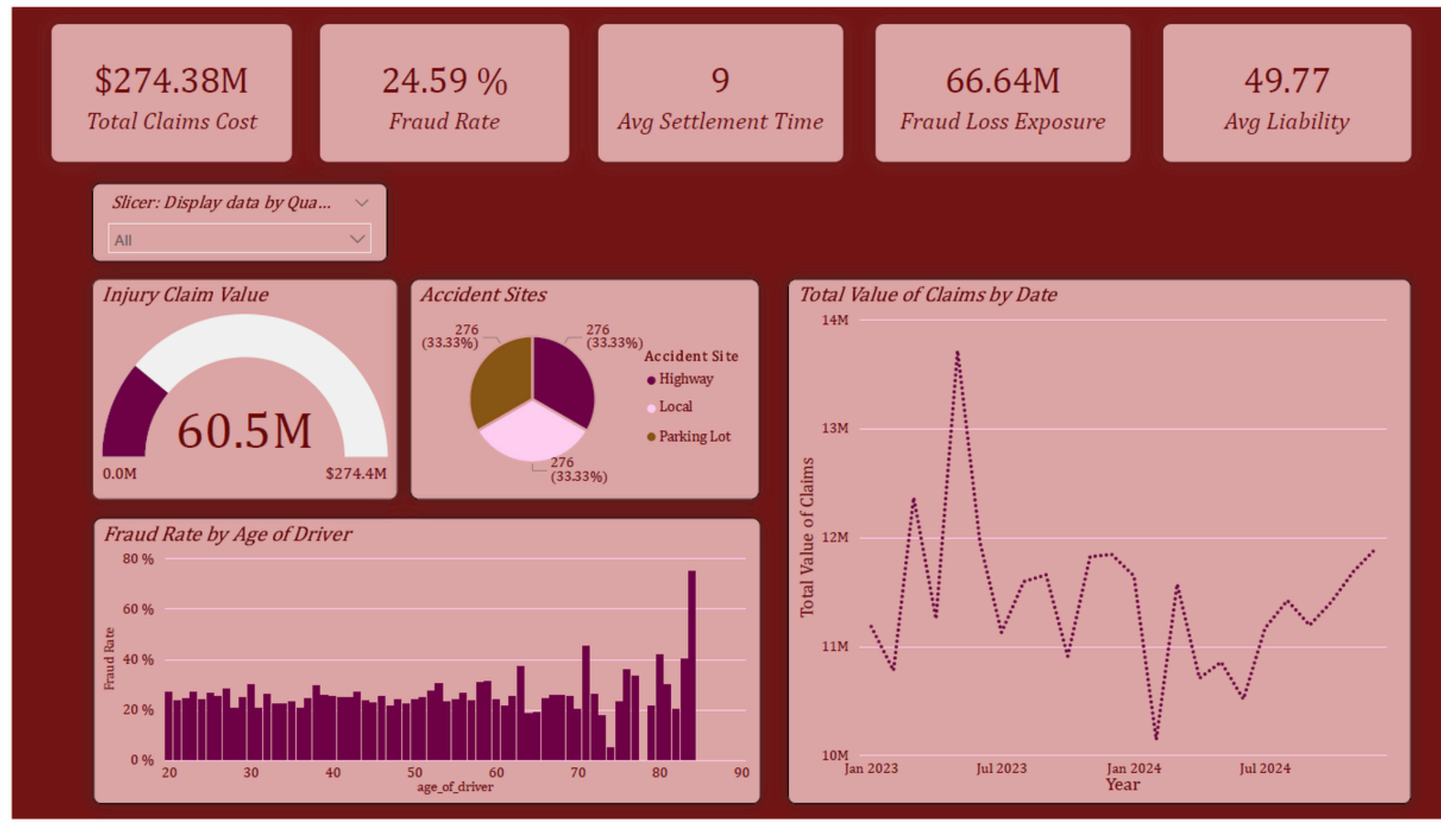
This dashboard provides a high-level view of the insurance portfolio's performance, highlighting key KPIs such as Total Claims Cost, Fraud Rate, Average Settlement Time, and Fraud Loss Exposure.

Deep Dive Analysis : General Overview
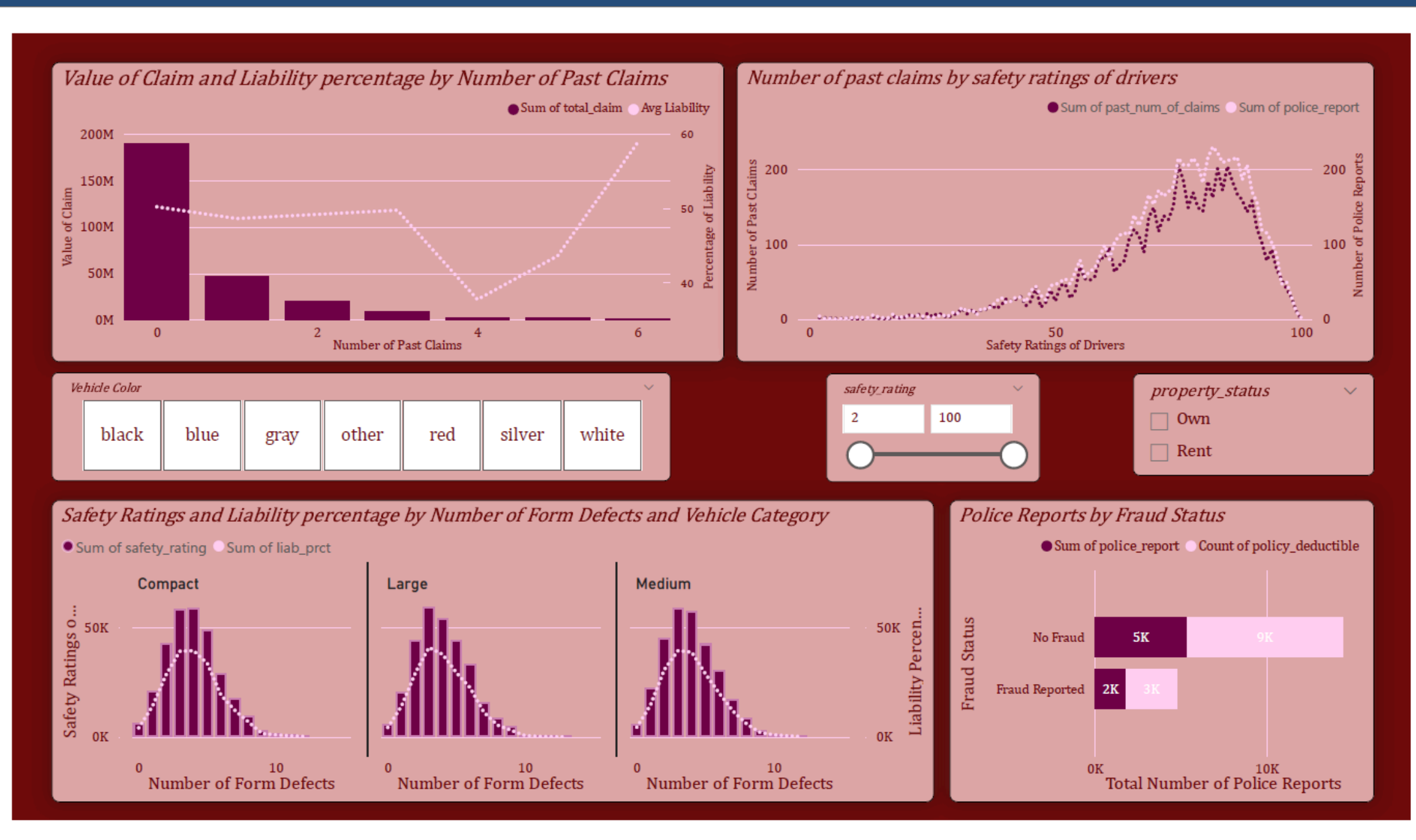
- Recidivism & Liability: Impact of claim history on costs
- Safety & Police Interaction: Link between safety ratings and police reports
- Vehicle Category: Safety and liability by vehicle class
- Fraud Verification: Police reports in fraud vs. legitimate claims

Deep Dive Analysis : Filtered Scenario
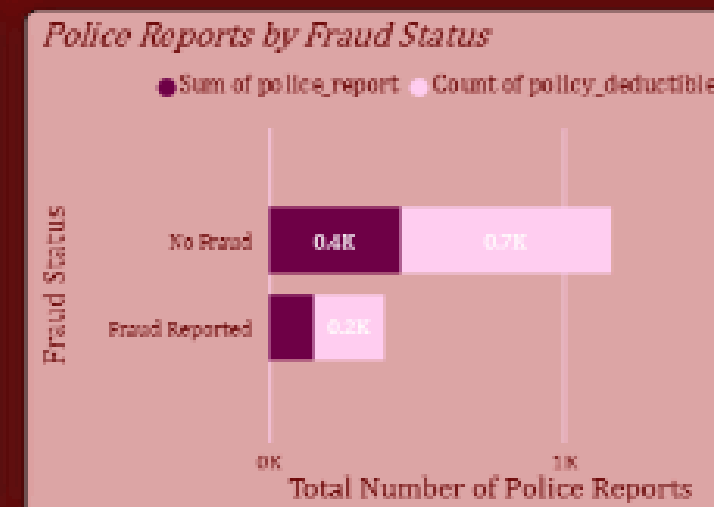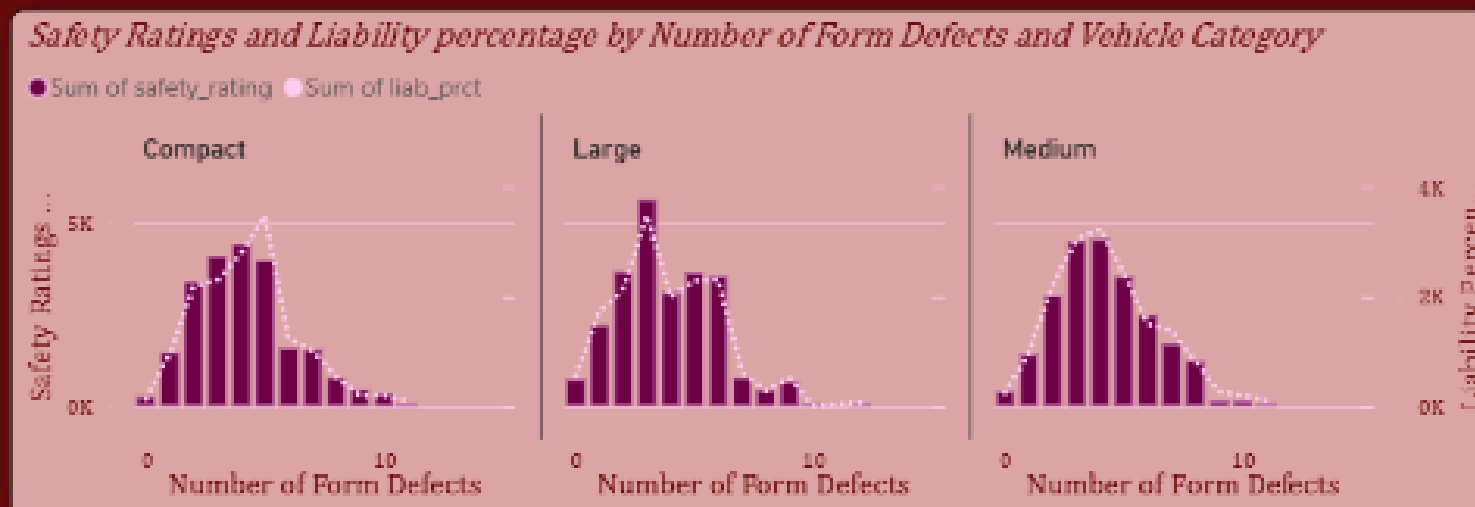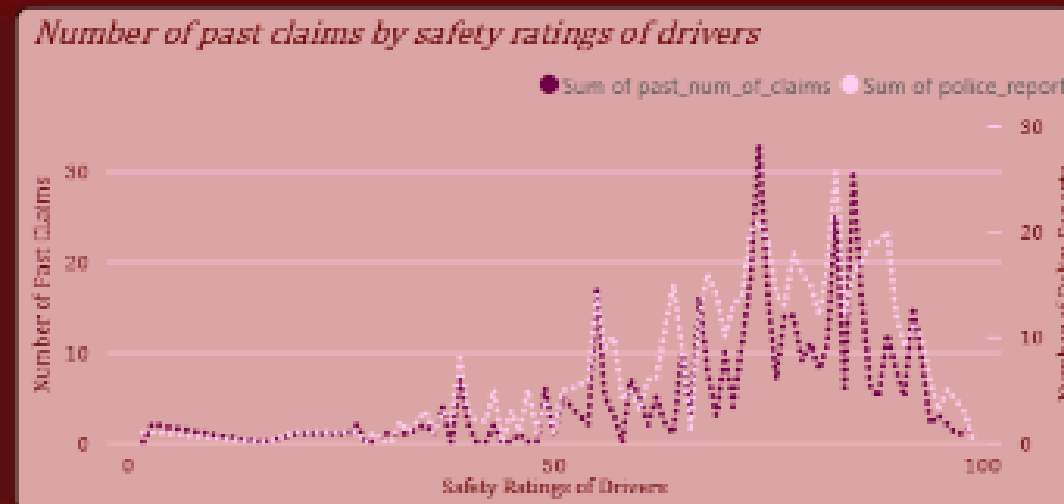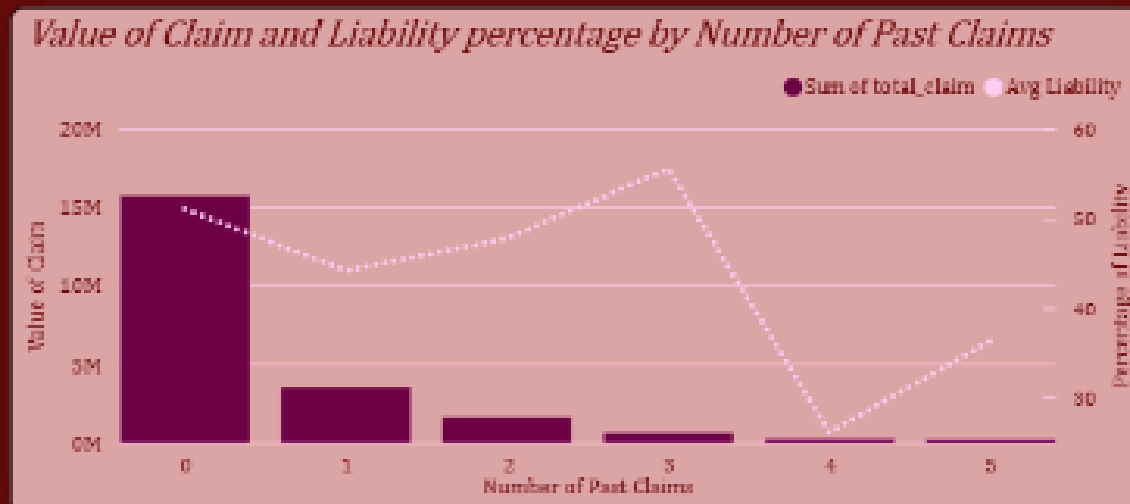
We isolated drivers who Own their vehicles and drive Red cars. This interaction instantly updates all metrics, revealing how specific demographics contribute to the overall risk profile.

Total Value of Claims by Location and Fraud Status

Operational & Claim Details
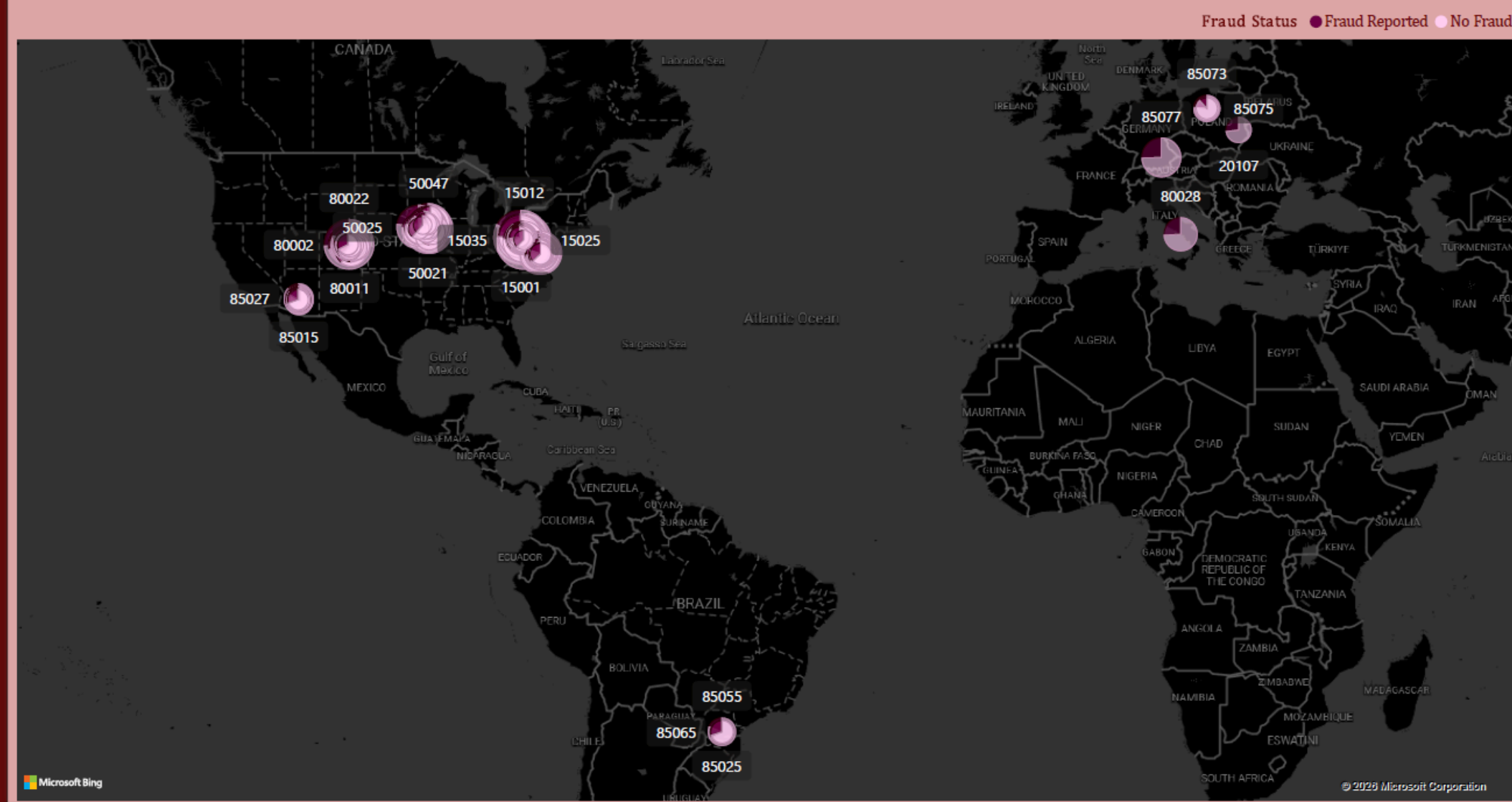
Using Azure Maps, this dashboard visualizes global and regional claim concentrations while using color-coded pie charts to instantly identify fraud hotspots across the US, Europe, South America, and East Asia.

Key Findings

Higher fraud risk among young drivers (18–30)

Luxury and Utility vehicles show higher fraud rates

Around 25% of claims flagged as fraudulent

Unwitnessed accidents are more likely fraudulent

Business Decisions

Automated Flagging → Implement a rule in the claims system to auto-flag any claim where Past Num of Claims > 0 for immediate review.

Premium Adjustments → Adjust risk premiums for the "Utility" vehicle category to reflect the higher fraud exposure.

SIU Prioritization → Direct Special Investigation Unit (SIU) resources specifically toward unwitnessed accidents involving high injury claims.

16

Key Findings

Higher fraud risk among young drivers (18–30)

Luxury and Utility vehicles show higher fraud rates

Around 25% of claims flagged as fraudulent

Unwitnessed accidents are more likely fraudulent

Conclusion

This BI project demonstrates how data-driven insights can reduce fraud exposure, improve operational efficiency, and support strategic decision-making in the insurance sector.

Thank You