Prepared by Nusreena chorath

Project

WARRANTY CLAIMS FRAUD PREDICTION

30 May, 2025

Introduction

● ● ● ● ●

This project aims to develop a machine learning model that can accurately classify warranty claims as either fraudulent or non-fraudulent, helping the company proactively detect suspicious claims and reduce financial loss.

● ● ● ● ●

🔍 PROBLEM STATEMENT

Warranty fraud is a significant concern in the automotive and manufacturing industries, leading to financial losses and reputational damage. Companies often face difficulty distinguishing between legitimate and fraudulent claims due to the complexity and volume of data.

key features and a few initial insights

Product Issues

- Which issue types (AC_100X or TV_200X) are most linked to fraud?
- Does a particular product category show higher fraud rates?

Source Analysis

- Are claims from Dealers more fraudulent than Manufacturers?

Geographic Distribution

- Which states or regions report more fraud?
- Where are high-claim values concentrated?

Customer Behavior
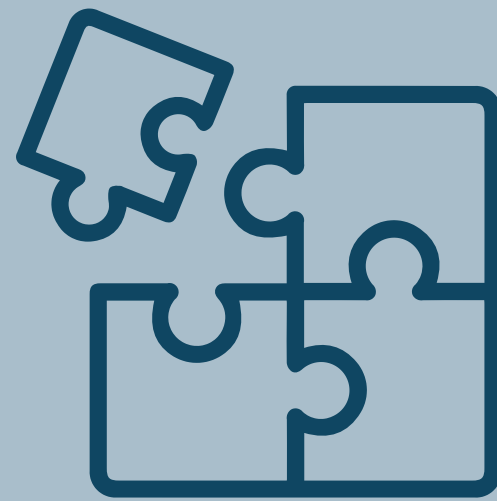
- Do Business profiles generate more fraudulent complaints?

Project Objectives

Analysis Phase

- Analyze the warranty claims based on their region, product, claim value and other features

Strategy Development

- This project holds the potential to assist in identifying and preventing warranty claims fraud,

Implementation Plan

- saving resources and enhancing the efficiency of warranty claim processing.

Methodology

Data Collection:

- The aim of this project is to analyze the warranty claims based on their region, product, claim value and other features to predict their authenticity. The dataset is taken from Kaggle. The dataset contains 358 rows and 21 columns.

DATASET OVERVIEW

- The dataset includes various features related to warranty claims, such as:
- Claim Amount
- Number of Repairs
- Time Since Last Service
- Vehicle Age
- Vehicle Type
- Claim History
- Service Center ID
- Fraudulent (Target variable – 0: Non-Fraud, 1: Fraud)

🧹 DATA PREPROCESSING

- Missing Values: Handled using mean/mode imputation.

- Outliers:Detected using IQR and treated appropriately.

- Categorical Encoding:Used Label Encoder and OneHot Encoder

- Feature Scaling: Applied StandardScaler to normalize numeric features.

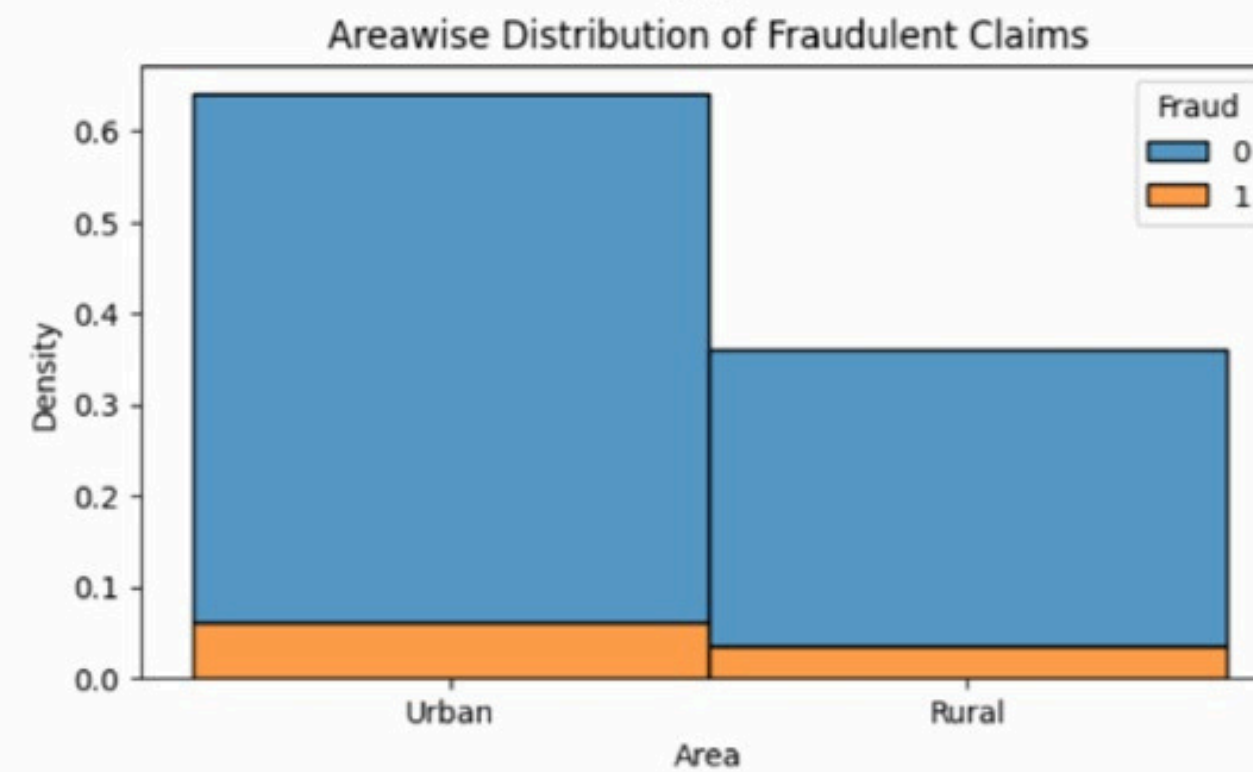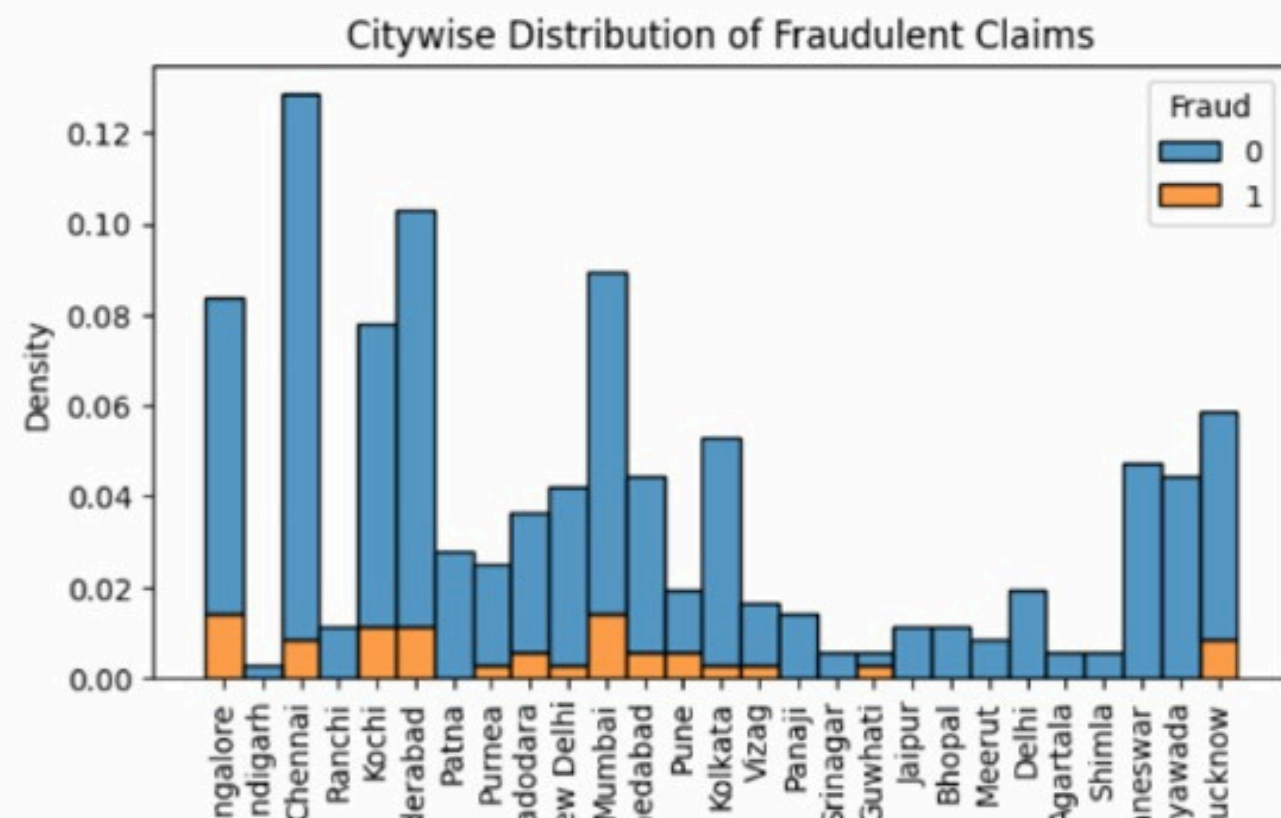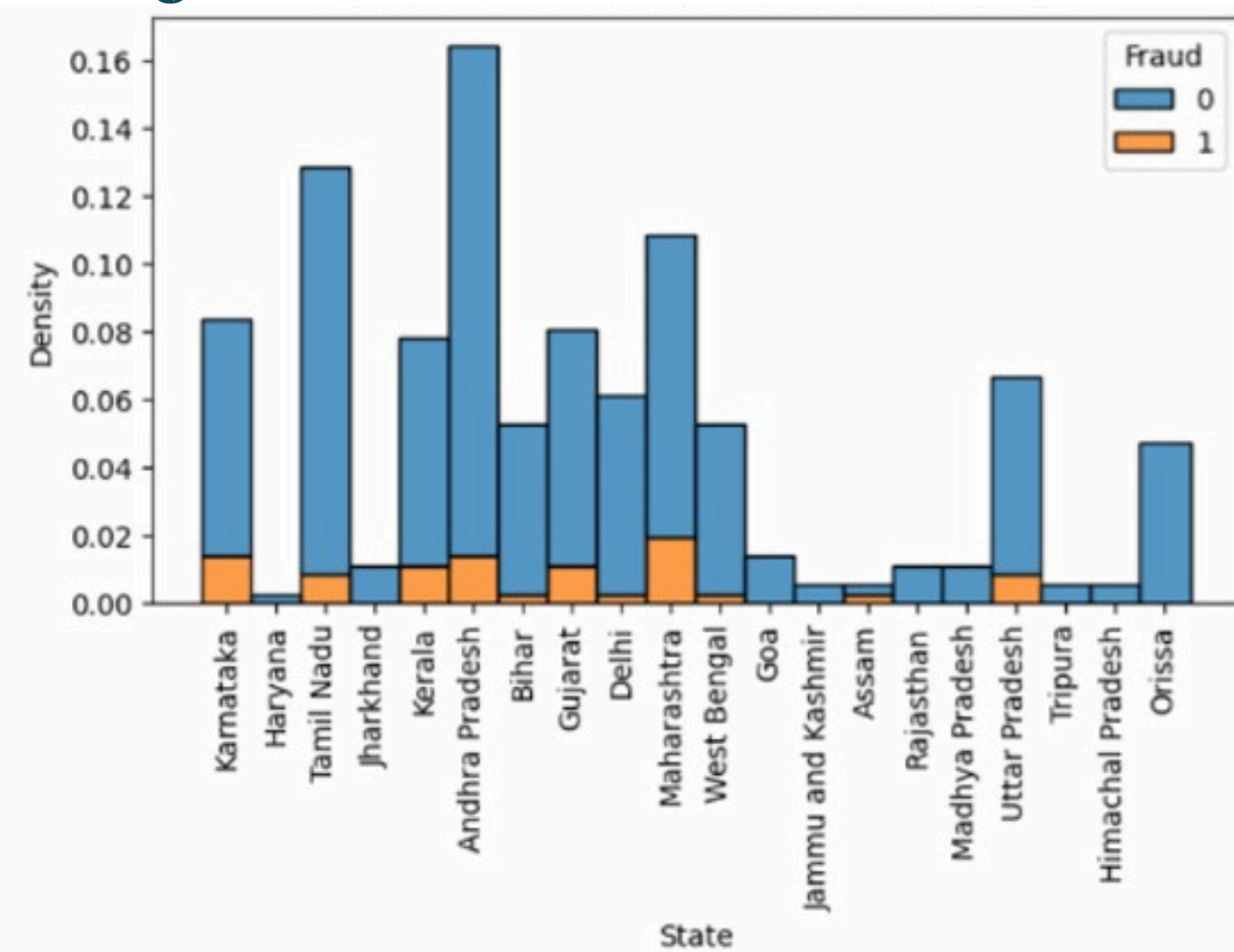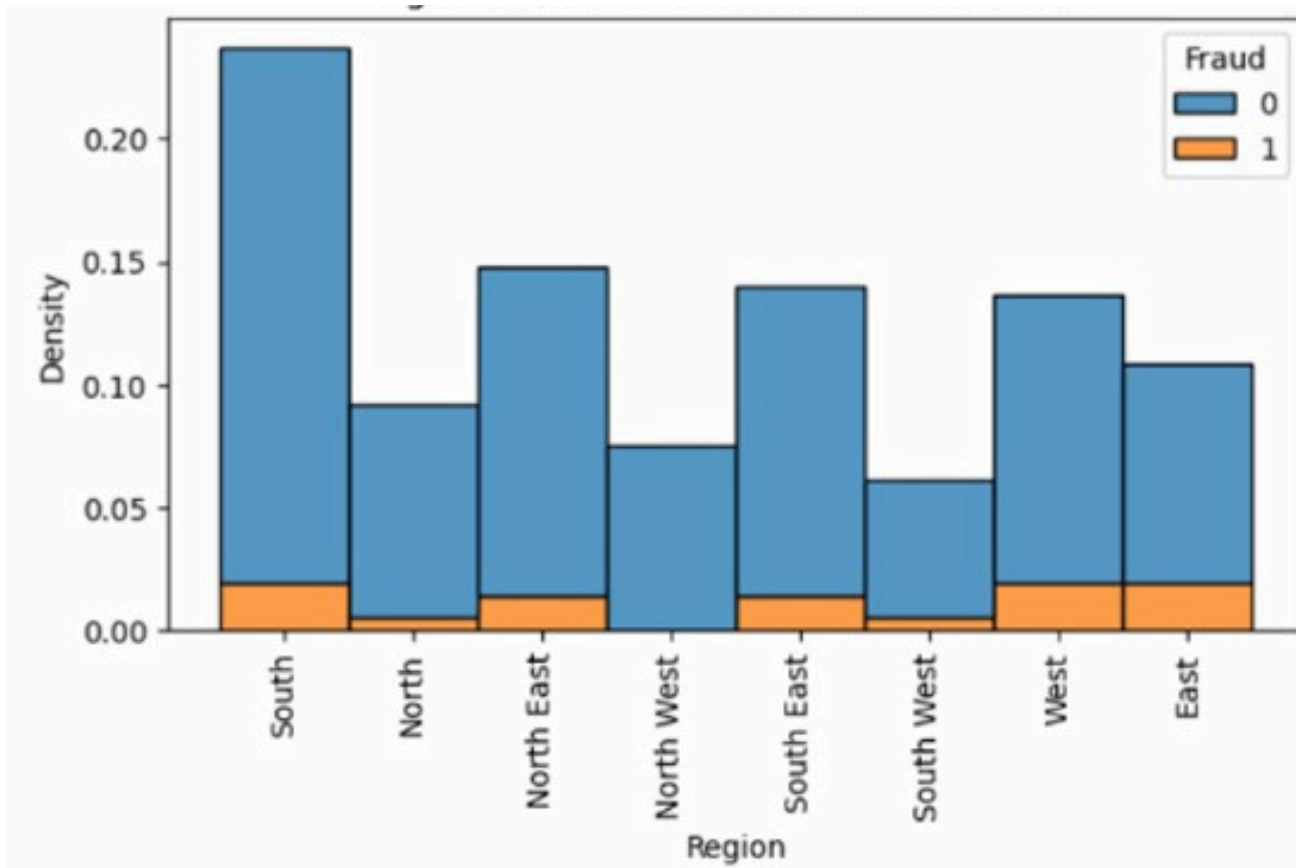- Train-Test Split: Data was split into 80% training and 20% testing.

EXPLORATORY DATA ANALYSIS (EDA)

- Imbalance observed in the target variable (Fraudulent).

- Claim amount and number of previous repairs are higher in fraudulent cases.

- Vehicle type and age also play a role in claim legitimacy.

Location based Distribution of Fraudulent Claims

Regional Distribution:

 South, North East, and South East regions have the highest number of warranty claims overal. West, East, and South regions report the highest number of fraudulent claims. The North West region has zero fraudulent claims.

State-wise Distribution:

States with highest warranty claims: Andhra Pradesh, Maharashtra, Tamil Nadu, Karnataka, and Gujarat. Haryana has the lowest number of warranty claims. States with highest fraudulent claims: Andhra Pradesh, Maharashtra, Tamil Nadu, Karnataka, and Gujarat. States with lowest fraudulent claims: Bihar, Delhi, West Bengal, and Assam.

City-wise Distribution:

Cities with highest claims: Chennai, Hyderabad, Bangalore, Mumbai, and Kochi. Cities with lowest claims: Chandigarh, Srinagar, Agartala, and Shimla. Cities with highest fraudulent claims: Chennai, Hyderabad, Bangalore, Mumbai, and Kochi. Cities with zero fraudulent claims: Chandigarh, Panaji, Meerut, Jaipur, and several others.

Urban vs Rural Areas:

Urban areas show a significantly higher number of claims. As a result, urban areas also contribute more to fraudulent claims compared to rural areas.

Consumer Profile and Fraudulent Claims

From this graph, it is clear that majority of the claims are from consumer who purchased the products for personal use. However,the consumers who purchases the products for business purpose have higher number of fraudulent warranty claims.

Consumer Profile distribution
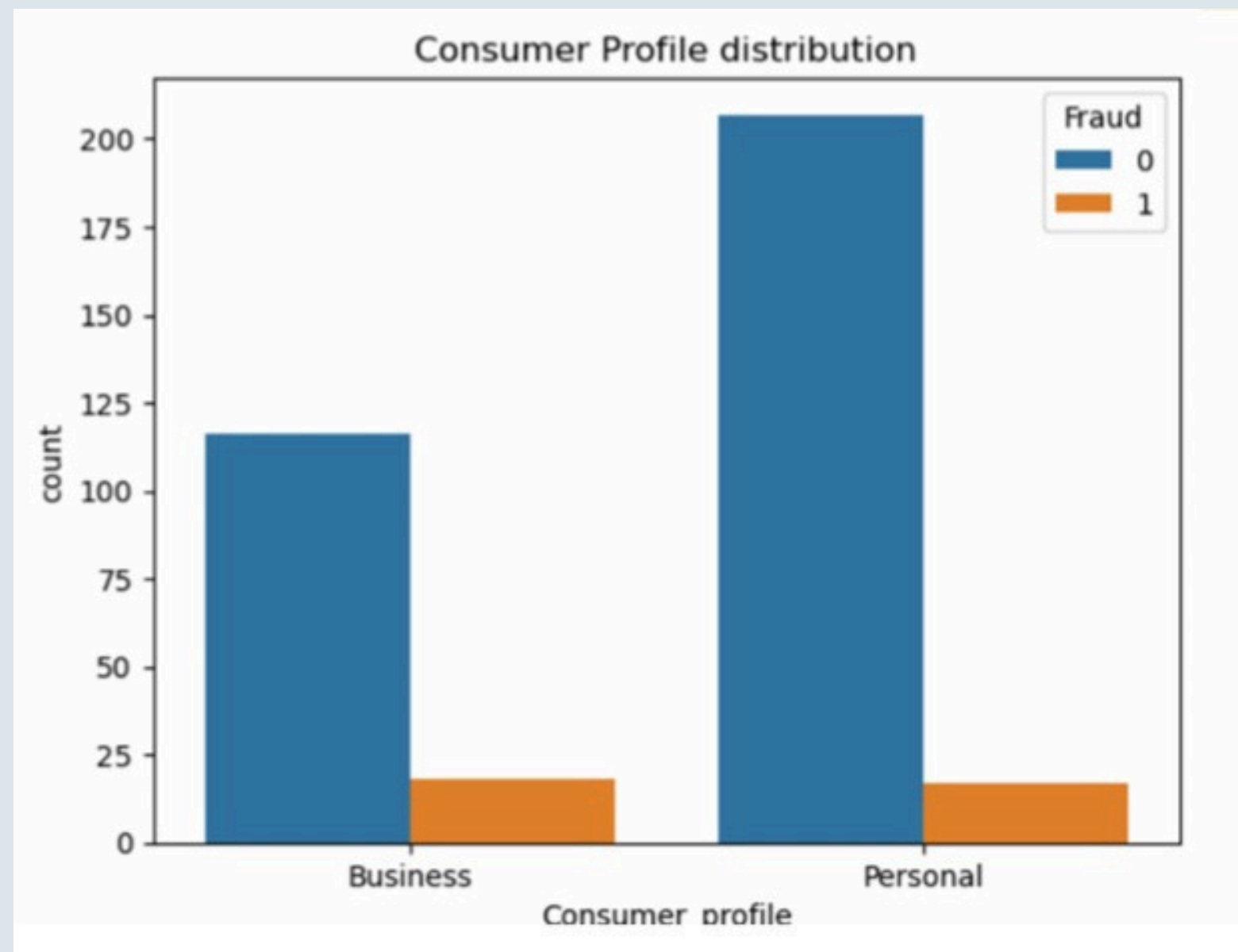
Product and Fraudulent Claims

This graph shows that the company has higher sales for the TV as compared to the AC, and ultimately the number warranty claims for TV is higher than AC. Moreover, the number of fraudulent claims for TV is also higher than AC.

Product and Fraud Distribution
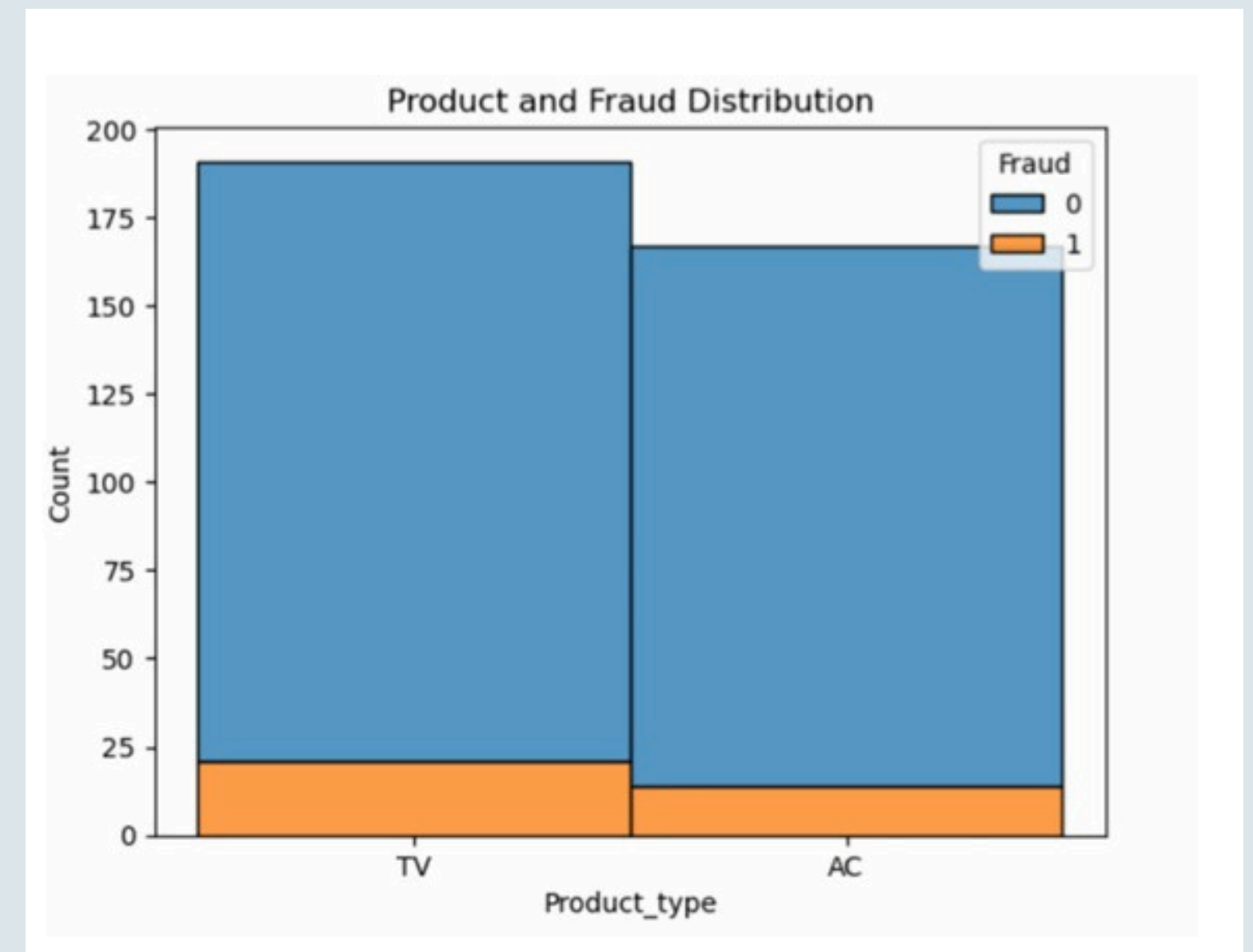
Conclusion

- Warranty claims are most frequent in the southern region of India, particularly in Andhra Pradesh and Tamil Nadu.
- Fraudulent claims are more common in urban regions such as Hyderabad and Chennai.
- The dataset includes claims for two products: TVs and ACs. TVs have higher warranty claims when purchased for personal use compared to ACs.
- Fraudulent claims for ACs occur even when there are no issues with the AC parts.
- Fraudulent claims for TVs can occur with or without issues in the TV parts.
- Fraudulent claims are more frequent when purchases are made through the manufacturer.
- Fraudulent claims tend to have higher claim values compared to genuine claims.
- Service center 13 had the highest number of fraudulent claims despite having fewer total warranty claims.
- Fraudulent claims are more frequent when the customer care call duration is less than 3-4 minutes.

Recommendations

- 🔄 Regularly update the model with new claim data to continuously improve performance and adapt to evolving fraud patterns.
- 🔍 Monitor flagged claims and refine the probability thresholds based on ongoing results and analyst feedback.
- 🧠 Integrate the model into the company's real-time warranty processing system for proactive fraud detection.
- 🧩 Use SHAP or LIME to ensure interpretability and explainability of model predictions, especially when used in production.
- 🛠️ Conduct a thorough audit of Service Centre 13 and 10 to investigate and rectify potential misuse or fraudulent behavior.
- 🛡️ Investigate the manufacturer sales channel for systemic issues or loopholes leading to high fraud rate
- 🧬 Focus on product part improvements, particularly AC_1001, AC_1002, TV_2001, and TV_2002, to reduce legitimate and fraudulent repair requests.

Thank you