Strategic Data Analysis Report: Insurance Risk & Claims

Date: October 26, 2023

Prepared By: Ayush Kumar Sahu

Subject: Comprehensive Review of Customer Demographics, Vehicle Data, and Claim Patterns

1. Executive Summary

This report provides a detailed analysis of our current insurance dataset, which encompasses critical information regarding policyholders, their vehicles, and their historical claims. The primary objective of this document is to structure the available data into actionable insights. By understanding the correlation between customer demographics (such as age and education), vehicle characteristics, and claim behaviors, we can develop more accurate risk profiles.

This analysis serves as a foundational document for future strategic initiatives, including premium optimization, fraud detection, and customer segmentation.

2. Customer Demographics: Understanding the Policyholder

To accurately assess risk, we must first understand who our customers are. The dataset provides a rich demographic profile that goes beyond simple identification.

Socio-Economic Indicators

We analyze **Education** and **Household Income** not just as financial markers, but as behavioral indicators. Industry data suggests that higher education levels and stable incomes often correlate with more cautious driving behaviors and consistent premium payments. By segmenting customers into income bands, we can tailor our marketing efforts—offering premium coverage packages to high-income households while providing budget-friendly options to cost-sensitive segments.

Personal Attributes & Stability

Personal details such as **Marital Status** and **Gender** allow us to assess lifestyle stability. Married individuals, for instance, are statistically viewed as "lower risk" due to perceived stability in daily routines. Furthermore, the **Birthdate** field allows us to calculate customer **Age**, a critical factor in insurance. Younger drivers (often under 25) typically face higher premiums due to inexperience, while older demographics may present different risks related to reflexes or health.

Household Structure

A unique aspect of this dataset is the inclusion of **Parental Status** and **Kids Driving**. A household with multiple teenagers driving is fundamentally different from a single-

driver household. The "Kids Driving" metric acts as a risk multiplier; as the number of young, inexperienced drivers in a household increases, the probability of minor accidents rises proportionally.

3. Vehicle Risk Profile: The Asset Analysis

The vehicle itself is a major variable in the claims equation. We categorize vehicle risk based on three main pillars: identification, condition, and usage.

Identification & Repair Costs

The **Car Make** and **Car Model** directly influence the "severity" of a claim. Luxury brands (e.g., BMW, Mercedes) carry higher repair costs compared to economy brands (e.g., Toyota, Ford). Consequently, a minor accident involving a luxury vehicle results in a significantly higher **Claim Amount** than the same accident in a budget vehicle.

Vehicle Age & Condition

The **Car Year** allows us to calculate the vehicle's age. Older vehicles present a dual challenge: they are more prone to mechanical failure (increasing roadside assistance claims) but are often cheaper to repair or replace than brand-new cars. Conversely, new cars have high replacement values but modern safety features that may reduce injury liability.

Usage Patterns

Perhaps the most critical risk factor is **Car Use**.

- **Commercial Use:** These vehicles spend the most time on the road, often during peak hours, leading to the highest exposure to accidents.
- **Commute:** These vehicles have moderate exposure, primarily during rush hour traffic.
- **Personal Use:** These vehicles generally have the lowest mileage and, subsequently, the lowest risk profile.

4. Geographic & Environmental Factors

Risk is not just about *who* is driving, but *where* they are driving. The **Coverage Zone** metric (e.g., Urban, Rural) provides context to the claims data.

- **Urban Zones:** Typically show a high **Claim Frequency** due to congestion and traffic density, though individual claim amounts may be lower (minor fender benders).
- **Rural Zones:** Typically show lower frequency but potentially higher severity due to higher speeds on open roads.

5. Financial Analysis: Claims & Profitability

The core of our business health lies in the relationship between **Claim Frequency** and **Claim Amount**.

Frequency vs. Severity

We must distinguish between "frequent claimants" and "severe claimants."

- **High Frequency / Low Severity:** A customer who files many small claims (e.g., scratches, minor bumps) may indicate poor driving habits.
- **Low Frequency / High Severity:** A customer who rarely files a claim but files a massive one (e.g., total loss of vehicle) represents a different type of financial shock.

Fraud Detection Capabilities

By cross-referencing these financial metrics with vehicle data, we can identify anomalies. For example, if we observe a high frequency of claims for a specific **Car Color** or unexpected claim amounts for older vehicles, these outliers can be flagged for special investigation by the fraud department.

6. Strategic Recommendations

Based on the data points outlined above, the following strategies are recommended for the upcoming quarter:

1. **Dynamic Pricing Model:** Move away from flat-rate pricing. Utilize **Age**, **Car Use**, and **Coverage Zone** to create a weighted risk score for every new applicant.
2. **Targeted Family Plans:** Create specific bundles for households with **Kids Driving**, offering safety courses in exchange for premium discounts.
3. **Portfolio Balancing:** Ensure our portfolio is not over-exposed to **Commercial** vehicles in high-risk **Urban** zones.

End of Report