

INSURANCE RISK & CLAIMS ANALYSIS

DOMAIN DOCUMENT

This database holds detailed information on automobile policyholders, vehicles and loss records. One record exists per individual customer with demographic information like birthdate, sex, marital status, education, household income and parent status. These characteristics are a good foundation on which to start discovering about customer profiles and how unique characteristics might affect driving habits and insurance risk.

The information also captures specific details about the insured car, such as car make, model, color and manufacturing year, as well as use type (personal, commercial, or commuting). Combined with the area covered, this information allows for estimation of vehicle and environment-related risk factors. For instance, an older commercial use car in a city area will have a different risk profile than that of a newer private car in a lower-density urban area.

For the claims and money side, the database contains claim amount (total incurred cost of claims), claim frequency (incurred claims), and income at the household level. Customer value and risk are determined using these. Repeat claimers might mean greater risk, and high-cost claimers might mean bad accidents or expensive repairs. The Kids Driving variable adds a new aspect, outlining driving behavior at the family level insofar as those with multiple young drivers are riskier from the point of view of the insurer.

Overall, this data set provides a comprehensive view of claims, vehicles, and customers. It can be utilized to create Power BI dashboards for claim analysis, customer segmentation, risk profiling, and premium optimization. The insights that can be gained from the data can be utilized to empower insurance companies to create more balanced pricing models, determine high-risk segments, detect potential fraud, and enhance customer targeting campaigns.

1. ID

- **Definition:** A unique identifier assigned to each policyholder or insurance record.
- **Type:** Numeric or text (Primary Key).
- **Details:**
 - o Ensures each policy/customer record is unique.

- o Not used directly in analysis but essential for data integrity, relationships, and avoiding duplicates.
- **Business Use:** Helps track individual policyholders across multiple datasets (e.g., linking claims, payments, or renewals).

2. Birthdate

- **Definition:** Date of birth of the policyholder.
- **Type:** Date.
- **Details:**
 - o Used to calculate the **Age** of the policyholder.
 - o Age is a strong factor in insurance risk assessment (younger drivers may have more accidents, older drivers may have slower reflexes).
- **Business Use:**
 - o Segment customers into age groups (e.g., <25, 25–40, 40–60, 60+).
 - o Age-based pricing of premiums and risk profiling.

3. Car Color

- **Definition:** The exterior color of the insured car.
- **Type:** Categorical (text).
- **Details:**
 - o Although color doesn't directly affect risk, it can influence **car visibility, theft probability, and customer preferences**.
- **Business Use:**
 - o Explore patterns (e.g., some studies suggest red cars might be involved in more speeding cases).
 - o Insights for fraud detection (unusual color-claim patterns).

4. Car Make

- **Definition:** The manufacturer/brand of the car (e.g., Toyota, Ford, BMW).
- **Type:** Categorical.
- **Details:**

- o Different makes have varying repair costs and accident likelihood.
- **Business Use:**
 - o Analyze claims by car brand.
 - o Identify which brands have higher claim frequency (riskier) vs lower claims (safer).
 - o Useful for underwriting and premium adjustments.

5. Car Model

- **Definition:** Specific model of the car (e.g., Toyota Corolla, BMW X5).
- **Type:** Categorical.
- **Details:**
 - o Provides deeper granularity beyond car make.
 - o Claim risk can vary significantly between models of the same make.
- **Business Use:**
 - o Compare claims for luxury vs economy models.
 - o Useful for **loss ratio** analysis by vehicle type.

6. Car Use

- **Definition:** The primary purpose of the car (e.g., Personal, Commercial, Commute).
- **Type:** Categorical.
- **Details:**
 - o Commercial and commuting cars usually have higher claim rates due to greater exposure (more time on the road).
- **Business Use:**
 - o Pricing policies differently for personal vs commercial use.
 - o Risk profiling based on usage patterns.

7. Car Year

- **Definition:** The manufacturing year of the car.
- **Type:** Numeric.
- **Details:**

- o Helps calculate **Car Age** = Current Year – Car Year.
- o Older cars may break down more often or have lower resale/repair costs.
- **Business Use:**
 - o Assess claim likelihood by vehicle age.
 - o Create segments like New (<3 years), Mid-age (3–10 years), Old (>10 years).

8. Coverage Zone

- **Definition:** The geographic risk area where the policyholder resides/uses the car (e.g., Zone A, Urban, Rural).
- **Type:** Categorical.
- **Details:**
 - o Location impacts accident probability, theft risk, and repair costs.
 - o Urban areas → higher claim frequency, rural areas → fewer accidents but possibly more severe damage.
- **Business Use:**
 - o Compare performance across regions.
 - o Useful for **mapping dashboards in Power BI**.

9. Education

- **Definition:** The highest education level attained by the policyholder (e.g., High School, Graduate, Postgraduate).
- **Type:** Categorical.
- **Details:**
 - o Education level often correlates with income and risk behavior.
- **Business Use:**
 - o Check claim frequency by education group.
 - o Insights for marketing campaigns and segmentation.

10. Gender

- **Definition:** Gender of the policyholder (Male, Female, Other).
- **Type:** Categorical.

- **Details:**
 - Gender differences in claim patterns may exist.
- **Business Use:**
 - Risk analysis by gender group.
 - Demographic breakdowns in dashboards.

11. Marital Status

- **Definition:** Marital status of the policyholder (Single, Married, Divorced, etc.).
- **Type:** Categorical.
- **Details:**
 - Married people are often seen as more stable, potentially lower risk.
- **Business Use:**
 - Compare claims between singles vs married customers.
 - Factor in pricing and risk adjustment.

12. Parent

- **Definition:** Whether the policyholder has children (Yes/No).
- **Type:** Boolean / Categorical.
- **Details:**
 - Parents may drive differently compared to non-parents.
- **Business Use:**
 - Segment claims and policies by parent status.
 - Combine with **Kids Driving** for more detailed household analysis.

13. Claim Amt

- **Definition:** Total monetary value of claims filed by the policyholder.
- **Type:** Numeric (currency).
- **Details:**
 - A critical measure of insurance losses.
- **Business Use:**
 - Used to calculate **total losses, average claim amount, and claim ratios.**

- o Identify high-cost claims for fraud detection.

14. Claim Freq

- **Definition:** The number of claims filed by the policyholder.
- **Type:** Numeric (integer).
- **Details:**
 - o Frequency of claims indicates risk behavior.
- **Business Use:**
 - o KPI for customer risk assessment.
 - o Combined with Claim Amt → identifies whether risk is due to high frequency or high severity.

15. Household Income

- **Definition:** Annual income of the policyholder's household.
- **Type:** Numeric.
- **Details:**
 - o Income level affects affordability of premiums and vehicle type.
- **Business Use:**
 - o Segment customers into income bands.
 - o Compare claim patterns across income groups.
 - o Target marketing/policy offers (e.g., premium insurance for high-income households).

16. Kids Driving

- **Definition:** Number of kids in the household who are licensed drivers.
- **Type:** Numeric (integer).
- **Details:**
 - o Values (1, 2, 3...) indicate the number of kids driving.
 - o More kids driving generally = higher accident probability.
- **Business Use:**
 - o Analyze claims by number of kids driving.

- o Group into categories (e.g., None, 1 Kid, 2 Kids, 3+ Kids).
- o Helps insurers estimate family household risk.