**Swedish Motor Insurance Claims Analysis Report**

**1. Executive Summary**

This report analyzes third-party motor claims data from Sweden (1977) to identify the key factors influencing claim payments and rates. Using R for data cleaning, visualization,, and regression modelling, we find that the number of claims and insured policy-years are the strongest predictors of total payment. Location (), car make, and bonus class also play significant roles. These insights support targeted risk management, premium setting, and branch expansion strategies.

**2. Data Overview**

* **Observations:** 2,182
* **Variables:**
  + **Kilometres:** Annual distance driven (5 categories)
  + **Zone:** Geographic region (7 categories)
  + **Bonus:** No-claims bonus class (7 categories)
  + **Make:** Car model group (9 categories)
* **Insured:** Number of insured policy-years (numeric)
  + **Claims:** Number of claims (numeric)
  + **Payment:** Total value of payments in SEK (numeric)
* **No missing values**
* **Categorical variables converted to factors for analysis**

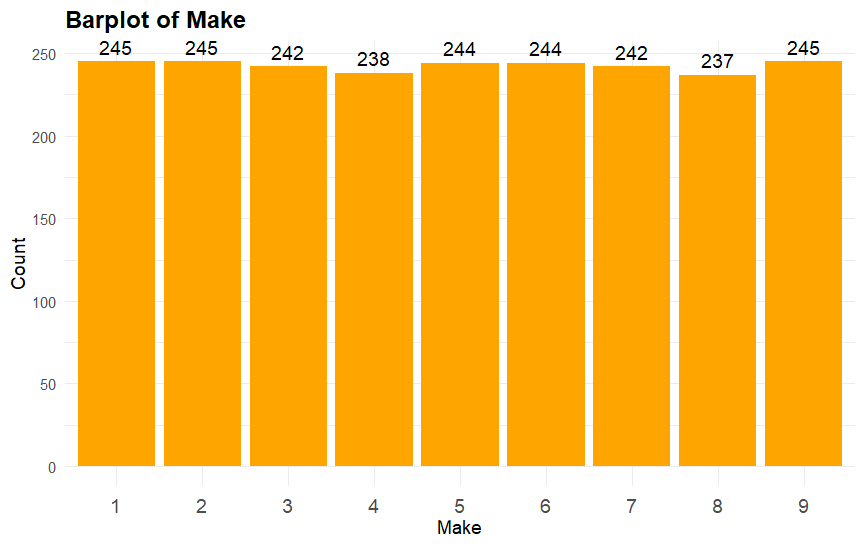
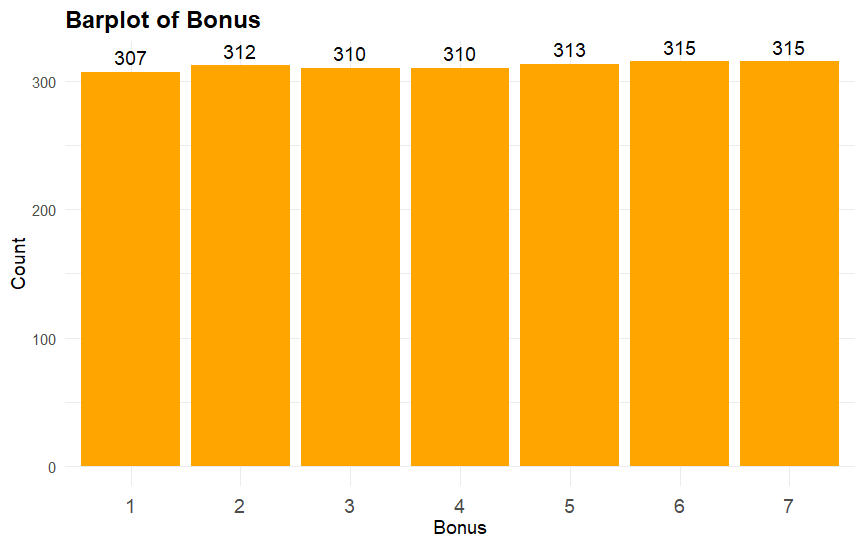
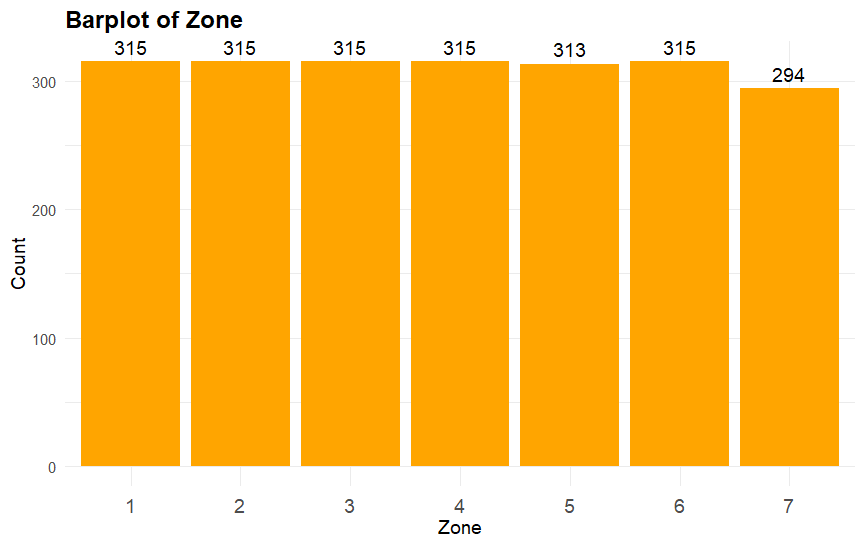
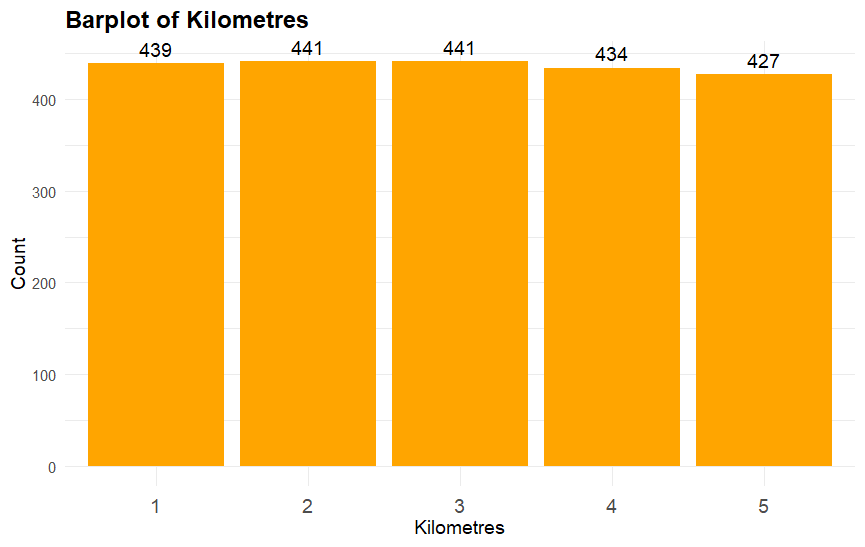
**3. Exploratory Data Analysis (EDA)**

**Distributions**

* All numeric variables (Insured, Claims, Payment) are highly right-sked with many low values and some very large outliers.
* Each categorical variable is well-represented across its levels.

*Visuals:*

* *Barplots of Kilometres, Zone, Bonus, Make*



**4. Correlation & Bivariate Relationships**

* **Correlation matrix:** Payment, Claims, and Insured are all strongly correlated (r > 0.9).
* **Scatterplots:** Both Payment vs Claims and vs Insured show strong linear relationships.

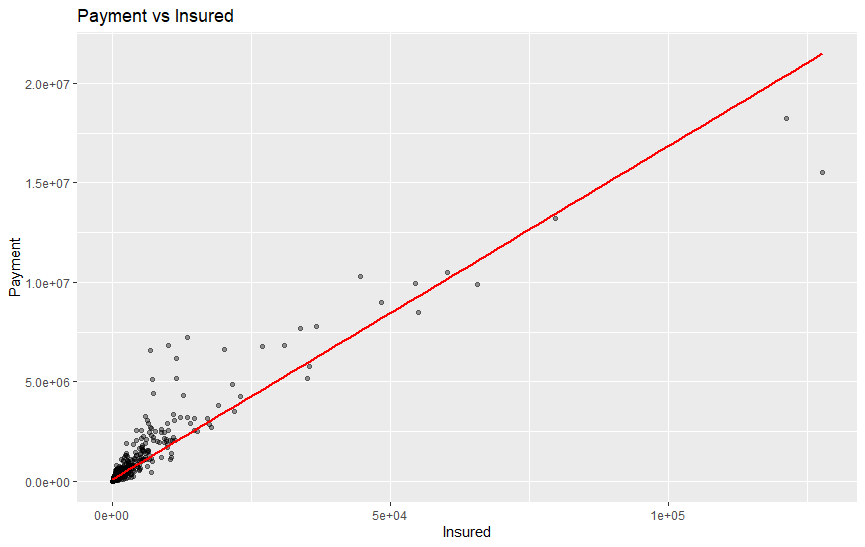
*Visuals:*

* *Correlation heat*

A blue circles with red text

AI-generated content may be incorrect.

* *Scatterplots with regression lines for Payment vs Claims and Payment vs Insured*

A graph with a red line

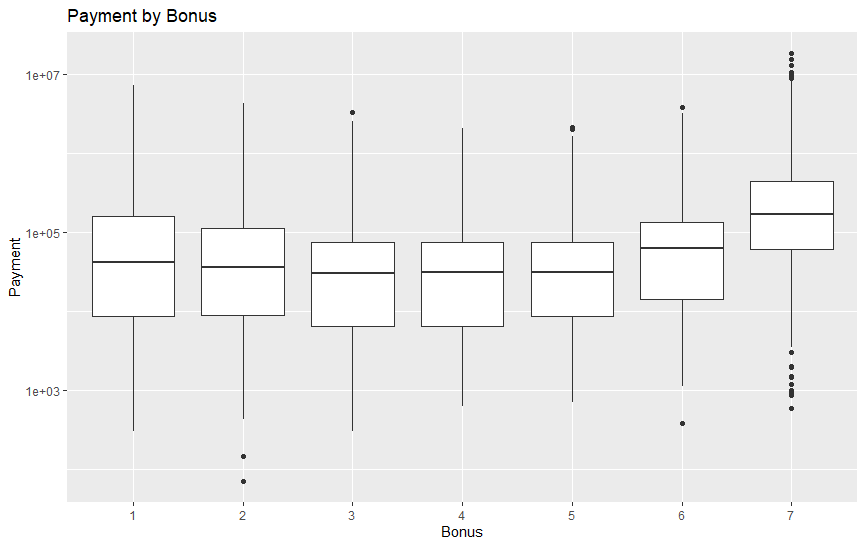
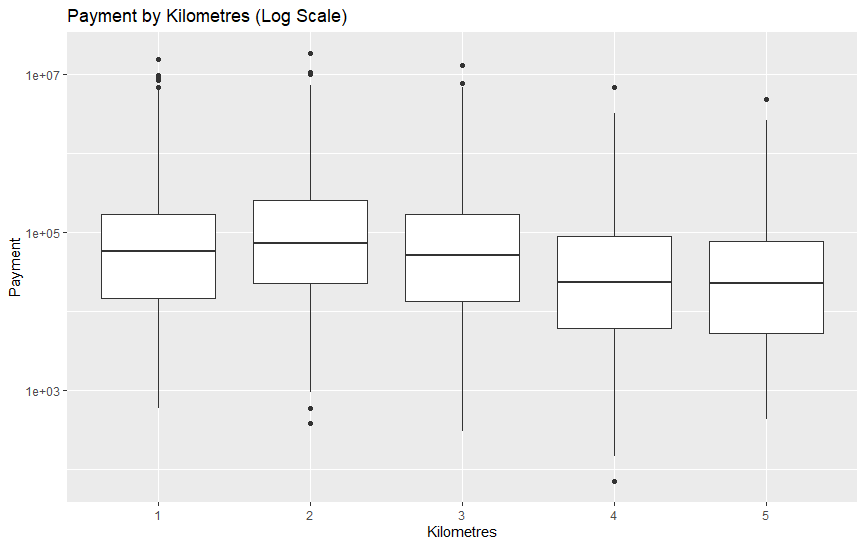
AI-generated content may be incorrect.

**5. Payment by Category (Log Scale)**

* **Boxplots (log scale):**
  + *Zone:* Zones 4 and 6 have higher median payments.
  + *Kilometres:* Higher kilometre groups show slightly higher payments, but overlap is substantial - *Bonus:* Bonus class 7 has the highest median and spread in payments, possibly due to larger portfolios.
  + *Make:* Make 9 (rare models) shows greater variability and higher payments.

*Visuals:*

* *Boxplots for Payment by Zone, Kilometres, Bonus, Make (all log scale)*A graph with a row of squares

  AI-generated content may be incorrect.A graph with a row of squares

  AI-generated content may be incorrect.

**6. Regression Modeling**

**Simple Model: Payment ~ Claims + Insured**

* **R² = 0.9951:** Explains nearly all variation in Payment.
* **Coefficients:** Both Claims and Insured are highly significant (p < 0.001).

**Full Model: Payment ~ All Predictors**

* **R² = 0.9954**
* **Significant predictors:** Claims, Insured, several Kilometres, Zone, Bonus, and Make levels.
* **Interpretation:** While Claims and Insured dominate, location, make, and bonus effects remain after controlling for volume.
* **Multicollinearity:** Vs are acceptable (all < 10).

*Visuals:*

* *Regression summary tables (appendix or as needed)*

**7. Aggregated Analysis**

* Aggregated totals by Zone, Kilometres, and Bonus reveal where most business and risk are concentrated.
* Example: “Zone 1, Kilometres 1, Bonus 1 has the highest aggregate payment.”

*Table:*

* *Top rows of aggregated summary table (Zone, Kilometres, Bonus, Total\_Insured, Total\_Claims, Total\_Payment)*

**8. Claim Rate Analysis**

* **Model:** ClaimRate ~ Insured + Zone + Kilometres + Bonus + Make
* **R² = 0.18:** Only 18% of variance explained.
* **Key findings:** Higher Zone numbers (less urban), higher Kilometres, and higher Bonus are with lower claim rates. Some Makes are significant.

*Visual:*

* *Histogram of Claim Rate*

A graph of a number of gray bars

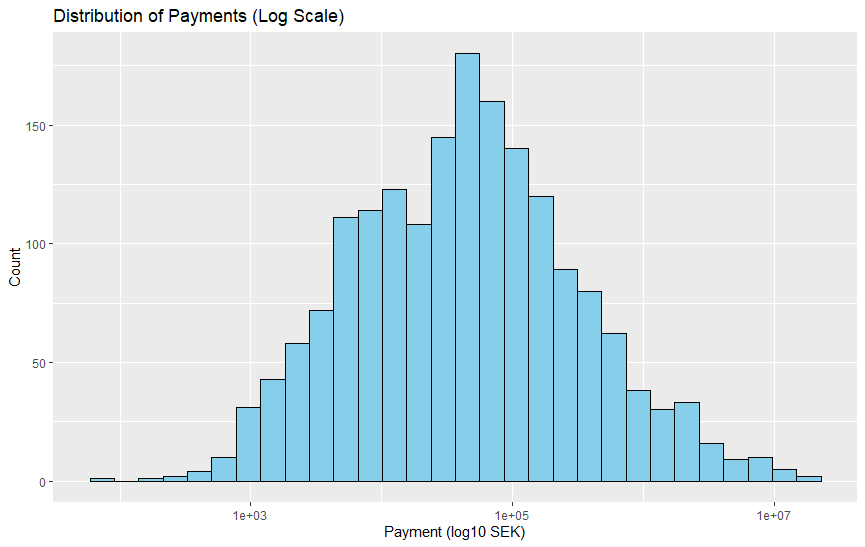
AI-generated content may be incorrect.

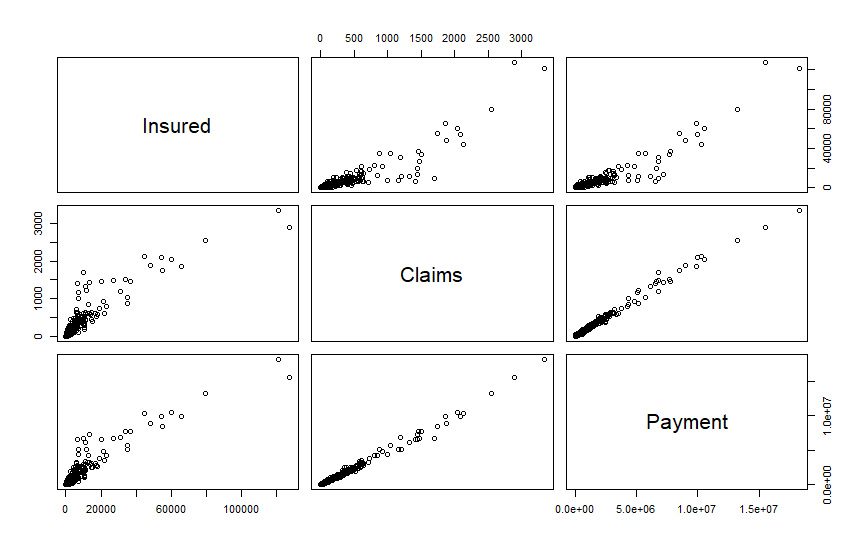
**9. Visualizations for Key Findings**

Below are visuals that summarize the main results:

* Histogram: Distribution of Payments (right-skewed)
* Barplot: Policy counts by Bonus class Scatterplot: Payment vs Claims (with regression line)
* Histogram: Claim Rate distribution

A graph of numbers and a number of points

AI-generated content may be incorrect.



**10. Recommendations**

* **Claims and Insured drive payments:** Focus on accurate forecasting and monitoring of these metrics.
* **Location, Make, and Bonus matter:** Consider these for branch planning, premium, and targeted marketing.
* **High-risk segments:** Identify and manage portfolios with high claim rates or payments.
* **Data-driven decisions:** Use aggregated and regression insights to inform expansion and pricing strategy.