Exploratory Data Analysis (EDA) Report

Identification of Premium Pricing Attributes for Home Insurance

---

1. Introduction

The objective of this Exploratory Data Analysis (EDA) is to analyze home insurance data to identify the key factors influencing **premium pricing** and **payment behavior**.
The analysis aims to understand customer and property characteristics that affect insurance risk, premium affordability, and payment reliability.

---

2. Dataset Description

The dataset contains information related to customers, properties, and insurance policies. It includes both numerical and categorical variables.

Key Variables:

- Customer Income
- Age / Age Group
- Property Value
- House Age
- Location (Urban / Rural)
- Premium Amount
- Claim Amount
- Payment Status

The dataset provides a comprehensive view of factors affecting insurance pricing and claims.

---

3. Data Cleaning and Preprocessing

Before analysis, the dataset was cleaned and prepared using the following steps:

- Checked for missing values and handled them appropriately
- Removed duplicate records
- Corrected data types for numerical and categorical variables

- Created derived variables such as **Age Group** and **Income Group**
- Ensured data consistency and accuracy
- After cleaning, the dataset was ready for exploratory analysis.

---

4. Univariate Analysis

Univariate analysis was performed to understand the distribution of individual variables.

Key Observations:

- Income values showed a wide range across customers
- Premium amounts varied significantly based on coverage and risk
- Claim amounts were right-skewed, indicating fewer high-value claims
- Most customers belonged to middle age groups

This step helped identify basic patterns and detect potential outliers.

---

5. Bivariate Analysis

Bivariate analysis was conducted to examine relationships between pairs of variables.

Key Relationships Observed:

- Income and Premium Amount showed a positive relationship
- Property Value and Claim Amount were positively correlated
- Older houses experienced higher claim frequency
- Urban locations showed higher premium values

This analysis provided insights into how different attributes interact with premium pricing and risk.

---

6. Advanced EDA Using Visualizations

Additional visual analysis was performed using simple plots to clearly identify customer segments related to **premium acceptance** and **payment difficulty**.

6.1 Income vs Premium Amount

Observation:
Customers with higher income tend to pay higher premiums.

Insight:
High-income customers are less price-sensitive and more willing to opt for comprehensive insurance plans.

6.2 Age Group vs Premium Amount

Observation:
Customers aged between 30 and 50 years pay higher premiums on average.

Insight:
This group is financially stable and more aware of insurance risks.

6.3 Property Value vs Claim Amount

Observation:
Higher property values are associated with higher claim amounts.

Insight:
Premium pricing should increase with property value to manage risk exposure.

6.4 Income Group vs Payment Status

Observation:
Low-income customers show a higher frequency of delayed or missed payments.

Insight:
Payment affordability is a major concern in lower-income segments.

6.5 Location vs Premium Amount

Observation:
Urban customers pay higher premiums compared to rural customers.

Insight:
Urban areas have higher risk exposure due to higher property density and costs.

6.6 House Age vs Claim Frequency

Observation:
Older houses generate more frequent insurance claims.

Insight:
Structural aging increases maintenance and damage-related risks.

---

7. Customer Segmentation Based on EDA

Customers Likely to Pay Higher Premiums:

- High-income customers
- Urban residents
- Middle-aged customers (30–50 years)
- Owners of high-value properties

Customers Facing Payment Difficulties:

- Low-income customers
- Customers with older houses
- Customers with frequent claims

---

8. Key Findings from Overall EDA

- Income is the strongest factor influencing premium affordability
- Property value directly impacts claim amount and premium pricing
- Location plays a significant role in determining insurance risk
- House age affects claim frequency
- Payment difficulties are more common among low-income customers

---

9. Business Recommendations

- Implement **risk-based and income-based premium pricing**
- Apply **location-specific premium adjustments**
- Increase premiums or inspections for **older houses**
- Offer **flexible payment options** to low-income customers
- Use **claim history** for better premium calculation

10. Conclusion

This EDA combines initial data exploration and advanced visual analysis to identify key premium pricing attributes in home insurance.
The findings support the adoption of **segment-based pricing strategies** to improve profitability while minimizing payment default risks.