📊 Project 1 – Key Insights

1. **Significant Claim Drivers Identified:**
    - The logistic regression model highlights that variables like **Claim_Amount**, **Hospital_ID**, and **Diagnosis_Code** are statistically significant in predicting claim outcomes (e.g., rejection or fraud).
2. **Age Shows Positive Correlation with Claim Amount:**
    - Correlation analysis indicates that **older customers tend to claim higher amounts**, which is consistent with increased health risks with age.
3. **Variation by Policy Type:**
    - Customers under **Corporate policies** generally show higher average claim amounts compared to Family or Individual plans, suggesting a potential for adverse selection or richer benefit plans.
4. **High-Value Hospitals:**
    - Certain hospitals appear repeatedly in high-claim data clusters. These may require further review for pricing agreements, overutilization, or fraud potential.
5. **Potential Fraud Risk Factors:**
    - The model output suggests certain **diagnosis codes and hospital patterns** are correlated with higher fraud flags, indicating the need for focused audit rules or ML-based scoring.
6. **Underwriting Strategy Alert:**
    - Younger customers with high claim frequency under individual policies may suggest misuse or poor underwriting controls.
7. **Actionable Strategy:**
    - **Strengthen fraud detection rules** using claim amount + provider + diagnosis patterns.
    - **Refine premiums** for Corporate and Individual plans based on age and claim history trends.
    - **Audit high-risk hospital-claim clusters** regularly.