Exploratory Data Analysis (EDA) Report: Delinquency Risk

# Step 1: Data Overview & Initial Assessment

• Notable Missing Data:  
 – Income: 7.8% missing.  
 – Credit Score: 0.4% missing.  
 – Loan Balance: 5.8% missing.  
  
• Key Anomalies:  
 – Some unusually high credit utilisation values (>100%).  
  
• Early Indicators of Risk:  
 – Customers with multiple missed payments.  
 – High credit utilisation and elevated debt-to-income ratios.

Initial data quality is generally strong, with only three columns requiring imputation. Patterns in missed payments and high utilisation suggest predictive power. No major format or duplication issues observed.

# Step 2: Handling Missing Data

|  |  |  |
| --- | --- | --- |
| Feature | Action Taken | Justification |
| Income | Imputed with mean | Income is a continuous variable; mean preserves scale. |
| Loan\_Balance | Imputed with median | Median handles skewed distribution better. |
| Credit\_Score | Imputed with mean | Slight missingness; mean reflects average customer risk. |

Missing values were imputed using statistical methods to preserve data integrity. Features with very low missing percentages were retained to avoid reducing sample size.

# Step 3: Detecting Patterns and Risk Factors

• Missed Payments – Strongest correlation with delinquency (≈ +0.65). Customers with ≥4 missed payments are highly delinquent.  
• High Credit Utilisation – Correlation ≈ +0.40. Median utilisation is much higher in delinquent group.  
• Debt-to-Income Ratio – Moderate correlation (≈ +0.35). Risk increases significantly above 40% DTI.  
• Low Credit Score – Negative correlation (≈ -0.32). Median score is significantly lower among delinquents.  
• Short Tenure – Accounts <2 years old show higher risk despite lower balances.  
  
Surprising Insight:  
• A small group with over-limit credit utilisation (>100%) had a very high delinquency rate (~70%) despite only moderate missed-payment history.

Exploratory pattern analysis confirms several strong delinquency drivers. Missed payments are the most reliable signal of delinquency, followed by high utilisation and DTI. Lower credit scores and short tenure modestly contribute to risk. One unexpected pattern was sudden risk elevation in over-limit accounts, which requires immediate review.