Exploratory Data Analysis (EDA) Report

Step 1: Review and Key Insights

The dataset contains 500 customer records with 19 features capturing demographic, financial, and behavioral factors. Initial review identified missing values in income (7.8%), credit score (0.4%), and loan balance (5.8%). High credit utilization and missed payments are prominent indicators of delinquency risk.

Notable Missing or Inconsistent Data

- Income: 39 missing values (7.8%)
- Credit Score: 2 missing values (0.4%)
- Loan Balance: 29 missing values (5.8%)

Key Anomalies

- Credit Utilization values > 0.9 indicate over-leveraged accounts
- Some accounts with 0 missed payments are still delinquent
- Monthly payment history includes mixed categorical values (Late/Missed/On-time)

Early Indicators of Delinquency

- Missed Payments
- Credit Utilization
- Debt-to-Income Ratio

Step 2: Missing Data Handling

Feature	Missing Count	Strategy	Justification
Income	39	Median Imputation	Median avoids bias due to skewed distribution
Credit_Score	2	Mean Imputation	Few missing; mean is sufficient
Loan_Balance	29	Predictive Imputation	Correlates with income & credit score

Step 3: Patterns and Risk Factors

- Missed Payments: Strong predictor of delinquency
- Credit Utilization: High ratios (>80%) indicate higher risk

- Debt-to-Income Ratio: High values correlate with delinquency
- Low Credit Score: Scores <450 often delinquent
- Loan Balance: Large balances relative to income indicate stress

Unexpected findings include delinquent accounts with zero missed payments, suggesting hidden variables (e.g., job loss). Monthly history shows alternating patterns between 'On-time' and 'Missed' indicating fluctuating customer risk.

Step 4: Summary

The dataset reveals clear delinquency risk patterns driven by missed payments, credit utilization, and debt ratios. Missing data primarily affects income and loan balance, addressed via median and predictive imputation. Categorical monthly payment histories will require encoding for modeling. These insights will guide feature engineering and predictive modeling in subsequent steps.