

LOAN DEFAULT

1.Business Problem Framing

Lenders need to monitor loan defaults to protect capital, set appropriate provisioning, and target collections and risk-based pricing. Predicting and tracking defaults reduces loss, improves capital allocation, and supports regulatory reporting.

2. Data Requirements (from CSV)

Columns detected (first 40): LoanID , Age , Income , LoanAmount , CreditScore , MonthsEmployed , NumCreditLines , InterestRate , LoanTerm , DTIRatio , Education , EmploymentType , MaritalStatus , HasMortgage , HasDependents , LoanPurpose , HasCoSigner , Default .

Data sample (first 4 rows):Only takes the important values

LoanID	Income	LoanAmount	CreditScore	Education	EmploymentType	LoanPurpose	InterestRa
I38PQUQS96	50432.00	124440.00	458	Master's	Full-time	Other	15.23
HPSK72WA7R	84208.00	129188.00	451	Master's	Full-time	Other	14.81
V2KKSFM3UN	31713.00	44799.00	743	High School	Unemployed	Auto	21.17
A9S62RQ7US	20437.00	9139.00	633	Bachelor's	Full-time	business	7.07

3. Key Performance Indicators (KPIs)

- Total records (loans): 255347
- Total loan principle (sum): 32576880572.00
- Average loan principal: 127578.87 (n=255347)
- Total outstanding balance: Column not found
- Default column used: Default
- Number of defaults: 29653
- Default rate: 11.61%
- Average credit score: 574.26 (n=255347)
- Share missing credit score: 0.00%
- Average term (units as in data): 36.03 (n=255347)
- Average interest / rate: 13.49 (n=255347)

- Top 5 borrowers by aggregate loan principle:
{'BZ6O5YSSA8': 249999, 'TGYZ4AXW7H': 249998, 'VIS6H9YBJM': 249997, '7OV9HHSV62': 249996, 'INBGJEMBW3': 249993}
- Recovery / LGD info: Not available - no recovery column or defaults not identified

4. Dashboard / PBIX inspection (best-effort)

PBIX is a proprietary file. For full structural analysis, open in Power BI Desktop to inspect visuals, measures, and data models.

5. Recommended KPIs to include on a loan-default dashboard

- Portfolio Default Rate: Percentage of loans that are in default over a period. Formula: defaults / total loans. High-level health indicator.
- Delinquency Rate (30/60/90 days): Share of loans past due by 30/60/90 days. Early warning signal for future defaults.

- Non-Performing Loan (NPL) Ratio: Outstanding principal of NPLs / total outstanding principal. Shows exposure at risk.
- Average Exposure at Default (EAD): Average outstanding balance for loans that defaulted. - Loss Given Default (LGD): Percent of exposure lost after recoveries. Formula: $(1 - \text{recoveries} / \text{exposure})$.
- Expected Loss (EL): $EL = PD * LGD * EAD$. Use modeled Probability of Default (PD) if available.
- Recovery Rate: Recoveries / exposure for defaulted loans. - Average Credit Score of Defaulters: Shows risk profile of defaulters vs non-defaulters.
- Concentration Metrics: Top 5 borrowers share of portfolio or top industry/region exposures.
- Vintage Default Rate: Defaults by origination cohort (e.g., loans originated in same month/year). Useful for trend analysis.
- Provision Coverage Ratio: Provisions set aside / expected loss or NPLs. Regulatory/finance metric.
- Time-to-default distribution: Histogram or percentiles of days from origination to default.

6. Next steps & Implementation Outline Feature engineering suggestions:

age of loan, months delinquent, payment_to_income_ratio (if income present), vintages, recent repayment behavior. ML approach: gradient-boosted trees (LightGBM/XGBoost) for PD modeling; survival analysis for time-to-default. Explainability: SHAP to explain drivers per-account.

7. Risks & Guardrails

- Data quality / missing values;
- Label leakage;
- Model drift;
- Fairness and regulatory compliance.

Use robust validation and monitoring.