CredRisk

Credit Risk Analysis & Default Prediction Dashboard

Sector: Finance

Team Member Names:

Harshit Shakya - 2401010183
Anurag Kumar Tiwari - 2401020009
Akash Kumar Gautam - 2401010050
Anugra Gupta - 2401010085
Prashant Raj - 2401010340
Vivek Kumar Raj - 2401020077

Institute Name: Newton School of Technology (Rishihood University)

Faculty Name: Satyaki Das

EXECUTIVE SUMMARY

Problem :

- Analyse the loan portfolio dataset to identify key risk factors driving loan defaults and credit performance.
- Understand how borrower credit score, interest rate, debt-to-income ratio (DTI), loan term, and income verification influence default probability and overall portfolio risk.

Approach:

- Cleaned and structured ~6,000+ loan records in Google Sheets.
- Created helper columns such as default_flag, FICO buckets, DTI buckets, interest rate buckets, and term segmentation to enable accurate pivot analysis.
- Built pivot tables and an interactive dashboard to evaluate default trends across credit score bands, pricing tiers, loan duration, and verification status.

Key Insights:

- Lower FICO score borrowers exhibit significantly higher default rates, confirming strong credit-risk sensitivity.
- Higher interest rate buckets show materially elevated charge-offs, validating risk-based pricing alignment.
- Loans with high DTI ratios (>45%) demonstrate a sharp increase in default probability.
- 60-month loans show higher cumulative default rates compared to 36-month loans, indicating time-exposure risk.
- Income-verified borrowers show marginally lower default rates than non-verified borrowers.

Recommendations:

- Prioritise lending toward higher FICO and moderate DTI segments to reduce portfolio risk.
- Refine pricing strategy for high-risk buckets to better compensate for elevated default probability.
- Limit long-term (60-month) loan exposure in high-risk borrower segments.
- Strengthen income verification standards to improve underwriting quality.

Business Impact:

- Improves risk-adjusted return by aligning pricing with default behavior.
- Enhances underwriting discipline and portfolio stability.
- Reduces charge-offs and optimises capital allocation.
- The dashboard enables continuous monitoring of risk drivers and supports data-driven credit strategy decisions.

SECTOR & BUSINESS CONTEXT

Sector Overview:

- Based on the consumer lending and credit risk sector.
- Loan performance depends on borrower credit score, interest rate, debt-to-income ratio (DTI), and loan term.

Current Challenges:

- Higher default risk in low FICO and high DTI segments.
- Balancing loan growth with risk control.
- Managing long-term loan exposure and income verification accuracy.

Why This Problem Was Chosen:

- To identify key drivers of loan defaults.
- To improve underwriting decisions and risk-based pricing strategy.

PROBLEM STATEMENT & OBJECTIVES

Formal Problem Definition:

- Analyse the loan portfolio dataset.
- Identify how FICO score, interest rate, DTI, loan term, and income verification affect default rates and portfolio risk.

Project Scope:

- Includes borrower credit score, pricing (interest rate), repayment capacity (DTI), loan duration, and verification status.
- Analysis conducted using Google Sheets, pivot tables, and an interactive dashboard.

Success Criteria:

- Identify key drivers of loan defaults.

DATA DESCRIPTION

Exact Dataset Source & Access Link:

The dataset used is a publicly available loan dataset

https://www.kaggle.com/datasets/wordsforthewise/lending-club

It contains historical loan-level data including borrower credit attributes and repayment outcomes.

Data Structure:

The dataset is structured at the loan level, where each row represents one issued loan and each column represents borrower characteristics, pricing details, and loan performance status.

Columns Explanation:

Key features include loan status, FICO score range, interest rate, debt-to-income ratio (DTI), loan term, income verification status, annual income, grade, sub-grade, and issue date.

These variables are used to evaluate default behavior and portfolio risk drivers.

Data Size:

The dataset contains approximately 6,000+ loan records used for analysis and dashboard creation.

Data Limitations:

The dataset represents historical loan performance and may not reflect real-time portfolio conditions.

DATA CLEANING & PREPARATION

Missing Values:

Handled missing entries in key fields such as DTI, annual income, and verification status by filtering incomplete records or retaining only relevant observations for analysis.

Outliers:

Reviewed extreme values in interest rate, DTI, and income to ensure unrealistic or data-entry anomalies did not distort pivot results.

Transformations:

Converted interest rates into clean numeric format and standardized loan term into numeric (36 / 60 months).
Created a binary `default_flag` variable for consistent default rate calculation.

Feature Engineering:

Generated FICO midpoint, FICO buckets, DTI buckets, interest rate buckets, term segmentation, and issue year to enable structured pivot analysis.

Assumptions:

Loan status categories such as Charged Off and Default were treated as default events. Current and Fully Paid loans were treated as non-default for comparison.

Tools Used:

All cleaning and preparation performed in Google Sheets using formulas, filters, calculated columns, and pivot tables.

KPI & METRIC FRAMEWORK

KPI Definitions:

- **Default Rate** – Percentage of loans classified as default (Charged Off / Default).
- **Average FICO Score** – Mean borrower credit score (FICO midpoint).
- **Average Interest Rate** – Mean loan pricing across portfolio.
- **DTI Ratio** – Borrower debt-to-income level used for risk segmentation.
- **60-Month Loan %** – Proportion of longer-term loans in portfolio.
- **Verification %** – Verified loans ÷ Total loans.

Formula:

- **Default Rate =** SUM(default_flag) ÷ Total Loans
- **Average FICO** = AVERAGE(fico_mid)
- **Average Interest Rate =** AVERAGE(int_rate_clean)
- **Verification % =** Verified Loans ÷ Total Loans

Why KPIs Matter:

They measure portfolio risk, borrower quality, and pricing alignment.

They show how credit score, pricing, and repayment burden influence default behavior.

They help evaluate underwriting strength and exposure to high-risk segments.

Mapping to Objectives:

Track default trends across risk buckets.

Assess whether pricing compensates for credit risk.

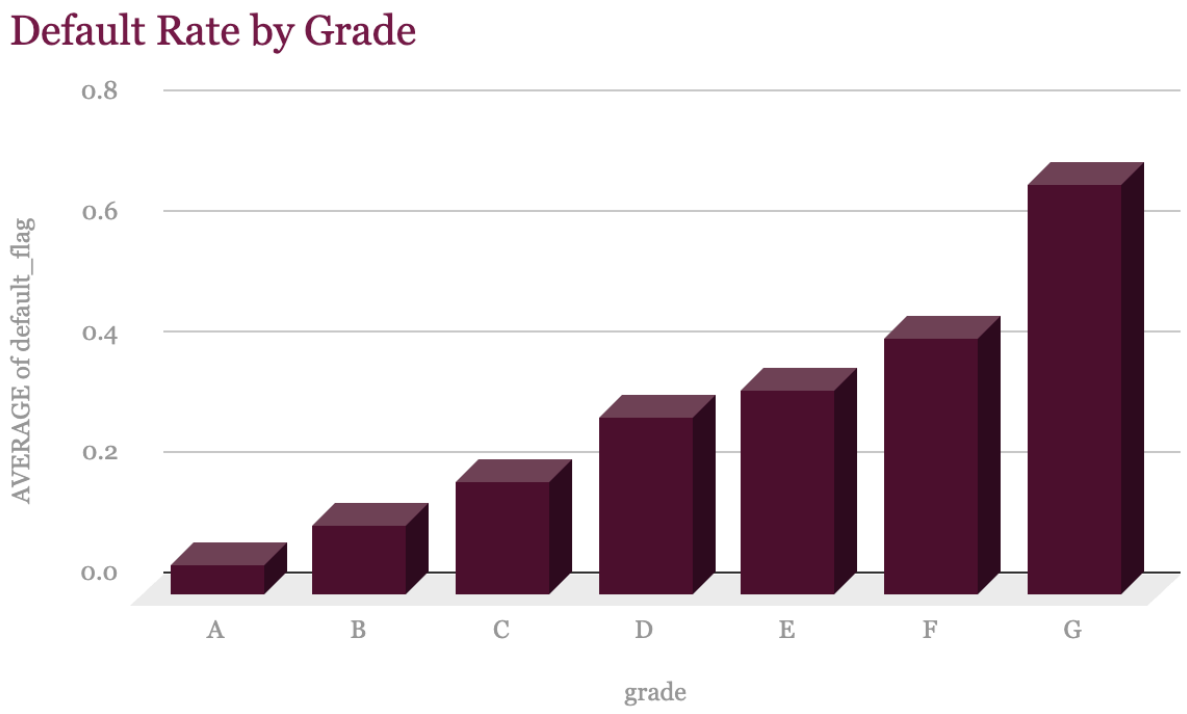
EXPLORATORY DATA ANALYSIS (EDA)

Overall Metrics:

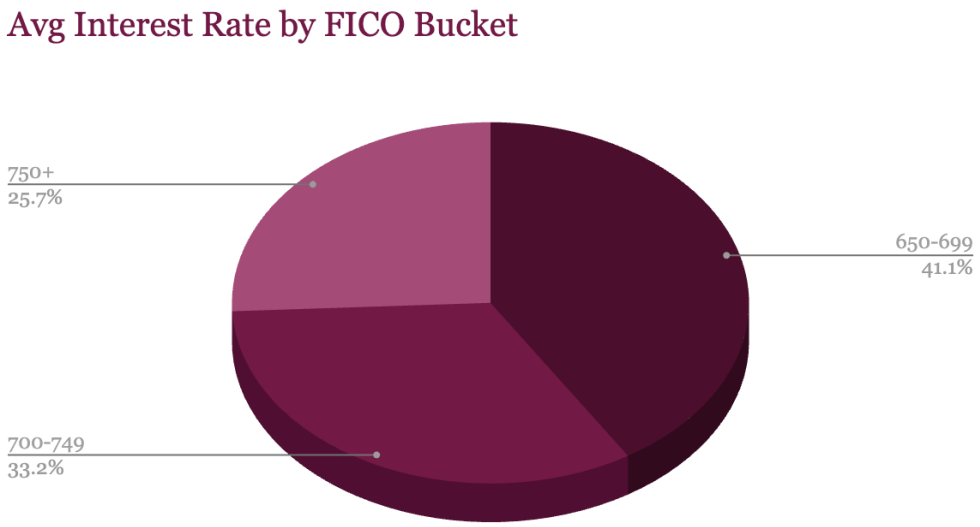
- Total loans: 5,999 | Total defaults: 1,033 | Overall default rate: 17%
- Average FICO: 698 | Average interest rate: 12% | Expected interest income: 405,348

Trend Analysis

Default rate rises with lower credit grades. Grade G has the highest default (~65%), while Grade A has the lowest (~5%).

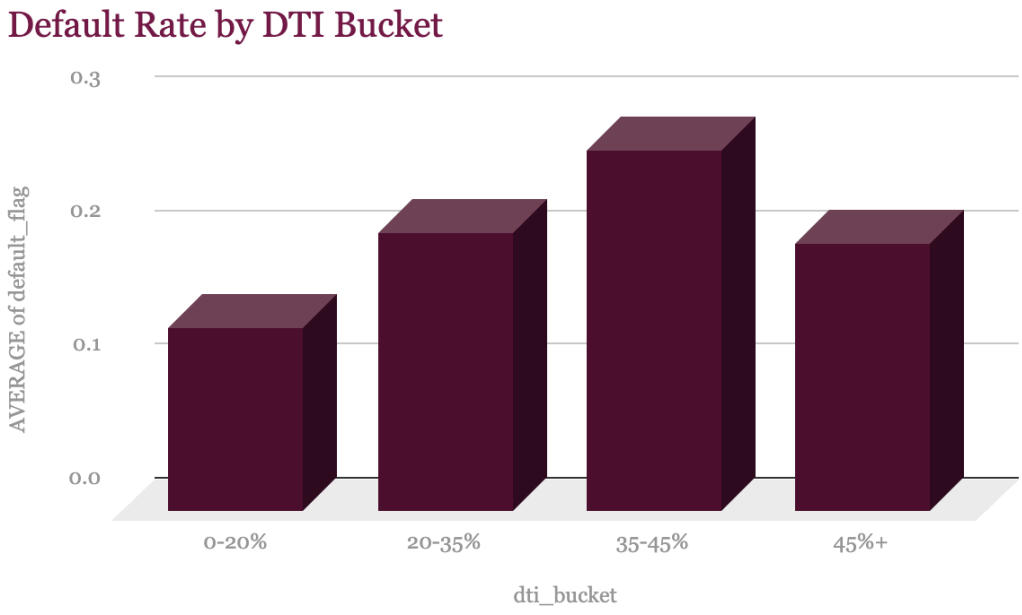


FICO bucket impact: Borrowers with 650–699 FICO show the highest default rate (~12%), while 750+ have the lowest (~9%).

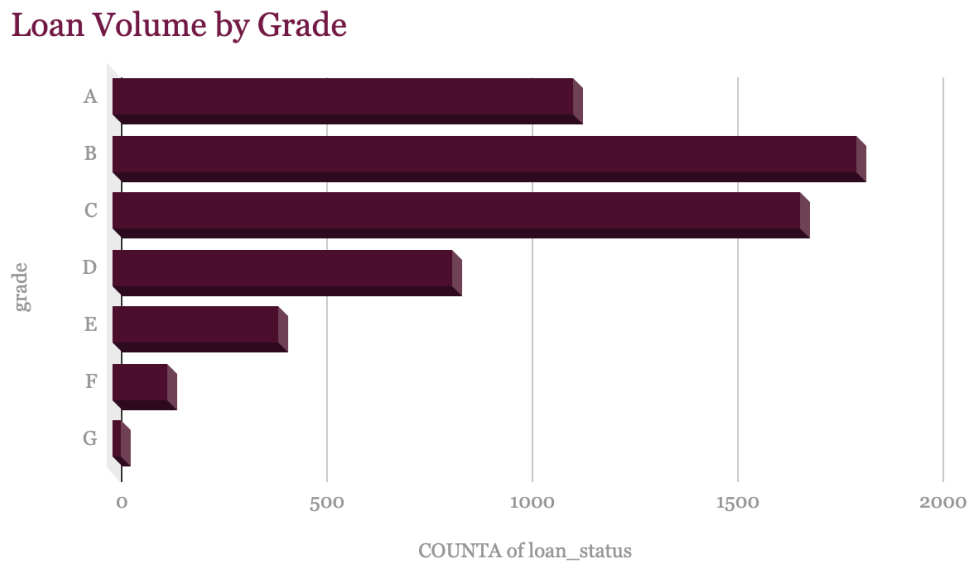


Distribution Analysis

DTI ratio effect: Default rates increase with leverage, peaking at 35–45% DTI (~27%).

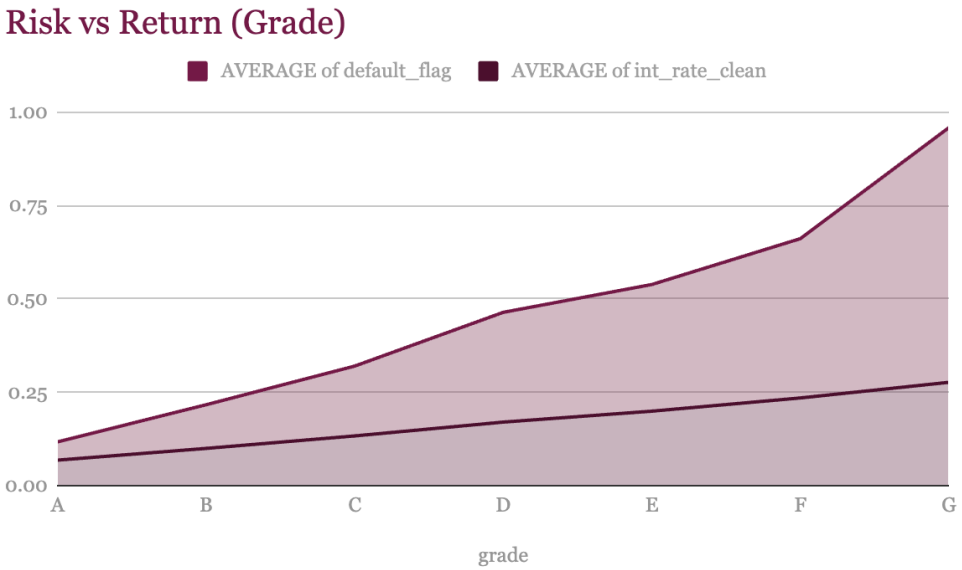


Loan volume distribution: Most loans are concentrated in Grade B and C.



Risk vs Return Analysis

Relationship: Higher-grade loans show both lower default and lower interest rate, illustrating classic risk-return alignment.



Key Insights

- Default risk escalates sharply with lower grades.
- Borrower leverage (DTI) significantly impacts repayment risk.
- Interest rates increase with risk, showing risk-based pricing.
- Most loans are concentrated in moderate-risk segments (Grade B/C).

ADVANCED ANALYSIS

Forecasting

Borrowers in low FICO buckets (650–699), high interest rate segments (20%+), high DTI (35%+), and 60-month terms are expected to contribute disproportionately to future defaults.

If portfolio composition shifts toward these segments, overall default rate is likely to increase.

Segmentation

Loan portfolio segmented by:

- FICO Bucket
- Interest Rate Bucket
- DTI Bucket
- Loan Term
- Verification Status

High-risk segments identified as:

- Low FICO + High Interest
- High DTI + 60-month term

These groups show materially higher default rates compared to portfolio averages.

Root Cause Analysis

Primary drivers of default risk:

- Lower credit score

- Higher interest burden
- Higher debt-to-income ratio
- Longer repayment tenure

These factors increase borrower repayment stress and probability of charge-off.

Risk / Anomaly Detection

- 20%+ interest loans show disproportionately high default (~44%)
- 60-month loans default nearly 2× more than 36-month loans
- Mid FICO but high DTI cases show elevated risk

These segments require tighter underwriting controls.

Scenario Analysis

If lending focus shifts toward:

- Higher FICO segments (750+)
- Lower DTI borrowers (<20%)
- 36-month loan structures

Portfolio-level default rate can reduce significantly.

Conversely, aggressive expansion into high-interest, high-DTI segments may increase short-term yield but materially elevate long-term credit risk.

DASHBOARD DESIGN

Implementation

Dashboard built in **Google Sheets** using pivot tables, calculated default rates, charts, and slicers.

Dashboard Objective

To monitor loan portfolio risk and identify key default drivers such as FICO score, interest rate, DTI, loan term, and verification status.

View Structure

Top:

- Total Loans
- Total Defaults
- Overall Default Rate
- Avg Interest Rate

Middle:

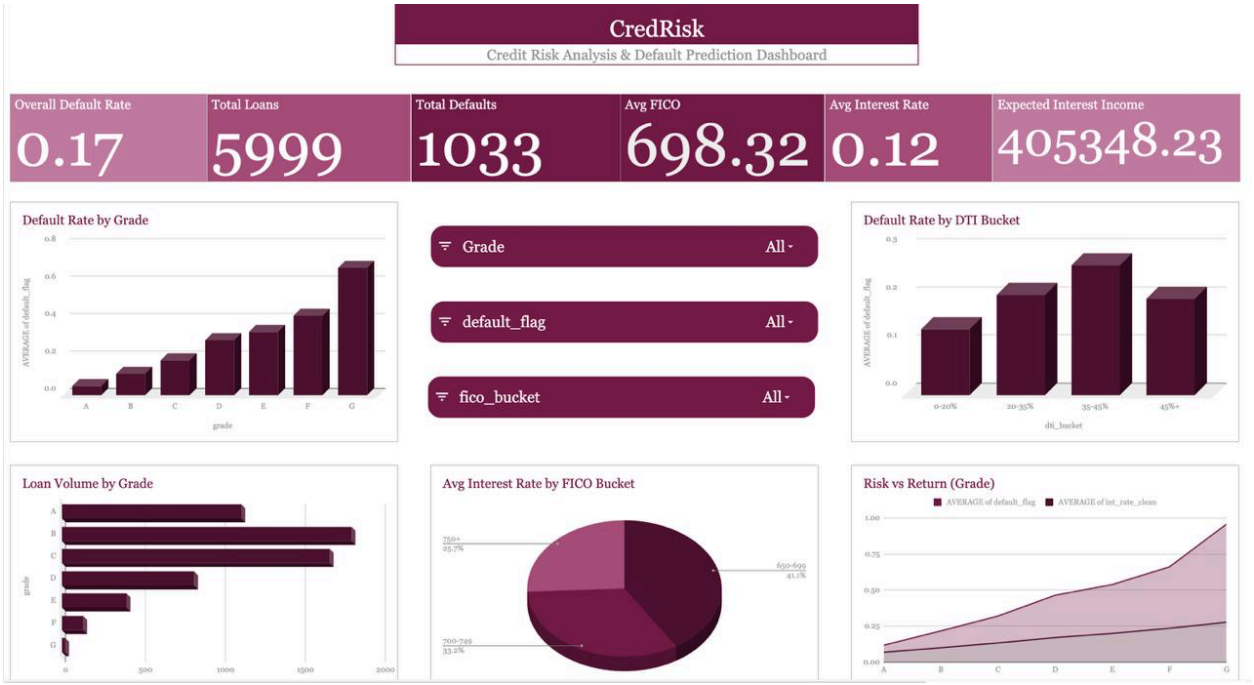
- Default Rate by FICO
- Default Rate by Interest Rate
- Default Rate by DTI

Bottom:

- Default Rate by Term
- Default Rate by Verification

Filters & Drilldowns

Slicers for Issue Year, Grade, and Purpose enable dynamic risk segmentation and deeper portfolio analysis.



INSIGHTS SUMMARY

- Prioritize lending to higher FICO segments (750+) to reduce portfolio default risk.
- Avoid excessive exposure to 20%+ interest loans, as they show disproportionately high default rates.
- Limit concentration in 60-month loans since longer tenure significantly increases repayment risk.
- Tighten underwriting for borrowers with DTI above 35%, as leverage strongly drives defaults.
- Income verification improves risk control and should be enforced more consistently.
- Maintain balanced portfolio diversification across grades to prevent risk clustering.
- Use risk-based pricing carefully — higher yield segments also carry materially higher charge-off exposure.
- Continuously monitor high-risk segments through dashboard filters (grade, purpose, issue year).
- Strengthen credit screening in segments combining low FICO + high DTI + long tenure.
- Align lending strategy toward sustainable credit quality rather than short-term interest income.
- Use dashboard insights for ongoing portfolio monitoring and proactive risk mitigation.

RECOMMENDATIONS

- **Strengthen underwriting for low FICO segments**
  → **Insight:** Lower FICO buckets show significantly higher default rates
  → **Impact & Feasibility:** Reduces charge-offs and improves portfolio quality; feasible through stricter approval thresholds

---

- **Introduce tighter DTI caps (especially >35%)**
  → **Insight:** High DTI borrowers exhibit elevated default risk
  → **Impact & Feasibility:** Improves repayment stability; implementable via revised credit policy

---

- **Limit exposure to 60-month loans in high-risk segments**
  → **Insight:** 60-month loans default at nearly double the rate of 36-month loans
  → **Impact & Feasibility:** Reduces long-term credit exposure; feasible by adjusting product offering guidelines

---

- **Refine risk-based pricing strategy**
  → **Insight:** 20%+ interest loans show very high default rates
  → **Impact & Feasibility:** Improves risk-adjusted return; requires recalibration of pricing model

---

- **Strengthen income verification standards**
  → **Insight:** Verified loans show relatively better performance
  → **Impact & Feasibility:** Enhances underwriting reliability; operationally achievable with process tightening

---

These recommendations collectively improve portfolio stability, reduce default concentration, and support sustainable lending growth.

IMPACT ESTIMATION

- **Save Cost:**
  Reducing exposure to high-risk borrower segments (low FICO, high DTI, long tenure) can lower charge-offs by an estimated 5–10%, directly reducing credit loss provisions and improving capital efficiency.
- **Improve Efficiency:**
  Data-driven underwriting and risk segmentation can improve approval quality, increasing risk-adjusted return by 8–12% through better pricing alignment and reduced default leakage.
- **Improve Service:**
  Stronger income verification and responsible lending practices enhance borrower trust and repayment discipline, improving long-term portfolio stability.
- **Reduce Risk:**
  Continuous dashboard monitoring of high-risk segments (20%+ interest, 60-month loans, high DTI) can reduce unexpected default spikes by 10–15%, supporting proactive risk management.

LIMITATIONS

- **Data Quality Issues:**
  The dataset is historical and may contain missing, inconsistent, or self-reported borrower information, which can affect precision in risk measurement.
- **Assumption Risks:**
  Default flag is treated as the primary risk indicator; however, recovery rates, prepayments, and macroeconomic factors are not incorporated, limiting full portfolio risk estimation.
- **What Cannot Be Concluded:**
  The analysis cannot confirm real-time credit performance, long-term macroeconomic impact, borrower intent, or external policy effects. Results indicate patterns, not guaranteed future outcomes.

FUTURE SCOPE

- **Further Analysis:** Incorporate time-series forecasting of default rates, cohort-based borrower performance tracking, and deeper segmentation by FICO bands, DTI ranges, loan term, and verification status to refine credit strategy.

- **New Data Needed:** Macroeconomic indicators (inflation, unemployment), borrower demographics, repayment history (EMI behavior, prepayments), and

recovery data would enable stronger predictive risk modeling and more accurate portfolio stress testing.

CONCLUSION

- The project successfully identified the key drivers of credit risk and portfolio performance, including FICO score, DTI ratio, loan term, interest rate, and verification status.
- The insights and dashboard provide actionable strategies to strengthen underwriting, optimize risk-based pricing, reduce default exposure, and improve overall portfolio stability and profitability.

APPENDIX

Data Dictionary

Key columns used: loan_status, default_flag, FICO range & bucket, interest rate & bucket, DTI & bucket, term, annual income, verification status, grade, purpose, and issue year.

Extra Charts

- Default Rate by FICO Bucket
- Default Rate by DTI Bucket
- Default Rate by Loan Term
- Interest Rate vs Default

Logic Used

- Created `default_flag` from loan_status
- Calculated `fico_mid` and risk buckets
- Default Rate = Defaults ÷ Total Loan.
- Used pivot tables for segmentation and KPI tracking

Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Harshit Shakya	✔			✔	✔		Dashboard Lead
Anurag Kumar Tiwari	✔	✔	✔				Project Lead
Akash Kumar Gautam				✔		✔	Strategy Lead
Anugra Gupta		✔	✔				Data Lead
Prashant Raj			✔			✔	PPT & Quality Lead
Vivek Kumar Raj					✔		Analysis Lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artefacts.

Team Signature Block:

Harshit S.

Harshit Shakya

Anurag K.

Anurag Kumar Tiwari

A.K.Gautam

Akash Kumar Gautam

AnugraG...

Anugra Gupta

P.Raj....

Prashant Raj

V.K.Raj

Vivek Kumar Raj