

# FINAL PROJECT REPORT

## Loan Default Risk Analytics

---

### Project Title

Predicting Loan Default Risk: A Data-Driven Credit Risk Analysis Framework

### Sector

Banking / Financial Risk Analytics

### Institute

Newton School of Technology, Rishihood University

---

### Team Members

- Preetish Ubhrani
  - Yashkumar Nimje
  - Sayooj S B
  - Pranav Singh
  - Kushal Tyagi
  - Vansh Khod
-

## 2. Executive Summary

### 2.1 Problem

Financial institutions incur significant financial losses due to loan defaults. Traditional credit evaluation methods often rely heavily on static indicators such as credit score and income, failing to capture deeper risk interactions across borrower profiles, loan types, and financial stress indicators like Loan-to-Value (LTV) and Debt-to-Income (DTI) ratios.

The absence of a centralized analytical dashboard limits the institution’s ability to monitor default behavior across segments, delaying identification of high-risk borrower groups.

---

### 2.2 Approach

We analyzed a dataset of **10,000 loan records** and developed a structured KPI-driven dashboard in Google Sheets to evaluate default risk across:

- Loan Type
- LTV Bands
- DTI Bands
- Credit Score
- Loan Category

Using pivot tables, calculated metrics, and risk segmentation techniques, we identified patterns that significantly influence default probability.

---

### 2.3 Key Metrics from Dashboard

| KPI                          | Value      |
|------------------------------|------------|
| Total Loans                  | 10,000     |
| Overall Default Rate         | 24.88%     |
| Median Interest Rate         | 3.99%      |
| Total Money Lost in Default  | 79,598,200 |
| Avg Credit Score (Defaulted) | 699.51     |

Avg Credit Score (Non-Defaulted) 702.03

---

## 2.4 Key Insights

- Nearly **1 in 4 loans (24.88%) defaulted**
  - DTI greater than 50% shows default rates above 40%
  - Business loans have significantly higher default risk than personal loans
  - LTV above 0.75 sharply increases default probability
  - Conventional loans dominate portfolio exposure (76.5%), creating concentration risk
- 

# 3. Sector & Business Context

## 3.1 Sector Overview

Credit Risk Analytics is a critical function within modern banking. Financial institutions use predictive risk modeling and segmentation frameworks to reduce non-performing assets (NPAs) and maintain portfolio health.

---

## 3.2 Current Challenges

- Rising consumer debt levels
  - High exposure in unsecured and high-leverage loans
  - Increasing Non-Performing Assets (NPAs)
  - Lack of granular borrower-level segmentation
- 

## 3.3 Why This Problem Was Chosen

Loan default risk directly impacts:

- Profitability
- Capital adequacy ratios
- Liquidity
- Financial stability

A data-driven segmentation dashboard enables proactive credit risk management instead of reactive recovery strategies.

---

## 4. Problem Statement & Objectives

### 4.1 Formal Problem Definition

“Which borrower financial characteristics most significantly influence loan default risk, and how can lenders reduce exposure using data-driven segmentation?”

---

### 4.2 Project Scope

**Dataset Size:** 10,000 loans

**Variables Analyzed:**

- Loan Type
- Loan Category
- LTV Ratio
- DTI Ratio
- Credit Score
- Interest Rate
- Default Status

**Tool Used:** Google Sheets (Data Cleaning and Dashboarding)

---

### 4.3 Success Criteria

- Identify high-risk DTI and LTV thresholds
  - Compare risk across loan categories
  - Quantify financial loss due to defaults
  - Build an interactive decision-support dashboard
-

# 5. Data Description

## 5.1 Dataset Source

Loan Default Dataset (Kaggle)

<https://www.kaggle.com/datasets/yasserh/loan-default-dataset>

---

## 5.2 Data Structure

Structured tabular dataset with 10,000 rows, 34 columns and multiple financial attributes.

---

## 5.3 Key Columns

- **ID:** Unique loan identifier
  - **Loan Amount:** Sanctioned principal amount
  - **Rate of Interest:** Applied interest rate (Median  $\approx$  3.99%)
  - **LTV (Loan-to-Value):** Loan amount divided by property value
  - **DTI (Debt-to-Income):** Percentage of income used for debt repayment
  - **Status:** Target variable (0 = Repaid, 1 = Default)
  - **Loan Category:** Business vs Personal
- 

## 5.4 Data Limitations

- Single-year dataset (2019 vintage)
  - Highly skewed financial distributions
  - No macroeconomic variables
-

## 6. Data Cleaning & Preparation

### 6.1 Missing Value Treatment

- Mode imputation for categorical features
  - Median imputation for skewed financial features
  - Grouped median imputation for Annual Income and Interest Rate
- 

### 6.2 Data Standardization

- Converted coded variables (cf/ncf → Conforming/Non-Conforming)
  - Standardized binary indicators
  - Converted text-based numeric values into analyzable formats
- 

### 6.3 Feature Engineering

- Created LTV Bands (0–0.25, 0.25–0.5, etc.)
  - Created DTI Buckets (0–9, 10–19, etc.)
  - Developed borrower risk segmentation logic
- 

## 7. KPI & Metric Framework

### 7.1 Overall Default Rate (24.88%)

Formula:

`Defaulted Loans / Total Loans`

Primary health indicator of portfolio quality.

---

### 7.2 Total Money Lost (79,598,200)

Formula:

`SUM(Loan Amount where Status = Default)`

Measures real financial exposure.

---

## 7.3 Credit Score Gap (2.52 Points)

702.03 – 699.51

Minimal difference indicates weak predictive strength.

---

## 7.4 Default Rate by Loan Category

- Business: 33.03%
- Personal: 23.55%

Business loans significantly underperform.

---

# 8. Exploratory Data Analysis (EDA)

## 8.1 LTV Sensitivity

Default remains stable below 0.75 LTV but spikes sharply near 1.0.  
High leverage significantly increases default probability.

---

## 8.2 DTI Stress Testing

DTI 50–59% → 42.40% default rate.  
This represents the borrower affordability breaking point.

---

## 8.3 Loan Type Mix

- Conventional: 76.5%
- Government: 14.1%
- Special: 9.4%

Portfolio concentration risk is high.

---

## 8.4 Correlation Insights

- LTV → Strong Positive correlation with Default
  - DTI → Strong Positive relationship
  - Credit Score → Weak relationship
- 

# 9. Advanced Analysis

## Risk Segmentation Matrix

### Segment A – Critical Risk

DTI > 50%  
LTV > 0.75

### Segment B – Moderate Risk

DTI 30–49%  
Medium LTV

### Segment C – Low Risk

DTI < 30%  
LTV < 0.5

---

## Financial Exposure Insight

Loss concentration exceeds 79 million, primarily clustered in high-DTI segments.

---

# 10. Dashboard Design

## Objective

To create a centralized “Risk Control Tower” for loan officers and credit managers.

---

## Structure

### Top Section – KPI Cards

- Total Loans
- Default Rate
- Median Interest Rate
- Total Financial Loss

### Middle Section – Risk Drivers

- Line Chart → Default by LTV
- Bar Chart → Default by DTI
- Donut Chart → Loan Type Distribution

### Bottom Section – Drilldowns

- LTV split by Loan Type
  - DTI split by Loan Type
  - Business vs Personal comparison
-

## 11. Insights Summary

1. DTI above 50% doubles default risk.
  2. LTV above 75% sharply increases exposure.
  3. Business loans carry structural risk.
  4. Portfolio is heavily concentrated in Conventional loans.
  5. Credit score alone is unreliable for this dataset.
  6. Risk clustering occurs at High DTI + High LTV intersection.
- 

## 12. Recommendations

1. Tighten DTI approval threshold above 50%.
  2. Cap LTV exposure above 75%.
  3. Introduce risk premium for Business loans.
  4. Diversify portfolio composition.
  5. Implement risk-based pricing using combined LTV + DTI scoring.
- 

## 13. Impact Estimation

Reducing high-DTI exposure by 30% could:

- Reduce default exposure by 10–15%
  - Save approximately 8–12 million in projected losses
- 

## 14. Limitations

- Single-year dataset
  - No employment/industry variables
  - No macroeconomic indicators
  - Correlation does not imply causation
-

## 15. Future Scope

- Logistic Regression modeling
  - Probability of Default (PD) framework
  - Machine Learning scoring
  - Early Warning delinquency detection system
- 

## 16. Conclusion

This analysis demonstrates:

- Debt burden (DTI) is the strongest predictor of default.
- Leverage (LTV) amplifies risk severity.
- Portfolio concentration increases systemic exposure.

Transitioning from static approval rules to dynamic segmentation-based underwriting can significantly improve portfolio health and reduce financial losses.

---

## 17. Appendix

### Selected Data Dictionary

- loan\_limit → Regulatory classification
  - approve\_in\_adv → Pre-approval indicator
  - loan\_purpose → Purchase / Refinance
  - business\_or\_commercial → Business / Personal
  - rate\_of\_interest → Interest band ~3.5–4.0%
  - loan\_to\_value\_ratio → Primary leverage metric
  - debt\_to\_income\_ratio → Affordability stress indicator
  - is\_loan\_default → Target variable
- 

## 18. Contribution Matrix

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

| Team Member      | Dataset Sourcing | Cleaning | KPI & Analysis | Dashboard | Report Writing | PPT    | Overall Role       |
|------------------|------------------|----------|----------------|-----------|----------------|--------|--------------------|
| Preetish Ubhrani | medium           | medium   | low            | low       | medium         | low    | Project Lead       |
| Yashkumar Nimje  | low              | medium   | medium         | low       | medium         | low    | Data Lead          |
| Sayooj S B       | low              | low      | medium         | medium    | low            | medium | Dashboard Lead     |
| Pranav Singh     | medium           | low      | low            | medium    | medium         | low    | Strategy Lead      |
| Kushal Tyagi     | low              | medium   | low            | medium    | low            | medium | Analysis Lead      |
| Vansh Khod       | medium           | low      | medium         | low       | low            | medium | PPT & Quality Lead |