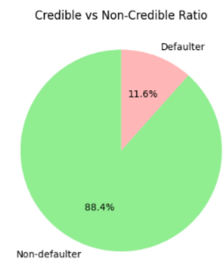
Credit Risk Management Model Report

1. Introduction

With the rapid growth of fintech platforms, individuals can now access microloans quickly and with minimal documentation. While this improves financial inclusion, it also increases the risk of loan defaults due to limited borrower verification. To mitigate this, fintech lenders must develop credit risk management models that distinguish between defaulters and non-defaulters effectively. This ensures responsible lending, minimizes financial losses, and allows risk-based interest rate pricing.

Dataset Summary

1. Total Records: 2,55,347 records
2. Default class distributions:
    - Default records: 2,25,694
    - Non-Default records: 29,653
3. Predictor Features Table



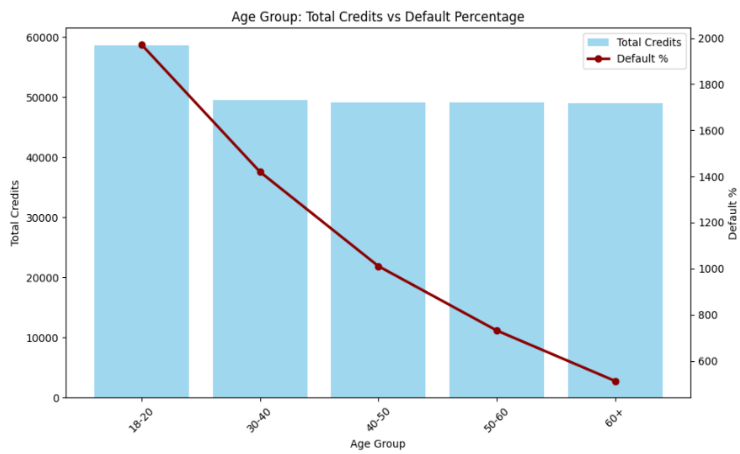
Feature	Description
MonthsEmployed	Number of months the borrower has been employed
LoanTerm	Duration of the loan (in months)
DTIRatio	Debt-to-Income ratio
Education	Education level (High School, Bachelor’s, Master’s, PhD)
EmploymentType	Type of employment (Full-time, Part-time, Self-employed, Unemployed)
MaritalStatus	Marital status (Single, Married, Divorced)
HasMortgage	Whether the borrower has a mortgage (Yes/No)
LoanPurpose	Purpose of the loan (Auto, Business, Other, etc.)
HasCoSigner	Whether the borrower has a cosigner (Yes/No)

2. Methodology

2.1 Exploratory Data Analysis (EDA)

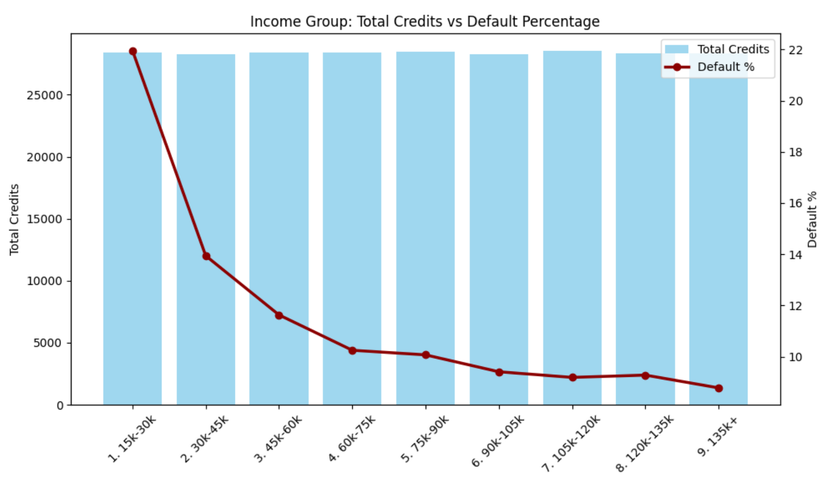
We analyzed borrower characteristics against default percentages/default ratios using grouping rather than raw correlations. Below are the insights with visual evidence:

1. Age Group vs Default Percentage



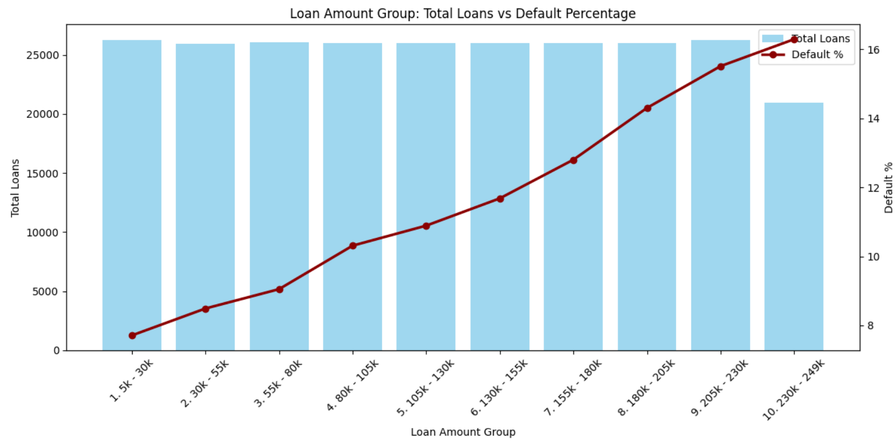
- Younger borrowers (18–20) have the highest default ratio.
- Default percentage decreases steadily with age, showing older borrowers are financially more responsible.

2. Income Group vs Default Percentage



- Borrowers with low income (15k–30k) default the most (~22%).
- As income increases, default ratio falls sharply, stabilizing around 9–10% for higher income groups.

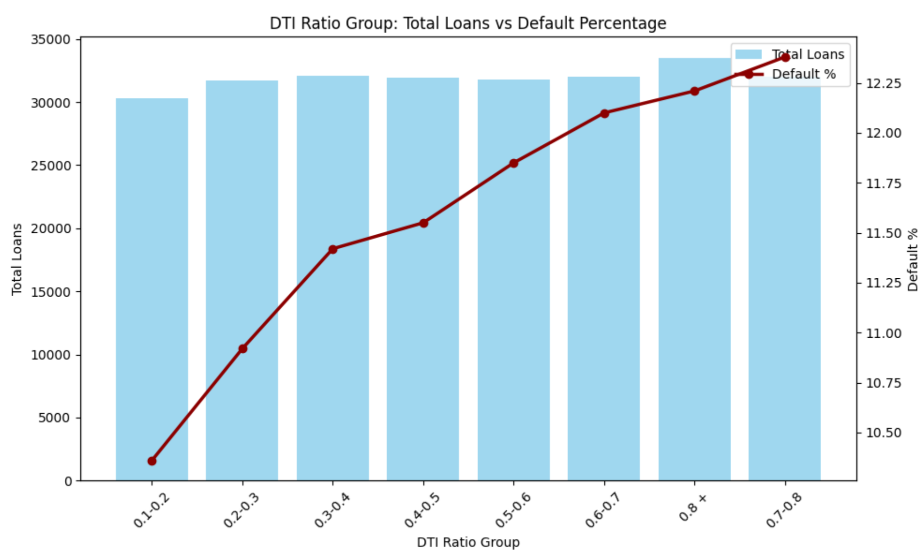
3. Loan Amount vs Default Percentage



- Smaller loans (5k–30k) show lower default percentages.
- As the loan amount increases, the default ratio rises, peaking at high-value loans (>200k).

However, Income and Loan Amount cannot individually predict, since high income indicates a low Default Ratio and high Loan Amount indicates a high Default Ratio. Thus, we'll use the ratio, Debt-to-Income Ratio.

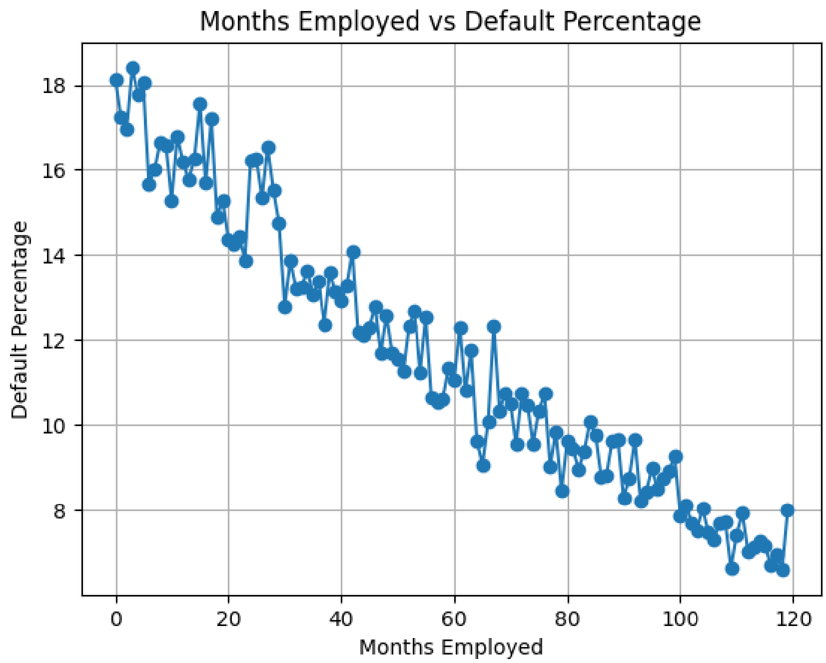
4. Debt-to-Income (DTI) Ratio vs Default Percentage



- A clear positive correlation: higher DTI ratios correspond to higher default percentages.
- Borrowers with DTI > 0.8 show the highest default probability.

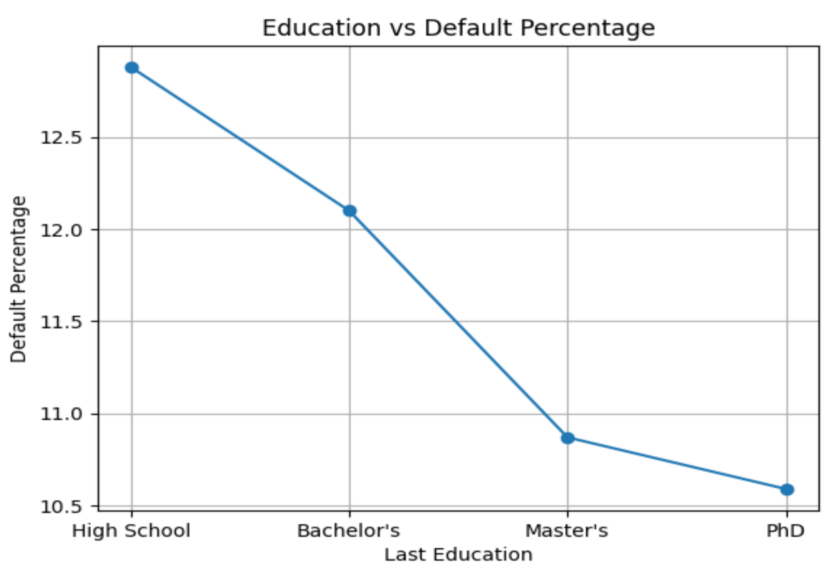
This indicates that loan amount and income should not be evaluated in isolation; instead, they must be considered together through the Debt-to-Income (DTI) ratio to accurately capture default risk.

5. Months Employed vs Default Percentage



- Default ratio declines steadily with employment tenure.
- Borrowers employed for longer than 5 years show much lower risk compared to new employees. This indicates that people with higher experience tend to be more financially stable. So, preference for high work experience can be given.

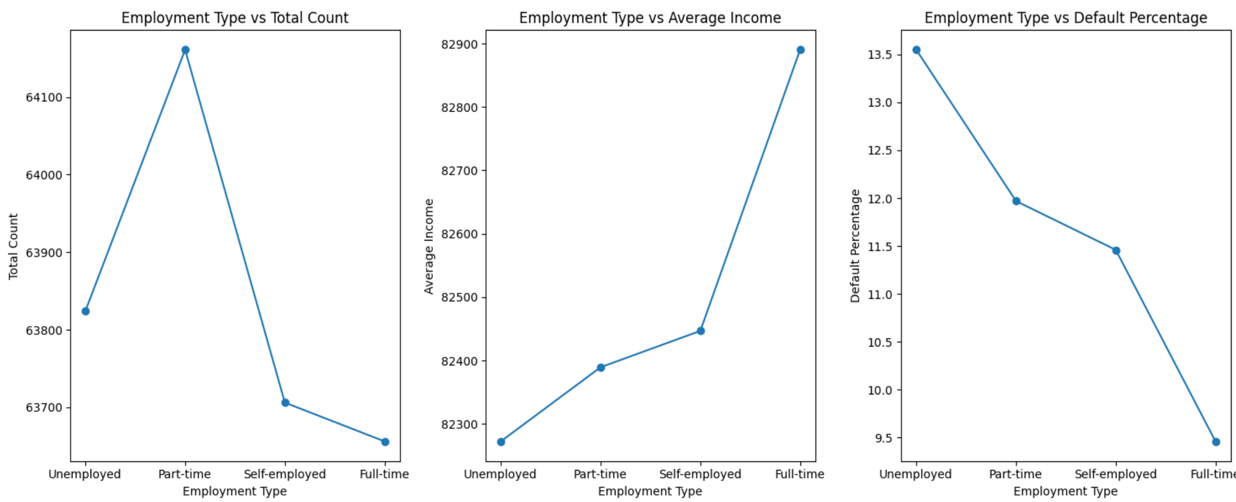
6. Education vs Default Percentage



- Borrowers with only a high school education have the highest default ratio (~12.8%).

- Default rate decreases with higher education levels (Bachelor’s → Master’s → PhD).

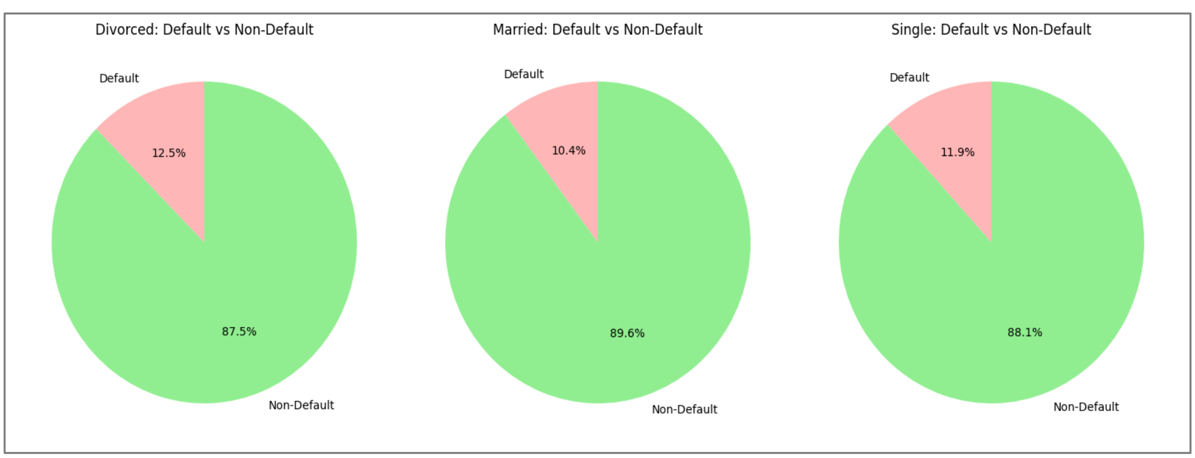
7. Employment Type vs Default Percentage

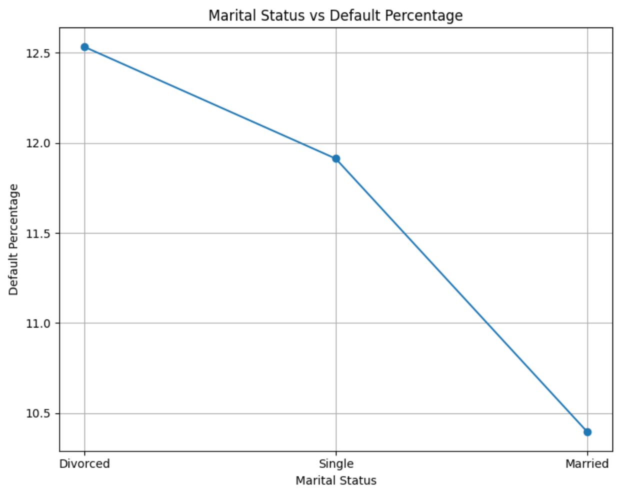


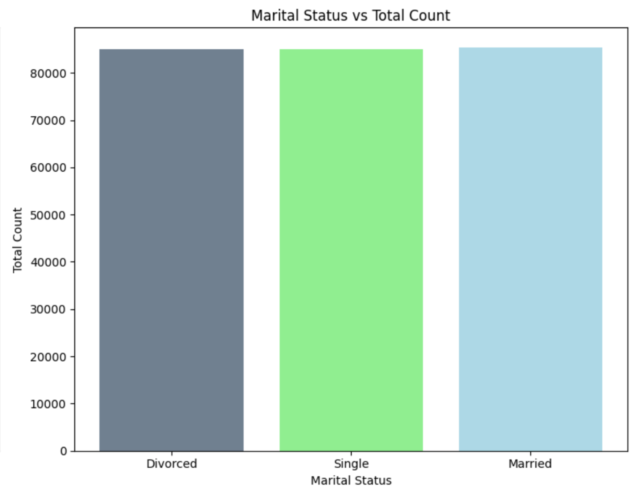
- Unemployed borrowers have the highest default percentage (~13.5%).
- Full-time employees are the lowest risk (~9.5%).
- Self-employed and part-time fall in the middle. With the income chart, we see that the self-employed have a higher average salary, indicating they are more reliable and hence have a lower default percentage than part-time.

However, Unemployed & Part-time employed individuals contribute to the majority of the potential customers. Losing them would mean losing business. So we could provide them with credit with a high-risk premium (higher interest rates %).

8. Marital Status vs Default Percentage

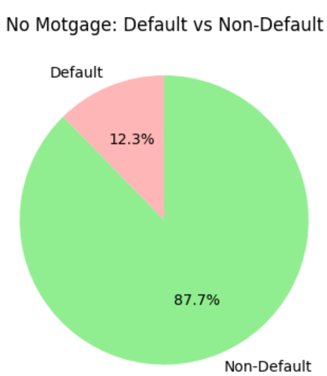


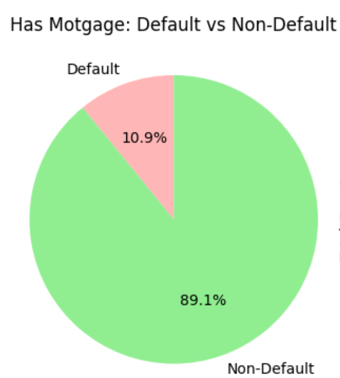


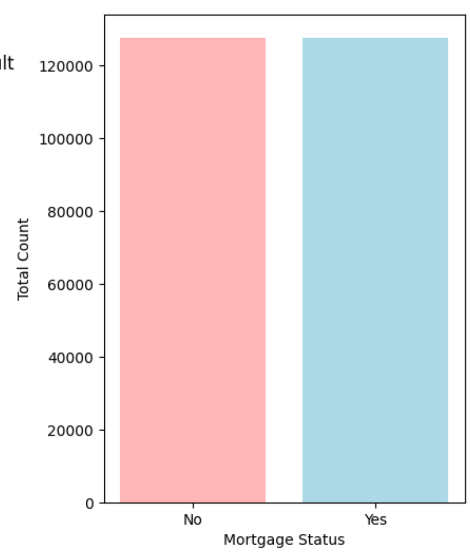


- Married borrowers show the lowest default ratio (10.4%).
- Single borrowers are slightly riskier (~11.9%).
- Divorced borrowers show the highest risk (~12.5%).
- The default Rates are lowest for stably married people, and highest for unstable families going through tough times like divorce.

9. Mortgage Status vs Default Percentage







- Borrowers with mortgages default less (~10.9%) than those without (~12.3%).
- Indicates that having financial obligations (like a mortgage) may signal greater responsibility.

Summary of EDA Findings

10. Default probability is higher for young, low-income, unemployed, less educated, divorced borrowers without mortgages.
11. Financial stability indicators (higher income, longer employment, mortgage, cosigner) reduce default risk.
12. DTI ratio and loan amount are strong financial predictors.

2.2 Feature Engineering

13. **Categorical Encoding**: Converted variables (education, cosigner, mortgage, employment type, marital status, loan purpose) into numerical form.

Example:

```
# mapping done in accordance to EDA
maping_of_education={'High School':0,"Bachelor's":1,"Master's":2,'PhD':3}
df['Education']=df['Education'].map(maping_of_education)
df.head()
```

14. **Imbalance Handling**: Managed imbalanced default/non-default data using **class weighting/resampling**.

Example:

```
from imblearn.over_sampling import SMOTE


X = df.drop('Default', axis=1)
y = df['Default']

smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X, y)
```

15. **Final Feature Set**: Combined demographic, financial, and employment-related features for modeling.

---

3. Model Development

3.1 Data Preparation

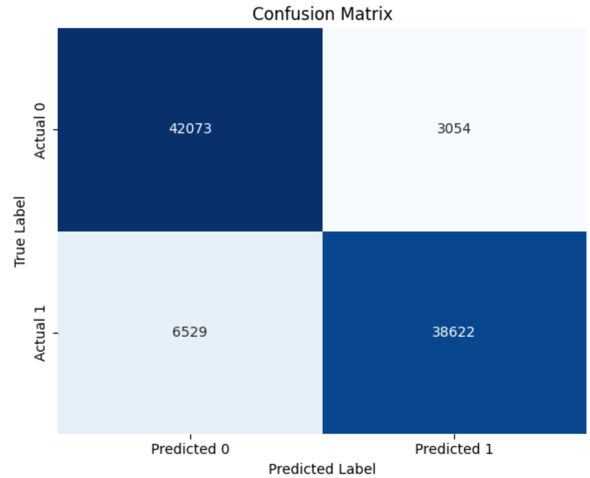
16. **Train-Test Split**: 80:20 ratio.
17. **Target Variable**: Default (1) vs Non-default (0).
18. **Predictors**: Engineered categorical + numerical features.

3.2 Random Forest Classifier

19. Ensemble model that builds multiple decision trees and aggregates predictions.
20. Reduces overfitting and handles both categorical and numerical data.
21. Provides feature importance to interpret predictors.

4. Model Performance

Metric	Non-Default 0	Default 1	Overall
Precision	0.87	0.93	-
Recall	0.93	0.86	-
F1-score	0.90	0.89	-
Accuracy	-	-	**0.89**



22. The model achieved **89% accuracy**.
23. Balanced precision and recall → model correctly identifies both defaulters and non-defaulters.

---

5. Application – Probability-Based Risk Pricing

The model outputs **default probabilities**, enabling **variable interest rates**:

24. Low-risk borrowers → lower interest rate.
25. High-risk borrowers → higher interest rate or loan rejection.

This ensures a **sustainable lending strategy** for fintechs, balancing profitability and financial inclusion.

Links

GitHub link: https://github.com/lakshit2508/Credit-Risk-Management-and-Client-analysis/tree/main
LinkedIn link: https://www.linkedin.com/in/lakshit-gupta-1b1283294/
Dataset link: https://www.kaggle.com/datasets/nikhil1e9/loan-default/data