

Credit Risk & Loan Performance Analysis

Data-Driven Insights for Smarter Lending Decisions

Shubham Kadtan

Tools Used: Excel | SQL (MySQL)

Executive Summary

- This project analyzes a financial loan dataset of over **38,000 customer records** to uncover patterns of **loan delinquency and profitability**.
- Using **SQL and Excel**, we examined key factors such as **loan term, annual income**, and **employment length** to identify which borrower segments are **high-risk** or **most profitable**.
- The insights enable financial institutions to:
 - Reduce non-performing assets (NPAs)
 - Strengthen lending policies
 - Improve return on investment through better risk segmentation

Problem Statement

- Financial institutions face significant losses due to **loan defaults** and poor **credit risk evaluation**.
- Understanding which borrower segments are more likely to **default** and which are **profitable** is critical for reducing **non-performing assets (NPAs)** and maximizing return on investment.
- This project addresses this challenge by analyzing historical loan data to detect risk patterns and improve lending decisions.

Project Goals

- Identify borrower segments with **high default rates**
- Analyze the relationship between **loan term**, **income**, and **employment** with default risk
- Calculate **interest earned** to find profitable segments
- Provide **data-driven recommendations** to reduce risk and increase profitability

Dataset Overview

| Feature | Description |
|---------------|---|
| loan_amount | Amount disbursed to borrower |
| loan_status | Loan status (Fully Paid, Charged Off, etc.) |
| annual_income | Borrower's annual income |
| emp_length | Employment length (converted to numeric) |
| term_months | Loan term (in months) |
| dti | Debt-to-income ratio |
| int_rate | Interest rate applied |
| Total Rows: | 38,576 |

Methodology

- Step 1: Data Exploration**

Reviewed 38,573 records and 26 columns to understand schema and missing values.

- Step 2: Data Cleaning**

- Removed/replaced nulls (e.g., emp_title, emp_length)
- Standardized fields like term, int_rate, and emp_length
- Cleaned columns for SQL compatibility (renamed, formatted)

- Step 3: Data Import**

Loaded cleaned dataset into **MySQL** for structured querying.

- Step 4: Data Analysis**

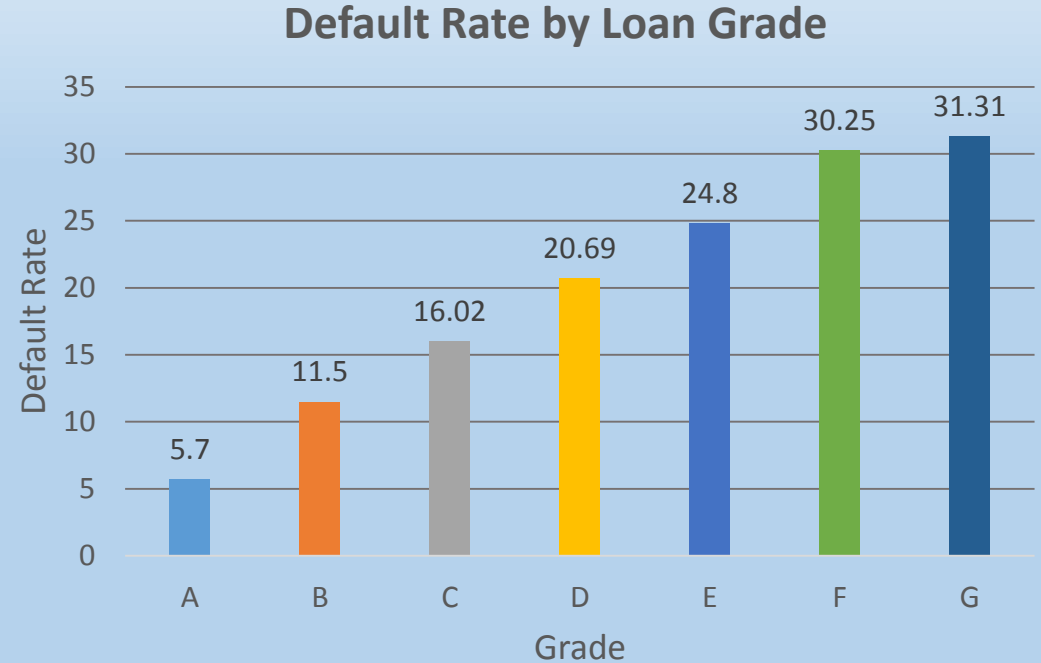
Wrote SQL queries to analyze default rates by multiple borrower attributes.

- Step 5: Visualization**

Created charts in **Excel** based on SQL outputs to highlight insights.

Which credit grades have the highest loan default rates?

```
SELECT grade,  
       count(*) AS total_loans,  
       sum(CASE  
           WHEN loan_status = "charged off" THEN 1  
           ELSE 0  
       END) AS defaulted_loans,  
       round(100* sum(CASE  
           WHEN loan_status = "charged off" THEN 1  
           ELSE 0  
       END) / count(*), 2) AS default_rate  
FROM loan_data  
GROUP BY grade  
ORDER BY default_rate;
```



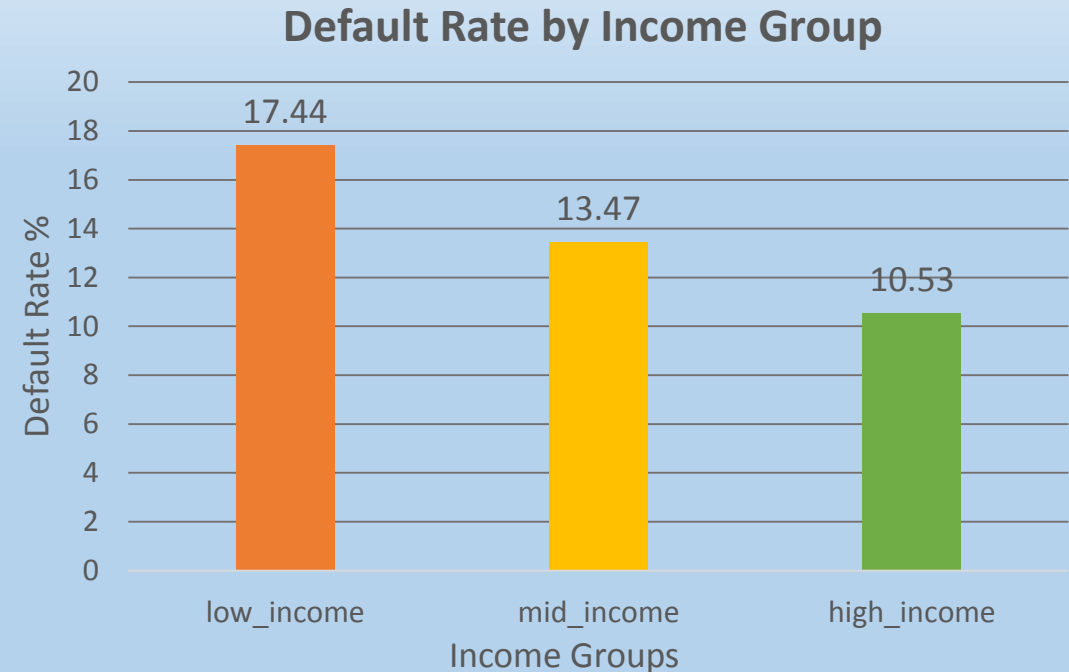
Default risk increases significantly as credit grade drops.

Borrowers in grades E, F, and G show **default rates above 24%**, with **Grade G peaking at 31.31%**, compared to just **5.7% for Grade A**.

These low-grade segments should be **flagged as high risk** and priced accordingly to offset potential losses.

How does income level impact loan default rates?

```
SELECT CASE
    WHEN annual_income < 30000 THEN "low_income"
    WHEN annual_income BETWEEN 30000 AND 70000 THEN "mid_income"
    ELSE "high_income"
END AS income_groups,
count(*) AS total_loans,
sum(CASE
    WHEN loan_status = "charged off" THEN 1
    ELSE 0
END) AS default_loans,
round(100 * sum(CASE
    WHEN loan_status = "charged off" THEN 1
    ELSE 0
END) / count(*), 2) AS default_rate,
round(avg(total_payment - loan_amount), 2) AS int_earned
FROM loan_data
GROUP BY income_groups;
```



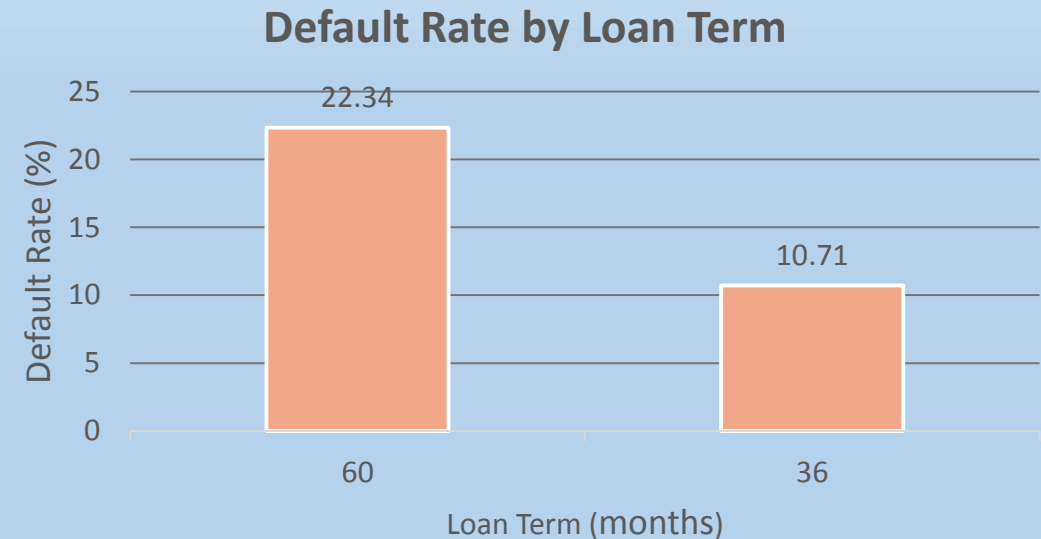
Borrowers in the **low-income group** have the **highest default rate at 17.44%**, compared to **13.47% for mid-income** and **10.53% for high-income** groups.

This suggests that **income level is inversely related to default risk** — lower-income borrowers are more likely to default.

Lenders should apply **stricter screening or offer smaller loan amounts** for low-income applicants to mitigate risk.

Are longer loan terms riskier in terms of default?

```
SELECT term_months,  
       count(*) AS total_loans,  
       sum(CASE  
           WHEN loan_status = "charged off" THEN 1  
           ELSE 0  
       END) AS default_loans,  
       round(100 * sum(CASE  
           WHEN loan_status = "charged off" THEN 1  
           ELSE 0  
       END) / count(*), 2) AS default_rate,  
       round(avg(total_payment - loan_amount), 2) AS int_earned  
FROM loan_data  
GROUP BY term_months;
```

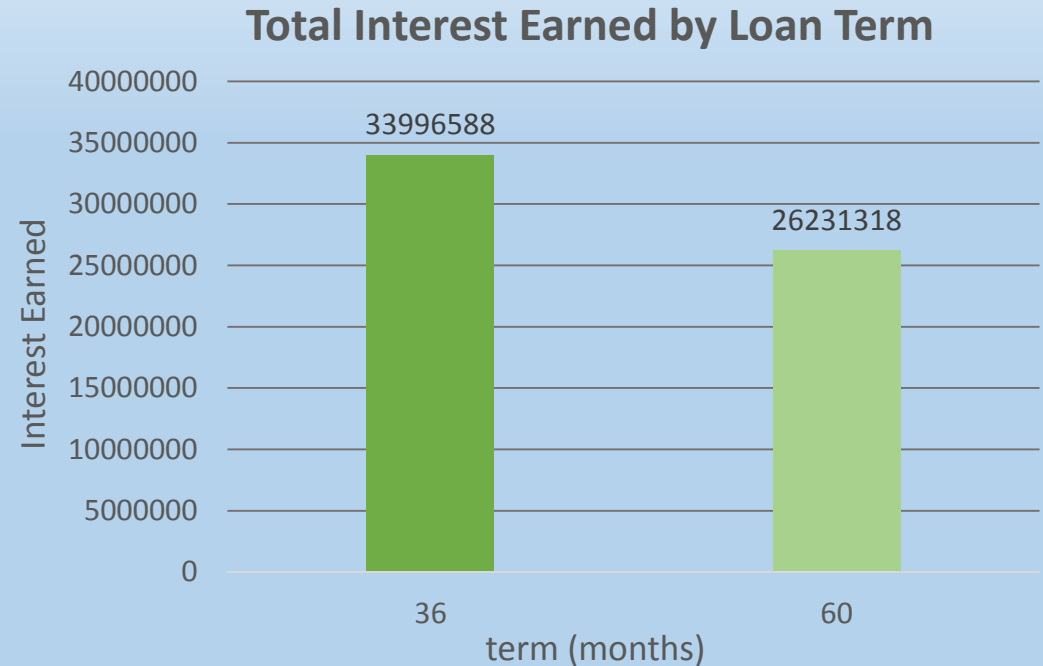


Loans with a **60-month term** have a **significantly higher default rate** compared to 36-month loans. This indicates that longer repayment periods are **riskier for lenders**, possibly due to increased uncertainty over time.

Lenders should apply **stricter credit checks** or **limit high-term loans to low-risk profiles**.

Which loan term generates more interest for the lender?

```
select
  term_months,
  SUM(loan_amount) AS total_loan_disbursed,
  sum(total_payment - loan_amount) as total_interest_earned
from loan_data
where loan_status = "fully paid"
group by term_months
order by term_months;
```

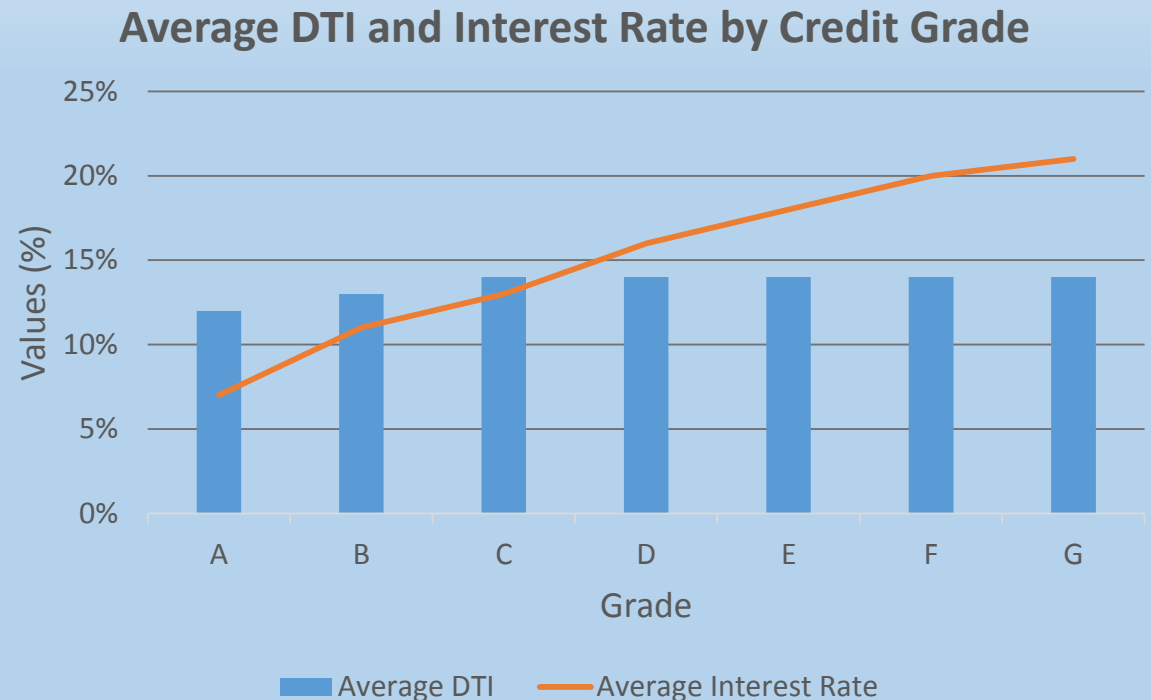


Although **60-month loans** have a **higher default risk**, they generate **significantly more interest** than 36-month loans.

This makes them **more profitable** if properly risk-managed, suggesting lenders should balance **return with risk** using stricter eligibility filters for longer-term loans.

Do interest rates and DTI vary across credit grades?

```
SELECT grade,  
       round(avg(dti), 2) AS average_dti,  
       round(avg(int_rate), 2) AS average_int_rate  
FROM loan_data  
GROUP BY grade  
ORDER BY average_int_rate;
```

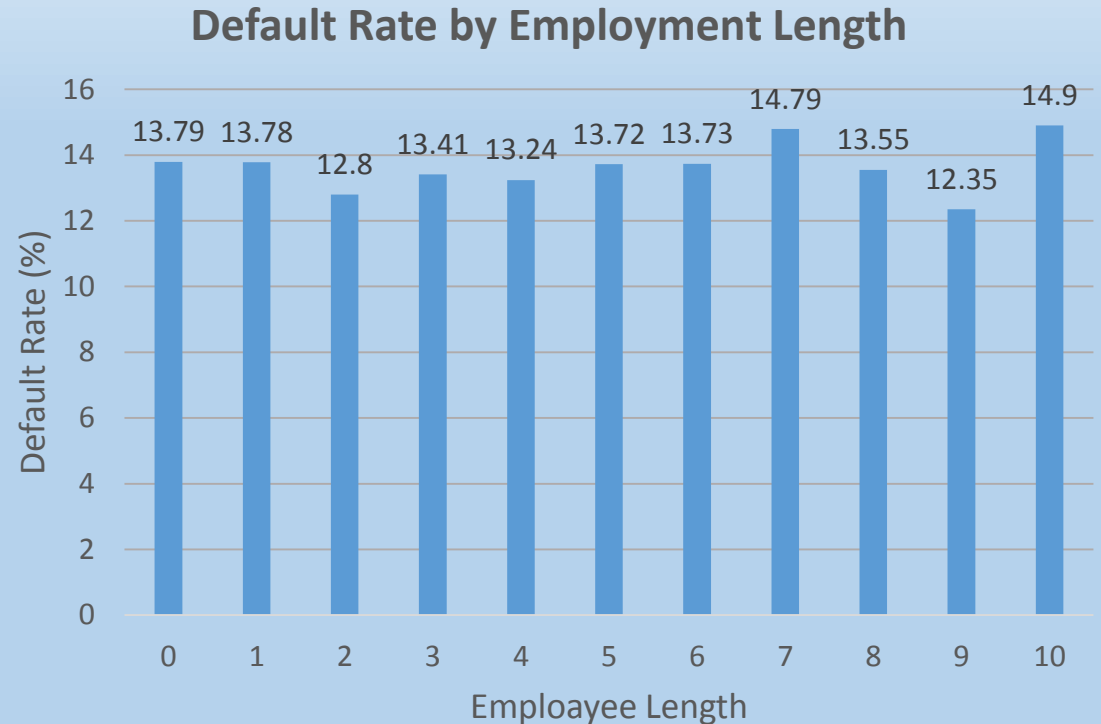


As credit grade worsens from A to G, the **interest rate increases sharply**, while **DTI (Debt-to-Income ratio)** stays relatively stable between **12%–14%**.

This indicates that **interest rates are strongly tied to credit grade risk**, not just borrower financial load. Lenders appear to **price risk effectively**, charging higher rates for lower grades regardless of DTI.

Does employment length affect loan default risk?

```
SELECT emp_length_years,  
       count(*) AS total_loans,  
       sum(CASE  
           WHEN loan_status = "charged off" THEN 1  
           ELSE 0  
       END) AS default_loans,  
       round(100 * sum(CASE  
           WHEN loan_status = "charged off" THEN 1  
           ELSE 0  
       END) / count(*), 2) AS default_rate,  
       round(avg(total_payment - loan_amount), 2) AS int_earned  
FROM loan_data  
GROUP BY emp_length_years  
ORDER BY default_rate;
```



The default rate remains **fairly consistent across all employment lengths**, ranging between **12.8% and 13.8%**, with no strong trend.

This suggests that **employment length alone is not a significant predictor of loan default**, and should be considered alongside other factors like income or credit grade.

Insights Summary

- **Default risk increases significantly from Grade A (5.7%) to Grade G (31.3%)** — lower-grade loans carry the highest risk.
- **Low-income borrowers have a 17.4% default rate**, much higher than high-income borrowers (10.5%).
- **60-month loans** have nearly **3× higher default rates** than 36-month loans, showing longer terms carry more risk.
- Despite the risk, **60-month loans generate more total interest**, making them more profitable if filtered properly.
- **Interest rates rise with lower grades**, but **DTI remains stable**, showing lenders already apply risk-based pricing.
- **Employment length has minimal effect** on default rates — not a strong standalone predictor.

Recommendations

- Avoid or adjust pricing for low-grade loans (E, F, G) due to high default rates above 24%. These should be offered only to **well-screened applicants** or at **higher interest rates**.
- Apply stricter screening to low-income borrowers, as they show a **default rate of 17.4%** — higher than mid- and high-income groups.
- Limit or risk-adjust 60-month loan offerings, since they carry **nearly 3× higher default risk** compared to 36-month terms.
- Offer **long-term loans only to low-risk borrowers**, as they generate **more interest** and can be **highly profitable** when defaults are low.

Thank
You