

Lending Club Assignment

Agenda

- Problem Statement
- Data Understanding
- Data Cleaning
- Data Analysis
- Insights
- Recommendations

Problem Statement

- Lending companies face 2 key risks:
 1. Ejecting good borrowers → lost opportunity.
 2. Accepting risky borrowers → financial loss.
- Goal: Perform EDA on loan dataset to identify patterns in borrower profiles and loan defaults.

Data Understanding

- Key Variables:
- Loan Details → loan_amnt, term, int_rate, installment
- Applicant Details → annual_inc, emp_length, home_ownership
- Target Variable → loan_status
- Note: Focus on **Fully Paid** vs **Charged Off**.
- Dataset Shape: {rows, columns}

Data Cleaning

- Removed % from **int_rate** and converted to numeric (int_rate_clean).
- Removed the word *months* from **term** and converted to numeric (term_clean).
- Checked for missing values across all columns and highlighted the top 10 columns with the most missing data.
- Counted and reported duplicate rows.

Data Analysis – Univariate

1. Loan Amount Distribution

1. Most loans between **5k–15k**.
2. Few borrowers take >25k → riskier segment.

2. Annual Income Distribution

1. Skewed: most borrowers earn <**150k**.
2. Some very high incomes (outliers) but not always safe.

3. Loan Term Distribution

1. **36 months** most common → lower risk.
2. **60 months** less common, usually larger and riskier.

4. Loan Status Distribution

1. Majority loans are **Fully Paid or Charged Off**.
2. Useful for identifying default patterns.

5. Loan Purpose Distribution

1. **Debt Consolidation** dominates.
2. Credit Card and “Other” purposes riskier.

Data Analysis – Bivariate

1. Interest Rate by Grade:

- Clear trend: **Low grades (E–G)** → high interest, higher risk.
- **A–C grades** → safer, lower interest rates.

2. Loan Amount by Term:

- **60-month loans** slightly higher amounts.
- Larger loans spread over longer terms = more risky.

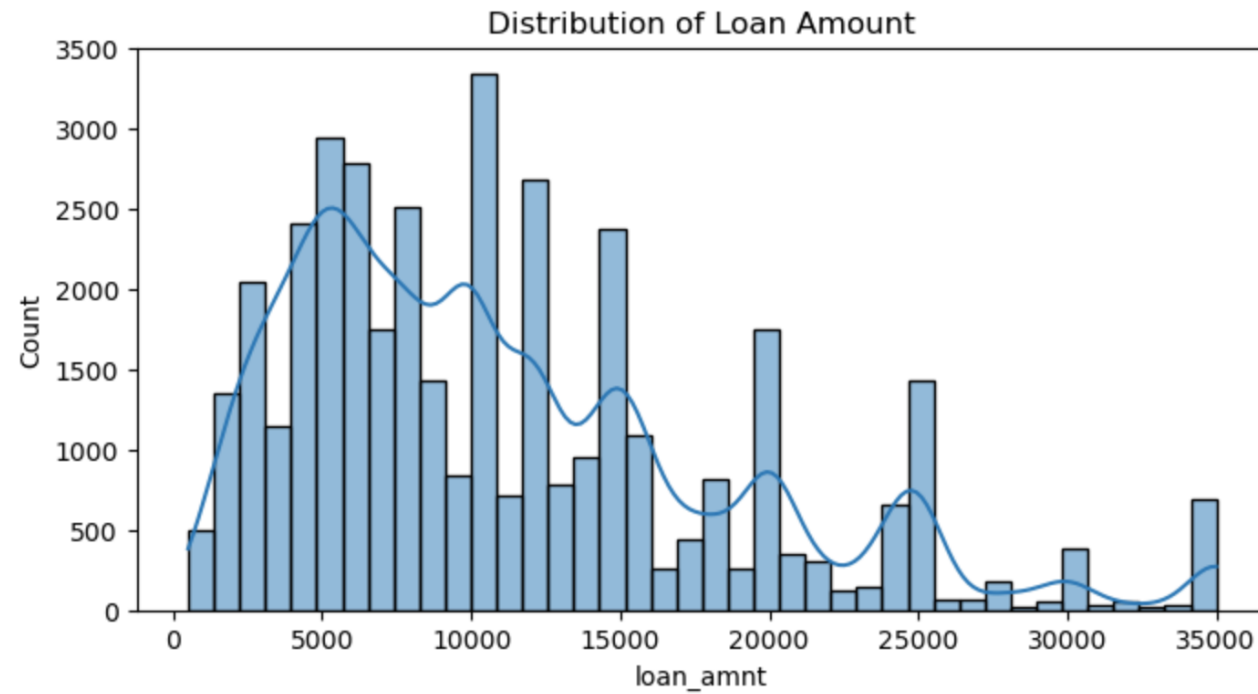
3. Annual Income vs Loan Amount (by Loan Status)

- Higher income borrowers tend to take larger loans.
- But defaults occur across income levels → income alone not enough to judge safety.

Insights

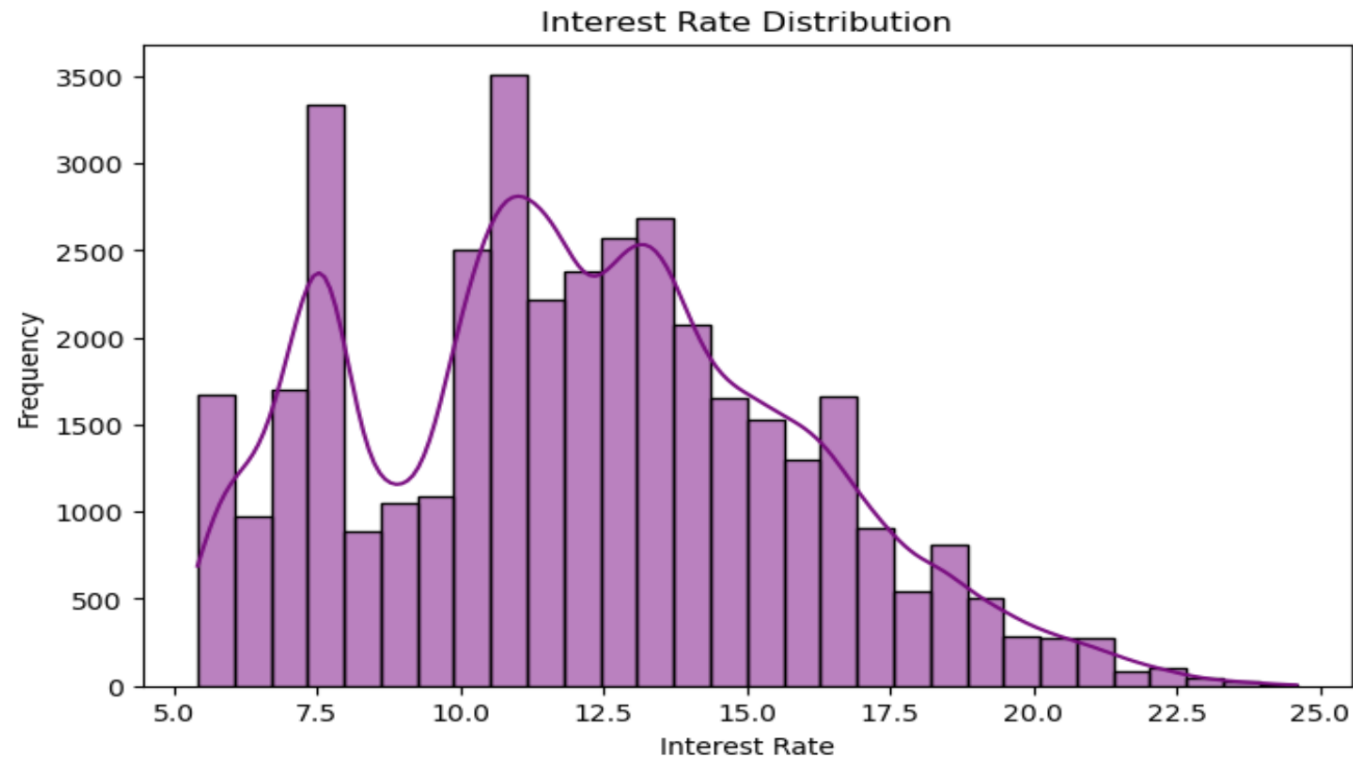
1. Most loans are for 36 months; fewer for 60 months.
2. Loan amounts are skewed; most between 5k–15k.
3. Annual income is skewed with outliers; majority <150k.
4. Loan status shows defaults (Charged Off) and Fully Paid dominate.
5. Debt consolidation is the most common loan purpose.
6. Interest rates mainly 10–15%, rising sharply for poor grades.
7. Loan amount, funded amount, and installment are highly correlated.
8. Higher income borrowers request higher loans, but risk exists across incomes.

Histogram for Loan Amount Distribution



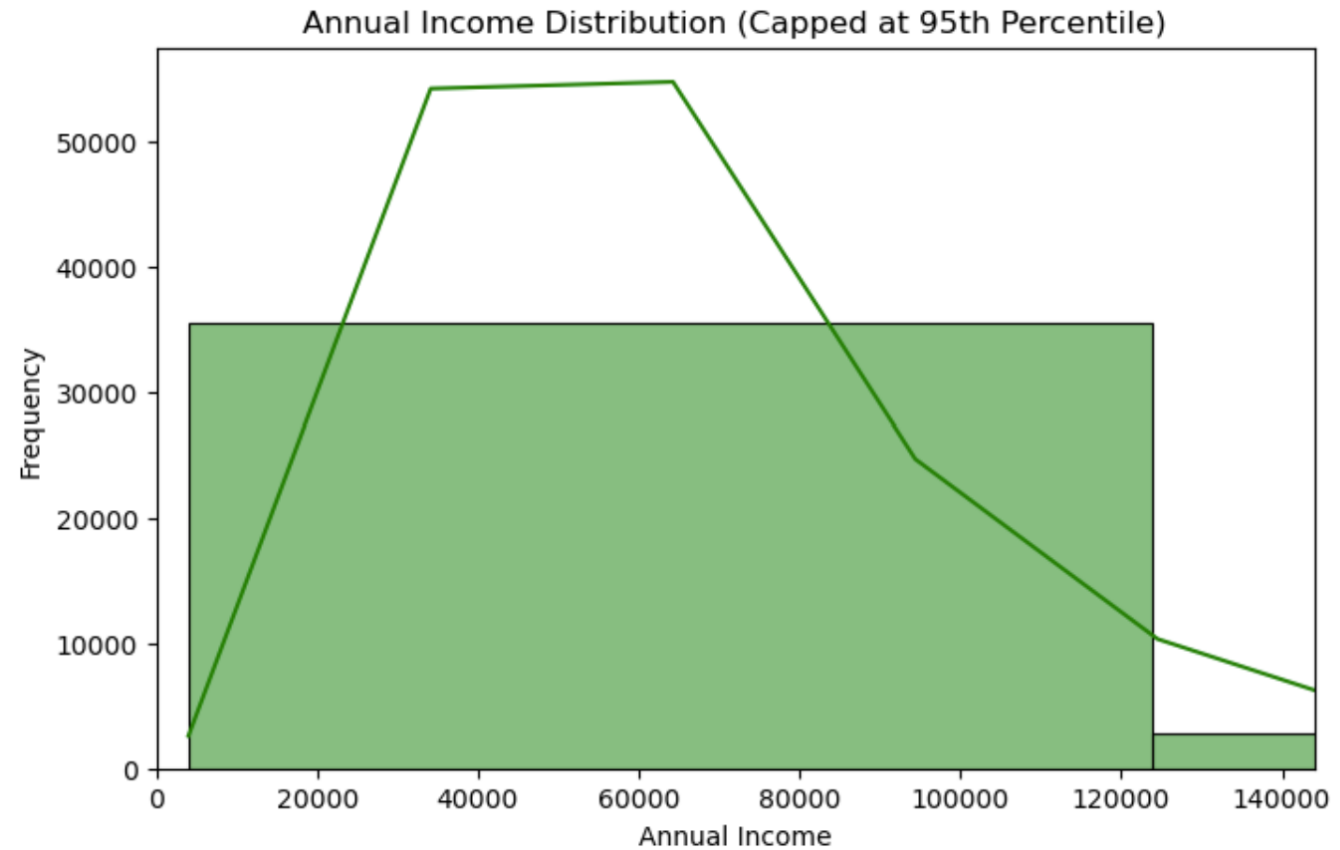
Insight: Most loans are between 5k–15k; skewed distribution, few >25k.

Histogram for Interest Rate Distribution



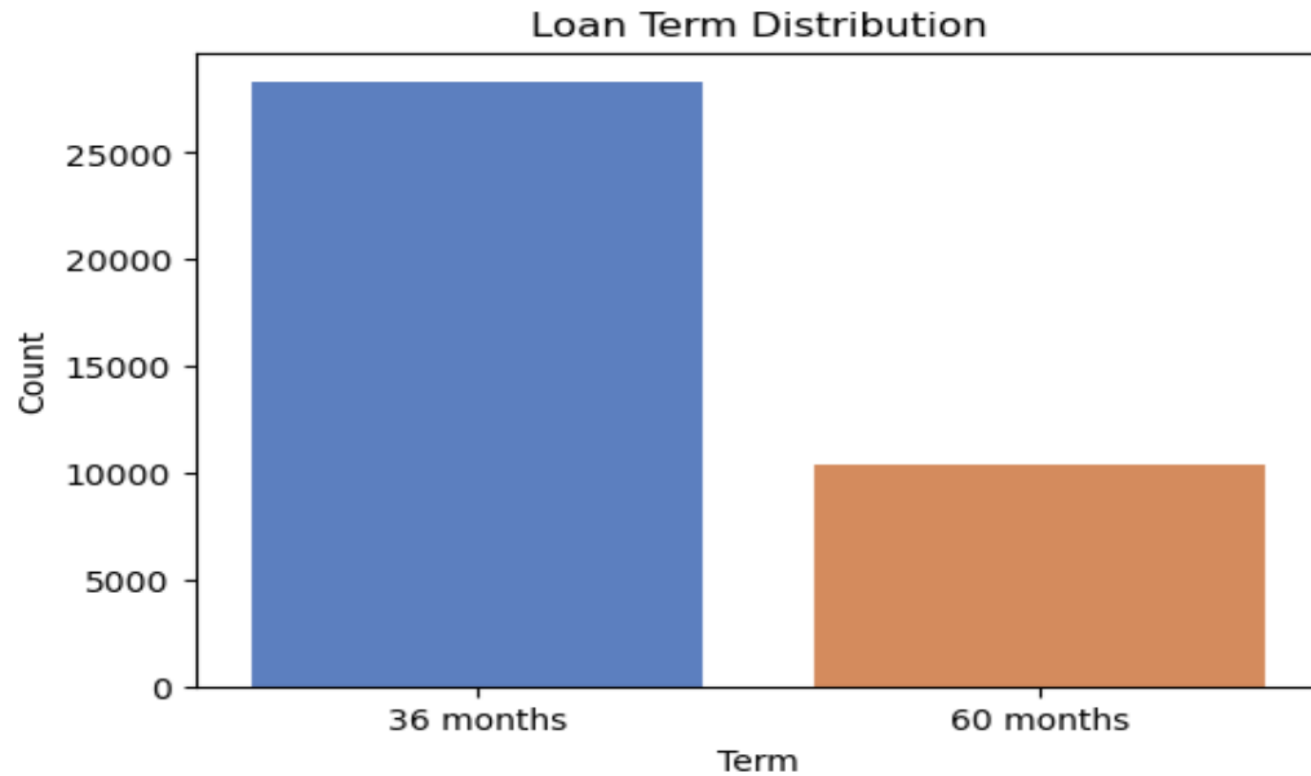
Insight: Most rates are 10–15%; higher rates linked to riskier borrowers.

Annual Income Distribution



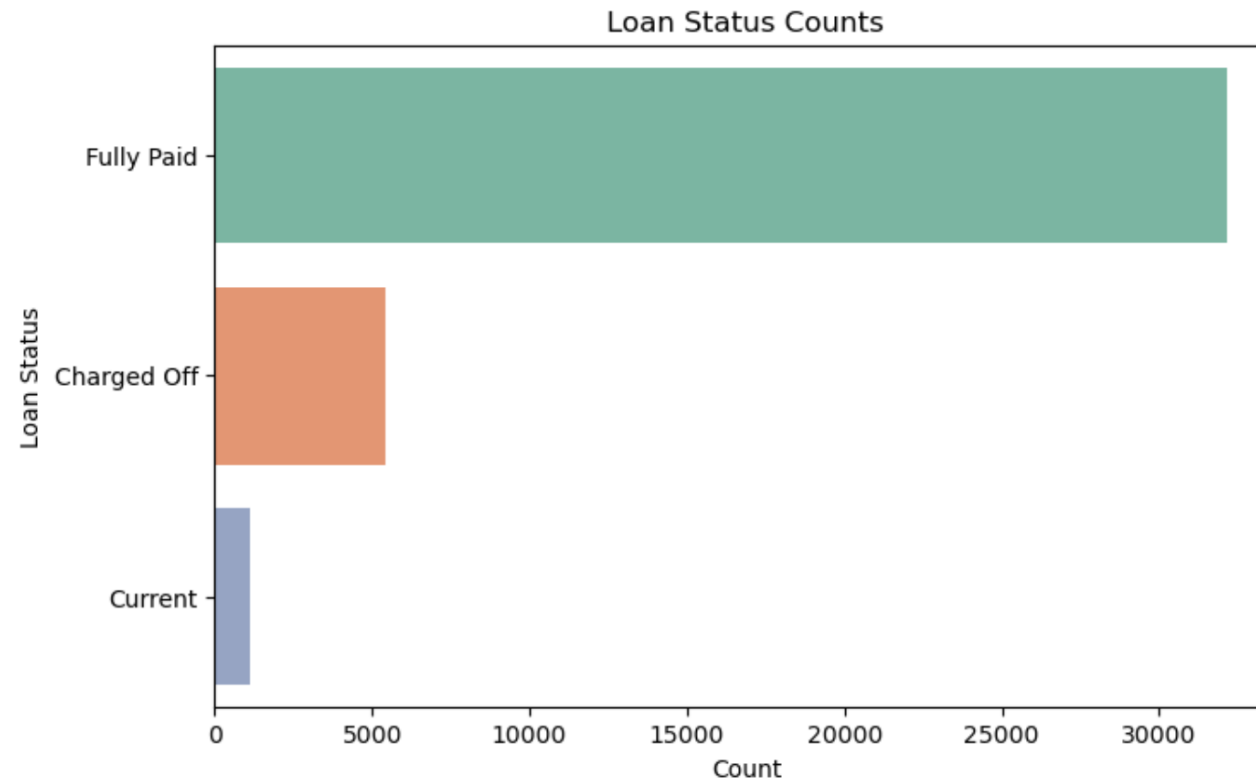
Insight: Income is skewed; majority earn <150k; extreme outliers exist.

Loan Term Distribution



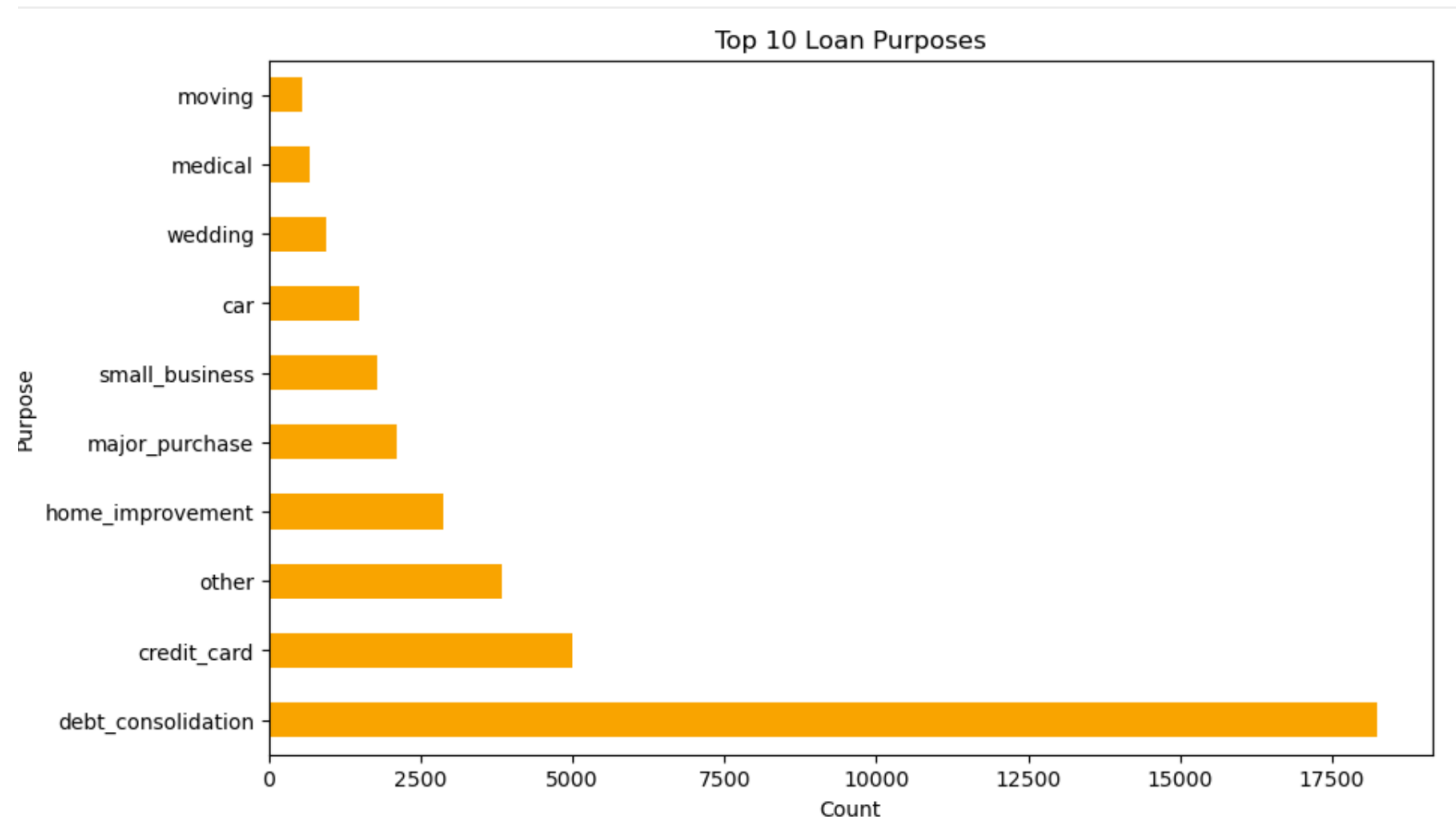
Insight: Majority loans are 36 months; fewer are 60 months.

Loan Status Distribution



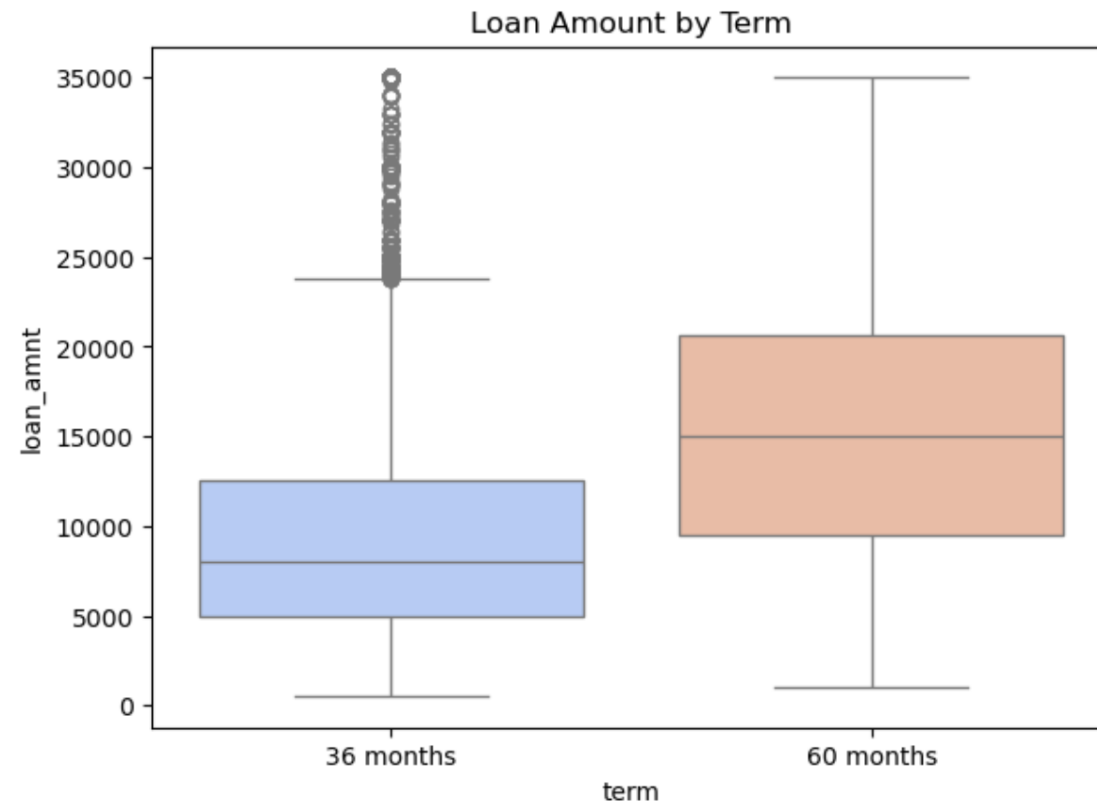
Insight: Fully Paid and Charged Off dominate; useful for default prediction.

Top 10 Loan Purposes



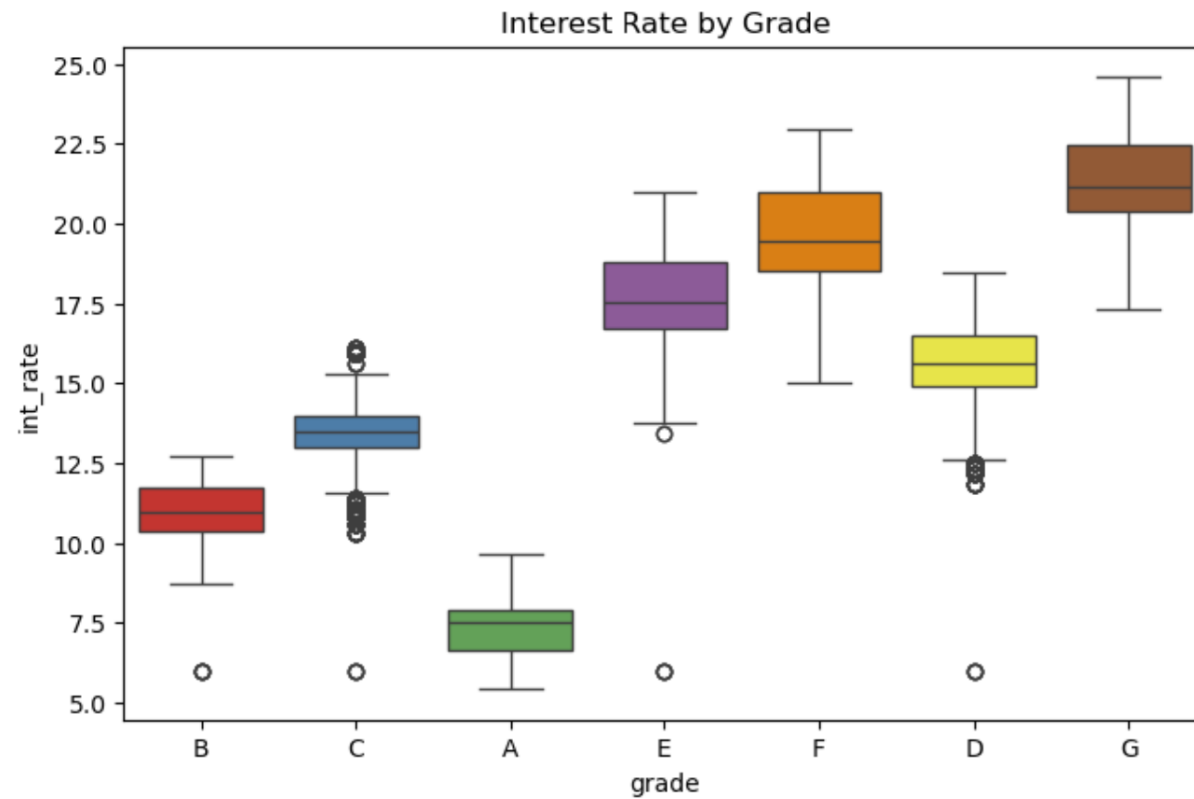
Insight: Debt Consolidation is top purpose, followed by Credit Card repayment.

Loan Amount by Term



Insight: 60-month loans slightly larger than 36-month loans.

Interest Rate by Grade



Insight: Clear trend; lower grade → higher interest.

Recommendations

- **Loan Term** → Prefer **36 months**; approve **60 months** only with strong profiles.
- **Loan Amount** → Safe: **5k–15k**; Risky: **>25k** → approve only if high income + good grade.
- **Annual Income** → Approve if **stable** and income-to-loan ratio is healthy.
- **Loan Purpose** → Prioritize **Debt Consolidation**; review “Other” carefully.
- **Loan Status Patterns** → Defaults linked to **low grades, high interest, 60-month terms**; safe loans = **good grade + 36 months**.

Thank You