Lending Club Case Study

Course: Executive Programme in ML and AI
Batch: ML67
Group: Vishal and Vinoth

Agenda

- Overview
- Univariate Analysis
- Types of derived metrics
- Bivariate Analysis
- Multivariate Analysis
- Summary

Overview

- Lending to ‘risky’ applicants is the leading cause of financial loss, known as credit loss.
- Credit loss occurs when a borrower defaults, refusing to repay the loan, resulting in a loss for the lender.
- Borrowers labeled as 'charged-off' are considered defaulters, contributing significantly to the lender’s financial loss.
- The primary goal is to identify these risky loan applicants by analyzing both consumer and loan attributes that are predictive of default.
- The company aims to understand the key factors, or driver variables, that indicate a higher likelihood of loan default.
- The insights gained from the exploratory data analysis (EDA) will help in assessing risk and reducing future financial losses.
- The dataset used for this analysis includes all loans issued from 2007 to 2011.

Univariate Analysis

Loan Status Distribution

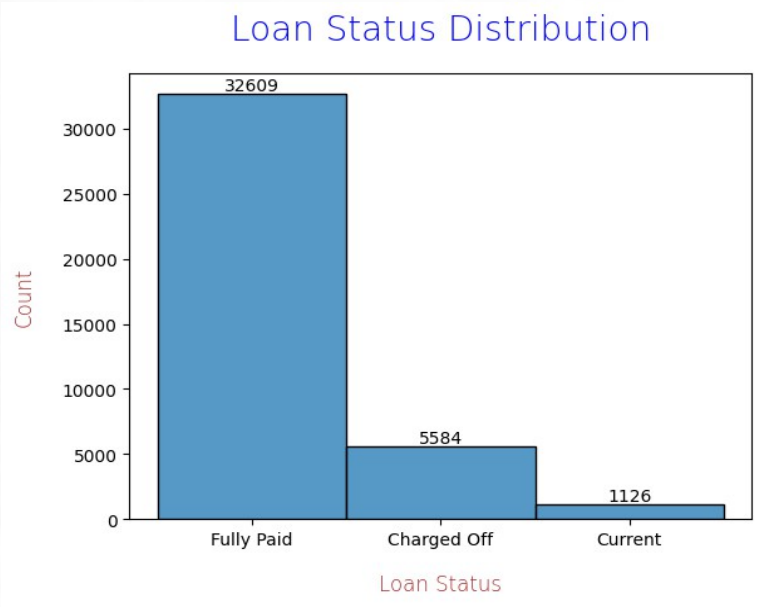
The dataset provided for analysis contains below loan statuses:

- **Fully paid**: Applicant has fully paid the loan
- **Current**: Applicant is in the process of paying the instalments
- **Charged-off**: applicant has defaulted on the loan

Out of the total approved loan applications, 5584 cases defaulted.

In next few slides, we will try to determine that attributes

that may have influenced these cases.

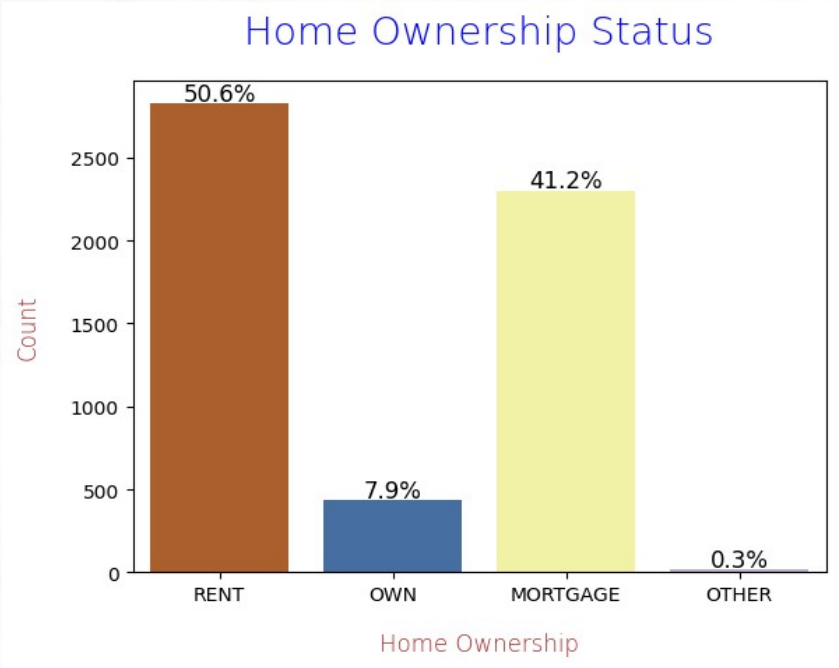


Home ownership analysis

- Rent – 50.6%
- Own – 7.9%
- Mortgage – 41.2%
- Other – 0.3%

The analysis shows that applicants

with RENTED home defaulted the most

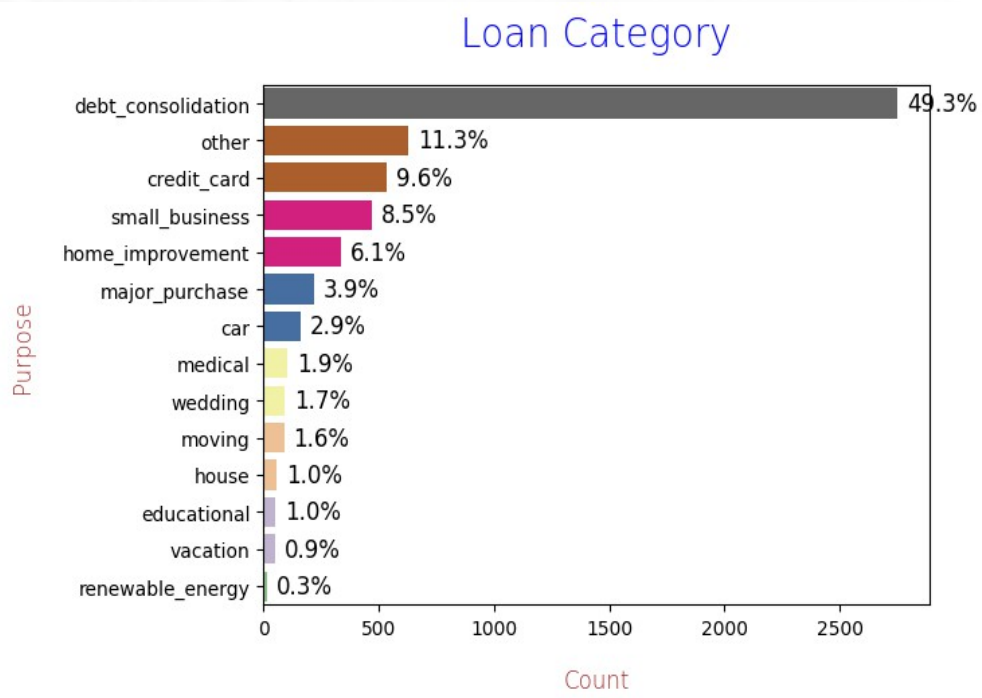


Loan Category Analysis

- Debt Consolidation – 49.3%

Applicants who took loan for

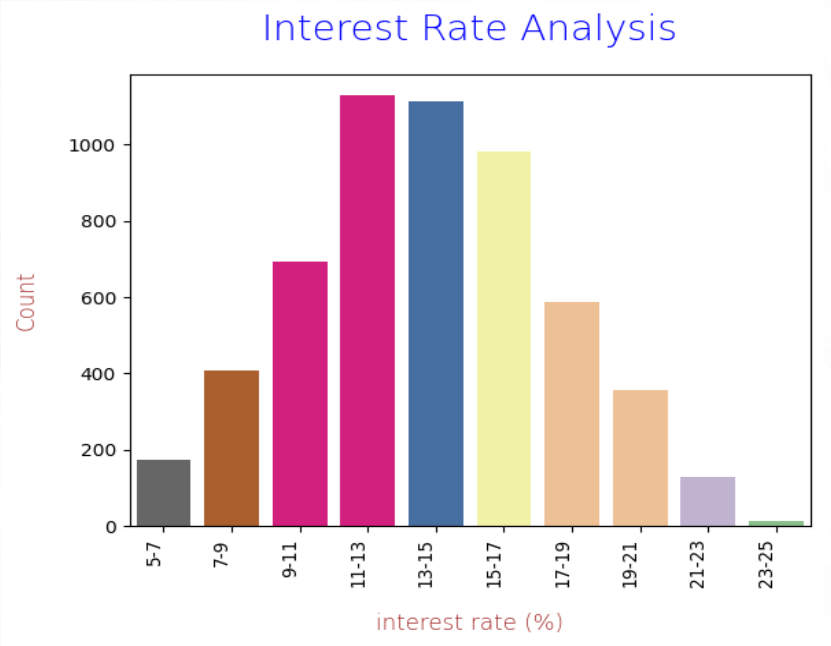
debt consolidation defaulted the most



Interest Rate Analysis

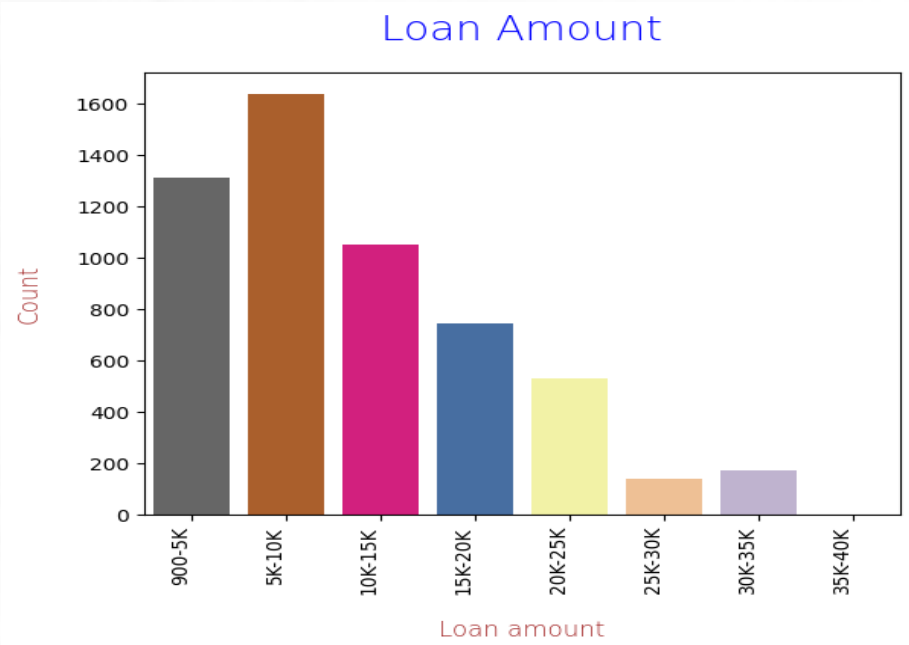
Applicants who got loan at rate of interest

between 11% to 17% defaulted the most



Analysis Based on Loan Amount

Applicants who took loan for the amount in range of 5000 to 10000 defaulted the most.



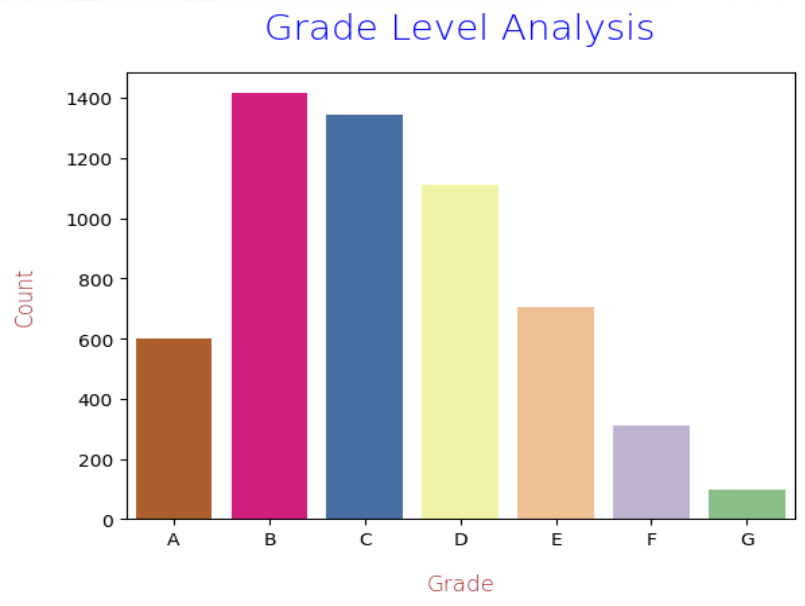
Loan Grade Analysis

Loan grading is a classification system

that involves assigning a quality score

to a loan based on a borrower's credit history.

Applicants having **Grades B,C,D** defaulted the most.

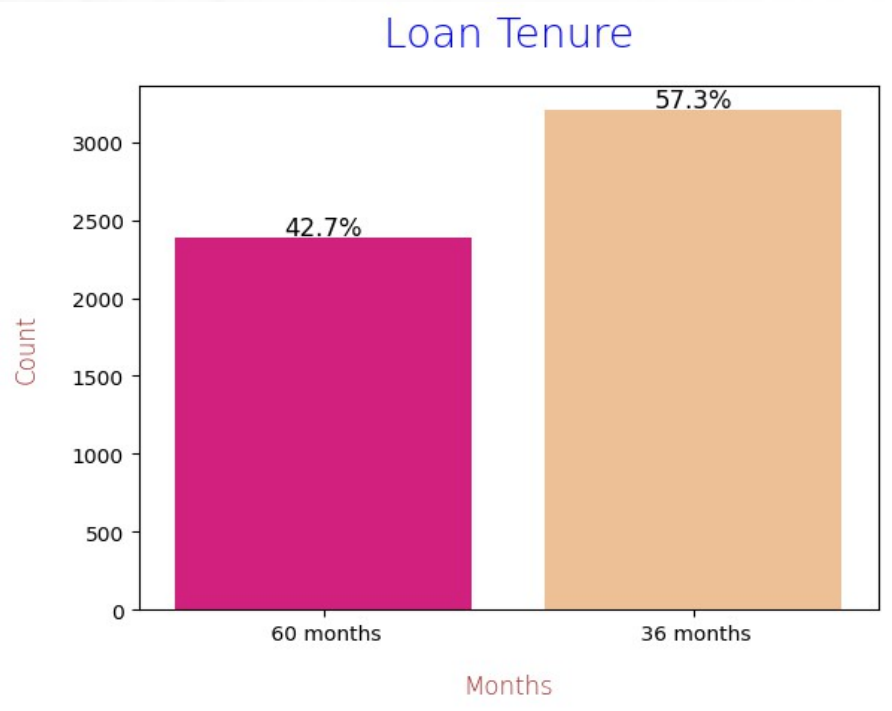


Loan Tenure Analysis

The analysis shows that applicants

who opted for 36 months loan tenure defaulted more

compared to 60 months.

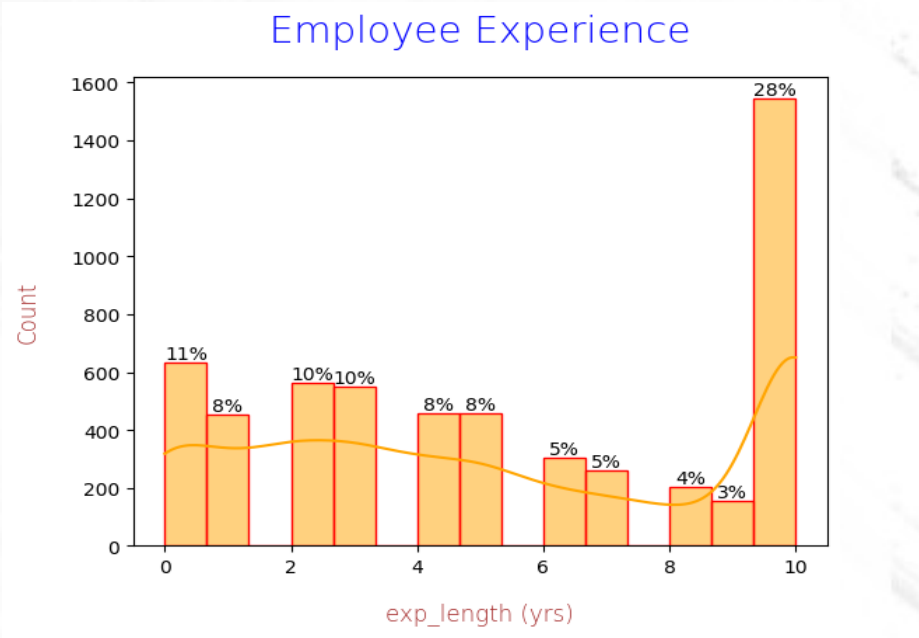


Types of Derived Metrics

Type-driven Metrics

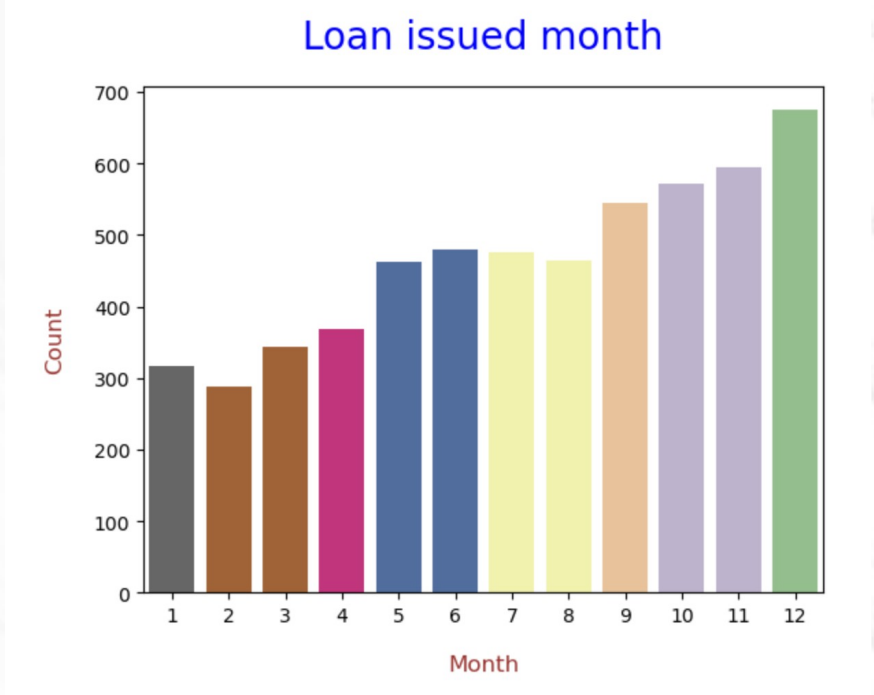
- 10+ Years experience: 28%

Applicants having experience of 10 or more years defaulted the most



Business-driven Metrics

Loans issued in December have the highest default rate



Data-driven metrics

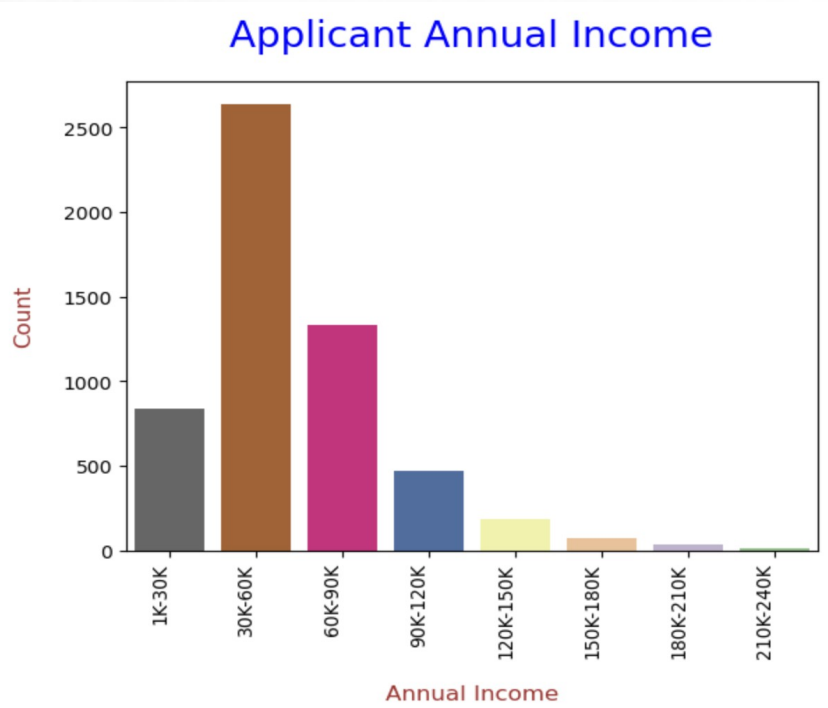
The analysis shows that applicants with annual income in t

range of 30000 to 60000 defaulted the most.

Interestingly, applicants in very low

income range (1000 to 30000) have fewer

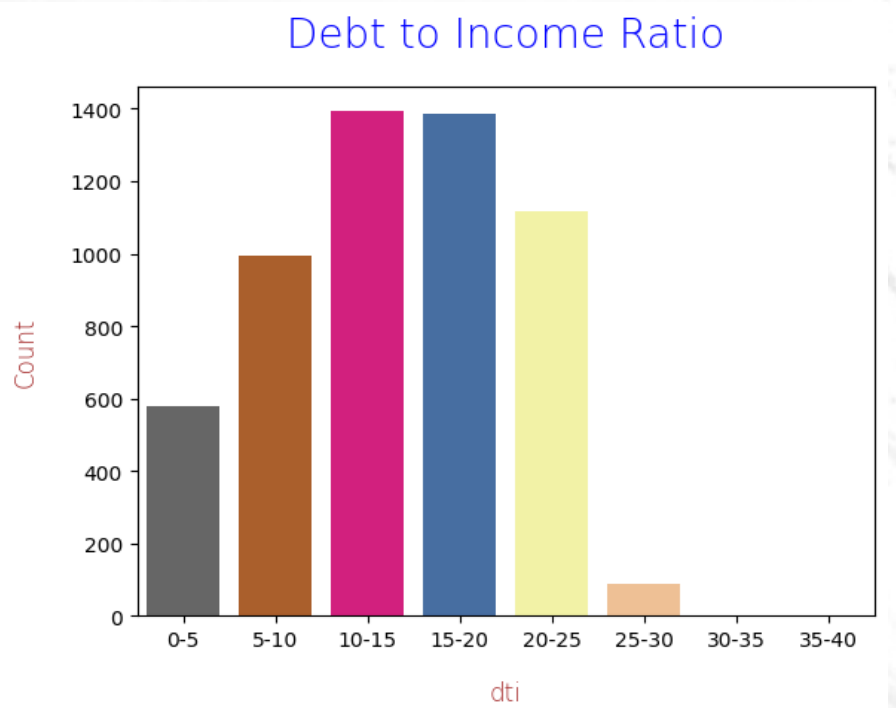
number of defaulted cases.




```
charged_off['inc_bins'] = pd.cut(charged_off['annual_inc'],
                                 bins=[1000, 30000, 60000, 90000, 120000, 150000, 180000, 210000, 240000],
                                 labels =['1K-30K','30K-60K','60K-90K','90K-120K','120K-150K', '150K-180K', '180K-210K', '210K-240K'],
                                 precision=0)
```

Data-driven metrics

Applicants with a monthly debt-to-income ratio between 10% and 20% have the highest default rate.

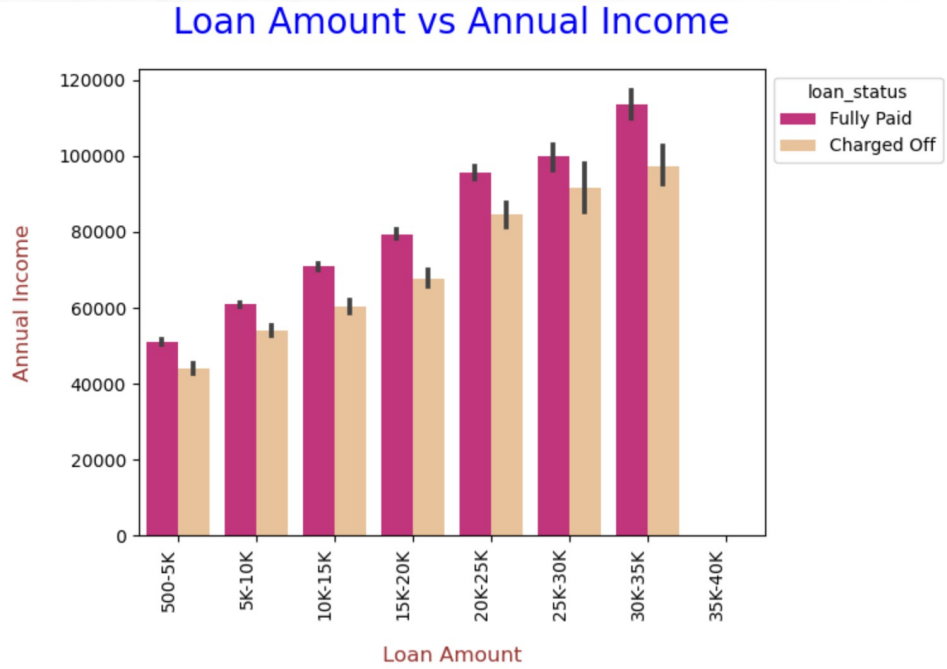



```
charged_off['dti_bins'] = pd.cut(charged_off['dti'],
                                 bins=[0, 5, 10, 15, 20, 25, 30, 35, 40],
                                 labels =['0-5','5-10','10-15','15-20','20-25', '25-30', '30-35', '35-40'],
                                 precision=0)
```

Bivariate Analysis

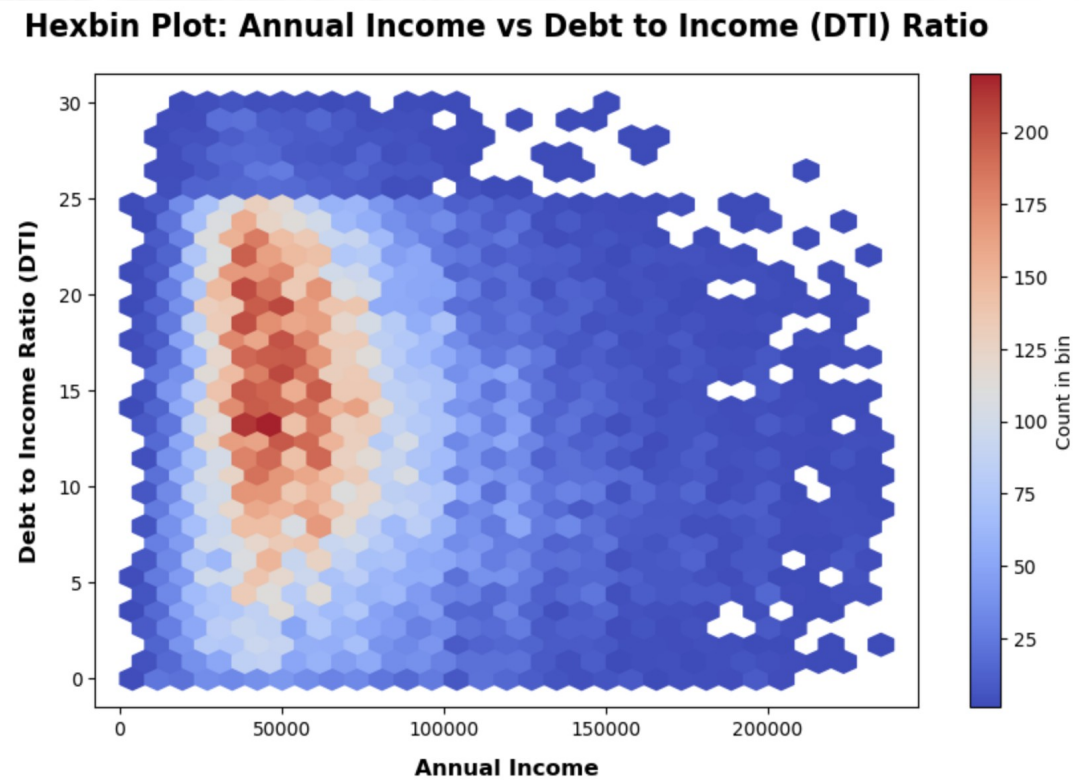
Loan Amount vs Annual Income

- As annual income increases the loan amount tends to increase as well.
- A correlation of 0.41 indicates a moderate positive relationship.



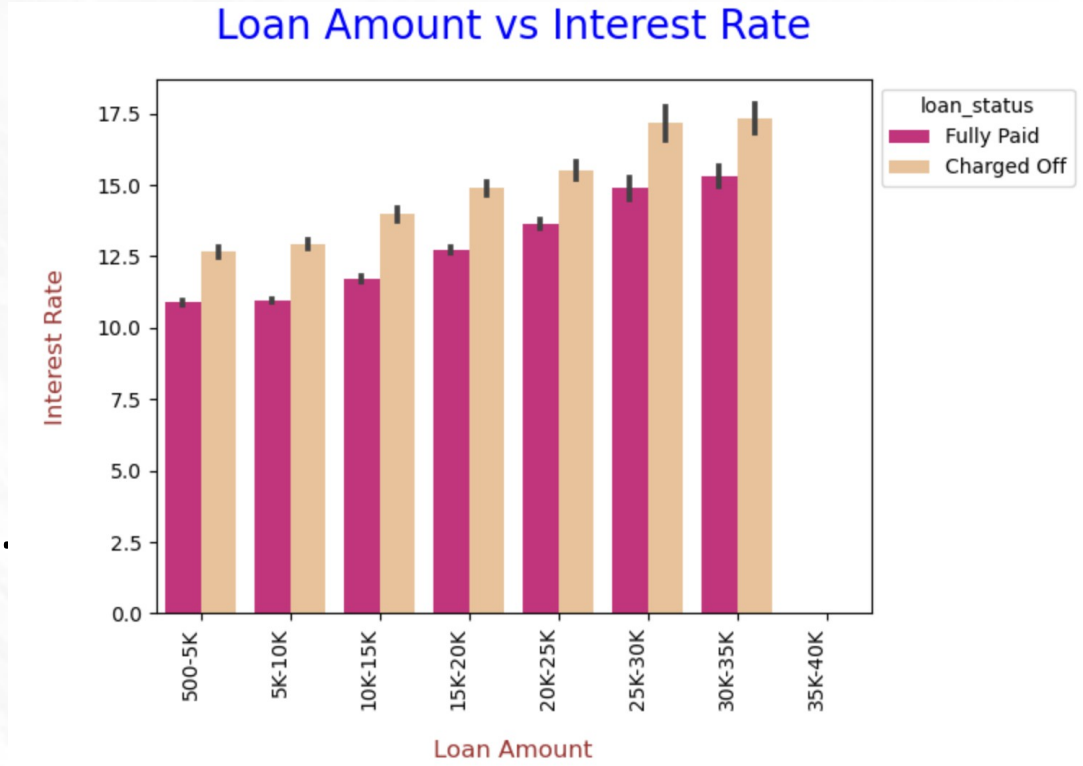
Annual income vs Debt to Income ratio

- **Debt management issues**: The dense orange-red region where income is between 30000 and 70000 and DTI is between 15 and 25.
- Borrowers with higher incomes able to manage their debt.



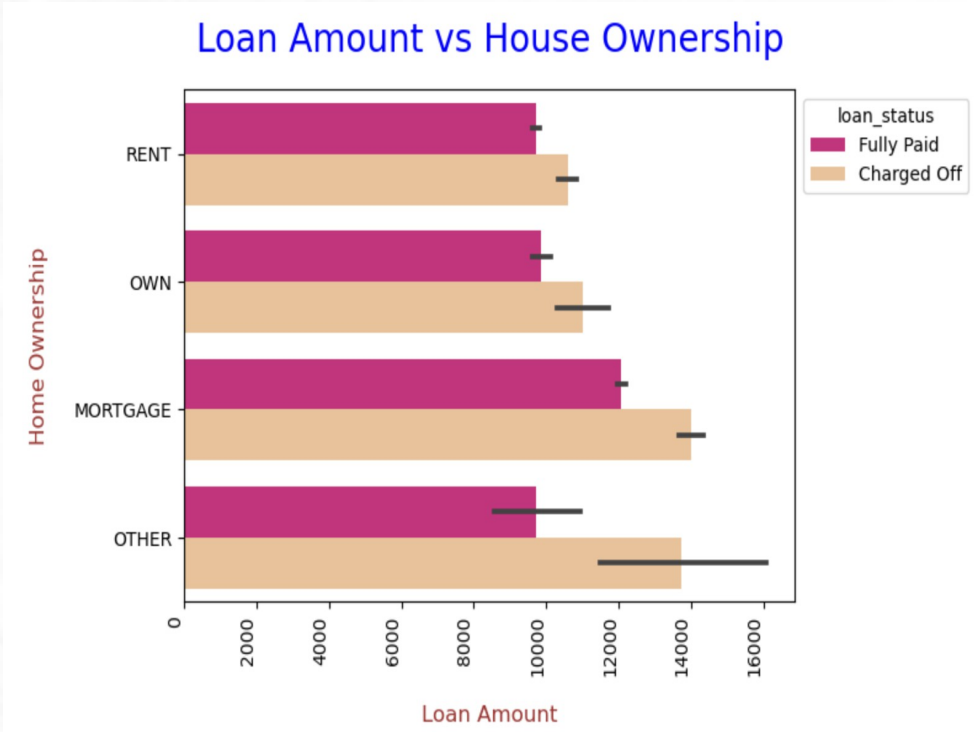
Loan Amount vs Interest Rate

- A correlation of **0.35** indicates a **moderate positive relationship** between two variables.



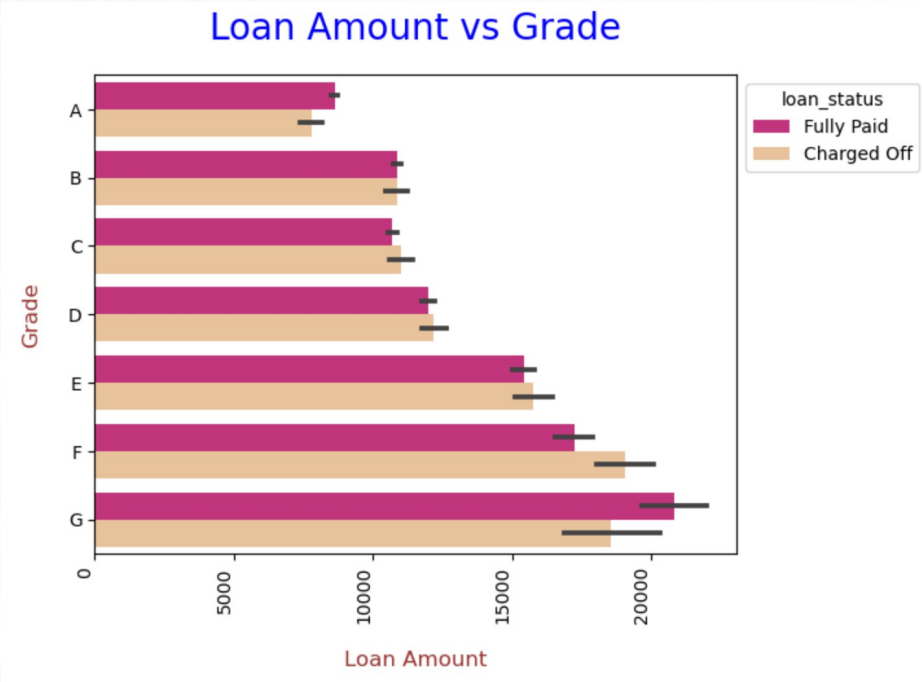
Loan Amount vs House Ownership

- Default cases highest for applicants with **mortgaged** or **other** ownership types



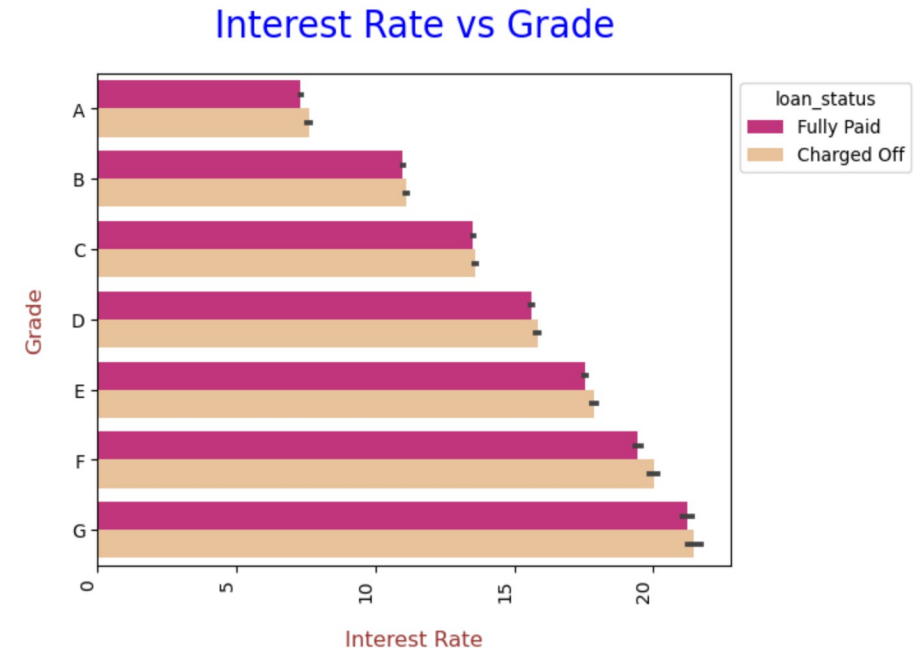
Loan Amount vs Grade

- High default rates are likely when the loan grade is **F** or **G,** and the loan amount exceeds approximately **17,000**



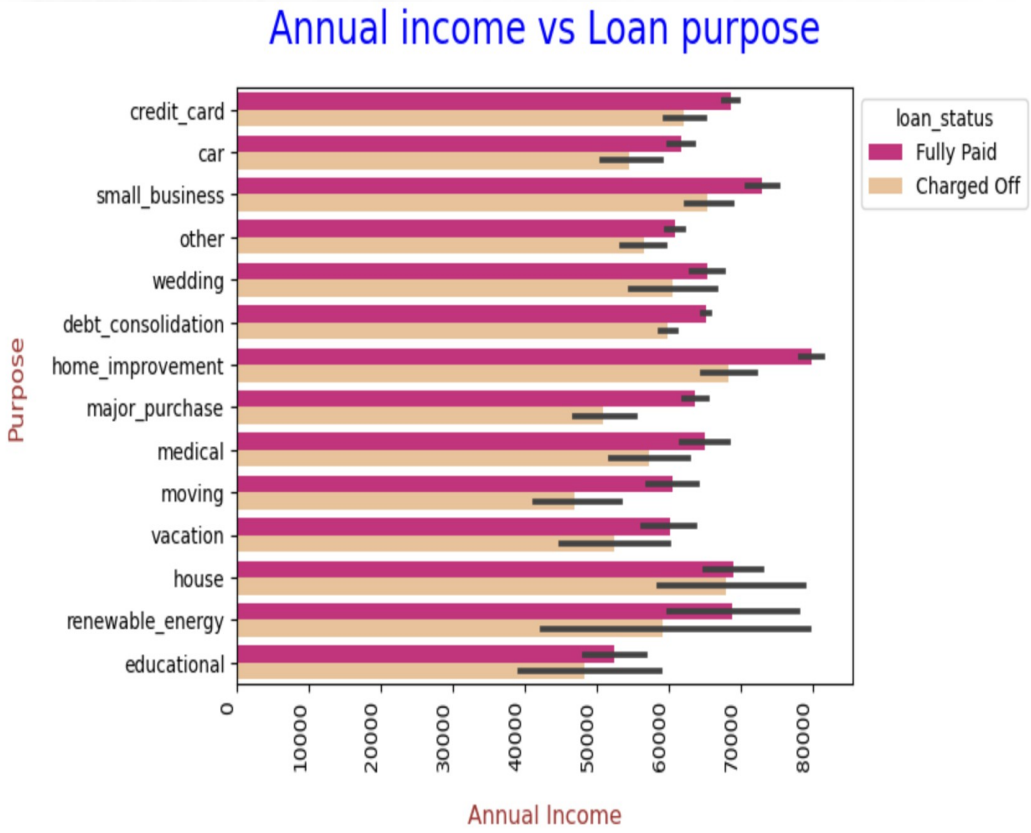
Interest Rate vs Grade

- As grade decreases (assuming A is highest and G is lowest), people are getting loans at higher interest rate. Defaulted cases are also increasing



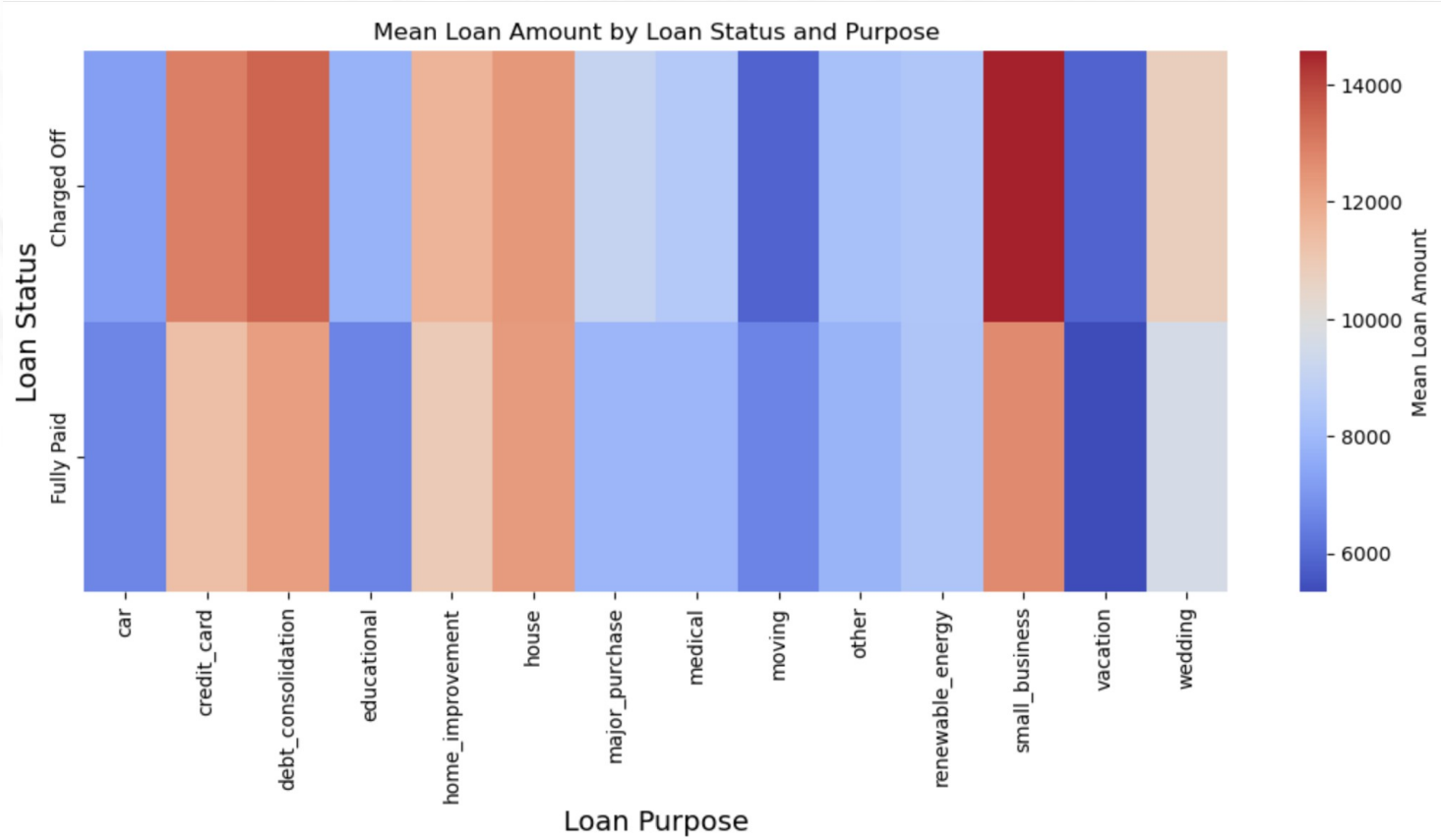
Annual income vs Loan purpose

- The highest default rates are observed in **credit cards, small business loans, home improvement loans**, and **house** mortgages.



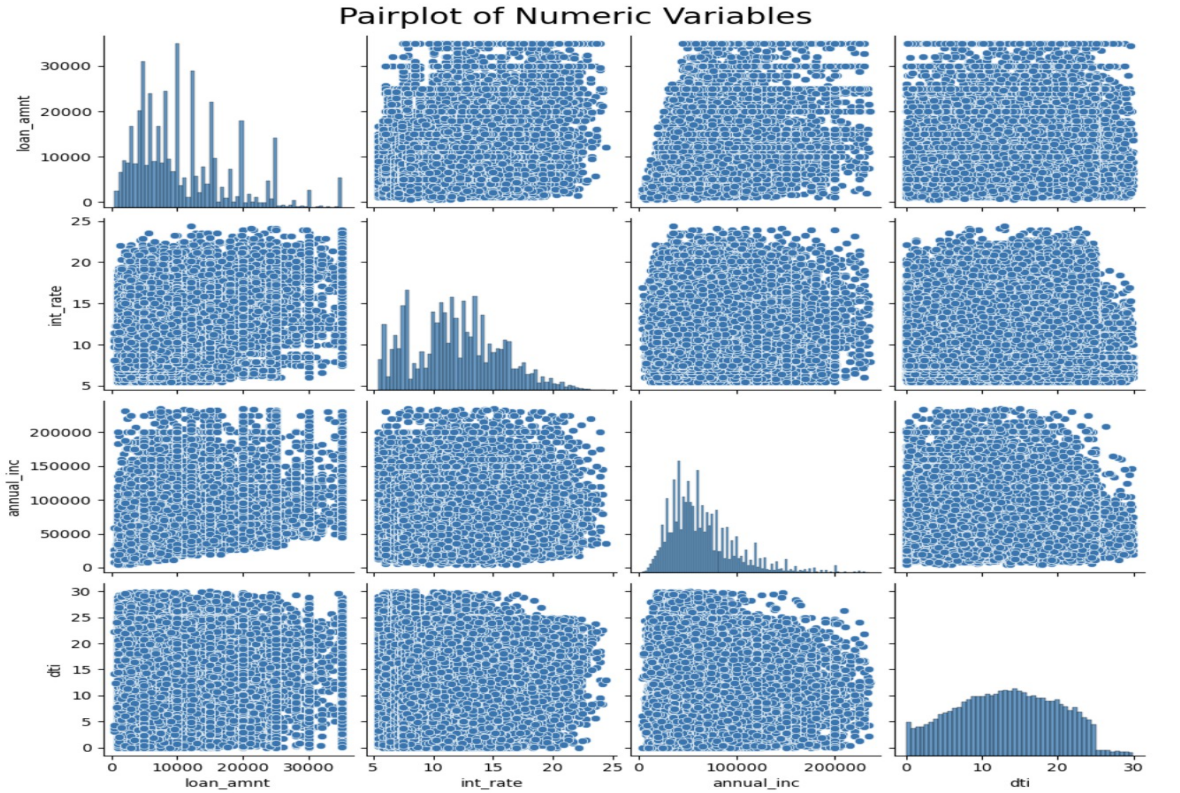
Multivariate Analysis

Mean Loan Amount by Loan Status and Purpose



Small business loans have defaulted, while credit cards and debt consolidation loans are on the verge of default.

Pair plot Analysis



Correlation Matrix:

	loan_amnt	int_rate	annual_inc	dti
loan_amnt	1.000000	0.298393	0.414847	0.071841
int_rate	0.298393	1.000000	0.066540	0.113477
annual_inc	0.414847	0.066540	1.000000	-0.107875
dti	0.071841	0.113477	-0.107875	1.000000

Summary

Univariate Analysis:

1. Applicants with rented homes accounted for the highest defaults (50%).
2. Debt consolidation loans led to the most defaults (49%).
3. Borrowers with an interest rate between 11% to 17% had the highest default rate.
4. Loans in the amount range of 5000 to 10000 were most prone to default.

Derived Metrics:

1. Type-driven Metric: Applicants with 10+ years of experience account for 28% of defaults, the highest among experience groups.
2. Business-driven Metric: Loans issued in December have the highest default rate.
3. Data-driven Metrics:
4. Applicants with an annual income between 30,000 and 60,000 defaulted the most.
5. Interestingly, those in the lower income range (1,000 to 30,000) showed fewer defaults.
6. Borrowers with a monthly debt-to-income ratio of 10% to 20% had the highest default rate.

Bivariate Analysis:

1. Annual Income vs Loan Purpose: The highest default rates are seen in loans for credit cards, small businesses, home improvement, and house mortgages.
2. Annual Income vs Debt-to-Income (DTI) Ratio: Borrowers with incomes between 30000 and 70000 and a DTI ratio between 15% and 25% show debt management issues.
3. Loan Amount vs Annual Income: A moderate positive correlation (0.41) indicates that as annual income increases, the loan amount tends to rise as well.
4. Loan Amount vs Grade: High default rates are observed when the loan grade is F or G and the loan amount exceeds 17000.