

LENDING CLUB CASE STUDY SUBMISSION

Group Members:

- 1. Sindhushree G S**
- 2. Krisalay Kumar**

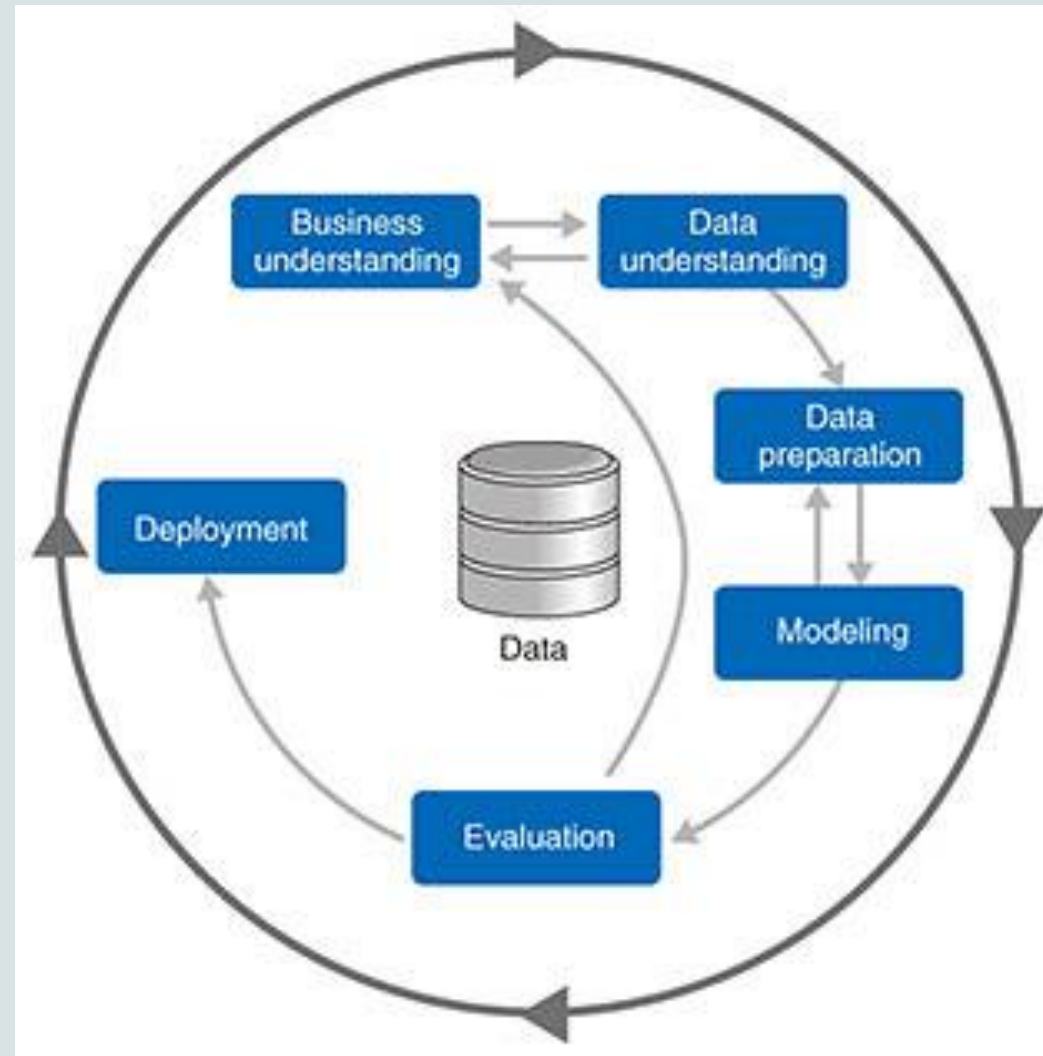
OBJECTIVE OF THE CASE STUDY

Identification of Loan Applicant traits that tend to 'default' paying back

Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena

Loan Lending Organizations may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants

Methodology used to solve the Problem



Data Cleaning Steps

1. **Duplicate Rows Check:** Check for duplicate rows in the dataset
2. **Missing values:** Treat missing values with appropriate approach.
3. Analyse & Delete Unnecessary Columns in data frame
4. **Standardize Values:** Ensure all observations under a variable have a common and consistent unit.
5. **Outlier Identification & Removal**

Analysis

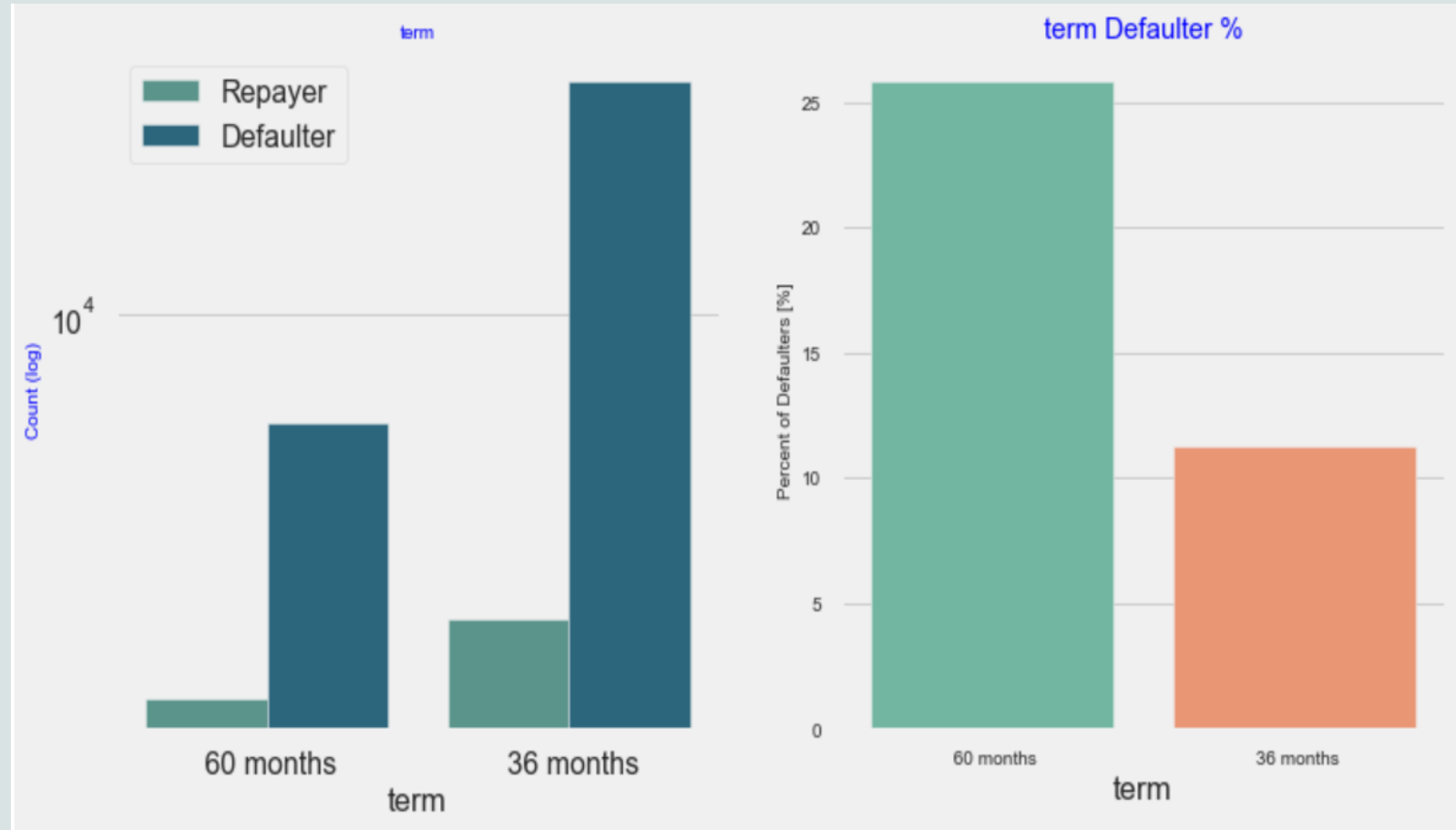
Strategy: The data analysis flow has been planned in following way :

- Imbalance in Data
- Categorical Data Analysis
 - Categorical segmented Univariate Analysis
 - Categorical Bi/Multivariate analysis
- Numeric Data Analysis
 - Bi-furcation of databased based on TARGET data
 - Correlation Matrix
 - Numerical segmented Univariate Analysis
 - Numerical Bi/Multivariate analysis

Loan term analysis

Observations:

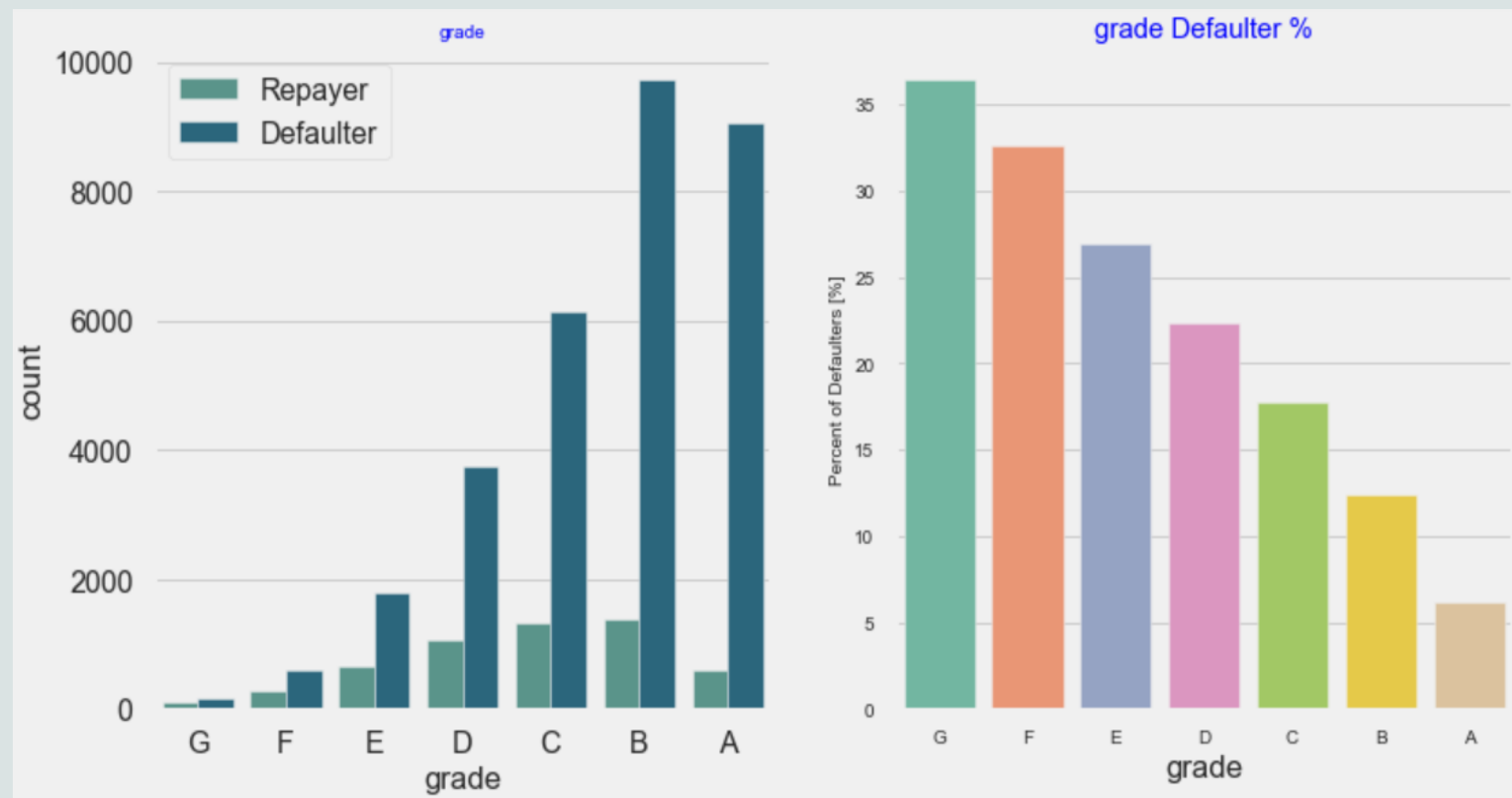
Term: Loans with 60 months tenure are just a small fraction from the total number of loans; in the same time, a larger amount of 60 months tenure loans, comparing with their frequency, are not repaid.



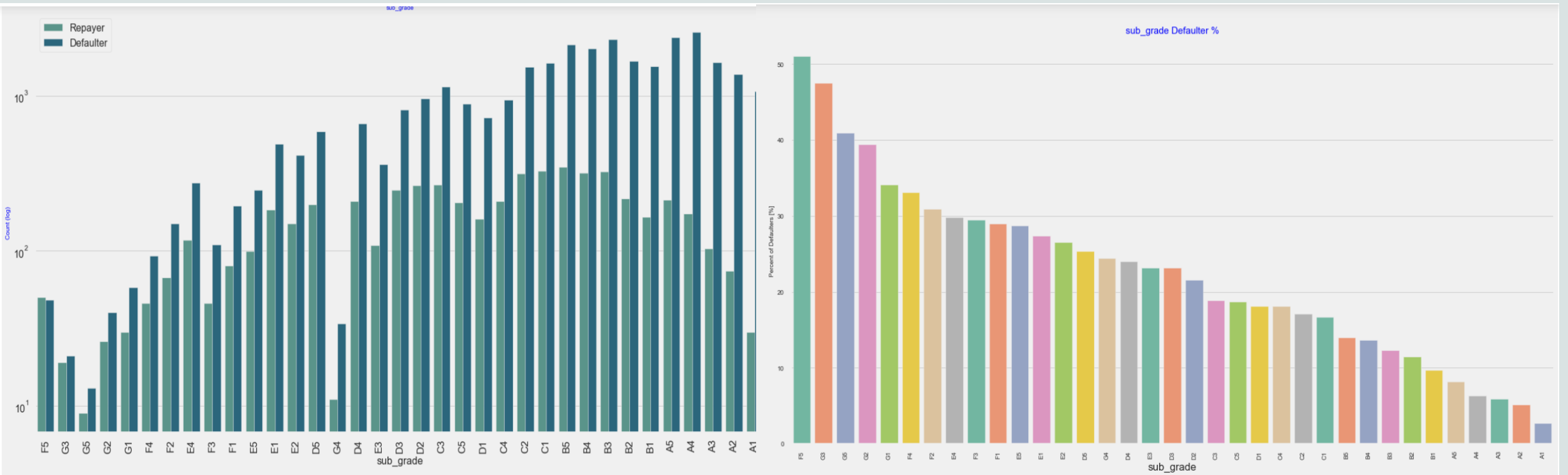
Grade analysis

Observations:

Grade: Loans with graders: A and B are in large fraction from the total number of loans. Based on the percentage of defaulted credits, grade G, F, and E have a higher chance of not returning their loans, comparing with other grades



Sub-grade analysis

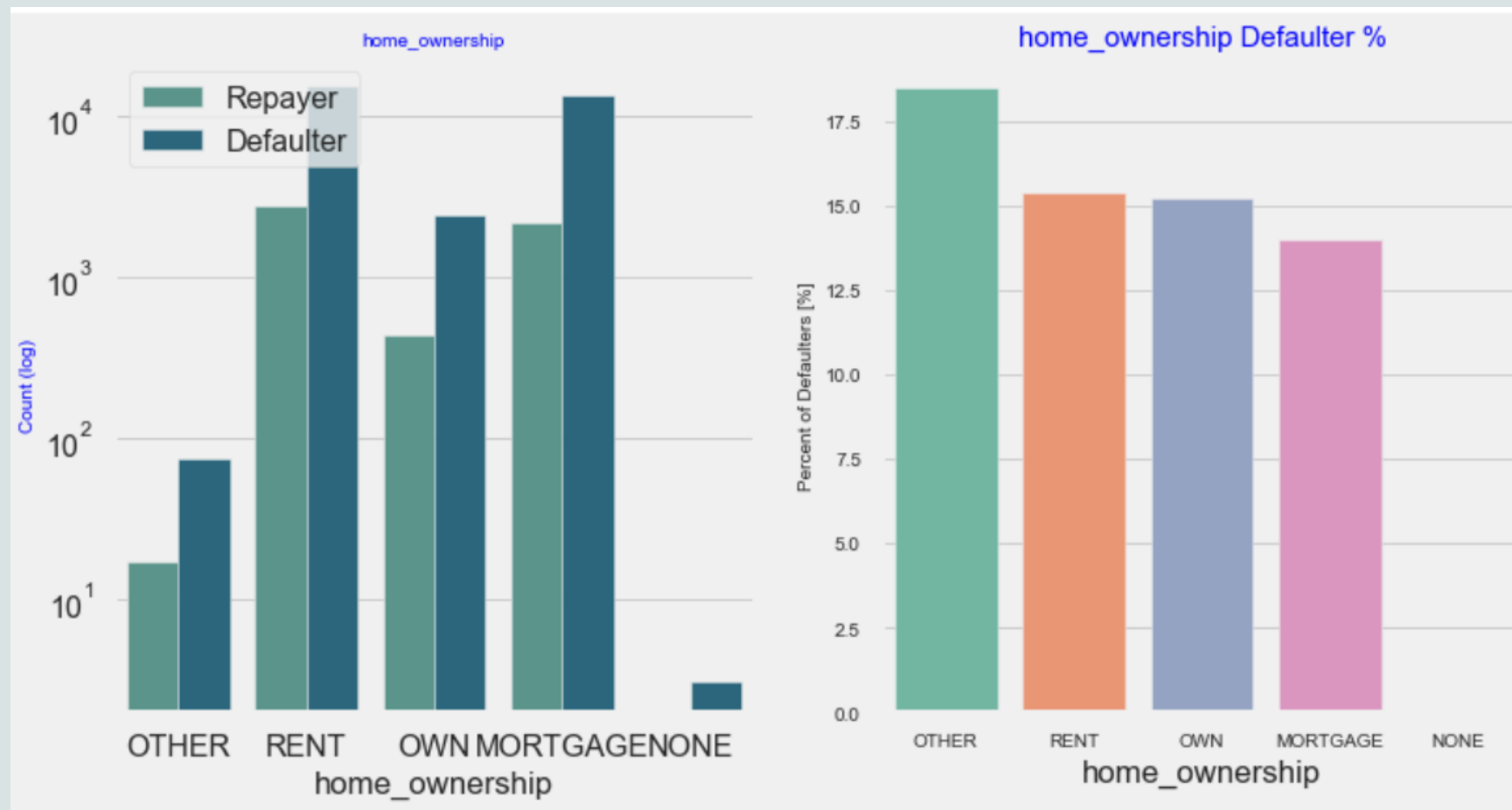


Observations: Sub Grade: Loans with sub grades: F5,G3, G5, G2, G1 has less change to get repaid.

Home ownership Analysis

Observations:

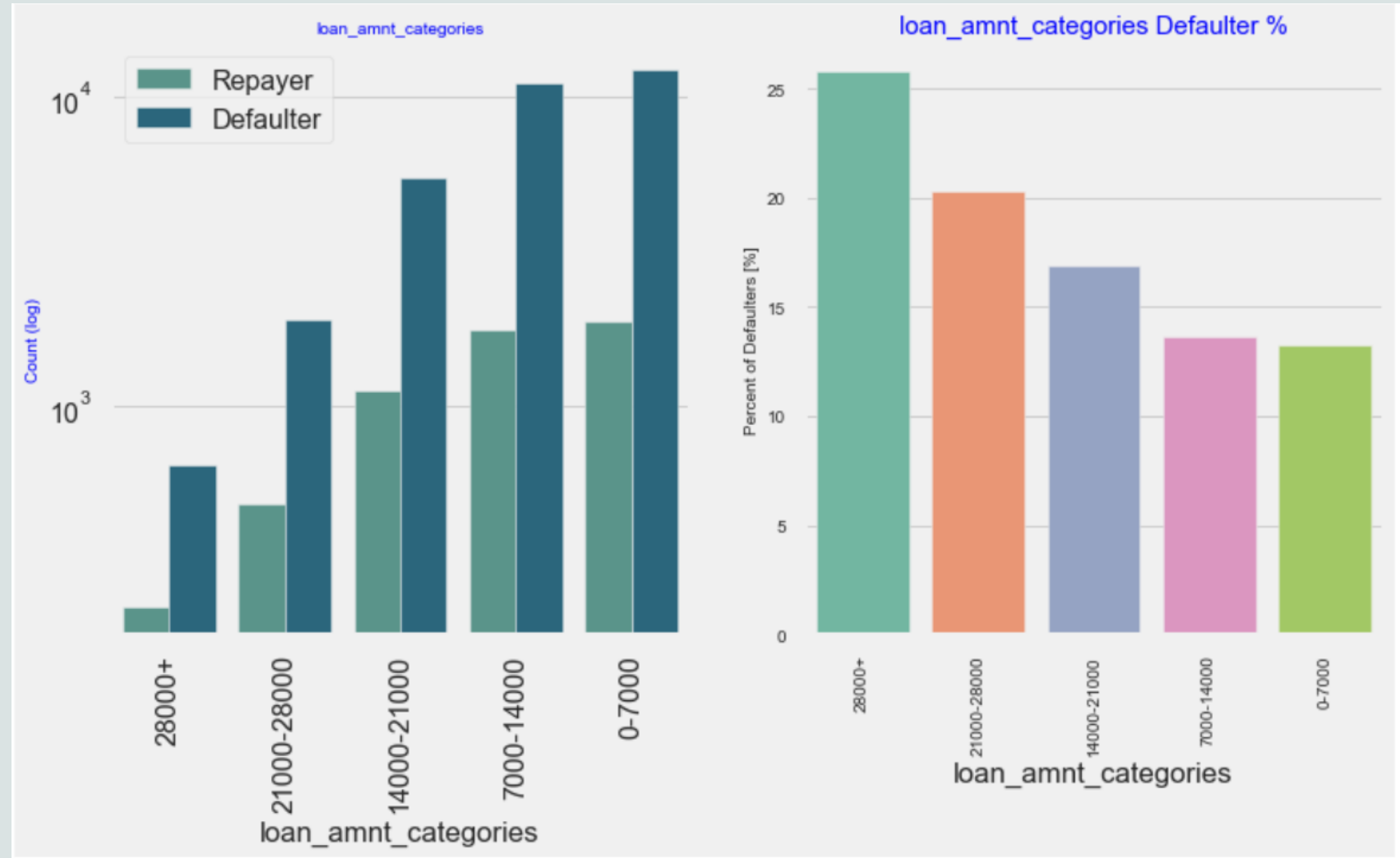
Home Ownership: Applicants whose home are Rented, Owned, and Mortgage comprises of most of the loan applications. But, most of the number of defaulters belong to the category "Other", but there is no significant difference between the defaulters.



Loan amount categories

Observations:

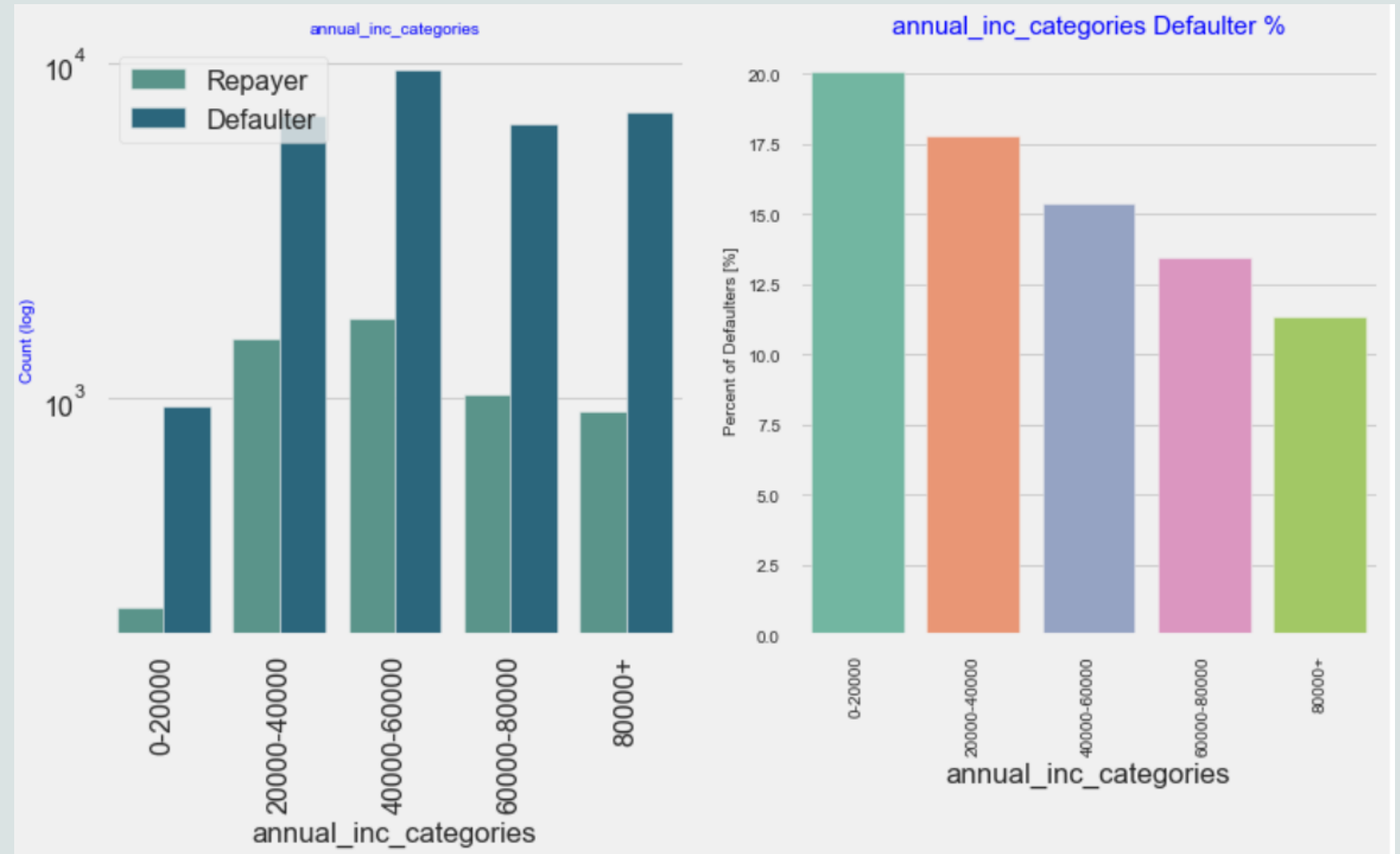
Loan Amount: Most of the applicants has applied the loan ranging from 0 to 21k. Applicants who has applied for loan amount more than 28k is more likely to default with ~27% probability.



Annual Income categories

Observations:

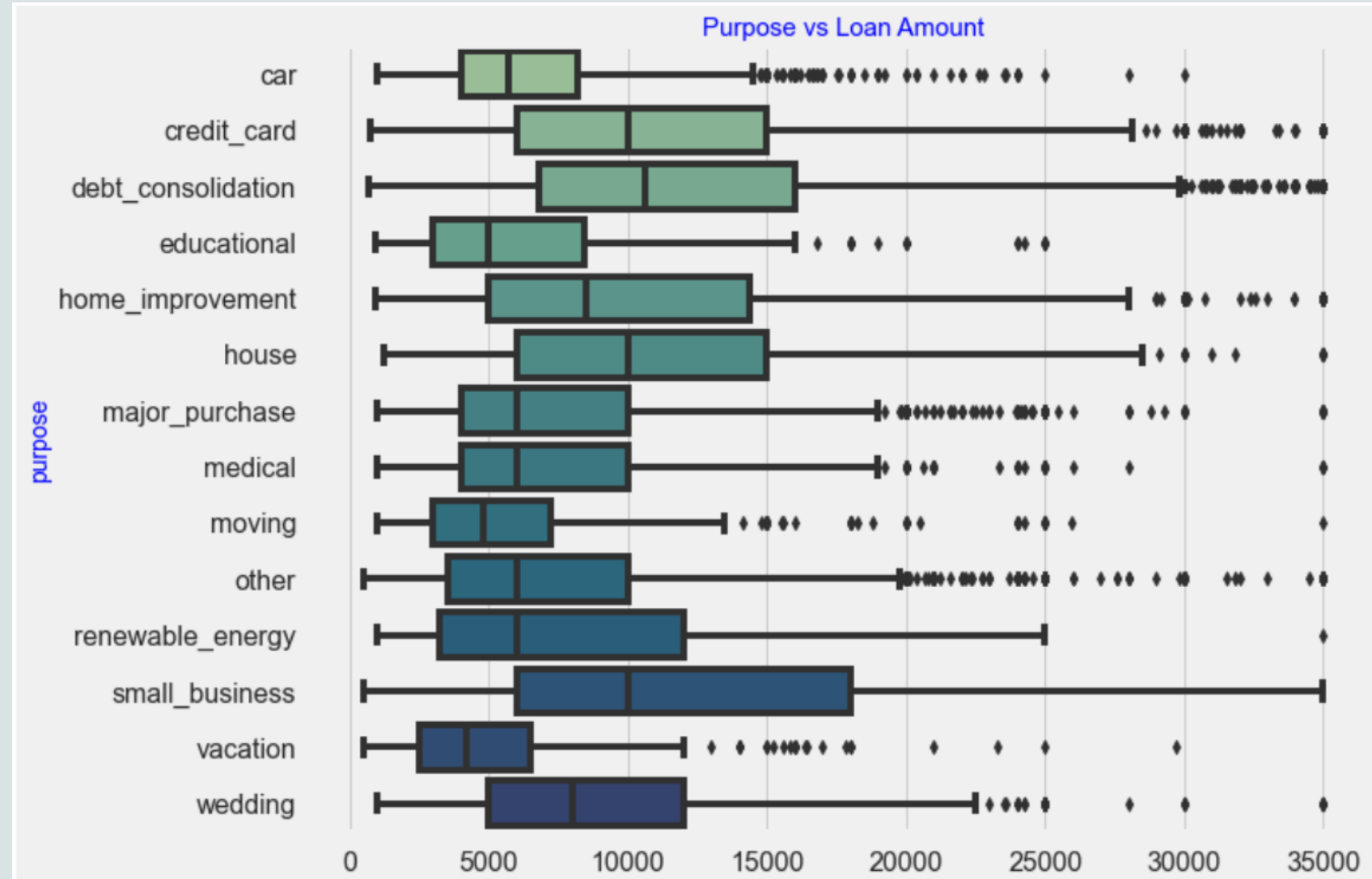
Annual Income: Most of the applicants whose annual income is more than 20k has applied for the loan. And the probability of getting default is more for the applicants whose annual income ranges in between 0 to 40k.



Bivariate Analysis: Analysis of loan amount and it's purpose

Observations:

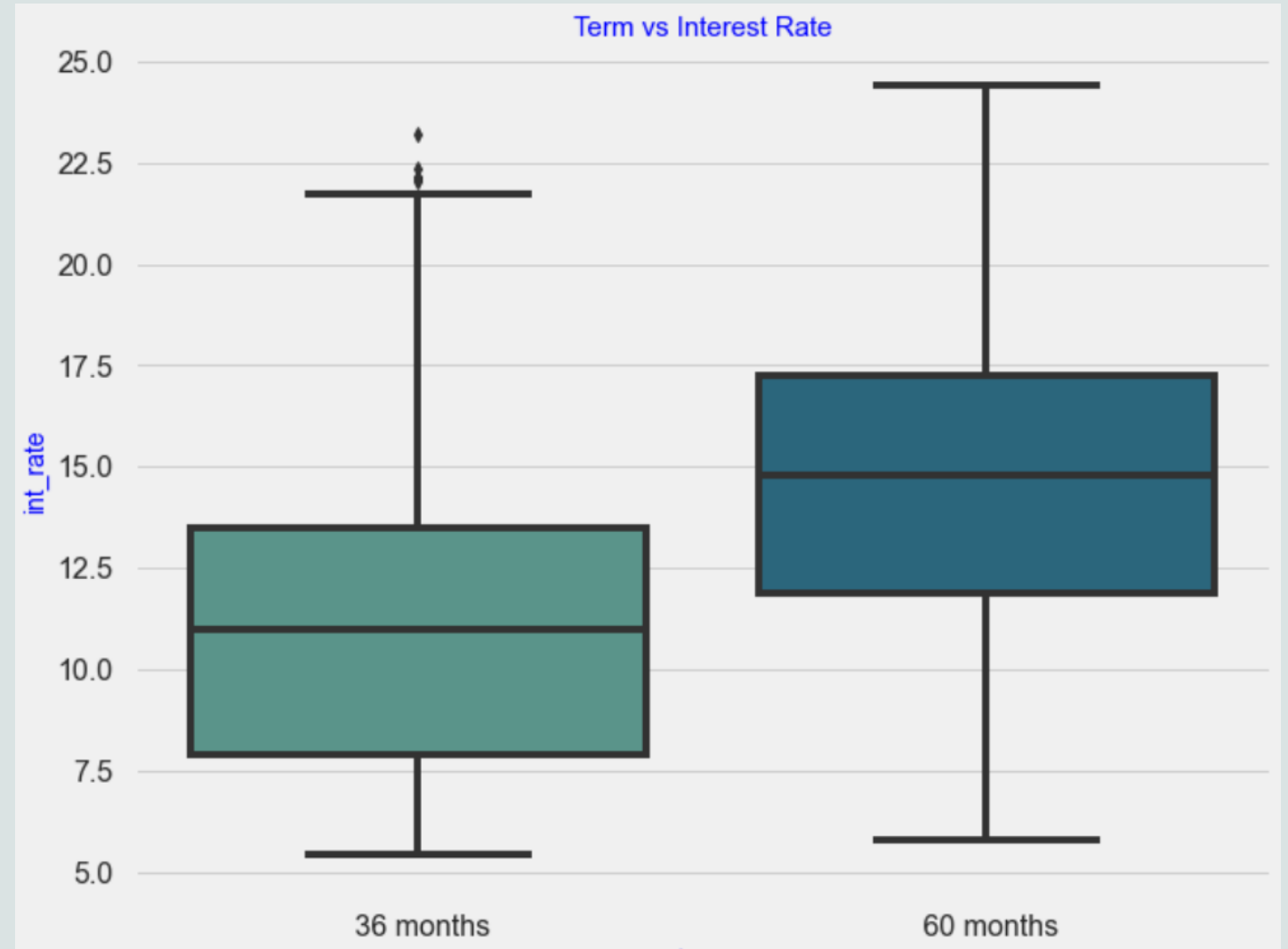
1. Median, 95th percentile, 75th percentile of loan amount is highest for loan taken for small business purpose among all purposes.
2. Debt consolidation is second and Credit card comes 3rd.



Loan term and it's interest rate

Observations:

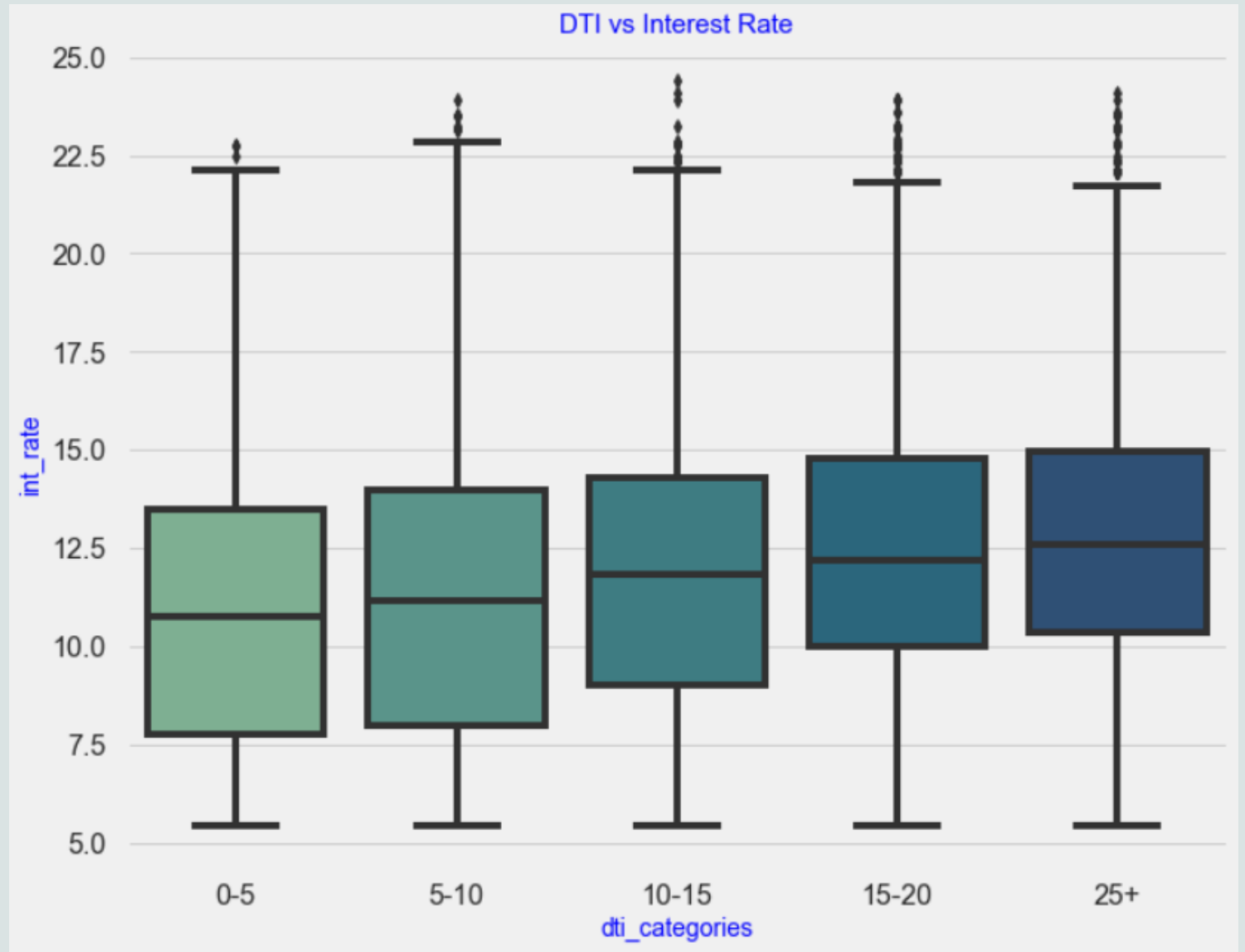
1. It is clear that average interest rate is higher for 60 months loan term.
2. Most of the loans issued for longer term had higher interest rates for repayment.



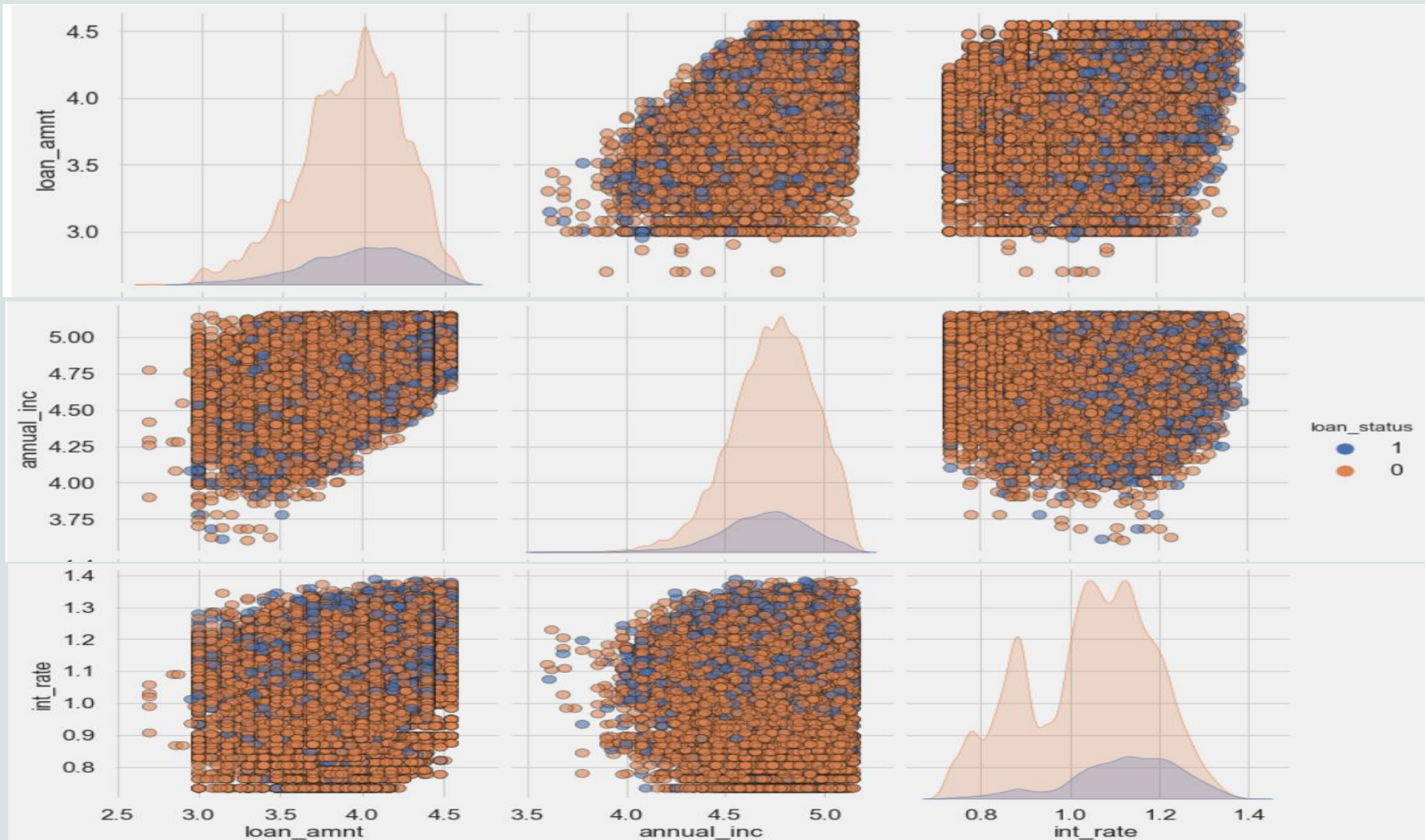
DTI and Interest rate

Observations:

1. We get lower interest rate when the DTI value is low.
2. These does not seem to be significant difference in the loan amount, but we can see that there is a slight increase in the interest rate when the DTI increase



Multivariate Analysis: Loan amount, Annual Income, Interest rate



Observations: The charged-off ratio tends to be higher for loans with increased interest rates. Additionally, there's a slight correlation between higher loan amounts and higher annual income. Yearly trends indicate a rising count of charged-off loans. Moreover, an increase in loan amount corresponds to higher interest rates within the dataset.

Conclusion

- **TERM:** 60 months term are more likely to default.
- **GRADE:** F and G graded loans are more likely to default.
- **SUB GRADE:** F5 G3 G5 G2 sub-grades has ~40+ percentage probability to default.
- **PURPOSE:** Small business owners has more probability to default as compared to other categories.
- **LOAN AMOUNT:** Applicants who requests for loan amount more than 28k are more likely to default.
- **ANNUAL INCOME:** The applicant's whose annual income is less than 20k has high chances to default.
- **INTEREST RATE:** Interest rate with value more than 16% is more likely to default
- **DTI:** If the DTI value is more than 15%, then there is more chances to default.