Lending Club Case Study

SUBMISSION BY :

NAREDLA RAMADEVI

RAHUL KUCHHADIA

SUBMISSION DATE :

25-DEC-2024

Problem Statement

The goal of this case study is to identify the driving factors behind loan defaults using Exploratory Data Analysis (EDA)

- The company faces two primary risks in loan approval decisions: loss of business and financial loss due to defaults.
- Identifying risky loan applicants (defaulters) is key to minimizing financial losses.
- Analyzing past loan data helps to uncover patterns that indicate a higher likelihood of default.
- The goal is to use EDA to identify driver variables behind loan defaults for better risk assessment and decision-making.

Solution Approach

Problem needs to be addressed as giving loans to who don’t repay will have a financial impact.

• **Data understanding** - Identification of the impacting columns for defaulters like interest rate, annual income of the borrower , loan status, home ownership.

Perform EDA

- • **Univariate analysis** -Performed univariate analysis on individual columns to identify the role of the columns independently.
- • **Segmented univariate analysis** - Performed segmented univariate analysis vs loan status to identify which group has defaulters.
- • **Bivariate analysis** - Performed bivariate analysis for loan status vs other grouped columns items to identify which set of group has more defaulters.

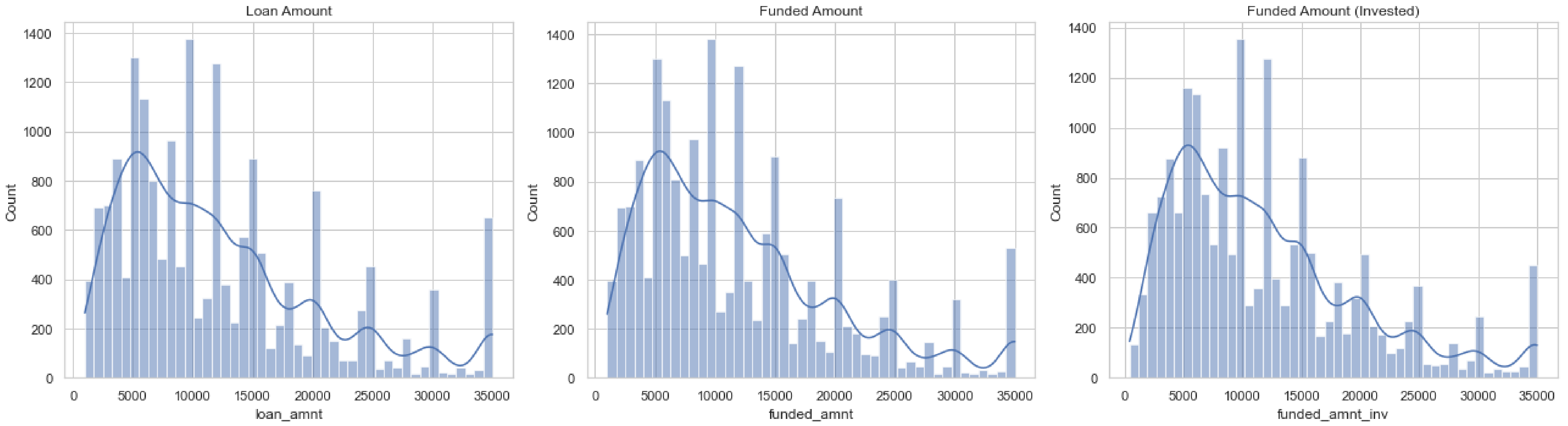
• Identification of factors influencing defaulting application based on above EDA

This solution approach gives indication to banks on loan defaulters and help them to build success strategy in customer focused expansion.

Data understanding and data clean up

- **Drop columns with excessive missing values**: Remove columns with a high percentage of missing data to maintain data quality.
- **Drop columns with unique values**: Eliminate columns with unique values (e.g., IDs) as they don't add value to analysis.
- **Remove duplicate rows**: Ensure the dataset contains only unique records to prevent biased results.
- **Cleanup unwanted characters**: Remove special symbols or whitespaces to standardize the data.
- **Convert data types**: Change columns to the appropriate data types for accurate analysis.

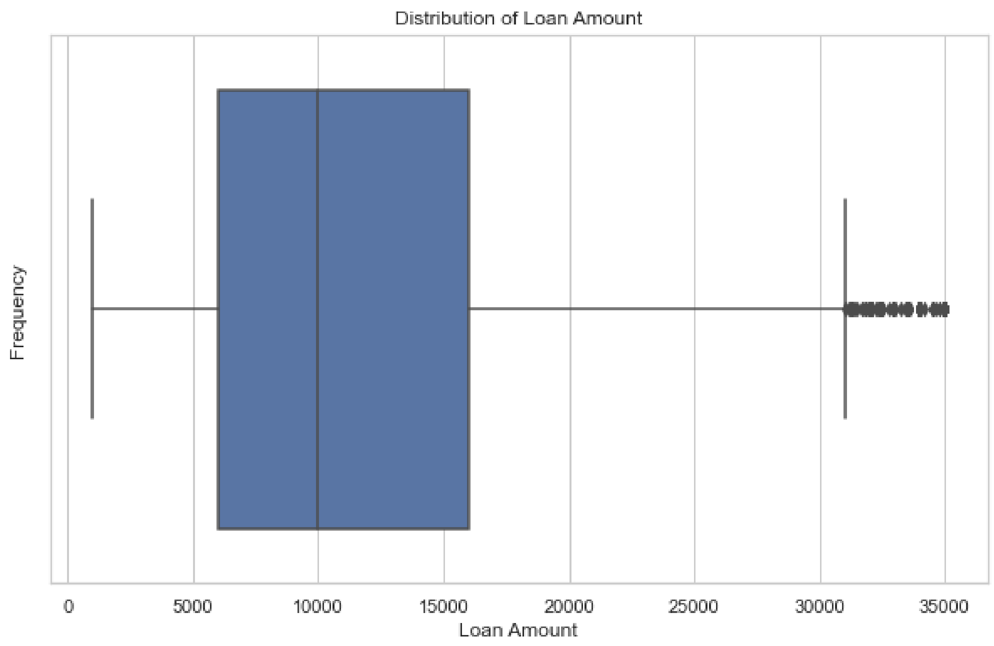
Loan_amt vs funded_amt vs funded_amnt_inv

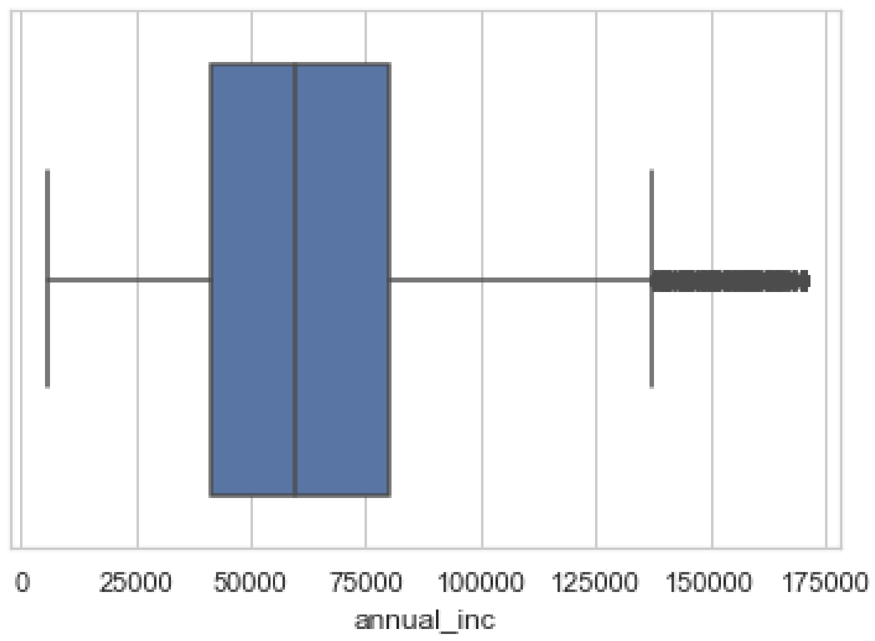


Observations -
There is no significant difference between loan amount to funded amount and funded amount invested
So funded amount and funded amount invested columns can be dropped and use only loan amount for further analysis.

Univariate Analysis

Univariate Analysis - Observations





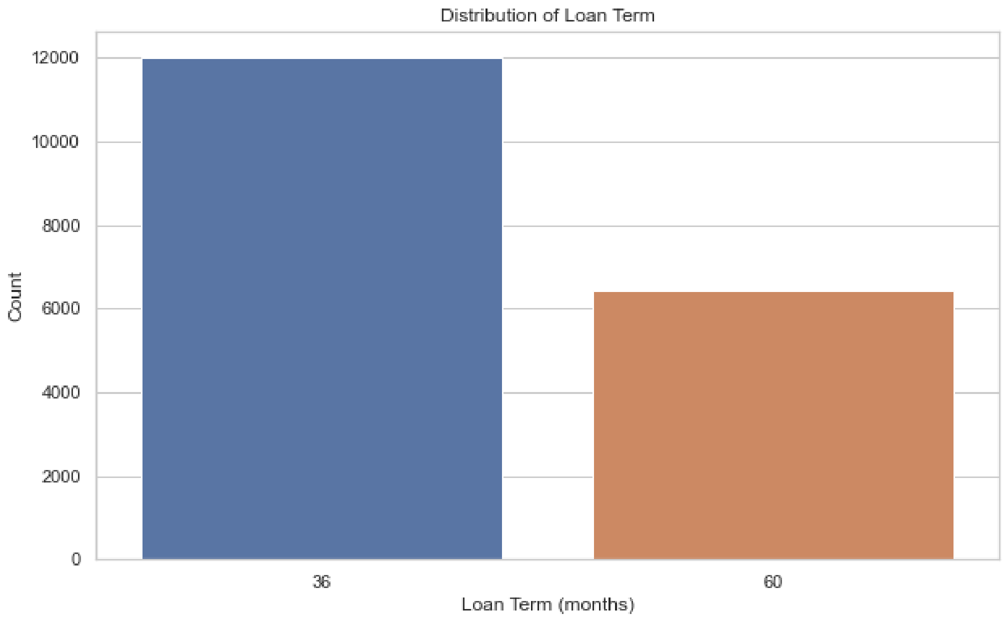
Loan Amount:

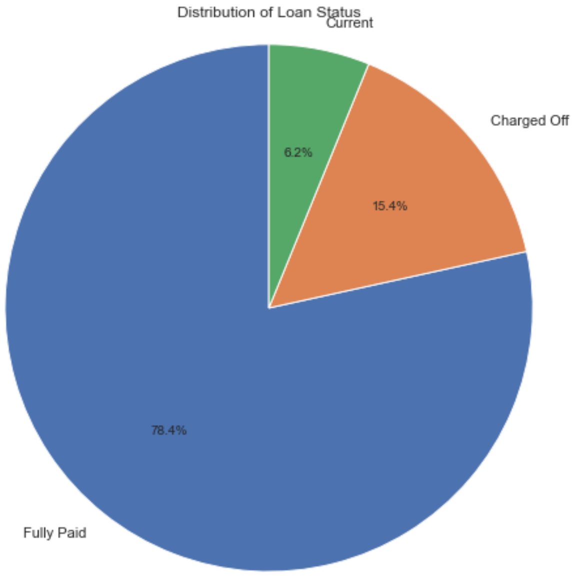
Majority of the loan amount are between 6K and 16K

Annual income:

Majority of the annual income is between 40K and 80K

Univariate Analysis - Observations





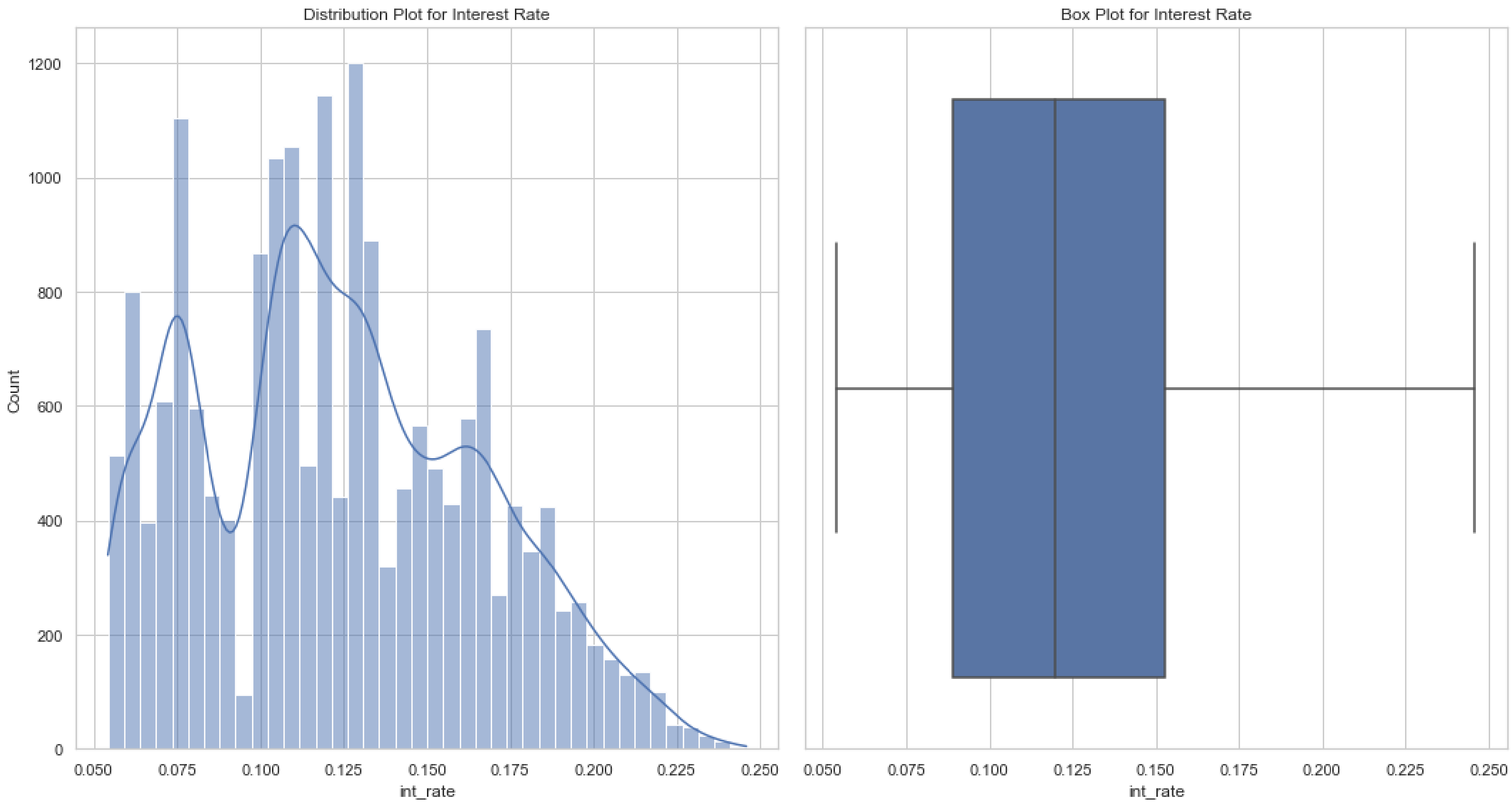
Loan Term:

Majority of the loans are 36 months term

Loan Status:

Majority of the loans are fully paid and few are currently going on.

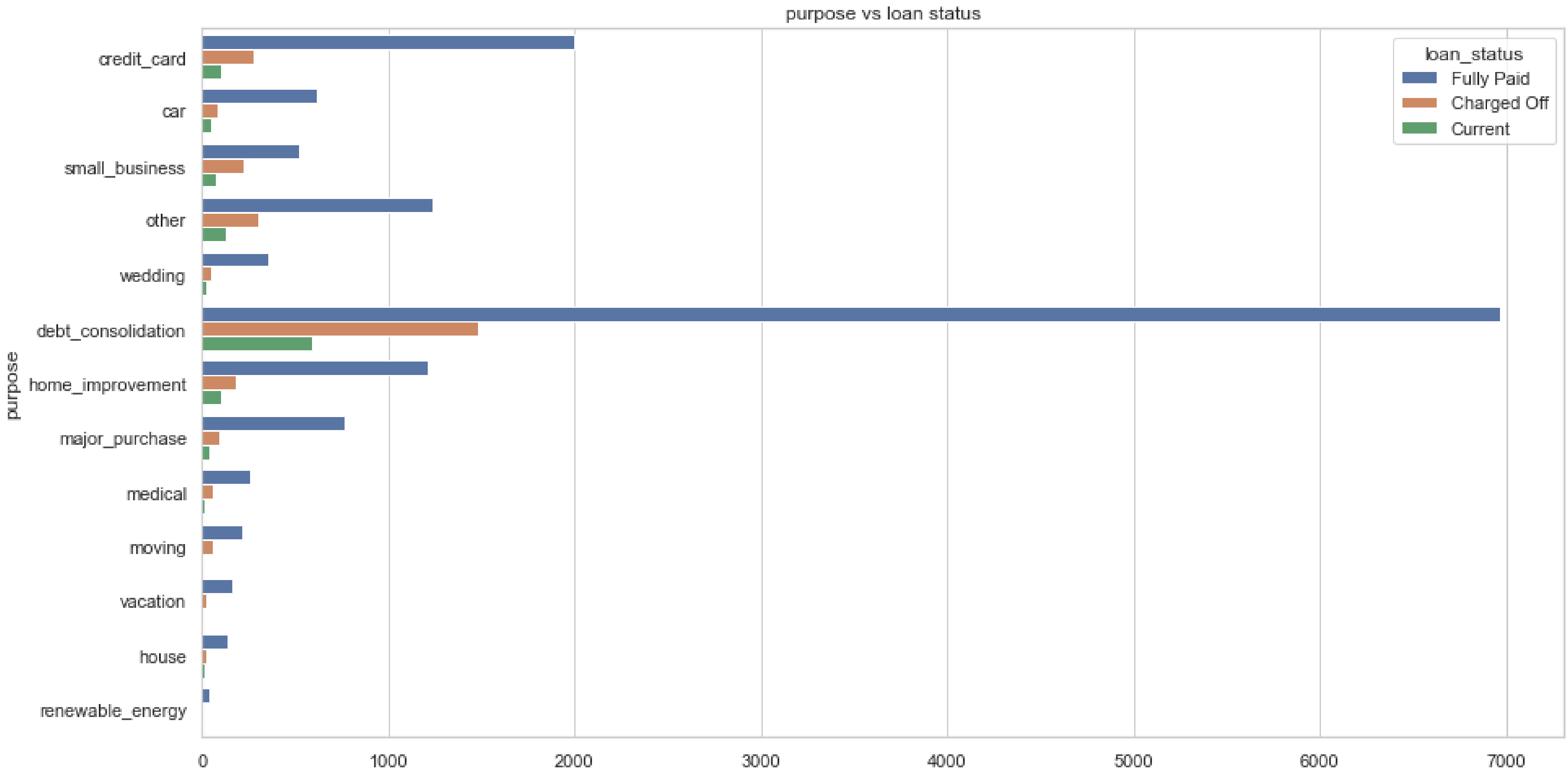
Univariate Analysis - Interest Rate



Majority of the interest rates lie between 10 and 15

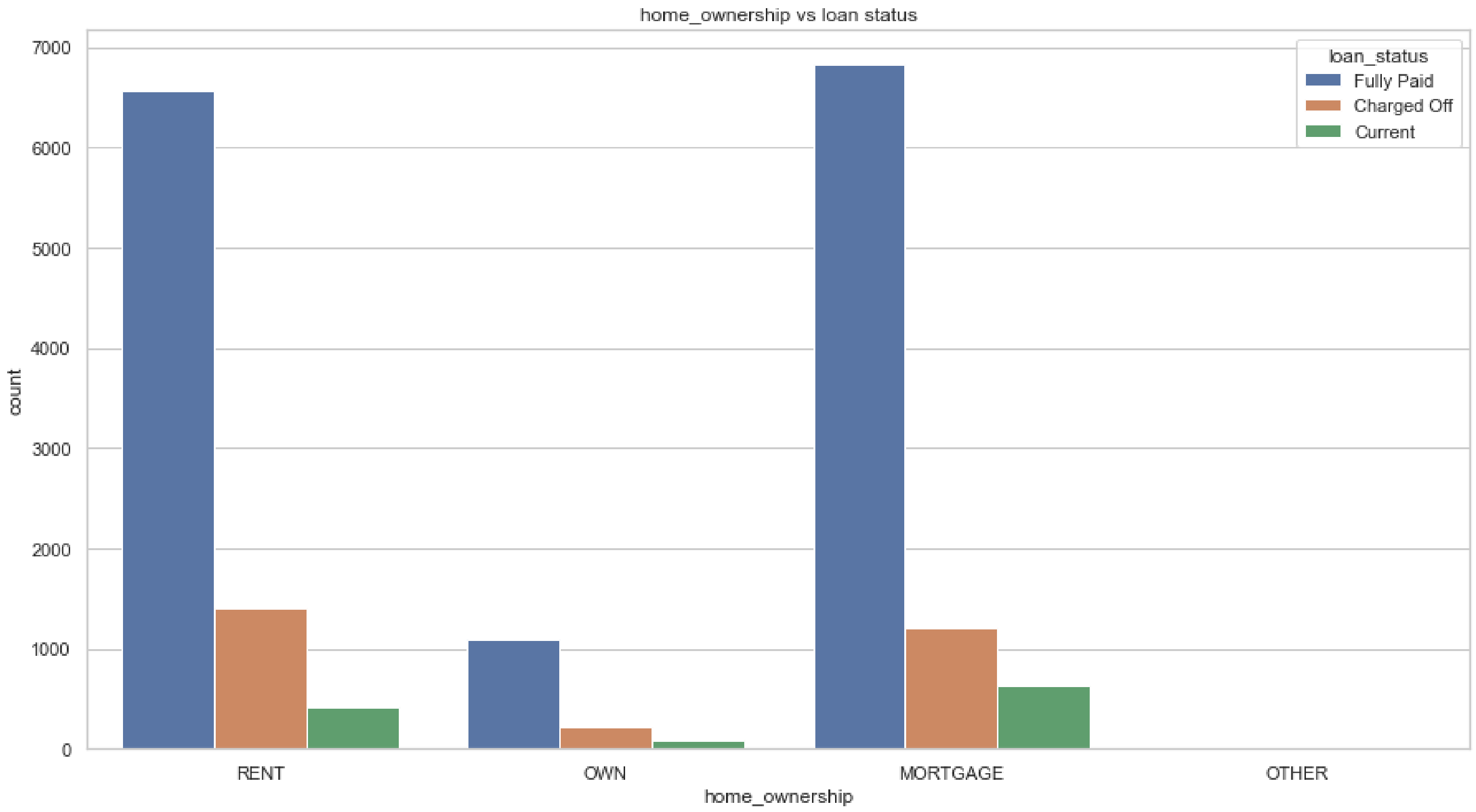
Segmented Univariate Analysis

Segmented Univariate Analysis
Purpose vs Loan status



- Majority of the charged off loans are with **debt_cosolidation** and **credit cards**.
- Loans under renewable energy, house, educational and vacation are least likely to be charged off

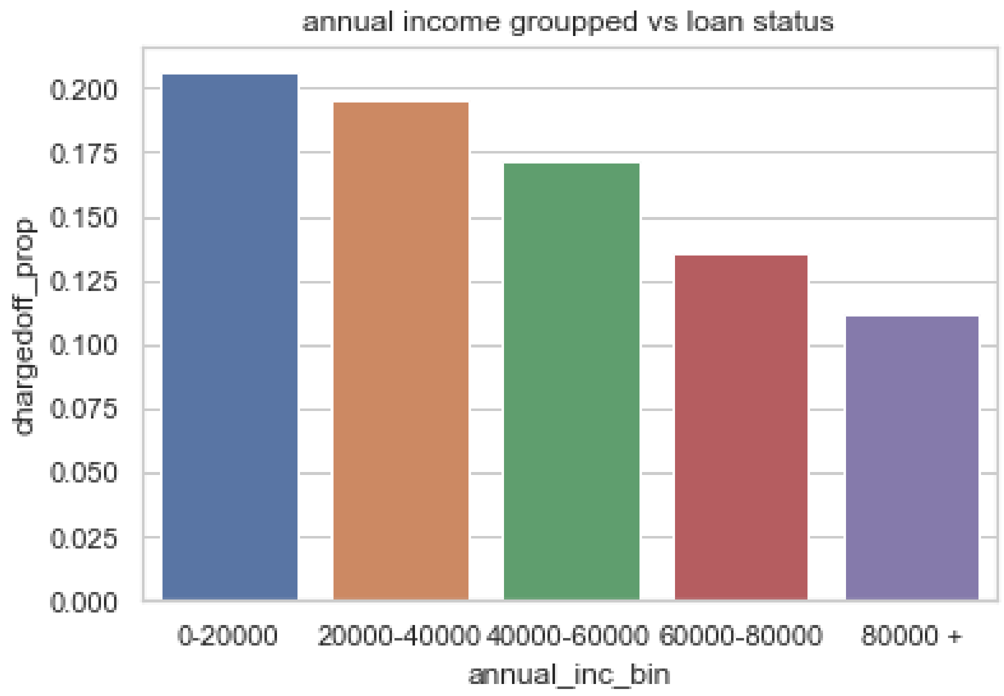
Segmented Univariate Analysis
Home ownership vs loan status

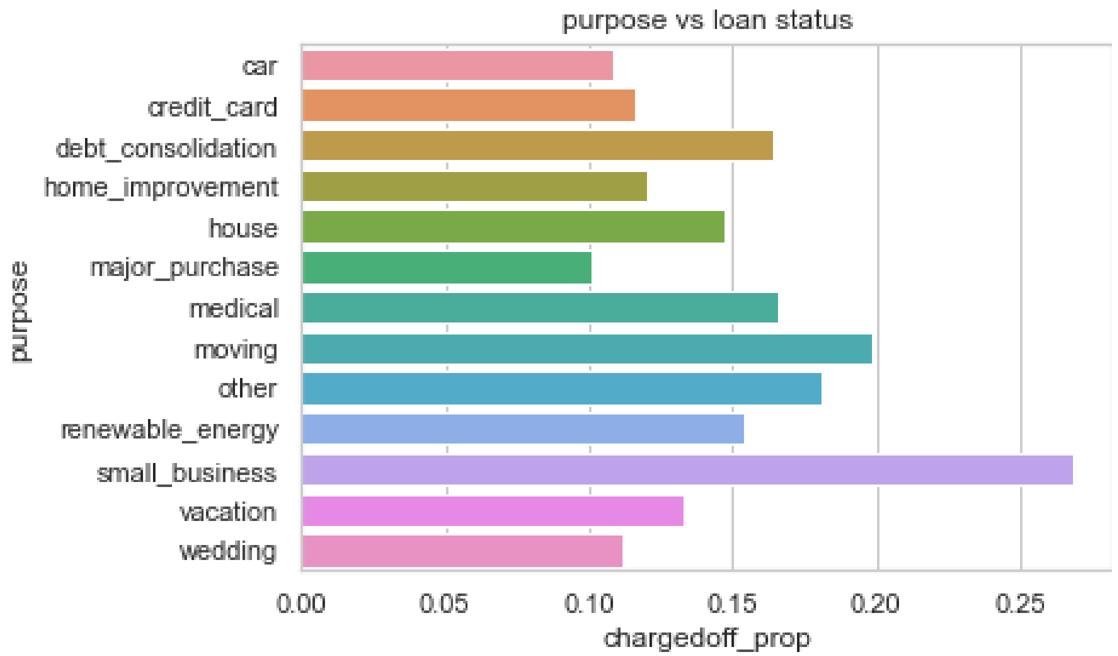


- Majority of the charged off accounts are where houses are in mortgage or in rent.
- People having own house and not in mortgage have very less chances of getting charged off

Bivariate Analysis

Bivariate Analysis - Observations





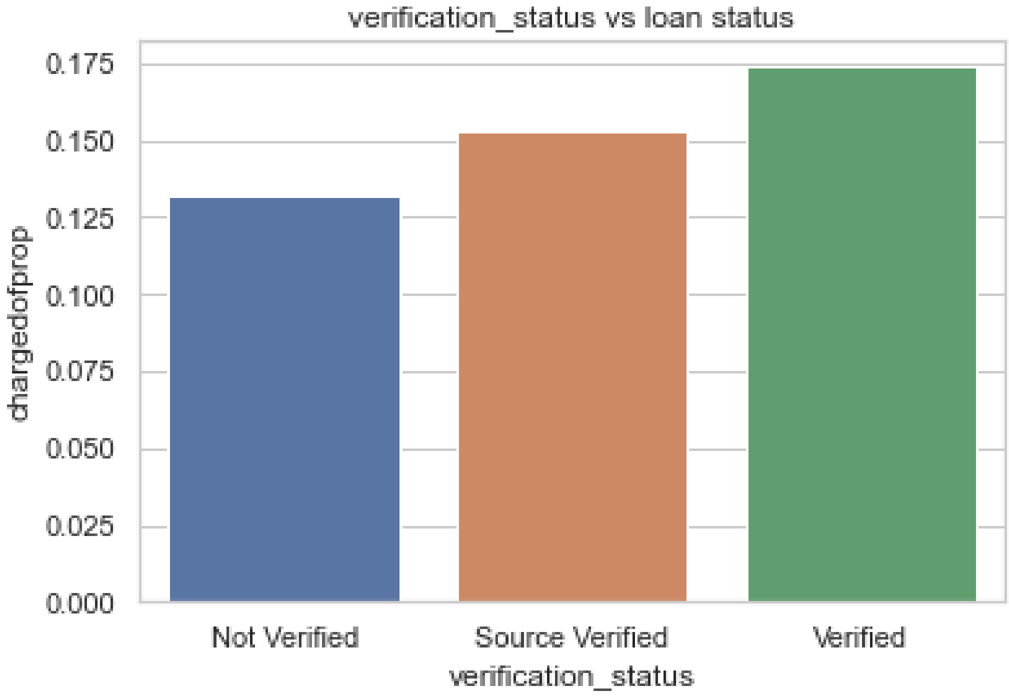
Annual income vs loan status:

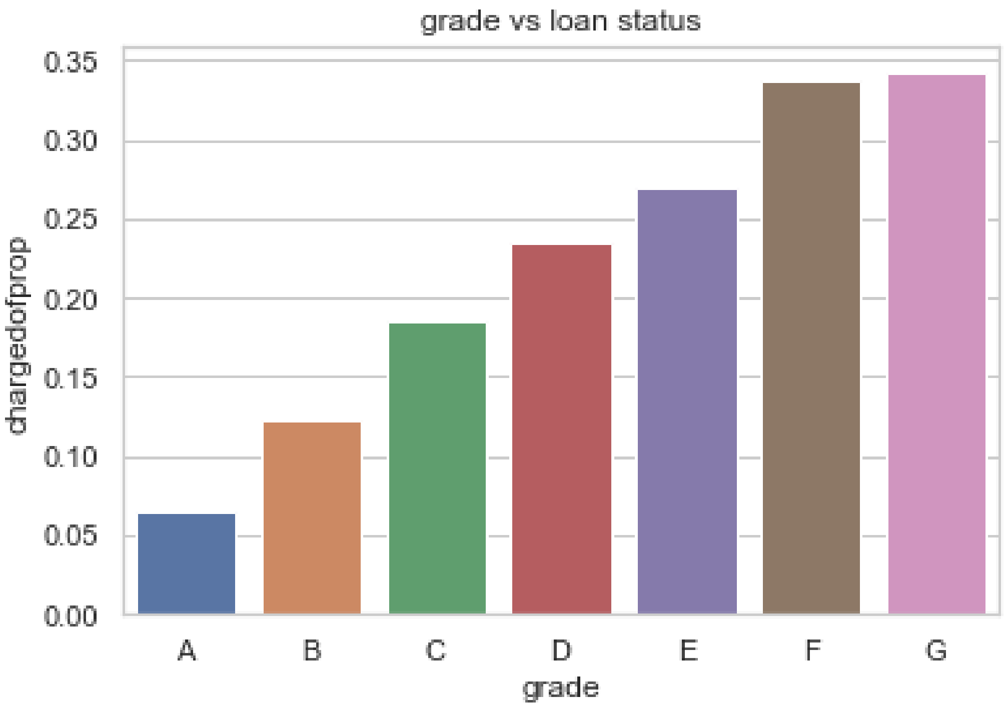
- Low-income grp people are most likely the charged off ones.
- With increasing in annual income, the charged of ones also decrease.

Loan purpose vs loan status:

- Small business purpose are mostly the charged off ones.
- People who have taken for wedding and major purchase are less likely to be the charged off ones.

Bivariate Analysis - Observations





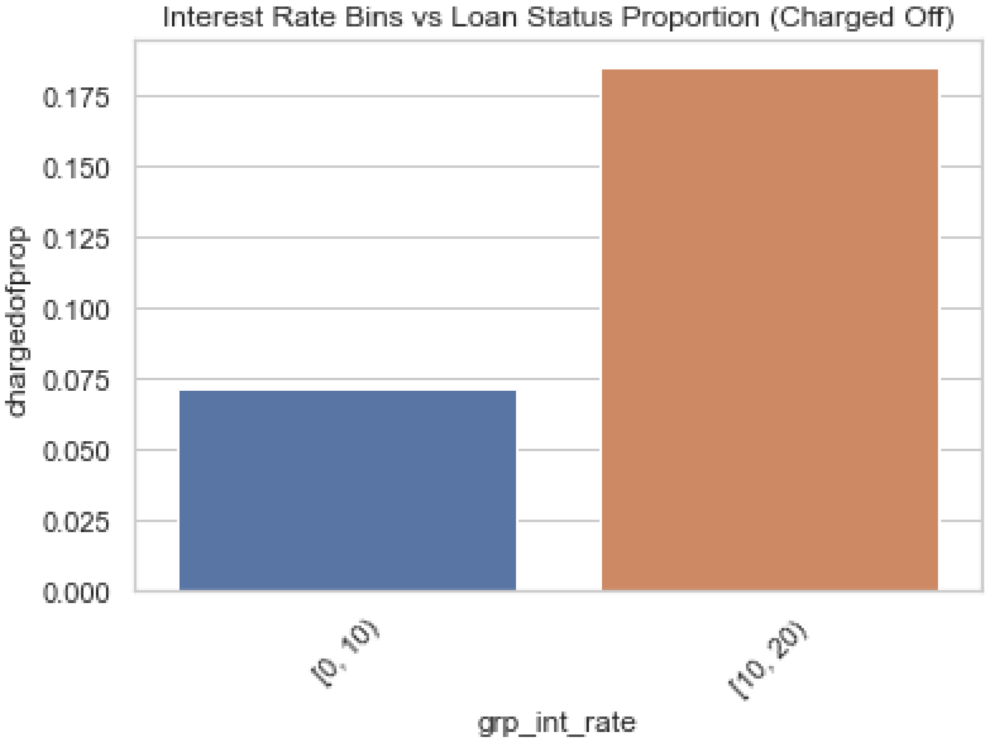
<u>Verification status vs loan status:</u>

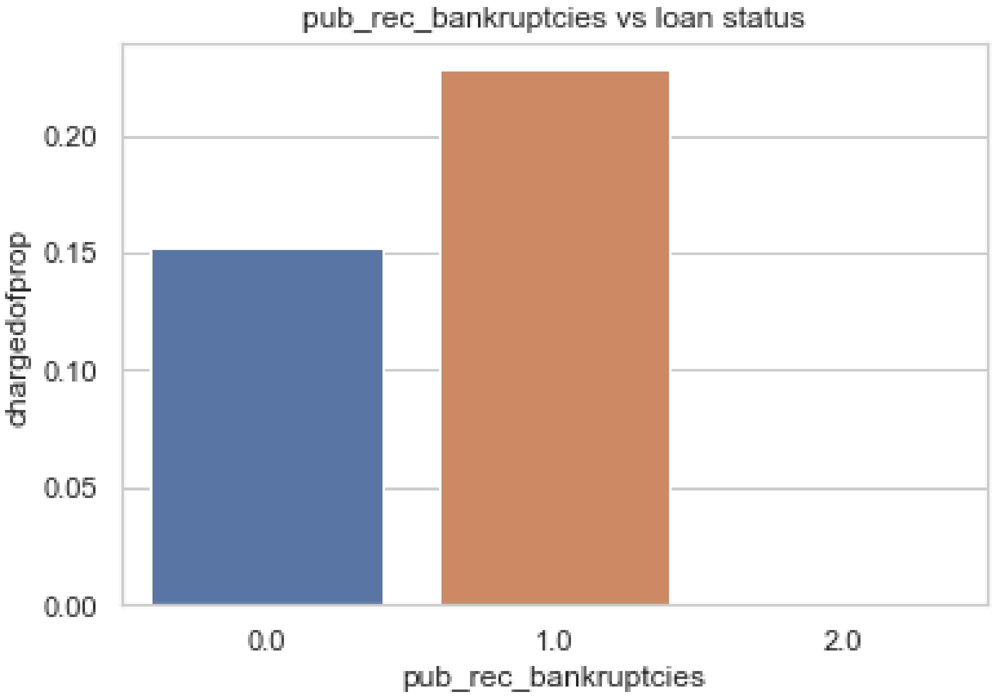
- Source verified loans are most likely to be defaulted, this is significant insight

<u>Grade vs loan status:</u>

- G and F groups are the most likely charged off ones.
- People in A group are less charged off.
- Increasing in grade the chances of getting charged off also increases.

Bivariate Analysis - Observations





<u>Interest rate vs loan status:</u>

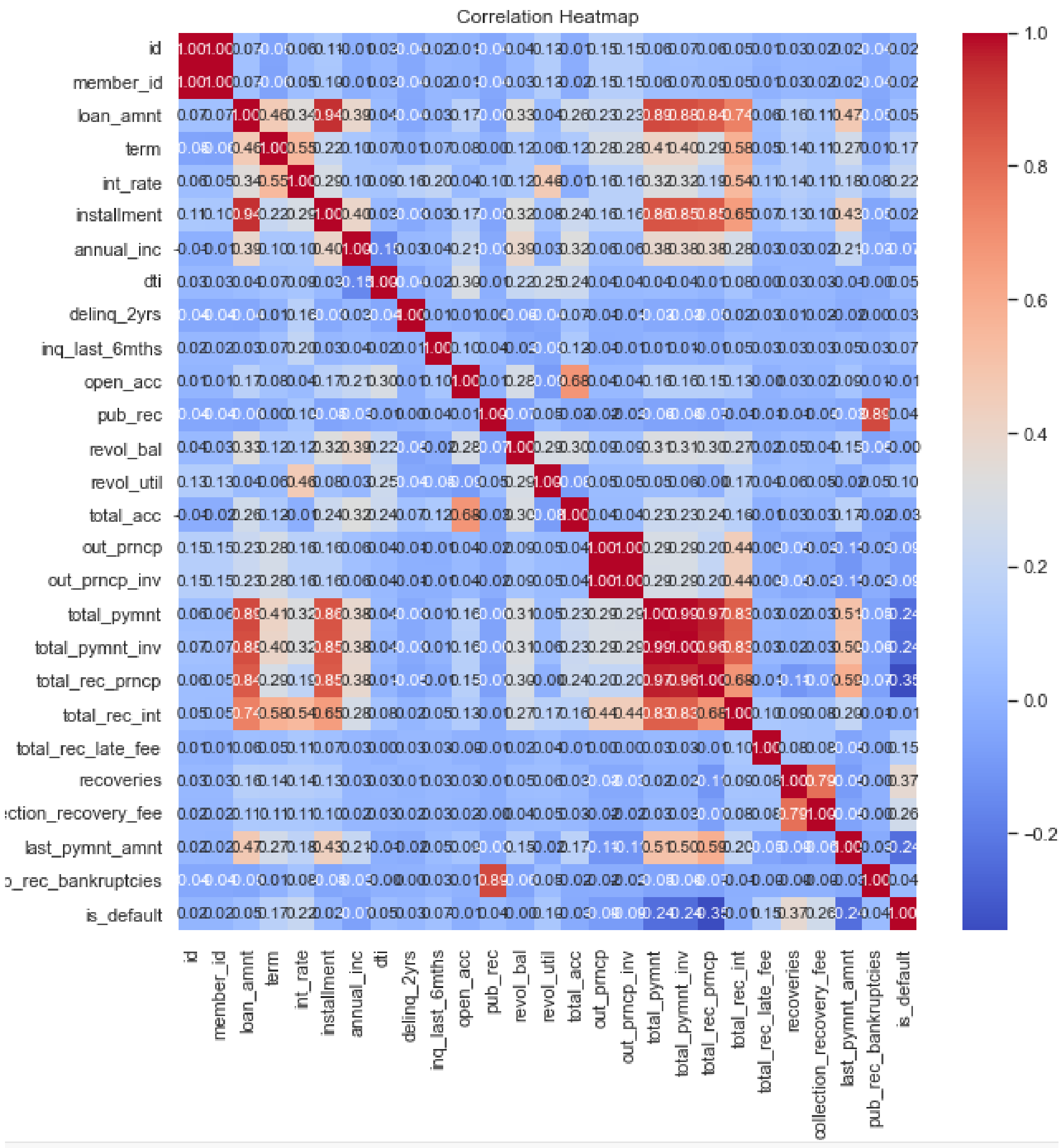
- loan with high interest rate are the ones charged off.
- Less interest rate the charged off rate also decreases.

<u>pub_rec_bankruptcies vs loan status:</u>

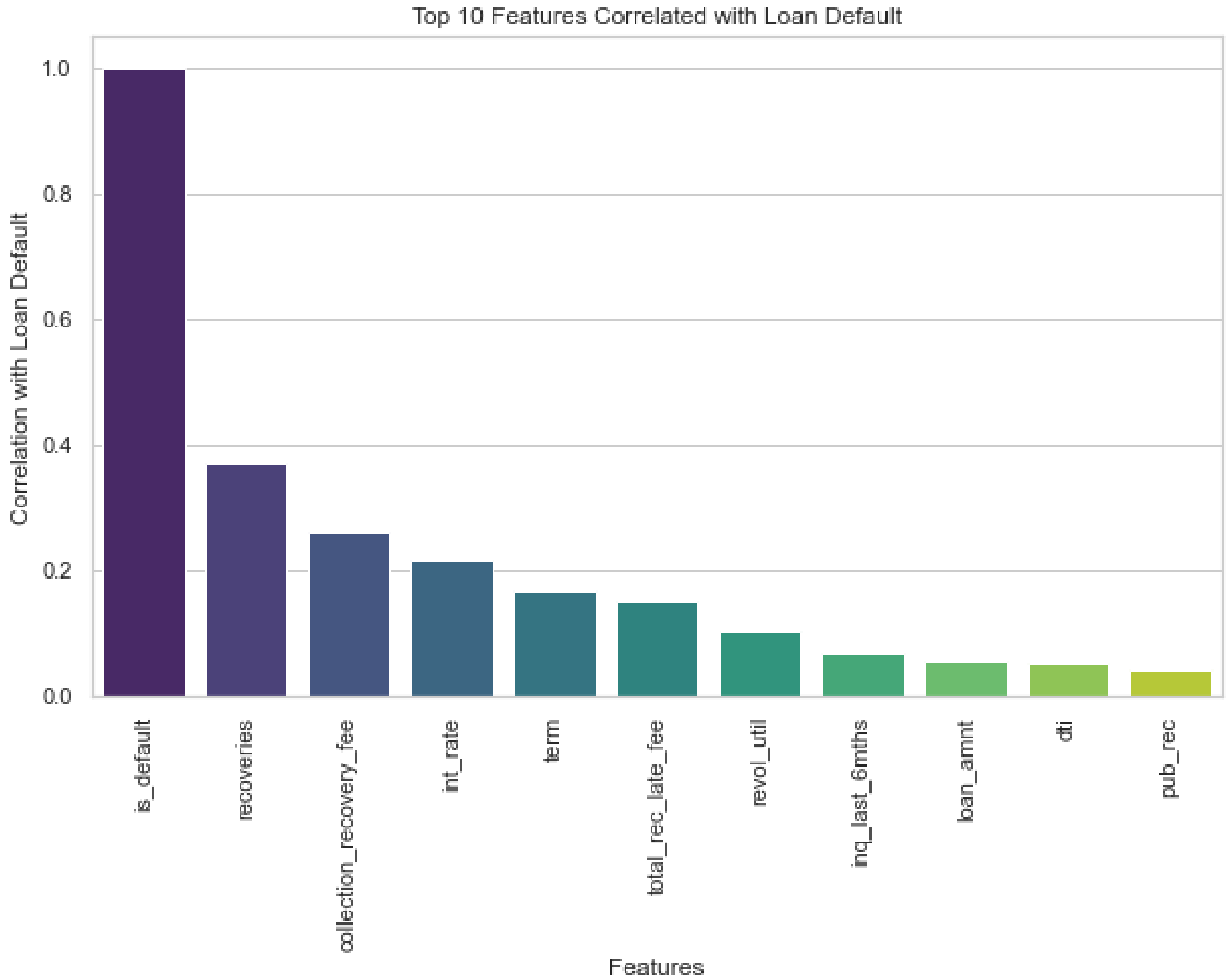
- People who have record of earlier pub_rec_bankruptcies will more like be defaulters again.

Multivariate Analysis

Multivariate Analysis- Correlation heat map



Top 10 features for loan defaults based on correlation map



Analysis insights

1. Higher interest rates on loans lead to a greater likelihood of defaults.
2. A lower annual income increases the chances of default.
3. The higher the group and sub-group (G and F), the greater the number of defaults.
4. Loans provided for small-scale businesses tend to have a higher number of defaulters.
5. An increase in the loan repayment term correlates with a higher rate of defaults.
6. Loans taken for debt consolidation and credit repayment are associated with higher default rates.
7. Borrowers who have defaulted in the past are more likely to default again.

Thank you