Lending Case Study

Upgrad Assignment 1 – Statistics & EDA

Problem Statement

- You work for a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company
- **Expected solution of this problem statement**: To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

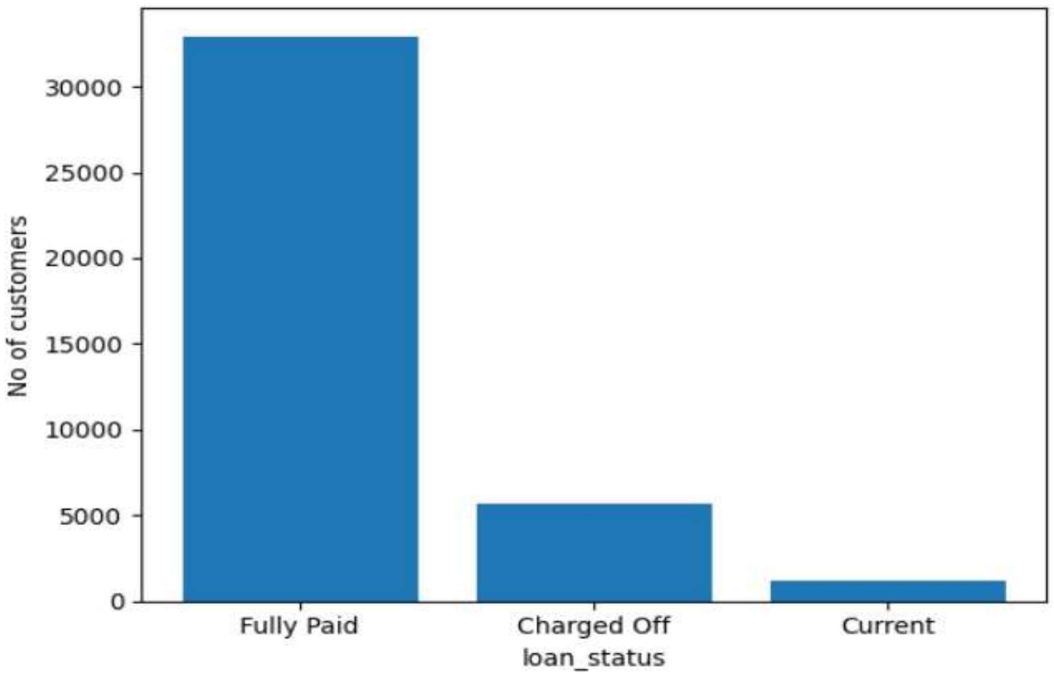
Data UNDERSTANDING & CLEANING

- Loan status, annual income, state, loan amount, loan term and employee length were the key fields selected to perform this analysis
- It is not ideal to remove outliers from financial institution/ banking data. Because defaulting pattern can be found majorly in outliers. Therefore, choosing not to remove outliers.
- Columns with 50% blanks and zeros are removed (54/111 columns were fully populated)
- Few columns data types were corrected – example : interest rate
- Some numerical fields were categorized to perform segmented bivariate analysis

DATA ANALYSIS OVERVIEW

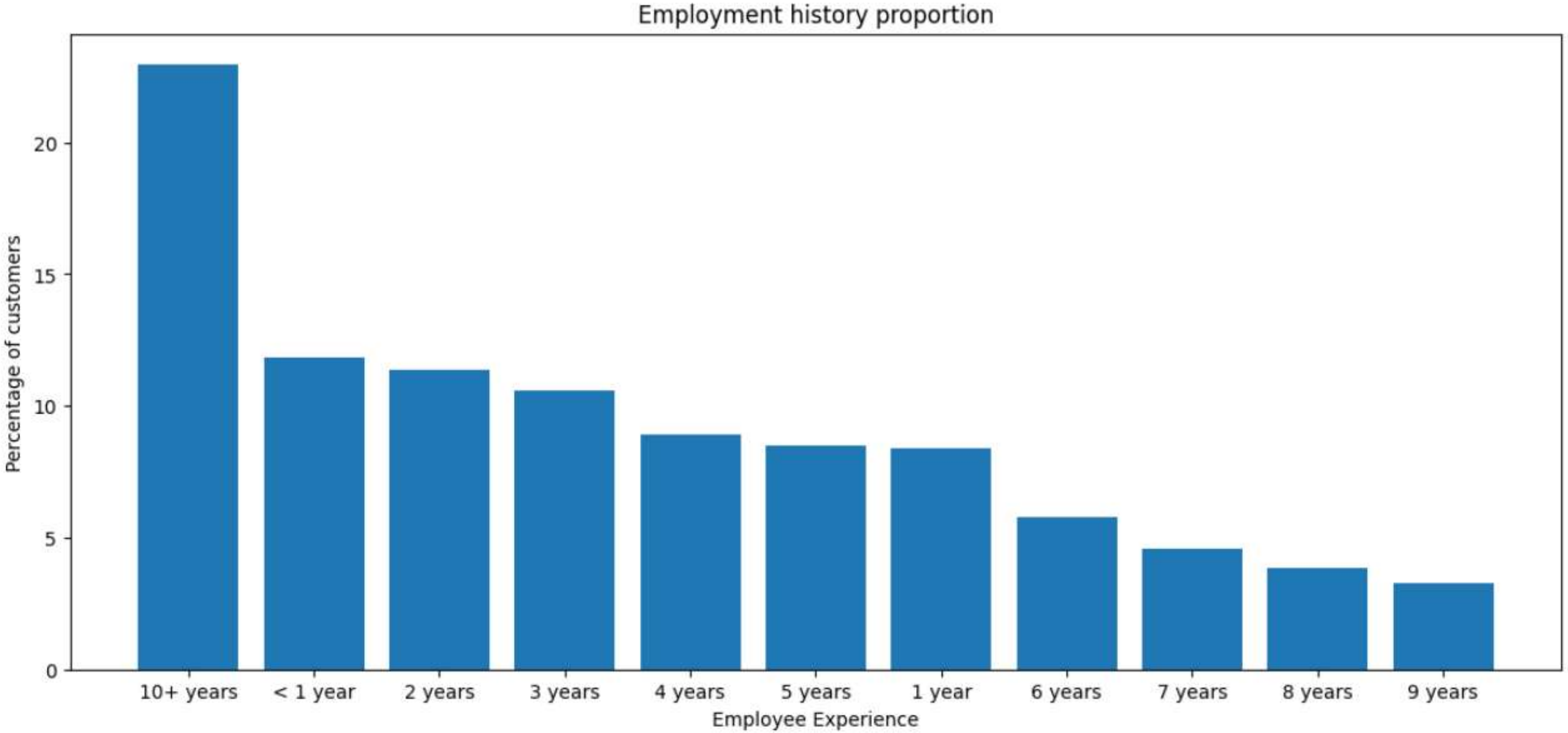
- Univariate Analysis
  - Loan Status Proportion
  - Employment history proportion of the customer base
- Bivariate Analysis + Segmented Analysis
  - State-wise customer charged off pattern
  - Defaulters and Non-defaulters rate aggregated on grades and sub-grades level
  - Distribution of loan amount across grades and sub-grades
  - Distribution of loan amount across loan status
  - Distribution of interest rate across loan status
- Multi-variate Analysis
  - Correlation between loan amount,annual income, DTI

Univariate Analysis – Loan Status



- The Overall default percentage of this lending company is 14.17%
- 83% of the loan take by the customers are fully paid and closed
- 3% of the loan repayment is still ongoing

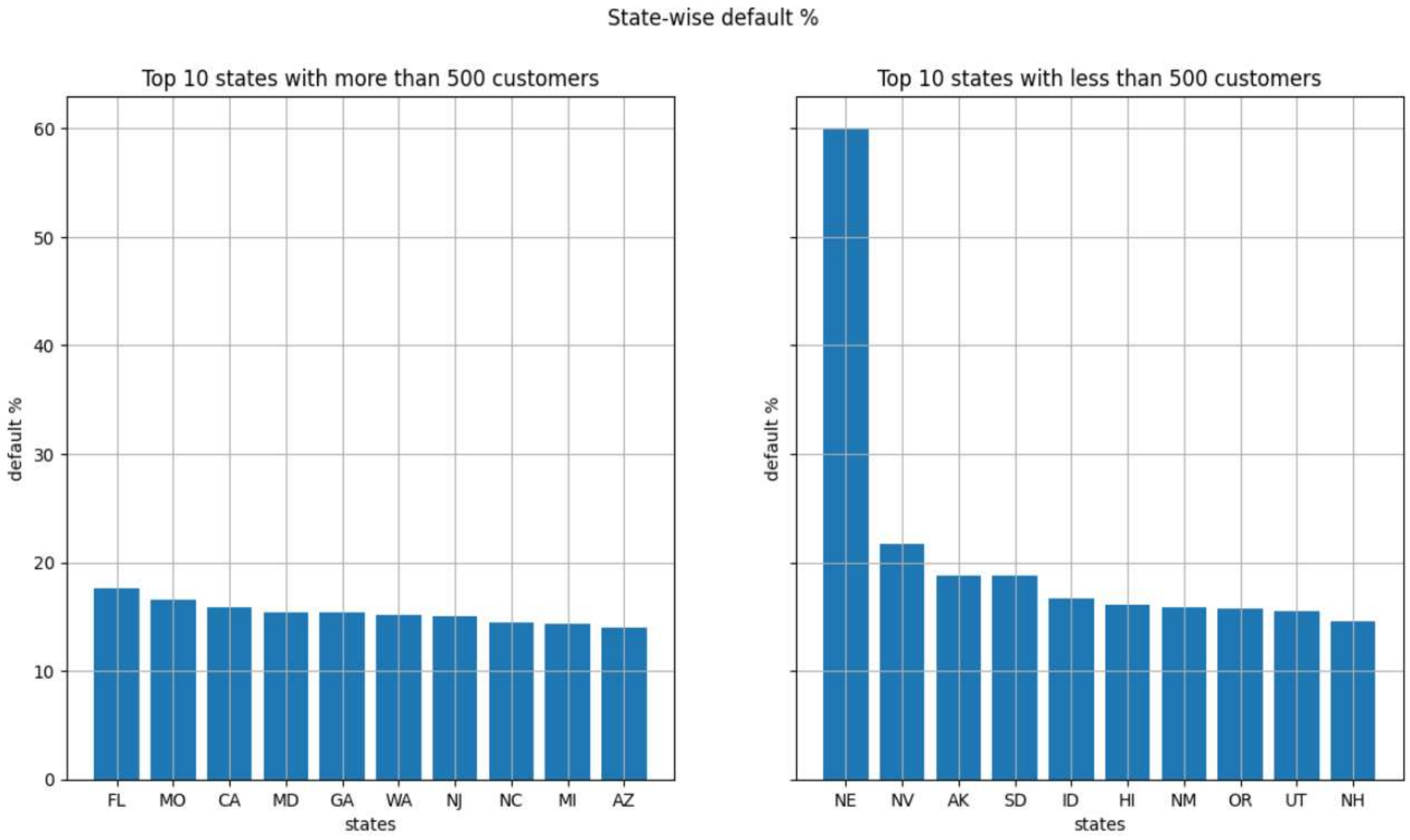
Employment history of customers



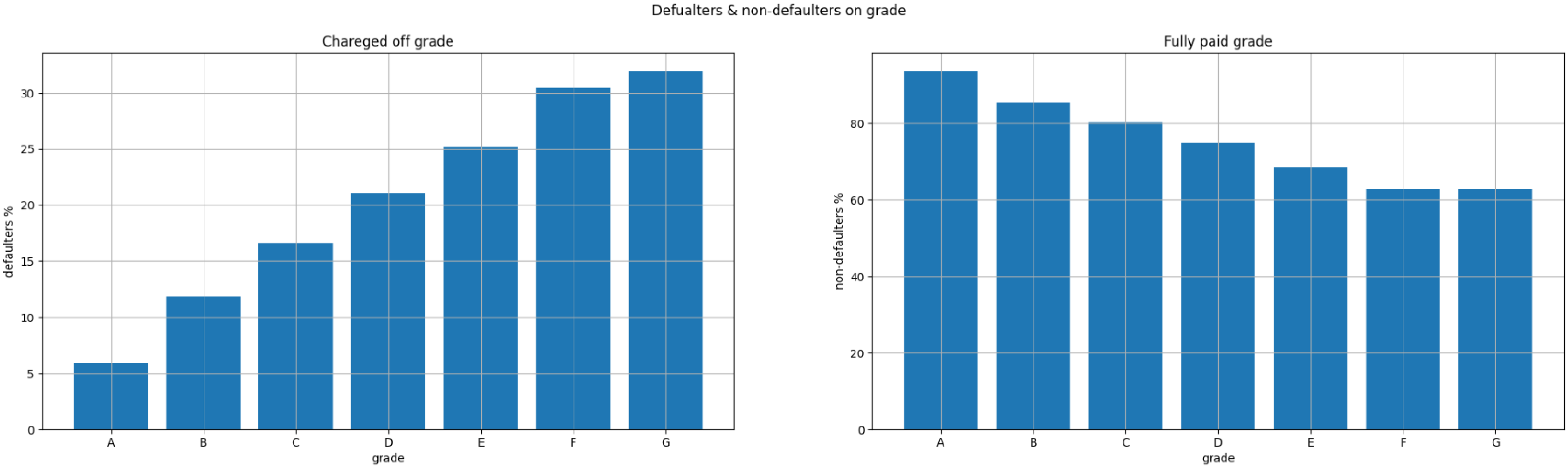
- Customers with more than 10 years+ of work experience applies for loan frequently compared to customers with experience ranging from 0 to 9 years

State-wise default rate
(Bivariate Analysis- Segmented)

- Top 10 states with huge customer base has their default rate within 20%
- Top 10 states with small customer base has an outlier for the state "NE" with 60% default rate
- Company should add additional verifications before approving any loans belonging to state "NE"

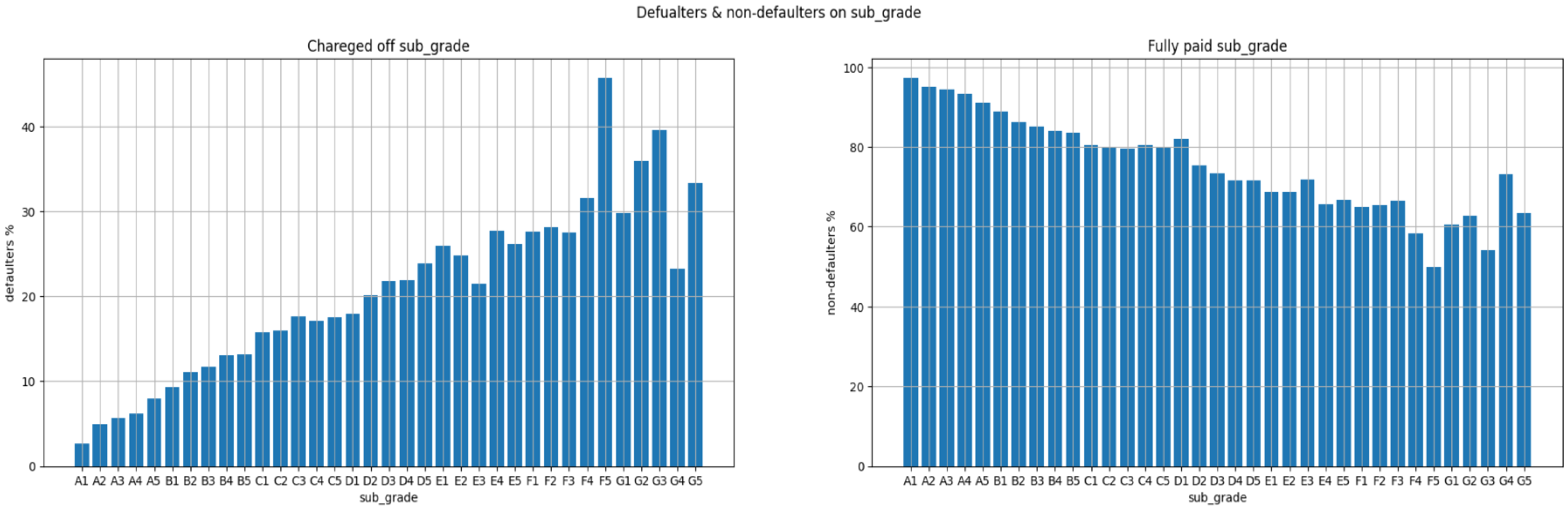


Grades – Defaulters VS Non-Defaulters
(Bivariate Analysis)

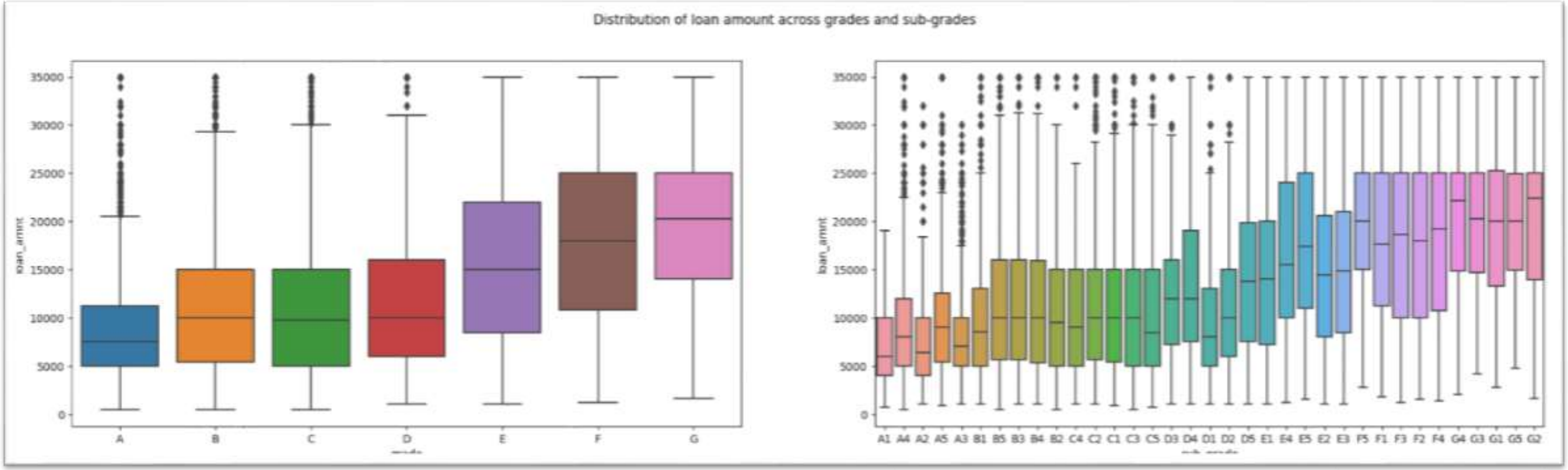


- Grade A has the lowest charged-off/ default rate and the highest non-defaulters rate vice versa for Grade G.
- Increase the loan offerings on the grade A to C.
- A to G grades are categorization by lending company based on loan amount in ascending order

- Sub-grades are descriptive version of grades with four levels on each grades. Similar story as the grades, A1 – A4 in the increasing order defaulters. Onboarding or approving loans of A1 category may boost up the profit of the company.



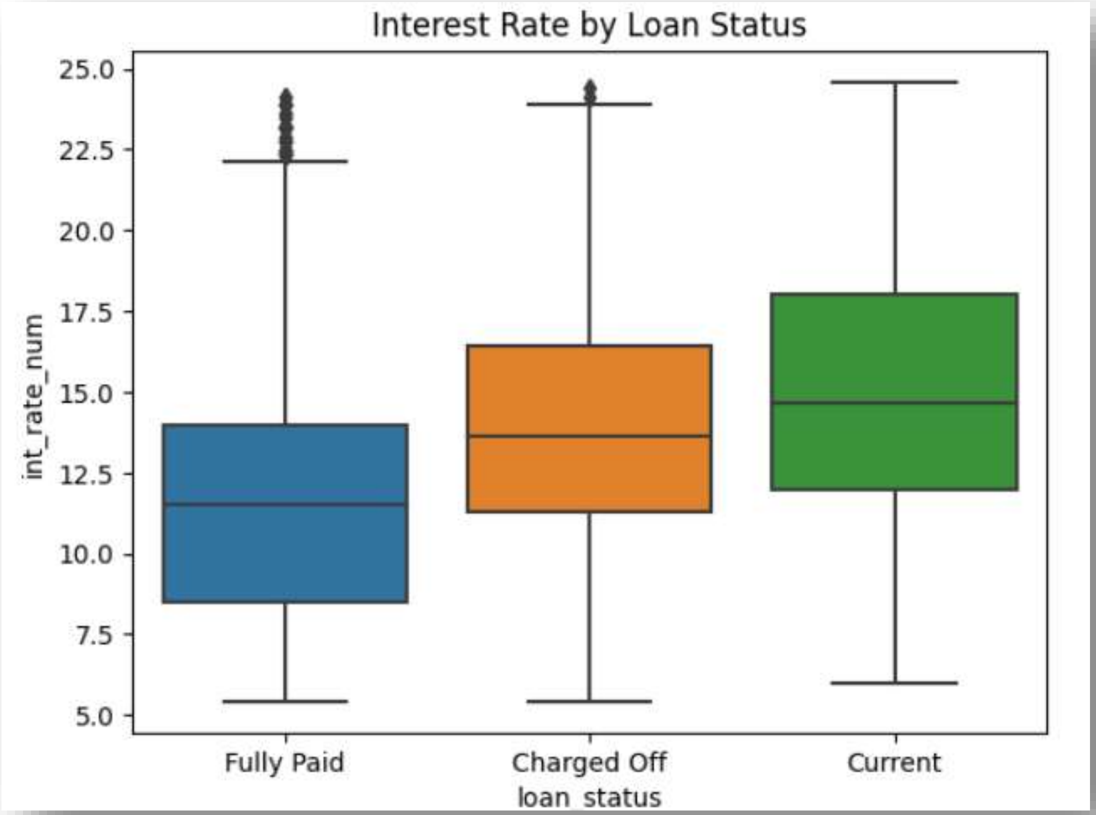
Loan amount Vs grades and sub-grades
(Bivariate Analysis)



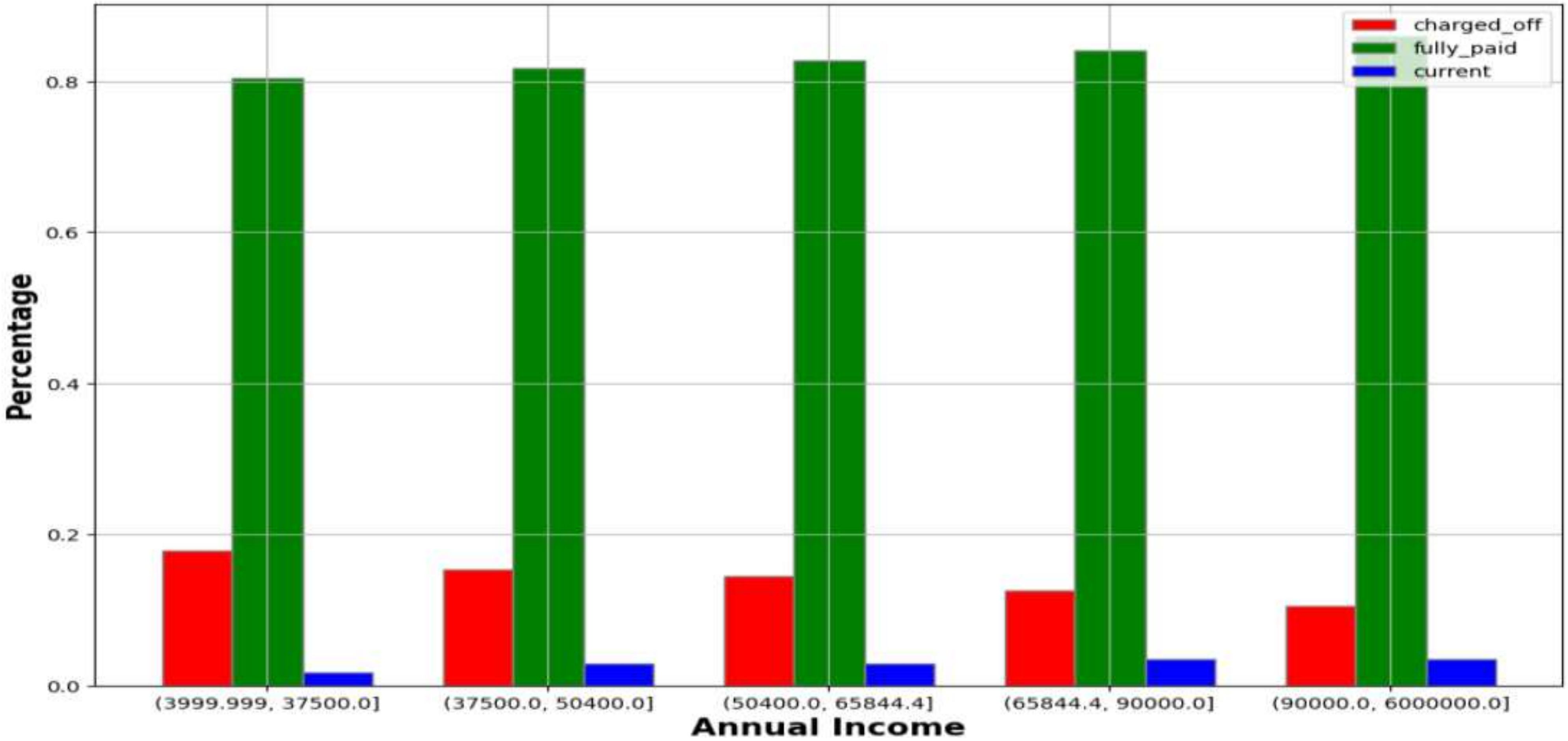
- Higher loan amount are lying under the grade and sub-grades of G's.
- Increasing the number of loans offered under A to C categories would bring more profit as the financial loss involved there is comparatively lesser than other grades

Interest Rate Vs loan status
(Bivariate Analysis)

- On an average defaulted customers has higher interest rate. This shows imposing higher interest rate leads to increased default rate
- There's a risk involved in ongoing (current) loan repayment as the rate of interest is higher. To motivate the customer to full repayment of loan, the company can minimize the interest rate after certain amount of principle amount coverage



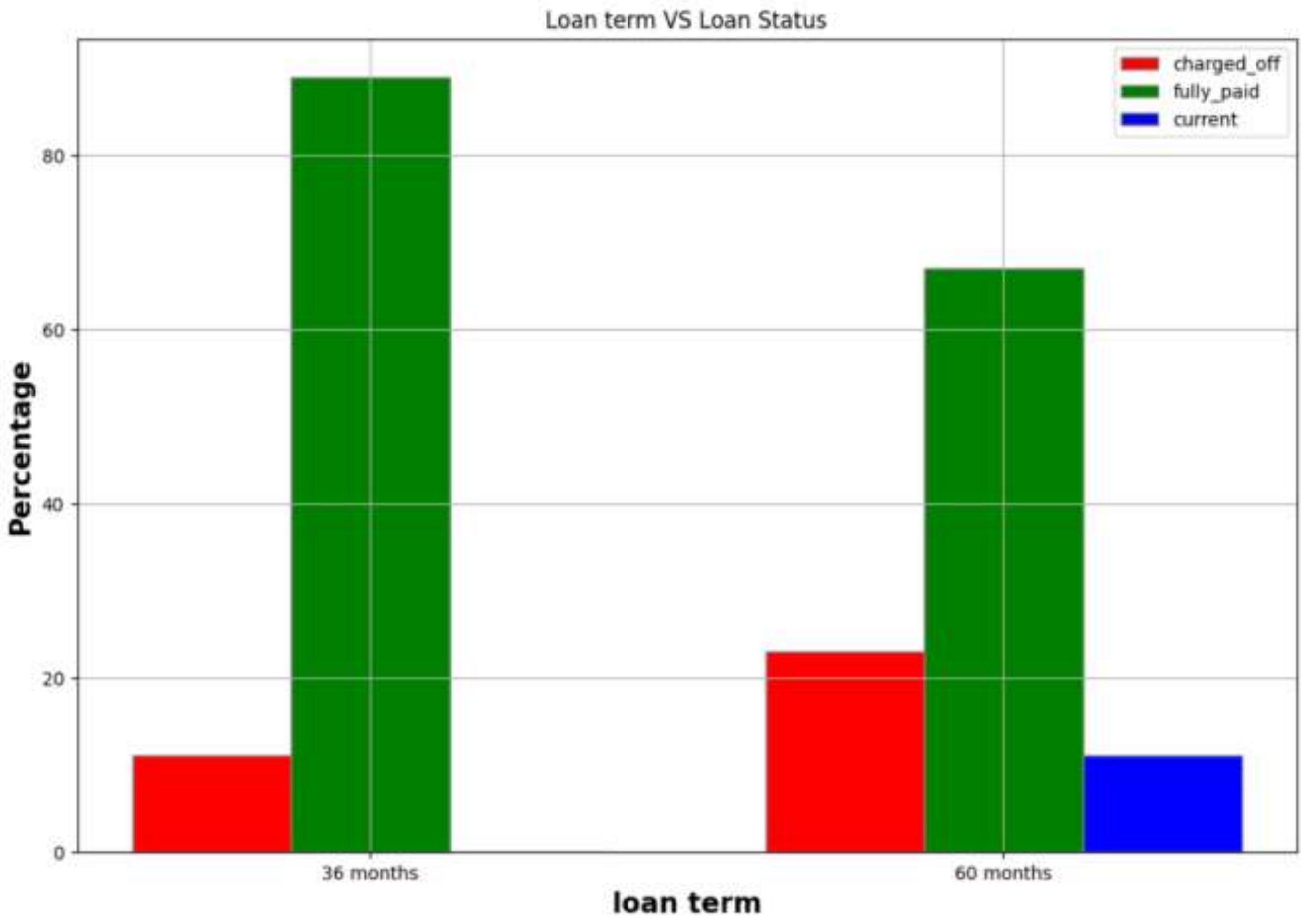
Loan status Vs annual Income categories
(Bivariate Analysis)



It is observed that low income category seems to have slightly higher default rate – Additional checks are recommended

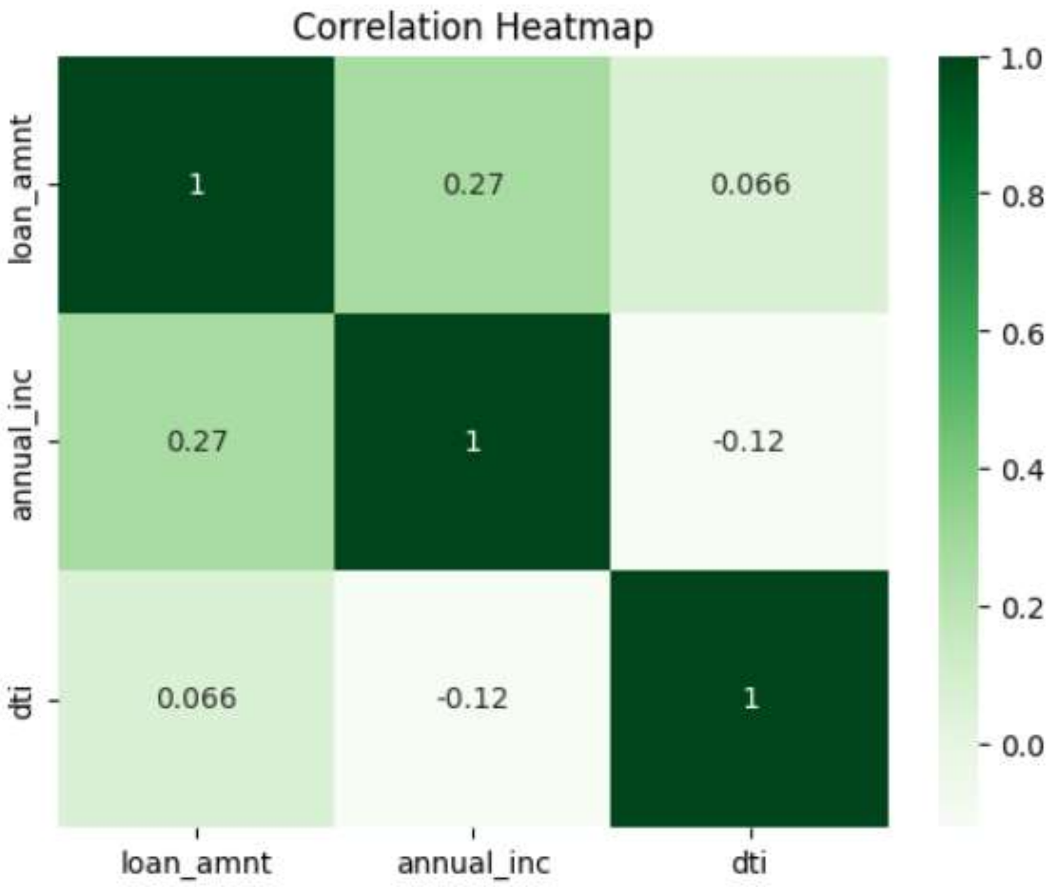
Loan Status Vs Loan term

- Loan repayment rate is higher for 36 months term.
- No ongoing loan repayments shows this area needs more investment in the future to bring profit the company.
- The default rate for 36 months is twice as less as compared to 60 months.



Correlation-plot
(Multivariate Analysis)

- There's a weak relationship seen amongst these three variables.
- Ideally, negatively correlated DTI and loan amount would fetch more gains to company.
- Generic understanding or assumption is annual income influences DTI and DTI influences loan amount. However, according to the distribution of the given data, it is uniform.



Recommendations/ Conclusions

- Perform additional checks and validations for states with a less number of loans i.e. NE state to ensure minimal defaulters .
- Increase the number of loans offered in the A to C grade categories, as these have a relatively lower default rate.
- Address the issue of high interest rates, which contribute to increased defaults. Consider implementing a reduction in interest rates after a specific duration for ongoing loan repayments.
- Loans with a 36-month term exhibit a lower default rate compared to 60-month terms. However, it is surprising that there are no current loans in this category. The company should promote 36-month loan terms and investigate the reasons behind the lack of ongoing loans in this segment.
- Prioritize aligning the principal loan amount with the borrower's annual income and debt-to-income ratio (DTI). For instance, approve higher loan amounts for customers with a DTI between 0 and 5, moderate amounts for DTI between 5 and 15, and lower amounts for DTIs exceeding 15 up to 30.