Lending Club Case Study

-By Parthasarathy Seshadri

Problem Statement

- Lending Club faces a critical challenge in minimizing credit losses stemming from loans to "risky" applicants.
- The objective is twofold: identifying applicants likely to repay for profitability while avoiding substantial financial losses associated with approving loans for those at risk of default.
- The primary focus is on making sound decisions in the loan approval process to strike a balance between generating profits through interest payments and preventing potential business loss by rejecting creditworthy applicants.
- The overall goal is to assist Lending Club in effectively navigating the delicate equilibrium between financial gain and risk mitigation.

Objective of EDA in the case study

- The objective of employing Exploratory Data Analysis (EDA) on Lending Club data is to uncover meaningful insights and patterns within the loan approval process.
- EDA aims to identify key factors influencing credit losses, enabling informed decision-making to mitigate risks, optimize profitability, and enhance the overall efficiency of lending operations.

Data understanding and Preparation process

- The data given as part of the case study contains information about past **Approved** loan whether they 'defaulted' or not.
- It has 3 status of loan which is Fully Paid, Current and Charged-Off.
- Other columns are related to
  - customer details
  - Loan related
  - Others which do not contribute for this analysis.

Preparation process

Importing data	Removing column with only NULL/NA	Removing column with single value	Removing non significant columns	Data conversion	Striping value to be more meaningful	Removing Outlier	Imputing data for better analysis	Creating Derived columns make analysis easier

EDA Analysis Process

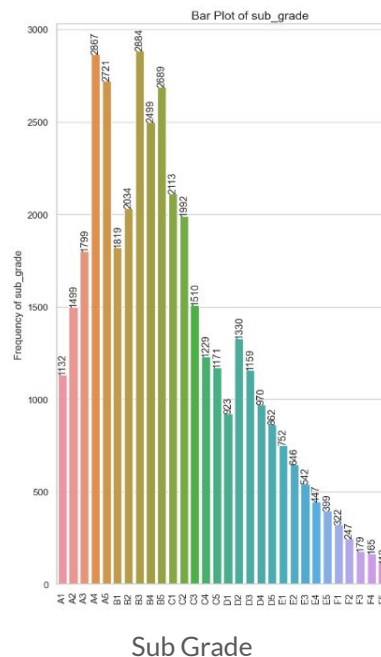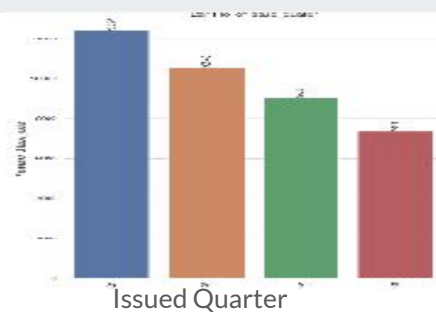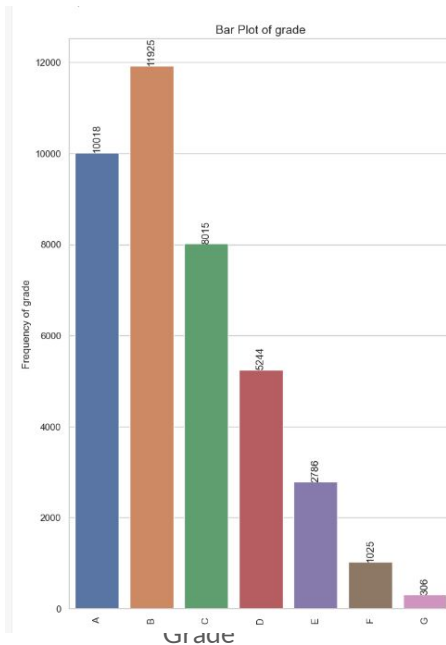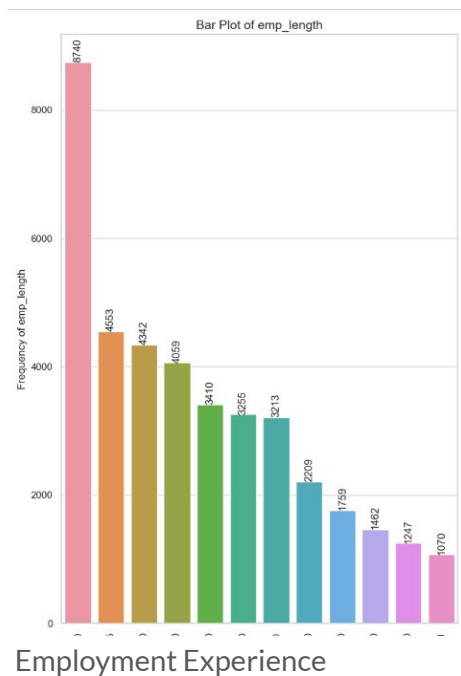| Data Loading & Preparation	Univariate Analysis	Segmented Univate Analysis	Bivariate analysis	Correlation Analysis	Conclusion

Univariate Analysis

Given data set divided into three types of variable

- Ordered category
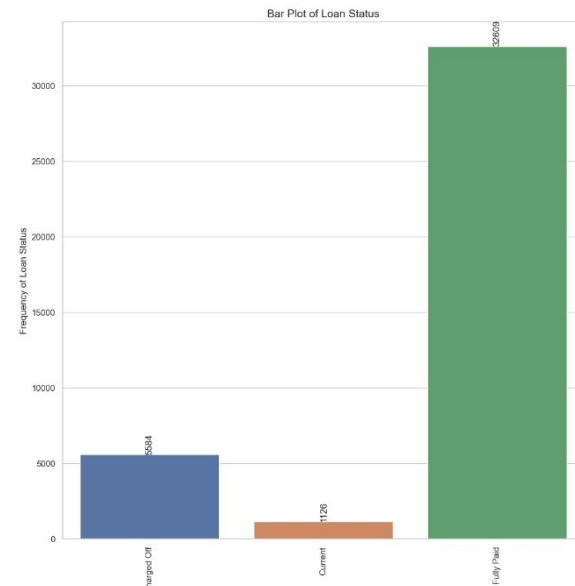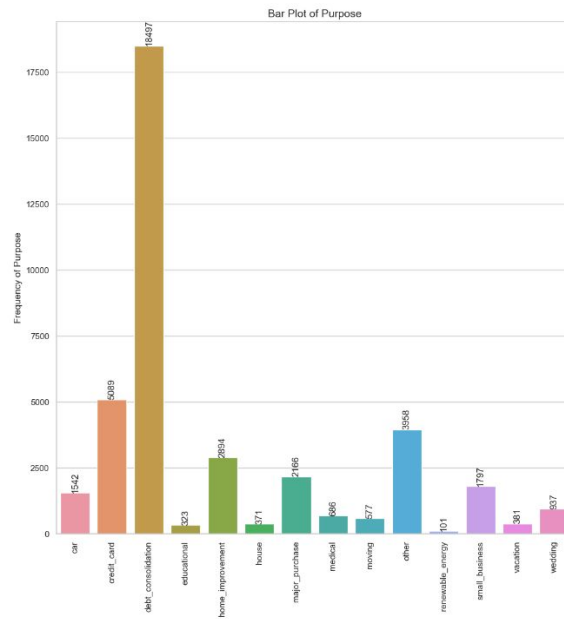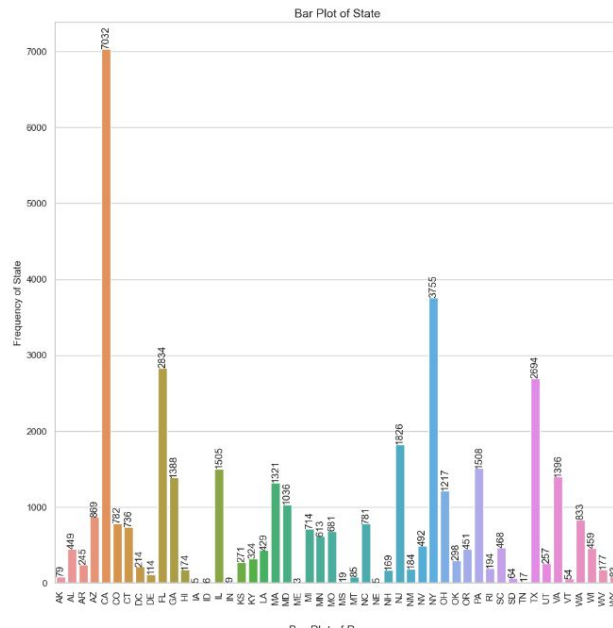- Unordered category
- Quantitative

Ordered Category

Issued Quarter

Issue Month

Employment Experience

Grade

Sub Grade

Issue Year
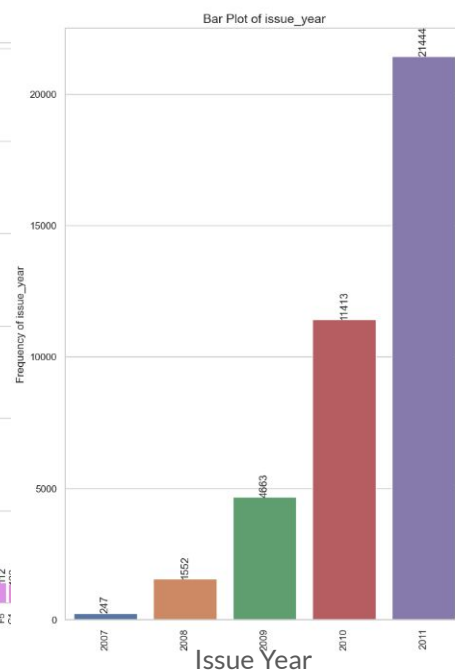
Ordered Category insight

- Majority of loans are with grade 'A' and 'B' compared to rest
- Majority of borrowers have working experience greater than 10 years. 11 kept for Unknown Category.
- Majority of the loans are given in last quarter of the year.
- Majority of the loans are given in the Year 2011.
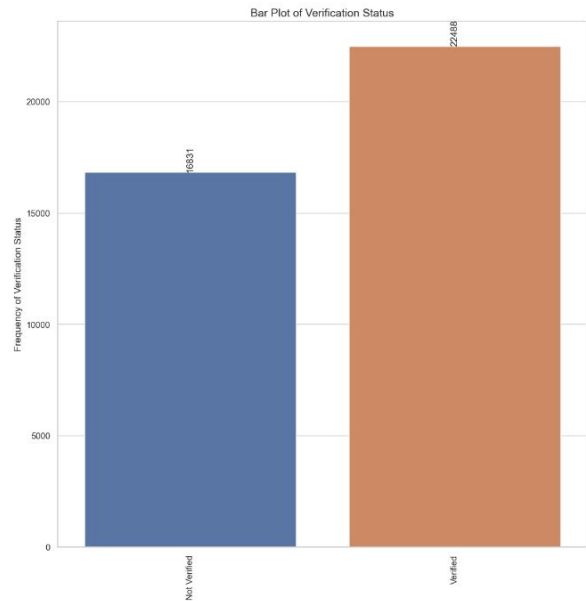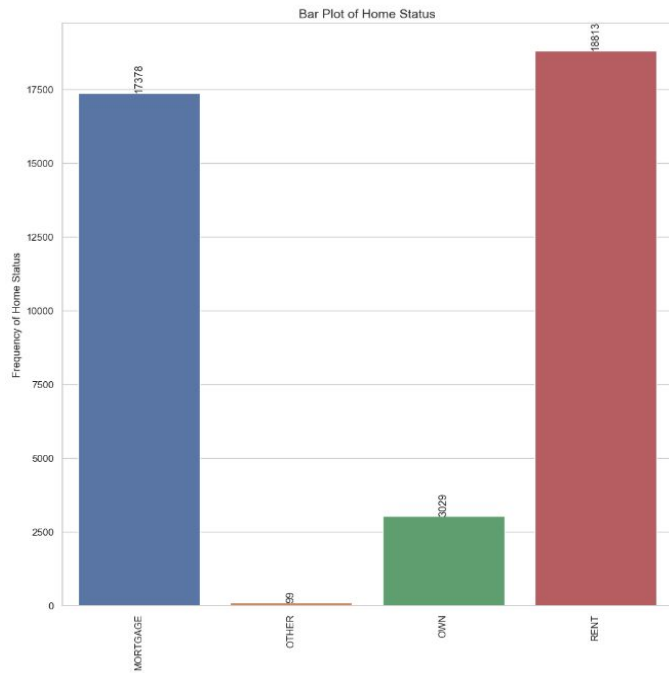- Majority of the loans are given in December.

UnOrdered categorical

# UnOrdered categorical
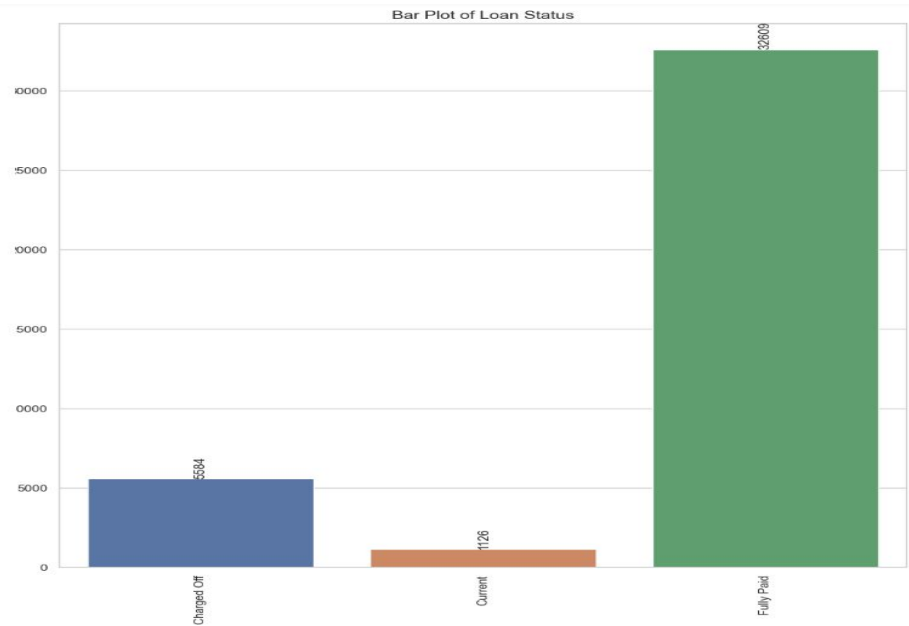
UnOrdered categorical

UnOrdered categorical insights

- The number of defaulted loan is 6 times less than the number of fully paid loan.
- A large percentage of loans are taken for debt consolidation followed by credit card.
- Majority of the borrowers are from the urban States like california, new york, texas, florida etc.
- Majority of borrowers don't possess property and are on mortgage or rent.
- Majority of the borrowers have no record of Public Recorded Bankruptcy.
- only 73% of the borrowers are verified by the company or have source verified.That is potential risk.

Quantitative variables
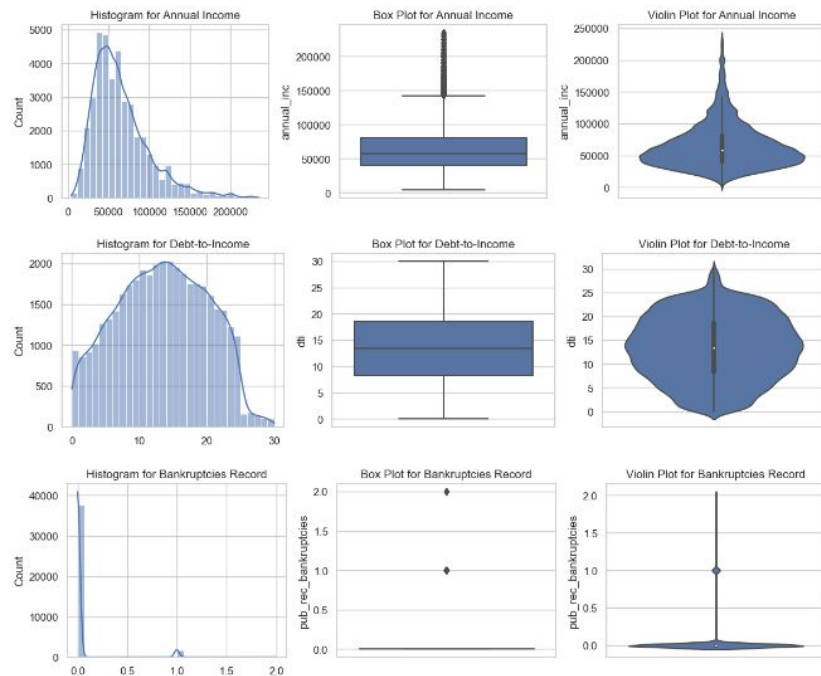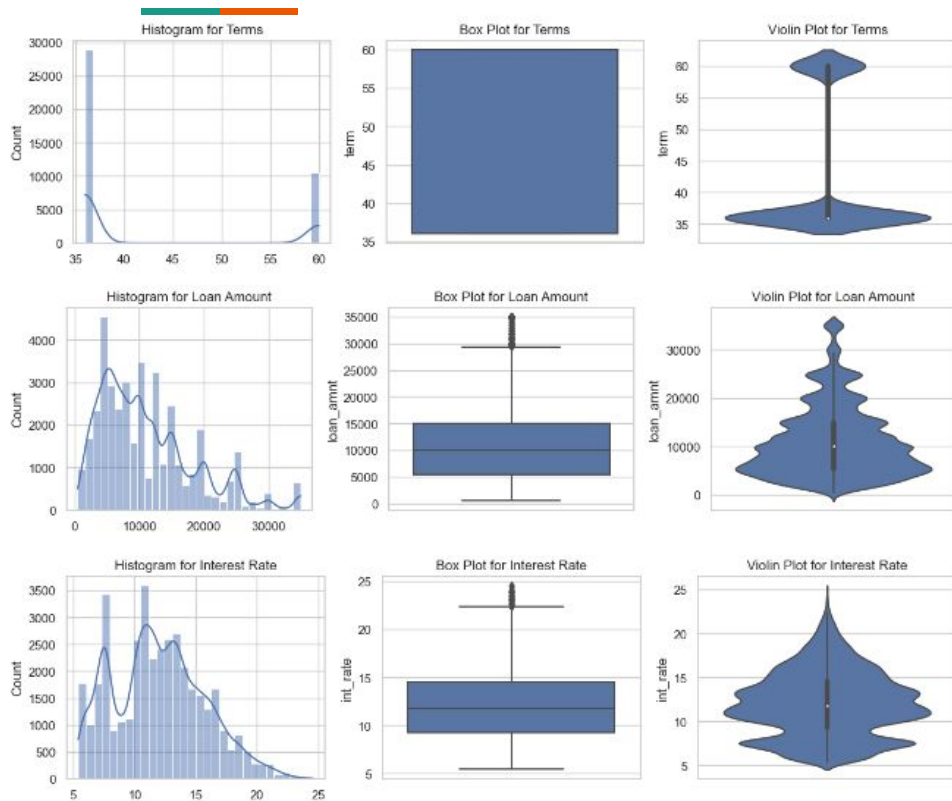
Quantitative variables
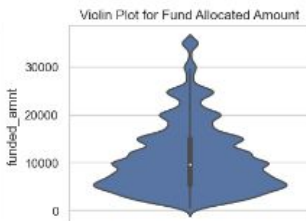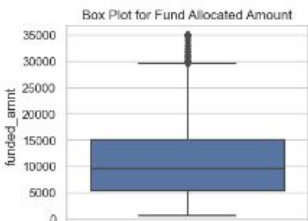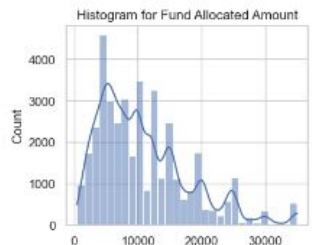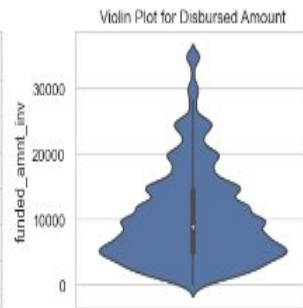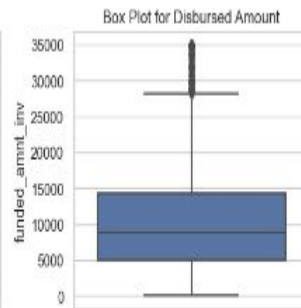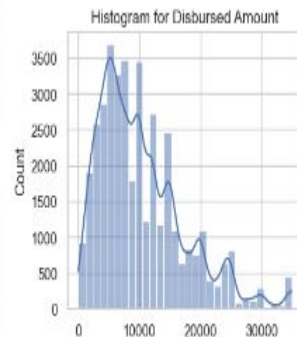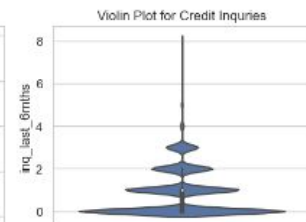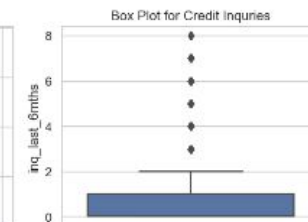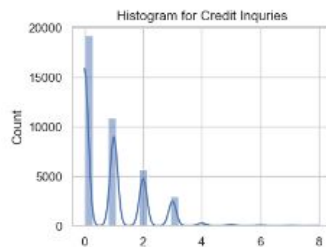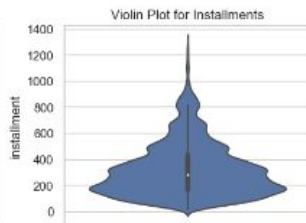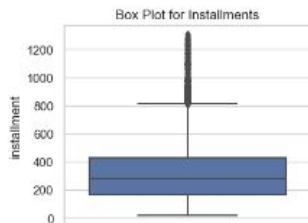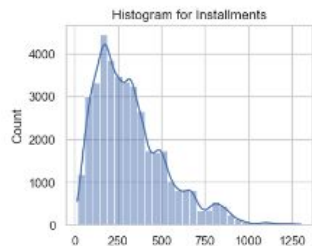
Quantitative variables Insights

- The interest rate is more crowded around 5-10 and 10-15.
- Majority of the borrowers have very large debt compared to the income registered, concentrated in the 10-15 DTI ratio.
- Majority of the borrowers has annual income around 50 K
- Loan Amount, Fund Amount and Disbursed amount follow the same pattern. so there is not much in the requested amount vs disbursed amount
- Number of credit inquiries in past 12 months(inq_last_12m), are less for most of the approved loans
- Most of the people have installment around 200
- The majority of loan has a term of 36 months compared to 60 months.

Segmented Univariate Analysis
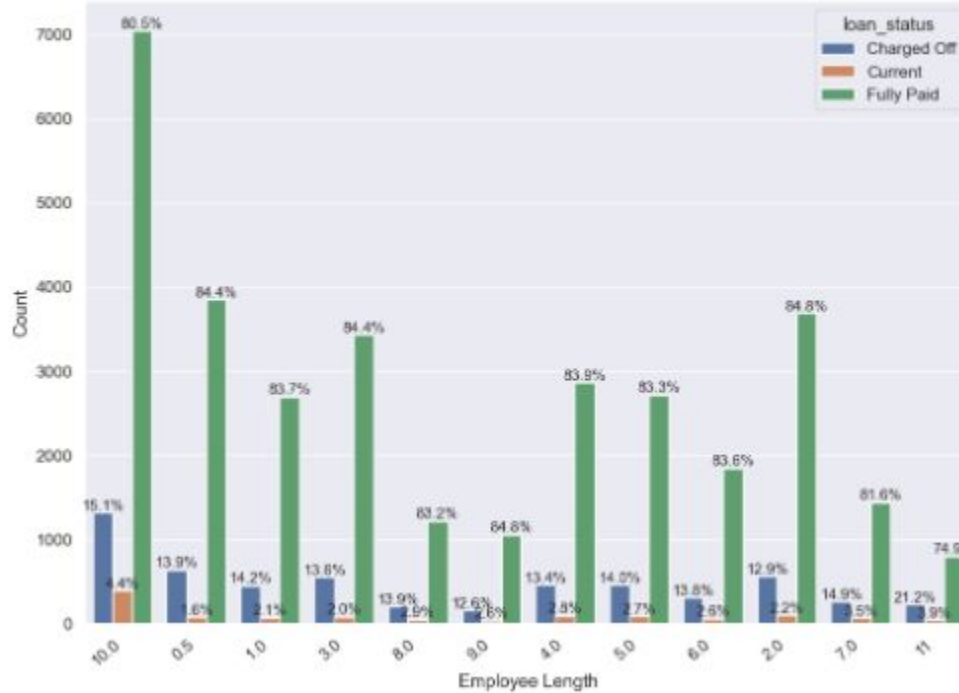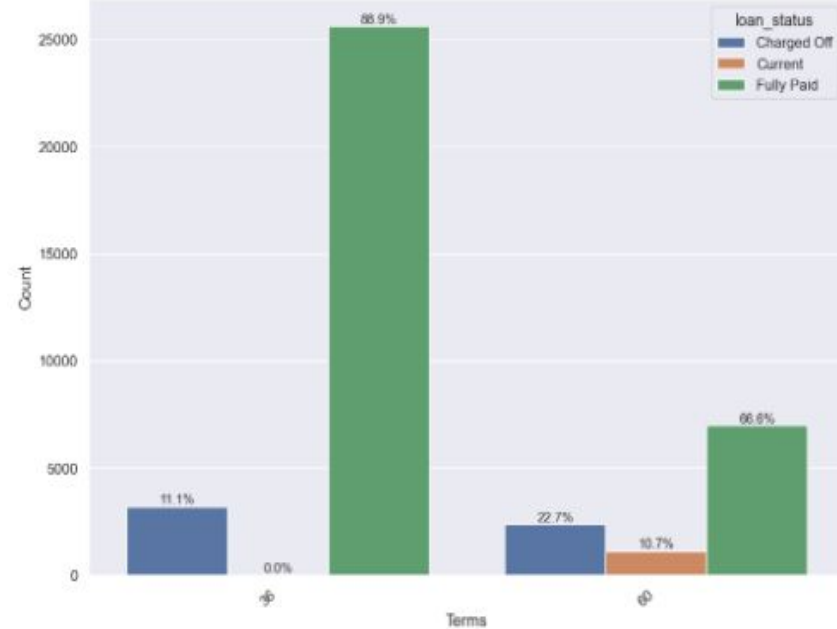
Segmented Univariate Analysis

Segmented Univariate Analysis
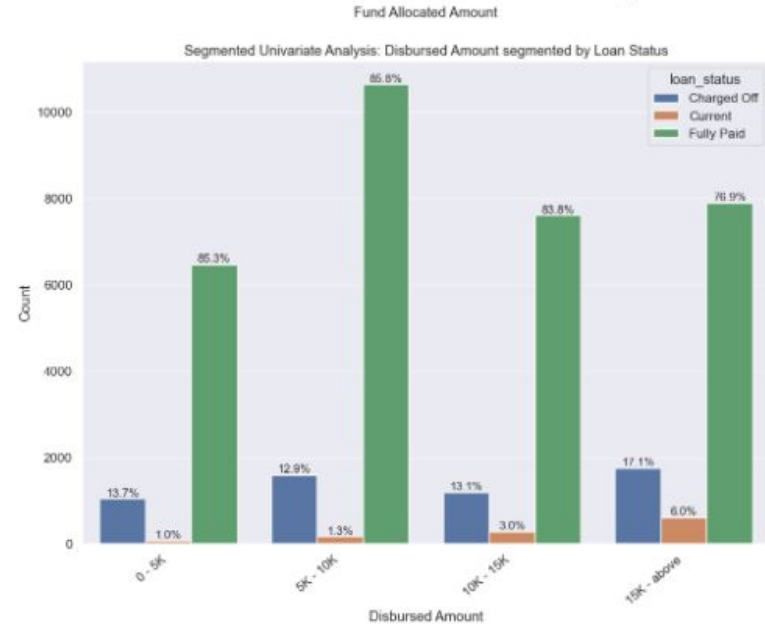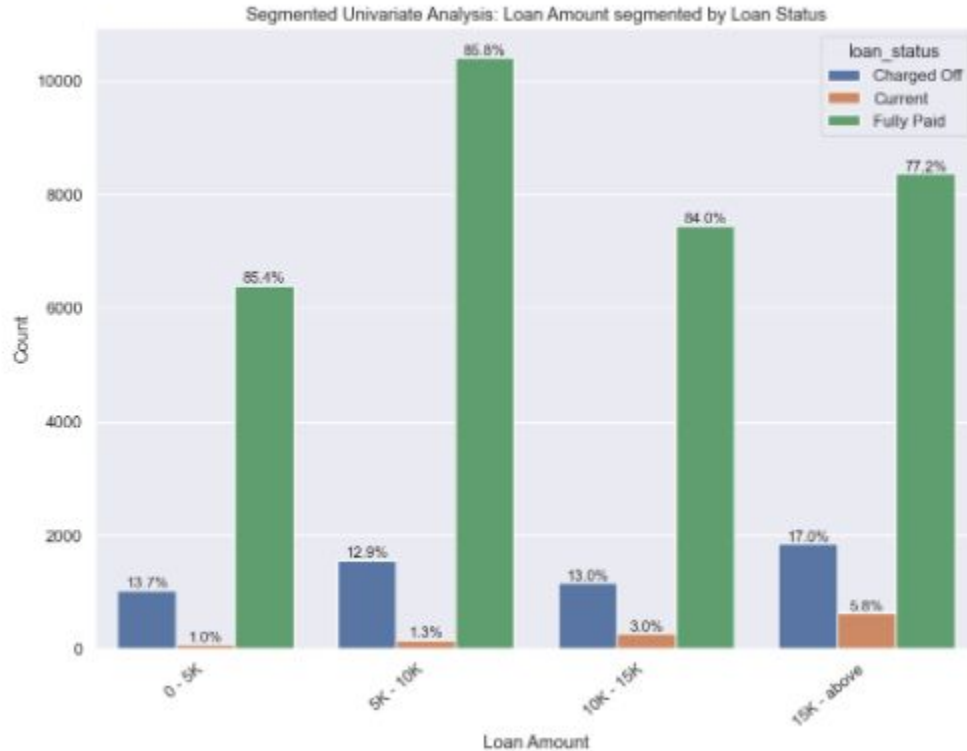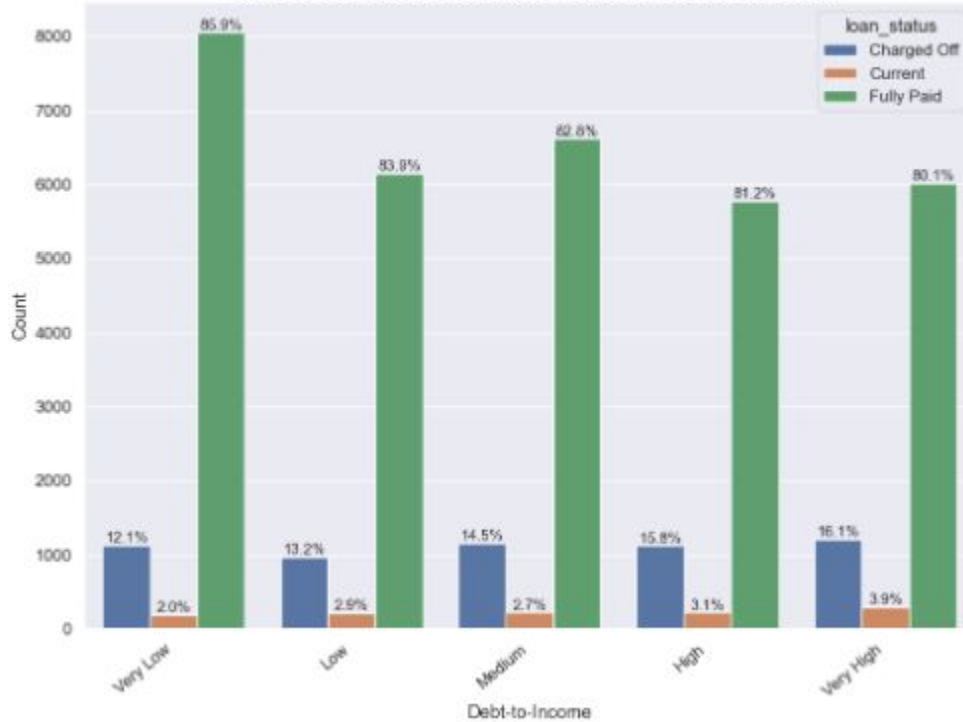
Segmented Univariate Analysis

Segmented Univariate Analysis: Interest Rate segmented by Loan Status

Segmented Univariate Analysis: Installments segmented by Loan Status

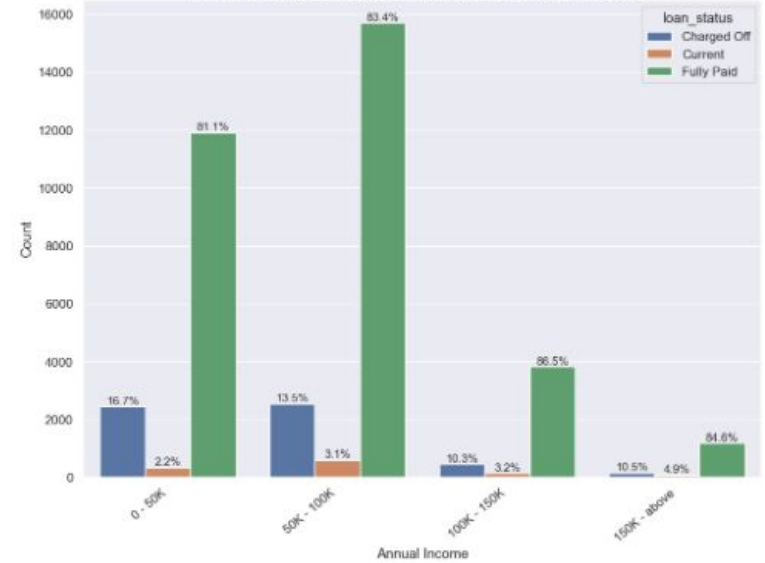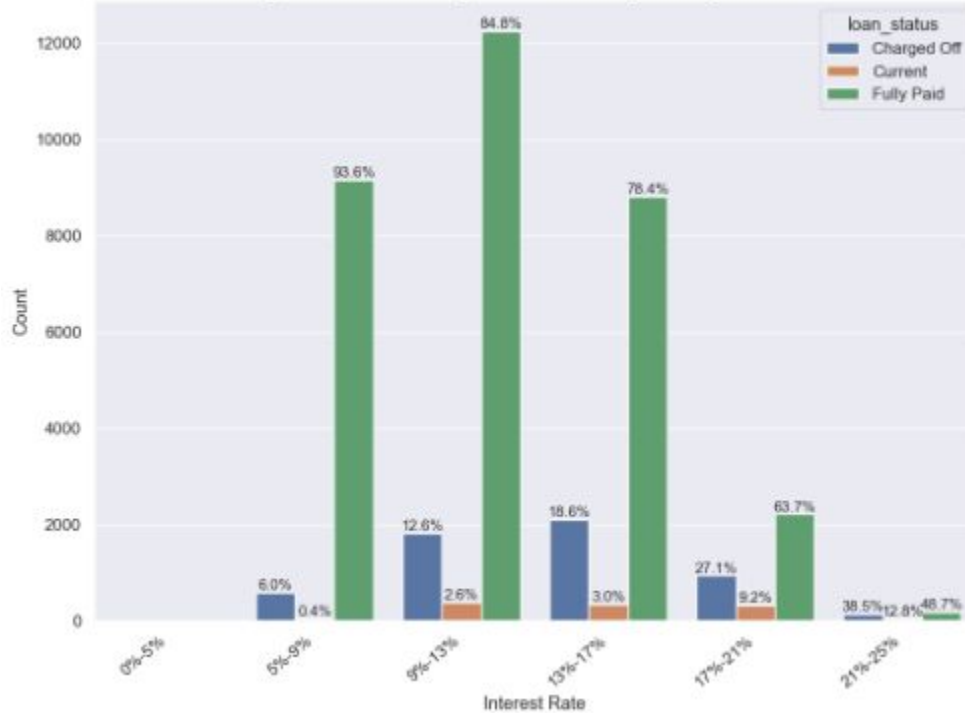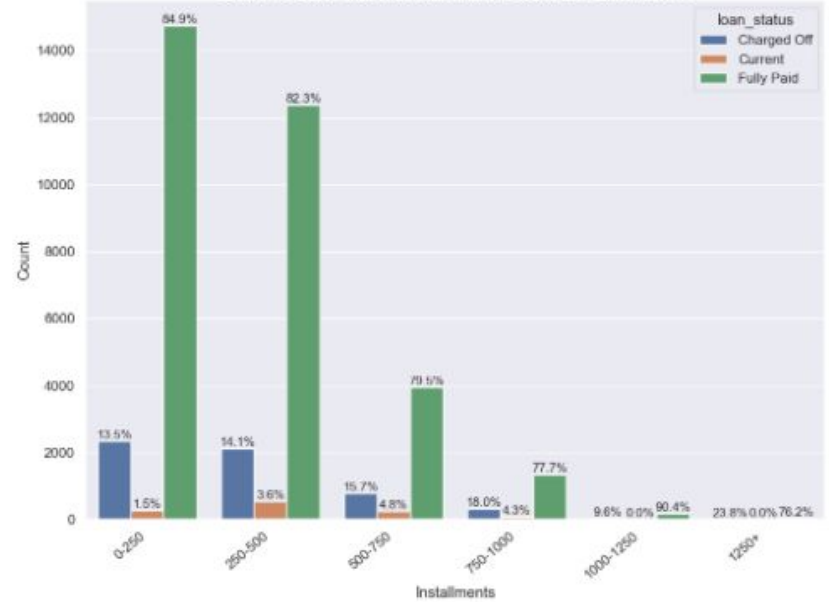# Segmented Univariate Analysis

Segmented Univariate Analysis: Credit Inquiries segmented by Loan Status
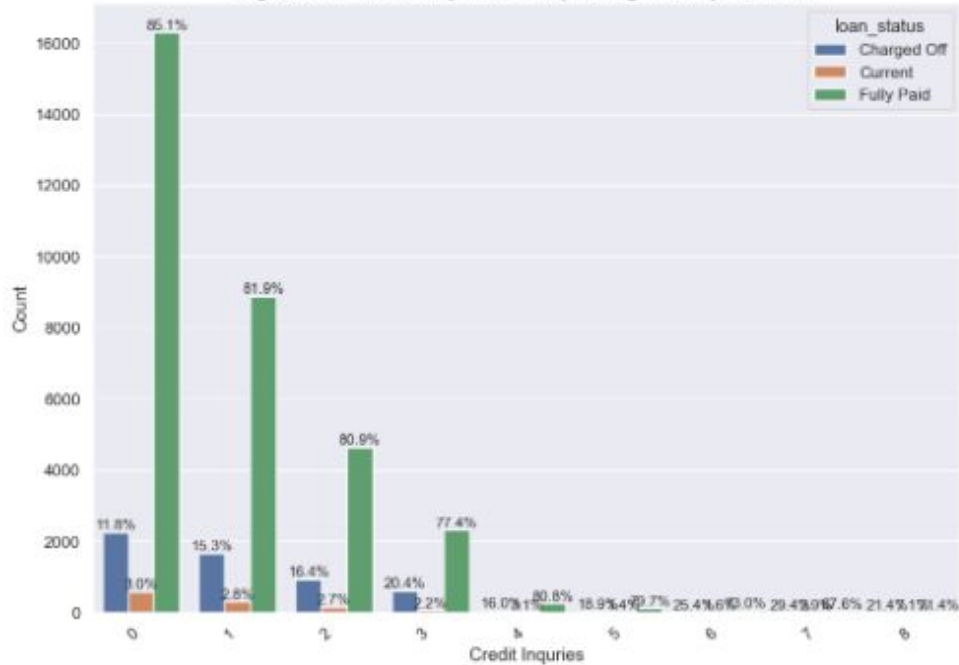
Segmented Univariate Analysis: Bankruptcies Record segmented by Loan Status
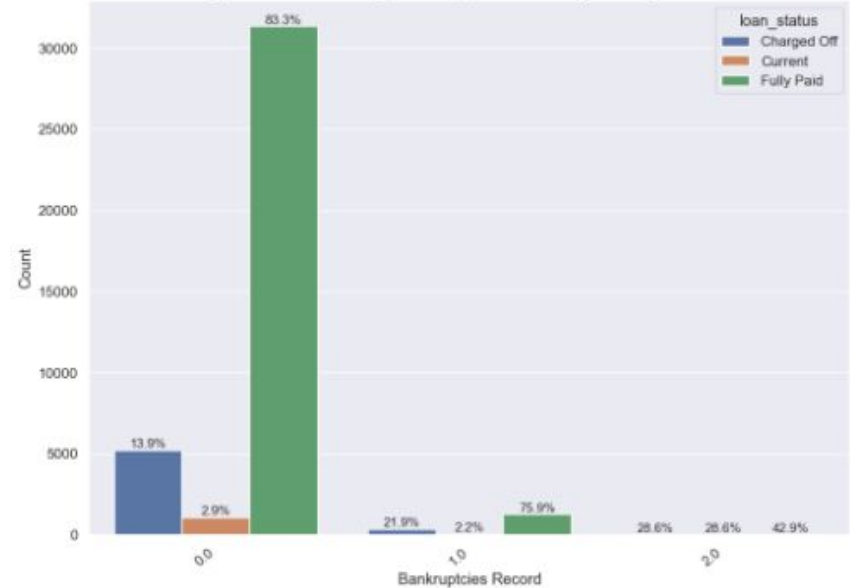
Segmented Univariate Analysis Insights

- Larger loans mean higher chance of defaulting.
- 60-month terms riskier, 36-month safer, with distinct grade distributions.
- Higher DTI has more default risk, especially in High and Very High Bucket.
- Default risk higher for those renting; lower for property owners.
- Lower income (<15K) has higher default probability.
- Defaulted number loan rise when interest rates increases.
- Employees with 10+ years experience have higher chances to default .
- Bankruptcies Record increases higher changes to default .
- Installments increases higher changes to default .
- More number of credit inquiries has higher changes to default .

Bivariate Analysis

Loan Grade vs. Interest Rate
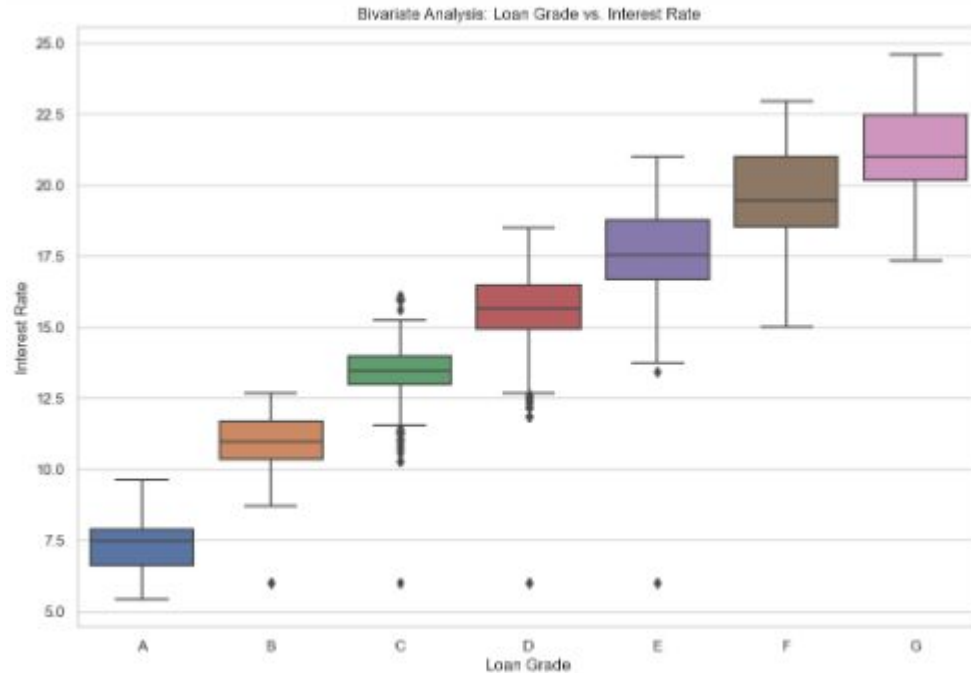
- Grade increase the Interest Rate increases
- This proves the interest calculation is based on the Grade

Bivariate Analysis

Loan Term vs. Loan Amount

Bivariate Analysis: Loan Term vs. Loan Amount
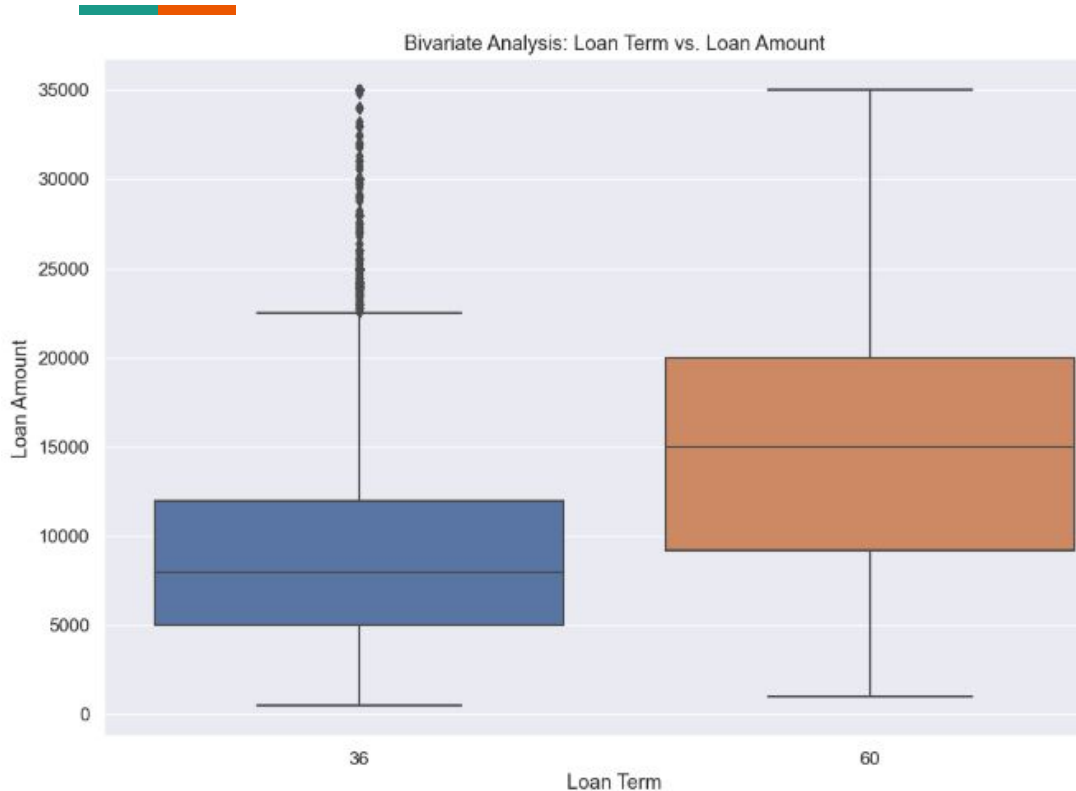
- More number of borrowers in the 36 Months term
- Mode of loan amount in 60 month term is more than 36 which is normal. Because borrower get more amount as they need more time to settle.

Bivariate Analysis
Loan Status vs. Loan Amount
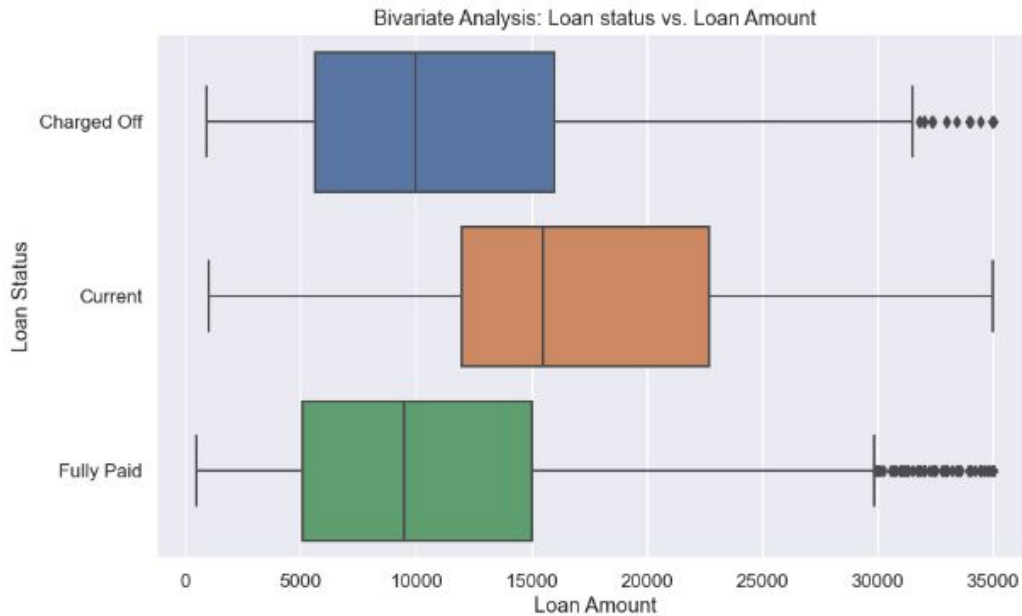
- Loan amount mostly lesser than 15K
- Higher it grows, chances of default is more

Bivariate Analysis

Loan Status vs DTI

Bivariate Analysis: DTI m vs. Loan status
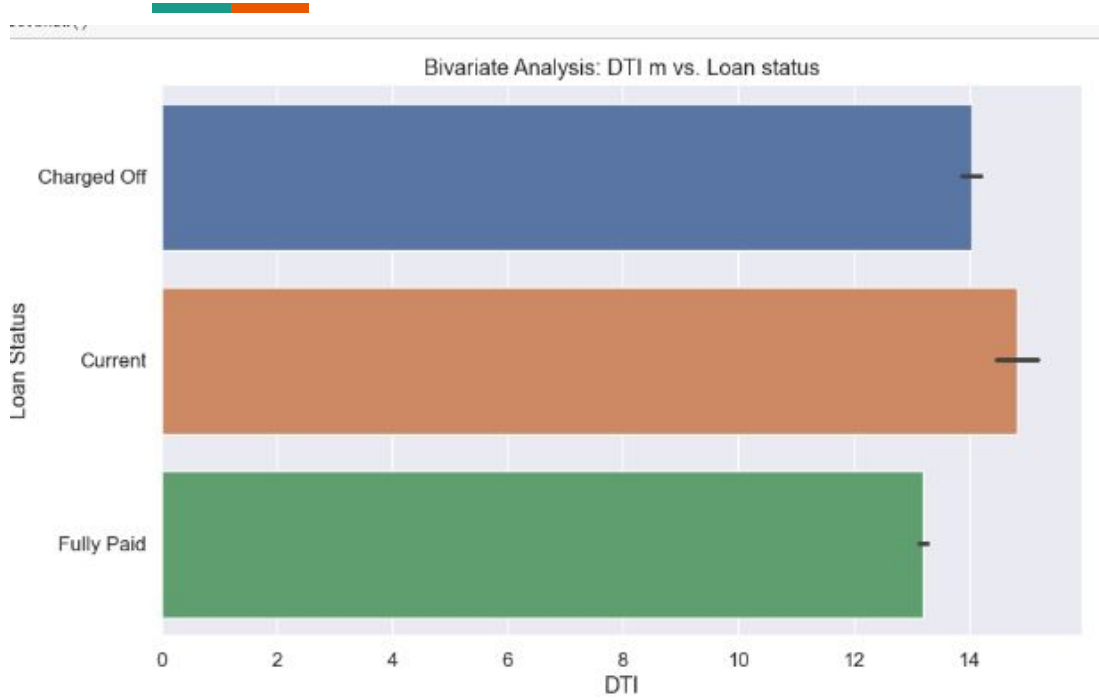
- Lower DTI is less riskier than higher DTI
- Based on this analysis, Current loans are more riskier than usual.

Bivariate Analysis Loan Status vs. Annual income

Bivariate Analysis: Annual income m vs. Loan status
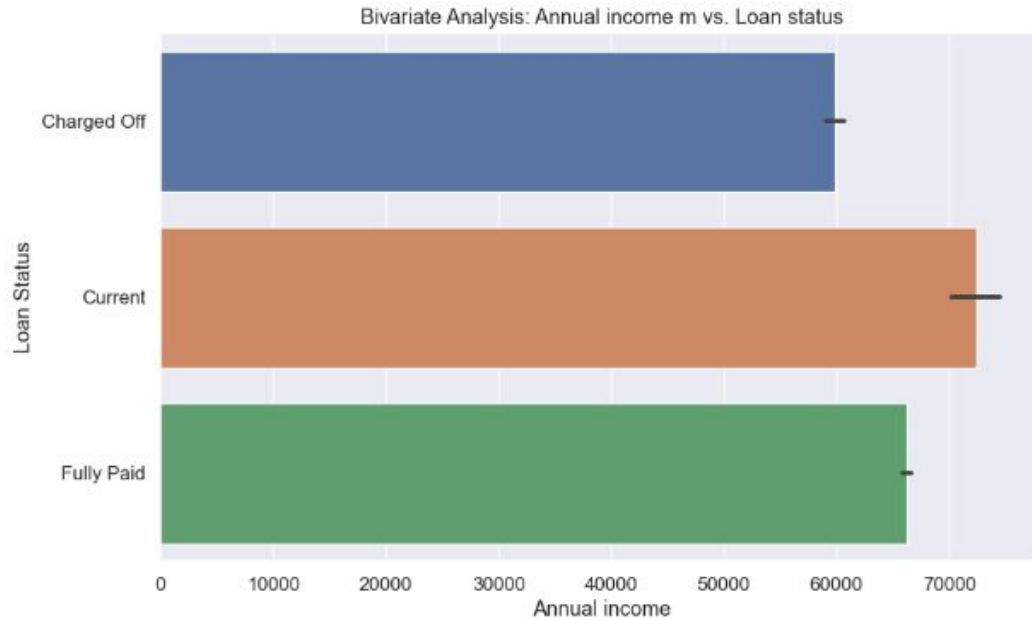
- High annual income is less the risk of default loan

Bivariate Analysis

Home status vs Disbursed amount segmented by Status:

Bivariate Analysis: Home Ownership vs. Loan Amount (Segmented by Loan Status)

- Borrowers living home in mortgage taking higher loan amount than others
- For all home status, Loan amount increases, it becomes risky.

Correlation Analysis

Correlation Analysis

- `installment` has a strong correlation with installment,funded_amnt, loan_amnt, and funded_amnt_inv
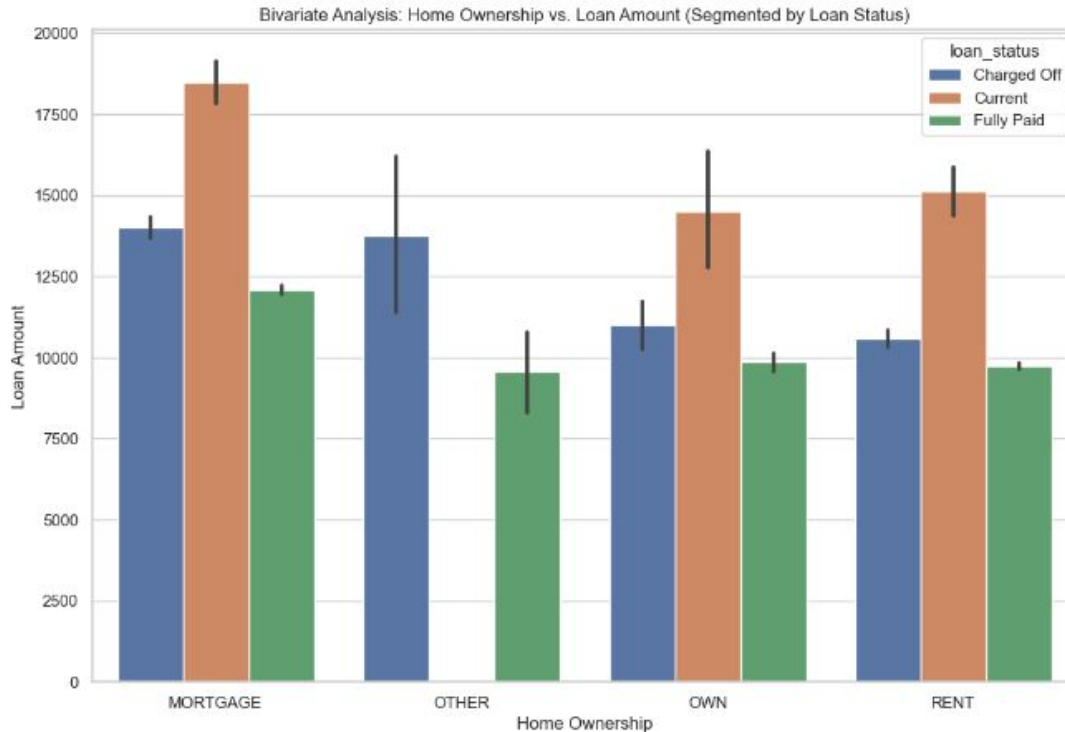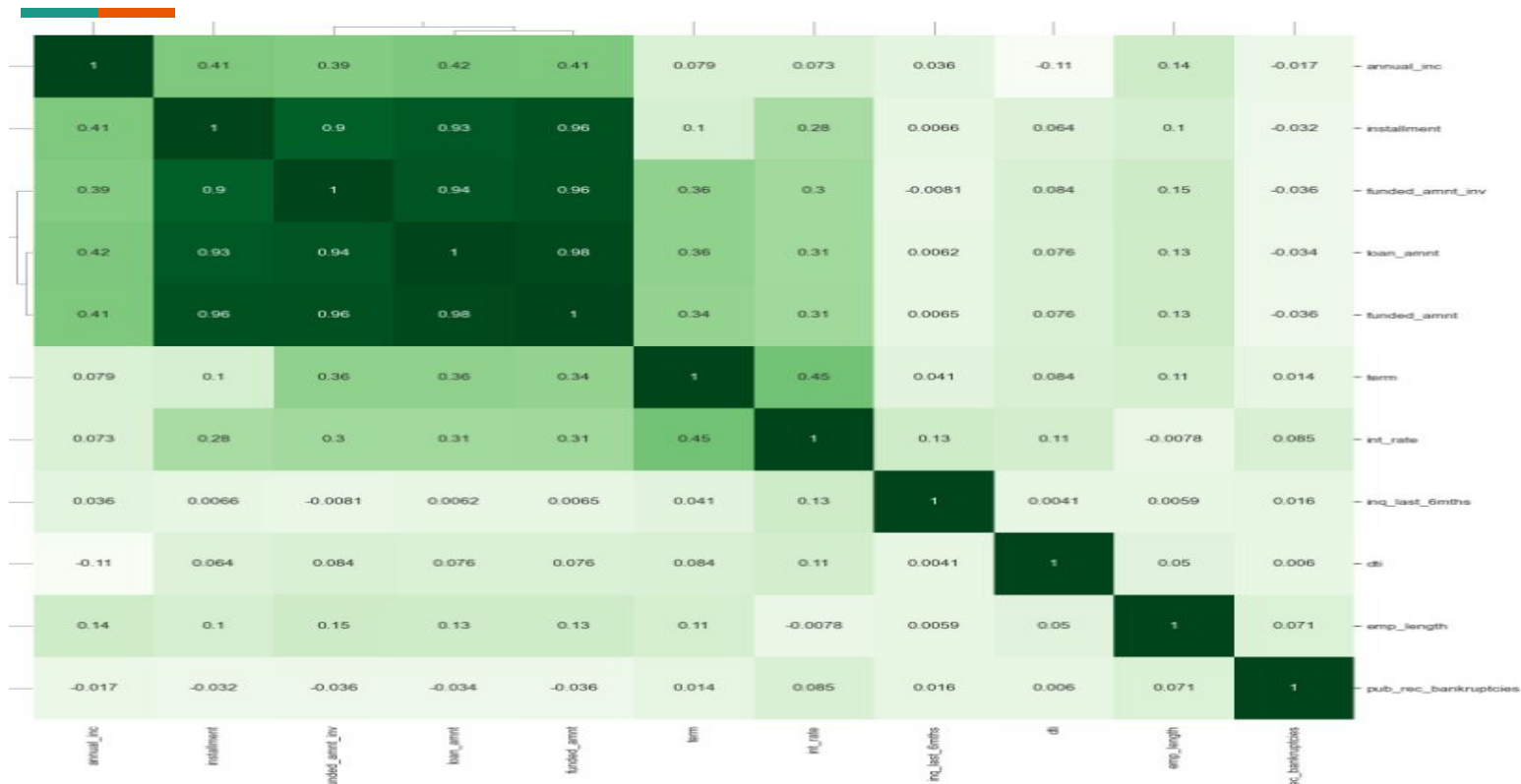- `term` has a strong correlation with interest rate
- `annual_inc` has a strong correlation with loan_amount
- `dti` has weak correlation with most of the fields
- `emp_length` has weak correlation with most of the fields. Employee Experience is not significant factor on load lending process.
- `loan_amnt` has strong correlation with installment,funded_amnt, and funded_amnt_inv
- `pub_rec_bankrupticies` has a negative correlation with annual_inc, funded_amnt, loan_amnt, and funded_amnt_inv
- `annual_inc` has a negative correlation with dti

Conclusion

- Factors for predicting defaulting
  - DTI being higher
  - Lesser Annual income
  - Longer term(60 month)
  - Number of Public record of Bankruptcies is higher
  - Home status as Mortgage or rent
  - Higher Interest Rate
  - Higher Installment.
- Other factor should be enforced
  - Enforce verification process for loan approval process. 20% loans are not verified
  - Borrower with higher grade(E and F) is risky as their interest rate is higher
  - Borrower with 10+ year experience , having higher risk than other.