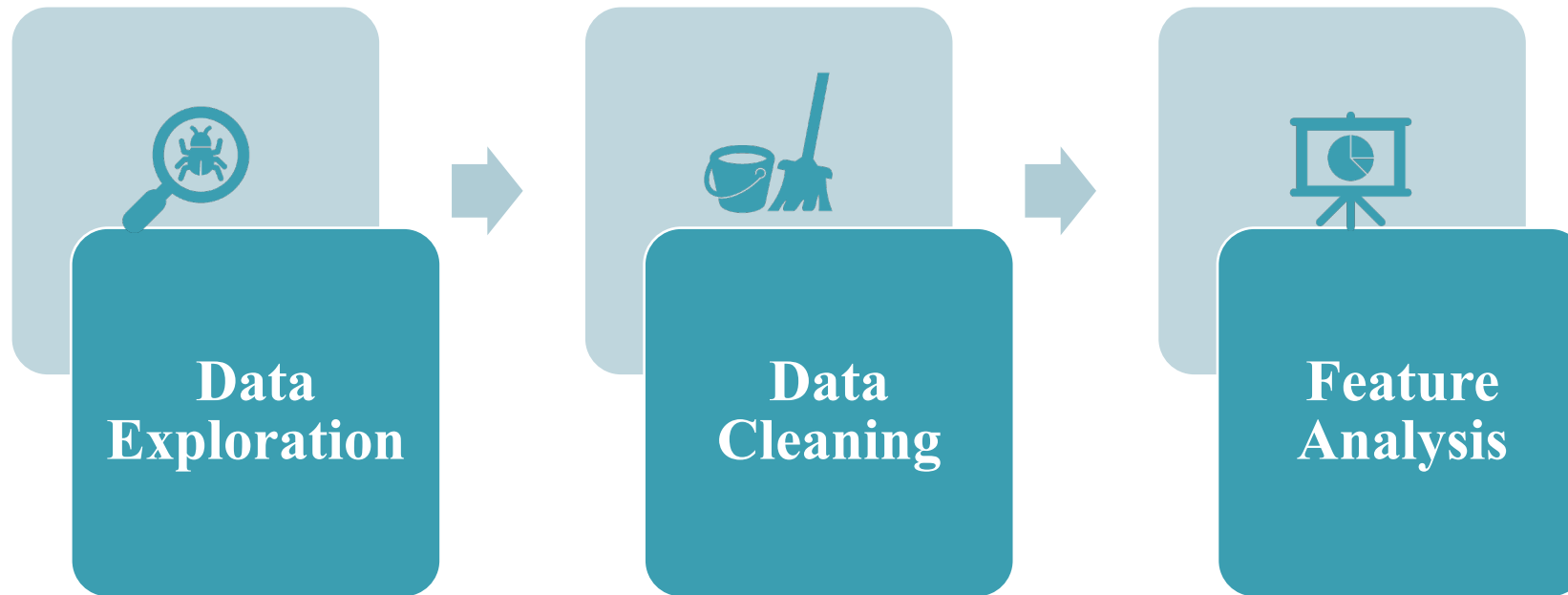# LENDING CLUB CASE STUDY

Detailed Analysis by

Samidh Bhagwat & Archishman Gantait

Business Objective

Perform EDA to understand how consumer attributes and loan attributes influence the tendency of default and identify driving factors behind loan default

Analysis Overview & Process

Data Exploration

Data Cleaning

Feature Analysis

Data Exploration

Data Overview - Understand the general structure of the dataset (Size, columns etc)

Data Types and Conversion - Check the data types of each column

Unique Values and Cardinality - Identify unique values in categorical columns

Data Cleaning

1

Handling Missing Values

Identify columns with missing values and decide on an appropriate strategy

2

Handling Irrelevant or Redundant Features

Identify and remove irrelevant or redundant columns that do not contribute meaningful information to the analysis

3

Handling Outliers

Outlier identification and treatment to ensure accuracy in the analysis

Feature Analysis - Univariate Analysis

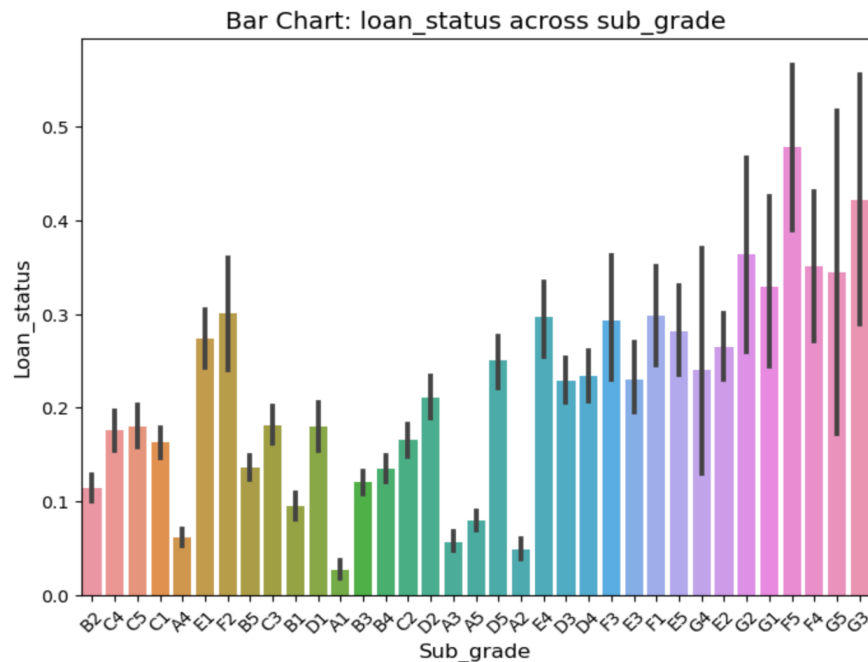**Univariate Analysis**

Bi-variate Analysis

UNIVARIATE ANALYSIS
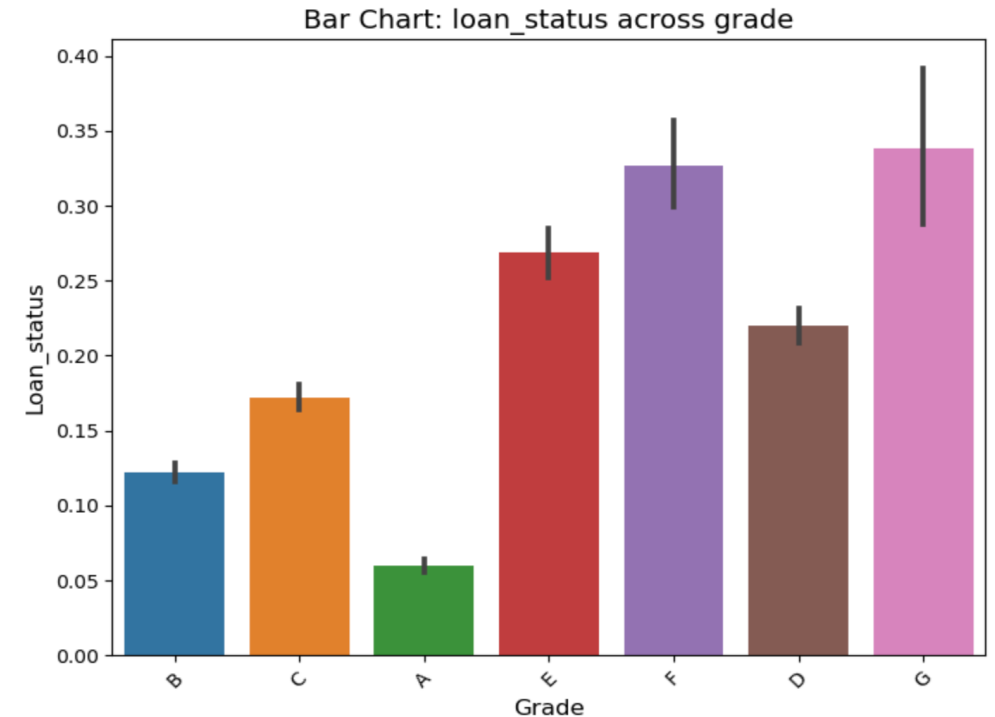
GRADE & SUB GRADE

- Grade "G", lowest grade of loans, as defined by LC have the highest rate of default.

- As we go down from sub grade A1 to A2 to A3, so on and so forth, the default rate increases
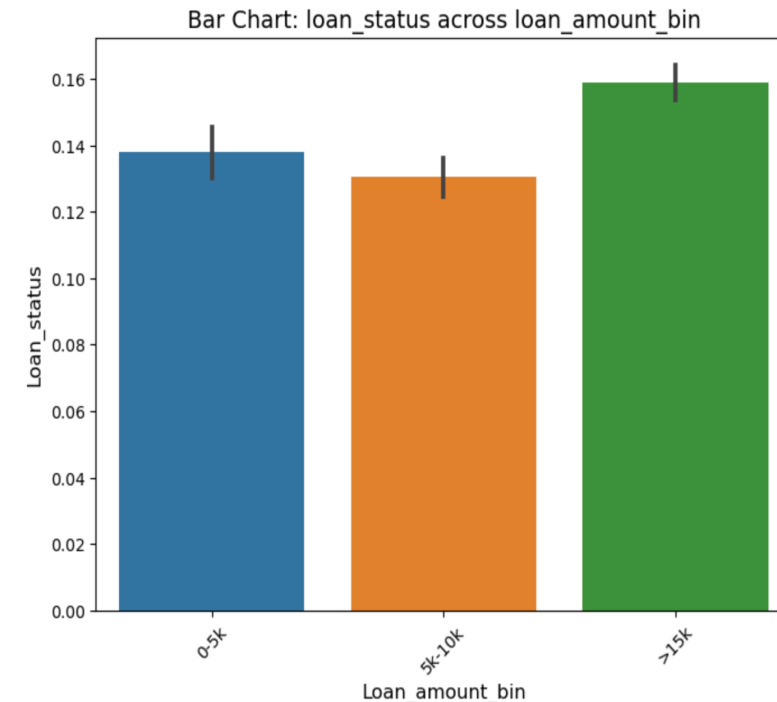
LOAN AMOUNT & LOAN TERM

- Loan with amount >15k have the highest default rate while for 5k-10k, the defaults are the lowest

- Defaults are higher when the term of loan is higher (60m)
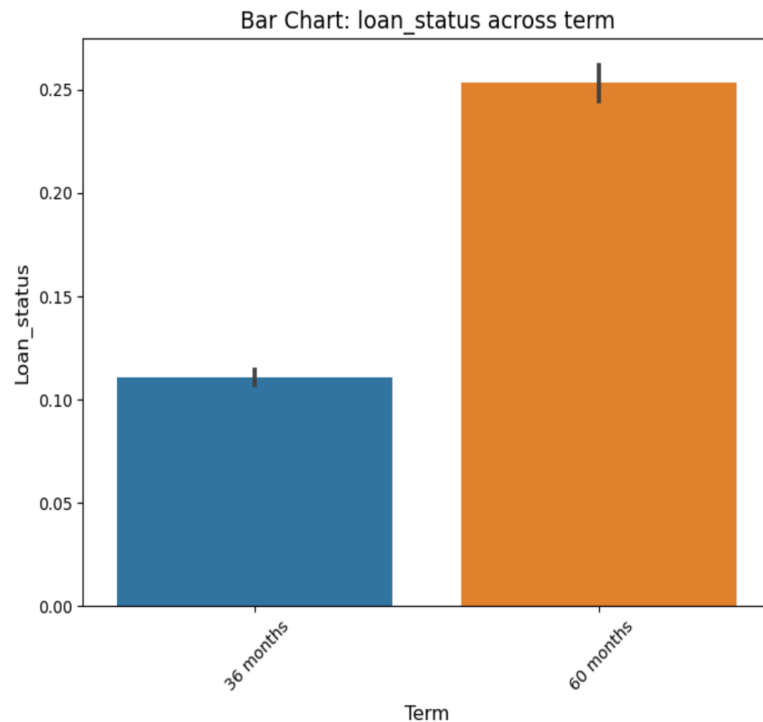
HOME OWNERSHIP & VERIFICATION STATUS

- Home Ownership doesn't really have a lot of impact on the overall default rate. The rate seem to be uniform across
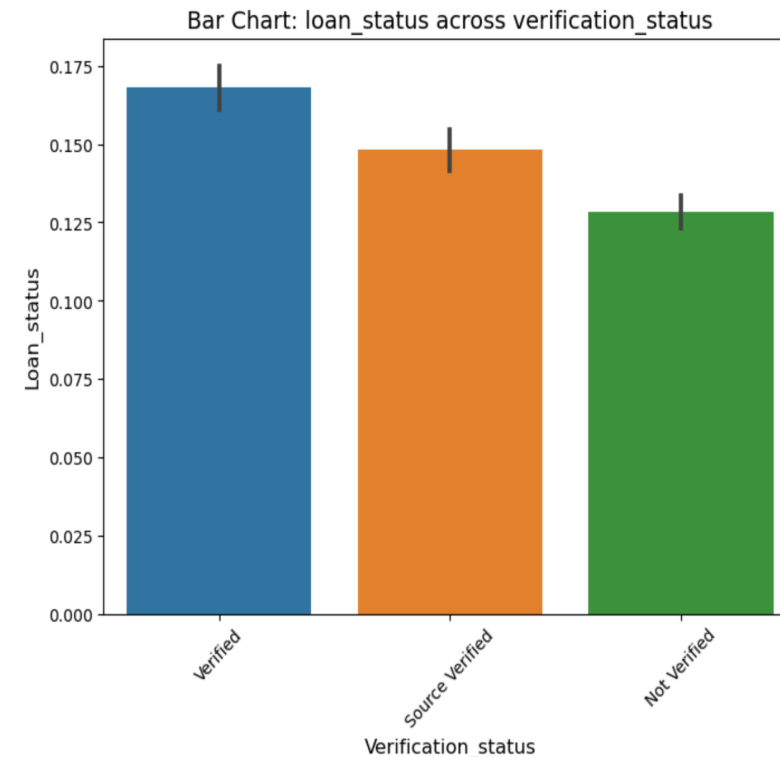
- Verified loans have a higher default level as compared to Non-verified loans. The verification process needs to be re-evaluated


```
In [31]: # home ownership: not a great discriminator
         plot_bar_chart('home_ownership','loan_status')
```
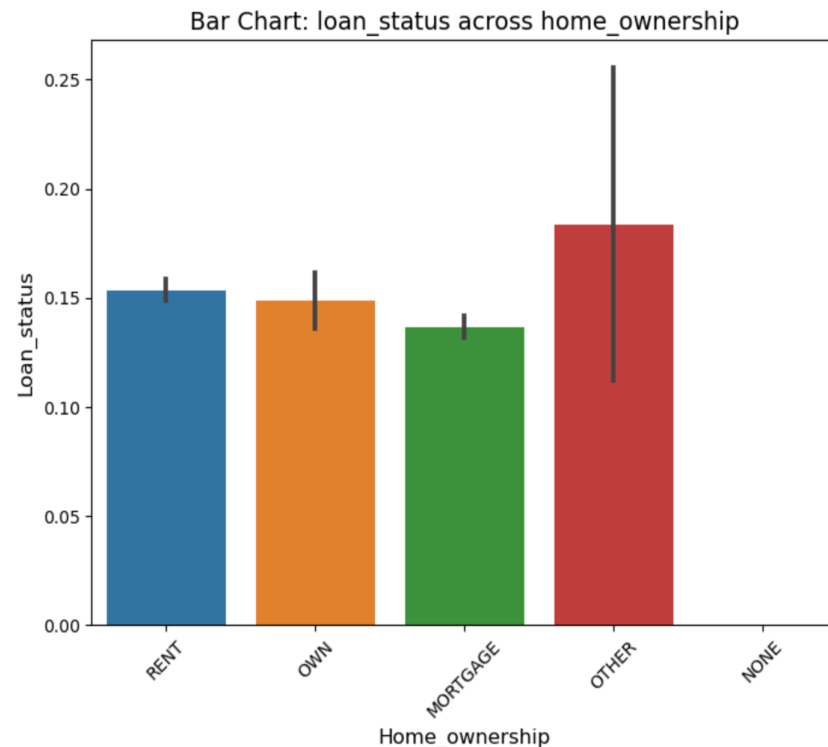


```
In [32]: # verification_status: surprisingly, verified loans default more than not verified
         plot_bar_chart('verification_status','loan_status')
```

PURPOSE & FUNDED AMOUNT

- Approving Loan request for purpose like credit card, car, home improvement and wedding can be considered as much safer than others as there are less defaulters

- The defaulters are more when funded amount is above 15k , however defaulters are less when funded amount id between 5k to 10k

Bar Chart: loan_status across purpose

Bar Chart: loan_status across funded_amnt_inv_bin
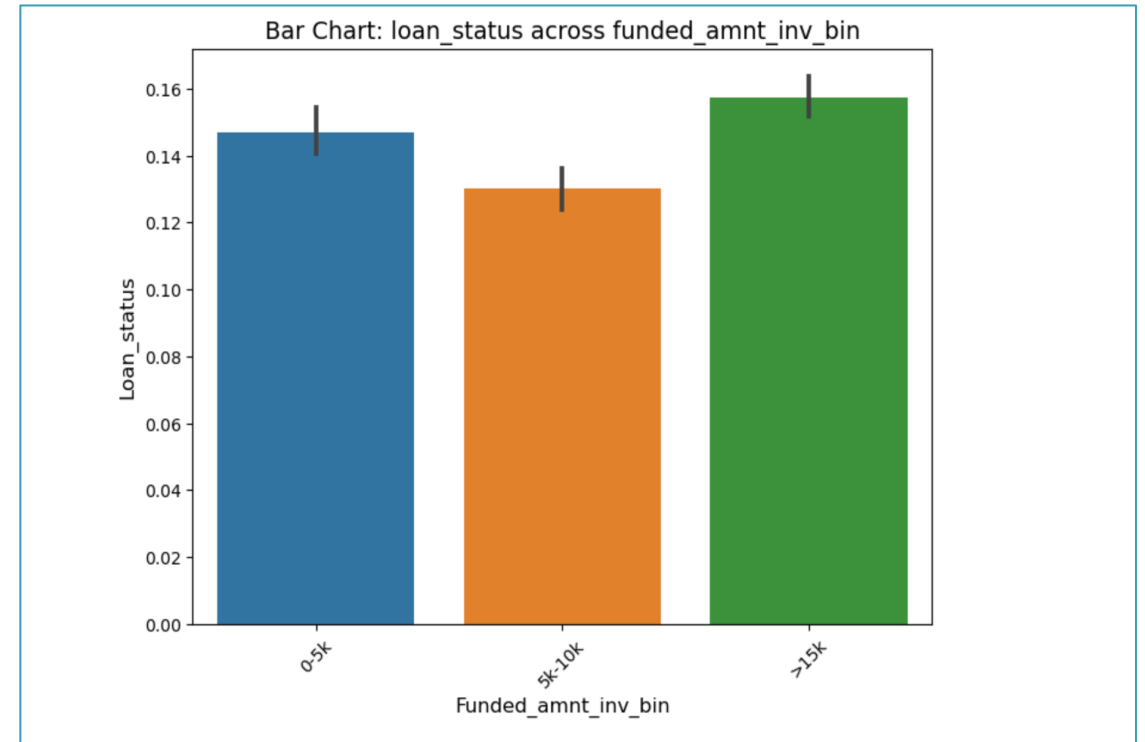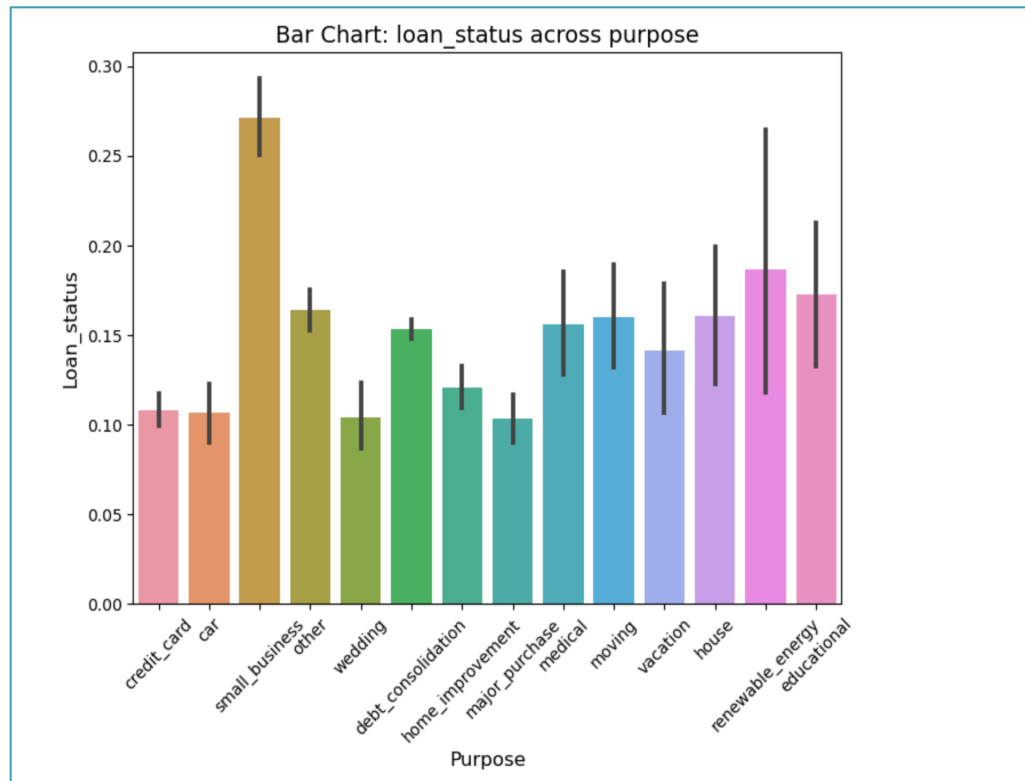
DEBT TO INCOME & INSTALLMENT

- Default rates are higher as the debt to income ratio increases , except for the 4-8 bracket , where defaulters are lower when compared with the previous segment of 0-4. A steady trend can be seen

- Default rate is high when increasing installment value. Borrowers who choose lower installments are less likely to default

Bar Chart: loan_status across dti_bin

Bar Chart: loan_status across installment_bin
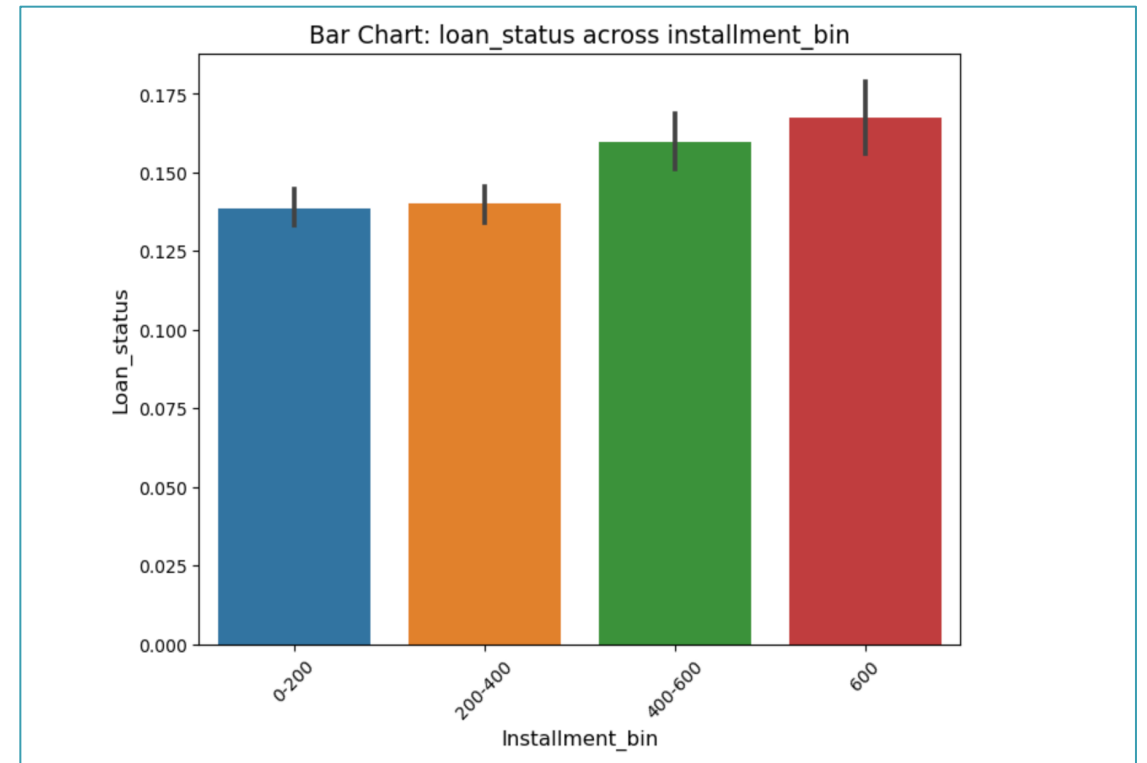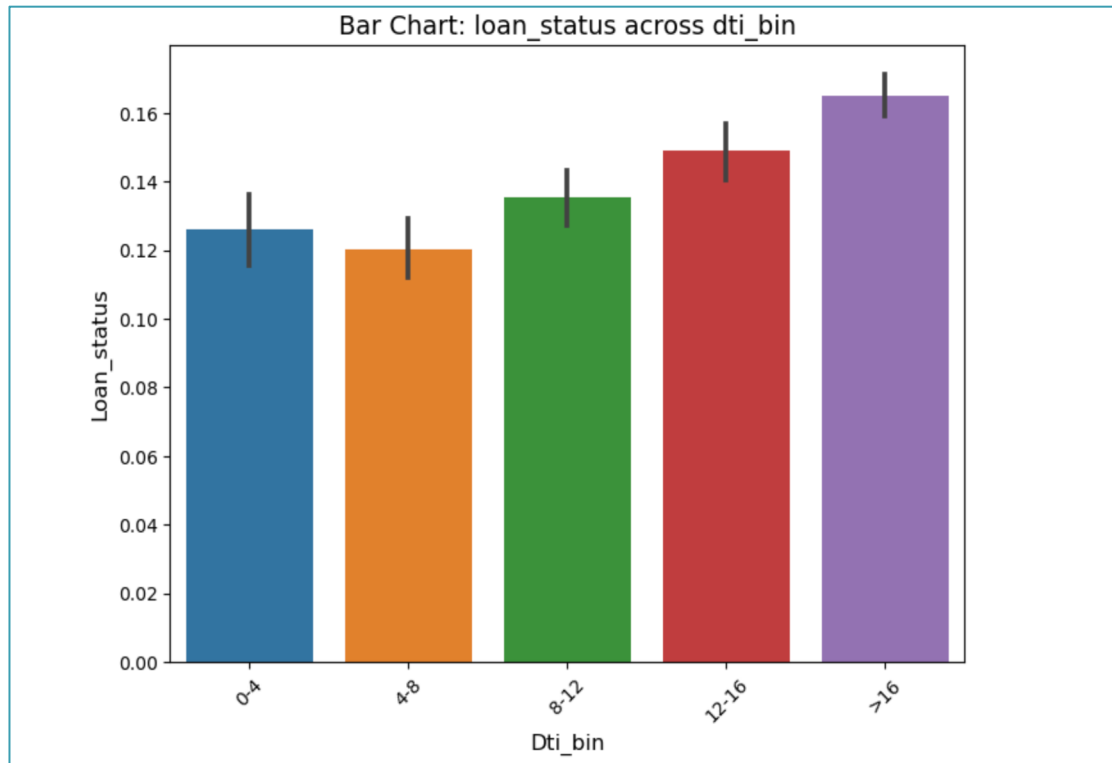
ANNUAL INCOME & EMPLOYEE LENGTH

- Defaulters tend to reduce with higher annual income. Income >125k has the lowest default rate.

- There is no such pattern in the data for decreasing or increasing years of employment. However, borrowers with 7- or 10-years experience show higher defaulters rate.
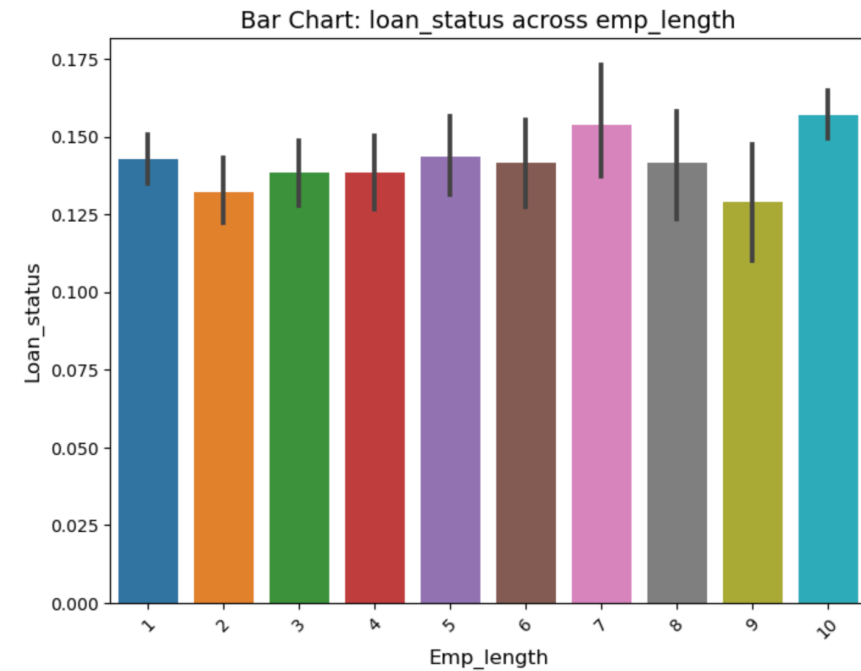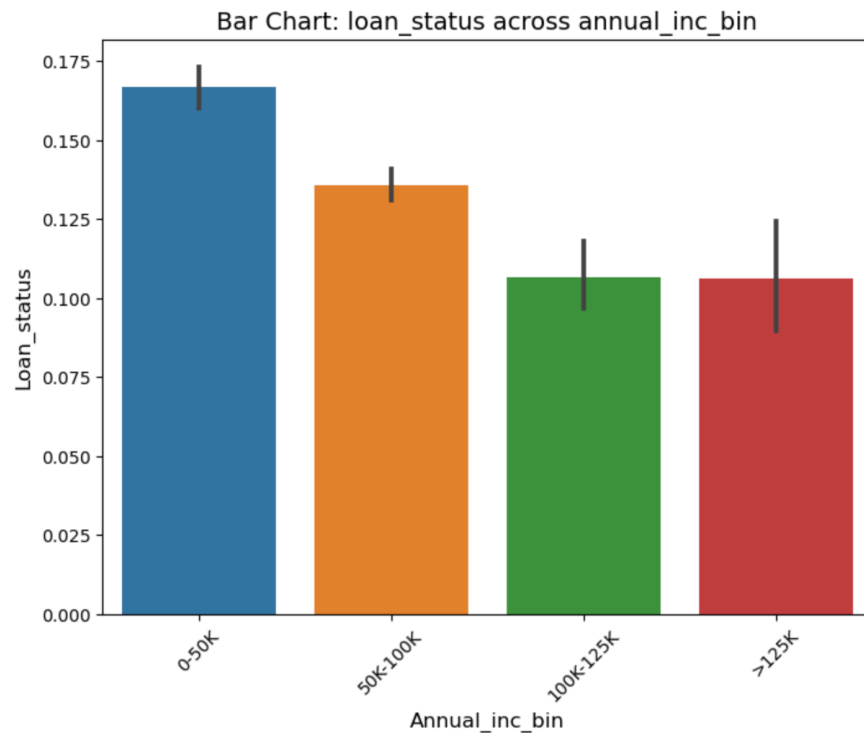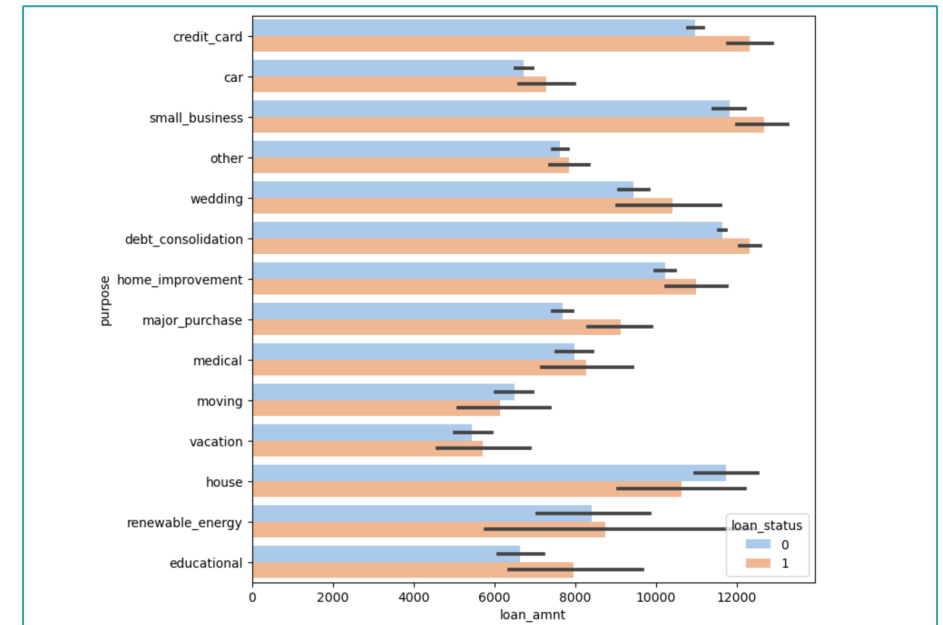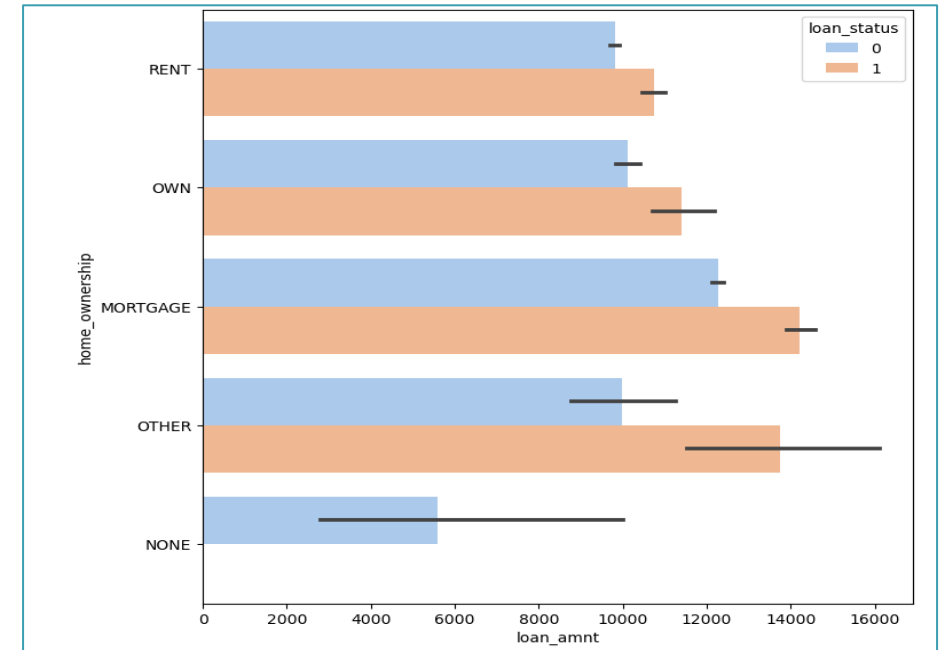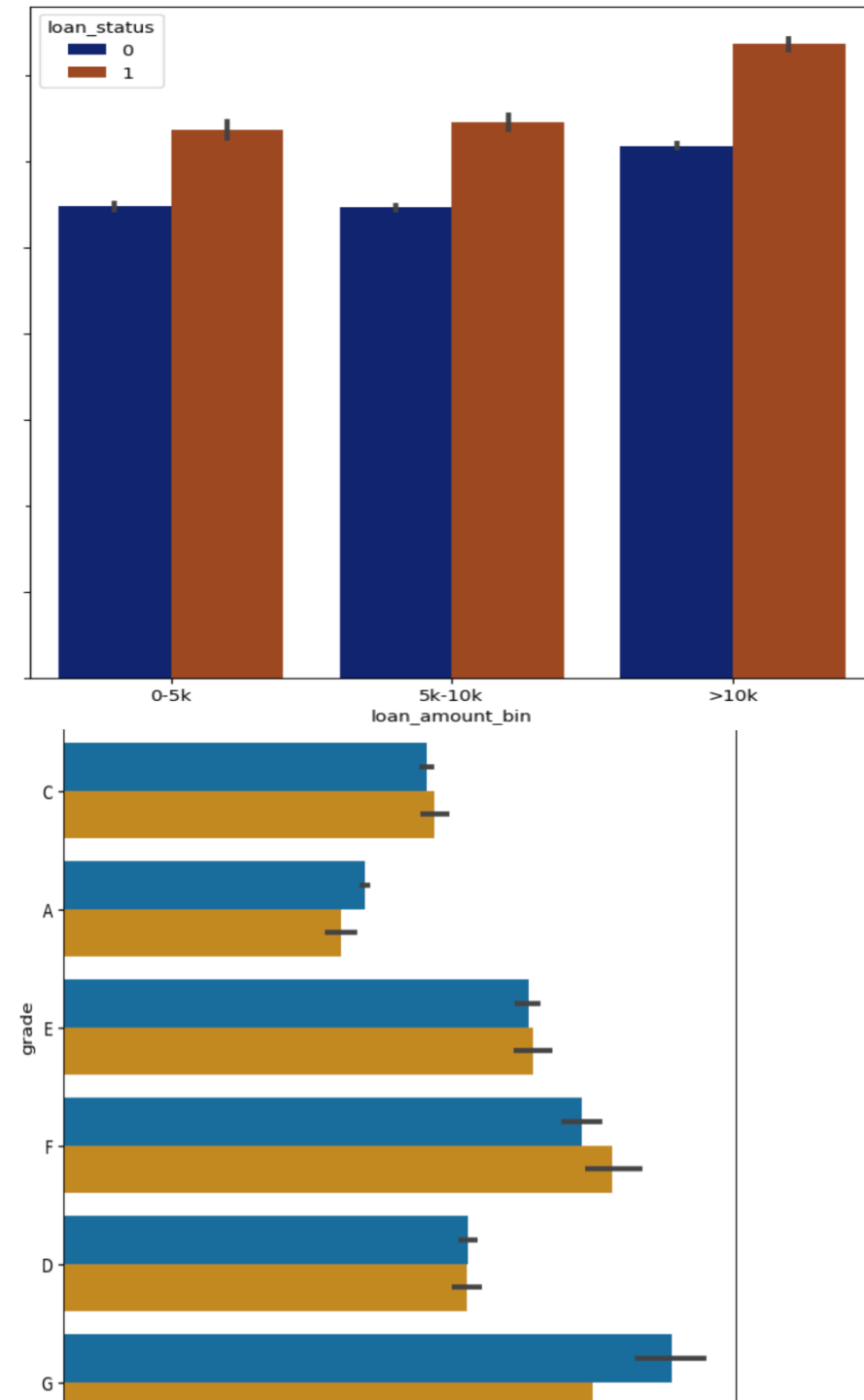
BI -VARIATE ANALYSIS

BI-VARIATE ANALYSIS ON LOAN AMOUNT WITH PURPOSE AND HOME OWNERSHIP

- The loan amount for loans for Vacation, moving, medical are the lowest, however, they have the least default rate too, making them least risky for the organization.

- Loans for small business have the biggest ticket size, suggesting a high revenue, however this high revenue comes with a highest rate of default

- borrowers living in a mortgaged house have the highest loan ticket size, however this comes with the highest risk too .

- As seen earlier, higher loan amount attracts higher risk

BI-VARIATE ANALYSIS ON LOAN AMOUNT WITH GRADE AND INTEREST RATE

- We saw earlier that Grade G has higher risk of defaulters , however there is still a bracket of loan amount that can be approved for grade G

- Grade F has the highest default rate, however the loan amount for these borrowers is significantly higher

- Higher loan amount, >10k have interest rates between 12-15%.
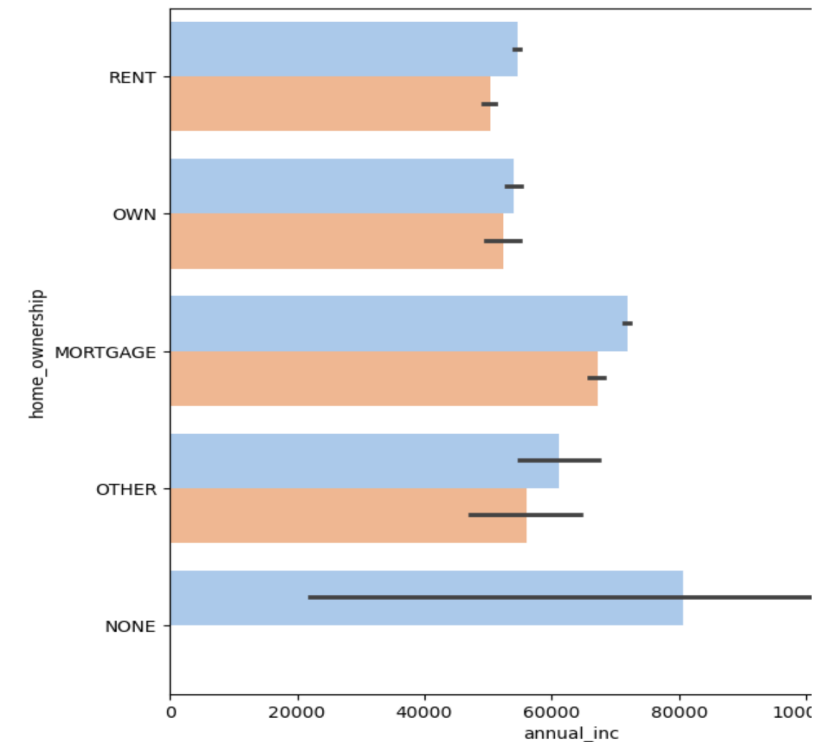
BI-VARIATE ANALYSIS ON LOAN AMOUNT WITH EMPLOYEE LENGTH

- For employees with longer work history, loan amount is higher, giving an impression of borrowers settled in their lives tend to have a higher loan eligibility

- The default rate is proportionately higher for these high value loans.

- E.g., for emp length 10 , above 12k there is a risk of loan not being repaid on time.
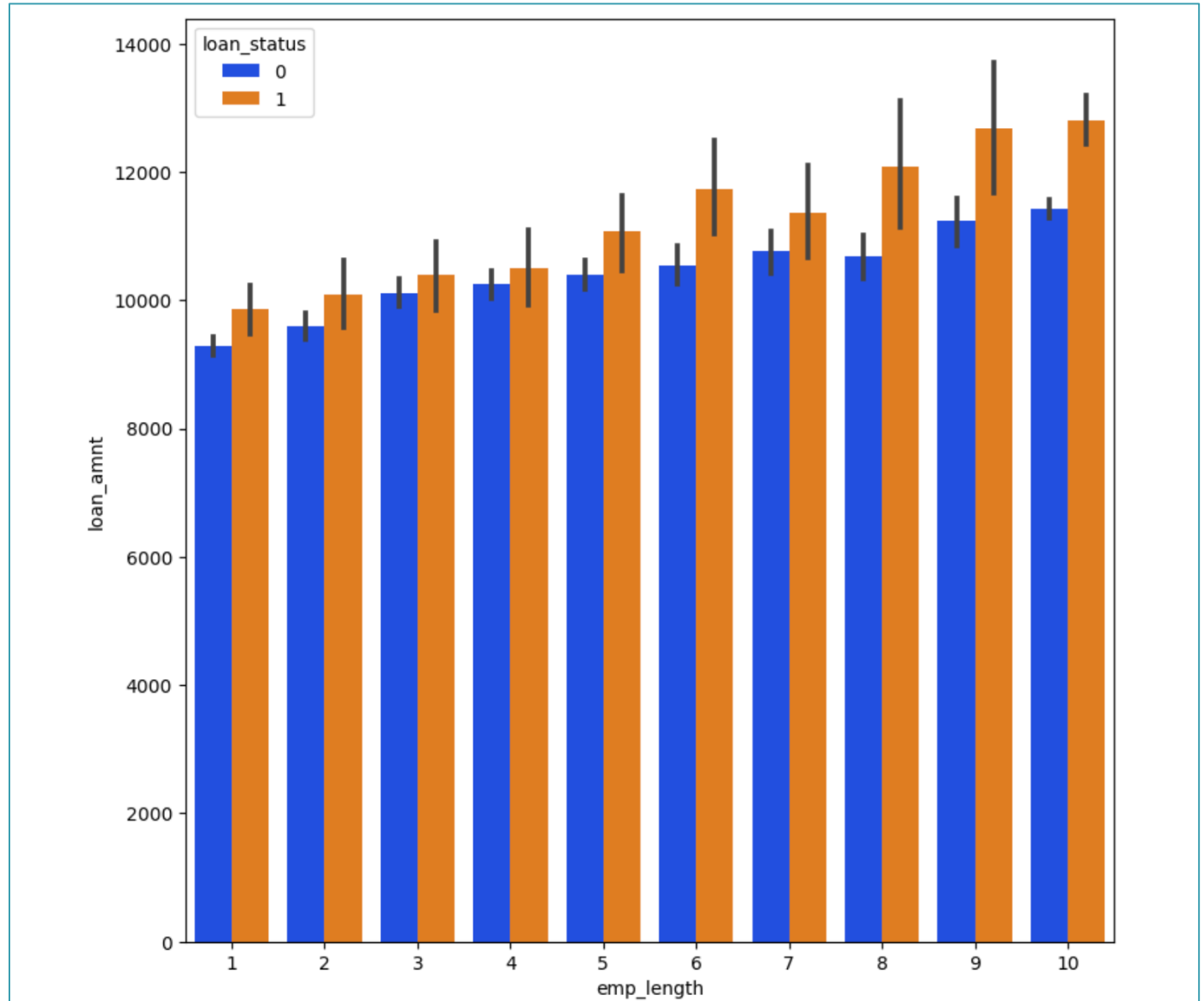
BI-VARIATE ANALYSIS ON ANNUAL INCOME WITH HOME OWNERSHIP & GRADE

- Within the types in home ownership , each type has an annual income post which it can be considered safe to approve loans.

- E.g. Any one applying for loan against mortgage with annual income above 80k have more chances of repaying loan on time.

- For each grade there are chances of loan being repaid when annual income exceeds a certain limit.

- E.g. for grade G , above 80k annual income people mostly repay loans.
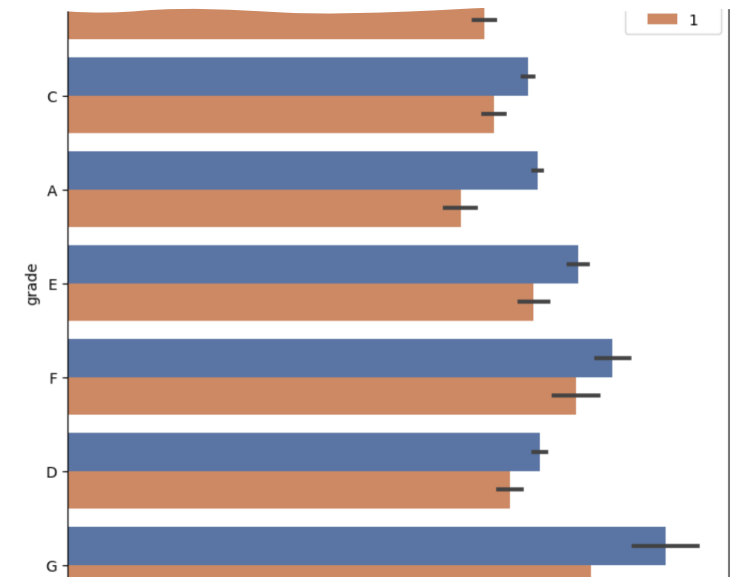
BI-VARIATE ANALYSIS ON ANNUAL INCOME WITH INTEREST RATE , LOAN TERM & LOAN AMOUNT

- For term 36 months , borrowers with annual income above 60k are more likely to repay loans. For term 60 months , borrowers with annual income above ~68k are more likely to repay loan.

- For each annual income bracket , a higher interest rate has more defaulter.

- Here for all the annual income range defined , interest rate over 12% has more default rate

- Borrowers with a higher income tend to get a higher loan as compared to the ones with a lower income

- This signifying the correlation between income and approved loan amount

- Borrowers with lower income tend to have a loan approved loan amount, in line with their income
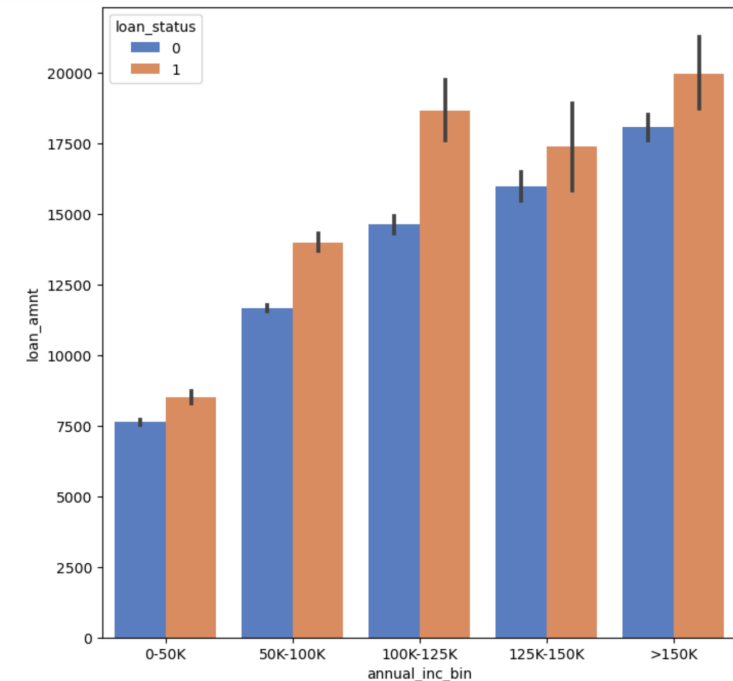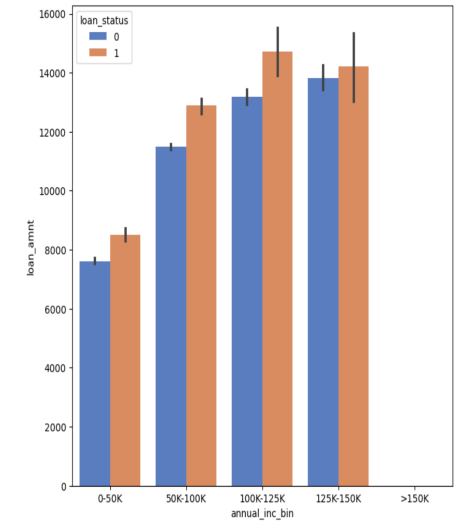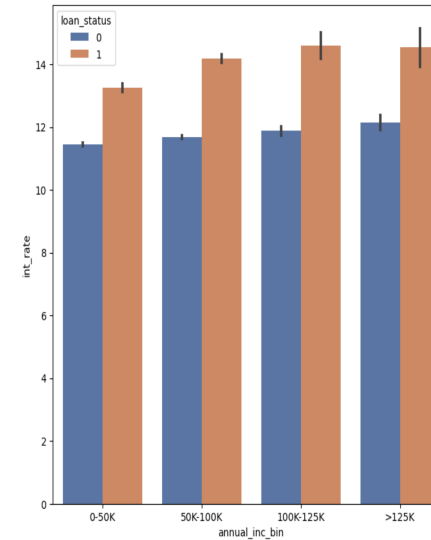
BI-VARIATE ANALYSIS ON INTEREST RATE & LOAN TERM

- For loans with higher term (60m), generally have higher rate of interest as compared to short tenure loans

- The default rate for longer term loans with higher interest rate is higher which suggests these are risky loans and LC is cognizant of this
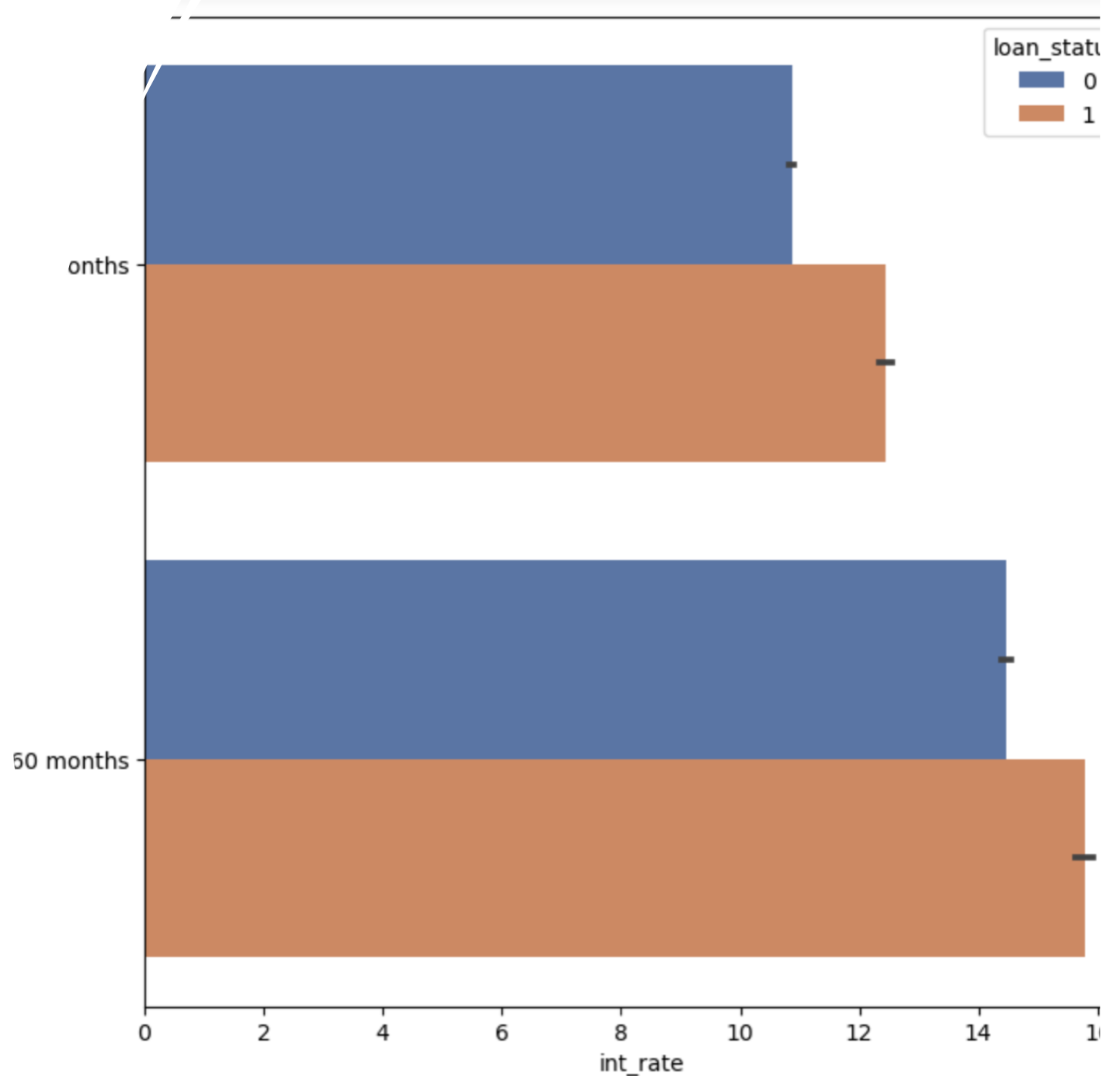
OVERALL RECOMMENDATIONS

- Low Risk borrowers - Borrowers with higher credit rating (A,B) could be given loans as they have a less likelihood of default. These borrowers can be offered loans at a competitive rate.

- Employment history - Borrowers with consistent employee history of >6-7 years could be less risky

- Smaller loans should be approved for low graded borrowers

- High risk borrowers should be given a short tenure loan

- Verification process for the loans should be little more stringent

- Applicants with higher income could be given more diversified loans