LENDING CLUB CASE STUDY

Jayaram TK

Sumit Khelwal

PROBLEM STATEMENT

- To work on the data of the loan applications to find out the strong indicators of default, to minimize the credit loss for the banking and financial sector companies.
- Perform analysis using EDA techniques to understand and pick the driving factors behind loan defaults.
- The analysis can be utilized by the banks for risk assessment during provision of loans.

ANALYSIS APPROACH

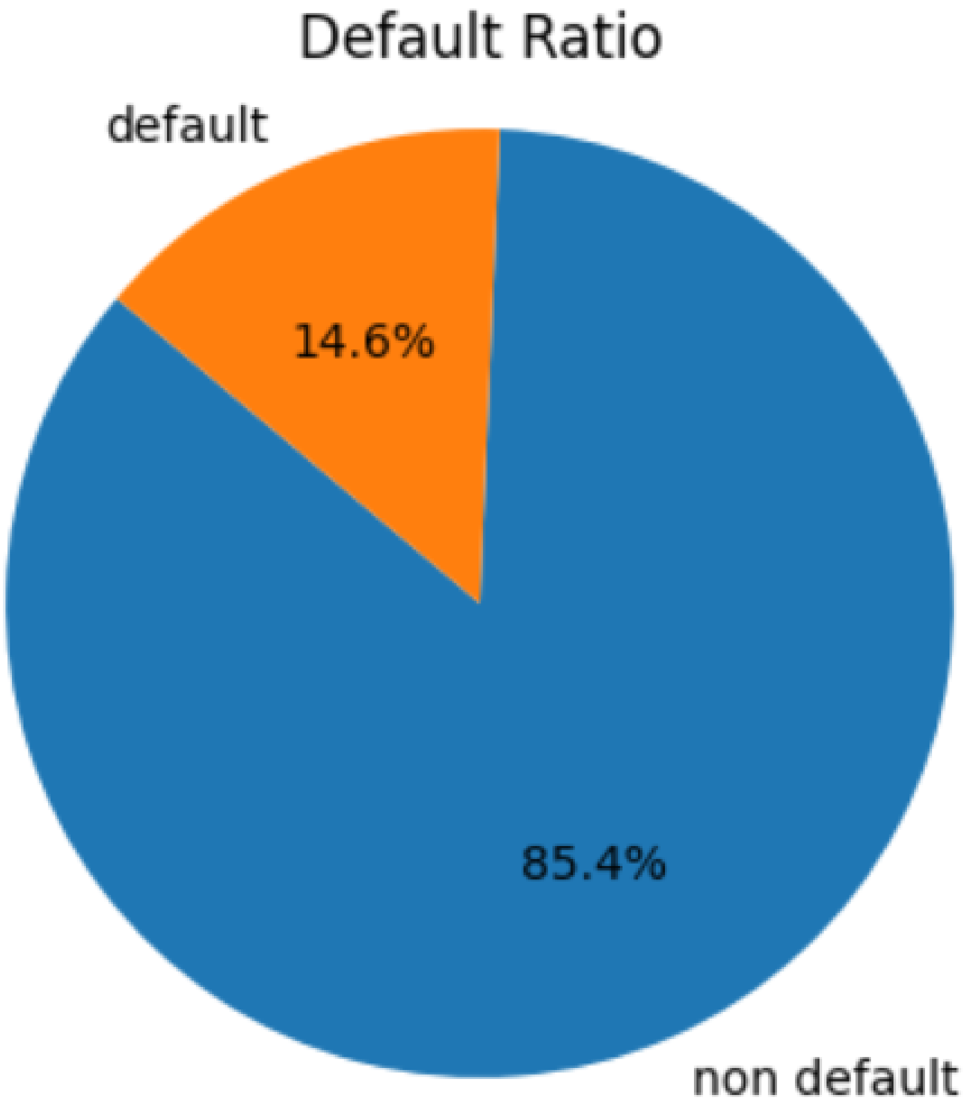
1. Initially reading of the dataset is done to observe the basics of the data like number of rows and types of columns
2. Features are categorized into 3 types: '**consumer attributes**', '**loan attributes**', and '**consumer behavior attributes**' . Only consumer attributes and loan attributes considered for analysis as consumer behavior attributes come to play only after loan is sanctioned. Here we must do risk analysis at the time of loan application.
3. Data cleaning is done at different stages. Initially the columns with more than 50% null value are removed.
4. Null values are filled with mode or mean as required at different columns.

CATEGORISATION OF ATTRIBUTES

CONSUMER	LOAN	CONSUMER BEHAVIOR
'emp_title', 'emp_length', 'home_ownership', 'annual_inc', 'verification_status', 'purpose', 'title', 'addr_state', 'zip_code', 'application_type', 'pub_rec_bankruptcies'	'loan_amnt', 'term', 'int_rate', 'installment', 'grade', 'sub_grade', 'issue_d', 'loan_status', 'dti', 'policy_code'	'last_pymnt_d', 'last_pymnt_amnt', 'last_credit_pull_d', 'delinq_2yrs', 'earliest_cr_line', 'inq_last_6mths', 'open_acc', 'pub_rec', 'total_pymnt', 'total_pymnt_inv', 'total_rec_prncp', 'total_rec_int', 'total_rec_late_fee', 'revol_bal', 'revol_util', 'total_acc', 'out_prncp', 'out_prncp_inv', 'recoveries', 'collection_recovery_fee' ETC

UNIVARIATE/BIVARIATE/MULTIVARIATE ANALYSIS

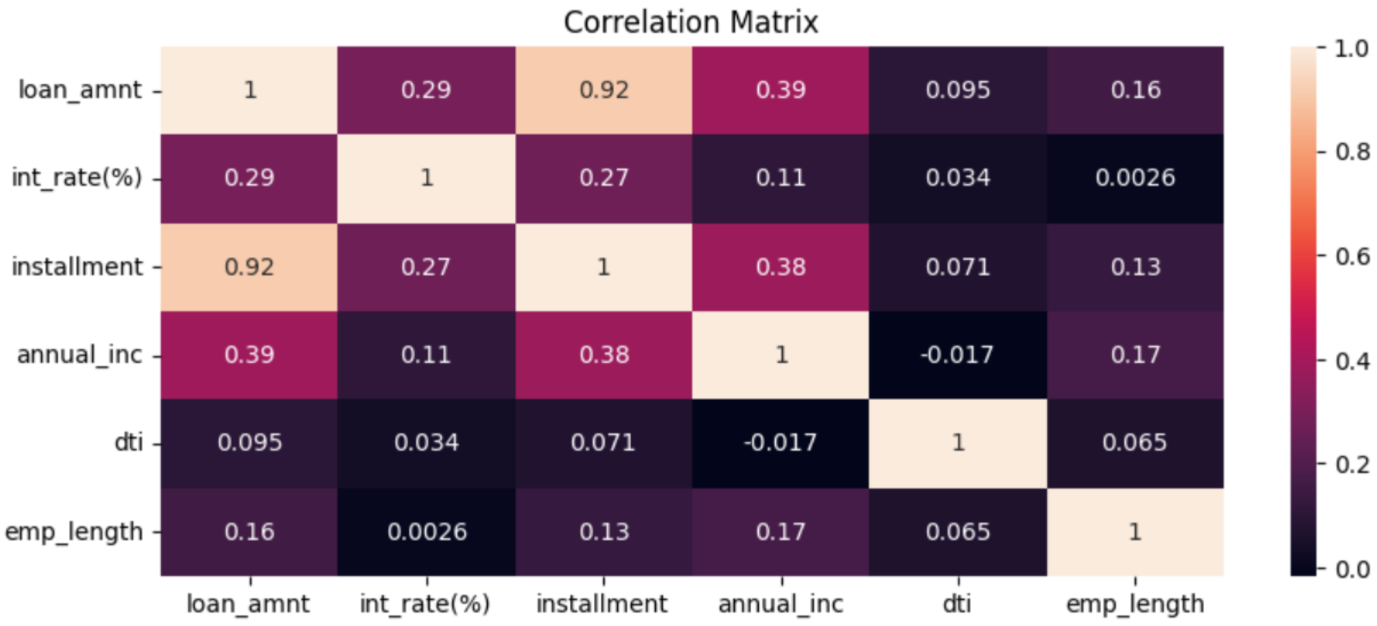
Graphs with observations are provided



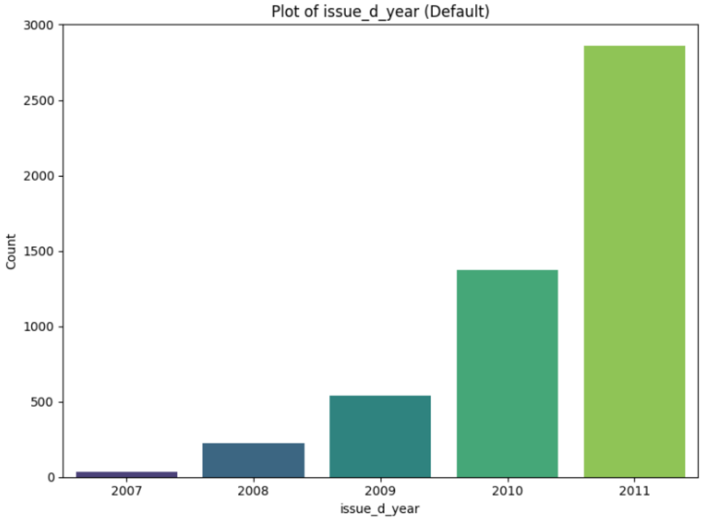
DEFAULT / NON DEFAULT PROPORTIONS

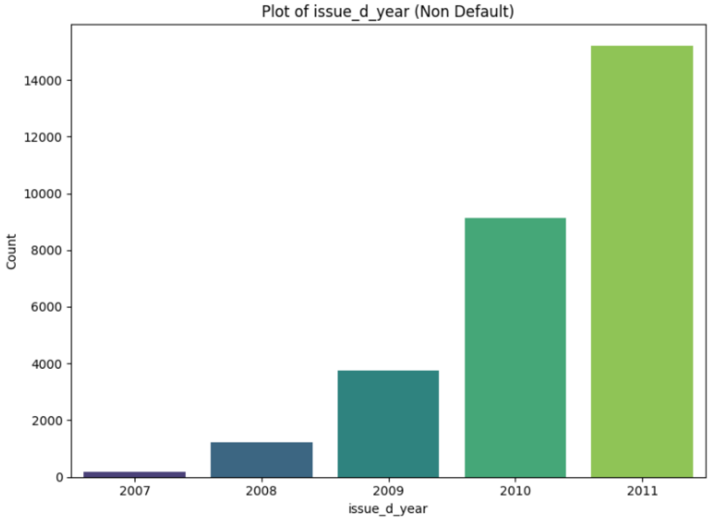
- In the data it is observed that only 14.6 % of the applications are default cases

CORRELATION MATRIX



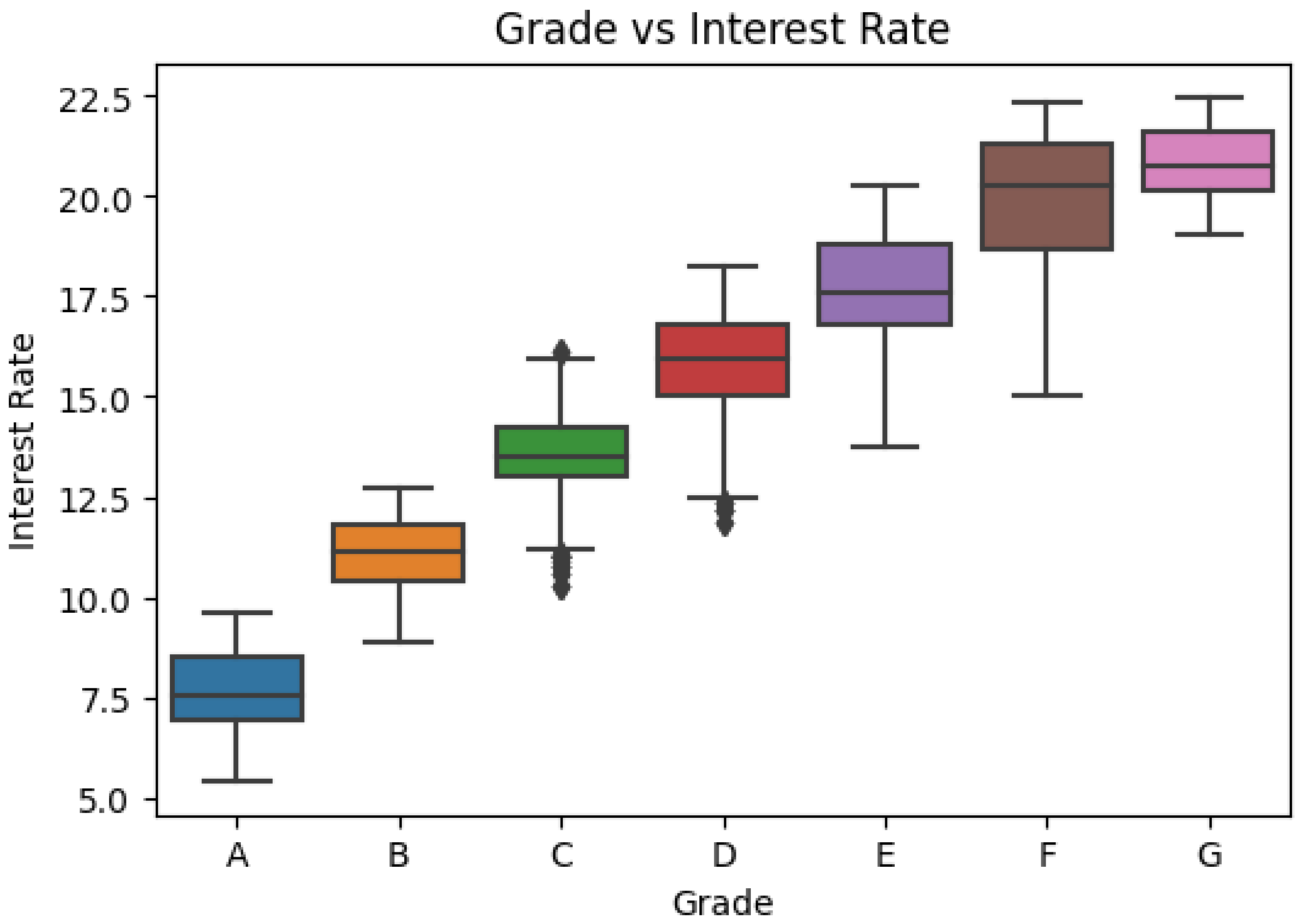
- There is correlation between loan amount and instalment which is quite obvious, as instalment tend to increase with increase in loan amount





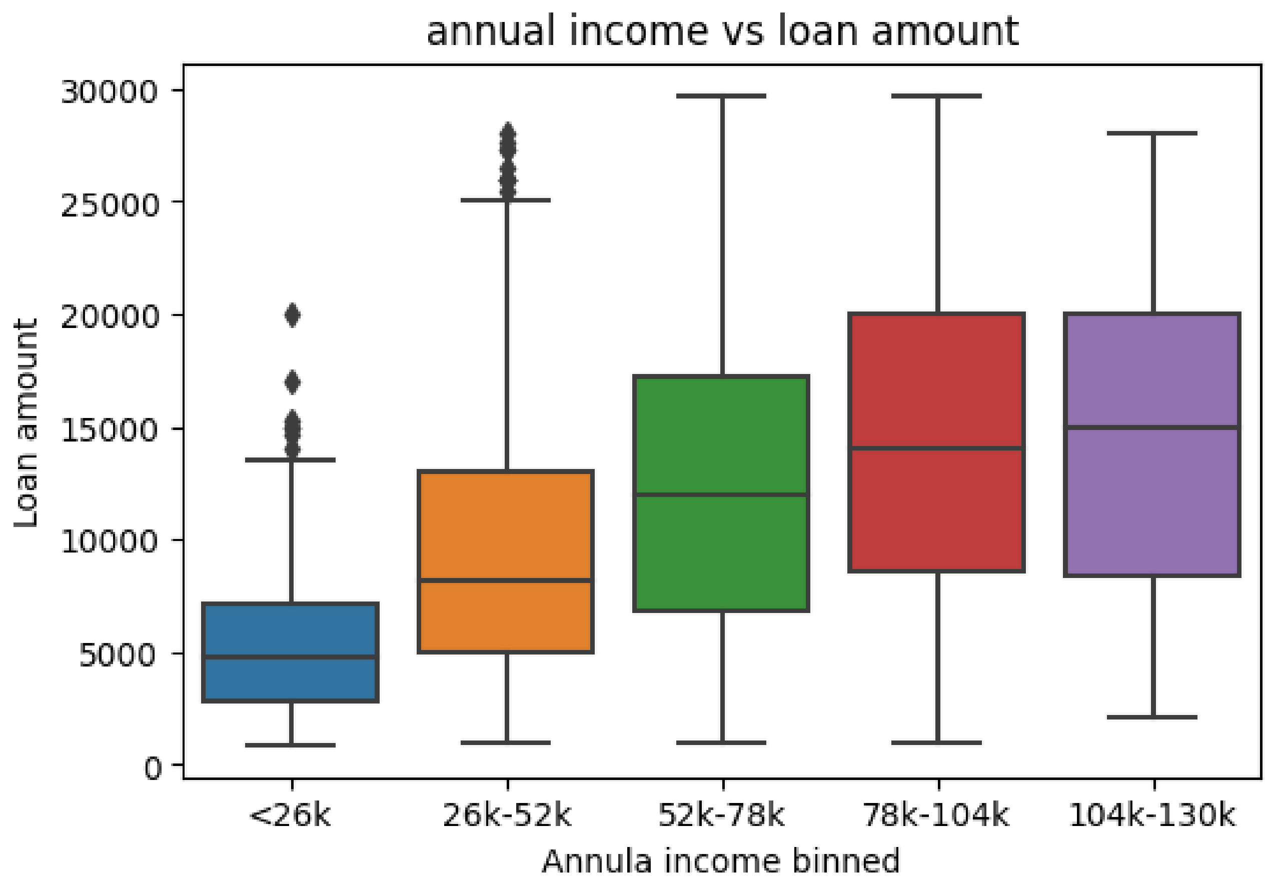
YEAR OF LOAN ISSUE

- Though it appears like the number of default loans increased with the year , this is pattern that is seen in Non default loans also
- Hence it can be inferred that volumes of loans itself is increased over time and nothing in specific can be told about default loans over time



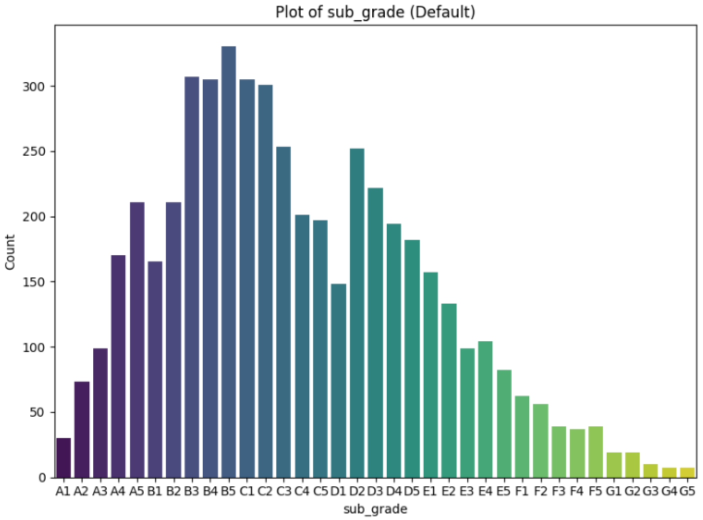
GRADES VS INTEREST RATES

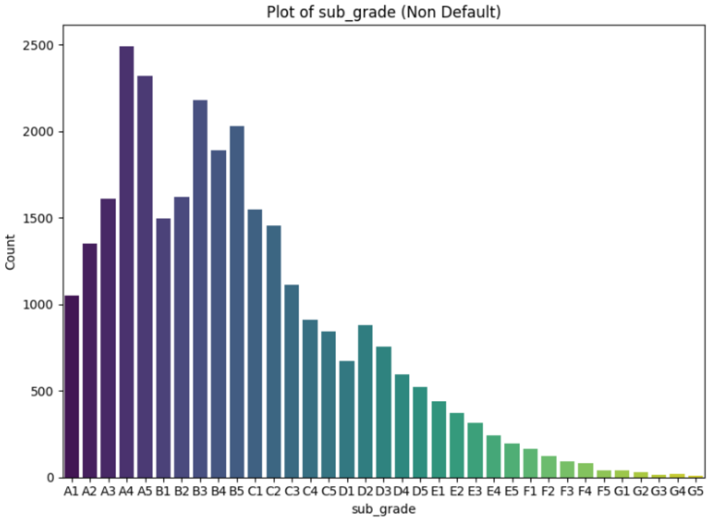
- As the loan's grade gets risker the interest rates have constantly increased



ANNUAL INCOME VS LOAN AMOUNT

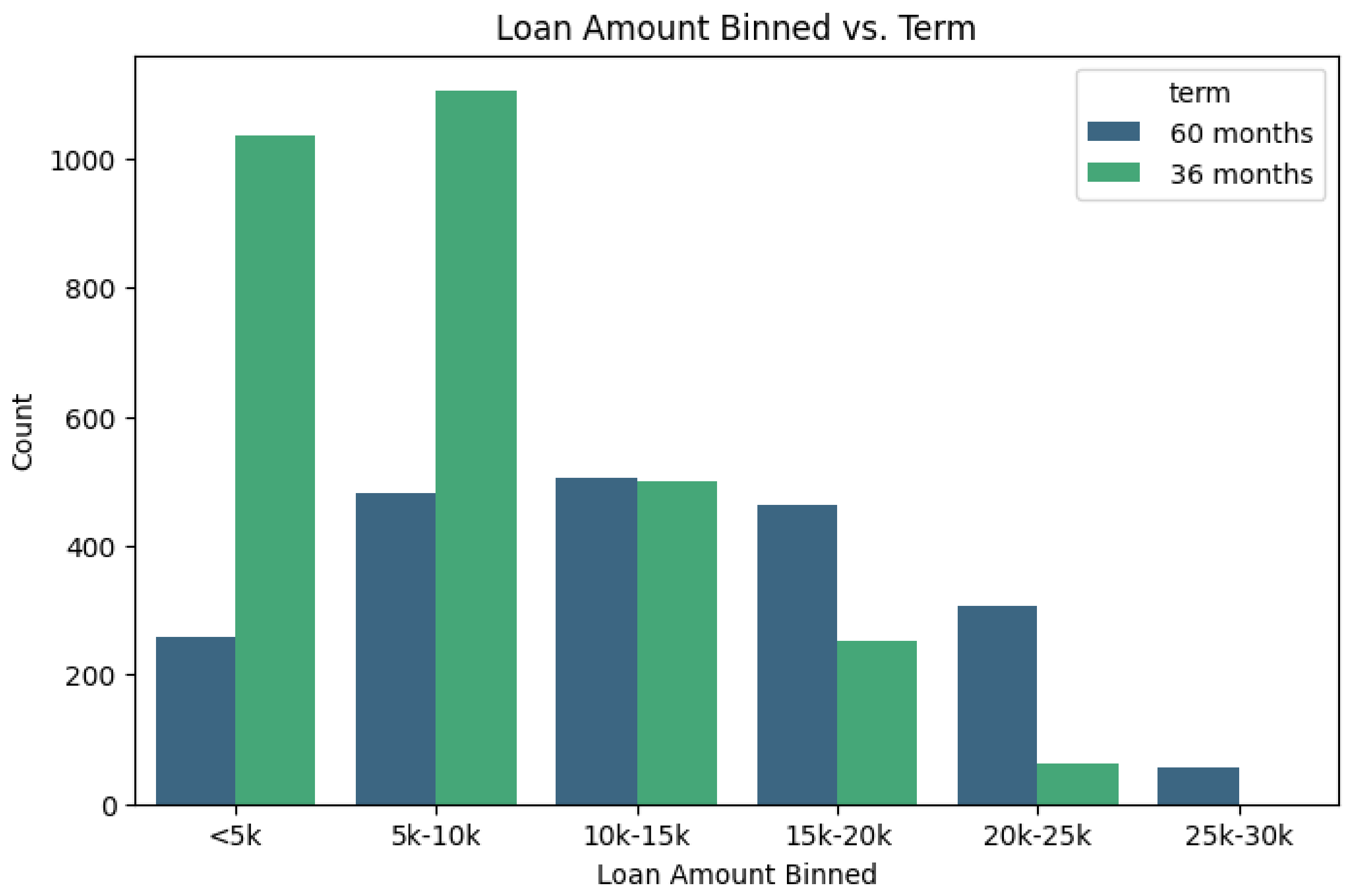
- The higher loan amounts are claimed by people in the higher income brackets





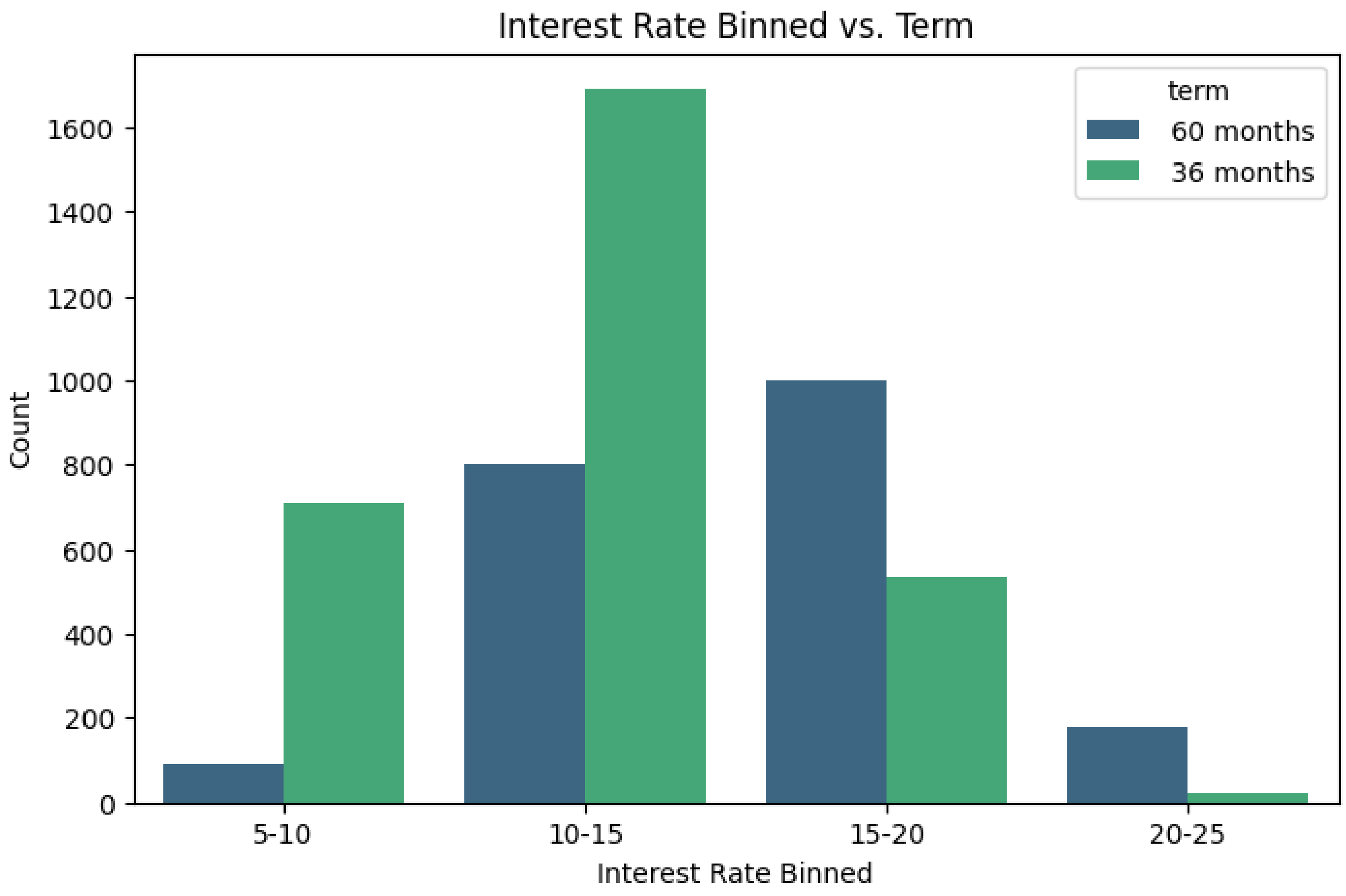
SUB-GRADE

- Loans with safer grades like A,B have less chances of default
- Loans with grades like D,E have more chances of default



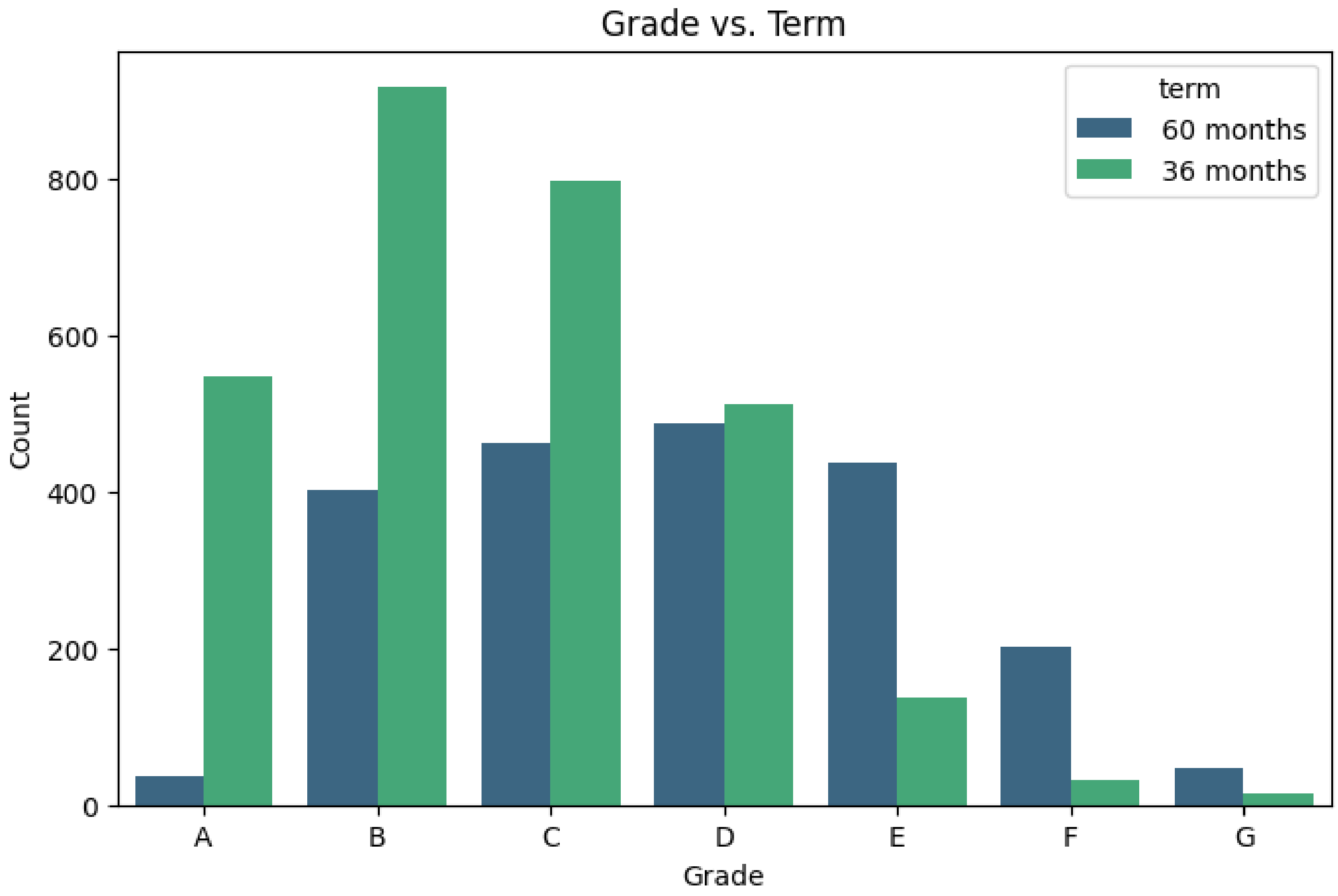
LOAN AMOUNT VS TERM

- In the lower loan amounts more people have preferred 36 month term, but at the higher loan amounts more people have preferred 60 month term



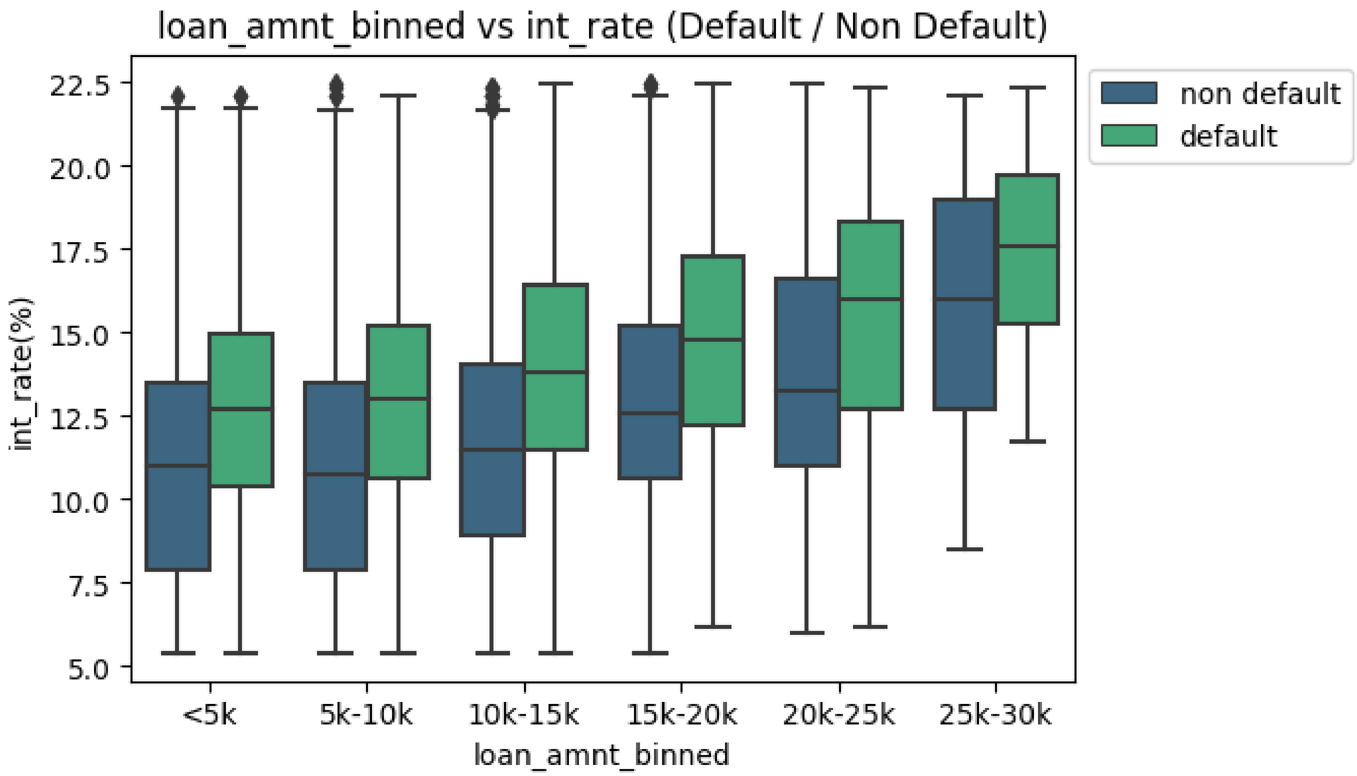
INTEREST RATE VS TERM

- In the lower interest rates more people have preferred 36 month term, but at the higher interest rates more people have preferred 60 month term



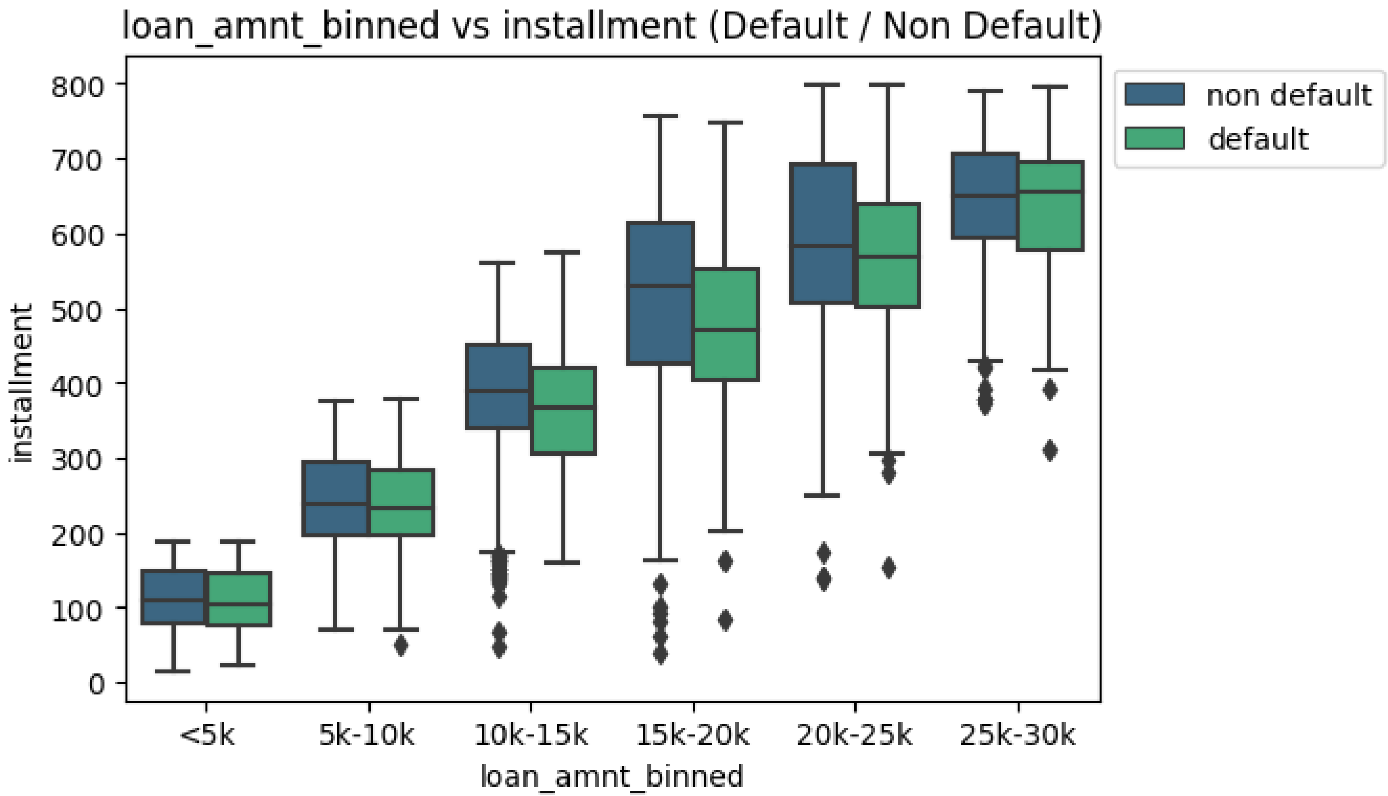
GRADE VS TERM

- For safer loans people have preferred 36 month term, but as the grade gets risky more people have preferred 60 month term



LOAN AMOUNT VS INTEREST RATES

- In all the annual income ranges the interest rate is highe for default cases
- Lower interest rates are paid back but higher interest rates are defaults in all income ranges



LOAN AMOUNT VS INSTALMENTS

- There is a casual observation that installments increase with increase in loan amount, which is quite expected

SUMMARY

- Major driving factors are:
  - Annual income
  - Term
  - Interest rate
  - Grade

RECOMMENDATIONS

- It's safer to give loans at 36 months than 60 months
- Loans of '5-10%' interest are less likely to default, whereas loans of '15-20%' and '20-25%' are more likely to default
- More safer grades(like grade A,B) are less likely to default, caution to be observed while giving loans of C,D,E,F,G grades
- Higher income people tend to take high amount loan, so to give high amount loans it's suggested to target high income people
- People of higher grades and higher term (60 month) are more likely to default, hence suggest them to take 36month term if they prefer higher grade loans
- In all loan amount ranges, the lower interest rates are paid back but higher interest rates are default. To avoid default, it suggested to decrease interest rates
- Year, home ownership, location have no special impact on the default

THANK YOU