Lending Club Case Study

By

Baskaran Murugesan

Varun Juyal

Purpose

- Lending club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

- The purpose of this data analysis is to measure the degree of risk that the borrower presents to the lender. It includes, based on the estimated probability assigning measurable numbers to the defaulters of the borrower. In this case study we have such dataset, and our objective is to determine the potential borrowers to determine their ability to meet debt obligations.
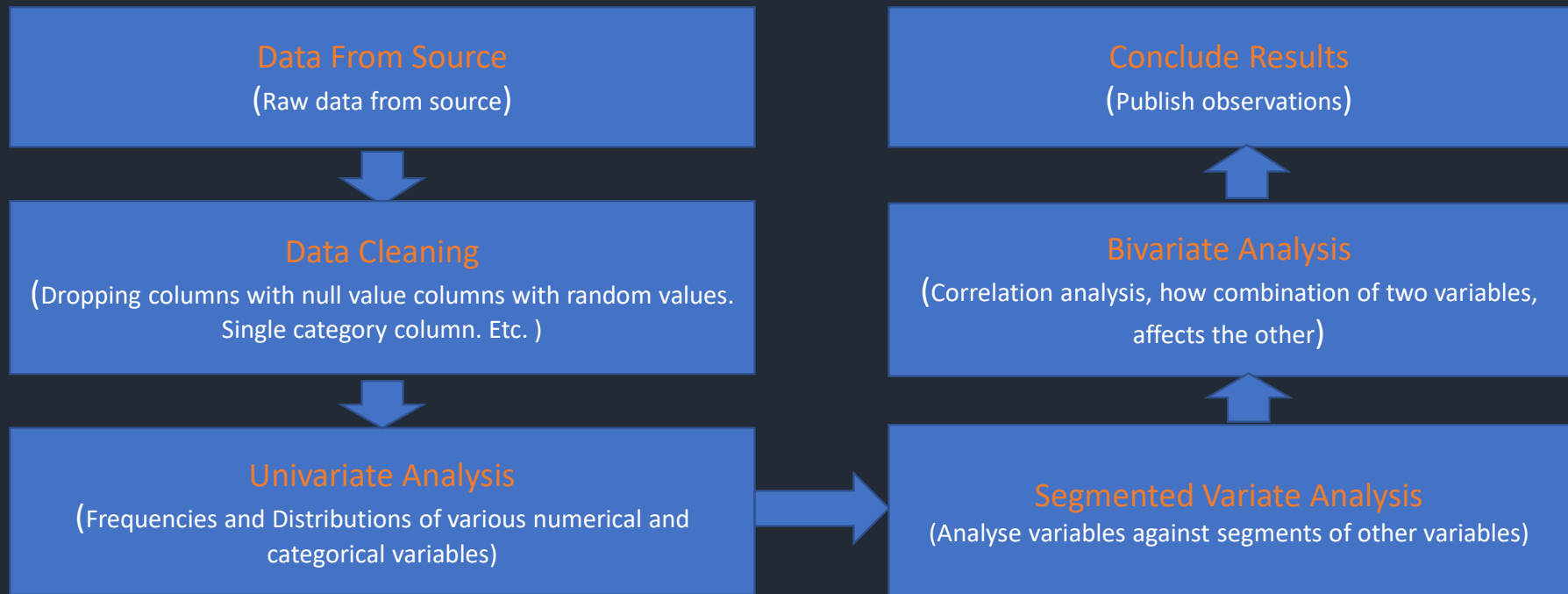
Objective

- The objective of data analysis, when a person applies for a loan, there are **two types of decisions** that the company should take. Depends on the observations and results from the data analysis either Loan should be **Accepted** or **Rejected**.

1.**Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:

  1. **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
  2. **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
  3. **Charged-off:** Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

2.**Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements
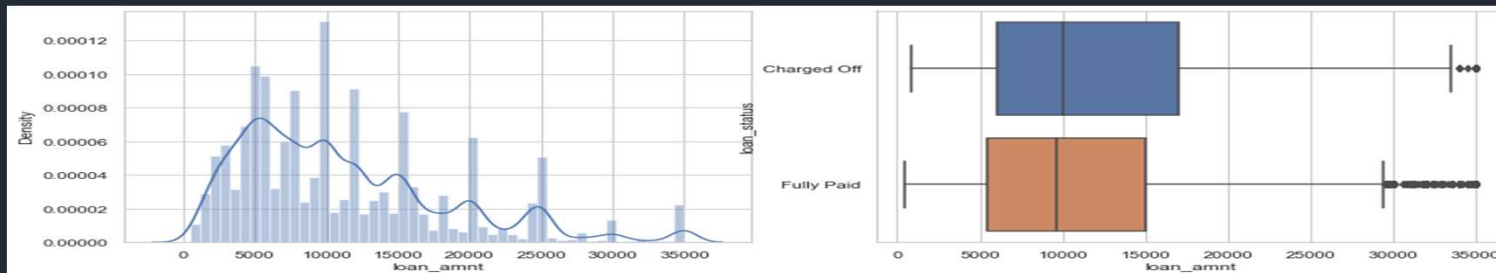
Analysis

Data From Source (Raw data from source)

↓

Data Cleaning (Dropping columns with null value columns with random values. Single category column. Etc.)

↓

Univariate Analysis (Frequencies and Distributions of various numerical and categorical variables)

→

Segmented Variate Analysis (Analyse variables against segments of other variables)

↑

Bivariate Analysis (Correlation analysis, how combination of two variables, affects the other)

↑

Conclude Results (Publish observations)

Analysis – Loan Status

- As per the data around **14%** of loans have been defaulted.

- Out of all defaulted loans only **57 %** percentage of loans have been recovered.
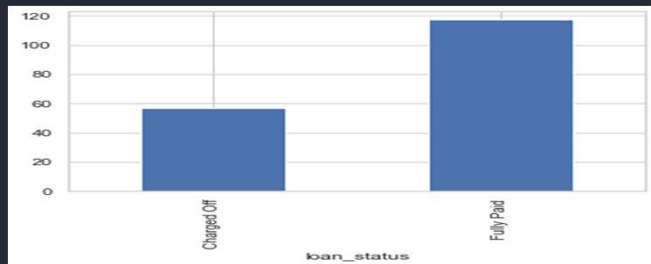
- Overall, the loan amount distribution is slightly right sided with mean greater than the median. Seventy five percent(75%)of the loans granted are below 15000.

- It is observed that there is a spike in loan amounts around each 5000 limit.

Univariate Analysis

- The percentage of charged off loans increases with the loan amount range. Even though, most of the loans are below 20000 amount. The higher loans, carry significant risk of default.

- As expected Grade A has less loan defaulters than the Grade F and Grade G.
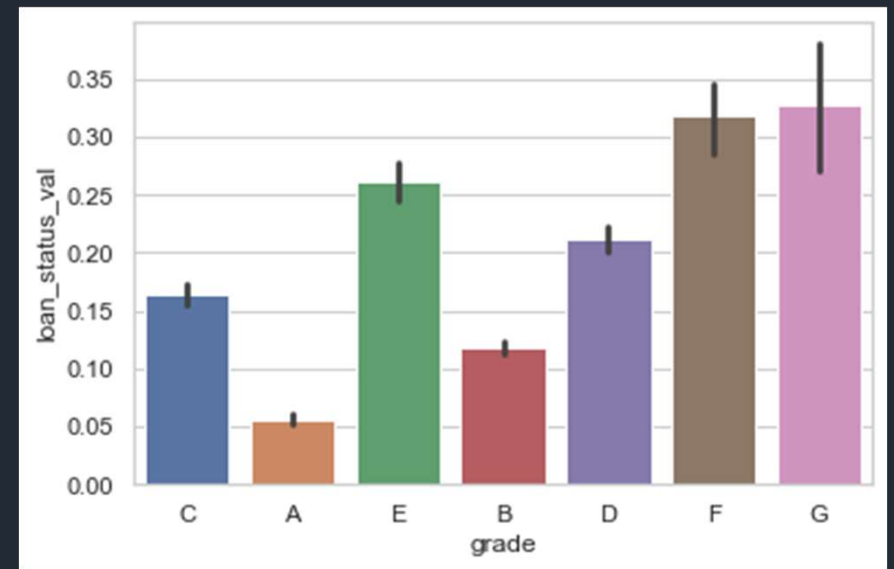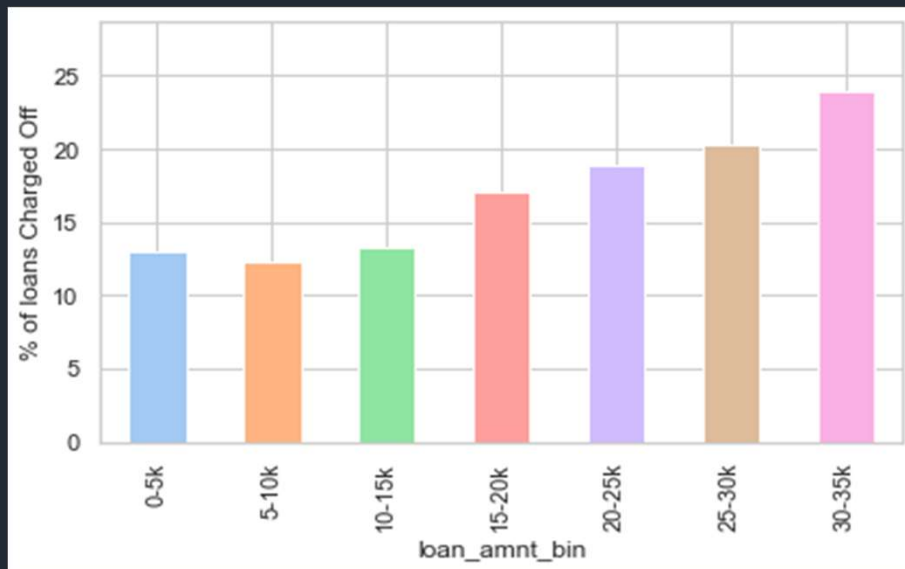
Univariate Analysis – continued…

- Charged off bar is higher for the long term loans. It is risky to give higher loan amount with long term.
- As expected the defaulter rate is lesser for A1 and the same is higher for F5. The defaulter rate increase as the sub grade goes down.
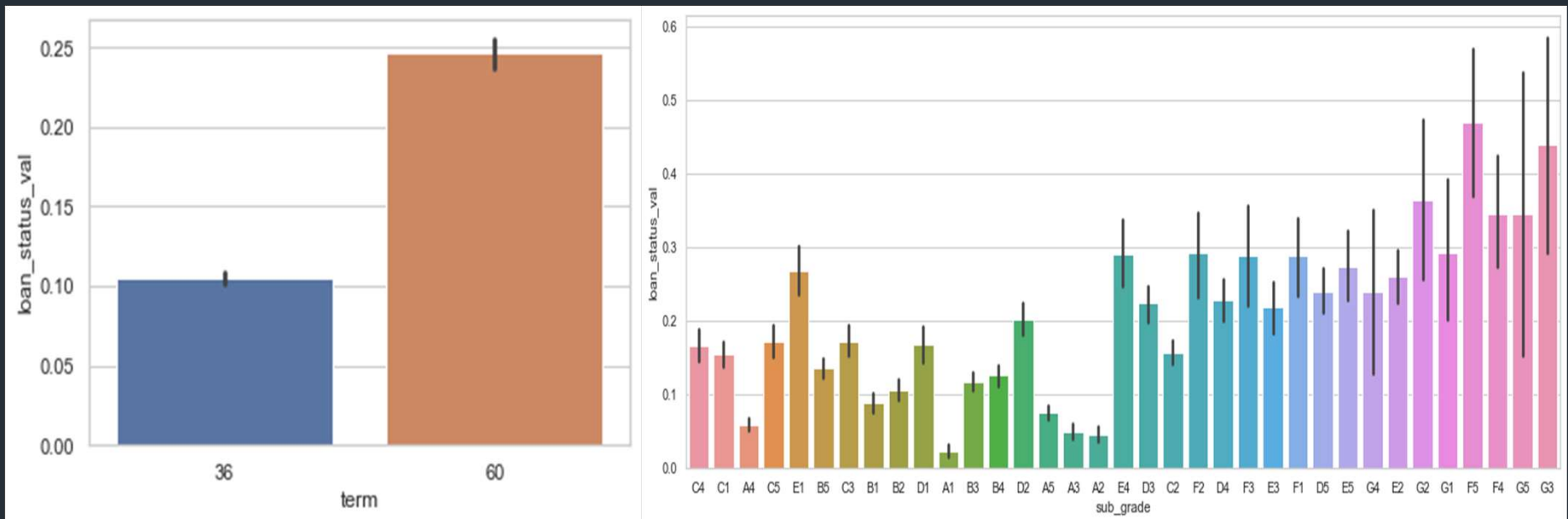
Univariate Analysis – continued...

- The variable home_ownership did not make much differences in defaulter rate.
- Interestingly the loans with Verified status has more loan defaulters than Not Verified.

Univariate Analysis – continued…

- While analyzing purpose of loans, Small business has more loan defaulters than the renewable energy and education loans.
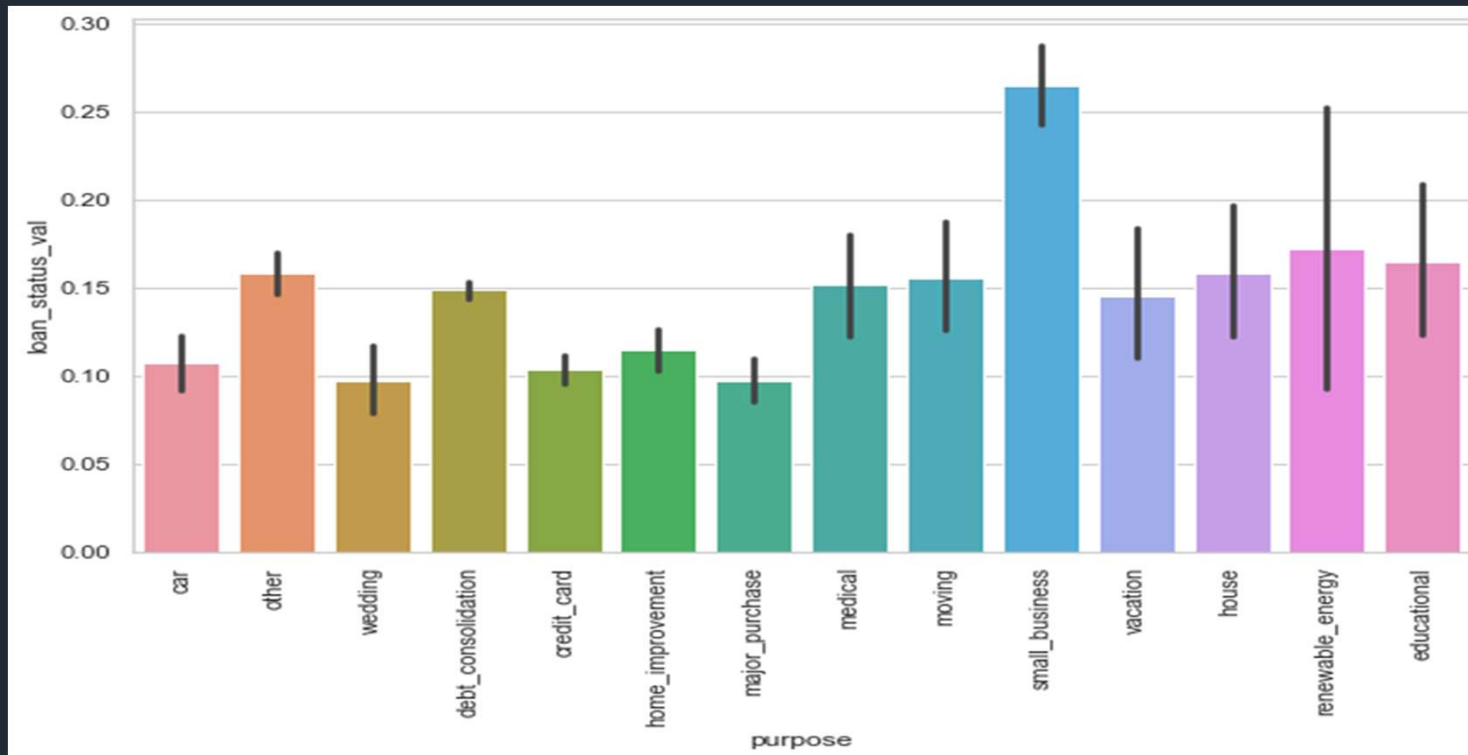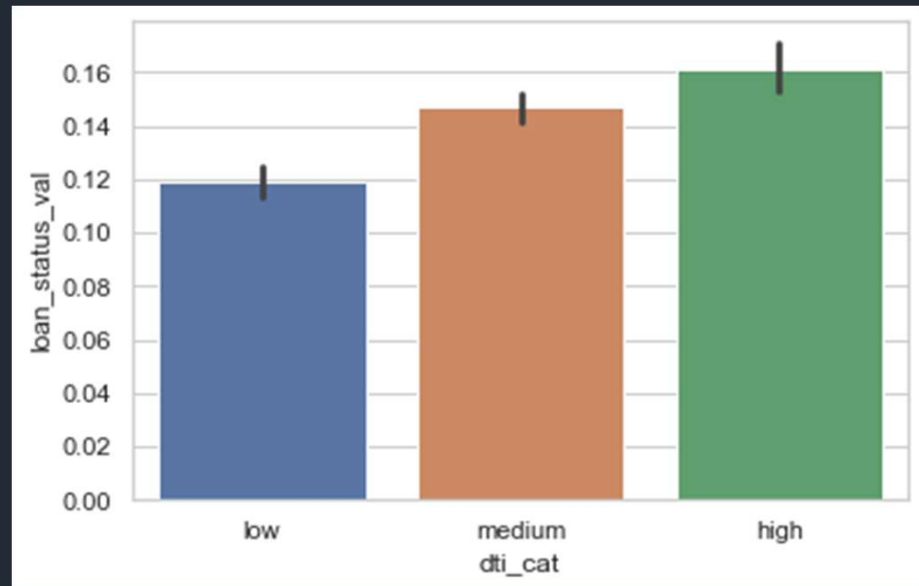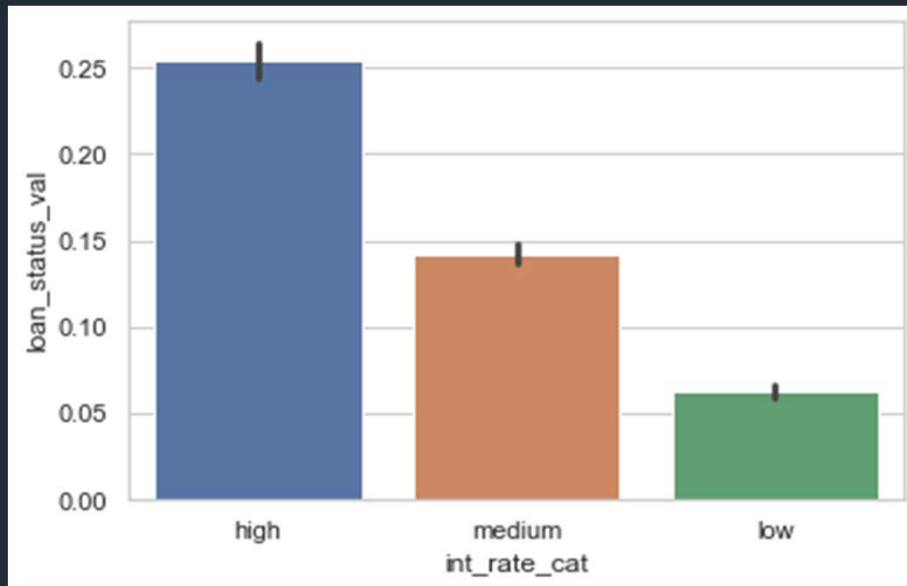
Univariate Analysis – continued…

- Loans distribution count has increased steadily across years.
- The default rate had continuously declined till 2010, and then suddenly increased in 2011.
- More loans has distributed in the second half of the year.
- When compared to defaulter status there is no much difference.
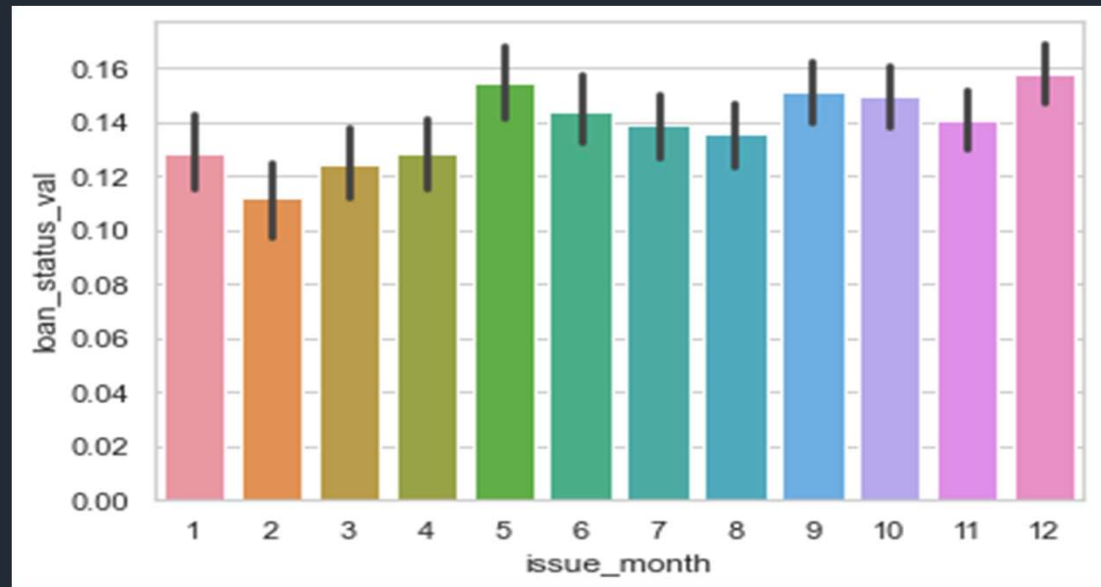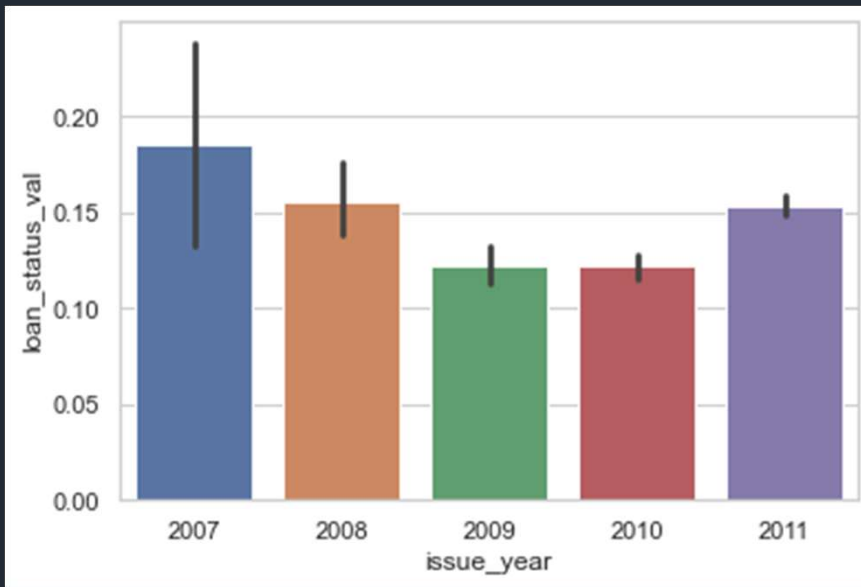
Univariate Analysis – continued...

- From the below chart, it is evident that the loan defaulters increases when the interest rate is high.
- As expected, high dti (Debit to income ratio) resulted in higher default rates.
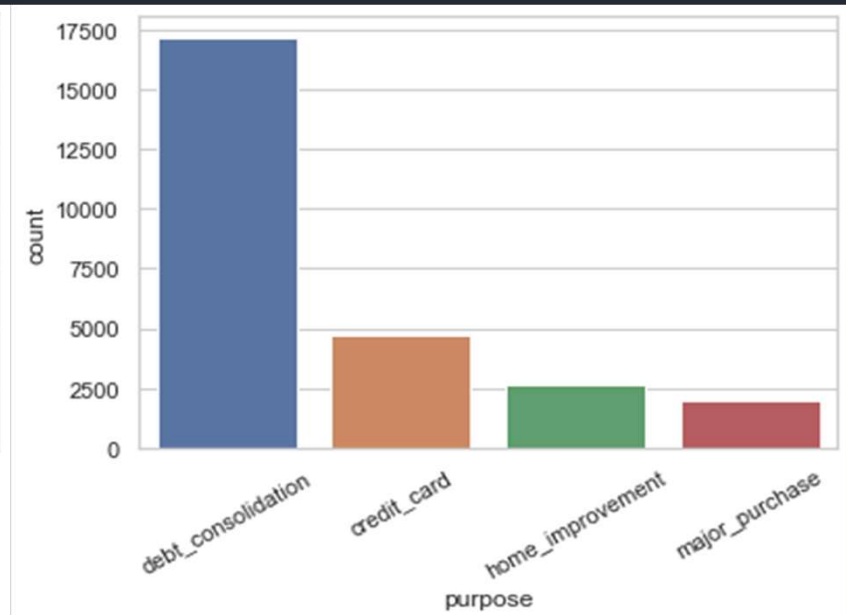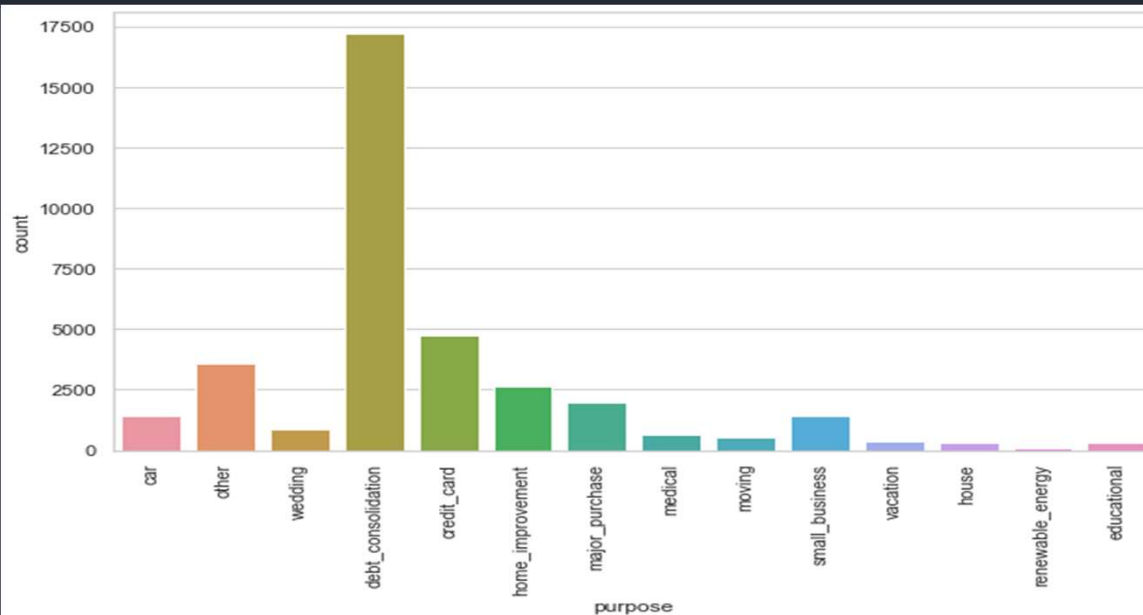
Univariate Analysis – End.

- The default rate increase as the installment amount goes up.
- The charged off rate is high when the annual income is low.
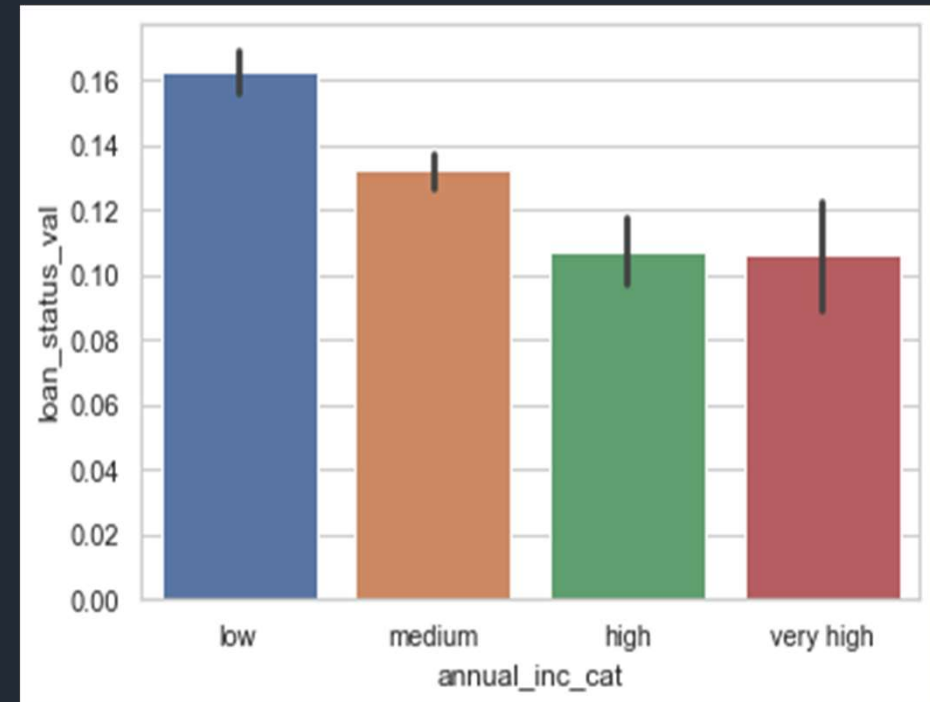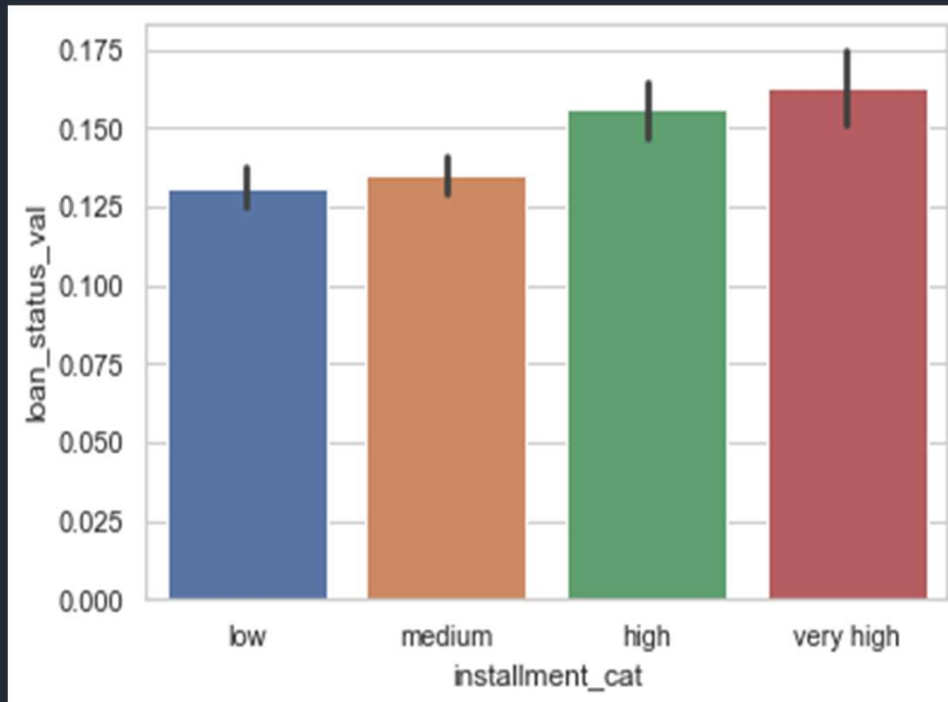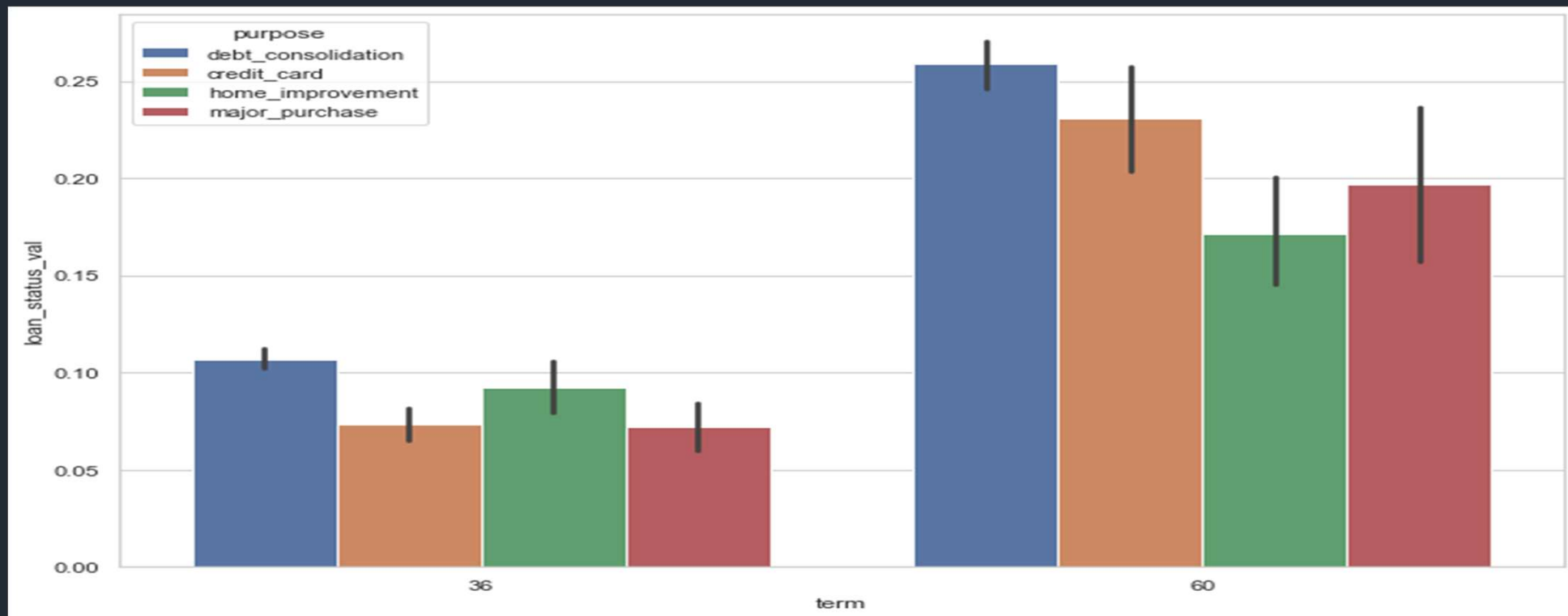
Segmented Univariate Analysis

So far we have seen different variables such as annual income, purpose of loan, grade, debit to income ratio and interest rate affected the loan defaults.
In the credit industry, one of the most important factors affecting default is the purpose of the loan - home loans perform differently than credit cards, credit cards are very different from debt consolidation loans etc.
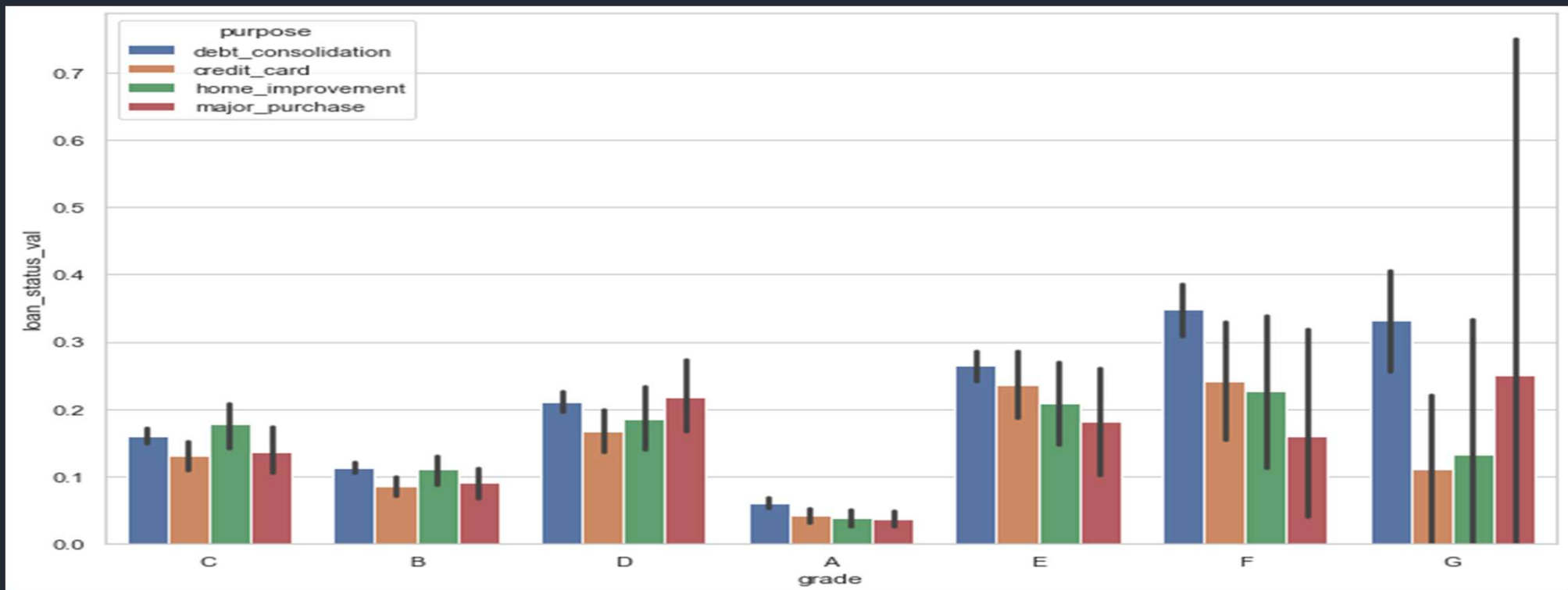
Segmented Univariate Analysis – continued...
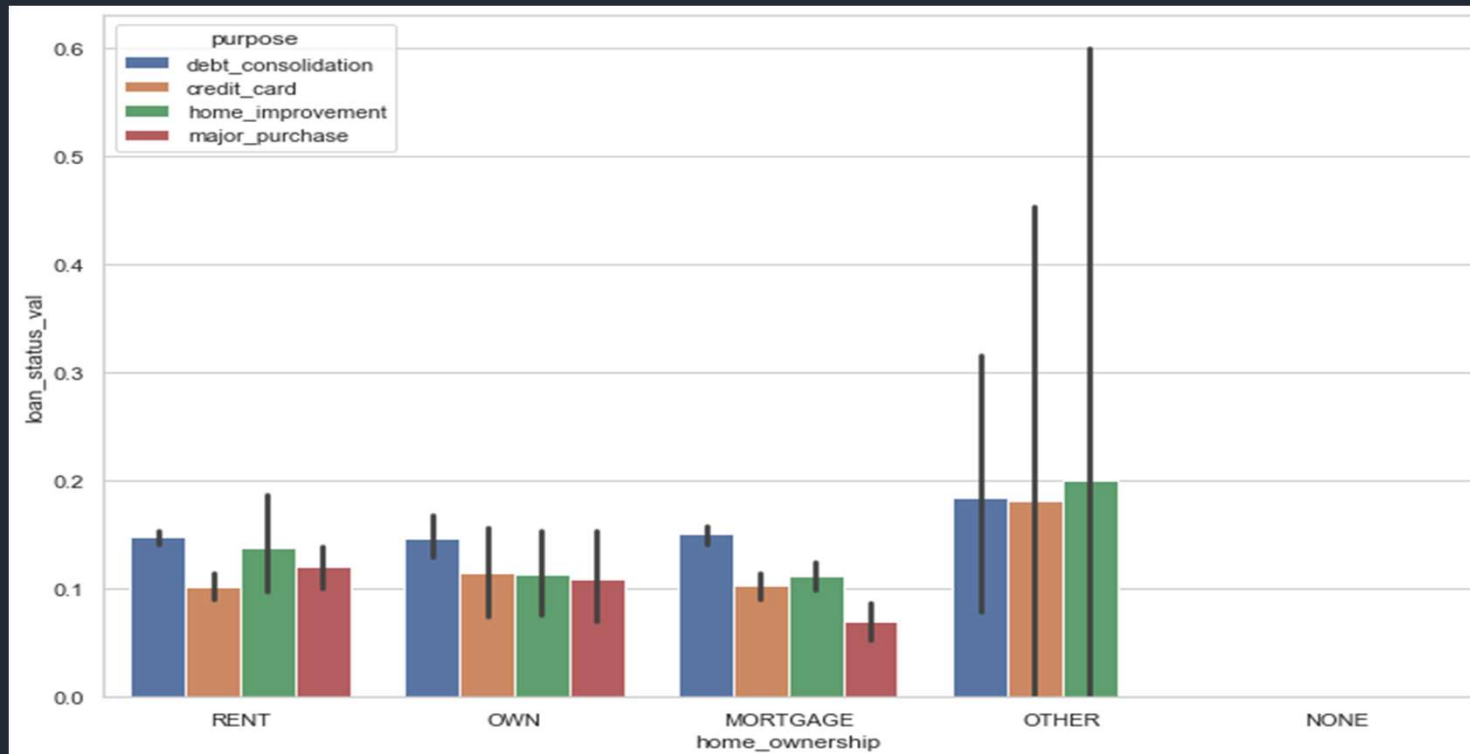
- Various purposes vs term

Segmented Univariate Analysis – continued…
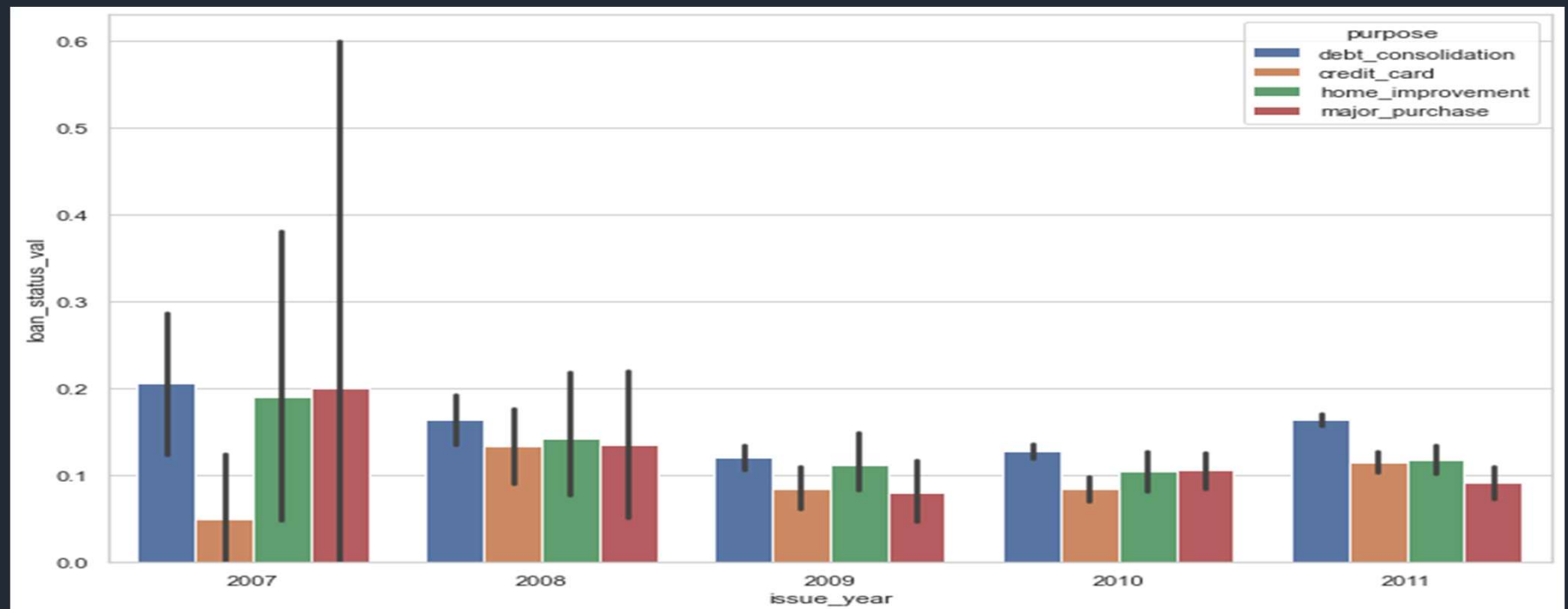
- Various purposes vs grade

Segmented Univariate Analysis – continued…

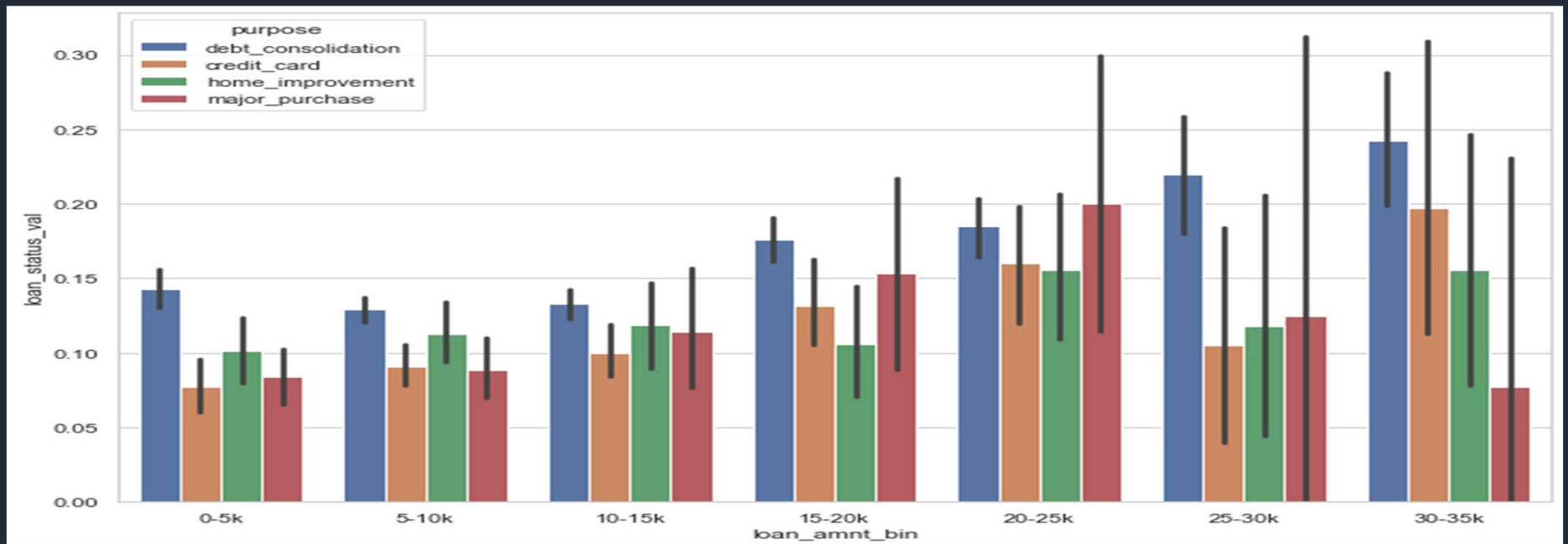- Various purposes vs home_ownership

Segmented Univariate Analysis – continued...
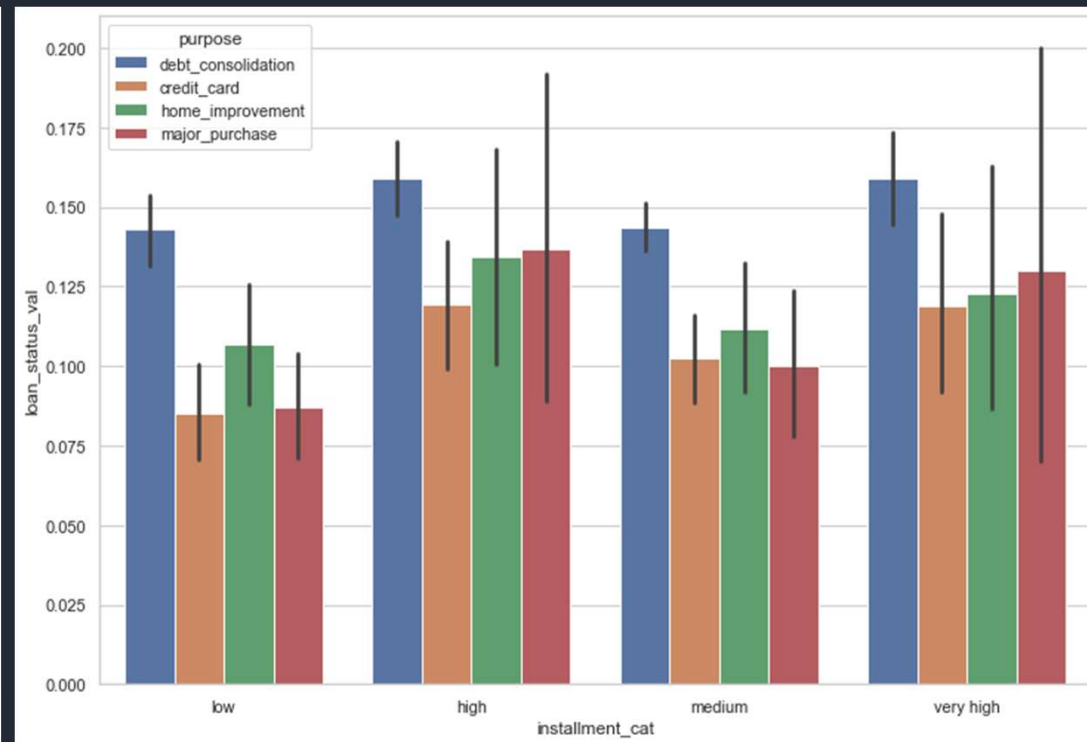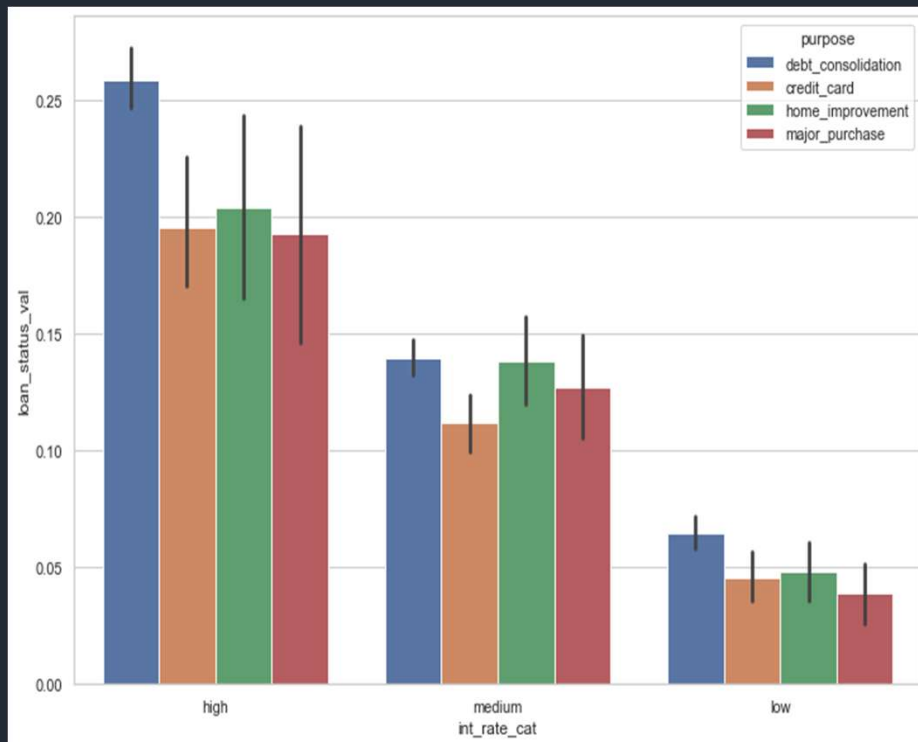
- Various purposes vs issue year

Segmented Univariate Analysis – continued...

- Various purposes vs loan_amnt_bin

Segmented Univariate Analysis – continued...

- Various purposes vs int_rate_cat
- Various purposes vs installment_cat

Segmented Univariate Analysis – End

- Various purposes vs annual_inc_cat

- In general, in segmented univariate analysis out of four different purposes debt consolidation loans have the highest default rates when compare to other categories.
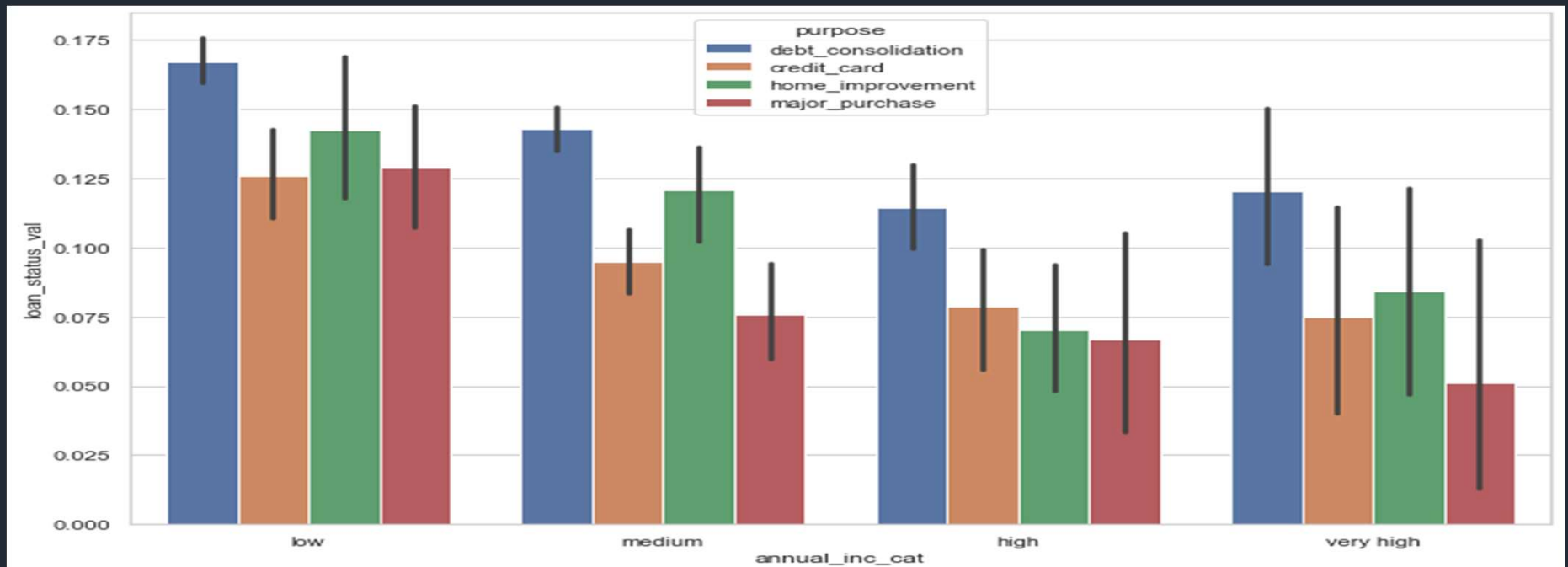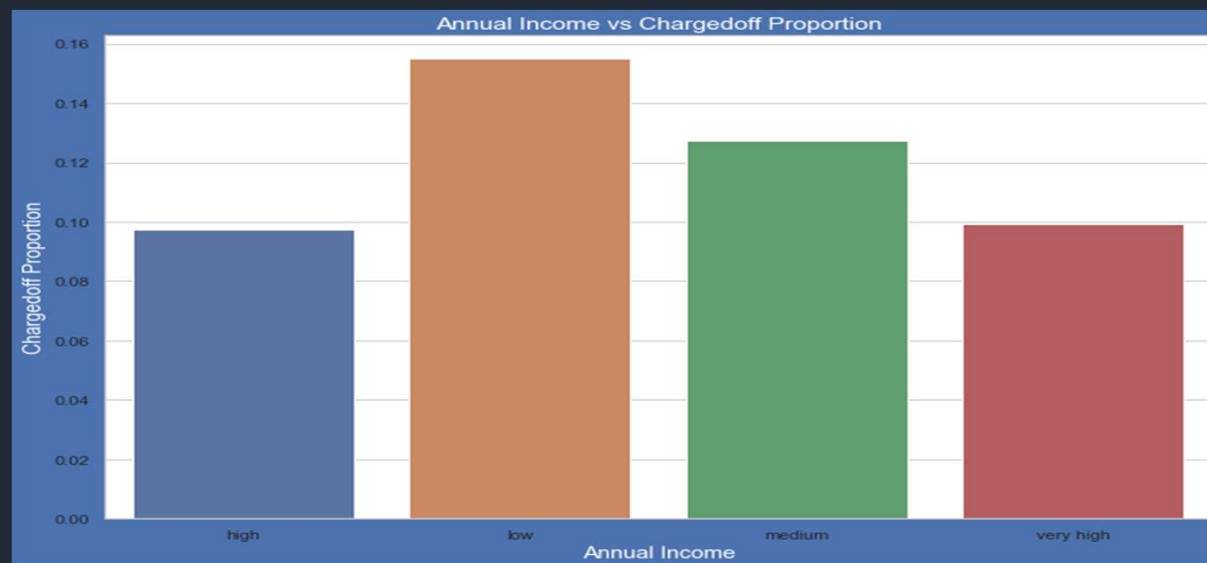
Bivariate Analysis

In this part of analysis, lets try to analyze the loan status against some important columns which might have played and important role in loan defaults. Lets try to find proportion of loan defaults in some categories.

Income range category vs Loan Status

- ✓ High income range has less chances of loan defaults.
- ✓ Low income range has high chances of loan default.
- ✓ Therefore, the observation is that, with increase in annual income loan default proportion got decreased.

Bivariate Analysis - continued

Purpose vs Loan status

- Renewable_energy where the loan defaulter proportion is higher when compare to other categories.
- small_business applicants have higher chances of getting loan defaults.
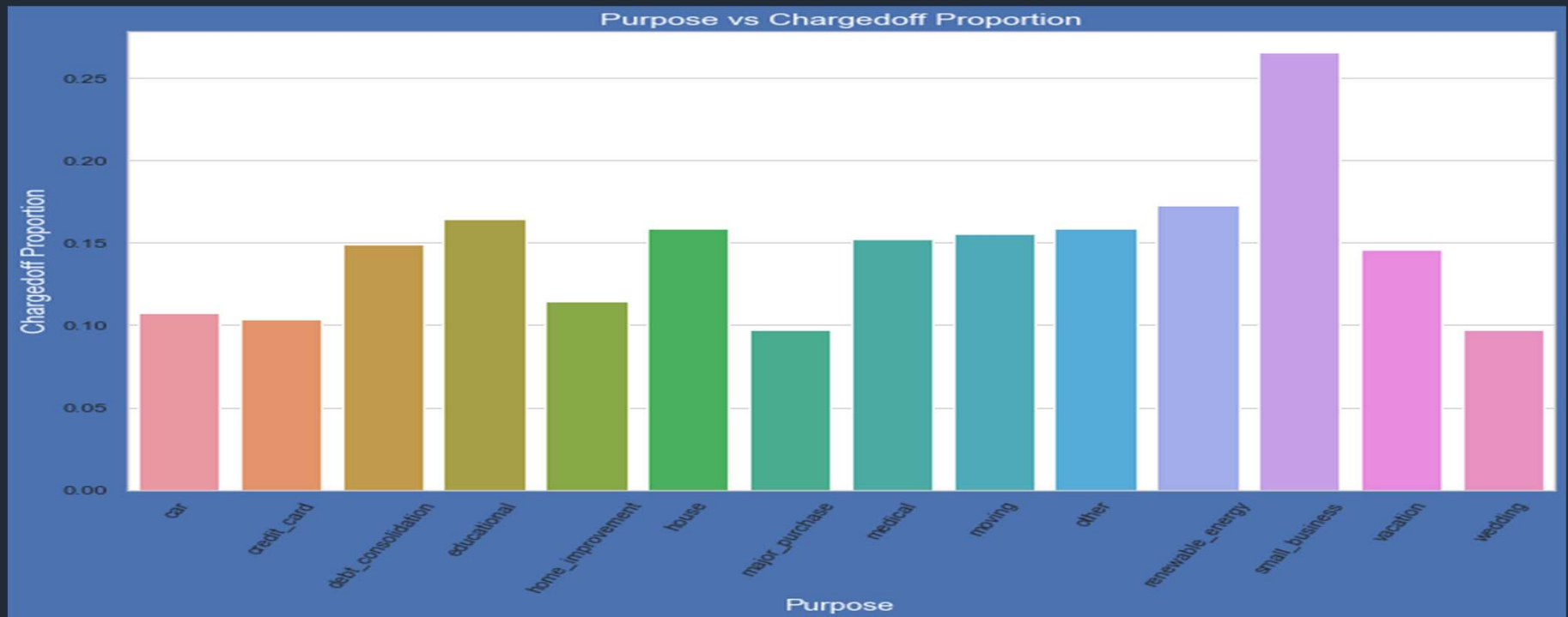
# Bivariate Analysis - continued

Grade vs Loan Status

- Grade A has very less loan defaults among all other grades
- Grade F has the highest loan defaults among all other grades
- It is clearly visible that the borrowers with top grade, brings less risk to the credit

Purpose vs Chargedoff Proportion
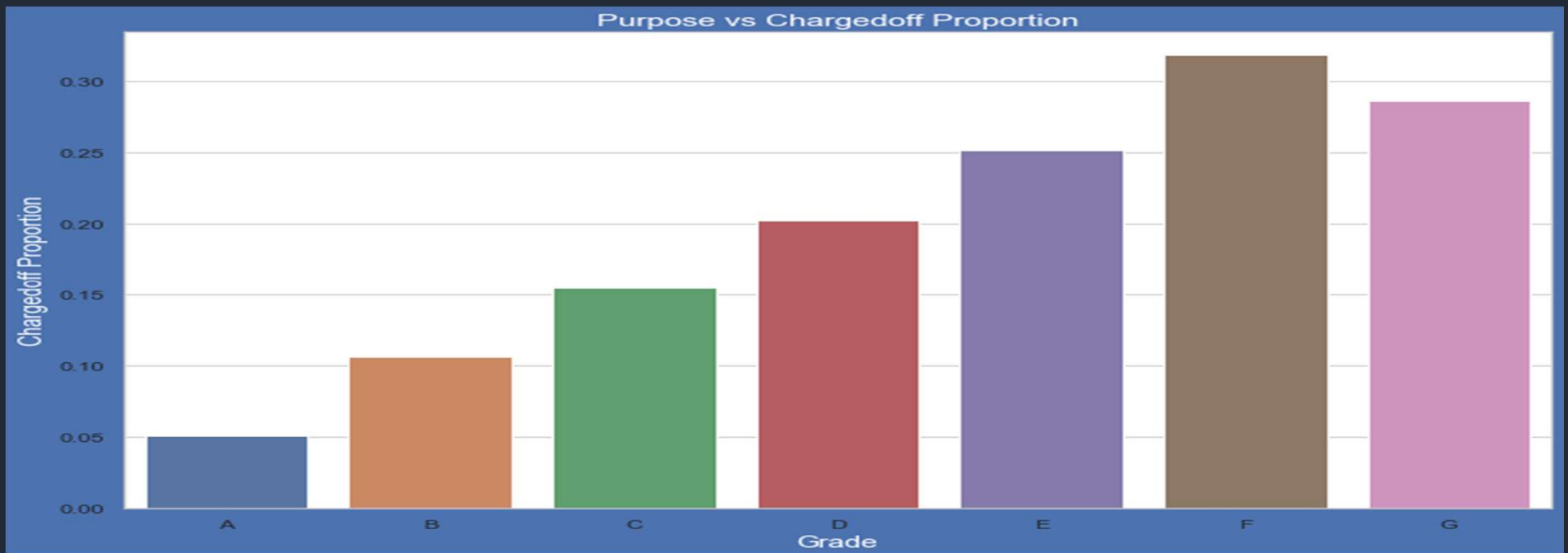
Bivariate Analysis - End

Interest Rate Category vs Loan Status

Loan default proportion is increasing with higher interest rates.
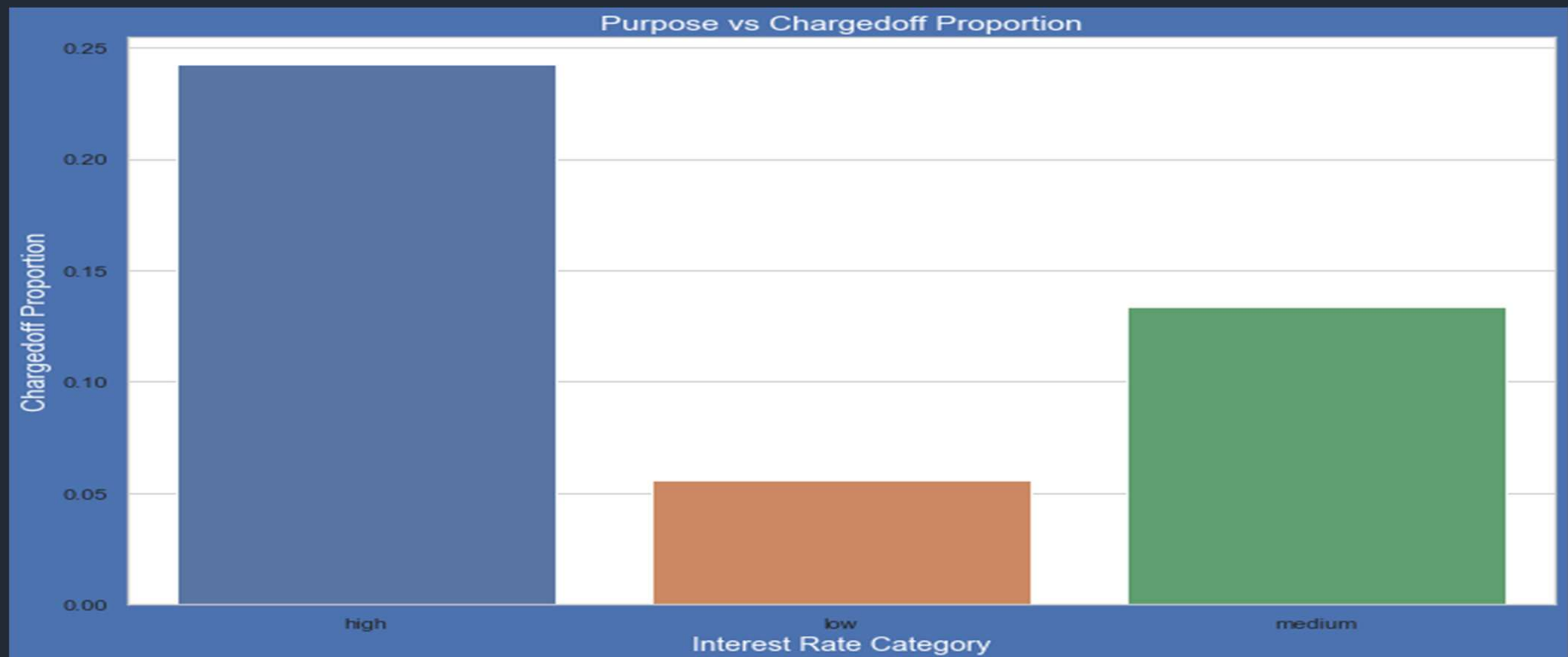
Conclusion:

- Reduce approving loans for small business.
- Increase approving loans for A grade borrowers, at the same time decrease loan approval for low grade borrowers, this would reduce credit risk.
- Stop approving loans for lower income borrowers.
- Similarly, stop approving loans for borrowers with high debit income ratio.
- Loans shall be approved if the loan amount is less than 20% of the annual income.
- Long term(60 months) loans has the higher rate of loan defaulters, so reduce approving long term loans.

Thank You