LENDING CLUB

CASE STUDY

Group Members:

Sai Shashank Kayasanoori

Harshita Sai Adagarla

UpGrad

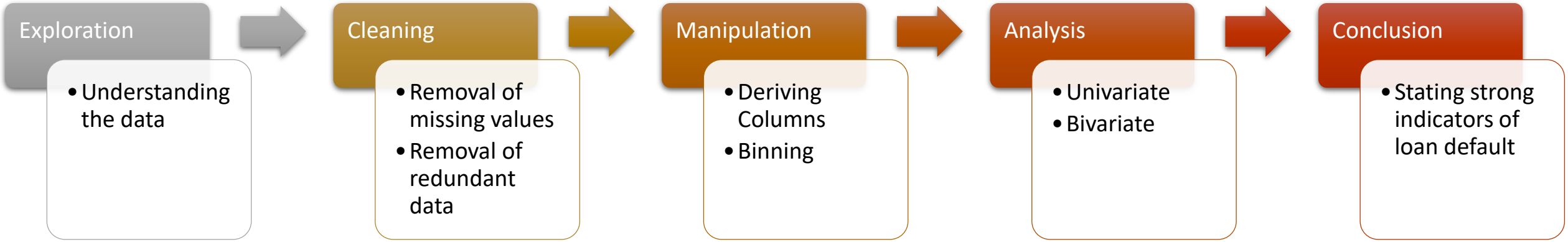
Background

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.

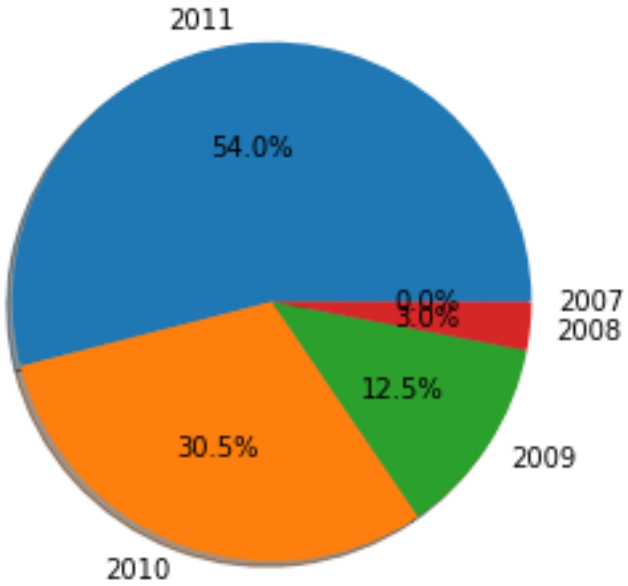
Objective

We are trying to summarise risk factors for investors that cause loan defaults with exploratory data analysis.

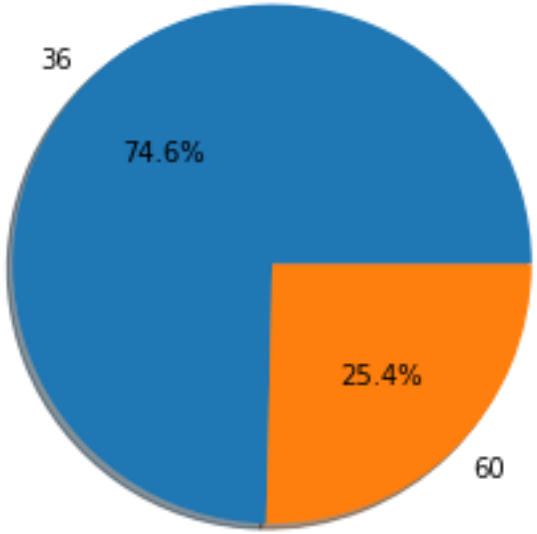
APPROACH


Exploration
• Understanding the data
Cleaning
• Removal of missing values
• Removal of redundant data
Manipulation
• Deriving Columns
• Binning
Analysis
• Univariate
• Bivariate
Conclusion
• Stating strong indicators of loan default

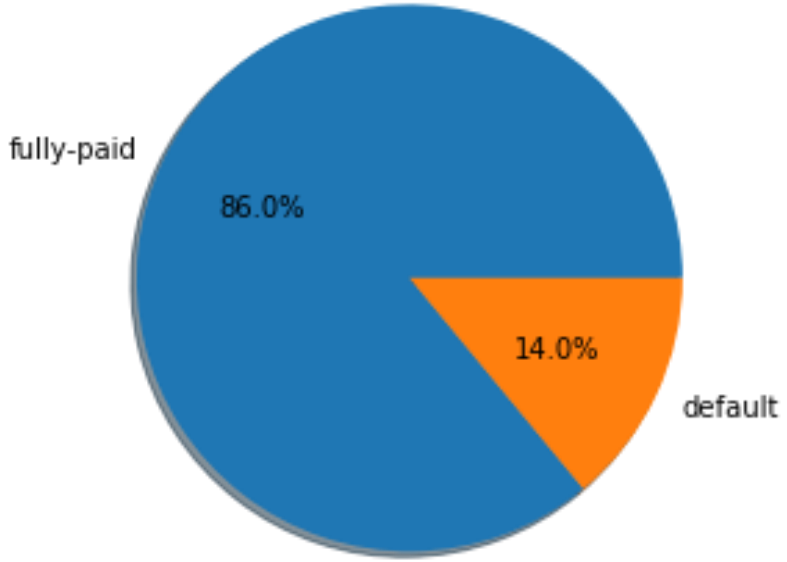
UNDERSTANDING DATA



- 2011 has highest number of loans counting up to 16000 and defaults counting up to 3800

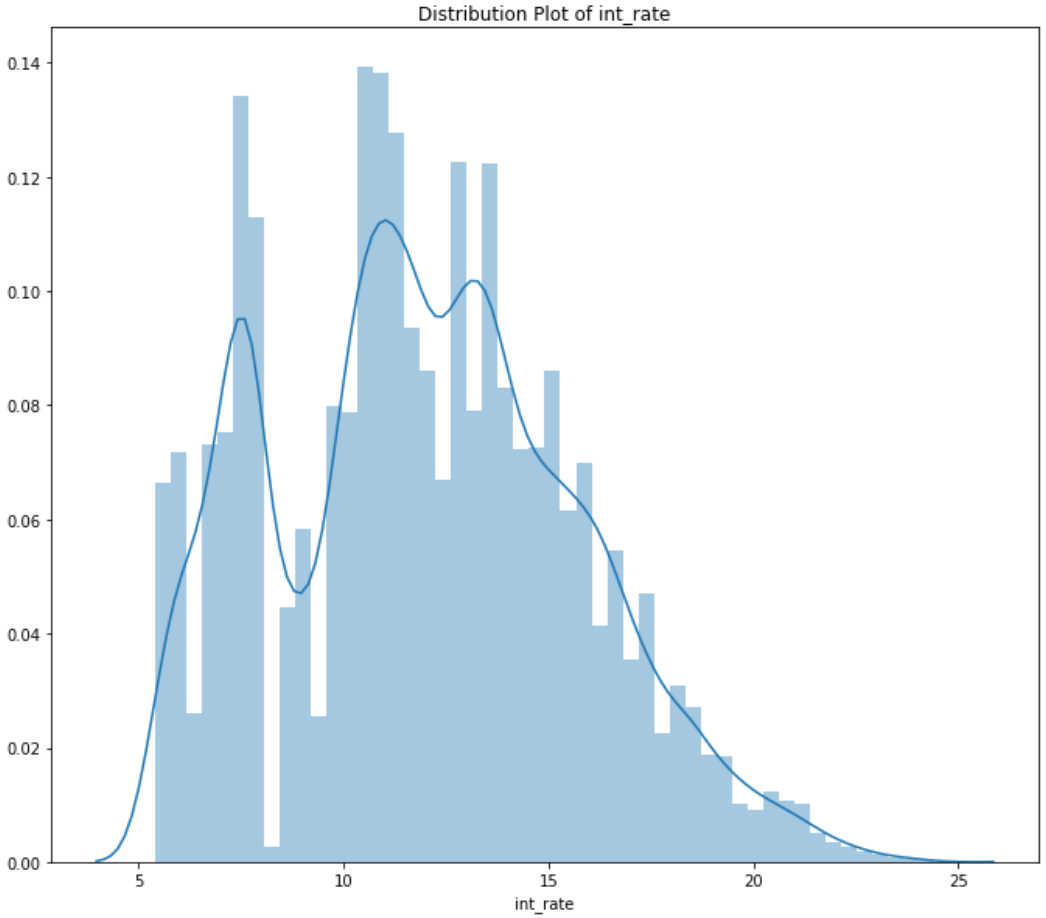


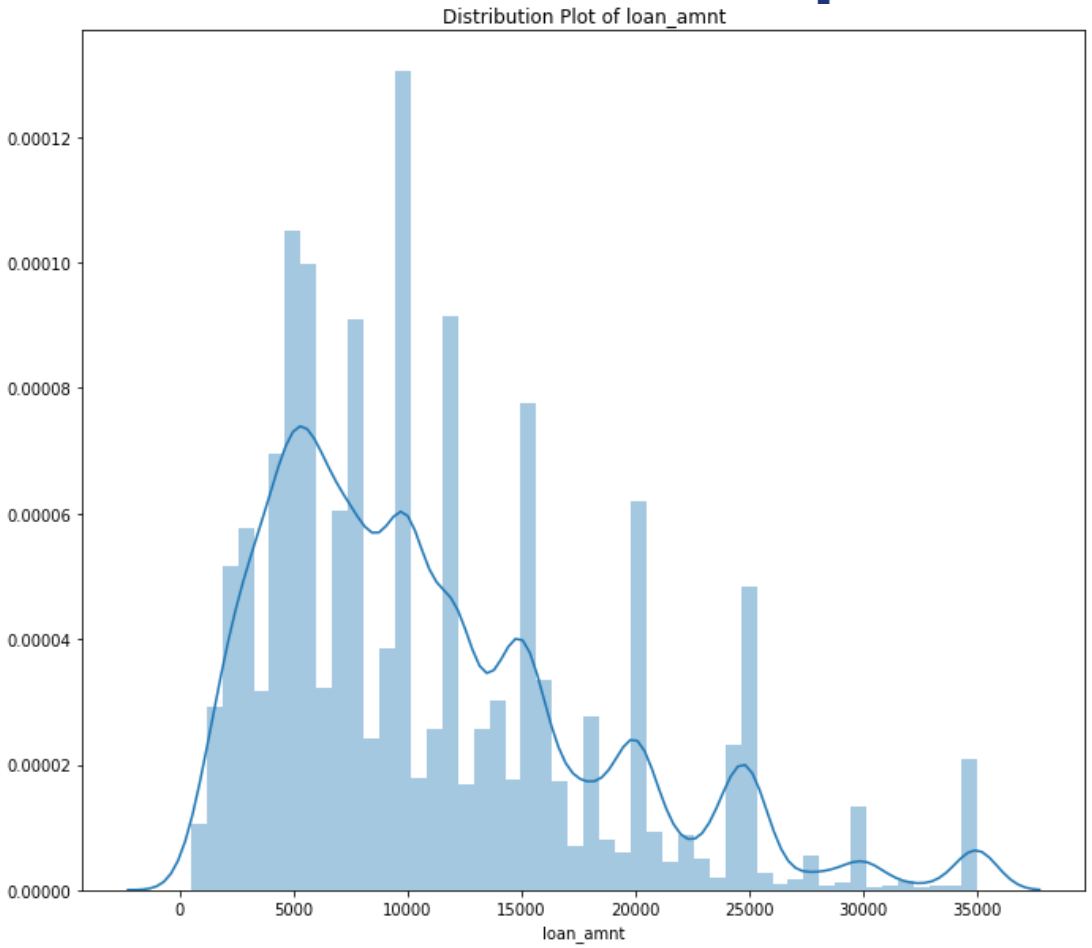
- 75 % of the loans given were for the term of 36 months counting up to 26136



- 14 % of the loans were defaulted counting up to 4909

OBSERVATIONS OF DATA DISTRIBUTION

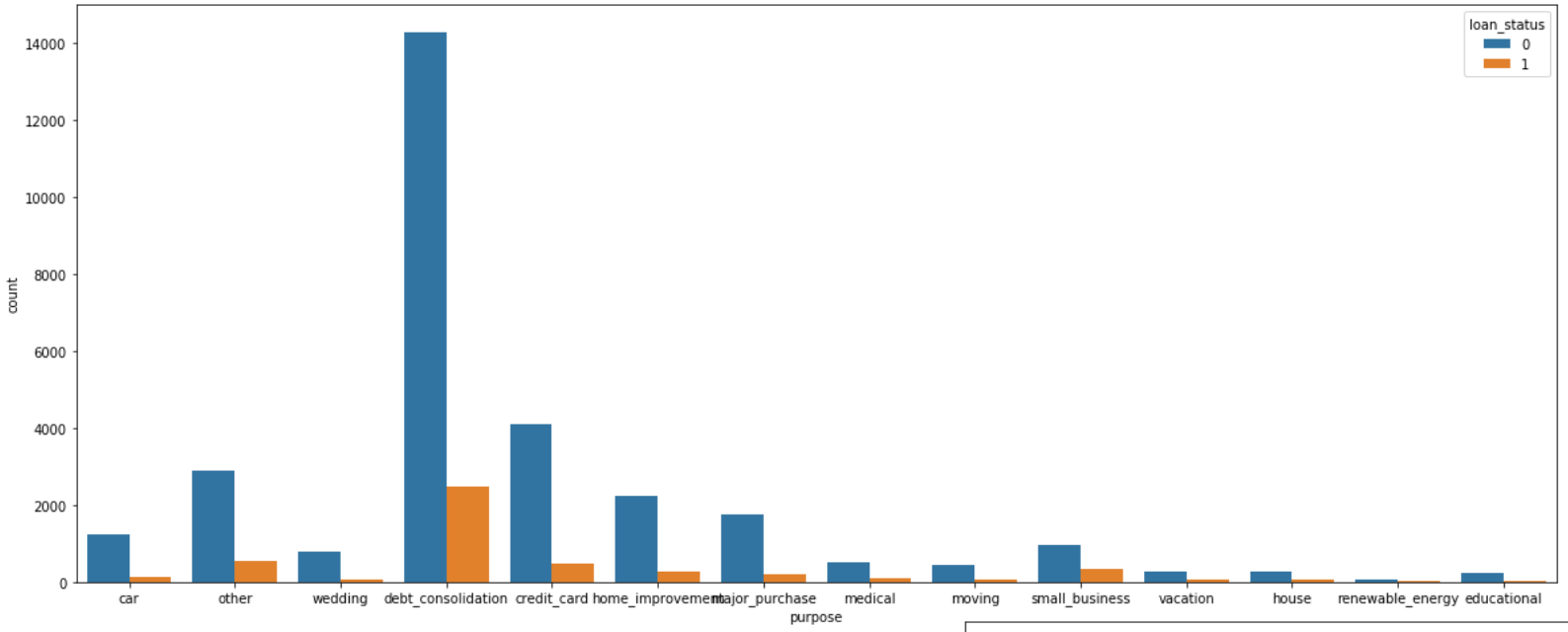




- Most interest rate are distributed between 11% - 14%
- Most loan amount and funded amount by investors are distributed between 5000 - 12000

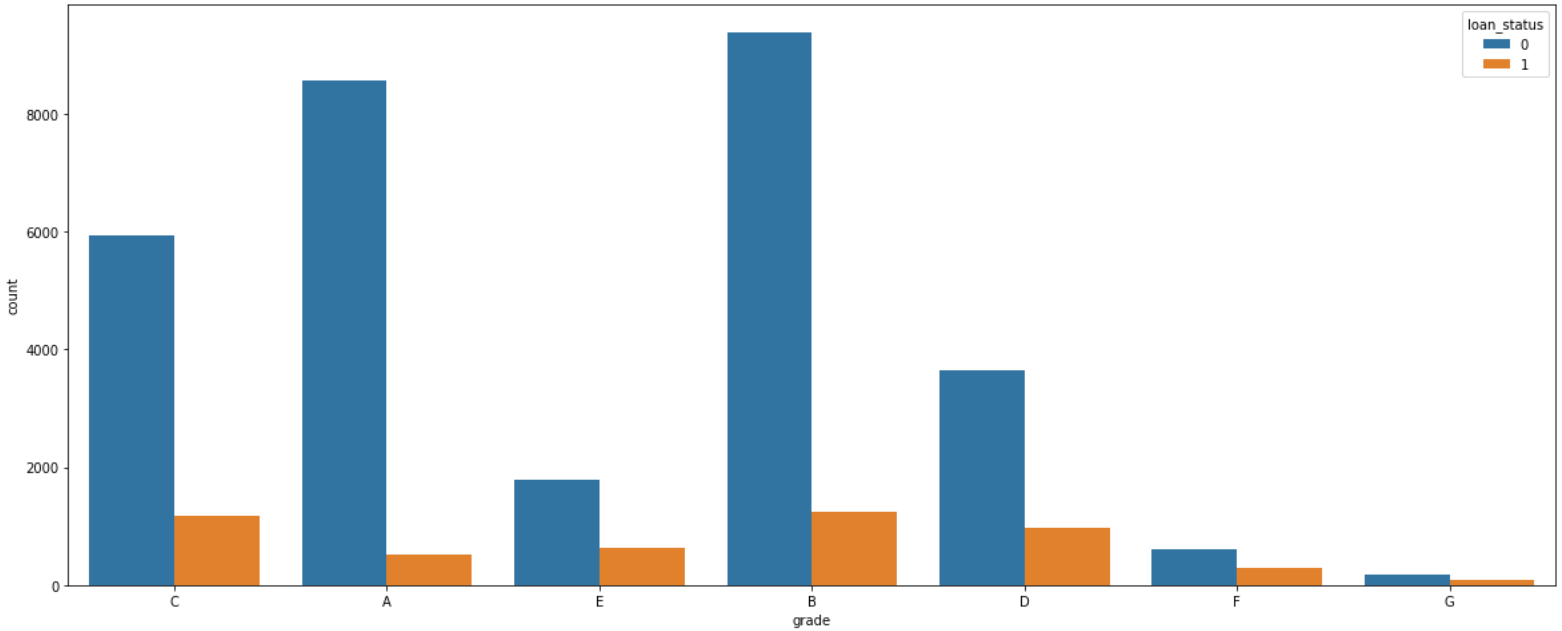
- Most annual income is distributed between 50000 – 90000
- loan amount - income ratio is high at 10 - 20

OBSERVATIONS OF DATA DISTRIBUTION

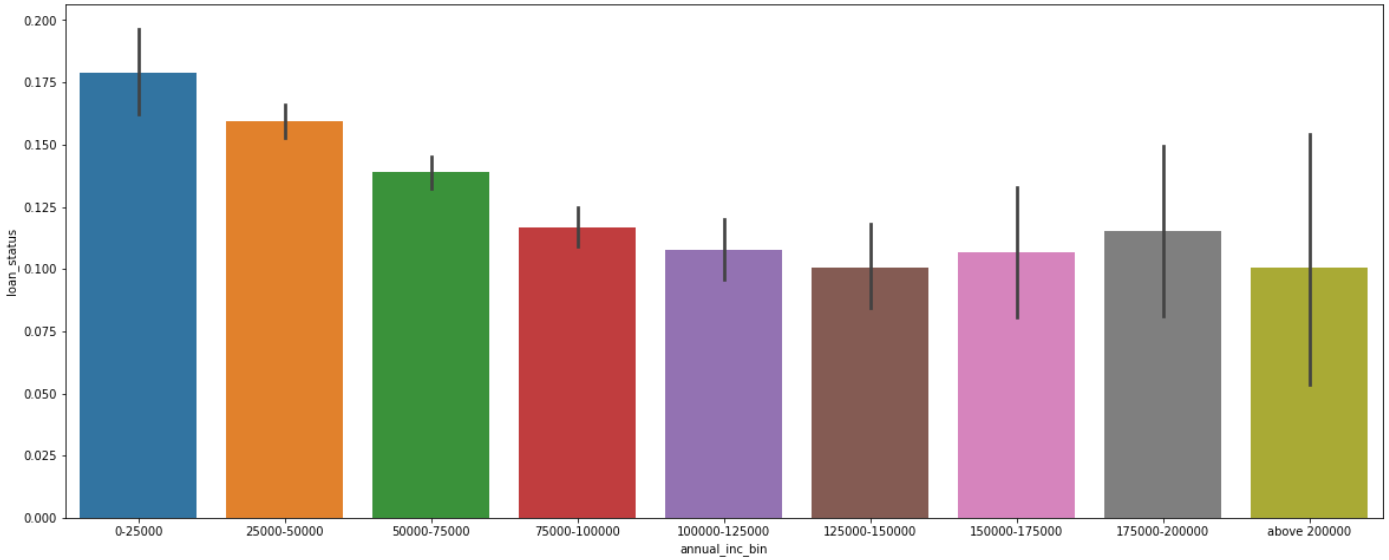


Most of the loans were taken with purpose debt_consolidation

- Most of the loans are graded B or A.
- There are more than 10000 loans under B grade



UNIVARIATE ANALYSIS



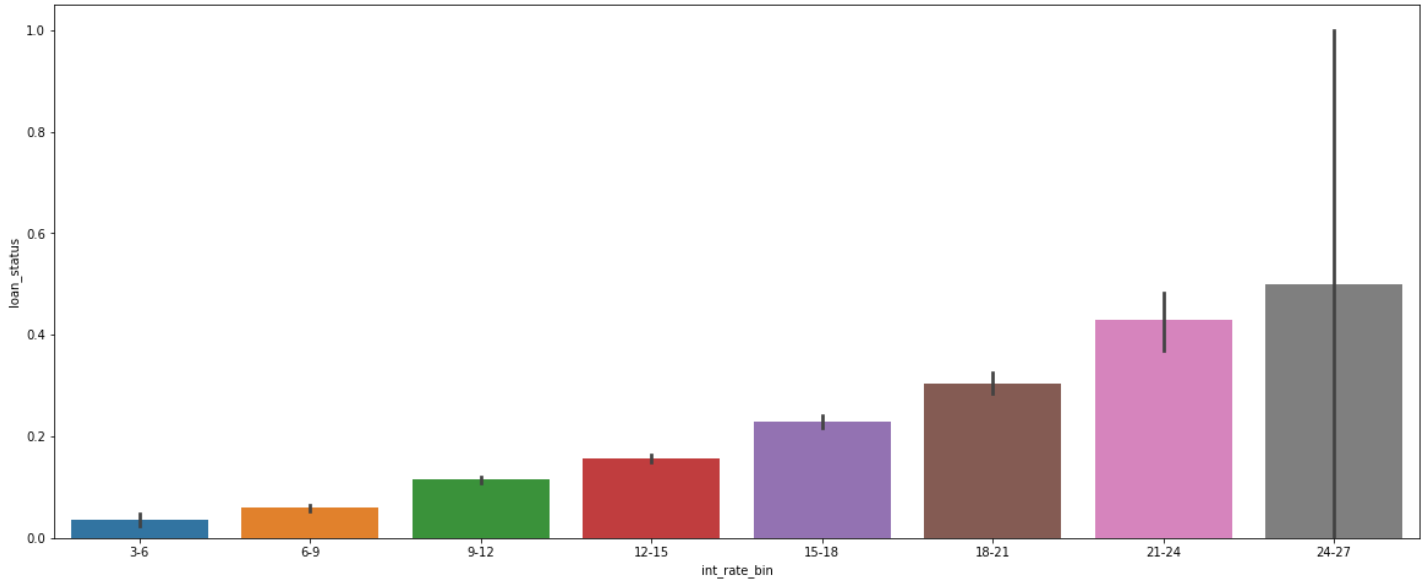
ANNUAL INCOME

Observation:
People with annual income less than 50000 tend to have high default rate.

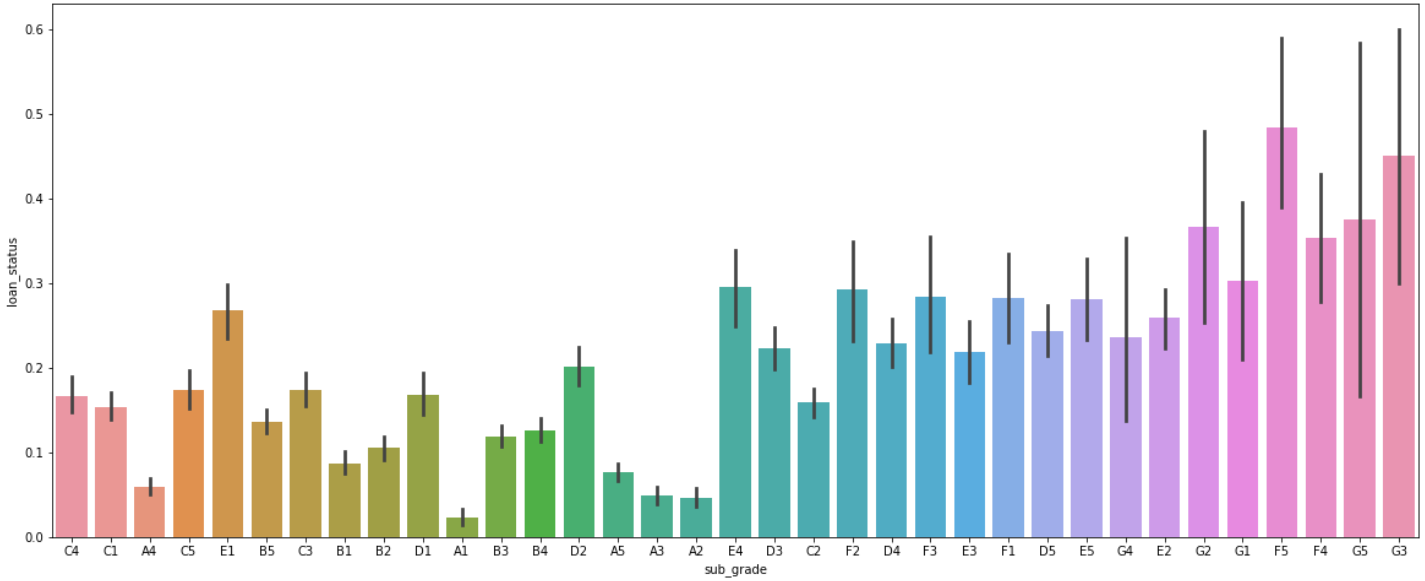
INTEREST RATE

Observation:

- The loan defaulters are increasing with interest rate
- There's a high chance of loan default if rate of interest is greater than 20



UNIVARIATE ANALYSIS



GRADE AND SUB-GRADE

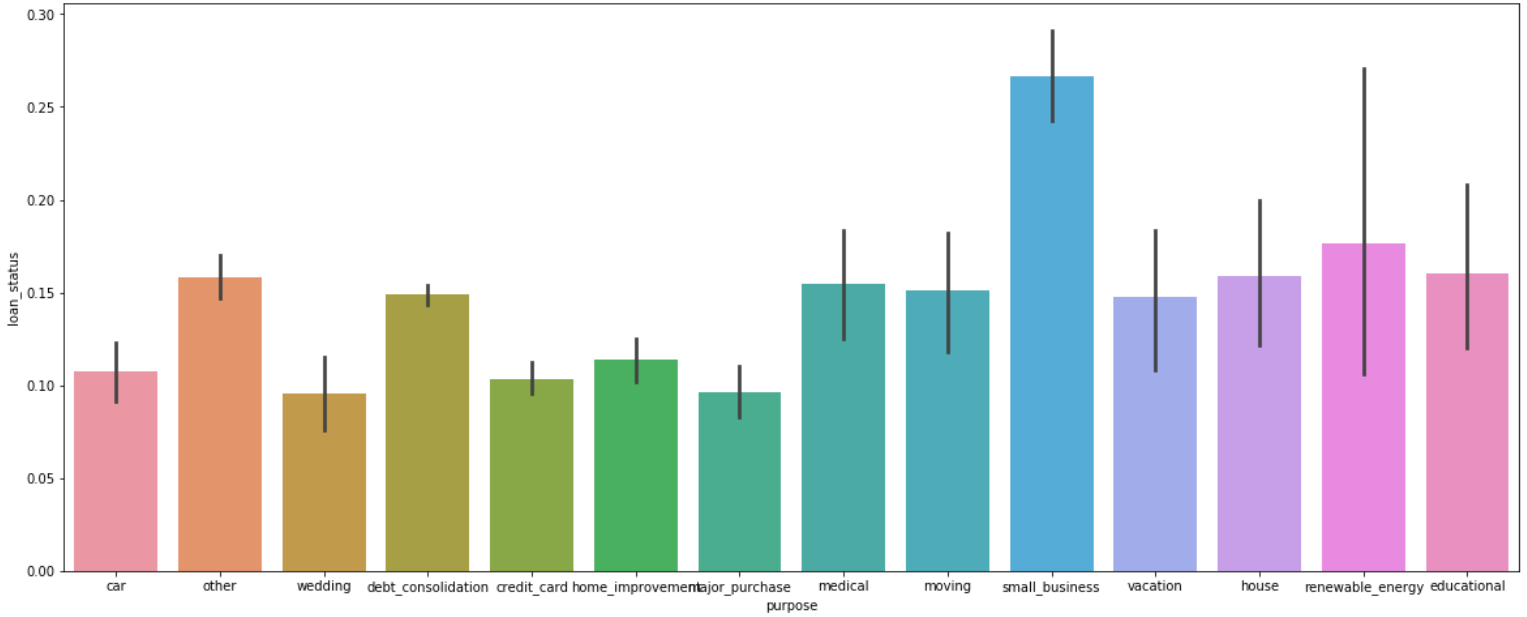
Observation:

- As expected A grade has lowest default rate and G, F have high default rate.
- F5 sub-grade tend to have highest default .

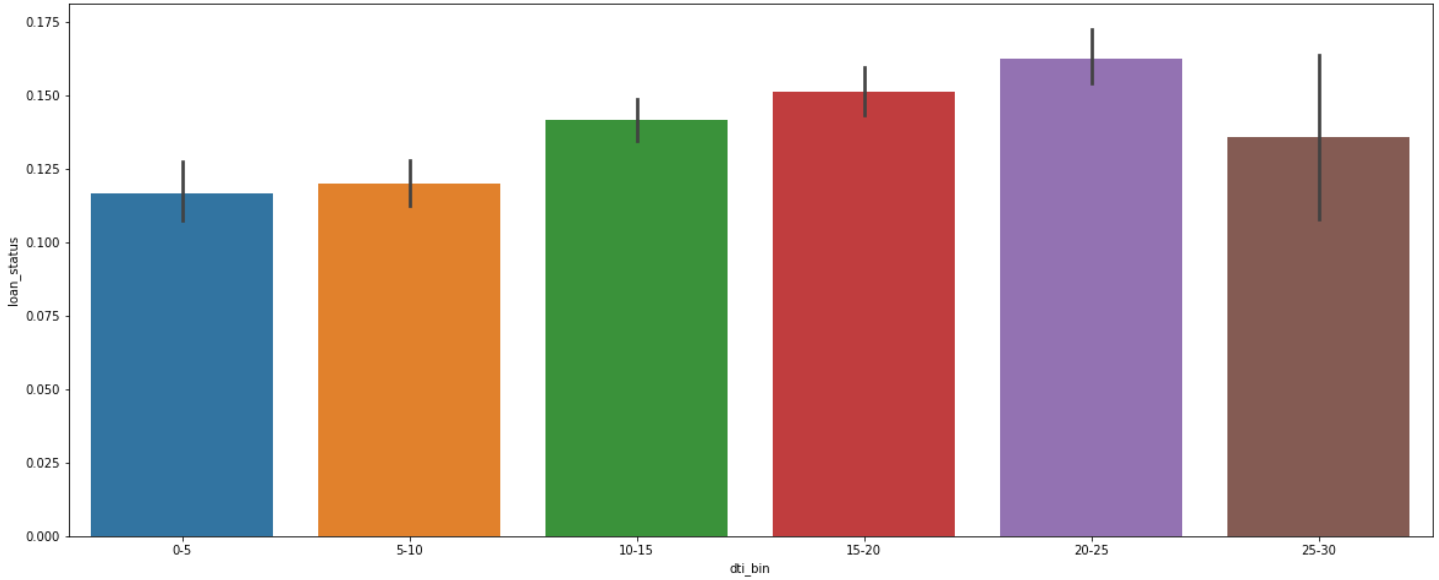
LOAN PURPOSE

Observation:

Loan taken with purpose ‘small business’ has highest loan default followed by ‘renewable energy ‘.



UNIVARIATE ANALYSIS

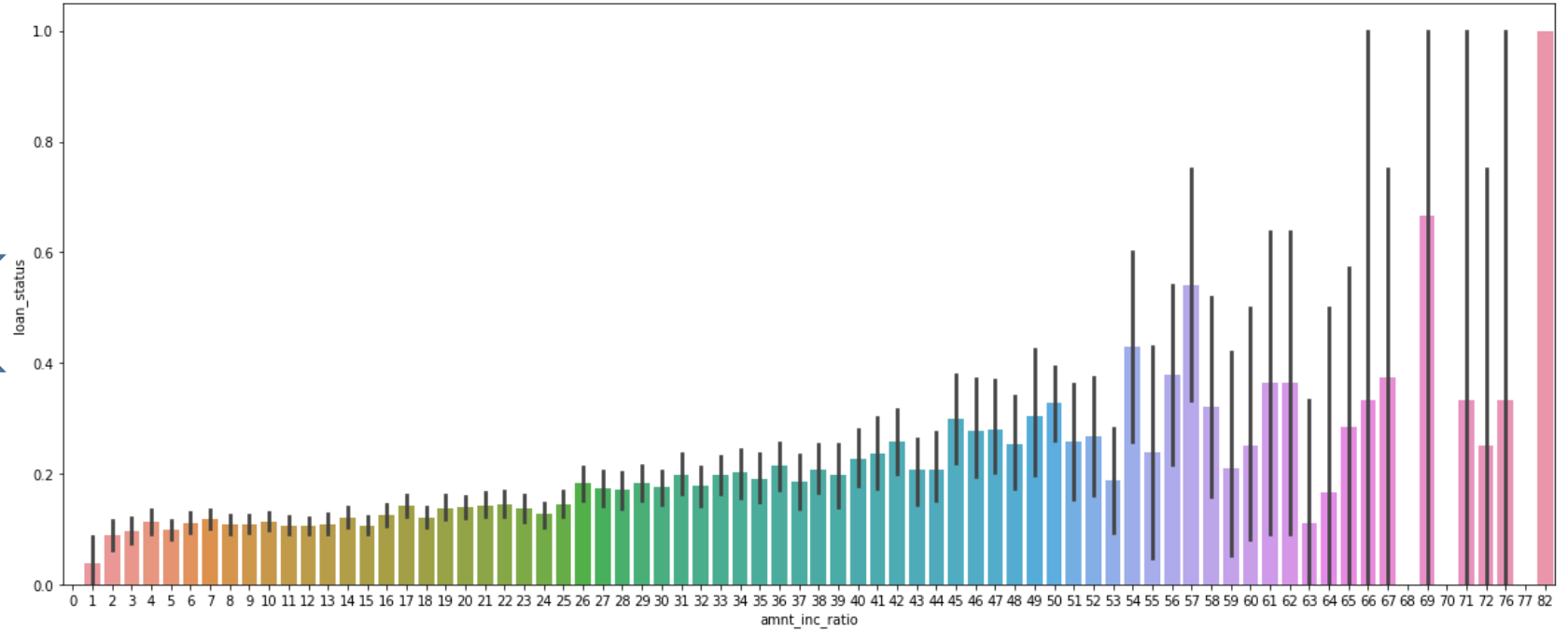


DTI

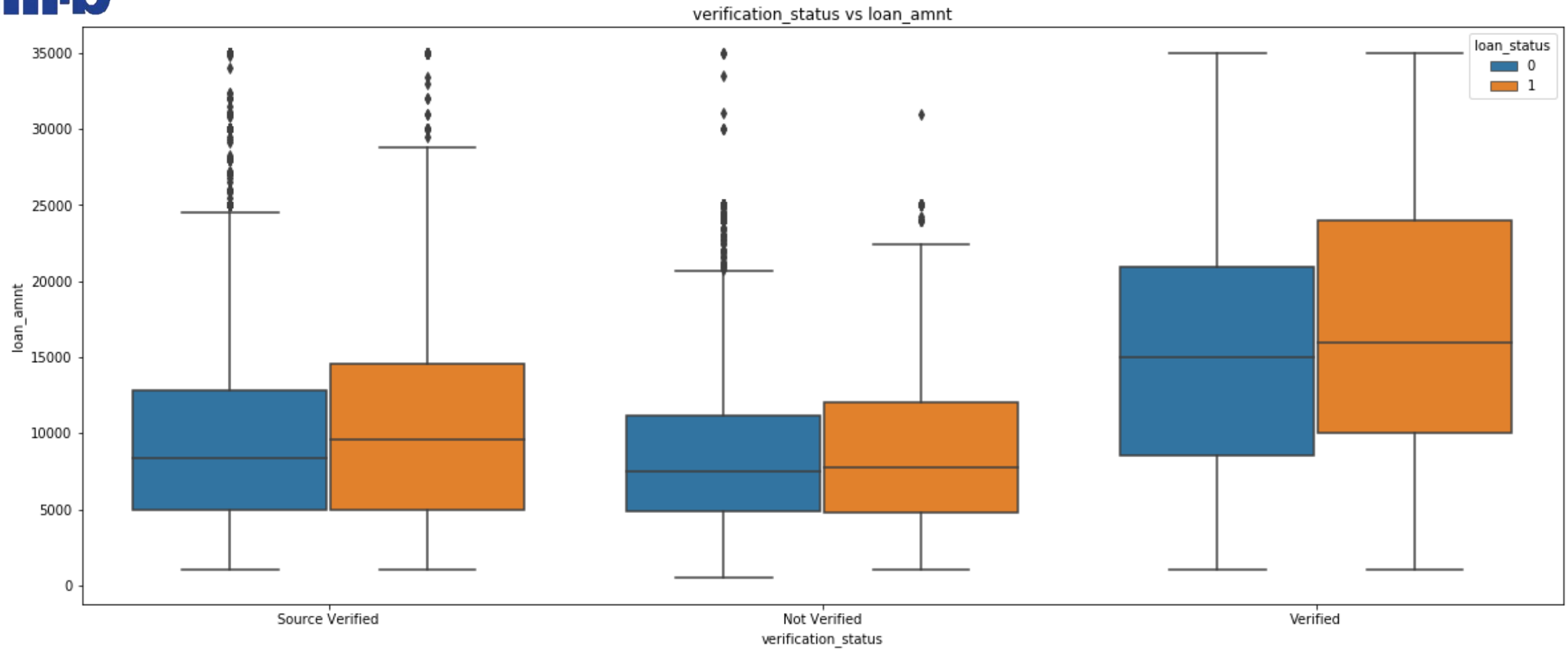
Observation:
Loan defaults are high if DTI is greater than 20

RATIO OF LOAN AMOUNT AND ANNUAL INCOME

Observation:
Loan defaults are high if ratio is greater than 25



BIVARIATE ANALYSIS

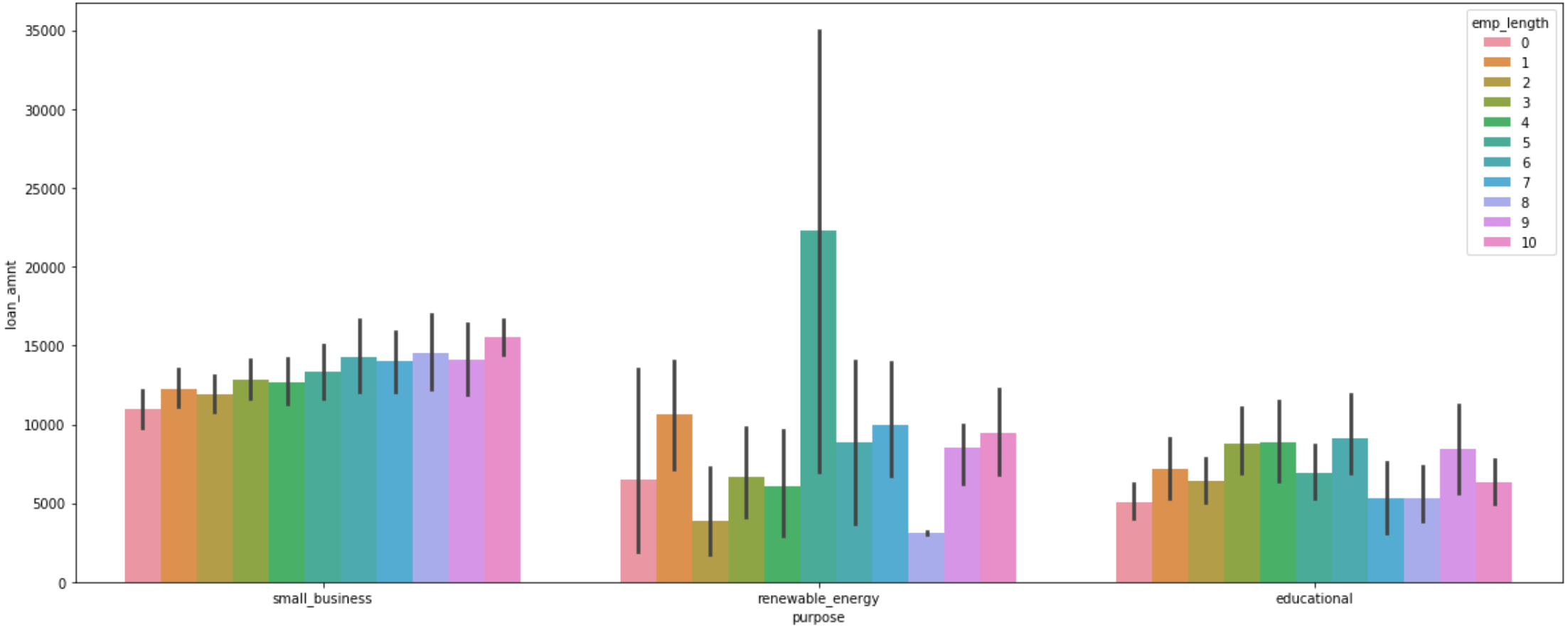


VERIFICATION STATUS VS LOAN AMOUNT SEGEMNTED BY LOAN STATUS

Observation:

Surprisingly verified / source verified loans tend to default more because these are the loans with high loan amounts, loans with high amounts are definitely verified and high loan amounts have high default rate

BIVARIATE ANALYSIS



PURPOSE WITH HIGH DEFAULT RATE VS LOAN AMOUNT SEGEMNTED BY EXPERIENCE

Observation:

We already know from previous univariate analysis that loan applications with purpose – ‘small business’ and ‘renewable_energy’ have high loan default . loan applicants with experience > 5 and with purpose ‘small business’ or ‘renewable_energy’ take huge amount of loan and they are highly prone to loan default

SUMMARY AND RECOMMENDATION

- **Driving variables behind loan default : Grade, Purpose, Interest rate, Annual income, home ownership**

Applications that can be considered With low loan default risk

- If loan amount – annual income ratio should be less than 25
- If grade is 'A' with any 'A' sub grade
- If Interest rate is < 6% and for 36 months term

Applications with high risk of loan default

- If Purpose mentioned is 'small business' Or If employee length > 5 and purpose mentioned is 'renewable_energy'
- If sub grade is F5 , G3
- If interest rate is > 20 %
- If revol_util > 90%
- If DTI is high
- Huge amounts of loans should not be given to applicants with annual income less than 25,000.