LENDING CLUB CASE STUDY

Group Members:

- Sachin Gupta
- Saibek Varghese Panicker

Problem Statement

A consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company must decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company

Using Exploratory to understand how consumer attributes and loan attributes influence the tendency of default.

Problem Solving Approach:

Data understanding

Data cleaning :

- Removal of unnecessary columns
- Handle missing data
- Reframing features acc to its datatypes.
- Filter unwanted outliers.

Data manipulation:

- Creation of derived cols
- Conversion of numerical cols to buckets for analysis

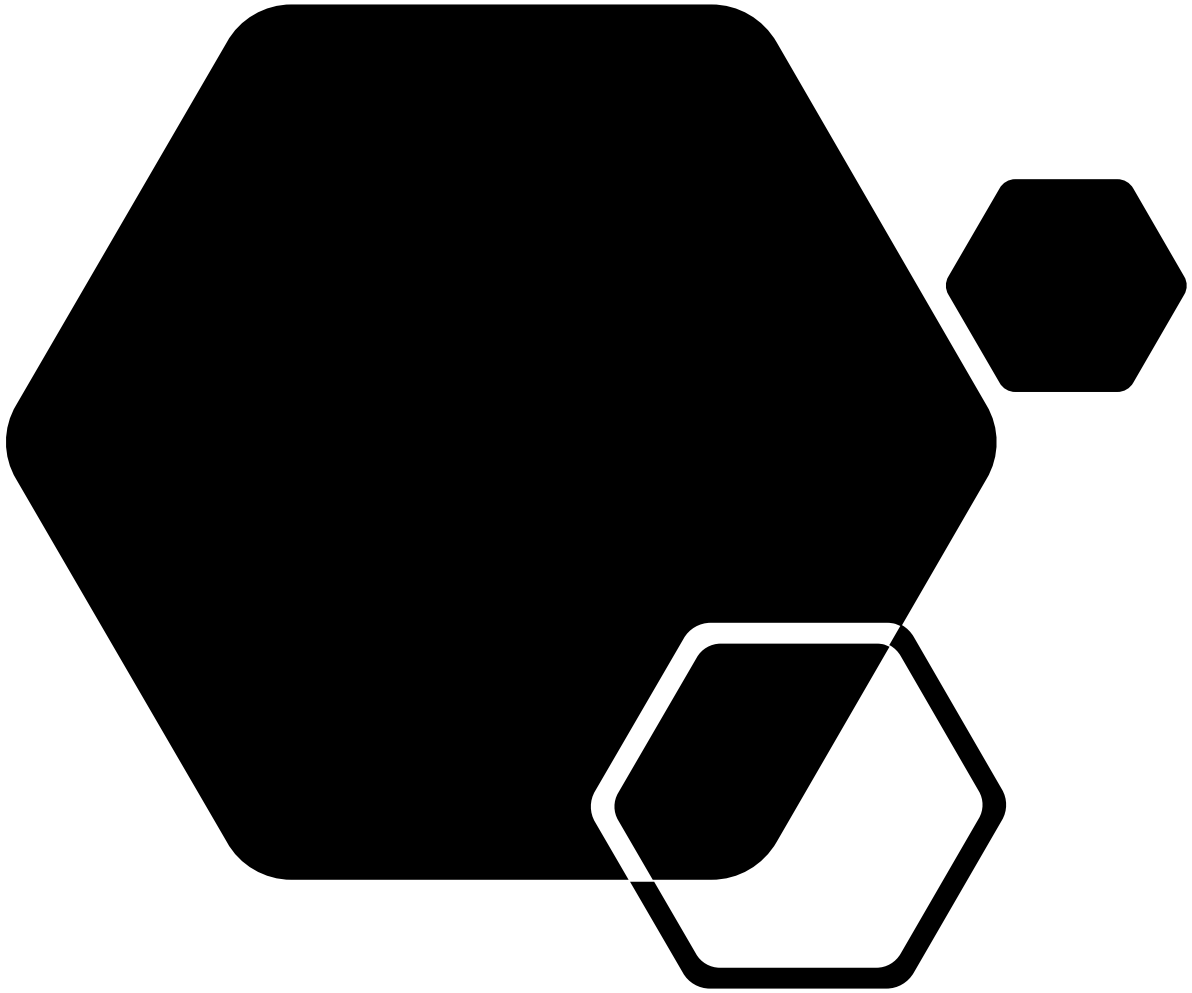
Univariate analysis

Bivariate analysis

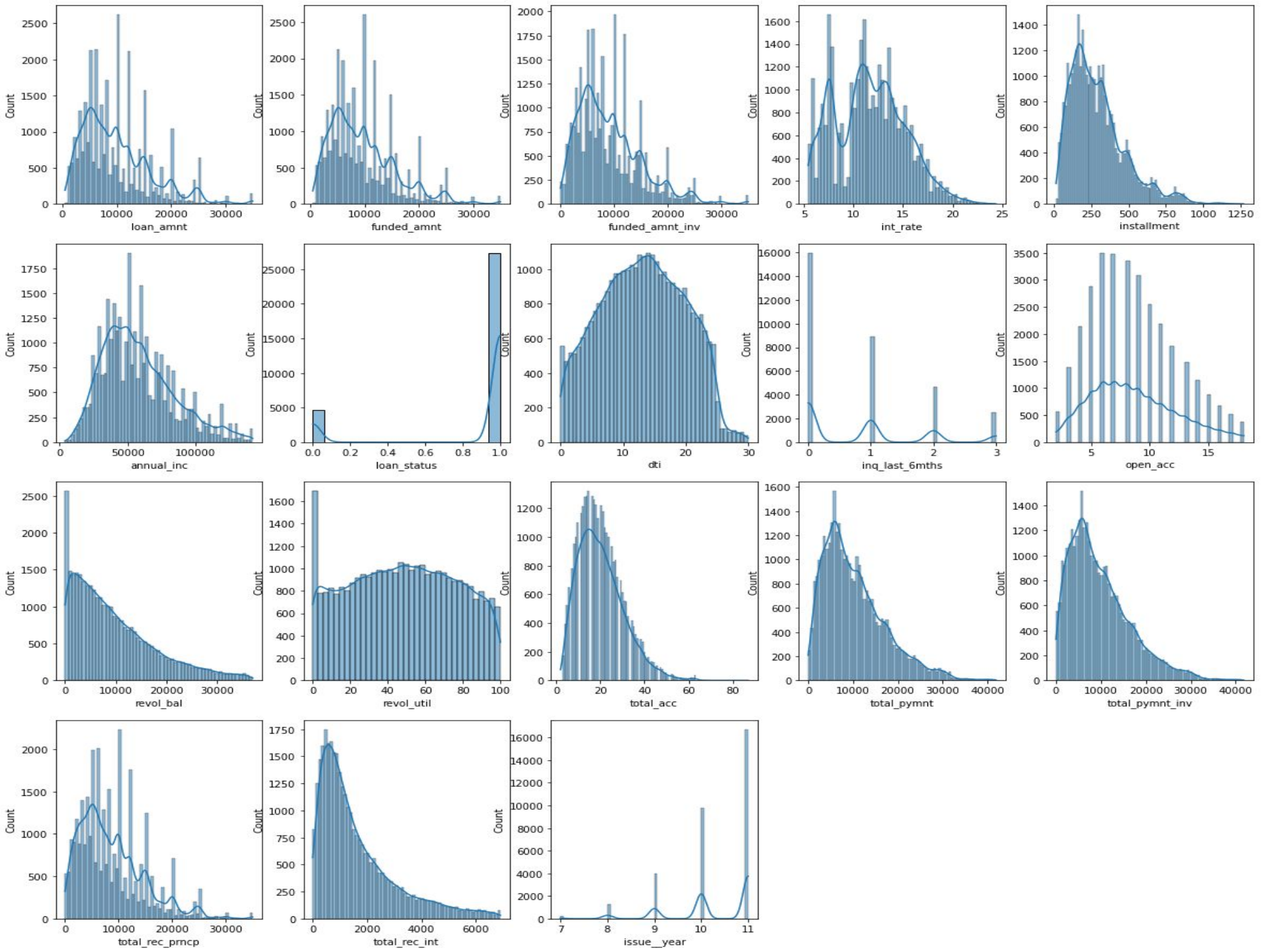
Conclusion

UNIVARIATE ANALYSIS

- **NUMERICAL UNIVARIATE ANALYSIS**
- **CATEGORICAL UNIVARIATE ANALYSIS**



UNIVARIATE ANALYSIS ON NUMERICAL FEATURES



Numerical Univariate Observation

1. **loan_amnt** : Most of the loan_amnt has been taken upto 10,000

2. From the observed plot we can say that **funded_amnt, funded_amnt_inv** is same as **loan_amnt**

3. **int_rate** : Most of the interest rate for the loan given lies between 10-15%

4. **installment** : Most of the installment lies in the range of 250 (Somewhat around Median)

5**. annual_inc** : looks like a normal distribution curve (Majority of the income lies between 35k - 60k)

6**. dti** : It's a purely Normal Distribution Graph (Maximum Data lies in it's median/mean)

7. **inq_last_6_months** : Most people not inquired in the last 6 months (approx. : 50%)

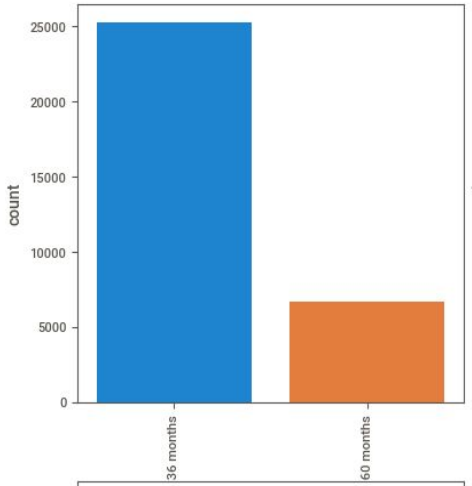
8. **open_acc** : Most of the people open credit lines lies between 5-10

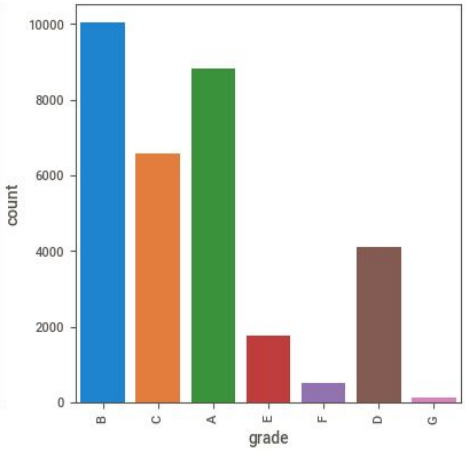
9. **total_acc** : Most of the people having total credit lines lies between 15-25

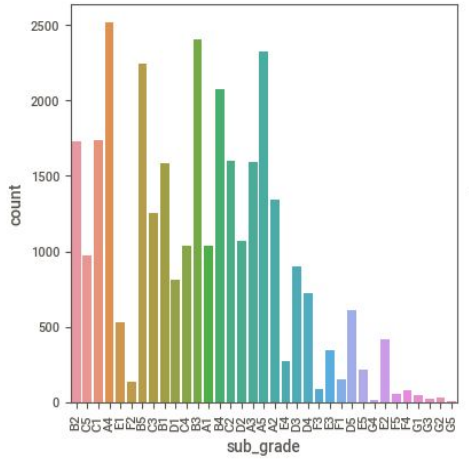
10. **issue_year** : As the years goes on the no. of people taking loans exponentially increases.

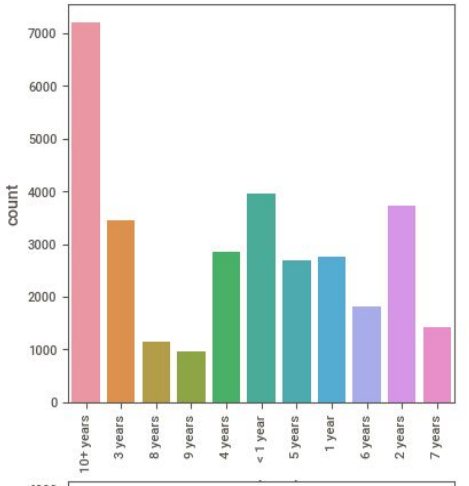
11. **loan_status**: 15% loan defaulters are present

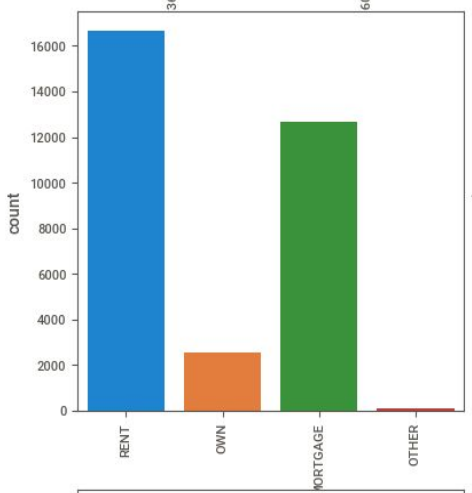
UNIVARIATE ANALYSIS ON CATEGORICAL FEATURES

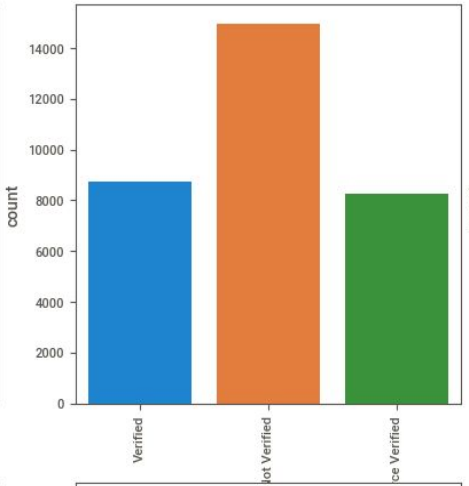


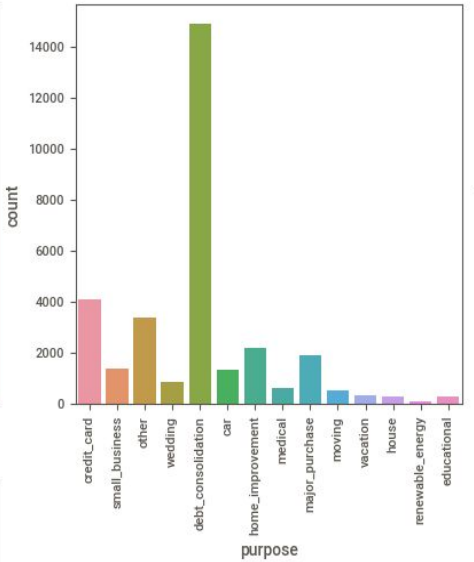


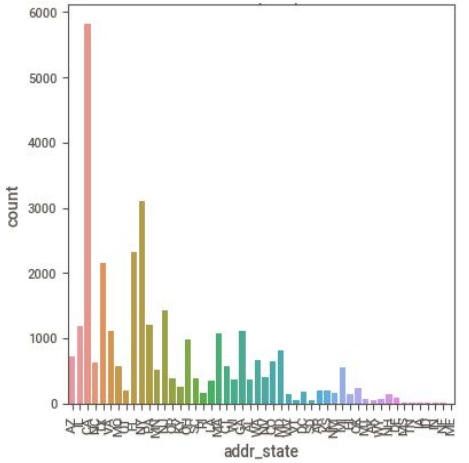


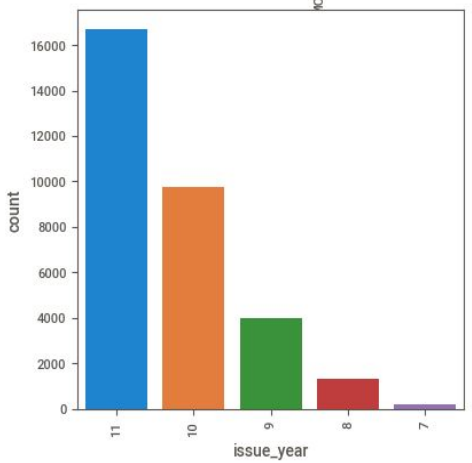


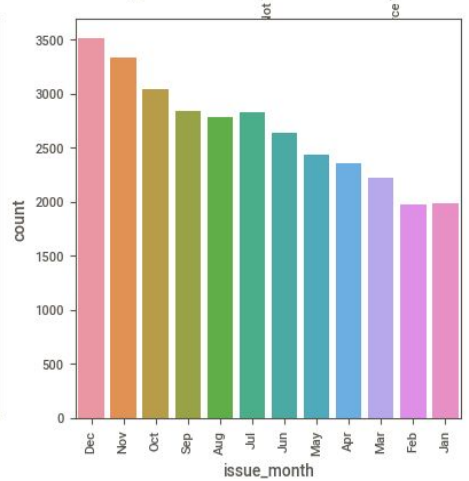












Categorical Univariate Observation

1**. term** : 36 Months is the most frequent time period.

2. **grade** : A and B are the most frequently occurring loan grade.

3. **emp_length** : 10+ Years is the most frequent one.

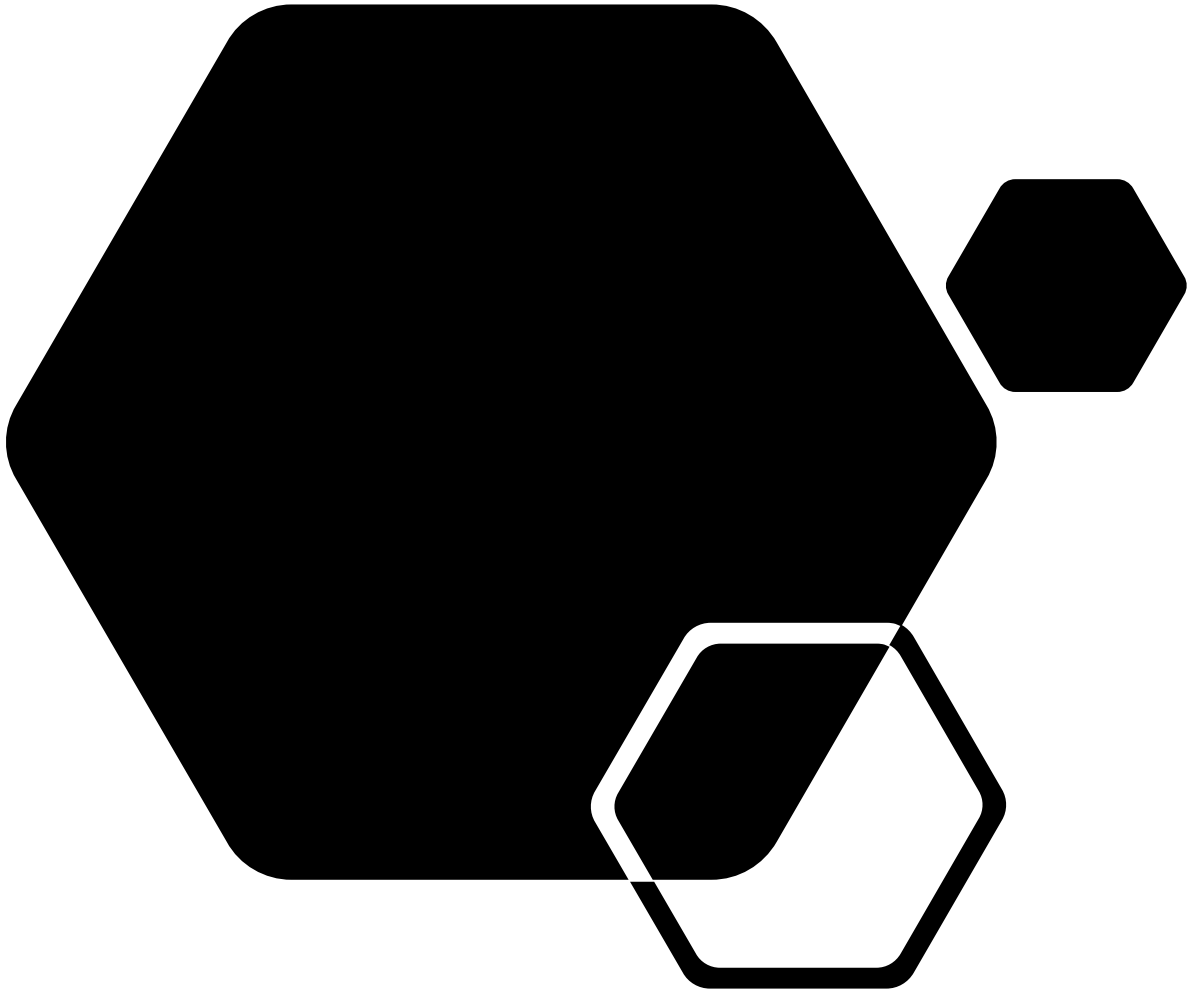
4. **home_ownership** : Max people Took loan who belongs to **RENT** Category.

5. **verification_status :** Maximum people are **not_verified** status.

6. **purpose** : Majority of the reason of took the loan is **debt_consolidation**.

7**.issue_month** : Observed that at the end of the year most people takes loan bcoz they will give some offer etc. to the customer etc.

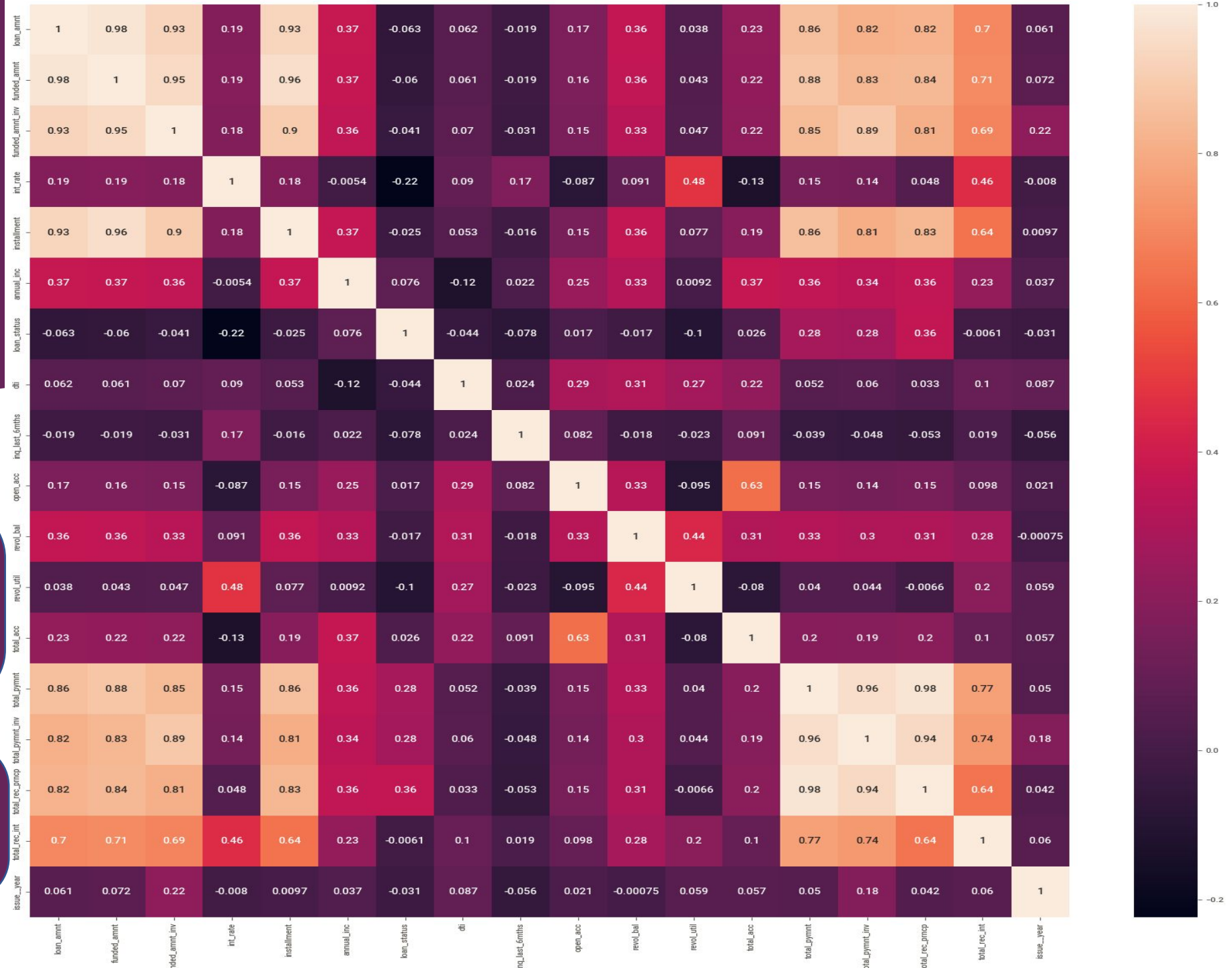
BIVARIATE ANALYSIS



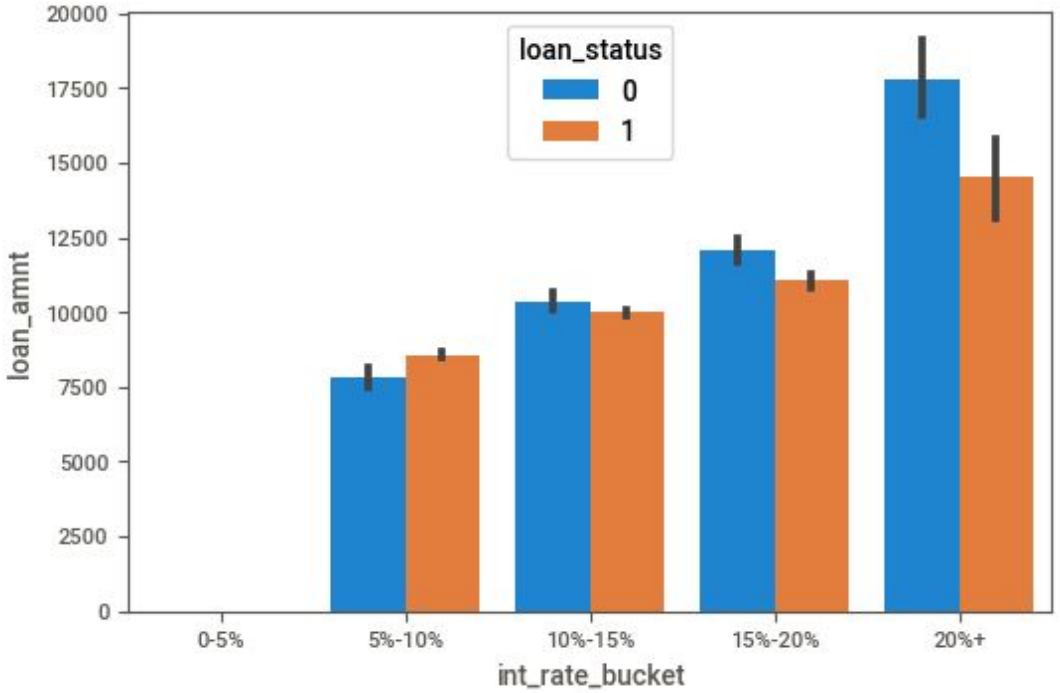
CORRELATION PLOT

Loan amount is highly correlated with funded_amnt, funded_amnt_inv and installment

total_pymnt is highly correlated with total_pymnt_inv and total_rec_prncp

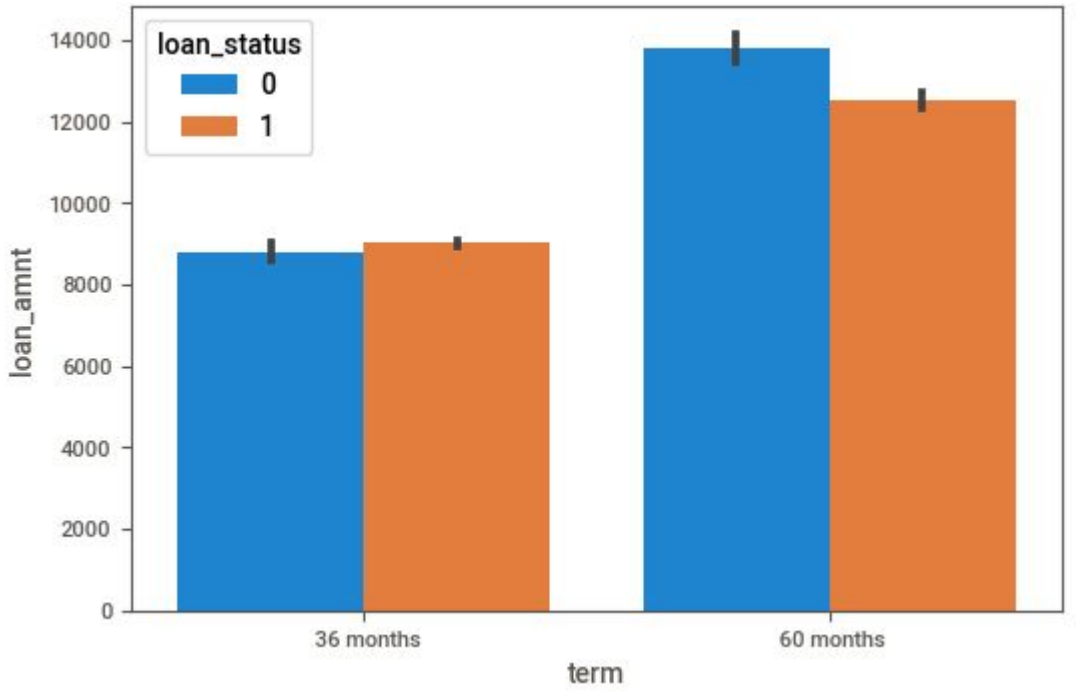


Loan Amount vs Interest rate bucket



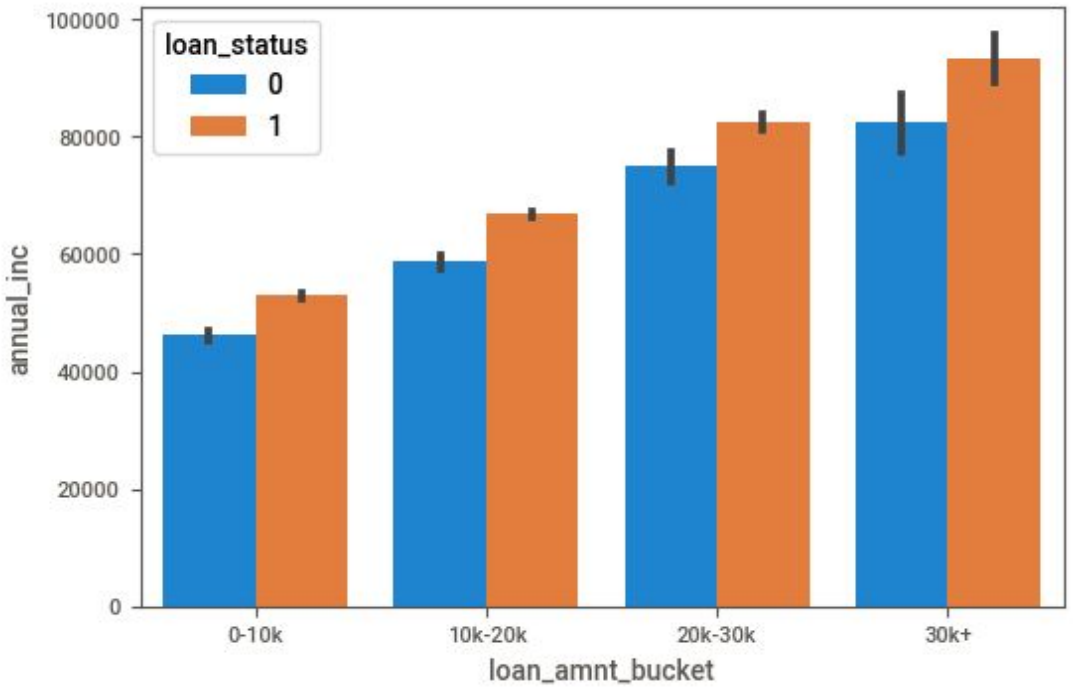
If interest rate and loan amount is higher then the defaulters are also higher.

Loan Amount vs Term

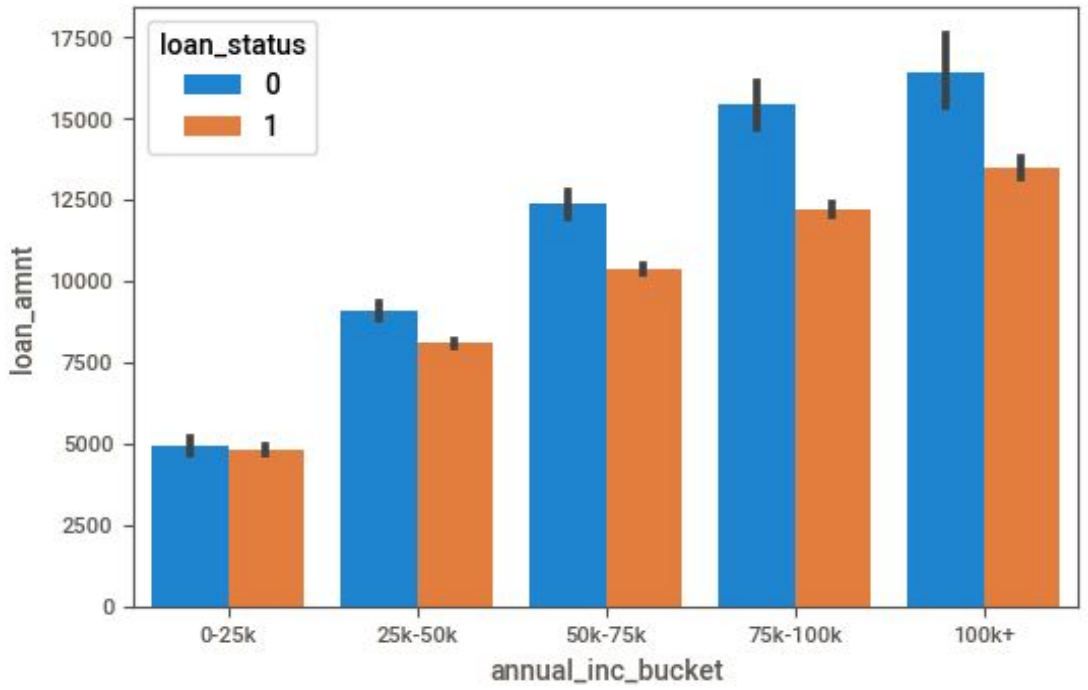


For 60 months term the loan amount taken is higher, hence the defaulter is higher.

Annual Income vs Loan amount bucket

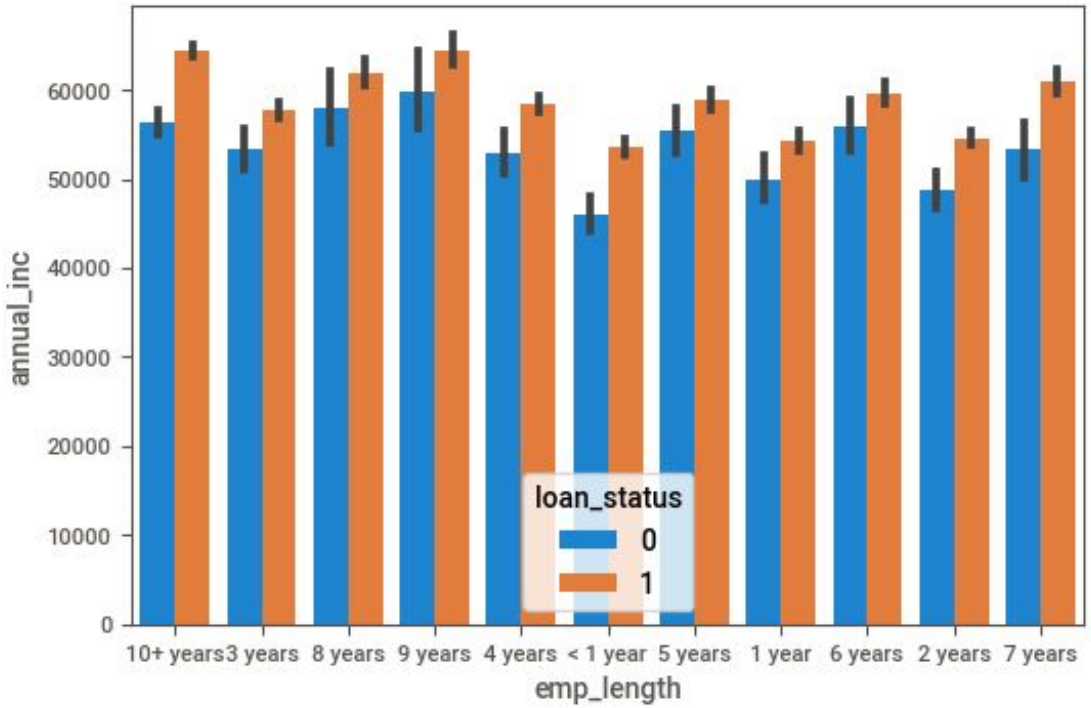


Loan Amount vs Annual income bucket

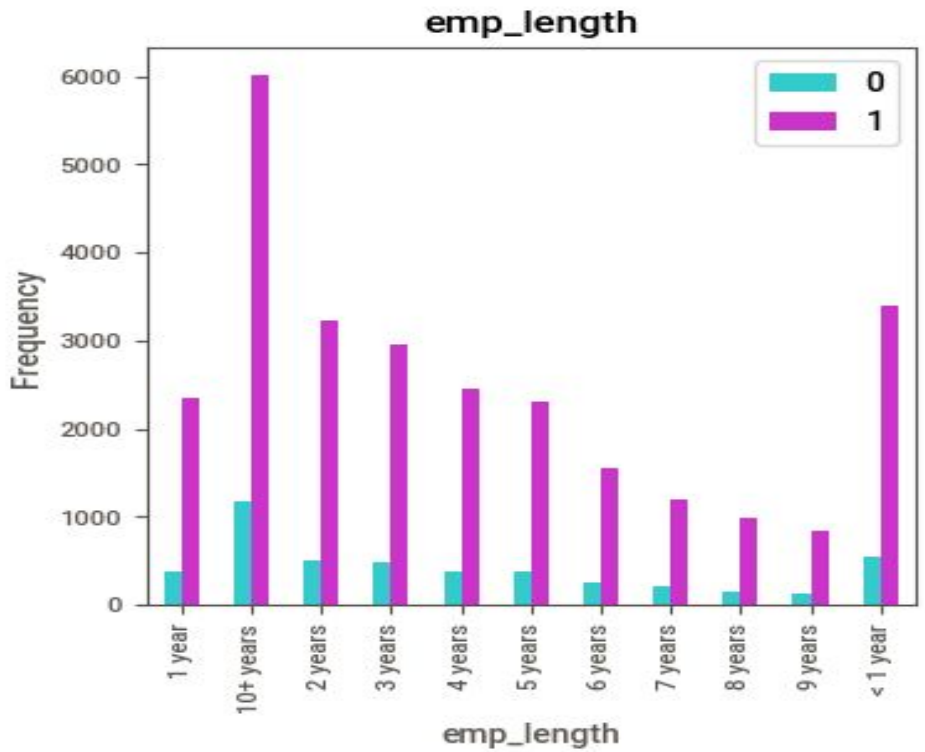


If the annual income is higher then people tend to take bigger loans where the defaulters considerably higher

Annual Income vs Loan amount bucket

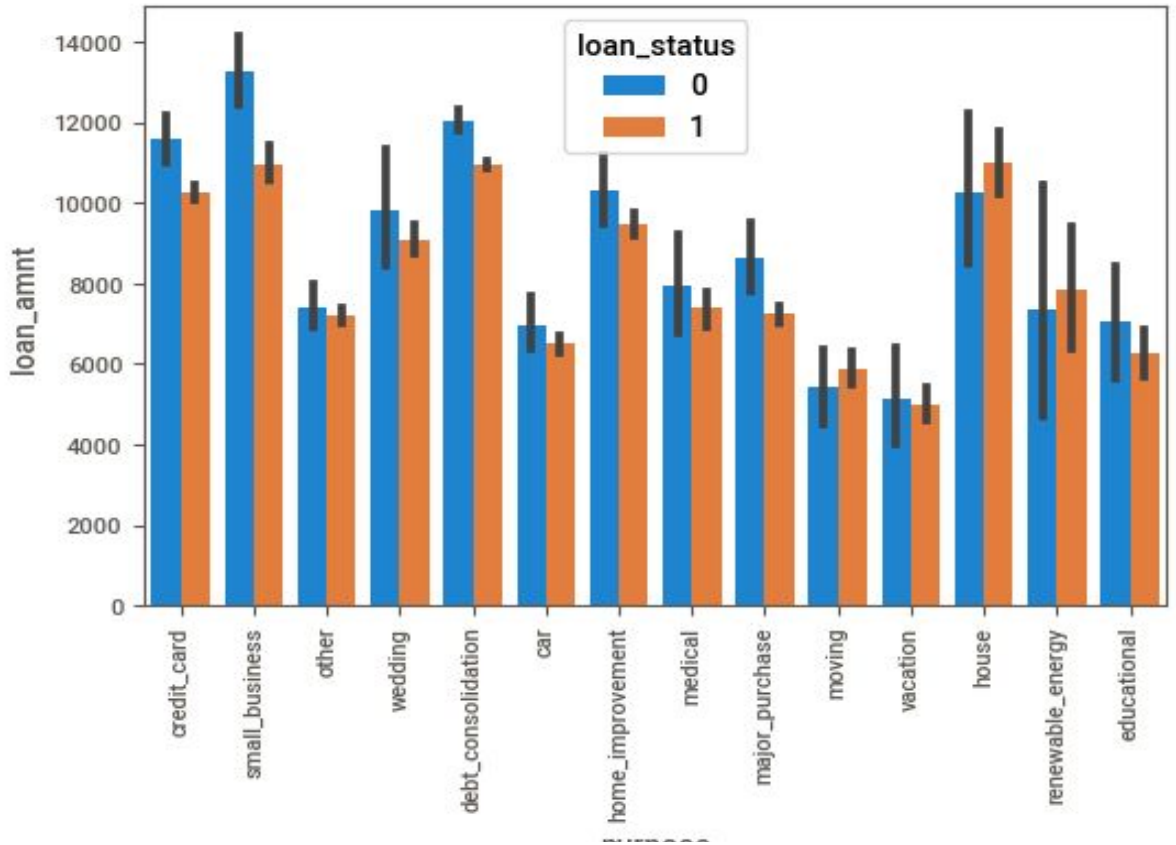


Emp length segmented analysis

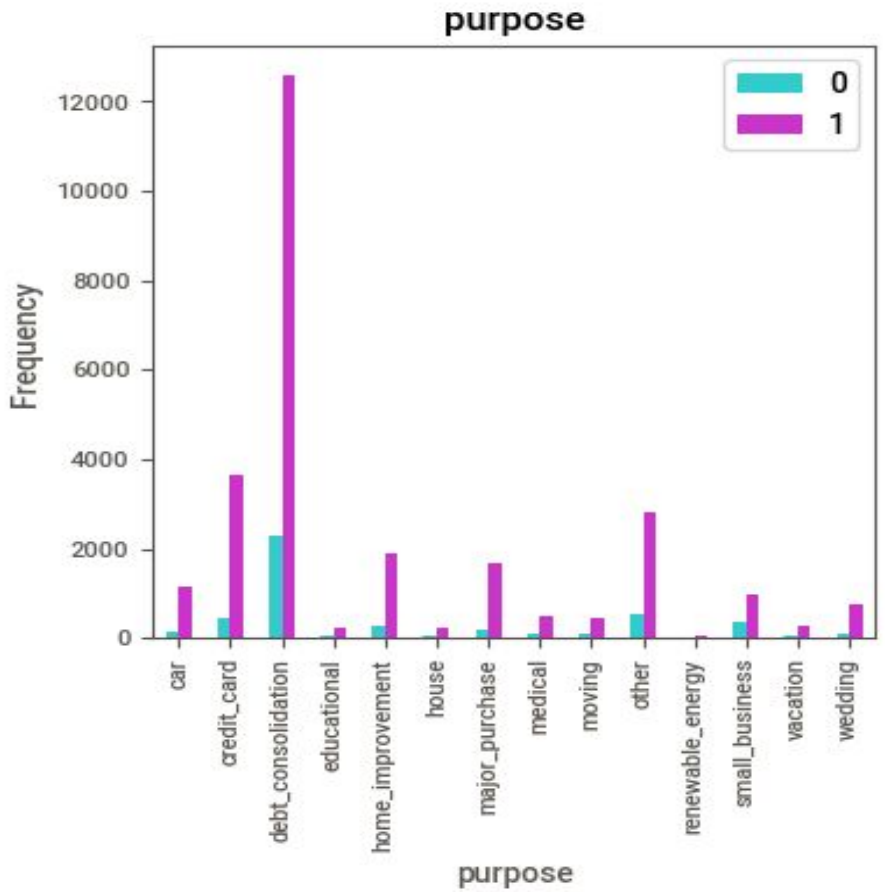


People with less than 1 year of exp and who takes a loan have a greater chance to be in a defaulter range.

Loan Amount vs Purpose

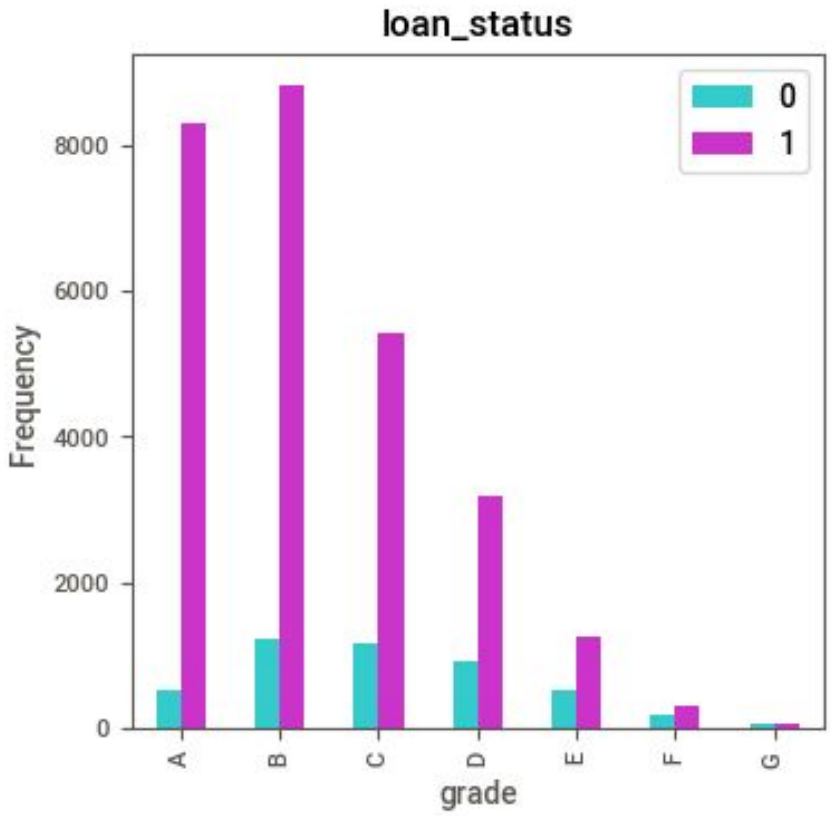


Purpose segmented analysis



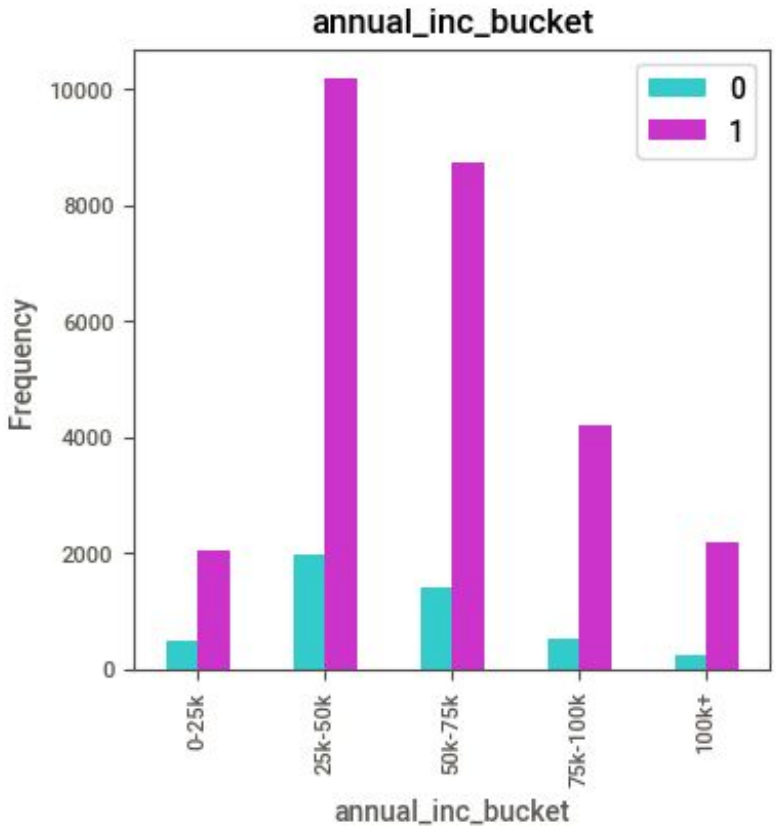
- **Max loan amount took for small_business which leads to more defaulters.**
- **People who took loan for debt_consolodation are higher rate of defaulters and the maximum people took loan from debt_consolidation as well.**

GRADE segmented



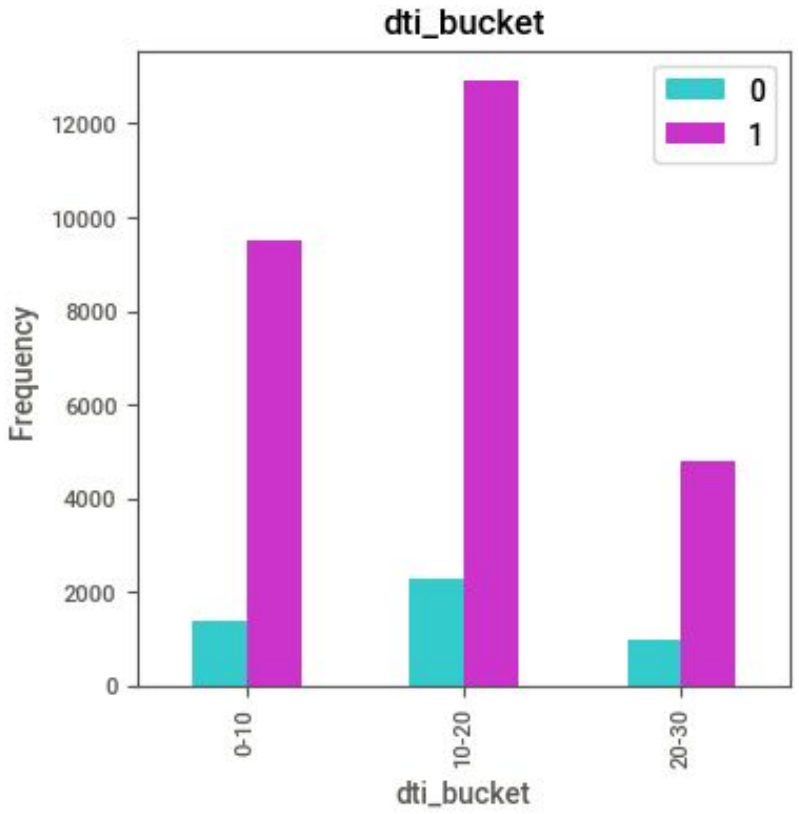
A and B grades have less defaulters as compared to others grade.

Annual income segmented



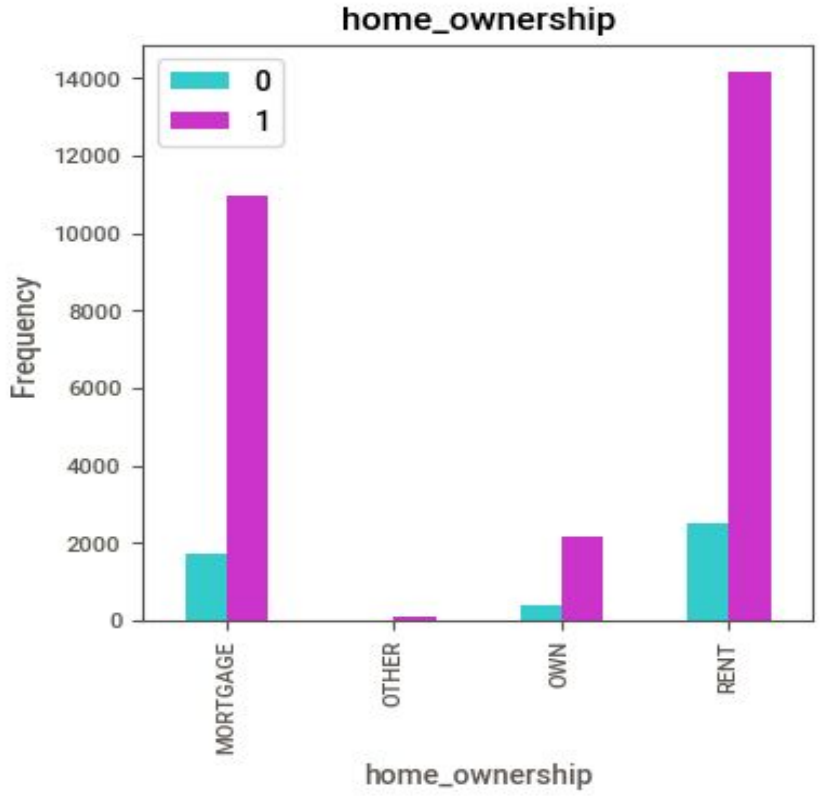
People having 25k-50k salary took more loan which leads to more defaulters.

GRADE segmented



People who comes under 10-20 dti_bucket tend to take more loans which leads to more defaulters.

Annual income segmented



People who own a owned house tend to take lesser loan which leads to less defaulters.

From the univariate analysis the most imp. features which leads to loan defaulters are as follows:

- Loan_amnt
- int_rate
- annual_inc
- grade
- emp_length
- purpose
- home_ownership

Final Observation

- Loan amount is directly proportional to defaulters
- Interest rate is directly proportional to defaulters
- For 60 months term the loan amount taken is higher, hence the defaulter is higher.
- As the year increases people tend to take more loans (exponentially increasing)
- People with less than 1 year of exp and who takes a loan have a greater chance to be in a defaulter range.
- People who own an owned house tend to take lesser loan which leads to less defaulters; people who have rented houses take more loans which tends to be defaulters
- People having 25k-50k salary took more loan which leads to more defaulters.