

Gramener Case Study

Group Name: Datalogists

1. Indira Kumari Saraswatula
2. Samadunnisha Sheik
3. Prudvi Bilakanti
4. Sampath Kanduri



Business Objectives



Gramener company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

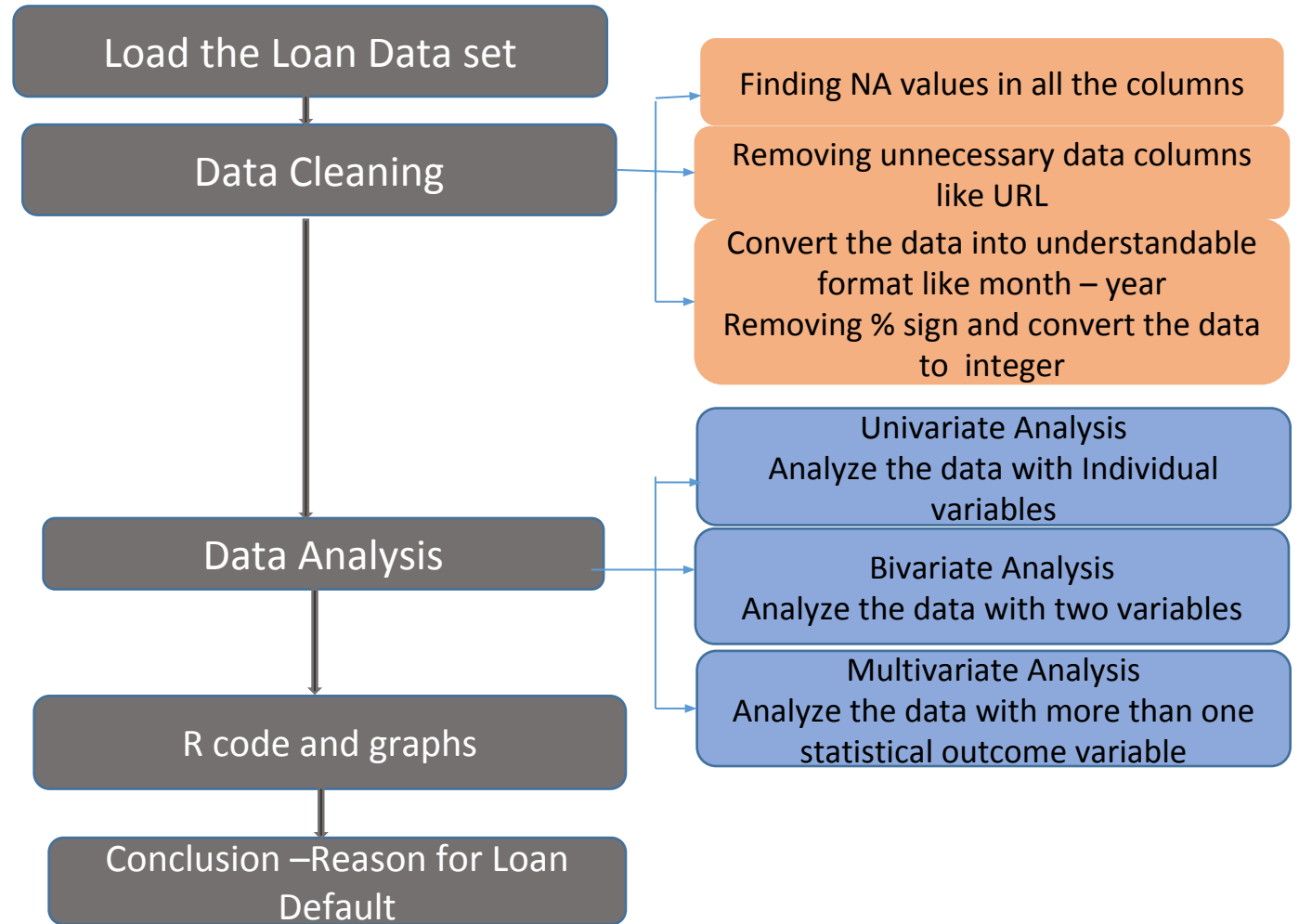
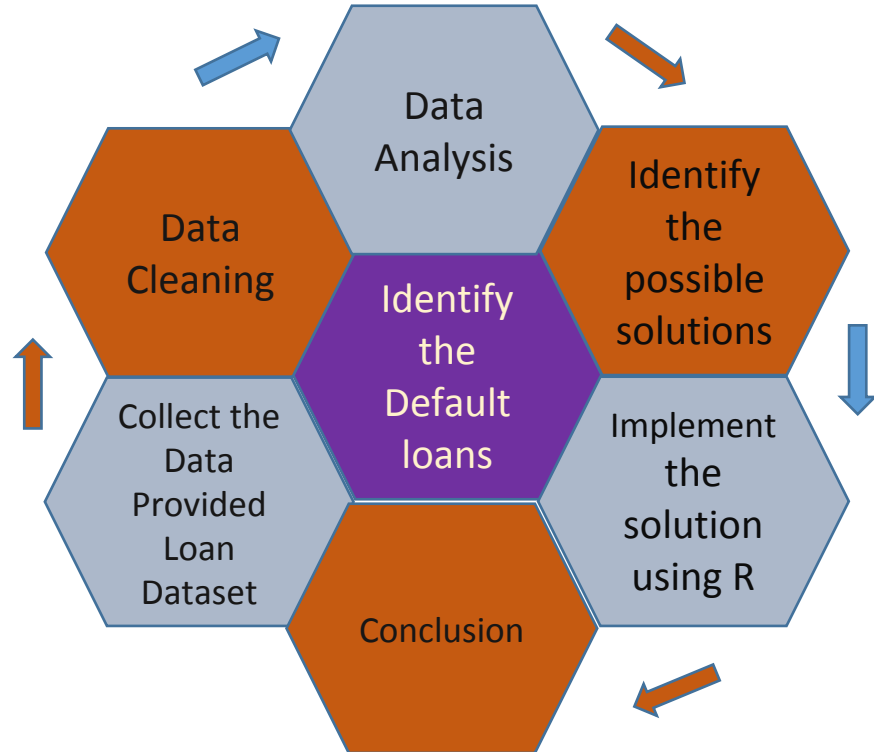
Objective:

Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

- The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default Identify the best suited investment type.
- We have to identify these risky loan applicants and driving factors, so company can utilize this knowledge for its portfolio and risk assessment.

Conditions/Constraints :

- Consider only customers who's loan status is Fully Paid and Charged Off.
- Understand the driving factors of the Charged off
- Do Univariate analysis , Bivariate Analysis and multivariate analysis to find the default loan.

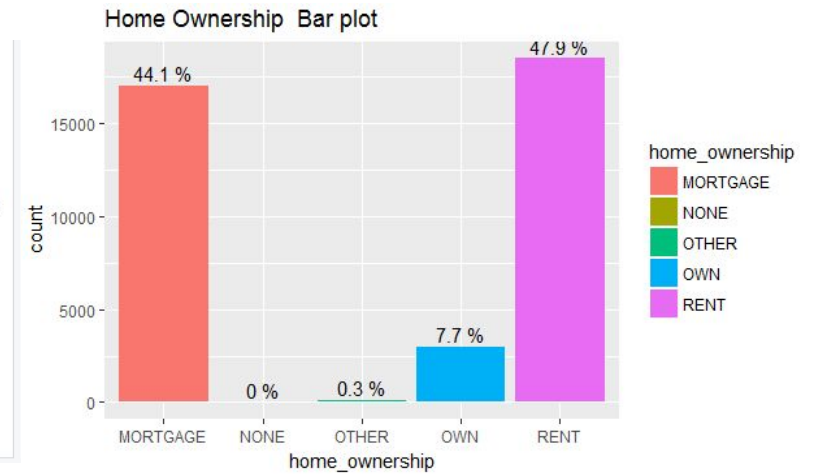
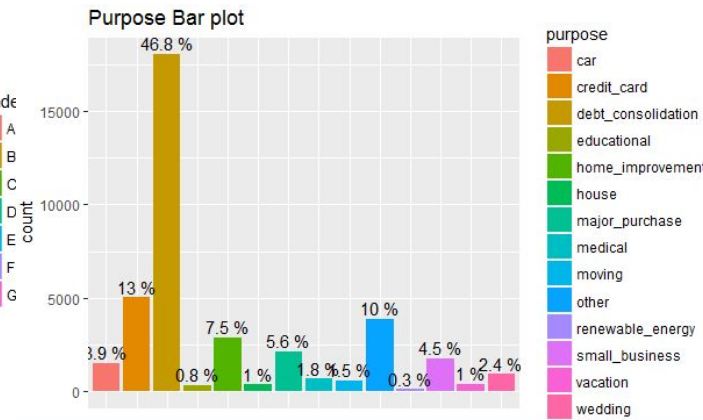
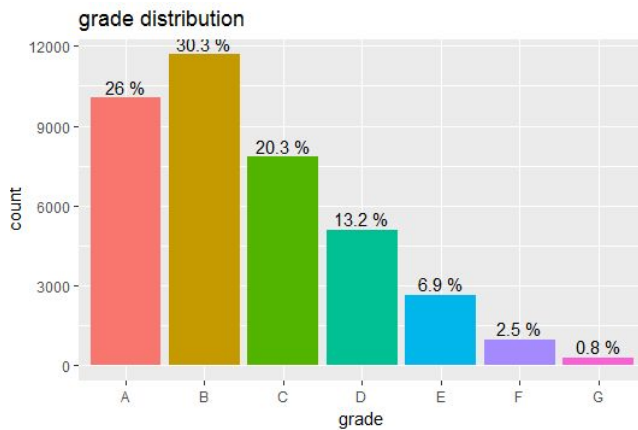


- Clean up has to be done for Target Data set of loan_status = 'Fully Paid' and Charged Off
- Remove Columns which has only NA as data. Example : annual_inc_joint, dti_joint, verification_status_joint
- Remove All columns which has only zero as data.
- Clean columns which doesn't provide any insights about why users are defaults
Eg: out_prncp', 'out_prncp_inv', 'next_pymnt_d', 'recoveries', 'id
- Convert Date columns to date format and extract month-year Eg : issue_d
- Format below variables for better understanding of data

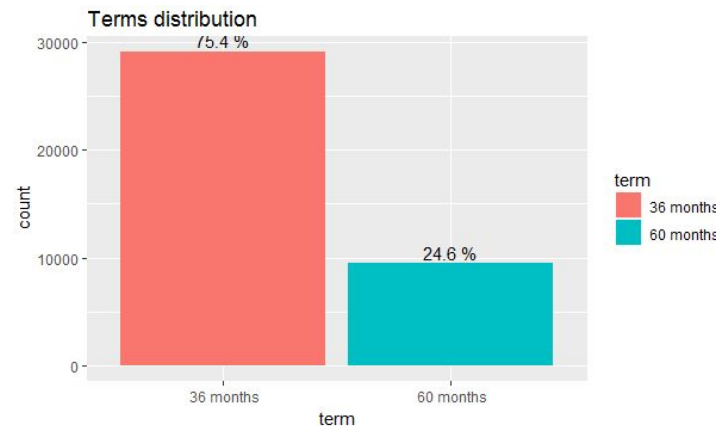
funded_amnt_inv
annual_inc
last_pymnt_d
Int_rate
installment
dti
total_pymnt
total_pymnt_inv
last_pymnt_amnt

Univariate Analysis of Data

- Analyze the data with categorical variable individually

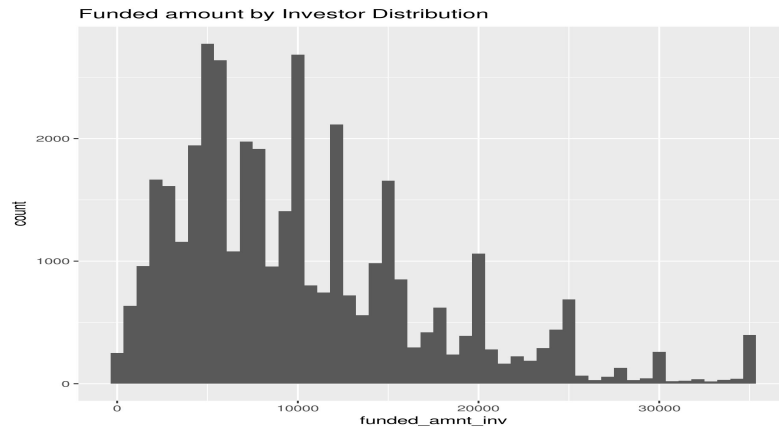


- 30.3% of dataset consist of **grade B**
- Most common Purpose for loan is **debt consolidation**



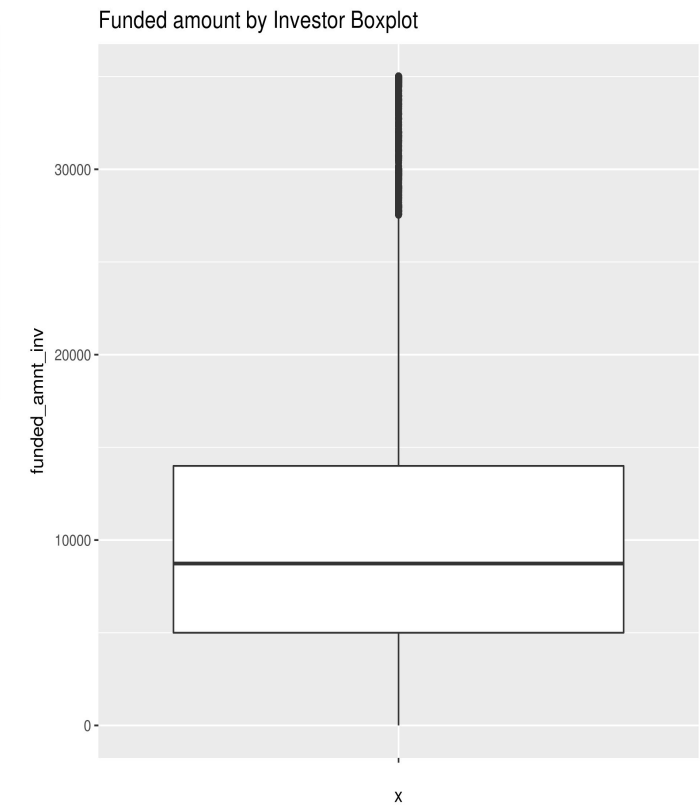
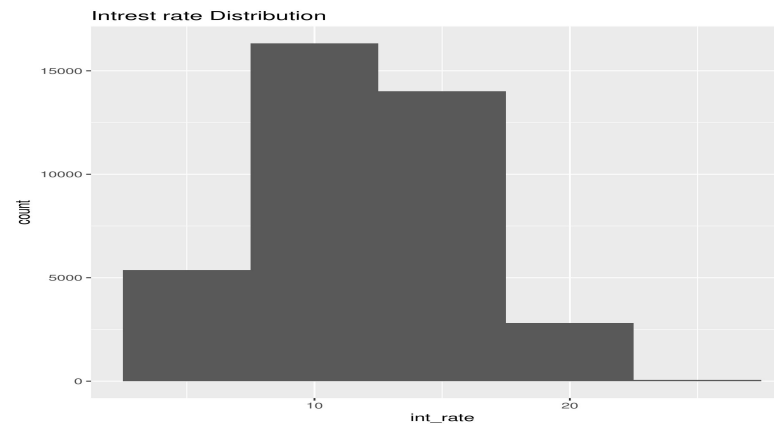
- Mortgage and Rent** are more in number
- Majority of people term is **36 Months**

Univariate Analysis of Data



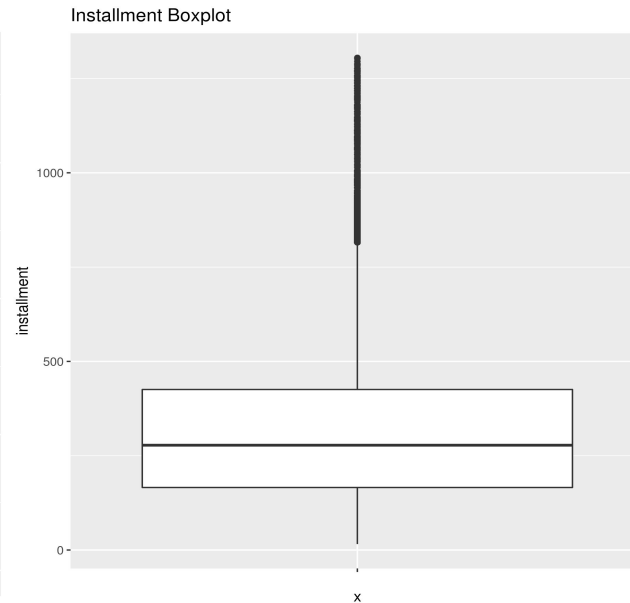
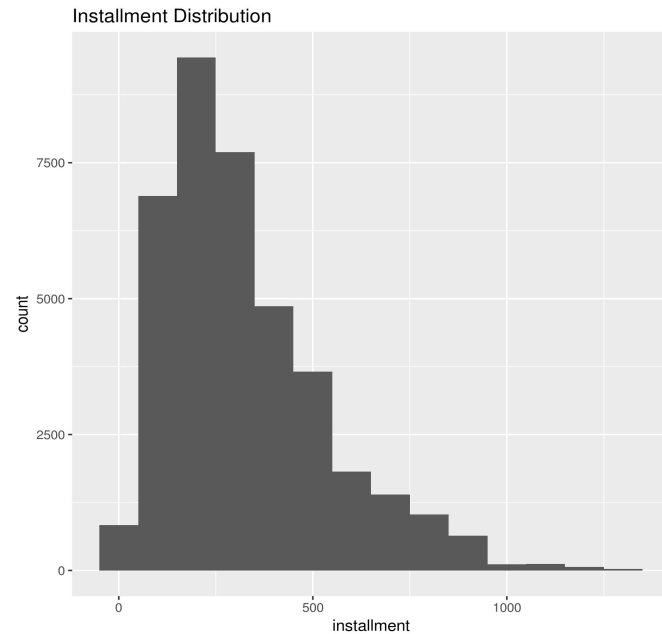
```
summary(df$funded_amnt_inv)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	5000	8733	10222	14000	35000



- **funded amount investor** has **outliers** which can be easily understood with boxplot at right, difference between 3rd quarter and max is more which also suggest presence of outliers
- Interest around 10 to 15 has high count than other rates.

Univariate Analysis of Data

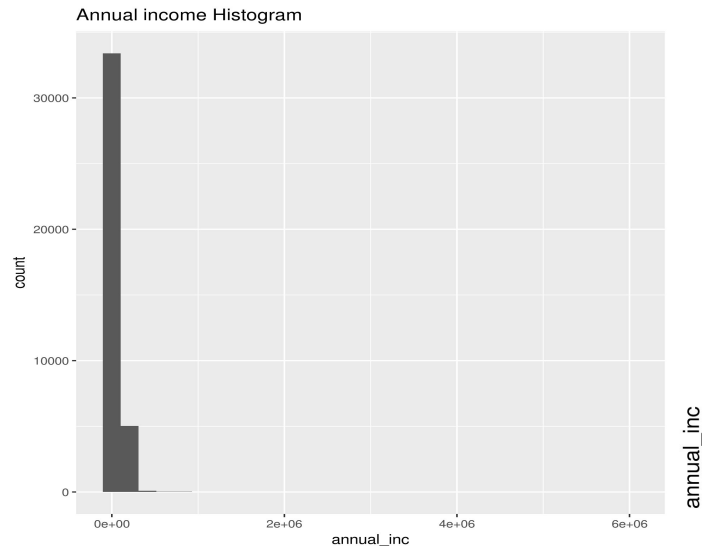


```
summary(df$installment)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.69	165.74	277.86	322.47	425.55	1305.19

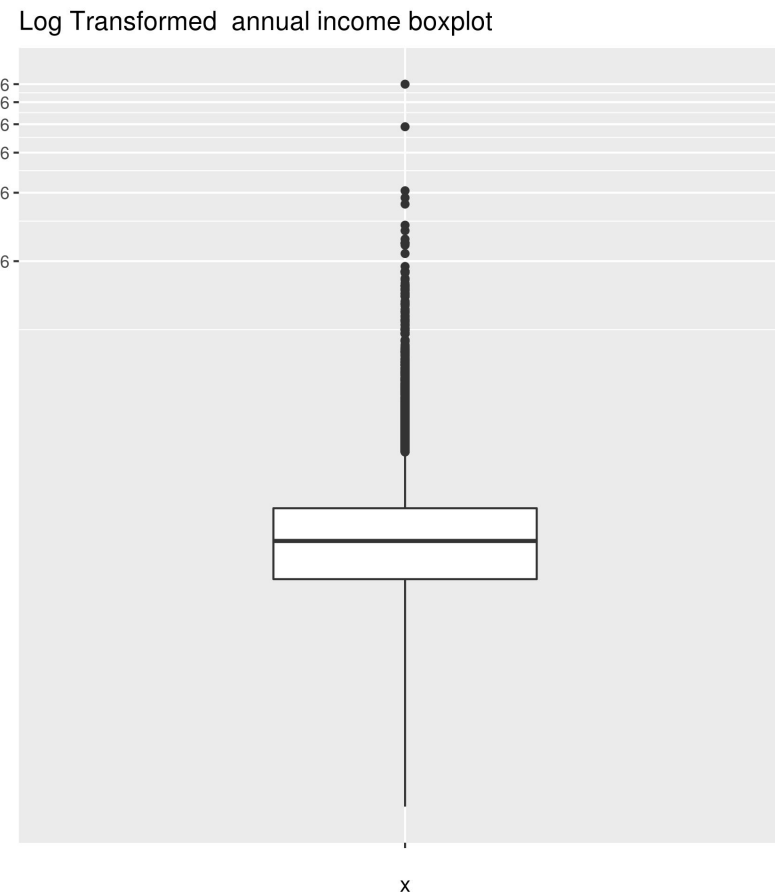
- Installment Distribution, Most of the installments are around **200 and 400**
- Difference in 3rd quarter and max is around 500 which tells us that they consist of **outliers**
- Some people pay **large installments** to **payoff** debt as **early** as possible

Univariate Analysis of Data

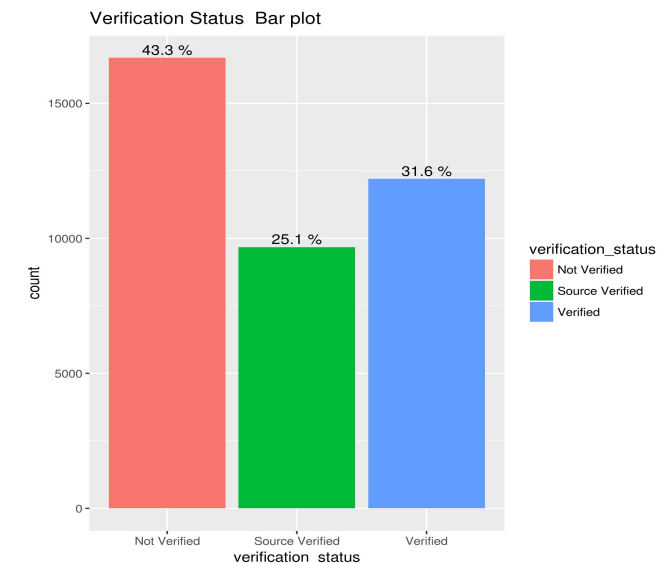


- **Annual income** has **skewed** distribution and it has large **outliers**

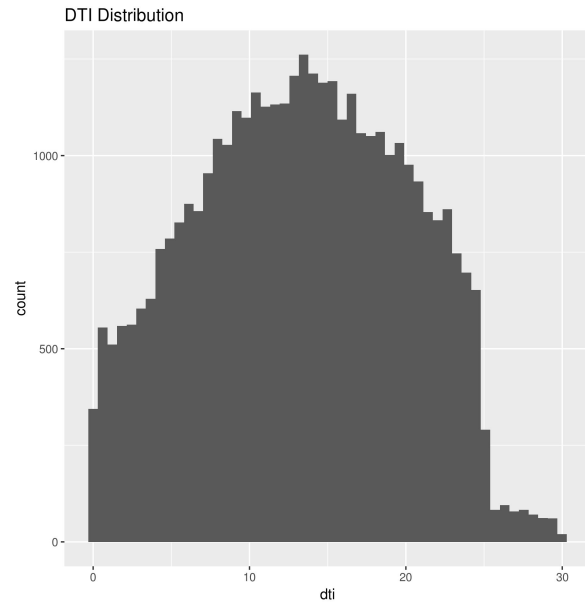
```
summary(df$annual_inc)
Min. 1st Qu. Median Mean 3rd Qu. Max.
4000  40000  58868  68778  82000 6000000
```



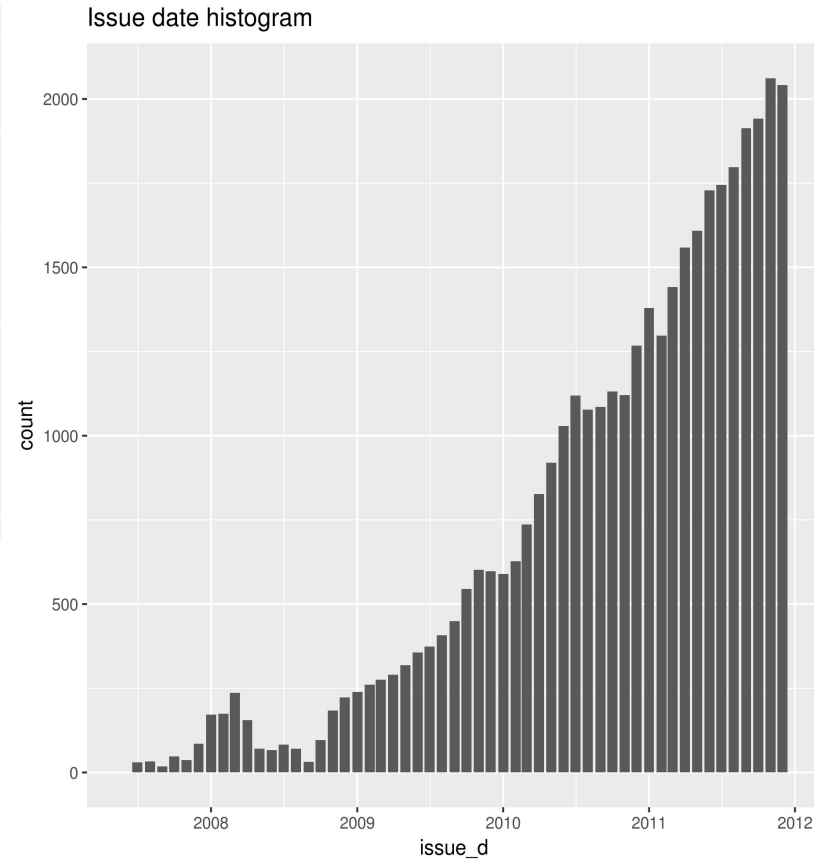
- Large values are hard to analyse we transform to look plot in different scale ,here we use **Log transformation**
- **Verification Status** has many people with **not verified**



Univariate Analysis of Data

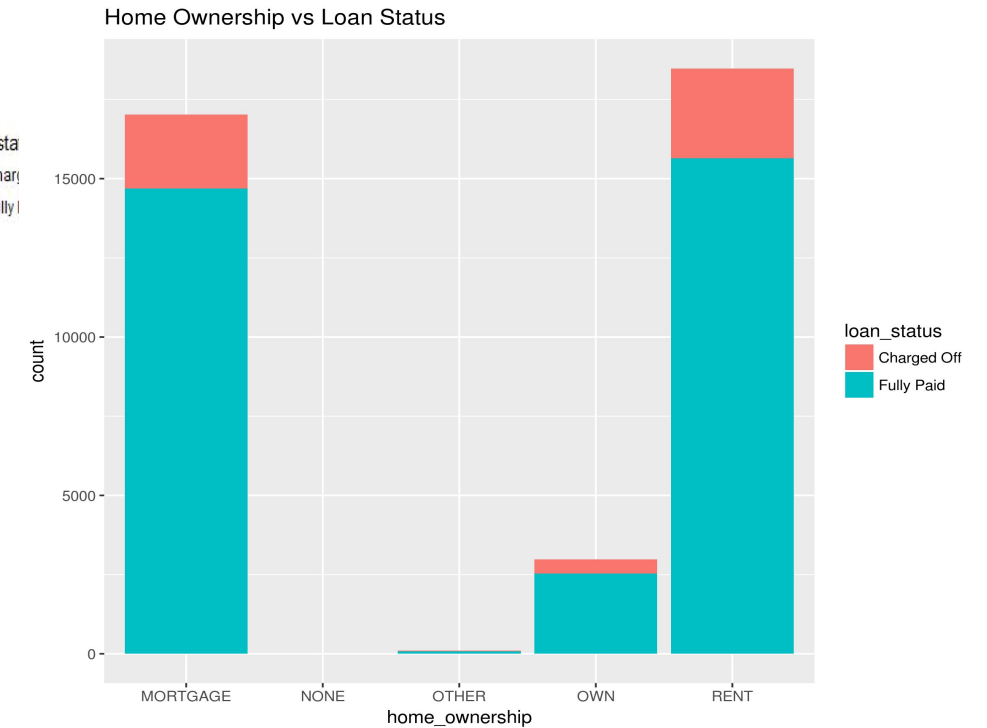
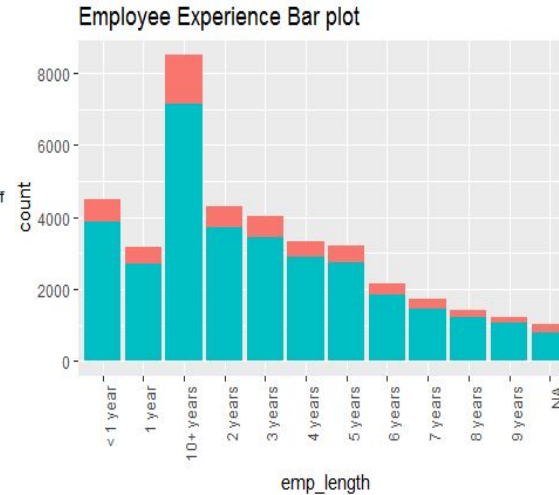
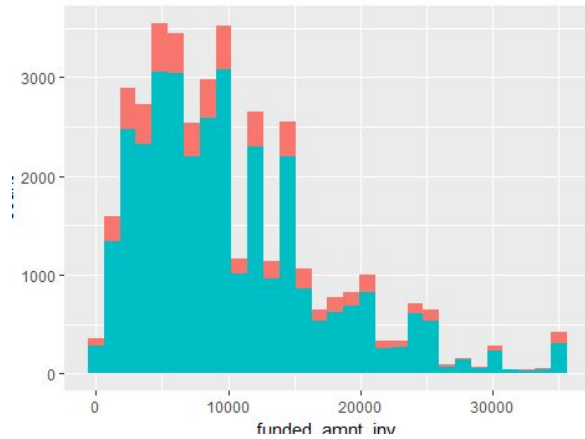


- Debt to income ratio there is sharp **drop off near 28**



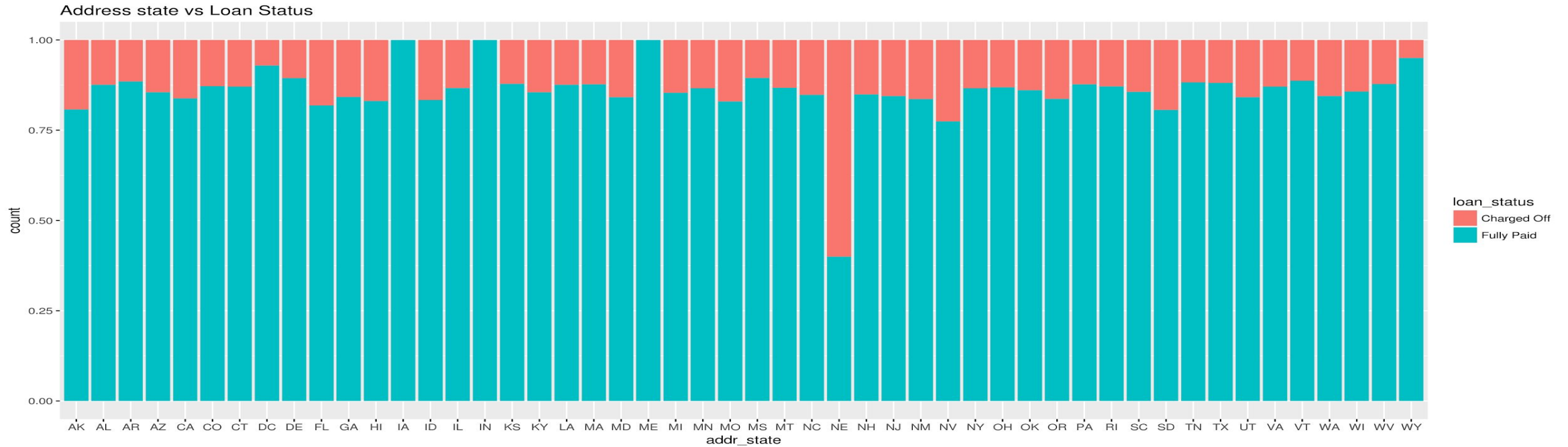
- As **years increases count increases** which indicates that customers for loan are increasing as year goes by

Bivariate Analysis of Data



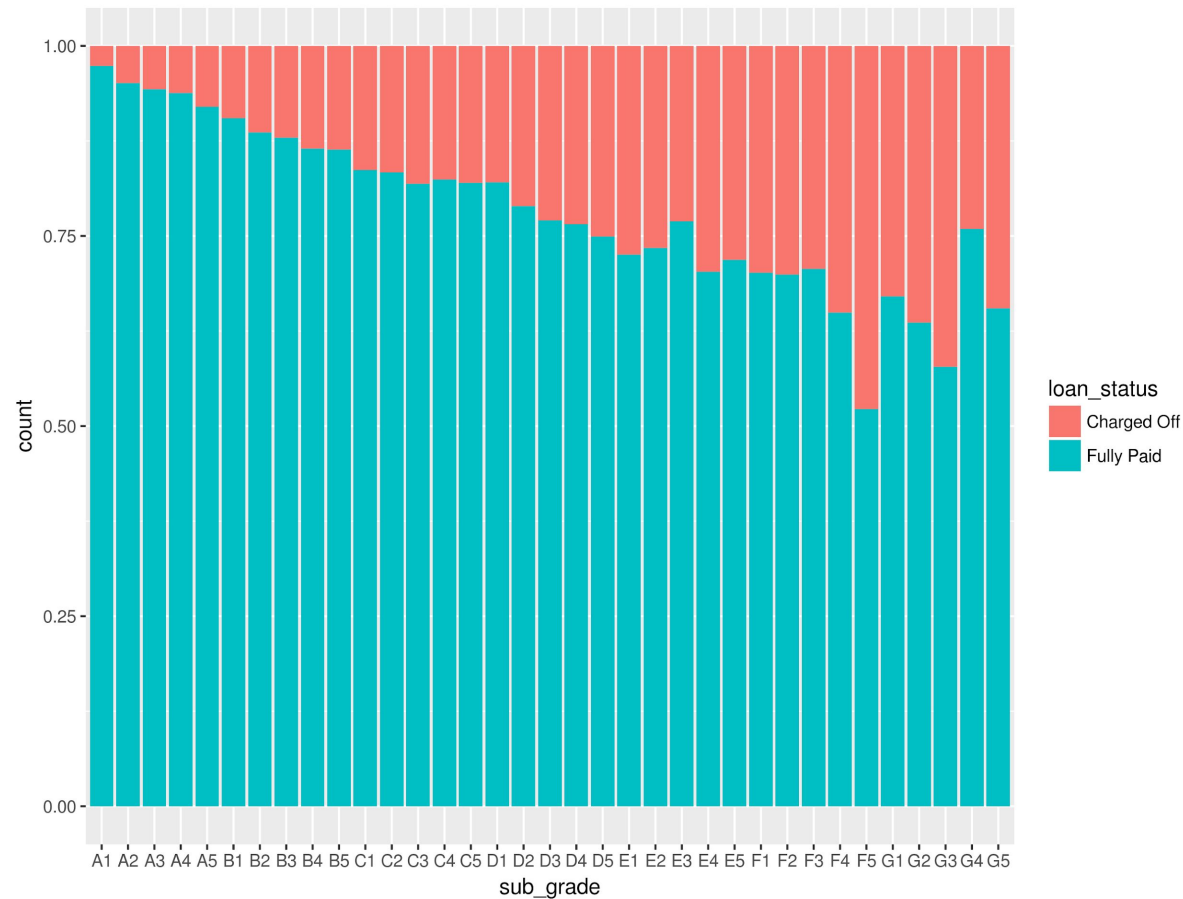
- Employee with **10+ years experience default** more than any other
- funded amount by investors doesn't have impact directly on loan status
- Most of the Defaulters are from Mortgage and Rent

Bivariate Analysis of Data



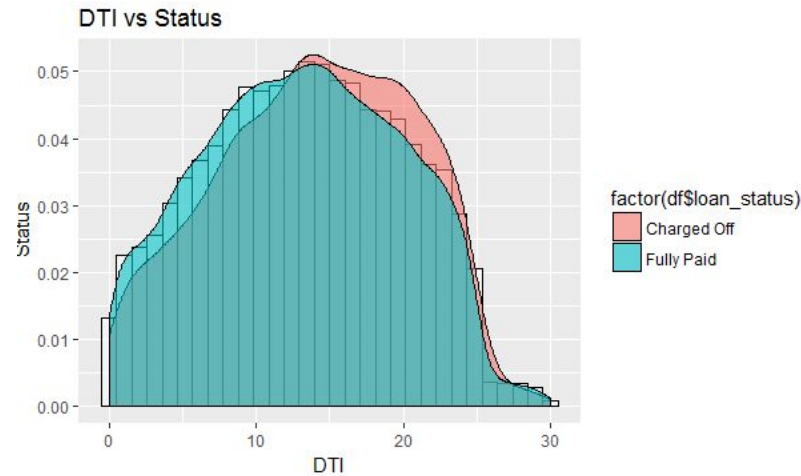
- **NE** as high amount of defaulters than any other in state
- **IA ME WY** states has less number of defaulters which can be good for banks to provide loans to all in that region

Bivariate Analysis of Data

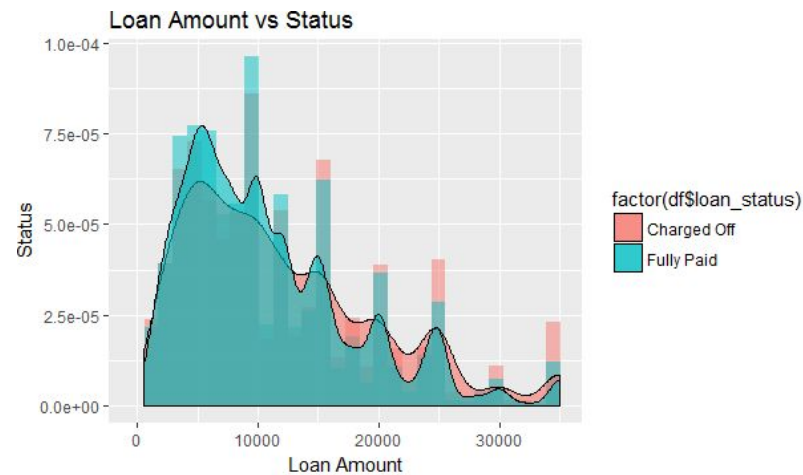


- **F1 , G3 , G5 has high percent of defaulters its better to avoid those sub grade**

Bivariate Analysis – Continuous variable



- As Debt to income ratio increases defaulters increases

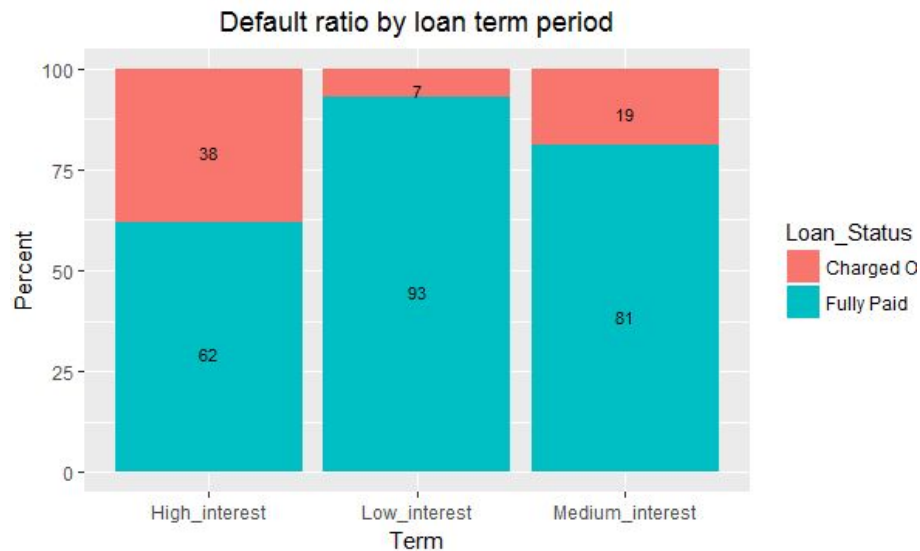


- As Loan amount increases defaulters will be high than non defaulters



- Finding Interest rate is affecting the loan default –
- Create a bins (define range) for interest rate

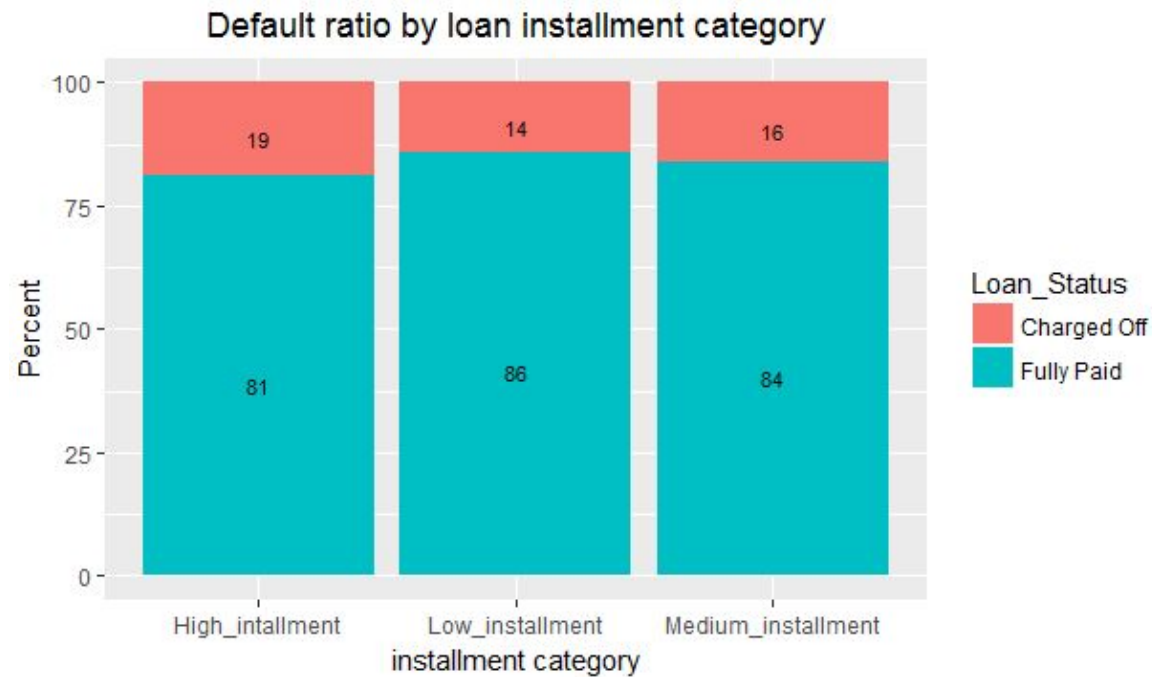
0 to 10	Low_interest
11 to 20	Medium_interest
>20	High interest



This observation
gives High Interest
rate is affecting the
loan defaulter

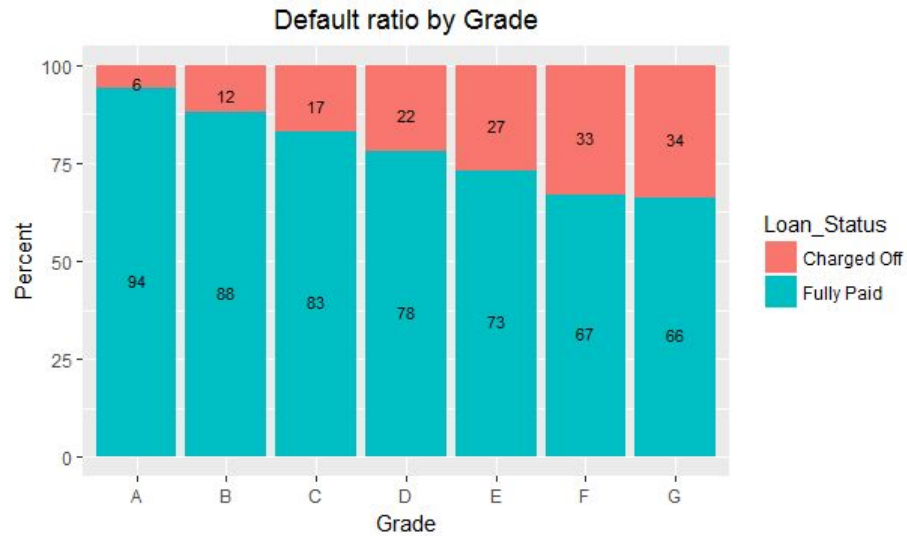
- 2) Finding installment is affecting the loan default
- Create a installment into buckets

0 to 450	Low_installment
451 to 900	Medium_installment
>900	High_intallment



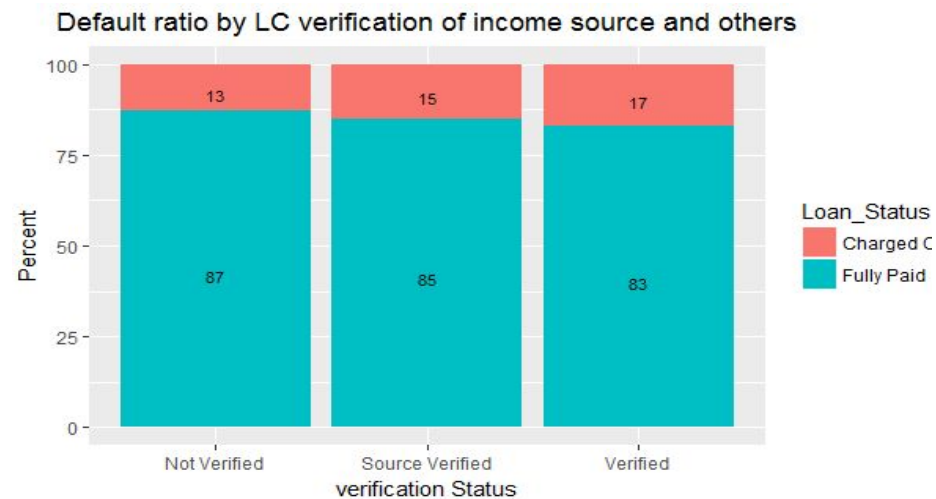
This observation gives High installment amount is affecting the loan defaulter

3) Grade Analysis



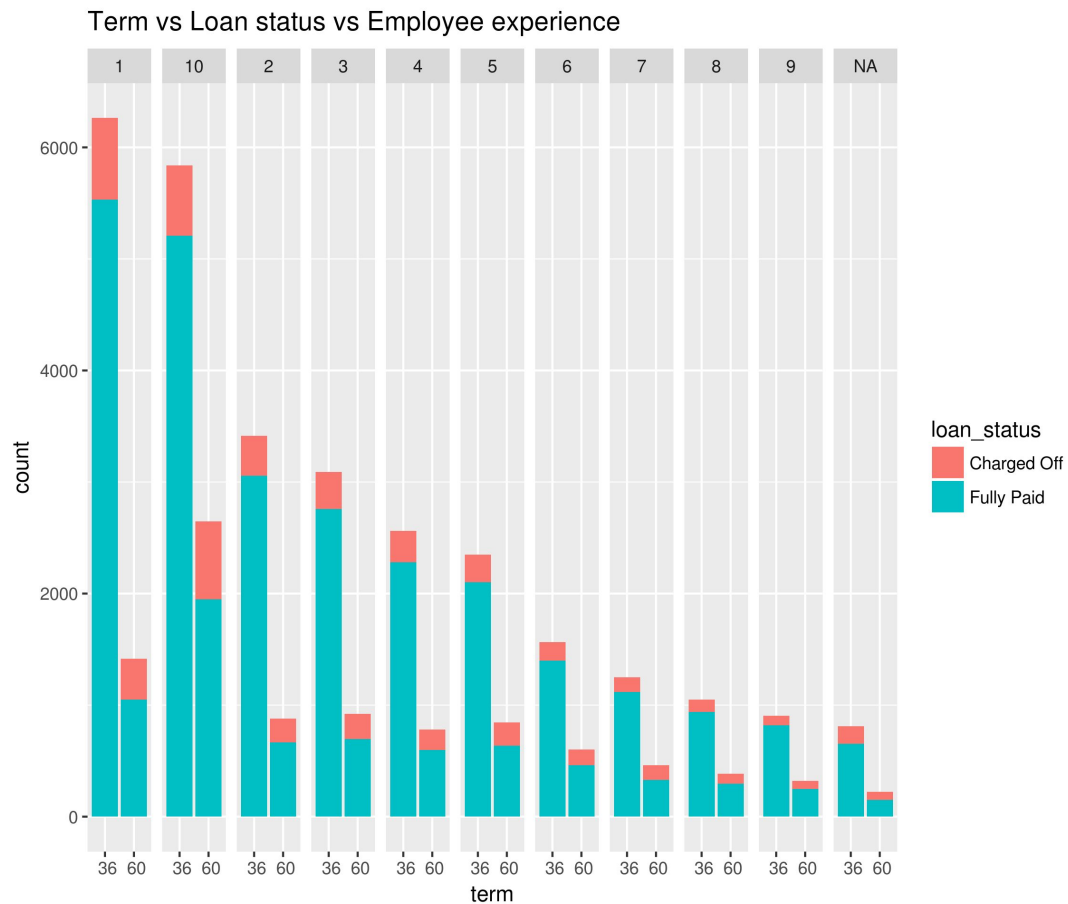
This observation gives G grade is affecting the loan defaulter

4) Verification Status Analysis



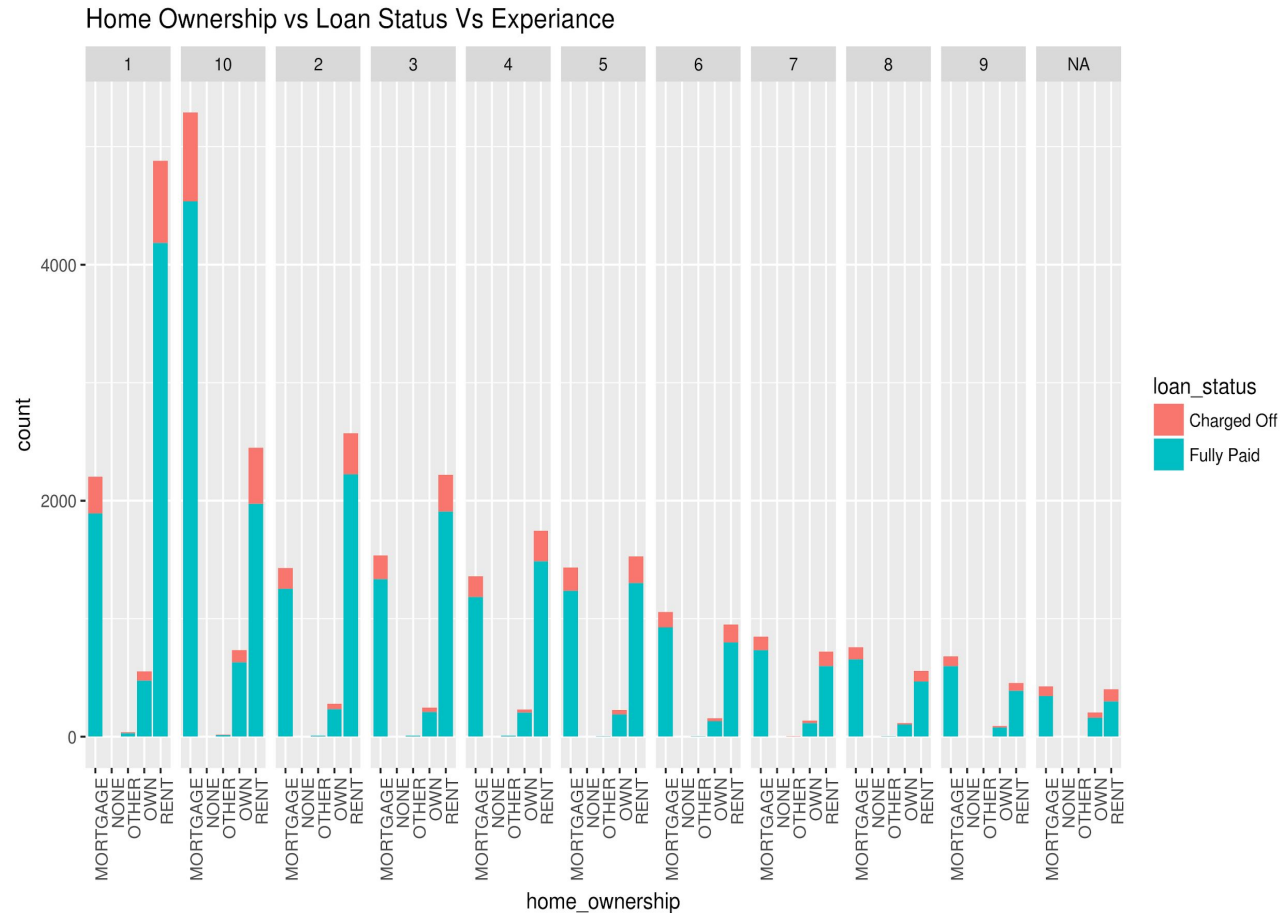
This observation gives verified is affecting the loan defaulter

Multivariate Analysis of Data



People having 10+ years of experience with month 60 have high default rate

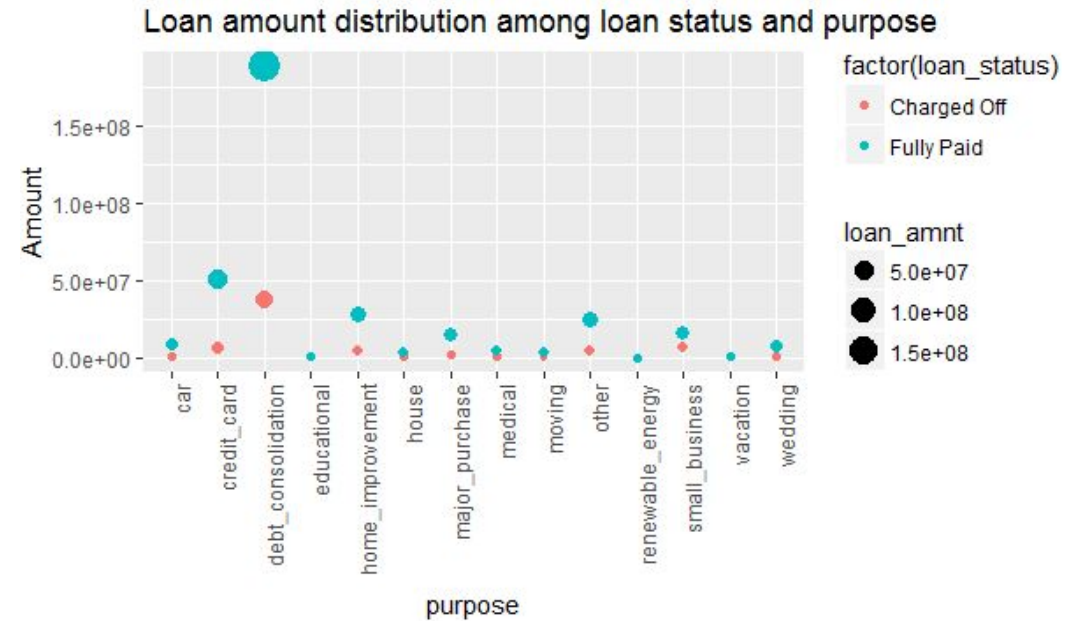
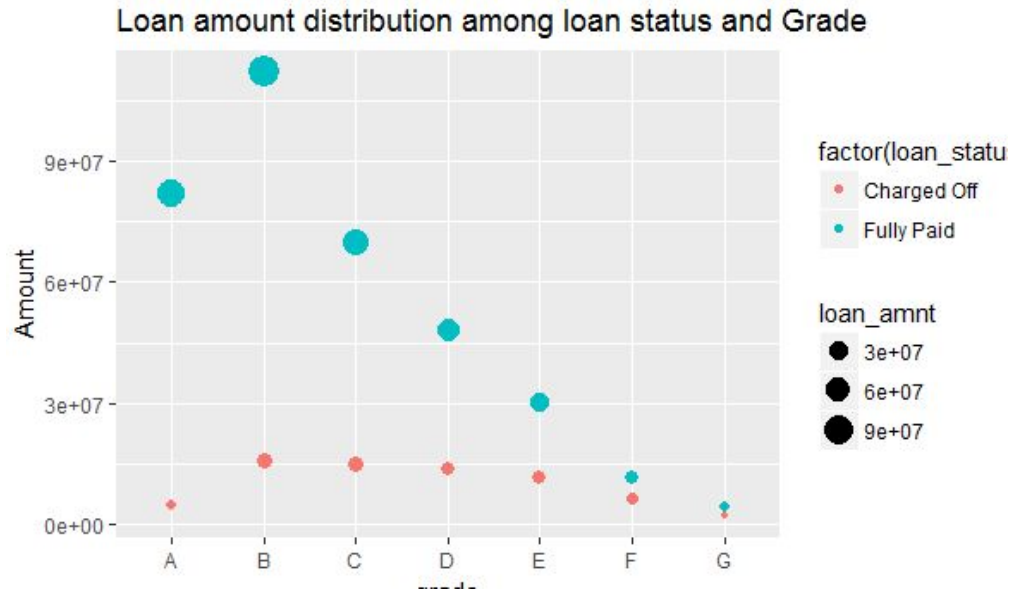
Multivariate Analysis of Data



People with 10 + years having ownership Mortgage have High Defaulters and 1 year with rent has high Defaulters

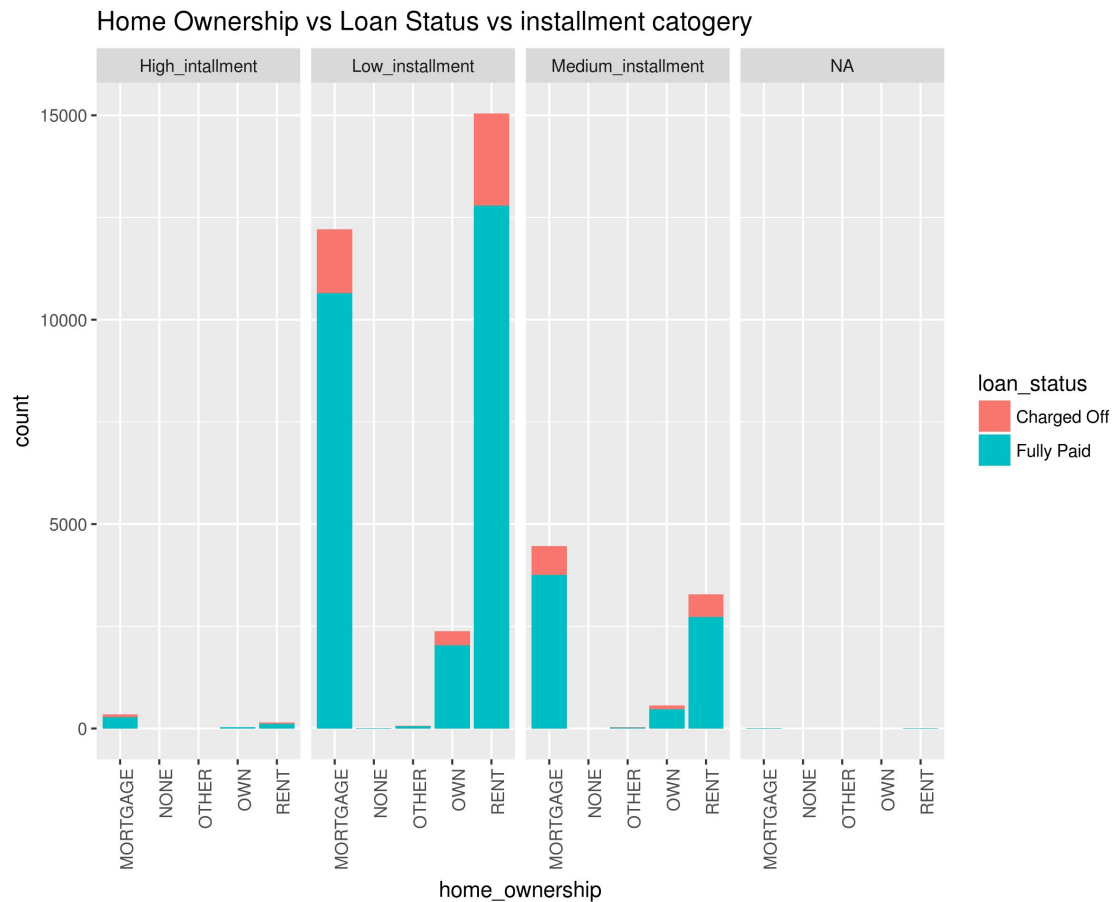
Multivariate Analysis of Data

- Analyze the income amount by two variables



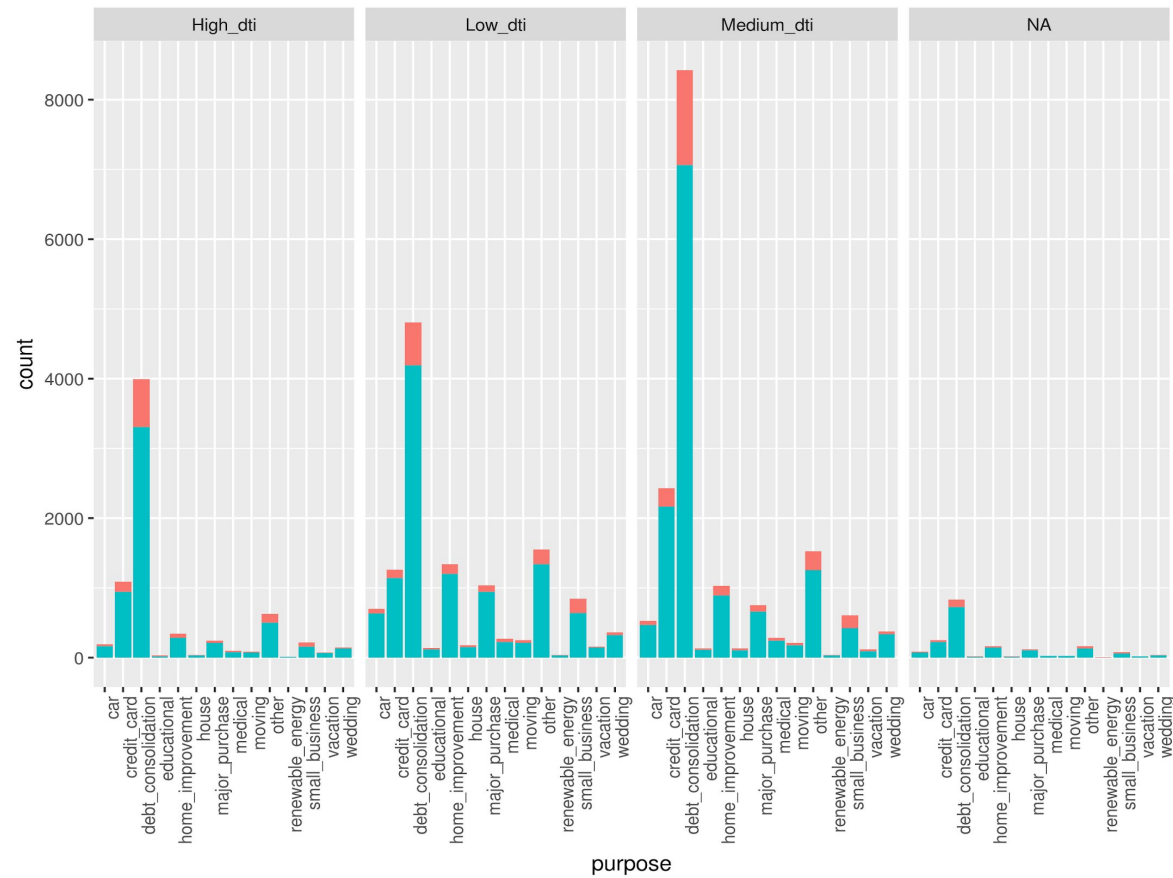
- Grade B,C,D with medium loan amount tend to default
- Debt consolidation with high loan amount have more chance of defaulting

Multivariate Analysis of Data



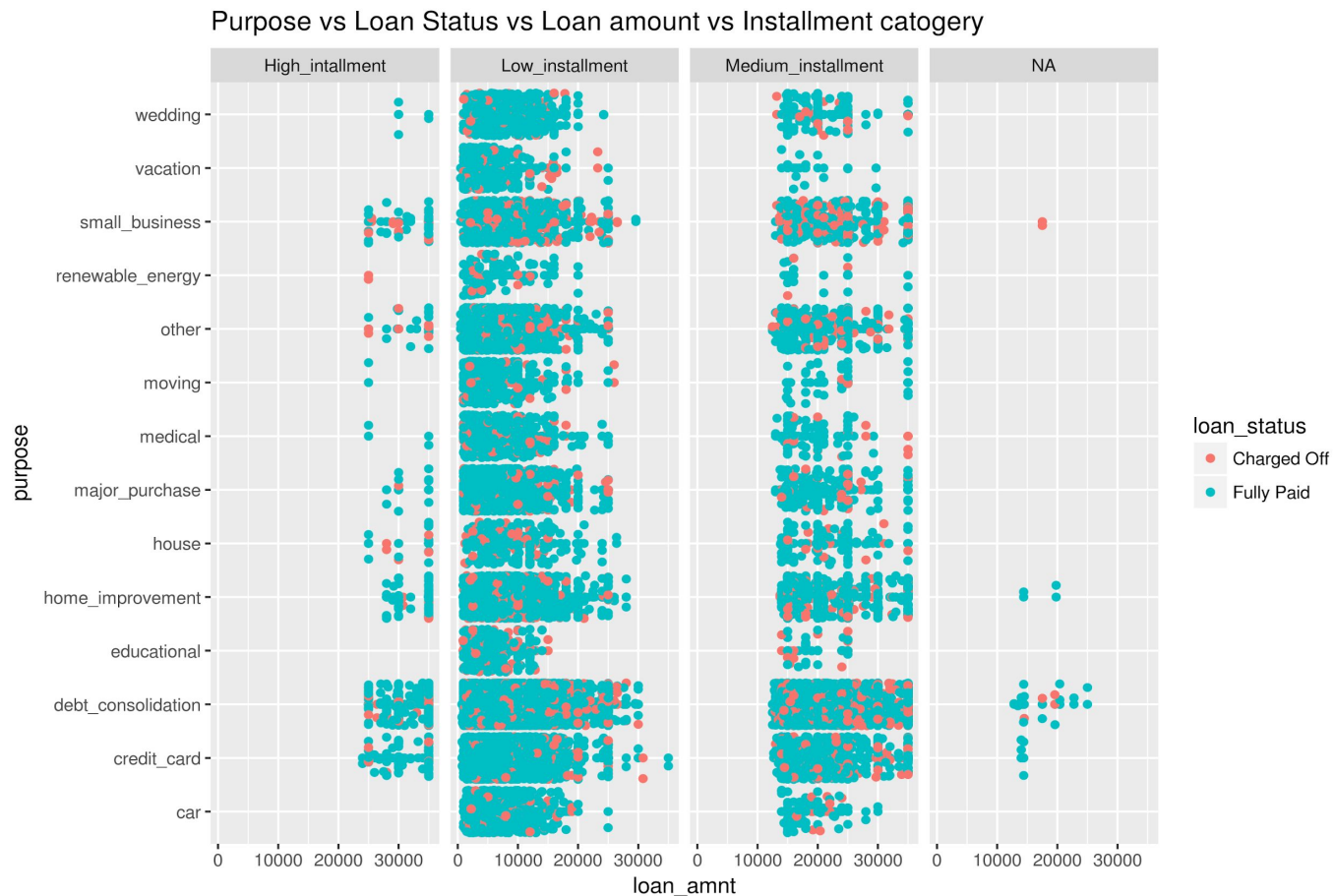
People paying low installment and having ownership mortgage and rent tend to default a lot

Multivariate Analysis of Data



People with medium dti and reason for loan debt consolidation tend to have high defaulters

Multivariate Analysis of Data



People paying low and medium installment having purpose debt consolidation and small business tend to default more

Conclusion & Recommendations

Loan amount vs Interest rate group	56 % of High interest rate that > 20 are effecting
Loan amount vs Purpose	61 % debt_consolidation for High interest rate are affected
Loan amount vs Verification Status	62 % Verified status for High interest rate are affected
Loan amount vs Grade	61% E and F Grades for High interest rate are affected
Loan amount vs Home ownership	62% Rent home for High interest rate are affected

- In EDA we have seen many features interrelated like purpose and loan status.It is recommended for company to avoid giving loans to debt consolidation and category having ownership (rent and mortgage)
- Also avoid low installments with purpose debt consolidation