Gramener Case Study

Content

1. Data Exploration
2. Problem statement and Data Cleansing
3. Univariate Analysis
4. Correlation Matrix
5. Bivariate Analysis
6. Multivariate Analysis
7. Recommendations

Data Exploration

1. Dataset contains 111 columns.
2. But Categorizing those as “Consumer Attributes” and “Loan Attributes” would considering only 25 of them for analysis.

Consumer Attributes

1. member_id
2. emp_title
3. emp_length
4. home_ownership
5. annual_inc
6. verification_status
7. url
8. Desc
9. Purpose
10. Title
11. zip_code
12. addr_state
13. application_type

Loan Attributes

1. id
2. loan_amnt
3. funded_amnt
4. funded_amnt_inv
5. term
6. int_rate
7. installment
8. grade
9. sub_grade
10. issue_d
11. loan_status
12. dti

Problem Statement & Data Cleansing

- **Problem Statement :**

  To identify the driving factor behind the Loan Defaults.

- **Cleansing Tasks :**

1. Deleting all irrelevant columns.
2. Checking NA values if those are their in required columns.
3. Standardizing the timestamp by adding “01-” default date before converting it into date time format.
4. Replacing “%” symbol from interest rate column so that it can be used as quantitative attribute for analysis.
5. Creating required derived columns.
6. Filtering data on loan status and considering only require no of rows.

Univariate Analysis

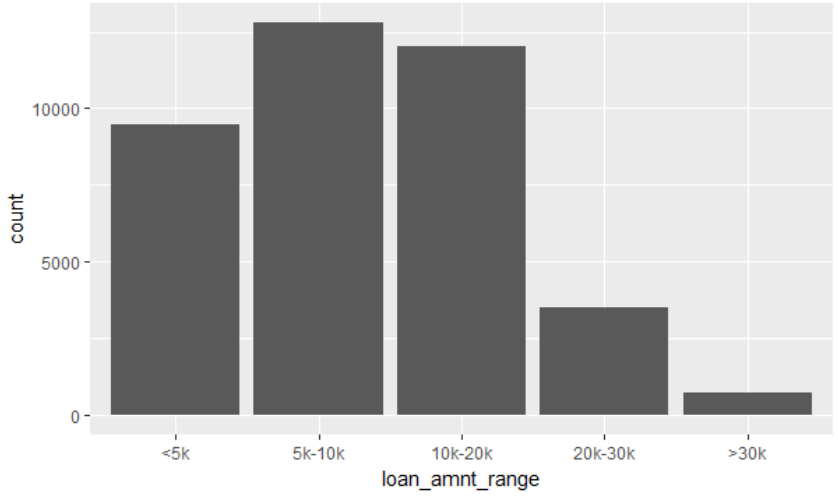
Loan Amount :

1. Creating Ranges as follows to ease out reading

- <5k
- 5k-10k
- 10k-20k
- 20k-30k
- >30k

2. Observation :

People are asking more loans in the range 5k to 10k



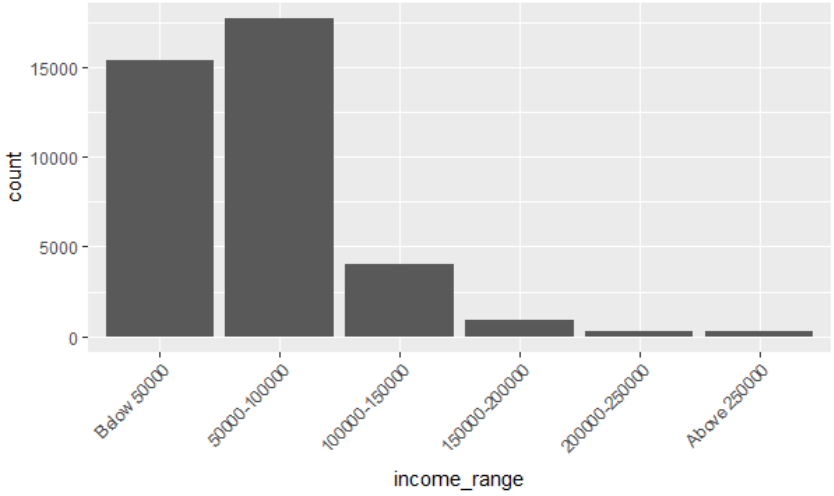
Annual Income :

1. Creating Ranges as follows to ease out reading

- Below 50000
- 50000-100000
- 100000-150000
- 150000-200000
- 200000-250000
- Above 250000

2. Observation :

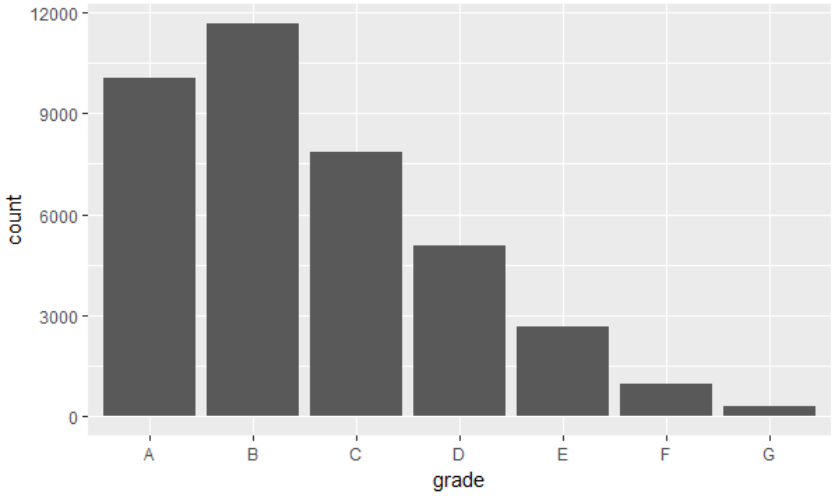
By below graph we can see that no of borrowers are more whose salary are either in rage of “50000-100000” or “Below 50000”



Univariate Analysis Contd..

Grade :

We Can see from the following graph that people with “Grade B” get most of the loan sanctioned.



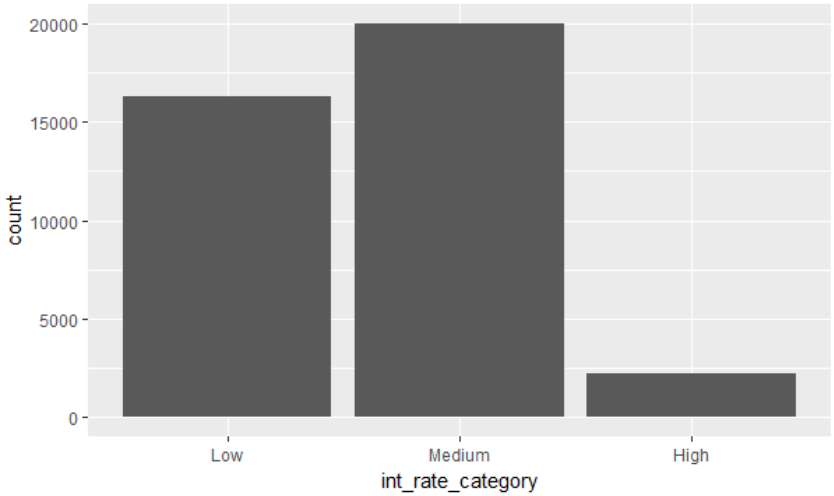
Interest Rate :

1. **Creating Ranges as follows to ease out reading**

- 5-11 : Low
- 11-18 : Medium
- 18-25 : High

2. Observation :

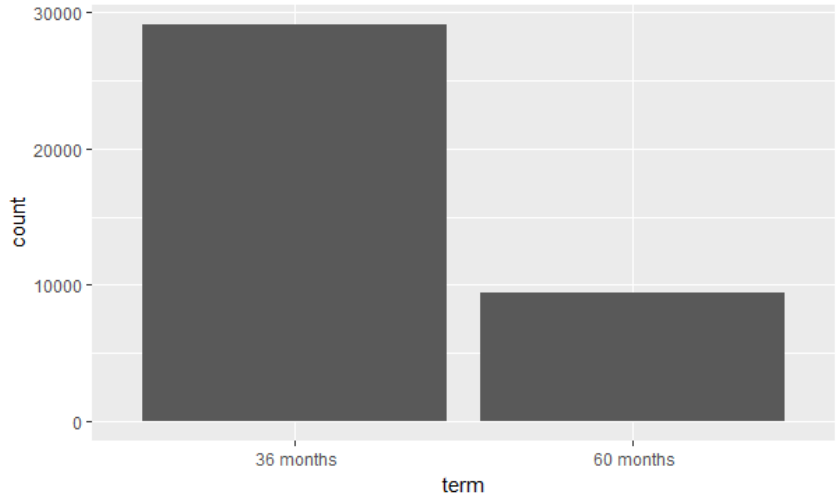
By below graph we can see that lots of borrower need to repay the amount with medium interest rate.



Univariate Analysis Contd..

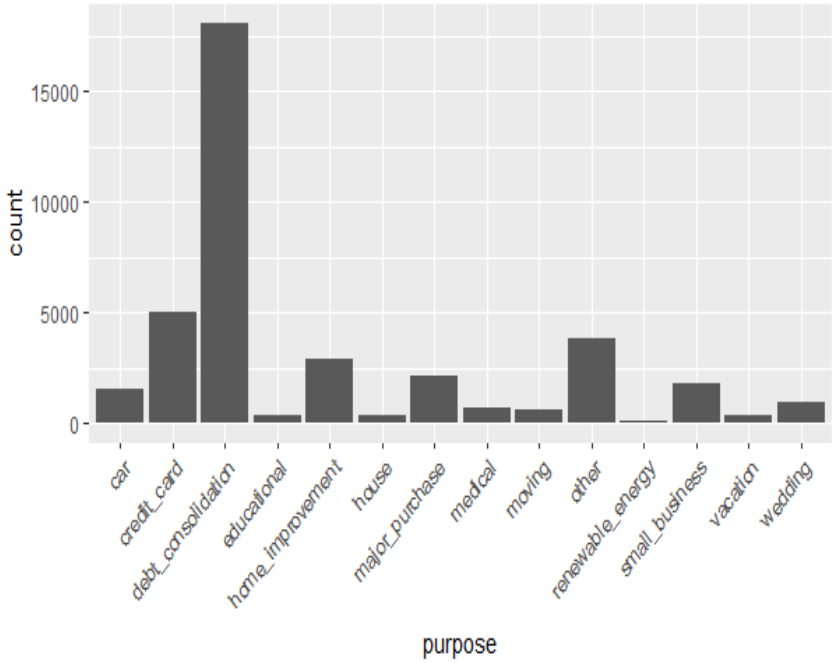
Term :

People prefer to pay the loan with interest within 36 months that is in 3 Years.

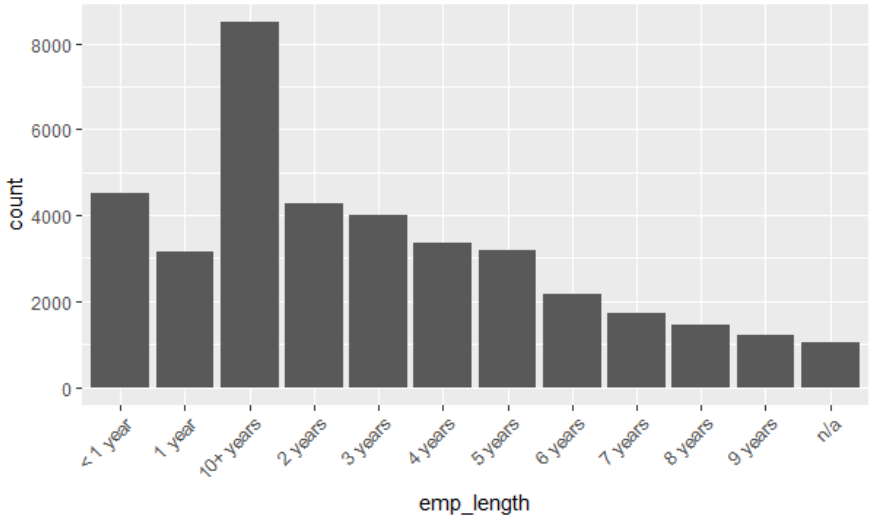


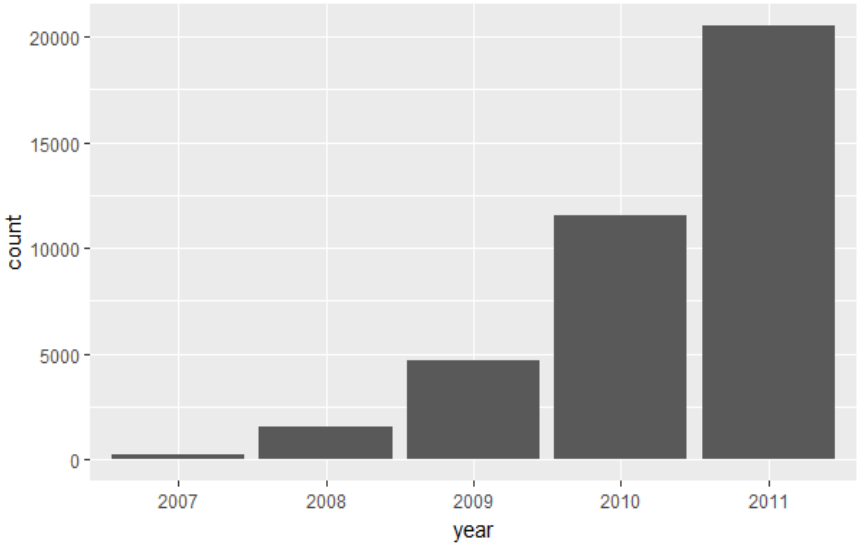
Purpose :

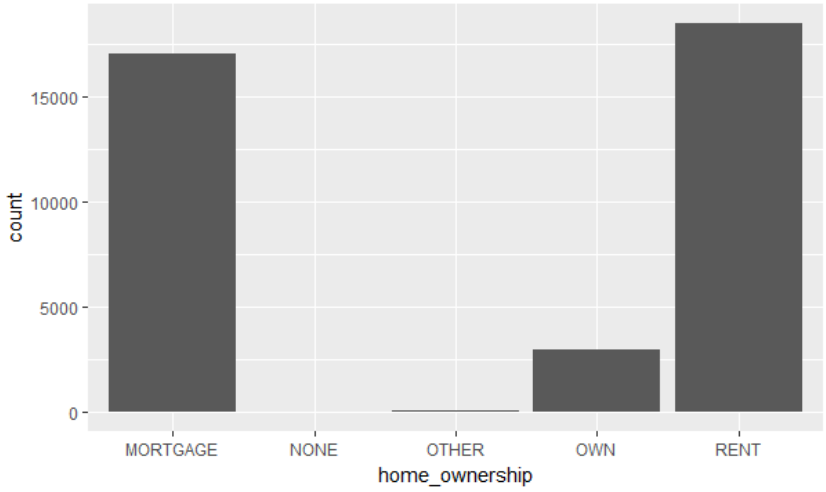
People with purpose debt_consolidation consist higher no of borrower.



Univariate Analysis Contd..







- We can see from emp_length graph that people with 10+ year are the highest borrowers
- People whose home ownership status is either “Mortgage” or “Rent” ask for more loans
- And as year pass by we can see that people borrowing loans increases.

Univariate Analysis Contd…

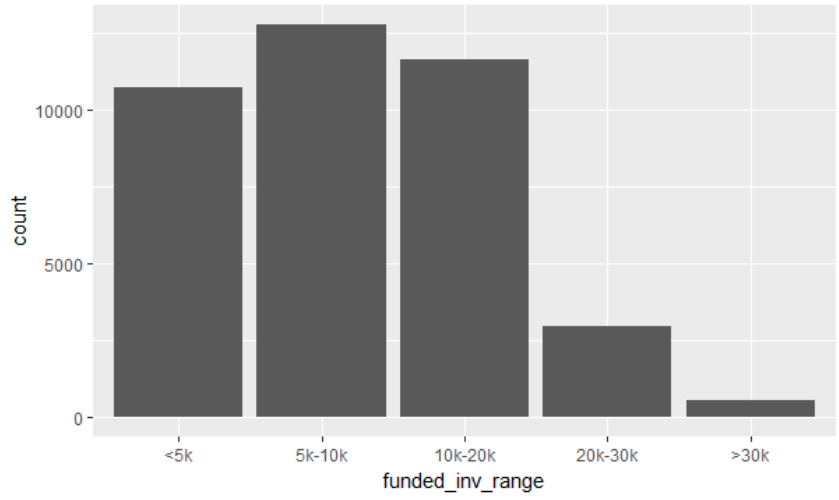
Funded amount by Investor:

1. Creating Ranges as follows to ease out reading

- <5k
- 5k-10k
- 10k-20k
- 20k-30k
- >30k

2. Observation :

People are asking more loans in the range 5k to 10k



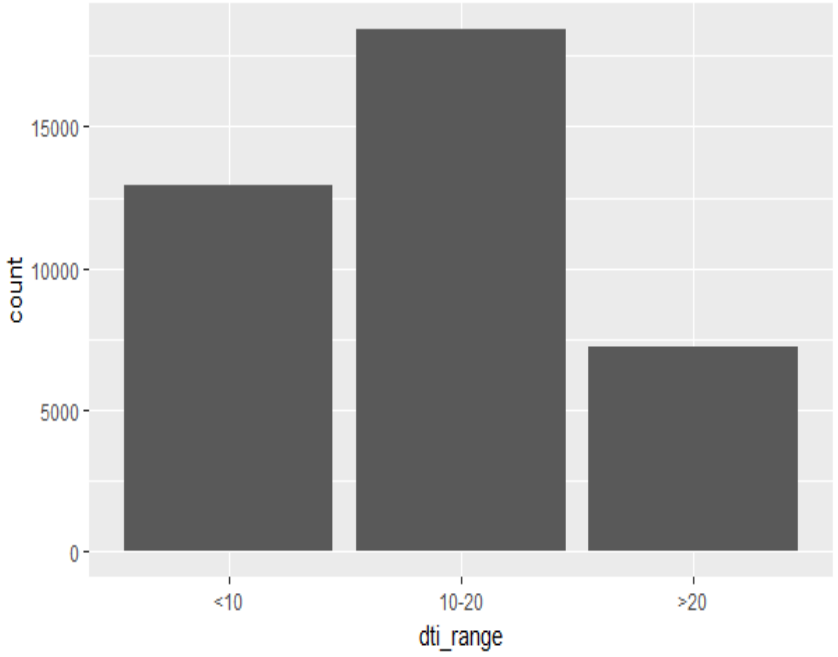
Dti :

1. **Creating Ranges as follows to ease out reading**

- <10
- 10-20
- >20

2. Observation :

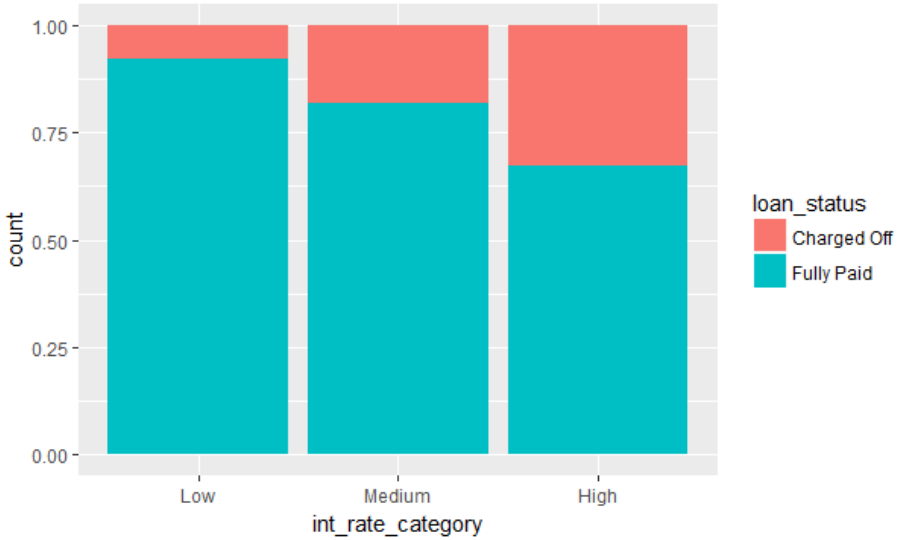
Mostly the dti ratio is between 10-20 or <10

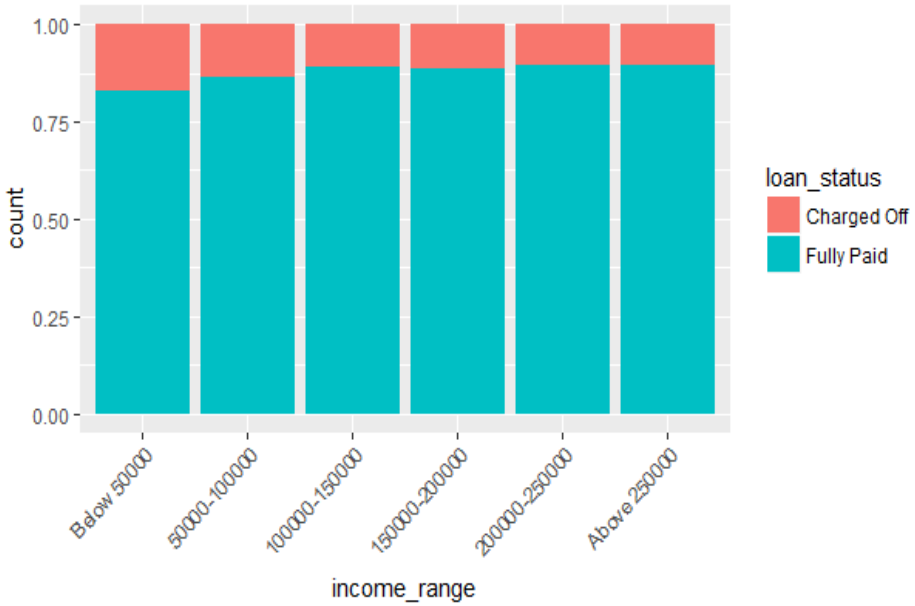


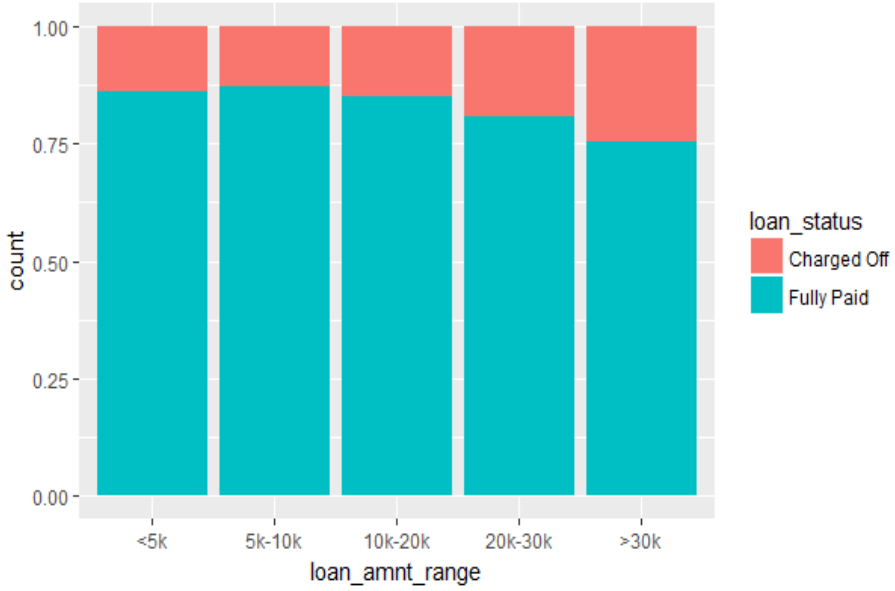
Correlation Matrix

- The observation is obvious funded amount is positively correlated to loan amount.
- Funded amount, funded amount investment and installments are dependent on each other.

Bivariate Analysis

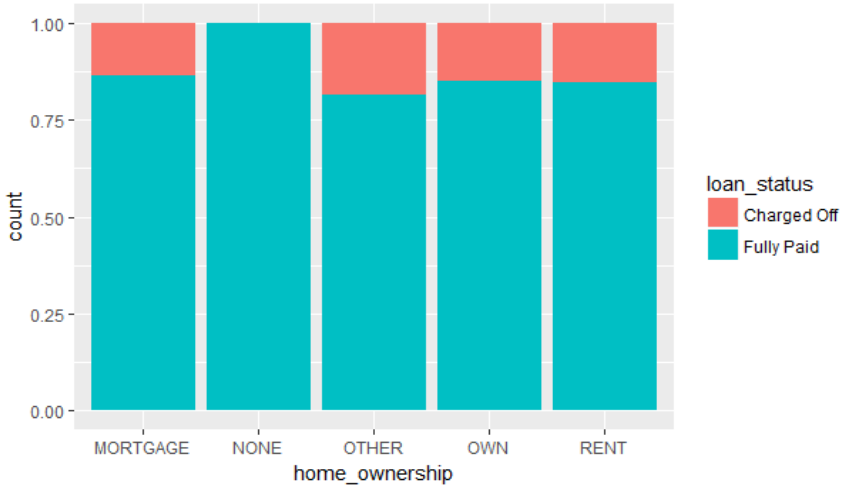


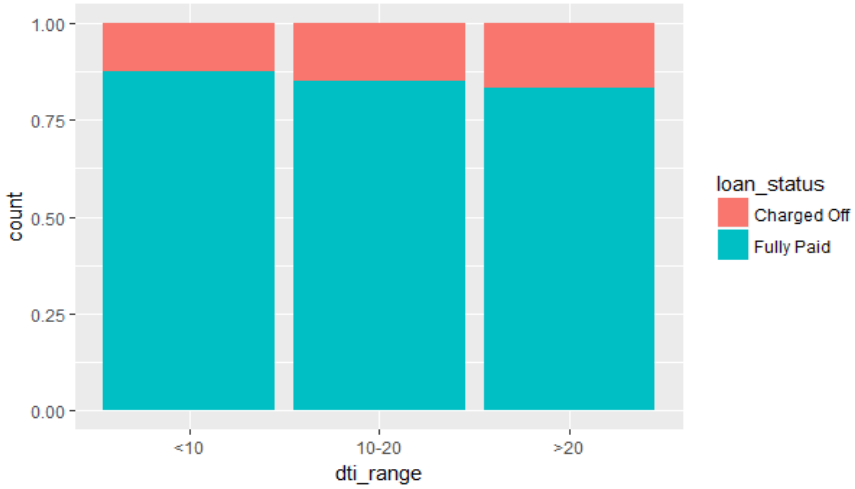


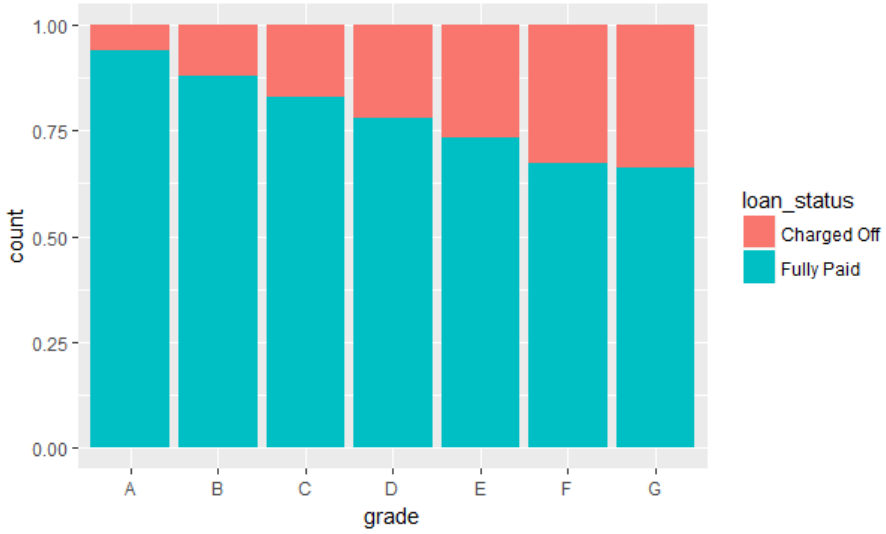


- Now in bivariate analysis focusing categorical variable and plotting it against the loan status to see how they react.
- So if you see interest rate category people with higher interest rate tends to default, same with loan amount. Higher the amount higher chances of default.
- Where as people who ask the loan below 50000 tends to defualt more as compare to others.

Bivariate Analysis Contd..

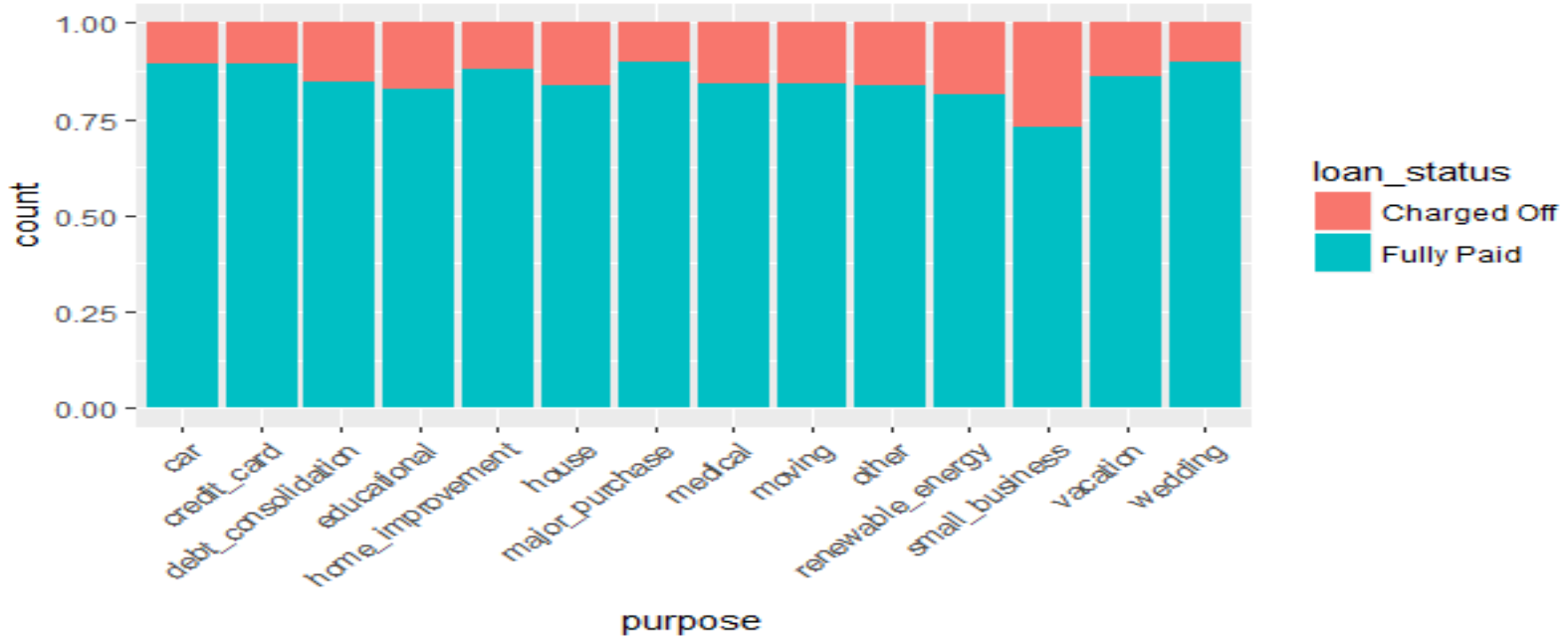


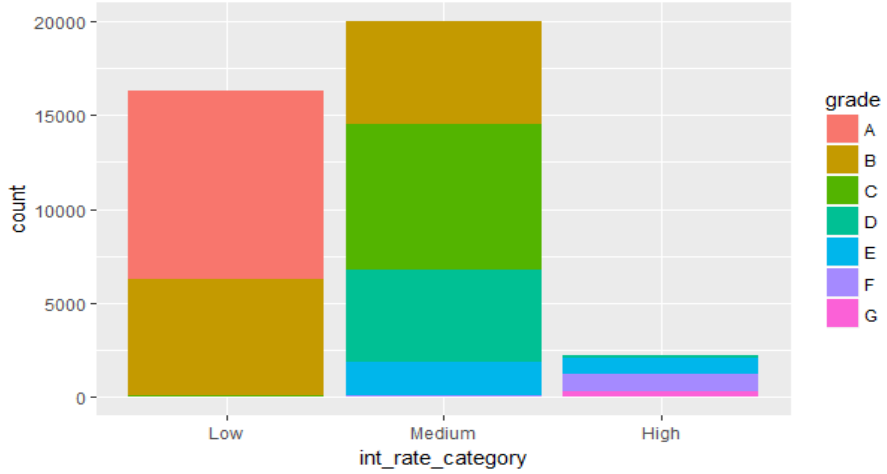




- Hire dti range higher amount of default.
- With respect grades we can see that as grades goes from A to G, people making defaults more. This variable is affecting the more.
- Where there is no significant changes in defaults with respect to home ownership.

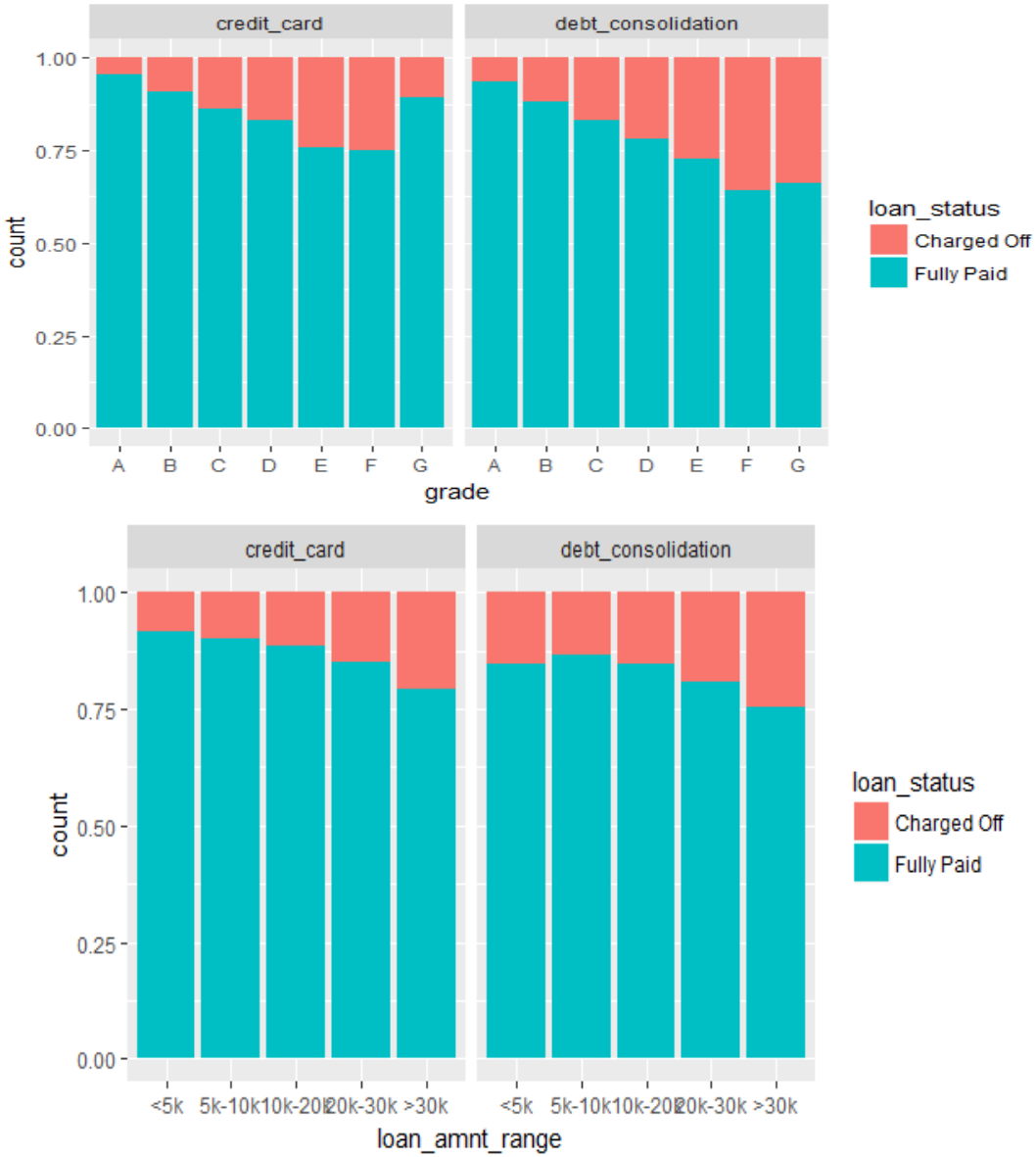
Bivariate Analysis Contd..





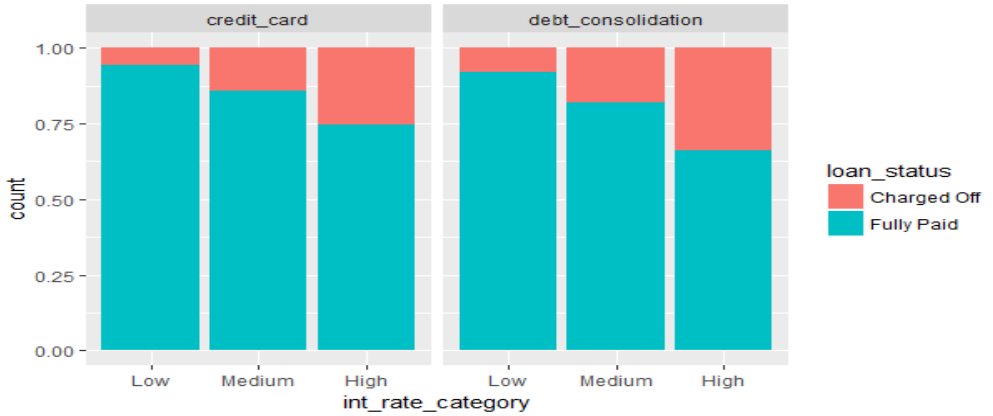
- By the Purpose graph we can clearly see that people taking loan for small businesses tend to default more.
- And from beside graph we can conclude that as there is a change in grade from A to G, the interest rate increases.

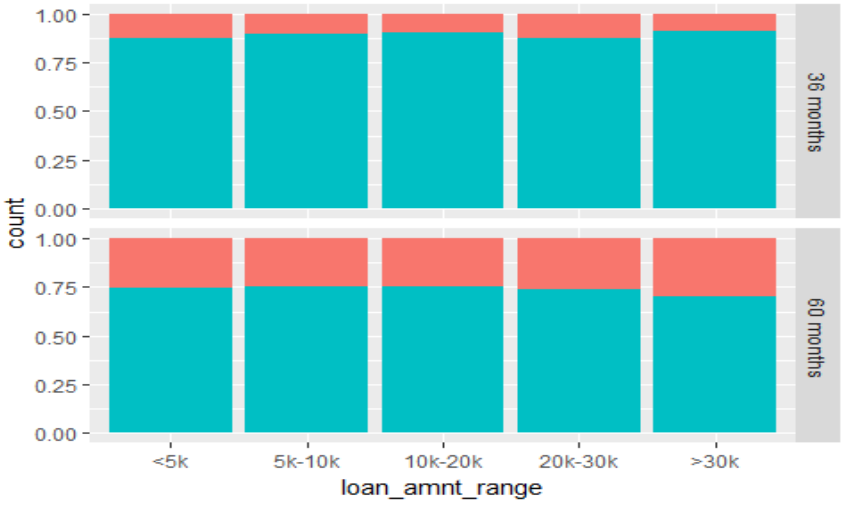
Multivariate Analysis



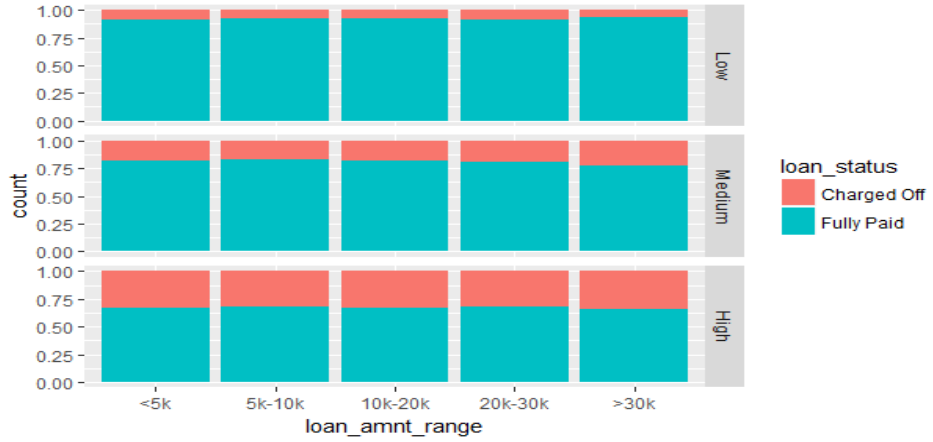
- Taking “Credit card” and “Debt Consolidation” into consideration as n of borrowers are more we have analysed behaviour of default changes across grades, loan amount and interest rate.
- In grades we can see that we can see that grade E & F are commonly affecting.
- Where as in loan amount people with more loan amount tends to default more.
- With interest rate for the both the purposes pattern is quite similar, which we can see in next slide.

Multivariate Analysis Contd….





- so from beside graphs we can conclude that more the loan amount, more the interest rate more likely to default. But pattern across loan range remains same for constant interest rate category.
- people with 60 months term and more loan amount tend to default more. But pattern across loan range remains constant for the each term individually.



Recommendations

- So to reduce people defaulting we can say that consumer finance company can cut down the interest to some margin so that borrower wont end up making default and company also wont loose business.
- Company should wisely think before lending money for the people whose grades are E & F as they are more likely to default.
- Also people with small business tend to make default more so lending money to such people should be decided on other parameters like income, grade, interest rate and also the loan amount.