Lending case Study

Problem Statement

- You work for a **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision 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

Below is the top 5 rows of the dataset .

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_p
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	

Shape of the data set

(39717, 111)

Duplicate Records

NUMBER OF DUPLICATE RECORDS ARE = 0

Null Values Check on the DataSet:

- Few Columns contains the null values, so we analysis, check weather these values are useful or not.

- If these columns affects the analysis then we have to keep them and treat them to get the accurate result or if
- These columns are of no use or did not play a role in impacting the analysis then we will remove it .

```
id                                0
member_id                         0
loan_amnt                         0
funded_amnt                       0
funded_amnt_inv                   0
                                ...
tax_liens                        39
tot_hi_cred_lim               39717
total_bal_ex_mort             39717
total_bc_limit                39717
total_il_high_credit_limit    39717
Length: 111, dtype: int64
```

- Now we have 48 columns out of which some correspond to the post approval of loan
- We are analyzing the user details and the driving factors of loan defaulting before approving loan.
- So we can safely remove the columns / variables corresponding to that scenario.
- Also there are some columns such as "id", "member_id", "url", "title", "emp_title", "zip_code", "last_credit_pull_d","addr_state".
- The above features or columns doesnt contribute to the loan defaulting in any way due to irrelevant information. So removing them.
- "desc" has description (text data) which we cannot do anythhing about for now. So removing the column.
- "out_prncp_inv" , "total_pymnt_inv " are useful for investors but not contributing to the loan defaulting analysis. So removing them.
- "funded_amnt" is not needed because we only need info as to how much is funded in actual. As we have "funded_amnt_inv" , we can remove the earlier column.

Employment Length

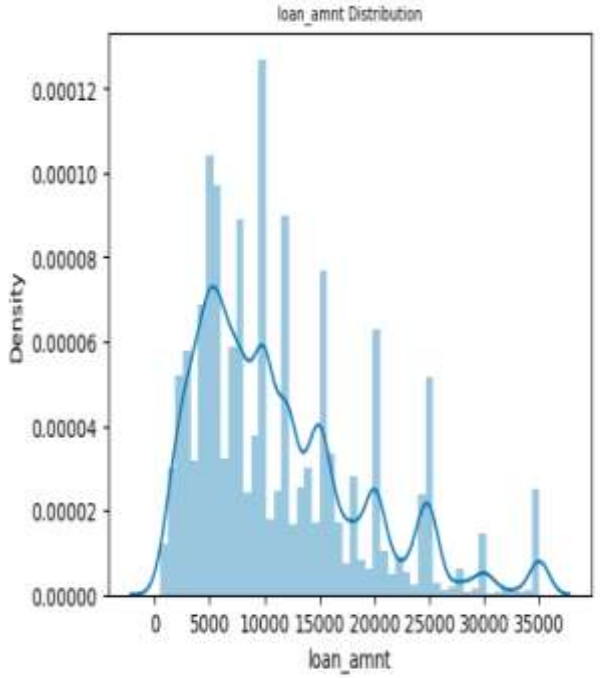
```
emp_length
10+ years    8879
< 1 year     4583
2 years      4388
3 years      4095
4 years      3436
5 years      3282
1 year       3240
6 years      2229
7 years      1773
8 years      1479
9 years      1258
Name: count, dtype: int64
```

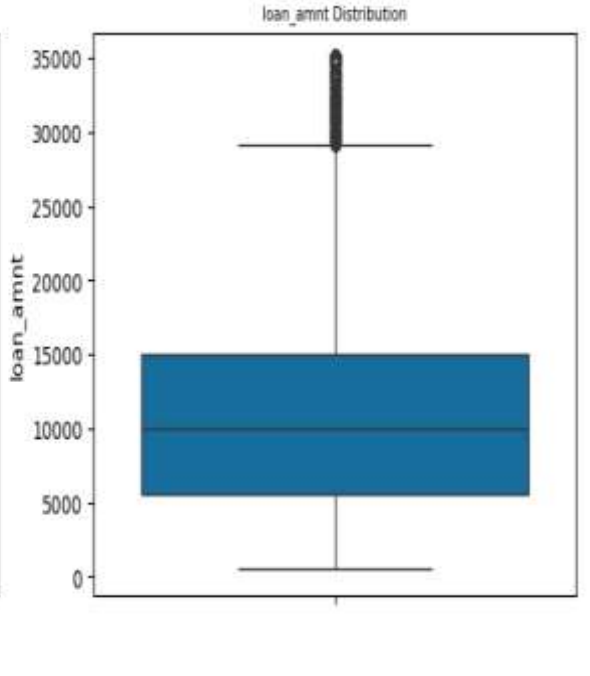
It is clear from above table that more than 8K having the employment length more

Inference using the graphs and plots

- Above plots represent the distribution of the loan amount:

- Mostly the distribution is Right skwed.
- And in boxplot there are too many outliers.
- As the loan amount plays the important role for further analysis so we will do outlier treatment in future .

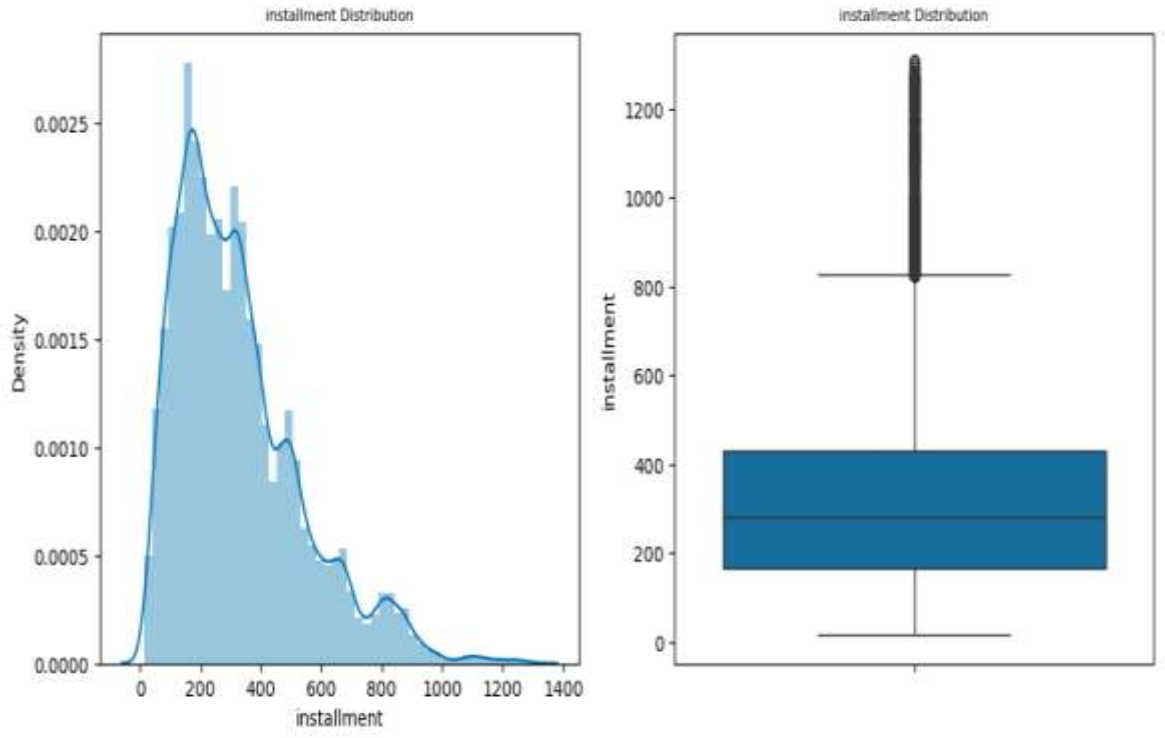




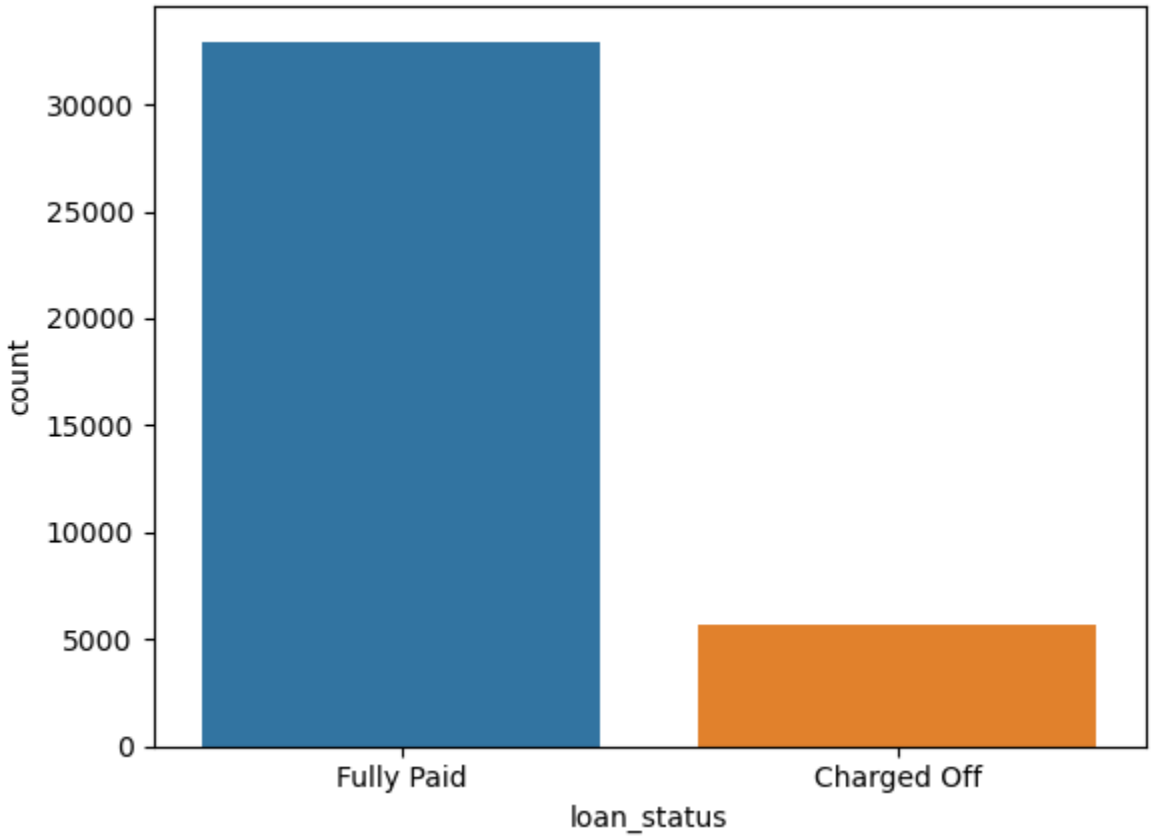
- Above picture shows the installment distribution here we can find that it is also right skwed and there are too many outliers present in the distribution and it will also require outlier treatment .

- By doing the above analysis on the continuous variable we conclude that many columns requires the outlier treatment here are few columns that are important and require outlier treatment
- total_acc, pub_rec, funded_amnt_inv, dti

- After Outlier treatment we get the values with in the range.



Loan Status :
The goal of the analysis is to see who is likely to default and this can only be said in case of either fully paid or charged off loans.

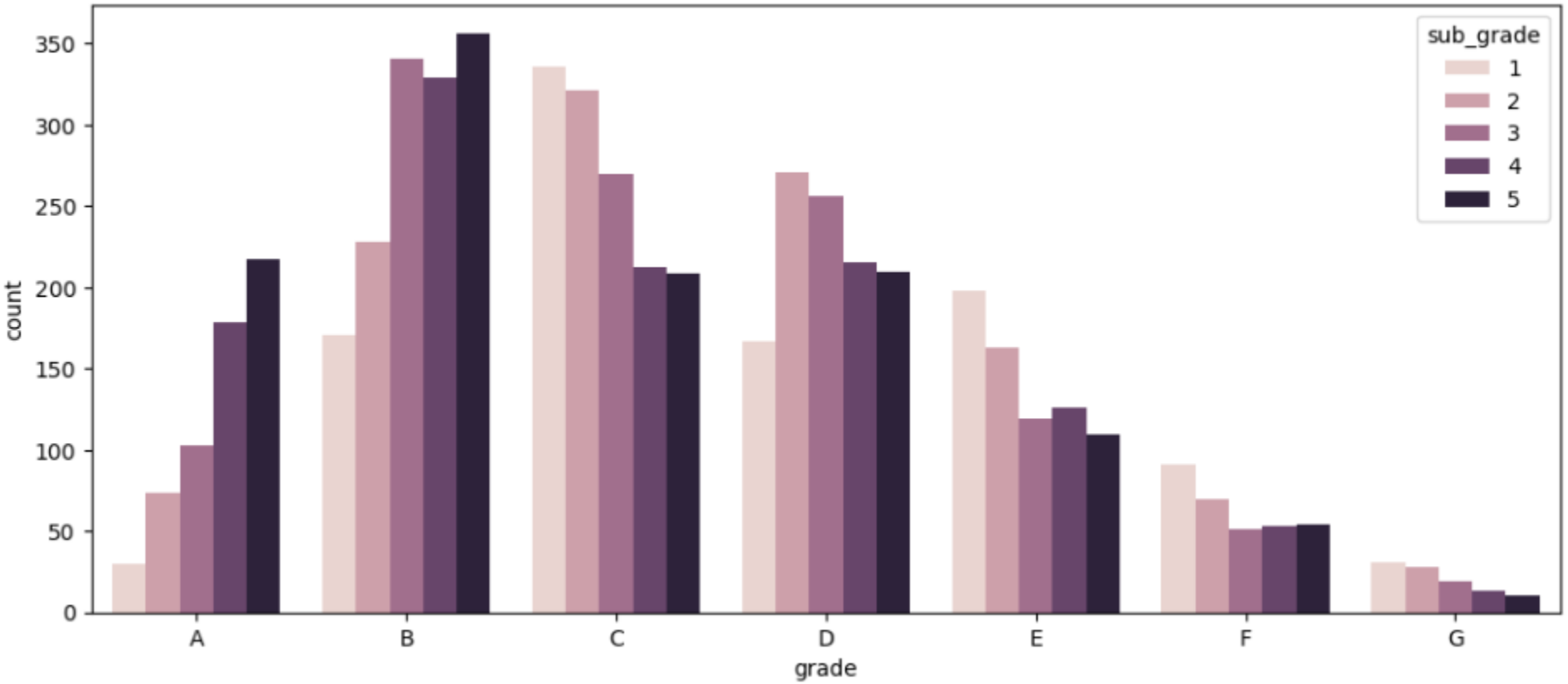


We remove the records with status current loan.

From the graphs it is clear that most of the loan is fully paid .

As we compare both approx. 15% is charged off as compared to fully paid loan

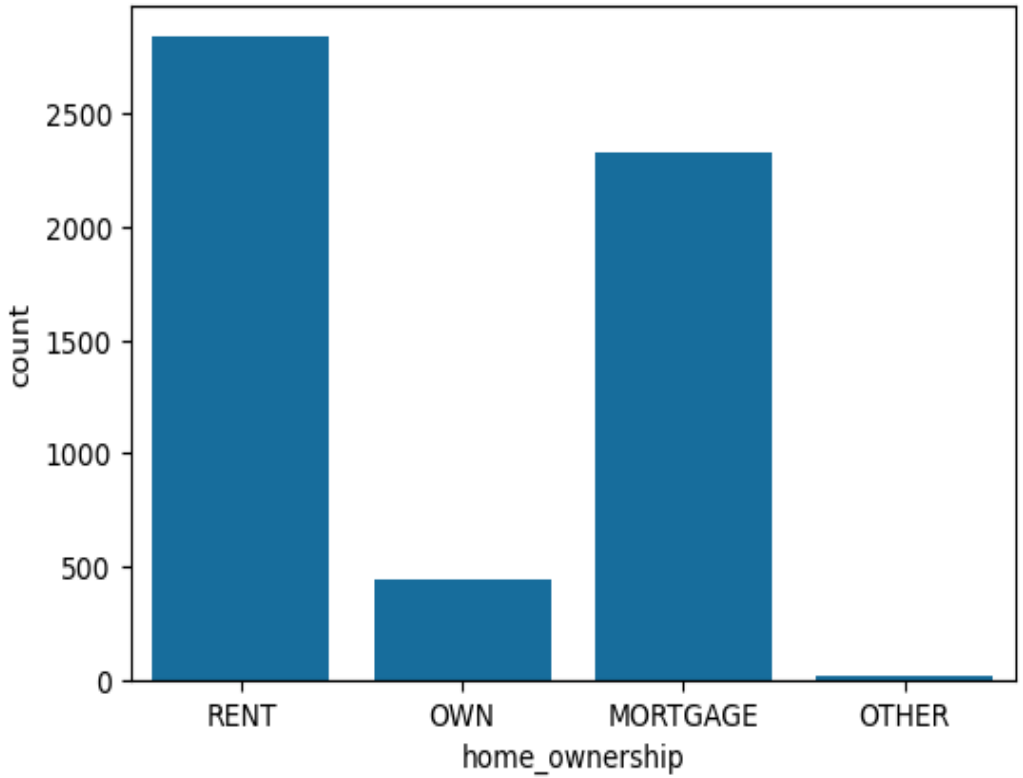
Grade and subgrade count of the loan status



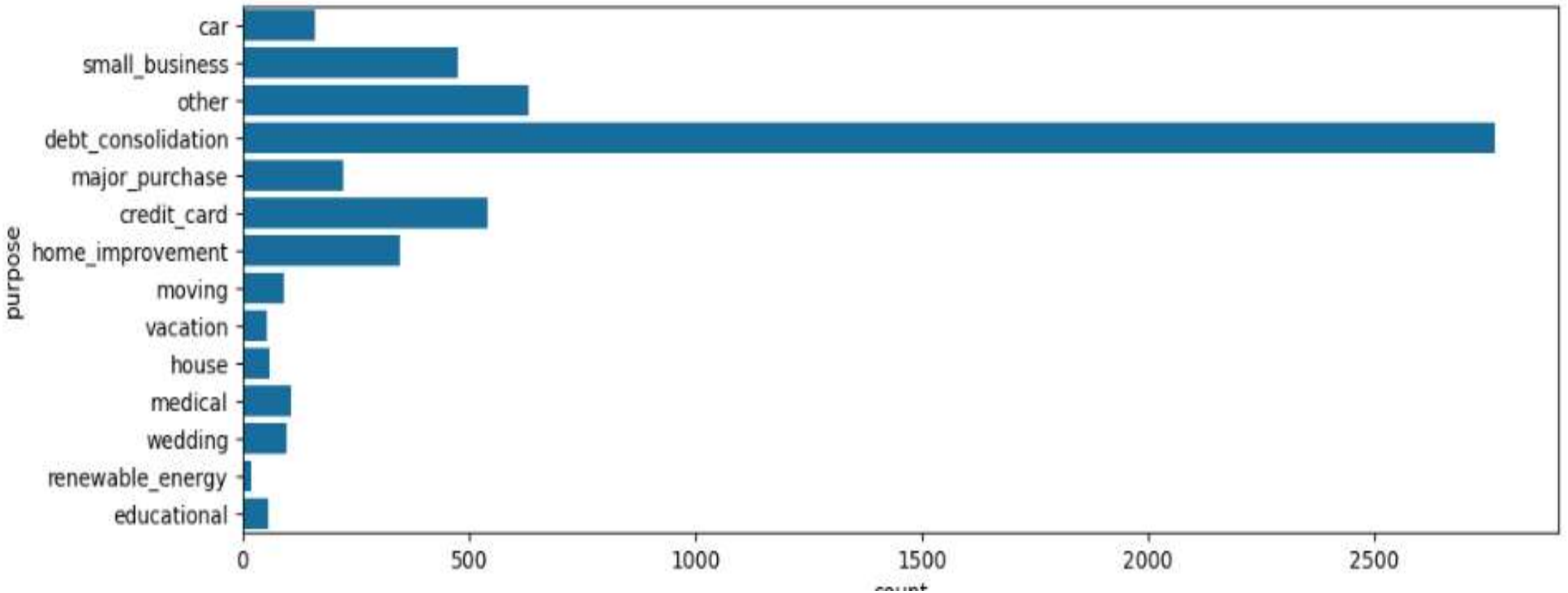
In the above graph is is shown that Grade B have the highest value count and in grade B 5 subgrade have the highest count

Ownership Count

- Form the graph we conclude that most of the loan taker are the rented person.
- Owners count is below 500
- Mortgage is the second highest in count



Purpose

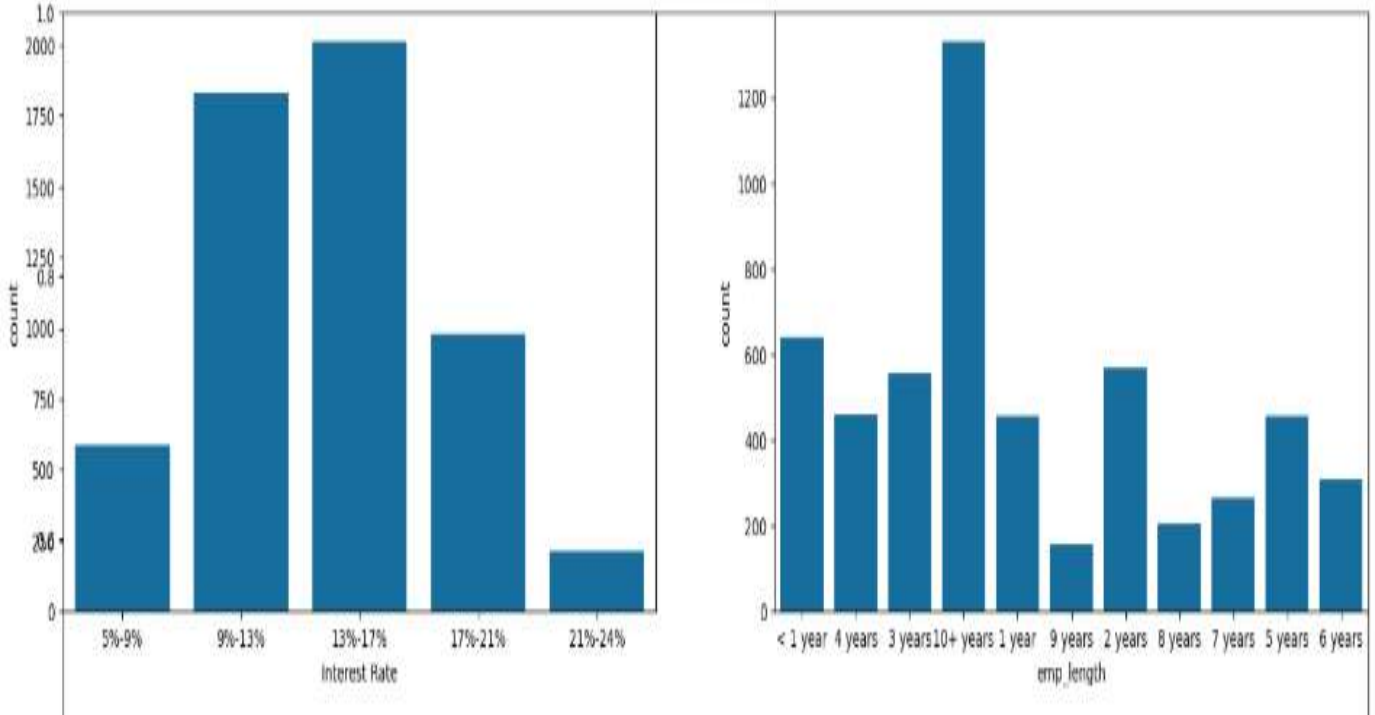


Majorly loan is taken to repay the multiple existing debt (debt consolidation)

Interest Rate Range and the employment length

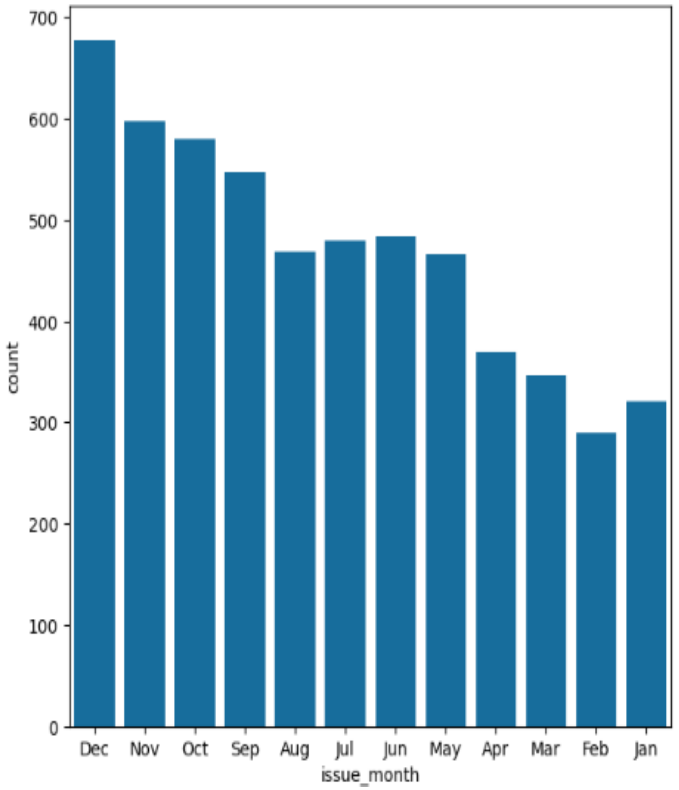
- As we can see the most of the lo given to customer are of range 13-17%

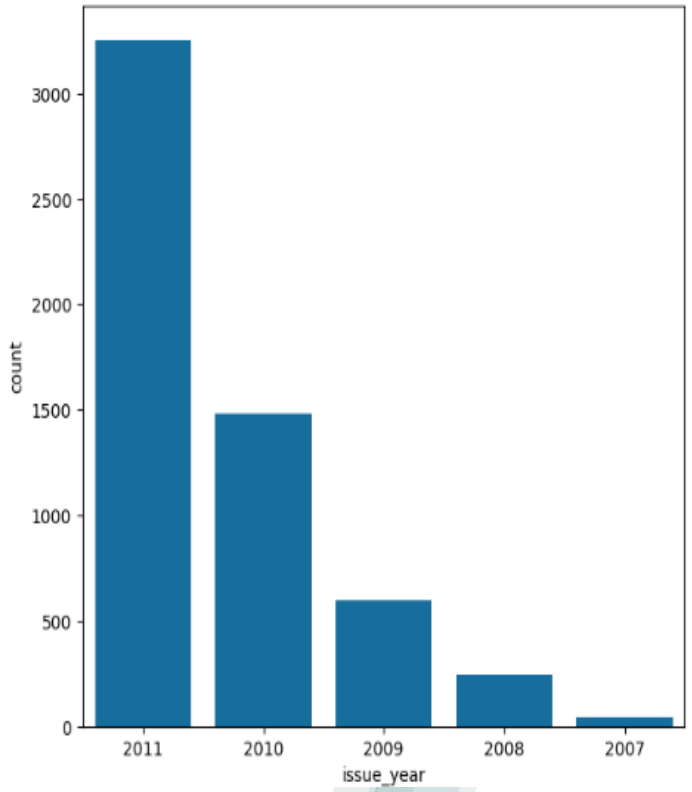
- And the most of the person are having more than 10 years of employment.



- Most of the loan taken by bank on December.

- In 2011 maximum person took the loan.

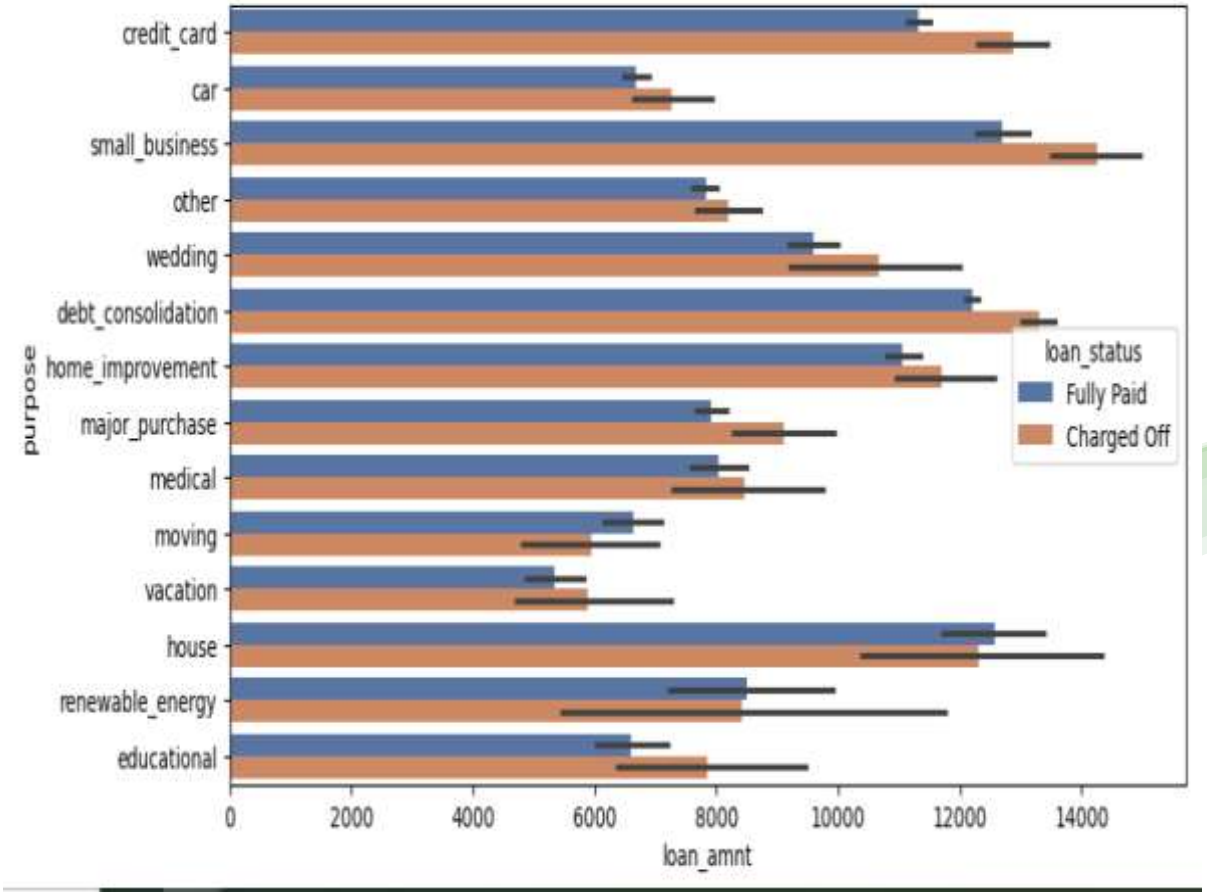




Reason to take loan and paid off by customer analysis

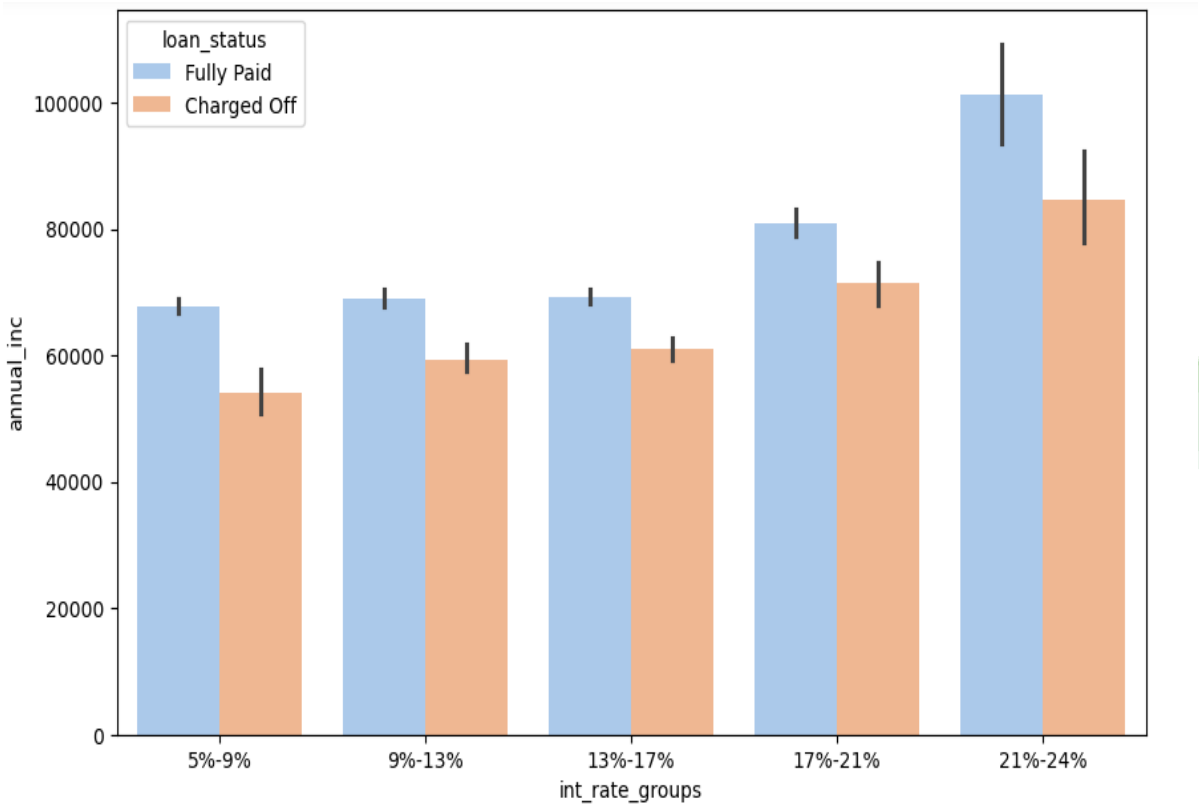
- As we can see from above plot it is clear that most of the loan amount taken for small business and the major part of the loan is charged off by the company.

- Most of the person took the loan for house had clear the loan on time.



Loan status with respect to interest rate

- Form above plot we conclude that major part of loans are fully paid in every interest rate range but in 21-24% maximum number of person paid off the loan



Tenure of the loan vs loan amount

- There are major loan taker took the tenure of 60 month and the there is sight difference in the loan fully paid and charged off.

- Just half of the loan amount is taken for 36 month tenure with respect to 60 month tenure