

Gramener Case Study

SUBMISSION

Group Name:

1. Anusha Kulkarni
2. Ramya sri sai Ch
3. Rudra sen
4. Husn Ara

Data Exploration

Data set contains total 111 columns but all these columns are not useful for analysis, so we need to do data refinement .

- After checking the null values we have removed the columns which are have $>70\%$ of null .
- We found what are the variable used for analysis
 1. Categorical variables
 2. Quantitative variables

Business Objective:

Is to identify the risky applicants, to reduce largest source of financial loss(called credit loss) to company.

Based on the business objective we should focus on Categorical variables and Quantitative variables

Variables Identified For Analysis

- Loan Status
- Term
- Interest rate
- Grade
- Sub Grade
- Employee Length
- Home Ownership
- Verification
- Annual Income
- Issue Date(only year we used)
- Purpose
- dti

Assumptions

- As current loans have not yet been completed we decide to ignore them during analysis.
- Outliers have not been considered during analysis.

Data Cleaning and Manipulation

Possible Data Inconsistencies :

1. Checking for duplicates on member id.
2. Cleaning the n/a values in the columns we identified that null values >70%
3. Dropping unnecessary columns which are not used for analysis.

Other Issues :

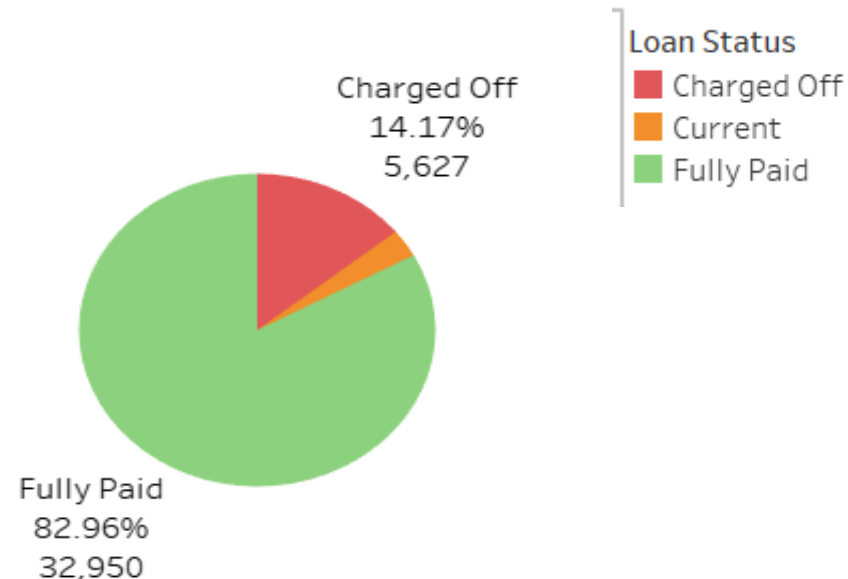
1. Converting the data type of int_rate column to float as it contains % in the cells
2. In employee length we should extract only the integers
3. Converting the data type of issue date to date time and extract only year

Analysis

From the data, observed that

- The defaulters are of 14.17% for Fully Paid, Current and Charged Off values.
- Whereas 15% for Fully Paid and Charged Off.

Using EDA have to identify ,the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.

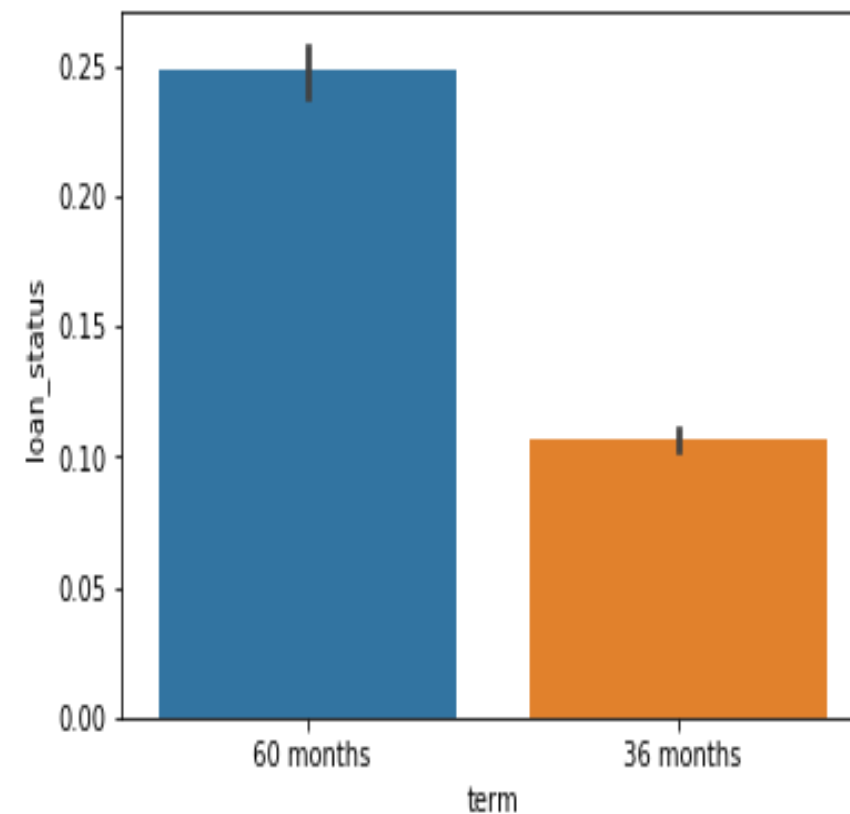


Term Analysis

Observations from term Analysis

- Defaulters ratio is more in 60 months term than 36 months

Count of loan_status	Column Labels				
Row Labels	Charged Off	Current	Fully Paid	(blank)	Grand Total %
36 months	3227		25869		29096 11.09087
60 months	2400	1140	7081		10621 22.59674
(blank)					
Grand Total	5627	1140	32950		39717



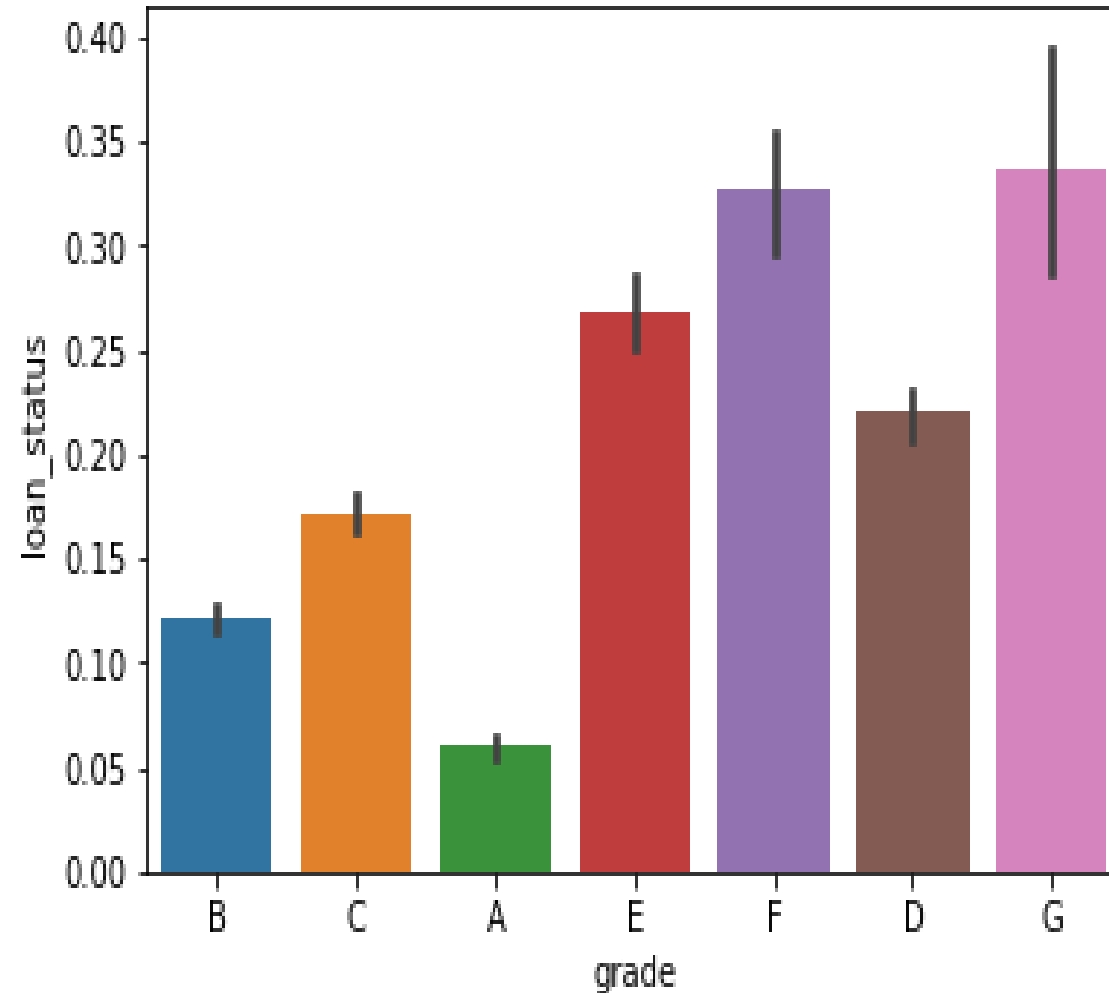
Grade Analysis

Observations from Grade Analysis

- The percentage of defaulters are more in Grade G

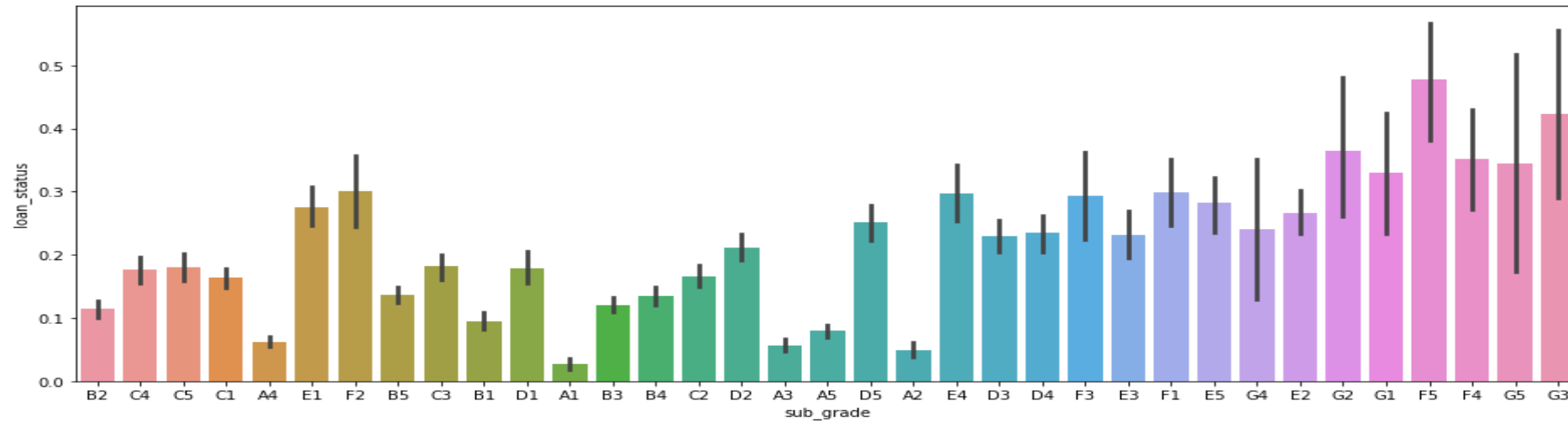
Verification Analysis

- Also observed that the percentage of bad loans not verified customers are also fall under grade G



Sub Grade Analysis

- Further analysis in Grade-G, identified that sub-grade F5 is more is count.



Conclusion :

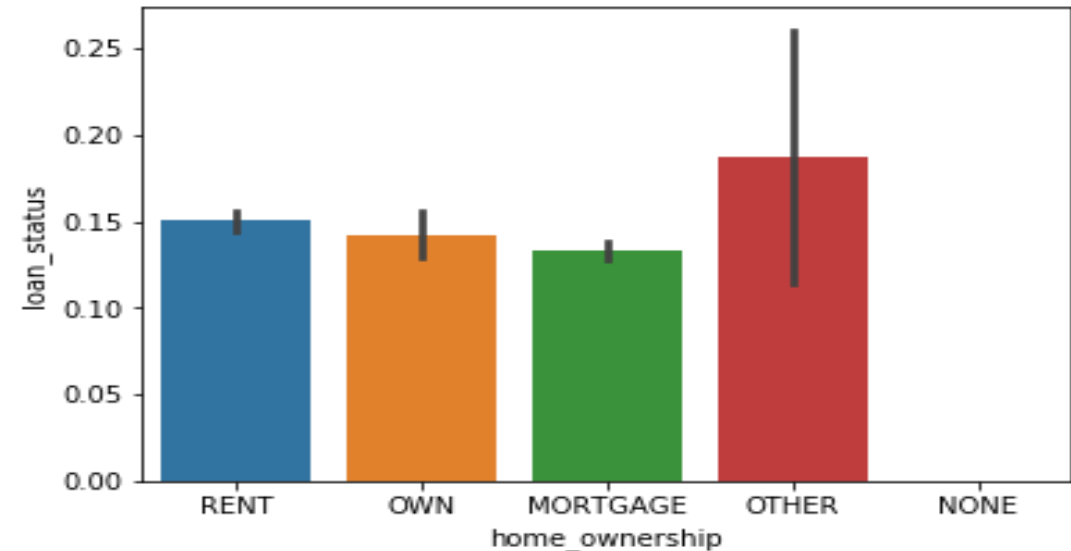
Sub Grade F5 under Grade G having more defaulters and indentified as not verified customers falling under this category.

Home Ownership Analysis

Observed that the more defaulters are falling under others ownership

Conclusion:

The company has to take more precautions while sanctioning loans to others

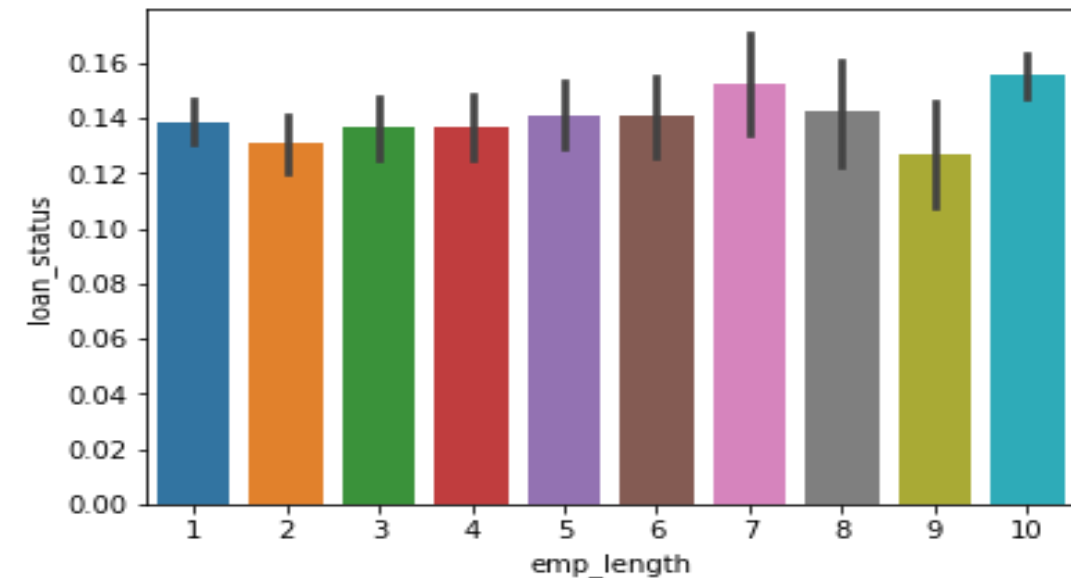


Employee Length Analysis

Observed that the more defaulters are having greater than 9 years experience.

Conclusion:

The company has to take care in providing loans to more than 9 years experience.

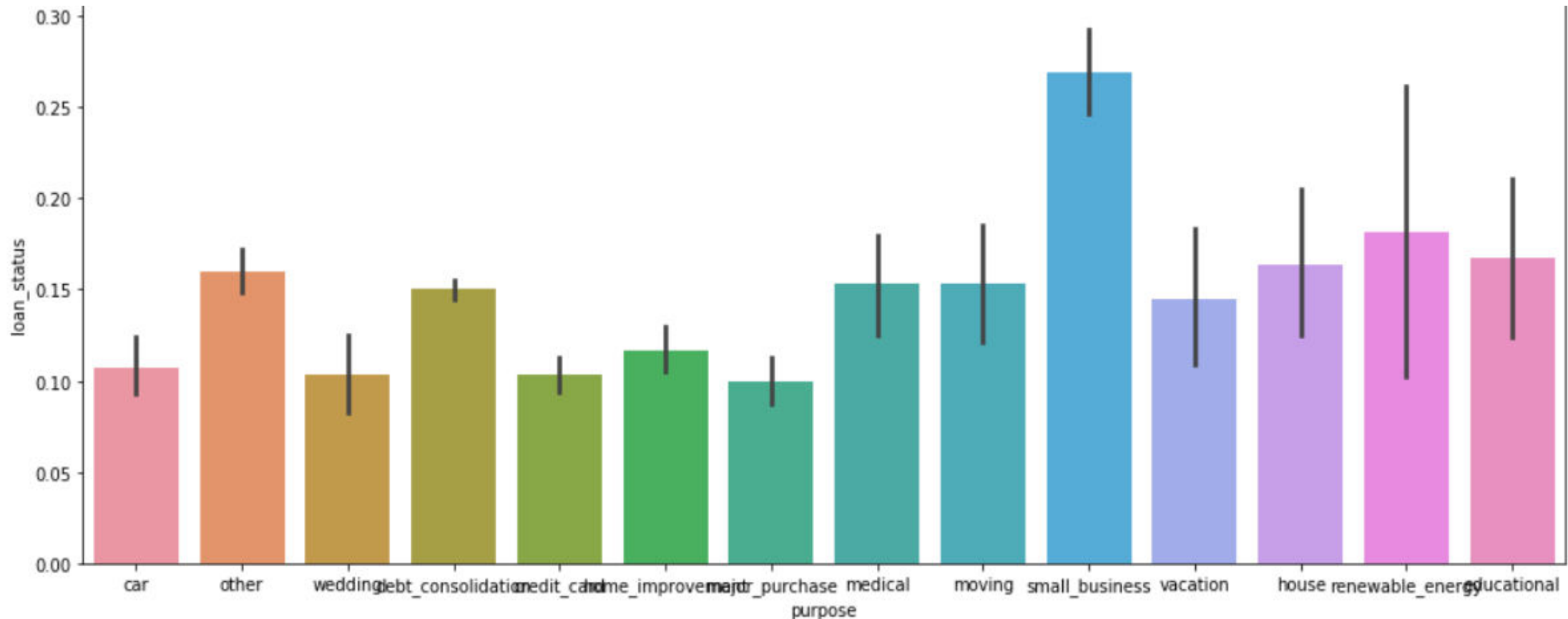


Purpose Analysis

Observed that the more defaulters are falling under customers having purpose with small business

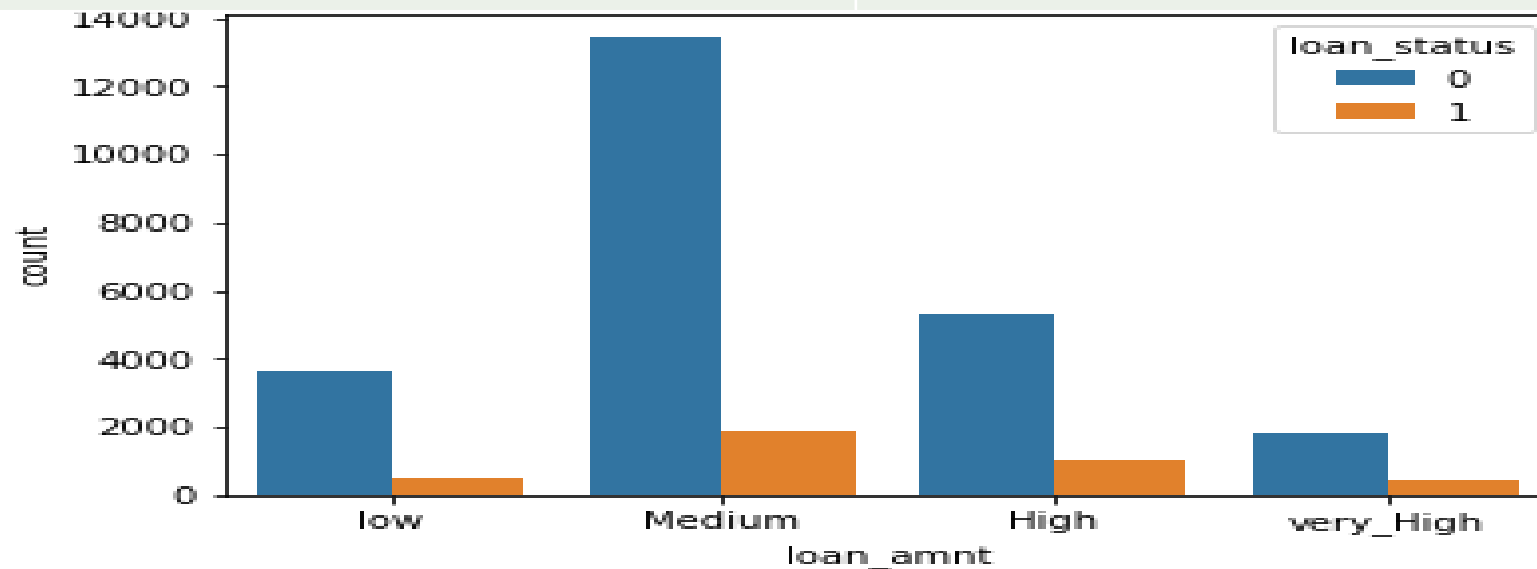
Conclusion:

Company can avoid loans to small business, purpose to reduce the amount of credit loss.



Binning loan amount into categories

LOAN AMOUNT	CATEGORIES
0-5000	LOW
5000-15000	MEDIUM
15000-25000	HIGH
25000-40000	VERY HIGH



The company should be careful while giving high loans

Conclusion:

Here are the suggestions company can utilise this knowledge for its portfolio and risk assessment.

- Company should take care in providing loans to customers with following behaviours :
 1. who are going for 60 month term.
 2. Falling under Grade G and A as sample ratio is varying
 3. Having Others house ownership.
 4. With more than 9 years experience.
 5. Having purpose of small business and debt consolidation.