

Bank Loan EDA

Gramener Case Study

Vijay Aditya T S

Problem Statement:

The case study given is related to risk analytics in the banking and financial services sector. Given the data, the objective of the process is to find out whether a customer is likely to repay his/her loan or default it. This will help the banks to understand data-driven decision making that can help to reduce the risk of losing money while lending to customers.

Strategies to be followed:

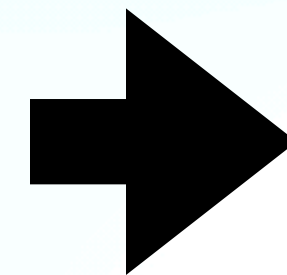
- Data Sourcing
- Data Cleaning
- Preparing the Data
- Visualising the results by plotting graphs
- Decision Making
- Identifying the factors which leads to loan default

Problem Solving Methodology

Stage 1

Data Sourcing, Data Cleaning and Preparation

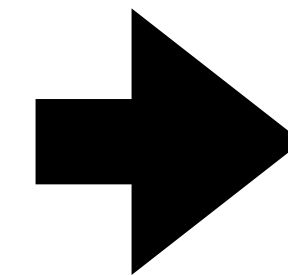
1. Read the data carefully from the source.
2. Clean data, if the dataset contains any null values or clean according to the necessity of the situation .
3. Prepare the data for analysis.



Stage 2

Data Analysis and Data Visualisation

1. Analyse the dataset by different charts / graphs by using:
 - Univariate analysis
 - Bivariate / Multivariate Analysis
2. Plot the outcome of data analysis through graphs and charts.

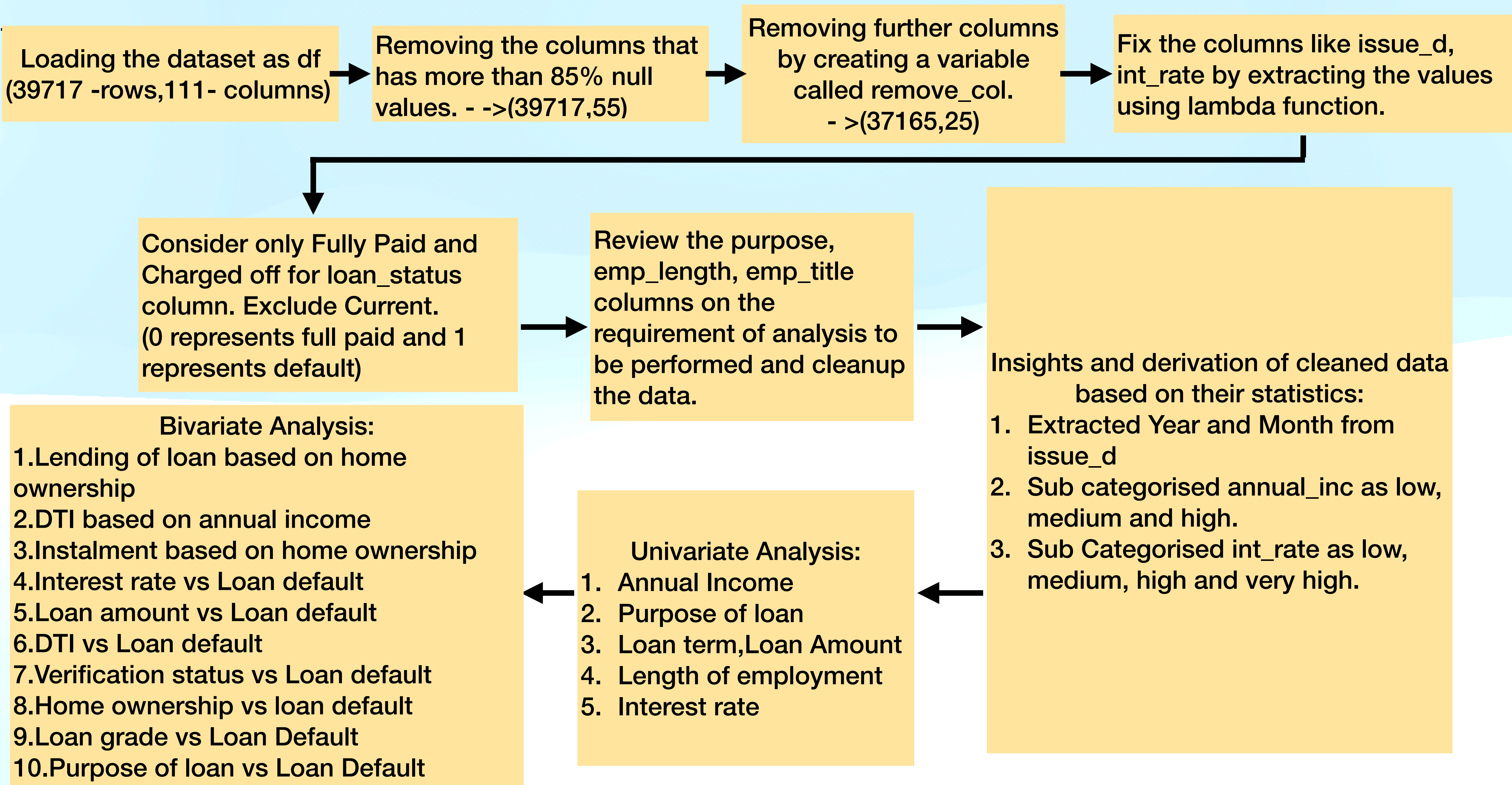


Stage 3

Decision and Report Making

In this stage, a data analyst has to prepare a detailed report on his findings with his data analysis.

Workflow of analysis



Assumptions

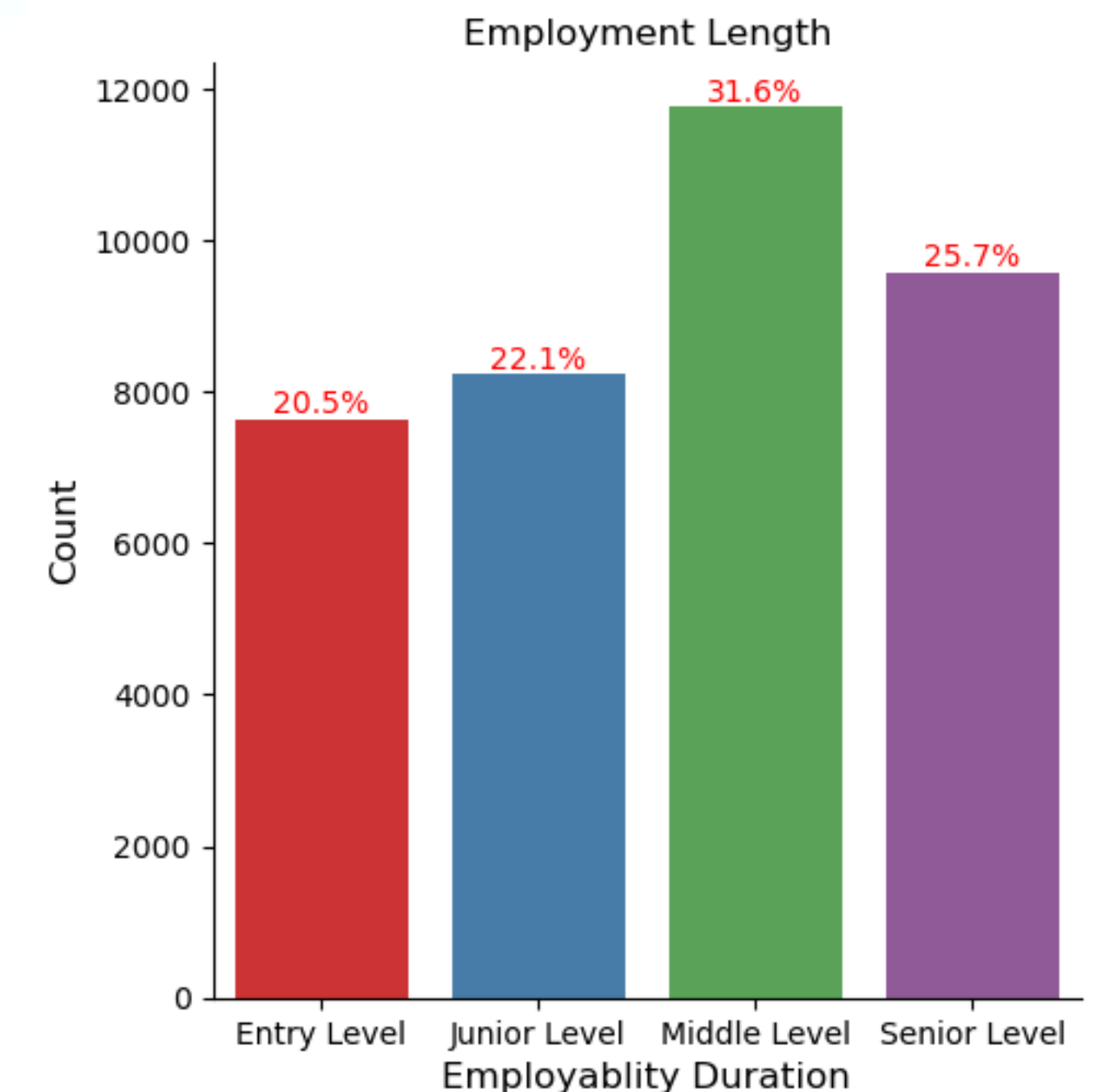
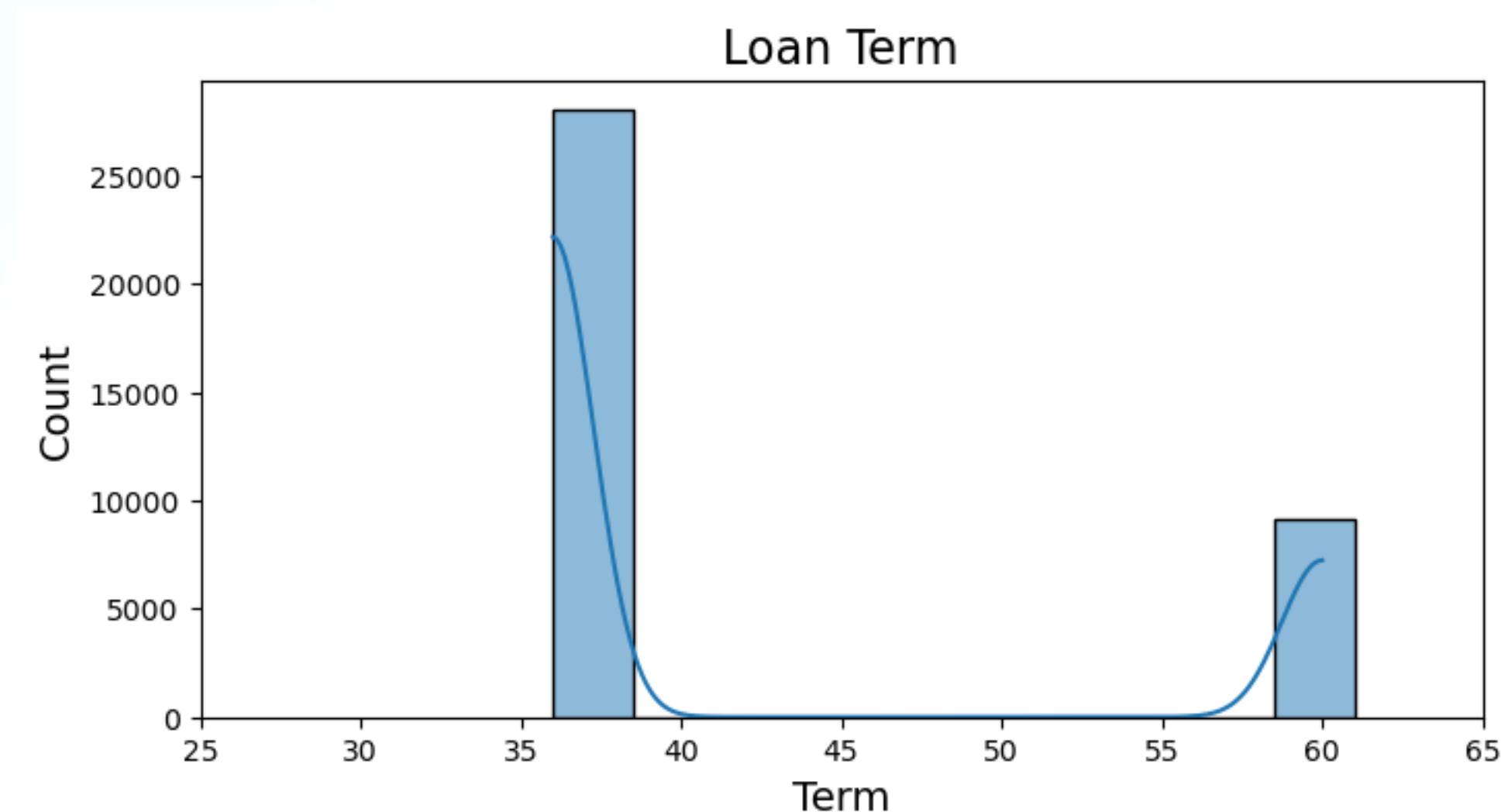
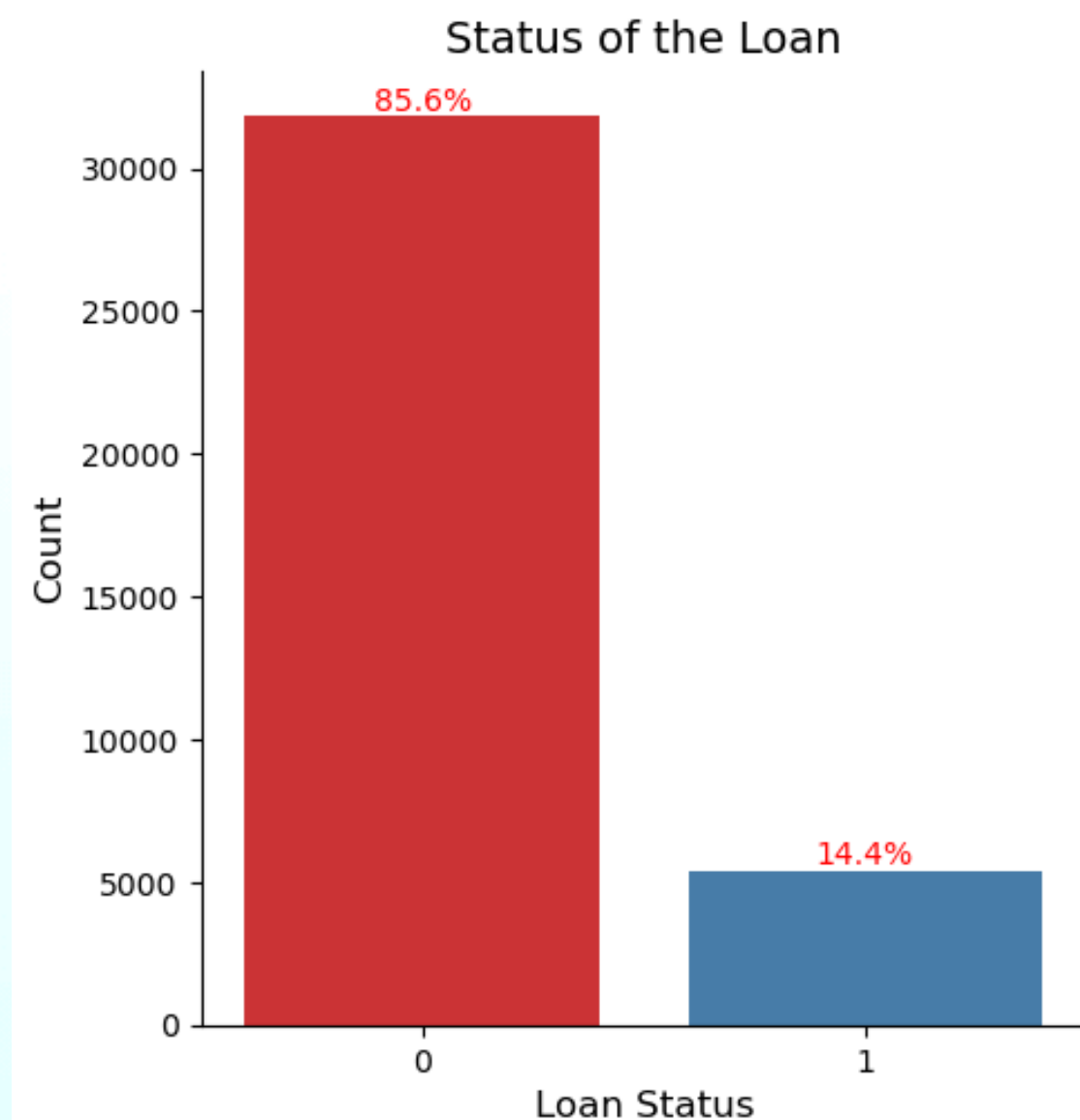
The following assumptions were made to categorise the following columns into sub categories such as the interest rate, loan amount, annual income, employment length and dti.

For interest rate:	For Debt to Income Ratio	For Annual Income
< 9% -> Low	< 8 .24 -> Low	<40906 -> Low
9% - 11% -> Medium	8.24 - 13.45 -> Medium	40906 - 59000 -> Medium
12% - 13% -> High	13.45 - 18.60 -> High	59000 -81600 -> High
>13 % -> Very High	> 18.60 -> Very High	81600> -> Very High
Loan Amount:	For Employment Length	
< 5400 -> Low	< 2 -> Entry Level	
5400 - 9800 -> Medium	2 - 3 -> Junior Level	
9800 - 15000 -> High	4 - 8 -> Middle Level	
> 15000 -> Very High	>9 -> Senior Level	

Analysis of Loan Status, Term and Employment Length

The inferences of the below graphs are as follows:

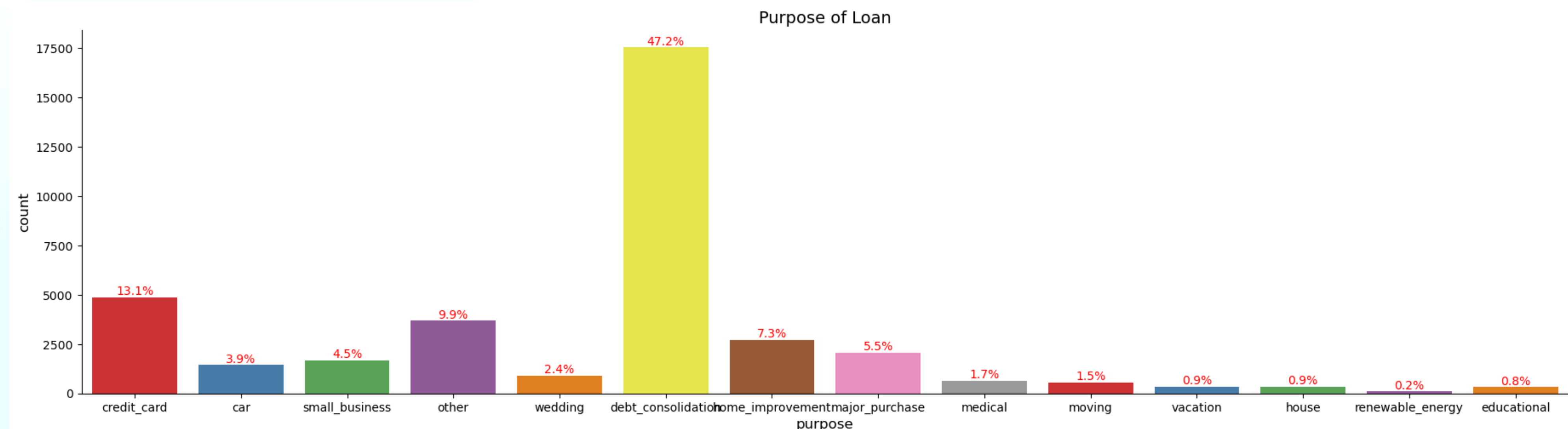
- As per the result of graph, 14.4% of customers have defaulted on loan.
- 3/4'th of the entire customers has opted for a loan term of 36 months rather than 60 months.
- The middle - level employed customer stands the highest with an percentage of 31.6 %. Entry level customers are slightly below junior level employees and bank should focus on increasing the entry level customers.



Analysis on Purpose of Loan

Inferences from the below graph are as follows:

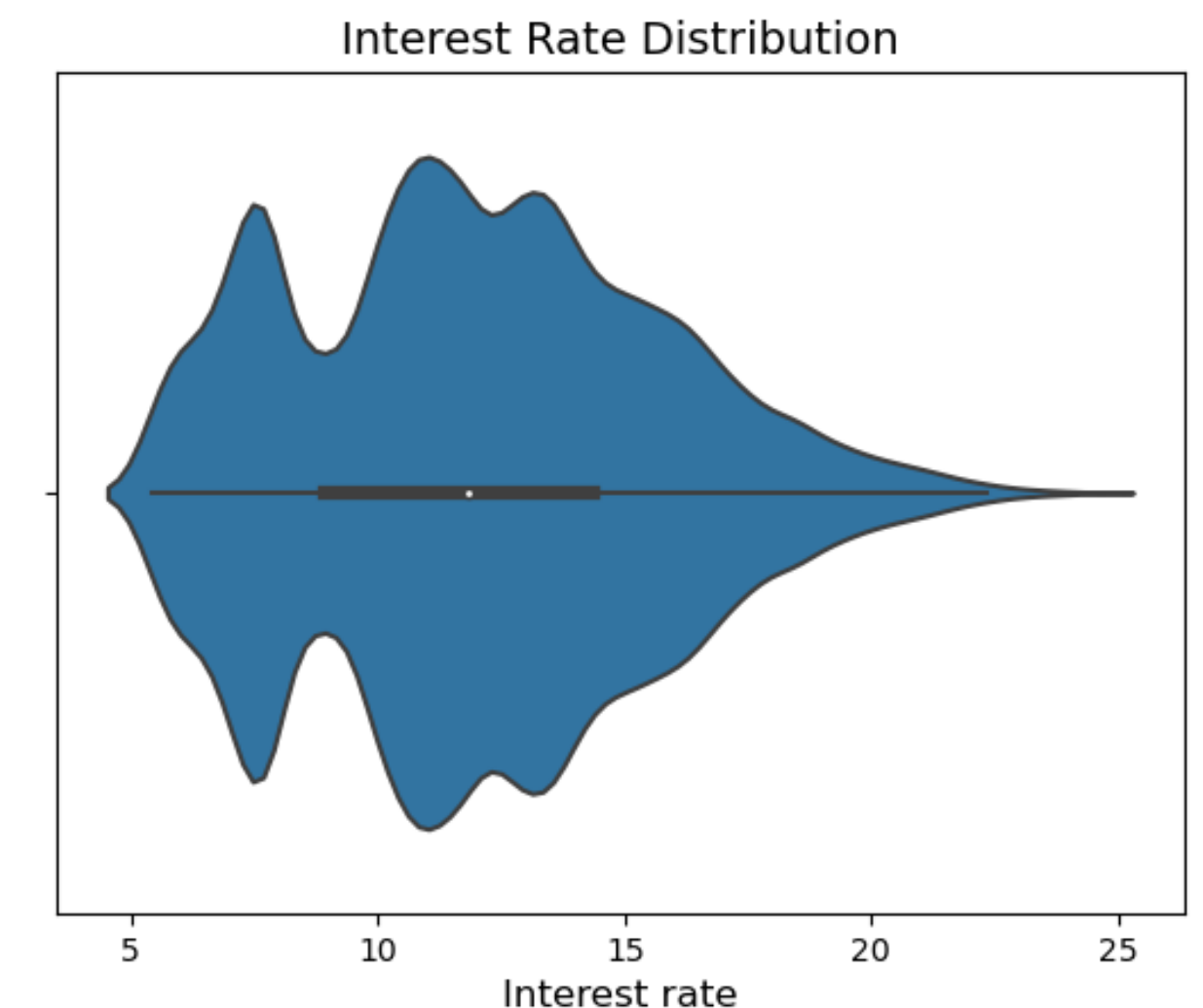
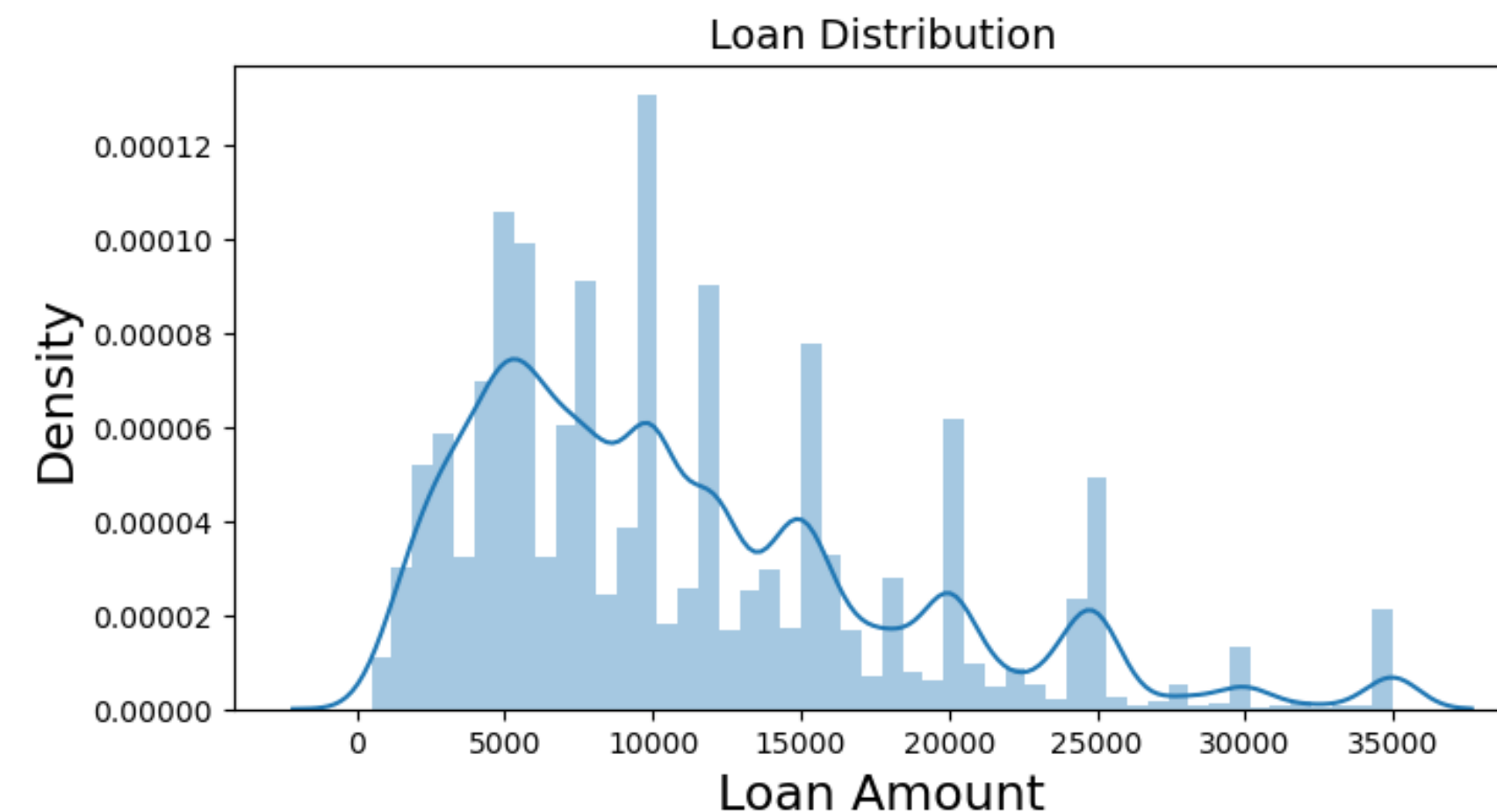
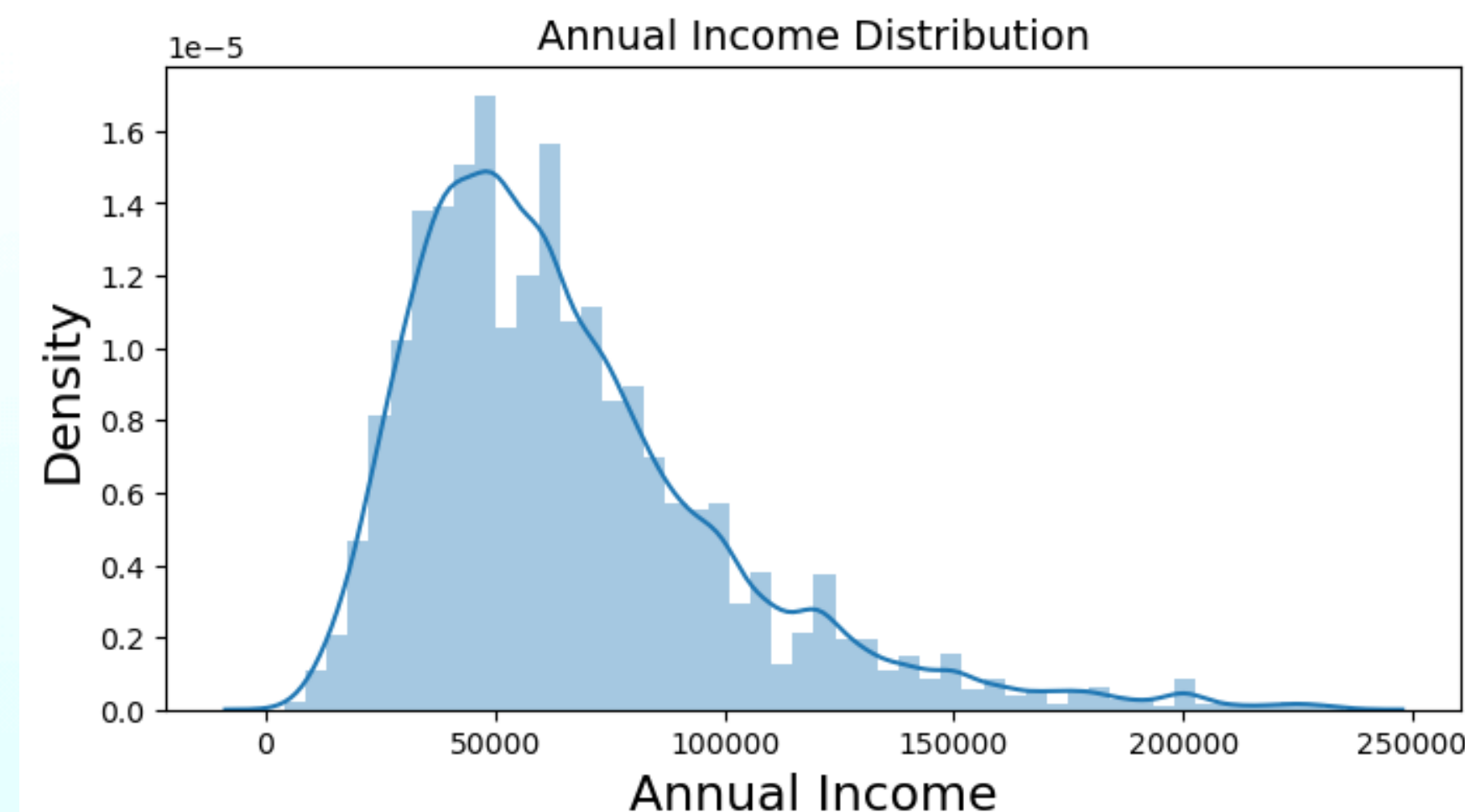
- Personal categories such as the educational, renewable energy, house and vacation are below 1 %.
- The majority of customers chose loan for their debt consolidation and credit card.
- The top 5 purpose are as follows:
 1. Debt consolidation - 47.2%
 2. Credit Card - 13.1%
 3. Home Improvement - 7.3%
 4. Major Purchases - 5.5%
 5. Small Business. - 4.5%



Analysis on Annual Income, Loan Amount and Interest Rate

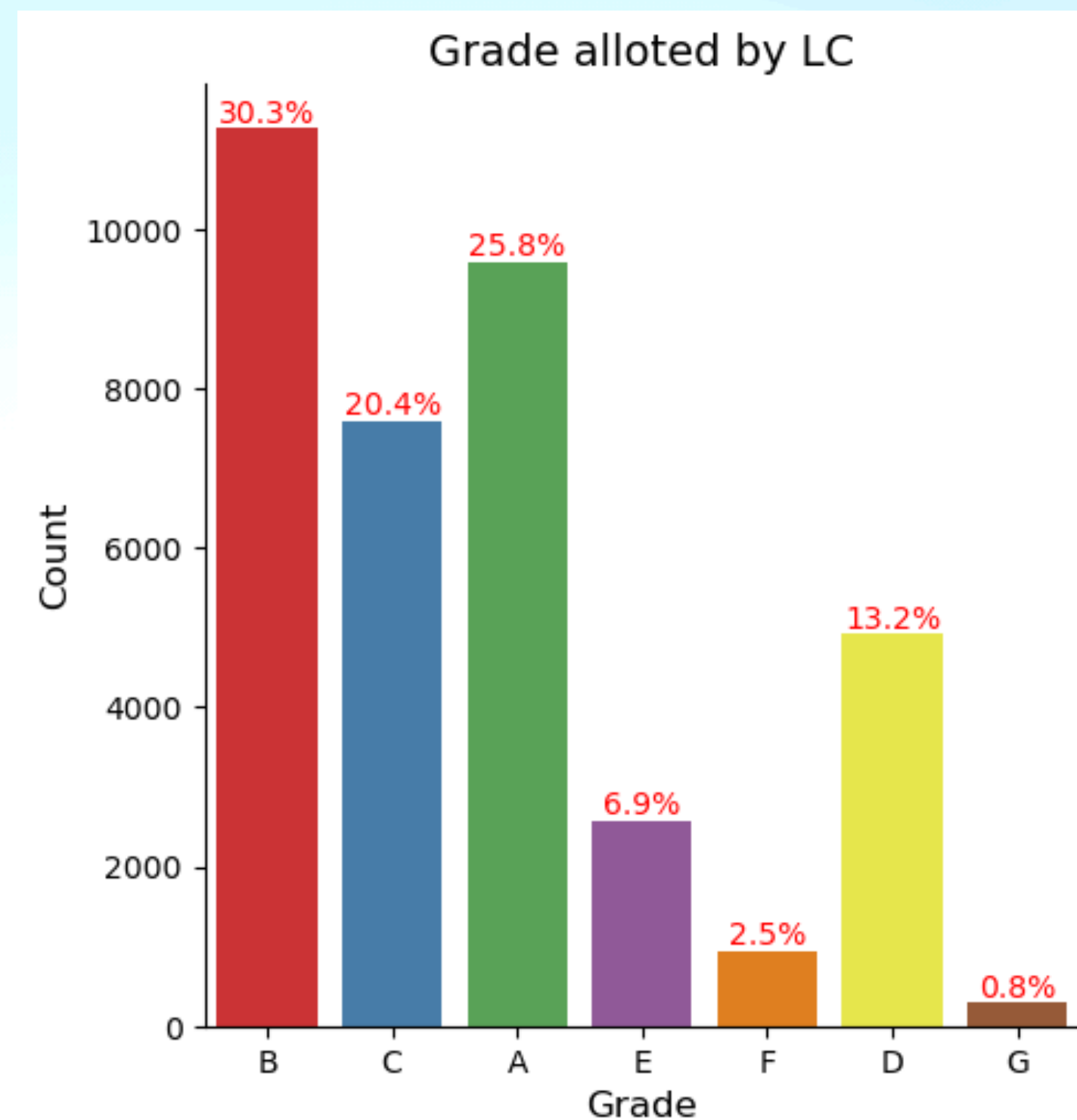
The inferences were drawn from the below graphs as follows:

- Majority of applicants annual income falls between 30000 USD to 80000 USD approximately .
- The loan amount is majorly spread between 4000 USD to 7000 USD approximately.
- The interest rate is spread between 8.94% to 14.35% approximately whereas majority of interest rate fall at 11.94%



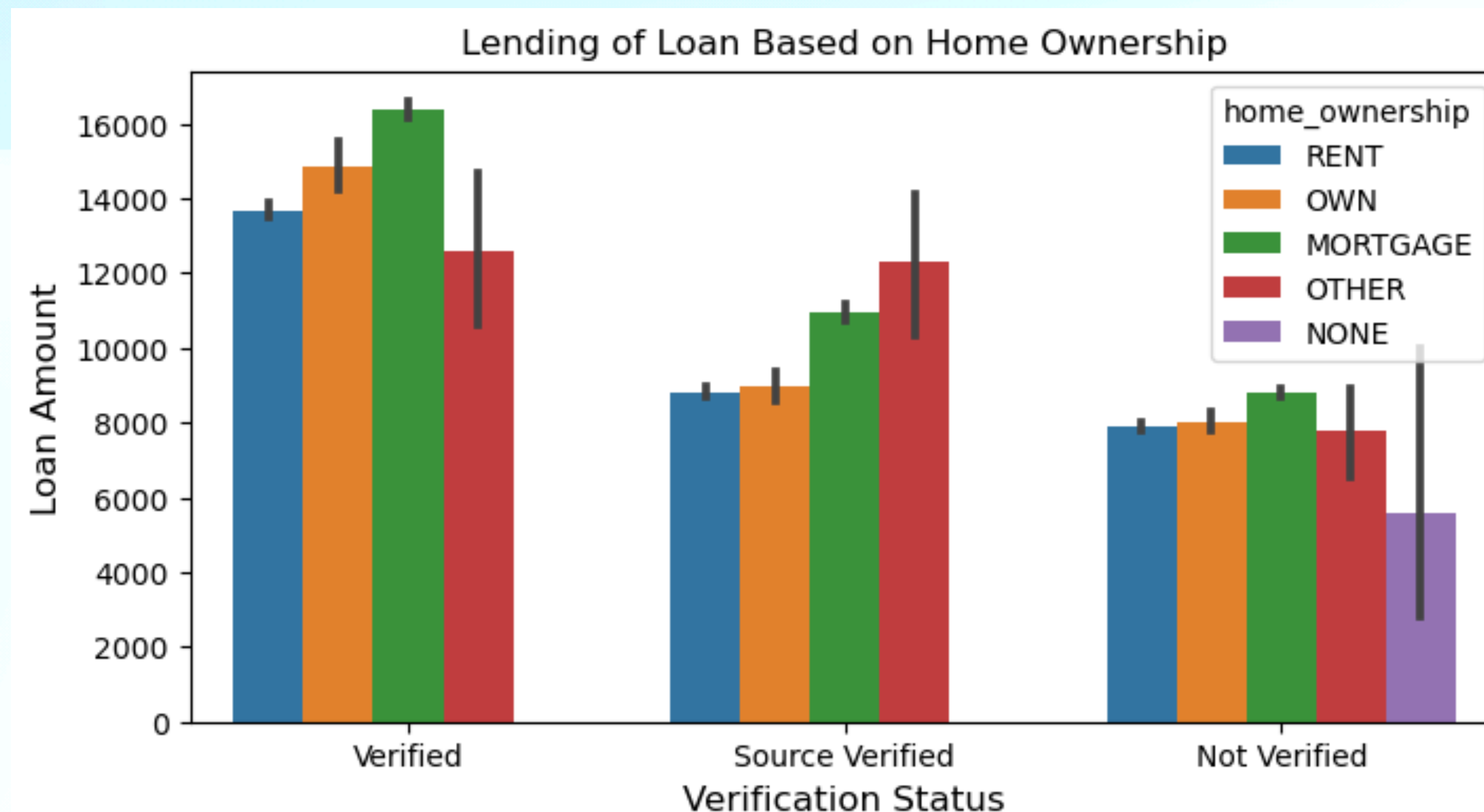
Analysis on Grade allotted by LC for customers

The below graph shows that majority of customers fall under Grade B with 30.3% and Grade A with 25.8%.



Lending of loan based on Verification and Home ownership

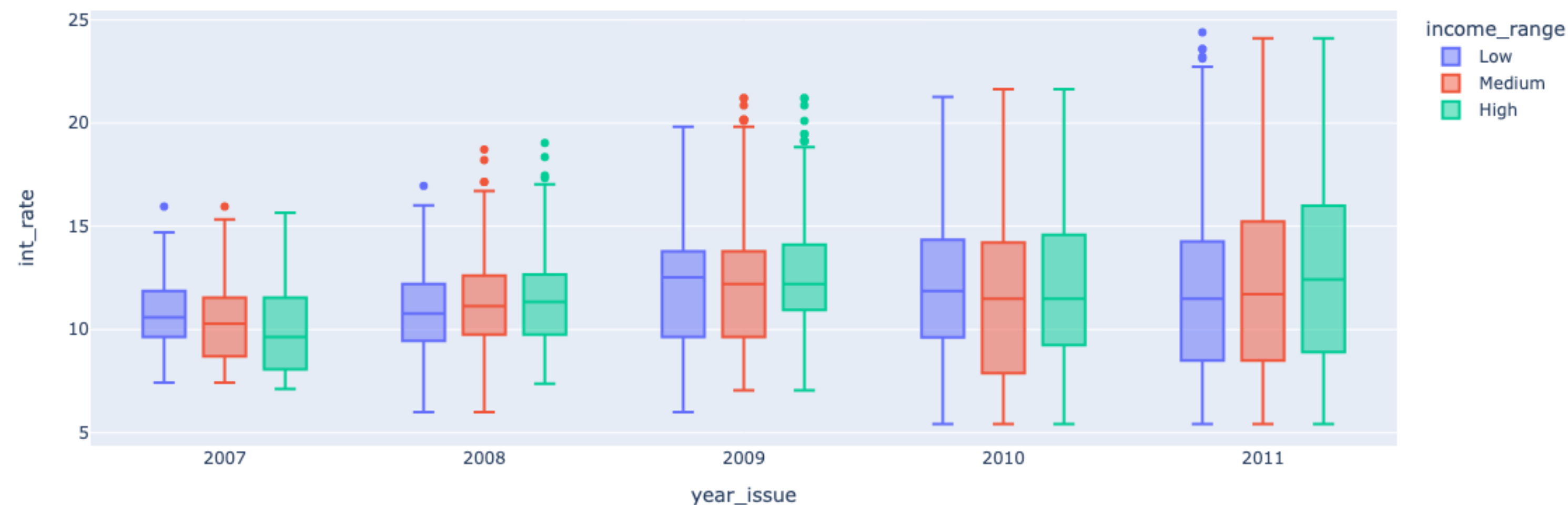
- The following graph below narrates that a customer with a verified status who has mortgaged will get highest loan disbursed.
- It is believed that the customers without home ownership and verification status, the chances are likely to receive funds very low or rejected.



Analysis of interest rate change every year

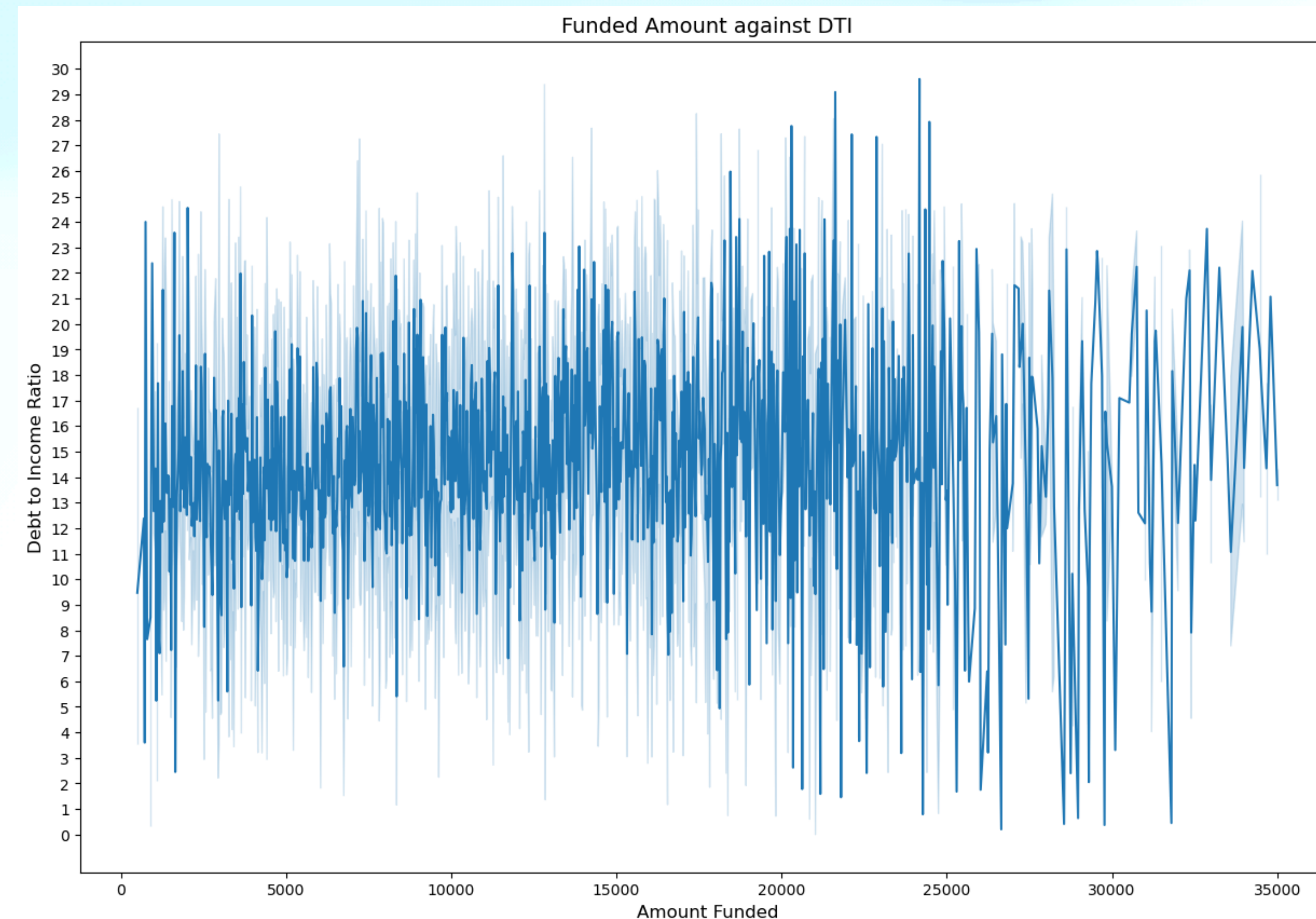
The below graph shows the statistical trend of interest rate changed over a period of time:

- 2010 and 2011 were the years, when interest rates were very low as compared to previous years.
- Though 2011 has low interest rate for the loan, the highest interest rate charged for a low income customer is 24.4%
- It is observed that from the year 2007, the interest rate has a steady growth with an increasing rate of interest for all income groups.



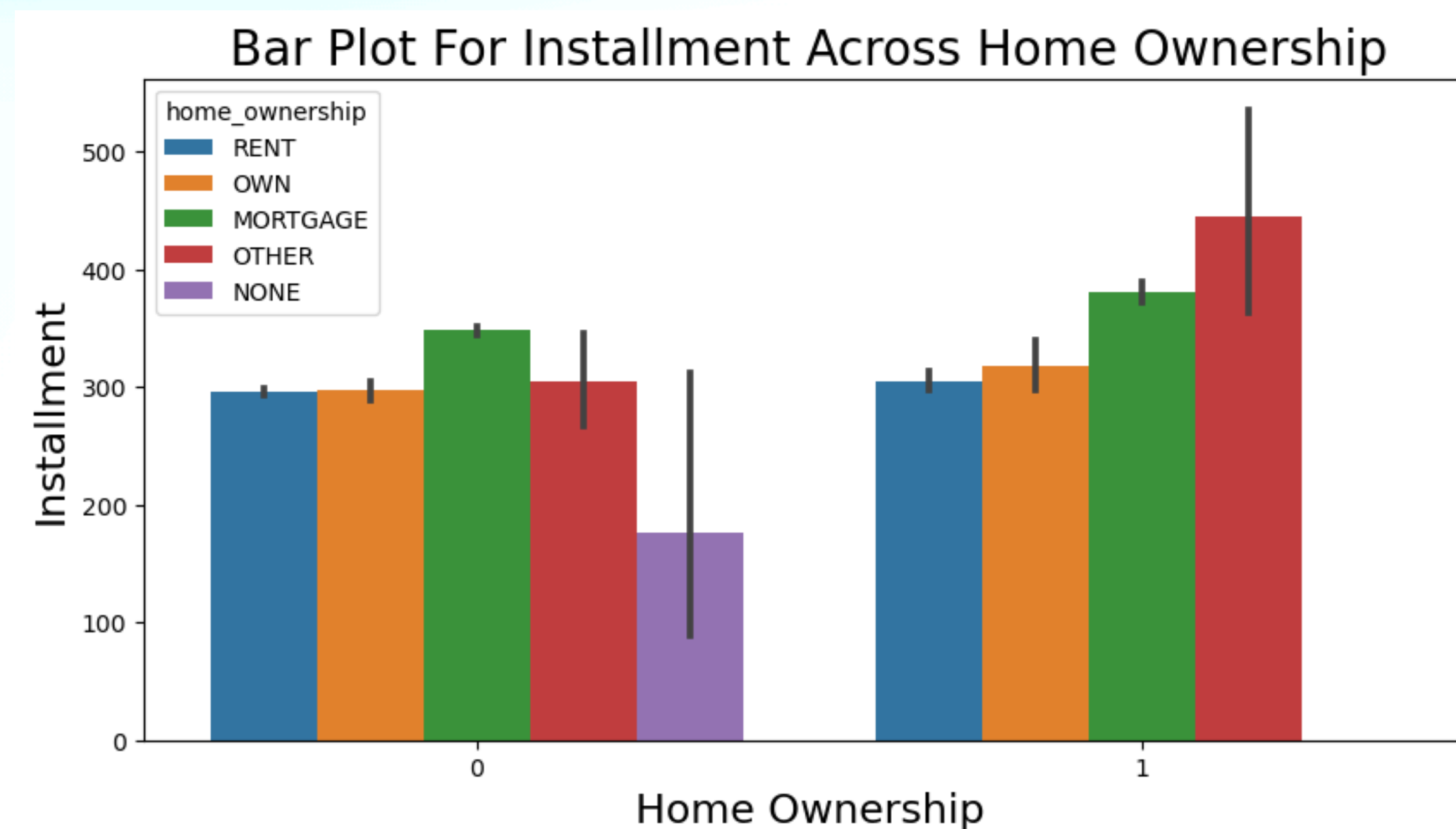
Analysis on Funded amount against DTI

- In the graph, we can observe a pattern between 6000 USD to 10000 USD where the DTI of a customer is between 12% to 17%.
- The highest ever recorded DTI of a customer is 29% where he has received a loan amount around 24000 USD. This shows that a customer belongs to high income group.



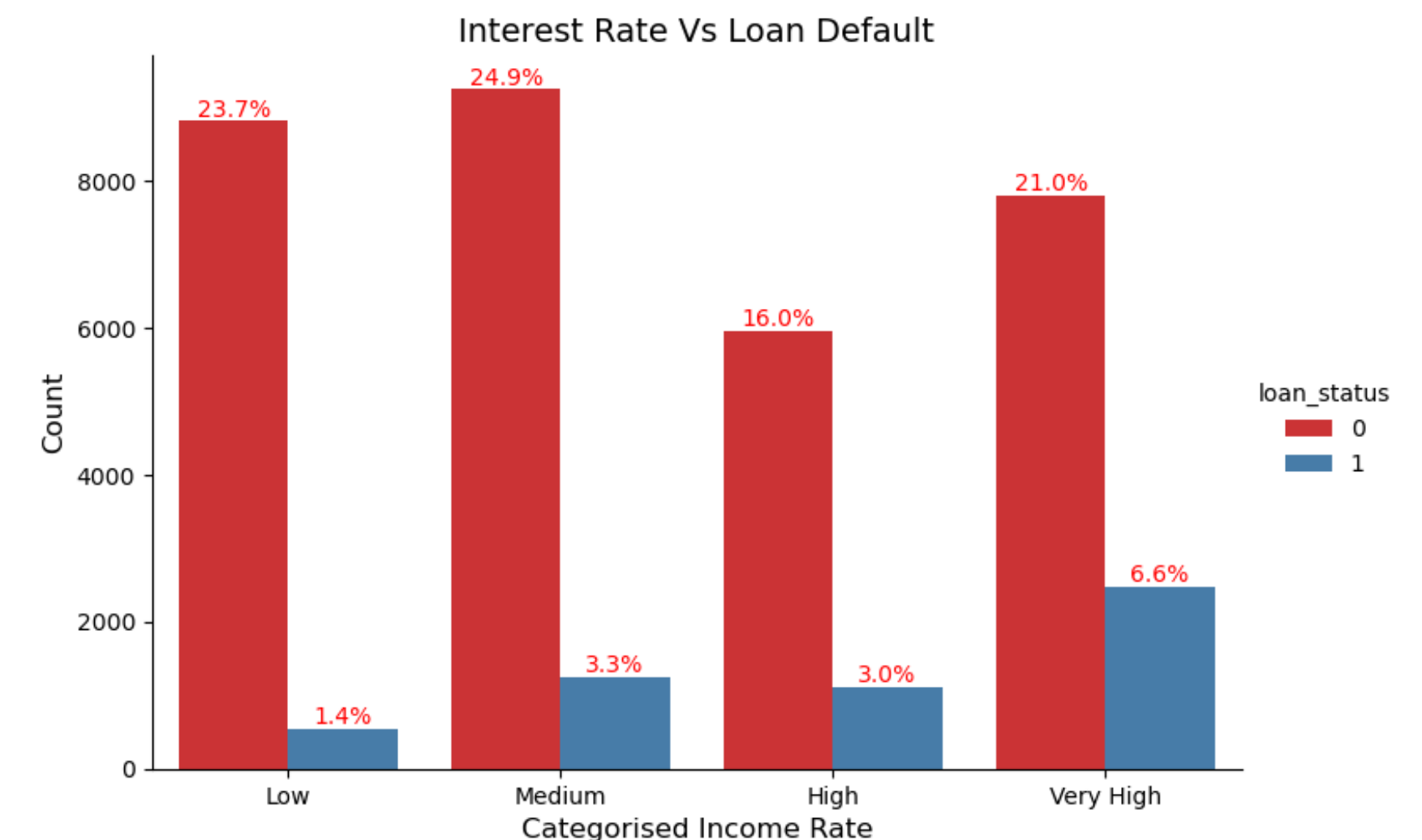
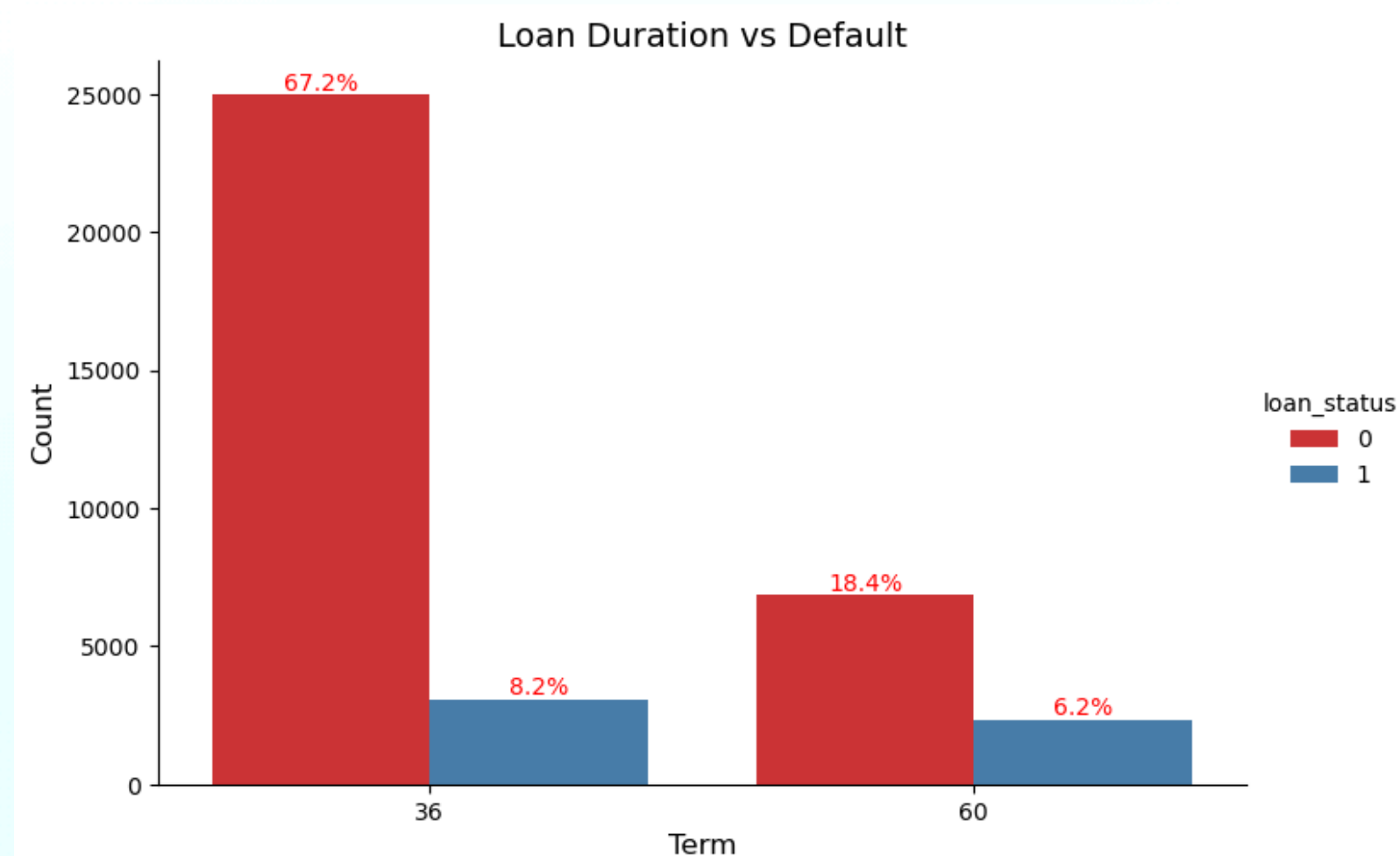
Analysis on Instalment paid by customer based on home ownership

- In the below graph, it is observed that the Mortgaged customers has defaulted the loan most.
- Customers who rent a house are equal by either making full payment or defaulting the loan.



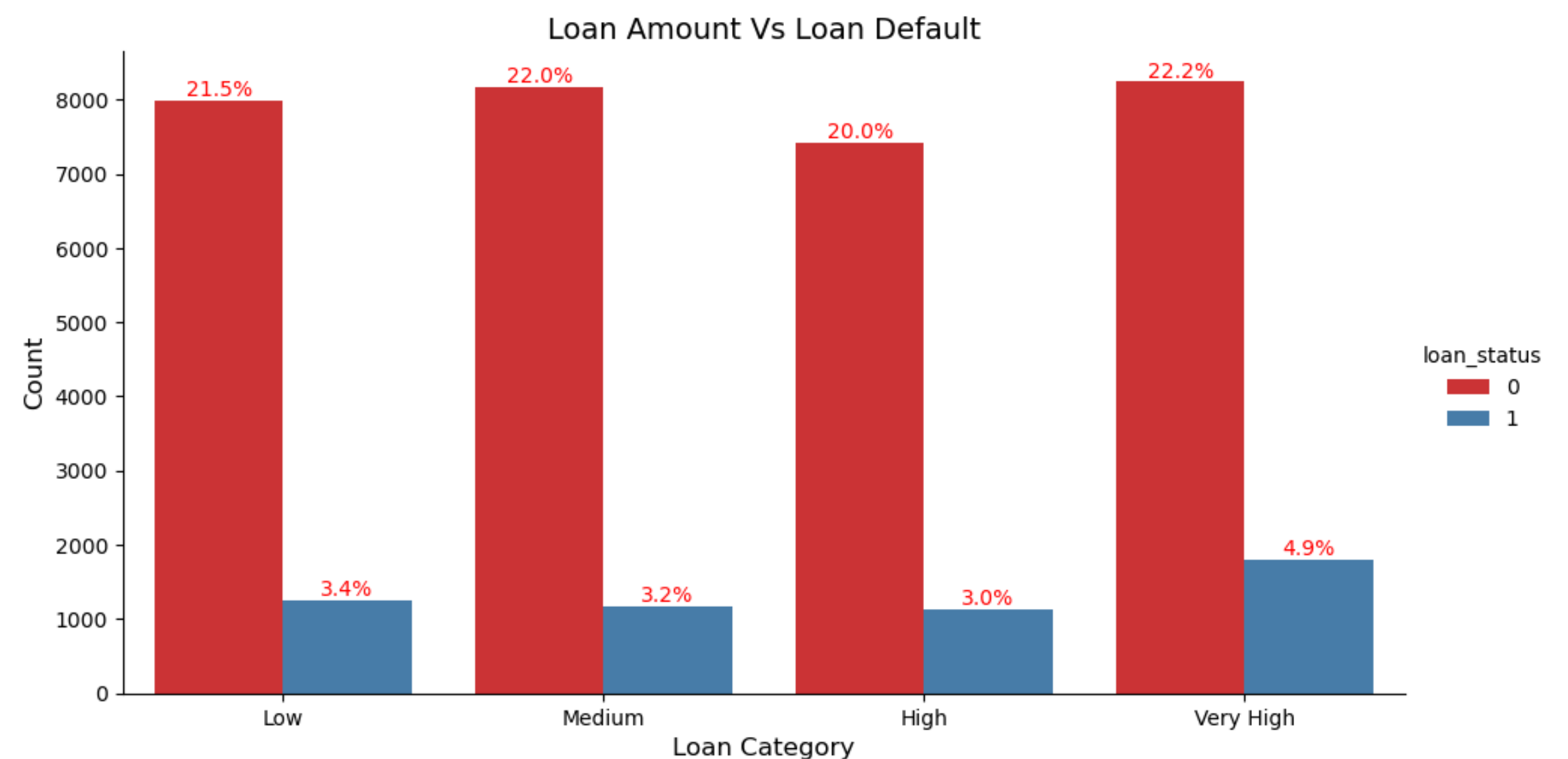
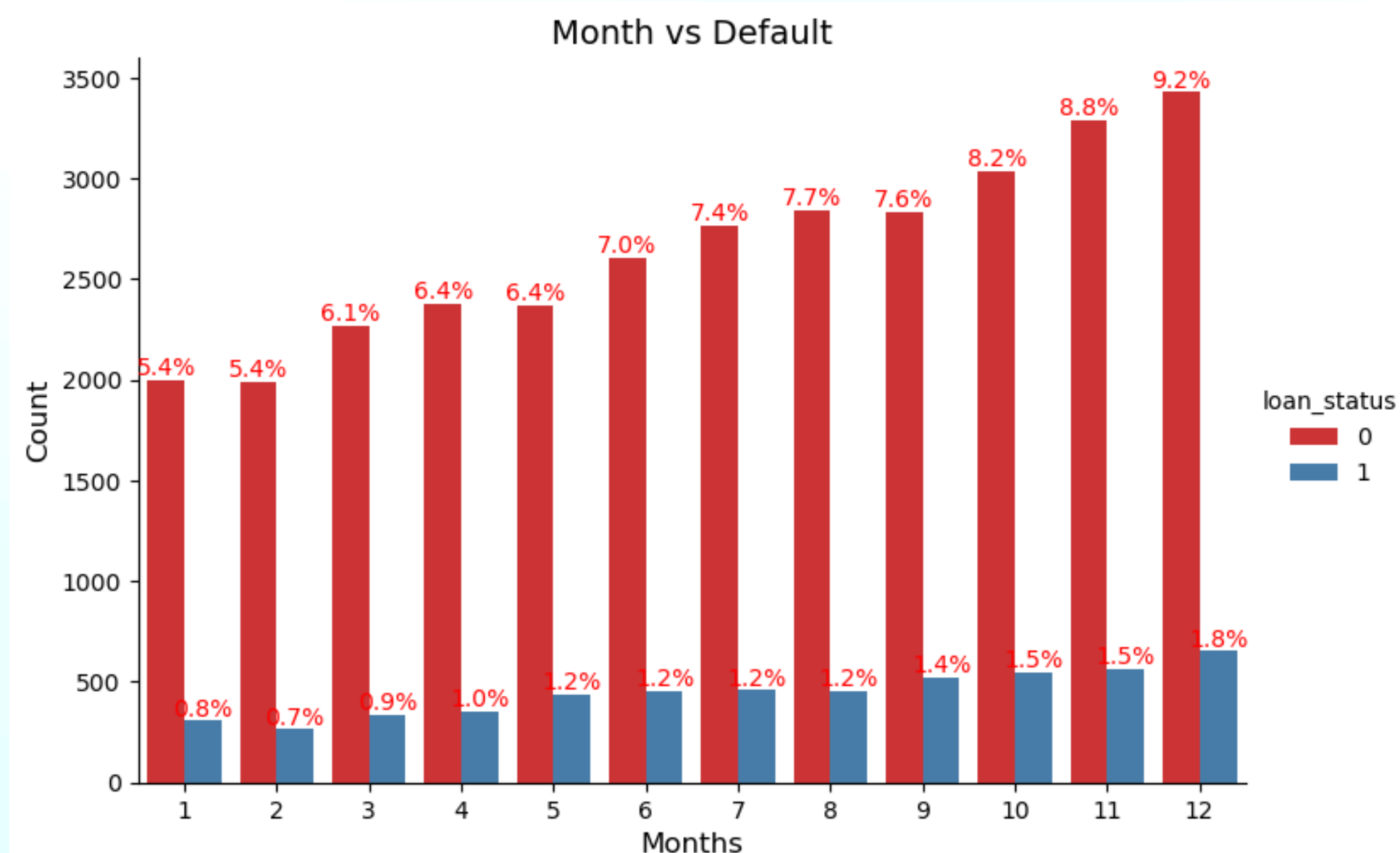
Customers who default based on Interest rate

- The key takeaway from the below graph is that, the customers who choose low interest rate for a period of 36 months are more likely to default on their loan.
- The approximate difference between 36 month and 60 month loan tenure is 48.8%. It means that majority of people stick to loan term which are less.
- The ratio of customers defaulting their loan based on interest rate is caused due to the increase in interest rate. For every increase in interest rate, there is a steady increase in default percentage.



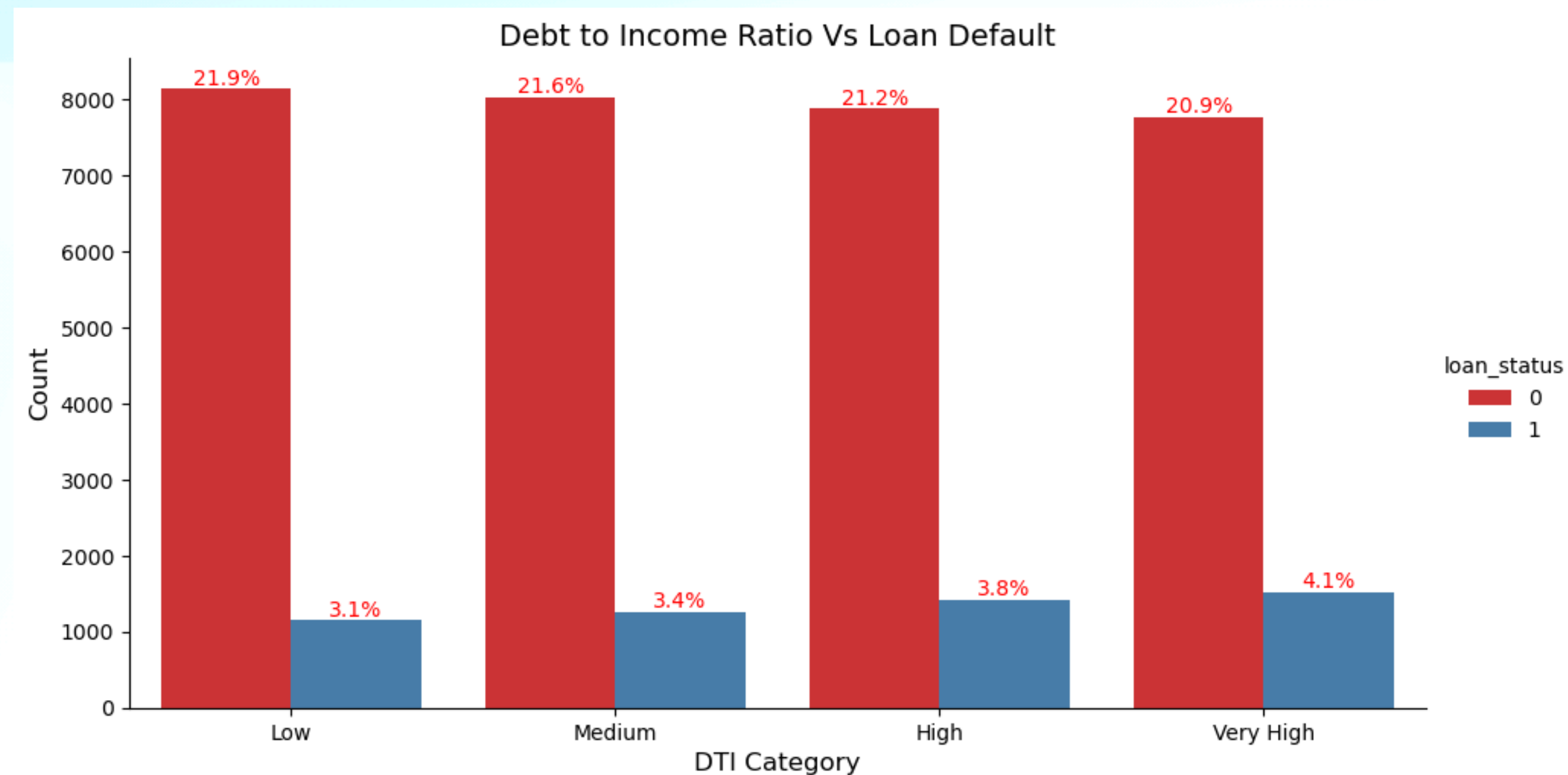
Analysis of Loan Amount against Term

- In the below graph it is observed that the customers start to default their loan when the loan amount start to increase. In our inference, customer has defaulted his loan 8000 USD and above.
- Customers are likely to default on their loan at the end of the year or at the end of the financial year. December has the highest default percentage (1.8%).



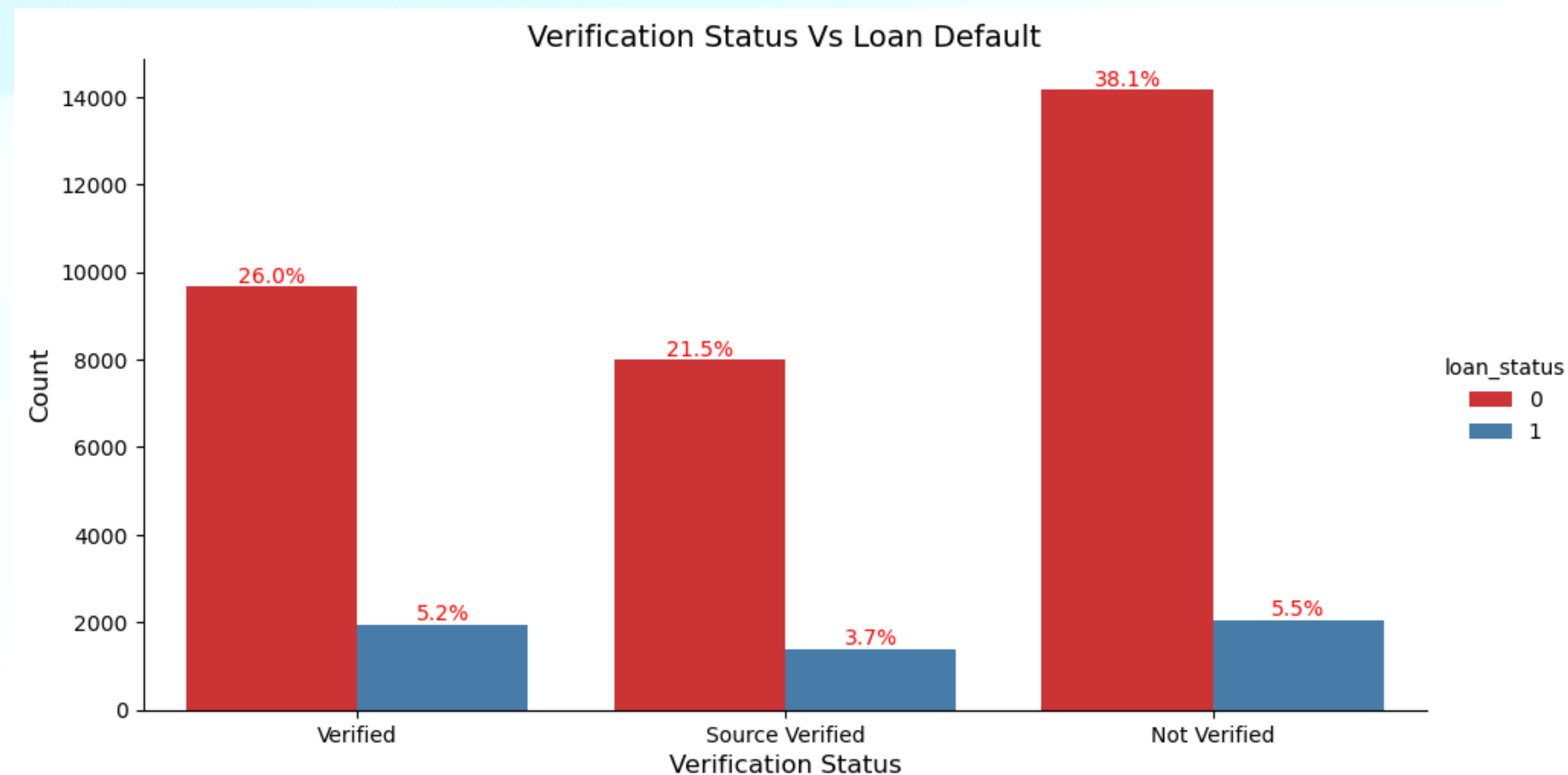
Default of loan based on DTI

- Persons with low and medium DTI are likely to make full payment on their loan.
- Customers with very high DTI percentage are most likely to default on their loan.



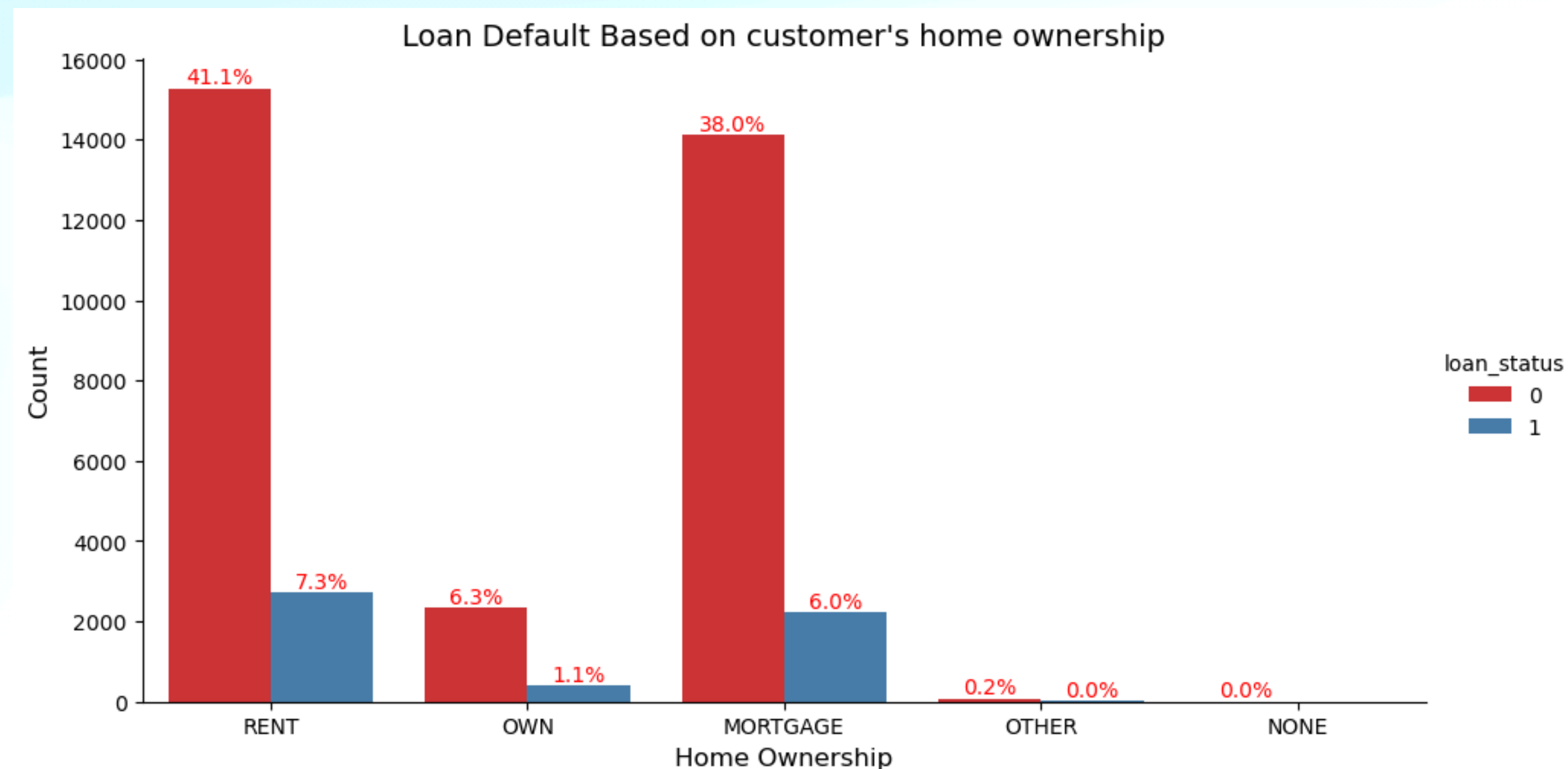
Loan default based on Verification status

- Though the customers who are not verified has significantly made the full payment rather than the verified customers, but the defaulting percentage of not verified stands at the top.



Loan Default based on Home ownership

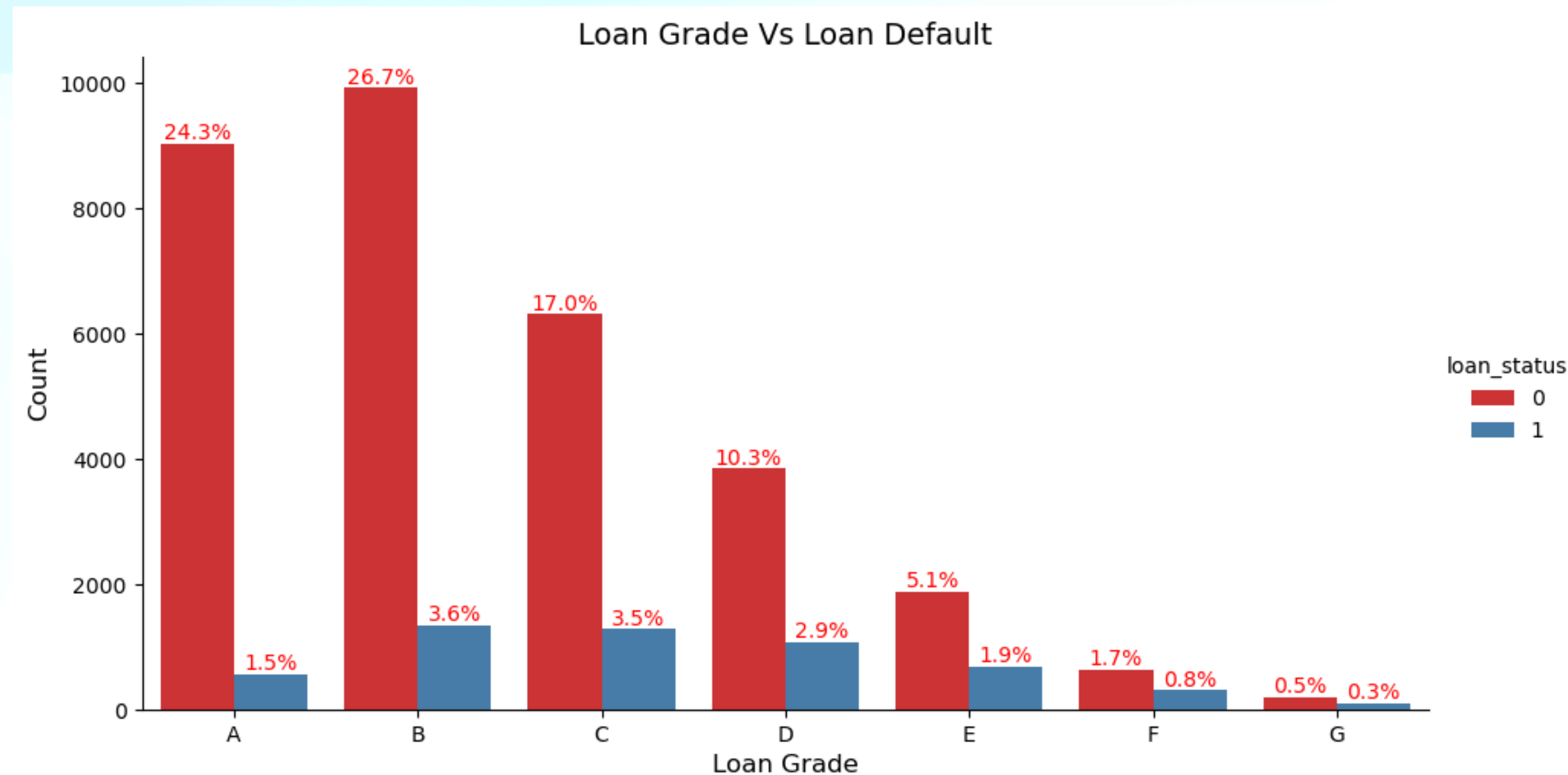
- Customers who don't own a property are the highest of defaulting the loan.
- Customer's who own a property has record of low defaulting their loan.
- Customers who rent are most likely to repay their loan rather than the customers who own a property.



Grade analysis on Loan Default

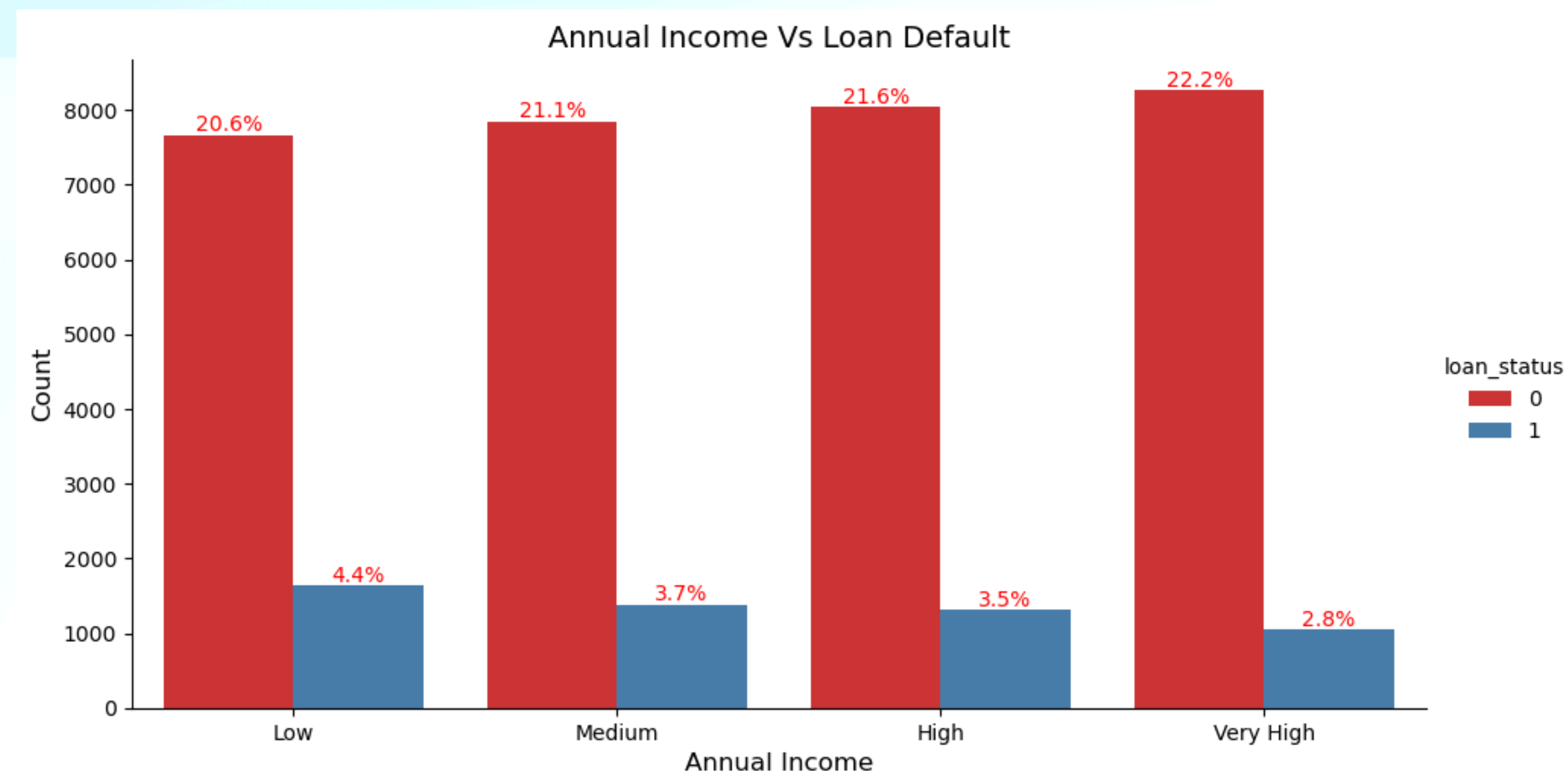
Loan Grade is the categorising the customers ranging from A to G by lending club where A is the highest score and G is the lowest score.

- In the below graph, customers who belong to Grade B and C are defaulted rather than other group.
- The Grade G has the lowest repayment and default record. It because either the borrowing population is low or the bank might rejected their applications.



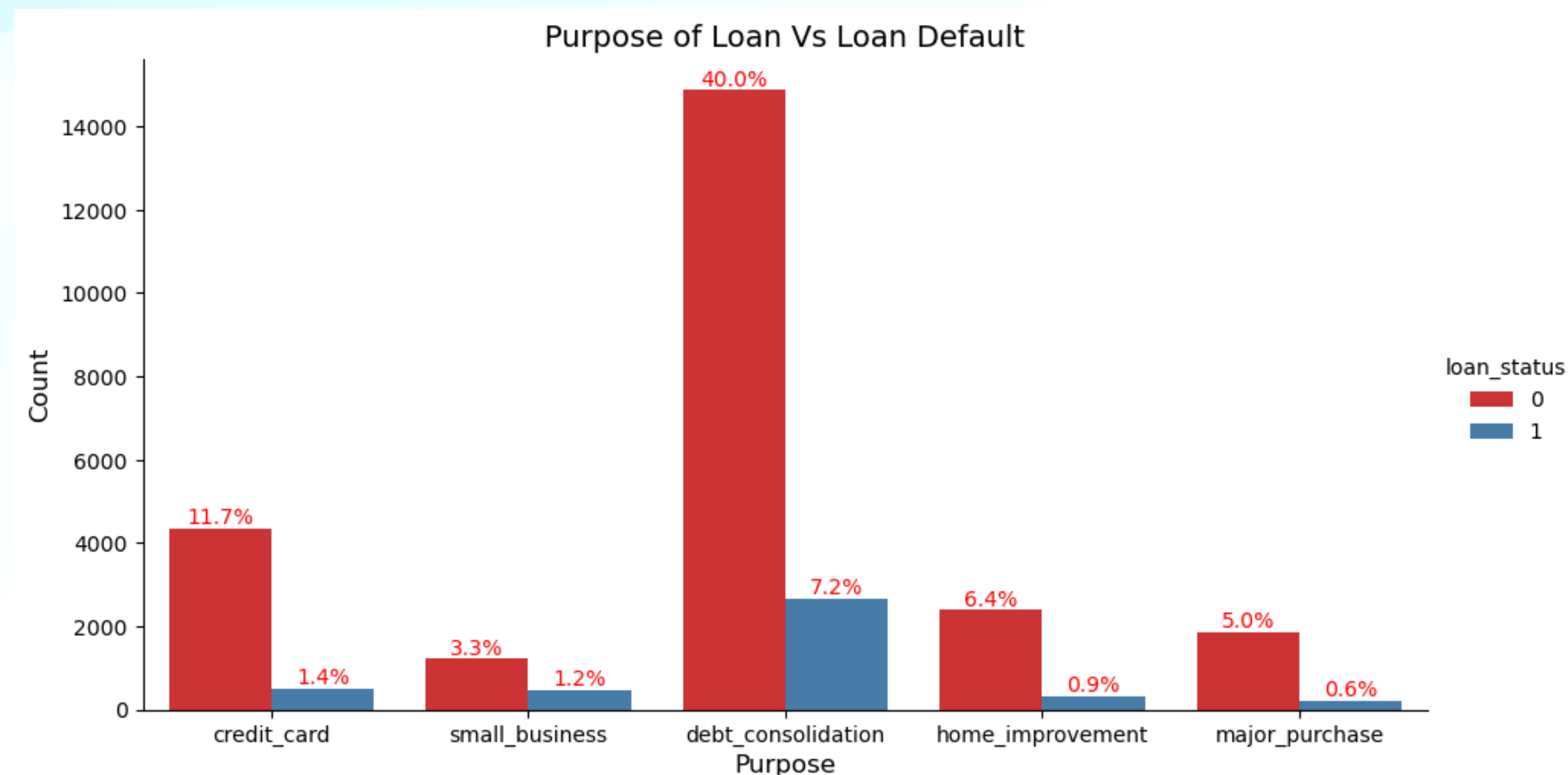
Annual Income Vs Default

- The below graph tells us that the customer's who belong to low income group are likely to default their own. Their reasons could be high interest rate, long term instalments, loan amount and various other factors.
- People who belong to very high annual income has considerably low default rate, since their DTI ratios are good.



Loan default based on the purpose of borrowing

Customers who have borrowed loan for their debt consolidation has defaulted their loan with 7.2%. In this inference, let us take that both debt consolidation and credit card can fall under one category as debt. The credit card default percentage is 1.4% . Lets combine both of them by adding up the percentage as 8.6% . From this percentage, let's consider that customers who take loan for debt are most likely to default rather than other categories.



Analysis Report

- With ample amount of parameters that we have inferred with the help of the given dataset, The loan amount has been compared against loan status with the following details:
 - Interest Rate at which the loan was taken.
 - Duration of the loan term.
 - Defaulting Month.
 - Debt income ratio.
 - Annual income of the customer.
 - Purpose of the loan.

After a thorough analysis of the dataset with the above parameters, it is advised that the bank should follow the criteria as below:

- Customers who approach loan for debt consolidation and credit card should be approved to customers who have very high annual income and DTI.
- Customers who are of low income group who pay very high interest rate should be watched carefully since the probability of defaulting is high for them.
- As per the data customer's tend to delinq or default at the end of every year end.
- Before lending the amount, the bank has to verify the customer if he owns a property and provide loans against security for staying on safer side.
- Banks can give priority to customers who belong in Grade B, since their repaying rate is better than Grade A and C.
- Low income group has taken a high interest loan with longer duration for small business will default their loan.