LENDING CLUB CASE STUDY

Team Details:

Group Facilitator: Shreeraksha S

shreeraksha.s123@gmail.com

Group Member: Shiva Sharaj H N

shivasharaj1ga16ee047@gmail.com

Table of Contents

- Introduction
- Business Understanding
- Business Objective
- Data Understanding
- Data Analysis(Visual Representation with Insights)
- Conclusion

INTRODUCTION

Solving this assignment will give you an idea about how real business problems are solved using EDA.

In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

BUSINESS UNDERSTANDING

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

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
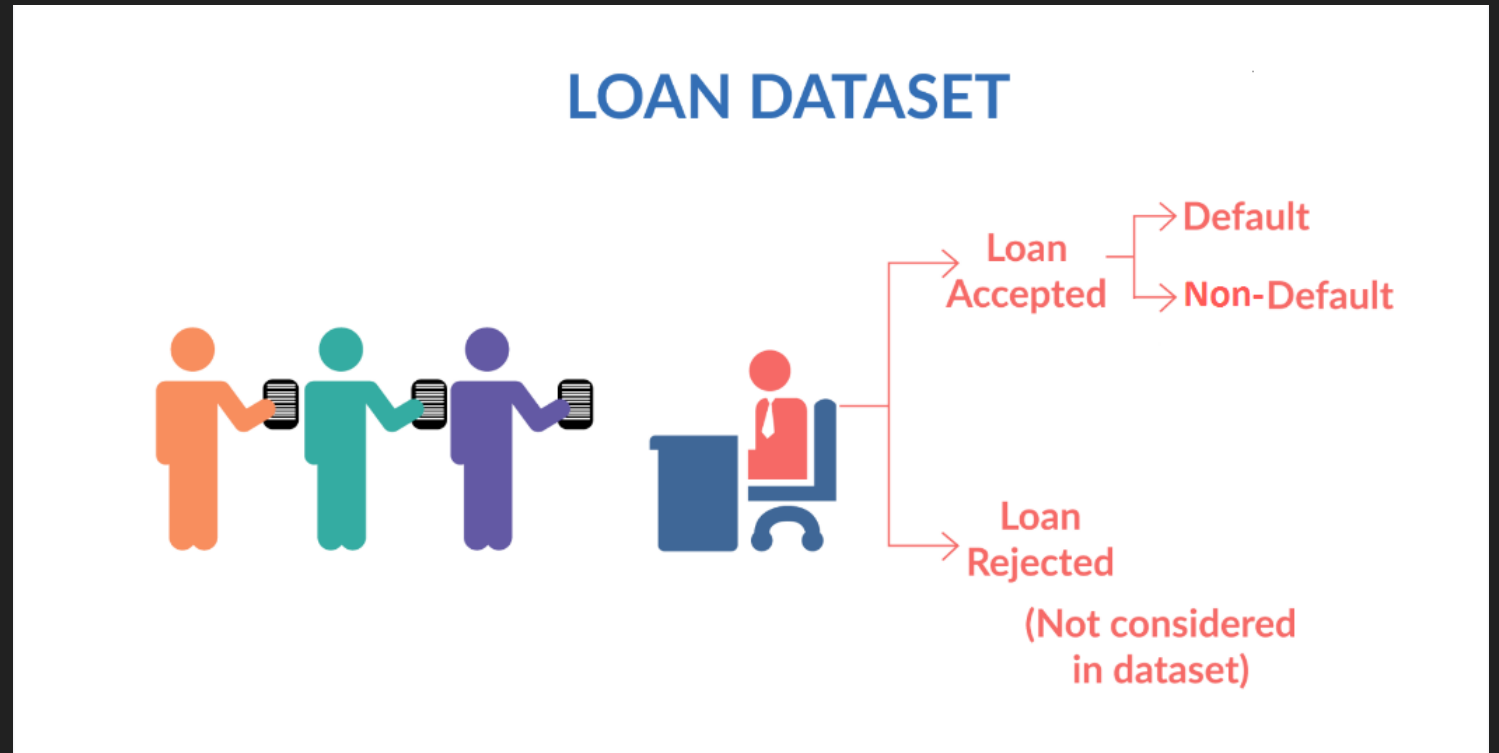
When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1.Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

    1.Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

    2.Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

    3.Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

2.Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

BUSINESS OBJECTIVE

This company is the largest online loan marketplace facilitating different types of loans.

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

DATA UNDERSTANDING

This dataset has 2 files as explained below:

1. **loan.csv** contains all the information of customer's loan application data and their financial history.

2. **Data_Dictionary.xlsx** is a Data Dictionary which describes the meaning of Data variables.

OUTLIERS DETECTION

boxplot on various amount columns

Loan_
Amnt

Funded_loan
_amnt

Funded_loan
_amnt_Inv

boxplot on various amount columns

Int_rate

DTI
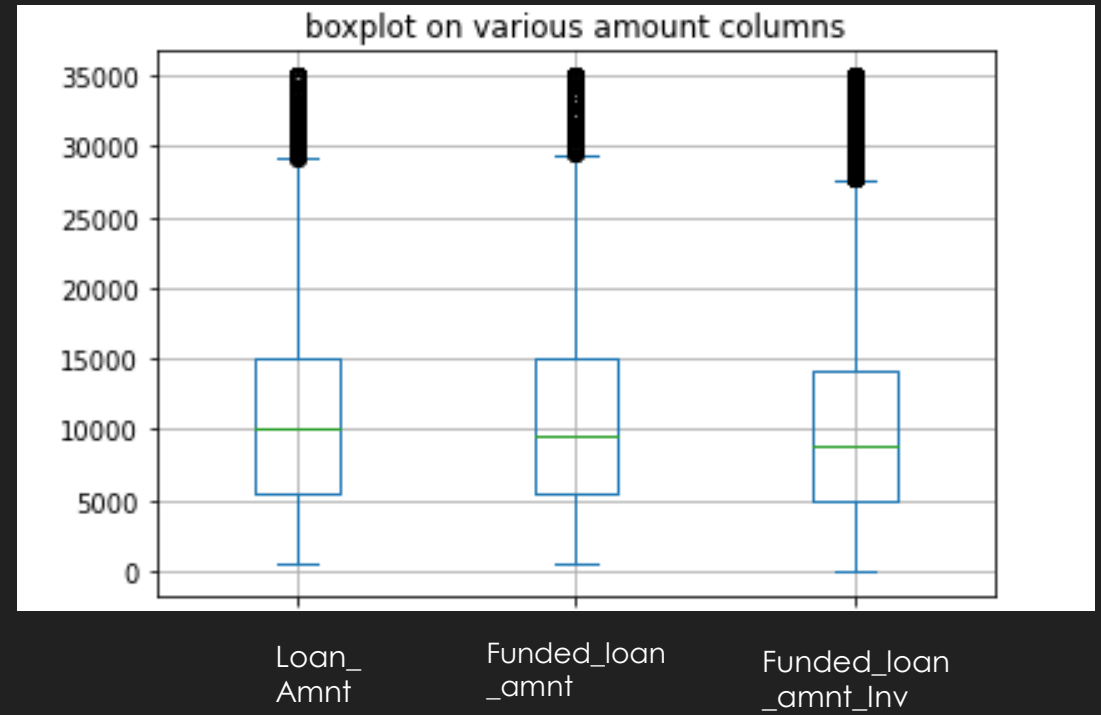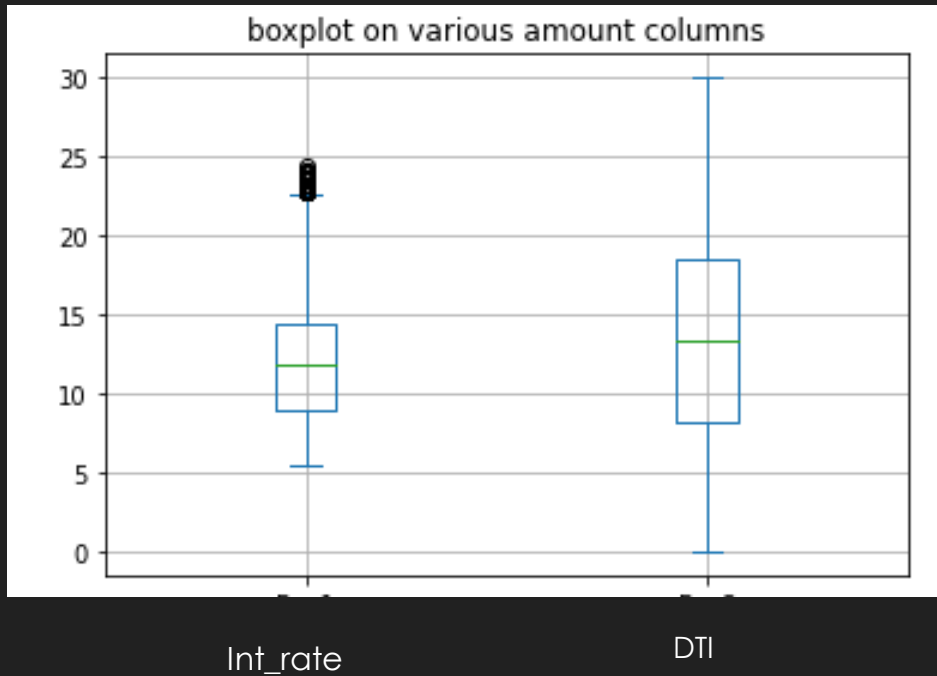
We can keenly observe the outliers in all the variables and get an understanding about the data.

It is important to detect Outliers , if not detected it may lead to wrong analysis of the data.

UNIVARIATE ANALYSIS

Insights: As the Grade of Customer goes from A to G, the chances of customer being a defaulter increases. This is expected because the grade is decided by Lending Club based on the credit history of the customer.

Grade vs Loan status
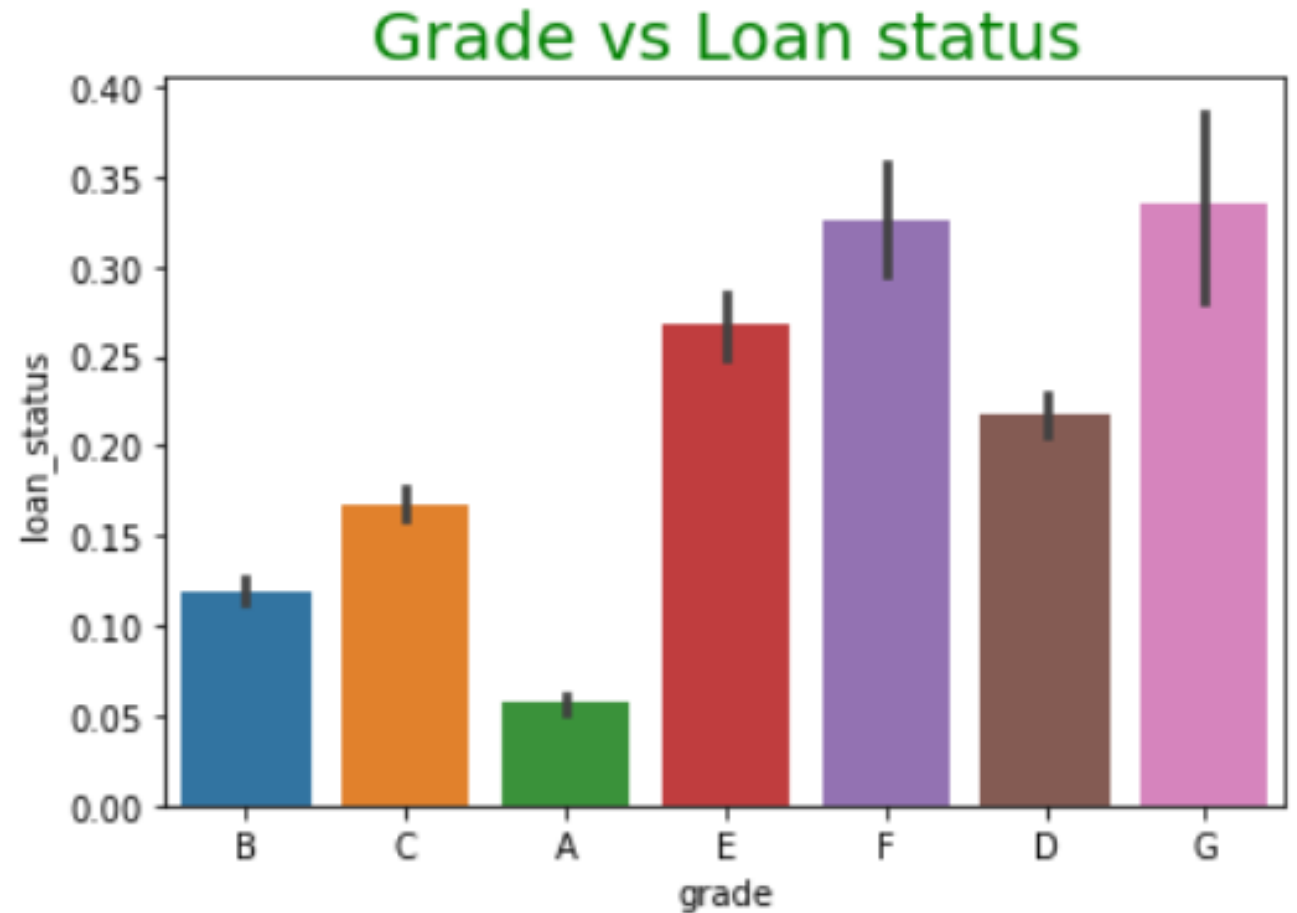
sub_grade vs Loan status
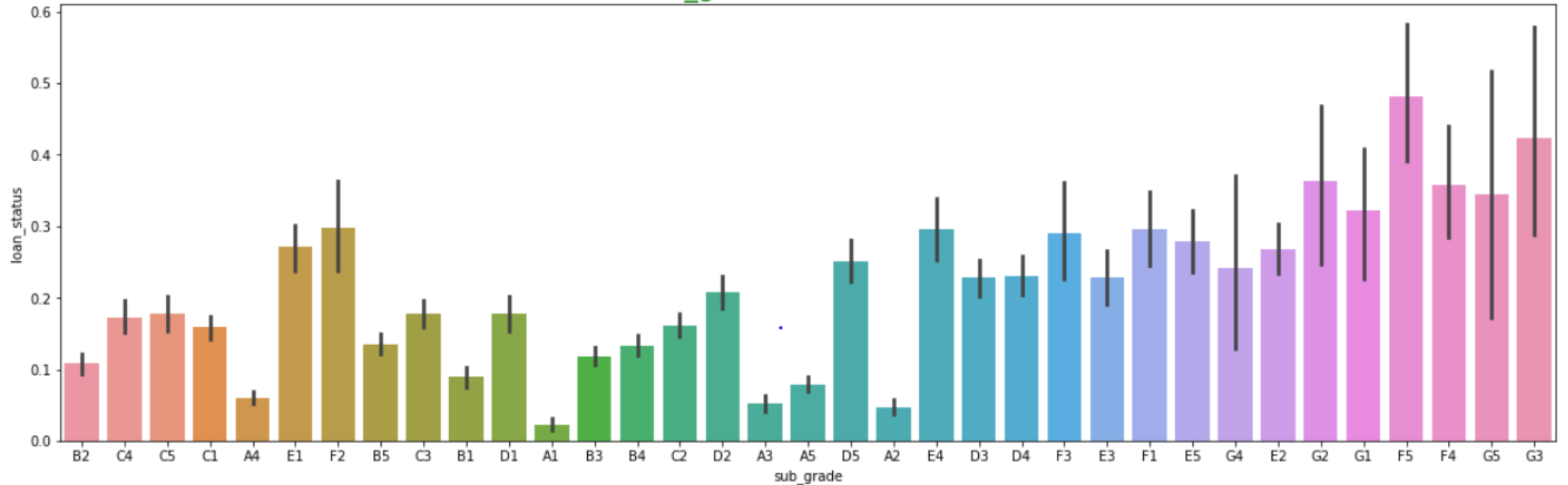
Insights: It is observed that the the chances of customer being a defaulter increases as sub-grade moves from A1-G5 roughly. It can be concluded that as the subgrade increases beyond F, the amount of risk involved increases.
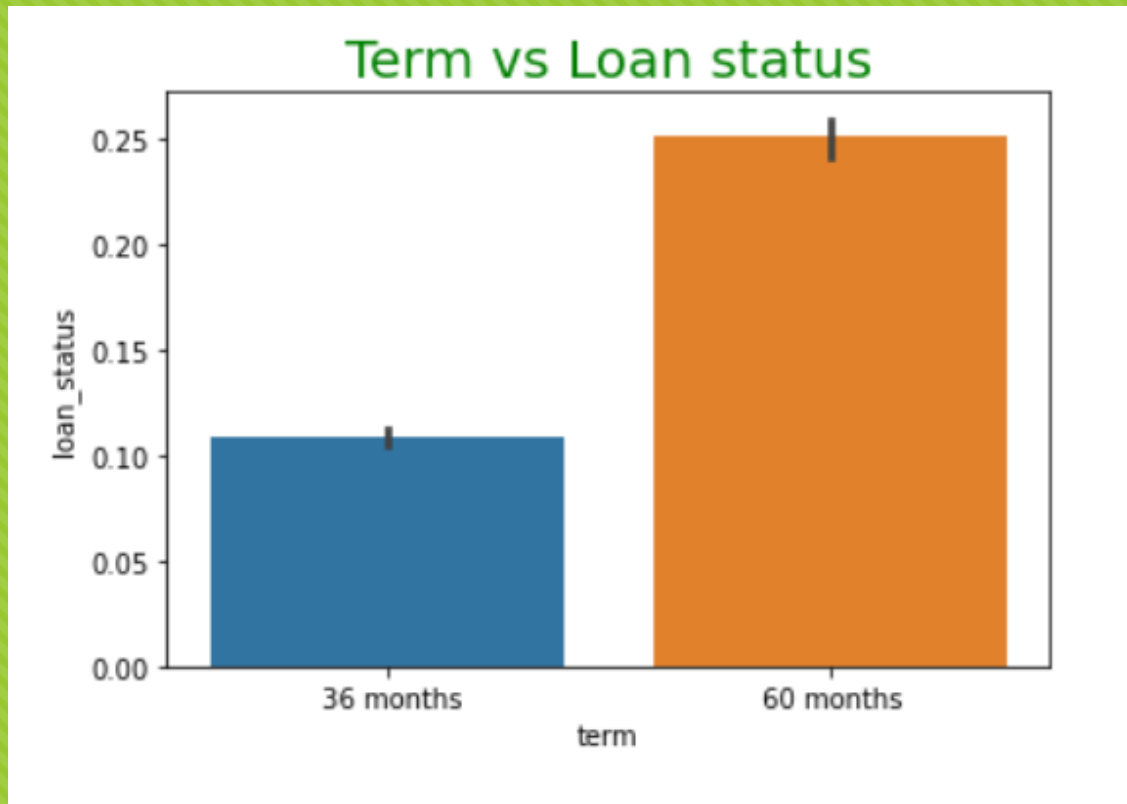
Figure 1
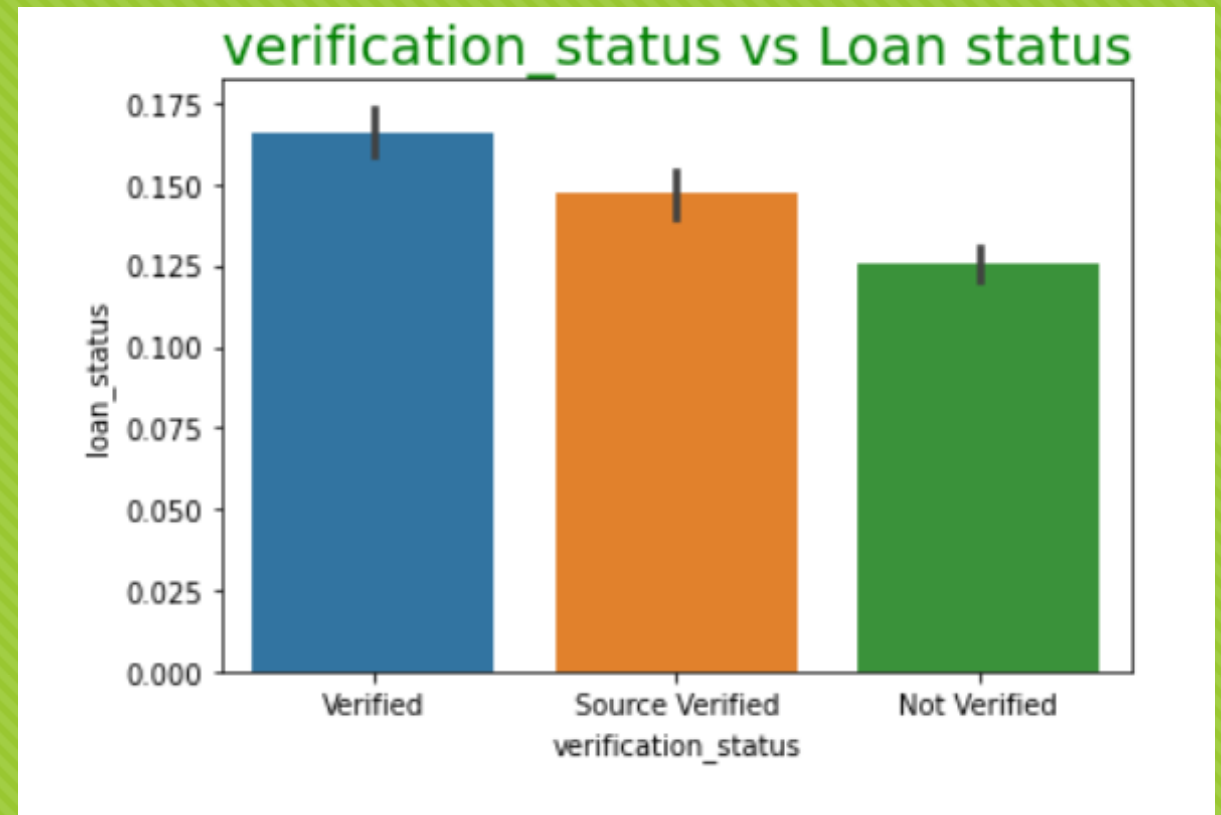
Figure 2

Insights:

Fig.1 : The defaulter's rate is more in customers who have opted for 60 months term period.

Fig.2 : It is observed that the verified customers are more defaulters than those who are not verified which is a surprising factor.
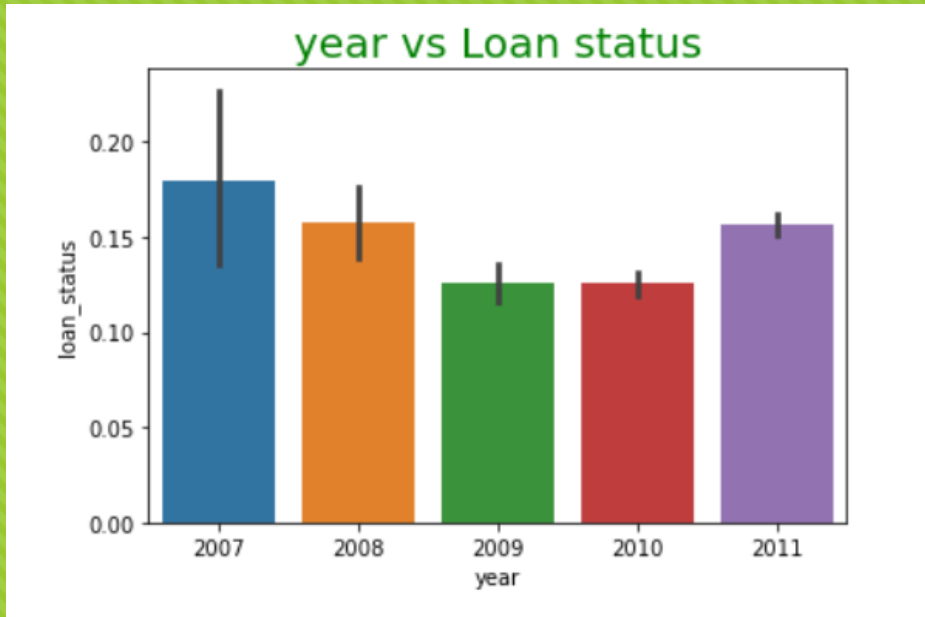
Figure 1

Figure 2

Insights:

Fig.1 : The defaulter's rate is decreasing from the year 2007, and there is sudden rise in the defaulter's rate in 2011. However we had the highest number of defaulters in 2007.

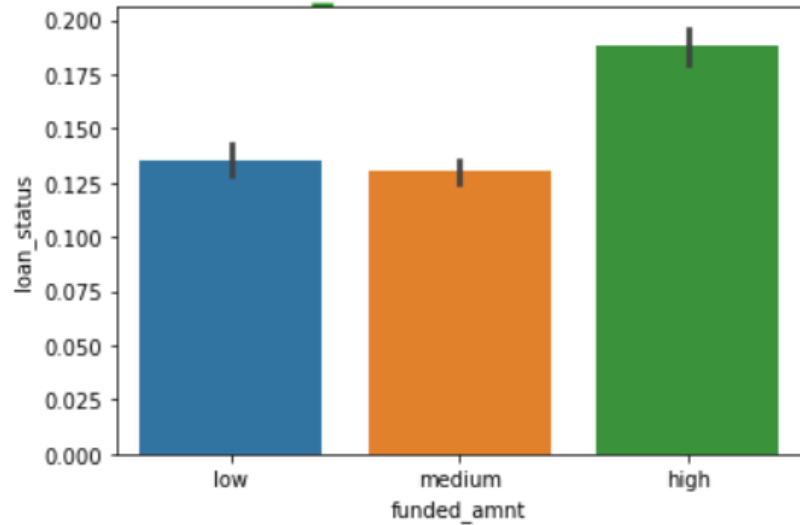Fig.2 : Not much can be inferred from Month vs. Loan status plot
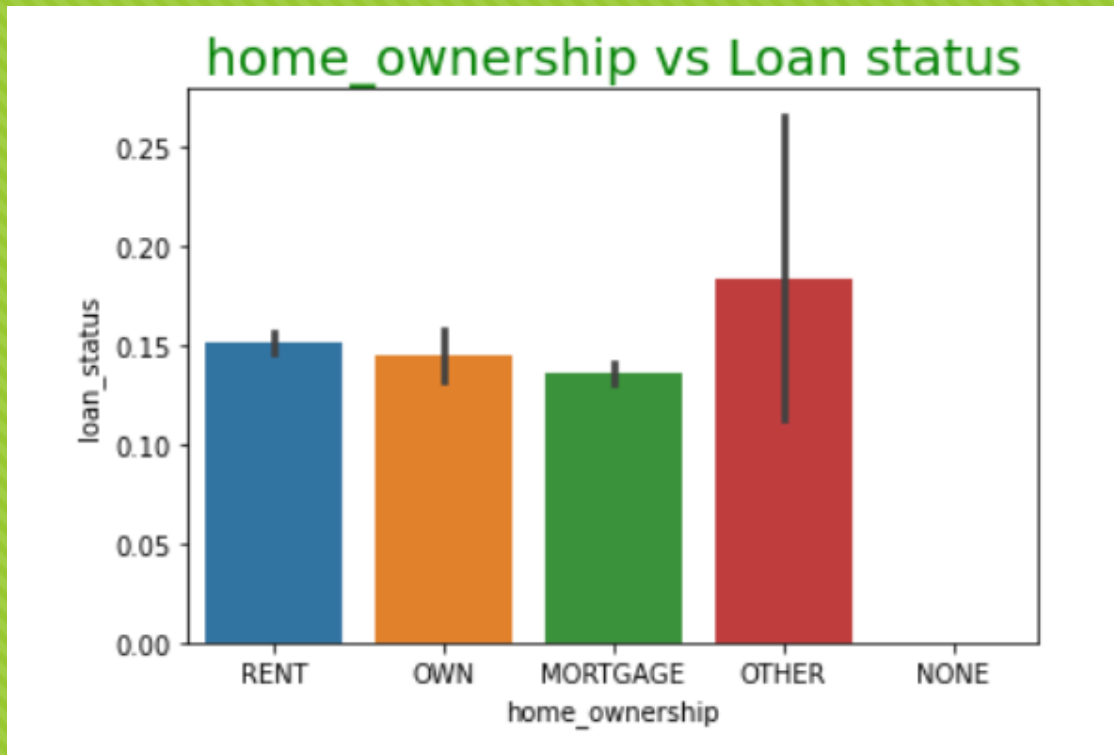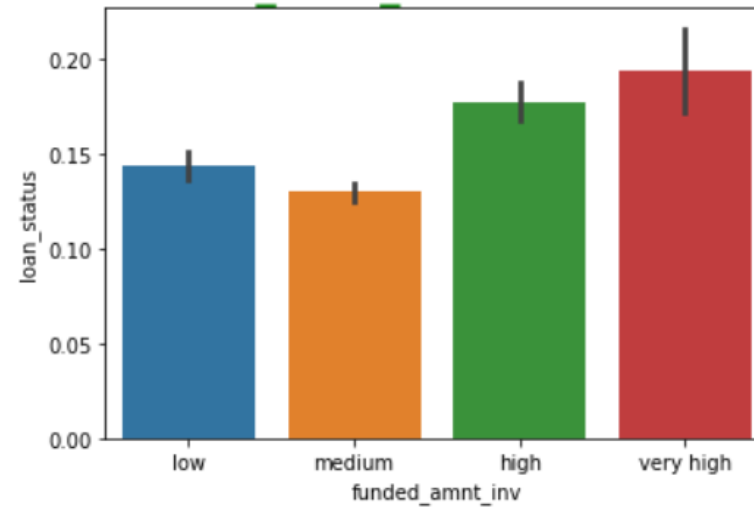
Figure 1
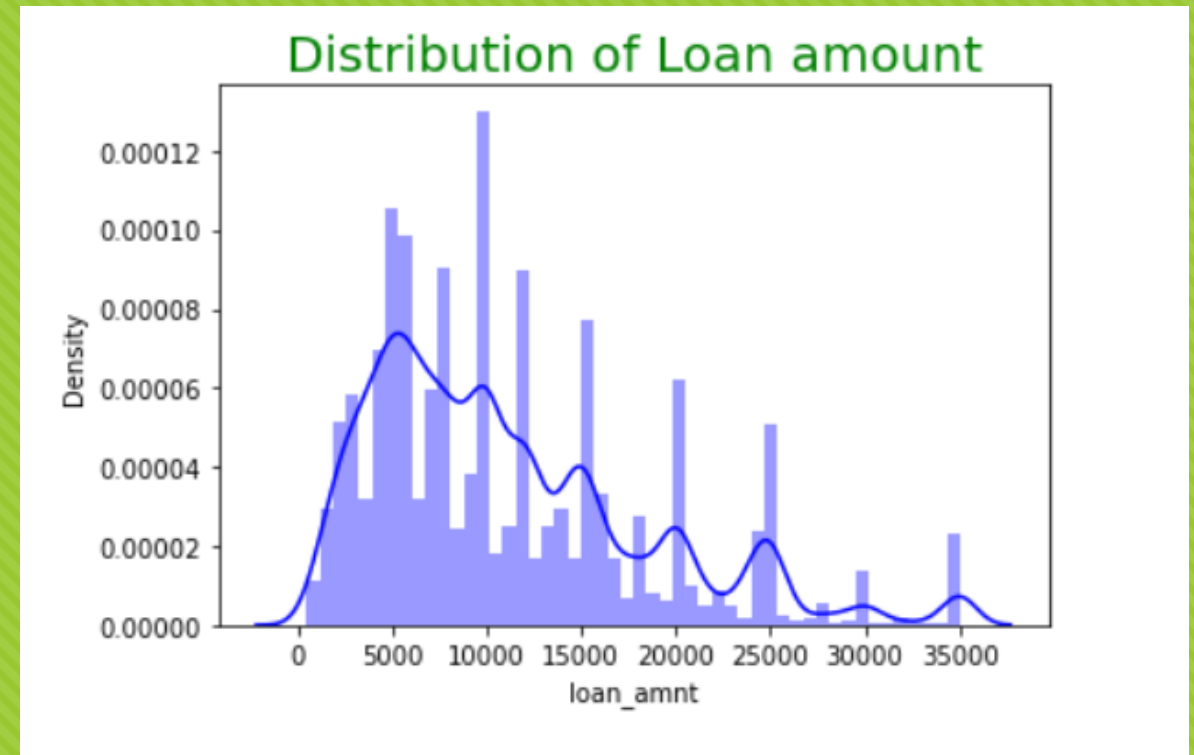
Figure 2

Insights:

Fig1: As there is not much difference in the barplot of home_ownership, it is not a great discriminator for our analysis.

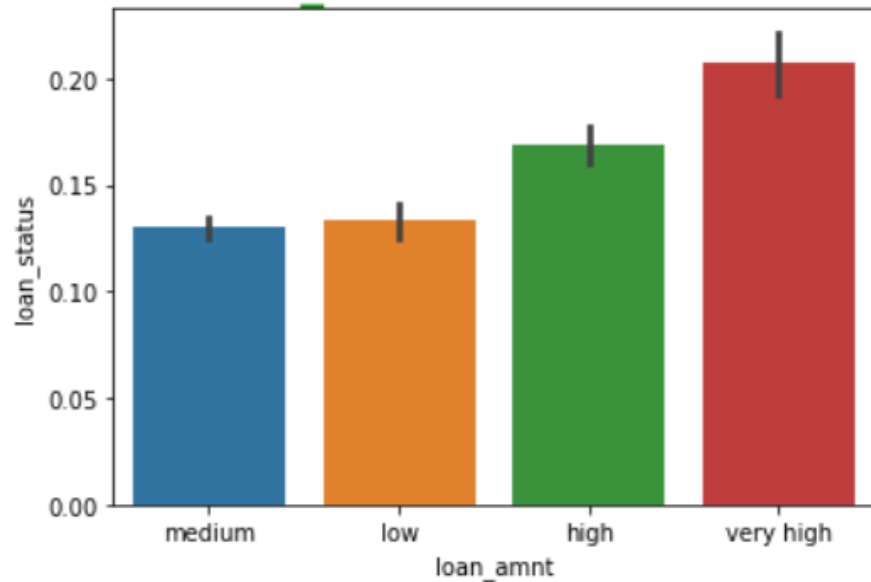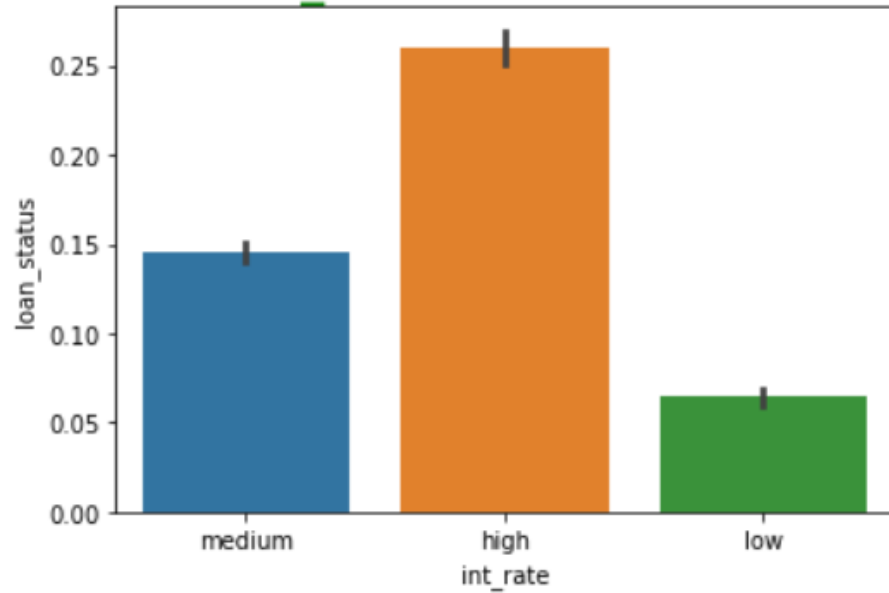**Fig2:** Number of customers taking Loan amount of 10,000 is the highest.

Loan_amnt vs Loan status : The customers who have higher loan amount are the highest in defaulters list.
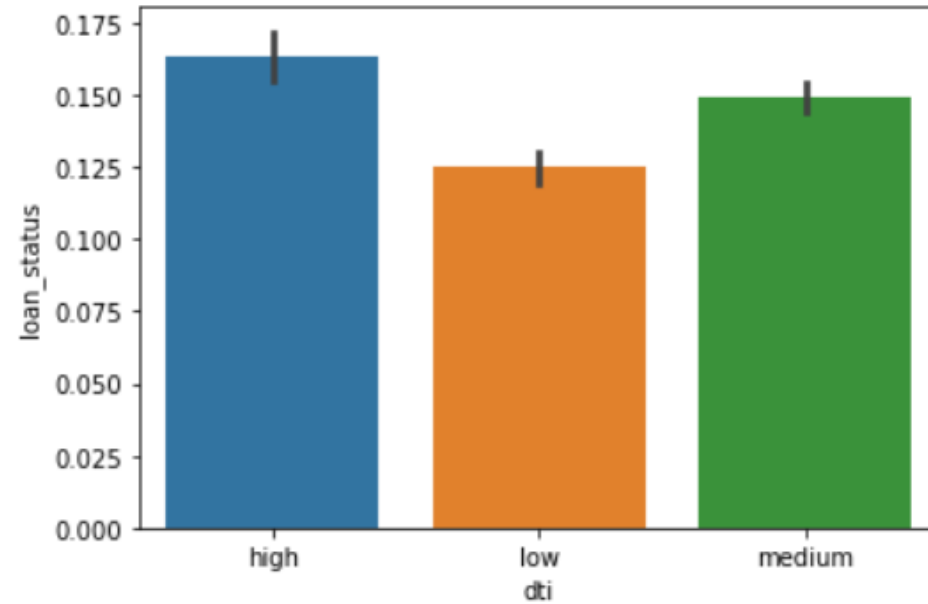
Funded_amnt vs Loan status : As a continuation to the previous plot, we can see here that higher the Funded amount, higher is the chances of customer being a defaulter.

Funded_amnt_inv vs Loan status : It also follows a similar pattern as the other two plots.
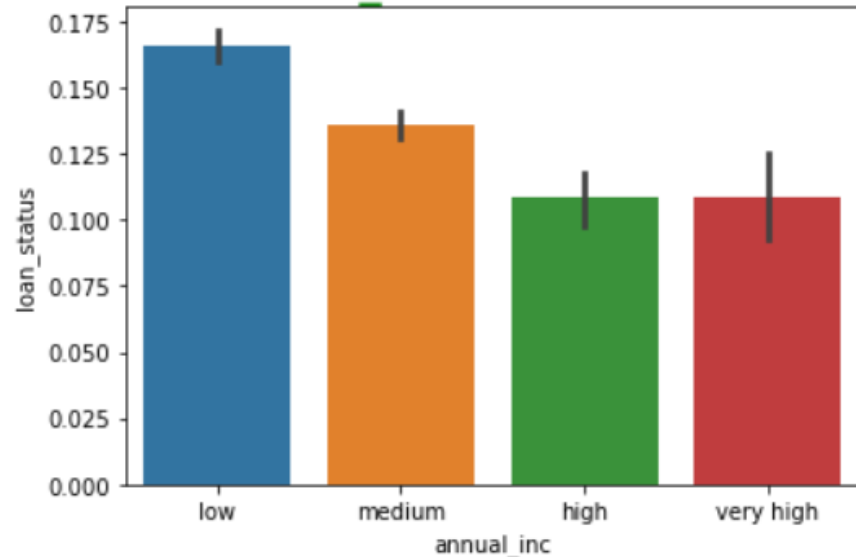
Int_rates vs Loan Status: Higher the interest rate, higher is the chances of being a defaulter, which is expected as banks levy higher interest rates for customers of high risk.

Dti vs Loan status: It is seen that the defaulters list is increasing for customers having higher Debt to income ratio(dti) which is expected.

Annual_inc vs Loan status: The customers with less annual income are the highest defaulters.

# SEGEMENTED UNIVARIATE ANALYSIS

Looking at the plot, we can infer that customers opting for 60 months loan term are most likely to be defaulters even on comparison with "Purpose of the loan " for different term periods of 36 months and 60 months.
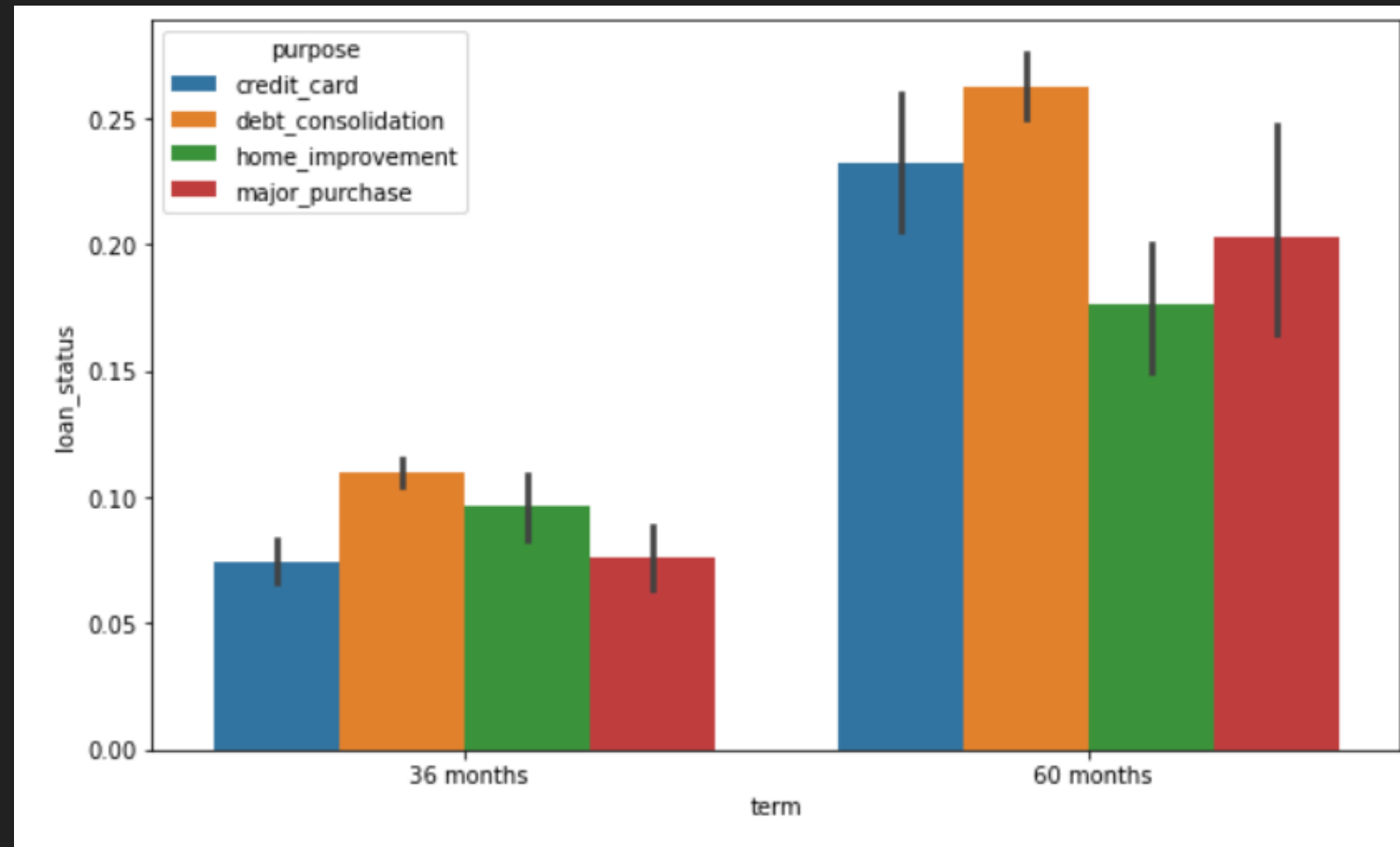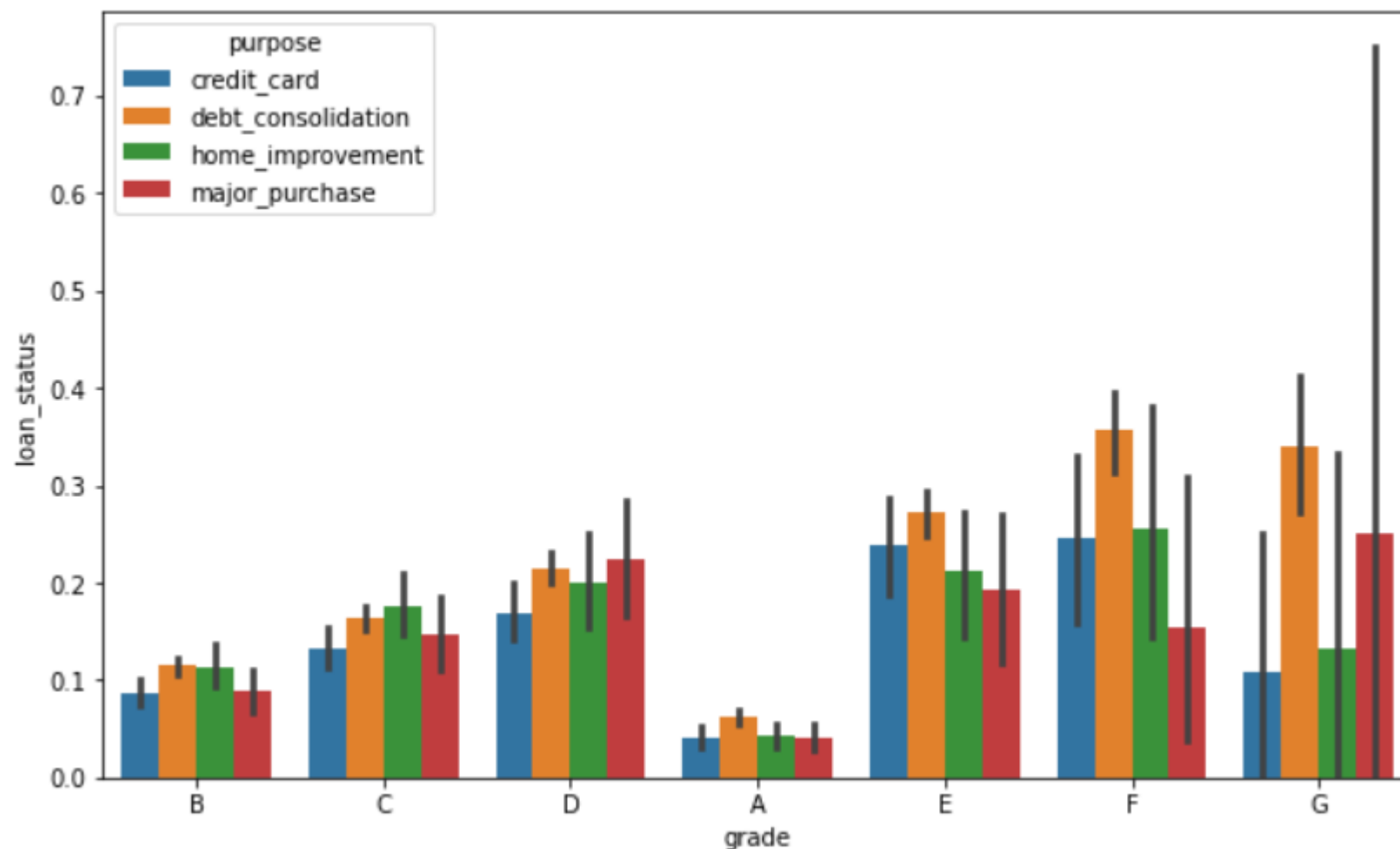
As the loan grade varies from A to G , there are high chances of customer being a defaulter.

From this we can infer that, If the customer is from either F or G grade and the purpose of loan is mentioned as "Debt consolidation", the customer is most likely to be a defaulter.
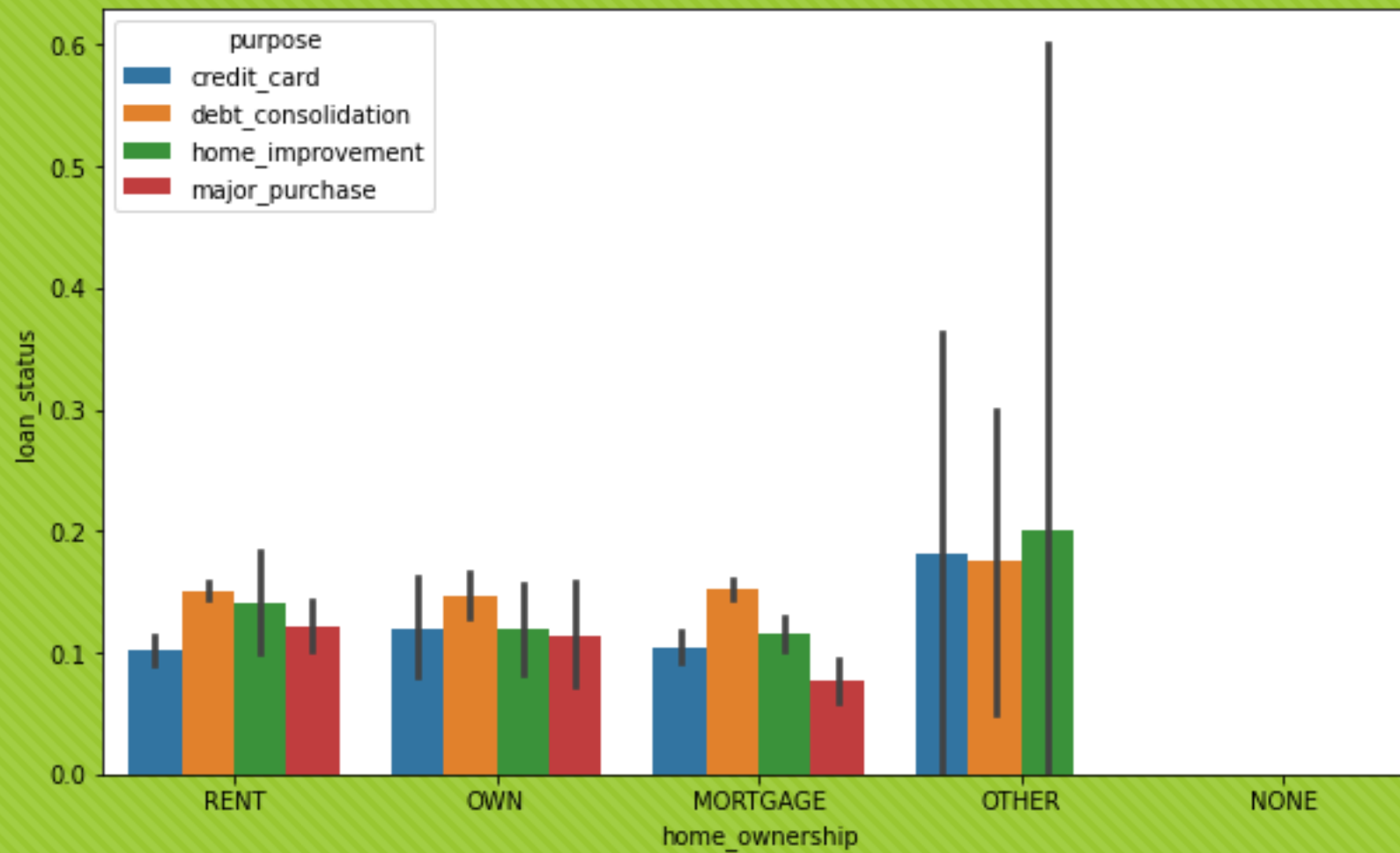
Insights: It's quite riskier to lend loan to customers with purpose of Loan as "Debt consolidation" and Home ownership status as "Other"
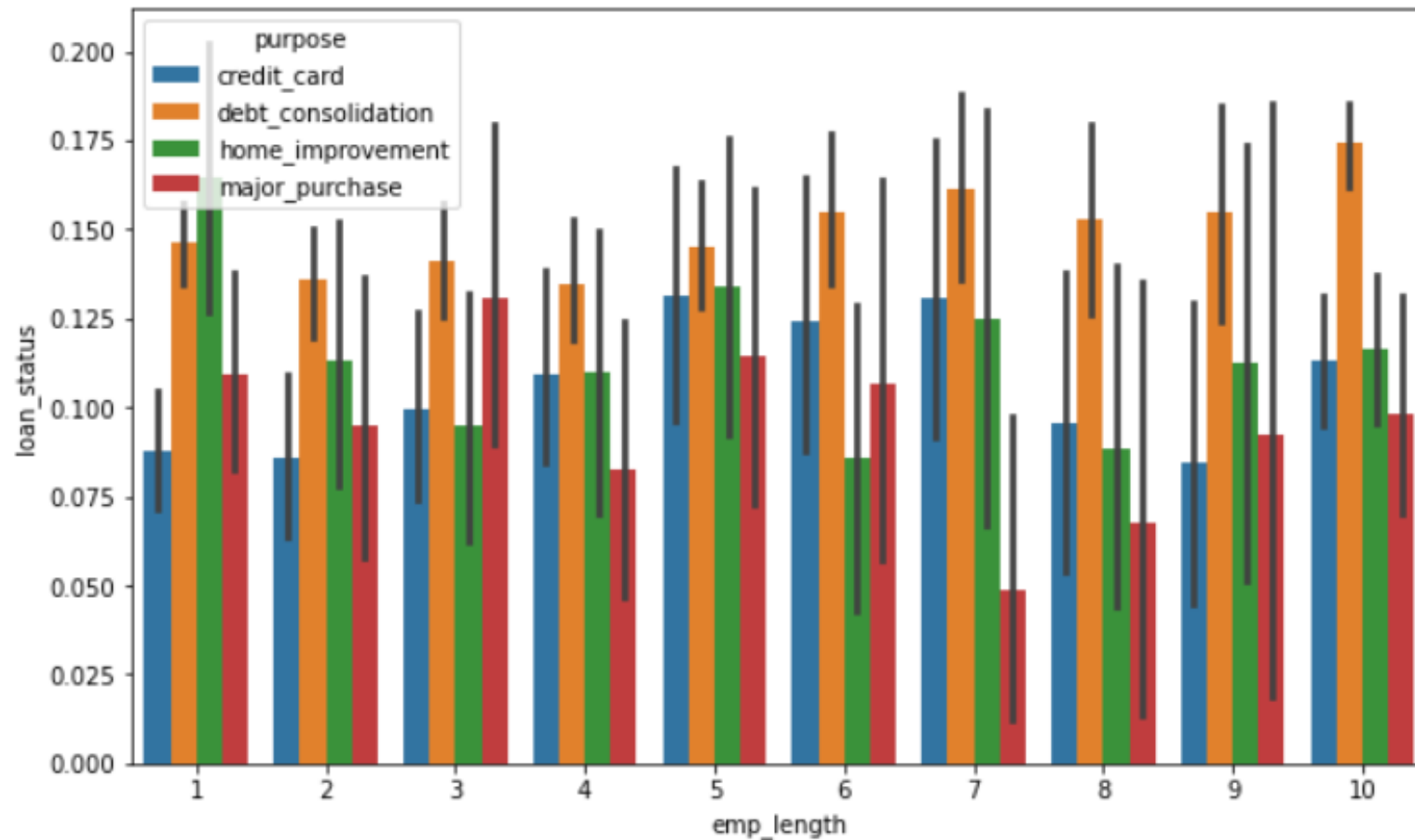
Insights: Not much can be inferred from this plot. Irrespective of the employment length of customer, the chances of customer being a defaulter remains high for customer with purpose as "Debt consolidation"
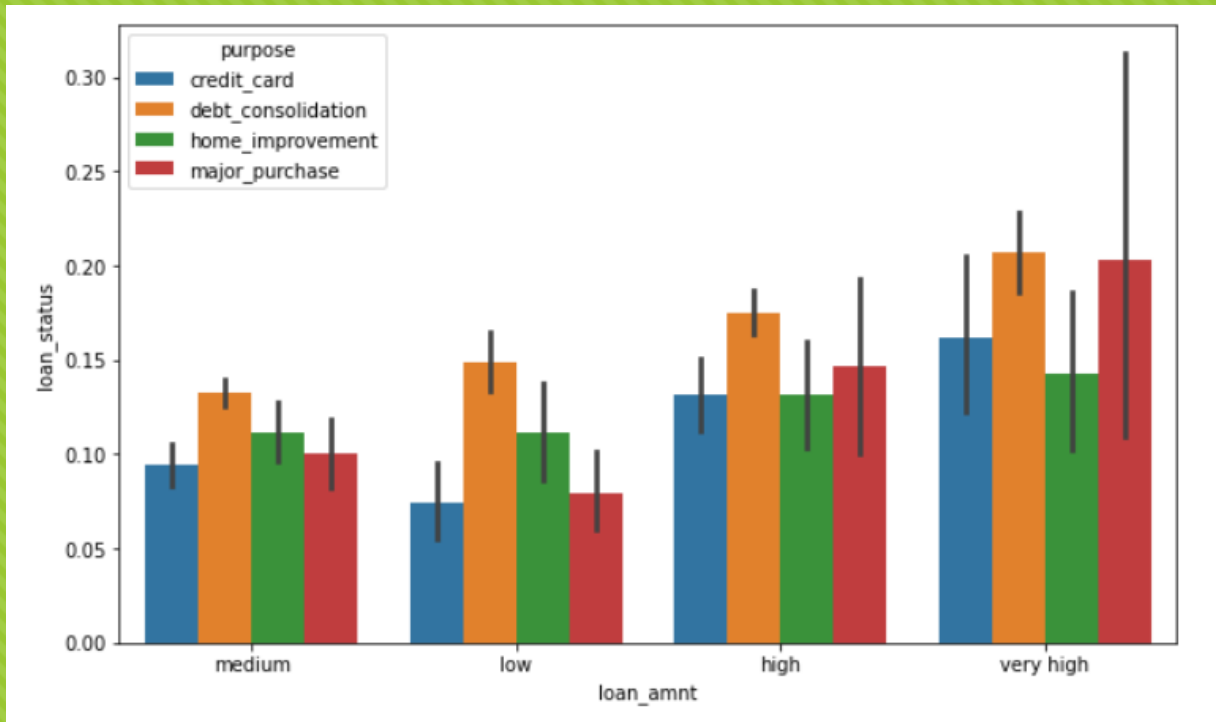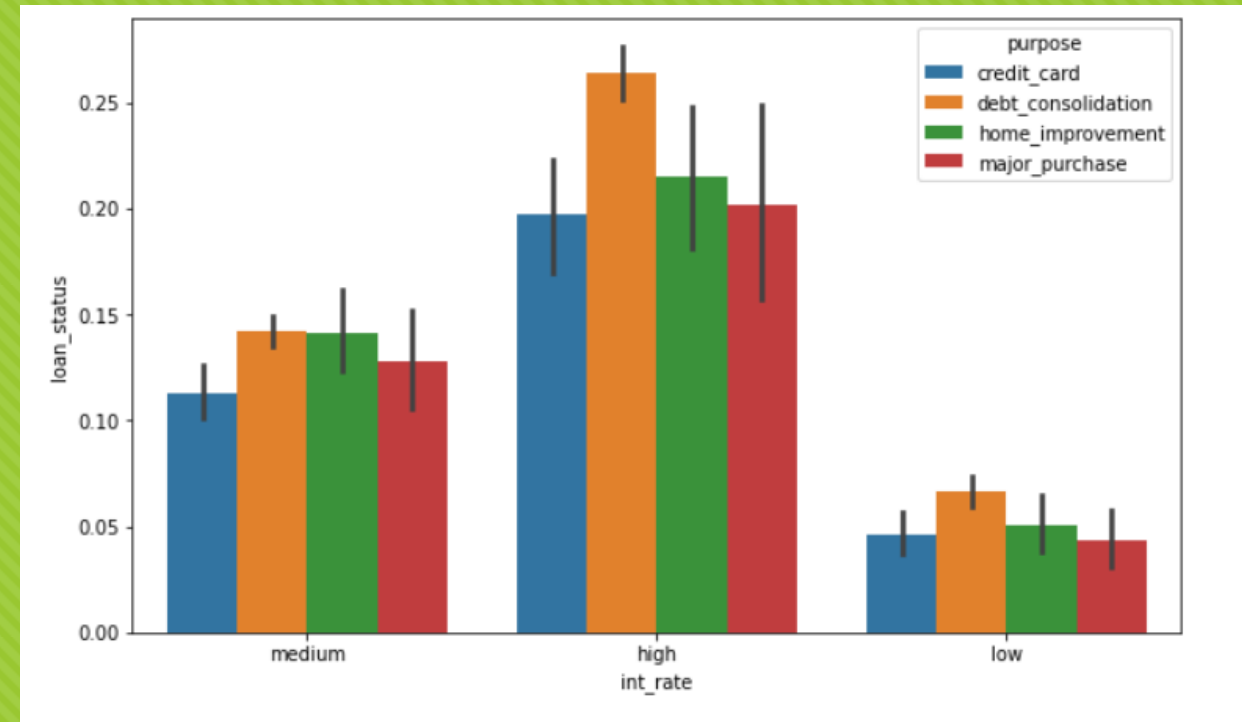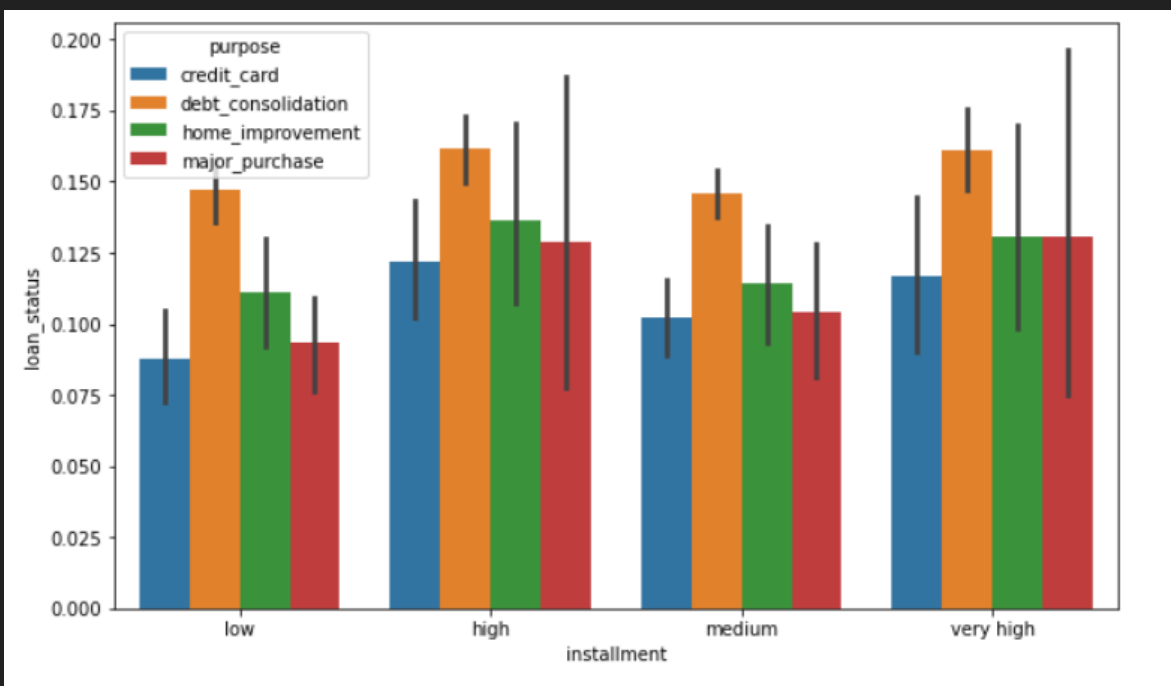
Figure 1

Figure 2

Fig.1 : As the amount of loan taken by a customer increases, the chances of him /her being a defaulter is high irrespective of the purpose of the loan.

Fig.2 : Interest rate seems to be directly proportional to chances of being a defaulter. Higher the interest rate, higher is the chance of him/her being a defaulter

Insights:
As the number of instalments opted by customer increases, higher is the chance of him/her being a defaulter.
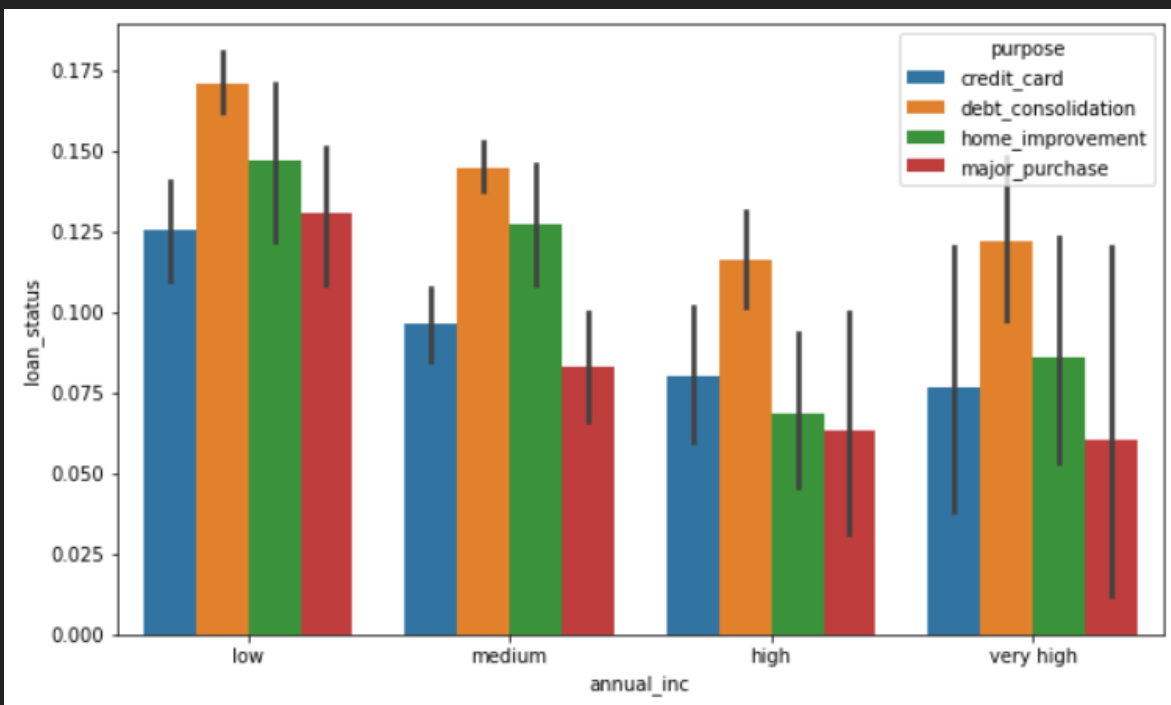
Lower the annual income, higher the chances of customer being a defaulter.

BIVARIATE ANALYSIS

Insights:

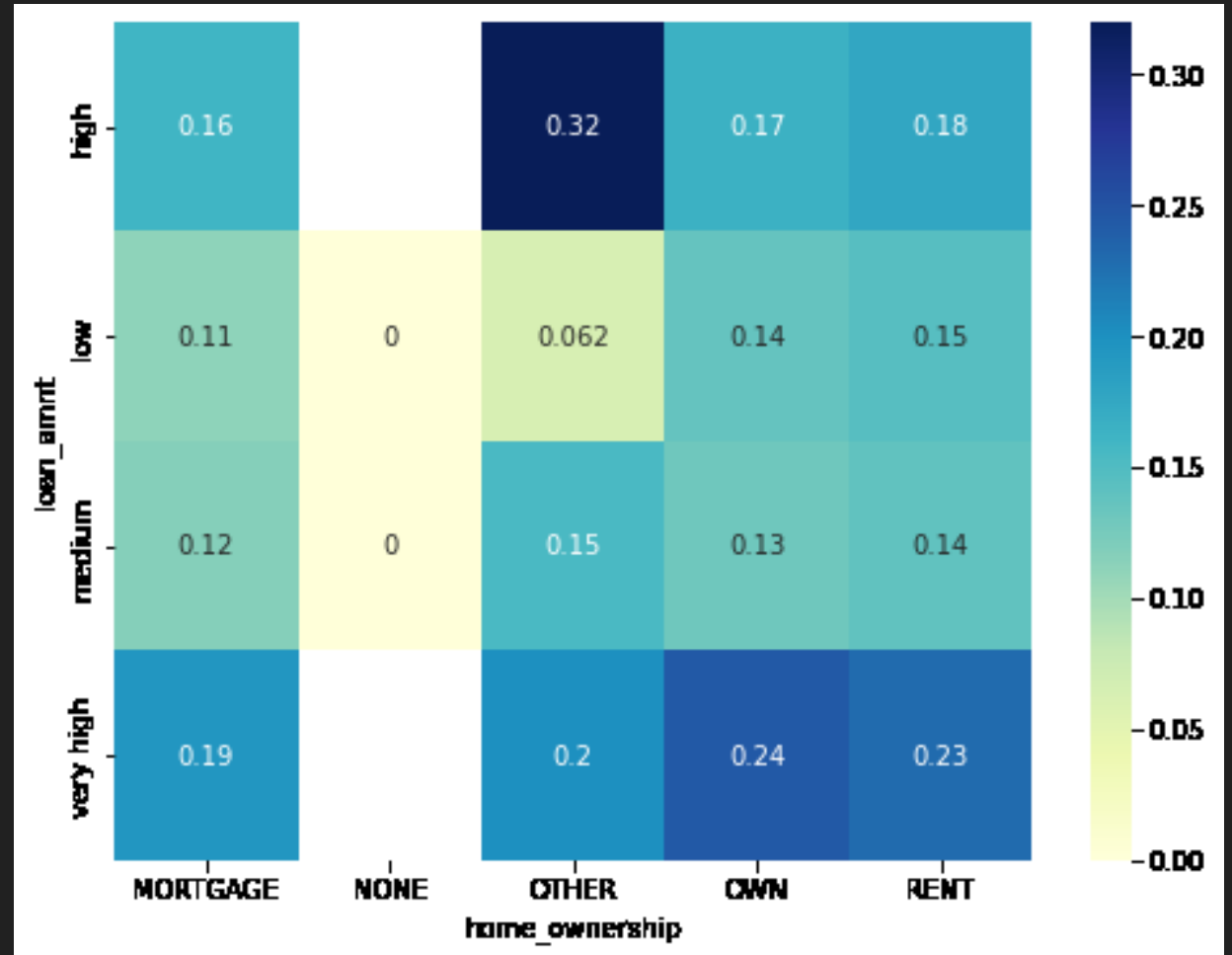Customers opting for very high loan amount irrespective of the type of home ownership have high chances of becoming defaulters.

Insights:

As per the above heatmap, The customers opting for 60 months loan period term irrespective of all purposes are having more chances of defaulters
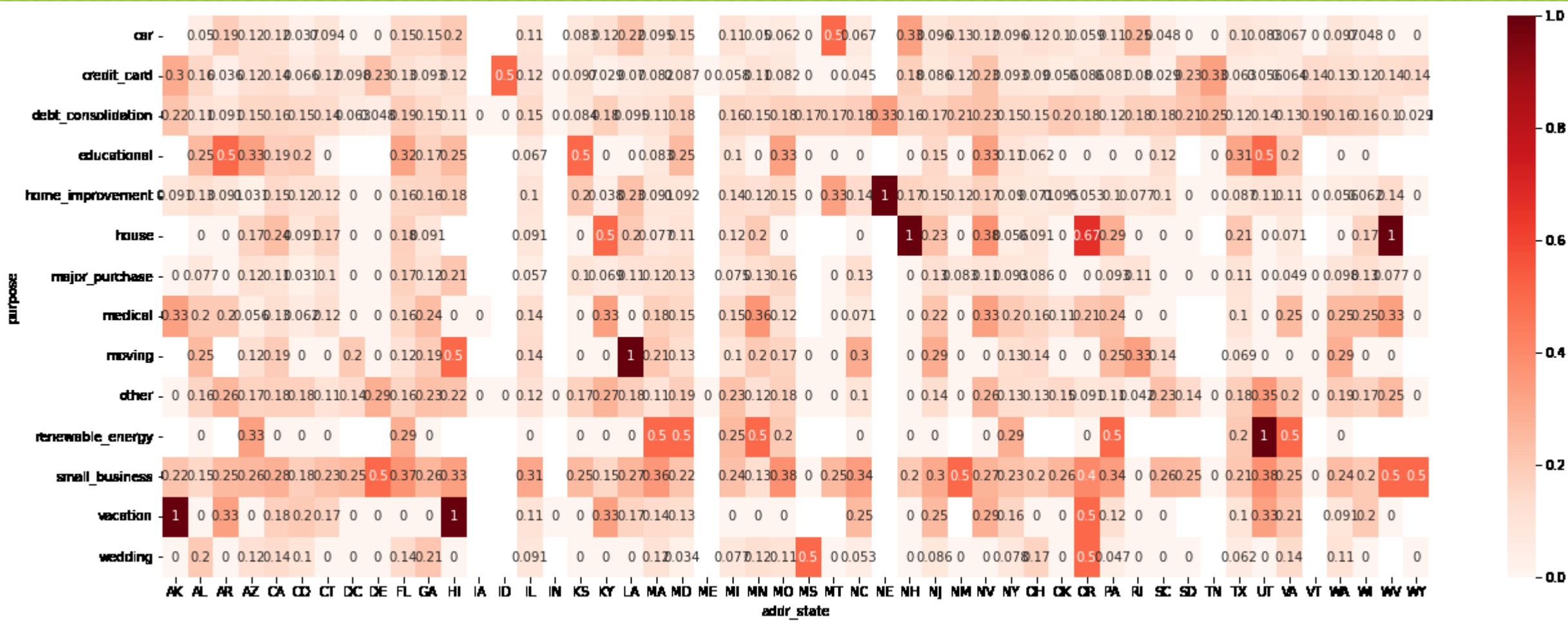
Insights:

As per the above plot, the darker the intersection of addr_state has with the purpose of the loan, the riskier the loan application is. Some of the examples are below:

- vacation loans in AK, HI, OR
- small business loans in DE, NM, WV, WY

- education loans in AR, KS, UT
- wedding loans in MS, OR
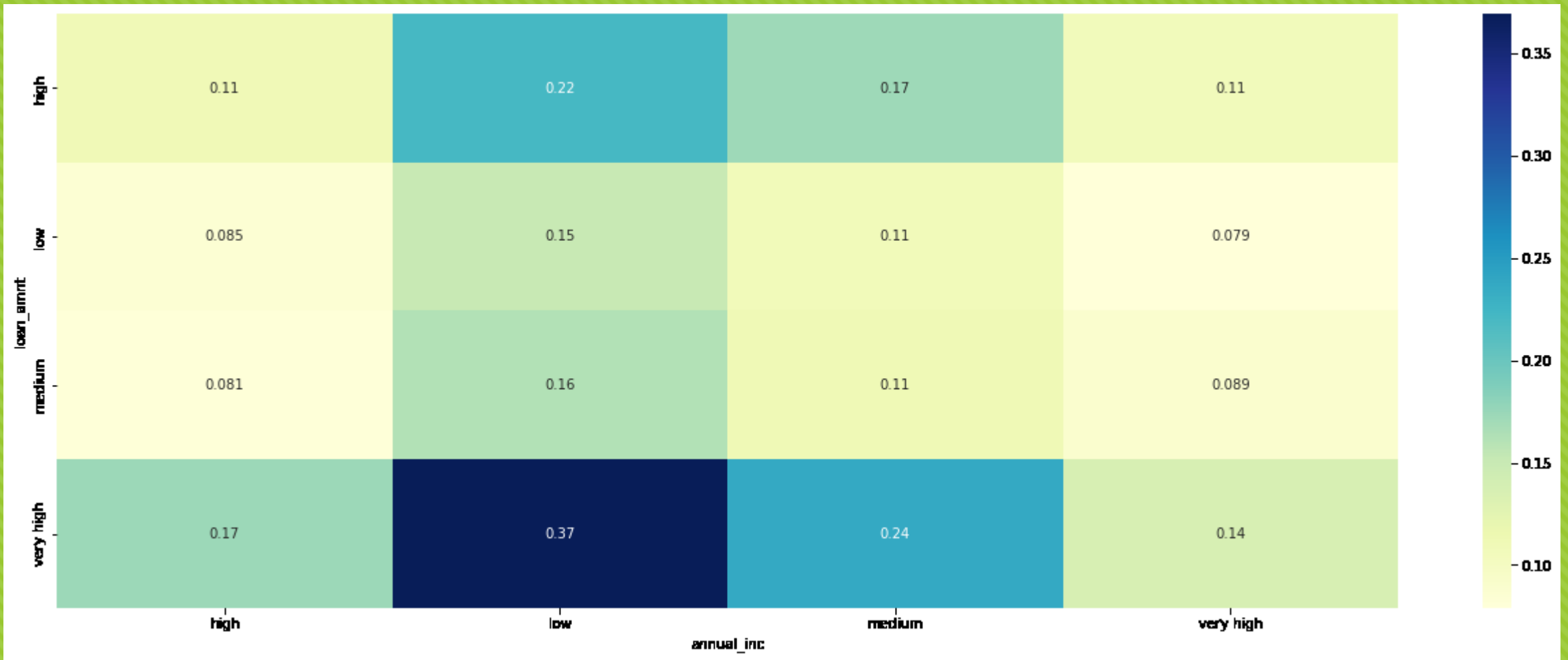
Insights:

From the above plot, we can infer that the customers having very low income but requesting very high loan amount are having higher chances of default rates.

Variables having combined impact on the Analysis

These variables on a combination have high risk:

- High loan amount & higher interest rates for lower income group
- High instalments and longer repayment term
- Home ownership (other) and loan purpose (car, moving or small business)
- Residential state and loan purpose
- Annual income and loan purpose

CONCLUSION

By the end of EDA for loan data set, the key factors influencing the defaulter's rate are as follows:

Minor Impact
Higher loan amount
Higher instalment amount
Lower annual income
Higher debt to income ratio (above 15%)
Applicant's address state (NV, SD, AK, FL, etc.)
Loan issue month (Dec, May, Sep)

Heavy impact
Higher interest rate
Repayment term (60 months)
Loan grade & sub-grade (F to G)
Loan purpose (small business, educational)

- Lending loan to customers with Grades and sub-grades beyond F is of higher risk and thorough verification of documents is required before lending loan.

RECOMMENDATIONS

- If customer is opting for the Repayment term of 60 months, the risk seems to be higher. However we can ask for some collateral deposit from the customer depending upon the amount of loan he is requesting for , to avoid losses for the Lending club.

- As the loan amount requested by the customer increases, the chances of him/her becoming defaulter increases. In this case the Lending club can do a thorough verification on the customer's annual income and the properties owned based on which "funded amount" and "funded amount investment" can be decided.

- If the Home Ownership status is "Other", analysis predicts it be having higher risk rate. So its recommended to investigate further on the property status.

RECOMMENDATIONS

- Loans taken with purpose mentioned as Debt consolidation possesses higher risk. So, its recommended to partially process the loan amount.

Example: 35-40% of the annual income of customer can be granted as the loan funded amount.

- If the DTI ratio is high and Annual income is low, it is recommended to reduce the funded amount investment for a particular customer.

Thank you!!

Team Details:

Group Facilitator: Shreeraksha S

shreeraksha.s123@gmail.com

Group Member: Shiva Sharaj H N

shivasharaj1ga16ee047@gmail.com