Lending Club Case Study

Problem Statement

Lending Club is the largest online loan marketplace and as any other organisation that lends money, it is faced with a tough decision of whether to approve the loan or not. Hence, the company wants to carry out credit risk analysis - which is a detailed inspection and review of the borrower's capability of repaying the loan back.

This paper is about identifying the driving factor that will help to quantify the credit risk of borrowers. Our first step is to understand the data which will be followed by cleaning the data and pre-process it before analysing it.

Data Understanding

The data consist of two main groups which are the customer attributes and the loan attributes. The customer attributes are further sub-categorised into customer loan behaviours and customer demographics.

Customer Attributes		**Loan Attributes**
Customer Demographics	Customer Loan Behaviors	desc
addr_state	collection_recovery_fee	funded_amnt
emp_length	delinq_2yrs	funded_amnt_inv
emp_title	earliest_cr_line	grade
home_ownership	inq_last_6mths	installment
pub_rec	last_credit_pull_d	int_rate
pub_rec_bankruptcies	open_acc	issue_d
zip_code	out_prncp	loan_amnt
	out_prncp_inv	loan_status
Financial Health	recoveries	purpose
annual_inc	revol_bal	sub_grade
dti	revol_util	term
	total_acc	title
	total_pymnt	url
	total_pymnt_inv	verification_status
	total_rec_int	
	total_rec_late_fee	
	total_rec_prncp	
	last_pymnt_amnt	
	last_pymnt_d	

In this study, the customer loan behaviour attributes will not be taken into account as these attributes are not available at the time of loan application and hence cannot help in the prediction. The target variable in this dataset would be loan_status.

Data Cleaning and Manipulation

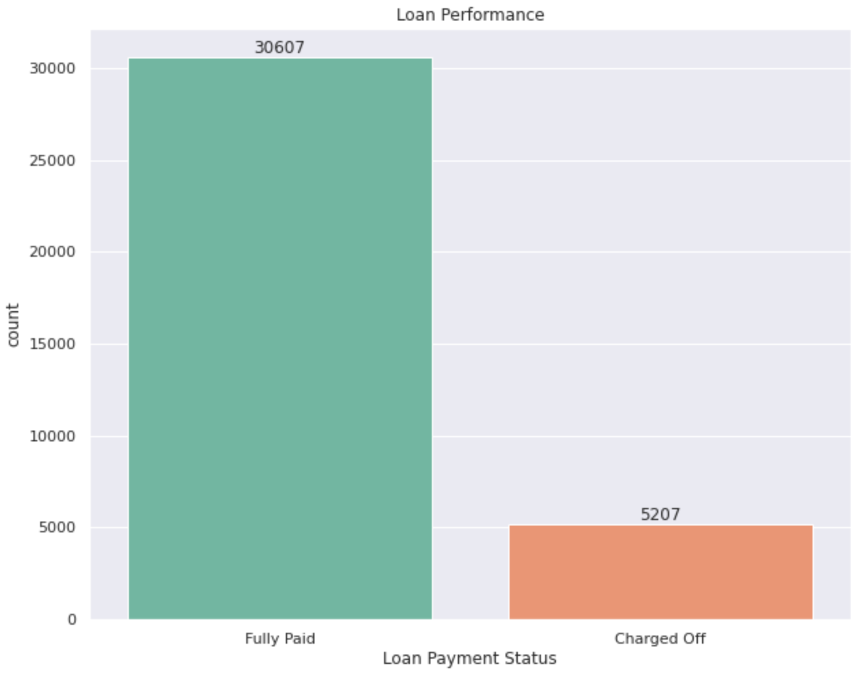
Data with a very high missing percentage were dropped from the dataset and new variables to help with the analysis were derived.

The following variables are derived because they are an important factor of the borrower's financial health. They are a strong indicator of the borrower's capability to repay the loan. Given that the loans are paid back in instalment, it is vital to know their monthly earning and spending. If the current debt plus the instalment exceeds the monthly income, it is safe to say that the borrower will have a hard time repaying the loan.

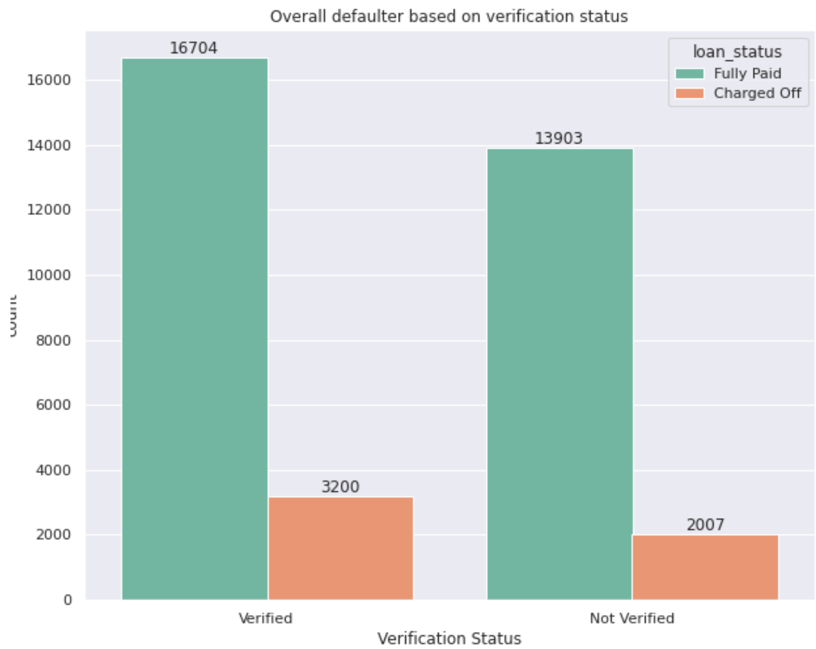
1. Borrower monthly income
2. Current monthly debt being paid by the borrower
3. Total debt including the loan instalment of each borrower
4. The DTI of each borrower in respect to the total debt

Data Analysis

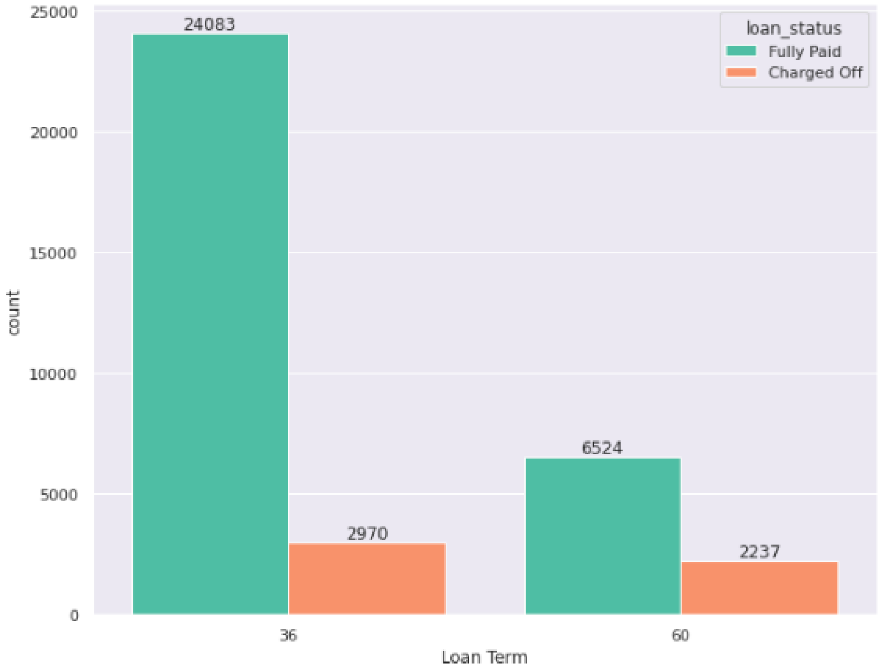
Univariate Analysis



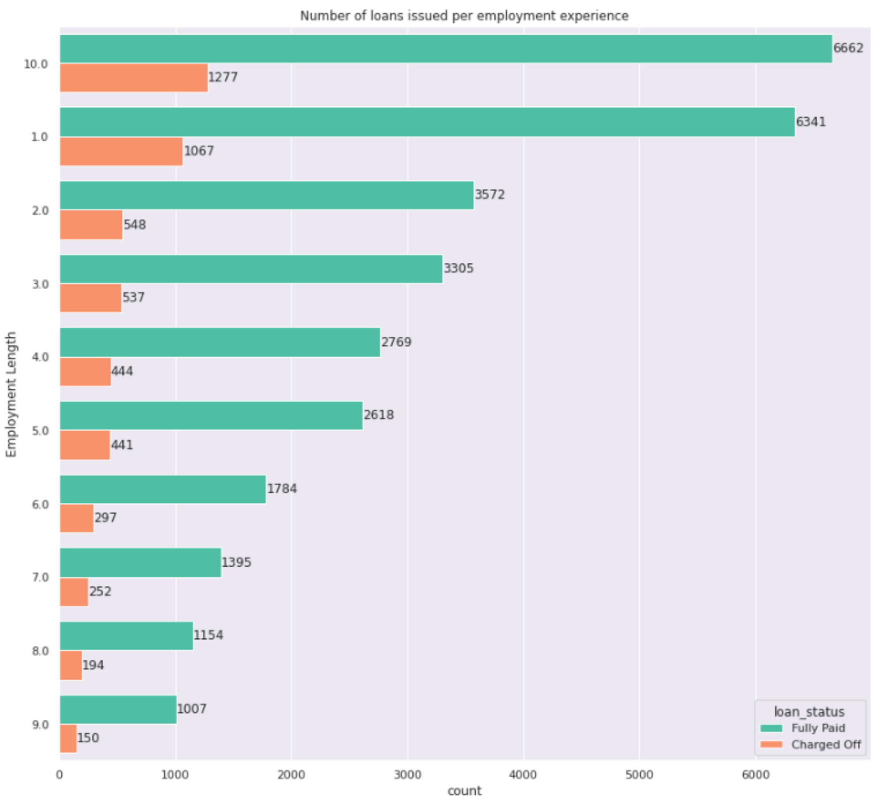
From the chart above we know that 85% of all previous loans were fully settled while 15% were defaulted. This is the variable that eventually will need to be predicted. We have to analyse the trend of those 15% and determine the driving factors.



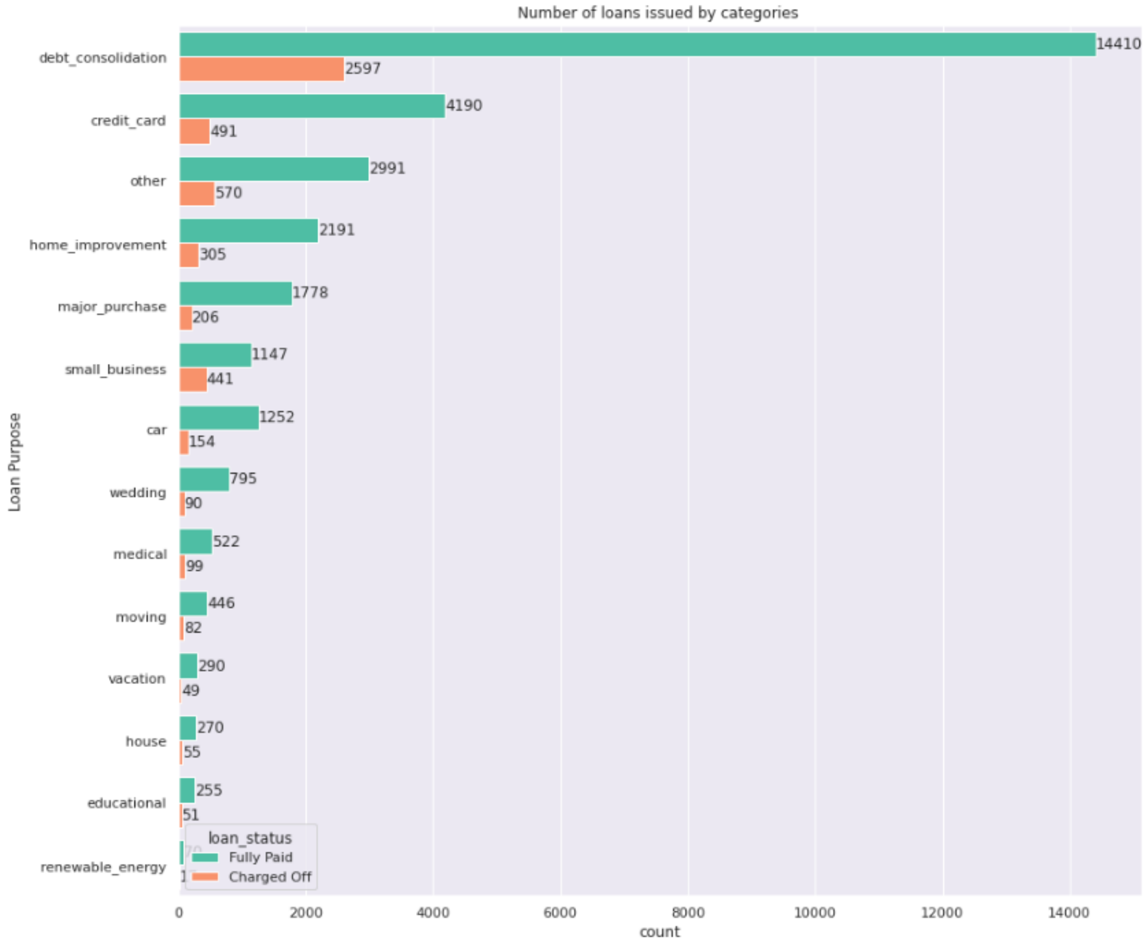
This illustrates that Income verification status is not really a relevant factor as most of the loans under each category were fully paid.



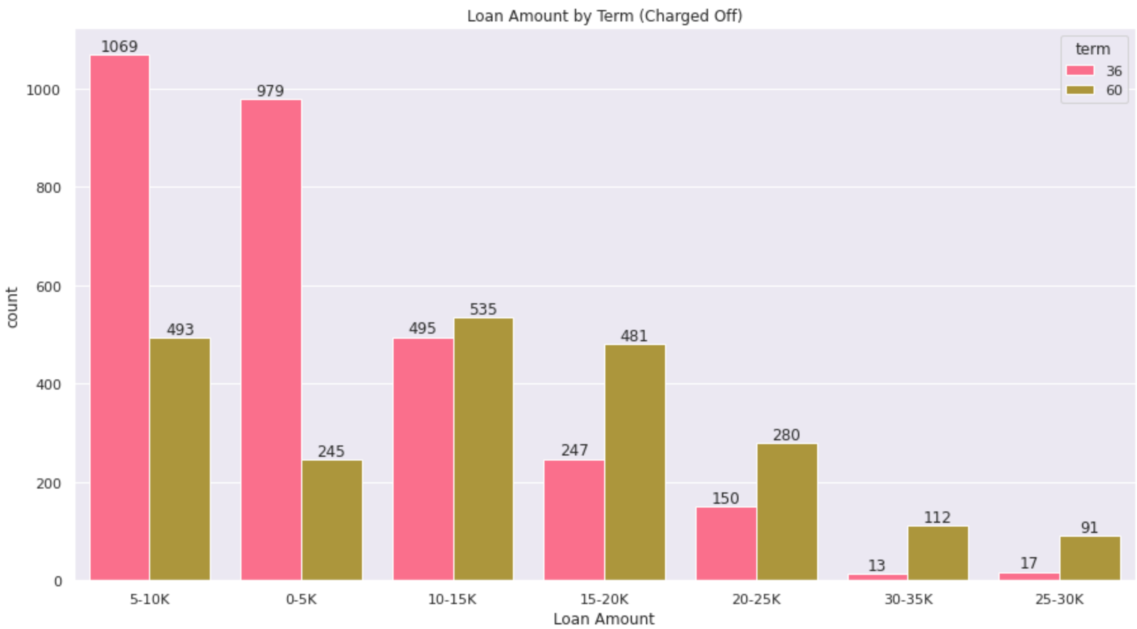
It is clear that 76% of loans issued were under 36 months term while 24% were under 60 months term. The 60 months term loans have a higher rate of defaulters which is 26 while the 36 month term has only 10% of borrowers who defaulted.



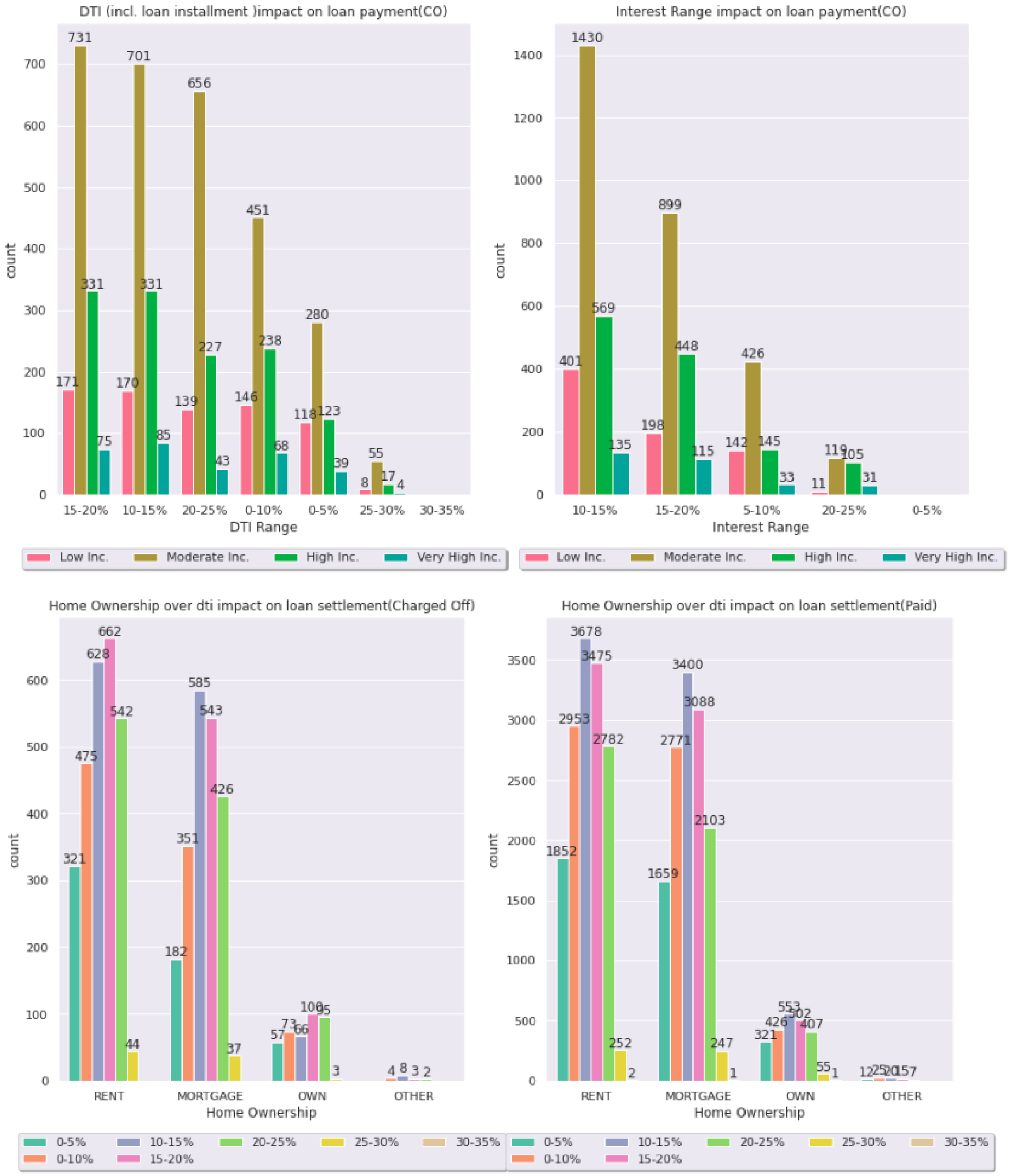
Most of the loans issued were by people having over 10 years of experience as well as those with less than 1 year. The purpose of most loans were for debt consolidation, credit card and home improvement.



Bivariate Analysis

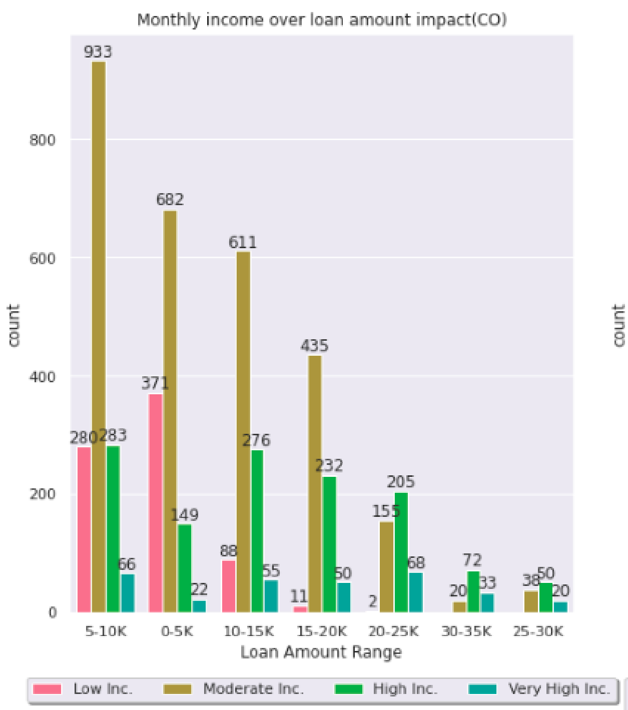


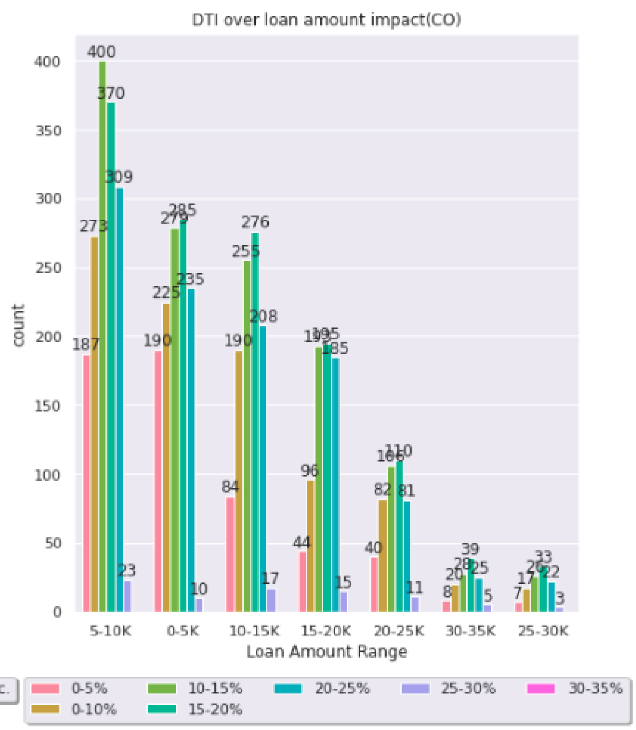
People who have borrowed between a range of 5 to 10 thousands have failed to repay the loan mostly being with a 36 months term.



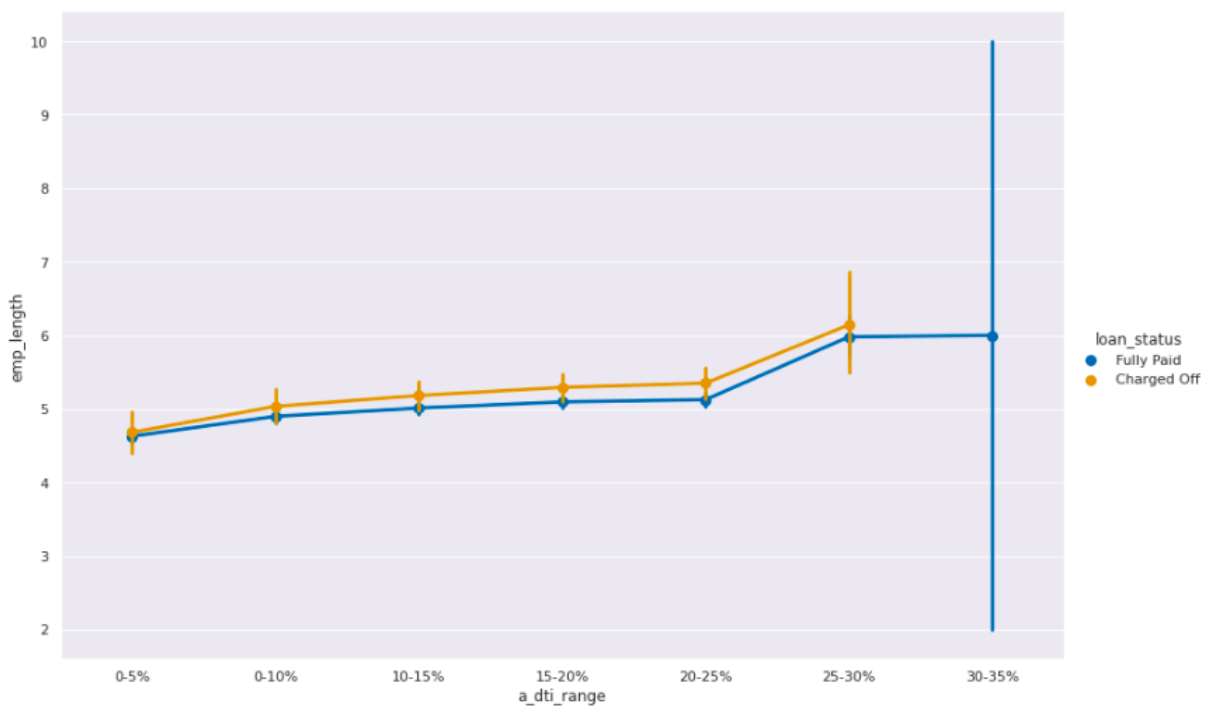
Monthly income was calculated as it is an important indicator of whether the applicant can pay the monthly instalment. Given that monthly income is a continuous value, it was grouped into 4 bins. Low income ranged between 0 - 2500, Moderate income ranged between 2500 to 5500 and High income ranged between 5500 to 9000 while very high income was between the range of 9000 to 12500.

Total dti was also calculated and it factored in the monthly instalment of the LC loan. As seen above, people whose DTI ranges between 15 to 20% and salary is between 2500 and 5500 tend to default the most. That could mostly be because of their mortgages. As depicted, very few defaulters actually own a house.

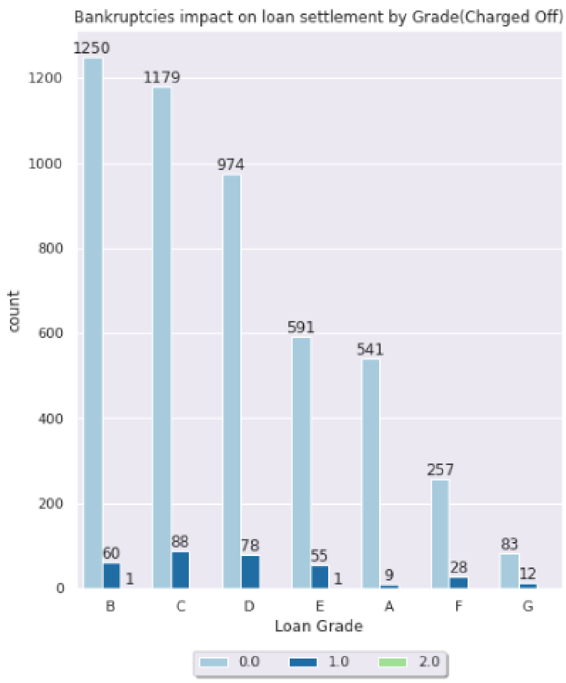


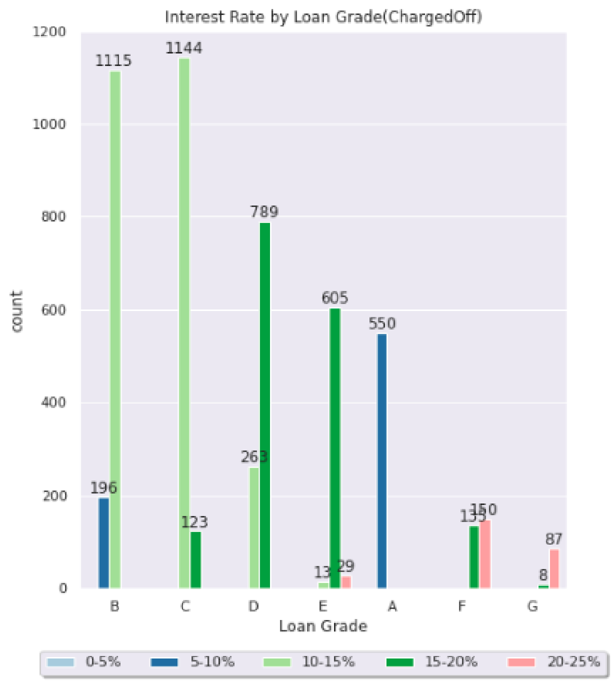


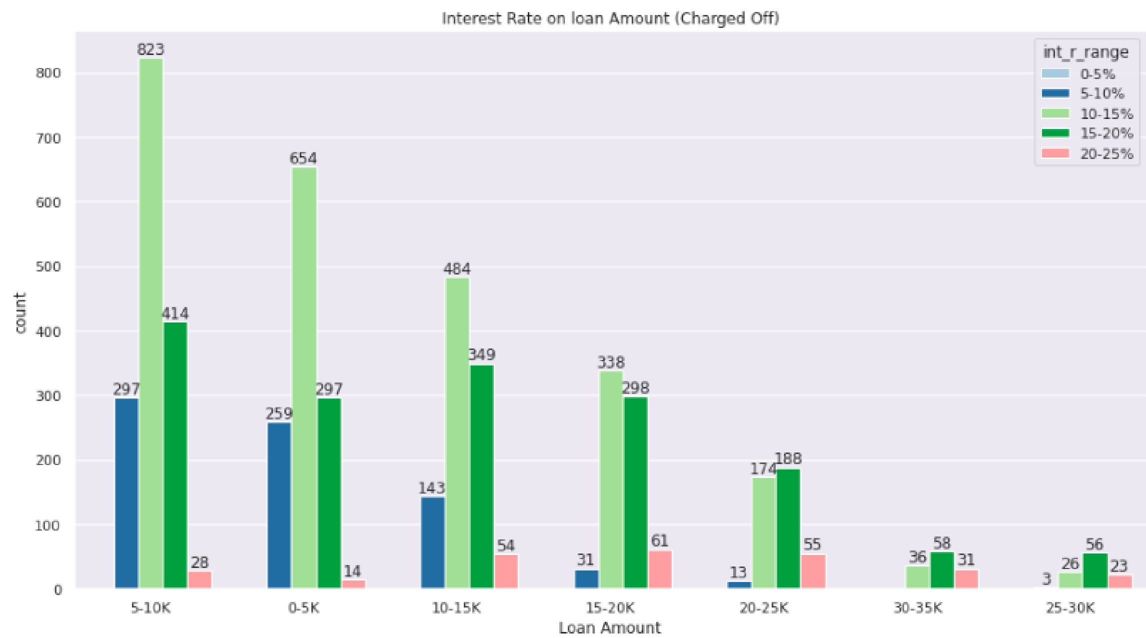
People with moderate income seem to have a higher default rate across all ranges of loan amounts. Loan amounts ranging from 5 thousands to 10 thousands have most defaulters. The DTI helps to give us an insight on those who have defaulted within the mentioned range, especially those who have a DTI of 10 to 15%.



People with 5 to 6 years of employment experience and a DTI of 20 to 30% have defaulted the most.

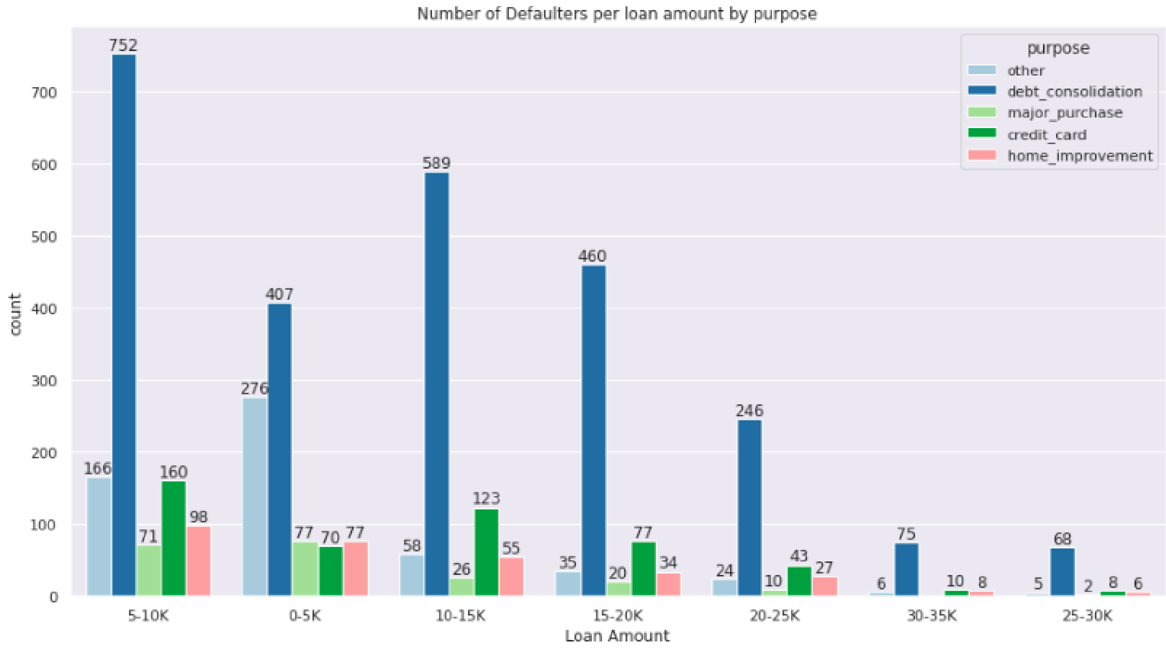




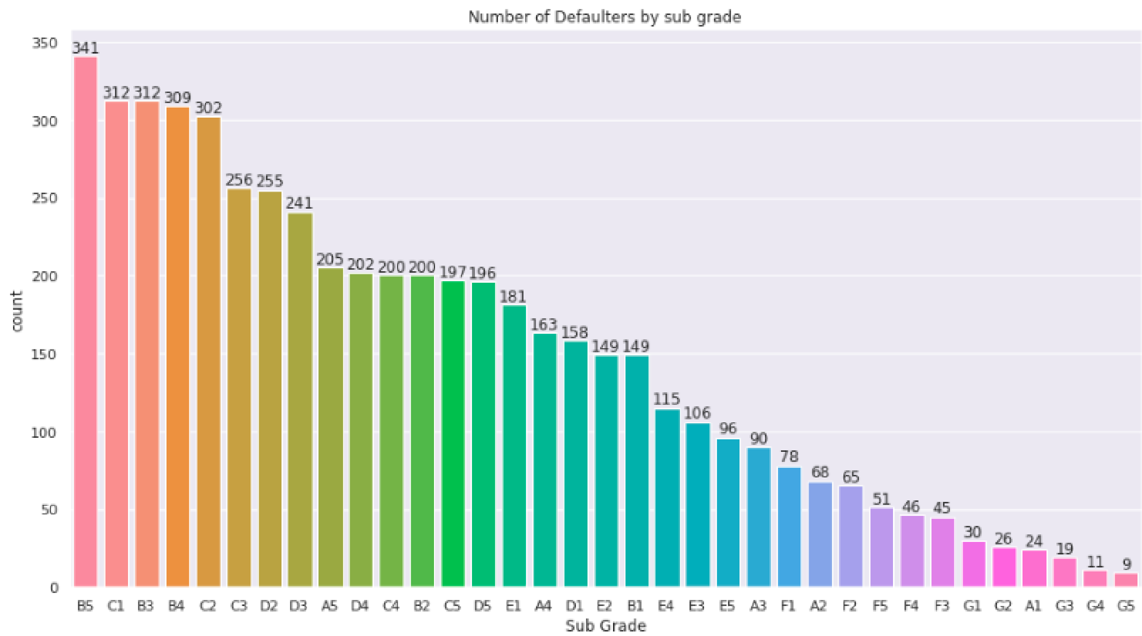


The next insight from the graph is that loans with Grade B and C have higher defaulter and that is because the interest rates for these grades are high. The most common interest rates for grade B and C are 10-15%. We have to consider that Grade B had more borrowers that had 1 public bankruptcy record. The amount of borrowers with 1 public bankruptcies record for Grade B is 473 and out of this, 12% have defaulted.

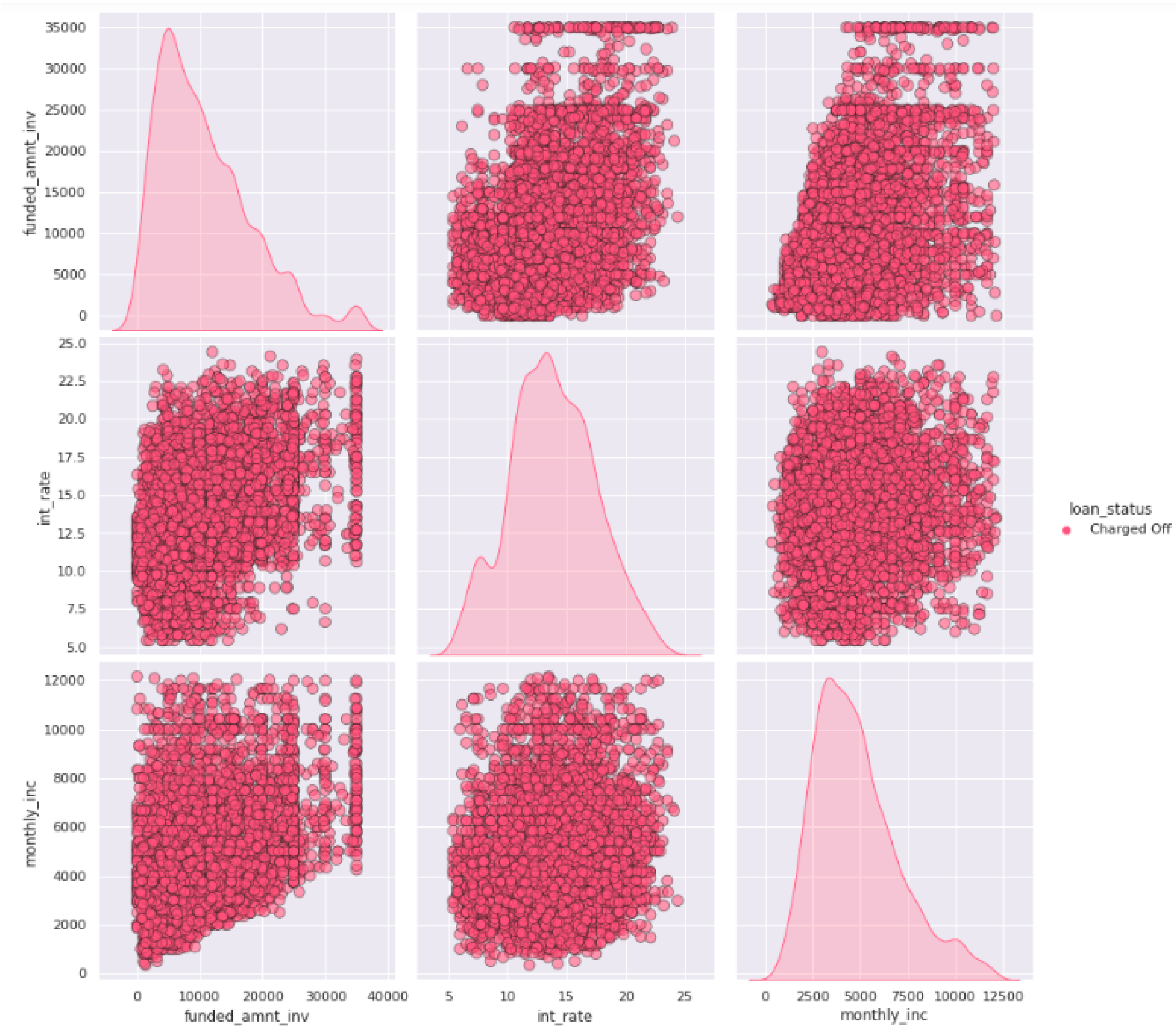
Moreover, the most defaulters have taken out a loan amount between 5 thousand to 10 thousands and the purpose is debt_consolidation.



If we analyse each loan that was charged off by subgrade, we notice that B5, C1, B4, C2 and C3 are the subgrades with most defaulters. Thus confirming that loans with Grade B and C always yield more defaulters.



Finally, we try to understand the relationship between loan amount funded, interest rate and monthly income.The higher the interest rate, the higher the probability of default.



Summary of Insights

The driving factors to identify borrowers who are most likely to default are

1. Loans with a 36 months term whose amount ranges between 5 thousand to 10 thousand.
2. DTI (including instalment) range between 15 to 20 percent and salary range between 2500 and 5500
3. Working experience between 5 to 6 years and DTI range between 20% to 30%.
4. Loan with Grade B and Grade C and subgrades B5, C1, B4, C2 and C3
5. At least 1 Public bankruptcy record.
6. Interest rate of 10 to 15 % for loans - mainly loans of Grade B and C.
7. Salary ranging from 2500 and 5500 and have rental property or mortgages.
8. Purpose of loan being debt consolidation.