GRAMENER CASE STUDY

Group Members :

1. Oshin Sharma
2. Chirag Sharma
3. Pranav Haldar
4. Ashutosh Kumar Singh

Overview

Background of Problem

A **consumer finance company** which specializes in lending various types of loans to urban customers needs to make a good business decision for **loan approval** based on the applicant's profile in order to enhance the profits.

Company Business Objectives

- The company wants to understand the **driving factors behind loan default** i.e. the variables which are strong indicators of default.
- The company intends to utilize this knowledge for its **portfolio and risk assessment**.

Aim

The aim of study 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.

Problem Analysis

The data contains details of about 4 years of loans that have been issued by the bank. The loans have been categorise into:

1. Fully Paid
2. Current
3. Charged Off

Based on their repayment status.

Assumptions:

1. Since on-going loans can't give insights to loan defaults. The current loan data has been ignored for the analysis
2. Removed outliers based on annual income and loan amount for the analysis to not be skewed

Problem Solving Methodology

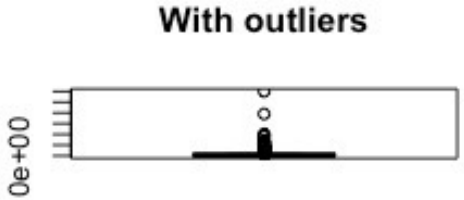
To ascertain the driving factors behind loan default, Study group has carried the activities for Cleaning out the data , prepared it and then carried out EDA to

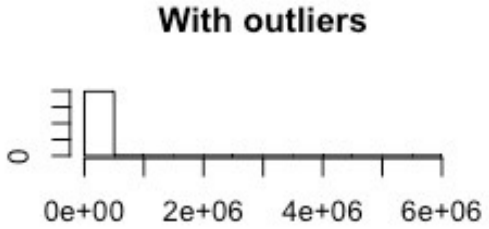
- Identify important variables/personal history parameters of loan/consumers
- Visualize the distribution of loan amounts for understand the lending trends over time
- Ascertain reasons for the default
- Study and visualize the demographic nature of loan defaults
- Compute and validate correlations between loan status and indicators for defaulting consumers

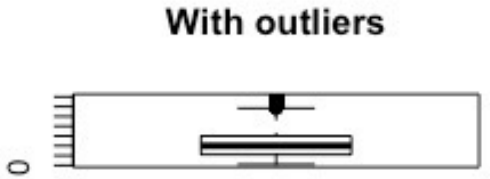
Steps Followed

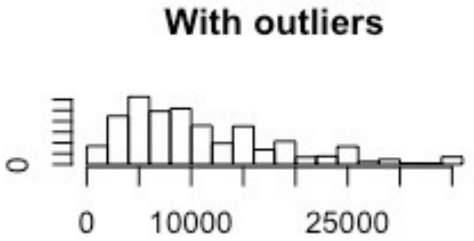
- Gather data for analysis
- Clean and Format the data for readability
- Extrapolate and Derive Metrics
- Filter out the outliers
- Filter data set on loan status
- Perform Univariate and Bivariate analysis on the data
- Visualise using plots and observe the patterns

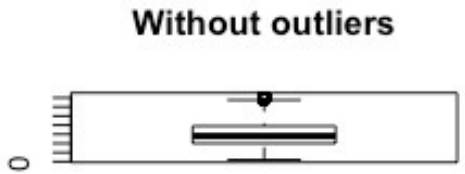
Outlier Analysis

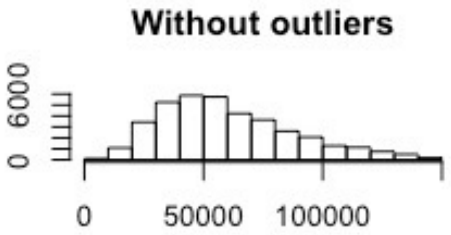


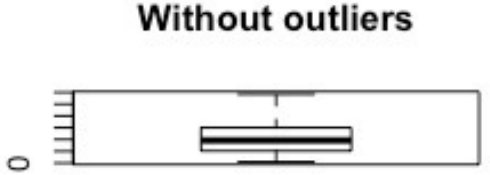


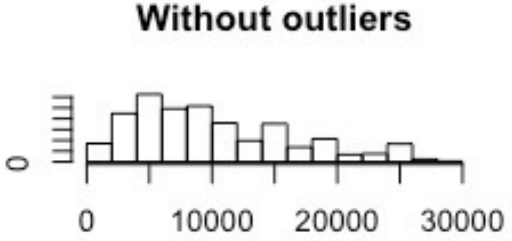










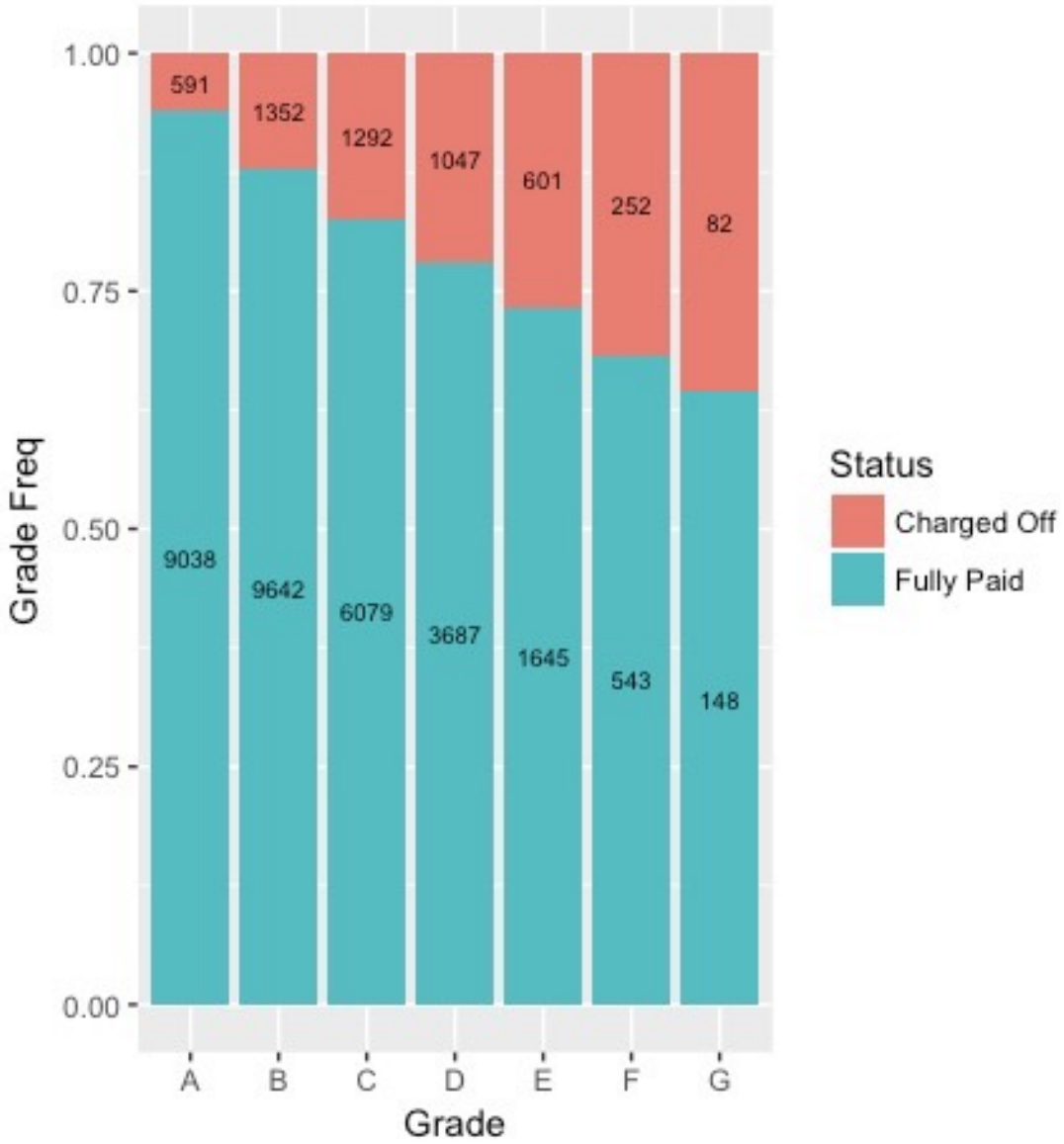


Annual Income

Loan Amount

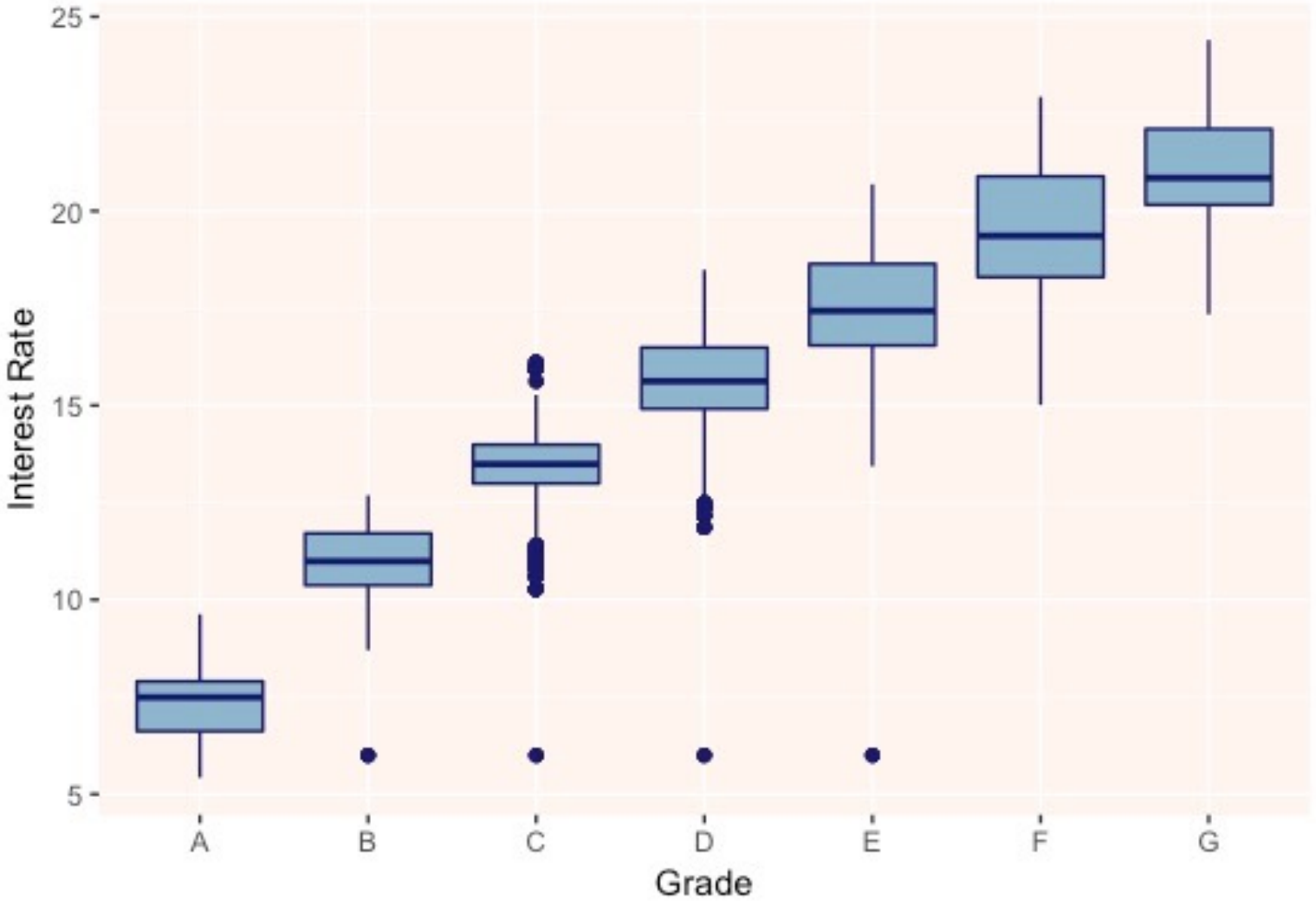
Removed outliers for Annual Income and Loan Amount to achieve close to normal distribution of the sample

Loan Grade and Loan Status



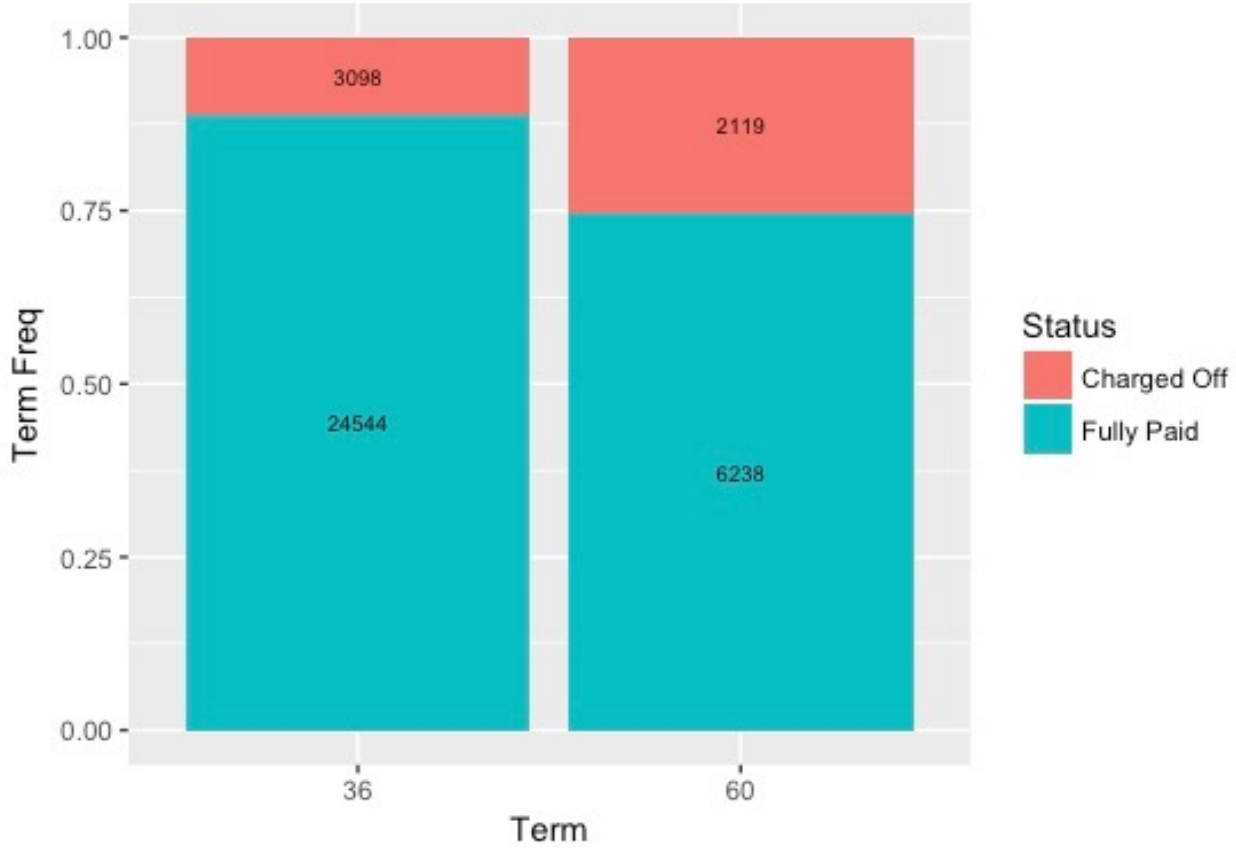
- It can be observed from the plot that the later loan grades have a higher proportion of loan default %age
- Thus it can be established that a later loan grade is more risky in comparison to an earlier loan grade with respect to the proportion of charged off loans

Loan Grade and Interest Rate



- Higher interest rates are being charged for later loan grades in line with high risk high gain theory thus higher chances for an loan approved at a higher rate to turn out to be charged off because of higher risk involved in such loans
- Thus banks should look at providing more loan to less risky grades at marginally higher or same rates if they want to reduce their risk of having to charge off a loan
- Banks are anticipating huge profits in risky loans given they pay it off completely but are facing even higher losses after having to charge them off because of defaults in loan repayment

Loan Term and Loan Status



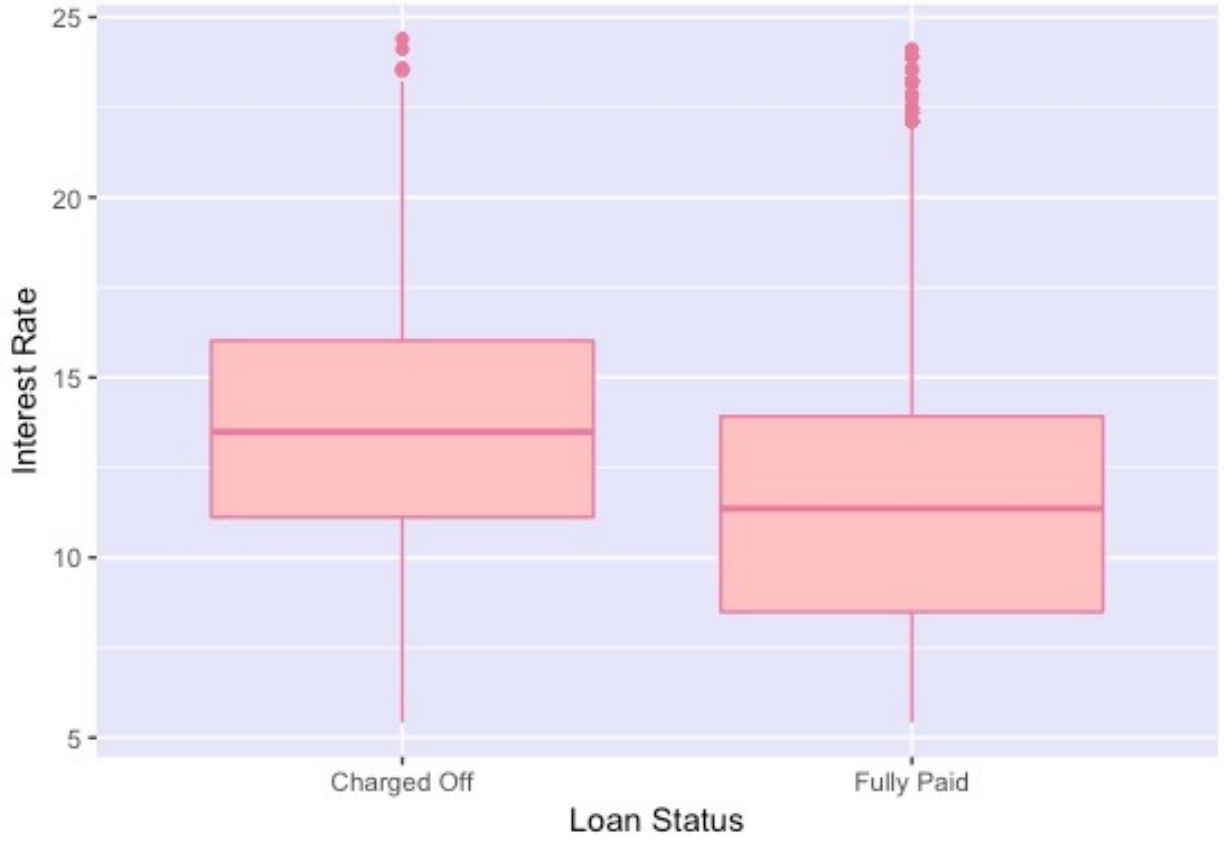
- There are two loan terms 36 months and 60 months; 36 month can be classified as short term loan and 60 months as long term on a relative scale
- We observe that the term of the loan plays a huge role in the final status of the loan as we found that a short term loan is far more likely to be re-paid in full
- Whereas the loan status is more likely to be charged off for a long term loan
- Also it is clearly evident from the plot that long term loans have higher proportion of defaults in comparison to short term loans

Loan Term and Loan Grade

		A	B	C	D	E	F	G	Row Total
36	count	9242	8855	5496	3031	782	187	49	27642
	%age	**33.43%**	**32.03%**	**19.88%**	**10.97%**	**2.83%**	**0.68%**	**0.18%**	**100%**
60	count	426	2425	2106	1901	1609	650	192	9309
	%age	**4.58%**	**26.05%**	**22.62%**	**20.42%**	**17.28%**	**6.98%**	**2.06%**	**100%**
Overall	count	9668	11280	7602	4932	2391	837	241	36951
	%age	**26.16%**	**30.53%**	**20.57%**	**13.35%**	**6.47%**	**2.27%**	**0.65%**	**100%**

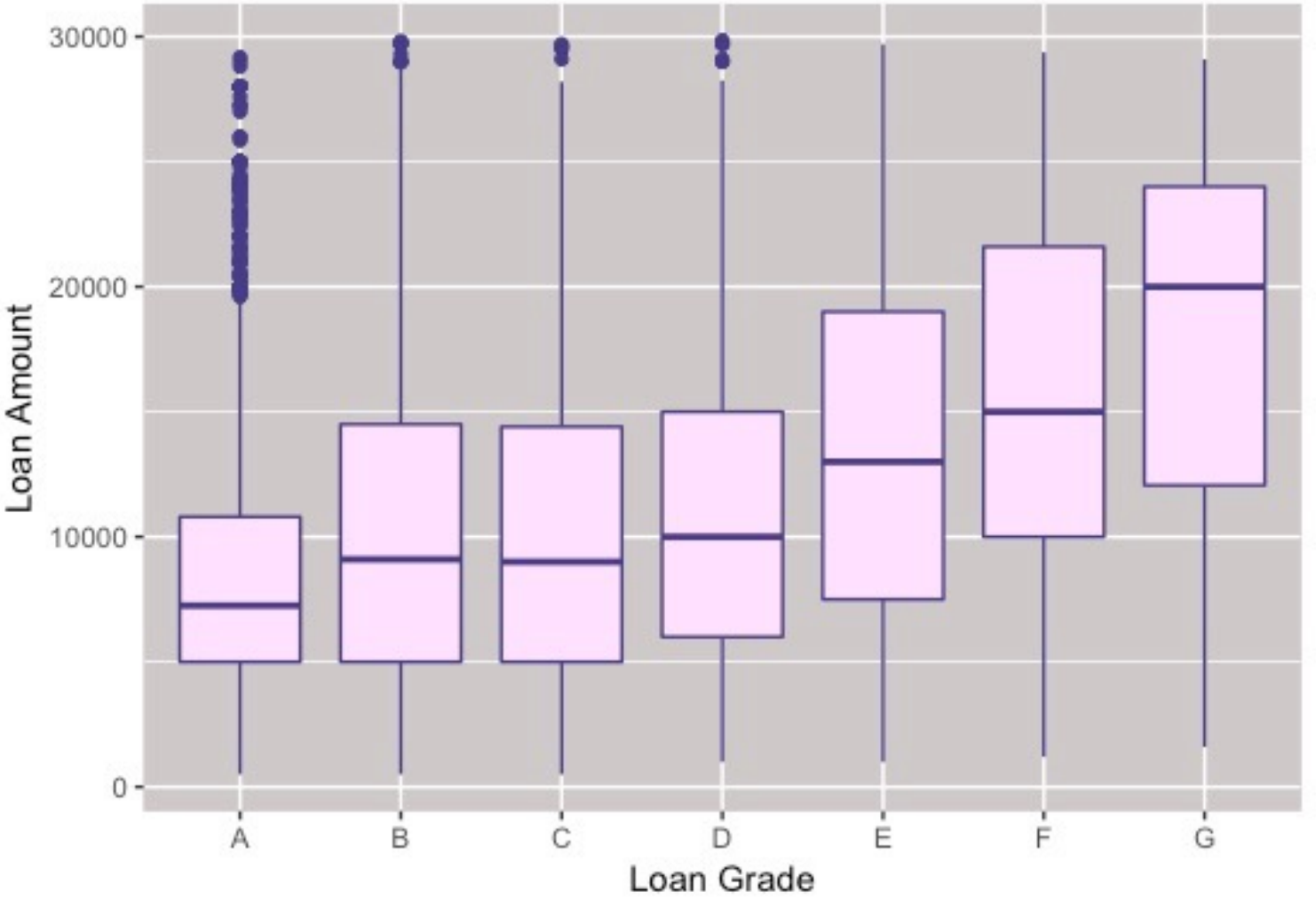
- High risk grades as established earlier are being given more long term loans whereas less risky grades are being given mostly short term loan thus making the long term prospects riskier and also larger in magnitude

Interest Rate and Loan Status



It can be observed from the above plot that charged off loans have significantly higher interest rates

Interest Rate and Loan Amount



- To maximise the profit even the higher loan amounts are being provided to later loan grades by the banks along with high interest rates
- This multiplies the risk of the bank as per the trends observed

Verification Status and Loan Status

		Charged Off	Fully Paid	Row Total
Not Verified	count	2098	14214	16312
	%age	**12.86%**	**87.14%**	**100%**
Source Verified	count	1348	7744	9092
	%age	**14.83%**	**85.17%**	**100%**
Verified	count	1771	8824	10595
	%age	**16.72%**	**83.28%**	**100%**
Overall	count	5217	30782	35999
	%age	**14.49%**	**85.51%**	**100%**

- There seems to be something not correct with the verification process because supposedly if the loan has been given after a proper verification process then the chances of it being charged off should be lesser in comparison to the unverified loans
- But as per the observations charged off loans are higher in proportion for verified loans (16.72%) and lesser for unverified loans (12.86%) when compared to the overall proportion of charged off loan (14.49%)
- Its an hypothesis which the bank should check at their end to verify the observation from data

Conclusion

- Parameters that were observed to have correlation with default loans based on the data provided are:

    1. Loan Term
    2. Interest Rates
    3. Loan Amount
    4. Loan Grade
    5. Verification Status (Anomaly)

- Using the above parameters we should be able predict with a fair probability if the current loan is going to be charged off or not

- This could be implemented by the Loan Company to implement a risk analysis practice in the assessment process