



GRAMENER CASE STUDY

SUBMISSION

Group Name:Group20Three

1. Deepankar
2. Arun
3. Sunil
4. Subodh



Business Understanding

Consumer Finance company specializes 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

AIM

- 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.

Ø Two types of Decisions that could be taken by the company when someone applies a loan:-

- **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
 - ✓ **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
 - ✓ **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'.
 - ✓ **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
- **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



Business Objective

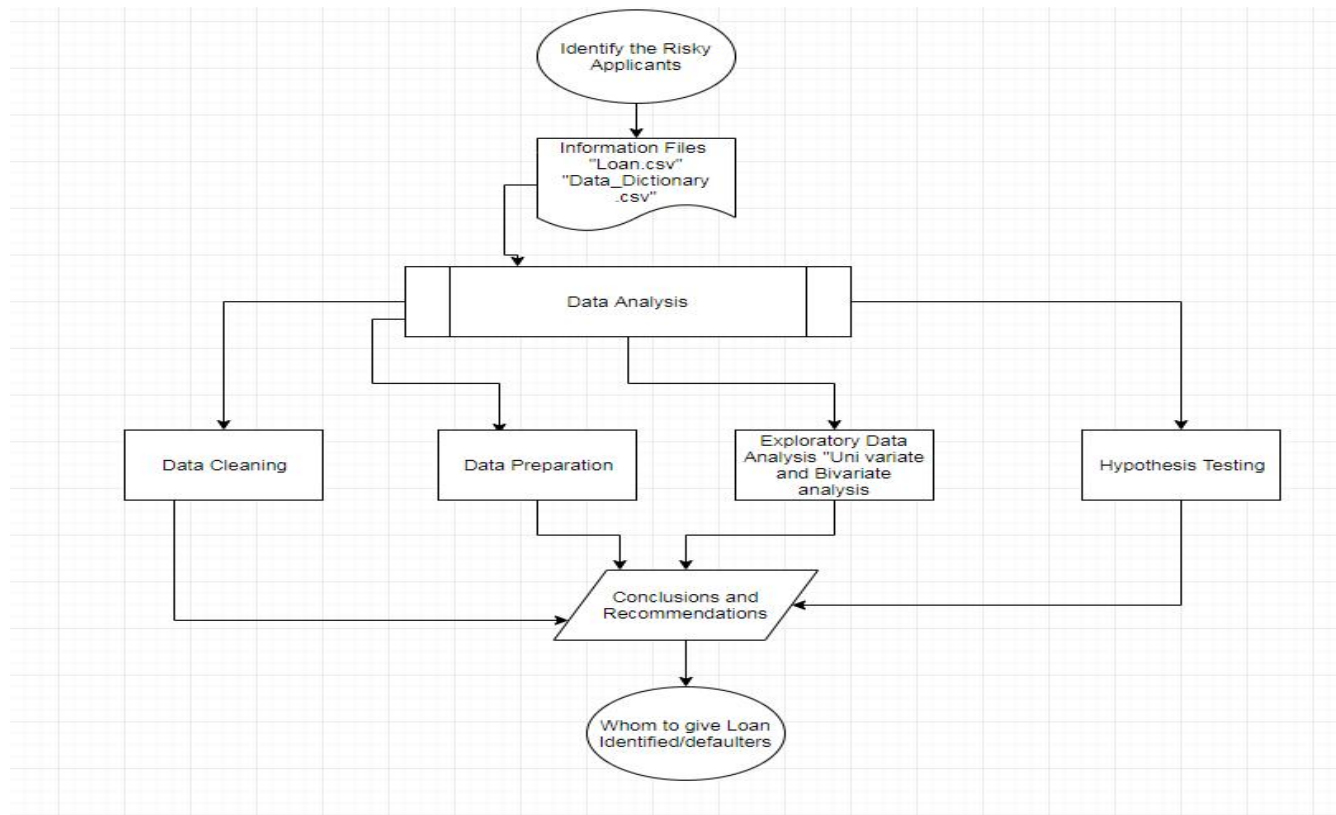
This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface

CREDIT LOSS:- The largest source of Financial loss is to lend loans to “risky” applicants is known as credit Loss. The 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'.

Main Objective:-

- Identification of risky applicants using EDA
- Understand the Variables which are strong indicators of Default

Problem solving methodology



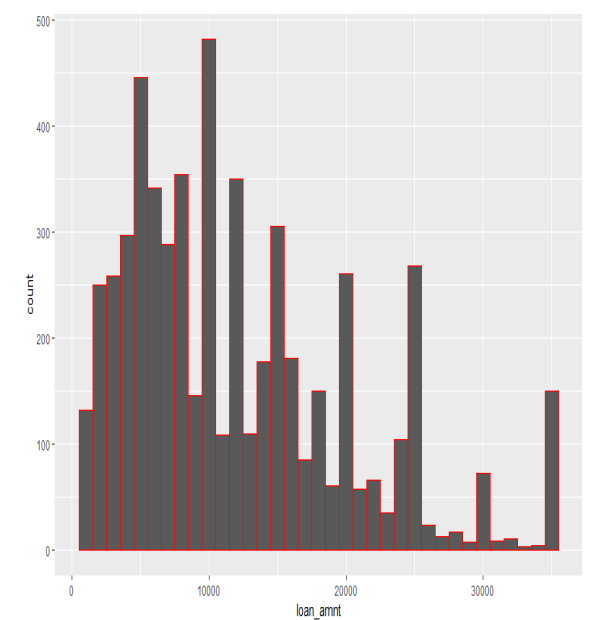
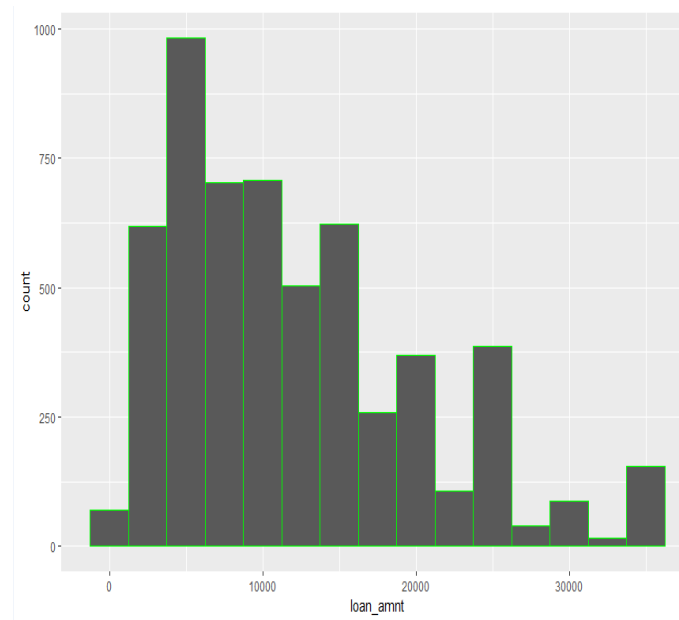
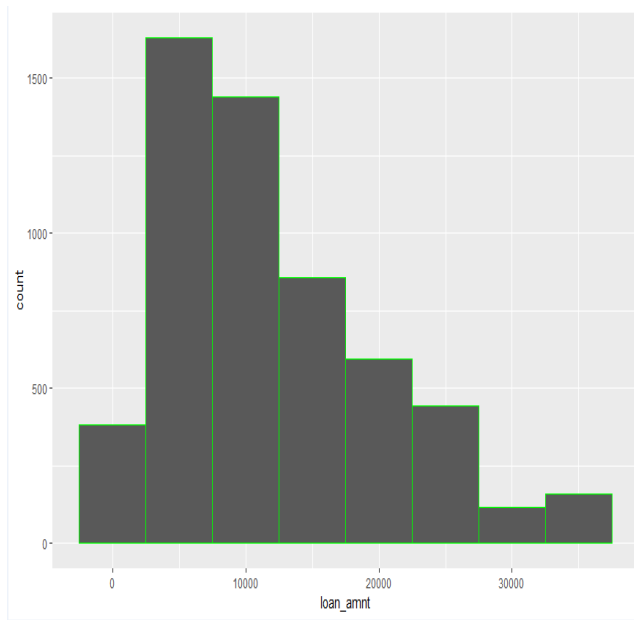


Analysis

Following Steps were involved for Data Analysis onto the platform “RSTUDIO”:-

- Import Data:- Import data from the file “Loan.csv”
- Cleaning the Data:- Correcting the data Collected, as well as checking and correcting of the “NA” values
- Identifying the duplicates with the help of “ID” as unique variable, removing the blanks rows from our driving variables
- Identifying what all parameters can be required for analysis and what all are not worth for analysis
- Export Data:- Tableau and R is used for plotting the data.

Results- EDA- Univariate Analysis



Ø Analyze loan_amnt range for which it is defaulted

Observation: Most defaults are with loan value from 5000 - 15000

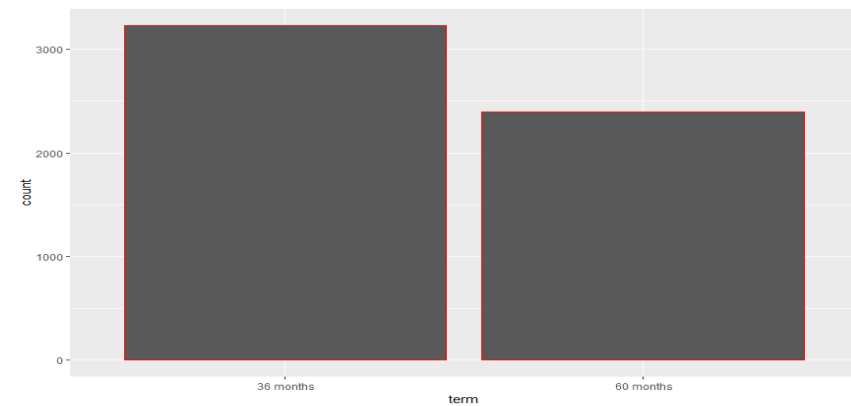
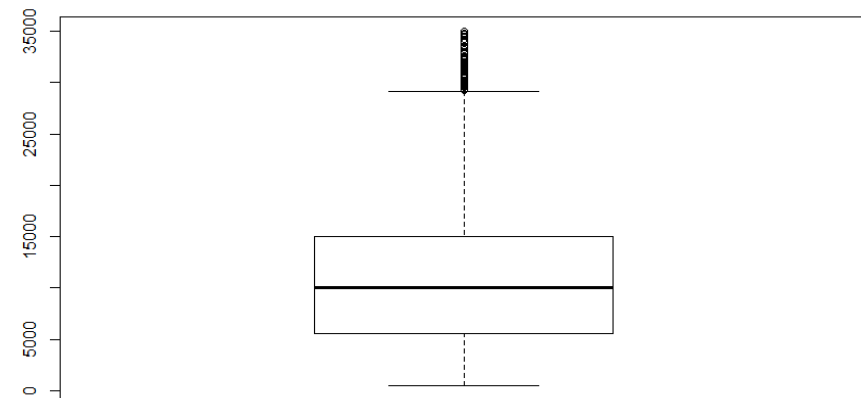
Results- EDA- Univariate Analysis

Ø Checked for the loan amount

Observation : Outliers present

Ø Tenure loans defaulted the most

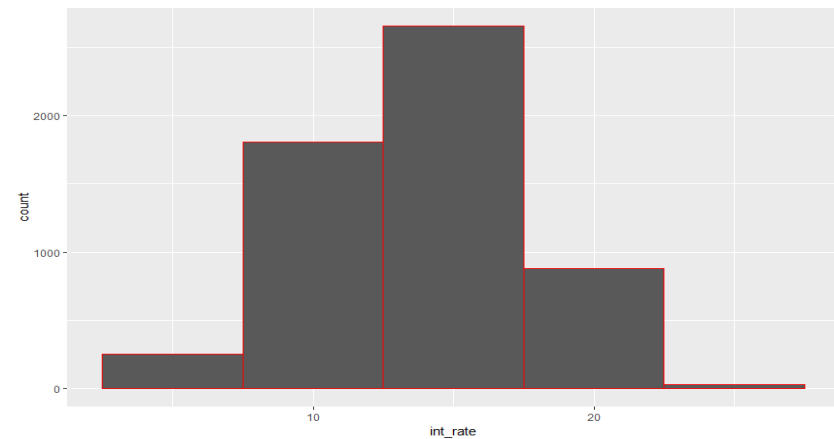
Observation : More defaults in loan with tenure '36 months'



Results- EDA- Univariate Analysis

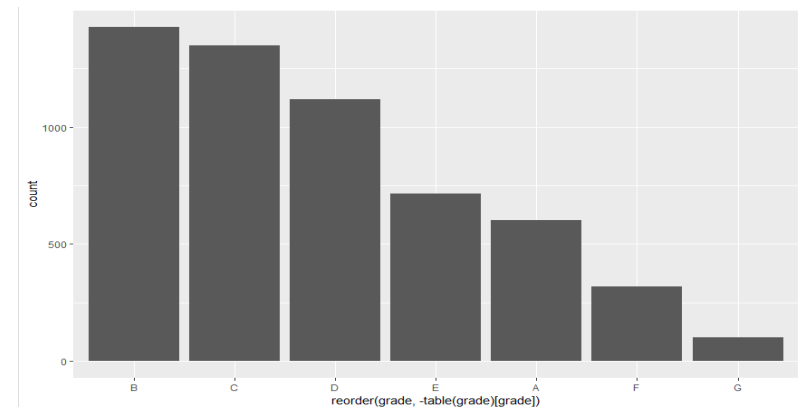
Ø impact of interest rate
on loan defaulting

Observation : interest
rates between 10-15%
Interest rate are majority
defaulters



Ø Grade & sub-grade

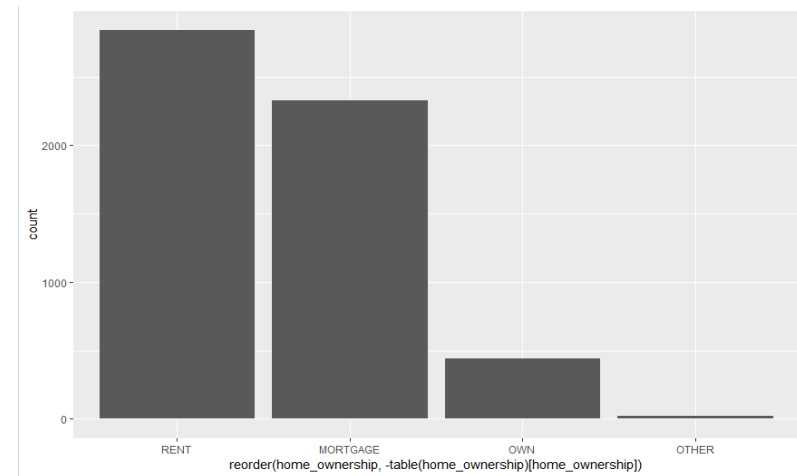
Observation : categories
"B", "C", "D" are max
defaulters



Results- EDA- Univariate Analysis

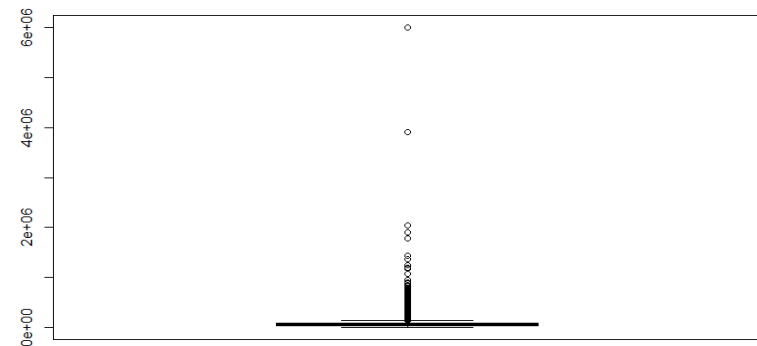
Ø Home ownership

Observation : 'RENT' & 'MORTGAGE' are the mostly defaulters



Ø Annual income

Observation : Outliers are present

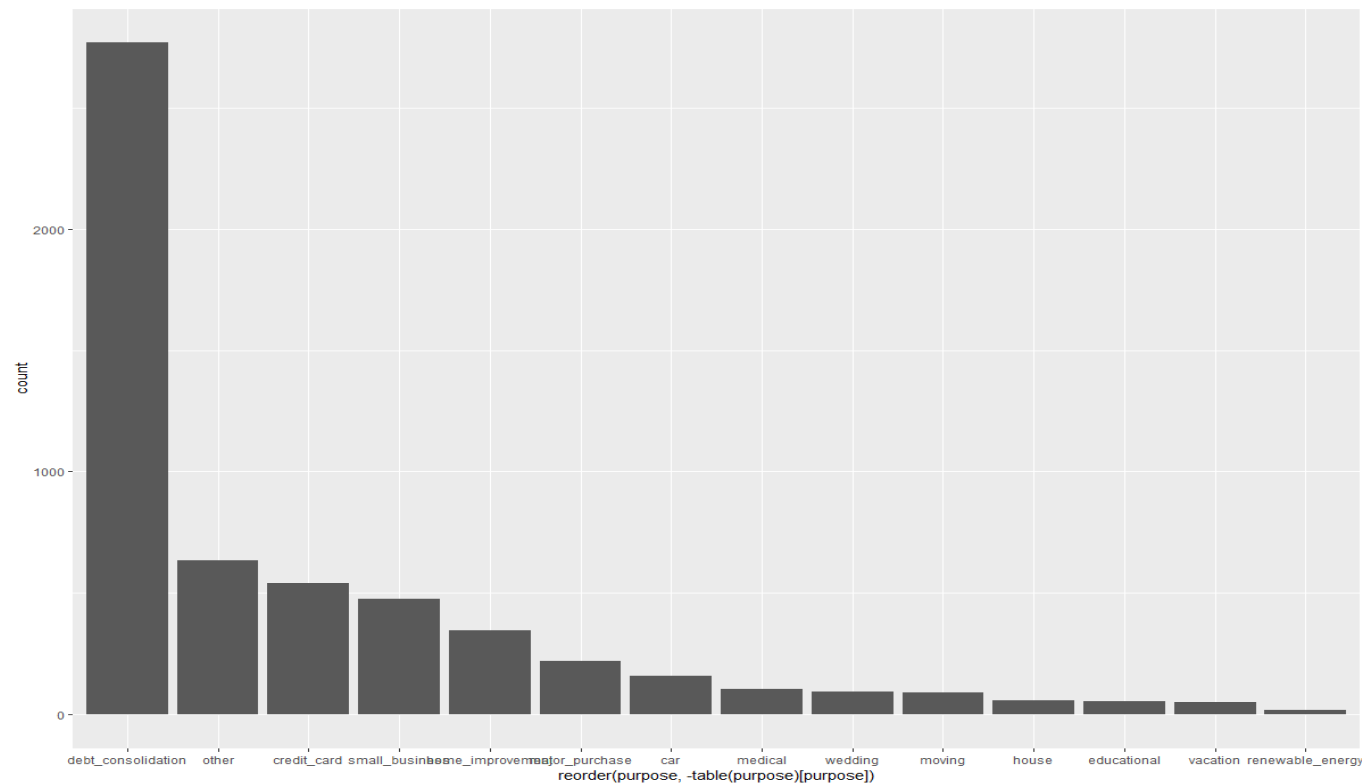


Results- EDA- Univariate Analysis

Ø Purpose of the loan

Observation : 'debt consolidation' is the largest factor in cases where people have defaulted

- debt_consolidation 2767
- other 633
- credit_card 542
- small_business 475



Results- EDA- Univariate Analysis

Ø The state major defaults are happening

Observation : CA

#CA 1125

#FL 504

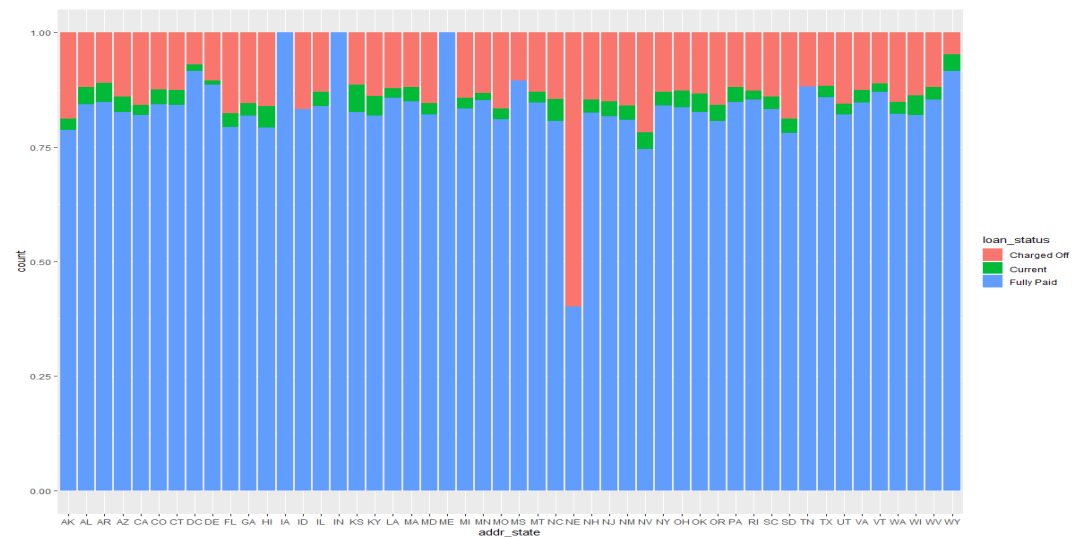
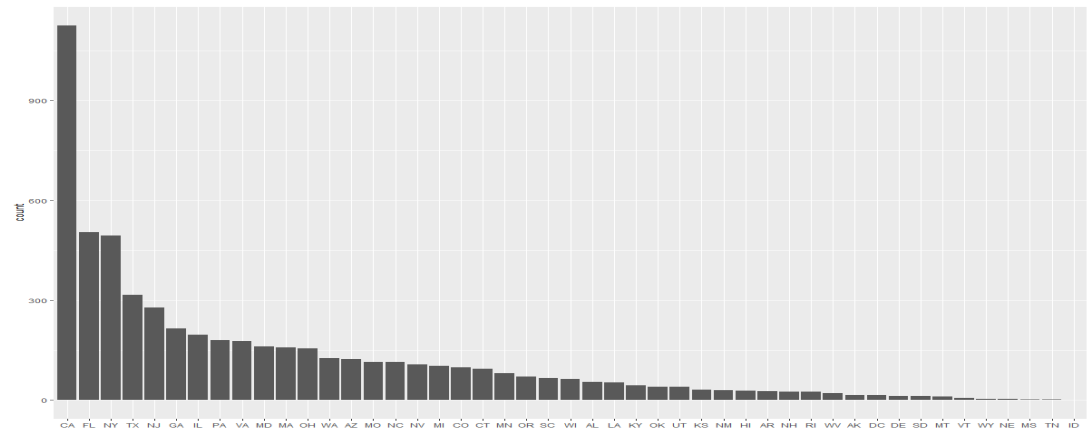
#NY 495

#TX 316

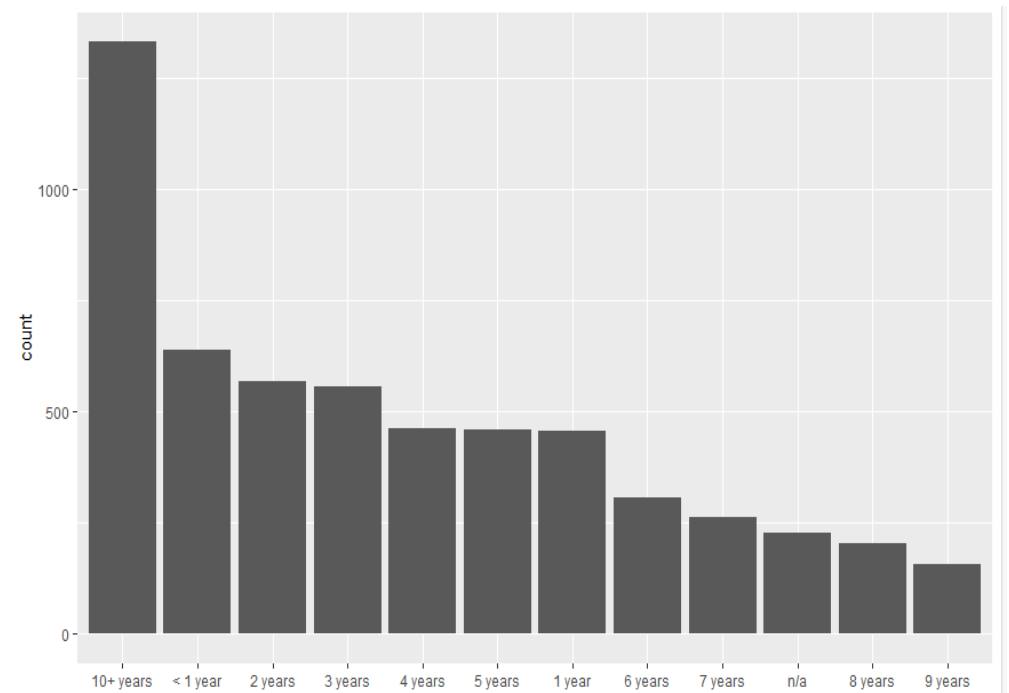
Ø The proportions of the default

Observation : NEVADA

seems to default more than 50% of times. After analyzing, we found that there are only 3 records
Not useful for analysis



Results- EDA- Univariate Analysis



Ø Employee length

Observation :

Most defaults in <1 year & 10+ years

Results- EDA- Bivariate Analysis

- Took variables for correlation
- annual_inc , loan_amnt , funded_amnt , dti

Observation: No correlation between "loan_amnt", "int_rate", "dti", "annual_inc" variables.

	#loan_amnt	int_rate	dti	annual_inc
#loan_amnt	1.00000000	0.30941527	0.06643935	0.27114855
#int_rate	0.30941527	1.00000000	0.11116168	0.05318516
#dti	0.06643935	0.11116168	1.00000000	-0.12273191
#annual_inc	0.27114855	0.05318516	-0.12273191	1.00000000

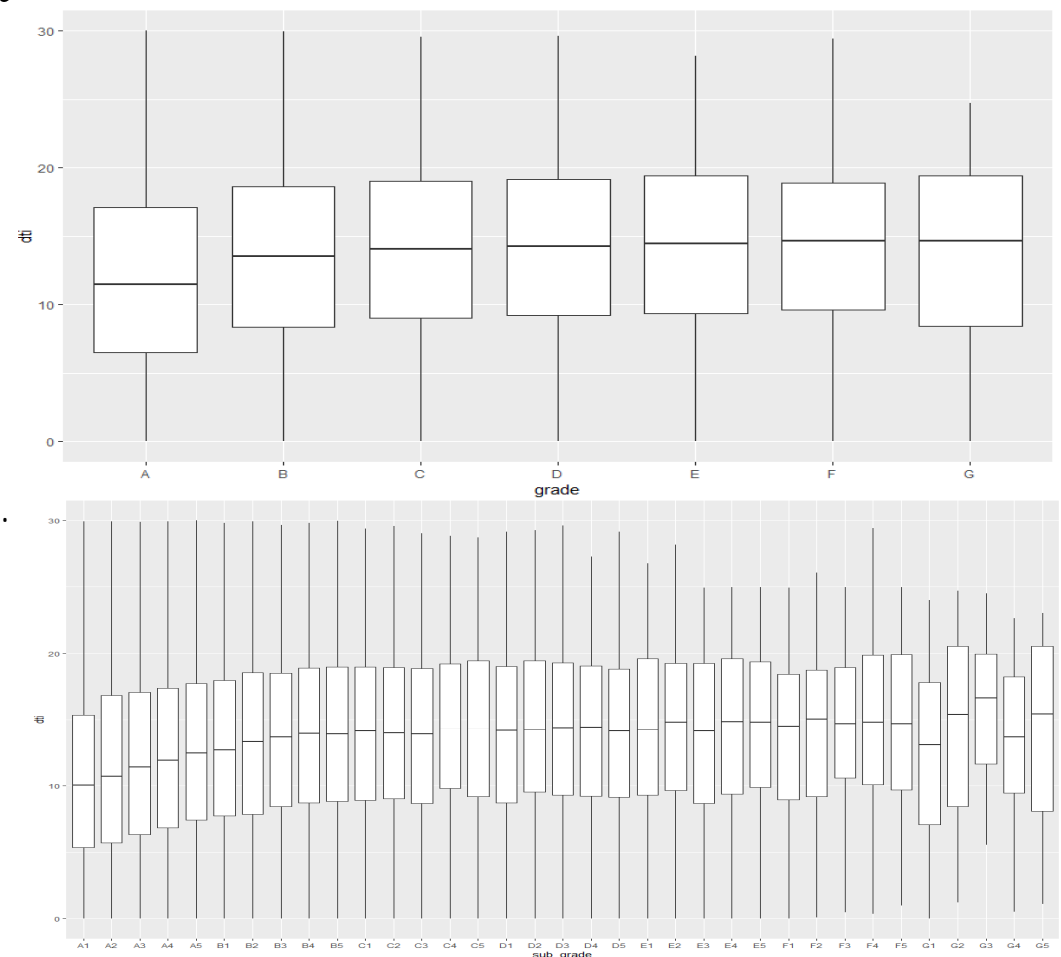
Results- EDA- Bivariate Analysis

- Took variables for correlation

Ø DTI

Ø Grade & Subgrade

Observation: People with higher dti has higher grades.
Hence, higher risk of loan defaulting.



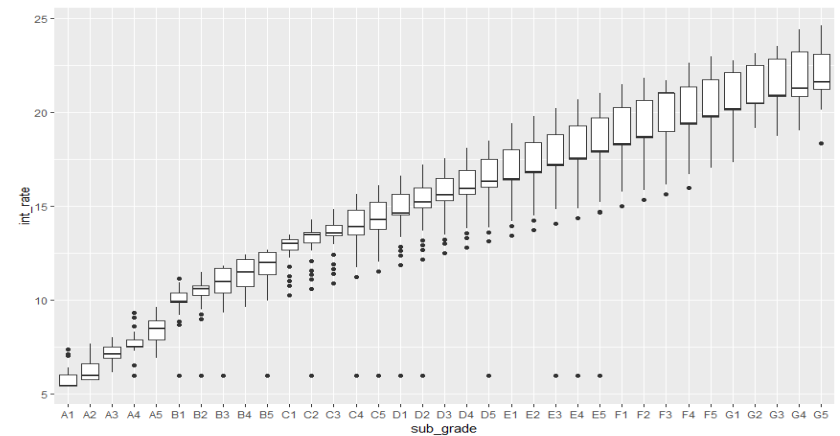
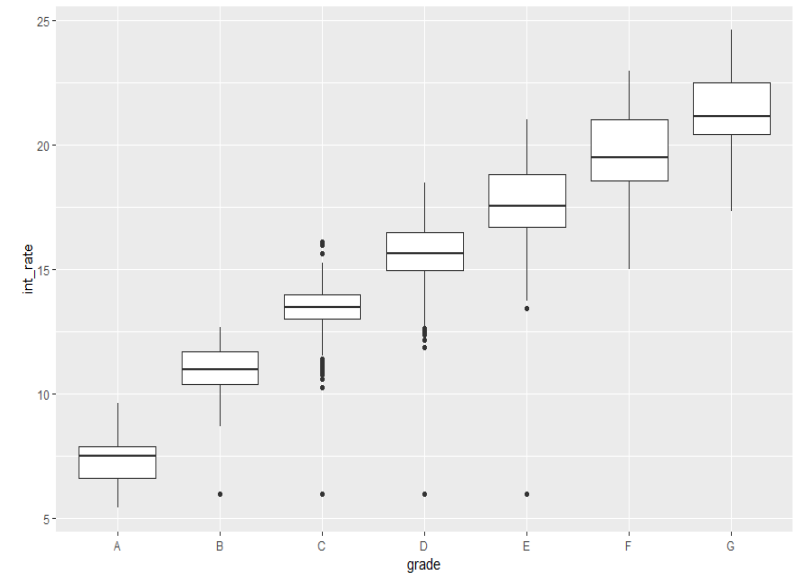
Results- EDA- Bivariate Analysis

- Took variables for correlation

Ø Interest rate

Ø Grade & Subgrade

Observation: Higher the grade or subgrade, higher the interest rate. Hence, higher risk of loan defaulting.



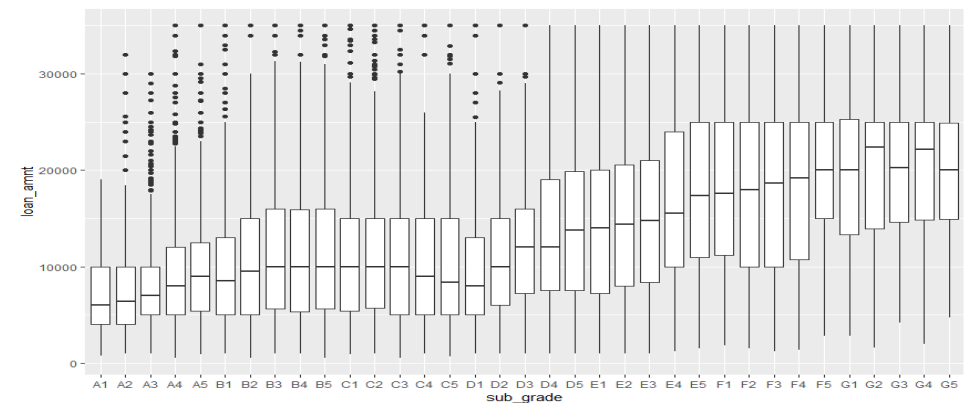
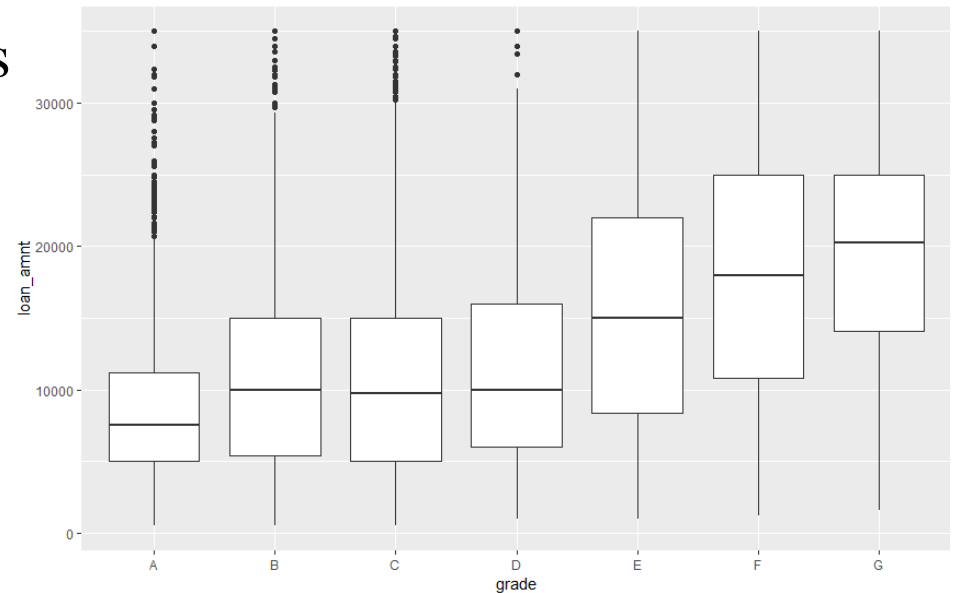
Results- EDA- Bivariate Analysis

- Took variables for correlation

Ø Loan Amount

Ø Grade & Subgrade

Observation: Higher grades has loans with higher amount. Hence, higher risk of loan defaulting.



Results- EDA- Hypothesis Testing- Excel

Ø Dti _defaulted

Ø Dti_nondefaulted

Observation:- People with higher dti has higher chances of loan default

t-Test: Two-Sample Assuming Unequal Variances		
	<i>dti-defaulted</i>	<i>dti-non defaulted</i>
Mean	14.00062378	13.20197976
Variance	43.36982501	44.71817432
Observations	5627	34090
Hypothesized Mean Difference	0	
df	7667	
t Stat	8.40947304	
P(T<=t) one-tail	2.4327E-17	
t Critical one-tail	1.645052395	
P(T<=t) two-tail	4.86539E-17	
t Critical two-tail	1.960273446	
<p>P value (one tail) is less than 5% significance value. So, there is significant difference between the mean. Hence, our observation is correct that people with higher DTI has higher chances of a loan default.</p>		

Ø Int_rate _defaulted

Ø Int_rate_nondefaulted

Observation:- People with higher Int_rate has higher chances of loan default

t-Test: Two-Sample Assuming Unequal Variances		
	<i>int_rate(0 - nondefault)</i>	<i>int_rate(1-defaulted)</i>
Mean	0.117166324	0.137974054
Variance	0.001329387	0.00132828
Observations	34090	5627
Hypothesized Mean Difference	0	
df	7604	
t Stat	-39.67509304	
P(T<=t) one-tail	0	
t Critical one-tail	1.645054042	
P(T<=t) two-tail	0	
t Critical two-tail	1.96027601	

Conclusions

- Ø The below mentioned points if avoided will be best to get away from the defaulters and is recommended.
- Most defaults are with loan value from 5000 – 15000
 - For loan_amount outliers are present
 - More defaults in loan with tenure '36 months'
 - interest rates between 10-15% Interest rate are majority defaulters
 - grades of categories "B", "C", "D" are max defaulters
 - 'RENT' & 'MORTGAGE' are the mostly defaulters
 - 'debt consolidation' is the largest factor in cases where people have defaulted
 - people with higher DTI default more
 - Most defaults in <1 year & 10+ years
 - People with higher dti has higher grades. Hence, higher risk of loan defaulting
 - Higher the grade or subgrade, higher the interest rate
 - Higher grades has loans with higher amount