Gramener Case Study

SUBMISSION

Group Name:

1. Abinash Bishoyi

2. Karthik Vecham

3. Balaji Nagaraja

Business Objective

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

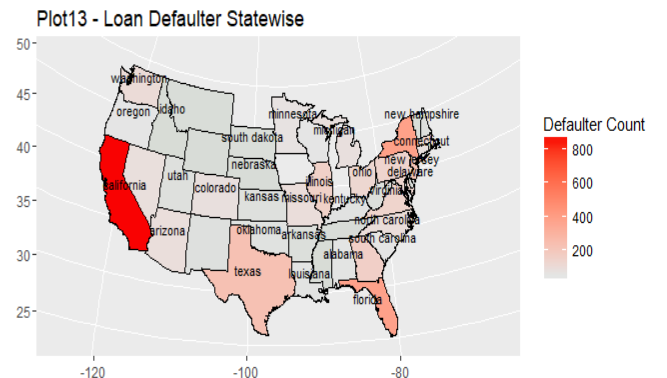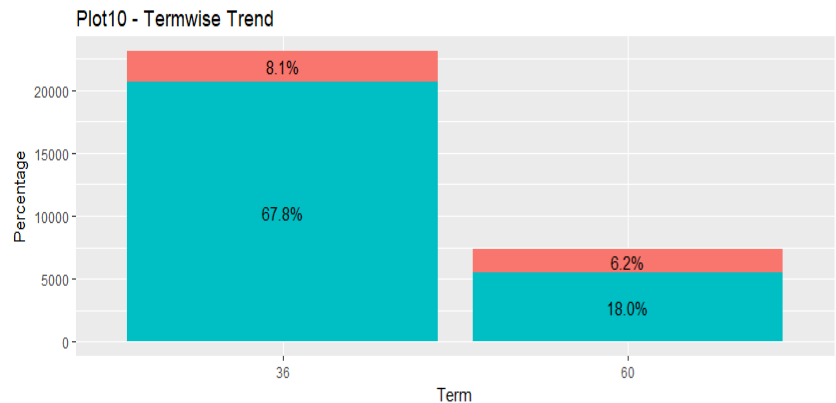Our aim is to understand how consumer attributes and loan attributes influencing the tendency of defaulting.
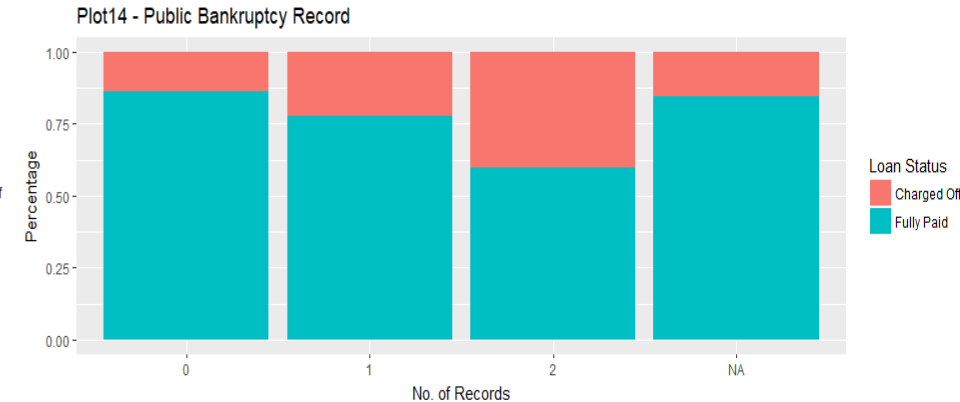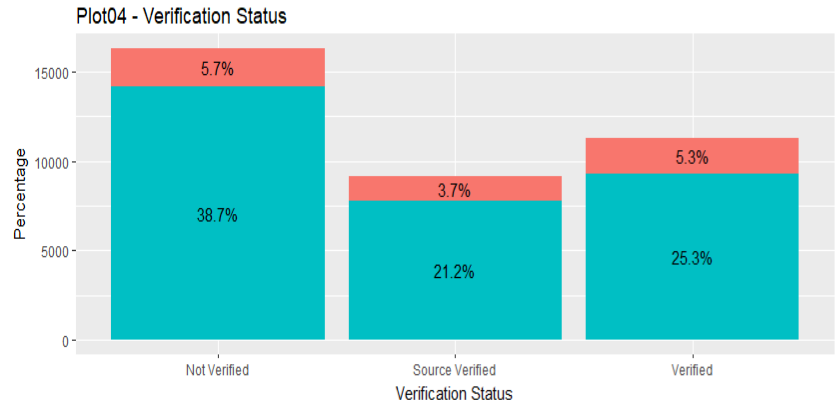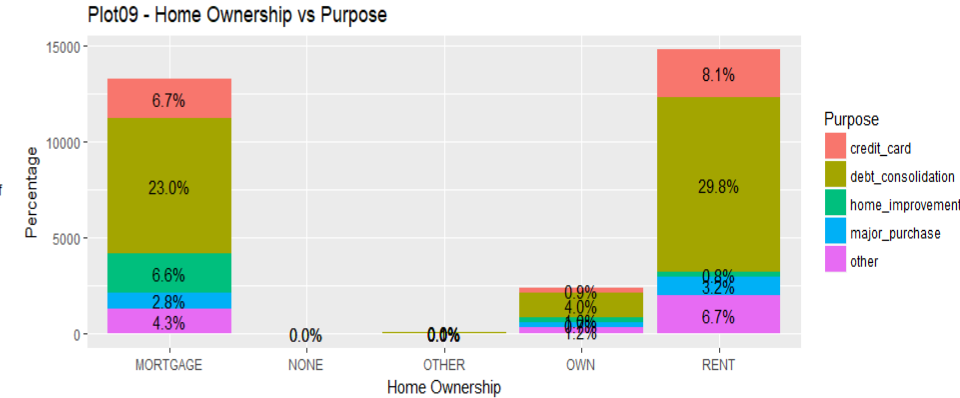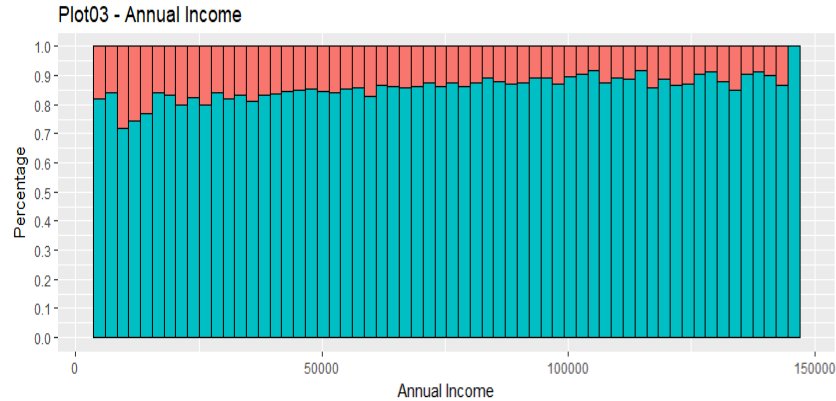
Constraints:

We consider only consumers whose loan application is approved.

Data Analysis Methodology

Univariate & Segmented Univariate Analysis
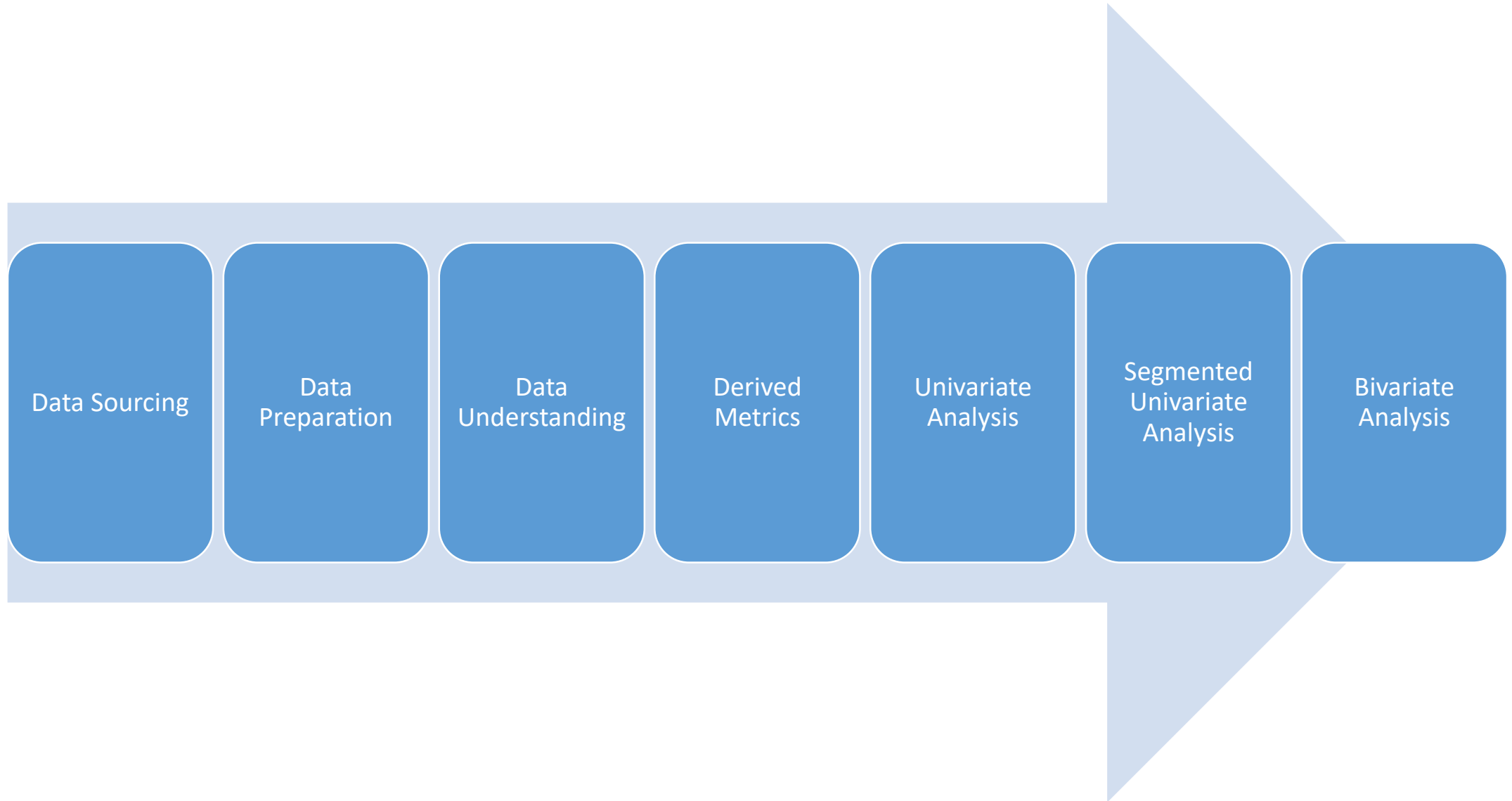
Inference :

- Annual Income does not show a consistent relationship with respect to loan status. ***Dropped from analysis***

- Home Ownership provides credible information about borrowers market value (Only 3 rows belong to "None" category). ***Should be considered for analysis***

- Verification status shows relationship with loan status, surprisingly verified loans seems to have more defaults. ***Should be considered for analysis***

- Public Bankruptcy Records shows strong relationship with loan status. ***Should be considered for analysis***

- Loan term shows strong relationship with loan status. ***Should be considered for analysis***

- Address State does not show a consistent relationship with respect to loan status. ***Dropped from analysis***

Plot03 - Annual Income

Plot09 - Home Ownership vs Purpose

Plot04 - Verification Status

Plot14 - Public Bankruptcy Record

Plot10 - Termwise Trend

Plot13 - Loan Defaulter Statewise

Univariate & Segmented Univariate Analysis

UpGrad

Inference :

- Loan Purpose plays an important role and shows strong relationship with loan status. Maximum number of loans are taken for debt_consolidation. *Should be considered for analysis.*

- Revolving Utilization shows strong relationship with loan status. *Should be considered for analysis.*

- Grade/Subgrade of the loan shows a very strong relationship with loan status. *Should be considered for analysis.*

- Loan Amount shows a very strong relationship with loan status. *Should be considered for analysis.*

Bivariate Analysis - Correlation
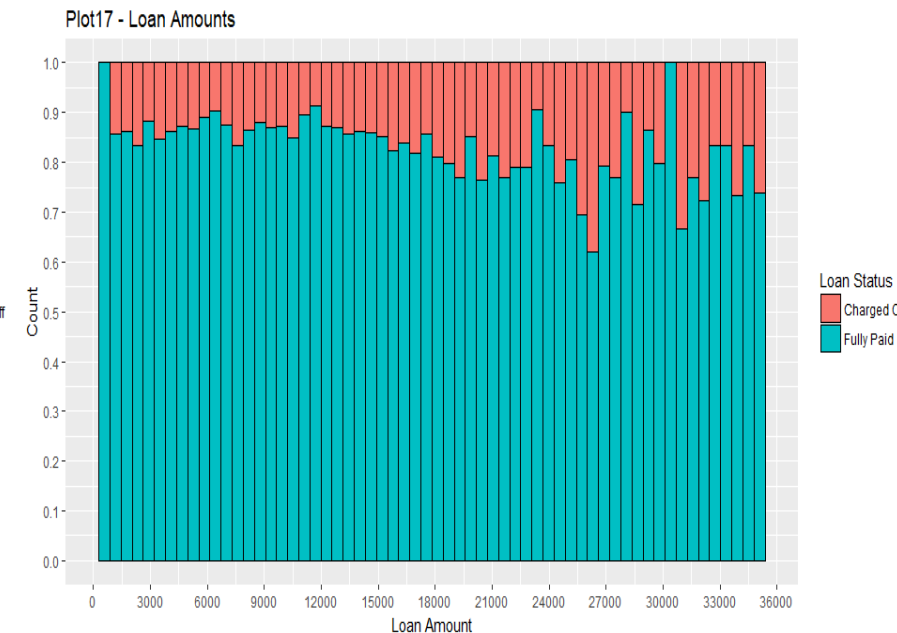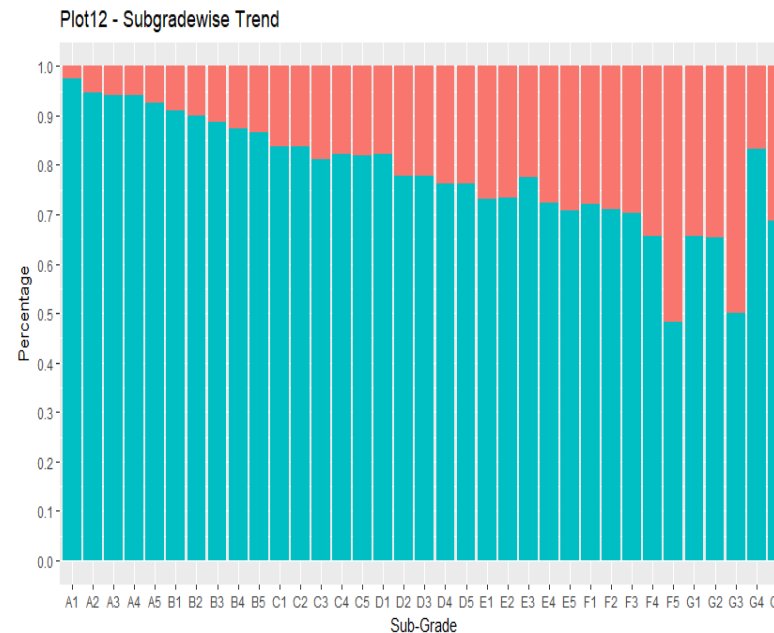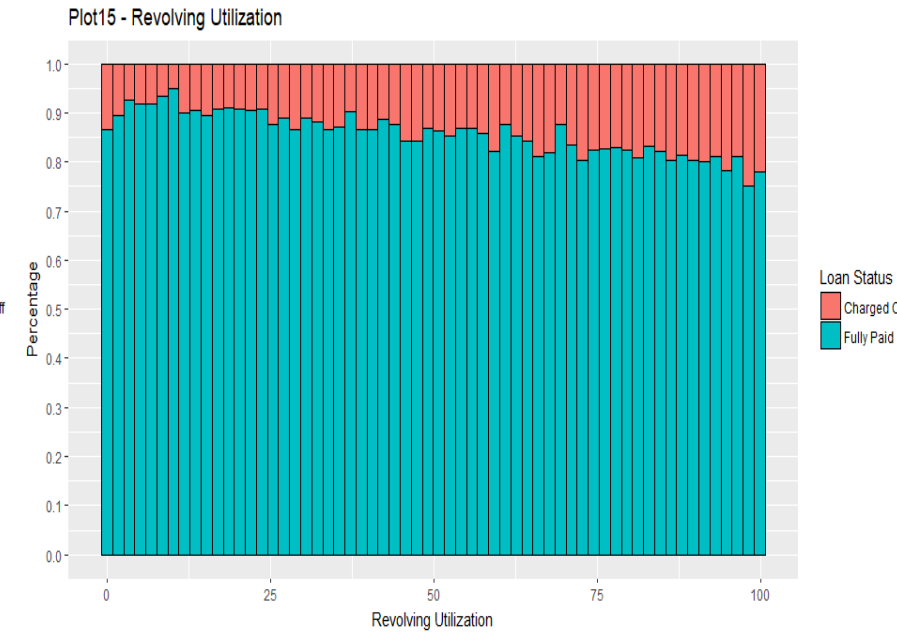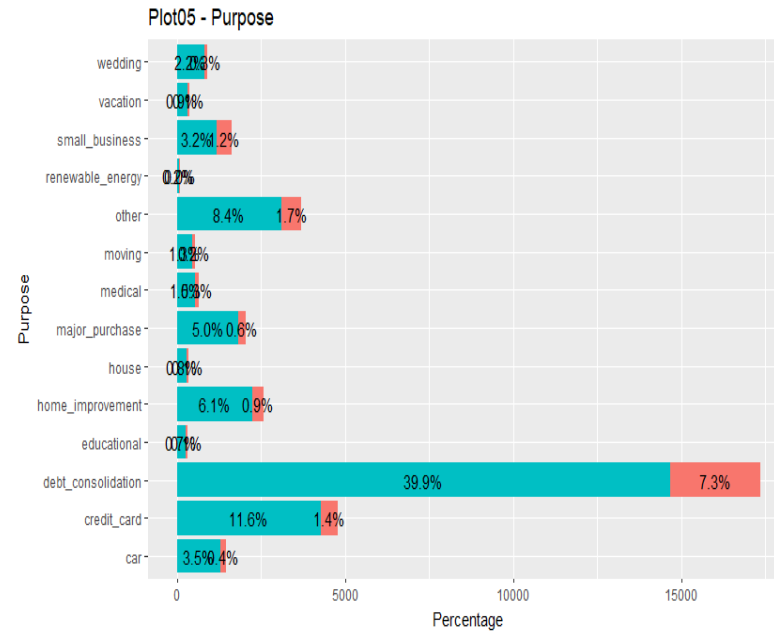
When a borrower comes with a loan request, following methodology would have been followed:

1. Based on the loan application submitted by the borrower, a specific grade is provided to the application
2. Based on the grade/sub-grade other variables (such as interest rate etc.) are decided.

In order to decide driving factor, it is important to check their relationship with grade/sub-grade as well.
If a variable's relation with grade/sub-grade is strong, then we may decide on considering only grade/sub-grade as the influencing parameter among the two variables.

Based on above thought process and co-relation matrix shown on the right, following conclusions can be drawn.

a) sub_grade - Interest Rate have strong co-relation
b) loan_amnt - Instalment have strong co-relation
c) revol_util - int_rate & sub_grade seem to have a moderate correlation.

Bivariate Analysis - Correlation
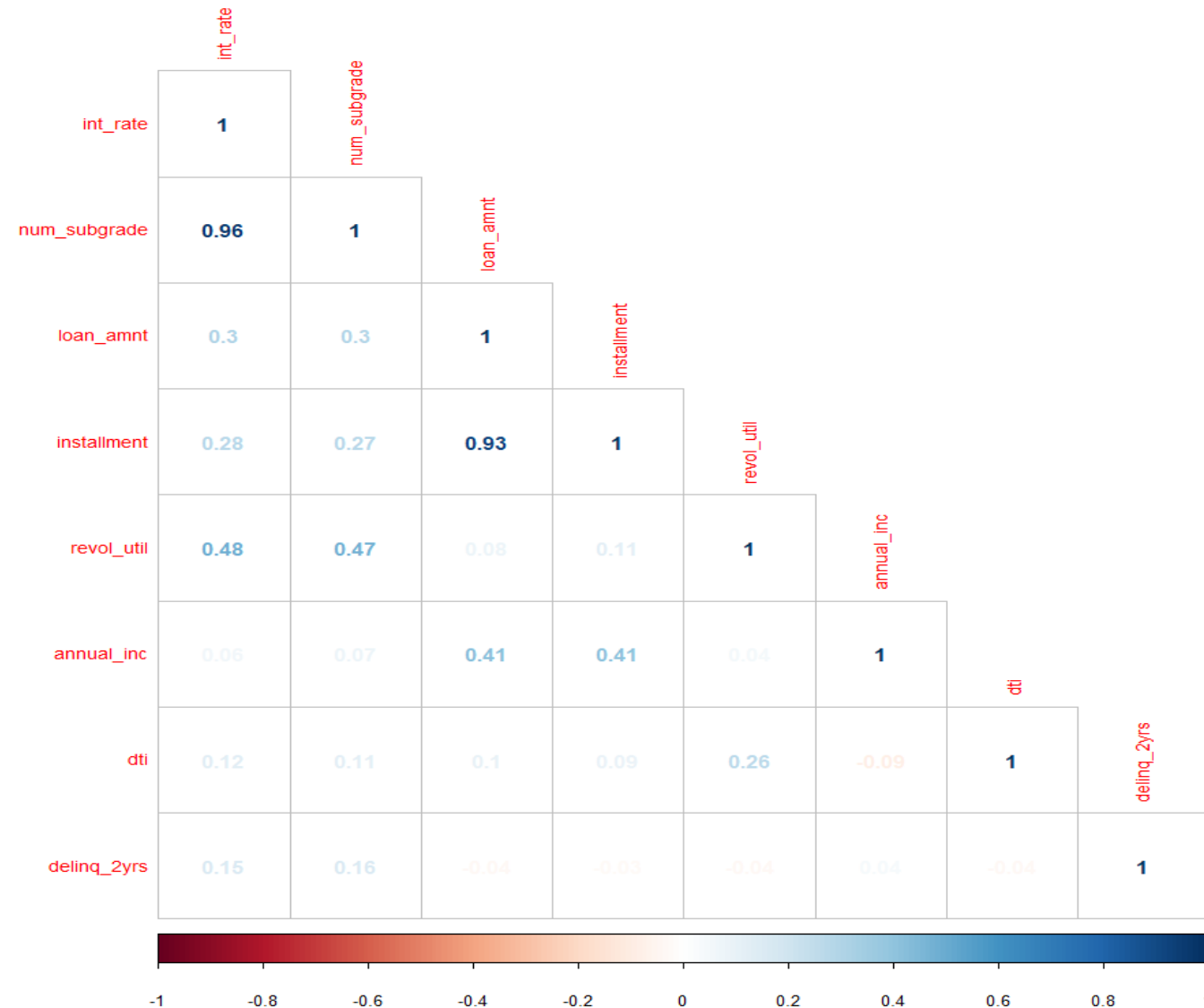
Grade/Sub- Grade vs Revolving Utilization

This co-relation seems to have exceptions, and hence binning should be done and revolving credit should be considered as an independent variable.

Conclusion:

It would be better to consider revol_util as a separate variable and bin it for easier analysis.

Plot18 - Grade Vs Revolving Utilization

Plot19 - Grade Vs Loan Amount
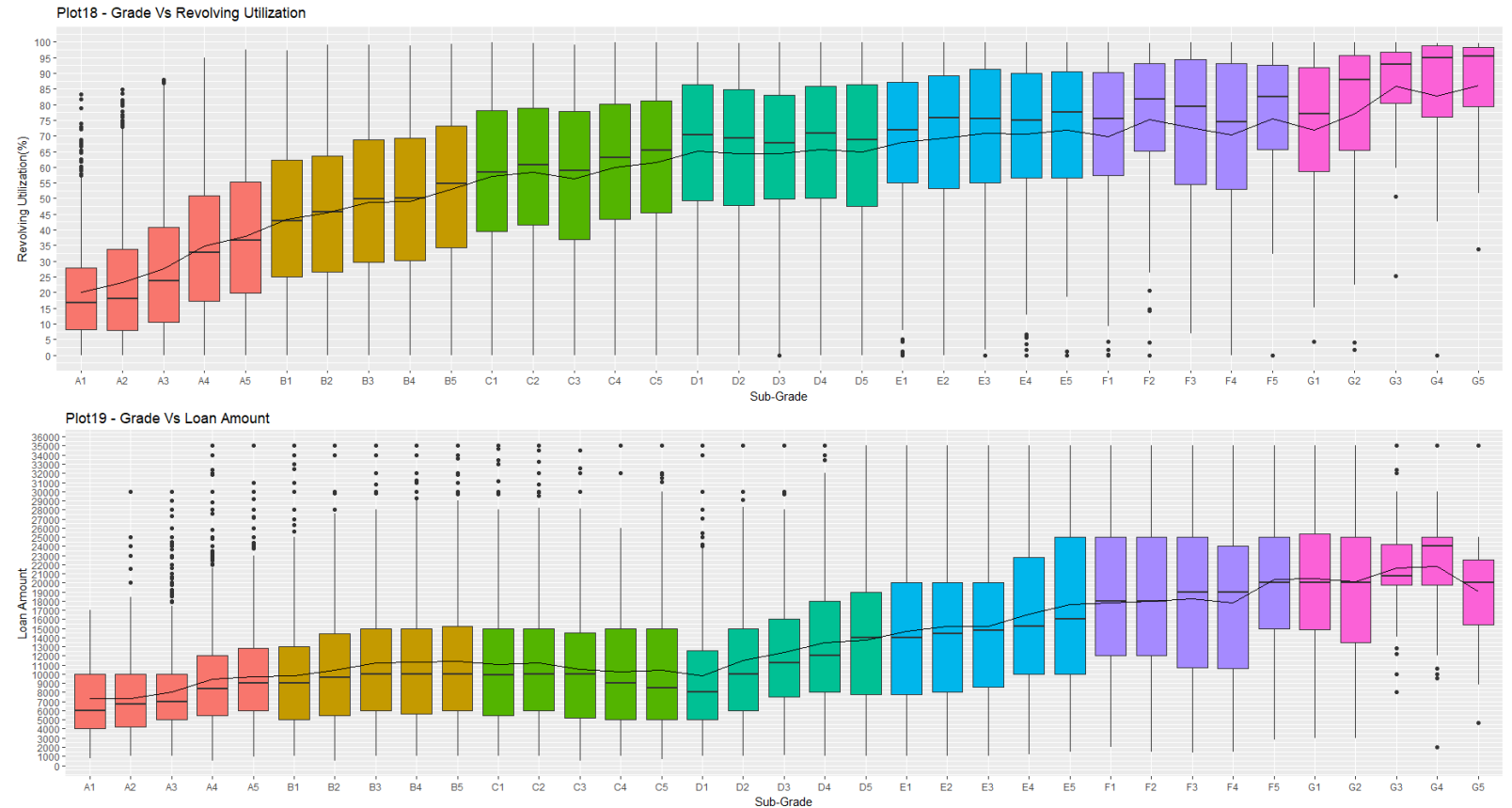
Grade Vs Loan Amount

A certain level of co-relation can be seen between Loan Amount and Grade of the loan, however this co-relation does not seem to be consistent.
Conclusion:
It would better to consider Loan Amount as a separate variable, and bin the loan amount for easier analysis.

Bivariate Analysis - Correlation

Grade/Sub-Grade vs DTI

DTI is the ratio calculated using the borrower's total monthly debt payments on the total debt obligations, divided by the borrower's self-reported monthly income.
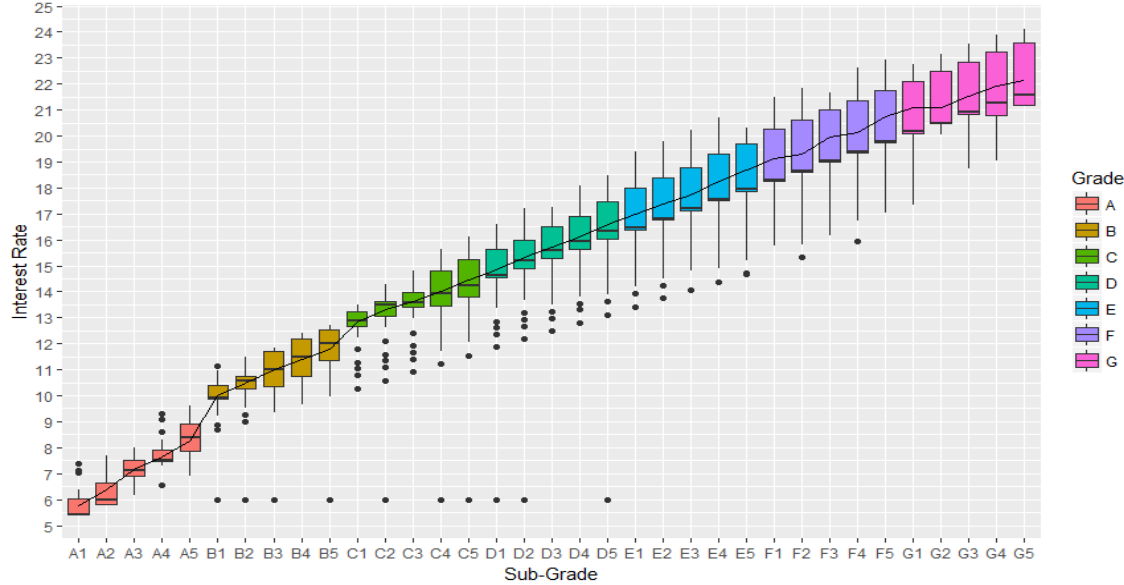
From Plot is clear that as the Sub-Grade increases (A1 – G5) the DTI value is also increasing.
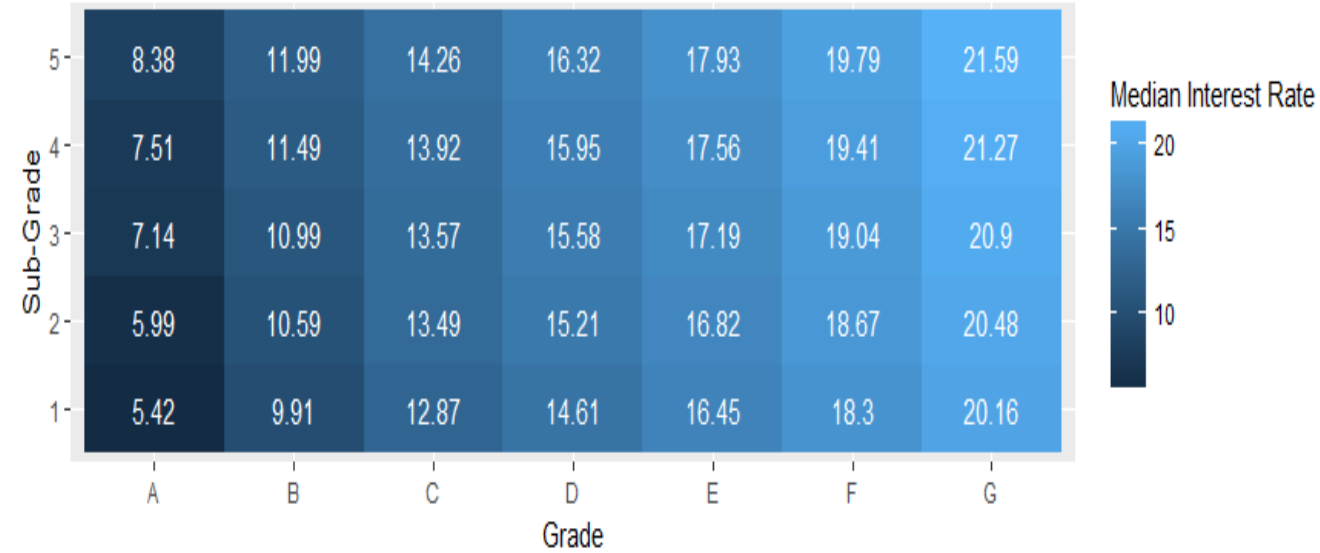
Conclusion:
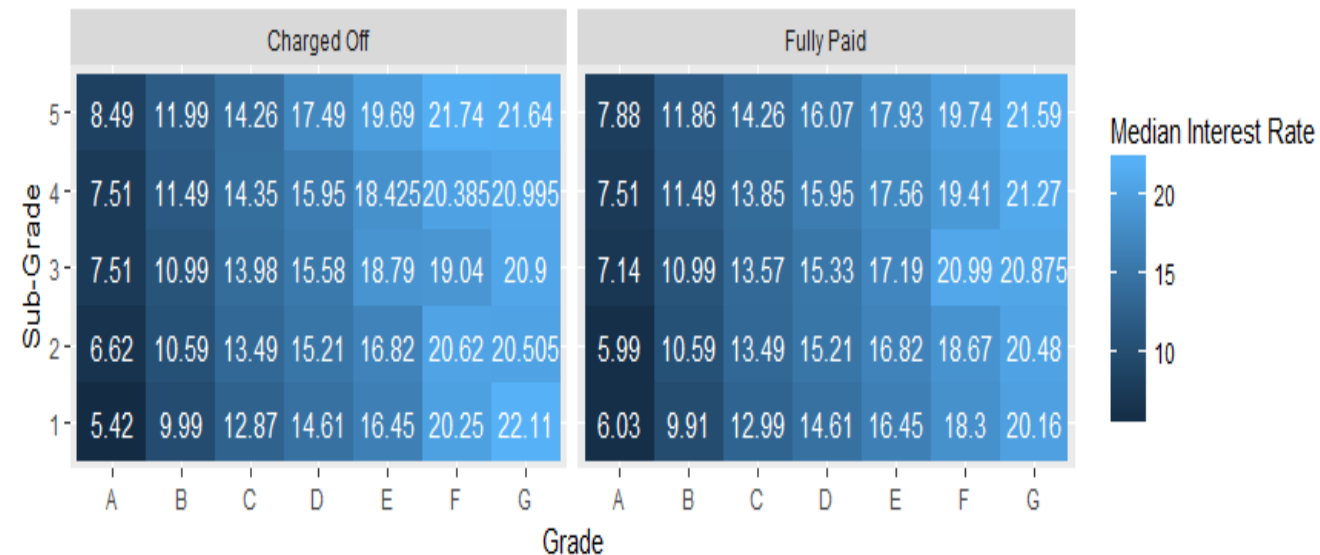DTI seems to be related to Sub-Grade hence we can take just Sub-grade as a reference variable.

Plot20 - DTI Vs Grade

Plot21 - DTI Vs Grade Vs Loan Status

Bivariate Analysis - Correlation
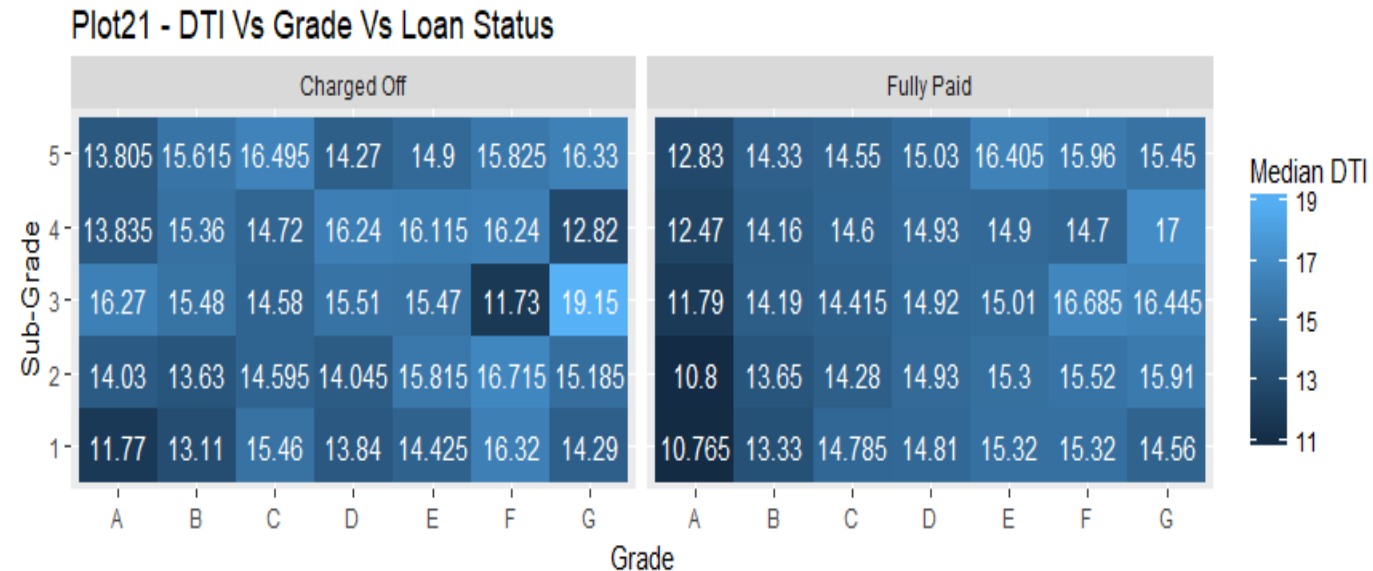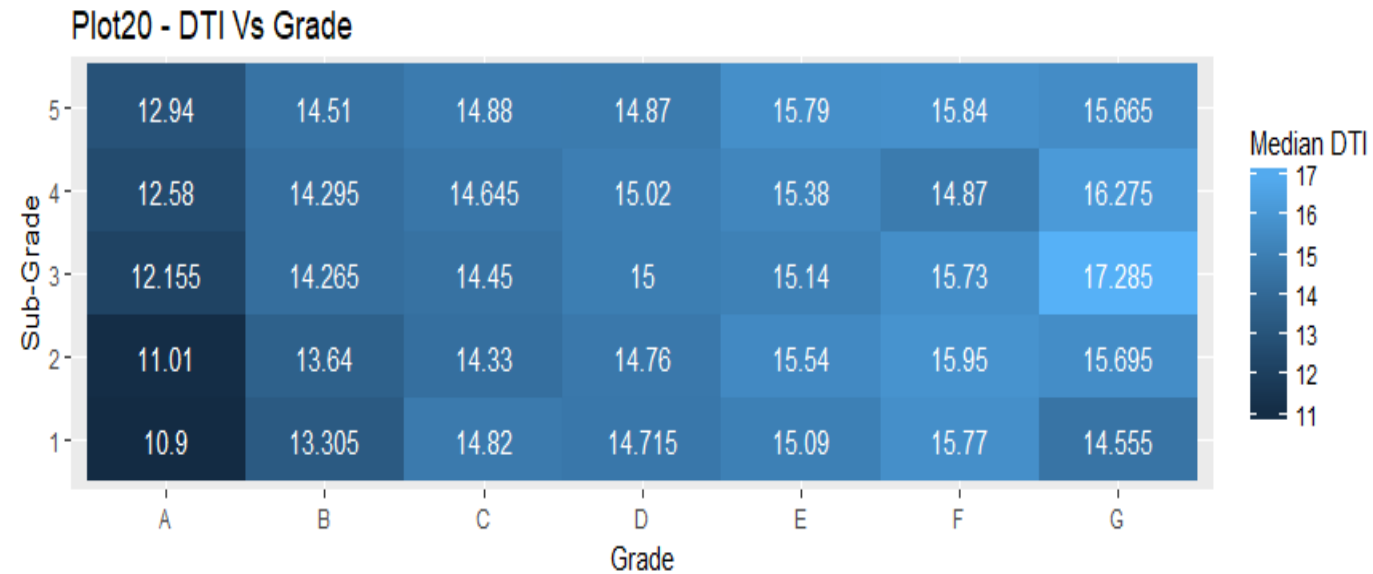
Plot22 - Grade Vs Interest Rate

Plot23 - Grade Vs Sub-Grade Vs Interest Rate

Grade/Sub-Grade vs Interest Rate

Loan interest rate shows a clear co-relation with subgrade feature.

Conclusion:

Since we are considering grade as an influencing variable, therefore, interest rate would be considered invariably. Hence we can leave away interest rate from being considered as an influencing variable.

Plot24 - Grade Vs Sub-Grade Vs Interest Rate Vs Loan Status

Conclusion

Based on EDA analysis, following variables can be said to be influencing the loan status to large extent.

Based on these variables we can predict the probability metrics as to whether currently running loans may end up as "Charged Off" or "Fully Paid".

Below list has been created on the basis of level of co-relation.
1. verification_status
2. purpose
3. home_ownership
4. term
5. pub_rec_bankruptcies
6. revol_util
7. sub_grade
8. loan_amnt