# EDA CASE STUDY

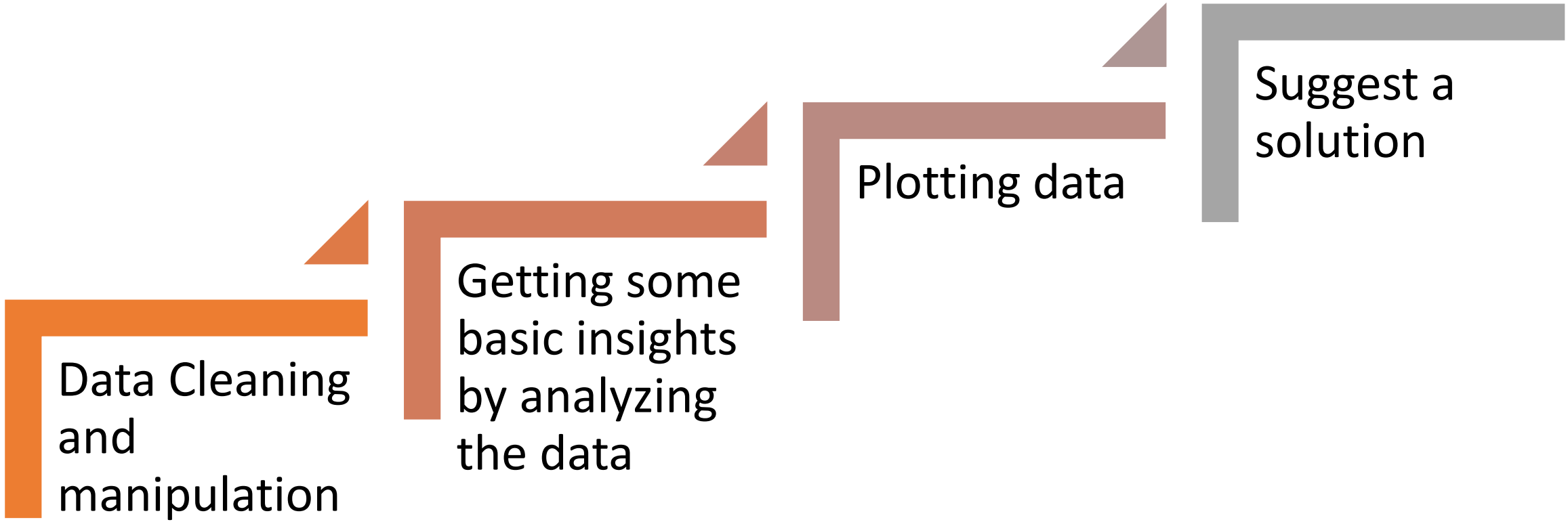
# SUBMISSION

Presented By:
Srikant Balaji
Ravi Prakash
Preksha Tripathi
Snehalakshmi Balasubramanian

# Abstract

- **Problem** : The objective of this case study is to determine the risks of a loan getting default based on the borrowers attributes
- **Data Available** : The loan data of customers from 2007 to 2011 is available for analysis. It contains data of loans that are either paid fully, Charged off (Defaulted) or In-progress.
- **Goal** : Determining the driving factors behind a loan default that helps perform portfolio analysis and risk assessment.
- **Technology** : R language for Analysing and plotting

# Problem solving methodology



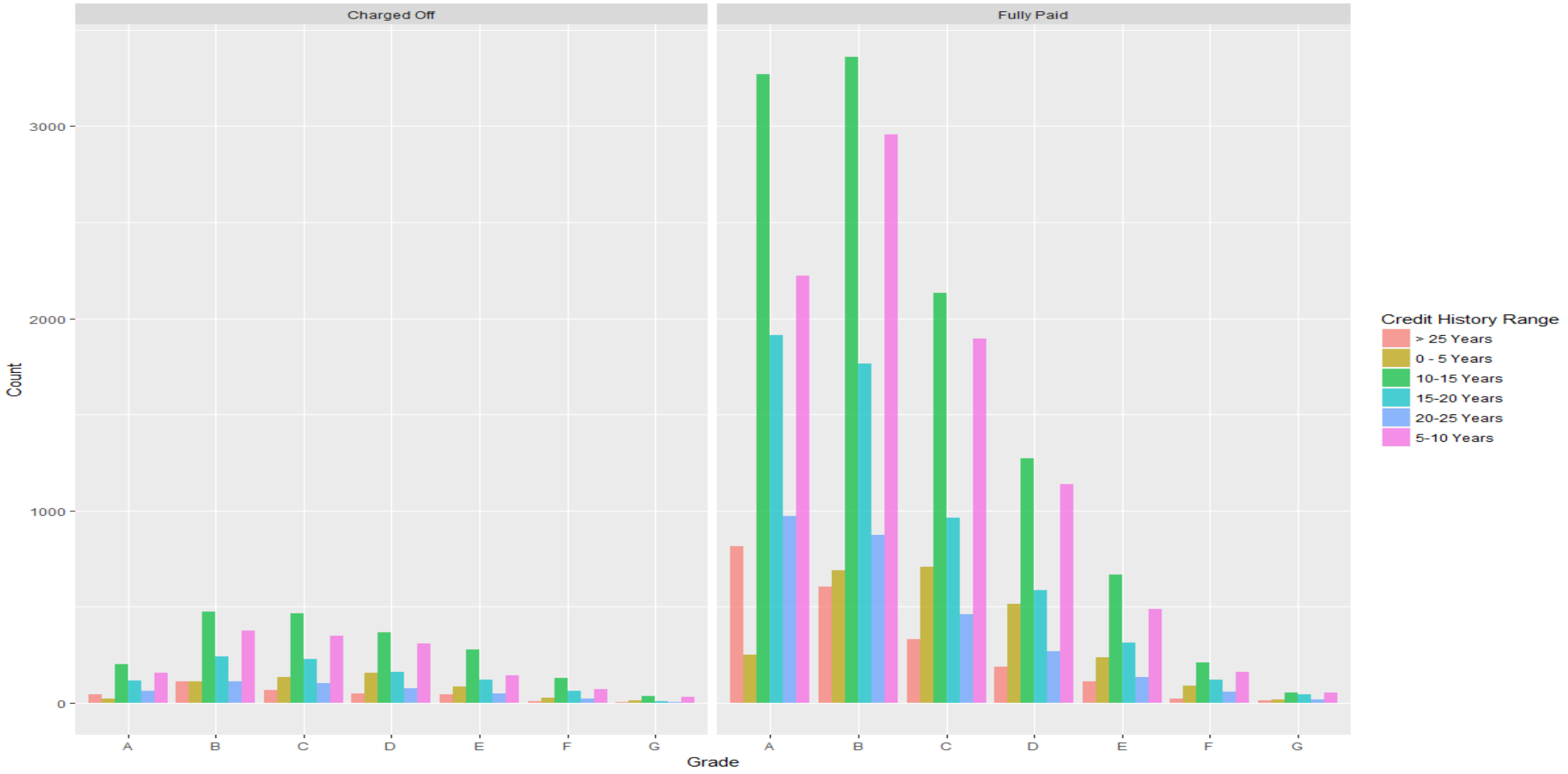
# Data Insights , Manipulation and Cleaning

- **Data Insights :**
  - ✓ The available data lists the details of loans from 2007 till 2011.
  - ✓ The loans are currently in-progress, paid off or Charged off .
  - ✓ There are a total of 111 fields in this file, with details of Borrower characteristics (such as Employment length, home ownership, Annual Income, Credit history, Delinquency etc.), Loan details (such as loan amount, purpose, interest rate) and LC assessment (Grade,SubGrade)
- **Data Issues and Cleaning :**
  - ✓ There are no duplicate rows.
  - ✓ Removing unwanted columns , several columns have null values and is of no use for the analysis.
  - ✓ Converting date format to an uniform format and splitting the month and year of the Issue data.
  - ✓ Handle NA or blank values

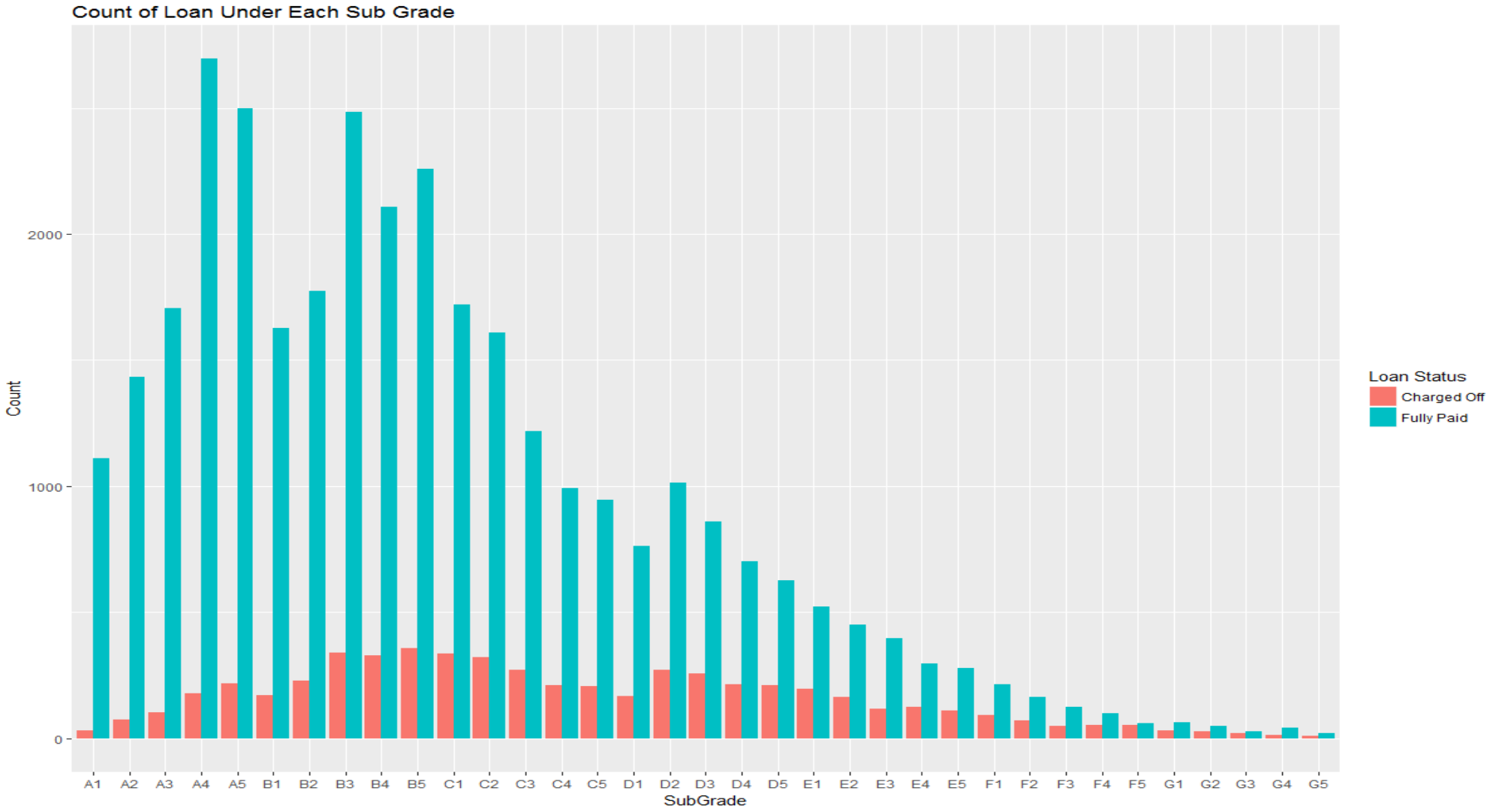
# Analysis of Key fields

| Loan Status | Median | | | | | | Average |
|---|---|---|---|---|---|---|---|
| | Funded Amount | Installment | Annual Income | Ratio of obligation to income | # of Credit accounts | Age of Credit | Derogatory records |
| Charged Off | 10000 | 294 | 53000 | 14.3 | 8 | 12 | 0.085 |
| Fully Paid | 9200 | 276 | 60000 | 13.2 | 9 | 13 | 0.050 |

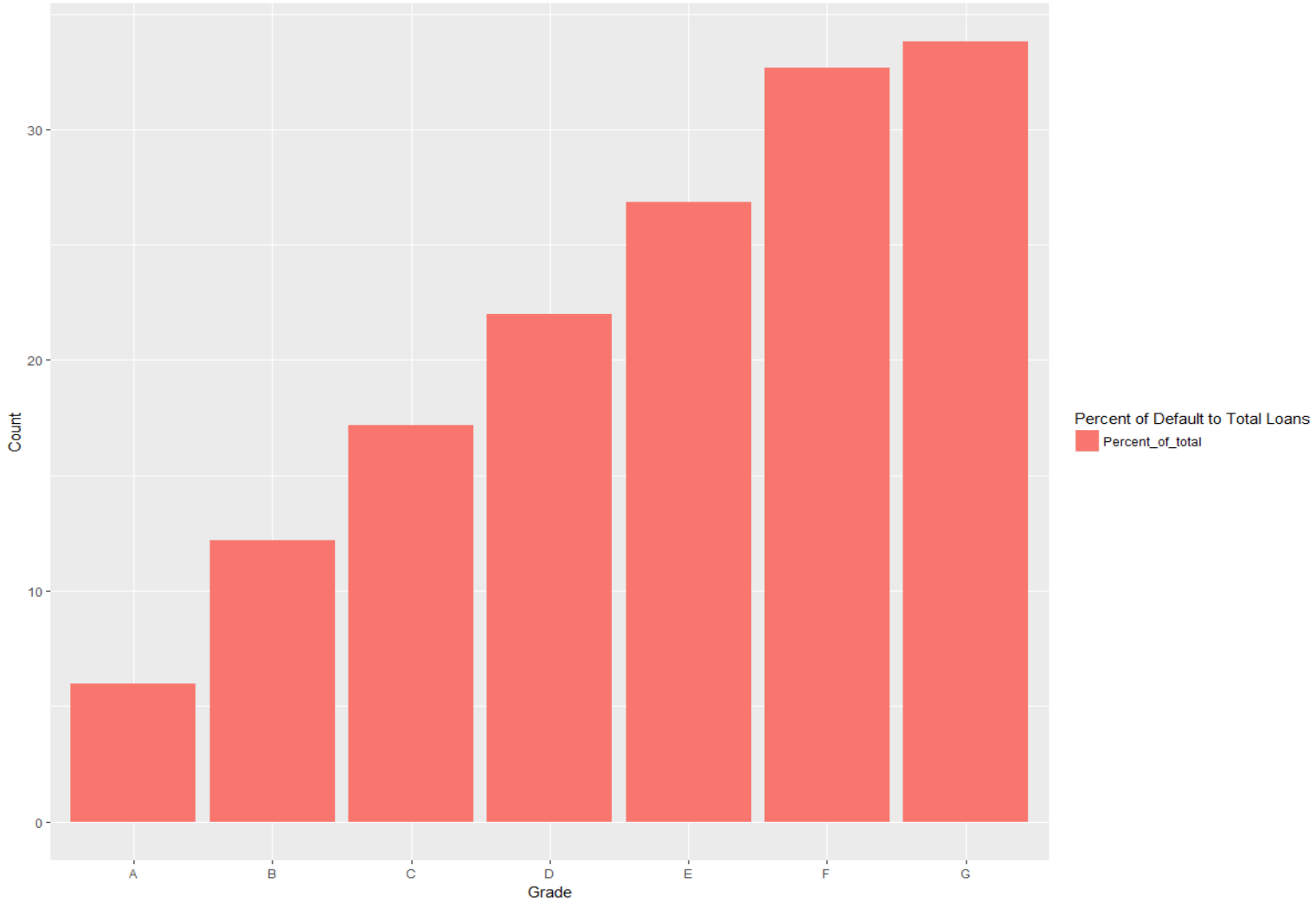
# Plot showing the number of loans for each grade, status and credit age



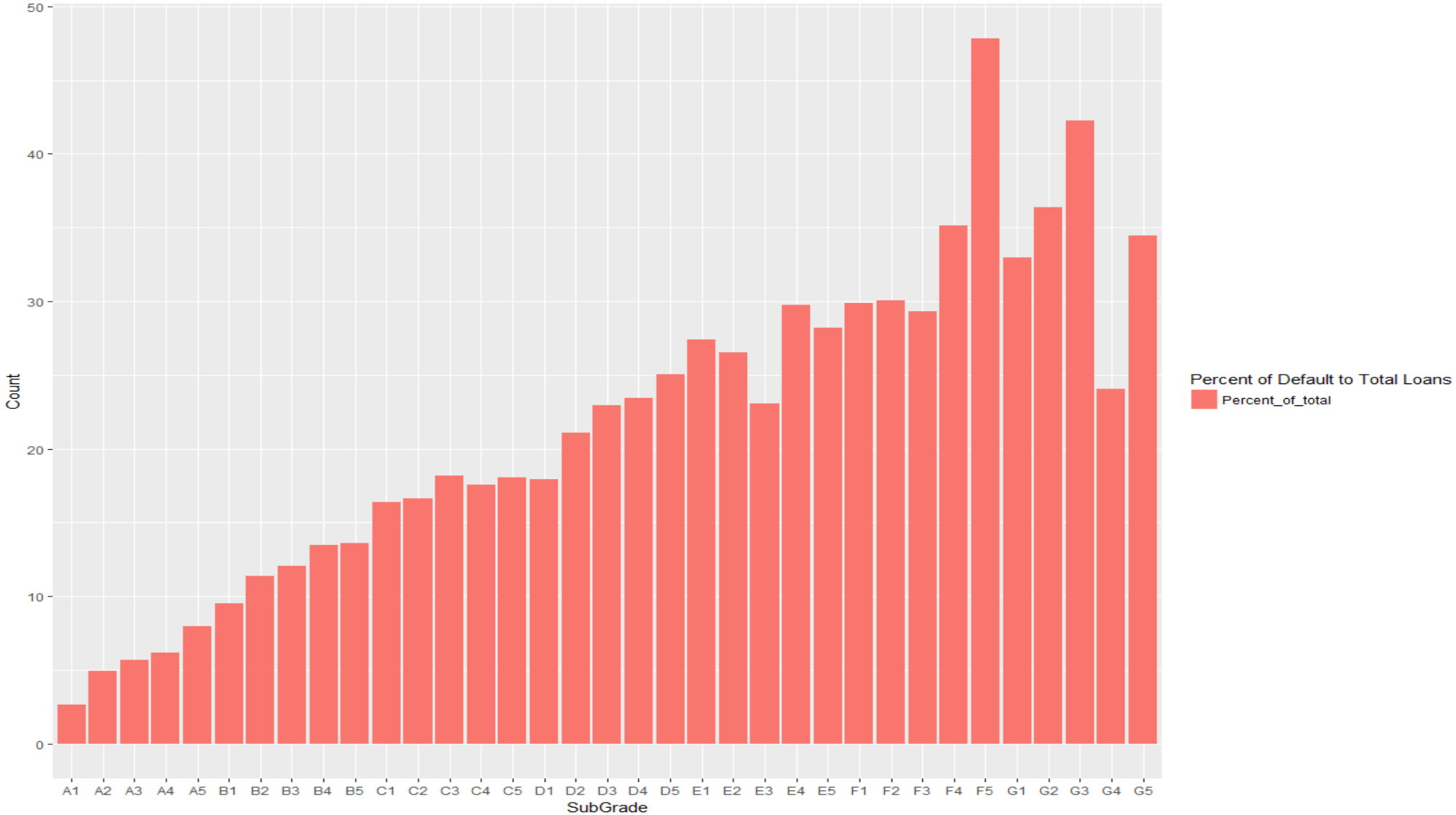
# Plot showing the number of loans for each subgrade and status



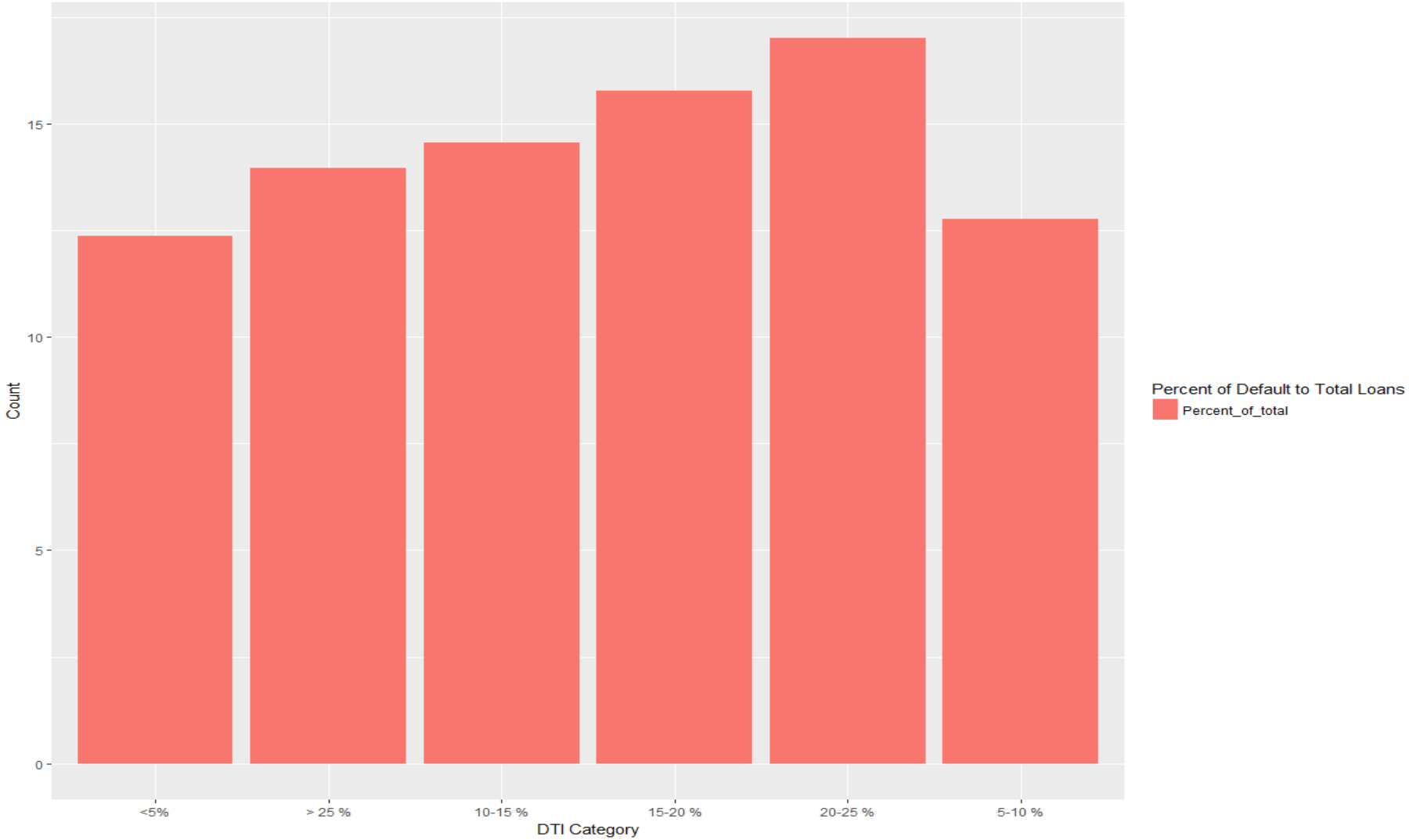
# Plot showing the ratio of defaults to total number of loans for each grade



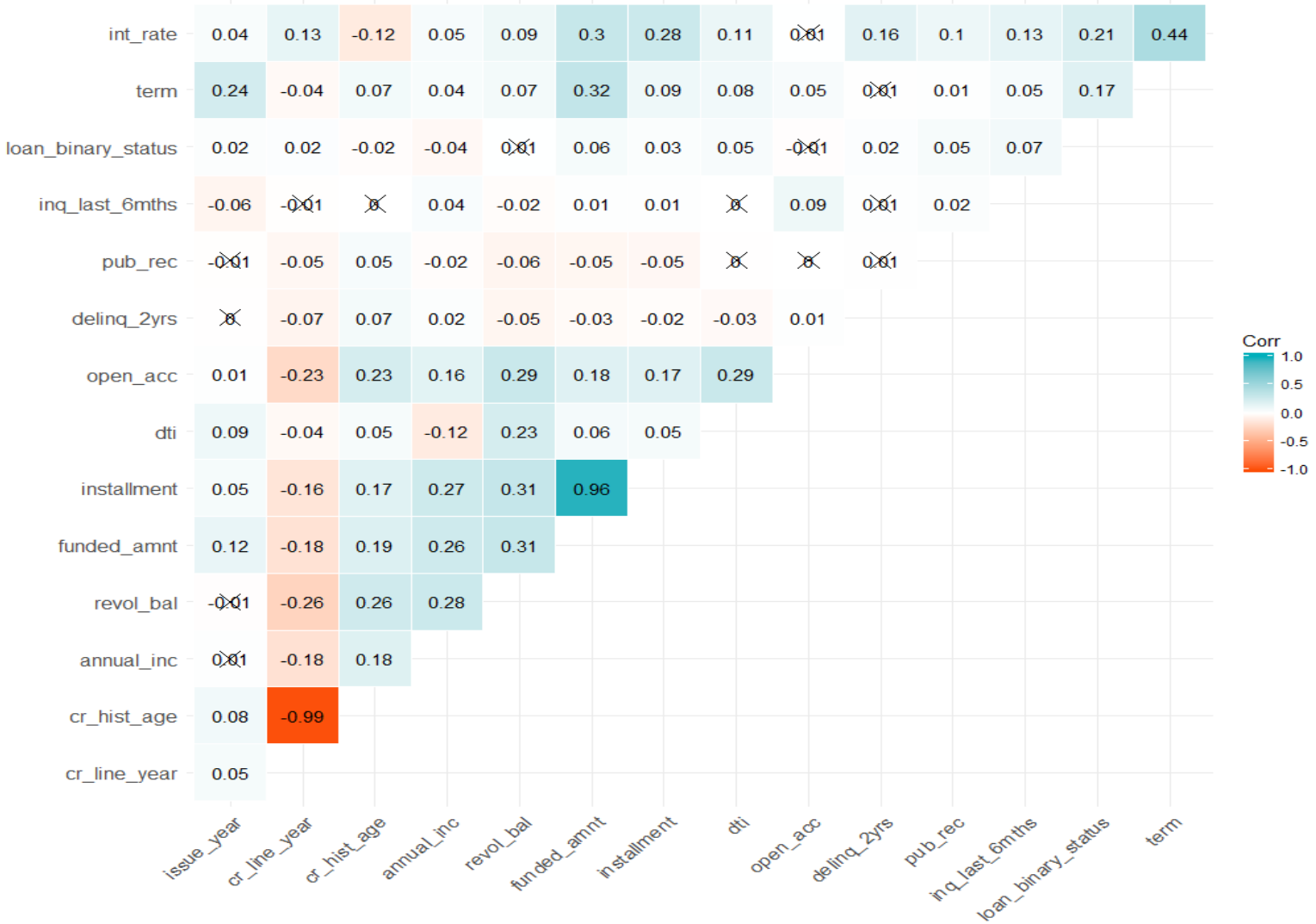
# Plot showing the ratio of defaults to total number of loans for each subgrade



# Plot showing the ratio of defaults to total number of loans for each DTI bucket



# Correlation plot showing the correlation of each variable to loan status



# Based on the analysis of the correlation plot, the following can be inferred :

- Interest Rate (a higher interest rate signals a higher chance of default )
- Annual Income (A higher annual income signals a lower default chance)
- Credit History age ( greater the age of credit history lower is the chance of default)
- DTI (higher the DTI higher chances of default)
- Term (Higher term has lower chances of default)

# Conclusions

| S.No | Field | Business Desc | Observations for default |
|---|---|---|---|
| 1 | grade | Assigned Loan Grade | Lower the grade (E,F etc.), higher the chance of default |
| 2 | subgrade | Assigned Loan Sub Grade | Within a grade, higher subgrade(4,5 etc.) signals higher chace of default |
| 3 | term | No. of Loan Payments | Higher term has lesser chances of default |
| 4 | int_rate | Interest Rate | Increasing interest rate, increases the chance of default |
| 5 | dti | Debt to Income Ratio | Increasing dti, increases the chance of default |
| 6 | pub_rec | No. Of Derogatory Remarks | Increasing derogatory remarks, increase the chance of default |
| 7 | open_acc | No. Of Credit Lines open | Lower the credit lines, higher the chance of default |
| 8 | funded_amnt | Loan Amount Approved | Higher the loan amount, higher the chance of default |
| 9 | Installment | Monthly Installment | Increasing installment, increases the chance of default |
| 10 | annual_inc | Annual Income | Lower the income, higher the chance of default |
| 11 | cr_hist_age | Credit History Age | Higher the age of credit history lower are the chances of default |

Thank You