

# Lending Club Case Study Assignment

## SUBMISSION

**Team:**

**Prasad Appana**

**Vibhor Malik**

## Problem Statement

- You work for a **Consumer Finance Company** which specializes in lending various types of loans to urban customers. Based on the past Loan Applications/Data, company wants to identify the 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.

## Constraints

- Past Loan Application Data is available only from **2007** to **2011**.

## Objective

- The objective is to identify **Risky Loan Applicants / Default Lenders** based on past data using **EDA**.
- Finance company wants to understand the **driving factors** (or driver variables) behind loan default, i.e. the **variables** which are strong indicators of default, thereby company can utilize this knowledge for its portfolio and risk assessment.



# Problem Solving Methodology



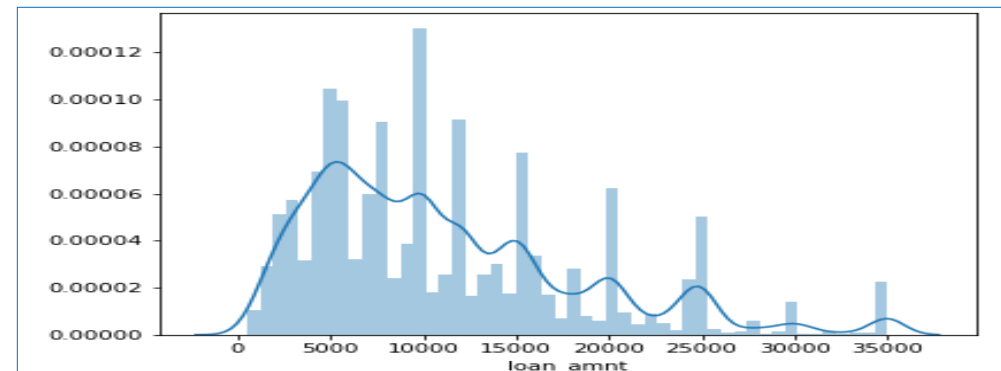
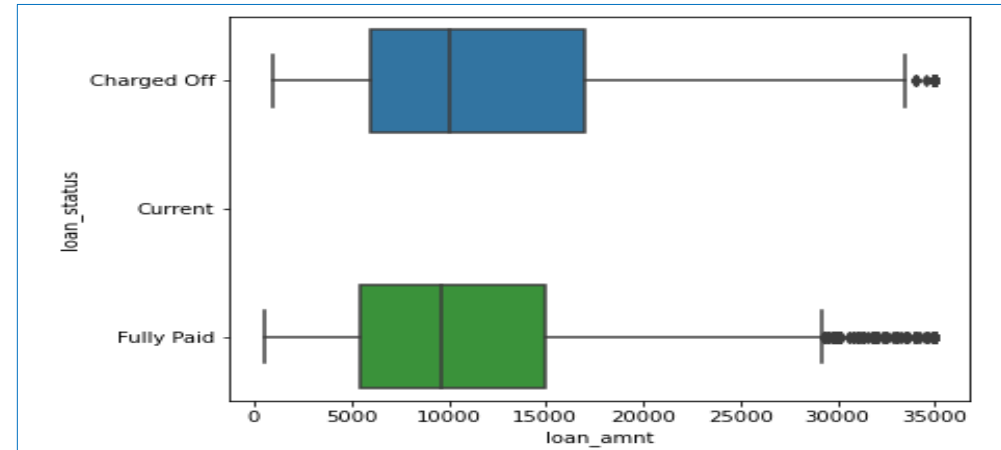
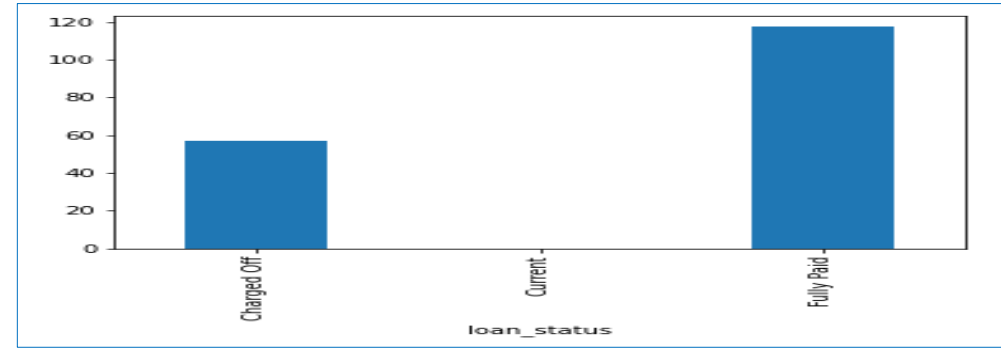
- Data understanding/Exploration
- Data cleaning
  - Cleaning Missing Values
  - Removing Redundant columns etc.
- Data Analysis
  - **Univariate analysis**
    - Check the Loan Status from Past Data, Keep only Fully Paid & Charged off.
    - From the above Data Cleaning Step, Identified about 30 Columns/Variables for this Analysis.
  - **Bivariate analysis**
    - Identify one or more variables and find out the relationship among these variables and analyze.
- Conclusions/ Driving Factors

## 1. Data Exploration

- ❖ It is observed that about 13.98% Default Rate from the given Data Set.
- ❖ 'loan\_status' is one of the Key Variable for Univariate Analysis.
- ❖ Bank was able to recover, 57.13% and 117.24% of Loan Amounts from the Defaulters & Fully Paid Customers, respectively.
- ❖ Most of the loans granted are less than 15000
- ❖ Charged loans shifted towards high average loan amount loans.

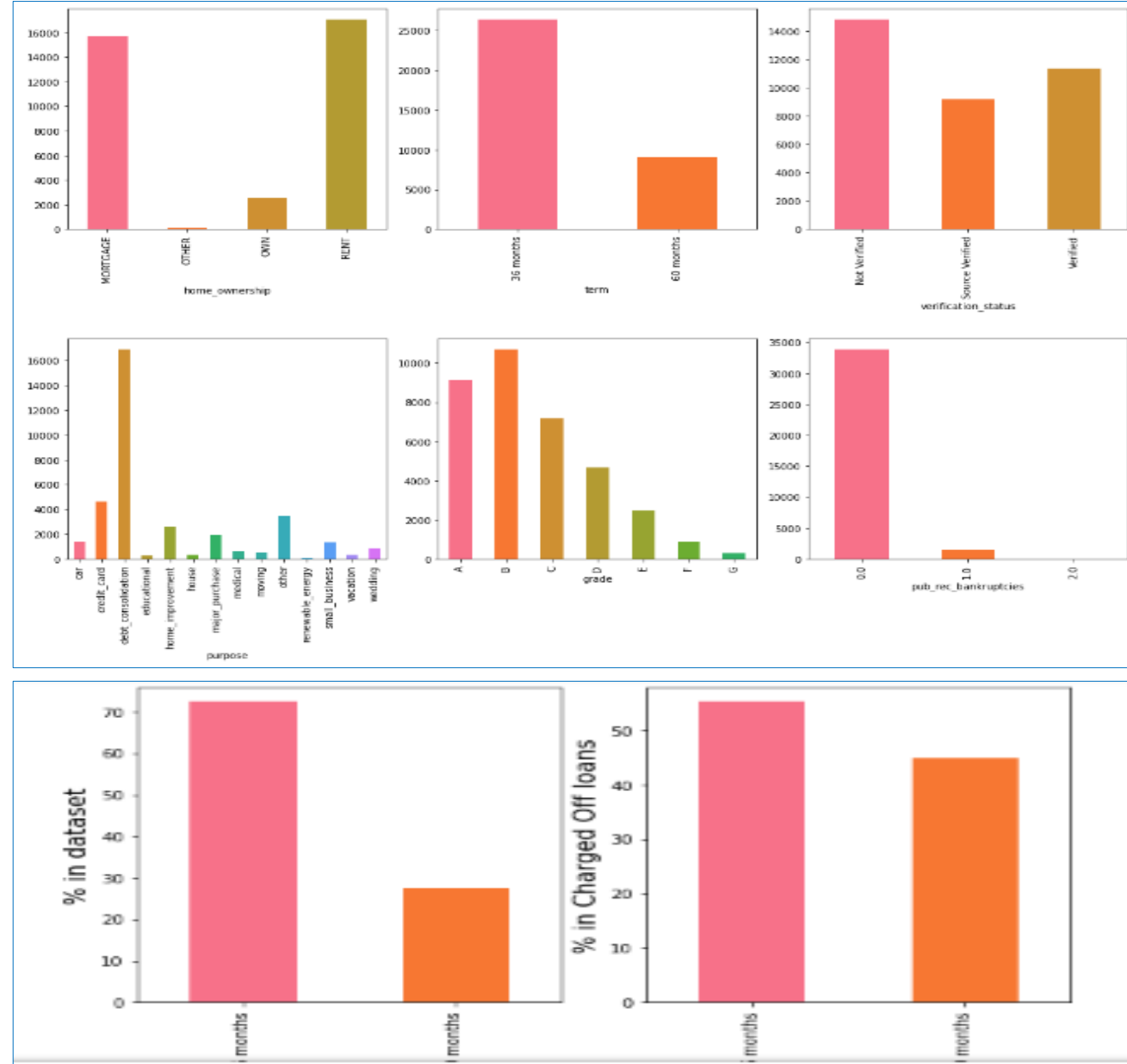
## 2. Key Variables considered for Univariate Analysis.

- loan\_amnt
- term
- Purpose
- installment
- home\_ownership
- annual\_inc
- verification\_status
- emp\_length
- grade and sub\_grade



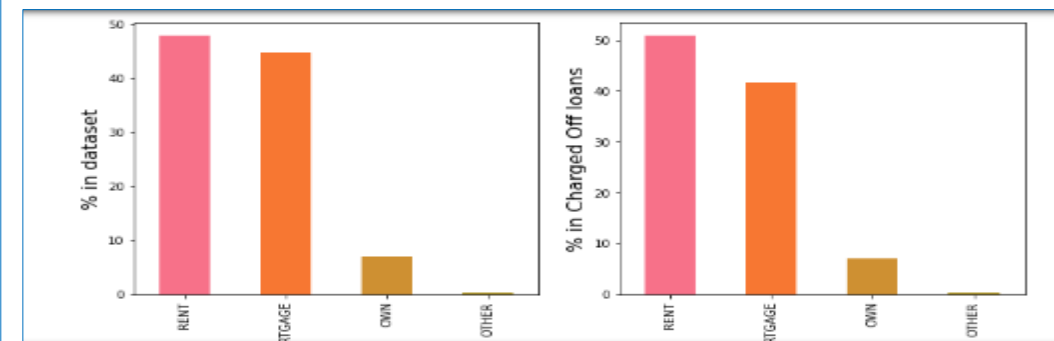
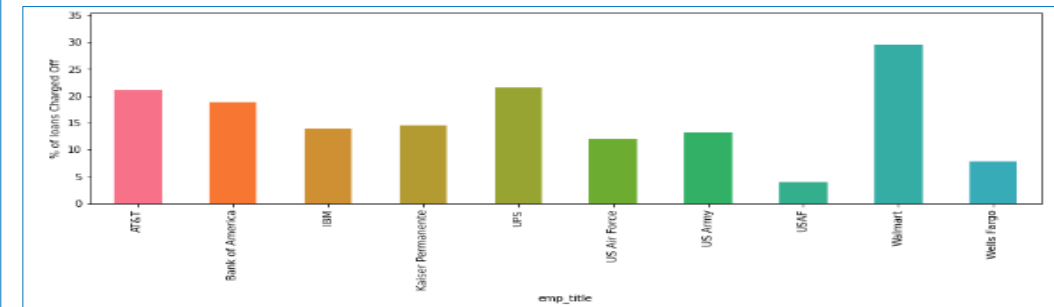
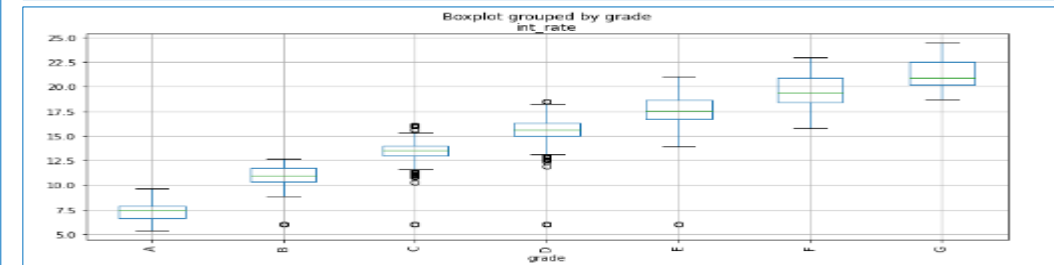
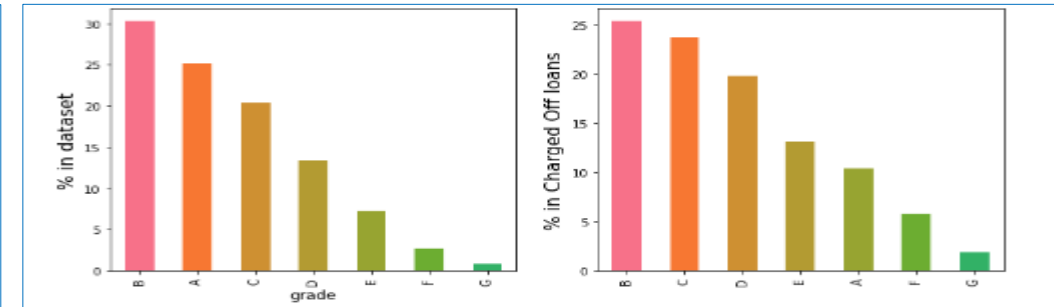
## Key Inferences

- ❖ Most Loans are taken with 36 Months Term.
- ❖ Very few people with Low Bankruptcy Records.
- ❖ Most Loans taken for the debt\_consolidation purpose(49%)
- ❖ Most of the loans are categorized under Grade B.
- ❖ Most of the People who took loan don't have their Own House.
- ❖ Default Rate is higher in case of those who opted 60 Months Term.
- ❖ The % age of Default Reate increases for those loan given to 2 ore more than 2 pub\_rec\_bankruptcies.
- ❖ Interest Rate of Defaulter is observed to be higher than Fully Paid.
- ❖ Higher the Installments, Higher is the Default Rate.



## Key Inferences – Contd.

- ❖ Grade A and Grade B Customers seems safe. Grade D, E, F and G have very less Default Rate.
- ❖ Grade E seems Risky, so Bank either can refuse loan or charge higher interest rates.
- ❖ The Lending Club charges higher interest rate for lower grade loans.
- ❖ Maximum Loans are taken from US Army, Bank of America and then followed by IBM, AT&T.
- ❖ Maximum Default Rate is from Walmart and UPS.
- ❖ It is observed that there is no impact on Default Rate due to employment length and home ownership.
- ❖ Default Rate is higher for Customers whose Annual Income is low.
- ❖ Number of Loans that are given to Not Verified are more compared to those who Verified.
- ❖ It is observed that Default Rate is Higher incase of those who have been Verified which is unexpected.  
**Suggestion : Lending Company need to reevaluate/strengthen their Verification Process**
- ❖ The Number of Loans increasing year by year, and it is observed that Number of Loans issued are quite high in the month of December.
- ❖ Default Rate is higher in 2007, Year of Loan, has no significant impact on Default Rate.
- ❖ Higher the DTI Rate, Higher is Default Rate.
- ❖ Those who make 2 or more than 2 inquiry is high default rates



## loan\_amnt vs time

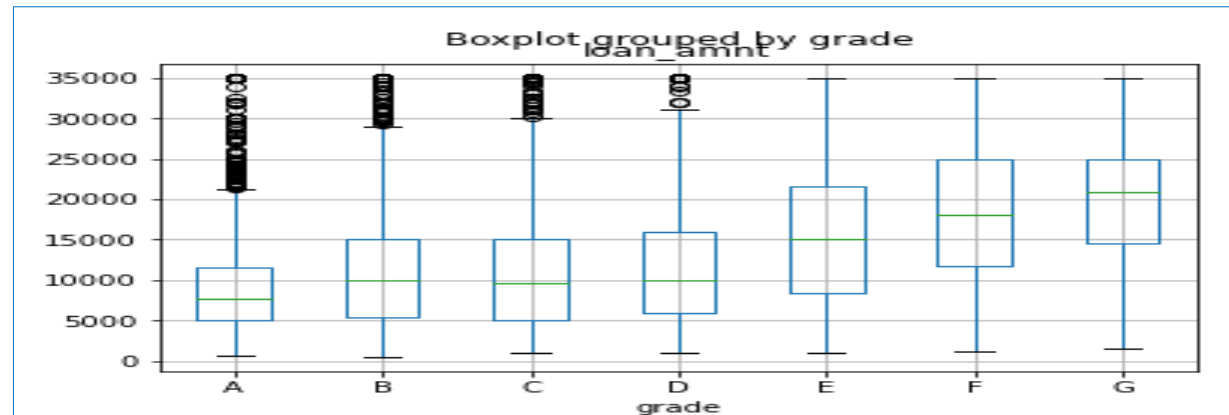
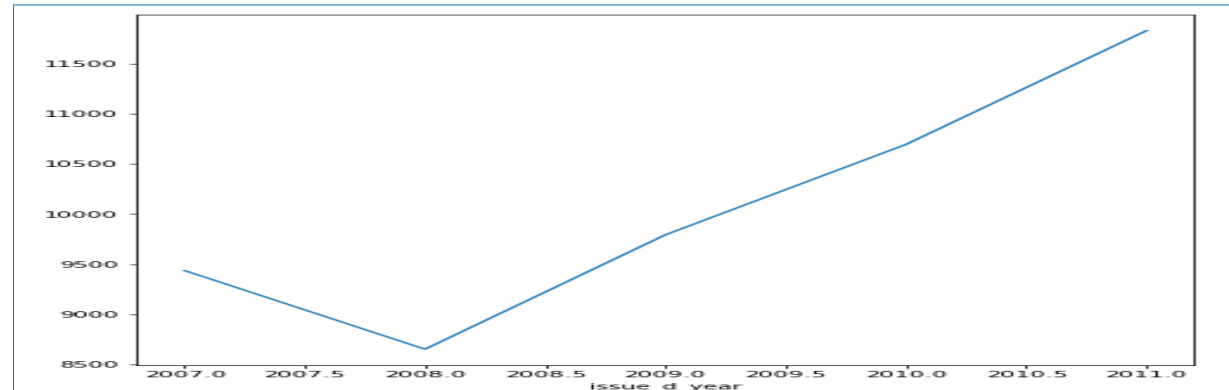
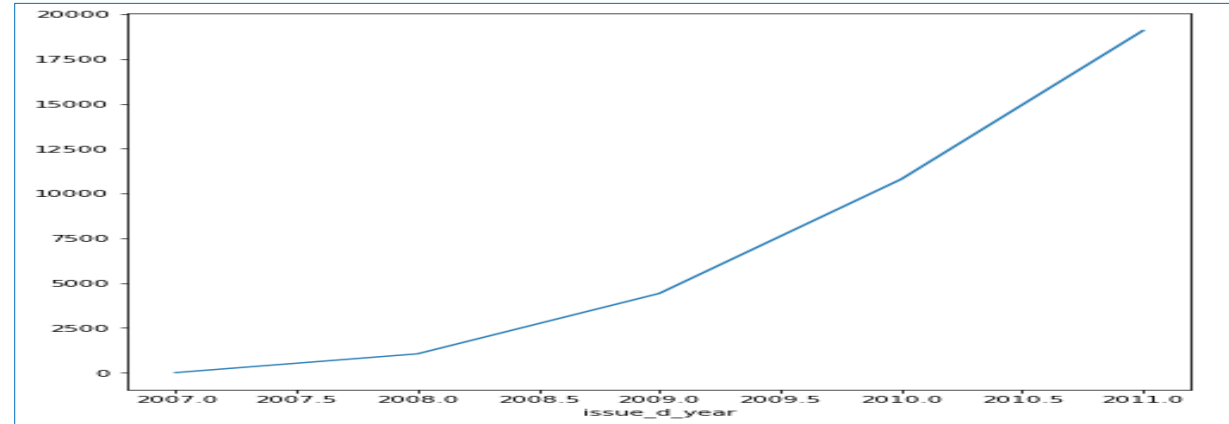
❖ Loan Amount given is increasing Year by Year.

## Avg. loan\_amnt vs time

❖ Average Loans taken in year 2008 is very less.

## loan\_amnt vs grade

❖ Loan amount taken is higher incase of Grades E, F and G



## loan\_amnt vs annual\_inc

- ❖ There are people whose average income is lower than 50000 taking loans 25000 or higher. **Those loans are risky for company**

## loan\_amnt vs verification\_status vs loan status

- ❖ Default Rate is high incase where it has been verified by the company.

## annual income vs grade

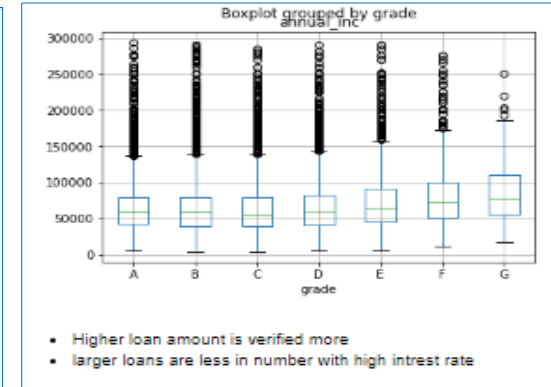
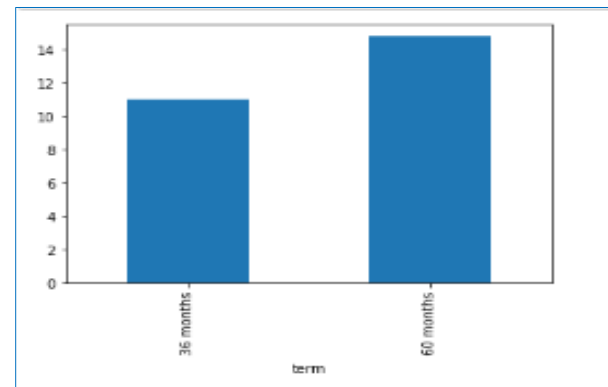
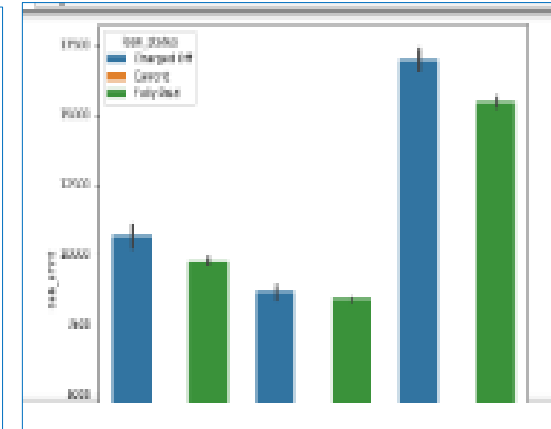
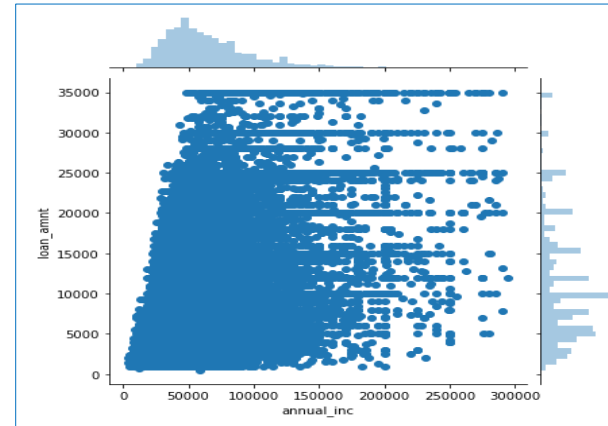
- ❖ Higher loan amount is verified more
- ❖ Larger loans are less in number with high interest rate

## Interest Rate and Term

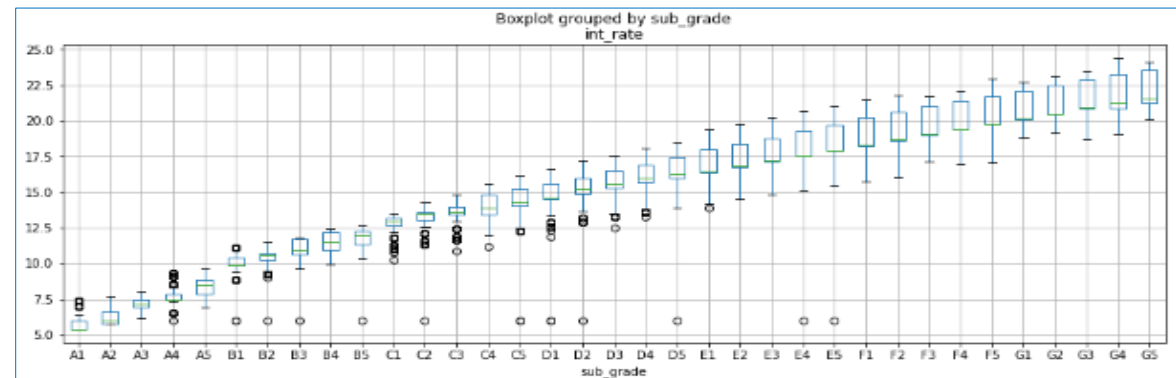
- ❖ Large amounts is given for loans for higher term. The rate of interest is also high

## Interest Rate vs Sub Grade

- ❖ Loans at higher interest rate are more charged off
- ❖ Interest rates directly with the subgrade. larger the subgrade higher the interest rate



- Higher loan amount is verified more
- Larger loans are less in number with high interest rate





## Interest Rate vs Sub Grade

- ❖ Loans at higher interest rate are more charged off
- ❖ Interest rates directly with the subgrade. larger the subgrade higher the interest rate

## Loan amount vs Grade vs Term

- ❖ Higher loan amount with lower grade for longer terms

## Loan amount Vs term Vs Loan Status

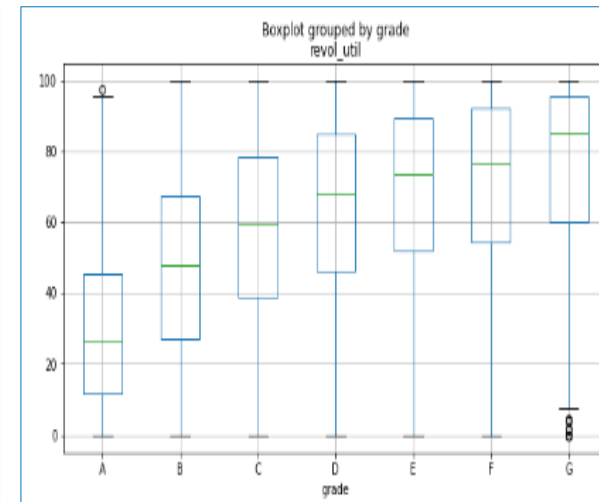
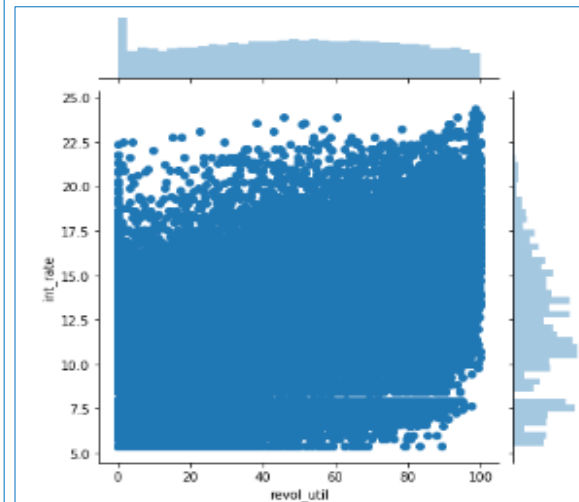
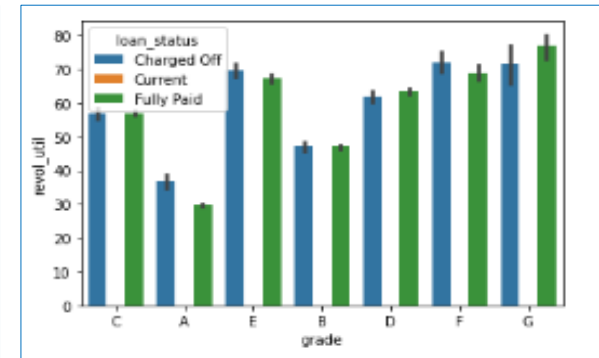
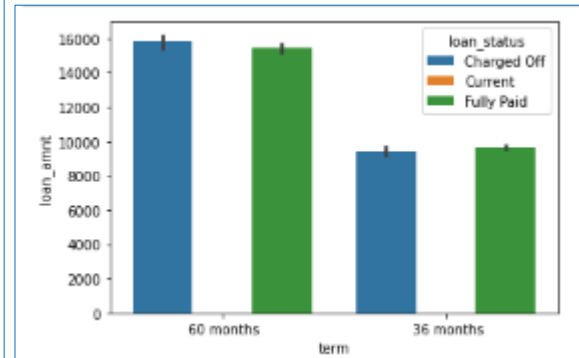
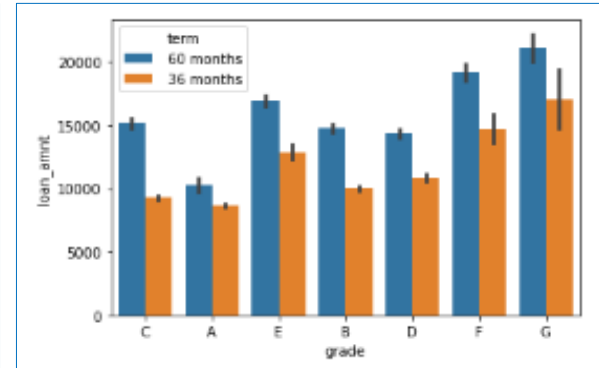
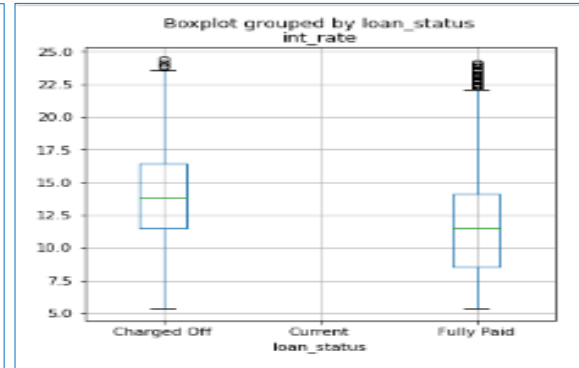
- ❖ Higher loan amount and with longer term and see higher default rate of loans

## revol\_util Vs grade Vs Loan Status

- ❖ revol\_util and grade are correlated. higher loan amounts related to with higher grades

## revol\_rate Vs int\_rate

- ❖ The dti increases proportionally with the number of credit line in clients file.

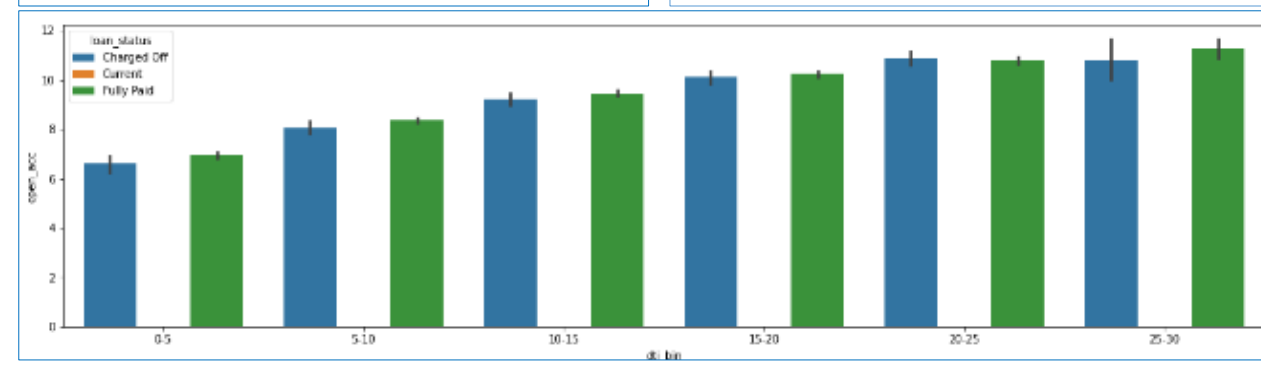
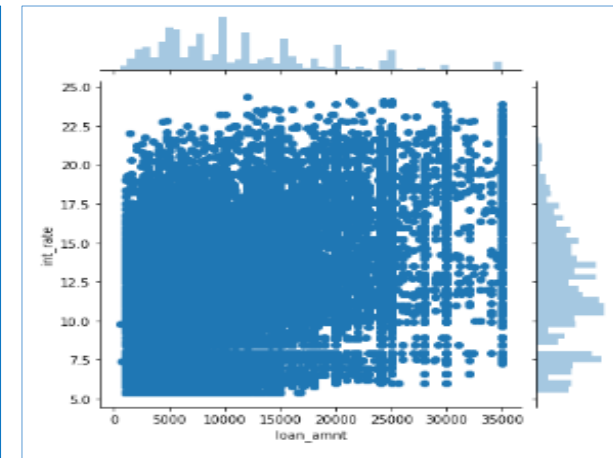
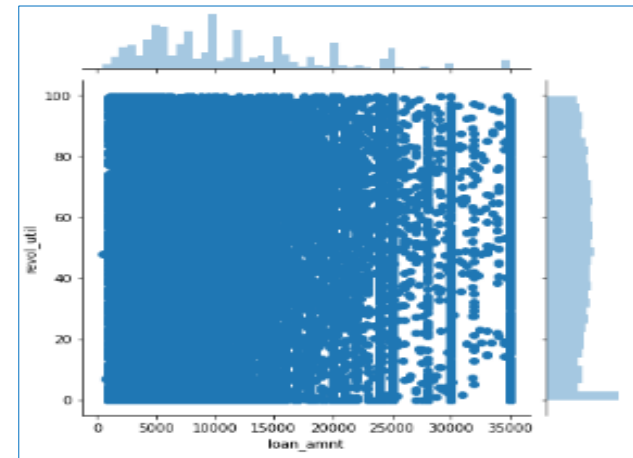
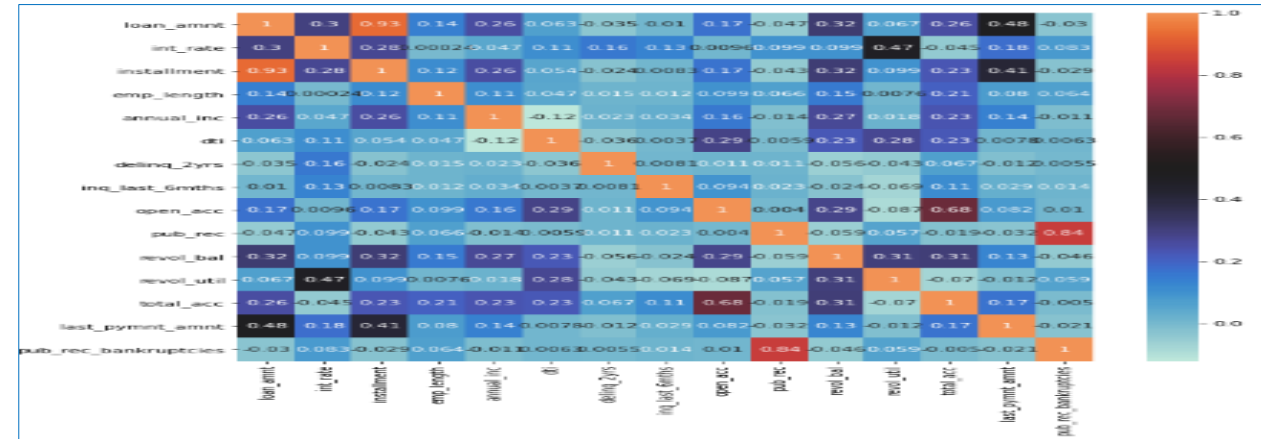


## revol\_util Vs Grade

- ❖ loan\_amnt is correlated to last\_paymnt\_amnt with r factor 0.44
- ❖ The company is charging higher interest from the risky loans. int\_rate is correlated to revol\_util with r factor of 0.47
- ❖ loan\_amnt revol\_bal are correlated with factor 0.35. It suggests that higher loan amount is being given to risky clients.

## dti Vs open\_acc

- ❖ The dti increases proportionally with the number of credit line in clients file.



## Driving factors (or driver variables) behind loan default :

- ❖ **Average Income** : People whose average income is lower than 50000 taking loans 25000 or higher.
- ❖ **Verification** : Default Rate is observed very high in those cases where it has been verified by the company, which is not expected. It is suggested strongly that company need to re-evaluate /strengthen their verification process.
- ❖ **Interest Rates** : Higher the Interest Rate, Higher is the Default Rate. Also, Interest Rate is directly proportional to Sub Grades, Higher the Subgrades, higher is the Default Rate.
- ❖ **Higher Term (60 Months)**: Default Rate is observed to be high, where people opted for Higher Term, Large Loan amounts and Higher Interest Rate.
- ❖ Higher the **DTI** Rate, Higher is Default Rate.
- ❖ Those who make 2 or more than **2 inquiry** have higher Default Rates
- ❖ People with **Grade E** are more Risky, so Bank either can refuse loan or charge higher interest rates.