ML C33 - Lending Club Case Study

Members:
Dhandapani Subramanian
Srinivas Raghupathi

Case Study - Overview

Business model background: Lending Club consumer finance company which collaborates the Investors and borrowers and helps investor to reach financial goals in one hand and facilitates borrowers to avail personal, business, car and medical loans on other hand. It offers vide range of financial products and services through technology driven platform.

Business objective of the study: Identify the risks of a loan application based on driving factors which are strong indicators of the loan default. The outcome of the study should recommend or guide credit approver to accept a loan application or reject it.

Data Analysis - Approach

Raw Data Source :
Loan data set contains the complete loan data for all the loans issued through the time period 2007 to 2011. Data dictionary illustrates meaning of each variable used in data set

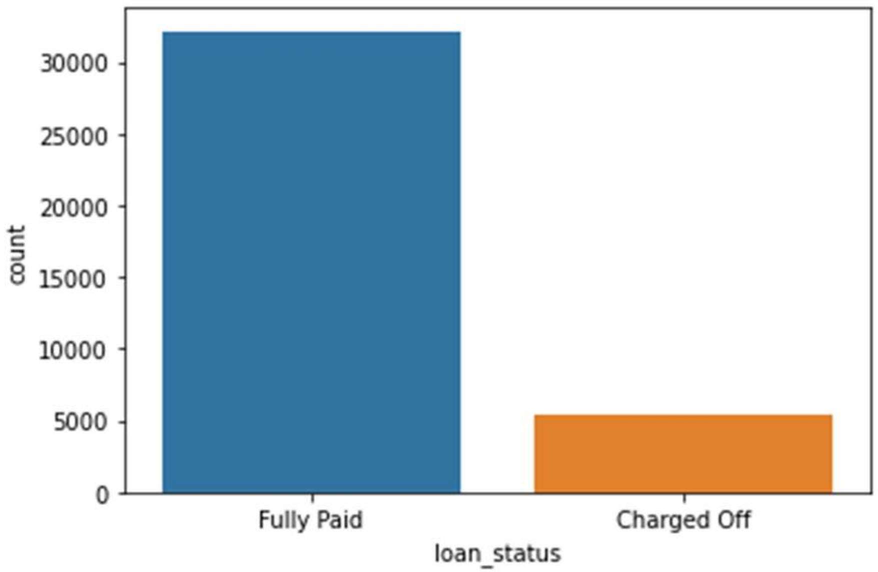
Data Clenching :

- Analyzed each column in the data set and their influence on target variable which is “loan status” and dropped wherever missing values are >30%.
- Dropped the complete columns which are not relevant for analysis to focus on more relevant data set.
- Dropped columns which had all unique values through out the data set as well as columns with all duplicate values.

EDA – Categorical Variables

Variable : Loan Status

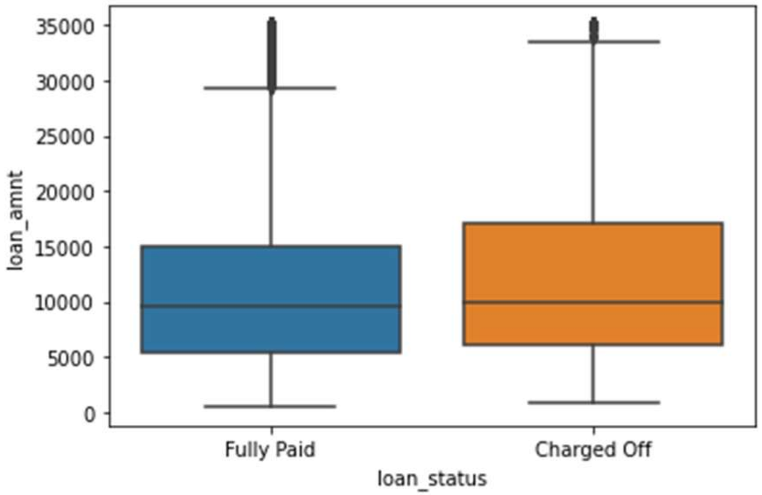
- The column loan illustrates Fully paid loan (vs) Charged off loans, around 85% loans were fully paid which indicates healthy customer base.



Fully paid: 85.6%
Charge off : 14.4%

EDA – Categorical Variables

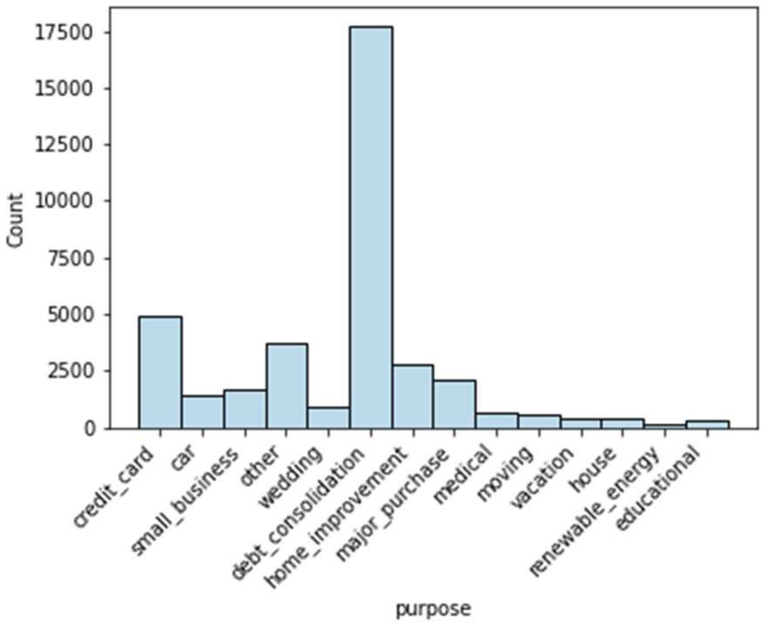
Variable : Loan Status

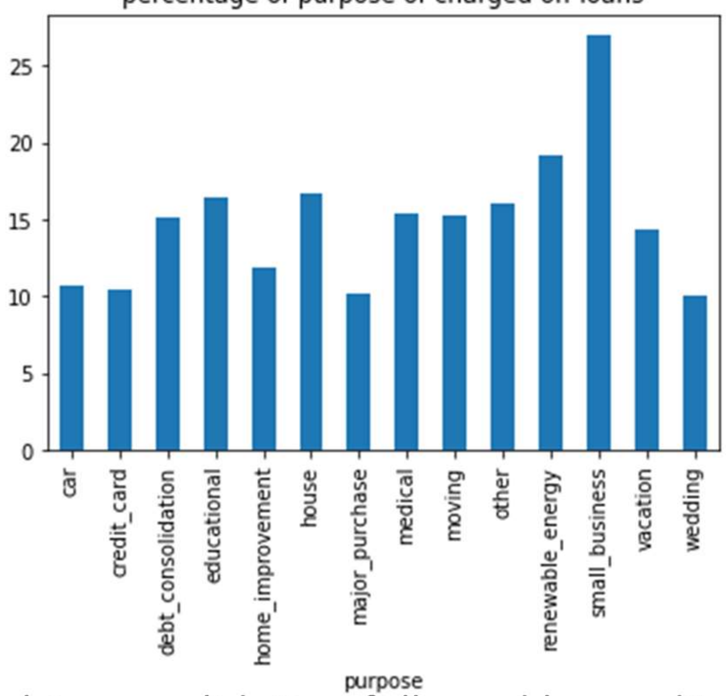


- Loan amount of “Charged off” loans are little higher than that of “fully paid” loans on an average.

EDA – Categorical Variables

Variable: Purpose of loan

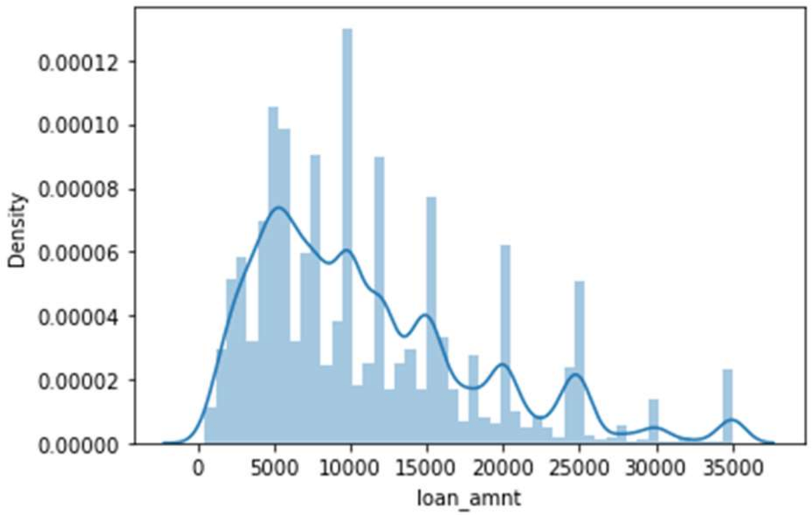


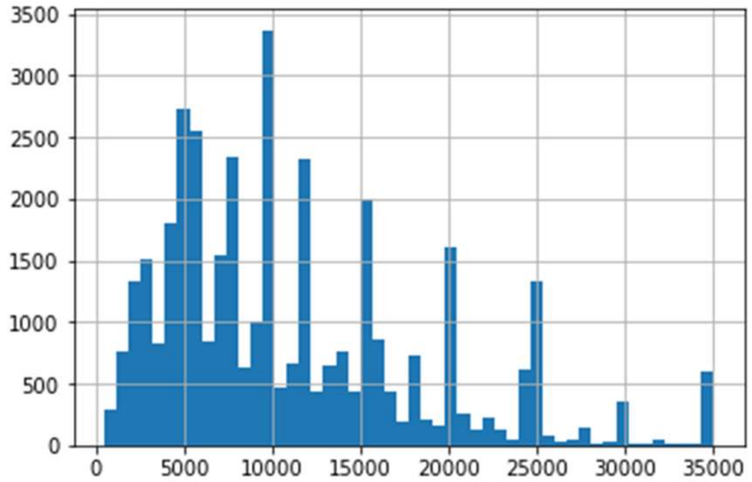


- We see majority of loans were taken for debt_consolidation followed by credit card. But they were mostly fully paid. Therefore, indicating a profitable deal for Lending Club.
- Small business is the category which shows the highest percentage of charged off loans. We recommend Loan Club to carefully analyze and reduce small business loans.
- We suggest to give a greater number of Loan for wedding, major purchase, car and credit card as they are most likely to be fully paid.

EDA – Numerical Variables

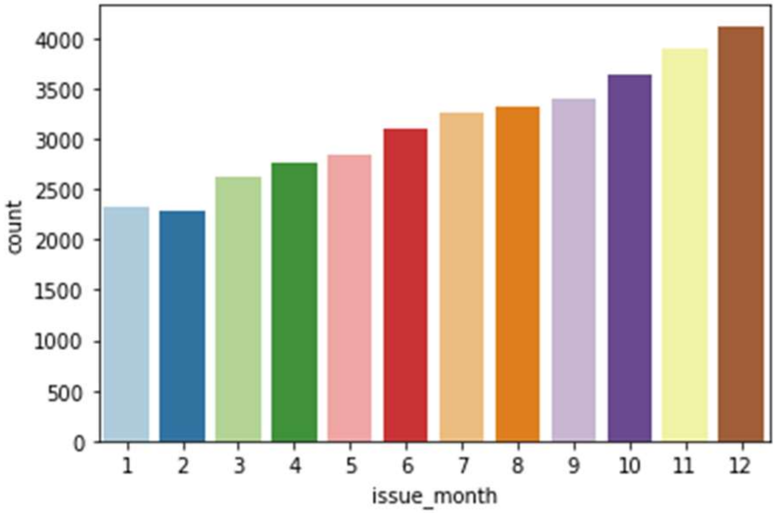
Variable: Loan amount

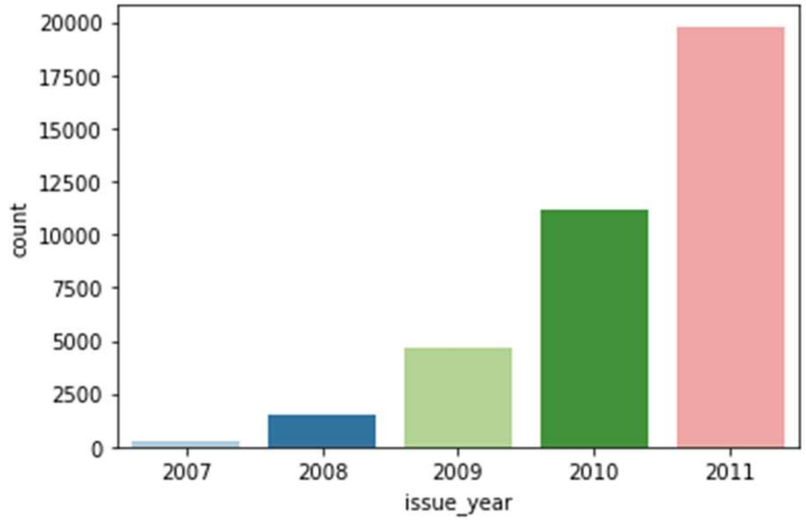




- Loan amount seems to be spiking at round figures around each $5000.
- Greater number of loans are taken for amount in between $5000 and $15000

EDA – Time Series – Month| Year



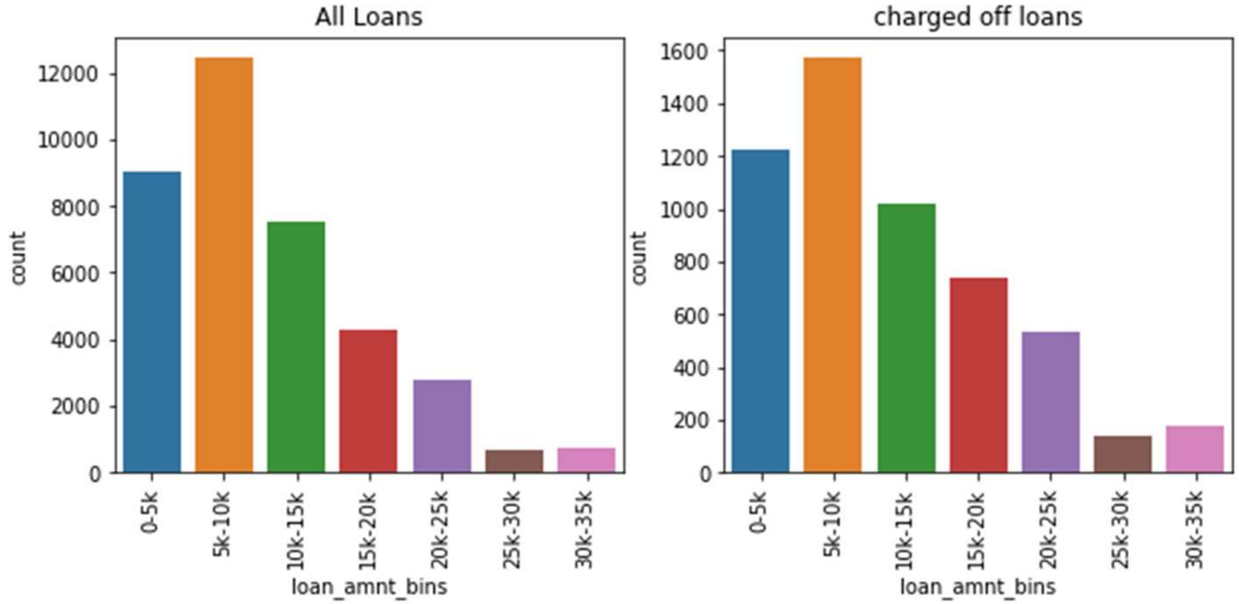


Lending Club has a significant growth year on year growth. The highest number of loans were discharged in year 2011. This shows a good profitable growth for Loan Club.

An observation, most of the loans were taken in year end around the month of December and November.

EDA – Bivariate

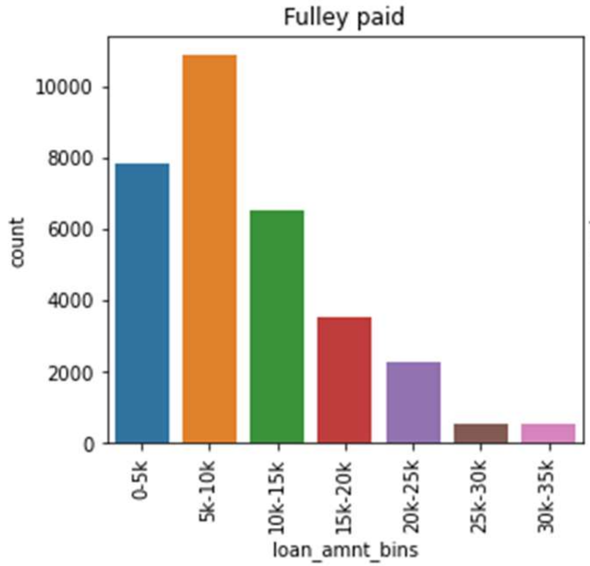
Variables: Loan Amount and Loan Status

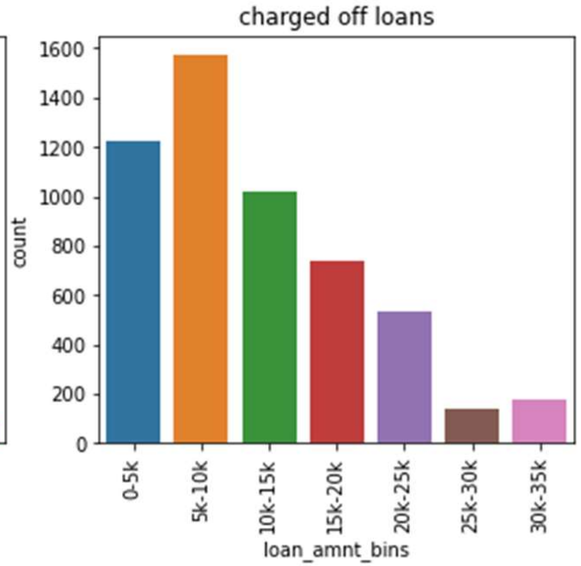


Loan amount segmented into buckets of 5000

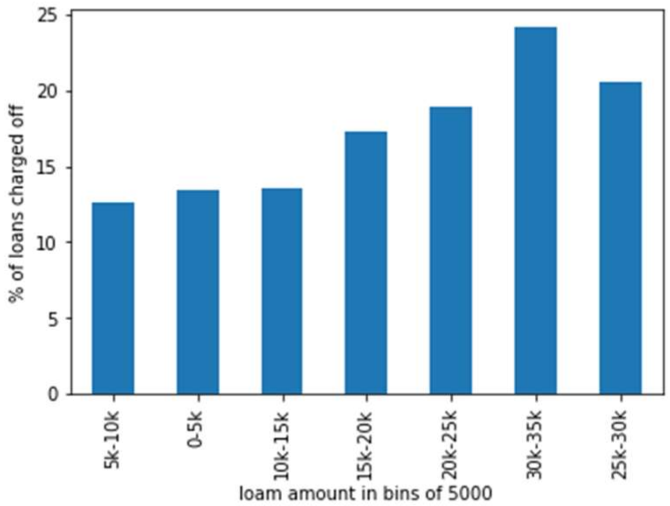
Major percentage of loans availed between $5K and $10K and "Charged off" loans are also distributed similarly.

EDA – Bivariate





Trend in fully paid vs charged off loans are similar for most of the smaller amounts while looks growing for 15k and above buckets of loan amount

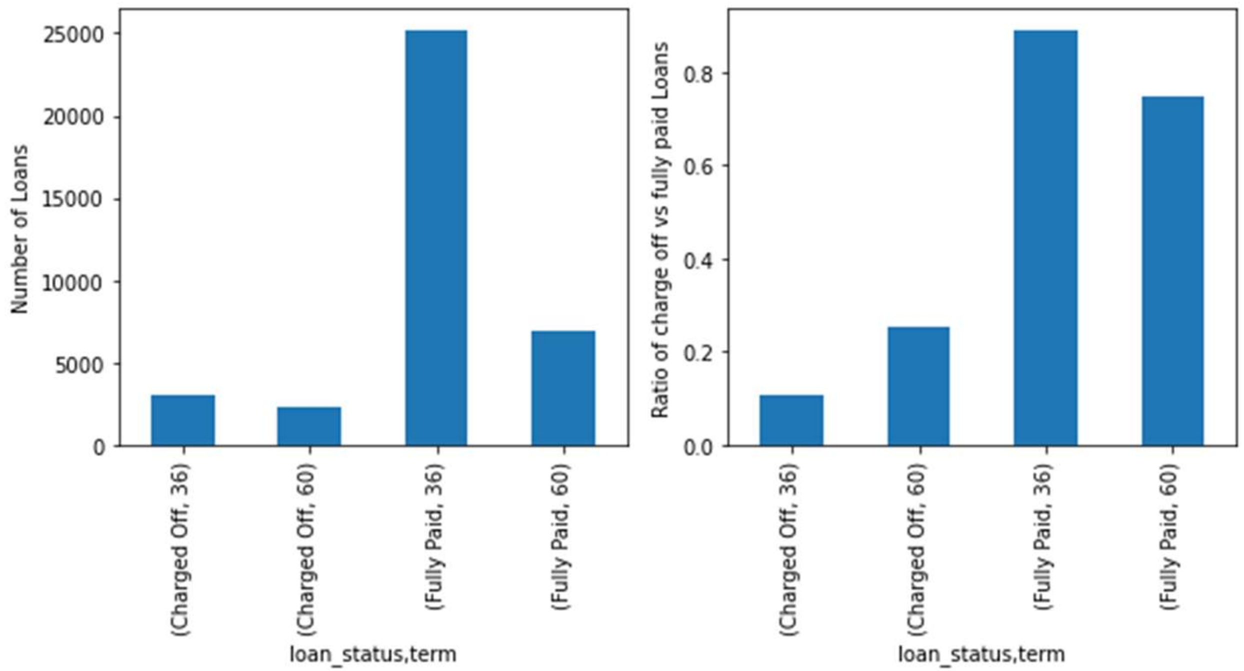


Percentage of charged off loans increases significantly as we go on higher loan amounts. Higher amount loans though less in number but still has a high chance of getting charged off.

Charged off percentage is high for loan amount more than $30K

EDA – Bivariate

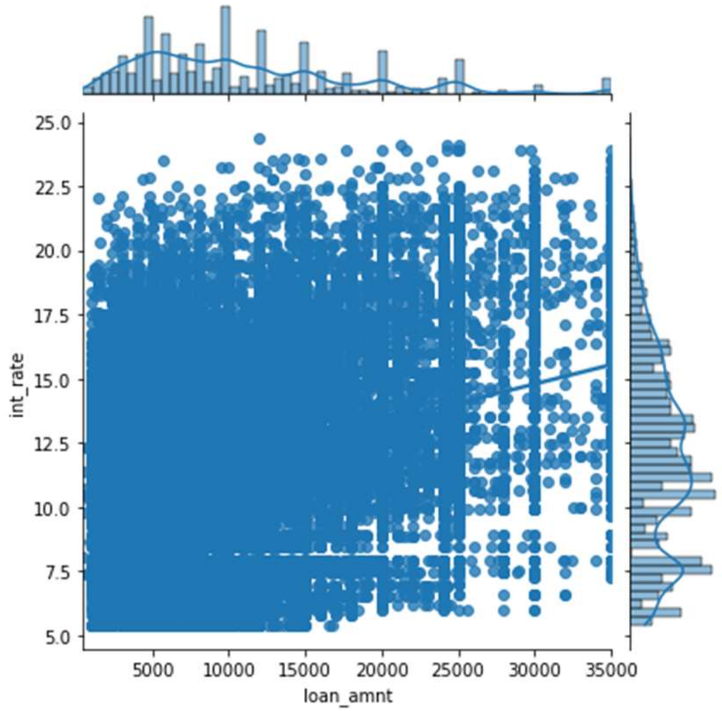
Variables: Duration or Term of loan (36|60 Months)

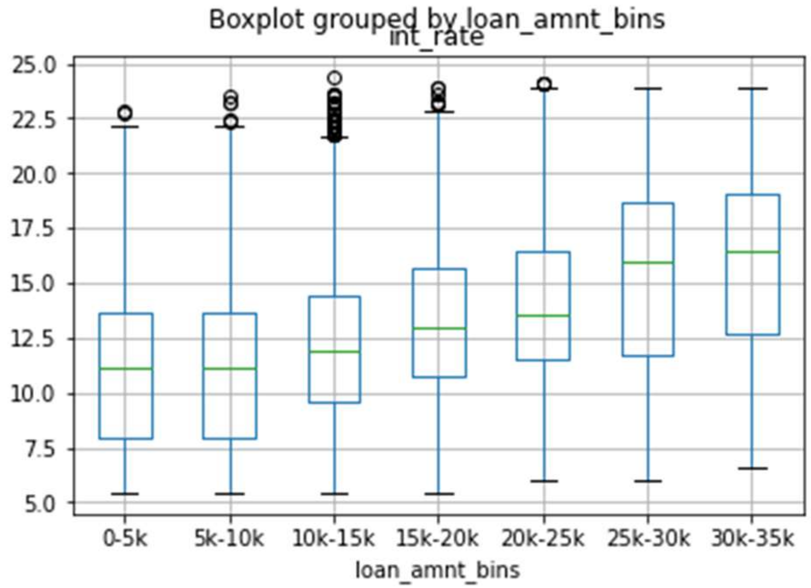


- One out of four loans with tenure 60 months end up in charge off.
- Loans of shorter term 36 months were seen more prominantly paid back. So we recomend promoting it.
- Recomadation would be to carefully analyse the loans for greater tenure, and reduce approvals for 60 months loans.

EDA – Bivariate

Variables: Interest rate and Loan Amount

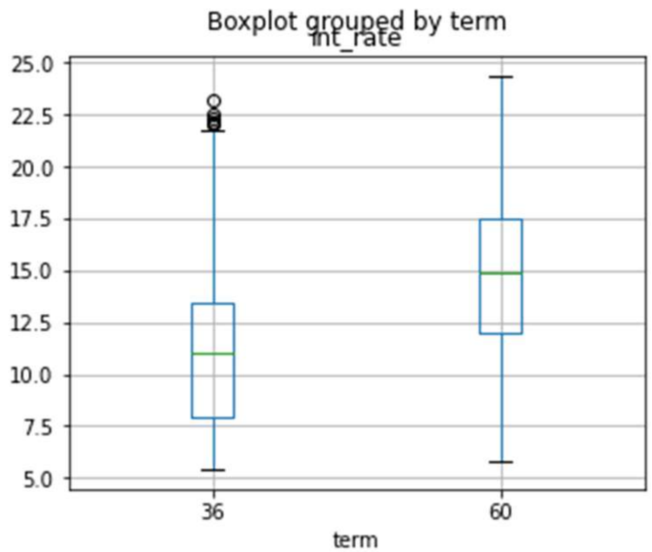


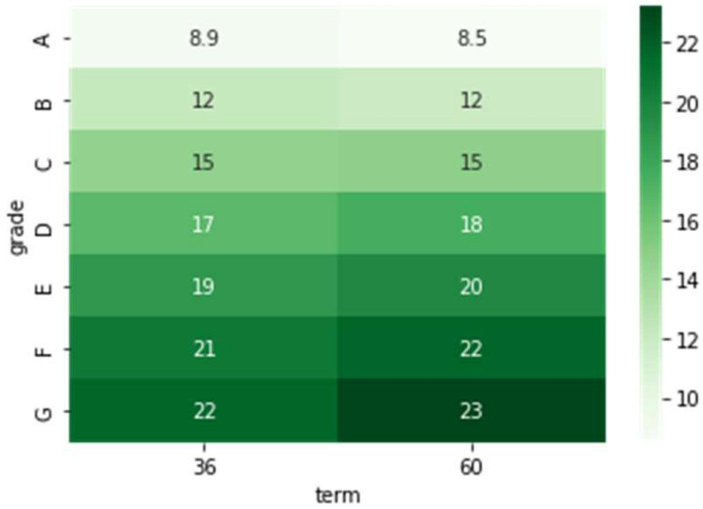


- Higher Interest rates are set for higher loan amounts. We see a direct raise in it. The regression line also shows an increasing trend.
- Boxplot also strengtens the fact that interest rate increases for increase in loan amount.

EDA – Multivariate

Variables: Interest rate, loan and loan status

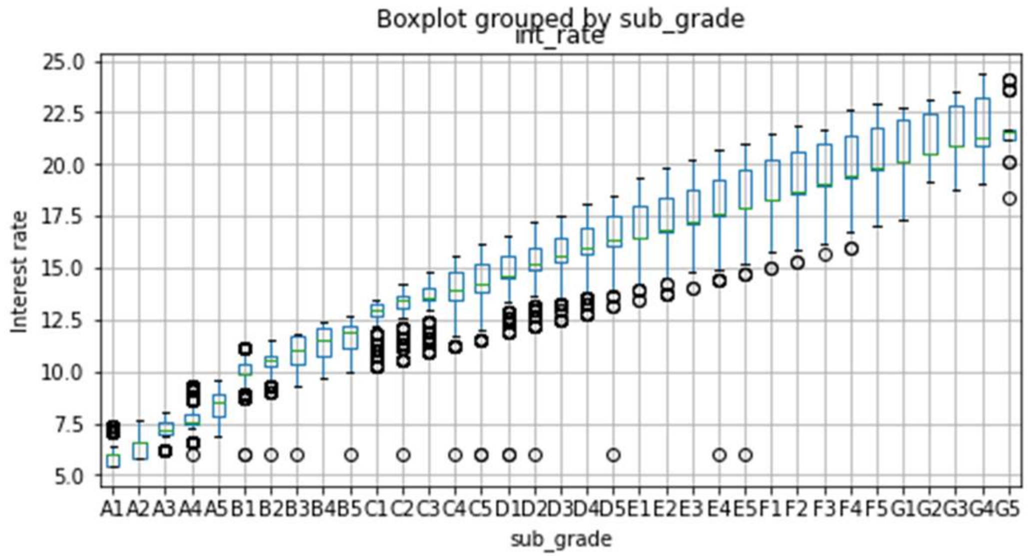


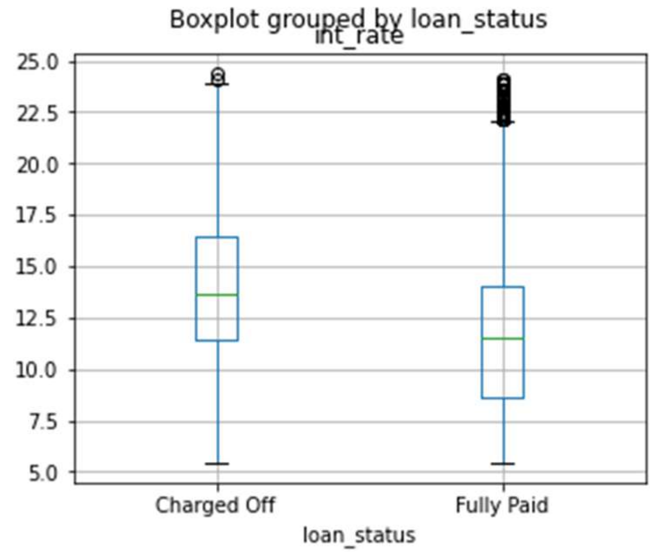


Interest rates are high for 60 month loans. Charge off percentage is high for high interest rate loans.
Higher grades end up with higher percent of interest rate.

EDA – Multivariate

Variables Interest rate, subgrade and loan status

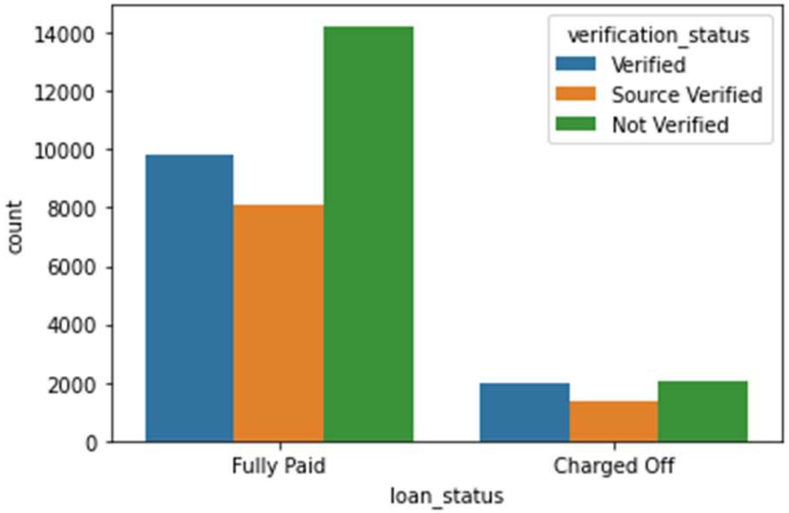


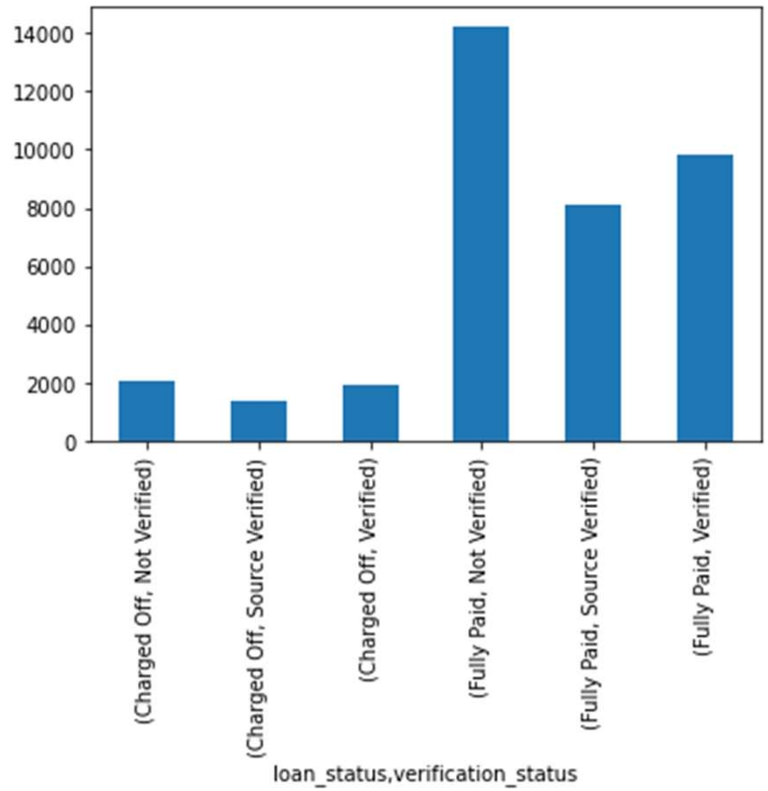


- Higher grade/subgrade indicate higher interest rate for the loan.
- Charged off loans also show a higher interest rate. So we can correlate that when interest rate increases, the chances of loan getting charged off increases as well.

EDA – Multivariate

Variables: loan amount, verification status and loan status

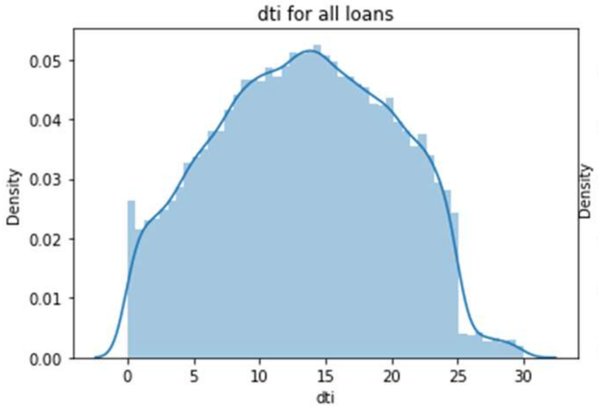


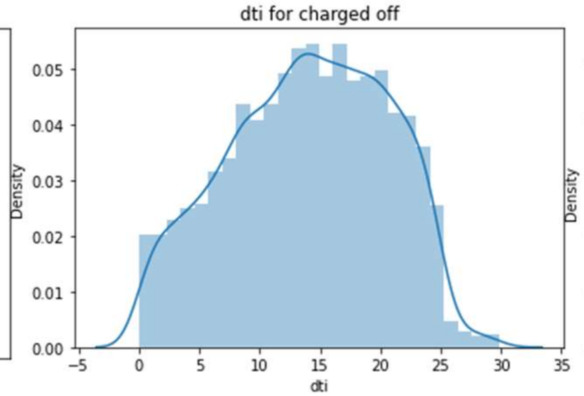


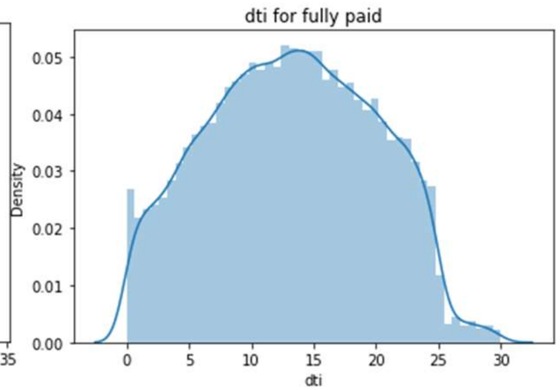
- Non verified status does not show any significant inference for charged off loans.

EDA – Multivariate

Variables: Debit to income (Dti) ratio, loan status and loan amount



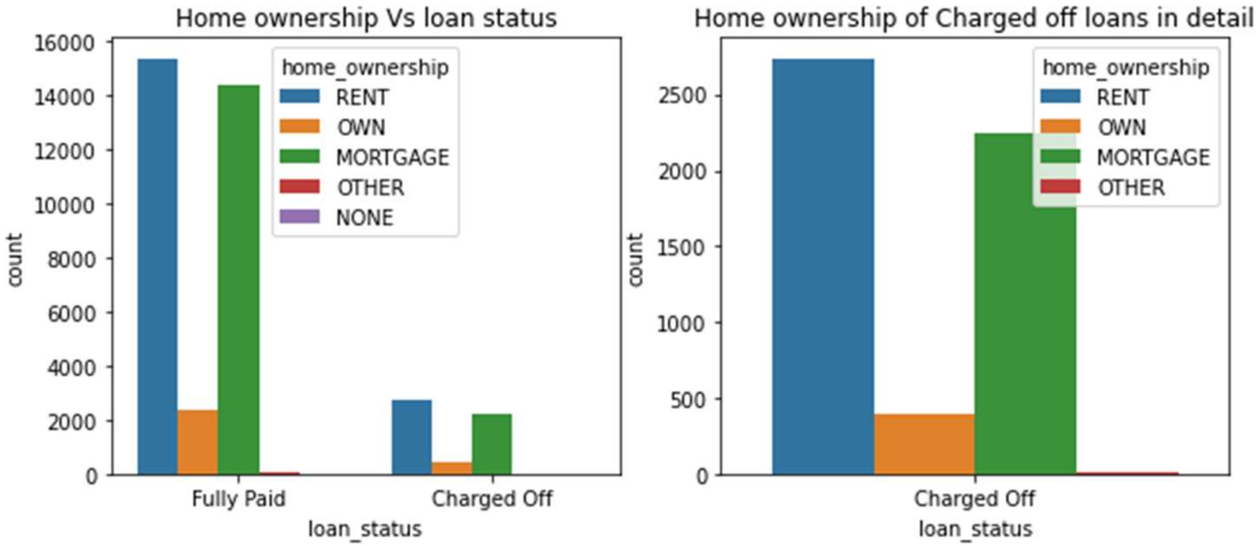




- Debit to income seems to similarly distributed for both fully paid as well as charged off loans after 15.
- **Dti less than 15 has more full paid loans.**

EDA – Bivariate

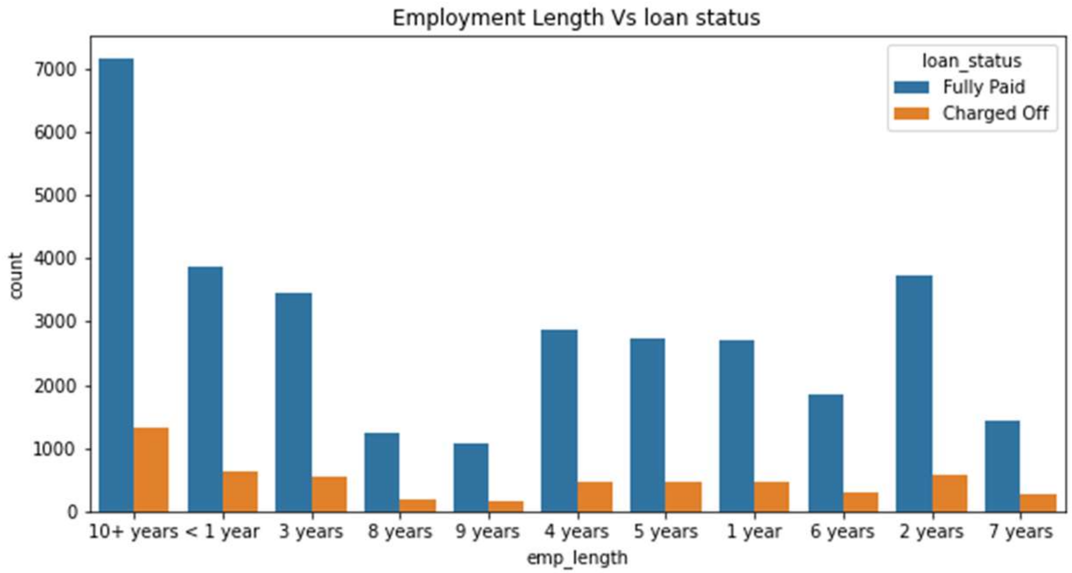
Variables: Home ownership and loan status

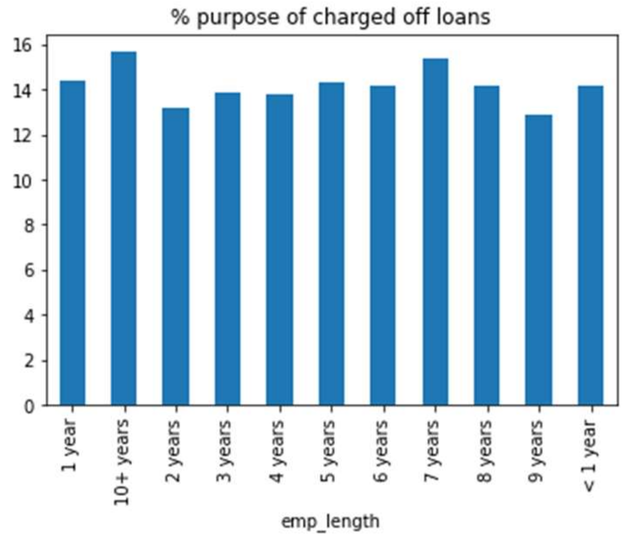


- People in Rented house and Mortgate seems to have more chances of charged off than that of people in own house.

EDA – Bivariate

Analyzing Employment length

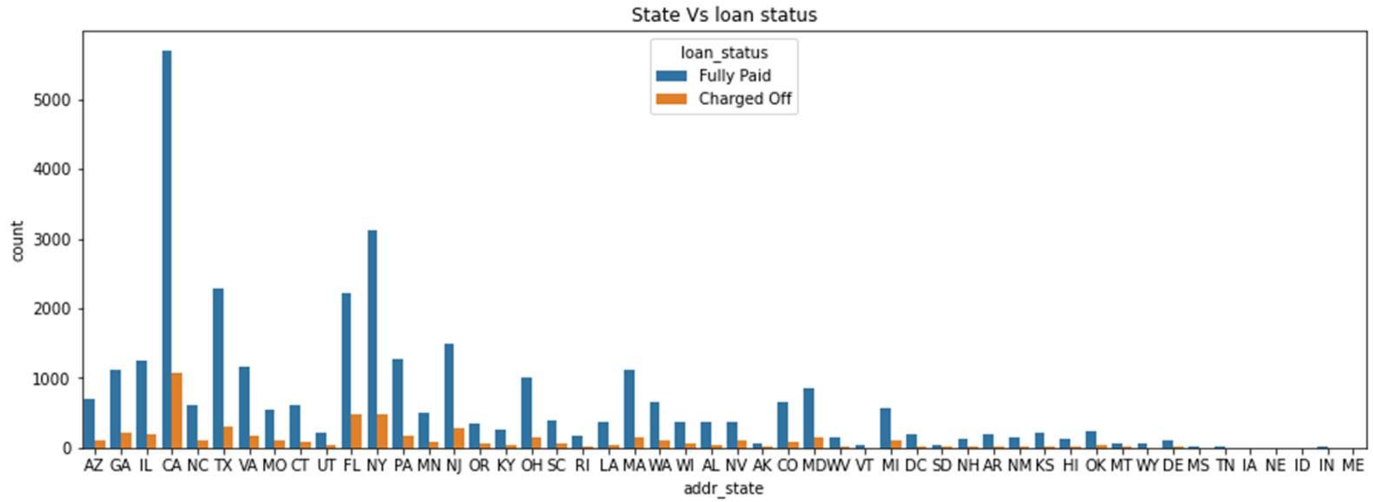


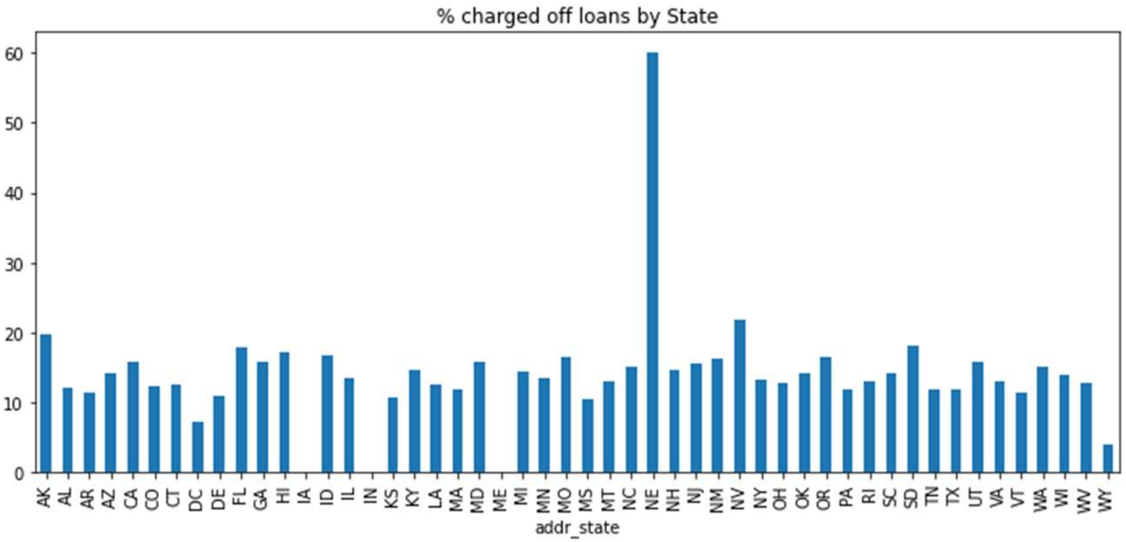


- People with employment length 10+ years tend to take more loans.
- Assuming the population would be more in 10+ years as it might contain, employees with 10 to 40/50 years of employment
- % of charged off loans also shows slightly higher for 10+ years followed by 7 years

EDA – Bivariate

Analyzing state wise loan status





- State CA has the highest number of loans.
- IA, IN and ME are those with 100% loan repayment rate. They are more profitable customers
- NE though having less number of loans, shows a high rate of charging off. Lending Club must double check before approving loans from this state

Data Analysis - Assumptions

- Loan status that represents “Current” rows are not relevant to our analysis as these could be Fully paid or charged off some point in time.
- Customer behavior variables were dropped of as they would not be available at the time of loan approval.
- Dropped column with more than 30% null values as imputing them will mislead.
- Desc column, though could give good insight, we dropped as it needs NLP.

Recommendations

- **Purpose of loan** – “Charged off” percentage is higher for “Small Business” loans. Would suggest to reduce number of loans for this category.
  We suggest to give a greater number of Loan for wedding, major purchase, car and credit card as they are most likely to be fully paid.
- **Loan Amount-** “Charged off” percentage is high for the loans >$30K, higher loan amounts needs further examinations.
- **High interest rate and 60 month term** – “Charged off” percentages is high where rate of interest is high and loan term is 60 months, Lending club can think of introducing 18 month and 24 month terms for higher interest rates loans.
  Reduce number of loans for 60 month term.
- **Home owner ship** - Rented house customers have high “Charged off” percentages, Lending club or investor should review the existing KYC procedures. Recovery would be a challenge during loan default.
- **Grade-** Higher grades like G, F & E show more risk of ending up in “Charged off”. Better to reduce number of loans or the loan amount for these grades. Promote more loans for grade A & B.