LENDING CLUB CASE STUDY SUBMISSION

Name : Kalpesh Krushna Zambare.

05/02/2023

Lending Club Loan Assessment for loan default

- Lending Club is interested in understanding patterns which indicates if a person is likely to default.
- This case study is to process the data set received from the Lending Club data from past loan applicants and identify patterns where the loan is successfully repaid or defaulted.
- The assessment helps lending companies to identify risky loan applicants and take early decisions saving credit loss to the company

Exploratory Data Analysis

The data analysis involves below flow for data analysis and problem solving

- **Load the loan data from lending club**
- **Data Learning**

As part of this step carefully understand each columns from the data set, it's possible impacts

- **Data Cleaning**

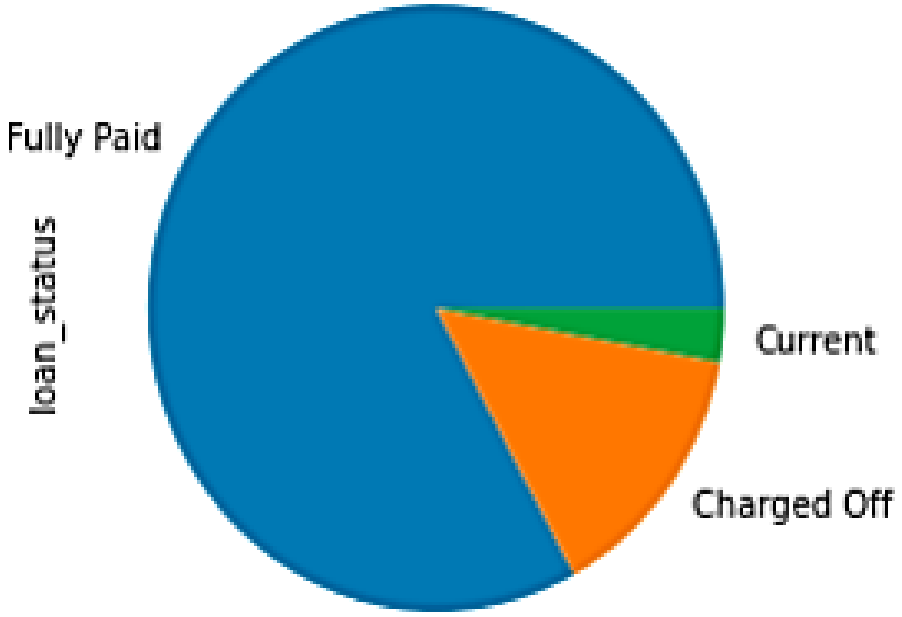
➢ Drop the columns from datasets which are all empty, to reduce the size of the data set
➢ Convert the data to lowercase, add numeric columns for the data to extrapolate graphs
➢ Identify and remove outliers(loan_amnt), Impute or Remove rows which are not relevant (current:2% of entire set)

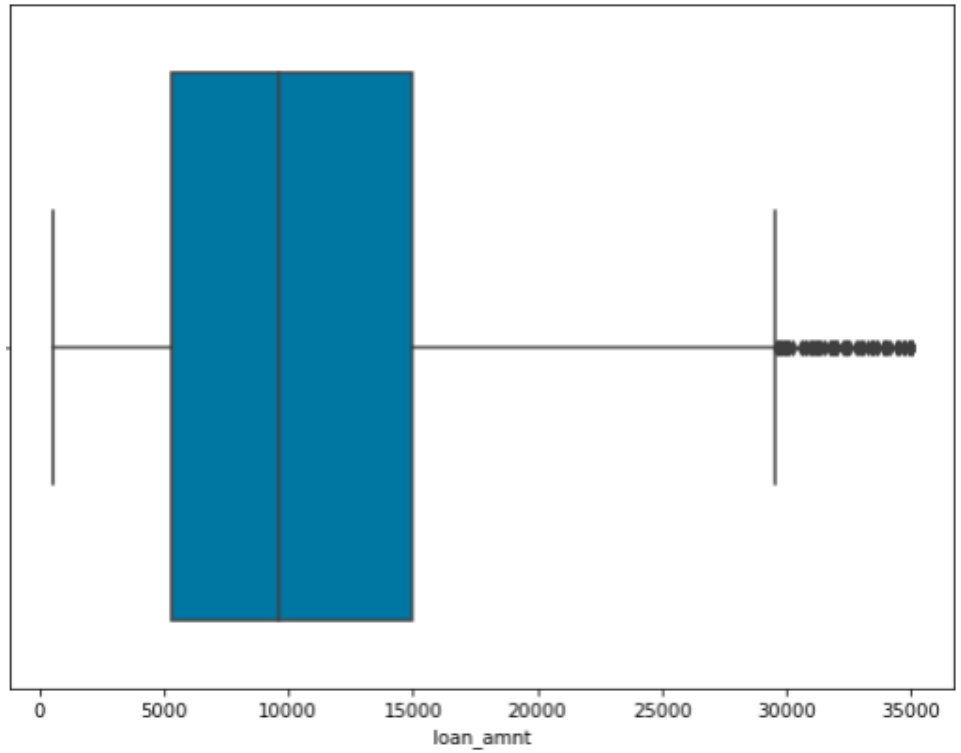
- **Perform univariate analysis using columns {purpose, term, addr_state, verification_status, sub_grade, emp_length, dti} on loan_status**
  - Note the observations from each of the analysis
  - Ignore columns which are of not much significance
  - Plot graphs to understand the behavior

- **Perform Bivariate anlaysis using columns {(term, purpose),(emp_length, purpose), (sub_grade,emp_length), (term,emp_length)}**
  - Note the observations from each of the analysis
  - Ignore combinations which are of not much significance, to avoid confusion
  - Plot graphs to understand the behavior
  - Plot other graphs and data anslysis as required to get more patterns on the dataset.

<Anlysis>

- As part of initial analysis get the data and cleanup the data sets for processing
- Drop rows with loan_status =“current” as it contributes only 2% of the dataset
- Remove loan_amnt outliers, as no much relevant data found in analysis form outlier





<Anlysis>

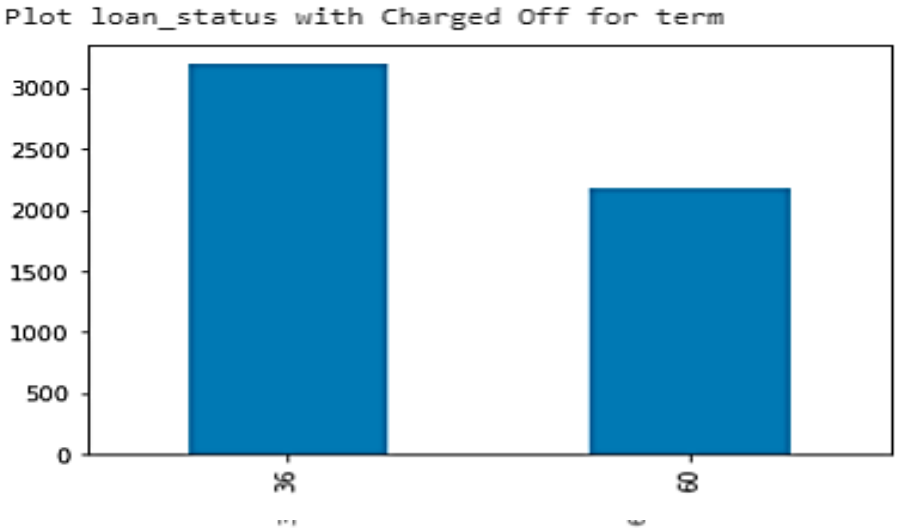
2 nd Part of Analysis involves univariate analysis of the Lending Club data for different columns

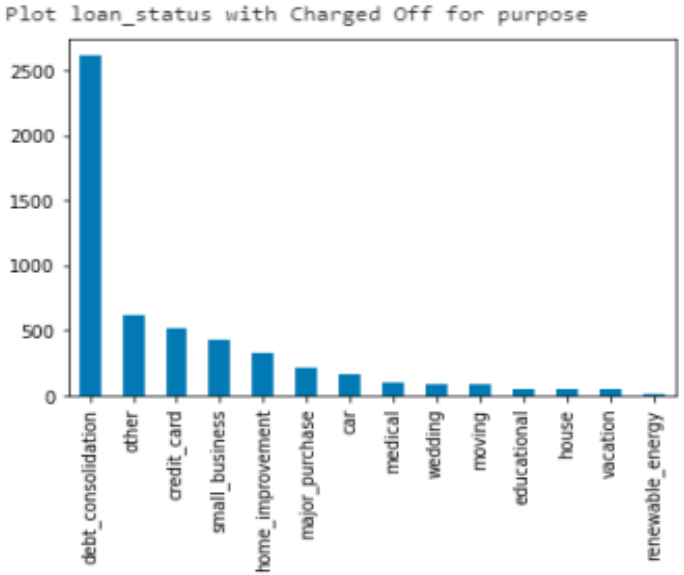
Identify pattern of loan default with purpose of loan.

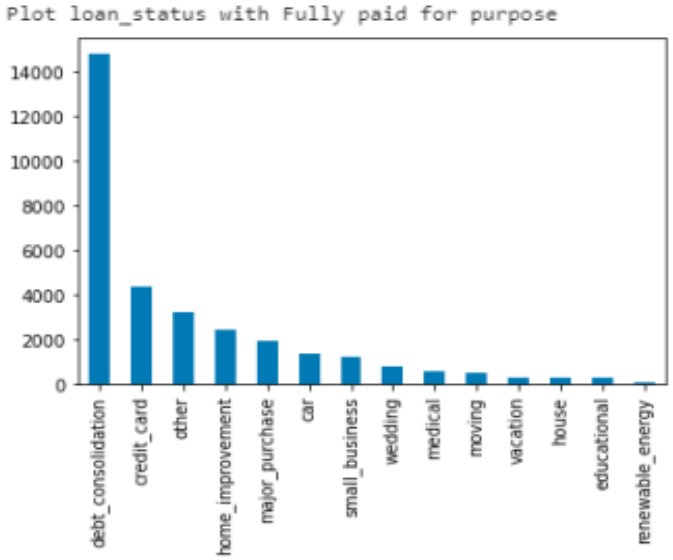
- Fully Paid loan has good success rate for "debt_consolidation"
- "credit_cart" has success rate of aprox 4000/(500+4000) ~= 86%

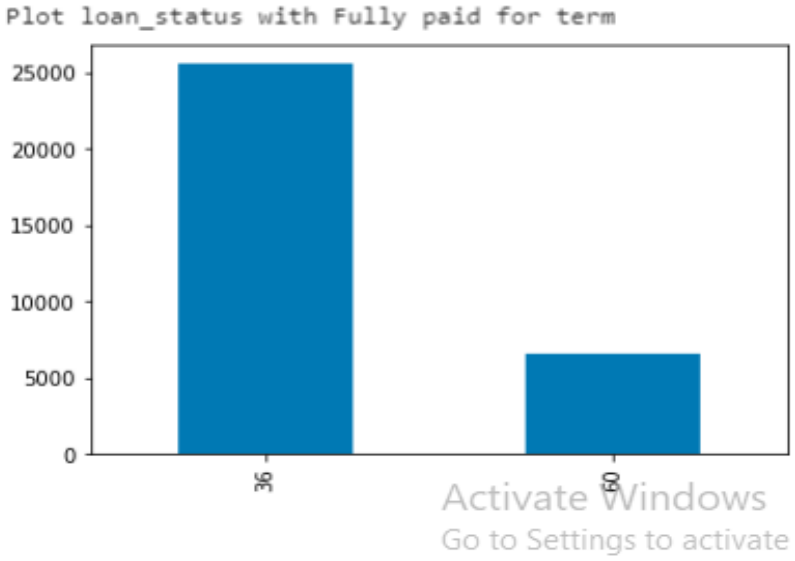
Identify pattern of loan default with loan term

- 36month seems to be the ideal duration to avoid loan defaulte









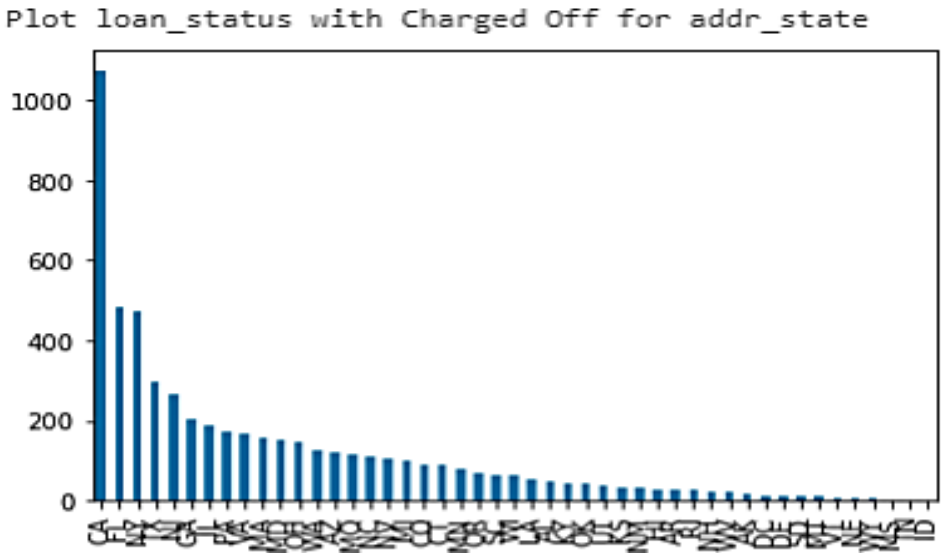
2 nd Part of Analysis involves univariate analysis of the Lending Club data for different columns

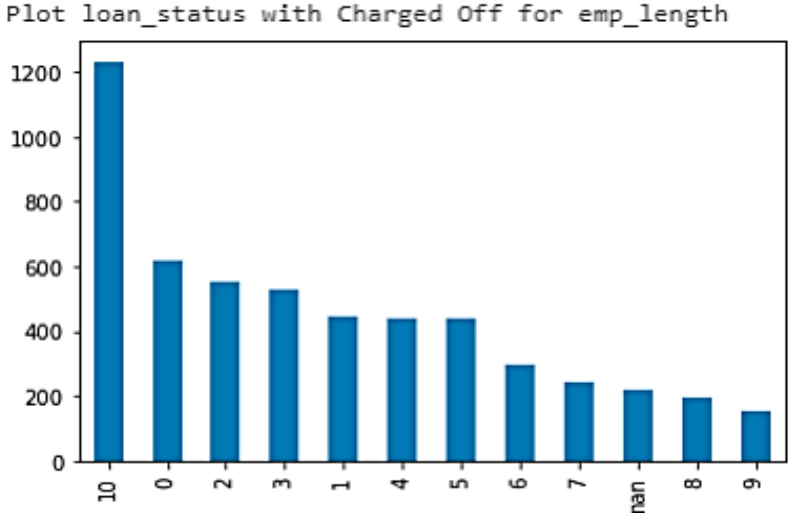
Identify pattern of loan default with addr_state.

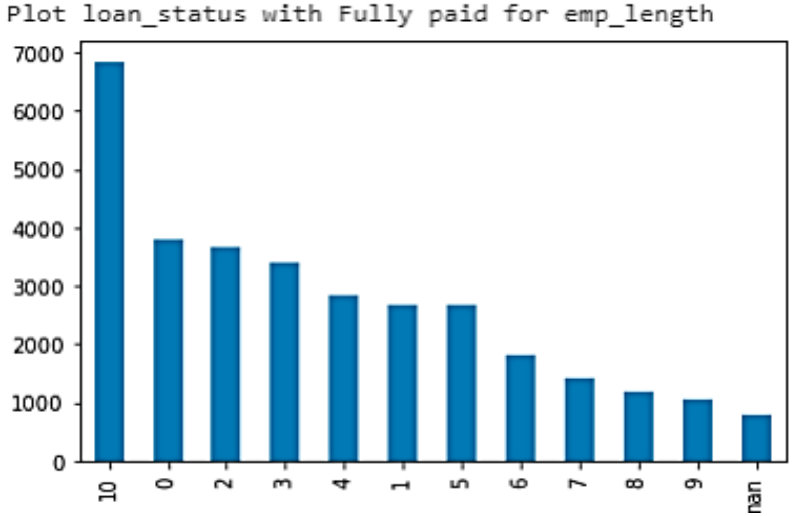
- NY has 3k/3.5k ~=85% successful repay
- CA has approx ~81% success repay

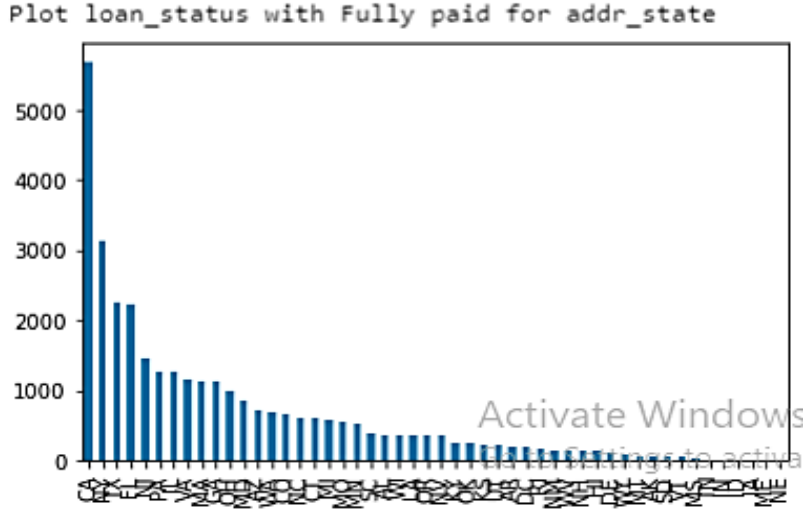
Identify pattern of loan default with employment length

- 10+ years of employment length are the largest consumer of the loans
- 0,2,3 years of employment length are having ~86% success
- 1,4,5 years of employment length have ~83% success









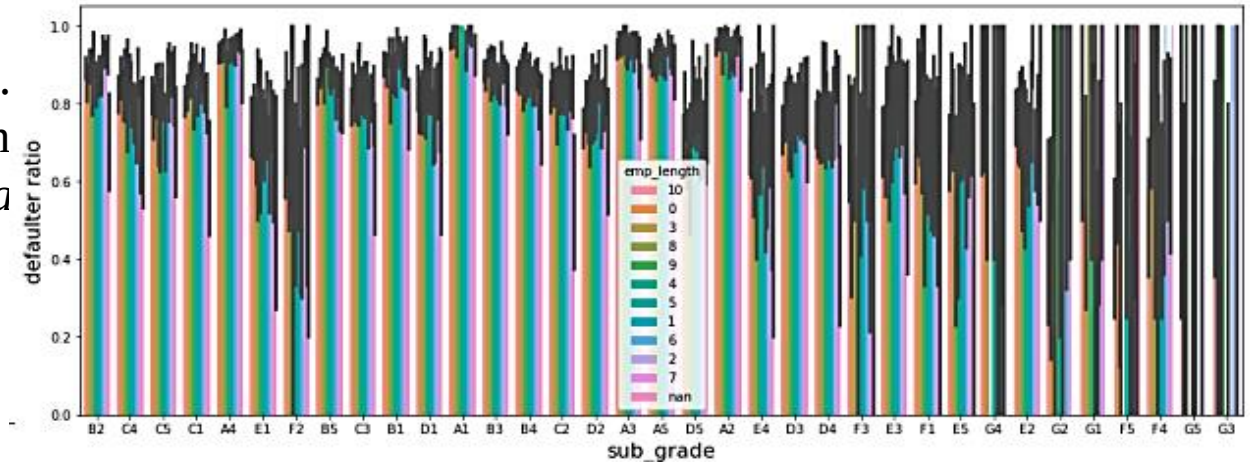
3 rd Part of the Analysis involves bivariate analysis involving two columns to get the impact on loan status

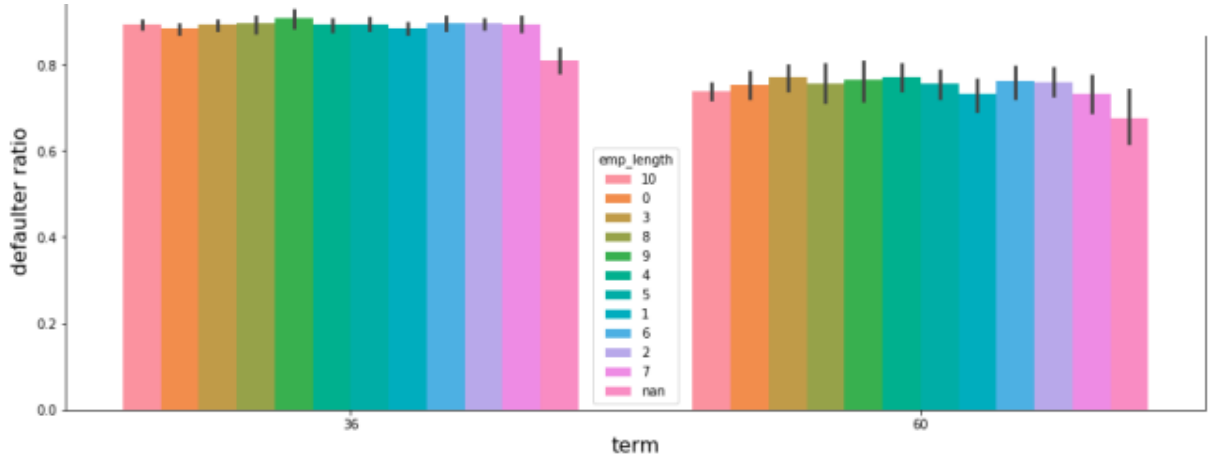
Employment length and Term has impact on loan status

- Not having nay significant impact on term and purpose.
- small_business is doing better with 36months term loan

User Grade and employment length has impact on loan sta

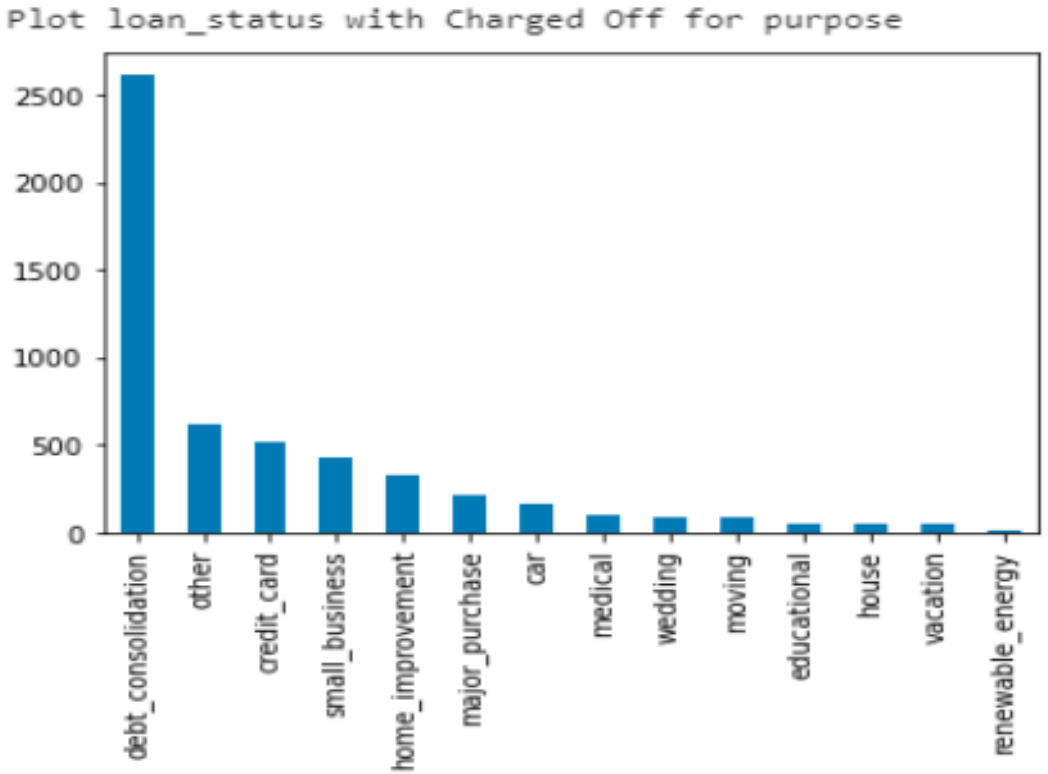
- G grade is not doing well on any emp length

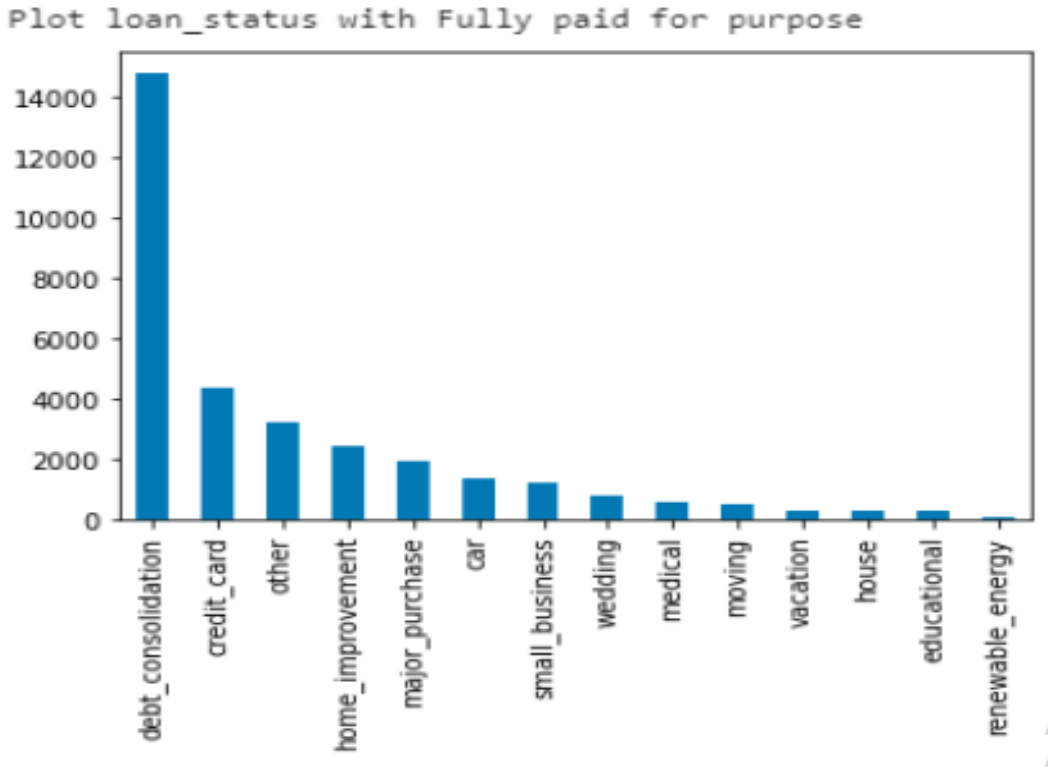




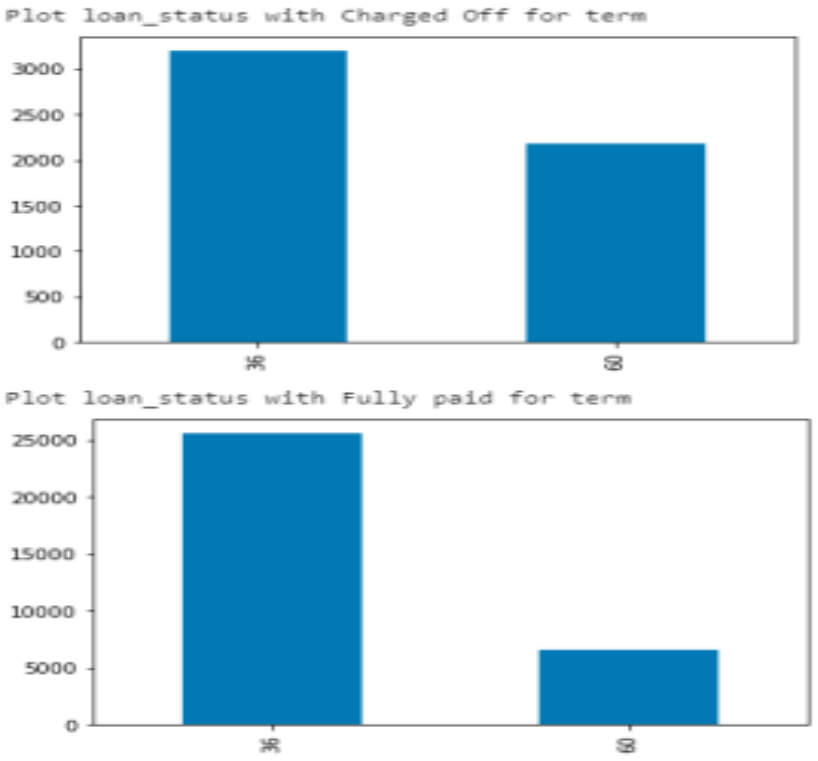
<Results>

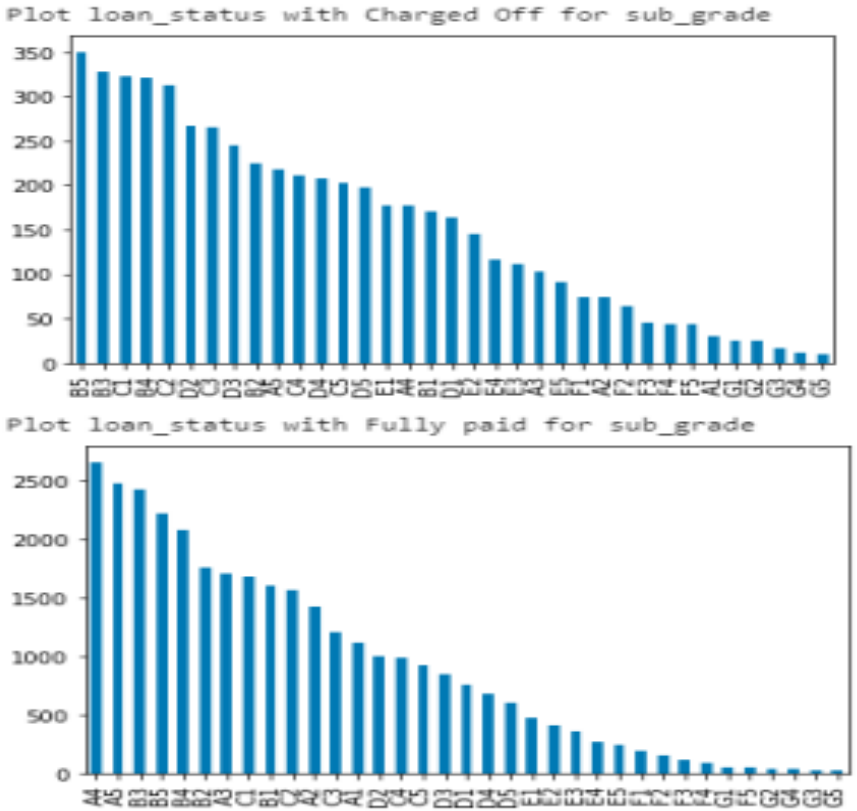
Plot 1: Histogram shows considering loan purpose which can be used to consider the loan application





Plot 2: Consideration of Member grades when processing loan application
Plot3: Considering loan term when processing application





<Conclusions>

- **Loan Purpose**
  - Fully Paid loan has good success rate for loan purpose for "debt_consolidation“
  - "credit_card" loan purpose has success rate of aprox 4000/(500+4000) ~= 86%
- **Loan Duraiton**
  - 36month seems to be the ideal duration to avoid loan defaulter
- **Address State**
  - On the large number of clients as members
  - NY has 3k/3.5k ~=85% successful repay
  - CA has approx ~81% success repay
- **Employment Length has significant impact on loan repayment**
  - 10+ years of employment length are the largest consumer of the loans
  - 0,2,3 years of employment length are having ~86% success
  - 1,4,5 years of employment length have ~83% success
- **Term and Purpose has no significant relation in loan repayment**
  - Not having nay significant impact on term and purpose, except for renewable_energy, educational loan
- **Purpose and loan duration has impact on the loan repayment**
  - small_business is doing better with 36months term loan

<Conclusions>

- **Term and Purpose has no significant relation in loan repayment**
  - Not having nay significant impact on term and purpose, except for renewable_energy, educational loan
- **Purpose and loan duration has impact on the loan repayment**
  - small_business is doing better with 36months term loan
- **Member grade with employment length has impact on the loan recovery**
  - G grade is not doing well on any emp length