LENDING CLUB CASE STUDY

GROUP MEMBERS:

SREE P

P. ARUNA DEVI

ABSTRACT

- This assignment will give us an idea about how real business problems are solved using EDA.
- In this case study, apart from applying the techniques you have learnt in EDA.
- Also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

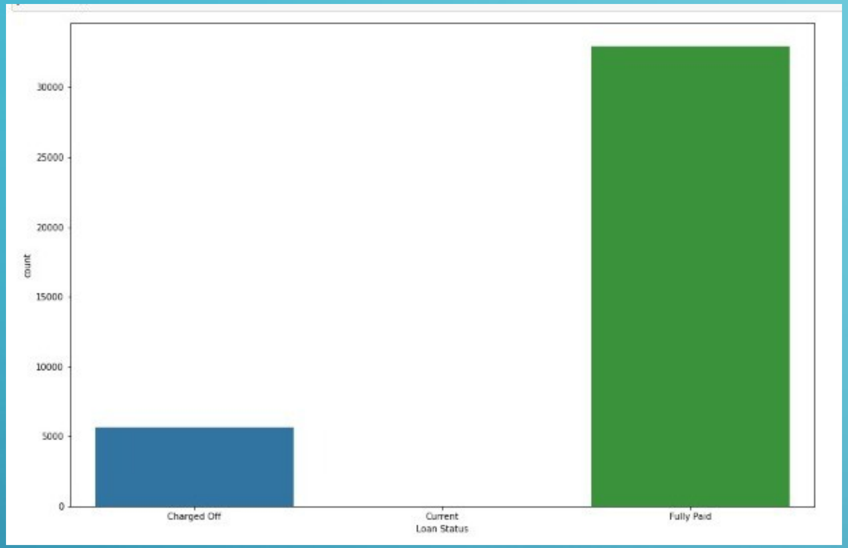
BUSINESS OBJECTIVES

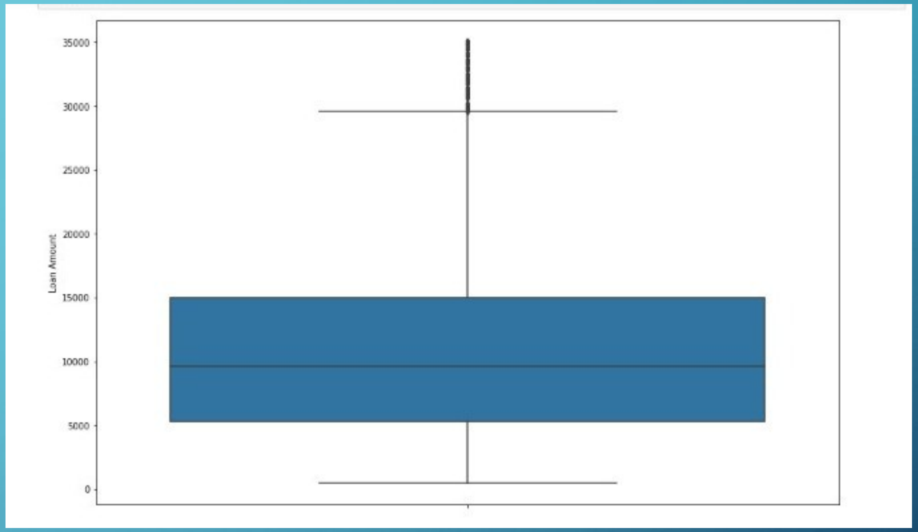
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- Identification of such applicants using EDA is the aim of this case study.
- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- The company can utilise this knowledge for its portfolio and risk assessment.

PROBLEM SOLVING METHODOLOGY

- Undertand the data dictionary.
  - Data cleaning or Data pre processing should be done.
  - If null values present over a certain threshold value then need to remove it.
  - Do Univariate analysis on the data.
  - Next is to do bivariate analysis
  - If needed perform multivariate analysis.

ANALYSIS

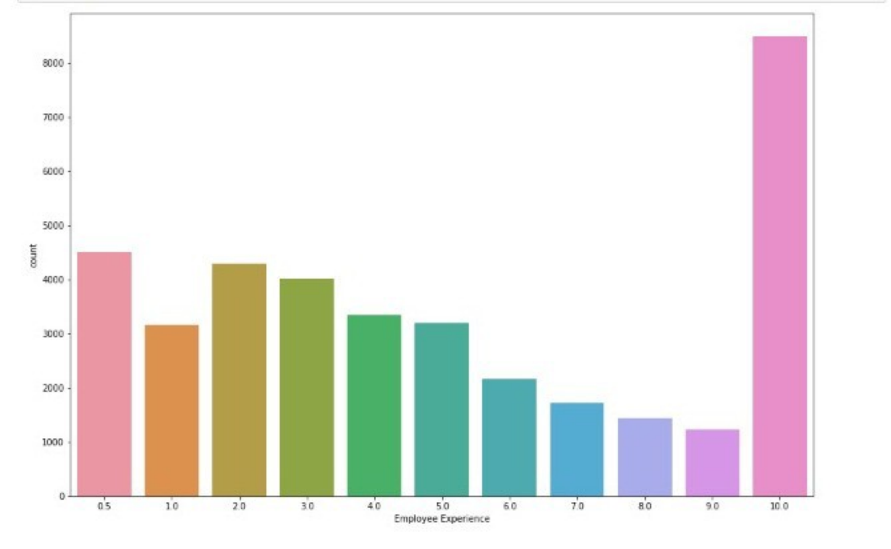


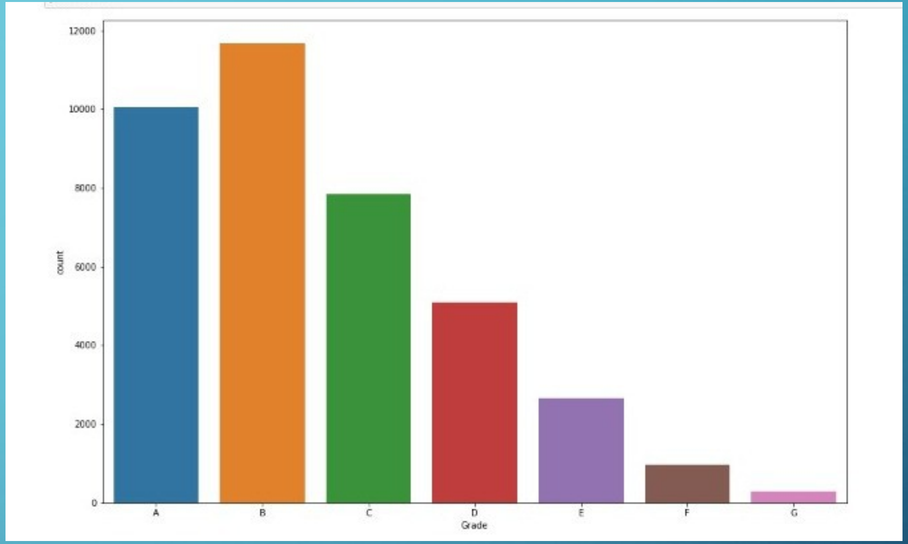


- Loan status insights:
  - From loan status we get to know that 5627 loans have been chargedoff.
  - Nearly 14.5 percent loans are charge off excluding current loans

- Loan amount insights:
  - Loan amount varies from 500 to 35000.
  - 10,000 is the most given loan amount it is given for 2248

ANALYSIS

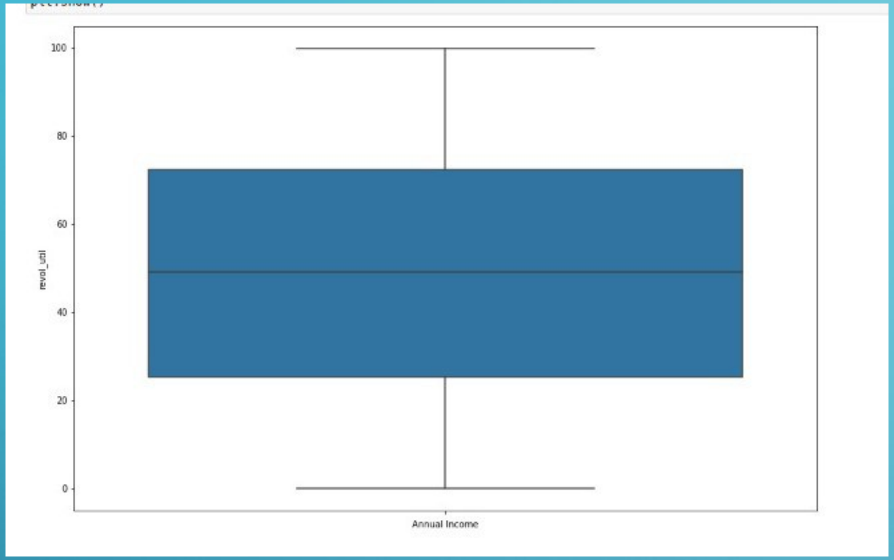


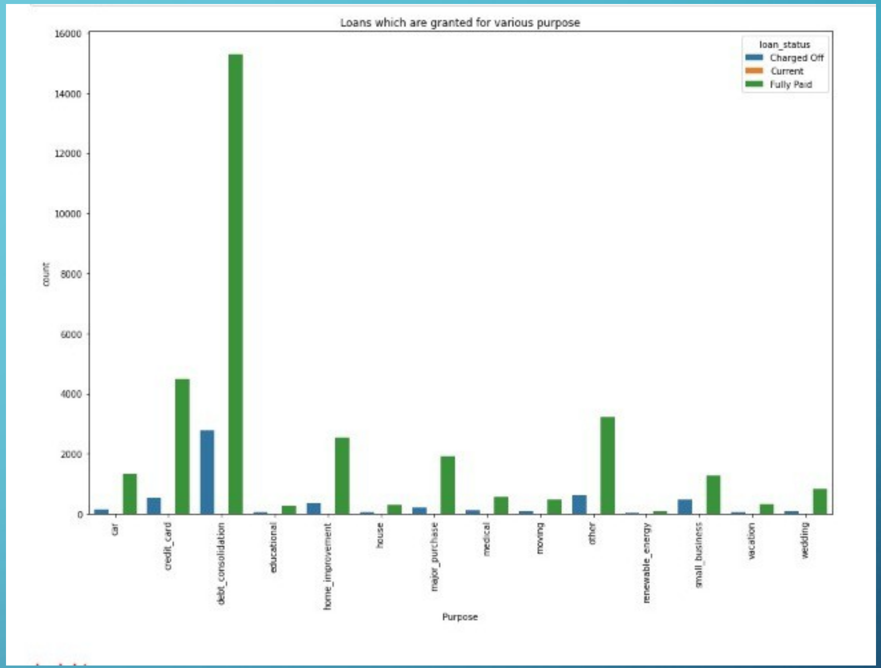


- Employee length insights:
  - Majority of the people who have taken loan have more that 10 years of work experience.
  - Minority of the people who have taken loan have 9 years of work experience.

- Grade insights:
  - Majority of the loans have grades A & B i.e 10045 and 11675 respectively.
  - Minority of the loans have grades G i.e 299.
  - Therefore most of the loans are graded high.

ANALYSIS

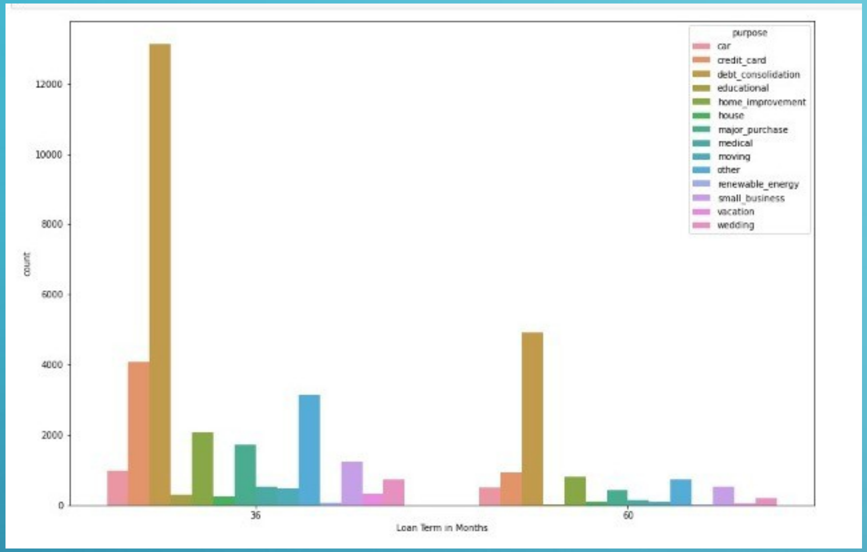


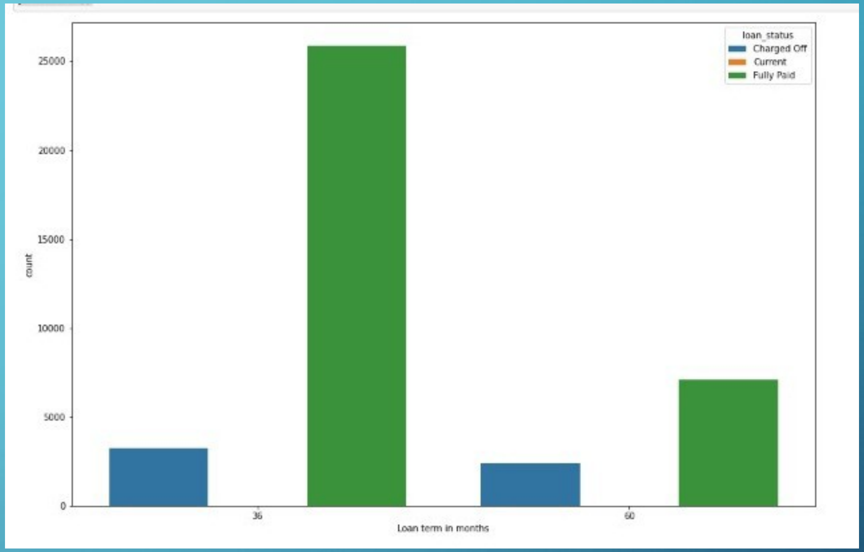


- Annual inc insights:
  - There are only few people above income 20L which are outliers

- Loans which are granted for various purpose:
  - From the above analysis it is evidend that the loans taken for small_business, renewable_energy and educational are the riskier ones.

ANALYSIS

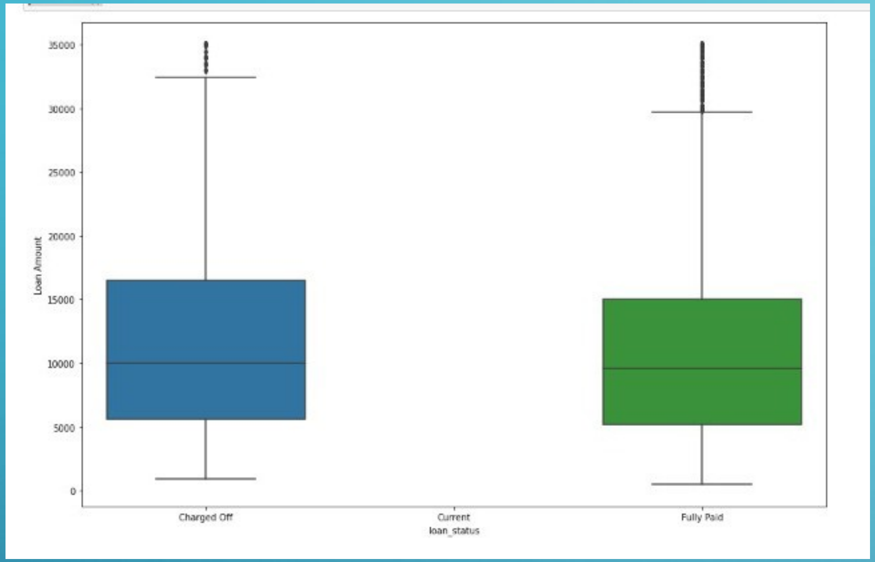




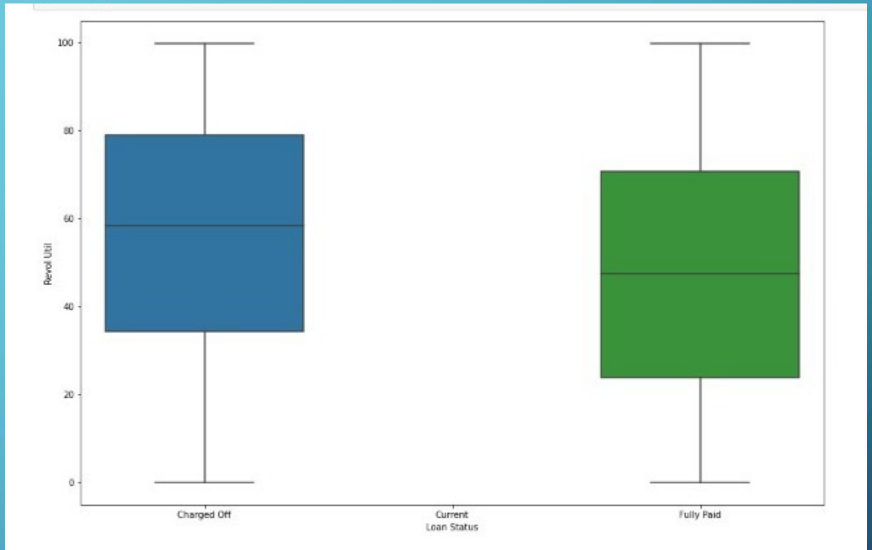
- Loan term in months and its purpose:
  - For the purpose of debt consolidation people get more loans and and term is 36 months

- Loan term and its respective loan status:
  - Percentage of defaulters in the term of 36 months is higher than 60 months

ANALYSIS



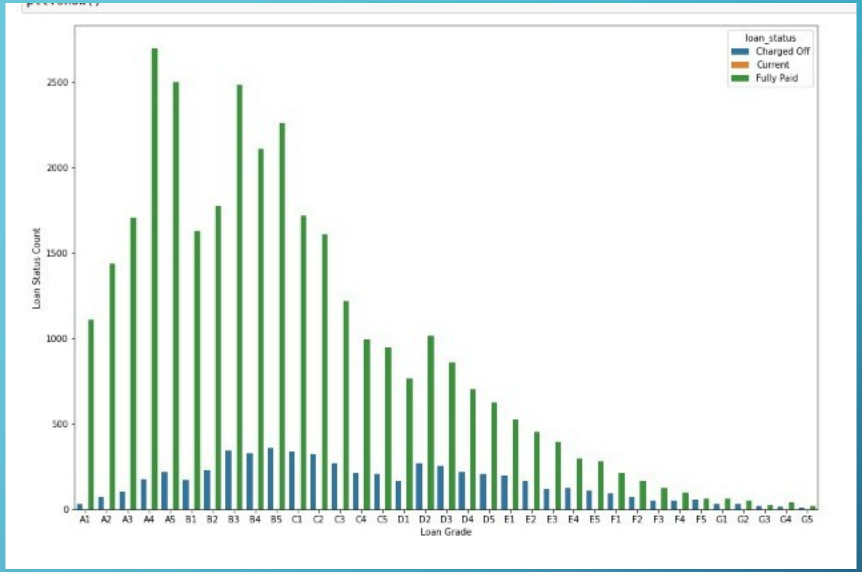
- Loan amount and its status:
  - Mean of both Charged off and Fully Paid are same.
  - 25% and 75% Loan amount of Fully paid and charged off are nearly equal

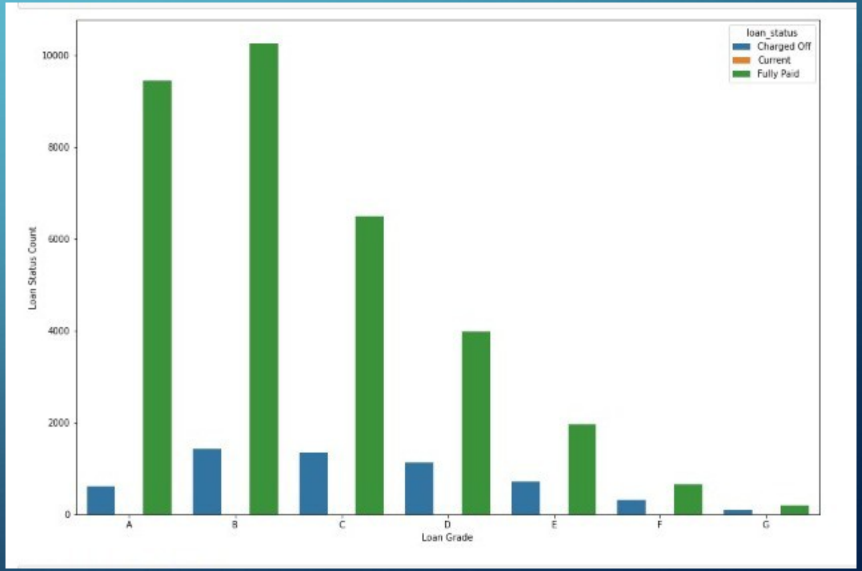


- Revolving Line Utilisation rate:
  - Higher revolving line utilisation rate above 58% increase the risk of defaulting the loan.

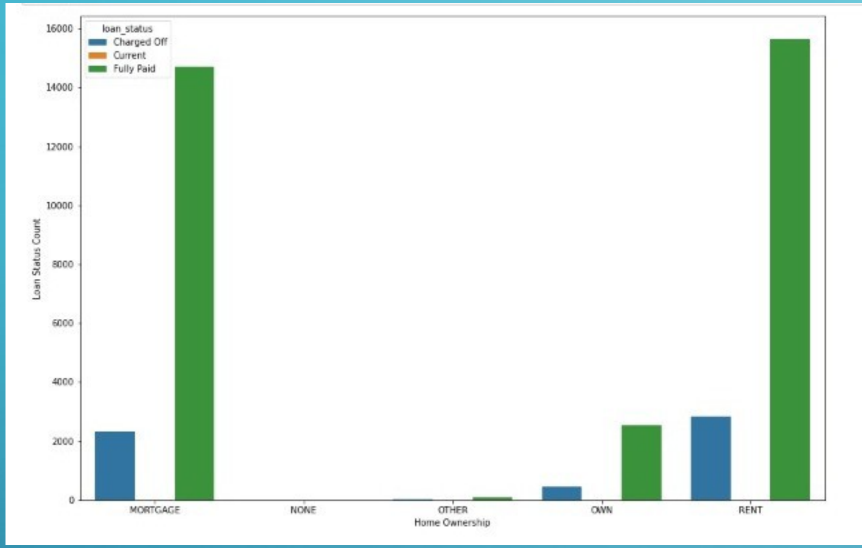
ANALYSIS

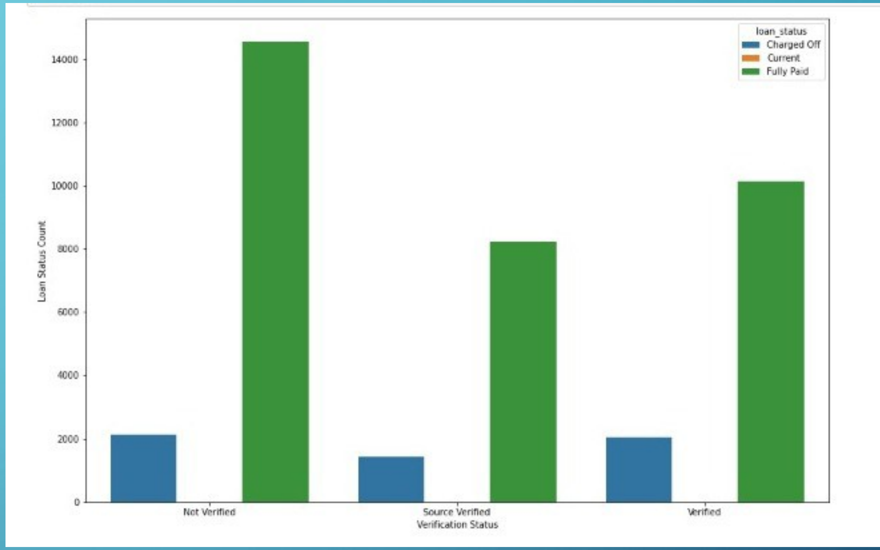
- Loan grade and it’s risk:
  - B,C D loan have more charged off
  - Loan grade and sub-grade from D to G have higher risk percentile





ANALYSIS

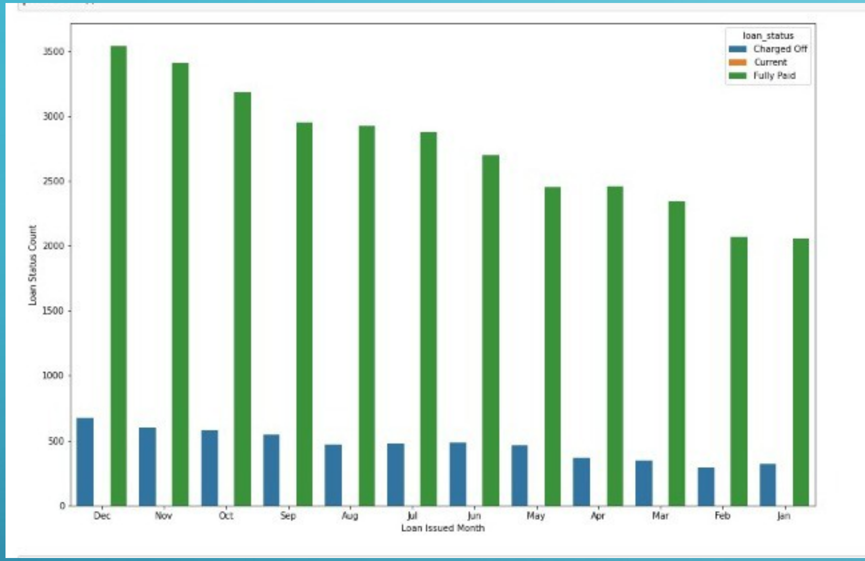


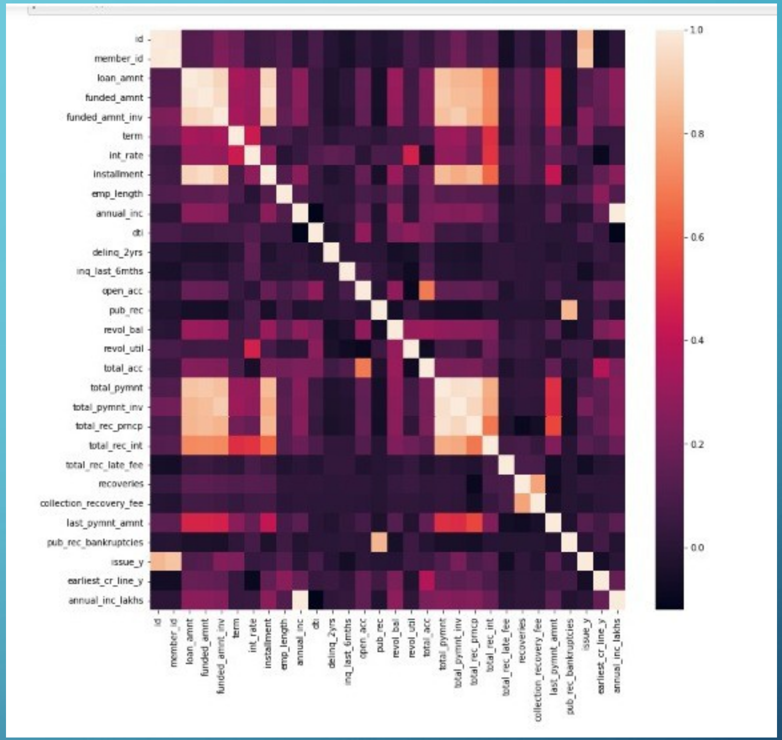


- Home ownership as driving factor:
  - People who are in the rented house are more likely to default the loan which is a risk factor

- Verification status and its role:
  - From the above graph we get to know that both verified ad non verified peoples loan are equally defaulted, from this we get to that we need to look into the verification process.

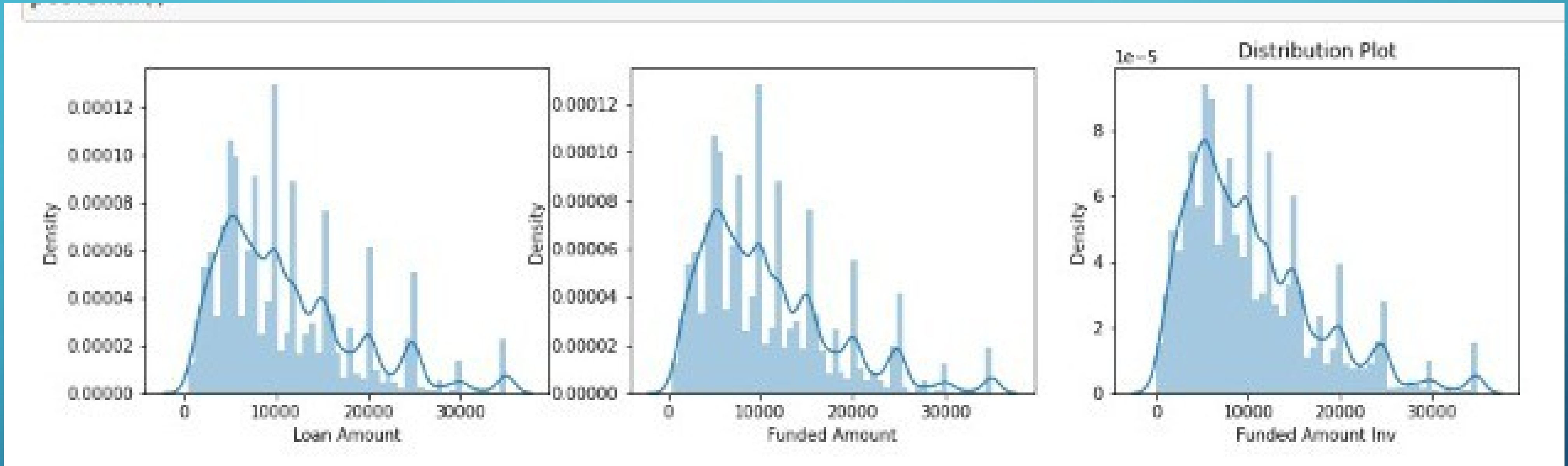
ANALYSIS





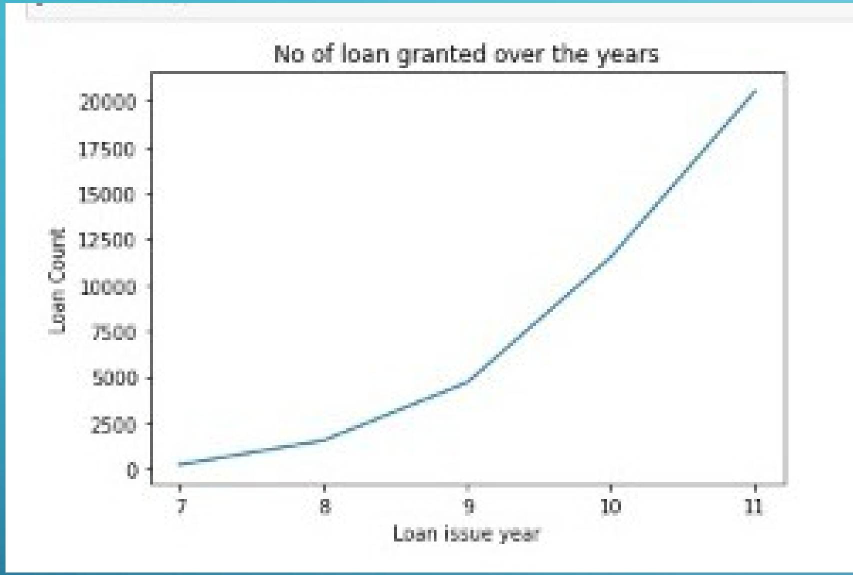
- Loan issued month and its role:
  - Loan issued in the month of DEC, SEP, MAY are more risker
- Bivariate Analysis

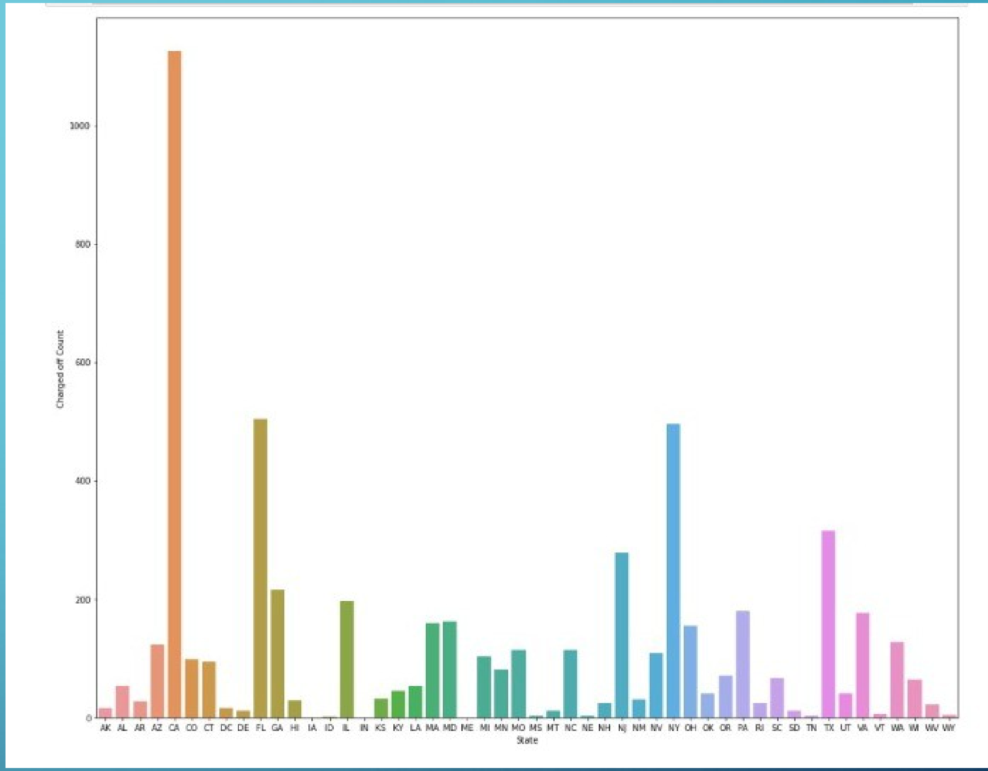
ANALYSIS



- Distribution Plot

ANALYSIS





- No of loans granted over the years:
  - From the above graph we get to known that loan given over the years have been increased

- Impact of Resident state on Loan:
  - People from CA are not paying the loan back , Charge off loan is high in that state

CONCLUSION

- From the EDA of the given loan data set we got to known that multiple factors depend on the driving factors.
- However, according to me some of the impactful factors are:
- Lesser Risk:
  - Applicant’s address state FL, NY, TX have some lesser amount of impact of defaulting loan.
  - Higher loan amount increases the risk.
  - From the above graph we get to know that both verified ad non verified peoples loan are equally defaulted, from this we get to that we need to look into the verification process.
  - Loan issue month int the Month Dec, May, Sep.

- Higher Risk:
  - Loan grade and sub-grade from D to G.
  - Loan purpose increases the risk some purposes are small business, renewable energy, educational.
  - Percentage of defaulters in the term of 36 months is higher than 60 months.
  - Higher interest rate above 13%.
  - People who are in the rented house are more likely to default the loan which is a risk factor.
  - Higher revolving line utilization rate above 58% increase the risk of defaulting the loan.
  - High loan amount and the high interest rate for lower income group.
  - More number of installment and longer repayment term.
  - People who are in the rented house are more likely to default the loan which is a risk factor.

THANK YOU