Lending Club Case Study

H.Yogajayalakshmi
MeenaSaranya

Problem Statement

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile

Risk Involved

- Two types of risks are associated with the bank's decision:
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss for business
  - If the applicant is not likely to repay the loan, then approving the loan may lead to a financial loss

Benefits

- How EDA is used in this business problems.
- It also develops a basic understanding of risk involve
- How the data is used to mitigate loss
- Data visualization and charts involved to provide better insights

Objective - To implement EDA technique on this problem and understand the insights to suggest business to improve their business

Steps involved in Analysis

Data Cleaning

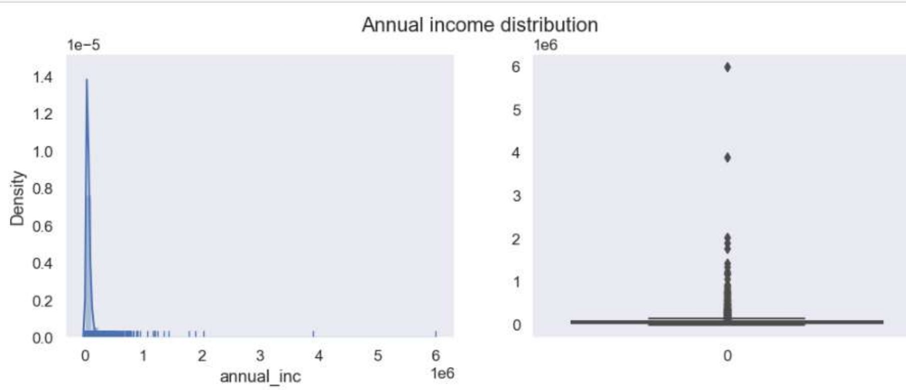
- Load the input data
- Remove the null or missing values in entire column
- Remove irrelevant columns
- Fill the missing value (NA,0)
- Remove Duplicate value/Retain the unique values
- Remove Outliers

Data Visualization

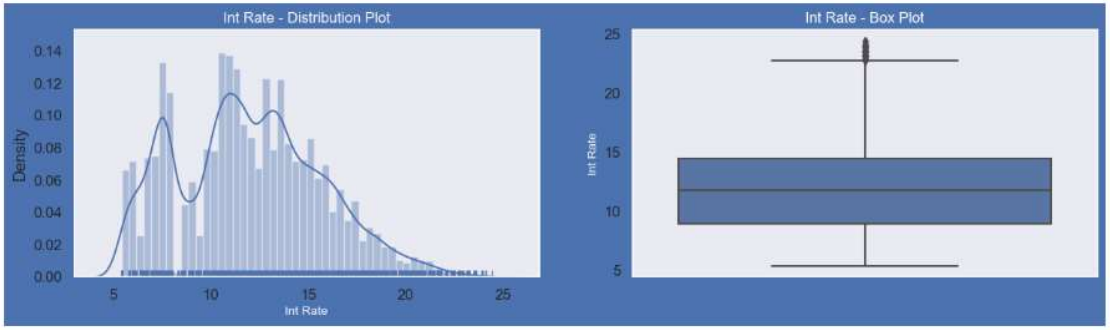
- Data Understanding
- Correlate the business Domain
- Univariate Analysis – Different single variable
- Bivariate Analysis – Involving two or more different variables
- Correlation Matrix - Heatmap

Insights & Recommendations to Business should be derived at end of this process

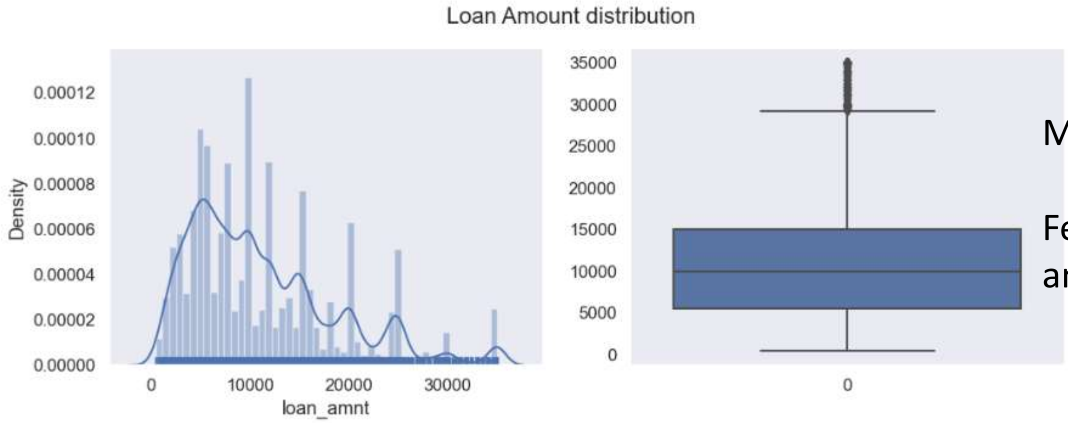
Univariate Analysis



More than 95% of people are having incomes below 235000.



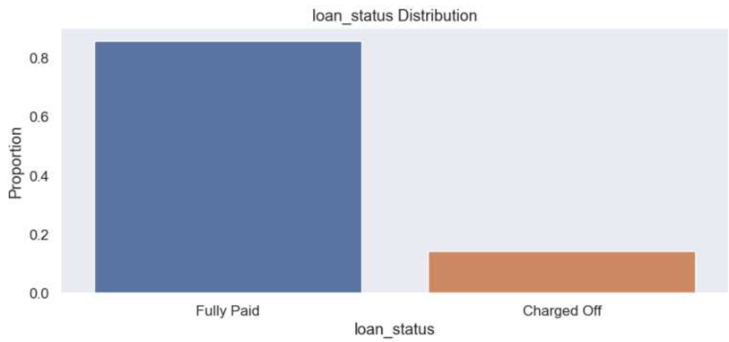
Most of interest rates lies between 9% to 14.5%.
Few customers took loan at higher rates of interest = 22.5%



More people took loan amount of 10000 (median = 10000)

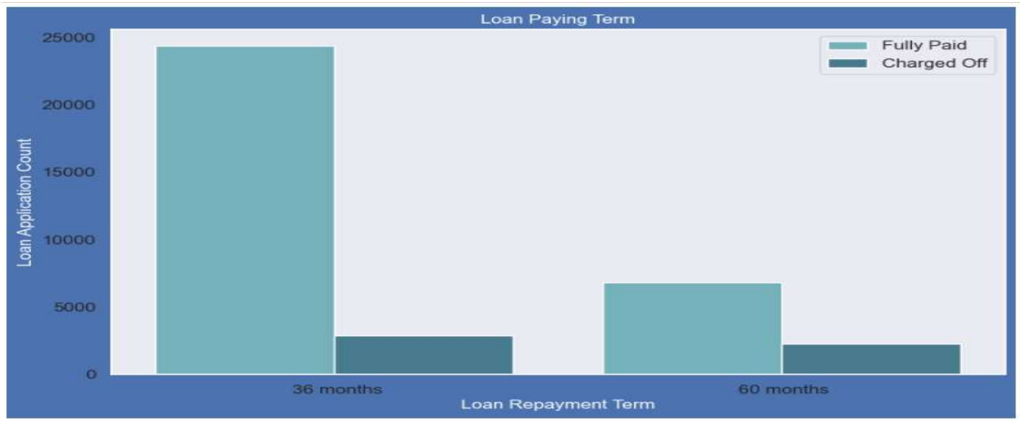
Few people were in range of more than 30000 loan amount.

Univariate Analysis

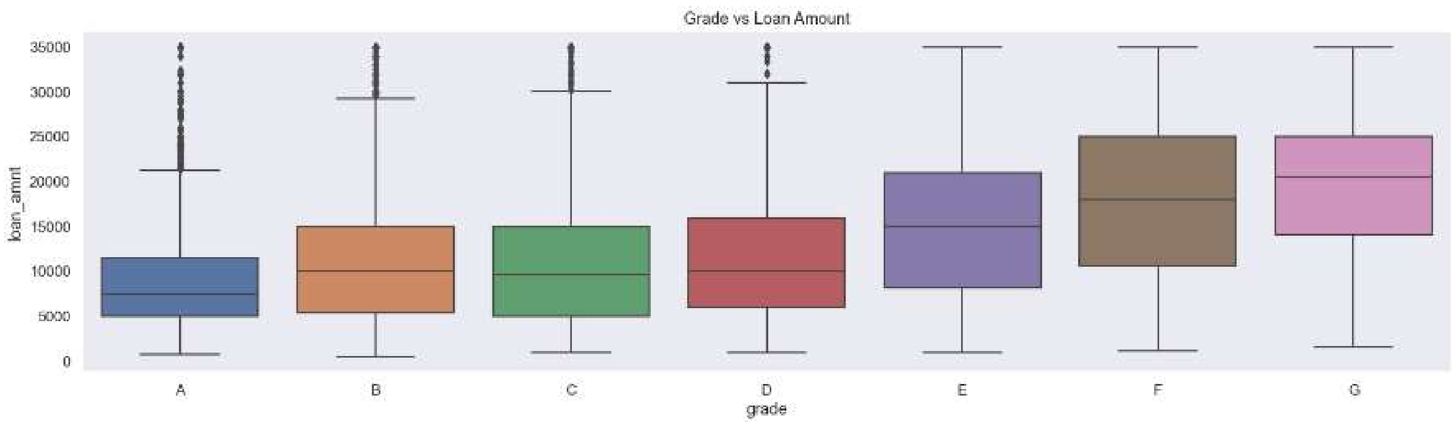


~85% of borrowers has paid the loan fully, 14% were defaulted/Charged off the loan.

Ordered Categorial Variable



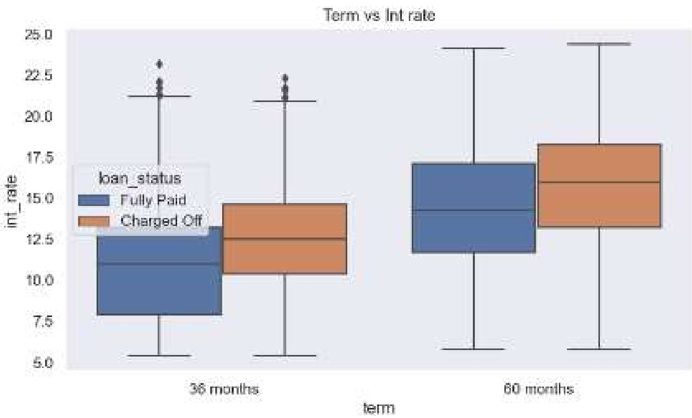
Term of 36 months had more number of Charged Off applicants as compared to applicants who had term of 60 months

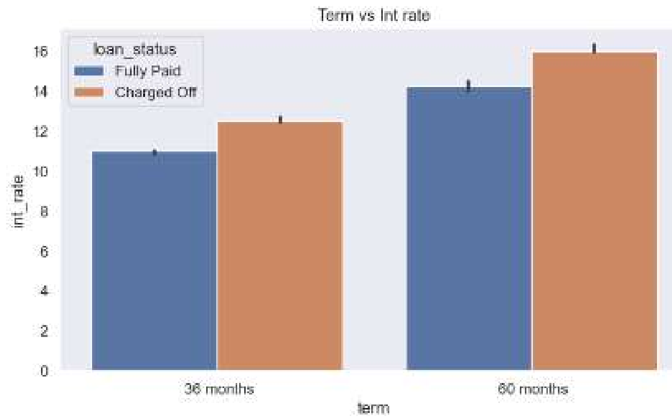


Grade 'F' and 'G' have taken max loan amount. As Grades were growing the loan amount was increasing

Bivariate Analysis

Terms vs Interest rate

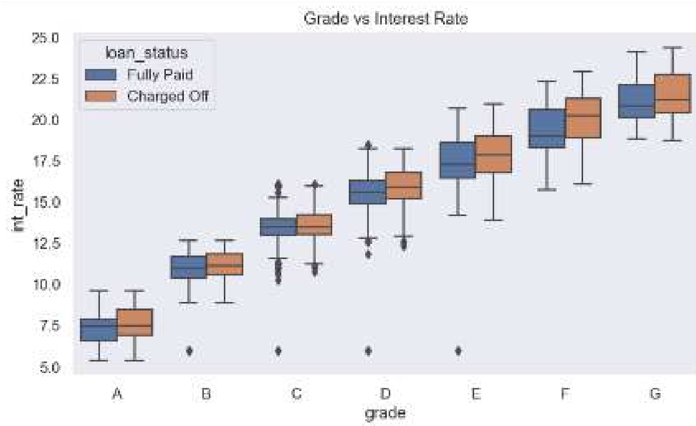


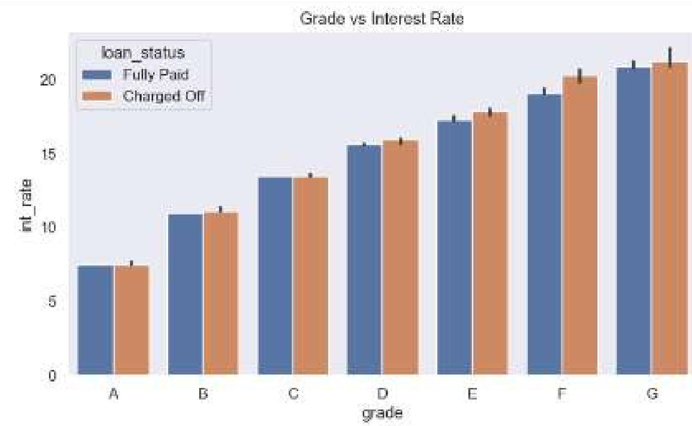


Higher interest rates the charged off was higher in both 36 and 60 months tenure.

Grade vs Interest rate

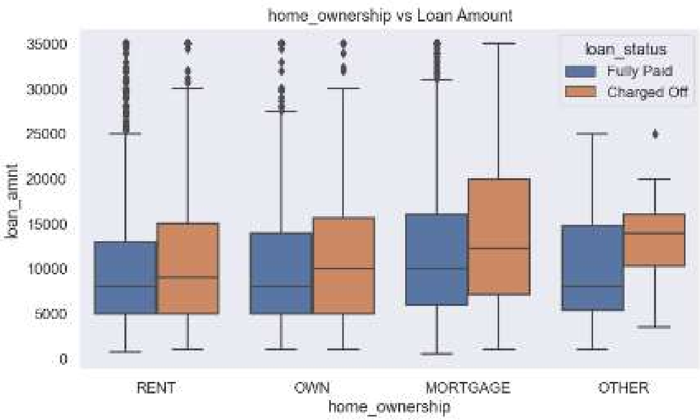
As grade decreases the interest rate increases and prone to default the loan.

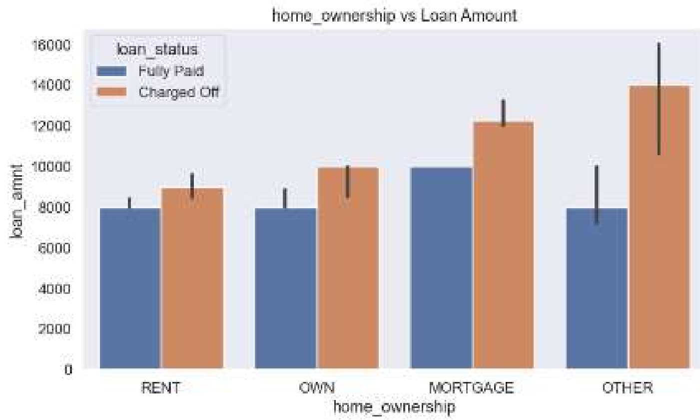




Bivariate Analysis

Home Ownership vs Loan amount

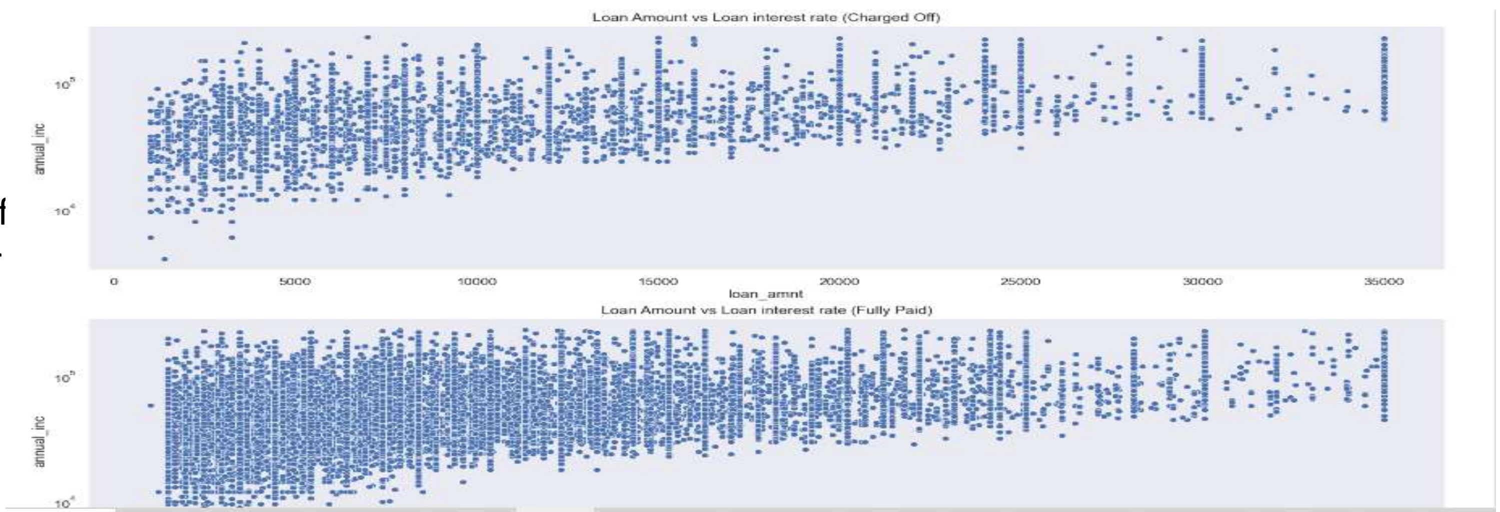




Customers who took higher loan amounts defaulted more than others.

Loan Amount Vs Interest Rate

Fully paid and Charged Of loans were having similar pattern.



Bivariate Analysis

Correlation Matrix

1. Loan amount, funded amount investor, funding amount are strongly correlated.
2.Annual income with DTI(Debt-to-income ratio) is negatively correlated

Recommendations

Factors driving the analysis

1. DTI
2. Grades
3. Annual income
4. Loan Status
5. HomeOwnership status

Observations – Key Points

Customer likely to default

1. Customers were having annual income in the range 40000-90000.
2. Customers who were having Public Recorded Bankruptcy.
3. Least grades like E,F,G which indicates high risk.
4. Applicants with high Debt to Income value.
5. Employee with working experience 10+ years

Thank you