LENDING CLUB CASE STUDY

Ruchir Dhingra
Tanupriya Chaurasia

OVERVIEW

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Loans can be taken for number of purposes like home loans, car loan, medical emergency loan, etc.

BUSINESS OBJECTIVE

The aim is to identify 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.

If we can identify borrowers who **default** cause the largest amount of loss to the lenders i.e. 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.

STEPS TAKEN

- Data Understanding
- Data Cleaning
- Data Analysis
- Inference

DATA UNDERSTANDING

- Types of variables present
- Classification on basis of – applicant's behaviour, loan information, statistics
- Classify each variable on data type – categorical, numerical
- Description about actual dataset given

DATA CLEANING

- **Dropping unnecessary Columns**
- **Dropping / Filling NULL values**
- **Separating Year and Month from date columns**
- **Detecting Outliers**
- **Removing Outliers**
- **Converting object variables into numeric or vice-versa as per requirements**

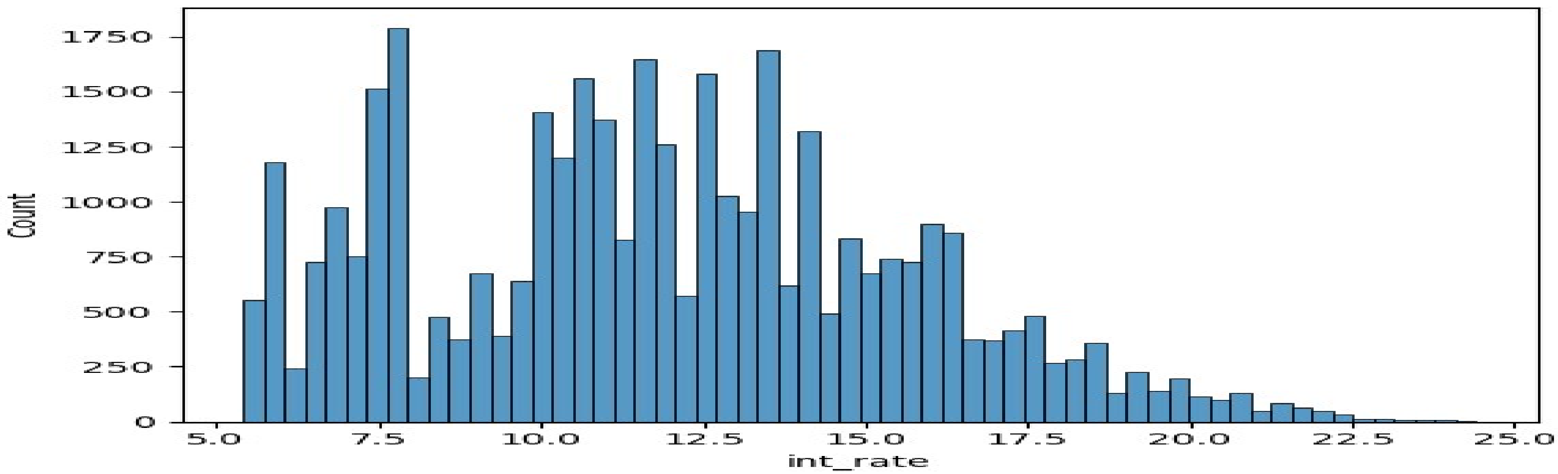
  - Total rows being 39717 and 111 columns
  - 54 columns deleted which have all null values
  - Dropping Behavior based columns
  - Segregating columns into Continuous, Categorical & Extra Columns
  - Dropping Behaviour based columns

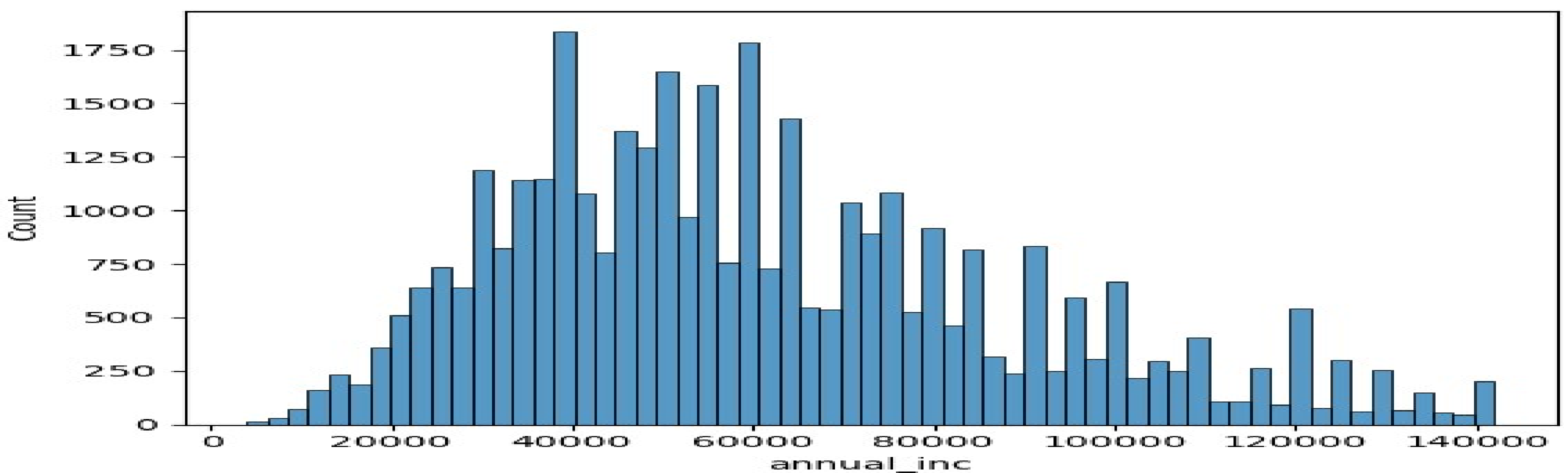
DATA ANALYSIS

- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis

Univariate Analysis

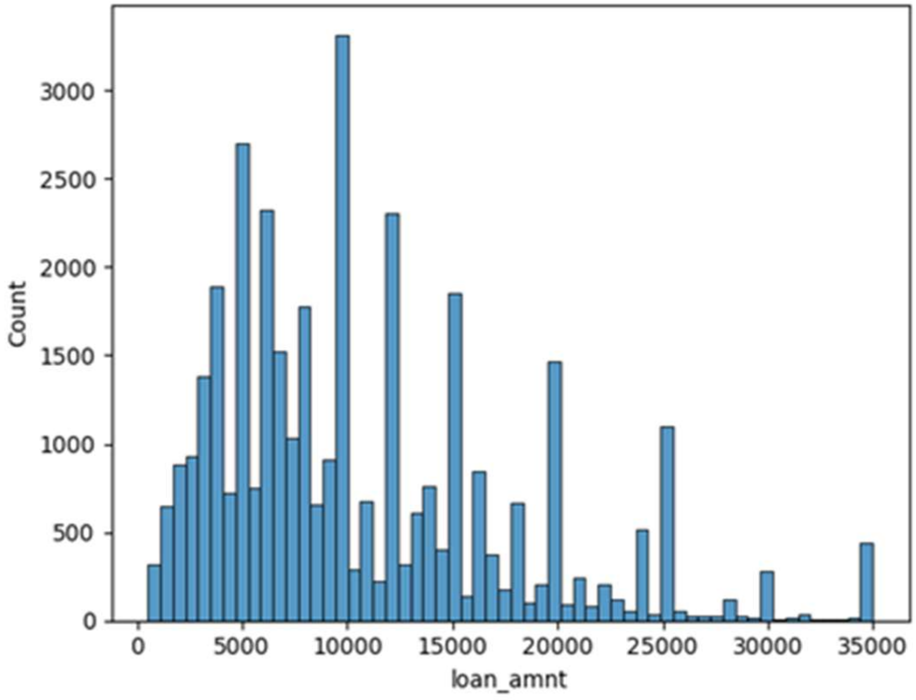
- Now, let us do analysis on each univariate continuous variable i.e, loan amount, funded amount, funded amount investment, interest rate, installment, annual income.

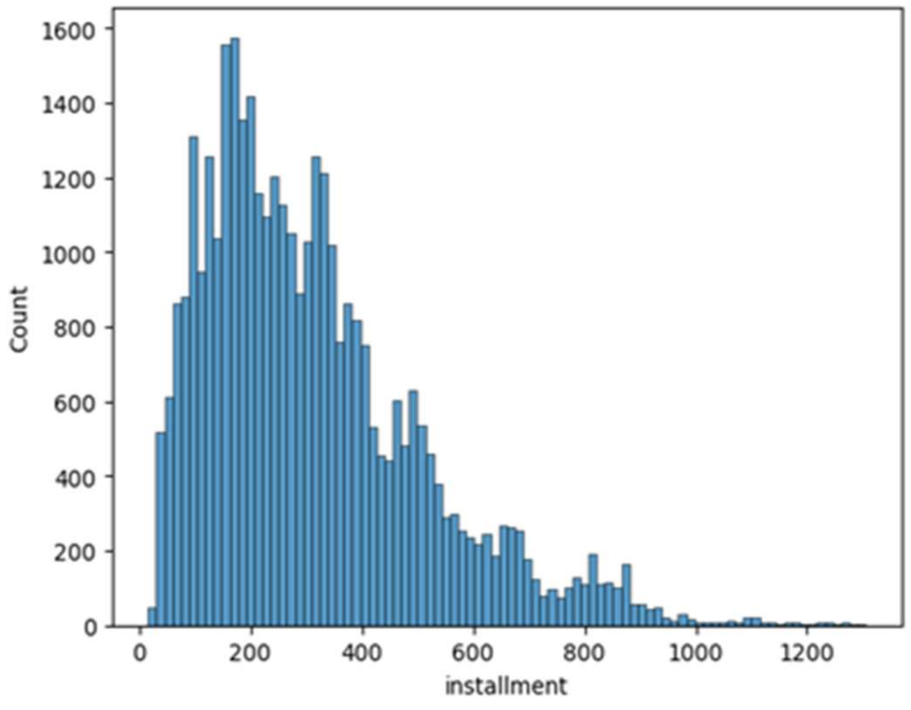




Univariate Analysis

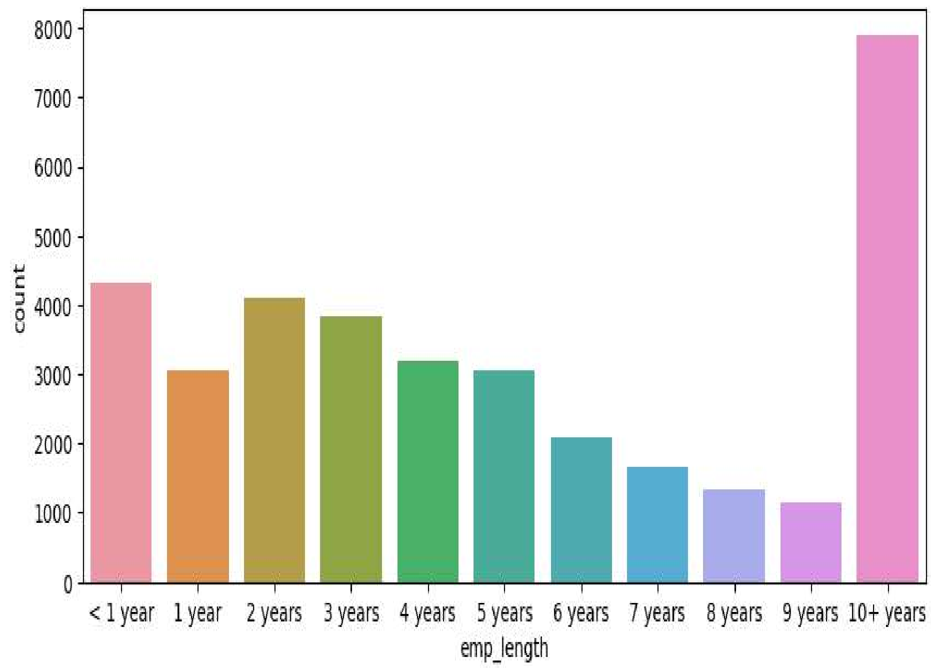
- **Loan amount has spikes in multiples of 5000, i.e., people prefer taking loan in multiple of 5000's**
- **Majority of applicants had around 100-250 rs as monthly installment**
- **Majority of applicants had annual income ranging from 40000-65000 installments.**

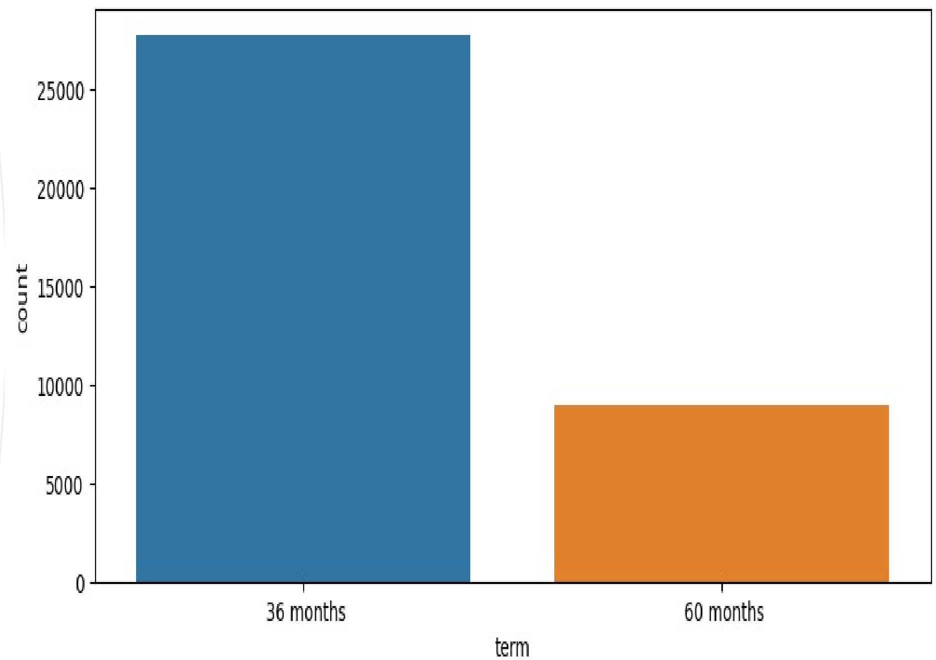




Univariate Analysis

Univariate Analysis on categorical columns i.e, term, grade, subgrade, loan status, purpose, home ownership, employment length, and verification status.

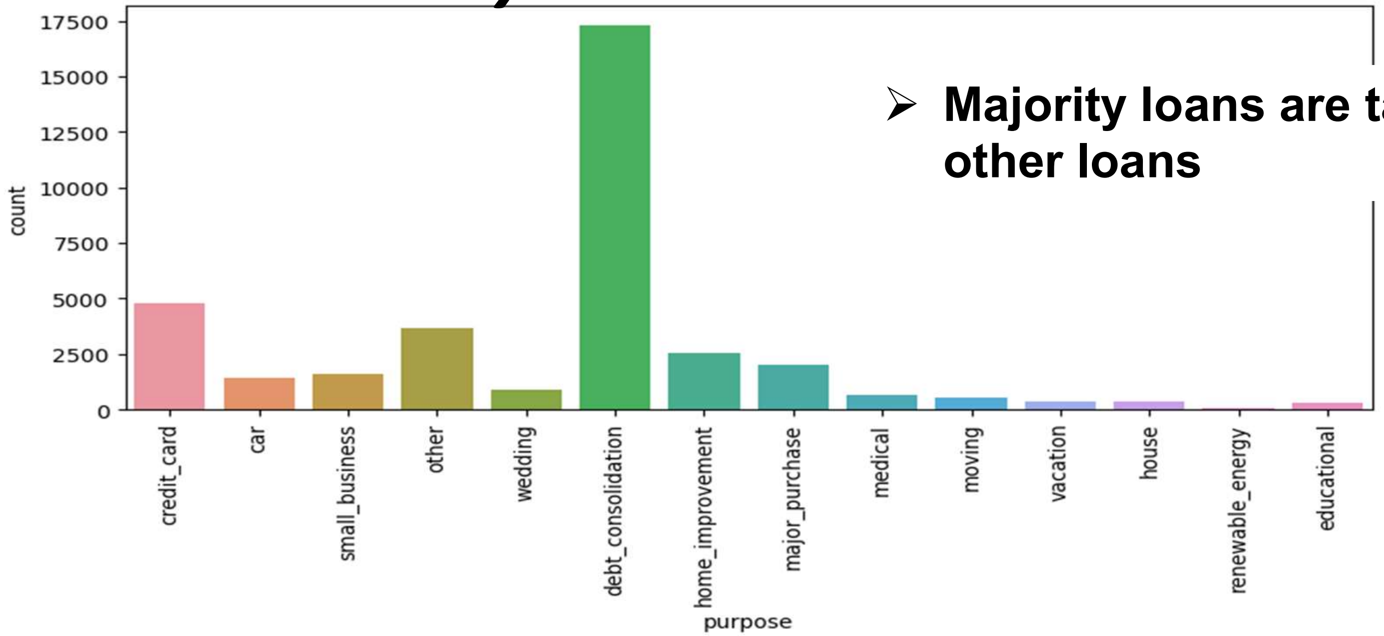




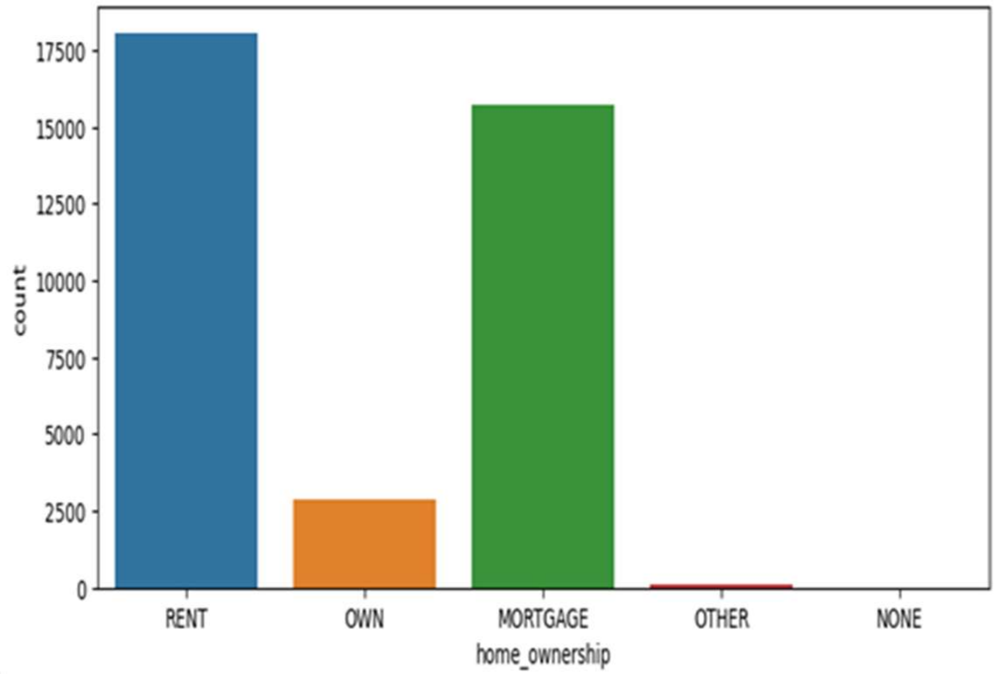
- **People employed recently ie in starting 0-5 years tend to take more loans than people with mature length in jobs**

- **Around 75% of applicants preferred to pay back the loan in 36 months**

Univariate Analysis



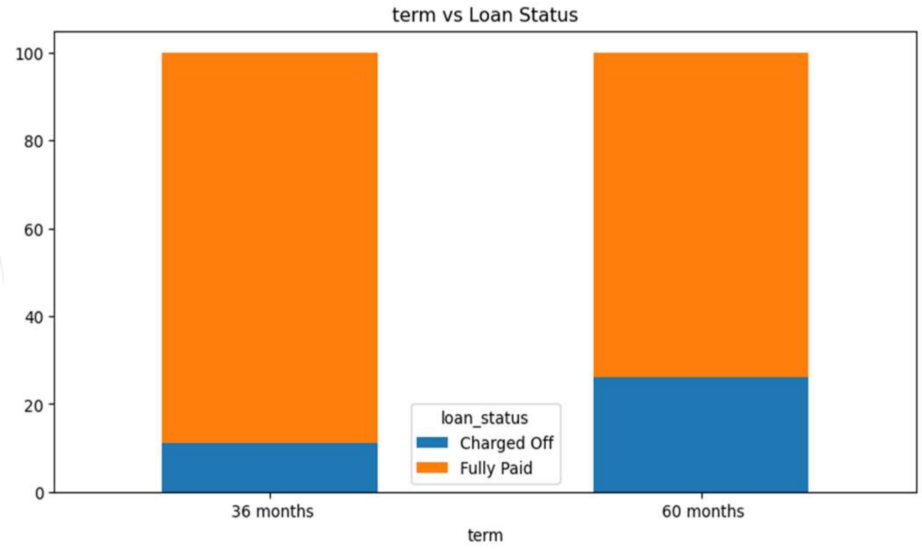
- **Majority loans are taken to pay off other loans**

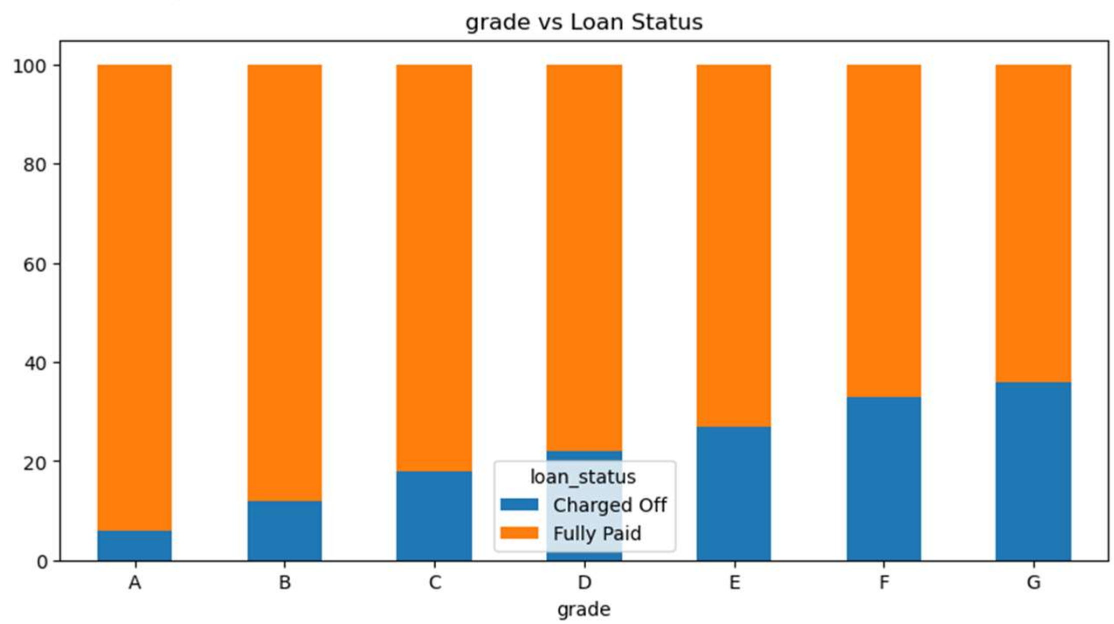


- **People having their own house are less likely to take loans**

Segmented Univariate Analysis

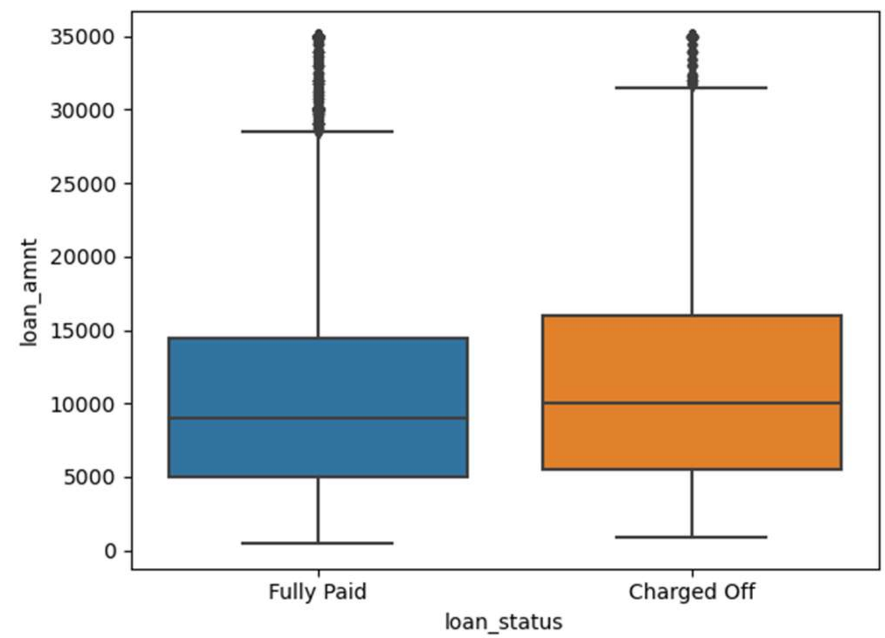
- **People having 60 months as loan payment tenure tend to default more**
- **People who default usually have a higher amount of loans in comparison to the ones who pays off**

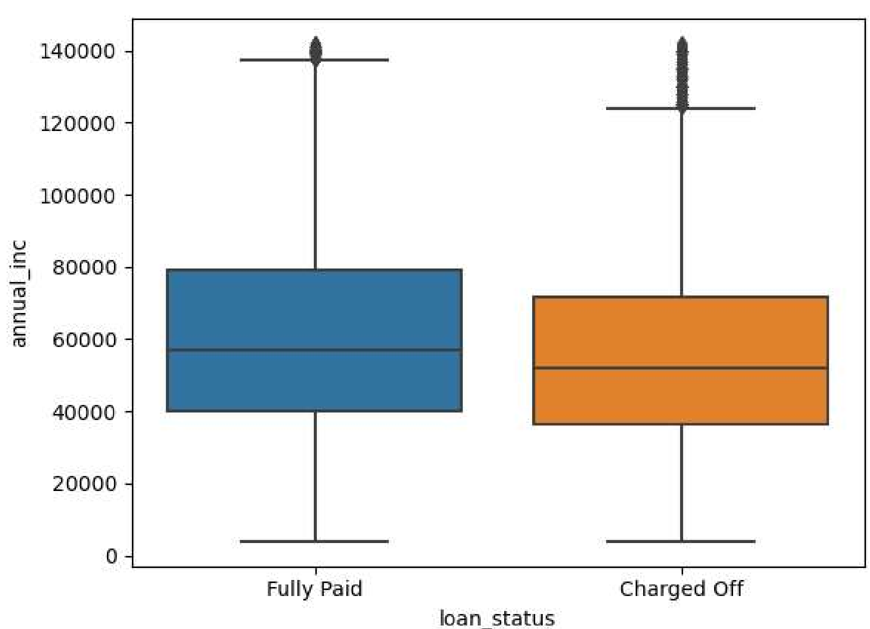


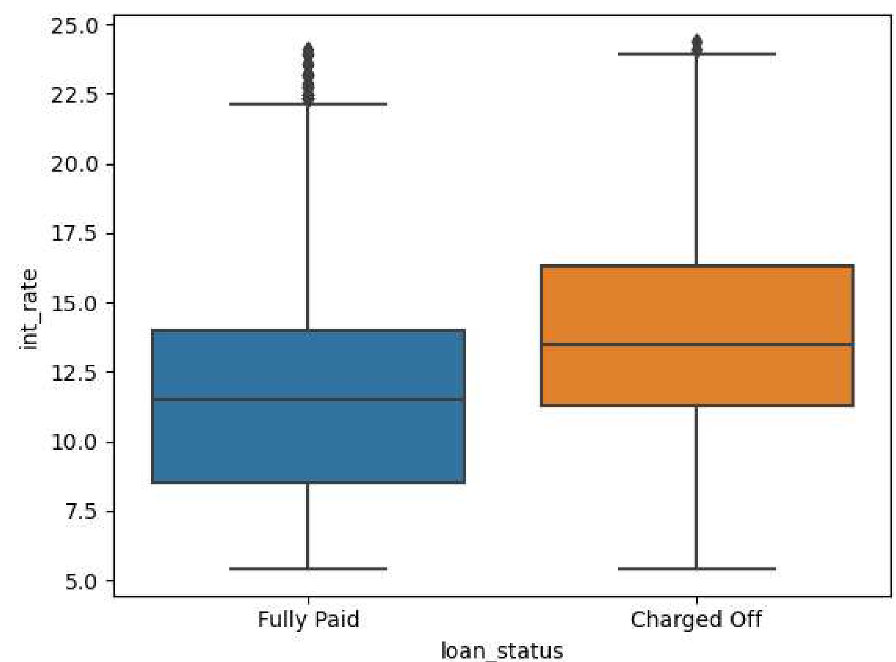


Segmented Univariate Analysis

- **People who default usually have higher amount of loans in comparison to the one's who pays off**
- **People having higher rate of interest tend to default on loan**
- **People having higher annual income tend to fully pay the loan**

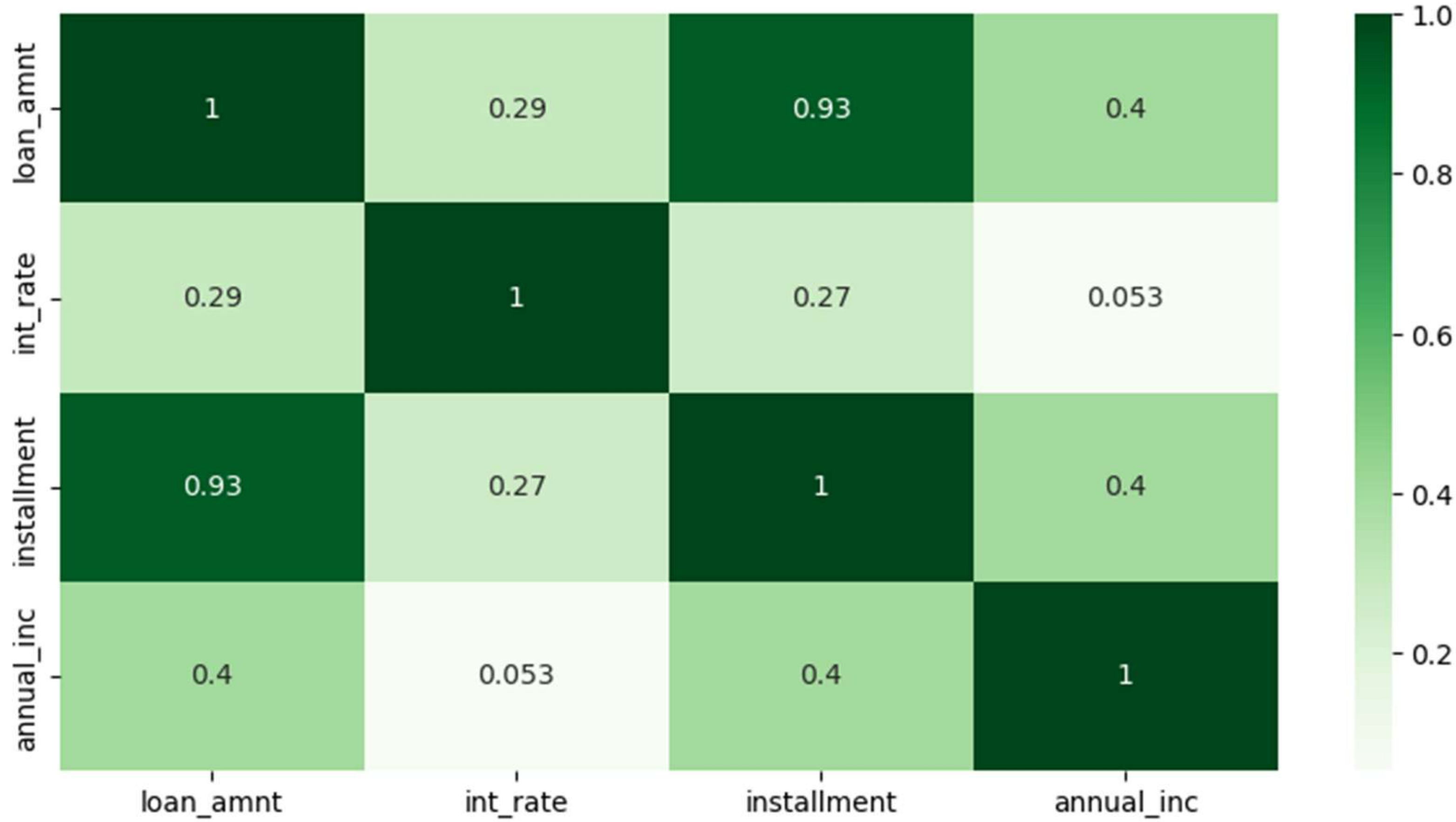






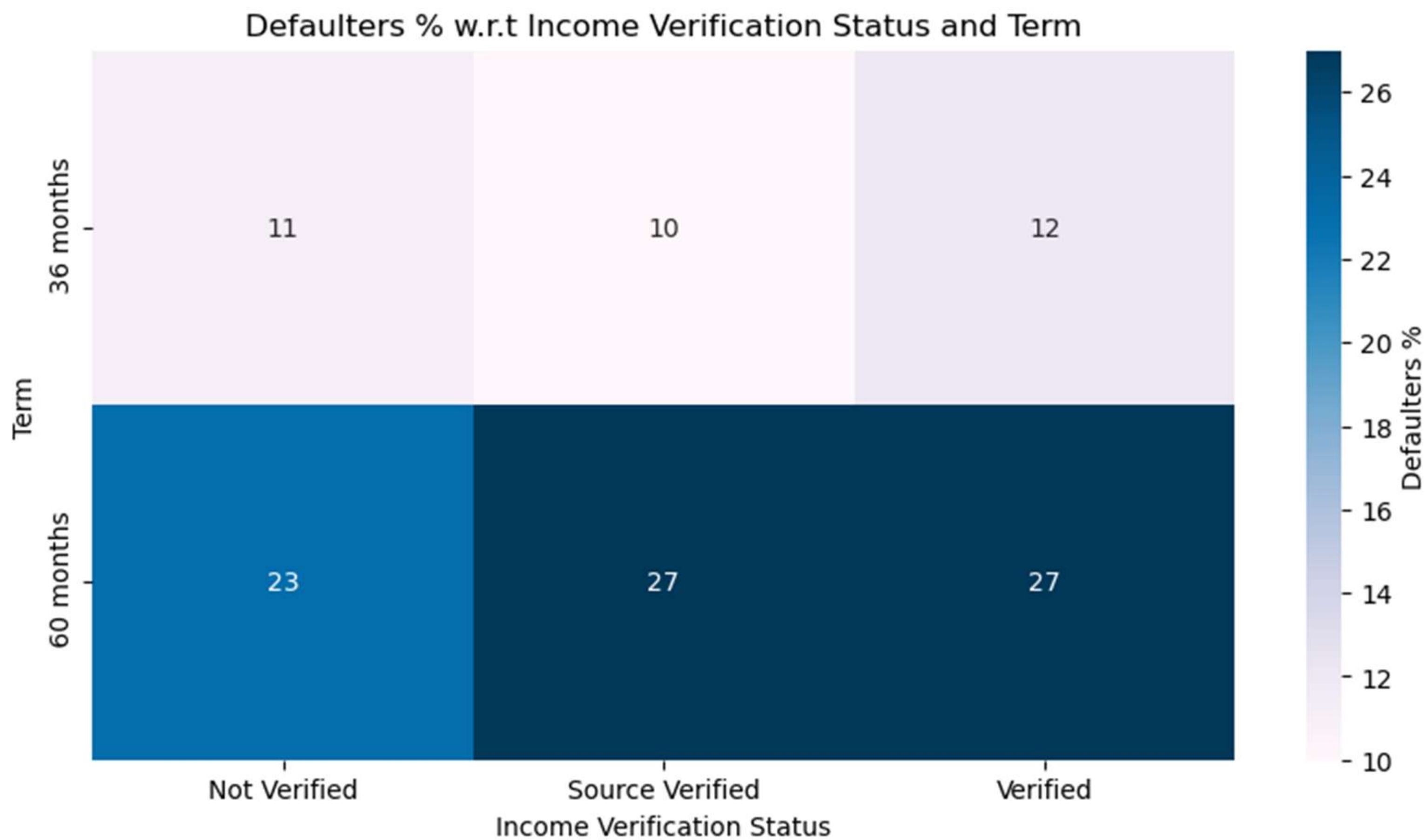
Bivariate Analysis

- **From the given correlation heatmap, we can conclude that there is a strong positive correlation between installment and loan amount**



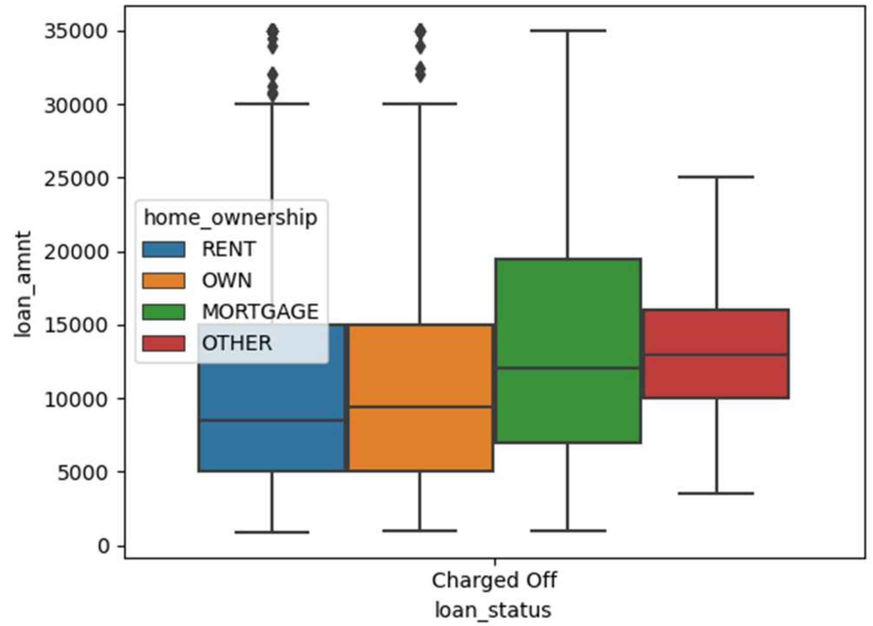
Bivariate Analysis

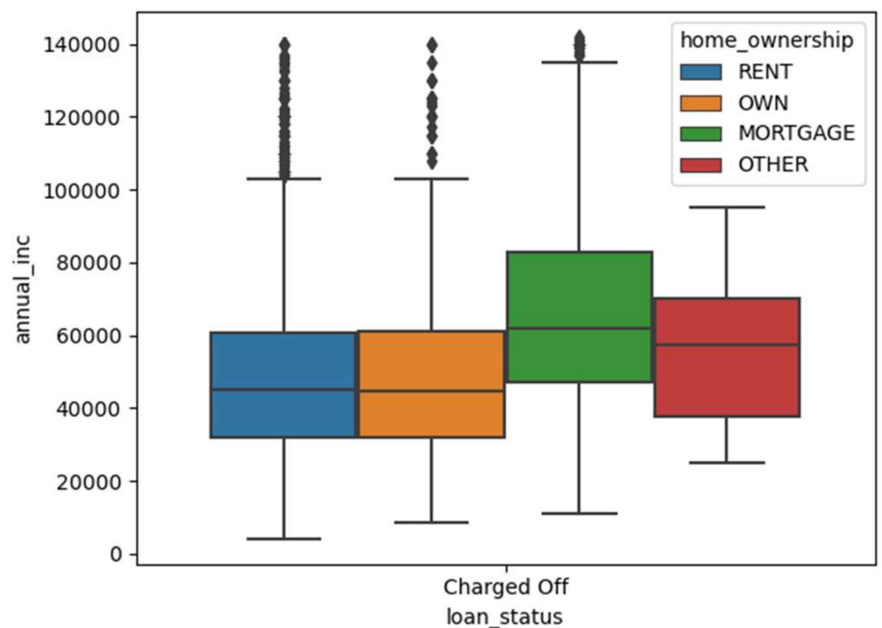
- **Applicants who have verified incomes tend to default more, which is quite surprising and one of the reason of same could be that verified income applicants are easily given bigger loans and also at higher interest rate which tends them to default loan**



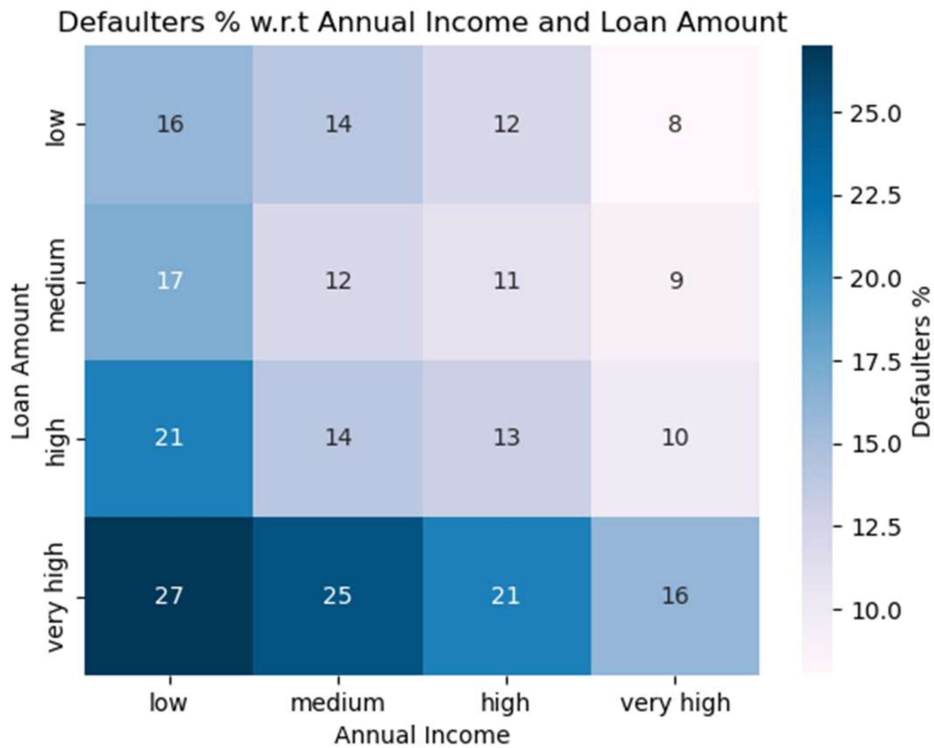
Bivariate Analysis

- Applicants having Mortgage home ownership and having income between 60,000 to 70,000 and have high amount of loan are likely to get the default.





Derived Metrics

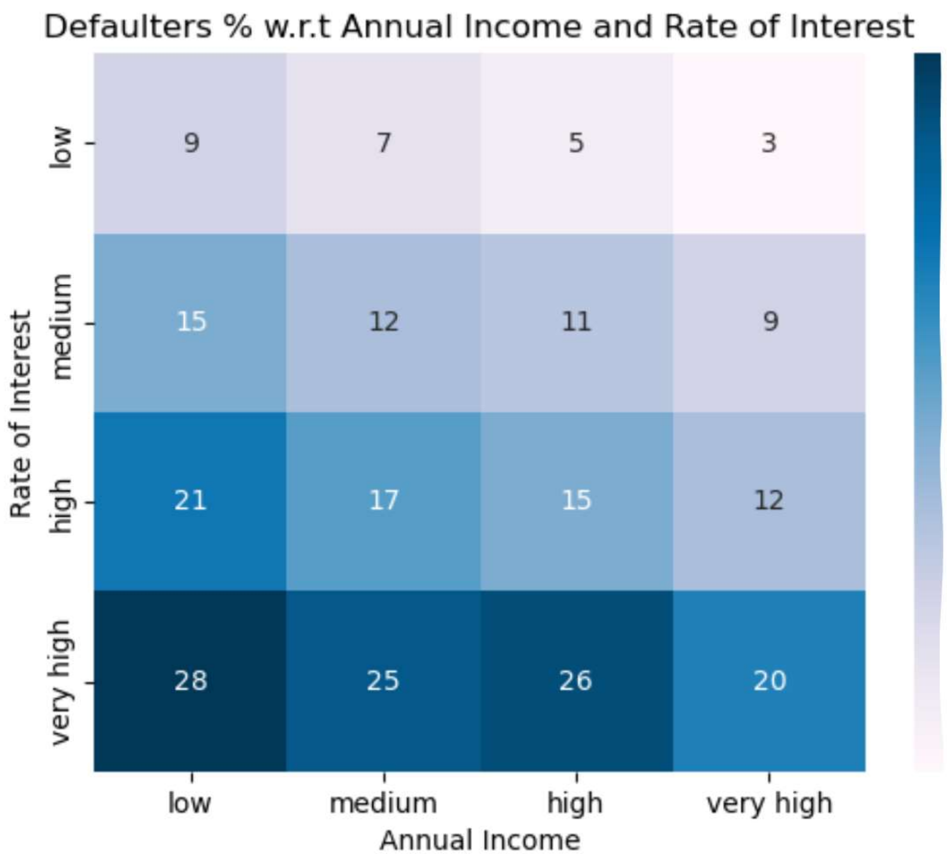


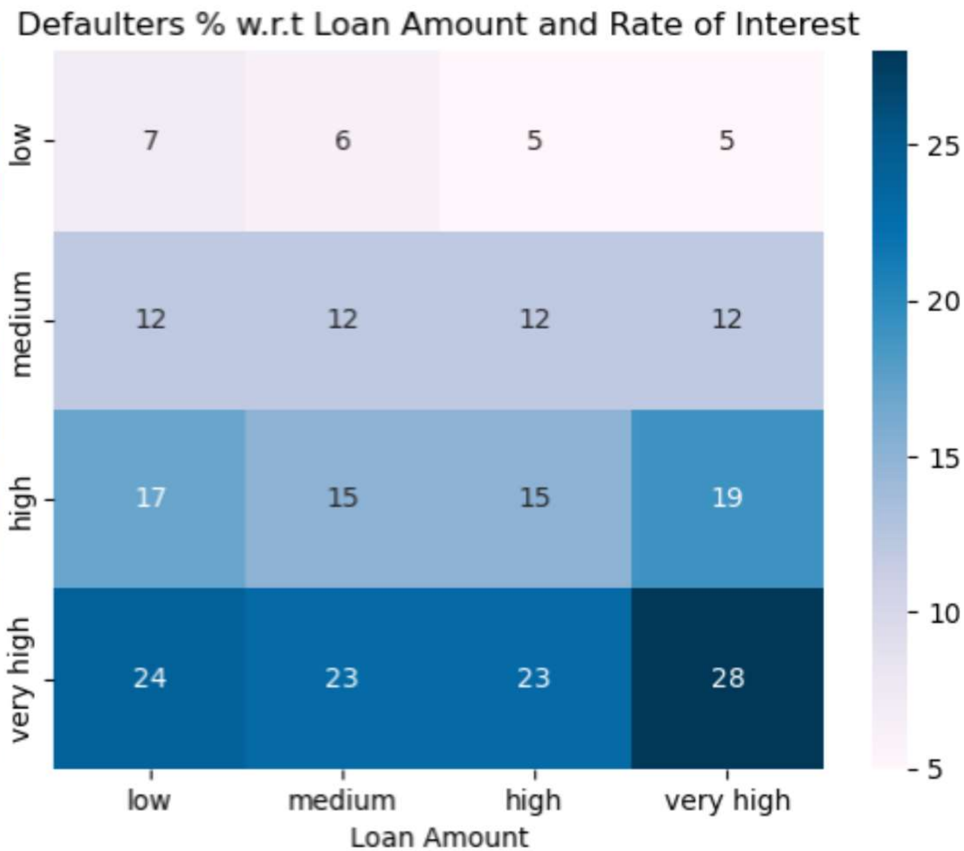
Have categorised a few domain based important variables into bins from low to very high on the basis of Quartiles.

- **As per the heatmap, an applicant defaults a loan when <u>loan amount is very high</u> no matter income earned by applicant, and also when an applicant has a <u>low annual income</u>**

Derived Metrics

- The % of <u>defaulters keeps incresing with increasing Rate of Interest</u> and <u>defaulters also increases with decrease in annual income.</u>
- A Loan is likely to be defaulted if it has <u>high or very high interest rates</u> especially when <u>loan amount is also very high</u>.





Summary

- Loan amount has spikes in multiples of 5000, ie, people prefer taking loan in multiple of 5000's
- Around 75% of applicants preffered to pay back loan in 36 months
- People employed recently ie in starting 0-5 years tend to take much loans than people with mature length in jobs
- Majority loans are taken to pay off other loans
- People with higher loan amount tend to pay in higher tenure ie 60 months
- Annual income had increase with their Employment Tenure
- People having mortgaged house tend to have bigger loan amounts in comparison to owned house owners and rented house owners
- People having Rented houses are less likely to get loan at lower interest rates
- People having mortgaged houses tend to have higher annual income
- People with verified incomes are easily given bigger loans
- People with verified incomes get loans at higher interests
- The rate of interest keeps increasing with grade going from A to G

Problem Statement – *To identify the applicants to whom loan can be granted to prevent loan defaults.*

- People who default usually have higher <u>amount of loans</u> in comparison to the one's who pays off
- People having higher <u>rate of interest</u> tend to default on loan ie, on average more than 11-12%
- People having higher <u>annual income</u> tend to fully pay the loan
- People having 60 months as loan <u>payment tenure</u> tend to default more
- Loans of <u>Grade</u> E, F & G tend to get much more defaulted, specially sub grade F5, G3, G5
- Loans given for the <u>purpose</u> of Small Business, Renewable Energy & Education have higher rate of getting defaulted, whereas, loan given for Weddings has highest chances of being fully paid
- People opting for 60 months loan tend to default specially if they took it for Small Business or Educational purpose.
- Applicants who have <u>verified incomes</u> tend to default more, which is quite surprising and one of the reason of same could be that verified income applicants are easily given bigger loans and also at higher interest rate which tends them to default loan
- If Interest Rate of Loan is high, people tend to default on loans irrelevant of any other factor
- People having lower income and opted for 60 months tenure tend much more to default loan

Solution & Driving Factors

1. **Interest Rate** – Lending Club shall not give loans at interest rate higher than 18% to prevent.
2. **Term** – Lending club shall grant majority loans in 36 months tenure as 60 months is more prone to default.
3. **Annual Income** – People having higher annual income tend to fully pay the loan, therefore people having income higher than 50,000 shall be given loan.
4. **Income Verification** – People having verified income also shall not be given higher loans with higher interest rate as that tends a person to default on loan.
5. **Grade** – It is recommended to discontinue loans classified as grade F and G, given they had significantly higher default rates.
6. **Purpose** – Loans extended to Small Businesses and Renewable energy initiatives will undergo increased evaluations due to a higher occurrence of loan defaults.

Thank You

By Ruchir Dhingra & Tanupriya Chaurasia