



Lending Club Case Study

Prepared By

- Saritha
- Saravanaraj

Table of Contents

- Problem Statement
- Objectives
- Loan Data statistics
- Analysis Approach
- Driver variables
- Case Study Analysis
- Conclusions
- References

Problem Statement

Lending Club is a consumer finance marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

In other words, **borrowers** who **default** cause the largest amount of **loss to the lenders**. In this case, the customers labeled as *'charged-off' are the 'defaulters'*. The core objective of the exercise is to **help the company minimize the credit loss**.

There are two potential sources of **credit loss** are:

1. Applicant **likely to repay the loan**, such an applicant will bring in profit to the company with interest rates.** Rejecting such applicants will result in loss of business**.
2. Applicant **not likely to repay** the loan, i.e. and will potentially default, then approving the loan may lead to a financial loss* for the company

Objectives

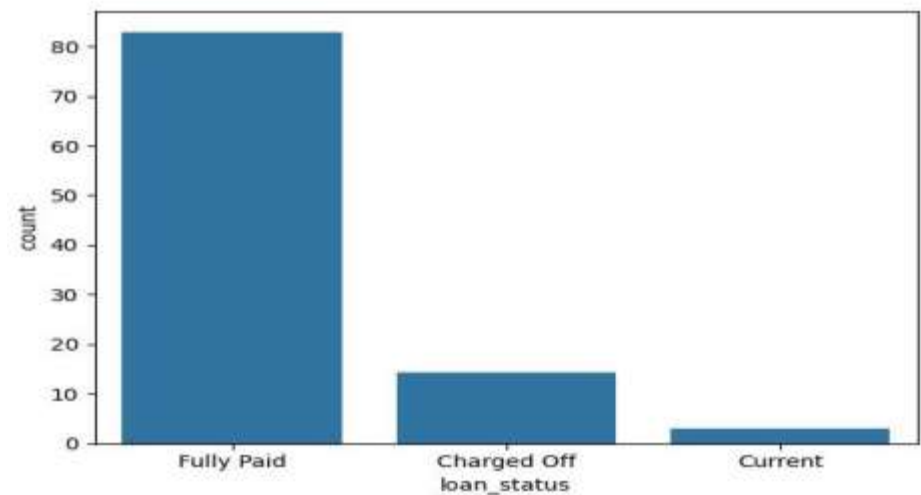
The goal is to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA using the given dataset, is the aim of this case study.

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 utilize this knowledge for its portfolio and risk assessment.

Loan Data statistics

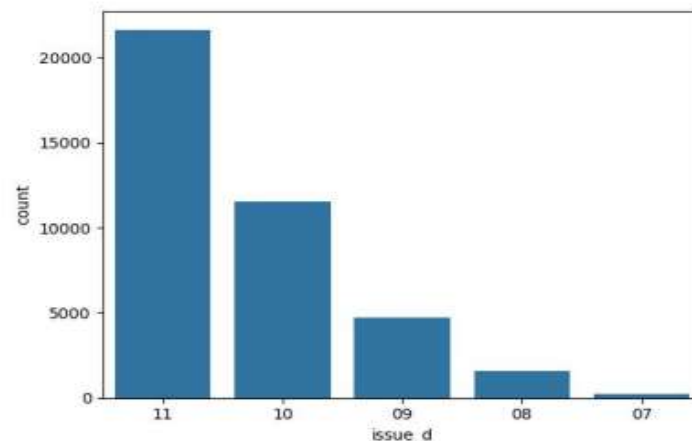
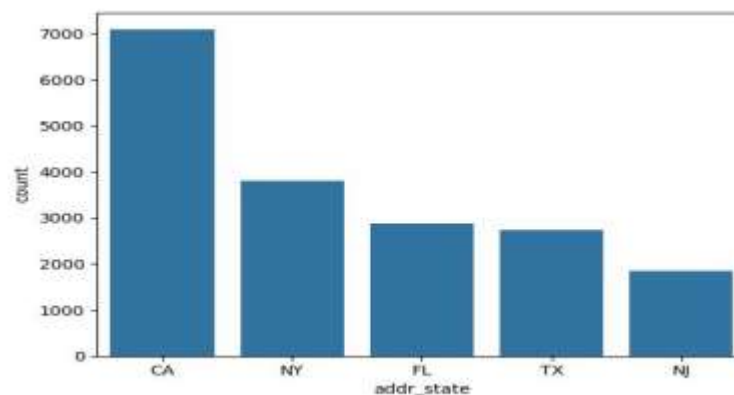
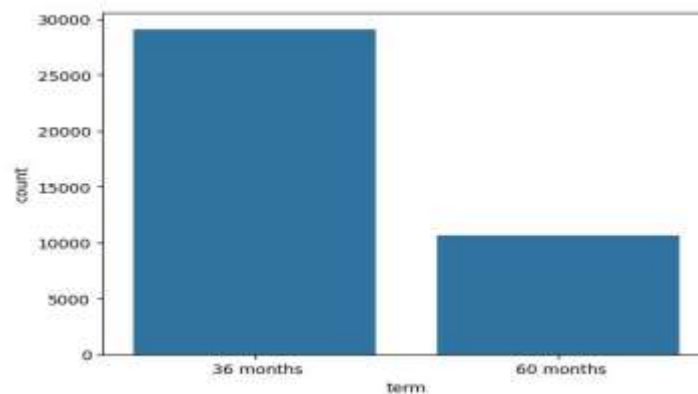
Total Number of Loans	39717	
Number of Paid Loans	32950	82% of total loans
Number of Loan Defaults	5627	14.16% of total loans
Number of Floating Loans	1140	2.8% of total loans



Loan Data statistics

A total of 445+ million has been disbursed.

- The available tenures for each loan stands at 36 and 60 months
- The interest range varies between 5% to 25% per annum
- The Primary Geography is United States with top 5 states being California, New York, Florida, Texas, New Jersey
- The Primary Loan Purposes include debt_consolidation, credit card, major_purchase and home improvement
- The disbursals are around 2007-2011 timeframe



Analysis approach

➤ **Clean Data**

- ☐ As part of this drop the columns having all null values and the columns having more null values
- ☐ Removing columns which are single valued as they are not contribute to analysis
- ☐ Analyze the column data type and correct them to required data type

➤ **Univariate analysis**

- ☐ Check distributions and frequencies of various numerical and categorical variables
- ☐ Created derived variables

➤ **Segmented Univariate Analysis**

- ☐ Analyze variables against segments of other variables
- ☐ Create derived variables

➤ **Bivariate Analysis**

- ☐ Do correlation analysis, check how one variable effects other variable or third variable
- ☐ Analyze the join distributions across different variables

➤ **Summarize results**

- ☐ Provide the summary of analysis results, insights and observations

Driver variables

These attributes are available at the time of the loan application and strongly helps in **prediction** of loan pass or rejection.

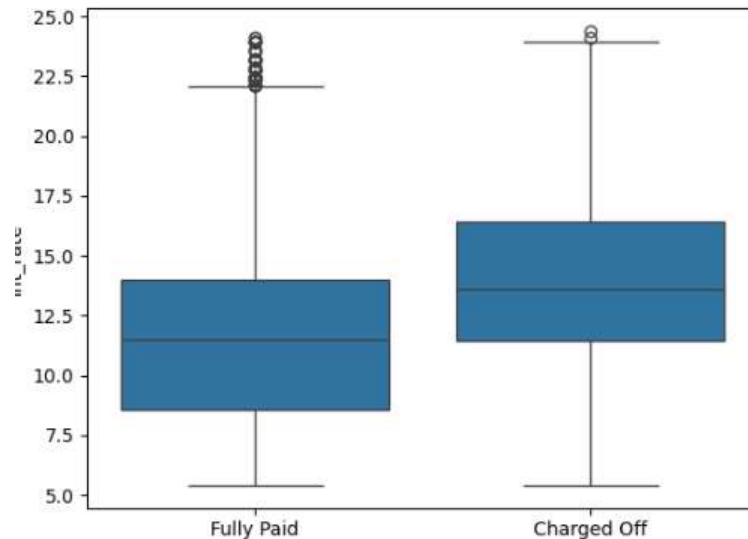
Loan Attributes

- Loan Ammount (loan_amt)
- Grade (grade)
- Term (term)
- Purpose of Loan (purpose)
- Loan Status
- Interest Rate (int_rate)

Customer Demographics

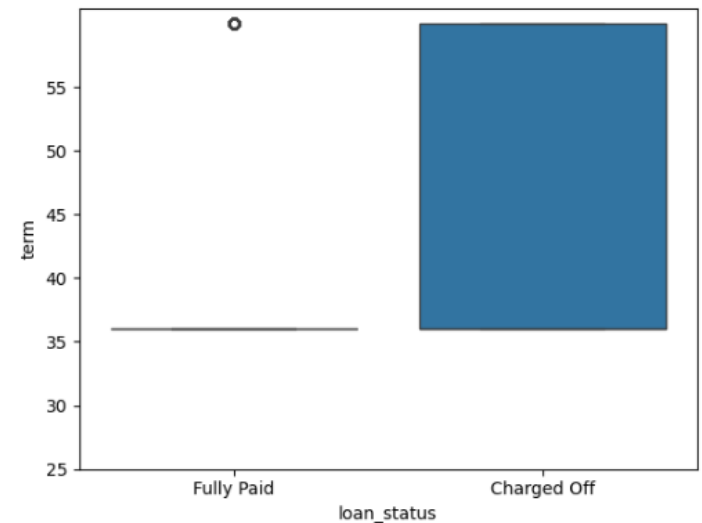
- Annual Income (annual_inc) - Annual income of the customer.
- Home Ownership(home_ownership) - Whether the customer owns a home or stays rented.
- Employment Length (emp_length) - Employment tenure of a customer
- Debt to Income (DTI) - The percentage of the salary which goes towards paying loan.
- State(addr_state) - Location of the customer.

Analysis on Interest Rate, Tenure



The charged off loans are having higher interest rates compared to that of Fully Paid Ones

The tenure of Charged Off variables are spreading to 60 months whereas all fully Paid loans are with 36 month tenure.

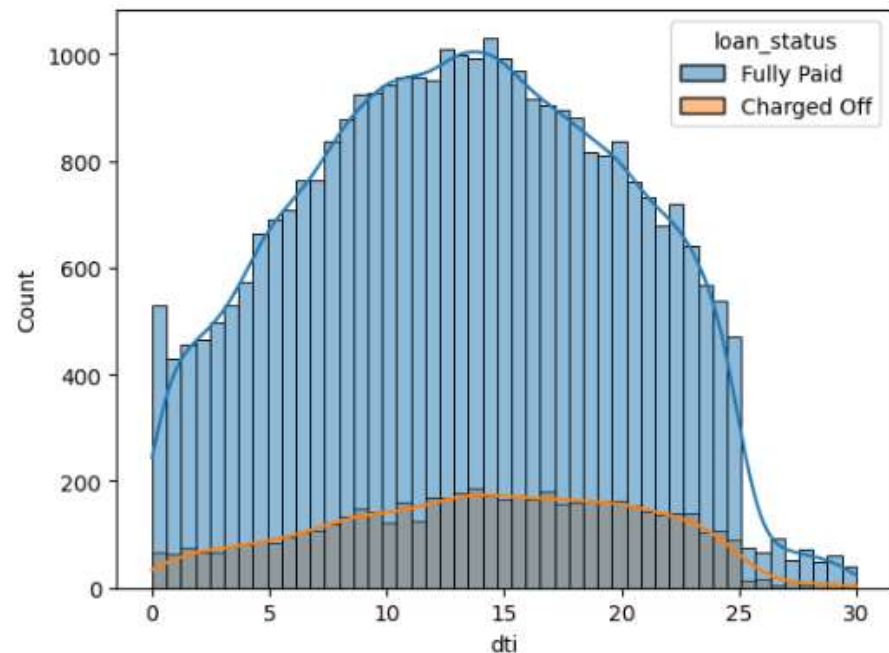


Analysis on Annual Income , Debt to Income

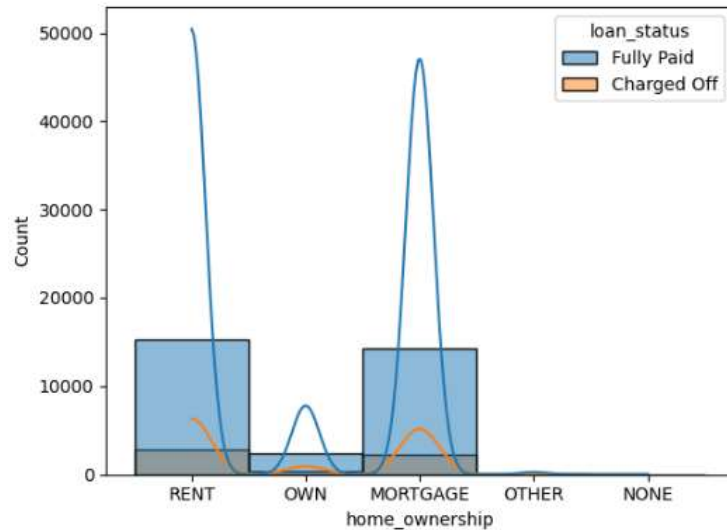
annual_inc_in_lakhs	
loan_status	
Charged Off	0.75
Fully Paid	0.85

The annual income of charged off loans are almost 10K Dollars less than the Fully paid counterparts

The dti ratio for the charged off loans seem to be higher almost 1 base points above Fully Paid Loans



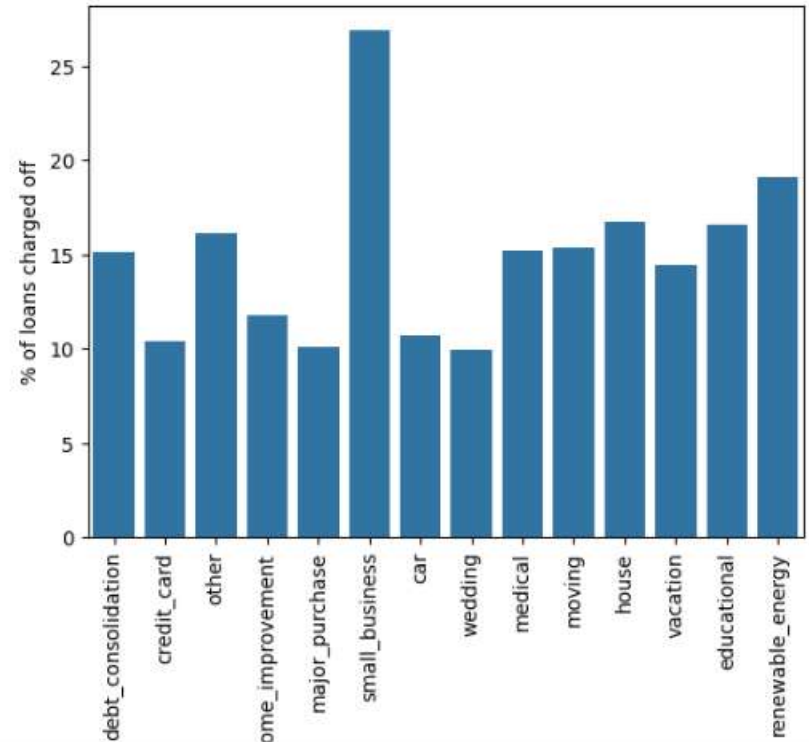
Analysis on Property Ownership



The customers in RENT and mortgage category seem to default more than the houses owned

Analysis on Loan Purpose

The Small Business segments seems to have higher charged off ratios compared to that of remaining loans which is above 25%



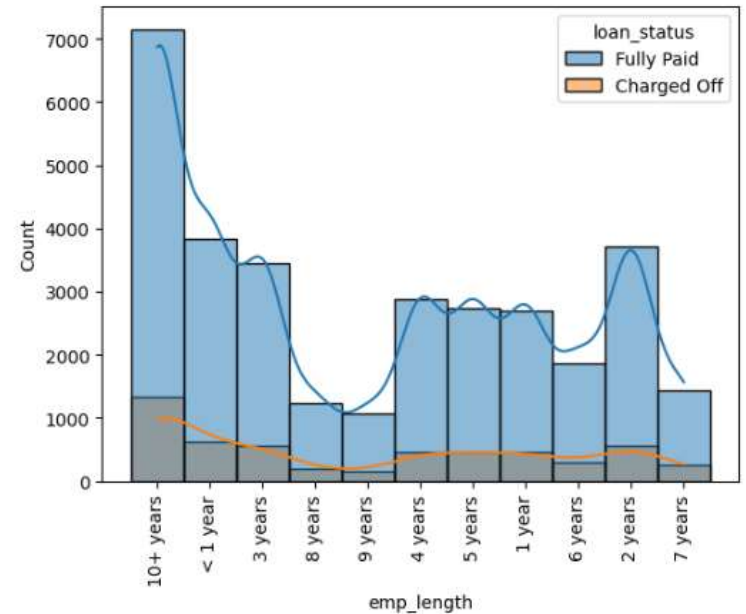
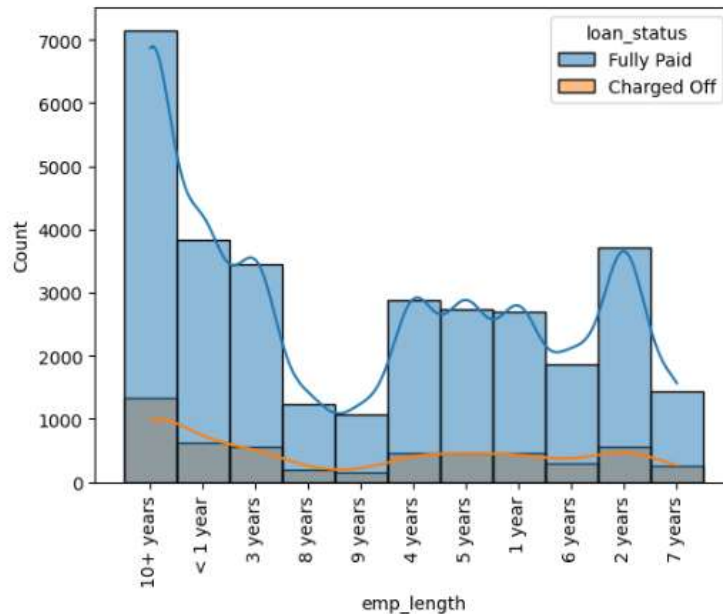
Analysis on Loan State

FL (Florida) has highest percentage of charging off loans, followed by CA and GA



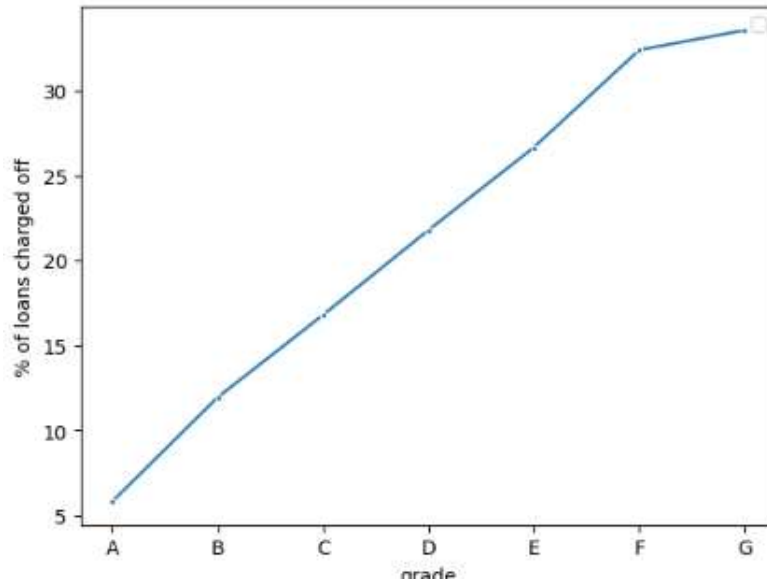
Analysis on Employment Length

The charged off contribution is high in 10 year followed by 2,3,4,5,1 year of emp length



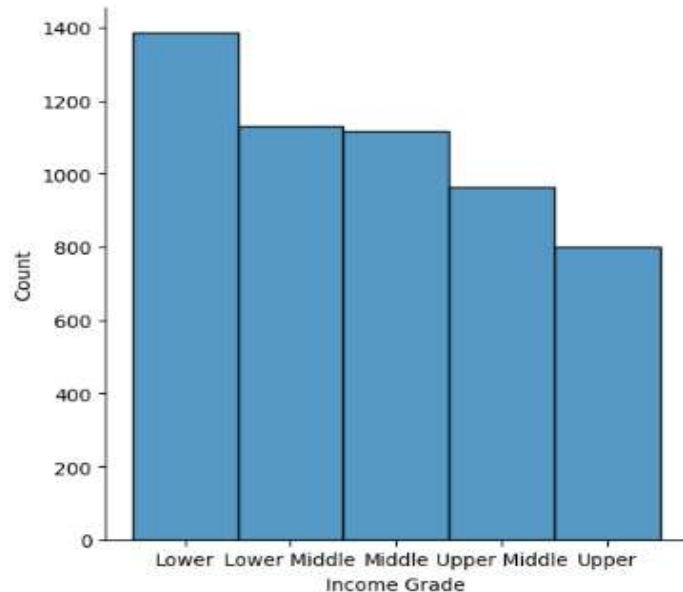
The Employment Length of 10 and above seems to have higher loans (7000 Loans), contributing higher % of charge offs

Analysis on Loan Grade



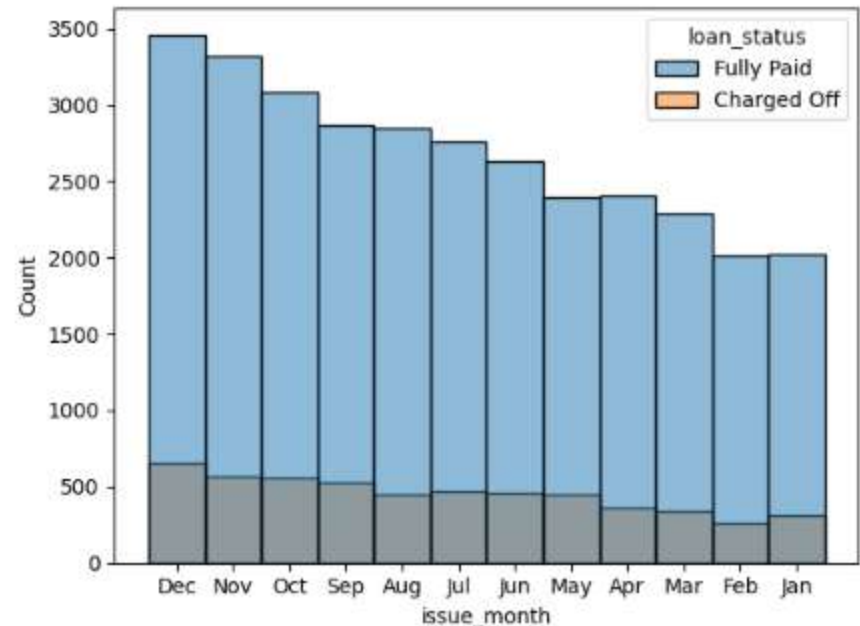
The Charged off loans are getting increased in % as the grade goes down

Segmented Analysis on Issued Month, Income Groups

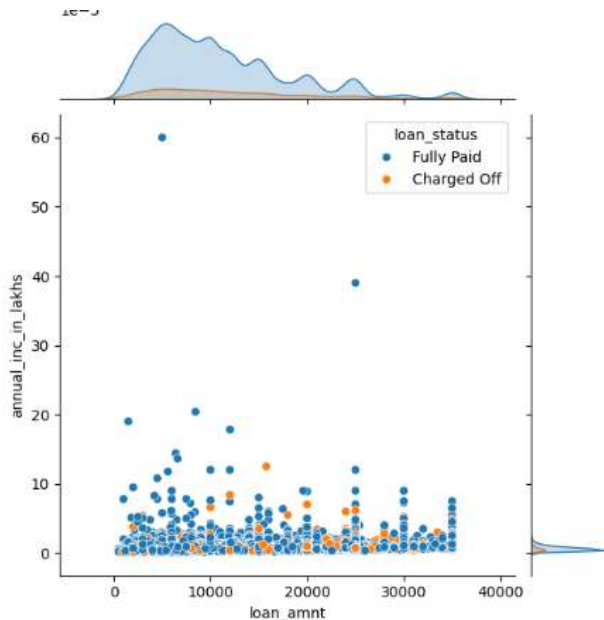


The number of charged off loans are very high in the low income group where as it is less in the high income groups

The month wise analysis of 'Issue date' shows most of the loans are being issued in Dec, November during US holidays

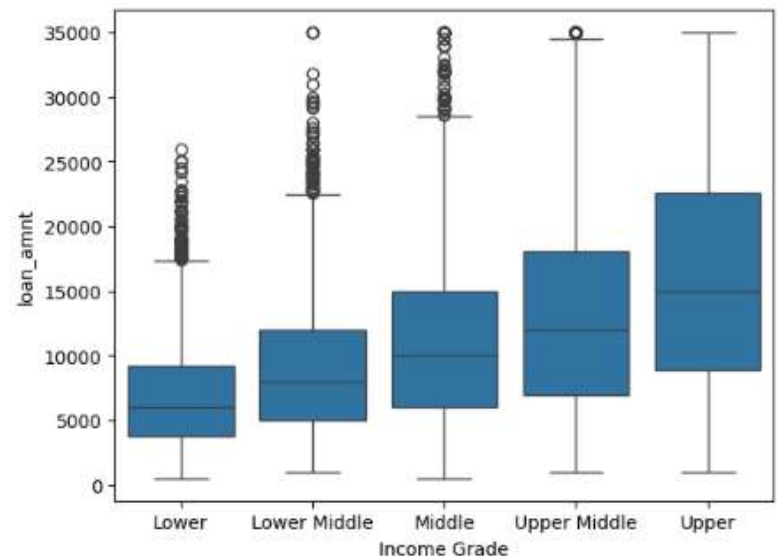


Analysis of Income vs loan status



Loan amount also gives us insights on many outliers in the 'Lower' Income region which are causing defaults upto 29% in the later part of analysis

As the income grade increases the loan amount IQR increases indicating a better distribution



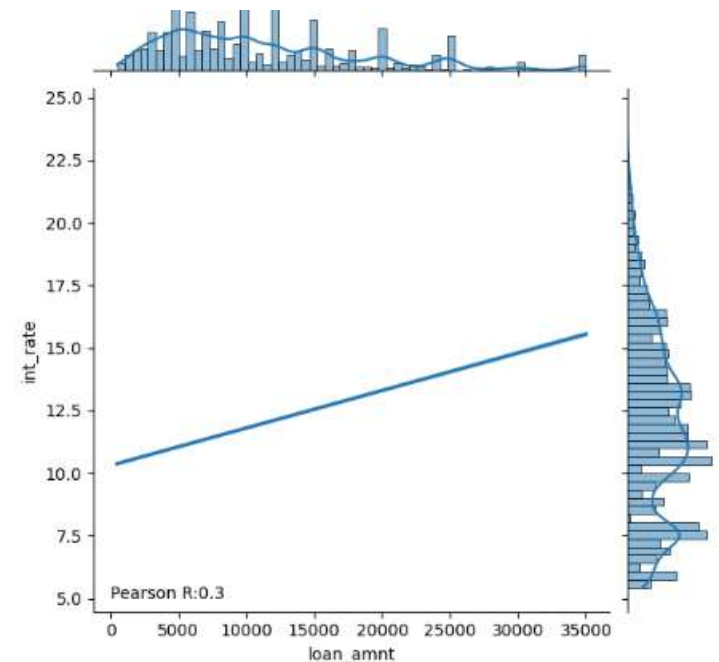
Analysis of Interest rate vs loan amount

% Charged Off (Very High Loan Amount)

int_rate_desc	
Very Less	3.859060
Less	8.913413
Moderate	15.089163
High	17.521090
Very High	28.783935

A strong positive correlation between Loan amount and interest rate

The % Loans getting charged off are increasing as the interest rate reaches towards higher end irrespective of the loan amount.

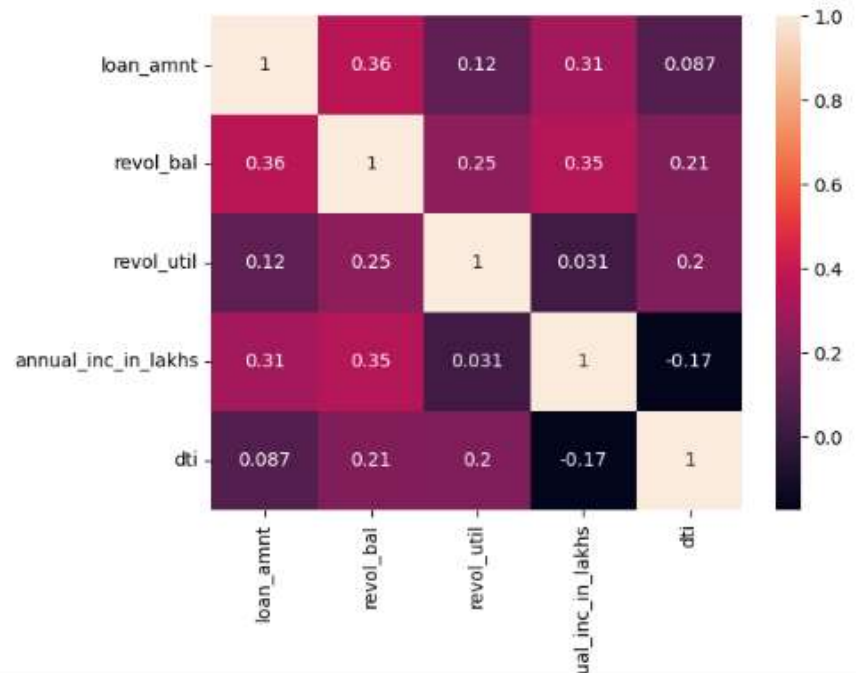


Analysis of credit card segment

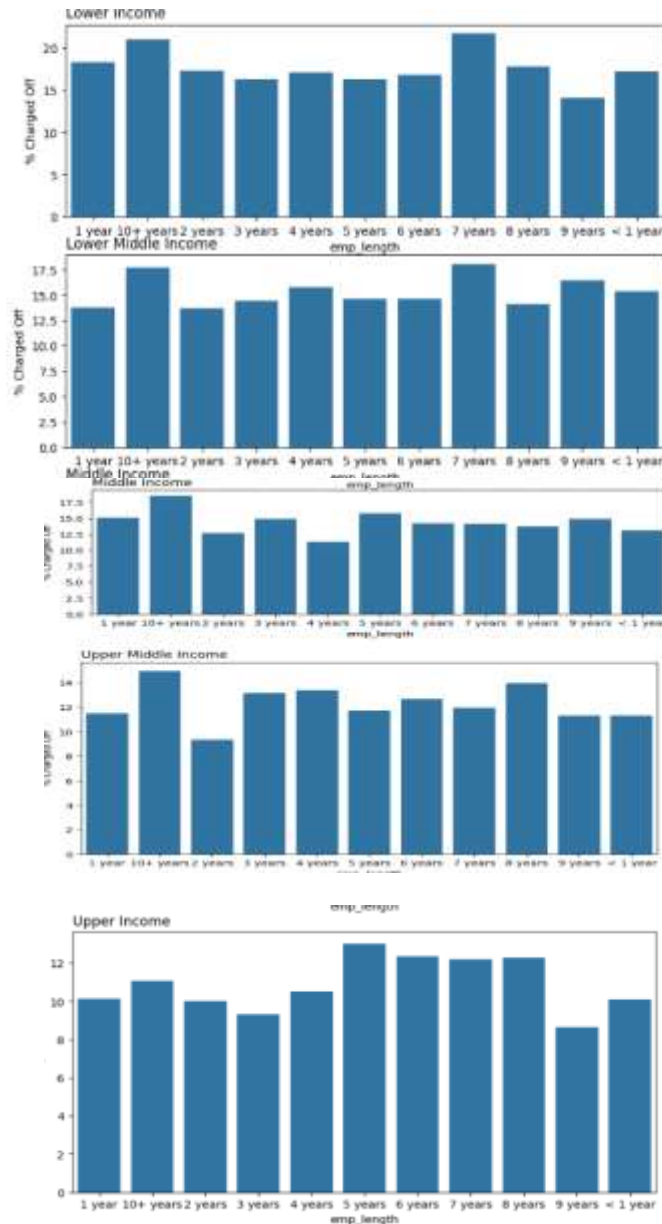
loan_status	dti	
	Charged Off	Fully Paid
purpose		
car	16.9850	16.1000
credit_card	20.1500	19.3125
debt_consolidation	19.9275	19.3200
home_improvement	17.2925	16.0200
medical	17.7675	17.6450
other	18.9650	17.6500
small_business	17.4300	16.1200

The charged off loans are relatively having higher debt to income ratio with credit card holding highest

Negative correlation between annual income & dti within credit card segment in heat map

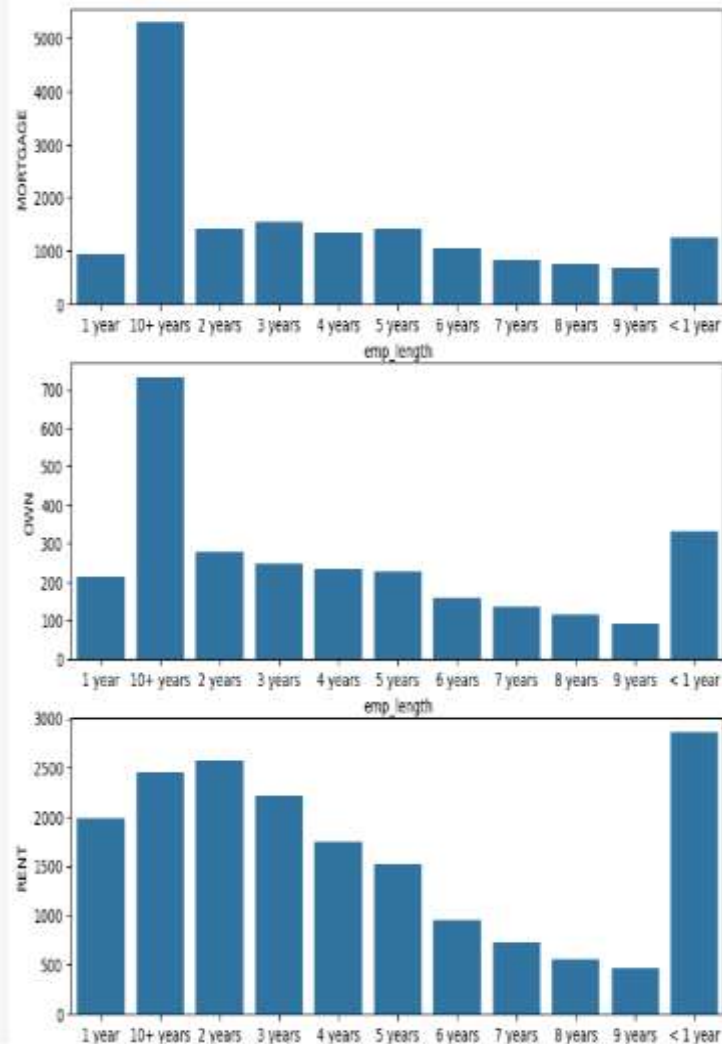


Analysis of employee length with different segments



- The 10+ Years segment is consistently having higher charge offs in each income segment apart from the Upper Income segment
- The charge off percentages increases to 20% as the income decreases within 10+ year bucket

Analysis of employee length with different segments



- The 10+ years have higher contribution in all 3 segments Mortgage , OWN and RENT
- The subsequent peaks are high in 2,3,4 and 1 year segment in MORTGAGE and OWN property segments, which probably is the reason for high charge off percentage in the univariate analysis
- RENT segment has highest contribution by customer with employment < 1 year , which could have been another reason for high charge offs, possibly the rent being burden

Conclusions

Recommendations for better quality Borrowers

- **Review the loans with higher interest rate or The tenure of loan spreading to 60 months. These loans having high risk**
- **Review before approving loans for the purpose of Small Business segments, these are having higher charged off ratios compared to that of remaining loans**
- **Review approving loans for employee with employment Length of 10 and above. They have higher loans, contributing higher % of charge offs**
- **Review loans having higher debt to income ratio with credit card holding.**

References

Git Hub URL:

[sarithajeevna/LendingClubCaseStudy \(github.com\)](https://github.com/sarithajeevna/LendingClubCaseStudy)



Thank you