

IIIT - B , UpGrad

LENDING CLUB CASE STUDY

Presentation by -

BALAMURALI RAGUPATHI
CHANDAN SHARMA BRK



Introduction

Lending Club Overview

Leading peer-to-peer lending platform connecting borrowers with investors. Specializes in personal, business, and other consumer loans.

Objective

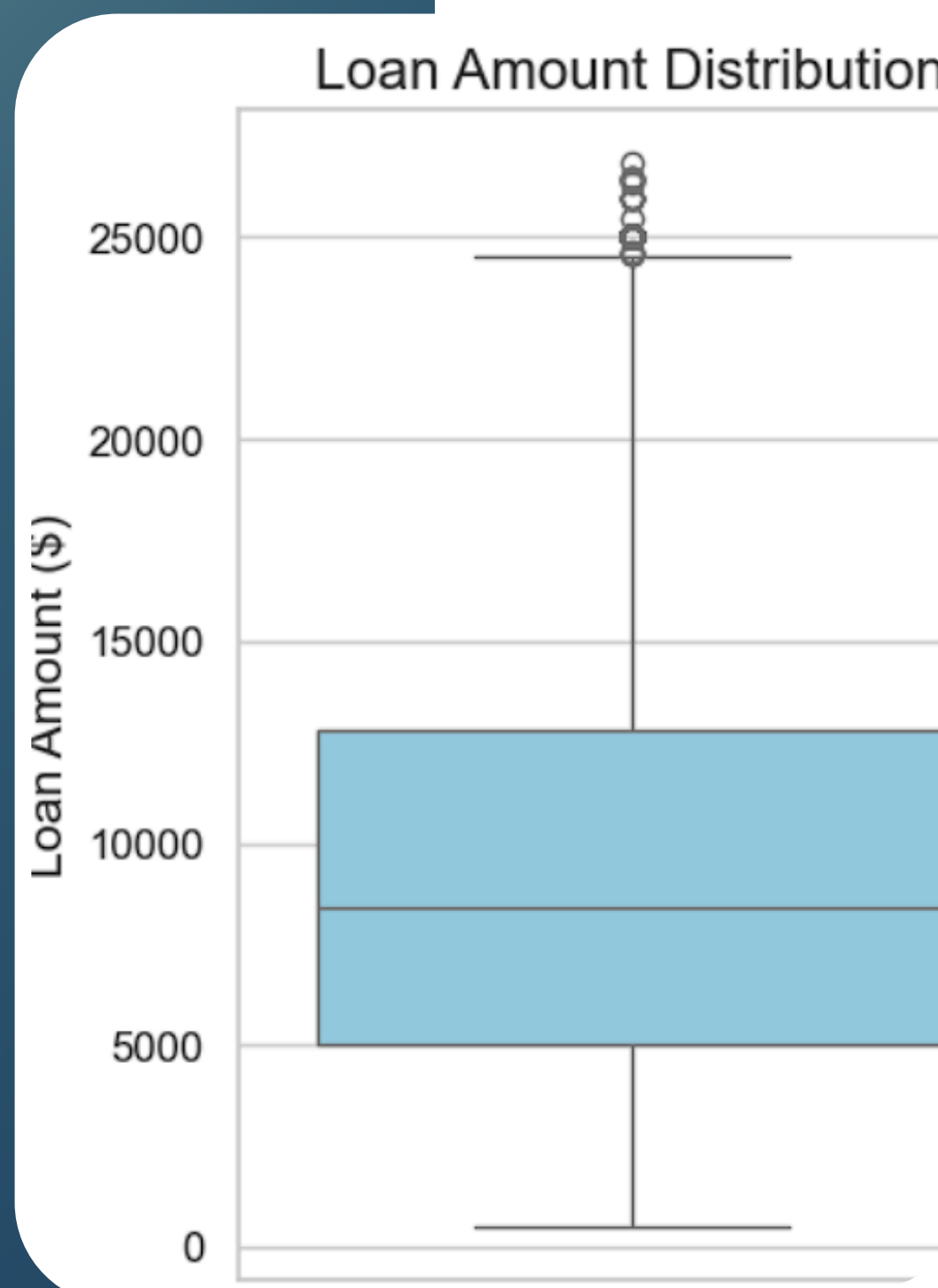
Identify key factors influencing loan defaults to enhance risk management and decision-making.

Importance

Effective risk management reduces defaults, improving profitability and customer satisfaction.

Engagement

Explore which borrower profiles and loan types are most prone to defaults.



Data Understanding

01

Dataset Overview:

Source: Lending Club public dataset

Records: 39717 loan applications

Attributes: 111 key variables including:

Loan Attributes: Loan amount, interest rate, term, grade

Borrower Attributes: Annual income, employment length, home ownership

Loan Status: Fully paid, charged off, current

02

Key Variables

Loan Amount: The total amount of money requested by the borrower.

Interest Rate: The rate of interest applied to the loan.

Term: The duration of the loan (36 or 60 months).

Grade: Lending Club's assigned grade for the loan's risk.

Annual Income: Borrower's yearly income.

Employment Length: Duration of the borrower's current employment.

Home Ownership: Home ownership status of the borrower.

Loan Status: Indicates whether the loan is fully paid, charged off, or current

03

Initial Observations

The dataset contains a mix of continuous and categorical variables. Presence of missing values in some attributes.

Variations in loan amounts, interest rates, and borrower profiles.

Data Cleaning

01

Data Cleaning Steps

- Removed 1,140 rows with loan_status marked as 'current' since they are not relevant for analysis.
- Eliminated 55 columns that contained only null or blank values.

02

Domain-Specific Data Exclusion

- Excluded 21 behavioral data columns that, based on domain knowledge, are not available during the loan approval process and therefore not relevant for the analysis

03

Redundant and Sparse Data Removal

- Removed 8 columns where all values were 1, as their uniqueness did not contribute to the analysis.
- Eliminated 2 columns with more than 50% missing values.

04

Derived Metrics and Final Dataset

- Created new columns issue_year and issue_month derived from issue_d
- Generated bucketed columns loan_amnt_b, annual_inc_b, int_rate_b, and dti_b from continuous data.

After data cleaning, the dataset had 36847 rows and 21 columns



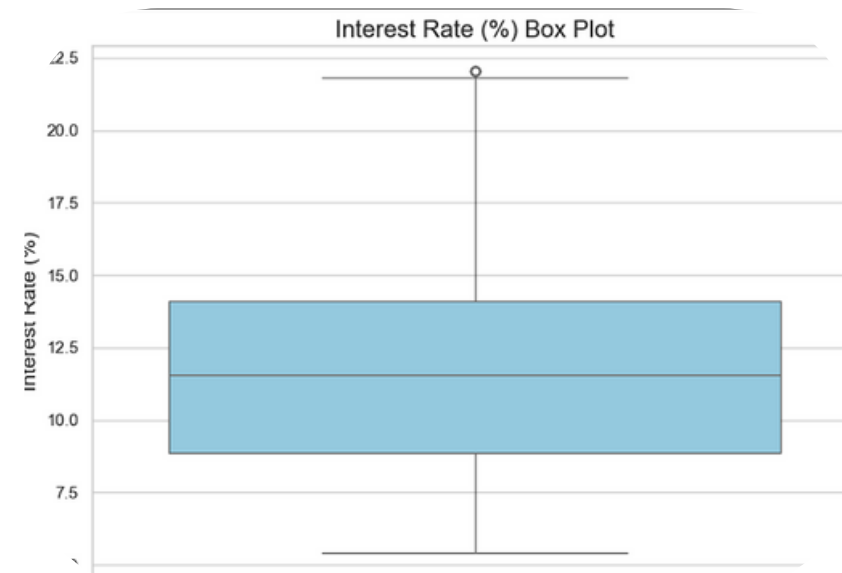
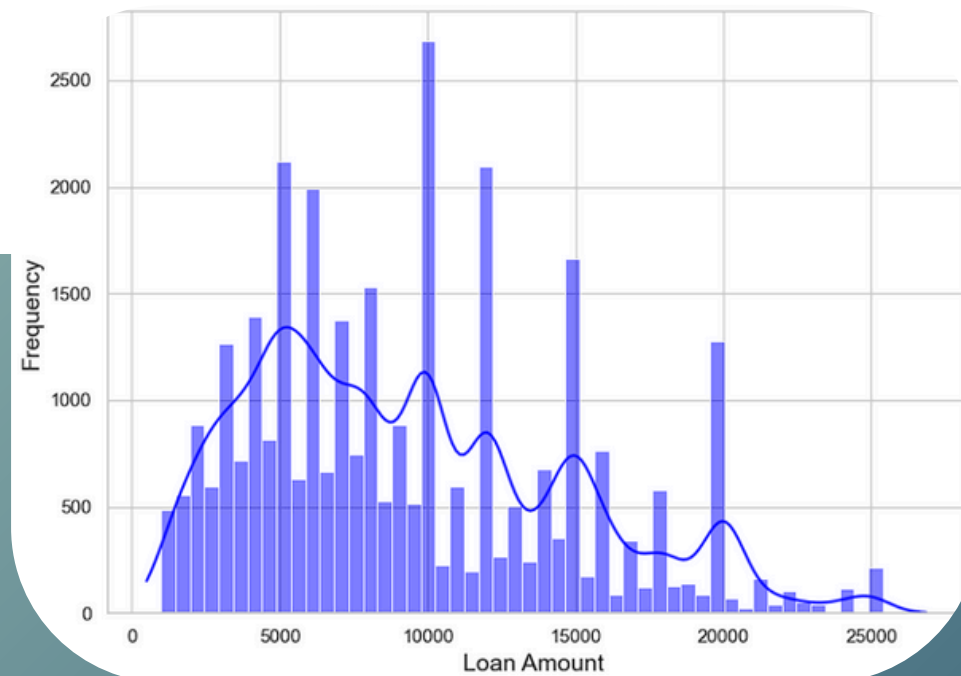
```
(36847, 21)
count      36847.000000
mean       11141.327000
std        7369.988000
min         500.000000
25%        5500.000000
50%       10000.000000
75%       15000.000000
```

Uni-Variate Analysis

01

Loan Amount Stats

- Most of the loan amount applied was in the **range of 5k-14k**.
- Max Loan amount applied was ~27k.
- The Annual income of most if applicants lies between 40k-75k.



02

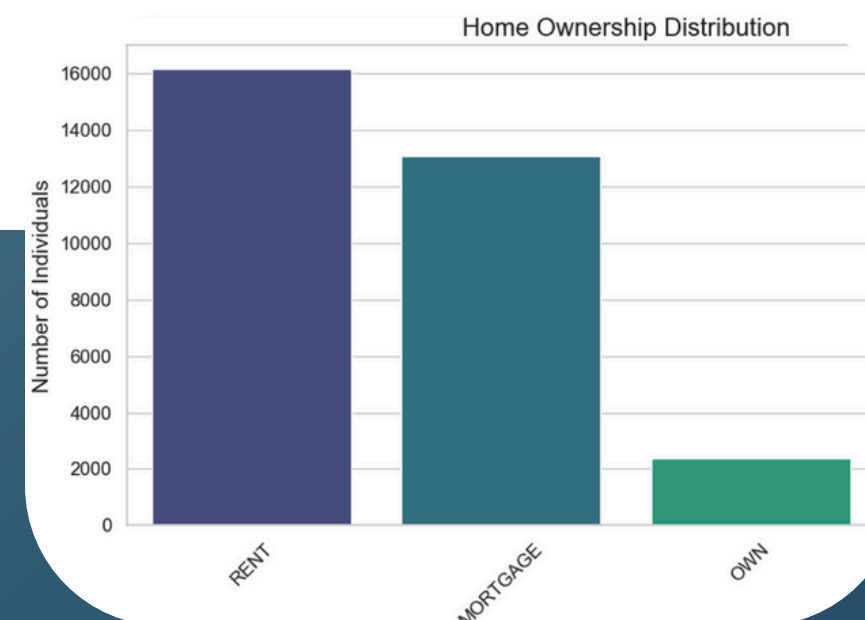
Loan Intrest Rate Stats

- Most of the applicant's rate of intrest is between in the range of 8%-14%.
- **Average Rate of intrest of rate is 11.7 %**

03

Home Ownership

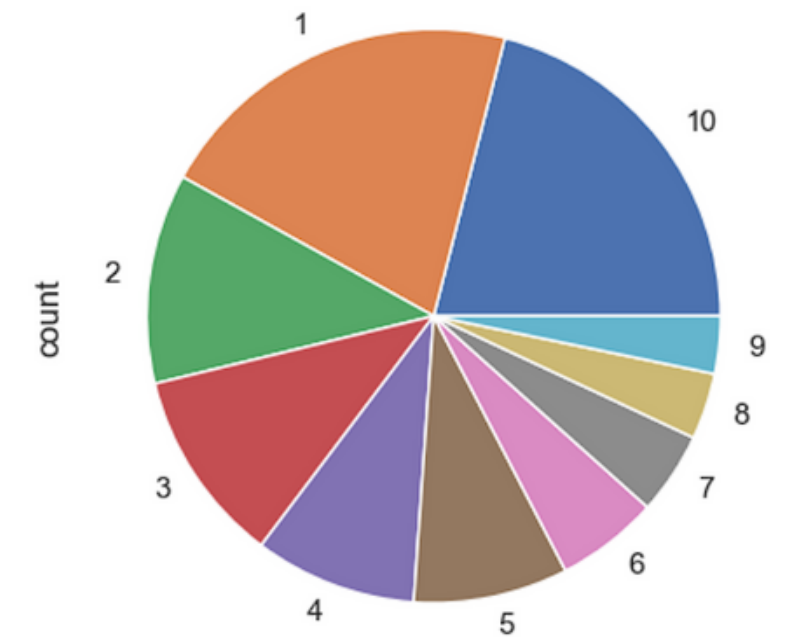
- Majority of loan applicants are either living on Rent or on Mortgage



04

Employee Work Experience

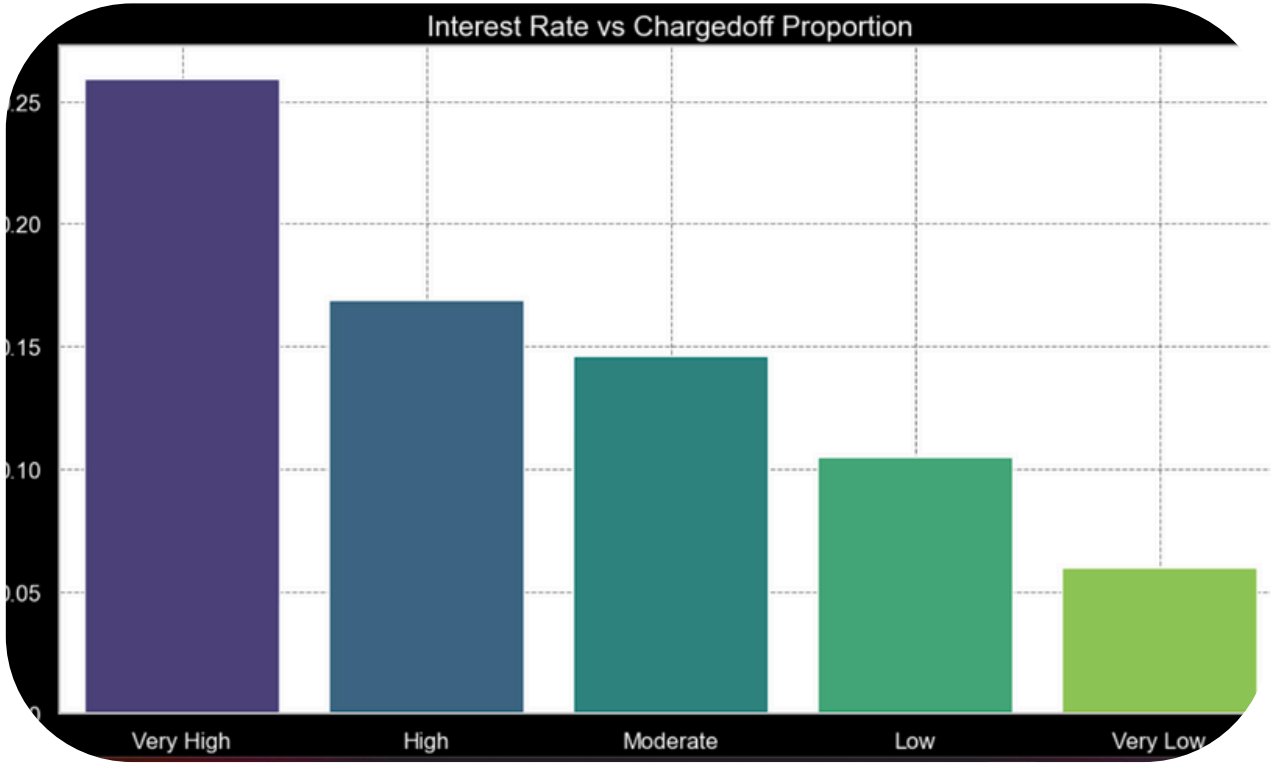
- Most of the applications are having 10+ yrs of Experience.



02

Intrest Rate v/s chargedoff

- Interest rates less than 10% or very low have a very low chance of being charged off. Interest rates start from a minimum of 5%.
- The proportion of charged off accounts increases with higher interest rates.



01

Annual Income Table

We could notice that annual income below 0 to 40K are more people who charged-off , also we could observe the income buckets vs loan taken

loan_status	annual_inc_b	Charged Off	Fully Paid	Total	Chargedoff_Proportion
0	0 - 40k	1566	7316	8882	0.176312
2	50k to 60k	753	4335	5088	0.147995
1	40k - 50k	787	4548	5335	0.147516
3	60k to 70k	451	3164	3615	0.124758
4	70k to 80k	364	2645	3009	0.120970
5	80k - above	547	5299	5846	0.093568

03

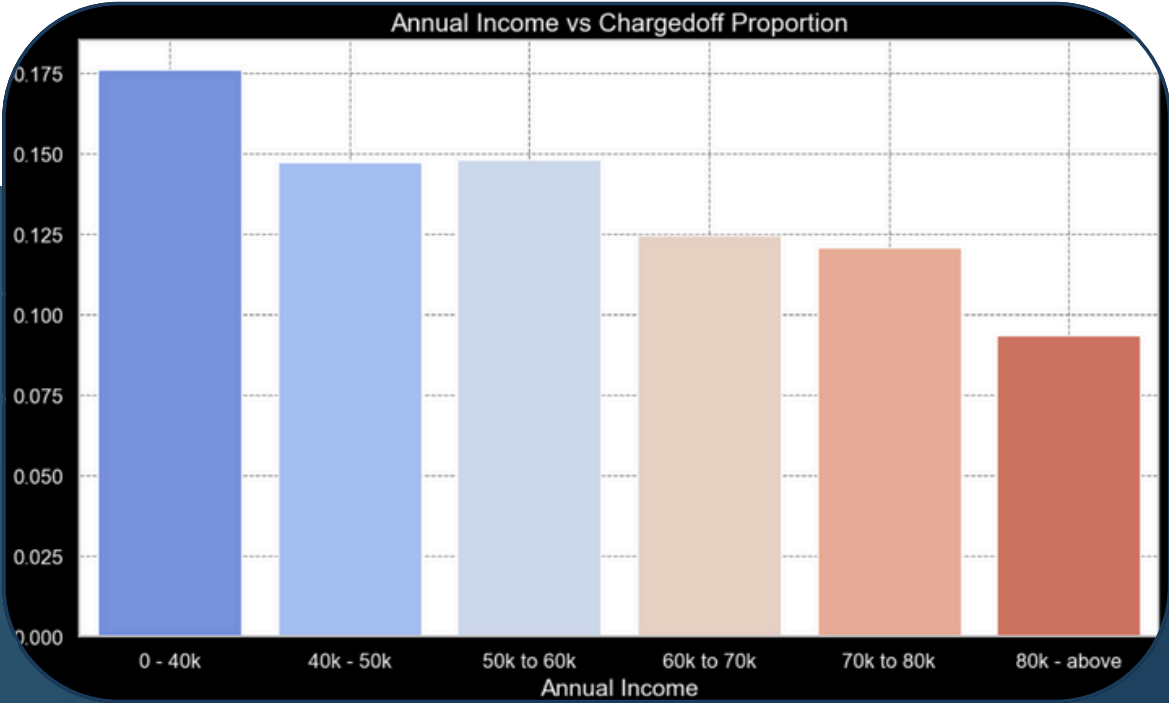
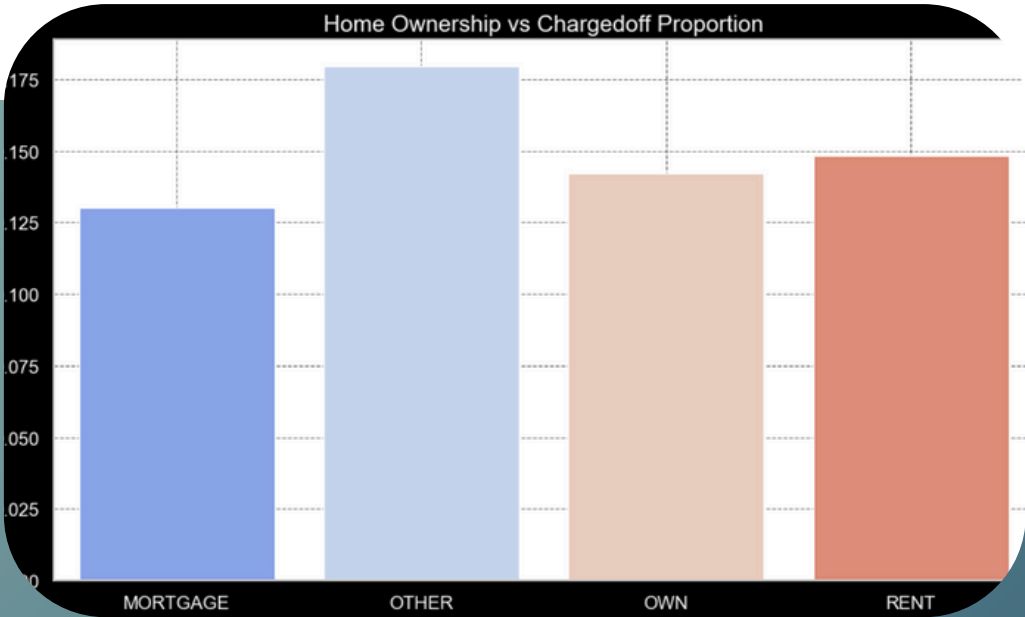
Home Ownership v/s chargedoff

Those who do not own a home have a higher chance of loan defaults

Annual Income v/s Chargedoff

- Income range 0-20000 has high chances of charged off.
- Notice that with increase in annual income charged off proportion got decreased.

04

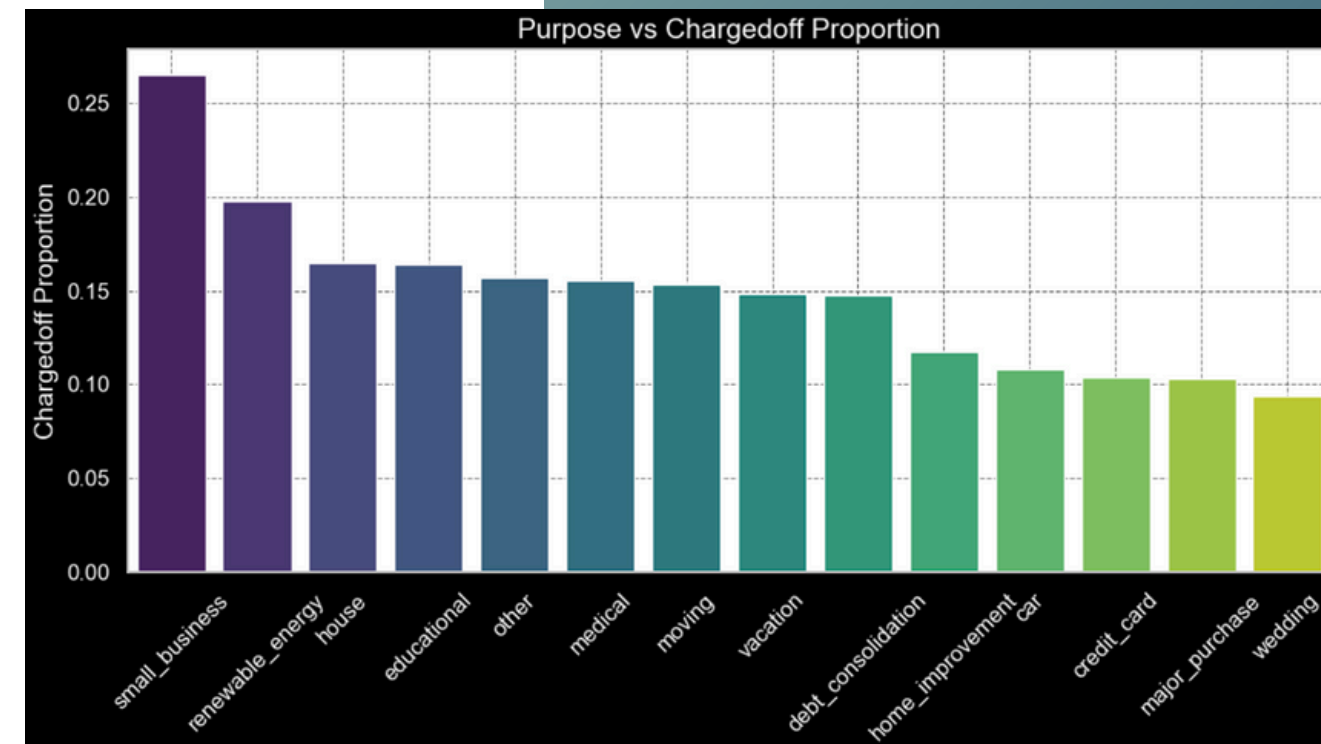


Bi- Variate Analysis - 2

01

Loan Purpose v/s Charged-off

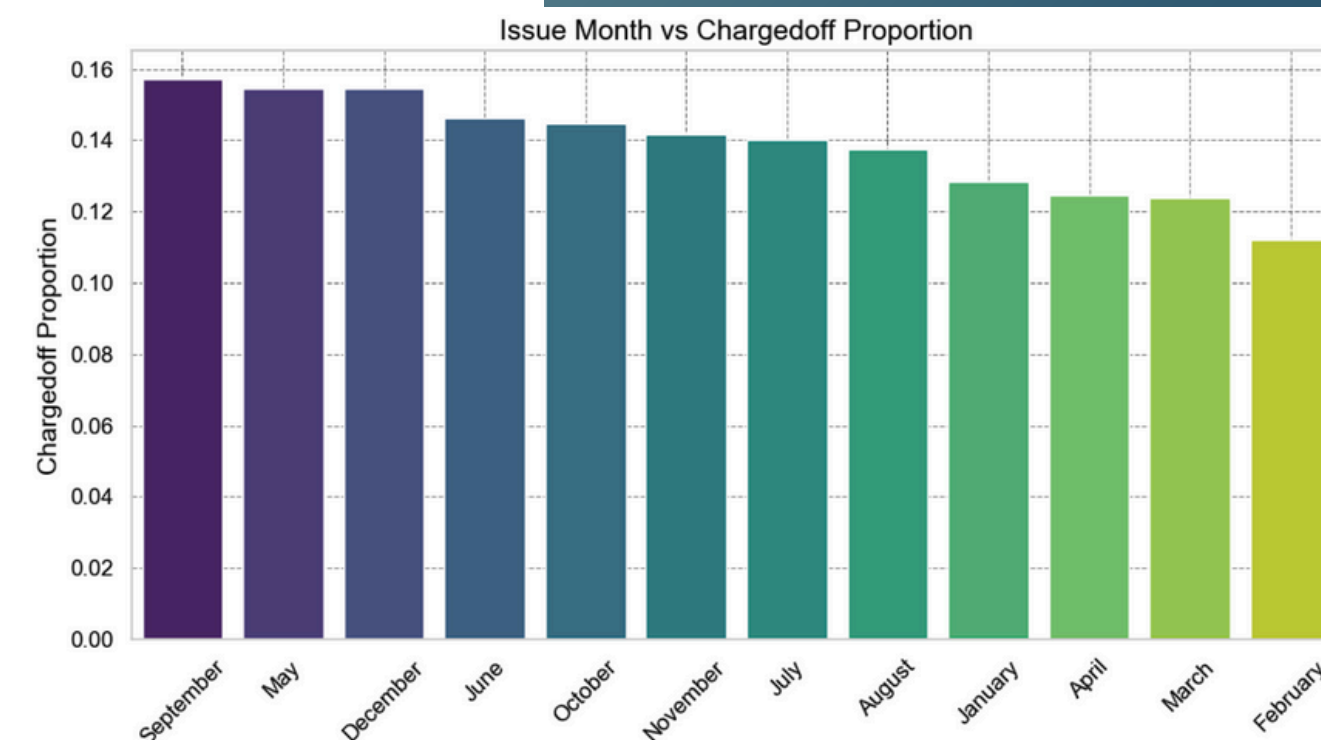
- Applicants who have a home loan have a lower chance of defaulting on their loans.
- Applicants who have a loan for a **small business** have a higher chance of defaulting on their loans



02

Issue Month v/s Charged-off

- Loans issued in May, September, and December have a high number of loan defaults.
- Loans issued in February have a high number of loan defaults.
- The majority of loan defaults come from applicants whose loans were approved from September to December.

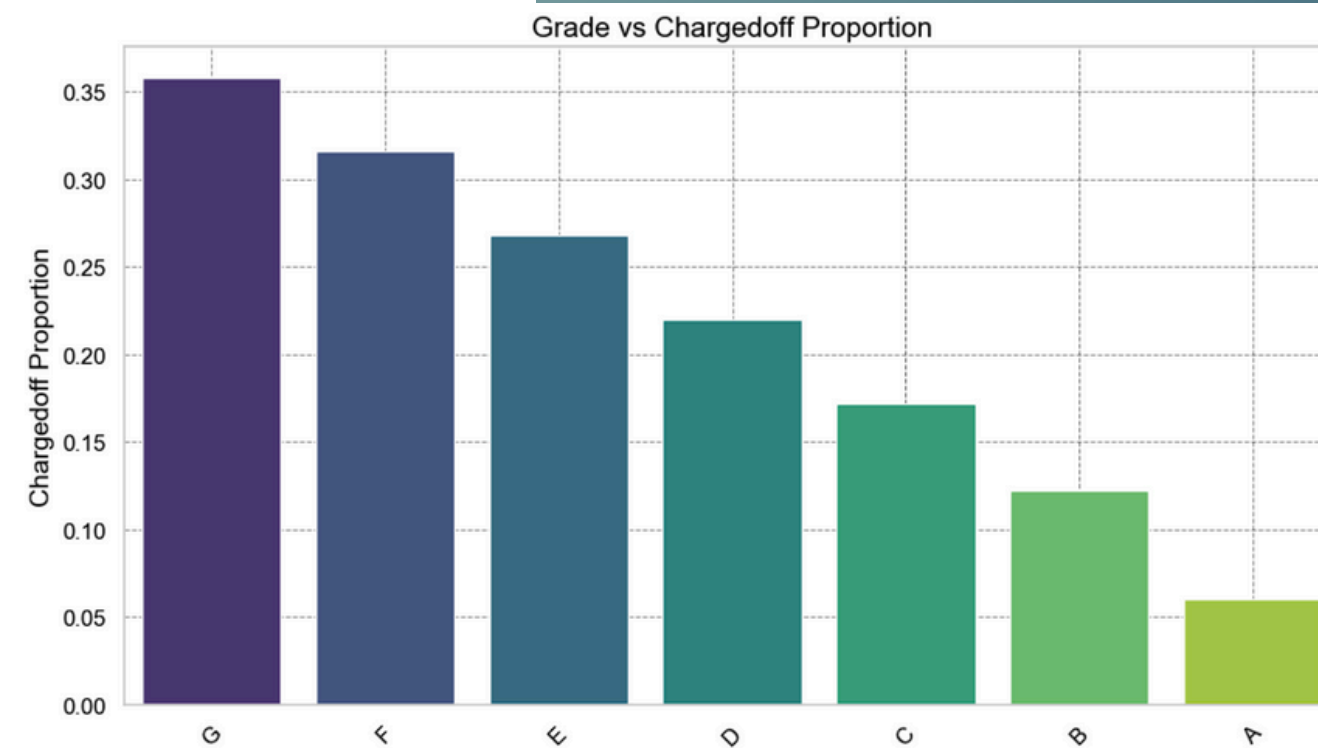


Bi - Variate Analysis - 3

01

Grade v/s Charged-off

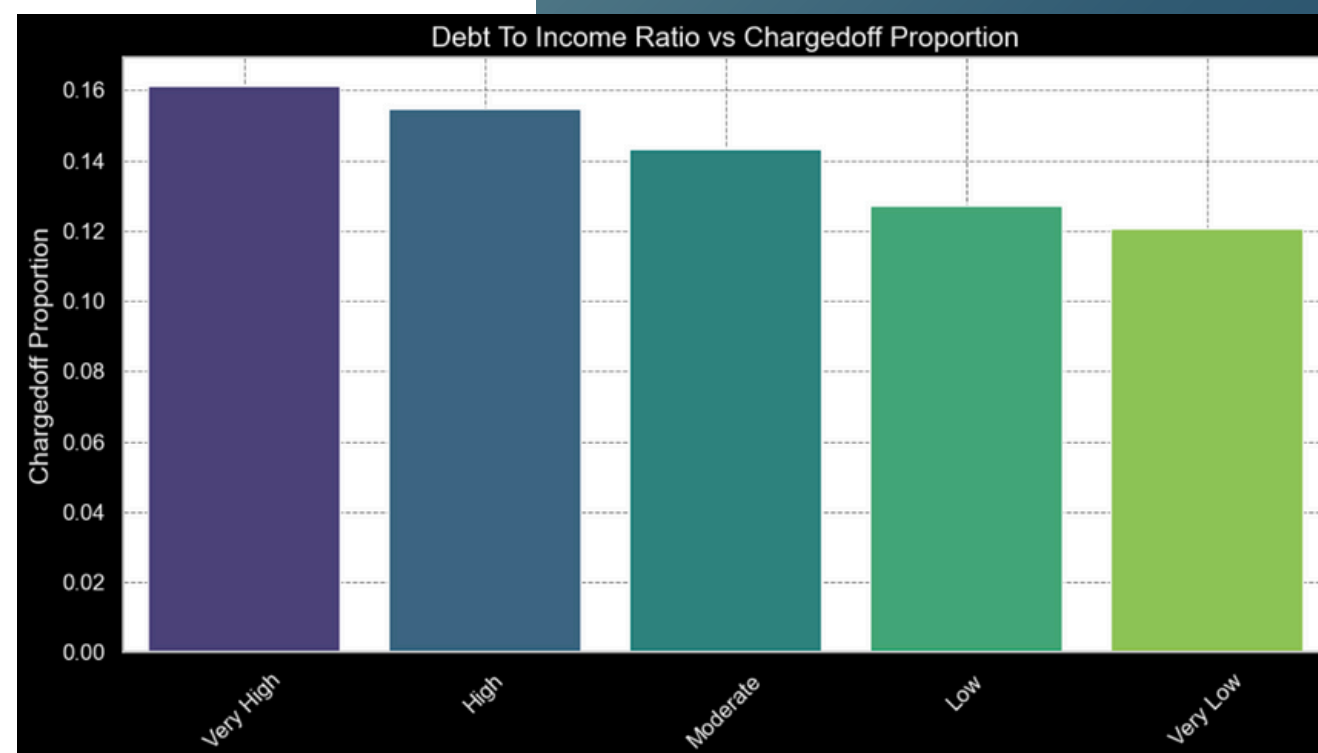
- Loan applicants with a loan grade of G have the highest loan defaults.
- The Lower grades(E,F,G) have higher chances of defaulting the loan than Higher ones(A,B)

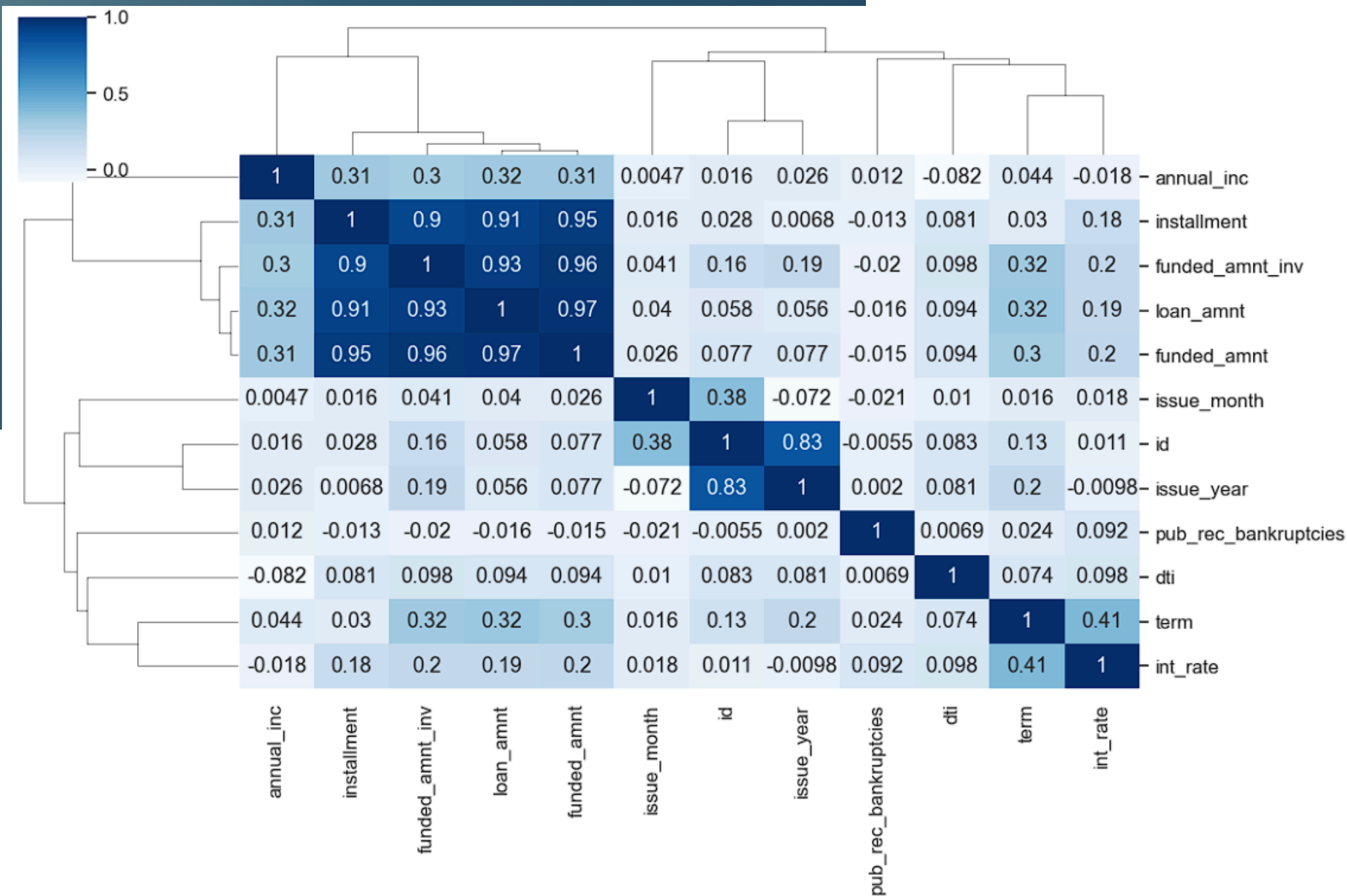


02

Debt-to-Income Ratio v/s Charged-off

- A high Debt-to-Income (DTI) ratio indicates a higher risk of defaults.
- A lower DTI ratio is associated with lower chances of loan defaults.





Co-Relation Analysis

Negative Correlation:

- Loan_amnt has negative correlation with pub_rec_bankruptcies
- Annual income has a negative correlation with dti

Strong Correlation:

- Term has a strong correlation with loan amount
- Term has a strong correlation with interest rate
- Annual income has a strong correlation with loan_amount



Recommendations

- Adjust loan terms and interest rates dynamically based on the borrower's risk profile determined during the application process
 - Implement post-approval monitoring of borrower behavior to detect early signs of financial distress or fraudulent activity.
 - Develop a risk model using the identified attributes, starting with a base interest rate.
 - Adjust the base interest rate by adding additional interest rates to account for delinquency risk as determined by the risk model.
-



Conclusion

- **Interest Rates:** Loans with **interest** rates of **13%-17%** have **higher default** risks. Adjusting rates based on **DTI ratios** may improve repayment alignment.
- **Low Annual Income:** Borrowers earning below ₹40,000 are more likely to default. Consider financial education or income-based loan limits.
- **Loan Grades B, C, and D:** These grades have higher default rates. Stricter risk assessment for these grades is recommended.
- **Term Length: 60-month** loans are more prone to default. **Assess risk** for longer terms and consider limiting maximum term lengths or reduce the interest rate for higher Tenure
- **Debt Consolidation:** This category has **high default** rates. Evaluate **debt consolidation** loans more carefully and consider adjusting interest rates or offering counseling.
- **High Loan Amounts:** Loans over **₹15,000** have higher **default** risks. Conduct thorough assessments and consider capping high-risk loan amounts.
- **Small Business Loans:** **Higher default** rates are seen in **small business** loans. Add more risk assessment parameters for these loans.
- **Public Derogatory Records:** More **derogatory** records increase **bankruptcy risk**. Ensure no public derogatory records before loan approval.

IIIT - B , UpGrad

THANK YOU

In conclusion, our comprehensive analysis of the Lending Club dataset has provided valuable insights into the key factors affecting loan performance. By leveraging these insights, we can develop more accurate risk models and make data-driven decisions to improve loan approval processes.
