upGrad

LENDING CLUB CASE STUDY

Course: Executive PG Programme in Machine Learning & AI - August 2023

Group Facilitator	Team Member
Name: Narendra Jha **Email ID**: narendra.jha@live.com **Phone no:** 9990227116	**Name:** Mohammad Monis **Email ID:** javedmonis07@gmail.com **Phone no:** 9642234275

TABLE OF CONTENTS

GENERAL INFORMATION

- Lending club case study is analysis of various factors on loan repayment. The insights of this analysis will help banks to identify if a person is likely to replay loan or not.
- The idea behind implementing this project is to understand how real business problems are solved using EDA. Apart from applying EDA techniques, we also learnt about risk analytics in banking and financial services.
- Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'. 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.
- Loan data set contains the complete loan data for all loans issued through the time period 2007 to 2011.

DRIVING FACTORS FOR LOAN DEFAULT RATE

Here are the key factors influencing the default rate. We will address each of them individually, examining their impact one by one.

- Loan Amount
- Loan Term
- Interest Rate
- Annual Income
- Home Ownership
- Loan Purpose
- Debt to Income Ratio
- Job Grade
- Address State
- Work Experience
- Loan Issue Date

upGrad

AFFECT OF ANNUAL INCOME ON LOAN DEFAULT RATE

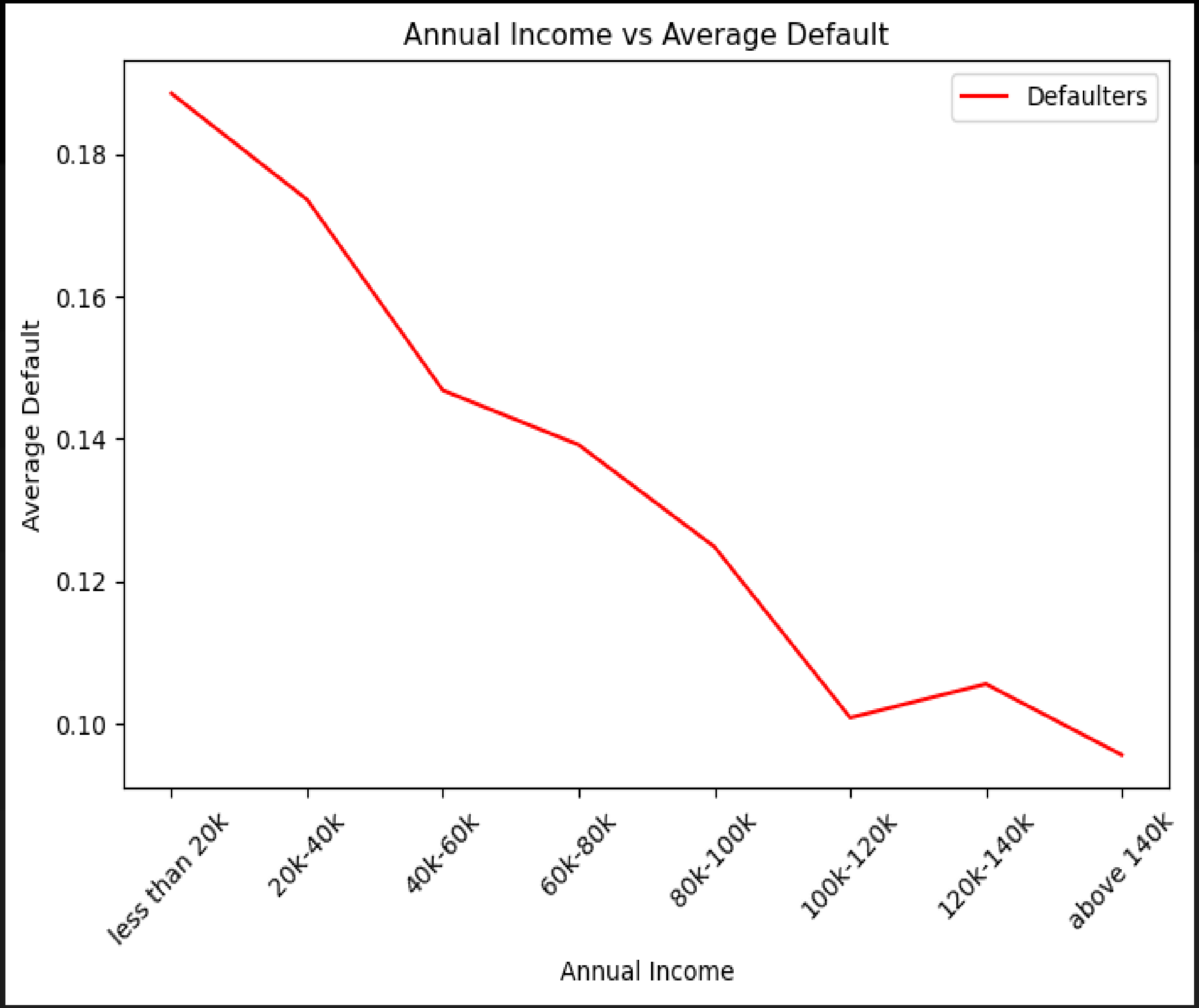
For studying the impact of annual income on loan default rate, we removed incomes which were not made by most people from the data set so that our analysis does not get skewed.

Incomes were clubbed into ranges of 20k i.e. ‘20k-40k’, ‘40k-60k’ and so on.

Defaulters were grouped by income range and the following trend was observed.

Conclusion: The loan default rate goes down with the increase in income.

Recommendation: This analysis suggests that giving loans to people with decent annual incomes is safer than people with low incomes.



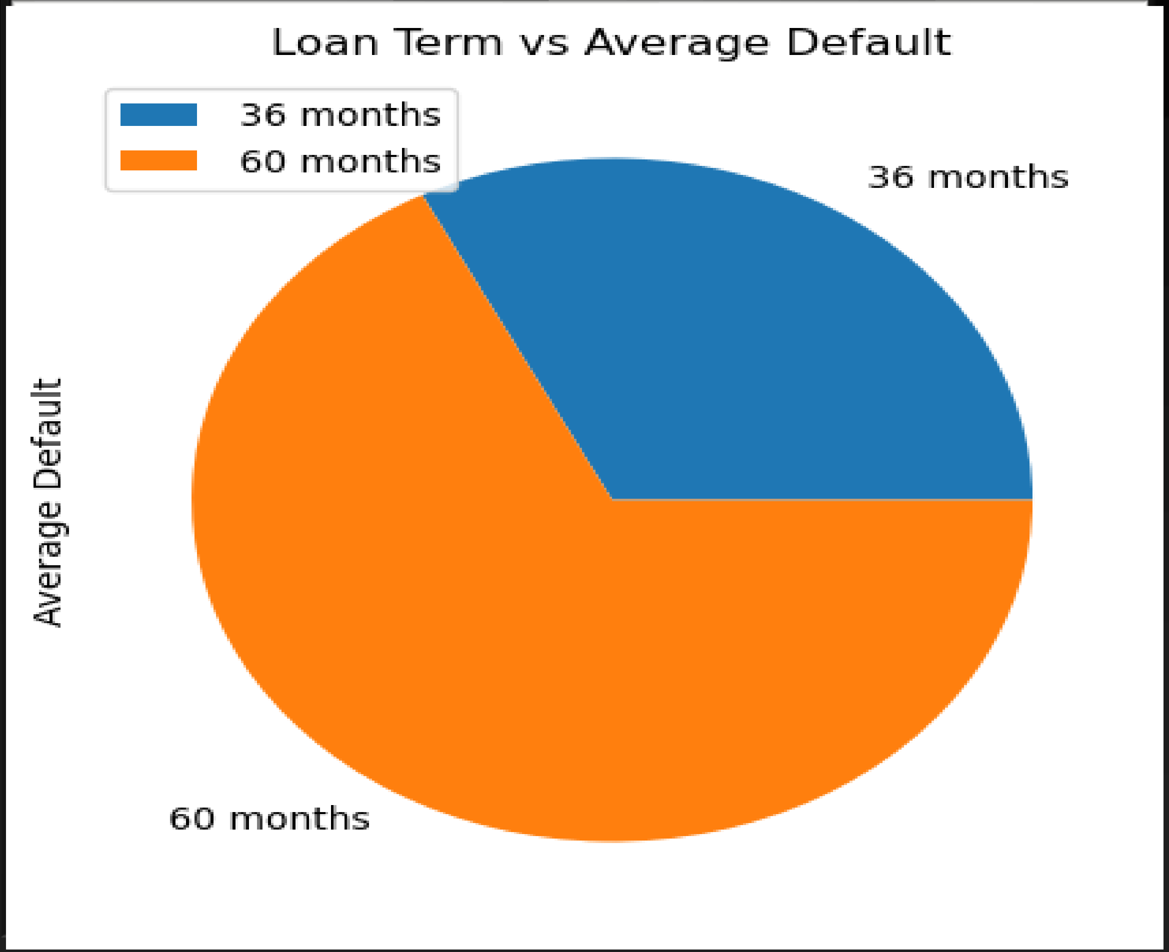
upGrad

AFFECT OF LOAN TERM ON DEFAULT RATE

To study the impact of the type of loan term i.e. 36 or 60 months, on default rate, we grouped defaulters by loan term and plotted a pie chart shown alongside.

Conclusion: The chart suggests that people who take loans for longer periods default more than people who take loans for short periods.

Recommendation: It is safer to give a loan for a short period than for a long period.



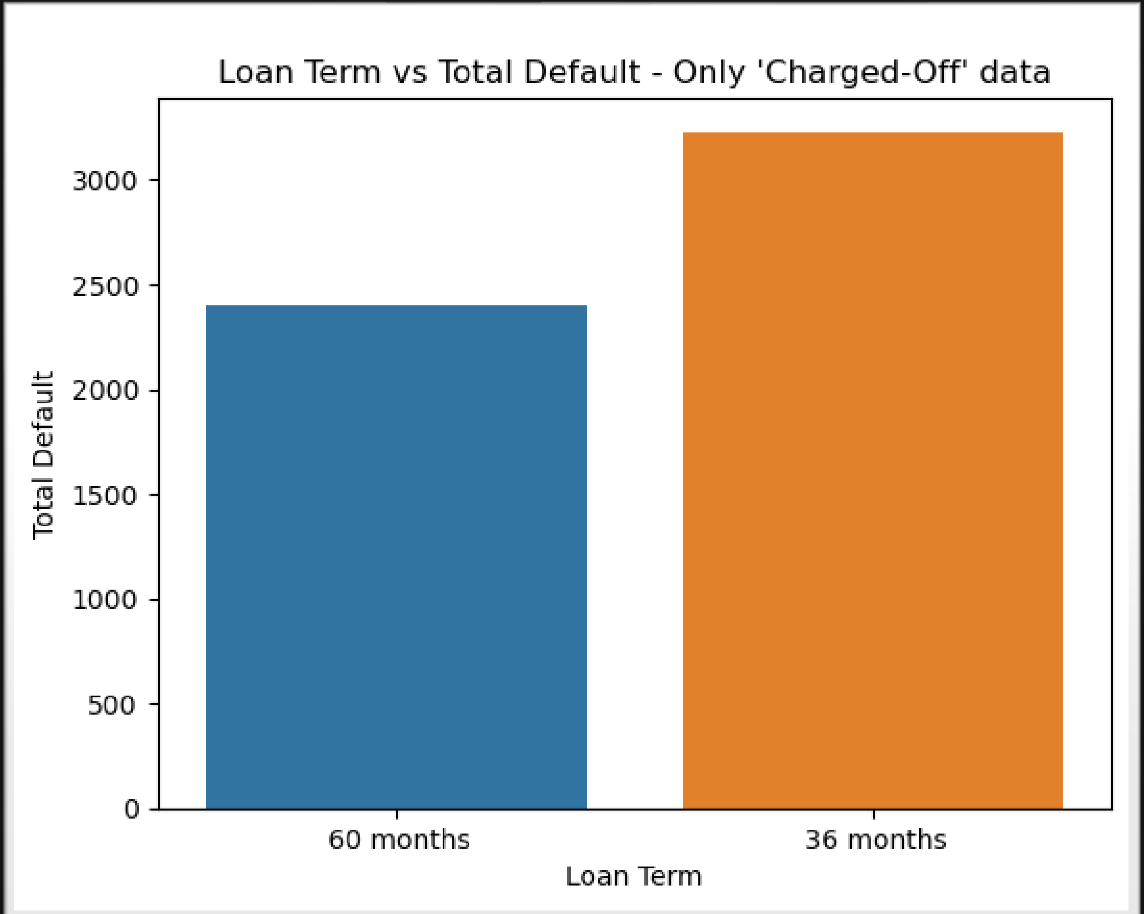
upGrad

AFFECT OF LOAN TERM ON DEFAULT RATE – *CHARGED OFF DATA*

To study the impact of type of loan term i.e. 36 or 60 months, on default rate we counted total defaulters by loan term and plotted a pie chart shown alongside. We considered only **'Charged-Off'** data to study it.

Conclusion: When we studied only 'Charged-Off' data then found that a significant portion of defaulters opted for 36 **months** loan term.

Recommendation: It is safer to give a loan for a short period than for long period.



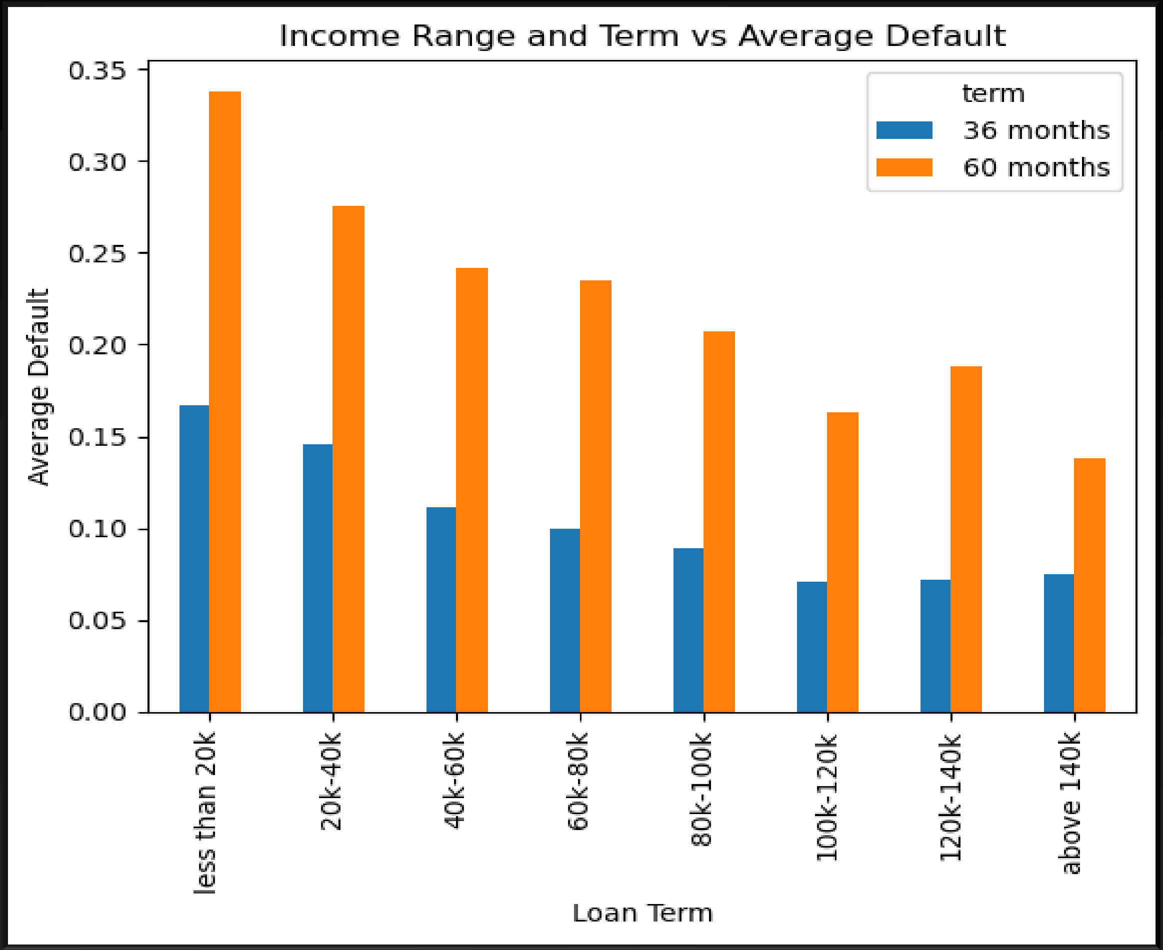
upGrad

AFFECT OF INCOME AND TERM ON DEFAULT RATE

Individuals To read the effect of income and terms on default rate we plotted a heatmap.

Conclusion: Individuals with lower annual income and longer loan tenures are more likely to default compared to those with higher annual income and shorter loan tenures.

Recommendation: It is safe to give loans to those with higher annual income and short periods.

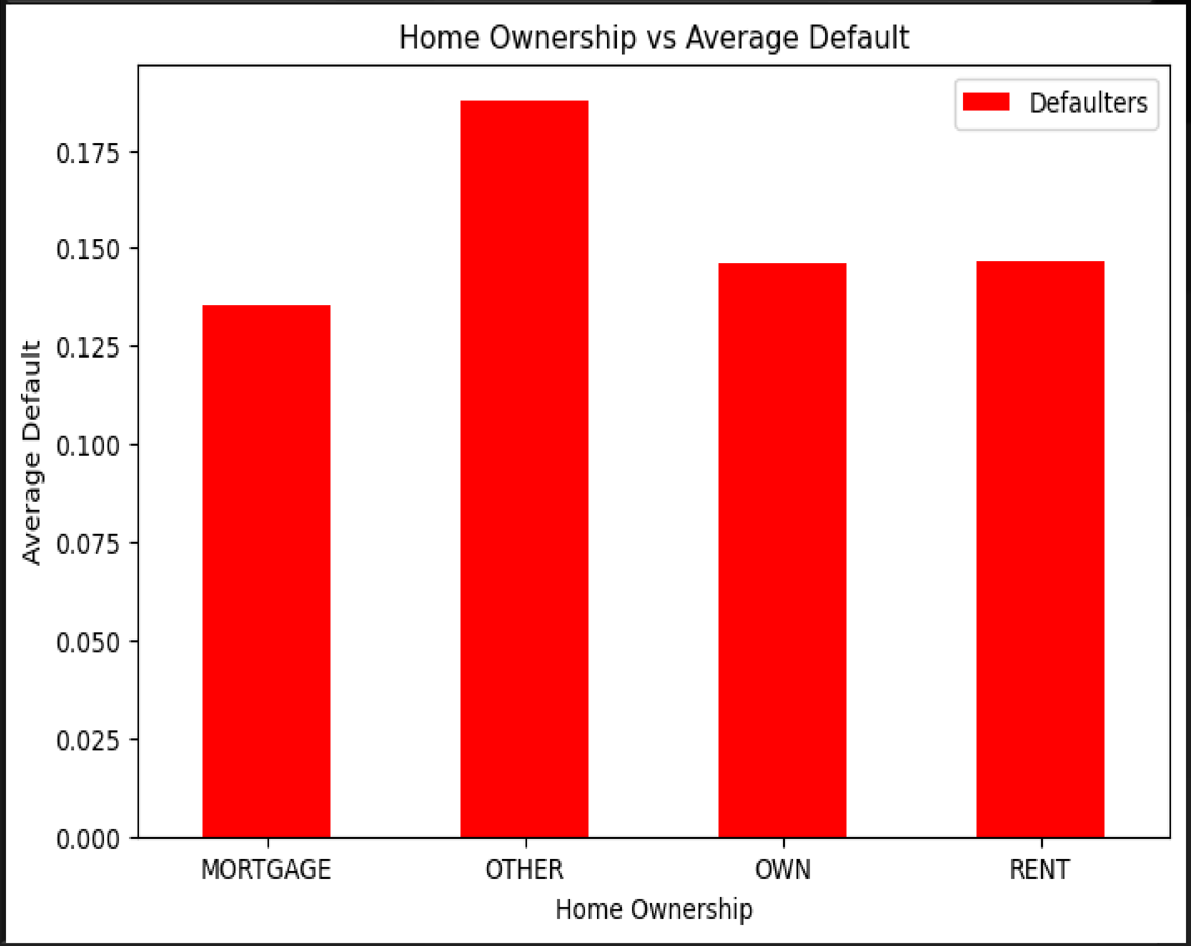


AFFECT OF HOME OWNERSHIP ON DEFAULT RATE

To study affect of home ownership i.e. own house, rented house, mortgaged house on loan default rate we removed home ownership 'NONE' from dataset because it seems that homeownership information was not available with banks for those records and then we grouped defaulters by home ownership and plotted a bar graph as shown along side.

Conclusion: There is no clear correlation between home ownership and default rate.

Recommendation: Homeownership status has no significant affect on default rate so it should not be considered as an important factor while lending out loans.

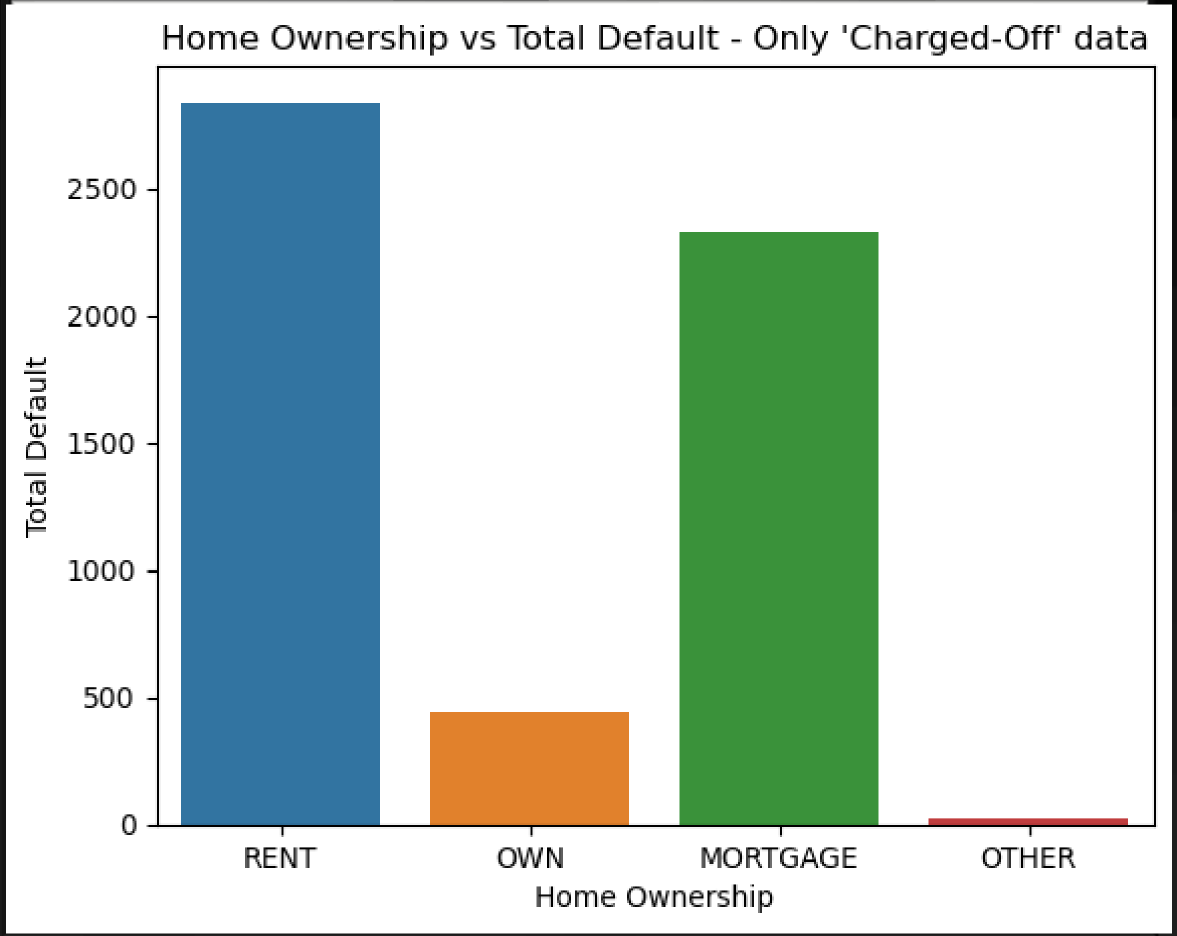


AFFECT OF HOME OWNERSHIP ON DEFAULT RATE – *CHARGED OFF DATA*

To study the affect of home ownership i.e. own house, rented house, mortgaged house on loan default rate we removed home ownership 'NONE' from the dataset because it seems that homeownership information was not available with banks for those records and then we counted total defaulters by home ownership and plotted a bar graph as shown along side. We considered only **'Charged-Off'** data to study it.

Conclusion: When we studied only 'Charged-Off' data then found that a significant portion of individuals who have defaulted on their payments reside in rented accommodations.

Recommendation: Banks need to exercise caution when granting loans to individuals residing in rented accommodations.

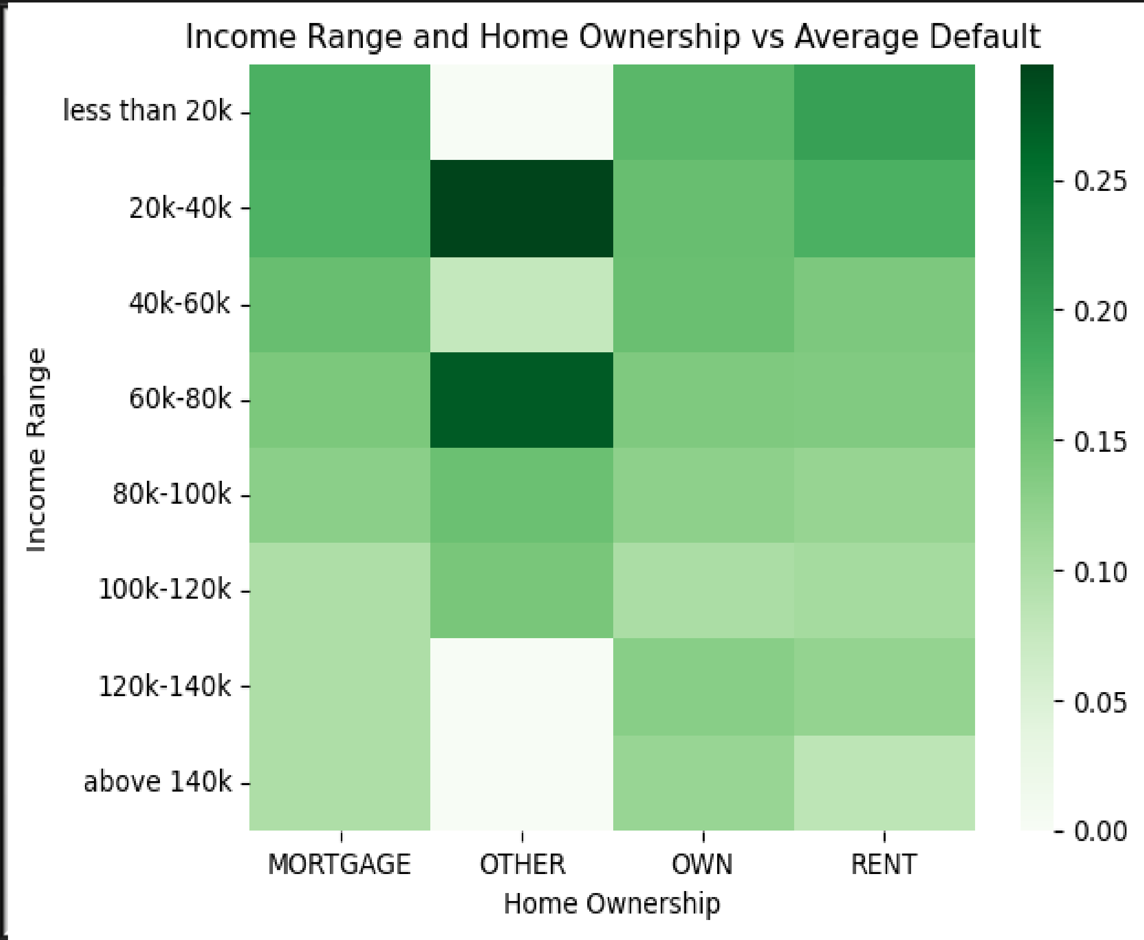


AFFECT OF INCOME AND HOME OWNERSHIP ON DEFAULT RATE

Individuals To read the effect of income and home ownership on default rate we plotted a heatmap.

Conclusion: There is no clear correlation between homeownership and annual income.

Recommendation: Homeownership status has no significant affect on default rate so it should not be considered as an important factor while lending out loans.

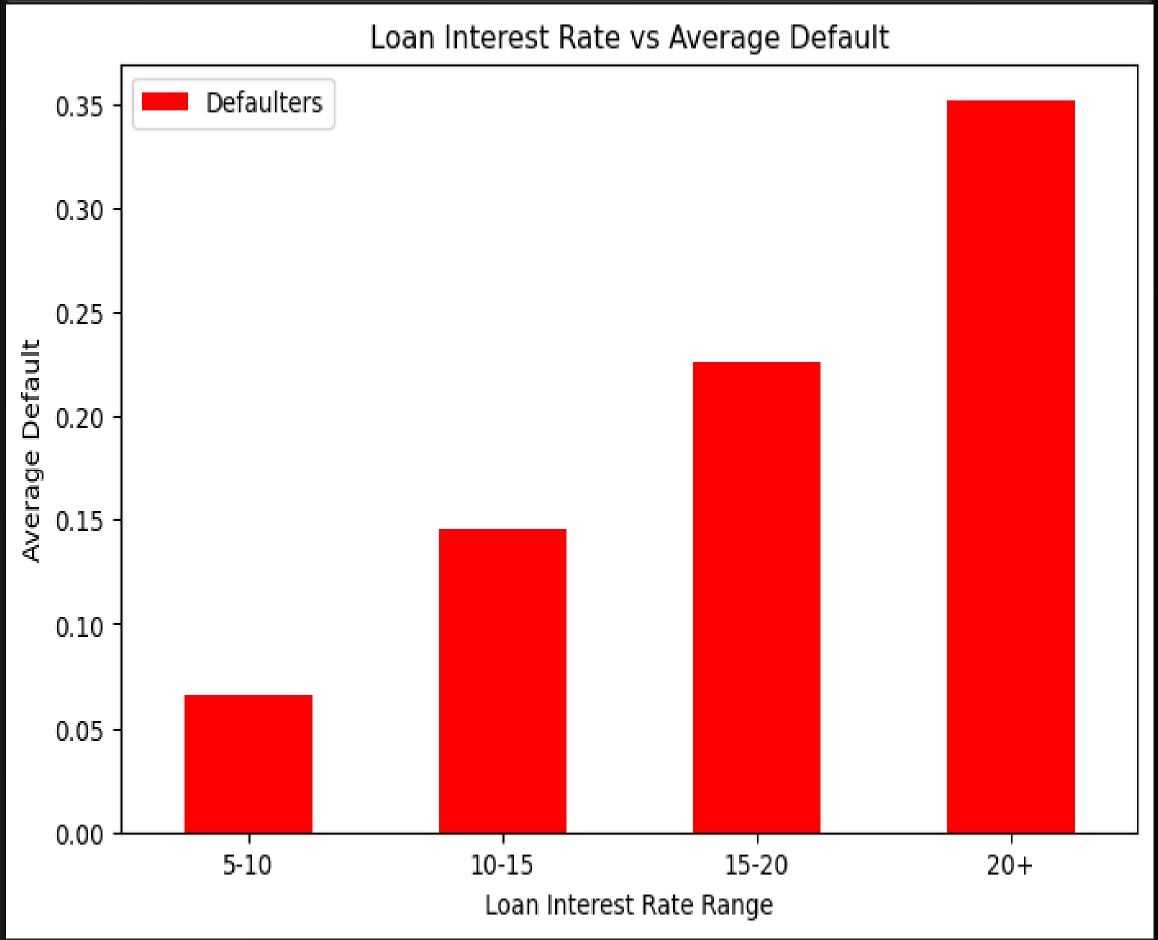


AFFECT OF INTEREST RATE ON LOAN DEFAULT RATE

To study affect of interest rate on loan default rate we divided interest rate into ranges i.e. 5-10%, 10-15%, 15-20% and 20+%. Then we grouped defaulters by interest range and plotted the bar graph shown alongside.

Conclusion: The default rate is proportional to the rate of interest.

Recommendation: It is safe to provide loans on a lower rate of interest than on a higher rate of interest.



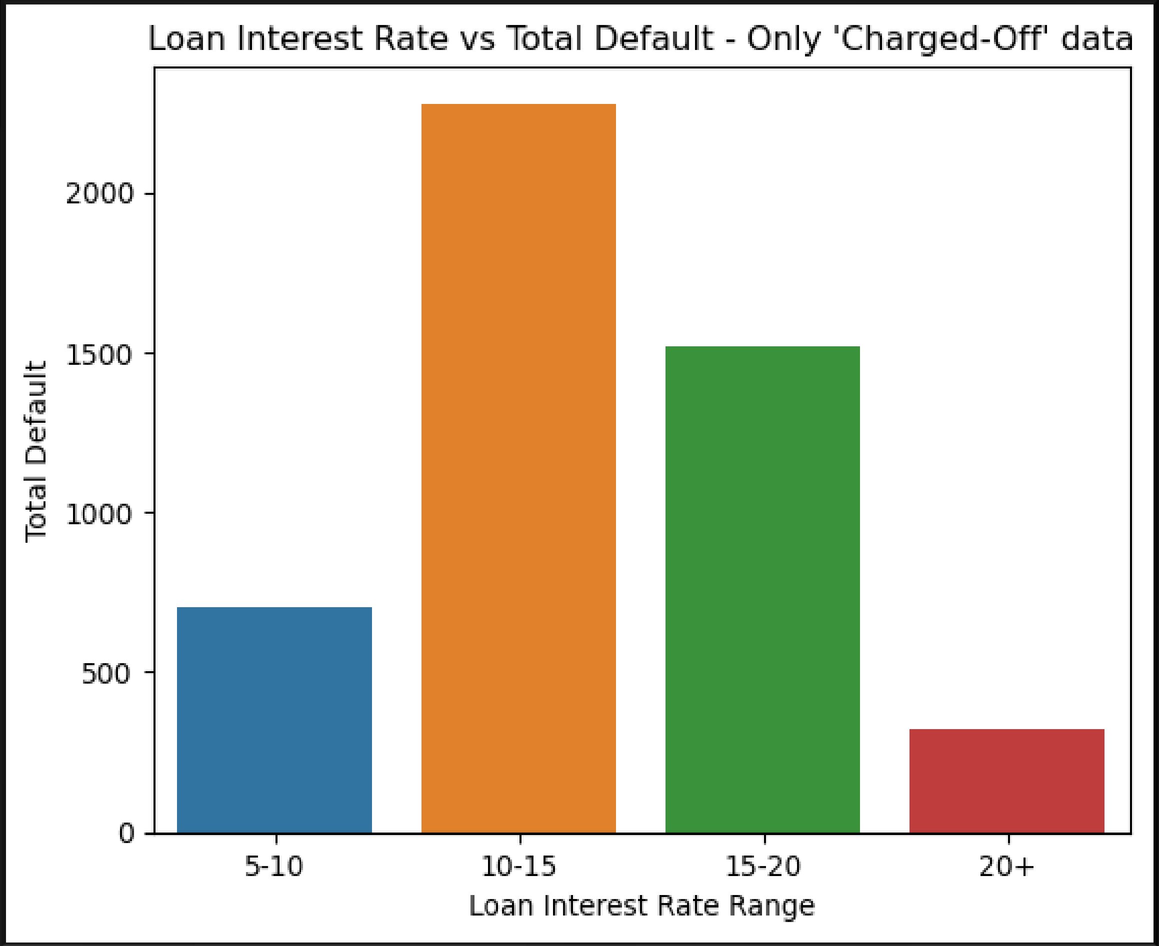
upGrad

AFFECT OF INTEREST RATE ON LOAN DEFAULT RATE – *CHARGED OFF DATA*

To study affect of interest rate on loan default rate we divided interest rate into ranges i.e. 5-10%, 10-15%, 15-20% and 20+%. Then we counted total defaulters by interest range and plotted bar graph shown alongside. We considered only **'Charged-Off'** data to study it.

Conclusion: When analysing exclusively the 'Charged-Off' data, we discovered that a notable number of people who failed to meet their payment obligations had taken out loans with a ratio ranging from **10-15%**.

Recommendation: Offering loans at both lower and higher interest rates is more secure than providing loans at moderate interest rates.



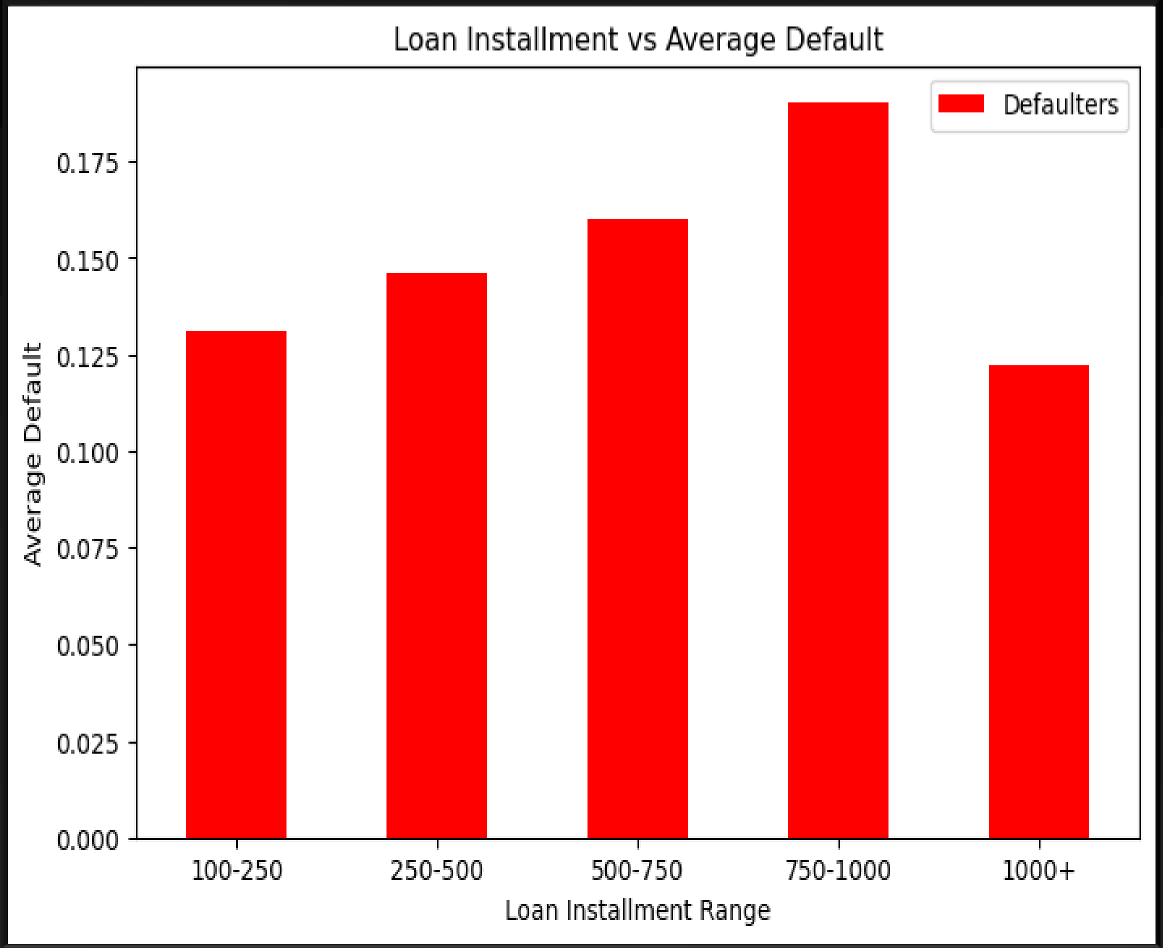
upGrad

AFFECT OF LOAN INSTALLMENT ON LOAN DEFAULT RATE

To study affect of Loan Installment on loan default rate we divided Loan Installment into ranges i.e. 100-250, 250-500, 500-750, 750-1000 and 1000+’. Then we grouped defaulters by Loan Installment range and plotted the bar graph shown alongside.

Conclusion: Loans with instalment amounts between **500 and 750** exhibit the highest rate of defaults. However, this information doesn't provide any useful insights on whether to approve or deny a loan application.

Recommendation: Loan Installment has no significant affect on default rate so it should not be considered as an important factor while lending out loans.

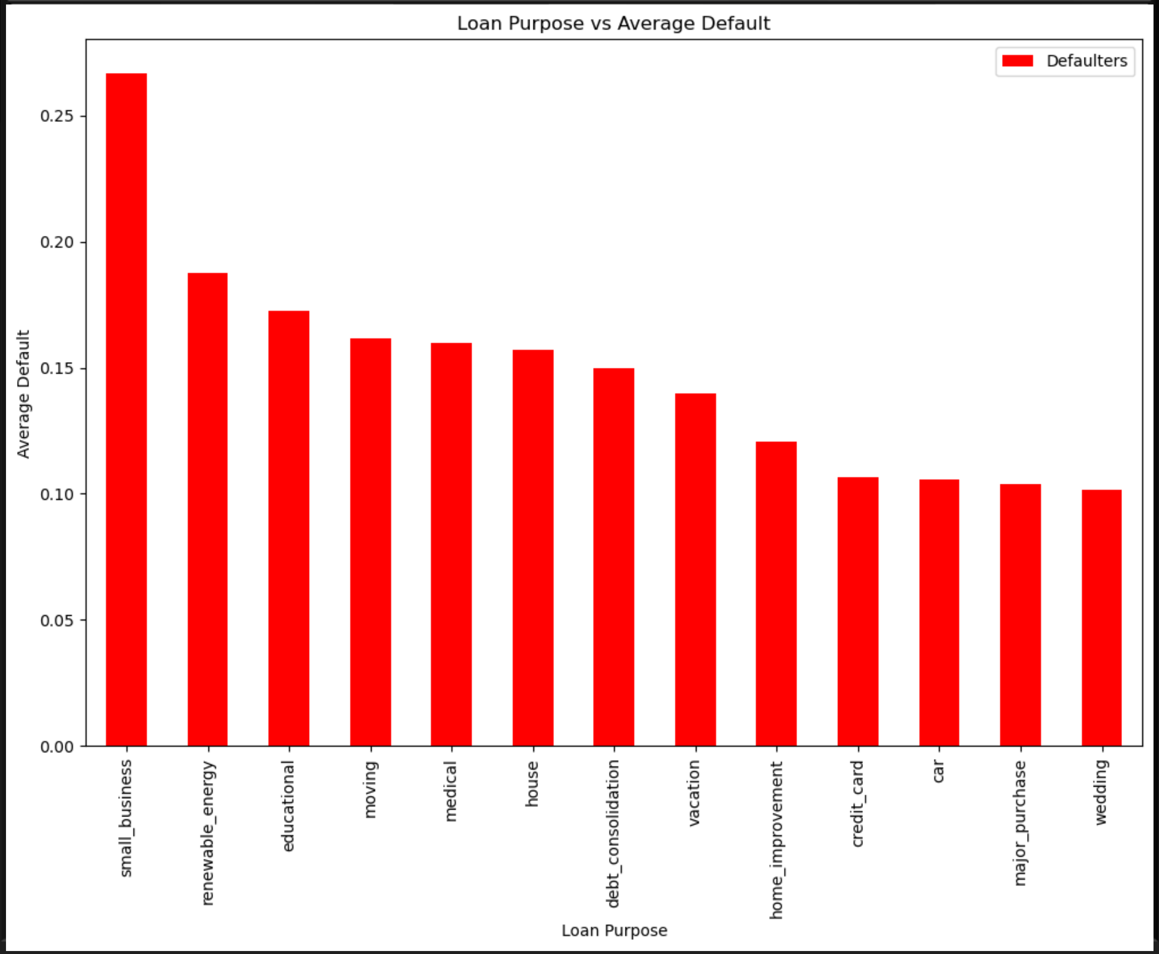


AFFECT OF LOAN PURPOSE ON LOAN DEFAULT RATE

We grouped default rates by the purpose of the loan and found results as shown alongside.

Conclusions: Loan taken for the purpose of **small business** has the highest default rate and loan taken for a **wedding** has the lowest default rate.

Recommendations: It is very risky to give loans to fund small businesses.

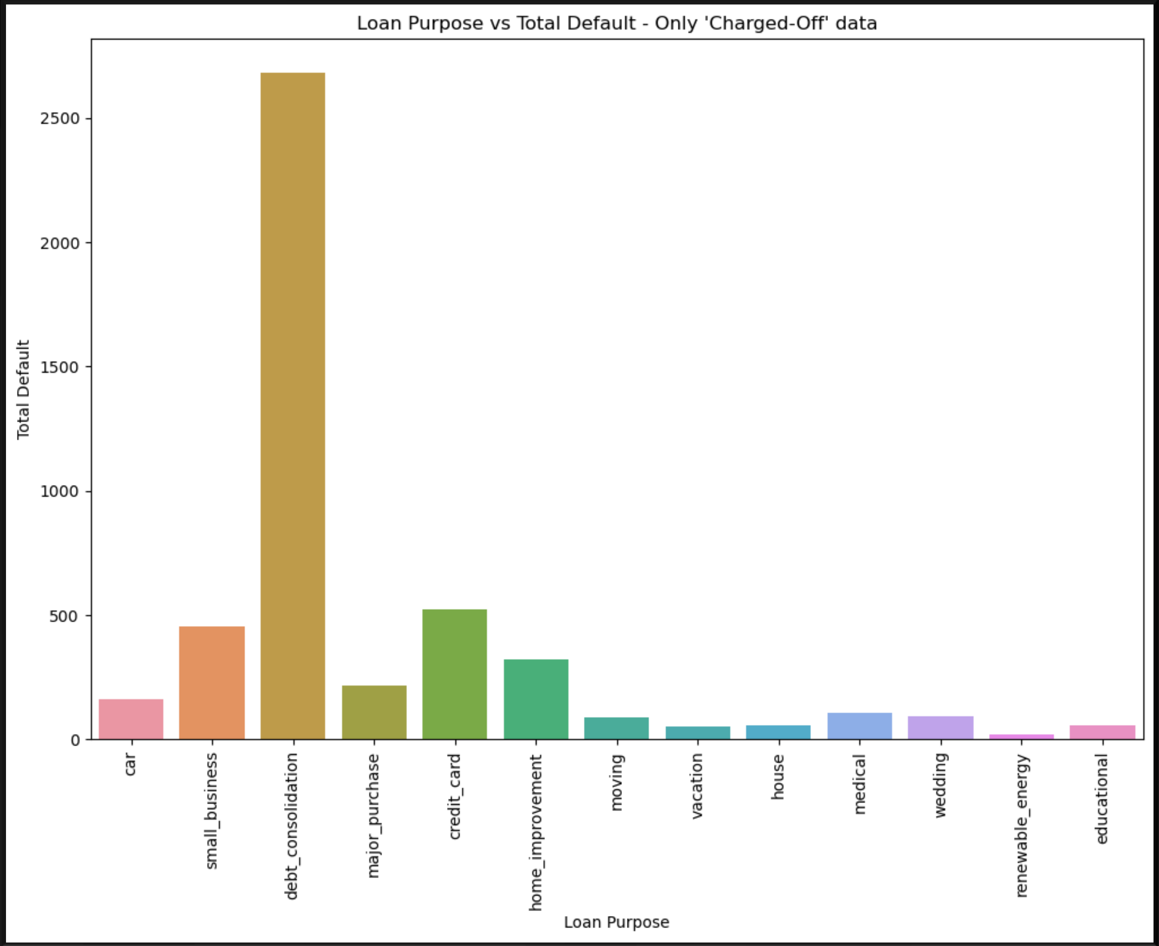


AFFECT OF LOAN PURPOSE ON DEFAULT RATE – *CHARGED OFF DATA*

We counted total default rates by the purpose of the loan and found results as shown alongside. We considered only **'Charged-Off'** data to study it.

Conclusions: When analyzing exclusively the 'Charged-Off' data, the default rate is highest for loans acquired for **Debt Consolidation** and lowest for those taken for **Renewable energy**.

Recommendations: Providing loans to support Debt Consolidation entails significant risks.

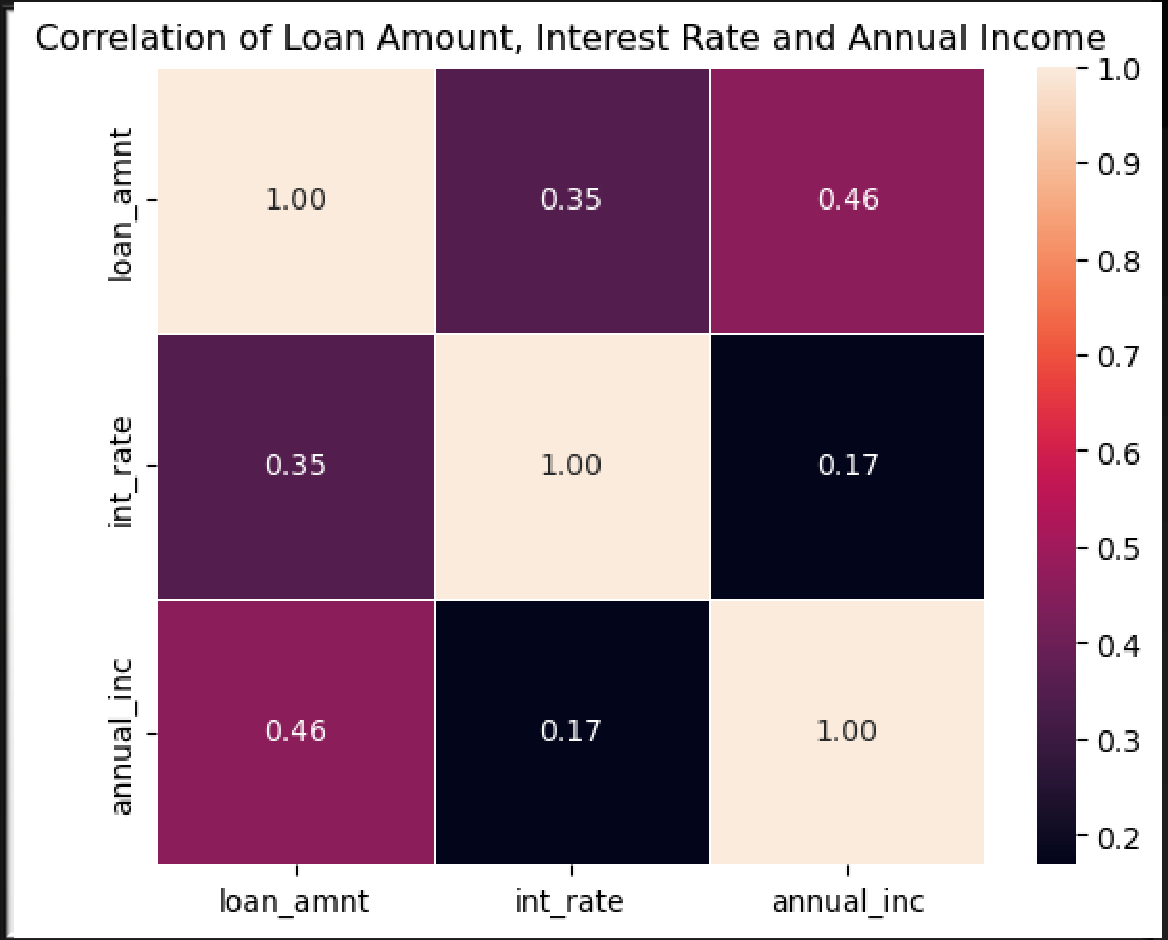


upGrad

CORRELATION OF LOAN AMOUNT, INTEREST RATE AND ANNUAL INCOME

Plotted heatmap to find a correlation of Loan Amount, Interest Rate and Annual Income.

Conclusions: This plot shows that the **loan_amnt**, **int_rate** and **annual_inc** are related together ie have a **positive correlation**.



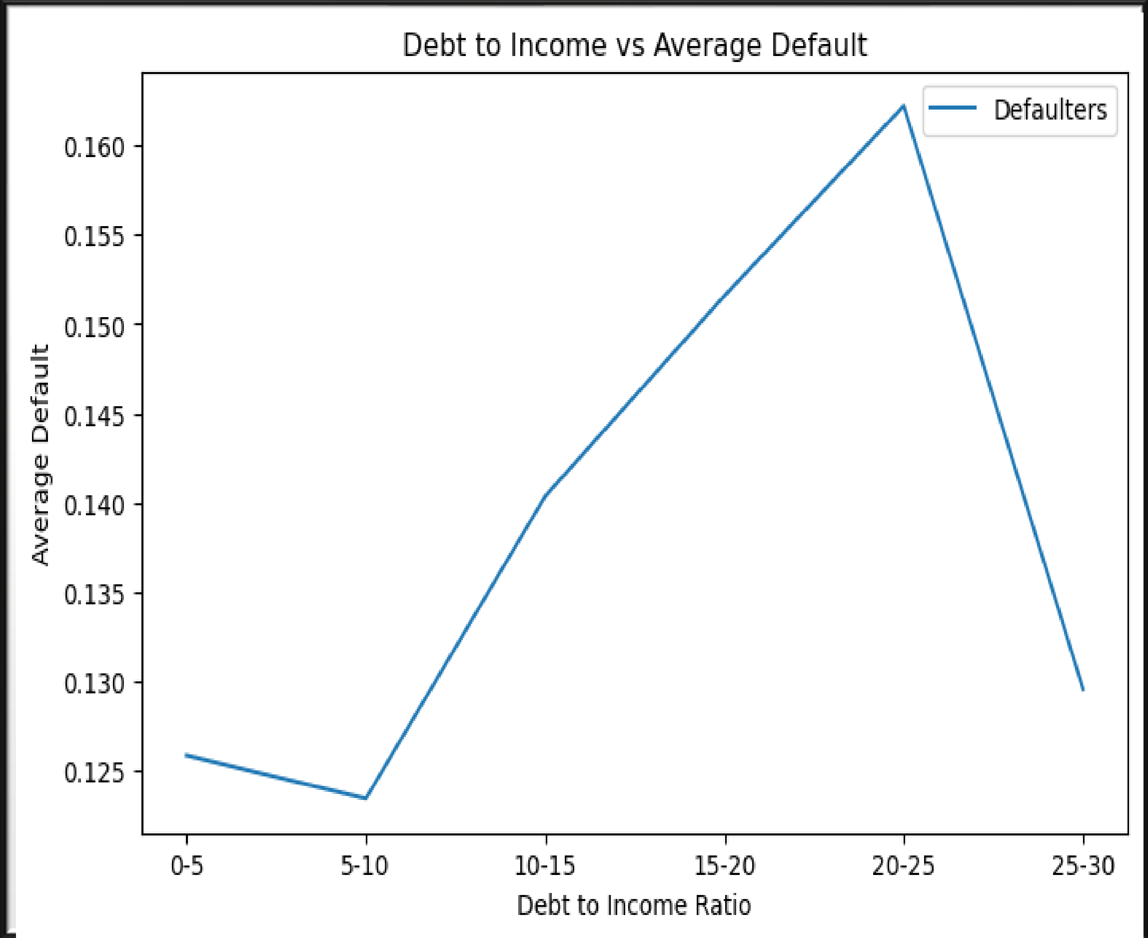
AFFECT OF DEBT TO INCOME RATIO ON LOAN DEFAULT RATE

People in loan dataset were having debt to income ratio between 0 to 30. We made 6 ranges of debt to income ratio i.e 0-5, 5-10, 10-15, 15-20, 20-25, 25-30 and we categorized each person in dataset by range of debt to income ratio they fall in.

On grouping defaulters by debt to income ratio ranges we get observation as shown alongside.

Conclusion: Default rate increases with an increase in debt to income ratio until 20-25 range than drops for 25-30 range.

Recommendation: It is safe to give loans to people with low debt to income ratio.

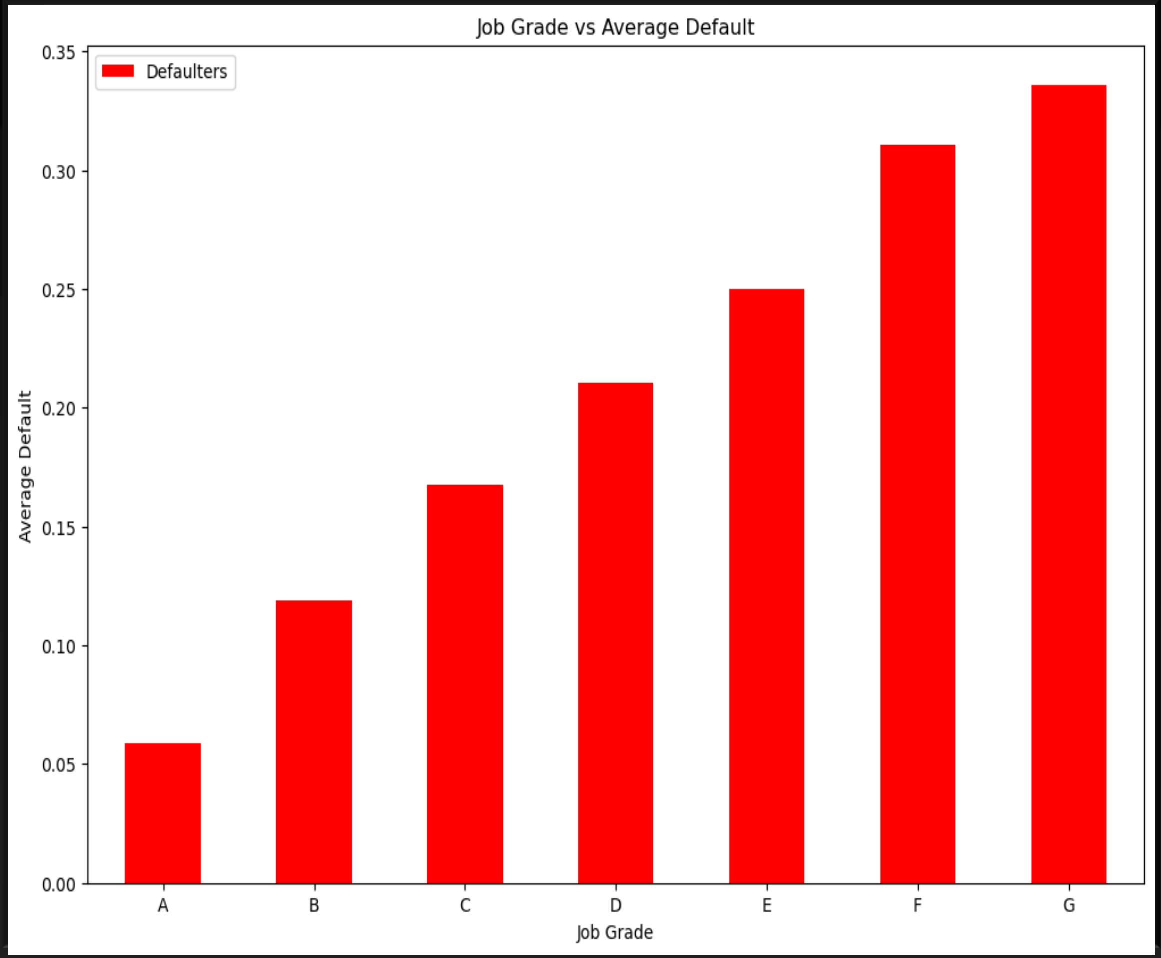


AFFECT OF JOB GRADE ON DEFAULT RATE

We grouped default rates by the job grade and found results as shown alongside.

Conclusions: The chart shows that the majority of individuals who defaulted on their payments belong to job **grade G**.

Recommendations: Granting loans to individuals with a job **grade G** might involve more risks.

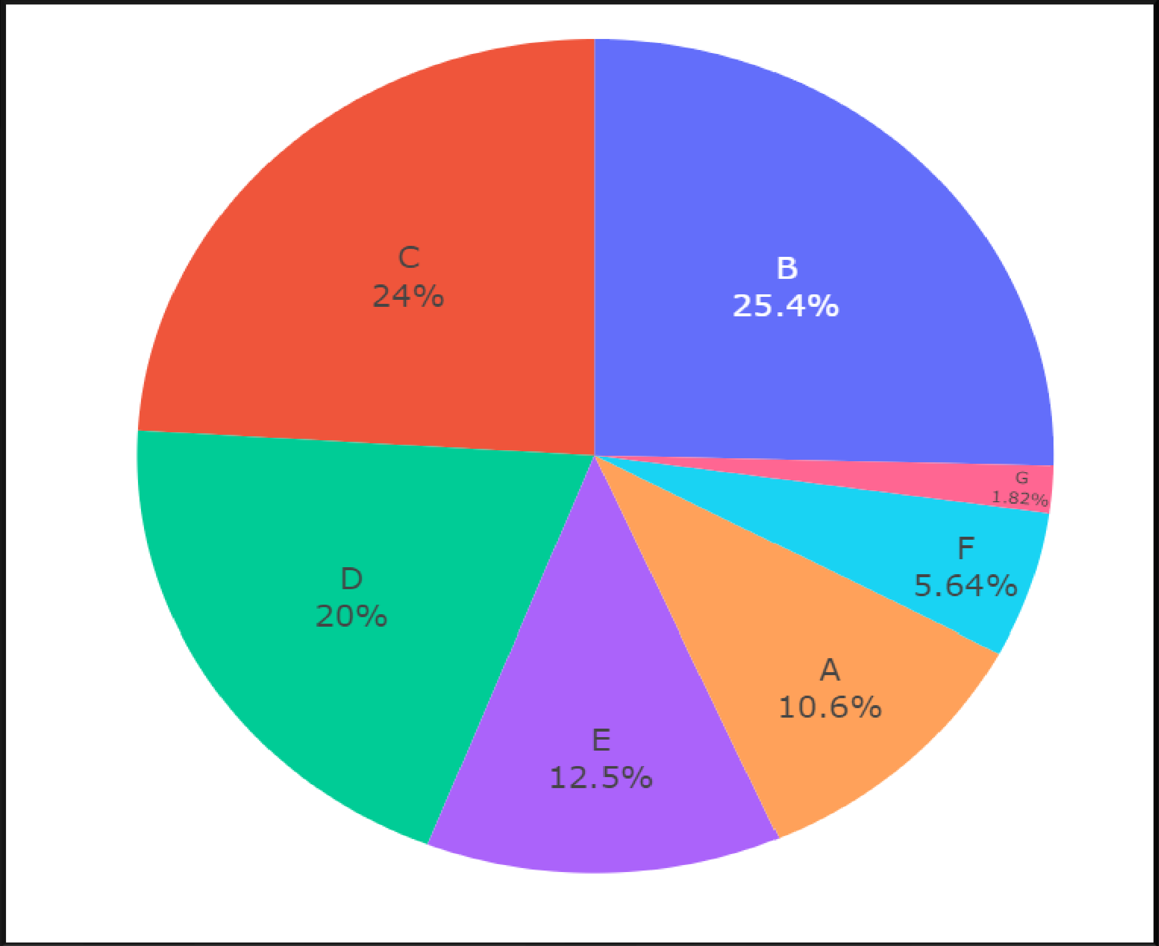


AFFECT OF JOB GRADE ON DEFAULT RATE – *CHARGED OFF DATA*

We counted default rates by the job grade and found results as shown alongside. We considered only **'Charged-Off'** data to study it.

Conclusions: However, Upon examining only the 'Charged-Off' data, we found that a small proportion of individuals who defaulted on their payments belong to job **grade G**.

Recommendations: Granting loans to individuals with a job **grade G** might involve less risks.

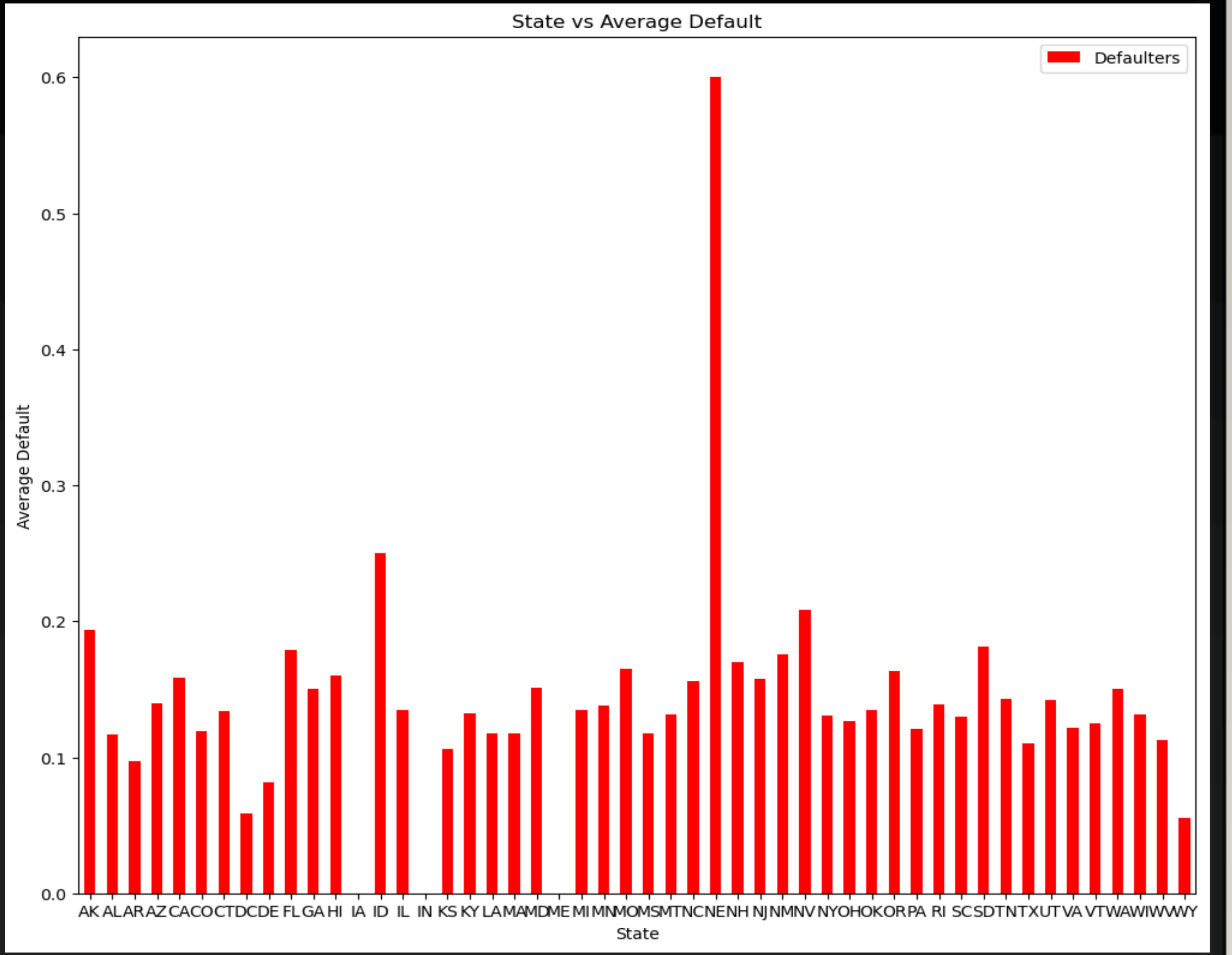


AFFECT OF STATE ON DEFAULT RATE

We grouped default rates by the state and found results as shown alongside.

Conclusions: The above chart shows that the majority of individuals who defaulted on their payments belong to the state **NE**.

Recommendations: Granting loans to individuals who belong to state **NE** might involve more risks.

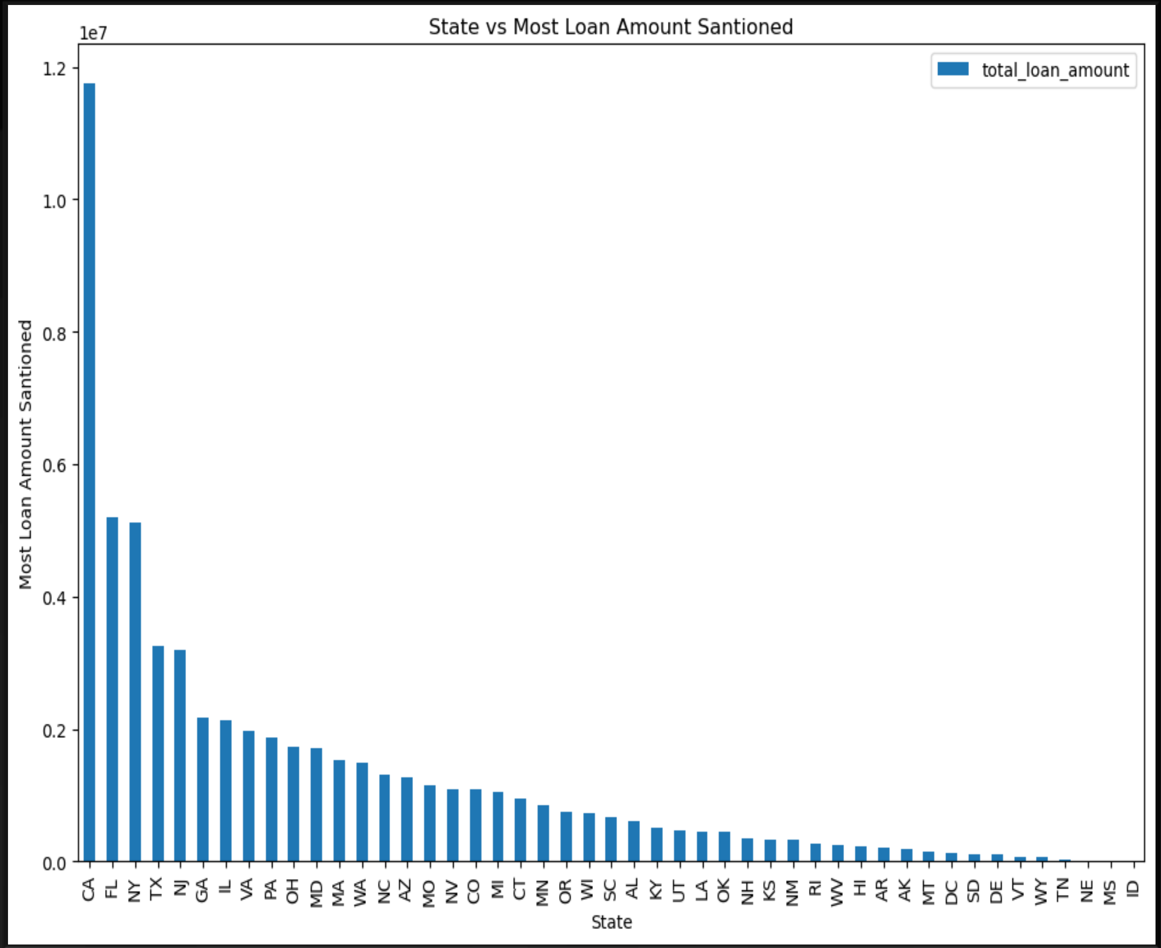


STATE VS MOST LOAN AMOUNT SANCTIONED

We sum the total loan amount sanctioned to each state and found results as shown alongside.

Conclusions: The chart above indicates that most people who defaulted took out larger loans in the state of **CA**.

Recommendations: Lending loans in the state of CA necessitates increased precautions.



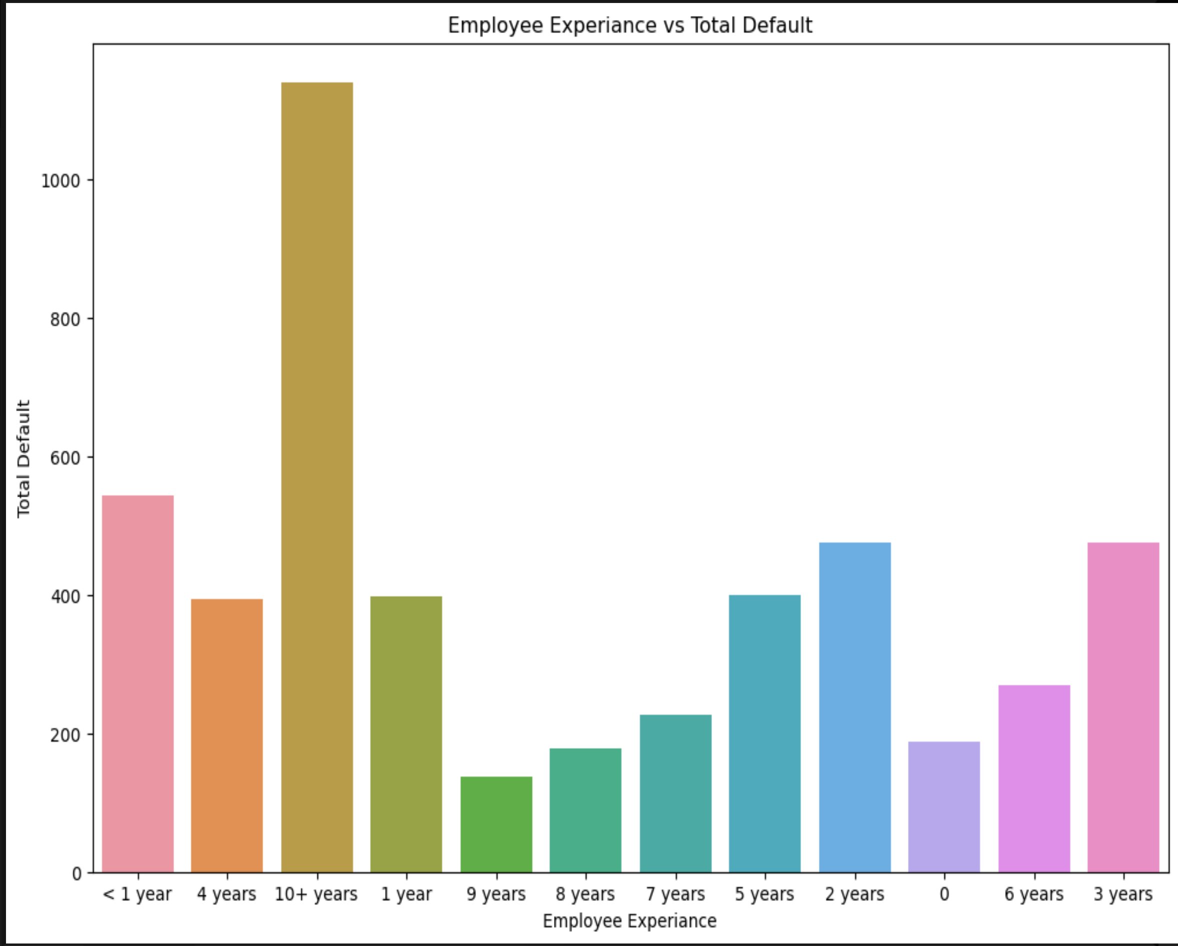
upGrad

AFFECT OF EMPLOYEE EXPERIENCE ON DEFAULT RATE

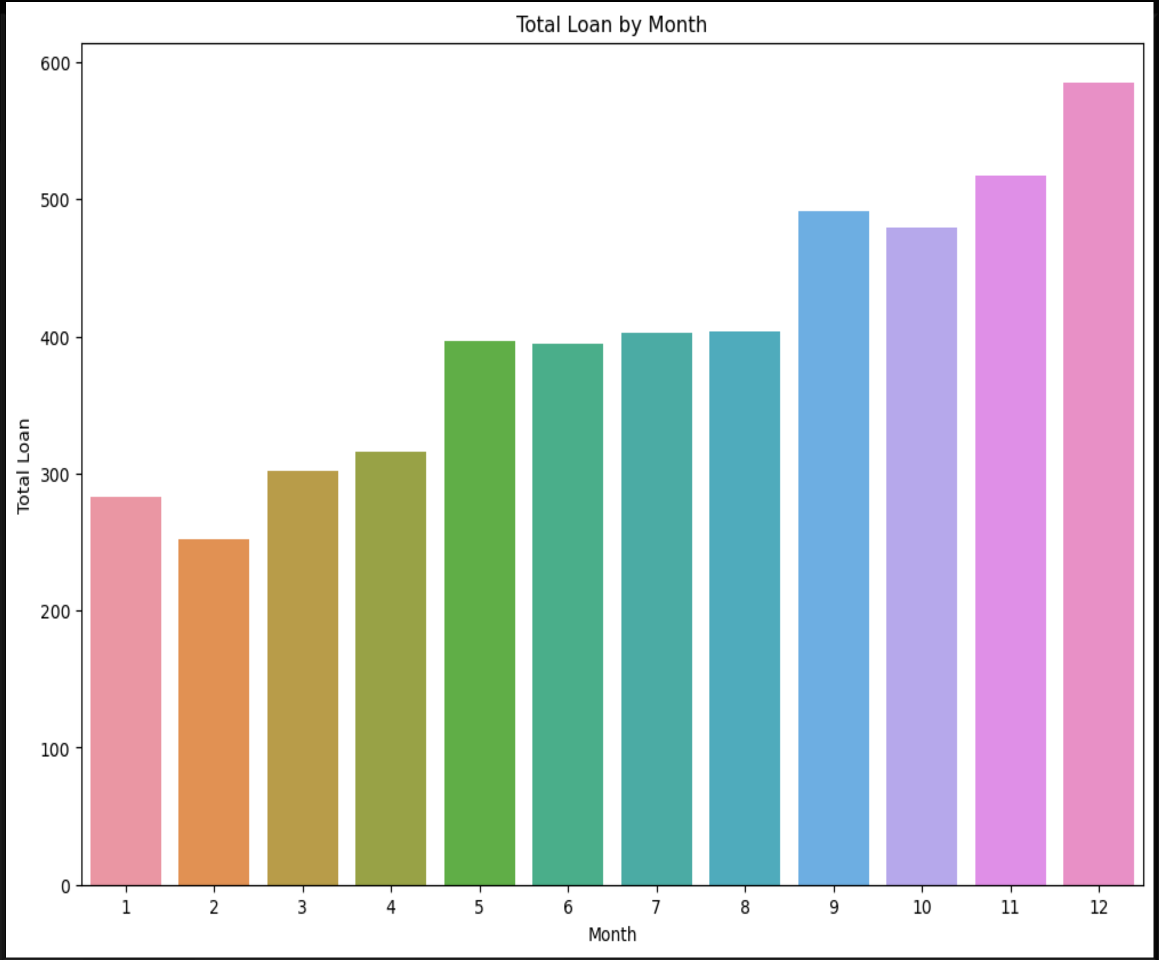
We counted total default by the employee's experience and found results as shown alongside.

Conclusions: The chart above shows that most people who failed to make payments have more than **10+ years** of experience.

Recommendations: Providing loans to individuals with more experience could potentially increase the associated risks.



TOTAL LOAN SANCTIONED BY MONTH



We separated month and year from the loan issue date column and then counted the total loans sanctioned in each month and found results as shown alongside.

Conclusions: The above chart shows that a majority of defaulted loans were approved in the month of **December**.

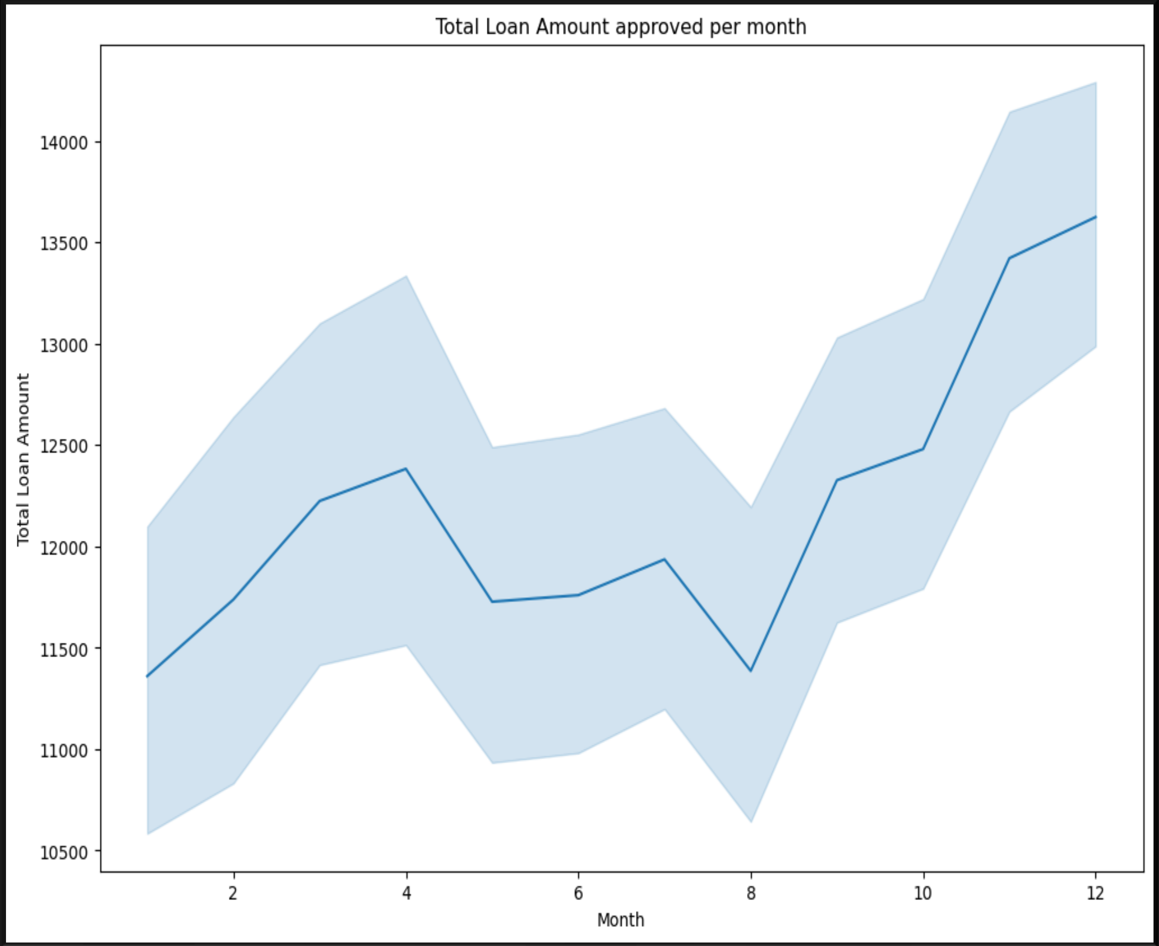
Recommendations: The decrease in application scrutiny at the year-end might be attributed to financial constraints. This situation can be enhanced through improvements.

TOTAL LOAN SANCTIONED BY MONTH

We separated month and year from the loan issue date column and then counted the total loans sanctioned in each month and found results as shown alongside.

Conclusions: The above chart shows that a majority of defaulted loans were approved in the month of **December**.

Recommendations: The decrease in application scrutiny at the year-end might be attributed to financial constraints. This situation can be enhanced through improvements.

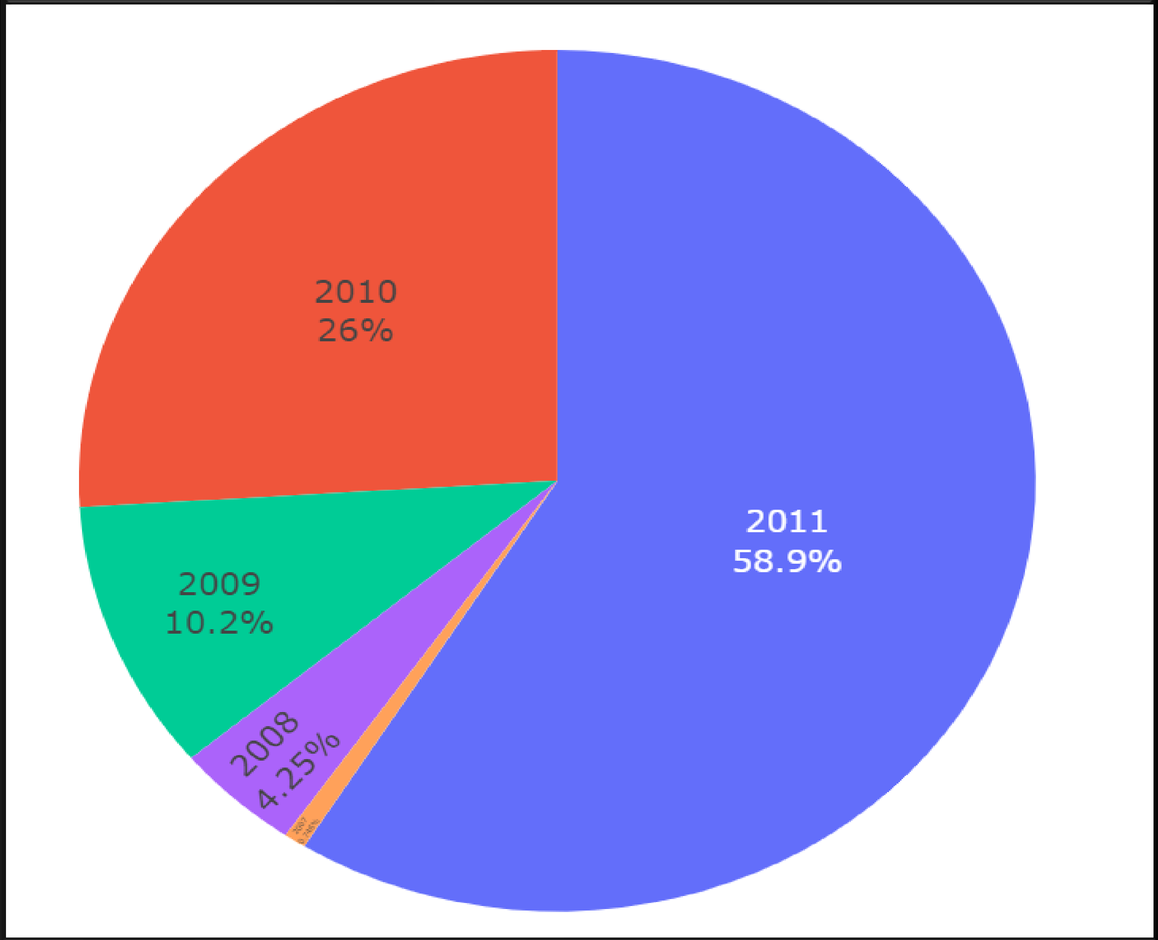


upGrad

TOTAL LOAN SANCTIONED BY YEAR

We separated month and year from the loan issue date column and then counted the total loans sanctioned in each year and found results as shown alongside.

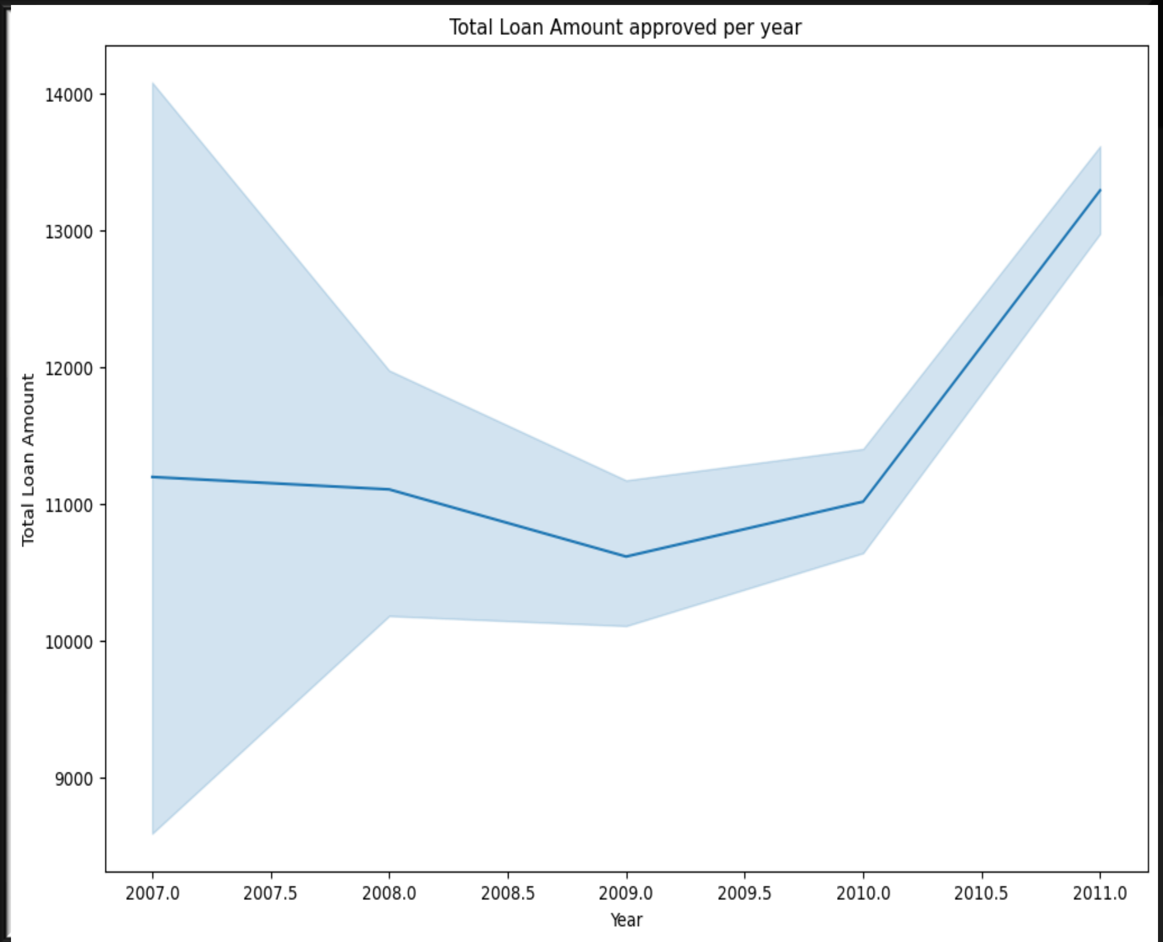
Conclusions: The above chart shows that a majority of defaulted loans were approved in the year of **2011**.



TOTAL LOAN SANCTIONED BY YEAR

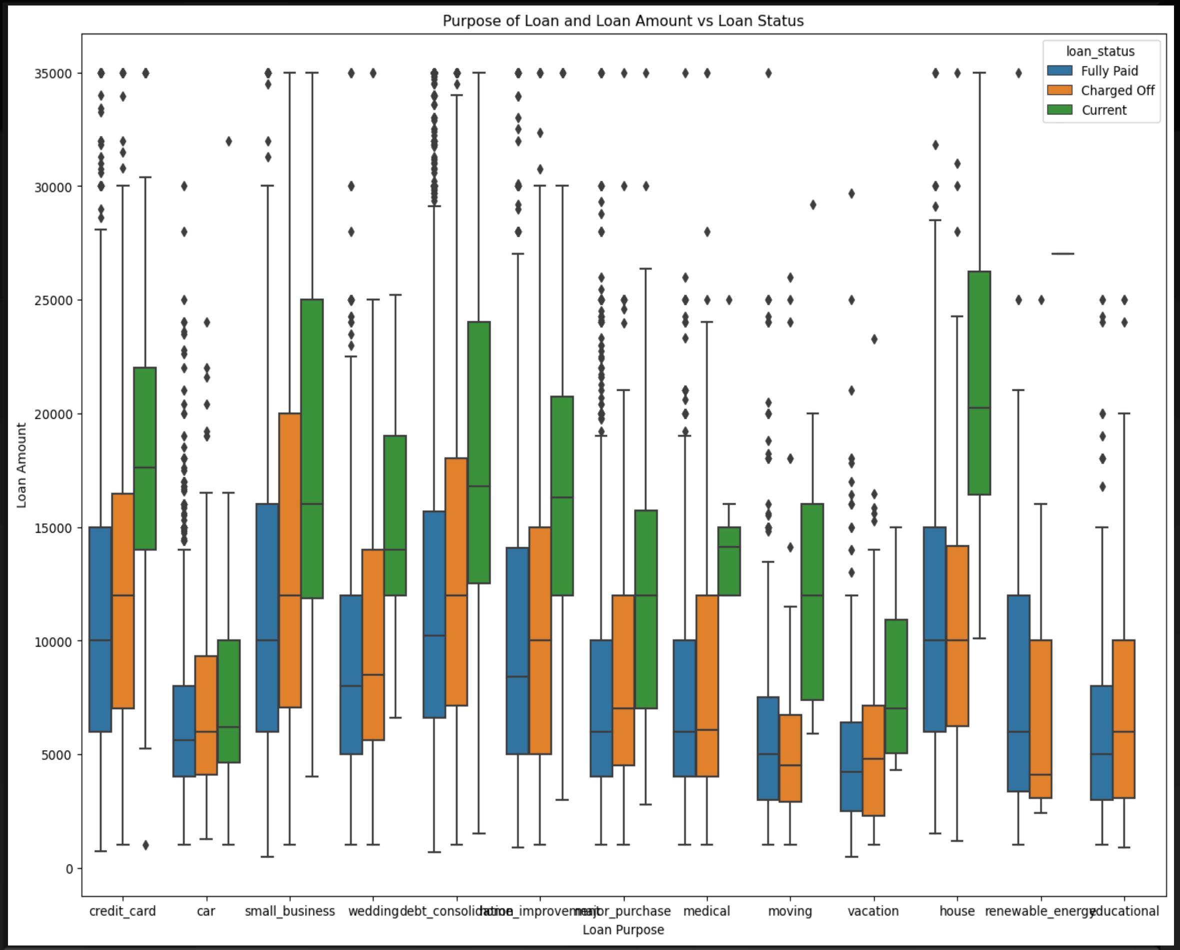
We separated month and year from the loan issue date column and then counted the total loans sanctioned in each year and found results as shown alongside.

Conclusions: The above chart shows that a majority of defaulted loans were approved in the year of **2011**.



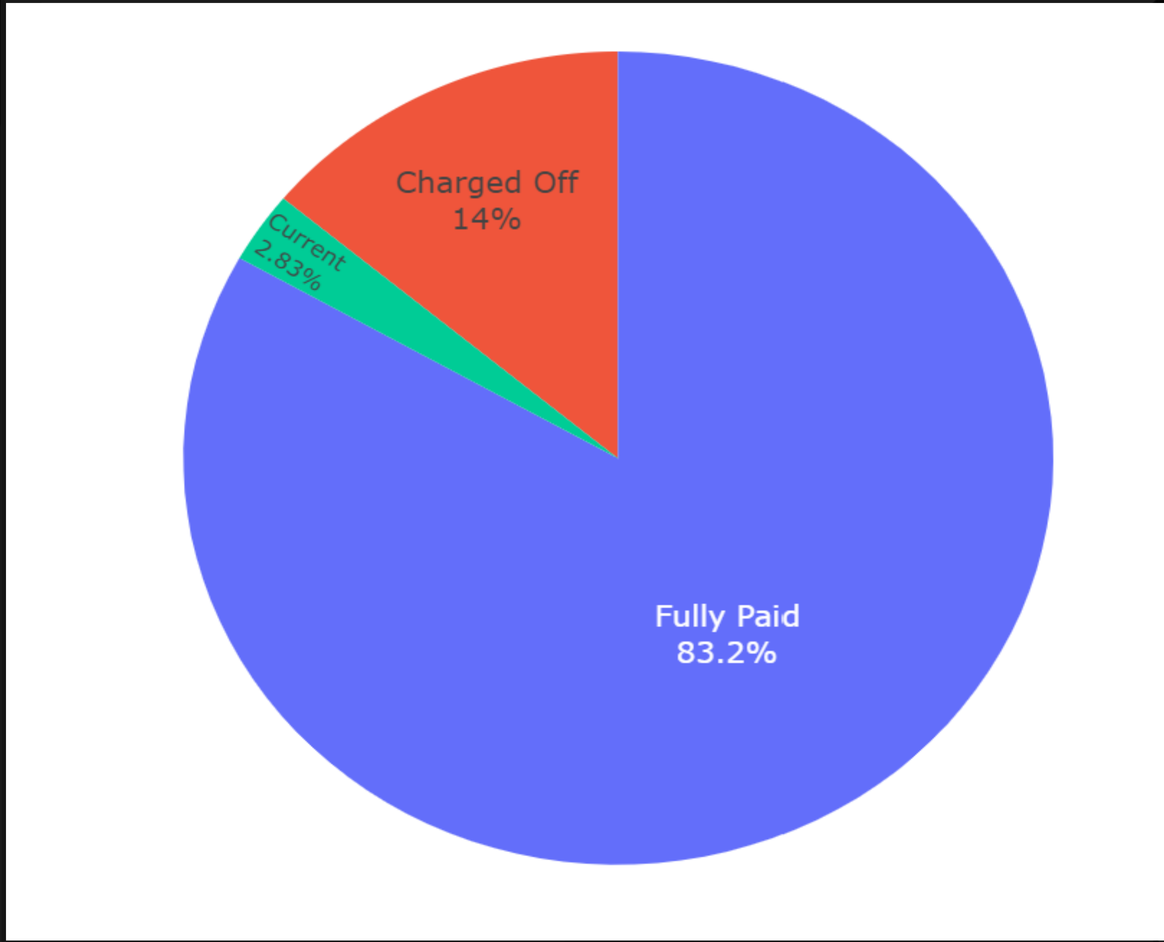
AFFECT OF LOAN PURPOSE AND LOAN AMOUNT ON DEFAULT RATE

Box Plot to see affect of **loan purpose** and **loan amount** on default rate.



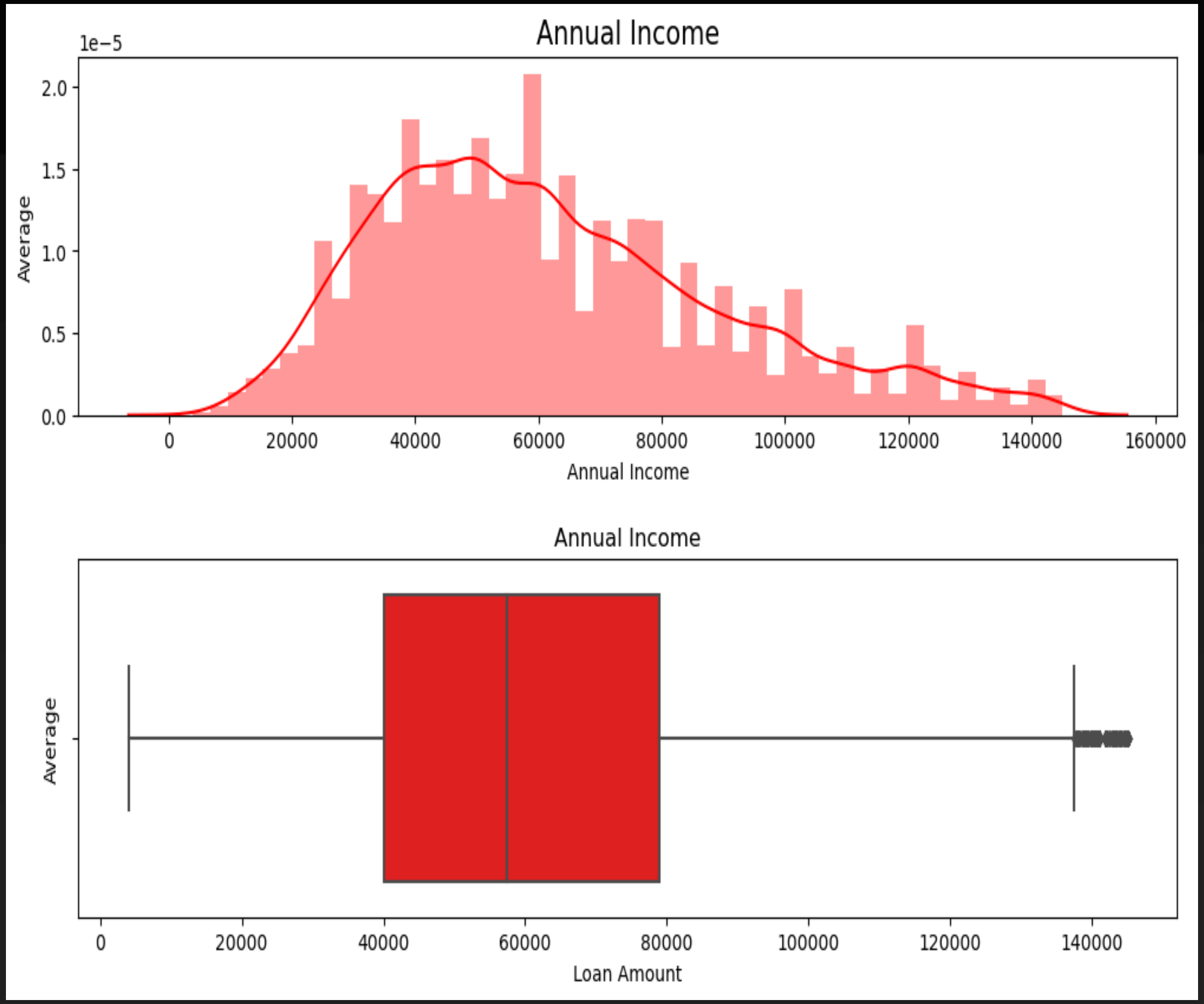
LOAN STATUS

83.2% of individuals have successfully repaid their loans, **2.83%** are current loans and **14%** loans have been declared as charged off.



ANNUAL INCOME

Majority of loans fall within the income range of **40000 to 80000**



TECHNOLOGIES USED

- Python
- Numpy
- Pandas
- Matplotlib
- Seaborn

ACKNOWLEDGEMENT

"We want to express our heartfelt gratitude to the teachers and peers at UpGrad/ IIITB for their consistently clear and informative sessions, as well as for patiently addressing our queries every day. I also extend my appreciation to the entire UpGrad team for providing us with this incredible learning platform."

THANK YOU