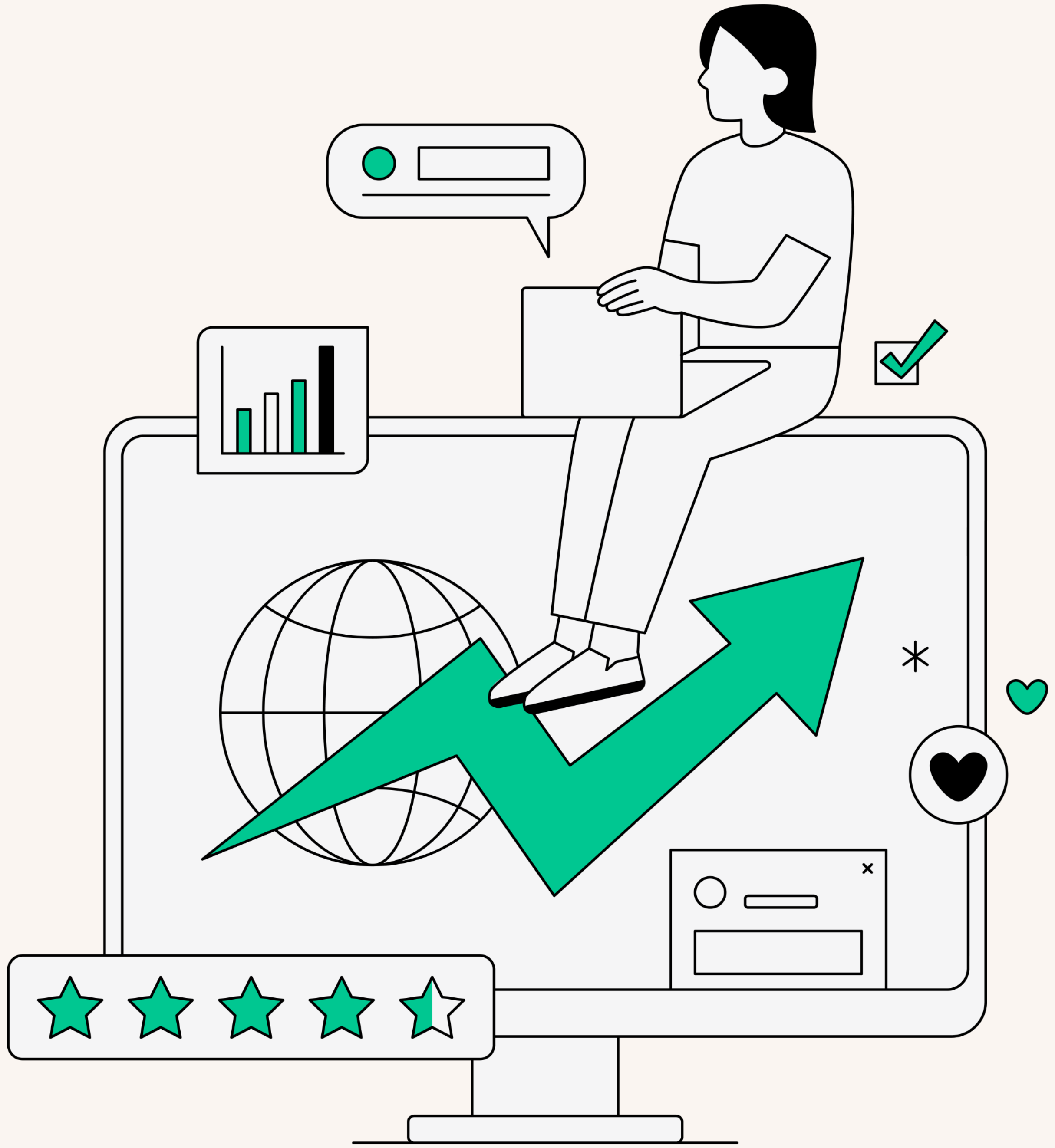
Presented by Devesh Parashar

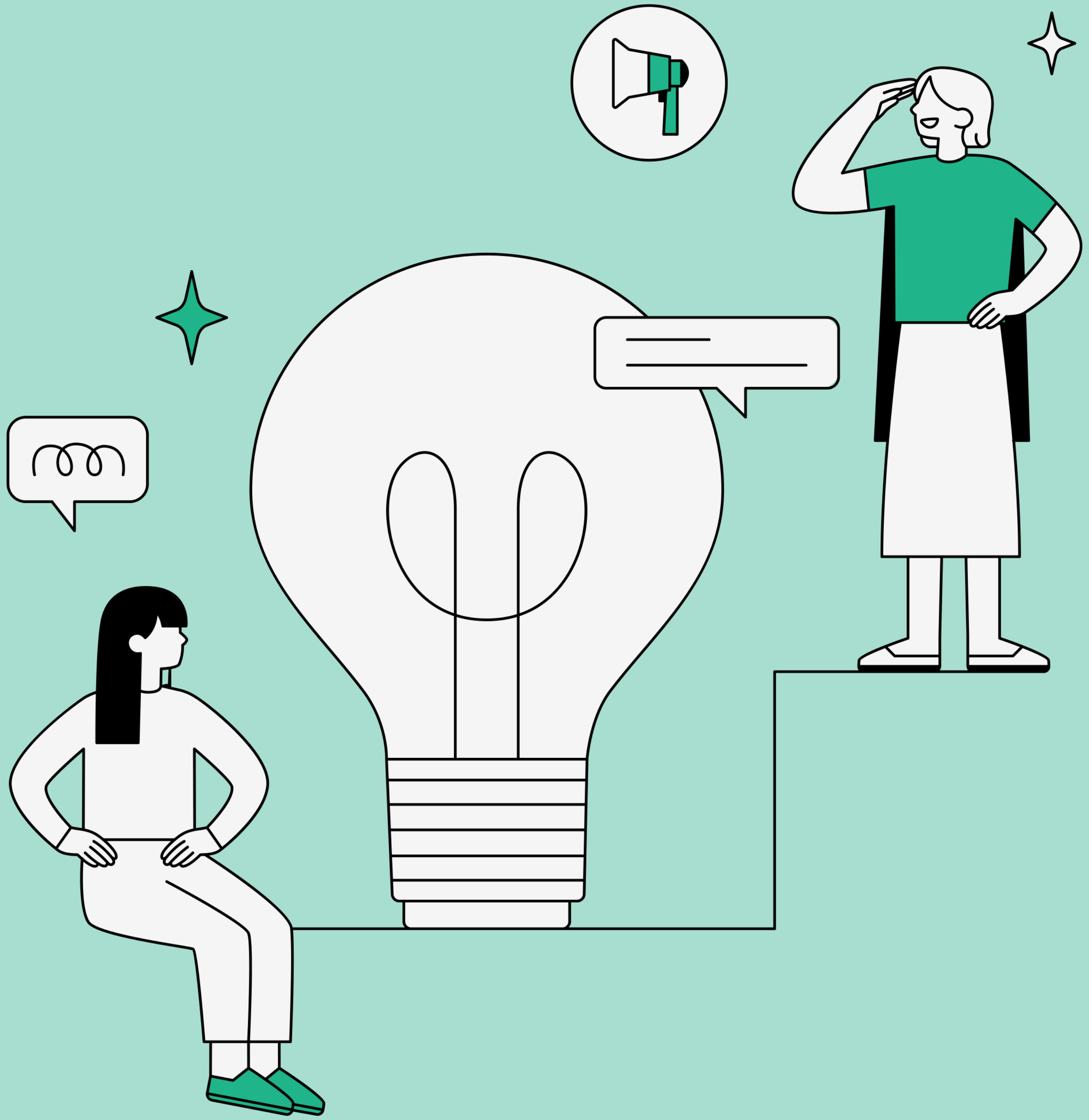
EDA CASE STUDY

CREDIT Risk Analysis



Business Objective

This case study aims to identify patterns that indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of the loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. **Identification of such applicants using EDA is the aim of this case study.**



Problem Statement

Two types of risks are associated with the bank's decision:

Case-1: If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

Case-2: If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

PROCESS

1 DATA UNDERSTANDING

2 DATA CLEANING

3 DATA ANALYSIS

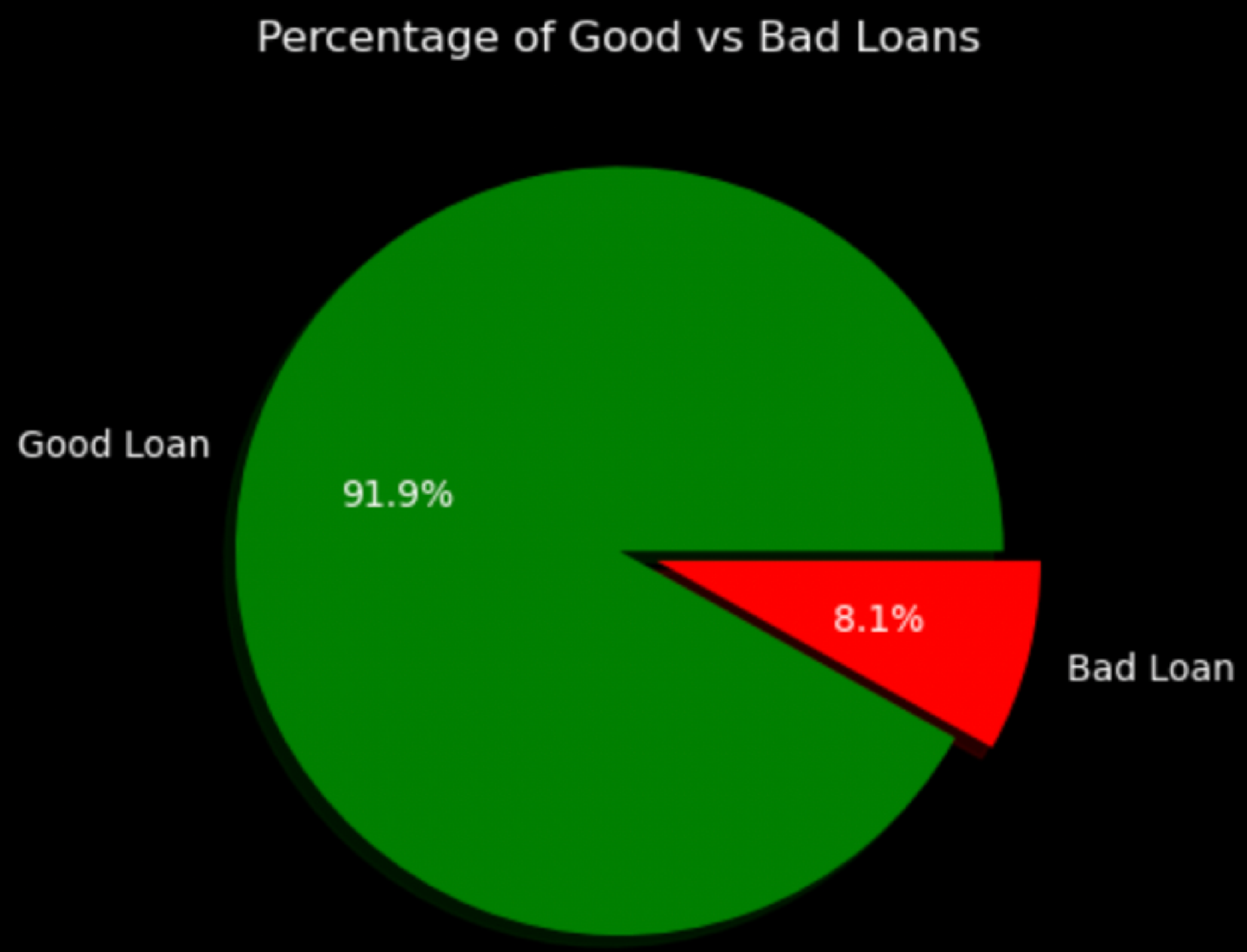
4 DATA VISUALISATION

5 UNCOVER INSIGHTS

Methodology used in the analysis

1. Used Python's Pandas library for EDA
2. Imputed missing values with mean, median, or mode based on the data types and column type.
3. Dealed with outliers by capping, or removing them.
4. Dropped unnecessary columns and columns with more than 40% null values.
5. Created new columns for better data analysis
6. Checked data imbalance and analysed accordingly
7. Merged current and previous applications' data frames
8. Found correlation between variables and picked the best ones
9. Used Maplotlib and Seaborn for data visualisation.

Good vs Bad Loan

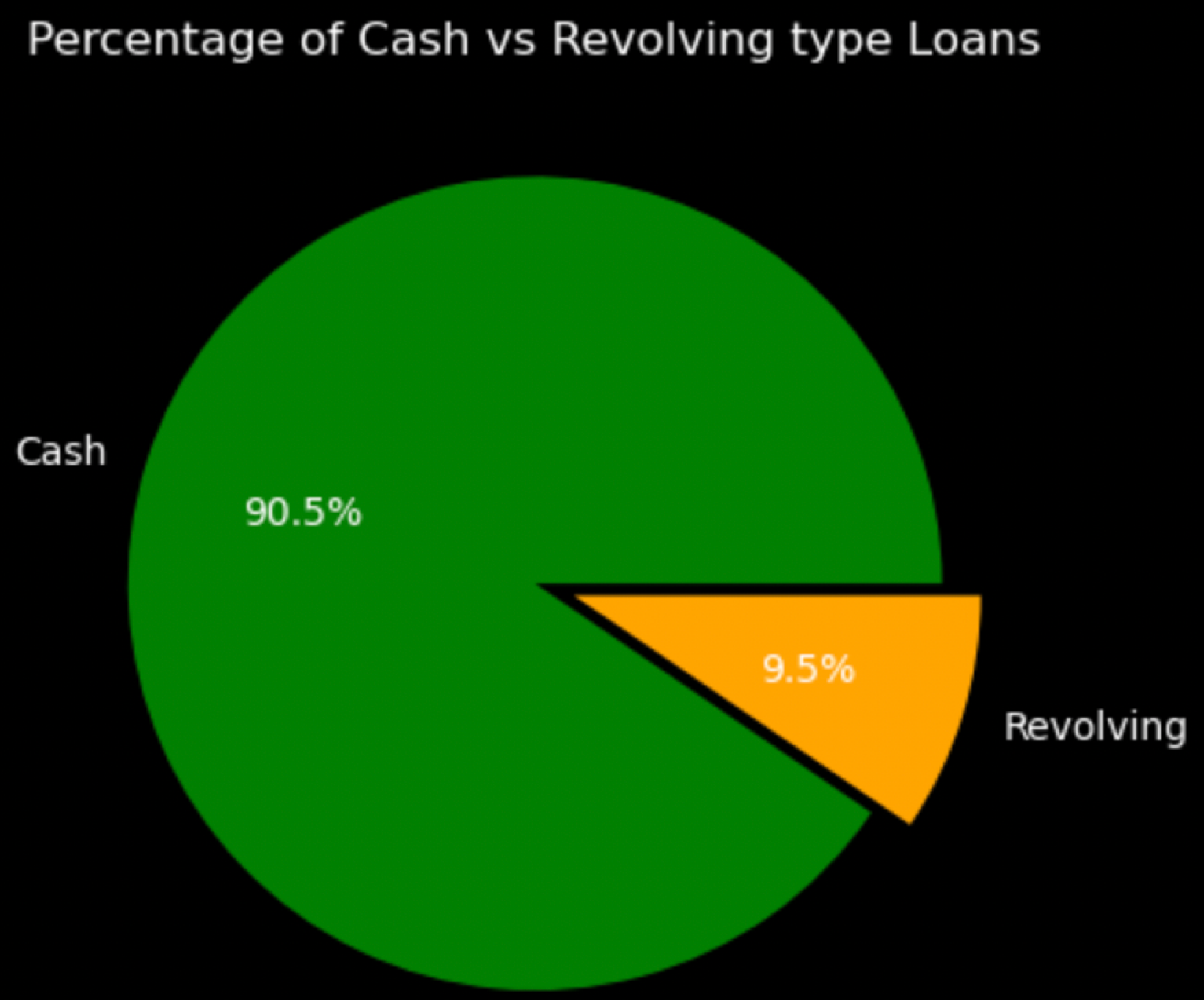


Good Loans - Applicant paying all installments on time.

Bad Loans - Applicant not paying installments and will likely default.

Bad Loans are 8.1% of total applicants

Cash vs Revolving Loans

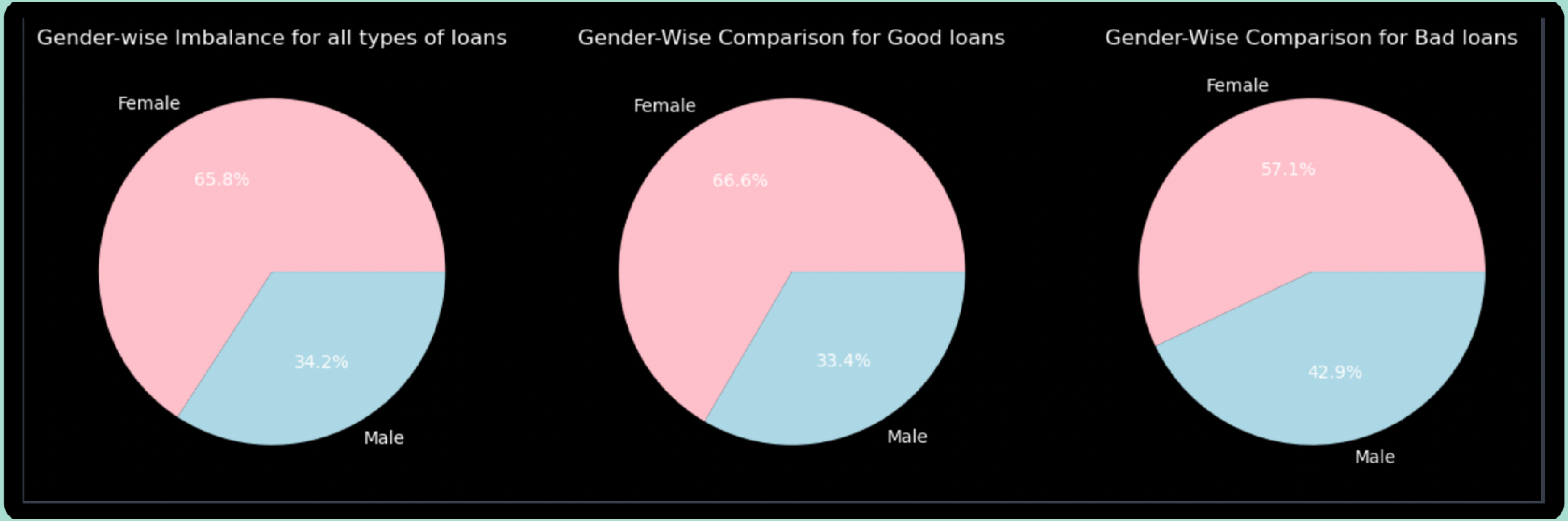


In our dataset

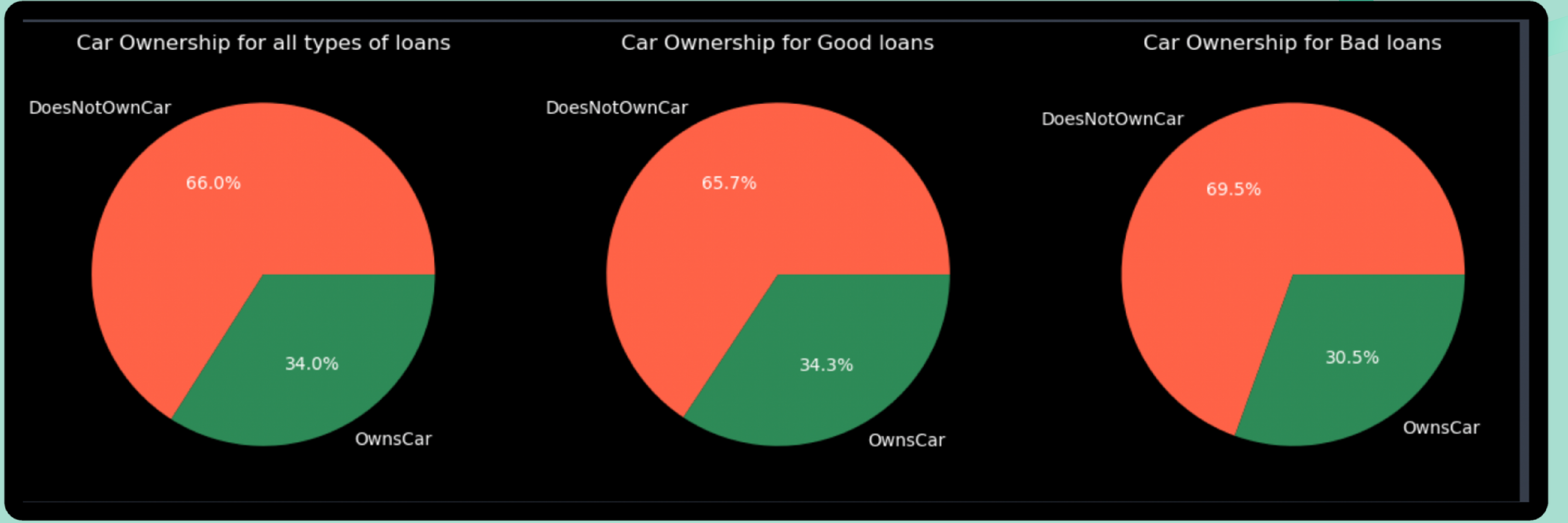
- 90.5% of loans are Cash loans.
- 9.5% only are Revolving loans.

Gender Imbalance in our Dataset

- There are clearly more females in our data.
- Percentage wise males are more likely to default than females.

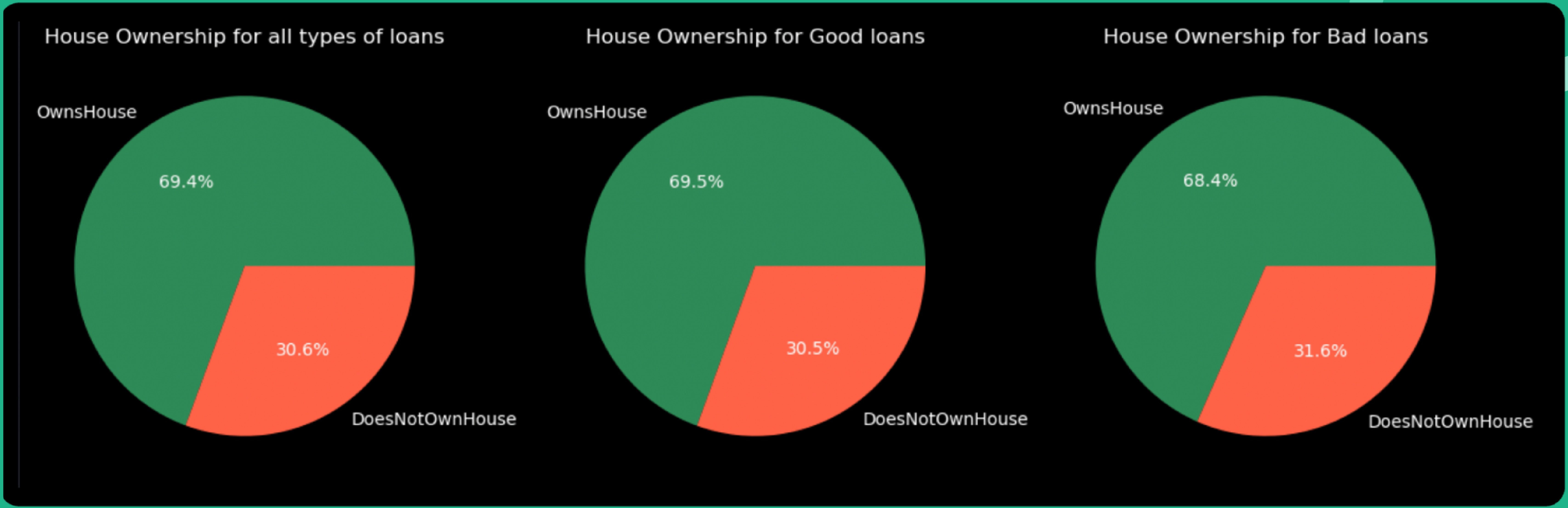


Car Ownership Data Imbalance



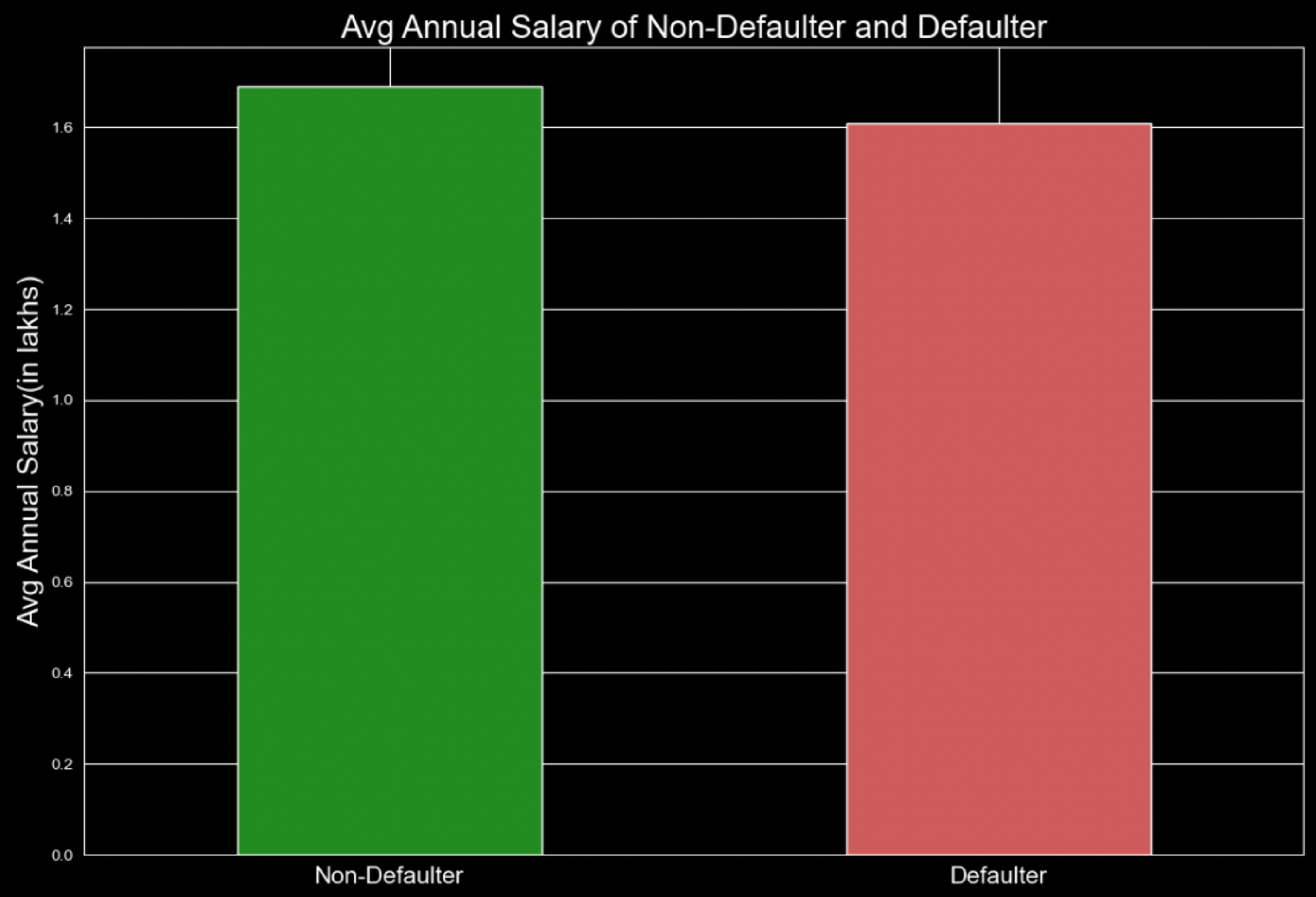
- 66% of applicants does not own car.

House Ownership Data Imbalance



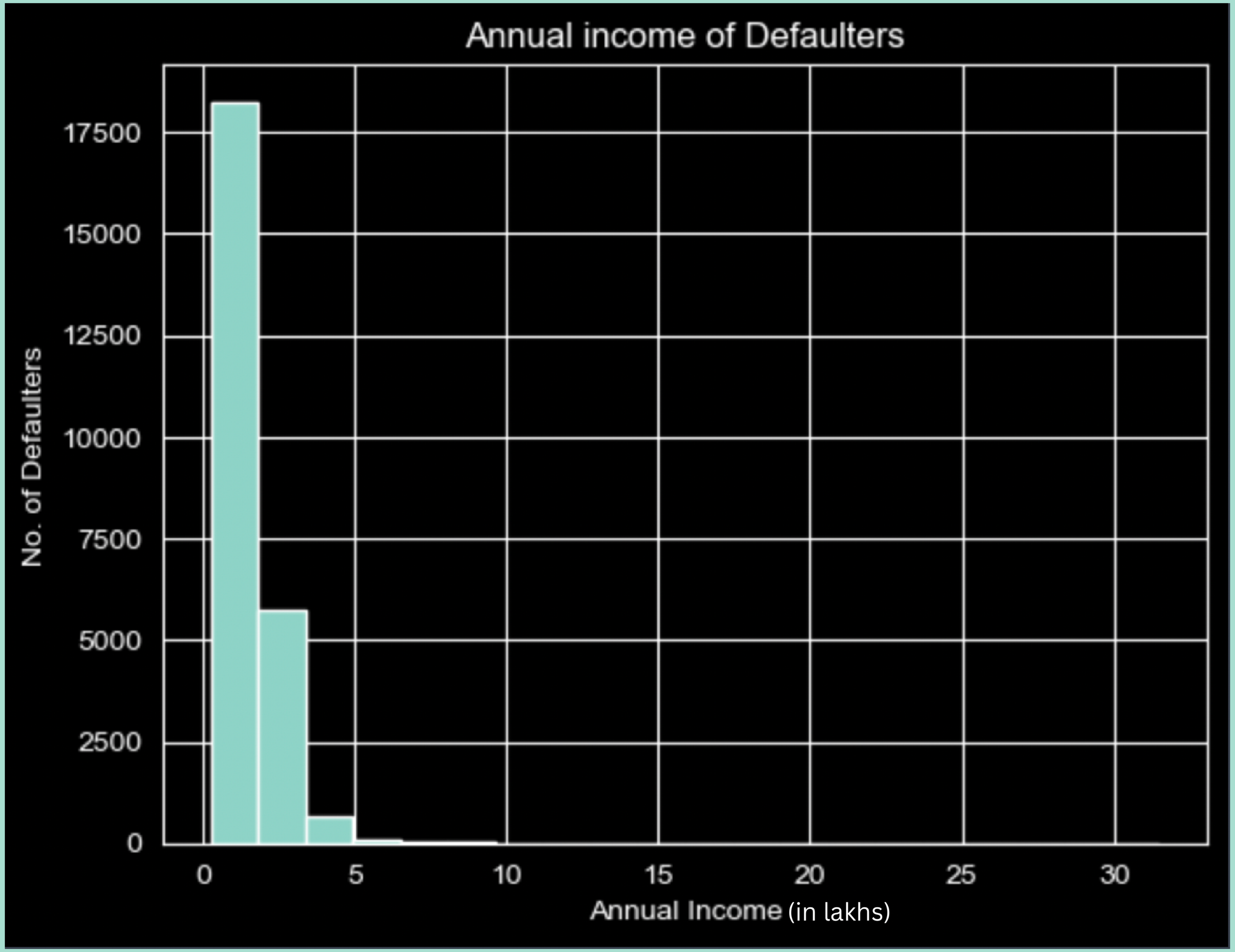
- 69.4% of applicants owns House

Avg Salary of Non-Defaulter vs Defaulter



On average, Salary of a Defaulter is slightly less than Non-Defaulter.

Annual Income of Defaulters

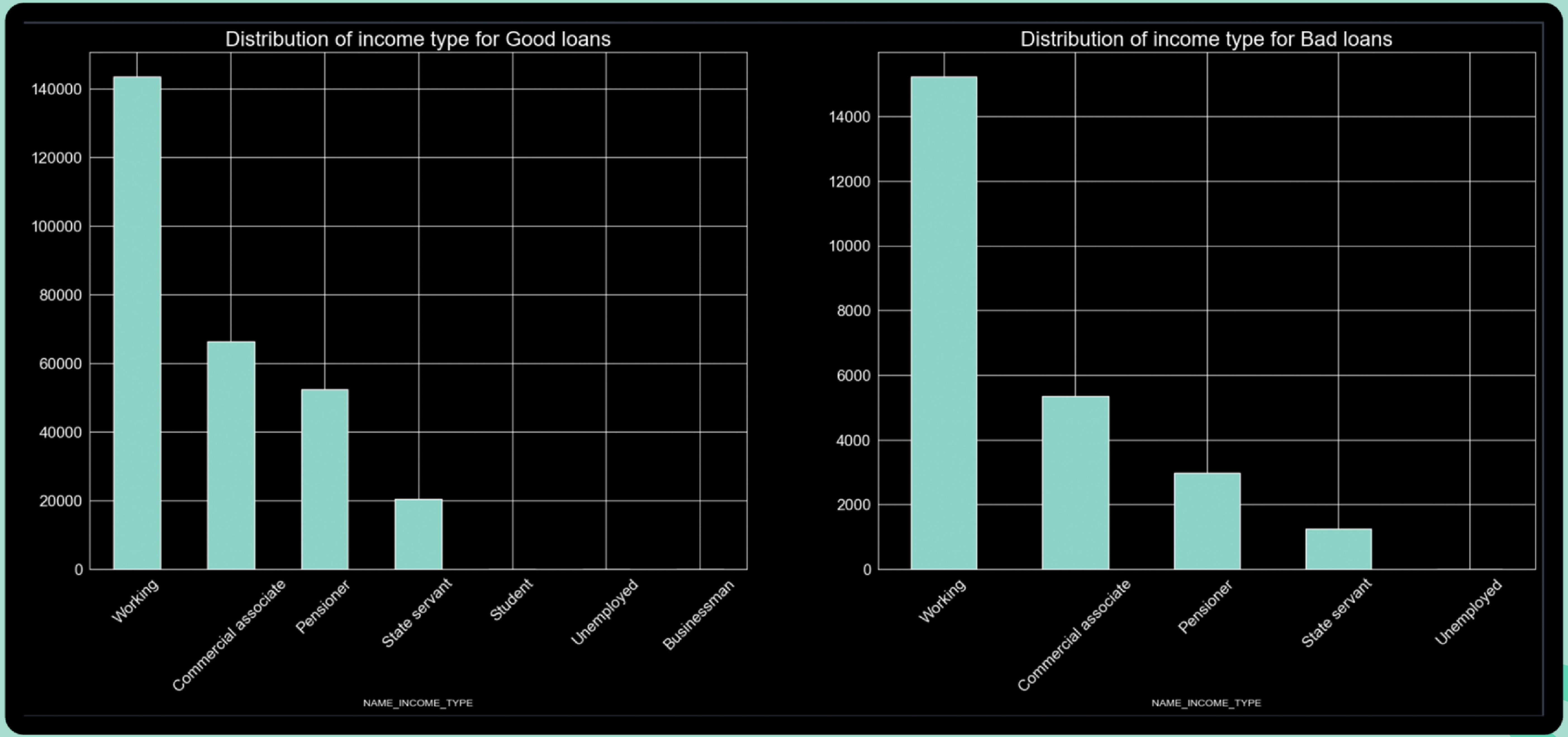


- Most defaults occur for people who have an income of less than **3 lakhs**.
- People with income of more than **5-7 lakhs** are less likely to default.

Age & Income trend of Non-Defaulter and Defaulter

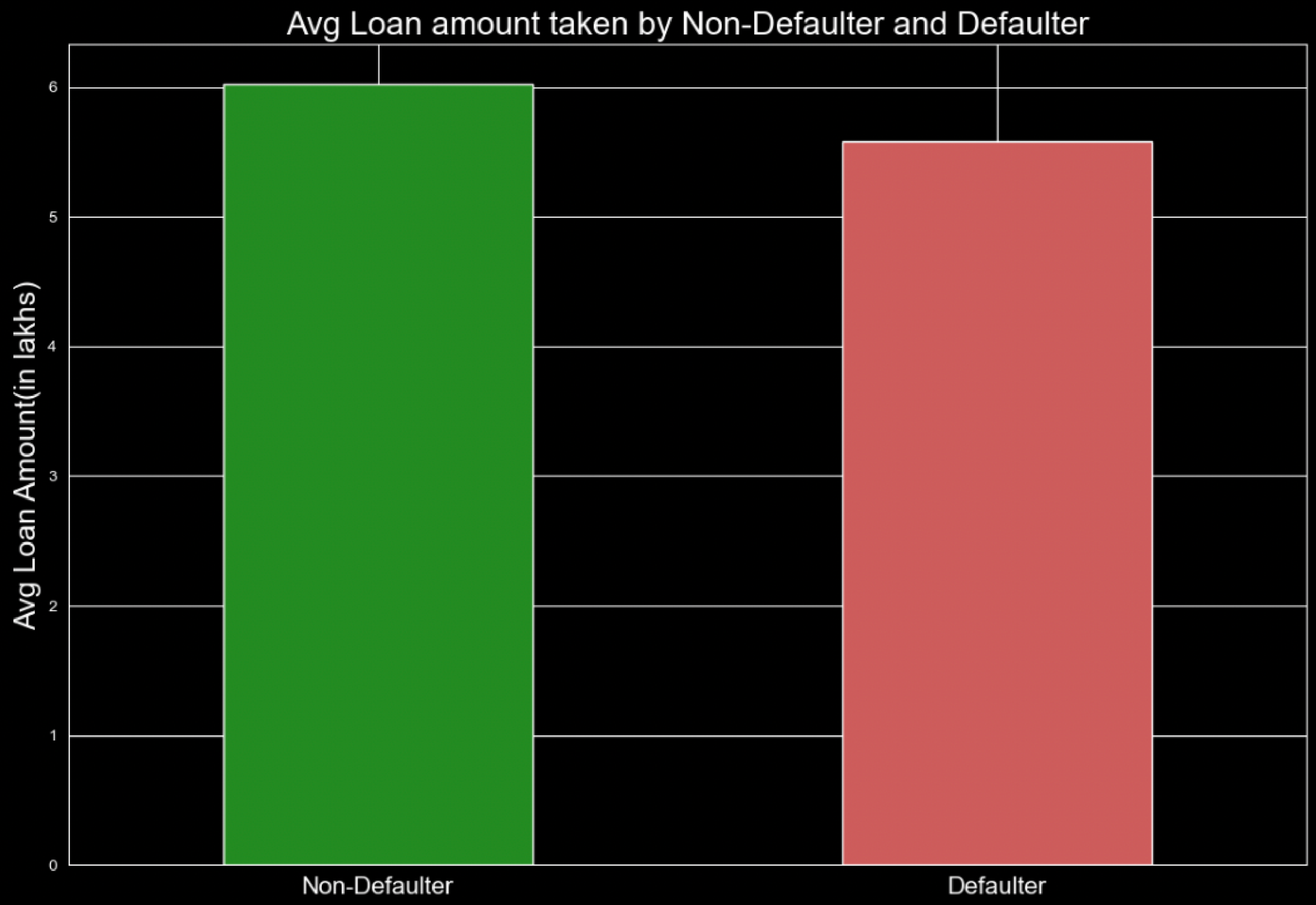
- The annual income trend rises until age 45 and then it starts falling slowly for both defaulter and non-defaulters.
- For defaulters, as I analysed earlier through bar charts, annual income in all age groups is slightly lower as compared to non-defaulter.

Defaulters vs Non-Defaulters based on Income Type



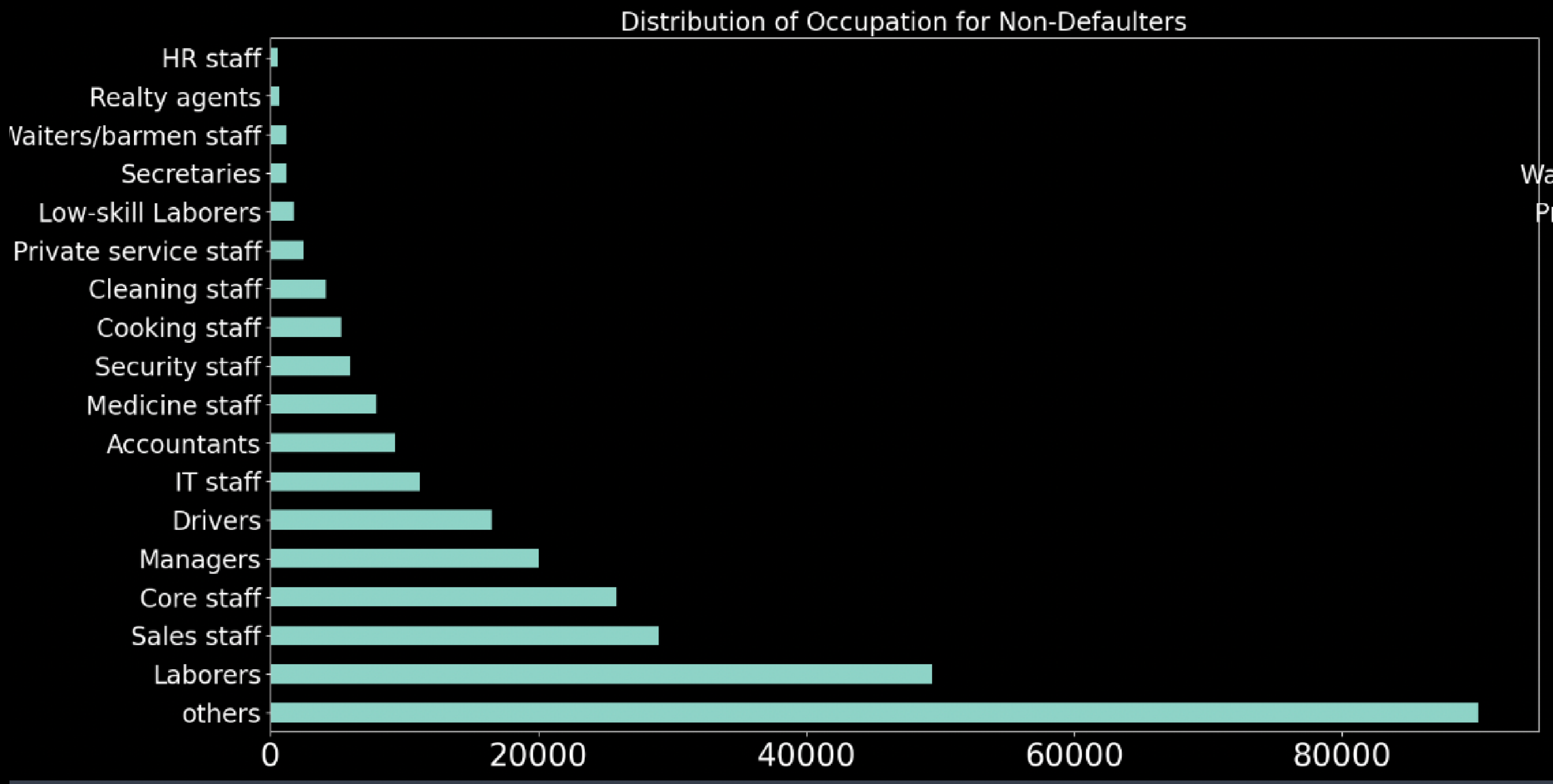
- There is a data imbalance, most applicants are working class, hence most defaults in absolute terms are from this class only.
- No student or businessman has defaulted

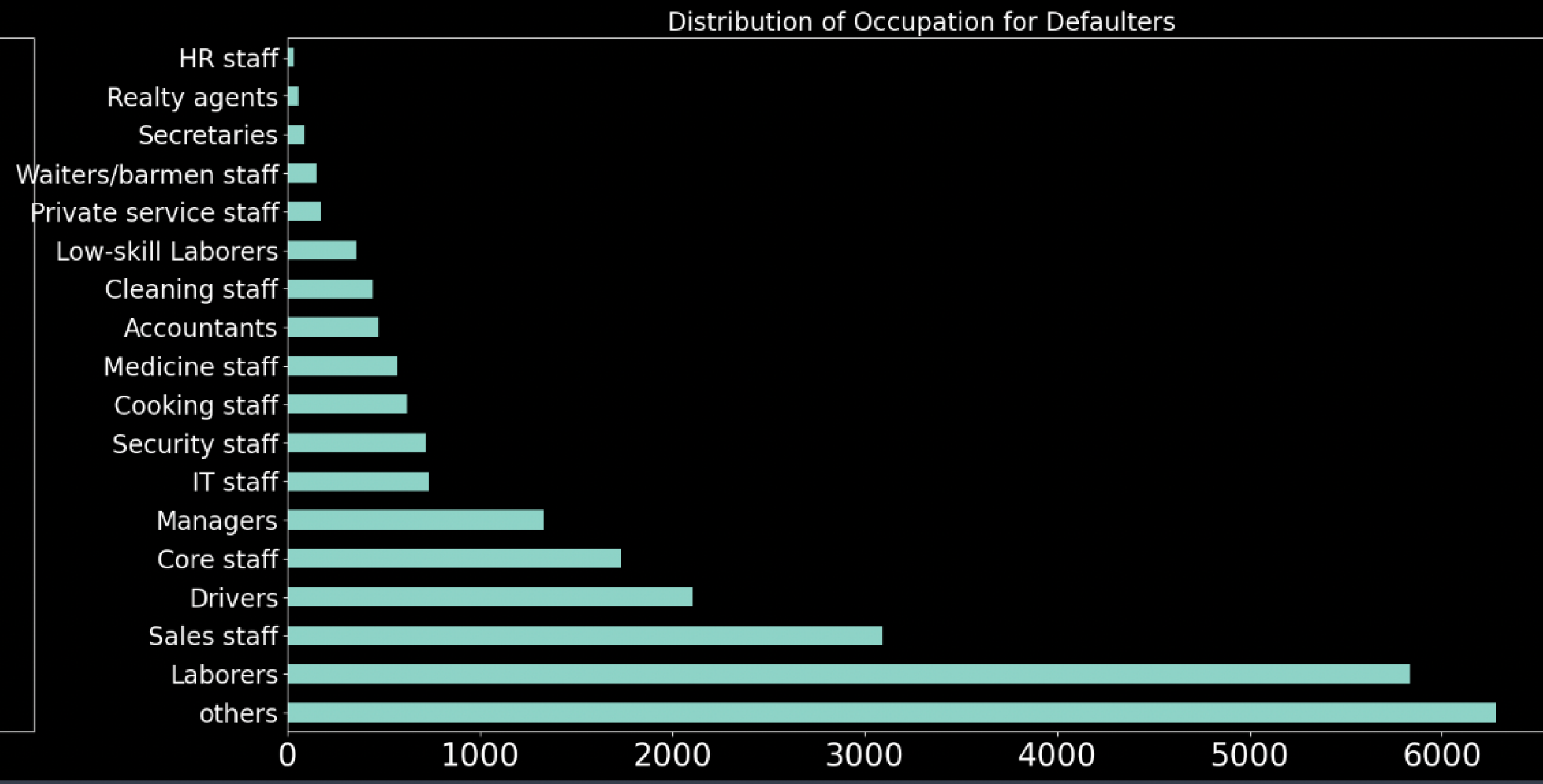
Avg Credit of Non-Defaulter vs Defaulter



- On average, Loan Amount taken by a Defaulter is slightly less than Non-Defaulter.

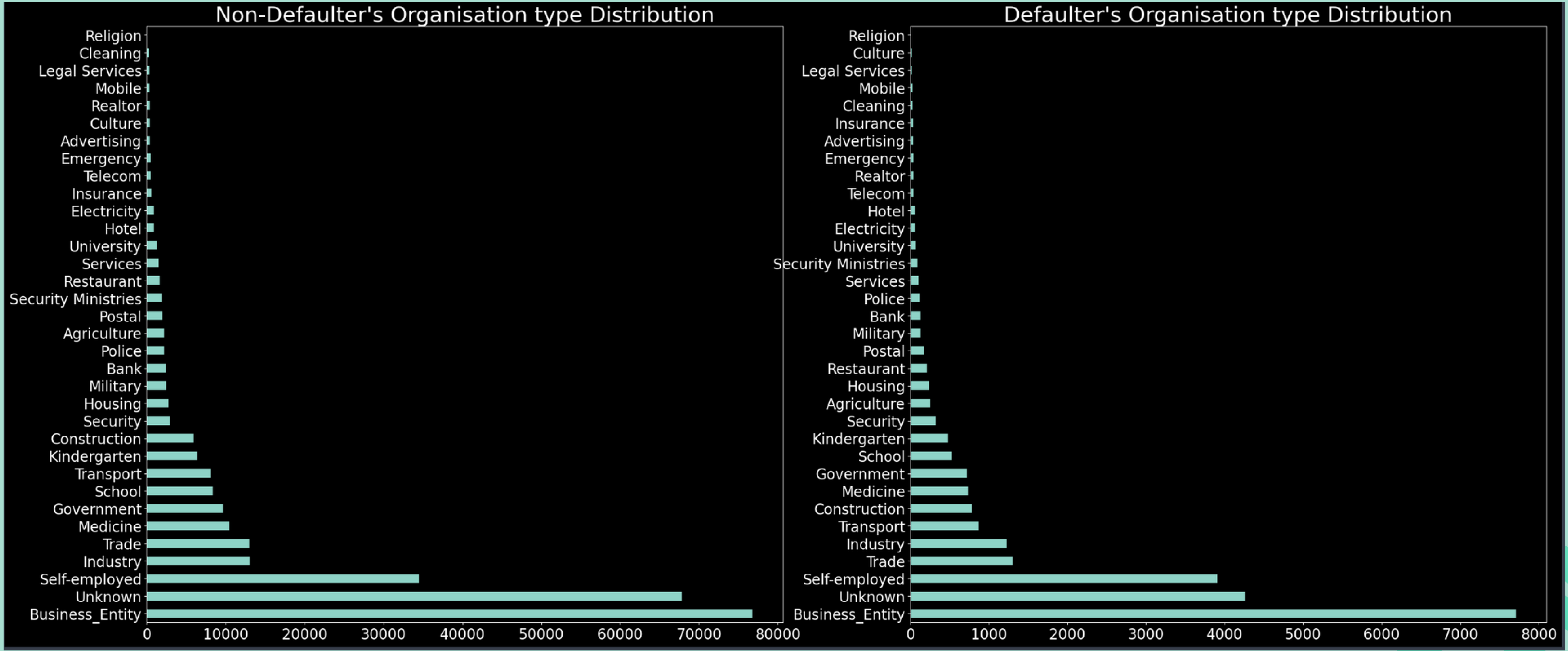
Defaulters vs Non-Defaulters based on their Occupation





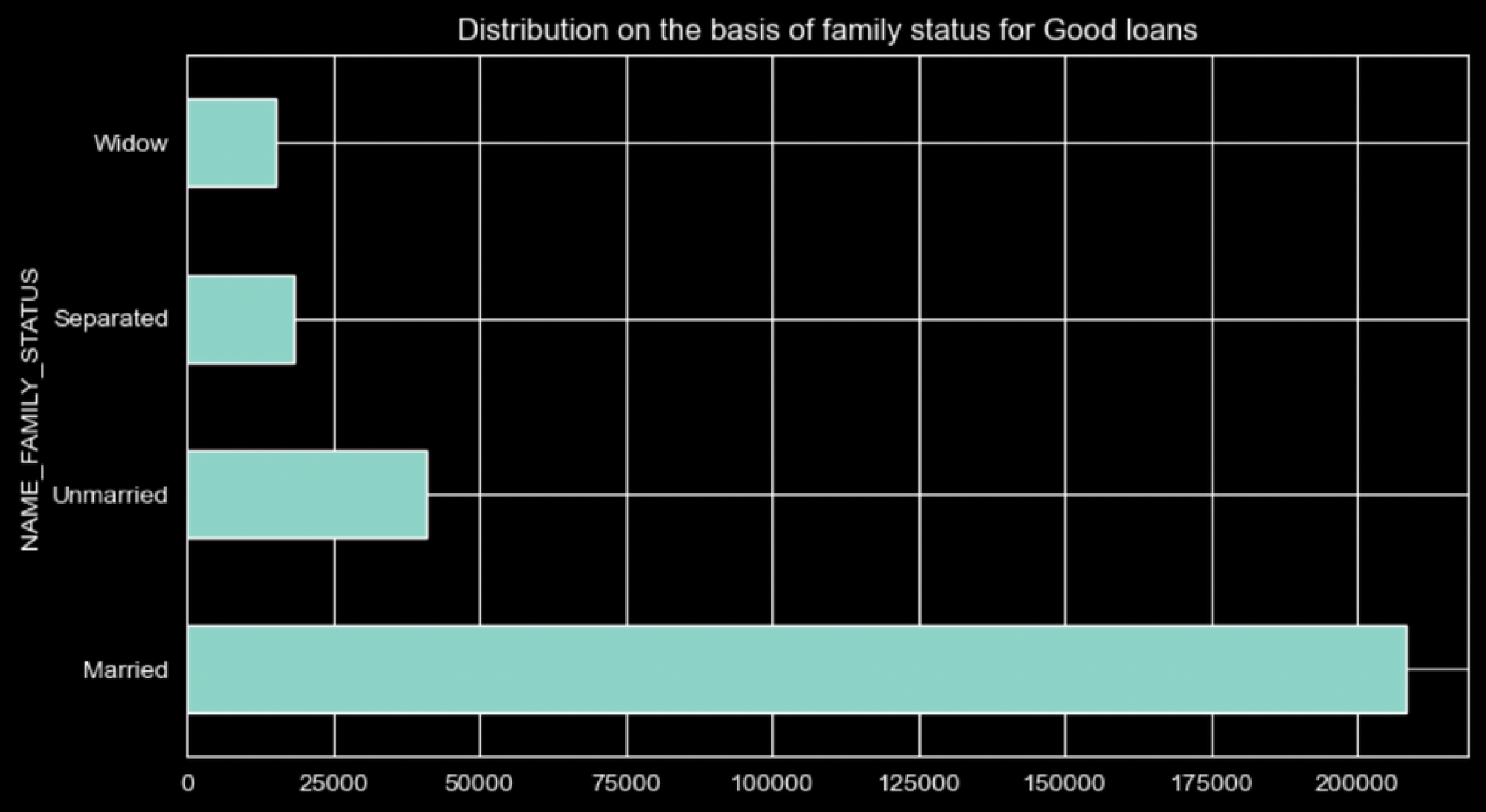
- Applicants who are in the occupation type Labourers are more likely to default than other occupation type applicants.

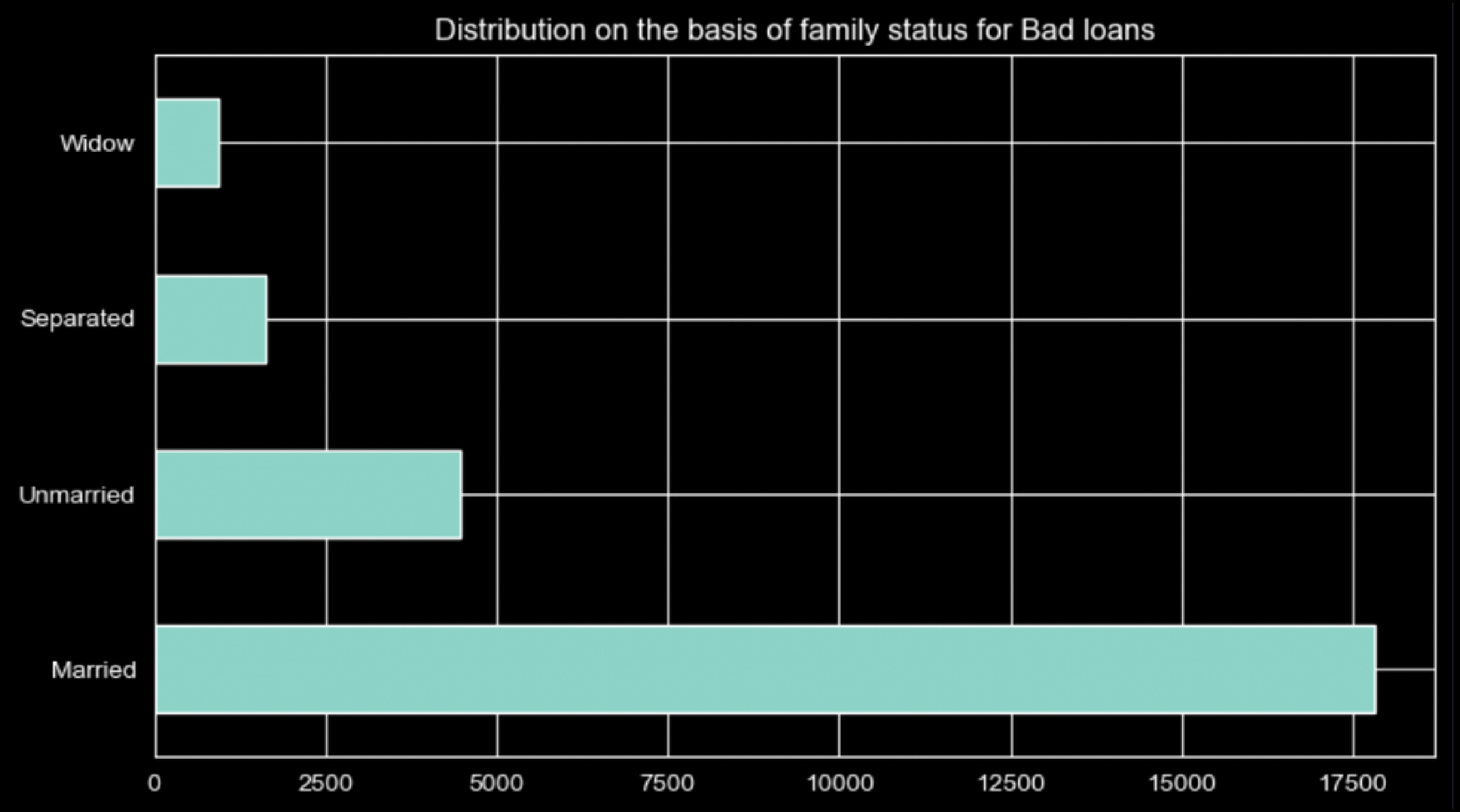
Defaulter vs Non-Defaulter based on their Organisation



- People in Construction and Restaurant organisation have the highest default rate which is 12%
- Most defaults are from Business Entity(absolute terms).

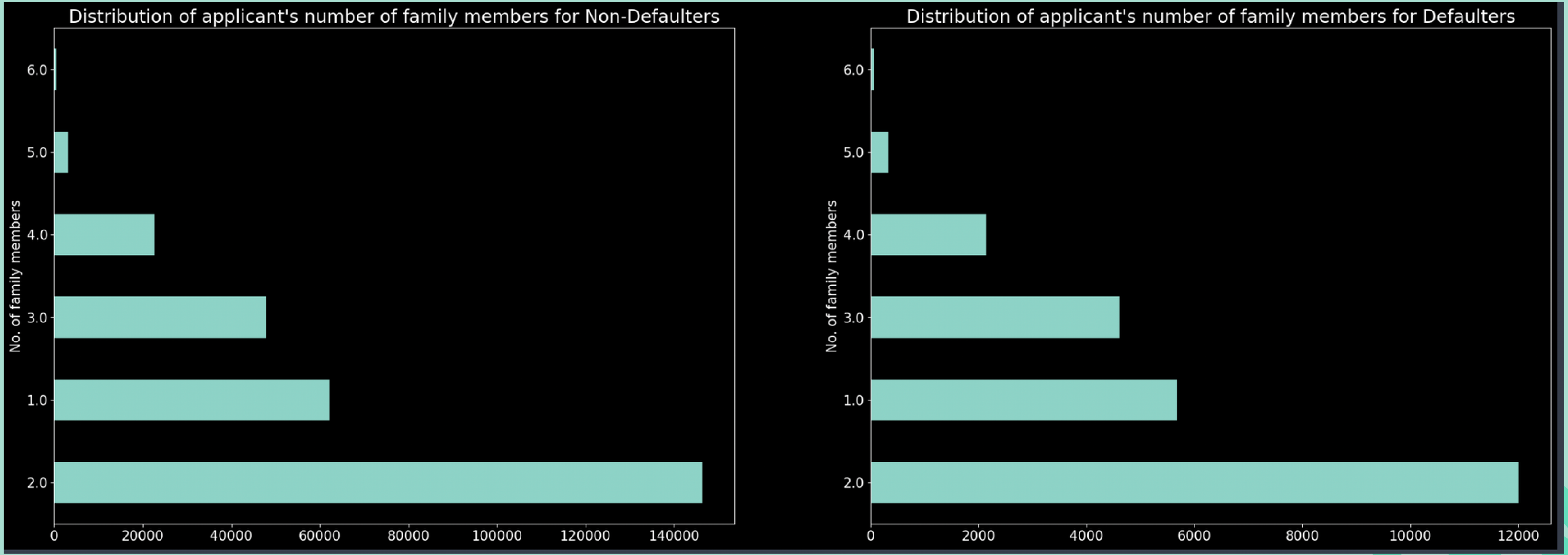
Non-Defaulter vs Defaulter based on Family Status





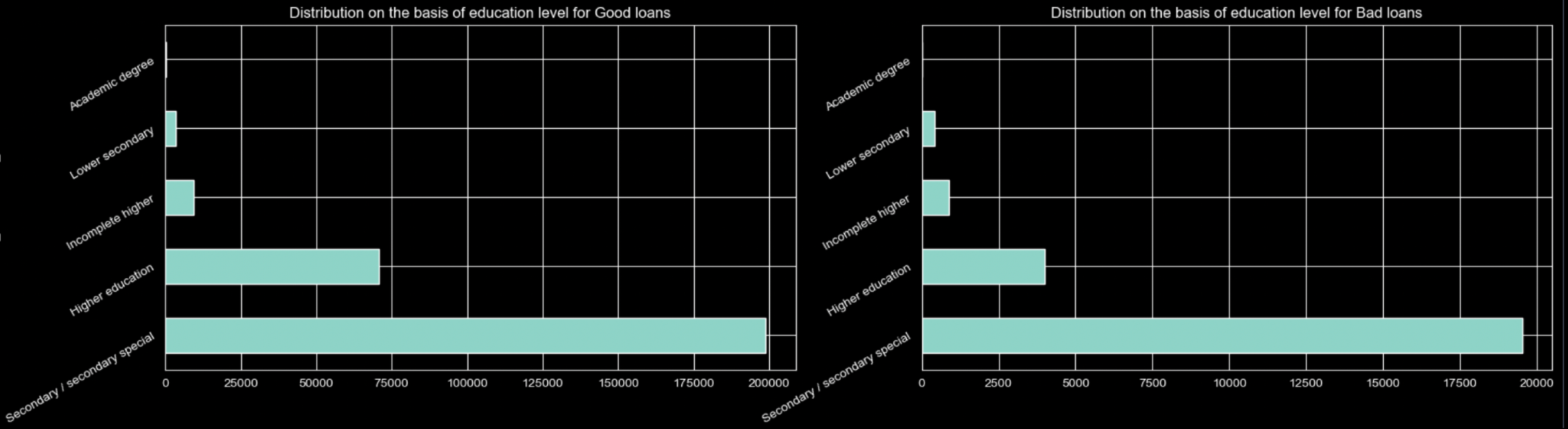
- We have more married people in our dataset.
- Unmarried people have the highest default rate of 9.8%.

Defaulter vs Non-Defaulter based on No. of family members



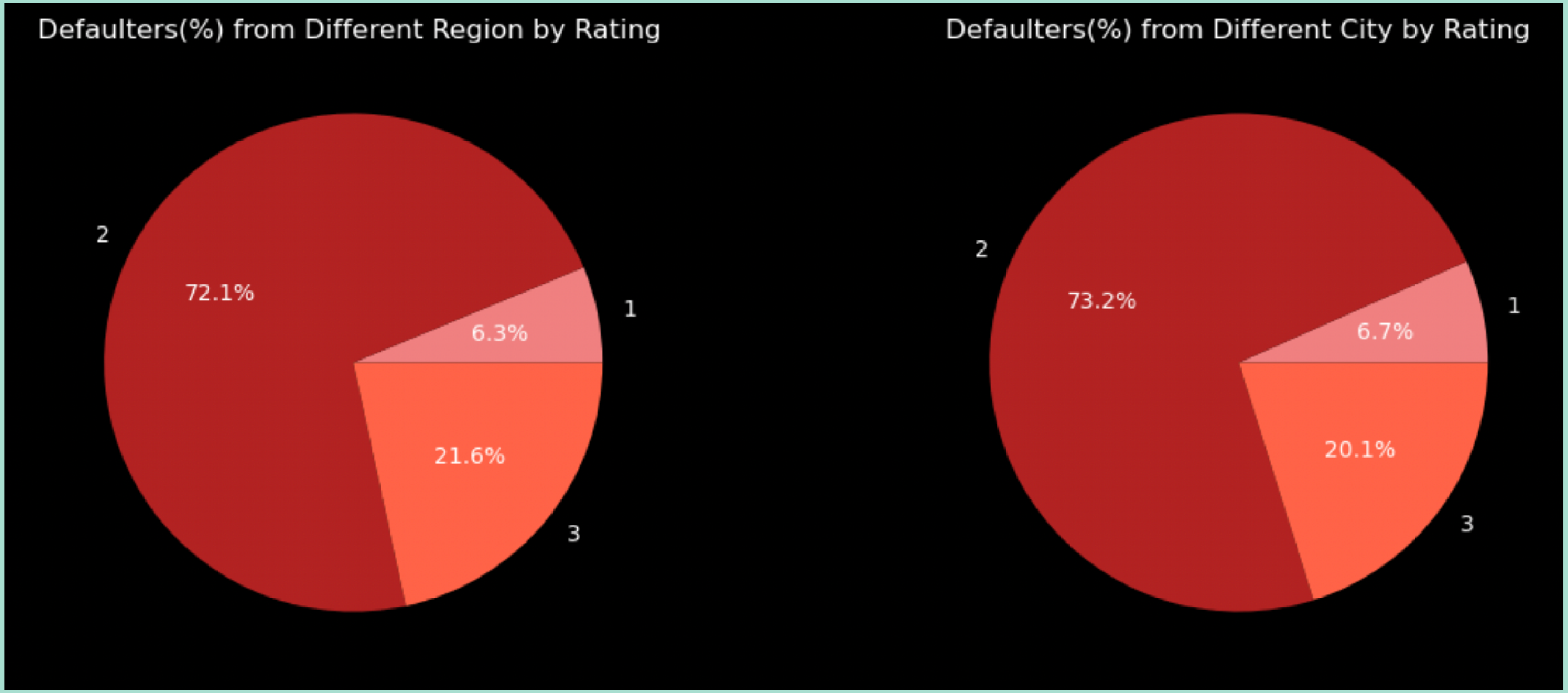
- People with 5 or more family members have 9.4% default rate which is the highest among all other categories.

Non-Defaulter vs Defaulter based on Education Level



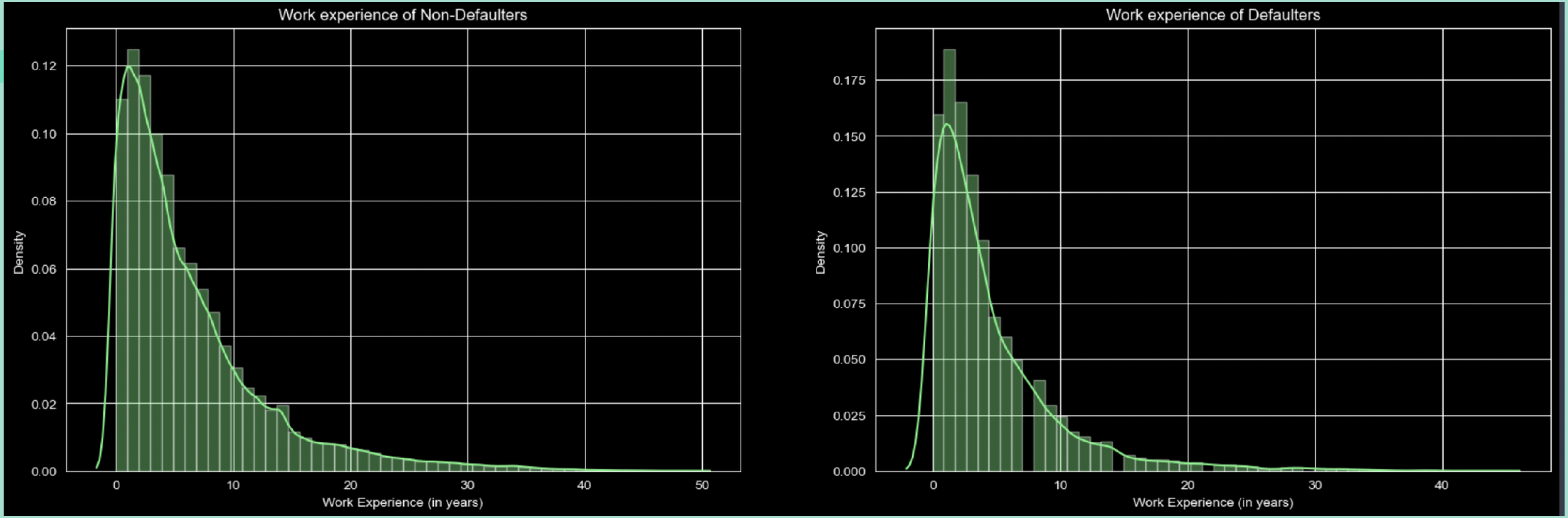
- We have more people in our dataset with secondary and less from academic degrees.
- The lowest default rate is for people with academic degrees (1.8%).
- The highest default rate is for people with lower secondary education (11%).

Defaulters from Different Regions and city by Rating



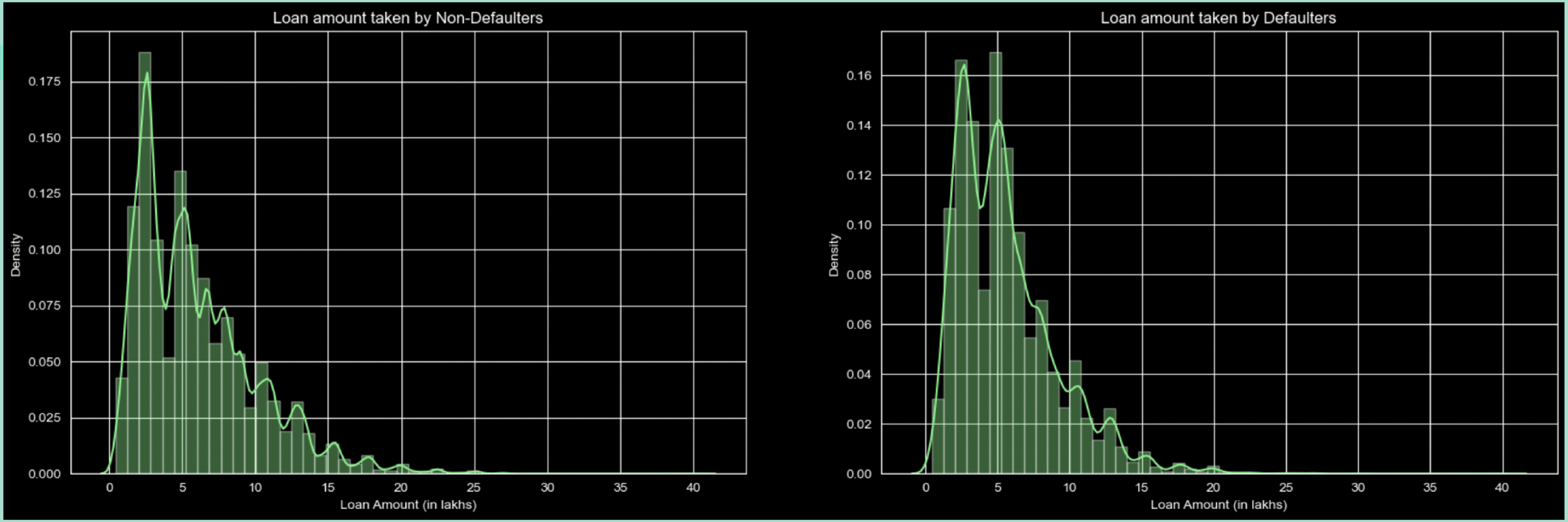
- Most defaulters are from region/city with rating of 2(absolute terms) and least from region/city with a rating of 1(absolute terms).
- But percentage-wise, Applicants from the city with a rating of 3 have the highest default rate of 11%.
- The lowest default rate is from cities with a rating of 1 i.e. only 4%

Work Experience of Non-Defaulter vs Defaulter



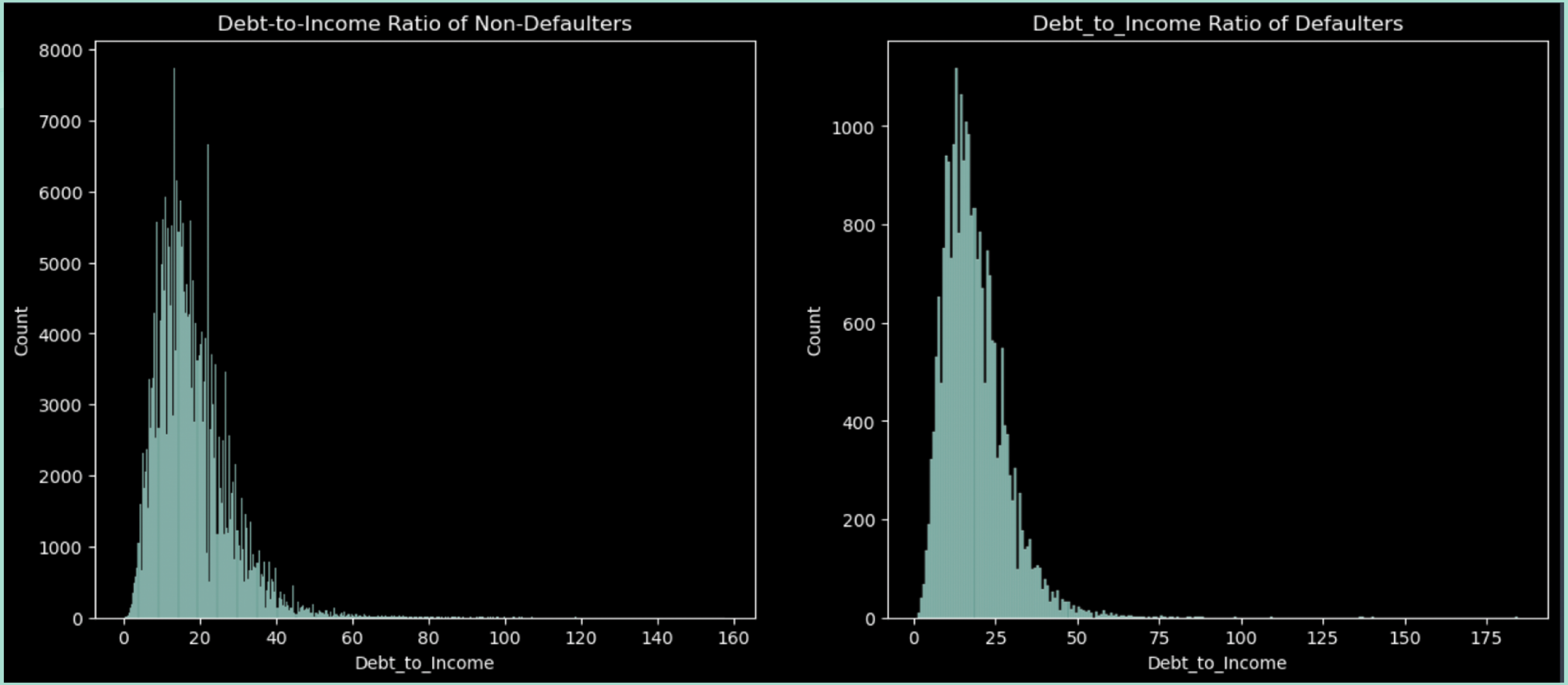
- Most applicants have work experience of 2-6 years.
- The percentage rate of defaulting is lowest for people with 40+ years of work experience.

Credit Taken by Non-Defaulter vs Defaulter



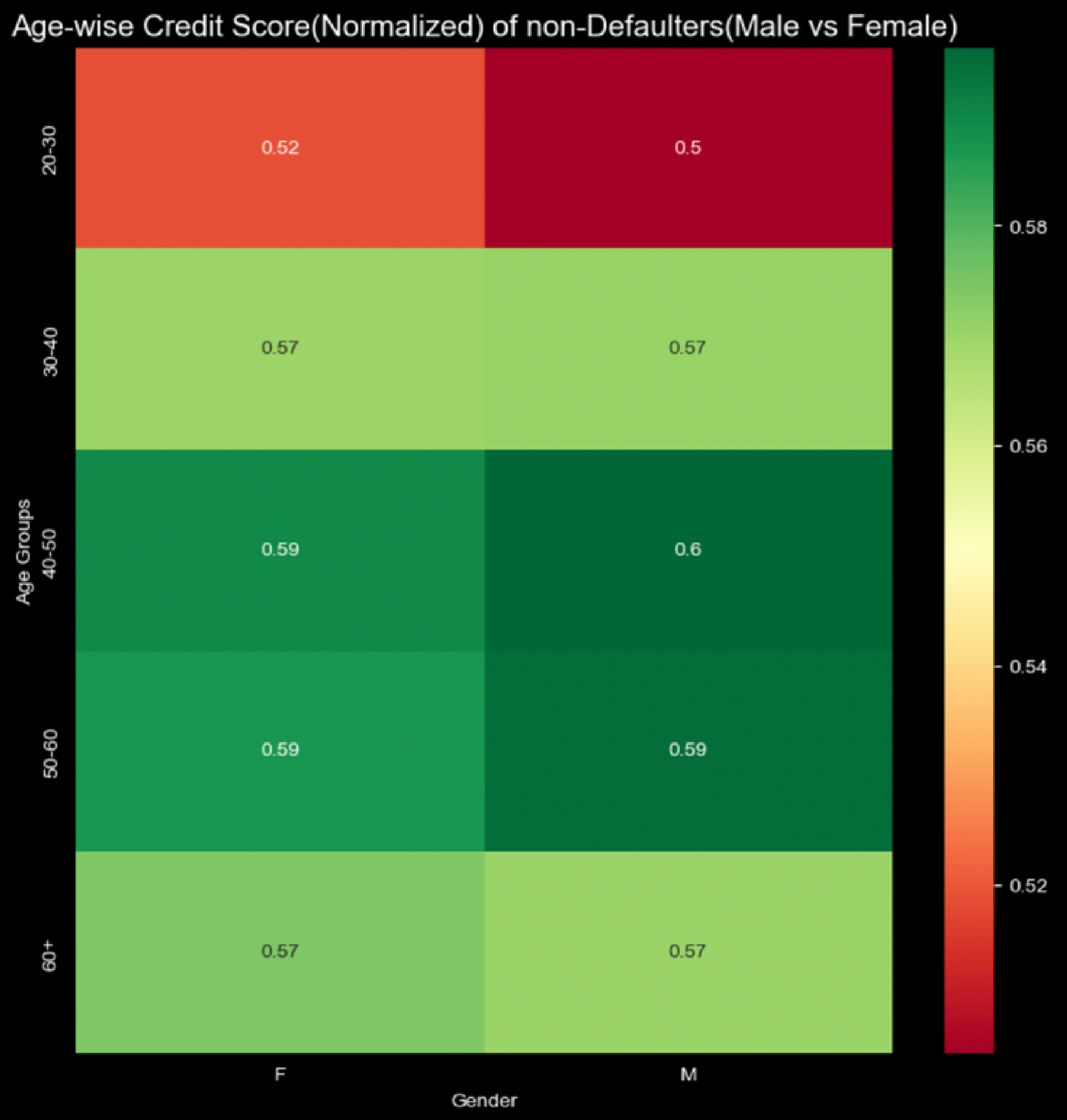
- Most applicants have taken loans in the range of 1-6 lakhs.
- People who have taken loans near 5lakhs are most likely to default.

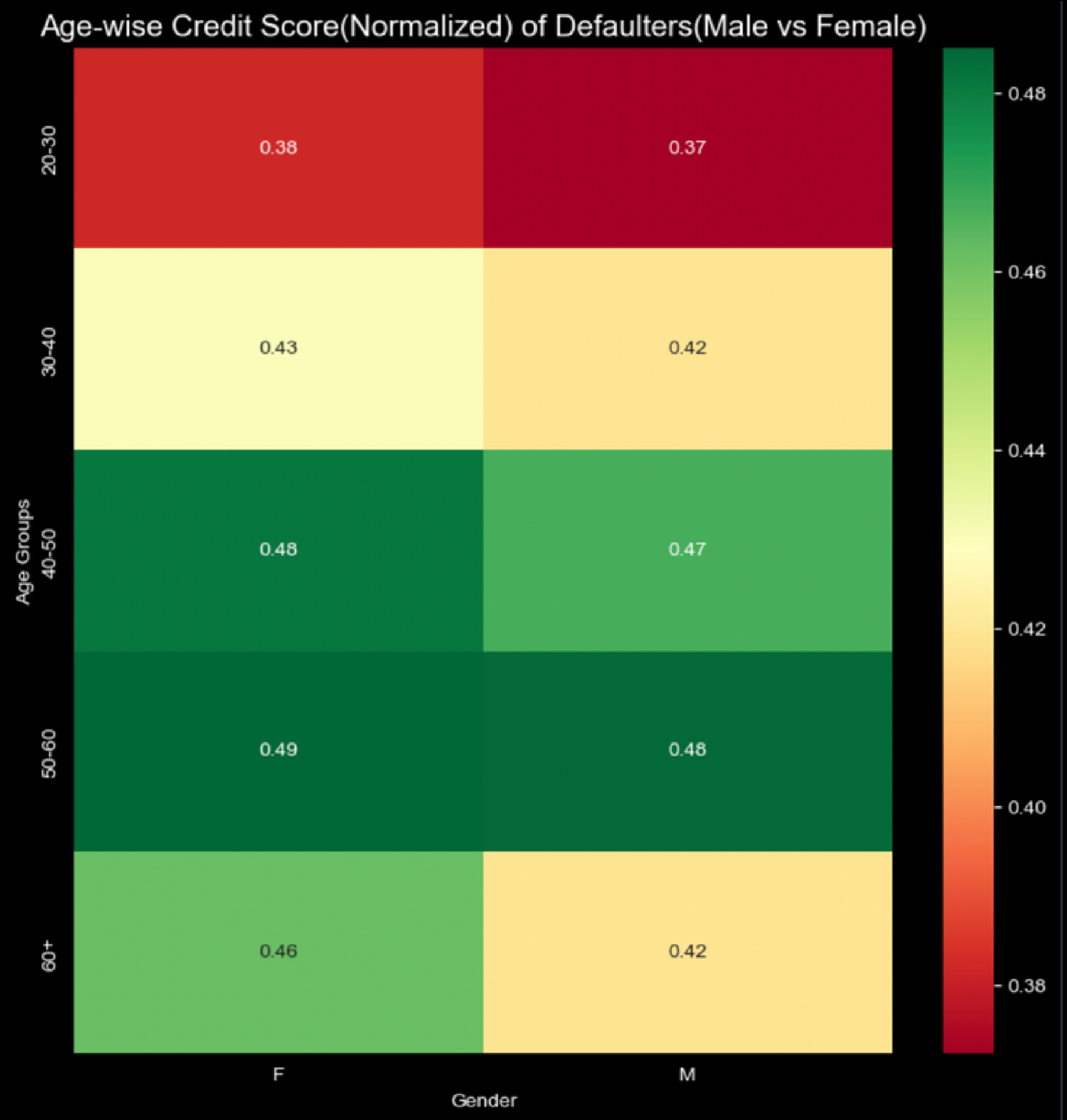
Debt-to-Income for Non-Defaulter vs Defaulter



- For all defaulters, the debt-to-income ratio i.e. percentage of income that goes to pay loan is higher than those of non-defaulters.

Age & Gender-Wise Credit Score of Non-Defaulter vs Defaulter

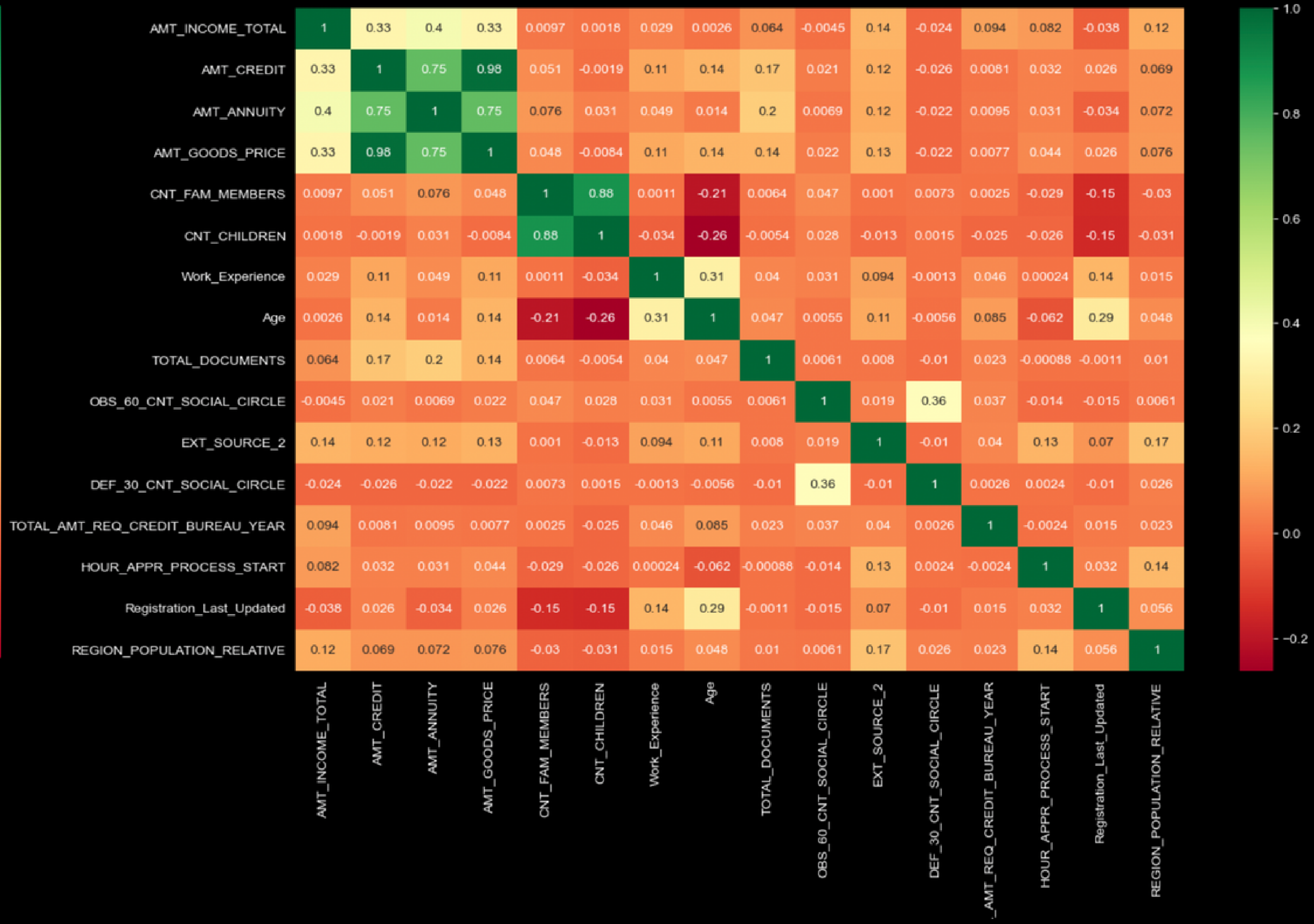




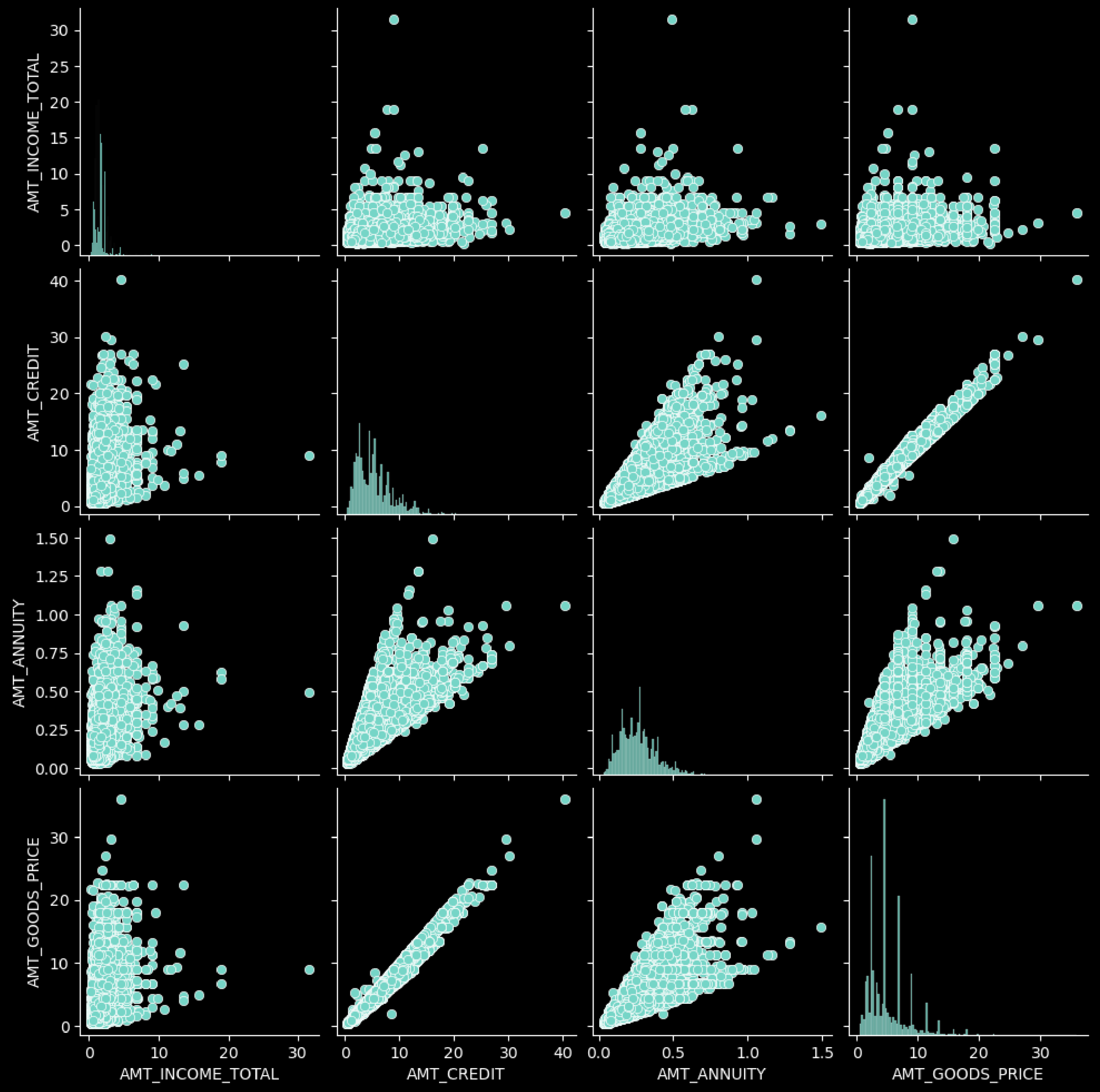
- Credit score of defaulters at all age stages, for both males and females, is lower than those of Non-defaulters.
- Credit score is lowest for people in ages ranging 20-30.

Non-Defaulter Correlation

Defaulter Correlation

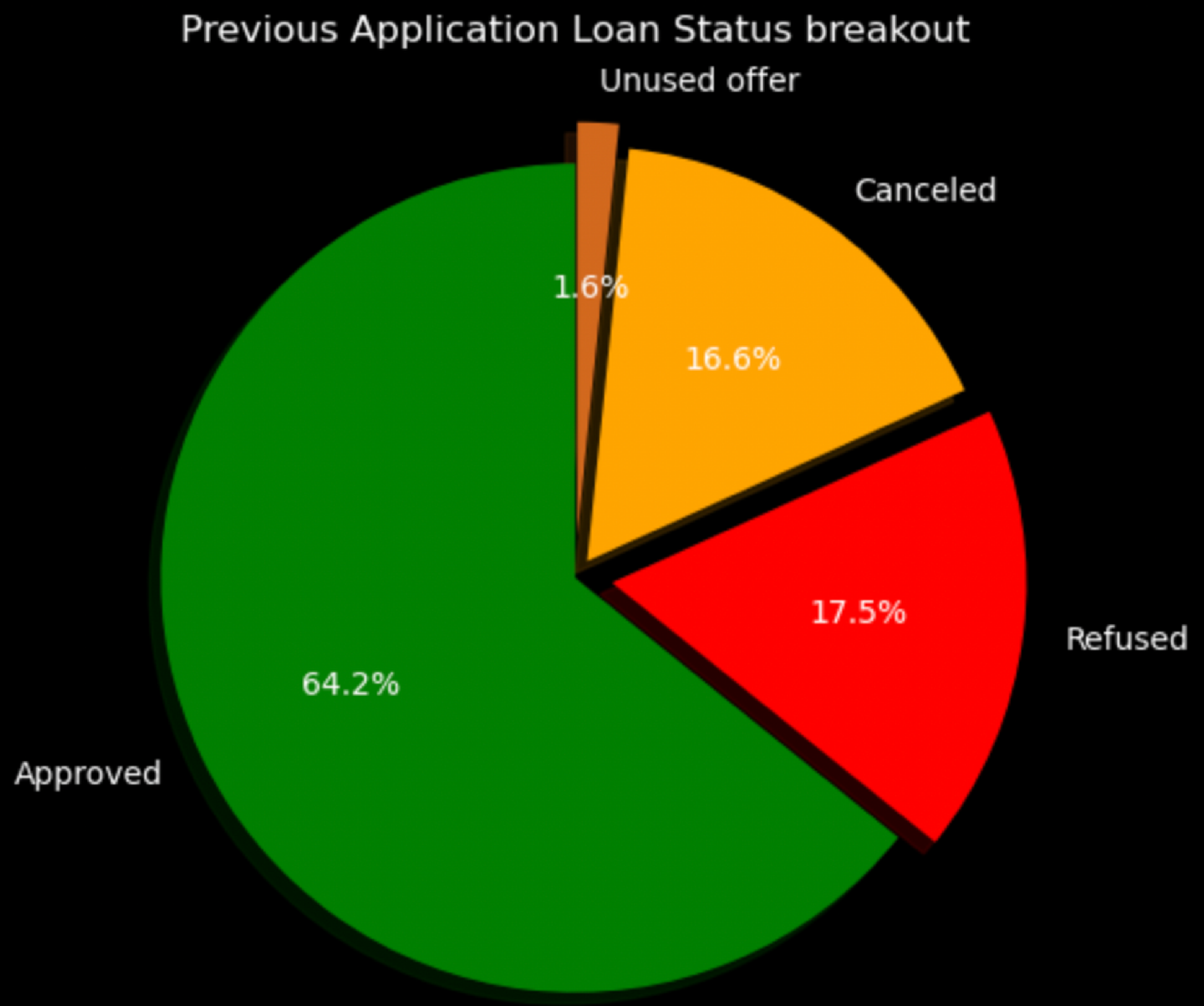


- Both charts show a similar correlation between multiple variables.
- All the positive linear correlations for defaulters are slightly less than that of Non-Defaulters.



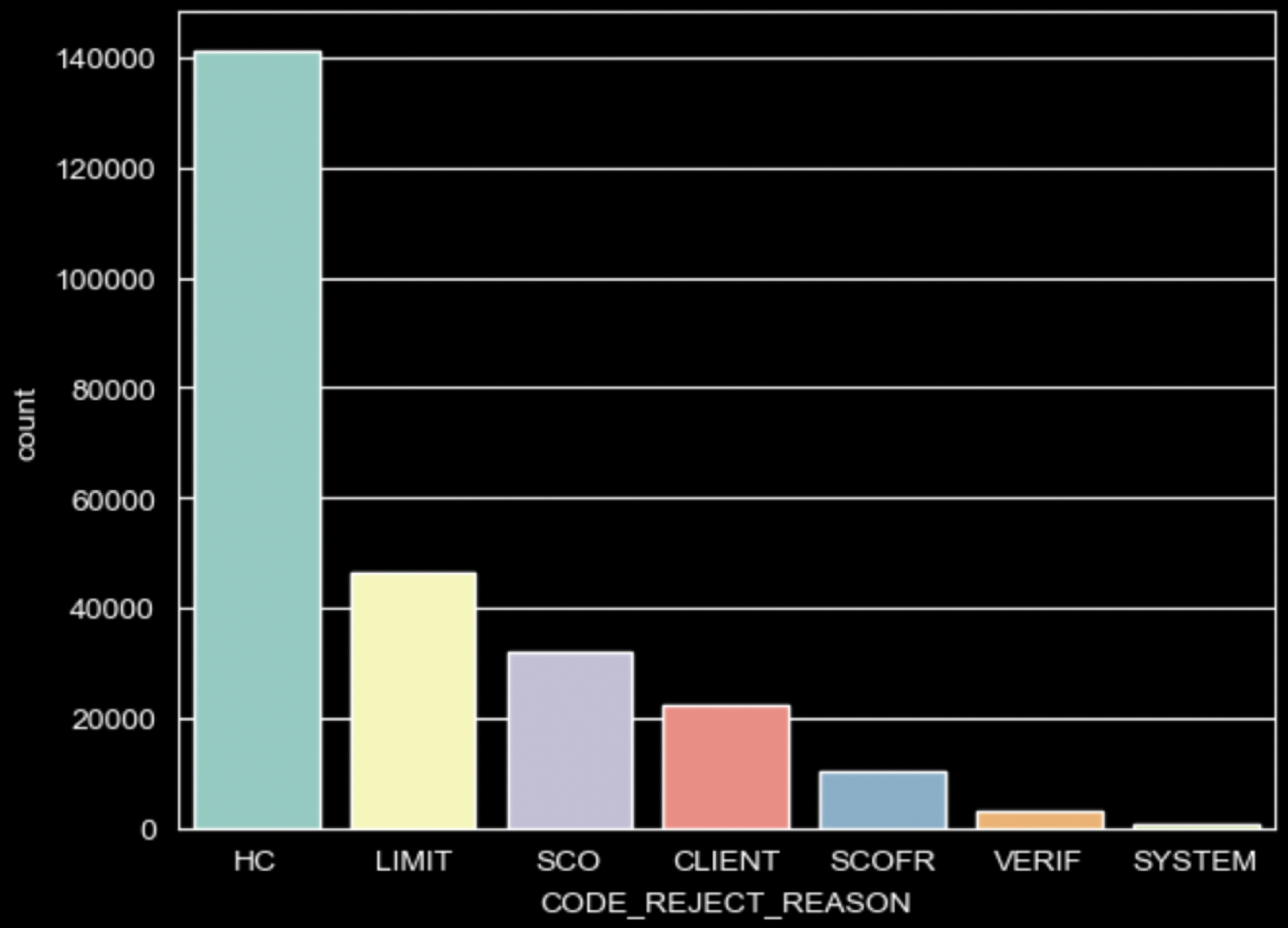
Defaulter’s Correlation between Numerical Variables

- There are no defaulters where the annuity is greater than 1.5 lakhs and the price of goods is more than 30 lakhs.
- There are fewer defaulters where the loan amount and goods price are more than 30lakhs
- We have a very strong positive linear relationship between the loan amount and the price of goods against which the loan has been taken. Both increase simultaneously.
- Another positive linear relationship is between an annuity and the price of goods against which the loan has been taken.
- Also, there is a positive linear relationship between loan amount and annuity.



Contract status of previous applications

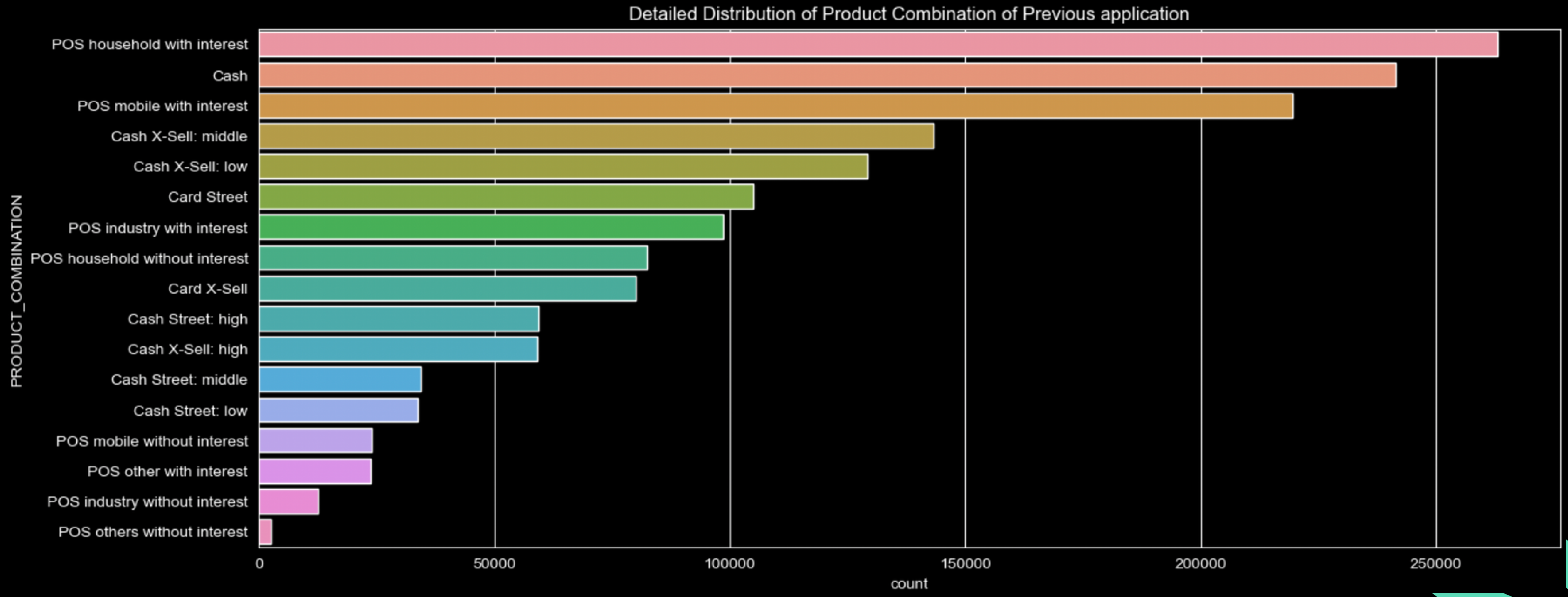
- For 64% of previous loan applicants, loan status was approved by the bank.
- 17.5% were Refused, 16.6% canceled and only 1.6% unused offer.



Distribution of Loan Rejection Reason by Bank

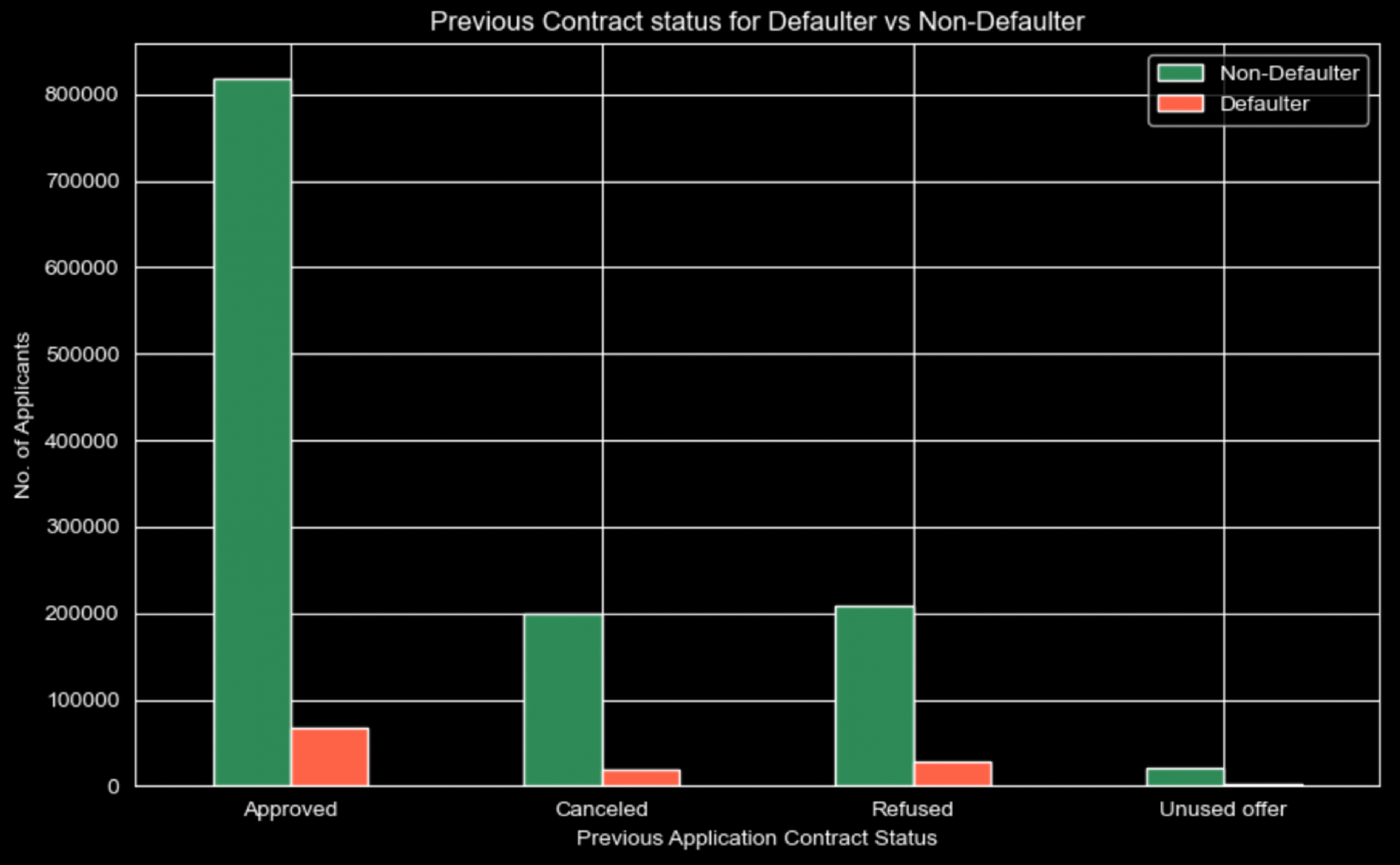
- Most applications have been rejected by banks because of 'HC' i.e. High risk.
- The least is due to the system default.

Distribution of Product Combinations of Previous Applications



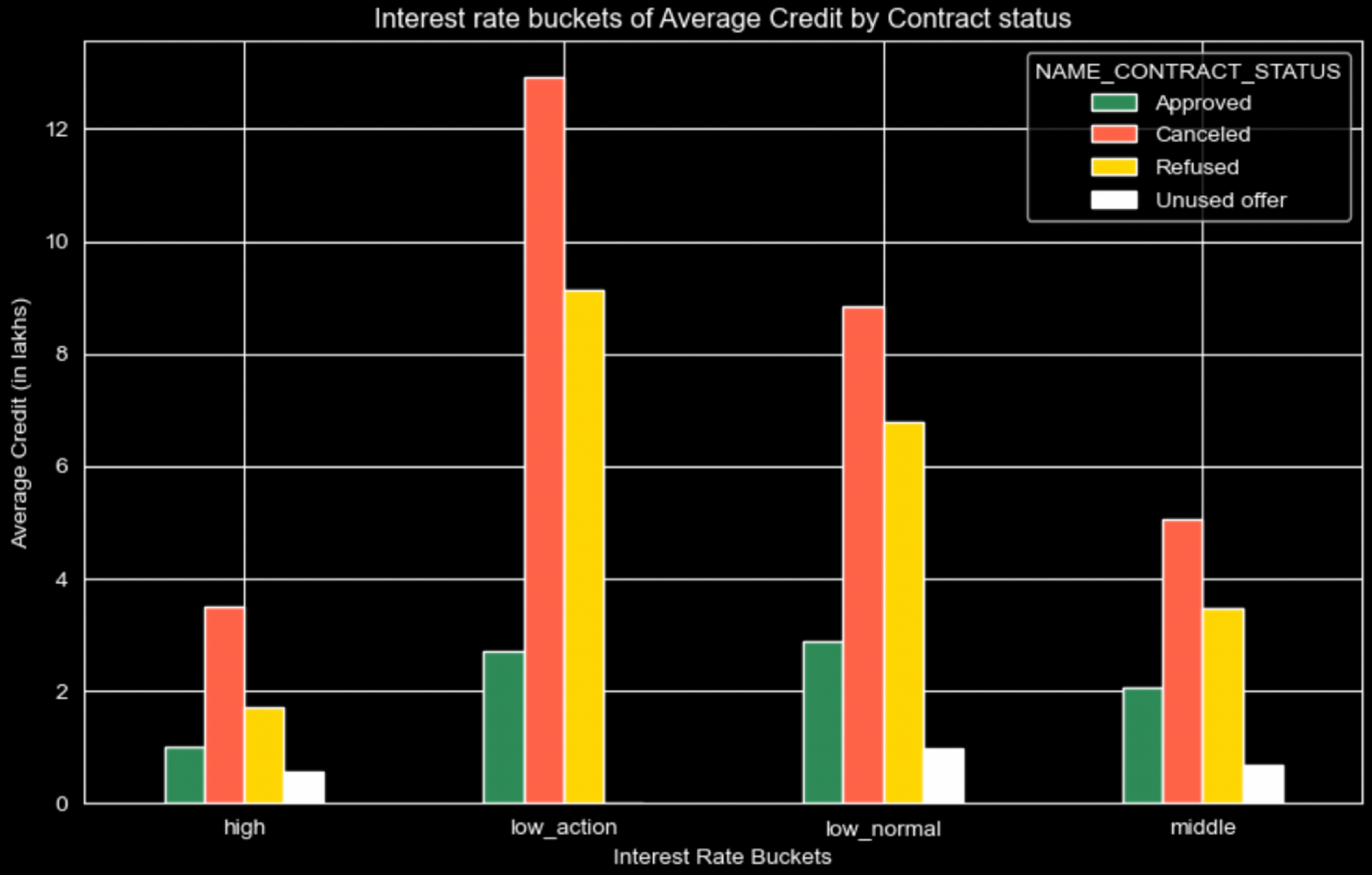
- Most products are in 'POS household with interest' followed by 'Cash'.
- Least in ‘POS others without interest’ and ‘POS industry without interest.’

Previous Contract Status for Non-Defaulters and Defaulters



- 90% of loans that were previously canceled by clients are currently being paid by them on time.
- 88% of loans that the company refused previously for the client are now being paid on time by the client.
- 91% of loans that the client canceled at some stage are also now being paid on time by the client.

Average Credit's Contract Status by Interest Rate



- In all interest buckets, the average credit for loans canceled by clients is very high.
- There are no unused offers in the low_action interest rate bucket.

Factors Influencing Good Loans

- People with academic degrees have the lowest default rate.
- Students and Businessmen have no defaults. One thing to note is that we have few students and businessmen in our dataset, so saying anything in general for businessmen or students would be wrong.
- People living in the city having a rating of 1 have the lowest percentage of defaults.
- People above the age of 50 have the lowest probability of default.
- People above 40 years of work experience are least likely to default.
- The average total income of people who don't default is more than those who default. Above 7 lakh of annual income, there are fewer chances of a person defaulting.

Factors Influencing Bad Loans

- Credit score is lowest for people in ages ranging 20-30. They tend to default more as compared to other age groups. People with low credit scores have a high chance of defaulting.
- Men even though less in number as compared to women in our dataset, are more likely to default than women.
- Unmarried people (single ones) are more likely to default(percentage-wise) than married people.
- People with lower secondary education have the highest default rate.
- Unemployed people are more likely to default than employed.
- People who live in a city with a rating of 3 have the highest percentage of defaults.
- Laborers have had the most defaults. So, giving loans to laborers should be avoided.
- People in Construction and Restaurant organization types are most likely to default.
- People with only 2-6 years of work experience have high default rates.
- Applicants with more family members i.e. 5 and 5+ have the highest default rate of 10%.
- People who have taken credit of around 5 lakh are most likely to default.
- Defualters' debt-to-income ratio i.e. percentage of income that goes to pay off loan is higher than those of non-defaulters.

Important Insights

- 90% of loans that were previously canceled by clients are being paid by them on time now. This means the company should market more to such clients who are canceling the loans either by decreasing interest rates or giving some sort of offer because they have good potential to pay it off.

- 88% of loans that the company refused previously to the client are now being paid by the client on time. This means the company should not refuse to offer the loan before a proper analysis of the client’s background.

- 91% of loans that the client canceled at some stage later are also now being paid on time by the client. This means that the company should note down the reason for which the client refused and work on it in the future to increase profits.

- The credit score of defaulters at all age stages, for both males and females, is lower than those of Non-defaulters. So, proper background checks should be done for people with low credit scores by the company before giving them loans.

Steps to avoid Defaults

- People who get loans for nearly Rs 500,000/- are defaulting more than others. Increasing/decreasing the interest rates near this credit amount is the step that the company should take. Also, communicating with the client is advisable.
- Since 90% of the applicants have a total Income of less than Rs 300,000/- and these people have the most defaults, hence, the company should monitor their accounts regularly and take necessary actions for these people for eg:- flexible repayment plans in case of financial hardships.
- People with 5 or more family members have the highest default rate of 9.4%. Hence, the company should take the necessary actions to avoid it.
- Contract Status Type:

1. We have a high number of applicants canceling the loan in all buckets of interest.
2. Most people are canceling the loan where the interest rate is very low(low_action). Also, there are no unused offers in this bucket as well which means either people canceled in the starting or they took the loan. In between nobody canceled the loan.
3. Banks should keep low_normal interest rates to attract most people towards loans.

Summary

- The majority of applicants opt for cash loans over revolving loans, with only a small percentage choosing revolving credit.
- Despite gender imbalance, male applicants are defaulting more as compared to female applicants.
- Car and house ownership alone don't significantly impact default rates since a large portion of applicants don't own cars and 1 in every 3 people don't own a house.
- The number of children an applicant has doesn't seem to correlate with default rates.
- Whether an applicant is accompanied or not during the loan application process doesn't influence default rates significantly except if accompanied by children, then the default rate is the least.
- Occupation type does play a role, with laborers being more prone to default compared to other occupations.
- People with more family members, particularly those with five or more, are more likely to default.
- People living in a city with a rating of 3 have the highest default rate and those with a rating of 1 have the lowest default rate.
- Organization types like Construction and Restaurant, have higher default rates compared to others.
- People with academic degrees are least likely to default and people with lower secondary are most likely to default.
- Credit scores tend to be lower for defaulters across all age groups and genders compared to non-defaulters.
- High debt-to-income ratios are associated with higher default rates.
- Loan amount, installment amounts, and price of goods have a positive correlation with each other.
- People with fewer years of work experience and lower annual incomes are more likely to default.

ONE LINER SUMMARY

Overall, there's a higher likelihood of loan default among applicants with lower credit scores, lower incomes, and higher debt-to-income ratios.

Presented by Devesh Parashar

Thank you very much!