CREDIT EDA ASSIGNMENT

ANALYSIS FOR BANK LOAN DEFAULTERS AND RISK ANALYSIS

Submitted By:- Kartikey Shukla
Batch:- February 2022

Problem Statement

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

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

- The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

> - The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
> - All other cases: All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

- **Approved**: The Company has approved loan Application
- **Cancelled**: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
- **Refused**: The company had rejected the loan (because the client does not meet their requirements etc.).
- **Unused offer**: Loan has been cancelled by the client but on different stages of the process.

Objective of Analysis:

- *This analysis aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as:*

> **a)** denying the loan,
>
> **b)** reducing the amount of loan,
>
> **c)** lending (to risky applicants) at a higher interest rate, etc.

- *This will ensure that the consumers capable of repaying the loan are not rejected.*
- *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 utilise this knowledge for its portfolio and risk assessment.*

Dataset 1: “application_data.csv”

Dataset Overview:-

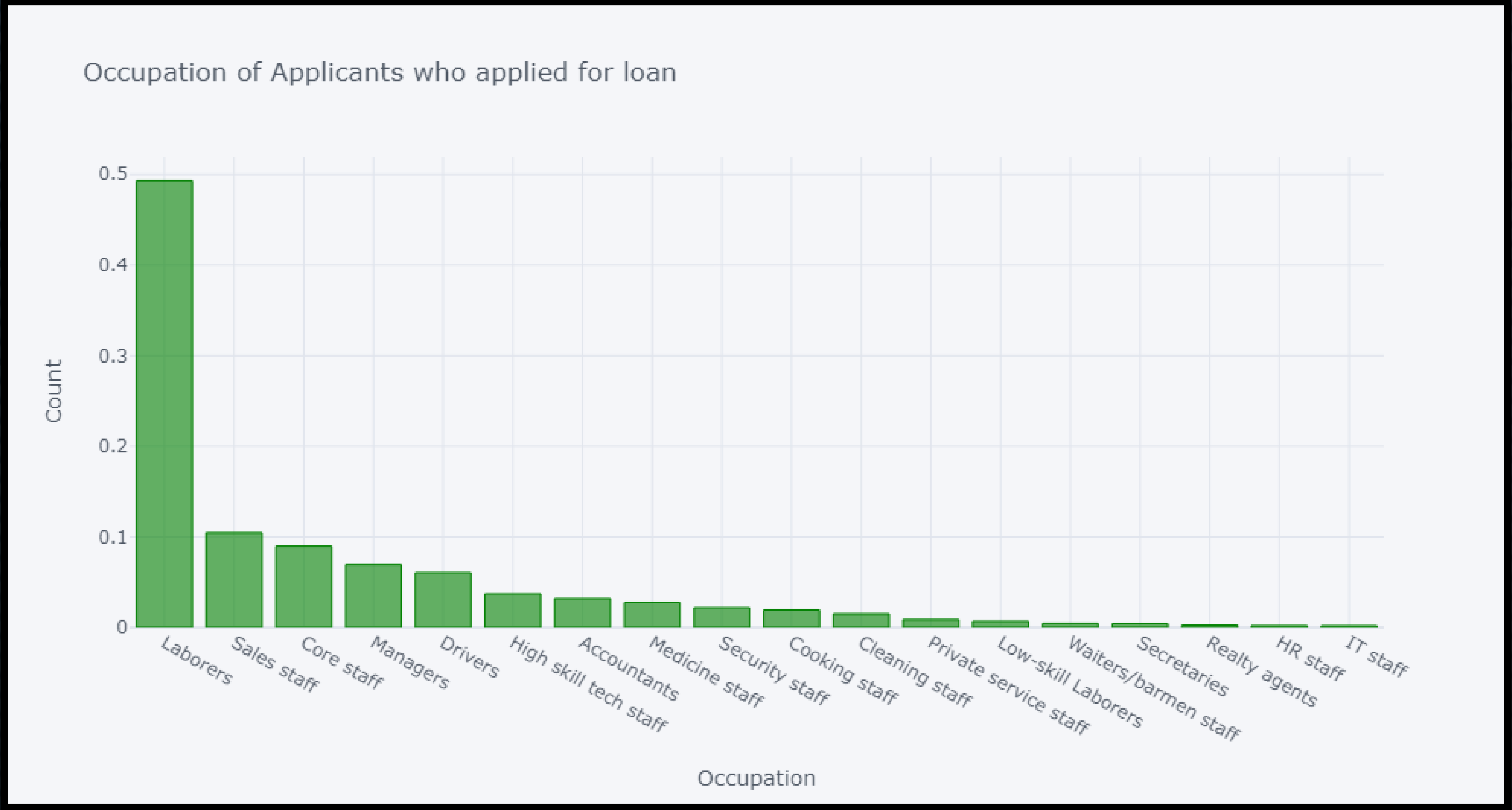
Number of variables	122
Number of observations	307511
Missing cells	9152465
Missing cells (%)	24.4%
Duplicate rows	0
Duplicate rows (%)	0.0%

Variable types

Numeric	70
Categorical	49
Boolean	3

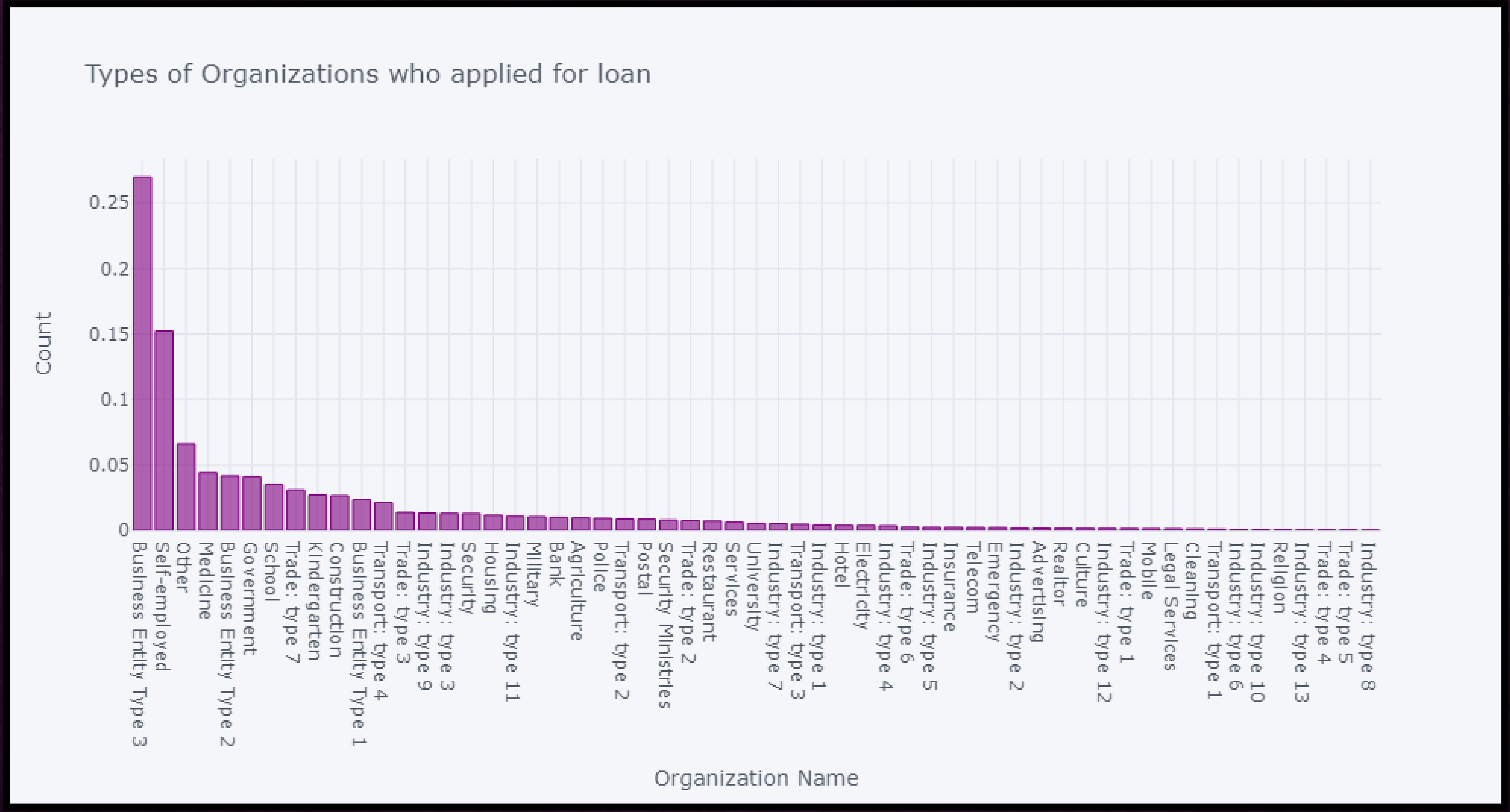
Exploratory Data Analysis

1. Distribution of Occupation Type of Clients Applying for Loan



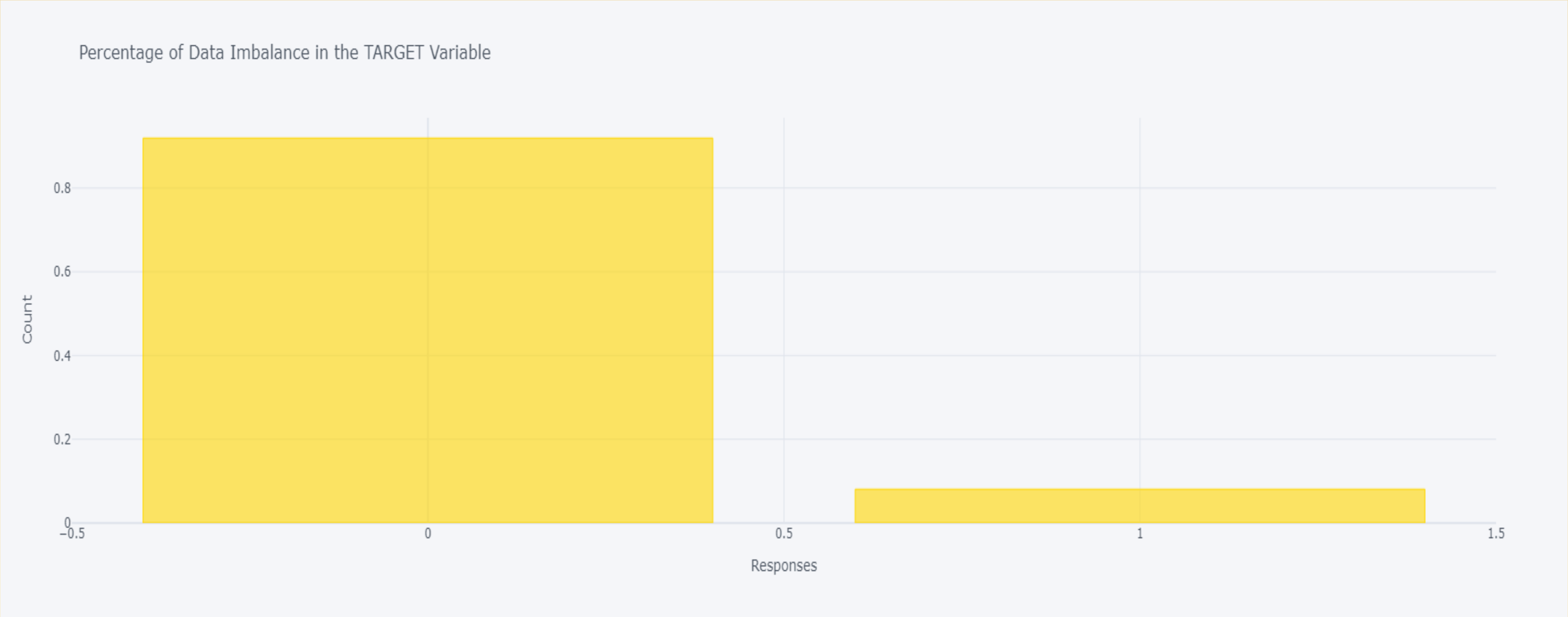
We can derive that majority of the clients applying for loan are "Laborers" followed by "Sales Staff". On the other hand, "IT Staff" & "HR Staff" are among the least.

2. Distribution of Types of Organizations applying for Loan



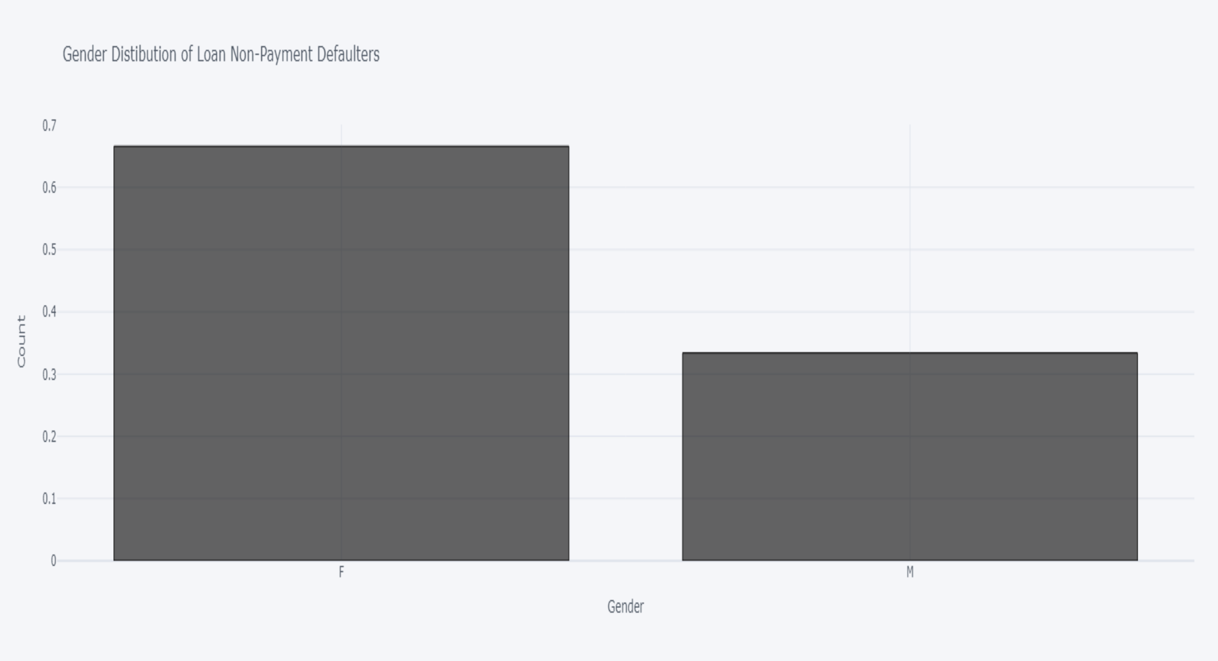
We can derive that majority of the Organizations applying for loan are “Business Entities” followed by “Self-Employed.” On the other hand, “Trade” & “Industry” are among the least.

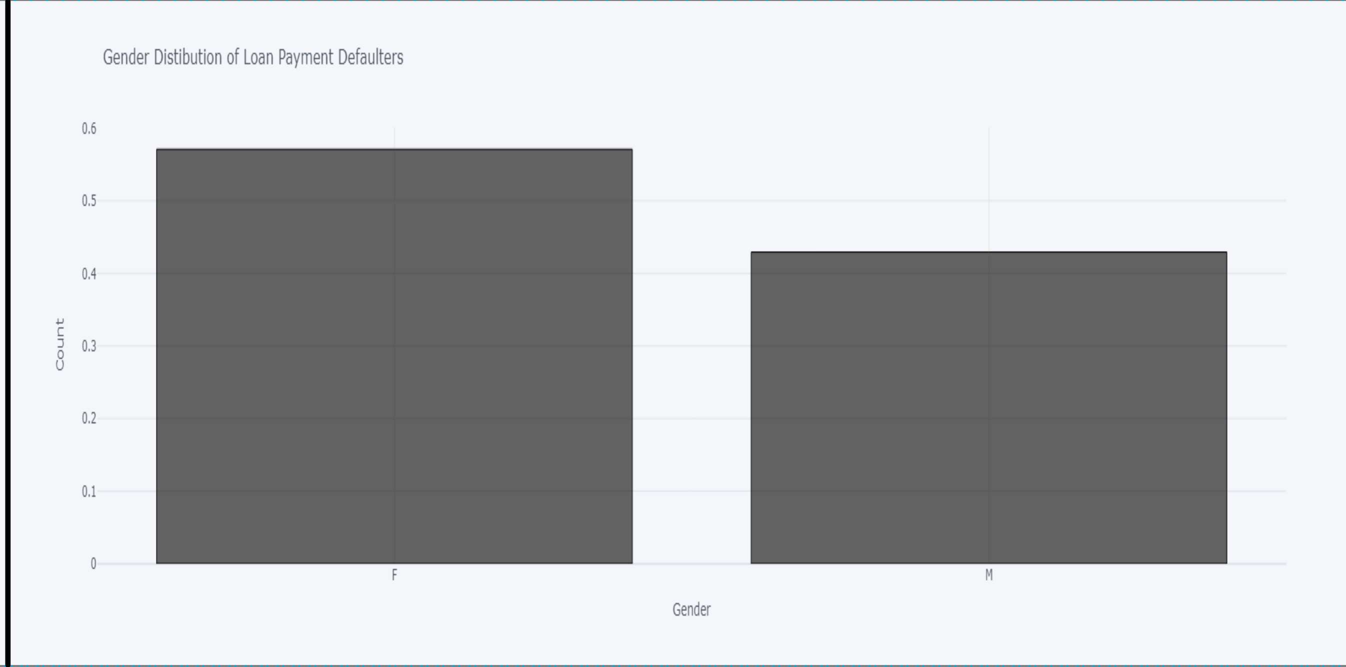
3. Imbalance in the Dataset



We can clearly see that there is a huge imbalance in the TARGET variable of the Dataset.
Imbalance ratio stands at :- 11:5

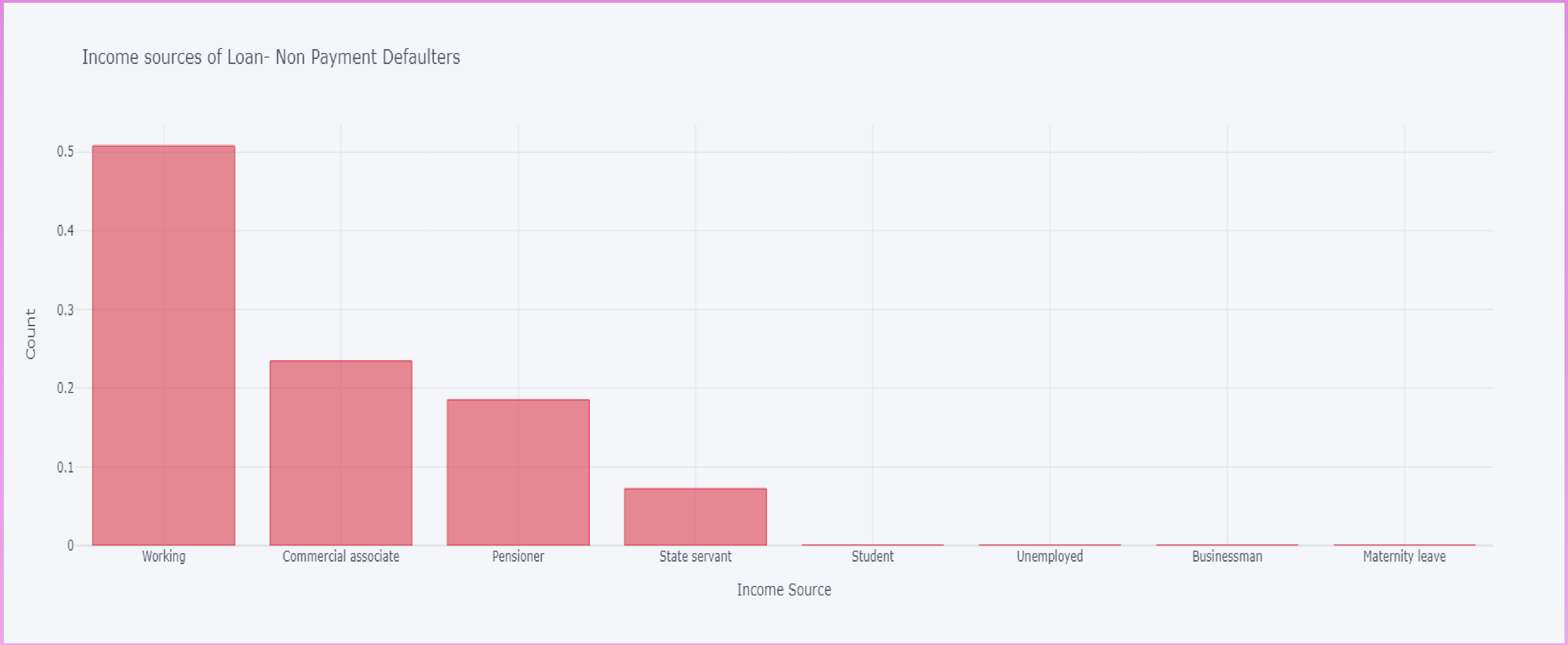
4. Gender Distribution of Loan Payment & Non-Payment Defaulters.



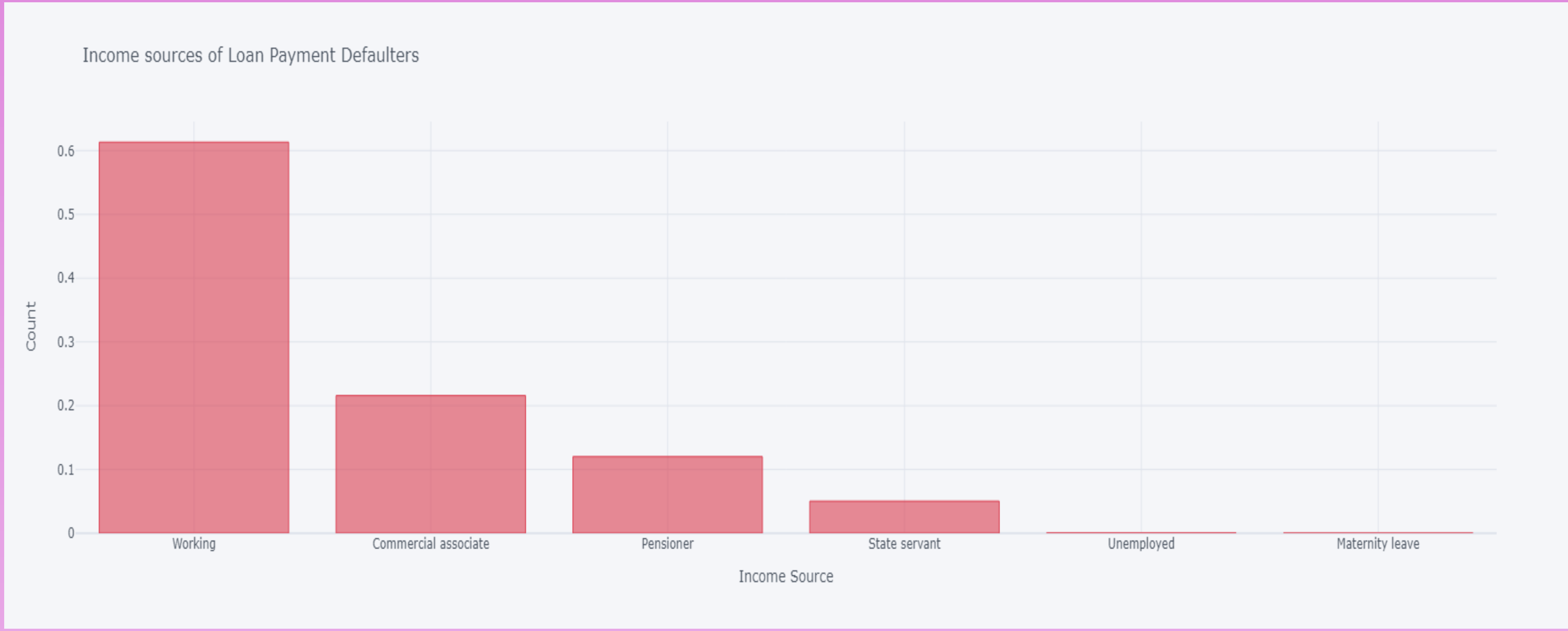


Comparing the Payment Defaulters and Non-Payment Defaulters on Gender basis, we observe that Females are the majority in both the cases although there is an increase in the percentage in Male Payment Defaulters from Non-Payment Defaulters.

5. Income Source Distribution of Loan Non-Payment Defaulters.

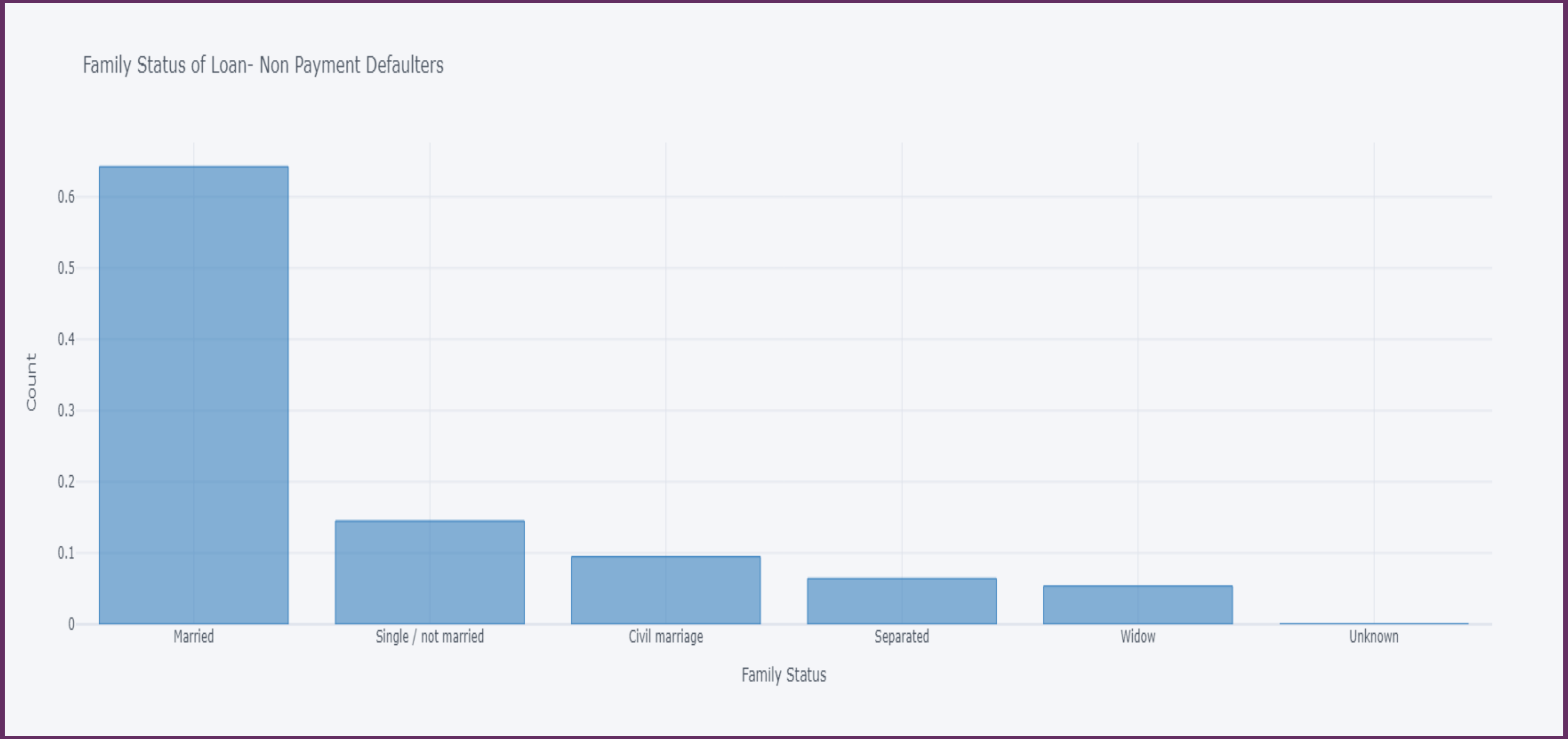


Income Source Distribution of Loan Payment Defaulters.

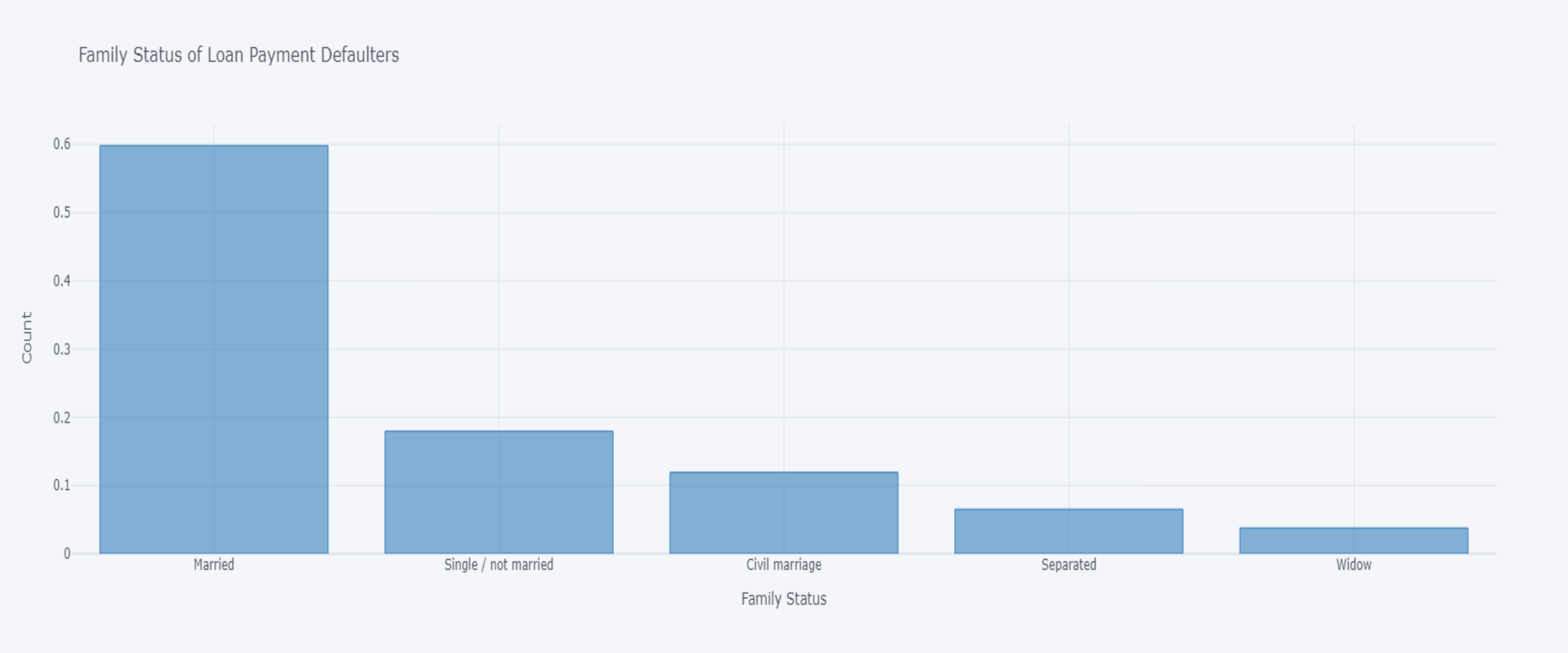


We observe a decrease in the percentage of Payment Defaulters who are pensioners and an increase in the percentage of Payment Defaulters who are working when compared the percentages of both Payment Defaulters and non-Payment Defaulters.

6. Family Status of Loan Non-Payment Defaulters

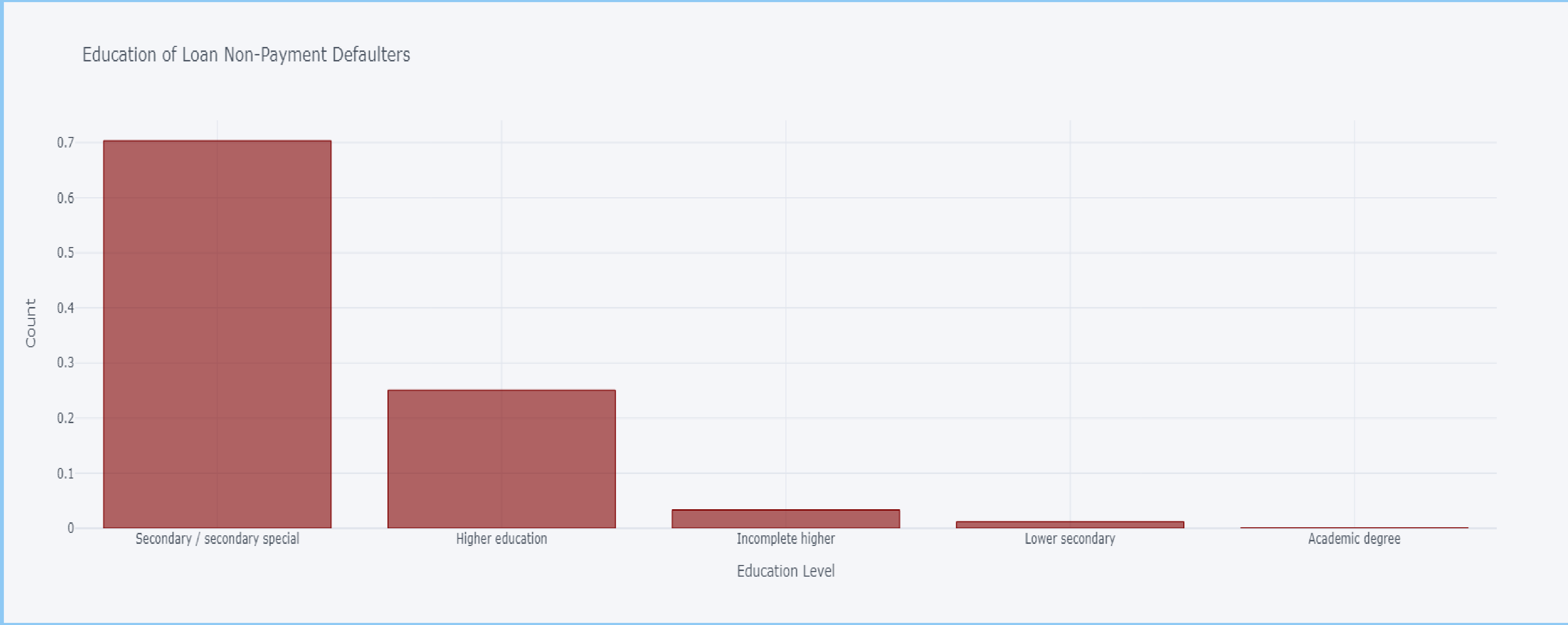


Family Status of Loan Non-Payment Defaulters

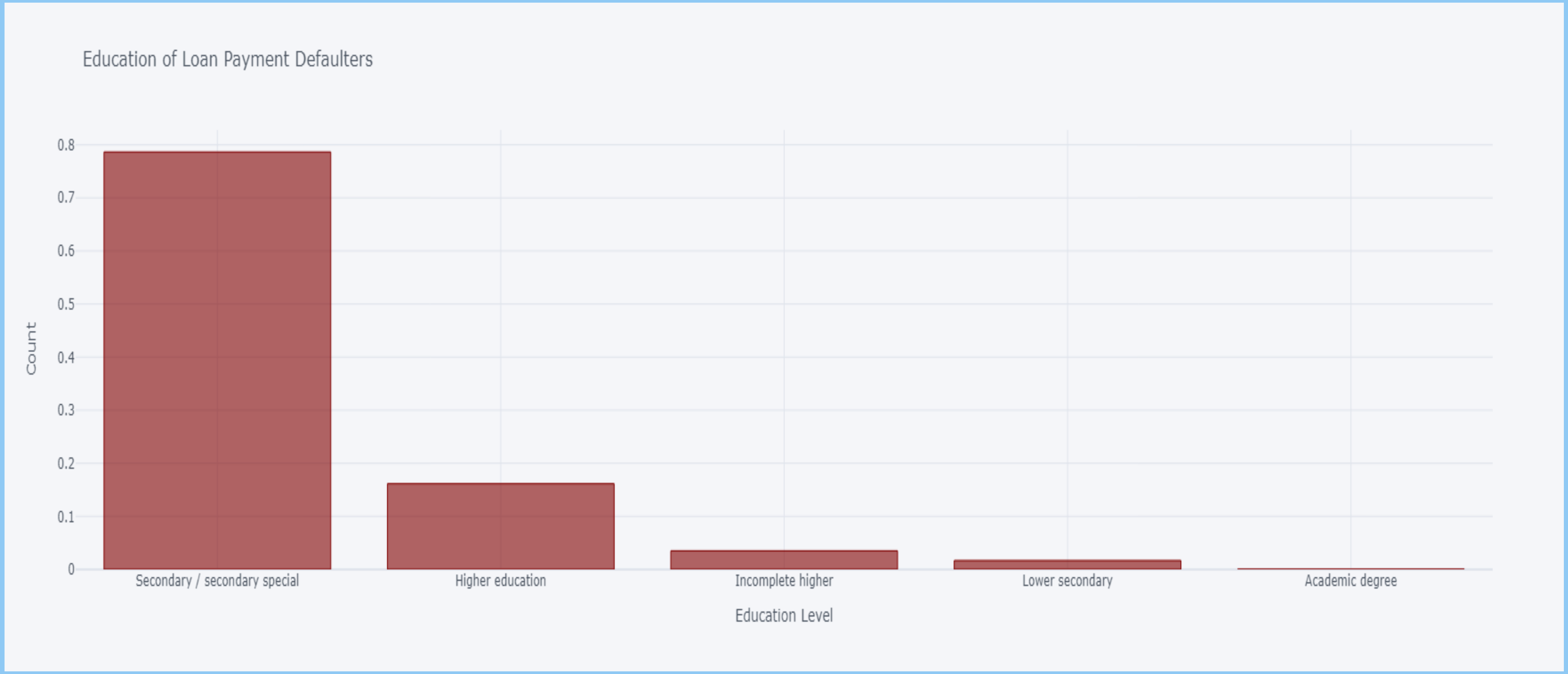


We observe a decrease in the percentage of married and widowed with Loan Payment Defaulters and an increase in the percentage of single and civil married with Loan Payment Defaulters when compared with the percentages of both Loan Payment Defaulters and Loan Non-Payment Defaulters.

7. Education Level of Loan Non-Payment Defaulters

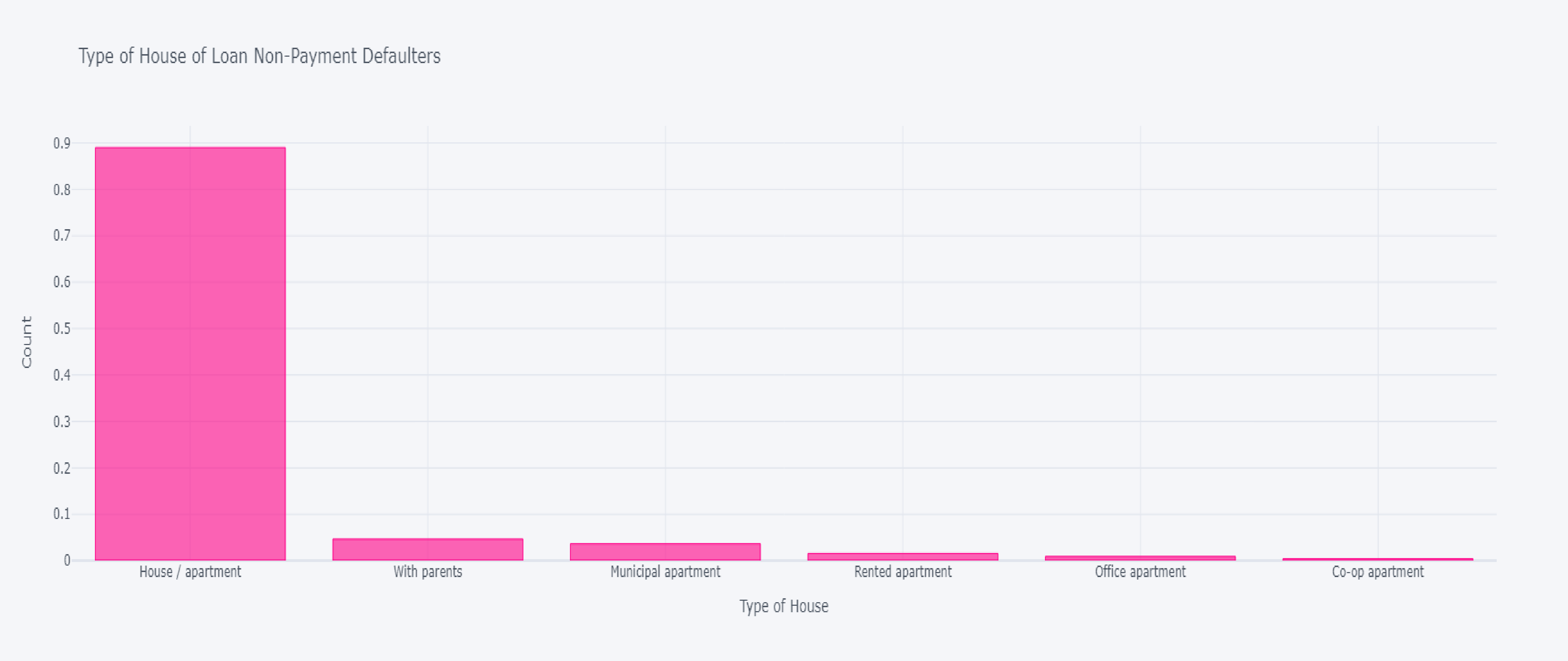


Education Level of Loan Payment Defaulters

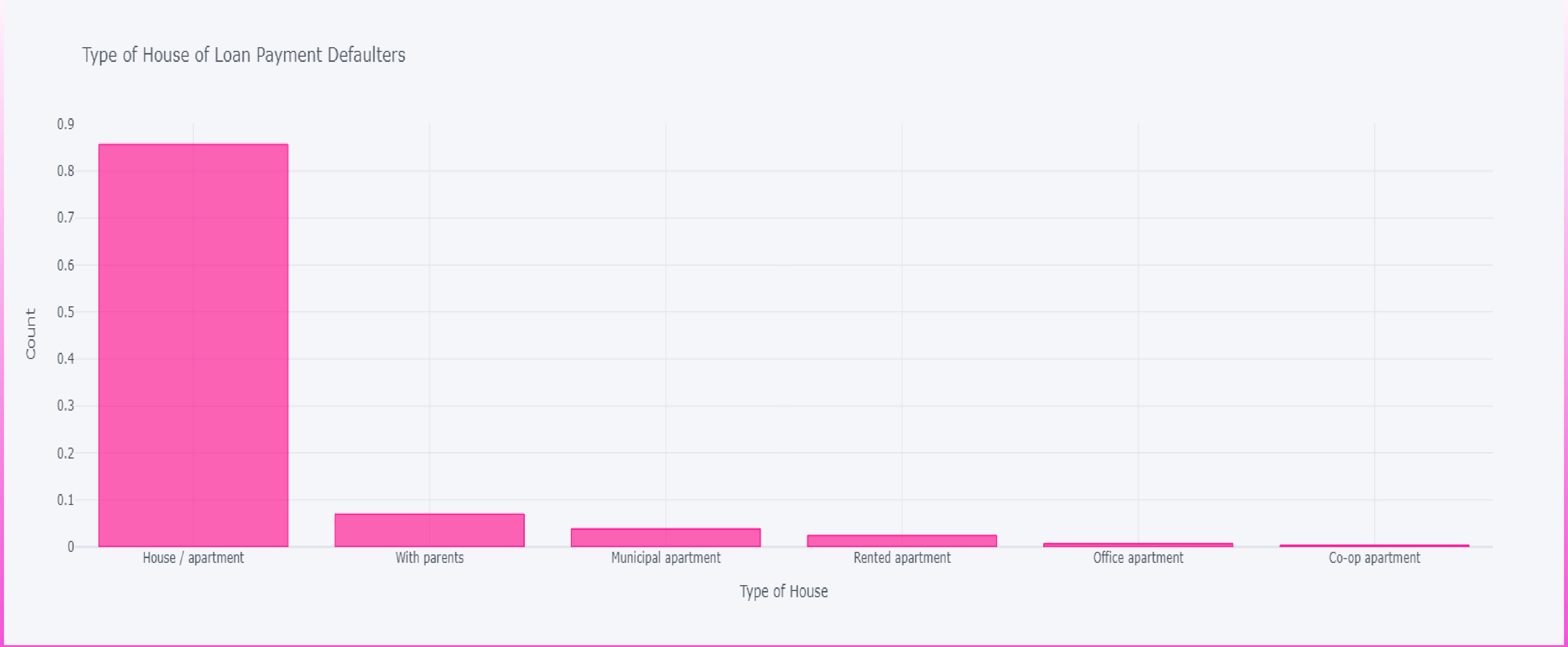


We observe an increase in percentage of Loan Payment Defaulters whose educational qualifications are secondary/secondary special and a decrease in the percentage of Loan Payment Defaulters who have completed higher education when compared with the percentages of Loan Payment Defaulters and Loan Non-Payment Defaulters.

8. Type of House / Residence of Loan Non-Payment Defaulters

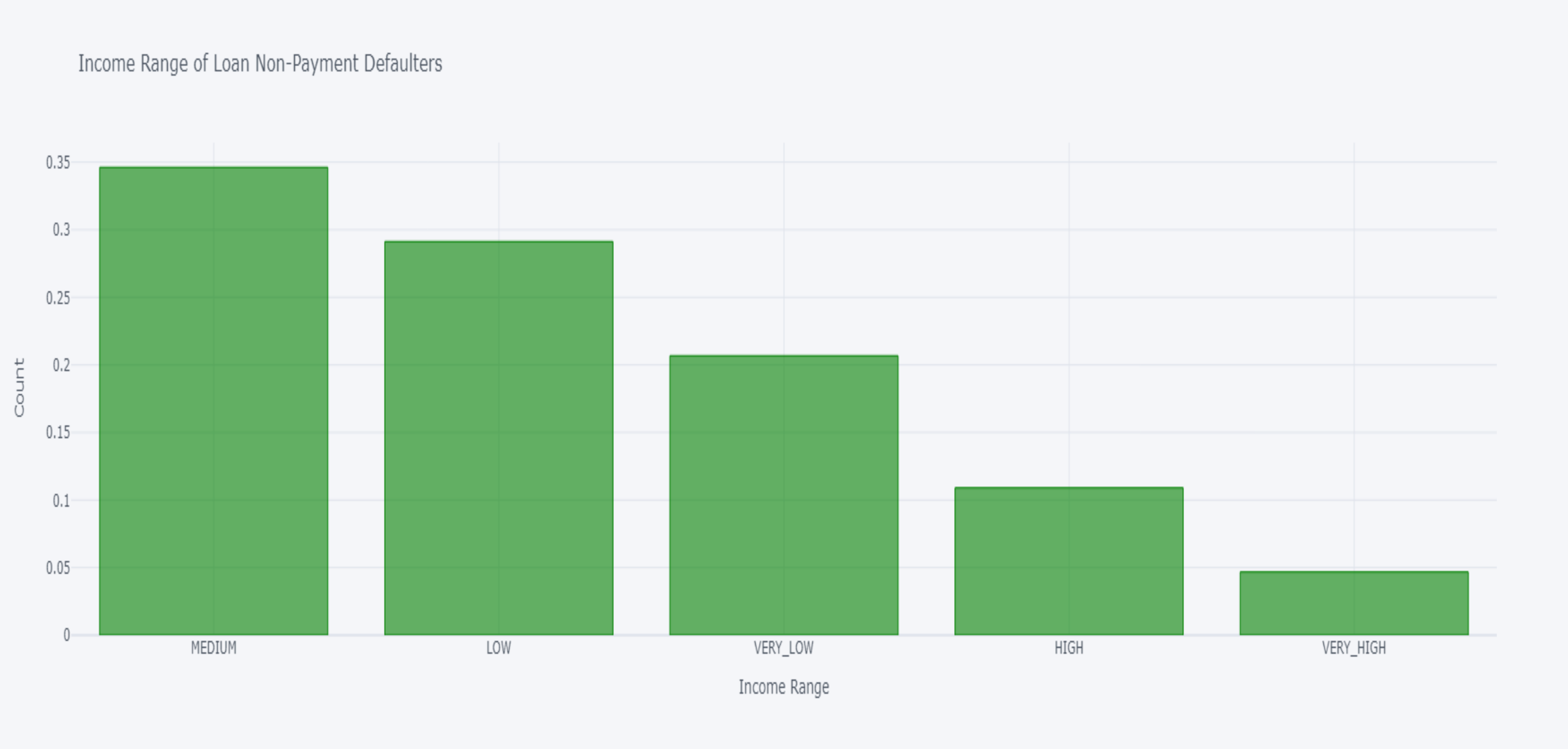


Type of House / Residence of Loan Payment Defaulters

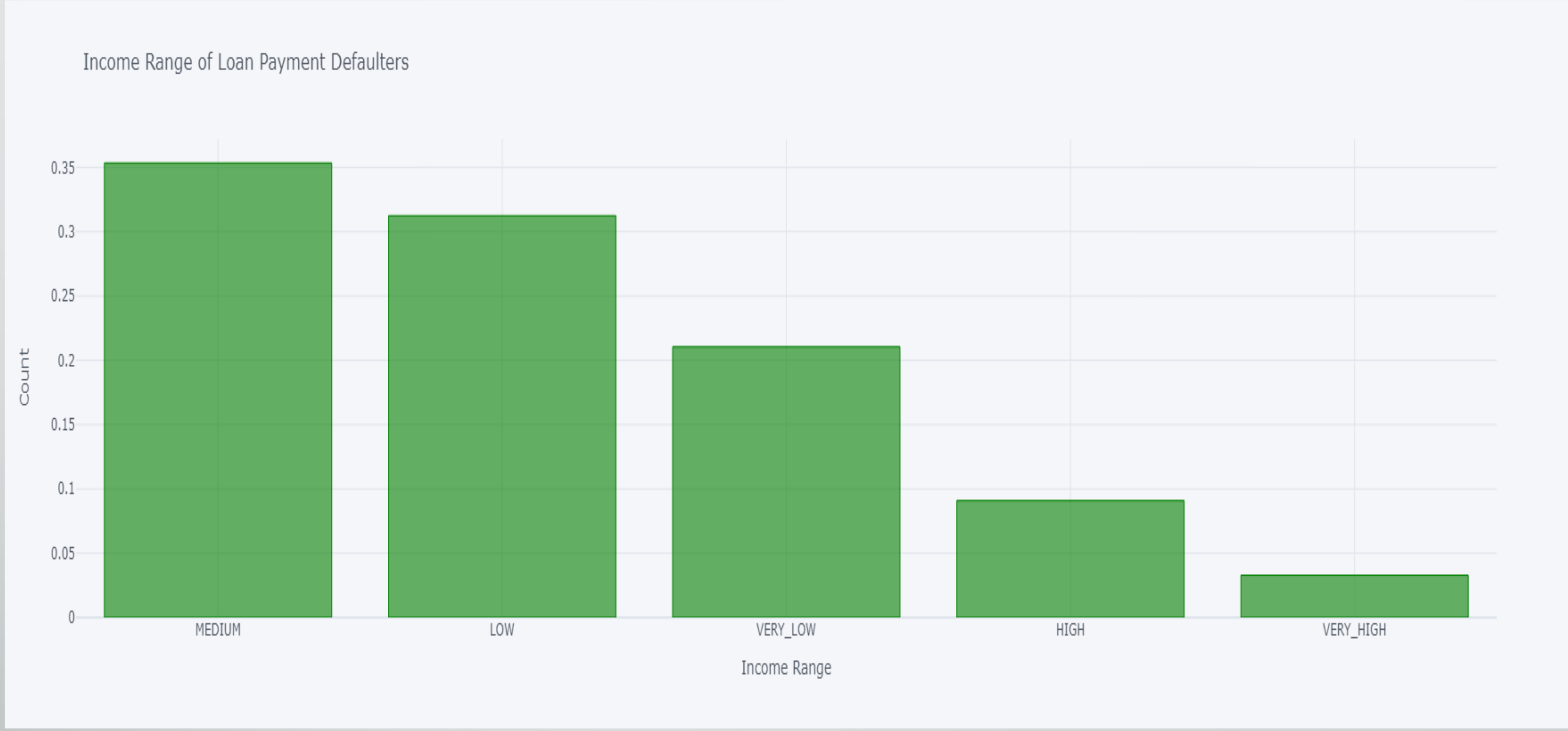


We observe an increase in the percentage of Payment Defaulters who live with their parents when compared to the percentages of Payment Defaulters and non-Payment Defaulters.

9. Income Range of Loan Non-Payment Defaulters



Income Range of Loan Payment Defaulters



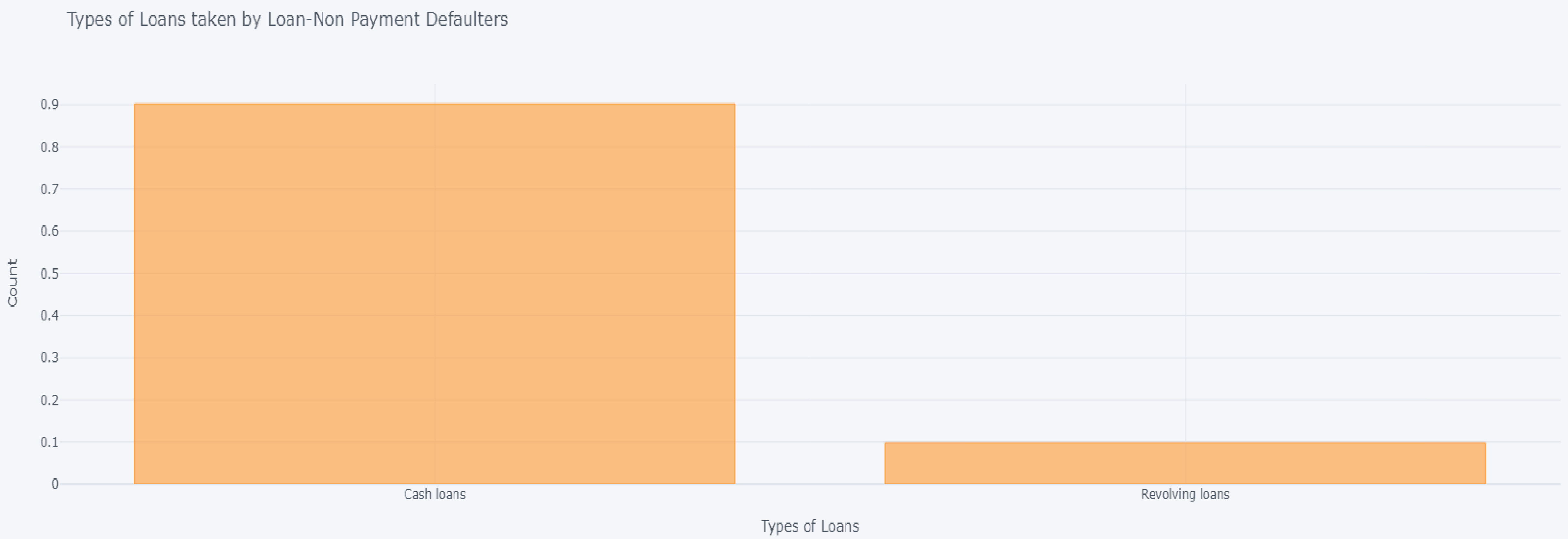
We observe an increase in the percentage of Loan Payment Defaulters whose income is low when compared with the percentages of Payment Defaulters and Loan-Non-Payment Defaulters.

10. Age of Loan Non-Payment Defaulters

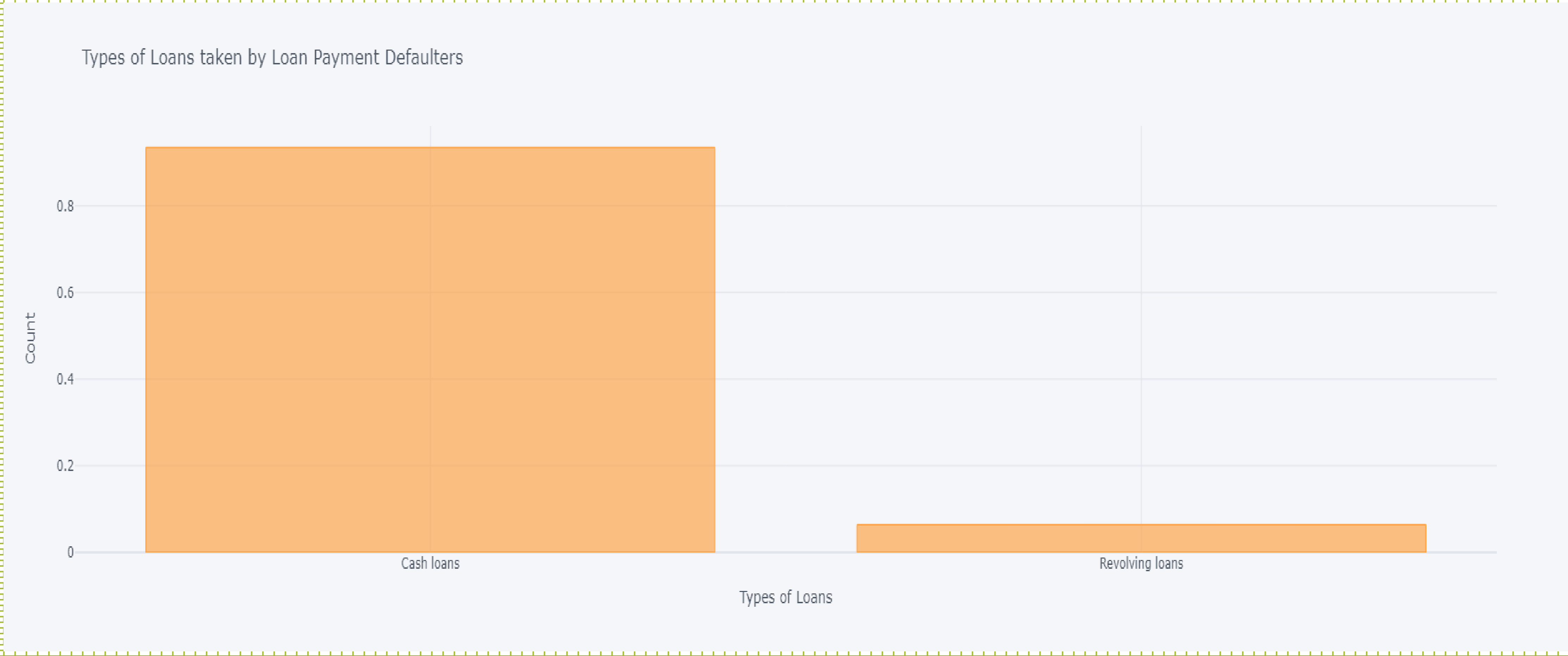
Age of Loan Payment Defaulters

We observe that there is an increase in the percentage of Loan Payment Defaulters who are young in age when compared to the percentages of Payment Defaulters and Loan-Non-Payment Defaulters.

11. Type of Loan taken by Loan Non-Payment Defaulters

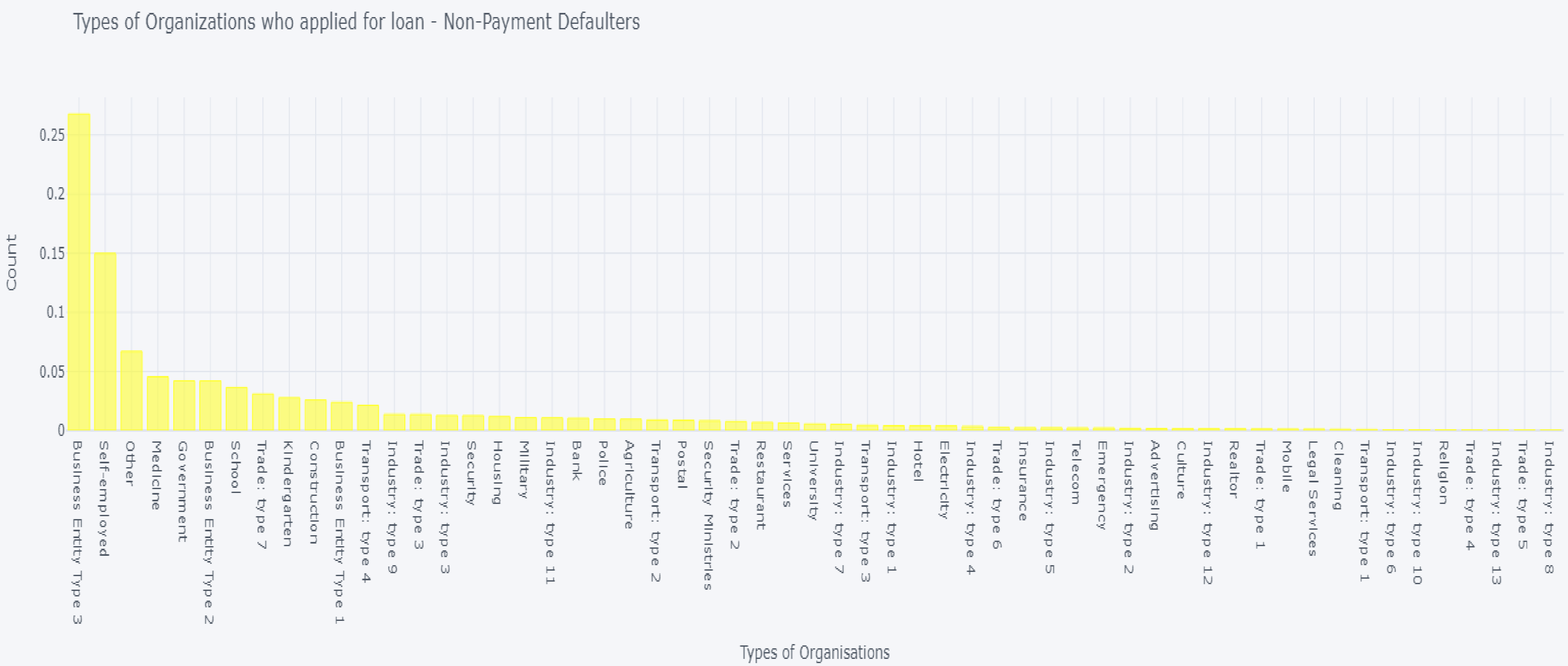


Type of Loan taken by Loan Payment Defaulters

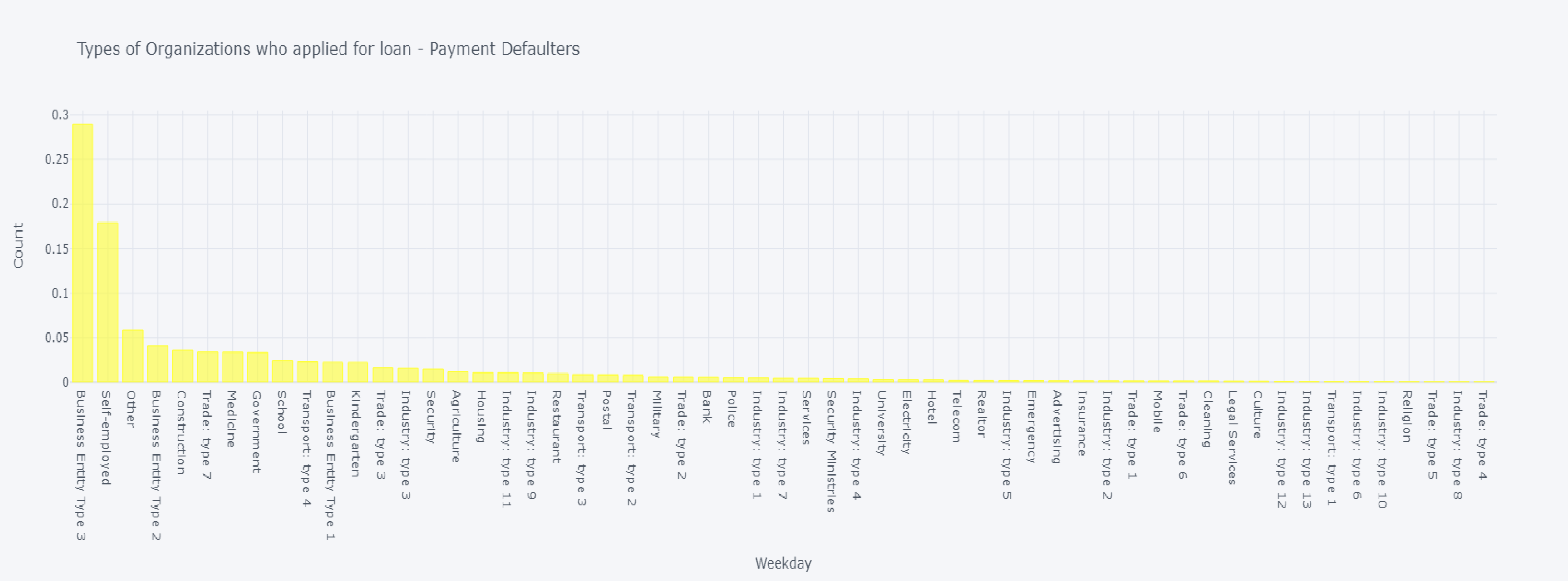


We can observe that cash loans are preferred by both Loan Payment Defaulters and Loan-Non-Payment Defaulters although there is a decrease in the percentage of Loan Payment Defaulters who opt for revolving loans.

12. Organization Type of Loan Non-Payment Defaulters

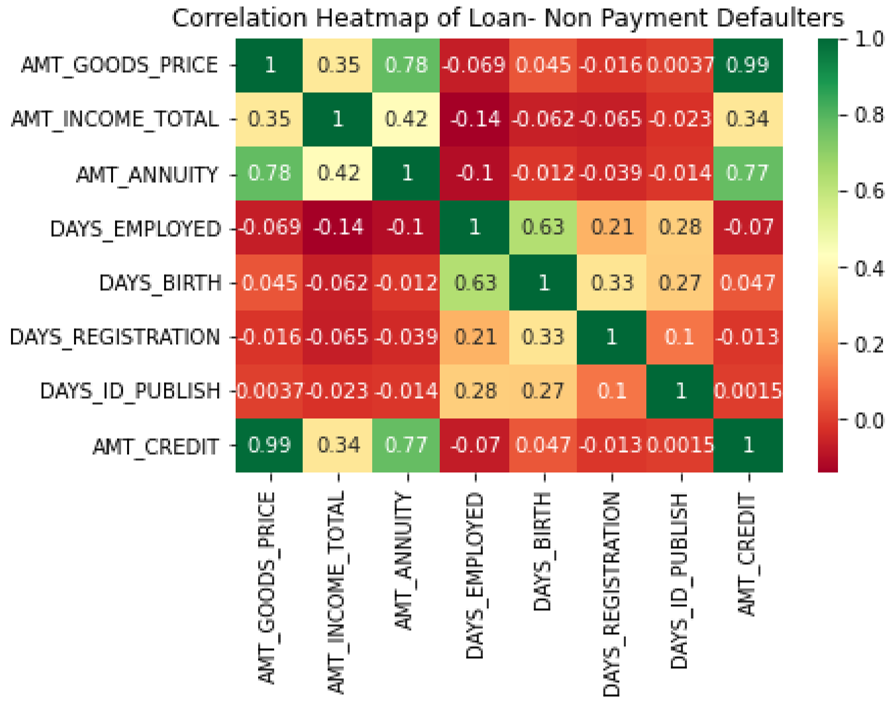


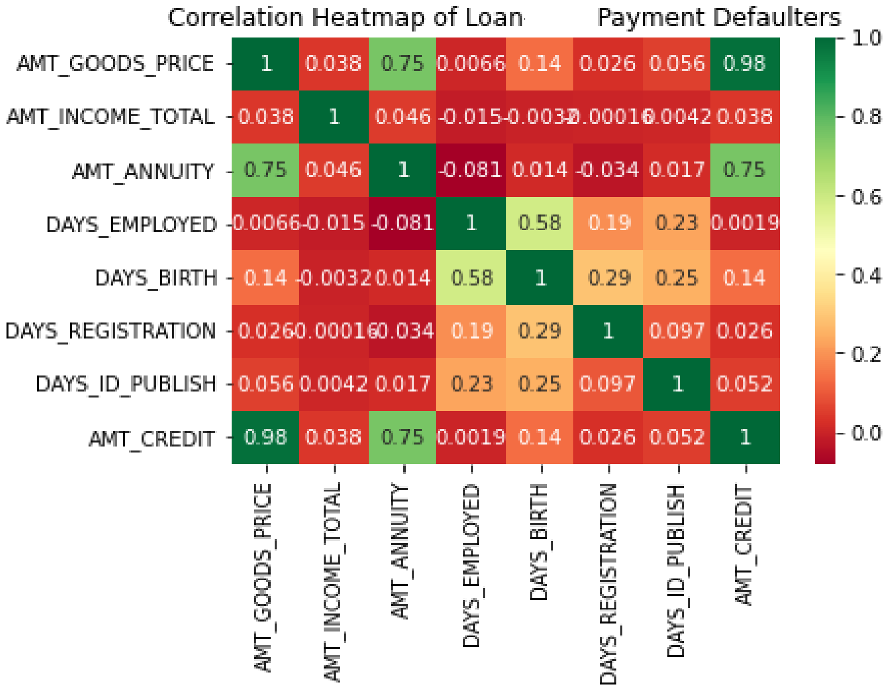
Organization Type of Loan Non-Payment Defaulters



Business Entity holds the highest percentage of both Loan Non-Payment & Loan Payment Defaulters. .

13. CORRELATION Heatmaps Between different variables for Loan Non-Payment Defaulters





- **Each square shows the correlation between the variables on each axis. Correlation ranges from -1 to +1. Values closer to zero means there is no linear trend between the two variables.**
- **The close to 1 the correlation is the more positively correlated they are; that is as one increases so does the other and the closer to 1 the stronger this relationship is.**
- **A correlation closer to -1 is similar, but instead of both increasing one variable will decrease as the other increases.**

Dataset 2: “previous_data.csv”

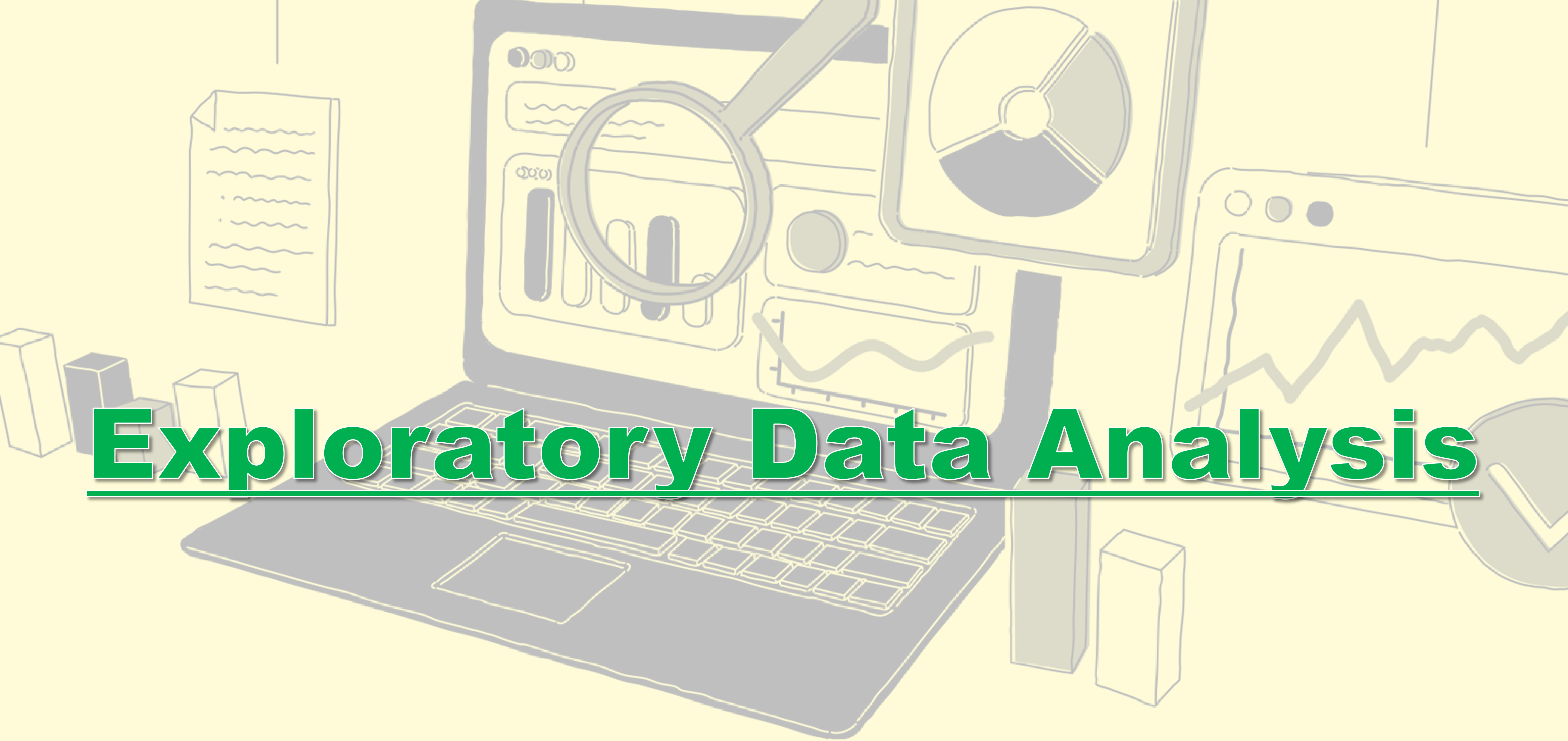
Dataset Overview:-

Number of variables	37
Number of observations	1670214
Missing cells	11109336
Missing cells (%)	18.0%
Duplicate rows	0
Duplicate rows (%)	0.0%

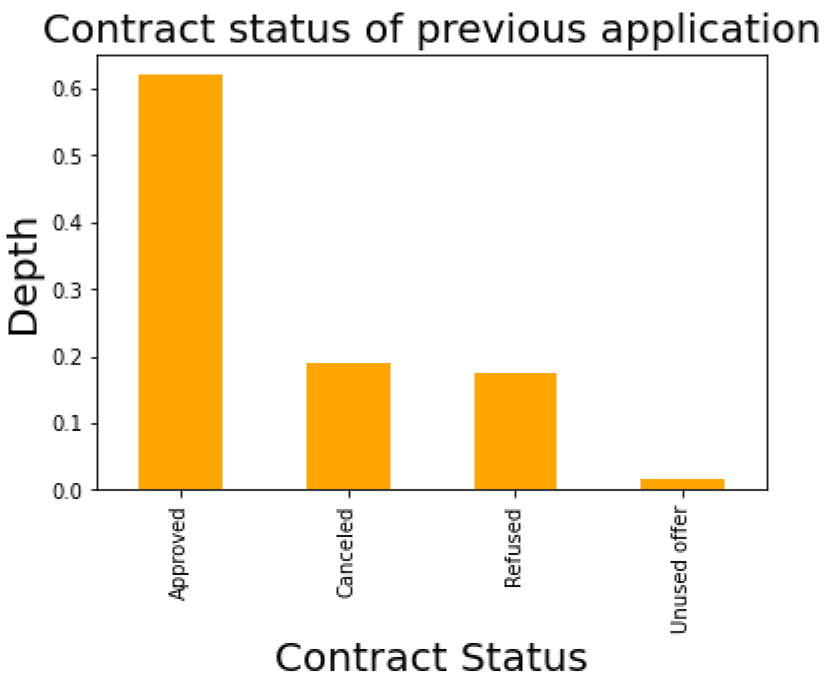
Variable types

Numeric	19
Categorical	17
Boolean	1

Exploratory Data Analysis

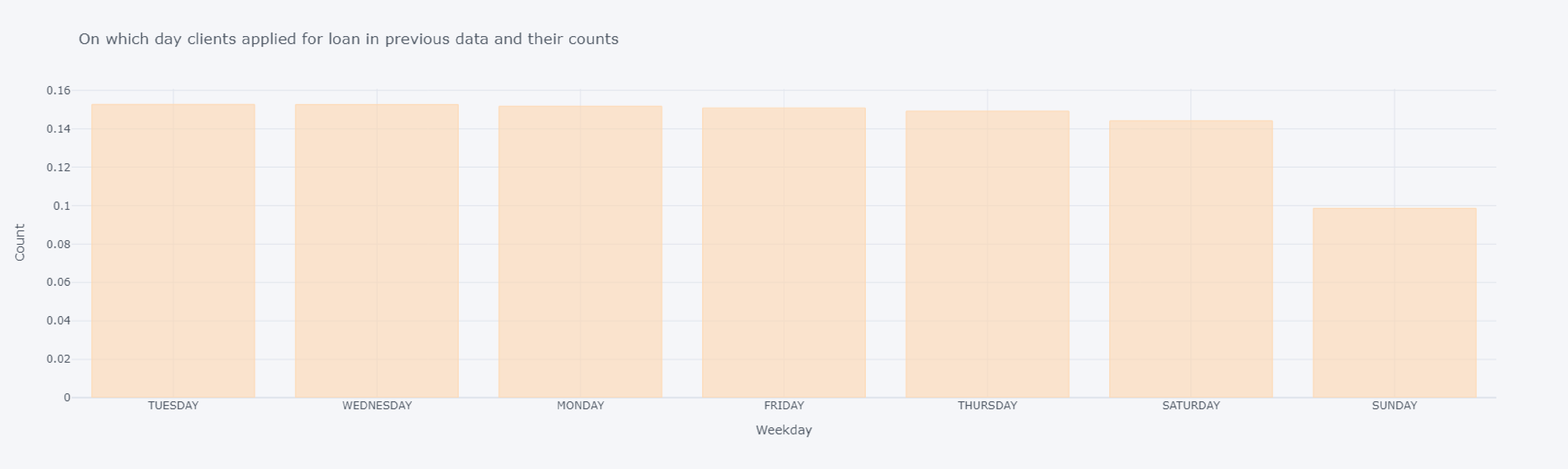


1. Contract Status



We can observe that majority of loans are approved and very less percentage of loans are unused offer in the previous application.

2. Day of the Week

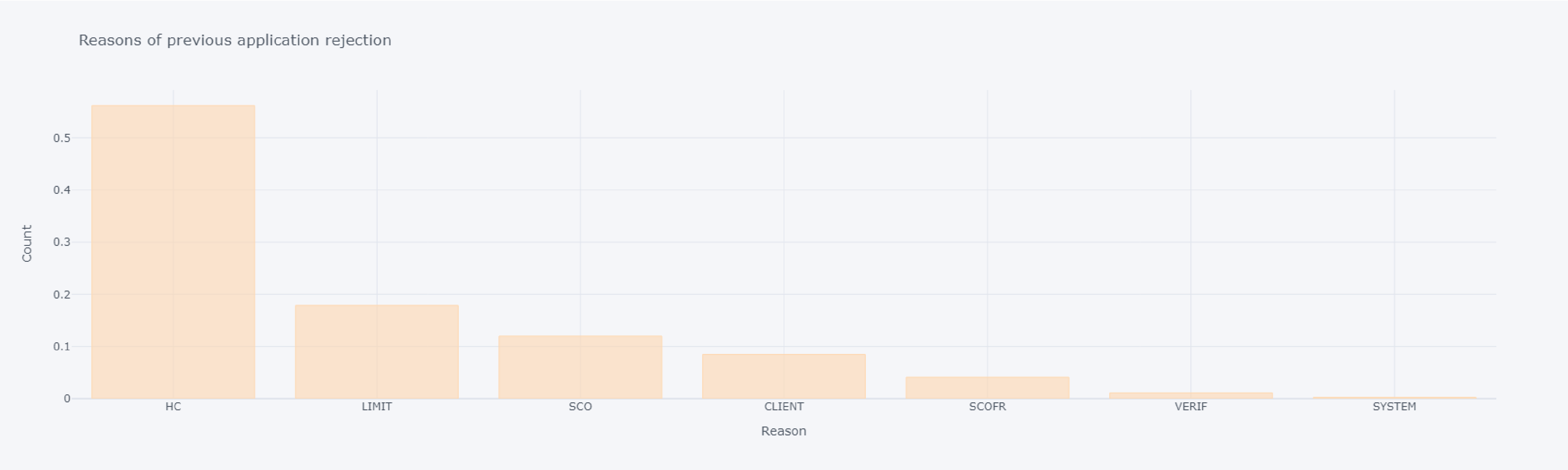


We observe that a high number of customers applied for loan on Weekdays & lower number of applicants apply in the weekends.

3. Payment Method

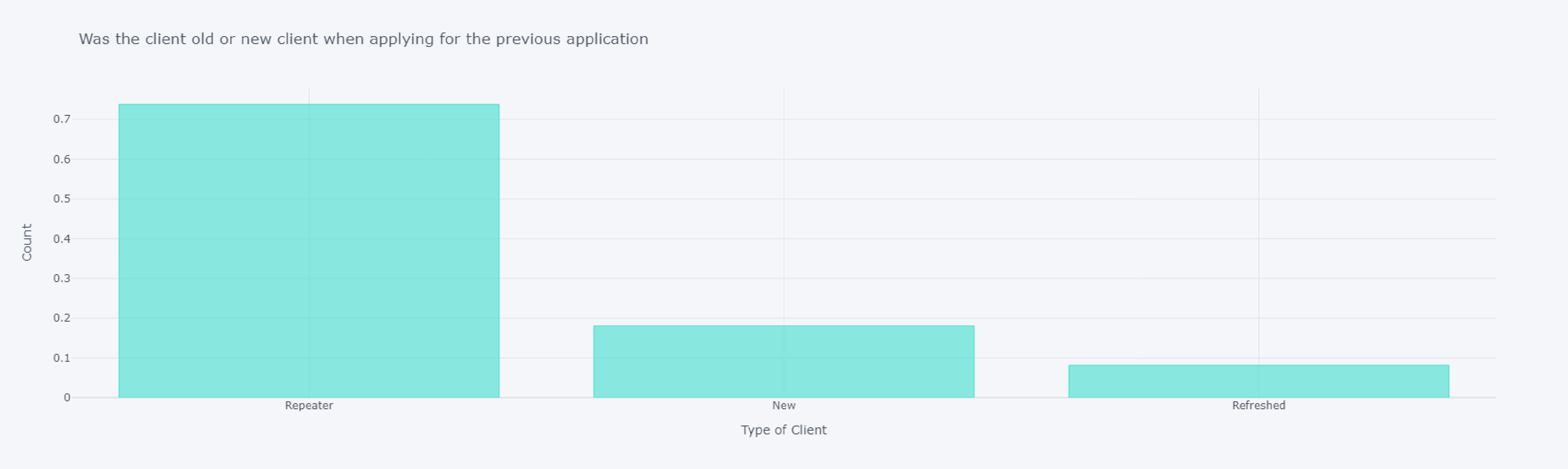
We observe that 99% of the clients chose to pay cash through bank and less than 0.1% of the clients chose to pay cashless from the account of the employer.

4. Reasons for previous application rejection



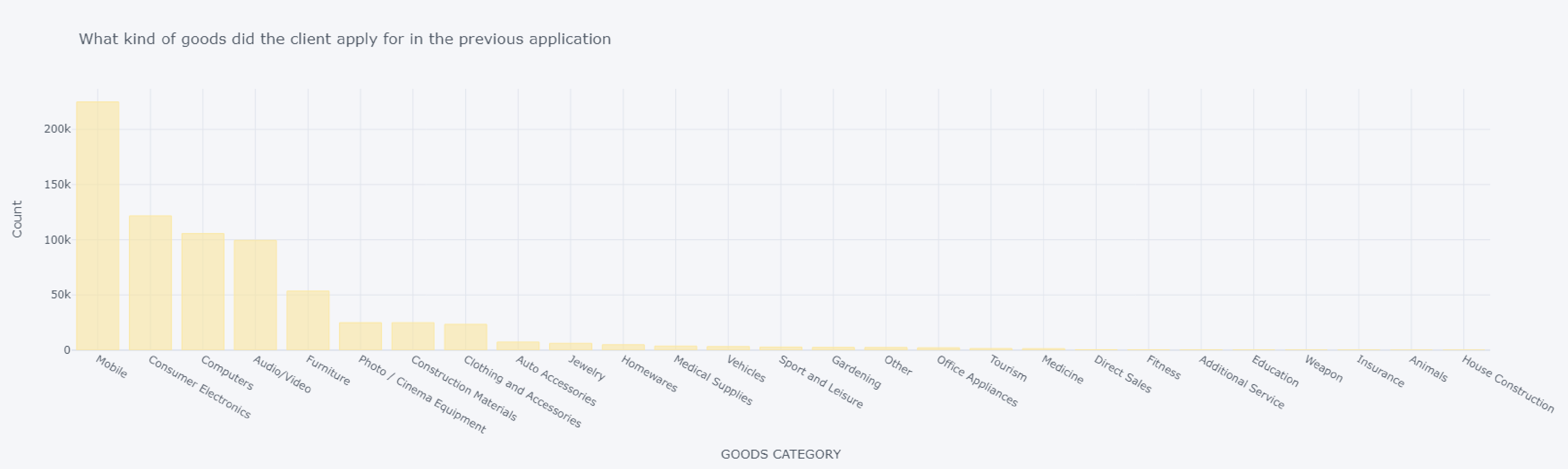
We observe that HC is the reason, majority of applications got rejected.

5. Client Type



We observe that majority of the clients are repeaters.

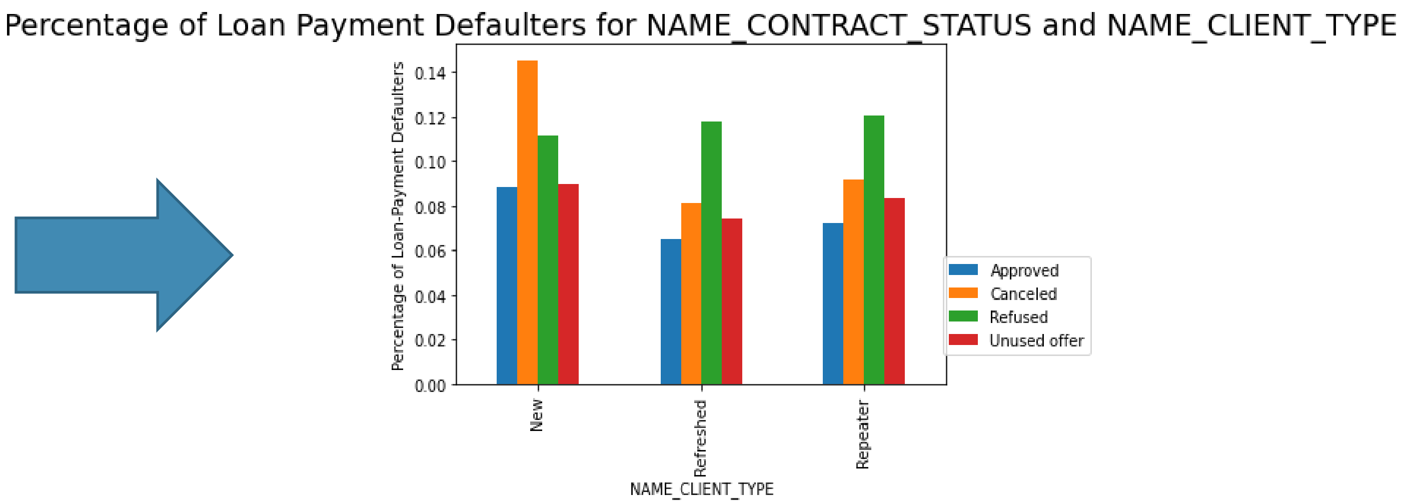
6. Type of Goods for which the loan was applied



We observe that majority of loans are for mobiles, consumer electronics, computers and furniture & very few are for Education, Fitness & House Construction.

7. Percentage of Loan Payment Defaulters with their Contract status and Client Type.

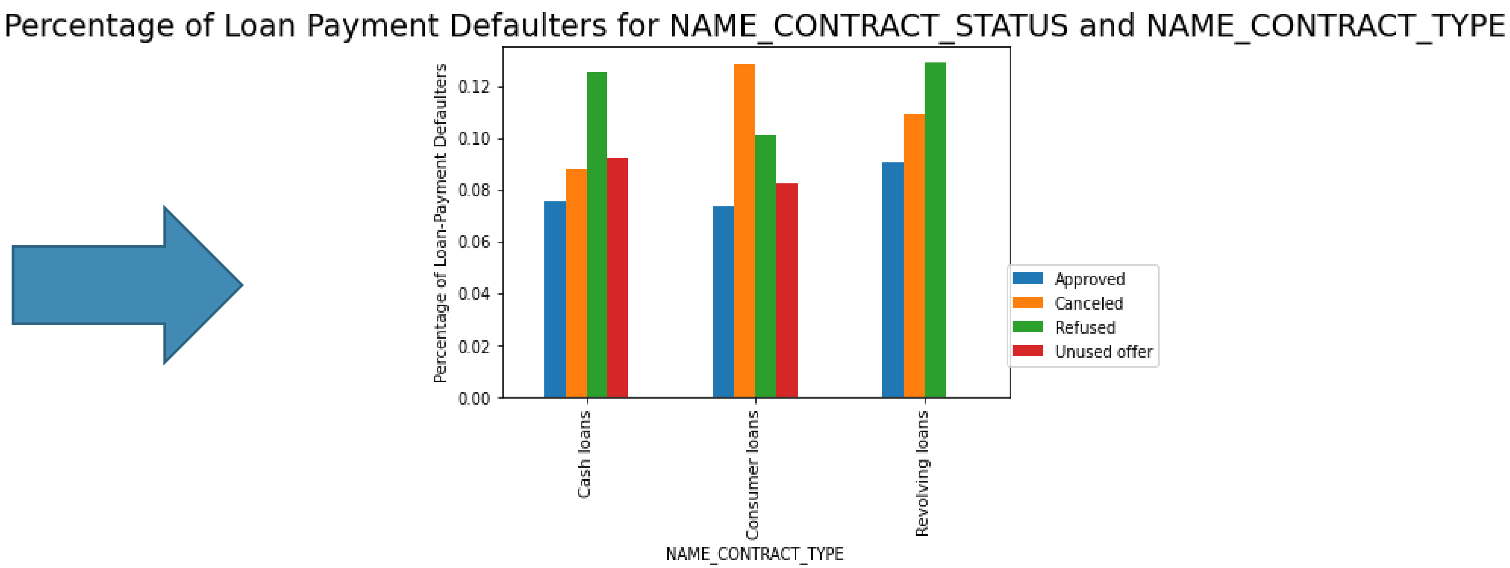
NAME_CONTRACT_STATUS	Approved	Canceled	Refused	Unused offer
NAME_CLIENT_TYPE				
New	0.088216	0.145205	0.110940	0.089448
Refreshed	0.065158	0.081098	0.117412	0.074324
Repeater	0.072144	0.091767	0.120596	0.083338



It can be observed from the above graph that Client who where 'New' and had 'Cancelled' previous application in Consumer Loans Category tend to have more percentage of Loan-Payment Defaulters in current application.

8. Percentage of Loan Payment Defaulters with their Contract status and Contract Type.

NAME_CONTRACT_STATUS / NAME_CONTRACT_TYPE	Approved	Canceled	Refused	Unused offer
Cash loans	0.075516	0.088401	0.12581	0.092593
Consumer loans	0.073853	0.128668	0.10135	0.082337
Revolving loans	0.090343	0.109254	0.12905	0.000000



It can be observed from the above graph that Clients with 'Revolving loans' and with 'Refused' previous application tend to have more % of Loan-Payment Difficulties in current application.

Conclusion

APPLICATION DATA

- Females account for a majority of Loan Defaulters.
- Bank should prefer clients with marriage status of Married and Widow, as there is a decrease in their percentage of loan payment defaults.
- Clients who live with their parents are the driving factors for Loan Defaulters.
- Clients with Income Range of Low & Medium are the driving factors for Loan Defaulters.
- Banks should prefer on clients with High or Very High Income Level.
- Bank should prefer clients with Higher Education Level.
- Clients with marriage status of Single & Civil Marriage are the driving factors for Loan Defaulters.
- Client with occupation type as 'Laborers' are the driving factors for Loan Defaulters.
- Client with education type as 'Lower Secondary' are the driving factors for Loan Defaulters.

PREVIOUS APPLICATTION

- Majority of the clients who applied for loans are Repeaters.
- Majority of previous loan applications are for Mobiles, Consumer Electronics, Computers & Furniture.
- Client with contract status as 'Refused' in previous application are the driving factors for Loan Defaulters.
- Client with contract type as 'Revolving loans' in previous application are the driving factors for Loan Defaulters.
- It can be observed from the graph that Clients with 'Revolving loans' and with 'Refused' previous application tend to have more % of payment difficulties in current application. Since the count of both 'Revolving loans' and 'Refused' is comparatively less(from the graphs in previous slide), clients with 'Revolving Loans' and 'Refused' previous application are driving factors for Loan Defaulters