Credit EDA Case Study

"Exploring Credit Data: Key Insights and Trends

Submitted by

Shravya Ramadugu

Purpose

To maximize profit, it is crucial to avoid two main issues:

▶ Approving loans for customers who are likely to default.

▶ Rejecting those who are likely to repay.

Properly addressing these challenges ensures we neither lose money on bad loans nor miss out on potential revenue from reliable borrowers.

Steps

1. Data understanding and sourcing

2. Data quality check

3. Outliers and Univariate analysis

4. Bivariate analysis

5. Merging datasets

6. Quality check on the merged dataset

7. Analysis on merged dataset

8. Deriving the risks
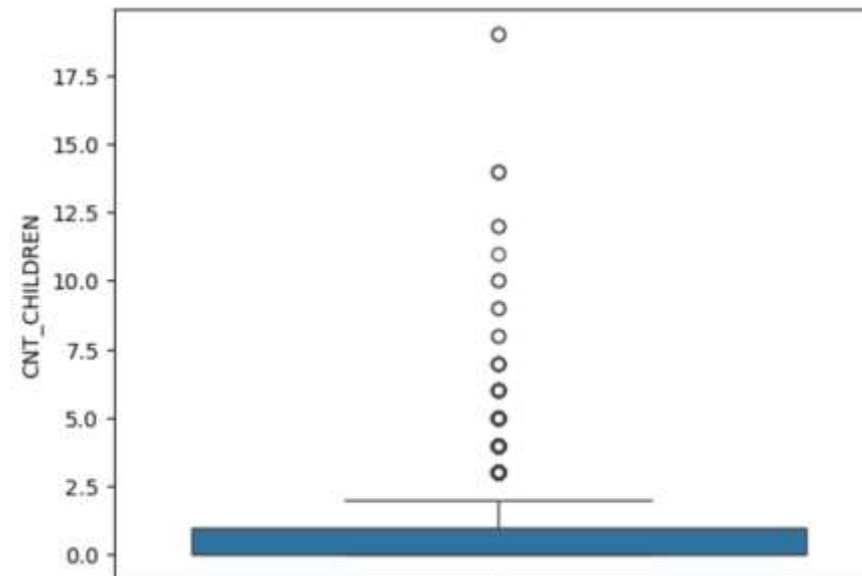
9. Conclusion

Outliers

"In most cases, it has been observed that the values exceed 3.5 million, as indicated by the boxplot.
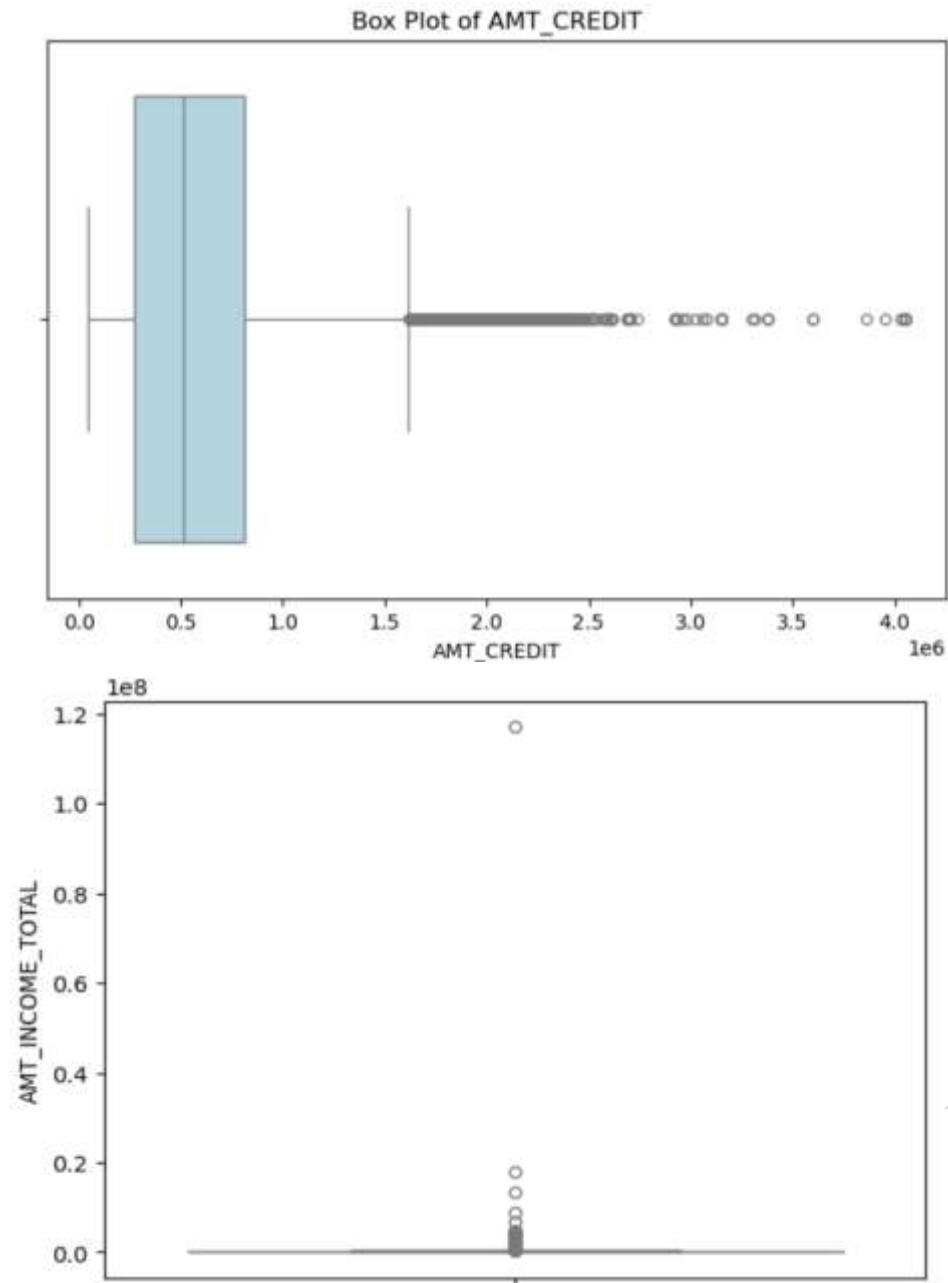
* We have data showing '1000 or more years of employment,' which is clearly unrealistic. This likely indicates a data entry error or an anomaly.
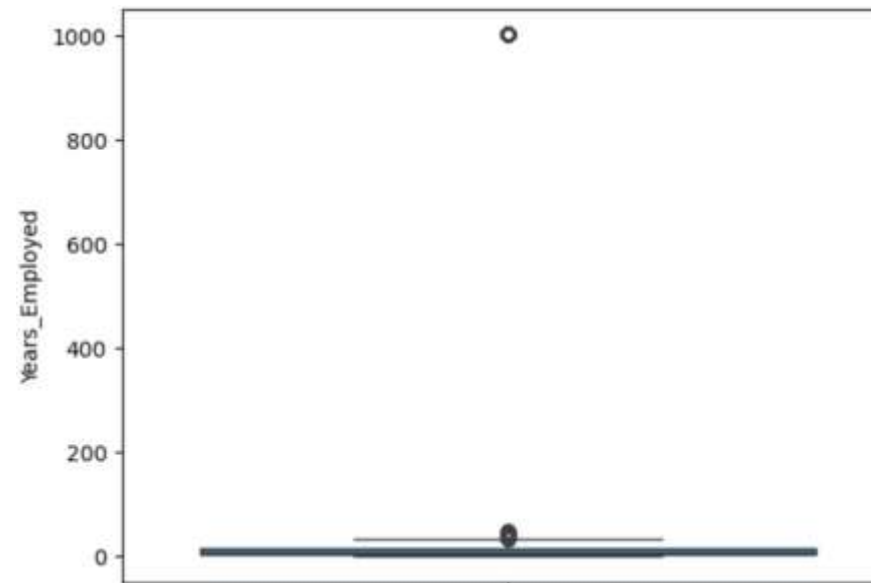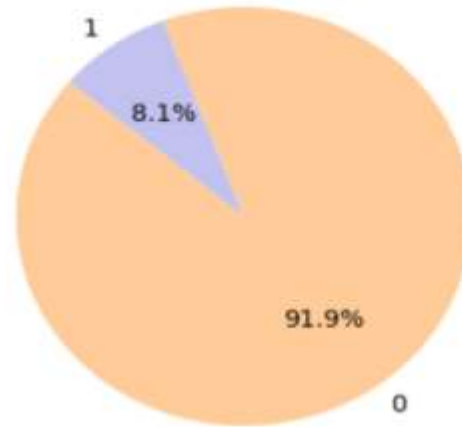
* "We've identified a data point showing 19 children, which is unusually high compared to other values.

"Based on this ratio, we can infer that there are approximately 11 instances of 0 for every single instance of 1. This distribution is clearly visible in the pie chart below."

Distribution of TARGET Variable

Correlation for Non-Defaulters and Defaulter

"By examining the heat map, we observe that while the correlations between most variables are similar, they are not identical. This indicates that the relationships among the variables are generally stable, with some variations. Most pairs of variables show consistent patterns, but there are a few exceptions where the correlations differ, highlighting some variability in the dataset."
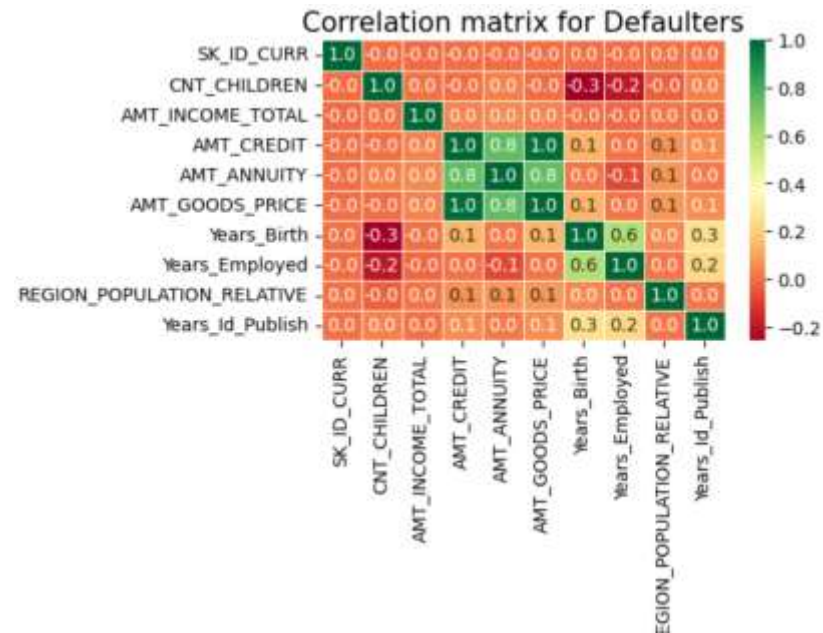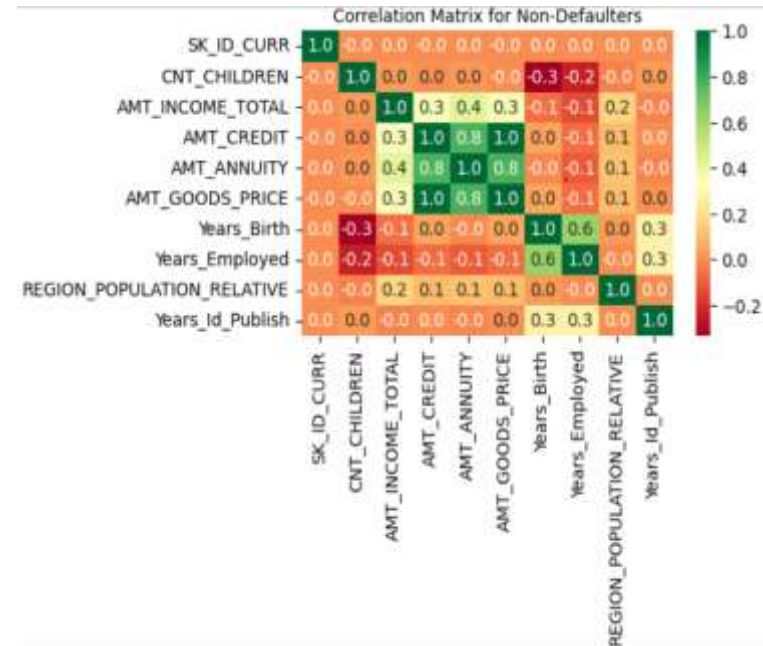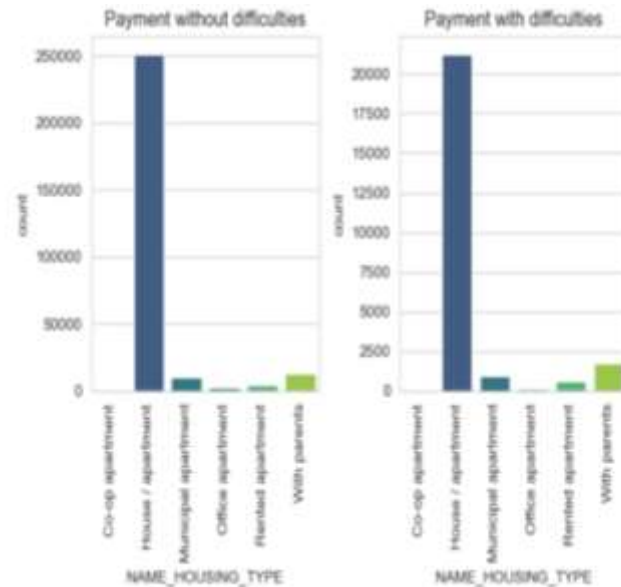
Analysis shows that females tend to apply for higher credit amounts compared to males.
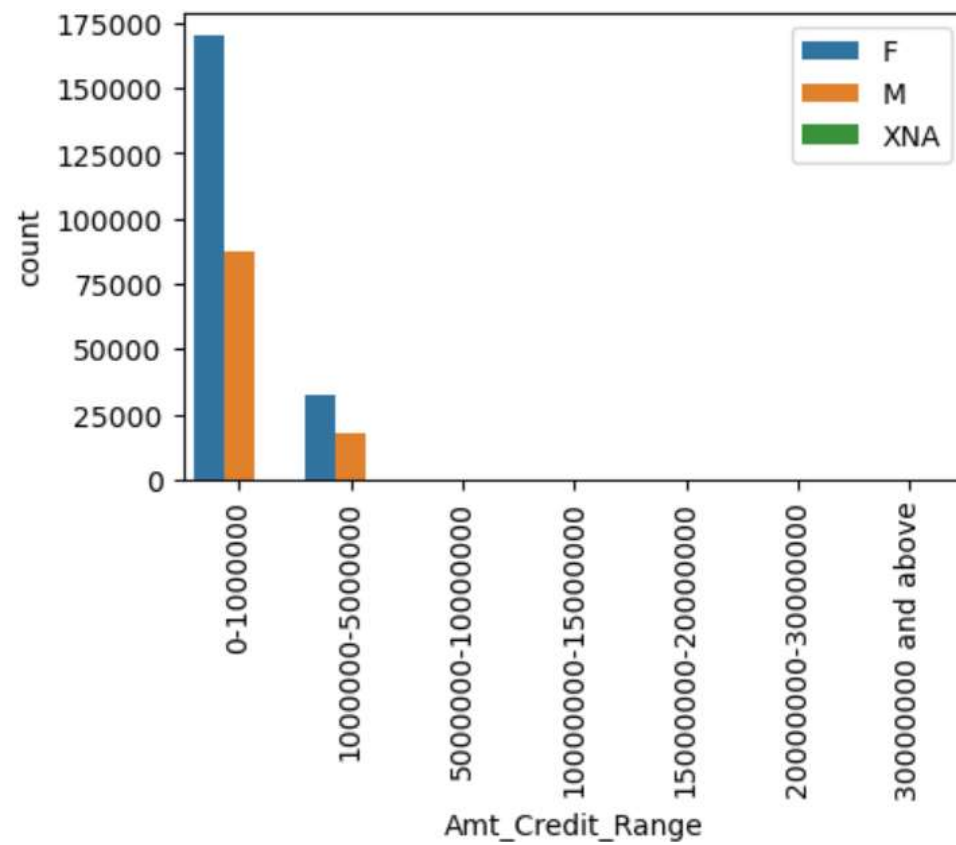
Univariate Analysis
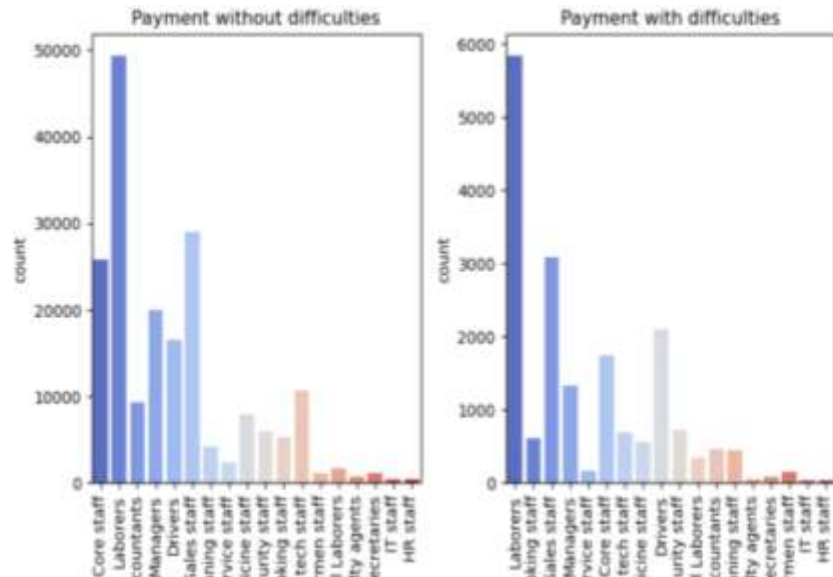
"We can see that laborers face more difficulty in repaying loans compared to other groups. Sales staff and core staff also experience challenges with loan repayments. Additionally, laborers represent a larger proportion of loan defaulters."

"The House/Apartment category emerges as the most frequently selected option in both instances, indicating that customers primarily prefer this category in each scenario."

Bivariate Analysis

From the below the graphs We can see that the proportion of customers with payment difficulties who have an academic degree is higher compared to those without payment difficulties. The distribution of other education types is similar between the two groups. Additionally, these customers tend to have higher credit amounts.
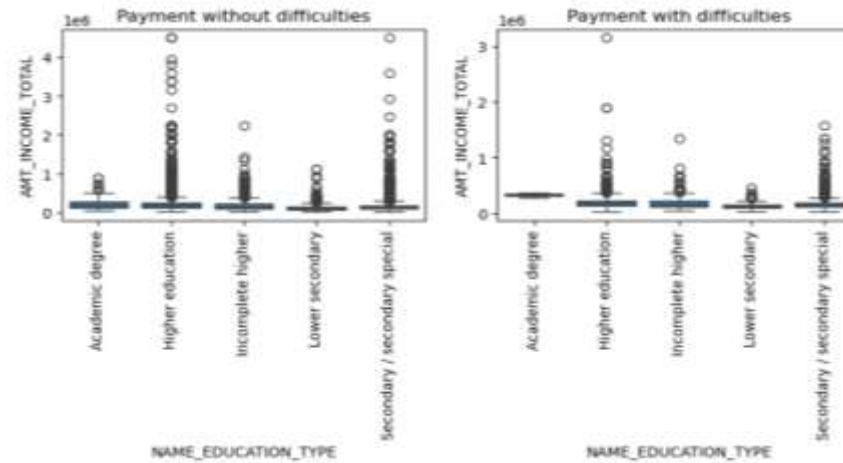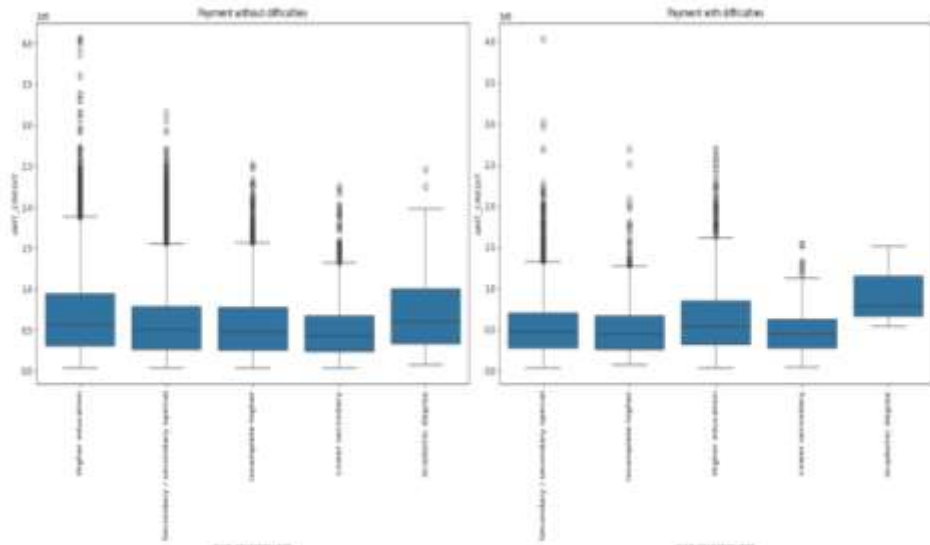
Below is the graph illustrating the contract statuses, including the numbers of approved, canceled, refused, and unused offer contracts.

Distribution of Contract Statuses
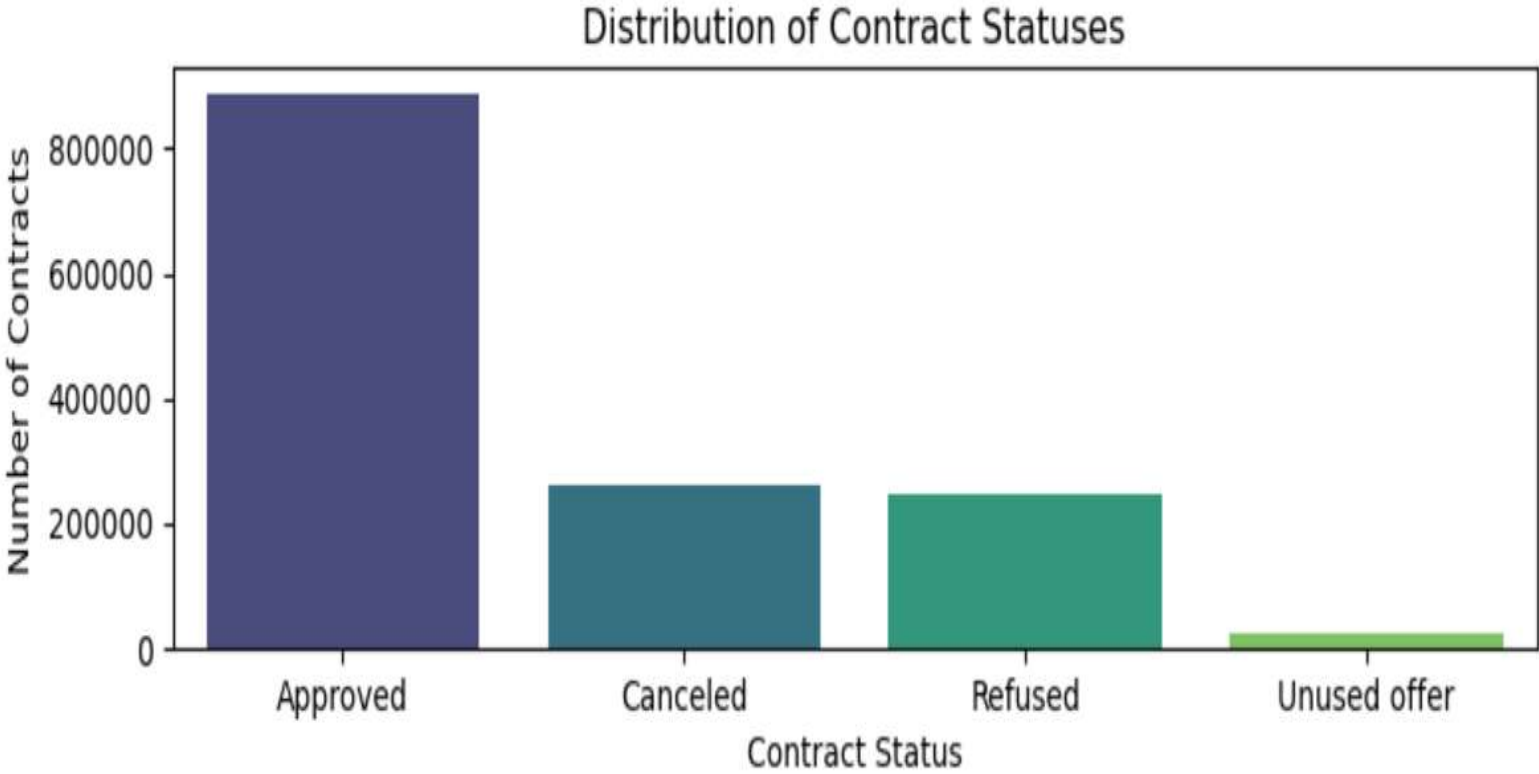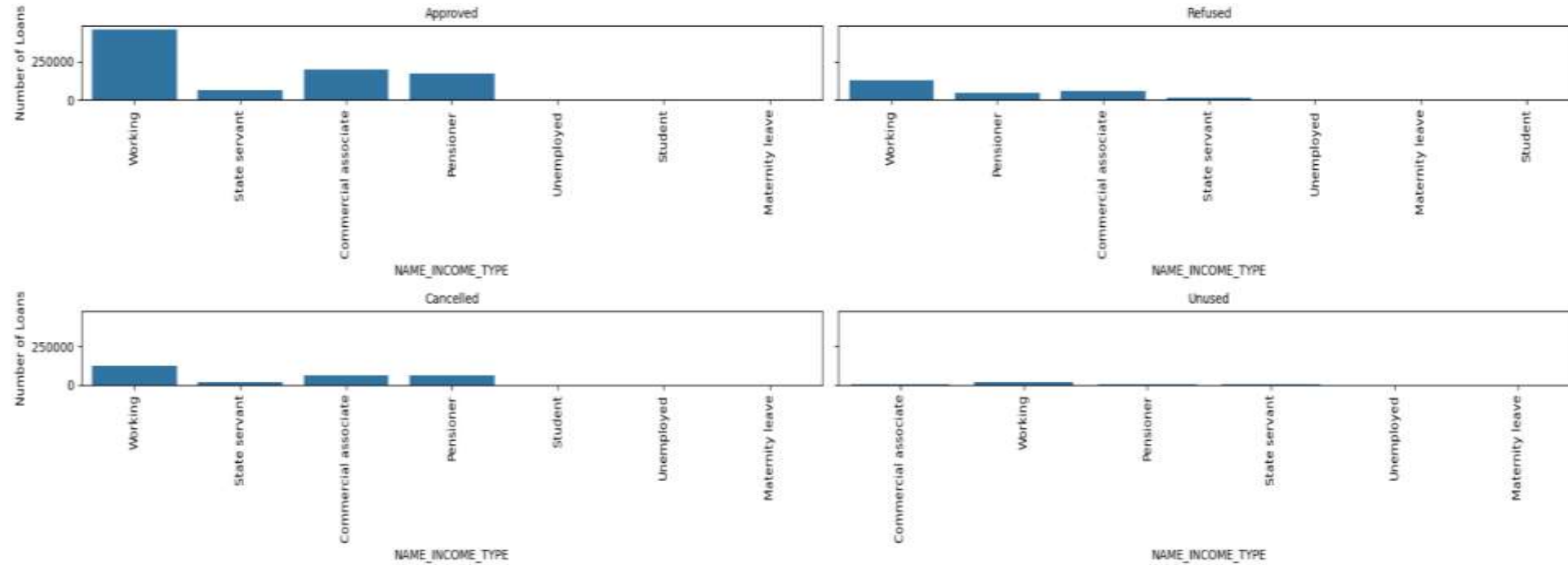
From the above graph

"We can observe that individuals in working roles and commercial associates tend to apply for more loans compared to others. To reduce unsuccessful payments, we should focus on students and pensioners, who generally have lower rates of payment difficulties."
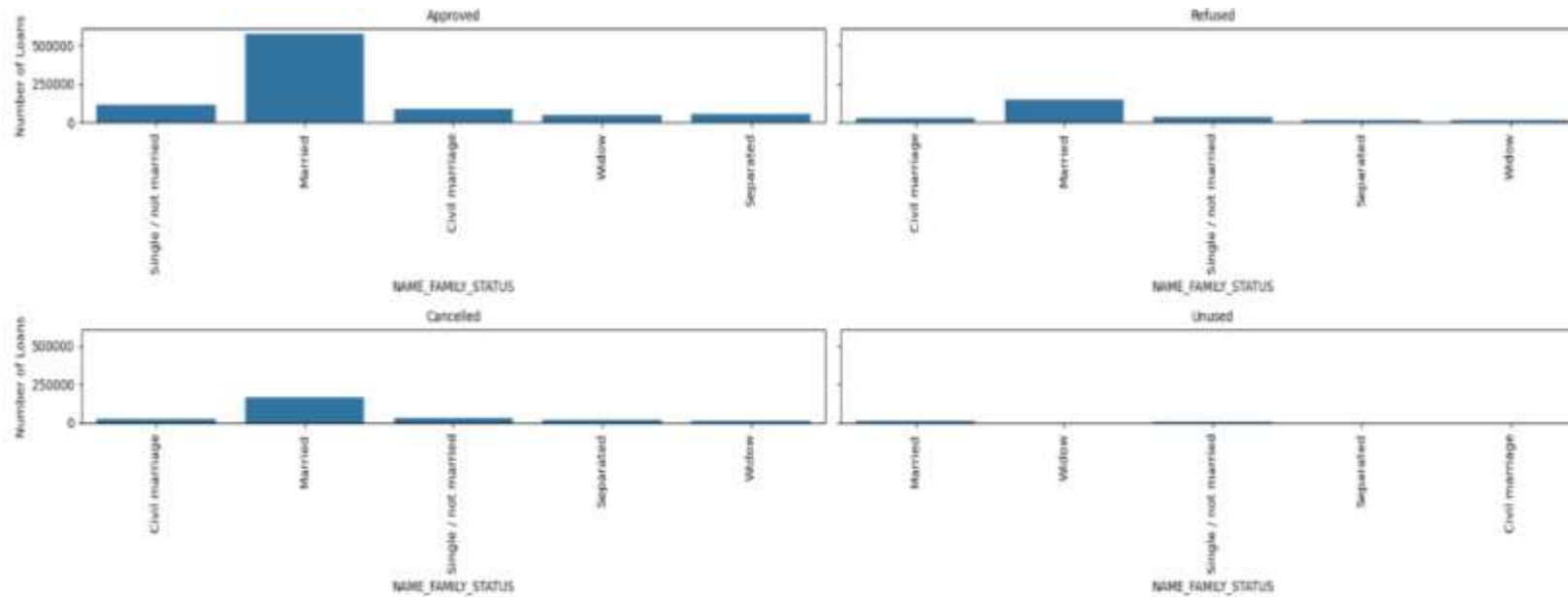
From the above graph
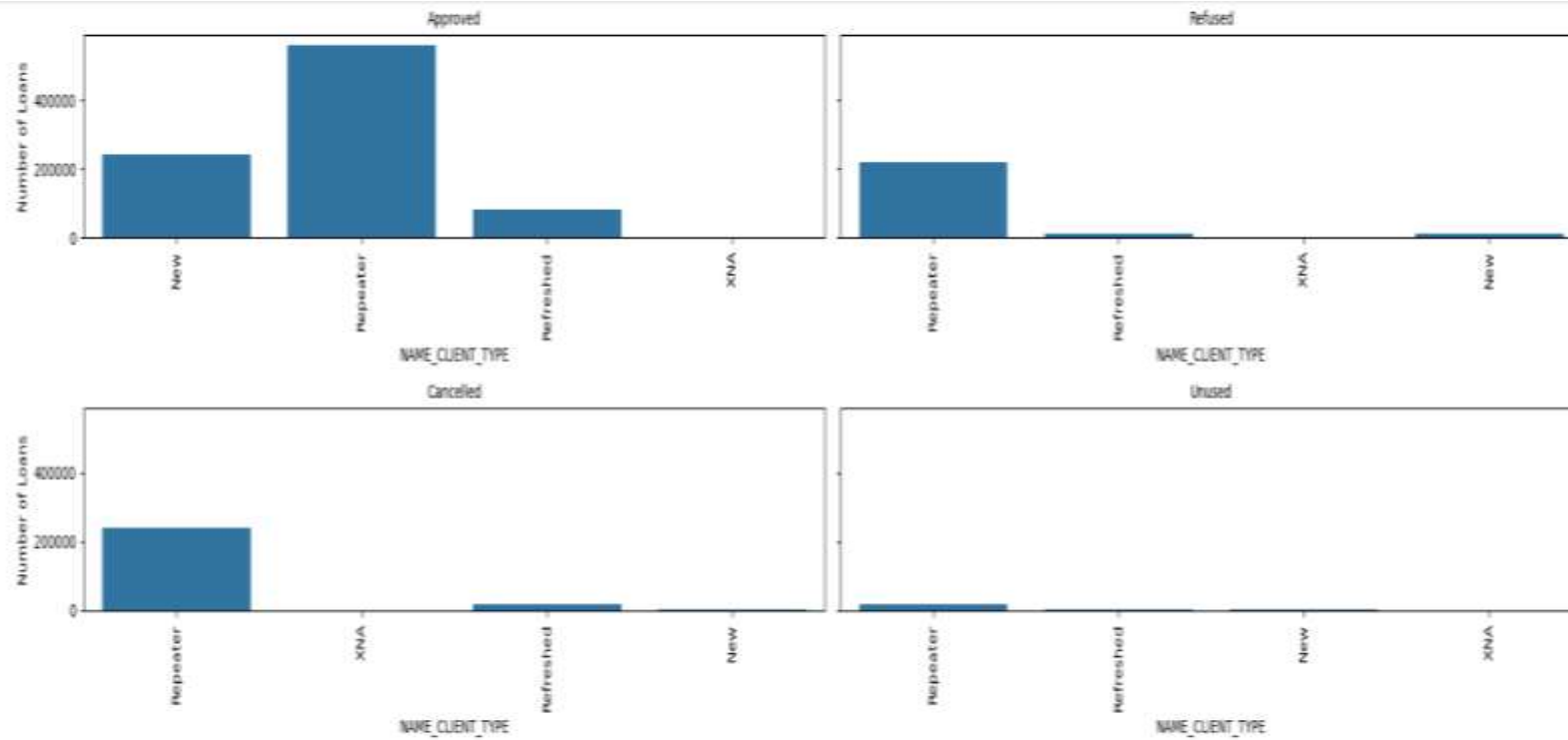we can see that married individuals have a higher number of approved, refused, and cancelled loans compared to other marital statuses."

From the above graph we can observe
Repeater loans have the highest number of refusals, but they also show higher rates of approvals, cancellations, and unused instances. In contrast, refreshed loans are less likely to default, as they have fewer cancellations and refusals.
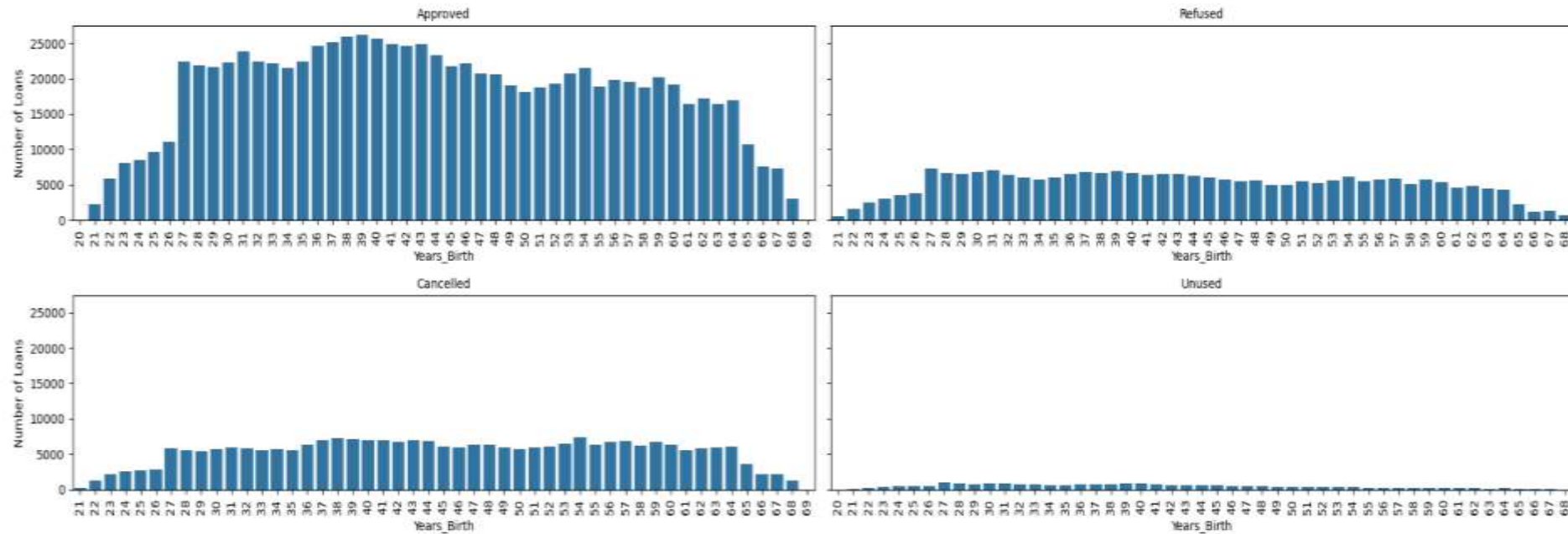
* Individuals aged 27 to 48 have the highest loan approval rates, while those aged 27 to 31 face the most refusals. Loan cancellations are less common among people in their early 20s and those over 65.

* Based on the data, individuals aged 27 to 48 are more likely to have successful loan repayments, given their higher approval rates. Conversely, those in their early 20s and over 65 tend to have fewer loan cancellations, indicating a stable repayment pattern. Students and pensioners also show a lower likelihood of default, suggesting they are more successful in repaying loans.

Likely Defaulting Borrowers

➢ People whose loans were refused in the past are more likely to default on future loans.

➢ Banks should probably focus less on the 'Working' income type, as it is likely associated with a higher rate of unsuccessful payments.

Conclusion

❑ Focus on 'State Servant' as they tend to repay loans more successfully.

❑ Consider applicants living 'With Parents' since they have fewer failed repayments.

❑ Give extra consideration to female applicants, as they have lower default rates.

❑ Students and pensioners are less likely to default.

❑ People who were approved for loans before are less likely to default, so past approval can help predict future repayment success.

Thank You