

CREDIT EDA ASSIGNMENT

Manjit Singh

Problem Statement

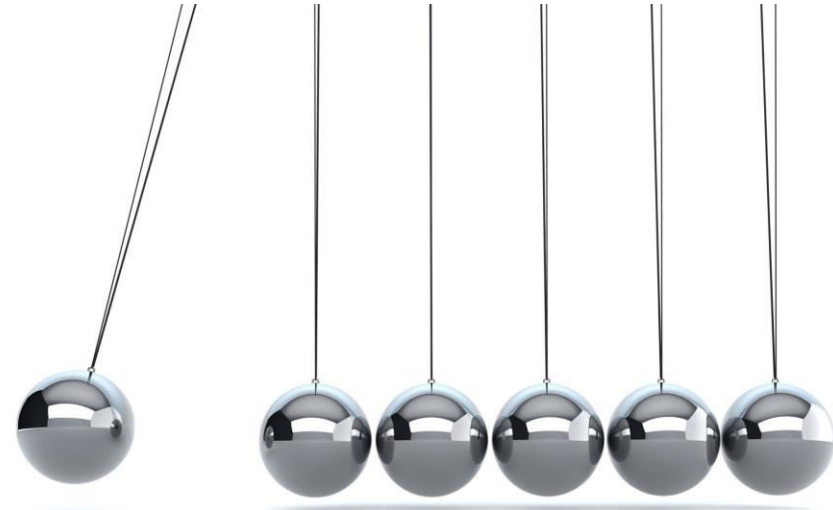
- Introduce the application of Exploratory Data Analysis (EDA) in a real life business scenario.
- Develop understanding of EDA techniques and risk analytics in banking and financial services.
- Highlight challenges faced by loan providers due to insufficient credit history of applicants.
- Discuss the two types of risks associated with loan approval decisions - loss of business or financial loss.
- Provide an overview of the dataset containing information about loan applications, including scenarios of payment difficulties and decision types.

Business Objective

- Identify patterns indicating difficulty in paying loan instalments to minimize rejection of capable consumers and mitigate default risks.
- Emphasize using EDA to understand the influence of consumer and loan attributes on default tendencies.
- Summarize the aim of identifying driving factors behind loan default for portfolio and risk assessment purposes.
- Suggest independent research on risk analytics to understand the significance of variables in the domain.

Approach Overview

- **Understanding Data:** Utilized the datasets and column descriptions to comprehend the provided data thoroughly.
- **Data Exploration:** Employed the pandas functions like shape, head, tail, describe, info, etc., to explore the dataset comprehensively.
- **Data Quality Check:** Performed checks for missing values and data integrity to ensure data reliability.



Data Cleaning and Preprocessing

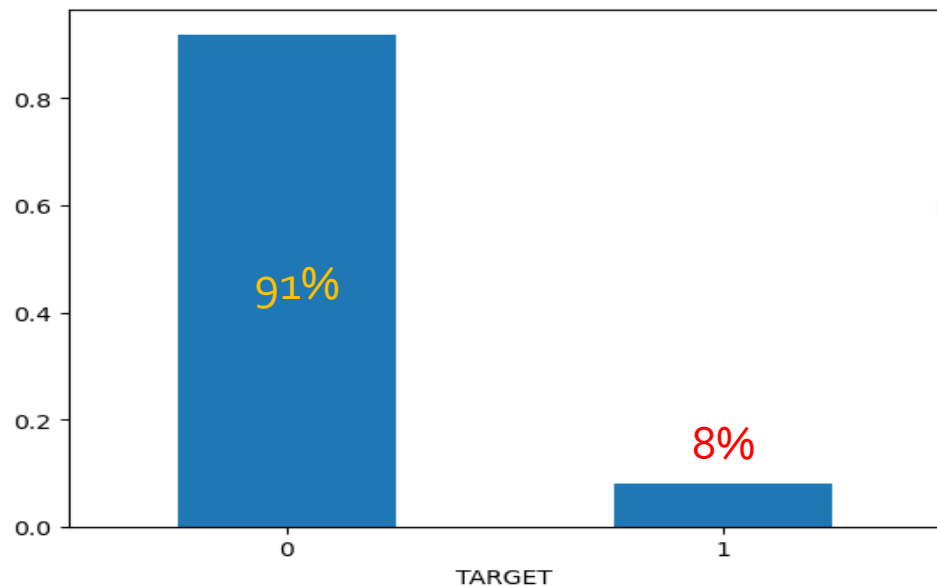
- **Handling Missing Values:** Dropped columns with missing values exceeding 50% and analyzed columns with less missing values (approximately 13%) using built-in functions.
- **Outlier Detection:** Identified and addressed the outliers in the dataset to enhance the accuracy of analysis.
- **Variable Binning:** Created bins for continuous variables to simplify analysis and interpretation.



Exploratory Data Analysis (EDA)

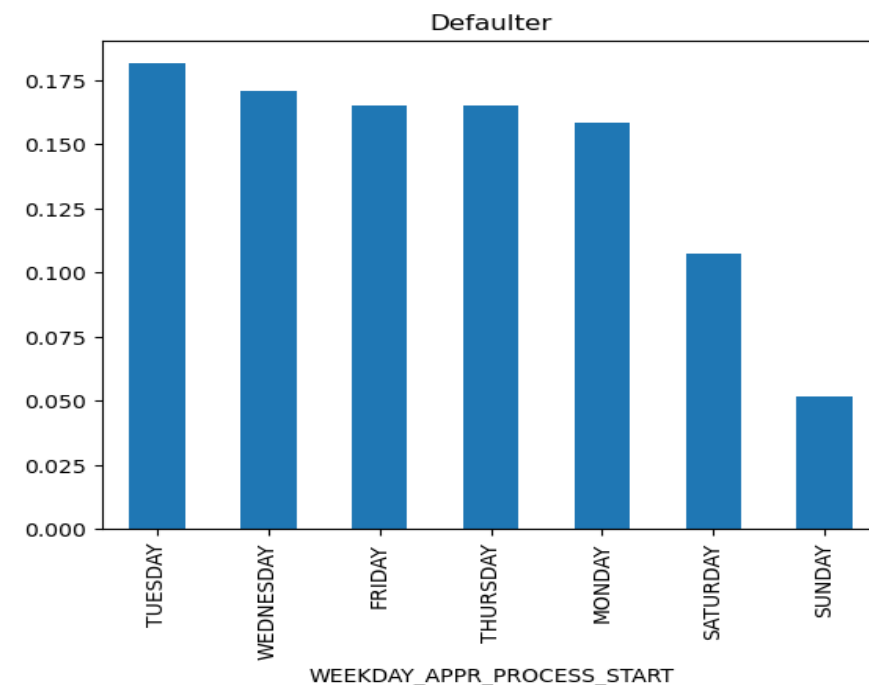
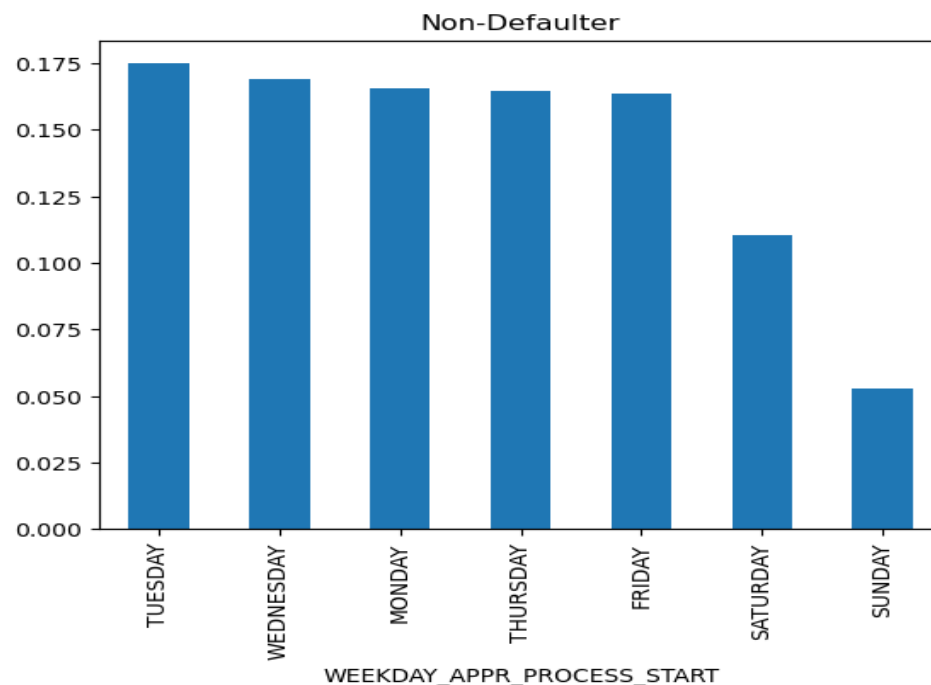
- **Graphical Analysis:** Utilized various visualization techniques such as box plots, scatter plots, etc., to analyze data distribution and relationships.
- **Statistical Analysis:** Applied statistical methods to uncover patterns and trends in the data.
- **Inference Generation:** Tried deriving meaningful insights from the analysis to understand how consumer and loan attributes influence the tendency to default.

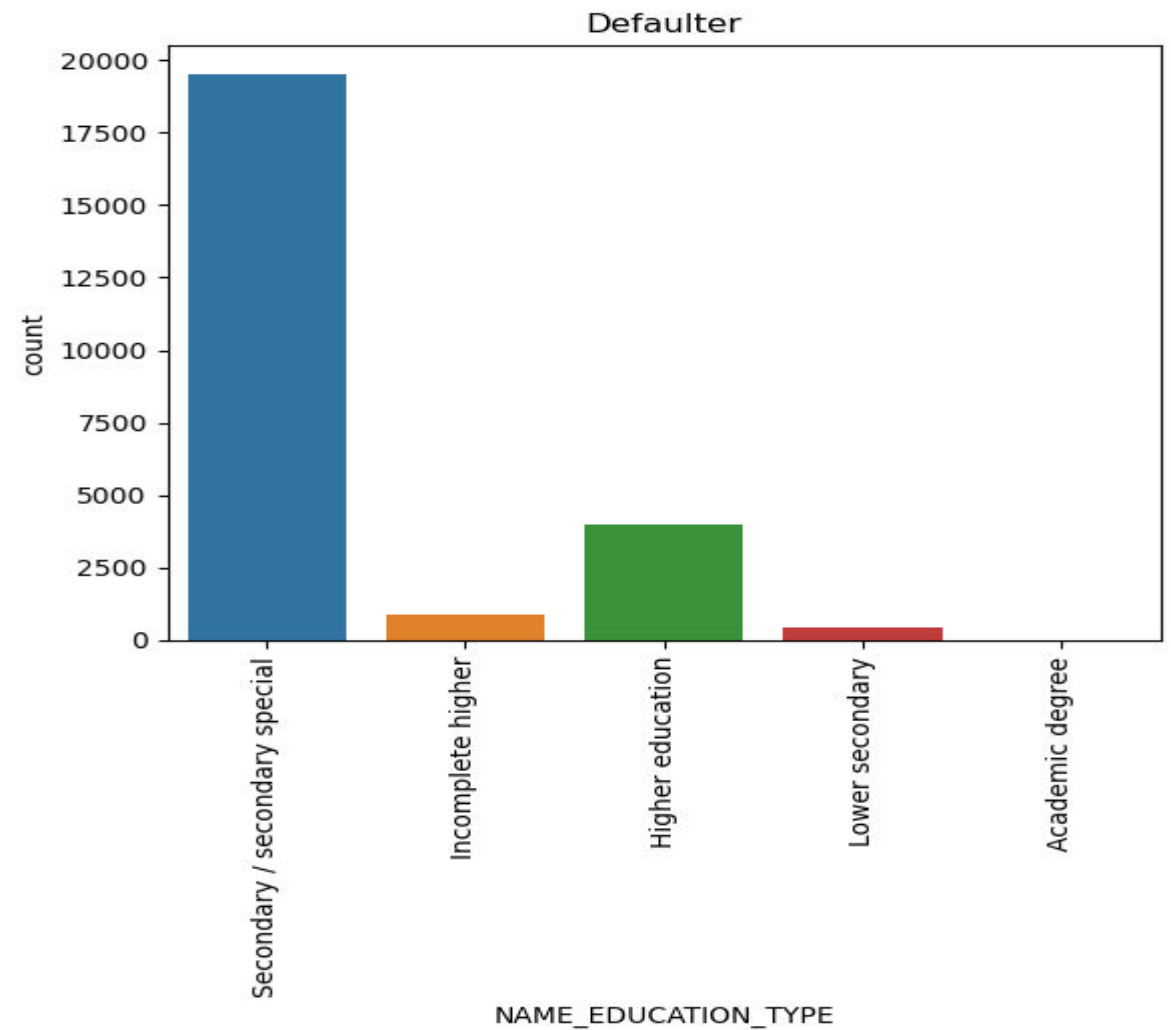
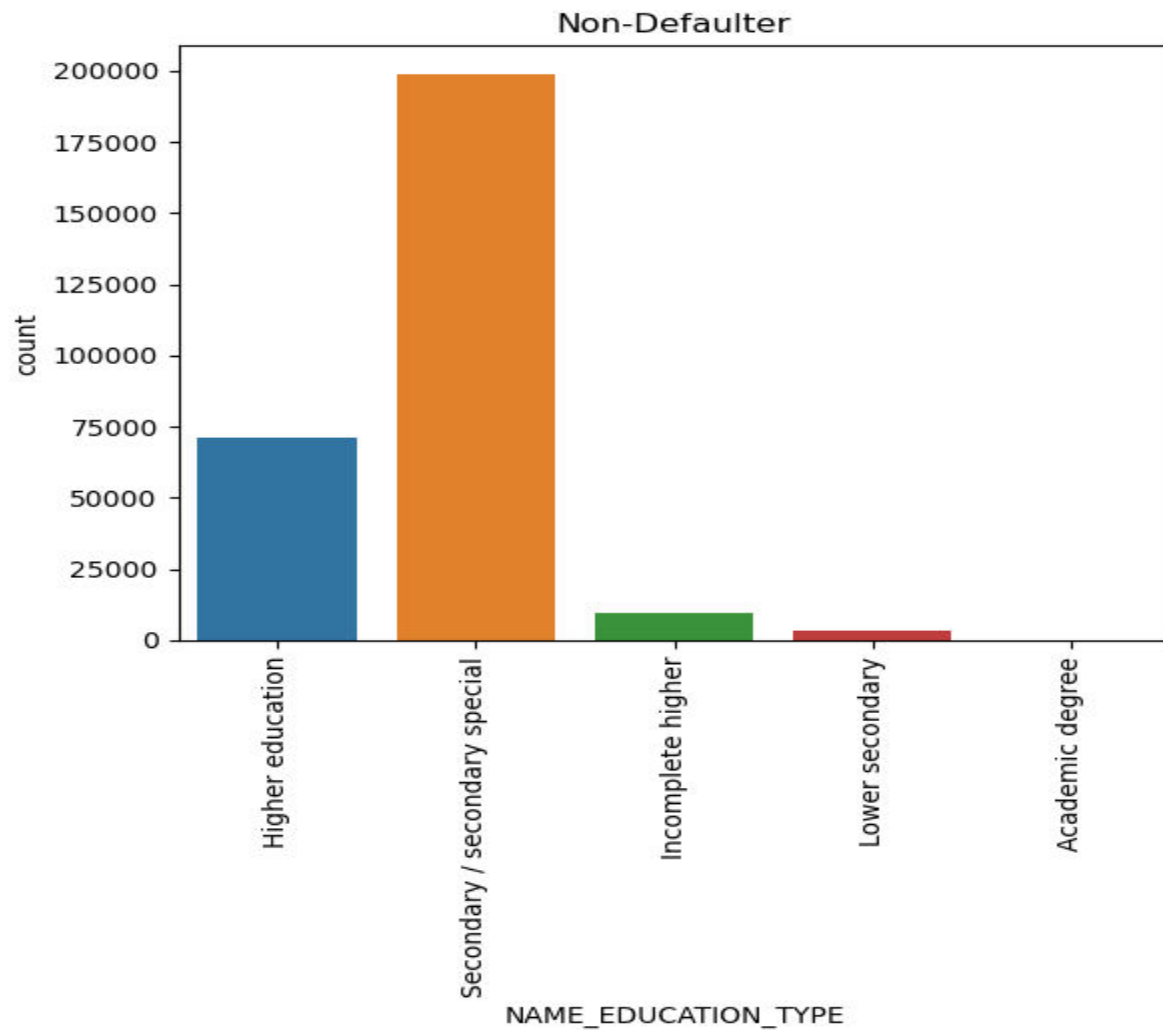




In my analysis, I discovered that nearly 1 in every 11 applicants encounters payment difficulties, with a stark difference in acceptance rates: 91.93% for the target group and only 8.07% for those facing payment difficulties. This highlights a significant challenge faced by a substantial portion of our applicant pool, emphasizing the importance of implementing effective solutions to address this issue.

Applying on Sundays lowers default risks, likely due to fewer applicants. Also weekends see fewer applications, indicating reduced banking activity.





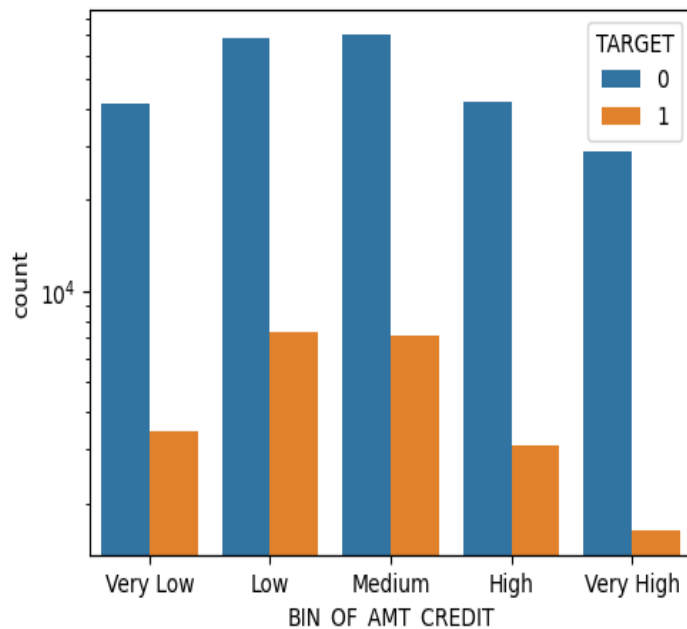
Individuals with secondary or special education qualifications are more likely to apply for loans, suggesting that financial accessibility extends to a broader range of educational backgrounds. This observation underscores the importance of inclusive lending practices and the potential impact of education level on borrowing behavior.



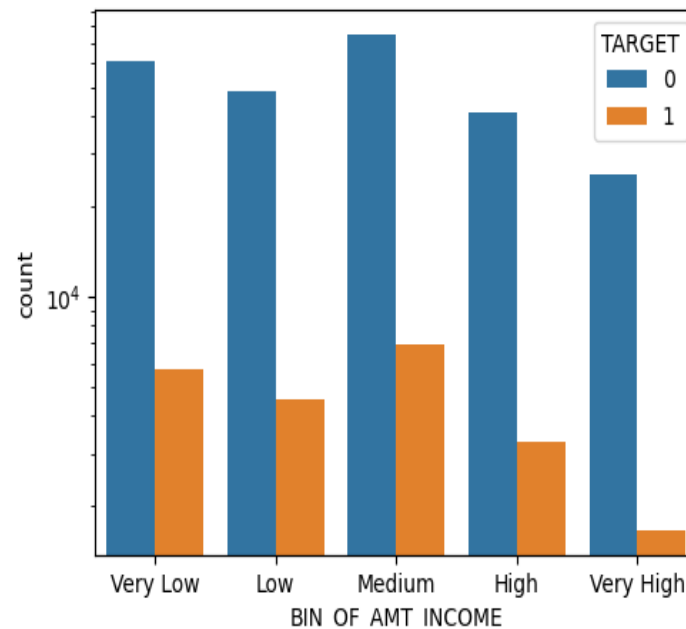
Individuals with access to higher credit limits exhibit a lower risk of default, showcasing their demonstrated trustworthiness and responsible management of debt.



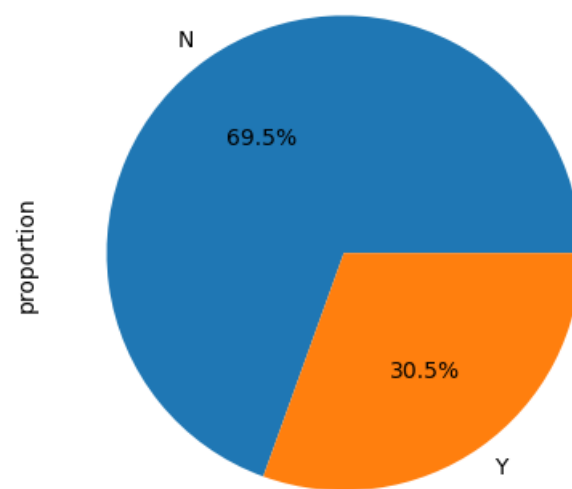
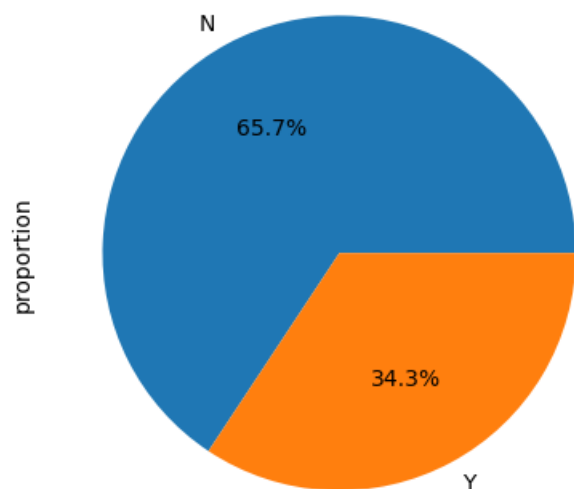
Non-Defaulter



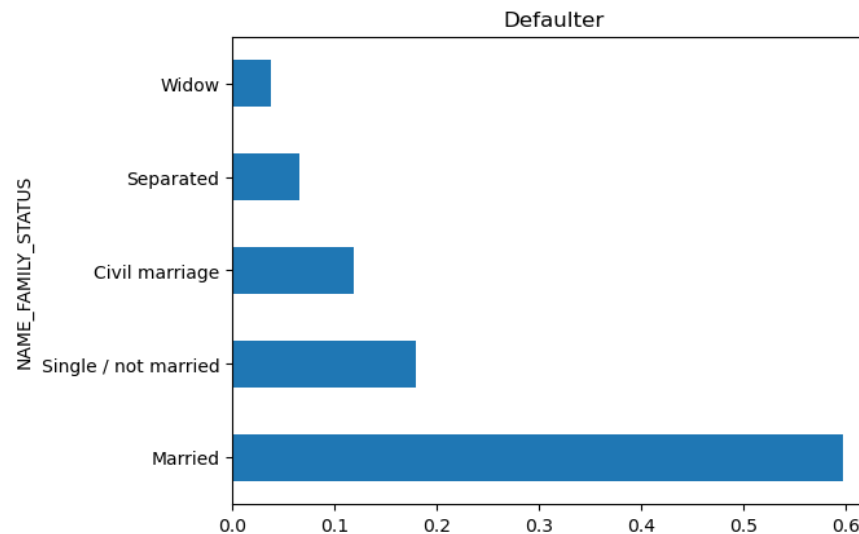
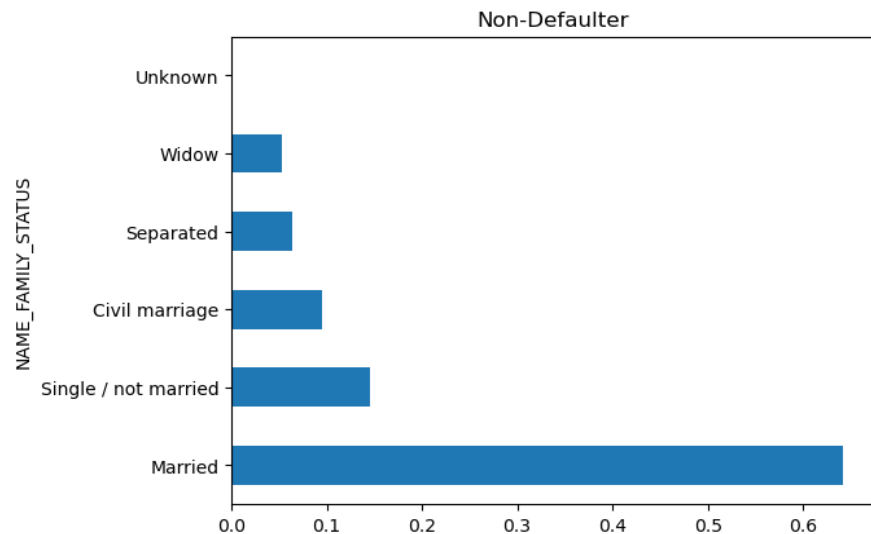
Defaulter



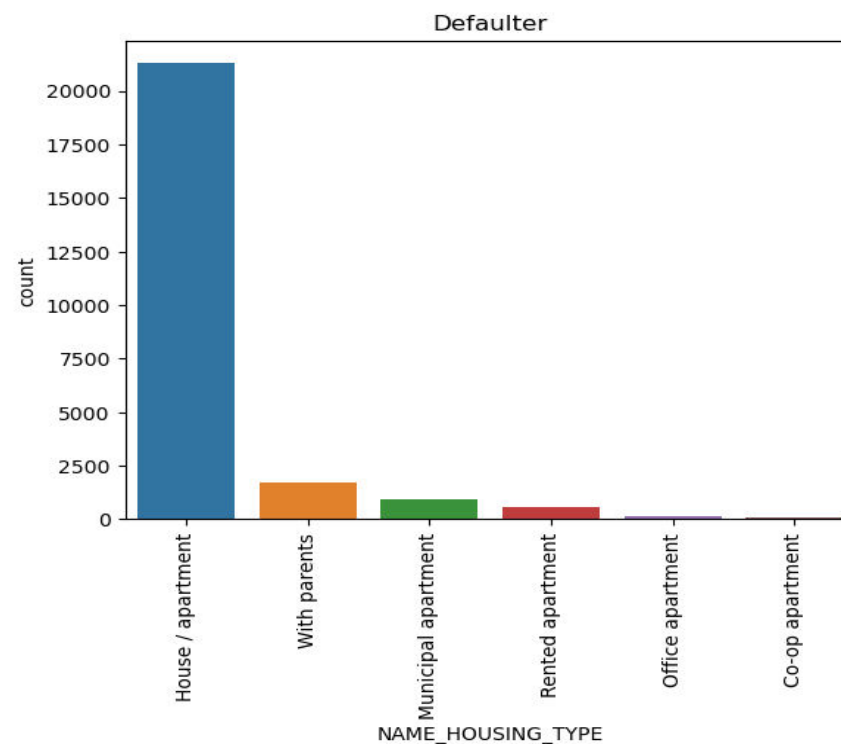
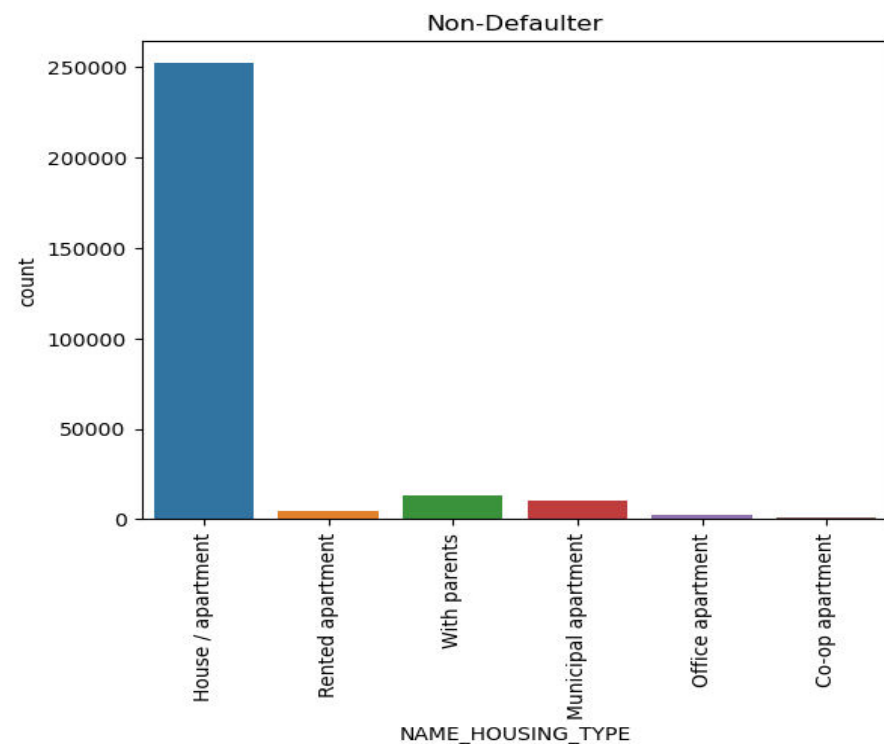
Medium income earners, often facing tighter financial constraints, are statistically more likely to default on loans, indicating underlying financial vulnerability.



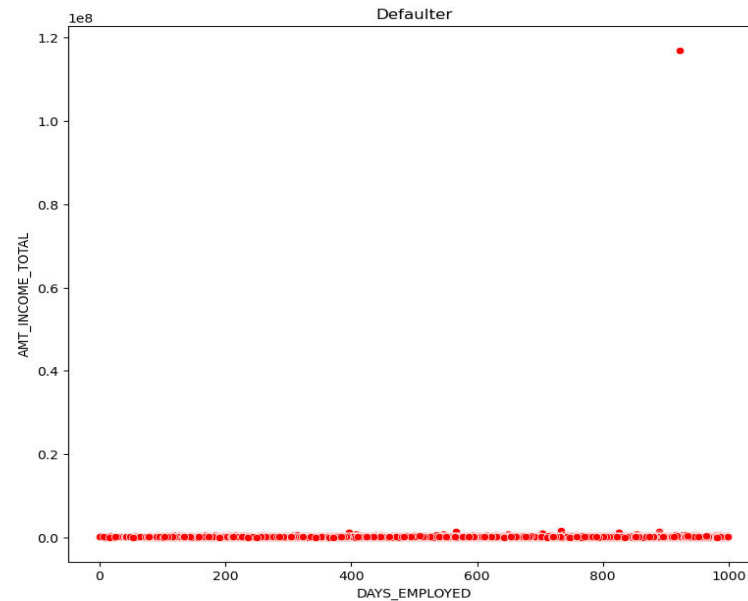
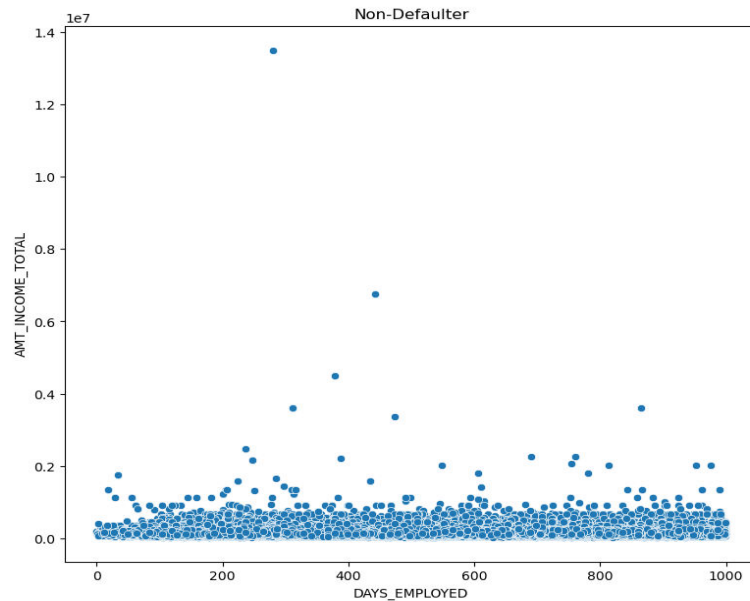
While car owners make up the majority of the population, a higher proportion of non-car owners are financially responsible based on loan default rates. This might be due to lower overall expenses or a different risk profile.



Married individuals often seek out more loans, suggesting a propensity for shared financial responsibilities and potentially a greater need for liquidity to manage joint expenses and investments.



Homeowners typically show a preference for loans, leveraging their property as collateral to access additional funds for various purposes.

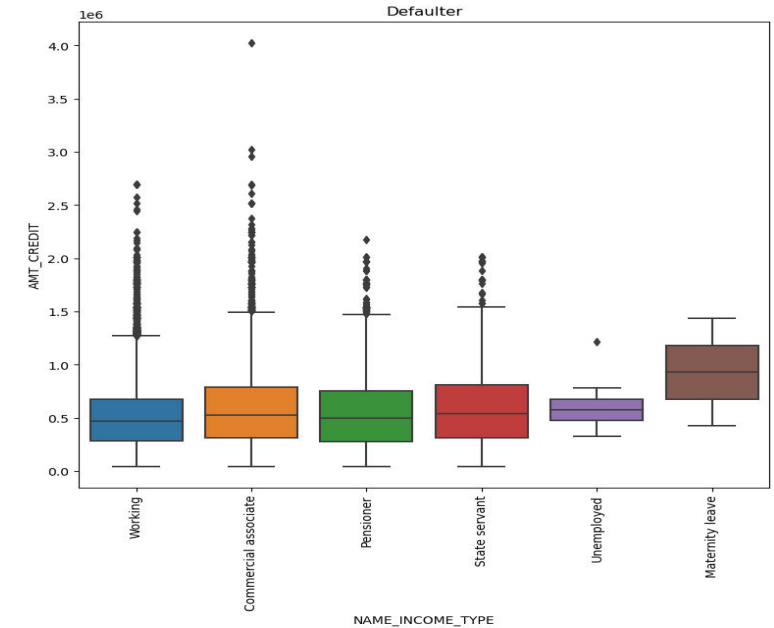
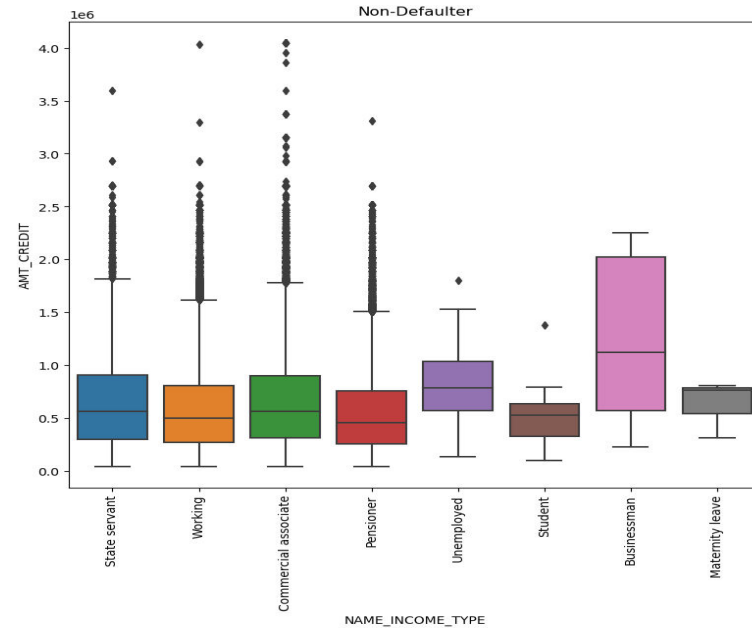


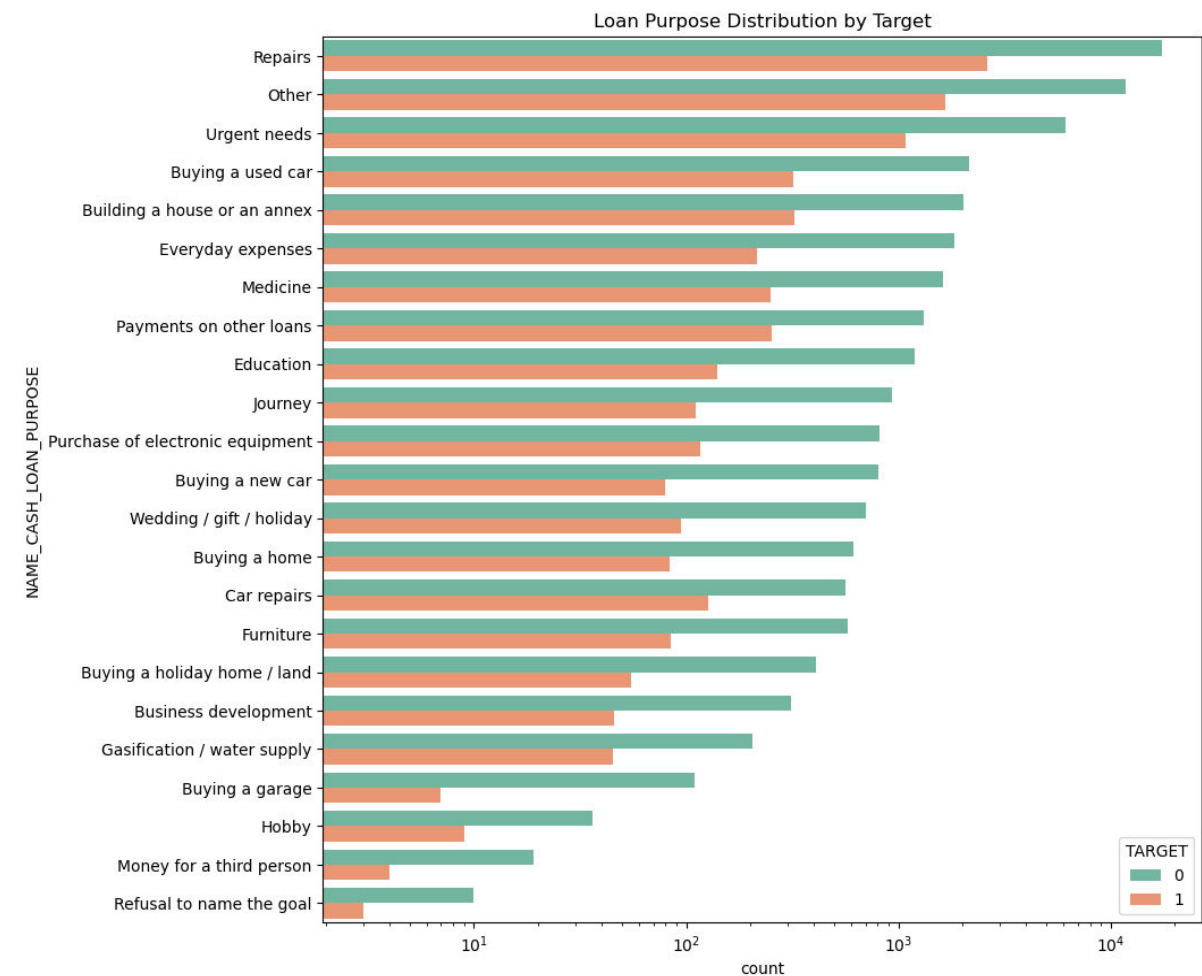
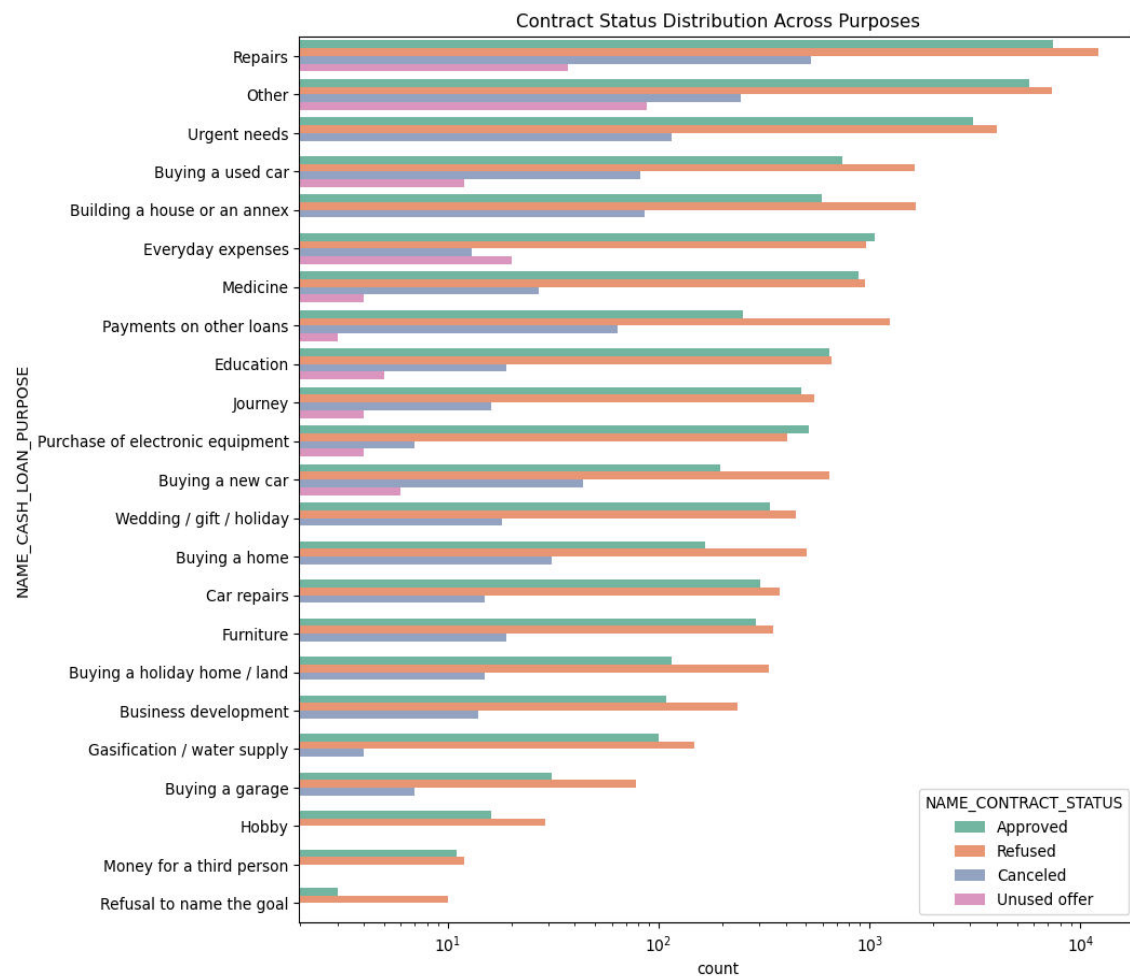
The plots reveal a consistent trend: defaulters, irrespective of employment duration, typically have substantially lower incomes.

In contrast, non-defaulters generally boast higher incomes, showcasing a discernible decrease in total income as years of employment rise.

Defaulters typically exhibit lower median credit amounts across various income brackets in comparison to non-defaulters.

Disparities in credit amounts among different income categories suggest the presence of potential factors influencing default probabilities across diverse income levels.





Analyzing both plots reveals intriguing insights into loan approval and repayment trends. Firstly, the graph depicting loan purposes demonstrates that 'Repairs' have the highest rejection rate, indicating challenges in securing loans for this purpose. Additionally, loans designated for 'Medicine', 'Everyday expenses', and 'Education' exhibit similar approval and rejection counts, suggesting a level of ambiguity or neutrality in their assessment. Secondly, the plot depicting timely payments unveils that loan purposes associated with 'Repairs' face notable difficulties in meeting payment deadlines, hinting at potential financial strains or unforeseen obstacles in this category. These observations underscore the nuanced dynamics of loan purposes and repayment behaviors.

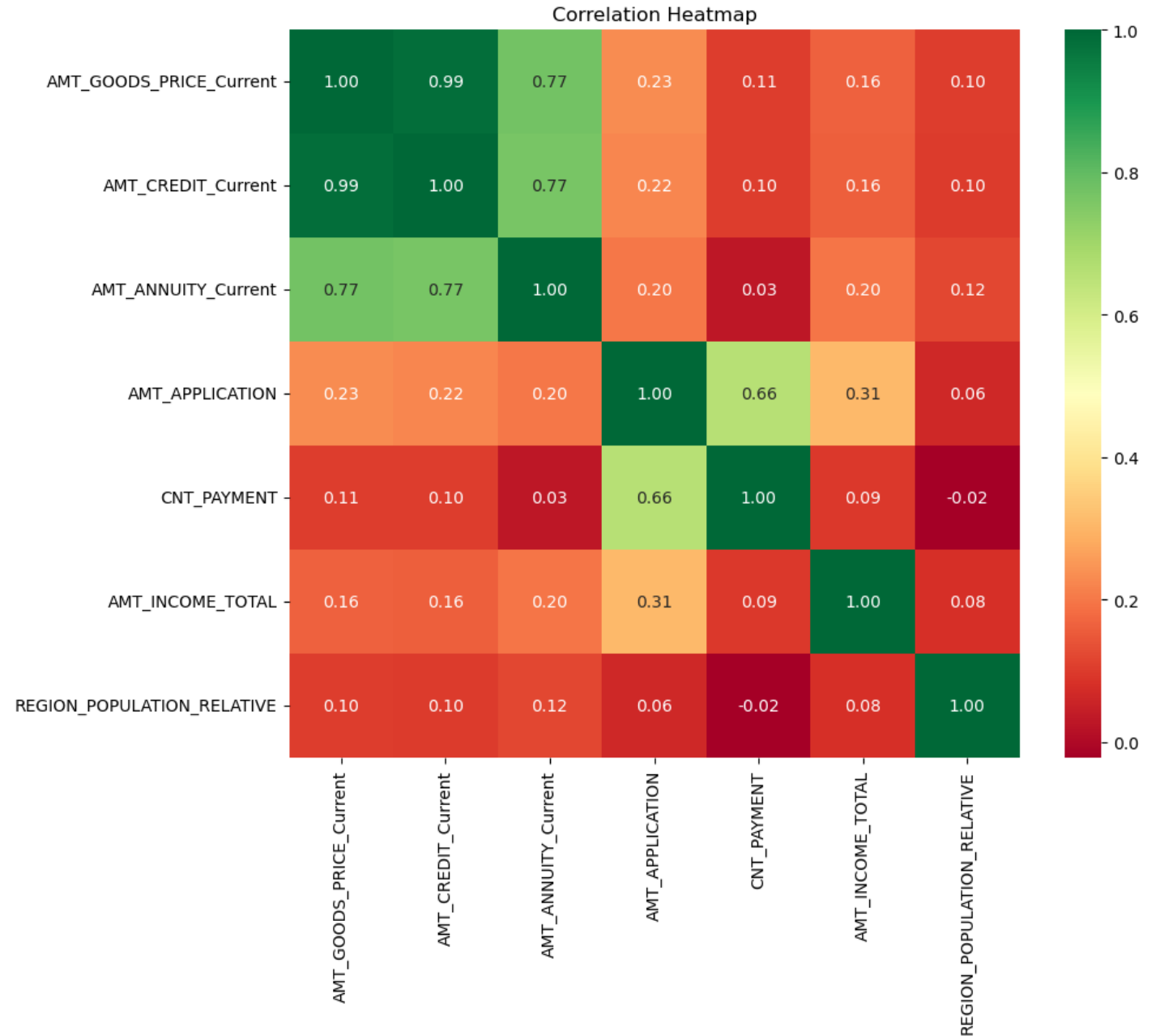
Based on the data analysis, it appears prudent for the bank to reassess its loan approval criteria concerning individuals residing in 'Co-op apartment' housing types. This subgroup exhibits notable difficulties in meeting payment deadlines, indicating a potential risk factor that warrants closer examination in loan evaluation processes.



Interpreting the Heatmap:

- A robust positive correlation is observed between the variables AMT_GOODS_PRICE_Current, AMT_CREDIT_Current, and AMT_ANNUITY_Current, implying a significant relationship among these factors.

- Conversely, a weaker positive correlation is noted between AMT_GOODS_PRICE_Current and AMT_INCOME_TOTAL, suggesting a less pronounced association between the current goods price and the total income.



Conclusion

- EDA, a valuable tool, identified patterns in loan applications impacting repayment risk.
- Key factors influencing risk include income level, housing type (co-op residents face higher repayment challenges), and loan purpose (purposes like "Buying a Car (New)" have higher approval rates compared to "Repairs").
- Data quality checks are crucial for reliable analysis, and diverse EDA techniques (graphical & statistical) were used to uncover these valuable insights.
- Further exploration of EDA is recommended to gain deeper understanding and improve loan approval processes.
- The bank should consider reevaluating loan criteria for co-op residents, following fair lending practices, and investigate potential implications of observed relationships between loan features.