CREDIT EDA CASE STUDY

CREDIT RISK ANALYSIS

- Paarthasarathy Rengasamy

Introduction:

- This assignment provides a practical application of Exploratory Data Analysis (EDA) within the context of risk analytics in banking and financial services.
- The primary objective is to apply EDA techniques to gain insights into customer data, specifically in the lending domain. Additionally, the assignment aims to foster a fundamental understanding of how data is instrumental in mitigating the risk of financial loss associated with lending activities.

Business Objectives:

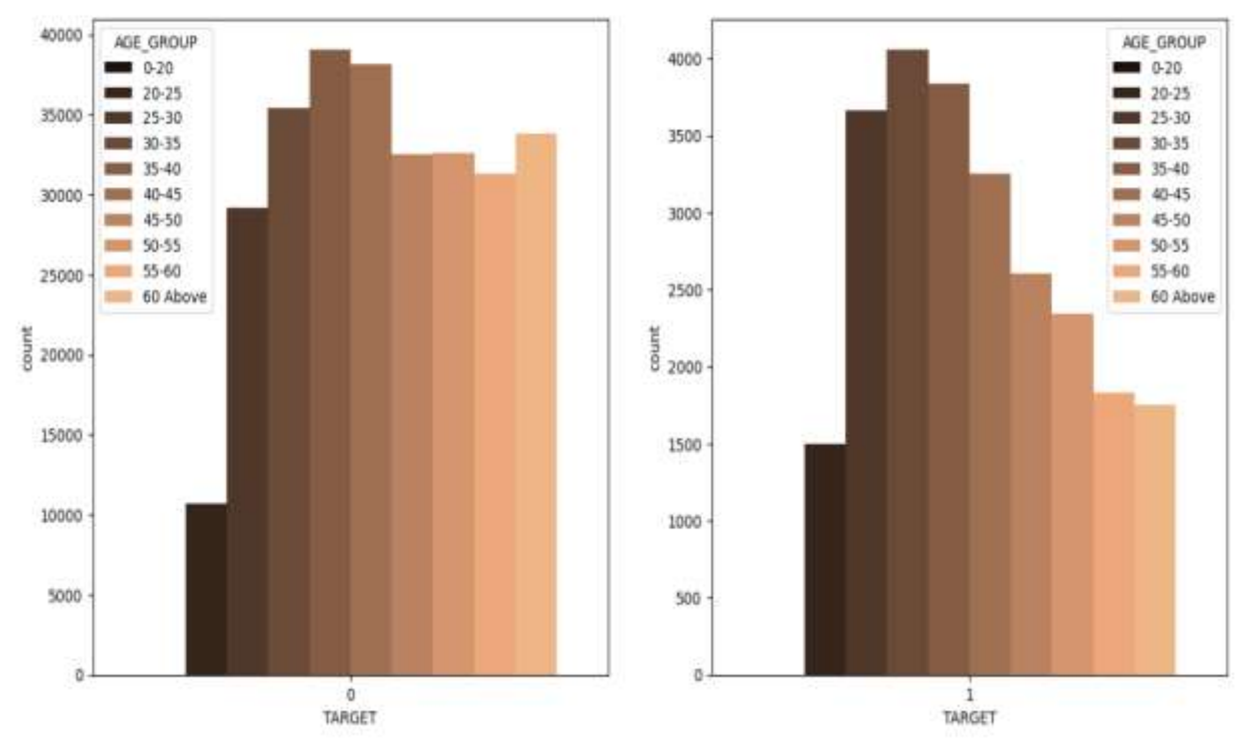
- ❖ The primary objective of this case study is to identify patterns that signify potential challenges in clients' loan repayment, with the overarching goal of informing strategic actions such as loan denial, adjusting loan amounts, or applying higher interest rates to risky applicants. The company aims to utilize Exploratory Data Analysis (EDA) to discern subtle data patterns, ensuring that consumers capable of repaying loans are not unfairly rejected.
- ❖ In essence, the company seeks to understand the key driver variables behind loan defaults, utilizing this knowledge to optimize its loan portfolio and enhance risk assessment capabilities. As part of developing a comprehensive understanding of the domain, stakeholders are encouraged to independently research risk analytics, focusing on variable types and their significance, to bolster analytical capabilities and support effective risk management decision-making.

Navigating Business Dynamics:

- In this case study, we delve into the loan provider's struggle to assess applications due to insufficient credit histories, potentially exploited by some consumers turning defaulters. As a member of a consumer finance company specializing in urban lending, our objective is to leverage exploratory data analysis (EDA) for analyzing data patterns.
- The overarching goal is to ensure fair loan approval for applicants capable of repayment. The company faces the dual risk of losing business when denying a capable applicant and incurring financial loss when approving a potential defaulter. The dataset includes scenarios of clients with payment difficulties and those who pay on time. The study aims to understand how consumer and loan attributes influence default tendencies.

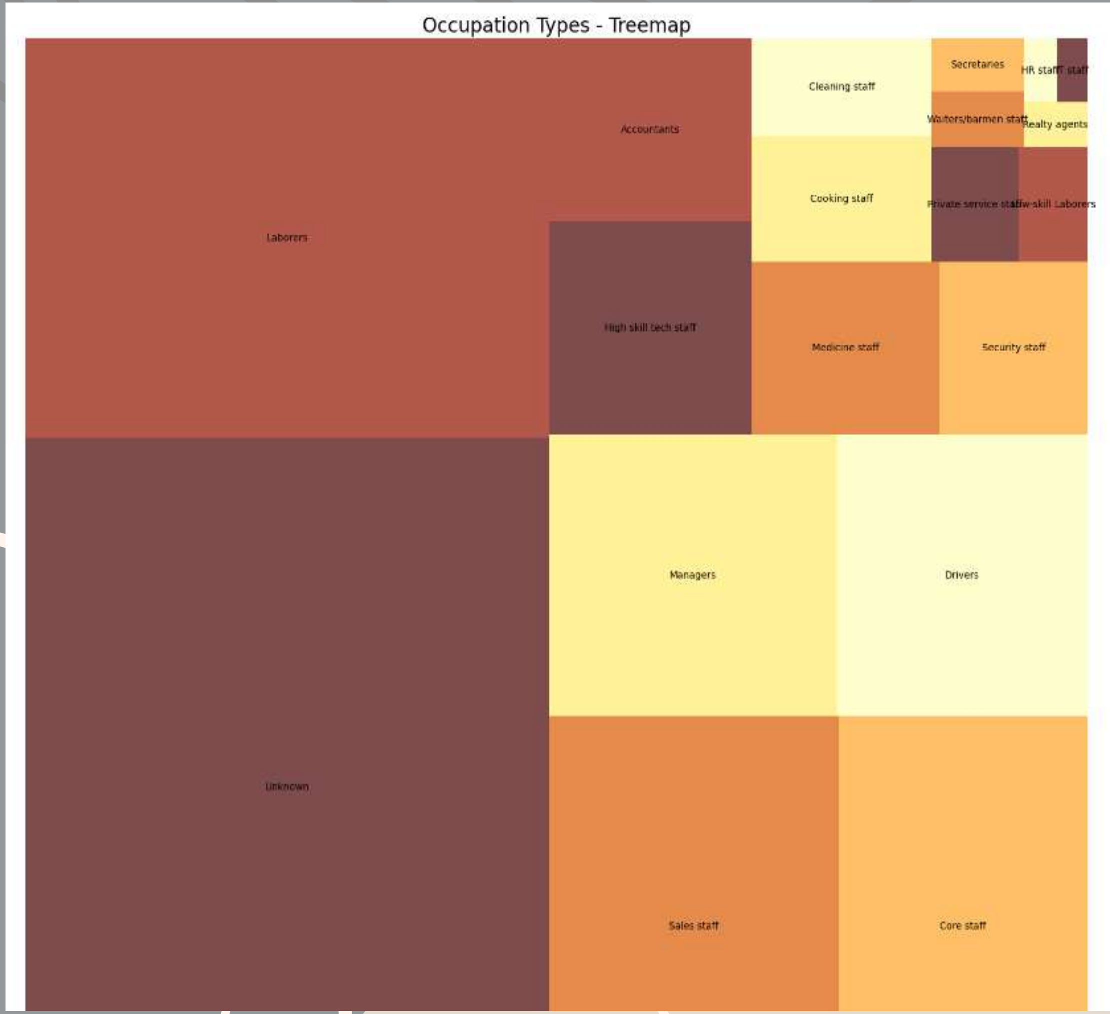
Exploring Age and Gender Demographics: A Preliminary Analysis

Age Distribution Analysis: Repayers vs. Defaulters



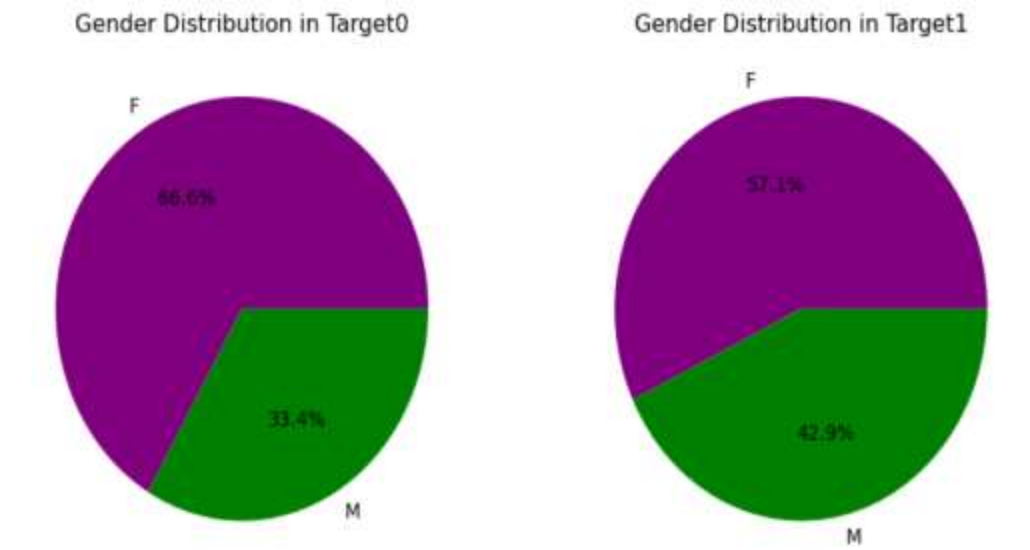
- ✓ The Age group (35-50 years) demonstrates the highest frequency of loan applications, both among defaulters and non-defaulters.
- ✓ This suggests a consistent trend of loan applications from individuals in the middle-age range
- ✓ This analysis reveals that a higher likehood of defaulters are among individuals aged 25 – 35 in comparison to other age group.
- ✓ On the contrary, Senior Citizens (60 years & above) and Very Young applicants (19-25 years) exhibit lower instances of payment difficulties, suggesting relative financial stability in these age categories.

Organizational Dominance Analysis:



- ✓ This analysis reveals the prevalence of specific categories within the dataset, showcasing that the 'Unknown' group constitutes the majority of values, holding the highest percentage.
- ✓ Following closely, the 'Laborers' group emerges as the second-highest category.
- ✓ This insight into the distribution of values sheds light on the significance of these two groups within the context of the dataset, highlighting potential areas for further investigation or targeted interventions.

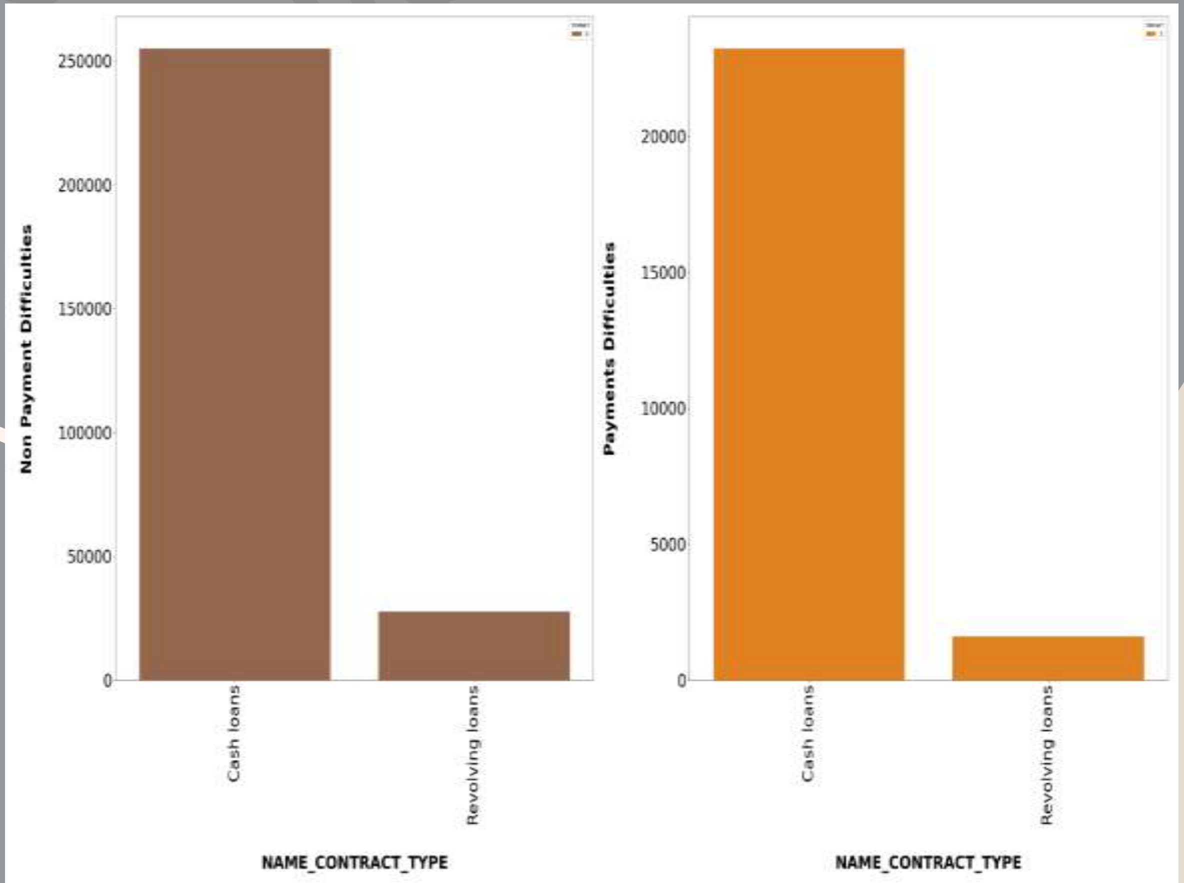
Gender Analysis in Loan Applications:



- ✓ Female clients exhibit a higher frequency of loan applications compared to male clients. A majority of female clients fall into the non-defaulter category, with a smaller proportion classified as defaulters.
- ✓ Conversely, female clients show a relatively higher percentage of defaulters also compared to non-defaulters. This gender-wise distribution provides insights into the loan application patterns and default tendencies within the dataset.
- ✓ Female clients display a higher loan application frequency, primarily non-defaulters, but with a notable proportion classified as defaulters, indicating distinctive gender-wise patterns in loan application and default tendencies within the dataset.

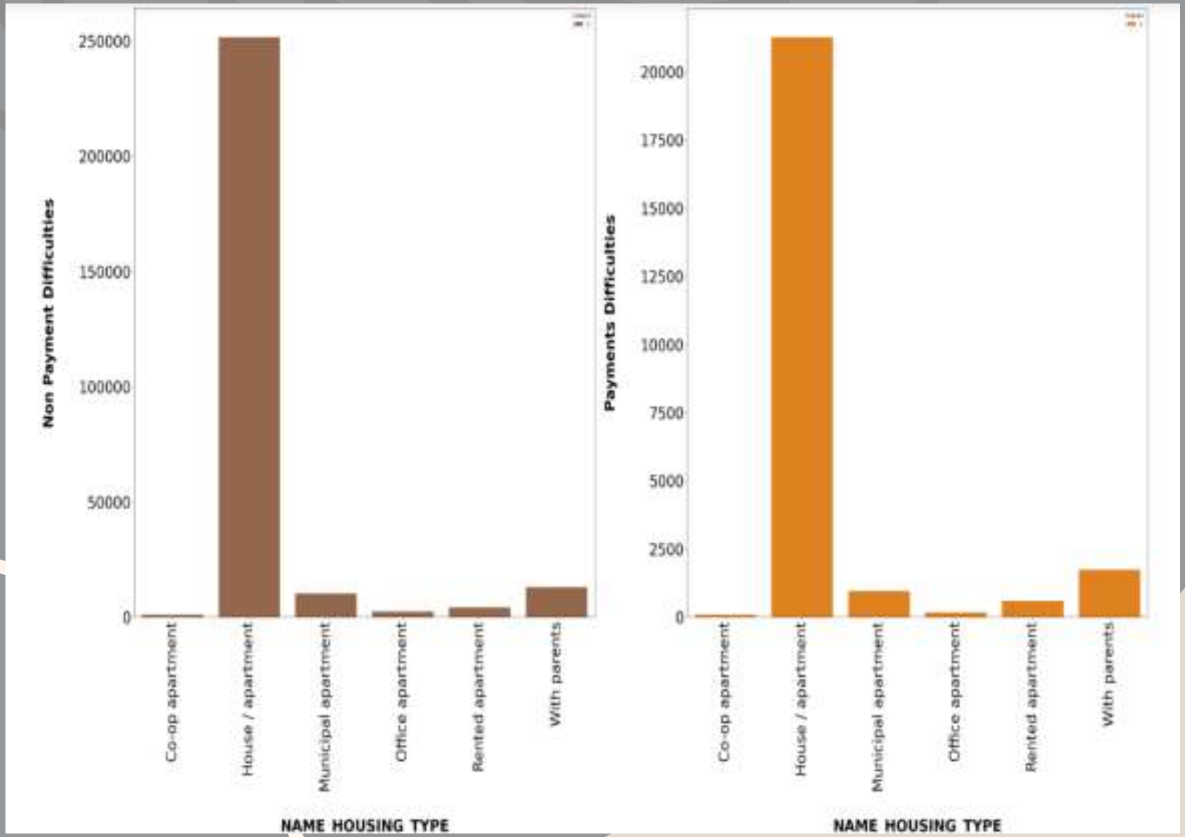
Exploring Univariate Analysis of Categorical Columns: Distribution and Insights

Contract Type Analysis:



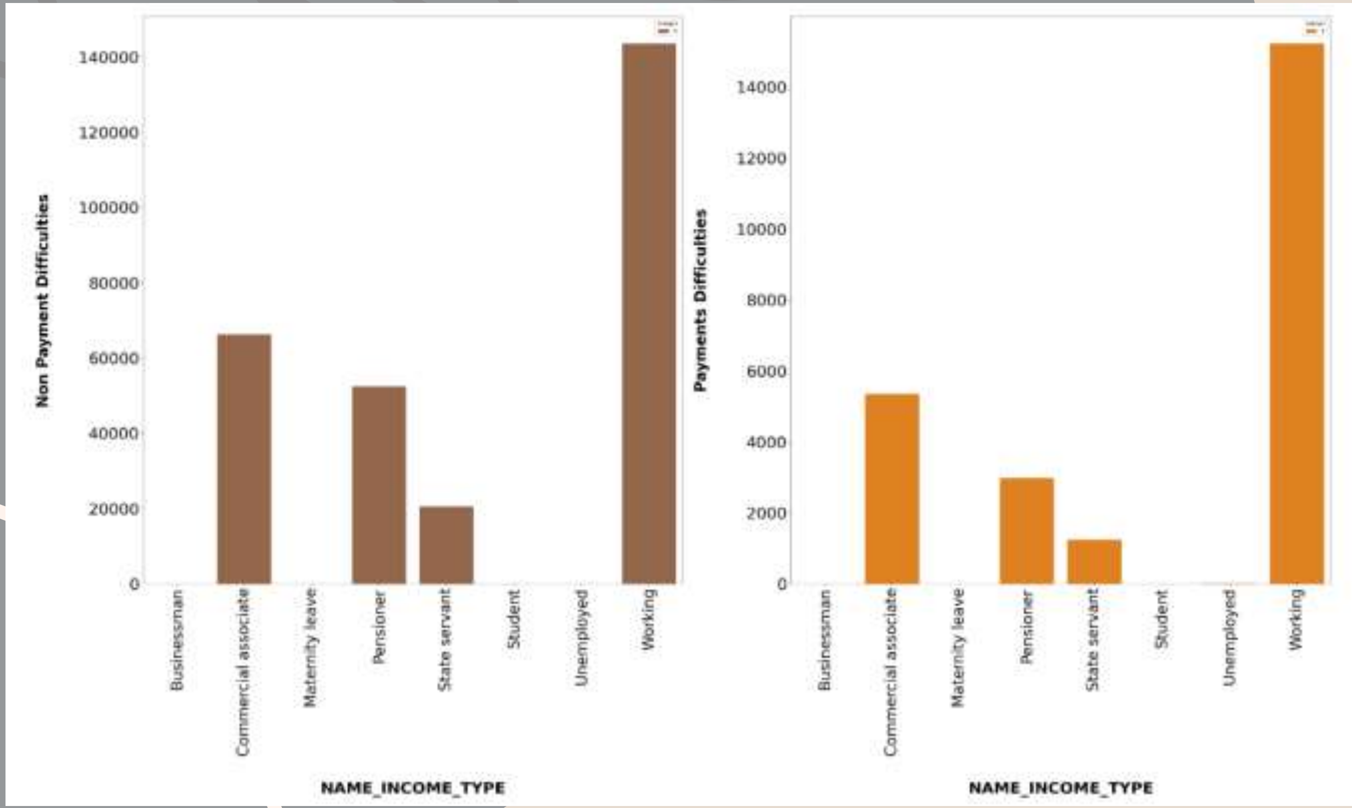
- ✓ Categorical analysis of payment difficulties based on the contract type (NAME-Contract_TYPE) reveals that a majority of cases, whether facing payment difficulties or not, are primarily handled through cash loans.

Housing Type Impact: Analyzing Payment Difficulties:



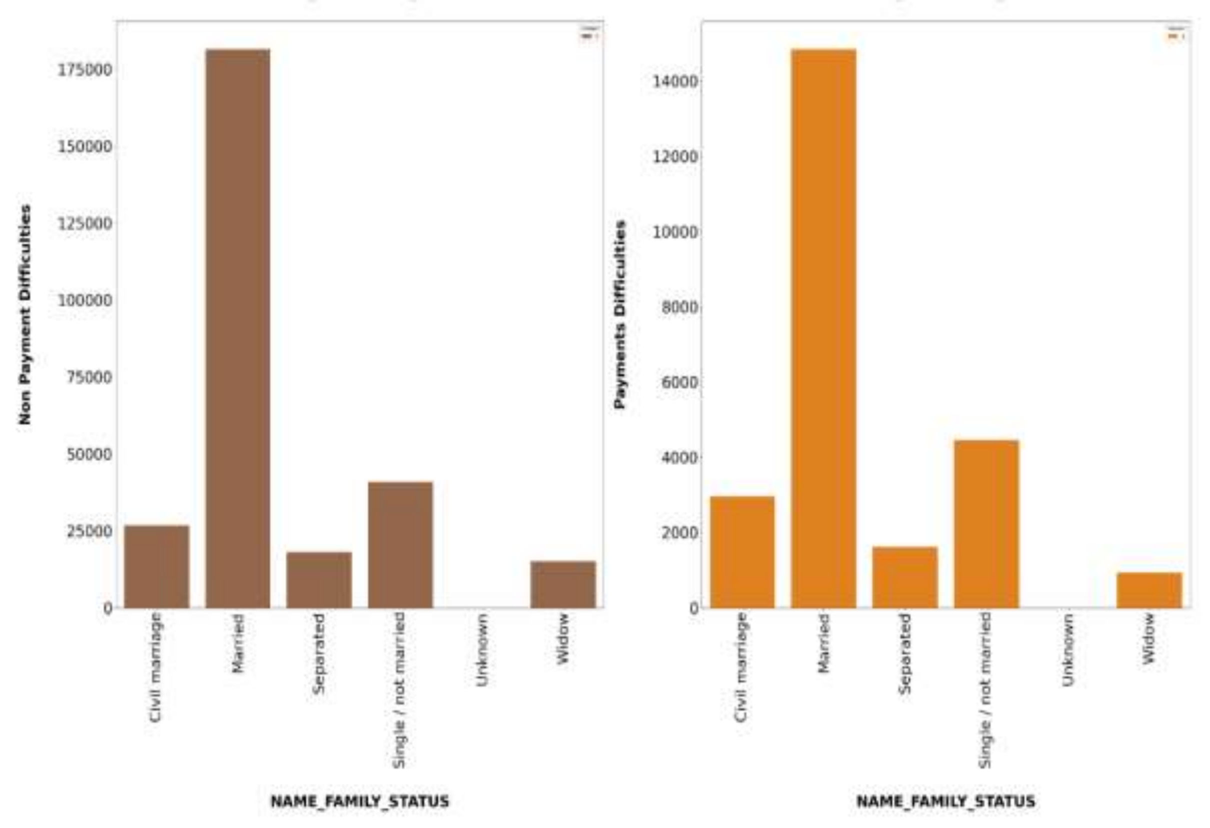
- ✓ This analysis explores payment difficulties among individuals with different contract types (NAME-Housing_TYPE), focusing on distinctions between apartment and house residents
- ✓ Insights from this study provide targeted information for addressing financial challenges based on housing situations.
- ✓ The primary objective is to discern whether residents in apartments or houses encounter a higher proportion of payment difficulties, shedding light on potential correlations between housing situations and financial challenges.

Analyzing Categorical Trends with NAME-INCOME_TYPE:



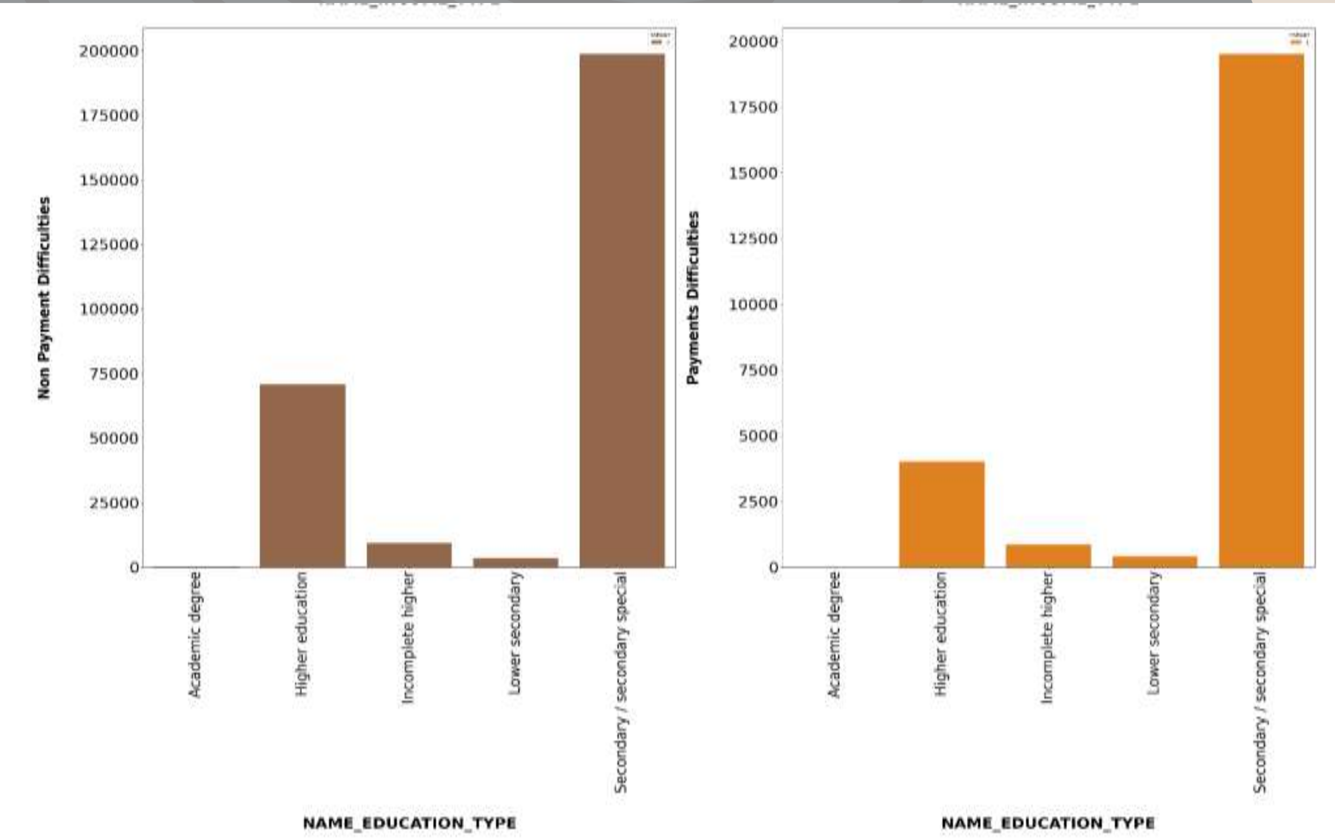
- ✓ In this categorical analysis utilizing the variable NAME-INCOME_TYPE, the examination focuses on payment difficulties and non-payment difficulties within different income types.
- ✓ Notably, the observation highlights a significant ratio of working people in both scenarios.
- ✓ The findings emphasize the considerable impact and relevance of the working population in the context of payment challenges and non-payment difficulties.

The Impact of Marital Status on Payment Difficulties:



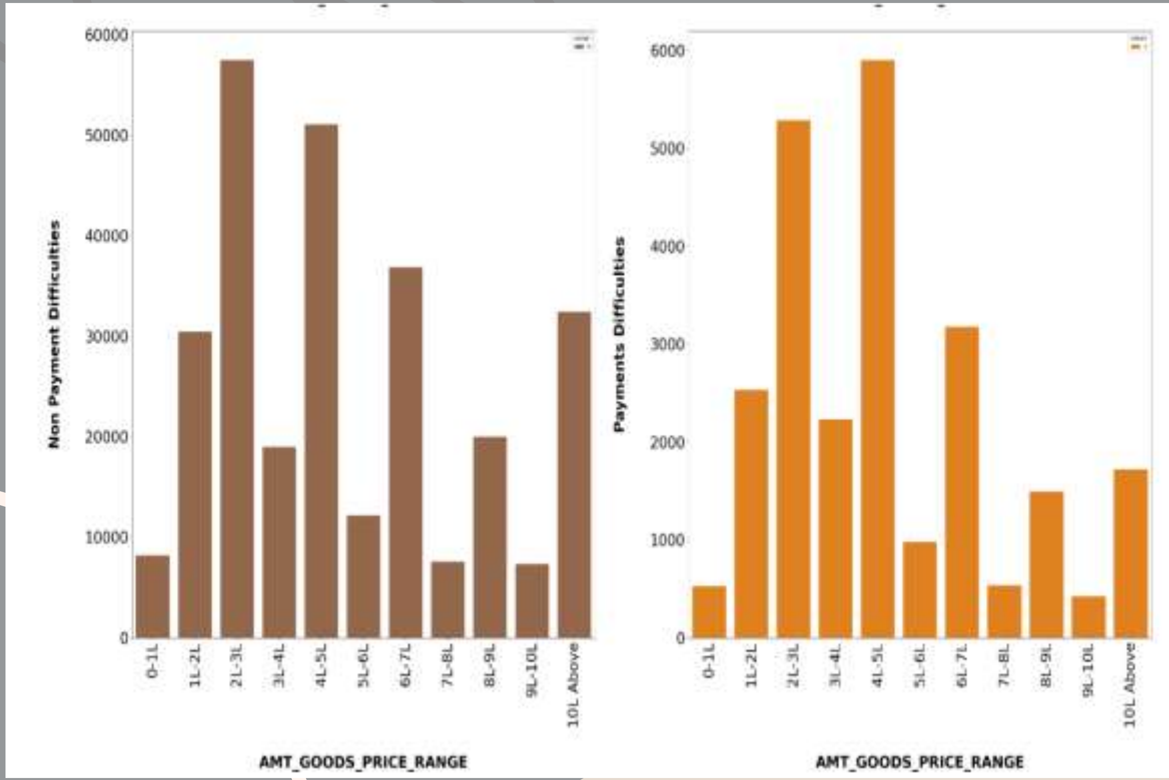
- ✓ The findings highlight a substantial presence of married individuals in both scenarios, indicating a noteworthy proportion of this demographic facing challenges in meeting financial obligations.
- ✓ The exploration sheds light on the importance of considering marital status in understanding patterns of payment behavior and financial struggles.

Education Type Analysis: Categorizing Payment Difficulties:



- ✓ In this categorical analysis, we explore the relationship between payment difficulties and non-payment difficulties with respect to different education types represented by the variable NAME-EDUCATION_TYPE.
- ✓ Notably, the investigation reveals that individuals with a secondary/secondary special education background experience a higher prevalence of payment difficulties compared to other education types.
- ✓ This observation underscores the importance of understanding the unique challenges faced by this specific educational demographic within the context of financial obligations and payment

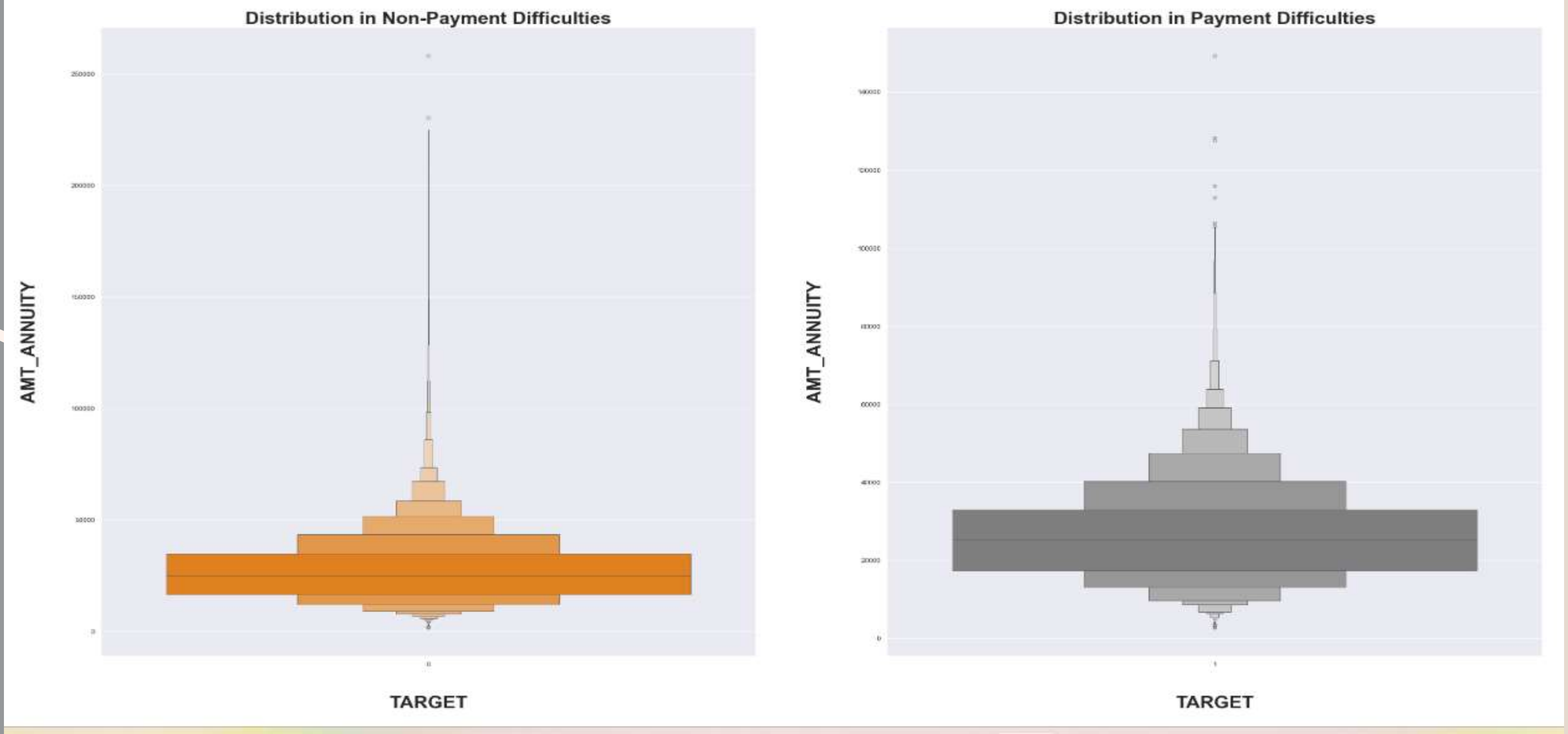
Payment Challenges in the Goods Price Range:



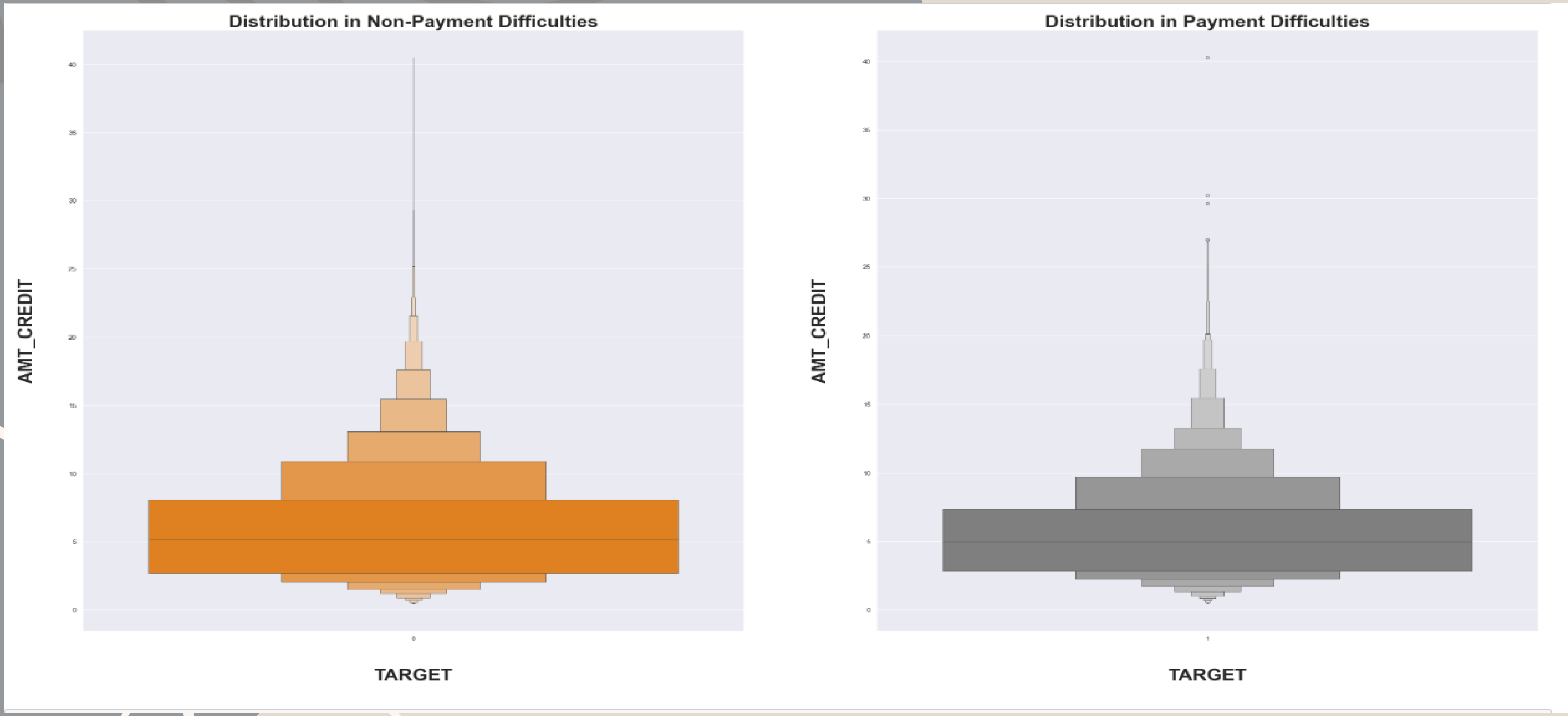
- ✓ The investigation reveals that individuals within this range encounter a higher prevalence of payment difficulties compared to other price brackets.
- ✓ This observation underscores the significance of understanding the distinctive challenges faced by individuals seeking goods within the specified price range in the context of financial commitments and payment outcomes.

Univariate Analysis of Numerical Columns with Respect to Loan Payment Difficulties

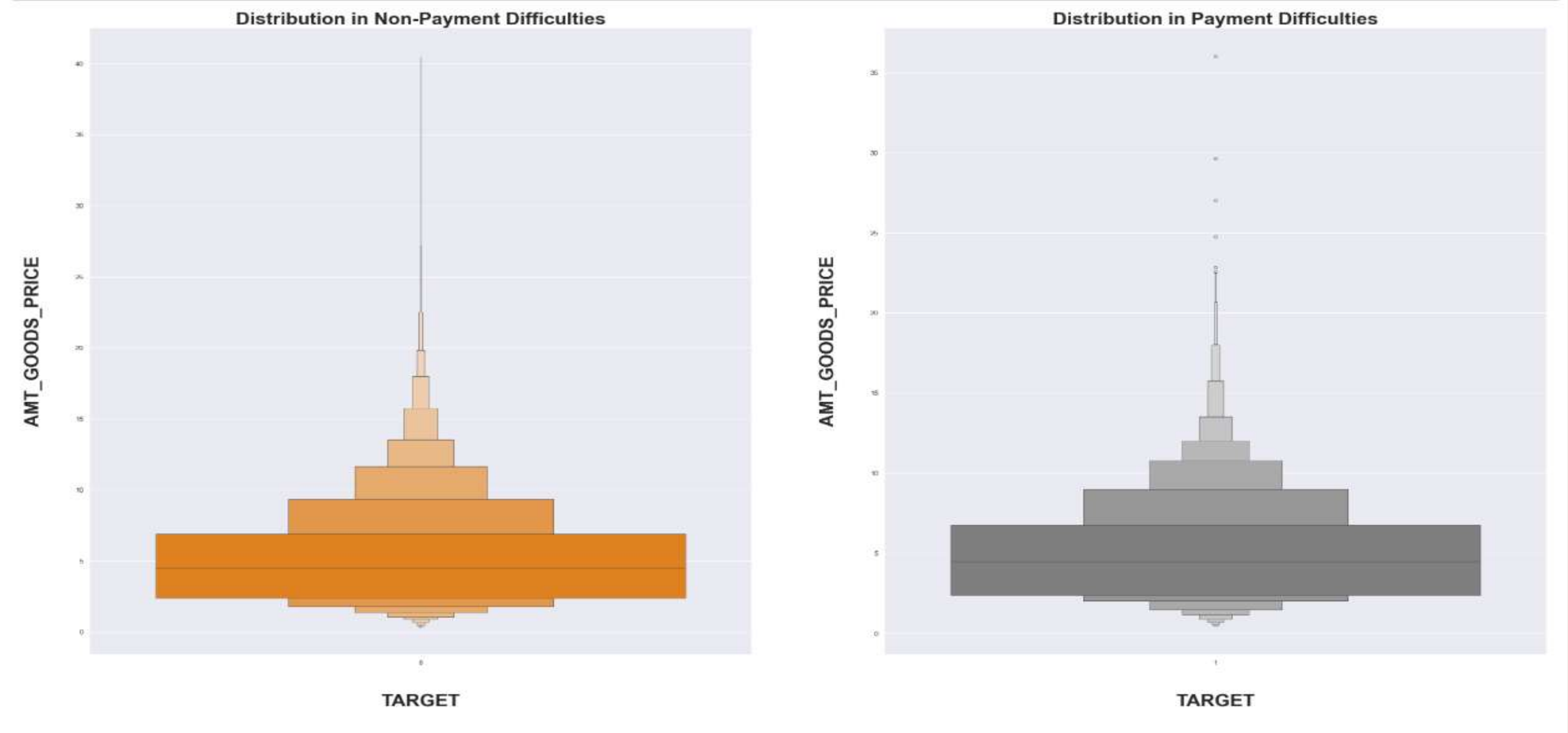
Comparison of AMT_ANNUITY distribution in loan applicants with and without payment difficulties



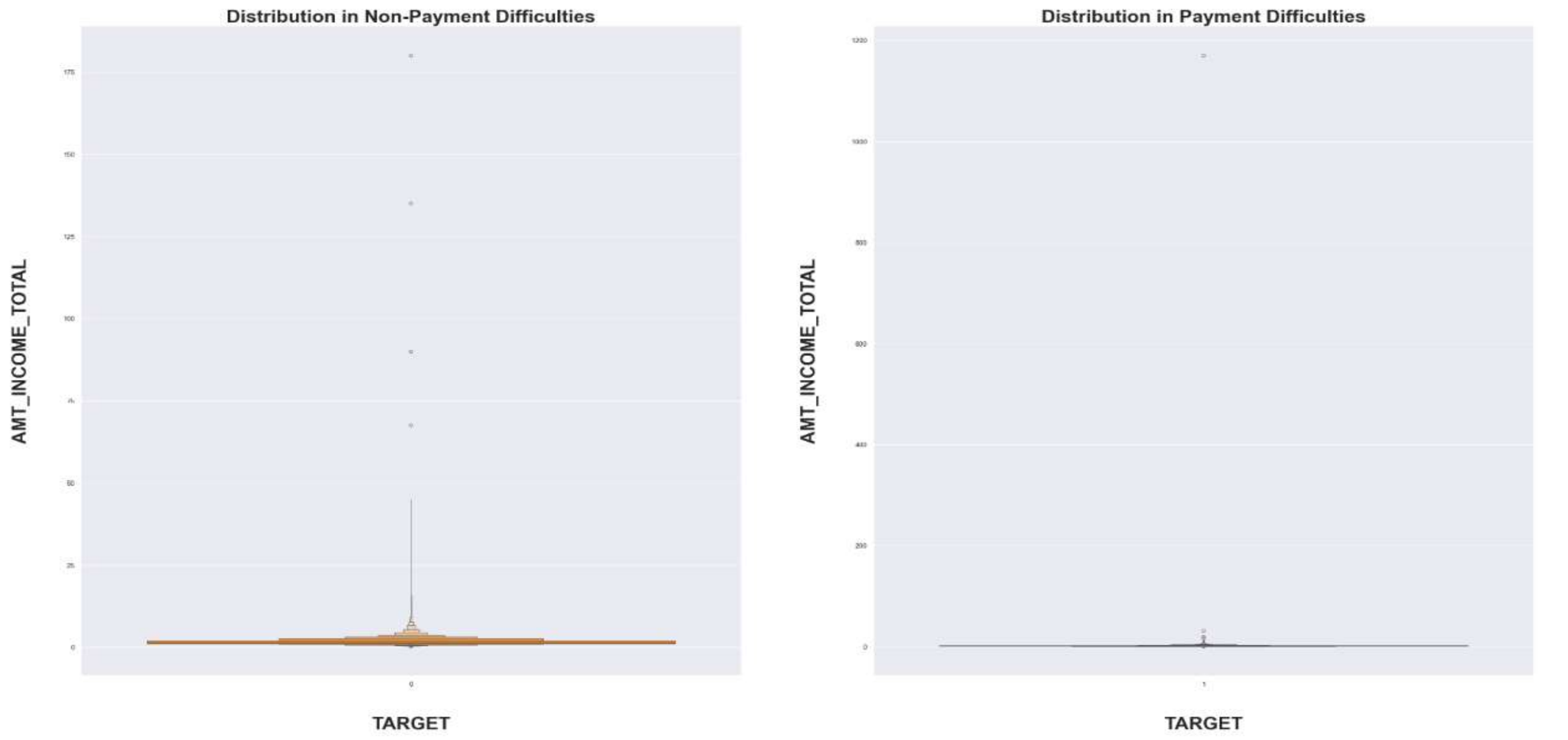
Comparison of AMT_CREDIT distribution in loan applicants with and without payment difficulties



Comparison of AMT_GOODS_PRICE distribution in loan applicants with and without payment difficulties



Comparison of AMT_INCOME_TOTAL distribution in loan applicants with and without payment difficulties



Boxen Plot Analysis:

Observation:

- ✓ *Boxen plots reveal distinct patterns in income, annuity, credit, and goods price for both target 0 and target 1.*

Income total.

- ✓ *The boxen plot for target 1 exhibits a more staggered and wider distribution compared to target 0, indicating greater variability in income among individuals facing payment difficulties.*

Annuity, credit, and goods price:

- ✓ *Similar boxen plot shapes for target 0 and target 1 suggest comparable distributions of annuity, credit, and goods price across both groups*

Overall shape:

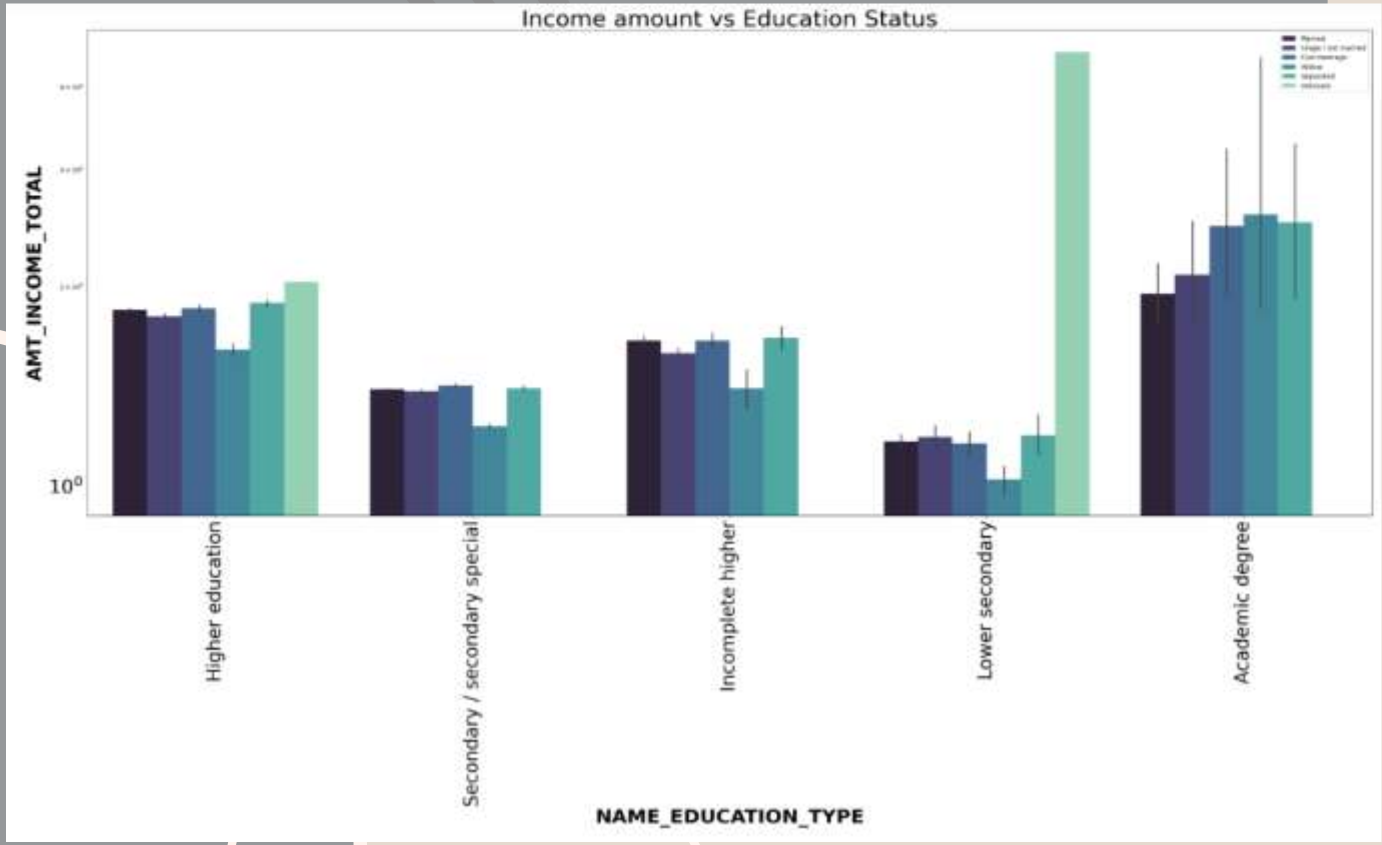
- ✓ *The wider curve shape in the boxen plot for target 1 signifies a broader and more spread-out distribution, implying higher variability in these financial metrics among individuals experiencing payment difficulties.*

Well-defined edges:

- ✓ *The narrower and well-defined edges in the boxen plot for target 0 suggest a more concentrated and predictable distribution in comparison to target 1.*
- ✓ *In summary, the boxen plot analysis provides a detailed view of the distributional characteristics of financial metrics, highlighting differences between individuals with and without payment difficulties.*

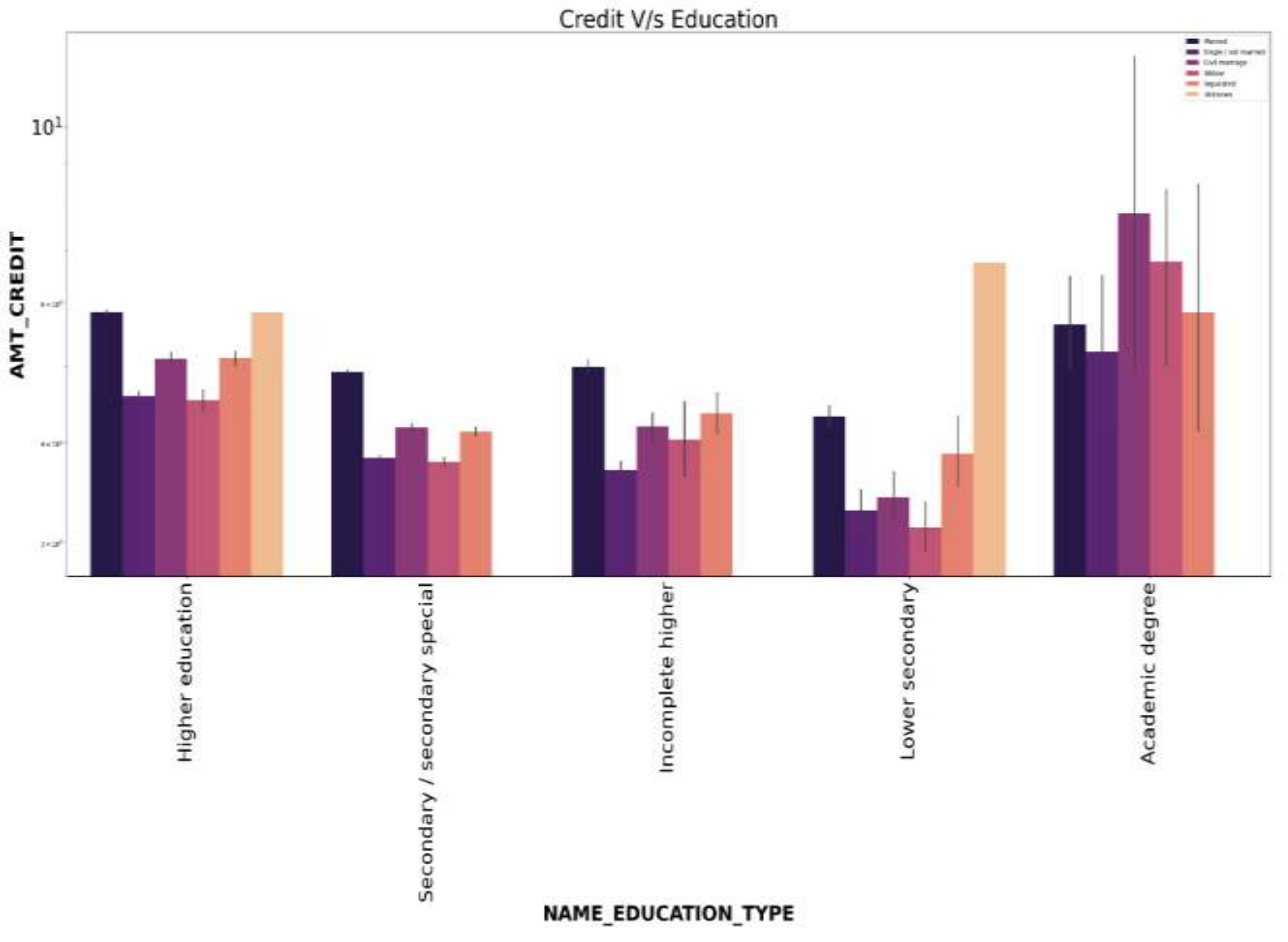
Bivariate Analysis: Numerical and Categorical Variables with Target Variables

Income Disparities Across Education Levels and Family Status:



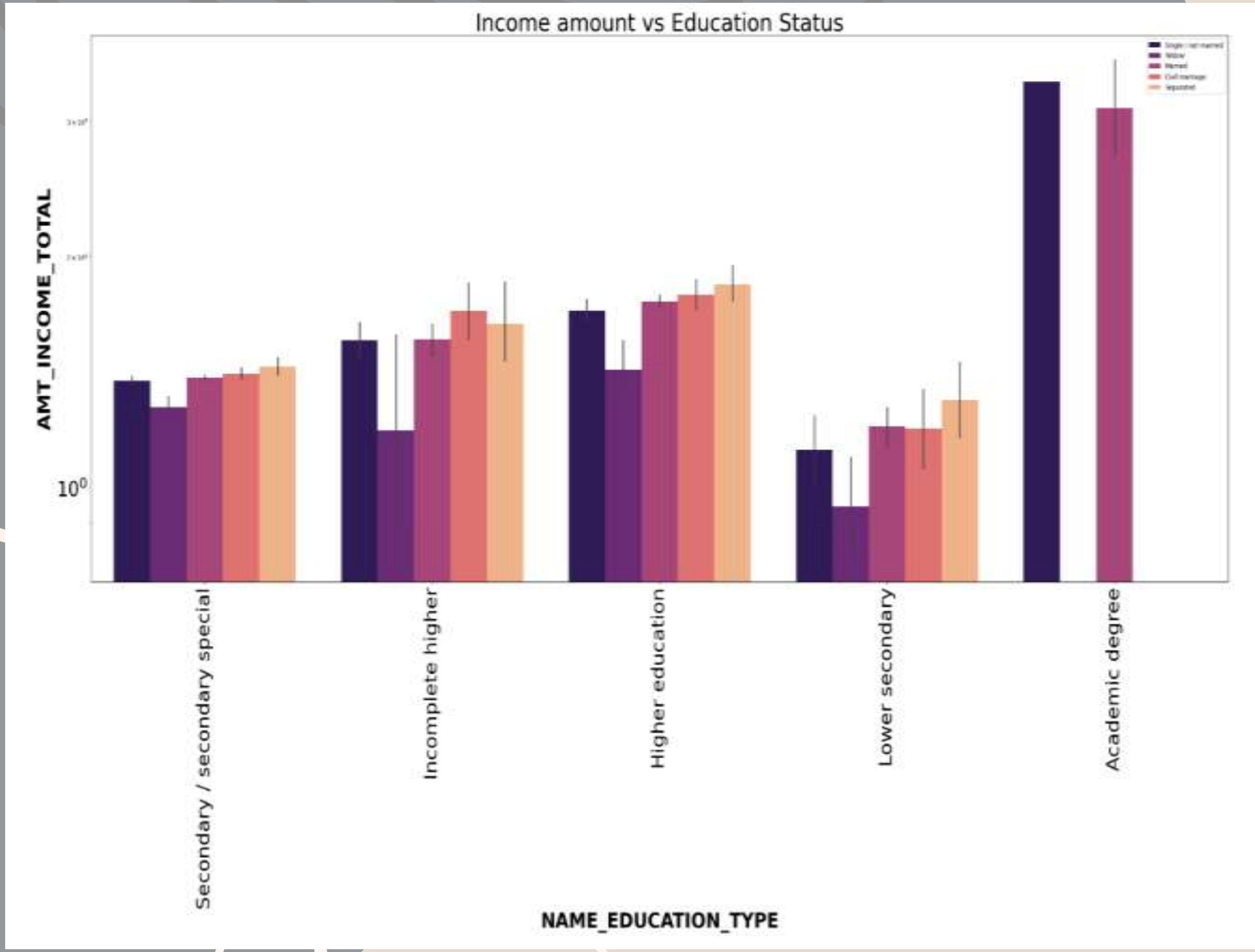
- ✓ Widow clients with academic degrees exhibit few outliers, lacking both the first and third quartiles.
- ✓ In terms of income, clients with higher education, incomplete higher education, lower secondary education, and secondary/secondary special education tend to have incomes below the first quartile, with a notable number of outliers.
- ✓ Additionally, clients with secondary/secondary special education can have higher incomes, indicating diverse income patterns among different education levels and family statuses.

Credit Allocation Based on Education and Family Status:



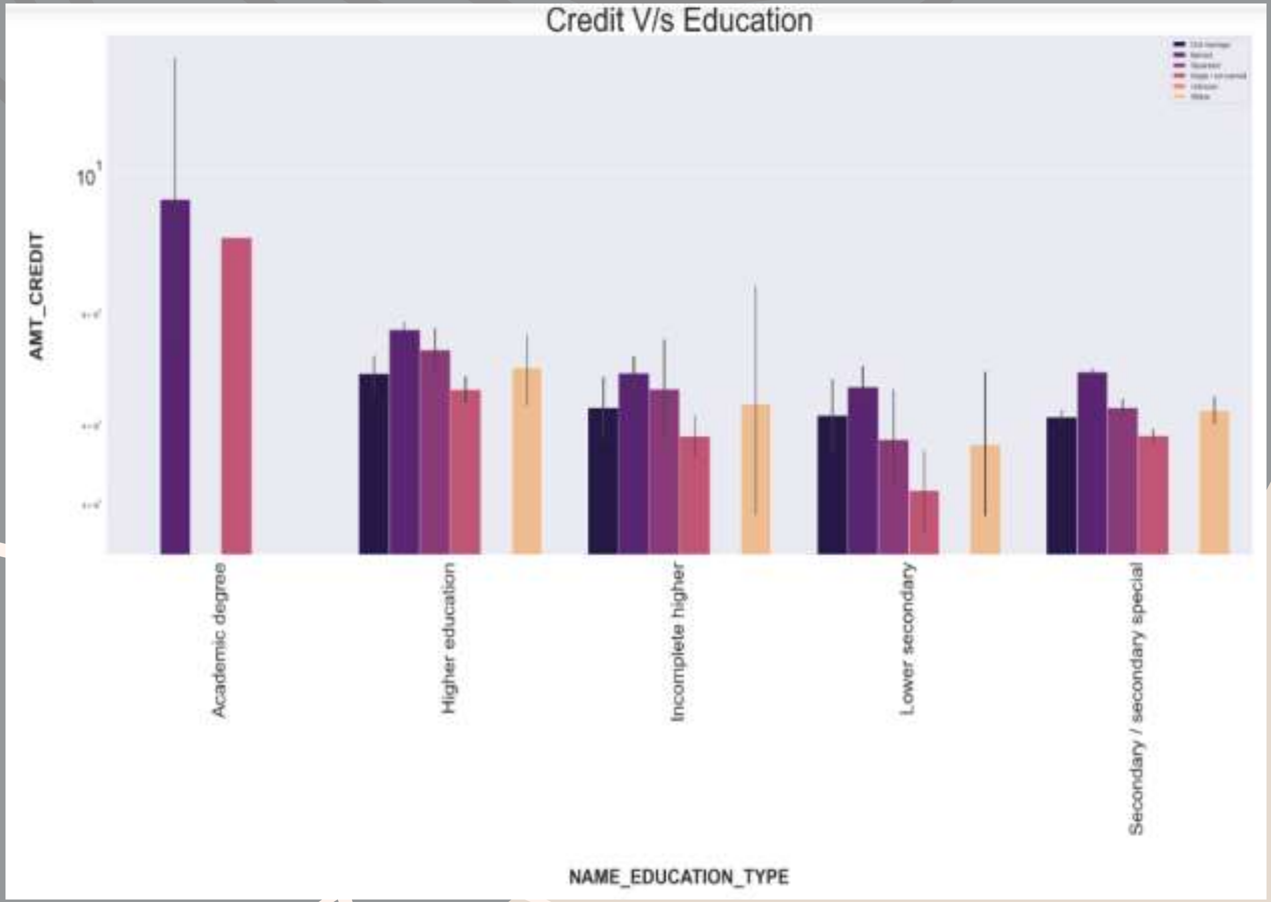
- ✓ The majority of clients have credit amounts below the 25th percentile. Notably, clients with an academic degree and those who are widowed tend to apply for higher credit loans.
- ✓ Additionally, clients with higher education, incomplete higher education, lower secondary education, and secondary/secondary special education also show a propensity for higher credit amounts, highlighting diverse borrowing patterns across different education levels and marital statuses.

Income Disparities Across Education Levels and Family Status for Defaulters:



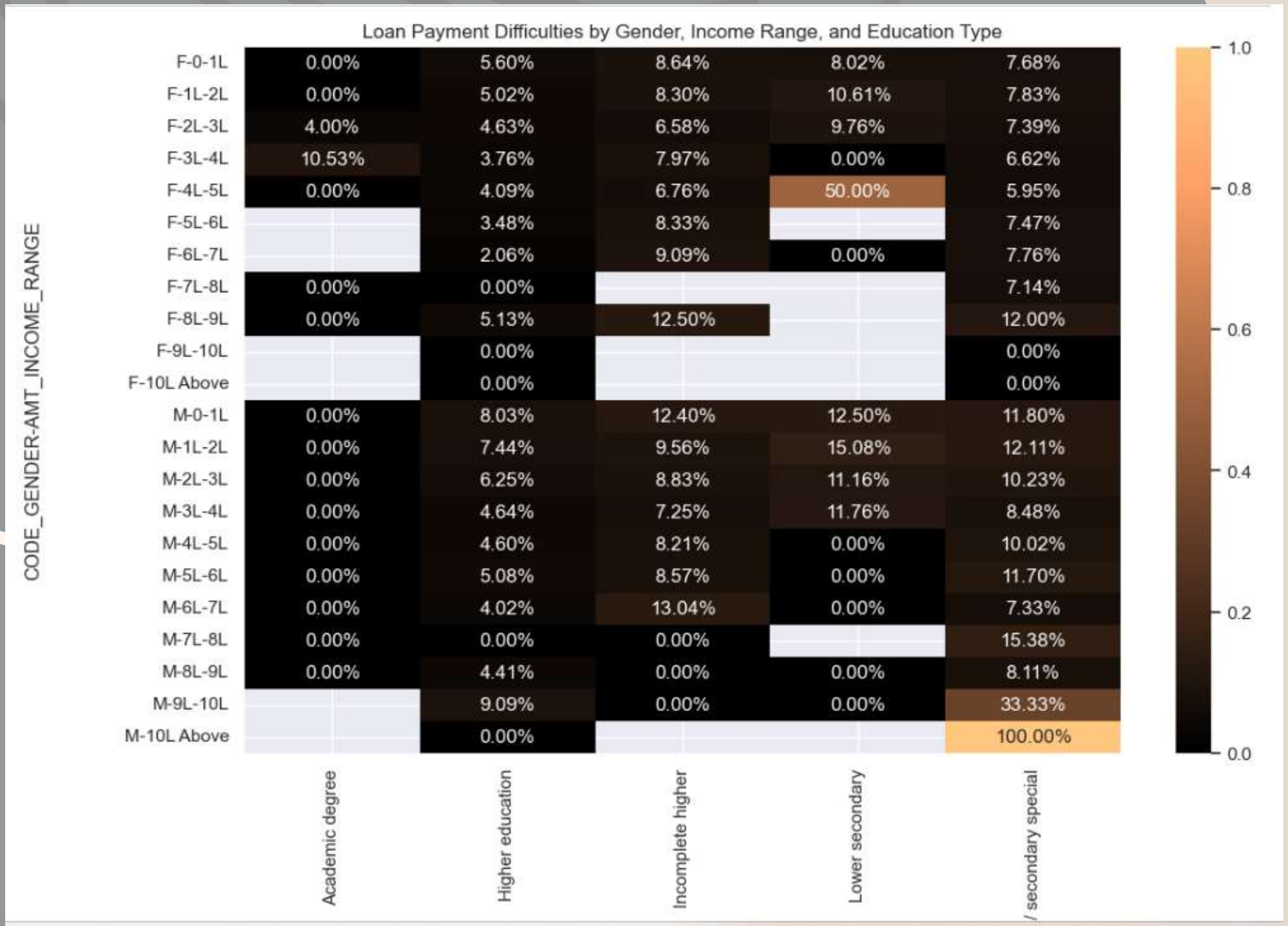
- ✓ Married clients holding academic degrees show notably higher incomes compared to counterparts with different education levels.
- ✓ Additionally, within the defaulter category, clients tend to have relatively lower incomes compared to those classified as non-defaulters.

Credit Allocation Based on Education and Family Status for Defaulters:



- ✓ Married clients with academic degrees exhibit a higher tendency to apply for higher credit loans.
- ✓ In contrast, single clients with academic degrees suggests a more concentrated credit range.
- ✓ Moreover, clients with higher education, incomplete higher education, lower secondary education, and secondary/secondary special education show a propensity for higher credit amounts, indicating diverse borrowing patterns across various education levels.

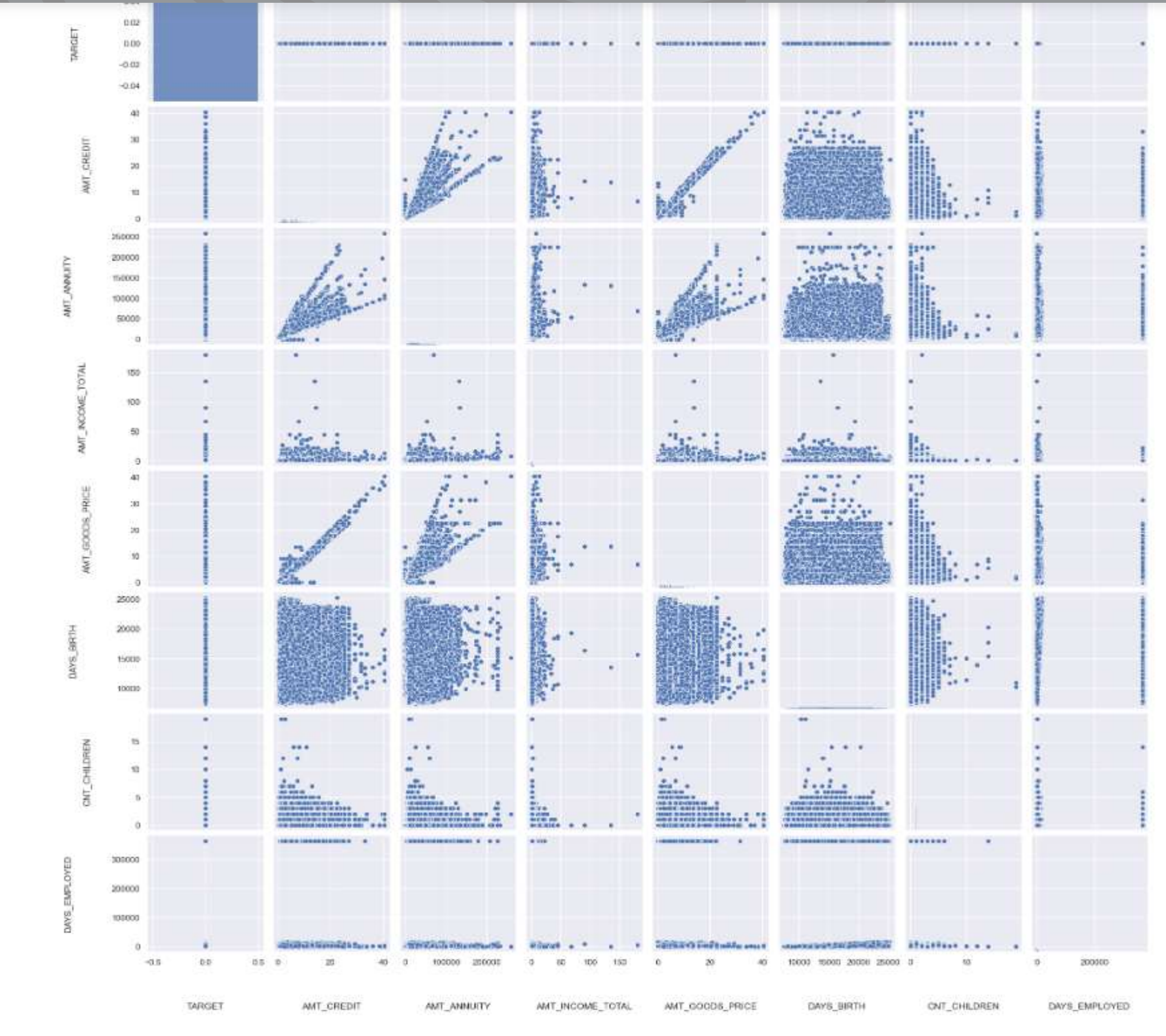
Loan Payment Difficulties by Gender, Income Range, and Education Type:



- ✓ These insights underscore the importance of tailored financial services, considering both education levels and income ranges. Implementing targeted interventions, such as financial education programs and flexible repayment options, can significantly contribute to mitigating default risks among identified high-risk demographic groups.
- ✓ Specifically, individuals with secondary and lower secondary education levels exhibit a significantly higher frequency of facing payment difficulties compared to other education categories.

Numeric Features and Correlations: Multivariate Insights

Pairwise Relationship Exploration for Non-Defaulters:



This exploration enhances comprehension of the non-defaulting group's financial and demographic characteristics

- ✓ **Trend Identification:**

  Positive or negative trends signify the relationships between variables, offering insights into their interdependencies.

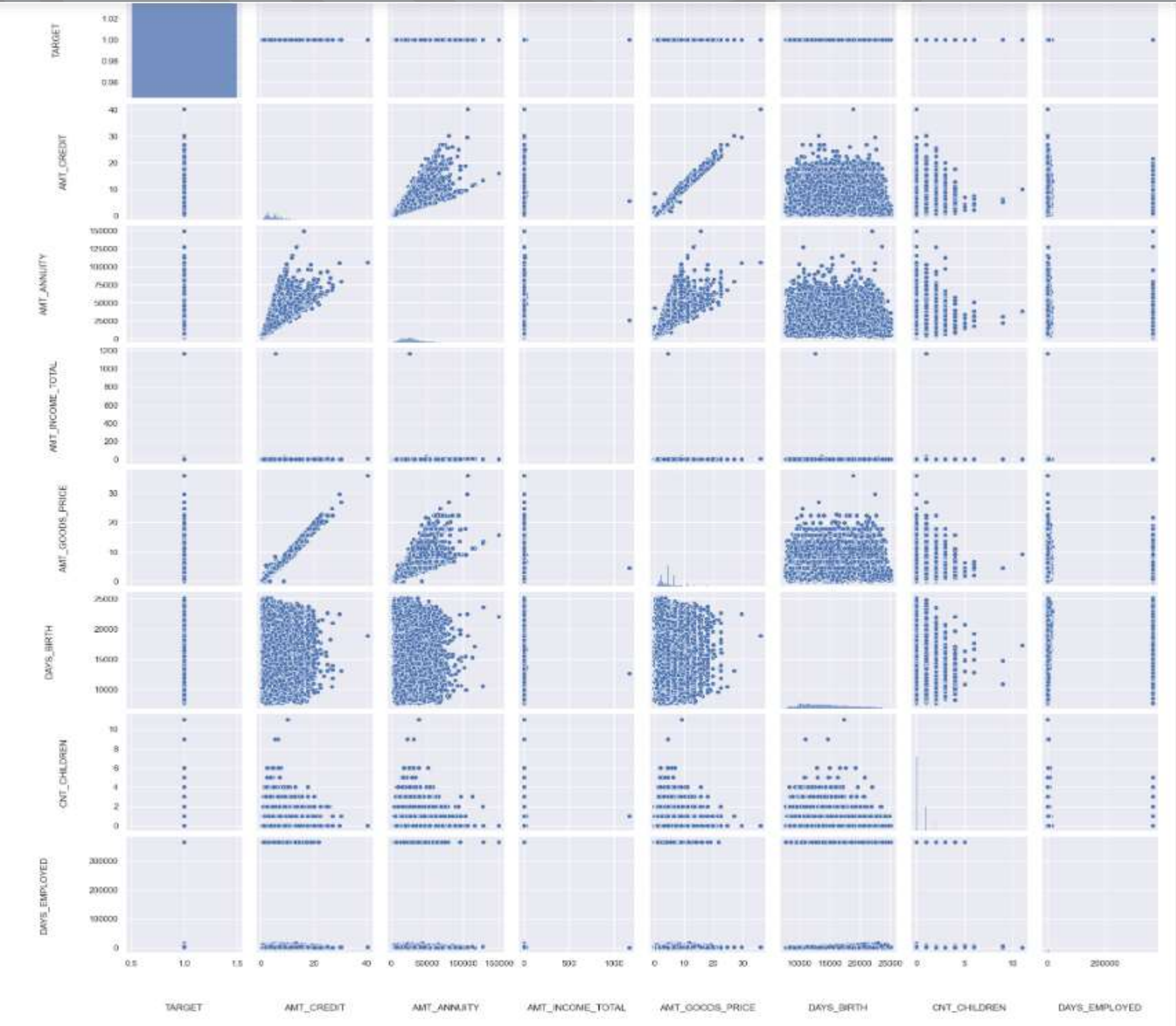
- ✓ **Cluster Analysis:**

  Data point clusters unveil distinct patterns or groups, aiding in the identification of subgroups with shared characteristics.

- ✓ **Distribution Insights:**

  Distributions along the diagonal unveil individual characteristics of each variable, contributing to a holistic view of their data patterns.

Pairwise Relationship Exploration for Defaulters:



- ✓ The pair plot analysis for defaulters provides insightful observations regarding the numerical features associated with loan payment difficulties.
- ✓ Furthermore, the scattered patterns in the distributions depicted in the pair plot underscore the heterogeneous nature of the financial and demographic profiles among defaulters.
- ✓ The absence of distinct trends or clusters in the data points suggests that individuals facing payment difficulties exhibit diverse characteristics across the numerical features under consideration.
- ✓ This understanding is crucial for developing a nuanced approach to addressing loan payment difficulties, taking into account the multifactorial nature of the challenges faced by individuals in this category.

Conclusion and Insights: Characteristics of Non-Defaulters - A Comprehensive Analysis

- ✓ Upon a thorough examination, it is evident that non-defaulters consistently exhibit specific demographic and socioeconomic characteristics. The significant trends include individuals in the middle-age group, engaged in working professions, possessing an academic degree and notably, having a married marital status.
- ✓ The middle-age group plays a significant role, standing out as a demographic factor associated with credit reliability.
- ✓ In terms of professions, working professionals consistently appear among non-defaulters, indicating a correlation between steady employment and creditworthiness.
- ✓ Education background also plays a crucial role, with individuals holding academic degrees, especially in higher education, demonstrating a notable prevalence among non-defaulters.
- ✓ Marital status further adds to the profile, as non-defaulters predominantly belong to the married category, suggesting a higher proportion of credit reliability within this demographic.
- ✓ Together, these factors provide valuable insights into the diverse yet interconnected attributes that consistently define individuals in the non-defaulter category, offering a holistic understanding for decision-making in credit assessments.

Thank You