Oeson Learning

CREDIT RISK ASSESMENT

Liudmila Stolbetskaia

Risk analysis is crucial in financial decisions, helping identify threats and prevent future losses. By combining historical data and forecasting, analysts predict and reduce financial risks

DATA ANALYST

Liudmila Stolbetskaia

Junior Data Analytics with 5+ years of experience in sales development and marketing, specialising in turning data into actionable insights

- EXCEL, PYTHON, POWER BI, SQL
- DATA CLEANING &VISUALISATION
- CRITICAL, ANALYTICAL, STRATEGIC THINKING
- TEAMWORK, LEADERSHIP, TIME MANAGEMENT
- PRESENTATION SKILLS, REPORT WRITING, DATA STORYTELLING

INTRODUCTION TO CREDIT RISK

Credit risk is the likelihood of financial loss that can occur if a borrower fails to repay a loan. It refers to the risk that a lender may not receive the owed principal and interest. Lenders can reduce credit risk by evaluating factors related to a borrower's creditworthiness, such as their existing debt and income.

UNDERSTANDING CREDIT RISK

RISK OF NON-REPAYMENT

Credit risk refers to the possibility of a lender losing money when lending funds to a borrower.

THE FIVE CS

Credit risk evaluated using the five Cs: credit history, repayment capacity, capital, loan conditions, and the associated collateral.

CREDIT SCORES

Credit score is one of the factors lenders use to determine the likelihood of you defaulting on a loan.

INTRODUCTION TO THE PROJECT

BUSSINES UNDERSTANDING

Loans are commonly used for various purposes, making it essential to evaluate an applicant's ability to repay to minimize financial risk. When an application is received, data analysis is performed using methods to assess the applicant's eligibility. The decision to approve or reject a loan is based on the applicant's risk profile. If the applicant is likely to repay, the loan is low risk. If not, approving the loan could lead to financial loss, making rejection the safer choice.

PROJECT OBJECTIVES

To analyse financial data to identify potential credit risks by examining customer behaviour, application trends, and relevant financial indicators.

GOALS OF THE PROGECT

- Identify key risk factors
- Minimise financial lose of the company
- Analyse customer behaviur
- Improve customer profile understanding

DATA DESCRIPTION

Application Data

- This dataset consists of 124 columns, which include various demographic, financial, and other client-related attributes.
- Key features include personal information (e.g., age, family status, income), financial details (e.g., credit amount, income, property ownership), and region-based features.

Previous Application Data

- This dataset contains 38 columns, which primarily focus on the previous loan applications, including details such as loan amount, approval status, payment types, and other loan-related features.

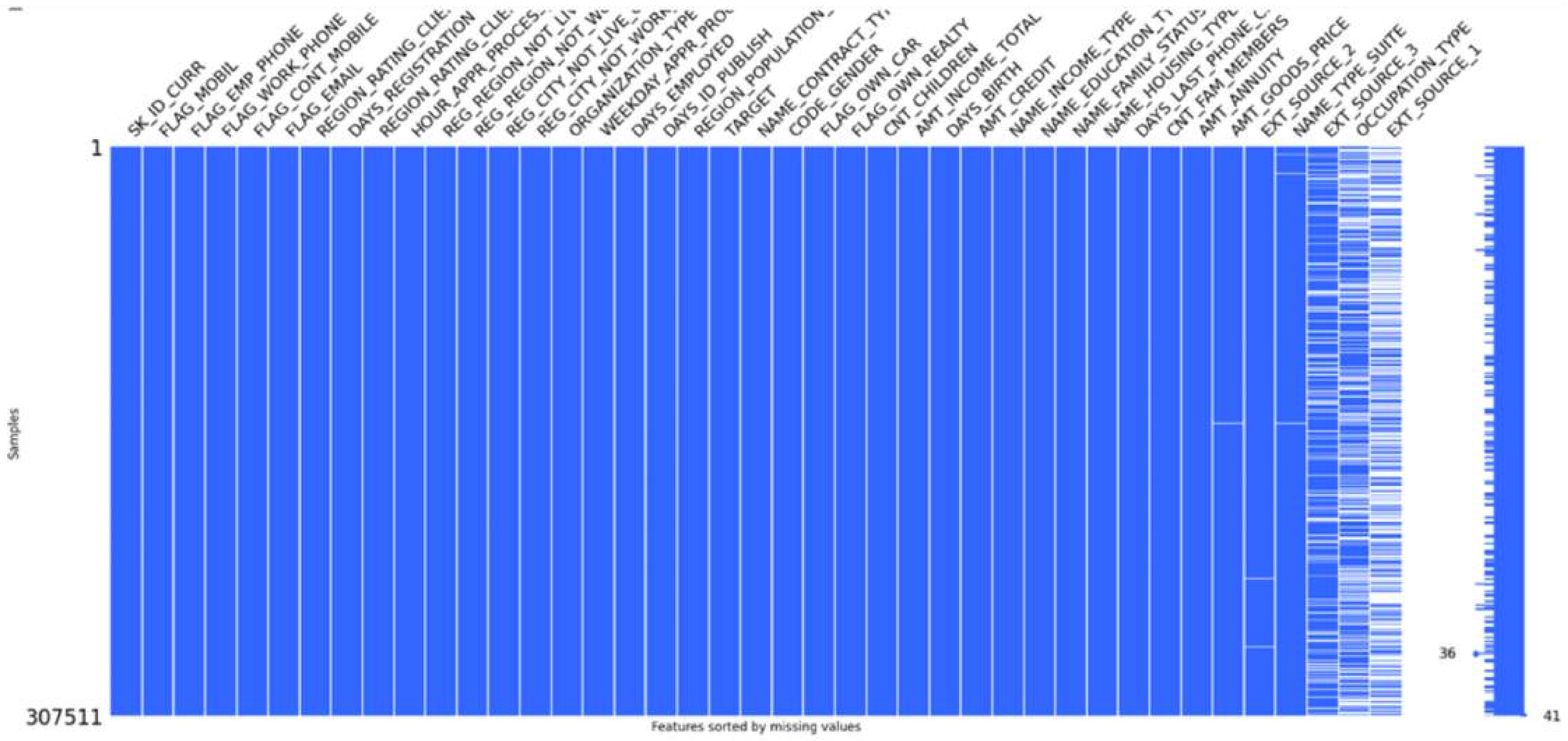
These datasets will be joined on the common column SK_ID_CURR, which represents the unique identifier for the applicants. By joining these datasets, we will combine the details of both the current and previous loan applications for each applicant, enabling a more comprehensive analysis of their behavior, creditworthiness, and other important factors.

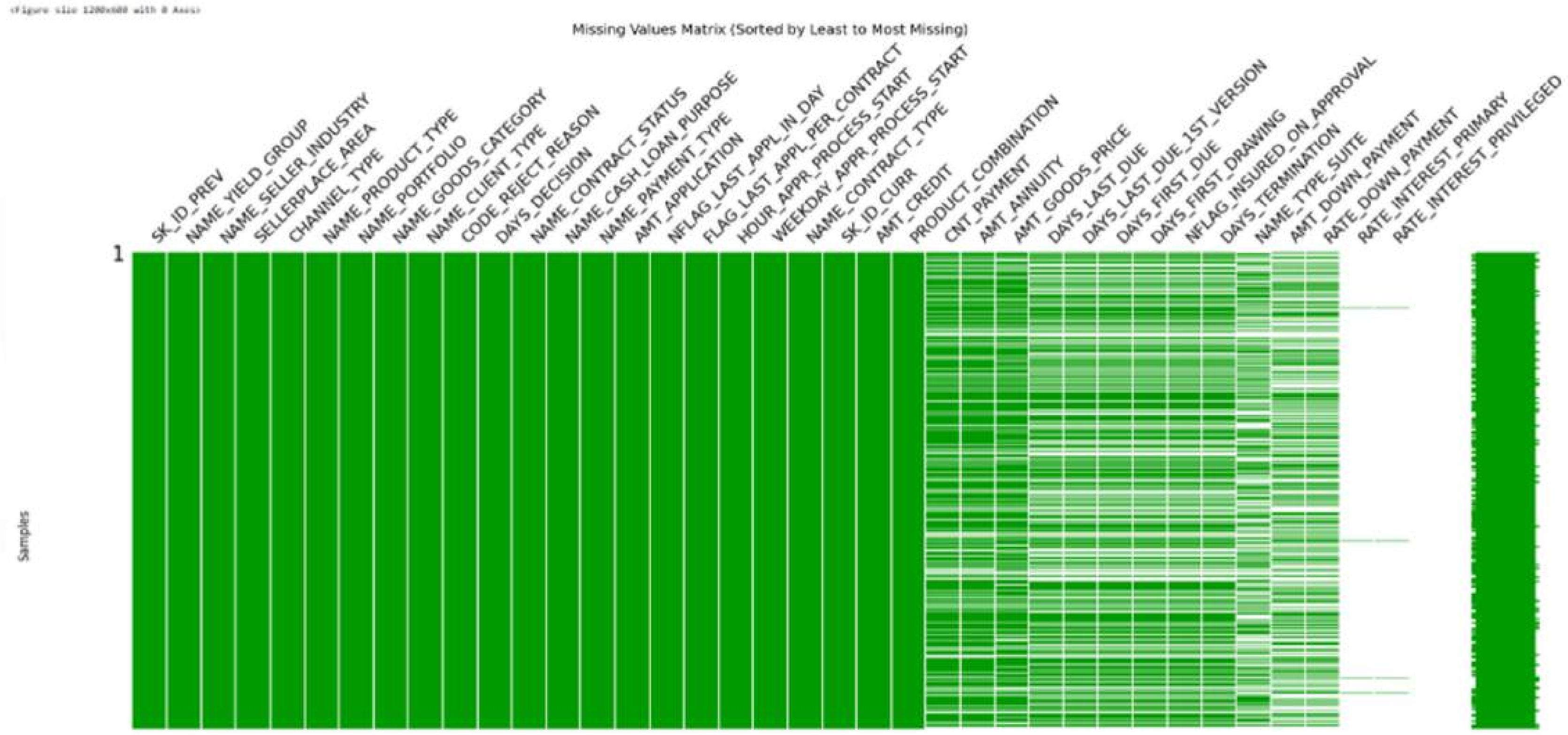
DATA CLEANING

Application & Previous Application

Before starting the analysis, data cleaning was applied to treat missing values in both tables.
The missing values were handled using functions like mean, median, and fillna.

Additionally, outliers were trimmed to ensure the quality and reliability of the data for accurate analysis



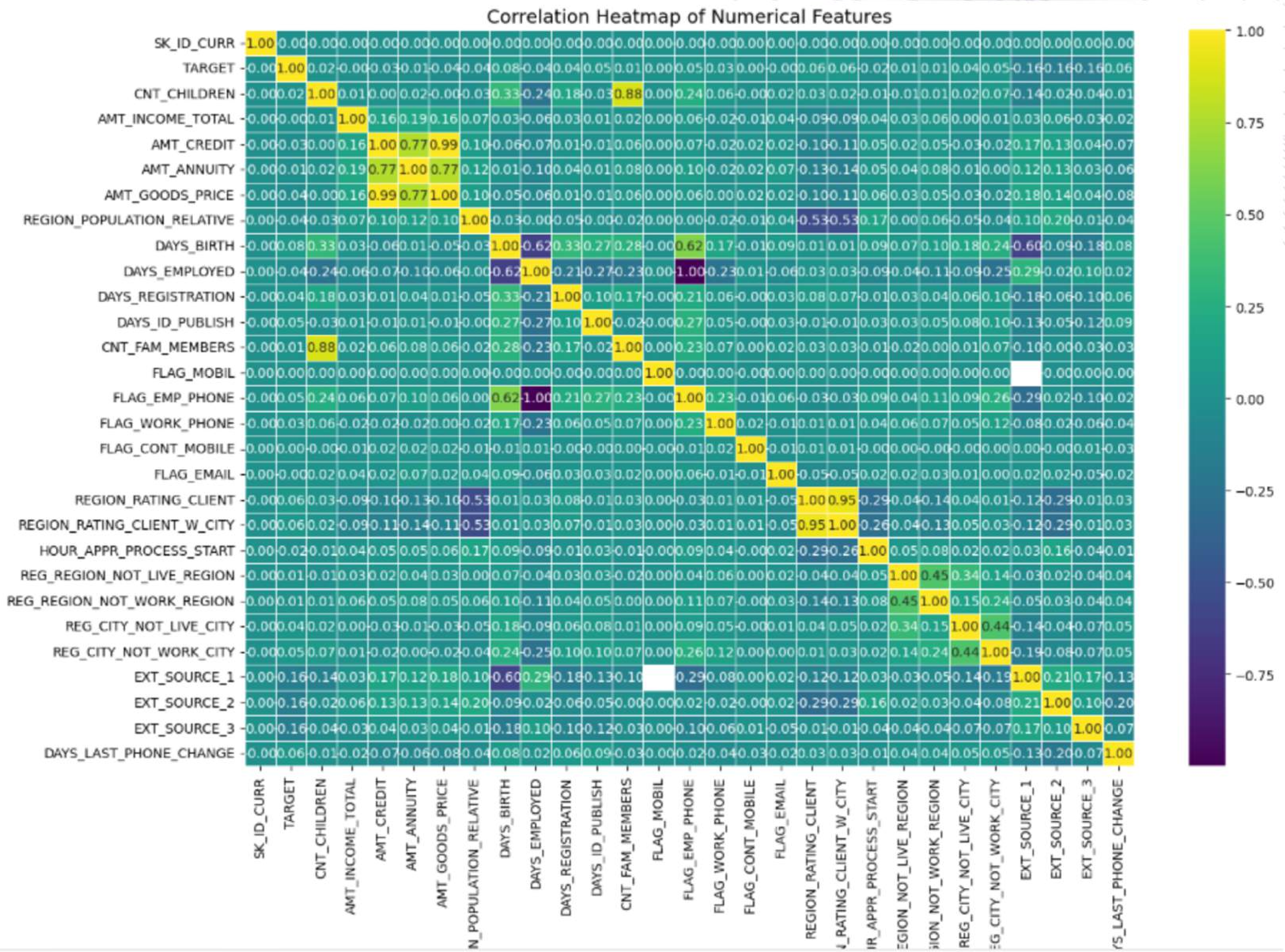


VISUALISATION INSIGHTS

APPLICATION DATA

A correlation matrix can effectively highlight redundancies. High correlation is typically considered above 80%, as this indicates a strong linear relationship between two variables.

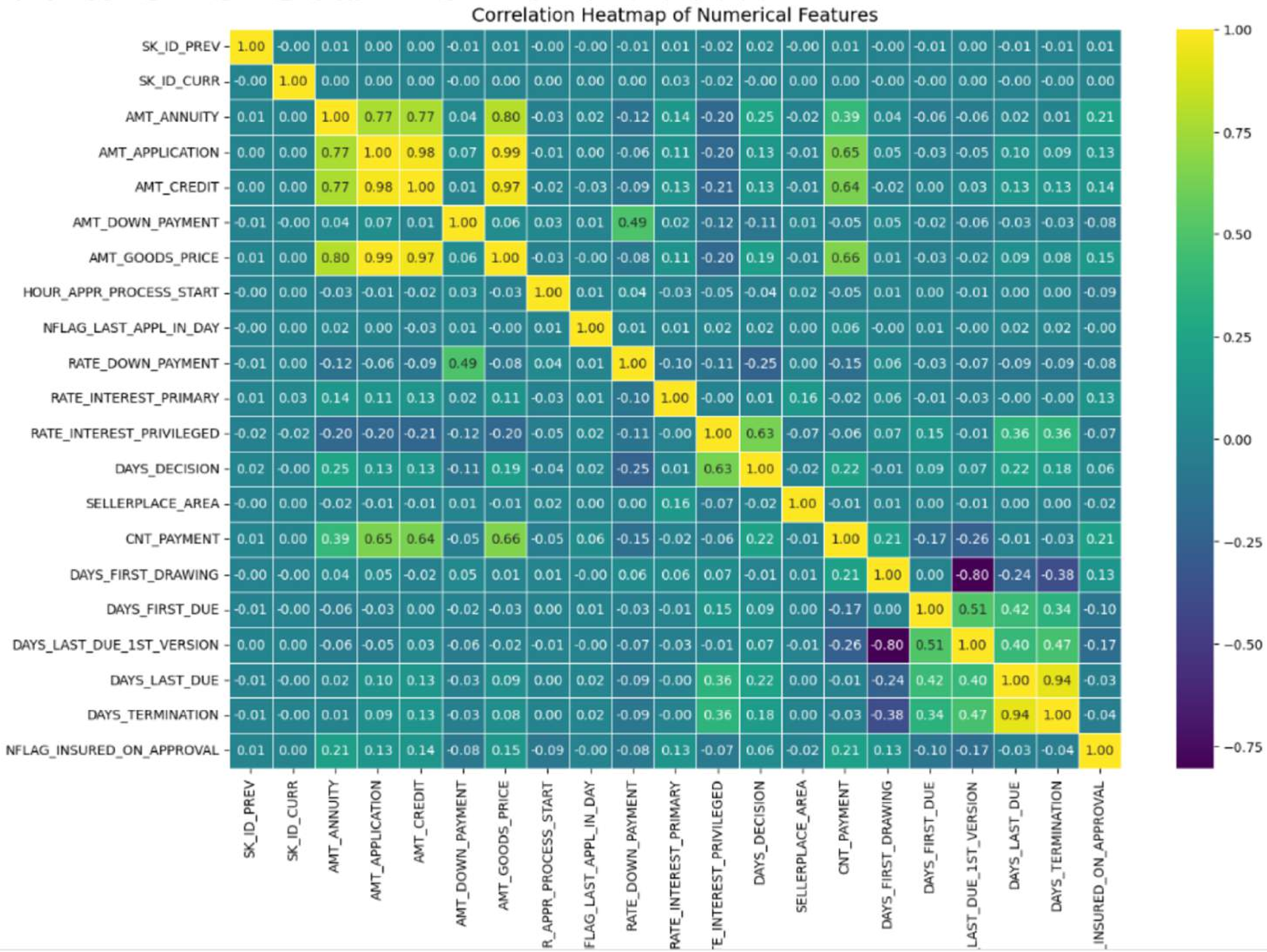
Particularly concerning the variables AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE, CNT_FAM_MEMBERS, and CNT_CHILDREN.



PREVIOUS APPLICATION

A correlation matrix can effectively highlight redundancies. High correlation is typically considered above 80%, as this indicates a strong linear relationship between two variables.

Particularly concerning the variables in this case are AMT_ANNUITY", "AMT_GOODS_PRICE", "RATE_INTEREST_PRIMARY", "RATE_INTEREST_PRIVILEGED", "DAYS_TERMINATION"



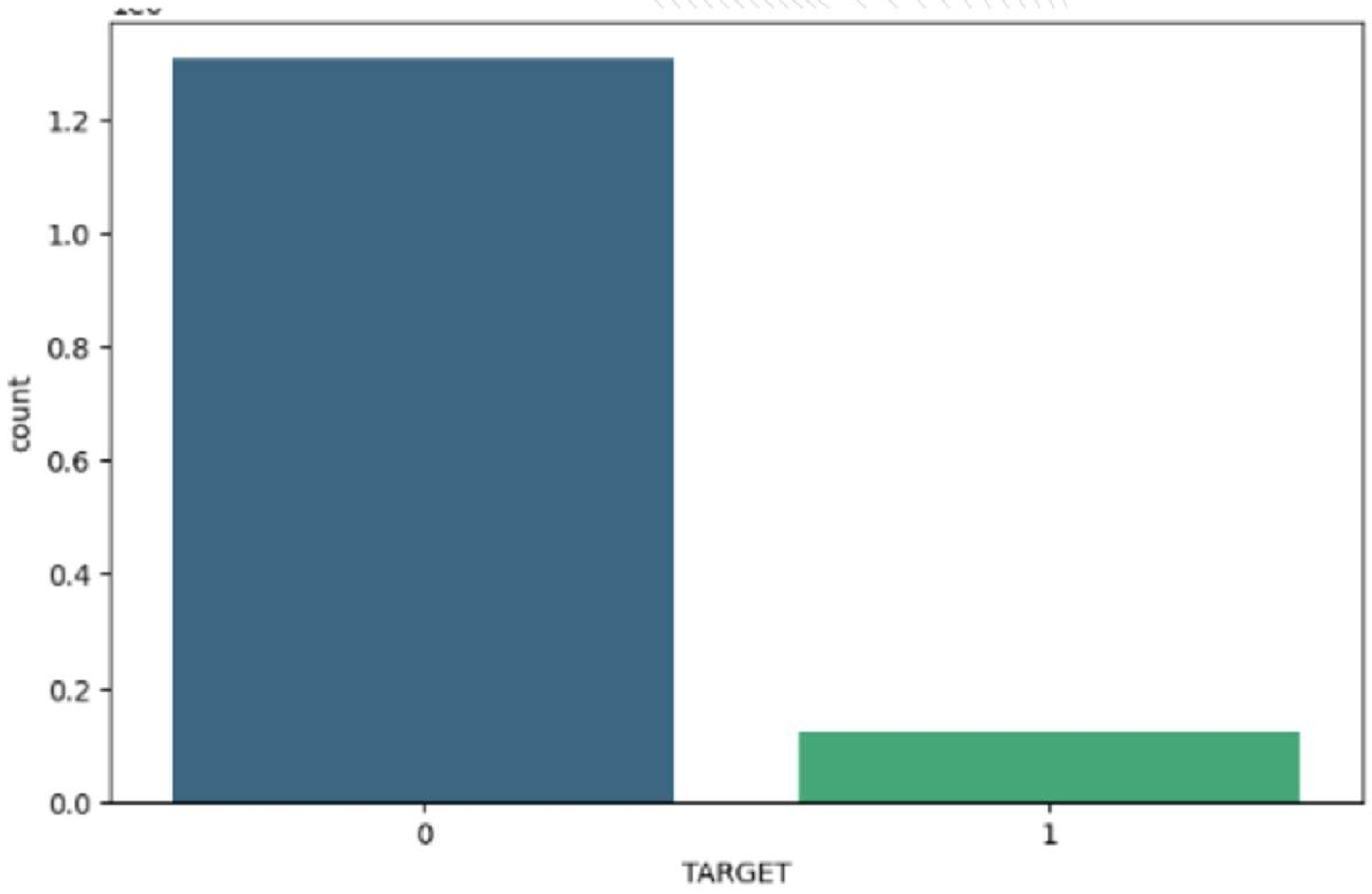
TARGET VARIABLE

TARGET

This tells us the proportion of clients who defaulted (1) versus those who repaid (0).

Insight: If the dataset is imbalanced (e.g., very few defaults), we may need special modeling techniques such as smote to handle the imbalance.

Visualisation indicates that there are more applications that were repaid than those that defaulted.

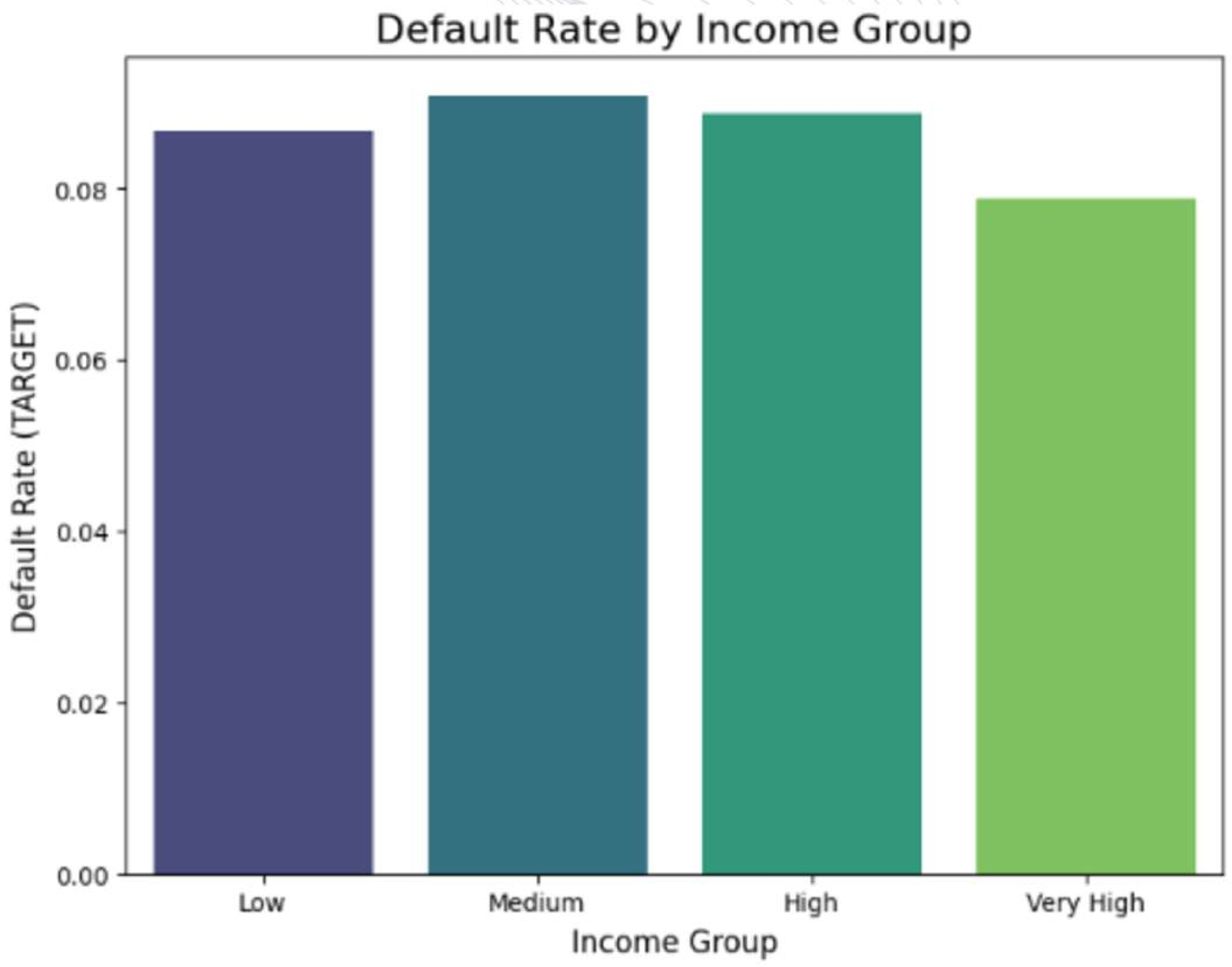


AMT_INCOME_TOTAL VS TARGET

Income is a major factor in a client's ability to repay a loan. If applicants have low income but take high loans, there could be a higher risk of default.

Income grouping provides us with insights into which applications failed based on income.

However, we can see that the income rate might not be entirely reliable, as people with low, medium, high, and very high income levels can still fail.

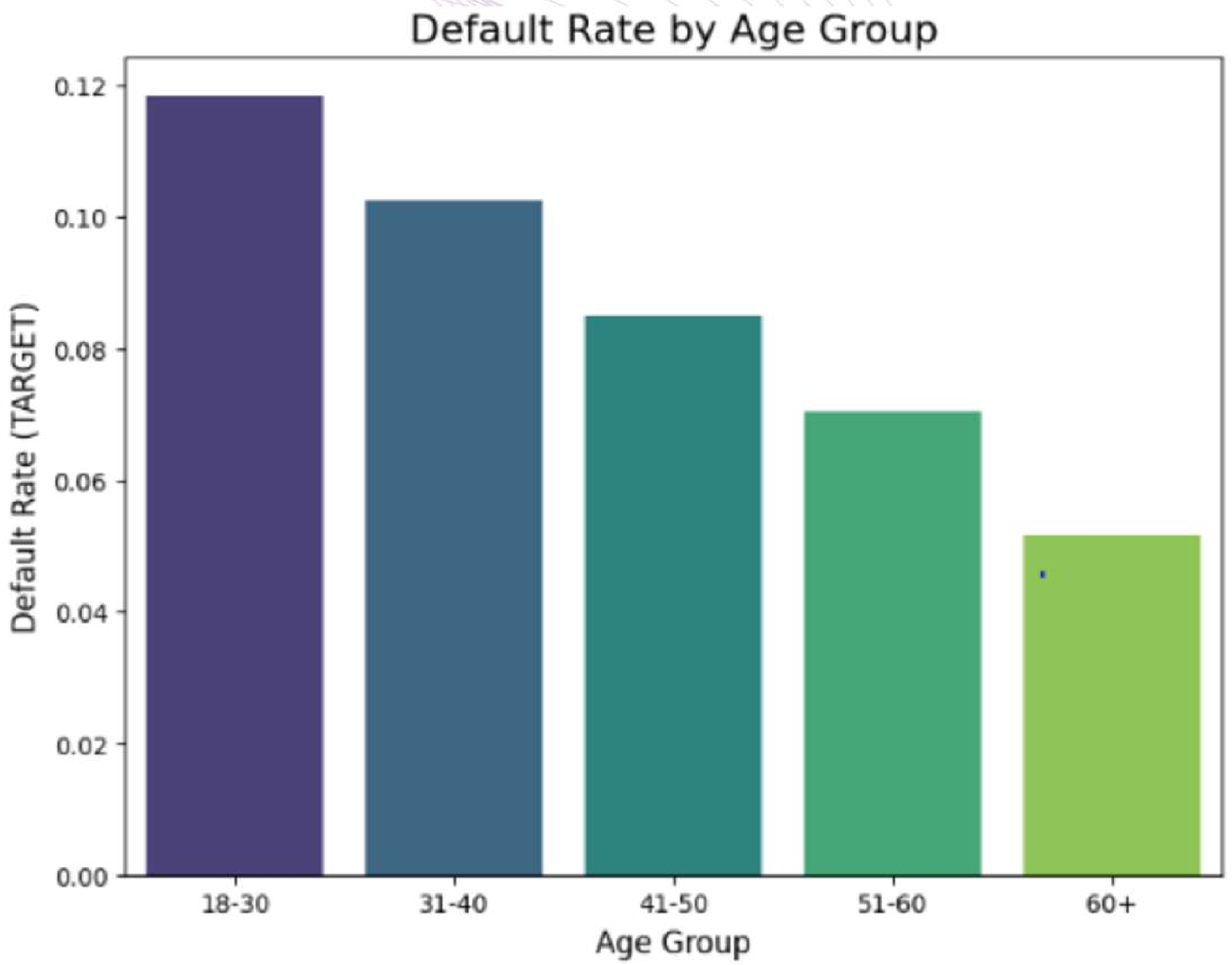


DAYS_BIRTH VS TARGET

The highest default rate is observed in the age group below 25 years, based on the DAYS_BIRTH vs. TARGET visualisation.

This is an important insight as it suggests that younger clients (those under 25) may present a higher credit risk

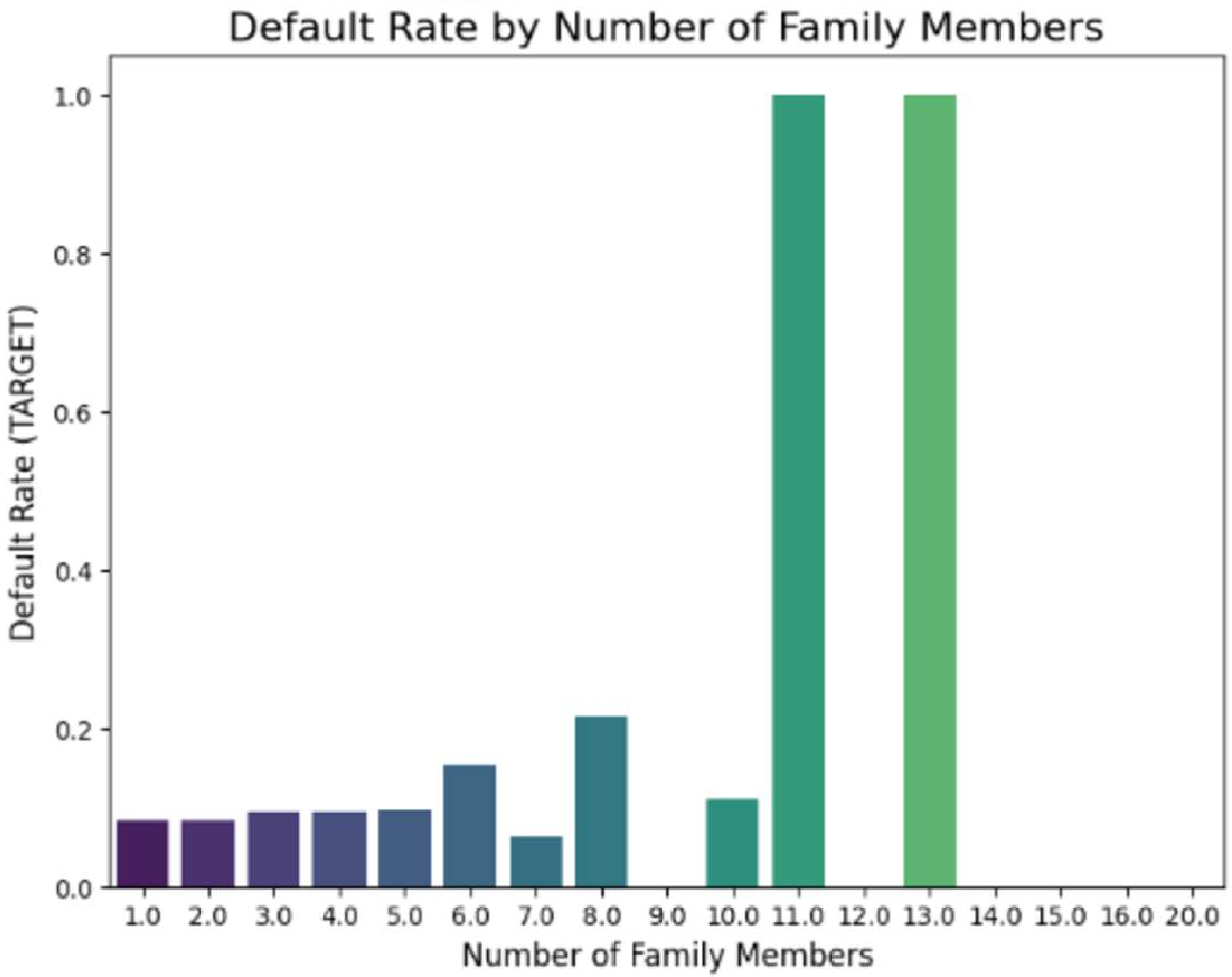
It is clear that older individuals are more likely to be financially stable and repay their loans. They may also have higher incomes compared to younger applicants



CNT_FAM_MEMBERS VS TARGET

The chart indicates that applicants with more than five family members are more likely to be refused a loan. This could be due to the financial burden of supporting a larger family.

In some cases, there may be two working adults, but as the number of family members increases, the available income for repaying the loan may decrease, leading to a higher risk of default



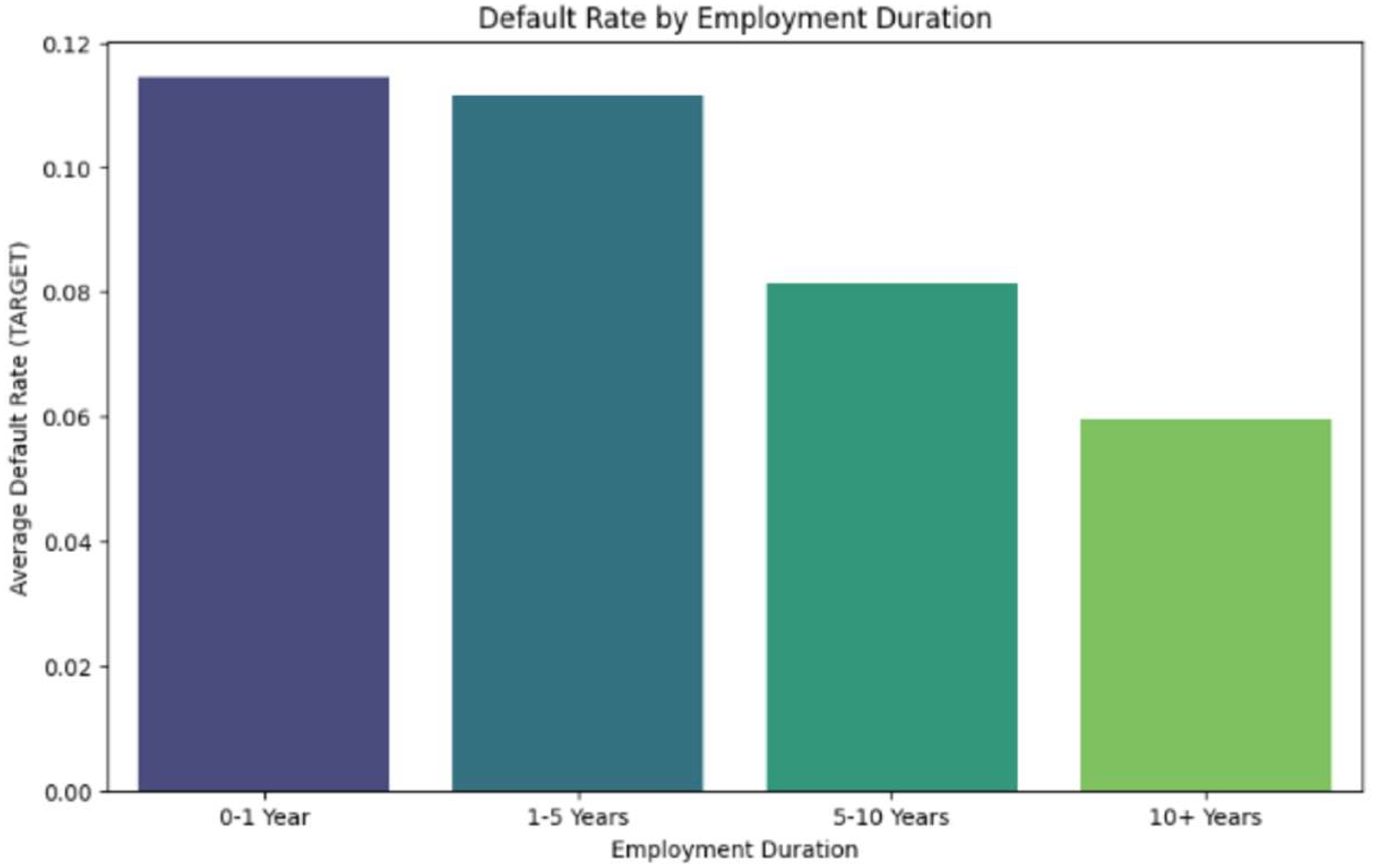
DAYS_EMPLOYED VS TARGET

Applicants with a shorter employment history are more likely to be denied a loan compared to those with a stable job history.

A stable and longer work tenure signals financial reliability and a consistent income source, which increases the likelihood of timely loan repayment.

In contrast, a shorter employment history may raise concerns about the applicant's financial stability and ability to sustain regular payments, making them a higher risk for default.

This makes DAYS_EMPLOYED a crucial factor in assessing loan risk

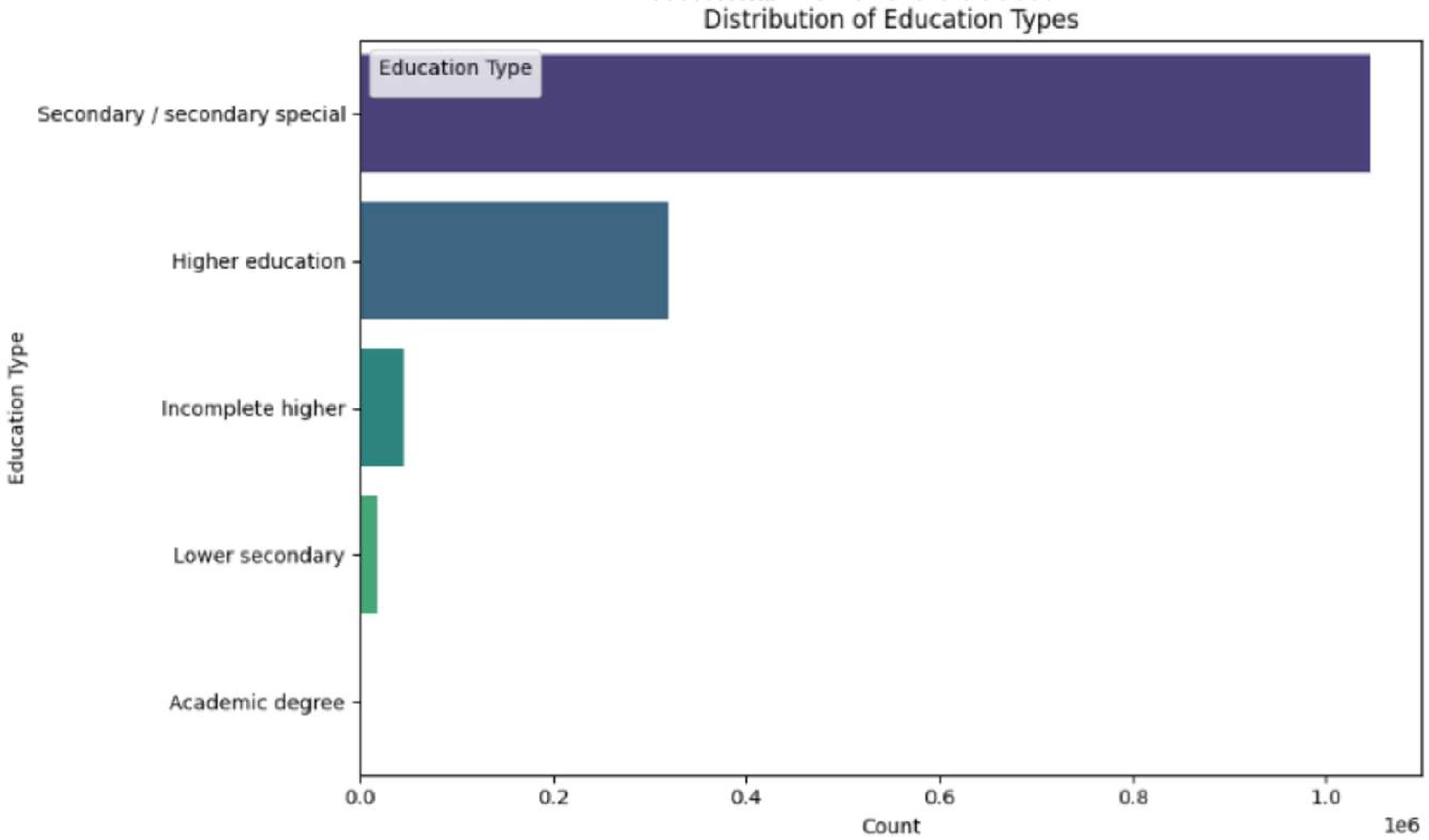


NAME_EDUCATION_TYPE VS TARGET

The NAME_EDUCATION_TYPE feature is important for risk assessment because it shows that while individuals with Secondary / Secondary Special and Higher Education tend to apply more, they also face a higher rejection rate.

In contrast, applicants with Incomplete Higher, Lower Secondary, or Academic Degree may have less applications.

Understanding the relationship between education and loan approval helps financial institutions adjust their risk models and refine criteria for accepting or denying applications based on a combination of education level and other key factors

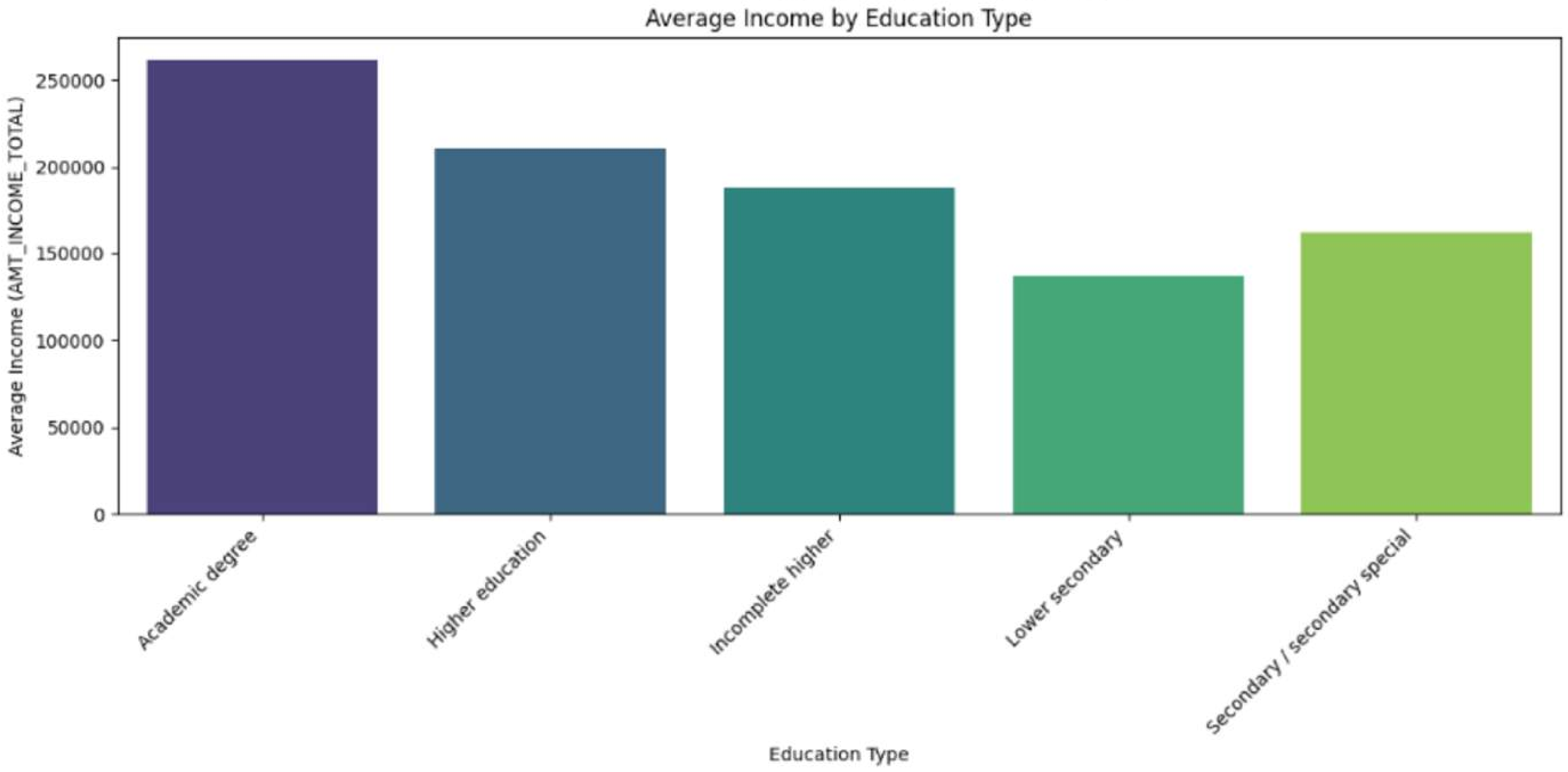


NAME_EDUCATION_TYPE VS AMT_INCOME_TOTAL

Education and Income: There's a clear relationship between education level and income, with those having higher education or academic degrees generally having higher incomes. This could reduce their need for loans.

Loan Approval: Despite higher incomes, people with higher education or secondary special education still tend to have higher loan approval rates, possibly due to their financial aspirations, career growth, or additional financial needs.

Lower Education Groups: Lower secondary individuals tend to have lower income and lower loan application rates, while incomplete higher education individuals still experience challenges with financial stability.



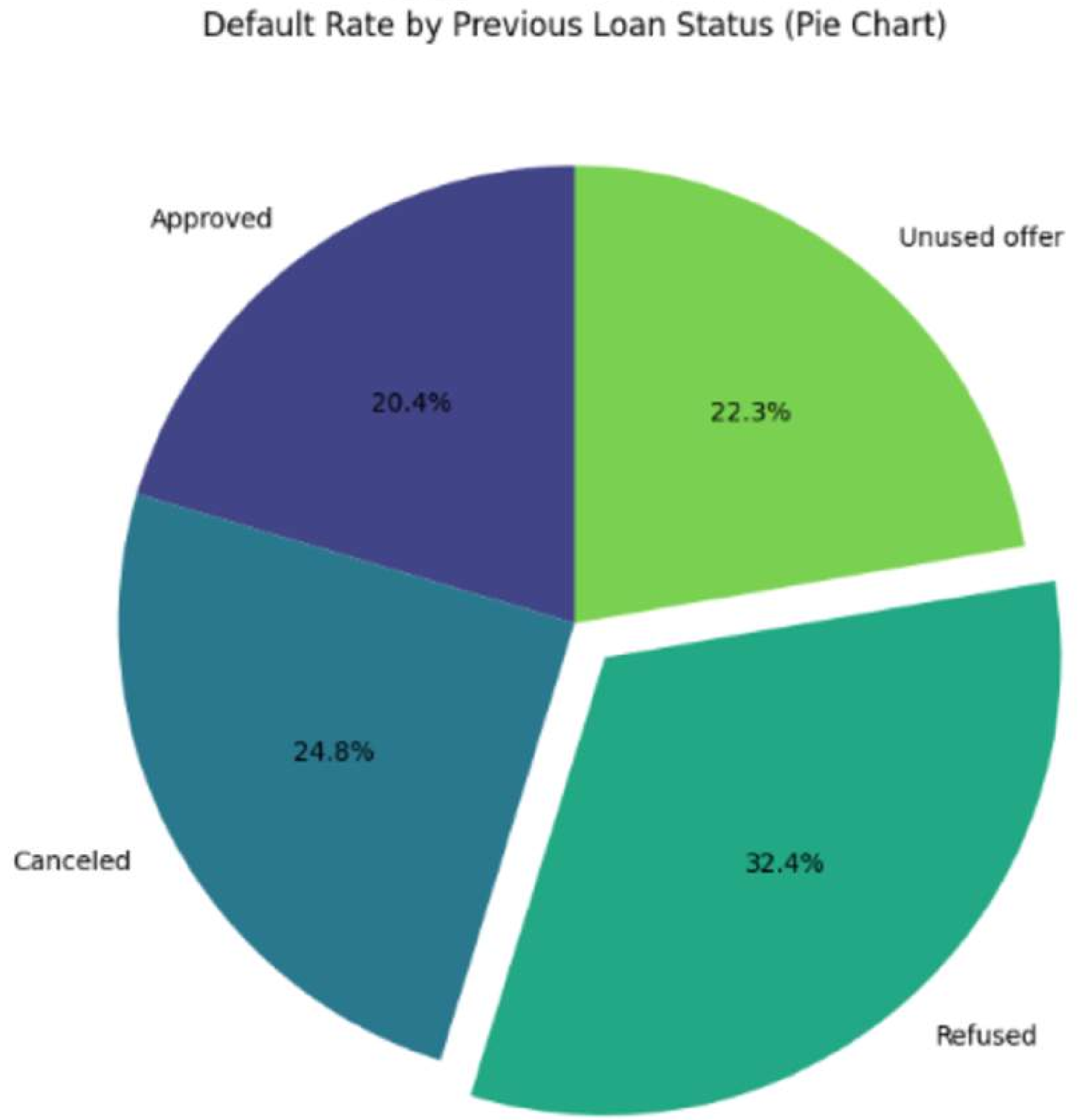
NAME_CONTRACT_STATUS VS TARGET

Refused Applicants as Potential Risk: Applicants with refused applications often display a higher default rate, making them an important group to monitor closely.

The refusal of an application could signal potential underlying issues with their creditworthiness, such as a poor financial history or inability to meet the bank's criteria.

This makes them a high-risk group for future defaults, and thus, they require more attention and thorough evaluation in future credit assessments.

By identifying these individuals early, banks can mitigate risk and implement precautionary measures to manage their exposure.

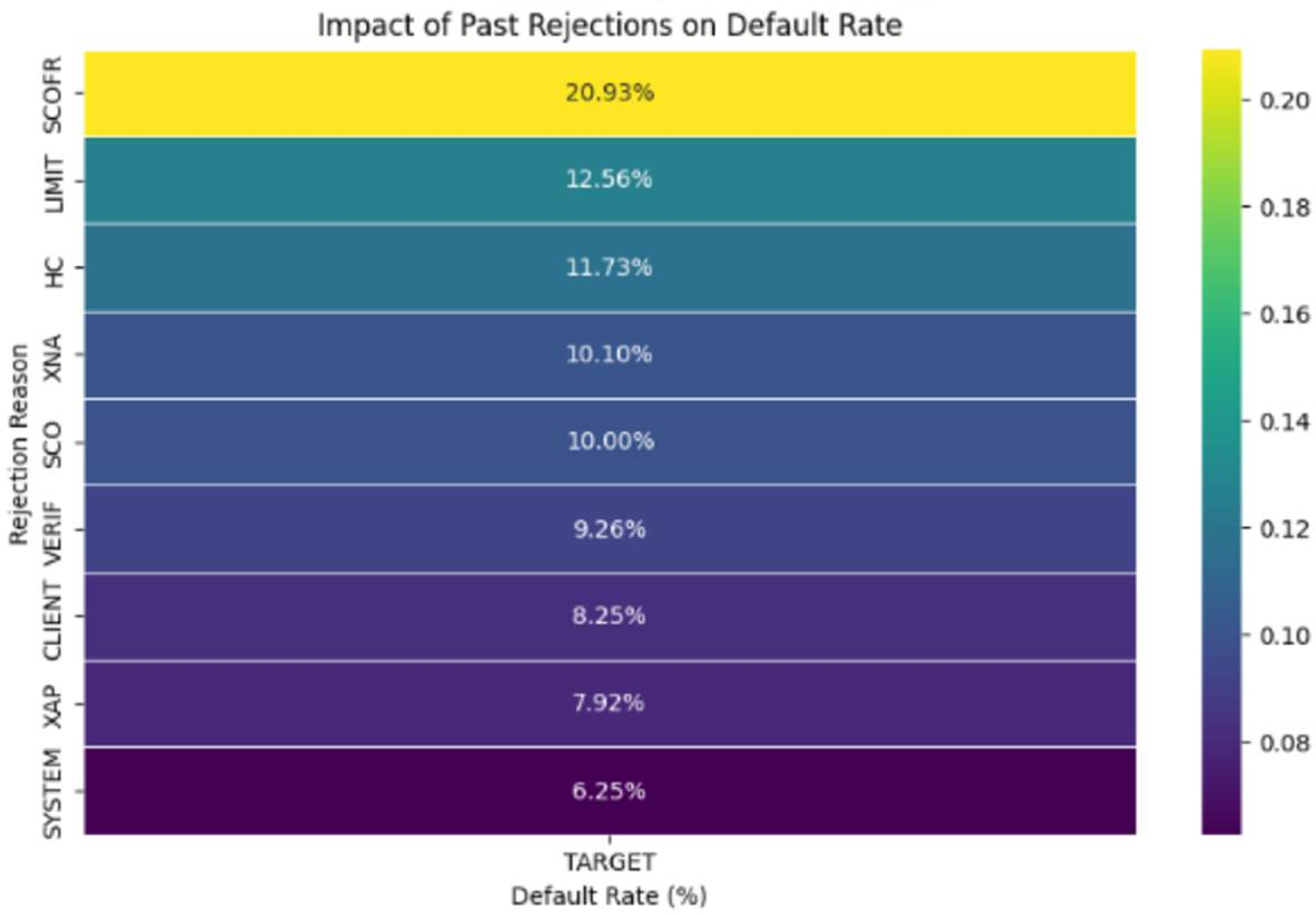


CODE_REJECT_REASON VS TARGET

CODE_REJECT_REASON feature is important for risk assessment because it helps to categorize why applicants were declined.

SCOFR (Score Fraud Risk) – If flagged for fraudulent behavior, applicants with this reason are a high-risk group. Monitoring them is crucial to prevent fraudulent activity.

These rejections in the range of 6.25% to 12.5% still represent important risk factors in the lending process, as they help identify applicants with potentially weak financial standing or issues with their application that could increase the risk of loan defaults. Monitoring these groups closely can help mitigate risk and refine lending practices.



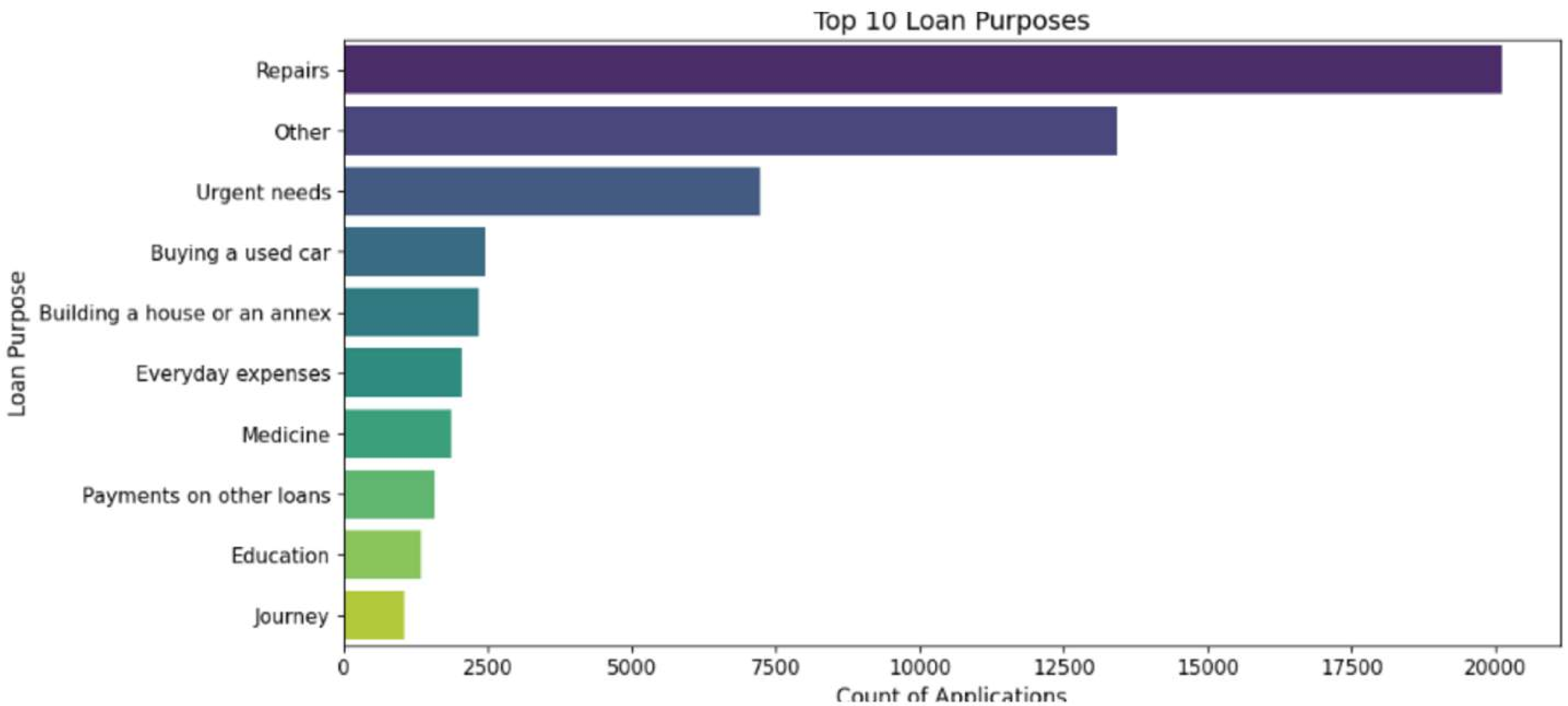
CODE_REJECT_REASON VS TARGET

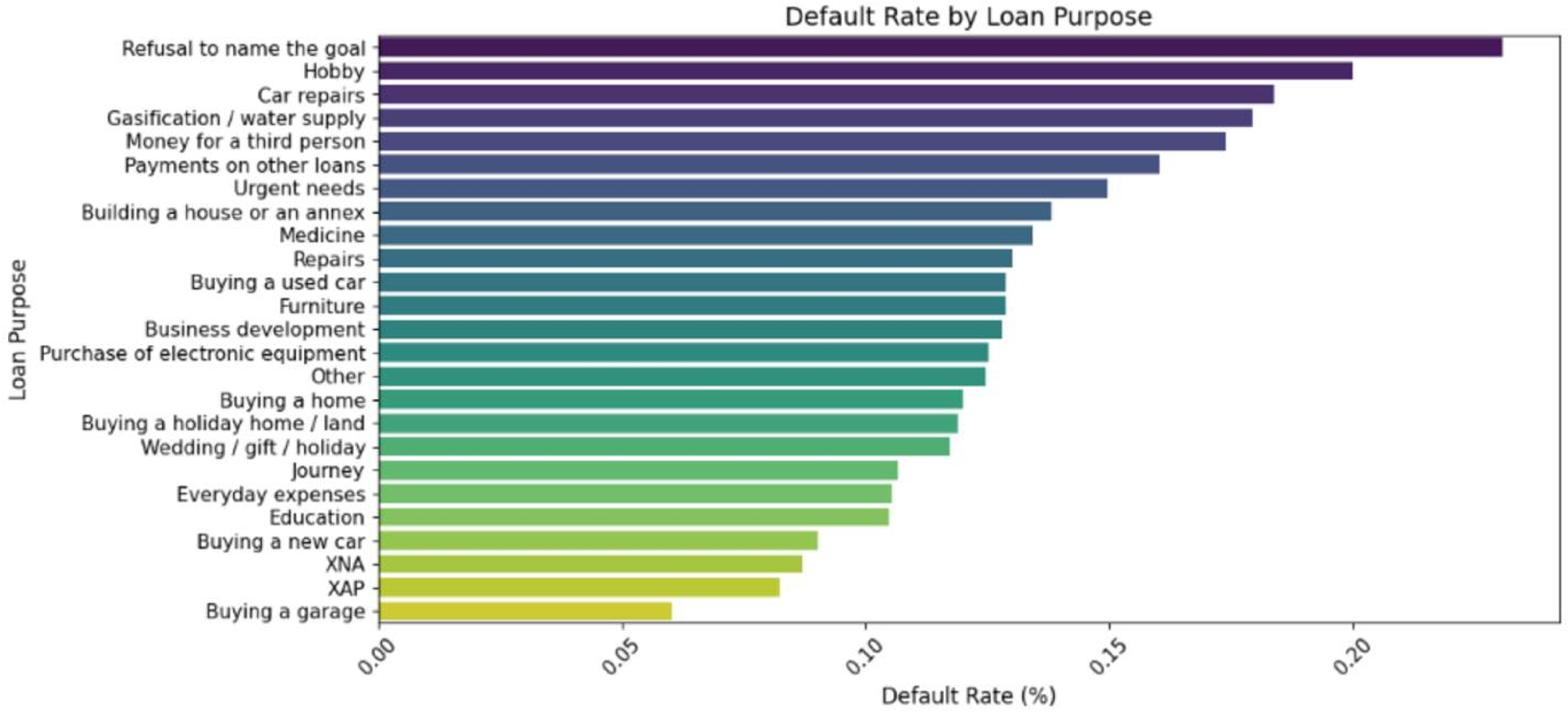
NAME_CASH_LOAN_PURPOSE is crucial for risk assessment because it indicates the reason for which an applicant is requesting the loan.

Applicants who take loans for investment purposes (e.g., business, education) may be viewed as lower risk, as they may have the potential to generate income or value from the loan

Conversely, loans for personal consumption (e.g., no purppose, car repairs, home renovations, buying goods) could suggest that applicants are less likely to generate an immediate return on investment, making them higher risk.

It is provides valuable context for understanding why a loan is being sought and can provide key insights into an applicant’s ability and likelihood to repay.





CONCLUSION

In this analysis, we explored key features that contribute to assessing the risk of loan applicants

- **Income Analysis:** Higher income typically correlates with lower default rates, but income alone isn't a reliable predictor of default risk. Applicants across all income levels (low to very high) can still default, indicating that additional factors must be considered.
- **Age and Employment History:** Younger applicants and those with shorter employment histories present higher credit risks. In contrast, stable, long-term employment increases the likelihood of loan repayment.
- **Family Size**: Larger families tend to have higher default rates, as a growing family can create more financial strain, especially with fewer earners.
- **Education:** Higher education levels are linked to better financial stability and repayment ability. However, we found that applicants with secondary or higher education had both high approval and rejection rates, suggesting that other financial factors are at play.
- **Loan Purpose:** Applicants requesting loans for personal consumption or discretionary spending show higher default rates. Loans for investment purposes tend to indicate more financially stable intentions.
- **Rejection Reasons:** Understanding rejection reasons such as low credit scores or fraud suspicion reveals high-risk applicants who are more likely to default.

LET'S CONNECT

https://www.linkedin.com/in/liudmila-stolbetskaia/

iudmila.stolbetskaia@gamil.com

https://github.com/MilaSoul

LIUDMILA STOLBETSKAIA

THANK YOU

FOR YOUR ATTENTION

This presentation provided valuable insights into credit risk assessment, which I developed during my internship with Oeson Learning.