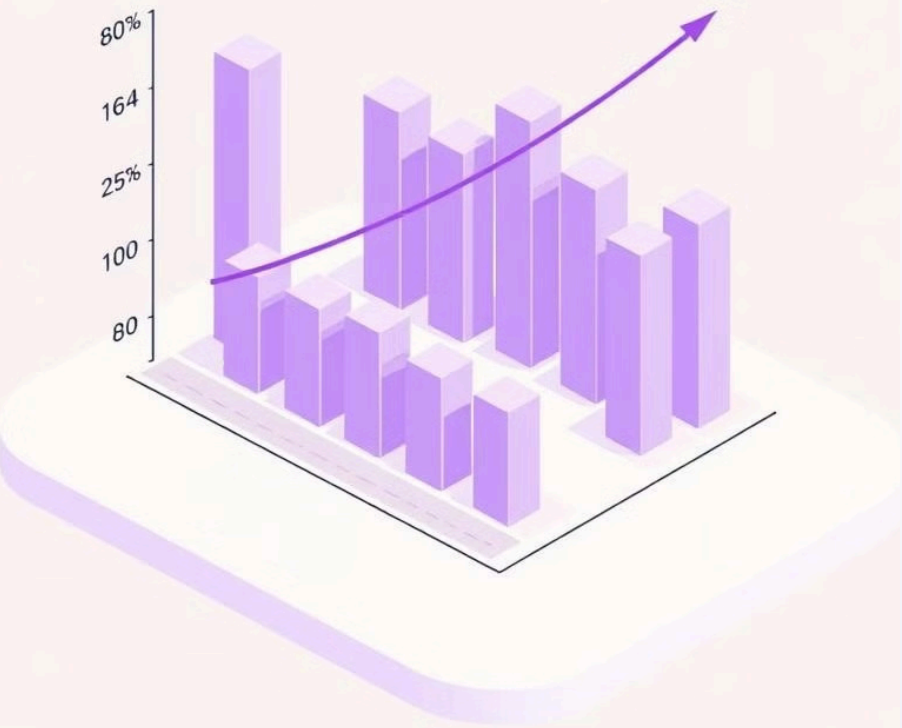


Loan Default Risk Analysis

Data-driven insights into loan default risk and strategies for improved risk management.

EDA done By **Soumak Majumdar**

Business Objective

Goal

To identify patterns in client data that indicate difficulty in repaying loan installments.

Application

Minimize financial losses due to loan defaults by identifying and mitigating risks associated with loan applications.

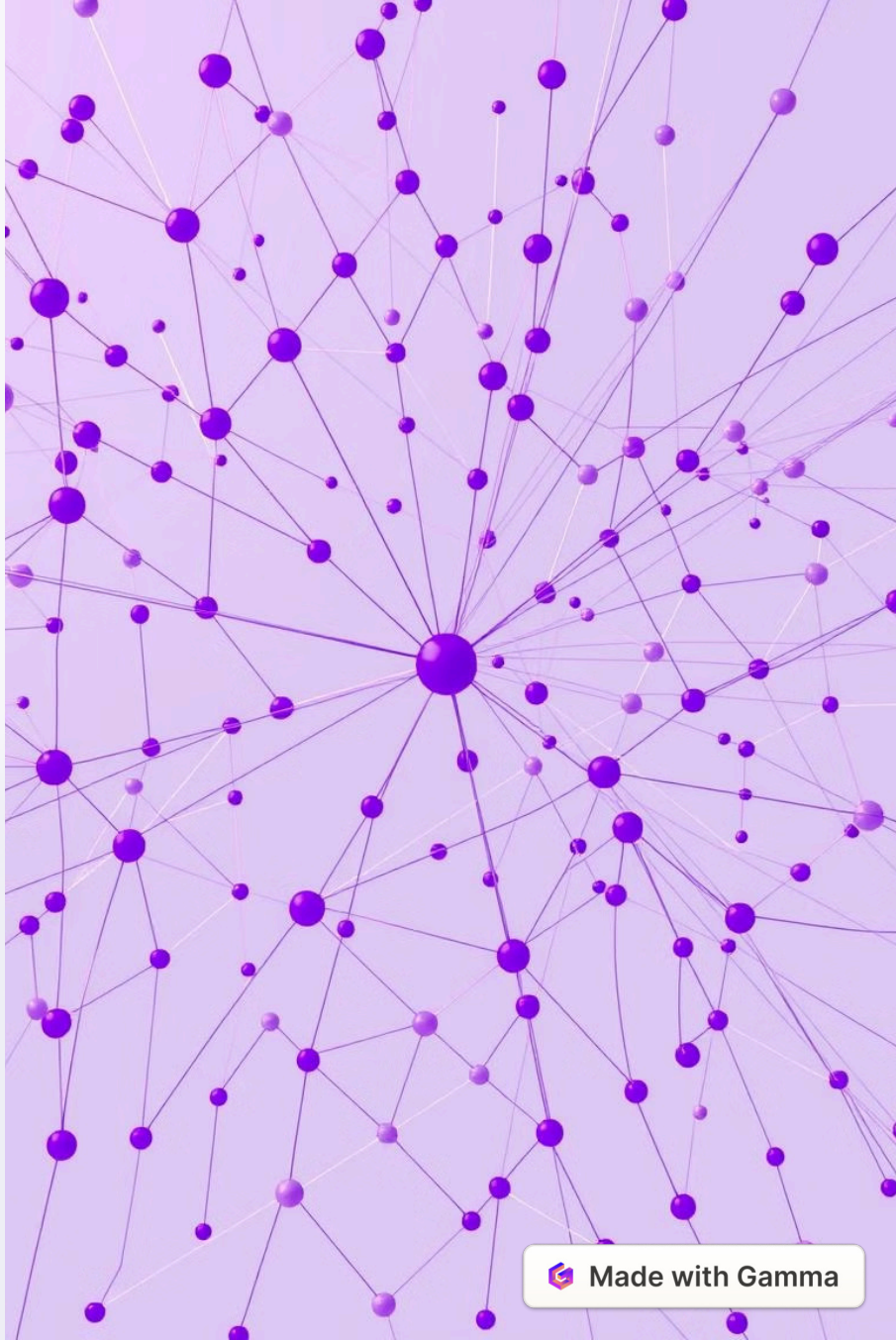
Data Overview

Application Data

Contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

Previous Application Data

Contains information about the client’s previous loan data. It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.

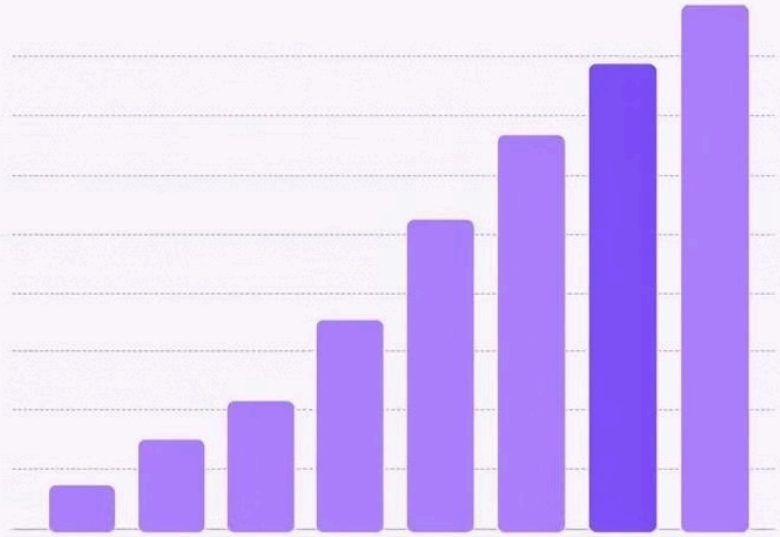


Univariate Analysis

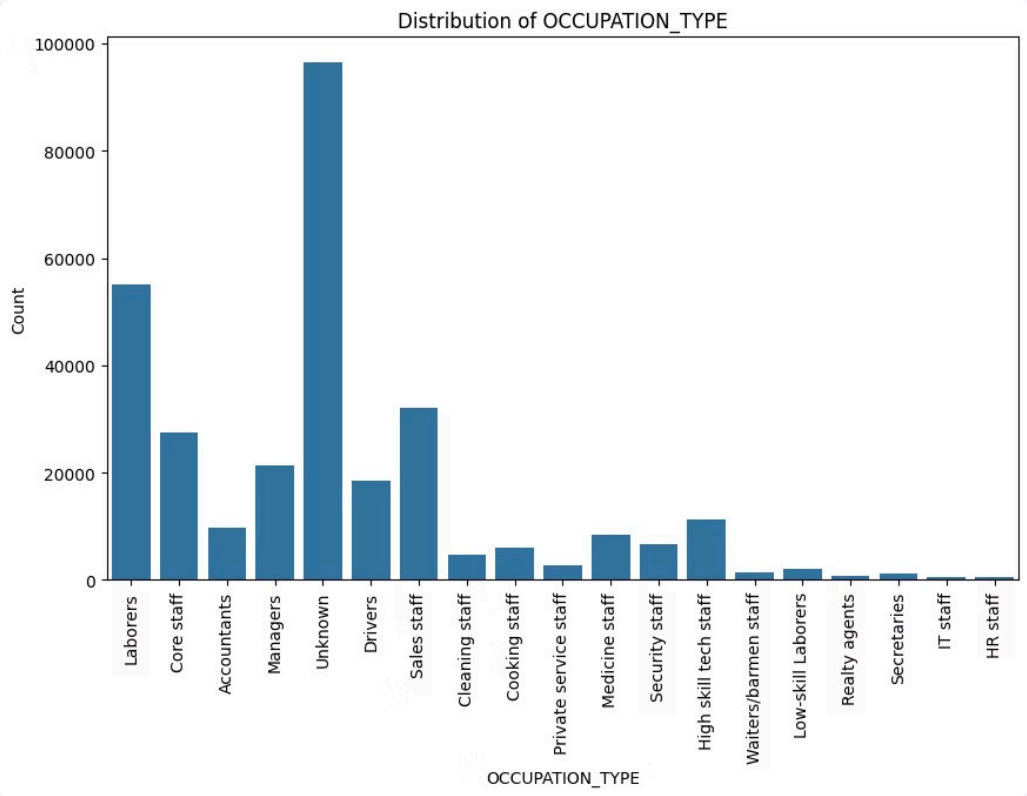
Key Insights

We can infer that the bank should prioritize risk management strategies for segments with potentially higher default risk, such as:

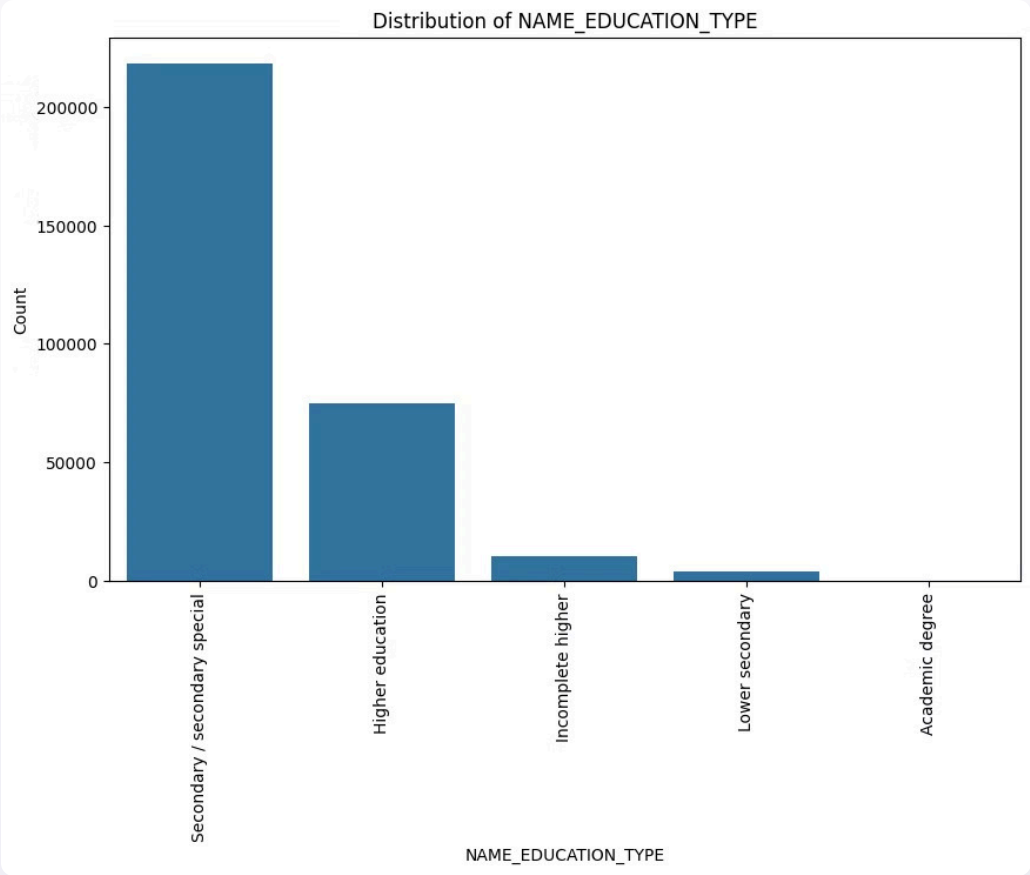
- Loans with high credit amounts and annuity payments.
- Loans for high-priced goods.
- Loans to clients with low income, shorter employment history, lower external scores, and lower education levels.
- Loans to clients who are unemployed, on maternity leave, or have unstable occupations.
- Loans to specific demographics (e.g., some family statuses, genders).



Univariate Analysis [Contd.]

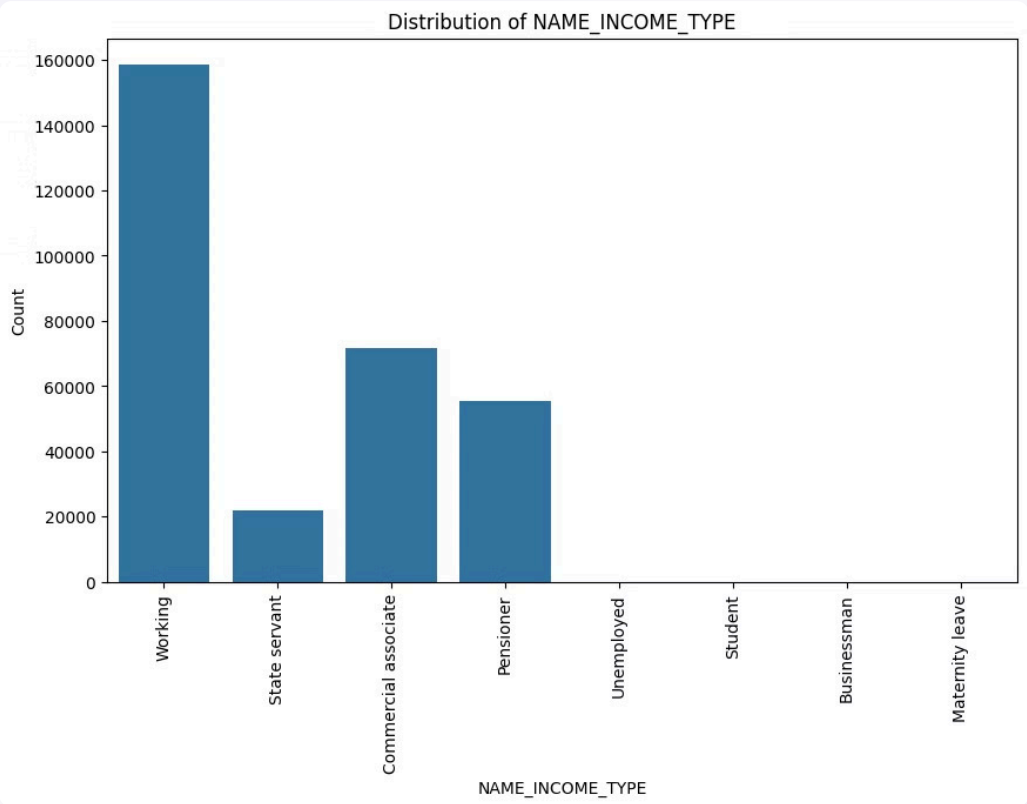


Distribution Of OCCUPATION_TYPE

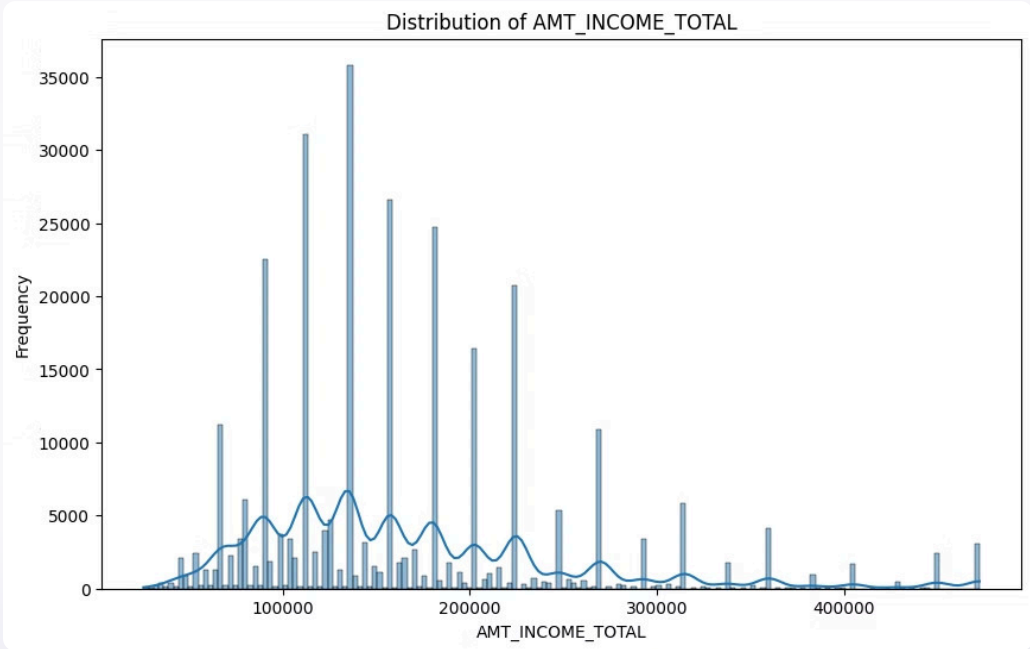


Distribution Of NAME_EDUCATION_TYPE

Univariate Analysis [Contd.]

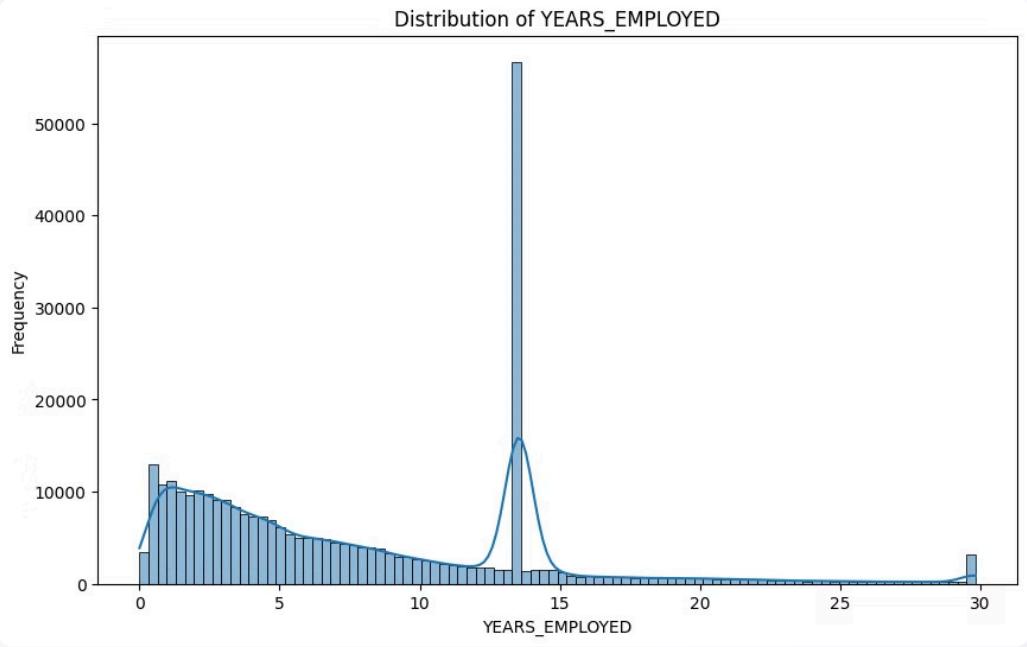


Distribution Of NAME_INCOME_TYPE

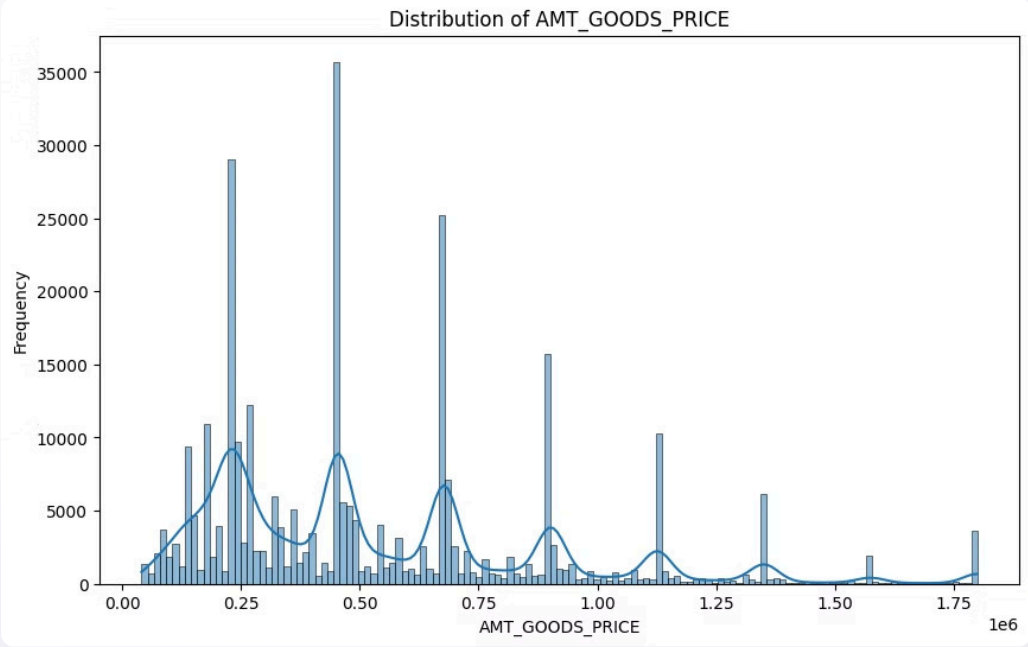


Distribution Of AMT_INCOME_TOTAL

Univariate Analysis [Contd.]

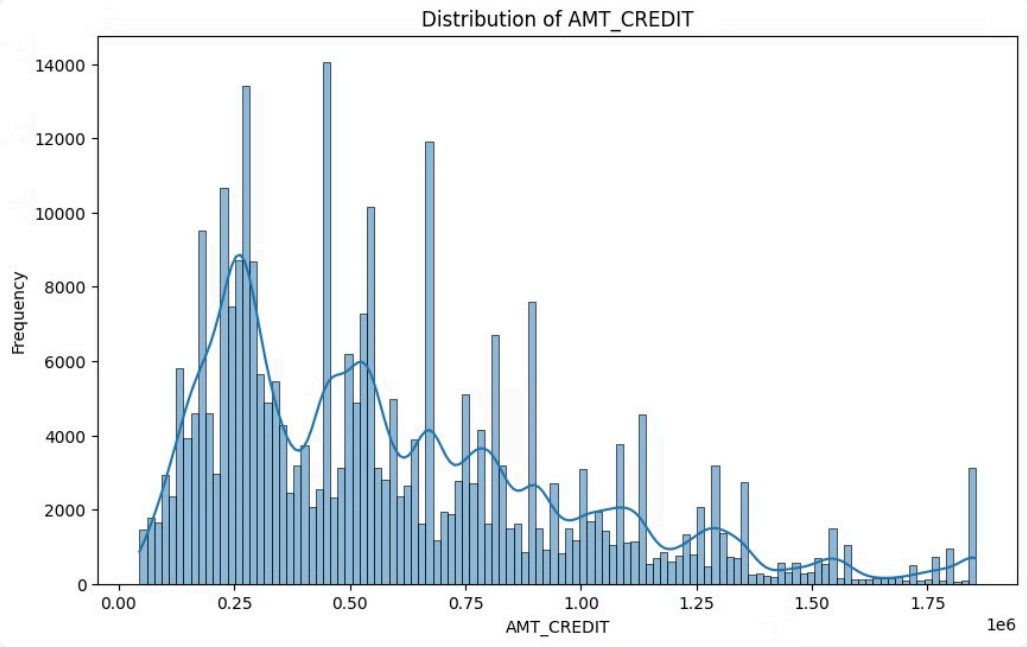


Distribution Of YEARS_EMPLOYED

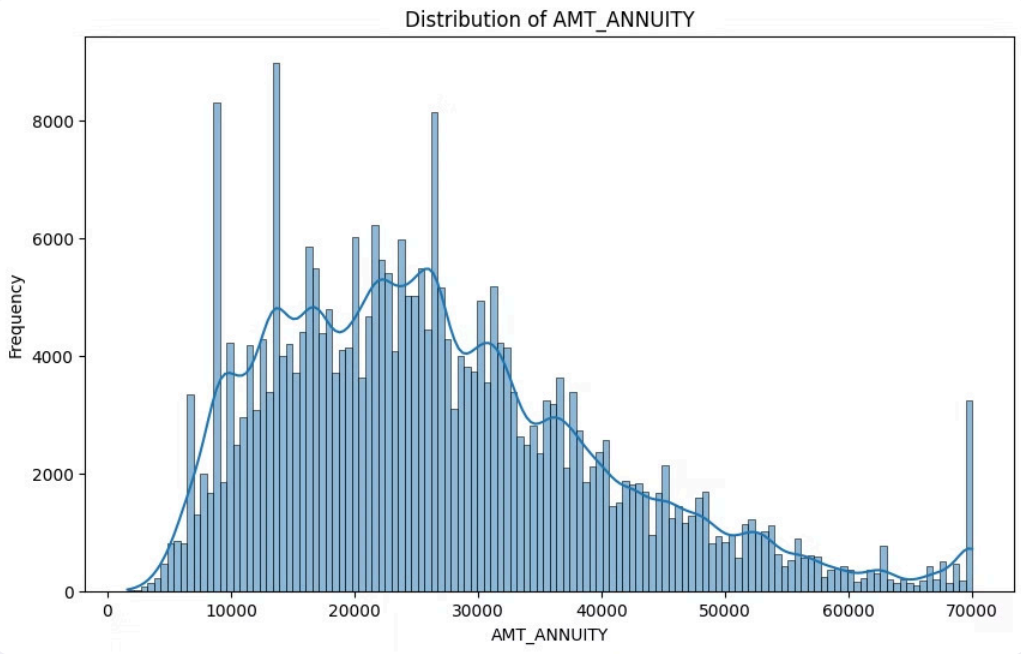


Distribution Of AMT_GOODS_PRICE

Univariate Analysis [Contd.]

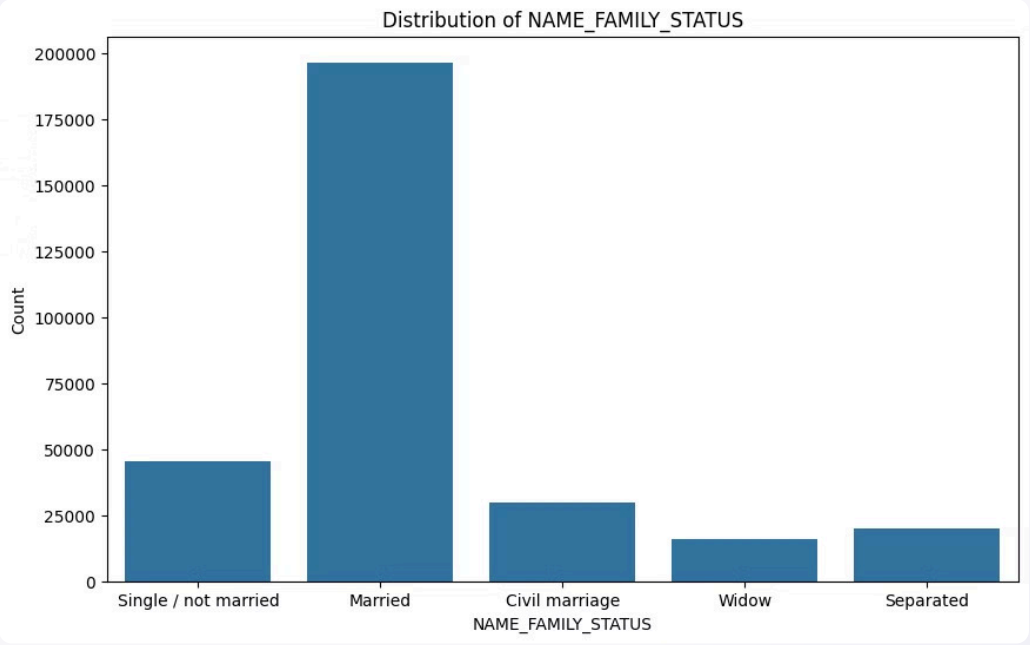


Distribution Of AMT_CREDIT

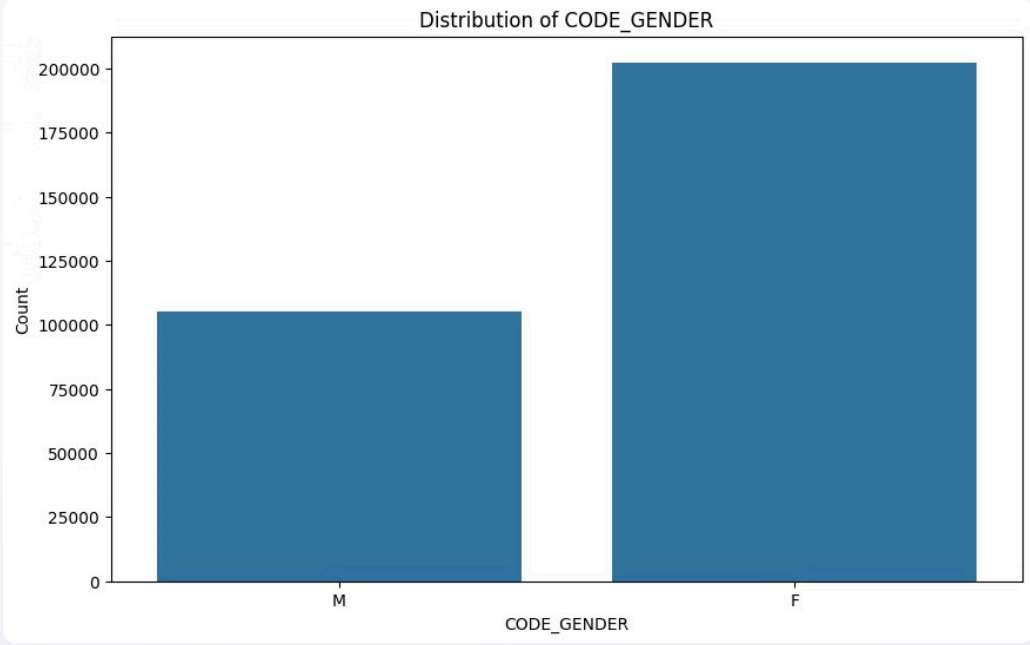


Distribution Of AMT_ANNUITY

Univariate Analysis [Contd.]



Distribution Of NAME_FAMILY_STATUS



Distribution Of CODE_GENDER

Bivariate Analysis

Strong Predictors

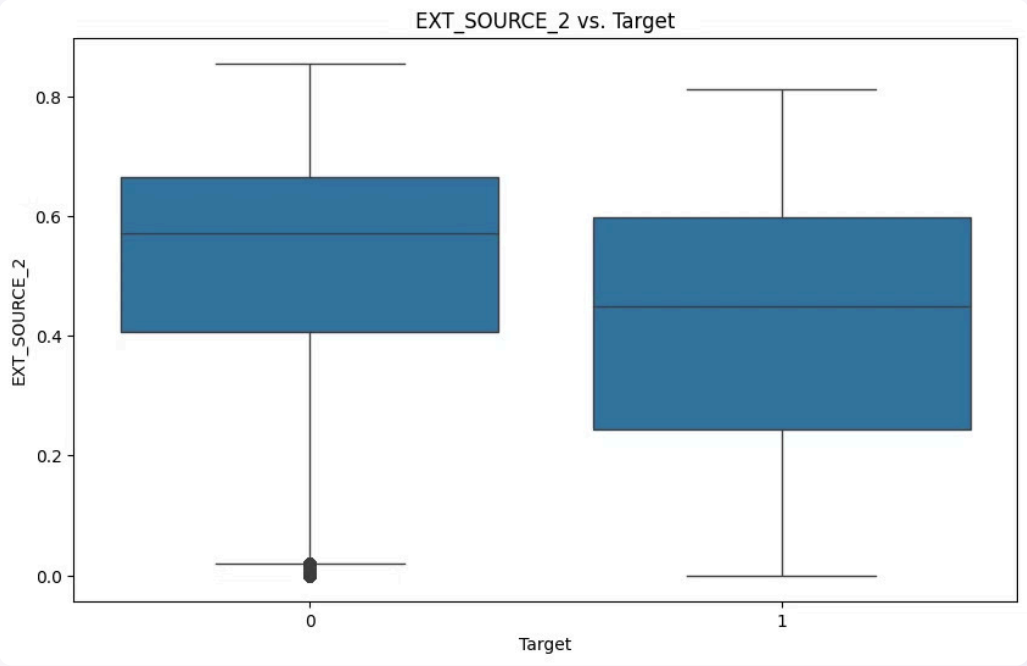
- EXT_SOURCE_2 and EXT_SOURCE_3 appear to be strong predictors of default risk.

Risk Factors

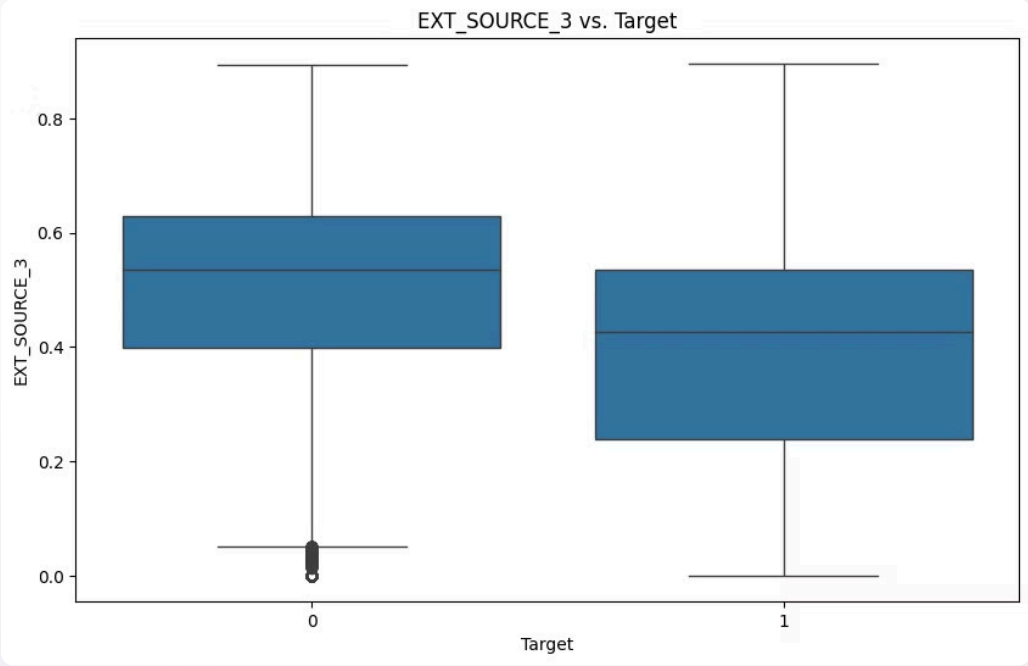
- Larger loan amounts (AMT_CREDIT_app) might be associated with increased risk.
- Clients with Cash loans, Refused or Canceled previous applications, and lower education levels might have higher default rates.
- Clients who are Working, Laborers, or who do not own a car or a house might have higher default risk.

Bivariate Analysis [Contd.]

Strong Predictors



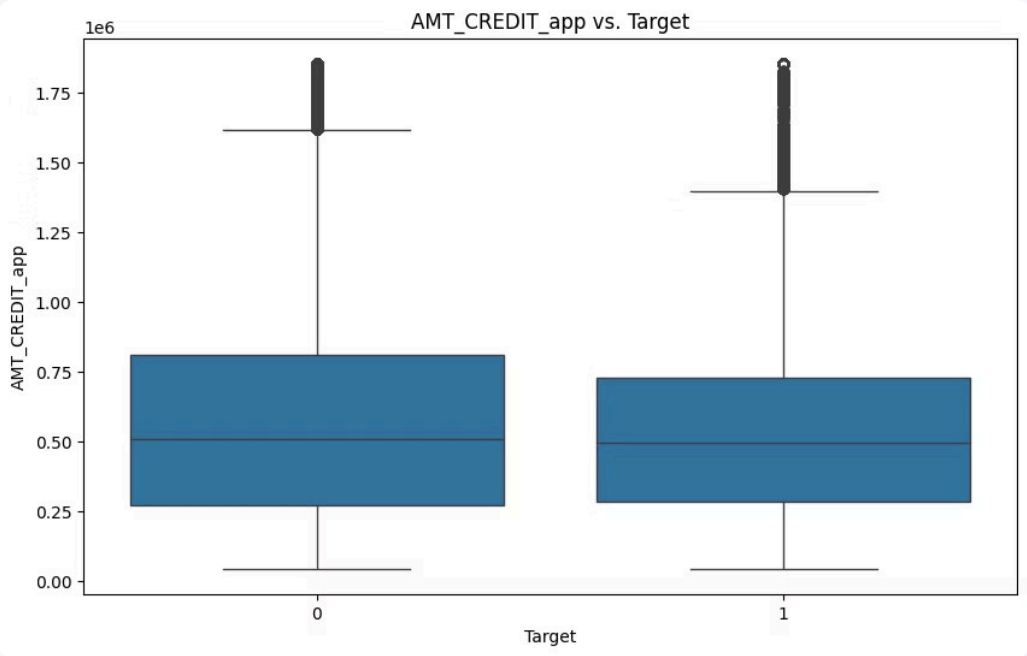
EXT_SOURCE_2 vs. Target



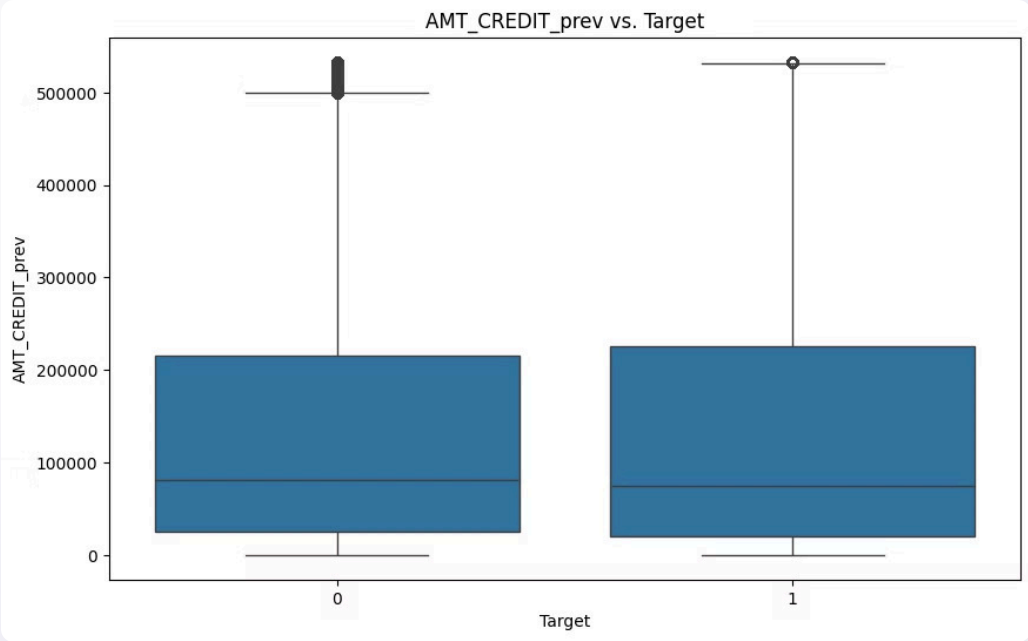
EXT_SOURCE_3 vs. Target

Bivariate Analysis [Contd.]

Risk Factors



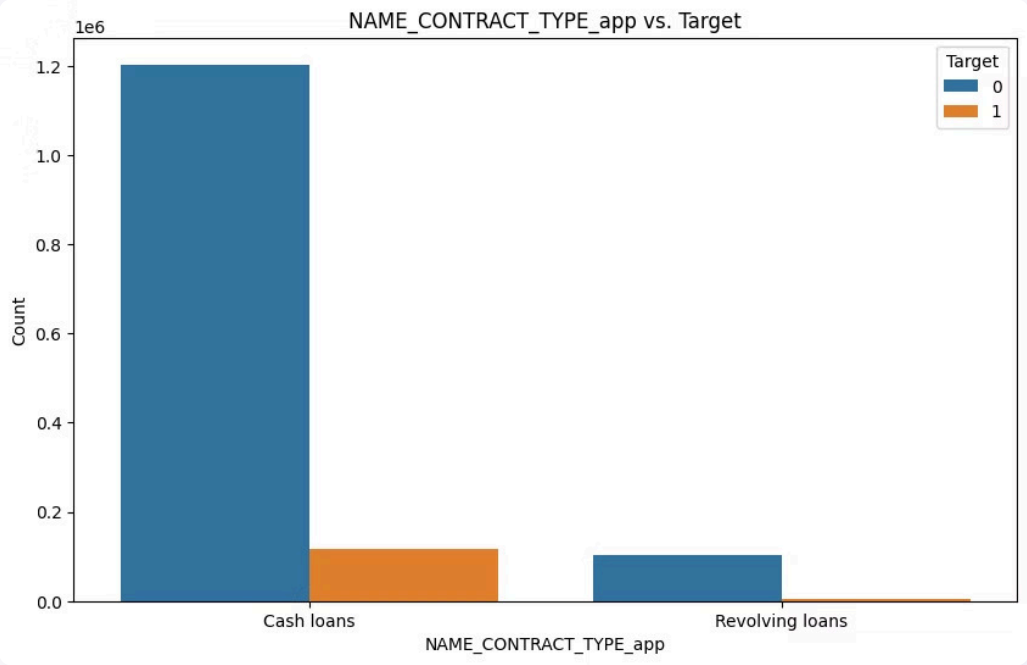
AMT_CREDIT_app vs. Target



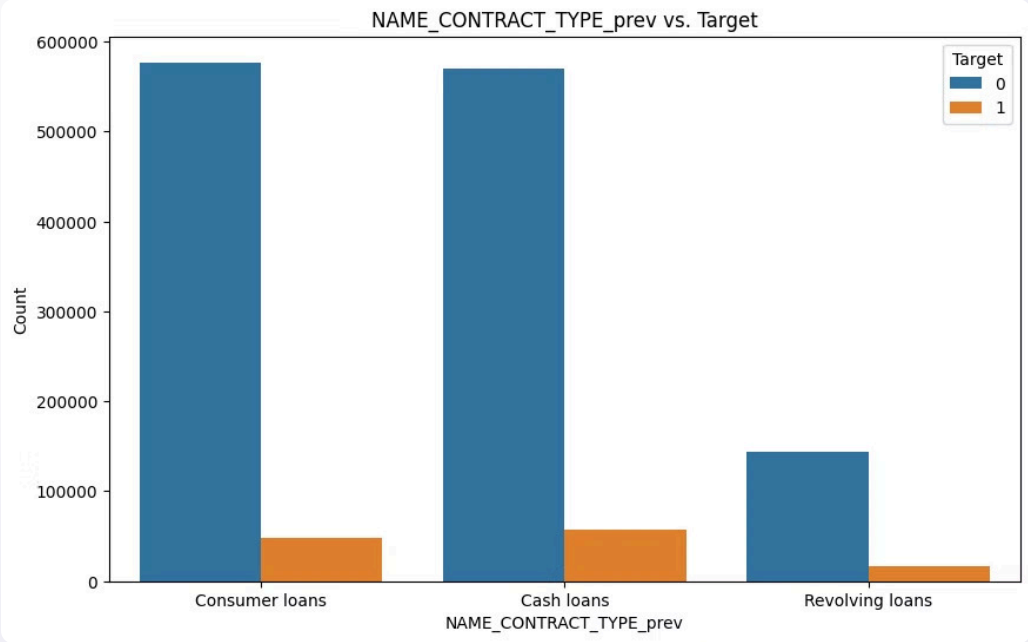
AMT_CREDIT_prev vs. Target

Bivariate Analysis [Contd.]

Risk Factors [Contd.]



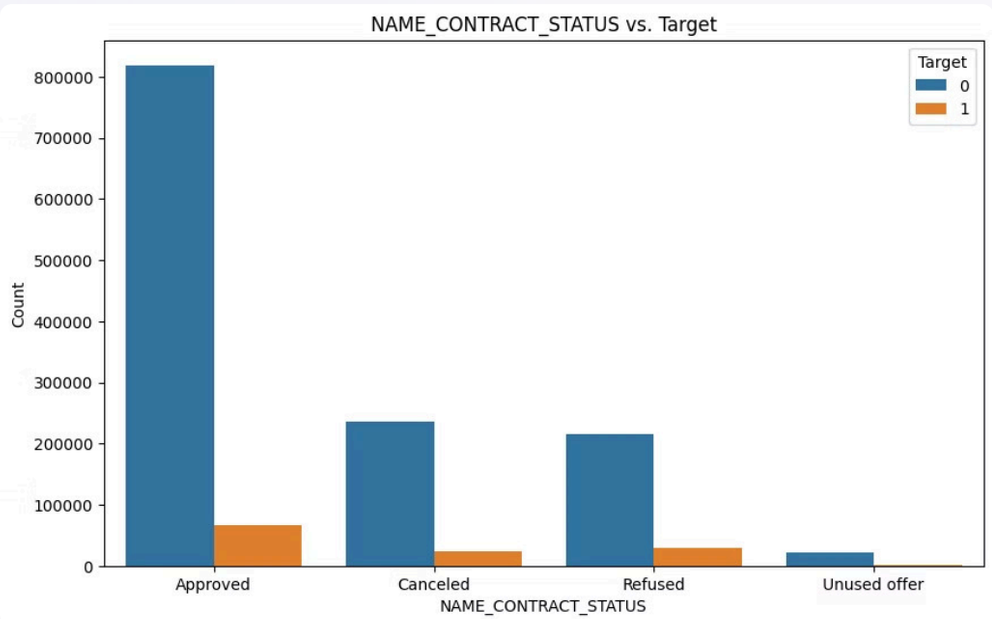
NAME_CONTRACT_TYPE_app vs. Target



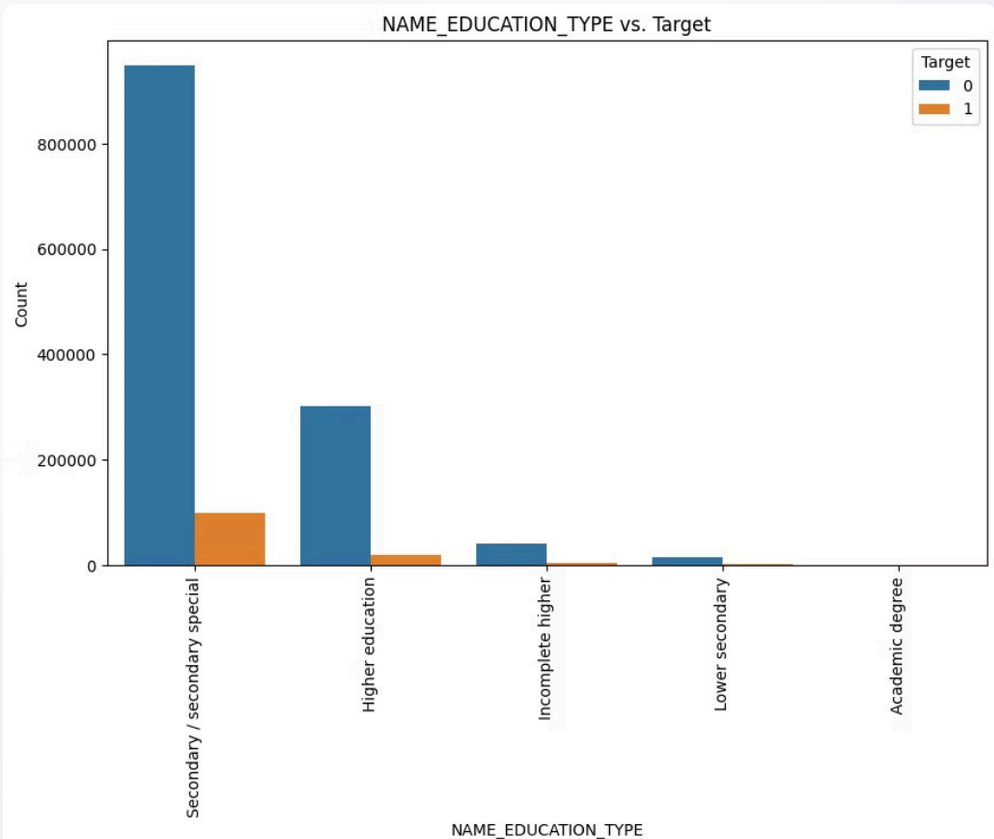
NAME_CONTRACT_TYPE_prev vs. Target

Bivariate Analysis [Contd.]

Risk Factors [Contd.]



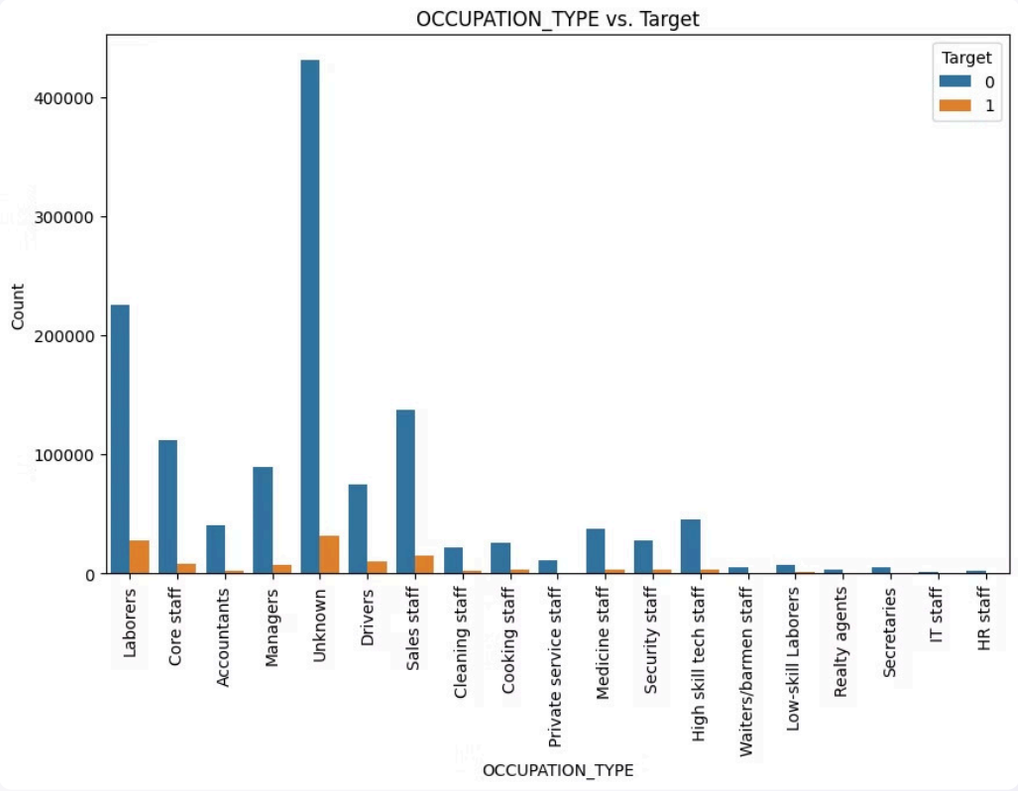
NAME_CONTRACT_STATUS vs. Target



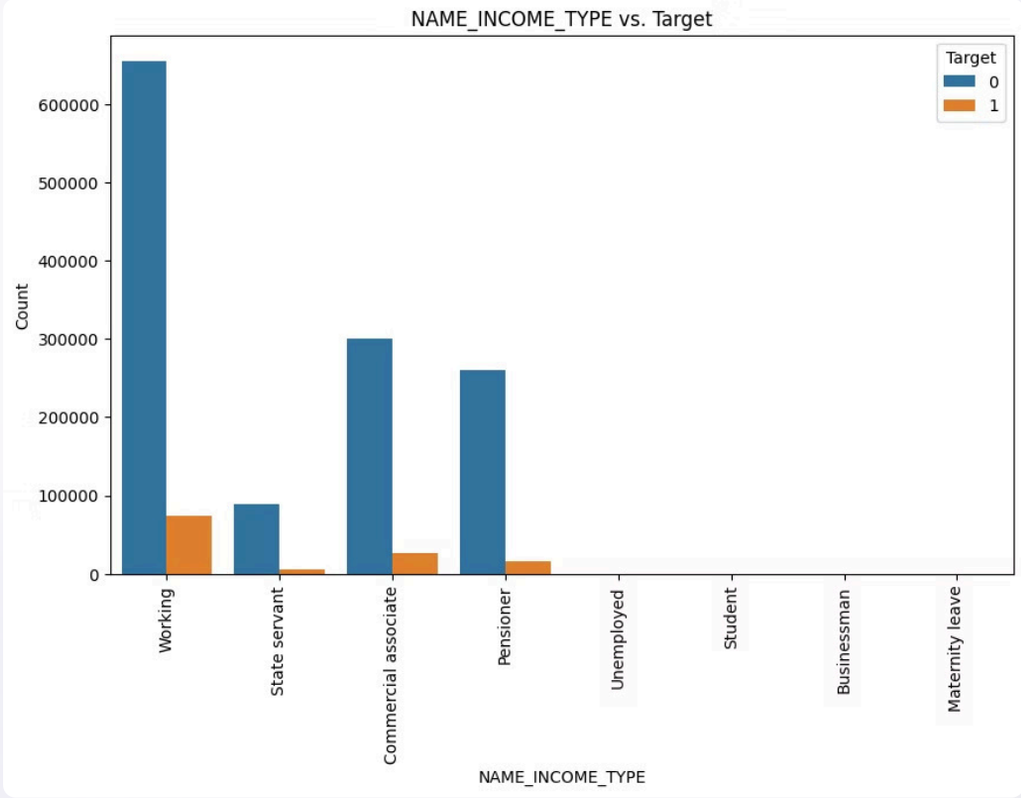
NAME_EDUCATION_TYPE vs. Target

Bivariate Analysis [Contd.]

Risk Factors [Contd.]



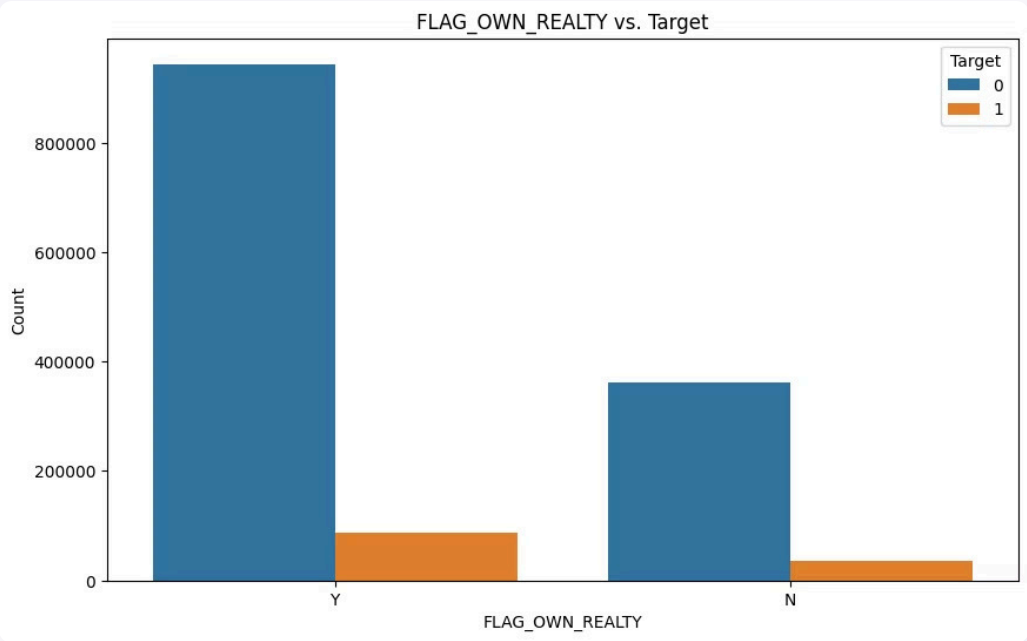
OCCUPATION_TYPE vs. Target



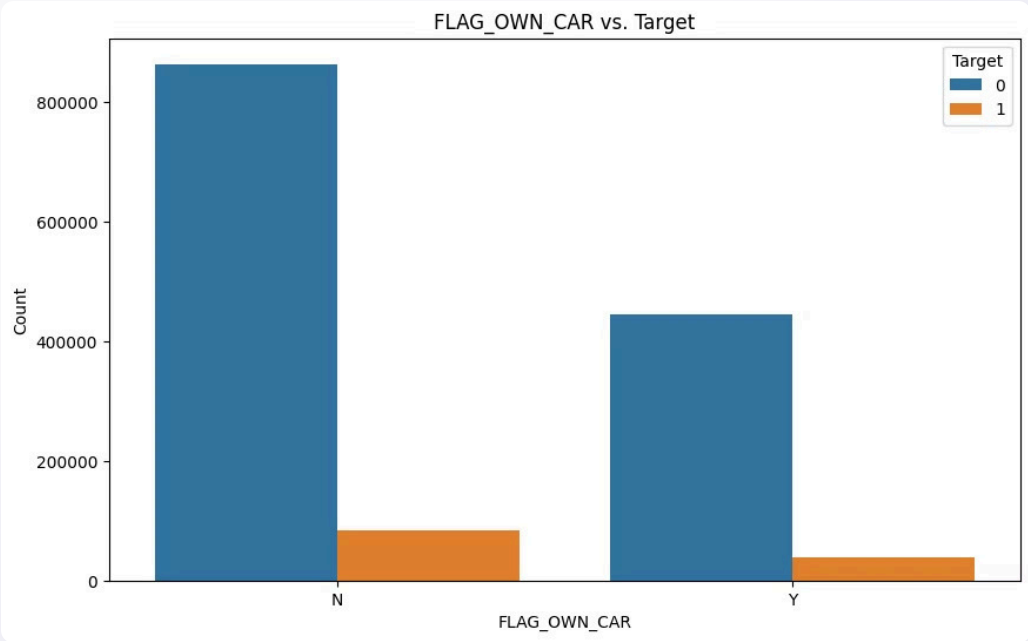
NAME_INCOME_TYPE vs. Target

Bivariate Analysis [Contd.]

Risk Factors [Contd.]



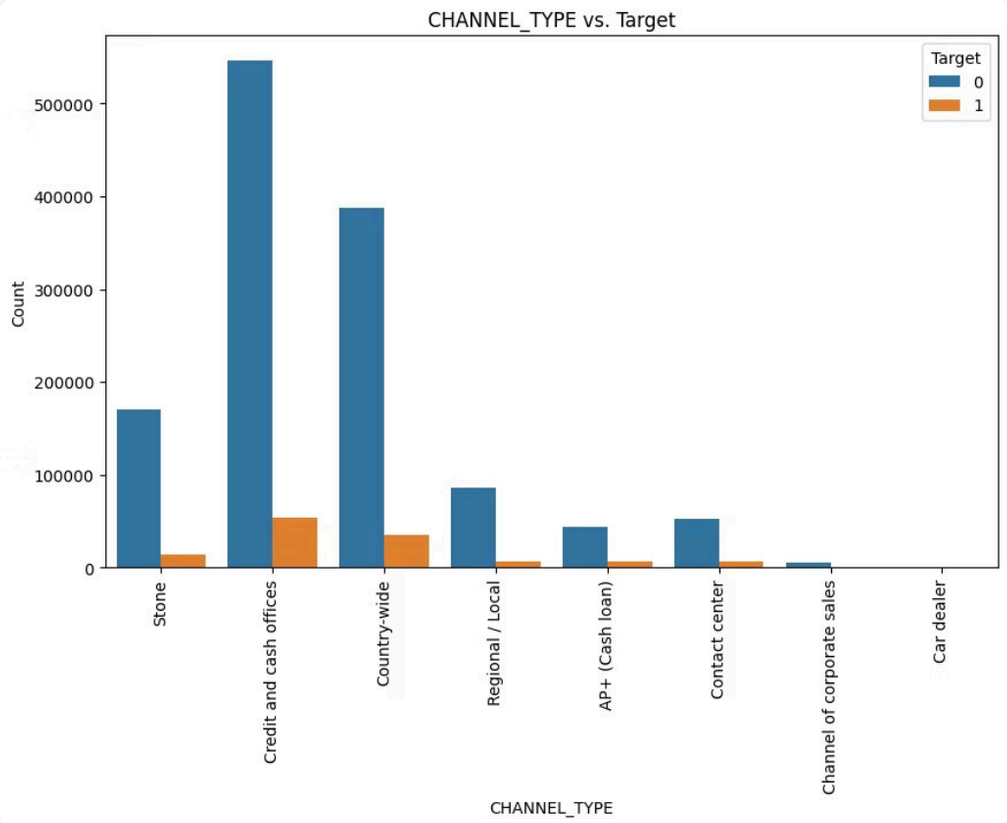
FLAG_OWN_REALITY vs. Target



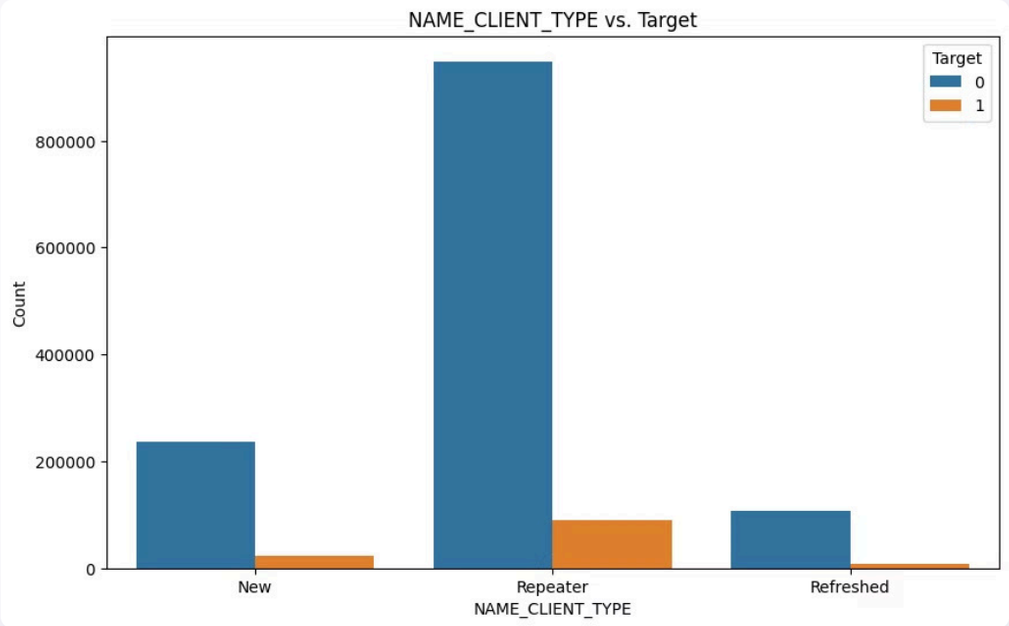
FLAG_OWN_CAR vs. Target

Bivariate Analysis [Contd.]

Risk Factors [Contd.]



CHANNEL_TYPE vs. Target



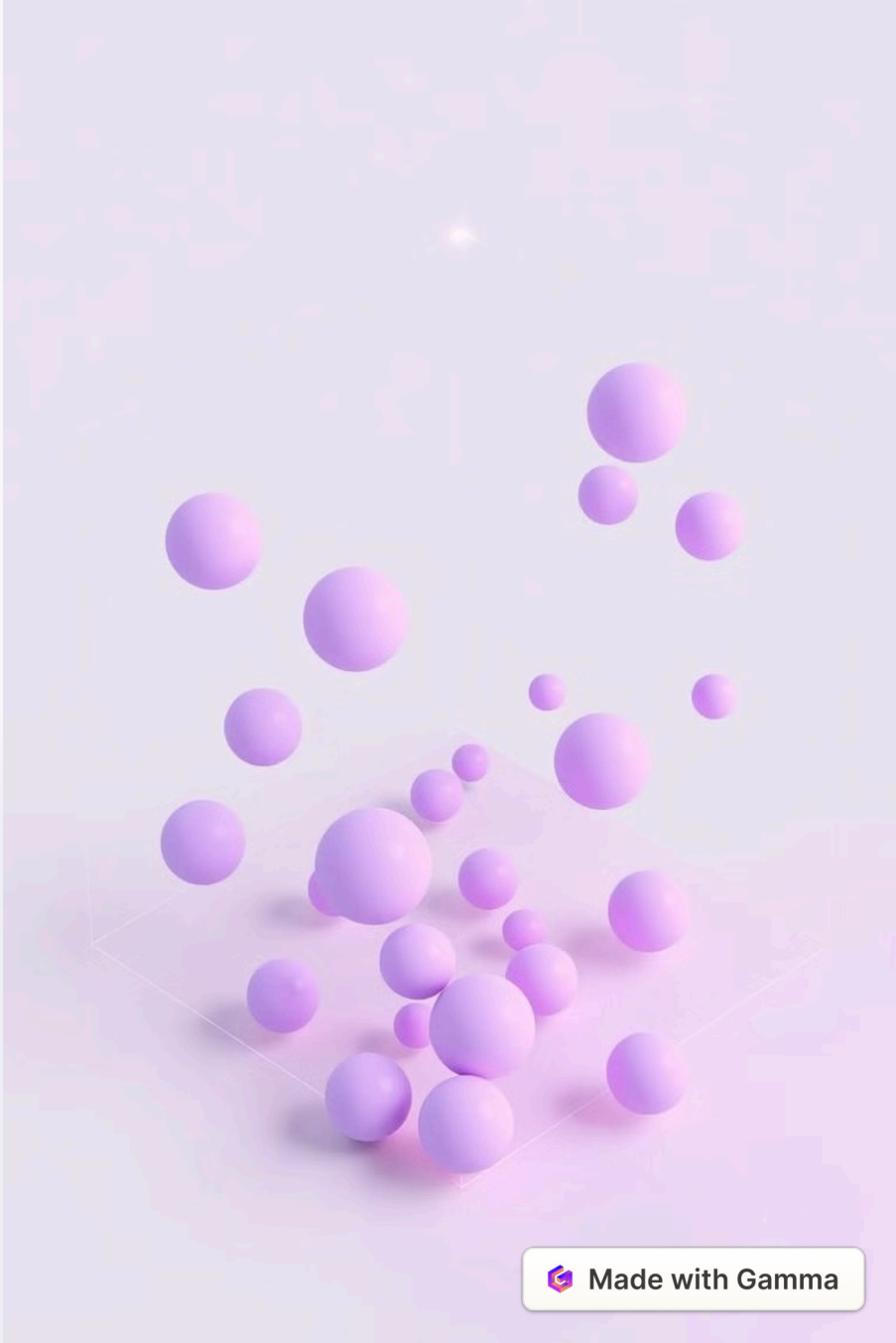
NAME_CLIENT_TYPE vs. Target

Multivariate Analysis

Key Insights

From our multivariate analysis, we observed the following patterns:

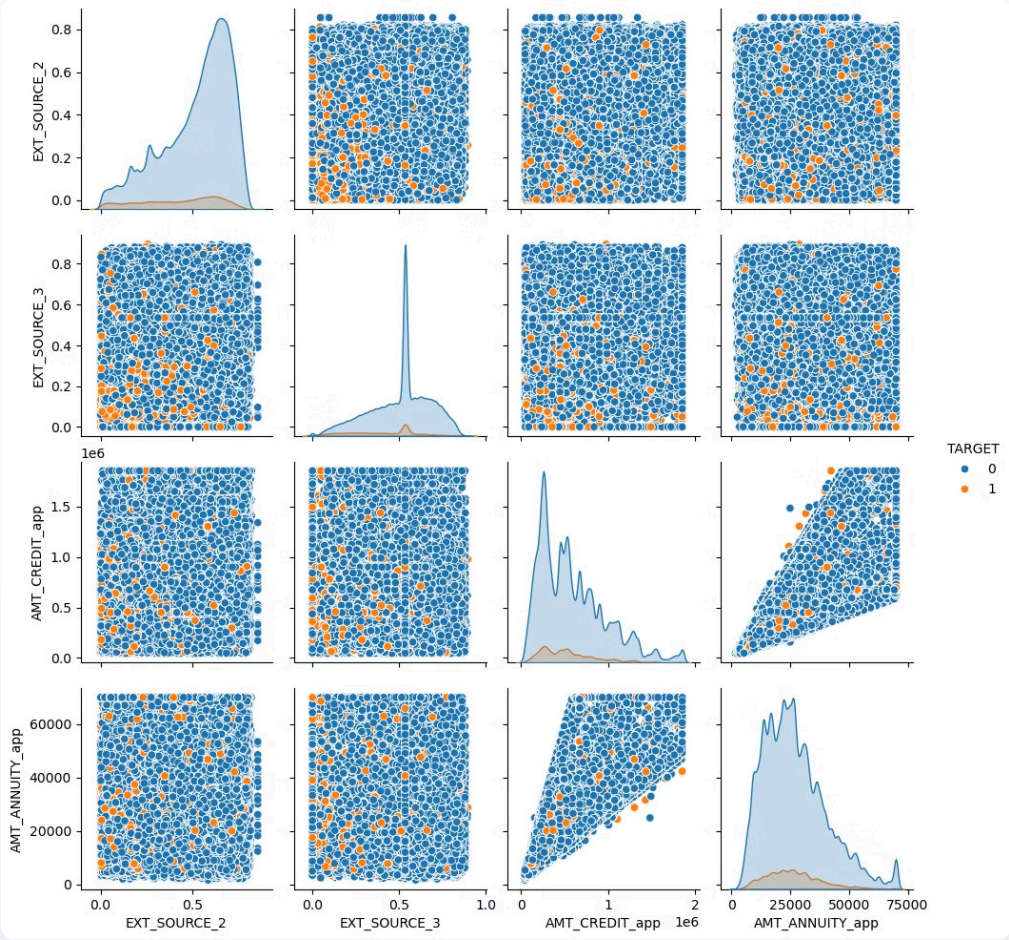
- `EXT_SOURCE_2` and `EXT_SOURCE_3` are crucial for risk assessment. The bank should prioritize understanding and leveraging these external scores in their risk management processes.
- **Loan Characteristics are important risk factors.** The bank should carefully analyze `loan amounts` and `annuity payments` in relation to `client income` and other relevant factors.



Multivariate Analysis [Contd.]

Pair Plot

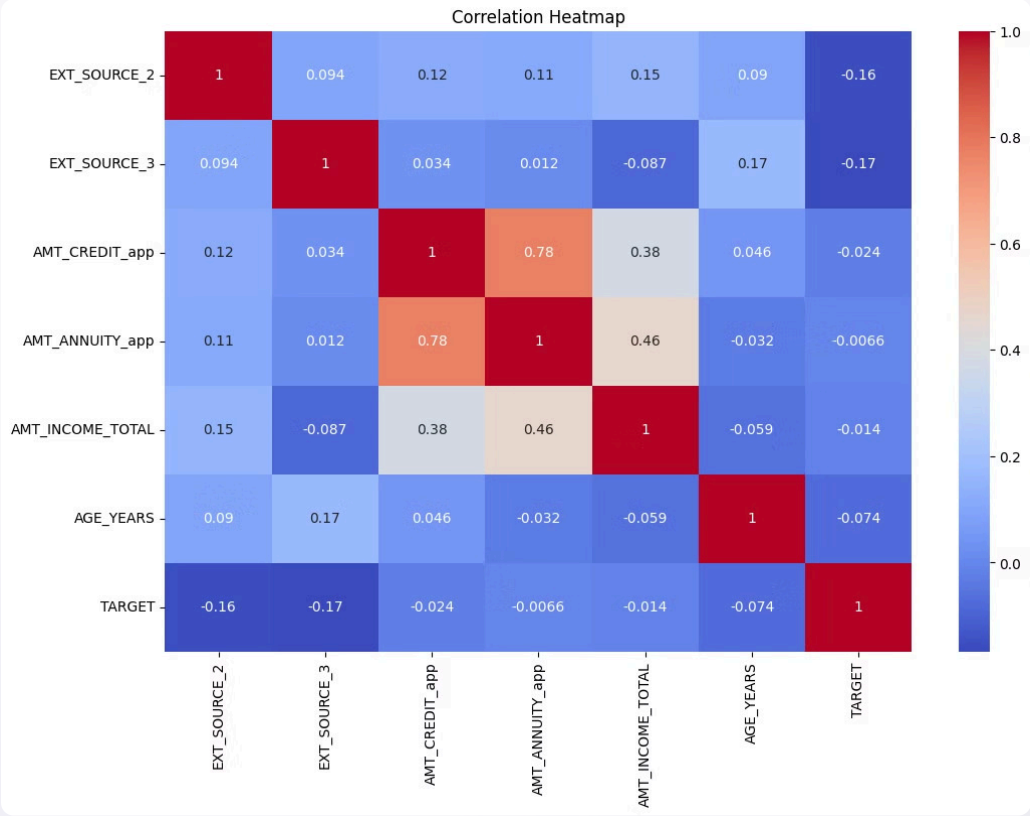
This pairplot confirms that lower values for EXT_SOURCE_2 and EXT_SOURCE_3 are strongly associated with an increased likelihood of default. Additionally, the plots suggest that clients with higher loan amounts (AMT_CREDIT_app) and annuity payments (AMT_ANNUITY_app) might also have a higher risk of default.



Multivariate Analysis [Contd.]

Correlation Heatmap

This heatmap confirms that EXT_SOURCE_2 and EXT_SOURCE_3 are strongly negatively correlated with the target variable, indicating they are significant predictors of default risk. Moderate correlations were observed between loan amount, annuity payment, and income, suggesting that larger loans and higher payments might be associated with increased risk.



Key Takeaways

1

Critical Predictors

EXT_SOURCE_2 and EXT_SOURCE_3 are crucial risk indicators.

2

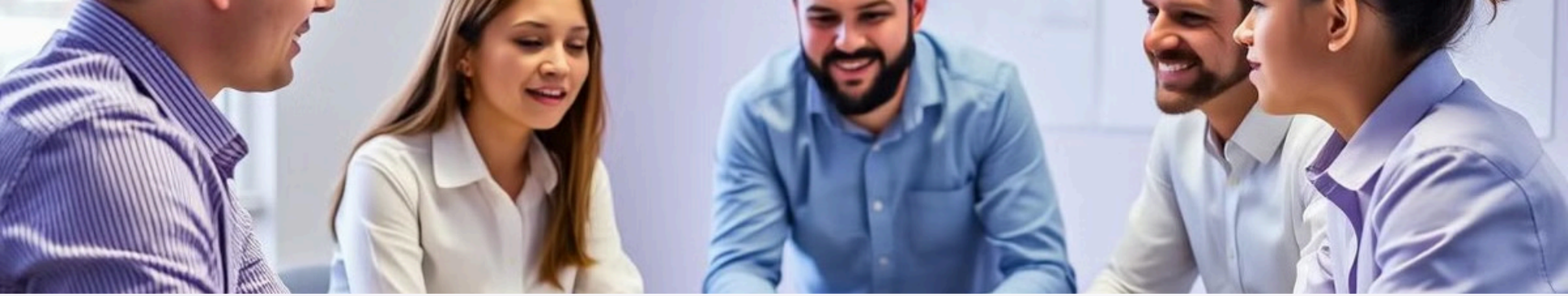
Significant Factors

Loan amount, annuity payment, and income influence default risk.

3

Other Influences

Regional factors, previous applications, and client demographics play a role.



Recommendations

1

Prioritize Risk Assessment

Implement robust scoring systems that incorporate EXT_SOURCE_2 and EXT_SOURCE_3 as key risk indicators.

2

Refine Loan Approval Criteria

Adjust loan approval criteria based on risk factors identified (e.g., loan amount, annuity payment, client characteristics).

3

Targeted Risk Management

Develop targeted risk management strategies for high-risk segments (e.g., clients with lower external scores, higher loan amounts, certain income types).

Conclusion

- **Key Takeaways:**
  - The EDA has provided valuable insights into the key drivers of loan default risk, including external scores, loan characteristics, client demographics, and regional factors.
  - These insights can be leveraged to enhance the bank's risk management framework and improve loan portfolio performance.
- **Looking Forward:**
  - Continuous monitoring and refinement of risk assessment models are crucial to adapt to evolving market conditions and client behavior.
  - Further research, such as exploring advanced machine learning models and incorporating alternative data sources, can further enhance risk prediction capabilities.