Evoastra

Major Project - 1

Credit risk analysis

BONDARA BANK

PROPOSAL OF PROJECT

In today's complex financial landscape, lending institutions and investors face significant challenges in evaluating the creditworthiness of borrowers. Effective credit risk analysis is crucial for mitigating potential losses, optimizing portfolio performance, and ensuring the stability of financial systems. This project aims to develop a robust framework for credit risk assessment, leveraging advanced analytics and machine learning techniques to identify, measure, and manage credit risk exposure. By exploring the interplay between borrower characteristics, market conditions, and credit outcomes, this project seeks to contribute to the development of more accurate and reliable credit risk models, ultimately enhancing the resilience and competitiveness of financial institutions.

Project Requirements:

Exploratory Data Analysis (EDA).

Goal is to find such trend from the data.Identify potential defaulter to mitigate risk of financal loss for the company.


Bondora

AIM

1. Identify potential credit risks: Recognize borrowers who may default on their payments.
2. Assess creditworthiness: Evaluate the ability of borrowers to repay their debts.
3. Measure credit risk exposure: Quantify the potential loss due to credit defaults.
4. Manage credit risk: Develop strategies to mitigate or minimize credit risk.
5. Optimize credit portfolios: Ensure a balanced and diversified portfolio to maximize returns while minimizing risk.
6. Predict credit defaults: Use statistical models and machine learning techniques to forecast potential defaults.
7. Evaluate credit risk mitigation techniques: Assess the effectiveness of collateral, guarantees, and other risk-reducing measures.
8. Comply with regulatory requirements: Meet regulatory capital requirements and industry standards for credit risk management.
9. Enhance credit decision-making: Provide insights for informed lending and investment decisions.
10. Monitor and review credit risk: Continuously assess and update credit risk assessments to reflect changing market condition

Problem Statement – I

This assignment introduces the application of Exploratory Data Analysis (EDA) in a real business scenario, specifically in risk analytics within banking and financial services. You'll use EDA to analyze loan application data to identify patterns that help minimize default risks while ensuring that creditworthy applicants are not rejected. Companies face challenges in lending due to clients with insufficient or no credit history, which can lead to defaults. The goal is to determine driving factors behind loan defaults by analyzing consumer and loan data, helping the company make informed decisions such as approving, denying, or adjusting loan terms. You will explore data from three files: 'application_data.csv' (current client information), 'previous_application.csv' (past loan data), and 'columns_description.csv' (data dictionary). The objective is to help the company optimize its lending process by better assessing risk and ensuring the right customers receive loans.

Problem Statement – II

In this task, you will analyze loan data, present your approach, and summarize key findings. Begin by identifying and addressing missing data, explaining your method. Detect and explain any outliers without needing to remove them. Investigate data imbalance in the target variable (payment difficulties), and analyze this using univariate and bivariate analysis. Segment the data and find the top 10 correlations for clients with payment difficulties and other cases. Use visualizations to highlight key insights and explain why certain variables are important in differentiating clients likely to default.

Objective

- 1. Risk Assessment: Evaluate the creditworthiness of a borrower or counterparty.
- 2. Loss Estimation: Estimate the potential loss in case of default.
- 3. Risk Classification: Categorize borrowers into different risk categories.
- 4. Risk Pricing: Determine the appropriate interest rate or risk premium to charge.
- 5. Risk Management: Identify and mitigate potential credit risks.
- 6. Regulatory Compliance: Meet regulatory requirements and standards.
- 7. Portfolio Optimization: Optimize credit portfolios to maximize returns while minimizing risk.

Understand the Data

Objective:

Familiarize with the dataset and understand variables.

Datasets:

application_data.csv: Information at the time of loan application.

previous_application.csv: Details of past loan applications.

columns_description.csv: Description of variables.

- Key Points:

Examine the meaning of variables.

Identify the relevance of each variable to the business problem

Understand the Data

- Objective:
  - Familiarize with the dataset and understand variables.

- Datasets:
  - application_data.csv: Information at the time of loan application.
  - previous_application.csv: Details of past loan applications.
  - columns_description.csv: Description of variable.

- Key Points:
  - Examine the meaning of variables.
  - Identify the relevance of each variable to the business problem.

DATA collection

- Objective:
  - Gather and compile data for analysis.
- Data Sources
  - Internal data from Bondora Bank's Ion applications.
- Key Points:
  - Ensure the completeness and accuracy of data collected.
  - Understand the historical context and relevance of the data.

Data cleaning

- Objective:
  - Identify and address data quality issues.
- Steps:
  - Handle missing values (imputation or removal).
  - Detect and manage outliers.
- Key Points:
  - Discuss the approach for missing data.
  - Importance of clean data for accurate analysis.

Data transformation

Objective:

Convert data into a suitable format for analysis.

Steps:

Normalize or scale variables if needed.

Encode categorical variables.

Key Points:

Transform data to enhance analytical accuracy.

Consider the impact of transformation on the analysis.

EDA

Exploratory Data Analysis (EDA) is the process of analyzing and summarizing datasets to uncover patterns, trends, relationships, and anomalies without making any prior assumptions. EDA is a crucial step in data analysis as it helps in understanding the data's structure, detecting outliers, and identifying important variables that could influence the outcome of interest.

EDA

uses of EDA in your project:

1. **Data Understanding**: Identify key variables and detect data quality issues (e.g., missing values, outliers).
2. **Pattern Identification**: Discover trends and relationships between variables that influence credit risk.
3. **Data Imbalance Detection**: Identify and address class imbalances (e.g., loan approvals vs. rejections).
4. **Key Variable Identification**: Find driver variables through correlation analysis and segmented analysis.
5. **Support Decision-Making**: Provide insights to minimize financial losses and optimize loan approval processes.
6. **Model Development**: Lay the foundation for building accurate credit risk prediction models.
7. **Visualization**: Create visual representations to communicate findings effectively.
8. **Hypothesis Generation**: Form hypotheses about factors contributing to loan defaults for further analysis.

DATA visualization

Objective:

Use visualizations to uncover patterns and insights.

Techniques:

Univariate analysis: Histograms, Boxplots.

Bivariate analysis: Scatter plots, Correlation matrices.

Key Points:

Visualize data distributions and relationships.

Identify key variables influencing loan defaults.

Descriptive statistics

Objective:

Summarize data characteristics through statistical measures.

Key Measures:

Mean, Median, Mode

Standard Deviation, Variance

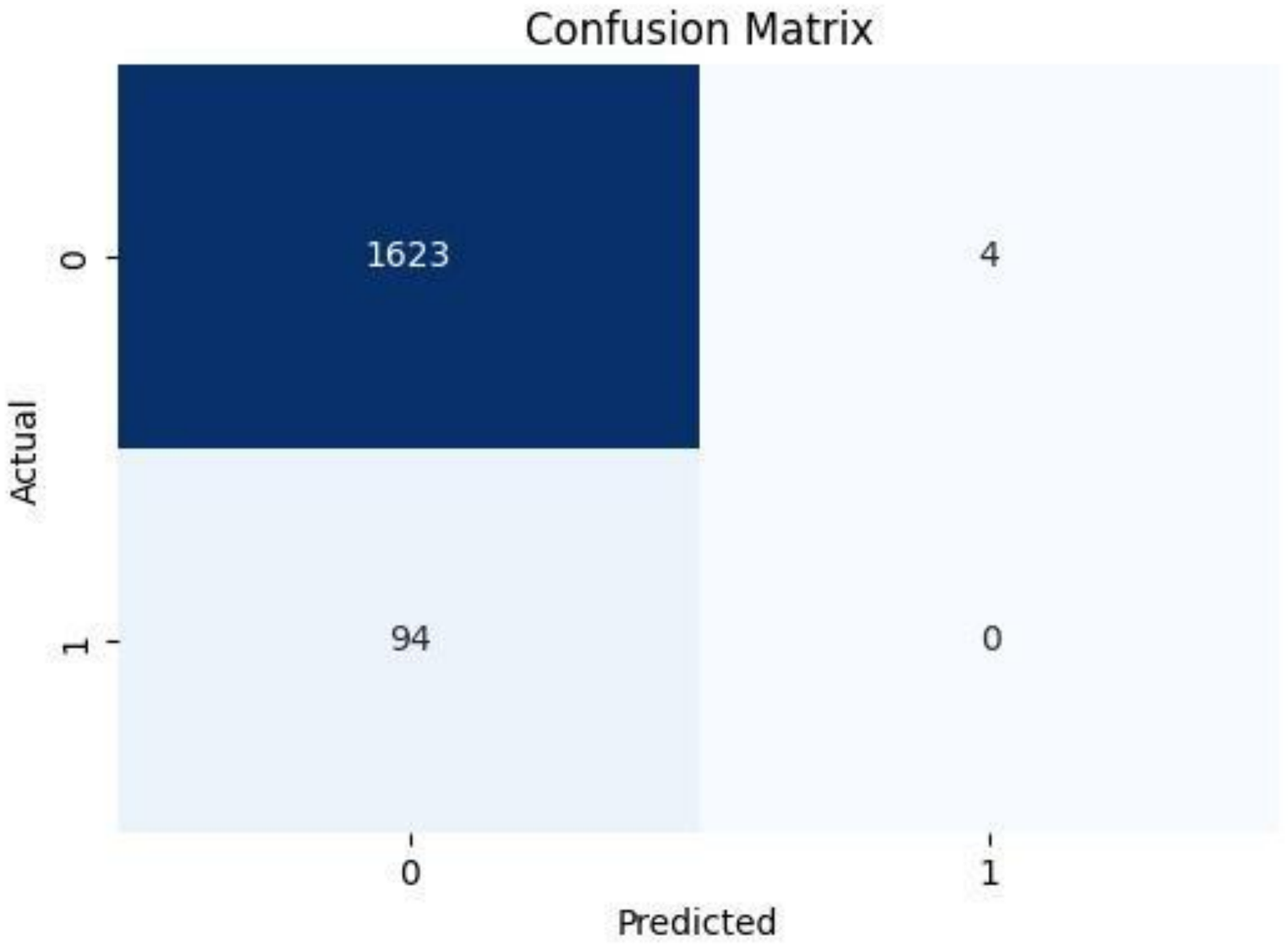
Correlation Coefficients

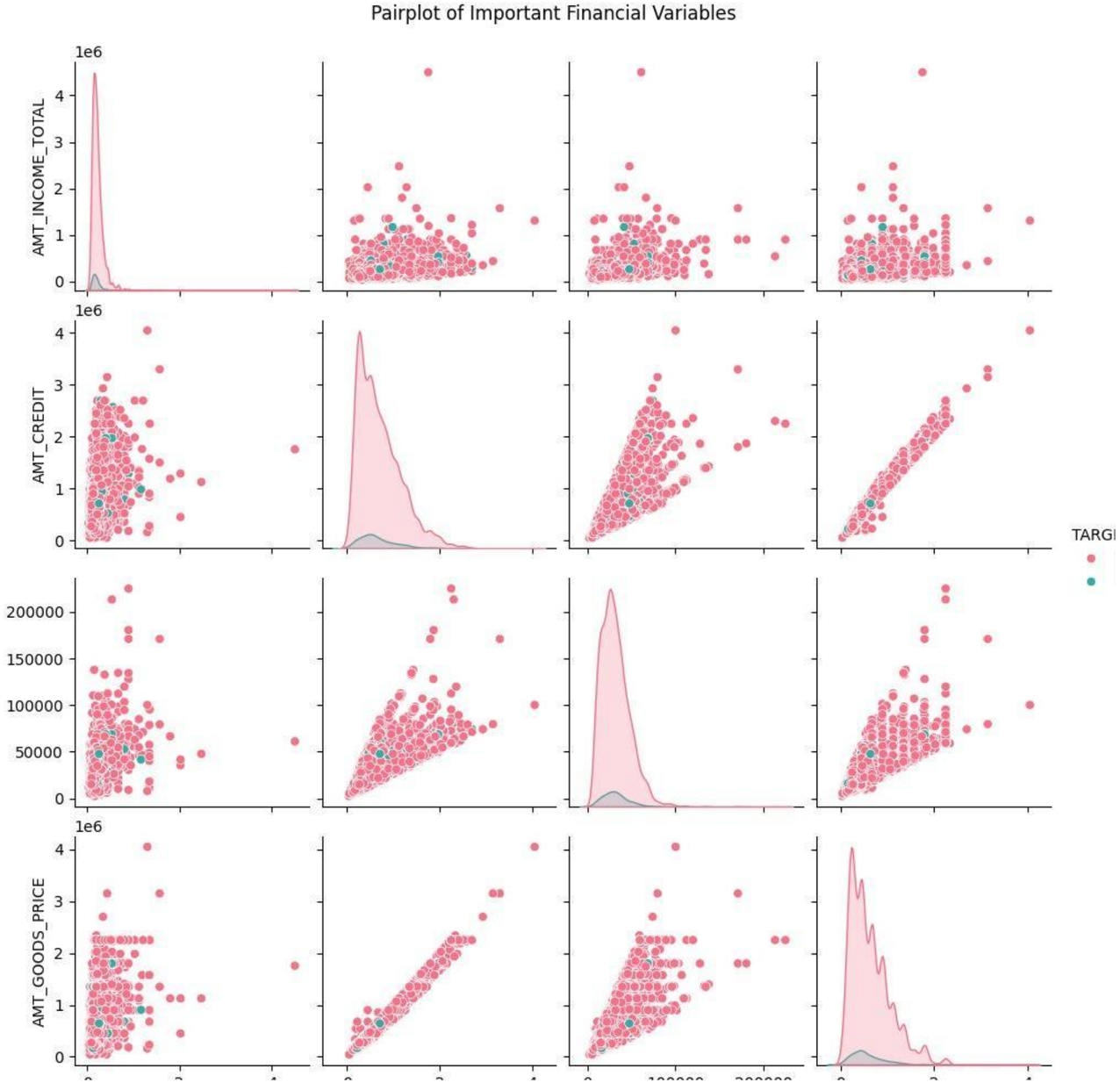
Key Points:

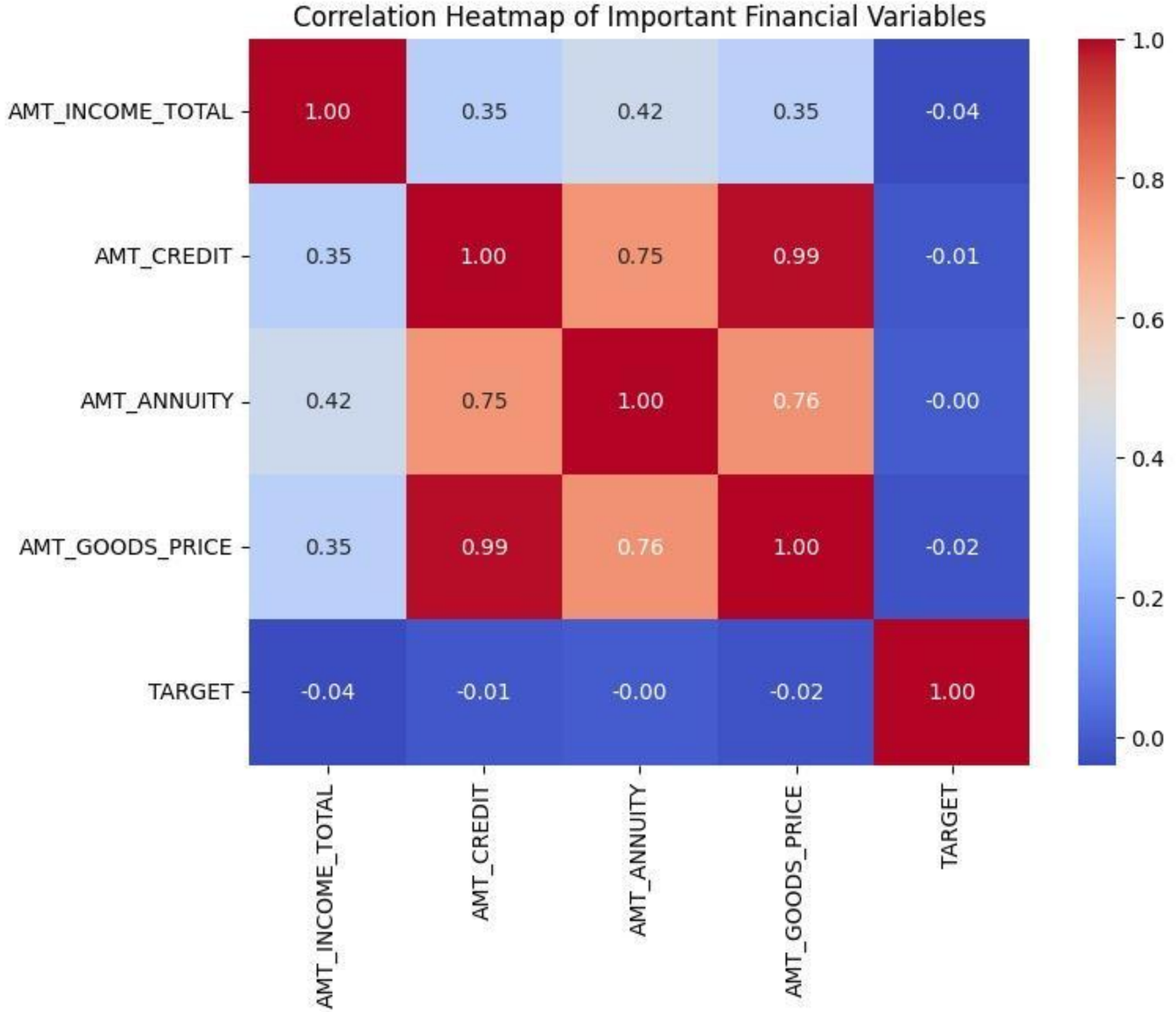
Use statistics to describe and summarize data.

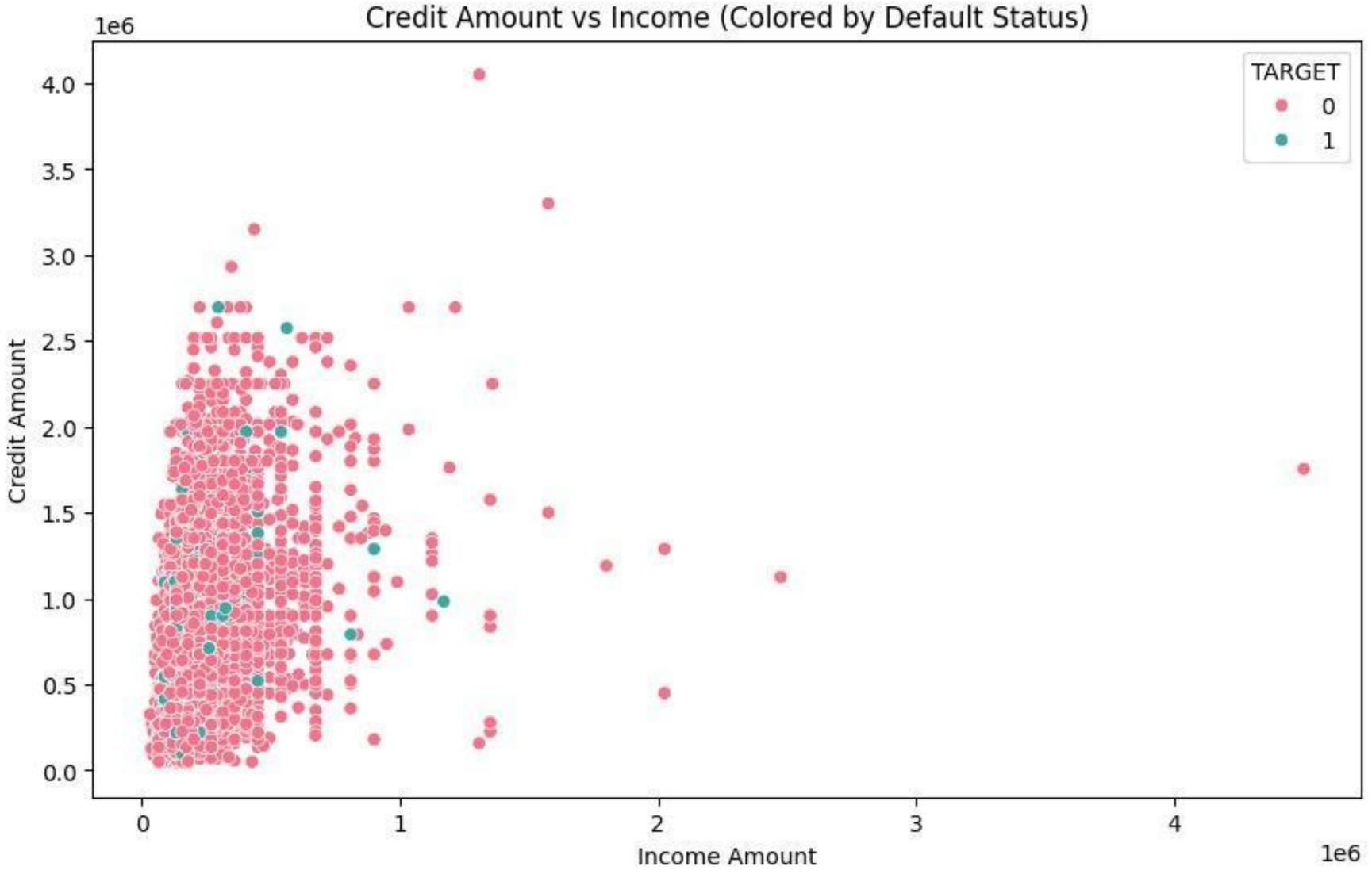
Highlight top correlations with loan default risk.

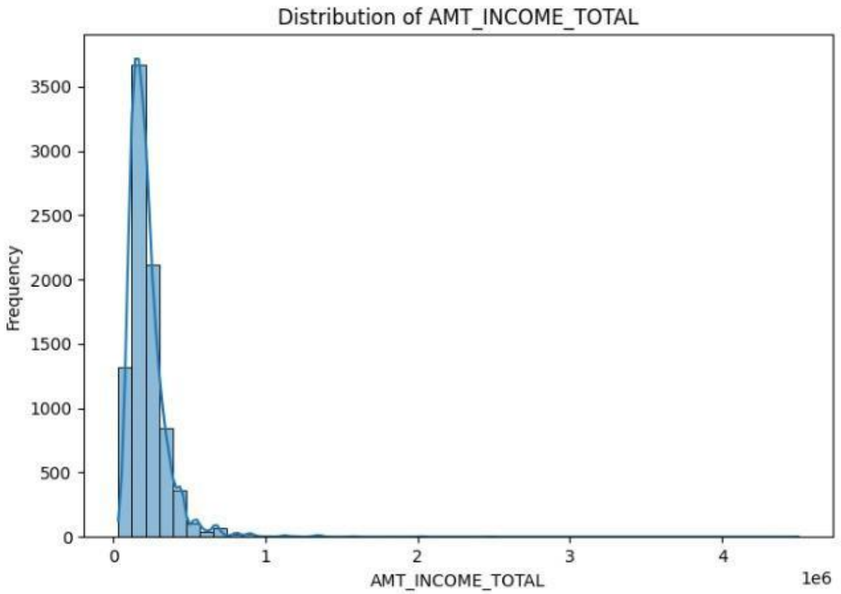

XGBoost Feature Importance
DAYS_BIRTH
DAYS_EMPLOYED
AMT_ANNUITY
AMT_CREDIT
AMT_INCOME_TOTAL
AMT_GOODS_PRICE
548.0
530.0
516.0
373.0
347.0
264.0
Features
0
100
200
300
400
500
600
F score

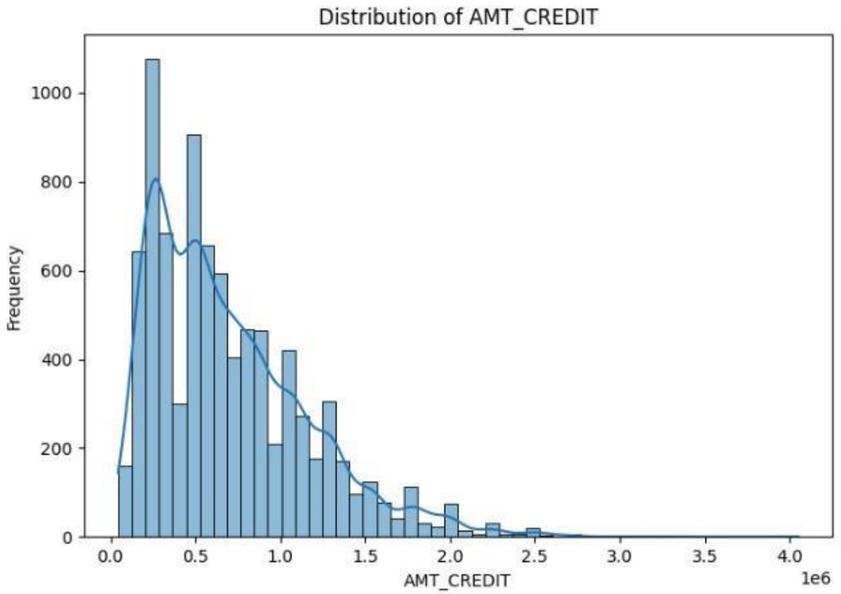

Confusion Matrix
1623
4
94
0
0
1
Actual
Predicted

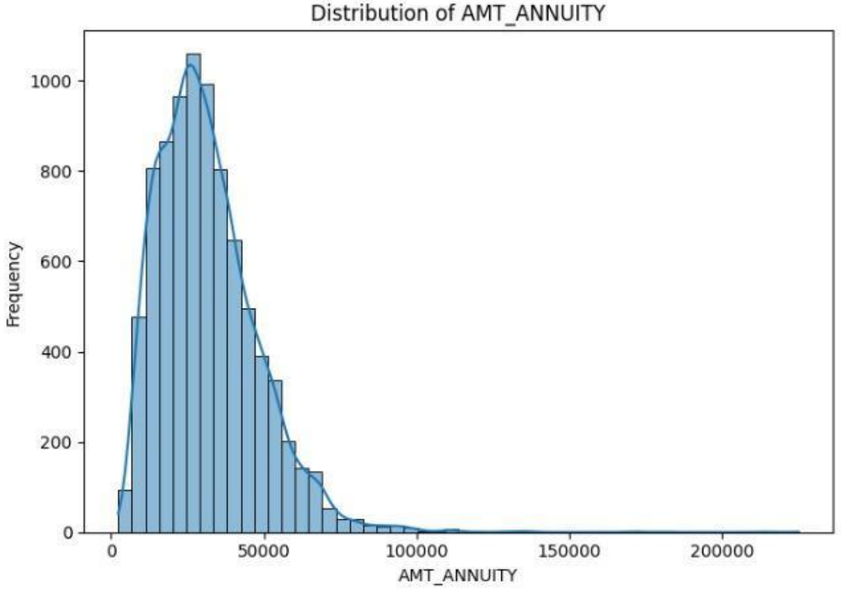

Pairplot of Important Financial Variables
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_GOODS_PRICE
1e6
200000
150000
100000
50000
0
1
2
3
4
TARGI

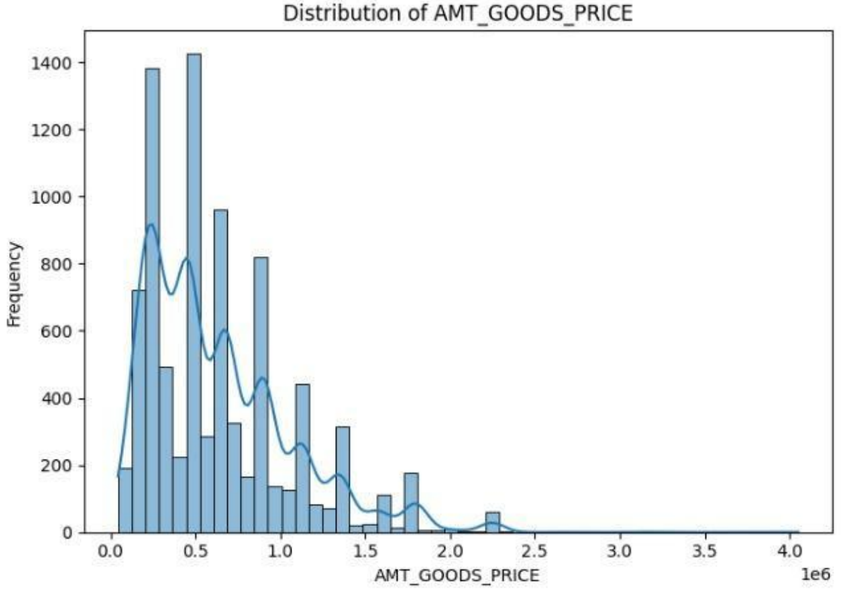

Correlation Heatmap of Important Financial Variables
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
TARGET
1.00 0.35 0.42 0.35 -0.04
0.35 1.00 0.75 0.99 -0.01
0.42 0.75 1.00 0.76 -0.00
0.35 0.99 0.76 1.00 -0.02
-0.04 -0.01 -0.00 -0.02 1.00
1.0
0.8
0.6
0.4
0.2
0.0

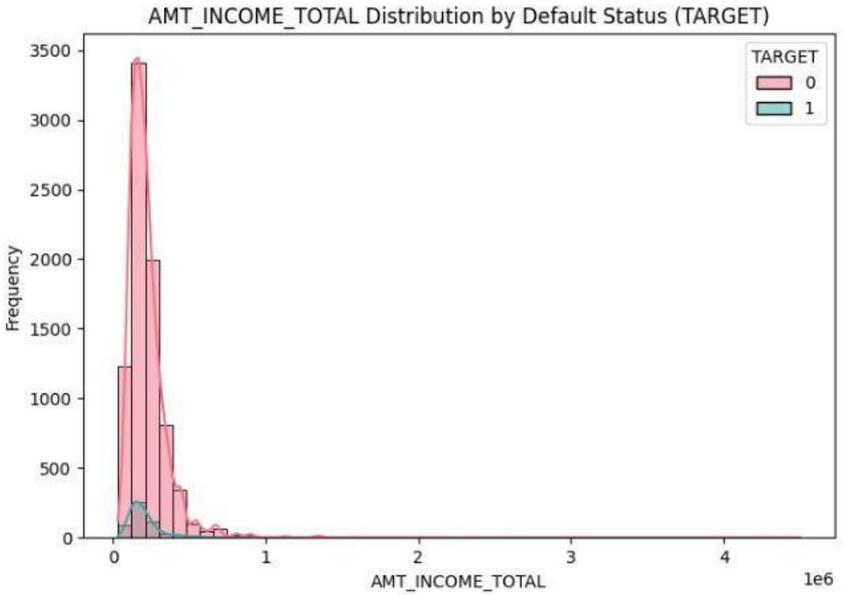

Credit Amount vs Income (Colored by Default Status)
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Credit Amount
TARGET
0
1
0
1
2
3
4
Income Amount
1e6

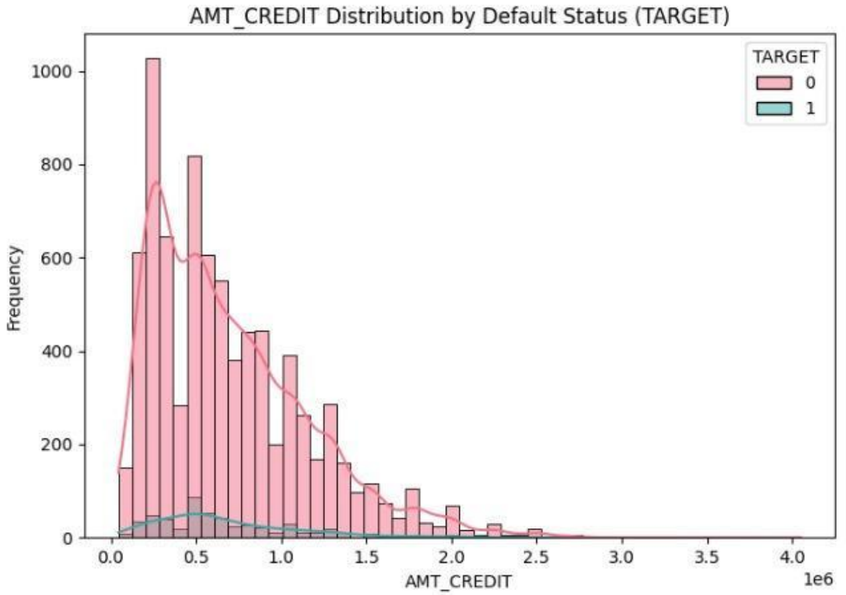

Distribution of AMT_INCOME_TOTAL
Frequency
AMT_INCOME_TOTAL
1e6

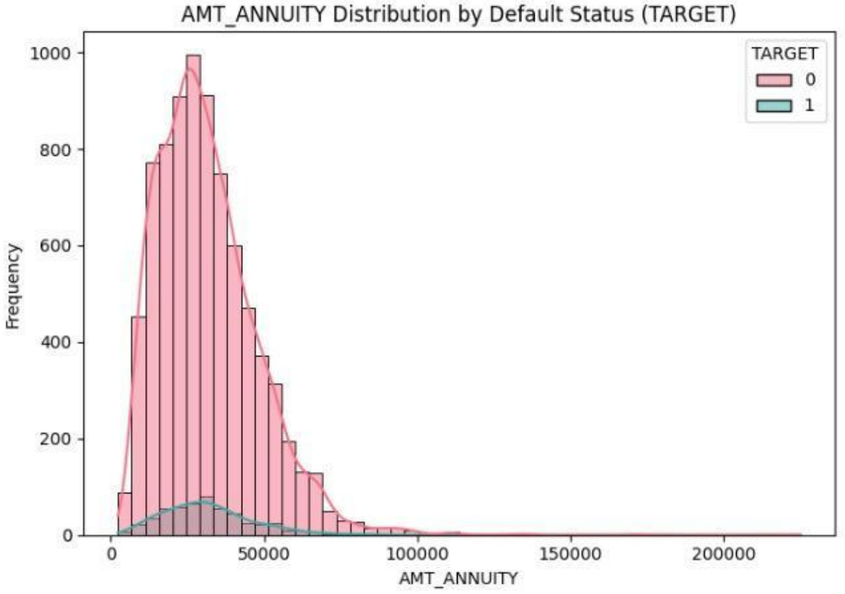

Distribution of AMT_CREDIT
Frequency
AMT_CREDIT
1e6

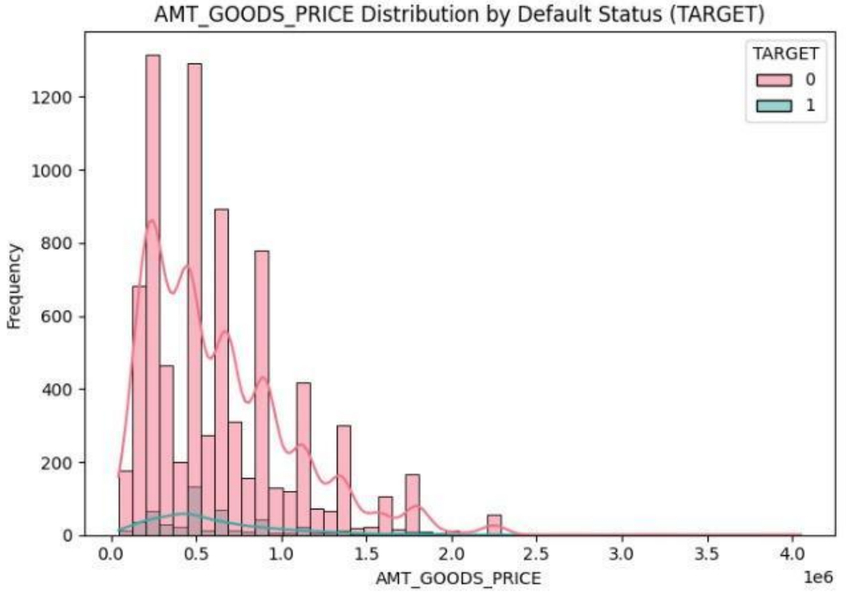

Distribution of AMT_ANNUITY
Frequency
AMT_ANNUITY

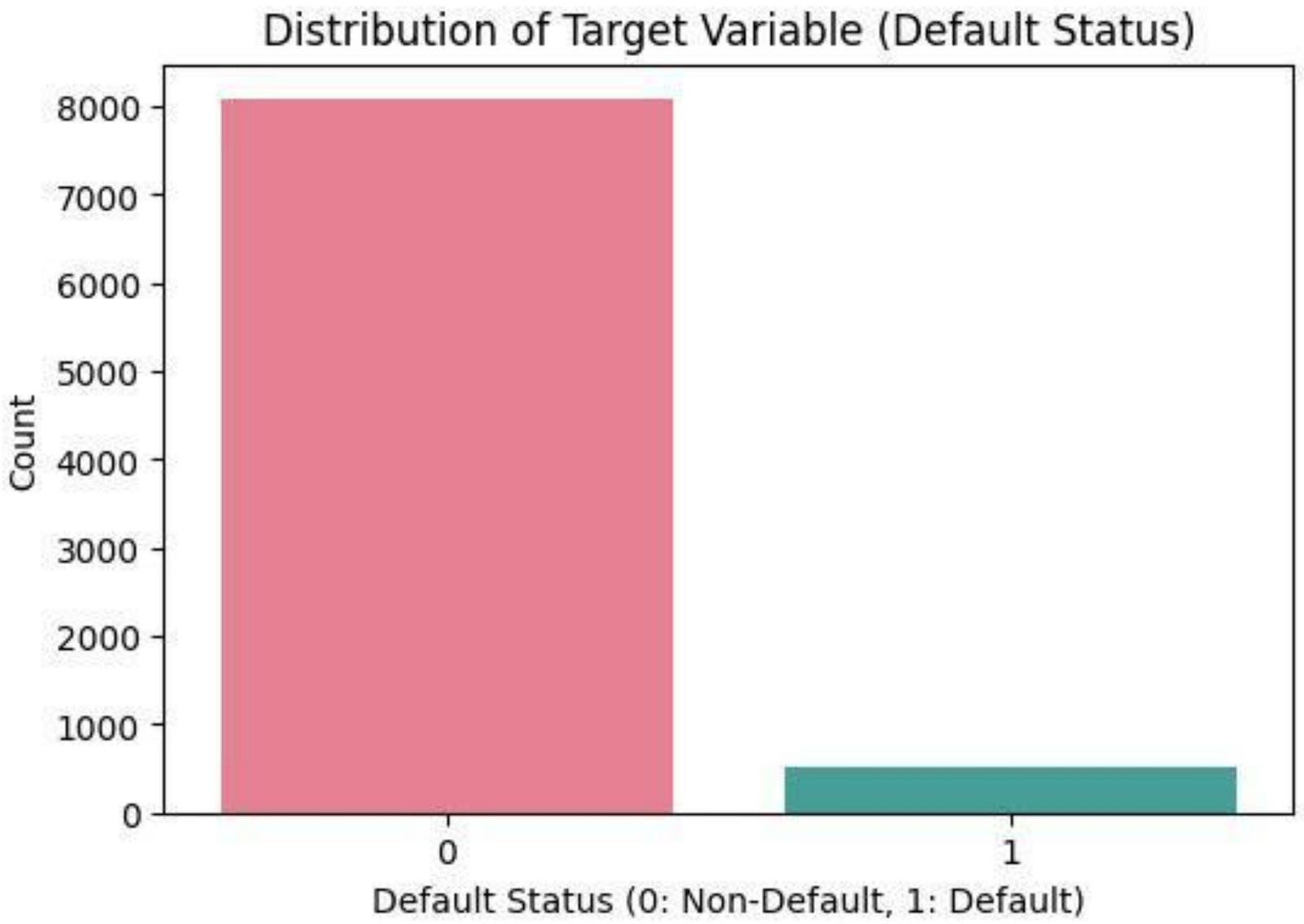

Distribution of AMT_GOODS_PRICE
Frequency
AMT_GOODS_PRICE
1e6


AMT_INCOME_TOTAL Distribution by Default Status (TARGET)
TARGET
0
1
Frequency
AMT_INCOME_TOTAL
1e6


AMT_CREDIT Distribution by Default Status (TARGET)
TARGET
0
1
Frequency
AMT_CREDIT
1e6


AMT_ANNUITY Distribution by Default Status (TARGET)
TARGET
0
1
Frequency
AMT_ANNUITY


AMT_GOODS_PRICE Distribution by Default Status (TARGET)
TARGET
0
1
Frequency
AMT_GOODS_PRICE
1e6


Distribution of Target Variable (Default Status)
Count
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1
Default Status (0: Non-Default, 1: Default)

Classification Report:

	precision	recall	f1-score	support
0	0.95	1.00	0.97	1627
1	0.00	0.00	0.00	94
accuracy			0.94	1721
macro avg	0.47	0.50	0.49	1721
weighted avg	0.89	0.94	0.92	1721

```
The individual is predicted to be a non-defaulter with a probability of 0.98.
```

Recomedations

In this project, you will use Exploratory Data Analysis (EDA) to help a consumer finance company minimize loan defaults while ensuring that creditworthy applicants are not rejected. By analyzing loan application data and identifying patterns, you will determine the factors that influence payment difficulties (default). Your approach includes handling missing data, detecting outliers, and addressing data imbalances. You'll conduct univariate and bivariate analysis, identifying the top correlations among key variables. The aim is to extract business insights that help in risk assessment, ensuring better lending decisions while reducing financial losses due to defaults.

Thank You

Presented By:

Team M

CONCLUSION

Findings:

Top 10 correlations with loan defaults.

Variables with significant impact on loan repayment behavior.

Implications:

How these insights can be used for better decision-making in loan approvals.