Credit EDA Case Study Assignment

By

Kaila Sai Pranava Karthik

Problem Statement

This case study aims to find trends and variables that affect a client's tendency to default on loans, or have trouble repaying them.
Financial institutions would benefit from these insights:

- Enhance loan approval decision-making procedures.
- By recognizing strong default signs, you can reduce the danger of issuing loans to high-risk candidates.
- Create plans to lower the number of non-performing loans by comprehending the important factors influencing repayment patterns.

To find the answers to these questions, we will examine the data using a variety of exploratory data analysis techniques.
Such an analysis ensures that the capability of clients to repay their loans is not denied credit, whereas those at high risk are identified effectively.

Assumptions

•I have retained the outliers as it was necessary to preserve data integrity.
•Columns with >40% missing values were dropped and other missing values were dealt by imputing techniques.
•Special values like 'XNA' and 'XAP' were either replaced or dropped based on its count and relevance.

Approach and Methodology

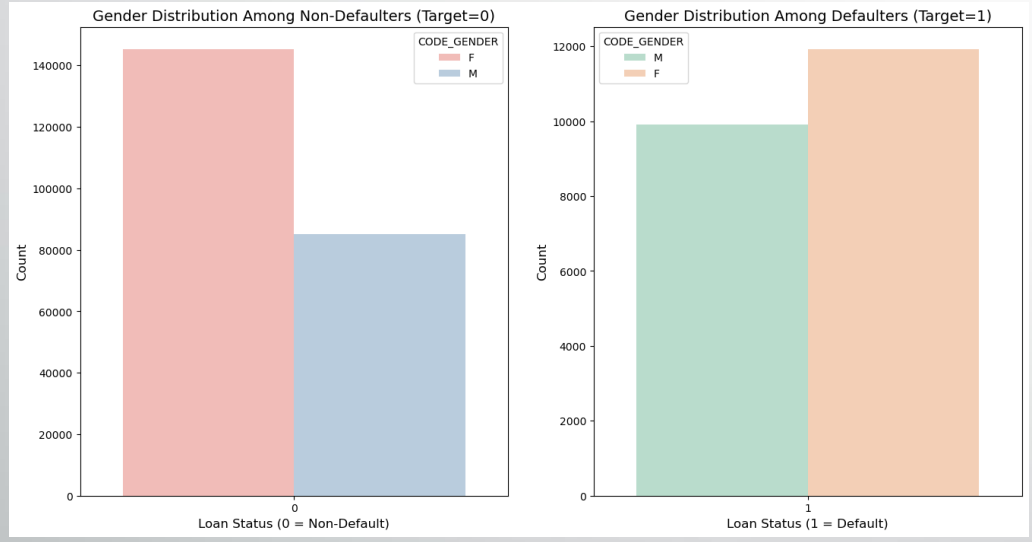
Performed these following steps for application, previous application, and merged dataset
Understanding the Domain: I had Identified the business problem and objectives.
Data Loading & Exploration: Loaded the datasets and checked their structure and metadata.
Data Cleaning:
•Identified and addressed missing values by dropping values >40% and using imputer techniques for others
•Dealt with special cases like XNA and XAP values either replace or dropped them accordingly.
•Removed some of the unnecessary columns based on relevance.
Outlier Detection & Handling:
•Identified outliers
•I had decided to retain outliers as they are crucial for analysis.
Data Imbalance Check:
•Calculated the imbalance ratio for the target variable (defaulters vs. non-defaulters).
Exploratory Data Analysis:
•Performed univariate, segmented univariate and bivariate analysis.

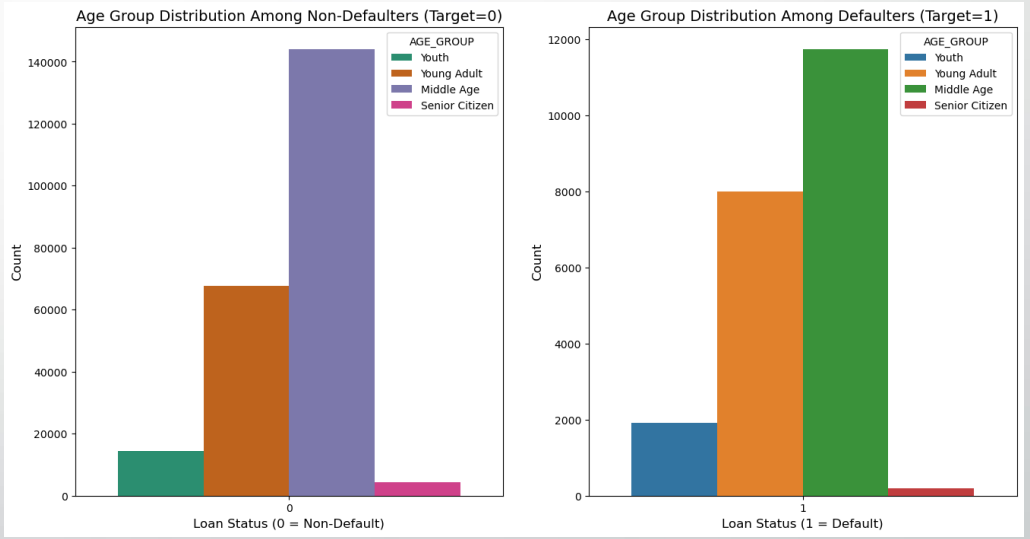
Approach and Methodology

Business Problem Understanding → Data Loading → Data Cleaning → Outlier Handling → Imbalance Check → Exploratory Data Analysis (EDA)

Graphs and Insights

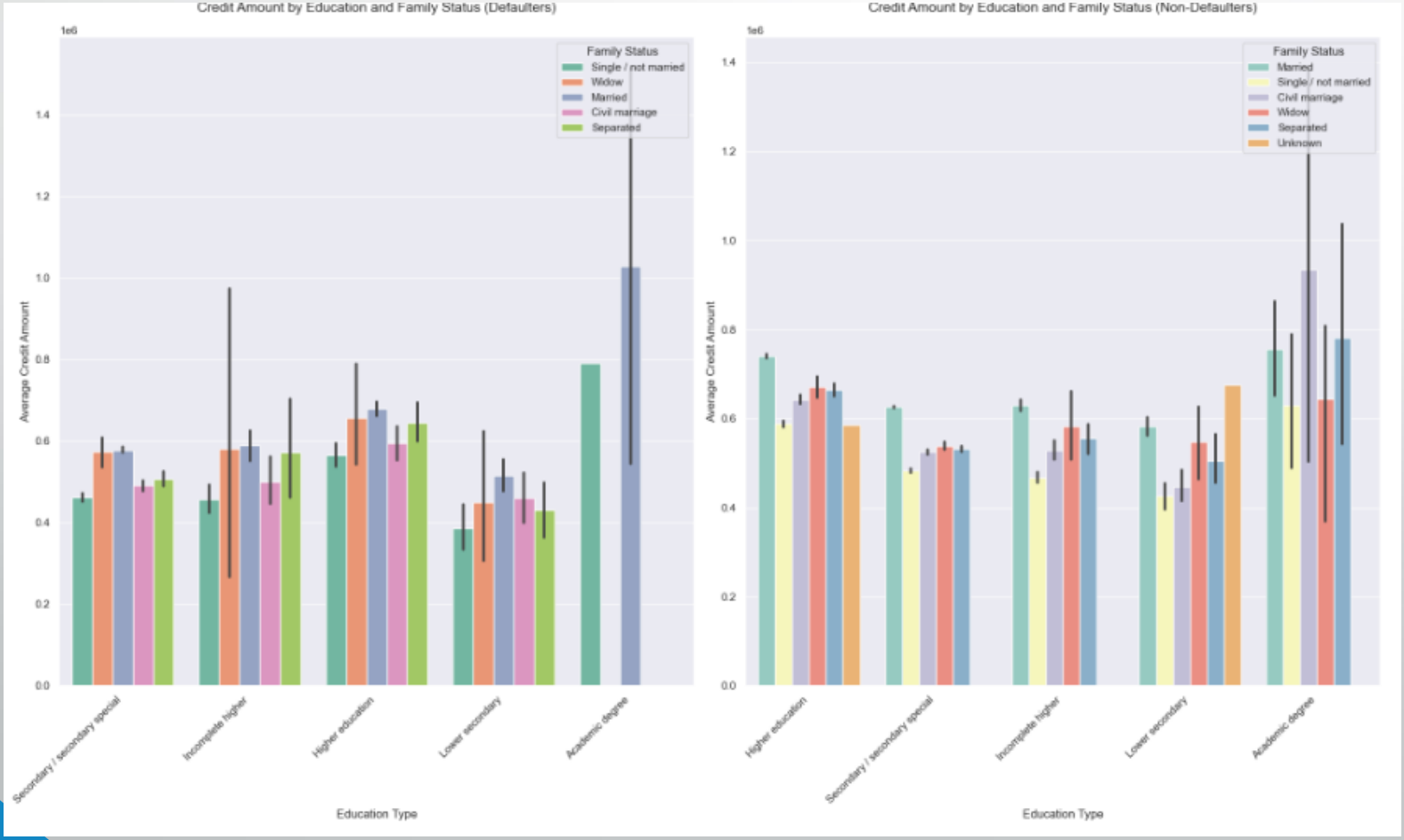
Application Dataset





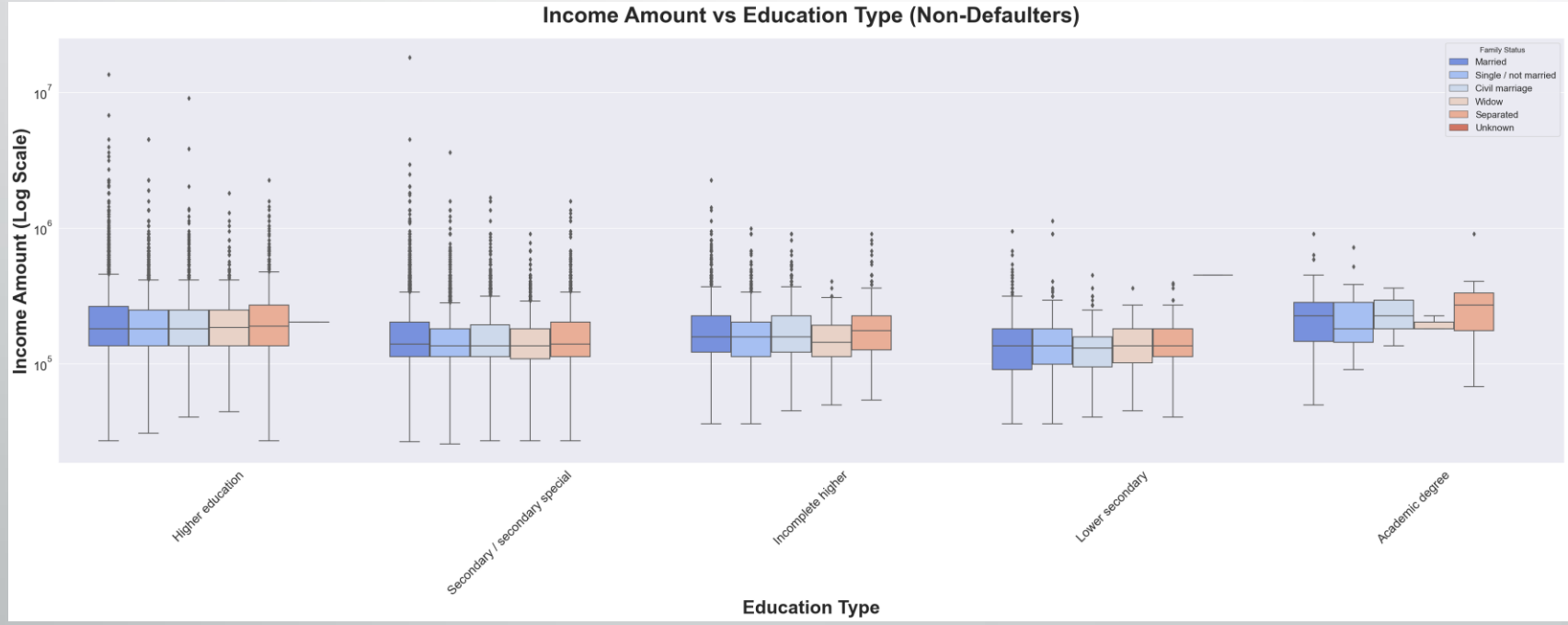
Graphs and Insights

Application Dataset



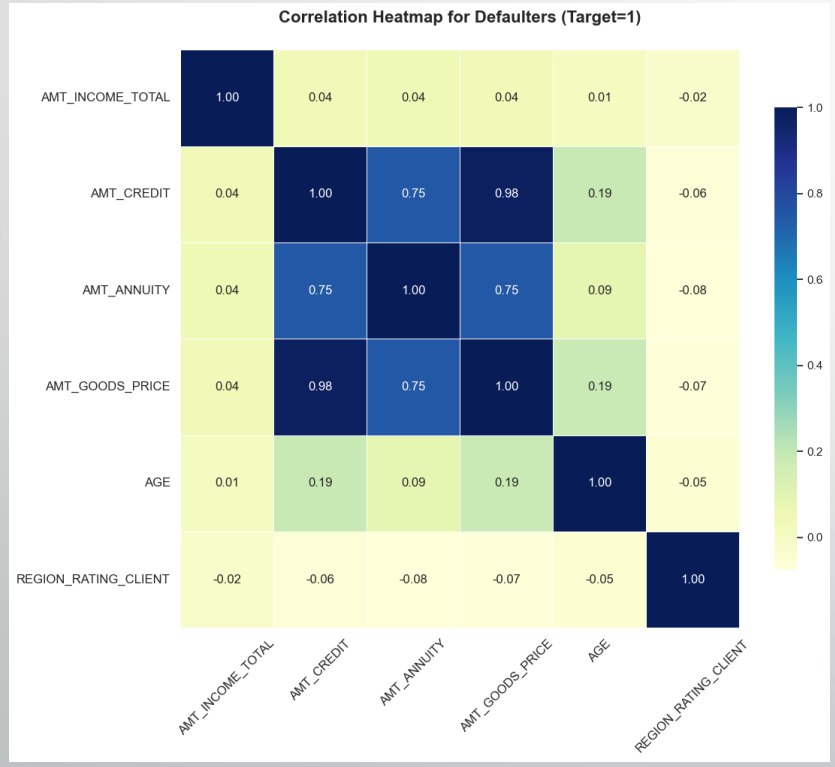
Graphs and Insights

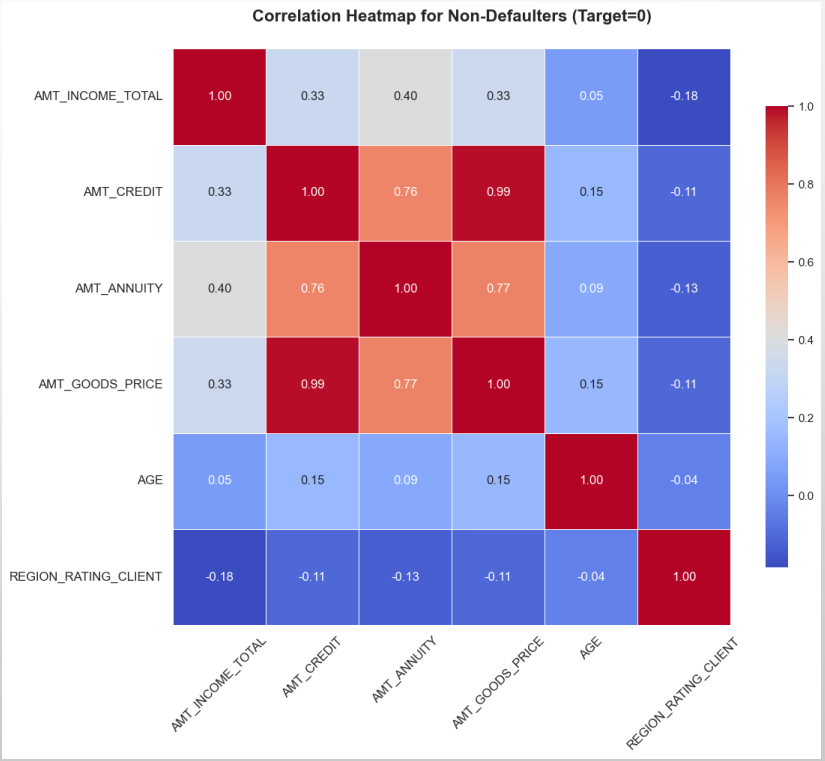
Application Dataset



Graphs and Insights

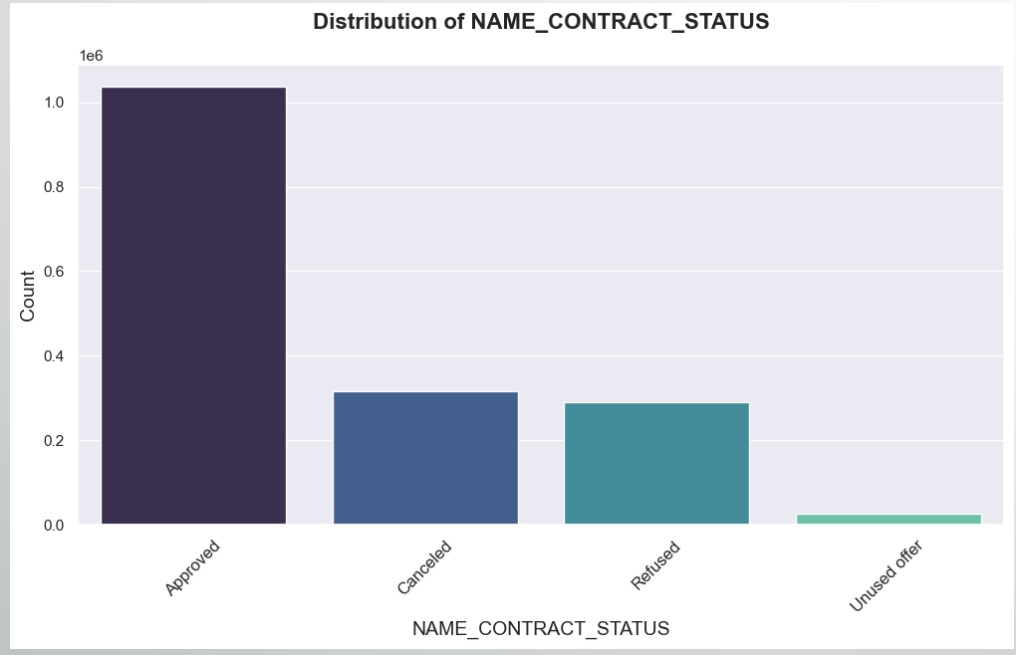
Application Dataset

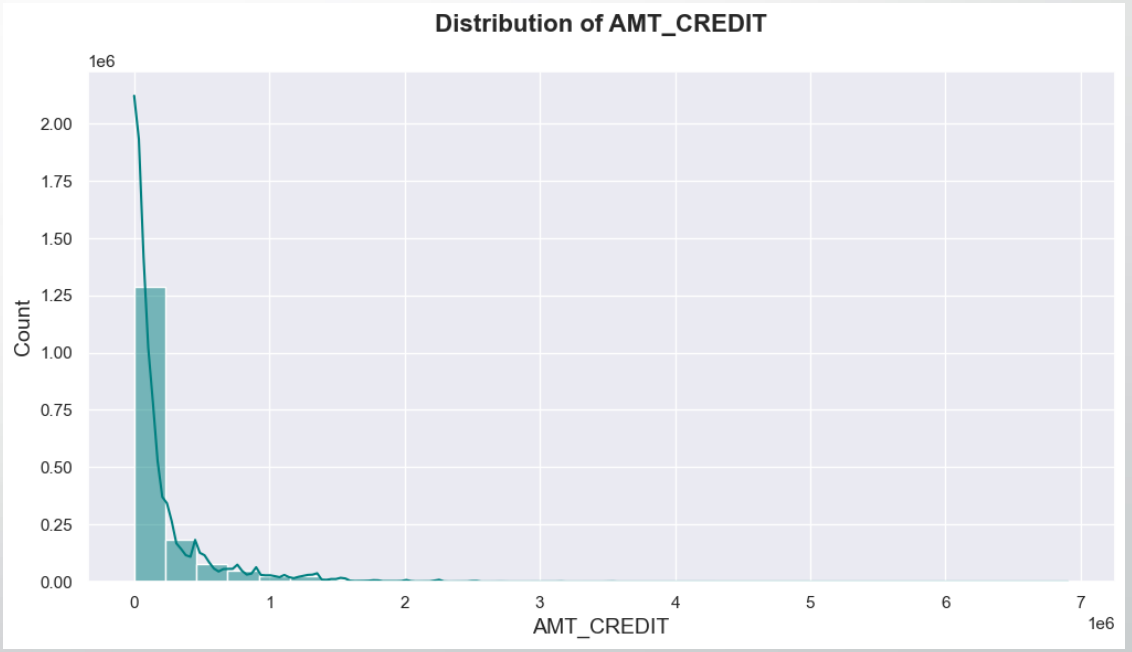




Graphs and Insights

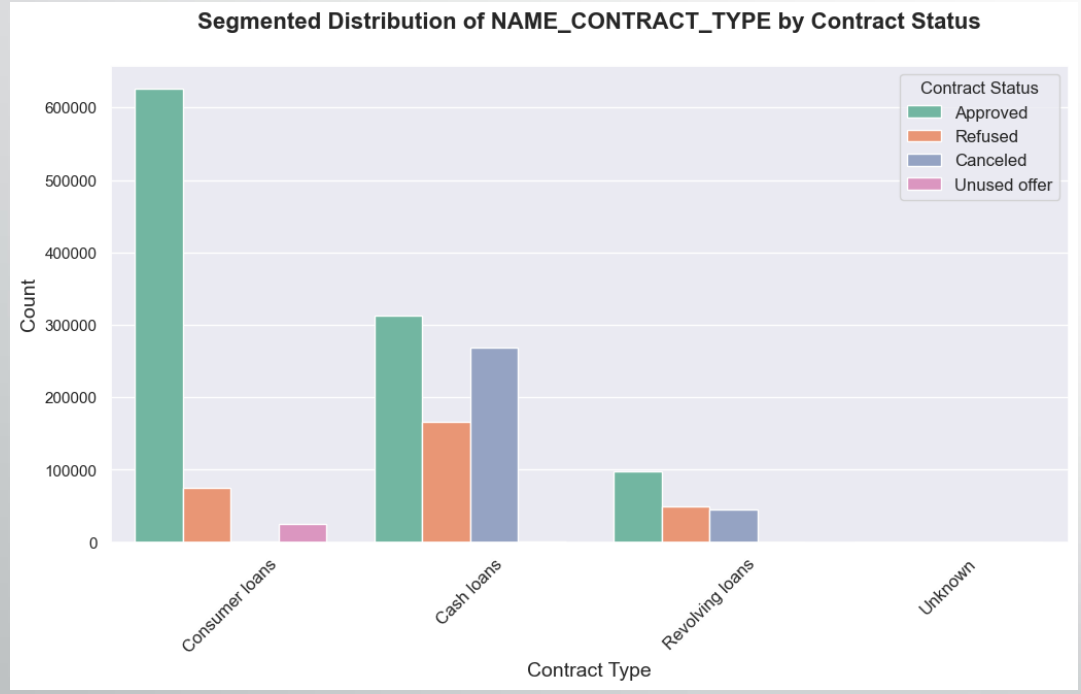
Previous Application Dataset

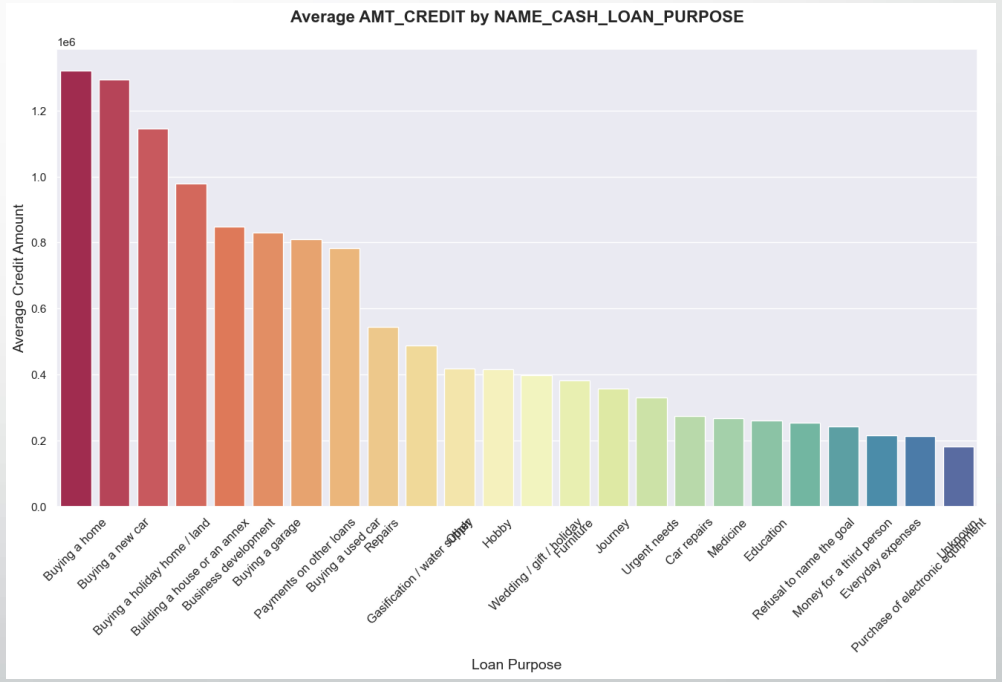




Graphs and Insights

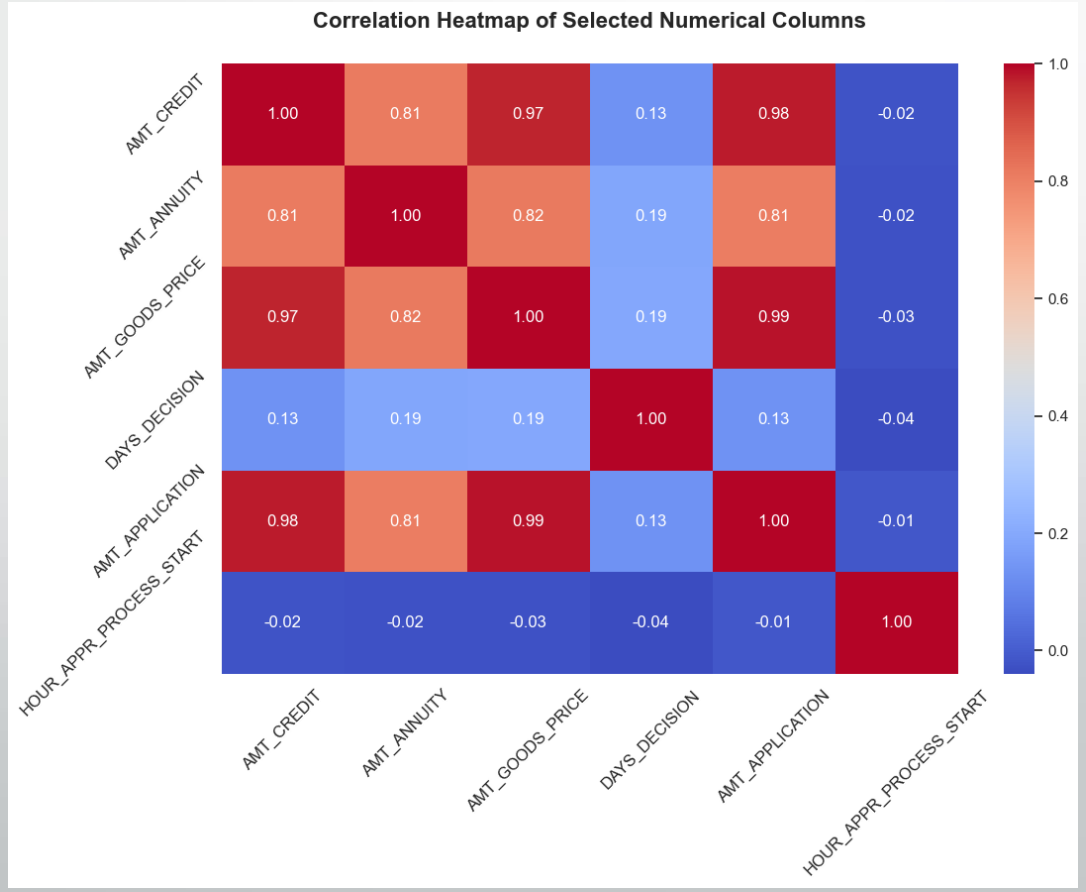
Previous Application Dataset





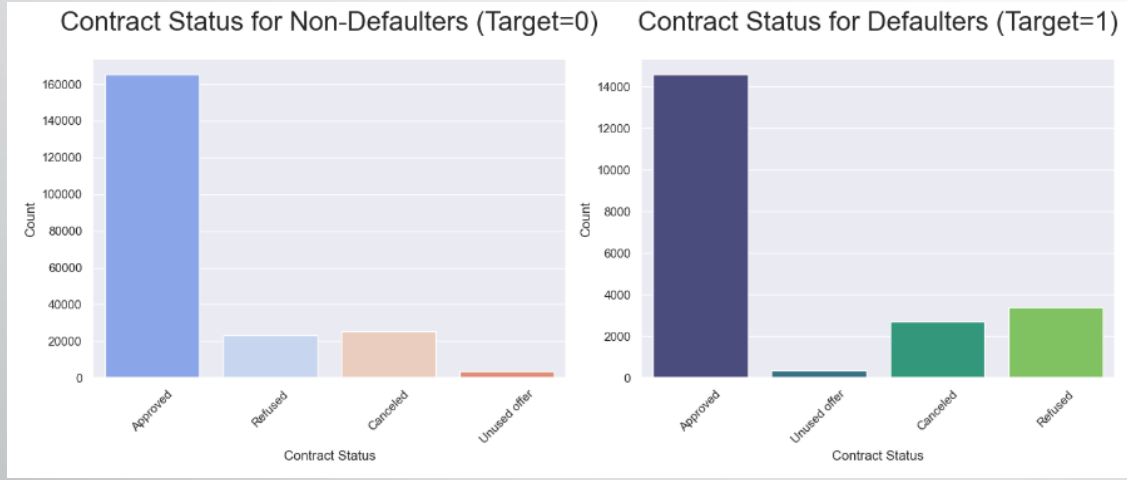
Graphs and Insights

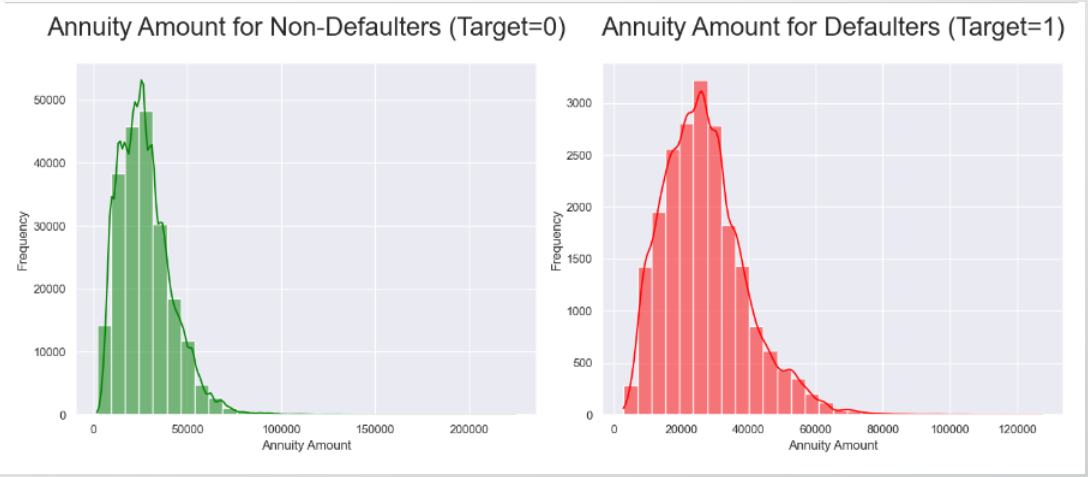
Previous Application Dataset



Graphs and Insights

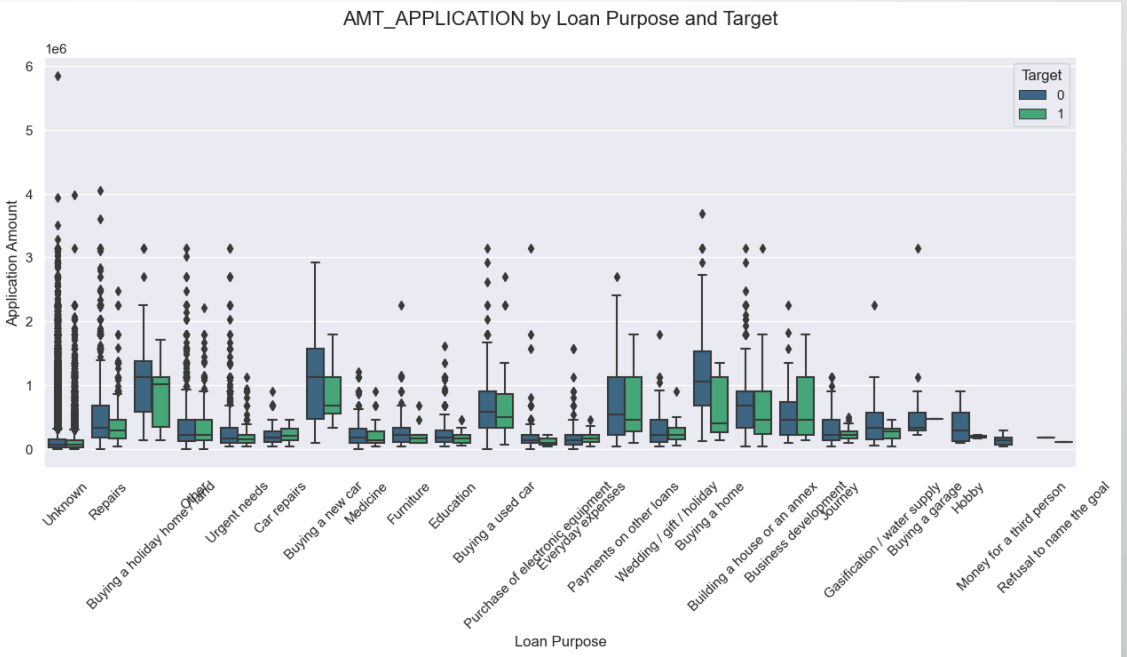
Merged Dataset





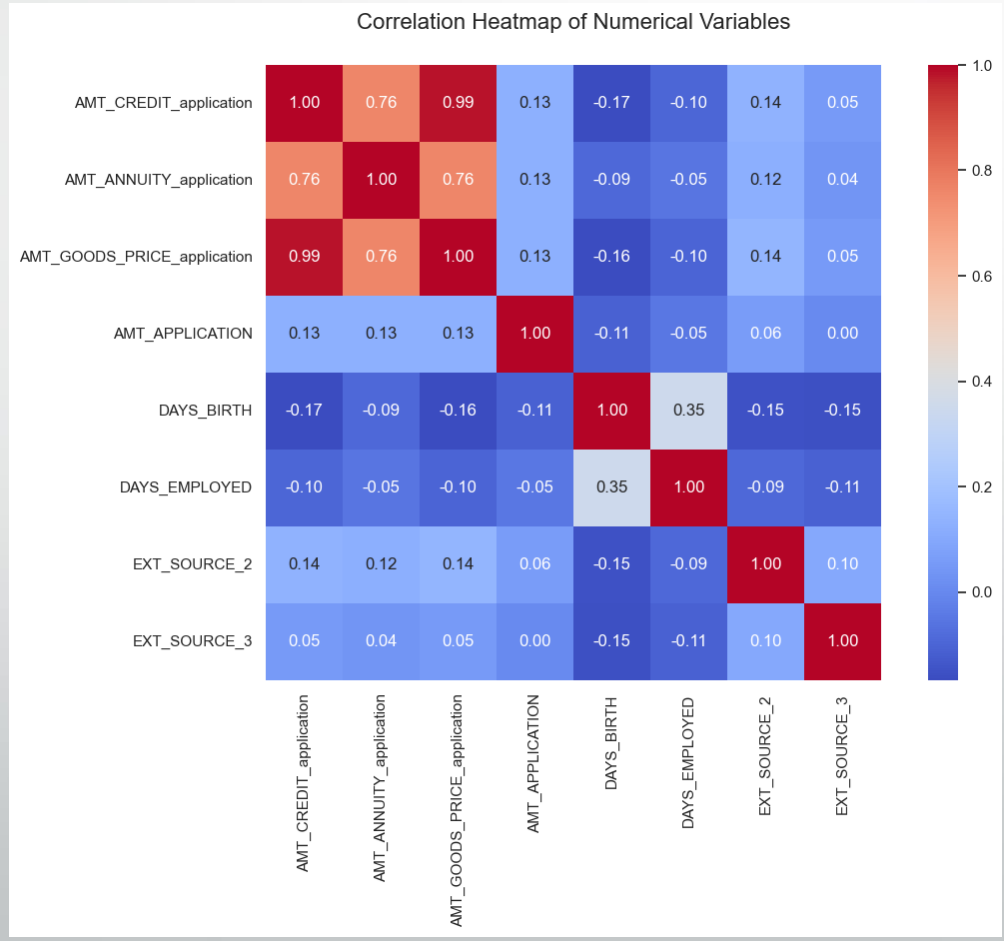
Graphs and Insights

Merged Dataset



Graphs and Insights

Merged Dataset



Conclusions and Recommendations

Insights:

1.Middle and low-income groups have higher default rates.
2.Significant data imbalance between defaulters (Target=1) and non-defaulters (Target=0).
3.Key risk indicators: AMT_CREDIT, AMT_INCOME_TOTAL, EXT_SOURCE scores
4.Family and education status impact loan default probability..
5.Consumer loans show higher default risk.
6.Education and family size influence default behavior.
7.Previous application patterns provide valuable insights.

Recommendations:

1.Develop risk models using income-to-credit ratios and external scores.
2.Customize loan policies for specific demographics.
3.Stricter policies for high-risk loan types.
4.Set credit limits to prevent over-leverage.
5.Use sampling techniques to address data imbalance.
6.Derive additional metrics for enhanced risk prediction.