Credit Risk Analysis

Problem Statement and Approach Overview

Presented by: Kiran Kumar Das

Date: October 27, 2024

Problem Statement & Expected Outcomes:

- Problem Statement: Analyze client data to identify credit risk indicators, manage missing values, detect outliers, and assess data imbalance.

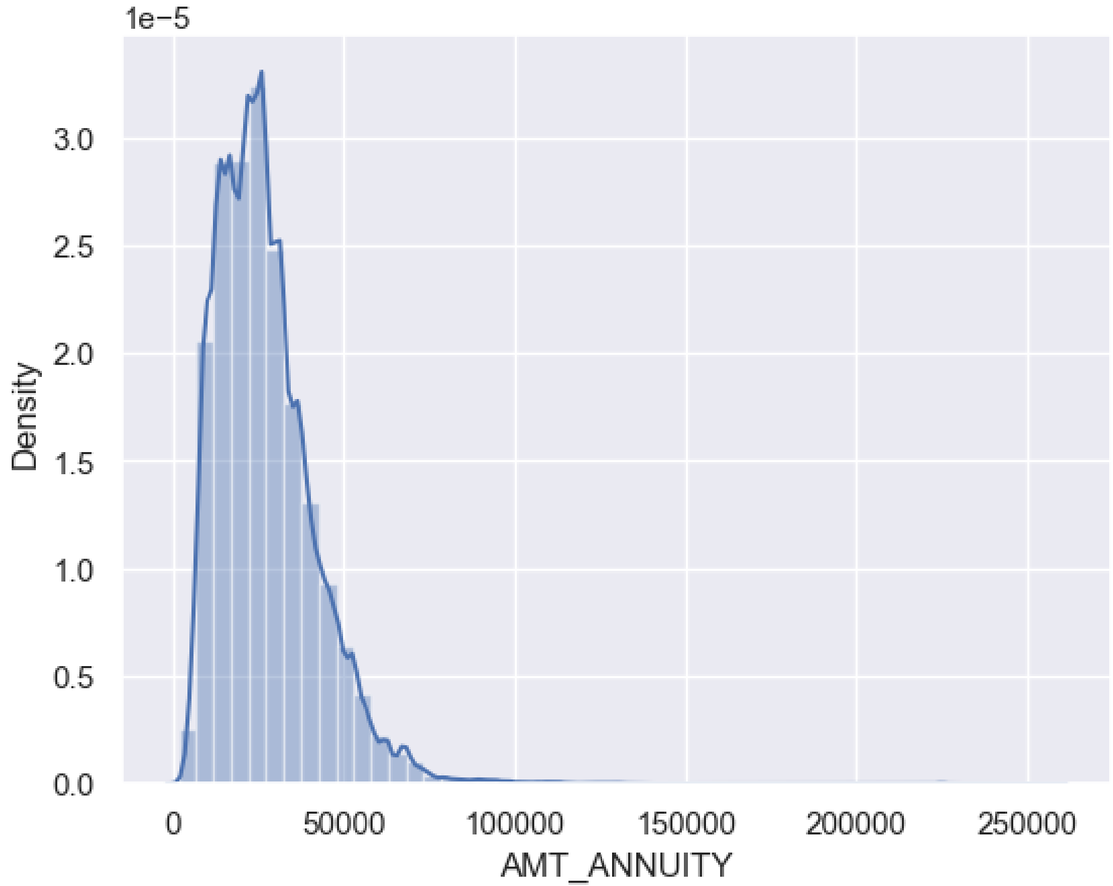
- Expected Outcomes: Key insights and visualizations to differentiate clients with payment difficulties from others.

- Methodology:

➢ Data exploration and cleaning.
➢ Detection of missing values, outliers, and data imbalance.
➢ Univariate and bivariate analysis for insights on variable influence.
➢ Segmentation-based correlation analysis targeting high-risk variables.

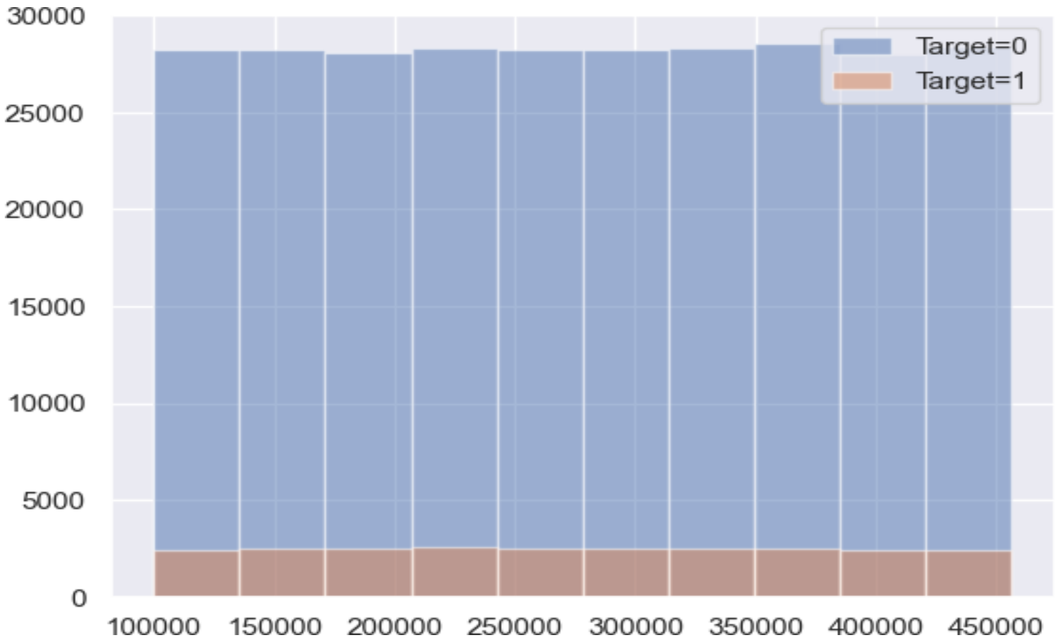
Analysis Approach Overview:

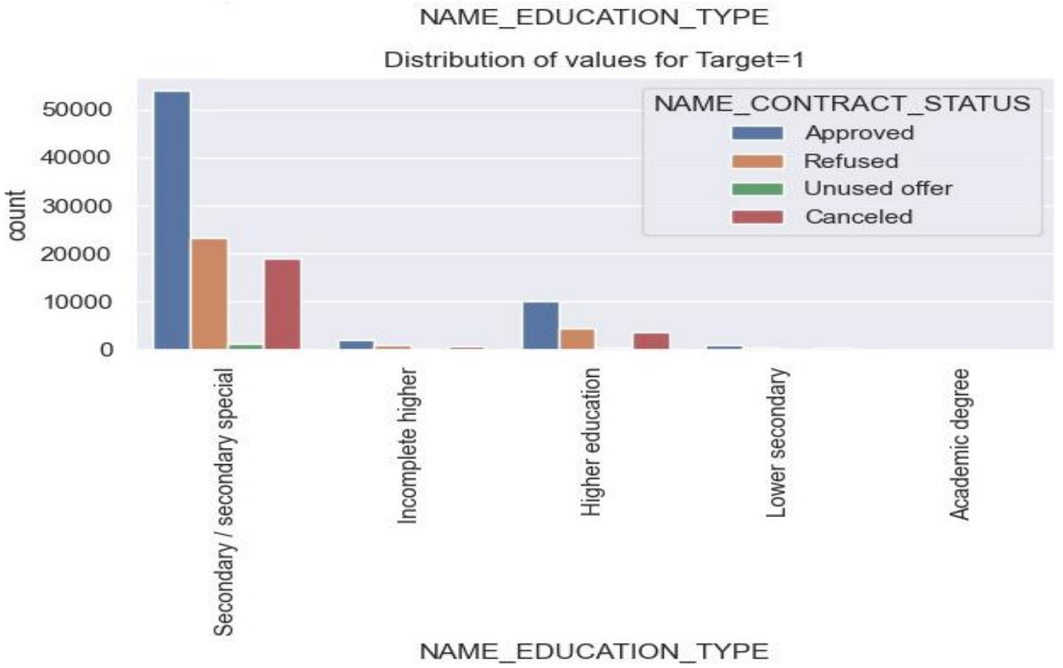
- Missing Data Handling: Remove or replace missing values appropriately.
- Outlier Detection: Detect outliers without removing data.
- Data Imbalance: Calculate imbalance ratios, focusing on the target variable.
- Correlation Analysis: Identify top correlations per target segment.
- Univariate & Bivariate Analysis: Explain trends in business terms.



Analysis Approach Overview:

- Univariate & Bivariate Analysis: Explain trends in business terms.



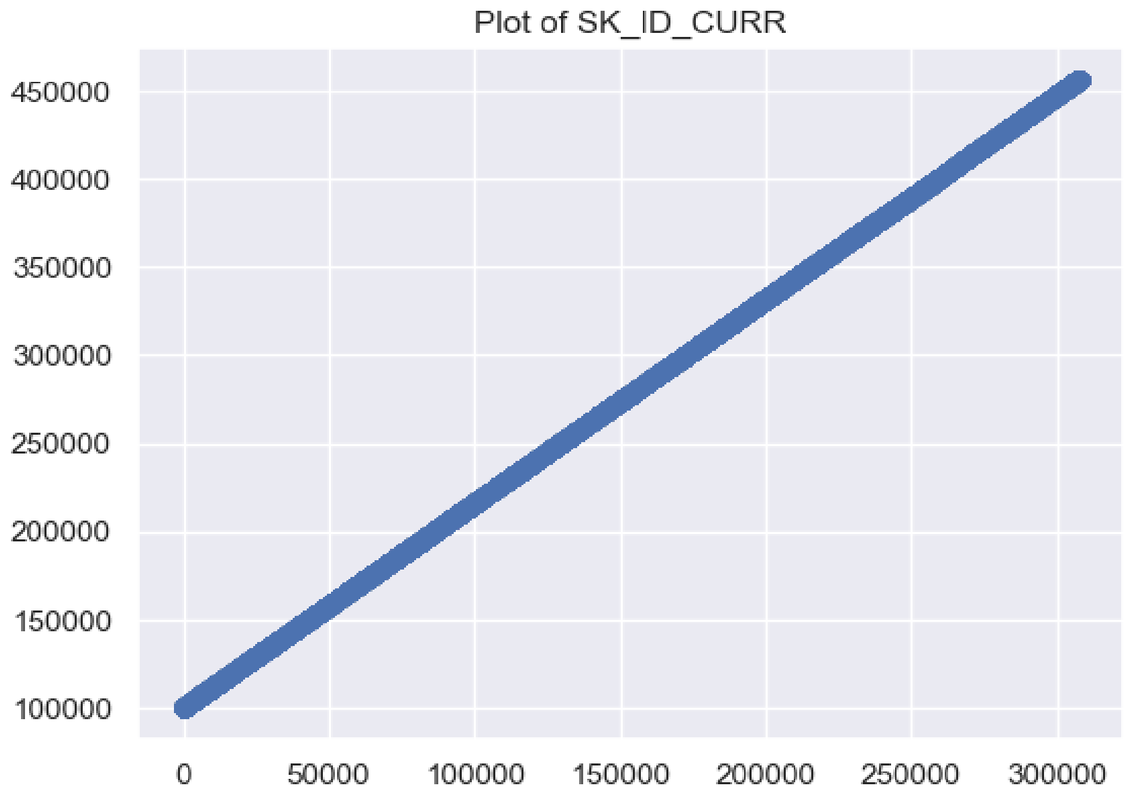


Missing Data Analysis:

- Methodology: Identified and managed missing data via column removal or substitution.
- Removed columns with >50% missing data, replaced others with calculated values.
- Example of high-missing columns: COMMONAREA_MEDI, FLOORSMIN_AVG, EXT_SOURCE_1

Outlier Analysis:

- Purpose: Identify extreme data points without removal.
- Outliers Found: Attributes like AMT_INCOME_TOTAL, AMT_CREDIT, and DAYS_EMPLOYED exhibit outliers.



Data Imbalance and Target Variable Analysis:

- Imbalance Ratio: 11.4:1 between clients with payment difficulties and others.
- Methodology: Analyzed using percentage and count-based visualizations.
- Suggested univariate and bivariate plots to highlight imbalance.

Key Univariate Analysis Insights:

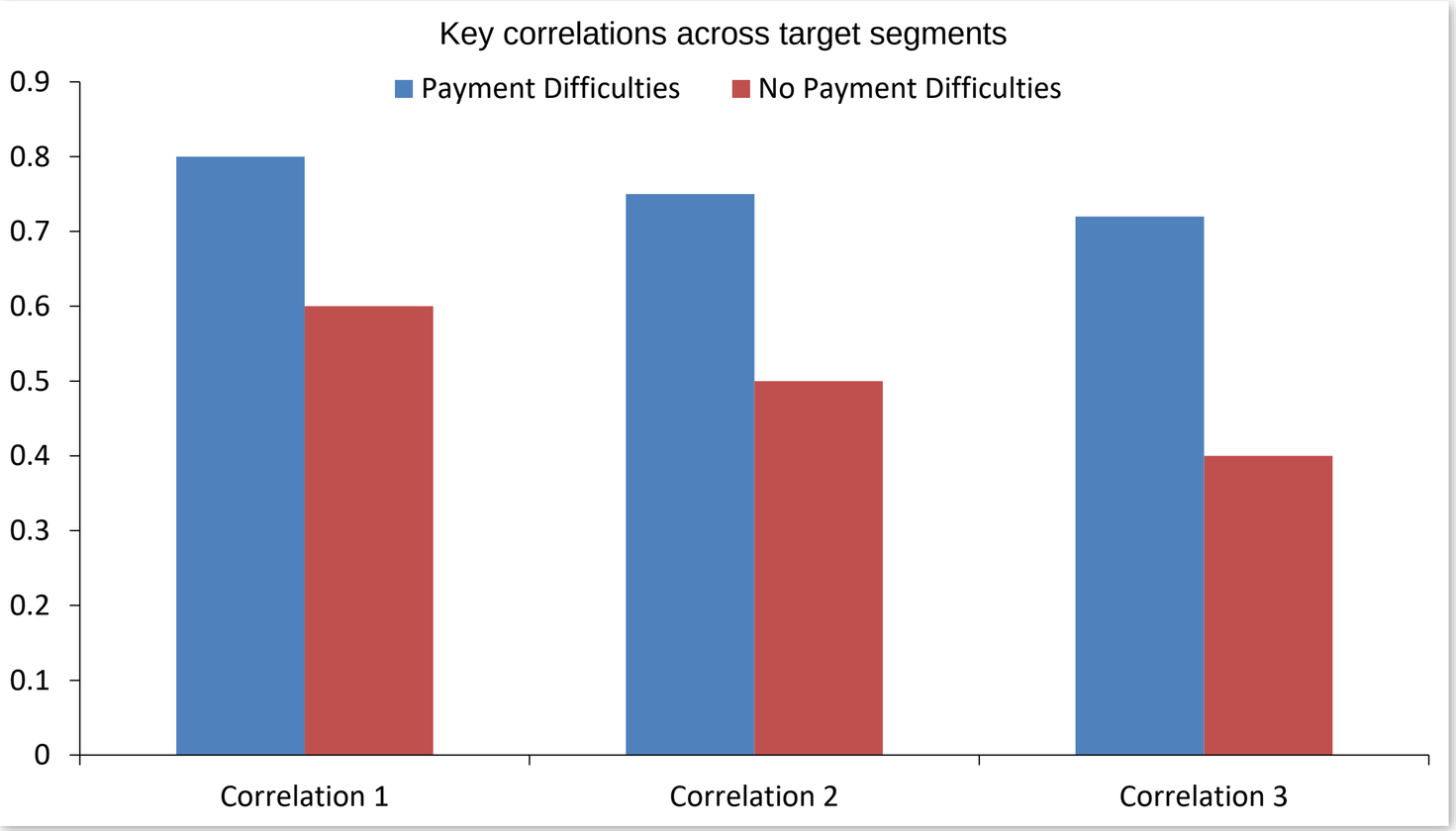
- Highlights: Important variables impacting client segmentation (e.g., AMT_INCOME_TOTAL, DAYS_BIRTH, EXT_SOURCE_2).
- Focus Variables: AMT_CREDIT, DAYS_BIRTH, and EXT_SOURCE_2.
- Insights: Age and income distributions reveal higher risk associations.

Bivariate Analysis Insights:

- Objective: Reveal variable relationships and their impact on credit risk.
- Key Findings: How variables like NAME_EDUCATION_TYPE and AMT_CREDIT relate to the target variable.
- Segmentation Insight: Risk patterns vary significantly between TARGET=1 and TARGET=0.

Top Correlation Analysis:

Top Correlation Analysis



- Objective: Identify top 10 correlations within each target group.
- Findings: For clients with payment difficulties, correlations among financial indicators; broader correlations in other clients.

Conclusion

- Summary: Key patterns identified to differentiate risk levels.
- Next Steps: Use insights to develop predictive models or enhance lending criteria.
- Recommendations: Implement focused risk assessments on younger clients with high external income scores.