CREDIT EDA

Prepared by KiranKumar H A

PROBLEM STATEMENT

Business Understanding

Loan providers often face challenges in assessing loan applications, particularly when applicants have insufficient or nonexistent credit histories. Some individuals exploit this situation by defaulting on loans. Our focus is on a consumer finance company specializing in urban lending.

When the company receives a loan application, it must make a crucial decision: whether to approve the loan. This decision carries two risks:

- **Risk 1**: If the applicant is likely to repay the loan but it's not approved, the company loses potential business.
- **Risk 2**: If the applicant is unlikely to repay, approving the loan may result in financial loss for the company.

PROBLEM STATEMENT

EDA Objectives

Our case study aims to identify patterns that signal clients' difficulties in paying their loan installments. This insight can guide actions such as:

- Denying loans to high-risk applicants.
- Reducing the loan amount for certain clients.
- Lending to risky applicants at higher interest rates.

Ultimately, we seek to ensure that clients capable of repaying their loans are not rejected. In essence, our objective is to understand the factors driving loan defaults, benefiting both the company and its clients.

OUR APPROACH

- Problem Understanding

  We focused on understanding the context and the critical target variable.
- Data Acquisition

  We started by acquiring the bank loan application dataset.
- Initial Data Assessment

  We assessed the dataset's structure and identified missing values.
- Data Cleaning

  We ensured data quality through meticulous cleaning.
- Data Imbalance Analysis

  We highlighted the significant data imbalance issue.

OUR APPROACH

- Data Transformation

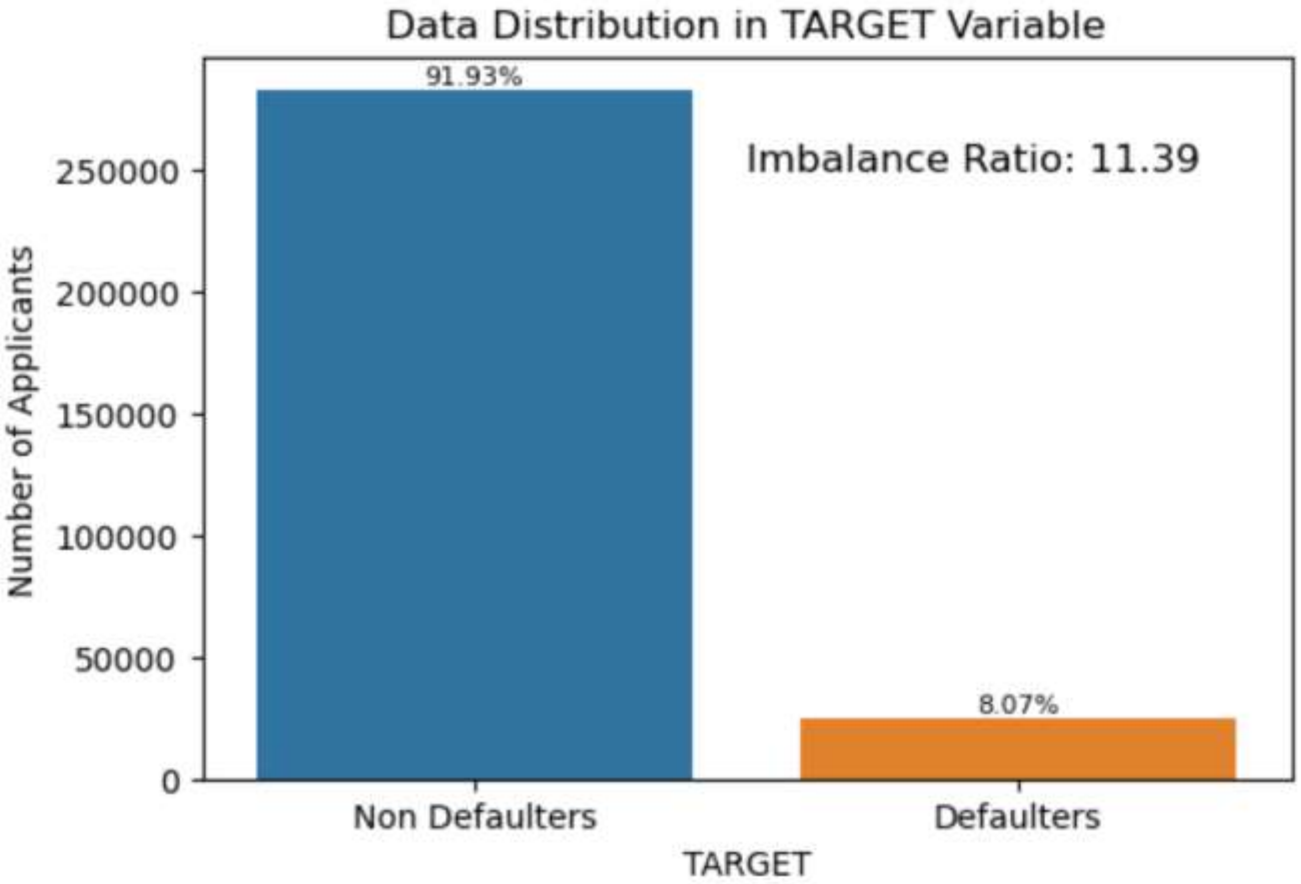
  We applied necessary transformations for analysis.
- Univariate Analysis

  We explored individual variables and their characteristics.
- Bivariate Analysis

  We examined relationships between variables, especially with the target variable.
- Inference and Visualization

  We used visualizations for data interpretation.
- Insights & Conclusion

  Our EDA provided valuable insights for further analysis and decision-making.



DATA IMBALANCE OVERVIEW

In application dataset, we are analyzing bank loan application data, with a crucial target variable that classifies clients into two categories: **defaulters** (late payment) and **non-defaulters** (regular payment).

The dataset exhibits a significant **data imbalance** issue, with the following statistics:

Approximately **91.93%** of the applicants fall into the **non-defaulter** category, indicating clients who make regular payments.

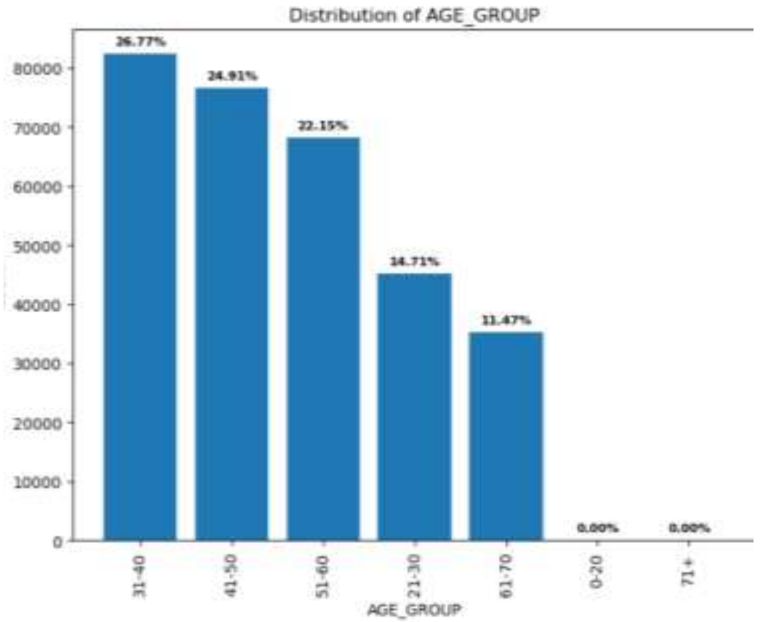
Conversely, the **defaulter** category, representing clients with late payments, comprises only **11.39%** of the dataset.

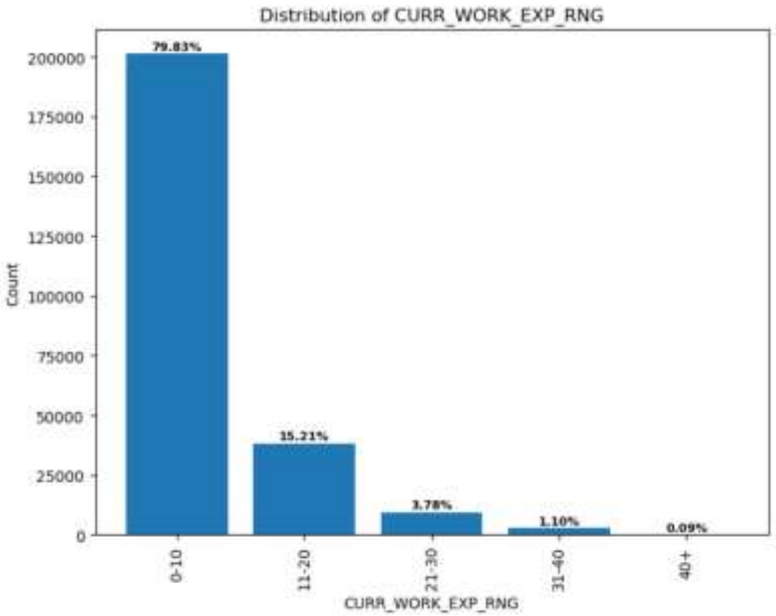
TOP 10 CORRELATION

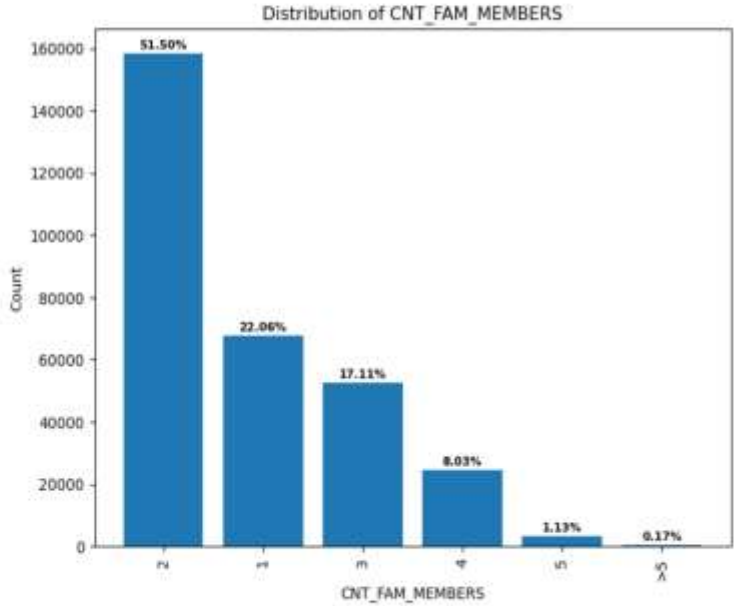
Defaulters		
Feature 1	Feature 2	Correlation
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.9983
AMT_GOODS_PRICE	AMT_CREDIT	0.9831
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.869
AMT_GOODS_PRICE	AMT_ANNUITY	0.7529
AMT_ANNUITY	AMT_CREDIT	0.7522
OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.3374
DEF_30_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.334
YEARS_REGISTRATION	YEARS_BIRTH	0.2891
DEF_60_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.2644
DEF_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.2612

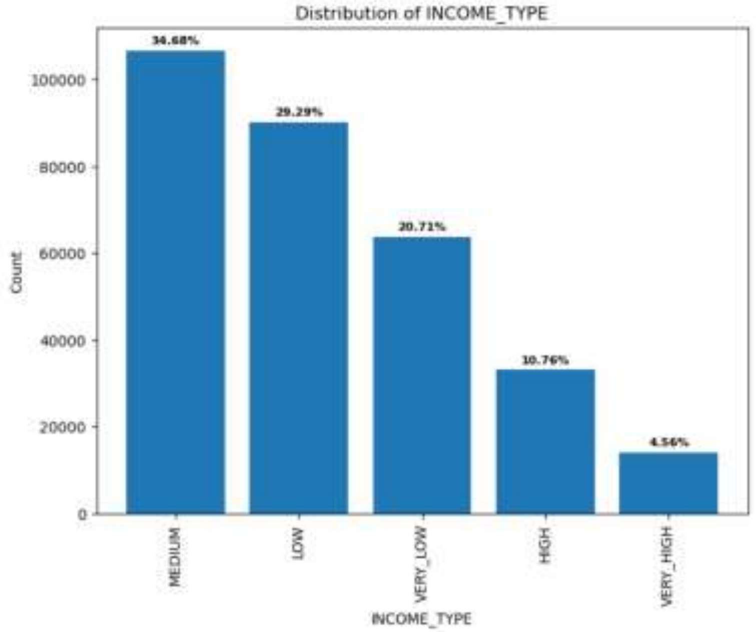
Non-Defaulters		
Feature 1	Feature 2	Correlation
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.9984
AMT_GOODS_PRICE	AMT_CREDIT	0.9873
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.8568
AMT_GOODS_PRICE	AMT_ANNUITY	0.7769
AMT_ANNUITY	AMT_CREDIT	0.7713
AMT_ANNUITY	AMT_INCOME_TOTAL	0.4189
AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.3497
AMT_CREDIT	AMT_INCOME_TOTAL	0.3428
YEARS_REGISTRATION	YEARS_BIRTH	0.3331
OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.307

We can observe that there is a strong correlation between below 5 pairs:





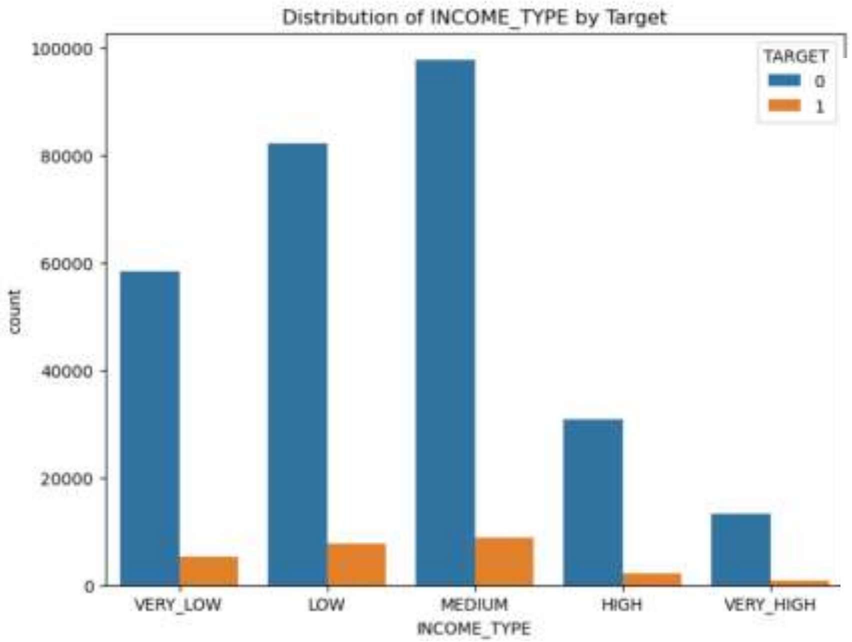


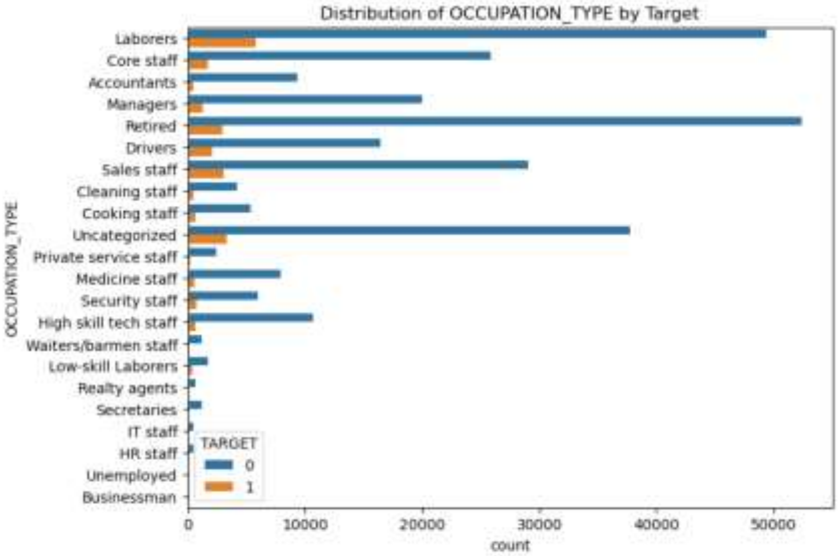


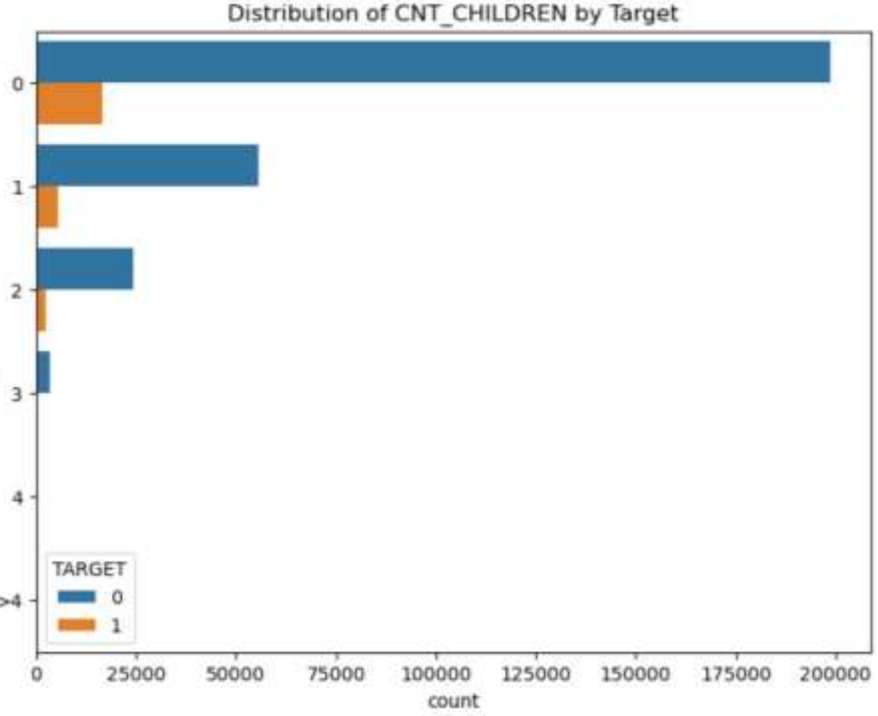
LOAN APPLICATION SUBMISSIONS

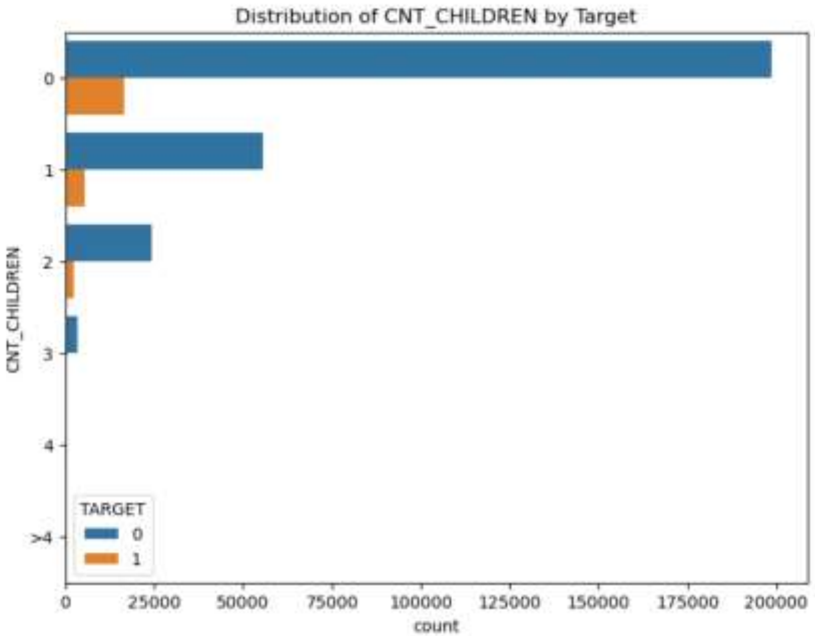
Below categories are more likely to apply loans –

- With very low to medium income levels
- With current job tenures within 0 to 10 years
- Households with two ,as the family size increases, the application rate decreases
- In the age range of 30 to 60 years



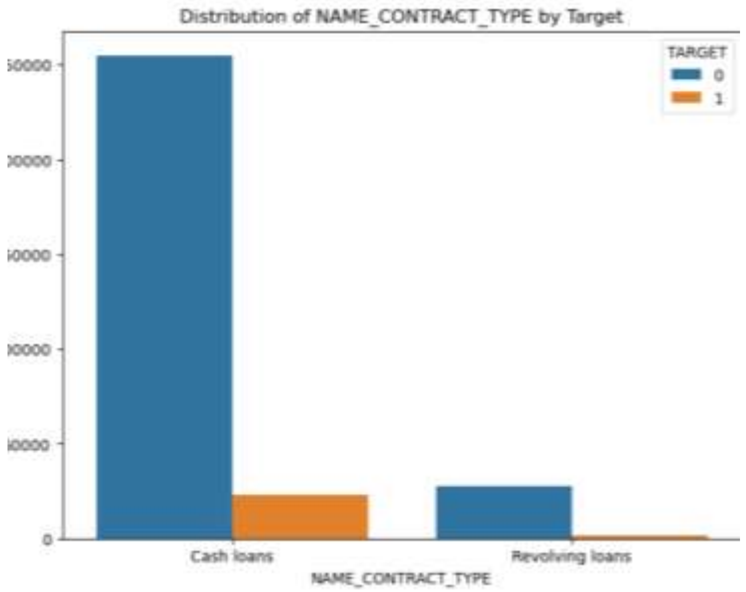


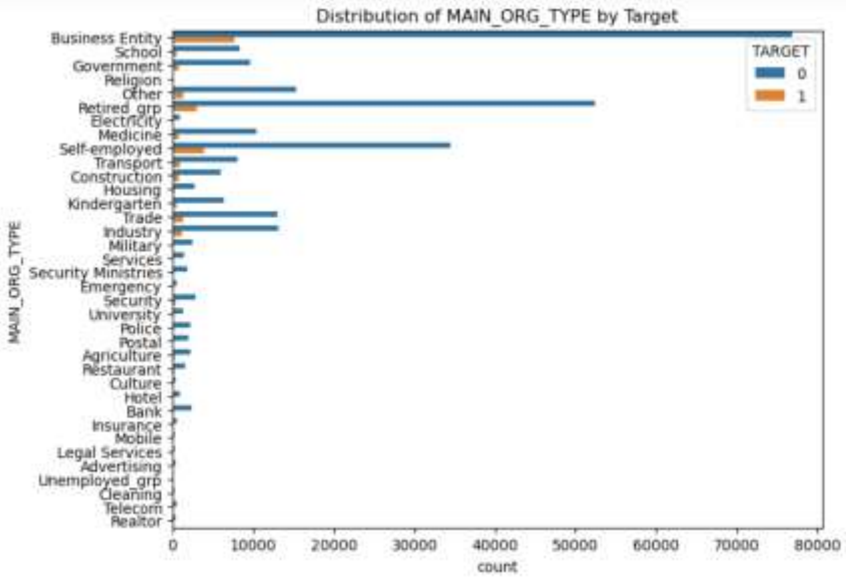


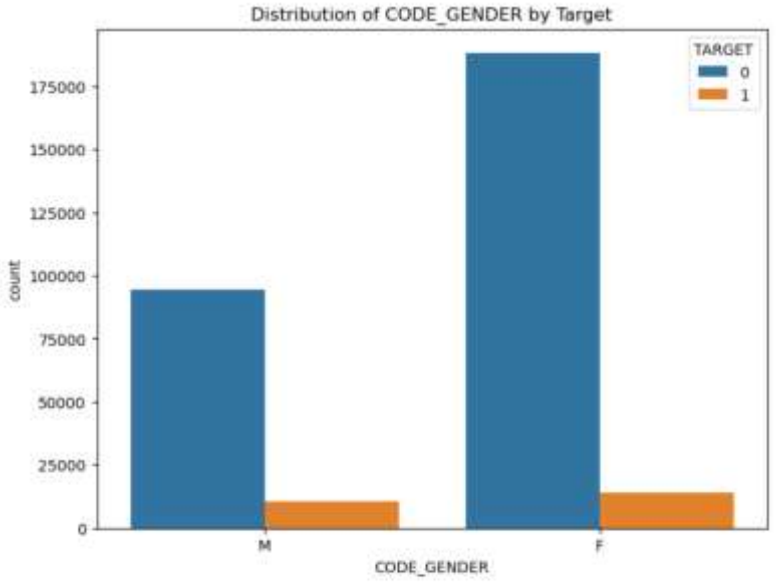


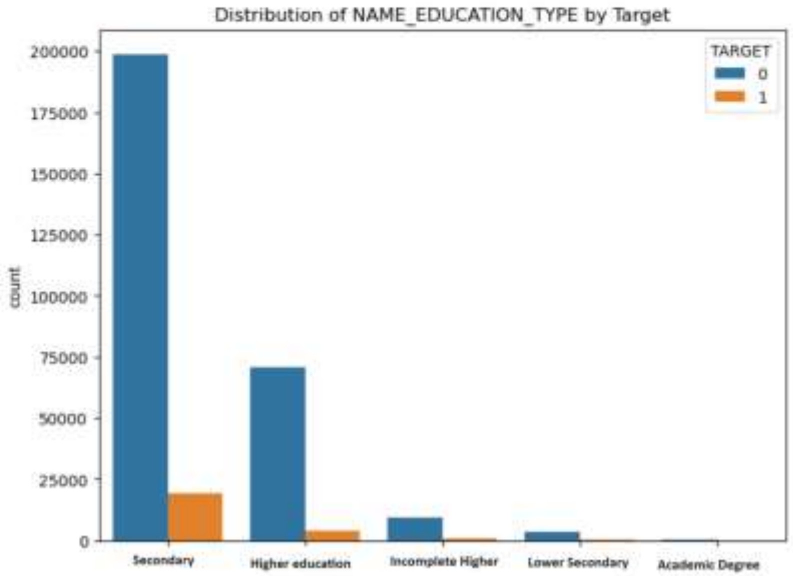
RECOMMENDATIONS

- Borrowers with a current employment duration of more than 40 years tend to have fewer late loan payment defaults.
- Individuals with income types categorized as Business and Student but have very few applications and have never defaulted on any payments.
- Clients who are Commercial Associates and Pensioner are good segment to give loans
- Clients who are Accountants by Occupation are less likely to default
- Borrowers with higher incomes are less likely to experience late loan payment defaults.
- Can prioritize for the client with less children and family member count









PRECAUTIONS

Males are more prone to late loan payment defaults compared to females.

Individuals aged 21 to 40 are more prone to late loan payment defaults compared to other age groups.

Loan defaults are more common in cash loans compared to revolving loans.

Working individuals and commercial associates are more prone to late loan payment defaults compared to pensioners and state servants.

Borrowers with lower secondary, secondary special, and incomplete higher education levels are more prone to defaults, while those with higher education levels are less likely to default.

THANK YOU