

EDA ASSIGNMENT

- **Title:** Exploratory Data Analysis of Credit Default Data
- **Your Name:** Mitesh Jaware

INTRODUCTION

Objective:

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

Datasets Used:

- `application_data.csv`
- `previous_application.csv`



DATA CLEANING

Missing Data Handling:

- Columns with >40% missing values were dropped.
- Null values in credit bureau-related columns replaced with 0.

Occupation Data:

- Missing values removed from both target datasets.

Column Removal:

- Irrelevant or excessive missing data columns removed.

Gender Column:

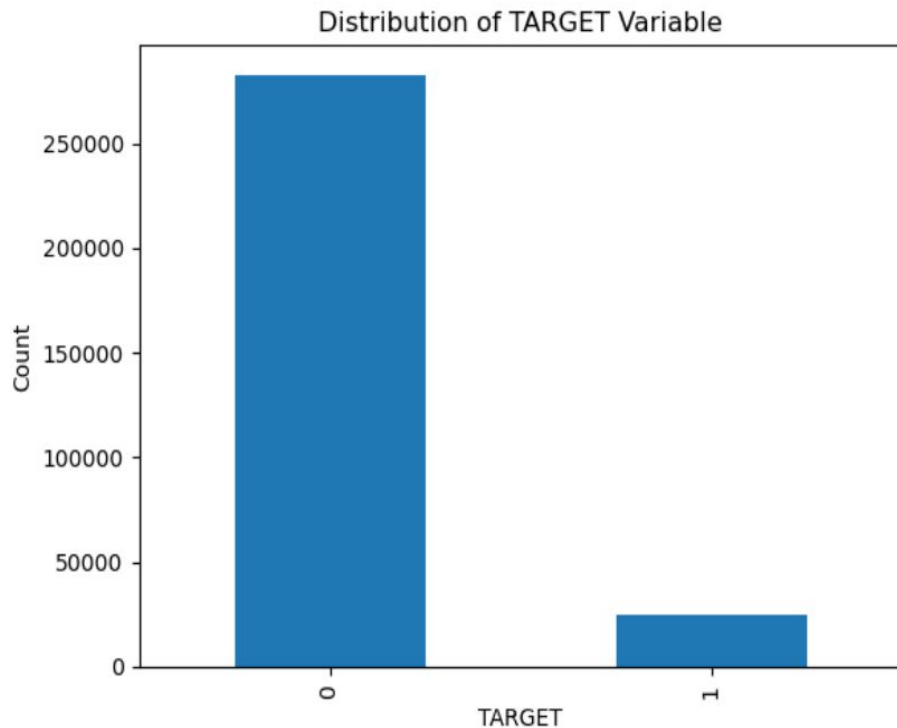
- Replaced 'XNA' with 'F'.



Univariate Analysis - Target Distribution

Insights:

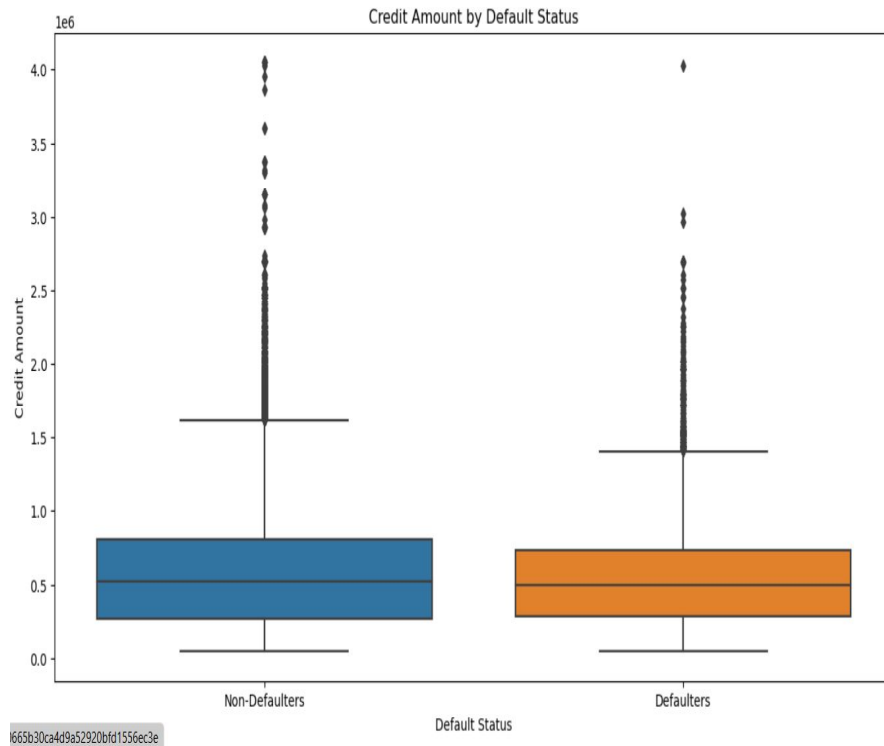
- Number of clients with and without payment difficulties.
- Imbalance observed between defaulters and non-defaulters.



Univariate Analysis - Credit Amount

Insights:

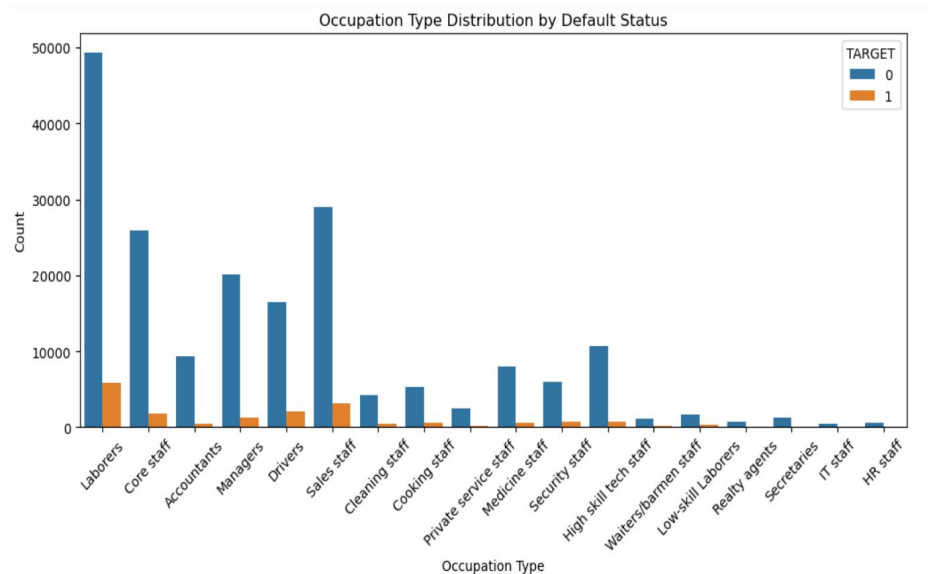
- Defaulters tend to have higher credit amounts.
- Significant spread in both defaulters and non-defaulters.



Univariate Analysis - Occupation Type

Insights:

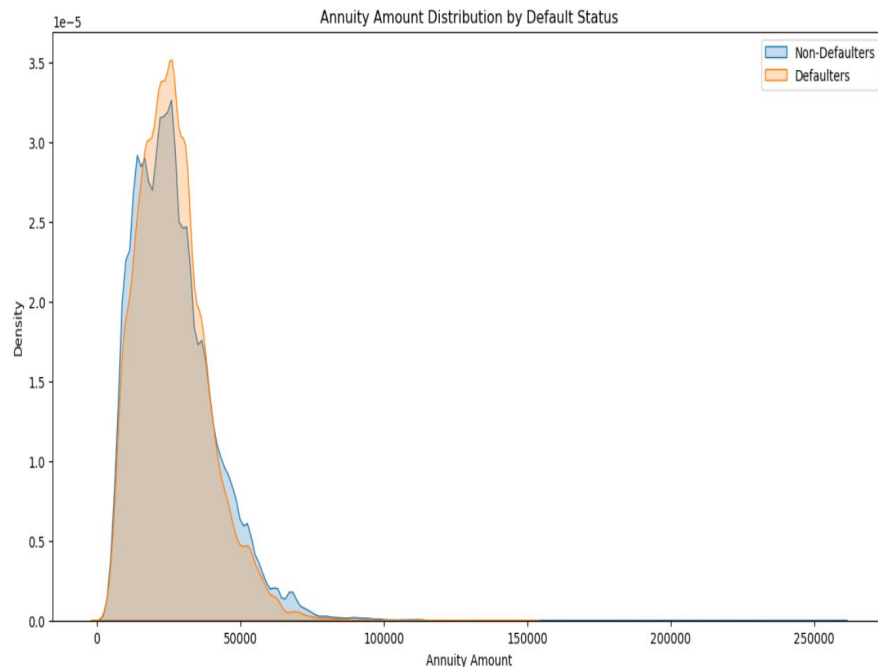
- Distribution of occupations among defaulters and non-defaulters.
- Some occupations like Laborers, core staff etc. are more prevalent among defaulters.



Univariate Analysis - Annuity Amount

Insights:

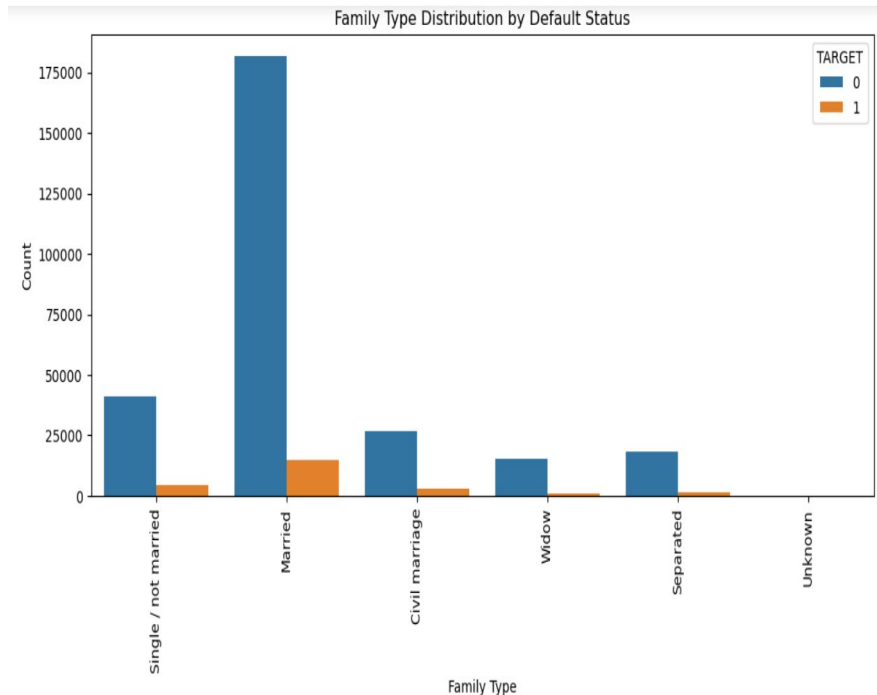
- Overlapping distributions for annuity amounts.
- Differences in density between defaulters and non-defaulters.



Univariate Analysis - Family Status

Insights:

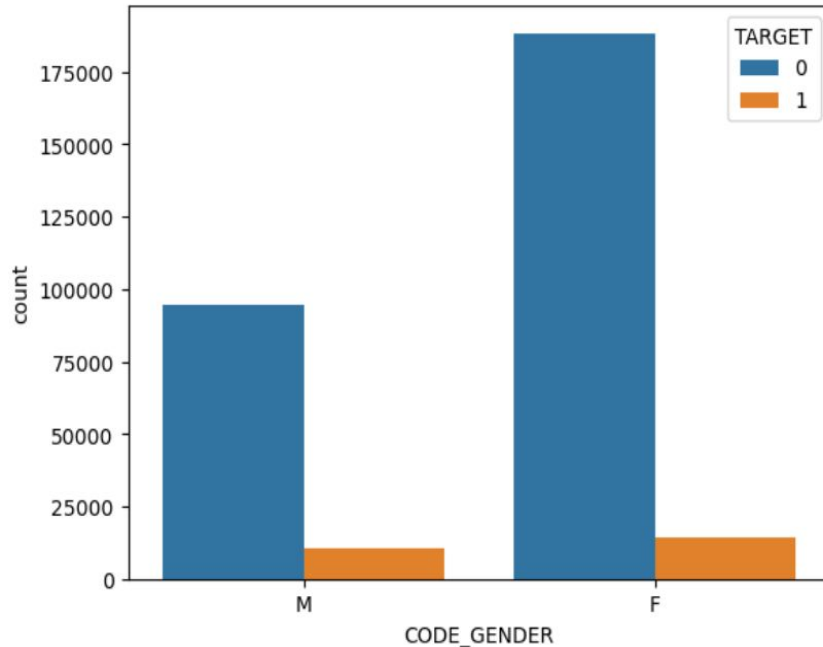
- Distribution of family statuses across default statuses.
- Certain family types are more common among defaulters.



Univariate Analysis - Gender Distribution

Insights:

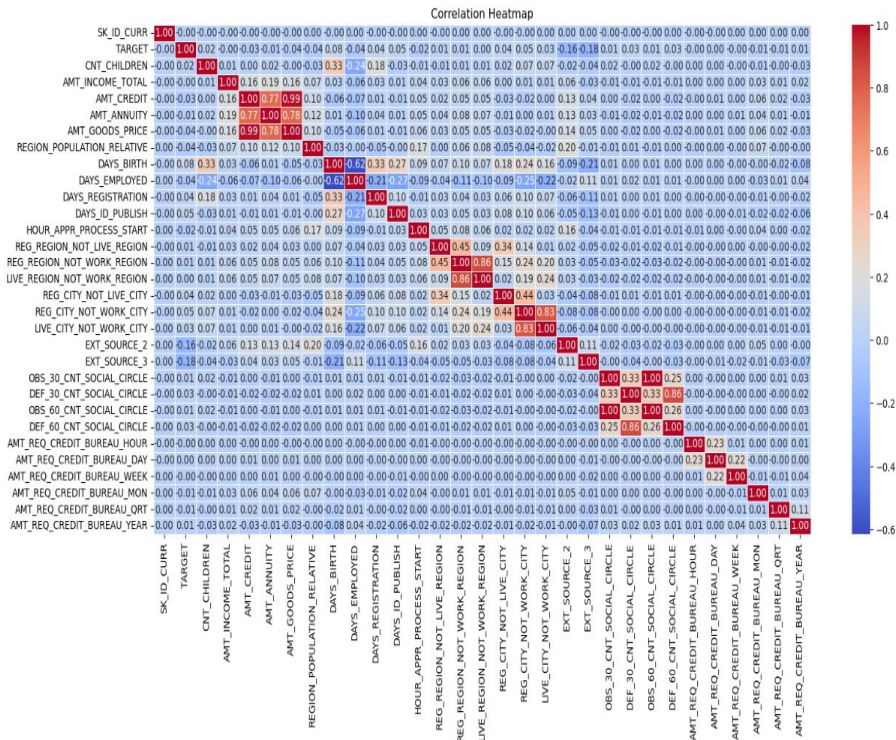
- Gender distribution among defaulters and non-defaulters.



Bivariate Analysis - Correlation Matrix

Insights:

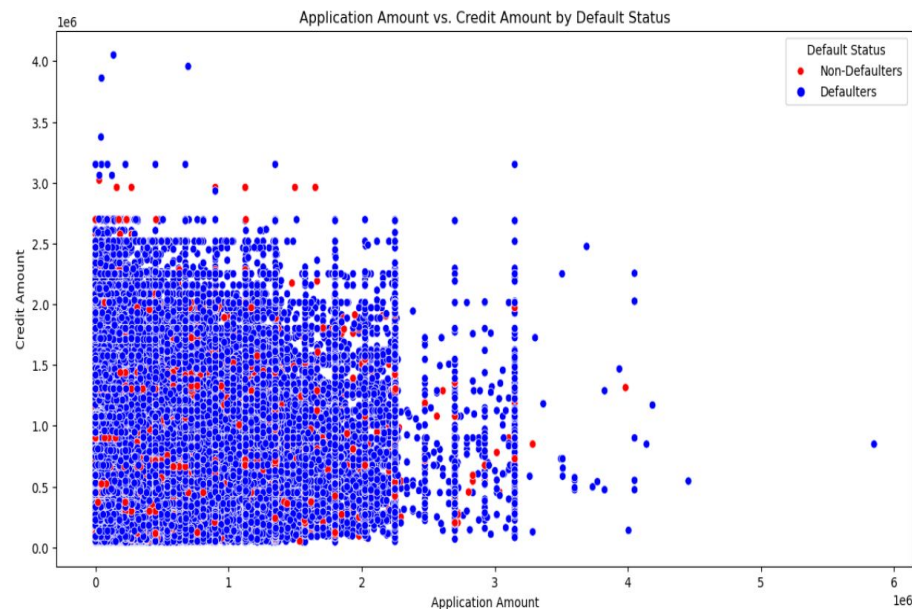
- Significant correlations between numerical features.
- Potential for multicollinearity.



Bivariate Analysis - Application Amount vs. Credit Amount

Insights:

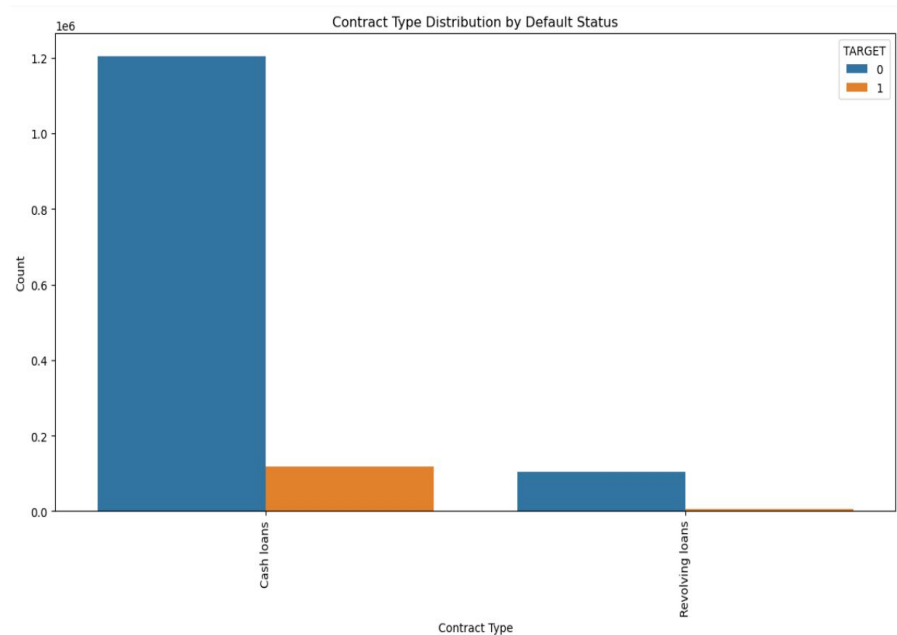
- Relationship between application and credit amounts.
- Differences between defaulters and non-defaulters.



Bivariate Analysis - Contract Type

Insights:

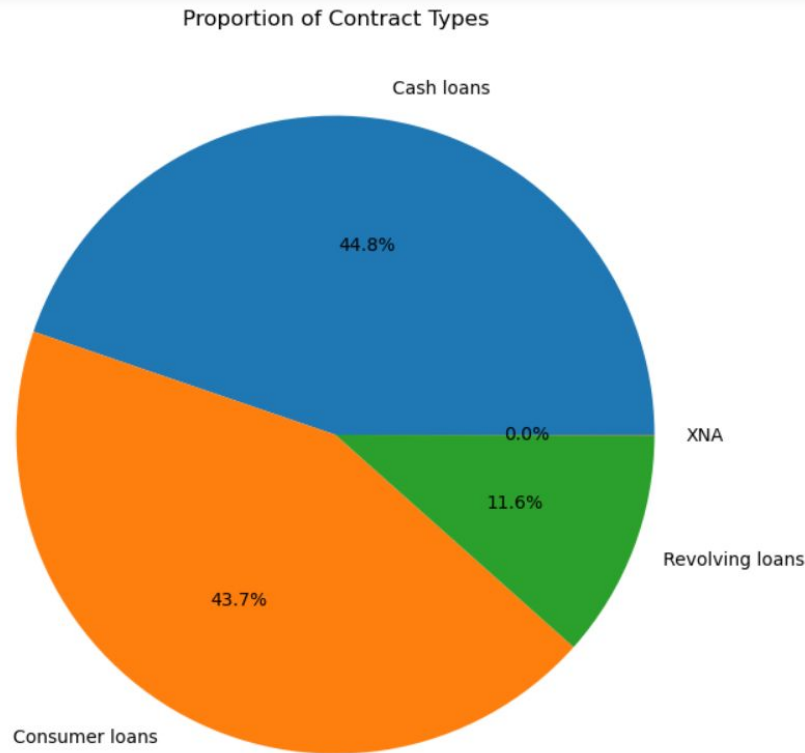
- Distribution of contract types among defaulters and non-defaulters.



Previous Application Data - Contract Types

Insights:

- Proportions of different contract types.



Conclusions

Key Findings:

- Defaulters tend to have higher credit amounts and different distributions for annuity amounts.
- Occupation type, family status, and gender show varied distributions across default statuses.
- Significant correlations among features that could impact modeling.
- Insights from previous application data provide additional context.

