Credit EDA Assignment

By Mohammed Muddasir

Introduction

- This assignment aims to give you an idea of applying EDA in a real business scenario. In this assignment, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Understanding

- Insufficient or non-existent credit history poses a challenge for loan providers in assessing creditworthiness and increases the risk of defaults.
- Some consumers may take advantage of this situation by intentionally defaulting on loans.
- The consumer finance company specializes in lending loans to urban customers.
- Exploratory Data Analysis (EDA) is being used to analyze patterns in the data.
- The goal of the analysis is to identify patterns that can help distinguish applicants who are capable of repaying the loan, thereby reducing the chances of rejecting potentially creditworthy individuals.

Business objectives

- The case study focuses on identifying patterns that indicate if a client may have difficulty paying their loan instalments.
- The identified patterns will be used to make informed decisions such as denying the loan, reducing the loan amount, or offering loans at higher interest rates to risky applicants.
- The goal is to ensure that consumers capable of repaying the loan are not rejected while managing the risk associated with lending.
- Exploratory Data Analysis (EDA) will be used to identify the driving factors or variables that strongly indicate loan default.
- The company aims to use this knowledge for portfolio management and risk assessment, allowing for better decision-making in loan approvals. Researching risk analytics and understanding variable types and their significance is recommended to gain domain knowledge.

Application Dataset Analysis

CLEANING THE DATA :

Data Cleaning is done for preparing the data for analysis. The following steps are taken for Cleaning the data :

1. Firstly, we have found out which all columns are relevant columns for analysis.

2. Secondly , we have to Identify and locate any null values within the dataset.

3. Then begin with Dropping the columns where null values are more or equal to 40%

ANALYSIS AND DATA CLEANING

Data cleaning is a crucial step in preparing data for analysis. Here are the key steps involved:

1. Identify the columns essential for analysis

2. Detect and handle null values in the dataset. By Removing all columns with a null value

3. Remove columns with a high percentage of null values, typically set at 40% or above.

4. we are done with the identification of the outliers. We have removed them and plotted them again to observer the difference

5. These steps are essential for ensuring data quality and integrity, allowing for accurate and reliable analysis.

ANALYSIS OF APPLICATION DATA

we have proceeded with the analysis of data. Checking the TARGET Variable If we plot the count of the TARGET Variable, we observe that there is a huge imbalance in our Target
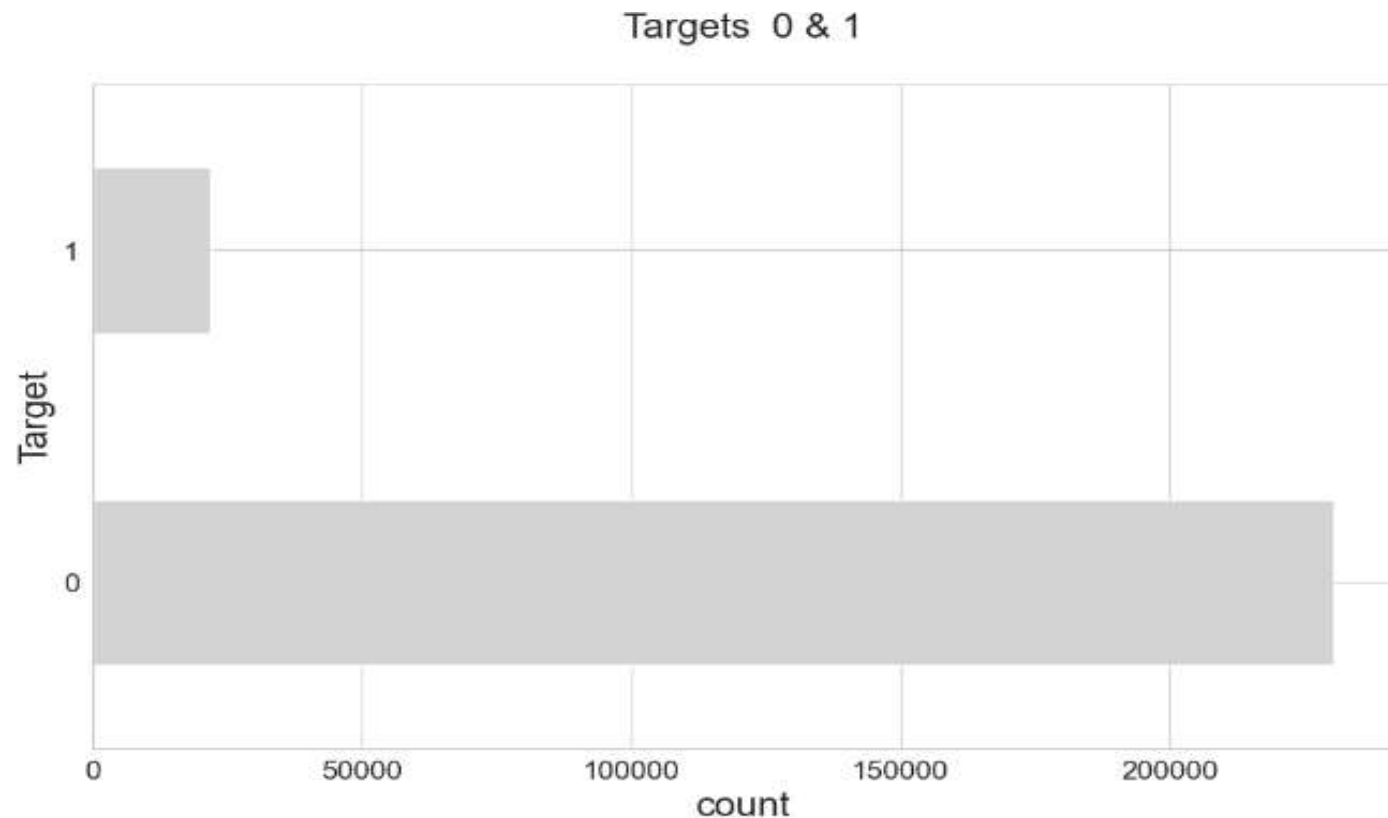
UNIVARIATE ANALYSIS

- Based on the graph, we can draw the following conclusions:
- Working women tend to have a higher credit count compared to other groups.
- Occupations such as 'State Servant', 'Working', and 'Commercial Associate' have higher credit counts compared to other occupations.
- Women on maternity leave have a lower credit count compared to other groups.
- These observations highlight the relationship between different variables and the credit counts, providing valuable insights into the data.
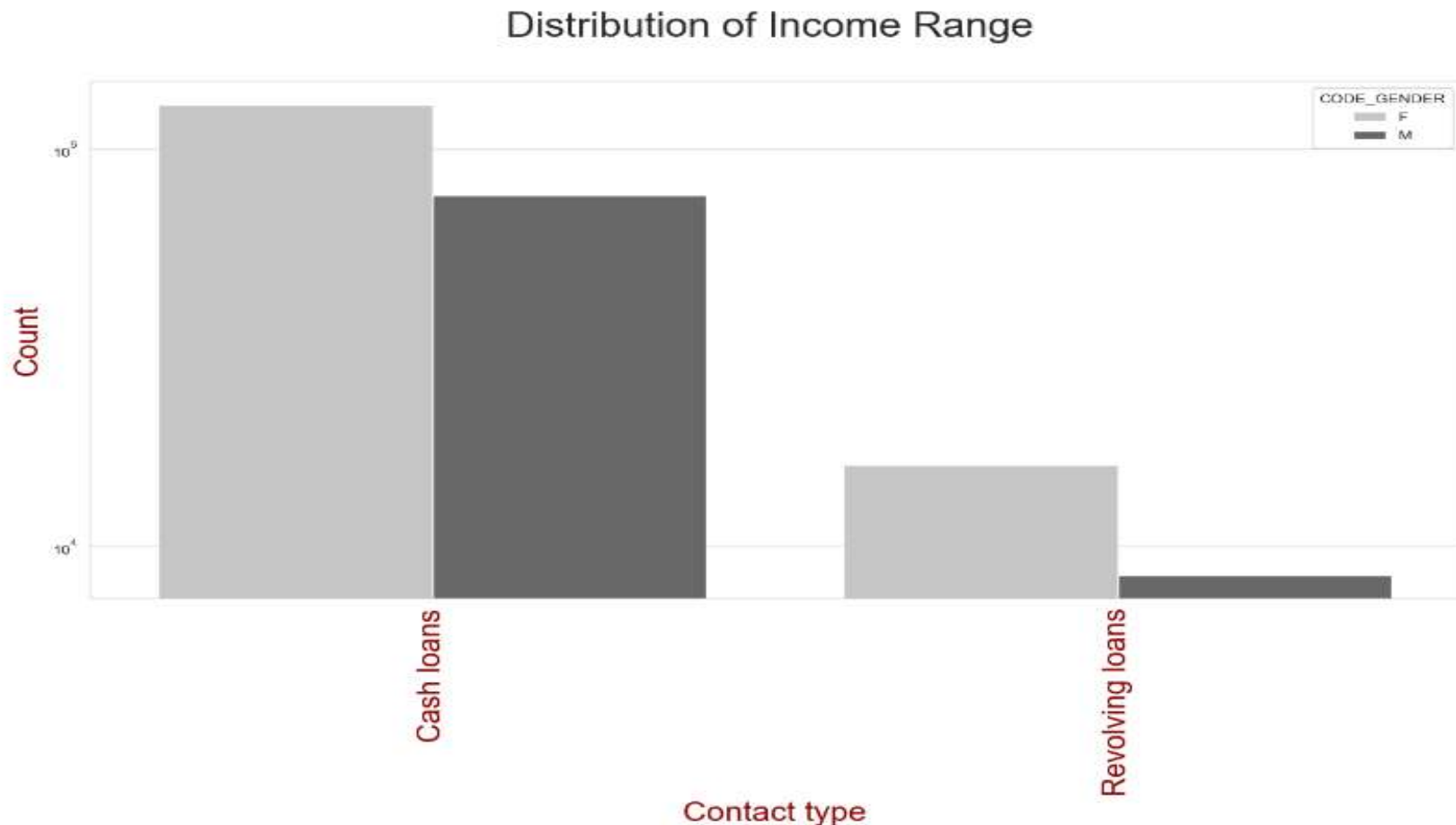
Distribution of Income Range
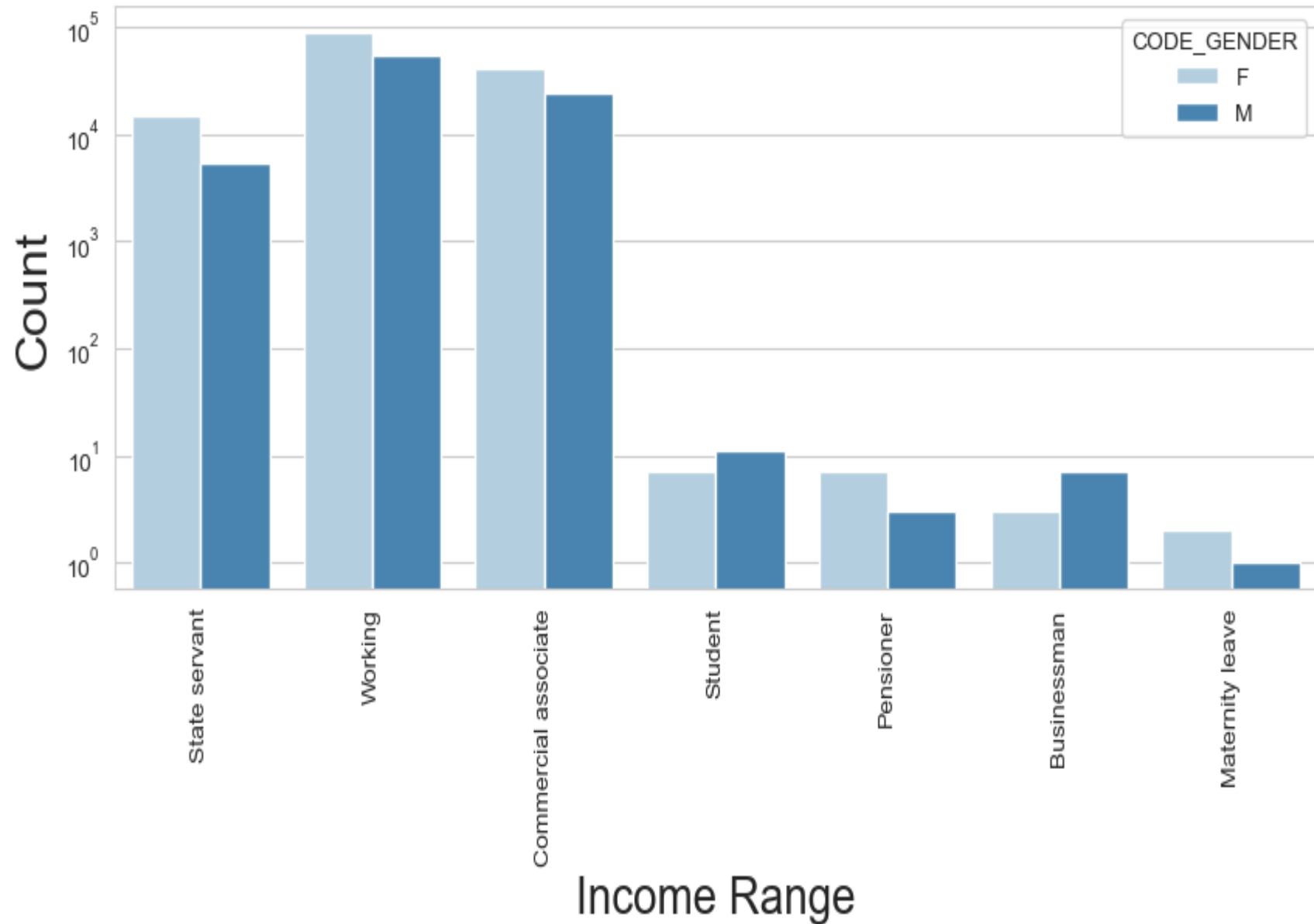
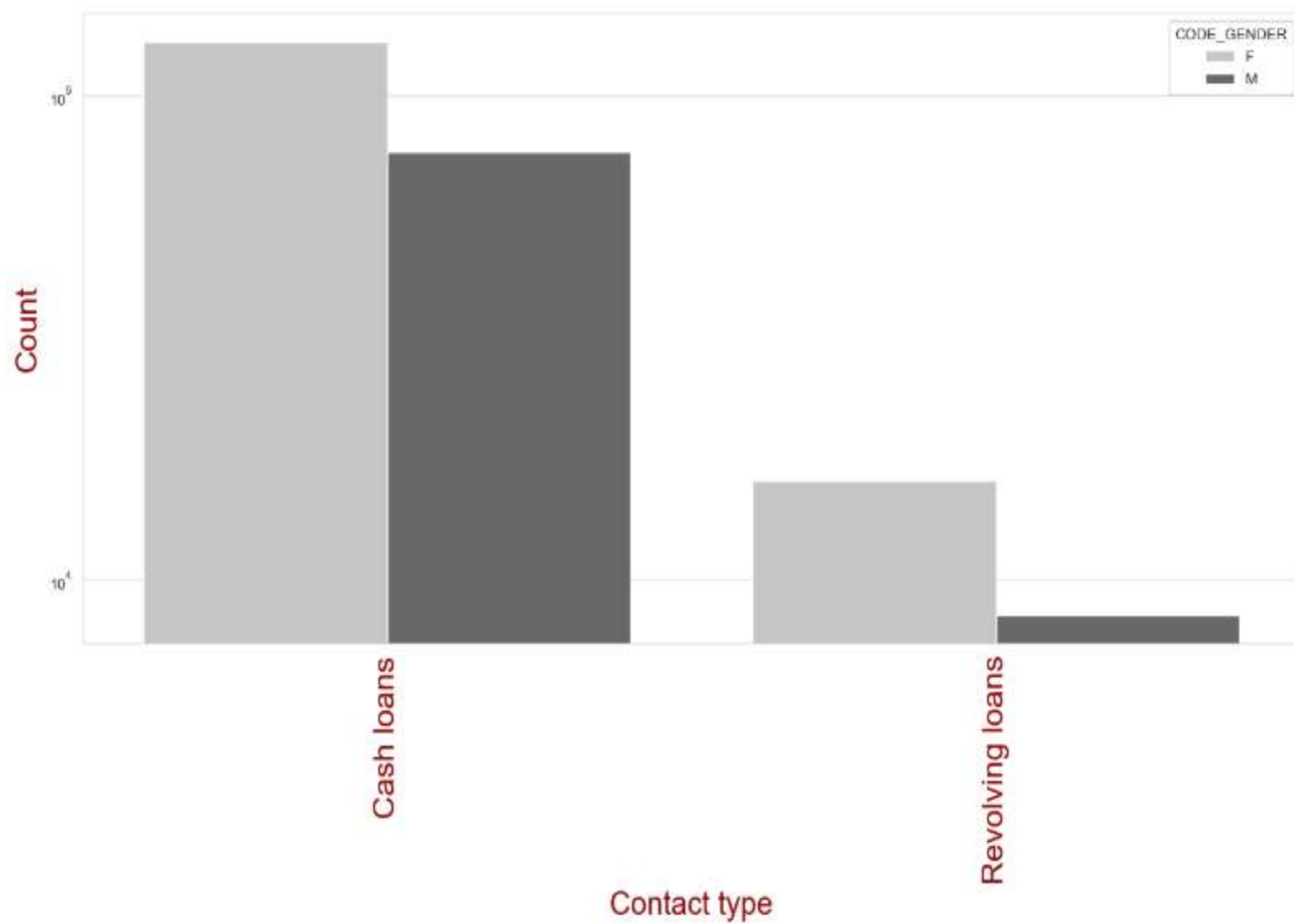Based on the graph, we can draw the following conclusions:

1. The 'cash loans' contract type has a higher number of credits compared to the 'revolving loans' contract type.
2. Females tend to apply for credit more frequently than males.
These observations highlight the disparities between different contract types and the gender distribution in credit applications, providing valuable insights into the data.


Distribution of Income Range

Distribution of Income Range

Finding the Outliers

Univariate analysis for Target =0

- Based on the graph, we can draw the following conclusions:

- The income amount of the clients appears to be evenly distributed,

- Additionally, there are some outliers present in the dataset, which are data points that's different from the majority of the data.

- These observations provide insights into the distribution of income amounts and the presence of outliers in the dataset.
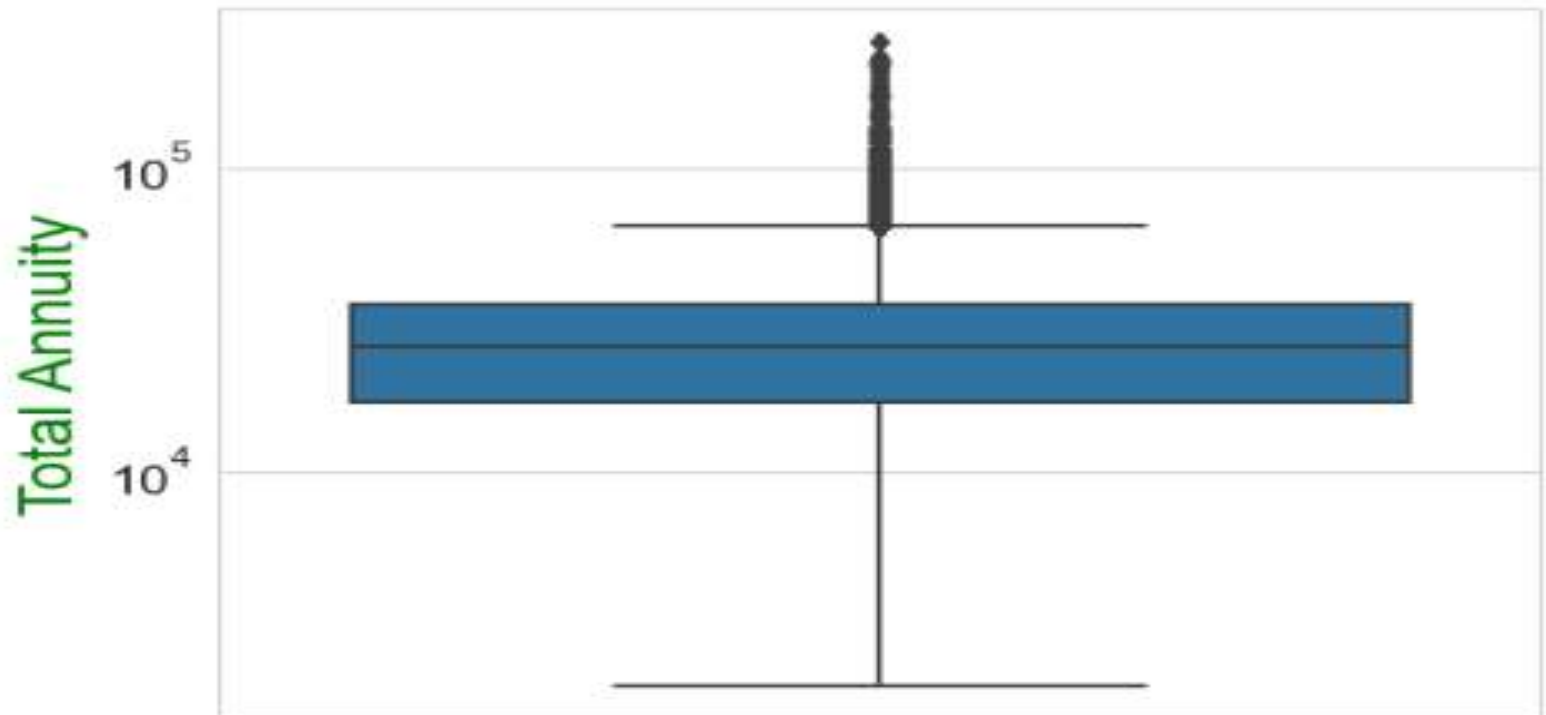
Distribution of Anrulty Amount
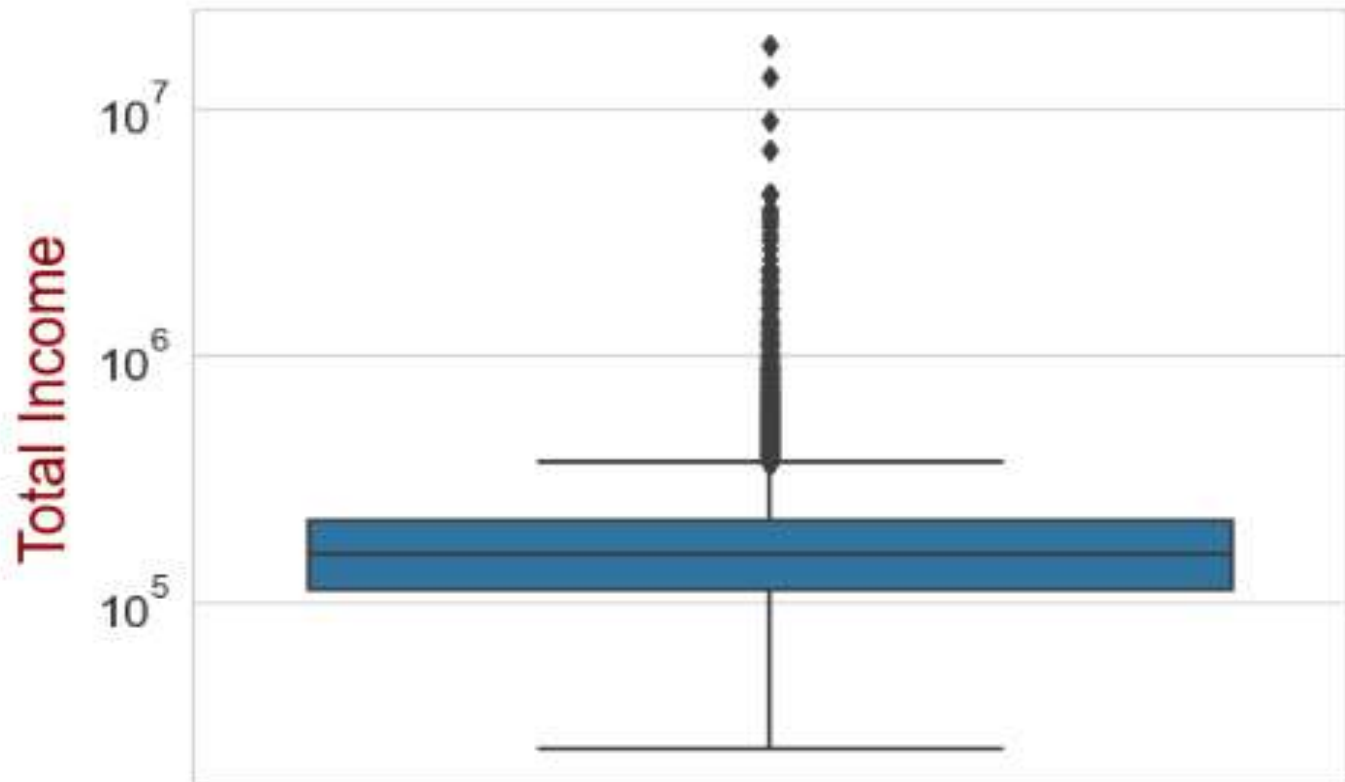
- Based on the graph, we can draw the following conclusions:

- The first quartile is bigger than the third quartile.
- There seems some outers in the Annuity boxplot.

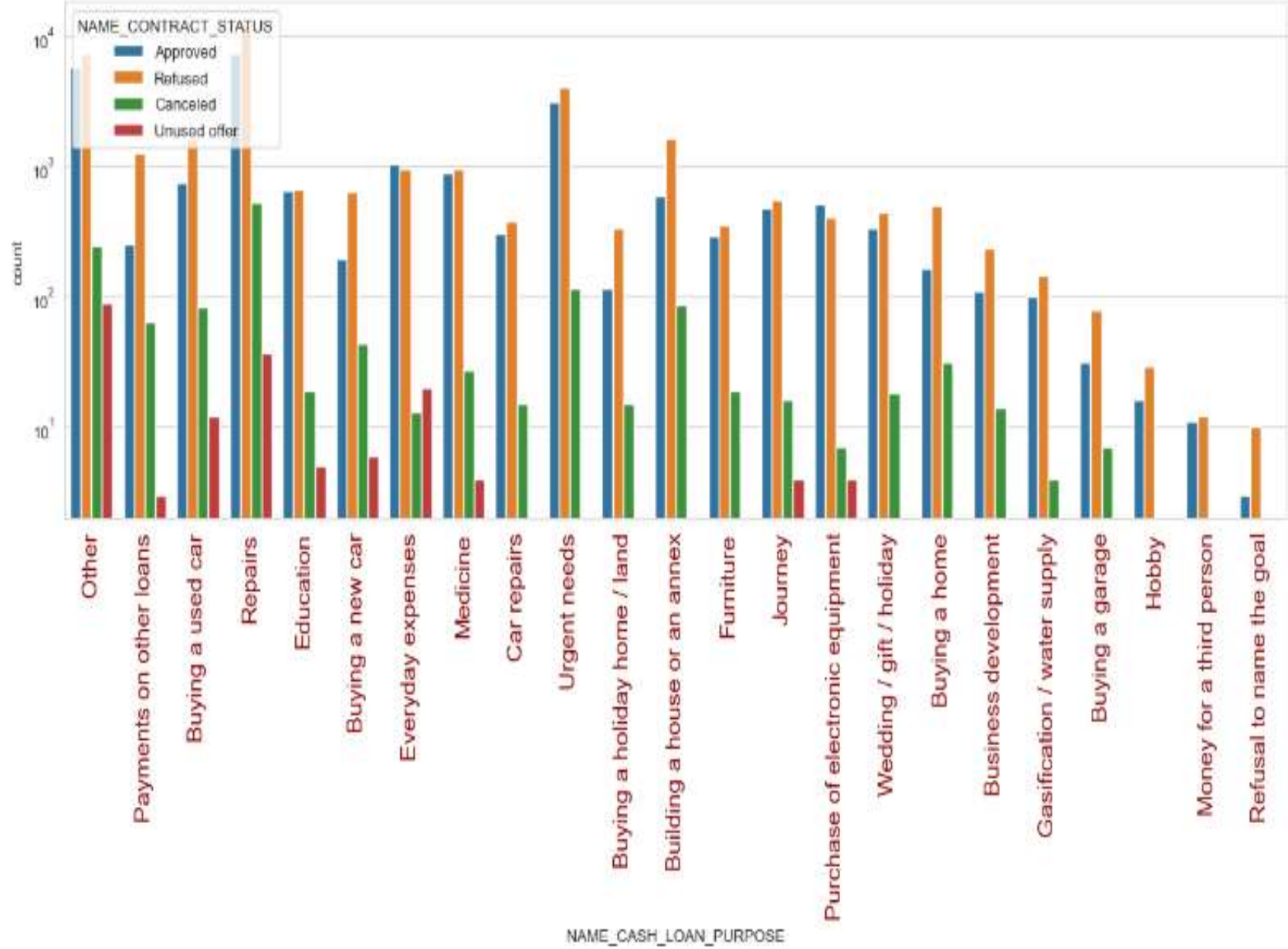Distribution of Anrulty Amount

Merging the Two datasets

Merging application_dataset and previous_application

- Based on the graph, we can draw the following conclusion:
- The majority of customers have applied for previous loans for the purpose of repairs, but it has the highest number of loan cancellations.

Based on the graph, we can draw the following conclusion:
most of the applications which were previously either canceled or refused 80-90% of them are repaire in the current data
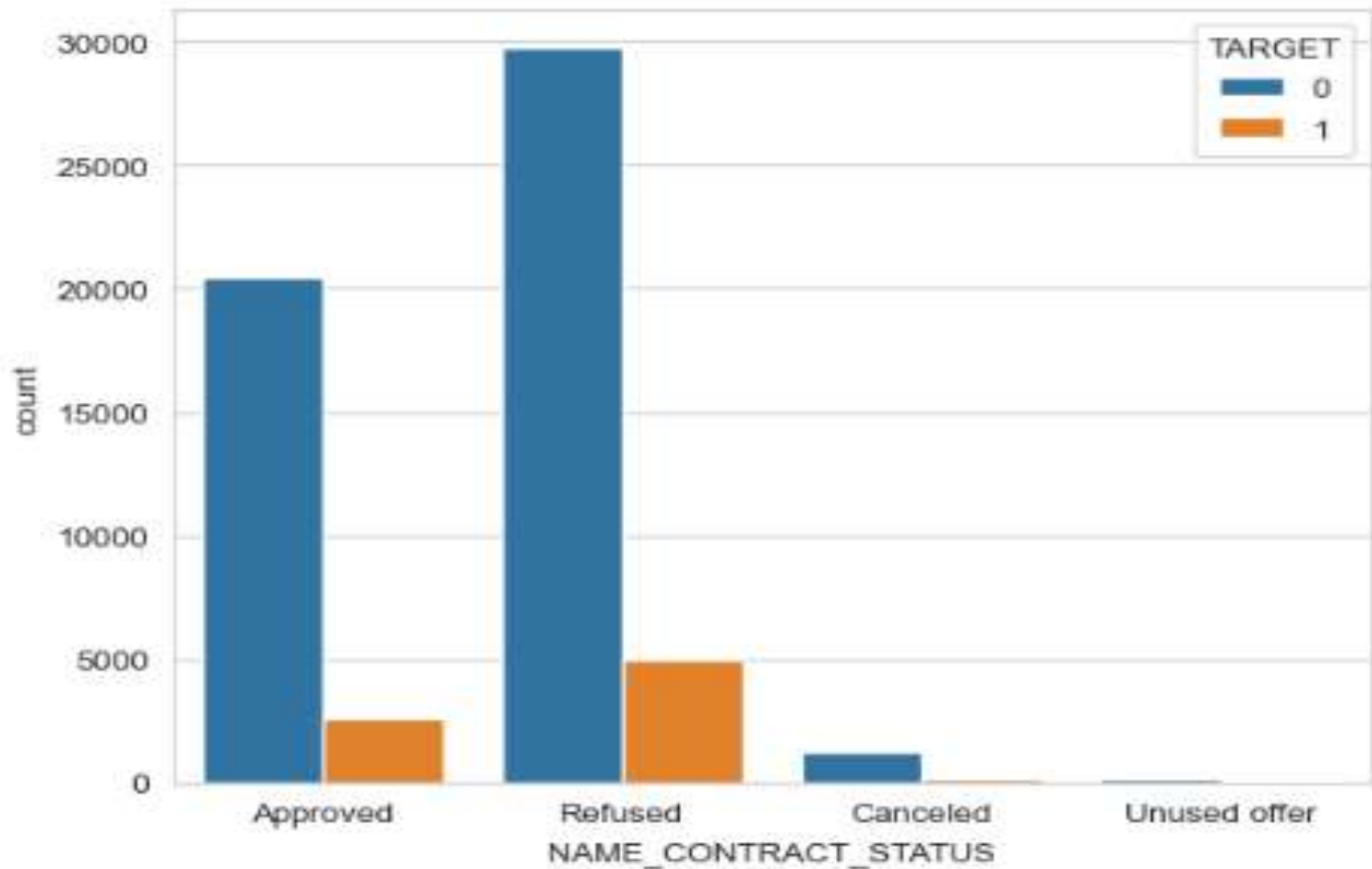
Final Conclusion :

- The majority of customers have taken cash loans.
- Customers who have taken cash loans are less likely to default.
- Most of the loans have been taken by females.
- The default rate for females is approximately 7%, which is lower and safer compared to males.
- The safest segments in terms of loan default are working individuals, commercial associates, and pensioners.

Thank you