Credit EDA Assignment

Mrs. Dhara Akshay Khamar

[DS 52 – Jan'23 Batch]

Introduction

- This case study aims to apply EDA in real business scenario of minimizing risk of losing money while lending money to customers

- Another objective is to develop basic understanding of risk analytics in banking and financial services.

Problem Statement

Two types of risks are associated with the bank's decision of either approving loan application or not :

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Business Objectives

- ❑ Identify patterns which indicate if a client has difficulty paying their instalments .
- ❑ These patterns may be used to for taking actions such as, denying the loan, reducing the amount of loan, lending at a higher interest rate, etc.
- ❑ The company wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default.
- ❑The company can utilize this knowledge for its portfolio and risk assessment.

Analysis Method

- Understanding of Current Loan Applications
- Data cleaning
  - Fixing rows / columns
  - Missing values handling
  - Outliers Analysis
  - Standardizing values
- Data Imbalance check
- Univariate, Segmented univariate, Bivariate analysis
- Identifying top Correlation variables for both Targets
- Analyzing Previous Loan Applications

Analysis Method | Data cleaning : Missing values & Outliers

❑ Missing values Handling

- ➢ Dropping columns having missing values more than 40 %
- ➢ Imputing Mean values where data distribution is fairly symmetrical
- ➢ Imputing Median values where data is either left or right skewed

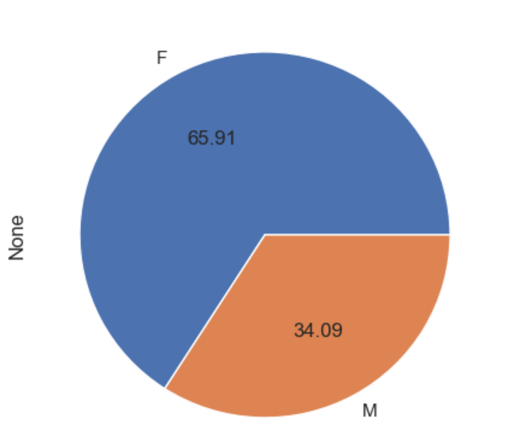
❑ Outliers Handling

- ➢ IQR based imputing for Right / Left skewed data distribution
- ➢ Numerical Continues data with large amount of outliers are Binned
- ➢ Trimming data if outliers are present in both directions

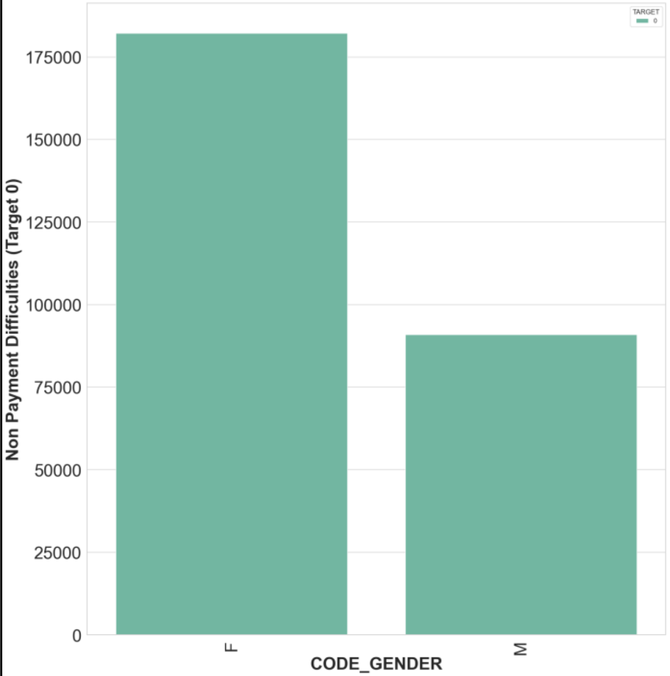
Univariate Analysis

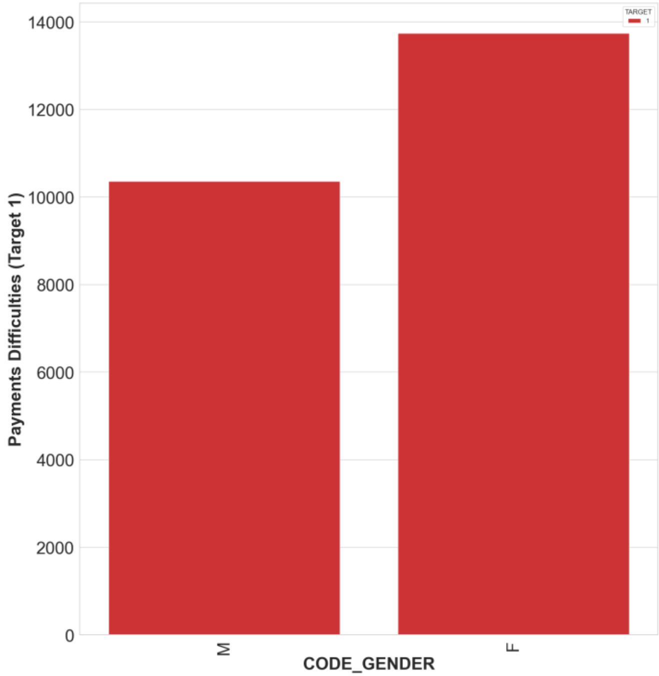
- **Variable** : Gender
- **Insights** :
  - ➢ Male vs. Female Applicants : ~66 % & ~34 % respectively
  - ➢ Male defaulters are higher compare to Female
  - ➢ (~11% and ~7% respectively)

Gender : Male **vs.** Female



Defaulters/Non-Defaulters **vs.** Gender

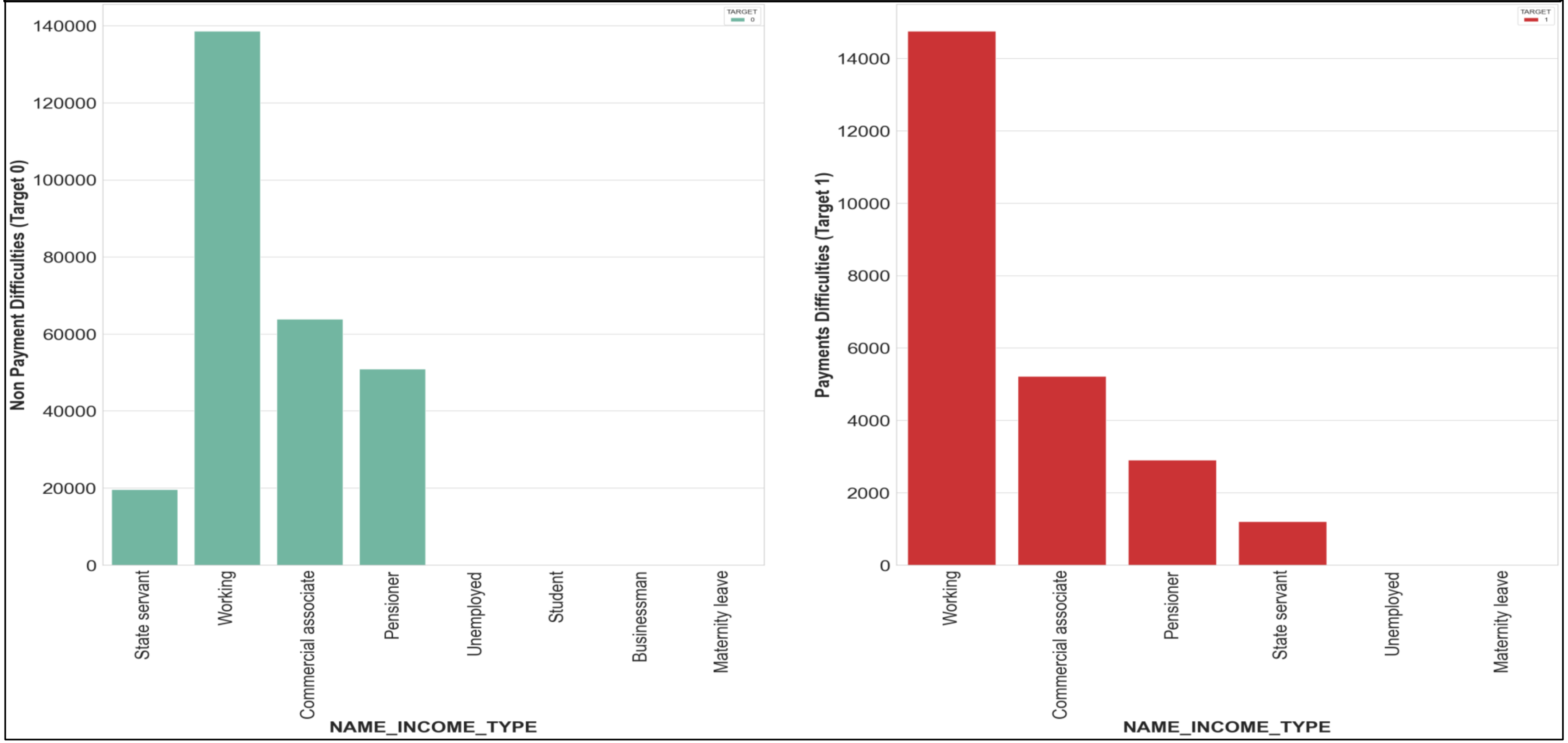




Univariate Analysis

- **Variable** : Profession

Profession **vs.** Default counts



- **Insights** :

➢ Working profession has slight more defaulters compare to others

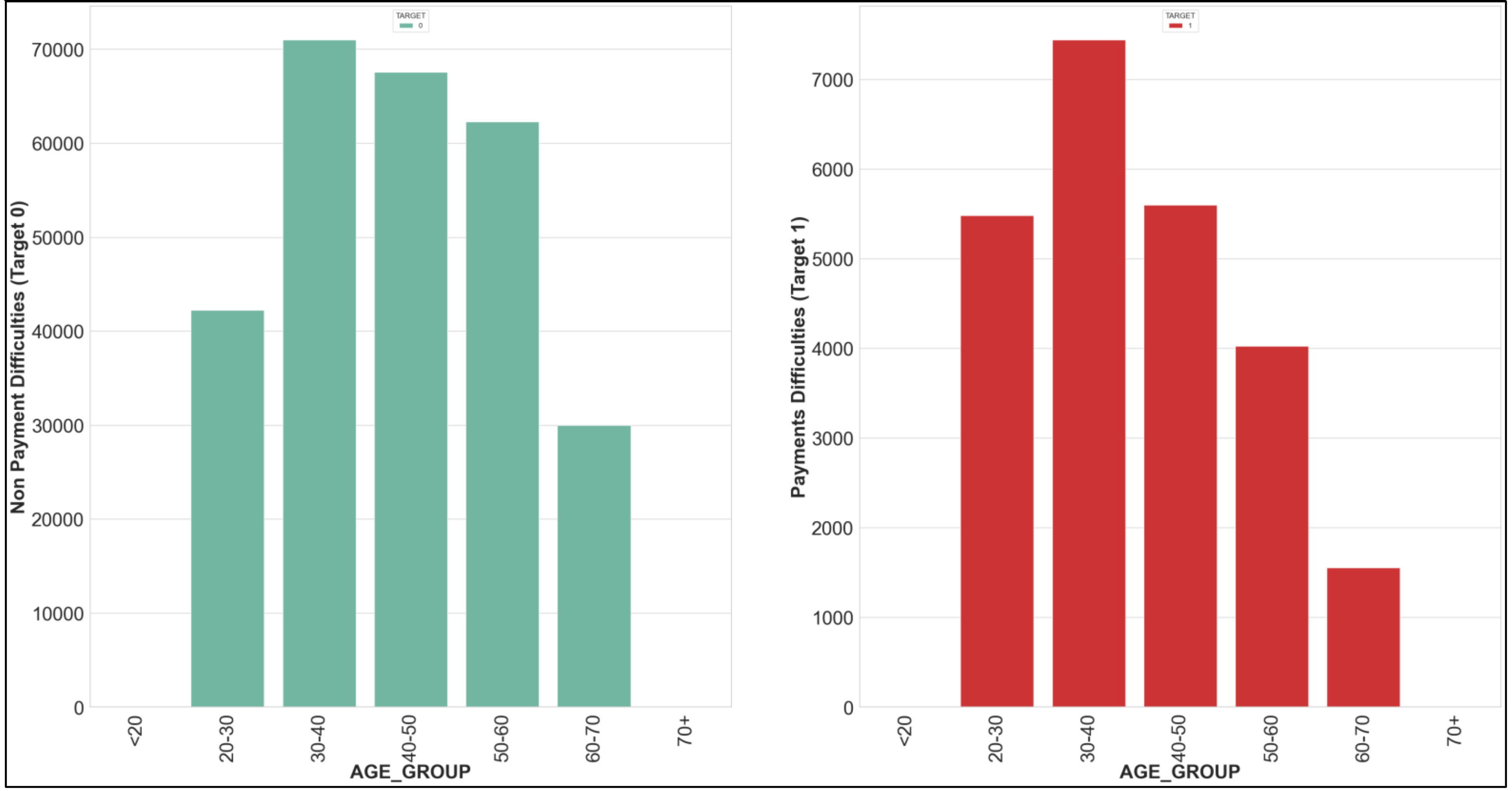
Univariate Analysis

- **Variable** : Profession

- **Insights** :

  - Applicants with payment difficulties reduces gradually as age increases.
  - Higher the age, lesser chances of payment difficulties.

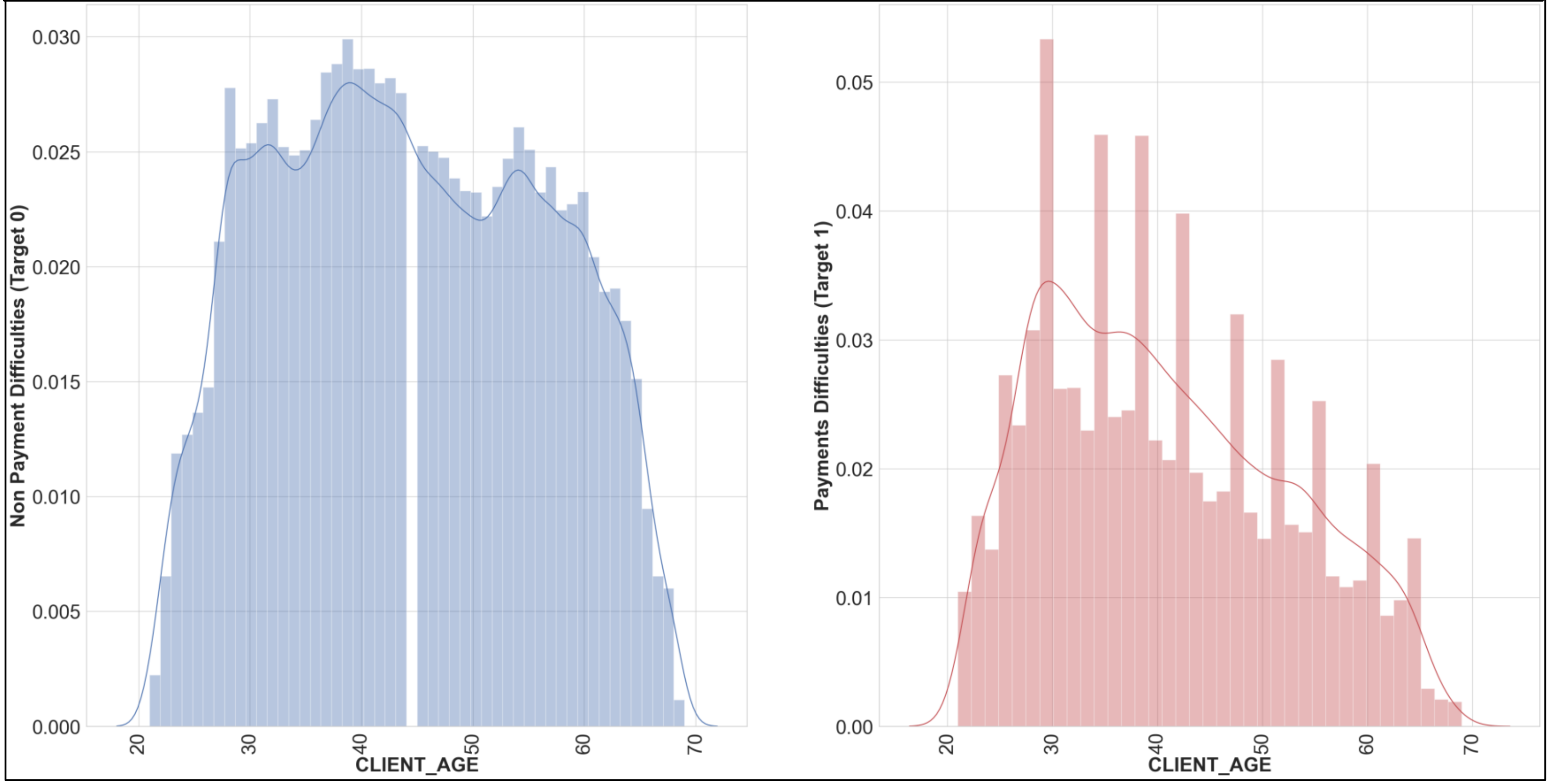
Age Group **vs.** Default counts



Univariate Analysis

- **<u>Variable</u> :** Age (yrs)

Age (yrs.) **vs.** Default counts



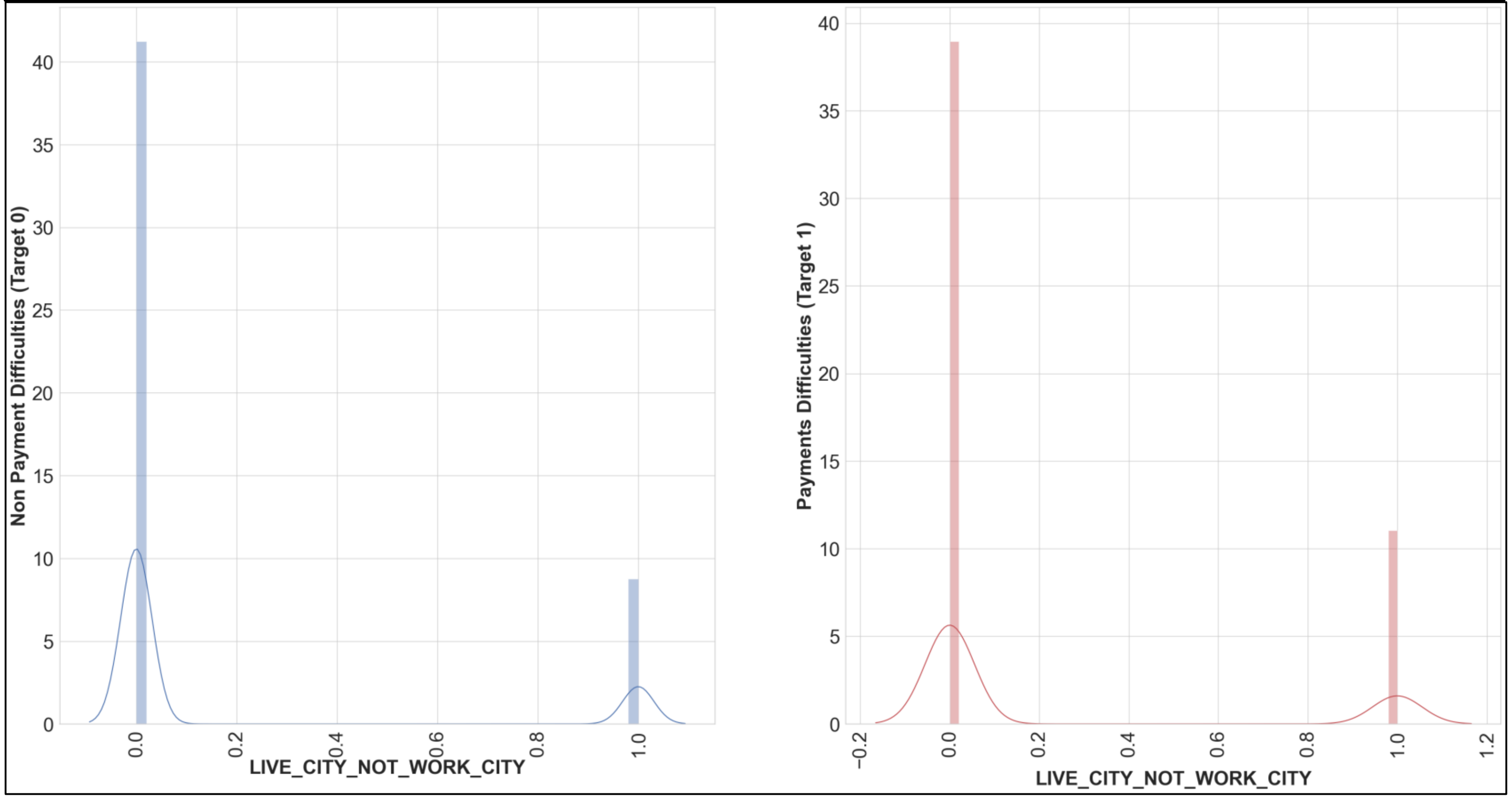
- **<u>Insights</u> :**
  - ➢ Applicant age in non defaulters is evenly distributed while in defaulters it is right skewed
  - ➢ Young applicants are more likely to become defaulters compared to senior citizens

Univariate Analysis

- **Variable :**

  Live & Work city : Not same

Live & Work city : Not same **vs.** Default counts



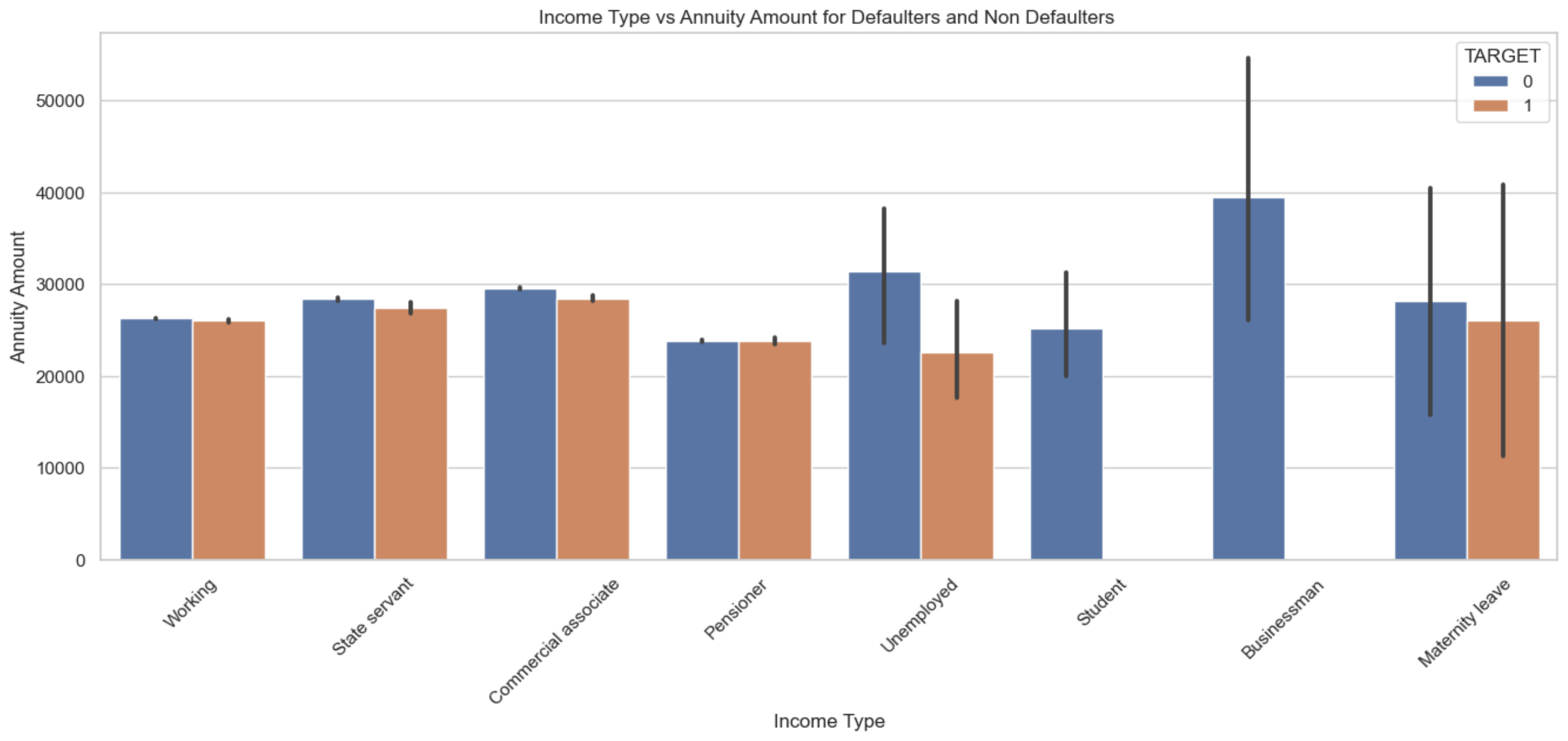
- **Insights :**

  ➢ Applicants who lives in different city than their work city has high default rate than who lives in same

Bivariate Analysis

- **<u>Variables</u> :**
  - ✓ Profession
  - ✓ Annuity

Profession **vs.** Annuity **vs.** Defaulters/Non-Defaulters



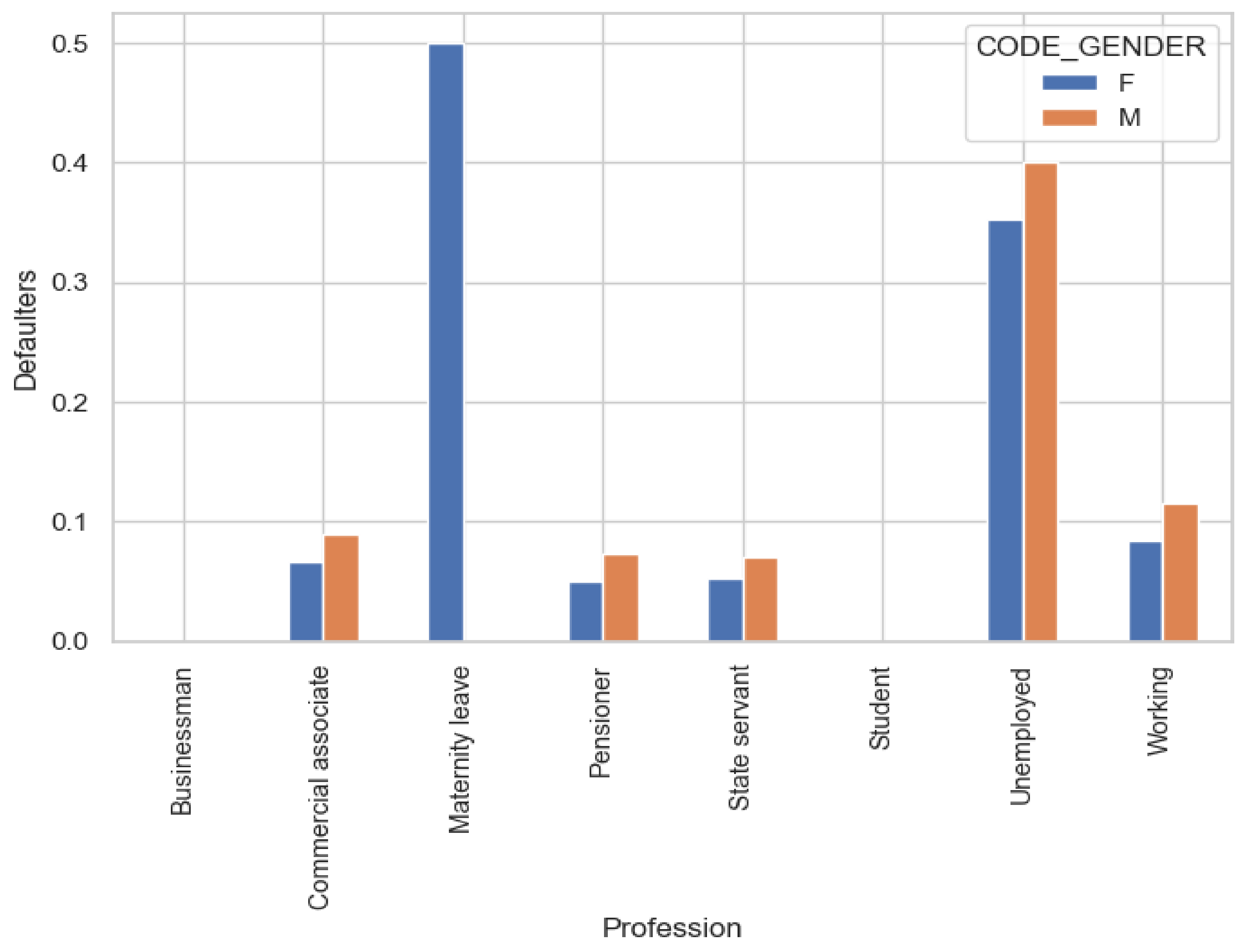
- **<u>Insights</u> :**
  - ➢ Businessman & Student applicants have nearly zero chance of payment difficulties.
  - ➢ Other income types have same defaulters ration irrespective of income type.

Bivariate Analysis

- **Variables** :
  - ✓ Profession
  - ✓ Gender

- **Insights** :
  - ➢ Unemployed applicants are more defaulted.
  - ➢ Applicants with maternity leave are expected to be defaulted more.

Profession **vs.** Gender **vs.** Defaulters/Non-Defaulters

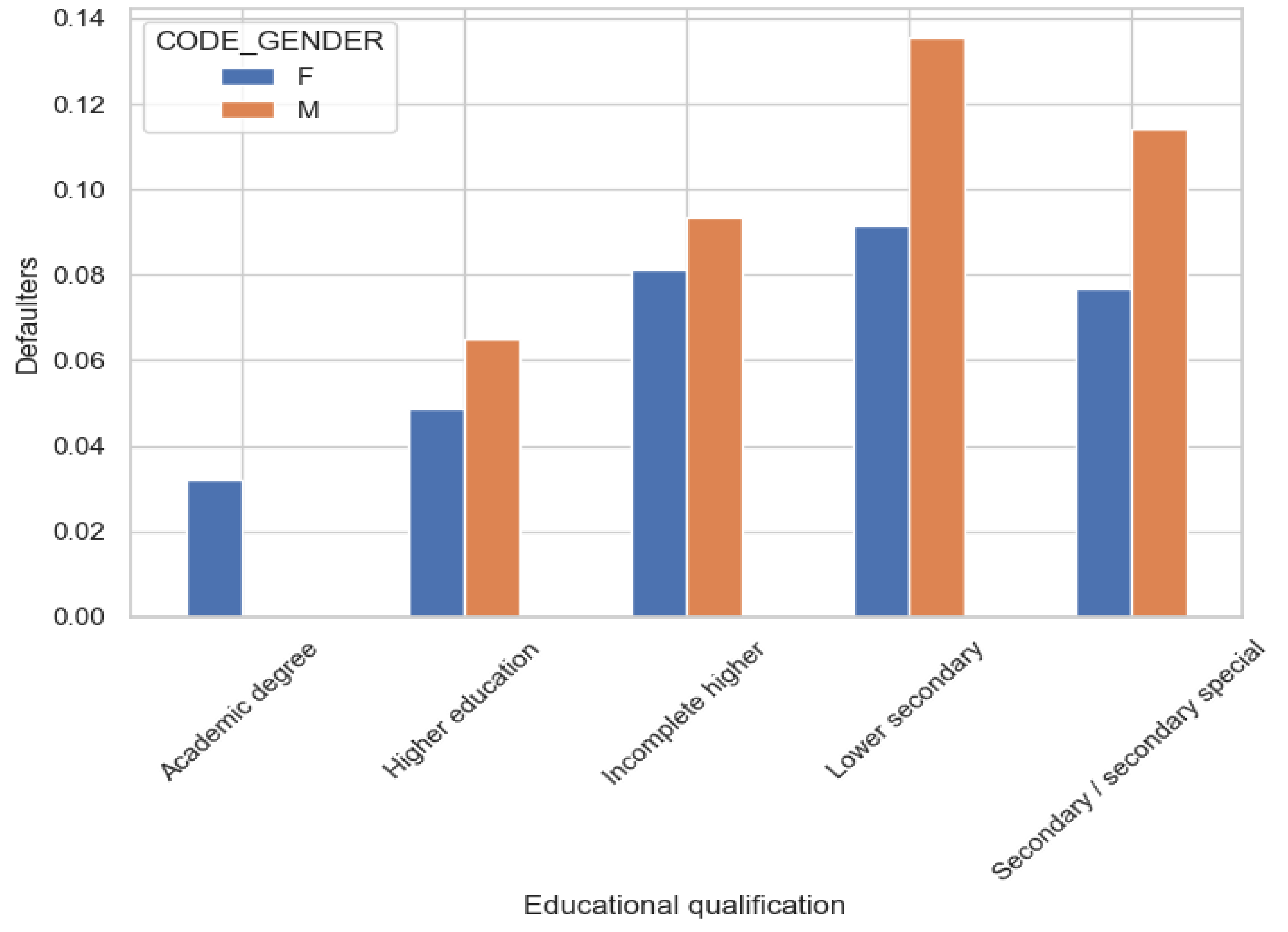


Bivariate Analysis

- **<u>Variables</u> :**
  - ✓ Gender
  - ✓ Education type

- **<u>Insights</u> :**
  - ➢ Applicants with secondary or lower secondary education are more likely to be defaulters.
  - ➢ Applicants with Academic degree are least among defaulters.
  - ➢ Irrespective of education Females are on lower side of defaulters.

Gender **vs.** Education type **vs.** Defaulters

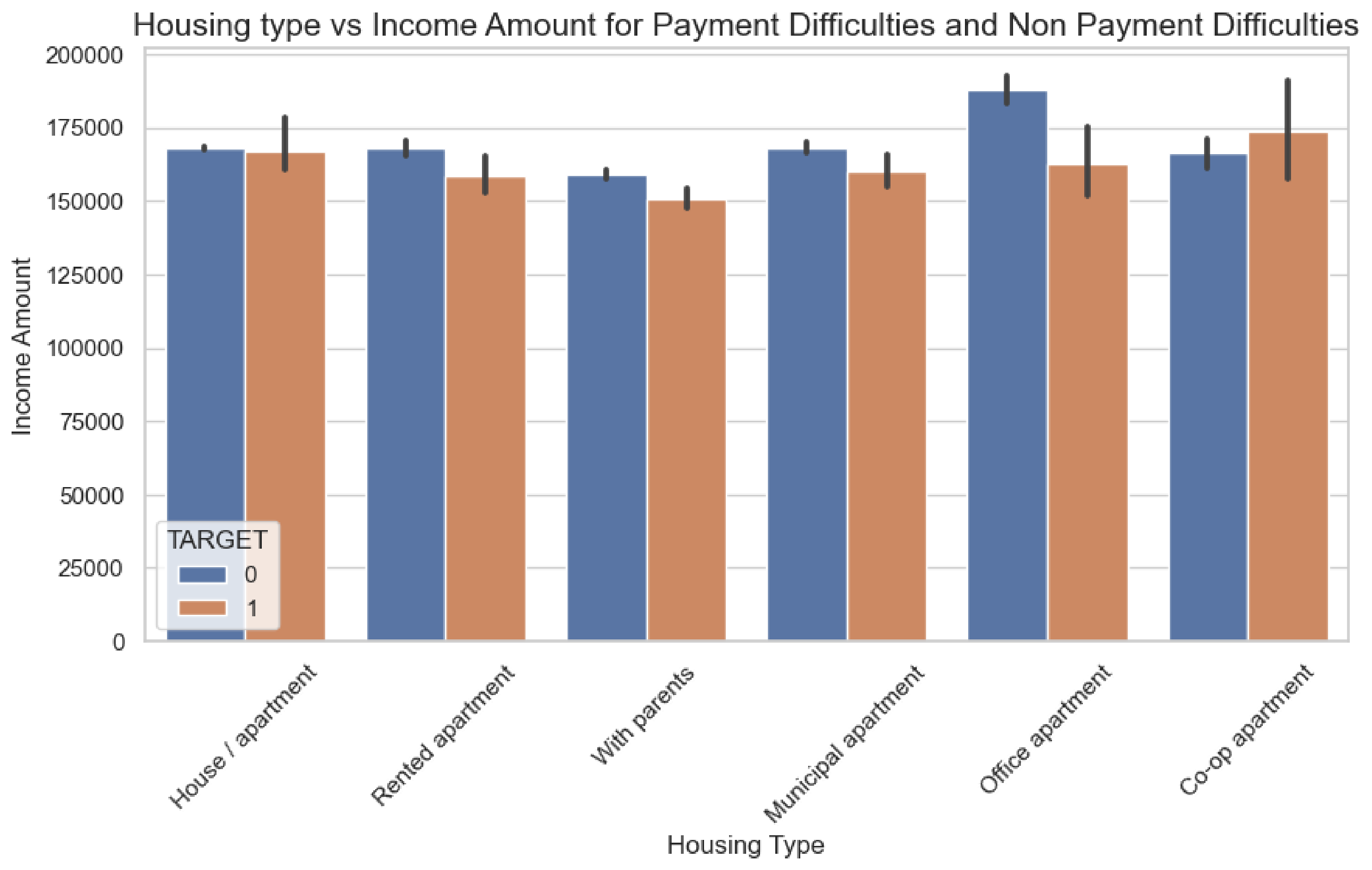


Bivariate Analysis

- **Variables :**
  - ✓ Housing Type
  - ✓ Income Amount

- **Insights :**
  - ➢ Co-op apartments applicants have higher avg. income then other type.
  - ➢ All other type defaulters have lower avg. income compared to non-defaulters

Housing Type **vs.** Income Amount **vs.** Defaulters

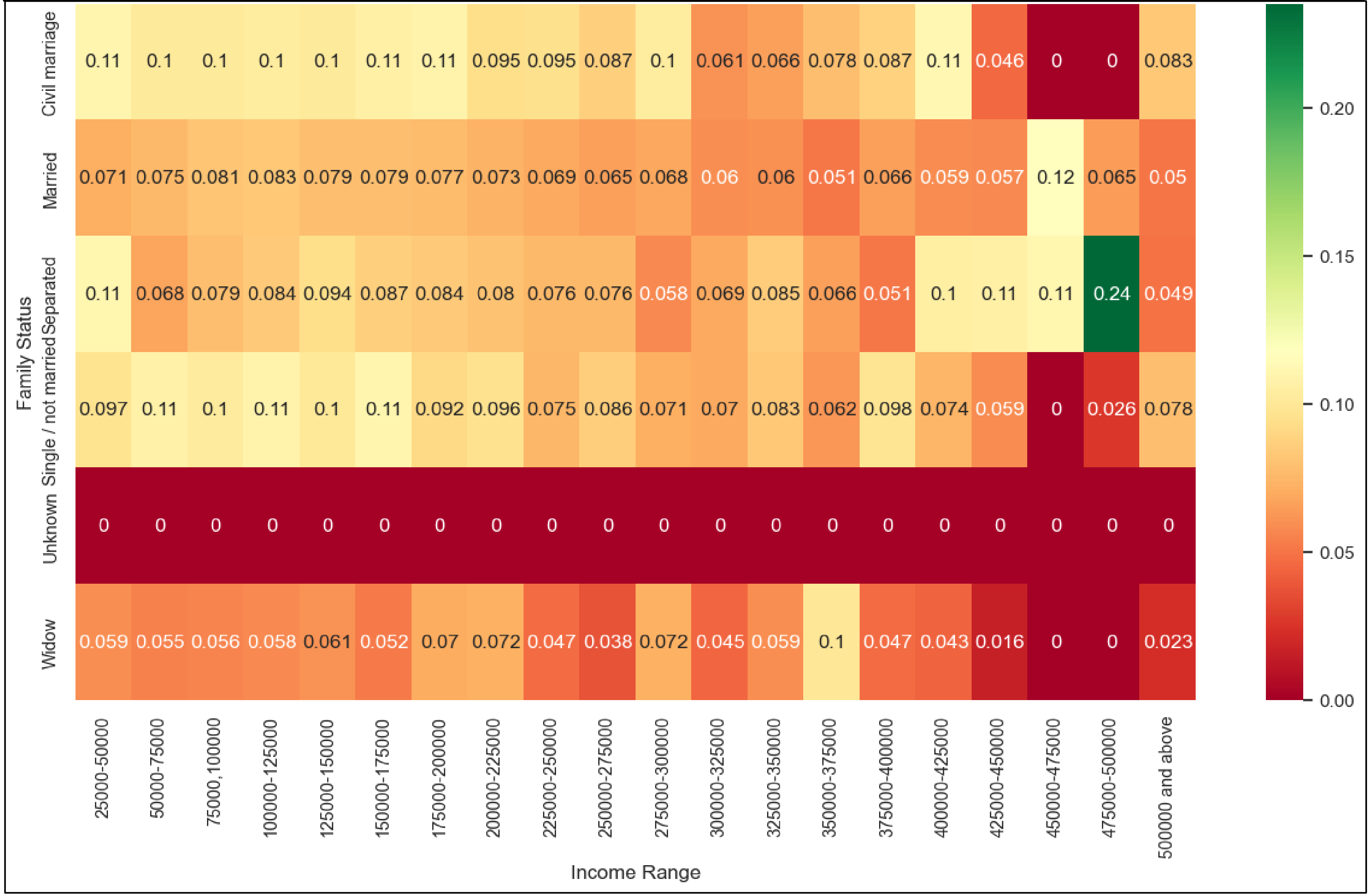


Bivariate Analysis

- **Variables :**
  - ✓ Family Status
  - ✓ Income Range

- **Insights :**
  - ➢ Applicants with separated status & income range of 475000-500000 are on higher side of defaulters.

Family Status **vs.** Income Range

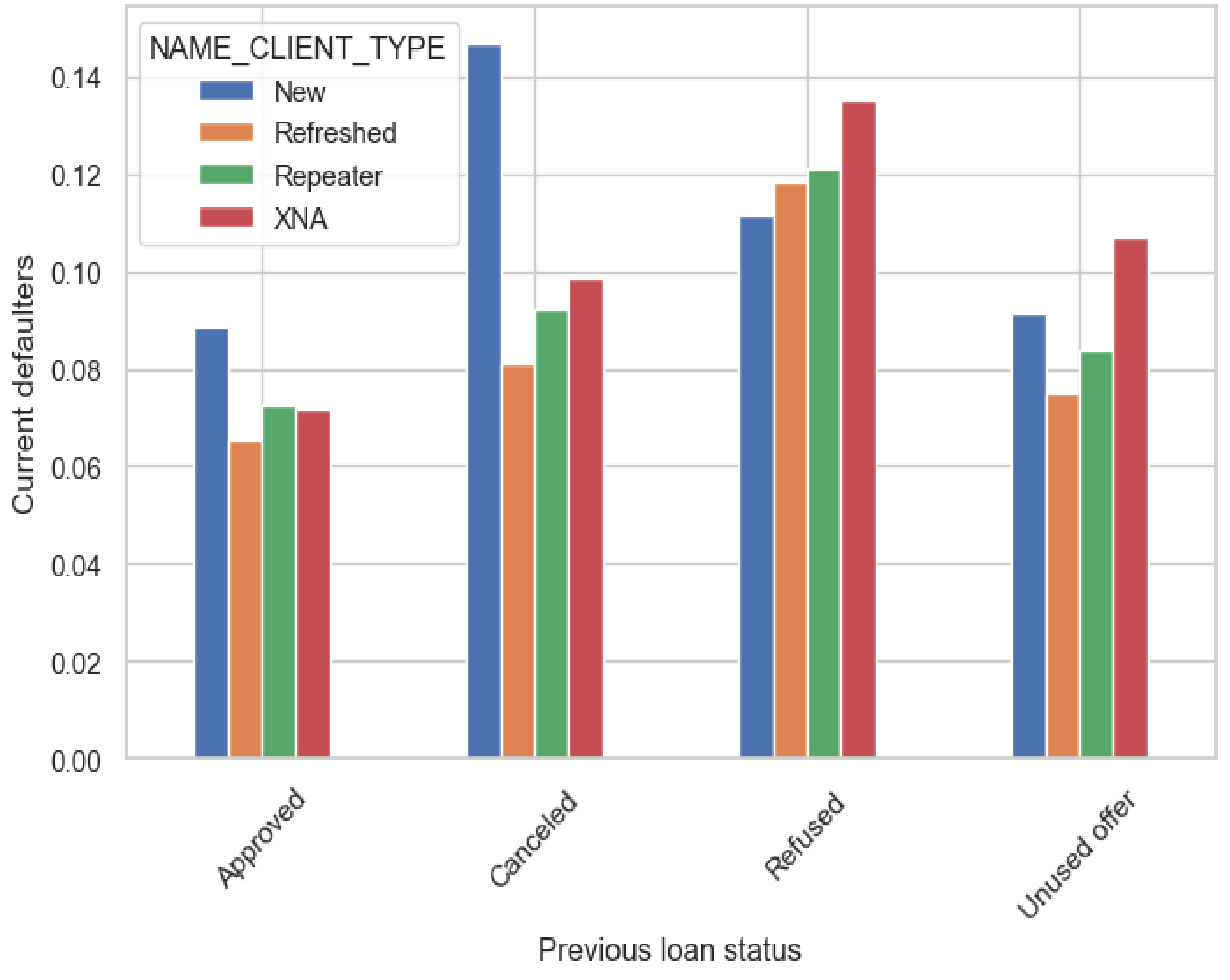


Bivariate Analysis | Previous Applicants

- **Variables :**
  - ✓ Loan Status (previous application)
  - ✓ Client type

- **Insights :**
  - ➢ Previously canceled loan has more of New clients compared to all loan status
  - ➢ Refreshed clients are less likely to be defaulters compared to others
  - ➢ For previously Approved status the New clients were more defaulted followed by Repeater.
  - ➢ For previously Refused applicants the Defaulters are more Repeater clients.

Loan Status **vs.** Client type

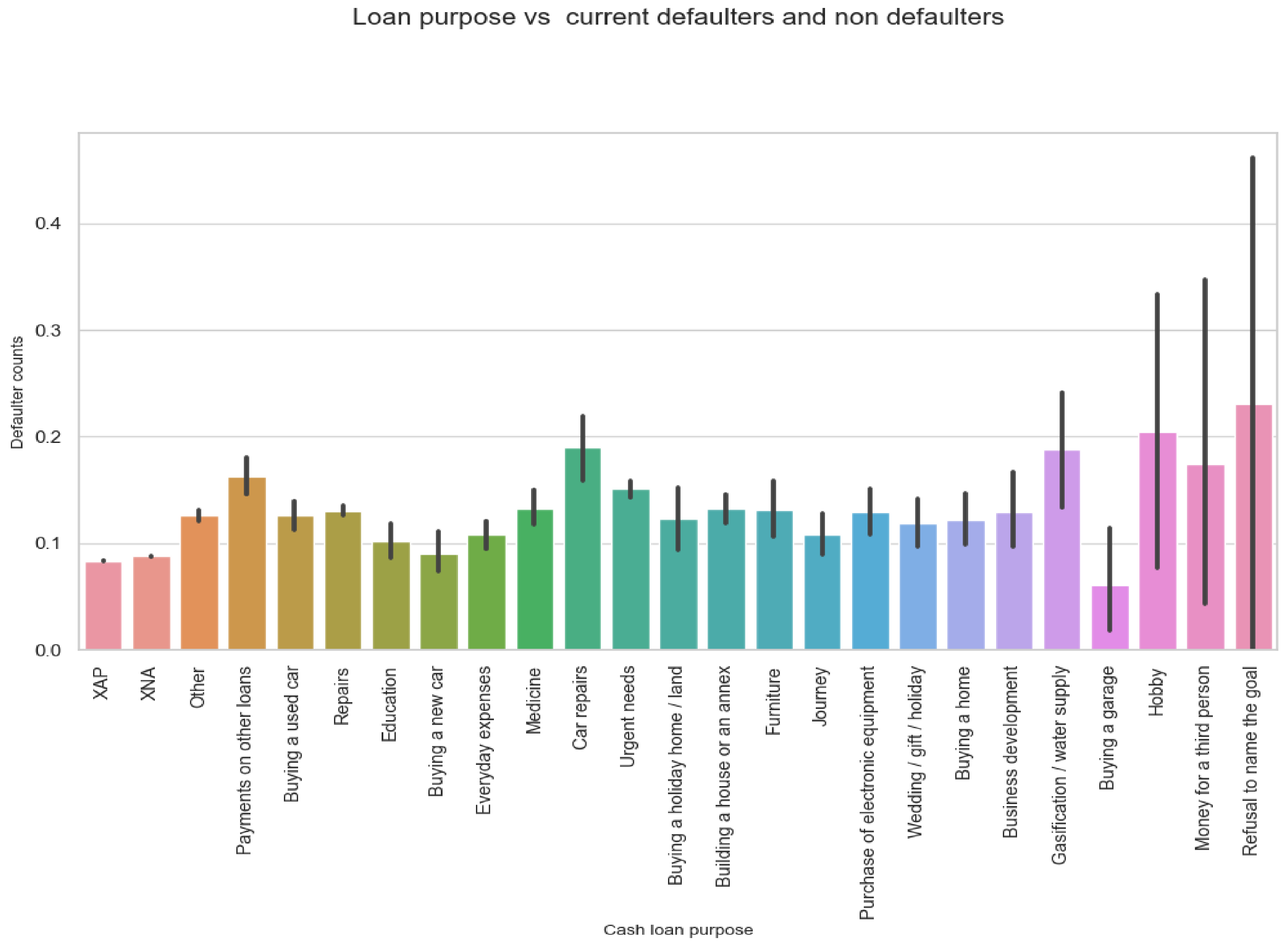


Bivariate Analysis | Previous Applicants

- **Variables** :
  - ✓ Loan Purpose (previous application)
  - ✓ Target

- **Insights** :
  - ➢ Loans with 'Un-declared purpose' are more likely to be default.
  - ➢ Loans for 'Hobby, Car repairs & Gasification' are also among more defaulted loans.

Loan Purpose **vs.** Target



Conclusion

❑ Likely Payment Difficulty Applicants

- Male applicants across all the categories compared to female
- Young age
- Lives in co-op apartments
- Having lower secondary/secondary education
- Unemployed or on maternity leaves
- Work & Permanent address : Not same
- Separated with higher income range
- New/Repeater with ‘Undisclosed/hobby purpose’

❑ Applicants to be targeted for providing loan

- Student or Business Person
- Senior citizens below certain amount of credit
- Refreshed loan applicants
- Applying for home loan / car loan / Business development with fair income range