IIT-B and Upgrad

EDA Credit Assignment

Presented by: Anish Deshpande

Agenda

- Business Overview
- Goals of EDA
- Steps involved in EDA
- Data Imbalance
- Univariate Analysis
- Bivariate Analysis
- Conclusion

IIT-B and Upgrad

Business Understanding

The loan providing companies are finding it hard to give loans to people due to insufficient credit history. They want to know the driving factors for clients to default

IIT-B and Upgrad

Goals of EDA

To do a credit risk analysis that will help the loan proving companies to make better decisions while approving loans to avoid financial losses

IIT-B and Upgrad

Steps involved in EDA

1. **Data Understanding**
    a. Understanding the data and Sourcing

2. **Data Cleaning**
    a. Handling Missing values
    b. Identifying outliers
    c. Converting the columns in right format
    d. Correcting the values in the columns

3. **Data Visualization**
    a. Univariate Analysis
    b. Bivariate Analysis
    c. Multivariate Analysis

4. **Data Cleaning for Previous application dataset**

5. **Merging the two datasets**

6. **Analysis of the merged dataset**

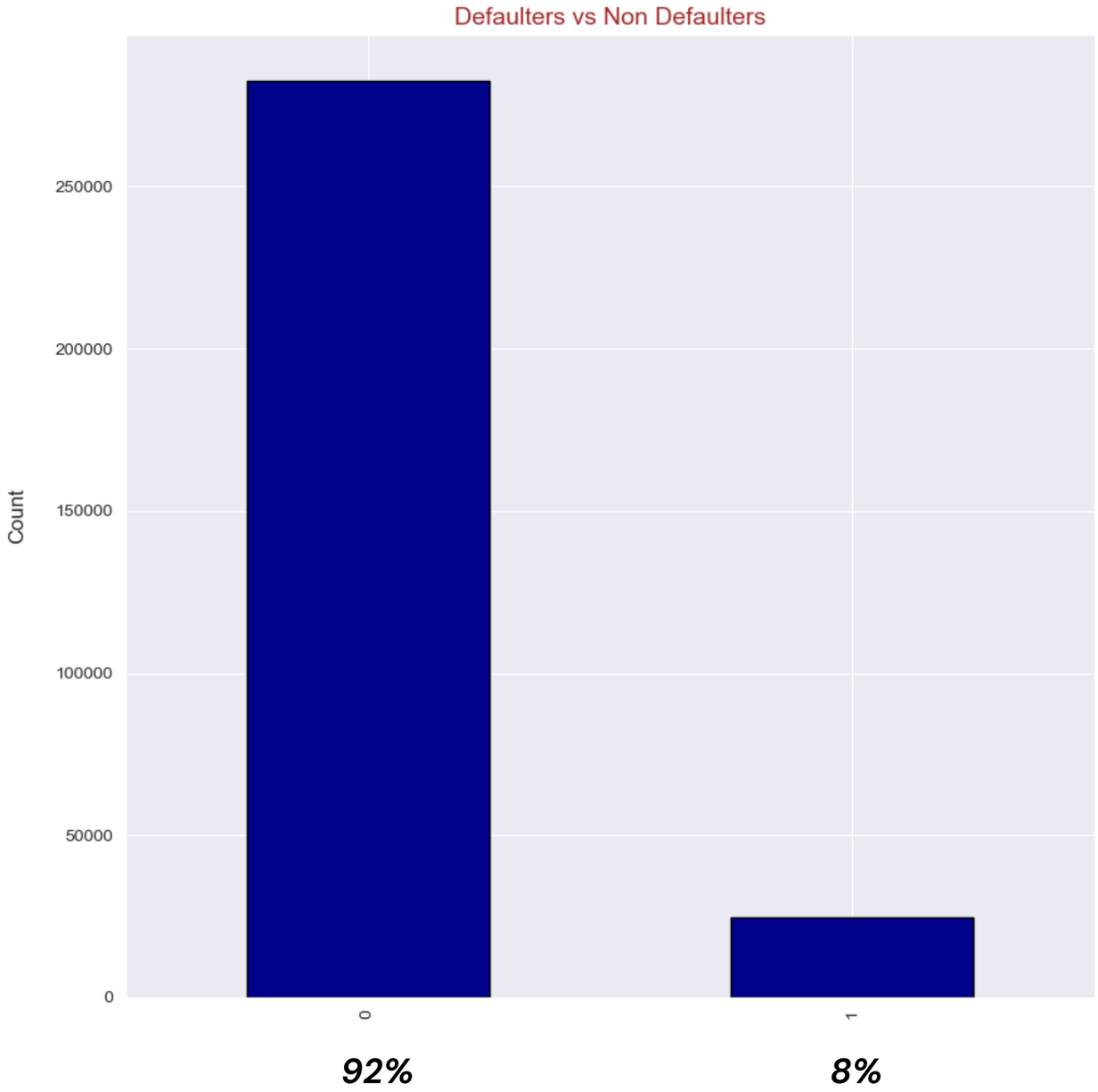
7. **Conclusion**

IIT-B and Upgrad

Data Imbalance

1 represents Defaulters

0 represents Non defaulters

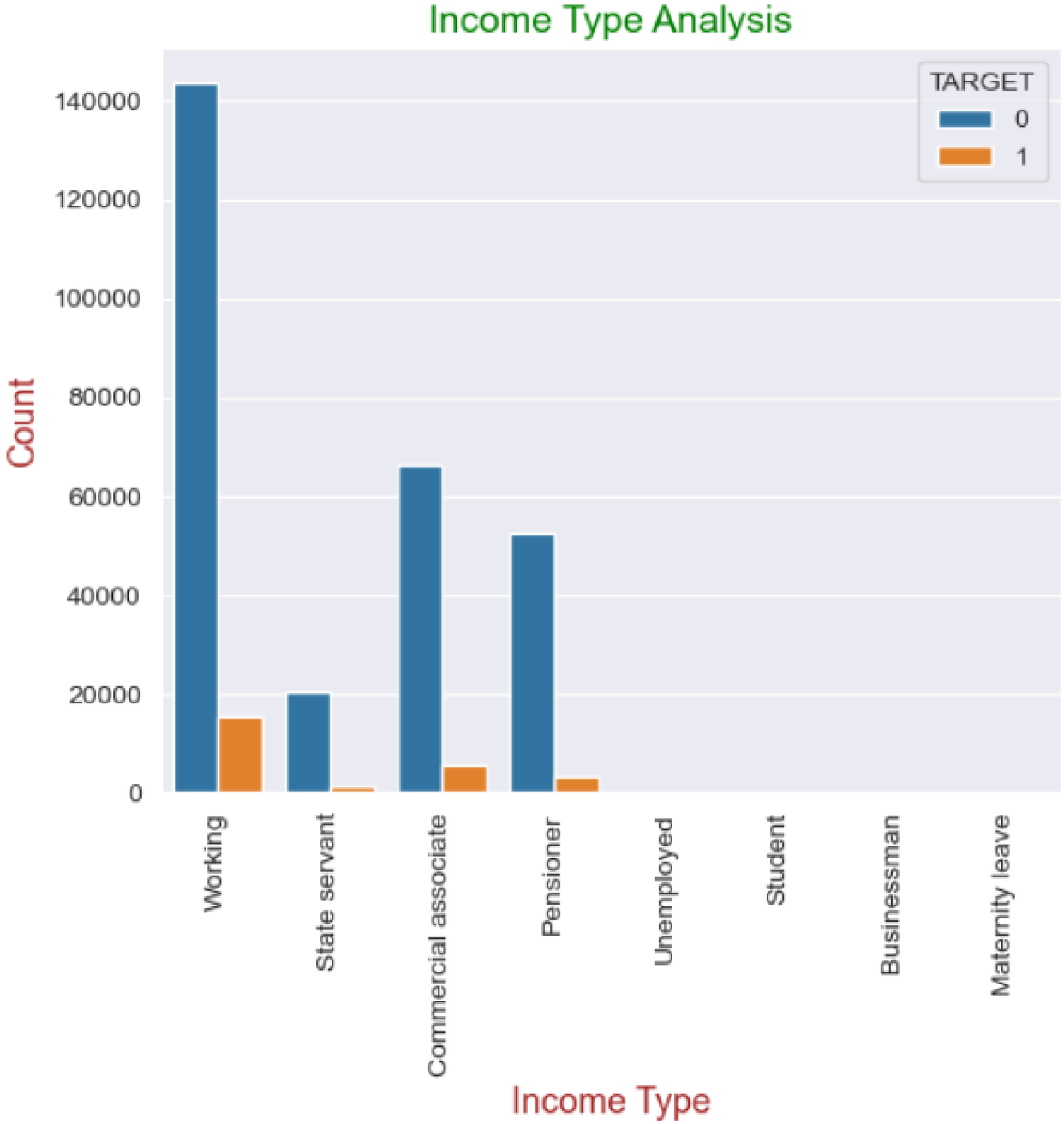


IIT-B and Upgrad

Univariate Analysis

Income Type Analysis

- Majority of the applicants are working
- Working category has the highest defaulters
- Students will not have any income as they are not yet working professionals

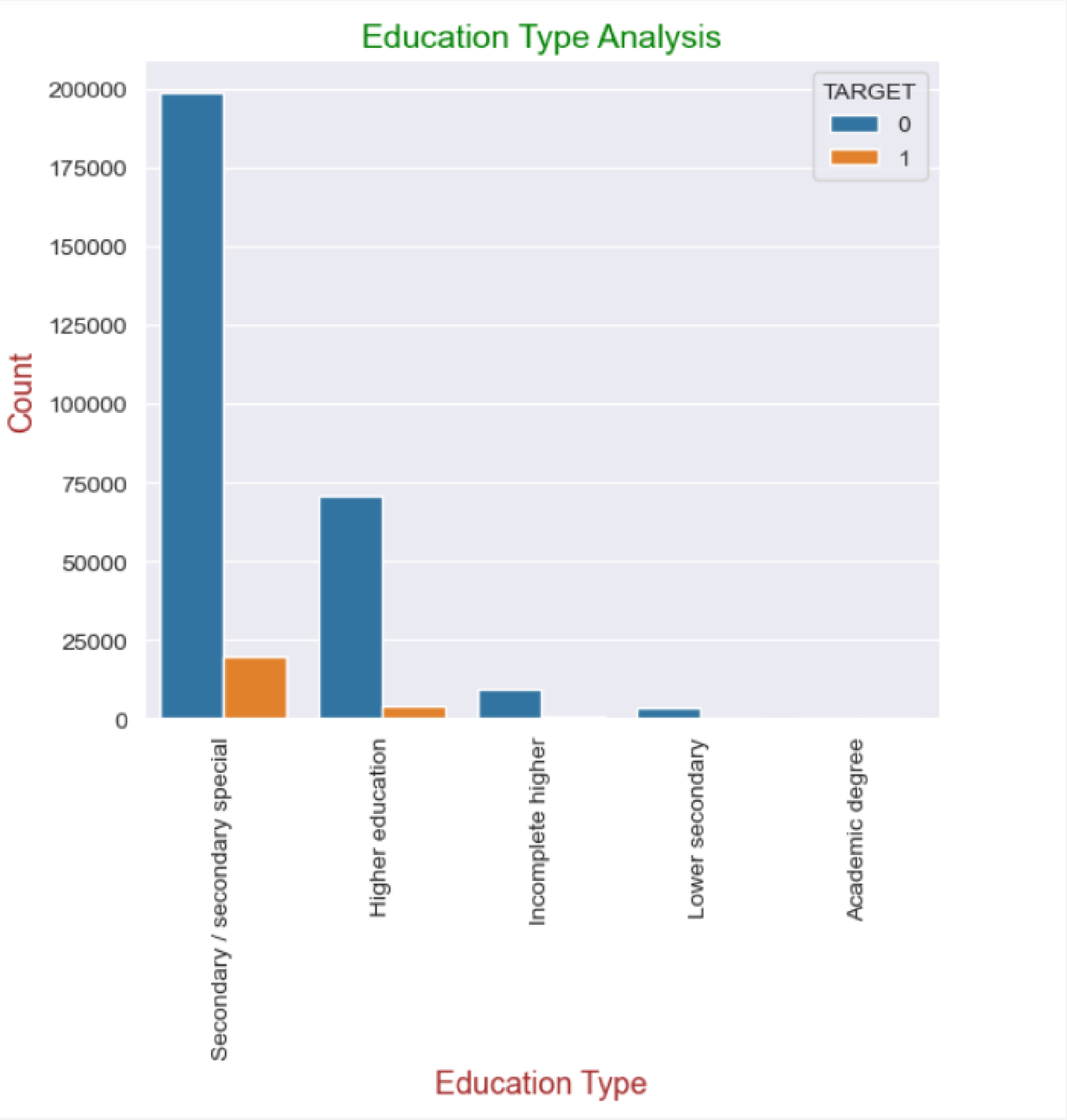


IIT-B and Upgrad

Univariate Analysis

Education Type Analysis

- Most people applied for a loan falls under Secondary / Secondary Education type which constitutes to 71%
- People with an Academic degree are less likely to be defaulters

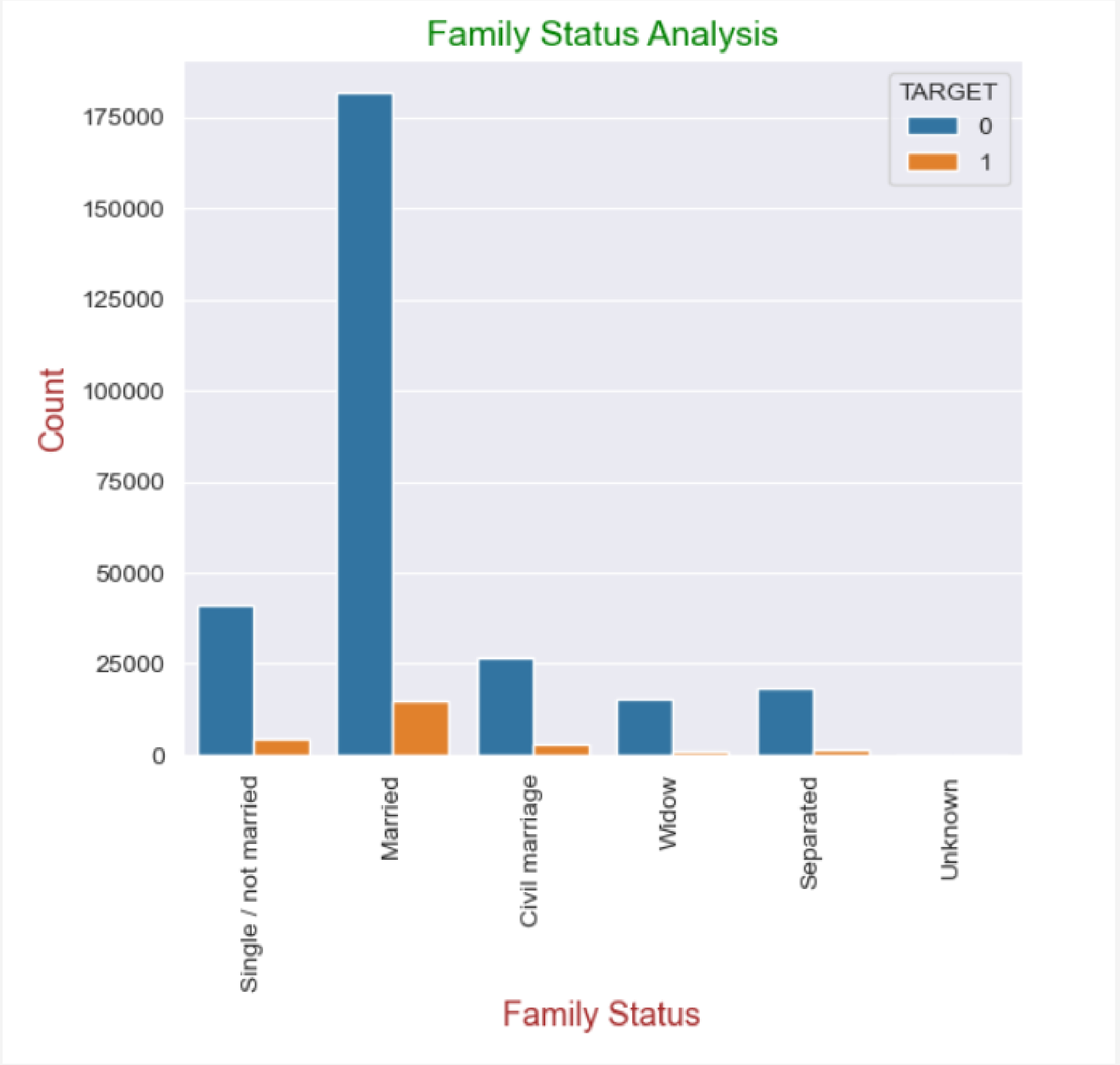


IIT-B and Upgrad

Univariate Analysis

Family Status Analysis

- Maximum people applied for a loan are married with a percentage of 64%
- Married clients are more likely to default and Widow clients and separated clients are less likely to default

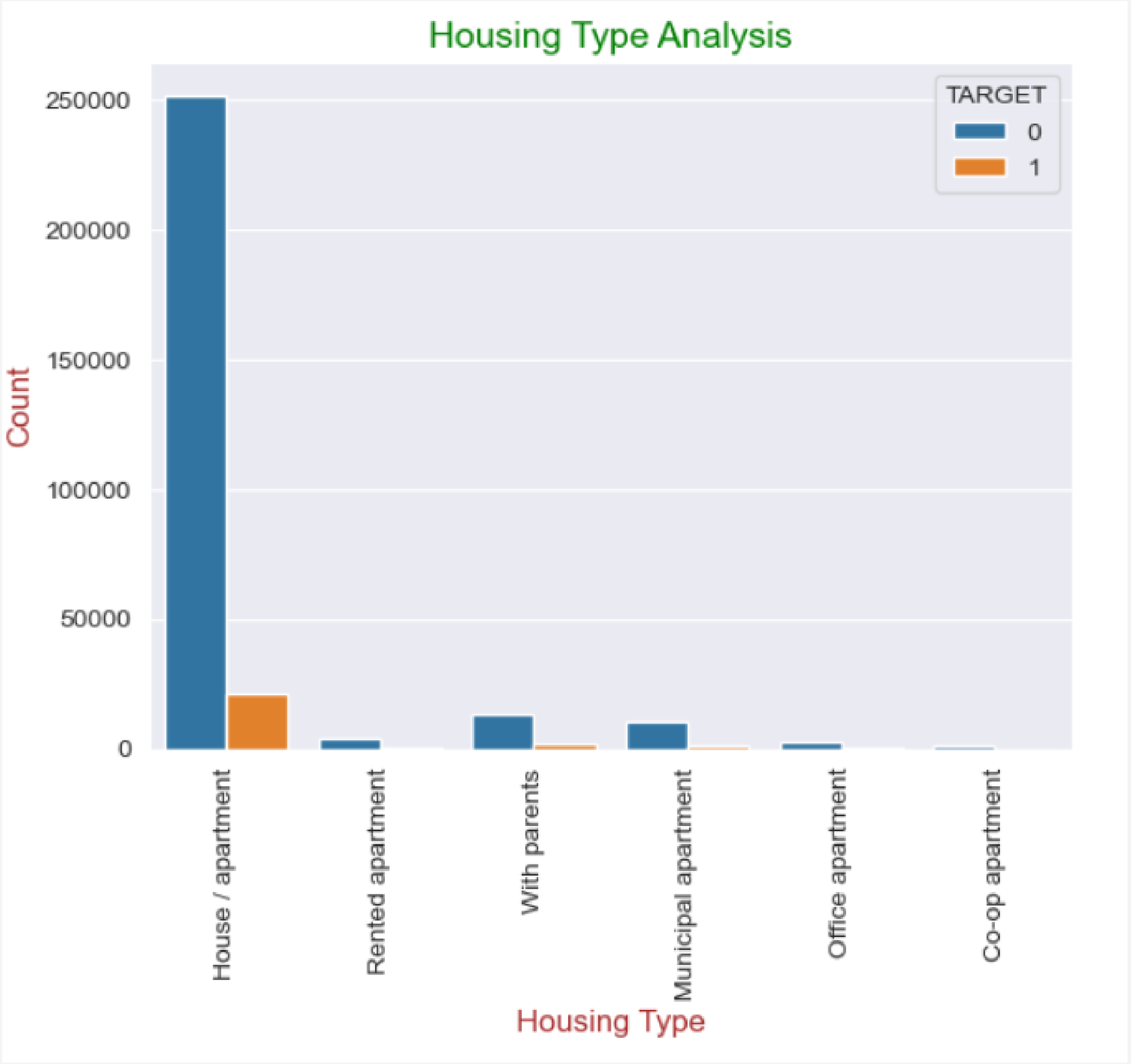


IIT-B and Upgrad

Univariate Analysis

Housing Type Analysis

- People living in a house or apartments are the highest defaulters
- People living in a municipal apartment, office apartment and rented apartments are the least defaulters

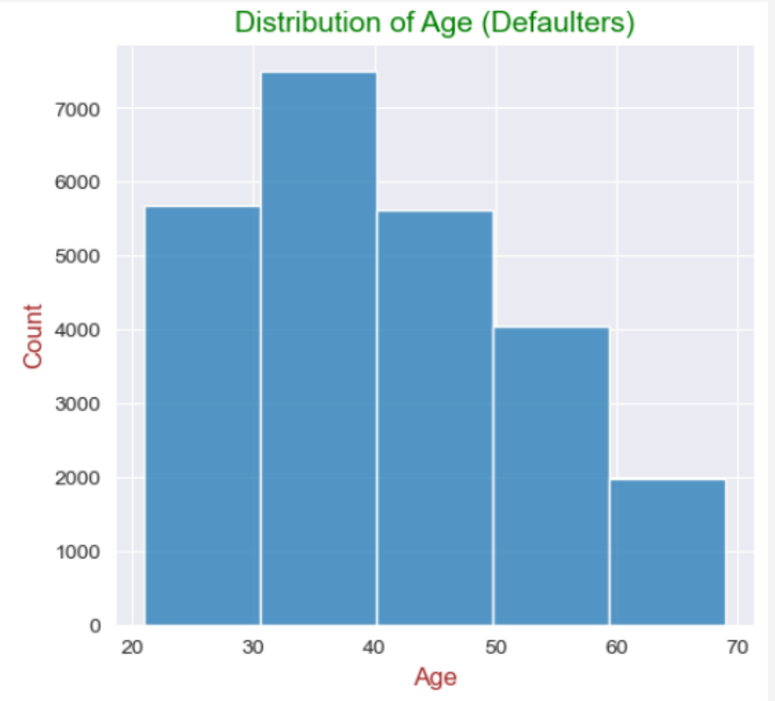


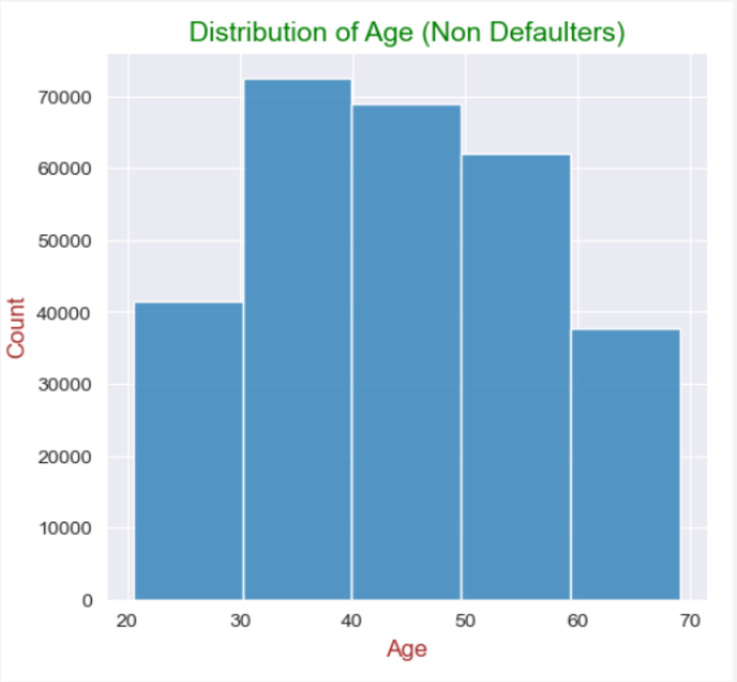
IIT-B and Upgrad

Univariate Analysis

Distribution of Age

- People who are aged 30 to 40 are the highest defaulters and Non defaulters
- People who are aged 40 to 60 are the highest non defaulters



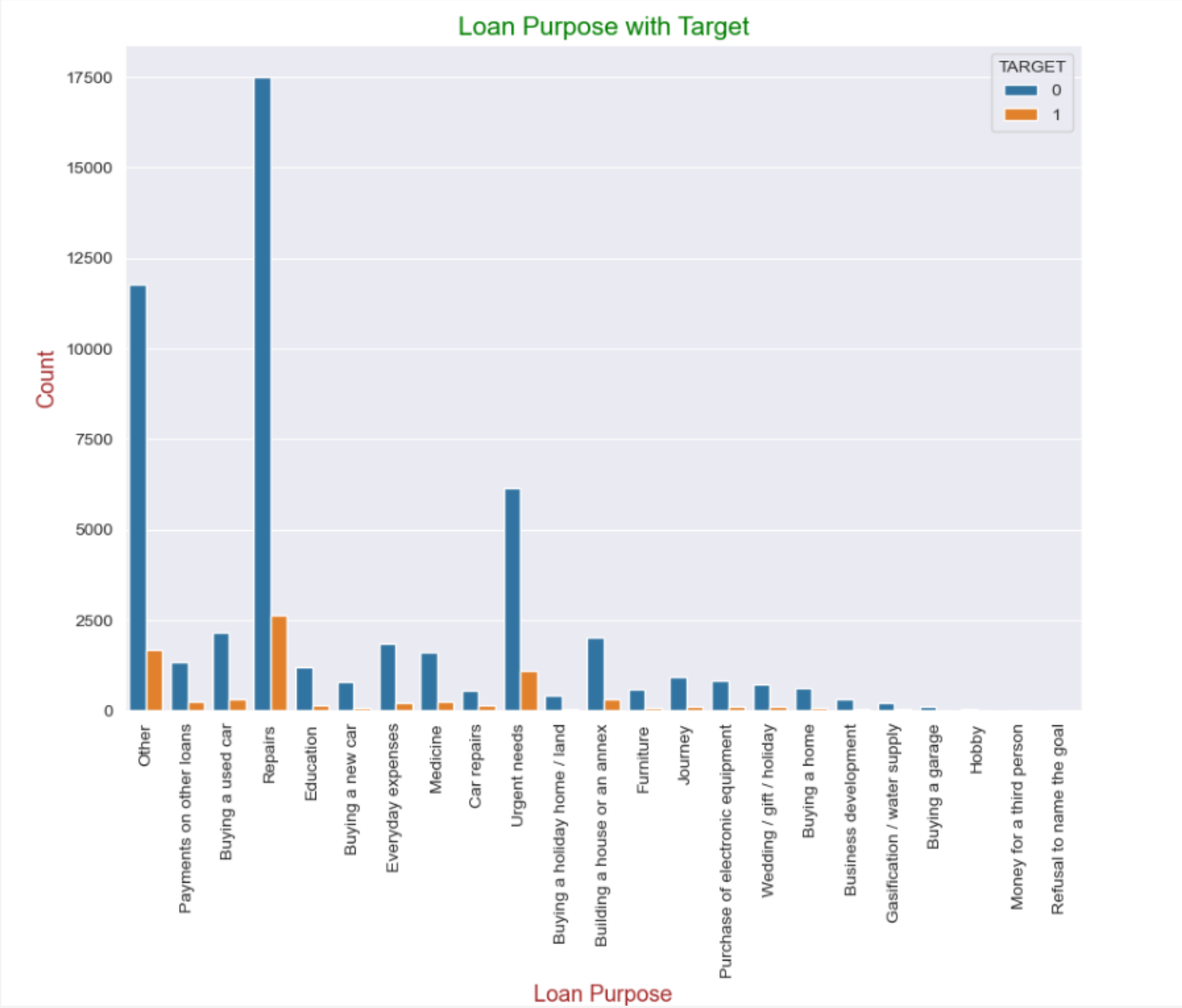


IIT-B and Upgrad

Univariate Analysis

Loan Type with target

- People taking loans for Repairs category tend to have payment difficulties



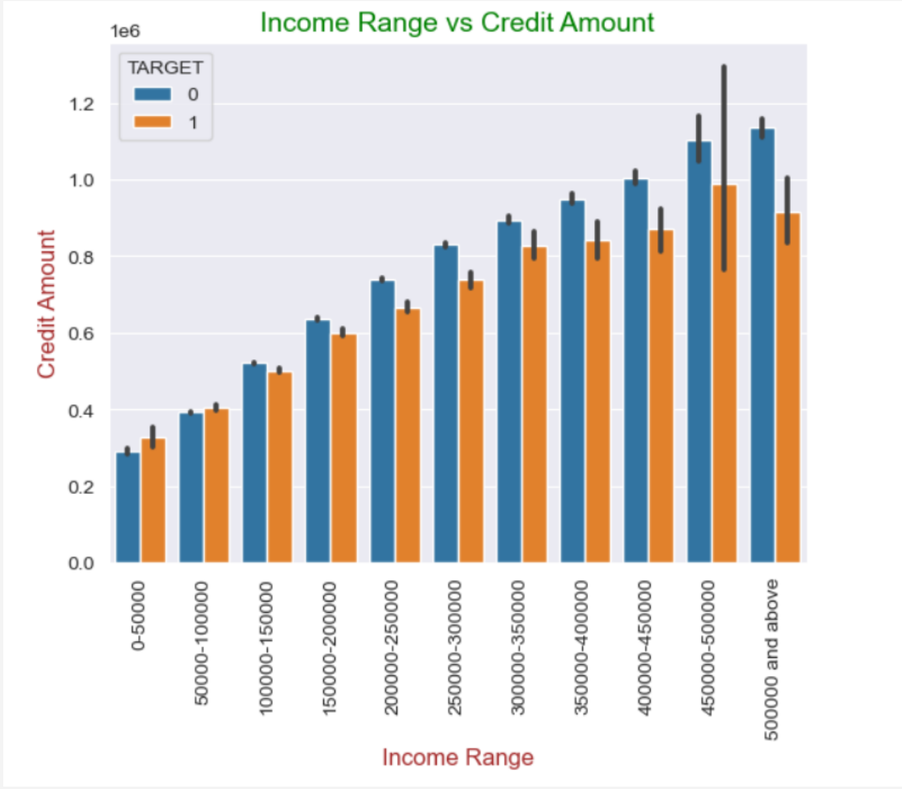
IIT-B and Upgrad

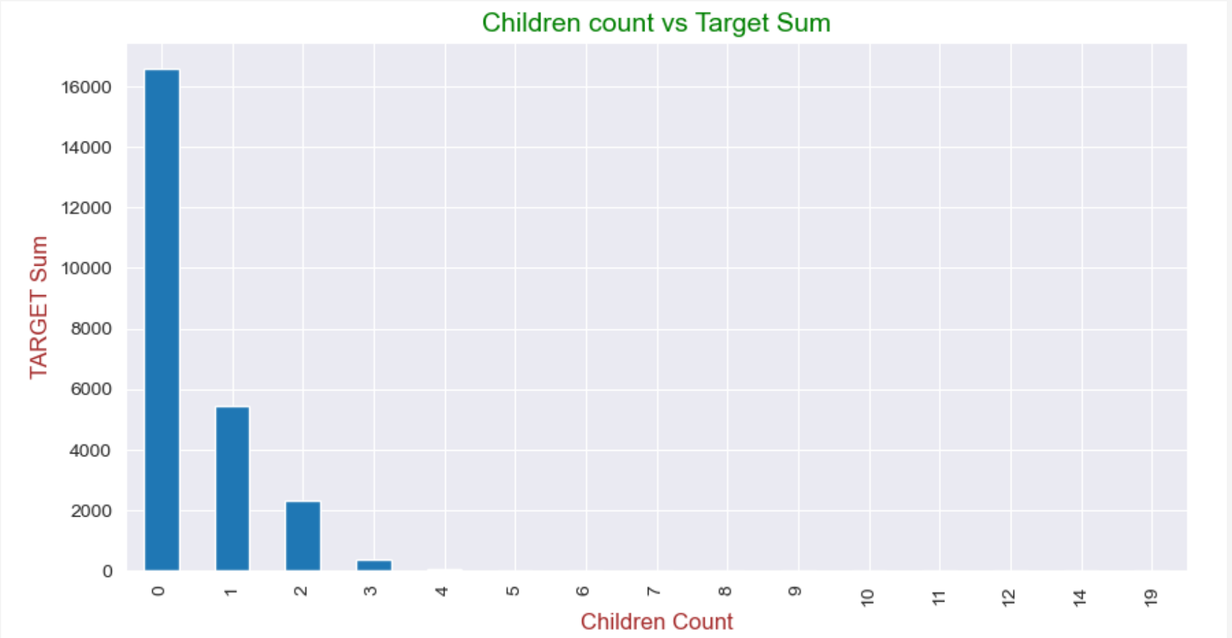
Bivariate Analysis

Income Range vs Credit Amt

Children count vs Target

- Clients with higher income range i.e. 4000000 and above are less likely to be defaulters
- Newly married couples or people with 0 children count are most likely to be defaulters

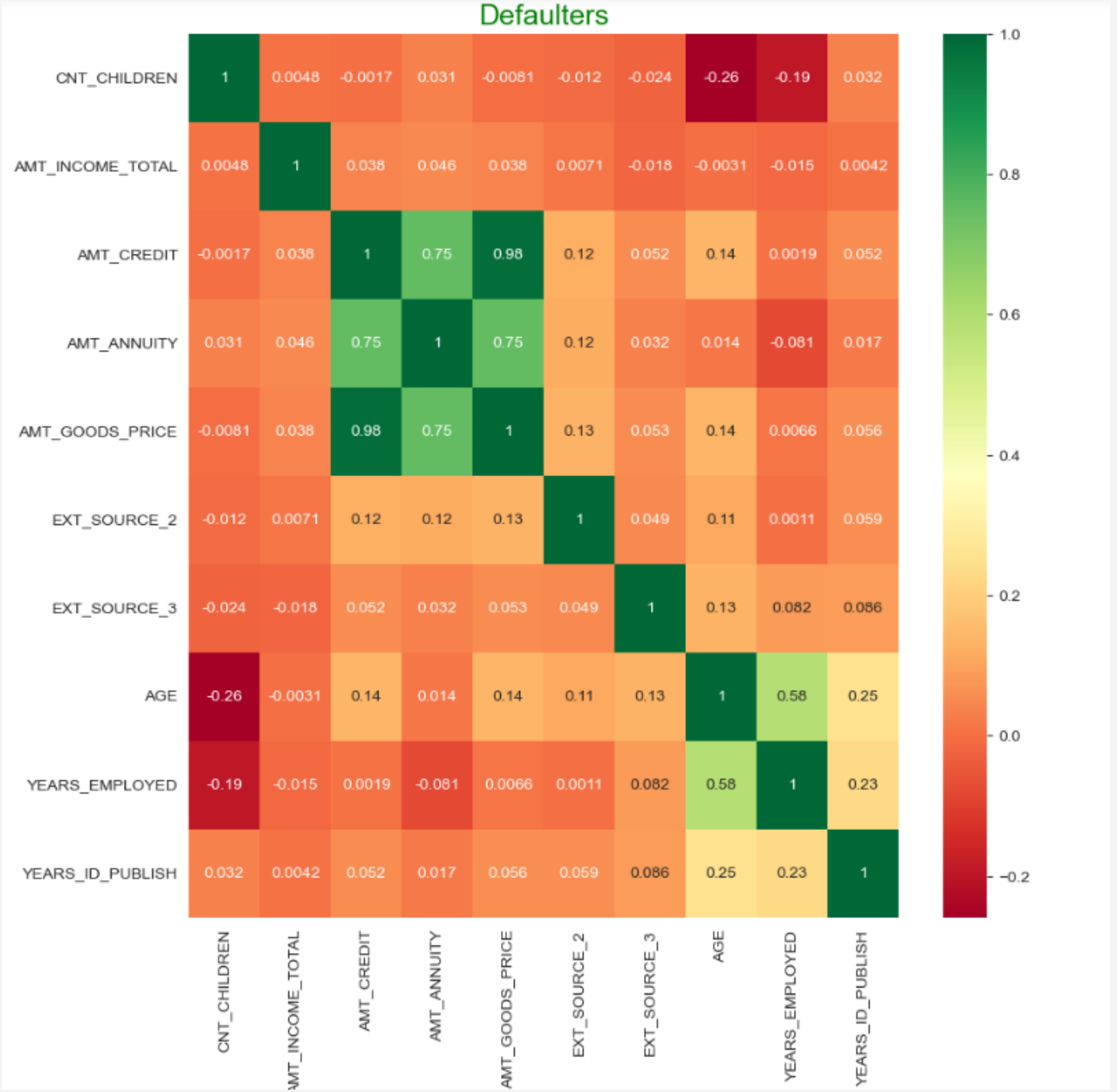




IIT-B and Upgrad

Multivariate Analysis

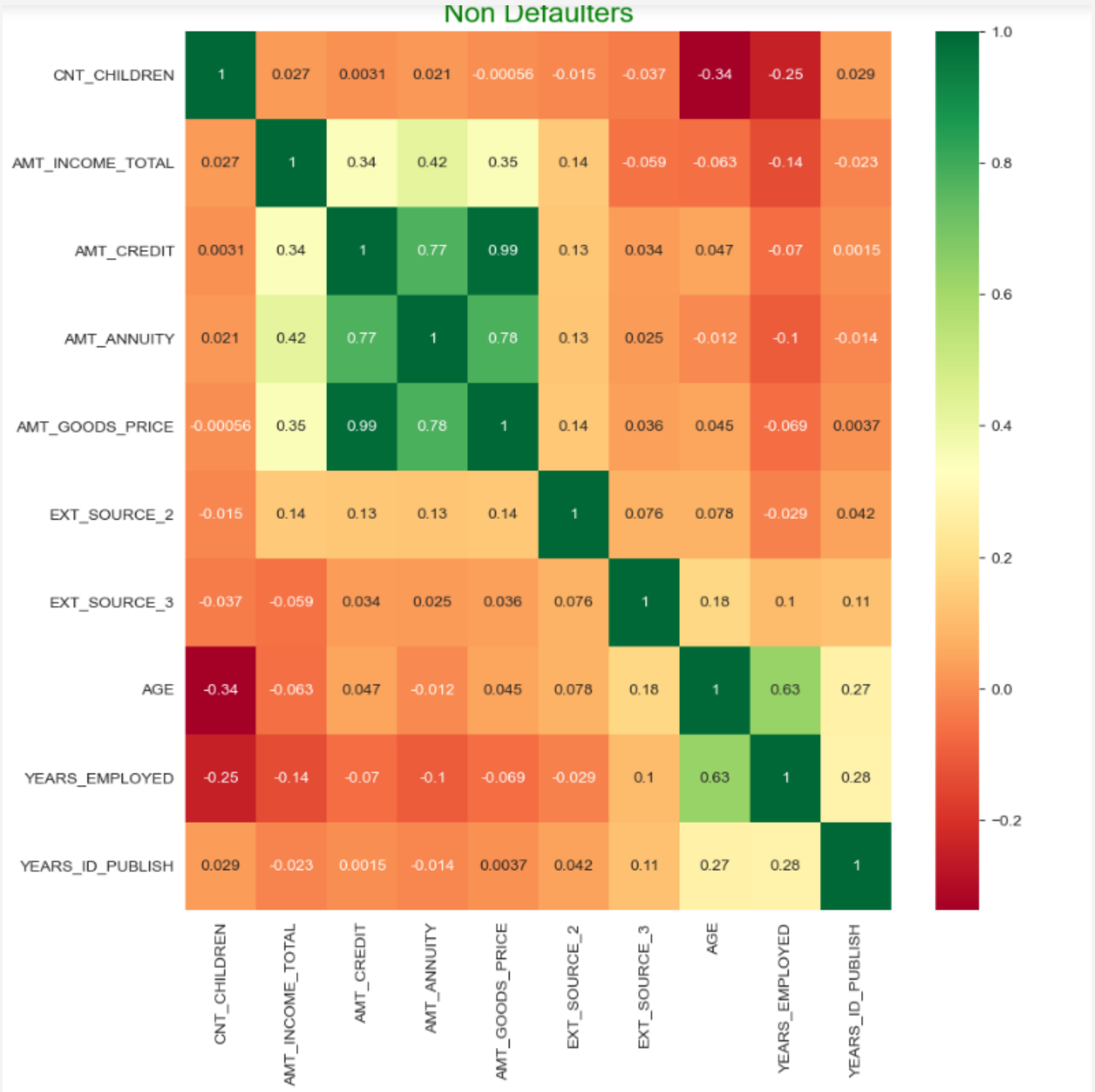
- There is a high correlation between Amt Annuity and Credit amount
- Very high positive correlation between Amt goods price and Credit amt
- Lower correlation between Income and Credit Amt.



IIT-B and Upgrad

Multivariate Analysis

- There is a high correlation between Amt Annuity and Credit amount
- Very high positive correlation between Amt goods price and Credit amt
- Moderate correlation between Income and Credit Amt.
- Moderate correlation between Income and Amt annuity



IIT-B and Upgrad

Conclusion

Recommended group where loans can be credited

- People aged 40 to 60
- Clients with higher income (400000 and above)
- Widowed Clients
- Clients with an Academic degree
- Clients with occupation type as Managers, Accountants and High skill tech staff
- Clients with higher children count

IIT-B and Upgrad

Conclusion

Risky Groups

- Clients opting loans for Repair purpose
- Previously refused groups
- Newly married couples or clients with 0 children
- Younger clients aged between 30 to 40
- Clients with occupation type as labors
- Clients with lower income range