

Loan Application Analysis

By Aakash Maskara (DS66 Batch)

Data Toolkit || Module 8 || Credit EDA Assignment

Date – 26th May, 2024

Contents

- Objective
- Assumptions
- Insights from Analysis of application_data.csv
- Insights from Analysis of merged files
- Conclusions & Observations
- Thank You

Objective

To perform EDA Analysis on Bank Data set and find insights on categories of clients who might be potential defaulters to loan. Also, find insights for various categories of clients who have less probability of defaulting a loan.

Input files received -

- application_data.csv
- previous_application.csv
- columns_description.csv

Assumptions

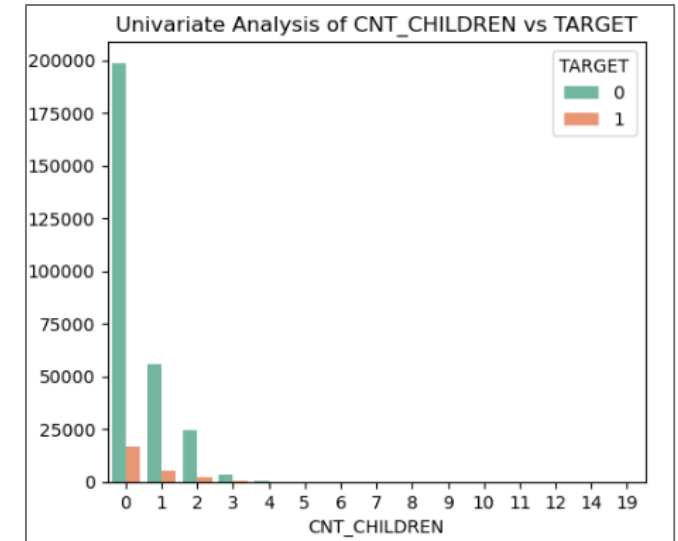
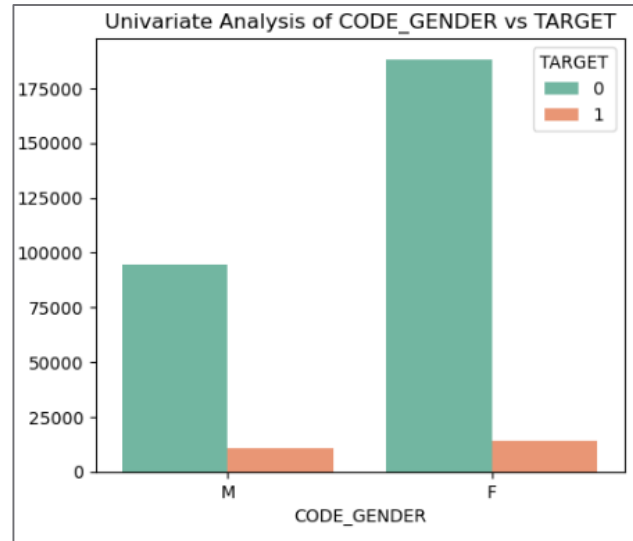
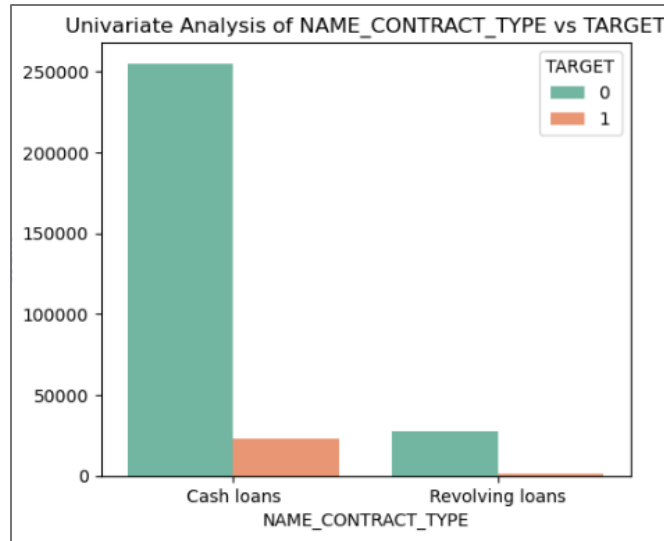
Following assumptions have been made for this analysis –

- Since the bank data consists of sensitive information where almost every parameters is a deciding factor of whether a client should be granted a loan or not, columns with null values > 40% has been dropped for this analysis.
- For columns with Null values < 40% have been filled with suitable values depending on the skewness of the data or depending on the fact if it's a categorical column or numerical column.
- Few columns have been dropped off considering they would not contribute much to the analysis compared to other potential parameters.
- All the 21 FLAG_DOCS have been combined into 1 single column which is a sum of all the FLAG_DOCS columns for 1 application (row).
- Wherever Outliers detected, they have been treated based on fact if they are potential clients for loan defaulters or not.

Insights from Analysis of application_data.csv

Univariate Analysis –

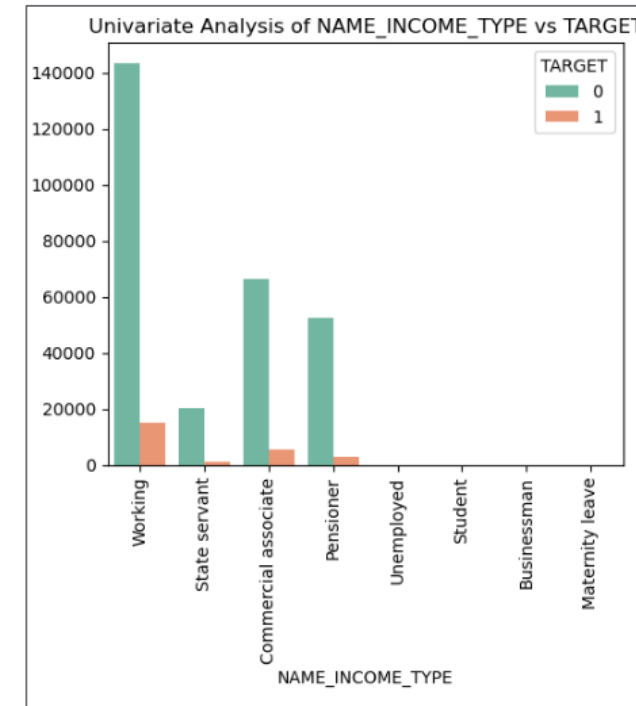
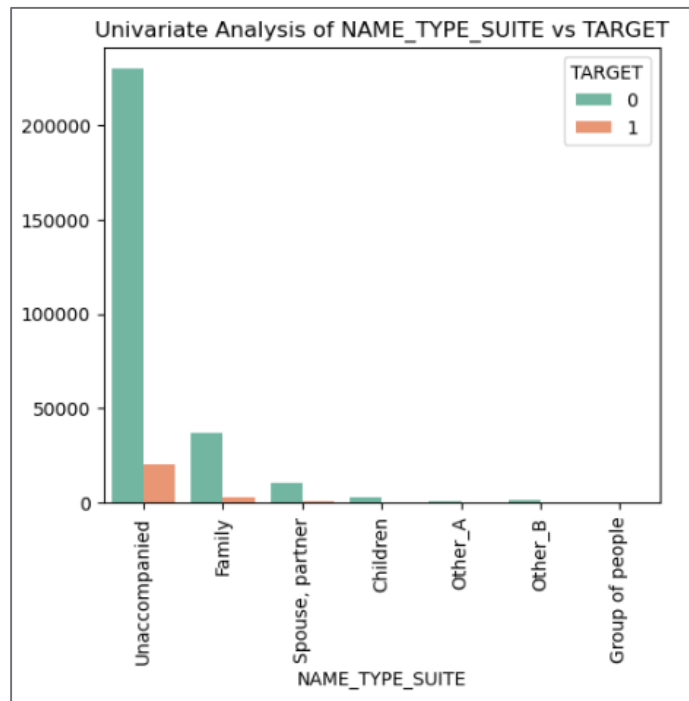
- The bank mostly provides cash loans, also there is a higher default ratio in the same.
- The bank provides loans mostly to females, a higher ratio of default has been observed in case of Male clients.
- Clients with no kids are the ideal Targets as their default ratio is very less compared to clients with 1-3 children.



Insights from Analysis of application_data.csv

Univariate Analysis –

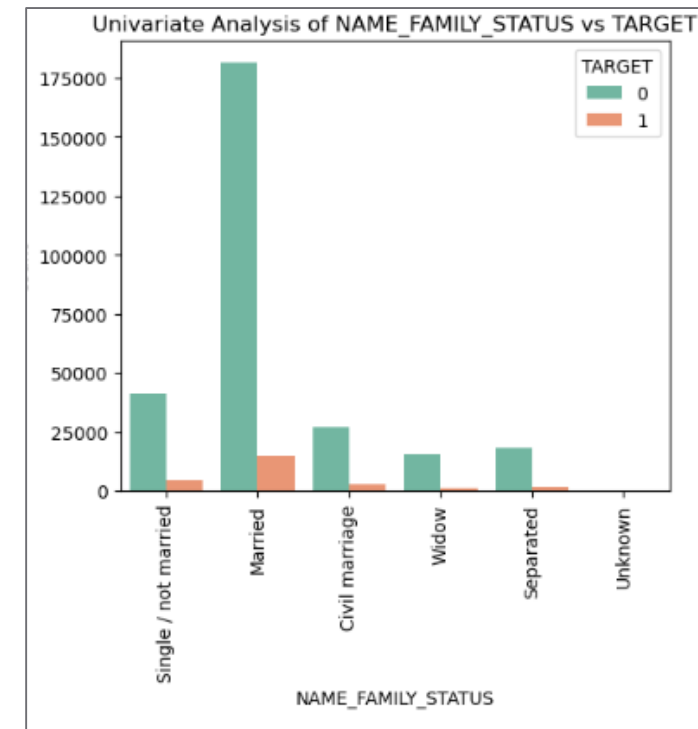
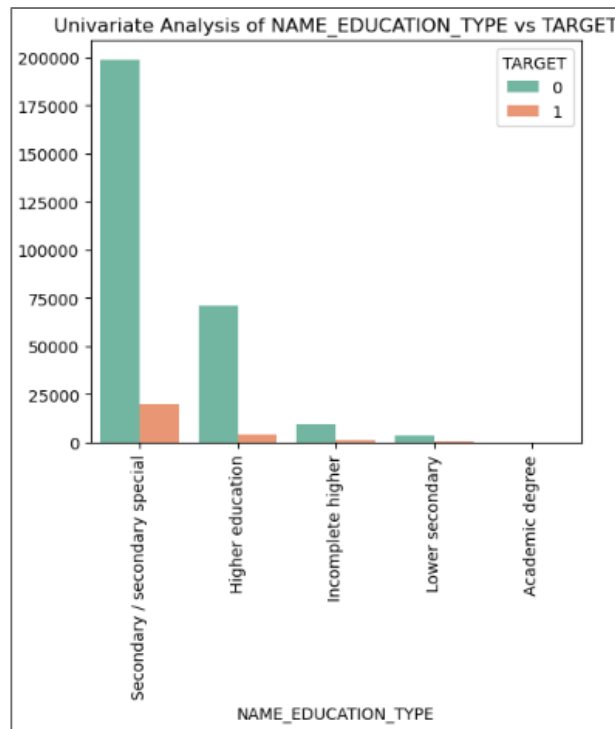
- Default rate seems lower for Unaccompanied clients compared to Family and Partners.
- Least number of defaulters are mostly observed in case of Pensioners.



Insights from Analysis of application_data.csv

Univariate Analysis –

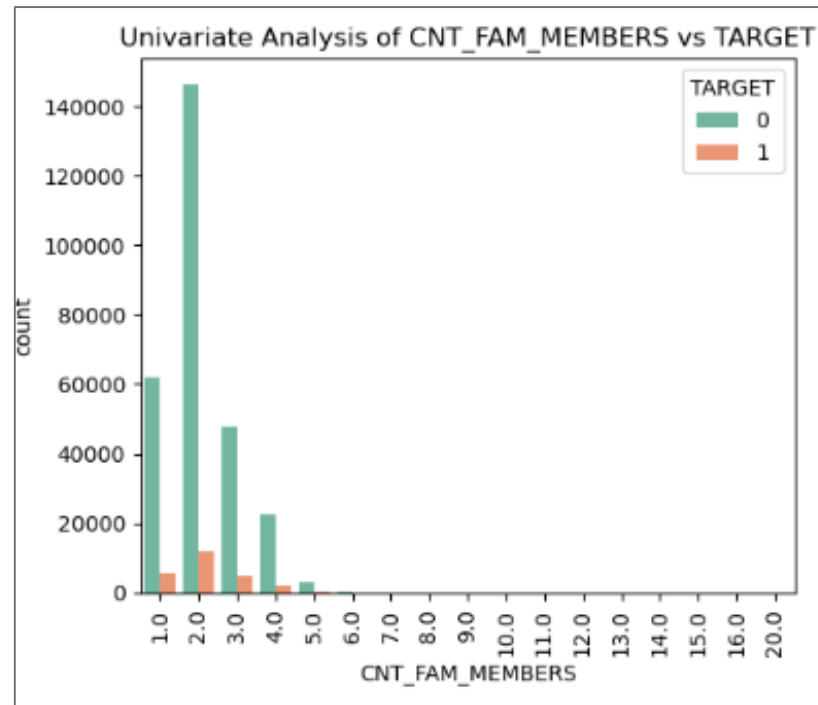
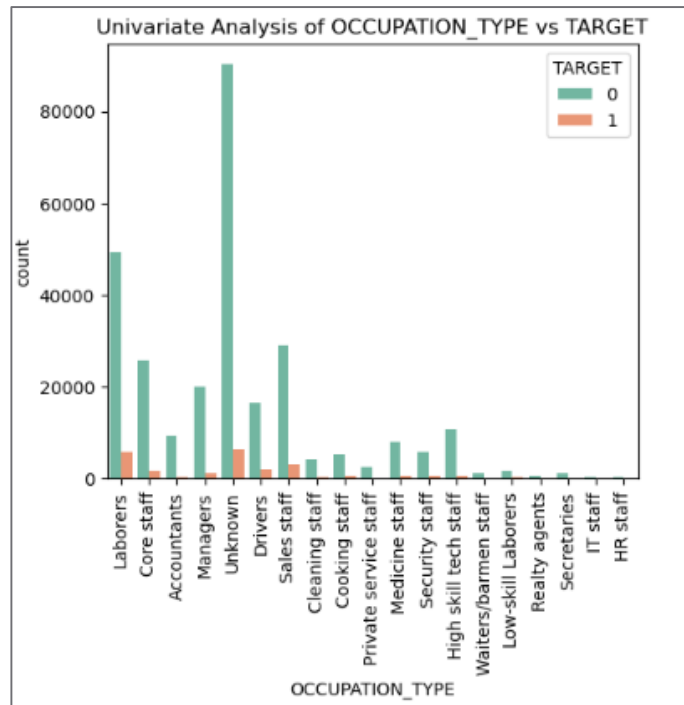
- Clients with Higher Education seem to have less defaults compared to clients who have lower education background.
- Single, widows, and separated clients seems to have minimum defaults.



Insights from Analysis of application_data.csv

Univariate Analysis –

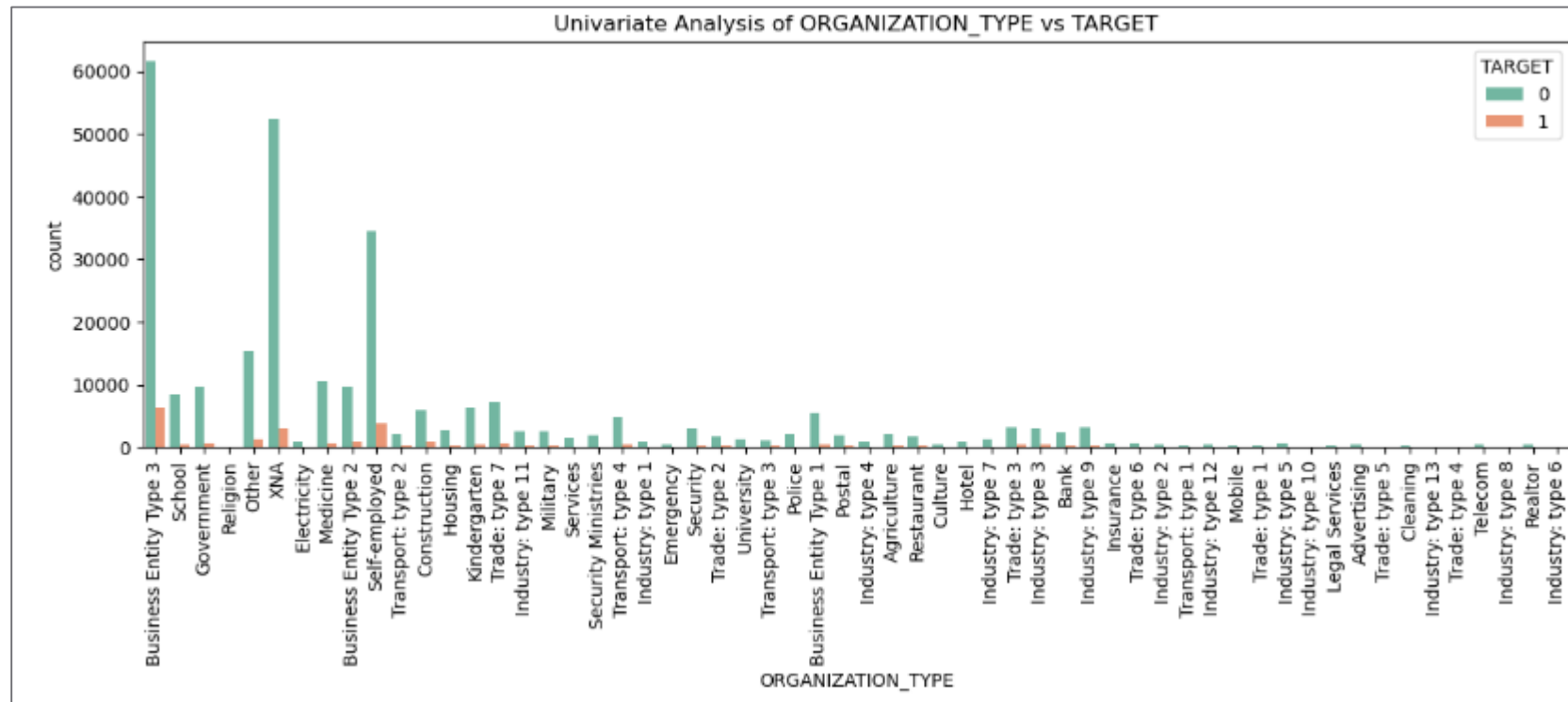
- Laborers seems to have the most number of defaults. On the other hand, Managers seems to have the least number of defaults.
- Default ratio is higher in case of family members greater than 2.



Insights from Analysis of application_data.csv

Univariate Analysis –

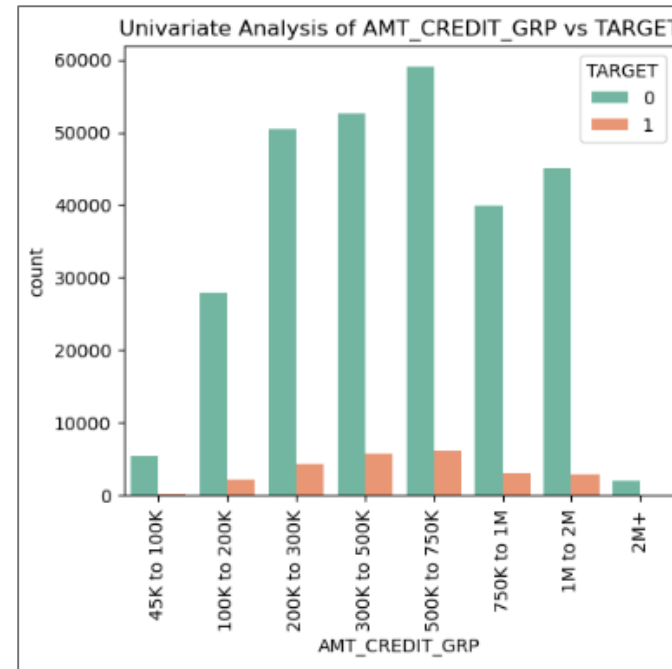
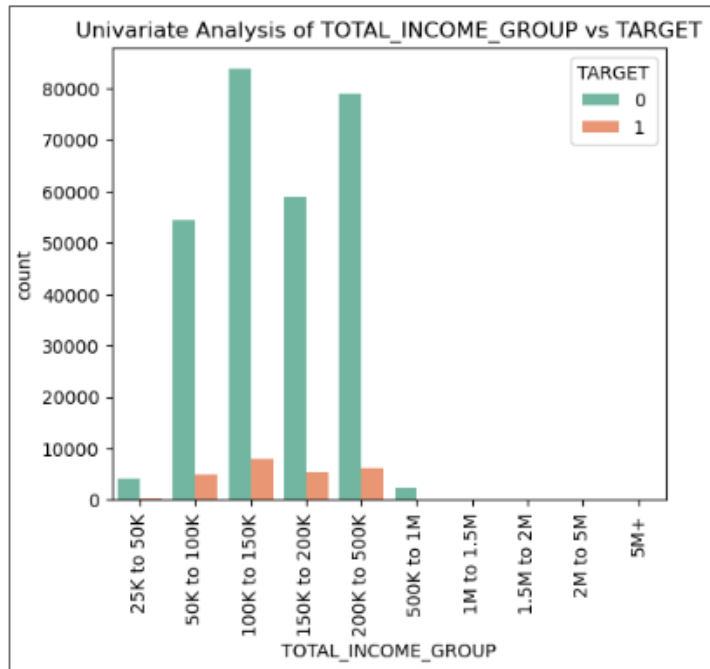
- Other than XNA, Business Entity Type 3 and Self Employed category seems to have the maximum number of defaults.



Insights from Analysis of application_data.csv

Segmented Univariate Analysis –

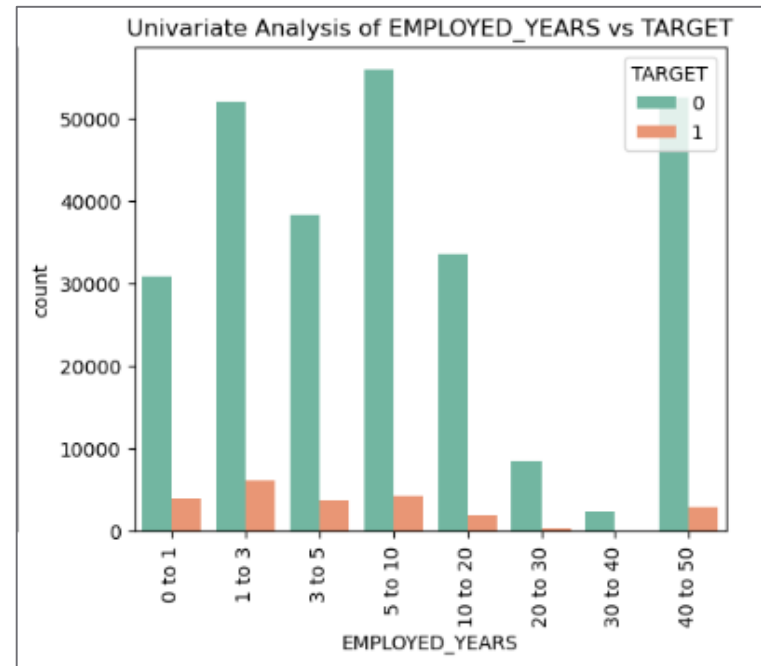
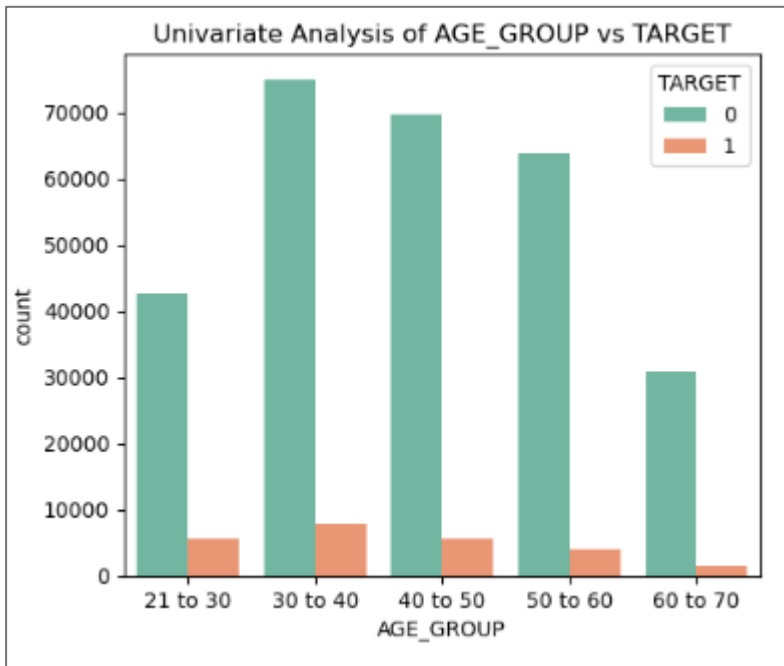
- Clients who have an income range in between 100K – 200K have the most defaults.
- Clients who have a credit amount range in between 200K – 750K seems to have the most defaults.



Insights from Analysis of application_data.csv

Segmented Univariate Analysis –

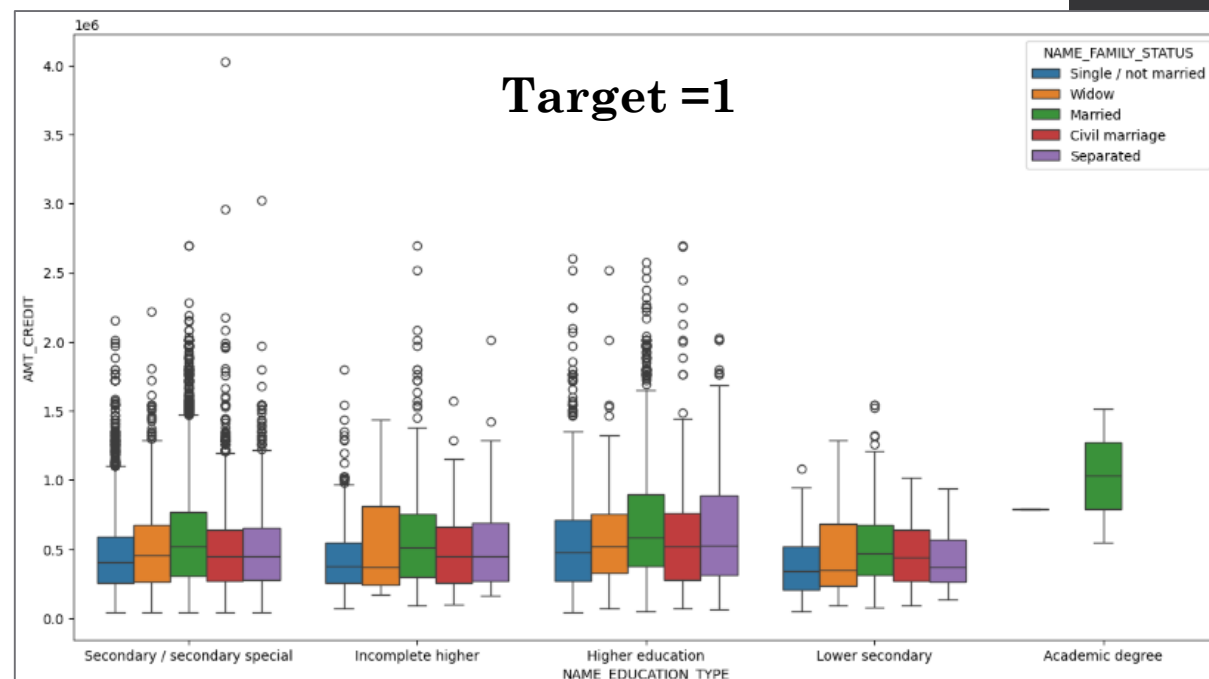
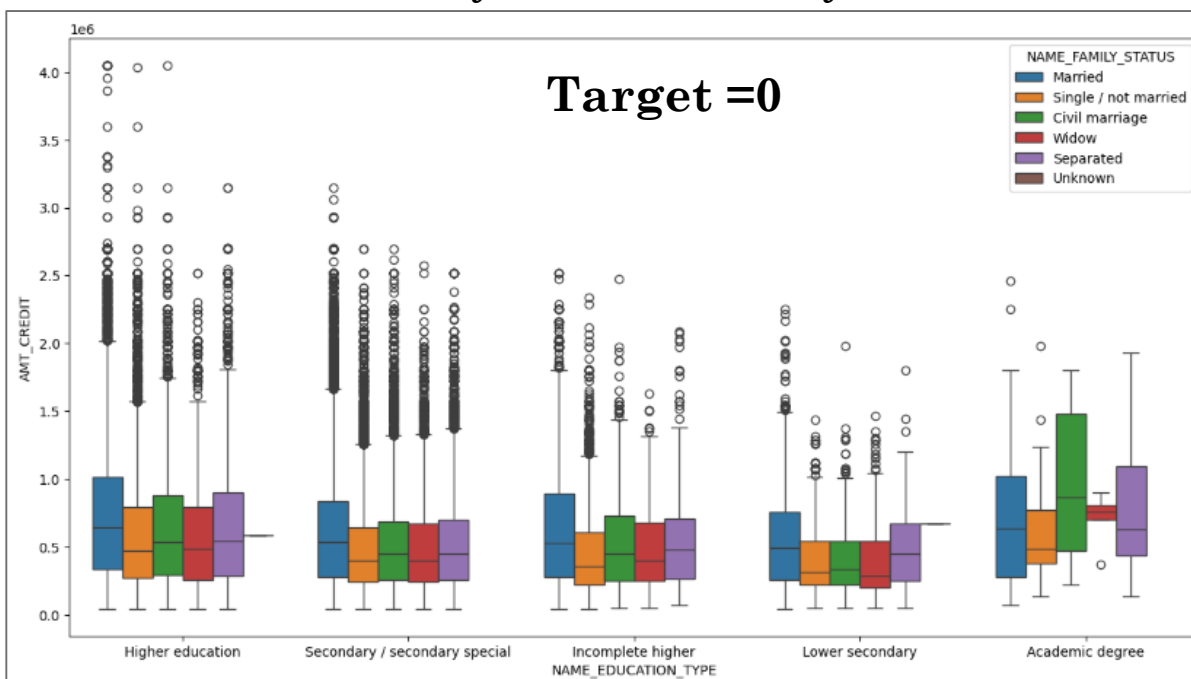
- Least chances of default are from the group of age between 60-70 years.
- Least chances of default are from the group of people who have work experience more than 20 years.



Insights from Analysis of application_data.csv

Segmented Univariate Analysis –

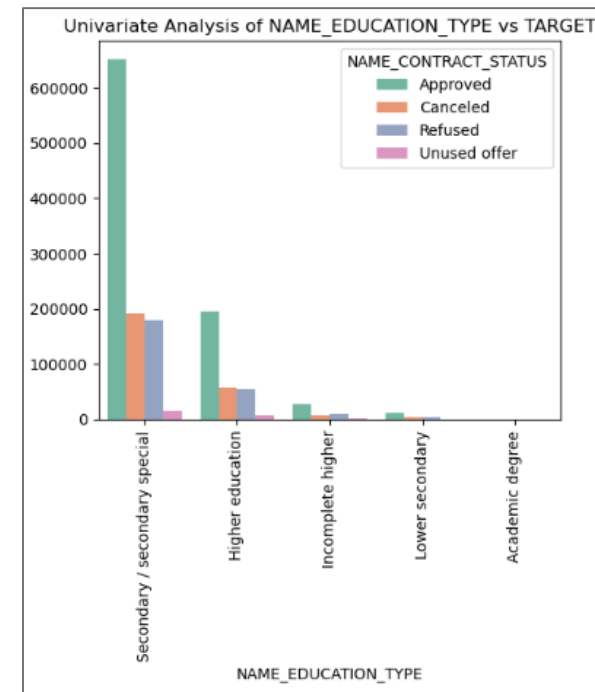
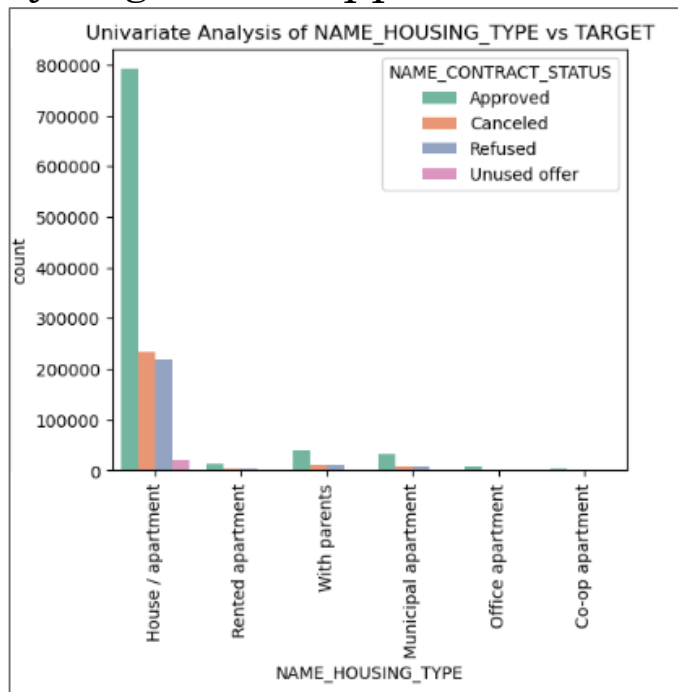
- Clients who are married and have received education above lower secondary are more likely to repay the loan.
- Clients who are married and have not received education above lower secondary are more likely to default the loan.



Insights from Analysis of Merged files

Univariate Analysis –

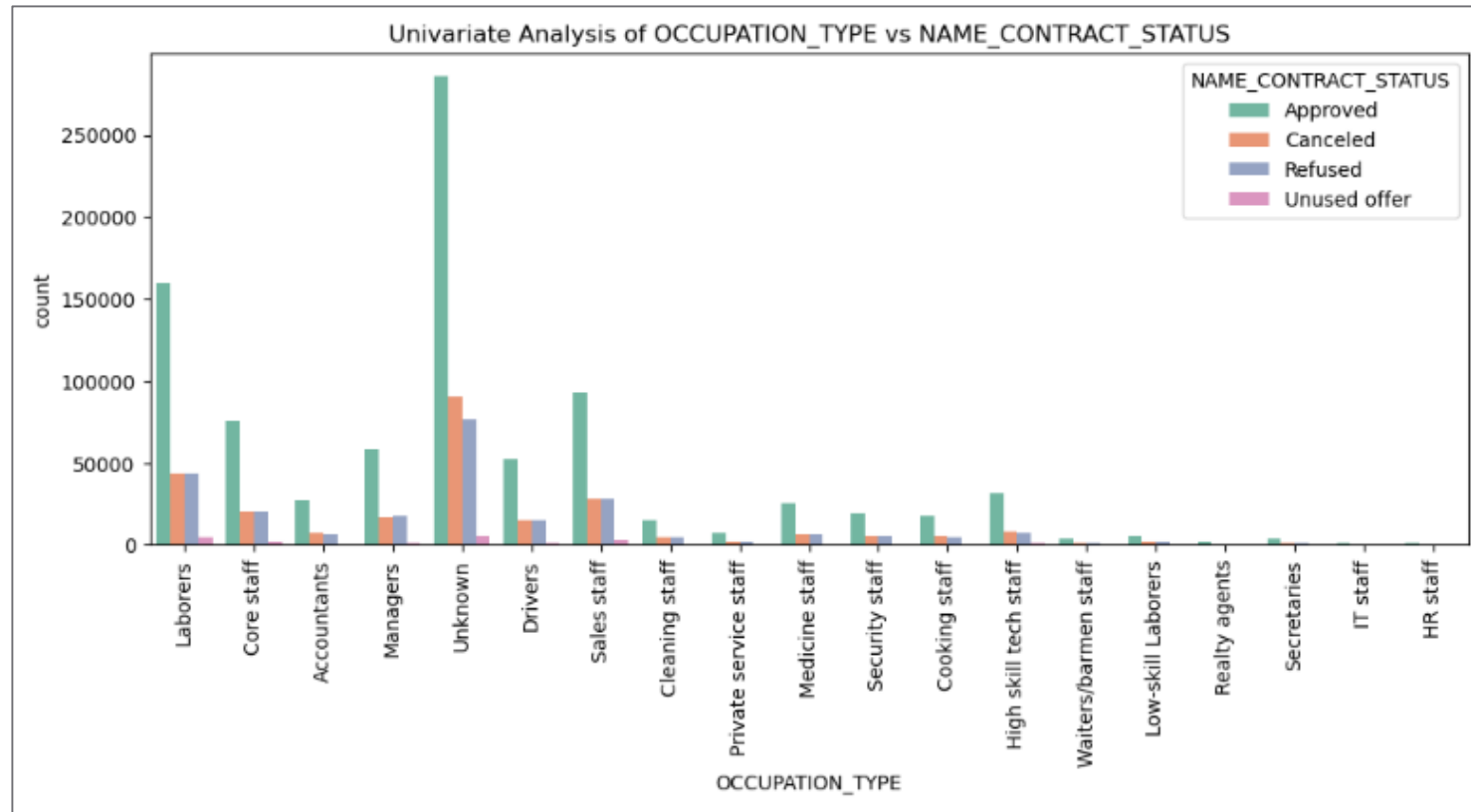
- From the plot it can be seen that owning a house can be a driving factor for loan approval.
- Clients who have completed their Secondary or higher education are more likely to get loan approvals.



Insights from Analysis of Merged files

Univariate Analysis –

- It seems like Laborers, core staff, and sales staff are more likely to receive approval on loan compared to other occupation types.



Conclusion & Observations

Potential categories of clients who might default on their loans -

- If the client is male,
- Is married and has children more than 2,
- Hasn't even completed Secondary education,
- Works as a laborer, or is a business man or is self employed,
- Has a family size greater than 4,
- Has an income in between 100K to 200K,
- Has a credit amount range in between 200K to 750K,
- Has an annuity in the range of 10K to 50K.
- Has age group in between 21 to 50.

Thank You