Exploratory Data Analysis

By Anugata Mahapatra.

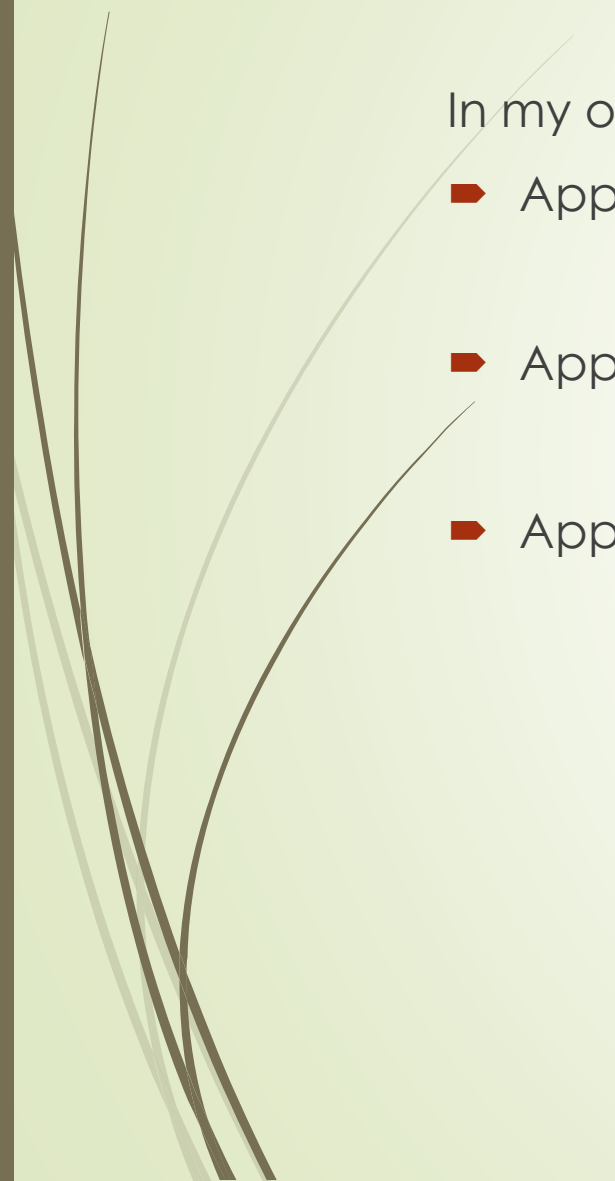**EXPLORATORY DATA ANALYSIS-CREDIT ASSIGNMENT**

Risk Analysis In Banking And Financial Services:

- Identify if a loan applicant has issue paying their installments.

- Actions to be taken, such as, Loan rejection, reduced loan amount.

- Understand the factors which are strong indicators of default.

Assumptions :

In my opinion-

- Applicants with higher education are less likely to default.

- Applicants with lower range of total income are more likely to default.

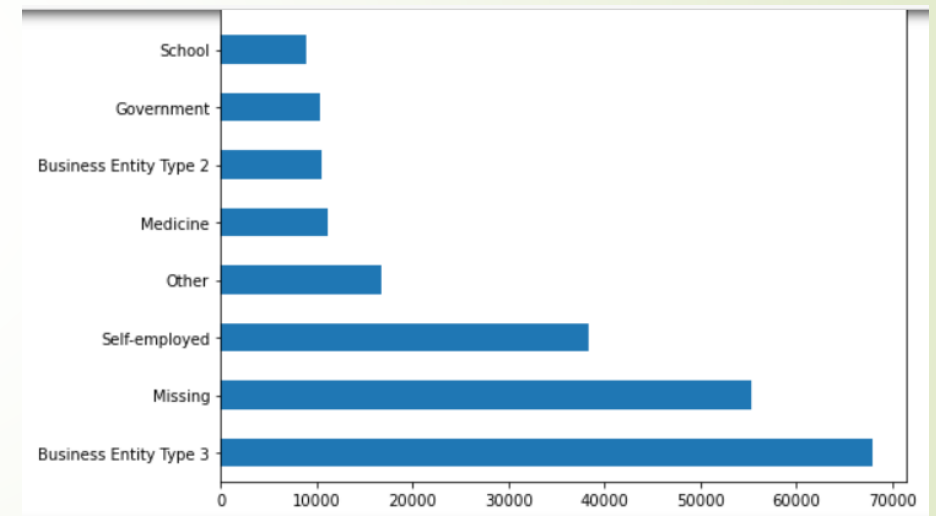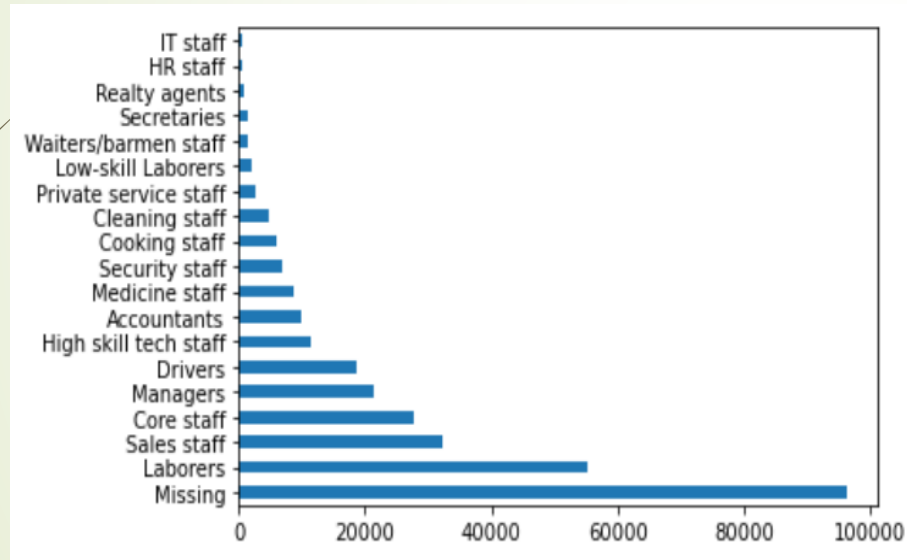- Applicants with a huge family are among defaulters.

Data Imbalance:

-The dataset is imbalanced with 92% non-defaulters and 8% of defaulters as analyzed from TARGET column
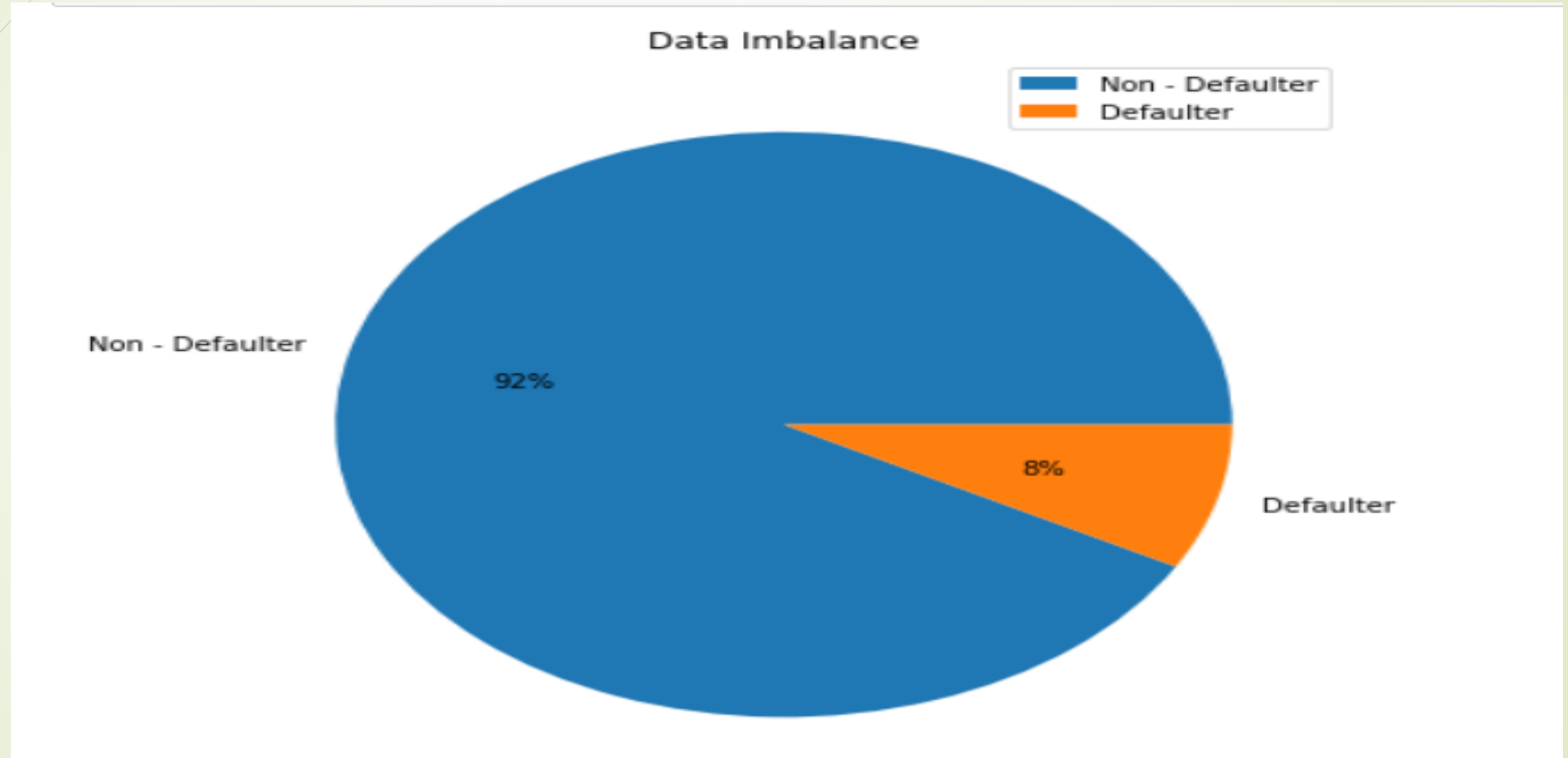
Applicants From Category Occupation and Organization:

-Laborers are the highest applicants(left).
-Maximum applicants belong to Business Entity Type 3(right).

Applicants From Category Credit Range and Total Income:

- Maximum applications belongs to 2L-3L Credit range followed by 10L Above range in the 2nd   rank
-Maximum applicants belongs to the range of 1L-2L Total Income.
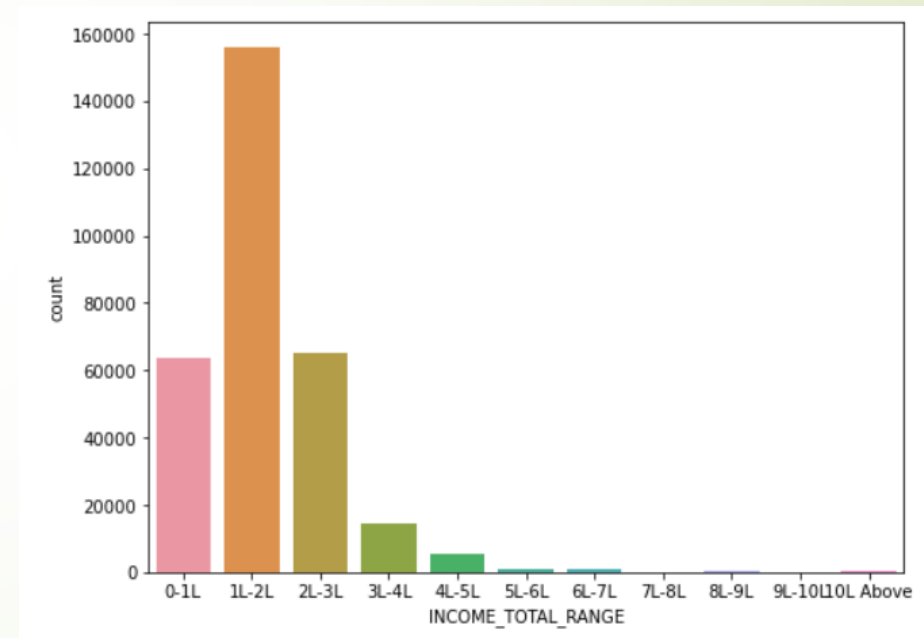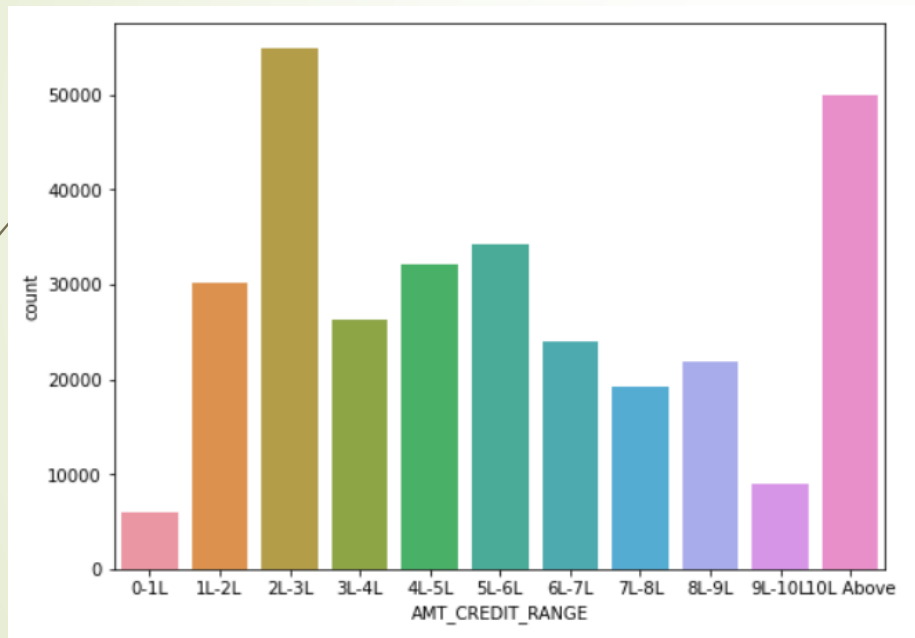
Applicants From Category INCOME TYPE and GENDER:

- Most of the applicants belong to Working type and Commercial Associates
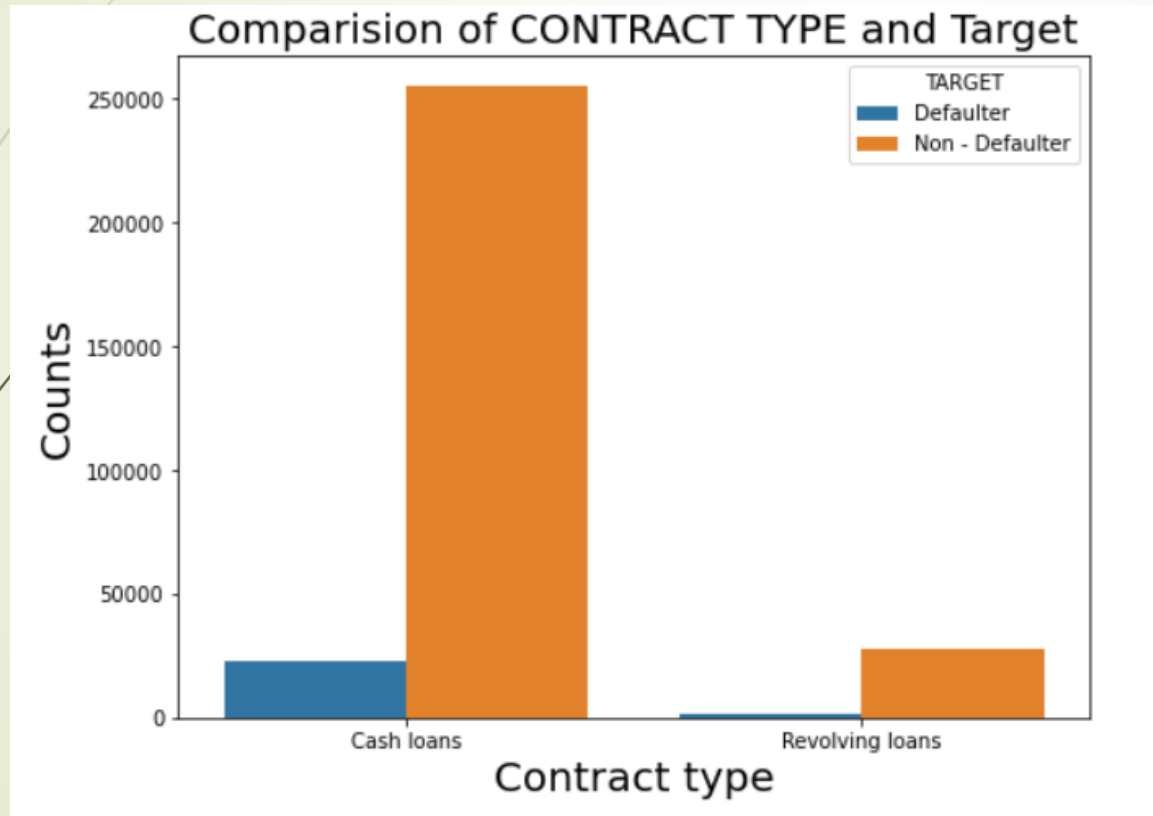- Maximum number of applicants are Females
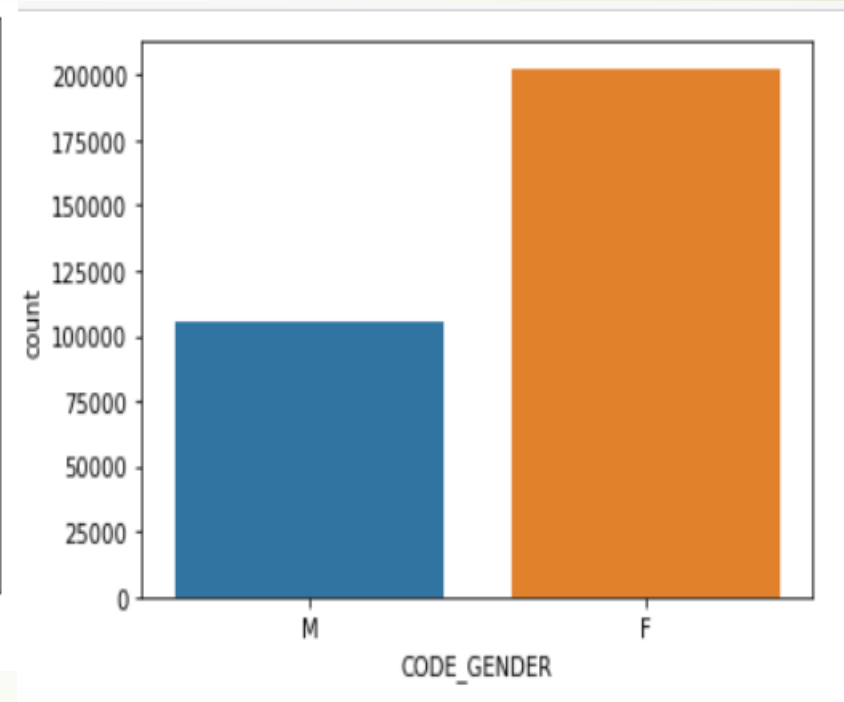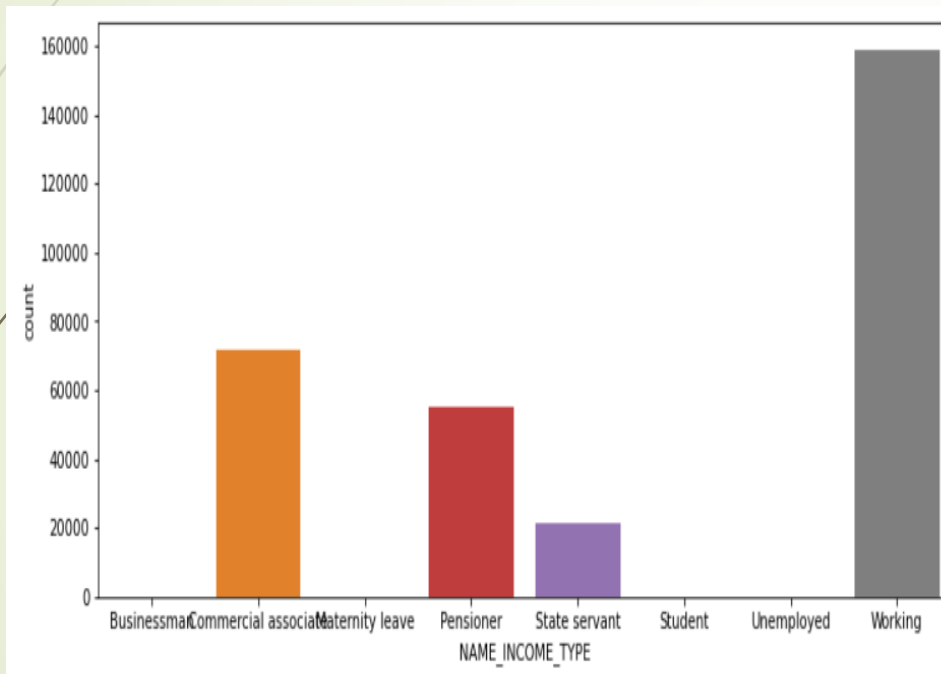
Applicants From Age Range, Education Vs. Target:

- Applicants with age 60 above are less likely to default.
- Most of the Applicants with Secondary Education are Non-Defaulters
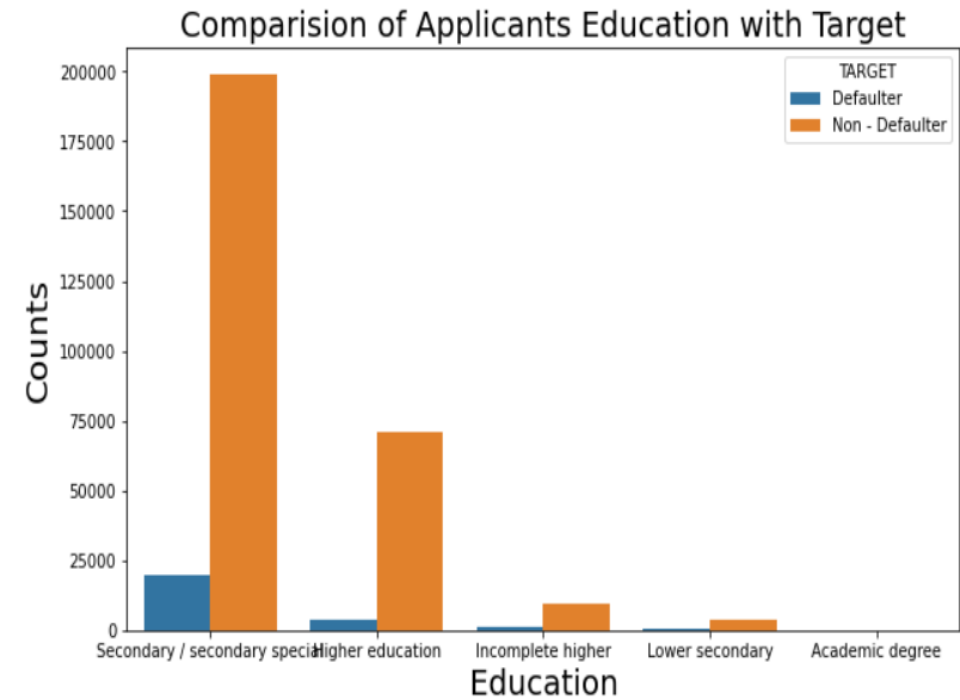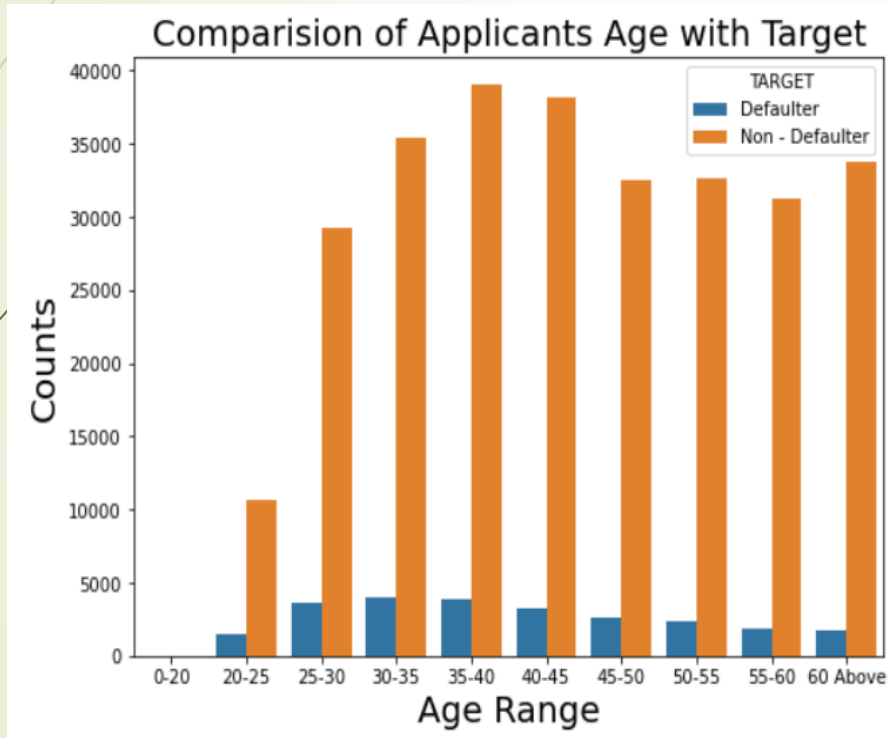
Contract Type Vs. Target:

-Bank should consider approving Cash Loans, as it has the highest rate of Non-Defaulters.

'Application_data' Dataset Analysis:

- As we can see, applicants with secondary education type are the highest non-defaulter applicants. My Conclusion from application_data.csv Analysis: NAME_INCOME_TYPE - Working and Commercial associates are the highest non-defaulters and Commercial associates are the highest defaulters.

- GENDER_CODE - Females have the highest non-defaulters rate as compared to Males

- INCOME_RANGE - People having annual income ranging in between 1L-2L are the highest non-defaulters as compared to people with annual income range of 0-1L.

- CONTRACT_TYPE - Maximum non-defaulter loans were 'Cash loans'.

- AGE_RANGE - Most of the Applicants with age Above 60 are non-defaulters.

- EDUCATION_TYPE - Applicants with secondary education type are the highest non-defaulters.

'Previous_application' Dataset Analysis:

- Maximum records were found with Missing 'Loan Purpose'
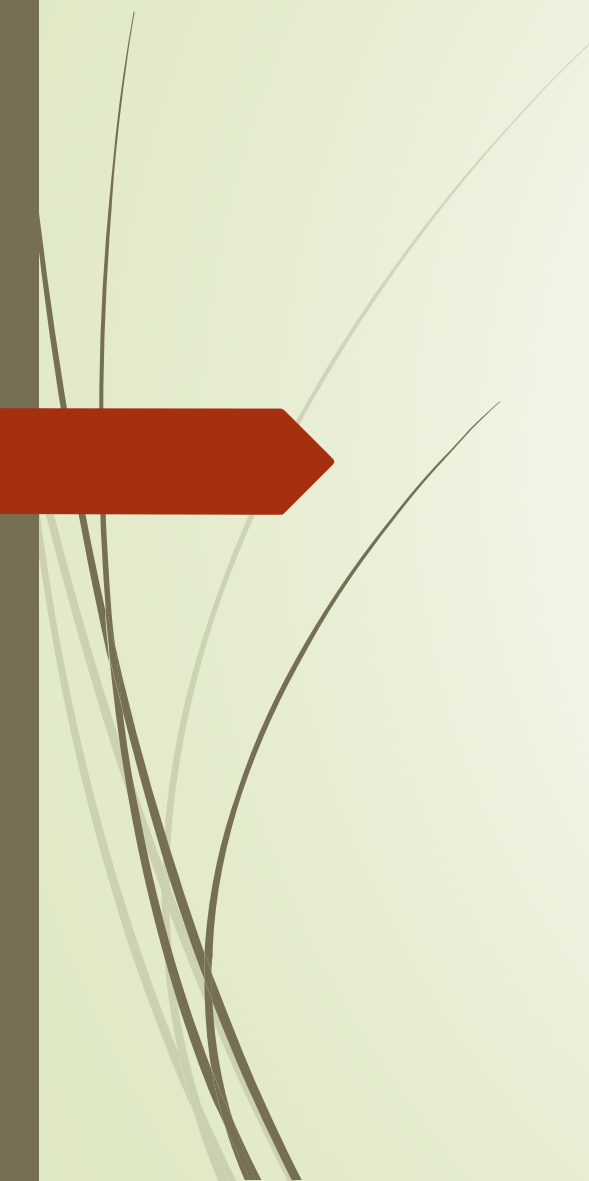-Among the existing records, Maximum applications have the purpose of 'Repair"
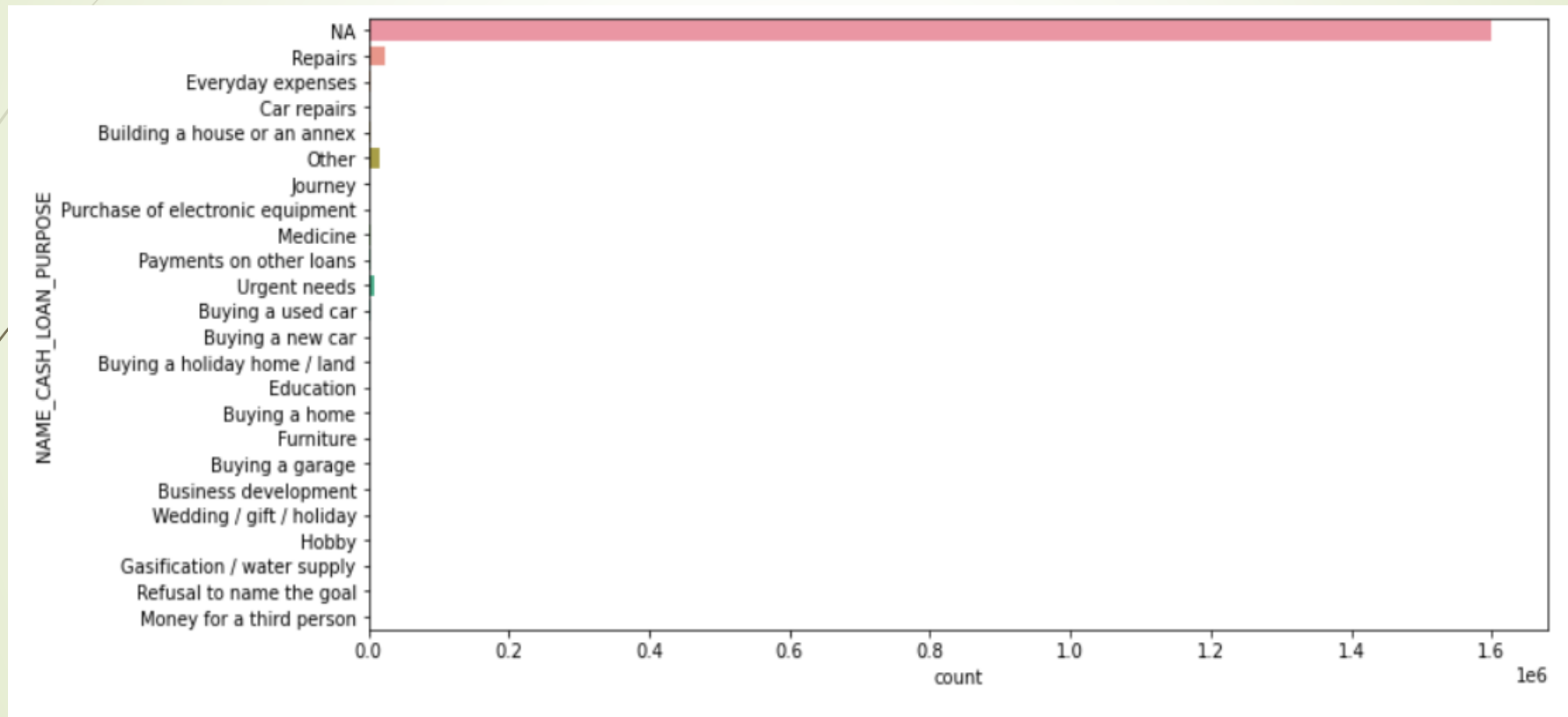
Total Income Vs. Children Count and Total Income Vs. Credit:

-From the left graph, people with more children and less income are likely to default more.
-From the right graph, can conclude that All the defaulters belong to lower income group.
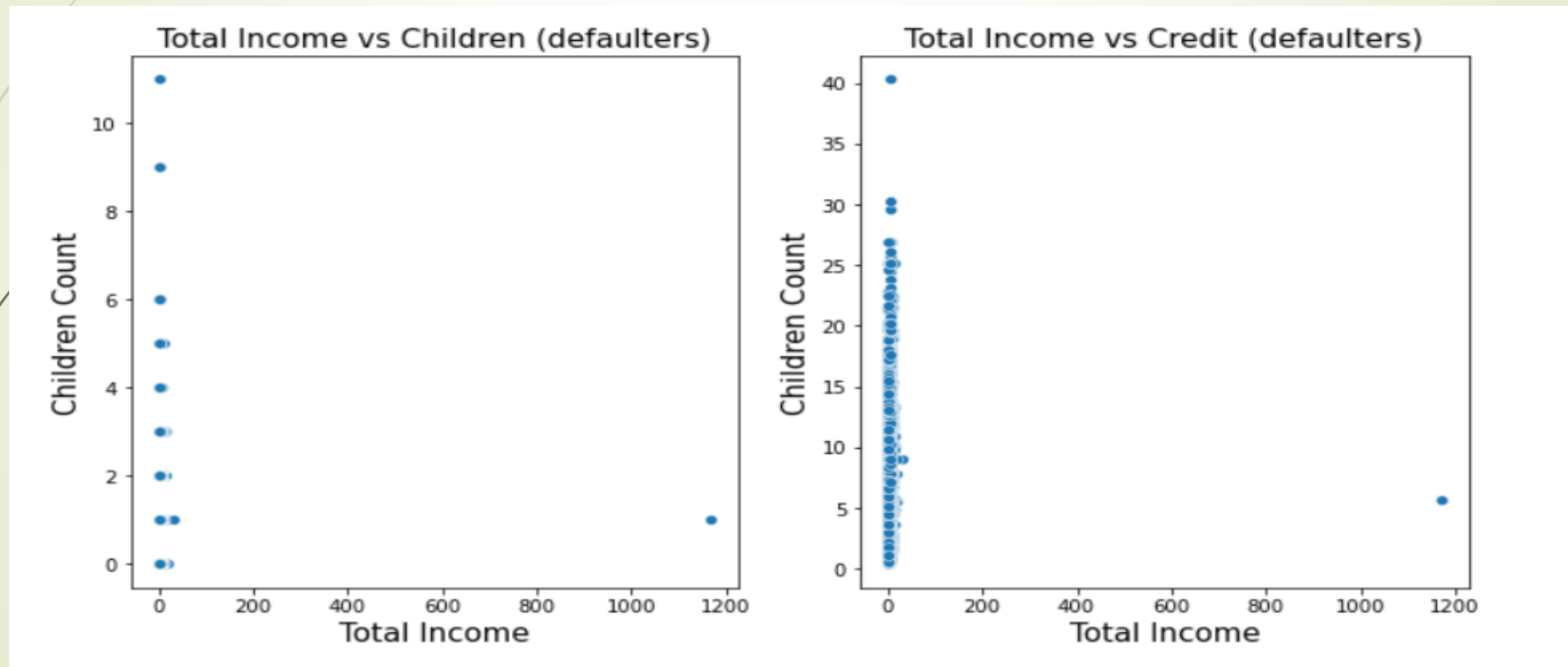
My Analysis:

-People with lower total income are most likely to default. Please, avoid approving these loans.

-Applications with applicant having Lower secondary education should be rejected.

-Maximum People with a secondary education are non-defaulters.

-Working type applicants are less likely to default.

-Applicants around age 30-45 are the highest among non-defaulters.

Thank You!