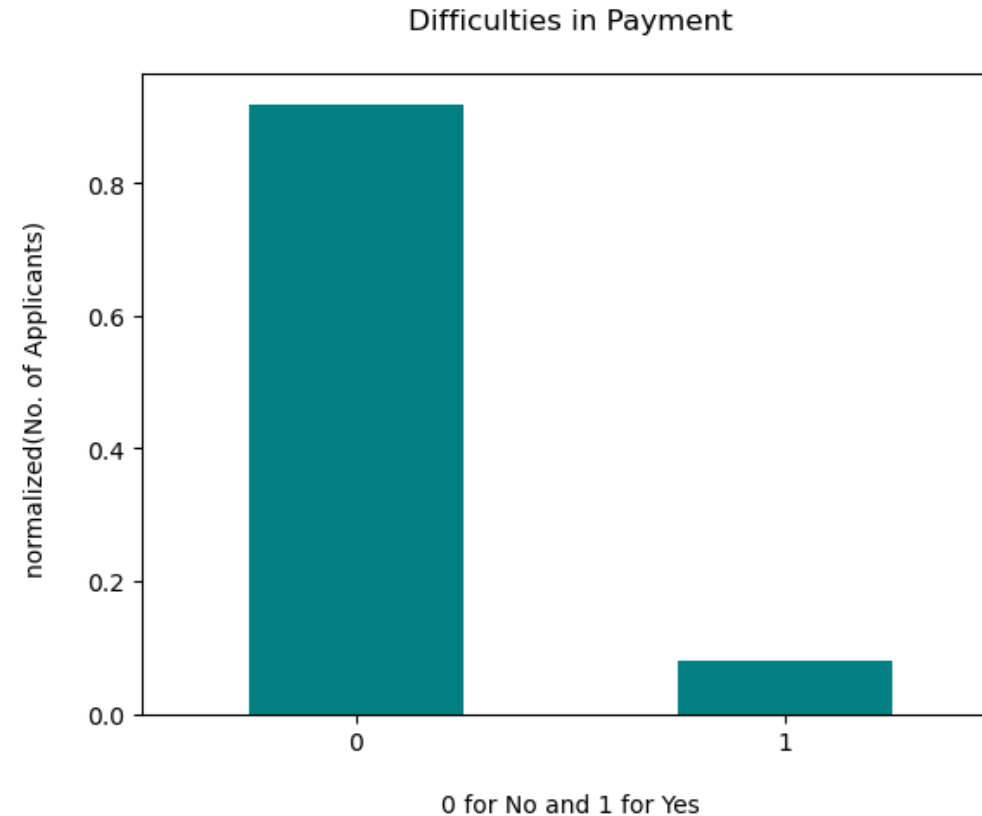


Exploratory Data Analysis for the Loan applicants

Objective : To help banks identify the Potential defaulters and credible applicants

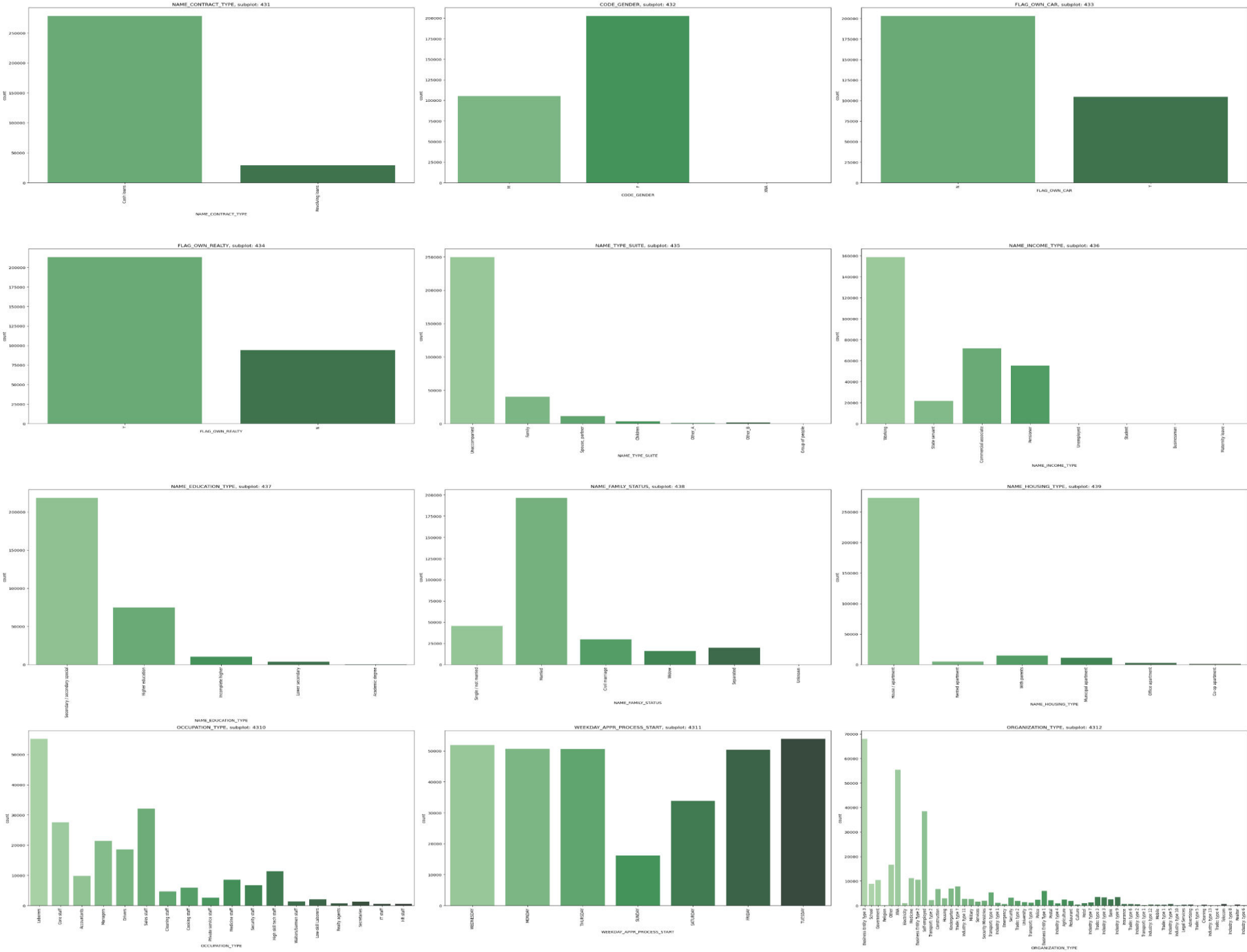
Analysis and Insights

The Provided dataset seems to be highly imbalanced based on Target Variables (0 being credible and 1 being defaulters)

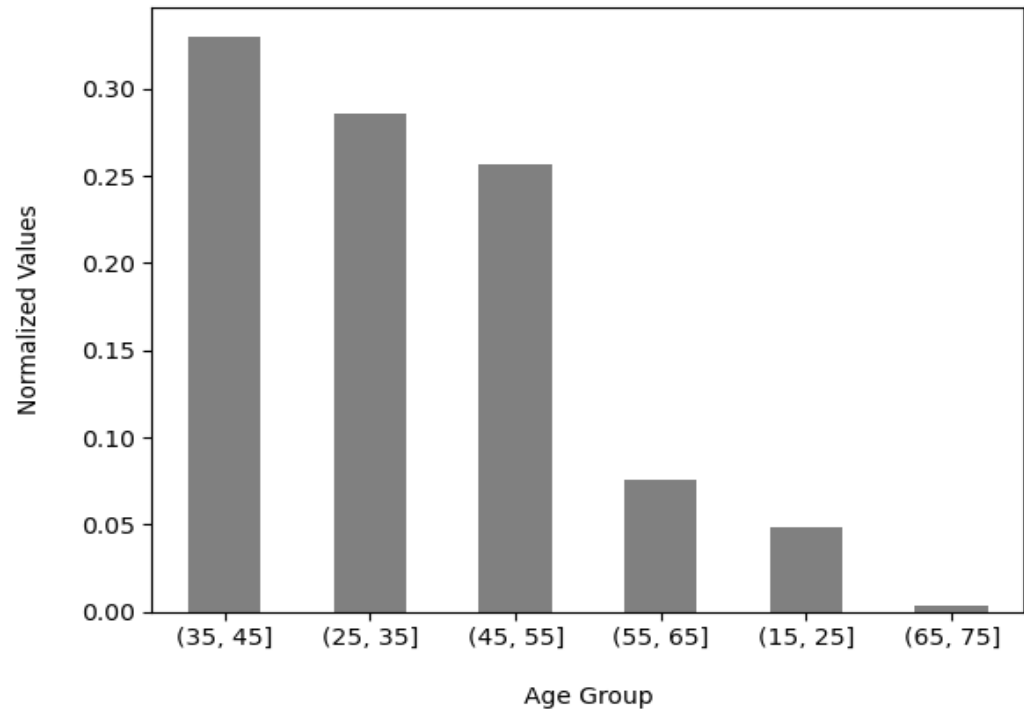
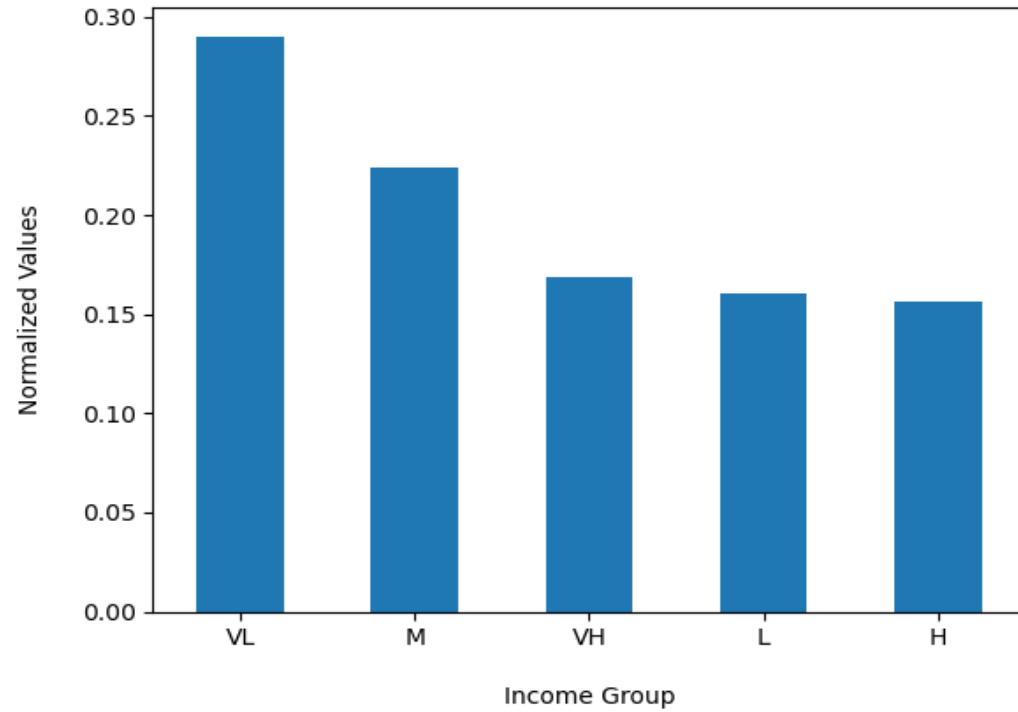


The above Plot and Ratio suggest that 1 out of every 11 applicants is going to have difficulties in repaying the loan

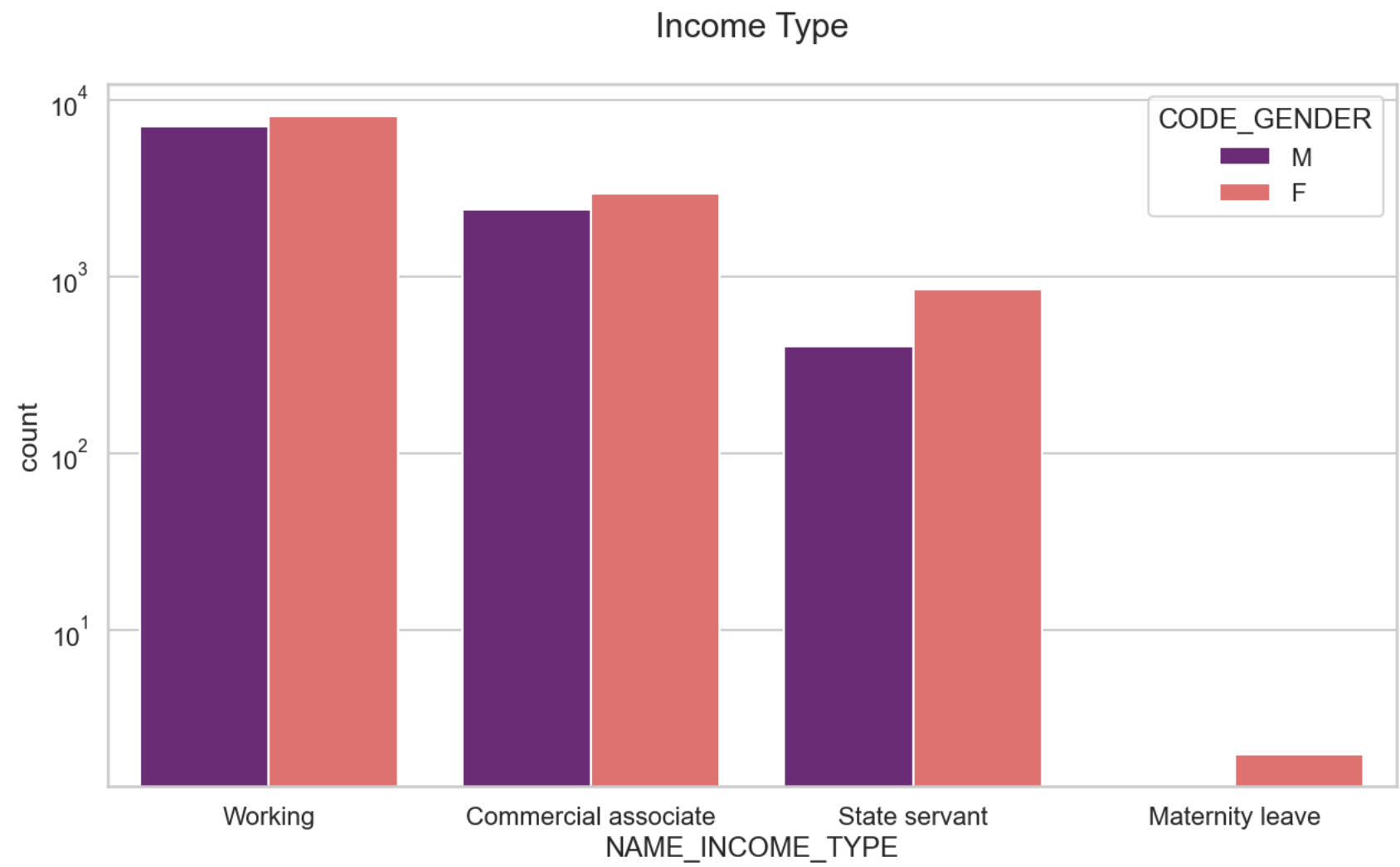
- Following Plots suggest :
- 1. Cash loans are highly preferred
 - 2. Majority of applicants are Married
 - 3. Working people are taking more loans
 - 4. Close to 60% applicants don't own a car



- . Very Low and Medium income groups tend to apply for loans
- . People between age 25-45 tend to apply for loans

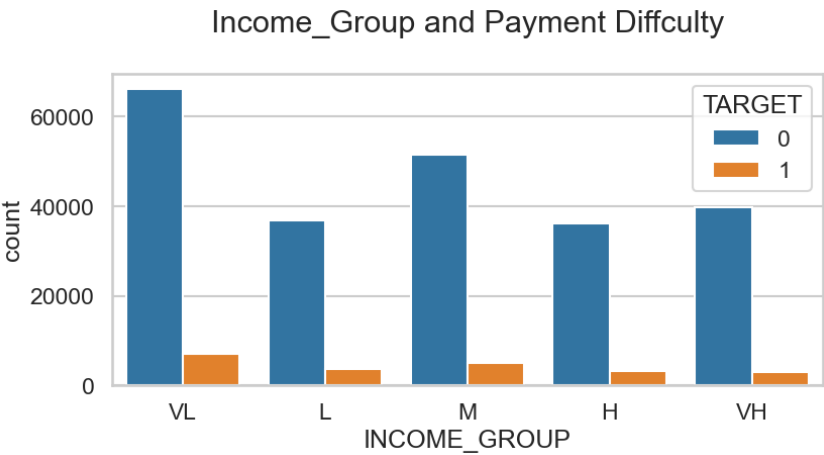
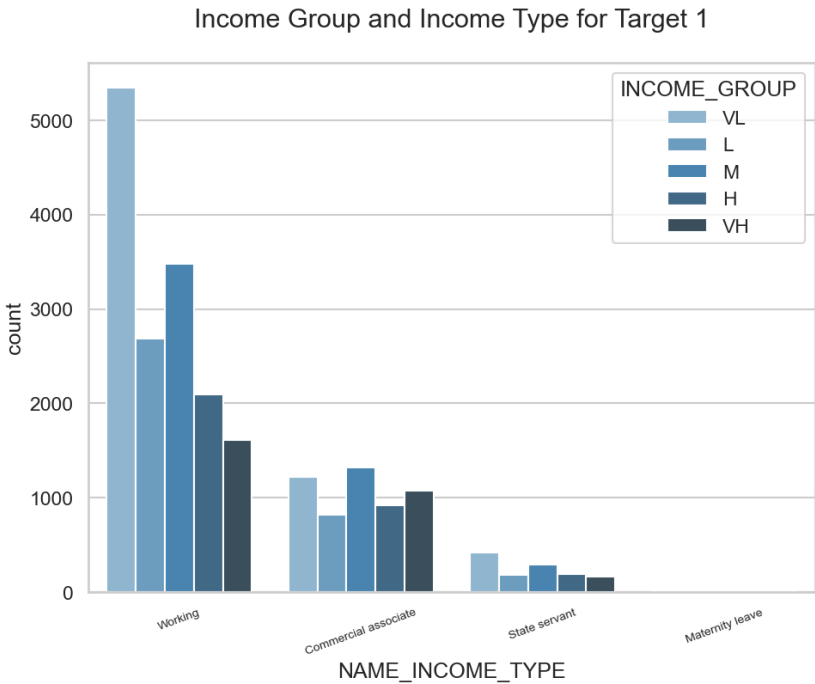
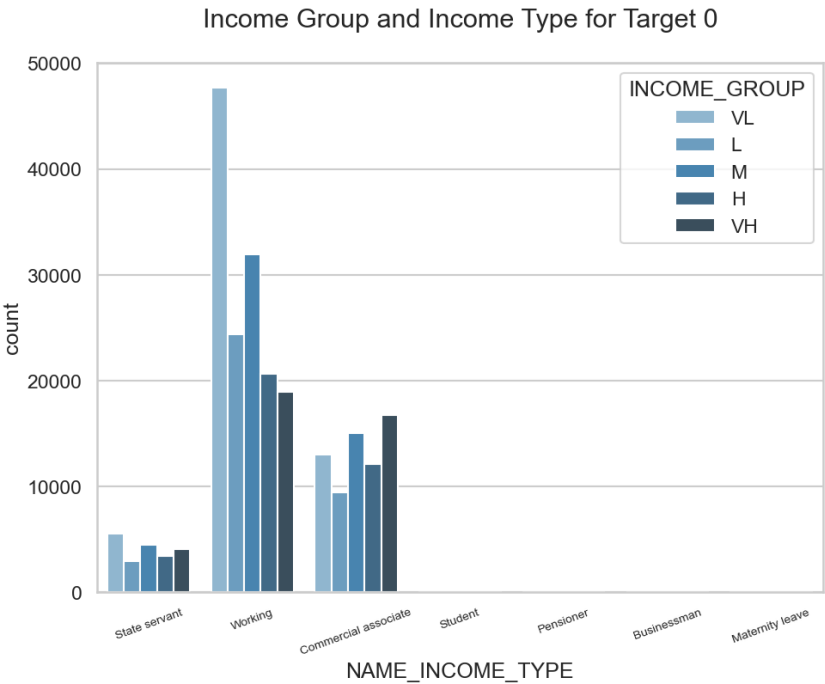


Applicants belonging to working class tend to default more on loans following by Commercial Associated and State servants



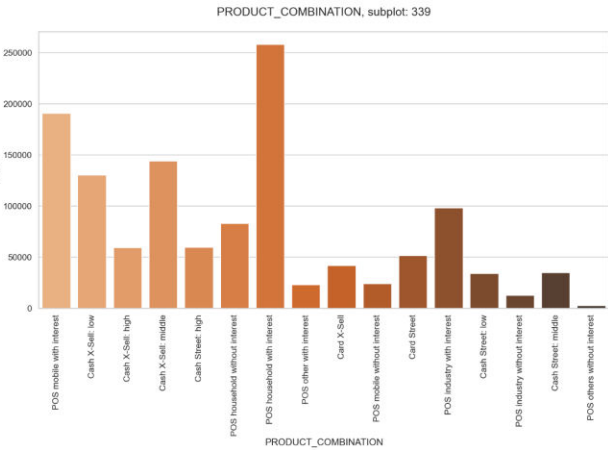
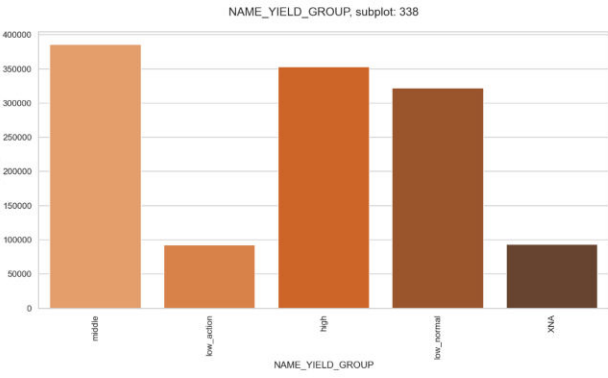
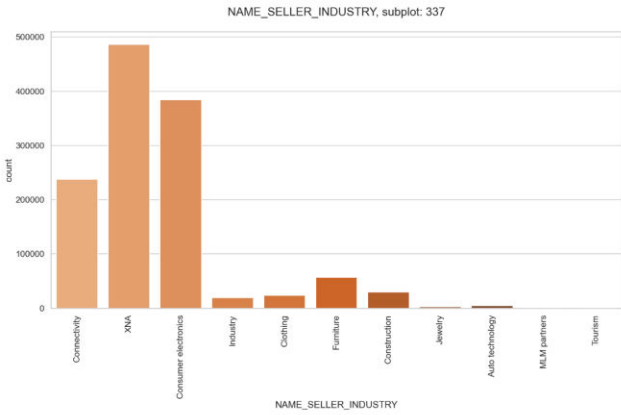
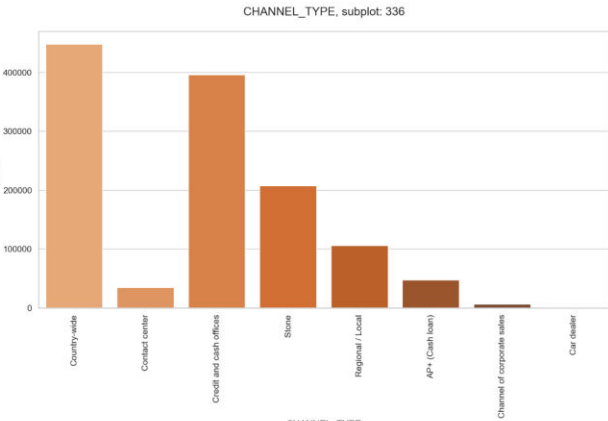
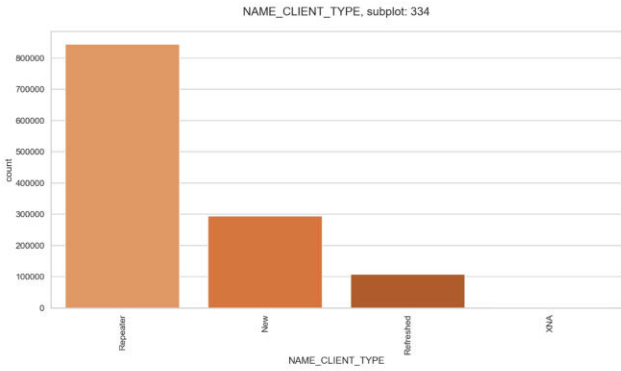
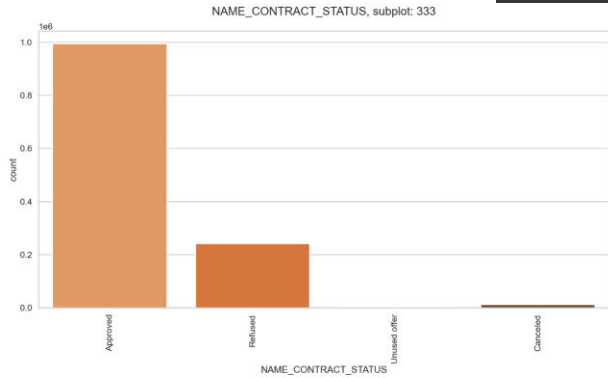
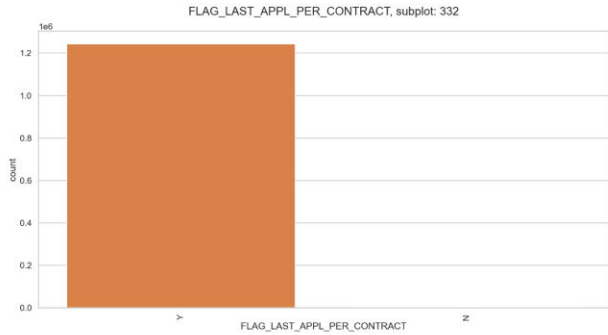
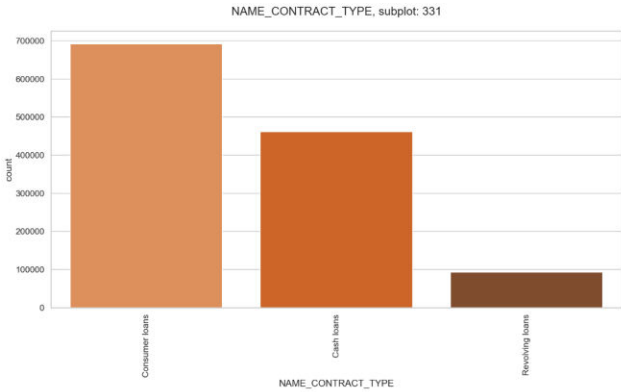
Following Plots shows the Income Group with respect to the Payment Difficulty

Applicants from the Very Low followed by the Medium income group tend to default more

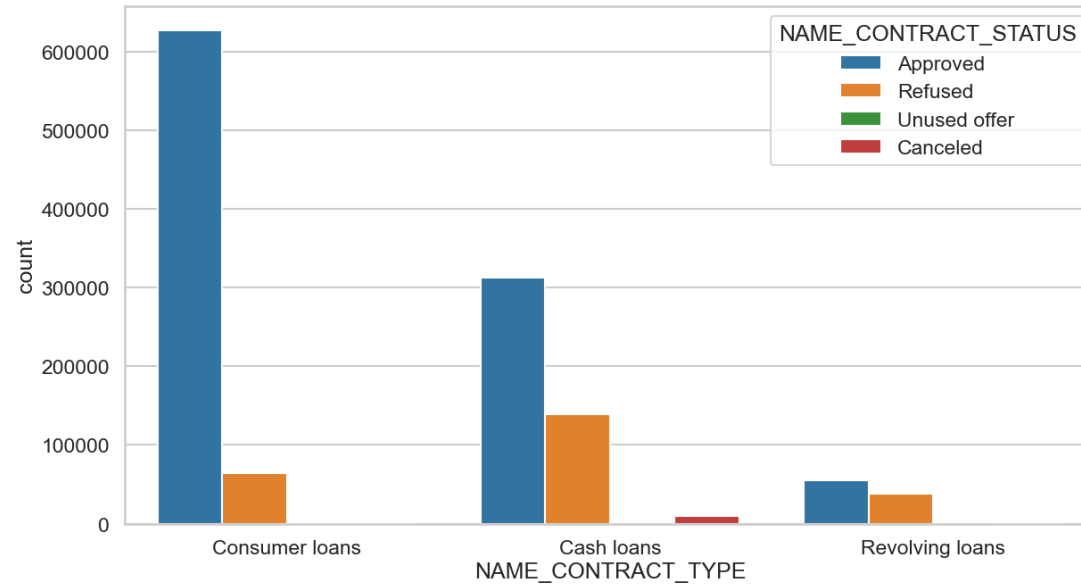


Plot of the categorical columns

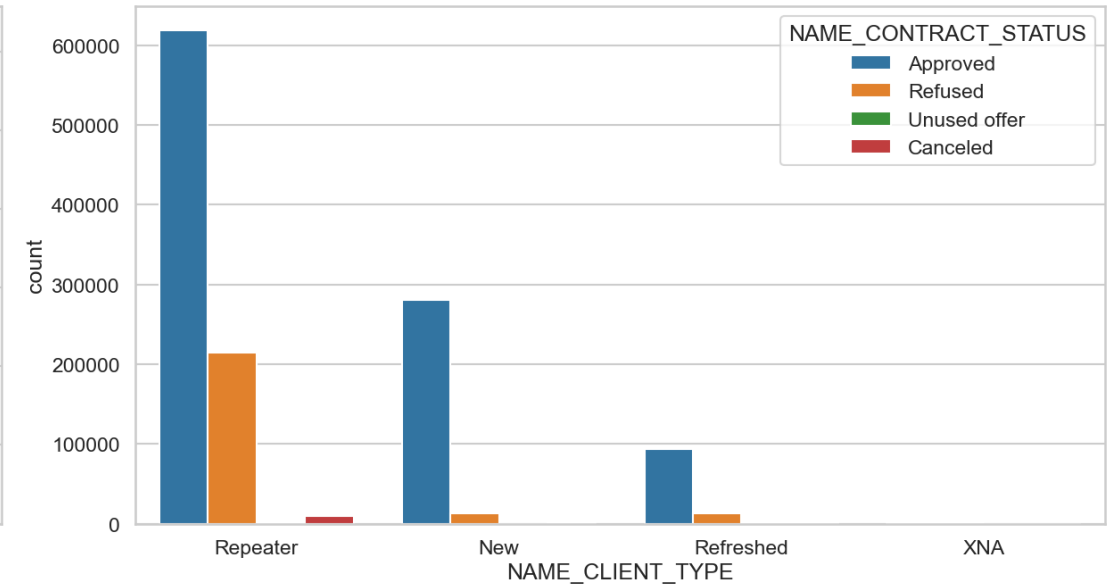
- 1. 55% of loans are consumer loan which was absent from the application.csv, 37% are cash loans and rest are revolving
- 2. Most records of approved loans(roughly 80%) and rest data is distributed in refused, cancelled and unused
- 3. Many are repeaters, NAME_CLIENT_TYPE also has some null values showing as XNA as per the plot
- 4. 55% of the applicants have taken loan for POS purchase.
- 5. Name seller industry has 37% XNA values, Consumer electronics is next highest category at 30%



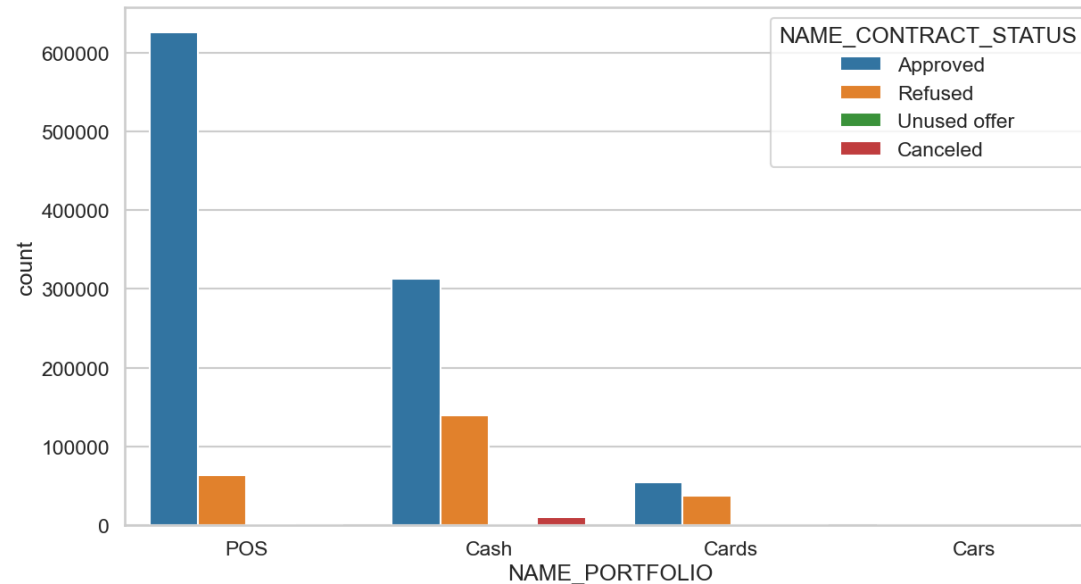
NAME_CONTRACT_TYPE



NAME_CLIENT_TYPE



NAME_PORTFOLIO



- Consumer loan has the largest share of applications irrespective of the status of loan
- Unused offer of the loan seems non existent at the moment
- Repeated applicants seem to have some refused status loan suggesting bank denied due to defaulting or already a loan repayment going on
- Cash loan refusal rate seems to be higher
- POS transactions seem to be consumer loans

Business model for Defaulters

1. Very low income groups followed by medium income groups tend to default more
2. Age group of 25 - 35 followed by age group of 35 to 45 tend to default more
3. Male applicants defaulting more as compared to the female applicants
4. Unemployed tends to default more which makes sense
5. Laborer's, Salesman, Drivers are some of the jobs that are more prone to defaulting
6. Business type 3 is the organization type that has defaulted in the past
7. Defaulters usually don't own real-estate
8. No of Bureau Hits in the past, zero or less hits is good
9. Income not matching to the Good Bought is a concern
10. Data suggested that previous applications that were not approved the loans can also default on their payments which is a concern. This indicates that the bank/financial company had earlier denied the loan and now approved it and is facing default in case of these applicants.

Business model for Credible applicants

1. Bank offered a lower credit/loan amount on the unused offers, that seems to be the case why offers are unused
2. Female applicants are less favorable to default and hence should be give a higher priority
3. Even though working applicants tend to default more (roughly 60%). still for the rest 40%, it feels business opportunity denied to not consider them. Proper scrutiny and parameters need to be enforced on these cases.
4. Previously denies applications also have cases where payments are coming on time in current application. This indicates that possibly wrong decisions were done in those cases.