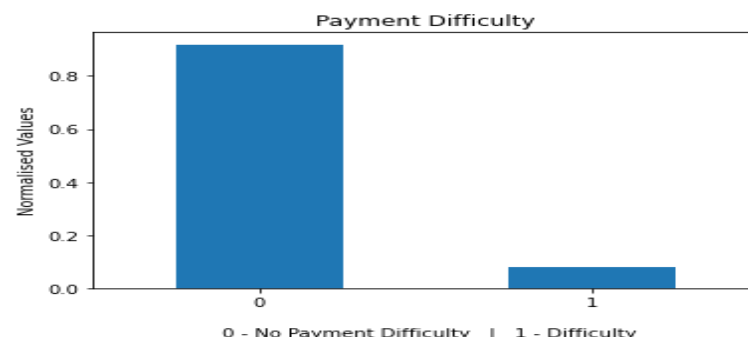
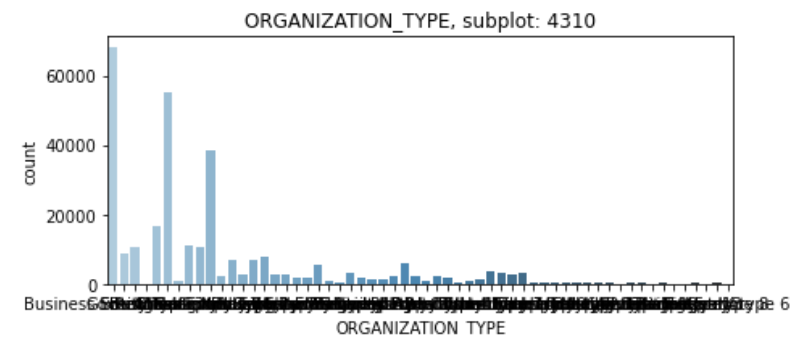
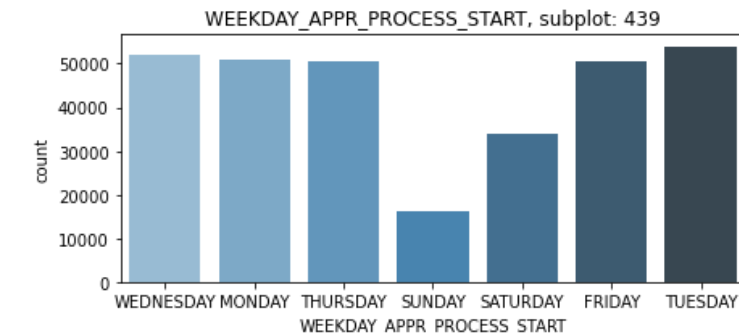
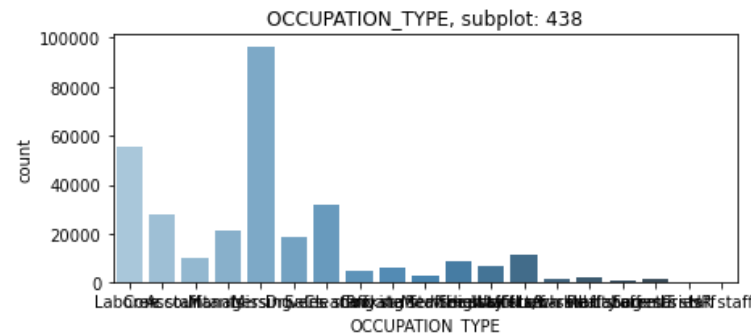
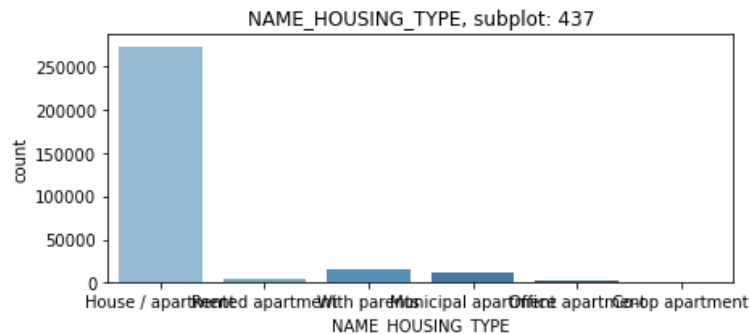
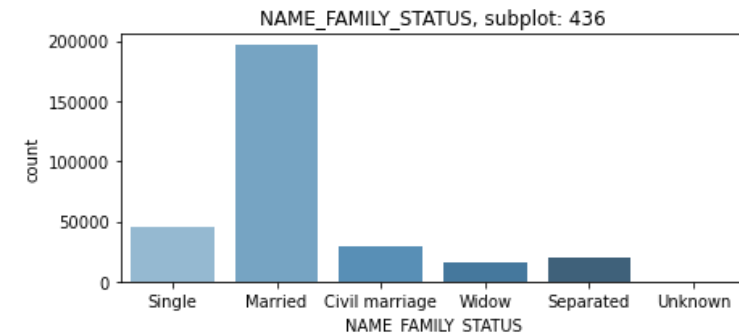
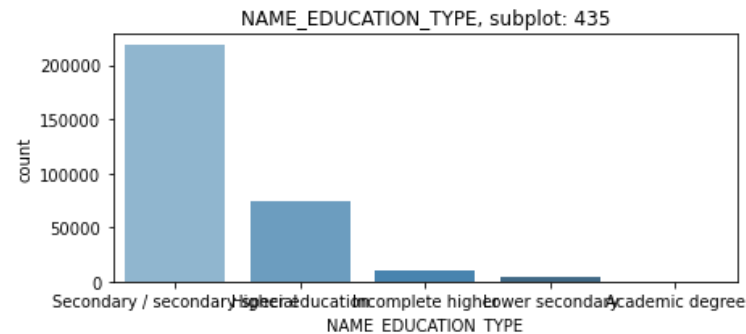
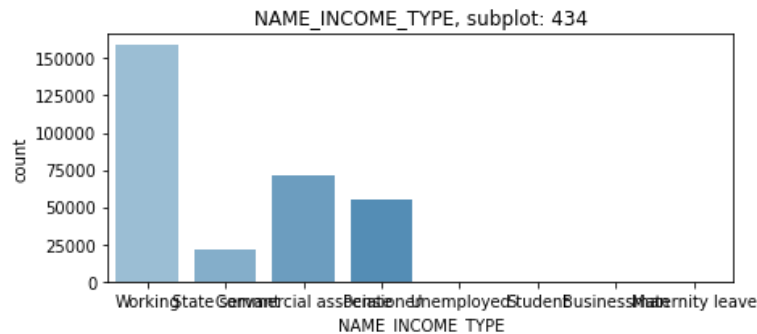
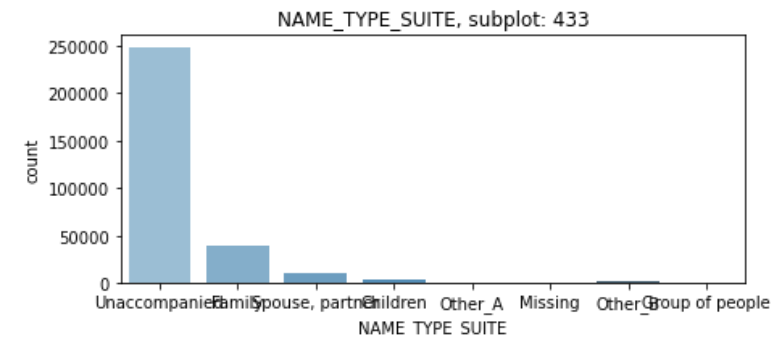
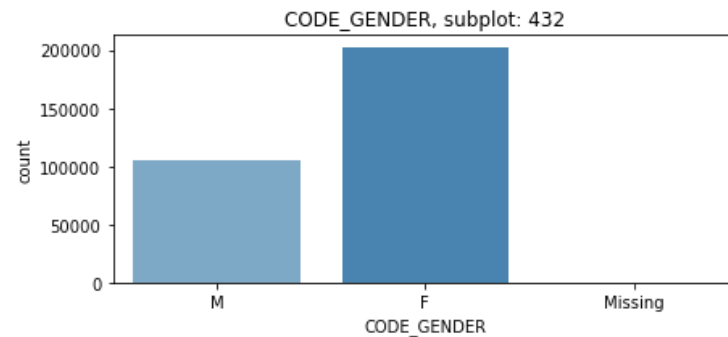
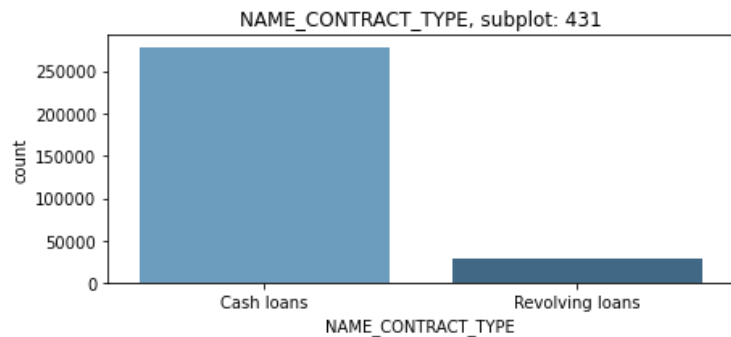


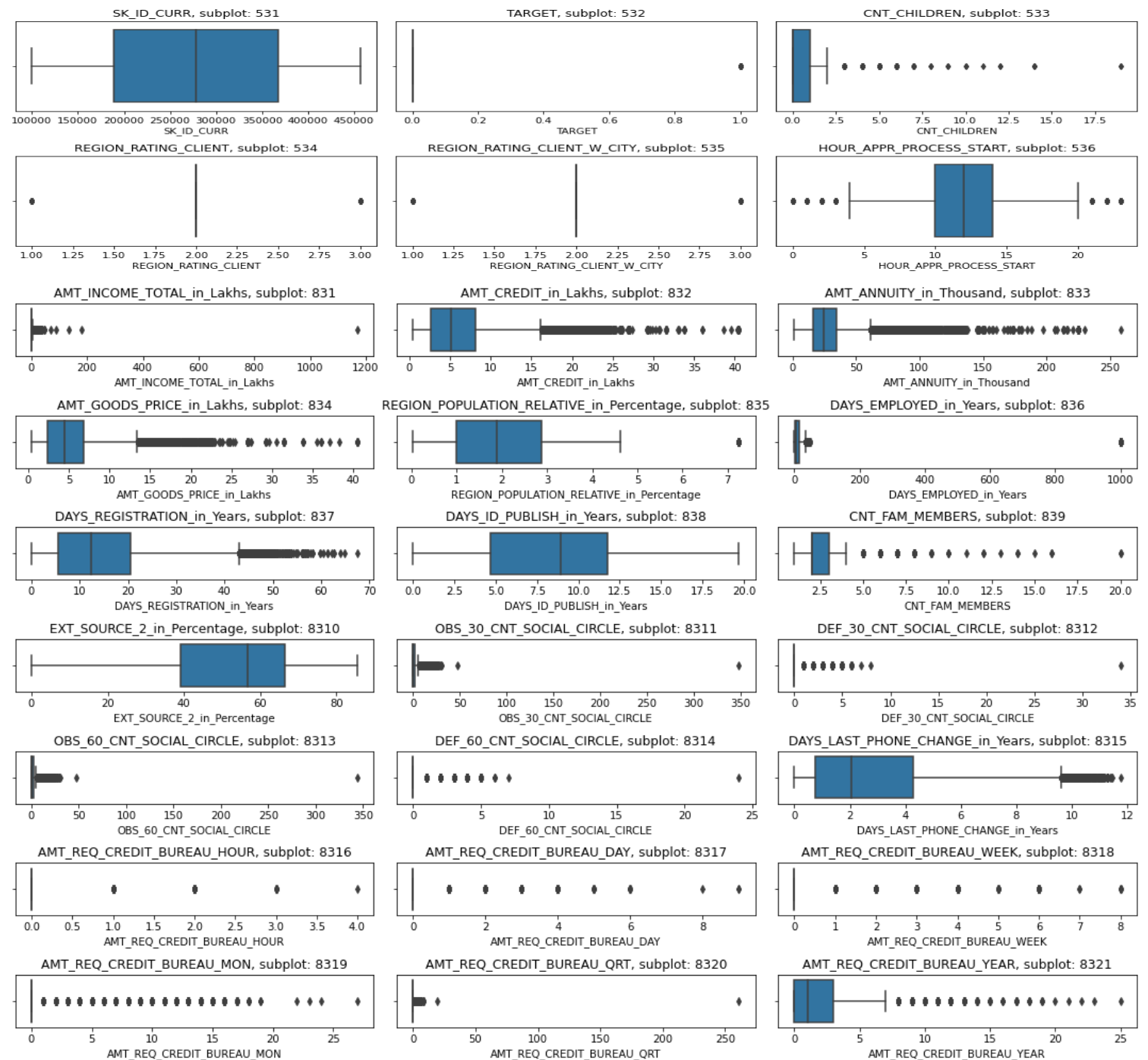
CREDIT EDA ASSIGNMENT

NAME : DEVANG Y. CHUDASAMA
BATCH : SEPTEMBER DC50
COURSE : EPGPDS

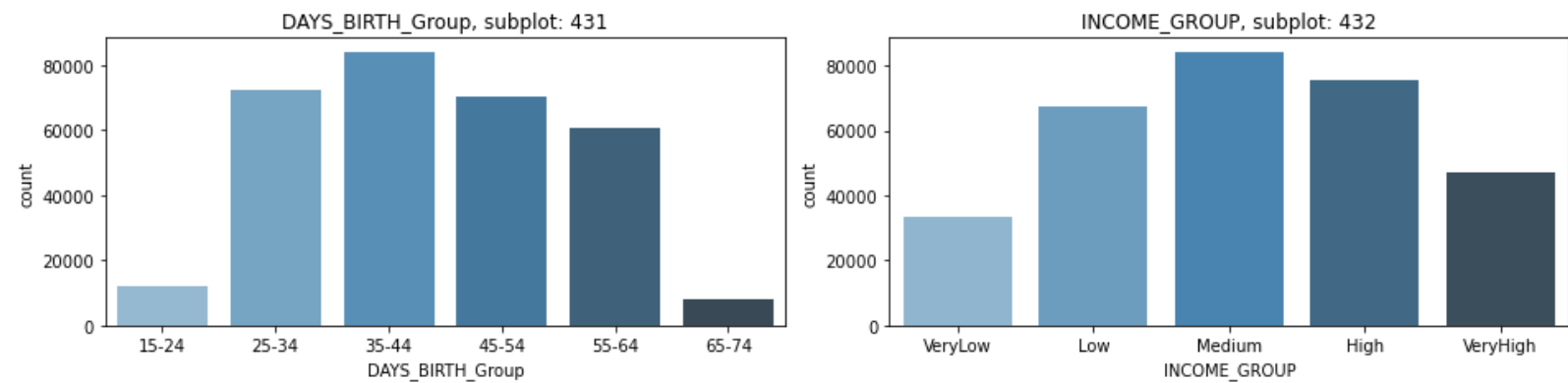
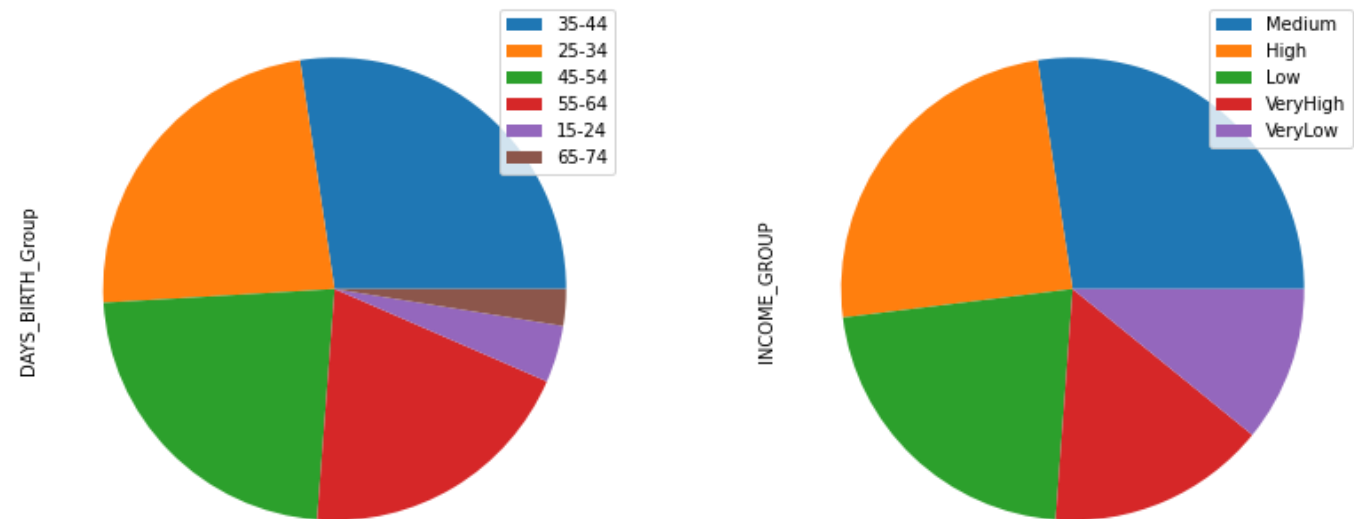


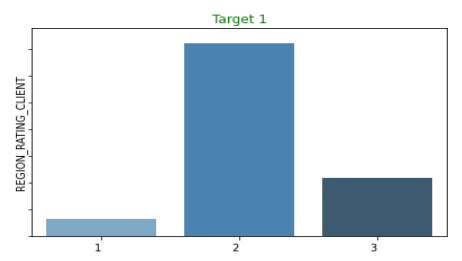
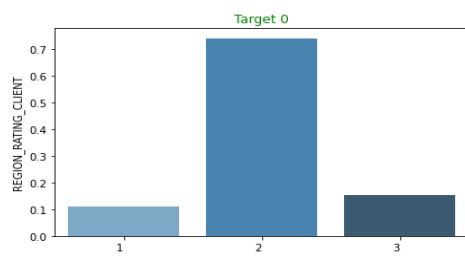
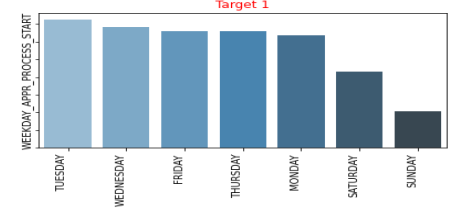
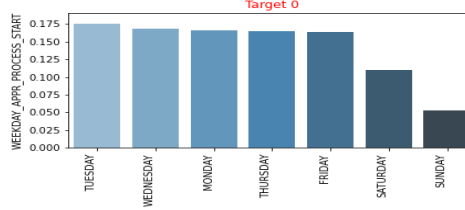
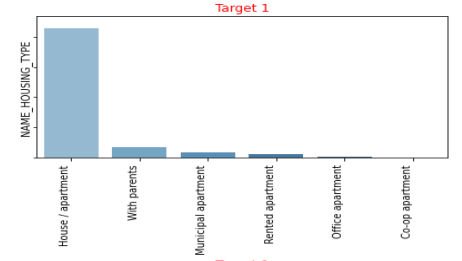
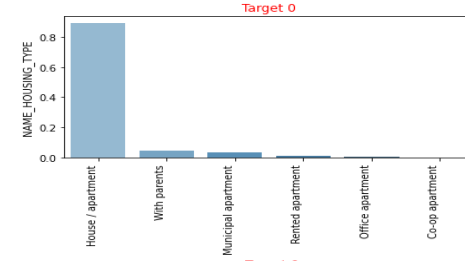
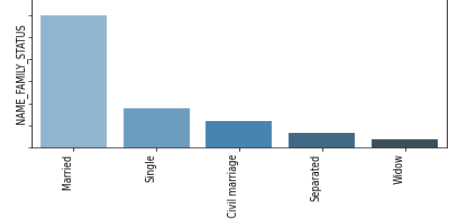
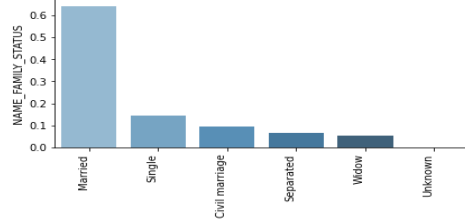
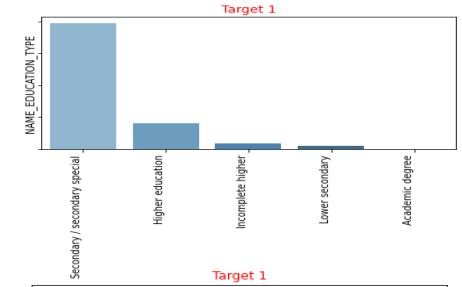
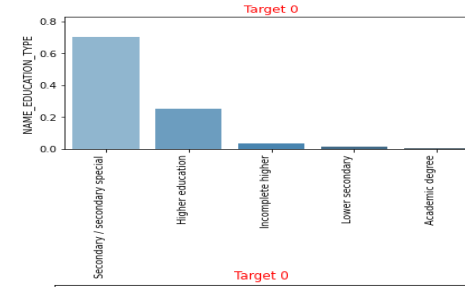
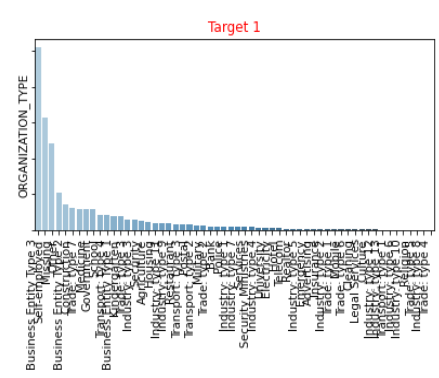
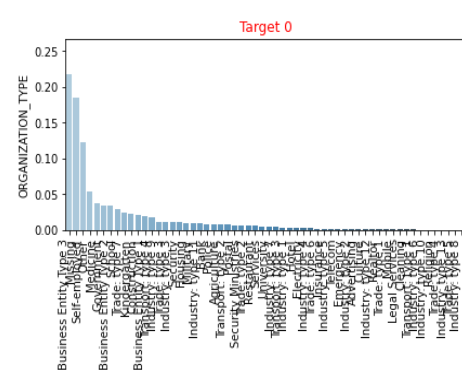
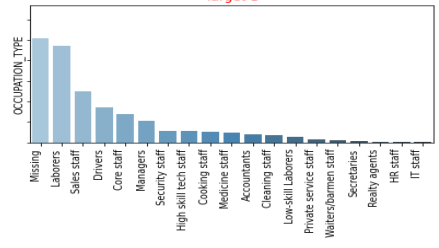
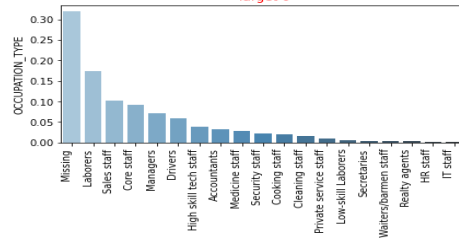
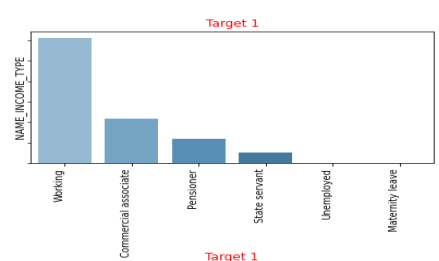
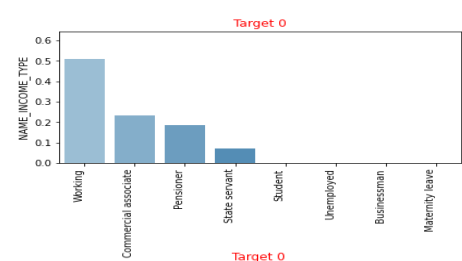
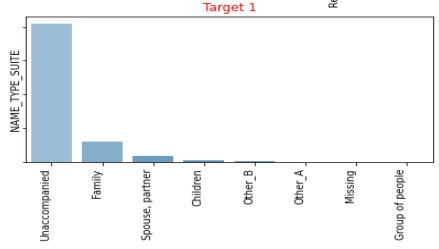
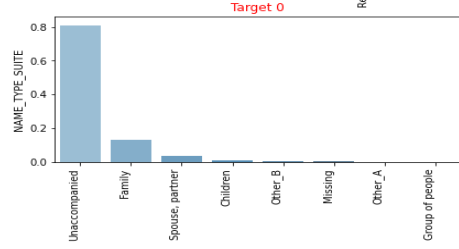
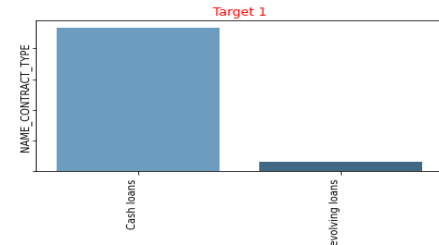
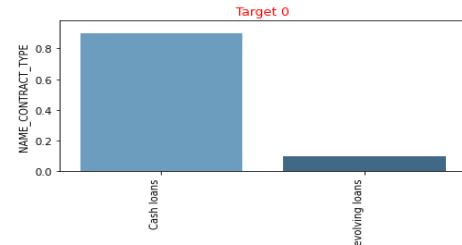
1. Cash loans offered are more than revolving loans, at 90%
2. 65% Females have taken loans in comparison to 34% male. This is very interesting and needs to be studied further
3. 81% applicants came Unaccompanied for loan application
4. 51% are Working
5. While most applicants are working class, 18% are pensioners
6. 71% have secondary education
7. 63% are married
8. 88% have House/Apartments
9. 31% have not mentioned their occupation type
10. 1 in every 11 applicant has payment difficulty.

Boxplot for all Columns which have int64 and float64 as dtype which gives information about mean, median, outliers etc

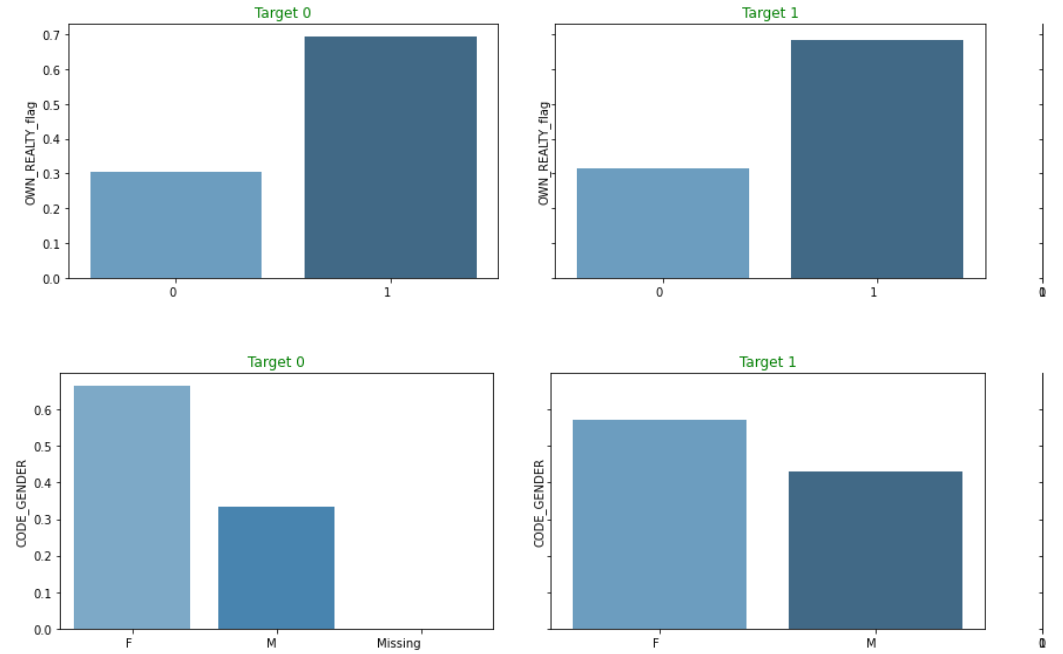
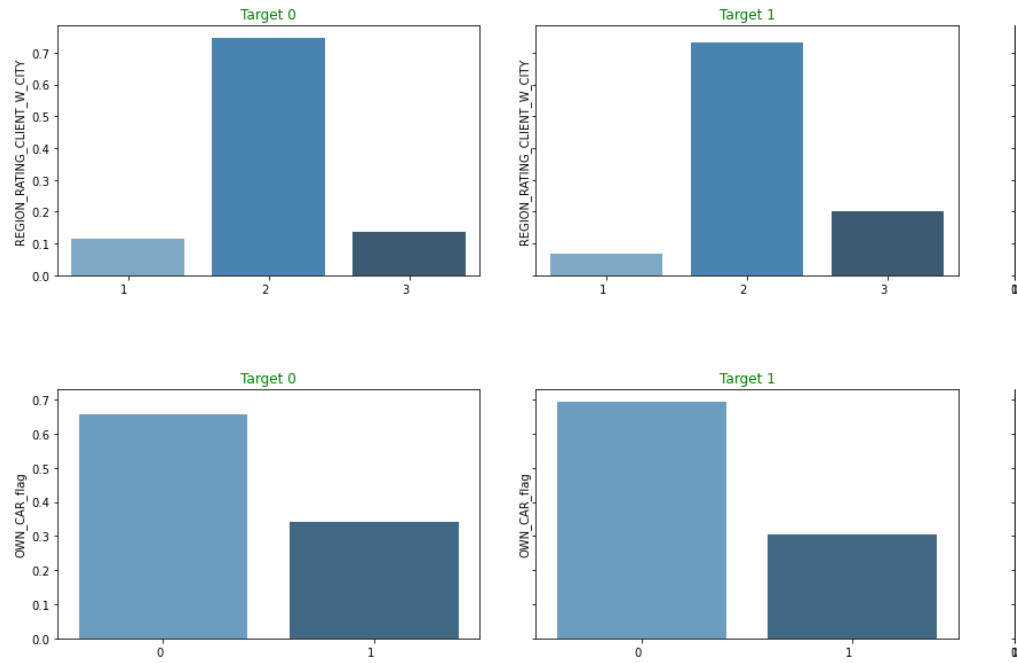


Boxplot for all Columns which have category as dtype and gives information
And here we have binned data for income group and age to visualize it properly

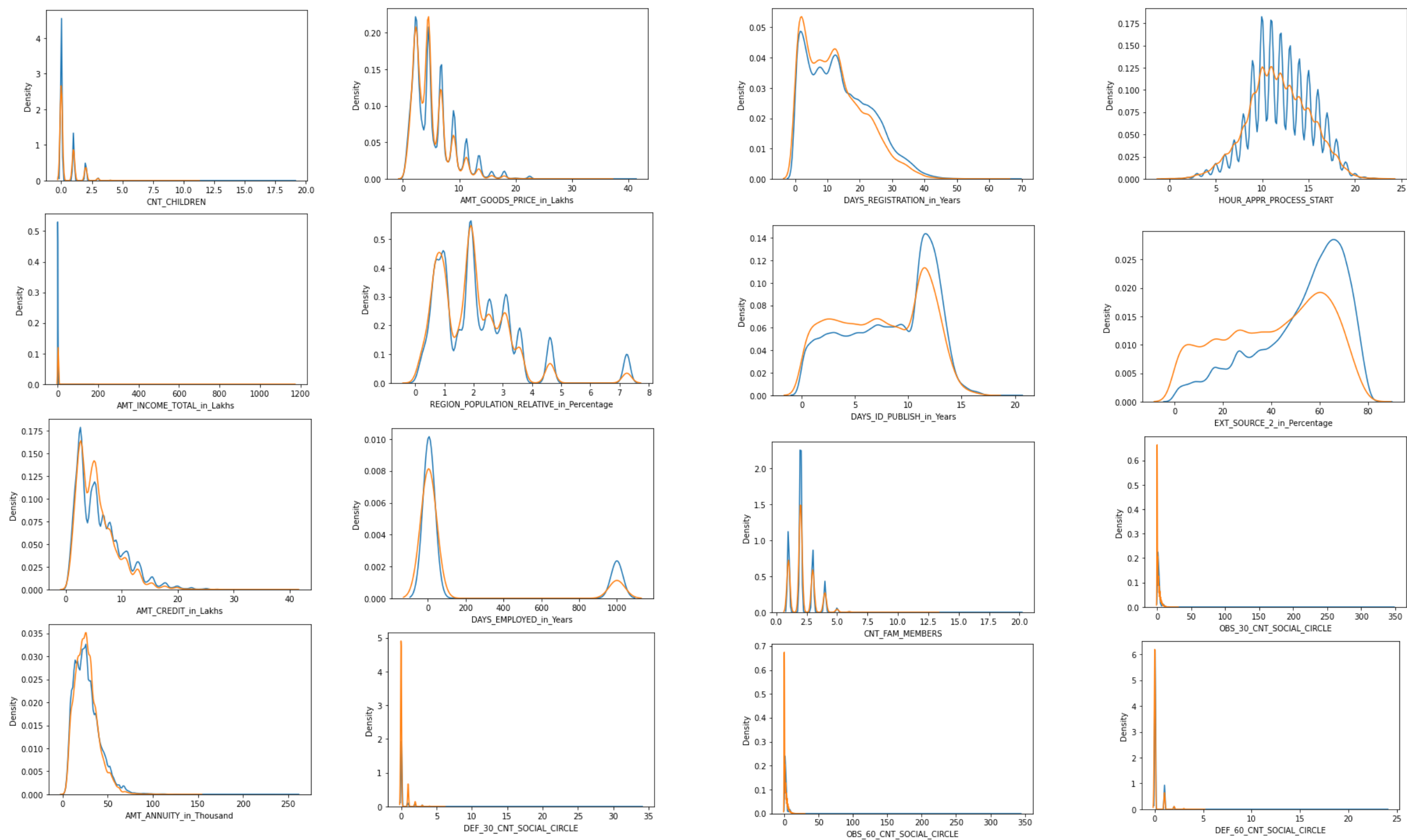


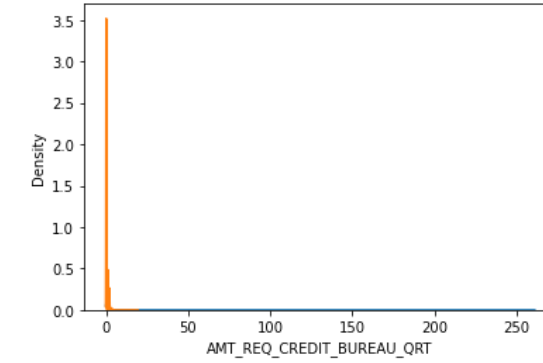
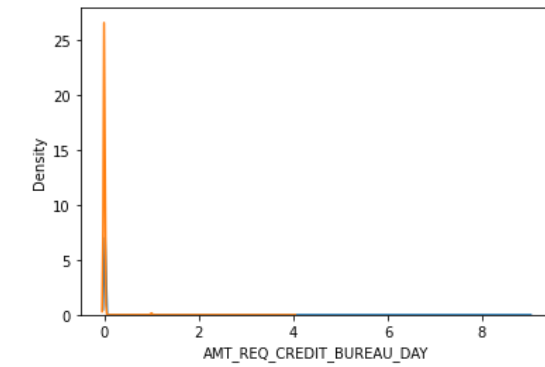
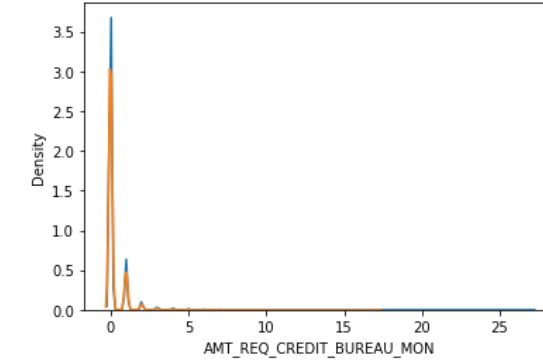
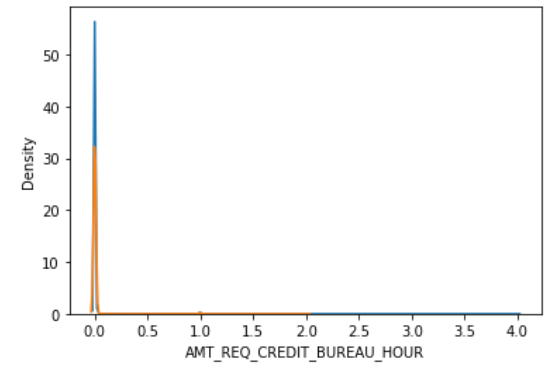
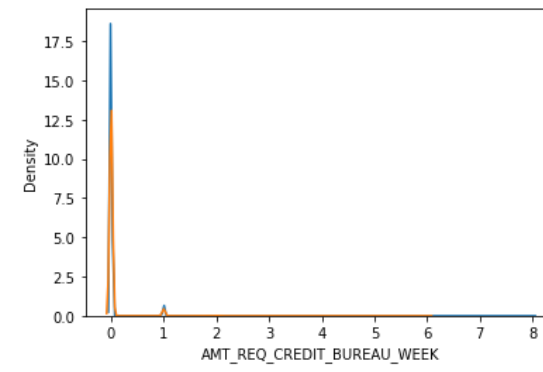
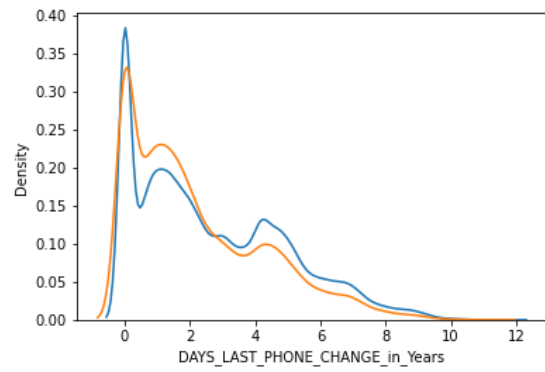


1. NAME_CONTRACT TYPE – Cash Loans For Target 0 is 85% and almost 95% for Target-1
2. NAME_TYPE_SUIT – Nearly 80% in Target 0 and Target 1 are applying for loan Unaccompanied.
3. NAME_INCOME_TYPE - 50% working in case of Target 0 and 60% in case of Target 1 are working income types.
4. NAME_EDUCATION_TYPE - In Target 0, applicants with Secondary Education has applied for loan is nearly 65%-70% and for Target 1 applicants with Secondary Education has applied for loan is nearly 75%-80%
5. NAME_FAMILY_STATUS - Married applicants - almost 60% have defaulted on payments
6. NAME_HOSUING_TYPE - 85-90% in Target 0 and Target 1 applicants are staying in "House/apartment".
7. OCCUPATION_TYPE - Labourers, sales staff, core staff, drivers constitute of 50% of defaulters. Labourers with the nearly 15%-20% is the highest percentage of applicants too. Many of the Applicant have not filled so it is Showing in Missing
8. Region_Rating – Mostly we can see for both the Target 0 and Target 1 rating given is 2



1. REG_CITY_NOT_LIVE_CITY, REG_CITY_NOT_WORK_CITY - Default ratio is higher for 1, ie different from permanent address
2. 'CODE_GENDER' - F have nearly 60% - 70% with more than M in Target 0 and nearly 50% - 60% of F in Target 1
3. OWN_REALTY_Flag - In Target 0 '1' have nearly 65%-75% and in Target 1 '1' have nearly 65%-75%
4. OWN_CAR_Flag - In Target 0 '0' have nearly 60%-65% and in Target 1 '1' have nearly 65%-75%

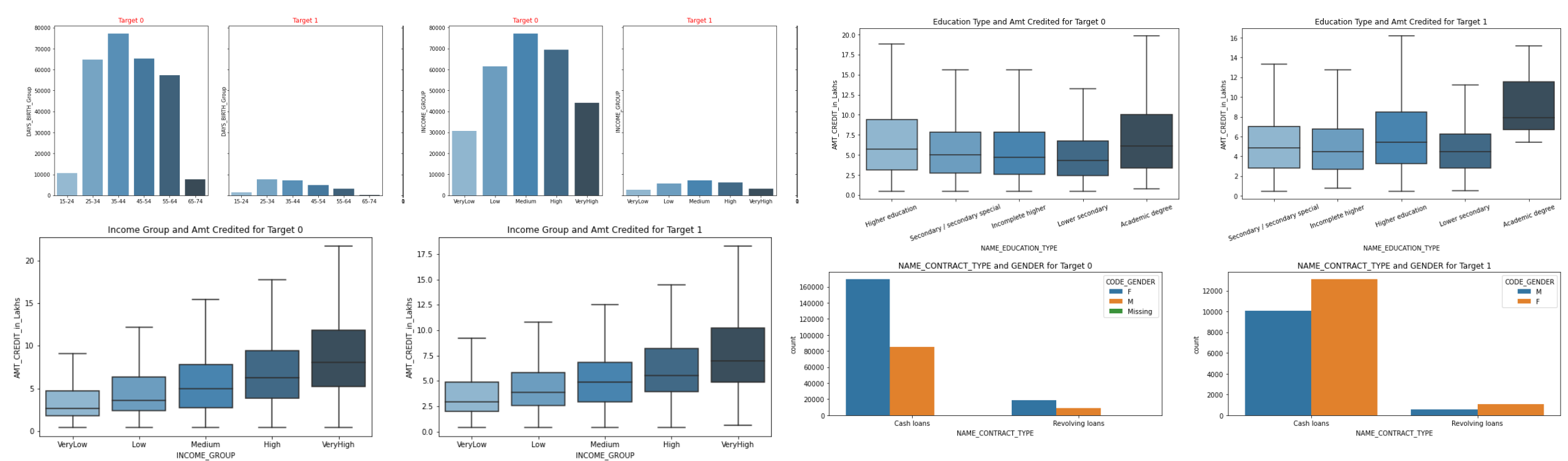




Blue Line is for Target 0 and Orange Line is for Target 1

1. AMT_CREDIT - Appears lower for Target 1, which is a good sign as lesser default loss to the company.
2. YEARS_EMPLOYED has large no of rows of incorrect data and hence the data representation is incorrect
3. EXT_SOURCE_2 is clearing showing that TARGET 0 has higher density of higher scores
4. OBS_30_CNT_SOCIAL_CIRCLE, DEF_30_CNT_SOCIAL_CIRCLE, OBS_60_CNT_SOCIAL_CIRCLE, DEF_60_CNT_SOCIAL_CIRCLE - Very clearly visible that for Target 1, 30 DPD and 60 DPD observed in social surrounding is higher.
5. DAYS_LAST_PHONE_CHANGE- More people from the Target 1 have changed their phone earlier than Target 1. Indicating intention issues in repaying loan
8. AMT_REQ_CREDIT_BUREAU_YEAR, month, week, hour - has lesser Target 1 at 0 hits. This could indicate they are looking getting loans from various financial companies

There are Many Observations we can see but these are kind of more important according to me



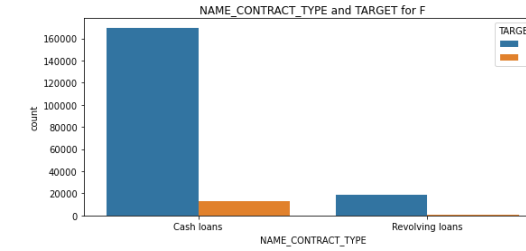
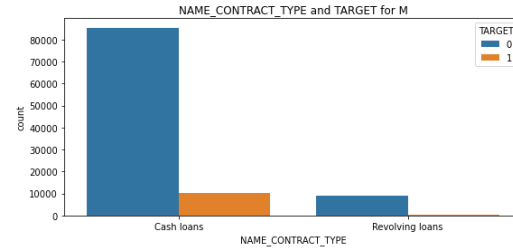
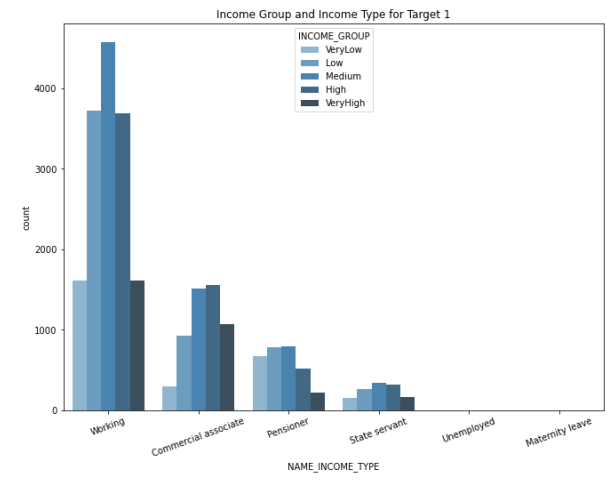
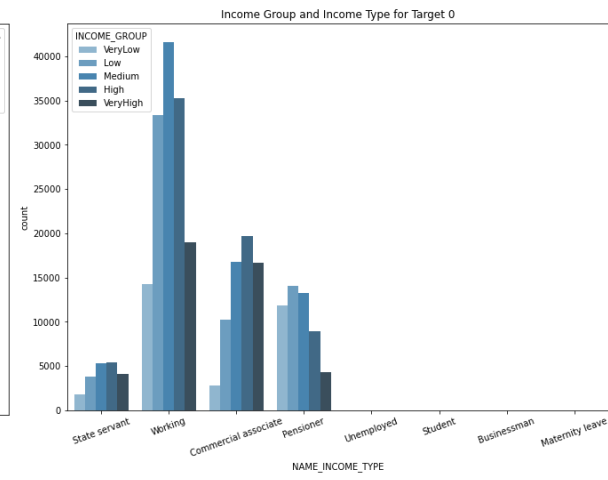
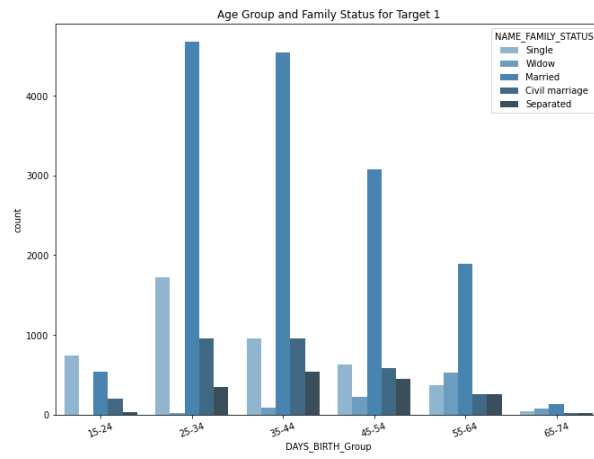
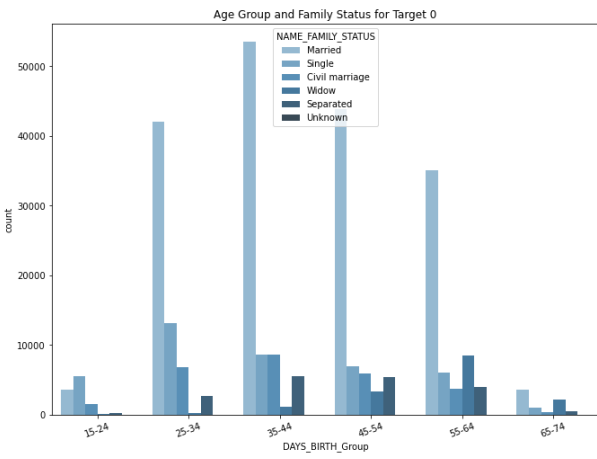
AGE_GROUP - 35-45 are more in Target 0. In Target 1- 25-25 have higher share. Age does seem like influencing default.

INCOME_GROUP - Medium income group have more count in Target 0 and Target 1

NAME_CONTRACT_TYPE - We can observe F have took Cash Loans More than M for Target 0 and same F have took Cash Loans more than M for Target 1

Income_Group and AMT Credited – We get information about Mean, Medium, Percentiles, Etc for every group for Target 0 and Target 1 We can infer that though the maximum no of loans is given to Medium income group. Default value per loan is highest in High income group as the AMT_CREDIT is higher too. The loan book of the financial institution can get affected due to higher amount not being paid back.

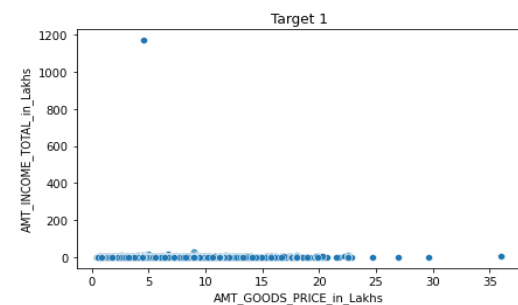
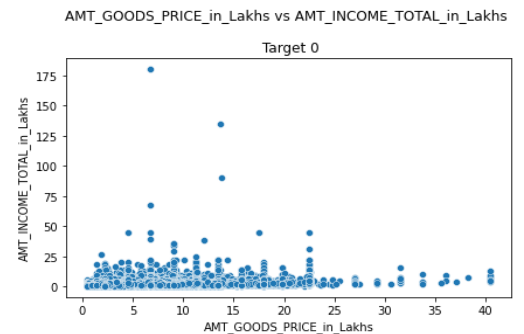
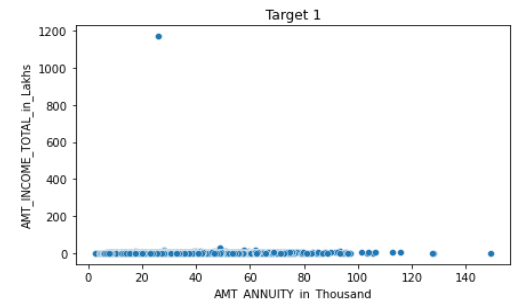
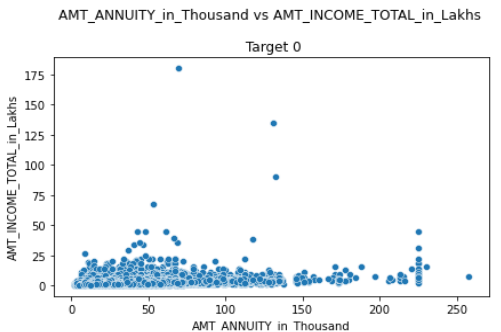
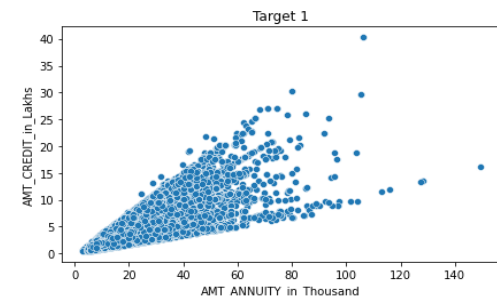
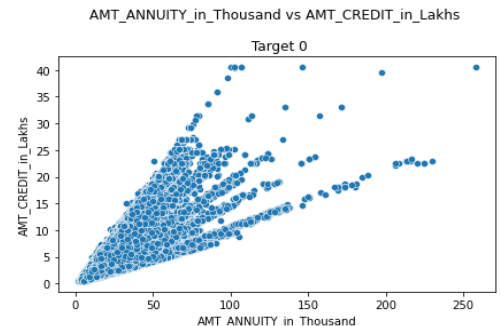
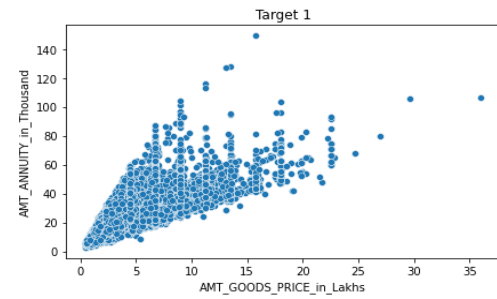
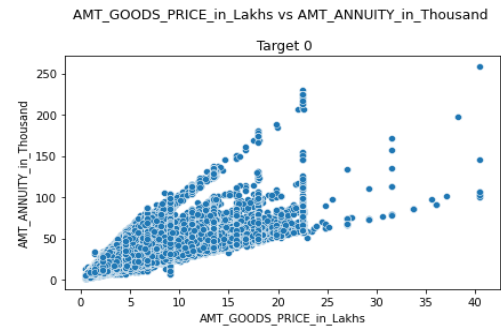
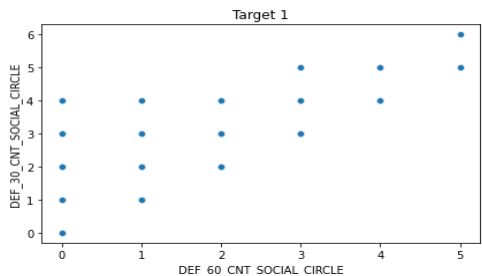
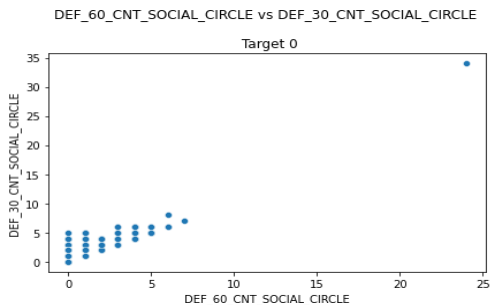
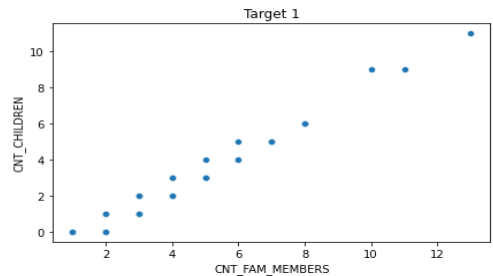
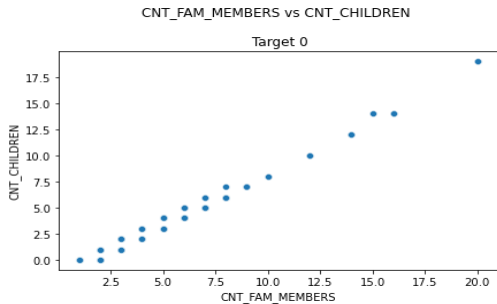
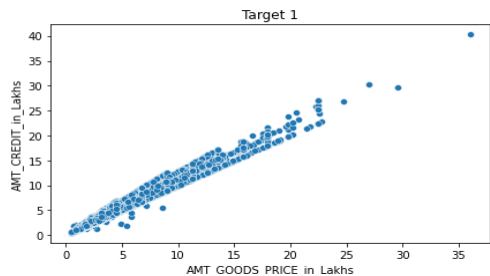
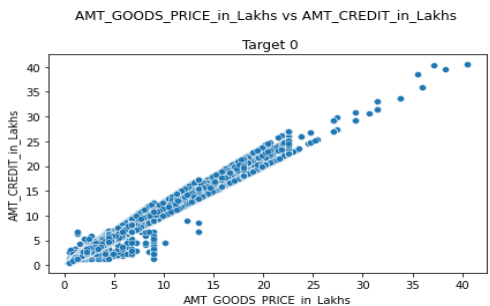
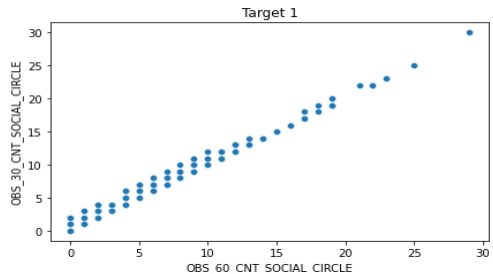
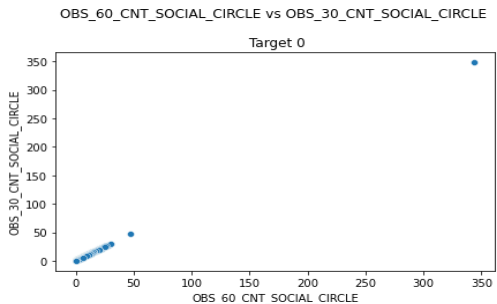
The company must devise a different set of rules and policies while approving higher income group loans.

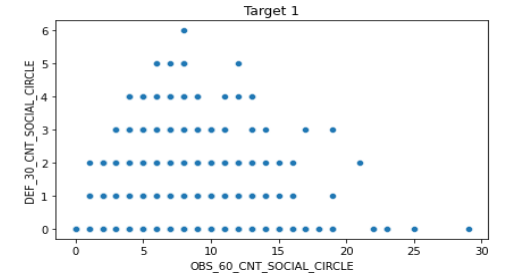
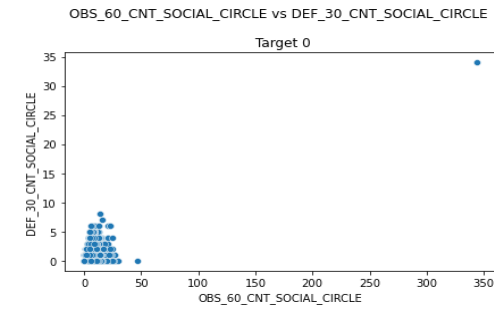
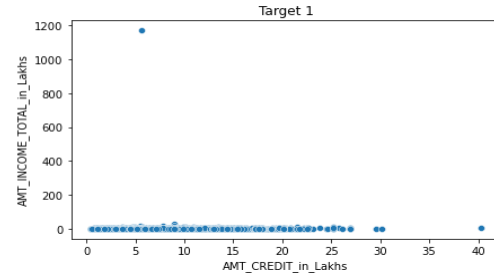
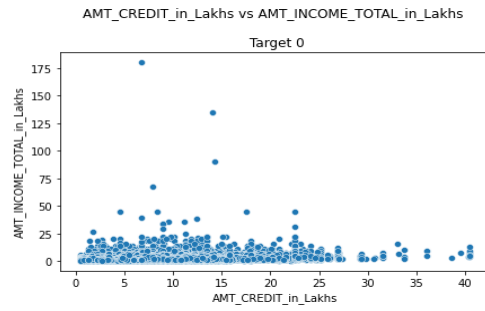


Married applicant in the age group 25-35 and 35-45 is the largest group of applicant with payment difficulties.

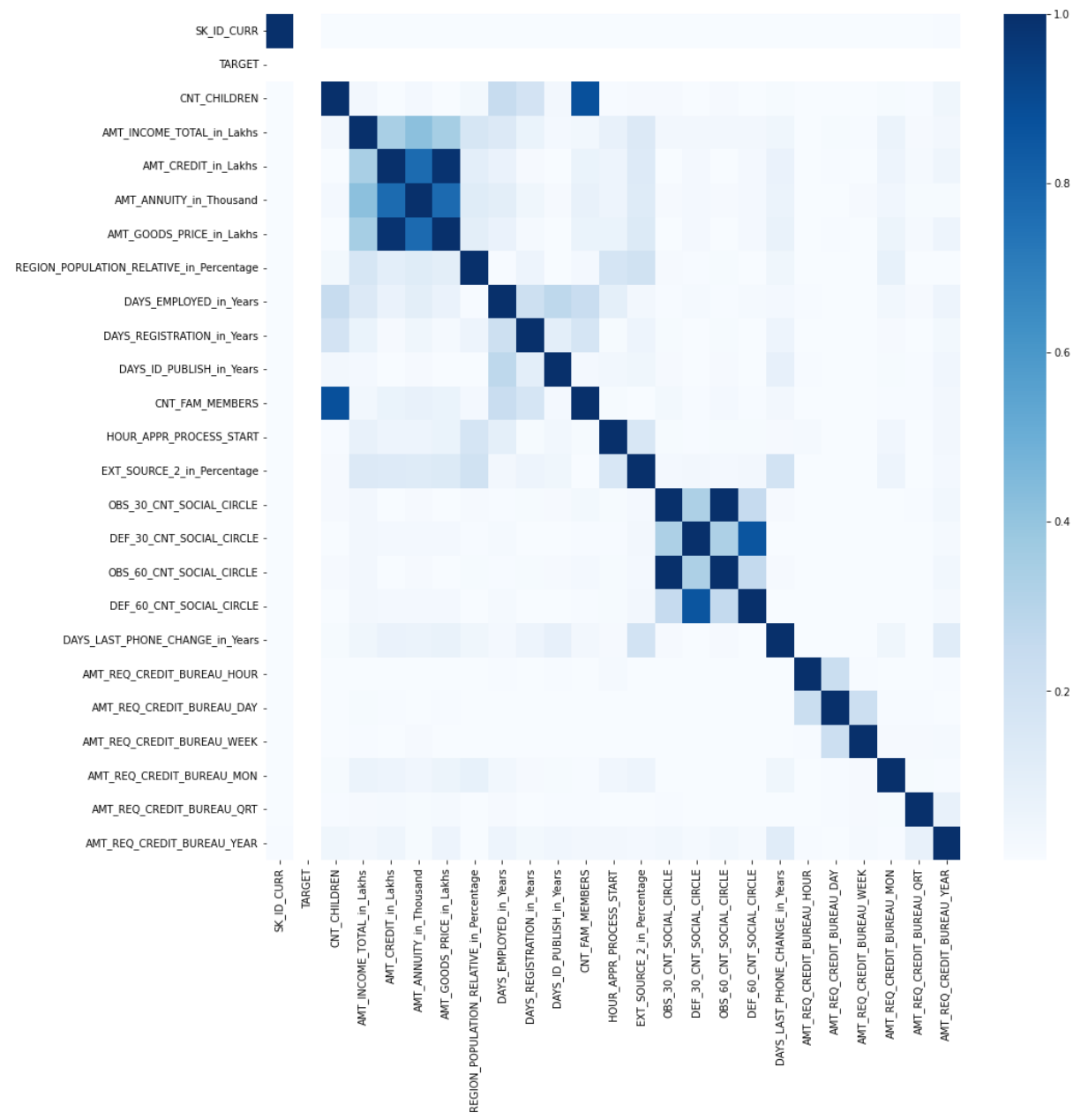
Medium income group with income type has almost 1 in 12 defaults. Higher than the avg 1 in 11 defaults.

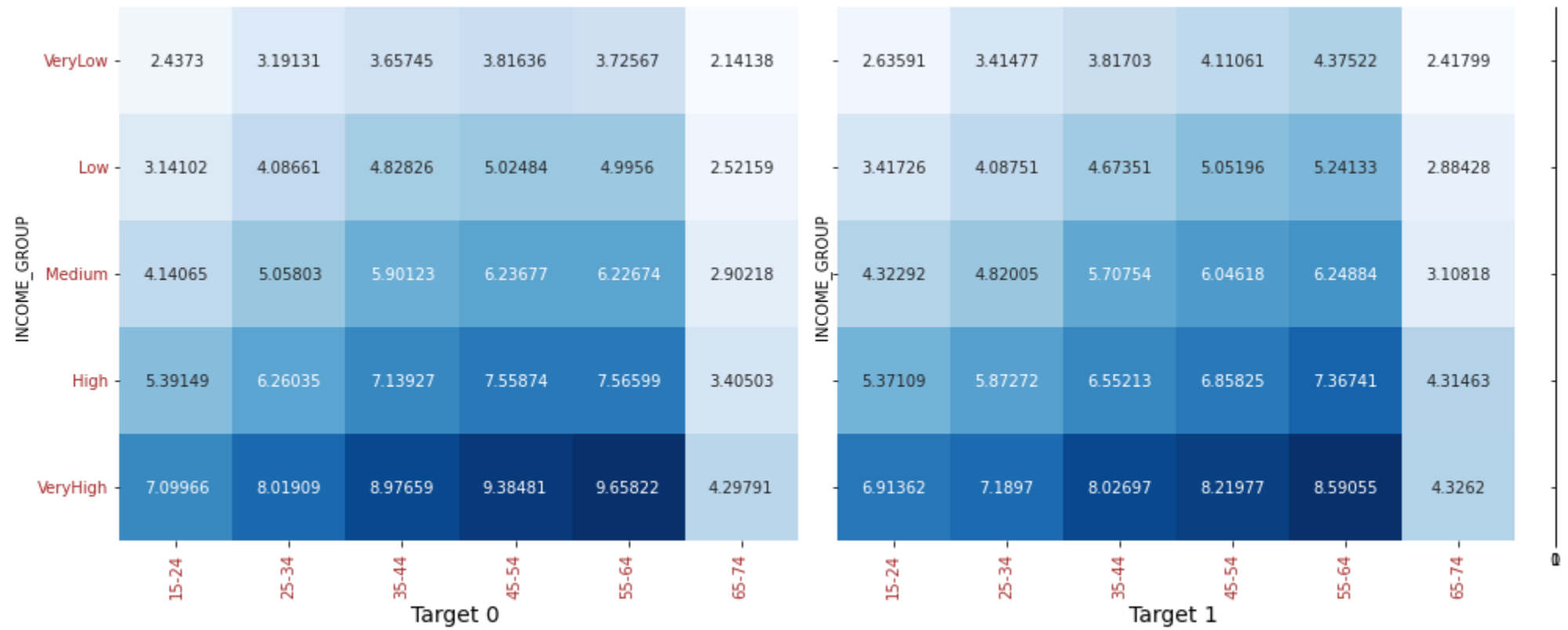
Now Name_Contract_Type we have divided for Male and Female



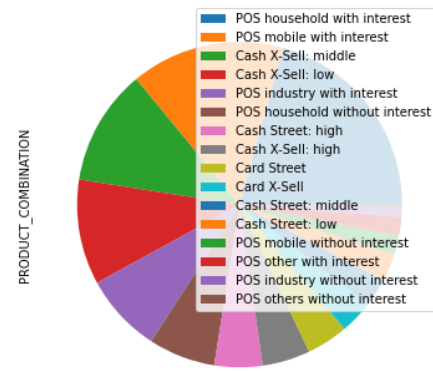
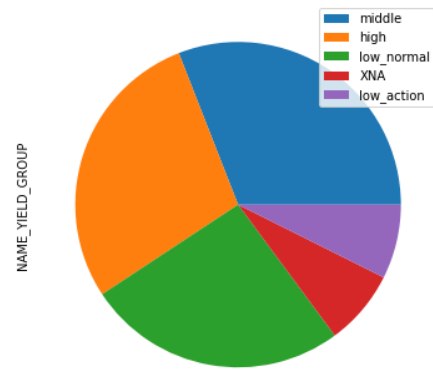
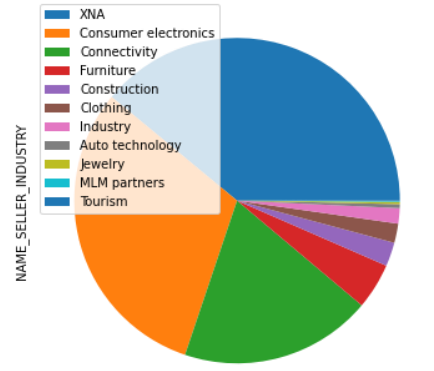
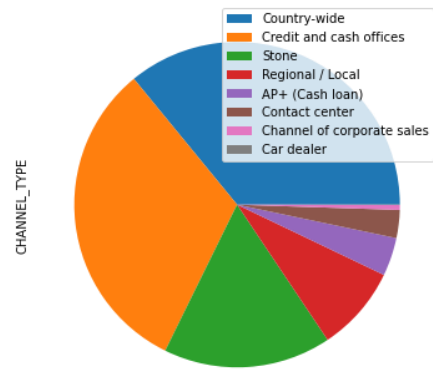
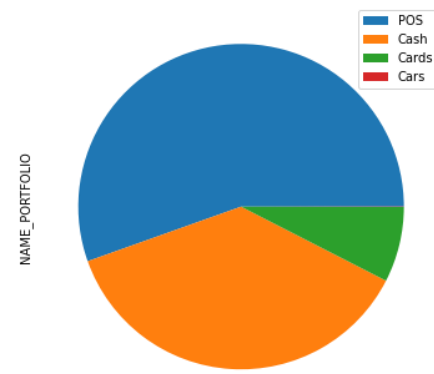
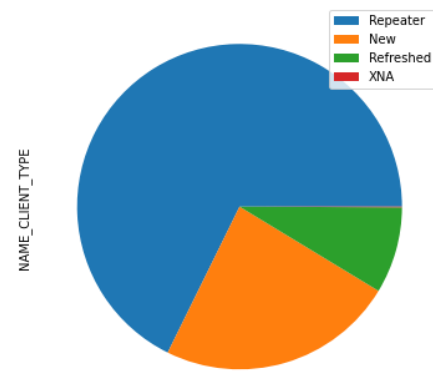
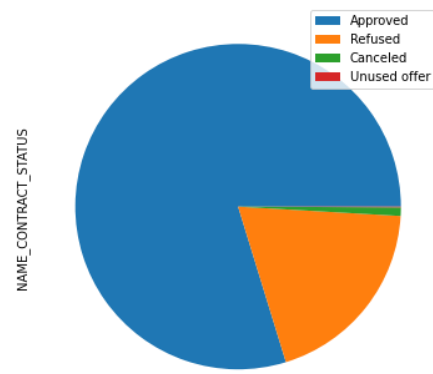
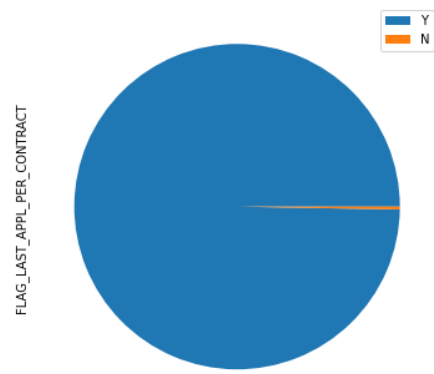
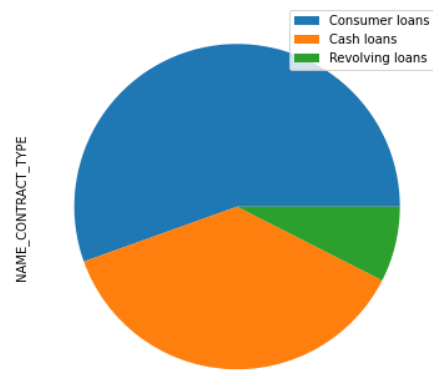


1. OBS_30_CNT_SOCIAL_CIRCLE', OBS_60_CNT_SOCIAL_CIRCLE' - denote the client's social surroundings with observable 30/60 DPD. These are definitely correlated. We can also see that its higher and steeper for Target 1, signifying that in approval process this parameter must be strongly looked into.
2. DEF_30_CNT_SOCIAL_CIRCLE - Trend is going up. But Target 1 has lesser data and hence graph is not dense.
3. Years employed has an outlier value of 999 and this is skewing the graph
4. AMT_CREDIT and AMT_GOOD PRICE don't seem to be increasing proportionately with AMT_INCOME for TARGET 1, thus possibly leading to default

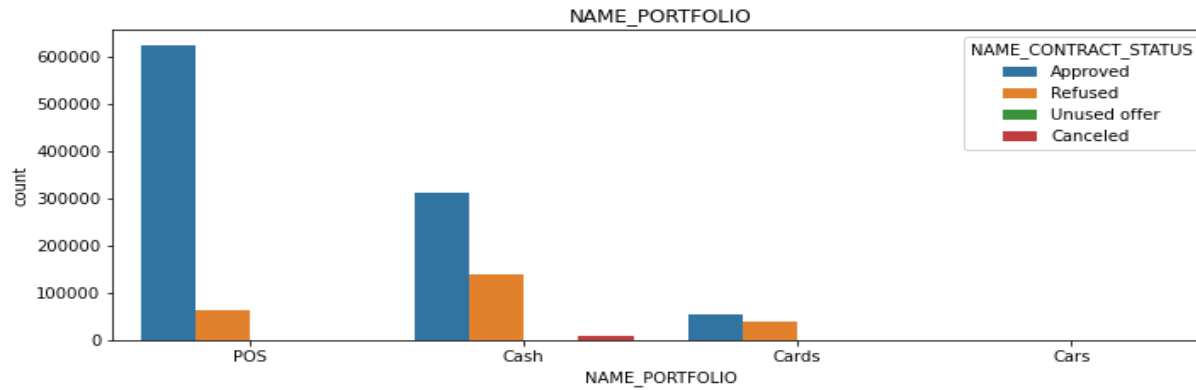
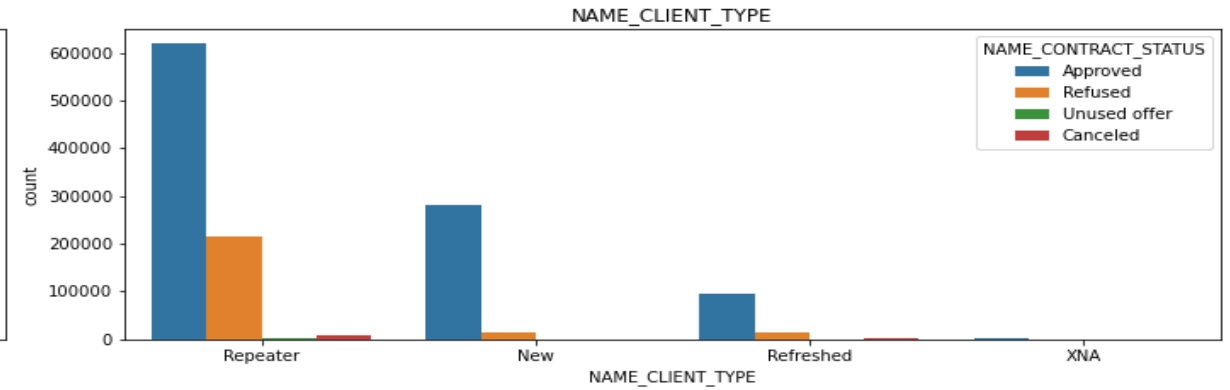
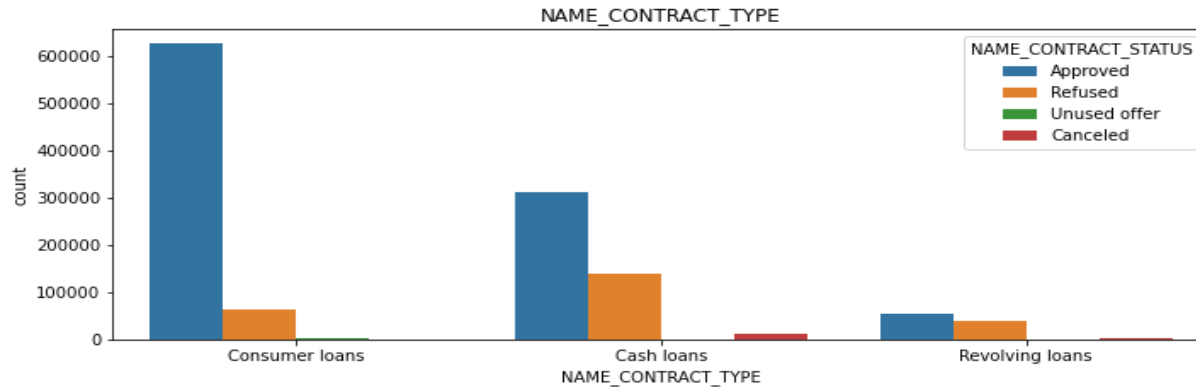




Age Group 55-65 in Very High income group has high amount credit. As explained above, this could result as loss in loan book

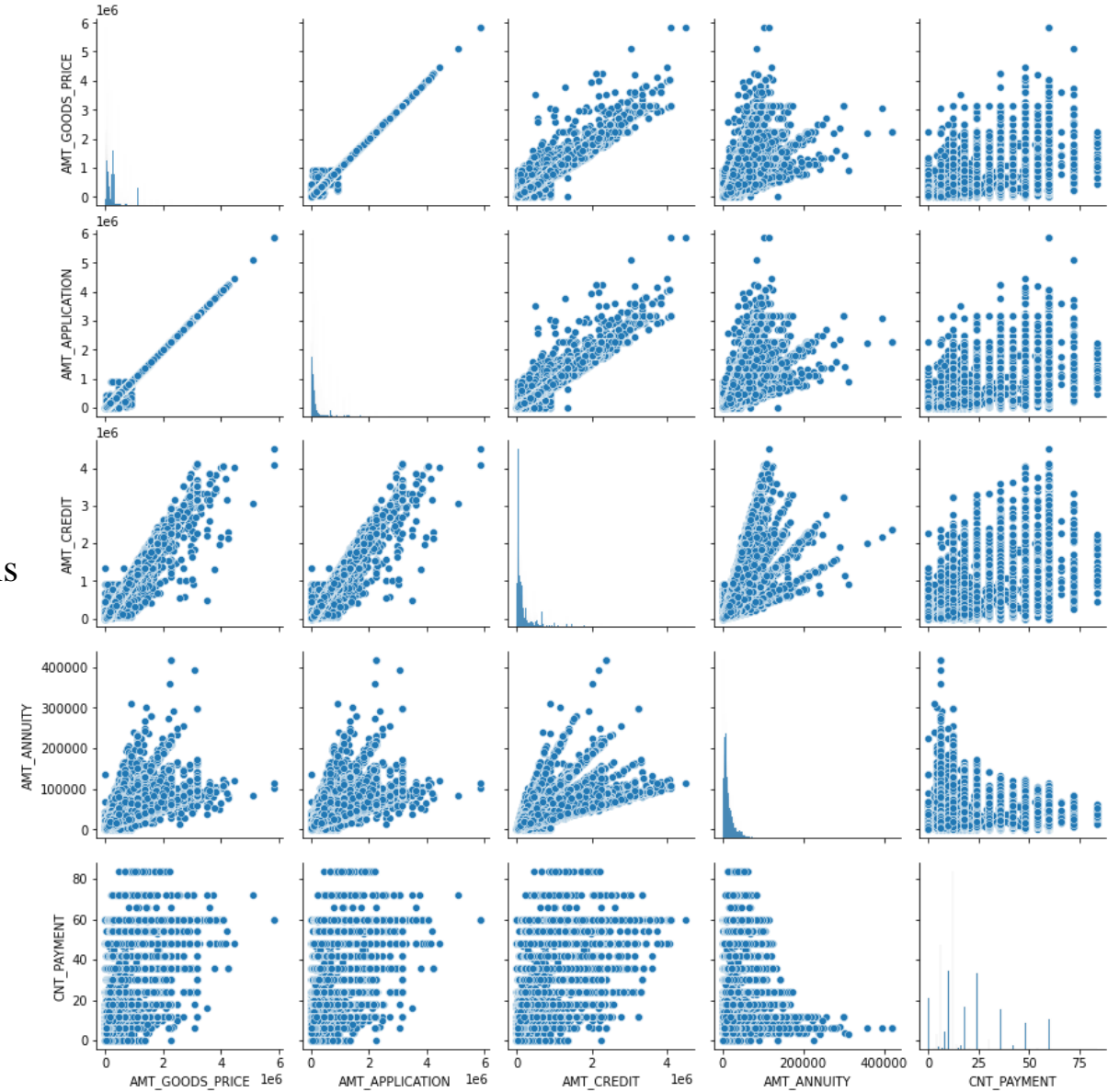


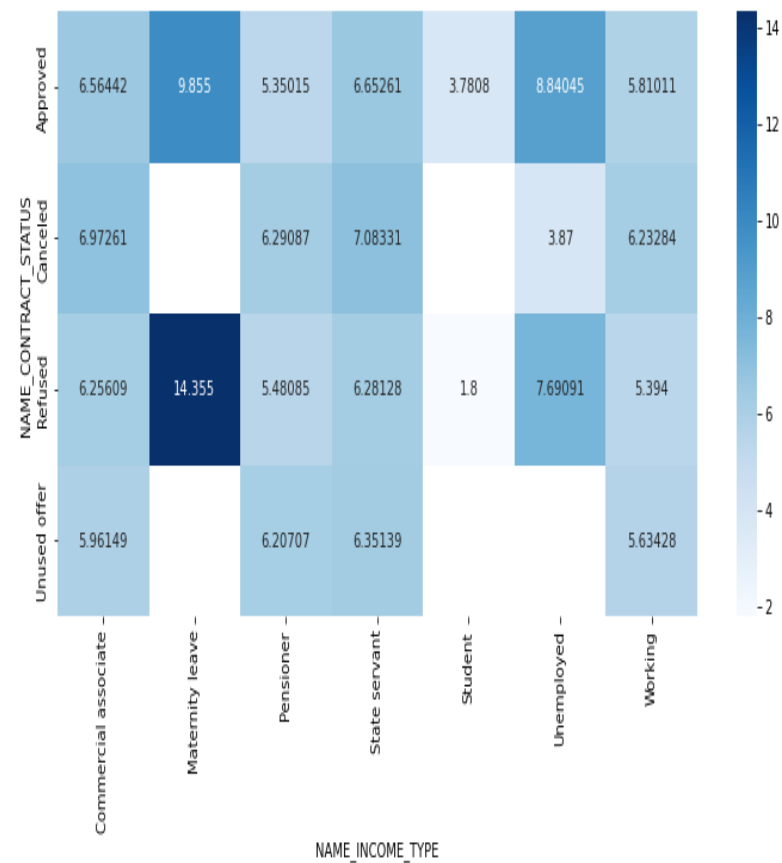
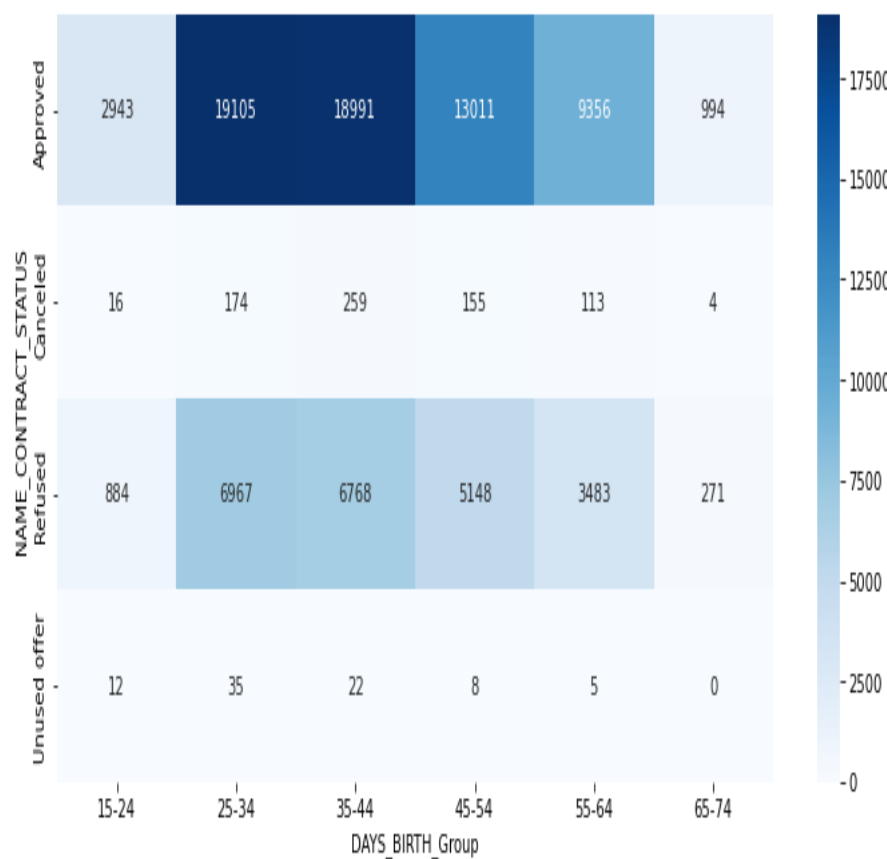
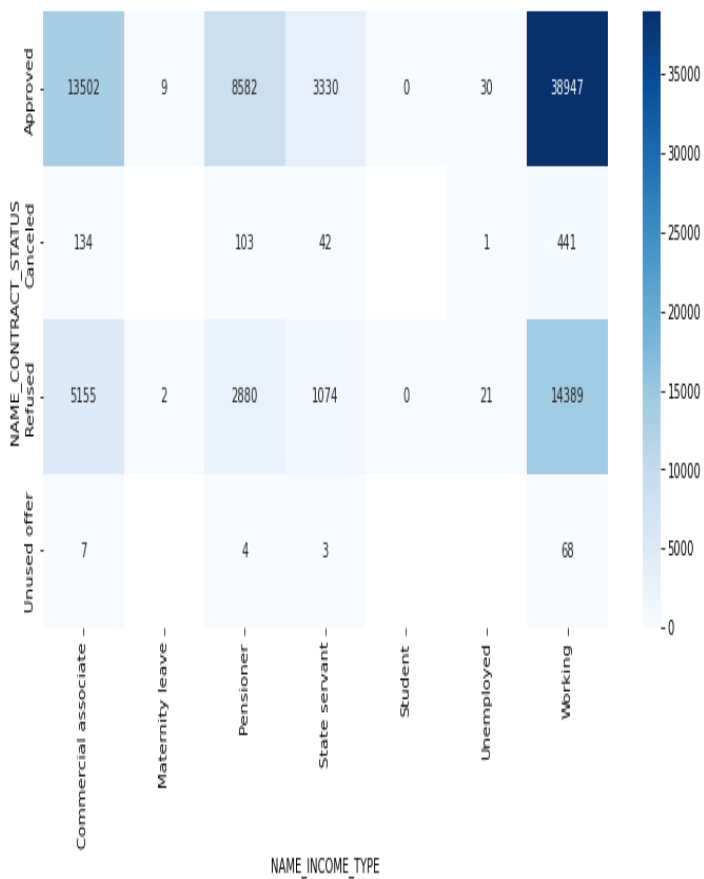
1. This data frame has a different type of loan called Consumer Loan, which was not there in Application data frame. 55% of loans are consumer loans. 37% cash loans and rest revolving
2. Approved loans are 79% and refused, cancelled, unused - rest.
3. 67% are repeaters. NAME_CLIENT_TYPE also has some null values showing as XNA
4. 55% of the applicants have taken loan for POS purchase.
5. Name seller industry has 38% XNA values, Consumer electronics is next highest category at 30%



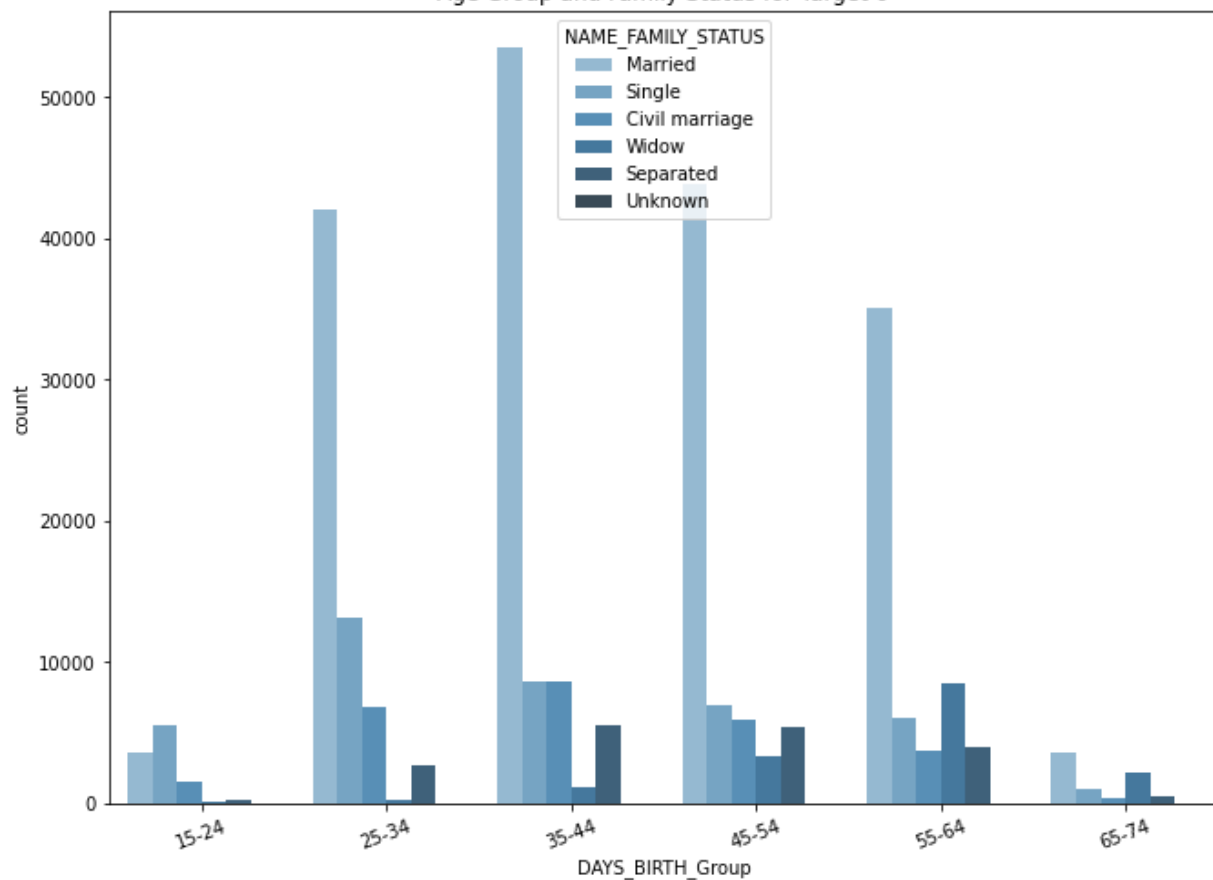
1. In approved category, consumer loan has largest no of applicants of nearly 6 lakhs.
2. There seem to be no cancelled loans in cash loan category than consumer loan.
3. More cash loans have been refused than consumer loans.
4. The bank has more repeaters in all approved, refused, unused, cancelled categories
5. POS transactions seem to be consumer loans and similar to point 2 - more cash loans have been refused than POS.

1. AMT_GOODS_PRICE, AMT_ANNUITY, AMT_APPLICATION - as expected have high correlation. Higher the value of good purchased more there will be need of loan and surely all these will correlate.
2. Similarly, AMT_Credit to AMT_GOOD_PRICE also the correlation is high.

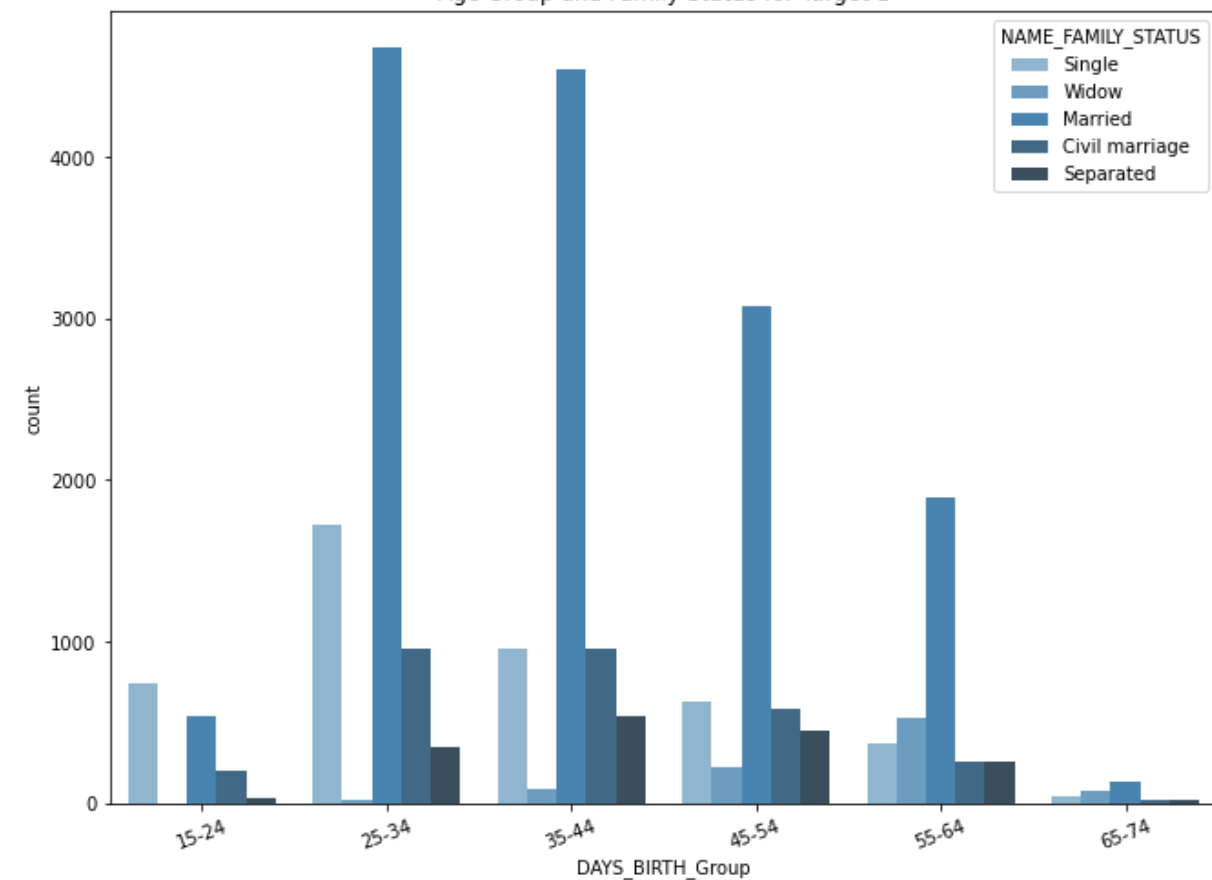




Age Group and Family Status for Target 0



Age Group and Family Status for Target 1



Some Factors Which I feel Bank need to consider to increase their Profit

- Medium income
- 25-35 years old, followed by 35-45 years age group
- Male
- Unemployed
- Labourers, Salesman, Drivers -Business type 3
- Own House - No