Loan Defaulter

CREDIT EDA CASE STUDY

Presented By : Juhi Chugh
Madhumitha R

Problem Statement

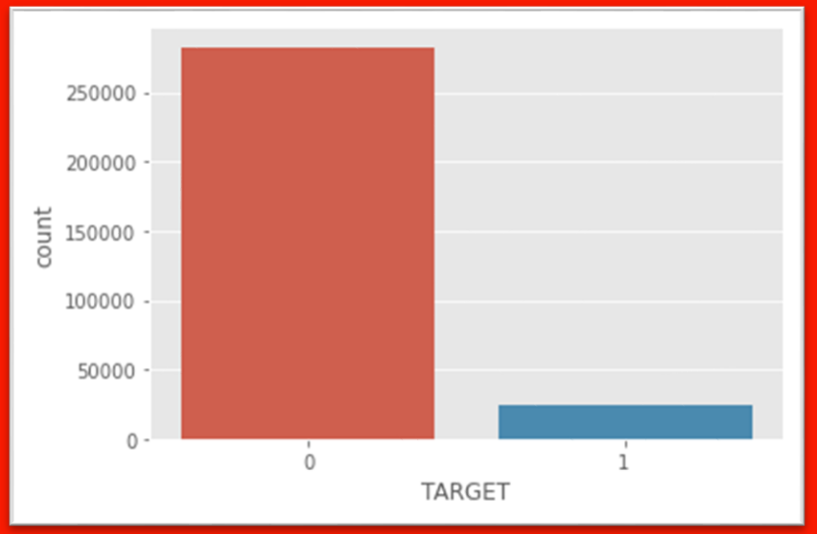
- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose we work for a consumer finance company which specializes in lending various types of loans to urban customers. We have used EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

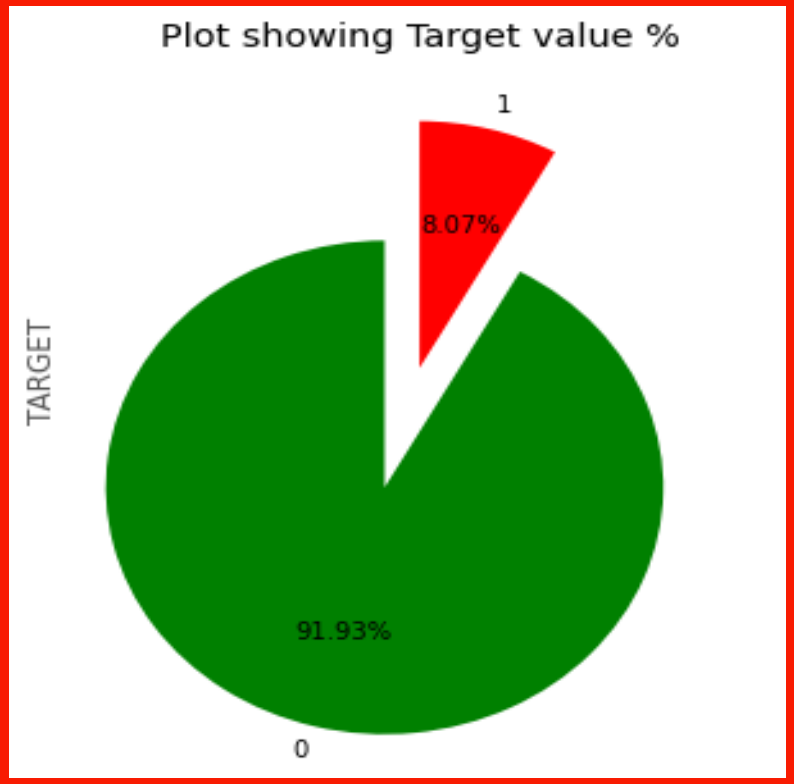
Objectives

- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants' using EDA is the aim of this case study.

DATA IMBALANCE

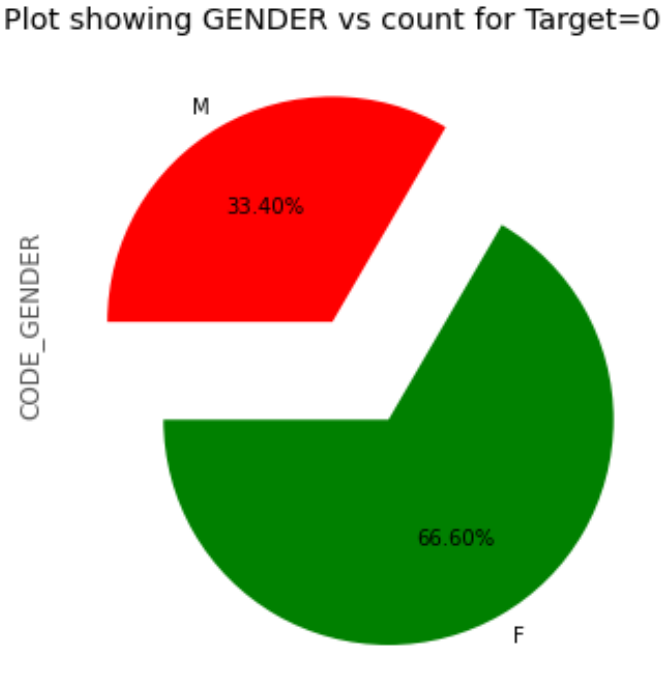
The Imbalance ratio is 11.39

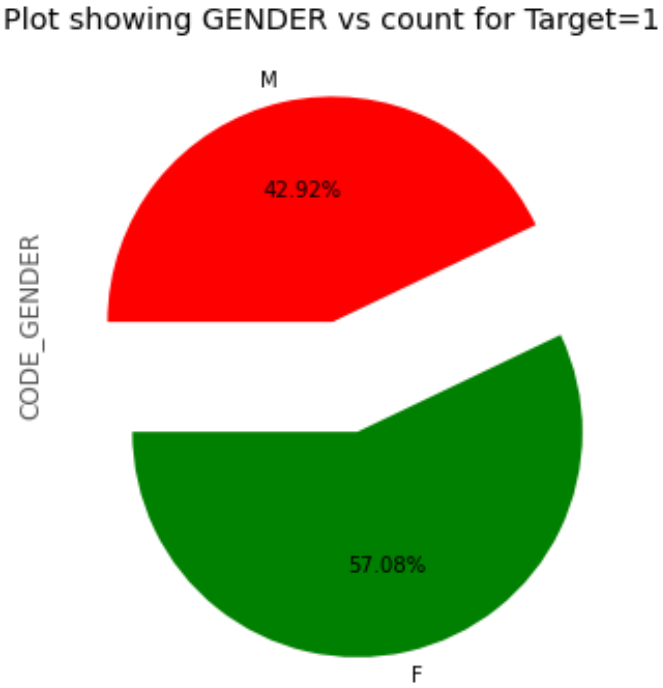




GENDER

The percentage of the Female clients is more compared to the Male.And Males have more chance of defaulting

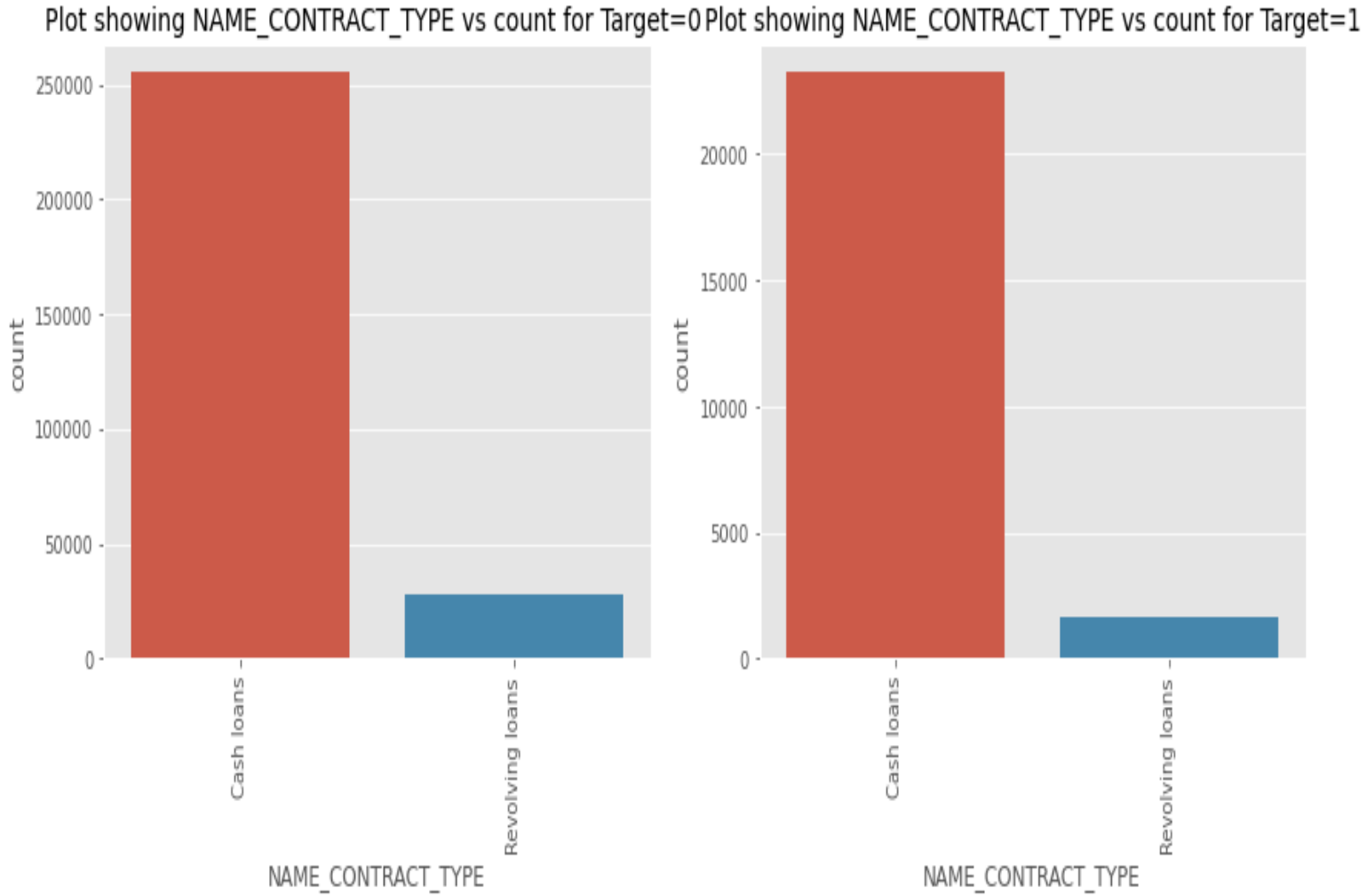


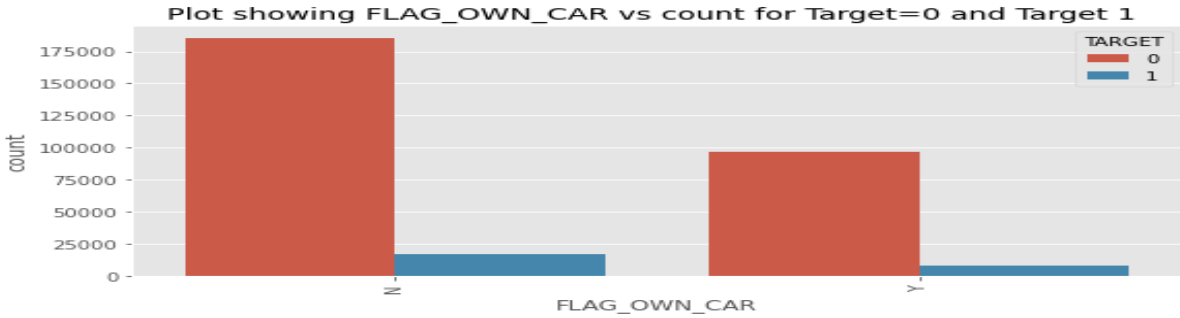


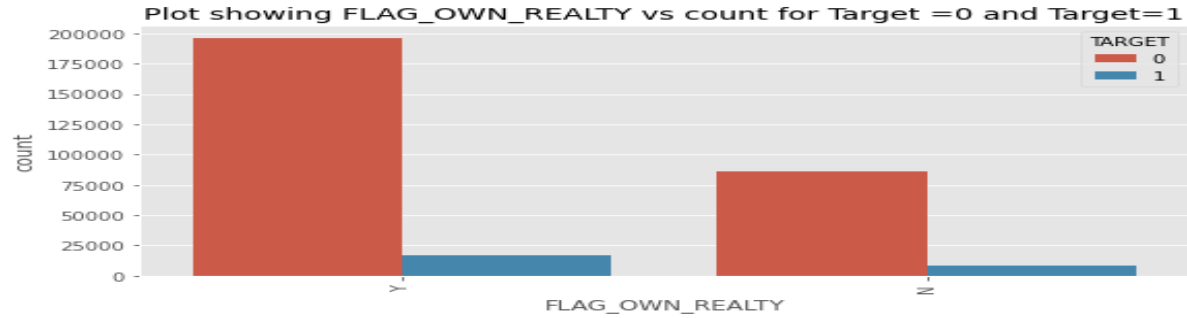
UNIVARIATE ANALYSIS – CATEGORICAL VARIABLES

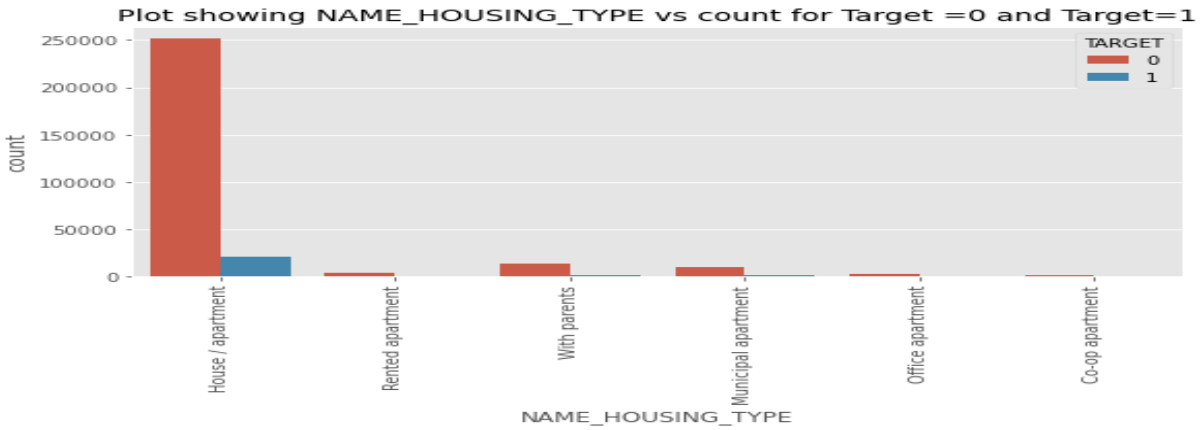
Contract Type

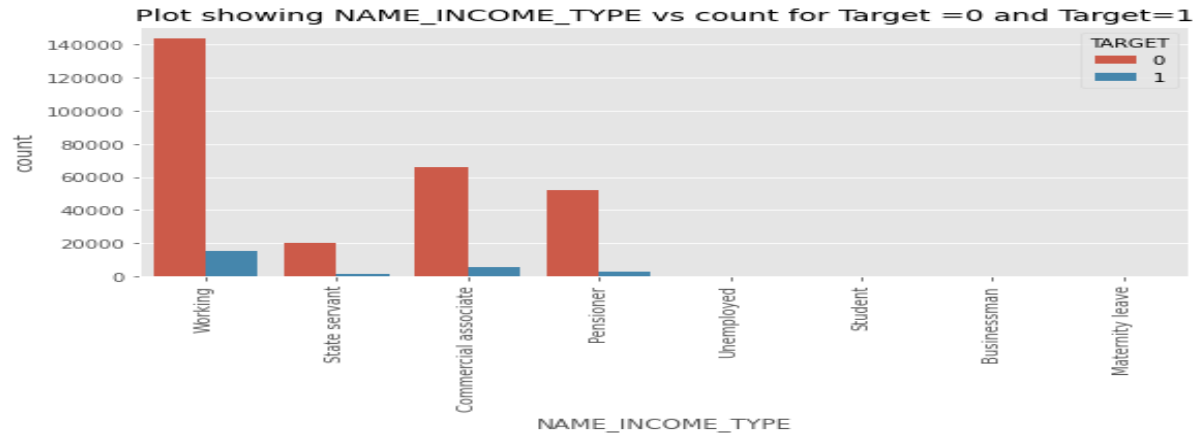
The percentage of client applying for Cash Loans is high.

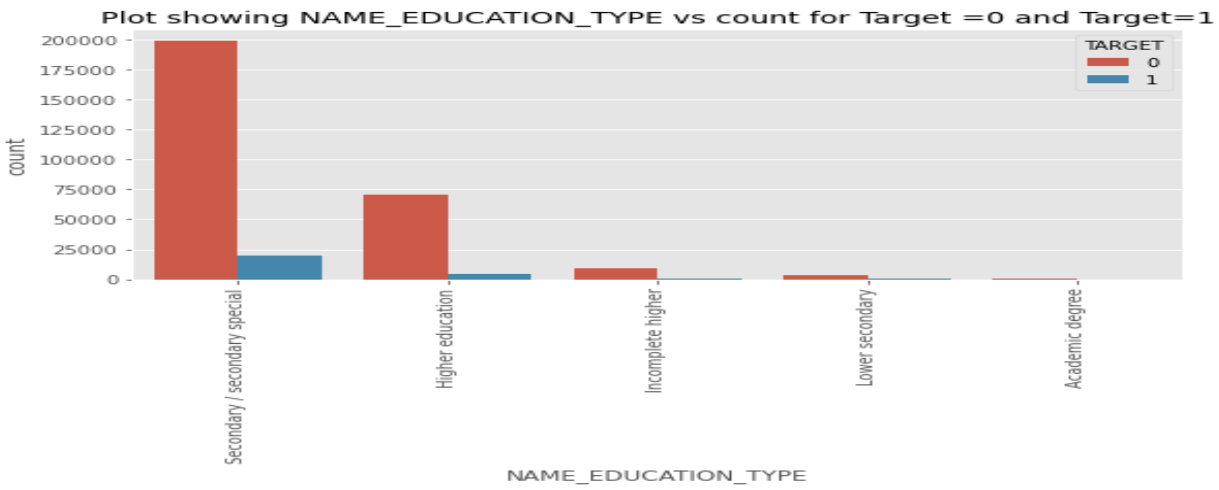


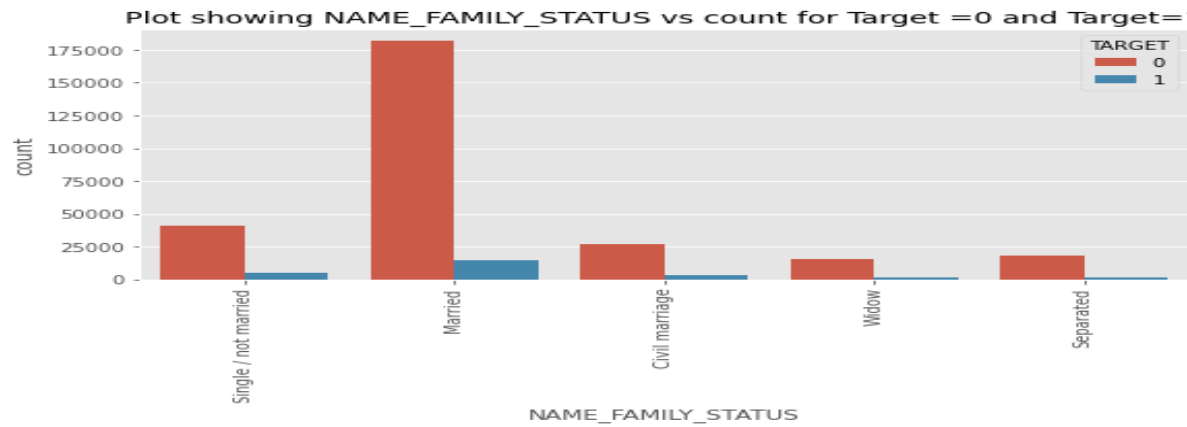

Plot showing FLAG_OWN_CAR vs count for Target=0 and Target 1
TARGET
0
1
count
FLAG_OWN_CAR
N
Y


Plot showing FLAG_OWN_REALTY vs count for Target =0 and Target=1
TARGET
0
1
count
FLAG_OWN_REALTY
Y
N


Plot showing NAME_HOUSING_TYPE vs count for Target =0 and Target=1
TARGET
0
1
count
NAME_HOUSING_TYPE
House / apartment
Rented apartment
With parents
Municipal apartment
Office apartment
Co-op apartment


Plot showing NAME_INCOME_TYPE vs count for Target =0 and Target=1
TARGET
0
1
count
NAME_INCOME_TYPE
Working
State servant
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave

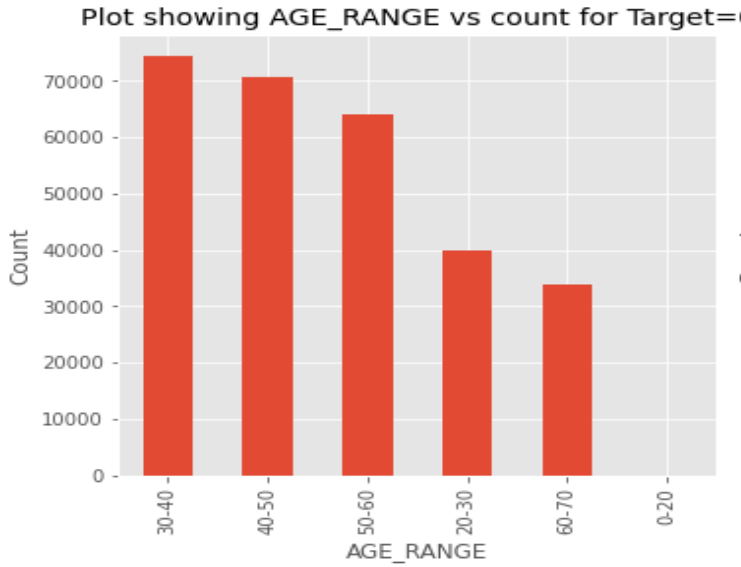

Plot showing NAME_EDUCATION_TYPE vs count for Target =0 and Target=1
TARGET
0
1
count
NAME_EDUCATION_TYPE
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree

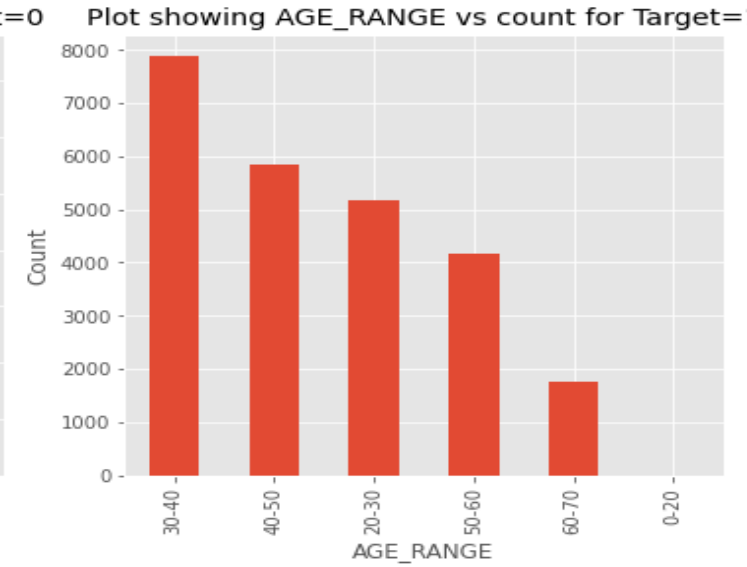

Plot showing NAME_FAMILY_STATUS vs count for Target =0 and Target=1
TARGET
0
1
count
NAME_FAMILY_STATUS
Single / not married
Married
Civil marriage
Widow
Separated

Housing Type
Income Type
Own Car
Realty
Education Type
Family Status

- The count of Clients who do not own a car is high. Hence we should target them but since the chance of defaulting is more should be more careful with the loan amount and other factors.
- The clients having own realty have more chance of defaulting.
- Clients living in Rented apartments ,Co-op apartments or office apartments have less chance of defaulting compared to others.
- Client who are working can be targetted.
- Clients who have secondary/higher education have a greater chance to take loans.
- Clients who are married have a greater chance to take loans.

AGE

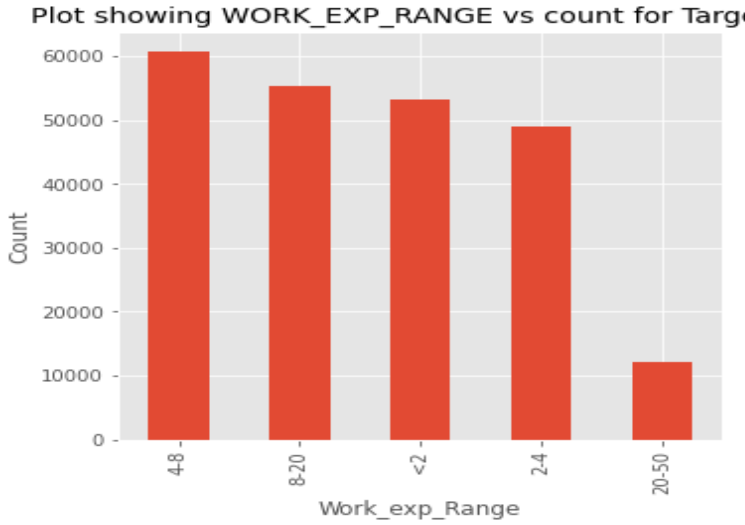


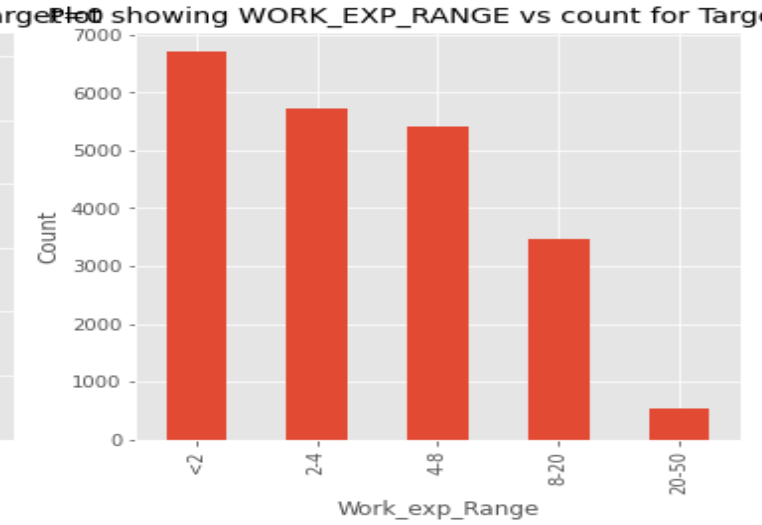


Target the clients in age group 30-50.Eventhough the age of 20-30 have applied for loan their percentage of defaulting is high when compared to other age group.

Clients who are having experience of 4-20 years have less chance of defaulting.While the clients with work experience less than 2 years or 2-4 years have higher chance of defaulting .We can target clients with greater work experience

WORK EXPERIENCE



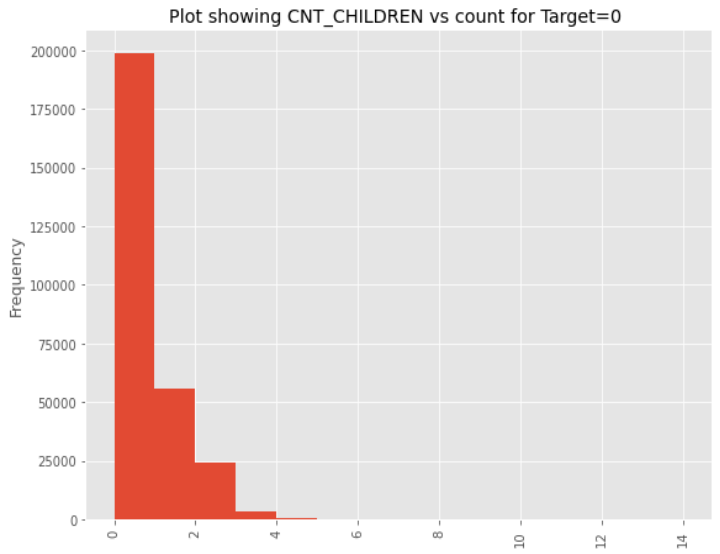


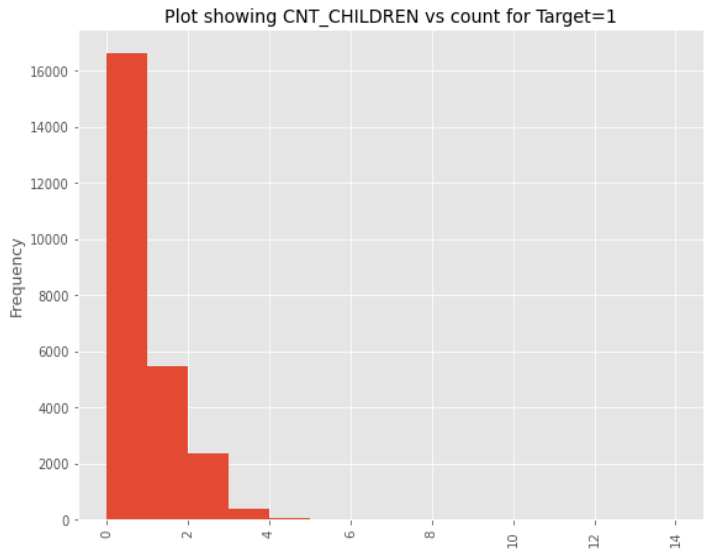
UNIVARIATE ANALYSIS –NUMERICAL VARIABLES

COUNT OF CHILDREN

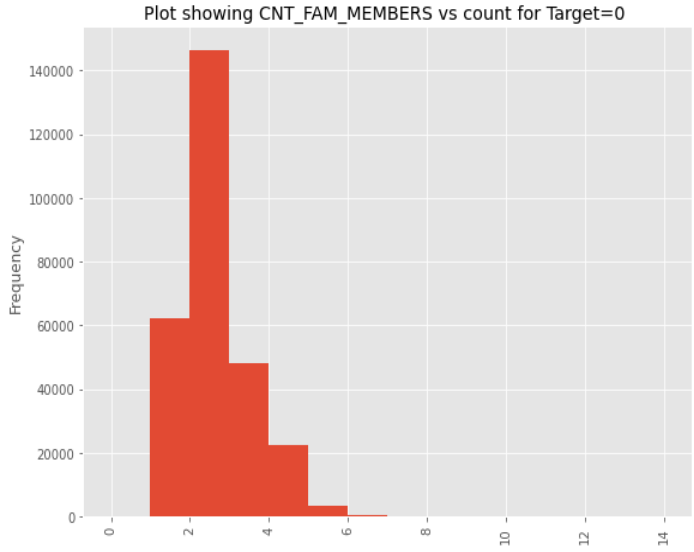
Clients who have no children have higher count of applying for loan.Also when the count of children increases, they have more chance of defaulting.

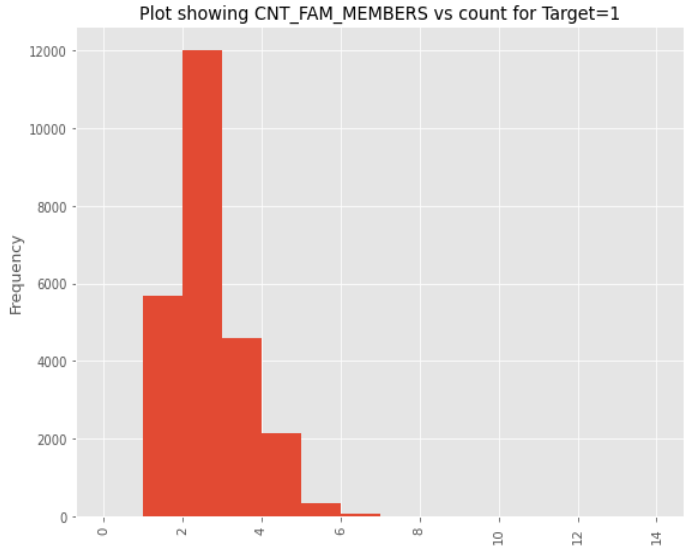
We can see that a family of 3 applies loan more often than the other families





COUNT OF FAMILY MEMBERS

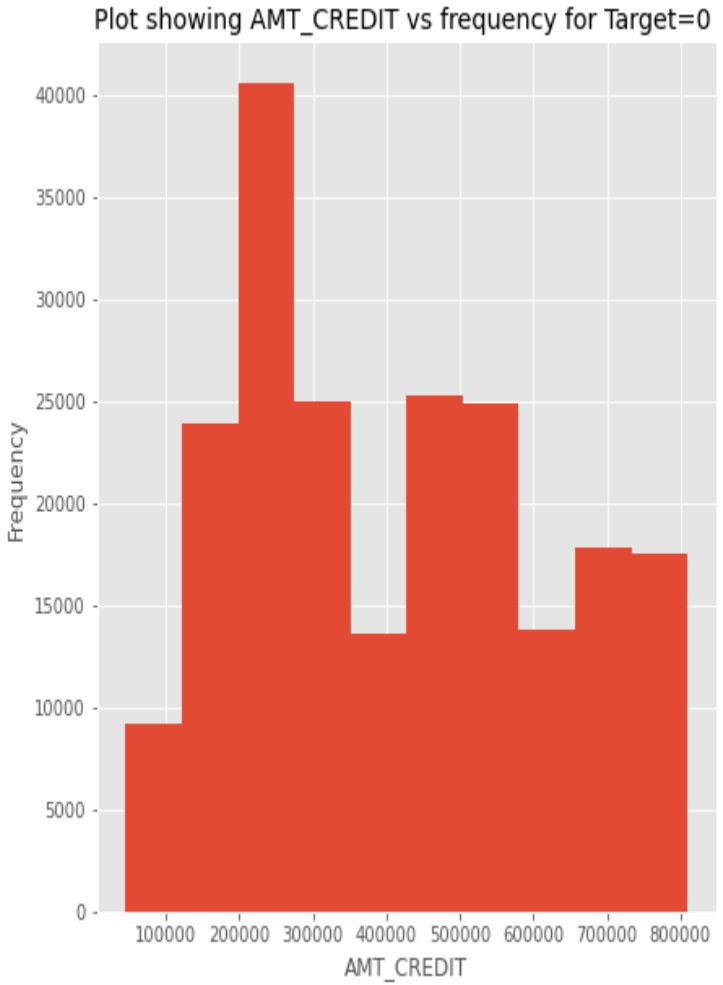


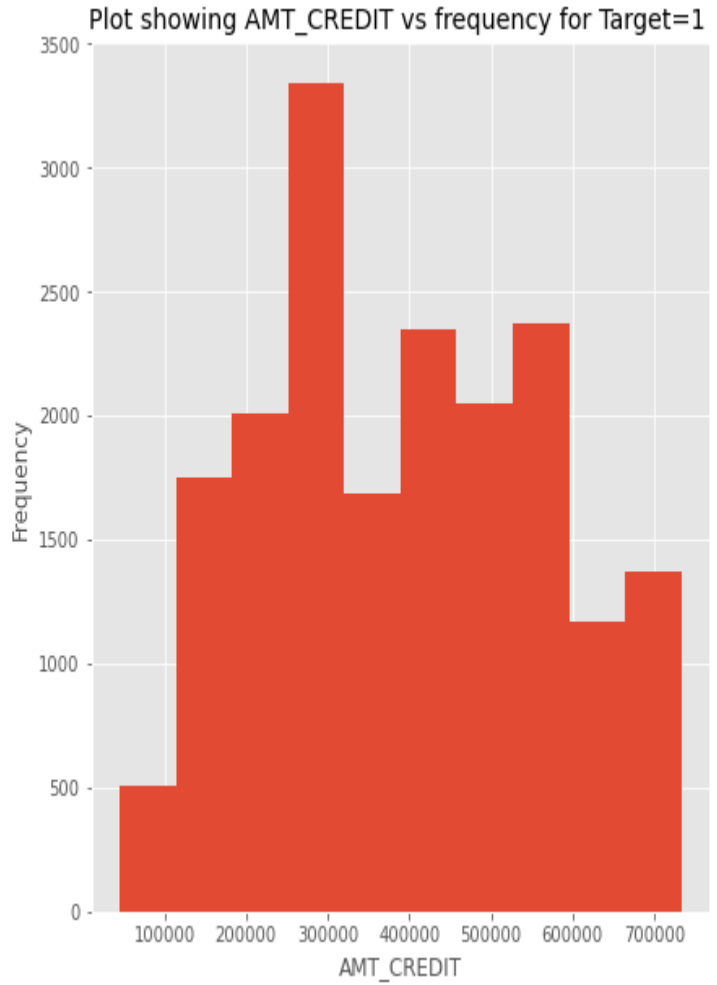


SEGMENTED UNIVARIATE ANALYSIS

CREDIT AMOUNT

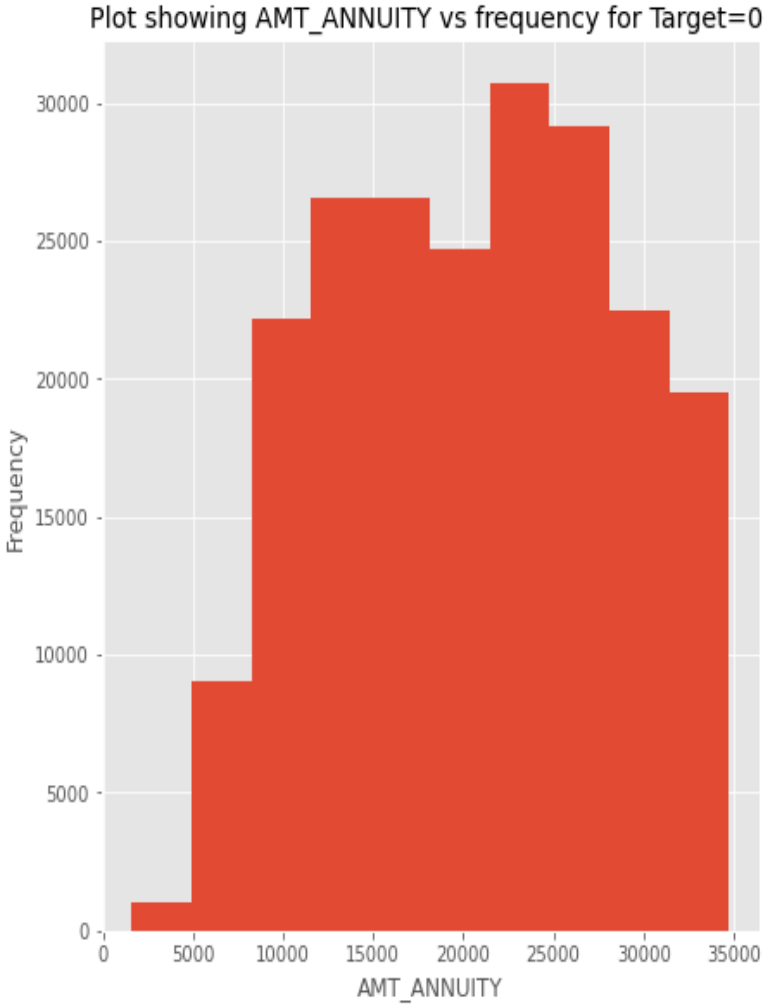
We can infer that as credit amount is above 300000 the chance of defaulting is slightly more

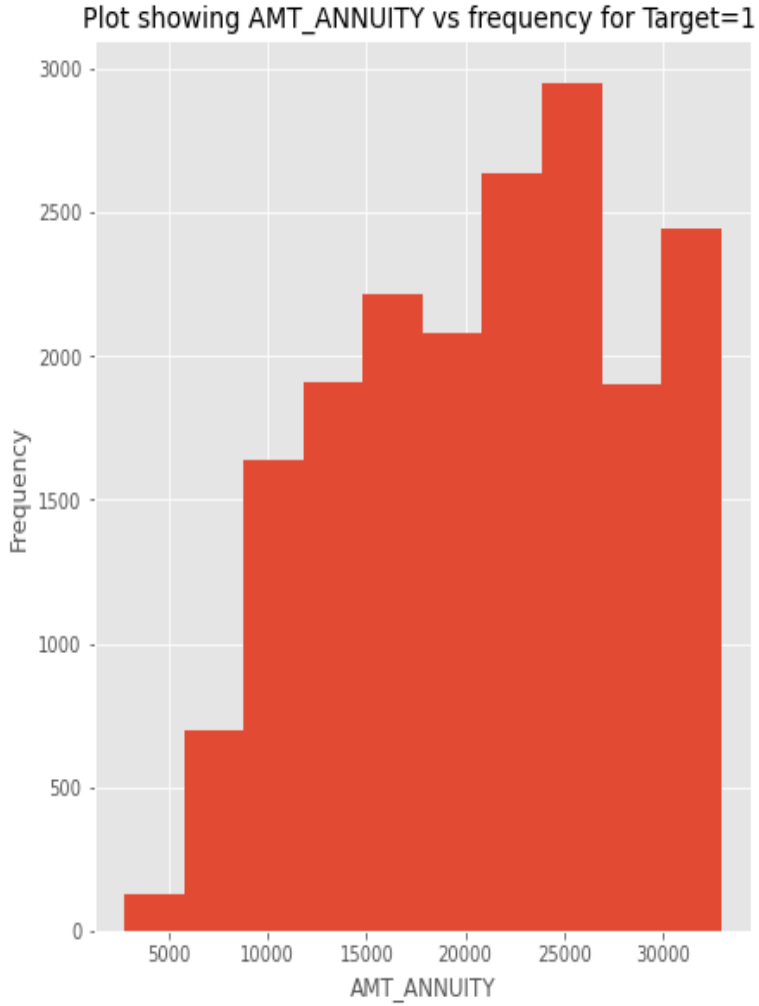




ANNUITY AMOUNT

Both the Target0 and Target1 data follow same pattern for AMT_ANNUITY.
Few exceptions occurs when the ANNUITY amount is 25000-30000.The defaulting is high around 25000 and 30000



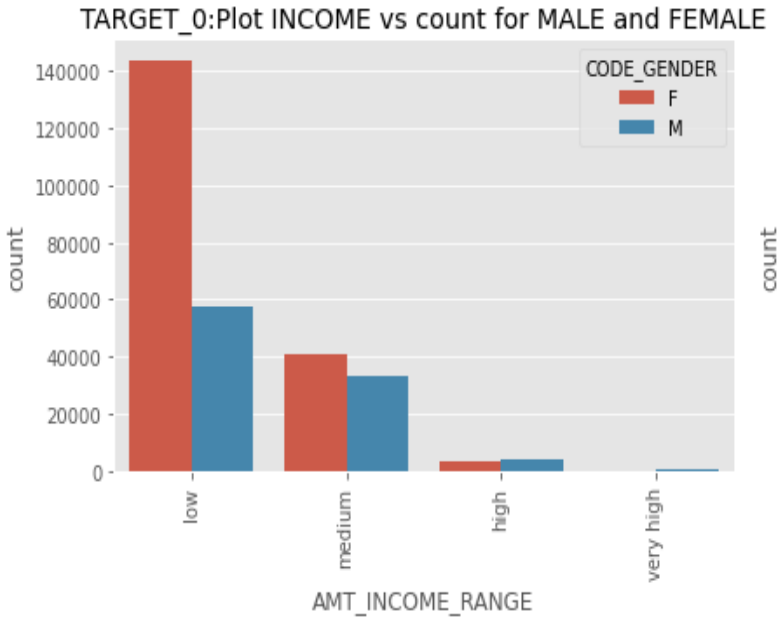


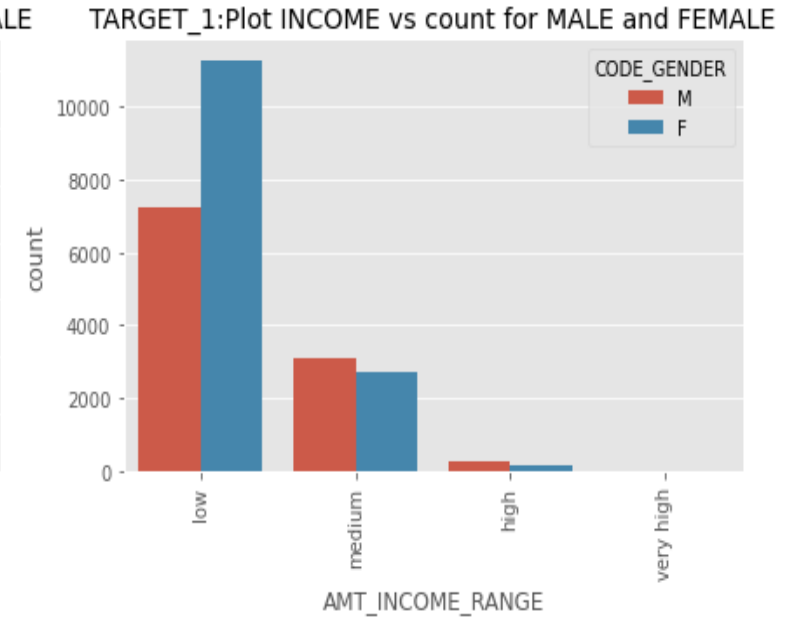
BIVARIATE ANALYSIS

INCOME AND GENDER

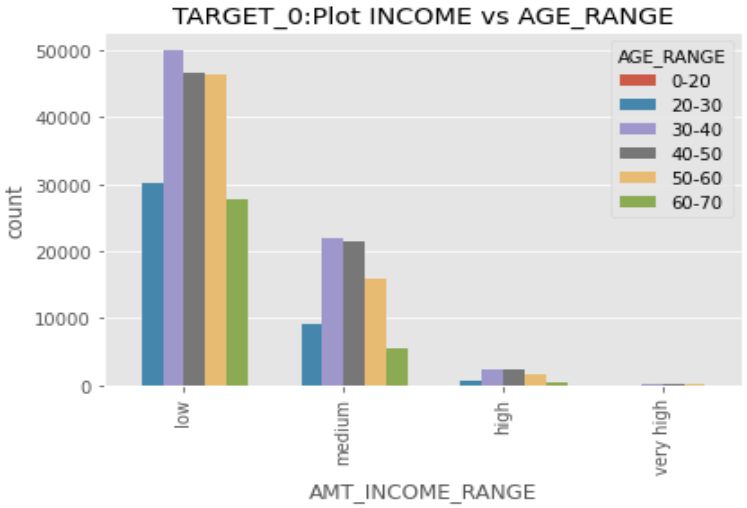
Female clients tend to have low income compared to male though their count for the loan application is high. Males with low income tend to default the payment more.

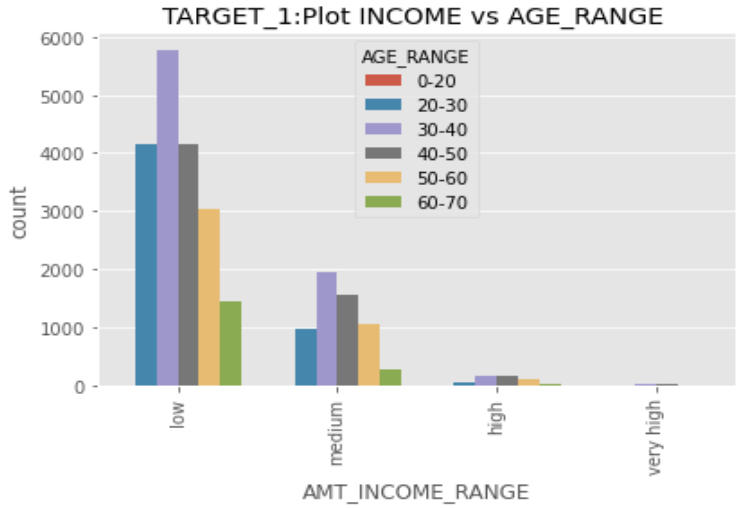
In the low level income range age group of 20-30 has a high chance of being a defaulter





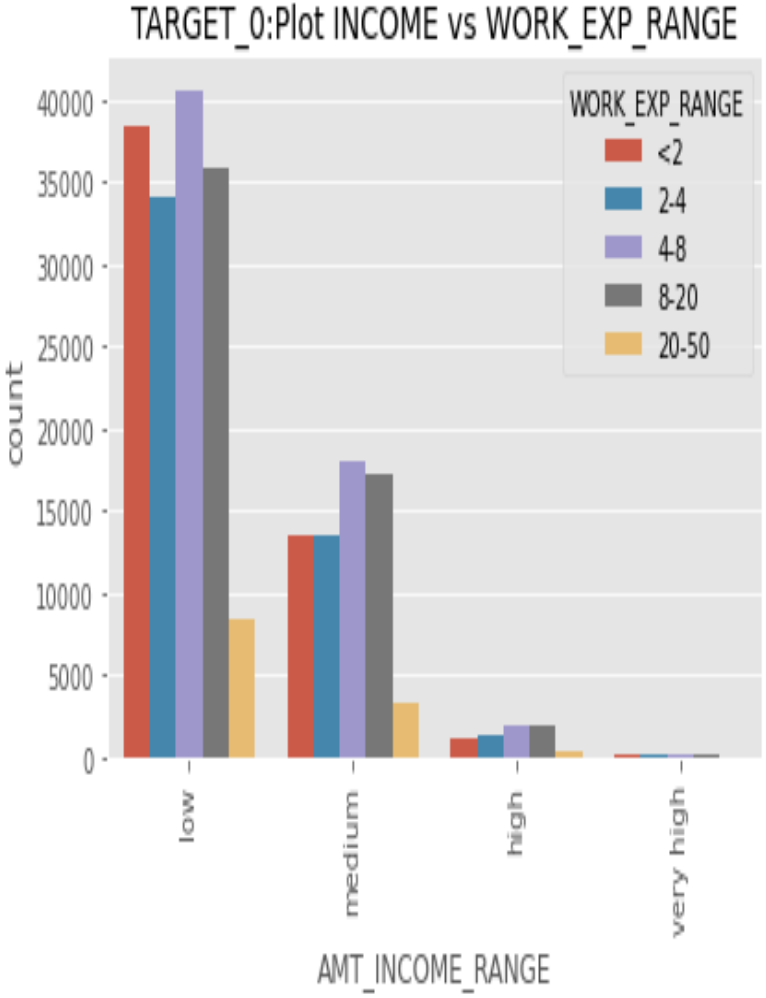
INCOME AND AGE

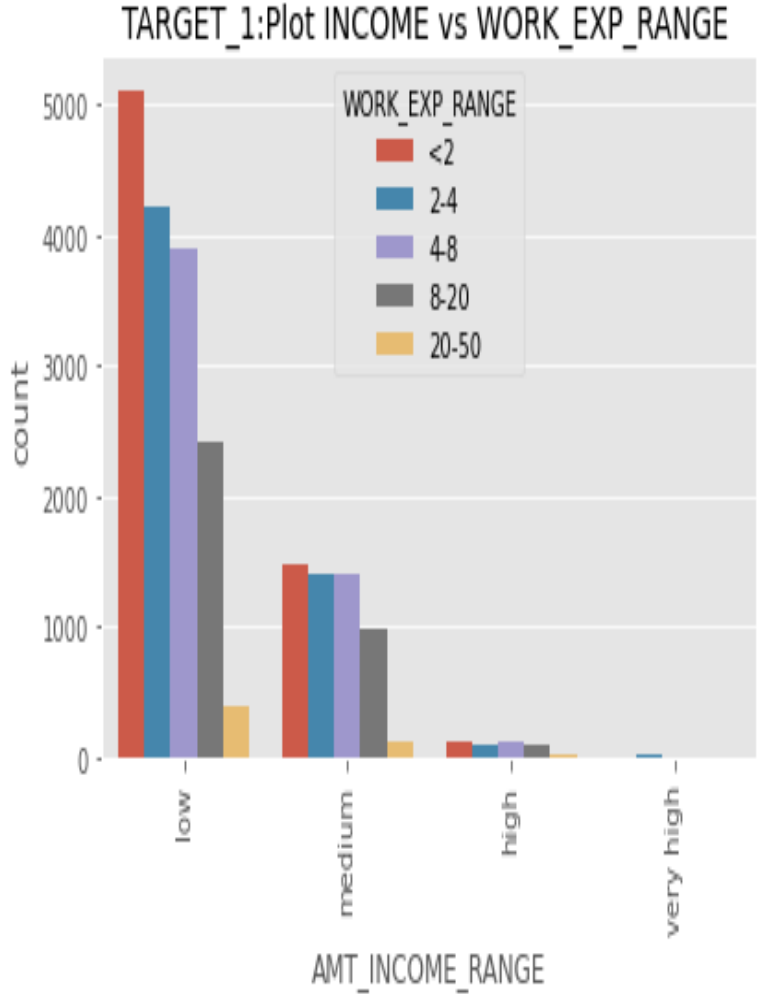




INCOME AND WORK EXPERIENCE

In the low-level income range clients with experience <2 years have a high chance of being a defaulter



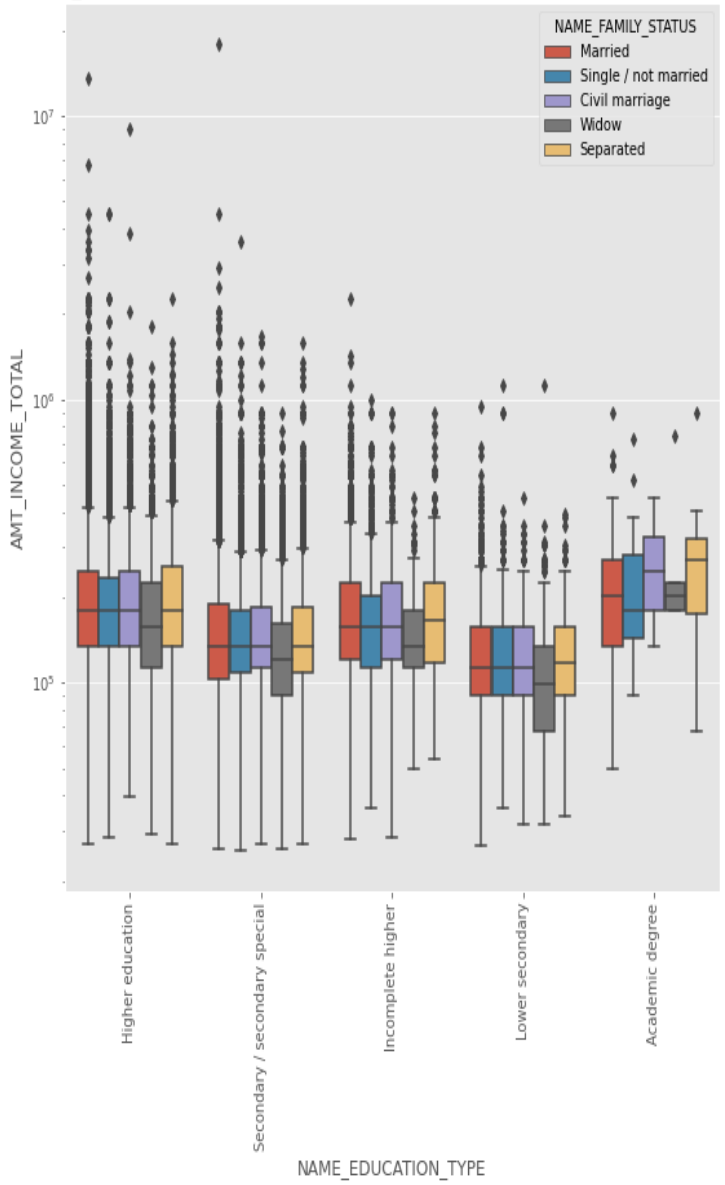


EDUCATION, INCOME, FAMILY_STATUS, ANNUITY and CREDIT AMOUNT

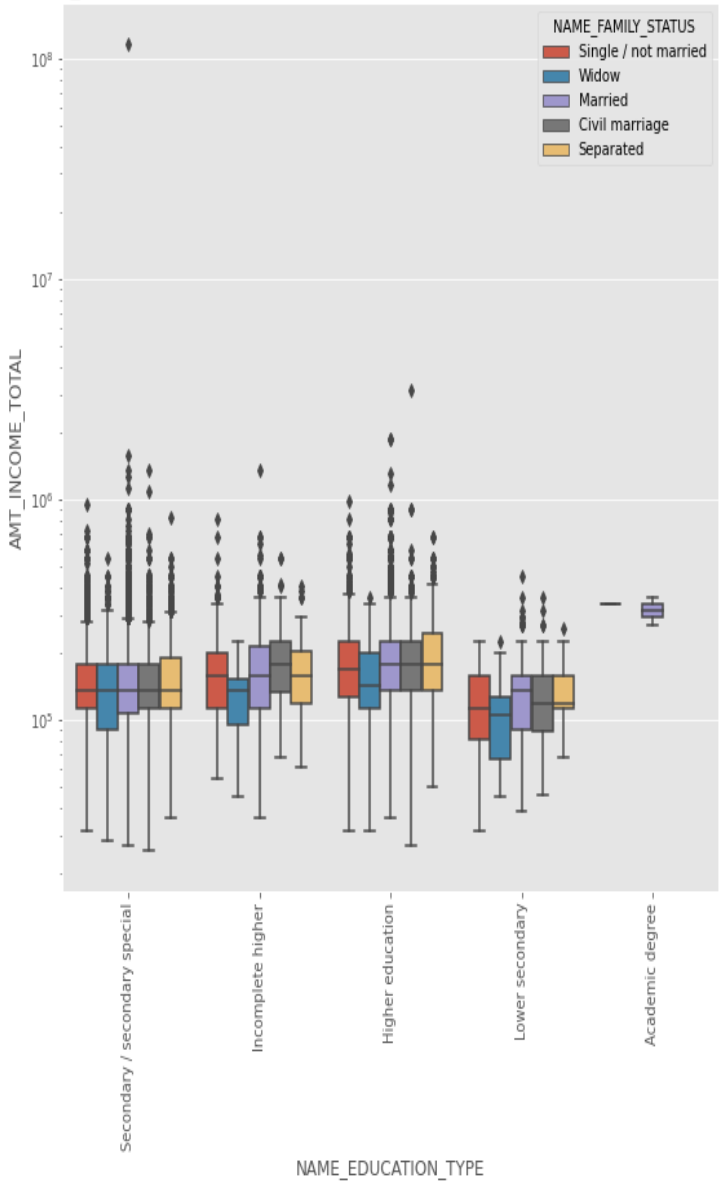
Clients with higher education and married have outliers for the Income amount value.They earn a high salary when compared to the Academic in all martial status

Clients with higher education and married have outliers for the Credit amount value, Income amount value, Annuity value. They have taken very high value for credit.

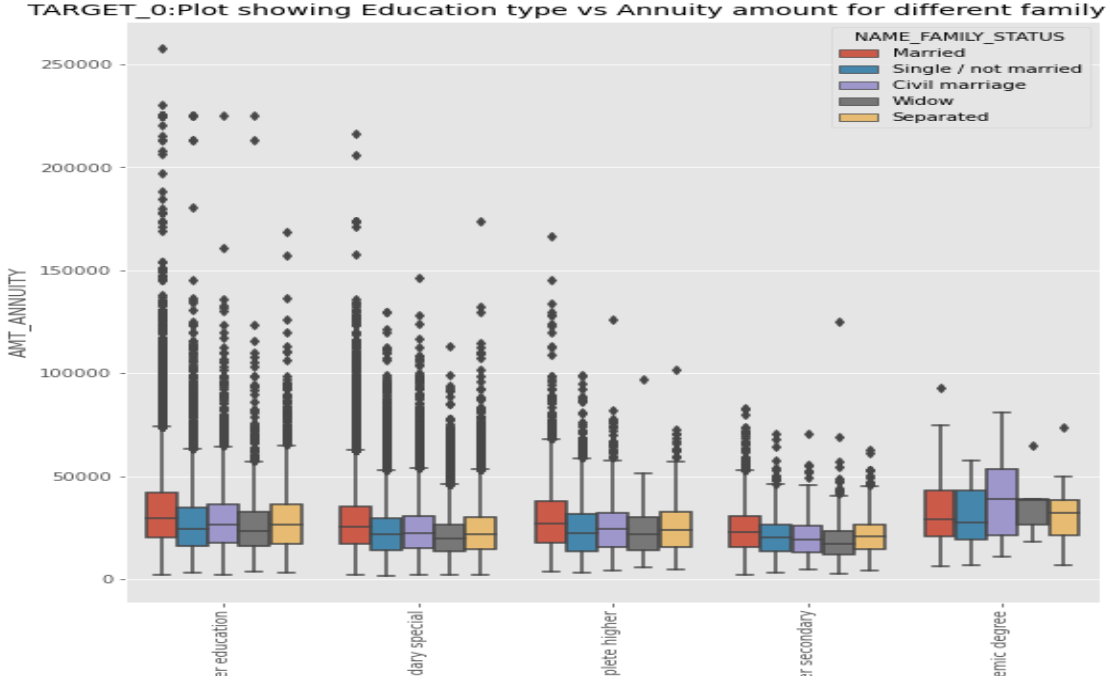
TARGET_0:Plot showing Education type vs Income amount for different family status



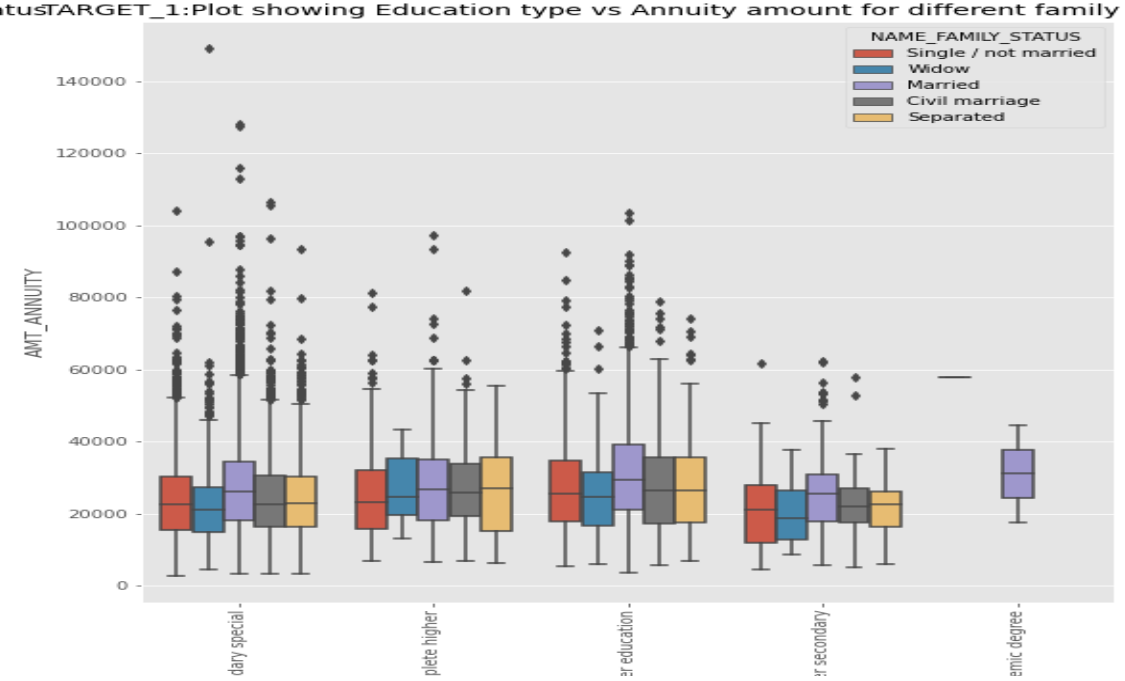
TARGET_1:Plot showing Education type vs Income amount for different family status



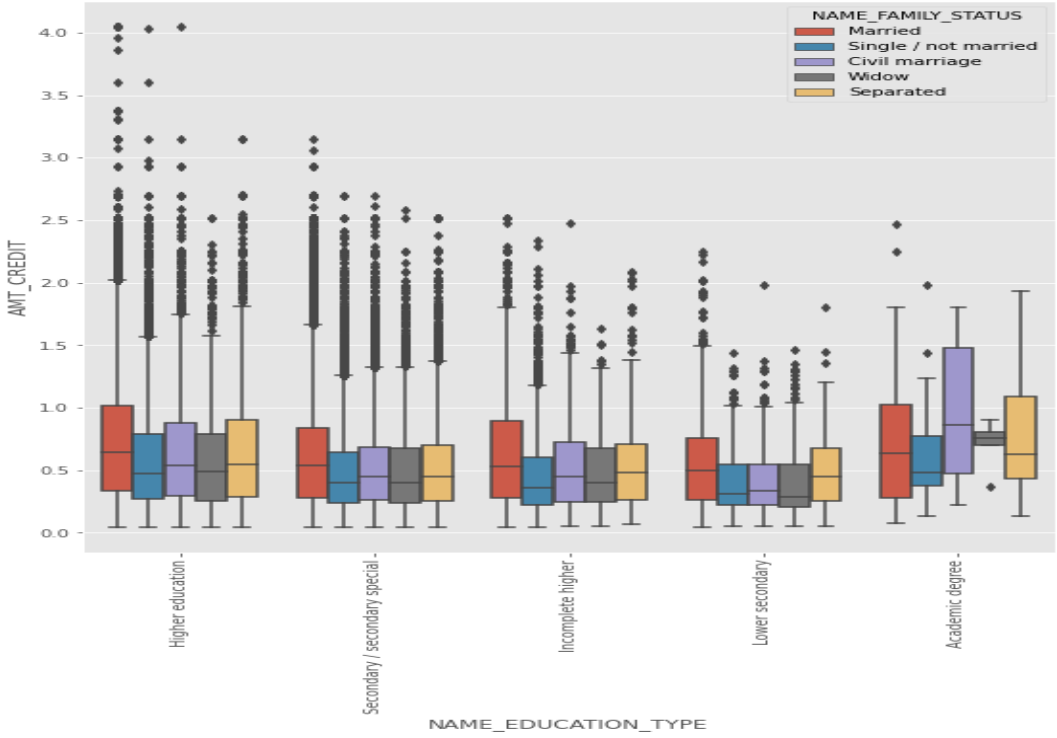
TARGET_0:Plot showing Education type vs Annuity amount for different family status



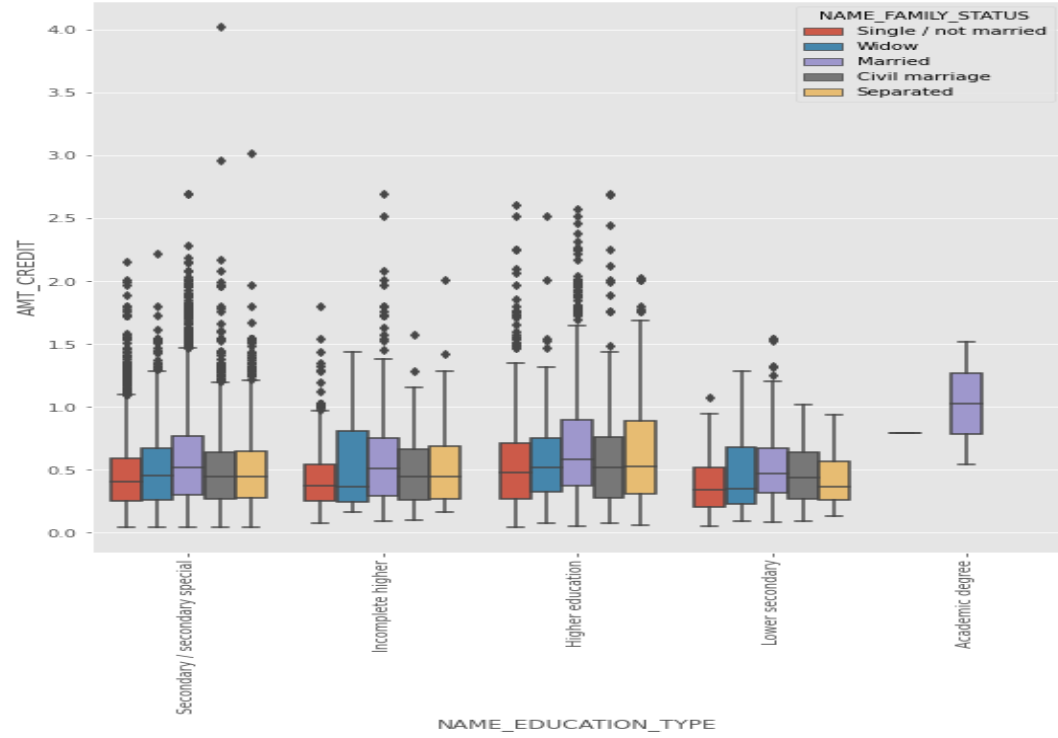
TARGET_1:Plot showing Education type vs Annuity amount for different family status



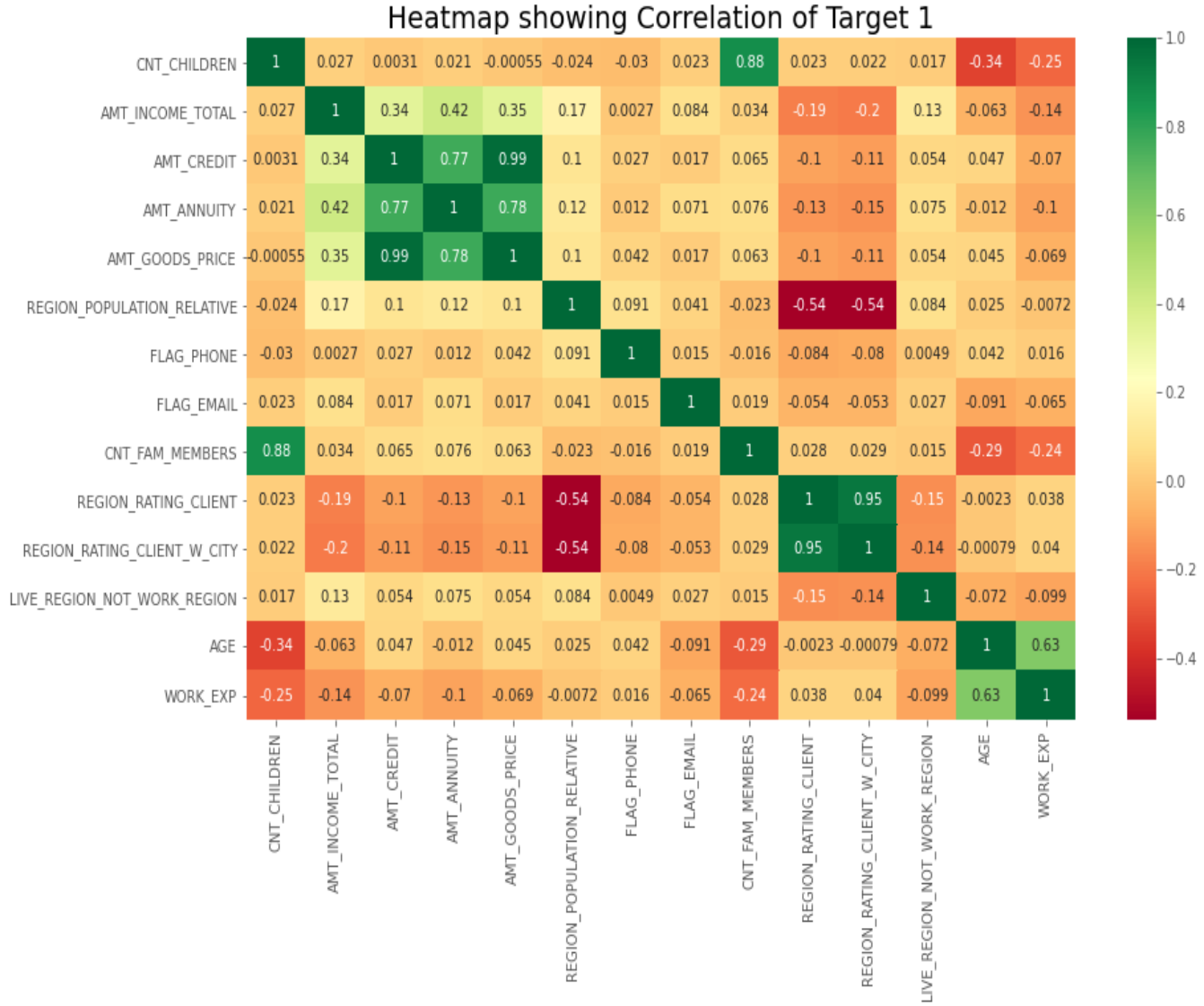
TARGET_0:Plot showing Education type vs credit amount for different family status



TARGET_1:Plot showing Education type vs credit amount for different family status



CORRELATION MATRIX

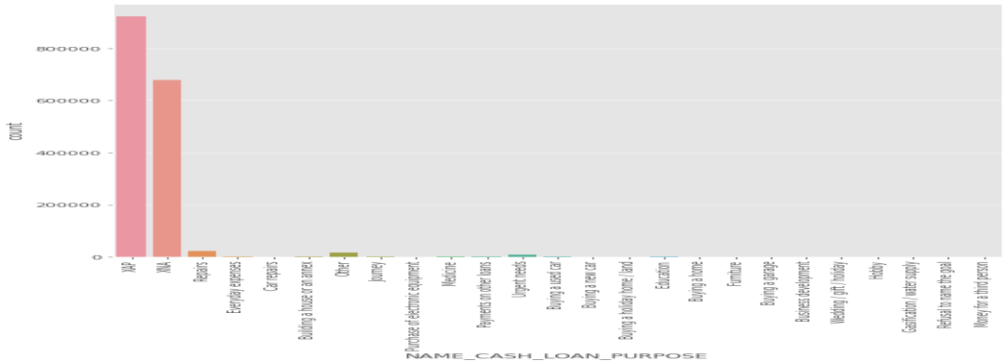


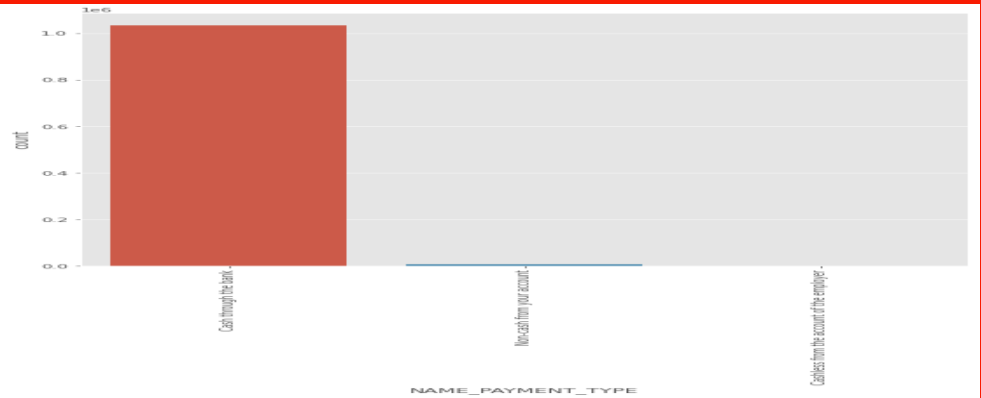
Top 5 Co-Relation Variable :

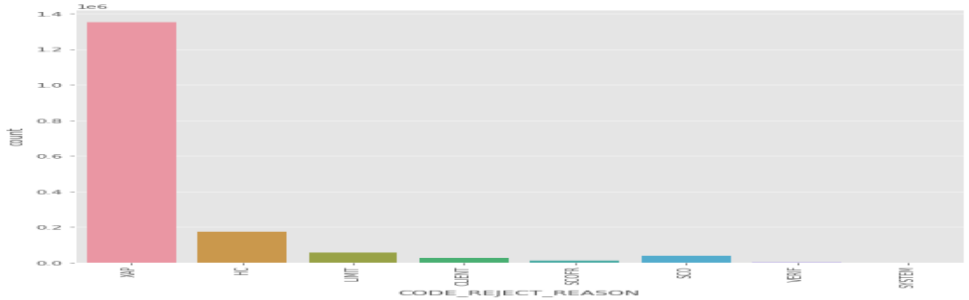
1. AMT_GOODS_PRICE & AMT_CREDIT
2. REGION_RATING_CLIENT_W_CLIENT & REGION_RATING_CLIENT
3. CNT_FAM_MEMBERS & CNT_CHILDREN
4. AMT_ANNUITY & AMT_GOODS_PRICE
5. AMT_ANNUITY & AMT_CREDIT

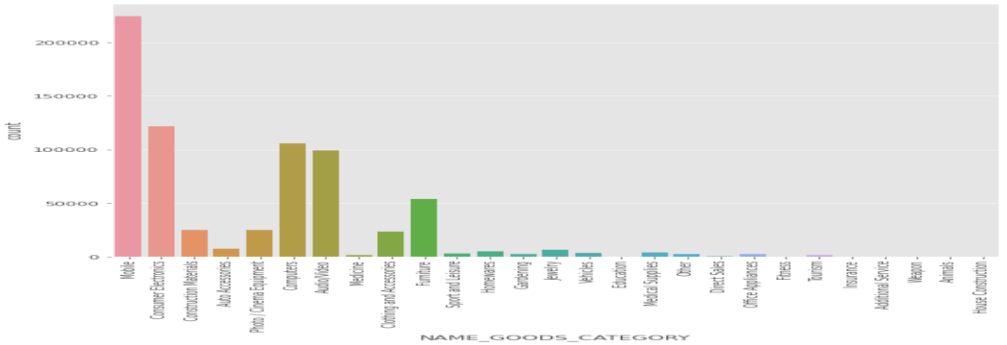
- We can infer the following from target0 and target1 heatmaps:
- The Age and Work Experience are positively correlated.i.e., as the Age increase the work experience increases . AS the age and work experience increase the annuity value tends to decrease. They are negatively correlated. As the credit value increases the annuity amt also increases As income increases credit and annuity increases As the children count increases the family members increases
- The Top 5 Co-Relation variables are same for Target 0 and Target 1

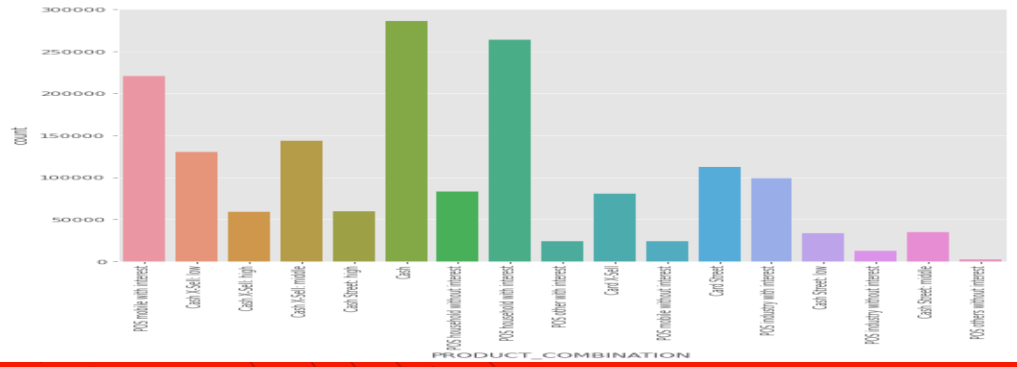
UNIVARIATE ANALYSIS- PREVIOUS APPLICATION

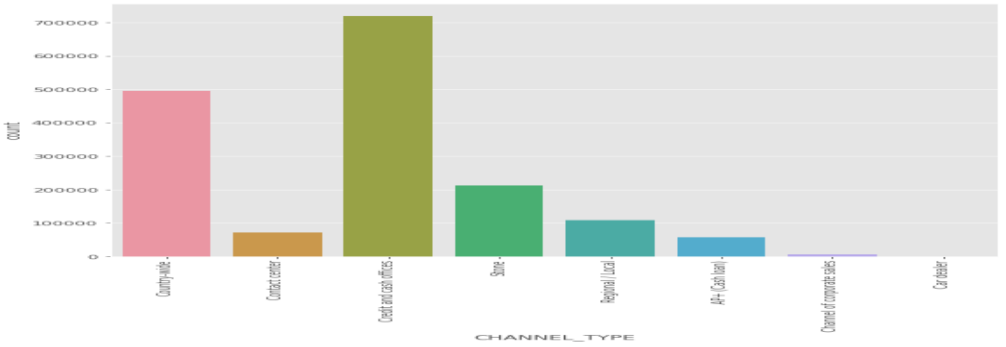












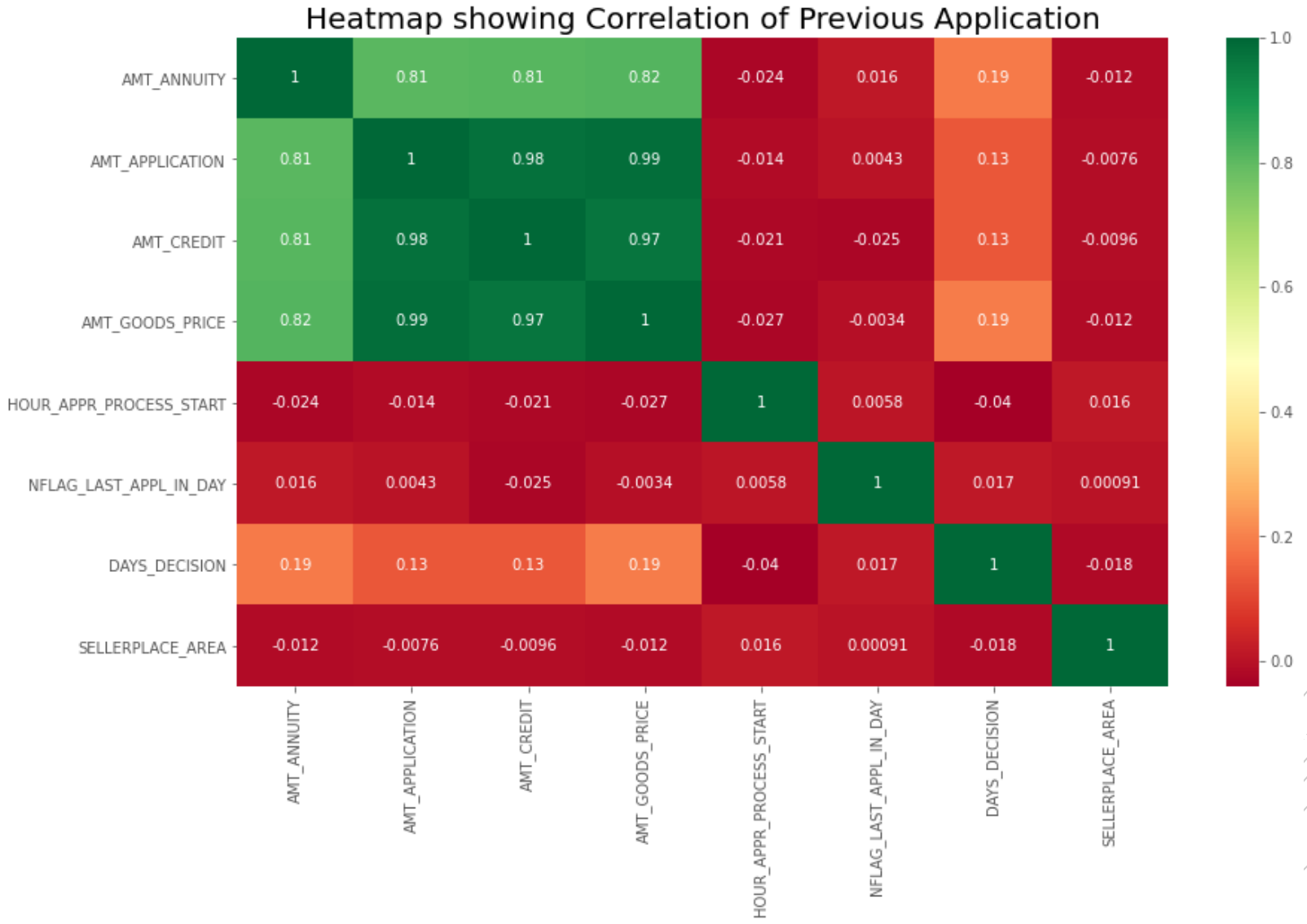
CASH LOAN PURPOSE
PAYMENT TYPE
REJECT REASON
GOODS CATEGORY,
CHANNEL TYPE,
PRODUCT COMBINATION

- Most Loan purpose was not recorded. XAP and XNA values are highest.
- Most clients preferred CASH as the mode of Payment.
- Primary reason for the Loan to get rejected is not recorded(XAP (81%)) followed by HC.
- Most clients applied for Mobile and Consumer electronics.
- Highest product combination is Cash followed by POS household with interest.
- Most clients were acquired from Credit and Cash Offices.

CO-RELATION HEATMAP FOR PREVOIUS APPLICATION

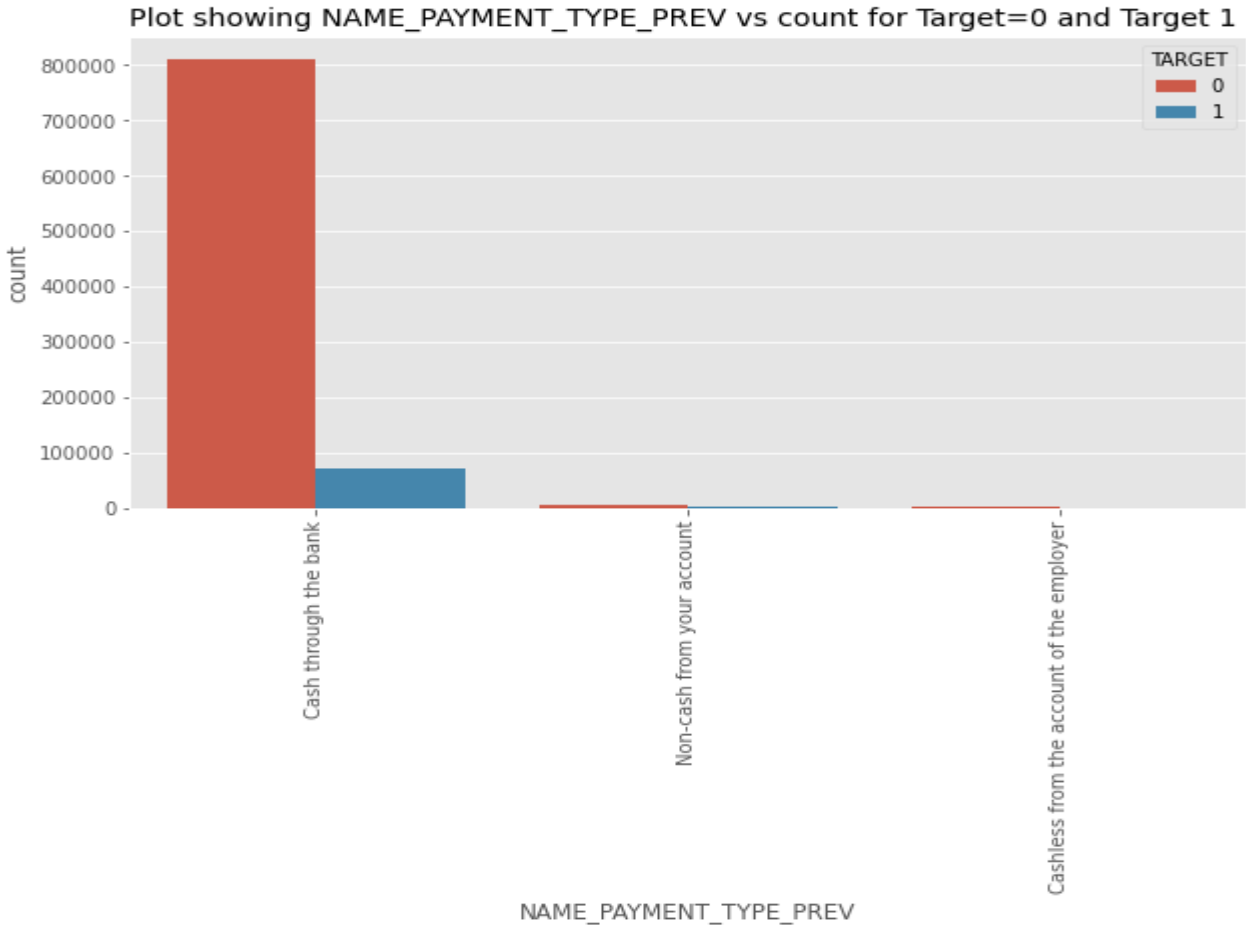
Top 5 Co-Relation Variable :

1. AMT_GOODS_PRICE & AMT_APPLICATION
2. AMT_APPLICATION & AMT_CREDIT
3. AMT_GOOD_PRICE & AMT_CREDIT
4. AMT_GOOD_PRICE & AMT_ANNUITY
5. AMT_CREDIT & AMT_ANNUITY

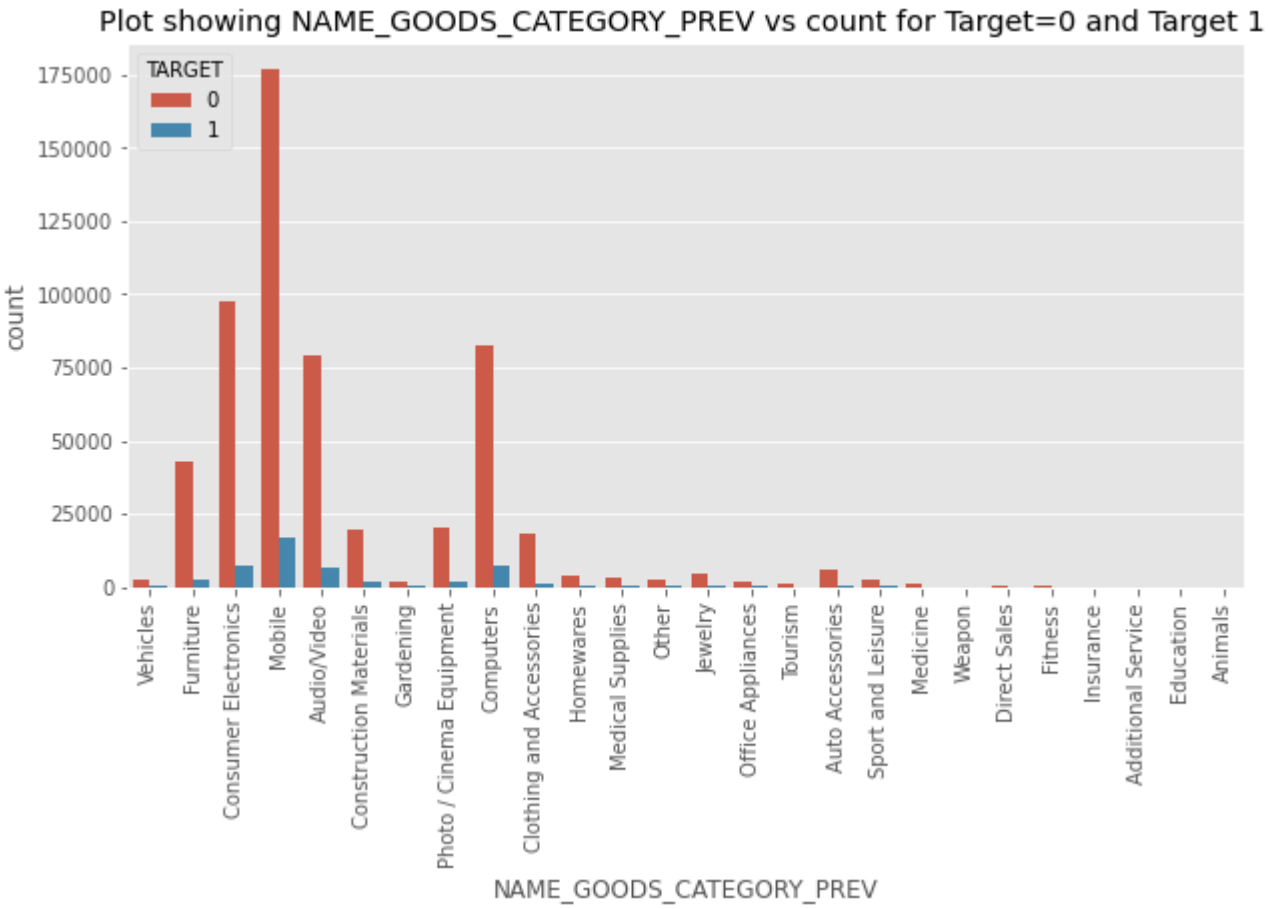


UNIVARIATE ANALYSIS – MERGED DATASET

PAYMENT TYPE

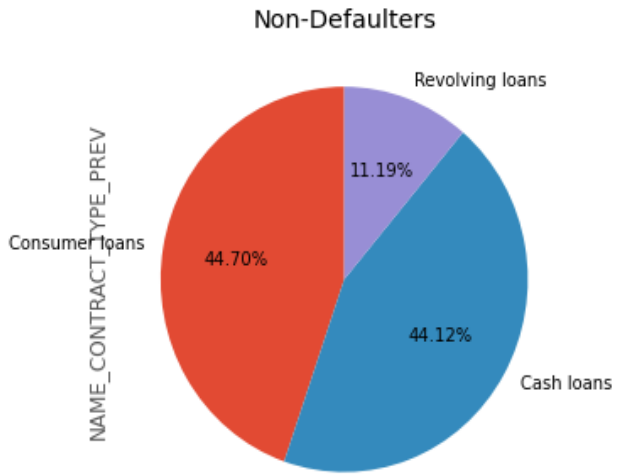


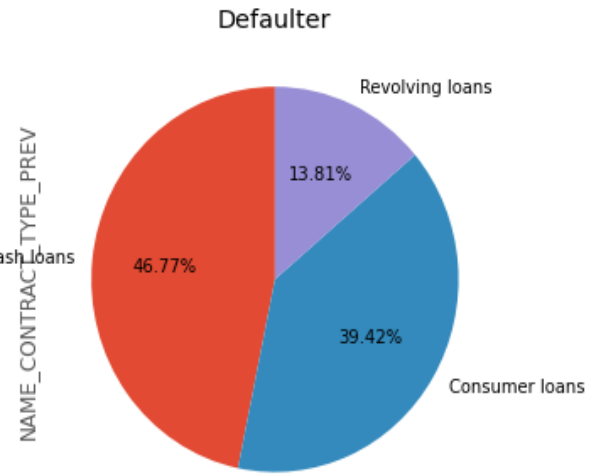
GOODS_CATEGORY



- Majority of payment for the prev application is made through cash. So the defaulters and non-defaulters have same payment type.
- For the prev application the defaulters and non-defaulters clients haven taken the higher count of loans for Mobiles followed by Consumer Electronics and Computers

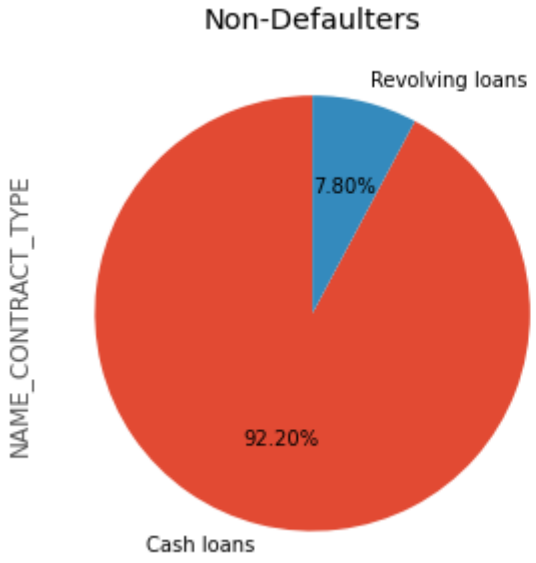
CONTRACT TYPE – PREVIOUS APPLICATION

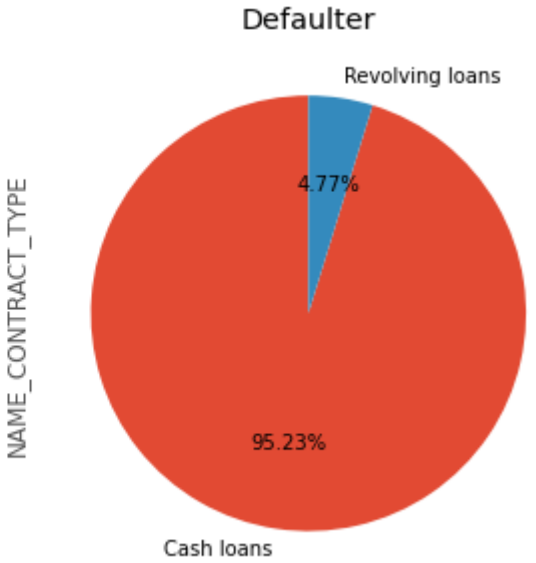




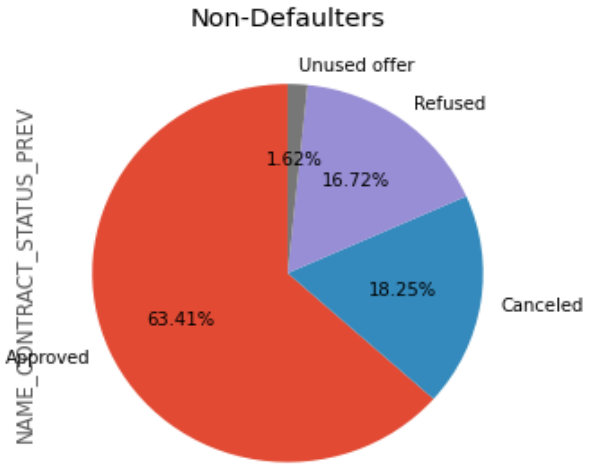
1. Current application has not consumer data unlike the previous application..
2. The non defaulters are high for Consumer loans from previous application.
3. Based on current application we have defaulters higher for cash loan compared to revolving loan.

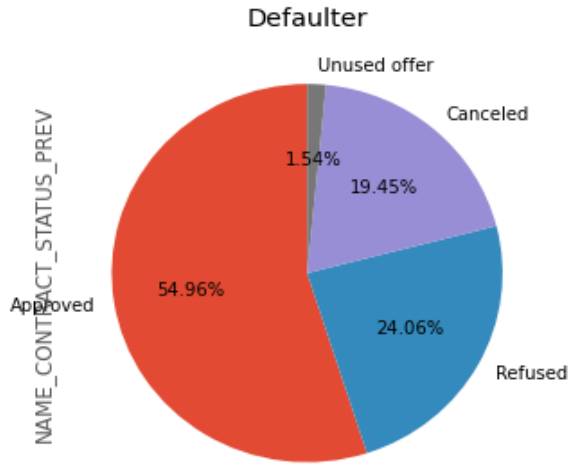
CONTRACT TYPE –CURRENT APPLICATION





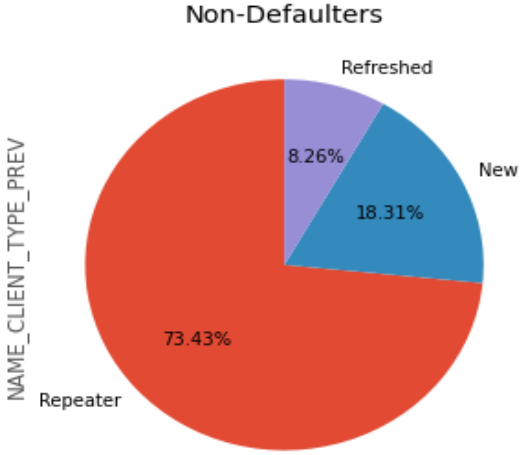
CONTRACT STATUS - PREV

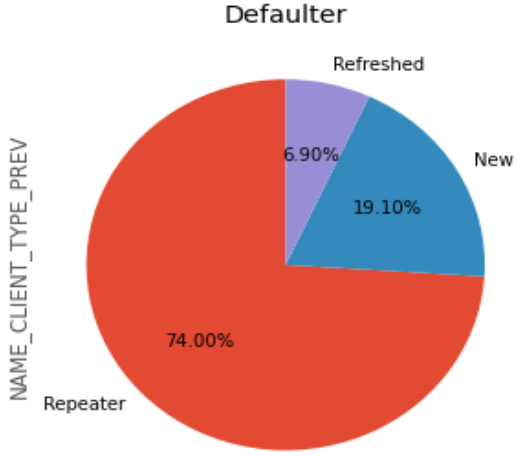




1 The percentage of defaulters from prev applications for refused is higher than that of the non-defaulters.
2.The percentage of non-defaulters from prev applications for approved is higher than that of the defaulters.
3.The percentages of defaulters previous applications from new and repeaters clients were higher than those of the non-defaulters.
4.The percentage of non-defaulters previous applications from refreshed clients was higher than those of the defaulters.

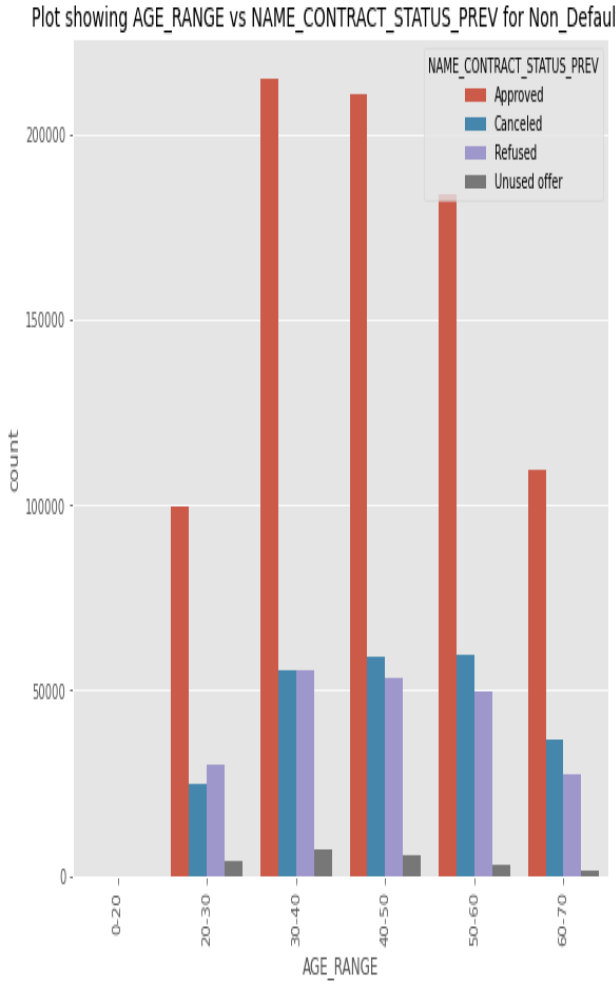
CLIENT TYPE - PREV

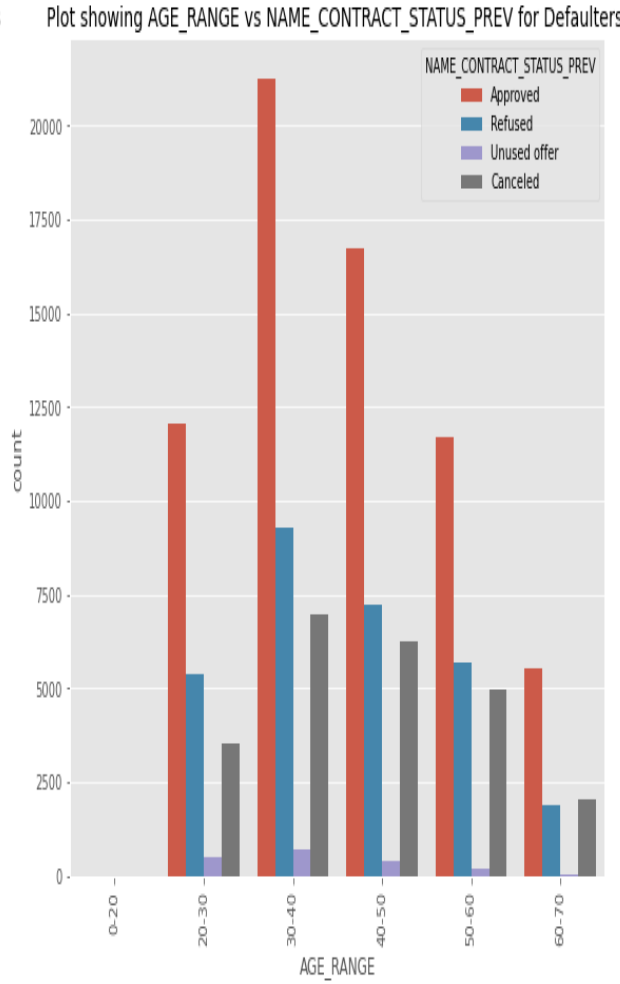




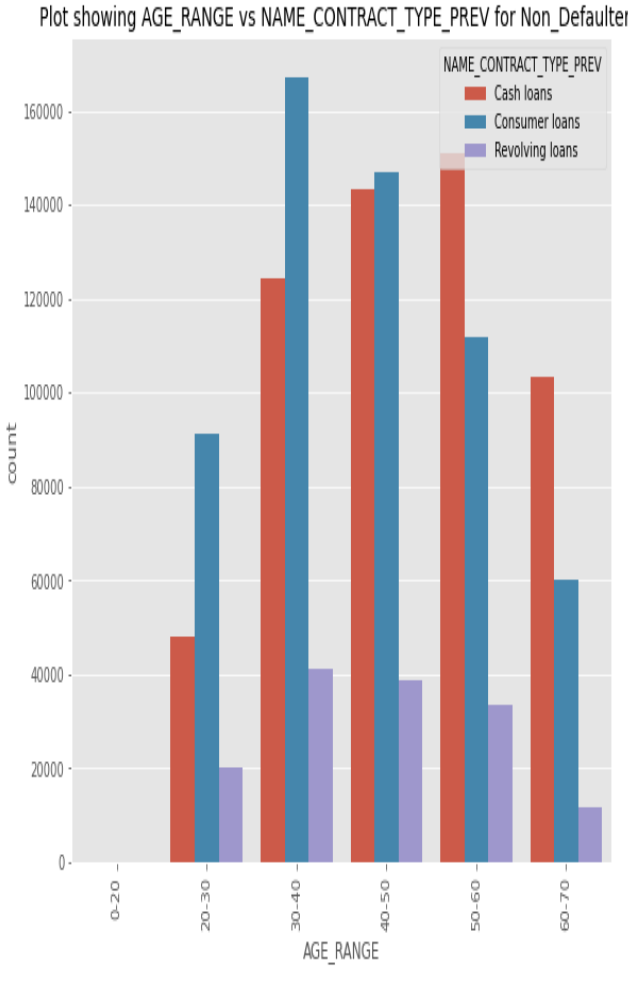
BIVARIATE ANALYSIS – MERGED DATASET

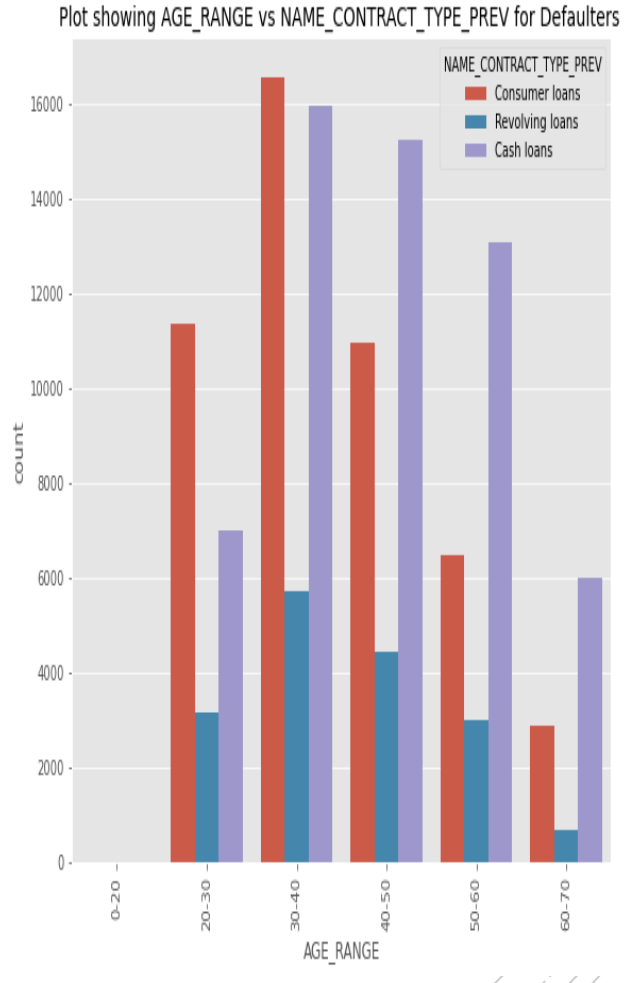
AGE , CONTRACT STATUS_PREV





AGE , CONTRACT TYPE_PREV

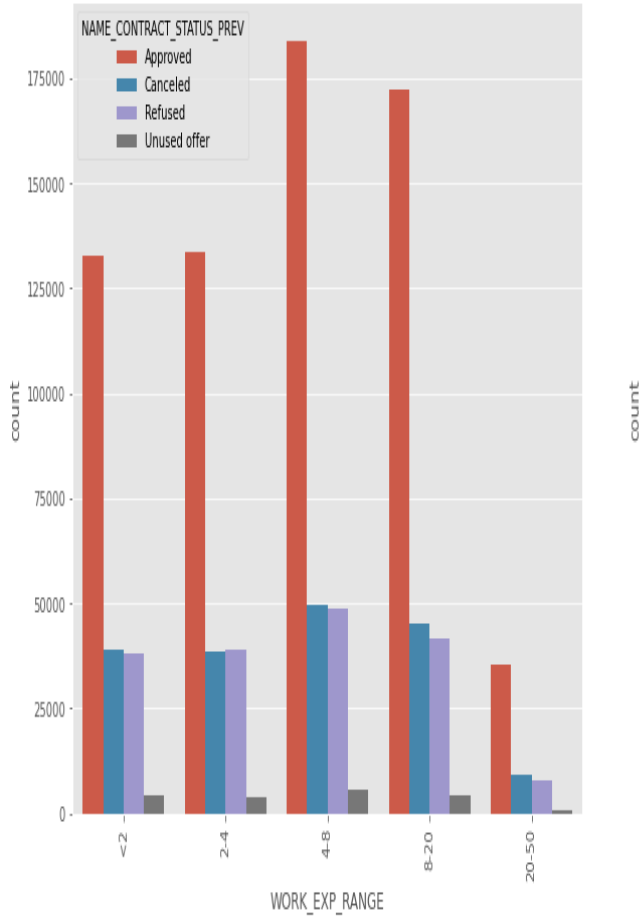




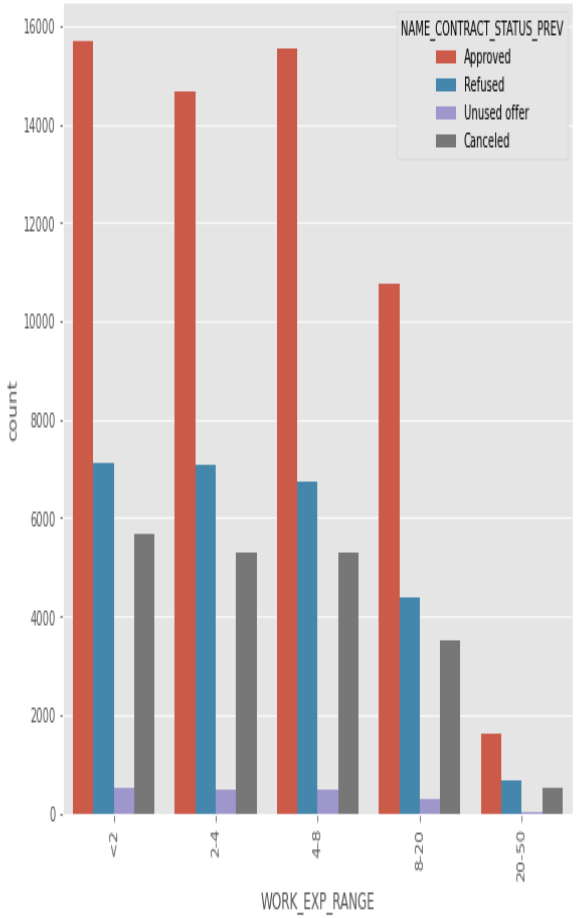
The clients in the age group of 30-60 having Approved and Unused offer from previous application status and those applied for Cash/Consumer loan tend to be non defaulters. The clients in the age group of 20-30 having Cancelled and Refused from previous application status and those applied for Revolving loans have more chance to be defaulters.

WORK EXPERIENCE & CONTRACT STATUS_PREV

Plot showing WORK_EXP_RANGE vs NAME_CONTRACT_STATUS_PREV for Non_Defaulters

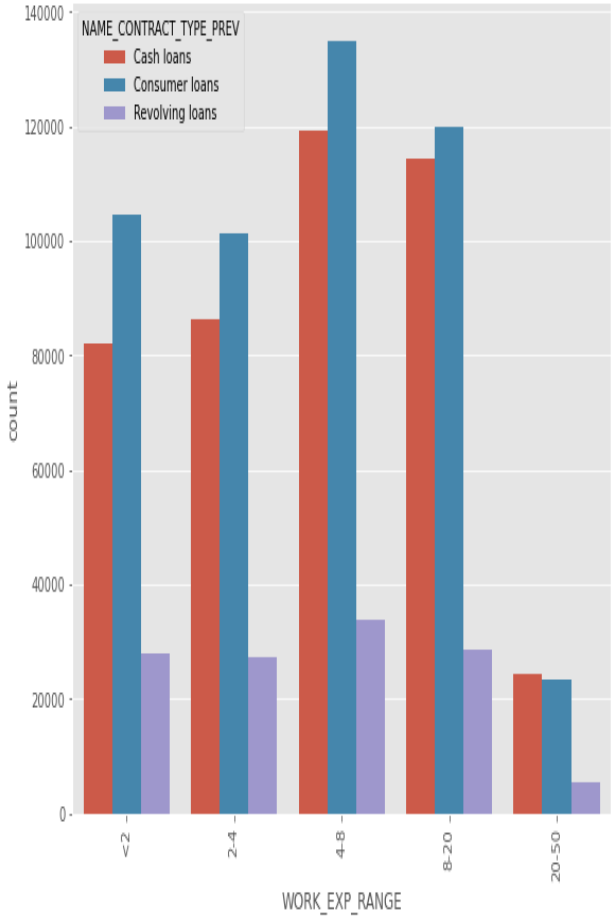


Plot showing WORK_EXP_RANGE vs NAME_CONTRACT_STATUS_PREV for Defaulters

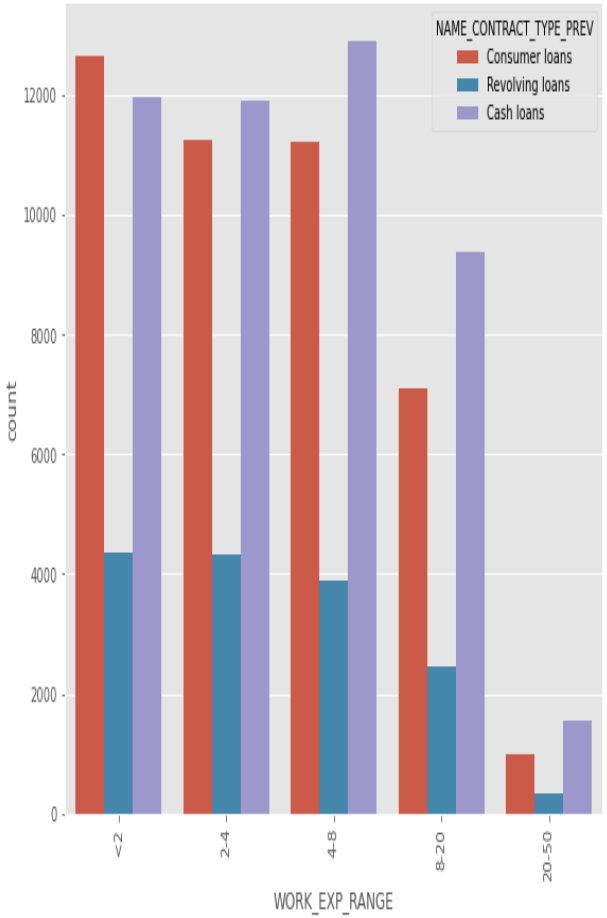


WORK EXPERIENCE & CONTRACT TYPE_PREV

Plot showing WORK_EXP_RANGE vs NAME_CONTRACT_TYPE_PREV for Non_Defaulters



Plot showing WORK_EXP_RANGE vs NAME_CONTRACT_TYPE_PREV for Defaulters



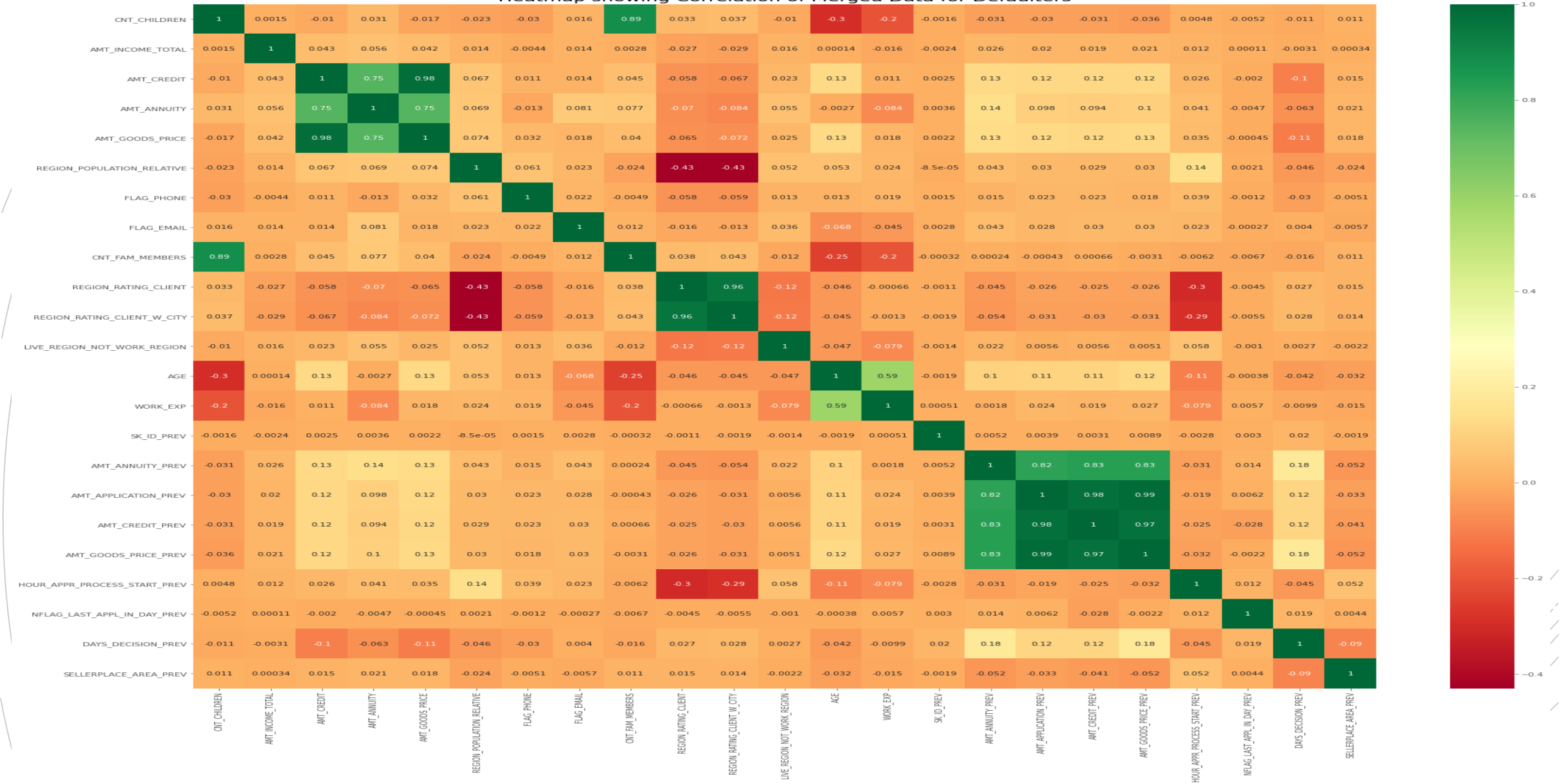
The clients in the Work Exp group of 4-20 having Approved from previous application status and those applied for Cash/Consumer loan tend to be non defaulters. The clients in the Work Exp group of 0-2 having irrespective of their status from previous application status and those applied for all loans types have more chance to be defaulters.

PAIRPLOT FOR DEFAULTERS

PAIRPLOT FOR NON - DEFAULTERS

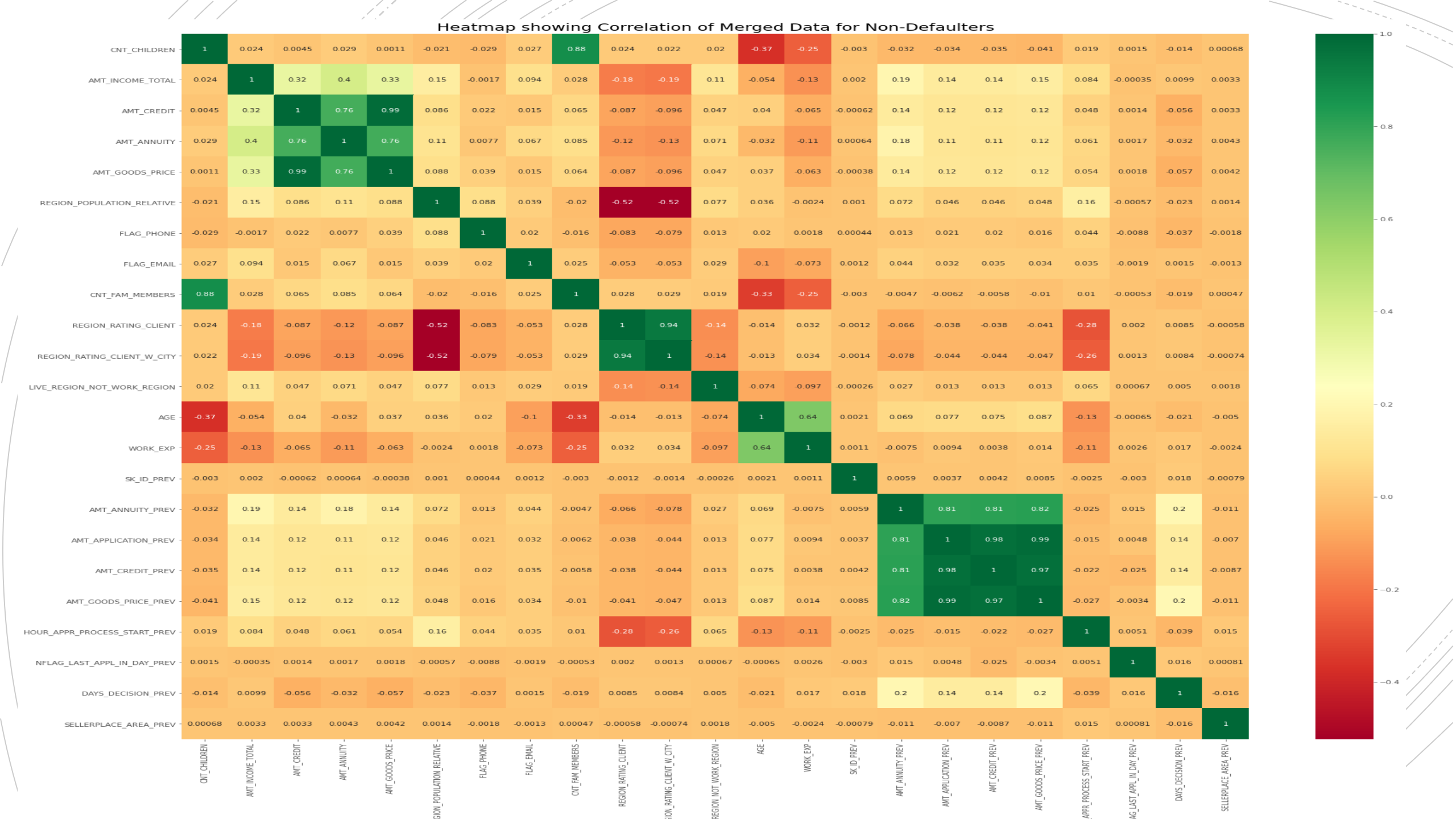
From the above two pairplots we can observe that:
High correlation exists between AMT_APPLICATION_PREV and AMT_GOODS_PRICE_PREV AMT_CREDIT_PREV and AMT_GOODS_PRICE_PREV AMT_APPLICATION_PREV and AMT_CREDIT_PREV AMT_GOODS_PRICE and AMT_CREDIT

CORRELATION MATRIX – MERGED DATASET



The top 10 correlations for defaulters

Attribute1	Attribute 2	Correlation Value
AMT_APPLICATION_PREV	AMT_GOODS_PRICE_PREV	0.985655
AMT_CREDIT	AMT_GOODS_PRICE	0.982525
AMT_APPLICATION_PREV	AMT_CREDIT_PREV	0.975377
AMT_CREDIT_PREV	AMT_GOODS_PRICE_PREV	0.969027
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956483
CNT_CHILDREN	CNT_FAM_MEMBERS	0.886300
AMT_ANNUITY_PREV	AMT_GOODS_PRICE_PREV	0.832264
AMT_ANNUITY_PREV	AMT_CREDIT_PREV	0.829383
AMT_ANNUITY_PREV	AMT_APPLICATION_PREV	0.816557
AMT_CREDIT	AMT_ANNUITY	0.745679


Heatmap showing Correlation of Merged Data for Non-Defaulters
CNT_CHILDREN 1 0.024 0.0045 0.029 0.0011 -0.021 -0.029 0.027 0.88 0.024 0.022 0.02 -0.37 -0.25 -0.003 -0.032 -0.034 -0.035 -0.041 0.019 0.0015 -0.014 0.00068
AMT_INCOME_TOTAL 0.024 1 0.32 0.4 0.33 0.15 -0.0017 0.094 0.028 -0.18 -0.19 0.11 -0.054 -0.13 0.002 0.19 0.14 0.14 0.15 0.084 -0.00035 0.0099 0.0033
AMT_CREDIT 0.0045 0.32 1 0.76 0.99 0.086 0.022 0.015 0.065 -0.087 -0.096 0.047 0.04 -0.065 -0.00062 0.14 0.12 0.12 0.12 0.048 0.0014 -0.056 0.0033
AMT_ANNUITY 0.029 0.4 0.76 1 0.76 0.11 0.0077 0.067 0.085 -0.12 -0.13 0.071 -0.032 -0.11 0.00064 0.18 0.11 0.11 0.12 0.061 0.0017 -0.032 0.0043
AMT_GOODS_PRICE 0.0011 0.33 0.99 0.76 1 0.088 0.039 0.015 0.064 -0.087 -0.096 0.047 0.037 -0.063 -0.00038 0.14 0.12 0.12 0.12 0.054 0.0018 -0.057 0.0042
REGION_POPULATION_RELATIVE -0.021 0.15 0.086 0.11 0.088 1 0.088 0.039 -0.02 -0.52 -0.52 0.077 0.036 -0.0024 0.001 0.072 0.046 0.046 0.048 0.16 -0.00057 -0.023 0.0014
FLAG_PHONE -0.029 -0.0017 0.022 0.0077 0.039 0.088 1 0.02 -0.016 -0.083 -0.079 0.013 0.02 0.0018 0.00044 0.013 0.021 0.02 0.016 0.044 -0.0088 -0.037 -0.0018
FLAG_EMAIL 0.027 0.094 0.015 0.067 0.015 0.039 0.02 1 0.025 -0.053 -0.053 0.029 -0.1 -0.073 0.0012 0.044 0.032 0.035 0.034 0.035 -0.0019 0.0015 -0.0013
CNT_FAM_MEMBERS 0.88 0.028 0.065 0.085 0.064 -0.02 -0.016 0.025 1 0.028 0.029 0.019 -0.33 -0.25 -0.003 -0.0047 -0.0062 -0.0058 -0.01 0.01 -0.00053 -0.019 0.00047
REGION_RATING_CLIENT 0.024 -0.18 -0.087 -0.12 -0.087 -0.52 -0.083 -0.053 0.028 1 0.94 -0.14 -0.014 0.032 -0.0012 -0.066 -0.038 -0.038 -0.041 -0.28 0.002 0.0085 -0.00058
REGION_RATING_CLIENT_W_CITY 0.022 -0.19 -0.096 -0.13 -0.096 -0.52 -0.079 -0.053 0.029 0.94 1 -0.14 -0.013 0.034 -0.0014 -0.078 -0.044 -0.044 -0.047 -0.26 0.0013 0.0084 -0.00074
LIVE_REGION_NOT_WORK_REGION 0.02 0.11 0.047 0.071 0.047 0.077 0.013 0.029 0.019 -0.14 -0.14 1 -0.074 -0.097 -0.00026 0.027 0.013 0.013 0.013 0.065 0.00067 0.005 0.0018
AGE -0.37 -0.054 0.04 -0.032 0.037 0.036 0.02 -0.1 -0.33 -0.014 -0.013 -0.074 1 0.64 0.0021 0.069 0.077 0.075 0.087 -0.13 -0.00065 -0.021 -0.005
WORK_EXP -0.25 -0.13 -0.065 -0.11 -0.063 -0.0024 0.0018 -0.073 -0.25 0.032 0.034 -0.097 0.64 1 0.0011 -0.0075 0.0094 0.0038 0.014 -0.11 0.0026 0.017 -0.0024
SK_ID_PREV -0.003 0.002 -0.00062 0.00064 -0.00038 0.001 0.00044 0.0012 -0.003 -0.0012 -0.0014 -0.00026 0.0021 0.0011 1 0.0059 0.0037 0.0042 0.0085 -0.0025 -0.003 0.018 -0.00079
AMT_ANNUITY_PREV -0.032 0.19 0.14 0.18 0.14 0.072 0.013 0.044 -0.0047 -0.066 -0.078 0.027 0.069 -0.0075 0.0059 1 0.81 0.81 0.82 -0.025 0.015 0.2 -0.011
AMT_APPLICATION_PREV -0.034 0.14 0.12 0.11 0.12 0.046 0.021 0.032 -0.0062 -0.038 -0.044 0.013 0.077 0.0094 0.0037 0.81 1 0.98 0.99 -0.015 0.0048 0.14 -0.007
AMT_CREDIT_PREV -0.035 0.14 0.12 0.11 0.12 0.046 0.02 0.035 -0.0058 -0.038 -0.044 0.013 0.075 0.0038 0.0042 0.81 0.98 1 0.97 -0.022 -0.025 0.14 -0.0087
AMT_GOODS_PRICE_PREV -0.041 0.15 0.12 0.12 0.12 0.048 0.016 0.034 -0.01 -0.041 -0.047 0.013 0.087 0.014 0.0085 0.82 0.99 0.97 1 -0.027 -0.0034 0.2 -0.011
HOUR_APPR_PROCESS_START_PREV 0.019 0.084 0.048 0.061 0.054 0.16 0.044 0.035 0.01 -0.28 -0.26 0.065 -0.13 -0.11 -0.0025 -0.025 -0.015 -0.022 -0.027 1 0.0051 -0.039 0.015
NFLAG_LAST_APPL_IN_DAY_PREV 0.0015 -0.00035 0.0014 0.0017 0.0018 -0.00057 -0.0088 -0.0019 -0.00053 0.002 0.0013 0.00067 -0.00065 0.0026 -0.003 0.015 0.0048 -0.025 -0.0034 0.0051 1 0.016 0.00081
DAYS_DECISION_PREV -0.014 0.0099 -0.056 -0.032 -0.057 -0.023 -0.037 0.0015 -0.019 0.0085 0.0084 0.005 -0.021 0.017 0.018 0.2 0.14 0.14 0.2 -0.039 0.016 1 -0.016
SELLERPLACE_AREA_PREV 0.00068 0.0033 0.0033 0.0043 0.0042 0.0014 -0.0018 -0.0013 0.00047 -0.00058 -0.00074 0.0018 -0.005 -0.0024 -0.00079 -0.011 -0.007 -0.0087 -0.011 0.015 0.00081 -0.016 1
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
FLAG_PHONE
FLAG_EMAIL
CNT_FAM_MEMBERS
REGION_RATING_CLIENT
AGE
WORK_EXP
SK_ID_PREV
AMT_ANNUITY_PREV
AMT_APPLICATION_PREV
AMT_CREDIT_PREV
AMT_GOODS_PRICE_PREV
DAYS_DECISION_PREV
SELLERPLACE_AREA_PREV
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

The top 10 correlations for Non-defaulters

Attribute1	Attribute 2	Correlation Value
AMT_APPLICATION_PREV	AMT_GOODS_PRICE_PREV	0.987533
AMT_CREDIT	AMT_GOODS_PRICE	0.986415
AMT_APPLICATION_PREV	AMT_CREDIT_PREV	0.975725
AMT_CREDIT_PREV	AMT_GOODS_PRICE_PREV	0.97165
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.944355
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878468
AMT_ANNUITY_PREV	AMT_GOODS_PRICE_PREV	0.816219
AMT_ANNUITY_PREV	AMT_CREDIT_PREV	0.811975
AMT_ANNUITY_PREV	AMT_APPLICATION_PREV	0.805867
AMT_ANNUITY	AMT_GOODS_PRICE	0.763433

FINAL RECOMMENDATIONS

Defaulters

The Clients below have greater chance of Defaulting.But it doesnot mean we cannot target them.Instead if we plan to target the below clients we have to do further detailed analysis on the credit amount range,Annuity amount limit ,Interest ,type of loan and other features and then further proceed with their loan.

- Males applied less for loans when compared Females. But they have more chance of defaulting
- Even though the age of 20-30 have applied for loan their percentage of defaulting is high when compared to other age group
- The clients with work experience less than 2 years or 2-4 years have higher chance of defaulting.
- The clients who has Cancelled and Refused from previous application status and those applied for Revolving loans have more chance to be defaulters.
- Clients having own realty have more chance of defaulting
- Client who own business and self employed tend to take loans more and also more the chance of defaulting.
- Clients with civil marriage and secondary/secondary special have chance of defaulting
- Client with more count of family members and children have greater chance of defaulting
- Client with low income have chance of defaulting

Non-Defaulters

We can target the client with following qualities as it increases our business and also reduce the chance of defaulting.

- Females have applied a lot for loans when compared to males their chance of defaulting is also less
- Target the people in age group 30-50.
- People who are having experience of 4-20 years have less chance of defaulting.
- Clients having Approved and Unused offer from previous application status and those applied for Cash/Consumer loan tend to be non defaulters.
- Clients having own car have less chance of defaulting
- Clients living in Rented apartments ,Co-op apartments or office apartments have less chance of defaulting
- Clients who are married and have higher education have a greater chance to take loans.
- Client who have more income and less credit value

THANK YOU