Credit EDA Case Study

ANALYSIS AND CONCLUSION

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.

Here, the aim is 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.

Information at Hand:
- Application Data: Information of the client at the time of application
- Previous Application Data: Information about the client's previous loan data

THE APPROACH..

Here, we have used EDA to find patterns in the given data set. Then, with the help of various visualizations we have studied how these patterns affect the target variables . Finally we have derived inferences with the help of the whole exercise.

Steps Followed :
- Cleaning of Application data
- Univariate and Bivariate analysis of Application data (Application Data.ipynb)
- Cleaning of Previous application data
- Univariate and Bivariate analysis of Previous Application data (Previous Application.ipynb)
- Merger of Application data and Previous Application data and analysis of merged data. (Previous Application.ipynb)
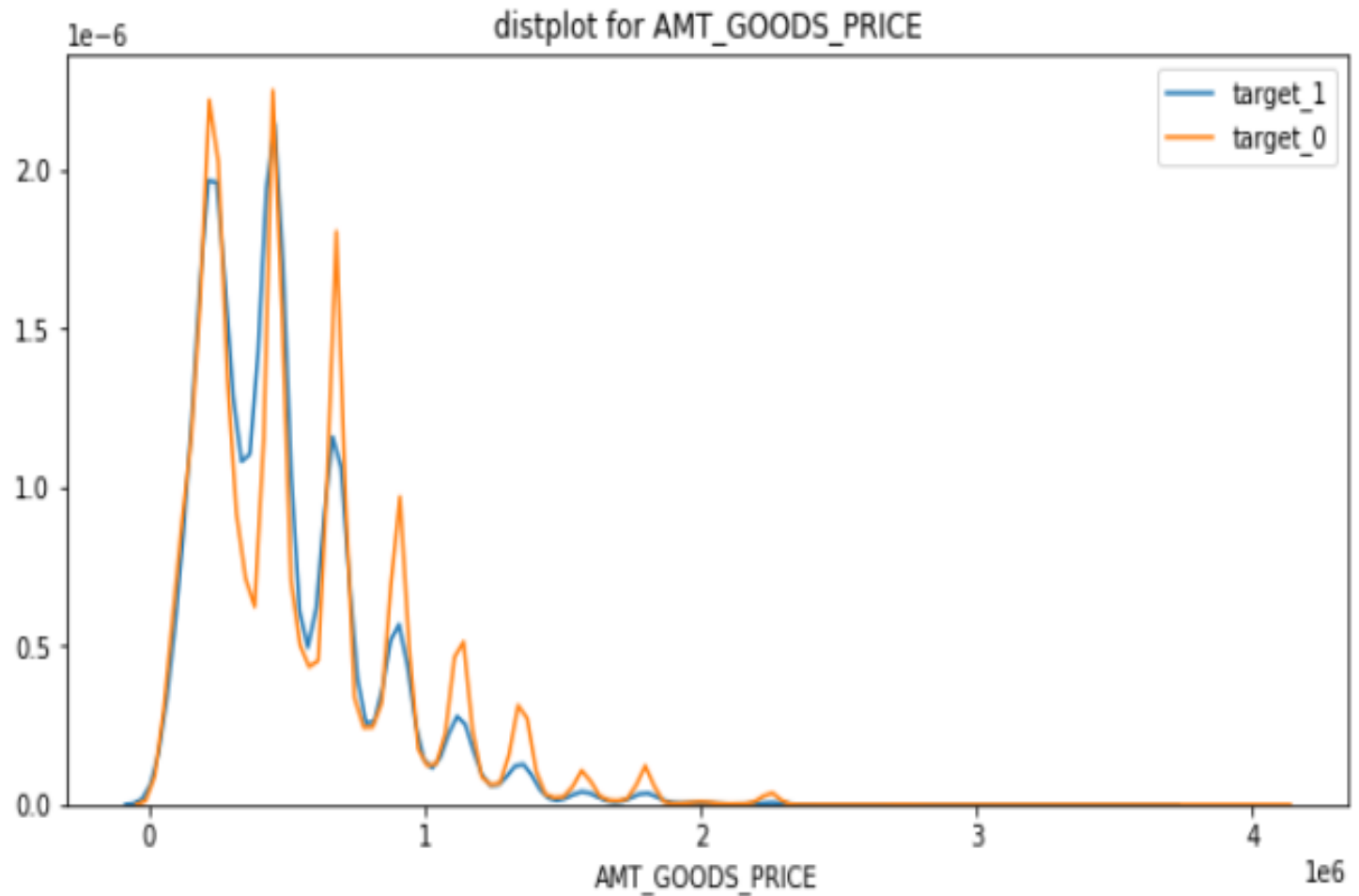- Conclusion

Application Data

UNIVARIATE ANALYSIS

CREDIT AMOUNT

- We can observe the distribution for AMT_CREDIT for target 0 and target 1

- Although there is a slight edge for target 0 over target 1, we cannot conclusively say that payment difficulties or on-time payment is correlated with AMT_CREDIT
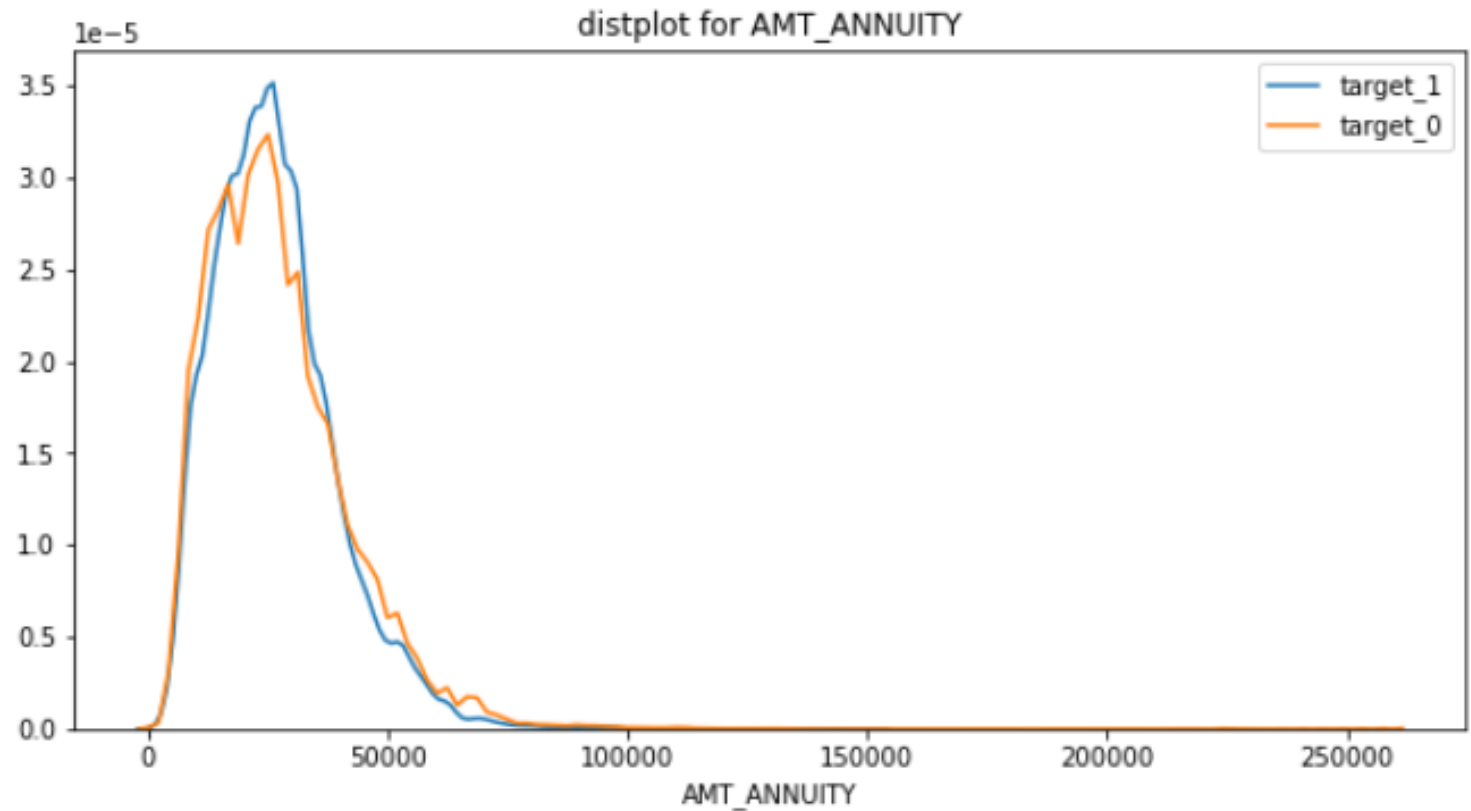

distplot for AMT_CREDIT
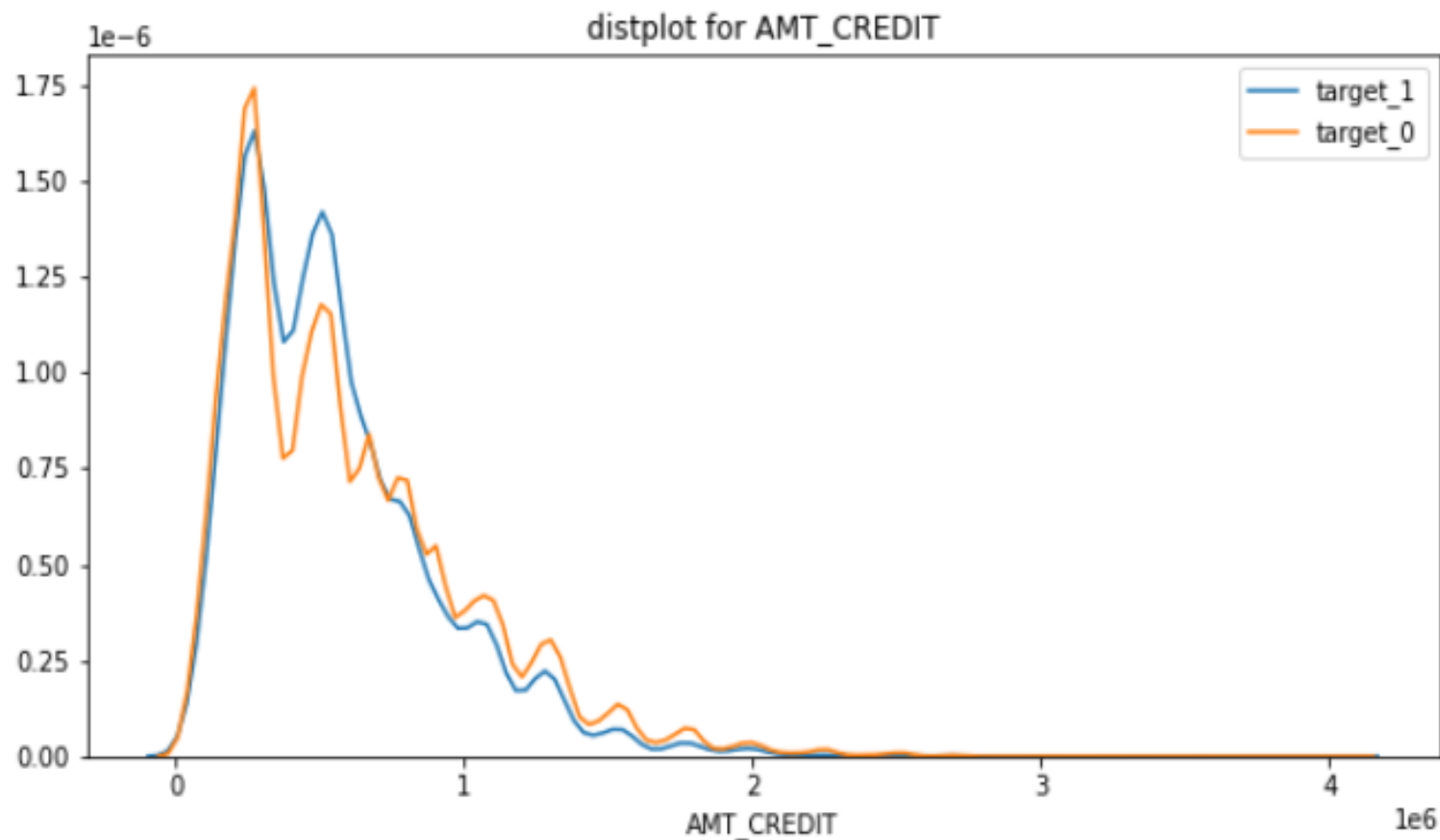
GOODS PRICE

- We can observe the distribution for AMT_GOODS_PRICE for target 0 and target 1

- Although there is an edge for target 0 over target 1 for AMT_GOODS_PRICE

distplot for AMT_GOODS_PRICE

ANNUITY AMOUNT

- We can observe the distribution for AMT_ANNUITY for target 0 and target 1

- Although there is a slight edge for target 0 over target 1, we cannot conclusively say that payment difficulties or on-time payment is correlated with AMT_ANNUITY



distplot for AMT_ANNUITY

AMOUNT INCOME TOTAL RANGE

- We can observe the distribution for AMT_ANNUITY for target 0 and target 1

- Although there is a slight edge for target 0 over target 1, we cannot conclusively say that payment difficulties or on-time payment is correlated with AMT_ANNUITY
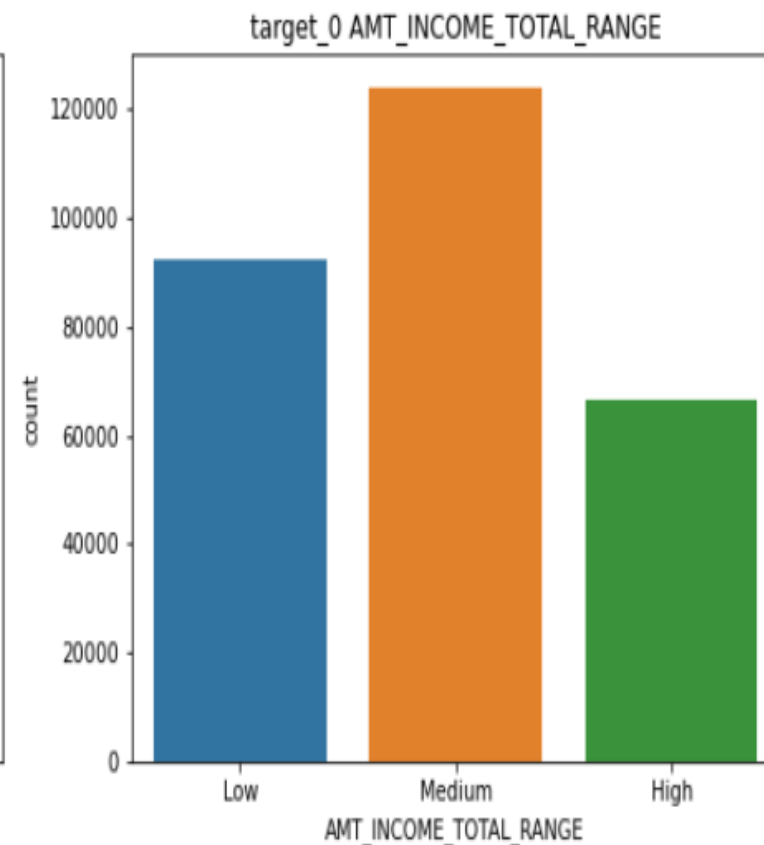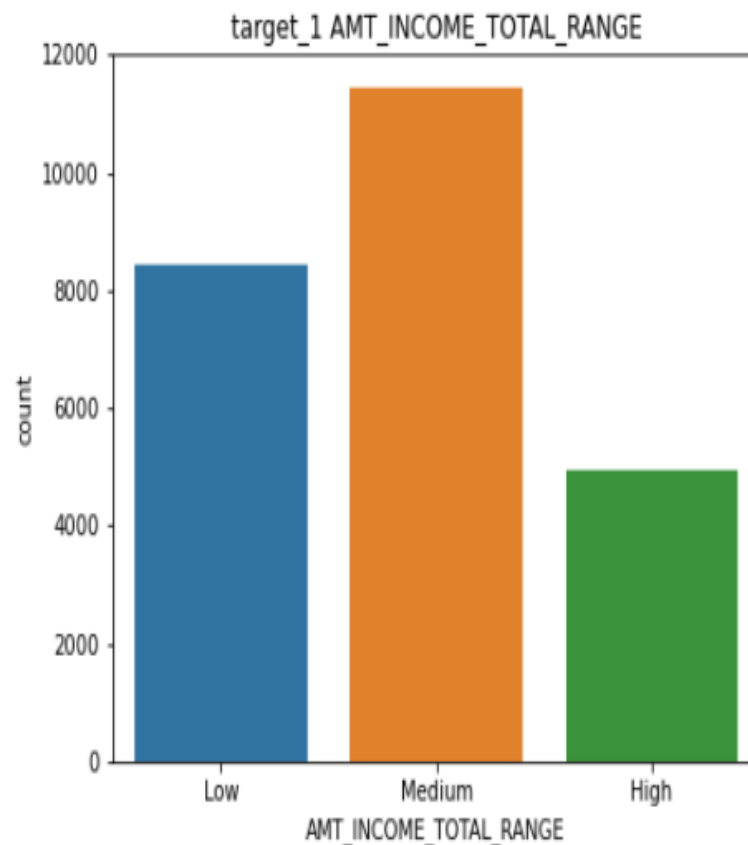
GENDER RATIO: Target 1 vs. Target 0

- From the pie-plot, we can observe that female has a better score of paying the loans on-time when compared to Men While 66.6% Female are able to pay on time, the men's ratio is 33.4% for the same
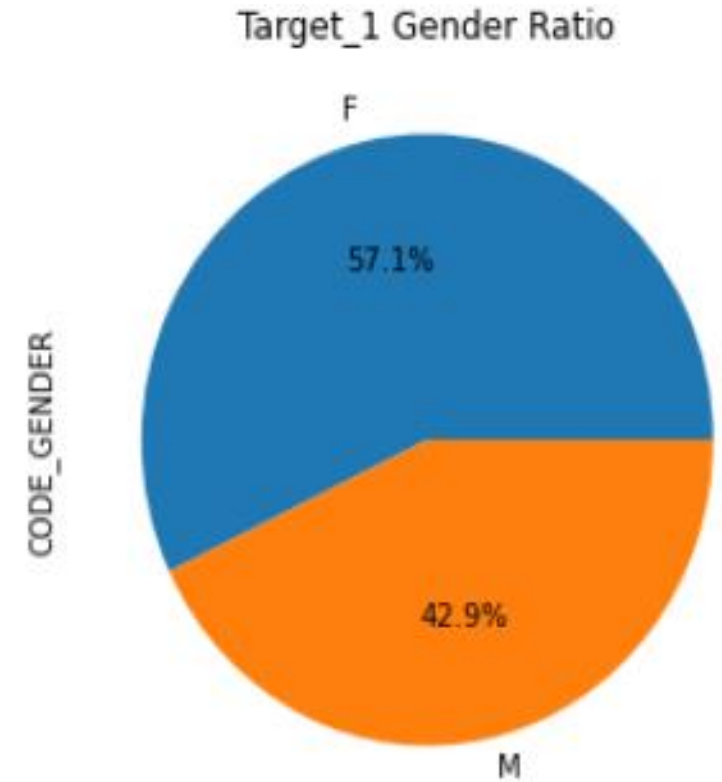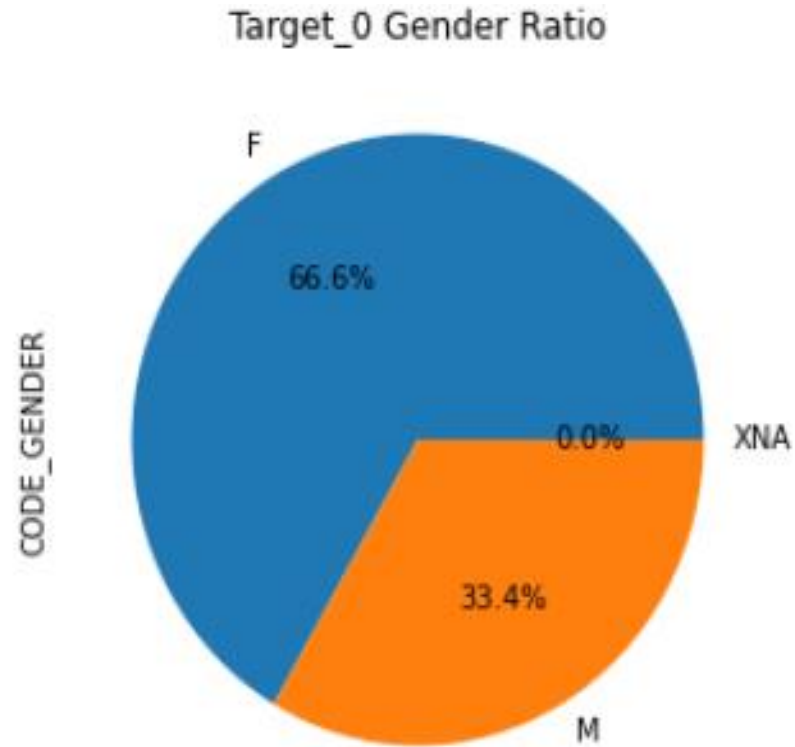
CONTRACT TYPE

The percentage of Cash Loans and Revolving Loans are almost same for both target 1 as well as target 0. Hence, this might not be the correct column for taking a decisive action
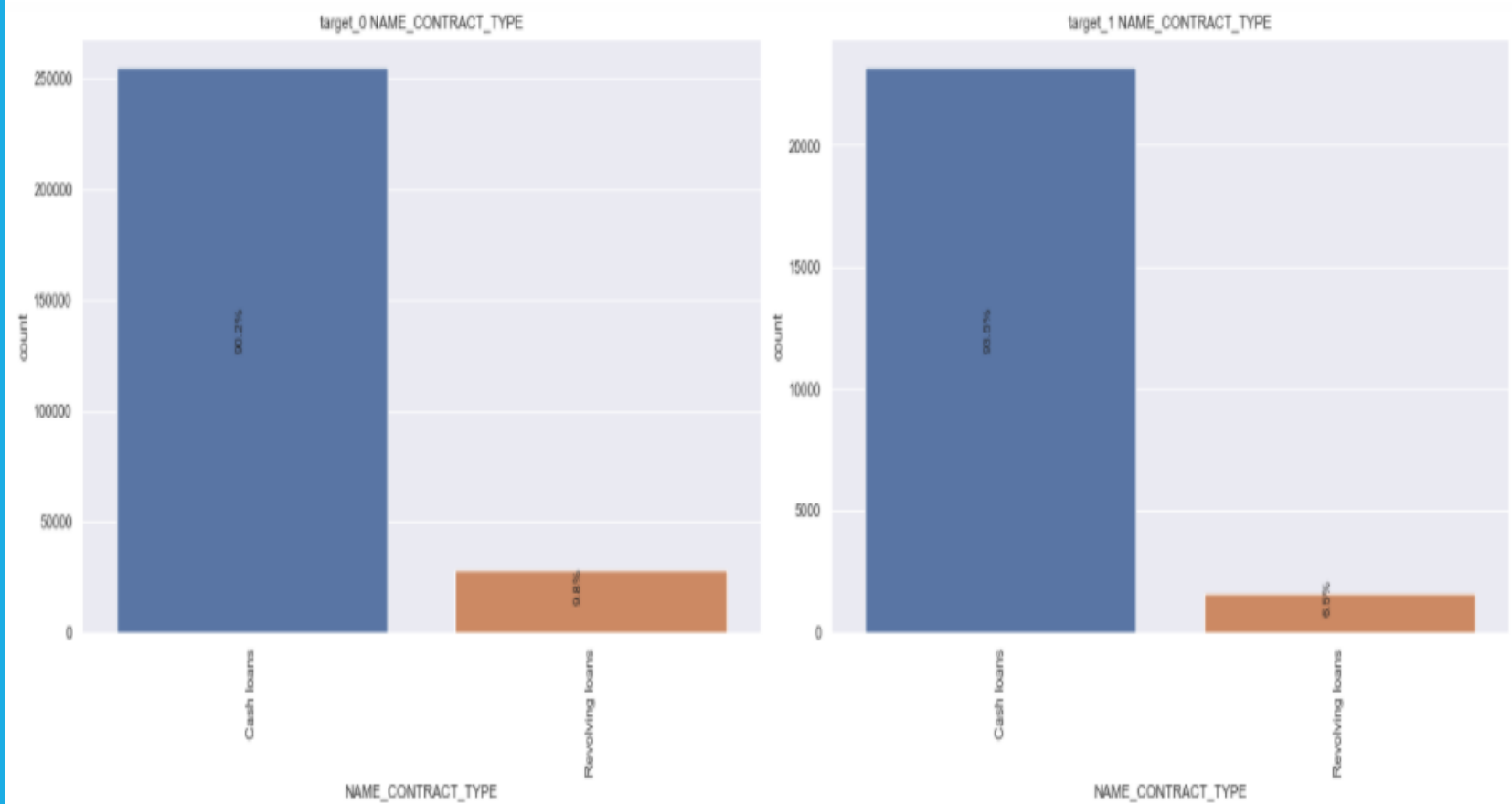
HEAT MAP – Target 0

A high correlation can be observed between the columns AMT_ANNUITY, AMT_GOODS_PRICE and AMT_CREDIT Top 10 Correlation between variables are shown in the next slide
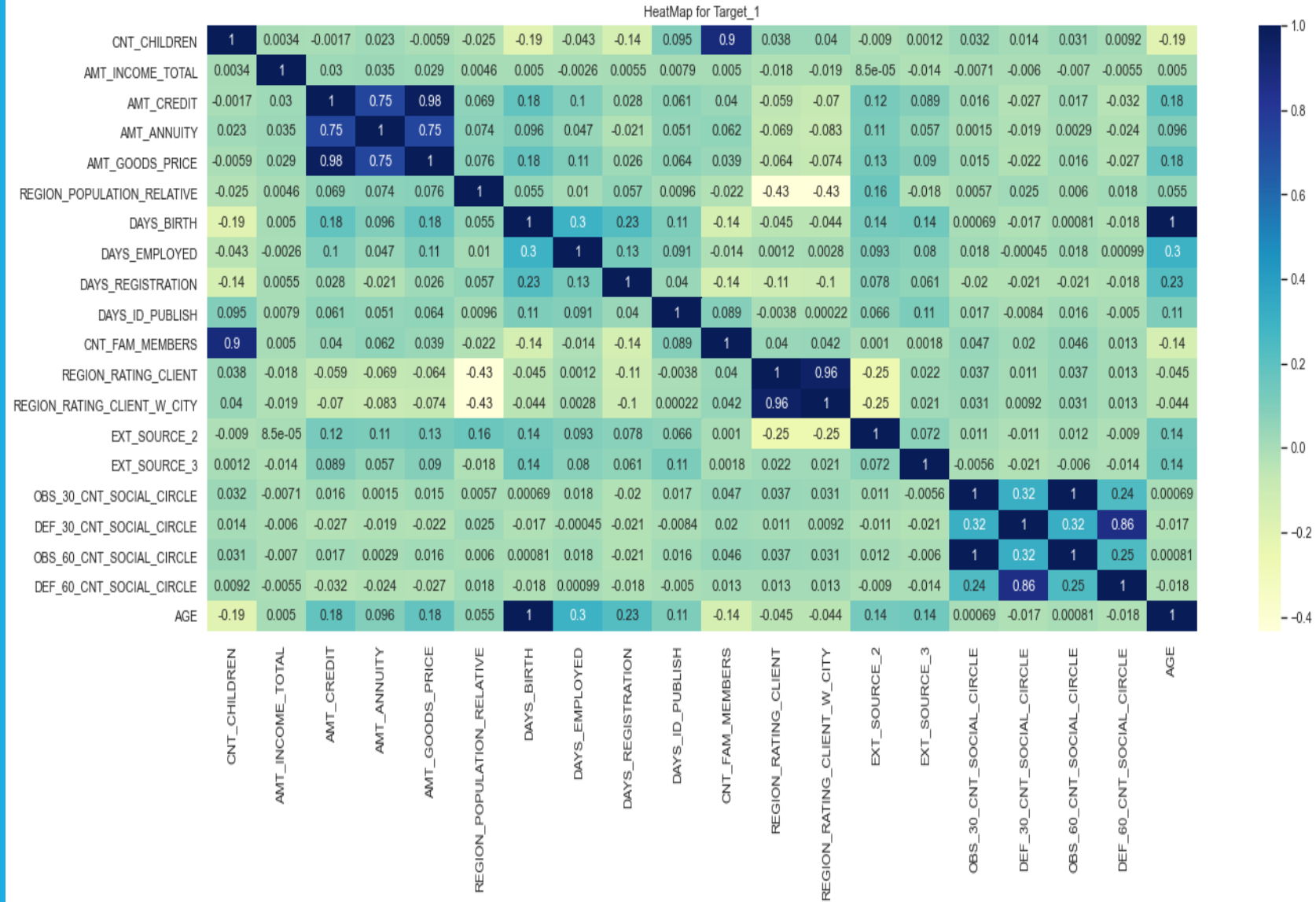

Heatmap for target_0
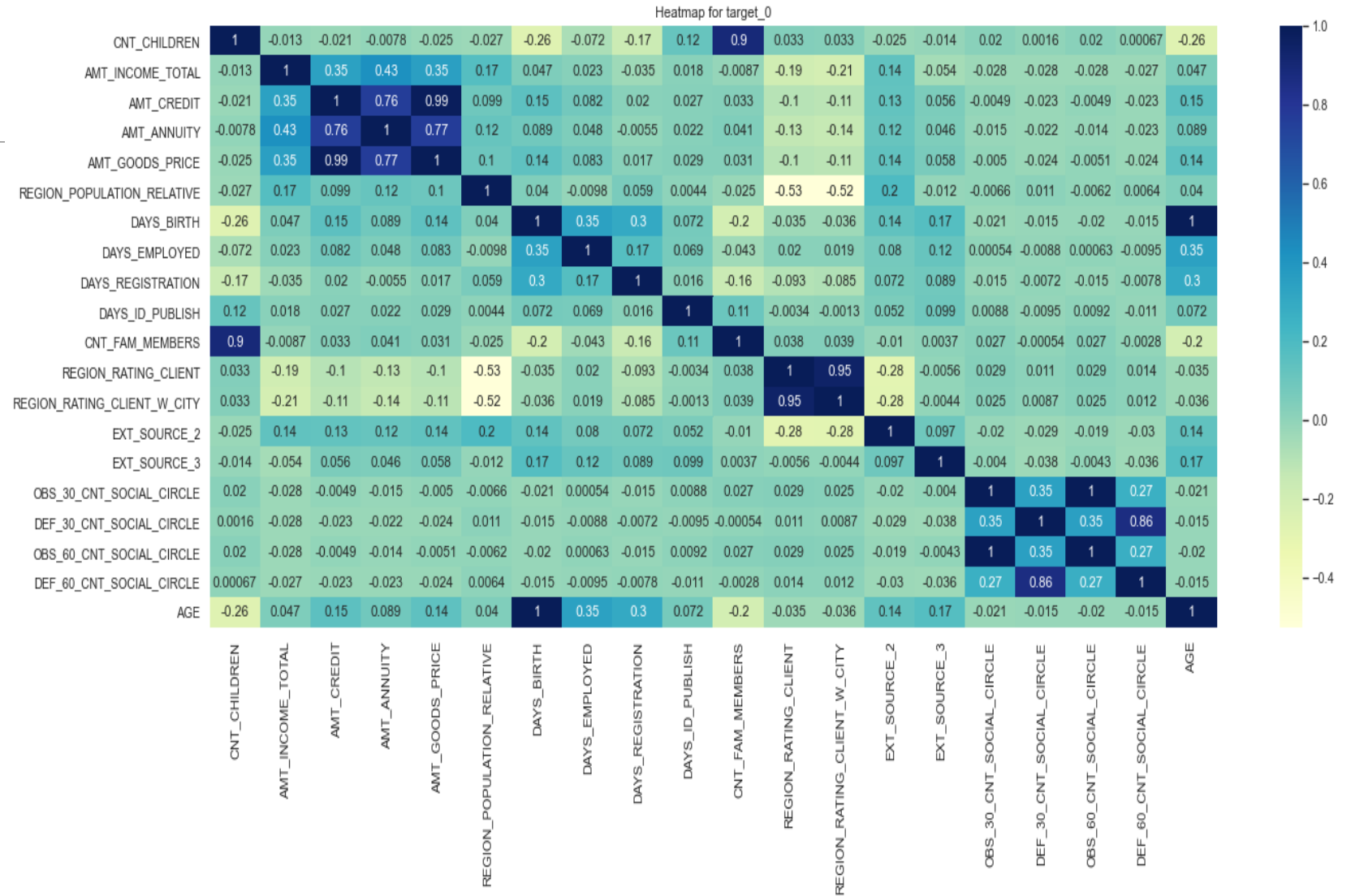
TOP 10 CORRELATION – Target 0

- The top correlation is between column OBS_60_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE

- AMT_CREDIT and AMT_GOODS_PRICE has the second highest correlation value

VAR1	VAR2	Correlation	Abs_Correlation
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998562	0.998562
AMT_GOODS_PRICE	AMT_CREDIT	0.986643	0.986643
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.949577	0.949577
CNT_FAM_MEMBERS	CNT_CHILDREN	0.896352	0.896352
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.864013	0.864013
AMT_GOODS_PRICE	AMT_ANNUITY	0.767564	0.767564
AMT_ANNUITY	AMT_CREDIT	0.763467	0.763467
REGION_RATING_CLIENT	REGION_POPULATION_RELATIVE	-0.527448	0.527448
REGION_RATING_CLIENT_W_CITY	REGION_POPULATION_RELATIVE	-0.523624	0.523624
AMT_ANNUITY	AMT_INCOME_TOTAL	0.425839	0.425839

HEAT MAP – Target 1

A high correlation can be observed between the columns AMT_ANNUITY, AMT_GOODS_PRICE and AMT_CREDIT
Top 10 Correlation between variables are shown in the next slide

HeatMap for Target_1

TOP 10 CORRELATION – Target 1

- The top correlation is between column OBS_60_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE

- AMT_CREDIT and AMT_GOODS_PRICE has the second highest correlation value

VAR1	VAR2	Correlation	Abs_Correlation
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998562	0.998562
AMT_GOODS_PRICE	AMT_CREDIT	0.986643	0.986643
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.949577	0.949577
CNT_FAM_MEMBERS	CNT_CHILDREN	0.896352	0.896352
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.864013	0.864013
AMT_GOODS_PRICE	AMT_ANNUITY	0.767564	0.767564
AMT_ANNUITY	AMT_CREDIT	0.763467	0.763467
REGION_RATING_CLIENT	REGION_POPULATION_RELATIVE	-0.527448	0.527448
REGION_RATING_CLIENT_W_CITY	REGION_POPULATION_RELATIVE	-0.523624	0.523624
AMT_ANNUITY	AMT_INCOME_TOTAL	0.425839	0.425839

Application Data

BIVARIATE ANALYSIS

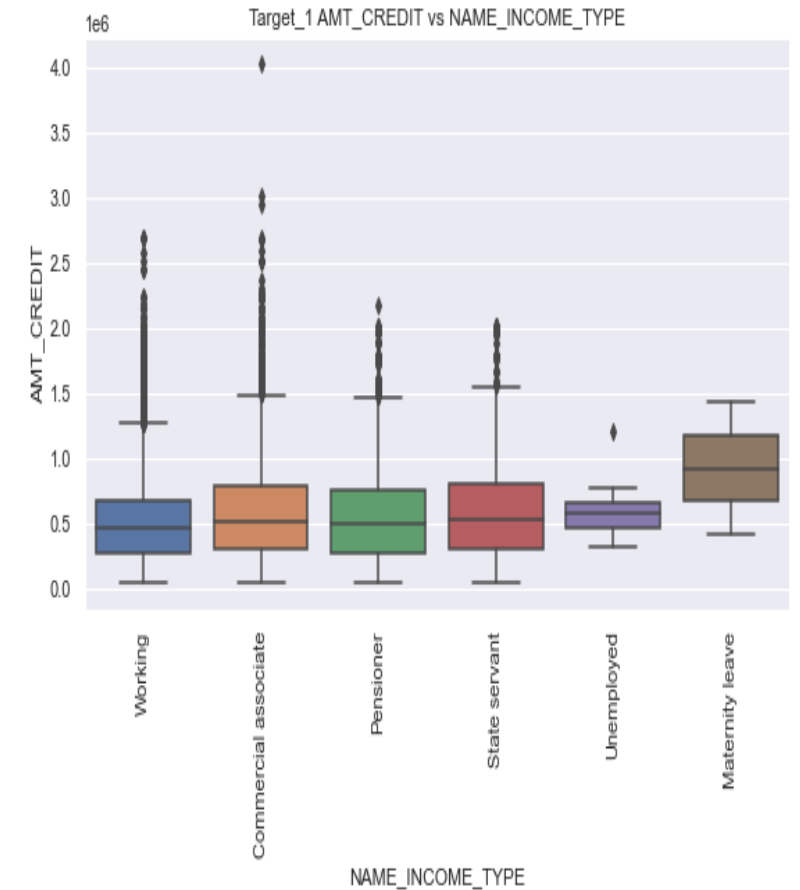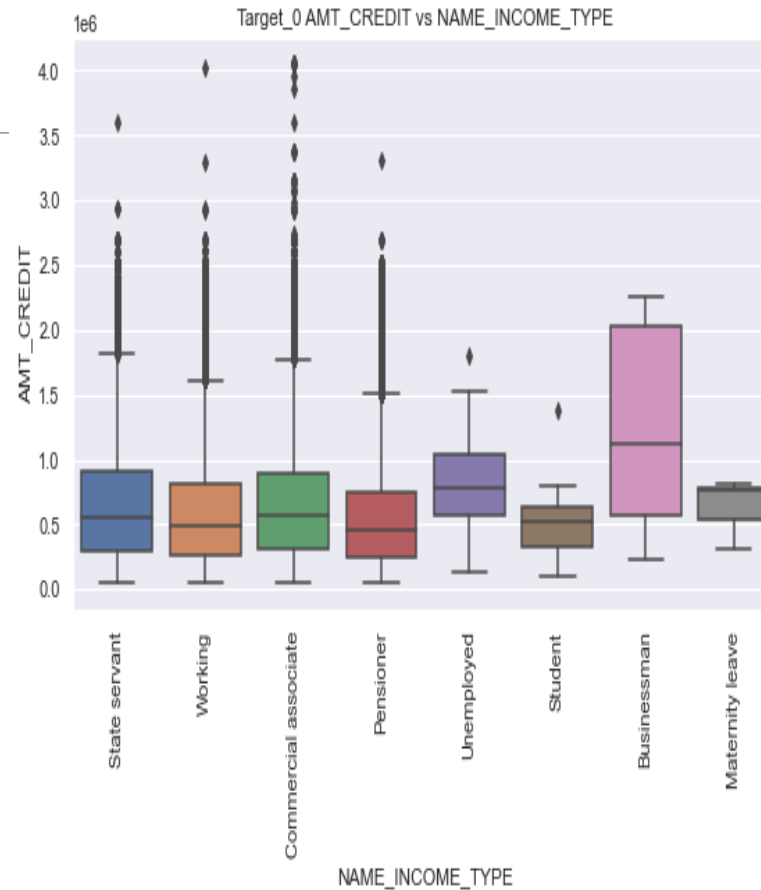# CREDIT AMOUNT vs. EDUCATION TYPE

The median for Loan Defaulters for Academic degree with higher credit is more and hence, we should recommend on approving lesser loan amount for Academic degree holders
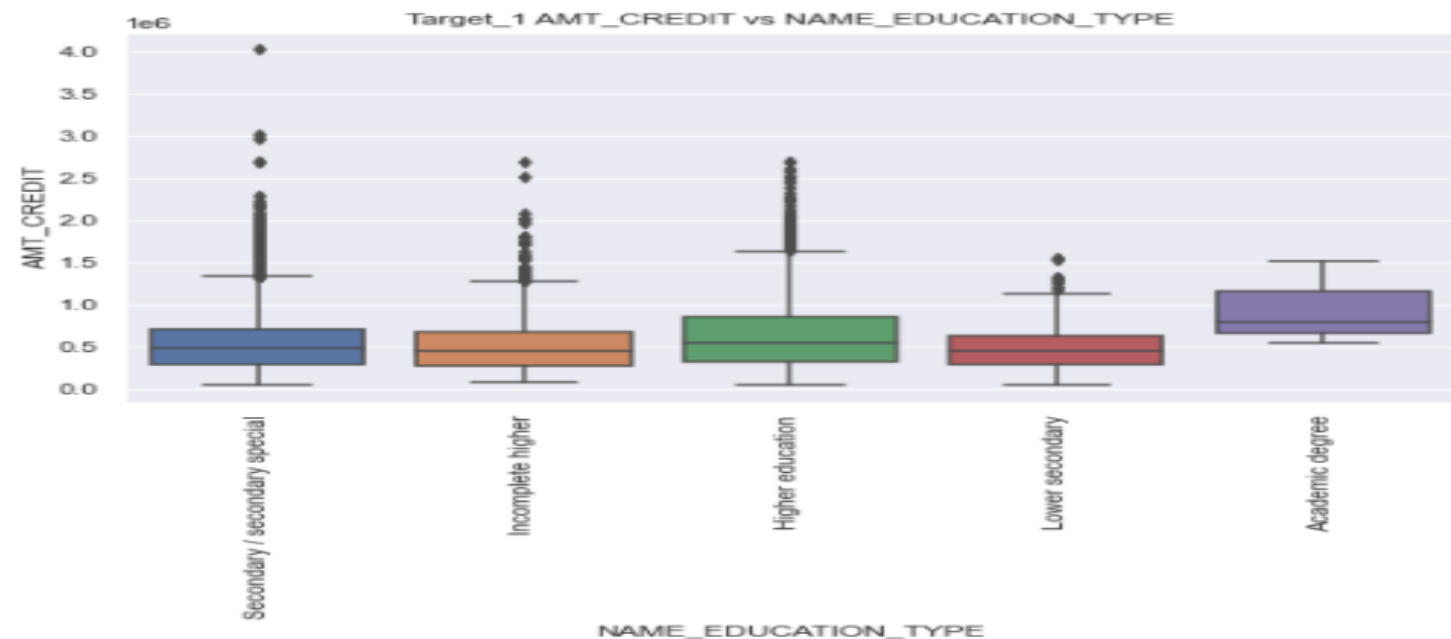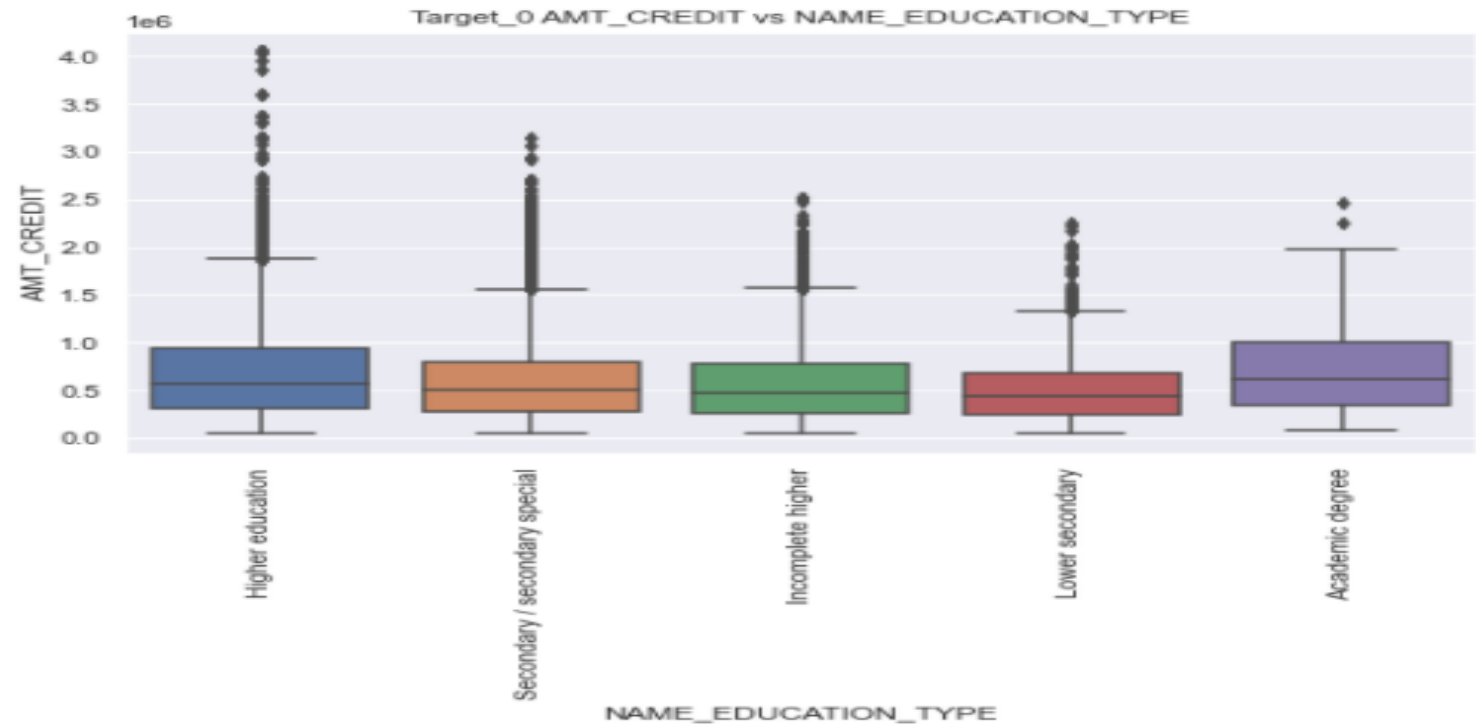
CREDIT AMOUNT vs. INCOME TYPE

Student and Businessman are not in the late payment difficulties and hence these class should be focused on for more loans
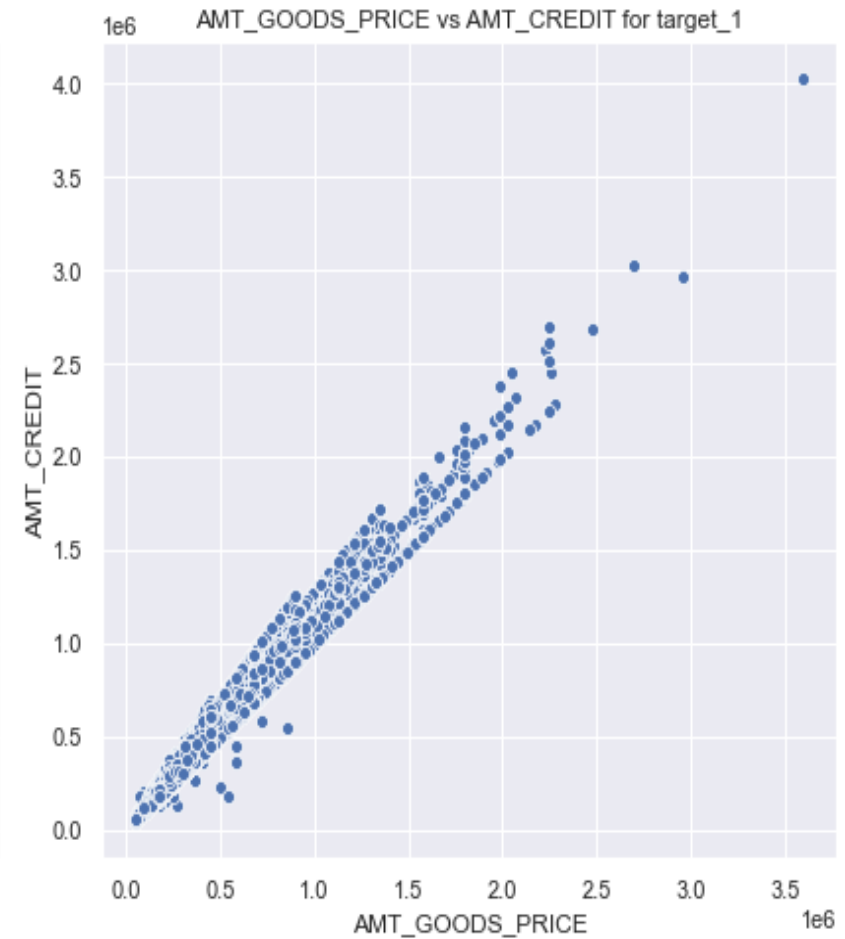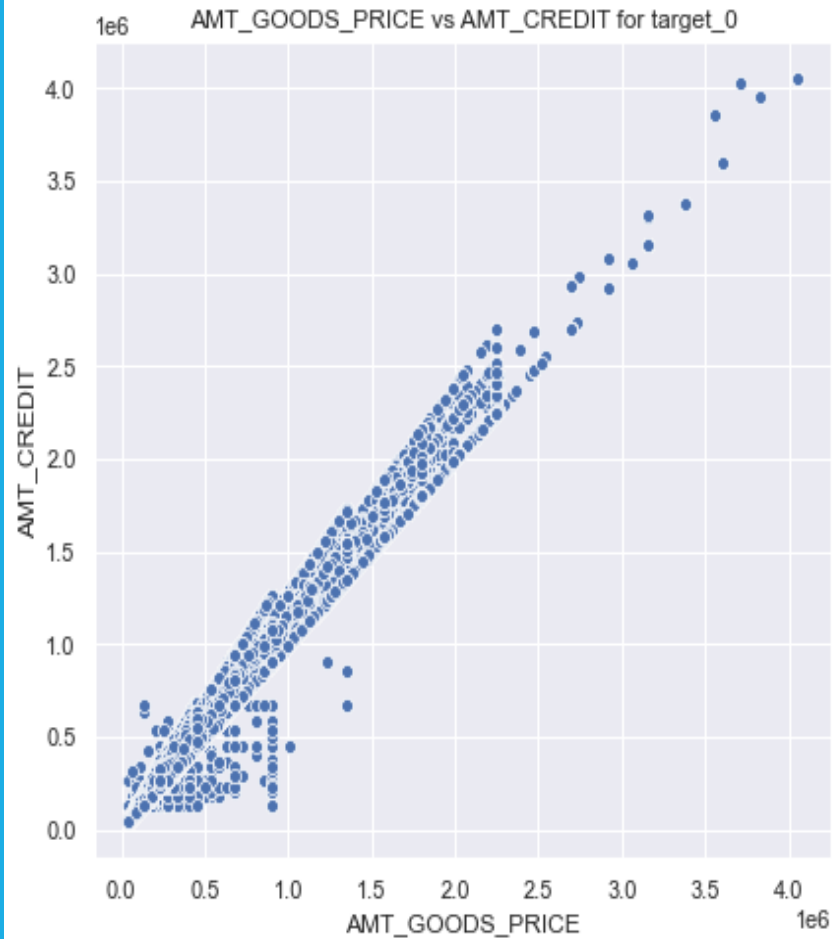
Average Loan amount of Businessman is higher than all other income type

GOODS PRICE vs. CREDIT AMOUNT

AMT_GOOD_PRICE in below 1000000 and AMT_CREDIT below 500000 is more likely to be paid on time
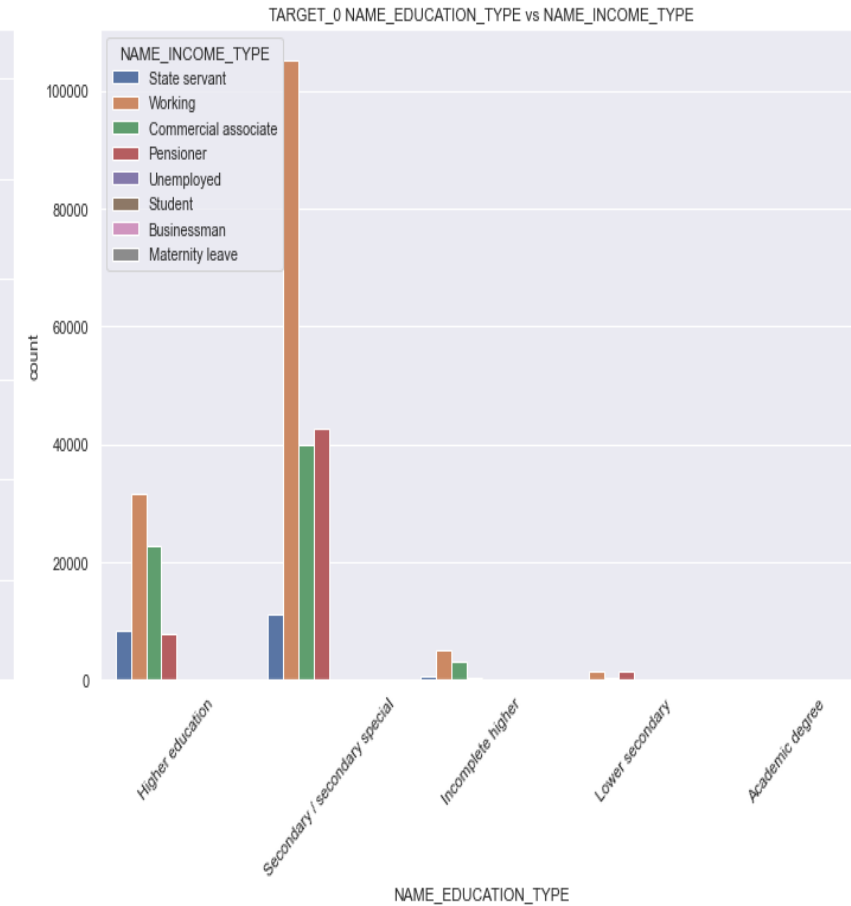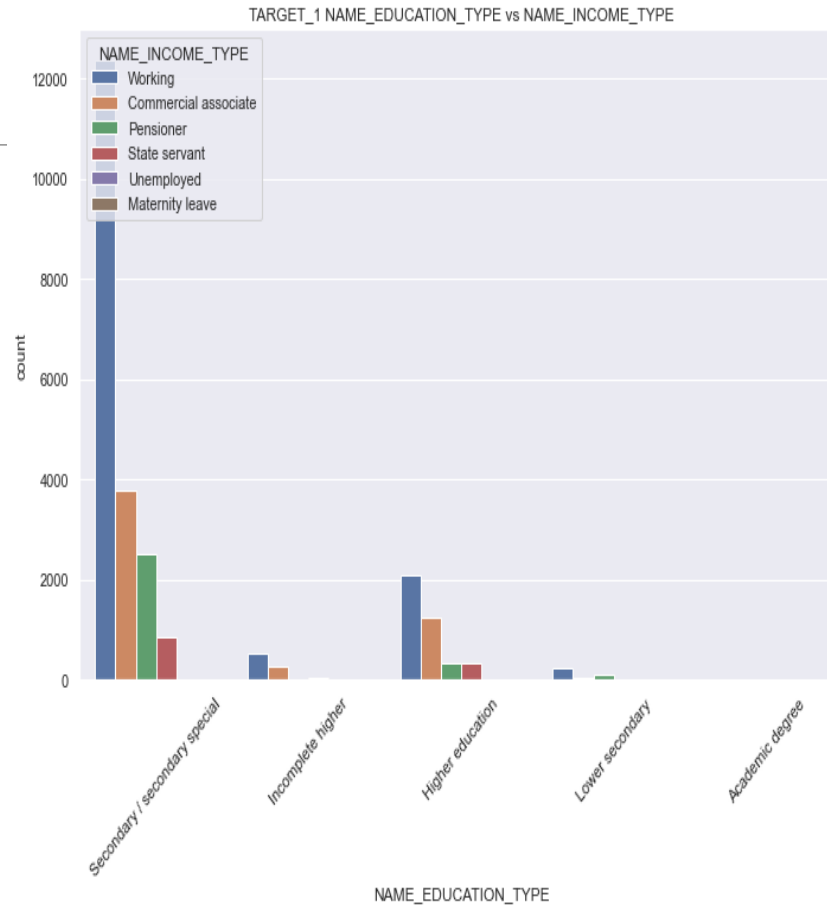
There is a high correlation between AMT_GOOD_PRICE and AMT_CREDIT

EDUCATION TYPE vs. INCOME TYPE

Student in Lower secondary Education takes less loans but they make the payments on-time

State servant with Higher education type are likely to pay more loans on-time than with Secondary education type

# Previous Application Data

UNIVARIATE ANALYSIS

DISTRIBUTION OF APPLICATION AMOUNT

- Lower loan amounts have been cancelled by client and higher loan amounts are cancelled by company.

- For the amount range of 2,500,000 to 4,000,000, number of loans canceled is greater or almost equal to number of loans approved.

- For the amount greater than 3,500,000, loans are mostly not approved.
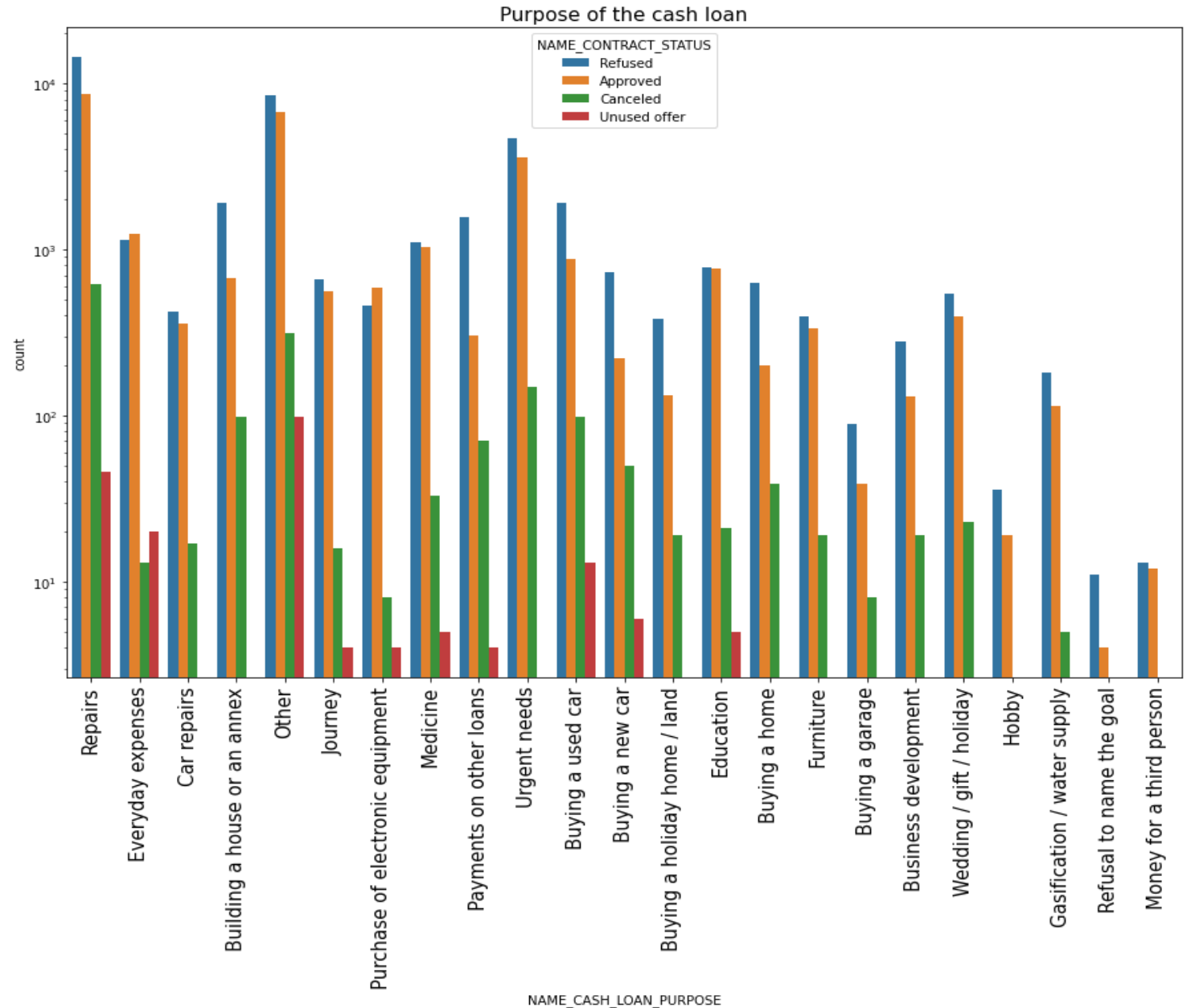

Distribution of Application Amount
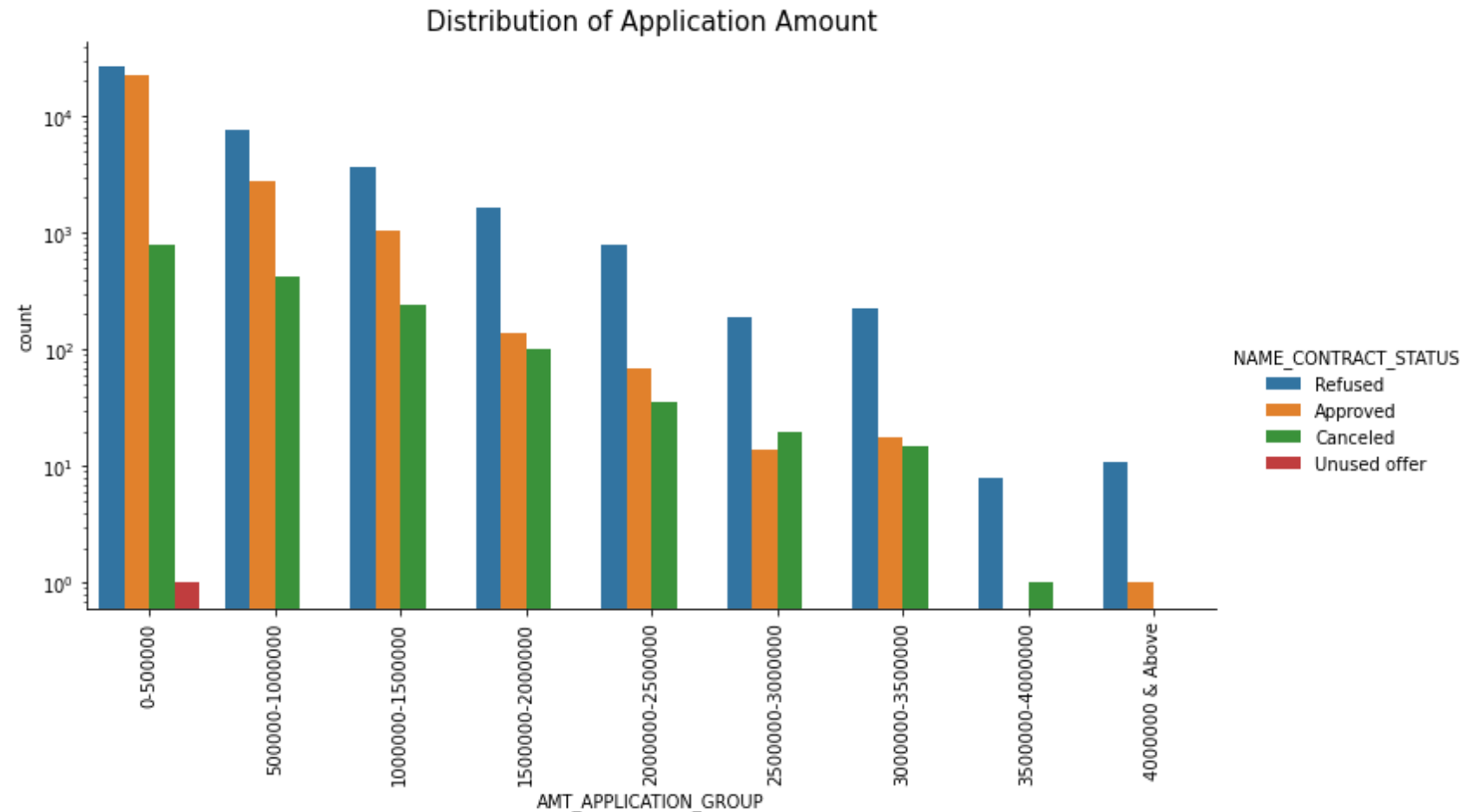
DISTRIBUTION OF ANNUITY AMOUNT

- Maximum number of loans are for low Annuity amount.

- Low annuity loans also have high number of Refused, Approved, and Canceled applications.


Distribution of Annuity Amount
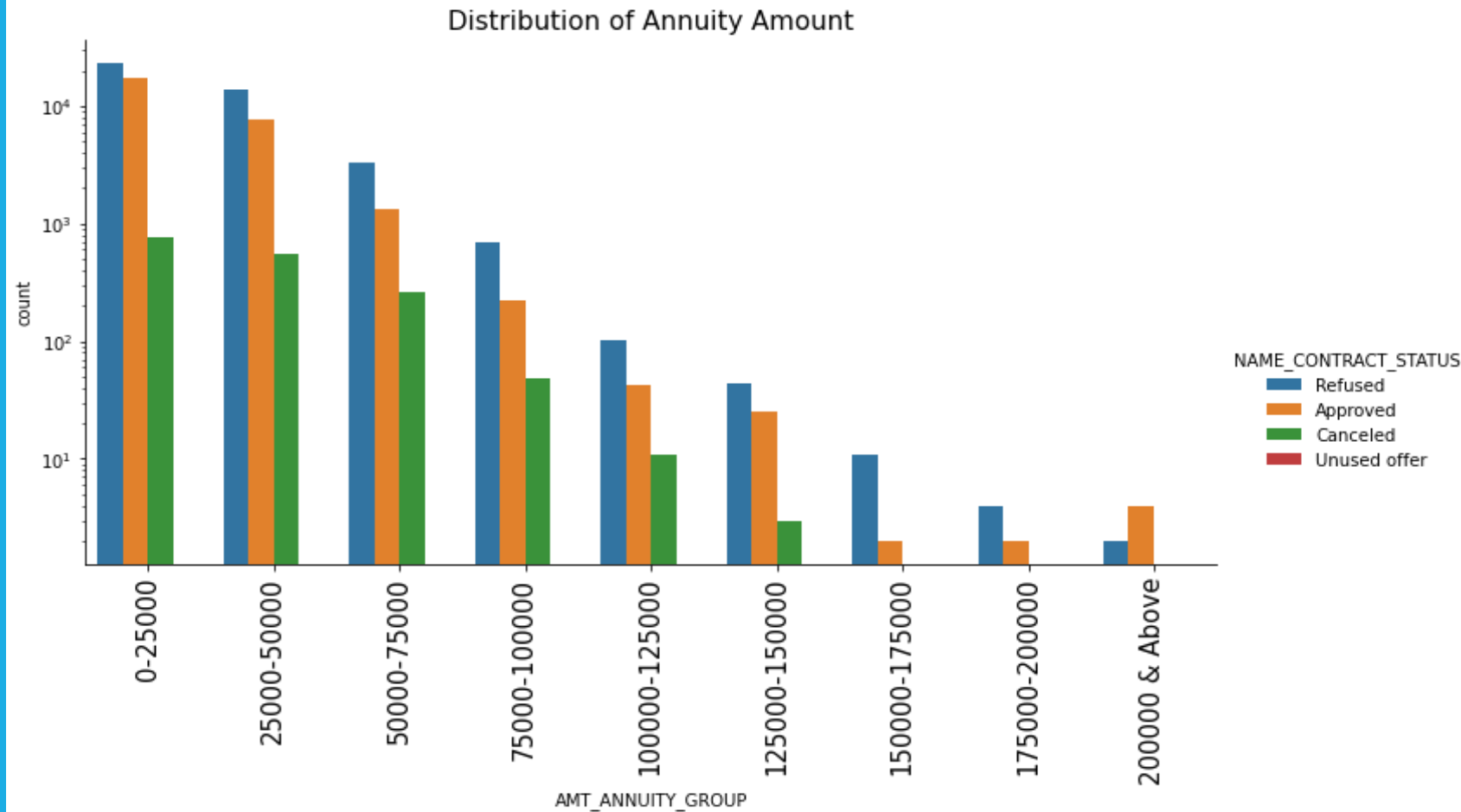
PURPOSE OF CASH LOAN

- Highest number of loans 'Approved', 'Refused' and 'Canceled' are for 'Repairs', 'Other' and 'Urgent Needs'

- For 'Everyday Expenses',' Purchase of Electronic Equipment' and 'Education', loans approved are more or almost equal to loans refused.

- In all the categories maximum clients cancel there application in approval phase only. This could be due to change in decision or worse pricing.

- Paying other loans and buying a new car and buying a home are having significantly higher rejections than approvals.

Purpose of the cash loan

CLIENT DISTRIBUTION TYPE

- Most of the new applicants get the loan approved.

- Repeaters have the highest number of loan refusals.
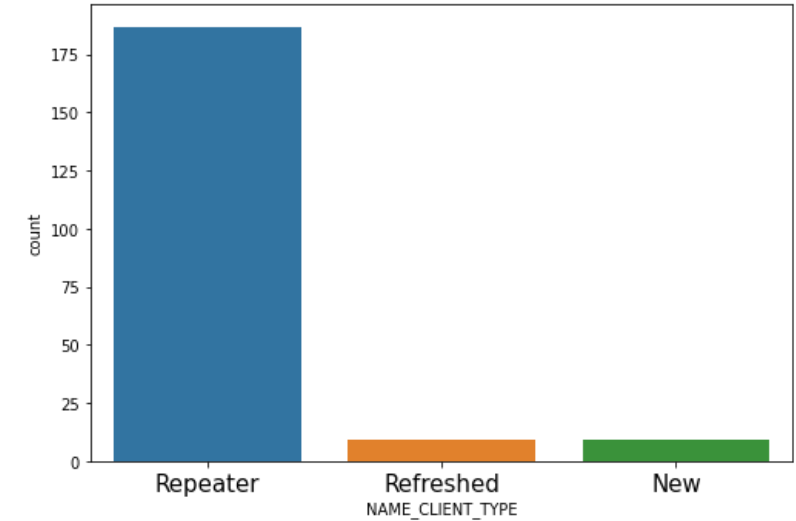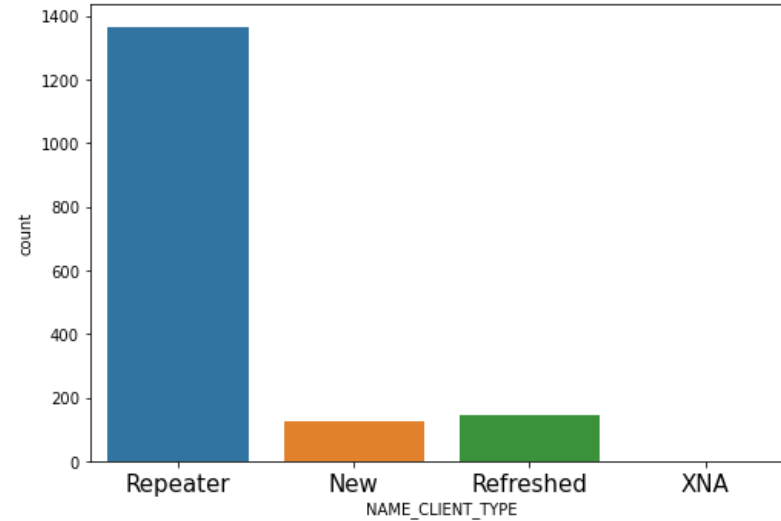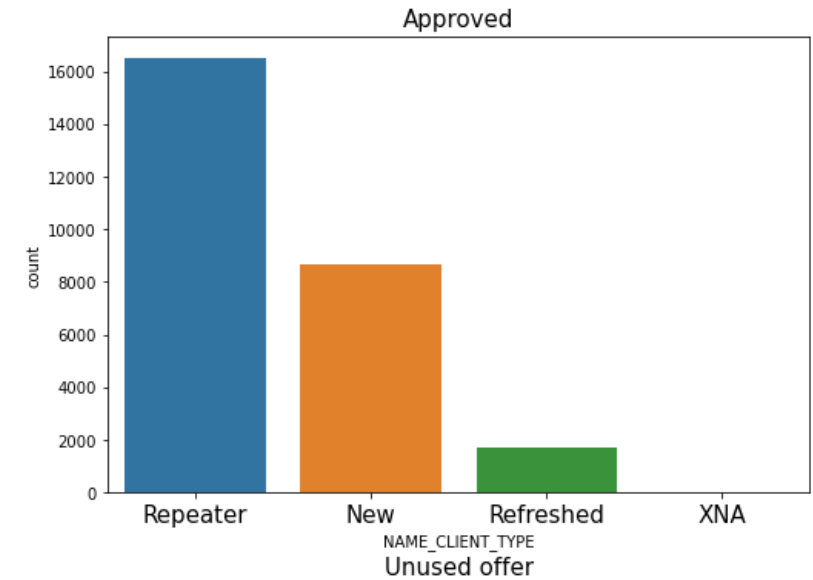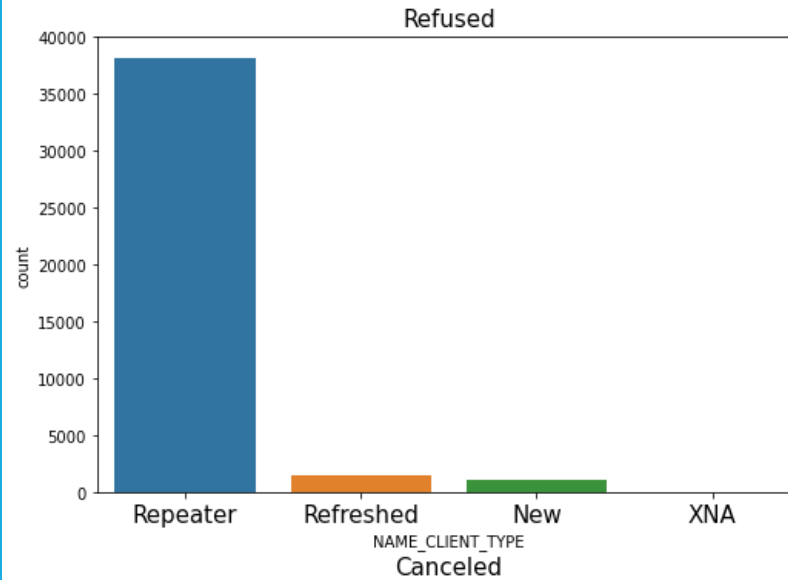
# Previous Application Data

BIVARIATE ANALYSIS

DISTRIBUTION OF PAYEMENT TERM IN MONTHS

- For shorter payment terms (up to 20 months), Approved loans are more than that of refuse ones.

- Number of loans canceled by client during approval phase increases with increasing tenure.

- Highest number of loans approved are for the tenure 10-20 months and same for loan refused is for 10-20 and 50-60 months.
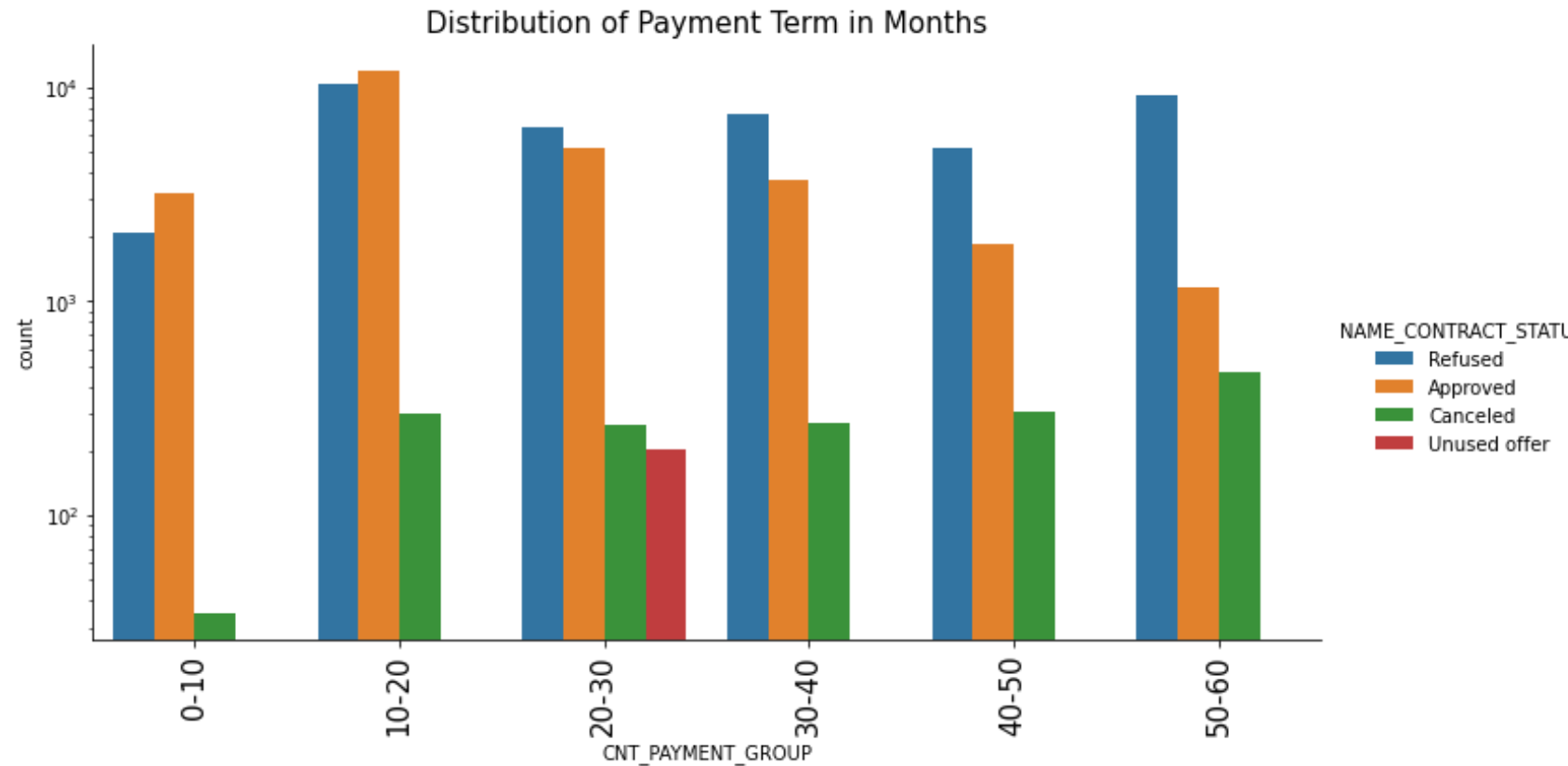
GROUPED INTEREST RATE

- Maximum number of loans approved are of high interest rate

- Maximum number of loans refused are of Low_Normal interest rate.

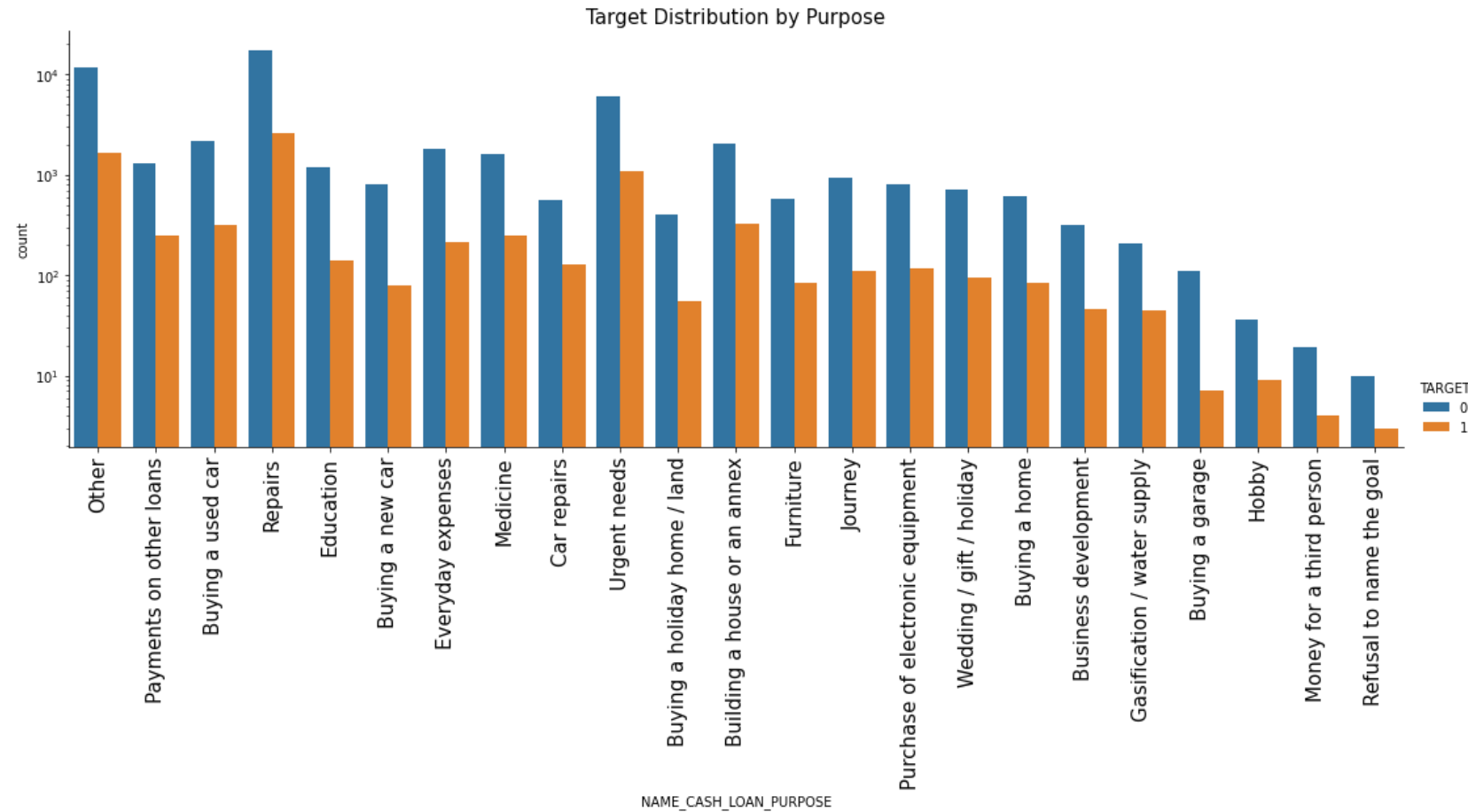- As the interest rate Increase, number of approved loans increase.
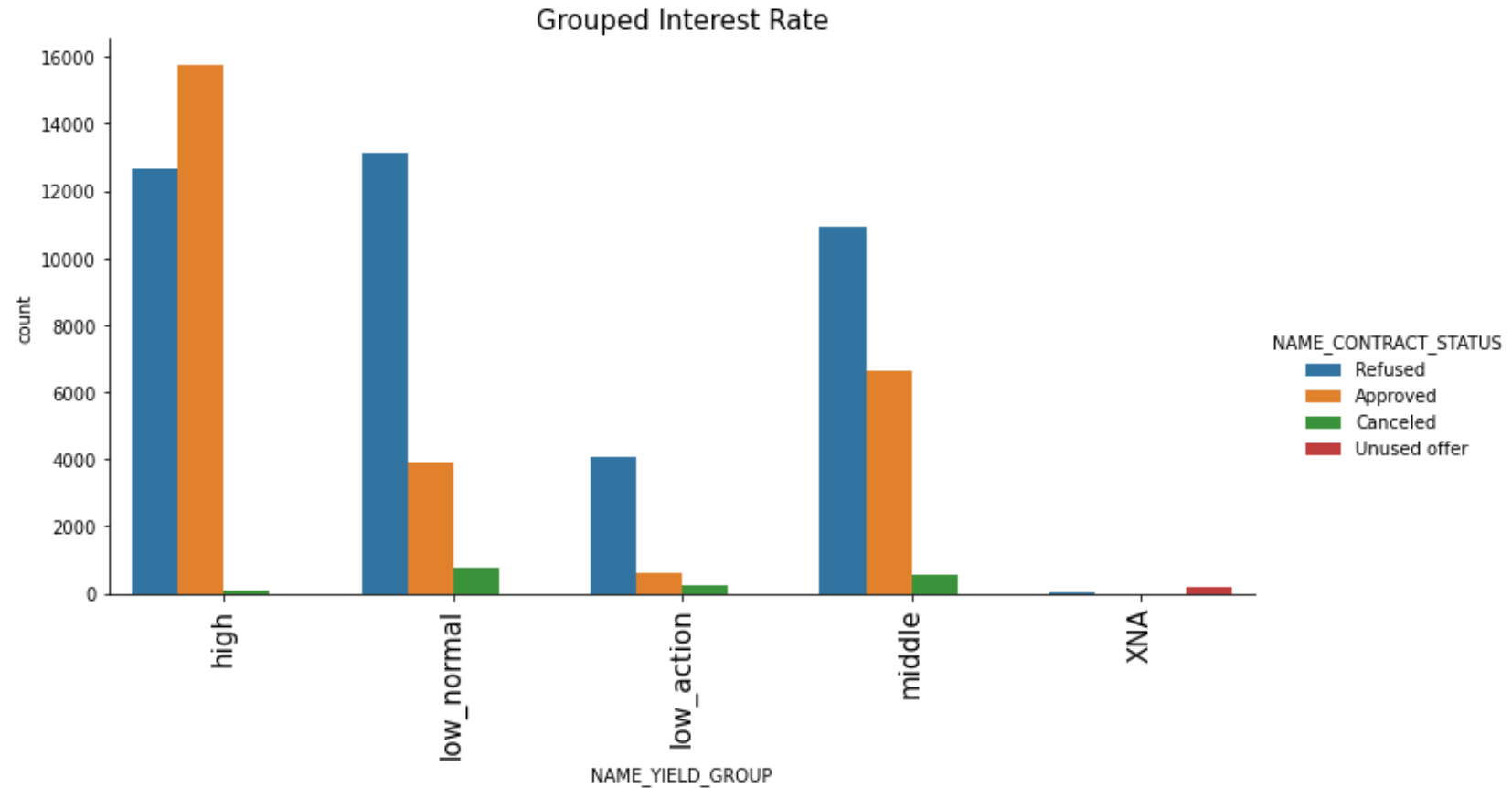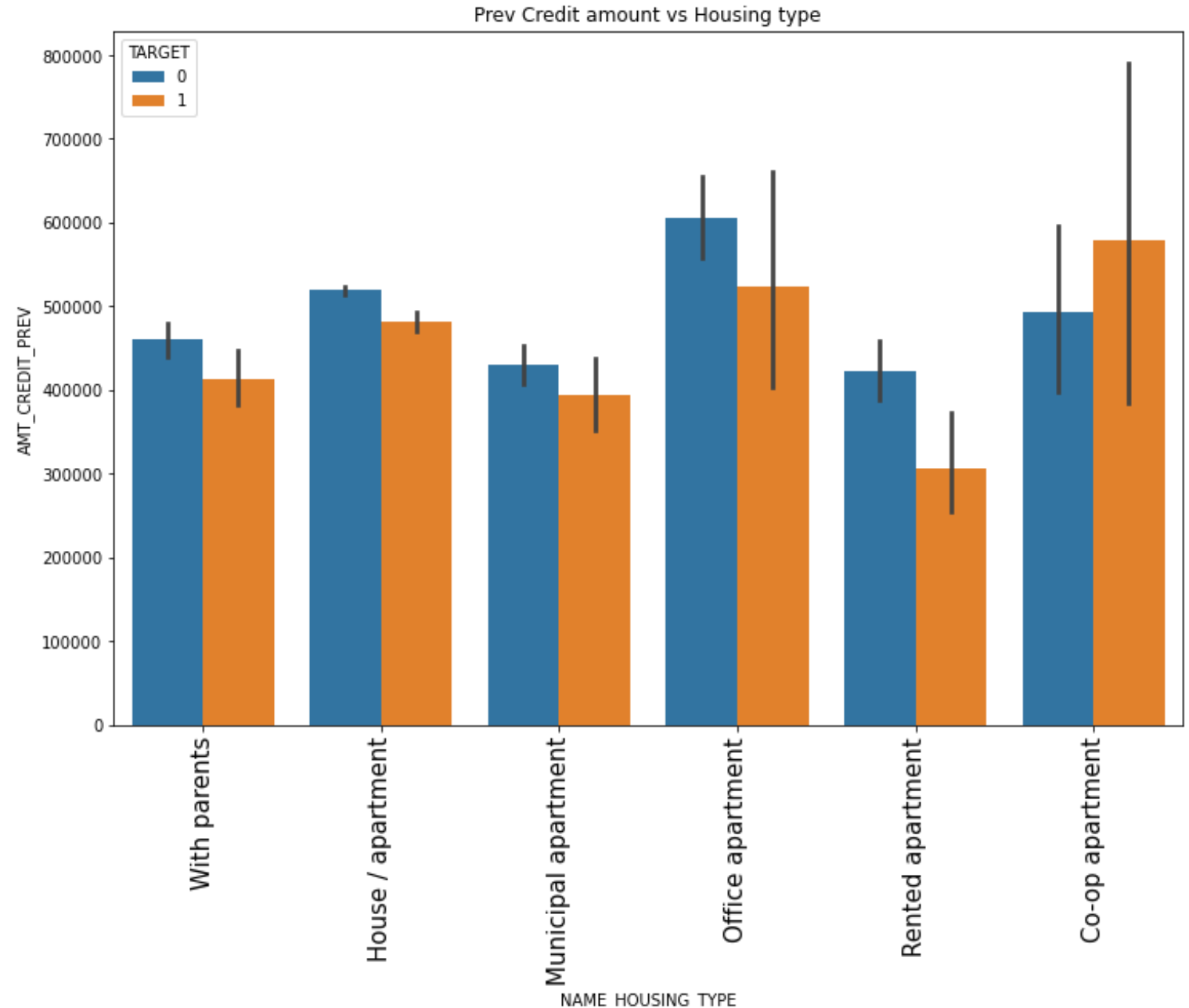
Merged Data

ANALYSIS

TARGET DISTRIBUTION BY PURPOSE

- Loan purposes with 'Repairs' are facing more difficulties in paying on time.

- Minimum payment difficulties are observed in Loan Purposes with 'Hobby', 'Buying a garage', 'Money for a third person' and ' Buying a new car'

Target Distribution by Purpose

PREV. CREDIT AMOUNT vs. HOUSING TYPE

- For Housing type, Office apartment is having higher credit of target 0 and Co-op apartment is having higher credit of target1. Hence, bank should avoid giving loans to clients having hosing type 'Co-op apartment 'and should focus on clients having housing type 'Office apartment'.

Conclusions

- Businessmen seem to take higher loan amount and make regular payment, however the number of loans taken by businessman are less and hence Bank should focus on providing more loans to Businessmen.

- Students seem to make successful On-Time payments and hence the focus should be on giving more loans to Students

- State servant with Higher education type are likely to pay more loans on-time than with Secondary education type

- Loan Defaulters for Academic degree with higher credit are more and hence, we should recommend on approving lesser loan amount for Academic degree holders

- Bank should focus less on loan purpose 'Repair' as they are having maximum number of delayed payments.

- Bank should avoid giving loans to clients having housing type 'Co-op apartment 'and should focus on clients having housing type 'Office apartment

- Bank should accommodate the loan terms under 20 months to increase customer base and avoid cancellation from client side.

- New applicants seem to get more loan approved than repeaters also female clients have better score of paying the loans on-time when compared to Men.

- In all the categories maximum clients cancel there application in approval phase only. This could be due to change in decision or worse pricing. Bank can look into this to increase the client base.