



EDA CREDIT ASSIGNMENT

SUBMITTED BY:-

YUVRAZ DHUNGANA

BATCH(DSC-46)

Reason for analysis:-

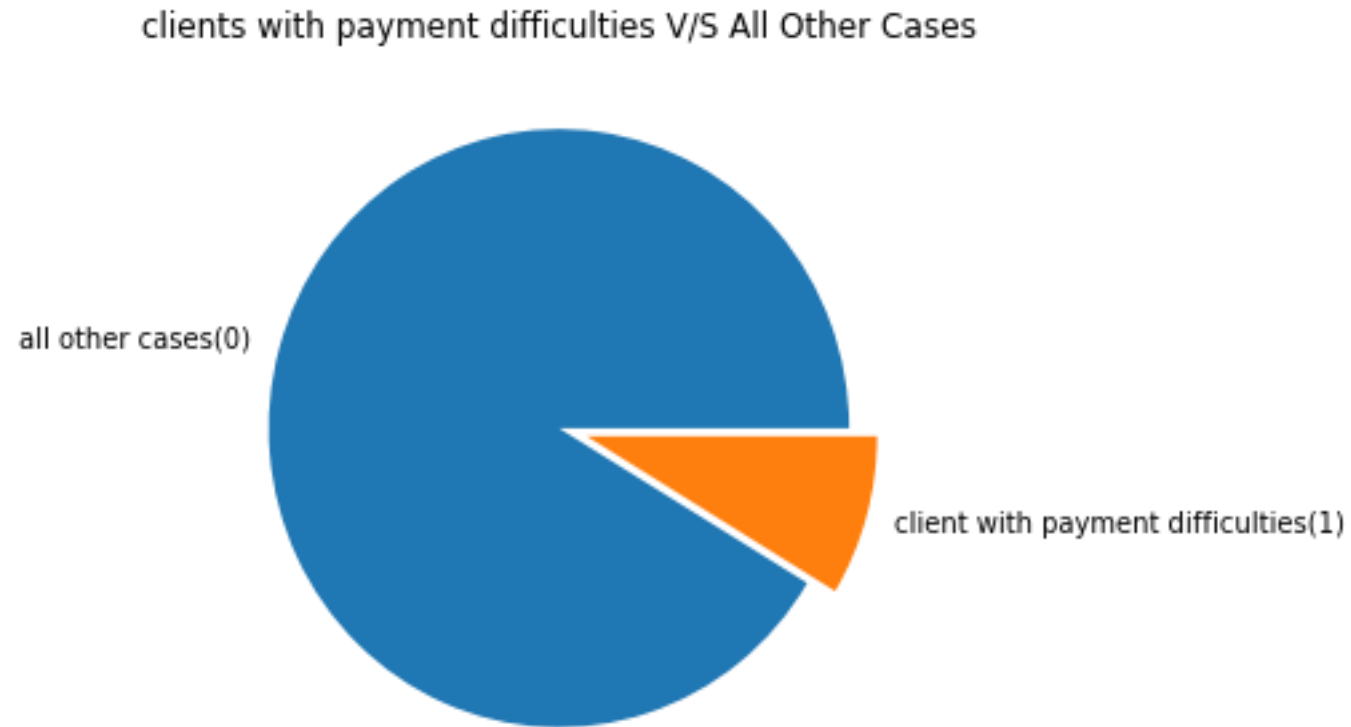
- 1.To help the bank to get to know probable defaulters of the bank based on their loan application*
- 2. This also helps the bank to control losses and avoid any big FINANCIAL calamity*

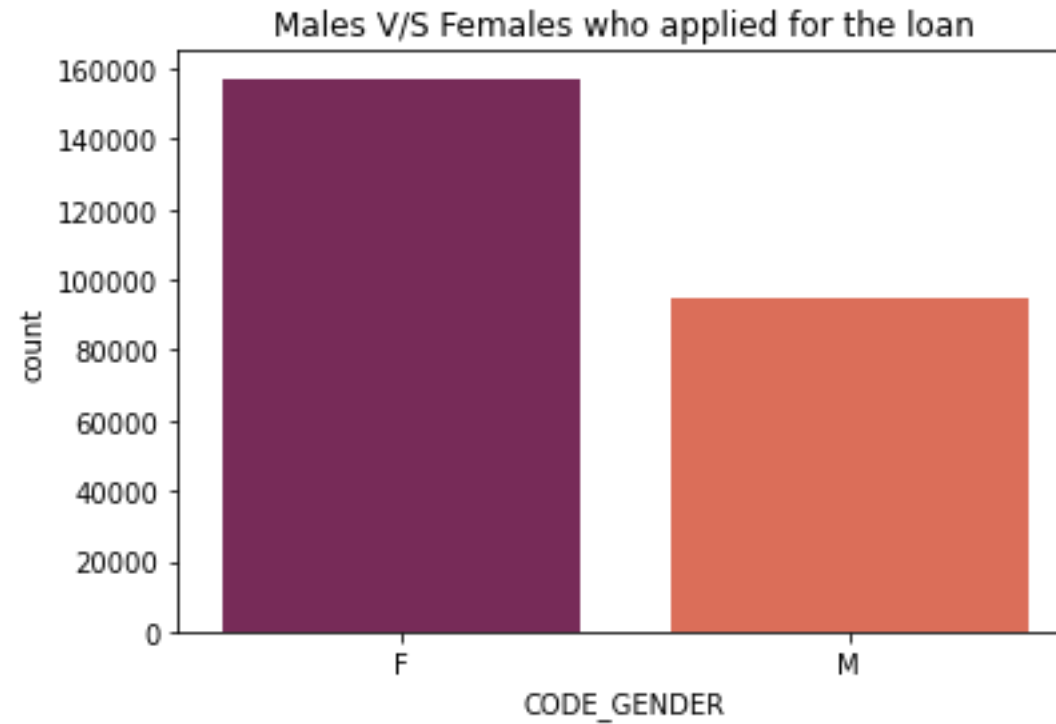
Steps involved in analysis :-

- *Imported necessary libraries like matplotlib.pandas,numpy and seaborn*
- *Checking for outliers using a boxplot and if the column is continuous not having any outliers then we can impute it using mean else median*
- *If the column is categorical than we can use mode to impute the null values*
- *Dropping the columns having more than 40% of null values*
- *Changing some of the data types to an logical data types.*
- *New application dataset where new application data is stored Previous application where previous application dataset is kept*
- *Dividing the previous application dataset in two, one with target variable of 1 and other with target variable of 0.*
- *Plotting and analyzing the data involving univariate analysis ,bivariate analysis and multivariate analysis.*
- **MERGING PREV APPLICATION AND NEW APPLICATION DATA SET FOR FURTHER ANALYSIS.**

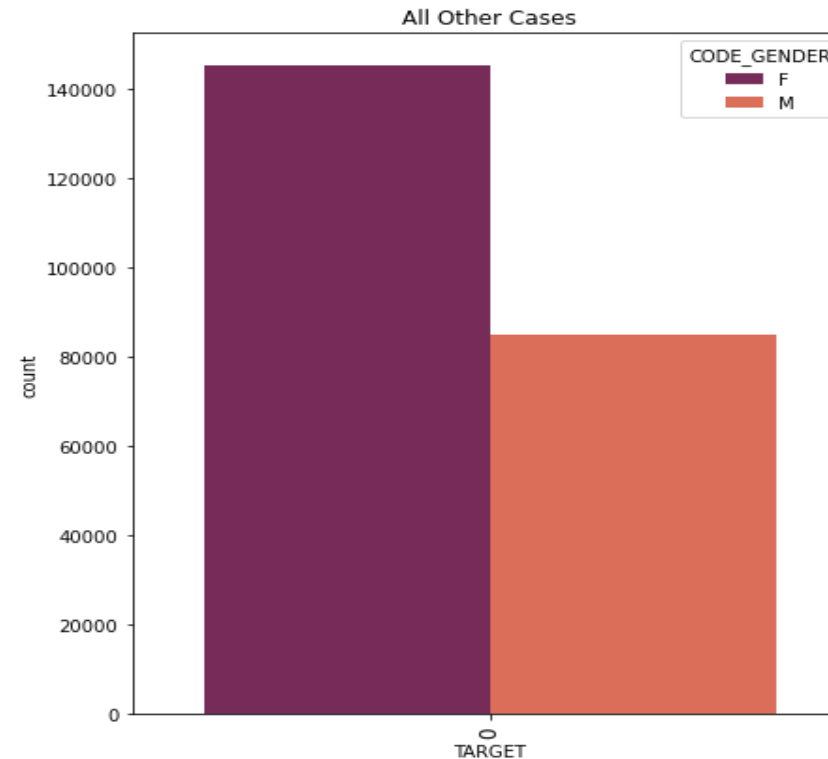
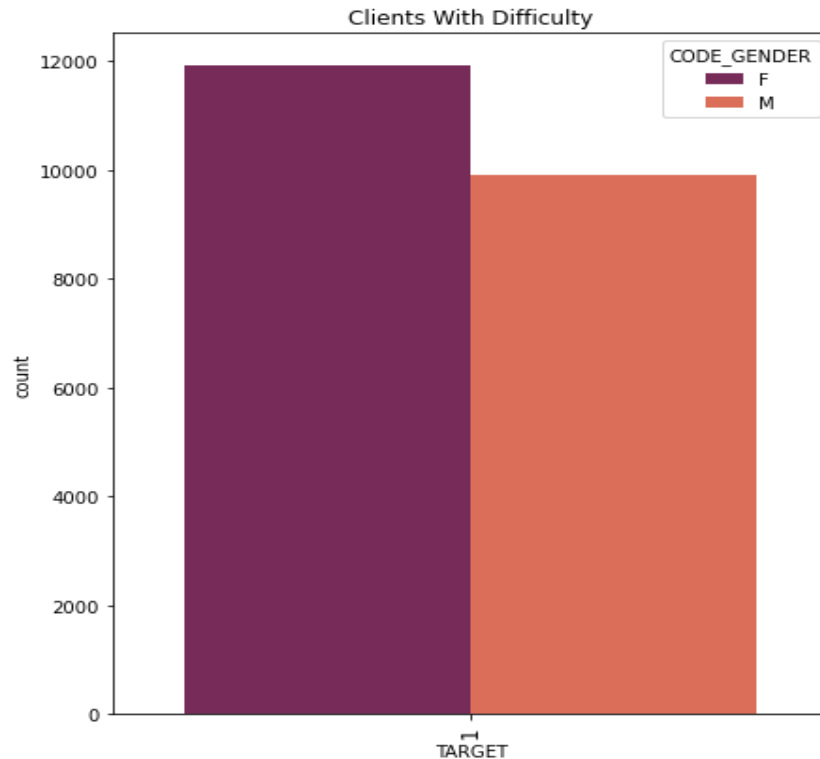
CHECKING FOR IMMBALANCE SO AS TO DIVIDE THE DATASET IN TWO PARTS

There is clear imbalance in the dataset where all other cases(0) is far more superior to those of client with payment difficulties(1)





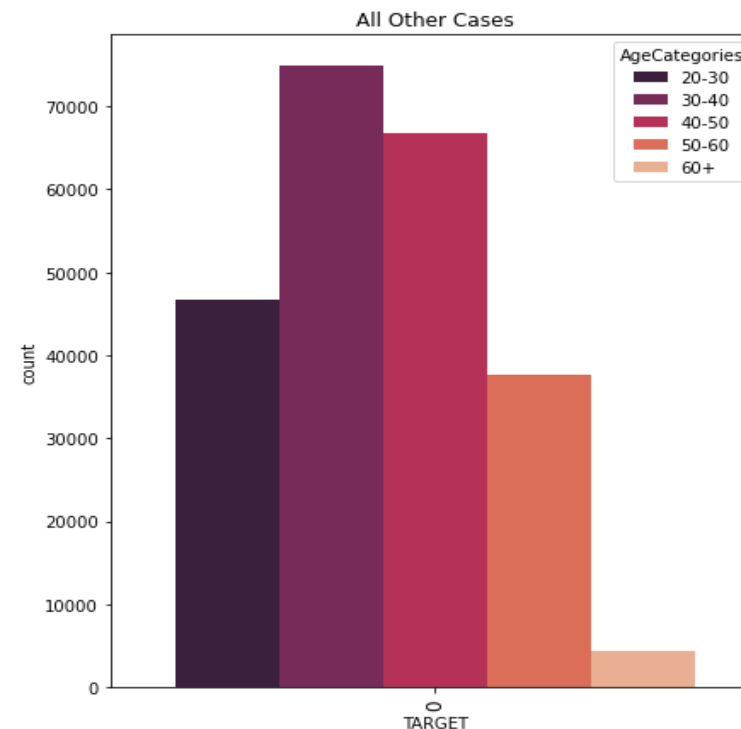
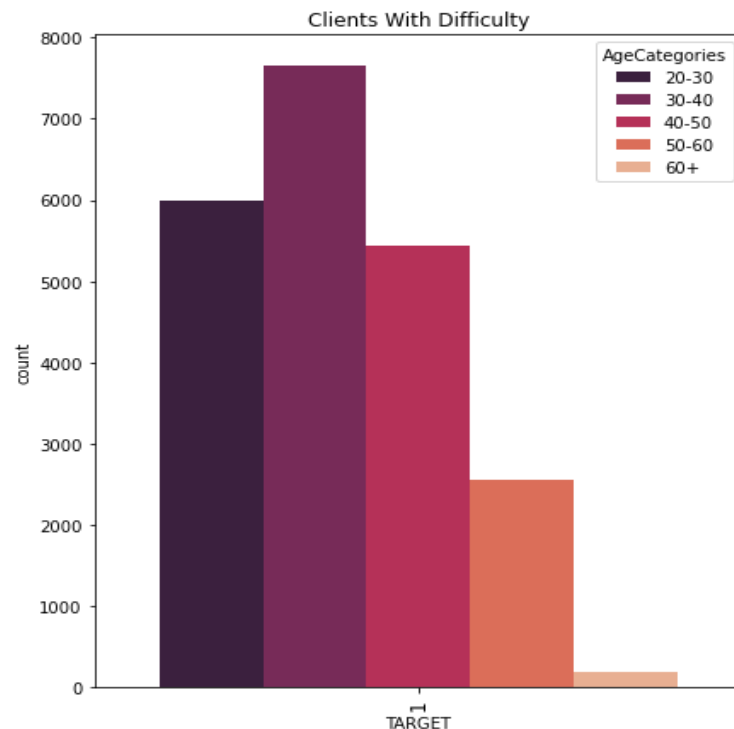
- Here we can see that count of females applying for loan is high as compare to males which can be both defaulter or non-defaulter
- so to overcome above issue we can distribute the data into two sets i.e. One for the client with payment difficulties(1) and other one for all other cases(0)



from above we can draw these points :-

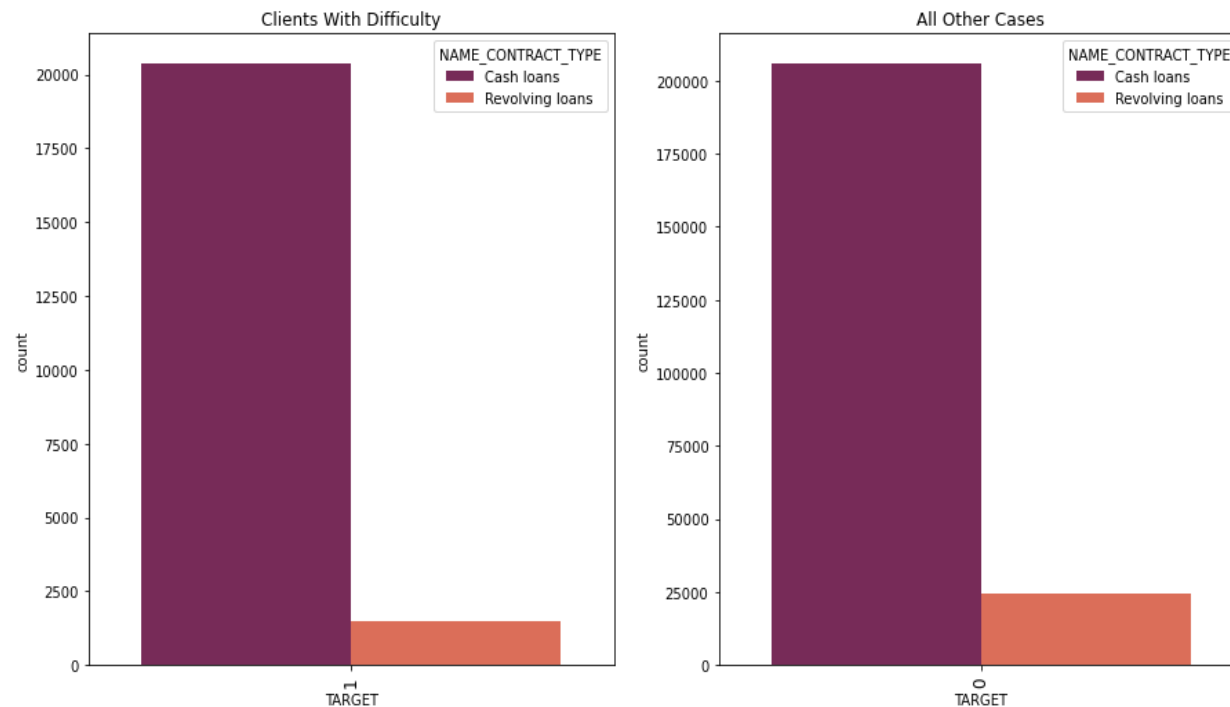
- female clients are greater in numbers for availing loan from the bank then male clients*
- Even count of female defaulting is high but they are less risky customer than male whose chances of defaulting is high as compared to females.*

- *age categories of who are availing more loan falls under the age group of 30-40 followed by 40-50 in both the cases*
- *20-30 years of age categories are also availing loan and is most likely to default hence they are risky age group to disburse the loan*
- *60+ age group or senior citizen are taking less loan comparatively but the repaying percentage is good comparatively with other age group that is they are creditable customer. Also the same goes for 50-60 years of age group.*
- *clients with difficulty in payments mostly fall under 30-40 followed by 40-50 years of age group age categories of 40-50 is also an profitable option for the bank because there percentage of repaying loan back is higher than their percentage of defaulting the loan.*



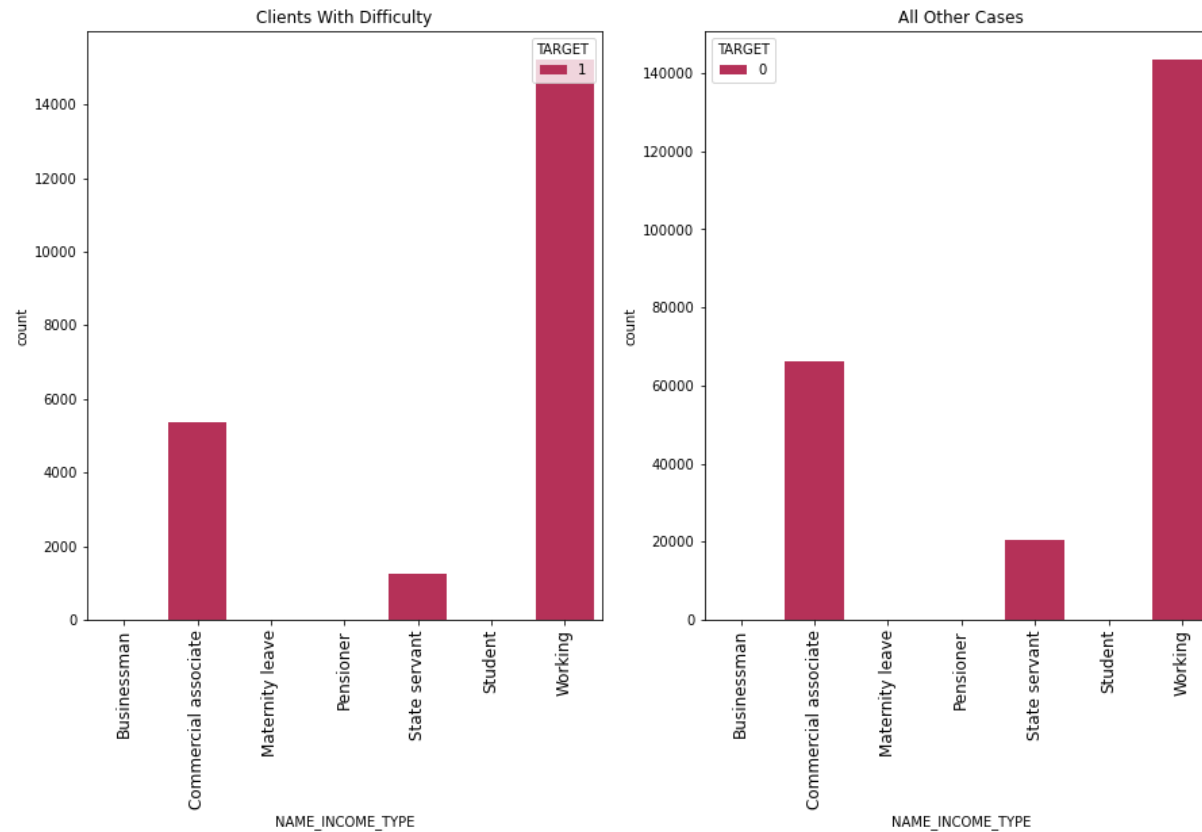
Analyzing NAME_INCOME TYPE COLUMN WITH RESPECT TO TARGET.

- ❖ *Here it is quite visible that cash loan is most disbursed loan by the bank in both the cases whereas revolving loans are in very less proportion.*
- ❖ *also the people availing revolving loans are facing less difficulty in repayments.*

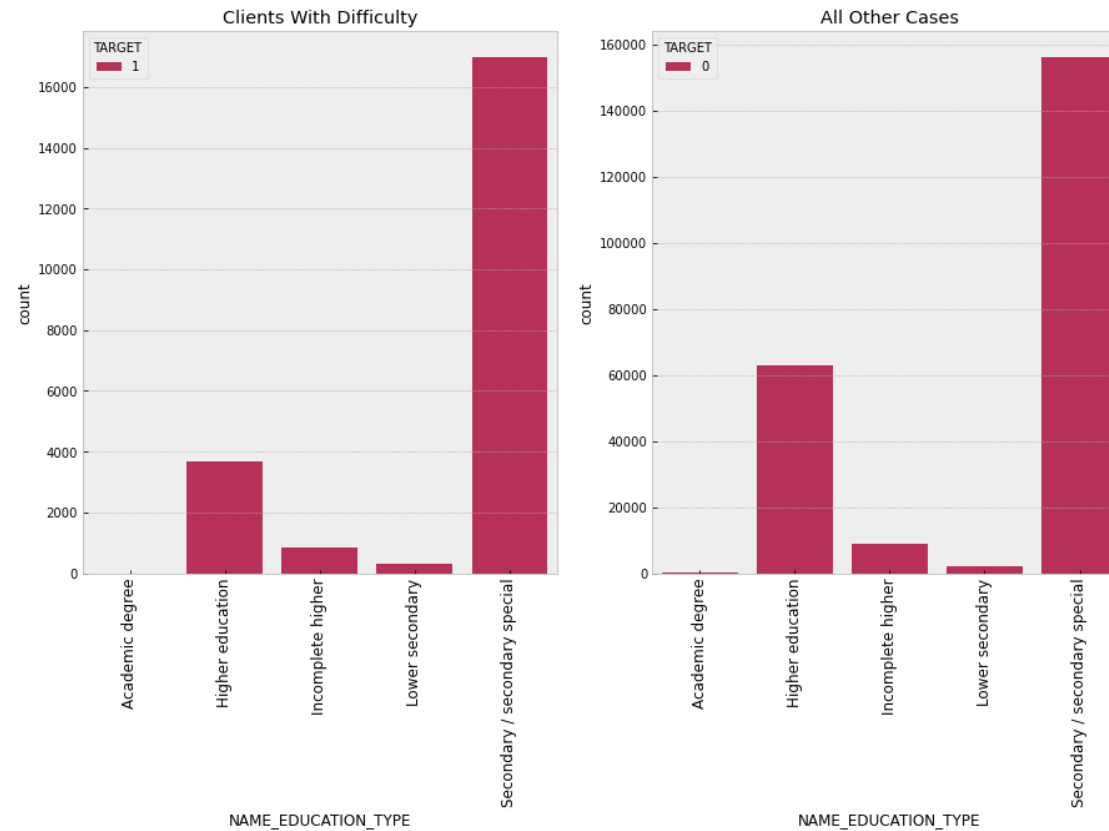


Analyzing NAME_INCOME TYPE COLUMN WITH RESPECT TO TARGET.

- *most loan availed are from working class category followed by commercial associates but working class is also very likely to default the loan thus making them a risky customers for the bank.*
- *commercial associates has minimum risk to default the loan.*
- *state servant avails less loan but is very likely to repay also which means its risk is minimal.*



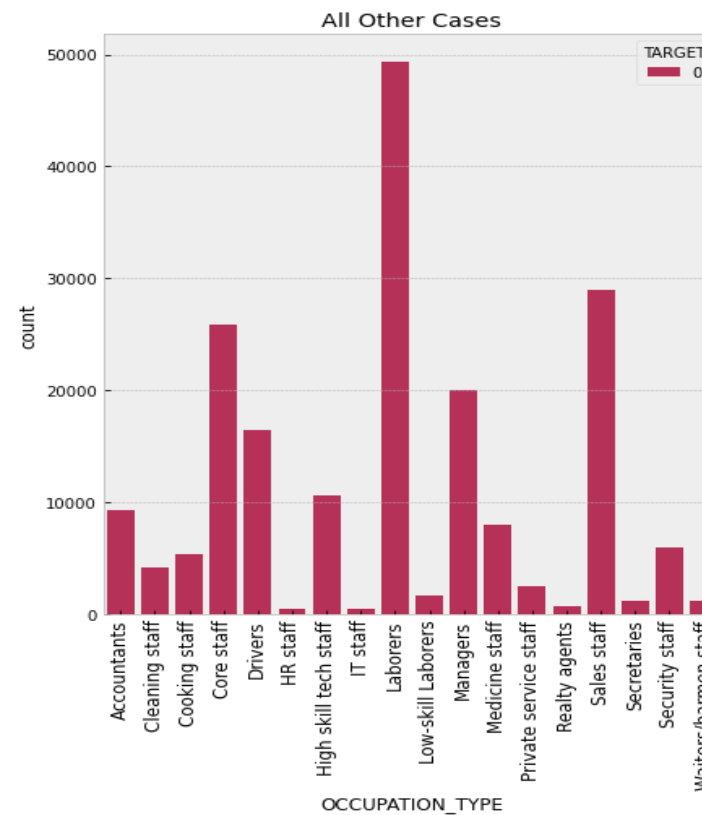
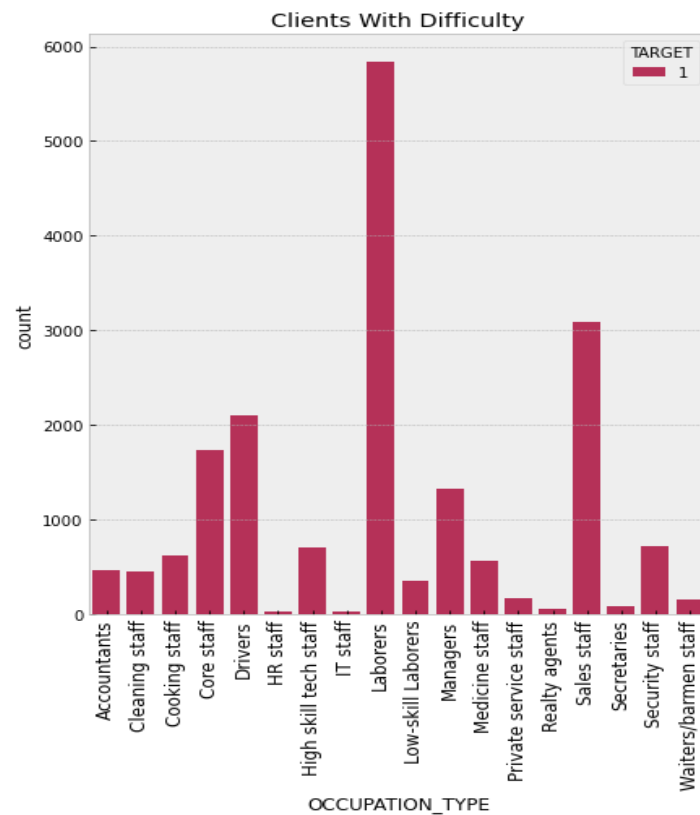
Analyzing NAME_EDUCATION TYPE COLUMN WITH RESPECT TO TARGET.



- *secondary/secondary special is the group who avails most loan from the bank but there risk of defaulting is also high*
- *minimal risk out of all is for higher education as there repayment probability is quite high thus making a good customer for the bank.*

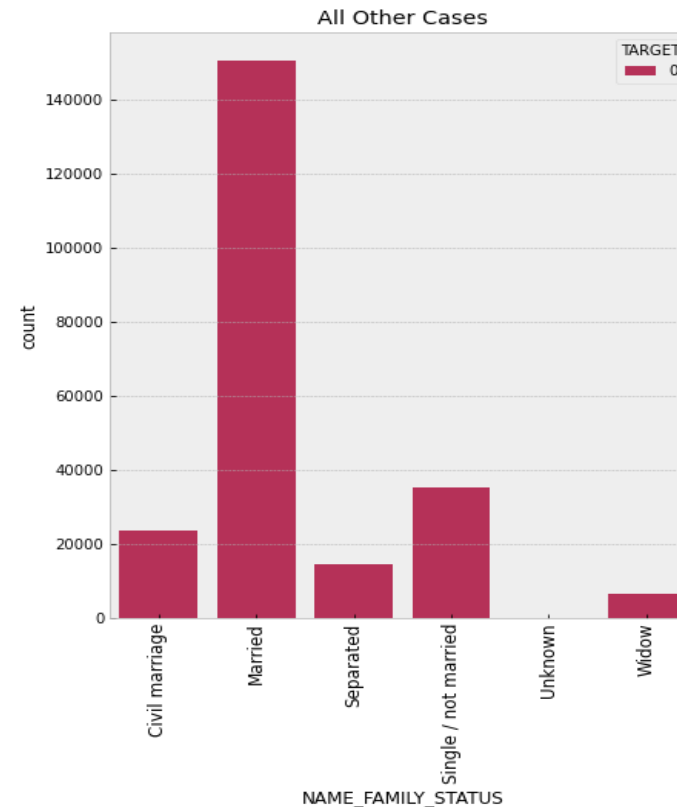
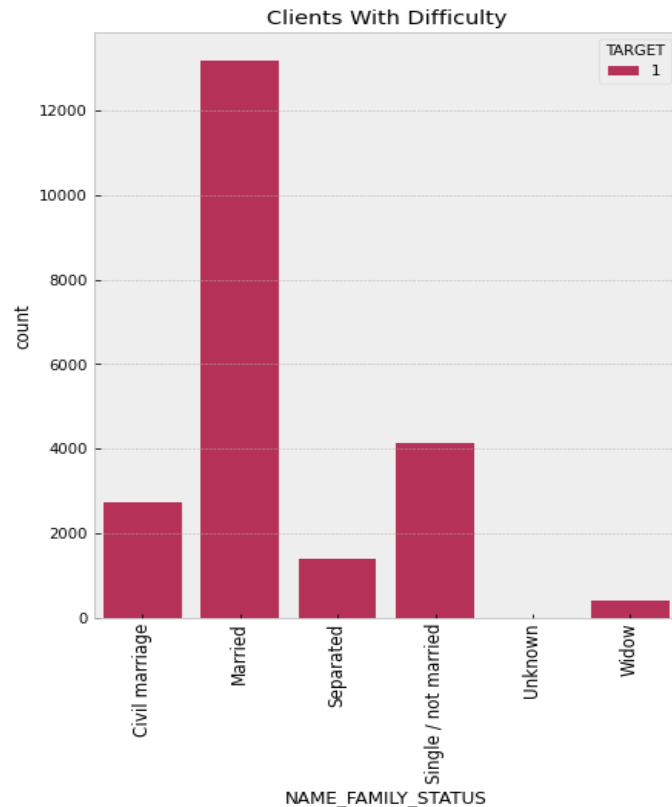
Analyzing OCCUPATION_TYPE COLUMN WITH RESPECT TO TARGET.

- most loan are taken by laborers in both the cases followed by sales staff.
- minimal risk group includes IT staff,Realty agents,Secretaries,HR staff in case of defaulting the loans as they take less loan and their repaying percentage is good.
- Risky customer comes under laborers followed by sales staff in case of defaulting



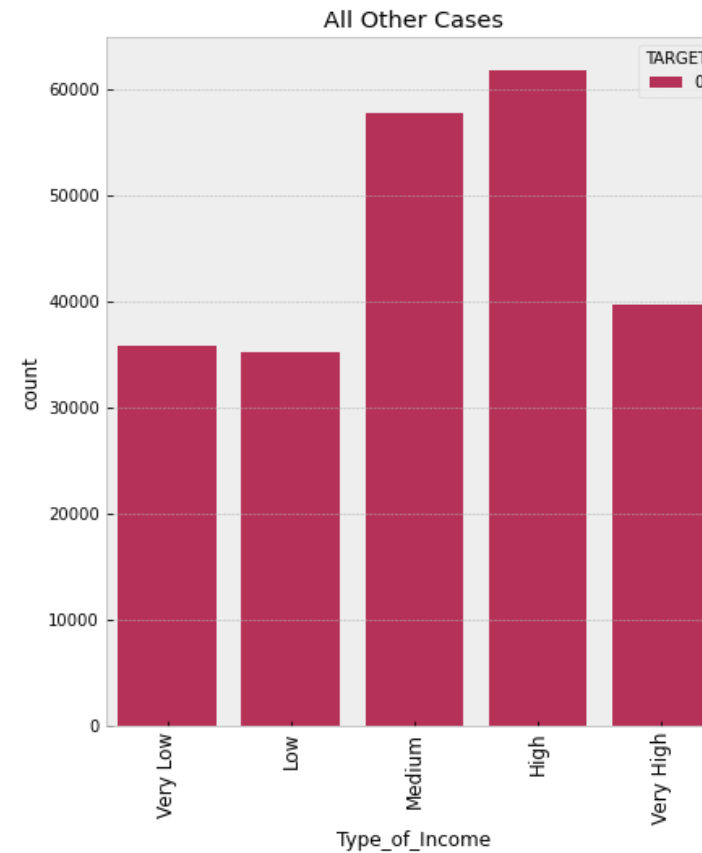
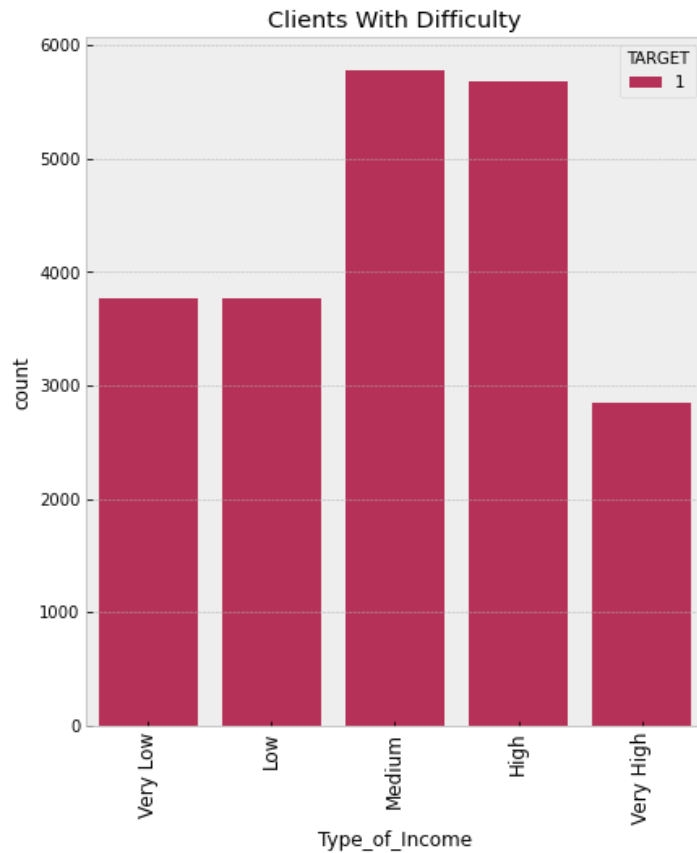
Analyzing OCCUPATION_TYPE COLUMN WITH RESPECT TO TARGET.

- Most loan are availed by the married peoples which is quite valid point as well, but there chances of defaulting is also very high.
- single/not married group comes in second place for taking loan but also shows positive sign in case of repayment. Hence making it less risky.
- Window's group risk is minimal in case of difficulty or defaulting



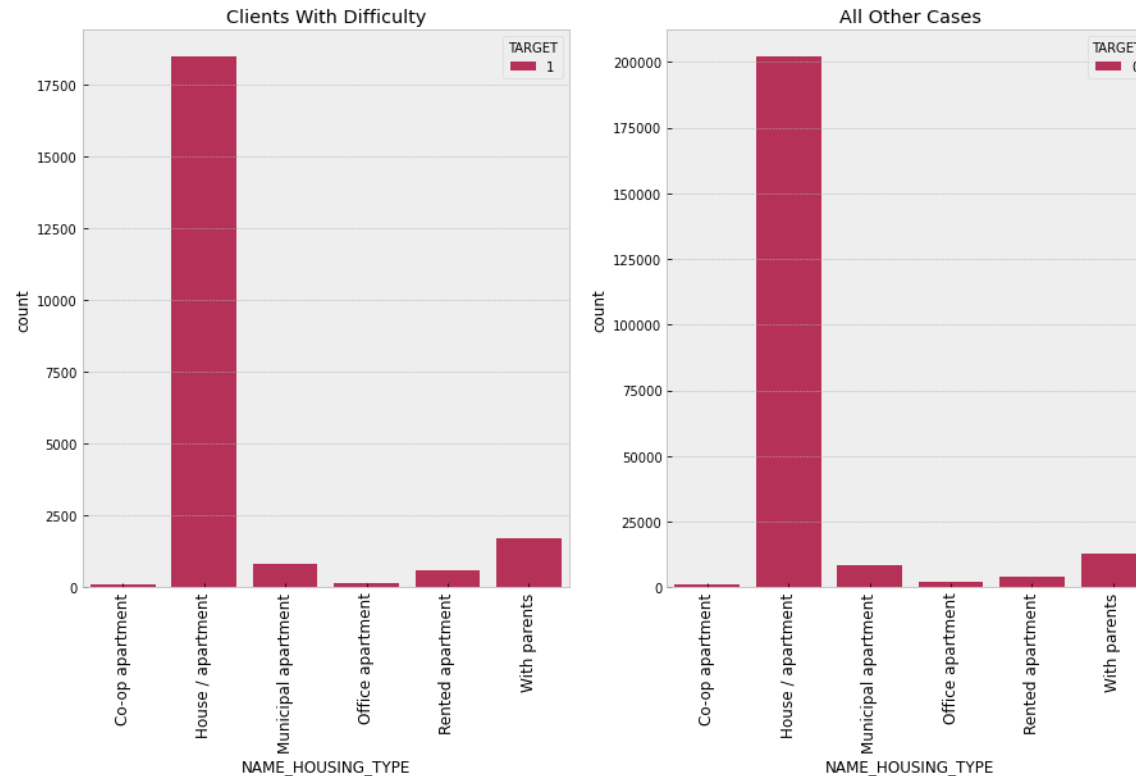
Analyzing Type_Of_Income COLUMN WITH RESPECT TO TARGET

Customers belonging to high income group and medium income group avails most loan out of all other categories.
Very high income group is more likely to repay the loans to the bank.
Very low and Low income group people are most likely to default their loans.



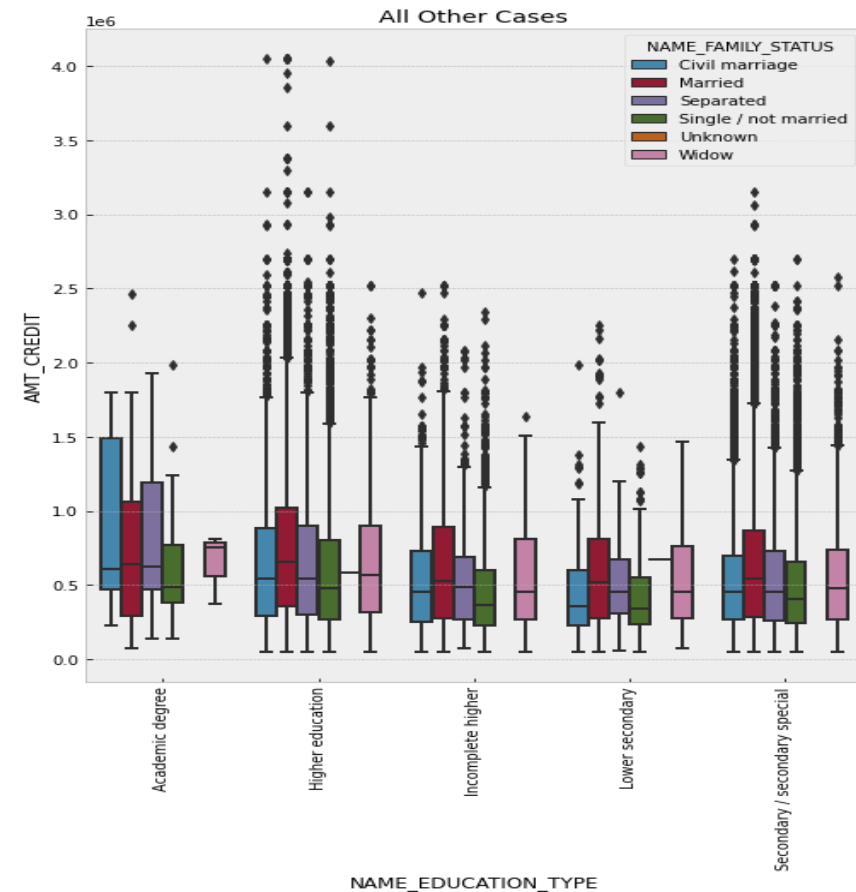
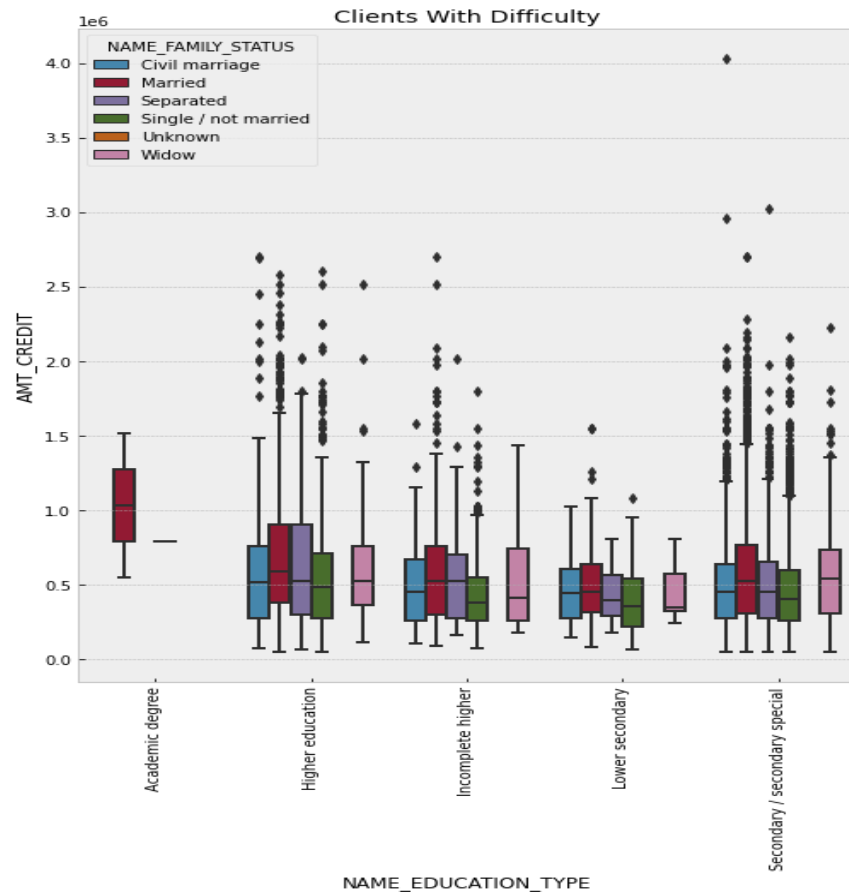
Analyzing NAME_HOUSING_TYPE COLUMN WITH RESPECT TO TARGET

- Most loans are taken by the people living in house/apartments for both the cases.
- most risky customer for the bank are the people living with their parents because there chances of defaulting is more than their chances of repayment.
- Also people living in rented apartments tends to be a risky customers for the bank.

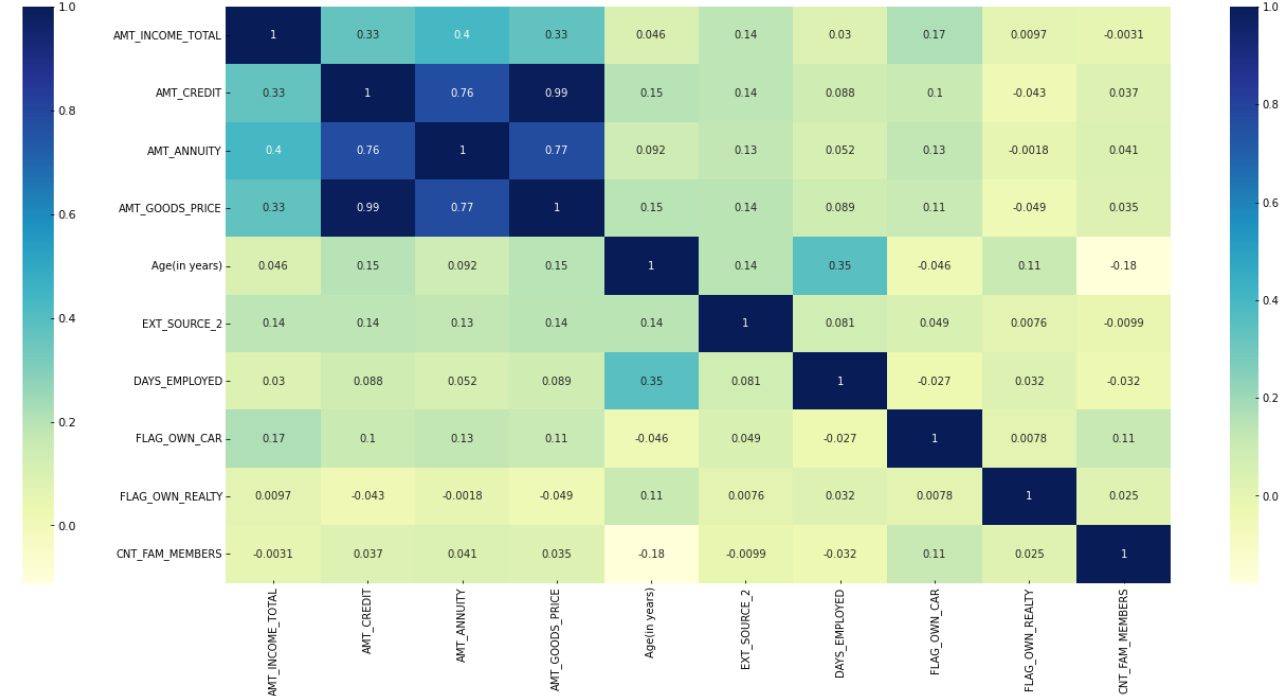
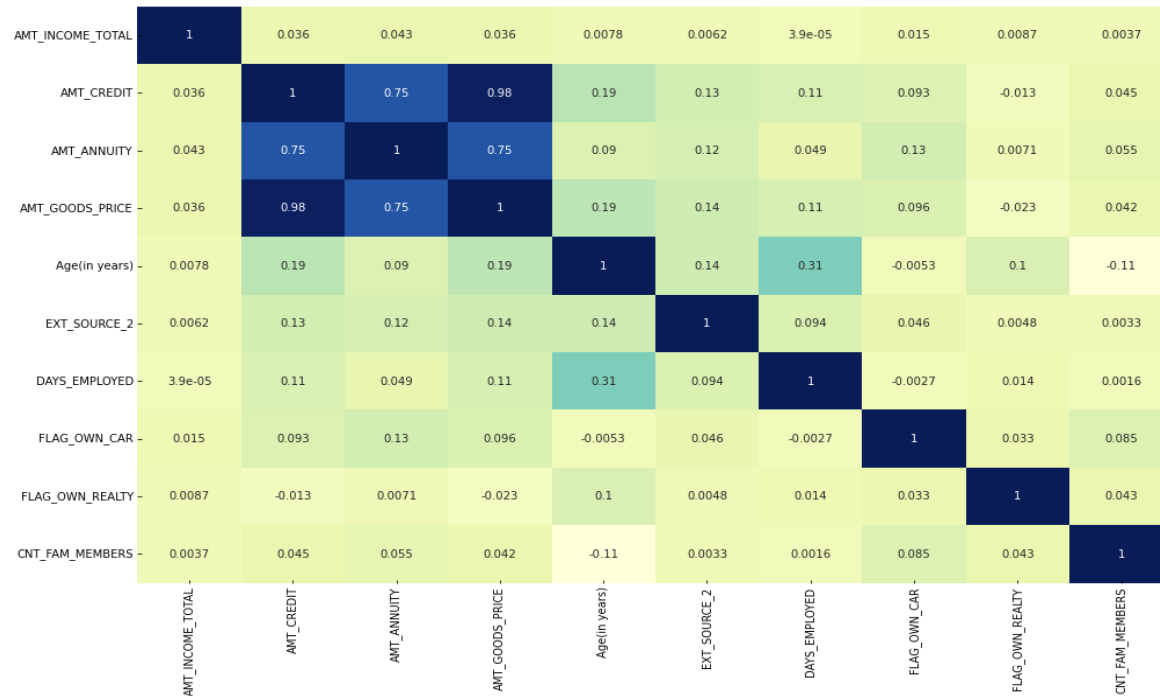


Analyzing NAME_EDUCATION_TYPE AND AMT_CREDIT COLUMN WITH RESPECT TO TARGET

People in Higher education and secondary education tends to high credit from the banks(as we can see outliers present in them in both the cases)
client in secondary special taking very high credit is risky for the banks as well.



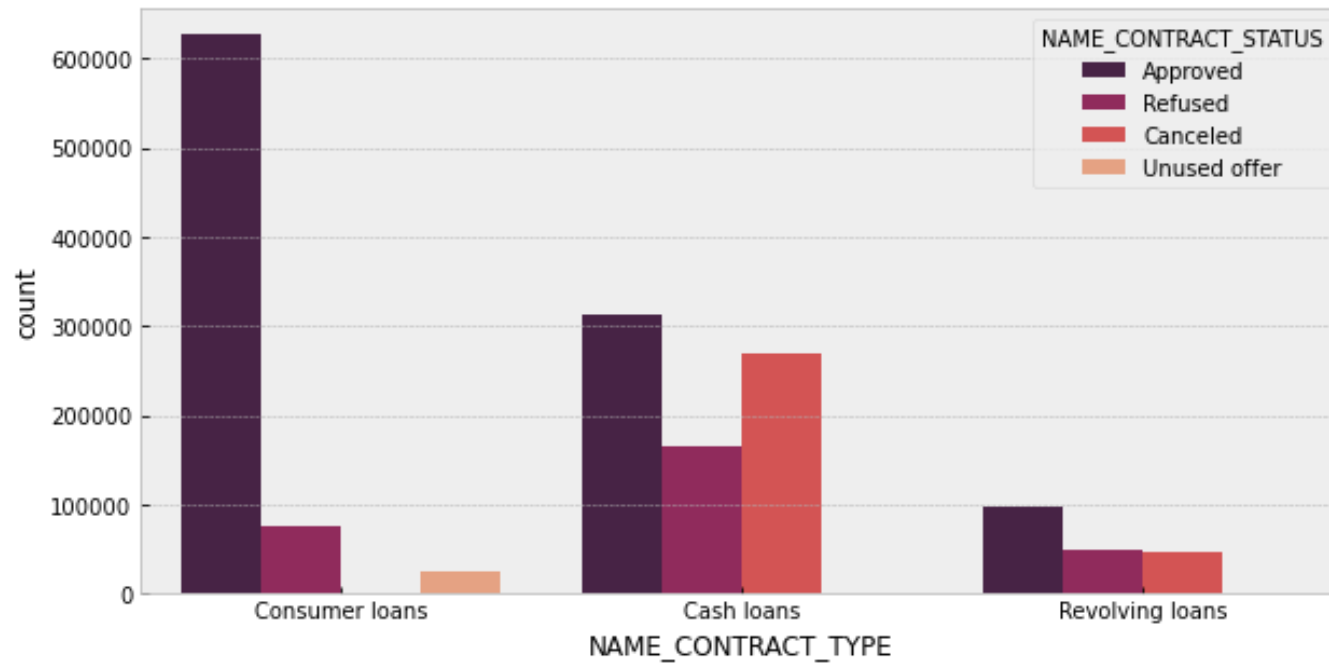
CORRELATION WITH RESPECT TO TARGET IN NEW APPLICATION DATASET



- AMT_CREDIT AND AMT_GOODS PRICE IS HIGHLY CORRELATED WITH AN POSITIVE CORR OF 0.98(FOR
- AMT_ANNUITY AND AMT_GOODS PRICE IS ALSO POSITIVELY CORRELATED FOR BOTH THE CASES
- EXT_SOURCE_2 IS ALSO POSITIVELY CORELATED WITH GOODS_PRICE
- (HERE LEFT ONE BEING CLIENTS WITH DIFFICULTY AND RIGH ONE BEING ALL OTHER CASES DATSET)

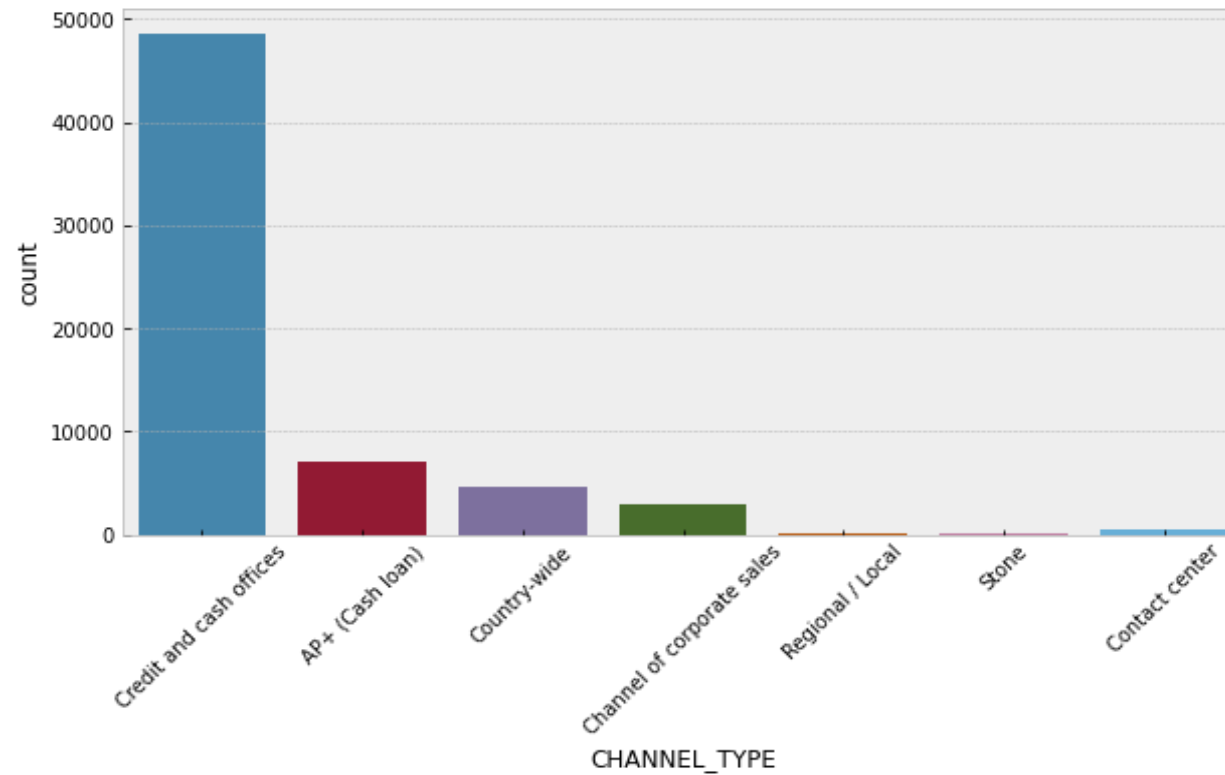
ANALYSIS ON PREVIOUS APPLICATION DATASET ON NAME_CONTRACT_TYPE COLUMN

- from here we can see that cash loans are most preferred loan application followed by consumers Loan
- Revolving loan being the least preferred.
- most approved loan is consumer loan even whereas cash loan are rejected heavily as compared to others type of loans

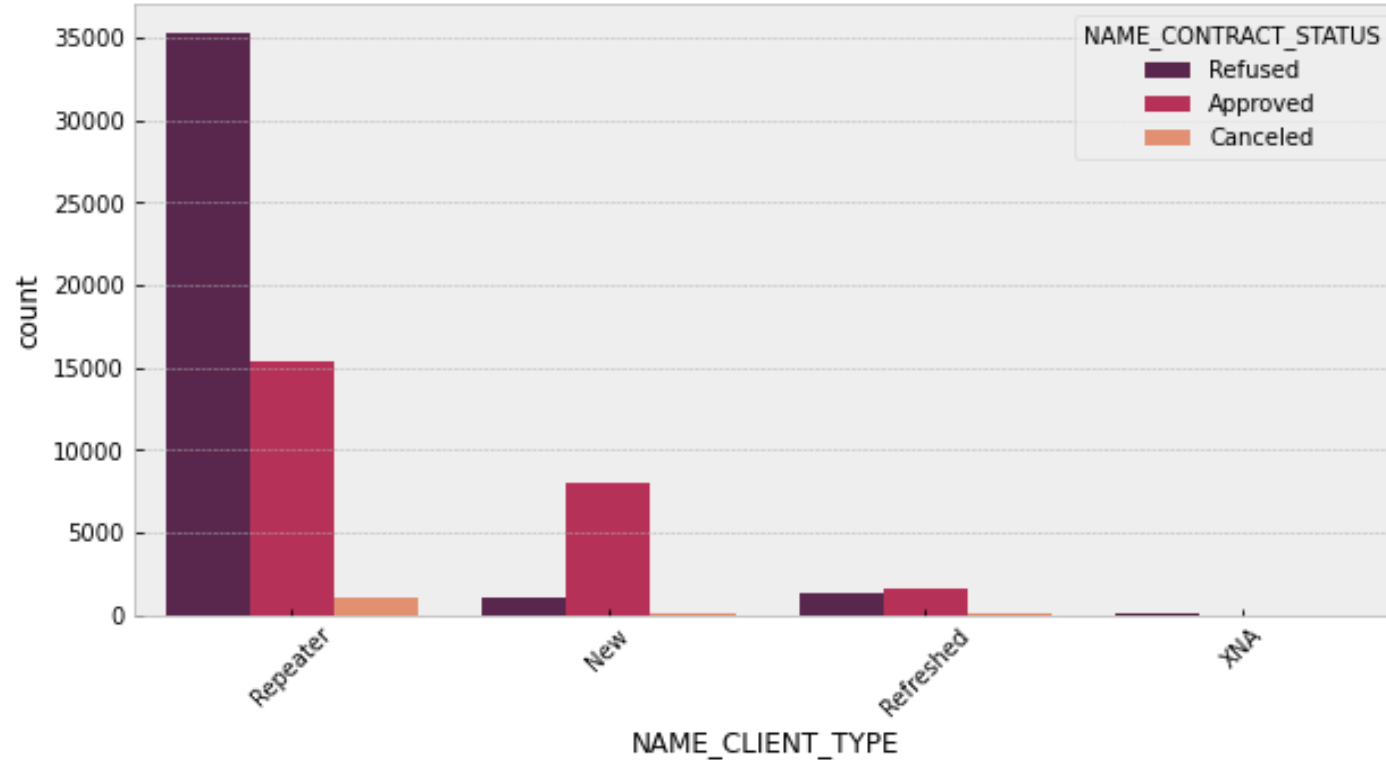


ANALYSIS ON PREVIOUS APPLICATION DATASET ON CHANNEL_TYPE COLUMN

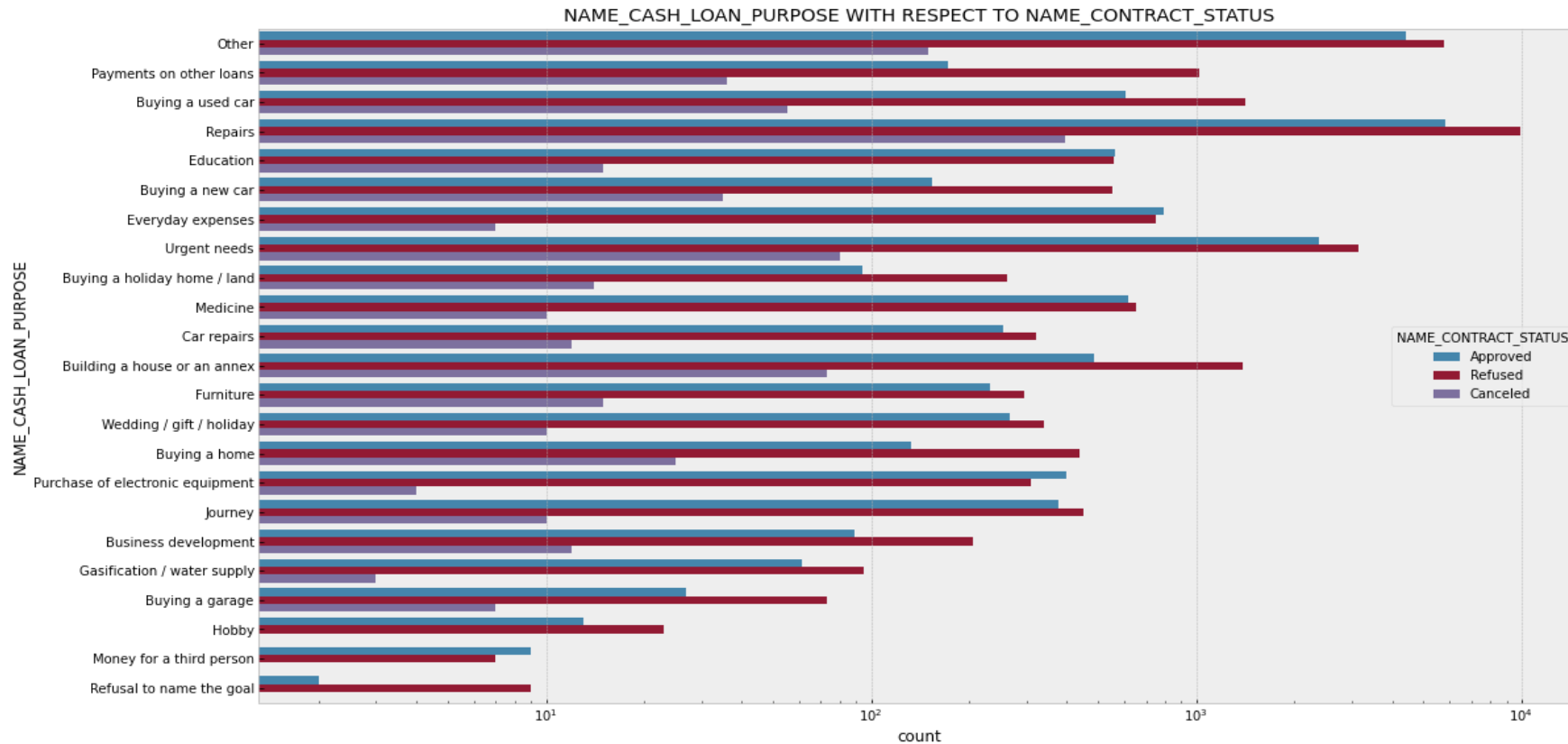
- most of the previous application client where acquired from credit and cash offices followed by country-wide
- Contact center being the least of all of them.



- most of clients are repeater and very few of them are new and refreshed comparatively
- repeaters groups is the most refused group out of all other groups but there approved percentage is also quite high

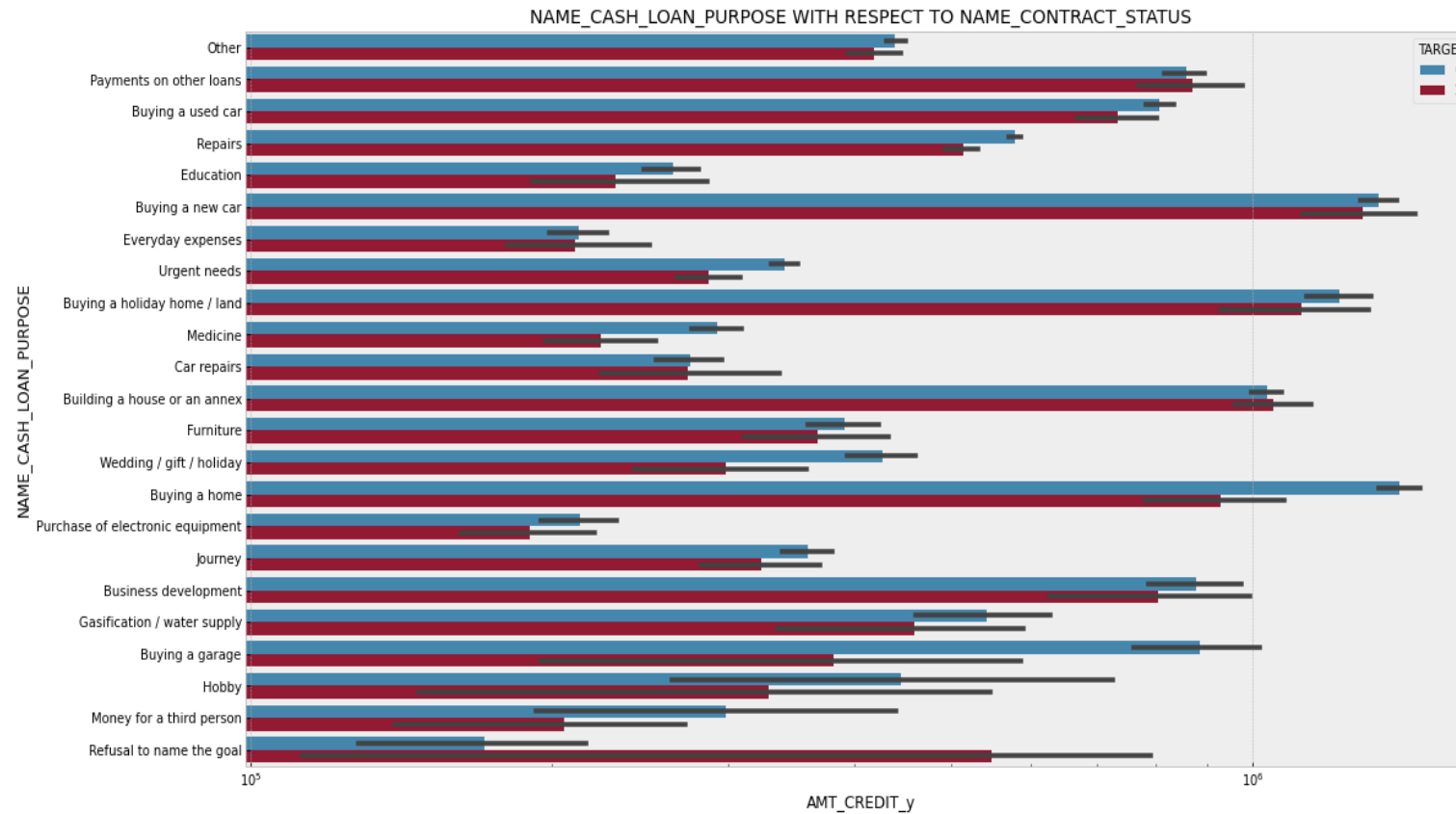


ANALYSIS ON MERGED APPLICATION DATASET ON NAME_CASH_LOAN PURPOSE COLUMN W.R.T NAME_CONTRACT_STATUS



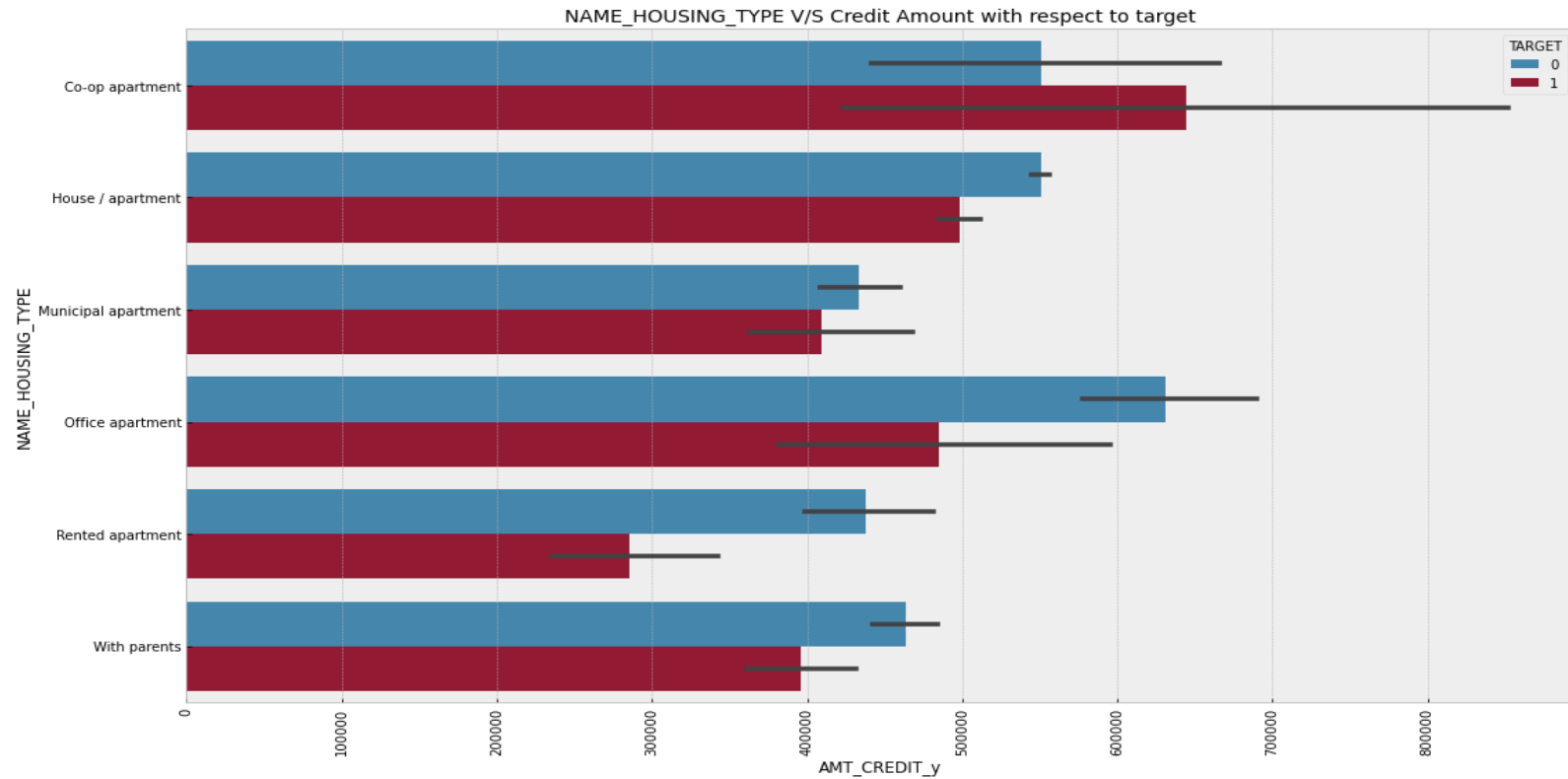
- Here we can see that repairs is where most of cash loan purpose rest but it also faces rejection the most followed by other group

BIVARIATE ANALYSIS ON MERGED APPLICATION DATASET ON NAME_CASH_LOAN V/S AMT_CREDIT COLUMN W.R.T NAME_CONTRACT_STATUS



- cash loan purpose for buying a new car taking high credit has 40% chances to default
- buying a home where credit amount is high in that case more customer usually pay off their loans

BIVARIATE ANALYSIS ON MERGED APPLICATION DATASET ON NAME_HOUSING_TYPE COLUMN V/S AMT_CREDIT W.R.T TARGET COLUMN.



- BANK should avoid loans to co-op apartment as they have most difficulty in repayments.

RECOMMENDED GROUP TO WHOM LOAN CAN BE CREDITED:-

- PEOPLE OF AGE GROUP OF MORE THAN 40 YEARS
- CLIENTS WITH HIGH INCOME CATEGORY
- OLD FEMALE CLIENT
- WIDOW CLIENT ARE ALSO MINIMAL RISK GROUP
- CLIENTS WORKING AS STATE SERVANT
- REVOLVING LOAN EVEN THOUGH LESS APPLICATION BUT GOOD PRODUCT FOR BANK

RISKY GROUPS ARE:-

- PEOPLE LIVING IN CO-OP APPARTMENTS
- SECONDARY EDUCATED CLIENTS TAKING HIGH CREDITS
- PEOPLE LIVING IN RENTED APPARTMENTS
- PEOPLE BELONGING TO WORKING CLASS