

EDA CASE STUDY

RISK ANALYTICS IN BANKING AND FINANCIAL SERVICES



PROBLEM STATEMENT

Two types of risks are associated with the bank's decision when the company receives a loan application

- ❑ If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- ❑ If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

Analysis of data set has been done using Python on Jupyter Notebook

APPROACH FOR ANALYSIS

□ Steps:

- ✓ Data Pre-Processing on application_data dataset
- ✓ Univariate Analysis
- ✓ Outlier analysis
- ✓ Top 10 correlation analysis for the client with payment difficulties and no difficulties
- ✓ Merging application_data and Previous_application Data Frame.
- ✓ Bi-Variate and Multi-Variate analysis on merged DataFrame.

SUB-TASKS PERFORMED IN EDA

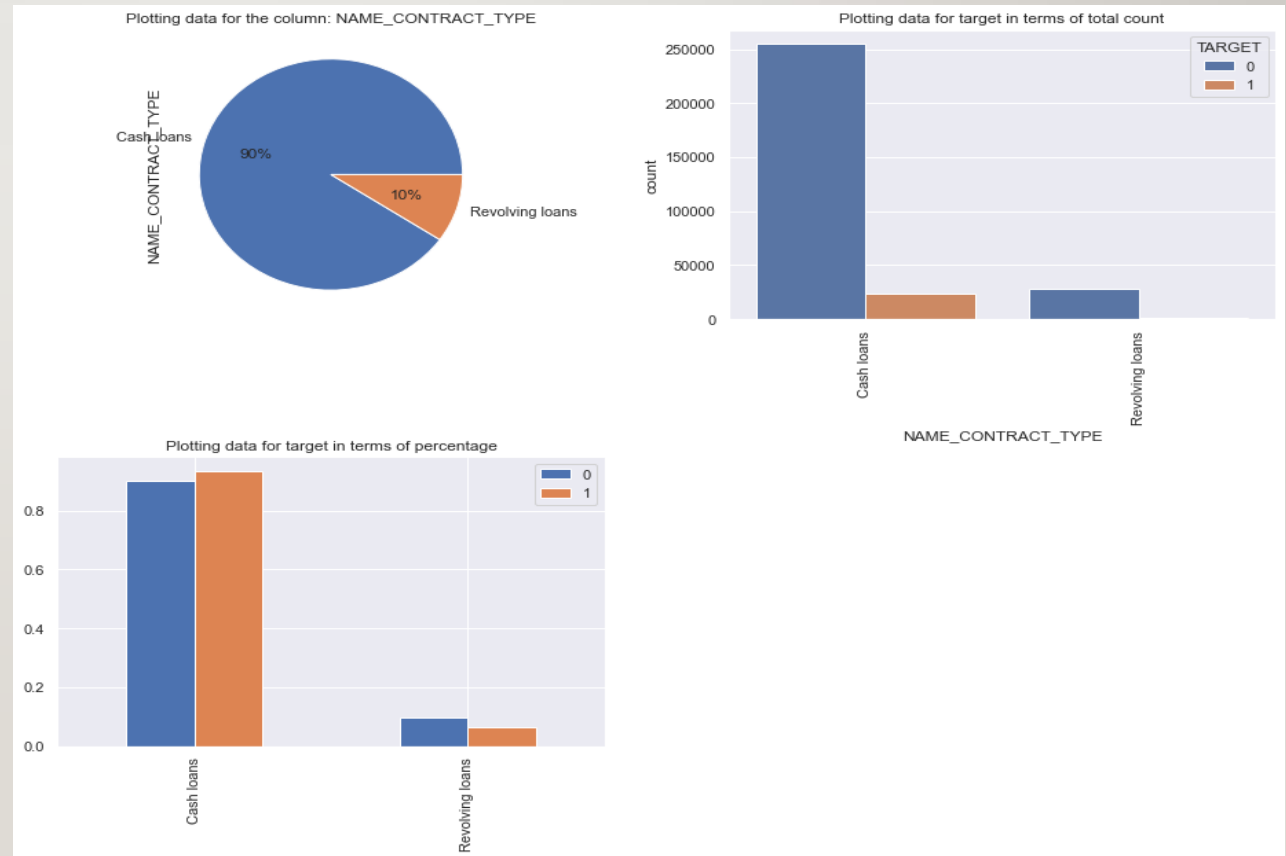
- ✓ Task1: Reading the data and finding the information about the data
- ✓ Task2: Inspecting Data for Data cleaning : Null values, which column to drop and which to impute
- ✓ Task3: Imputing Values(Categorical-Mode; Numeric- depending on type of distribution)
- ✓ Task4: Checking Datatypes for columns
- ✓ Task5: Checking the Data, after data cleaning operation
- ✓ Task6: Check Data Imbalances
- ✓ Task7: Univariate Analysis
- ✓ Task8: Checking for Outliers
- ✓ Task9: Finding Co-relation in the data
- ✓ Task10: Bivariate Analysis of numerical variables
- ✓ Task11: Analysis of Previous Application v/s Defaults and finding co-relations

UNIVARIATE ANALYSIS FOR TARGET 0 AND TARGET 1

DISTRIBUTION FOR NAME_CONTRACT_TYPE

Points to be concluded from the graph on the right.

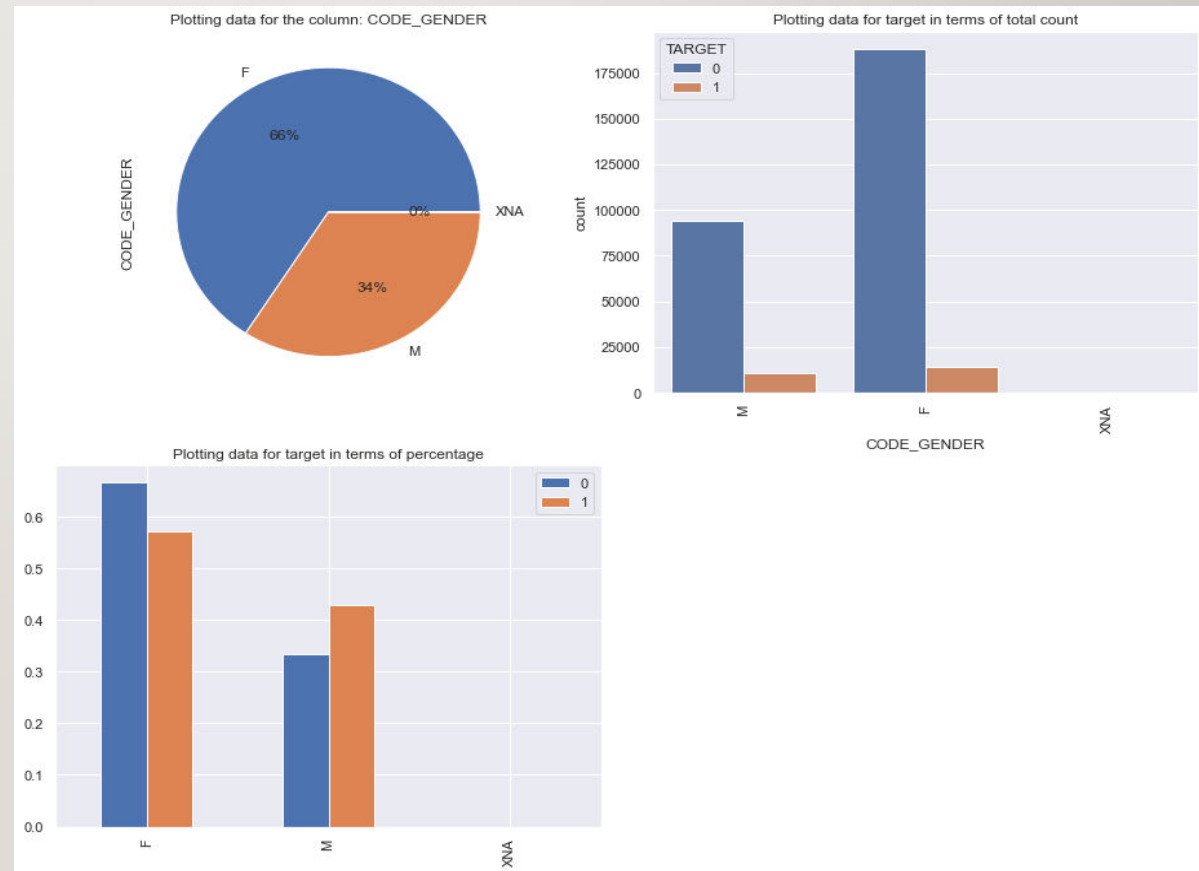
- ✓ For contract type 'cash loans' is having higher number of credits than 'Revolving loans' contract type.
- ✓ More people are availing cash loans as compared to Revolving loans. They seem to have difficulties in repayment of loan as compared to those availing revolving loans



DISTRIBUTION FOR CODE_GENDER

Points to be concluded from the graph on the right.

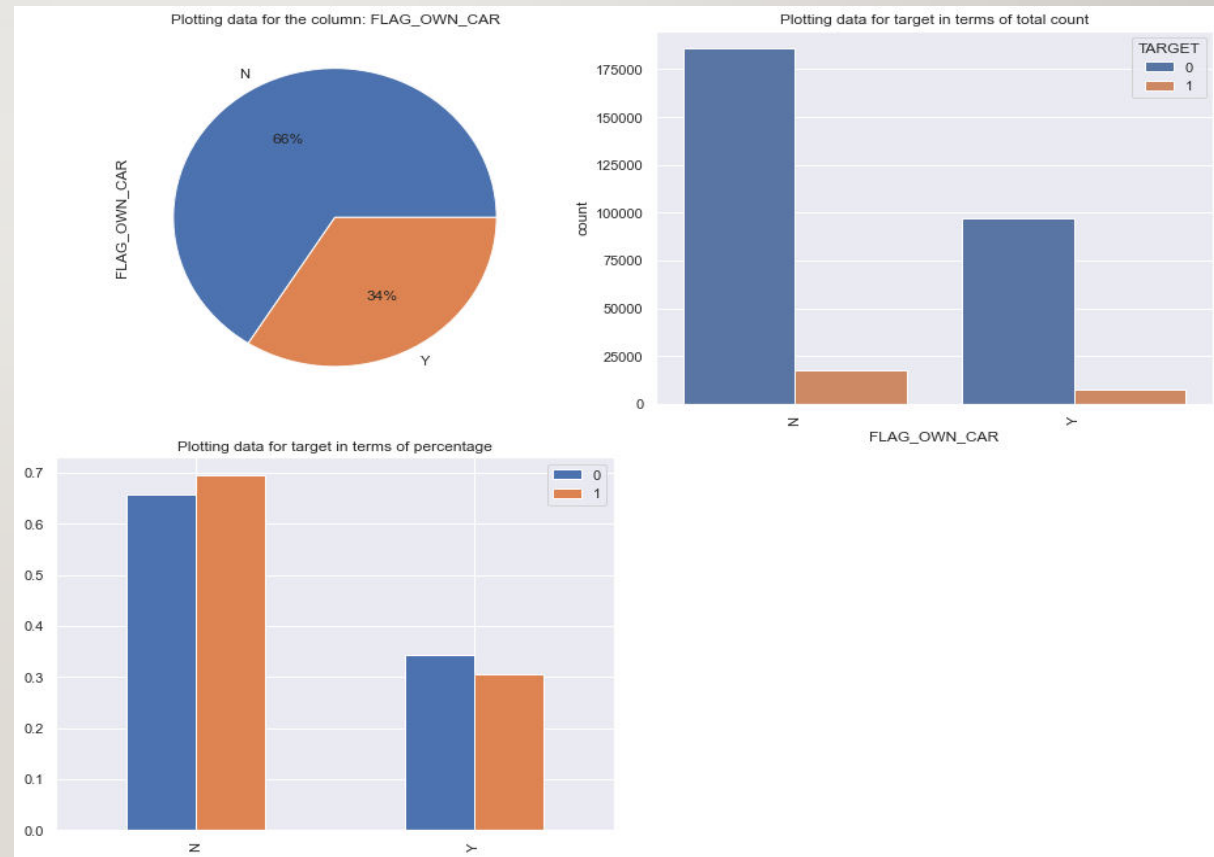
- ✓ The clients availing loans are Females in majority.
- ✓ However, More males seems to be defaulter as compared to Females.



DISTRIBUTION FOR FLAG_OWN_CAR

Points to be concluded from the graph on the right.

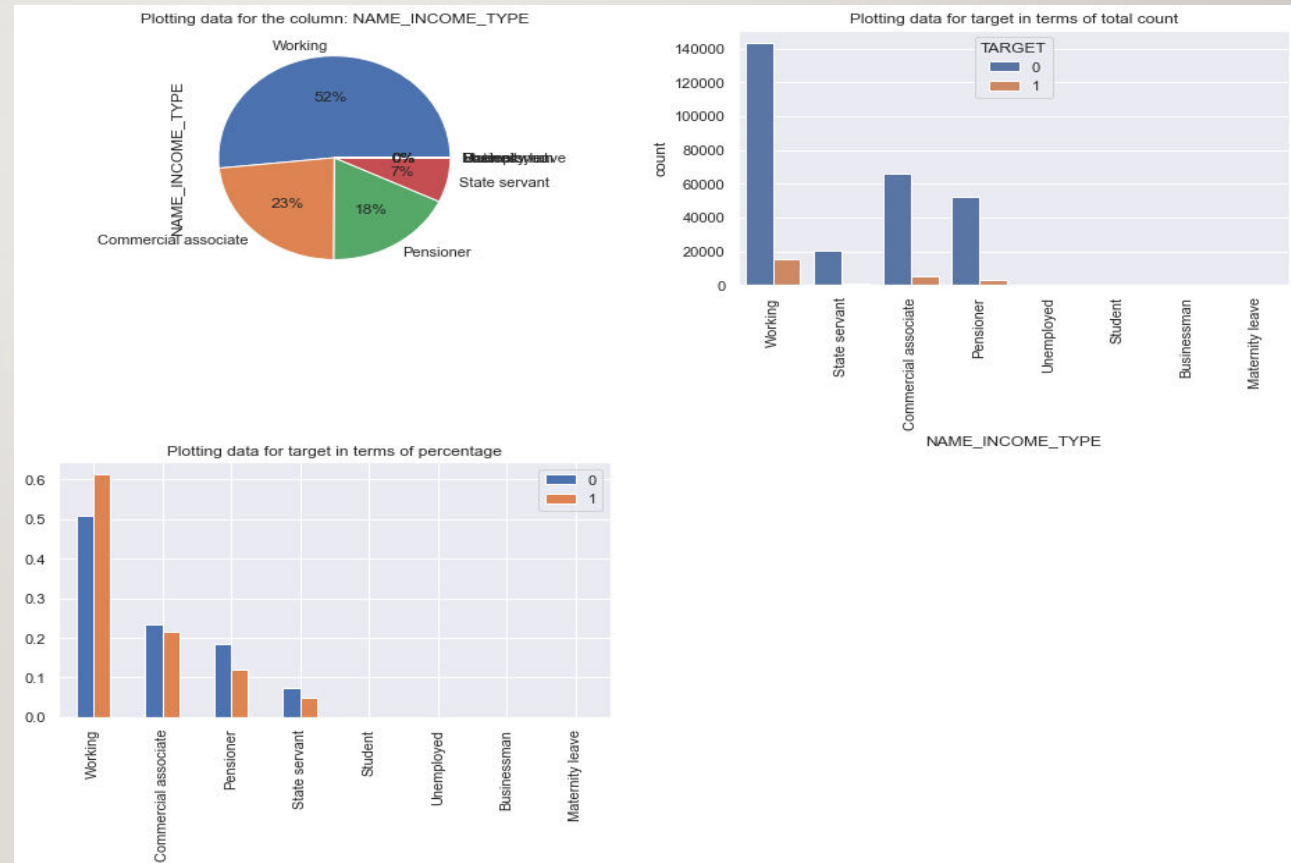
- ✓ The clients who don't own cars tends to be threat of becoming defaulter.



DISTRIBUTION FOR NAME_INCOME_TYPE

Points to be concluded from the graph on the right.

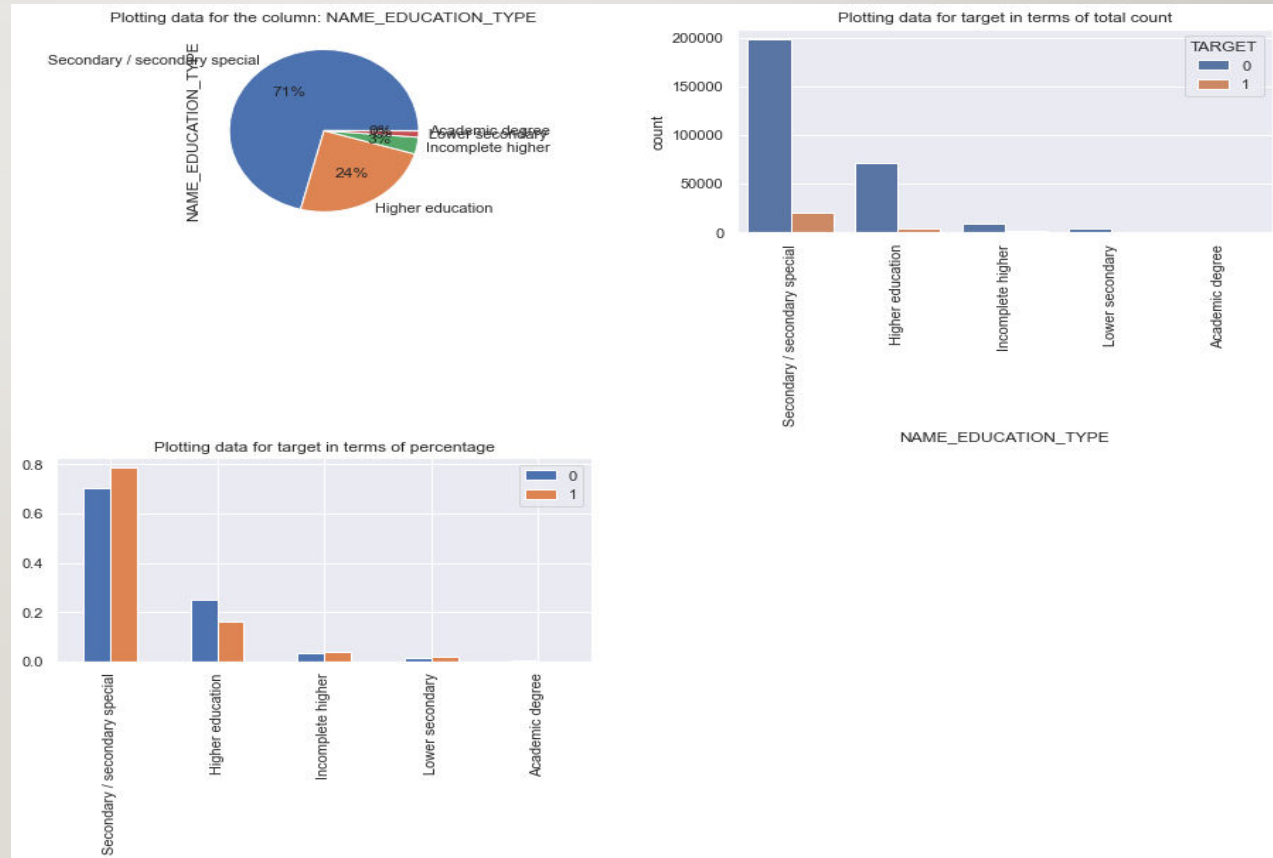
- ✓ Majority of people availing loans are from working category, however they are also more likely to be defaulter.
- ✓ Less number of credits for income type student, pensioner, Businessman and Maternity leave



DISTRIBUTION FOR NAME_EDUCATION_TYPE

Points to be concluded from the graph on the right.

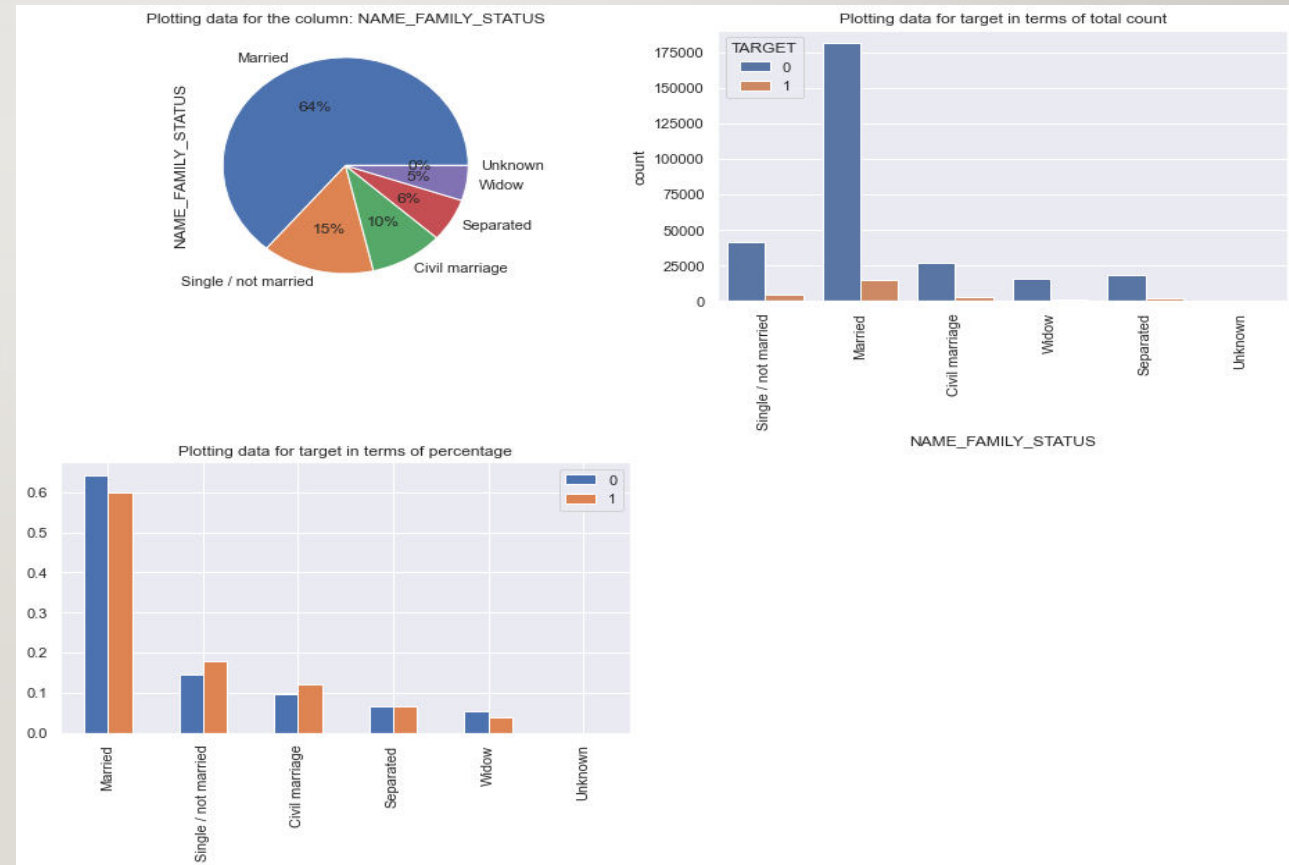
- ✓ Majority of client who are Secondary / Secondary Special are likely to apply for loan.
- ✓ These people are likely to be defaulters



DISTRIBUTION FOR NAME_FAMILY_STATUS

Points to be concluded from the graph on the right.

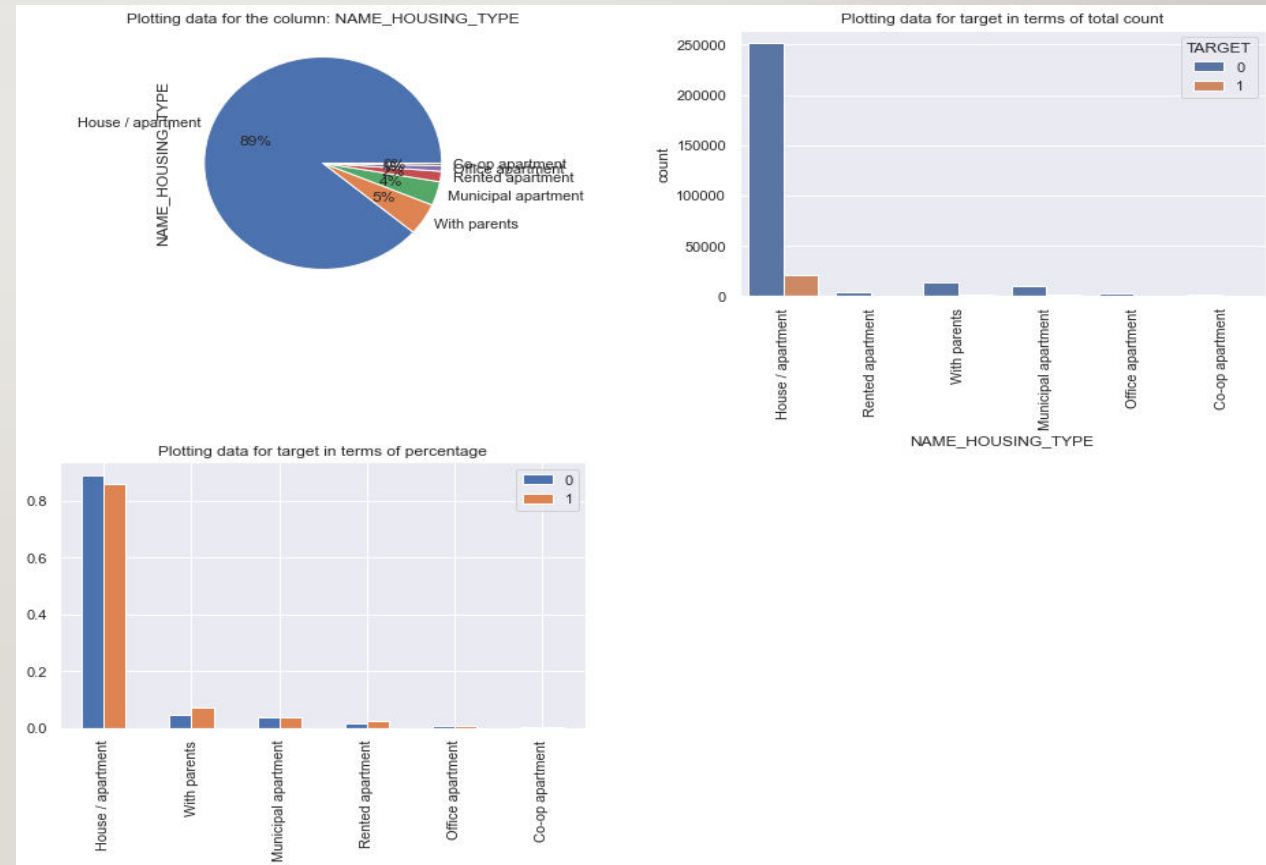
- ✓ Majority of people availing loans are married and also they are also good with loan repayment.
- ✓ Separated people are Equally likely to be defaulters as well as pay on time.



DISTRIBUTION FOR NAME_HOUSING_TYPE

Points to be concluded from the graph on the right.

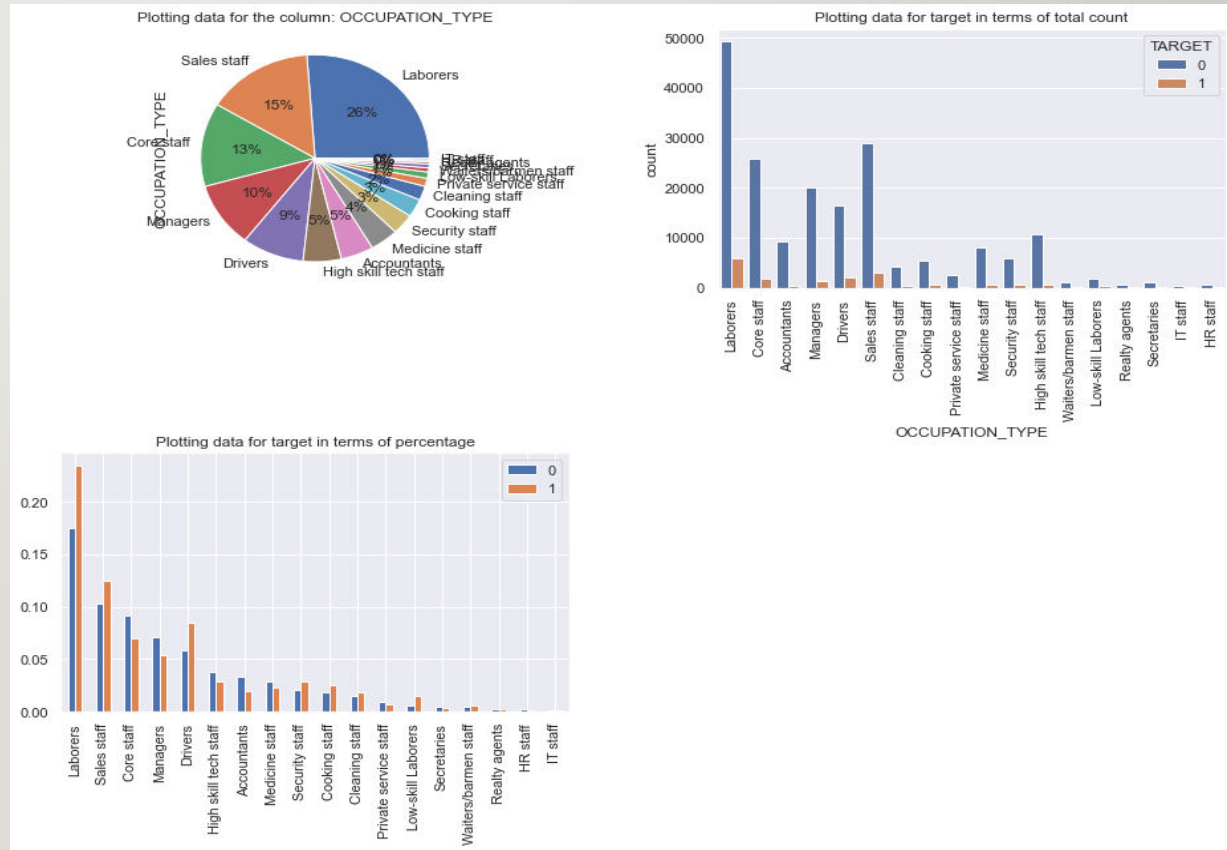
- ✓ The people who owns house/ apartments are more likely to avail loans.



DISTRIBUTION FOR OCCUPATION_TYPE

Points to be concluded from the graph on the right.

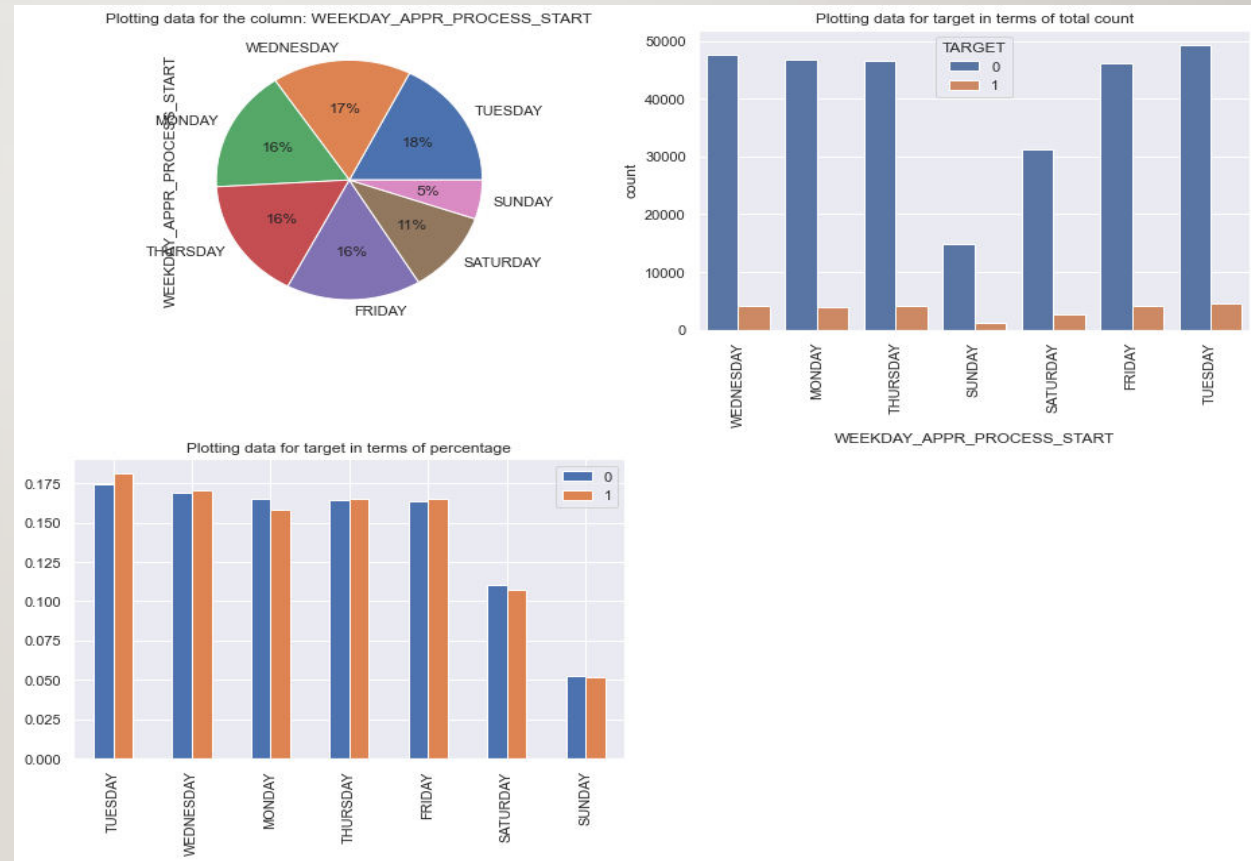
- ✓ The people who have their occupation type as 'labourers' are availing more loans
- ✓ However, they are also more likely to be defaulter.



DISTRIBUTION FOR WEEKDAYS_APPR_PROCESS_START

Points to be concluded from the graph on the right.

- ✓ Majority of clients apply for loans in weekdays.
- ✓ However, people who apply for loans on Sunday are less likely to be defaulters.

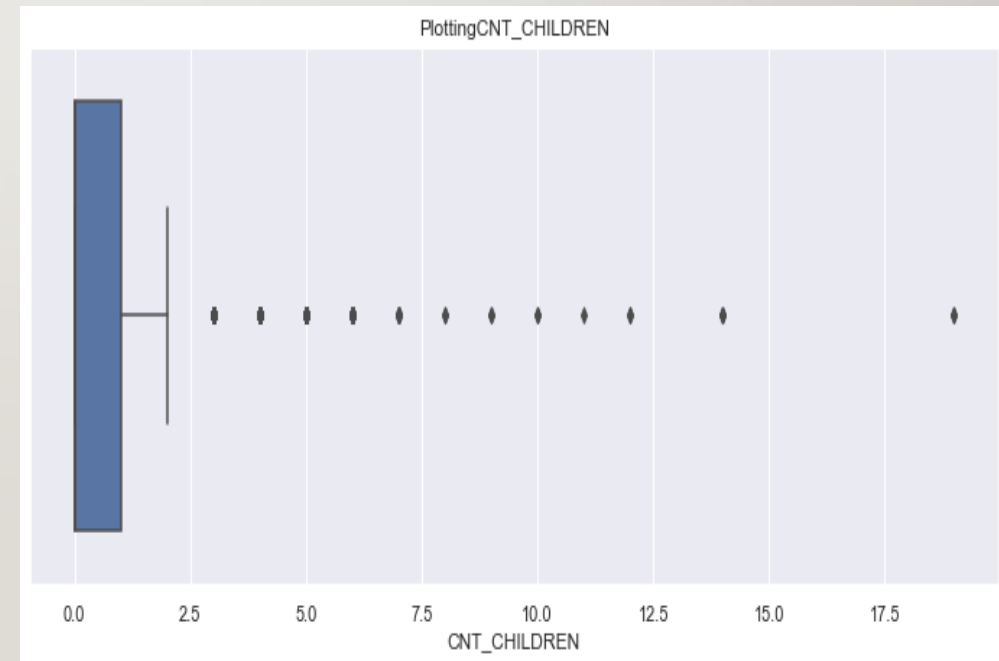
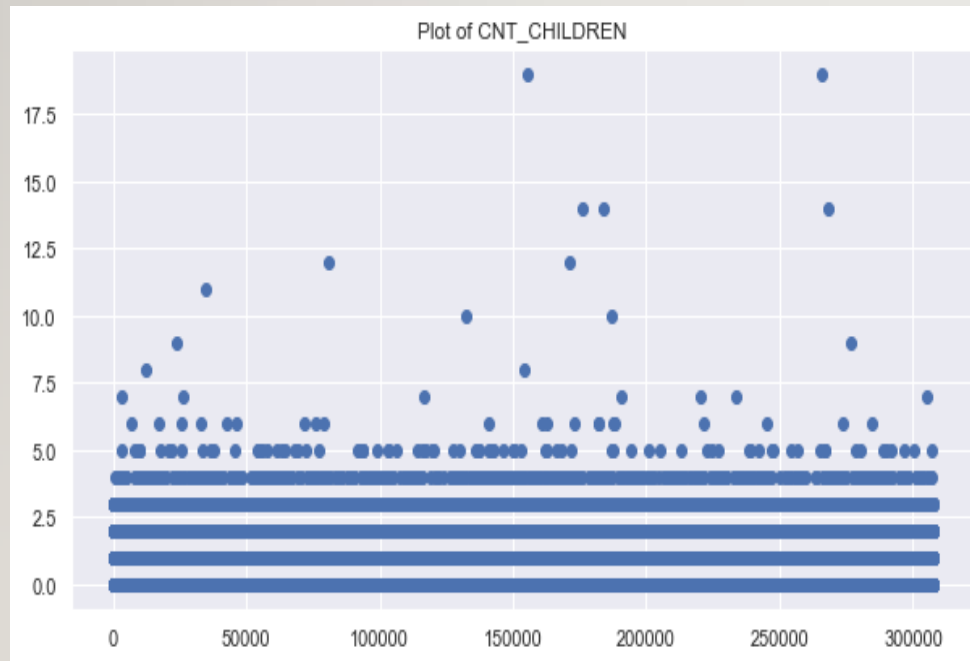


ANALYSIS FOR THE OUTLIERS

SCATTER PLOT & BOX PLOT CNT_CHILDREN

Below points can be concluded from the graph:

- There are two cases where no of children is 19 and there are 3 cases where no of children is 14 which seems very illogical ,so we believe these are outliers and should be treated.

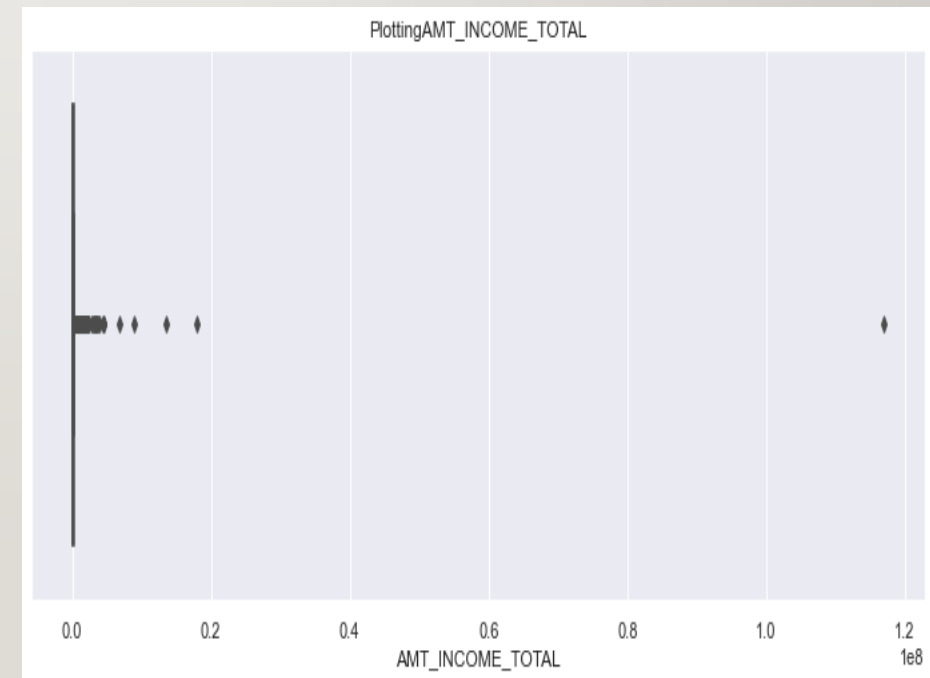
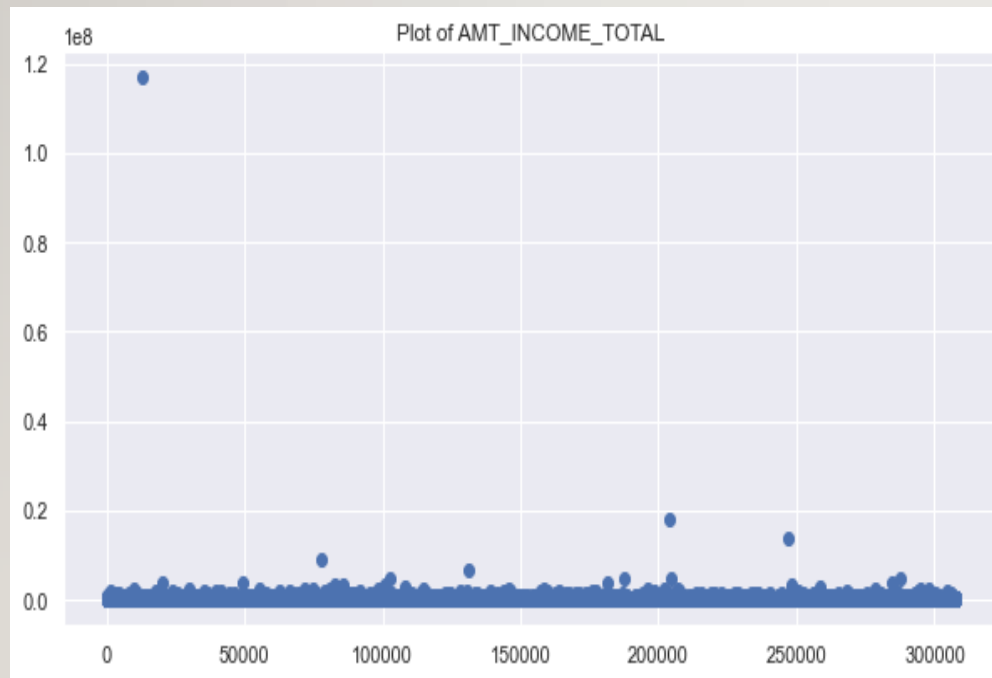


SCATTER PLOT & BOX PLOT AMT_INCOME_TOTAL

Below points can be concluded from the graph:

- As max amount_income is way above the mean and 75th quartile range. Hence, it is an outlier.

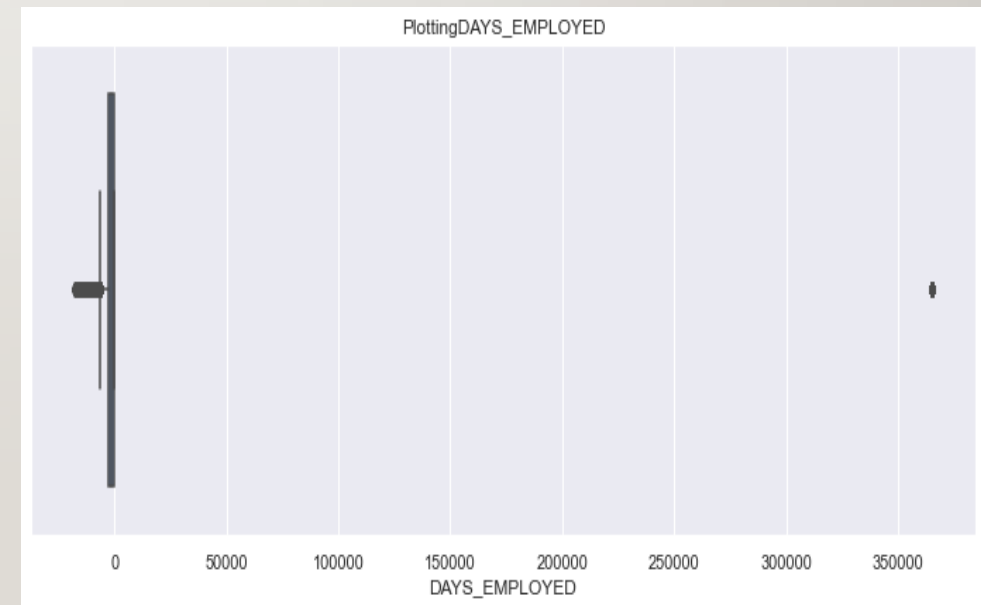
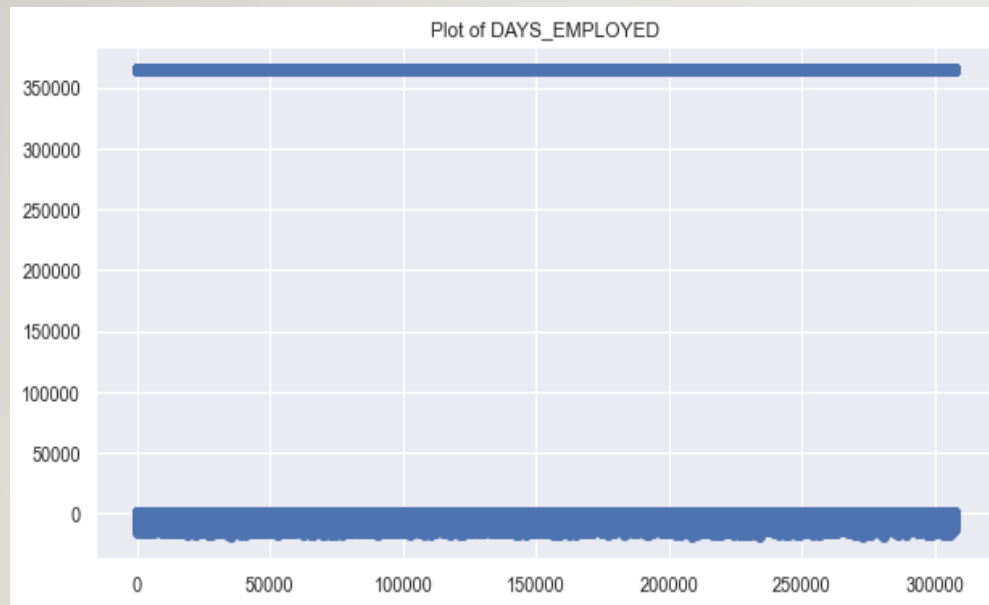
As this data point may influence the data, it should be treated.



SCATTER PLOT & BOX PLOT DAYS_EMPLOYED

Below points can be concluded from the graph:

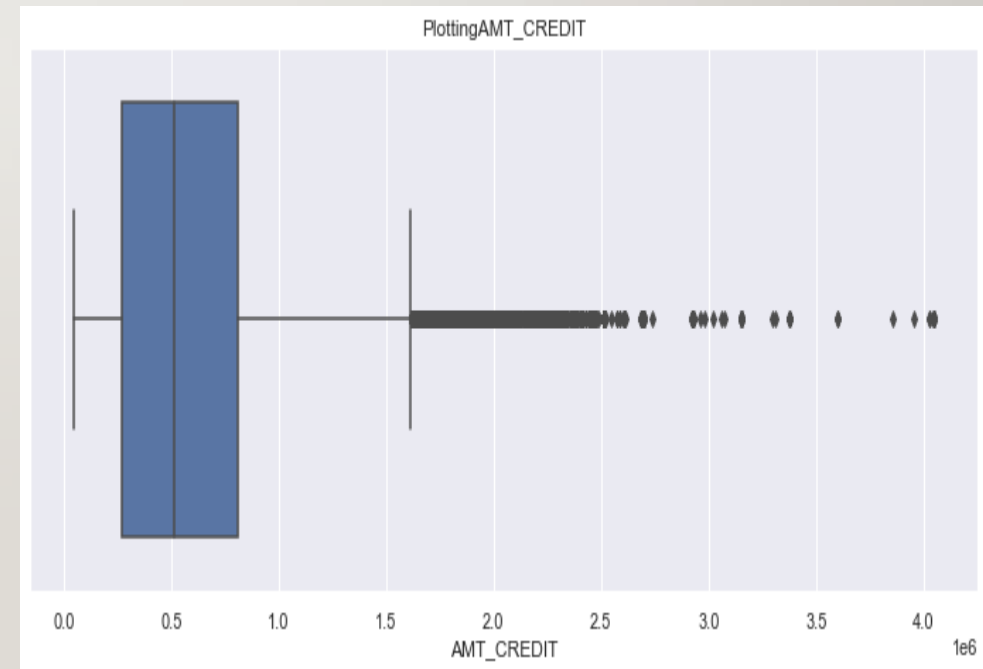
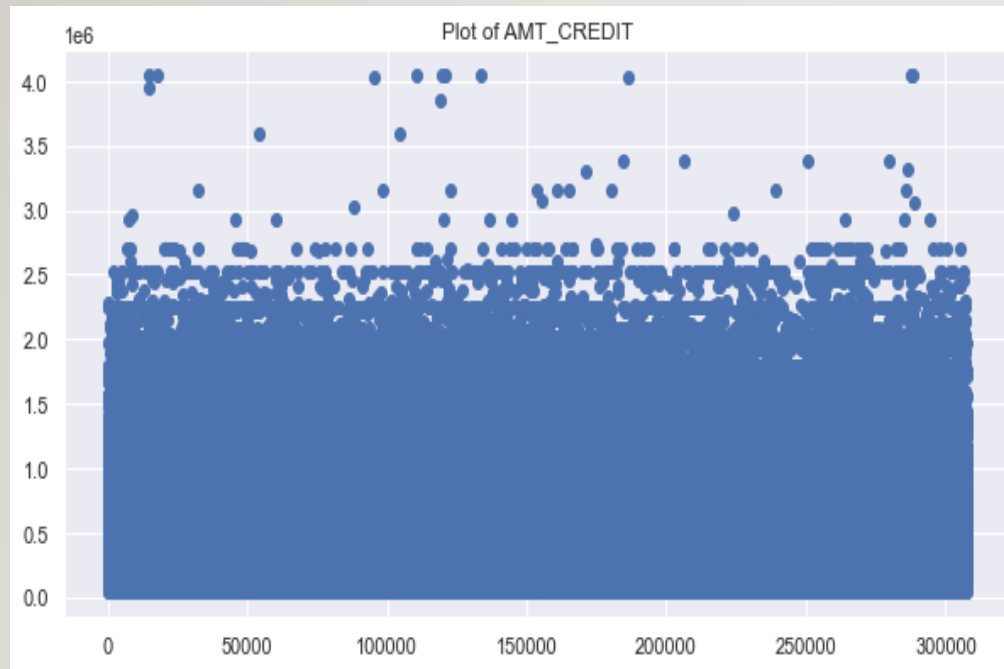
- there is one value which is present here 365243. It might be due to manual error while data entry plus all the values in this column are negative except for this value, hence we will treat this as an outlier.



SCATTER PLOT & BOX PLOT AMT_CREDIT

Below points can be concluded from the graph:

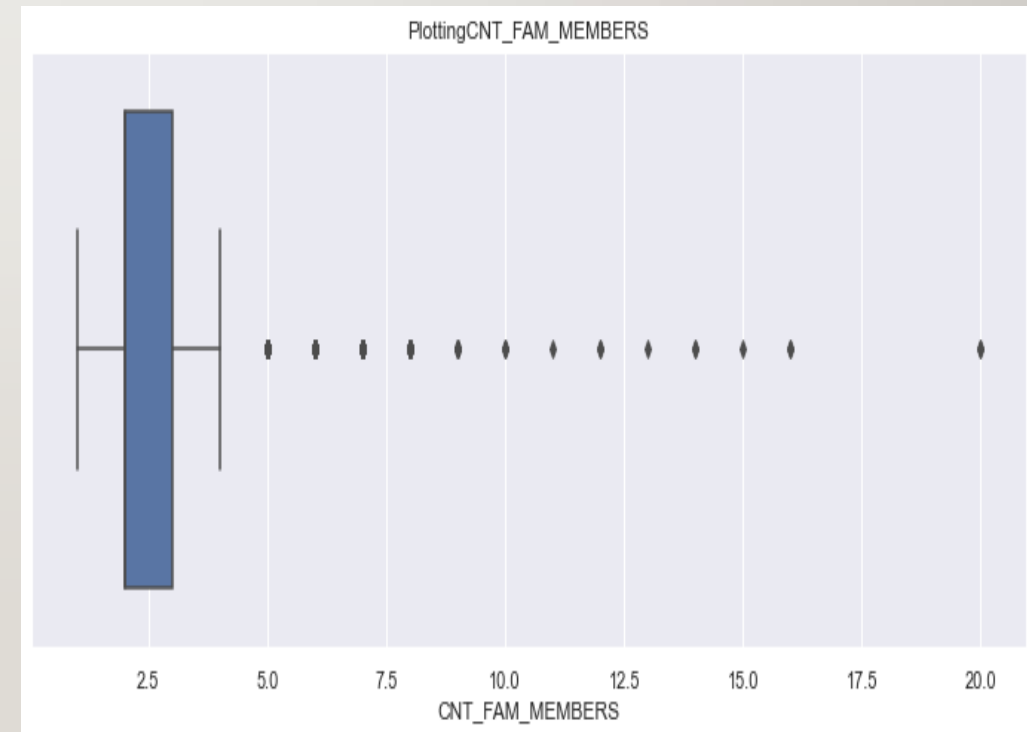
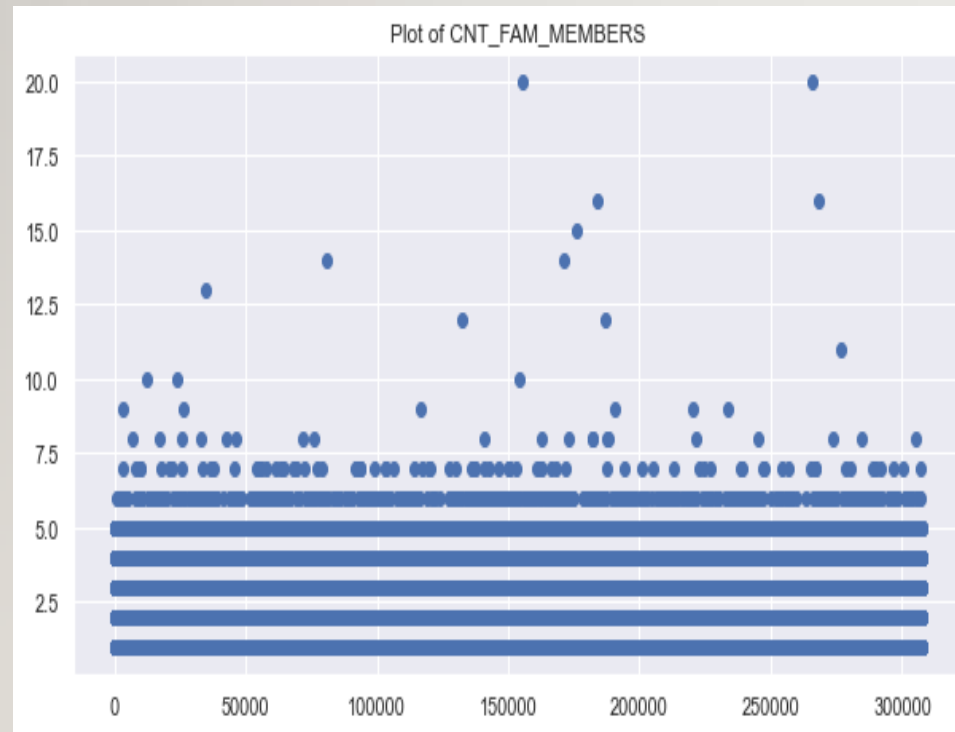
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile. Hence points greater than 3.5 may be treated as outliers.



SCATTER PLOT & BOX PLOT CNT_FAM_MEMBERS

Below points can be concluded from the graph:

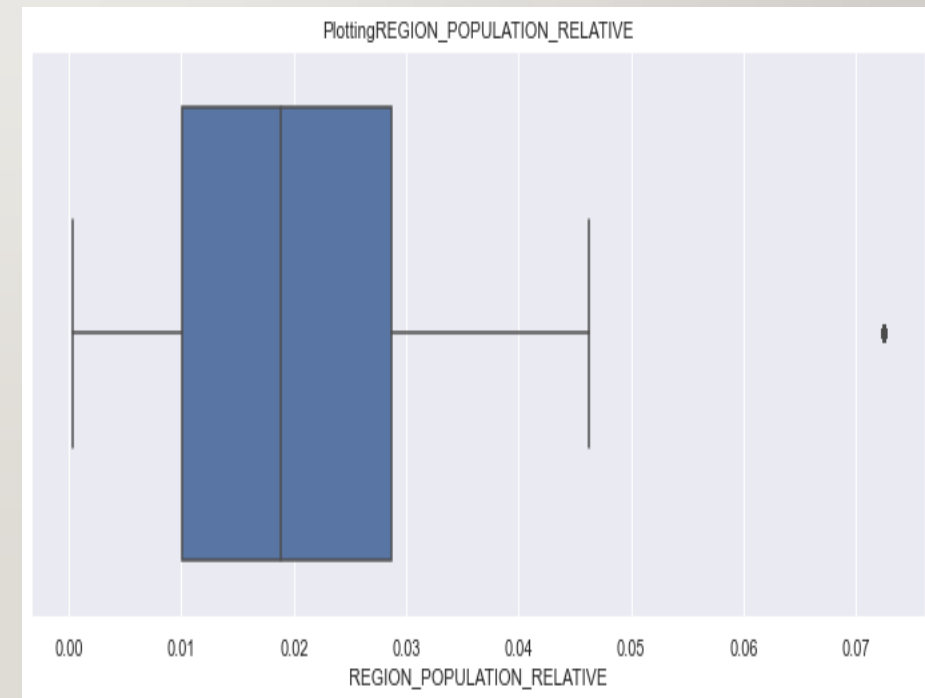
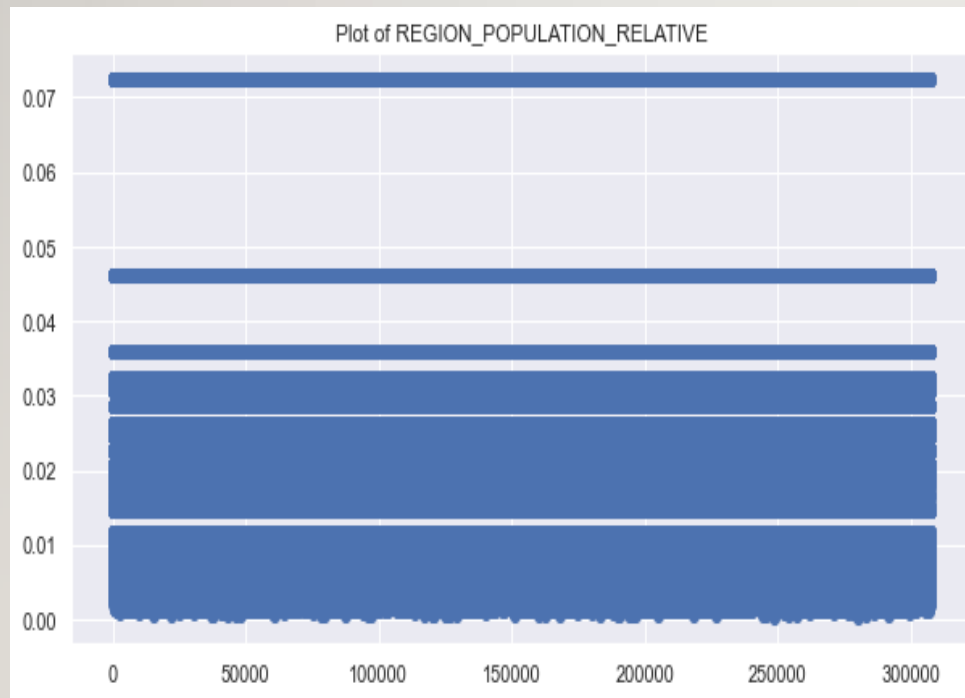
- Now a days it is impossible to have a family of 20 members so we think this is also an outliers



SCATTER PLOT & BOX PLOT REGION_POPULATION_RELATIVE

Below points can be concluded from the graph:

- There are some outliers present in densely populated where loan amount is also high more than 75% quartile



CO- RELATION FOR TARGET 0 AND TARGET I

INSIGHT FROM CO-RELATION IN SAME DATA SET

Co-relations between numeric columns for Pays On Time

Top10 Correration who Pays On Time

REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950149
YEARS_BEGINEXPLUATATION_MEDI	YEARS_BEGINEXPLUATATION_MODE	0.962133
YEARS_BEGINEXPLUATATION_AVG	YEARS_BEGINEXPLUATATION_MODE	0.971086
FLOORSMAX_MODE	FLOORSMAX_AVG	0.986569
AMT_GOODS_PRICE	AMT_CREDIT	0.987022
FLOORSMAX_MEDI	FLOORSMAX_MODE	0.988955
YEARS_BEGINEXPLUATATION_MEDI	YEARS_BEGINEXPLUATATION_AVG	0.993594
FLOORSMAX_MEDI	FLOORSMAX_AVG	0.997253
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998510
FLAG_EMP_PHONE	DAYS_EMPLOYED	0.999758

dtype: float64

INSIGHT FROM CO-RELATION IN SAME DATA SET

Co-relations between numeric columns for Defaulters

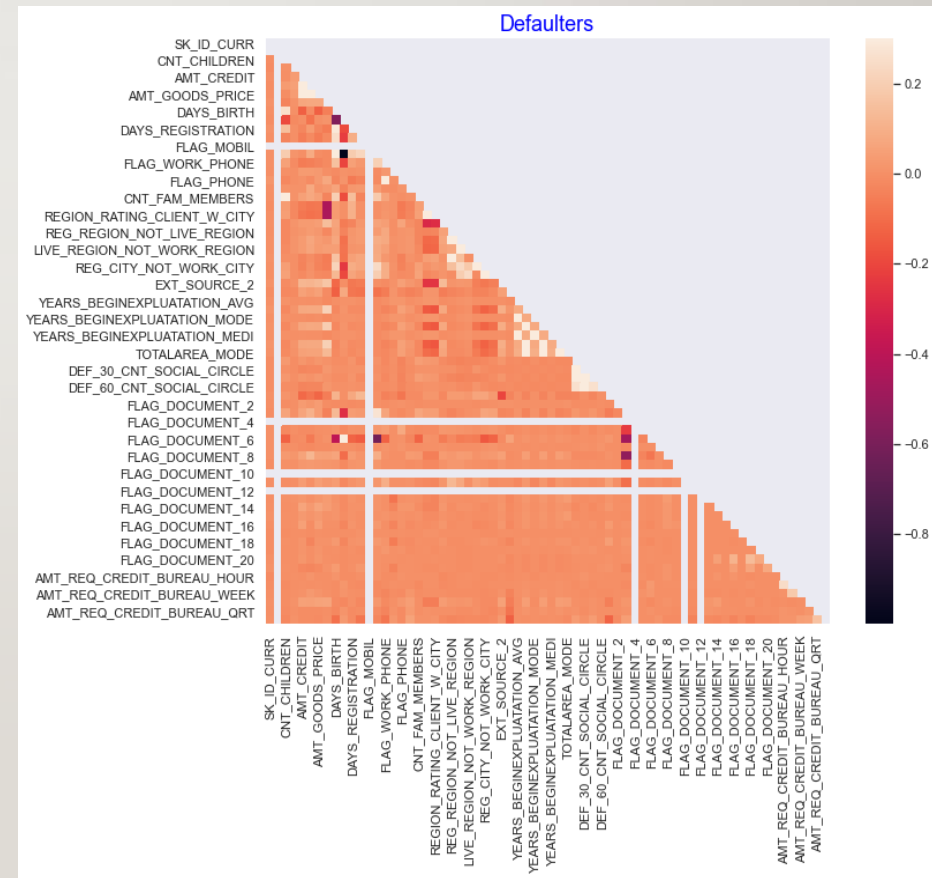
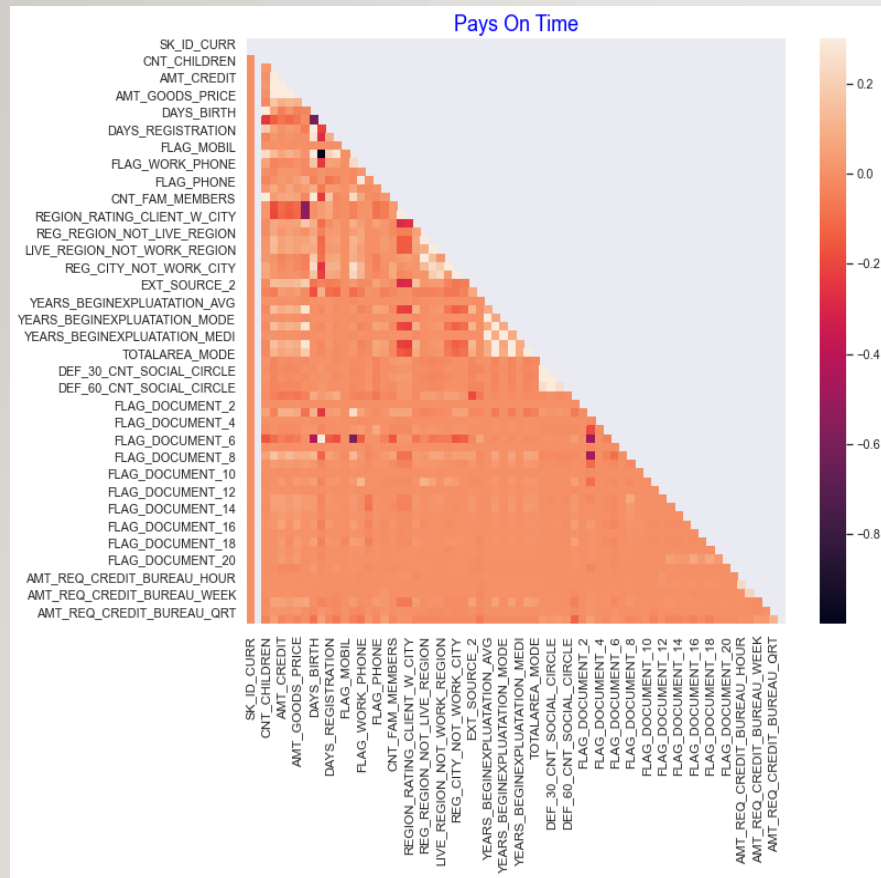
Top10 Correlation who don't Pays On Time

REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956637
YEARS_BEGINEXPLUATATION_MEDI	YEARS_BEGINEXPLUATATION_MODE	0.978163
YEARS_BEGINEXPLUATATION_AVG	YEARS_BEGINEXPLUATATION_MODE	0.980546
AMT_GOODS_PRICE	AMT_CREDIT	0.982783
FLOORSMAX_MODE	FLOORSMAX_AVG	0.986935
	FLOORSMAX_MEDI	0.989472
YEARS_BEGINEXPLUATATION_AVG	YEARS_BEGINEXPLUATATION_MEDI	0.996139
FLOORSMAX_AVG	FLOORSMAX_MEDI	0.997295
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998270
DAYS_EMPLOYED	FLAG_EMP_PHONE	0.999702

dtype: float64

INSIGHT FROM CO-RELATION IN SAME DATA SET

Co-relation Heatmap



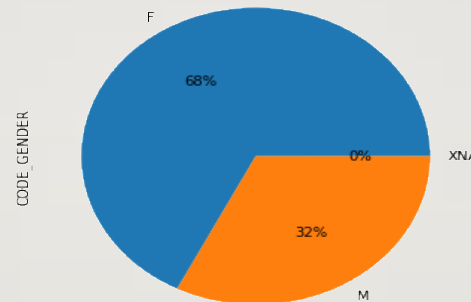
BIVARIATE ANALYSIS

CODE_GENDER v/s NAME_CONTRACT_TYPE

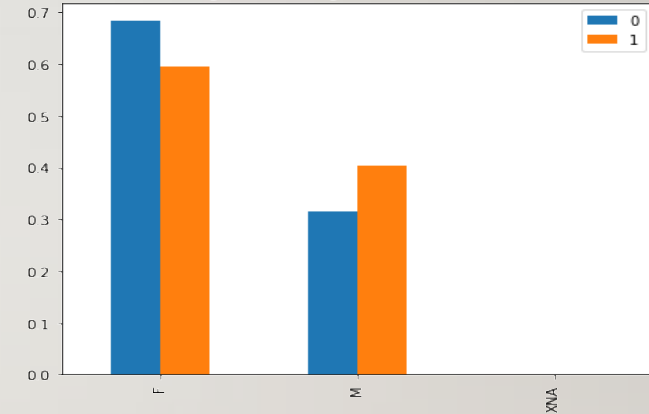
Points to be concluded from the graph on the right.

- More Females seems to avail cash loans as compared to revolving loans and are less likely to be defaulter as compared to Males.

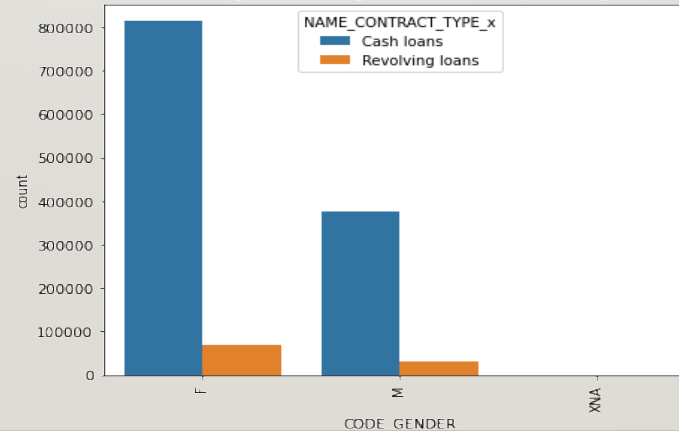
Plotting data for the column: CODE_GENDER



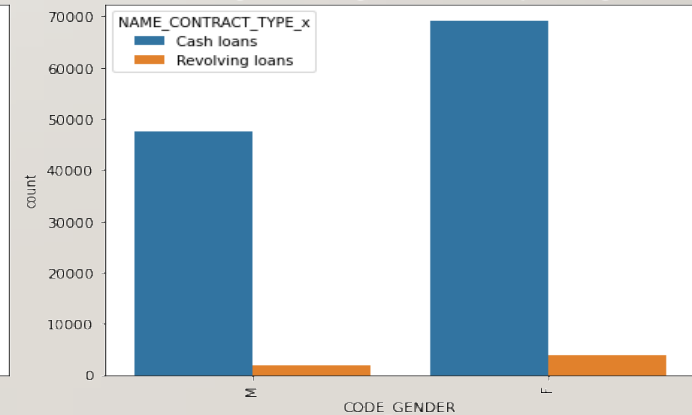
Plotting data for target in terms of total count



Plotting data for Target=0 in terms of percentage



Plotting data for Target=1 in terms of percentage

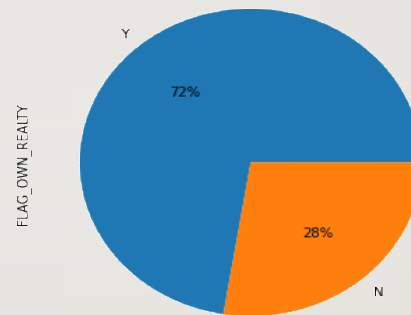


FLAG_OWN_REALTY v/s NAME_CONTACT_TYPE

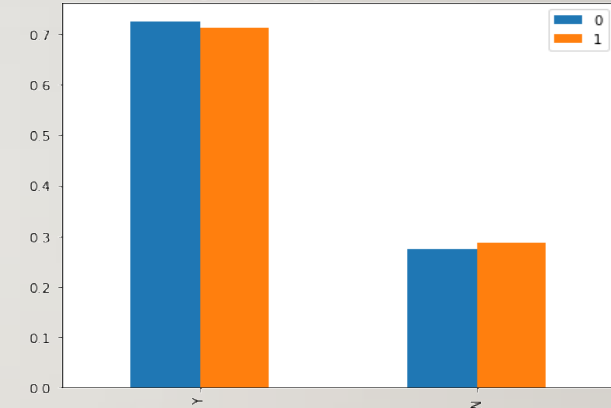
Points to be concluded from the graph on the right.

- People who own realty prefers cash loans more as compared to those who don't own realty and also are less likely to be defaulter.

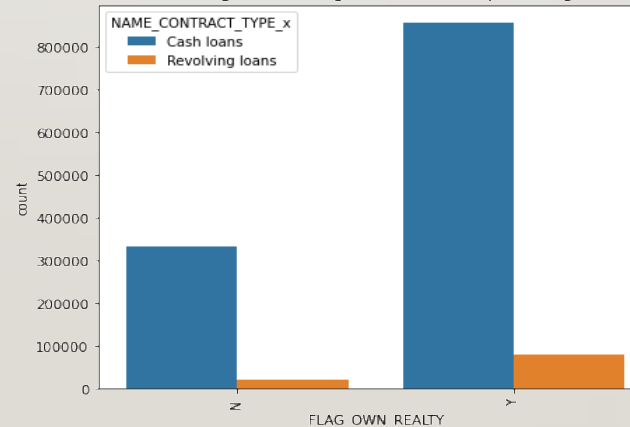
Plotting data for the column: FLAG_OWN_REALTY



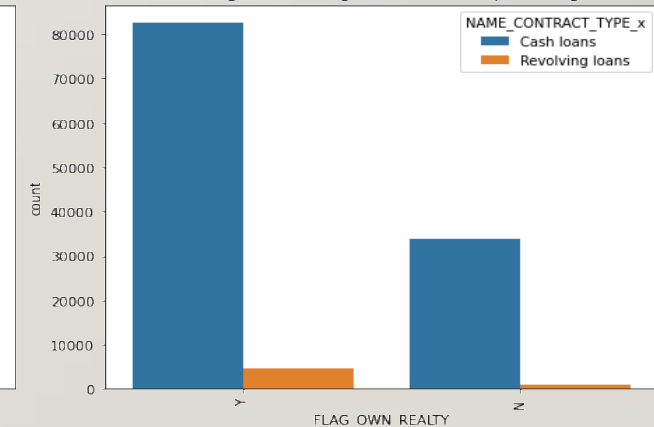
Plotting data for target in terms of total count



Plotting data for Target=0 in terms of percentage



Plotting data for Target=1 in terms of percentage



NAME_EDUCATION_TYPE v/s NAME_CONTACT_TYPE

Points to be concluded from the graph on the right.

- People with Secondary/secondary special education prefers cash loans over revolving loans.



NAME_FAMILY_STATUS v/s NAME_CONTACT_TYPE

Points to be concluded from the graph on the right.

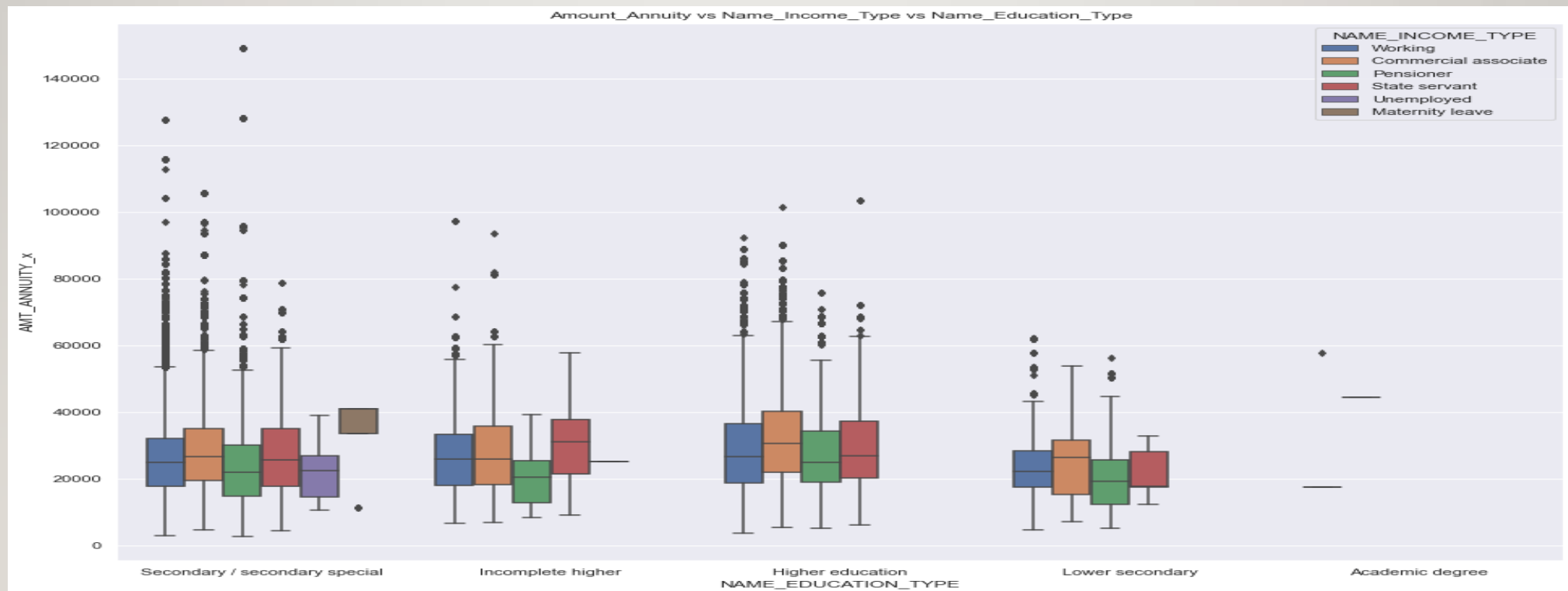
- Married clients tend to avail cash loans and are seems to have no difficulties in repaying it.



MULTI-VARIATE ANALYSIS

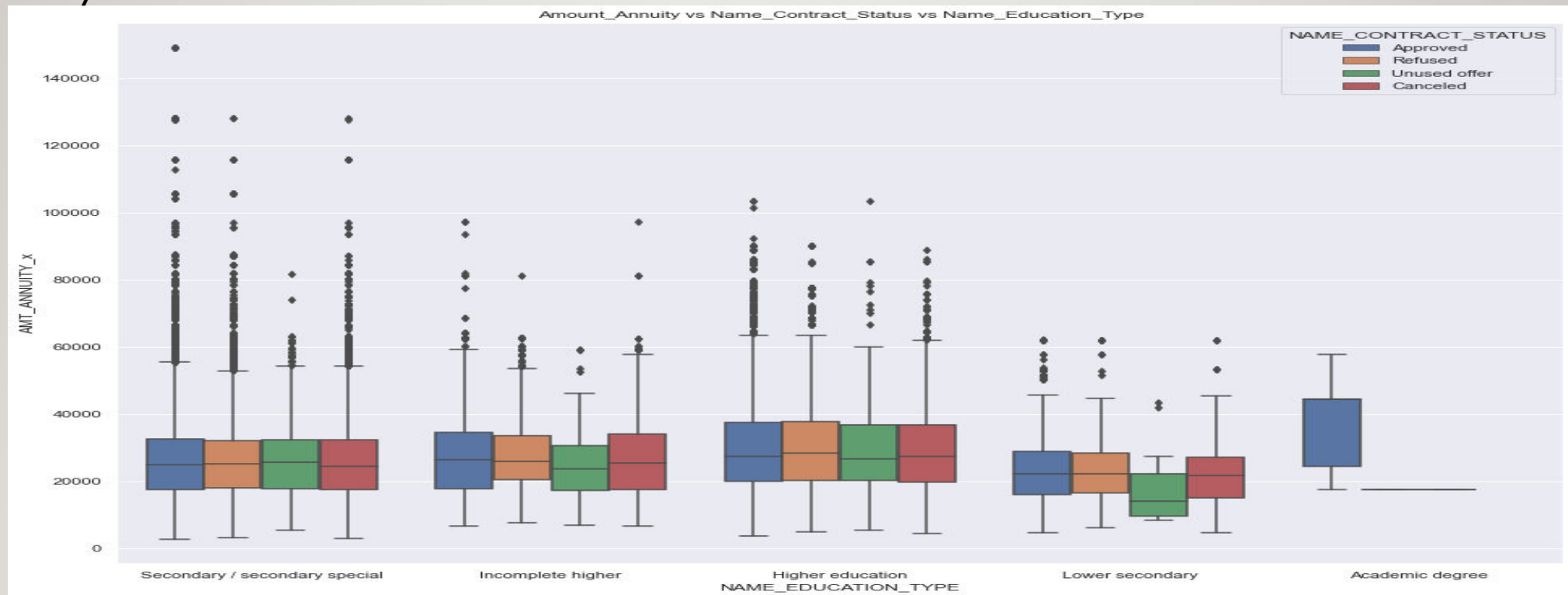
AMOUNT_ANNUITY vs NAME_INCOME_TYPE vs NAME_EDUCATION_TYPE

The clients having secondary education and are on maternity leave are seems to higher amount annuity. However, they are more likely to be defaulter.



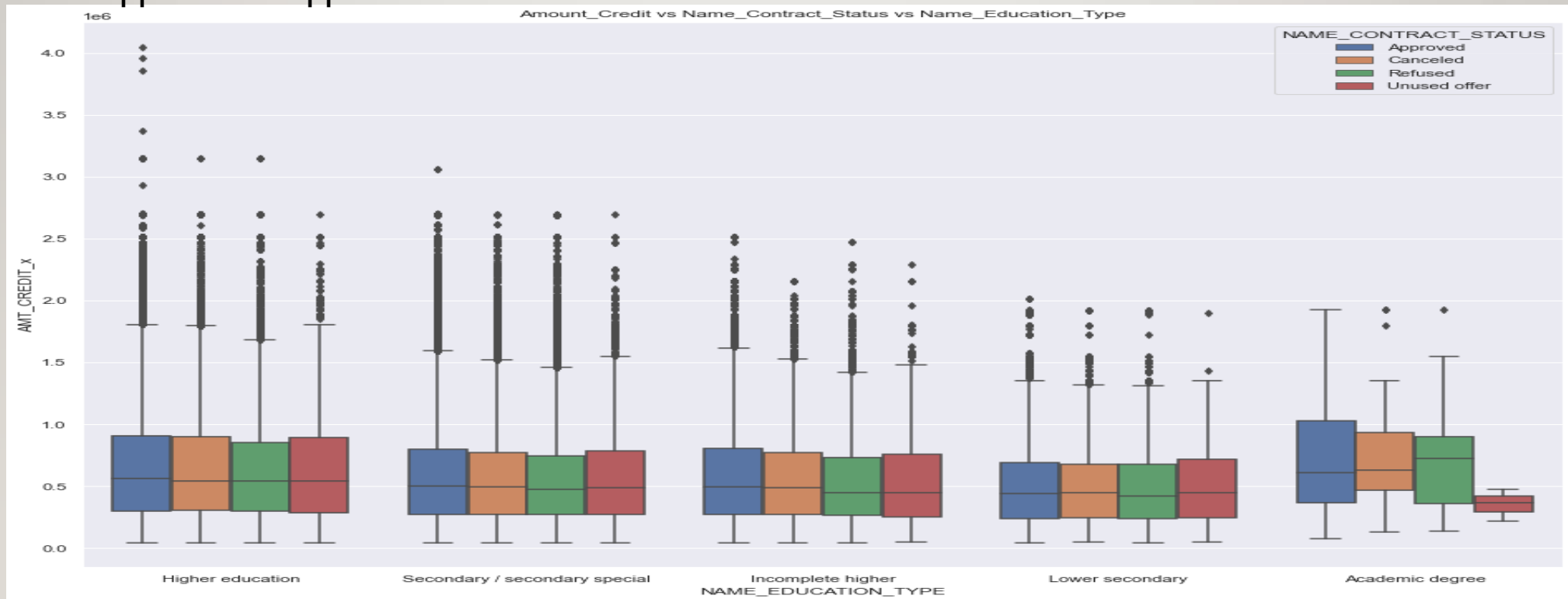
AMOUNT_ANNUITY vs NAME_CONTRACT_STATUS vs NAME_EDUCATION_TYPE

The people having lower secondary education tends to have lowest amount annuity and are more likely do not uses sanctioned loan.



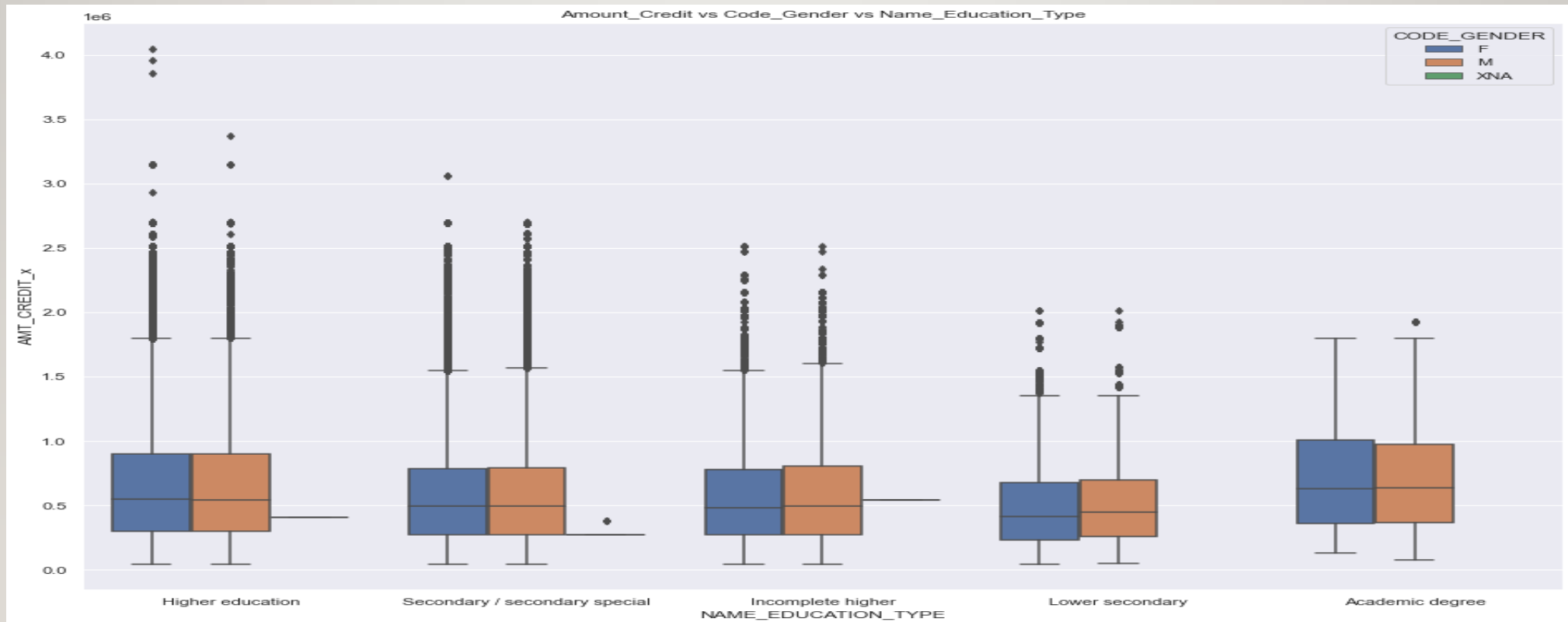
AMOUNT_CREDIT vs CODE_GENDER vs NAME_EDUCATION_TYPE

The people with academic degree tends to have highest amount credit and are more likely to get their applications approved.



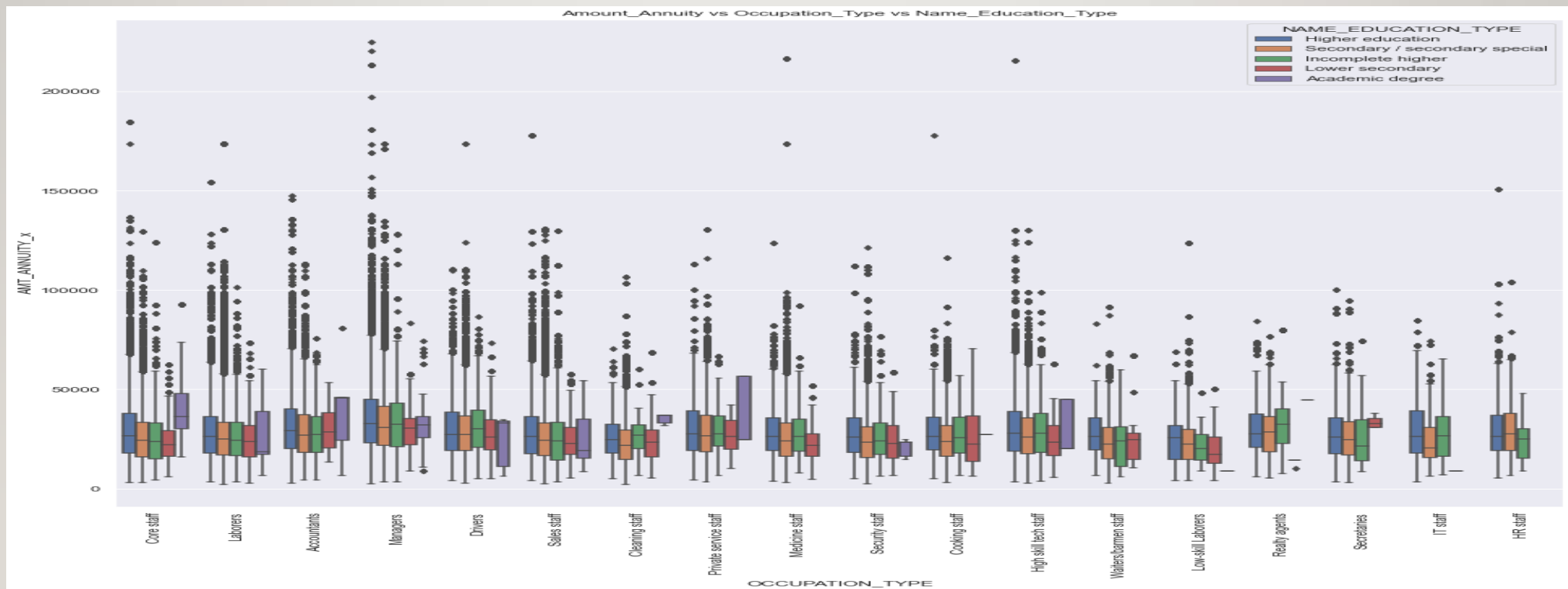
AMOUNT_CREDIT vs CODE_GENDER vs NAME_EDUCATION_TYPE

Females with academic degree have higher amount of credit and are less likely to be defaulter.



AMOUNT_ANNUIITY vs OCCUPATION_TYPE vs NAME_EDUCATION_TYPE

The private staff employees having academic degree seems to have highest amount annuity as compared to people with different occupation type.



SUMMARY

-
- ✓ More people are availing cash loans as compared to Revolving loans.They seems to have difficulties in repayment of loan as compared to those availing revolving loans
 - ✓ The clients availing loans are Females in majority. However,more males seems to be defaulter as compared to Females.
 - ✓ The clients who don't own cars tends to be threat of becoming defaulter.
 - ✓ Majority of people availing loans are from working category, however they are also more likely to be defaulter.
 - ✓ Majority of people availing loans are married and also they are also good with loan repayment.
 - ✓ The people who owns house/ apartments are more likely to avail loans.
 - ✓ The people who have their occupation type as labourers are availing more loans however,they are also more likely to be defaulter.
 - ✓ Majority of clients apply for loans in weekdays. However,people who apply for loans on sunday are less likely to be defaulters.
 - ✓ Banks should focus more on contract type 'Student' ,'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.
 - ✓ Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.
 - ✓ Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.
 - ✓ Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.

Many More...

THANK YOU