CREDIT EDA CASE STUDY

BANK LOAN DEFAULT ANALYSIS

SUBMITTED BY,

KRITHIGA K.

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 you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- 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 not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.
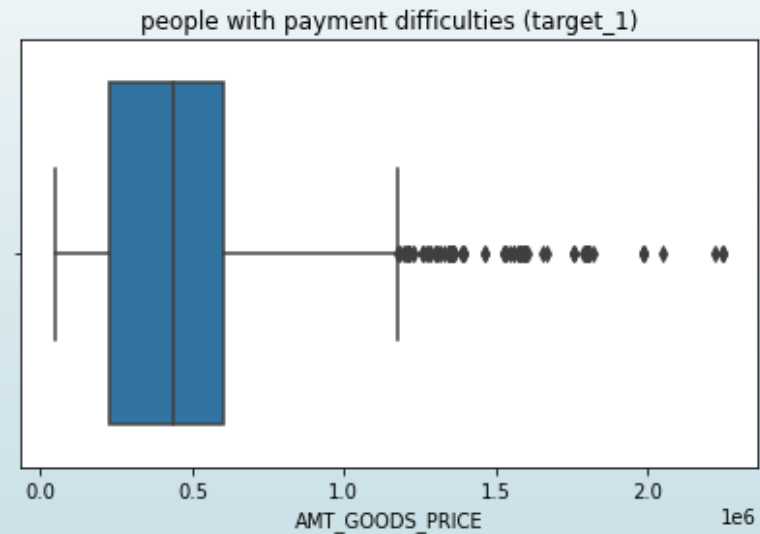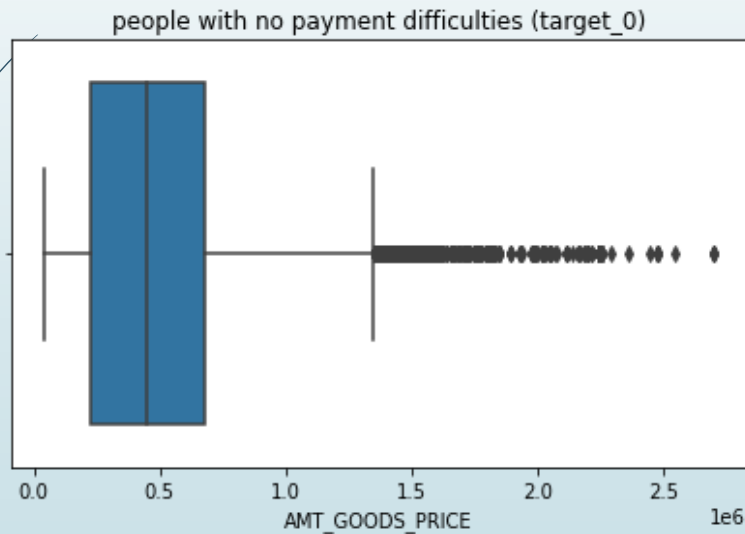
UNIVARIATE ANALYSIS :

- In both the target_0 and target_1, Females count are more than Males in AMT_INCOME_TOTAL.

- In target_0 the number of Males are very high as compare to the number of Males of target1.
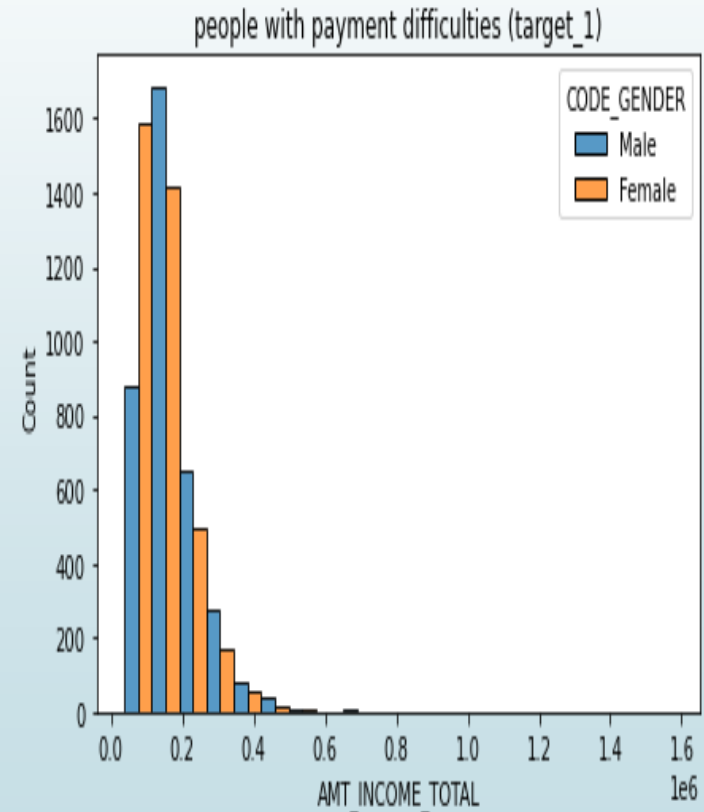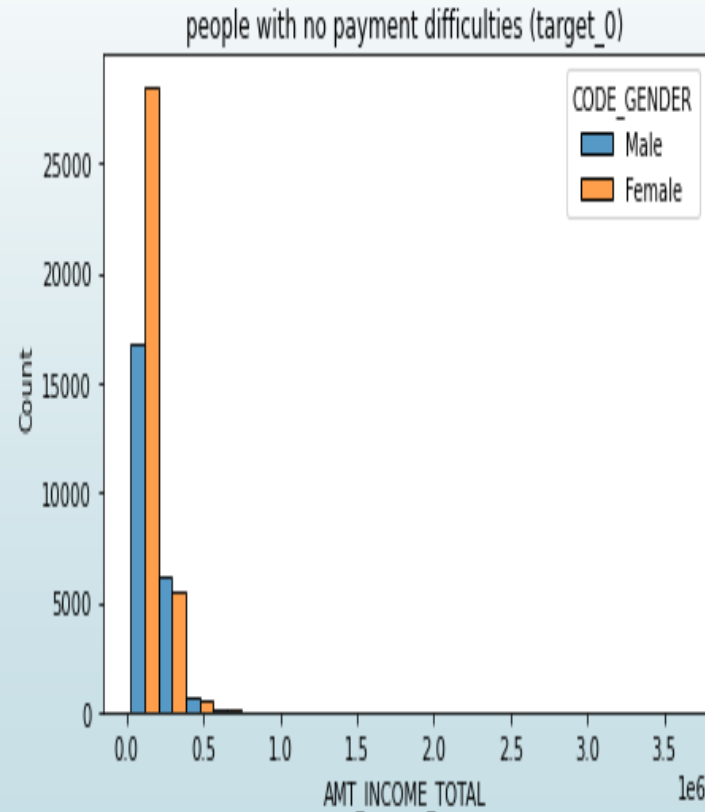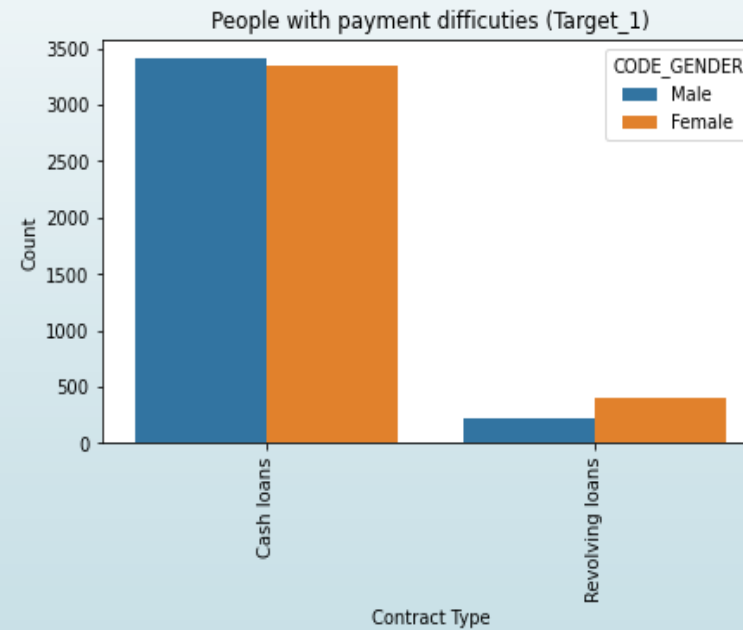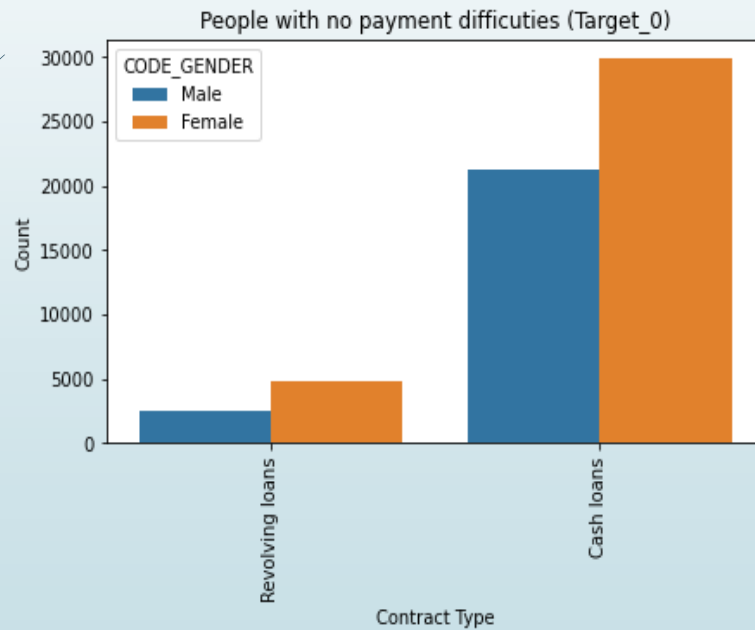
UNIVARIATE ANALYSIS :

- In both the target_0 and target_1 there are outliers in AMT_GOODS_PRICE.

- Where in target_1 we can see the third quartile is slightly bigger than other quartiles, and in target_0 the quartiles seems equal.
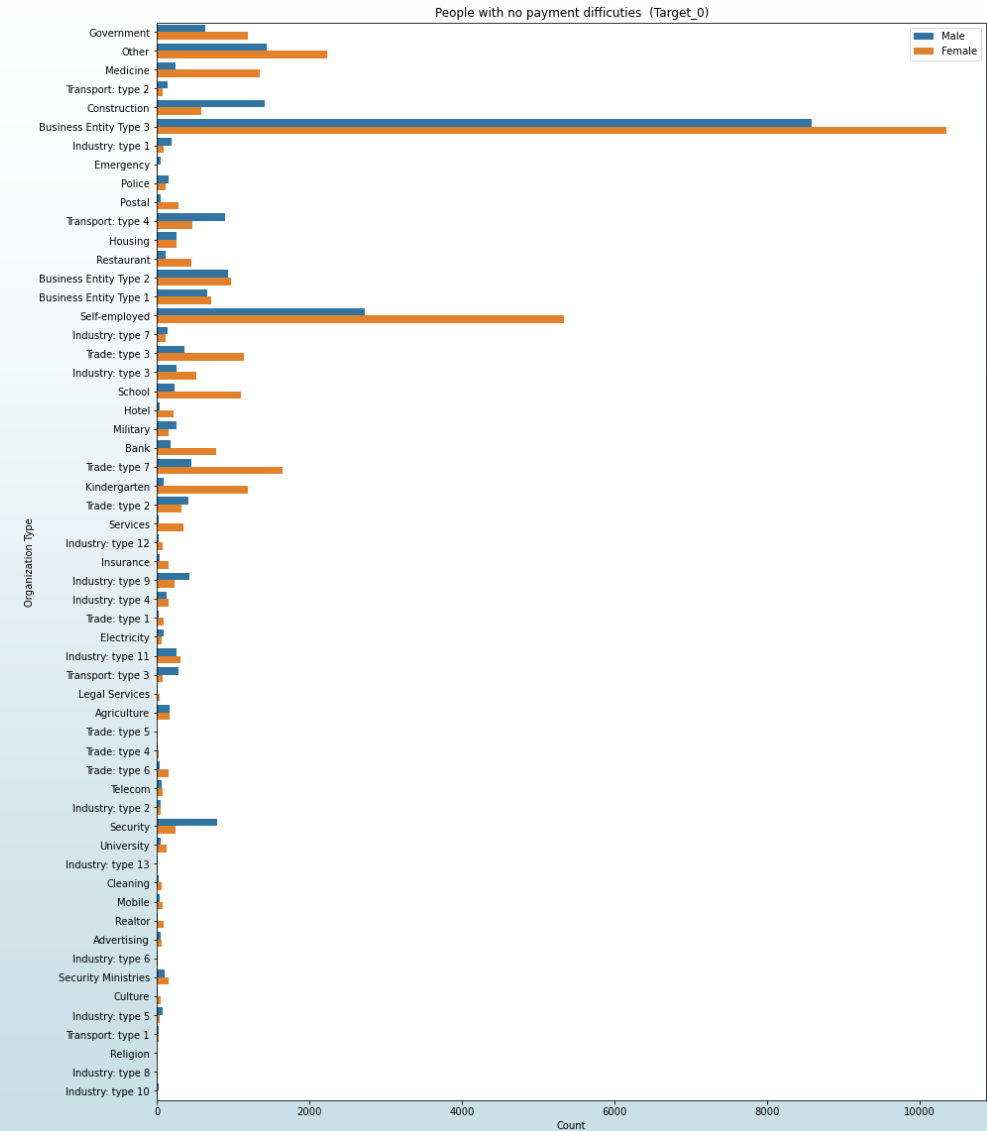
UNIVARIATE ANALYSIS :

- In both the target_0 and target_1 the requirement of Cash Loans are high as compare to Revolving Loans, regardless of the genders.

- Female Count is higher in both the target_0 and target_1 regardless of the contract type.

UNIVARIATE ANALYSIS :

- We have highest number of application from the Organization Type of Business Entity Type-3 in target_0.

- Lowest number of application from the Organization Type are Industry Type-6 and Industry Type-10 in target_0.

People with no payment difficuties (Target_0)

UNIVARIATE ANALYSIS :

- We have highest number of application from the Organization Type of Business Entity Type-3 in target1.

- Lowest number of application from the Organization Type are Cleaning and Industry Type-12 in target1.
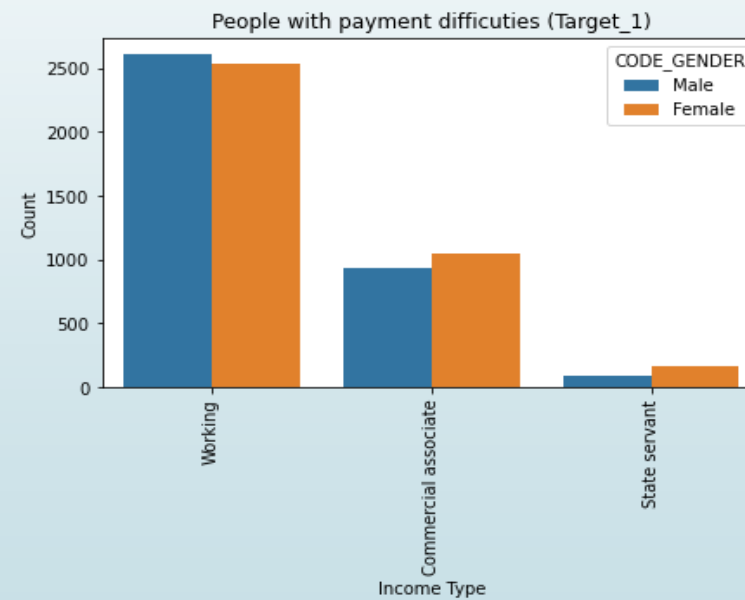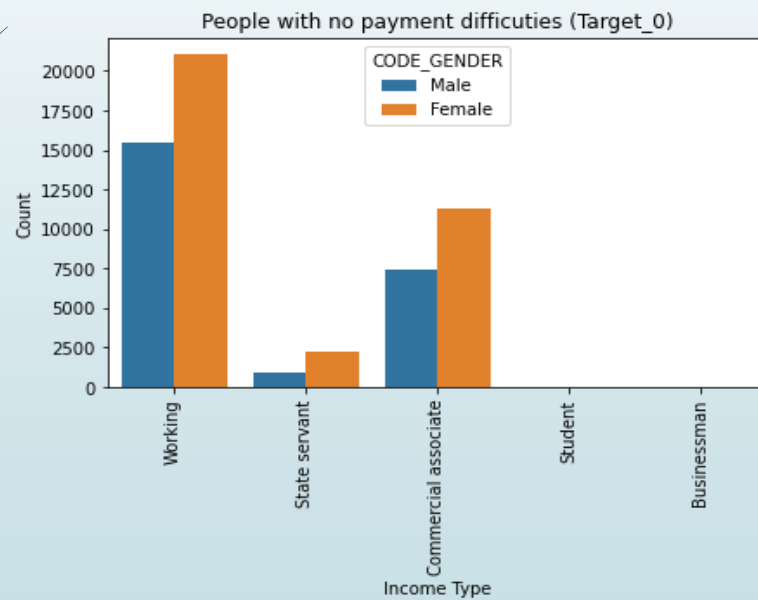

People with payment difficuties (Target_1)
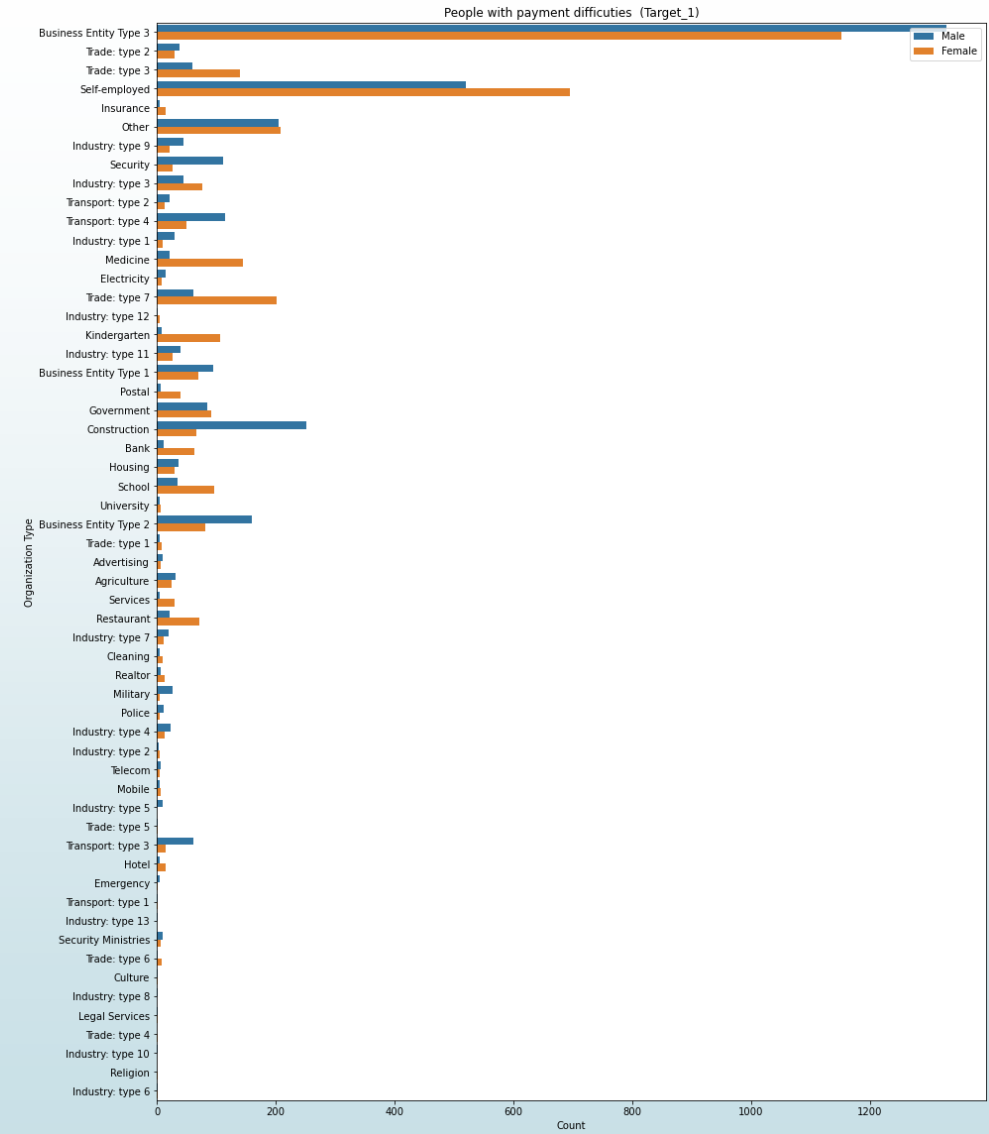
UNIVARIATE ANALYSIS :

- the median values of both the targets of Amount Annuity lies at 30000.
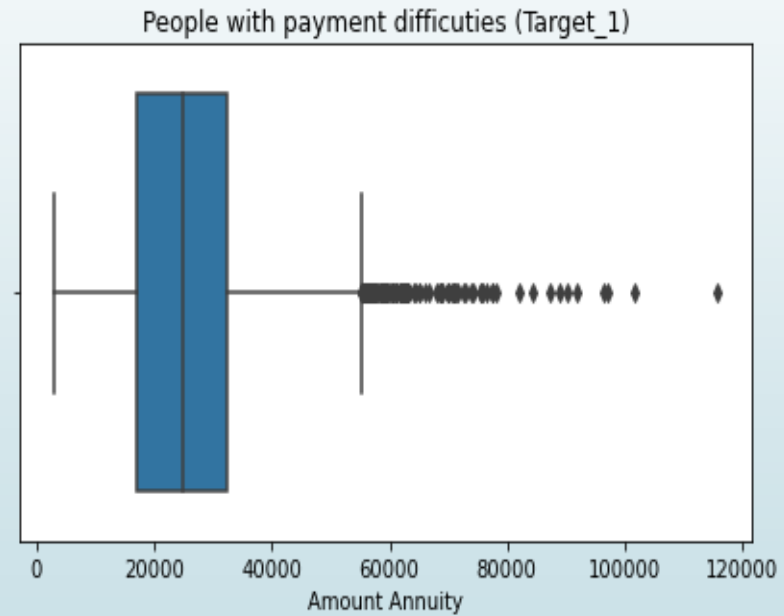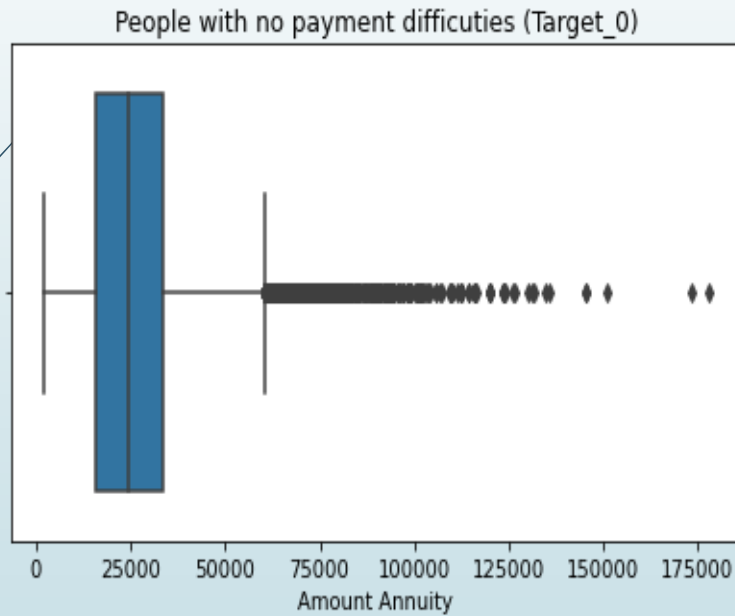
UNIVARIATE ANALYSIS :

- The Student, and Businessman are present in the income type of target_0 but they are absent in target_1.

- The number of Females is high as compared to Males in the income type of State Servant, and Commercial Associate irrespective of the targets.
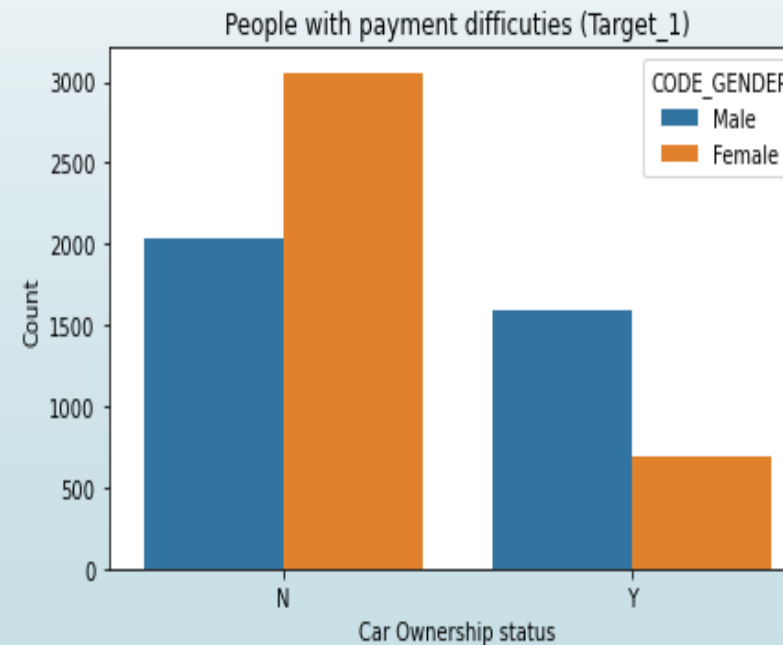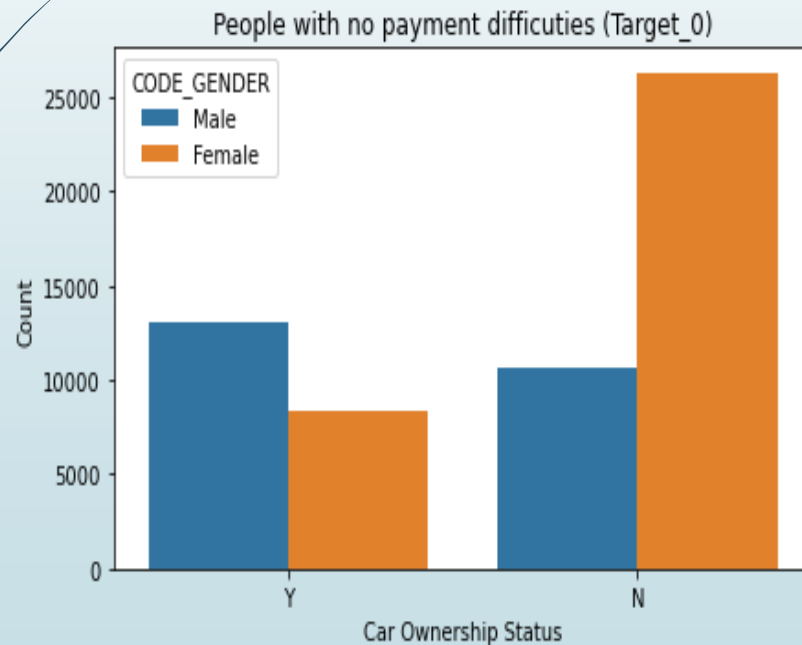
UNIVARIATE ANALYSIS :

- The number of people with payment difficulties (target_1) and non-payment difficulties (target_0) have Females in higher number as compared to Males.
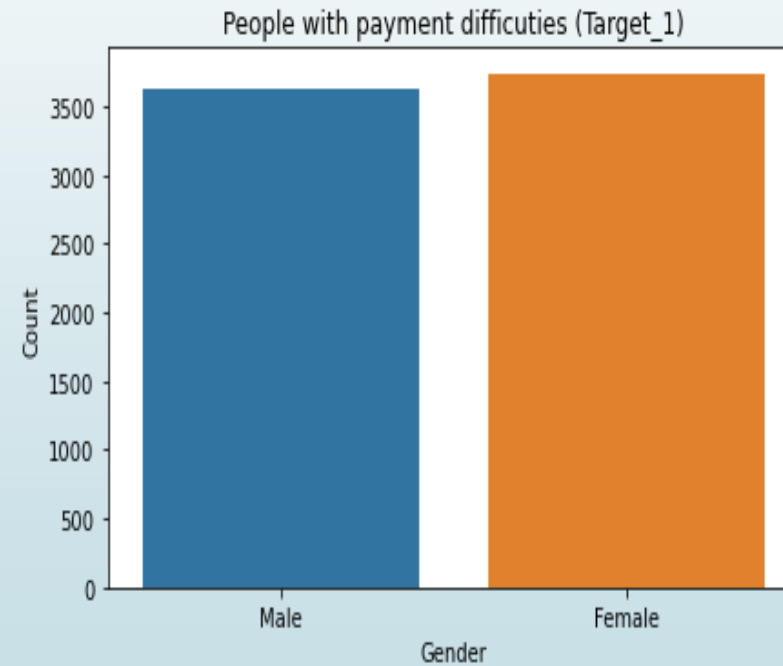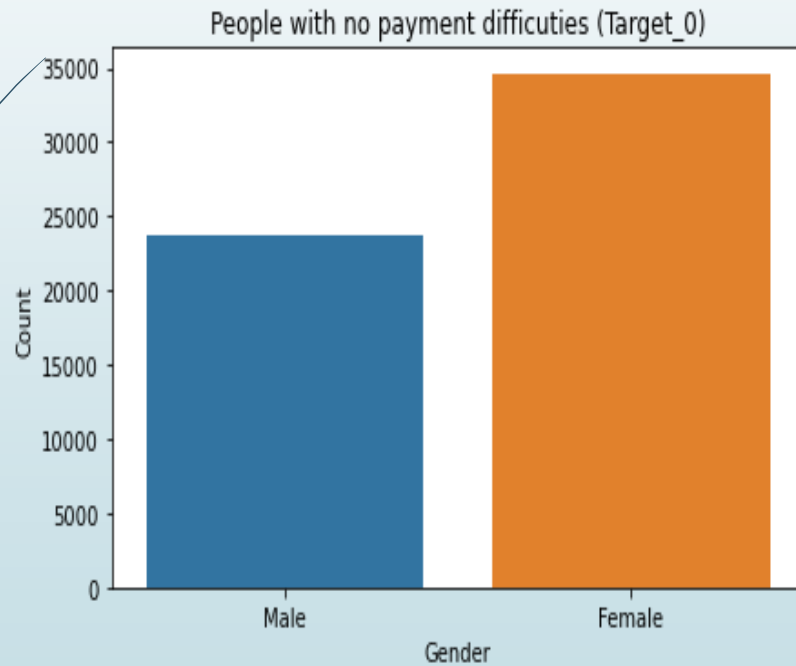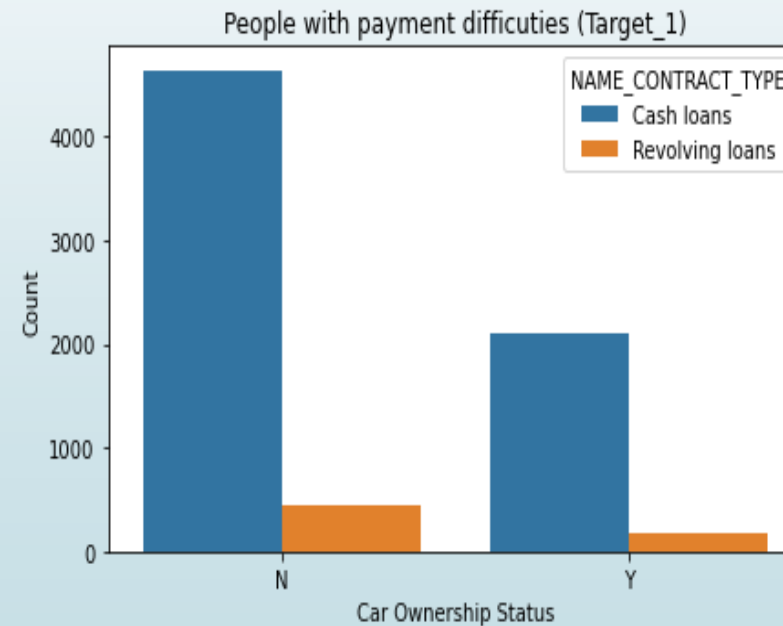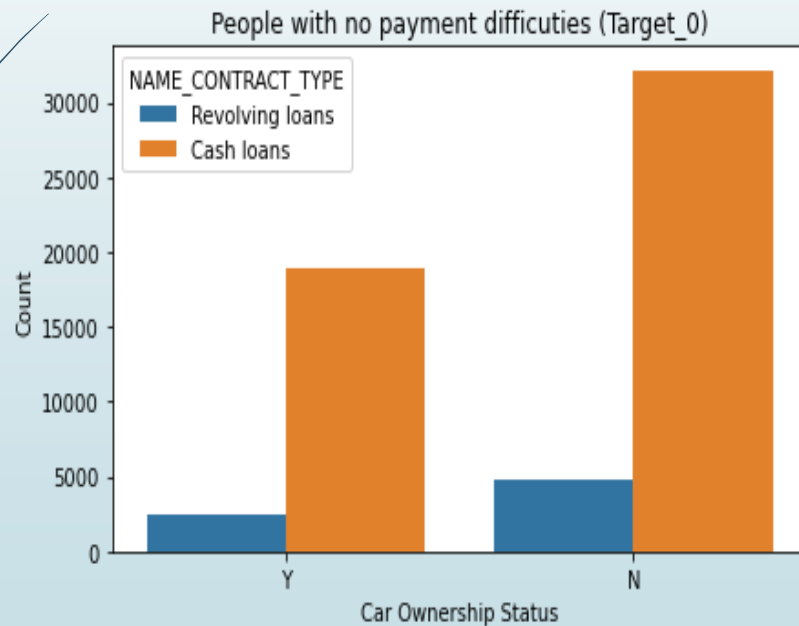
UNIVARIATE ANALYSIS :

- The Males have higher number of Ownership of Cars in both the target_0 and target_1 as compared to Females.

- the distribution of Non-Ownership of Cars shows maximum number of Females in both the target_0 and target_1.
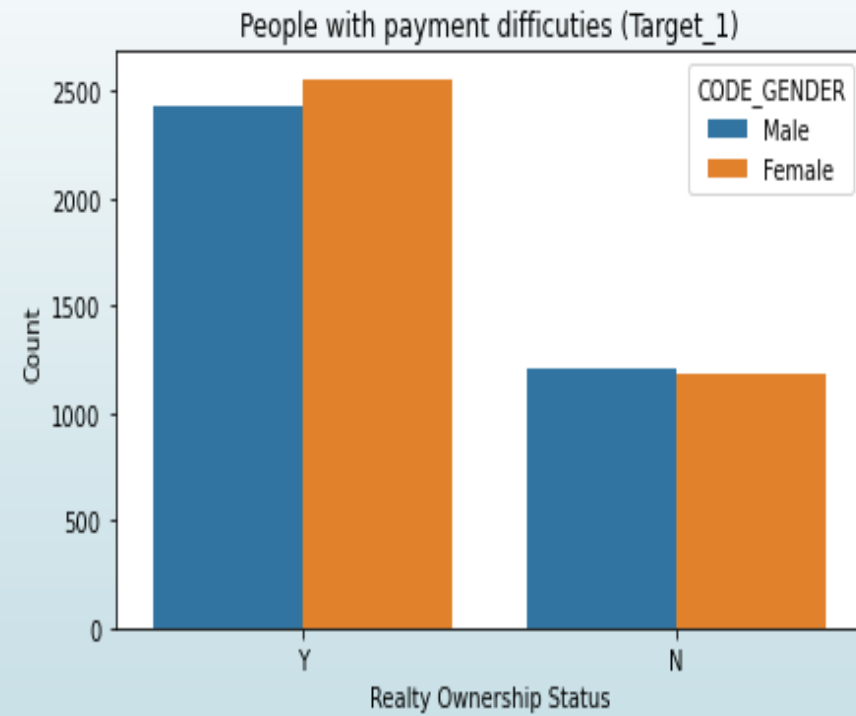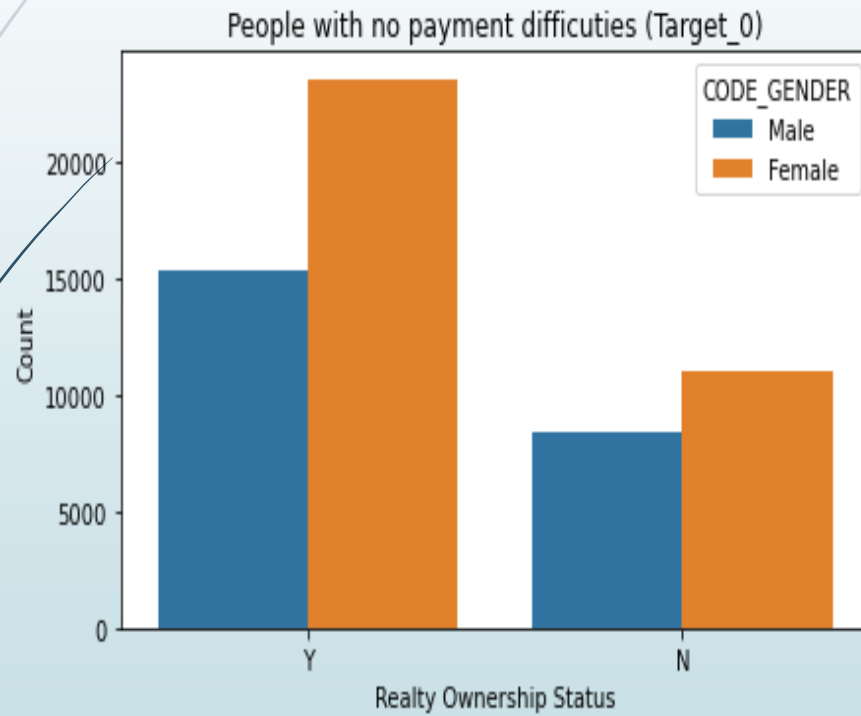
UNIVARIATE ANALYSIS :

- Those who have Ownership of Cars in both the target_0 and target_1 shows high application in Cash Loans.

- Also those who have Non-Ownership of Cars in both the target_0 and target_1 also shows high application interest in Cash Loans, but the number of applications for Cash Loans are higher for those who have no ownership of cars as compared to those who have the cars.
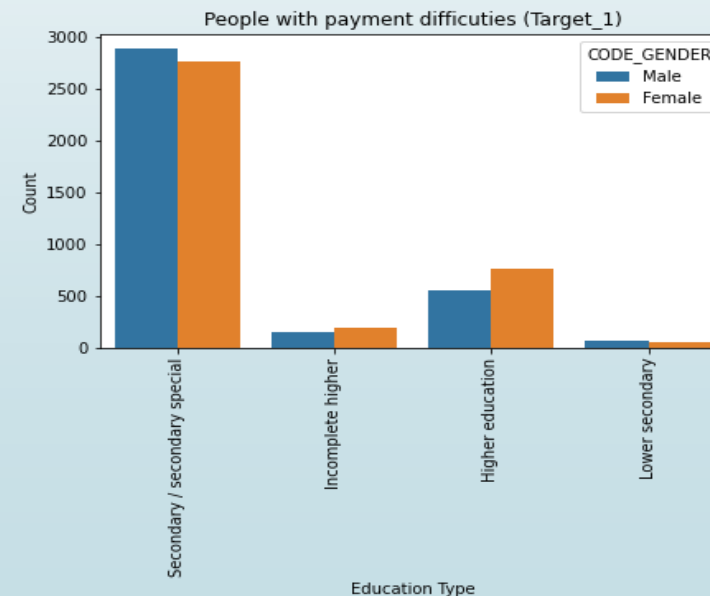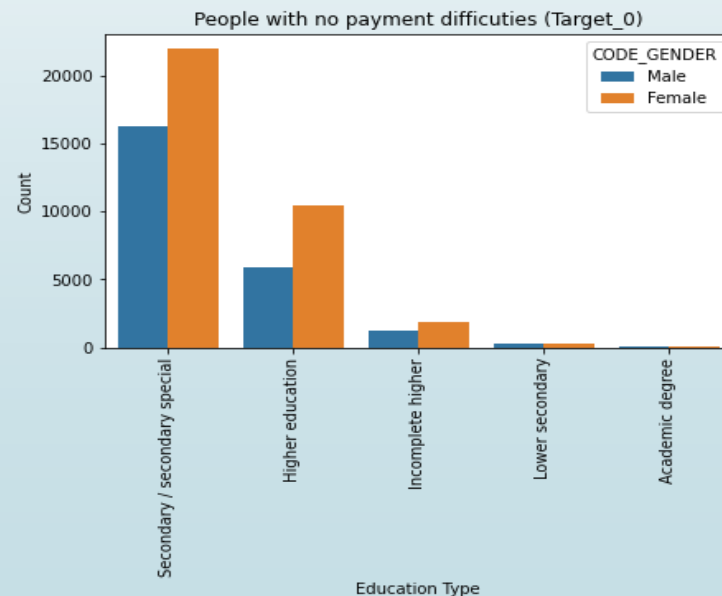
UNIVARIATE ANALYSIS :

- Realty Ownership in both the target_0 and target_1 have Females in higher number as compare to Males.
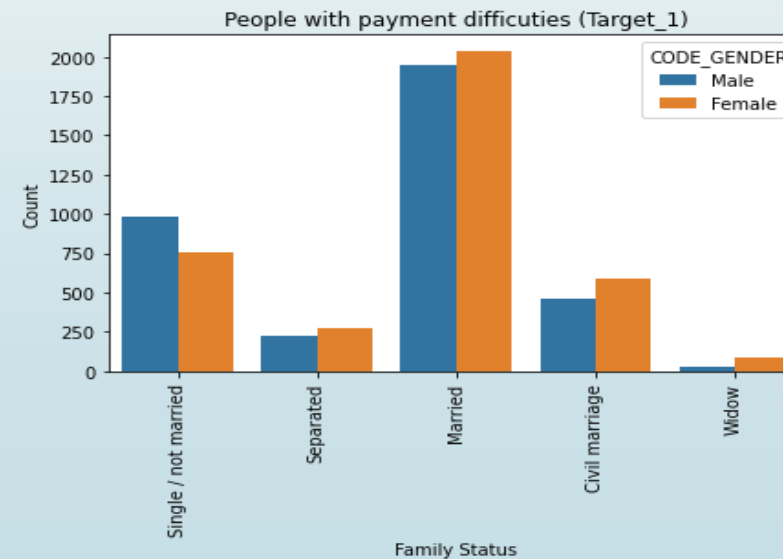
UNIVARIATE ANALYSIS :

- In both the target_0 and target_1 we have the highest number of application from Secondary Special.

- Distribution of Females in both the target_0 and target_1 is high in all types of Education.

- In target_0 we can see the distribution of Males are decreased as compared to Females in all types of Education.

- In target_1 we can see the distribution of Males and Females in all types of Education have only slight difference in application.
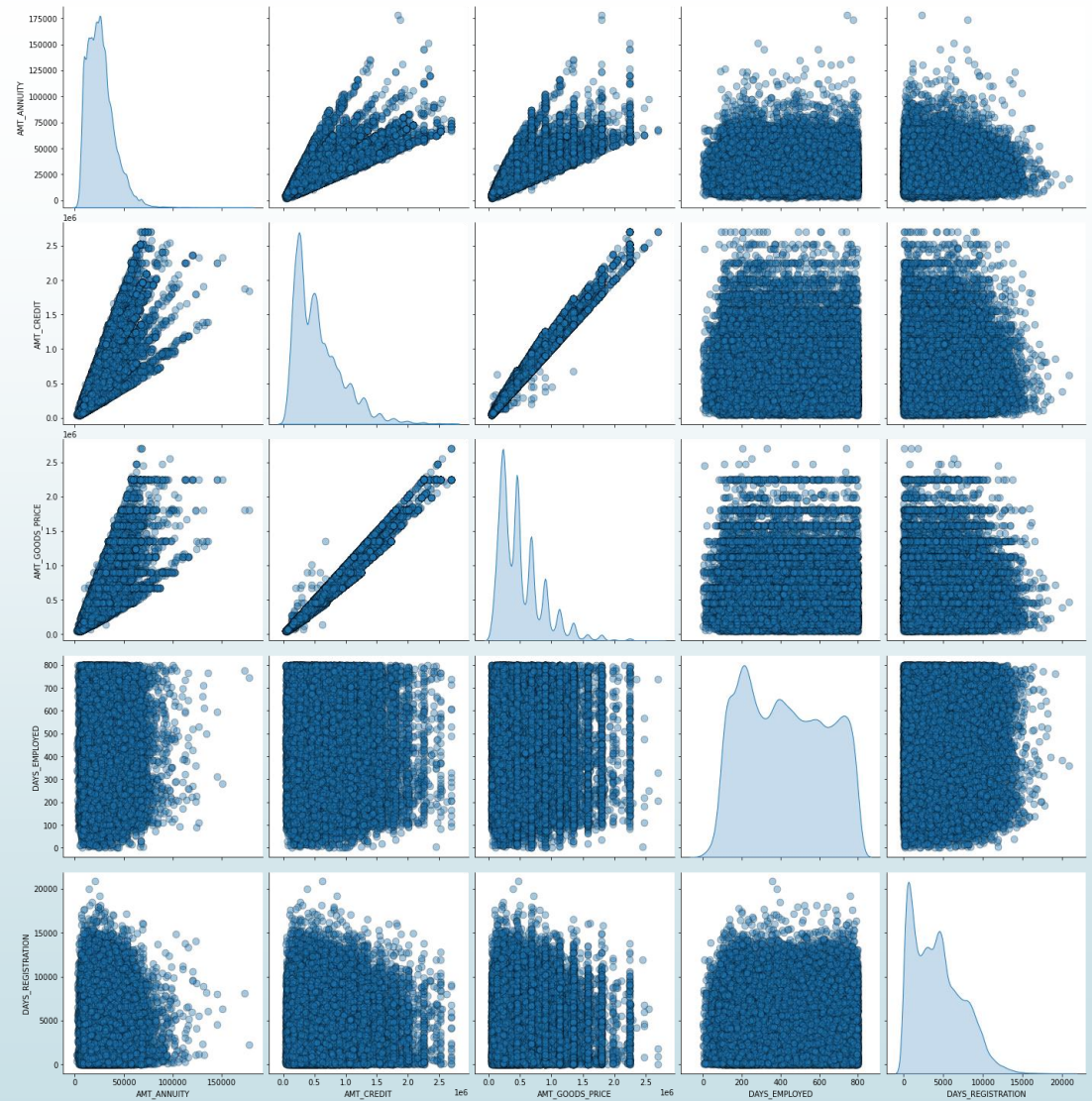
UNIVARIATE ANALYSIS :

- In both the target_0 and target_1 the highest distribution are of Married peoples and the lowest distribution is of Widow's, irrespective of their gender.

- We have Females in target_1 in higher numbers as compare to Males, which means Females of all type of Family Status are more likely to default.

- We can also see in target_0 the distribution of Females are higher as compared to Males, so that's why the target1 have also high distribution of Females.
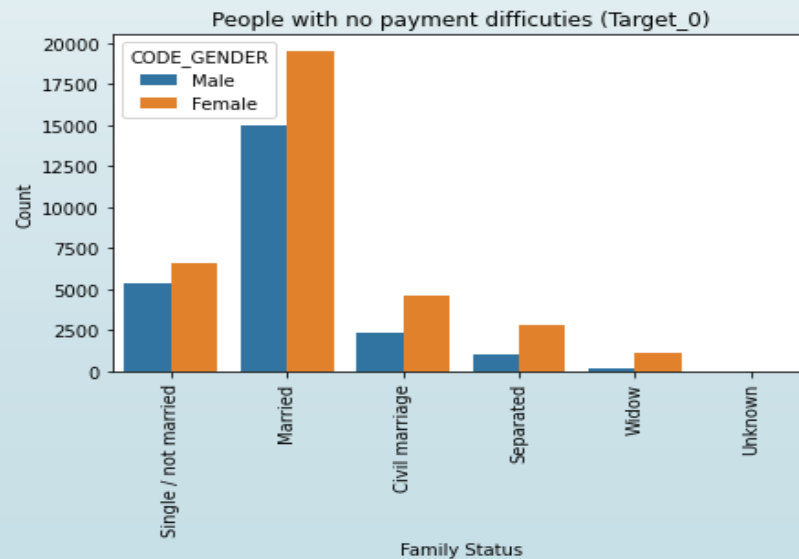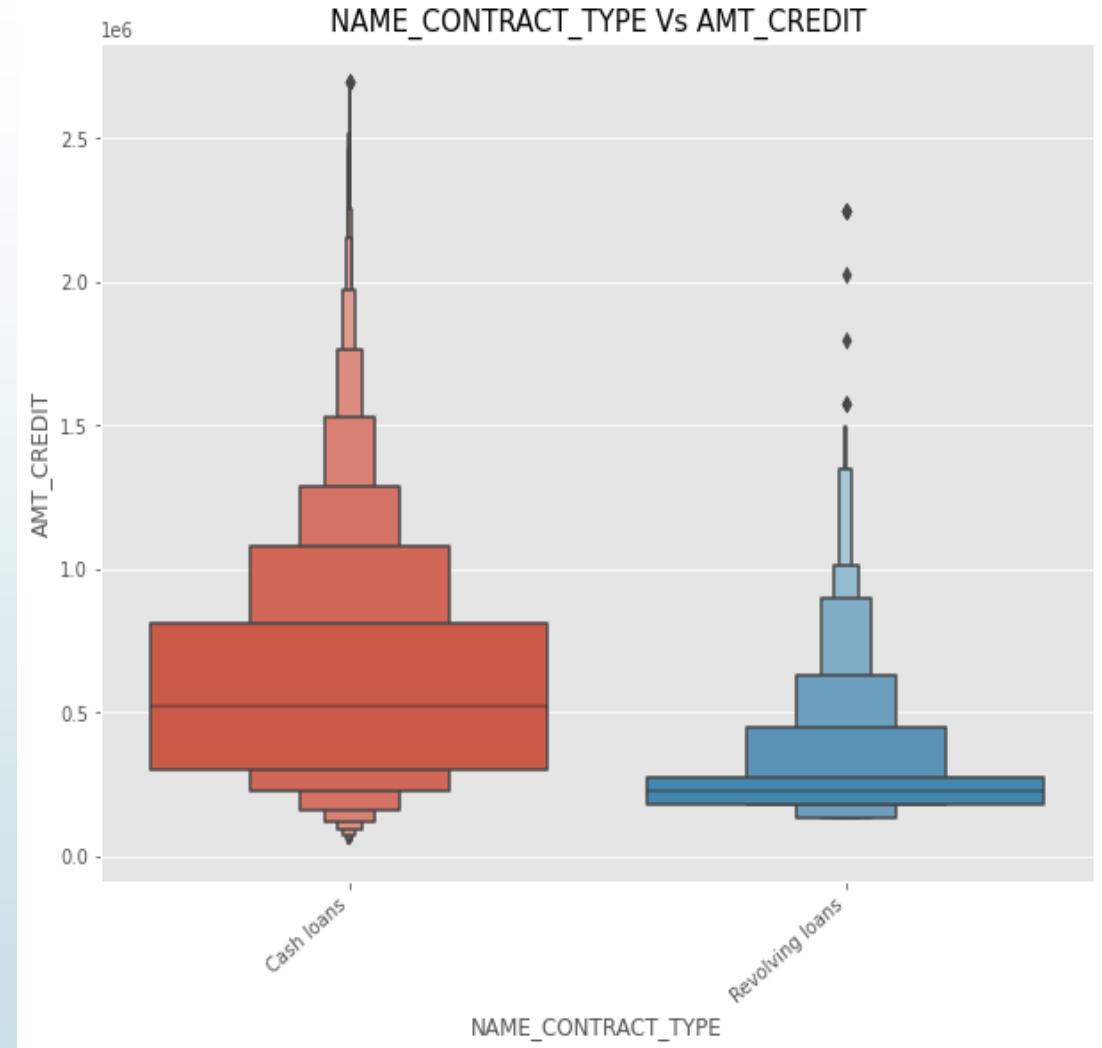
BIVARIATE ANALYSIS :
(NUMERIC – NUMERIC)

- There is not such correlation can be observed from the above pairplots

- Linear correlation present between AMT_GOODS_PRICE vs AMT_CREDIT , AMT_GOODS_PRICE vs AMT_ANNUITY
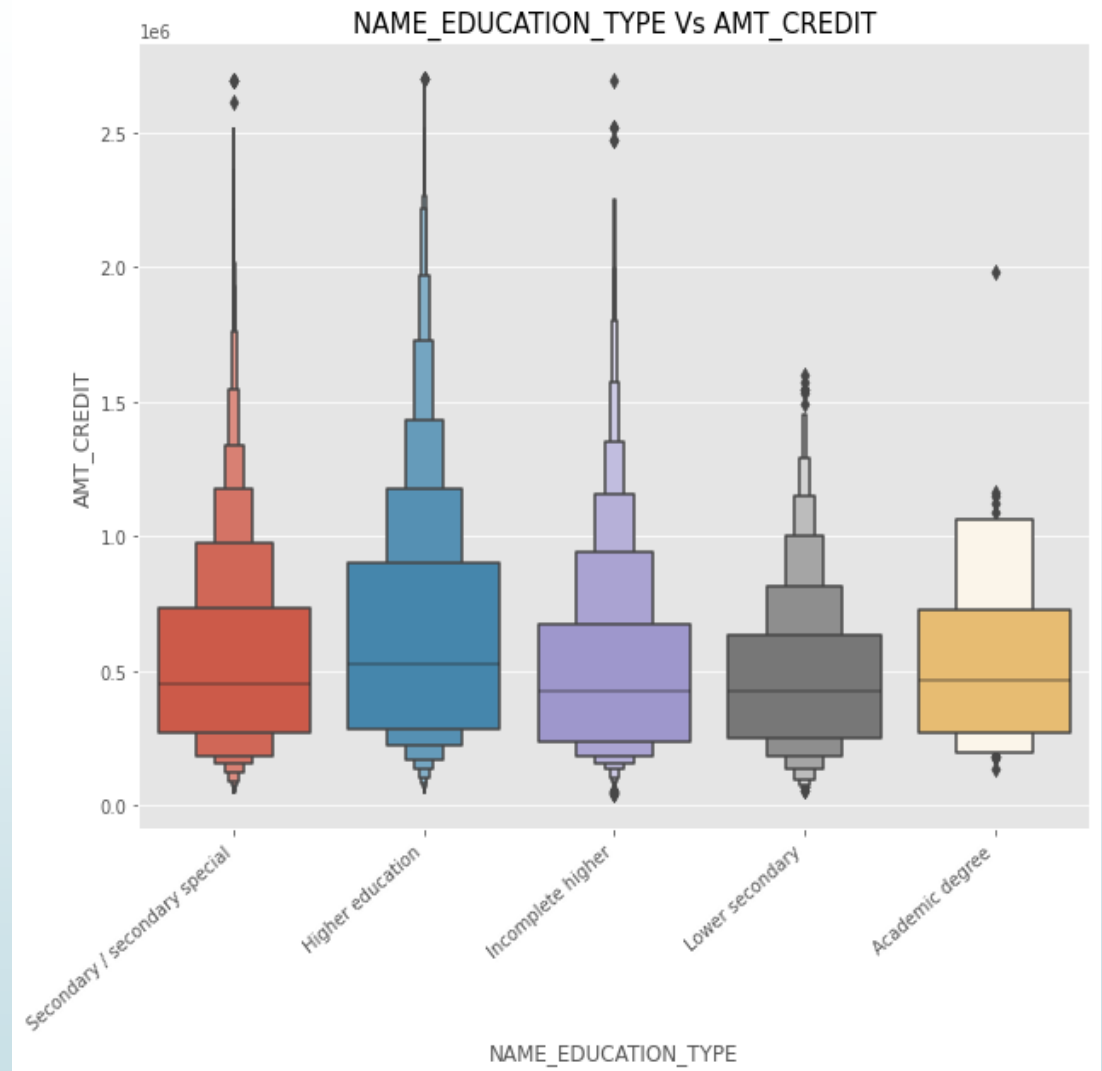
BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- There is huge difference between 75th quartile in cash loans and revolving loans

- There are some outliers present in cash loan

- Number of clients is very high for cash loans than revolving loan

NAME_CONTRACT_TYPE Vs AMT_CREDIT

BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- median is approximately same for each education type

- Number of outliers are present for each education type

BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- Number of outliers are more for working and commercial associate as compared to state servant

- Median is heighest for commercial associate

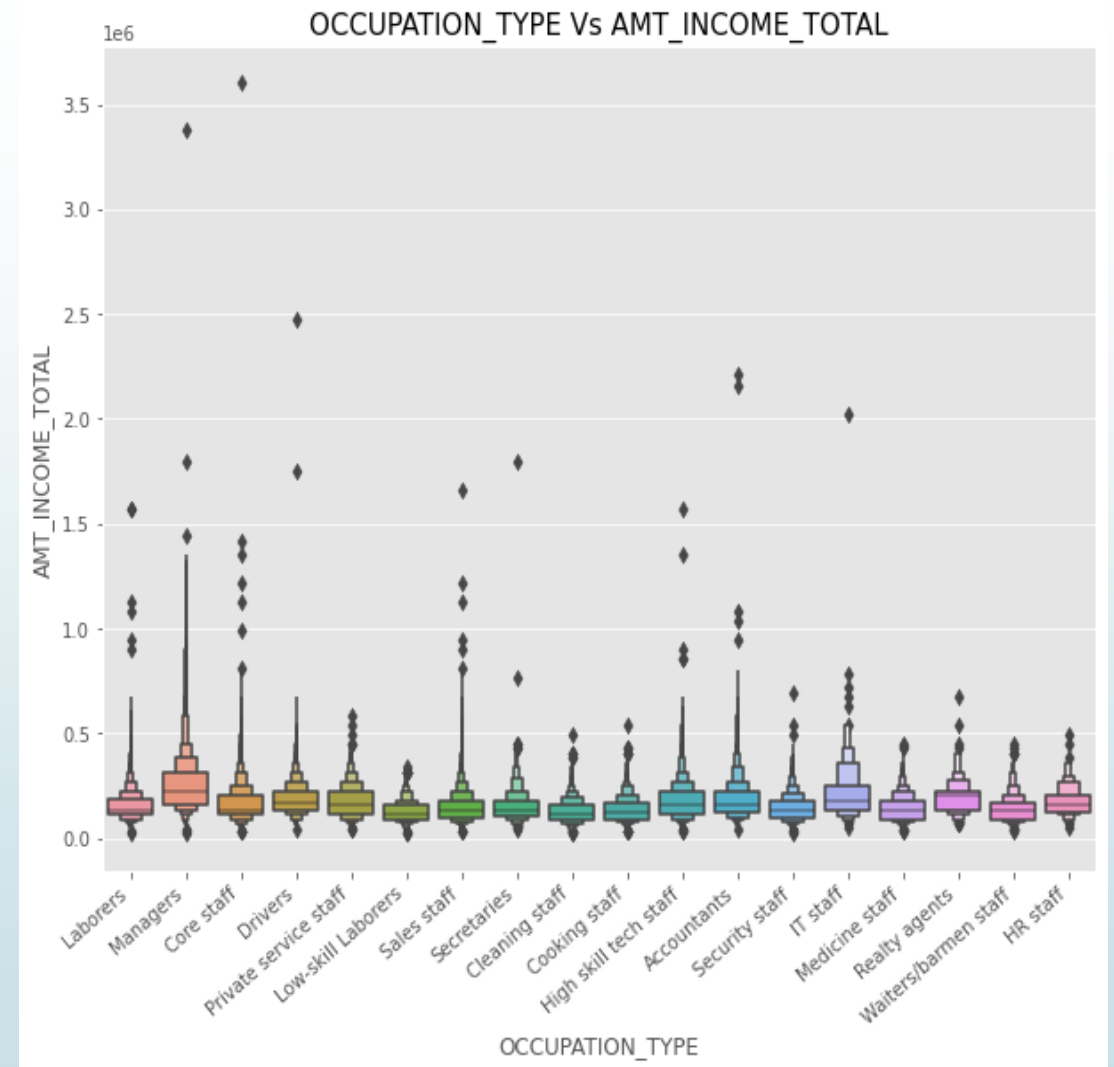NAME_INCOME_TYPE Vs AMT_CREDIT
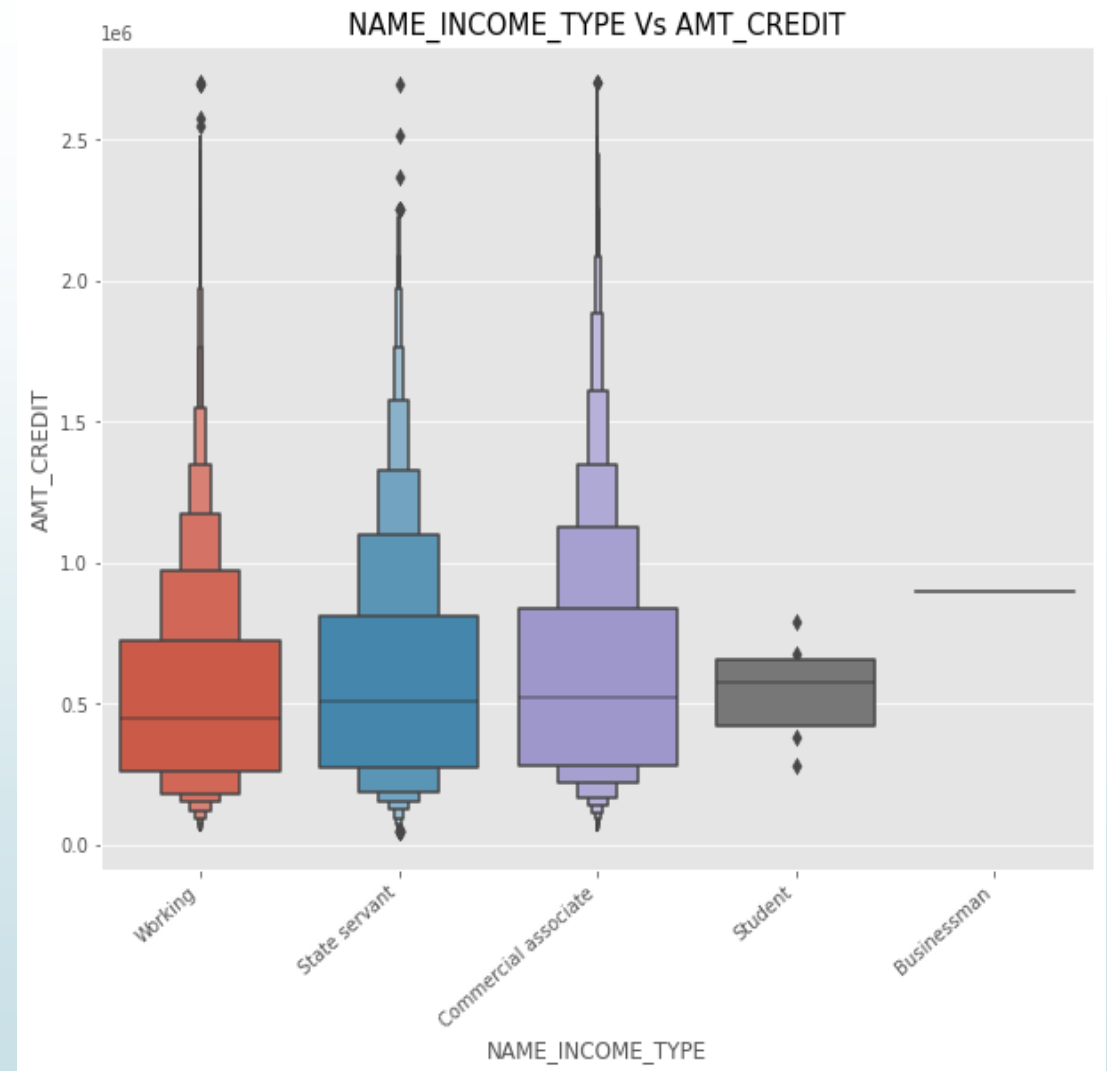
BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- Number of outliers is also more for each occupation type

- Median of occupation type Managers is highest
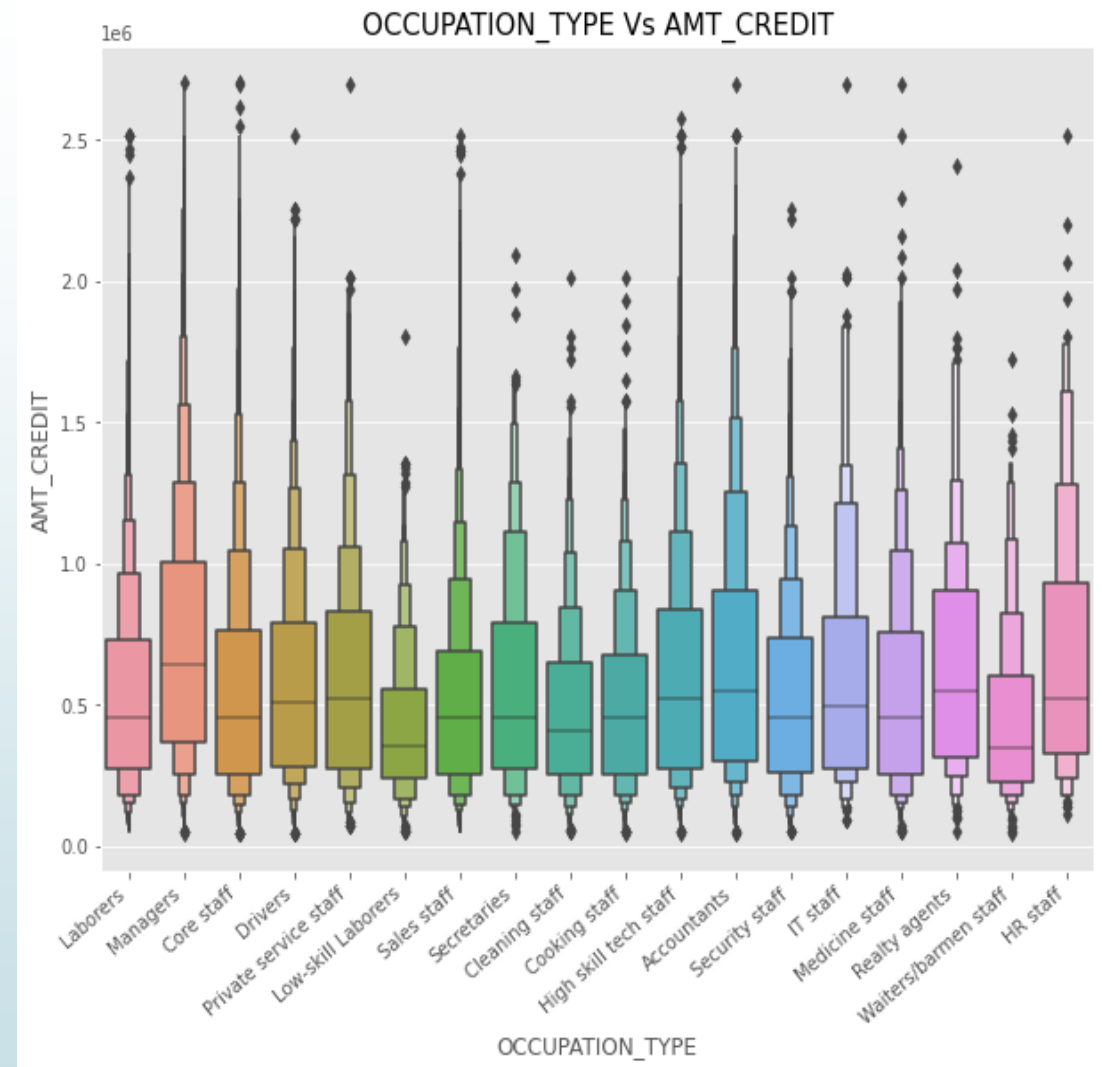
BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- Occupation type Managers have highest income total

- Cleaning staff, cooking staff ,Low skill laborers have lowest income total
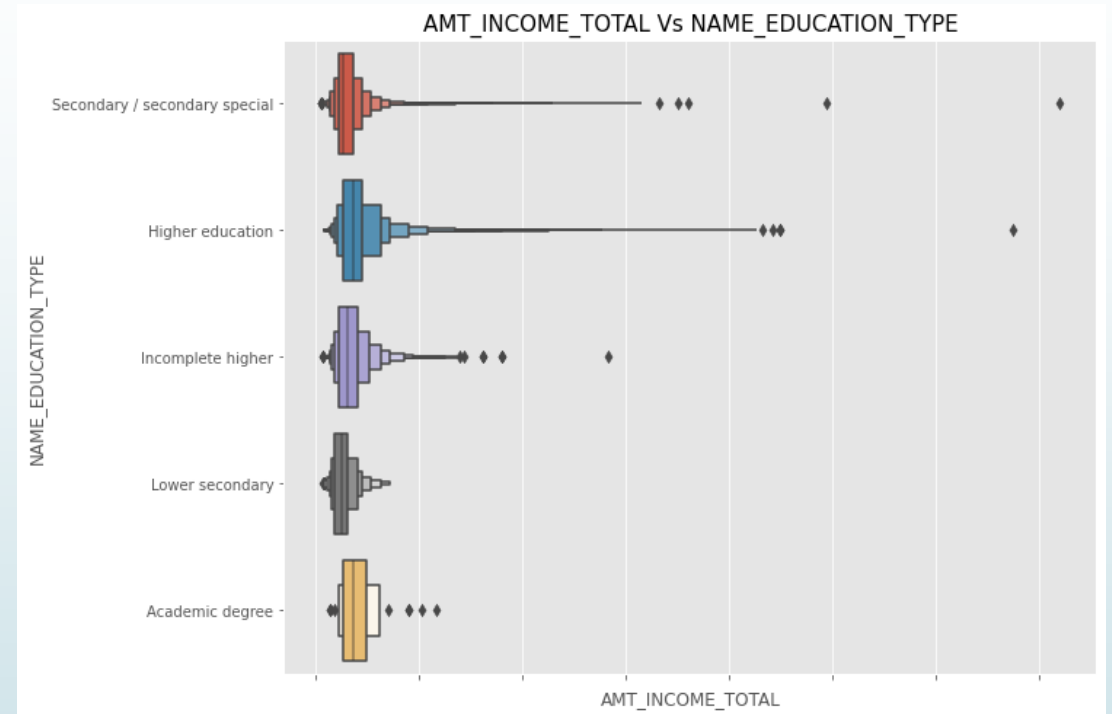
BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- Number of outliers are more for male gender

- Median of income total is more of male gender as compared with female gender


AMT_INCOME_TOTAL Vs CODE_GENDER
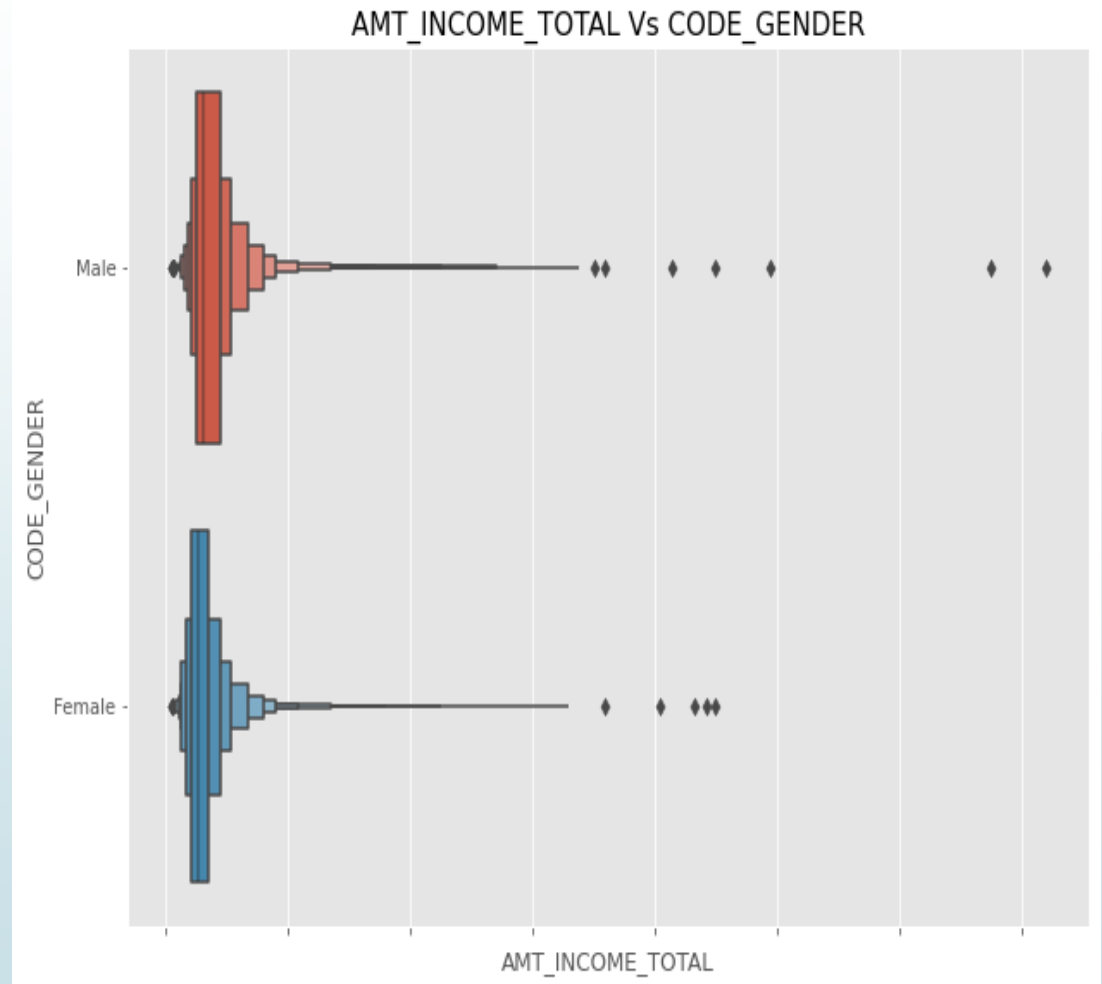
BIVARIATE ANALYSIS : (NUMERIC – CATEGORICAL)

- More outliers present in the education type Higher education and secondary special

- Median of income total is lowest for education type lower secondary

BIVARIATE ANALYSIS :
(CATEGORICAL – CATEGORICAL)

- Clients with Education type academic are not available in data

- Clients having education type Secondary/secondary special have highest number of cars
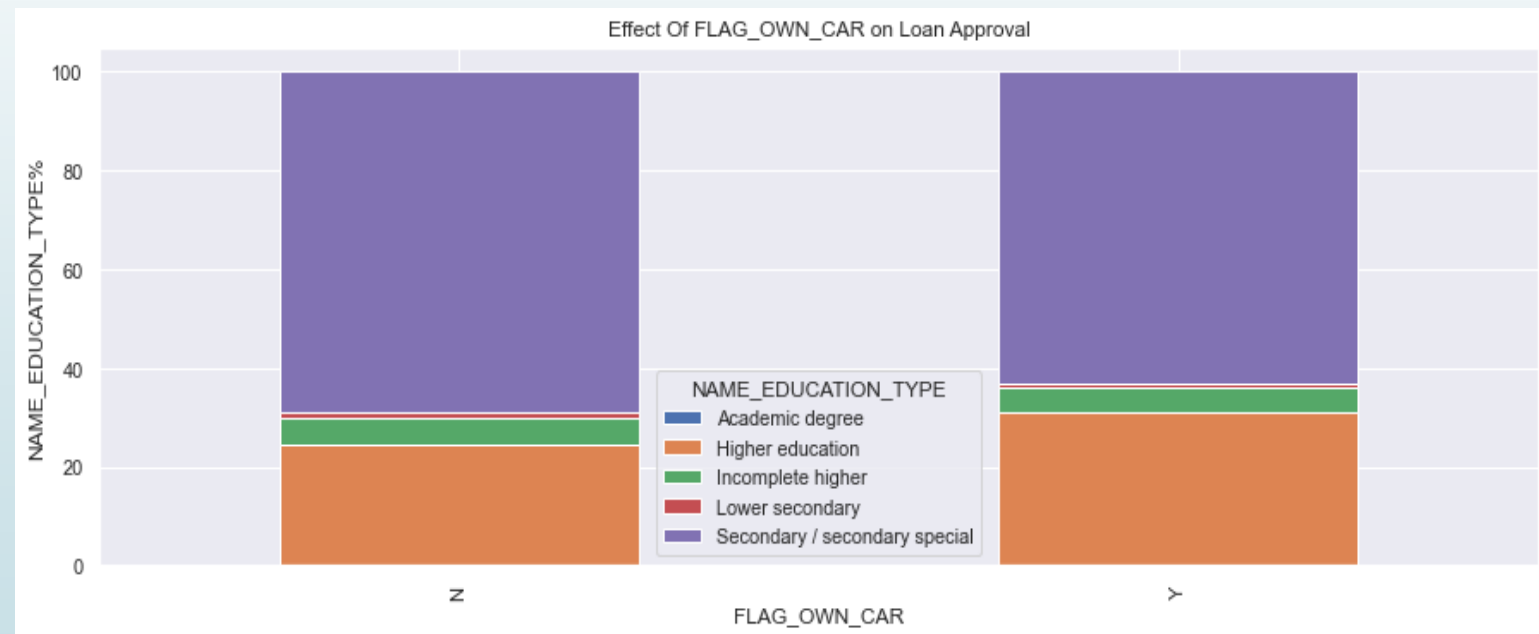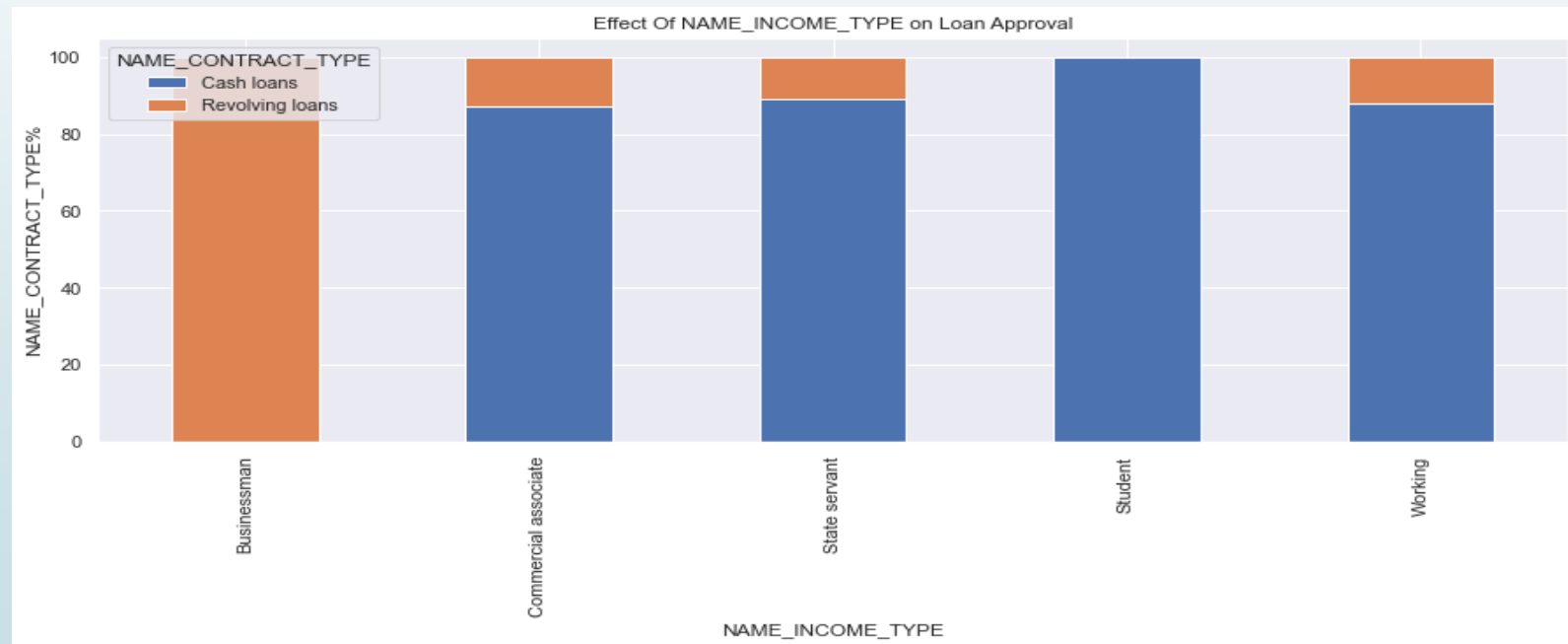
BIVARIATE ANALYSIS :
(CATEGORICAL – CATEGORICAL)

- Very few Female and male clints are present having occupation type HR staff and IT staff

- Proportion of male and female clients with occupation type managers is some what same.

- Proportion of female clients is more having occupation type Core staff, Sales staff , Accountants , cleaning staff , cooking staff and unknown

- Very less male clients present in Reality agents, Private service staff, Medicine staff, Waiters/barmen staff

- Number of male clients is more in Drivers, Security staff and Laborers
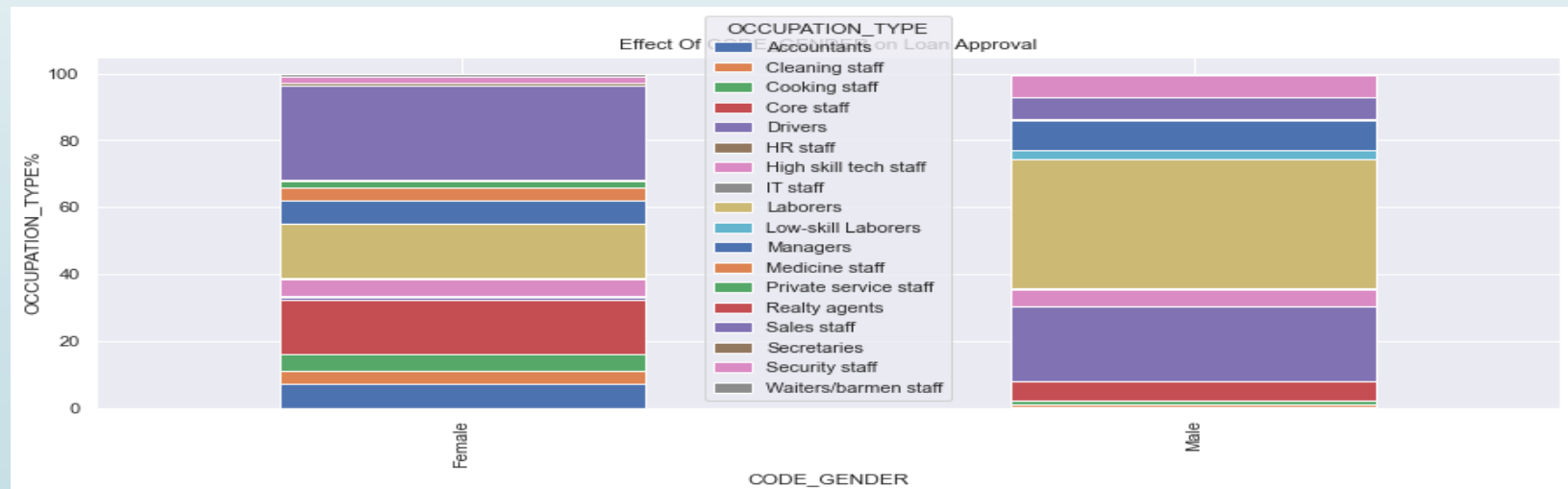
BIVARIATE ANALYSIS :
(CATEGORICAL – CATEGORICAL)

- There is not any correlation between income type Businessman and Student

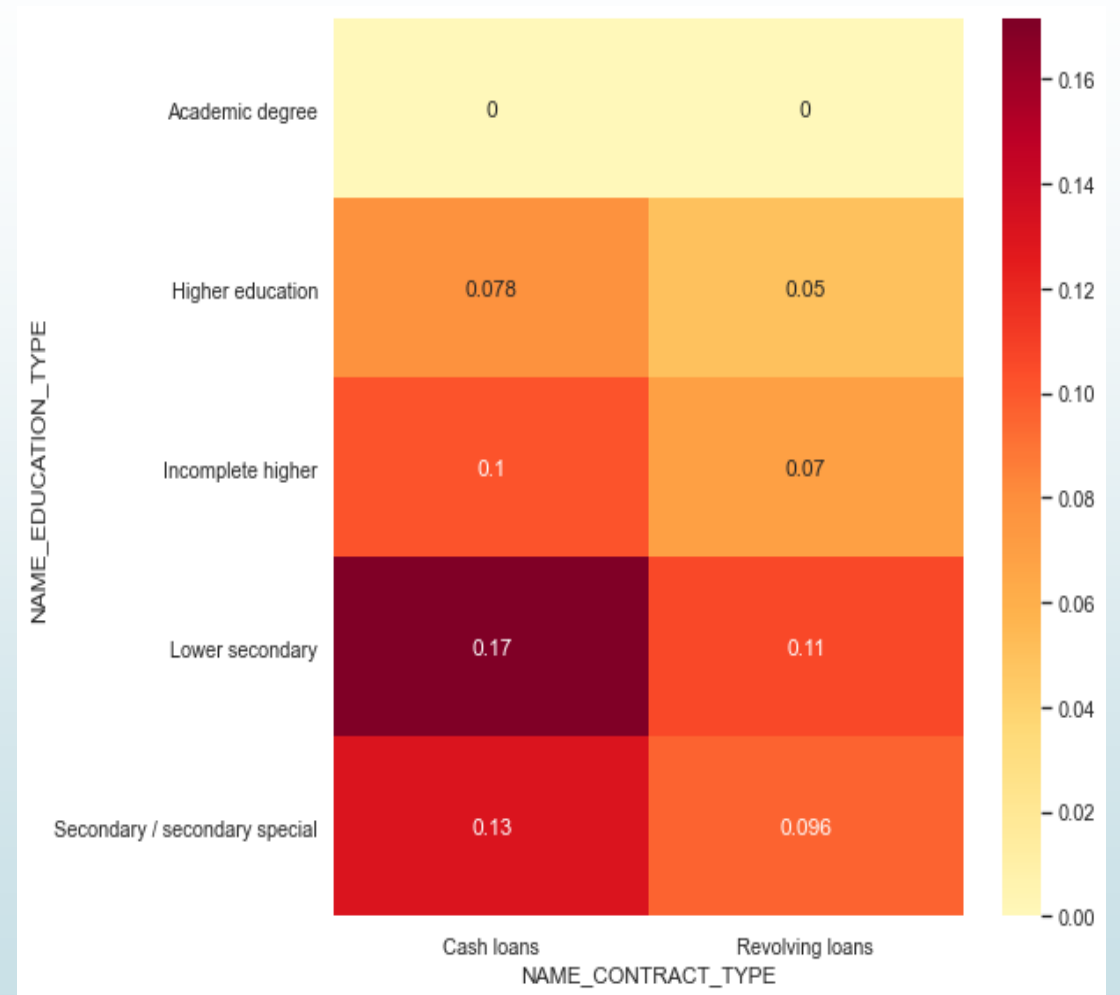- Number of cash loan client is very high for students

MULTIVARIATE ANALYSIS :

- Low-skill Laborers and cash loans have highest correlation of 0.22

- IT staff and revolving loans have no correlation at all.
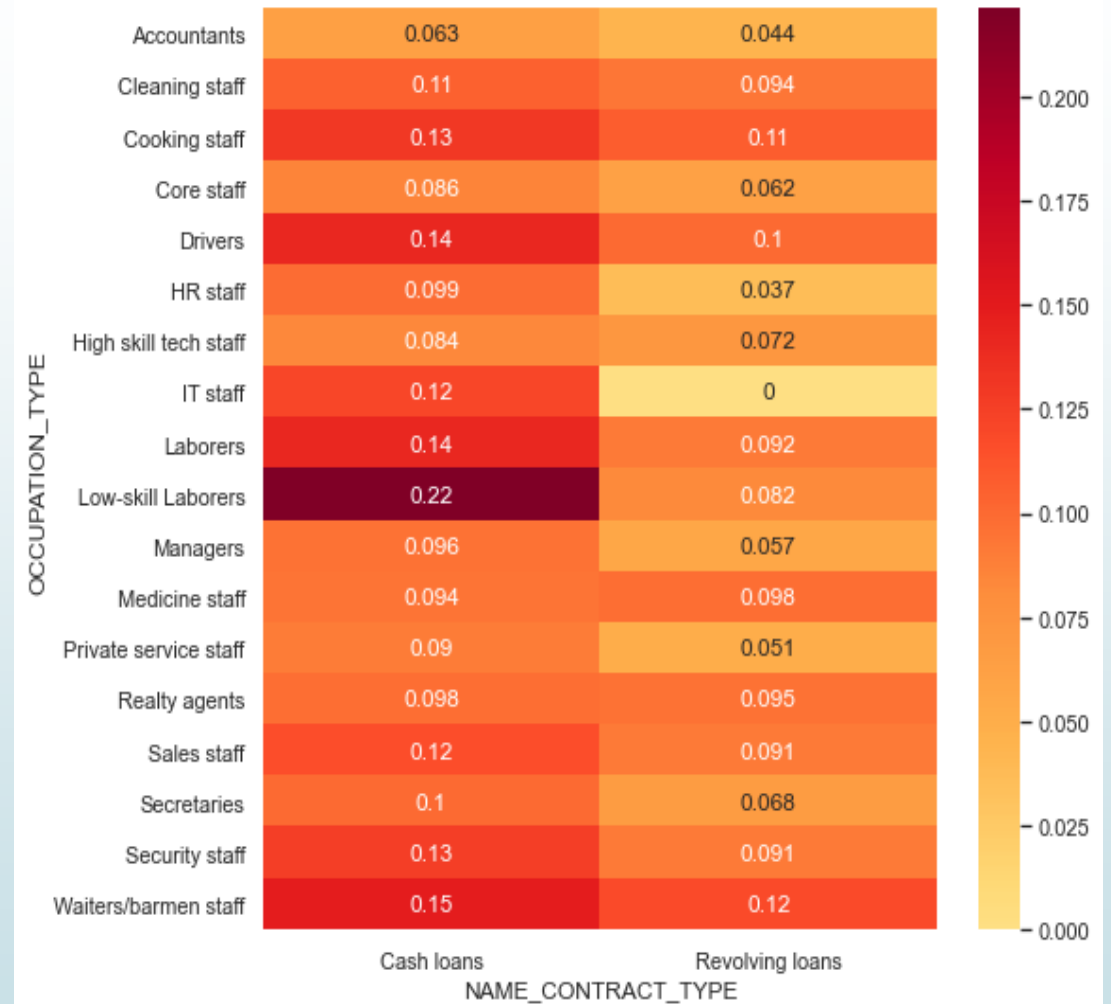
MULTIVARIATE ANALYSIS :

- Highest correlation is between lower secondary and cash loans

- Higher education and revolving loans have lower correlation of 0.05

MULTIVARIATE ANALYSIS :

- Higher correlation is between realty agents and incomplete higher

- Medicine staff and incomplete higher have lower correlation of 0.029
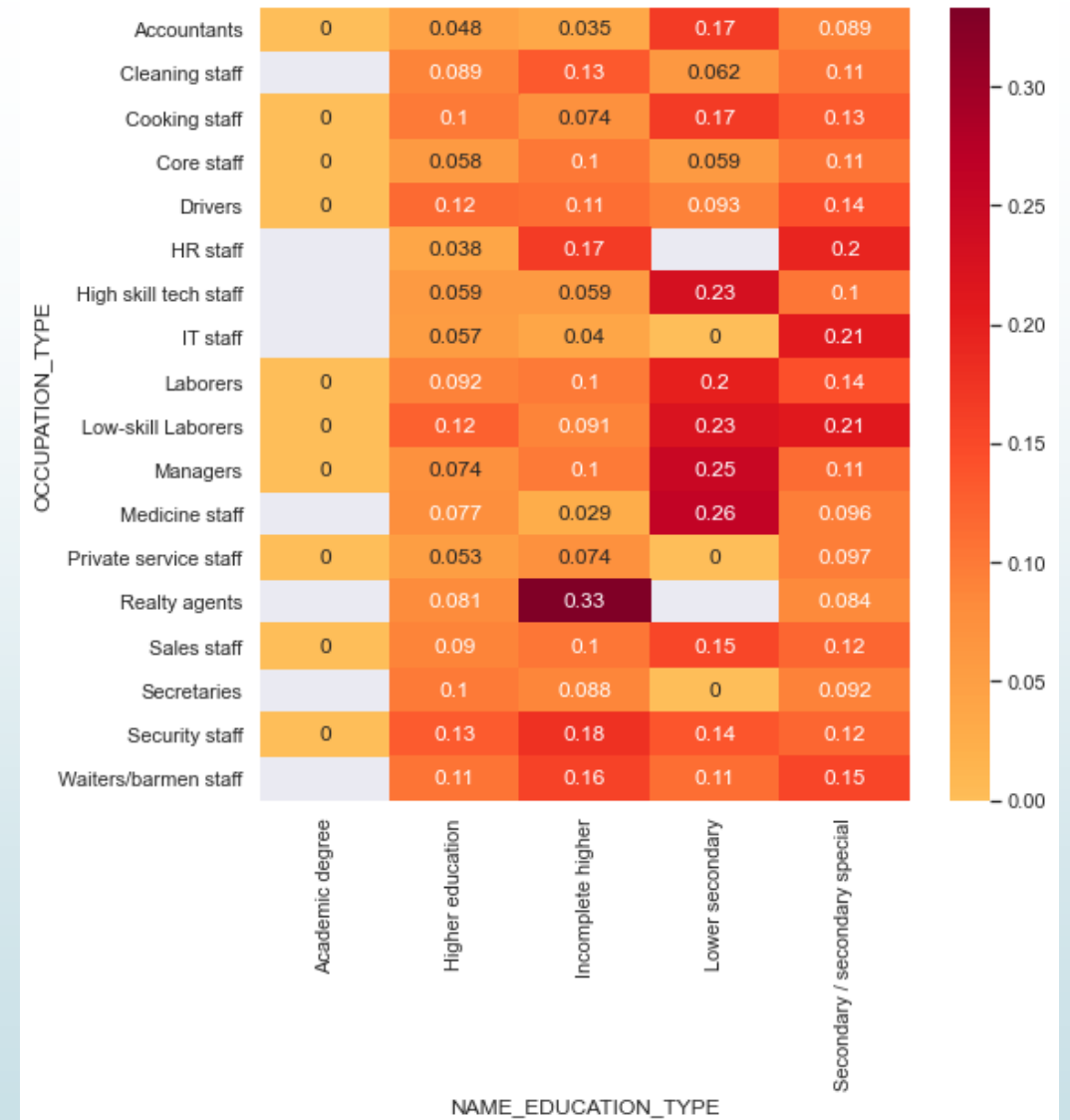
MULTIVARIATE ANALYSIS :

- AMT_INCOME_TOTAL and AMT_CREDIT have higher correlation.

- EXT_SOURCE_2 and TARGET have negative correlation of -0.18
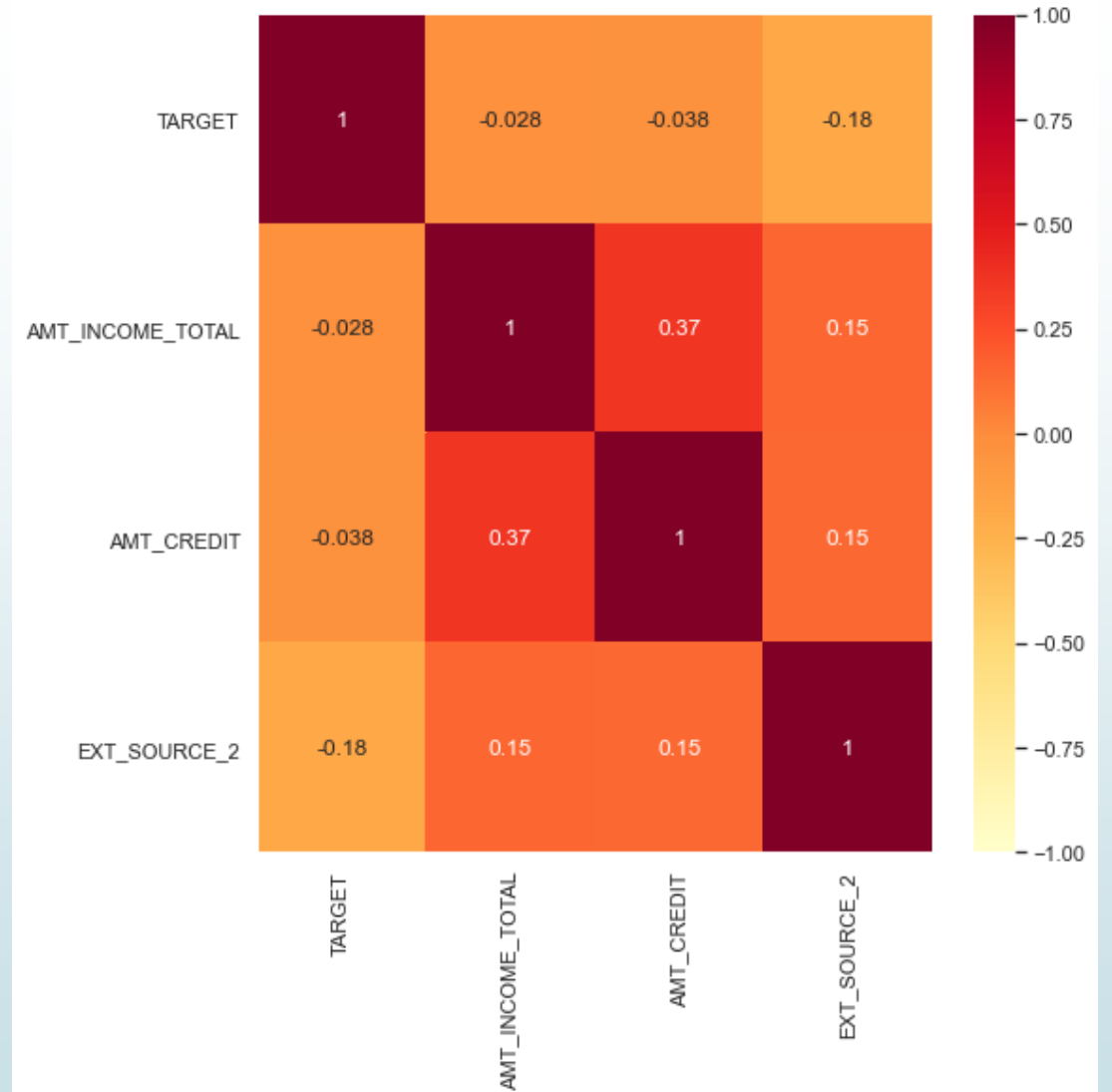
CORRELATION :

- AMT_GOODS_PRICE and AMT_CREDIT have highest correlation around 0.99

- AMT_GOODS_PRICE and AMT_ANNUITY is also showing good correlation

People without payment problem
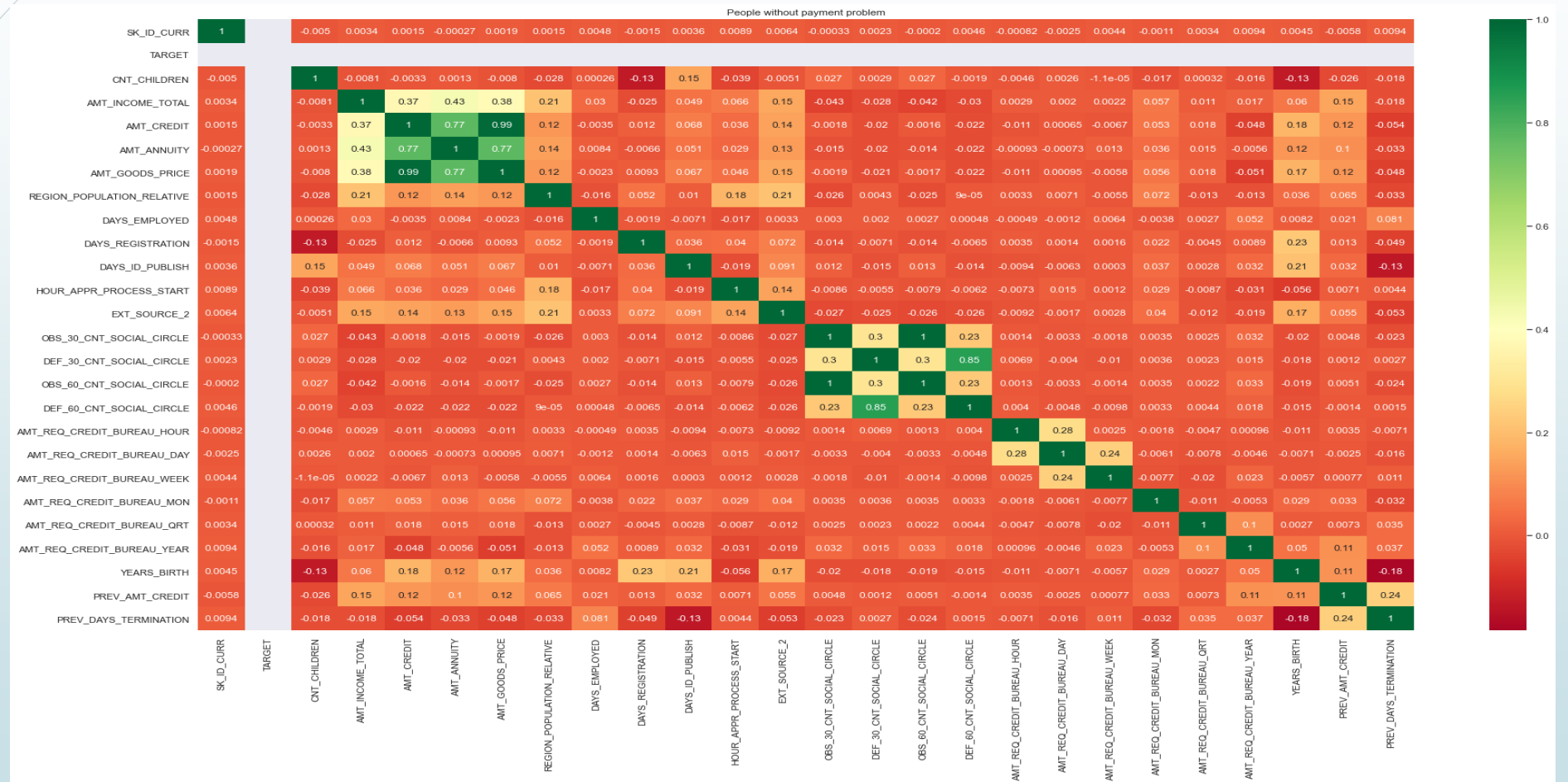
CORRELATION :

- AMT_CREDIT and AMT_GOODS_PRICE have heighest correlation around 0.98

People with payment problem