Application Data Analysis

Problem Statement

1

Final credit amount differs with the client application amount.

2

Identify defaulters is the one of
the most important problem.

Assumptions

01
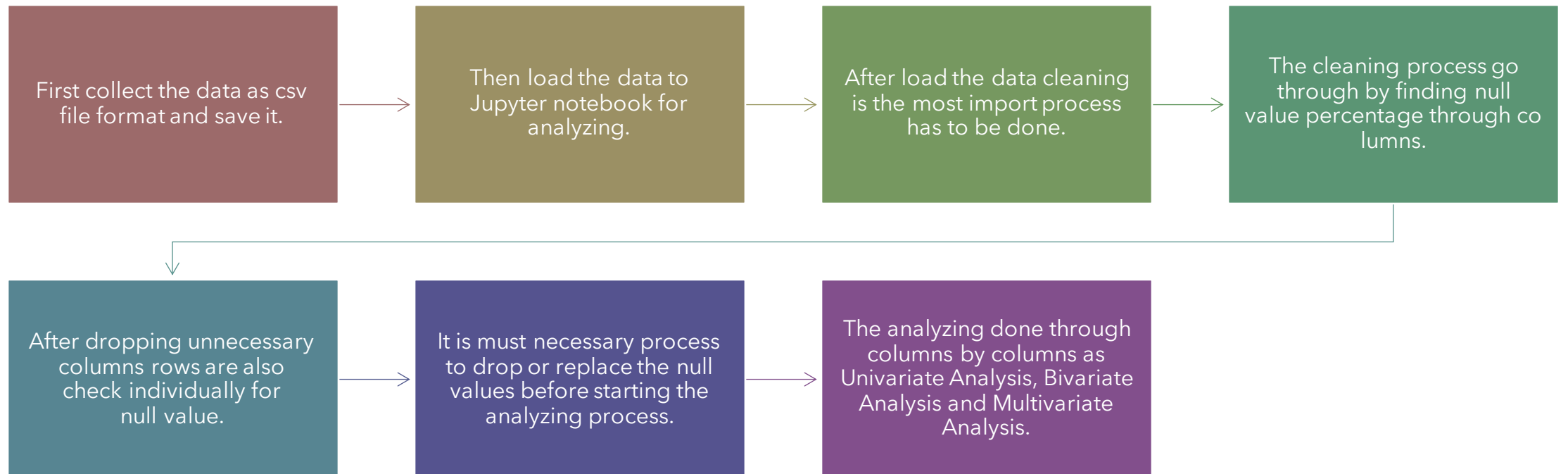The data has **307510** row entries and **122** columns to be analyzing.

02
After over through with the data, it has a huge amount of null value to be cleaning.

03
After go through the data set it has a lot of different parts to be analysis.

Overall Approach

First collect the data as csv file format and save it.

Then load the data to Jupyter notebook for analyzing.

After load the data cleaning is the most import process has to be done.

The cleaning process go through by finding null value percentage through co lumns.

After dropping unnecessary columns rows are also check individually for null value.

It is must necessary process to drop or replace the null values before starting the analyzing process.

The analyzing done through columns by columns as Univariate Analysis, Bivariate Analysis and Multivariate Analysis.

Overall Approach

The data holds 307511 rows entries and 122 columns to be Analyse.

After cleaning the data(dropping the null values) 304531 number of rows and 65 columns left.

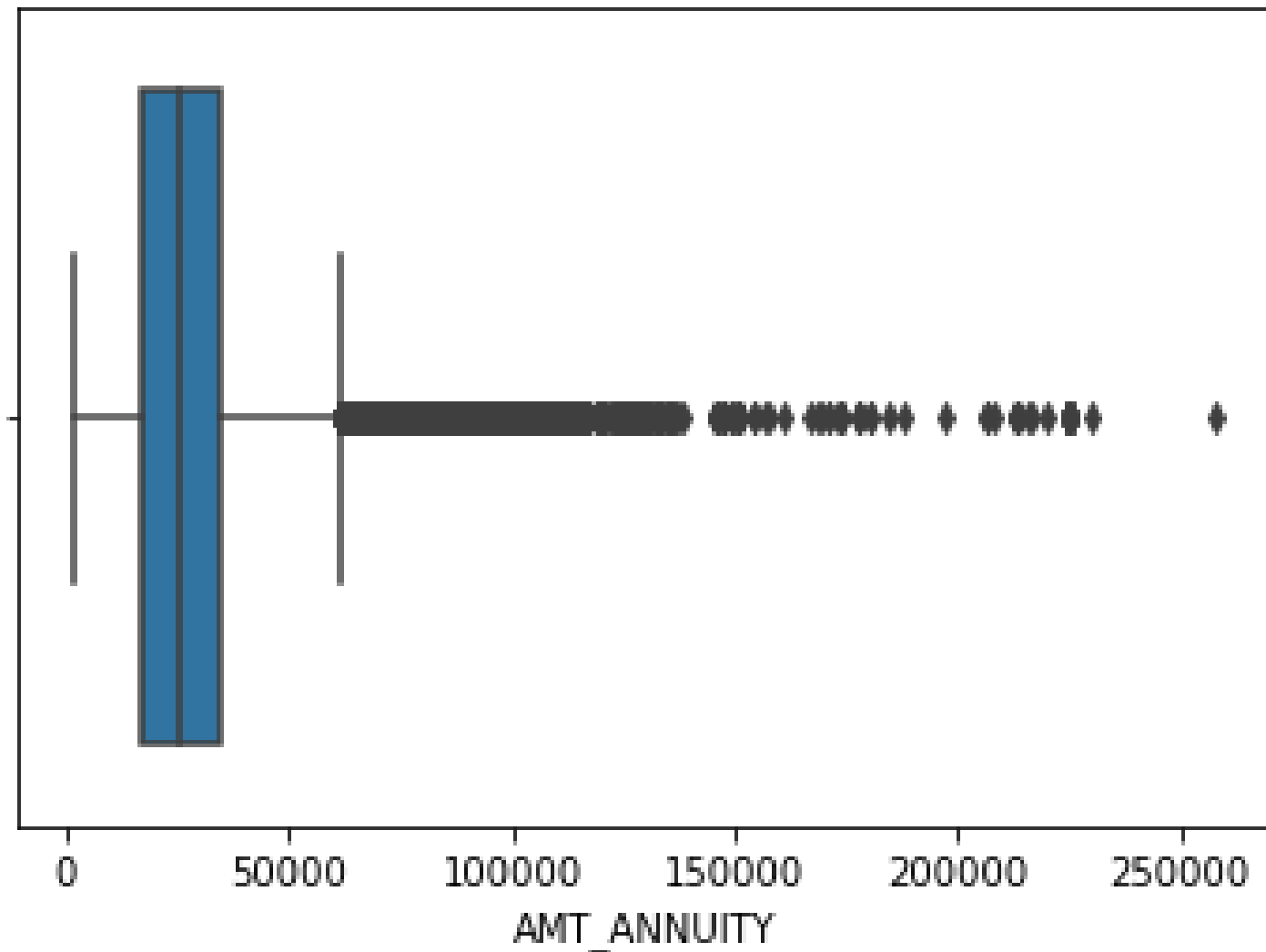Almost 50% of the columns than holds more than 10% of null values.

Rows holding null value also cleaning individually.
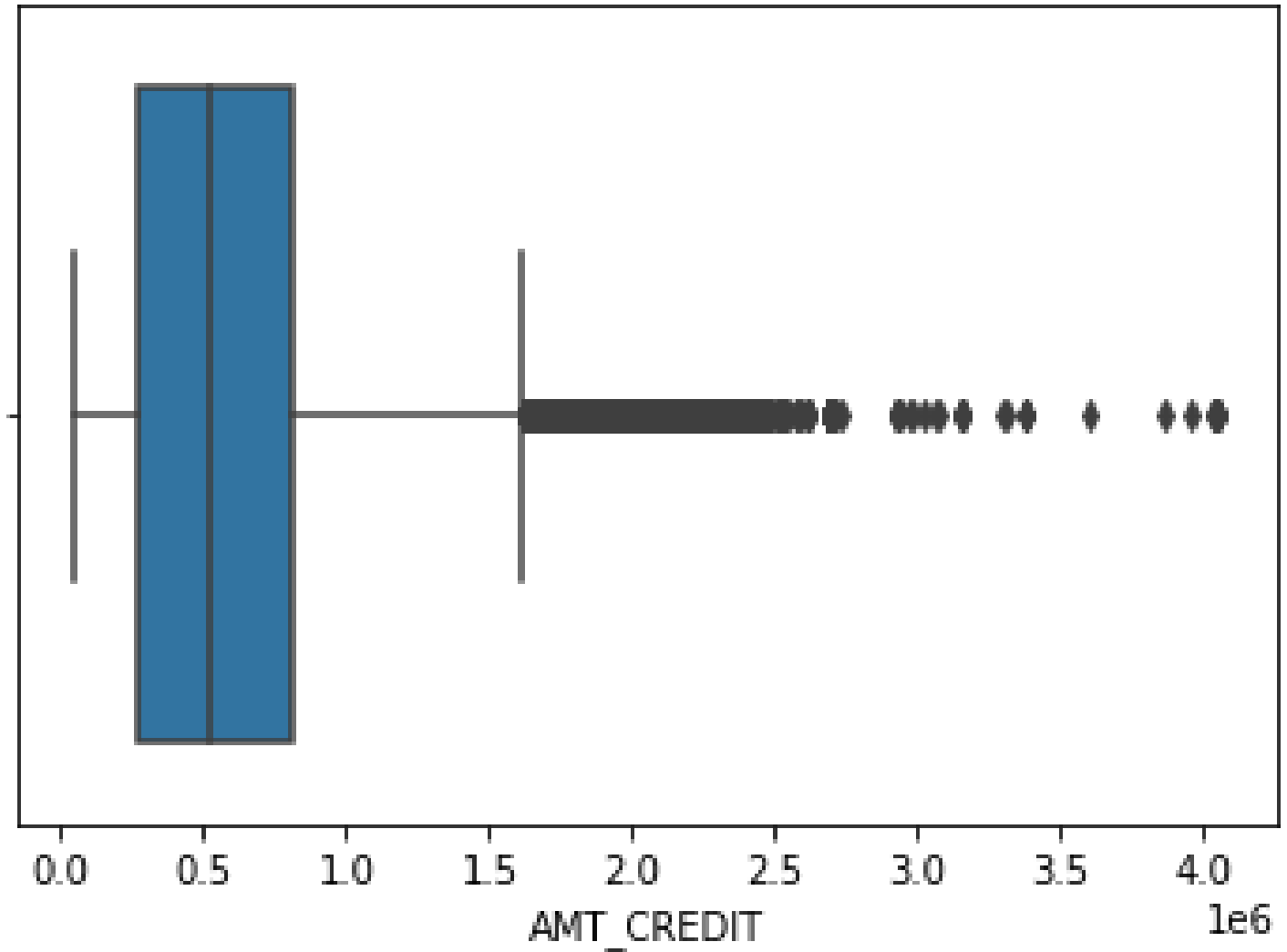
Relevant results with graphs

AMT_ANNUITY

count	304531.000000
mean	27146.369655
std	14480.104699
min	1615.500000
25%	16573.500000
50%	24939.000000
75%	34641.000000
max	258025.500000

AMT_CREDIT

count	3.045310e+05
mean	5.995592e+05
std	4.021453e+05
min	4.500000e+04
25%	2.700000e+05
50%	5.172660e+05
75%	8.086500e+05
max	4.050000e+06

AMT_INCOME_TOTAL

count	3.045310e+05
mean	1.686634e+05
std	2.378910e+05
min	2.565000e+04
25%	1.125000e+05
50%	1.476000e+05
75%	2.025000e+05
max	1.170000e+08
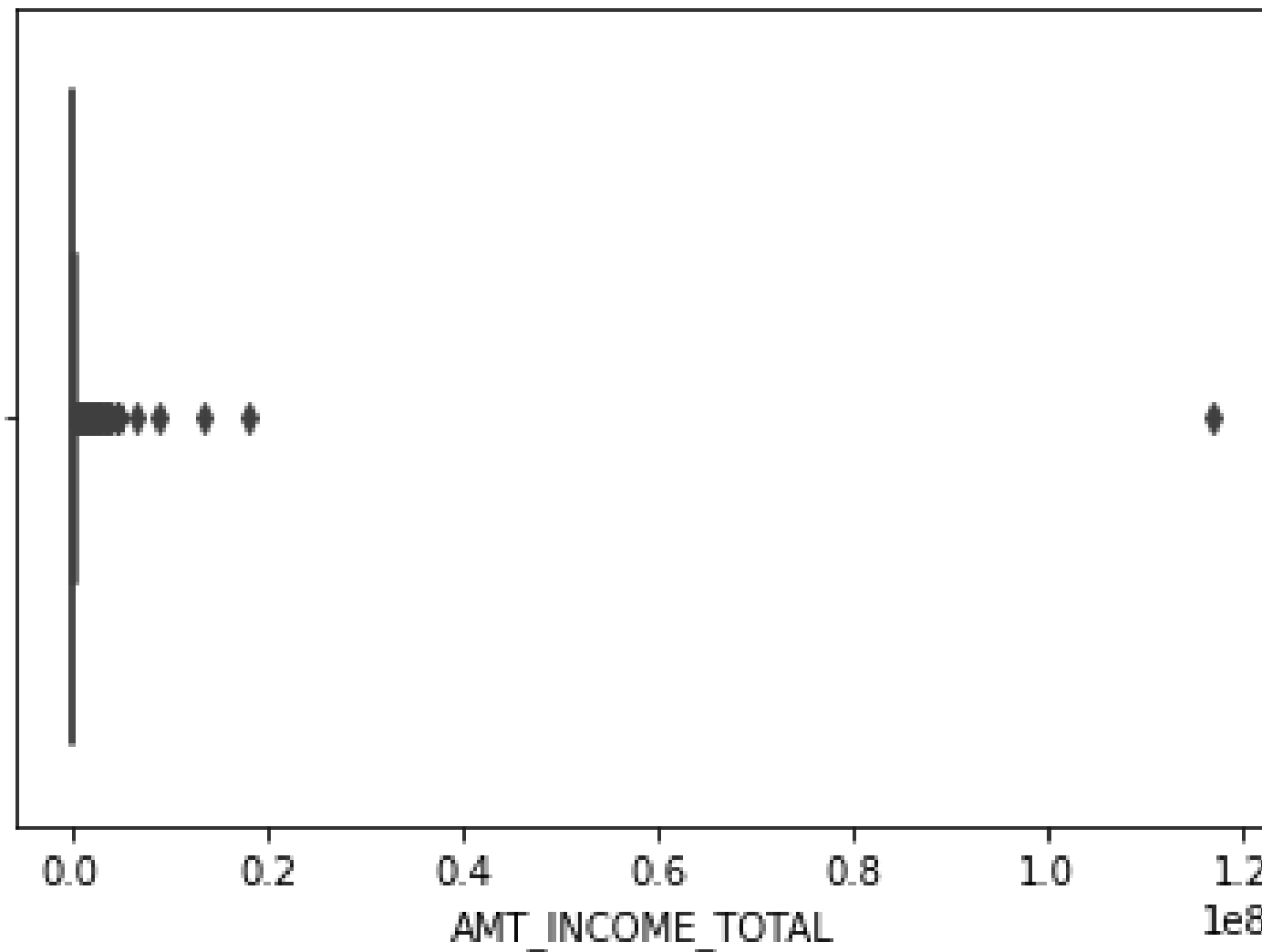
AMT_GOODS_PRICE

count	3.045310e+05
mean	5.384949e+05
std	3.690524e+05
min	4.050000e+04
25%	2.385000e+05
50%	4.500000e+05
75%	6.795000e+05
max	4.050000e+06
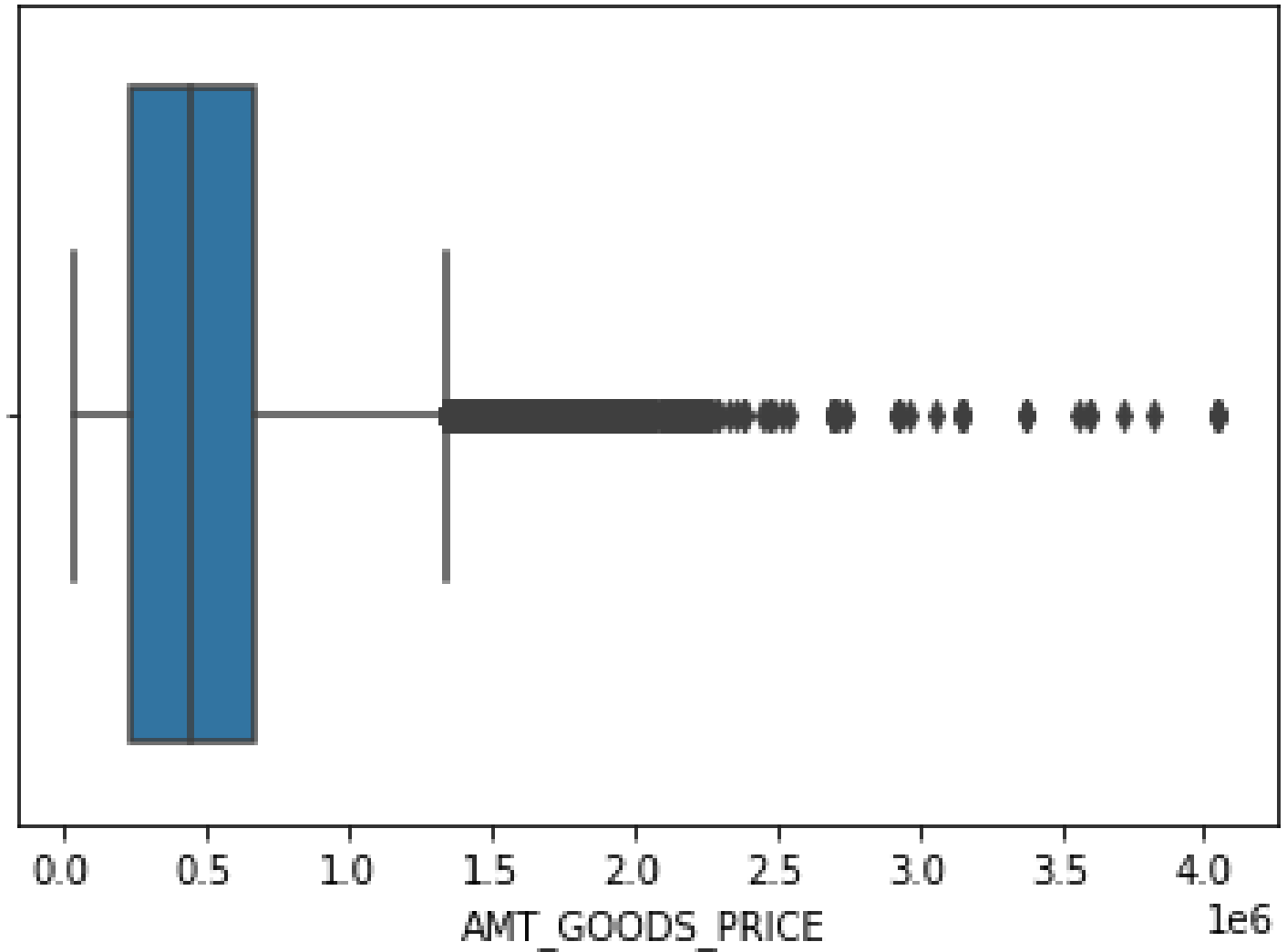
Scatterplot

- Amount ANNUITY and GOODS PRICE amount are positively co-related with each other.

- CREDIT amount and ANNUITY amount are co-related with each other with positive co-relation.
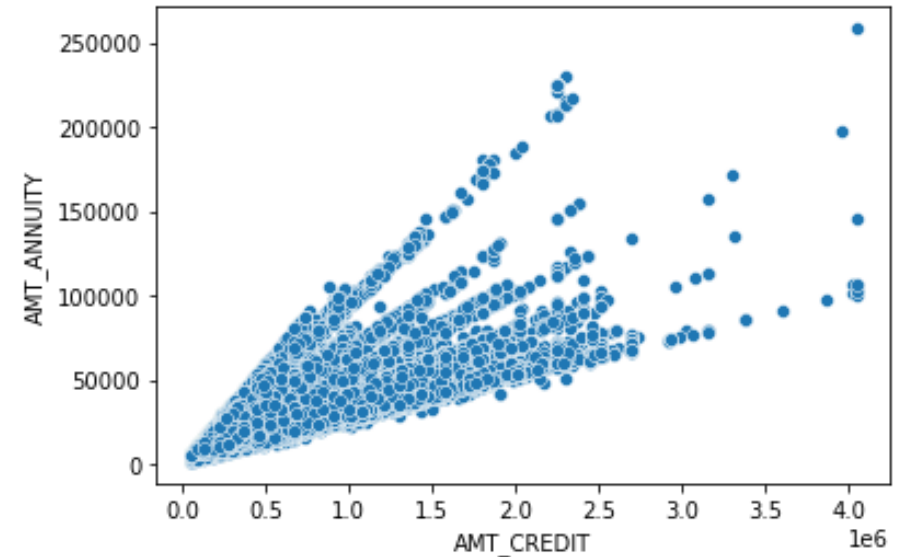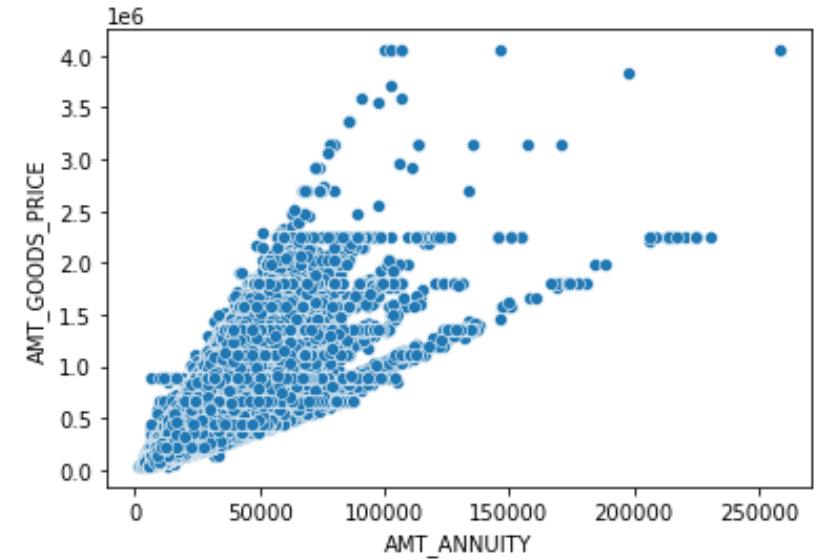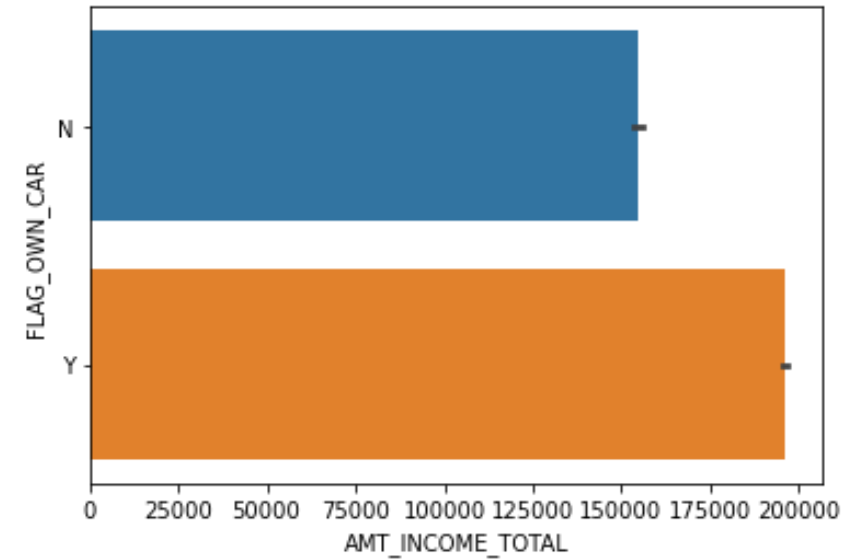
Scatterplot

- GOODS PRICE amount and CREDIT loan amount are highly positively co-related with each other.

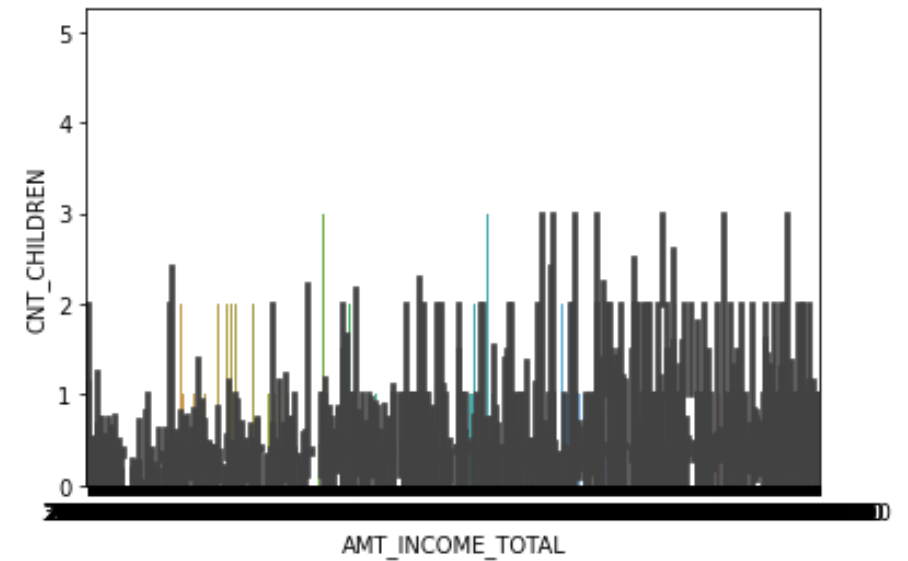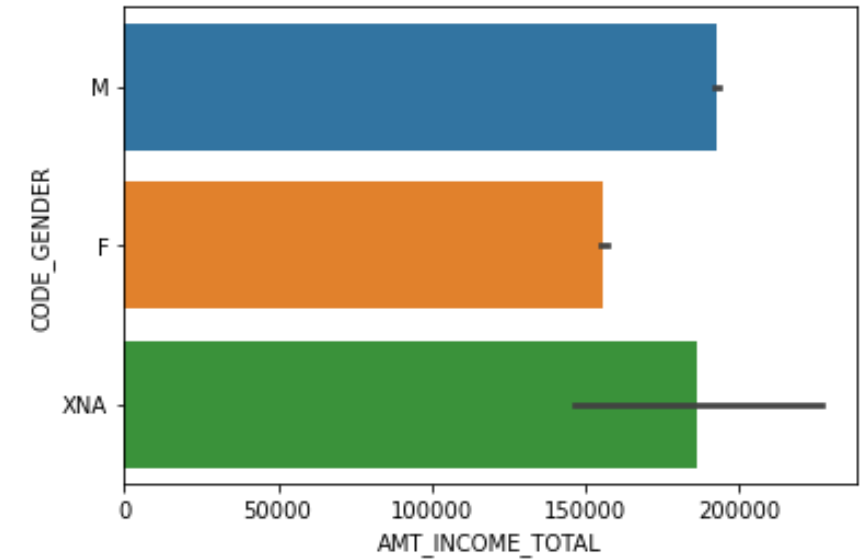- FAM MEMBERS count and ANNUITY amount are almost negatively co-related with each other.
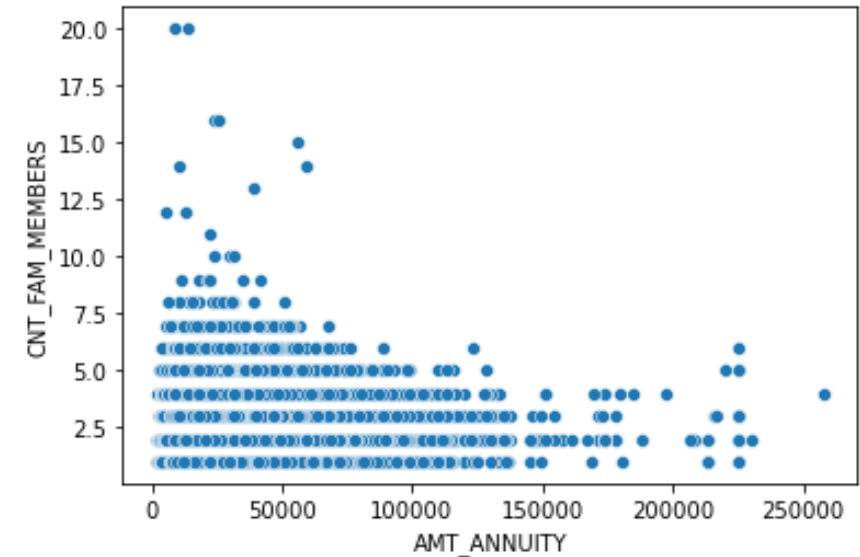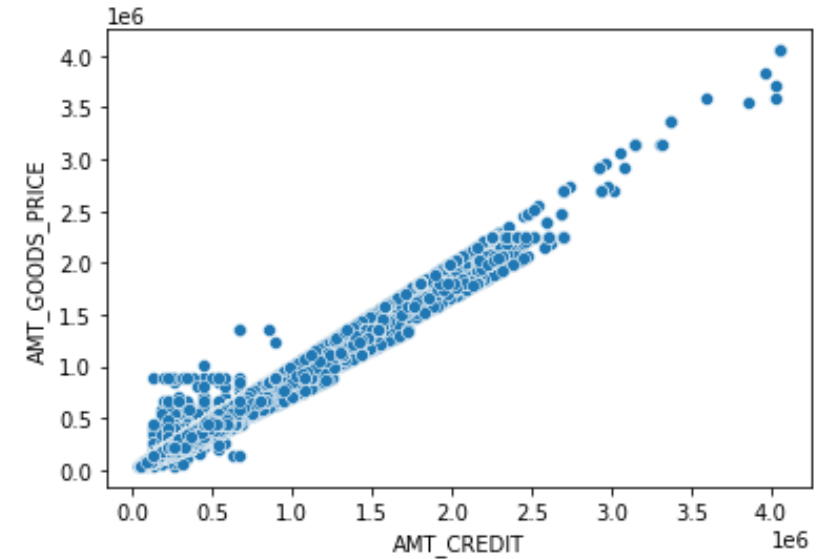
Barplot

- Majority of CAR owner clients has INCOME TOTAL almost 200000.

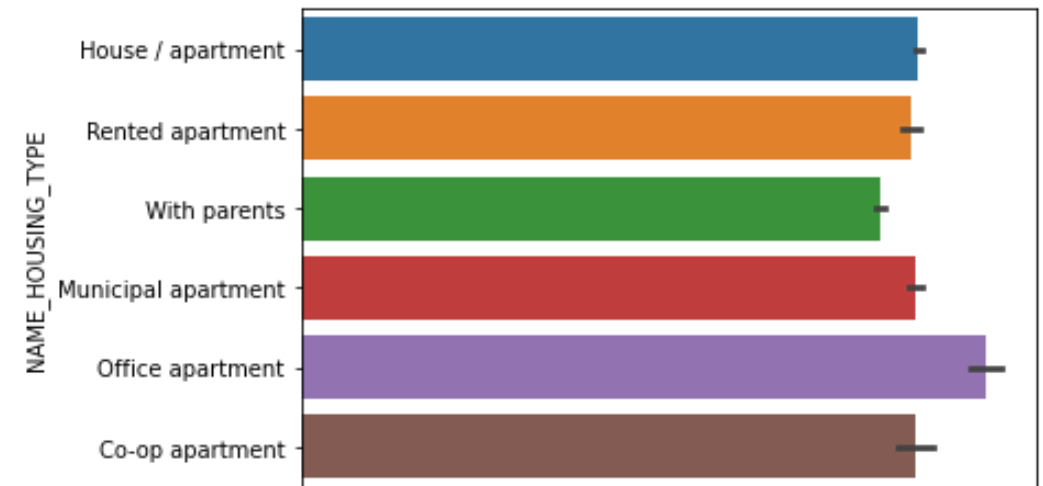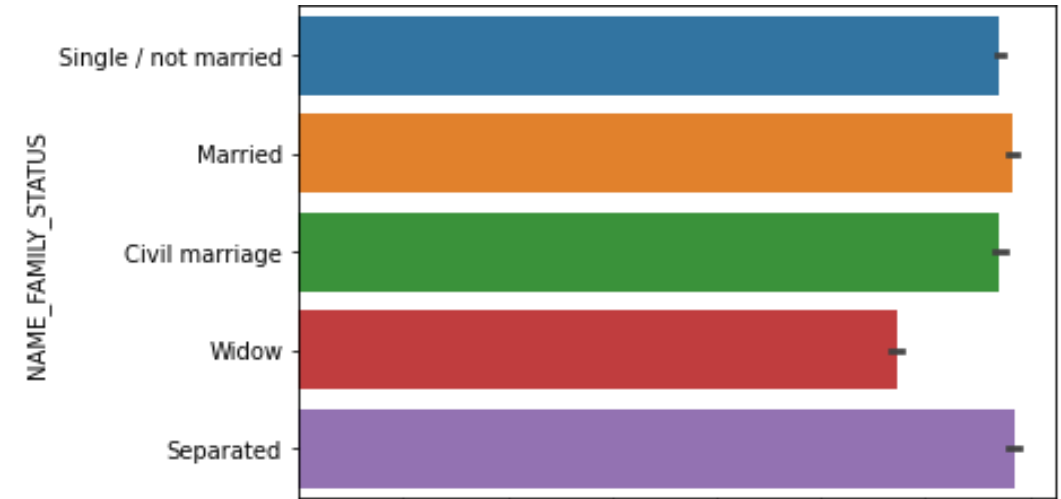- ACADEMIC DEGREE holder clients has the highest INCOME TOTAL .
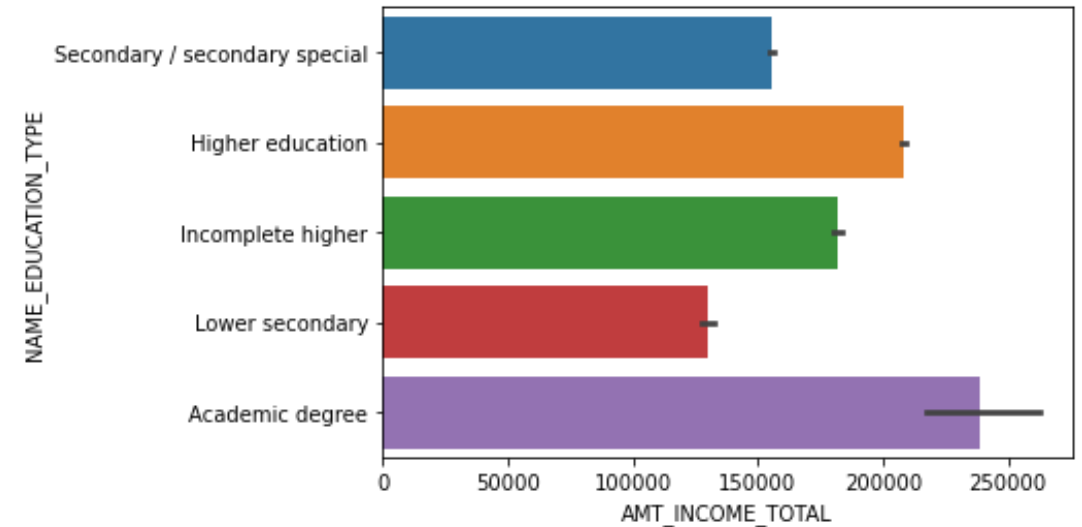
Barplot

- Majority of MALE clients have the highest INCOME TOTAL.

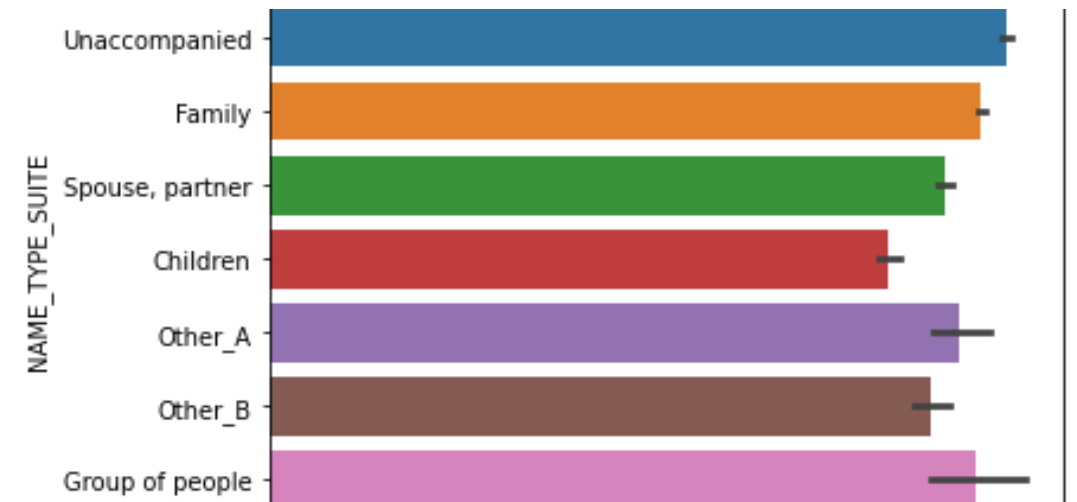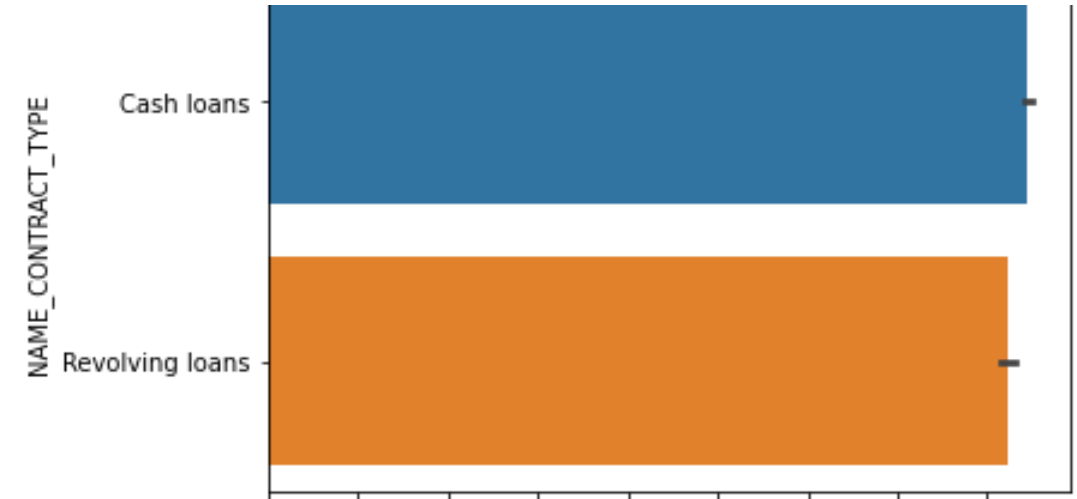- Majority of clients have no children or single one.

Barplot

- Separated and Married clients have the highest INCOME TOTAL in majority.

- Those clients HOUSING TYPE as Office apartment have highest INCOME TOTAL.
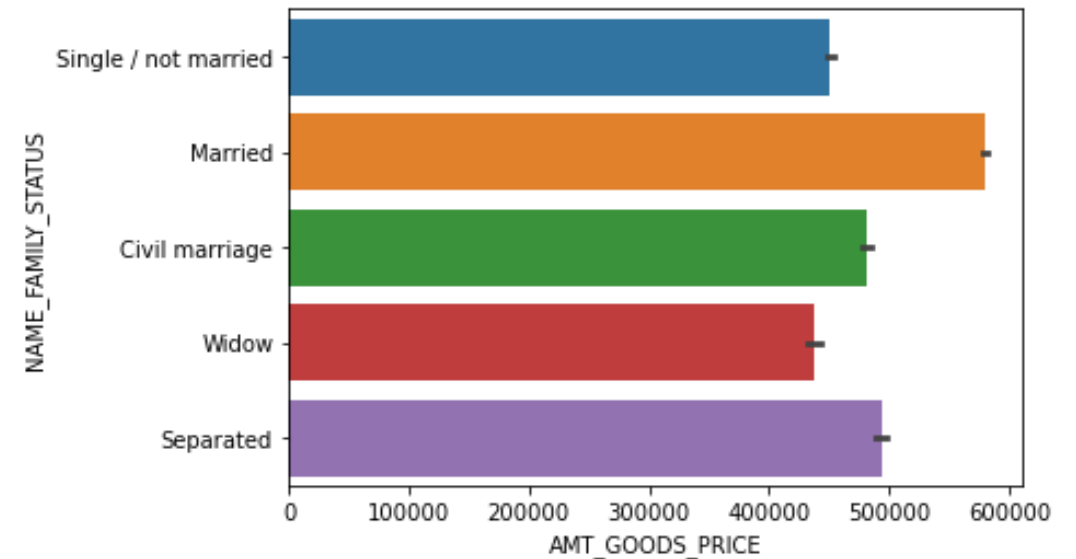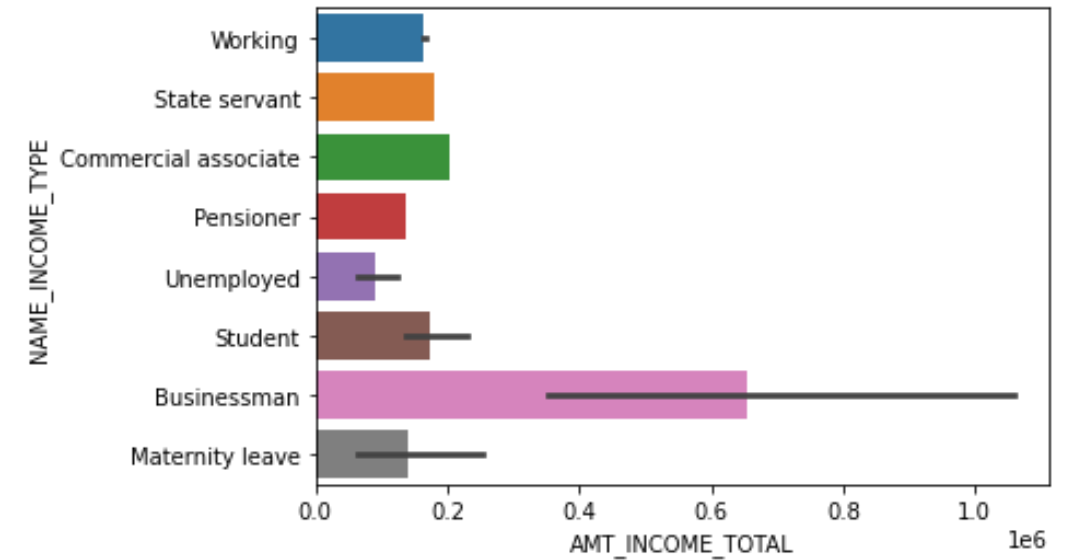
Barplot

- Majority of the clients prefer Cash as a type of loan CONTRACT.

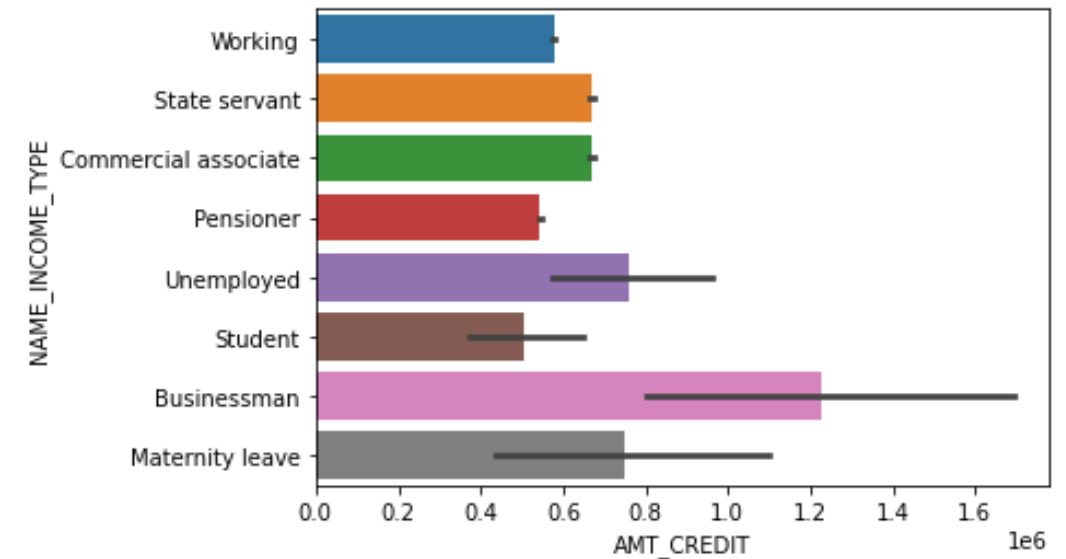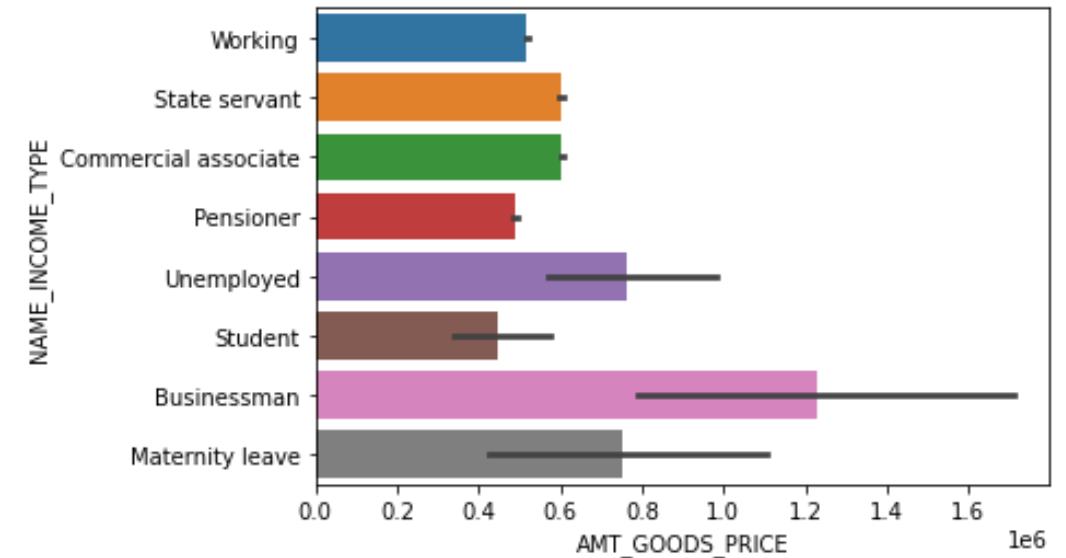- Majority of the clients are Unaccompanied during loan procedure.

Barplot

- Business is the best income type for highest INCOME TOTAL.
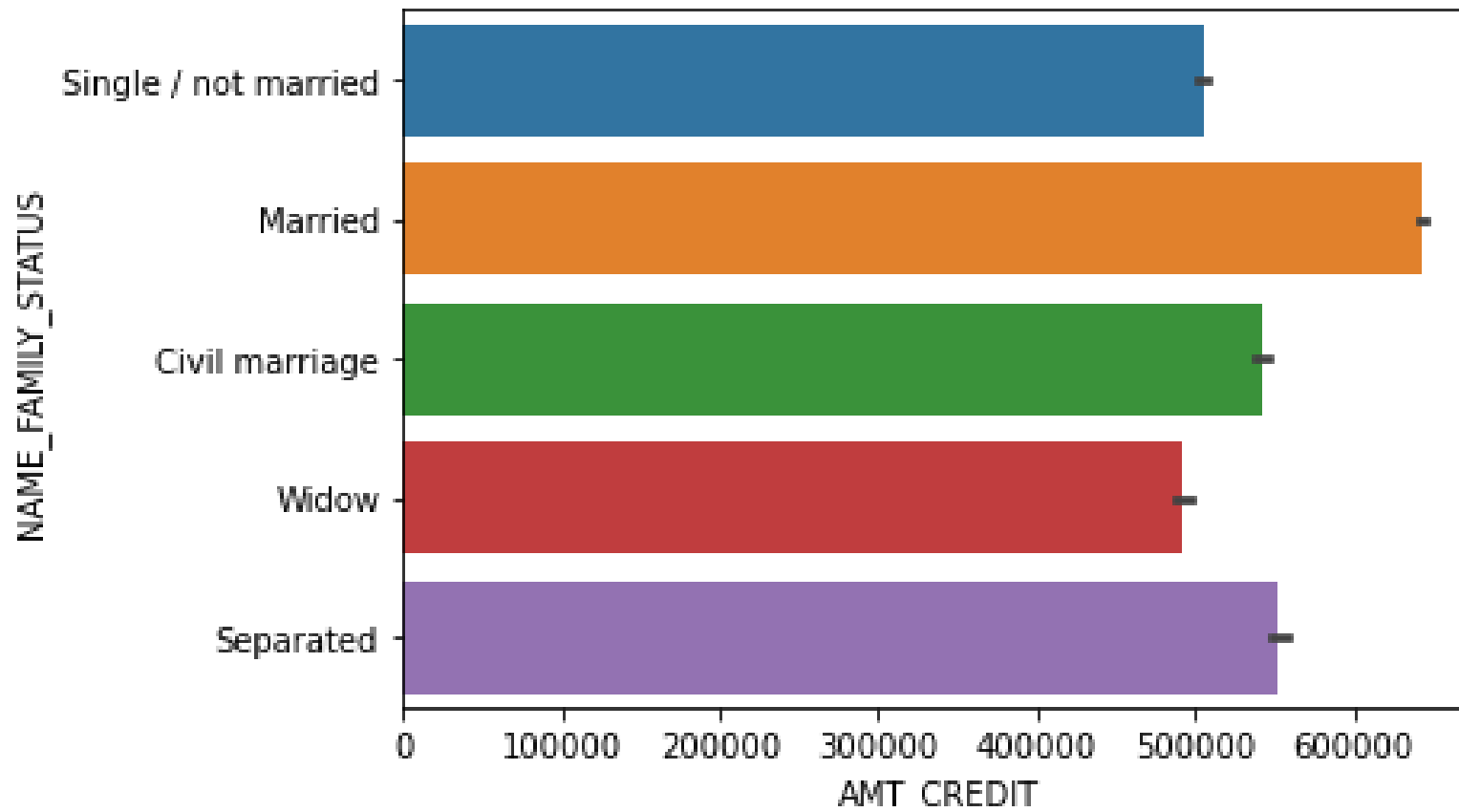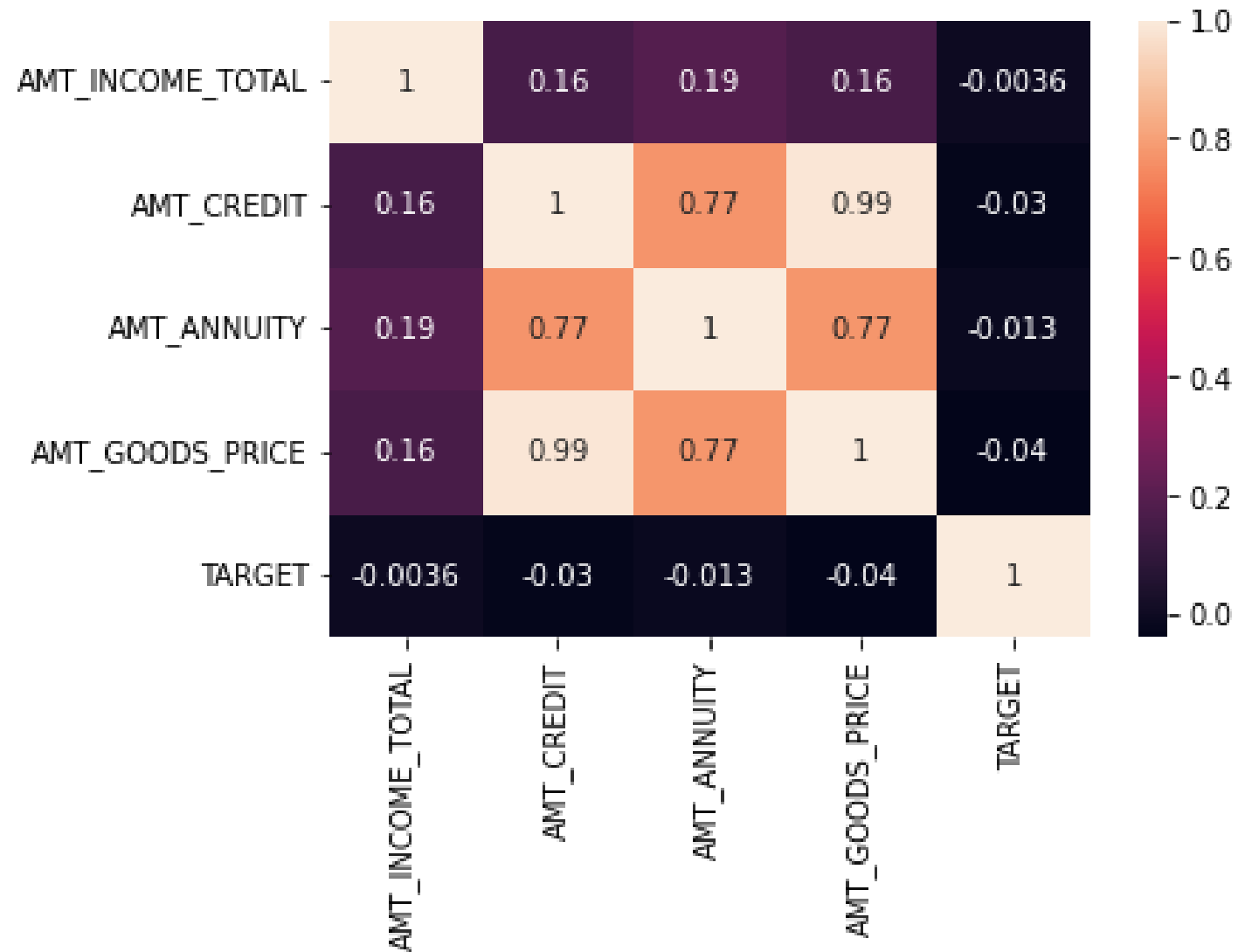- Married clients have the highest GOODS PRICE amount in demand.

Barplot

- Business man clients has the highest GOODS PRICE in demand.

- Business man clients has the highest loan amount CREDIT status among majority.

Married clients has the highest loan amount CREDIT.

ANNUITY amount, CREDIT amount and GOODS PRICE are co-related with each other.

Conclusion and Recommendation

Majority of the clients are Male Married and has their own car.

Clients has Business credit with highest amount of loans.

Total Income amount more than 150000 will not be the loan defaulters.

Business owners as client has the highest loan amount application and good price demand.

Academic degree holders are the highest income clients and not be the defaulters for the bank.