Previous Application Data Analysis

Problem statement

- Final credit amount differs with the client application amount.

Assumptions

- The data has 1670214 row entries and 37 columns to be analyzing.
- After over through with the data, it has a huge amount of null value.

Over all 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 Null values has to be deleted before starting the analyzing process.
- The analyzing done through Univariate Analysis, Bivariate Analysis and Multivariate Analysis step by step.

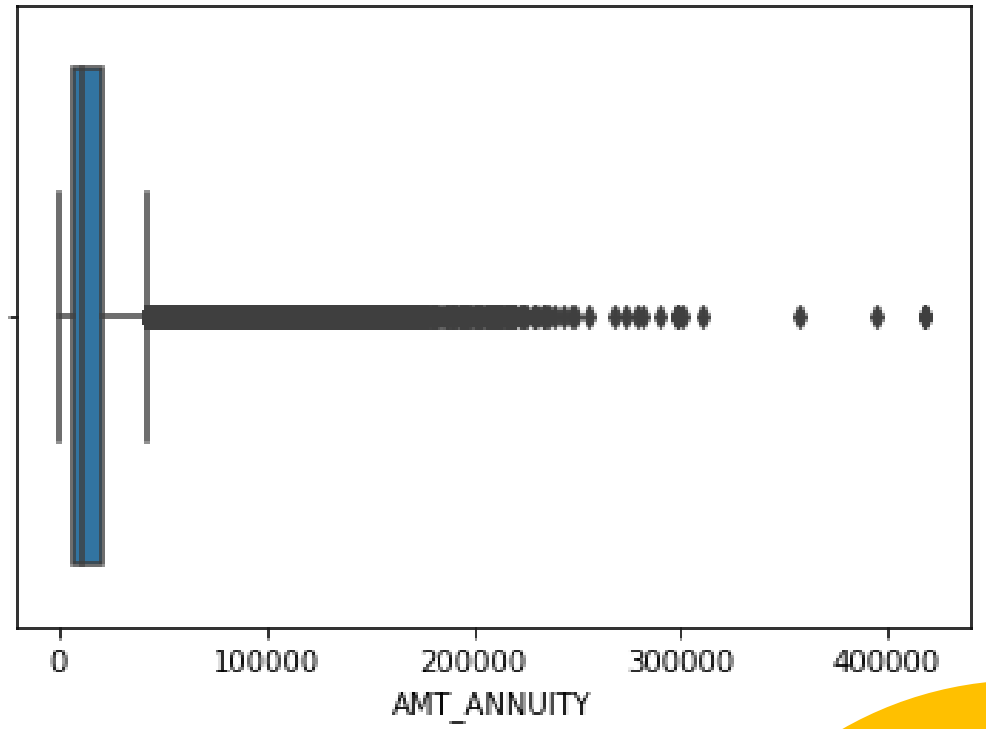
Over all approach

- The data holds 1670214 rows entries and 37 columns.
- After cleaning the data(dropping the null values) 1670214 number of rows and 26 columns left.
- More than 10 columns holds more than 40% of null values.

Relevant result with graphs

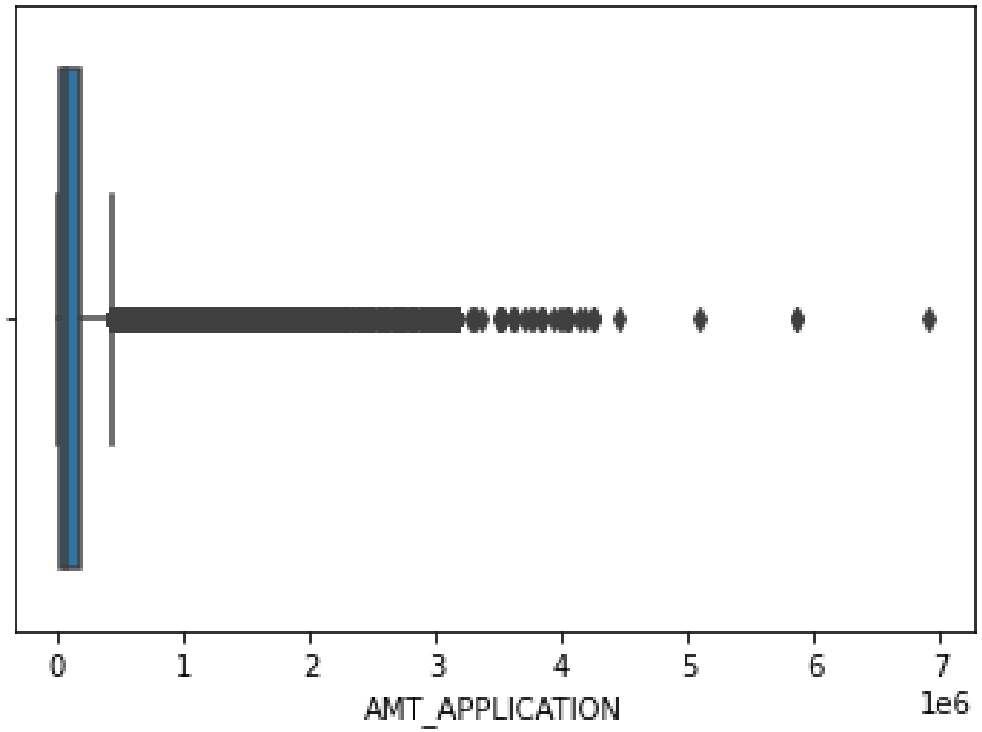
AMT_ANNUITY

count	1.297979e+06
mean	1.595512e+04
std	1.478214e+04
min	0.000000e+00
25%	6.321780e+03
50%	1.125000e+04
75%	2.065842e+04
max	4.180581e+05



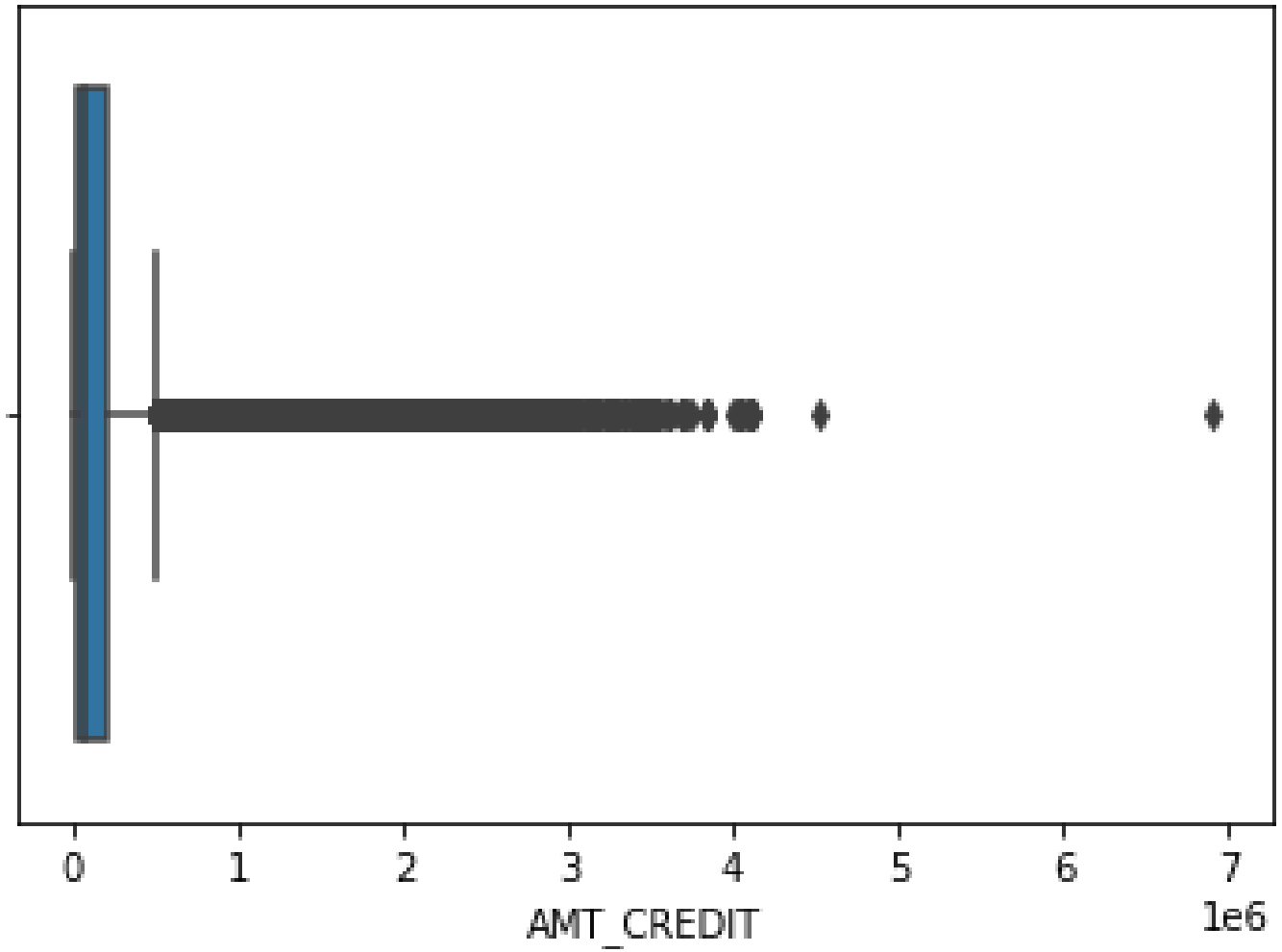
AMT_APPLICATION

count	1.670214e+06
mean	1.752339e+05
std	2.927798e+05
min	0.000000e+00
25%	1.872000e+04
50%	7.104600e+04
75%	1.803600e+05
max	6.905160e+06



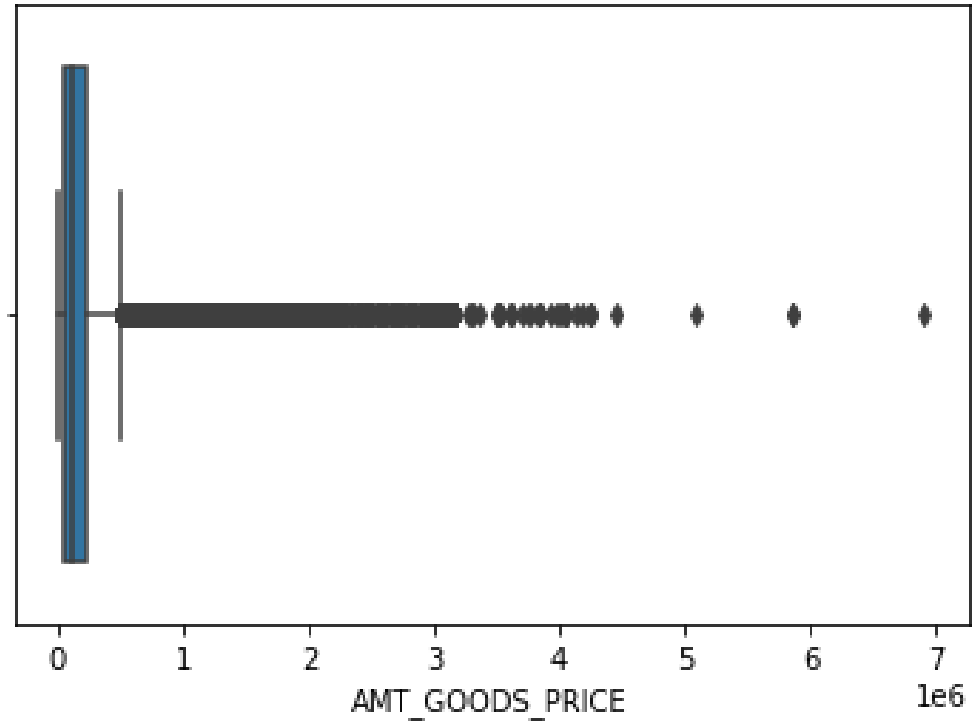
AMT_CREDIT

count	1.670213e+06
mean	1.961140e+05
std	3.185746e+05
min	0.000000e+00
25%	2.416050e+04
50%	8.054100e+04
75%	2.164185e+05
max	6.905160e+06



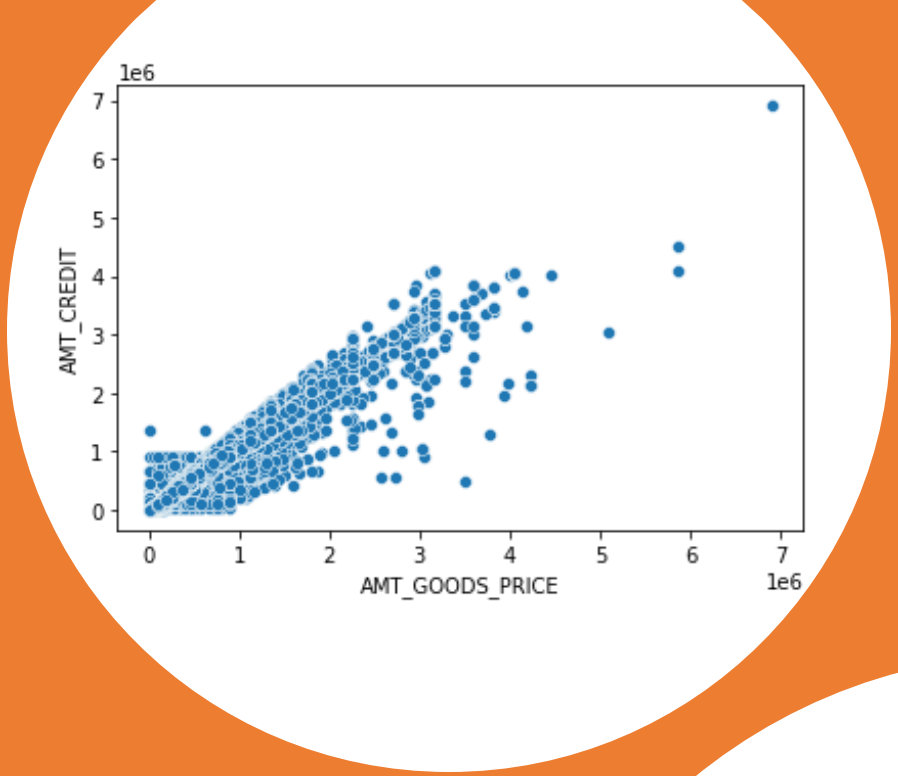
AMT_GOODS_PRICE

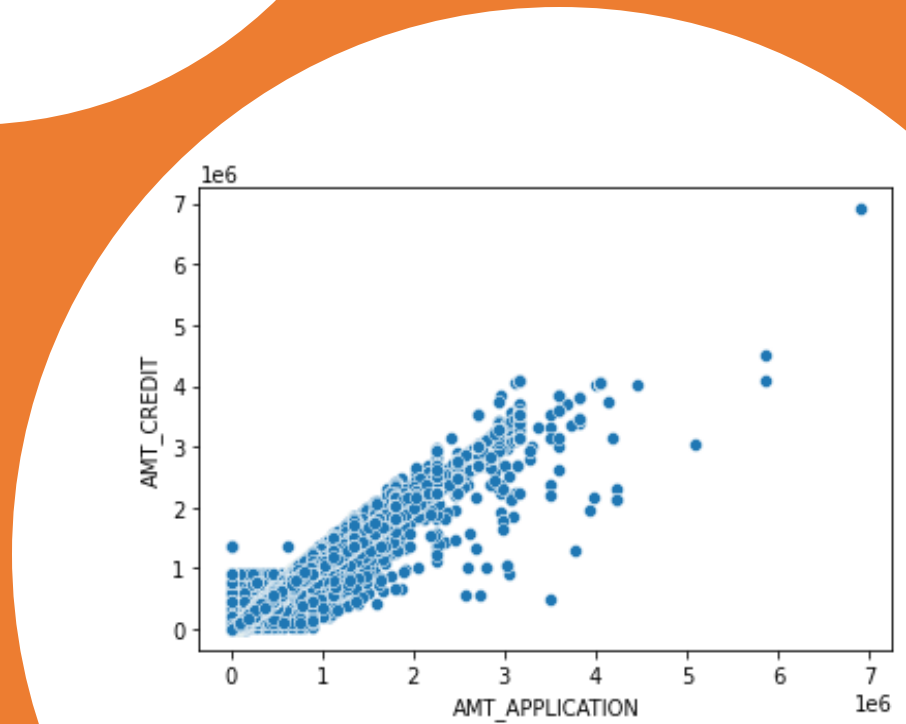
count	1.284699e+06
mean	2.278473e+05
std	3.153966e+05
min	0.000000e+00
25%	5.084100e+04
50%	1.123200e+05
75%	2.340000e+05
max	6.905160e+06



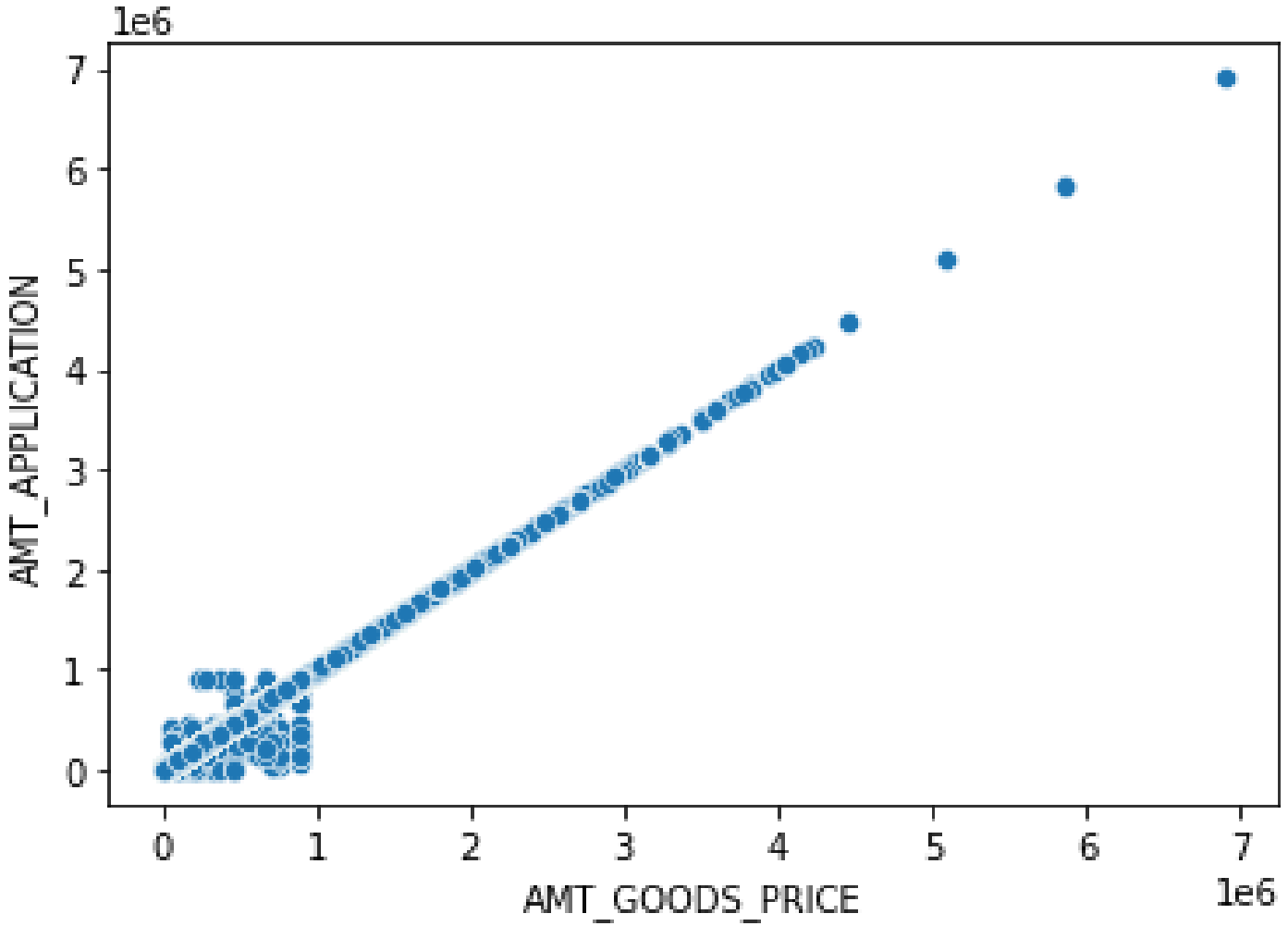
AMT_CREDIT and AMT_GOODS_PRICE are positive co-related with each other's.

AMT_CREDIT and AMT_APPLICATION are positive co-related with each other's.





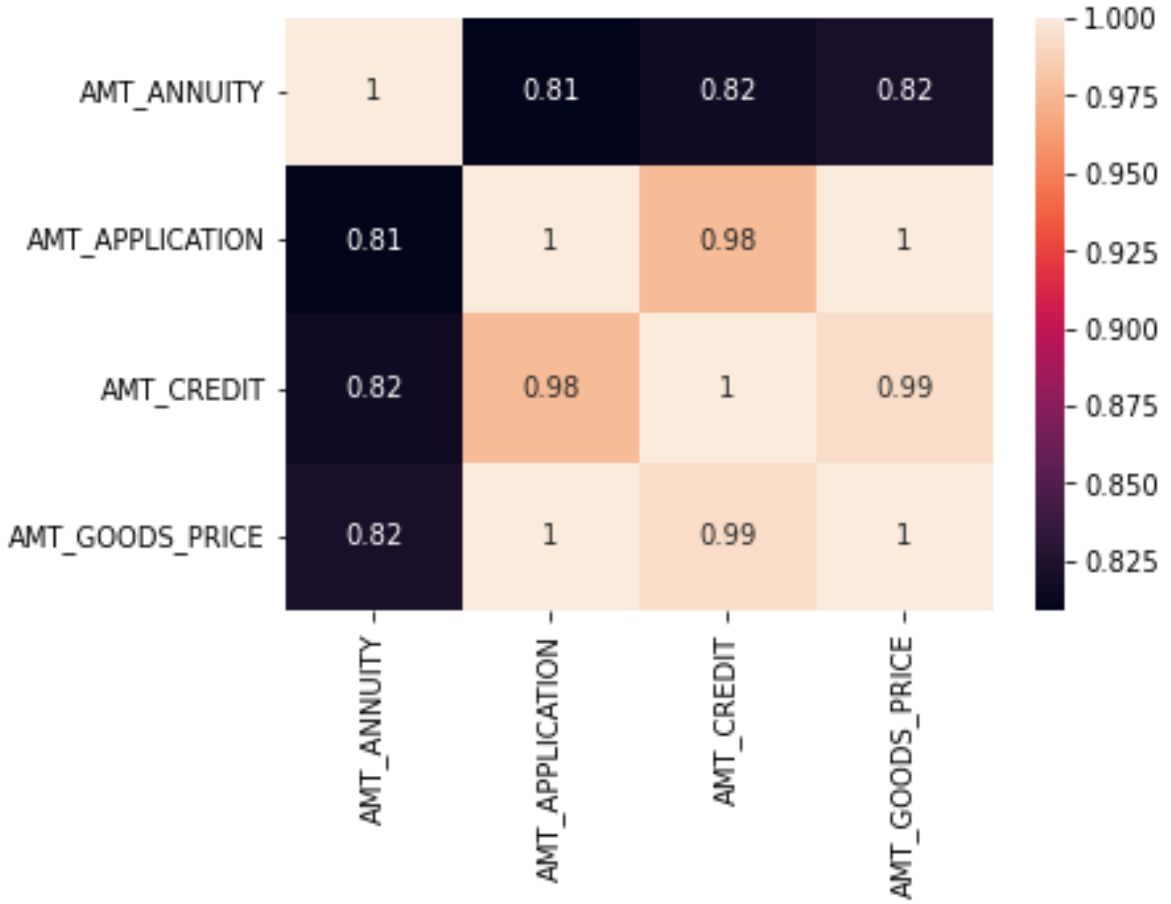
AMT_APPLICATION and AMT_GOODS_PRICE highly co-related with each other's.



Multivariate analysis

- AMT_ANNUTY
- AMT_APPLICATION
- AMT_CREDIT
- AMT_GOODS_PRICE

Accept AMT_ANNUITY other 3columns are highly co-related with each other's.



Conclusion and Recommendation

- AMT_APPLICATION, AMT_CREDIT and AMT_GOODS_PRICE are highly co-related with each other's.
- Other columns also have some linear co-relation with this three columns.