CREDIT EDA CASE STUDY

BY – 1. ANIKET CHAWARE

2. SHUBHAM GOUR

INTRODUCTION

This case study aims to give us an idea of applying EDA in a real business scenario. In this case study, we develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

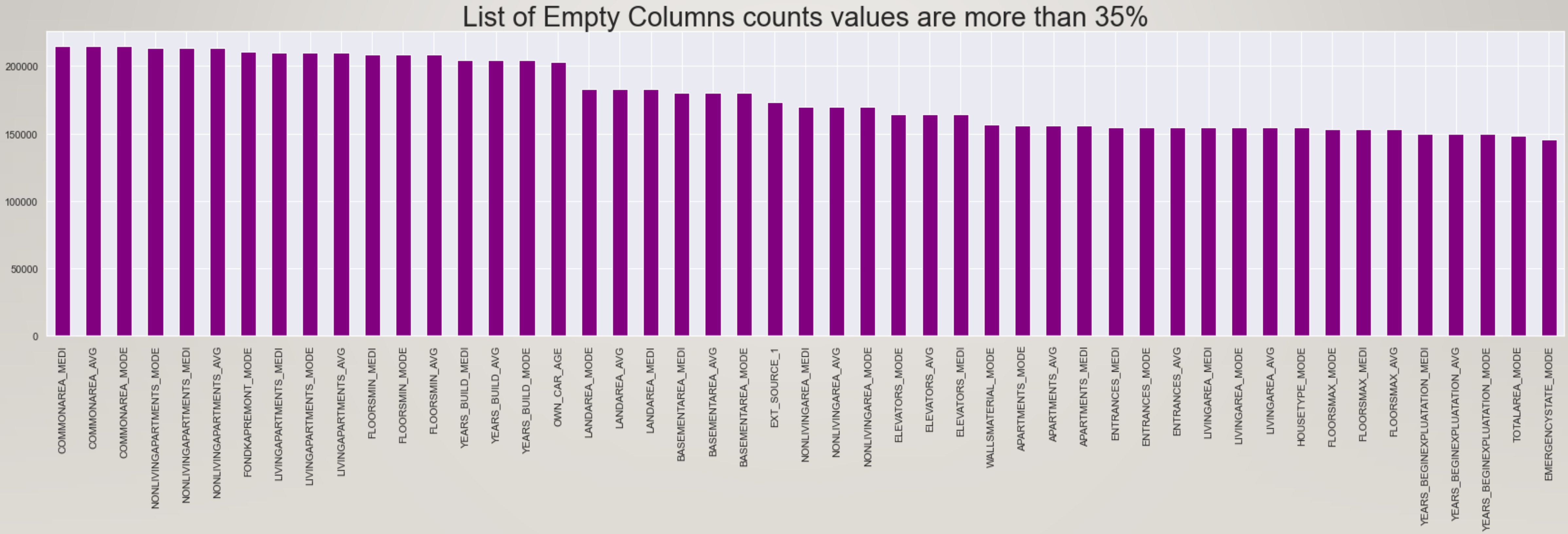
PROBLEM STATEMENT

To identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate.This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

- 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.

DATA CLEANING

LIST OF EMPTY COLUMNS VALUES ARE MORE THAN 35%



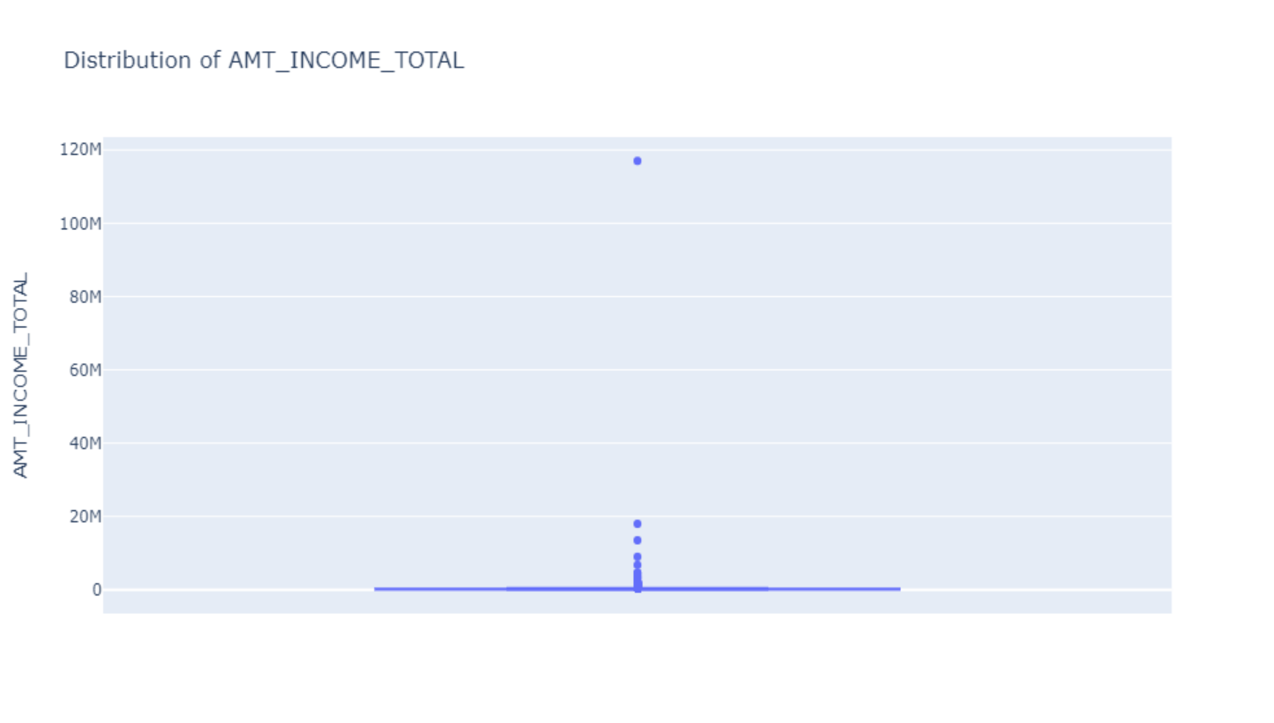
List of Empty Columns counts values are more than 35%

FINDING OUTLIERS

DISTRIBUTION OF AMT_INCOME_TOTAL

Observtion :

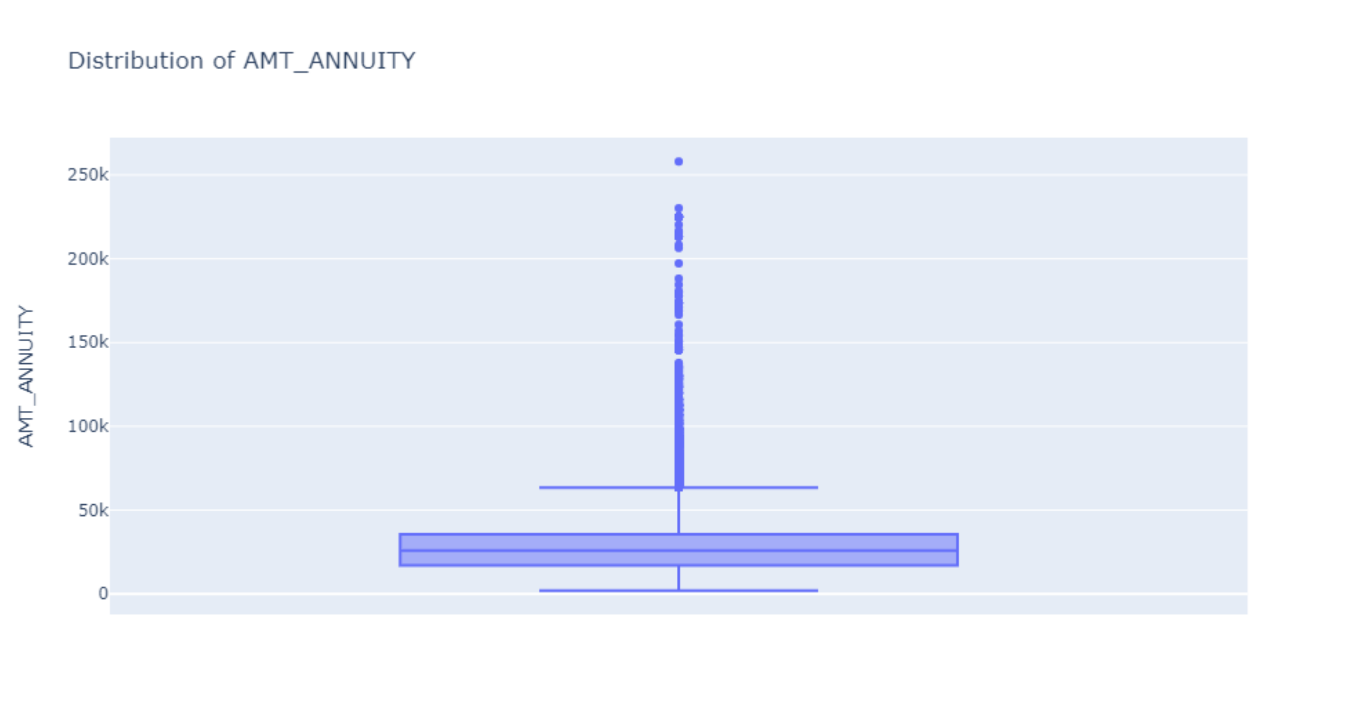
Here, in the column 'AMT_INCOME_TOTAL' which tells us the income of the client. We observe a value of 117M which is surely an outlier.



DISTRIBUTION OF AMT_ANNUITY

Observation:

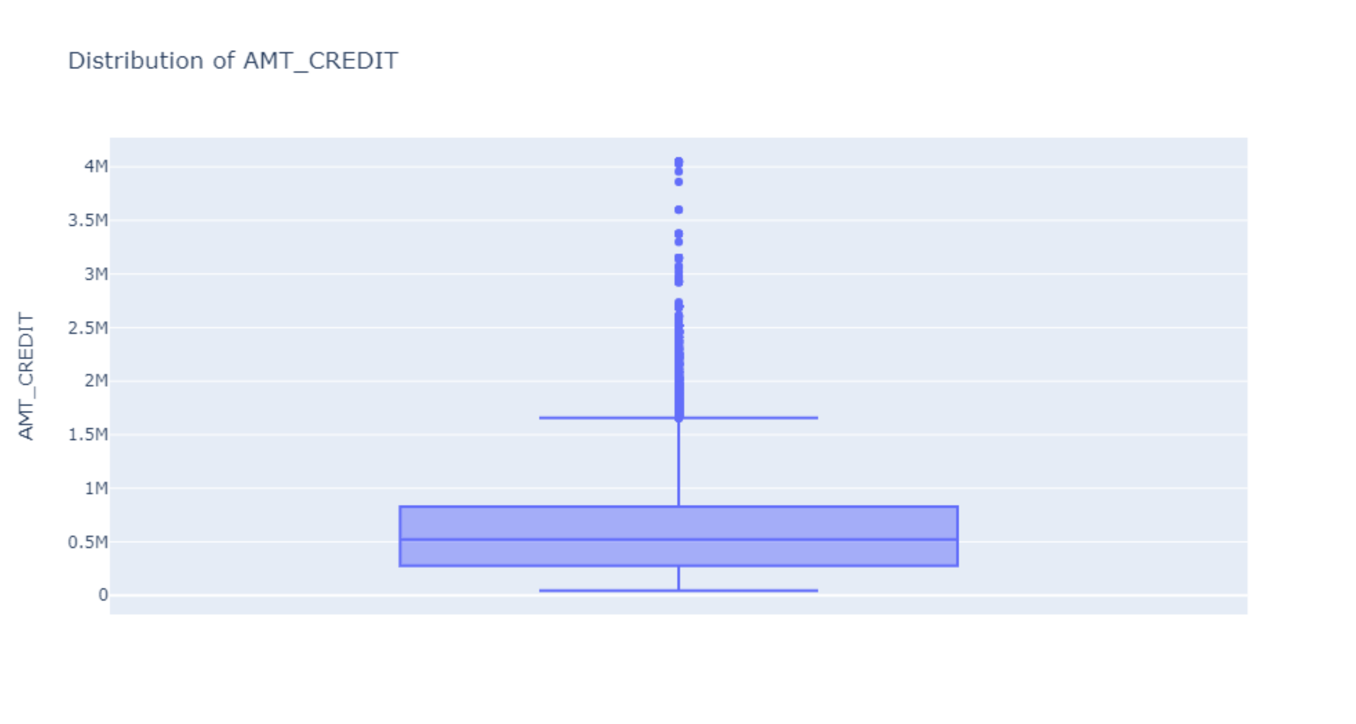
Here, in the column 'AMT_ANNUITY' which tells the loan annuity. We observe a value which is greater that 258000 which is surely an outlier.



DISTRIBUTION OF AMT_CREDIT

Observation:

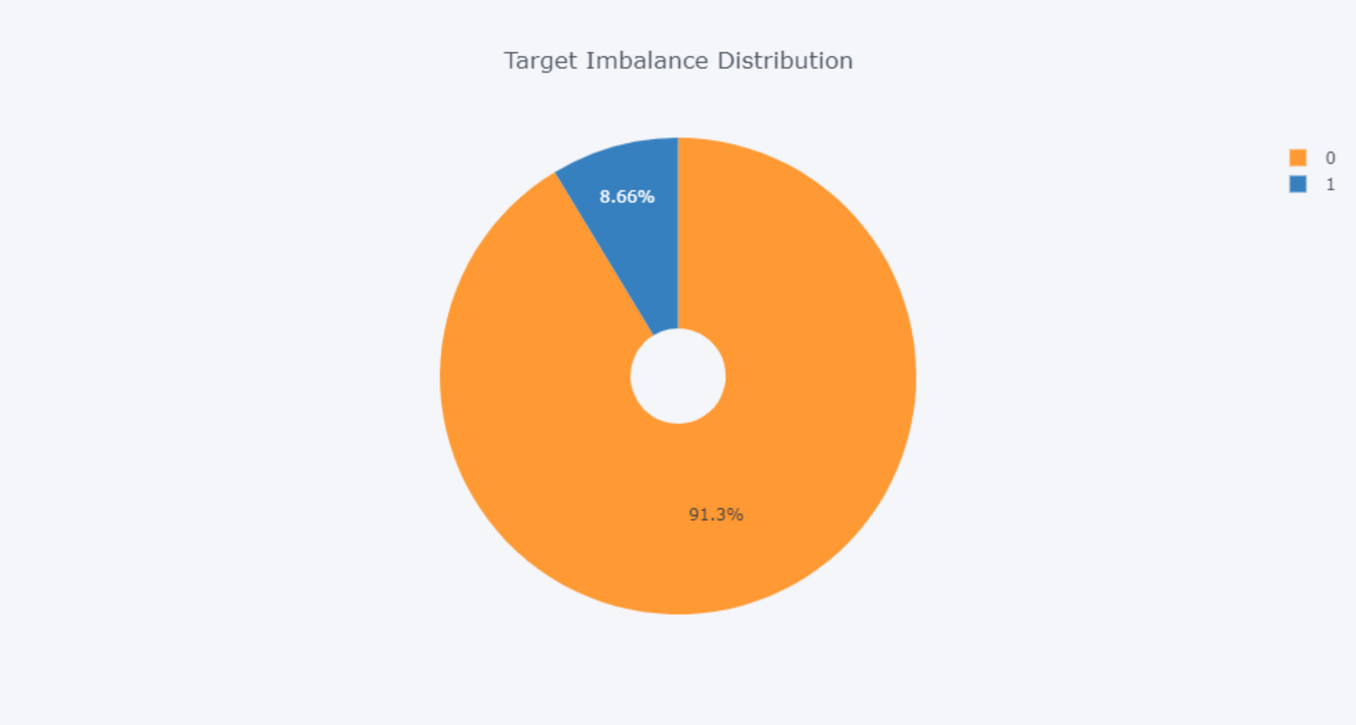
Here, in the column 'AMT_CREDIT' which tells the Credit Amount of the client. We observe a value which is greater that 4.05M which is surely an outlier.



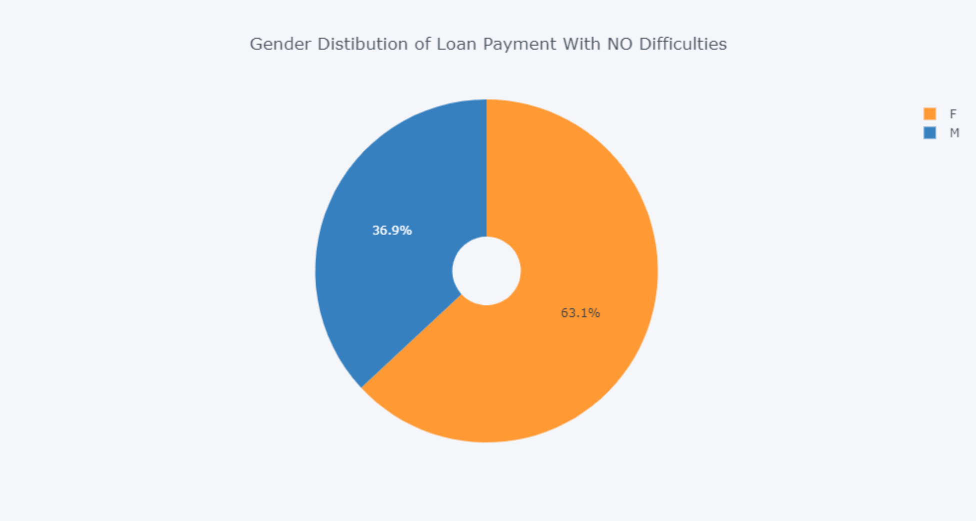
TARGET IMBALANCE DISTRIBUTION

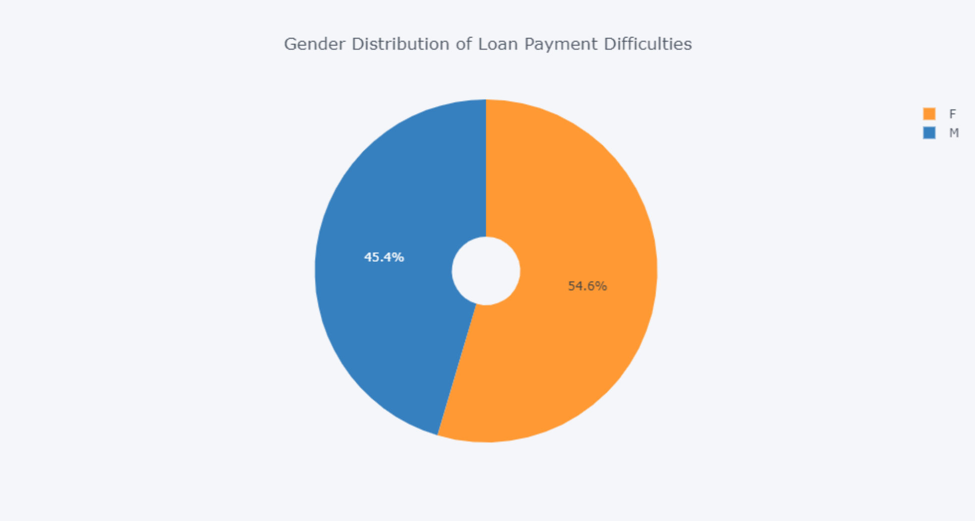
Observation:

Figure shows the Distribution of Target Imbalance and the Imbalance ratio is 10.55



GENDER DISTRIBUTION

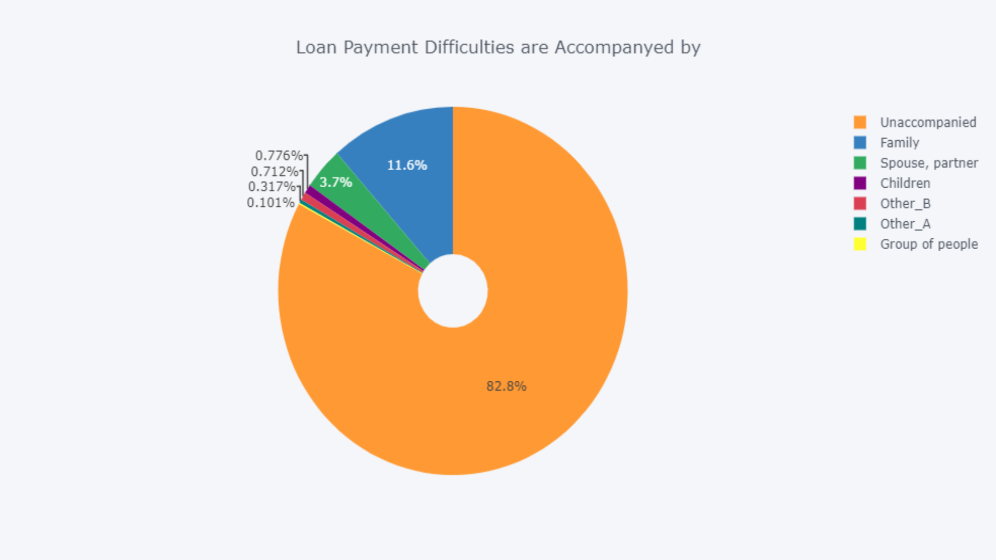




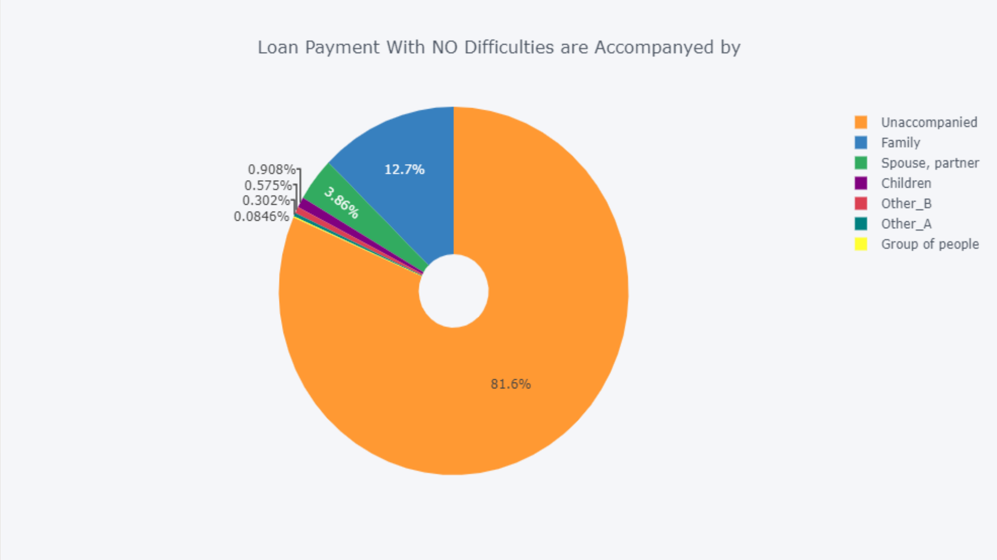
Observation :

Comparing the Payment Difficulties and Non Payment Difficulties on the basis of Gender, we observe that Females are the majority in both the cases although there is an increase in the percentage in Male Payment Difficulties from Non-Payment Difficulties.

INCOME SOURCES



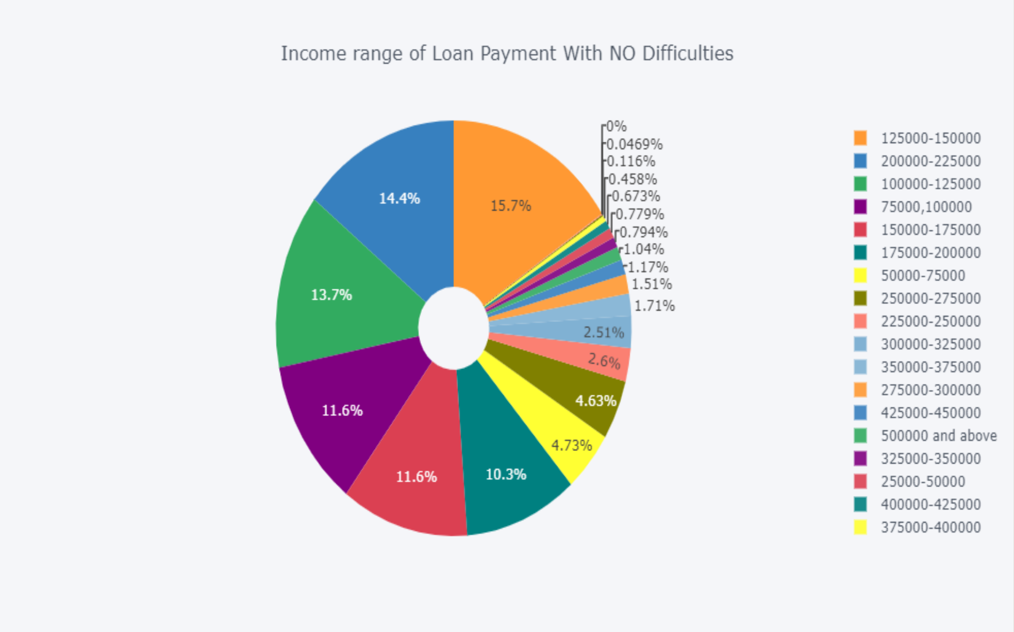
Loan Payment Difficulties are Accompanyed by



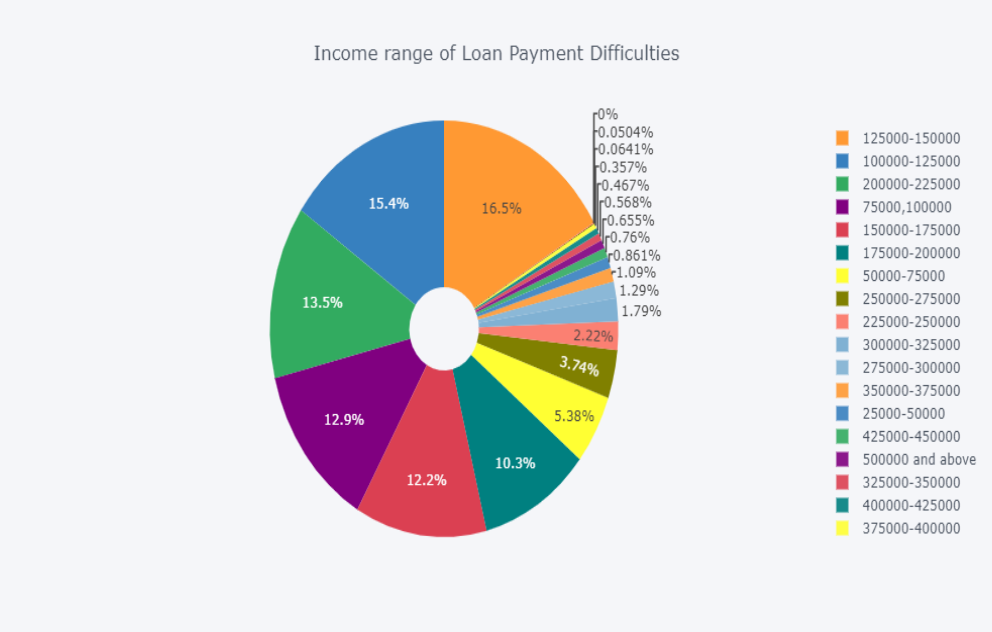
Loan Payment With NO Difficulties are Accompanyed by

Obseravtion :
We observe a decrease in the percentage of Payment Difficulties who are pentioners and an increase in the percentage of Payment Difficulties who are working when compared the percentages of both Payment Difficulties and non-Payment Difficulties.

INCOME RANGE



Income range of Loan Payment With NO Difficulties



Income range of Loan Payment Difficulties

Observation :

We observe an increase in the percentage of Loan Payment Difficulties whose income is low when compared with the percentages of Payment Difficulties and Loan-Non Payment Difficulties

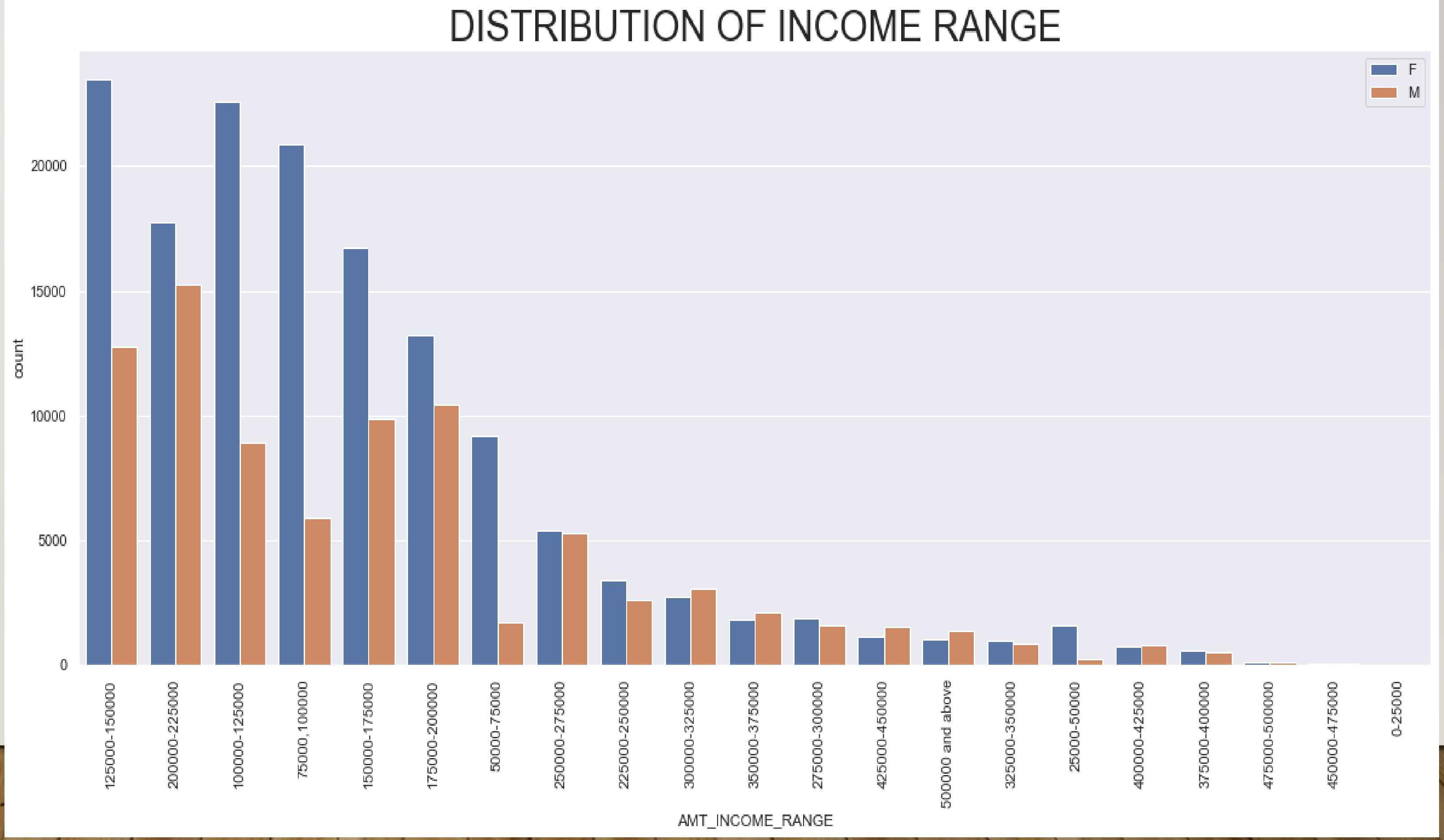
Categorical Analysis of Numerical Variables on the basis of 'Target' Variable

TARGET 0

DISTRIBUTION OF INCOME RANGE

Observation:

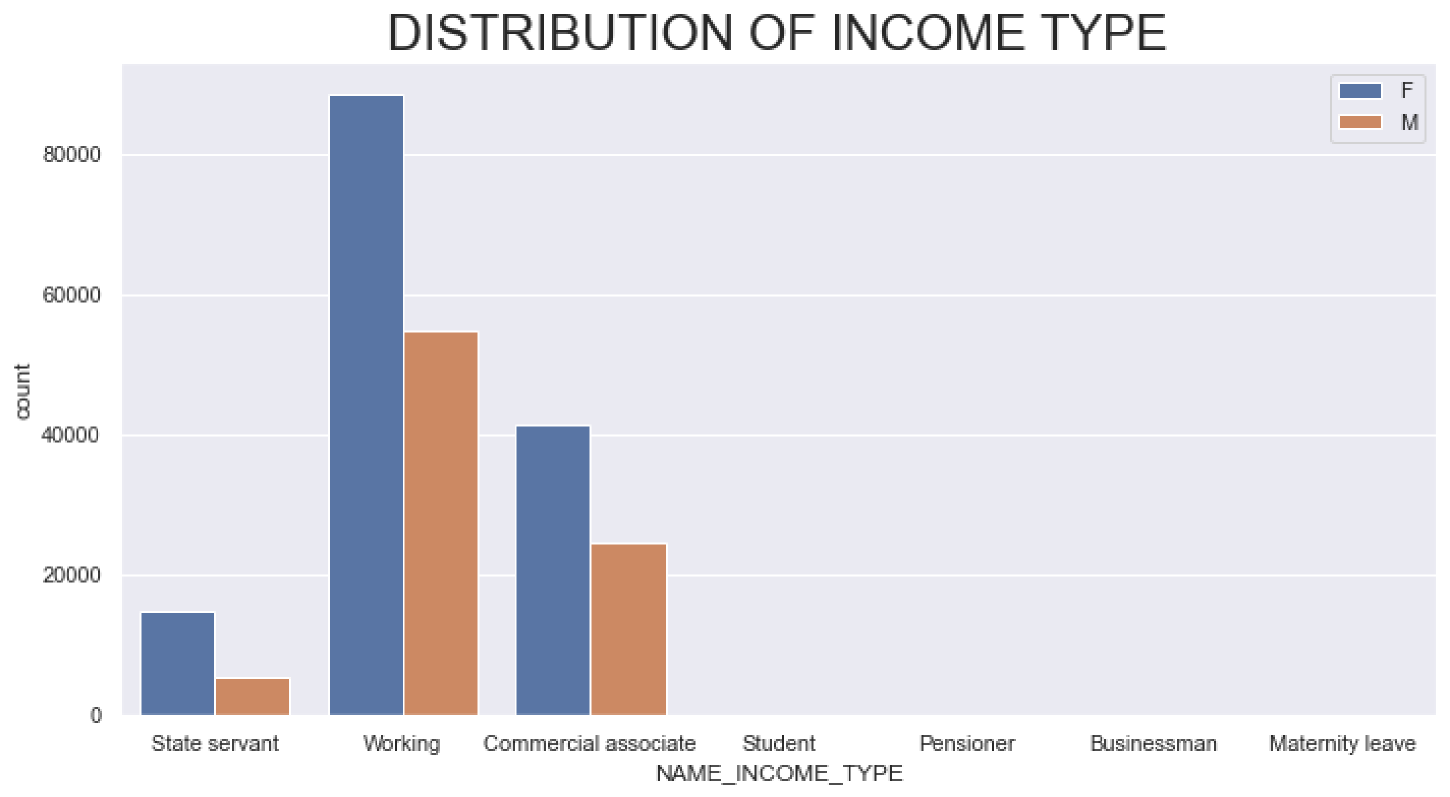
1. Female counts are higher than male.
2. Income range from 100000 to 200000 is having more number of the credits.
3. This graph show that females are more than male in having credits for that range.
4. Very less count for income range 400000 and above.



DISTRIBUTION OF INCOME TYPE

Observation:.

1. ‘Working’, ’Commercial Associate’, and ‘State Servant’ has the high the number of credits than others.
2. Females are having more number of credits than male.
3. Less number of credits for income type ‘Student’ ,'Pensioner’, ‘Businessman’ and ‘Maternity leave’.

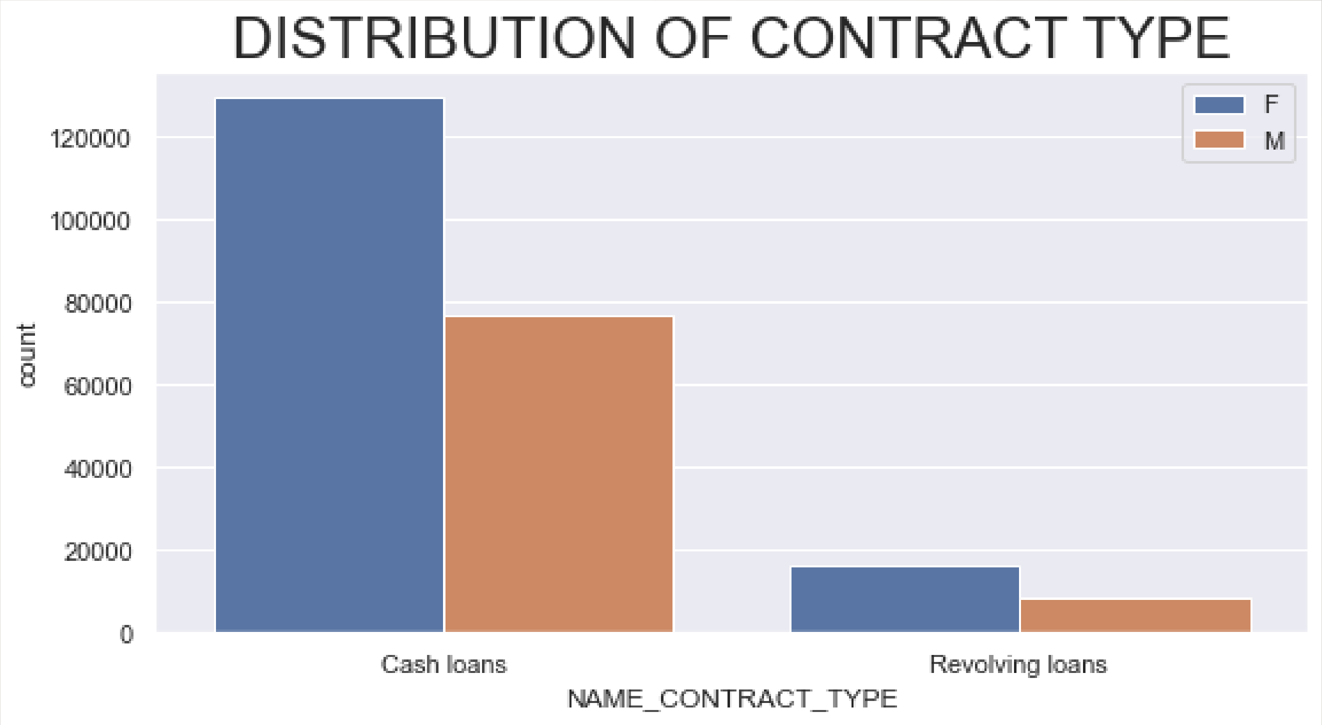


DISTRIBUTION OF CONTRACT TYPE

Observation:

1.‘Cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.

2.Females are leading in this.

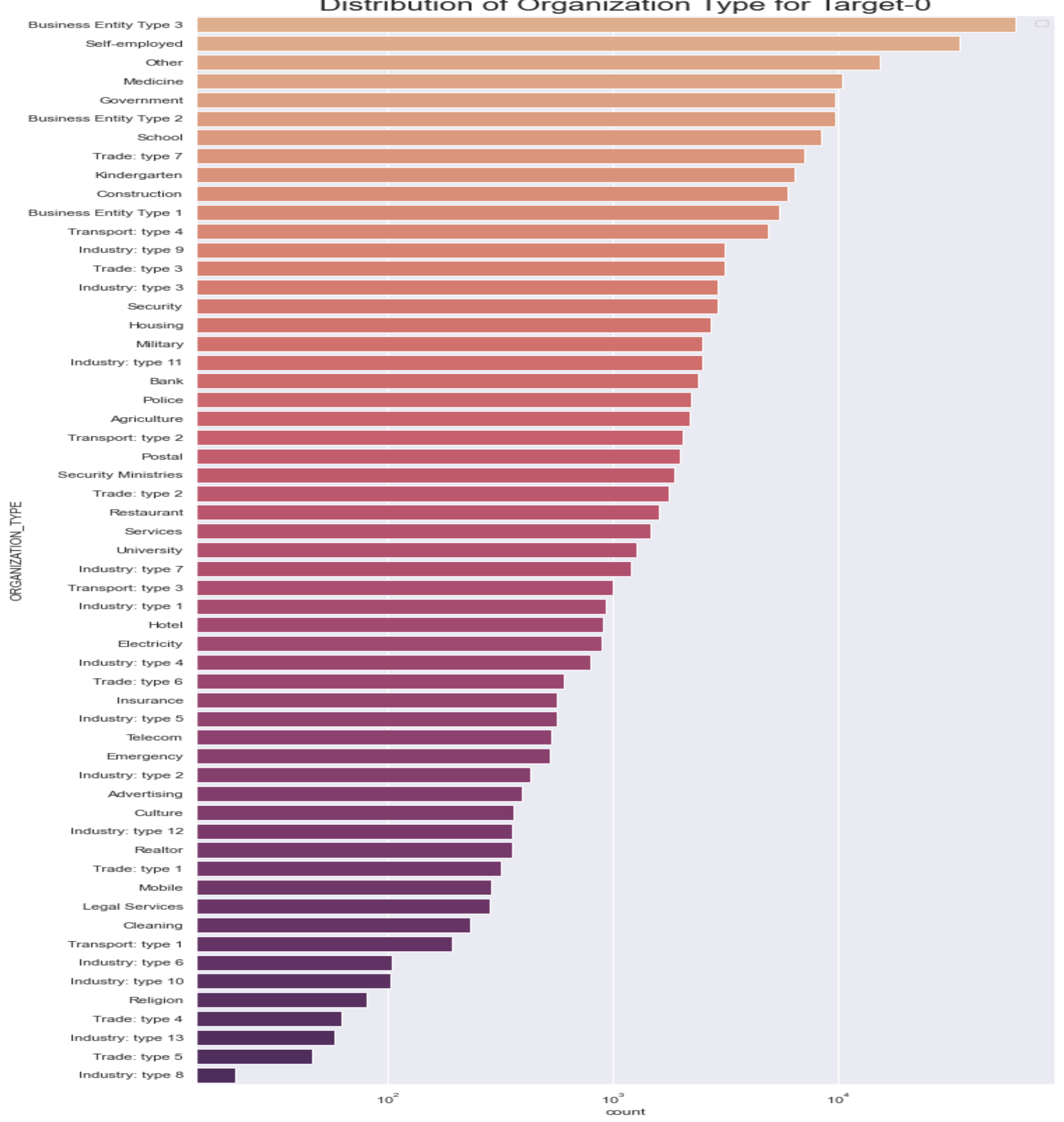


DISTRIBUTION OF ORGANIZATION TYPE

Observation:

1. Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’, ‘Other’ , ‘Medicine’ , ‘Government’ and 'Business entity Type 2'.

2. Very Less Clients from 'Industry Type 8'



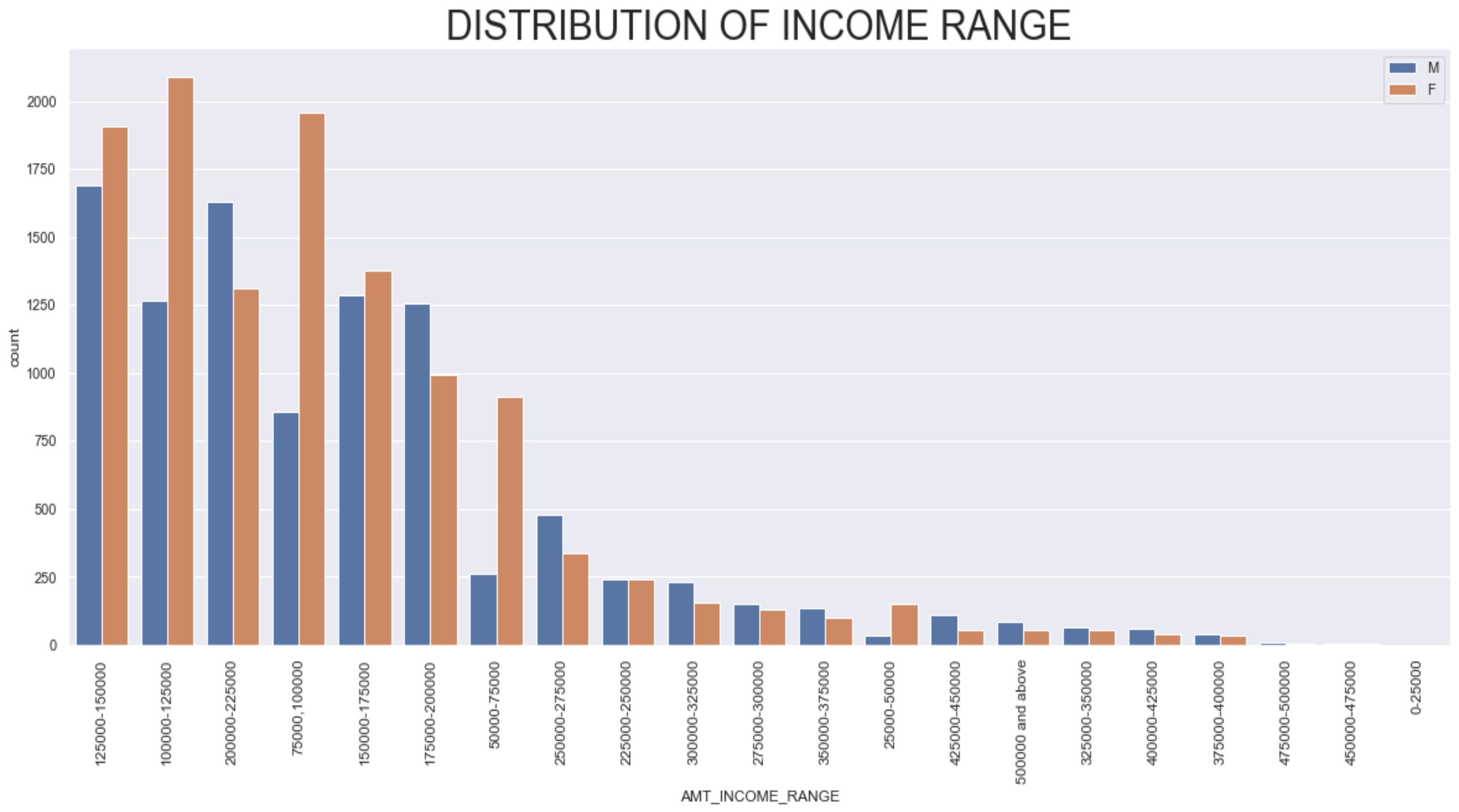
Categorical Analysis of Numerical Variables on the basis of 'Target' Variable

TARGET 1

DISTRIBUTION OF INCOME RANGE

Observation:

1. Female counts are higher than male.
2. Income range from 100000 to 200000 is having more number of credits.
3. This graph show that females are more than male in having credits for that range.
4. Very less count for income range 400000 and less.

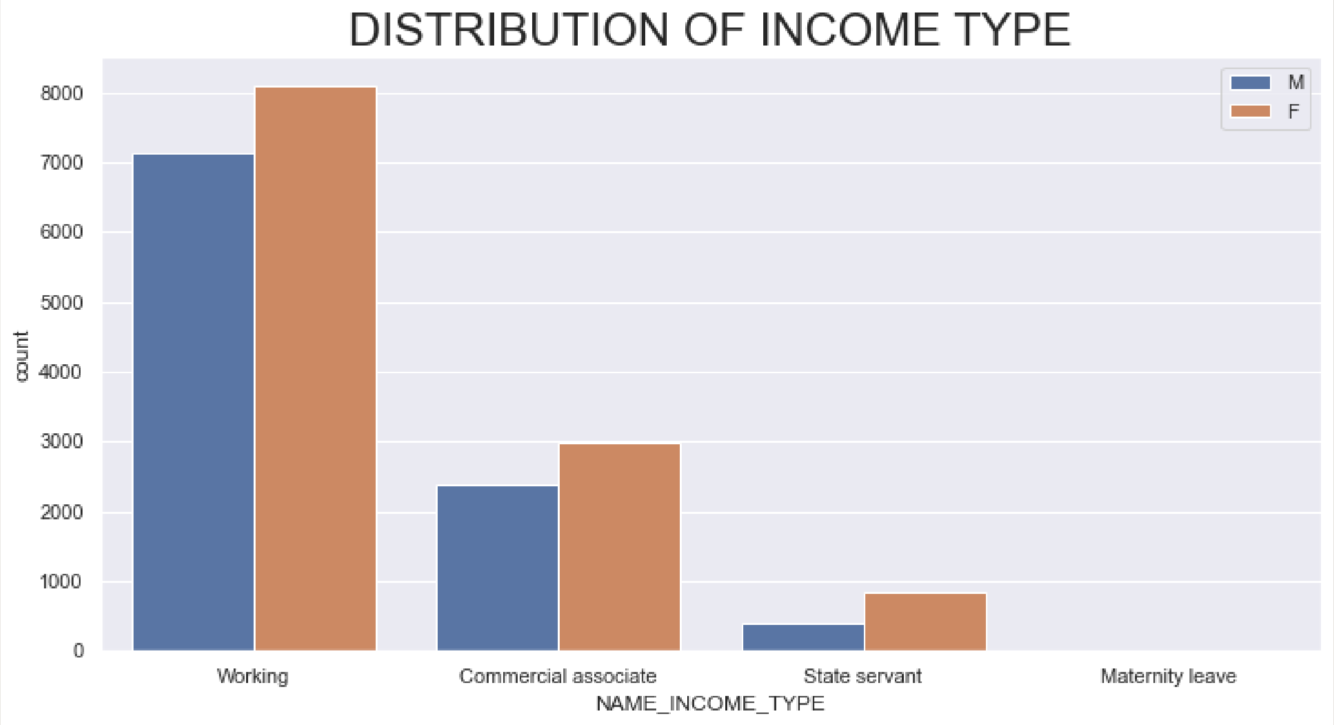


DISTRIBUTION OF INCOME TYPE

Observation:

1. ‘Working’, ’Commercial associate’, and ‘State Servant’ has higher number of credits than other i.e. ‘Maternity leave.

2. Females are having more number of credits than male.

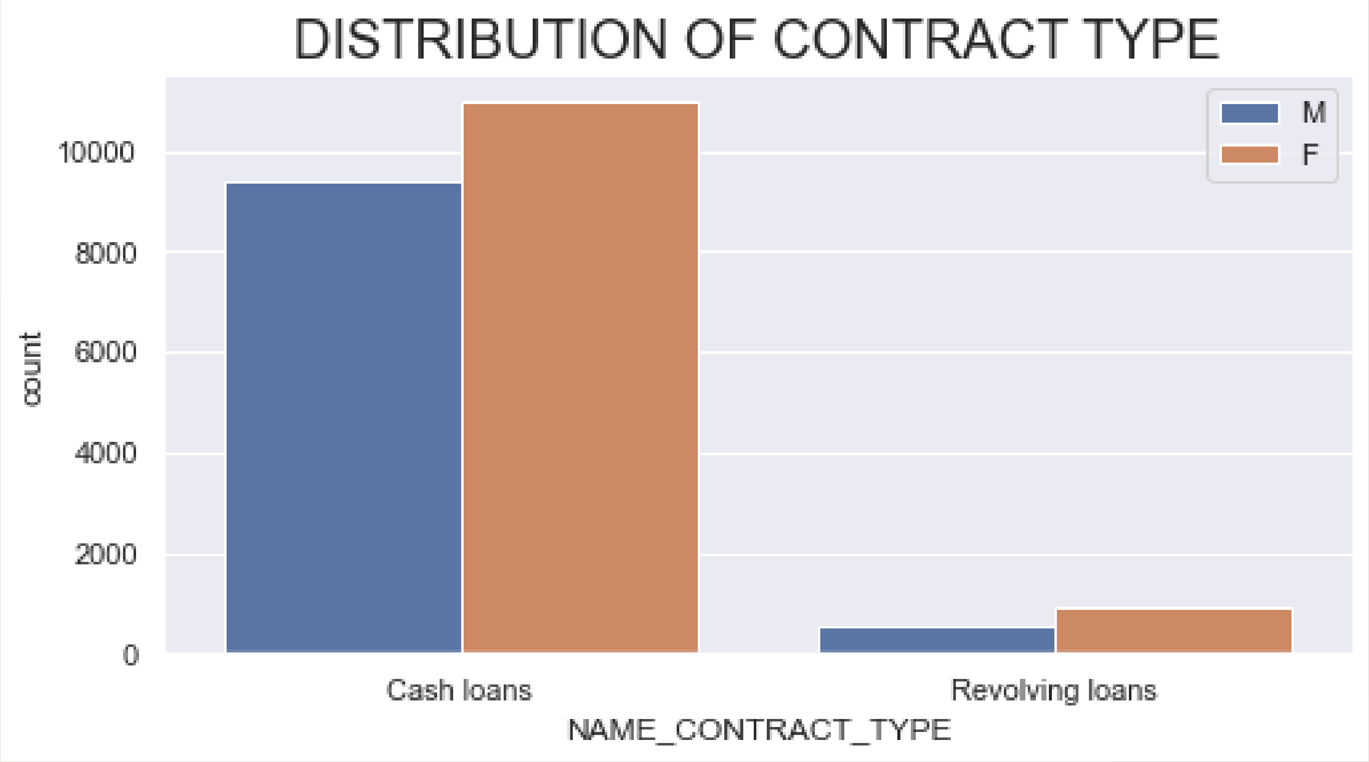


DISTRIBUTION OF CONTRACT TYPE

Observation:

1.'Cash loans' is having higher number of credits than 'Revolving loans' contract type.

2.Females are leading for applying credits.

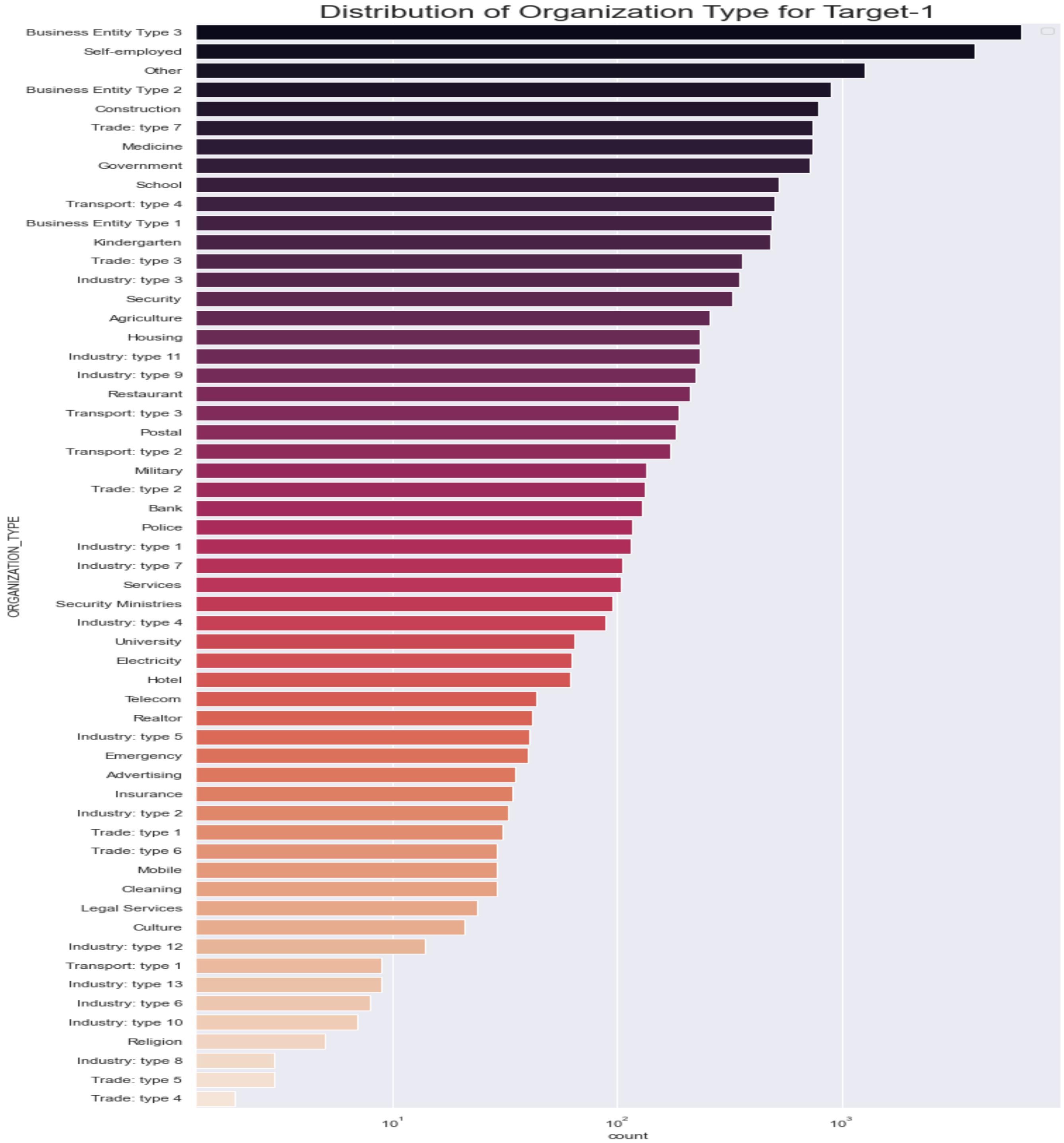


DISTRIBUTION OF ORGANIZATION TYPE

Observation:

1.Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ ,‘Self employed’, ‘Other’ and 'Business entity Type 2'.

2.Very Less Clients from 'Trade Type 4' , 'Trade Type 5' and 'Trade type 8’



CORRELATION FOR TARGET 0

Observation:

1. Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.

2. Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.

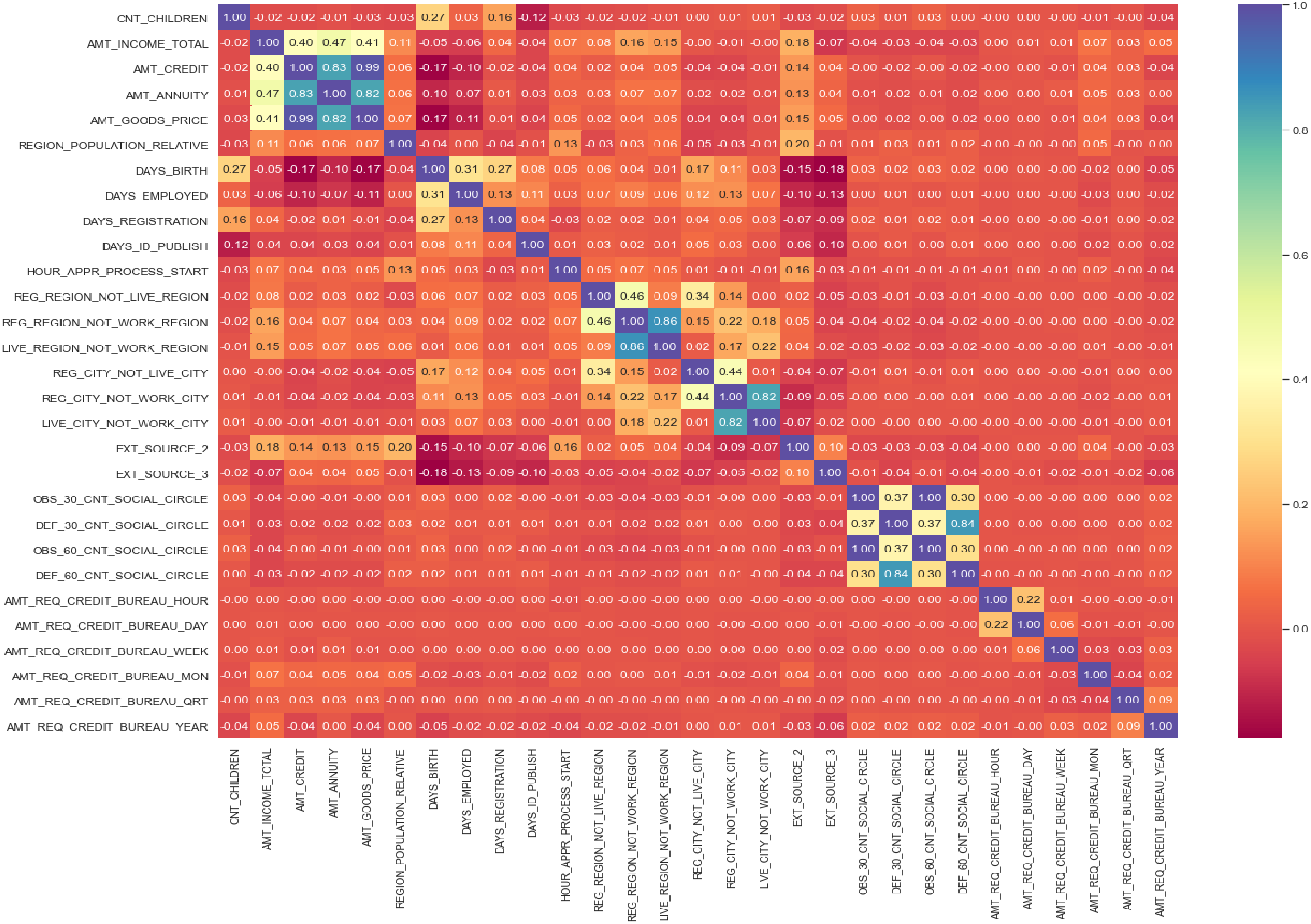
3. Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.

4. Less children client have in densely populated area.

5. The income is higher in densely populated area.

6. Credit amount is higher in densely populated area.

CORRELATION FOR TARGET 0



	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	HOUR_APPR_PROCESS_START	REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	EXT_SOURCE_2	EXT_SOURCE_3	OBS_30_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY	AMT_REQ_CREDIT_BUREAU_WEEK	AMT_REQ_CREDIT_BUREAU_MON	AMT_REQ_CREDIT_BUREAU_QRT	AMT_REQ_CREDIT_BUREAU_YEAR
CNT_CHILDREN	1.00	-0.02	-0.02	-0.01	-0.03	-0.03	0.27	0.03	0.16	-0.12	-0.03	-0.02	-0.02	-0.01	0.00	0.01	0.01	-0.03	-0.02	0.03	0.01	0.03	0.00	-0.00	0.00	-0.00	-0.01	-0.00	-0.04
AMT_INCOME_TOTAL	-0.02	1.00	0.40	0.47	0.41	0.11	-0.05	-0.06	0.04	-0.04	0.07	0.08	0.16	0.15	-0.00	-0.01	-0.00	0.18	-0.07	-0.04	-0.03	-0.04	-0.03	0.00	0.01	0.01	0.07	0.03	0.05
AMT_CREDIT	-0.02	0.40	1.00	0.83	0.99	0.06	-0.17	-0.10	-0.02	-0.04	0.04	0.02	0.04	0.05	-0.04	-0.04	-0.01	0.14	0.04	-0.00	-0.02	-0.00	-0.02	-0.00	0.00	-0.01	0.04	0.03	-0.04
AMT_ANNUITY	-0.01	0.47	0.83	1.00	0.82	0.06	-0.10	-0.07	0.01	-0.03	0.03	0.03	0.07	0.07	-0.02	-0.02	-0.01	0.13	0.04	-0.01	-0.02	-0.01	-0.02	0.00	0.00	0.01	0.05	0.03	0.00
AMT_GOODS_PRICE	-0.03	0.41	0.99	0.82	1.00	0.07	-0.17	-0.11	-0.01	-0.04	0.05	0.02	0.04	0.05	-0.04	-0.04	-0.01	0.15	0.05	-0.00	-0.02	-0.00	-0.02	-0.00	0.00	-0.01	0.04	0.03	-0.04
REGION_POPULATION_RELATIVE	-0.03	0.11	0.06	0.06	0.07	1.00	-0.04	0.00	-0.04	-0.01	0.13	-0.03	0.03	0.06	-0.05	-0.03	-0.01	0.20	-0.01	0.01	0.03	0.01	0.02	-0.00	-0.00	-0.00	0.05	-0.00	0.00
DAYS_BIRTH	0.27	-0.05	-0.17	-0.10	-0.17	-0.04	1.00	0.31	0.27	0.08	0.05	0.06	0.04	0.01	0.17	0.11	0.03	-0.15	-0.18	0.03	0.02	0.03	0.02	0.00	0.00	-0.00	-0.02	0.00	-0.05
DAYS_EMPLOYED	0.03	-0.06	-0.10	-0.07	-0.11	0.00	0.31	1.00	0.13	0.11	0.03	0.07	0.09	0.06	0.12	0.13	0.07	-0.10	-0.13	0.00	0.01	0.00	0.01	-0.00	0.00	-0.00	-0.03	0.00	-0.02
DAYS_REGISTRATION	0.16	0.04	-0.02	0.01	-0.01	-0.04	0.27	0.13	1.00	0.04	-0.03	0.02	0.02	0.01	0.04	0.05	0.03	-0.07	-0.09	0.02	0.01	0.02	0.01	-0.00	0.00	-0.00	-0.01	0.00	-0.02
DAYS_ID_PUBLISH	-0.12	-0.04	-0.04	-0.03	-0.04	-0.01	0.08	0.11	0.04	1.00	0.01	0.03	0.02	0.01	0.05	0.03	0.00	-0.06	-0.10	-0.00	0.01	-0.00	0.01	0.00	0.00	-0.00	-0.02	-0.00	-0.02
HOUR_APPR_PROCESS_START	-0.03	0.07	0.04	0.03	0.05	0.13	0.05	0.03	-0.03	0.01	1.00	0.05	0.07	0.05	0.01	-0.01	-0.01	0.16	-0.03	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.00	0.02	-0.00	-0.04
REG_REGION_NOT_LIVE_REGION	-0.02	0.08	0.02	0.03	0.02	-0.03	0.06	0.07	0.02	0.03	0.05	1.00	0.46	0.09	0.34	0.14	0.00	0.02	-0.05	-0.03	-0.01	-0.03	-0.01	-0.00	-0.00	0.00	0.00	-0.00	-0.02
REG_REGION_NOT_WORK_REGION	-0.02	0.16	0.04	0.07	0.04	0.03	0.04	0.09	0.02	0.02	0.07	0.46	1.00	0.86	0.15	0.22	0.18	0.05	-0.04	-0.04	-0.02	-0.04	-0.02	-0.00	-0.00	-0.00	0.00	-0.00	-0.02
LIVE_REGION_NOT_WORK_REGION	-0.01	0.15	0.05	0.07	0.05	0.06	0.01	0.06	0.01	0.01	0.05	0.09	0.86	1.00	0.02	0.17	0.22	0.04	-0.02	-0.03	-0.02	-0.03	-0.02	-0.00	-0.00	-0.00	0.01	-0.00	-0.01
REG_CITY_NOT_LIVE_CITY	0.00	-0.00	-0.04	-0.02	-0.04	-0.05	0.17	0.12	0.04	0.05	0.01	0.34	0.15	0.02	1.00	0.44	0.01	-0.04	-0.07	-0.01	0.01	-0.01	0.01	0.00	0.00	-0.00	-0.01	0.00	0.00
REG_CITY_NOT_WORK_CITY	0.01	-0.01	-0.04	-0.02	-0.04	-0.03	0.11	0.13	0.05	0.03	-0.01	0.14	0.22	0.17	0.44	1.00	0.82	-0.09	-0.05	-0.00	0.00	-0.00	0.01	0.00	-0.00	-0.00	-0.02	-0.00	0.01
LIVE_CITY_NOT_WORK_CITY	0.01	-0.00	-0.01	-0.01	-0.01	-0.01	0.03	0.07	0.03	0.00	-0.01	0.00	0.18	0.22	0.01	0.82	1.00	-0.07	-0.02	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.00	0.01
EXT_SOURCE_2	-0.03	0.18	0.14	0.13	0.15	0.20	-0.15	-0.10	-0.07	-0.06	0.16	0.02	0.05	0.04	-0.04	-0.09	-0.07	1.00	0.10	-0.03	-0.03	-0.03	-0.04	-0.00	0.00	-0.00	0.04	-0.00	-0.03
EXT_SOURCE_3	-0.02	-0.07	0.04	0.04	0.05	-0.01	-0.18	-0.13	-0.09	-0.10	-0.03	-0.05	-0.04	-0.02	-0.07	-0.05	-0.02	0.10	1.00	-0.01	-0.04	-0.01	-0.04	-0.00	-0.01	-0.02	-0.01	-0.02	-0.06
OBS_30_CNT_SOCIAL_CIRCLE	0.03	-0.04	-0.00	-0.01	-0.00	0.01	0.03	0.00	0.02	-0.00	-0.01	-0.03	-0.04	-0.03	-0.01	-0.00	0.00	-0.03	-0.01	1.00	0.37	1.00	0.30	0.00	-0.00	-0.00	0.00	0.00	0.02
DEF_30_CNT_SOCIAL_CIRCLE	0.01	-0.03	-0.02	-0.02	-0.02	0.03	0.02	0.01	0.01	0.01	-0.01	-0.01	-0.02	-0.02	0.01	0.00	-0.00	-0.03	-0.04	0.37	1.00	0.37	0.84	-0.00	-0.00	-0.00	0.00	-0.00	0.02
OBS_60_CNT_SOCIAL_CIRCLE	0.03	-0.04	-0.00	-0.01	-0.00	0.01	0.03	0.00	0.02	-0.00	-0.01	-0.03	-0.04	-0.03	-0.01	-0.00	0.00	-0.03	-0.01	1.00	0.37	1.00	0.30	0.00	-0.00	-0.00	0.00	0.00	0.02
DEF_60_CNT_SOCIAL_CIRCLE	0.00	-0.03	-0.02	-0.02	-0.02	0.02	0.02	0.01	0.01	0.01	-0.01	-0.01	-0.02	-0.02	0.01	0.01	-0.00	-0.04	-0.04	0.30	0.84	0.30	1.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.02
AMT_REQ_CREDIT_BUREAU_HOUR	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.01	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	1.00	0.22	0.01	-0.00	-0.00	-0.01
AMT_REQ_CREDIT_BUREAU_DAY	0.00	0.01	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.01	-0.00	-0.00	-0.00	-0.00	0.22	1.00	0.06	-0.01	-0.01	-0.00
AMT_REQ_CREDIT_BUREAU_WEEK	-0.00	0.01	-0.01	0.01	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.02	-0.00	-0.00	-0.00	-0.00	0.01	0.06	1.00	-0.03	-0.03	0.03
AMT_REQ_CREDIT_BUREAU_MON	-0.01	0.07	0.04	0.05	0.04	0.05	-0.02	-0.03	-0.01	-0.02	0.02	0.00	0.00	0.01	-0.01	-0.02	-0.01	0.04	-0.01	0.00	0.00	0.00	-0.00	-0.00	-0.01	-0.03	1.00	-0.04	0.02
AMT_REQ_CREDIT_BUREAU_QRT	-0.00	0.03	0.03	0.03	0.03	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.02	0.00	-0.00	0.00	-0.00	-0.00	-0.01	-0.03	-0.04	1.00	0.09
AMT_REQ_CREDIT_BUREAU_YEAR	-0.04	0.05	-0.04	0.00	-0.04	0.00	-0.05	-0.02	-0.02	-0.02	-0.04	-0.02	-0.02	-0.01	0.00	0.01	0.01	-0.03	-0.06	0.02	0.02	0.02	0.02	-0.01	-0.00	0.03	0.02	0.09	1.00

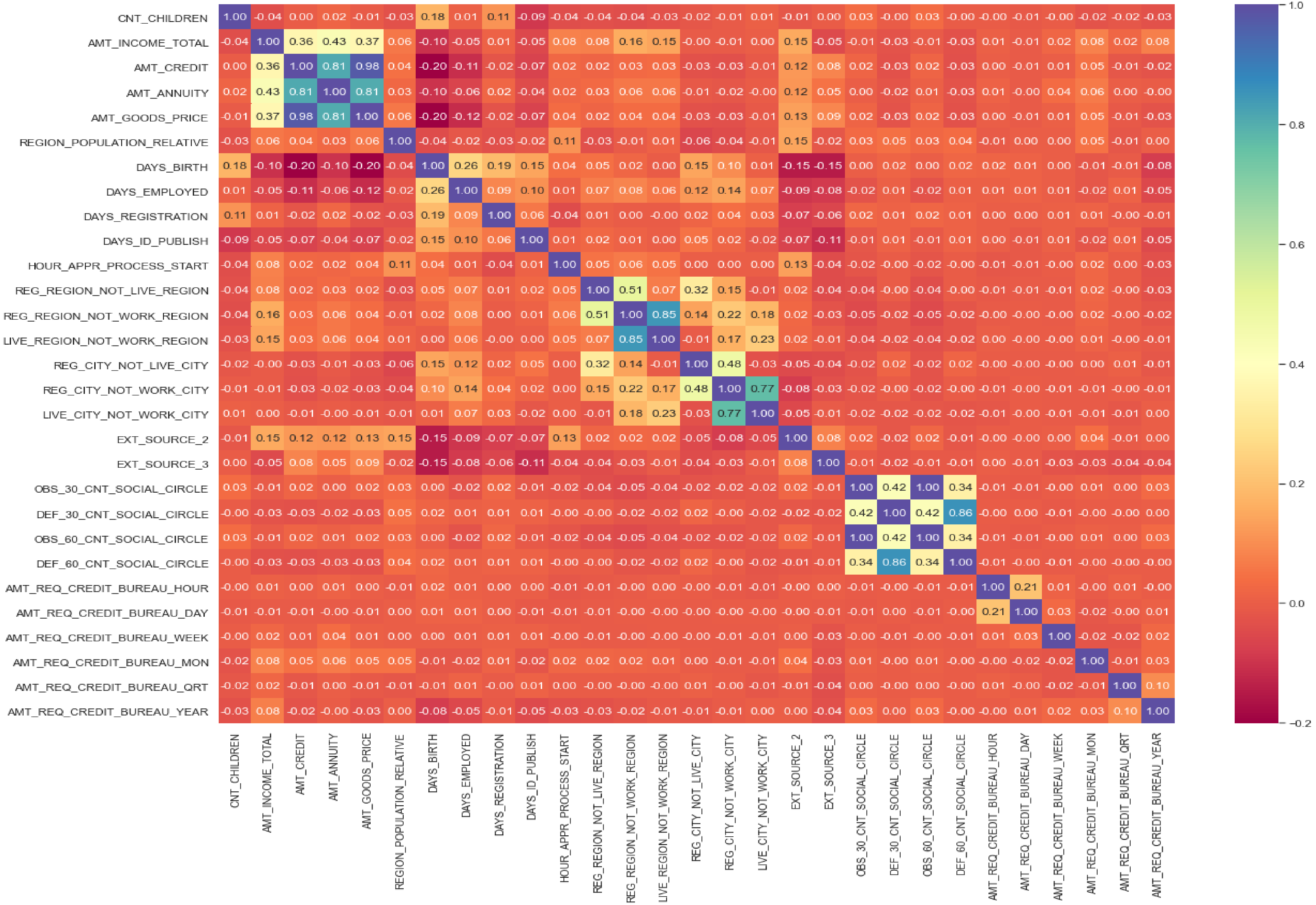
CORRELATION FOR TARGET 1

Observation:

Heat map for Target 1 is also having little bit same observation just like Target 0. But for few points are different. They are listed below.

1. Client's permanent address does not match to contact address which are having less children and vice-versa.
2. Client's permanent address does not match to work address which are having less children and vice-versa.

CORRELATION FOR TARGET 1



	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	HOUR_APPR_PROCESS_START	REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	EXT_SOURCE_2	EXT_SOURCE_3	OBS_30_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY	AMT_REQ_CREDIT_BUREAU_WEEK	AMT_REQ_CREDIT_BUREAU_MON	AMT_REQ_CREDIT_BUREAU_QRT	AMT_REQ_CREDIT_BUREAU_YEAR
CNT_CHILDREN	1.00	-0.04	0.00	0.02	-0.01	-0.03	0.18	0.01	0.11	-0.09	-0.04	-0.04	-0.04	-0.03	-0.02	-0.01	0.01	-0.01	0.00	0.03	-0.00	0.03	-0.00	-0.00	-0.01	-0.00	-0.02	-0.02	-0.03
AMT_INCOME_TOTAL	-0.04	1.00	0.36	0.43	0.37	0.06	-0.10	-0.05	0.01	-0.05	0.08	0.08	0.16	0.15	-0.00	-0.01	0.00	0.15	-0.05	-0.01	-0.03	-0.01	-0.03	0.01	-0.01	0.02	0.08	0.02	0.08
AMT_CREDIT	0.00	0.36	1.00	0.81	0.98	0.04	-0.20	-0.11	-0.02	-0.07	0.02	0.02	0.03	0.03	-0.03	-0.03	-0.01	0.12	0.08	0.02	-0.03	0.02	-0.03	0.00	-0.01	0.01	0.05	-0.01	-0.02
AMT_ANNUITY	0.02	0.43	0.81	1.00	0.81	0.03	-0.10	-0.06	0.02	-0.04	0.02	0.03	0.06	0.06	-0.01	-0.02	-0.00	0.12	0.05	0.00	-0.02	0.01	-0.03	0.01	-0.00	0.04	0.06	0.00	-0.00
AMT_GOODS_PRICE	-0.01	0.37	0.98	0.81	1.00	0.06	-0.20	-0.12	-0.02	-0.07	0.04	0.02	0.04	0.04	-0.03	-0.03	-0.01	0.13	0.09	0.02	-0.03	0.02	-0.03	0.00	-0.01	0.01	0.05	-0.01	-0.03
REGION_POPULATION_RELATIVE	-0.03	0.06	0.04	0.03	0.06	1.00	-0.04	-0.02	-0.03	-0.02	0.11	-0.03	-0.01	0.01	-0.06	-0.04	-0.01	0.15	-0.02	0.03	0.05	0.03	0.04	-0.01	0.00	0.00	0.05	-0.01	0.00
DAYS_BIRTH	0.18	-0.10	-0.20	-0.10	-0.20	-0.04	1.00	0.26	0.19	0.15	0.04	0.05	0.02	0.00	0.15	0.10	0.01	-0.15	-0.15	0.00	0.02	0.00	0.02	0.02	0.01	0.00	-0.01	-0.01	-0.08
DAYS_EMPLOYED	0.01	-0.05	-0.11	-0.06	-0.12	-0.02	0.26	1.00	0.09	0.10	0.01	0.07	0.08	0.06	0.12	0.14	0.07	-0.09	-0.08	-0.02	0.01	-0.02	0.01	0.01	0.01	0.01	-0.02	0.01	-0.05
DAYS_REGISTRATION	0.11	0.01	-0.02	0.02	-0.02	-0.03	0.19	0.09	1.00	0.06	-0.04	0.01	0.00	-0.00	0.02	0.04	0.03	-0.07	-0.06	0.02	0.01	0.02	0.01	0.00	0.00	0.01	0.01	-0.00	-0.01
DAYS_ID_PUBLISH	-0.09	-0.05	-0.07	-0.04	-0.07	-0.02	0.15	0.10	0.06	1.00	0.01	0.02	0.01	0.00	0.05	0.02	-0.02	-0.07	-0.11	-0.01	0.01	-0.01	0.01	0.00	-0.01	0.01	-0.02	0.01	-0.05
HOUR_APPR_PROCESS_START	-0.04	0.08	0.02	0.02	0.04	0.11	0.04	0.01	-0.04	0.01	1.00	0.05	0.06	0.05	0.00	0.00	0.00	0.13	-0.04	-0.02	-0.00	-0.02	-0.00	-0.01	-0.01	-0.00	0.02	0.00	-0.03
REG_REGION_NOT_LIVE_REGION	-0.04	0.08	0.02	0.03	0.02	-0.03	0.05	0.07	0.01	0.02	0.05	1.00	0.51	0.07	0.32	0.15	-0.01	0.02	-0.04	-0.04	-0.00	-0.04	-0.00	-0.01	-0.01	-0.01	0.02	-0.00	-0.03
REG_REGION_NOT_WORK_REGION	-0.04	0.16	0.03	0.06	0.04	-0.01	0.02	0.08	0.00	0.01	0.06	0.51	1.00	0.85	0.14	0.22	0.18	0.02	-0.03	-0.05	-0.02	-0.05	-0.02	-0.00	-0.00	-0.00	0.02	-0.00	-0.02
LIVE_REGION_NOT_WORK_REGION	-0.03	0.15	0.03	0.06	0.04	0.01	0.00	0.06	-0.00	0.00	0.05	0.07	0.85	1.00	-0.01	0.17	0.23	0.02	-0.01	-0.04	-0.02	-0.04	-0.02	0.00	-0.00	-0.00	0.01	-0.00	-0.01
REG_CITY_NOT_LIVE_CITY	-0.02	-0.00	-0.03	-0.01	-0.03	-0.06	0.15	0.12	0.02	0.05	0.00	0.32	0.14	-0.01	1.00	0.48	-0.03	-0.05	-0.04	-0.02	0.02	-0.02	0.02	-0.00	-0.00	-0.00	0.00	0.01	-0.01
REG_CITY_NOT_WORK_CITY	-0.01	-0.01	-0.03	-0.02	-0.03	-0.04	0.10	0.14	0.04	0.02	0.00	0.15	0.22	0.17	0.48	1.00	0.77	-0.08	-0.03	-0.02	-0.00	-0.02	-0.00	-0.01	-0.00	-0.01	-0.01	-0.00	-0.01
LIVE_CITY_NOT_WORK_CITY	0.01	0.00	-0.01	-0.00	-0.01	-0.01	0.01	0.07	0.03	-0.02	0.00	-0.01	0.18	0.23	-0.03	0.77	1.00	-0.05	-0.01	-0.02	-0.02	-0.02	-0.02	-0.01	-0.00	-0.01	-0.01	-0.01	0.00
EXT_SOURCE_2	-0.01	0.15	0.12	0.12	0.13	0.15	-0.15	-0.09	-0.07	-0.07	0.13	0.02	0.02	0.02	-0.05	-0.08	-0.05	1.00	0.08	0.02	-0.02	0.02	-0.01	-0.00	-0.00	0.00	0.04	-0.01	0.00
EXT_SOURCE_3	0.00	-0.05	0.08	0.05	0.09	-0.02	-0.15	-0.08	-0.06	-0.11	-0.04	-0.04	-0.03	-0.01	-0.04	-0.03	-0.01	0.08	1.00	-0.01	-0.02	-0.01	-0.01	0.00	-0.01	-0.03	-0.03	-0.04	-0.04
OBS_30_CNT_SOCIAL_CIRCLE	0.03	-0.01	0.02	0.00	0.02	0.03	0.00	-0.02	0.02	-0.01	-0.02	-0.04	-0.05	-0.04	-0.02	-0.02	-0.02	0.02	-0.01	1.00	0.42	1.00	0.34	-0.01	-0.01	-0.00	0.01	0.00	0.03
DEF_30_CNT_SOCIAL_CIRCLE	-0.00	-0.03	-0.03	-0.02	-0.03	0.05	0.02	0.01	0.01	0.01	-0.00	-0.00	-0.02	-0.02	0.02	-0.00	-0.02	-0.02	-0.02	0.42	1.00	0.42	0.86	-0.00	0.00	-0.01	-0.00	-0.00	0.00
OBS_60_CNT_SOCIAL_CIRCLE	0.03	-0.01	0.02	0.01	0.02	0.03	0.00	-0.02	0.02	-0.01	-0.02	-0.04	-0.05	-0.04	-0.02	-0.02	-0.02	0.02	-0.01	1.00	0.42	1.00	0.34	-0.01	-0.01	-0.00	0.01	0.00	0.03
DEF_60_CNT_SOCIAL_CIRCLE	-0.00	-0.03	-0.03	-0.03	-0.03	0.04	0.02	0.01	0.01	0.01	-0.00	-0.00	-0.02	-0.02	0.02	-0.00	-0.02	-0.01	-0.01	0.34	0.86	0.34	1.00	-0.01	-0.00	-0.01	-0.00	-0.00	-0.00
AMT_REQ_CREDIT_BUREAU_HOUR	-0.00	0.01	0.00	0.01	0.00	-0.01	0.02	0.01	0.00	0.00	-0.01	-0.01	-0.00	0.00	-0.00	-0.01	-0.01	-0.00	0.00	-0.01	-0.00	-0.01	-0.01	1.00	0.21	0.01	-0.00	0.01	-0.00
AMT_REQ_CREDIT_BUREAU_DAY	-0.01	-0.01	-0.01	-0.00	-0.01	0.00	0.01	0.01	0.00	-0.01	-0.01	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.01	0.00	-0.01	-0.00	0.21	1.00	0.03	-0.02	-0.00	0.01
AMT_REQ_CREDIT_BUREAU_WEEK	-0.00	0.02	0.01	0.04	0.01	0.00	0.00	0.01	0.01	0.01	-0.00	-0.01	-0.00	-0.00	-0.00	-0.01	-0.01	0.00	-0.03	-0.00	-0.01	-0.00	-0.01	0.01	0.03	1.00	-0.02	-0.02	0.02
AMT_REQ_CREDIT_BUREAU_MON	-0.02	0.08	0.05	0.06	0.05	0.05	-0.01	-0.02	0.01	-0.02	0.02	0.02	0.02	0.01	0.00	-0.01	-0.01	0.04	-0.03	0.01	-0.00	0.01	-0.00	-0.00	-0.02	-0.02	1.00	-0.01	0.03
AMT_REQ_CREDIT_BUREAU_QRT	-0.02	0.02	-0.01	0.00	-0.01	-0.01	-0.01	0.01	-0.00	0.01	0.00	-0.00	-0.00	-0.00	0.01	-0.00	-0.01	-0.01	-0.04	0.00	-0.00	0.00	-0.00	0.01	-0.00	-0.02	-0.01	1.00	0.10
AMT_REQ_CREDIT_BUREAU_YEAR	-0.03	0.08	-0.02	-0.00	-0.03	0.00	-0.08	-0.05	-0.01	-0.05	-0.03	-0.03	-0.02	-0.01	-0.01	-0.01	0.00	0.00	-0.04	0.03	0.00	0.03	-0.00	-0.00	0.01	0.02	0.03	0.10	1.00

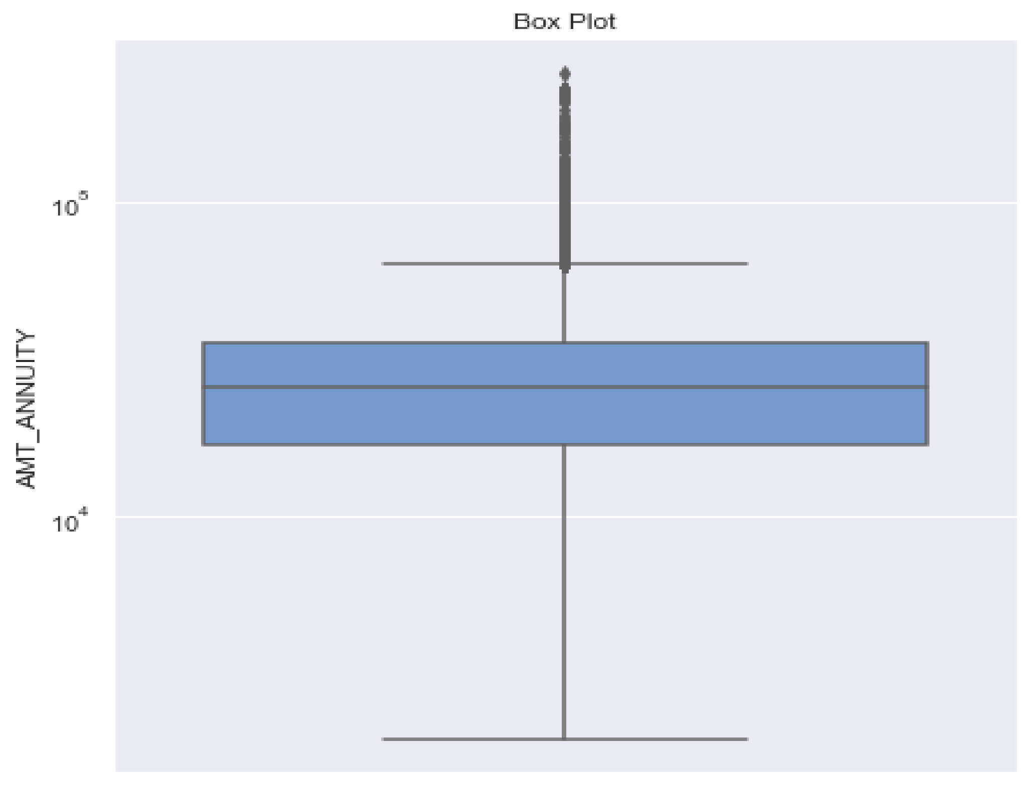
Univariate Analysis of Numerical Variables

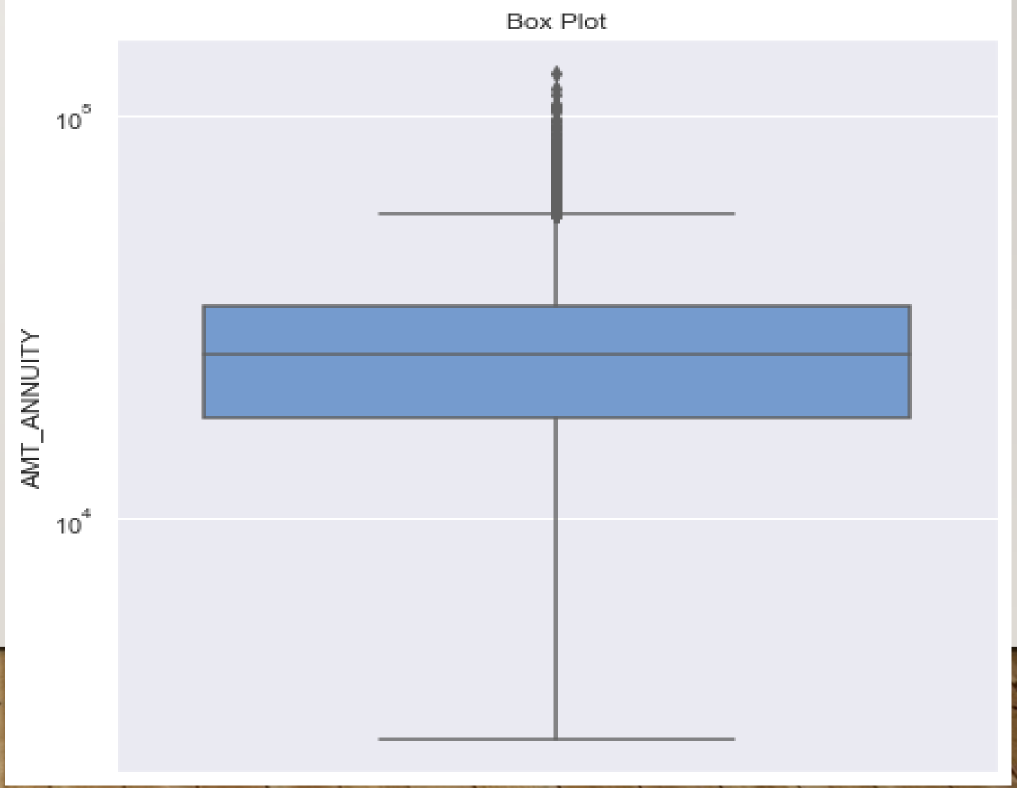
LOAN ANNUITY

Observations:

1. Some outliers are noticed in income amount.

2. The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.



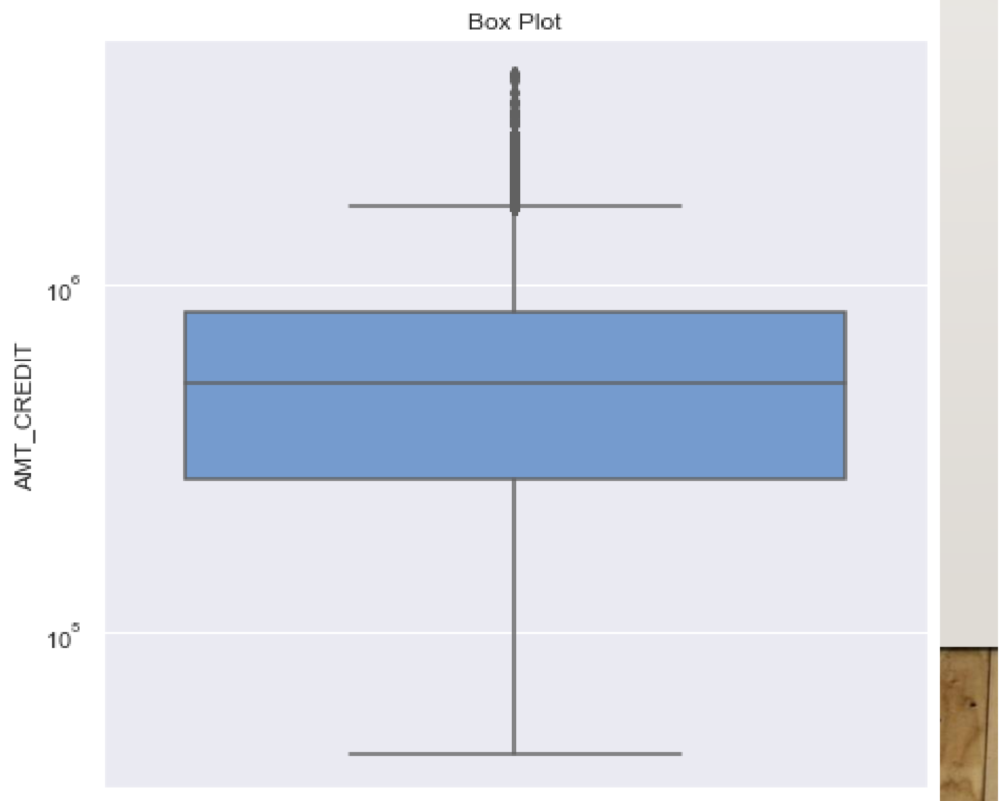


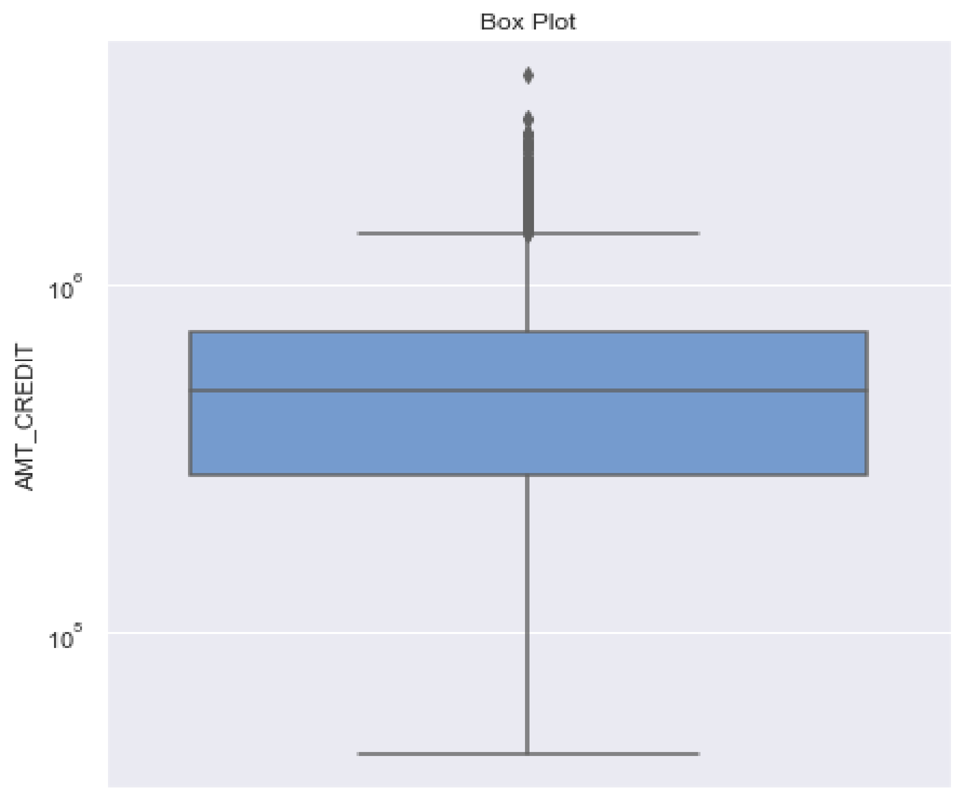
CREDIT AMOUNT

Observations:

1. Some outliers are noticed in income amount.

2. The first quartile is bigger than third quartile for credit amount which means most of the credit clients are present in the first quartile.



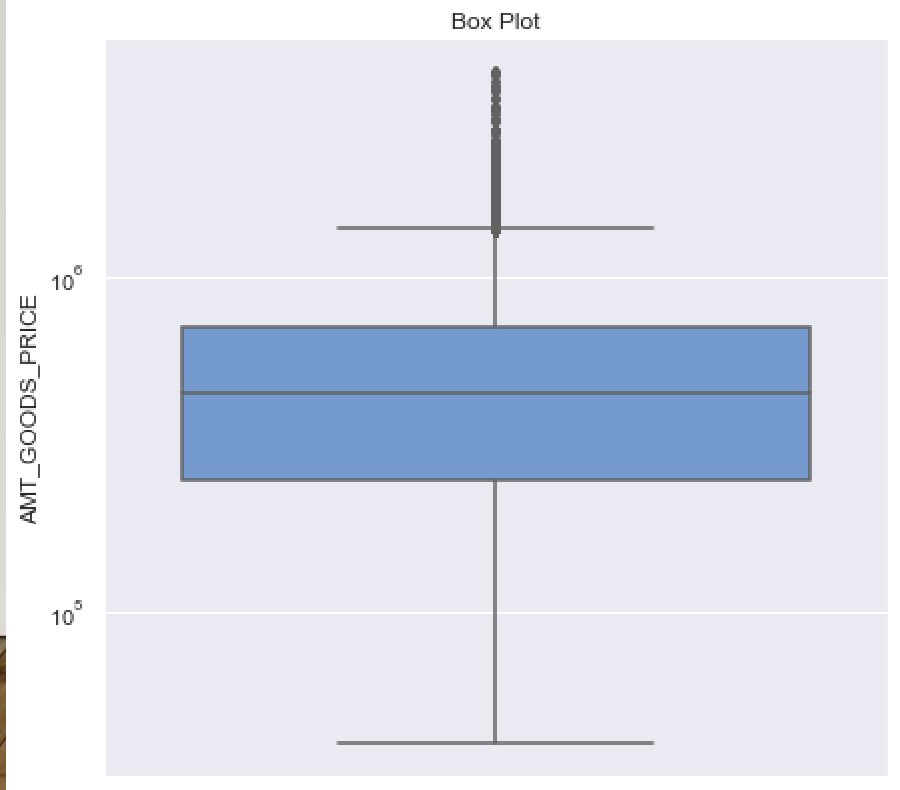


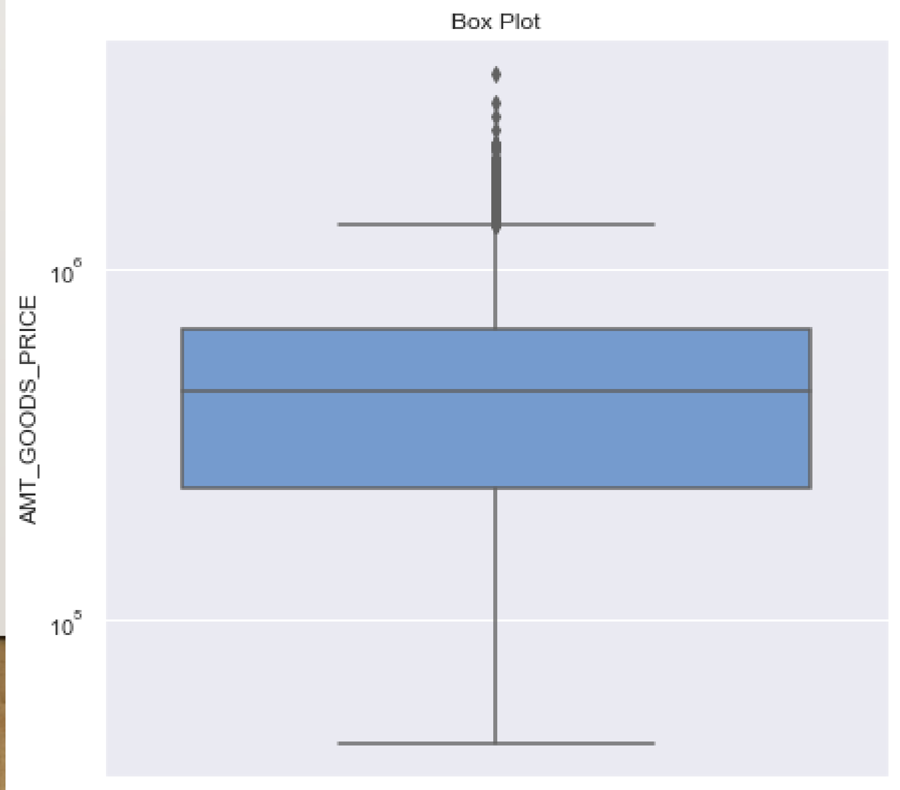
GOODS PRICE

Observations:

1. Some outliers are noticed in income amount.

2. The first quartile is bigger than third quartile for goods amount which means most of the goods clients are from first quartile.



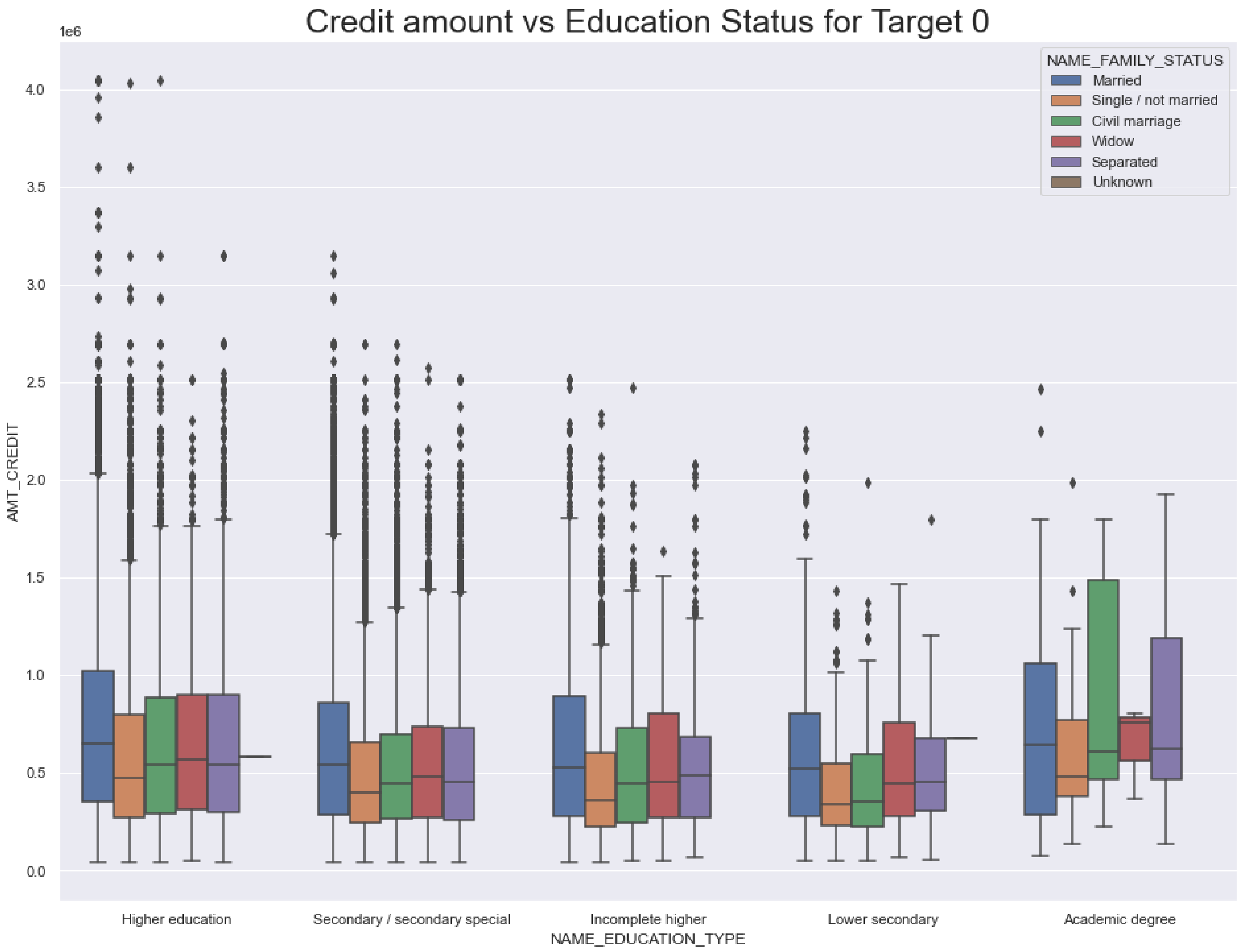


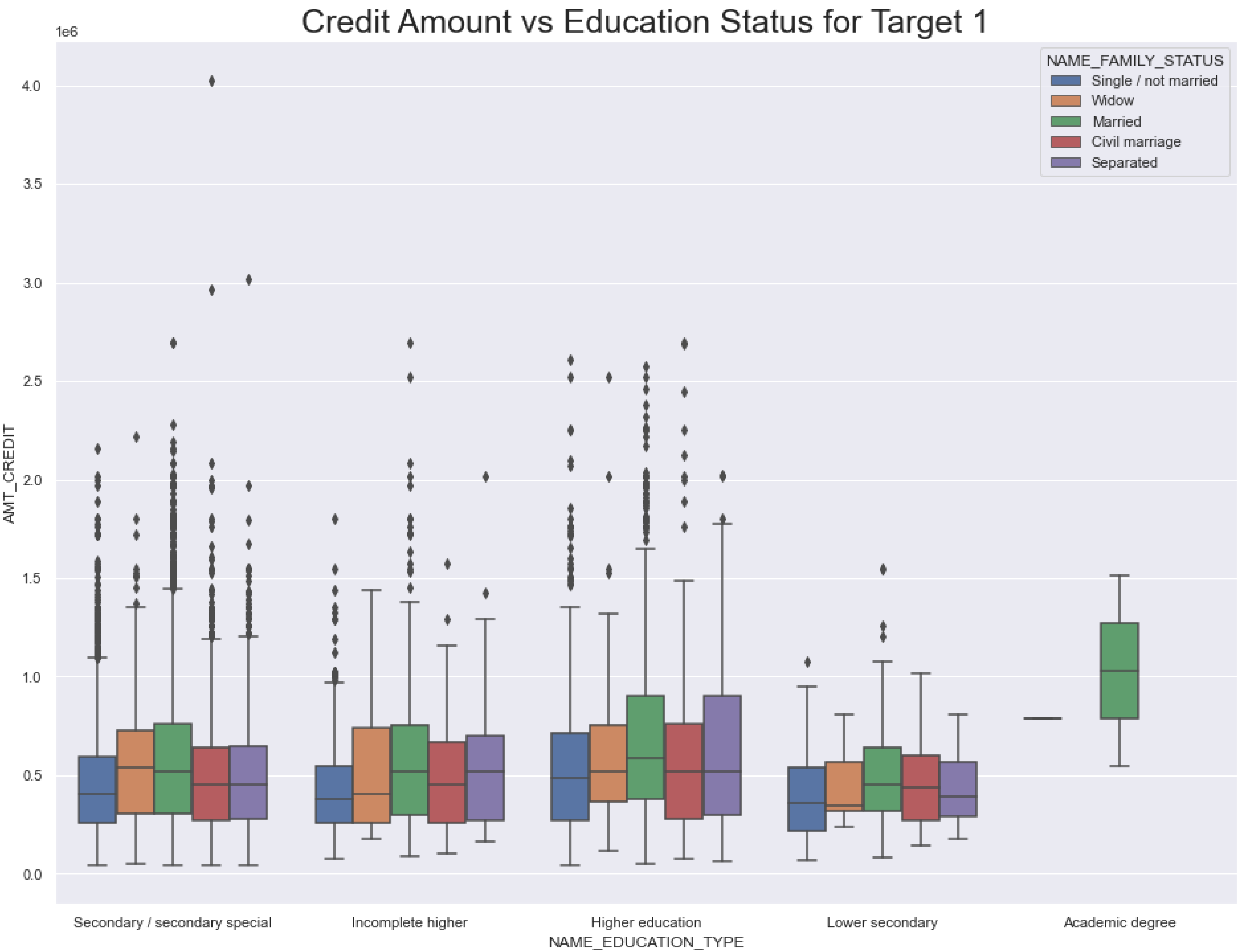
Bivariate Analysis for Numerical Variables

CREDIT AMOUNT VS EDUCATION STATUS

Observation :

- Here we can see that the range of customers without payment of Academic degree is higher than the customer of with payment. And the rest of the Education type is almost same for both the cases.


Credit amount vs Education Status for Target 0
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
AMT_CREDIT
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Widow
Separated
Unknown
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

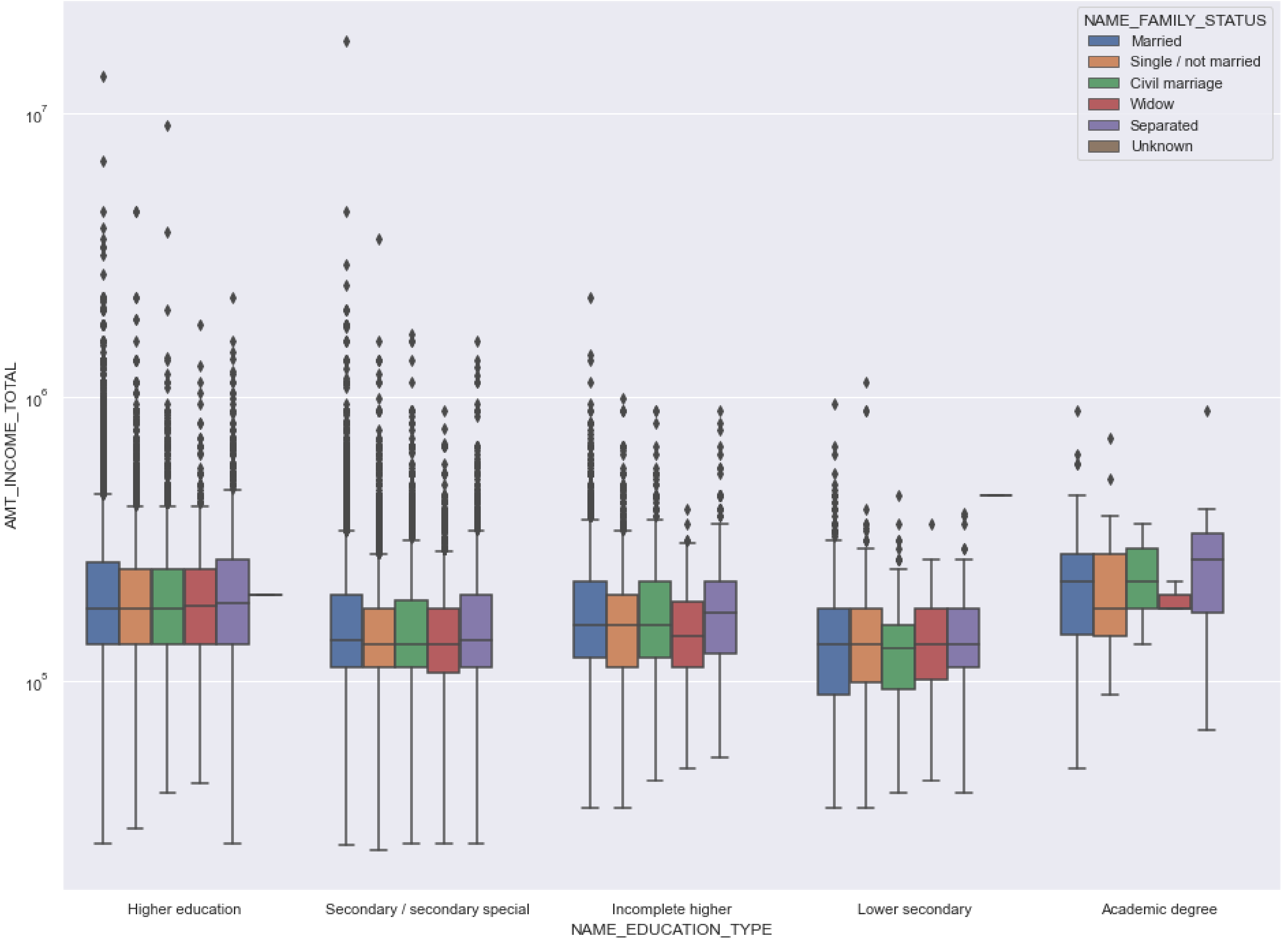

Credit Amount vs Education Status for Target 1
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
AMT_CREDIT
NAME_FAMILY_STATUS
Single / not married
Widow
Married
Civil marriage
Separated
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

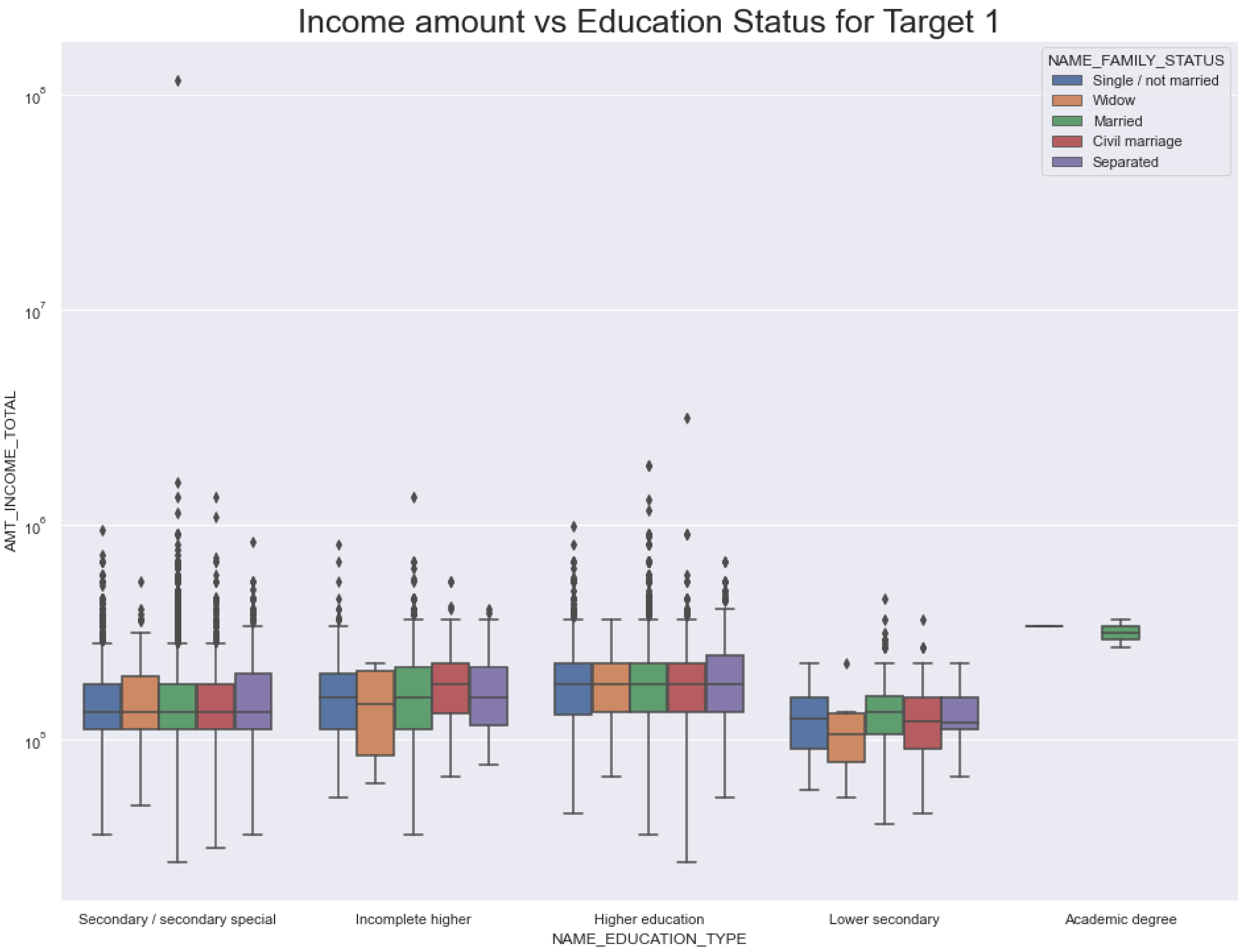
INCOME AMOUNT VS EDUCATION STATUS

Observation:

Here we can see that the customers without payment is having more outliers as compare to the customer with payment.

Income amount vs Education Status Target 0





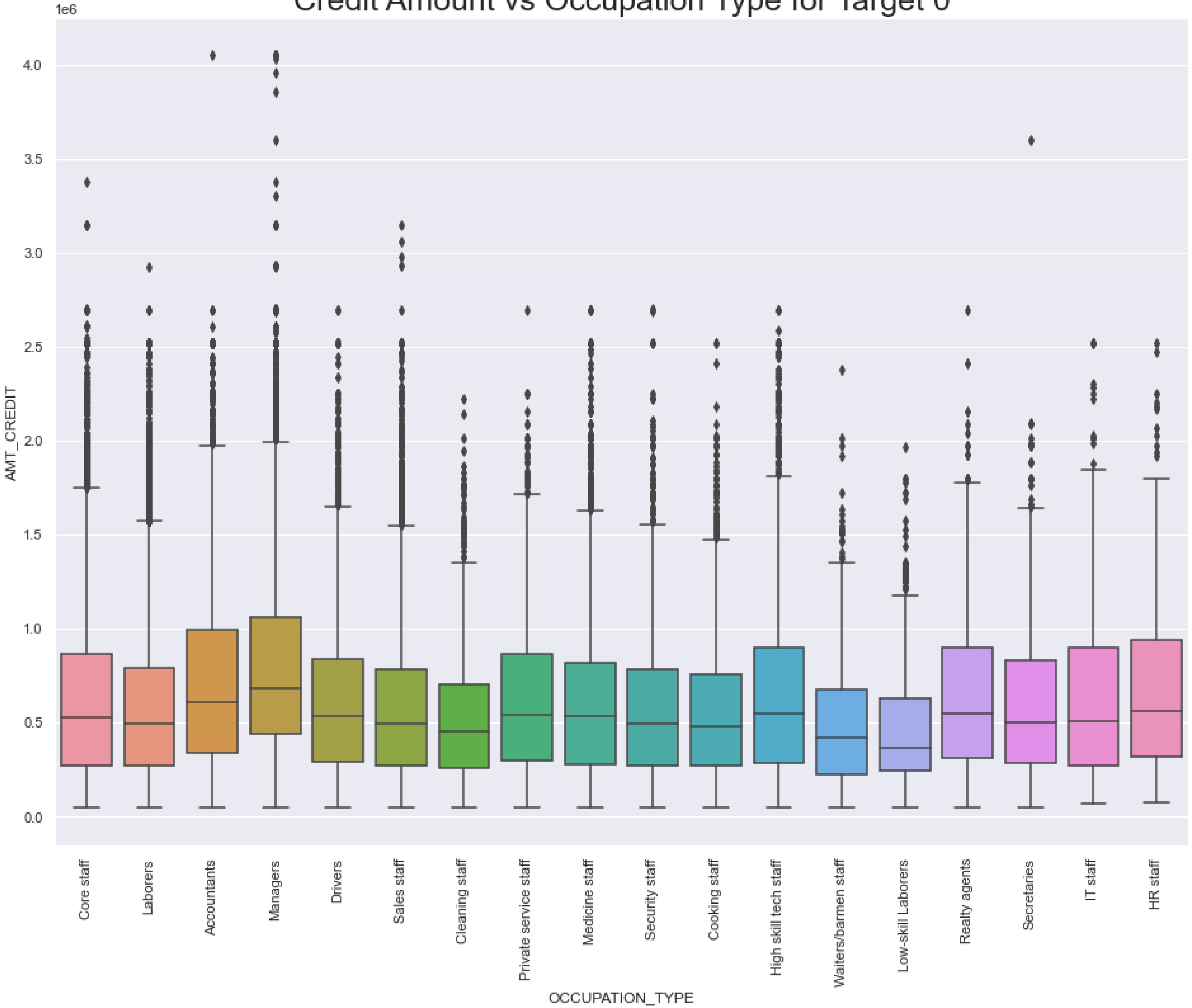
Income amount vs Education Status for Target 1

CREDIT AMOUNT VS OCCUPATION TYPE

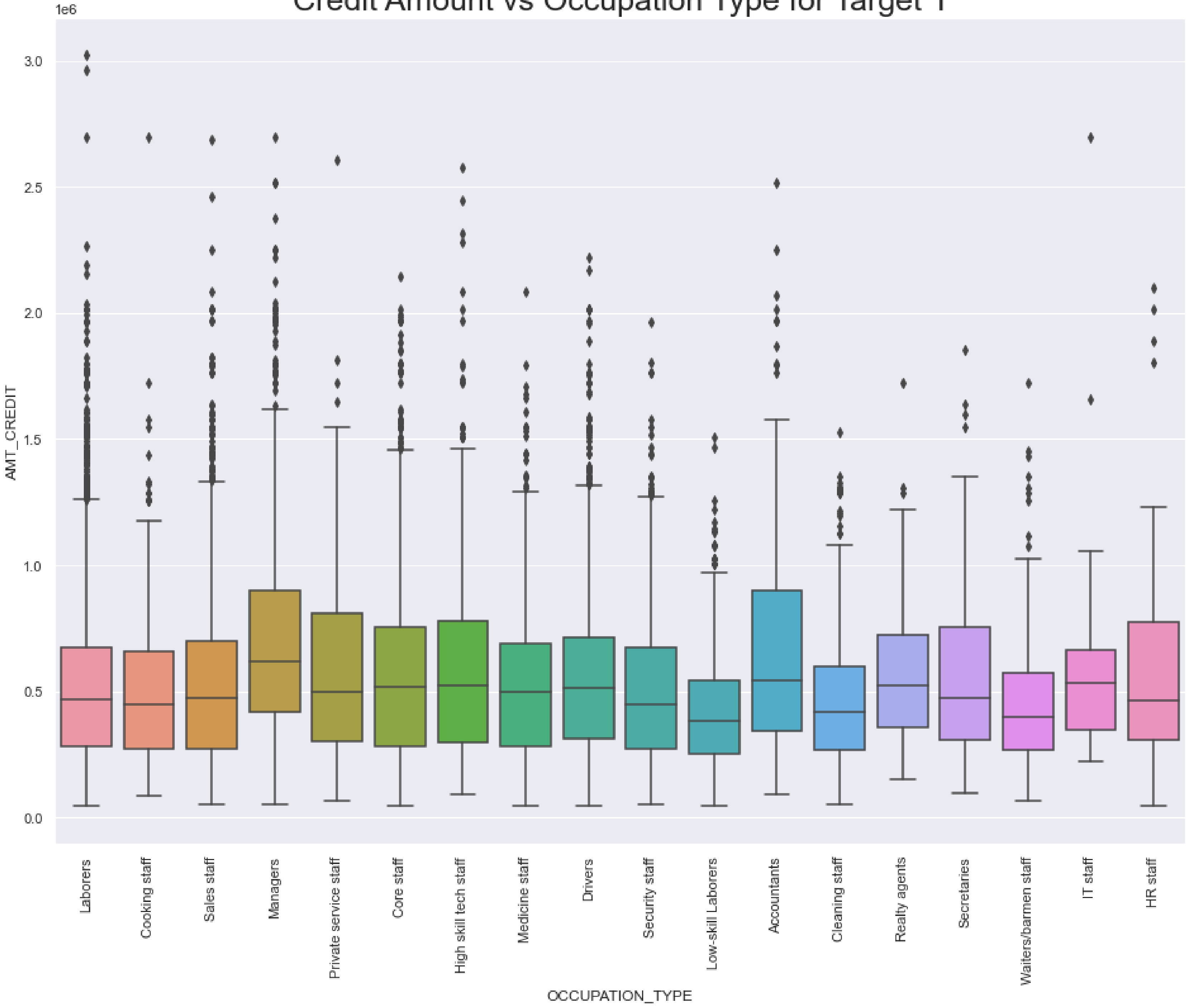
Observation :

Here we can see that the range of the customers without payment more as compare to the customers with payment.

Credit Amount vs Occupation Type for Target 0



Credit Amount vs Occupation Type for Target 1



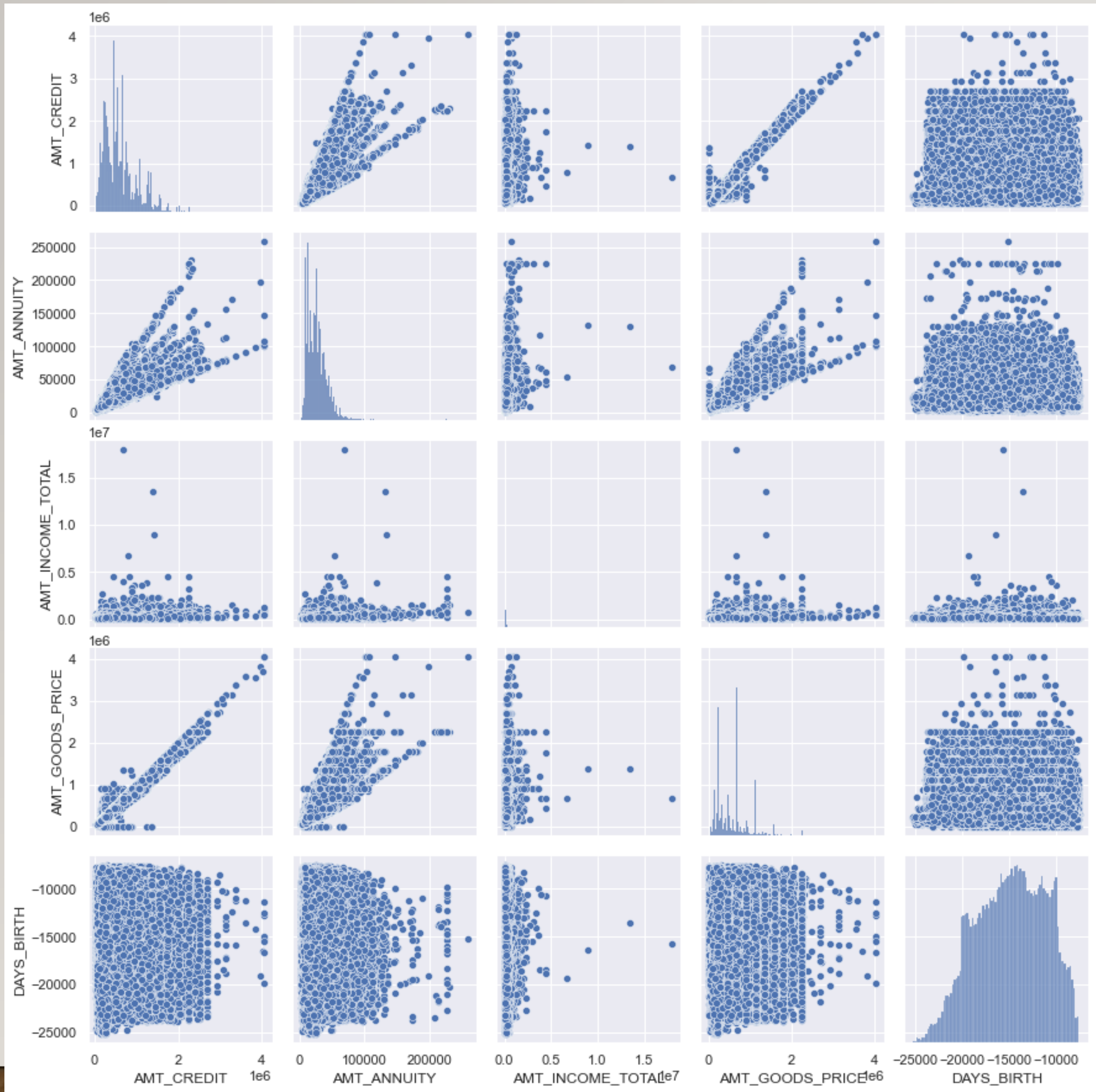
Bivariate Analysis of Numerical vs Numerical Variables

PAIR PLOT FOR TARGET 0 AND TARGET 1

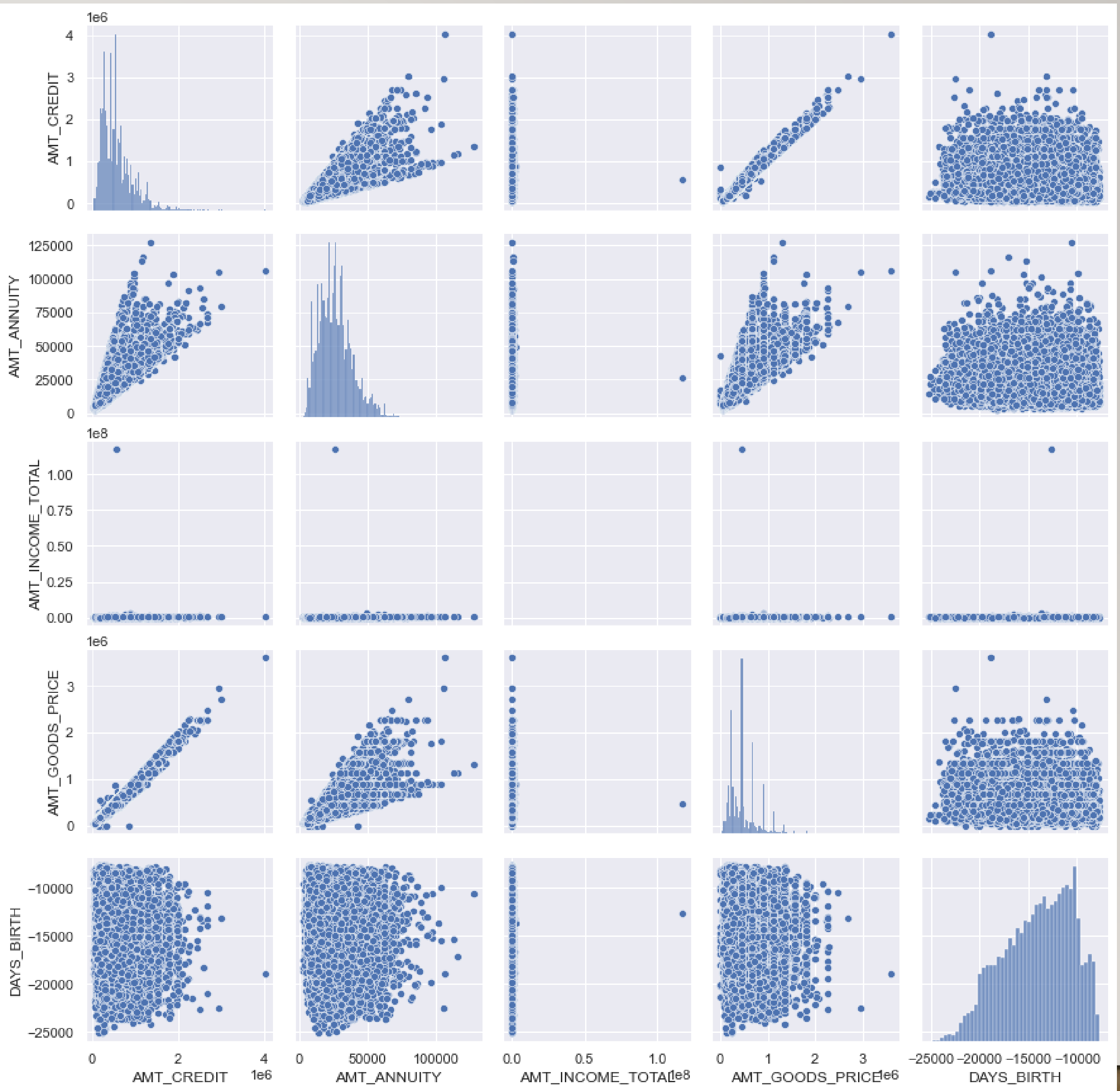
Observation:

We observe that there is a high correlation between credit amount and goods price. There appears to be some deviancies in the correlation of Loan-Payment Difficulties and Loan Payment with No Difficulties such as credit amount v/s income.

Target 0

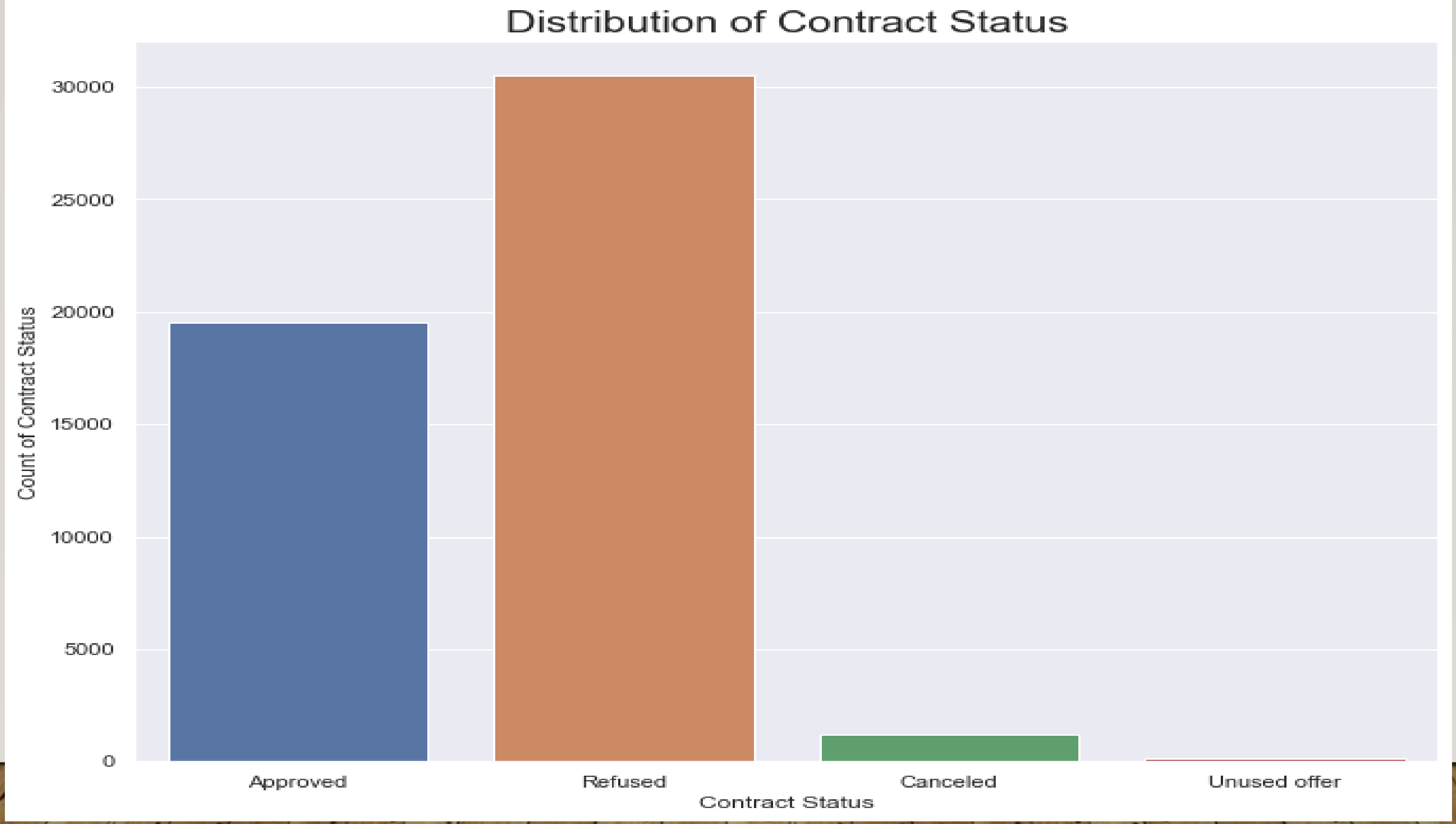


Target 1



Some Univariate Analysis on Previous Application Data

DISTRIBUTION OF CONTRACT STATUS



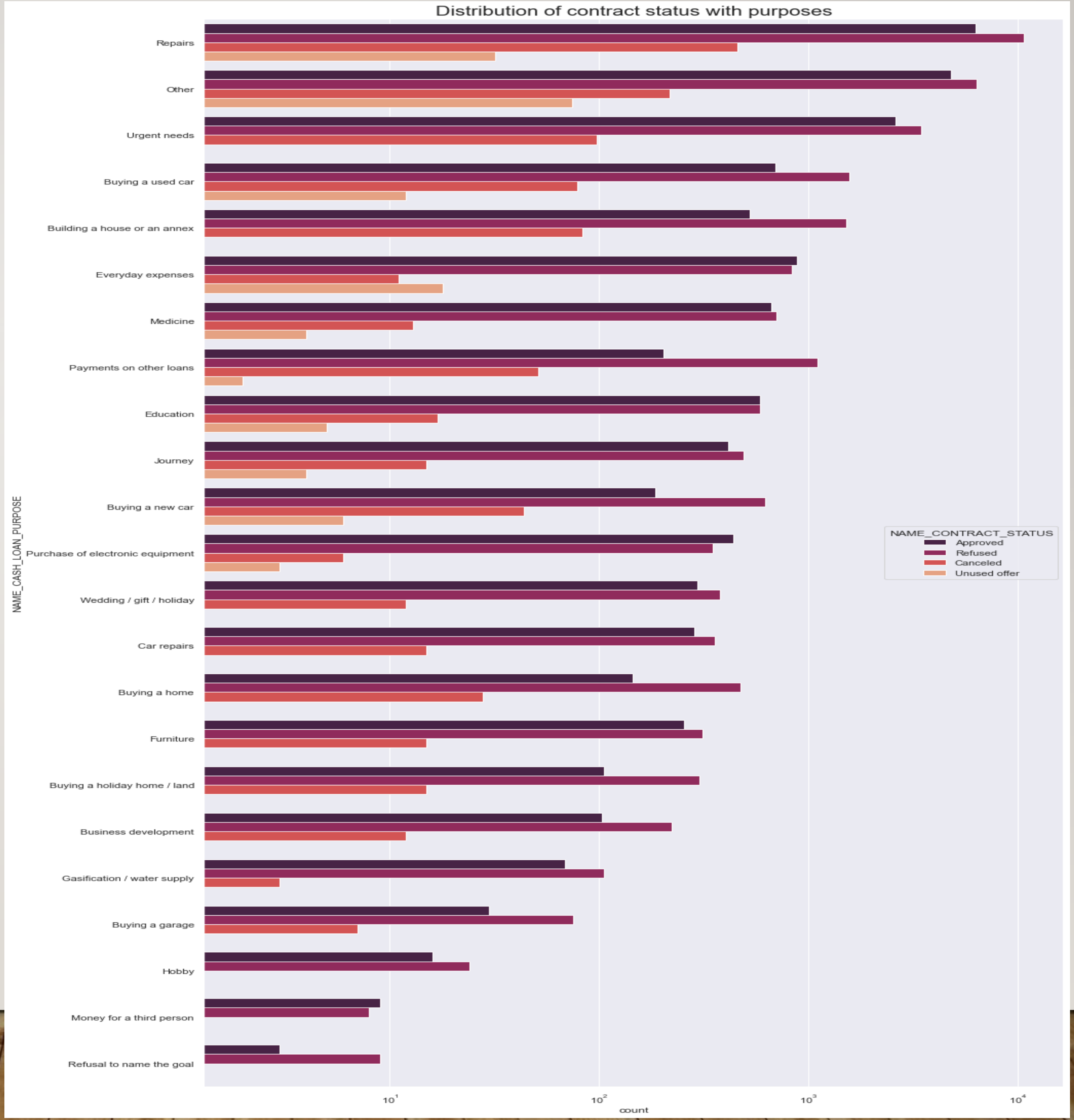
DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

Observation:

1. Most rejection of loans came from purpose 'repairs'.

2. For education purposes we have equal number of approves and rejection

3. Paying other loans and buying a new car is having significant higher rejection than approves.

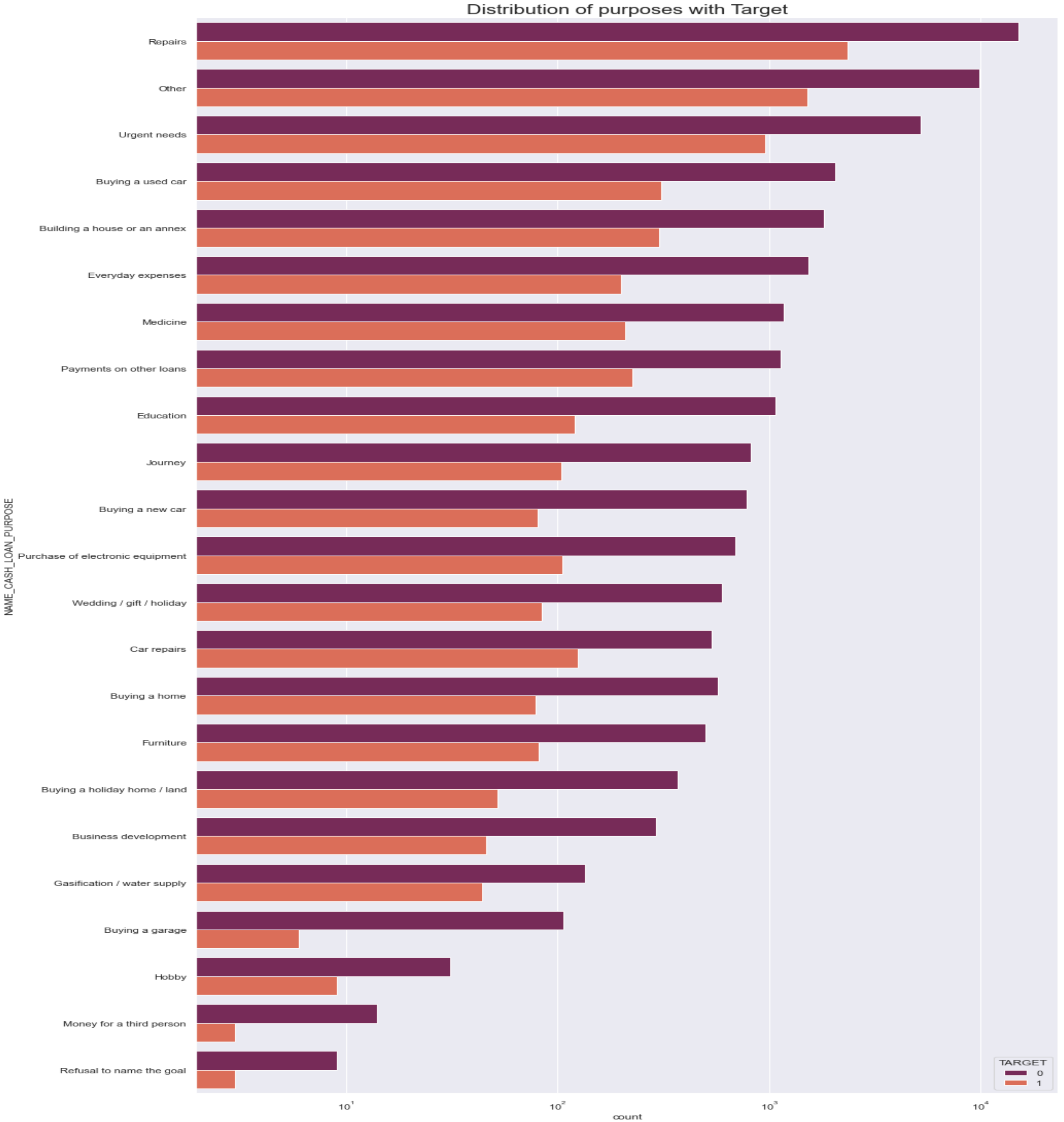


DISTRIBUTION OF PURPOSES WITH TARGET

Observation:

1. Loan purposes with 'Repairs' are facing more difficulties in payment on time.

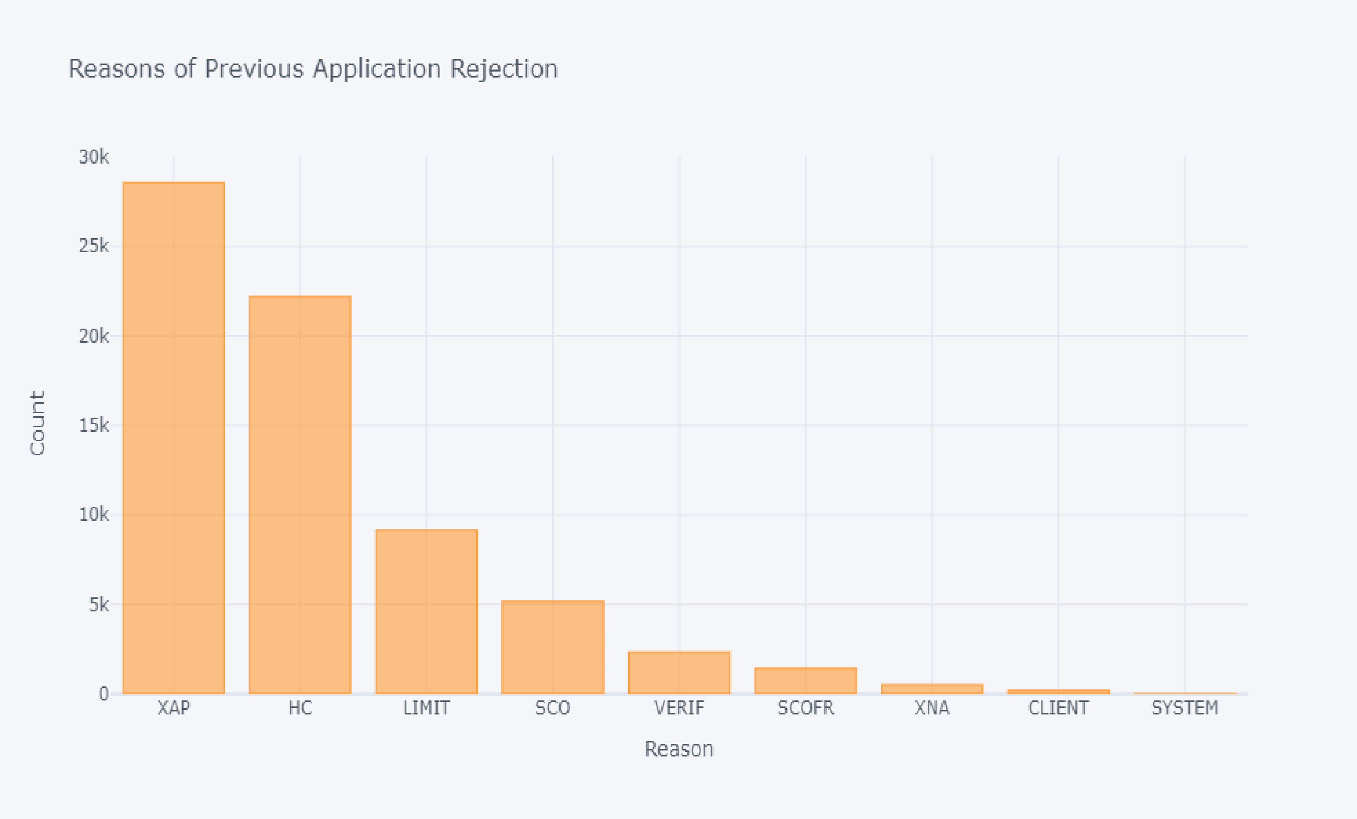
2. There are few places where loan payment is significant higher than facing difficulties. They are 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education' Hence we can focus on these purposes for which the client is having for minimal payment difficulties.



REASONS OF PREVIOUS APPLICATION REJECTION

Observation:

We observe that XAP is the reason majority of applications got rejected.



Bivariate Analysis

CONTRACT STATUS

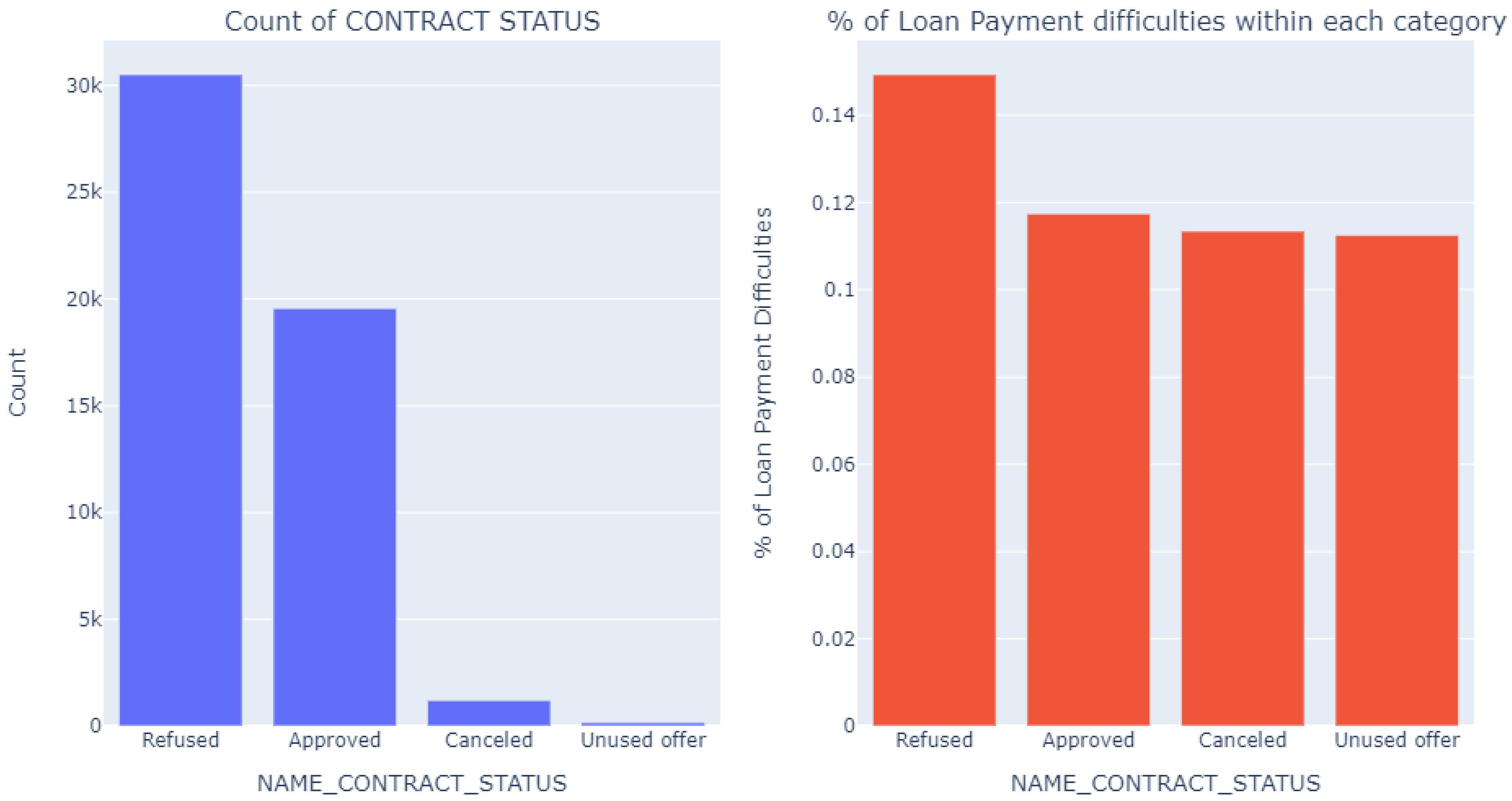
Observation:

From the first graph it can be seen that most of the contracts from previous application have been Approved.

It can be clearly seen from the second graph that the -

1. 'Refused' contracts from previous application are the ones who have maximum % of Loan-Payment Difficulties from current application.
2. 'Approved' contracts from previous application are the ones who have minimum % of Loan-Payment Difficulties from current application.

CONTRACT STATUS



CASH LOAN PURPOSE

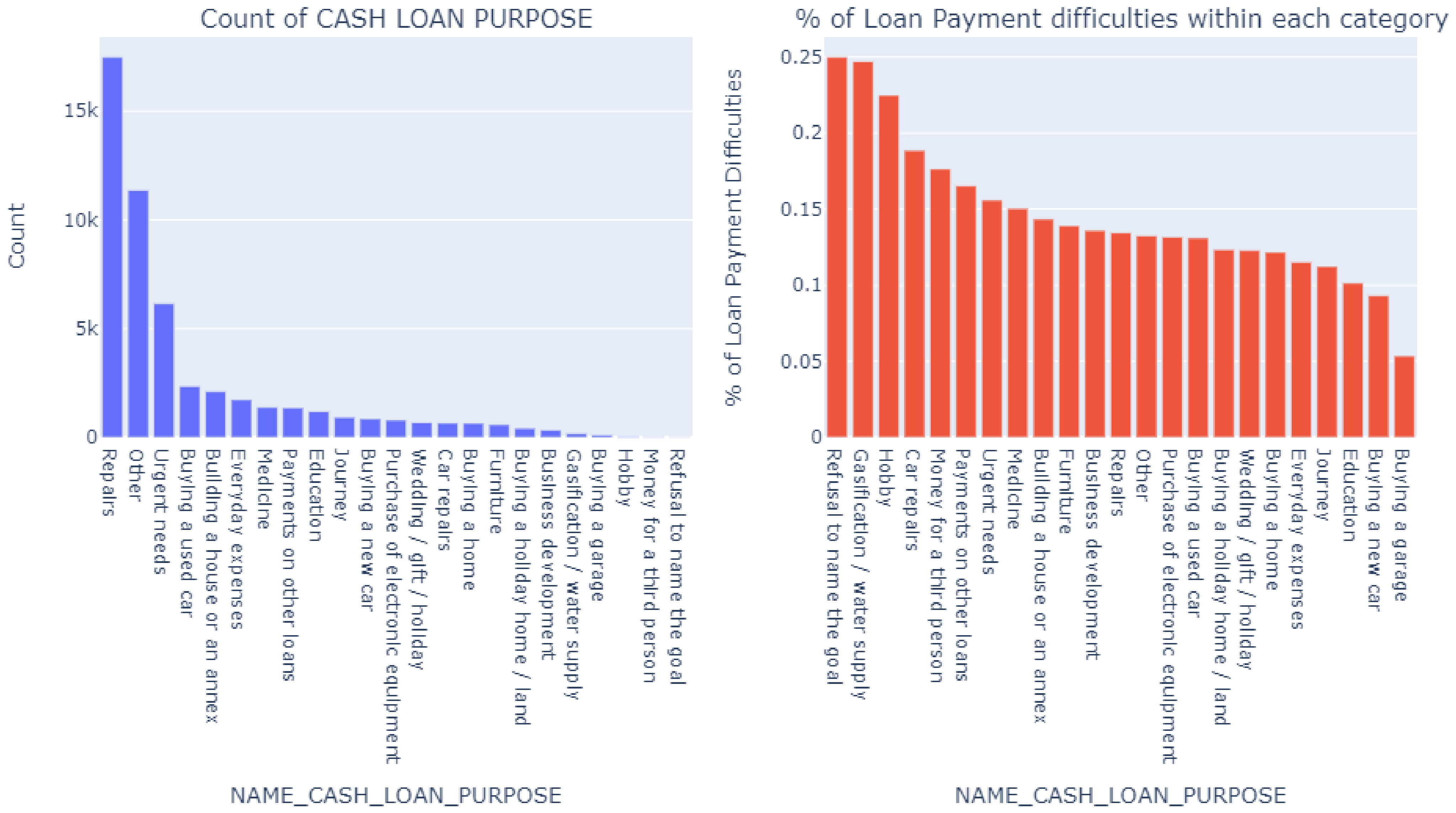
Observation:

From the first graph it can be seen that purpose of cash loan from previous data was maximum for 'Repairs'

It can be clearly seen from the second graph that the

1. 'Refusal to name the goal' for cash loan from previous application are the ones who have maximum % of Loan-Payment Difficulties from current application.

CASH LOAN PURPOSE



CONCLUSION

- Banks should focus more on contract type of ‘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.

THANK YOU..!!