

EDA ASSIGNMENT PRESENTATION

BY – Yogini Madhukar Shewale

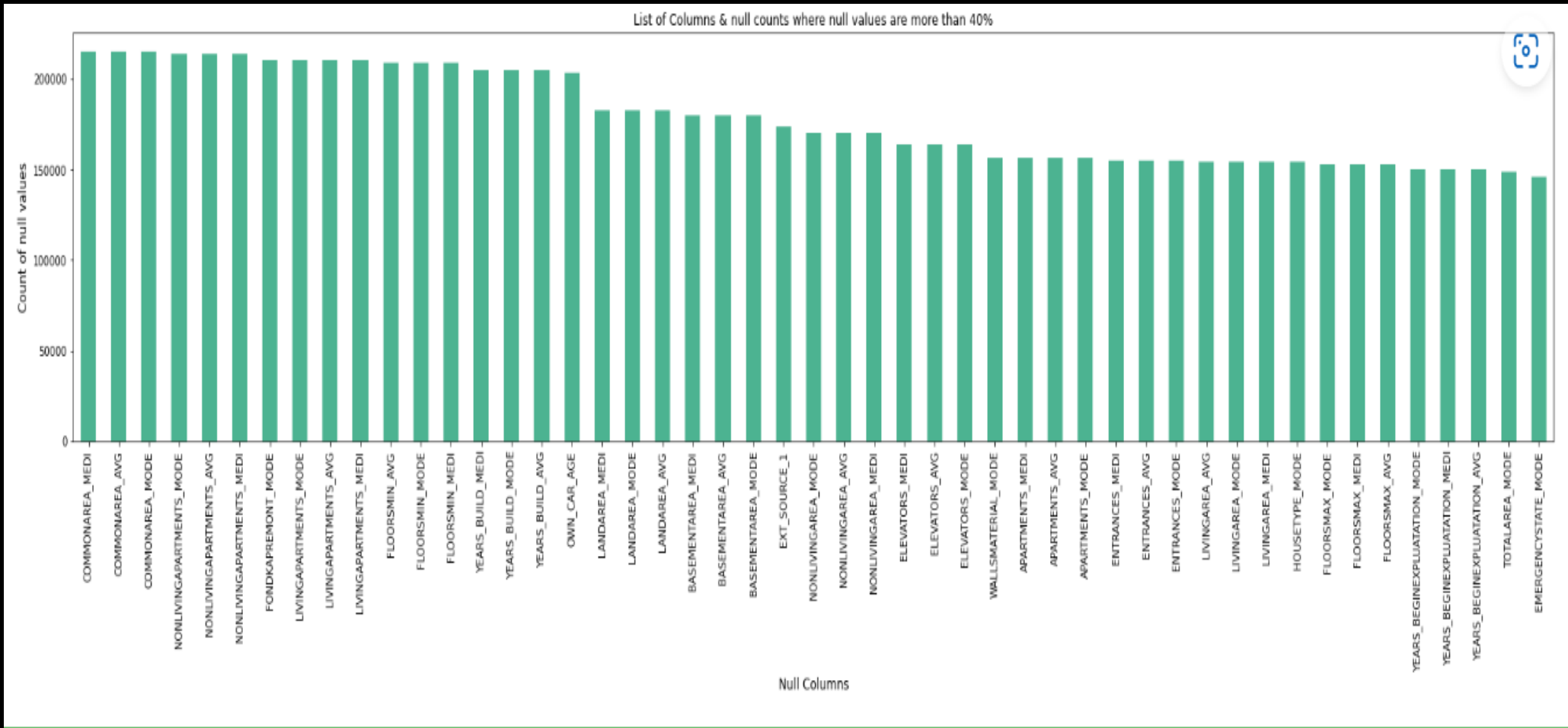
CASE STUDY ON APPLICATION DATA

DESCRIPTION OF APPLICATION DATA:

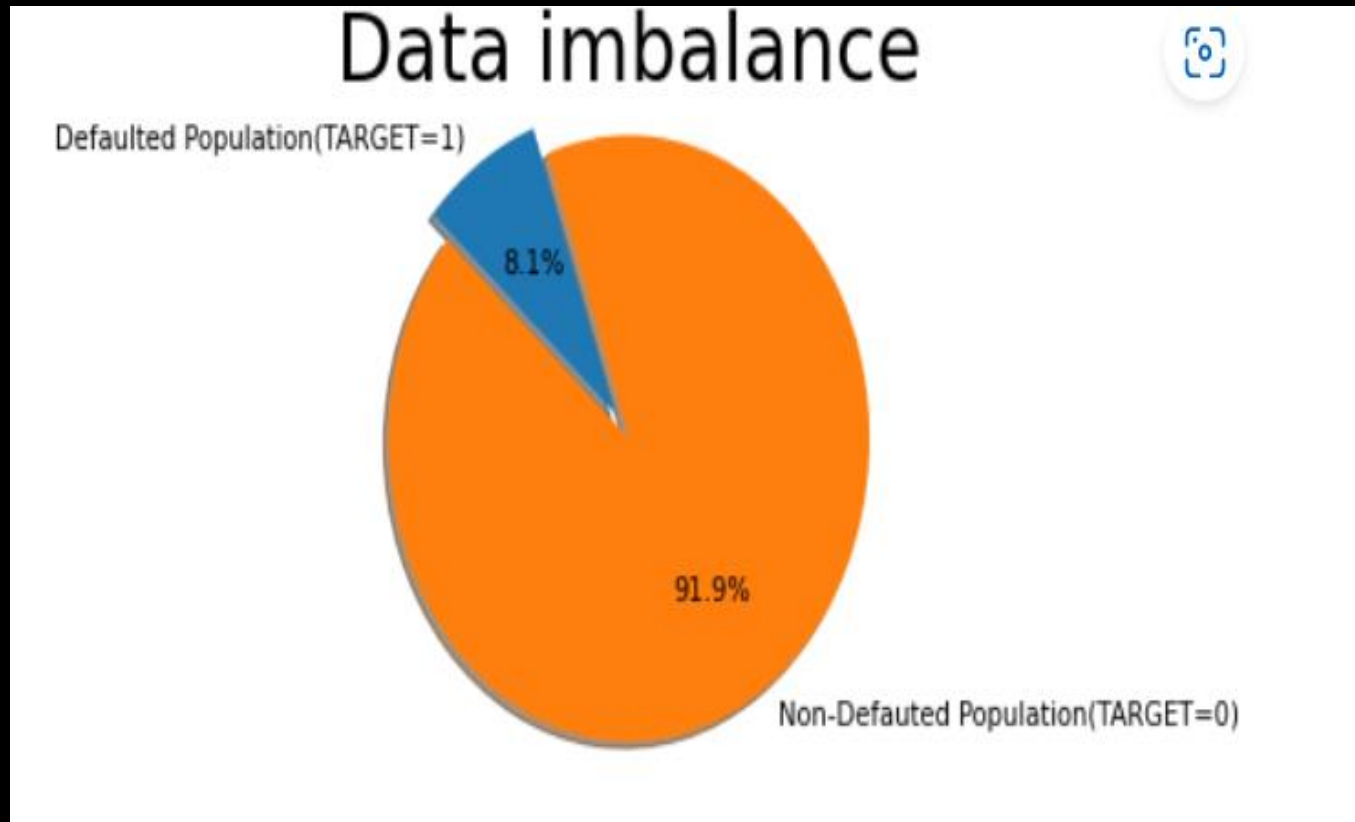
	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT.
0	100002	1	Cash loans	M	N	Y	0	202500.0	
1	100003	0	Cash loans	F	N	N	0	270000.0	1
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	
3	100006	0	Cash loans	F	N	Y	0	135000.0	
4	100007	0	Cash loans	M	N	Y	0	121500.0	
...	
307506	456251	0	Cash loans	M	N	N	0	157500.0	
307507	456252	0	Cash loans	F	N	Y	0	72000.0	
307508	456253	0	Cash loans	F	N	Y	0	153000.0	
307509	456254	1	Cash loans	F	N	Y	0	171000.0	
307510	456255	0	Cash loans	F	N	N	0	157500.0	

307511 rows × 122 columns

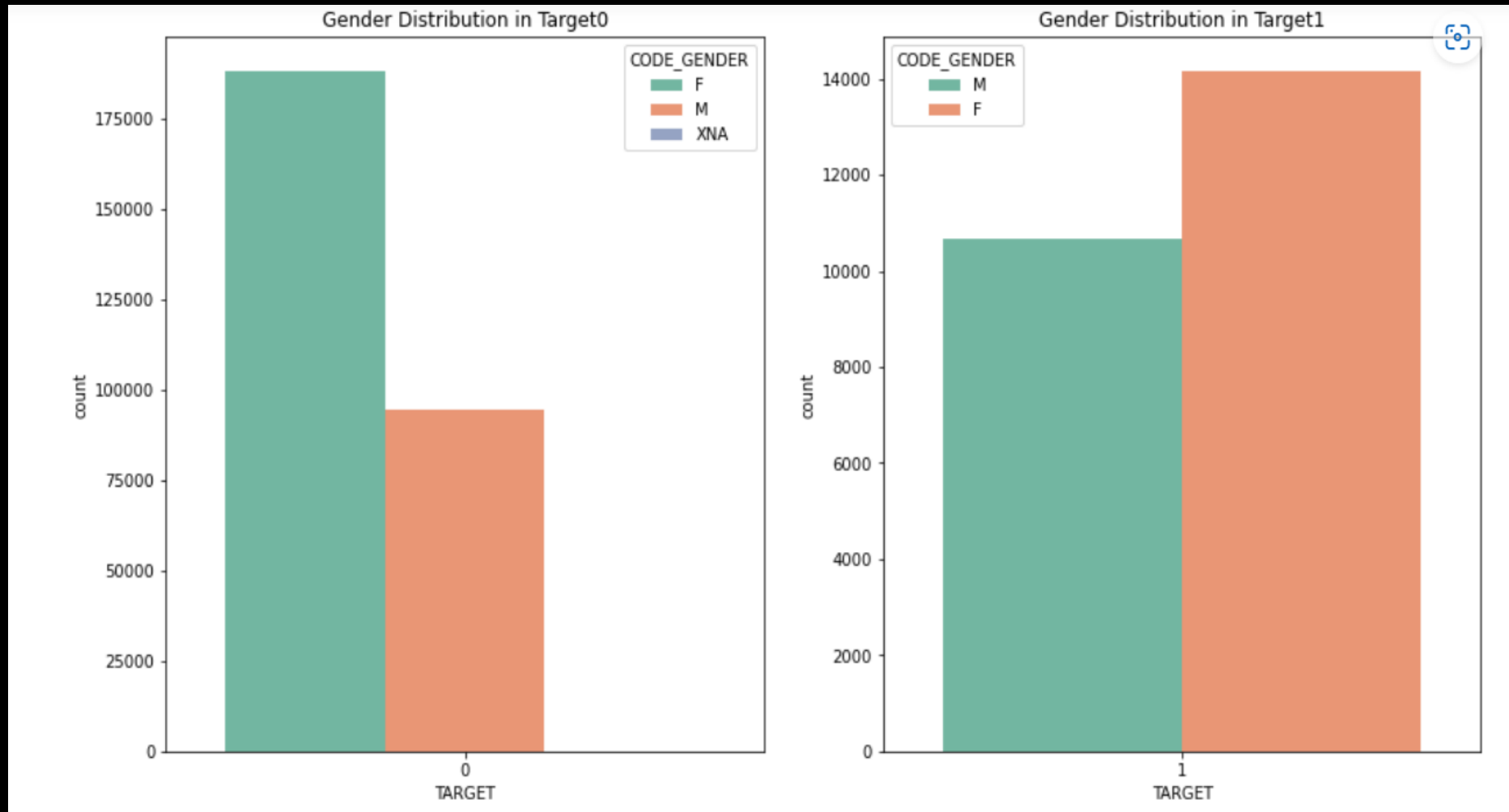
CHECKING NULL VALUES IN DATA SET AND PRESENTING THEM USING GRAPH:




CHECKING FOR % OF IMBALANCING BEFORE ANALYSING THEM:

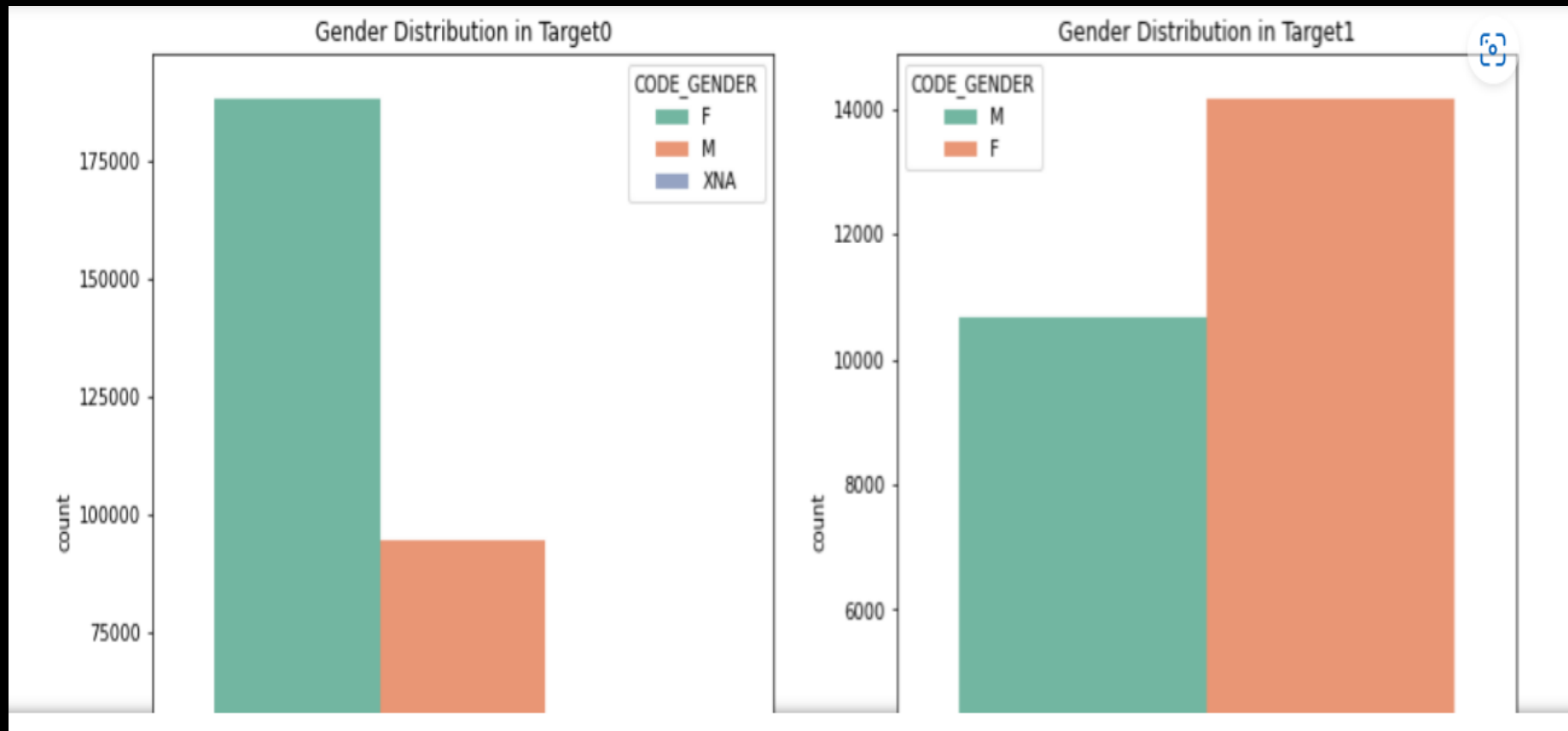


ANALYSING DATA BASES ON TARGET VARIABLES:



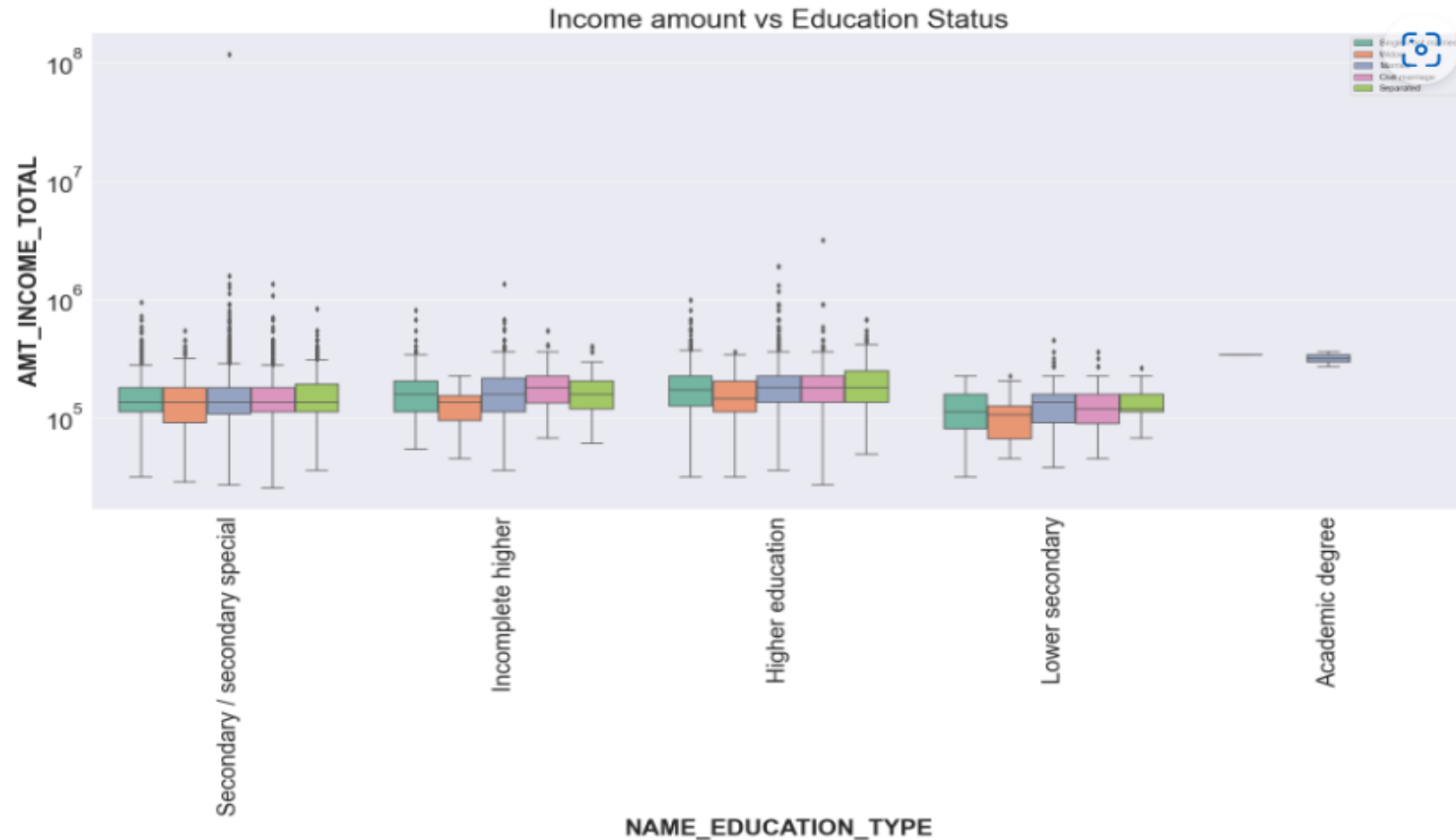
- 
- Female candidates have more % of loans than males
 - 57% females are defaulters while 66.6% are non defaulters
 - 42% males are defaulters while 33.4% are non defaulters

AGE DISTRIBUTION BASED ON TARGET 1 & 2:

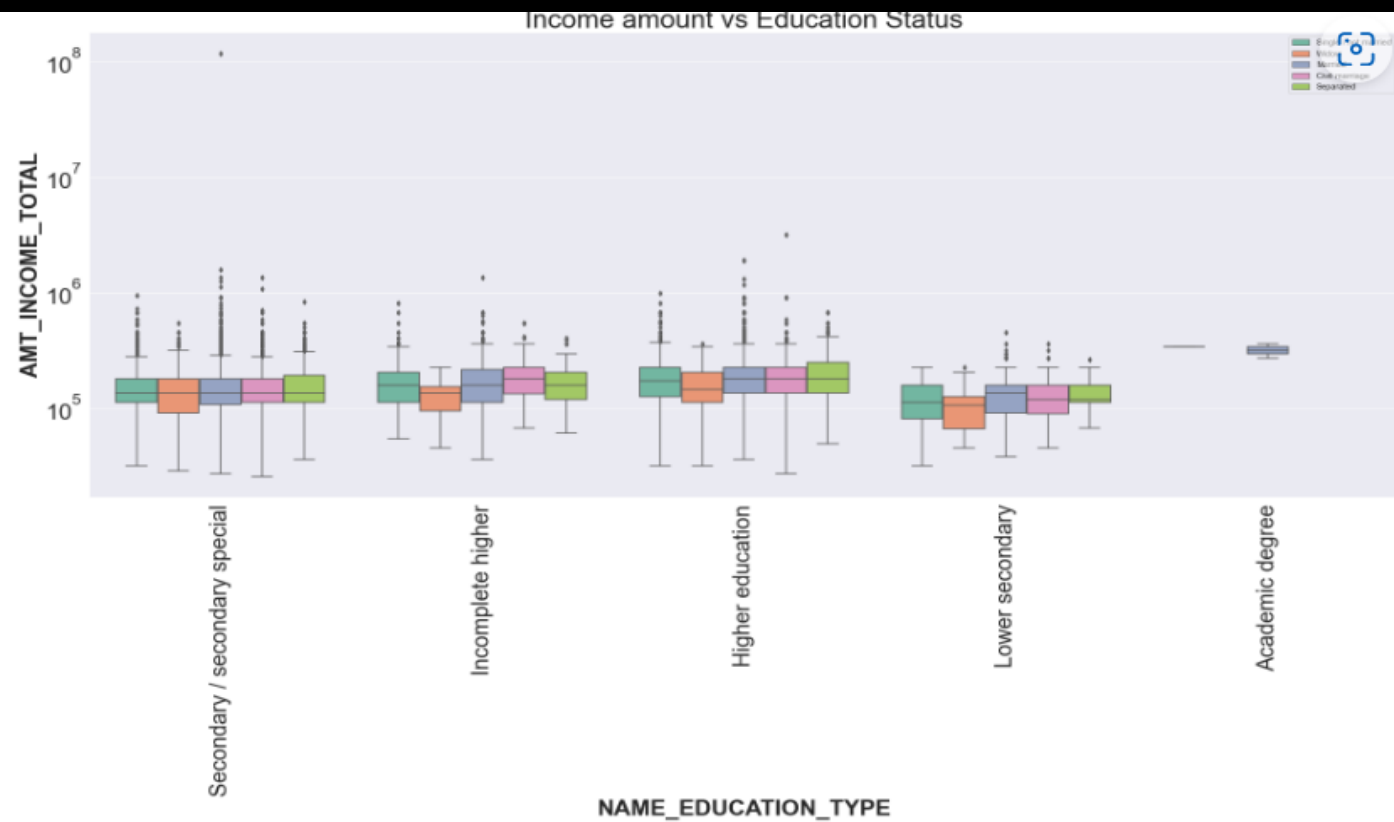


BIVARIATE ANALYSIS FOR TARGET 1 & 2:

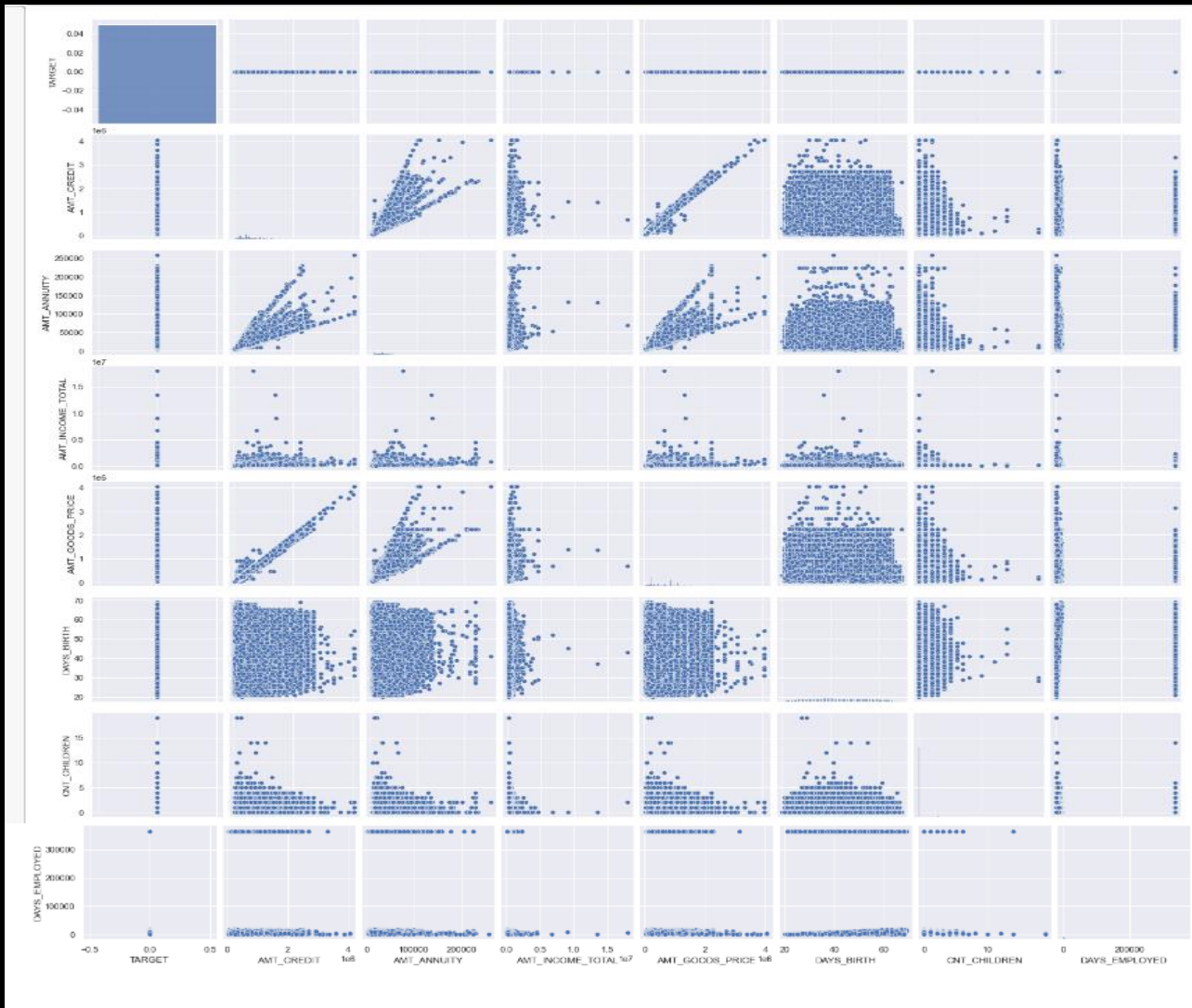
TARGET 1:



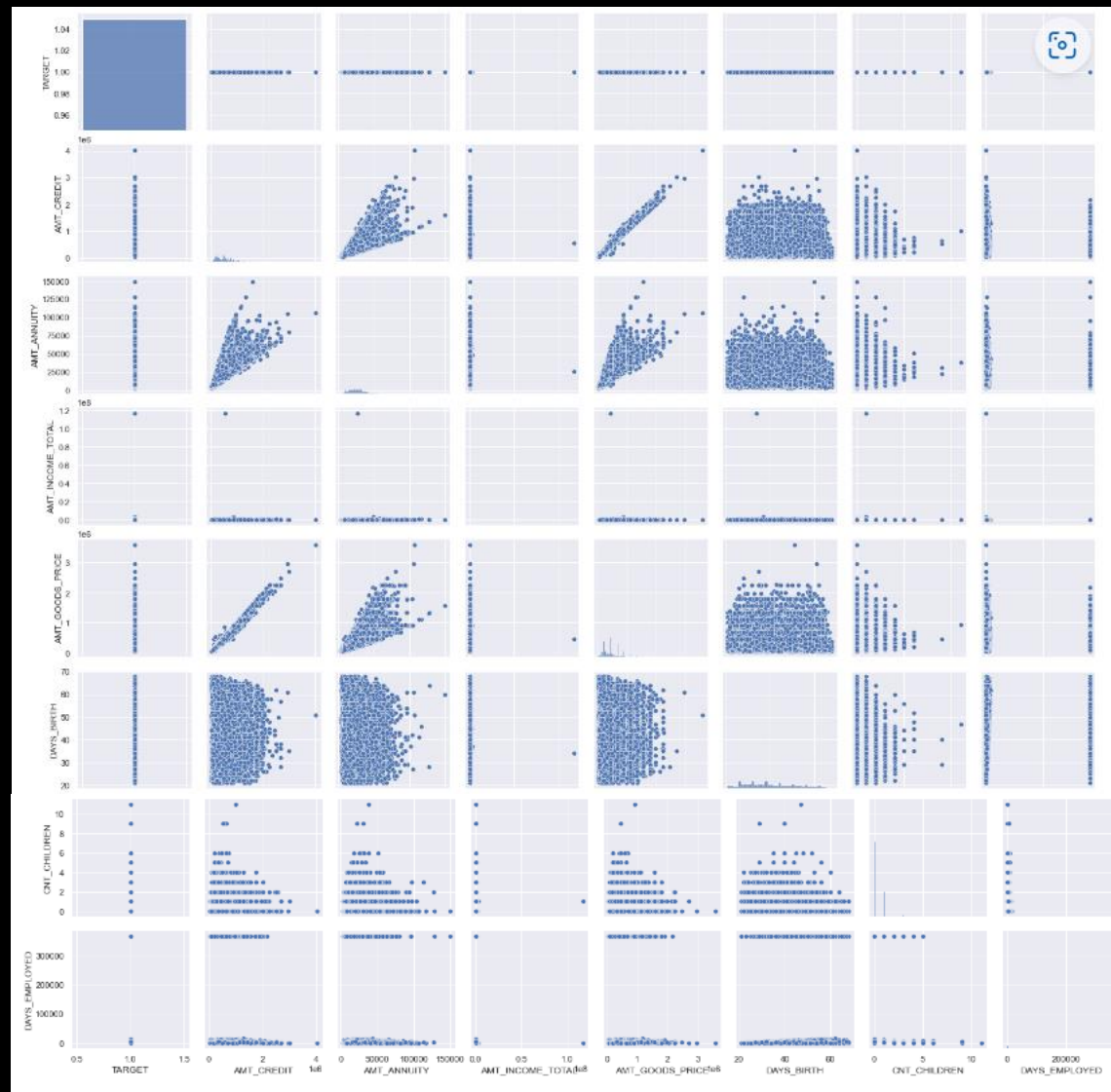
TARGET 2:



CORRELATION BETWEEN NUMERICAL VARIABLES USING PAIRPLOTS: TARGET 1



TARGET 2:



CASE STUDY ON PREVIOUS APPLICATION:

DESCRIPTION OF DATA:

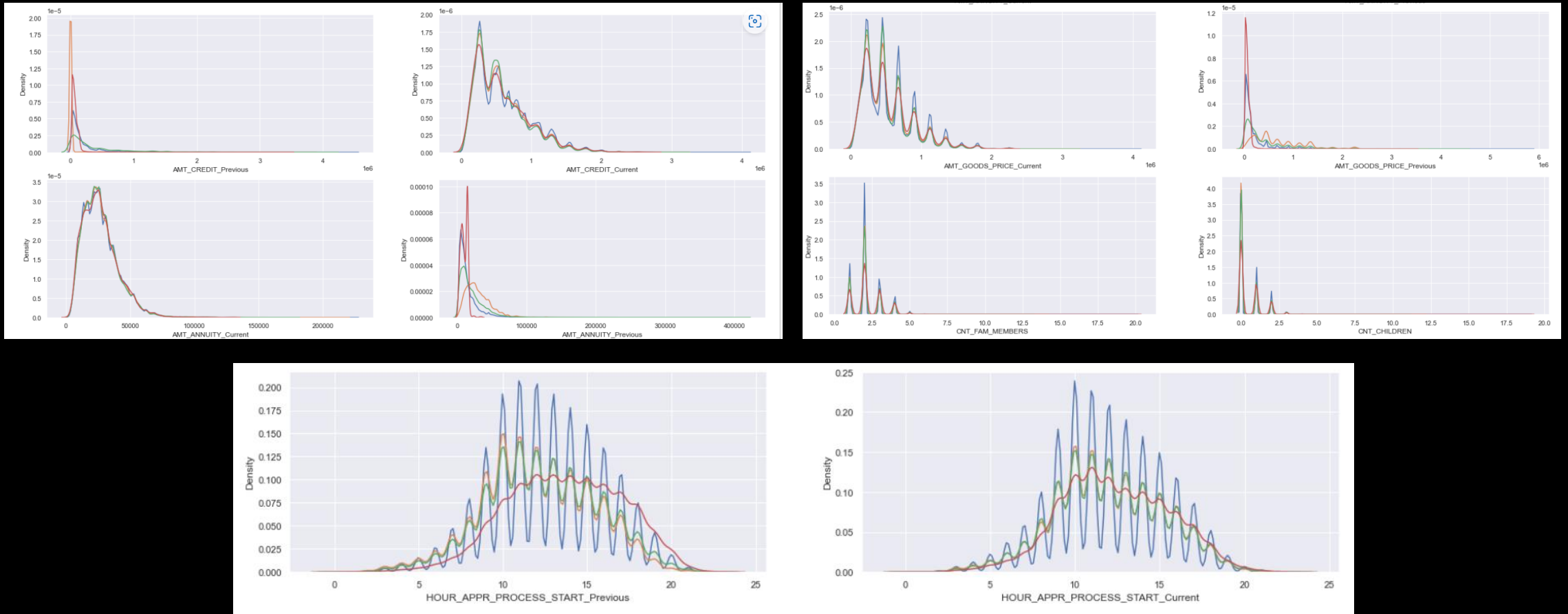
	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE
0	2030495	271877	Consumer loans	1730.430	17145.0	17145.0	0.0	17145.0
1	2802425	108129	Cash loans	25188.615	607500.0	679671.0	NaN	607500.0
2	2523466	122040	Cash loans	15060.735	112500.0	136444.5	NaN	112500.0
3	2819243	176158	Cash loans	47041.335	450000.0	470790.0	NaN	450000.0
4	1784265	202054	Cash loans	31924.395	337500.0	404055.0	NaN	337500.0
...
1670209	2300464	352015	Consumer loans	14704.290	267295.5	311400.0	0.0	267295.5
1670210	2357031	334635	Consumer loans	6622.020	87750.0	64291.5	29250.0	87750.0
1670211	2659632	249544	Consumer loans	11520.855	105237.0	102523.5	10525.5	105237.0
1670212	2785582	400317	Cash loans	18821.520	180000.0	191880.0	NaN	180000.0
1670213	2418762	261212	Cash loans	16431.300	360000.0	360000.0	NaN	360000.0

1670214 rows × 37 columns

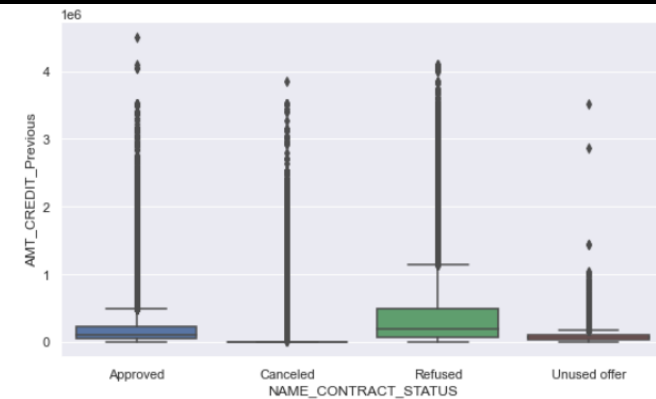
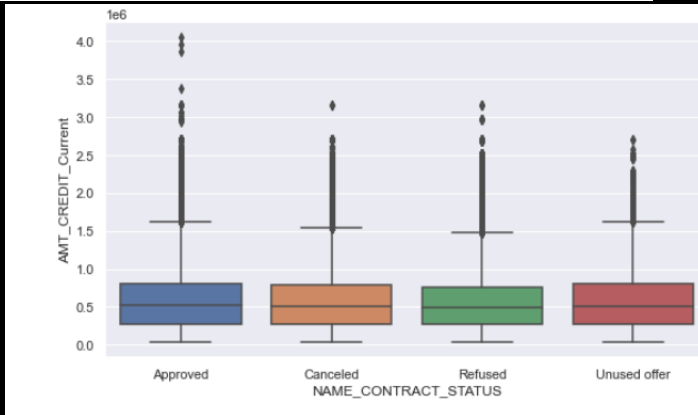
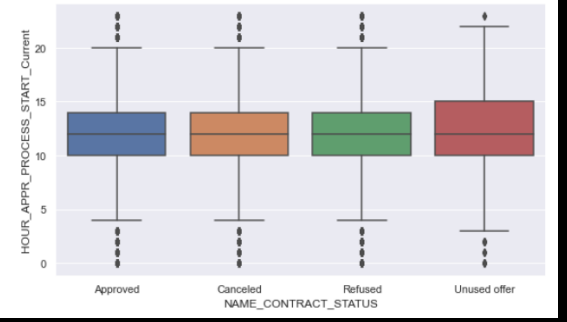
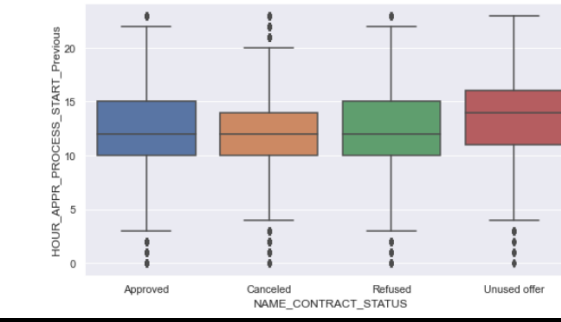
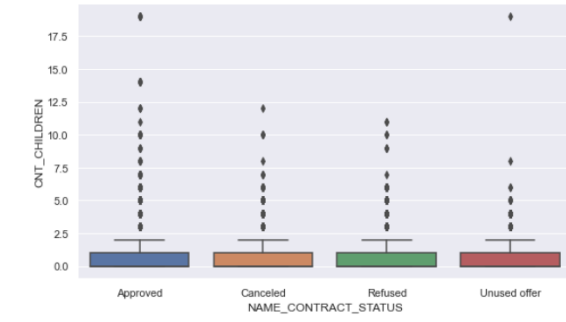
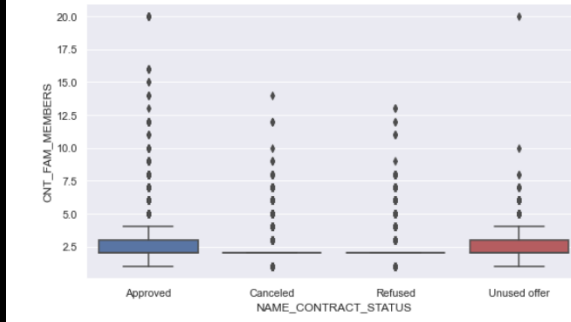
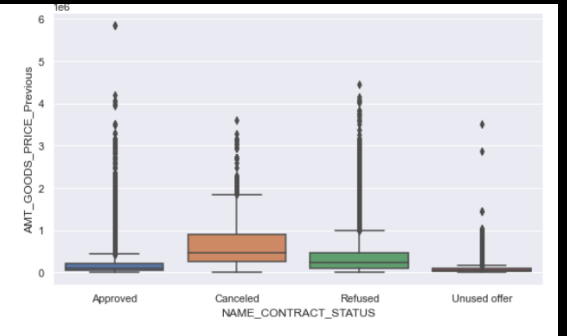
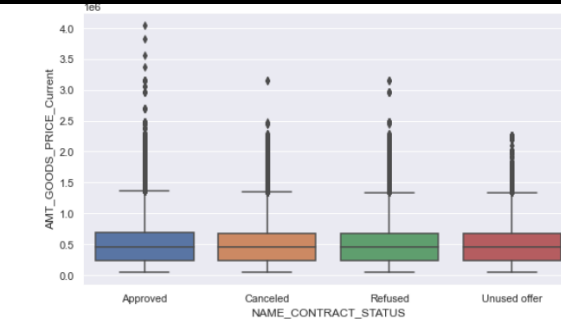
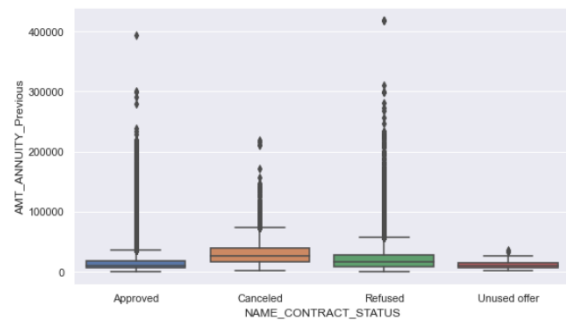
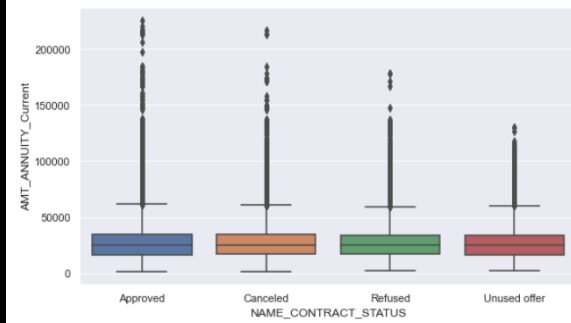
MERGING INP1 & INP2:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE_Current	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	A
0	100002	1	Cash loans	M	N	Y	0	202500.0	
1	100003	0	Cash loans	F	N	N	0	270000.0	
2	100003	0	Cash loans	F	N	N	0	270000.0	
3	100003	0	Cash loans	F	N	N	0	270000.0	
4	100004	0	Revolving loans	M	Y	Y	0	67500.0	
5	100006	0	Cash loans	F	N	Y	0	135000.0	
6	100006	0	Cash loans	F	N	Y	0	135000.0	
7	100006	0	Cash loans	F	N	Y	0	135000.0	
8	100006	0	Cash loans	F	N	Y	0	135000.0	
9	100006	0	Cash loans	F	N	Y	0	135000.0	
10	100006	0	Cash loans	F	N	Y	0	135000.0	

UNIVARIATE NUMERICAL ANALYSIS:

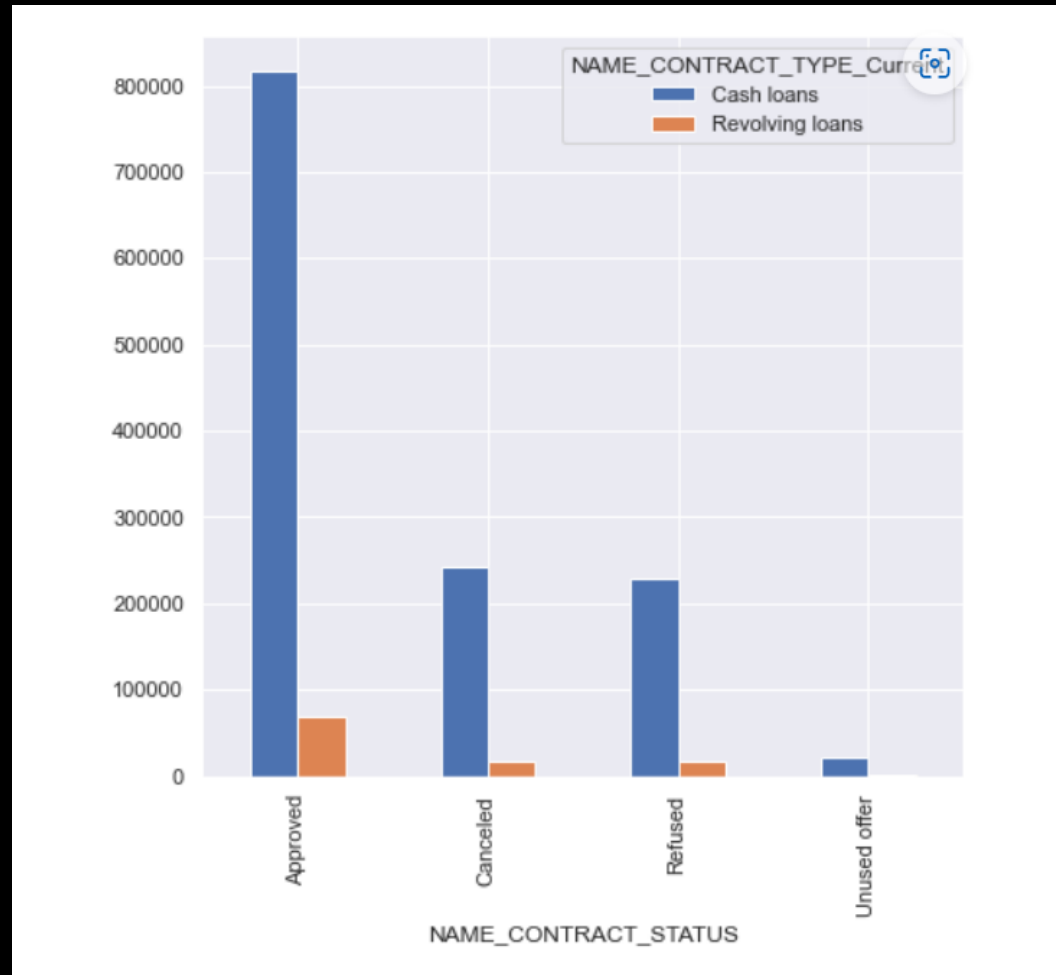


BI-VARIATE CONTINUOUS PLOTS:

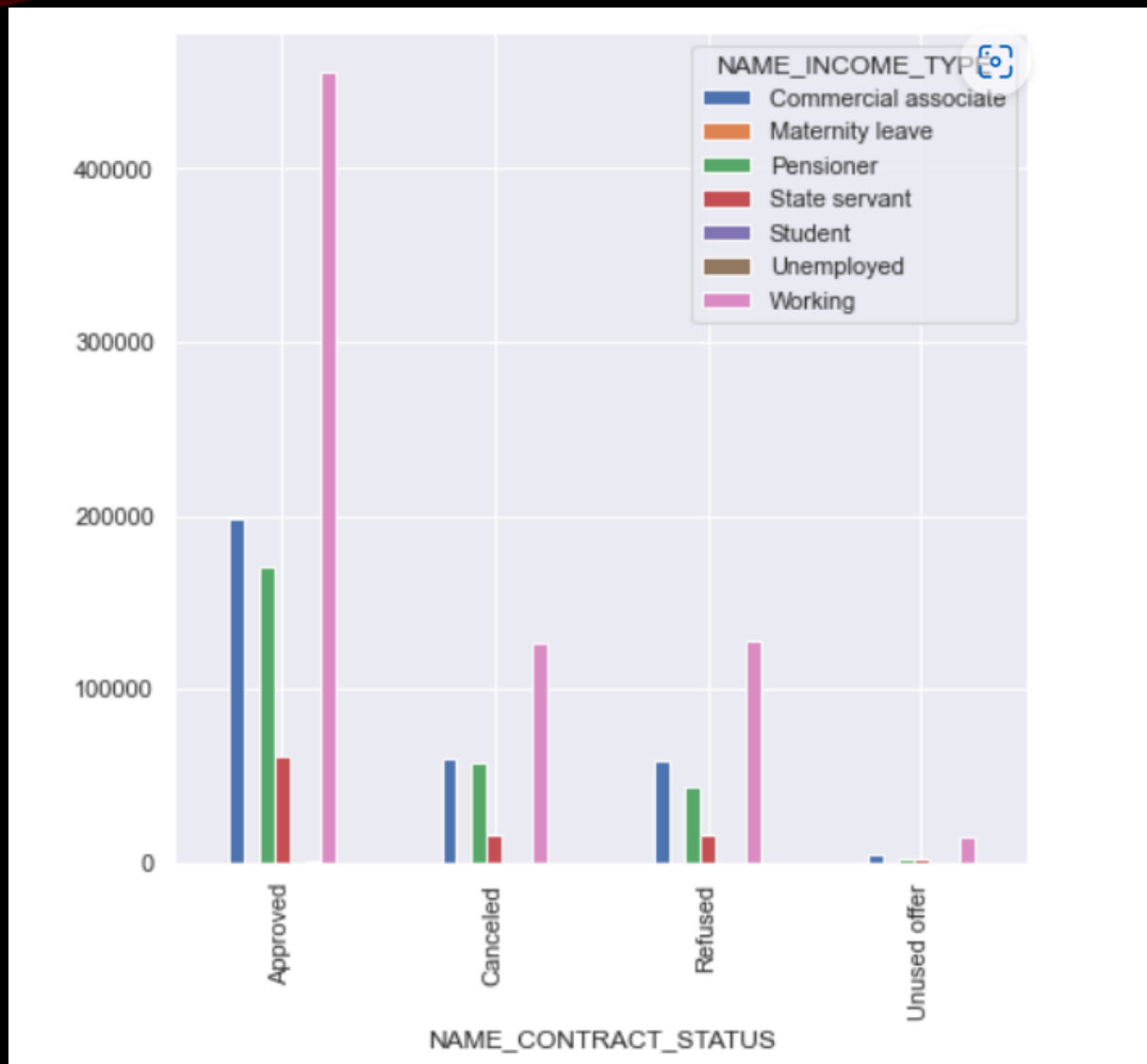


CATEGORIAL BIVARIATE ANALYSIS:

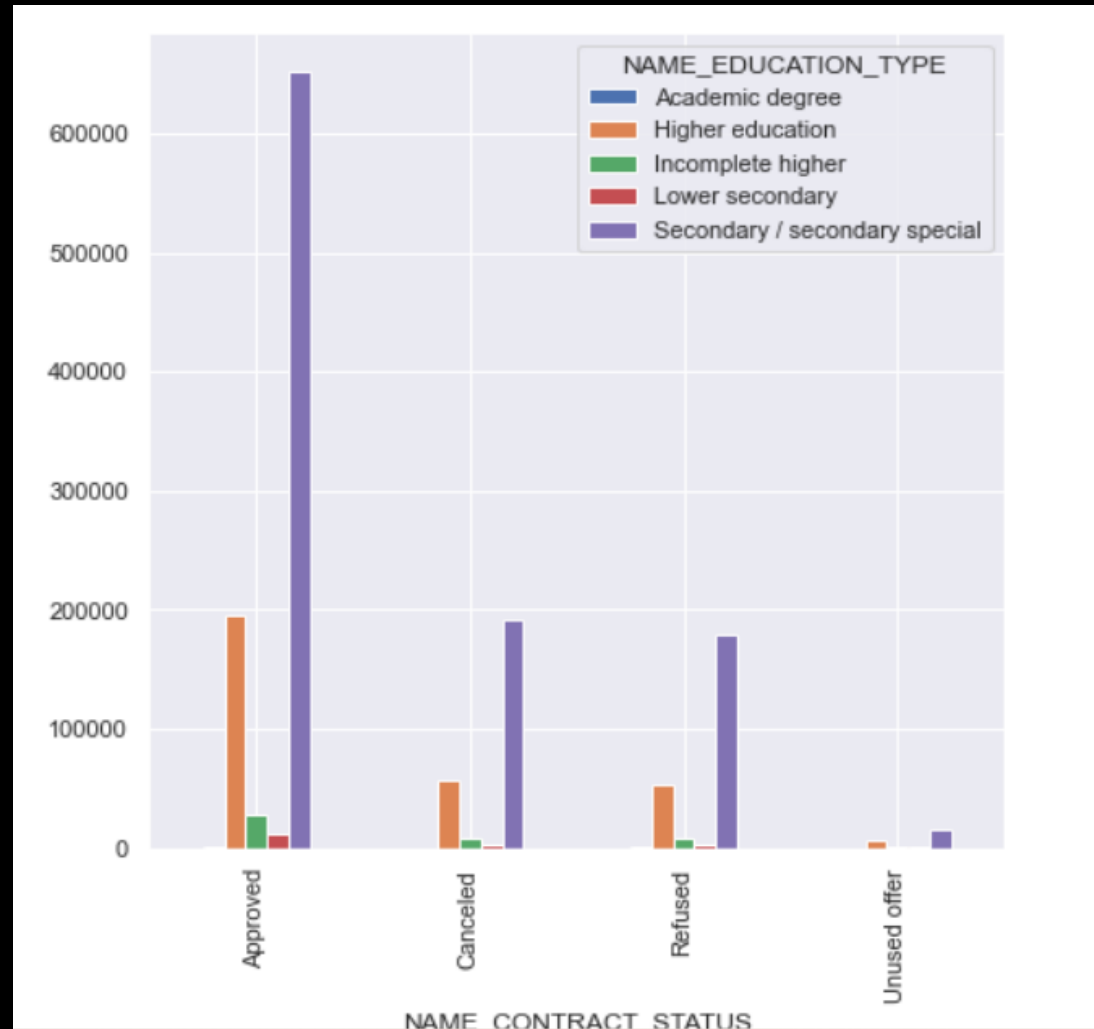
CASH LOANS HAVE HIGHEST APPROVAL OF LOANS:



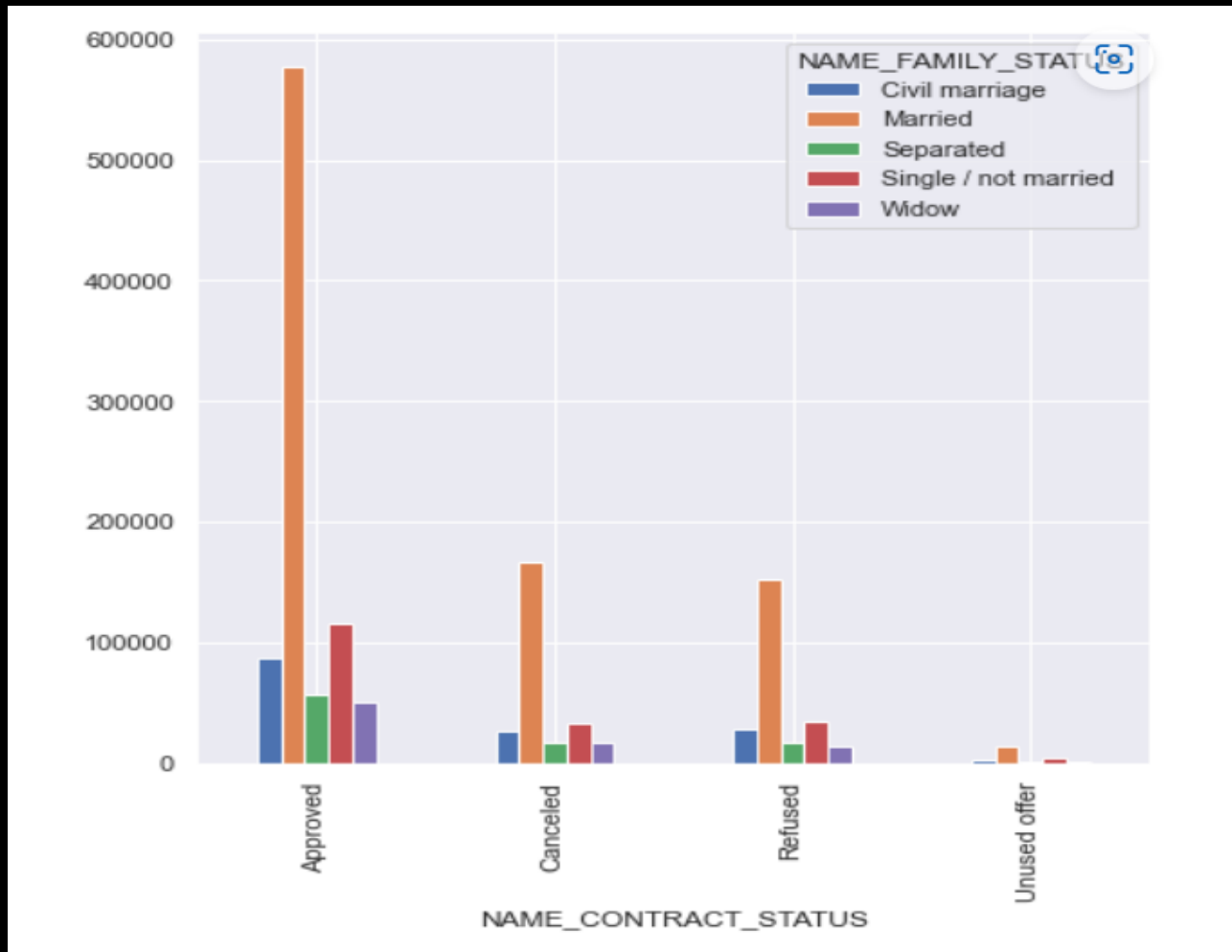
HIGHEST NUMBER OF APPROVALS FOR WORKING APPLICANT:



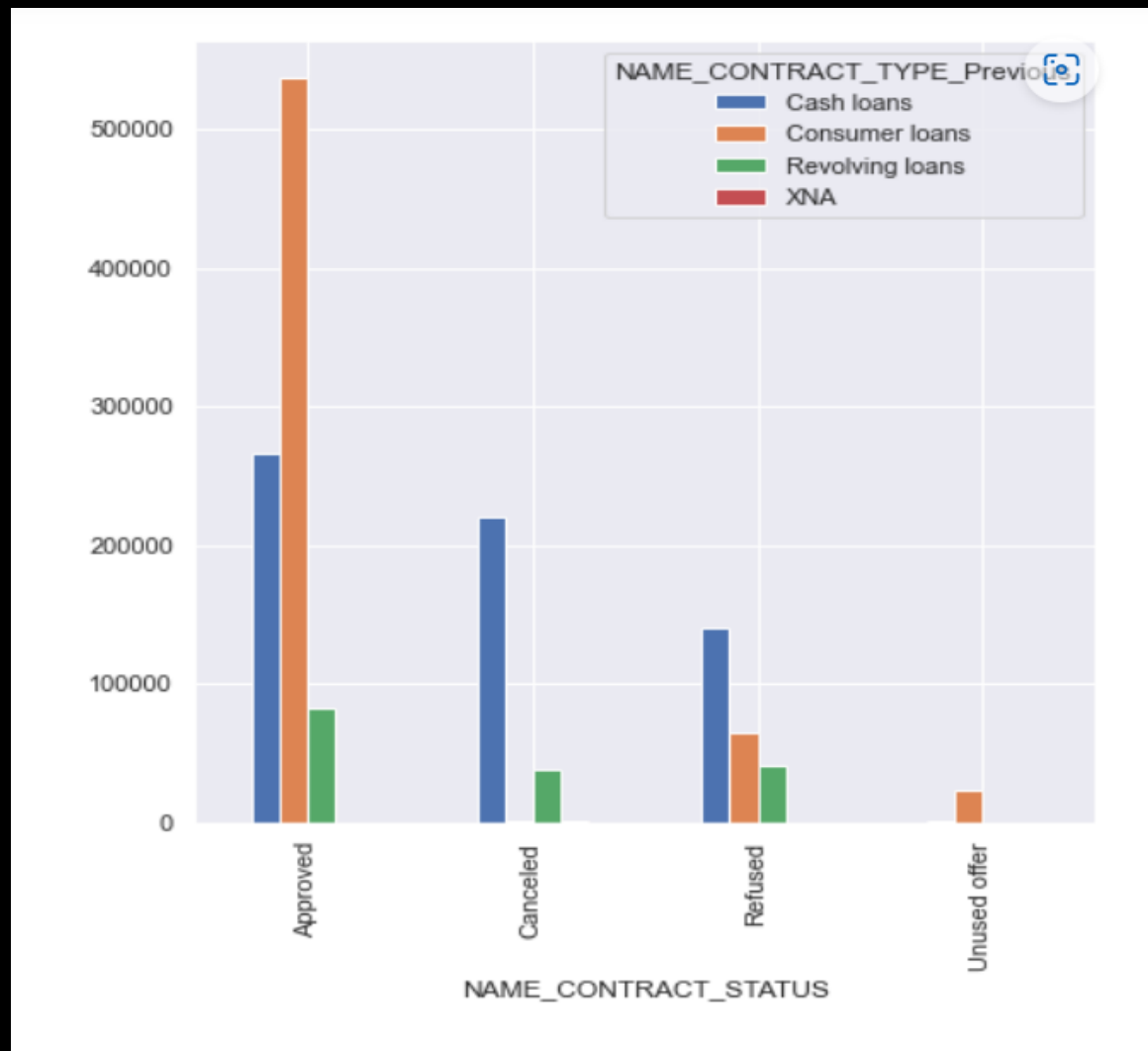
HIGHEST NUMBER OF APPROVALS FOR SECONDARY/SECONDARY SPECIAL EDUCATED APPLICANT



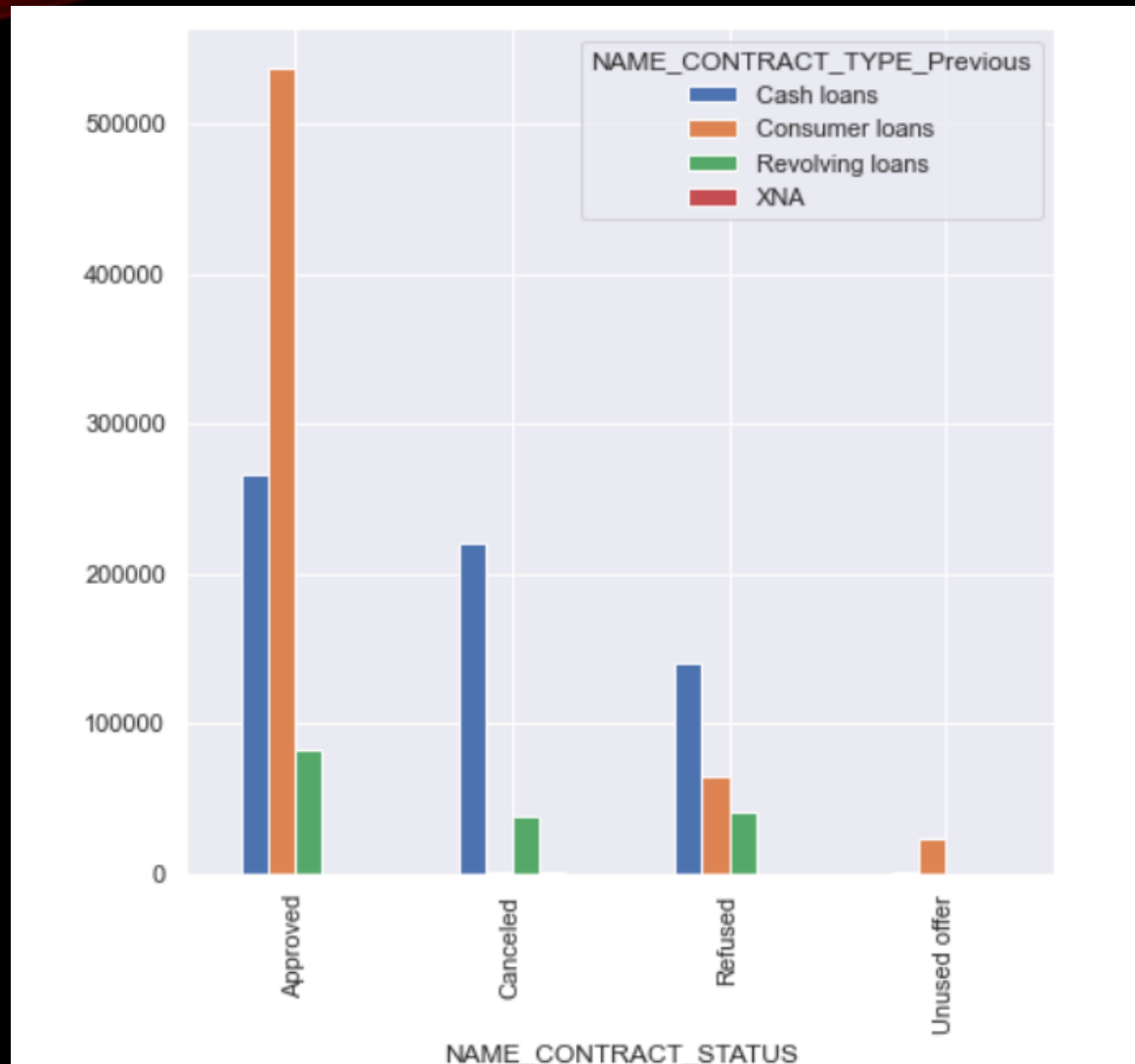
HIGHEST NUMBER OF APPROVALS FOR MARRIED APPLICANT



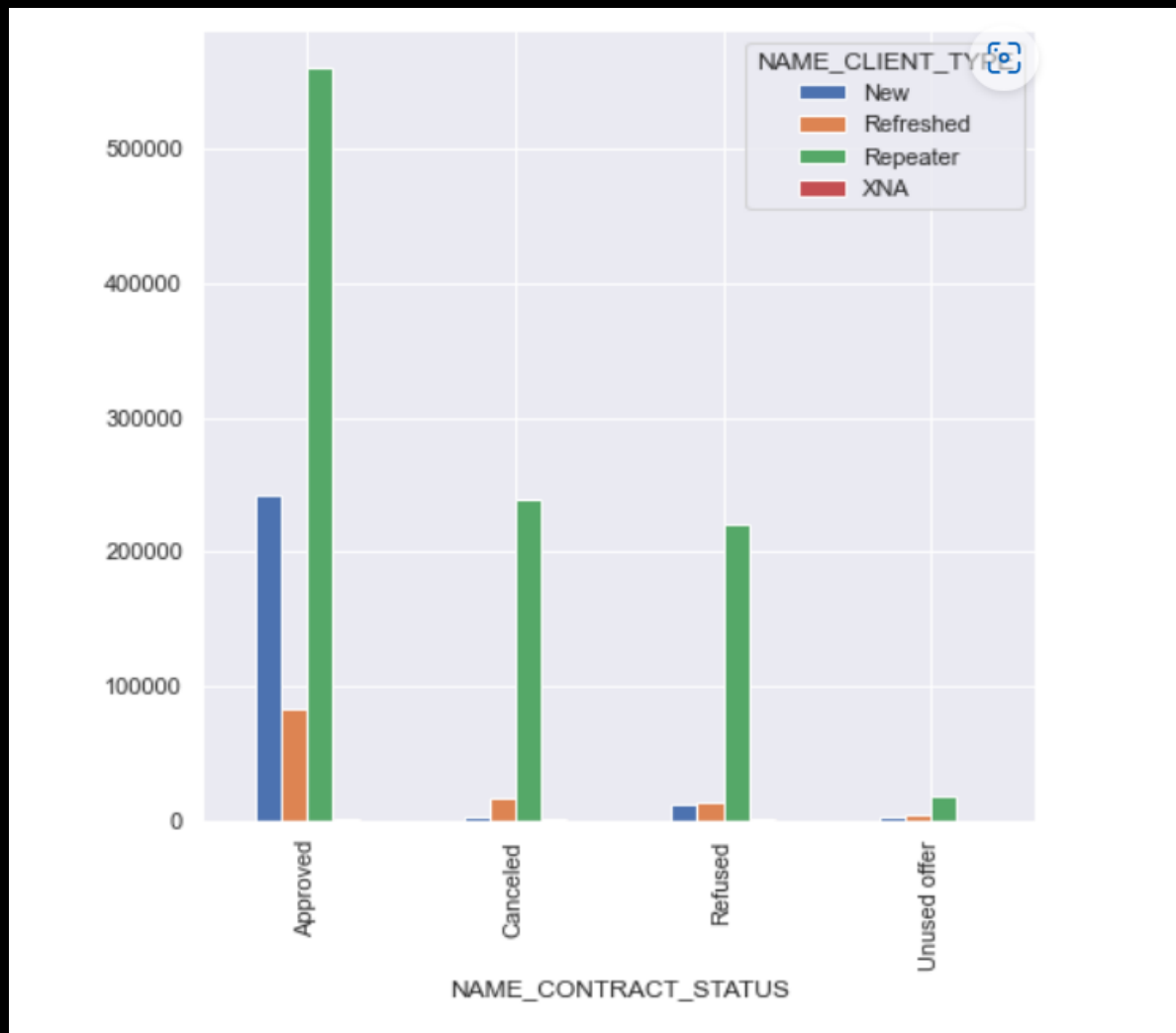
HIGHEST NUMBER OF APPROVALS FOR HOUSE/APARTMENT OWNER




HIGHEST NUMBER OF APPROVALS FOR CONSUMER LOANS



REPEATED APPLICATIONS GOT APPROVED MOST NUMBER OF TIMES



- 
- .AMT_CREDIT_Previous has highest refused cases and AMT_CREDIT_Current is similar for all 4 cases.
 - Near family(2-3 people in family) get highest approval.
 - Previously most of the applications were cancelled or refused but now Refused/Cancelled/Approved/Unused all four have similar situation for AMT_GOODS_PRICE.

-THANK YOU

