

BANK LOAN ANALYSIS

- **PROJECT DESCRIPTION:** The project is about to carry out an exploratory data analysis(EDA) for a bank. Here we are going to consider the different types of EDA(univariate, bivariate and multivariate analysis) in order to carry out the analysis and find the solution to problem statement.
- **APPROACH:** My first approach was to analyse the dataset, clean the dataset finding the blanks and missing values ,imputing the missing values with the appropriate method(mean, median ,mode).Then I tried to find the outliers in the dataset, there are some anomalies such as negative values which need either to be deleted or standardized. After all these I used pivot tables and basic charts to visualise the data. Moreover, insights were drawn based on my understandings
- **TECH -STACK USED:** Microsoft excel 2019 was used for the same .Pivot tables and Charts were also used for visualizing the data. Microsoft word was used for writing the project's description
- **INSIGHTS:** following insights were drawn based on understanding and capabilities:
 - Cash loans have higher percentage as compared to revolving loan
 - Most of the people have no problems regarding the loan payment
 - The defaulters of cash loans are higher than the revolving ones
 - Most of the labourers have problems in loan payment
- **RESULTS:** Following results were obtained while doing the project:

1. Importing the dataset in excel

A	B	C	D	E	F	G	H	I	J	K	L
SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE
100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied
100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family
100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied
100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied
100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied
100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner
100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied
100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied
100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children
100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied
100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied
100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children
100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied
100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied
100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied
100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family
100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied
100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied
100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A
100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied
100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied
100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied
100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied
100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied
100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied
100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied

A	B	C	D	E	F	G	H	I	J
SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HOURL_APPR_PROCESS_START
2030495	271877	Consumer loans	1730.43	17145	17145	0	17145	SATURDAY	11:00
2802425	108129	Cash loans	25188.615	607500	679671		607500	THURSDAY	11:00
2523466	122040	Cash loans	15060.735	112500	136444.5		112500	TUESDAY	11:00
2819243	176158	Cash loans	47041.335	450000	470790		450000	MONDAY	11:00
1784265	202054	Cash loans	31924.395	337500	404055		337500	THURSDAY	11:00
1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	11:00
2315218	175704	Cash loans		0	0			TUESDAY	11:00
1656711	296299	Cash loans		0	0			MONDAY	11:00
2367563	342292	Cash loans		0	0			MONDAY	11:00
2579447	334349	Cash loans		0	0			SATURDAY	11:00
1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	11:00
2257824	161140	Cash loans	13832.775	211500	246397.5		211500	FRIDAY	11:00
2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	11:00
1397919	321676	Consumer loans	7654.86	53779.5	57564	0	53779.5	SUNDAY	11:00
2273188	270658	Consumer loans	9644.22	26550	27252	0	26550	SATURDAY	11:00
1232483	151612	Consumer loans	21307.455	126490.5	119853	12649.5	126490.5	TUESDAY	11:00
2163253	154602	Consumer loans	4187.34	26955	27297	1350	26955	SATURDAY	11:00
1285768	142748	Revolving loans	9000	180000	180000		180000	FRIDAY	11:00
2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	11:00
1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	11:00
1506815	166490	Cash loans	25454.025	450000	491580		450000	MONDAY	11:00
1182516	267782	Cash loans	20361.6	405000	451777.5		405000	SATURDAY	11:00
1172842	302212	Cash loans		0	0			TUESDAY	11:00
1172937	302212	Cash loans	39475.305	1129500	1277104.5		1129500	THURSDAY	11:00
1555330	199353	Cash loans		0	0			SATURDAY	11:00
1543131	275707	Cash loans	22619.52	229500	241920		229500	THURSDAY	11:00

A	B	C	D	E
Column1	Table	Row	Description	Special
2	1 application_data	SK_ID_CURR	ID of loan in our sample	
3	2 application_data	TARGET	Target variable (1 - client with payment difficulties: he/she had	
4	5 application_data	NAME_CONTRACT_TYPE	Identification if loan is cash or revolving	
5	6 application_data	CODE_GENDER	Gender of the client	
6	7 application_data	FLAG_OWN_CAR	Flag if the client owns a car	
7	8 application_data	FLAG_OWN_REALTY	Flag if client owns a house or flat	
8	9 application_data	CNT_CHILDREN	Number of children the client has	
9	10 application_data	AMT_INCOME_TOTAL	Income of the client	
10	11 application_data	AMT_CREDIT	Credit amount of the loan	
11	12 application_data	AMT_ANNUITY	Loan annuity	
12	13 application_data	AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the lo	
13	14 application_data	NAME_TYPE_SUITE	Who was accompanying client when he was applying for the lo	
14	15 application_data	NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,...	
15	16 application_data	NAME_EDUCATION_TYPE	Level of highest education the client achieved	
16	17 application_data	NAME_FAMILY_STATUS	Family status of the client	
17	18 application_data	NAME_HOUSING_TYPE	What is the housing situation of the client (renting, living with	
18	19 application_data	REGION_POPULATION_RELATIVE	Normalized population of region where client lives (higher num normalized	
19	20 application_data	DAYS_BIRTH	Client's age in days at the time of application	time only relative to the application
20	21 application_data	DAYS_EMPLOYED	How many days before the application the person started curri	time only relative to the application
21	22 application_data	DAYS_REGISTRATION	How many days before the application did client change his reg	time only relative to the application
22	23 application_data	DAYS_ID_PUBLISH	How many days before the application did client change the id	time only relative to the application
23	24 application_data	OWN_CAR_AGE	Age of client's car	
24	25 application_data	FLAG_MOBIL	Did client provide mobile phone (1=YES, 0=NO)	
25	26 application_data	FLAG_EMP_PHONE	Did client provide work phone (1=YES, 0=NO)	
26	27 application_data	FLAG_WORK_PHONE	Did client provide home phone (1=YES, 0=NO)	
27	28 application_data	FLAG_CONT_MOBILE	Was mobile phone reachable (1=YES, 0=NO)	
28	29 application_data	FLAG_PHONE	Did client provide home phone (1=YES, 0=NO)	
29	30 application_data	FLAG_EMAIL	Did client provide email (1=YES, 0=NO)	

• Highlighting the blank cells:

A	B	C	D	E	F	G	H	I	J
SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HOURL_APPR_PROCESS_ST
2030495	271877	Consumer loans	1730.43	17145	17145	0	17145	SATURDAY	
2802425	108129	Cash loans	25188.615	607500	679671		607500	THURSDAY	
2523466	122040	Cash loans	15060.735	112500	136444.5		112500	TUESDAY	
2819243	176158	Cash loans	47041.335	450000	470790		450000	MONDAY	
1784265	202054	Cash loans	31924.395	337500	404055		337500	THURSDAY	
1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	
2315218	175704	Cash loans		0	0			TUESDAY	
1656711	296299	Cash loans		0	0			MONDAY	
2367563	342292	Cash loans		0	0			MONDAY	
2579447	334349	Cash loans		0	0			SATURDAY	
1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	
2257824	161140	Cash loans	13832.775	211500	246397.5		211500	FRIDAY	
2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	
1397919	321676	Consumer loans	7654.86	53779.5	57564	0	53779.5	SUNDAY	
2273188	270658	Consumer loans	9644.22	26550	27252	0	26550	SATURDAY	
1232483	151612	Consumer loans	21307.455	126490.5	119853	12649.5	126490.5	TUESDAY	
2163253	154602	Consumer loans	4187.34	26955	27297	1350	26955	SATURDAY	
1285768	142748	Revolving loans	9000	180000	180000		180000	FRIDAY	
2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	
1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	
1506815	166490	Cash loans	25454.025	450000	491580		450000	MONDAY	
1182516	267782	Cash loans	20361.6	405000	451777.5		405000	SATURDAY	
1172842	302212	Cash loans		0	0			TUESDAY	
1172937	302212	Cash loans	39475.305	1129500	1277104.5		1129500	THURSDAY	
1555330	199353	Cash loans		0	0			SATURDAY	
1543131	275707	Cash loans	22619.52	229500	241920		229500	THURSDAY	

2. Finding the occurrences of blanks cells(column wise) and deleting them if they are greater than 50%

	A	B	C	D	E	F
1	Field name	Occurrence of the null value	percentage Occurrence	dropping column		
2	AMT_ANNUITY	12	0%	No	no of columns to be dropped	46
3	AMT_GOODS_PRICE	0	0%	No	no of columns to be left	21
4	NAME_TYPE_SUITE	0	0%	No		
5	OWN_CAR_AGE	#REF!	#REF!	#REF!		
5	OCCUPATION_TYPE	0	0%	No		
7	CNT_FAM_MEMBERS	0	0%	No		
3	EXT_SOURCE_1	#REF!	#REF!	#REF!		
3	EXT_SOURCE_2	0	0%	No		
0	EXT_SOURCE_3	0	0%	No		
1	APARTMENTS_AVG	#REF!	#REF!	#REF!		
2	BASEMENTAREA_AVG	#REF!	#REF!	#REF!		
3	YEARS_BEGINEXPLUATATION_AVG	0	0%	No		
4	YEARS_BUILD_AVG	#REF!	#REF!	#REF!		
5	COMMONAREA_AVG	#REF!	#REF!	#REF!		
6	ELEVATORS_AVG	#REF!	#REF!	#REF!		
7	ENTRANCES_AVG	#REF!	#REF!	#REF!		
8	FLOORSMAX_AVG	0	0%	No		
9	FLOORSMIN_AVG	#REF!	#REF!	#REF!		
0	LANDAREA_AVG	#REF!	#REF!	#REF!		
1	LIVINGAPARTMENTS_AVG	#REF!	#REF!	#REF!		
2	LIVINGAREA_AVG	#REF!	#REF!	#REF!		
3	NONLIVINGAPARTMENTS_AVG	#REF!	#REF!	#REF!		
4	NONLIVINGAREA_AVG	#REF!	#REF!	#REF!		
5	APARTMENTS_MODE	#REF!	#REF!	#REF!		
6	BASEMENTAREA_MODE	#REF!	#REF!	#REF!		
7	YEARS_BEGINEXPLUATATION_MODE	0	0%	No		

previous_application | previous_application (2) | application_data (2) | Cleaning_application_data | Sheet1 | Targets ...

#Ref is because these fields were deleted from the dataset

- Imputing the missing values(using mean , median and mode)

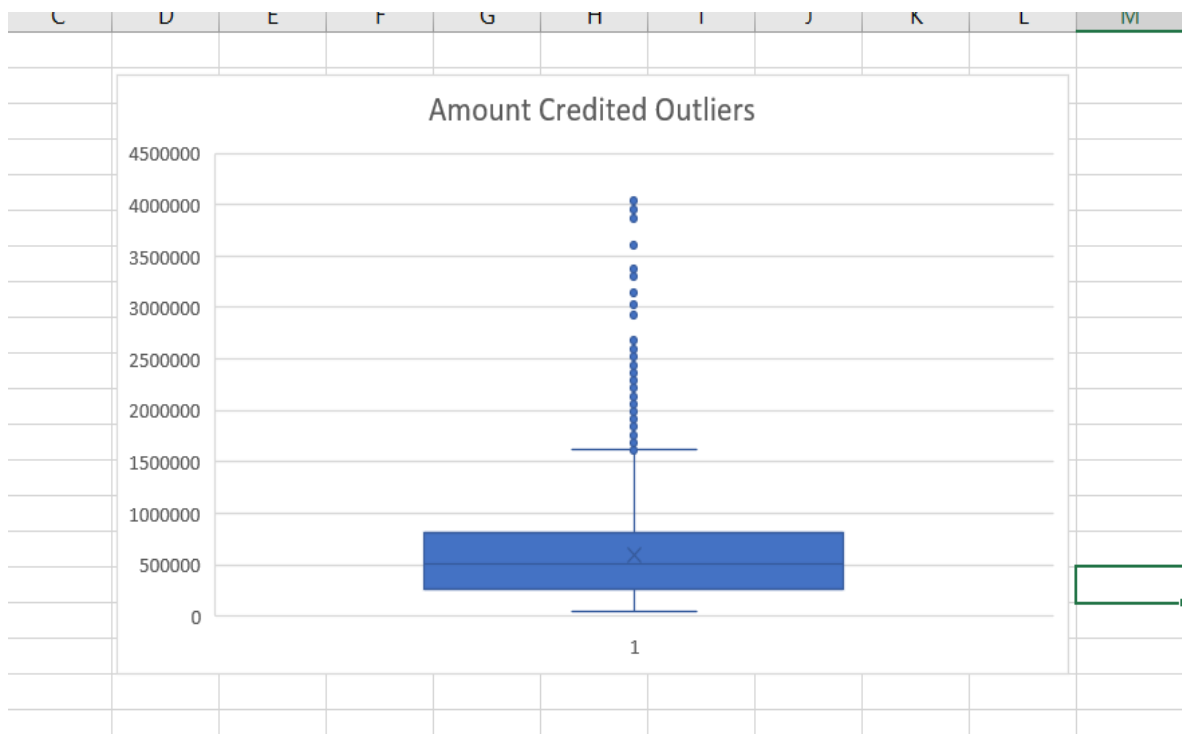
A	B	C	D	E	F	G	H	I	J	K	L
SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE
100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied
100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family
100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied
100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied
100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied
100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner
100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied
100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied
100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children
100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied
100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied
100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children
100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied
100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied
100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied
100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family
100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied
100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied
100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A
100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied
100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied
100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied
100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied
100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied
100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied
100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied

previous_application (2) | application_data (2) | Cleaning_application_data | Sheet1 | Targets and amt annuity | Sh ...

	AM	AN	AO	AP	AQ	AR	AS	AT	
	LIVE_CITY_NOT_WORK_CITY	ORGANIZATION_TYPE	EXT_SOURCE_2	EXT_SOURCE_3	YEARS_BEGINEXPLUATATION_AVG	FLOORSMAX_AVG	YEARS_BEGINEXPLUATATION_MODE	OBS_30_CNT_SOCIAL_CIRCLE	DEF_3
2		0 Business Entity Type 3	0.262948593	0.13937578	0.9722	0.0833	0.9722		2
3		0 School	0.622245775	0.53527625	0.9851	0.2917	0.9851		1
1		0 Government	0.555912083	0.729566691	0.977734858	0.226281907	0.9871		0
5		0 Business Entity Type 3	0.65044169	0.53527625	0.977734858	0.226281907	0.9871		2
5		1 Religion	0.322738287	0.53527625	0.977734858	0.226281907	0.9871		0
7		0 Other	0.354224732	0.621226338	0.977734858	0.226281907	0.9871		0
3		0 Business Entity Type 3	0.723999852	0.492060094	0.977734858	0.226281907	0.9871		1
3		1 Other	0.714279286	0.54065445	0.977734858	0.226281907	0.9871		2
0		0 XNA	0.205747288	0.751723715	0.977734858	0.226281907	0.9871		1
1		0 Electricity	0.746643629	0.53527625	0.977734858	0.226281907	0.9871		2
2		0 Medicine	0.651862333	0.363945239	0.977734858	0.226281907	0.9871		0
3		0 XNA	0.555183162	0.652896552	0.977734858	0.226281907	0.9871		0
4		0 Business Entity Type 2	0.715041819	0.176652579	0.977734858	0.226281907	0.9871		0
5		0 Self-employed	0.566906613	0.77008707	0.977734858	0.226281907	0.9871		0
6		0 Transport: type 2	0.642656205	0.53527625	0.977734858	0.226281907	0.9871		0
7		0 Business Entity Type 2	0.346633981	0.678567689	0.977734858	0.226281907	0.9871		0
8		0 Government	0.23637784	0.062103038	0.977734858	0.226281907	0.9871		0
9		0 Construction	0.683513346	0.53527625	0.977734858	0.226281907	0.9871		4
0		0 Housing	0.706428403	0.556727426	0.977734858	0.226281907	0.9871		8
1		0 Kindergarten	0.58661714	0.477649155	0.977734858	0.226281907	0.9871		0
2		0 Self-employed	0.113374513	0.53527625	0.977734858	0.226281907	0.9871		0
3		0 Trade: type 7	0.233766958	0.542445144	0.977734858	0.226281907	0.9871		0
4		0 Self-employed	0.457142972	0.358951229	0.977734858	0.226281907	0.9871		0
5		0 XNA	0.624304737	0.669056695	0.977734858	0.226281907	0.9871		0
6		1 Business Entity Type 3	0.786179309	0.565607981	0.977734858	0.226281907	0.9871		1
7		0 Business Entity Type 3	0.651405637	0.461482391	0.977734858	0.226281907	0.9871		0

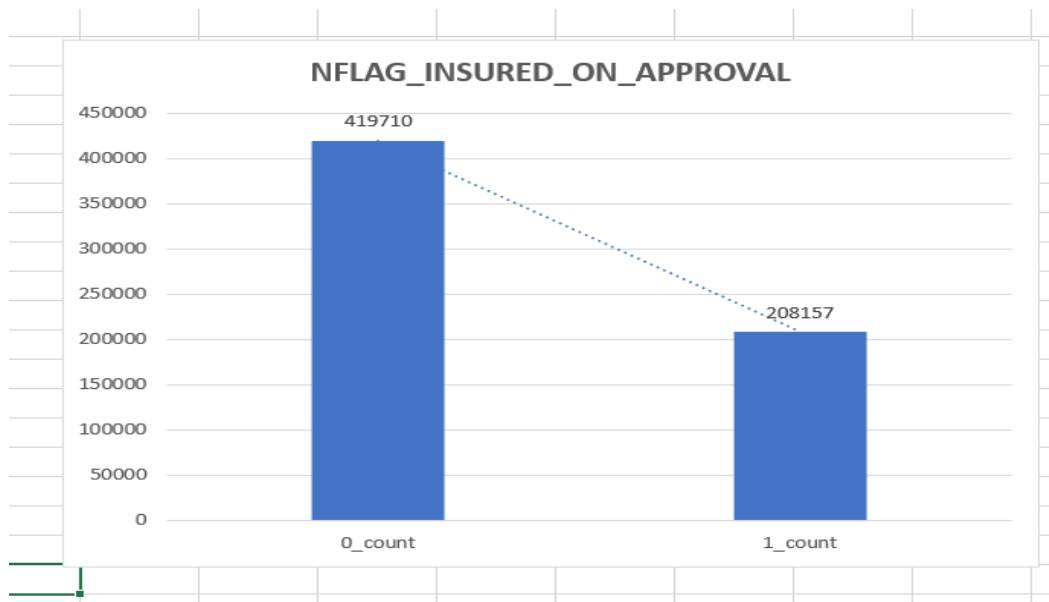
3. Identifying the outliers in the dataset

- In Amount Credited you can see a significant amount of outliers from the mean value of the entire dataset. It is an outlier as a significant data point is shown in the Box Plot as outlier towards the amount 4500000.

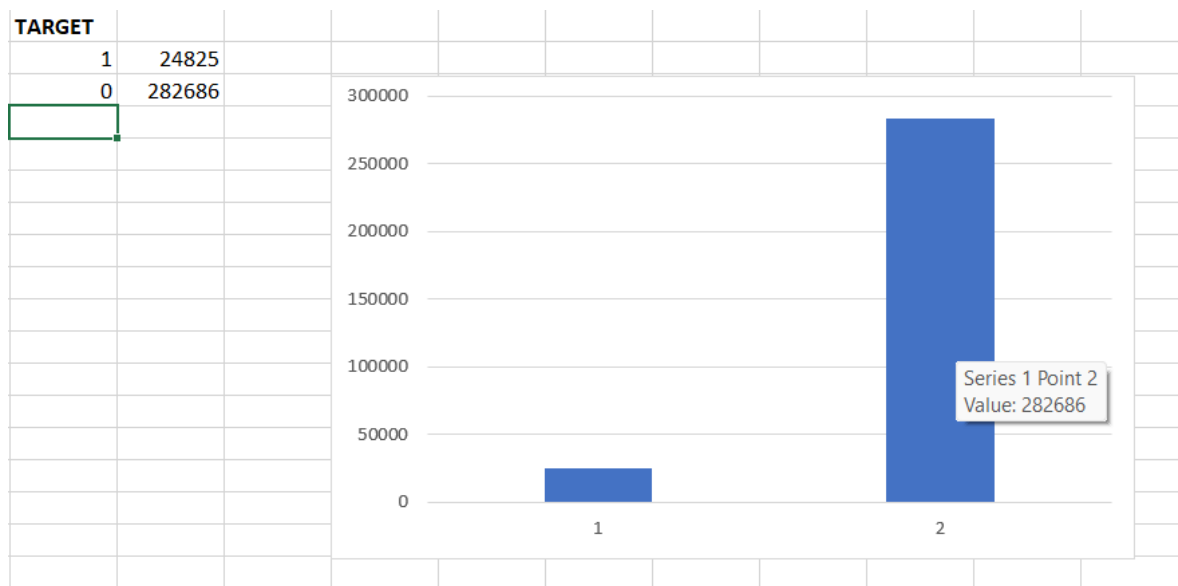


4. Identify if there is imbalance in the data

The ratio of NFLAG_INSURED_ON_APPROVAL is 665527:331622

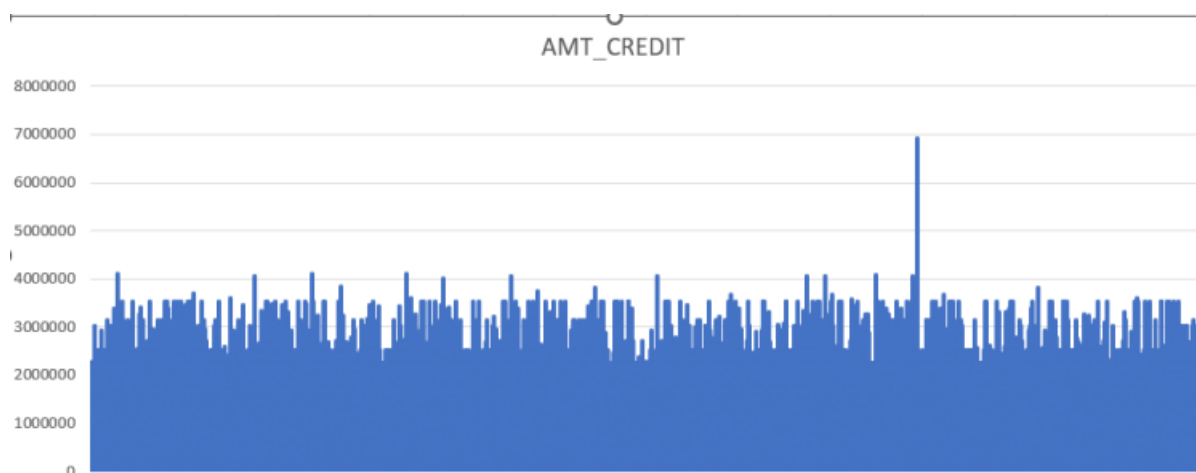


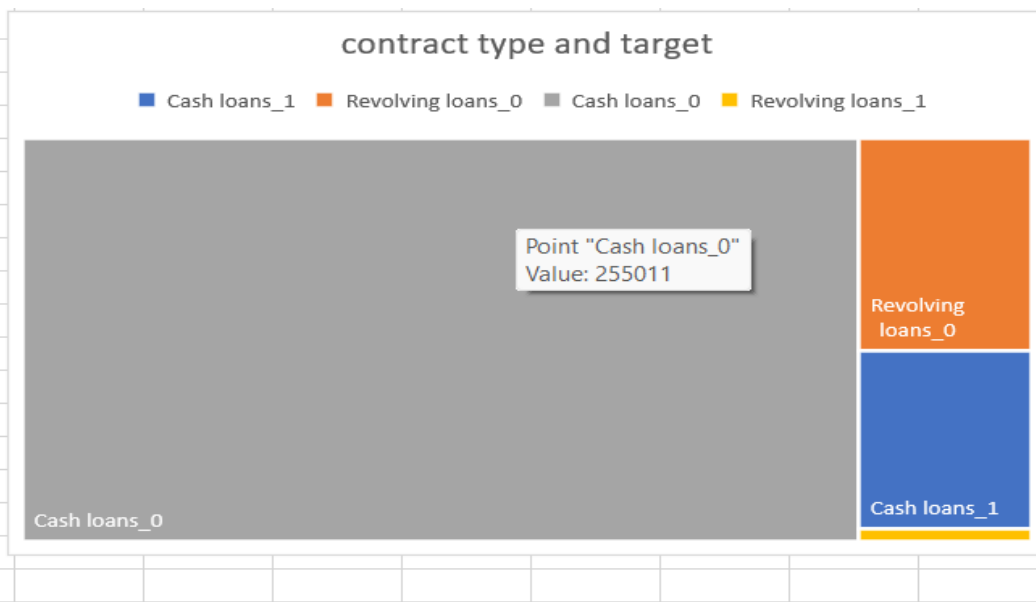
there is also imbalance between the targets 0 and 1



5. Explain the results of univariate ,segmented variate and bivariate analysis

- The results of univariate, segmented and bivariate analysis using visual means are as follows:





A	B	C	D	E	F	G	H
NAME_CONTRACT_TYPE	Cash loans						
Sum of AMT_ANNUIITY	Column Labels						
Row Labels	F	M					
100003	35698.5						
100006	29686.5						
100007		21865.5					
100008		27517.5					
100009	41301						
100010		42075					
100011	33826.5						
100014	21177						
100015	10678.5						
100016	5881.5						
100017		28966.5					
100018	32778						
100019		20160					
100020		26149.5					
100023	17563.5						
100025	37561.5						
100026	32521.5						
100027	23850						
100029		12703.5					
100030	11074.5						
100032		23827.5					
100033		57676.5					
100035	24592.5						
100036	25033.5						
100037	20893.5						
100039		39069					
100040	32895						
100041	44509.5						

TARGET

0

1

Targets and am

A	B	C	D	E	F	G	H
NAME_CONTRACT_TYPE	Cash loans						
Sum of AMT_ANNUIITY	Column Labels						
Row Labels	F	M					
100002		24700.5					
100031	27076.5						
100047		35028					
100049	16258.5						
100096	14593.5						
100112		64107					
100130	30802.5						
100160		36747					
100181	12667.5						
100192	21037.5						
100214	28516.5						
100246	26995.5						
100273	66262.5						
100286		17298					
100295		31032					
100300		22468.5					
100301		31131					
100326		10849.5					
100336	28692						
100396		25330.5					
100401	16663.5						
100424		12433.5					
100439		22891.5					
100452		36391.5					
100472		26509.5					
100477		22599					
100490		38686.5					
100540	26901						

TARGET

0

1

A	B	C	D	E	F	G	H
NAME_CONTRACT_TYPE	Revolving loans						
Sum of AMT_ANNUIITY	Column Labels						
Row Labels	F	M	XNA				
100004		6750					
100012		20250					
100021	13500						
100022	7875						
100024		21375					
100034		9000					
100046		27000					
100052	9000						
100058	6750						
100068		12375					
100079		13500					
100080	22500						
100088	6750						
100095	6750						
100098		13500					
100119	9000						
100126		9000					
100129	6750						
100134	9000						
100140	33750						
100143		13500					
100154		9000					
100174	11250						
100182	6750						
100194	15750						
100226	13500						
100230	20250						
100234	9000						

TARGET

0

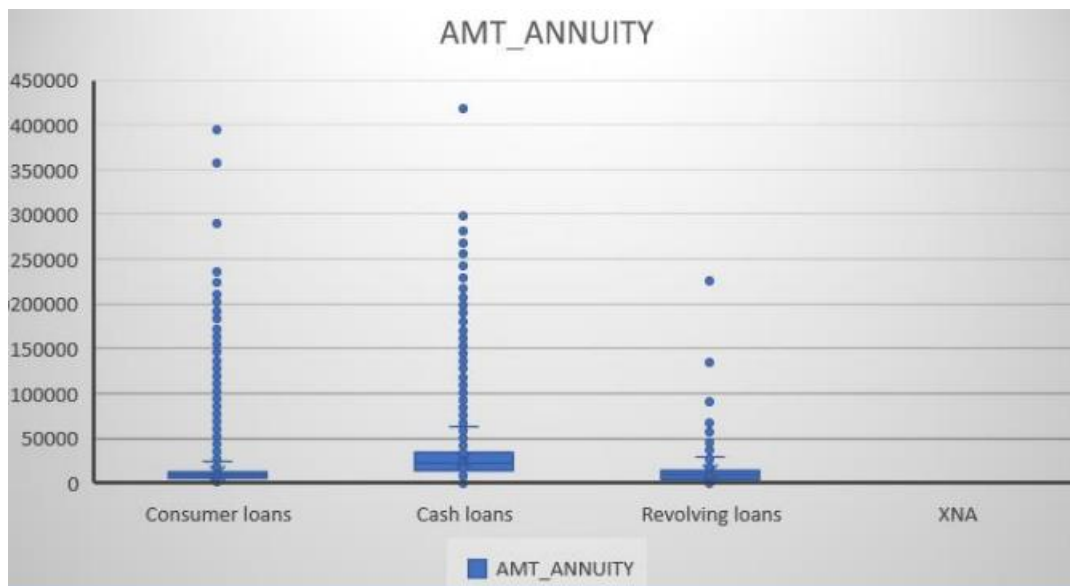
1

	A	B	C	D	E	F	G	H
1	NAME_CONTRACT_TYPE	Revolving loans						
2								
3	Sum of AMT_ANNUITY	Column Labels						
4	Row Labels	F	M					
5	100209		27000					
6	100282	6750						
7	100485		40500					
8	100567		9000					
9	101165	11250						
10	101330		9000					
11	101575		16875					
12	101705		13500					
13	101929	22500						
14	102128		10125					
15	102208		7875					
16	102955	13500						
17	102994	12375						
18	103059	9000						
19	103113		9000					
20	103134		16875					
21	103407	9000						
22	103422	13500						
23	103439		16875					
24	103810	6750						
25	104107		13500					
26	104315	13500						
27	104806	13500						
28	105885	9000						
29	105917	9000						
30	106066		10125					
31	106508		10125					
32	106518	9000						

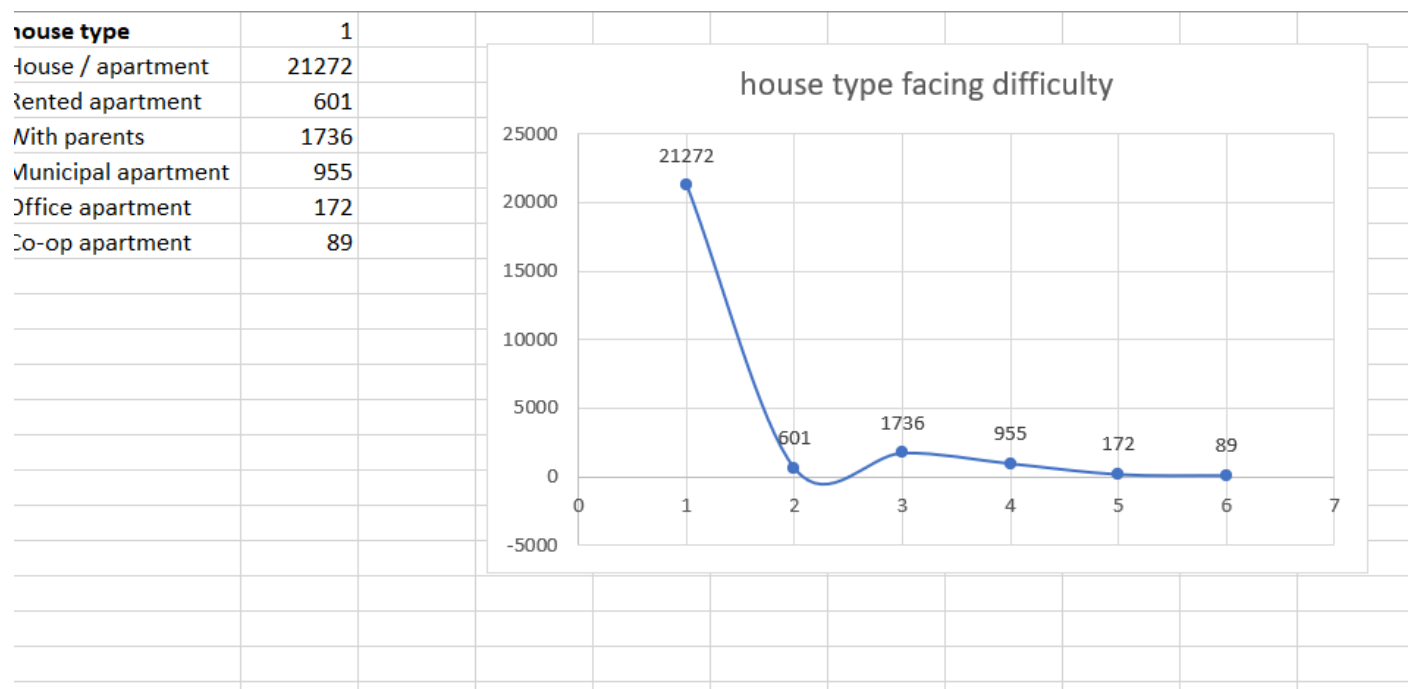
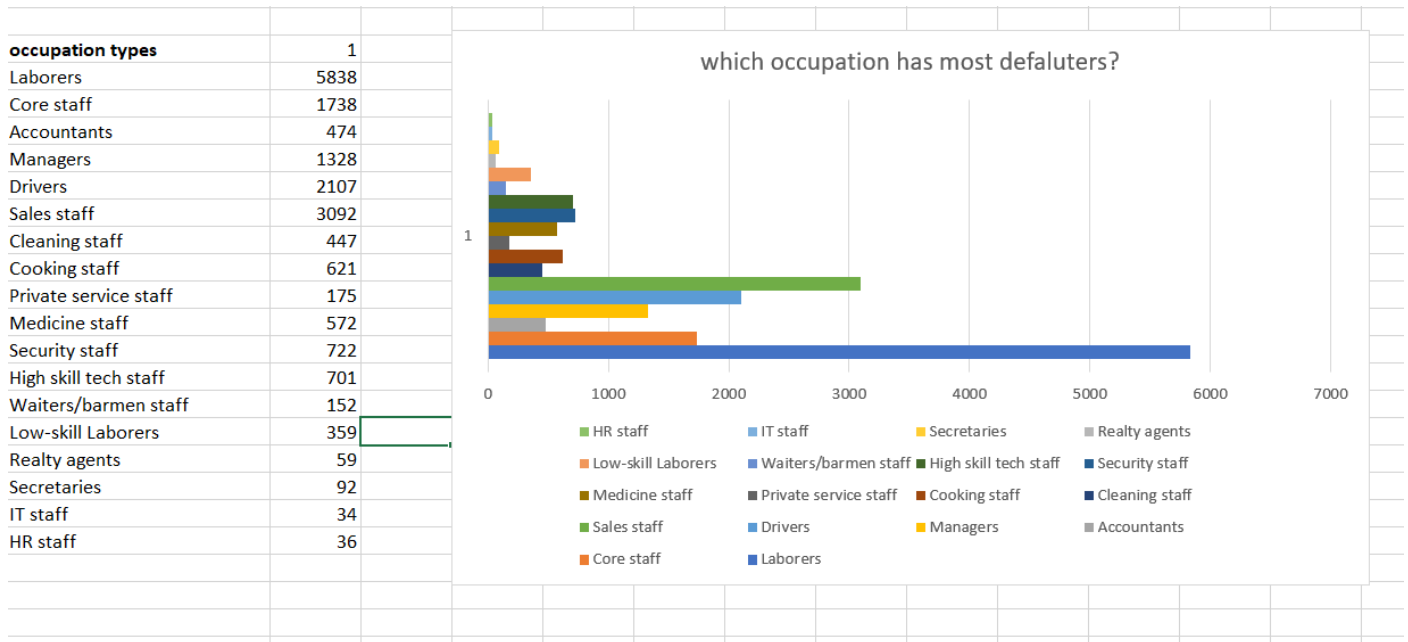
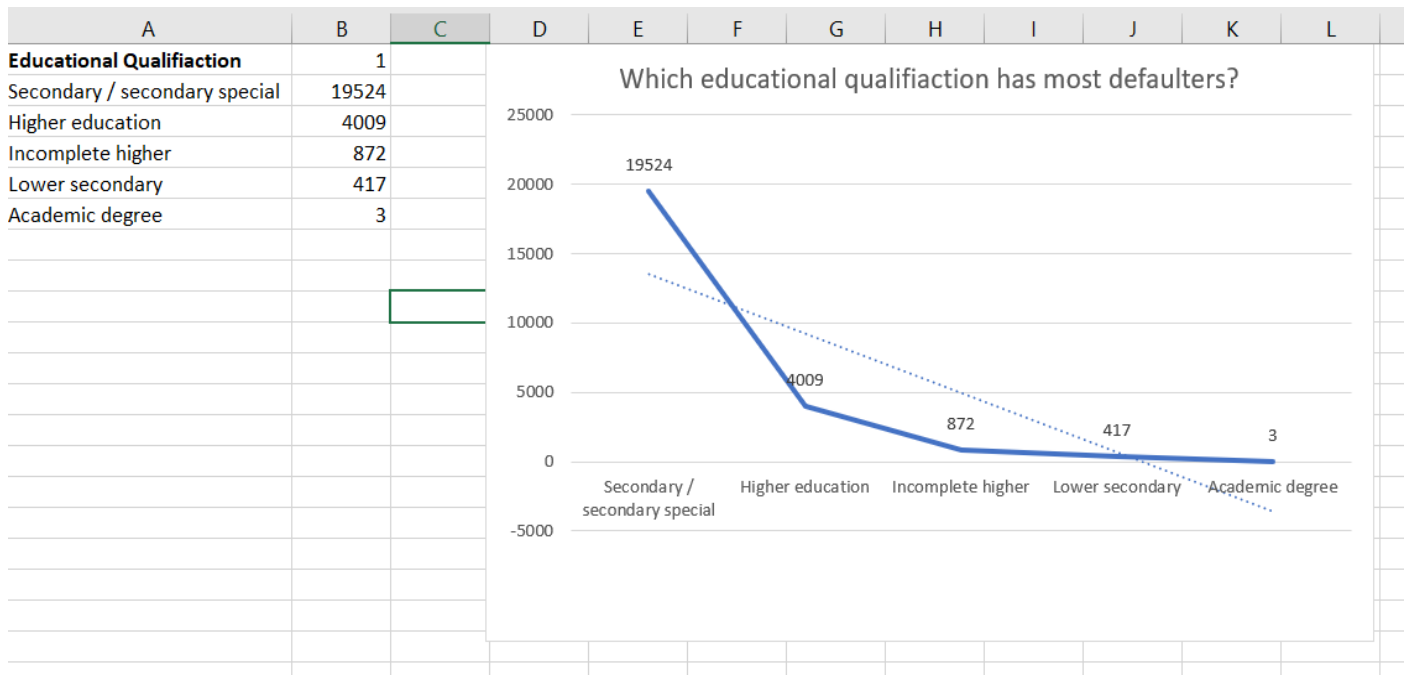
TARGET

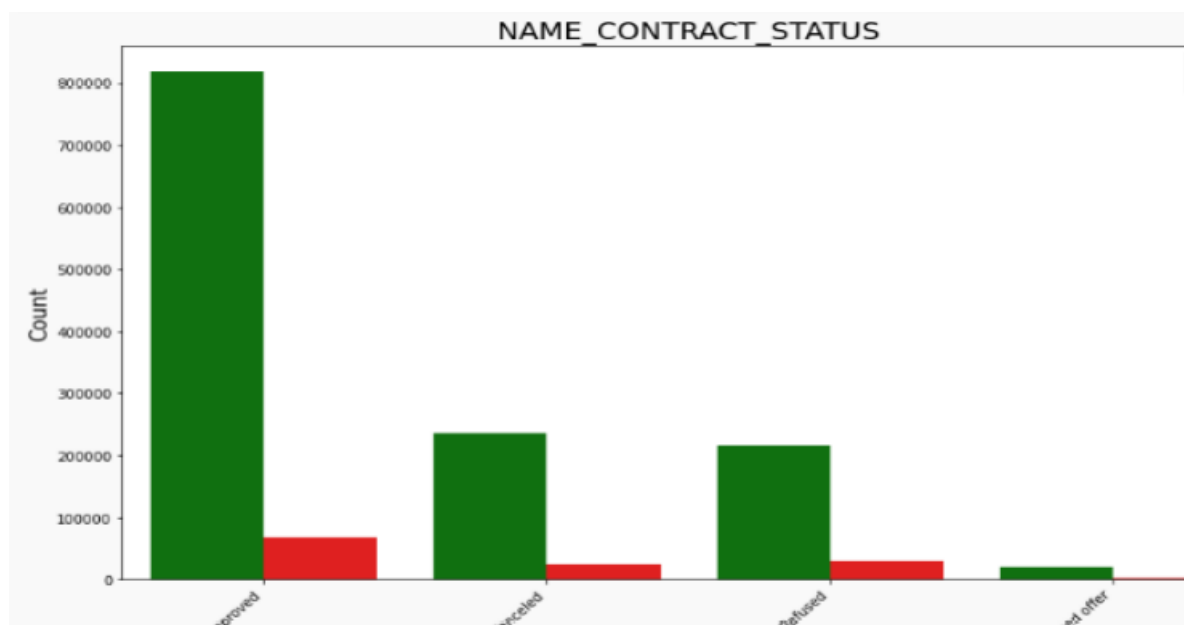
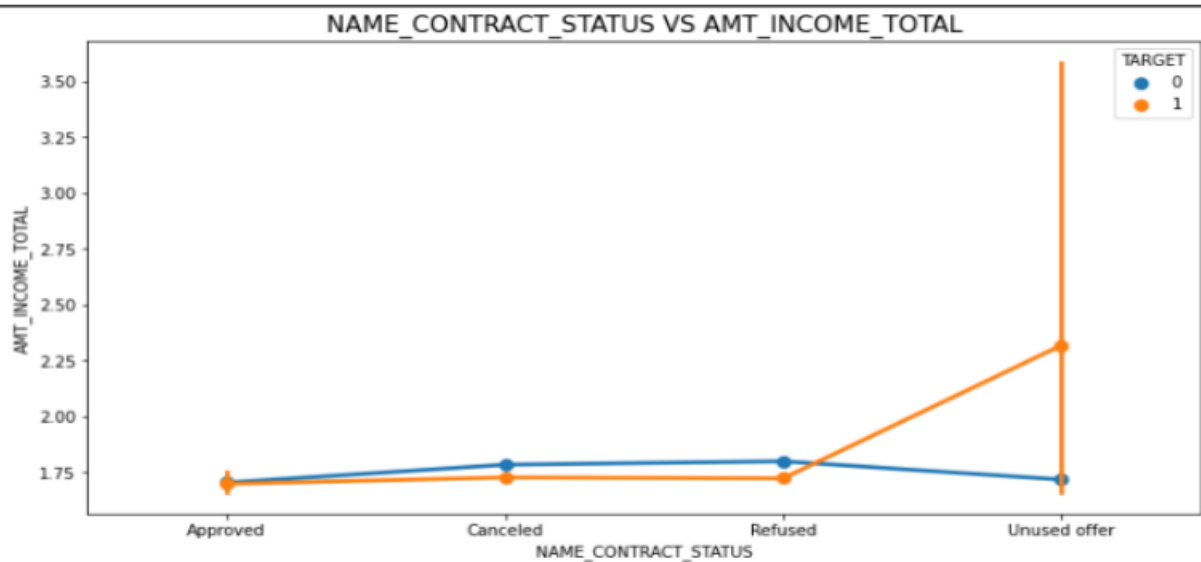
0

1



6. Find the top 10 correlations for Client with payment difficulties and all other case





7. Conclusions: After performing the analysis ,we can rectify whether a client will repay the loan or not. Also, the people who are likely to face problem in loan repayment are labourers .Also people with Secondary /secondary special education might face problem in loan repayment. Moreover, those who are living in house/apartment are facing difficulty in loan repayment (may be because of extra home loan, EMIs and so on).people opting for cash loan faces difficulty in doing the same

FACTORS DECIDING THE REPAYER

- NAME_EDUCATION_TYPE: Academic degree has less defaults.
- NAME_INCOME_TYPE: Student and Businessmen have no defaults.
- REGION_RATING_CLIENT: RATING 1 is safer.
- ORGANIZATION_TYPE: Clients with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%
- DAYS_BIRTH: People above age of 50 have low probability of defaulting
- DAYS_EMPLOYED: Clients with 40+ year experience having less than 1% default rate
- AMT_INCOME_TOTAL: Applicant with Income more than 700,000 are less likely to default
- NAME_CASH_LOAN_PURPOSE: Loans bought for Hobby, buying garage are being repaid mostly.
- CNT_CHILDREN: People with zero to two children tend to repay the loans.

FACTORS DECIDING THE DEFAULTERS

- **CODE_GENDER:** Men are at relatively higher default rate
- **NAME_FAMILY_STATUS :** People who have civil marriage or who are single default a lot.
- **NAME_EDUCATION_TYPE:** People with Lower Secondary & Secondary education
- **NAME_INCOME_TYPE:** Clients who are either at Maternity leave OR Unemployed default a lot.
- **REGION_RATING_CLIENT:** People who live in Rating 3 has highest defaults.
- **OCCUPATION_TYPE:** Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.
- **ORGANIZATION_TYPE:** Organizations with highest percent of loans not repaid are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%). Self-employed people have relative high defaulting rate, and thus should be avoided to be approved for loan or provide loan with higher interest rate to mitigate the risk of defaulting.
- **DAYS_BIRTH:** Avoid young people who are in age group of 20-40 as they have higher probability of defaulting
- **DAYS_EMPLOYED:** People who have less than 5 years of employment have high default rate.
- **CNT_CHILDREN & CNT_FAM_MEMBERS:** Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.
- **AMT_GOODS_PRICE:** When the credit amount goes beyond 3lakhs, there is an increase in defaulters

SUGGESTIONS

- As we can see that many defaulters are labourers ,therefore bank should provide special privileges to those people like charging less interests(after proper documentation).
- For the people who have repaid the after cancellation ,feedback should be taken and proper reasons should be recorded
- Clients who were refused by bank for loan earlier have now turned into a repaying client. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans.