

# Bank Loan Case Study

A bank loan case study analyzes a bank's loan portfolio to identify trends and patterns. This information can be used to improve the bank's lending practices and reduce risk.

## Identification of Missing Data and Dealing with it Appropriately:

Here we are using Excel formulas to see the missing data in the given data set which is “=COUNTBLANK(A:A)” We will be applying this to all the columns in the data set to find the number of missing cells in the column.

SK_ID	TARGET	NAME	CODE	FLAG_C	FLAG_N	CNT	AMT	AMT	AMT	AMT	NAME	NAME	NAME	NAME	NAME	REGION	DAYS	DAYS	DAYS	DAYS	OWN	FLAG
727	100837	0 Revolving F	N	Y	2	45000	135000	6750			Commercial Secondary Married	House / as	0.02461	-10072	-381	-519	-1834					
5940	106955	0 Revolving F	N	N	0	157500	450000	22500			Working Secondary Married	House / as	0.010006	-11993	-2921	-1289	-1948					
6428	107494	0 Revolving F	N	N	0	67500	202500	10125			Working Higher edu Married	House / as	0.008474	-9727	-2712	-4132	-800					
6706	107822	0 Revolving F	N	N	1	121500	180000	9000			Working Secondary Civil marri	House / as	0.011657	-11079	-899	-3765	-572					
7650	108913	0 Revolving M	N	Y	0	180000	450000	22500			Working Higher edu Single / no	House / as	0.032561	-9986	-1847	-4762	-506					
7863	109190	1 Revolving F	N	N	0	121500	270000	13500			Working Higher edu Single / no	With parer	0.006296	-12390	-640	-6365	-3597					
7998	109322	0 Revolving M	N	Y	0	112500	180000	9000			Working Secondary Single / no	House / as	0.00823	-12420	-1610	-6443	-2463					
10822	112595	0 Revolving F	N	Y	0	90000	270000	13500			Working Secondary Married	House / as	0.0228	-18193	-242	-1462	-1731					
11290	113148	0 Revolving F	N	Y	4	225000	135000	6750			State servi Secondary Single / no	House / as	0.003541	-17610	-4662	-8233	-1148					
13011	115162	0 Revolving F	N	Y	1	157500	450000	22500			State servi Higher edu Married	House / as	0.04622	-14005	-1404	-2284	-2282					
14702	117150	0 Revolving M	N	N	1	135000	270000	13500			Working Secondary Married	House / as	0.02461	-9962	-1676	-4321	-2592					
15956	118618	0 Revolving F	N	N	1	90000	270000	13500			Commercial Higher edu Married	House / as	0.008625	-10233	-2216	-4304	-2191					
18938	122080	0 Revolving M	N	Y	0	157500	180000	9000			Working Lower sec Married	House / as	0.003813	-9727	-2137	-4394	-2301					
19181	122374	0 Revolving F	N	Y	0	67500	180000	9000			Working Secondary Married	House / as	0.006008	-10777	-3330	-706	-873					
19924	123233	0 Revolving F	N	N	0	90000	180000	9000			Working Secondary Single / no	House / as	0.00496	-12920	-100	-5986	-4363					
21196	124697	0 Revolving F	N	Y	0	202500	270000	13500			Working Higher edu Married	Co-op apa	0.025164	-16944	-686	-9883	-471					
21341	124859	0 Revolving M	N	Y	0	225000	675000	33750			Working Secondary Married	House / as	0.035792	-17658	-5986	-12170	-1212					
21549	125096	0 Revolving F	N	N	1	45000	135000	6750			Working Secondary Civil marri	Municipal	0.031329	-7727	-1159	-7704	-381					
25394	129532	0 Revolving M	N	Y	1	382500	270000	13500			Commercial Higher edu Single / no	House / as	0.04622	-8993	-641	-2332	-1676					
26401	130692	0 Revolving M	N	N	0	135000	495000	24750			Pensioner Higher edu Single / no	House / as	0.020713	-22009	365243	-342	-1397					
26739	131077	0 Revolving F	N	Y	0	90000	202500	10125			Working Secondary Married	House / as	0.018634	-14069	-2646	-2821	-1141					
27006	131385	0 Revolving M	N	Y	0	180000	450000	22500			Working Secondary Married	House / as	0.022625	-16559	-496	-3284	-111					
28204	132778	0 Revolving F	N	Y	0	45000	315000	15750			Pensioner Secondary Married	House / as	0.019101	-23588	365243	-3752	-4715					
29062	133761	0 Revolving F	N	Y	2	81000	247500	12375			Working Secondary Married	House / as	0.031329	-11751	-2601	-975	-4392					
30167	135010	0 Revolving F	N	Y	1	121500	405000	20250			Commercial Higher edu Married	House / as	0.026392	-10168	-759	-4235	-2840					
30297	135159	0 Revolving M	N	Y	0	225000	450000	22500			Working Secondary Married	House / as	0.022625	-12943	-2214	-7000	-4355					
32146	137268	0 Revolving M	N	Y	0	112500	180000	9000			Working Secondary Married	House / as	0.005002	-14063	-1593	-6647	-4723					
36253	141975	0 Revolving F	N	Y	0	90000	202500	10125			Pensioner Secondary Widow	House / ac	0.011657	-20277	365243	-11951	-2809					

	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL
1	LIVE_C	ORGAN	EXT_SQ	EXT_SQ	EXT_SQ	APARTI	BASEM	YEARS	YEARS	COMM	ELEVAT	ENTRAI	FLOOR	FLOOR	LANDAI	LIVING	LIVING	NONLIV	NONLIV	APARTI	BASEM	YEARS	YEARS	COMM	ELEVAT
5	0	Government		0.55591	0.72957																				
6	0	Business Entity Type		0.65044																					
7	1	Religion		0.32274																					
8	0	Other		0.35422	0.62123																				
9	0	Business E	0.77476	0.724	0.49206																				
10	1	Other		0.71428	0.54065																				
11	0	XNA	0.58733	0.20575	0.75172																				
12	0	Electricity		0.74664																					
13	0	Medicine	0.31976	0.65186	0.36395																				
14	0	XNA	0.72204	0.55518	0.6529																				
18	0	Business E	0.11563	0.34663	0.67857																				
19	0	Government		0.23638	0.0621																				
20	0	Construction		0.68351																					
22	0	Kindergarten		0.58662	0.47765																				
24	0	Trade: typ	0.43771	0.23377	0.54245																				
30	1	Industry: type 11		0.54112	0.65941																				
31	0	Military	0.6004	0.68501	0.5245																				
36	0	Self-employed		0.32174	0.41185																				
39	0	Other	0.84276	0.6817	0.75441																				
40	1	Transport: type 1	0.80459	0.7198	0.72239																				
44	1	Industry: type 1		0.03733																					
45	0	Self-empic	0.46821	0.6742	0.39968																				
47	1	Business Entity Type		0.63152																					
48	0	Self-employed		0.53554																					
49	0	XNA		0.24766	0.47585																				
50	1	Business Entity Type		0.08924	0.72814																				
51	0	Governme	0.71043	0.58919	0.46686																				
52	1	Transport: type 2		0.52393																					
55	0	Emergency	0.24456	0.70595	0.7993	0.0227		0.9806				0.1034	0.0417				0.0114		0.0226	0.0231		0.9806			

	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS
1	APARTI	BASEM	YEARS	YEARS	COMM	ELEVAT	ENTRAI	FLOOR	FLOOR	LANDAI	LIVING	LIVING	NONLIV	NONLIV	FONDKI	HOUSE	TOTALA	WALLSI	EMERG	OBS_3C	DEF_3C	OBS_6C	DEF_6C	DAYS_L	FLAG_C
5																				0	0	0	0	-815	
6																				2	0	2	0	-617	
7																				0	0	0	0	-1106	
8																				0	0	0	0	-2536	
9																				1	0	1	0	-1562	
10																				2	0	2	0	-1070	
11																				1	0	1	0	0	
12																				2	0	2	0	-1673	
13																				0	0	0	0	-844	
14																				0	0	0	0	-2396	
18																				0	0	0	0	-925	
19																				0	0	0	0	-3	
20																				4	0	4	0	-2811	
22																				0	0	0	0	-1850	
24																				0	0	0	0	0	
30																				2	0	2	0	-2	
31																				0	0	0	0	-849	
36																				2	0	2	0	-697	
39																				0	0	0	0	-2411	
40																				0	0	0	0	-1541	
44																				4	0	4	0	-458	
45																				1	0	1	0	-1480	
47																				1	0	1	0	-556	
48																				1	1	1	1	-462	
49																				0	0	0	0	-649	
50																				4	0	4	0	-1247	
51																				0	0	0	0	-784	
52																				1	1	1	1	-1498	
55	0.0229		0.9806				0.1034	0.0417				0.0116	0.023		block of fli	0.0139	Stone, brk No		4	0	4	0	-466		

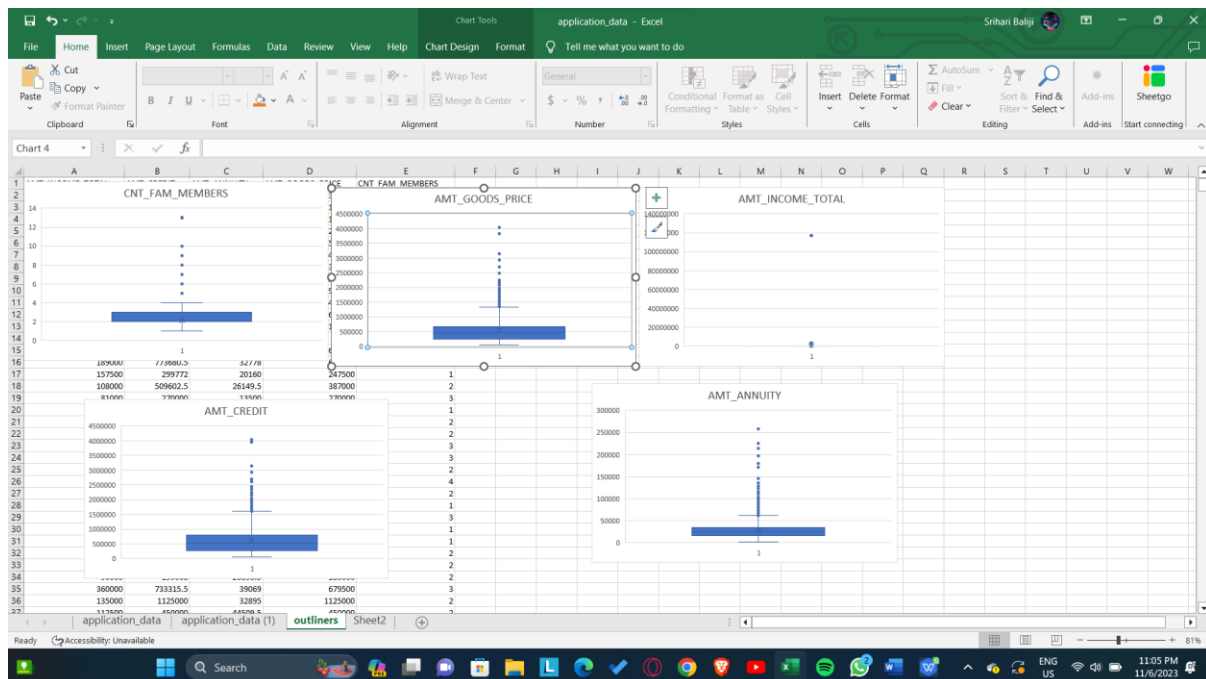
From the above images, we can say that there are moves and more missing cells in the data from

APARTMENTS\_AVG to EMERGENCYSTATE\_MODE

More than 35% of the cells in this columns are empty

## Identify Outliers in the Dataset:

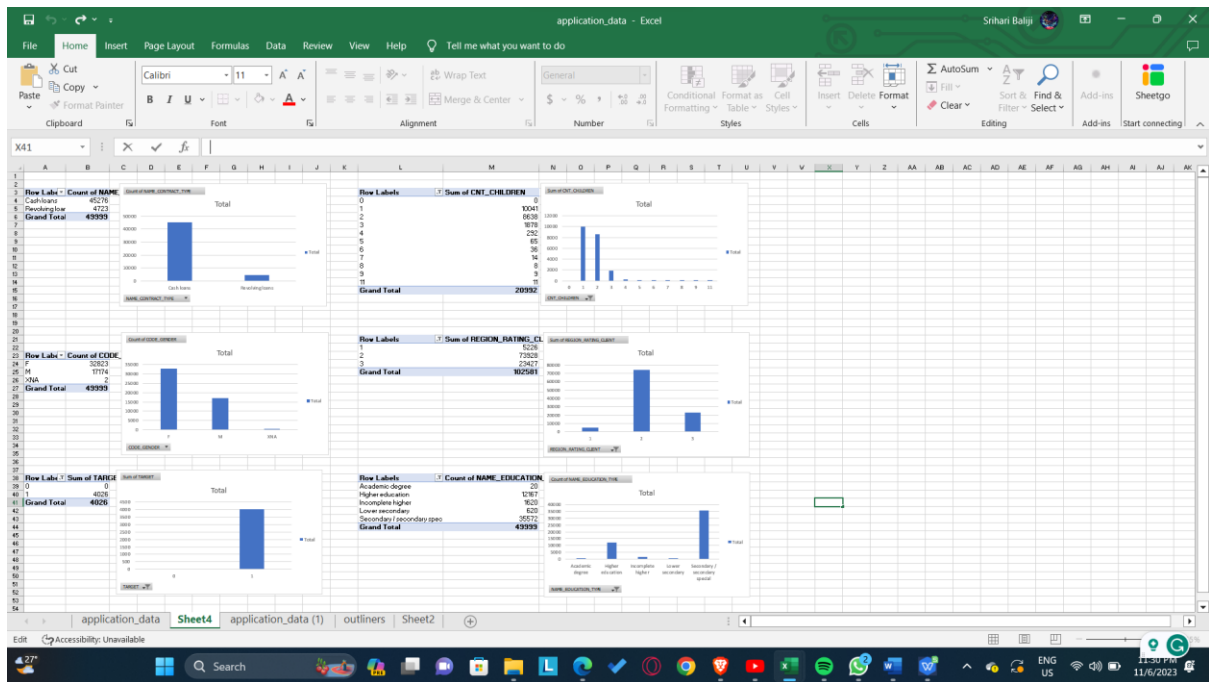
Here Outliers can significantly impact the analysis and distort the results. You need to identify outliers in the loan application dataset. By using the created plot box for identification of the outliers



## Analyzing of Data Imbalance:

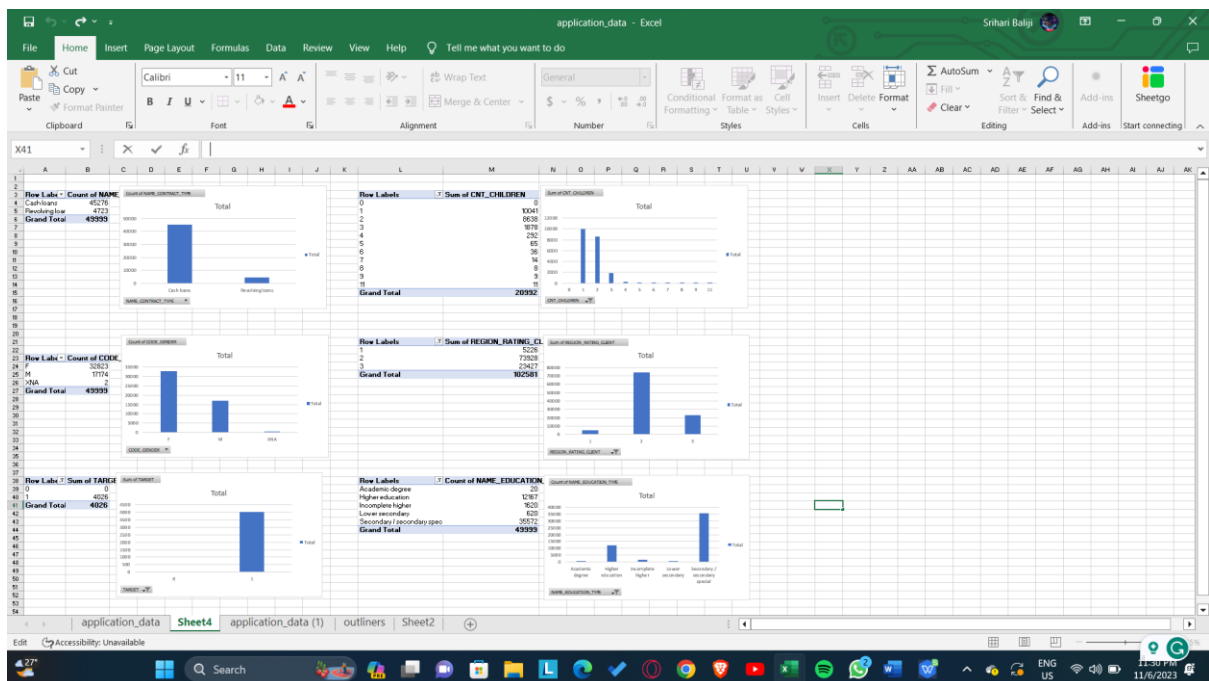
Data imbalance can affect the accuracy of the analysis, especially for binary classification problems. Understanding the data distribution is crucial for building reliable models.

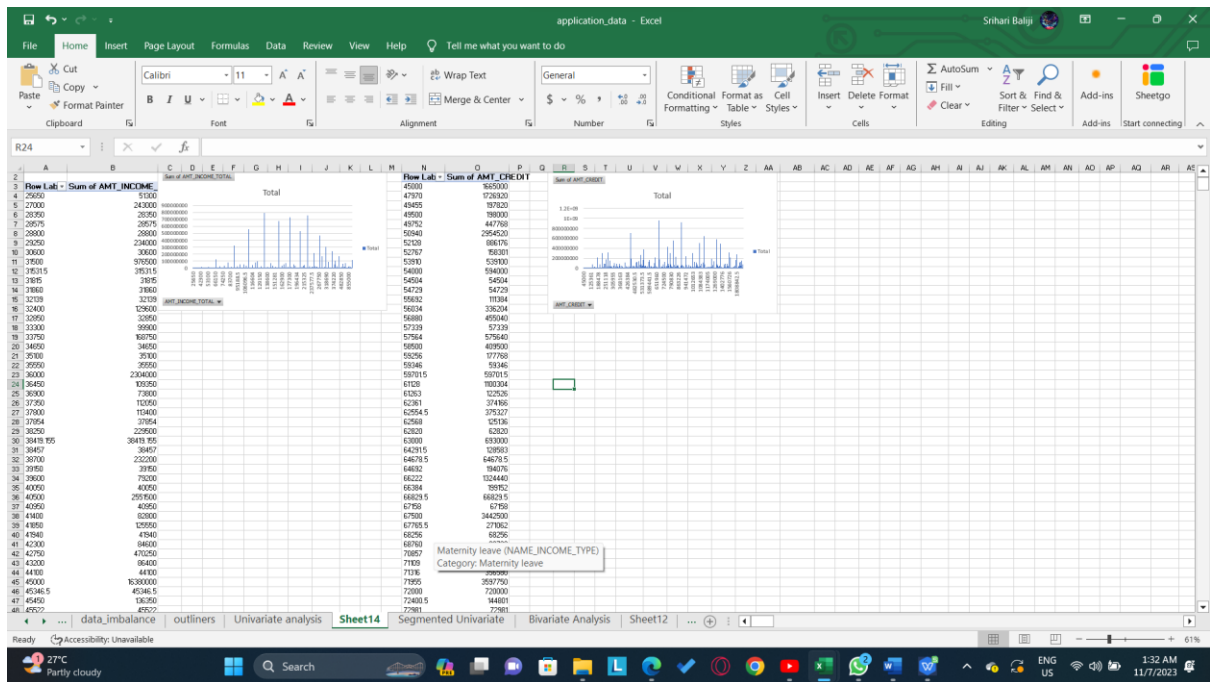
Here we use a Pivot table to show the imbalance of data



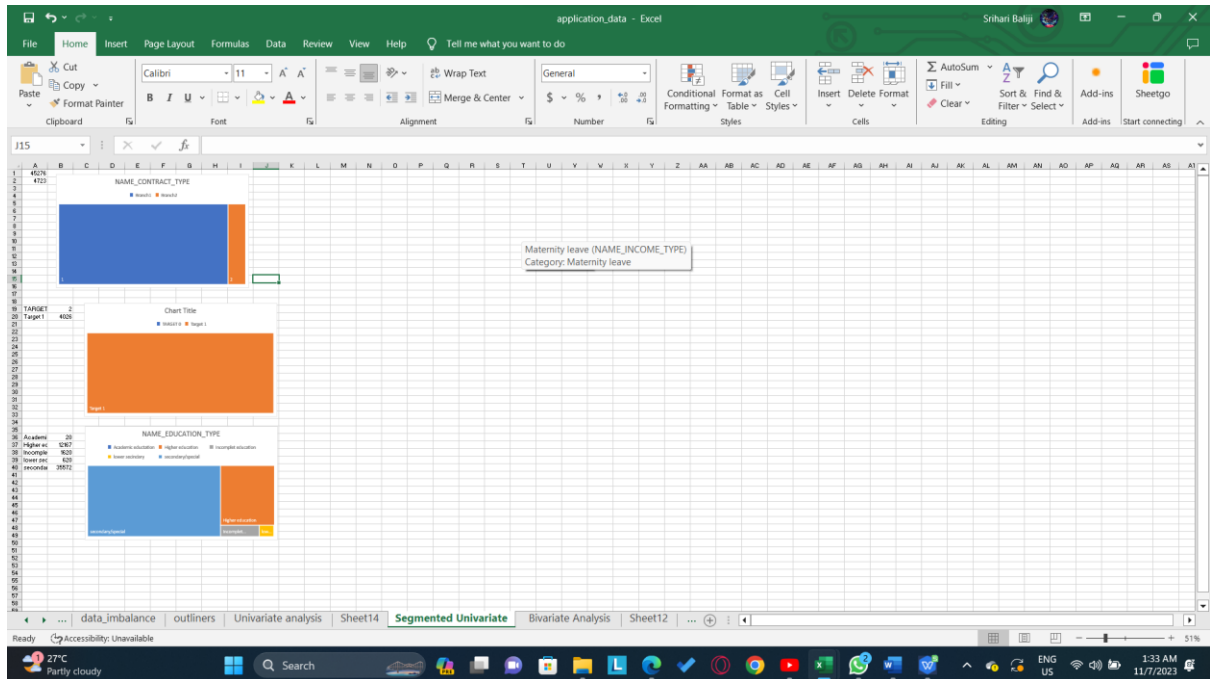
**Perform Univariate, Segmented Univariate, and Bivariate Analysis:** To gain insights into the driving factors of loan default, it is important to conduct various analyses on consumer and loan attributes.

## Univariate:

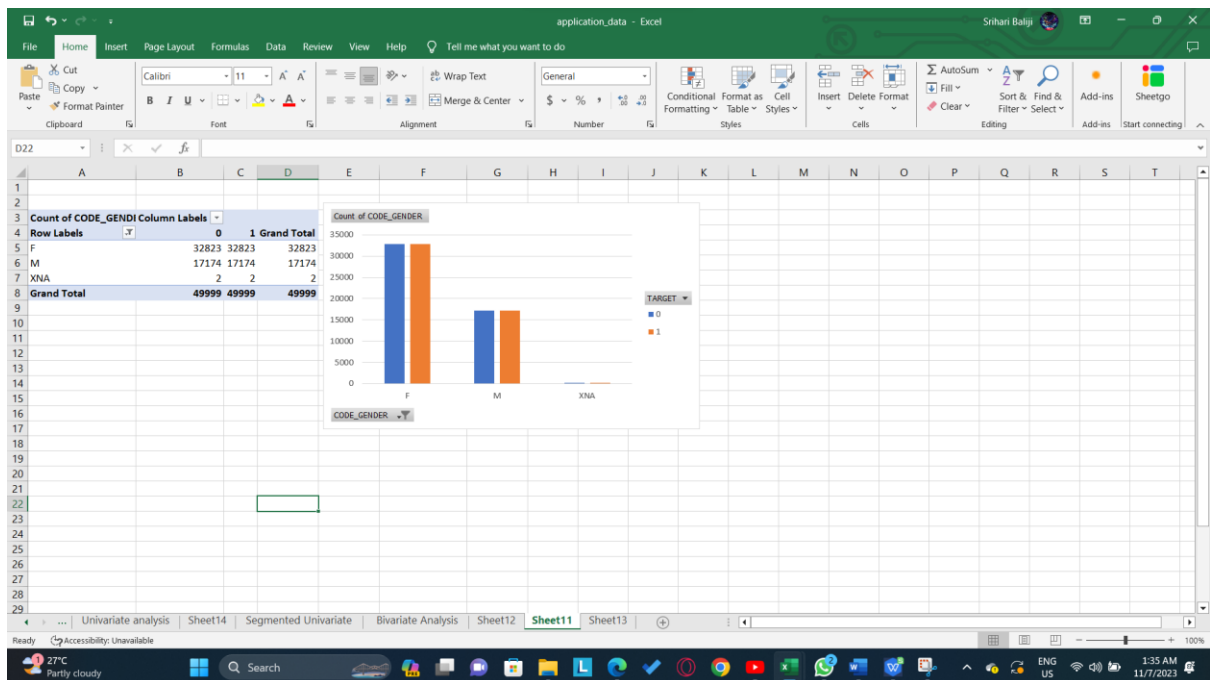
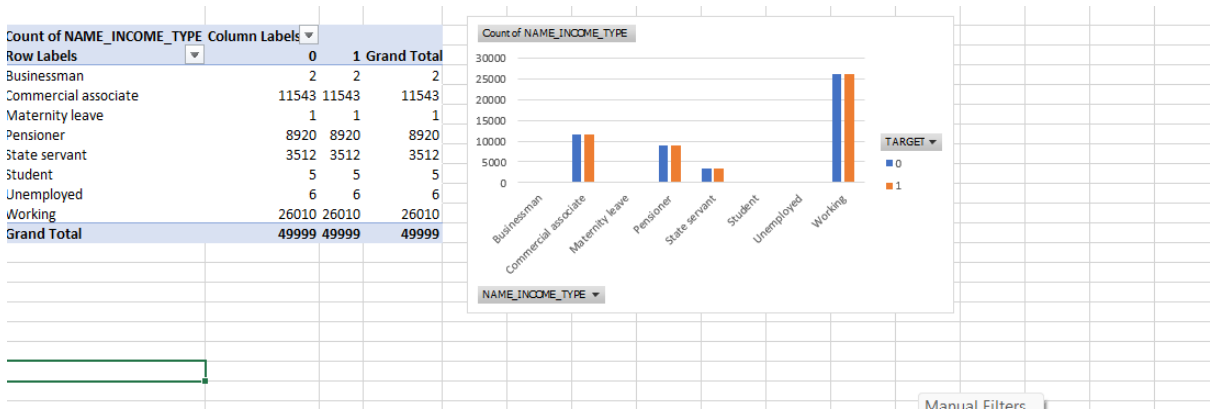
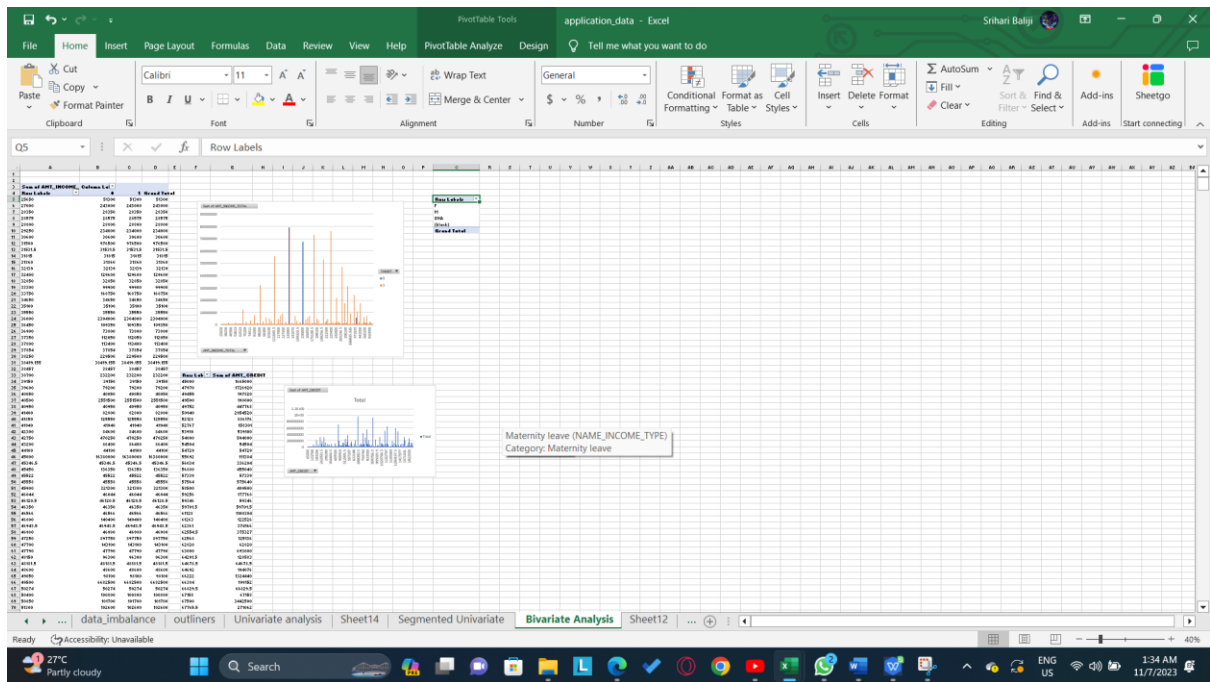


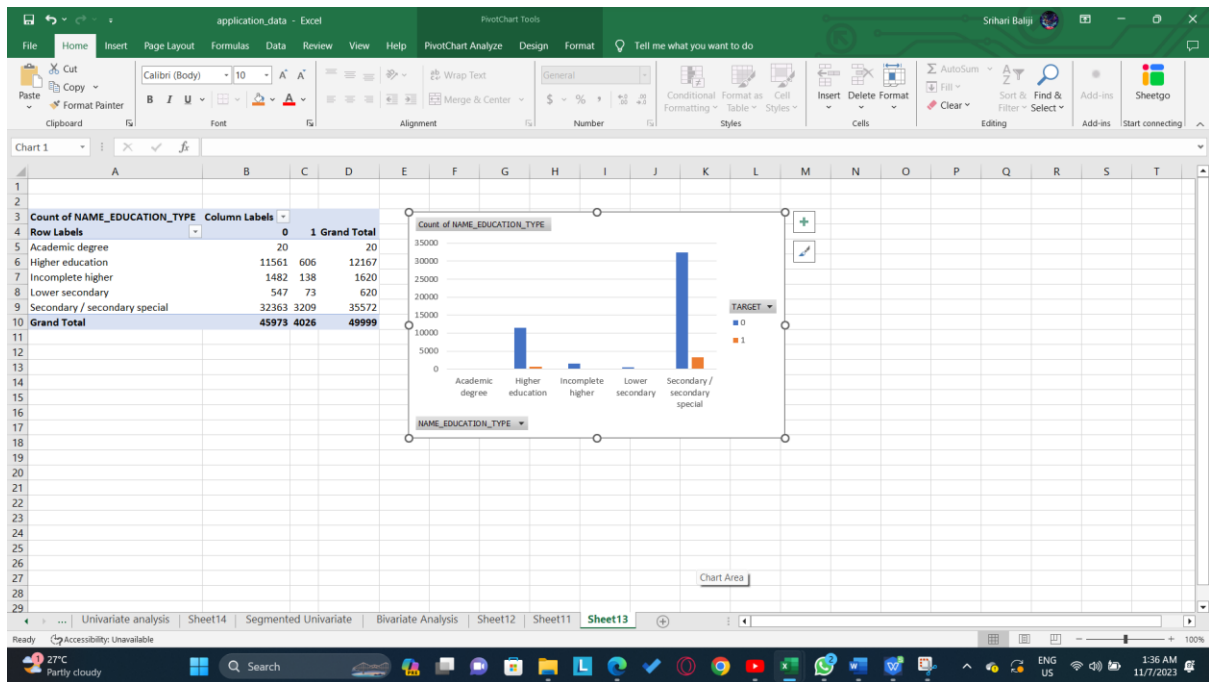


## Segmented Univariate:



## Bivariate Analysis:





**Identify Top Correlations for Different Scenarios:** Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.

	CNT_CHI LDREN	AMT_INCOM E_TOTAL	AMT_CR EDIT	AMT_AN NUITY	AMT_GOOD S_PRICE	REGION_POPULATI ON_RELATIVE	Age	Years Employe d	CNT_FAM_M EMBERS	REGION_RATI NG_CLIENT	EXT_SOU RCE_2	EXT_SOU RCE_3
CNT_CHILDREN	1	0.028	0.003	0.021	-0.001	-0.024	-0.337	-0.041	0.879	0.023	-0.015	-0.037
AMT_INCOME_TOTAL	0.028	1	0.343	0.419	0.349	0.168	-0.063	0.035	0.034	-0.187	0.140	-0.059
AMT_CREDIT	0.003	0.343	1	0.771	0.987	0.100	0.047	0.084	0.065	-0.103	0.129	0.033
AMT_ANNUITY	0.021	0.419	0.771	1	0.777	0.121	-0.013	0.053	0.076	-0.132	0.127	0.025
AMT_GOODS_PRICE	-0.001	0.349	0.987	0.777	1	0.104	0.045	0.085	0.063	-0.104	0.136	0.036
REGION_POPULATION_RELATIVE	-0.024	0.168	0.100	0.121	0.104	1	0.025	-0.009	-0.023	-0.539	0.198	-0.011
Age	-0.337	-0.063	0.047	-0.013	0.045	0.025	1	0.226	-0.286	-0.002	0.078	0.175
Years Employed	-0.041	0.035	0.084	0.053	0.085	-0.009	0.226	1	-0.010	0.016	0.074	0.097
CNT_FAMILY_MEMBERS	0.879	0.034	0.065	0.076	0.063	-0.023	-0.286	-0.010	1	0.028	-0.001	-0.023
REGION_RATING_CLIENT	0.023	-0.187	-0.103	-0.132	-0.104	-0.539	-0.002	0.016	0.028	1	-0.291	-0.004
EXT_SOURCE_2	-0.015	0.140	0.129	0.127	0.136	0.198	0.078	0.074	-0.001	-0.291	1	0.076
EXT_SOURCE_3	-0.037	-0.059	0.033	0.025	0.036	-0.011	0.175	0.097	-0.023	-0.004	0.076	1



	CNT_CHI LDREN	AMT_INCOM E_TOTAL	AMT_CR EDIT	AMT_AN NUITY	AMT_GOOD S_PRICE	REGION_POPULATI ON_RELATIVE	Age	Years Employe d	CNT_FAM_M EMBERS	REGION_RATI NG_CLIENT	EXT_SOU RCE_2	EXT_SOU RCE_3
CNT_CHILDREN	1	0.005	-0.002	0.031	-0.008	-0.032	-0.259	-0.044	0.886	0.041	-0.012	-0.024
AMT_INCOME_TOTAL	0.005	1	0.038	0.046	0.038	0.009	-0.003	-0.001	0.007	-0.021	0.007	-0.018
AMT_CREDIT	-0.002	0.0381	1	0.752	0.983	0.069	0.135	0.098	0.051	-0.059	0.121	0.052
AMT_ANNUITY	0.031	0.0464	0.752	1	0.753	0.072	0.014	0.041	0.076	-0.074	0.116	0.032
AMT_GOODS_PRICE	-0.008	0.0376	0.983	0.753	1	0.076	0.136	0.104	0.047	-0.066	0.131	0.053
REGION_POPULATION_RELATIVE	-0.032	0.0091	0.069	0.072	0.076	1	0.048	0.016	-0.030	-0.443	0.170	-0.010
Age	-0.259	-0.0031	0.135	0.014	0.136	0.048	1	0.281	-0.203	-0.034	0.108	0.134
Years Employed	-0.044	-0.0010	0.098	0.041	0.104	0.016	0.281	1	-0.010	-0.005	0.089	0.056
CNT_FAM_MEMBERS	0.886	0.0067	0.051	0.076	0.047	-0.030	-0.203	-0.010	1	0.044	0.002	-0.029
REGION_RATING_CLIENT	0.041	-0.0215	-0.059	-0.074	-0.066	-0.443	-0.034	-0.005	0.044	1	-0.250	0.014
EXT_SOURCE_2	-0.012	0.0071	0.121	0.116	0.131	0.170	0.108	0.089	0.002	-0.250	1	0.049
EXT_SOURCE_3	-0.024	-0.0182	0.052	0.032	0.053	-0.010	0.134	0.056	-0.029	0.014	0.049	1

The correlation is calculated between two columns using the formula `correl(array1,array2)` the highest is marked using green markings