

# BANK LOAN CASE STUDY



# Agenda

- Project Description
- Approach
- Tech-Stack Used
- Insights
- Result



# PROJECT DESCRIPTION

- This project analyzes bank loan application data to uncover valuable insights. The success of a loan application—measured by whether it's approved and repaid—is vital for financial institutions. By examining factors like customer attributes, loan amount, credit history, and repayment behavior, we can gather meaningful information to improve the decision-making process for loan approvals.
- The data provided is in an Excel sheet, and the first step is to clean and prepare it for analysis.



# TECH STACK USED

- **Software: Microsoft Excel**

A spreadsheet application by Microsoft used for data analysis, visualization, and automation with formulas, charts, and VBA.

- **Operating System: Windows**

A widely used OS by Microsoft, known for its user-friendly interface, multitasking, and software compatibility.



# APPROACH

- To analyze any data, we should first clean the data like empty blanks, duplicates , null values , design ,Outlier Detection etc.
- So I started cleaning the given data
- I followed the below steps for cleaning the data
  - 1)Firstly I have converted the raw data into table format.

The team has given (122) columns and 50,000 records in Excel to work.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	SK_ID_CUR	TARGET	NAME_CC	CODE_GE	FLAG_OW	FLAG_OW	CNT_CHIL	AMT_INCC	AMT_CREI	AMT_ANNA	AMT_GOC	NAME_TY	NAME_IN	NAME_ED	NAME_FA	NAME_HC	REGION_F	DAYS_BIR	DAYS_EMI	DAYS_REG	DAYS_ID_
2	100002	1	Cash loan	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccomp	Working	Secondary	Single / n	House / a	0.018801	-9461	-637	-3648	-2120
3	100003	0	Cash loan	F	N	N	0	270000	1293503	35698.5	1129500	Family	State serv	Higher ed	Married	House / a	0.003541	-16765	-1188	-1186	-291
4	100004	0	Revolving	M	Y	Y	0	67500	135000	6750	135000	Unaccomp	Working	Secondary	Single / n	House / a	0.010032	-19046	-225	-4260	-2531
5	100006	0	Cash loan	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccomp	Working	Secondary	Civil marri	House / a	0.008019	-19005	-3039	-9833	-2437
6	100007	0	Cash loan	M	N	Y	0	121500	513000	21865.5	513000	Unaccomp	Working	Secondary	Single / n	House / a	0.028663	-19932	-3038	-4311	-3458
7	100008	0	Cash loan	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, p	State serv	Secondary	Married	House / a	0.035792	-16941	-1588	-4970	-477
8	100009	0	Cash loan	F	Y	Y	1	171000	1560726	41301	1395000	Unaccomp	Commerci	Higher ed	Married	House / a	0.035792	-13778	-3130	-1213	-619
9	100010	0	Cash loan	M	Y	Y	0	360000	1530000	42075	1530000	Unaccomp	State serv	Higher ed	Married	House / a	0.003122	-18850	-449	-4597	-2379
10	100011	0	Cash loan	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary	Married	House / a	0.018634	-20099	365243	-7427	-3514
11	100012	0	Revolving	M	N	Y	0	135000	405000	20250	405000	Unaccomp	Working	Secondary	Single / n	House / a	0.019689	-14469	-2019	-14437	-3992
12	100014	0	Cash loan	F	N	Y	1	112500	652500	21177	652500	Unaccomp	Working	Higher ed	Married	House / a	0.0228	-10197	-679	-4427	-738
13	100015	0	Cash loan	F	N	Y	0	38419.16	148365	10678.5	135000	Children	Pensioner	Secondary	Married	House / a	0.015221	-20417	365243	-5246	-2512
14	100016	0	Cash loan	F	N	Y	0	67500	80865	5881.5	67500	Unaccomp	Working	Secondary	Married	House / a	0.031329	-13439	-2717	-311	-3227
15	100017	0	Cash loan	M	Y	N	1	225000	918468	28966.5	697500	Unaccomp	Working	Secondary	Married	House / a	0.016612	-14086	-3028	-643	-4911
16	100018	0	Cash loan	F	N	Y	0	189000	773680.5	32778	679500	Unaccomp	Working	Secondary	Married	House / a	0.010006	-14583	-203	-615	-2056
17	100019	0	Cash loan	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary	Single / n	Rented ap	0.020713	-8728	-1157	-3494	-1368
18	100020	0	Cash loan	M	N	N	0	108000	509602.5	26149.5	387000	Unaccomp	Working	Secondary	Married	House / a	0.018634	-12931	-1317	-6392	-3866
19	100021	0	Revolving	F	N	Y	1	81000	270000	13500	270000	Unaccomp	Working	Secondary	Married	House / a	0.010966	-9776	-191	-4143	-2427
20	100022	0	Revolving	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary	Widow	House / a	0.04622	-17718	-7804	-8751	-1259
21	100023	0	Cash loan	F	N	Y	1	90000	544491	17563.5	454500	Unaccomp	State serv	Higher ed	Single / n	House / a	0.015221	-11348	-2038	-1021	-3964



**A. Identify Missing Data and Deal with it Appropriately:** Task: Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.

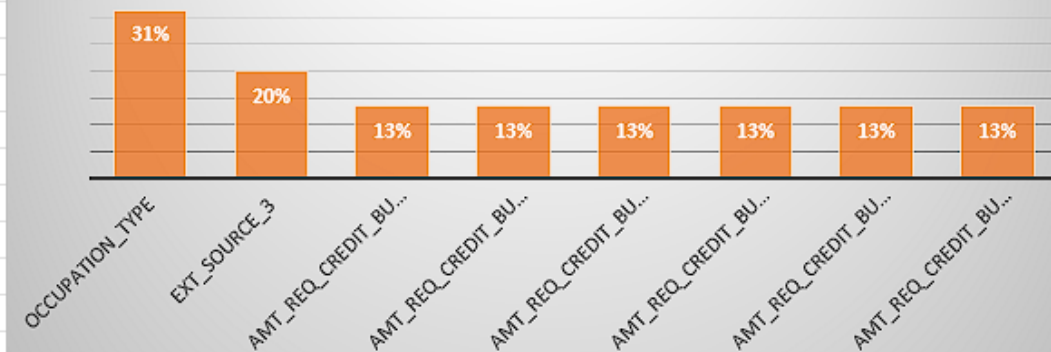
[illegible]



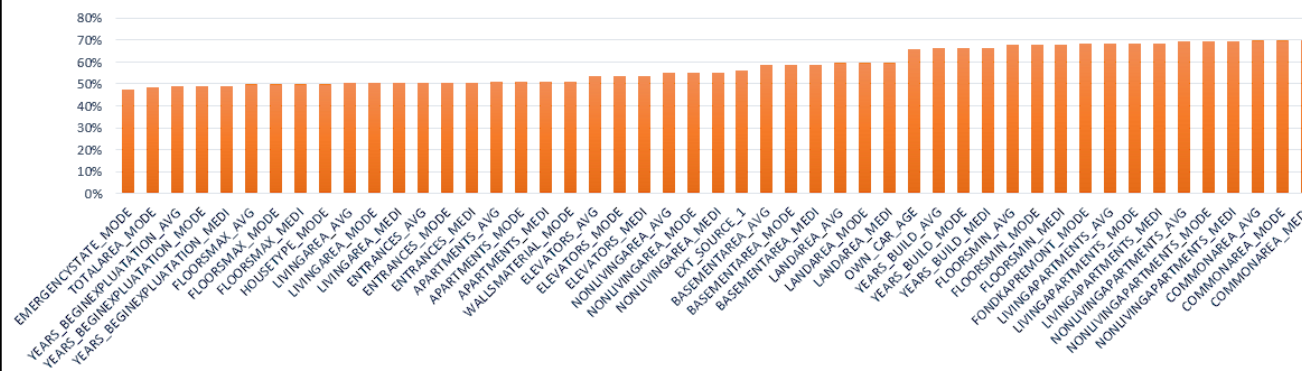
# Identify Missing Data and Deal with it Appropriately

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Columns	Blanks	Blank %		Columns with >35% Blank values	49 col					Columns with < 35% blank values	8 columns	
2	SK_ID_CURR	0	0%		Column Name	Blank %					Column Name	Blank %	
3	TARGET	0	0%		EMERGENCYSTATE_MODE	47%					OCCUPATION_TYPE	31%	
4	NAME_CONTRACT_TYPE	0	0%		TOTALAREA_MODE	48%					EXT_SOURCE_3	20%	
5	CODE_GENDER	0	0%		YEARS_BEGINEXPLUATATION_AVG	49%					AMT_REQ_CREDIT_BUREAU_HOUR	13%	
6	FLAG_OWN_CAR	0	0%		YEARS_BEGINEXPLUATATION_MODE	49%					AMT_REQ_CREDIT_BUREAU_DAY	13%	
7	FLAG_OWN_REALTY	0	0%		YEARS_BEGINEXPLUATATION_MEDI	49%					AMT_REQ_CREDIT_BUREAU_WEEK	13%	
8	CNT_CHILDREN	0	0%		FLOORSMAX_AVG	50%					AMT_REQ_CREDIT_BUREAU_MON	13%	
9	AMT_INCOME_TOTAL	0	0%		FLOORSMAX_MODE	50%					AMT_REQ_CREDIT_BUREAU_QRT	13%	
10	AMT_CREDIT	0	0%		FLOORSMAX_MEDI	50%					AMT_REQ_CREDIT_BUREAU_YEAR	13%	
11	AMT_ANNUITY	1	0%		HOUSETYPE_MODE	50%							
12	AMT_GOODS_PRICE	38	0%		LIVINGAREA_AVG	50%							
13	NAME_TYPE_SUITE	192	0%		LIVINGAREA_MODE	50%							
14	NAME_INCOME_TYPE	0	0%		LIVINGAREA_MEDI	50%							
15	NAME_EDUCATION_TYPE	0	0%		ENTRANCES_AVG	50%							
16	NAME_FAMILY_STATUS	0	0%		ENTRANCES_MODE	50%							
17	NAME_HOUSING_TYPE	0	0%		ENTRANCES_MEDI	50%							
18	REGION_POPULATION_RELATIVE	0	0%		APARTMENTS_AVG	51%							
19	DAYS_BIRTH	0	0%		APARTMENTS_MODE	51%							
20	DAYS_EMPLOYED	0	0%		APARTMENTS_MEDI	51%							
21	DAYS_REGISTRATION	0	0%		WALLSMATERIAL_MODE	51%							
22	DAYS_ID_PUBLISH	0	0%		ELEVATORS_AVG	53%							
23	EMERGENCYSTATE_MODE	23698	47%		ELEVATORS_MODE	53%							
24	FLAG_MOBIL	0	0%		ELEVATORS_MEDI	53%							
25	FLAG_EMP_PHONE	0	0%		NONLIVINGAREA_AVG	55%							
26	FLAG_WORK_PHONE	0	0%		NONLIVINGAREA_MODE	55%							
27	FLAG_CONT_MOBILE	0	0%		NONLIVINGAREA_MEDI	55%							

Columns with < 35% blank values ==> 8 columns



Columns with >35% Blank values ----> 49 Columns



## Identify Missing Data and Deal with it Appropriately

0.588593689	0.519097338	0	0	0	0	-738
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	DAYS_LAST_PHONE_CHANGE
0.262948593	0.13937578	2	2	2	2	-1134
0.622245775	0.519097338	1	0	1	0	-828
0.555912083	0.729566691	0	0	0	0	-815
0.65044169	0.519097338	2	0	2	0	-617
0.322738287	0.519097338	0	0	0	0	-1106
0.354224732	0.621226338	0	0	0	0	-2536
0.723999852	0.492060094	1	0	1	0	-1562
0.714279286	0.54065445	2	0	2	0	-1070
0.205747288	0.751723715	1	0	1	0	0
0.746643629	0.519097338	2	0	2	0	-1673
0.651862333	0.363945239	0	0	0	0	-844
0.555183162	0.652896552	0	0	0	0	-2396
0.715041819	0.176652579	0	0	0	0	-2370
0.566906613	0.77008707	0	0	0	0	-4
0.642656205	0.519097338	0	0	0	0	-188
0.346633981	0.678567689	0	0	0	0	-925
0.23637784	0.062103038	0	0	0	0	-3
0.683513346	0.519097338	4	0	4	0	-2811
0.706428403	0.556727426	8	0	8	0	-239
0.58661714	0.477649155	0	0	0	0	-1850
0.113374513	0.519097338	0	0	0	0	-296
0.233766958	0.542445144	0	0	0	0	0
0.457142972	0.358951229	0	0	0	0	-468
0.624304737	0.669056695	0	0	0	0	-795
0.786179309	0.565607981	1	1	1	0	-4
0.651405637	0.461482391	0	0	0	0	0
0.54847716	0.190705948	10	1	10	0	-161

NAME_TYPE_SUITE	Row Labels	Count of NAME_TYPE_SUITE
Unaccompanied	Children	542
Family	Family	6549
Unaccompanied	Group of people	36
Unaccompanied	Other_A	137
Unaccompanied	Other_B	259
Spouse, partner	Spouse, partner	1849
Unaccompanied	Unaccompanied	40627
Unaccompanied	Grand Total	49999
Children		
Unaccompanied	"Unaccompanied" is counted most of the time hence Mode of this column is "Unaccompanied"	
Unaccompanied		
Children		
Unaccompanied	We can also use "UNIQUE" function	
Unaccompanied		
Unaccompanied		
Family		
Unaccompanied		

Columns marked in Yellow had missing values,I filled them as per below mentioned approach-

1)NAME\_TYPE\_SUITE column's missing value filled by Its Mode which is- Unaccompanied

2)I Filled OCCUPATION\_TYPE column's missing value by Unknown

3)Numeric columns missing value cell filled by their

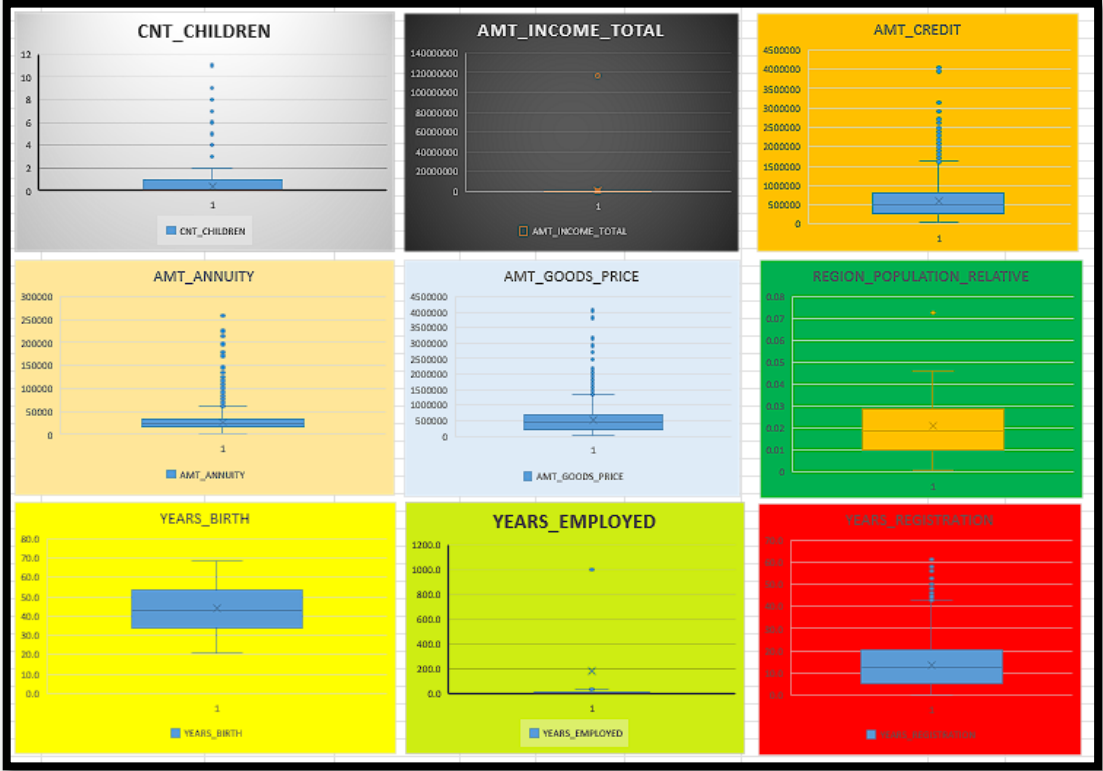




**B. Identify Outliers in the Dataset:** Outliers can significantly impact the analysis and distort the results. You need to identify outliers in the loan application dataset.

ANS:

REGION_POPULATION_RELATIVE	YEARS_BIRTH	YEARS_EMPLOYED	YEARS_REGISTRATION	YEARS_ID_PUBLISH	FLAG_MOBIL
0.000533	52.37	2.15	27.3	7.14	1
0.000533	51.05	1.43	9.3	5.97	1
0.000533	37.54	7.90	17.1	11.88	1
0.000533	28.18	2.61	10.6	8.07	1
0.000533	65.55	1000.67	6.3	5.54	1
0.000533	59.15	1000.67	31.4	12.03	1
0.000533	64.72	1000.67	9.5	11.28	1
0.000938	33.33	8.74	6.8	9.96	1
0.000938	42.50	1000.67	7.4	13.06	1
0.000938	64.35	1000.67	35.5	7.52	1
0.000938	50.18	11.26	13.8	5.13	1
0.000938	32.93	9.40	11.5	12.10	1
0.000938	50.36	25.11	11.7	5.23	1
0.000938	62.35	1000.67	26.9	11.81	1
0.000938	57.19	17.16	9.9	10.99	1
0.001276	23.82	5.33	13.2	3.75	1
0.001276	43.49	21.81	27.1	14.07	1
0.001276	49.71	0.30	0.0	3.36	1
0.001276	38.06	3.12	5.7	2.90	1
0.001276	26.96	5.03	1.0	0.41	1
0.001276	34.31	11.78	3.8	10.85	1
0.001276	42.30	4.55	6.1	11.59	1
0.001276	26.70	1.05	11.5	6.65	1
0.001276	22.63	0.56	10.4	2.43	1
0.001276	27.48	3.66	12.9	1.65	1
0.001276	49.94	3.53	1.1	4.70	1



Columns	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	YEARS_BIRTH	YEARS_EMPLOYED	YEARS_REGISTRATION
Q1	0	112500	270000	16456.5	238500	0.010006	33.9123288	2.556164384	5.473972603
Q3	1	202500	808650	34596	679500	0.028663	53.8191781	15.66575342	20.44931507
IQR	1	90000	538650	18139.5	441000	0.018657	19.9068493	13.10958904	14.97534247
Upper Limit	2.5	337500	1616625	61805.25	1341000	0.0566485	83.6794521	35.33013699	42.91232877
Lower Limit	-1.5	-22500	-537975	-10752.75	-423000	-0.0179795	4.05205479	-17.10821918	-16.9890411
IQR = Q3-Q1									
Upper Limit= Q3+(1.5*IQR)									
Lower Limit=Q1-(1.5*IQR)									

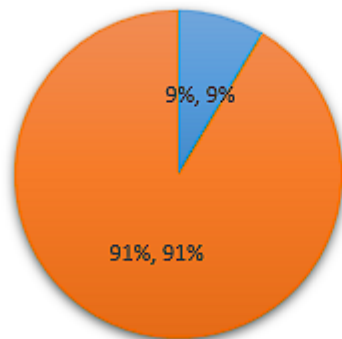


## C. Analyze Data Imbalance:

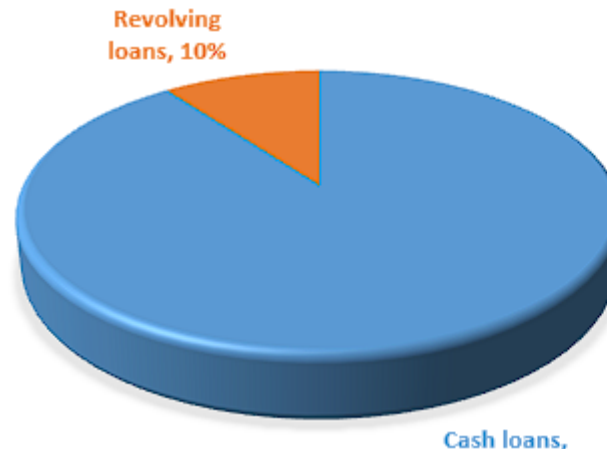
Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.

Target	Count of TARGET		Loan Type	Count of Loan Type(Number of Clients)		Loan Type	% Client	
0	37549		Cash loans	36892		Cash loans	90%	<-- H2/H4
1	3520		Revolving loans	4177		Revolving loans	10%	<-- H3/H4
Grand Total	41069		Grand Total	41069				
Target	% of Target		Max	37549				
Defaulters	9%	<--E3/E4	Min	3520				
Non-Defaulters	91%	<--E2/E4	Ratio of Data Imbalance(Min/Max)	0.09				

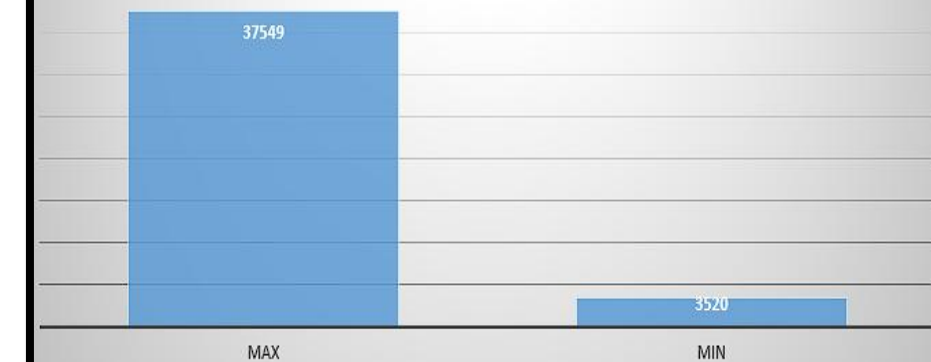
% of Defaulters Vs Non-Defaulters



% LOAN TYPE

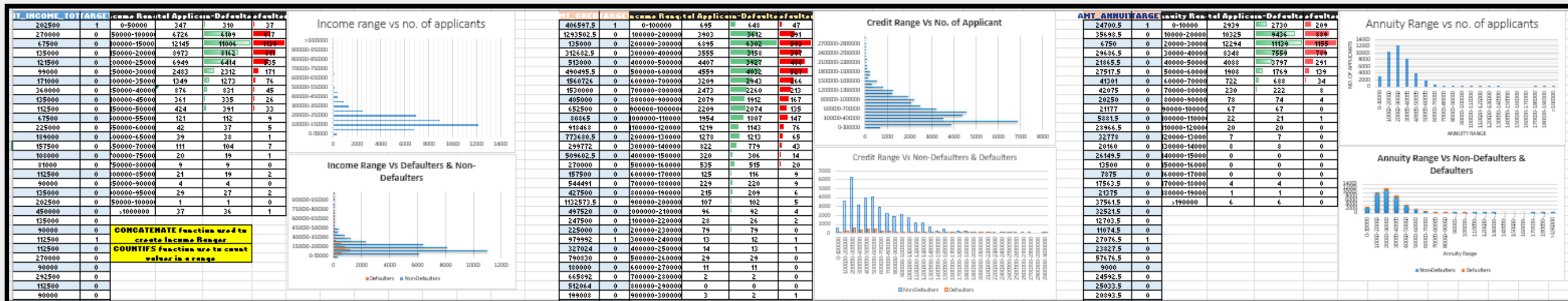


Non-Defaulters Vs Defaulters



# D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:

Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.

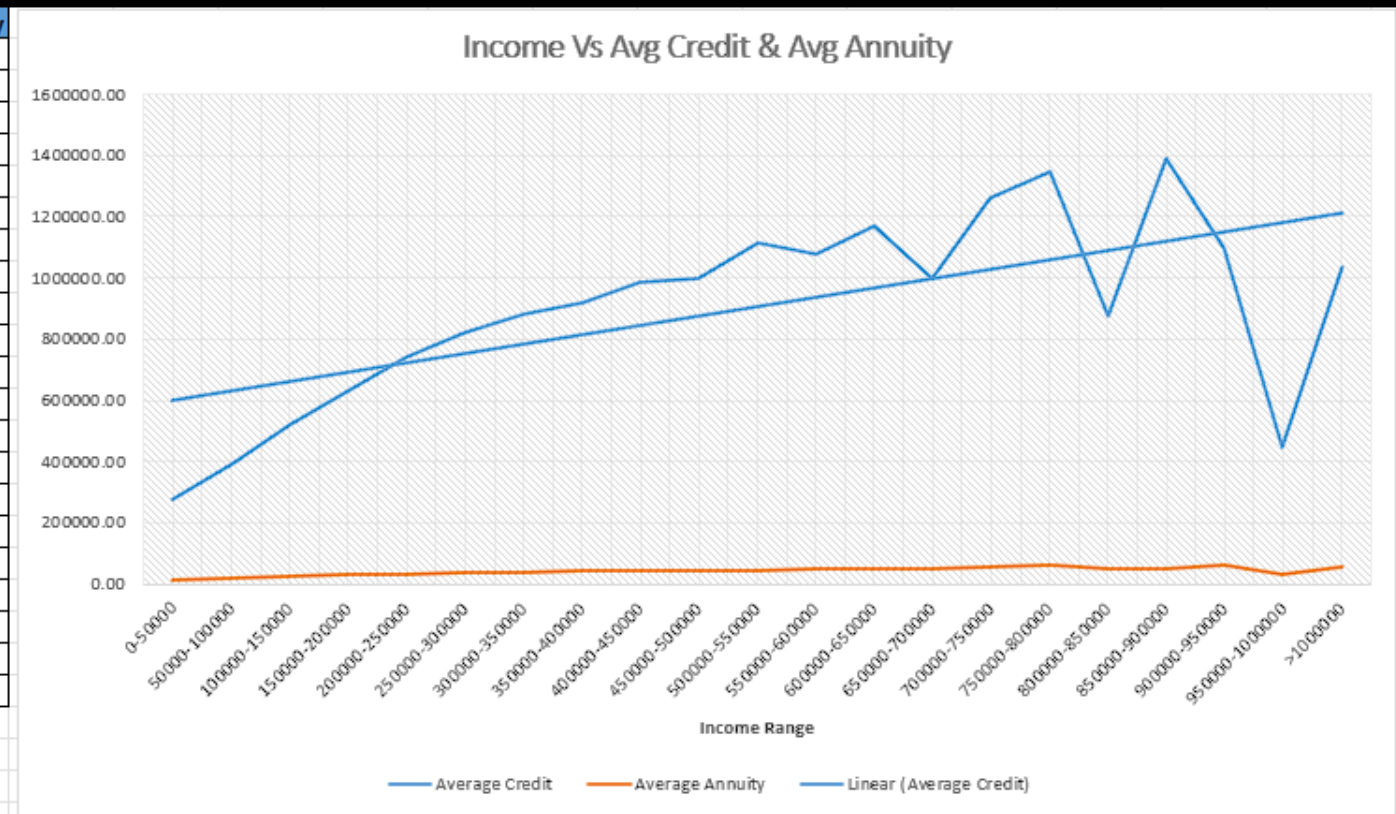


[illegible][illegible]



## BIVARIATE ANALYSIS:

AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	Income Range	Average Credit	Average Annuity
80397	315000	17100	0-50000	277298.05	13640.21
112500	270000	13500	50000-100000	393349.85	18704.74
90000	180000	9000	100000-150000	519709.47	24009.44
144000	630000	17325	150000-200000	630878.77	28654.87
202500	180000	9000	200000-250000	740833.61	33007.89
180000	180000	9000	250000-300000	821826.37	36100.31
97875	225000	9787.5	300000-350000	884090.21	39301.63
67500	202500	10125	350000-400000	920791.10	40810.46
67500	135000	6750	400000-450000	985704.88	44097.51
99000	180000	9000	450000-500000	997944.62	44784.76
157500	706500	28773	500000-550000	1112433.21	45984.46
67500	180000	9000	550000-600000	1074844.07	46788.96
76500	157500	7875	600000-650000	1171325.88	49857.69
135000	180000	9000	650000-700000	1000031.84	47379.41
243000	180000	9000	700000-750000	1259983.35	53482.28
180000	773680.5	31639.5	750000-800000	1344940.00	58803.00
157500	180000	9000	800000-850000	876760.07	47799.86
112500	180000	9000	850000-900000	1388400.75	47631.38
112500	180000	9000	900000-950000	1093675.66	58976.22
90000	180000	9000	950000-1000000	450000.00	30073.50
112500	180000	9000	>1000000	1037054.80	54840.04
90000	180000	9000			
67500	180000	9000			
135000	180000	9000			
112500	180000	9000			
90000	180000	9000			





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

•**Task:** Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.

FOR NON\_DEFAULTERS:

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	YEARS_BIRTH	YEARS_EMPLOYED
AMT_INCOME_TOTAL	1.000						
AMT_CREDIT	0.360	1.000					
AMT_ANNUITY	0.431	0.761	1.000				
AMT_GOODS_PRICE	0.368	0.986	0.765	1.000			
REGION_POPULATION_RELATIVE	0.189	0.097	0.116	0.100	1.000		
YEARS_BIRTH	0.050	0.161	0.099	0.155	0.049	1.000	
YEARS_EMPLOYED	0.036	0.095	0.062	0.096	-0.006	0.352	1.000
YEARS_REGISTRATION	-0.034	0.024	0.001	0.020	0.065	0.305	0.176
YEARS_ID_PUBLISH	0.023	0.044	0.032	0.044	0.004	0.108	0.083
FLAG_MOBIL	0.003	0.004	0.001	0.004	0.004	0.009	0.005
FLAG_EMP_PHONE	0.002	-0.003	-0.001	-0.002	0.004	-0.008	0.000
FLAG_WORK_PHONE	-0.081	-0.034	-0.053	-0.010	-0.019	-0.041	0.014
FLAG_CONT_MOBILE	-0.016	0.027	0.026	0.025	-0.006	0.002	0.001
FLAG_PHONE	0.012	0.023	0.008	0.040	0.092	0.039	0.057
FLAG_EMAIL	0.078	0.005	0.058	0.005	0.041	-0.067	-0.031
CNT_FAM_MEMBERS	-0.001	0.036	0.043	0.034	-0.031	-0.180	-0.036
REGION_RATING_CLIENT	-0.207	-0.106	-0.130	-0.108	0.045	-0.050	0.015
REGION_RATING_CLIENT_W_CITY	-0.223	-0.115	-0.144	-0.116	-0.541	-0.049	0.012

## FOR DEFAULTERS:

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	YEARS_BIRTH	YEARS_EMPLOYED	YEARS_REGISTRATION	YEARS_ID_PUBLISH	FLAG_MOBIL	FLAG_EMP_PHONE
AMT_INCOME_TOTAL	1.000										
AMT_CREDIT	0.312	1.000									
AMT_ANNUITY	0.371	0.745	1.000								
AMT_GOODS_PRICE	0.314	0.982	0.746	1.000							
REGION_POPULATION_RELATIVE	0.097	0.056	0.066	0.061	1.000						
YEARS_BIRTH	0.088	0.194	0.086	0.188	0.013	1.000					
YEARS_EMPLOYED	0.023	0.105	0.054	0.113	-0.002	0.306	1.000				
YEARS_REGISTRATION	-0.003	0.043	-0.013	0.041	0.047	0.240	0.151	1.000			
YEARS_ID_PUBLISH	0.038	0.054	0.050	0.059	0.008	0.125	0.099	0.044	1.000		
FLAG_MOBIL	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	
FLAG_EMP_PHONE	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
FLAG_WORK_PHONE	-0.099	-0.054	-0.056	-0.026	-0.027	-0.060	0.011	-0.022	0.012	#DIV/0!	#DIV/0!
FLAG_CONT_MOBILE	-0.055	0.030	0.040	0.027	0.002	0.026	0.009	0.001	0.006	#DIV/0!	#DIV/0!
FLAG_PHONE	0.008	0.032	0.007	0.046	0.062	0.035	0.071	0.058	0.039	#DIV/0!	#DIV/0!
FLAG_EMAIL	0.080	-0.006	0.098	-0.005	0.052	-0.059	-0.032	0.017	-0.018	#DIV/0!	#DIV/0!
CNT_FAM_MEMBERS	-0.029	0.049	0.047	0.044	-0.017	-0.107	0.011	-0.132	0.111	#DIV/0!	#DIV/0!
REGION_RATING_CLIENT	-0.160	-0.048	-0.065	-0.054	-0.437	-0.051	-0.004	-0.127	-0.028	#DIV/0!	#DIV/0!
REGION_RATING_CLIENT_W_CITY	-0.169	-0.059	-0.085	-0.062	-0.439	-0.047	0.000	-0.116	-0.020	#DIV/0!	#DIV/0!
HOOR_APPR_PROCESS_START	0.060	0.048	0.039	0.060	0.158	-0.037	0.016	0.065	0.003	#DIV/0!	#DIV/0!
REG_REGION_NOT_LIVE_REGION	0.054	0.006	0.029	0.007	0.000	-0.025	-0.040	-0.001	-0.019	#DIV/0!	#DIV/0!
REG_REGION_NOT_WORK_REGION	0.104	0.027	0.063	0.029	0.021	-0.032	-0.071	0.000	-0.023	#DIV/0!	#DIV/0!
LIVE_REGION_NOT_WORK_REGION	0.106	0.039	0.073	0.040	0.064	-0.015	-0.050	0.000	-0.014	#DIV/0!	#DIV/0!
REG_CITY_NOT_LIVE_CITY	-0.016	-0.051	-0.023	-0.051	-0.035	-0.126	-0.097	-0.033	-0.046	#DIV/0!	#DIV/0!
REG_CITY_NOT_WORK_CITY	-0.033	-0.039	-0.019	-0.044	-0.046	-0.108	-0.131	-0.063	-0.028	#DIV/0!	#DIV/0!
LIVE_CITY_NOT_WORK_CITY	-0.018	-0.004	-0.003	-0.010	-0.026	-0.034	-0.070	-0.037	0.010	#DIV/0!	#DIV/0!
EXT_SOURCE_2	0.117	0.113	0.103	0.127	0.157	0.164	0.101	0.085	0.048	#DIV/0!	#DIV/0!
EXT_SOURCE_3	-0.059	0.048	0.022	0.050	-0.028	0.117	0.076	0.030	0.057	#DIV/0!	#DIV/0!
OBS_30_CNT_SOCIAL_CIRCLE	-0.020	0.031	0.013	0.029	-0.012	0.006	0.045	-0.010	0.027	#DIV/0!	#DIV/0!
DEF_30_CNT_SOCIAL_CIRCLE	-0.046	-0.037	-0.035	-0.031	0.019	0.001	-0.003	-0.001	0.025	#DIV/0!	#DIV/0!

## TOP CORRELATIONS :

Top 5 Correlation (Non-Defaulters)				Top 5 Correlation (Defaulters)		
Variable 1	Variable 2	Correlation		Variable 1	Variable 2	Correlation
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998		OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998
AMT_GOODS_PRICE	AMT_CREDIT	0.986		AMT_GOODS_PRICE	AMT_CREDIT	0.982
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.948		REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.951
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861		DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.891
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.853		LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.806



## INSIGHTS

- **Missing Data:** Identify gaps, decide whether to fill them (e.g., with averages) or leave them out.
- **Outliers:** Spot extreme values, check if they're errors or meaningful insights.
- **Data Imbalance:** Check class distribution (e.g., defaults vs. non-defaults), calculate proportions, and address imbalance if needed.
- **Analysis:**
  - **Univariate:** See individual variable distributions.
  - **Segmented:** Compare groups (e.g., defaults vs. non-defaults).
  - **Bivariate:** Find relationships between variables.
- **Correlations:** Spot variables most tied to loan defaults, segment insights for targeted analysis.



# RESULT

In this project, I analyzed loan application data, which included customer attributes, loan amounts, credit history, repayment statuses, and other financial variables. I explored how each factor affects loan approvals, defaults, and the assessment of financial risks. The project provided valuable insights into the relationships between various factors, contributing to a better understanding of loan repayment patterns and effective risk management strategies



## REFERENCE LINK:

**The link for excel sheet file**



[Bank\\_loan\\_case\\_study](#)





