PROJECT 6

BANK LOAN ANALYSIS

PROJECT DESCRIPTION

- Embark on a journey to revolutionize bank loan analysis. Discover innovative methodologies, cutting-edge technologies, and data-driven insights that will enhance the efficiency, accuracy, and compliance of the analysis process. Join us in shaping the future of financial evaluation and unlocking new possibilities. WELCOME TO OUR PRESENTATION ON **“BANK LOAN ANALYSIS” !**

WHY BANK LOAN ANALYSIS MATTERS?

- Securing the right financing empowers businesses to invest in expansion, technology, and talent.
- A well-analyzed loan application increases the likelihood of approval and favorable terms.
- Understanding the financial implications ensures that the chosen loan aligns with the company's long-term goals.

CASH LOANS VS REVOLVING LOANS

CASH LOANS

- Provide a fixed amount of money
- Require regular fixed payments
- Typically have a fixed interest rate
- One-time borrowing
- Commonly used for large purchases

REVOLVING LOANS

- Provide a credit limit
- Allow flexible payments
- Interest charged on outstanding balance
- Can be used repeatedly
- Commonly used for ongoing expenses

APPROACH

- **Data Collection** -Gather relevant financial statements, loan applications, credit reports, and other supporting documents.
- **Risk Assessment** -Analyze the borrower's credit history, income stability, cash flow, and collateral to determine the risk involved in granting the loan.
- **Ratio Analysis-** Compute and interpret key financial ratios,such as the debt-to-income ratio and loan-to-value ratio,to assess the borrower's financial health.
- **Decision Making -** Based on the analysis, make informed decisions regarding loan approval, interest rates, and loan terms.

TECH STACK USED

Microsoft Excel

- Versatile tool for collecting and organising data.
- Used for data analysis including sorting, filtering and statistical calculations.
- Used for creating visualisations

Microsoft powerpoint

Finalized report is visualized in the form of presentation.

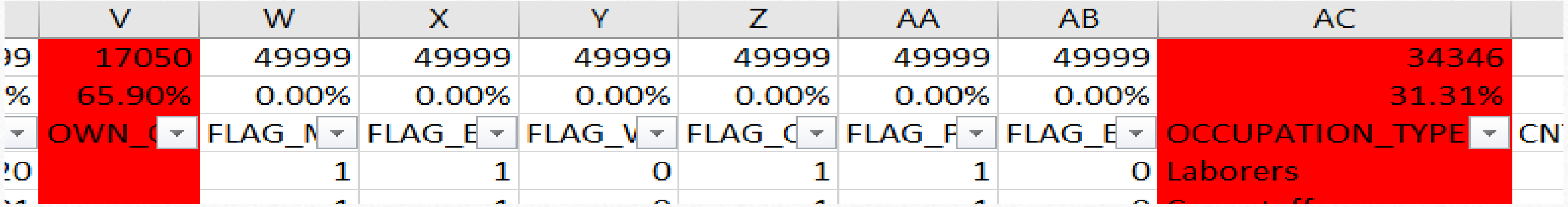
EXCEL SHEET LINK - https://drive.google.com/file/d/1tYrIfIGPXGi8QkPhMIIzJic2OXv_vRAr/view?usp=gmail

A. HANDLING DATA

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.

❖ Deleting columns with more than 30% missing values.

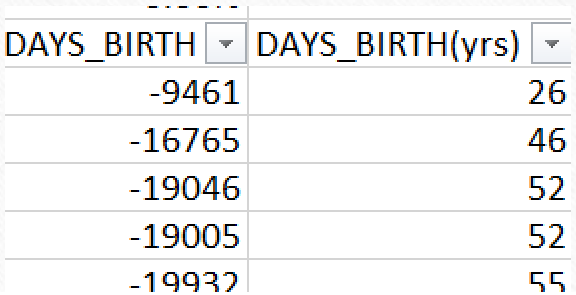
- Finding the total values in the column by using =COUNTA(A3:A50001)
- Then for finding the no. of missing values , using =1-B1/A1 , Where B1 = total no. of values that are present in the column , A1 = total no. of rows including blanks
- Dragging this formula to all the formulas
- Then converting this number to percentage



	V	W	X	Y	Z	AA	AB	AC	
99	17050	49999	49999	49999	49999	49999	49999	34346	
%	65.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	31.31%	
	OWN_(	FLAG_M	FLAG_E	FLAG_V	FLAG_C	FLAG_F	FLAG_E	OCCUPATION_TYPE	CN
20		1	1	0	1	1	0	Laborers	

❖ Converting days to years for better understanding and visualising the data in the columns(days_birth , days_employed ,days_id_publish, days_registration).

- Formula used - =ABS(R4/365)
- Dragging the formula to all the values in the column
- Then converting them to number and removing the decimals

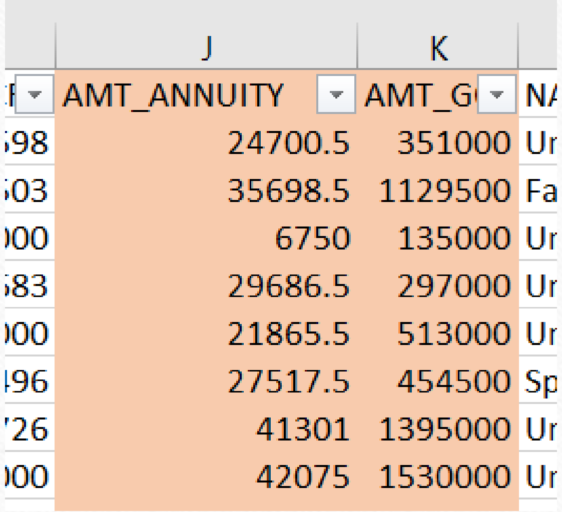


DAYS_BIRTH	DAYS_BIRTH(yrs)
-9461	26
-16765	46
-19046	52
-19005	52
-19932	55

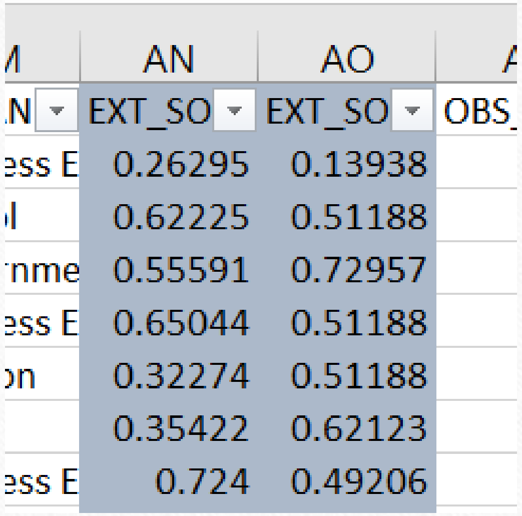
A. HANDLING DATA

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.

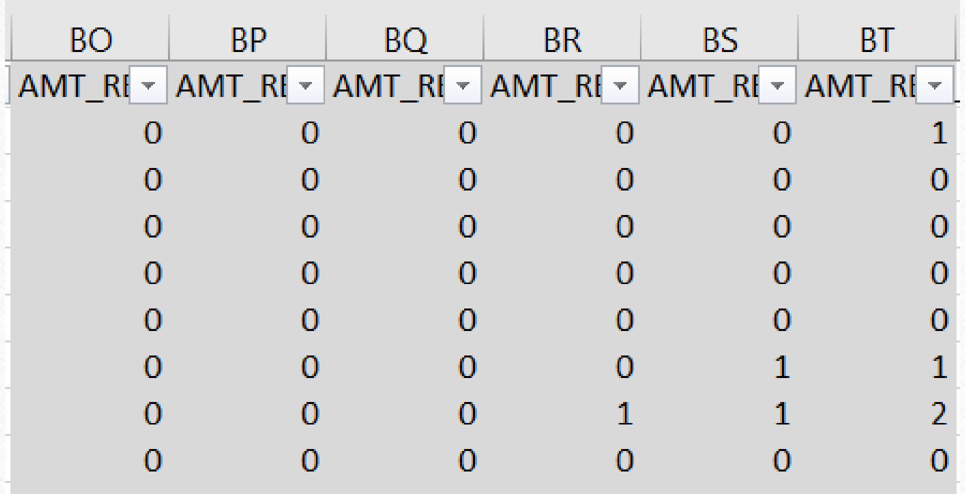
- ❖ Using the mean , median , mode imputation accordingly – filling missing values
- Mean imputation on columns (EXT_SOURCE_2,EXT_SOURCE_3)
- Median imputation on columns (AMT_ANNUITY,AMT_GOODS_PRICE)
- Mode imputation on columns (AMT_REQ_CREDIT_BUREAU_HOUR, AMT_REQ_CREDIT_BUREAU_DAY, AMT_REQ_CREDIT_BUREAU_WEEK , AMT_REQ_CREDIT_BUREAU_MON, AMT_REQ_CREDIT_BUREAU_QRT, AMT_REQ_CREDIT_BUREAU_YEAR)



	J	K	
	AMT_ANNUITY	AMT_G	NA
98	24700.5	351000	Ur
03	35698.5	1129500	Fa
00	6750	135000	Ur
83	29686.5	297000	Ur
00	21865.5	513000	Ur
96	27517.5	454500	Sp
26	41301	1395000	Ur
00	42075	1530000	Ur



M	AN	AO	A
N	EXT_SO	EXT_SO	OBS_
ess E	0.26295	0.13938	
l	0.62225	0.51188	
nme	0.55591	0.72957	
ess E	0.65044	0.51188	
on	0.32274	0.51188	
	0.35422	0.62123	
ess E	0.724	0.49206	

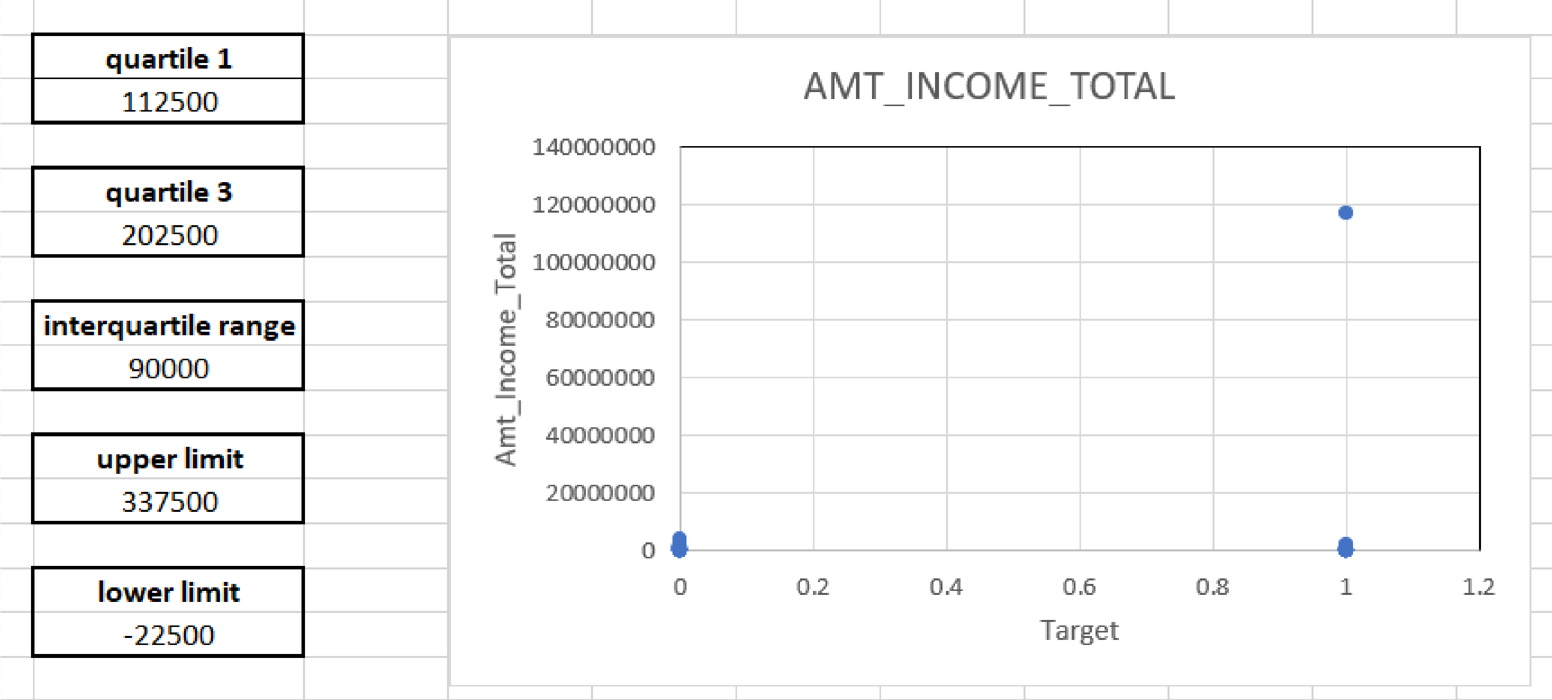


BO	BP	BQ	BR	BS	BT
AMT_RE	AMT_RE	AMT_RE	AMT_RE	AMT_RE	AMT_RE
0	0	0	0	0	1
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	1	1
0	0	0	1	1	2
0	0	0	0	0	0

B. IDENTIFYING OUTLIERS

Task: Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables

- for amt_income_total and target , we can clearly identify the outlier at (1,120000000)

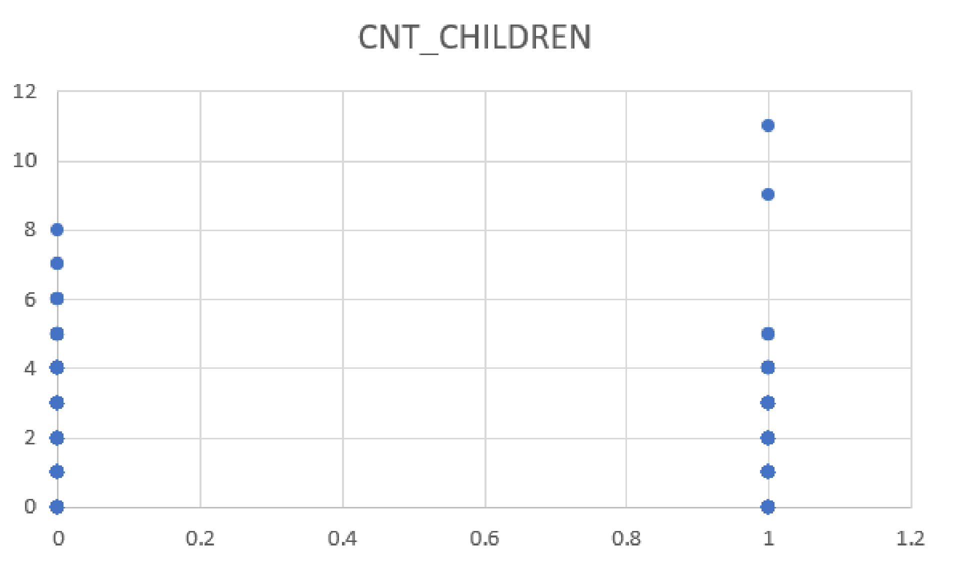


quartile 1	112500
quartile 3	202500
interquartile range	90000
upper limit	337500
lower limit	-22500

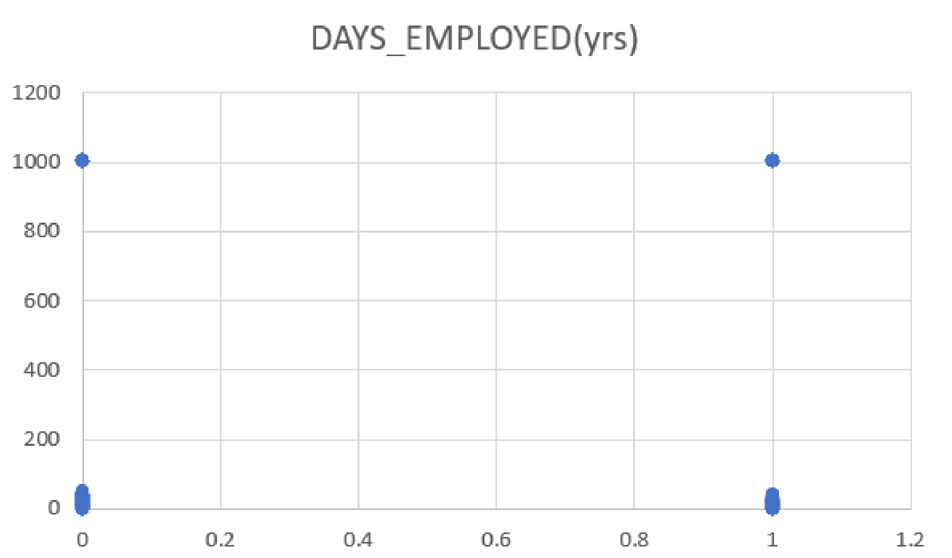
B. IDENTIFYING OUTLIERS

Task: Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables

- ❖ For cnt_children , we can clearly see outliers at (1,9) ,(1,11).



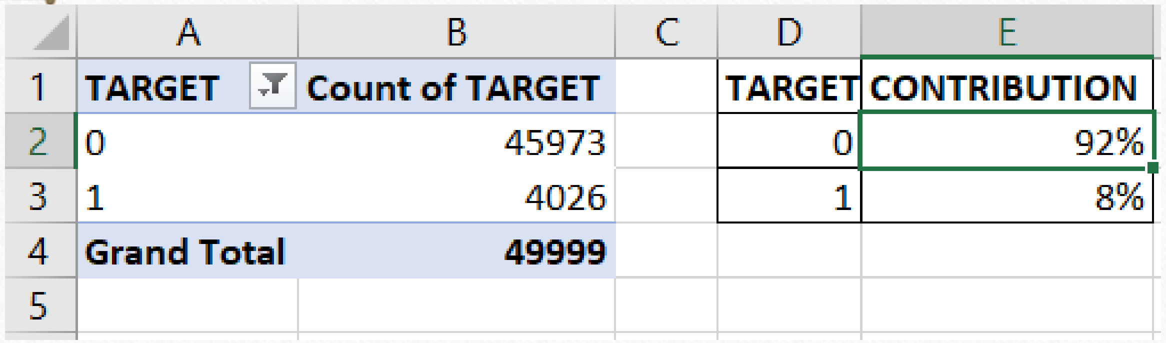
- ❖ For days employed , outliers can be seen at (0,1001) , (1,1000)



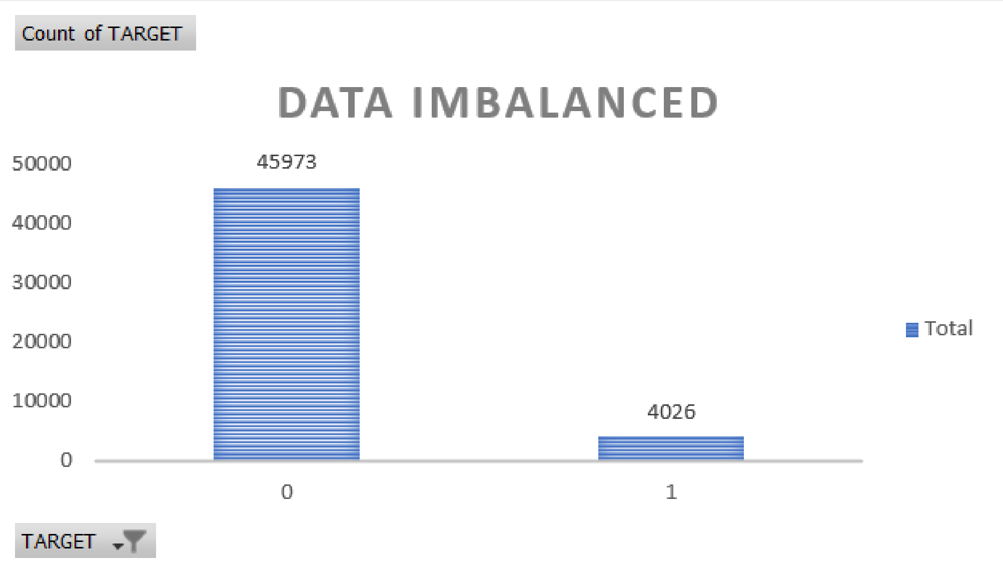
C. DATA IMBALANCE

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

- Using the target column which consists of 2 values 0 and 1.
- Making a pivot table for the count of 0's and 1's
- Then making a table for contribution of each value i.e. 0 and 1.
- Formula used = b2/b4 and =b3/b4 where b2=count of 0 in column 'target' , b3 = count of 1 in the column 'target' and b4 is the total no. of values in the column 'target' --- Then converting it into percentage
- Then making a bar chart to visualize it.



	A	B	C	D	E
1	TARGET	Count of TARGET		TARGET	CONTRIBUTION
2	0	45973		0	92%
3	1	4026		1	8%
4	Grand Total	49999			
5					

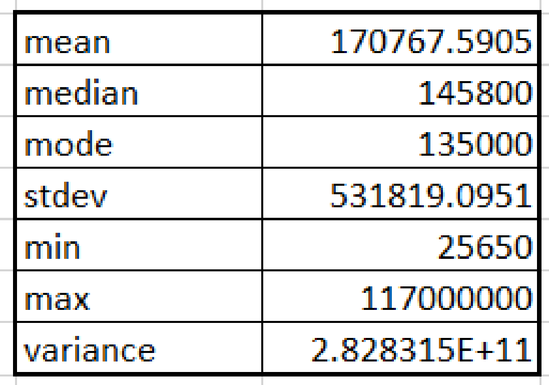


D. UNIVARIATE, SEGMENTED UNIVARIATE AND BIVARIATE ANALYSIS

Task: 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.

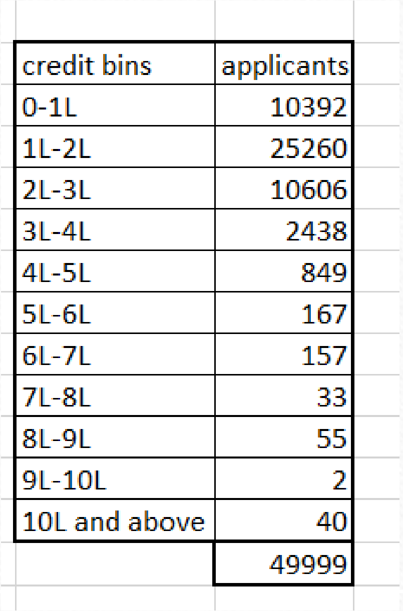
(i) UNIVARIATE ANALYSIS -

FORMULA USED =COUNTIFS(B:B,">=0", B:B,"<=100000") to find the applicants with income in the particular ranges

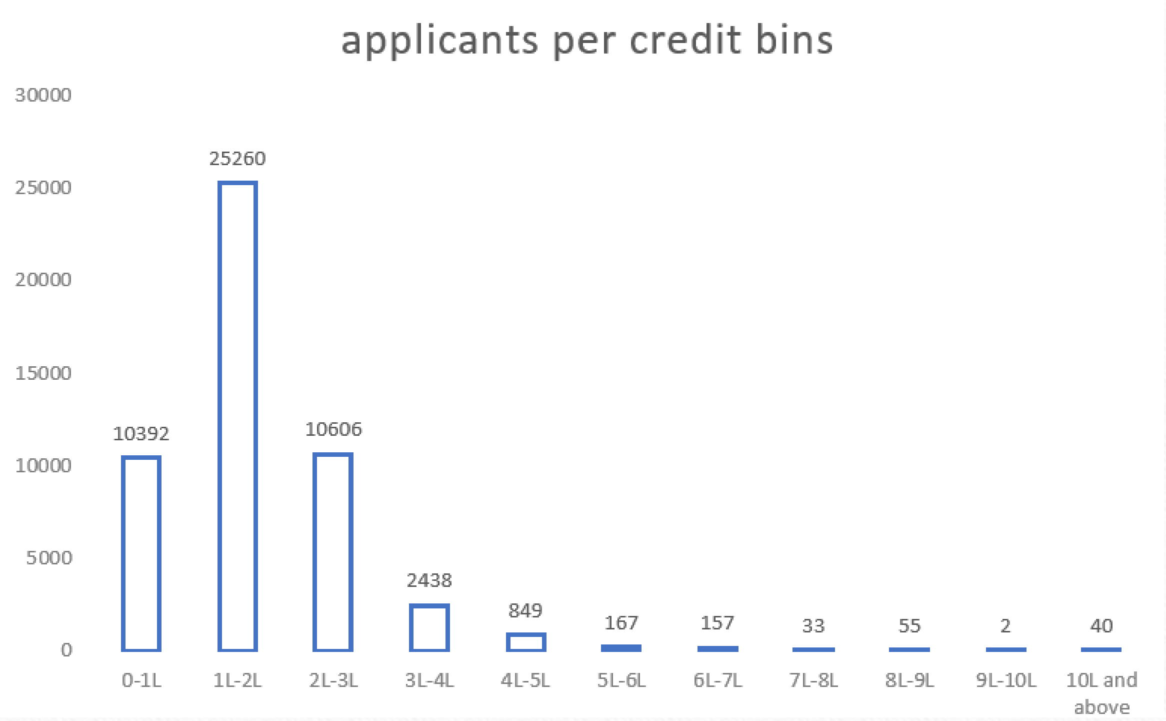


mean	170767.5905
median	145800
mode	135000
stdev	531819.0951
min	25650
max	117000000
variance	2.828315E+11

For amt_income_total

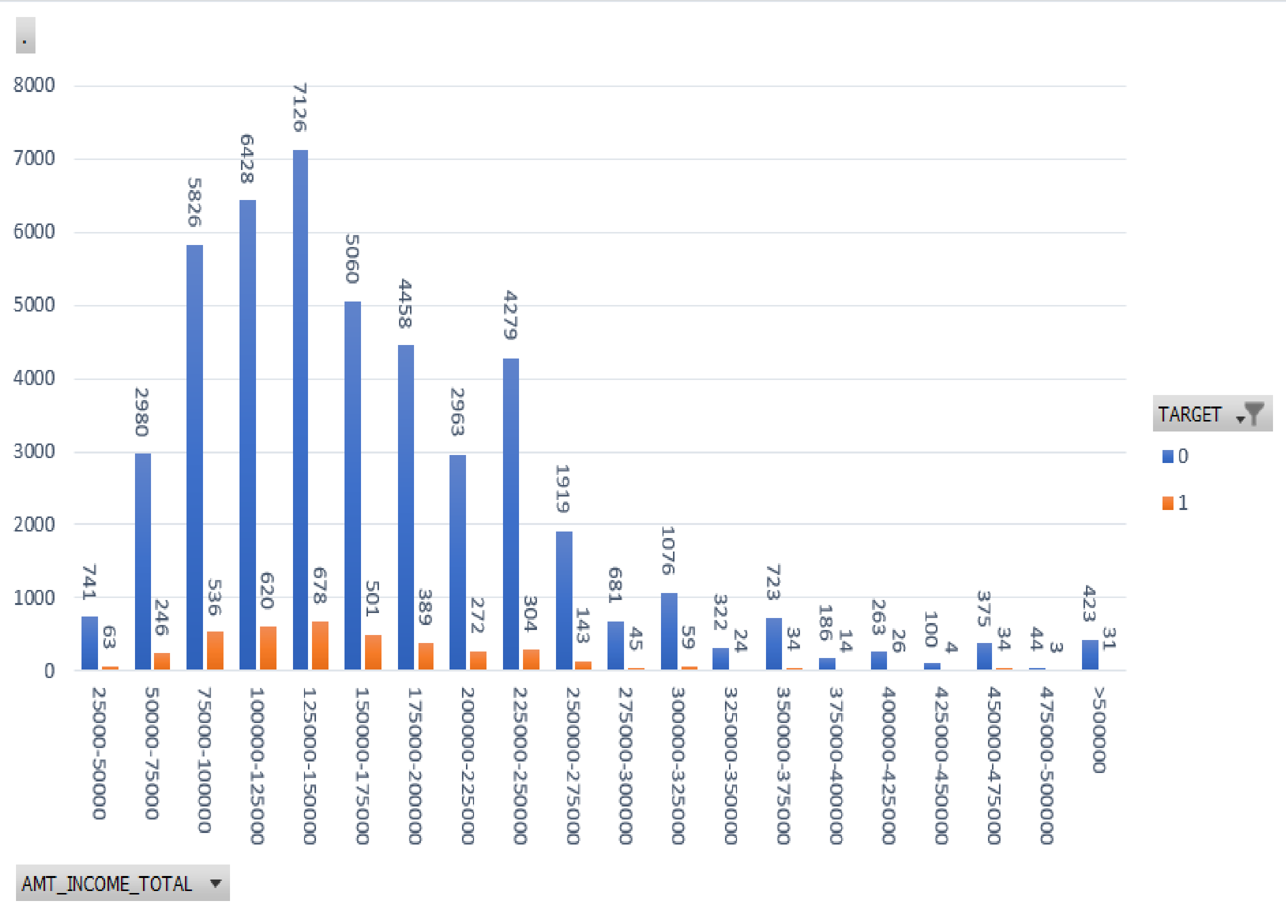


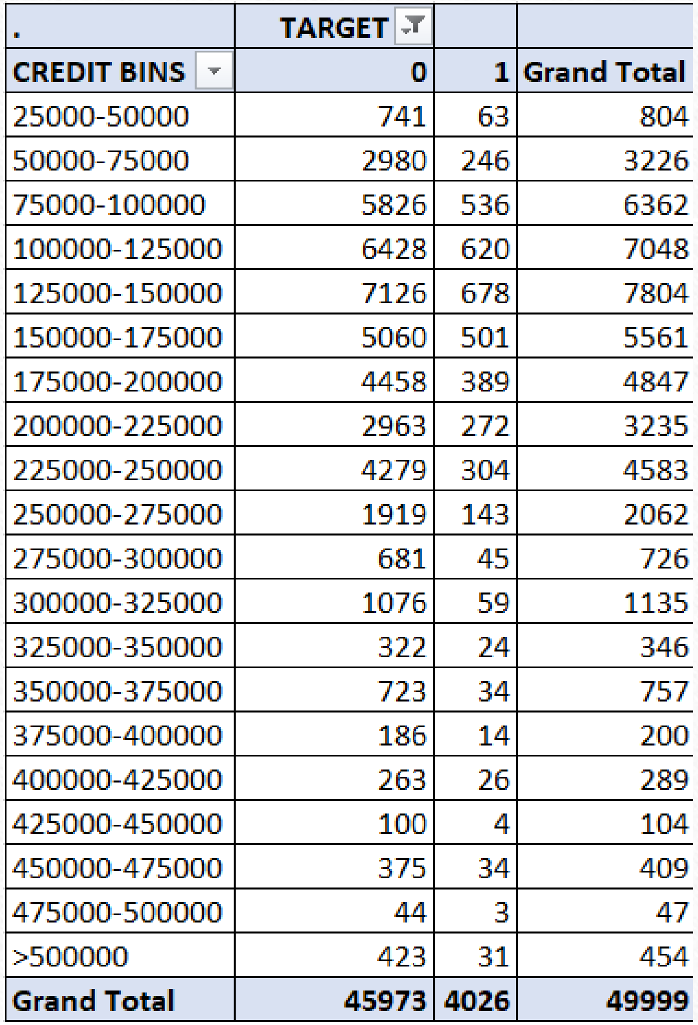
credit bins	applicants
0-1L	10392
1L-2L	25260
2L-3L	10606
3L-4L	2438
4L-5L	849
5L-6L	167
6L-7L	157
7L-8L	33
8L-9L	55
9L-10L	2
10L and above	40
	49999



(ii) SEGMENTED UNIVARIATE ANALYSIS –

- Making a pivot table with the use of target and income columns
- Grouping the income as columns and making the ranges for income column

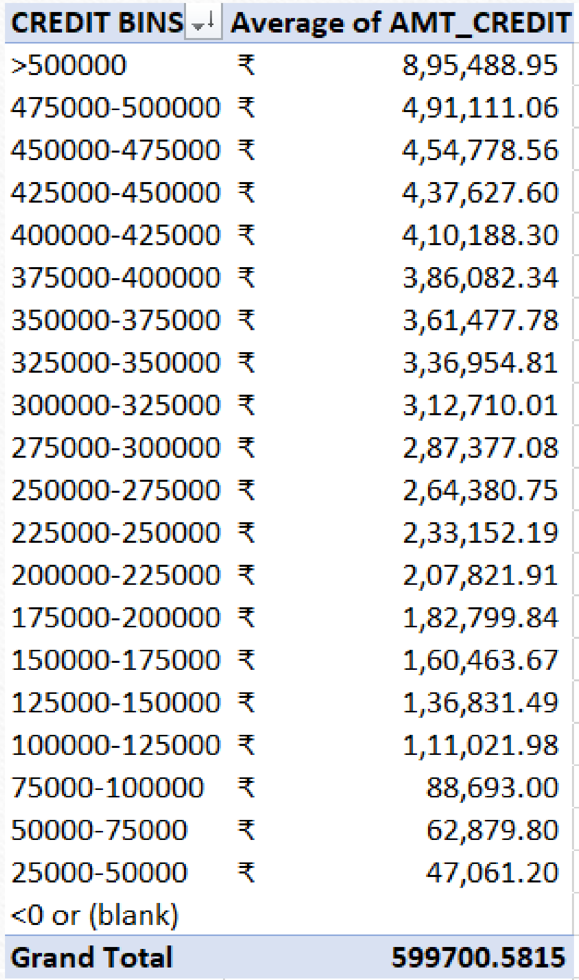




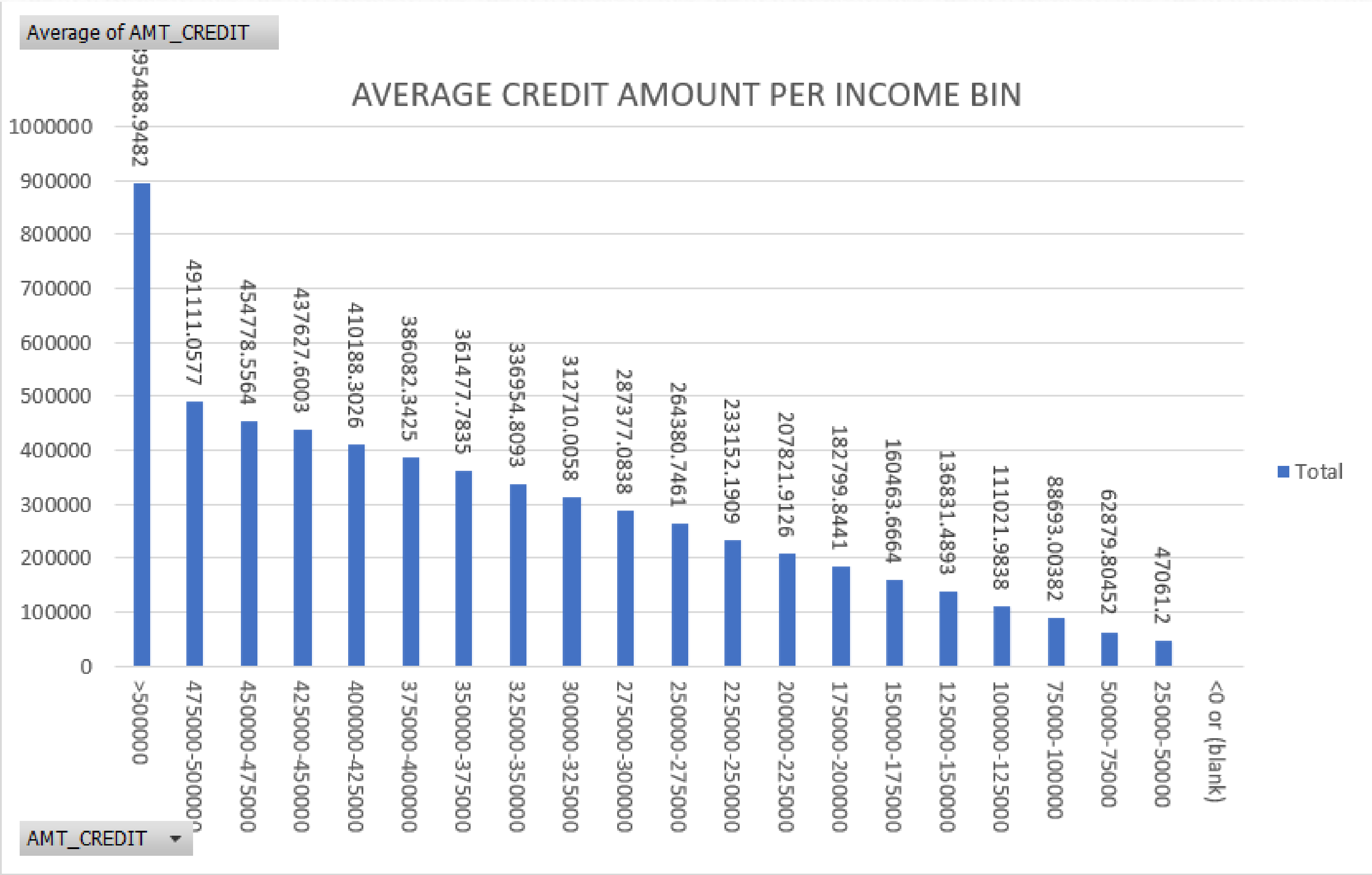
.	TARGET		
CREDIT BINS	0	1	Grand Total
25000-50000	741	63	804
50000-75000	2980	246	3226
75000-100000	5826	536	6362
100000-125000	6428	620	7048
125000-150000	7126	678	7804
150000-175000	5060	501	5561
175000-200000	4458	389	4847
200000-225000	2963	272	3235
225000-250000	4279	304	4583
250000-275000	1919	143	2062
275000-300000	681	45	726
300000-325000	1076	59	1135
325000-350000	322	24	346
350000-375000	723	34	757
375000-400000	186	14	200
400000-425000	263	26	289
425000-450000	100	4	104
450000-475000	375	34	409
475000-500000	44	3	47
>500000	423	31	454
Grand Total	**45973**	**4026**	**49999**

(iii) BIVARIATE ANALYSIS –

- ❖ Making a pivot table with the AMT_CREDIT columns
- ❖ Grouping the credit for making a continuous range and finding the average of credit



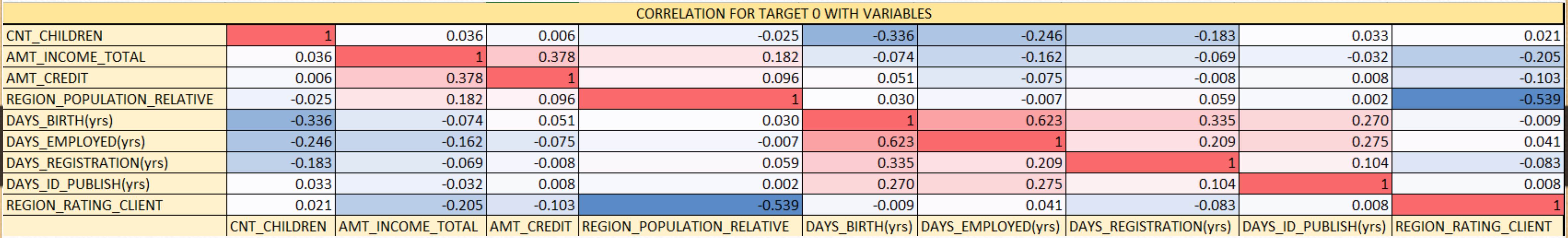
CREDIT BINS	Average of AMT_CREDIT
>500000	₹ 8,95,488.95
475000-500000	₹ 4,91,111.06
450000-475000	₹ 4,54,778.56
425000-450000	₹ 4,37,627.60
400000-425000	₹ 4,10,188.30
375000-400000	₹ 3,86,082.34
350000-375000	₹ 3,61,477.78
325000-350000	₹ 3,36,954.81
300000-325000	₹ 3,12,710.01
275000-300000	₹ 2,87,377.08
250000-275000	₹ 2,64,380.75
225000-250000	₹ 2,33,152.19
200000-225000	₹ 2,07,821.91
175000-200000	₹ 1,82,799.84
150000-175000	₹ 1,60,463.67
125000-150000	₹ 1,36,831.49
100000-125000	₹ 1,11,021.98
75000-100000	₹ 88,693.00
50000-75000	₹ 62,879.80
25000-50000	₹ 47,061.20
<0 or (blank)	
Grand Total	**599700.5815**



E. CORRELATION

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.

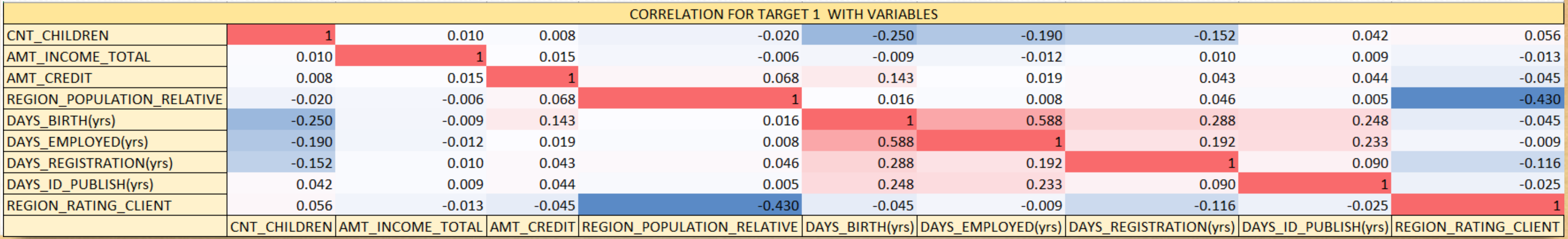
- Correlation with variables (TARGET = 0 i.e. all other cases)



CORRELATION FOR TARGET 0 WITH VARIABLES

CNT_CHILDREN	1	0.036	0.006	-0.025	-0.336	-0.246	-0.183	0.033	0.021
AMT_INCOME_TOTAL	0.036	1	0.378	0.182	-0.074	-0.162	-0.069	-0.032	-0.205
AMT_CREDIT	0.006	0.378	1	0.096	0.051	-0.075	-0.008	0.008	-0.103
REGION_POPULATION_RELATIVE	-0.025	0.182	0.096	1	0.030	-0.007	0.059	0.002	-0.539
DAYS_BIRTH(yrs)	-0.336	-0.074	0.051	0.030	1	0.623	0.335	0.270	-0.009
DAYS_EMPLOYED(yrs)	-0.246	-0.162	-0.075	-0.007	0.623	1	0.209	0.275	0.041
DAYS_REGISTRATION(yrs)	-0.183	-0.069	-0.008	0.059	0.335	0.209	1	0.104	-0.083
DAYS_ID_PUBLISH(yrs)	0.033	-0.032	0.008	0.002	0.270	0.275	0.104	1	0.008
REGION_RATING_CLIENT	0.021	-0.205	-0.103	-0.539	-0.009	0.041	-0.083	0.008	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH(yrs)	DAYS_EMPLOYED(yrs)	DAYS_REGISTRATION(yrs)	DAYS_ID_PUBLISH(yrs)	REGION_RATING_CLIENT

- Correlation with variables (TARGET = 1 i.e. clients with payment difficulties)



CORRELATION FOR TARGET 1 WITH VARIABLES

CNT_CHILDREN	1	0.010	0.008	-0.020	-0.250	-0.190	-0.152	0.042	0.056
AMT_INCOME_TOTAL	0.010	1	0.015	-0.006	-0.009	-0.012	0.010	0.009	-0.013
AMT_CREDIT	0.008	0.015	1	0.068	0.143	0.019	0.043	0.044	-0.045
REGION_POPULATION_RELATIVE	-0.020	-0.006	0.068	1	0.016	0.008	0.046	0.005	-0.430
DAYS_BIRTH(yrs)	-0.250	-0.009	0.143	0.016	1	0.588	0.288	0.248	-0.045
DAYS_EMPLOYED(yrs)	-0.190	-0.012	0.019	0.008	0.588	1	0.192	0.233	-0.009
DAYS_REGISTRATION(yrs)	-0.152	0.010	0.043	0.046	0.288	0.192	1	0.090	-0.116
DAYS_ID_PUBLISH(yrs)	0.042	0.009	0.044	0.005	0.248	0.233	0.090	1	-0.025
REGION_RATING_CLIENT	0.056	-0.013	-0.045	-0.430	-0.045	-0.009	-0.116	-0.025	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH(yrs)	DAYS_EMPLOYED(yrs)	DAYS_REGISTRATION(yrs)	DAYS_ID_PUBLISH(yrs)	REGION_RATING_CLIENT

RESULT

- **Handled Missing Values-** Missing values were effectively managed to ensure accurate analysis.
- **Identified Outliers -** Outliers were identified and addressed to prevent skewing of results.
- **Univariate and Bivariate Analysis -** Detailed analysis was conducted on both individual variables and their relationships.
- **Data Imbalance -** Imbalance in the dataset was identified and accounted for in the analysis.
- **Correlation Analysis -** Correlations between variables were examined to understand their relationships.