Bank Loan Case Study

Link For Excel Sheet:

https://docs.google.com/spreadsheets/d/1mlateqxKvMcygMzOsjjxkTRCFzyoXbFu/edit?usp=sharing&ouid=107365393175079460343&rtpof=true&sd=true

Excel file contains different worksheets which have results of different tasks in them.

Project Description:

This project aims to analyze a dataset containing information about various bank loan applications. The goal is to gain insights about approval of bank loans, such as the relation between income and credit. The data provided has various missing or null Data, our task is to handle those missing values appropriately, by either deleting or imputing these data. There are various outliers in data, we have to find these outliers. We also have to check for data imbalance and perform various analyses on data, such as univariate and bivariate analysis. Finding correlation between various parameters would help us understand what factors affect most in bank loan application approval. Thus, by employing statistics and Excel formulas, we will extract meaningful conclusions to help understand the factors that contribute to a bank loan getting approved.

Approach:

As an individual working on this project, I followed a structured approach to analyze data about bank loan applications. I began by carefully examining the provided database and familiarizing myself with its structure and columns. I tried to find columns which had the most significance in the dataset. I handled missing values by eliminating columns which had most empty cells, and were not significant. And imputed data into cells that were necessary for analysis. Then, I utilized Excel fundamentals to retrieve the necessary information for each task, employing appropriate functions and statistical methods. I focused on data accuracy and quality throughout the project, ensuring reliable results. By leveraging my Excel skills and maintaining a systematic workflow, I successfully executed the project and created a comprehensive report that fulfilled the objectives of providing marketing insights and investor metrics.

Tech-Stack Used:

For this project, I utilized Microsoft Excel as the primary software tool.

Insights:

Task 1:

Identify Missing Data and Deal with it Appropriately (Data Cleaning):

To find data having missing values we utilized COUNTA function in Excel, which returns no. of cells which are not blank.

Formula:

=COUNTA(A4:A50002)

This gave us the number of rows in the TARGET column, which is the total number of rows which we have to consider for analysis.

C17 : Cash loans

	A	B	C	D	E	
1	49999	49999	49999	49999	49999	
2	0	0	0	0	0	
3	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_
4	100002	1	Cash loans	M	N	Y
5	100003	0	Cash loans	F	N	N
6	100004	0	Revolving loans	M	Y	Y
7	100006	0	Cash loans	F	N	Y
8	100007	0	Cash loans	M	N	Y
9	100008	0	Cash loans	M	N	Y
10	100009	0	Cash loans	F	Y	Y

The columns which had missing data in them were found out by using the formula:

=(100-(V1/$A1)*100)

This formula gives us the percentage of missing values in the column.

Alignment | Number | Styles | Cells

AO	AP	AQ	AR	AS	AT	AU	AV
49999	21827	49873	40055	24614	20800	25605	16760
0	56.3451269	0.25200504	19.88839777	50.77101542	58.39916798	48.78897578	66.47932959
RGANIZATION_TYPE	EXT_SOURCE_1	EXT_SOURCE_2	EXT_SOURCE_3	APARTMENTS_AVG	BASEMENTAREA_AVG	YEARS_BEGINEXPLUATATION_AVG	YEARS_BUILD_AVG
usiness Entity Type 3	0.083036967	0.262948593	0.13937578	0.0247	0.0369	0.9722	0.6192
chool	0.311267311	0.622245775		0.0959	0.0529	0.9851	0.796
overnment		0.555912083	0.729566691				
usiness Entity Type 3		0.65044169					
eligion		0.322738287					
ther		0.354224732	0.621226338				
usiness Entity Type 3	0.774761413	0.723999852	0.492060094				
ther		0.714279286	0.54065445				
NA	0.587334047	0.205747288	0.751723715				
lectricity		0.746643629					
edicine	0.319760172	0.651862333	0.363945239				
NA	0.72204445	0.555183162	0.652896552				
usiness Entity Type 2	0.464831117	0.715041819	0.176652579	0.0825		0.9811	
elf-employed		0.566906613	0.77008707	0.1474	0.0973	0.9806	0.7348
ransport: type 2	0.721939769	0.642656205		0.3495	0.1335	0.9985	0.9796
usiness Entity Type 2	0.115634337	0.346633981	0.678567689				
overnment		0.23637784	0.062103038				
onstruction		0.683513346					
lousing		0.706428403	0.556727426	0.0278	0.0617	0.9881	0.8368
indergarten		0.58661714	0.477649155				
elf-employed	0.565654882	0.113374513		0.0722	0.0801	0.9781	0.7008
rade: type 7	0.43770902	0.233766958	0.542445144				
elf-employed		0.457142972	0.358951229	0.0907	0.0795	0.9786	0.7076
NA		0.624304737	0.669056695	0.1443	0.0848	0.9876	0.83

We highlighted columns with missing values by using conditional formatting.

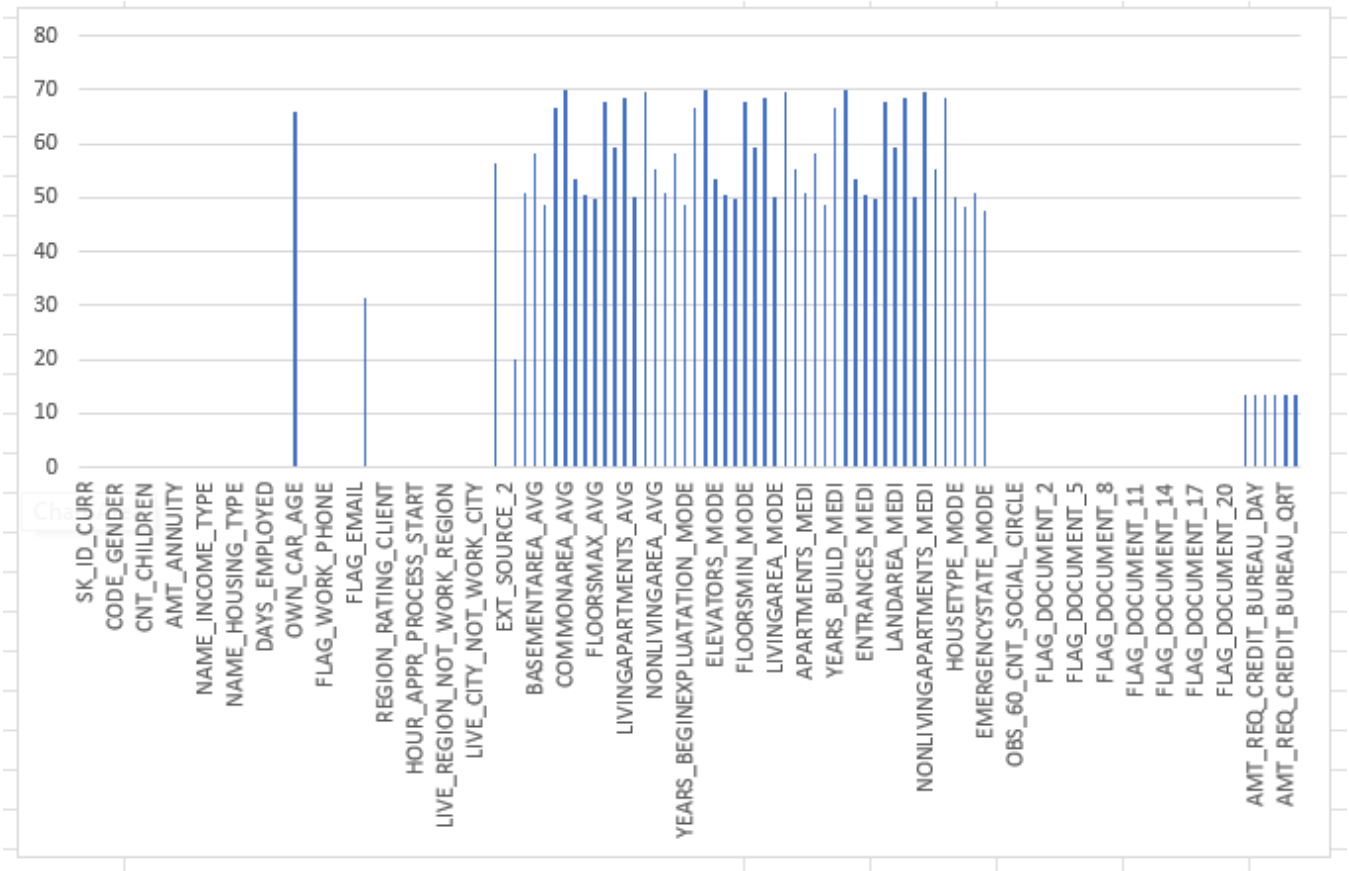
We found no. of columns with missing values greater than 10% by this formula:

=COUNTIF(A2:DR2,">10")
=COUNTIF(A2:DR2,"<10")

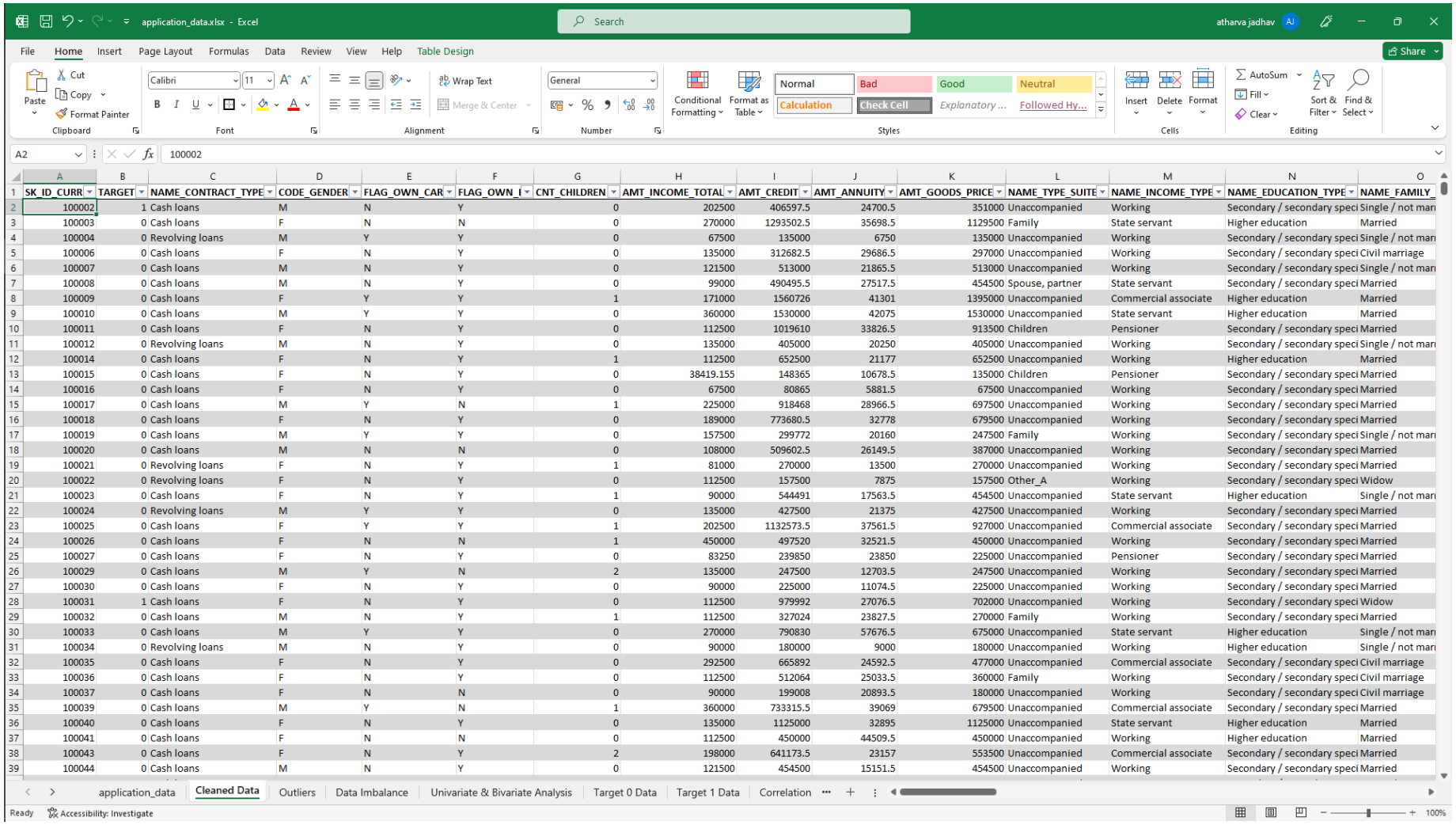
No. of columns with missing data more than 10%	57
No. of columns with missing data less than 10%	65

We plotted a bar graph to better understand the number of columns containing missing values

To Visualize columns and their respective missing values.



We saved all 65 columns with least missing values in a new worksheet called Cleaned Data.



SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_I	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	NAME_FAMILY_
100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied	Working	Secondary / secondary speci	Single / not mar
100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family	State servant	Higher education	Married
100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied	Working	Secondary / secondary speci	Single / not mar
100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied	Working	Secondary / secondary speci	Civil marriage
100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied	Working	Secondary / secondary speci	Single / not mar
100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner	State servant	Secondary / secondary speci	Married
100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate	Higher education	Married
100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied	State servant	Higher education	Married
100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary / secondary speci	Married
100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied	Working	Secondary / secondary speci	Single / not mar
100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied	Working	Higher education	Married
100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children	Pensioner	Secondary / secondary speci	Married
100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied	Working	Secondary / secondary speci	Married
100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied	Working	Secondary / secondary speci	Married
100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied	Working	Secondary / secondary speci	Married
100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary / secondary speci	Single / not mar
100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / secondary speci	Married
100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied	Working	Secondary / secondary speci	Married
100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary / secondary speci	Widow
100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher education	Single / not mar
100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied	Working	Secondary / secondary speci	Married
100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied	Commercial associate	Secondary / secondary speci	Married
100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied	Working	Secondary / secondary speci	Married
100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied	Pensioner	Secondary / secondary speci	Married
100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied	Working	Secondary / secondary speci	Married
100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied	Working	Secondary / secondary speci	Married
100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompanied	Working	Secondary / secondary speci	Widow
100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family	Working	Secondary / secondary speci	Married
100033	0	Cash loans	M	Y	Y	0	270000	790830	57676.5	675000	Unaccompanied	State servant	Higher education	Single / not mar
100034	0	Revolving loans	M	N	Y	0	90000	180000	9000	180000	Unaccompanied	Working	Higher education	Single / not mar
100035	0	Cash loans	F	N	Y	0	292500	665892	24592.5	477000	Unaccompanied	Commercial associate	Secondary / secondary speci	Civil marriage
100036	0	Cash loans	F	N	Y	0	112500	512064	25033.5	360000	Family	Working	Secondary / secondary speci	Civil marriage
100037	0	Cash loans	F	N	N	0	90000	199008	20893.5	180000	Unaccompanied	Working	Secondary / secondary speci	Civil marriage
100039	0	Cash loans	M	Y	N	1	360000	733315.5	39069	679500	Unaccompanied	Commercial associate	Secondary / secondary speci	Married
100040	0	Cash loans	F	N	Y	0	135000	1125000	32895	1125000	Unaccompanied	State servant	Higher education	Married
100041	0	Cash loans	F	N	N	0	112500	450000	44509.5	450000	Unaccompanied	Working	Higher education	Married
100043	0	Cash loans	F	N	Y	2	198000	641173.5	23157	553500	Unaccompanied	Commercial associate	Secondary / secondary speci	Married
100044	0	Cash loans	M	N	Y	0	121500	454500	15151.5	454500	Unaccompanied	Working	Secondary / secondary speci	Married

Task 2:

Identify Outliers in the Dataset:

To find Outliers in the Dataset we utilized functions like QUARTILE, IQR, and conditional formatting to identify potential outliers.

We first copied the columns of interest into a new worksheet for finding Outliers.

Columns Copied are:

SK_ID_CURR	TARGET	AMT_INCOME_TOTAL	CNT_CHILDREN	DAYS_EMPLOYED	DAYS_EMPLOYED(ABS)
100002	1	202500	0	-637	637
100003	0	270000	0	-1188	1188

We used the QUARTILE function to find quartile 1 and quartile 3, along with the IQR and upper limit and lower limit ranges.

Formulae:

=QUARTILE.EXC(Table5[AMT_INCOME_TOTAL],1)

=QUARTILE.EXC(Table5[[#All],[AMT_INCOME_TOTAL]],3)

=I4-I2 (IQR)

=I4+1.5*I6 (Upper limit)

=I2-1.5*I6 (Lower limit)

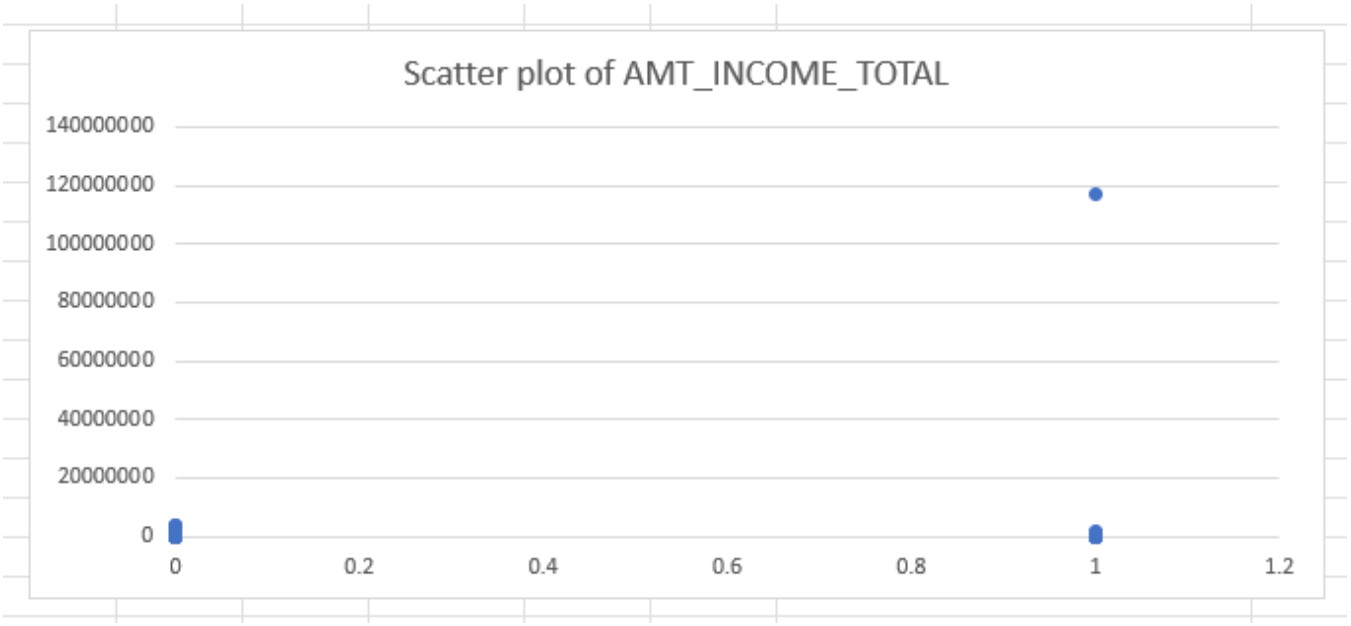
=COUNTIF(C2:C50000,">337500") (Count of elements outside limits)

Outliers in AMT_INCOME_TOTAL					
Quartile 1	112500				
		Upper Limit	337500	Count of elements above upper limit	2295
Quartile 3	202500				
		Lower Limit	-22500		
IQR	90000				

We used Conditional Formatting to highlight the cells which contain values outside the limits.

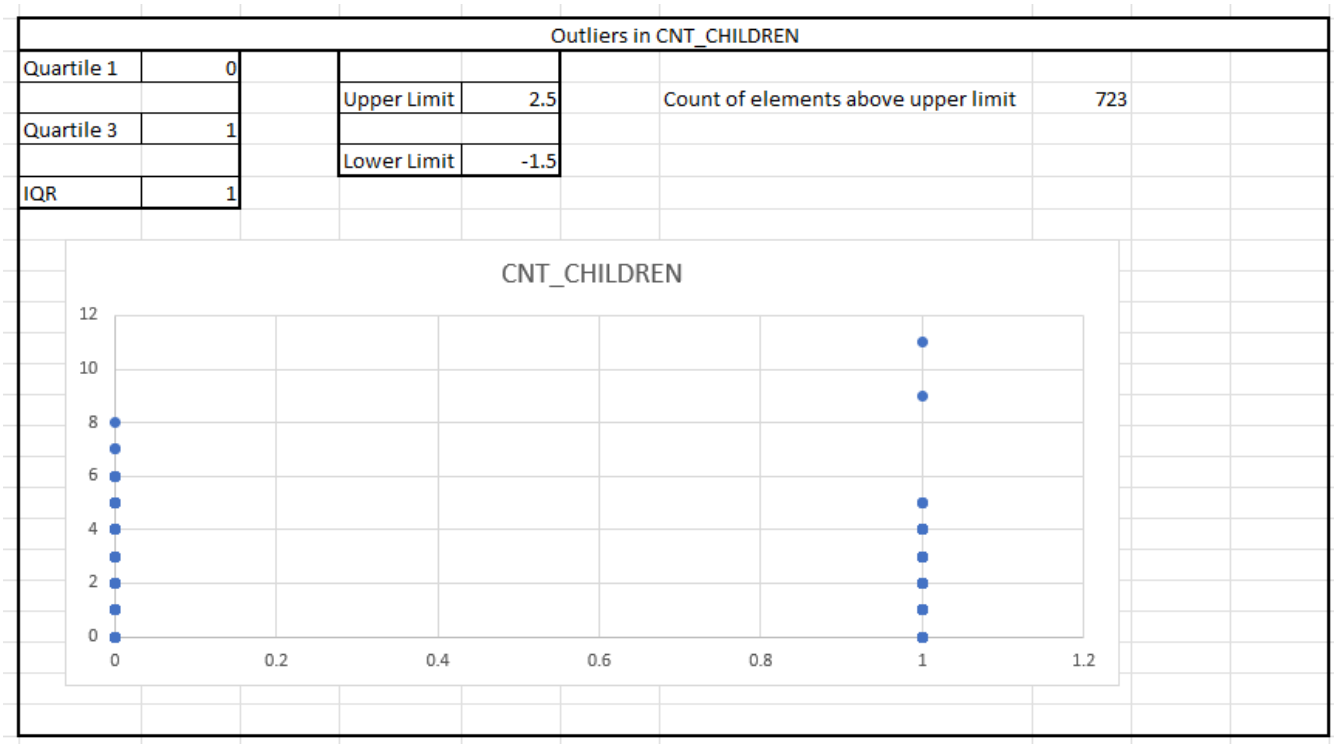
100007	0	121500
100008	0	99000
100009	0	171000
100010	0	360000
100011	0	112500
100012	0	135000
100014	0	112500

We also plotted A scatter plot to visualize the outliers

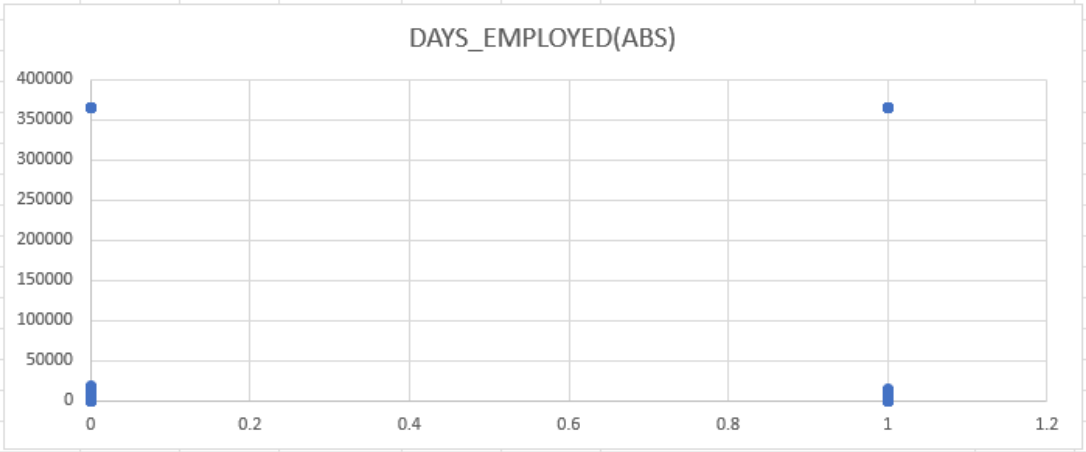


In the above plot the point which lies outside the general trend, and is very much out of the scope can be called an outlier.

Similar Steps were done for other columns and following results were obtained.



Outliers in DAYS_EMPLOYED					
Quartile 1	933				
		Upper Limit	12895.5	Count of elements above upper limit	9082
Quartile 3	5718				
		Lower Limit	-6244.5		
IQR	4785				



	A	B	C	D	E	F
1	SK_ID_CURR	TARGET	AMT_INCOME_TOTAL	CNT_CHILDREN	DAYS_EMPLOYED	DAYS_EMPLOYED(ABS)
2	100002	1	202500	0	-637	637
3	100003	0	270000	0	-1188	1188
4	100004	0	67500	0	-225	225
5	100006	0	135000	0	-3039	3039
6	100007	0	121500	0	-3038	3038
7	100008	0	99000	0	-1588	1588
8	100009	0	171000	1	-3130	3130
9	100010	0	360000	0	-449	449
10	100011	0	112500	0	365243	365243
11	100012	0	135000	0	-2019	2019
12	100014	0	112500	1	-679	679
13	100015	0	38419.155	0	365243	365243
14	100016	0	67500	0	-2717	2717
15	100017	0	225000	1	-3028	3028
16	100018	0	189000	0	-203	203
17	100019	0	157500	0	-1157	1157
18	100020	0	108000	0	-1317	1317
19	100021	0	81000	1	-191	191
20	100022	0	112500	0	-7804	7804
21	100023	0	90000	1	-2038	2038
22	100024	0	135000	0	-4286	4286
23	100025	0	202500	1	-1652	1652
24	100026	0	450000	1	-4306	4306
25	100027	0	83250	0	365243	365243
26	100029	0	135000	2	-746	746
27	100030	0	90000	0	-3494	3494
28	100031	1	112500	0	-2628	2628
29	100032	0	112500	1	-1234	1234
30	100033	0	270000	0	-1796	1796
31	100034	0	90000	0	-1010	1010
32	100035	0	292500	0	-2668	2668
33	100036	0	112500	0	-1104	1104
34	100037	0	90000	0	-4404	4404
35	100039	0	360000	1	-2060	2060
36	100040	0	135000	0	-4585	4585
37	100041	0	112500	0	-1275	1275
38	100043	0	198000	2	-768	768
39	100044	0	121500	0	-1288	1288

In the above image all highlighted cells are outliers.

Task 3:

Analyze 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.

To find Data Imbalance we find the number of each element in the TARGET column. For doing this we use the COUNTIF formula.

Formulae:

=COUNTIF(B:B,0)

=COUNTIF(B:B,1)

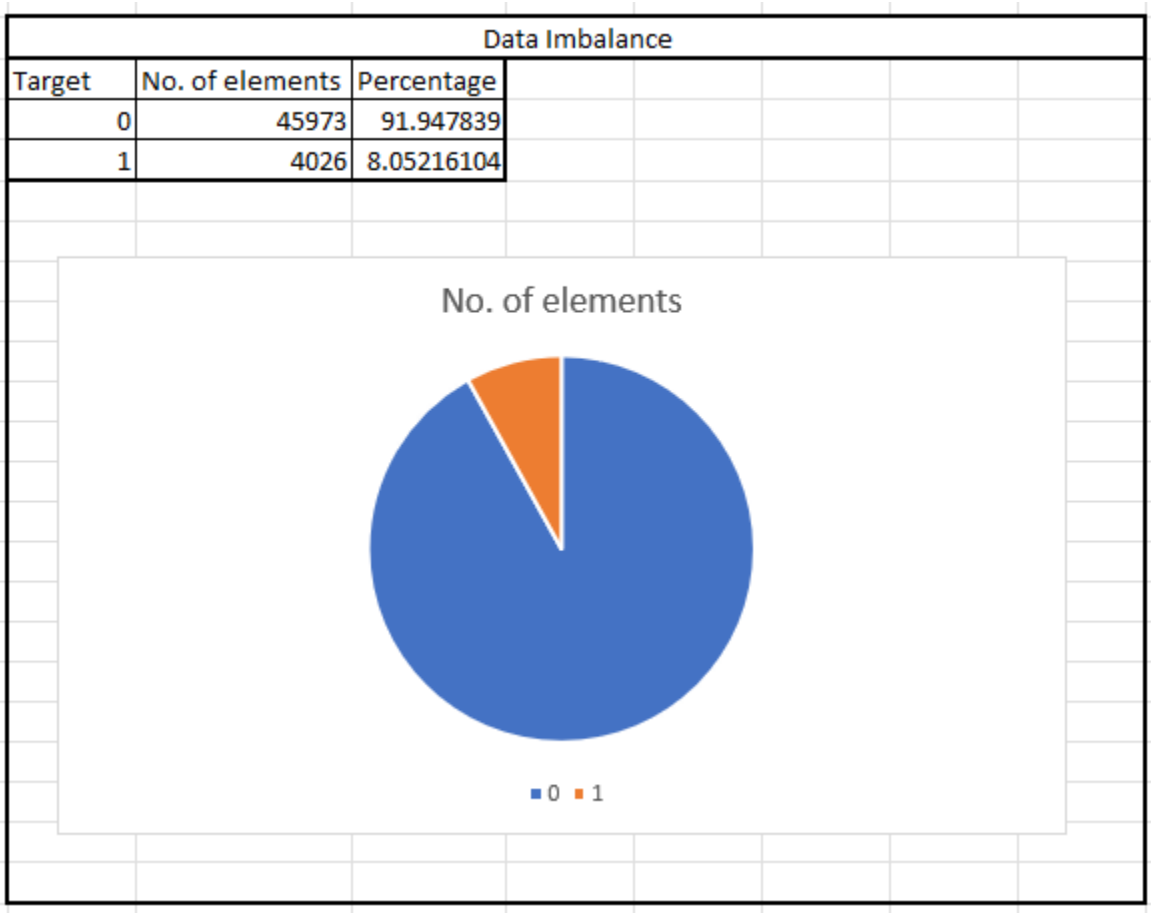
=F4/49999*100 (Percentage)

=F5/49999*100 (Percentage)

We also plot this Data in Pie Chart to visualize the Data Imbalance.

Data Imbalance

Target	No. of elements	Percentage
0	45973	91.947839
1	4026	8.05216104



As we can see the no. of 0 in TARGET is very large compared to no. of 1. This will result in a very large data imbalance. Which might skew the results and give less accurate results.

Task 4:

Perform Univariate, Segmented Univariate, and Bivariate Analysis:

To perform Univariate/ Segmented Univariate analysis, we have to utilize functions such as COUNT, AVERAGE, or MEDIAN to find out the total number of applicants over a particular range or how much credit one shall receive according to their income, and other such relations.

We start by selecting two columns, Credit and Income, we have selected this columns as they have higher correlation. We find maximum and minimum values of these columns excluding outliers.

Maximum Income	117000000	Maximum Credit	4050000
Excluding Outlier	3825000		
Minimum Income	25650	Minimum Credit	45000

This Data helps us to define ranges to find how many applicants fall in each range.

We Define Ranges on particular Intervals

Income Ranges
25000-50000
50000-75000
75000-100000
100000-125000
125000-150000
150000-175000
175000-200000
200000-225000
225000-250000
250000-275000
275000-300000
300000-325000
325000-350000
350000-375000
375000-400000
400000-425000
425000-450000
450000-475000
475000-500000
500000+

Credit Ranges
0-200000
200000-400000
400000-600000
600000-800000
800000-1000000
1000000-1200000
1200000-1400000
1400000-1600000
1600000-1800000
1800000-2000000
2000000-2200000
2200000-2400000
2400000-2600000
2600000-2800000
2800000-3000000
3000000-3200000
3200000-3400000
3400000-3600000
3600000-3800000
3800000+

We find number of applicants over these ranges by utilizing functions such as:

=FREQUENCY(C:C,X5:X23)

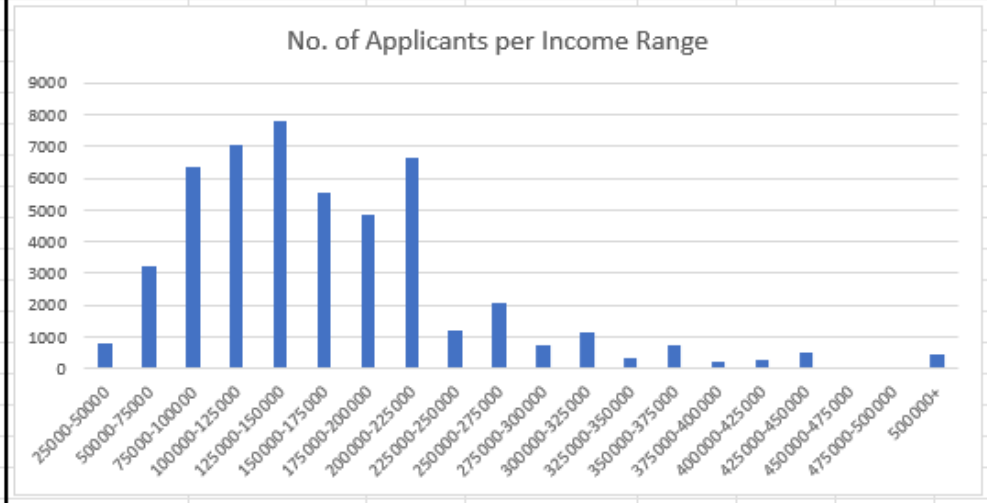
=FREQUENCY(D:D,Y5:Y23)

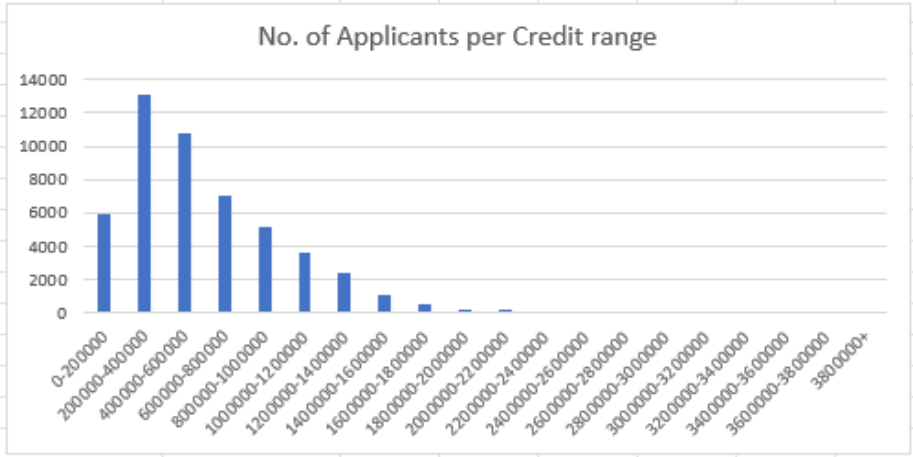
We also plot bar charts to visualize the frequency of applicants in each range.

Univariate Analysis

Income Ranges	No. of Applicants
25000-50000	804
50000-75000	3226
75000-100000	6362
100000-125000	7048
125000-150000	7804
150000-175000	5561
175000-200000	4847
200000-225000	6612
225000-250000	1206
250000-275000	2062
275000-300000	726
300000-325000	1135
325000-350000	346
350000-375000	757
375000-400000	200
400000-425000	289
425000-450000	492
450000-475000	21
475000-500000	47
500000+	454

Credit Ranges	No. of Applicants
0-200000	5900
200000-400000	13105
400000-600000	10782
600000-800000	6971
800000-1000000	5095
1000000-1200000	3615
1200000-1400000	2408
1400000-1600000	1038
1600000-1800000	554
1800000-2000000	208
2000000-2200000	151
2200000-2400000	103
2400000-2600000	49
2600000-2800000	13
2800000-3000000	3
3000000-3200000	1
3200000-3400000	0
3400000-3600000	0
3600000-3800000	0
3800000+	3

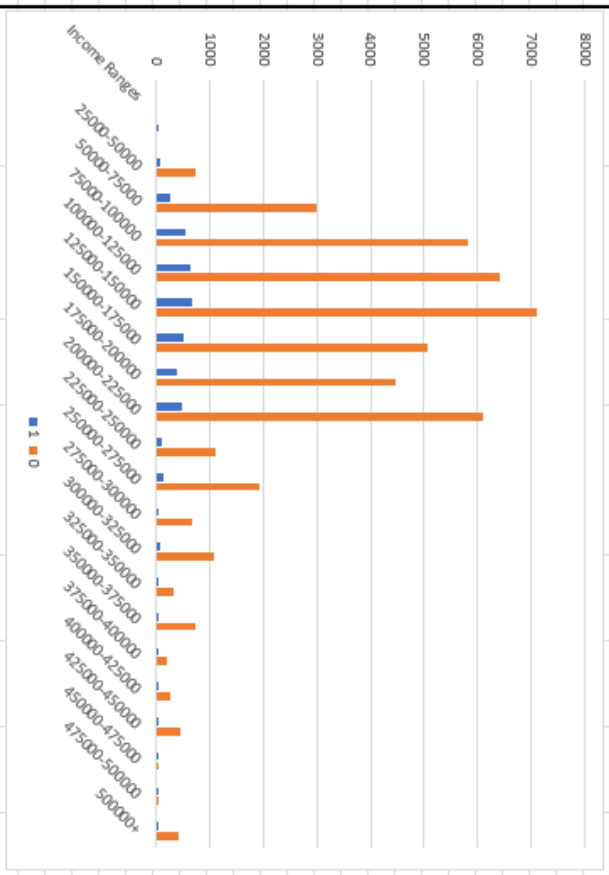




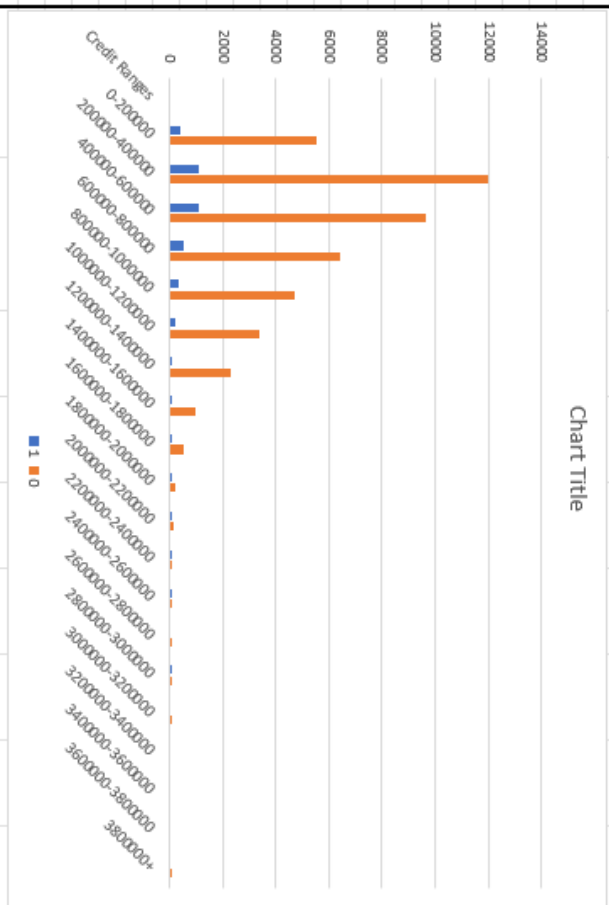
Similarly for segmented univariate analysis we split the Data into two classes according to TARGET.

Segmented Univariate Analysis

Income Ranges	Target	
	1	0
25000-50000	63	741
50000-75000	246	2980
75000-100000	536	5826
100000-125000	620	6428
125000-150000	678	7126
150000-175000	501	5060
175000-200000	389	4458
200000-225000	491	6121
225000-250000	85	1121
250000-275000	143	1919
275000-300000	45	681
300000-325000	59	1076
325000-350000	24	322
350000-375000	34	723
375000-400000	14	186
400000-425000	26	263
425000-450000	36	456
450000-475000	2	19
475000-500000	3	44
500000+	31	423



Credit Ranges	Target	
	1	0
0-200000	390	5510
200000-400000	1126	11979
400000-600000	1129	9653
600000-800000	557	6414
800000-1000000	365	4730
1000000-1200000	246	3369
1200000-1400000	124	2284
1400000-1600000	45	993
1600000-1800000	23	531
1800000-2000000	10	198
2000000-2200000	8	143
2200000-2400000	1	102
2400000-2600000	1	48
2600000-2800000	0	13
2800000-3000000	1	2
3000000-3200000	0	1
3200000-3400000	0	0
3400000-3600000	0	0
3600000-3800000	0	0
3800000+	0	3



To perform Bivariate Analysis, we need to find the average of credit per income range, for that we use the AVERAGEIF function.

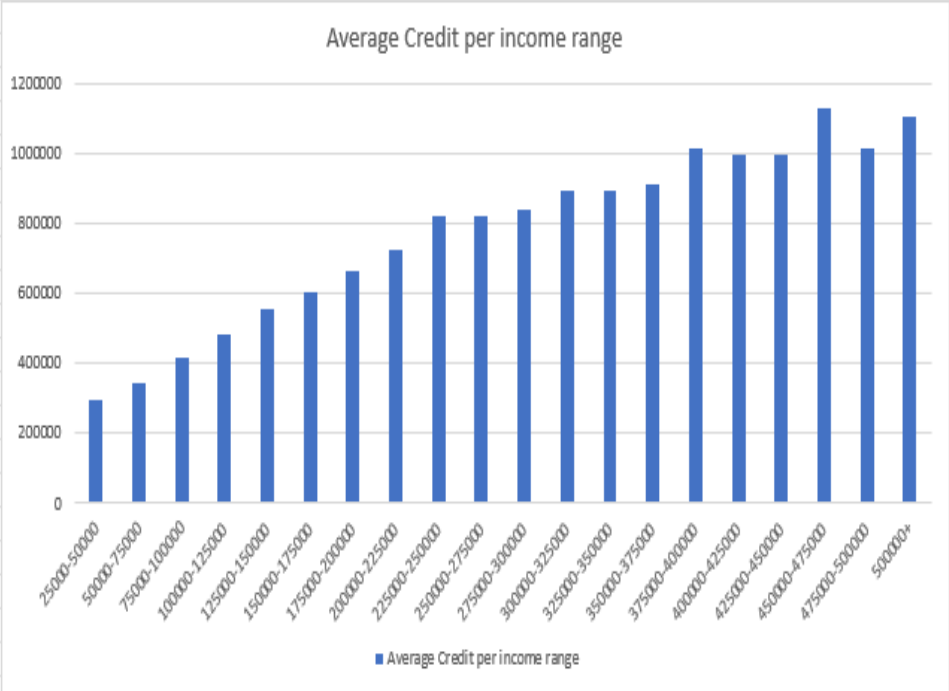
=AVERAGEIFS($D$2:$D$50000,$C$2:$C$50000,">"&X4,$C$2:$C$50000,"<="&X5)

Above formula checks two conditions, if element is greater than lower limit and smaller than upper limit, and only then is considered for average.

We plotted a Bar Graph Similar to above analysis

Bivariate Analysis

Income Ranges	Average Credit per income range
25000-50000	297752.0765
50000-75000	345240.3585
75000-100000	417267.8771
100000-125000	483568.8073
125000-150000	553042.1642
150000-175000	602034.4016
175000-200000	667004.421
200000-225000	727198.4449
225000-250000	822956.3582
250000-275000	820255.3451
275000-300000	842725.6488
300000-325000	892300.0718
325000-350000	892332.6503
350000-375000	910363.0482
375000-400000	1016814.375
400000-425000	999208.199
425000-450000	999153.6402
450000-475000	1132882.5
475000-500000	1015150.404
500000+	1105365.122



Task 5:

Identify Top Correlations for Different Scenarios:

To find Correlation of different columns we utilized the CORREL function of Excel. We first separated the data into three tables, one having only 0 Target, one having 1 Target and both combined Target. We found correlation tables for all these by using the CORREL function, and made it better for visualization using conditional formatting.

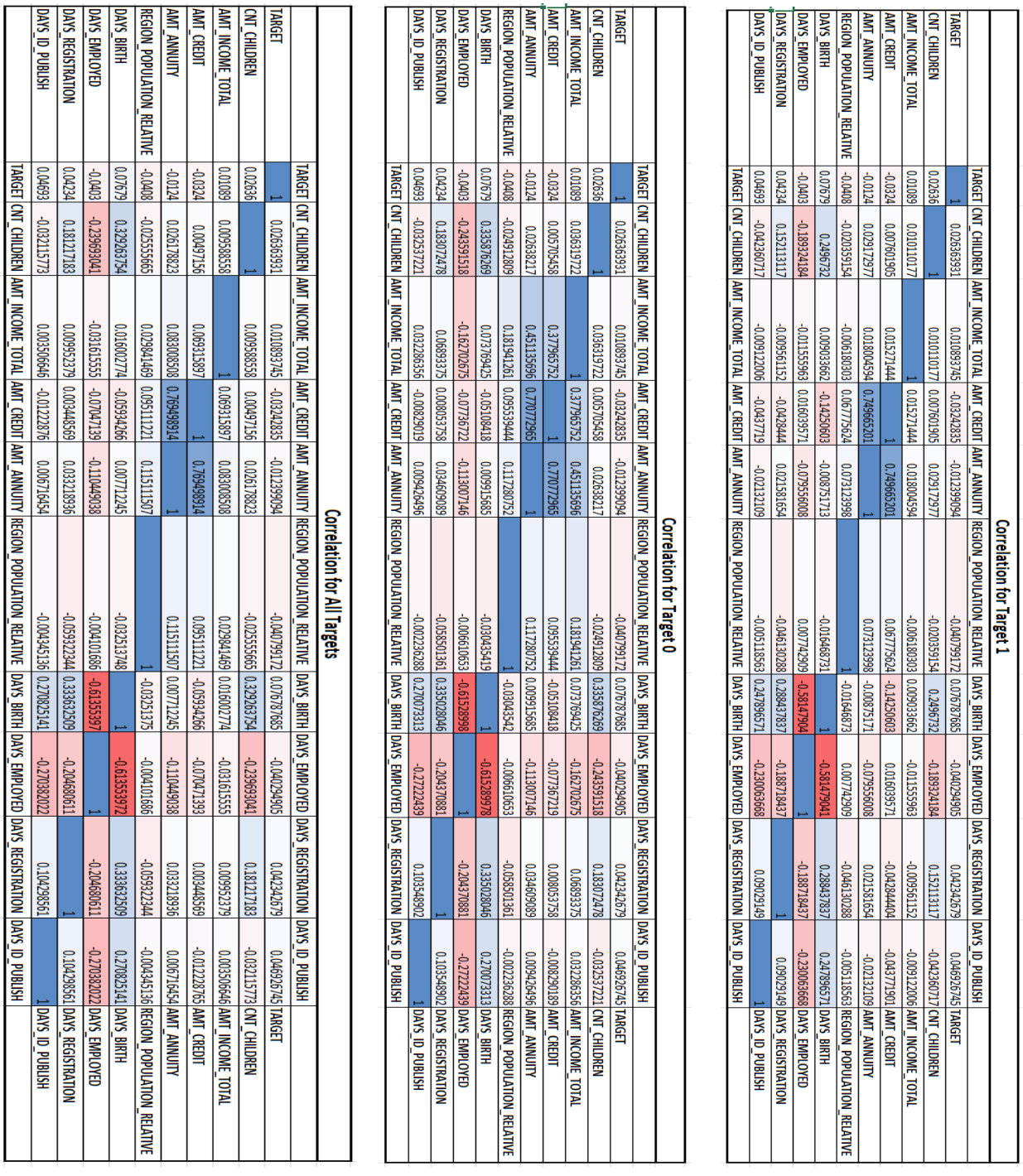
Formulae:

=CORREL($C:$C,B:B)

=CORREL('Target 0 Data'!C:C,'Target 0 Data'!$F:$F)

=CORREL('Target 1 Data'!B:B,'Target 1 Data'!$F:$F)

For better visualization heatmaps of correlation matrix were created.



Correlation for All Targets

	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH
TARGET	1	0.026363931	0.010893745	-0.03242835	-0.012399094	-0.040799172	0.078787685	-0.040294905	0.042342679	0.046926745
CNT_CHILDREN	0.026363931	1	0.009588558	0.004971561	0.026178823	-0.025555665	0.329263754	-0.239693041	0.181217183	-0.032115773
AMT_INCOME_TOTAL	0.010893745	0.009588558	1	0.069315897	0.083008308	0.029841469	0.016002774	-0.031615555	0.009952379	0.003506646
AMT_CREDIT	-0.03242835	0.004971561	0.069315897	1	0.769498914	0.095111221	-0.059342066	-0.070471393	0.003448569	-0.012228765
AMT_ANNUITY	-0.012399094	0.026178823	0.083008308	0.769498914	1	0.115111507	0.007712245	-0.110449038	0.033218936	0.006716454
REGION_POPULATION_RELATIVE	-0.040799172	-0.025555665	0.029841469	0.095111221	0.115111507	1	-0.032513752	-0.004101686	-0.059322344	-0.004345136
DAYS_BIRTH	0.078787685	0.329263754	0.016002774	-0.059342066	0.007712245	-0.032513752	1	-0.613539972	0.333632509	0.270825141
DAYS_EMPLOYED	-0.04029491	-0.239693041	-0.031615555	-0.070471393	-0.110449038	-0.004101686	-0.613539972	1	-0.204680611	-0.270382022
DAYS_REGISTRATION	0.042342679	0.181217183	0.009952379	0.003448569	0.033218936	-0.059322344	0.333632509	-0.204680611	1	0.104298561
DAYS_ID_PUBLISH	0.046926745	-0.032115773	0.003506646	-0.012228765	0.006716454	-0.004345136	0.270825141	-0.270382022	0.104298561	1

Correlation for Target 0

	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH
TARGET	1	0.026363931	0.010893745	-0.03242835	-0.012399094	-0.040799172	0.078787685	-0.040294905	0.042342679	0.046926745
CNT_CHILDREN	0.026363931	1	0.009588558	0.004971561	0.026178823	-0.025555665	0.329263754	-0.239693041	0.181217183	-0.032115773
AMT_INCOME_TOTAL	0.010893745	0.009588558	1	0.069315897	0.083008308	0.029841469	0.016002774	-0.031615555	0.009952379	0.003506646
AMT_CREDIT	-0.03242835	0.004971561	0.069315897	1	0.769498914	0.095111221	-0.059342066	-0.070471393	0.003448569	-0.012228765
AMT_ANNUITY	-0.012399094	0.026178823	0.083008308	0.769498914	1	0.115111507	0.007712245	-0.110449038	0.033218936	0.006716454
REGION_POPULATION_RELATIVE	-0.040799172	-0.025555665	0.029841469	0.095111221	0.115111507	1	-0.032513752	-0.004101686	-0.059322344	-0.004345136
DAYS_BIRTH	0.078787685	0.329263754	0.016002774	-0.059342066	0.007712245	-0.032513752	1	-0.613539972	0.333632509	0.270825141
DAYS_EMPLOYED	-0.04029491	-0.239693041	-0.031615555	-0.070471393	-0.110449038	-0.004101686	-0.613539972	1	-0.204680611	-0.270382022
DAYS_REGISTRATION	0.042342679	0.181217183	0.009952379	0.003448569	0.033218936	-0.059322344	0.333632509	-0.204680611	1	0.104298561
DAYS_ID_PUBLISH	0.046926745	-0.032115773	0.003506646	-0.012228765	0.006716454	-0.004345136	0.270825141	-0.270382022	0.104298561	1

Correlation for Target 1

	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH
TARGET	1	0.026363931	0.010893745	-0.03242835	-0.012399094	-0.040799172	0.078787685	-0.040294905	0.042342679	0.046926745
CNT_CHILDREN	0.026363931	1	0.010110177	0.007601905	0.029172977	-0.020359154	0.249673 2	-0.189324184	0.152113117	-0.042360717
AMT_INCOME_TOTAL	0.010893745	0.010110177	1	0.015271444	0.018004594	-0.006180303	0.009033662	-0.011555963	-0.009561152	-0.009122006
AMT_CREDIT	-0.03242835	0.007601905	0.015271444	1	0.749665201	0.067775624	-0.142506033	0.016039571	-0.042844404	-0.043771901
AMT_ANNUITY	-0.012399094	0.029172977	0.018004594	0.749665201	1	0.073123998	-0.008751171	-0.079556008	0.021581654	-0.021321094
REGION_POPULATION_RELATIVE	-0.040799172	-0.020359154	-0.006180303	0.067775624	0.073123998	1	-0.016468731	0.007742909	-0.046130288	-0.005118563
DAYS_BIRTH	0.078787685	0.249673 2	0.009033662	-0.142506033	-0.008751171	-0.016468731	1	-0.581479041	0.288437837	-0.024796571
DAYS_EMPLOYED	-0.040294905	-0.189324184	-0.011555963	0.016039571	-0.079556008	0.007742909	-0.581479041	1	-0.188718437	-0.230063668
DAYS_REGISTRATION	0.042342679	0.152113117	-0.009561152	-0.042844404	0.021581654	-0.046130288	0.288437837	-0.188718437	1	0.090291491
DAYS_ID_PUBLISH	0.046926745	-0.042360717	-0.009122006	-0.043771901	-0.021321094	-0.005118563	0.247896571	-0.230063668	0.090291491	1

Results:

While working on this project, I have gained a better understanding of Bank Loan Application Process and Analytics and Advanced Excel methodologies. By analyzing Application Data, I was able to provide insights on various aspects such as Cleaning the Data, Outliers in the Data, Data Imbalance, Univariate and Bivariate Analysis, and correlation between various parameters in bank loan application.

This project has helped me enhance my Excel skills, particularly in functions and data visualization to derive meaningful insights. It has also improved my ability to interpret data and provide actionable recommendations based on the analysis.