BANK LOAN CASE STUDY

PROJECT DESCRIPTION:

This project aims to discover the crucial factors that can help predict whether the person to whom the loan is granted will default or not.

APPROACH:

1) Performing Exploratory Data Analysis (EDA)

   application_data:

   All the columns with more than 40% null values are removed as this unknown data might skew the results.

   Columns left with less than 40% null values:

E	F	G	H	I	
AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	H
4219.155	20430	23881.5	20430	FRIDAY	
11110.815	56835	53851.5	56835	MONDAY	
5801.715	80055	48555	80055	WEDNESDAY	
4149.9	34582.5	34582.5	34582.5	TUESDAY	
16288.425	101790	91611	101790	SATURDAY	
17939.205	194427	174982.5	194427	TUESDAY	
	0	0		MONDAY	
	0	0		WEDNESDAY	
22980.6	360000	398016	360000	WEDNESDAY	
18772.425	337500	399870	337500	SATURDAY	
9722.295	46800	36958.5	46800	MONDAY	
43195.185	431995.5	388795.5	431995.5	SUNDAY	
7412.895	63000	67158	63000	SATURDAY	
9208.935	151209	160155	151209	SATURDAY	
21416.85	214191	192771	214191	SATURDAY	
5485.275	28210.5	26896.5	28210.5	FRIDAY	
3959.1	36576	35635.5	36576	THURSDAY	
32242.5	1125000	1125000	1125000	MONDAY	
36111.6	1260000	1260000	1260000	WEDNESDAY	
3951.225	87612.3	87610.5	87612.3	WEDNESDAY	
18000	360000	360000	360000	FRIDAY	
	0	0		THURSDAY	
	0	0		FRIDAY	
22176.405	180000	216418.5	180000	TUESDAY	
	0	0		WEDNESDAY	
24909.39	360000	409896	360000	FRIDAY	
26.87847337	0	0	27.36976181	0	

Y	Z	AA
CNT_PAYMENT	NAME_YIELD_GROUP	PRODUCT_COMBINATION
6	low_normal	POS household with interest
6	high	POS household with interest
10	middle	POS industry with interest
10	middle	POS industry with interest
6	low_normal	POS industry with interest
12	middle	POS industry with interest
	XNA	Card Street
	XNA	Cash
24	middle	Cash X-Sell: middle
48	middle	Cash X-Sell: middle
4	low_normal	POS other with interest
10	low_normal	POS household with interest
12	high	Cash X-Sell: high
24	middle	POS household with interest
10	low_normal	POS household without interest
6	high	POS mobile with interest
10	low_normal	POS household without interest
60	low_normal	Cash Street: low
60	low_normal	Cash Street: low
24	low_action	POS household without interest
0	XNA	Card X-Sell
	XNA	Cash
	XNA	Cash
12	middle	Cash X-Sell: middle
	XNA	Cash
36	high	Cash X-Sell: high
26.87847337	0	0.01600032

EXT_SOURCE2 and EXT_SOURCE3 have been removed due to no significant correlation found between the target variable and them.

AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE
23107.5	495000	Unaccompanied
28503	810000	Unaccompanied
27630	413442	Unaccompanied
26451	567000	Unaccompanied
25317	225000	Unaccompanied
21865.5	477000	Family
29817	459000	Unaccompanied
14184	467874	Unaccompanied
10575	198000	Unaccompanied
29385	252000	Unaccompanied
13500	270000	Unaccompanied
15498	270000	Unaccompanied
36130.5	454500	Unaccompanied
21330	270000	Unaccompanied
26446.5	900000	Unaccompanied
27000	540000	Family
13500	270000	Unaccompanied
9000	180000	Unaccompanied
31653	900000	Unaccompanied
7875	157500	Unaccompanied
22500	450000	Unaccompanied
45936	1206000	
47794.5	1125000	Unaccompanied
26316	900000	Unaccompanied
34897.5	733500	Family
14751	225000	Unaccompanied
0.00200008	0.076059326	0.38400768

AMT_ANNUI	AMT_GOODS_PRI	NAME_TYPE_SUI
10575	198000	Unaccompanied
29385	252000	Unaccompanied
13500	270000	Unaccompanied
15498	270000	Unaccompanied
36130.5	454500	Unaccompanied
21330	270000	Unaccompanied
26446.5	900000	Unaccompanied
27000	540000	Family
13500	270000	Unaccompanied
9000	180000	Unaccompanied
31653	900000	Unaccompanied
7875	157500	Unaccompanied
22500	450000	Unaccompanied
45936	1206000	Unknown
47794.5	1125000	Unaccompanied
26316	900000	Unaccompanied
34897.5	733500	Family
14751	225000	Unaccompanied
0	0	0

AB	AC
OCCUPATION_TYPE	CNT_FAM_MEMBEI
Managers	2
Unemployed	1
Laborers	2
Core staff	1
Accountants	2
Core staff	2
Drivers	1
Laborers	3
Unemployed	2
Sales staff	1
Laborers	2
Unemployed	1
Unemployed	5
Unemployed	1
Waiters/barmen staff	1
Laborers	2
Unemployed	3
Cleaning staff	2
Unemployed	2
0	0

Here, all the column's null values have been treated by the following ways:

AMT_ANNUITY: Paid all at once

AMT_GOODS_PRICE: No goods kept on loan

NAME_TYPE_SUITE: Unknown

OCCUPATION_TYPE- Unemployed

CNT_FAM_MEMBERS- No family

OBS_30_CNT_SOCIAL_CIRCI	DEF_30_CNT_SOCIAL_CIRCI	OBS_60_CNT_SOCIAL_CIRCI	DEF_60_CNT_SOCIAL_CIRCI	DAYS_LAST_PHONE_CHANG
0	0	0	0	-1151
0	0	0	0	0
3	1	3	1	-305
4	0	4	0	-990
1	1	1	1	-602
1	0	1	0	-1798
0	0	0	0	0
0	0	0	0	-824
0	0	0	0	-353
3	0	3	0	-961
0	0	0	0	-533
7	0	7	0	-1226
2	0	2	0	-2338
0	0	0	0	0
0	0	0	0	0
0	0	0	0	-2714
				-550
7	0	7	0	-330
0	0	0	0	-1120
0	0	0	0	0
0	0	0	0	-1737
0	0	0	0	-14
1	0	1	0	-1930
0	0	0	0	-712
1	0	1	0	-1989
1	0	1	0	-295
0.337139532	0.337139532	0.337139532	0.337139532	0.00200008

Here, blank values have been filled by the median of their respective column.

In the DAYS_LAST_PHONE_CHANGE, null values have been filled with “UsingSamePhone”.

AMT_REQ_CREDIT_BUREAU_HOU	AMT_REQ_CREDIT_BUREAU_DA	AMT_REQ_CREDIT_BUREAU_WEE	AMT_REQ_CREDIT_BUREAU_MO	AMT_REQ_CREDIT_BUREAU_QR	AMT_REQ_CREDIT_BUREAU_YEA
0	0	0	0	0	0
0	0	0	0	0	3
0	0	0	0	2	0
0	0	0	0	0	1
0	0	0	0	0	3
1	0	0	0	1	3
0	0	0	0	0	1
0	0	0	0	0	0
0	0	0	0	0	5
0	0	0	0	0	0
0	0	0	0	0	6
0	0	0	0	2	5
0	0	0	0	0	8
0	0	0	0	0	0
0	0	0	0	0	4
0	0	0	0	1	5
0	0	0	0	0	1
0	0	0	0	2	0
0	0	0	0	0	4
0	0	0	0	0	1
0	0	0	0	0	2
0	0	0	0	0	0
0	0	0	0	2	5
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	2
0	0	0	0	2	4
0	0	0	0	0	0
15.56454409	15.56454409	15.56454409	15.56454409	15.56454409	15.56454409

Again, the null values have been treated by calculating the median for their respective columns.

AMT_REQ_CREDIT_BUREAU_HOU	AMT_REQ_CREDIT_BUREAU_DA	AMT_REQ_CREDIT_BUREAU_WEE	AMT_REQ_CREDIT_BUREAU_MO	AMT_REQ_CREDIT_BUREAU_QR	AMT_REQ_CREDIT_BUREAU_YEA
0	0	0	0	0	3
0	0	0	0	0	0
0	0	0	0	2	0
0	0	0	0	0	1
0	0	0	0	0	3
1	0	0	0	1	3
0	0	0	0	0	1
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	5
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	6
0	0	0	0	2	5
0	0	0	0	0	8
0	0	0	0	0	0
0	0	0	0	0	4
0	0	0	0	1	5
0	0	0	0	0	1
0	0	0	0	2	0
0	0	0	0	0	4
0	0	0	0	0	1
0	0	0	0	0	2
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	2	5
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	2
0	0	0	0	2	4
0	0	0	0	0	0
0	0	0	0	0	0

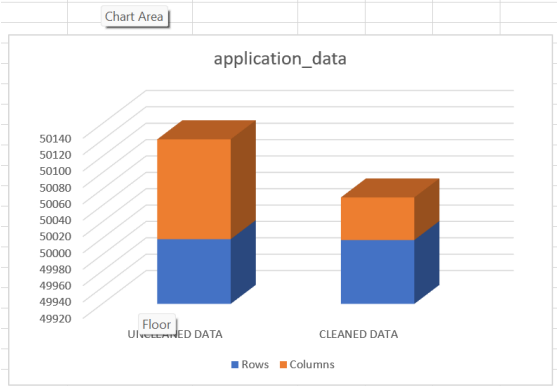
<u>previous_application</u>:

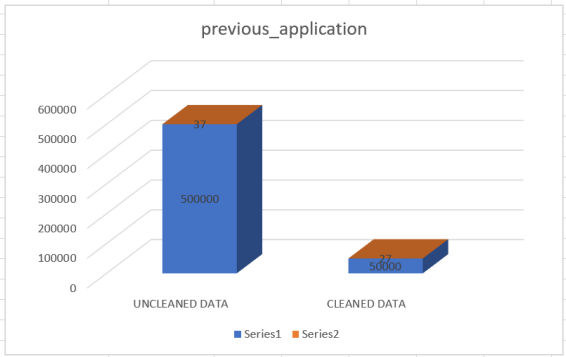
EDA has been performed on the data in the same way.

RESULT:

LABELS	UNCLEANED DATA	CLEANED DATA
Rows	49999	49998
Columns	122	52

LABELS	UNCLEANED DATA	CLEANED DATA
ROWS	500000	50000
COLUMNS	37	27





2) Identifying outliers in the dataset

To identify the outliers, the Interquartile Range has been calculated for each column.

Further, “conditional formatting” has been performed on all columns and the cells with “red box and red text” indicate outliers.

application_data:

FORMULA:

```
=PERCENTILE.INC(B$2:B$50000, 0.25)
```
(Q1)

```
=PERCENTILE.INC(B$2:B$50000, 0.75)
```
(Q3)

```
=B$50005-B$50004
```
(IQR=Q3-Q1)

```
=B$50004-1.5*B$50006
```
Outlier: Q1- 1.5*IQR

```
=B$50005+1.5*B$50006
```
Outlier: Q3+1.5*IQR

	A	B	C	D	E	F	G	H	I	J	K	L
1		SK_ID_CU	TARG	NAME_CONTRACT_TYI	CODE_GEND	FLAG_OWN_C	FLAG_OWN_REAL	CNT_CHILDR	AMT_INCOME_TOT	AMT_CRED	AMT_ANNUI	AMT_GOODS_PRI
50000												
50001	COUNT	0	0	0	0	0	0	0	0	0	0	0
50002	MEAN	129012.63	0.0805					0.419856794	170768.7559	599707.185	27107.14641	539066.3223
50003	MEDIAN	129012.63	0					0	145800	514777.5	24939	450000
50004	Q1	114570.25	0					0	112500	270000	16456.5	238500
50005	Q3	143437.75	0					1	202500	808650	34596	679500
50006	IQR	28867.5	0					1	90000	538650	18139.5	441000
50007	LOWER	71269	0					-1.5	-22500	-537975	-10752.75	-423000
50008	UPPER	186739	0					2.5	337500	1616625	61805.25	1341000
50009												

	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	REGION_POPULATION_RELATI	DAYS_BIR	DAYS_EMPLOY	DAYS_REGISTRATIC	DAYS_ID_PUBLI	FLAG_MO	FLAG_EMP_PHO	FLAG_WORK_PHO	FLAG_CONT_MOBI	FLAG_PHO	FLAG_EM/	OCCUPATION_TYPE	CNT_FAM_MEMBEI
50000													
50001	0	0	0	0	0	0	0	0	0	0	0	0	0
50002	0.020798196	-16021.879	63213.38378	-4977.082423	-2996.766431	0.99998	0.82149286	0.199267971	0.997979919	0.27773111	0.05566223		2.158906356
50003	0.01885	-15731	-1221	-4490	-3261	1	1	0	1	0	0		2
50004	0.010006	-19644	-2786	-7463	-4297	1	1	0	1	0	0		2
50005	0.028663	-12378.25	-292	-1998	-1722	1	1	0	1	1	0		3
50006	0.018657	7265.75	2494	5465	2575	0	0	0	0	1	0		1
50007	-0.0179795	-30542.625	-6527	-15660.5	-8159.5	1	1	0	1	-1.5	0		0.5
50008	0.0566485	-1479.625	3449	6199.5	2140.5	1	1	0	1	2.5	0		4.5
50009													

AE	AF	AG	AH	AI	AJ	AK	AL
REGION_RATING_CLIE	REGION_RATING_CLIENT_W_CI	WEEKDAY_APPR_PROCESS_STAI	HOUR_APPR_PROCESS_STAI	REG_REGION_NOT_LIVE_REGIC	REG_REGION_NOT_WORK_REGIC	LIVE_REGION_NOT_WORK_REGIC	REG_CITY_NOT_LIVE_CI
0	0	0	0	0	0	0	0
2.051662066	2.030721229		12.05264211	0.0150006	0.049921997	0.039641586	0.079963199
2	2		12	0	0	0	0
2	2		10	0	0	0	0
2	2		14	0	0	0	0
0	0		4	0	0	0	0
2	2		4	0	0	0	0
2	2		20	0	0	0	0

REG_CITY_NOT_WORK_CI	LIVE_CITY_NOT_WORK_CI	ORGANIZATION_TY	OBS_30_CNT_SOCIAL_CIRC	DEF_30_CNT_SOCIAL_CIRC	OBS_60_CNT_SOCIAL_CIRC	DEF_60_CNT_SOCIAL_CIRC	YEAR_LAST_PHONE_CHANC	DAYS_LAST_PHONE_CHANC
0	0	0	0	0	0	0	0	
0.232169287	0.179707188		129013.2106	129013.2106	129013.2106	129013.2106	2.641816138	-964.26289
0	0		129076	129076	129076	129076	2.067123288	-754
0	0		114570.5	114570.5	114570.5	114570.5	0.739726027	-157
0	0		143438.5	143438.5	143438.5	143438.5	4.309589041	-27
0	0		28868	28868	28868	28868	3.569863014	130
0	0	0	4	4	4	4	-4.615068493	-3527
0	0	0	20	20	20	20	9.664383562	1684

DAYS_LAST_PHONE_CHANC	AMT_REQ_CREDIT_BUREAU_HOU	AMT_REQ_CREDIT_BUREAU_DA	AMT_REQ_CREDIT_BUREAU_WEE	AMT_REQ_CREDIT_BUREAU_MO	AMT_REQ_CREDIT_BUREAU_QR	AMT_REQ_CREDIT_BUREAU_YEA
0	0	0	0	0	0	0
-964.2628905	0	0	0	0	0	0
-754.5	0	0	0	0	0	0
-1573	0	0	0	0	0	0
-270	0	0	0	0	0	0
1303	0	0	0	0	0	0
-3527.5	0	0	0	0	0	0
1684.5	0	0	0	0	0	0

CNT_CHILDR	AMT_INCOME_TOT	AMT_CRED	AMT_ANNUI	AMT_GOODS_PRI
2	157500	232344	11992.5	157500
0	112500	717003	27913.5	598500
0	121500	189351	14782.5	153000
0	315000	814041	23800.5	679500
0	225000	1125000	47794.5	1125000
0	67500	348264	20124	315000
0	162000	521280	41926.5	450000
0	112500	225000	11488.5	225000
1	180000	1288350	37800	1125000
0	337500	2169913.5	59800.5	1939500
0	337500	719946	33489	643500
3	37800	269550	18364.5	225000
1	90000	225000	13045.5	225000
0	180000	1288350	37800	1125000
0	441000	701730	72036	675000
1	135000	765000	25281	765000
0	99000	188685	16272	157500
0	112500	490495.5	27387	454500
1	202500	182983.5	13810.5	166500
0	96750	360000	19660.5	360000
0	81000	283419	15952.5	234000
1	225000	1540588.5	58810.5	1377000
0	135000	405000	20250	405000
0	157500	418500	22041	418500
2	157500	545040	26640	450000
1	180000	400500	31122	400500
0	180000	916470	26928	765000
0	90000	540000	28768.5	540000
1	103500	331834.5	17073	252000
0	90000	284400	19134	225000
0	67500	938304	31140	810000
0	360000	135000	6750	135000
1	76500	639000	30739.5	639000
0	225000	900000	26446.5	900000
0	225000	942300	30528	675000
0	90000	405000	19480.5	405000
0	135000	668304	26631	540000
2	157500	270000	13500	270000
1	202500	1120068	37017	1003500

previous_application:

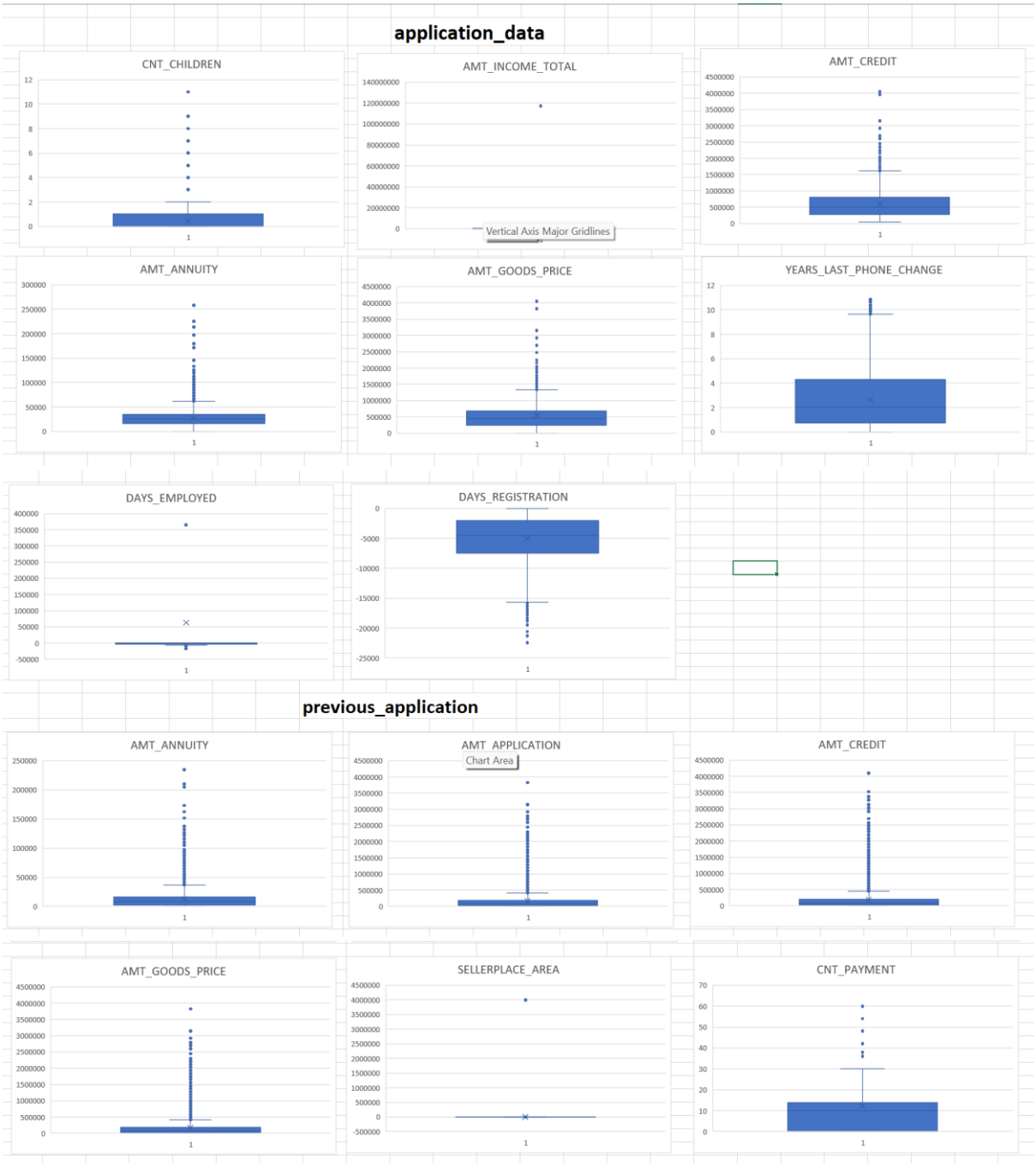
Same steps as above have been applied here.

	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HOUR_APPR_PROCESS_START	FLAG_LAST_APPL_PER_CONTRACT	NFLAG_LAST_APPL_IN_DAY	NA
COUNT	0	0	0	0	0	0	0	0	0	0	0	
Q1	1457920	189919.5		1457920	22045.5	26055	1457920		10		1	
Q3	2388632	368527.5		2388632	180000	198105.75	2388632		15		1	
IQR	930712	178608		930712	157954.5	172050.75	930712		5		0	
LOWER	61852	-77992.5		61852	-214886.25	-232021.125	61852		2.5		1	
UPPER	3784700	636439.5		3784700	416931.75	456181.875	3784700		22.5		1	

DAYS_DECISION	N
0	
-1335	
-292	
1043	
-2899.5	
1272.5	

SELLERPLACE_AREA	NAME_SELLER_INDUSTRY	CNT_PAYMENT	NAME_YIELD_GROUP	PRODUCT_COMBINATION	NFLAG_INSURED_ON_APPROVAL
0	0	0	0	0	0
-1		1457920			1457920
100		2388632			2388632
101		930712			930712
-152.5		61852			61852
251.5		3784700			3784700

RESULT:



3) Analyzing data imbalance

To find the data imbalance, the count of “defaulters and others” has been found in “application_data” & count of “approved, canceled, refused, and unused offer” has been calculated. Further, its ratio has been calculated to understand the data distribution.

FORMULA:

```
=COUNTIF(previous_application4[NAME_CONTRACT_STATUS],"Approved")
```

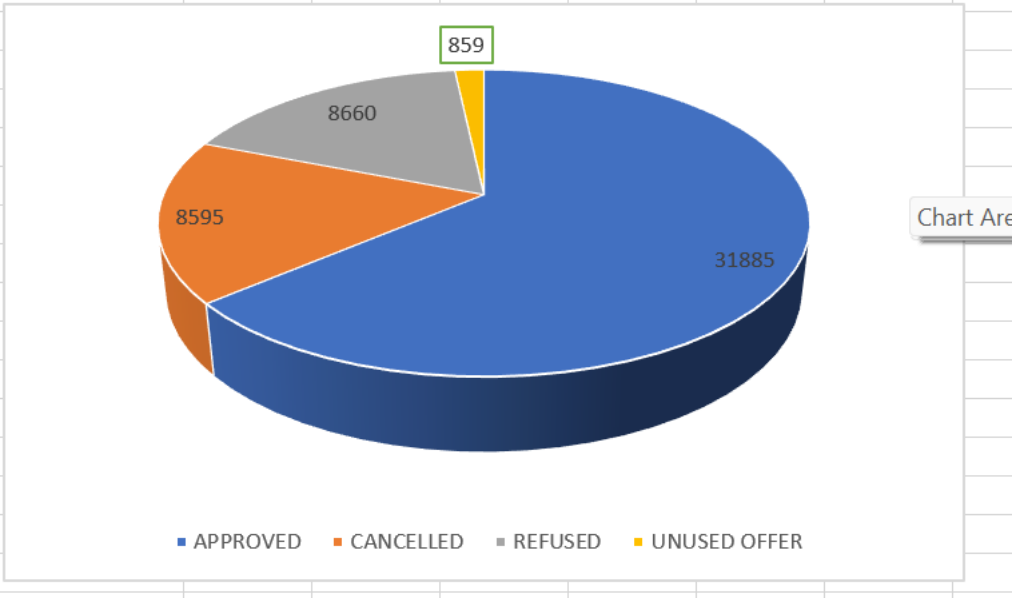
```
=COUNTIF('application_data (2)'!C2:C49999,"1")
```

```
=B2/$B$5
```

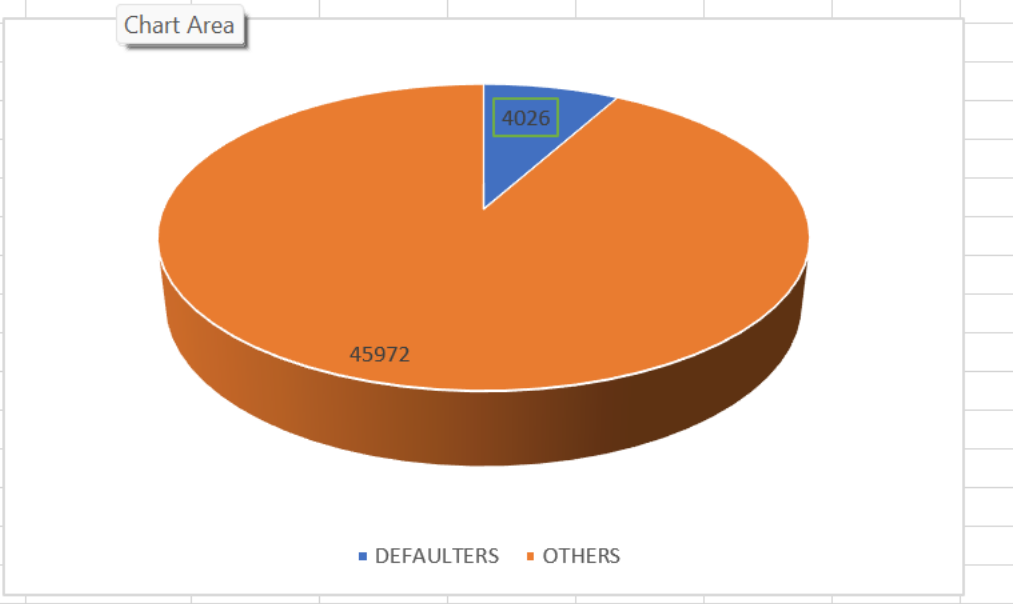
(For data imbalance ratio)

RESULT:

LABELS	COUNT	RATIO
APPROVED	31885	37.11874
CANCELLED	8595	10.00582
REFUSED	8660	10.08149
UNUSED OFFER	859	1

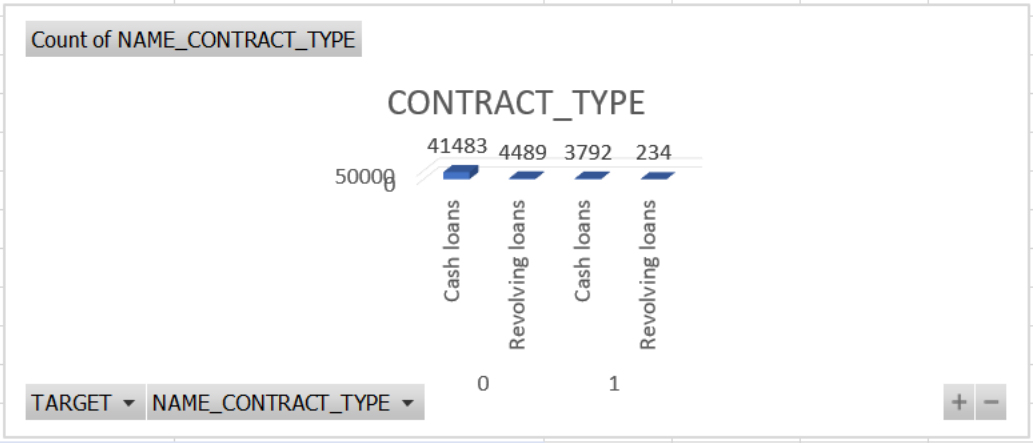


LABELS	COUNT	RATIO
DEFAULTERS	4026	1
OTHERS	45972	11.41878

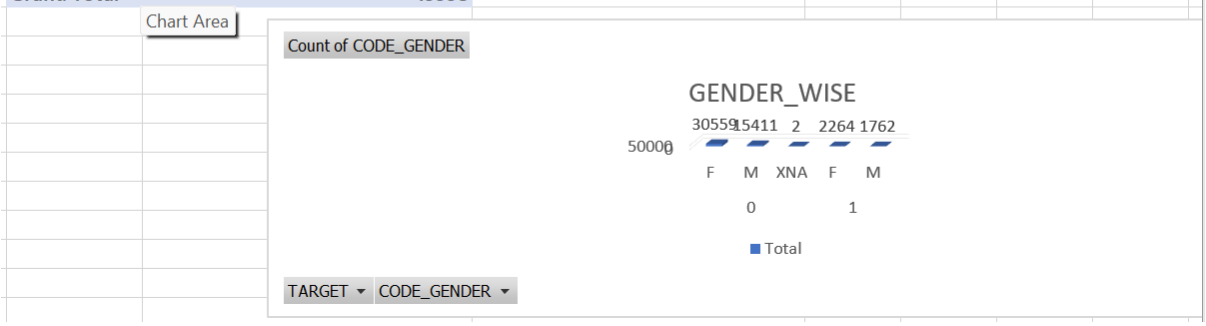


4) Performing Univariate, Segmented Univariate, and Bivariate Anlaysis:
   All three analysis have been done to better understand the data and their relationships.
   To find out the relationship between the target variable and other variables, a pivot table has been created and further charts have been plotted for visualization purposes.
   RESULT:

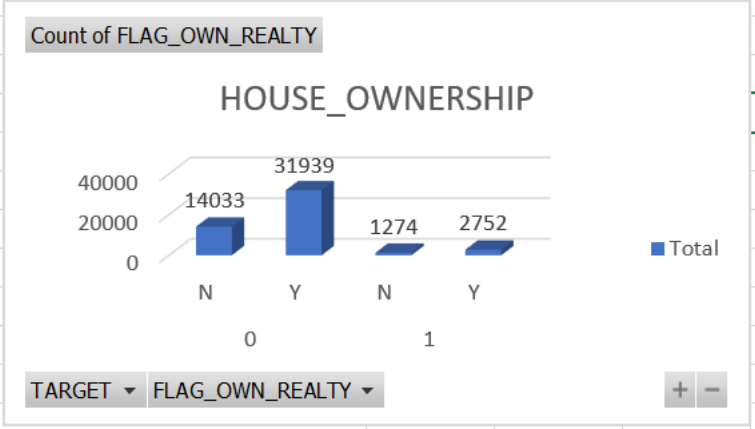
Row Labels	Count of NAME_CONTRACT_TYPE
0	**45972**
Cash loan	41483
Revolving	4489
1	**4026**
Cash loan	3792
Revolving	234
Grand Total	**49998**



Row Labels	Count of CODE_GENDER
0	**45972**
F	30559
M	15411
XNA	2
1	**4026**
F	2264
M	1762
Grand Total	**49998**

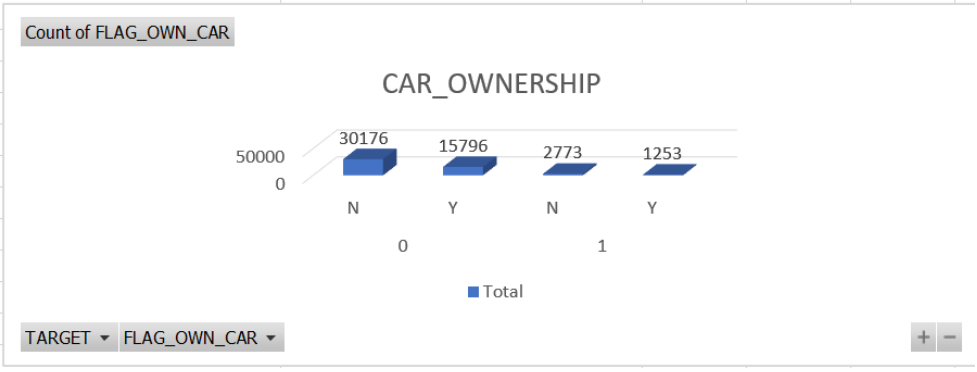


Row Labels	Count of FLAG_OWN_REALTY
0	45972
N	14033
Y	31939
1	4026
N	1274
Y	2752
Grand Total	49998

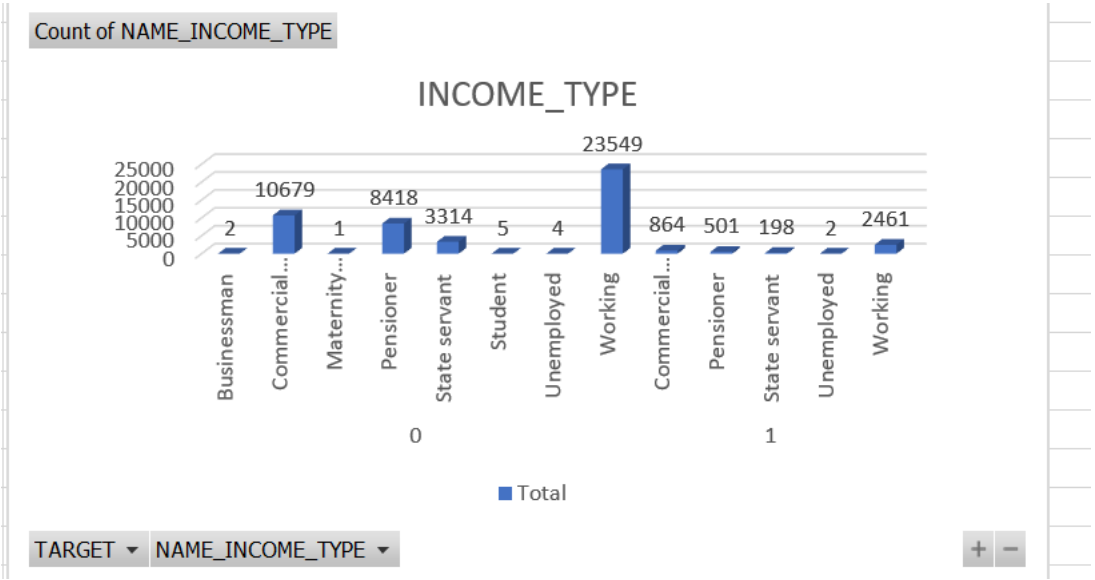


Row Labels	Count of FLAG_OWN_CAR
0	45972
N	30176
Y	15796
1	4026
N	2773
Y	1253
Grand Total	49998

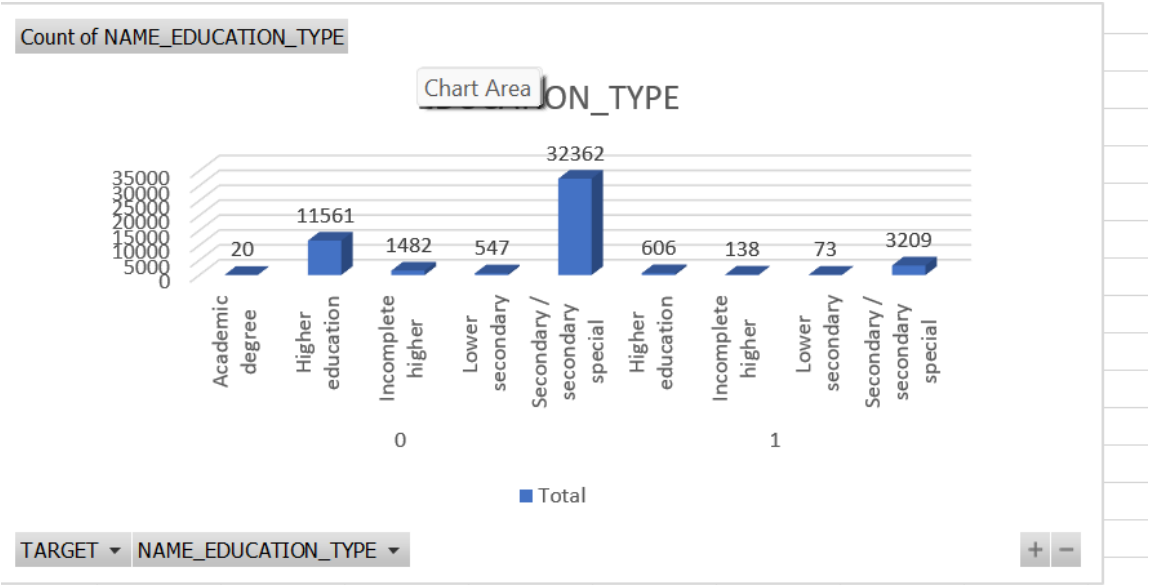
Chart Area



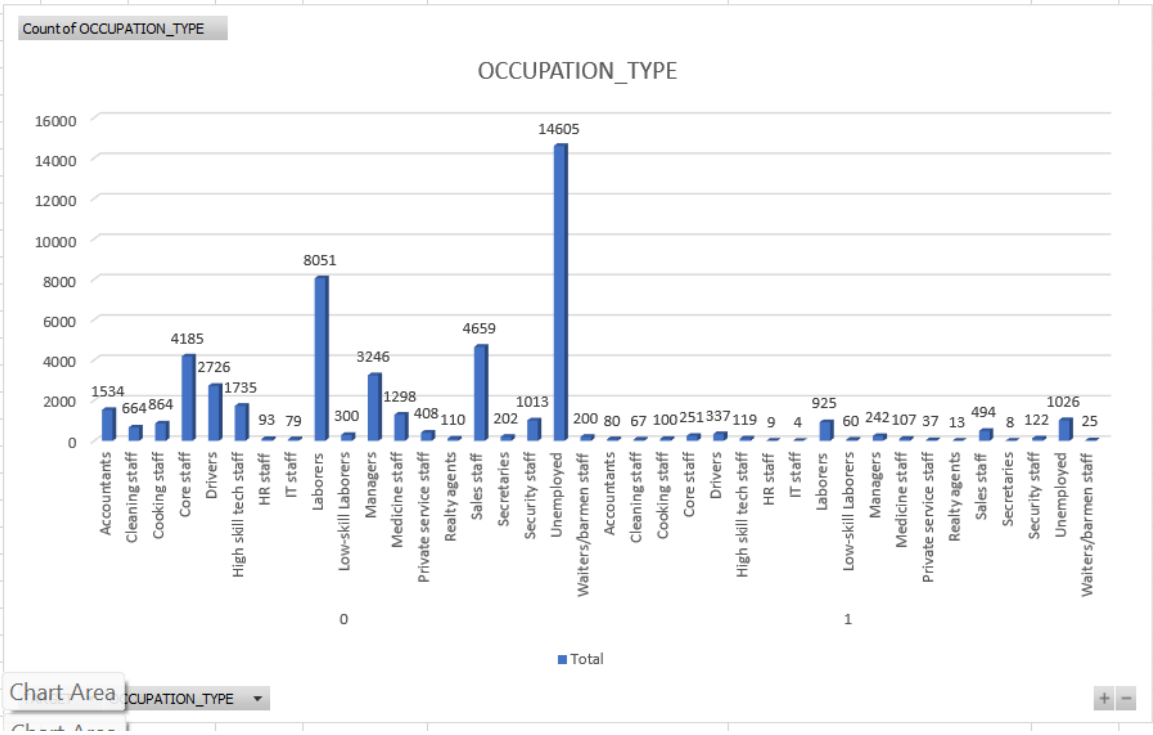
Row Labels	Count of NAME_INCOME_TYP
0	**45972**
Businessn	2
Commerci	10679
Maternity	1
Pensioner	8418
State serv	3314
Student	5
Unemploy	4
Working	23549
1	**4026**
Commerci	864
Pensioner	501
State serv	198
Unemploy	2
Working	2461
Grand Total	**49998**



Row Labels	Count of NAME_EDUCATION_TYPE
0	**45972**
Academic degree	20
Higher education	11561
Incomplete higher	1482
Lower secondary	547
Secondary / secondary special	32362
1	**4026**
Higher education	606
Incomplete higher	138
Lower secondary	73
Secondary / secondary special	3209
Grand Total	**49998**

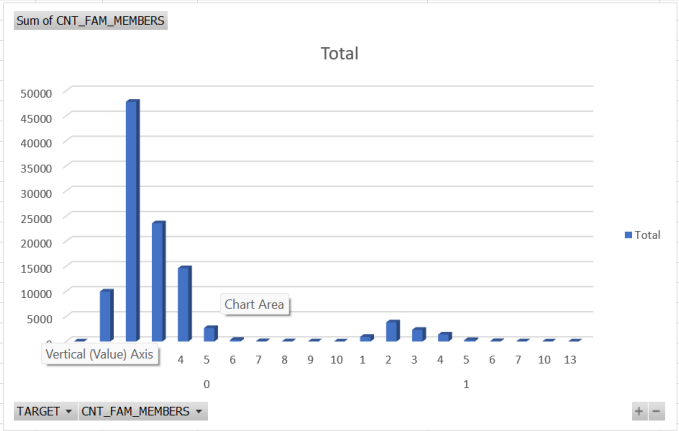


Row Labels	Count of OCCUPATION_TYPE
0	**45972**
Accountar	1534
Cleaning s	664
Cooking s	864
Core staff	4185
Drivers	2726
High skill	1735
HR staff	93
IT staff	79
Laborers	8051
Low-skill l	300
Managers	3246
Medicine	1298
Private se	408
Realty ag	110
Sales staf	4659
Secretarie	202
Security st	1013
Unemploy	14605
Waiters/b	200
1	**4026**
Accountar	80
Cleaning s	67
Cooking s	100
Core staff	251
Drivers	337
High skill	119
HR staff	9
IT staff	4
Laborers	925
Low-skill l	60
Managers	242
Medicine	107
Private se	37
Realty ag	13
Sales staf	494
Secretarie	8
Security st	122
Unemploy	1026
Waiters/b	25
Grand Total	**49998**



Grand Total	49998

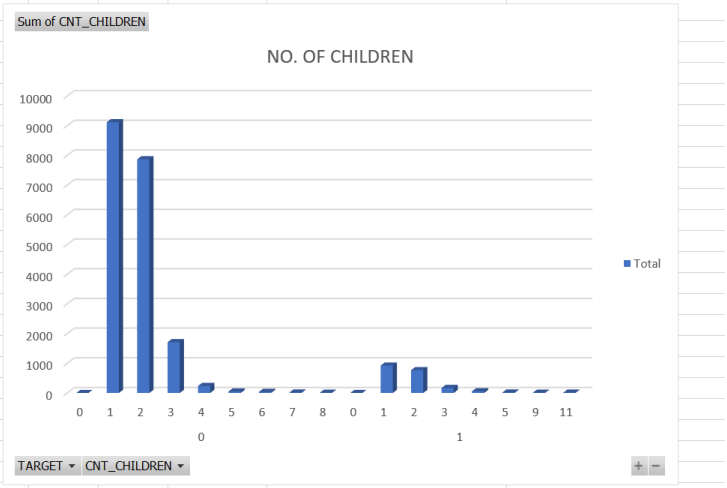
Row Labels	Sum of CNT_FAM_MEMBERS
0	**99088**
0	0
1	9951
2	47800
3	23574
4	14604
5	2690
6	330
7	63
8	48
9	18
10	10
1	**8853**
1	922
2	3812
3	2331
4	1396
5	270
6	78
7	21
10	10
13	13
Grand Total	**107941**



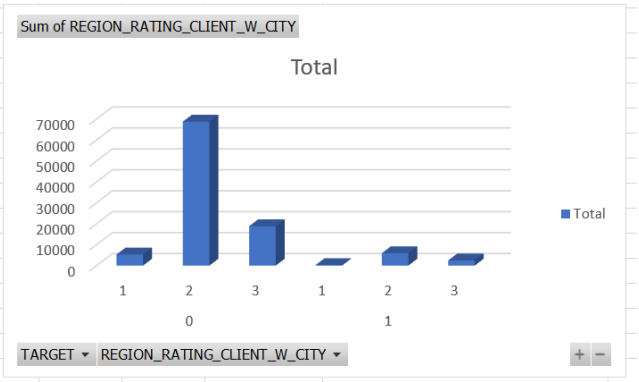
Sum of CNT_CHILDREN

Grand Total	107941

Row Labels	Sum of CNT_CHILDREN
0	**19042**
0	0
1	9118
2	7870
3	1710
4	236
5	50
6	36
7	14
8	8
1	**1950**
0	0
1	923
2	768
3	168
4	56
5	15
9	9
11	11
Grand Total	**20992**



Row Labels	Sum of REGION_RATING_CLIENT_W_CITY
0	**92898**
1	5320
2	68756
3	18822
1	**8634**
1	241
2	5924
3	2469
Grand Total	**101532**



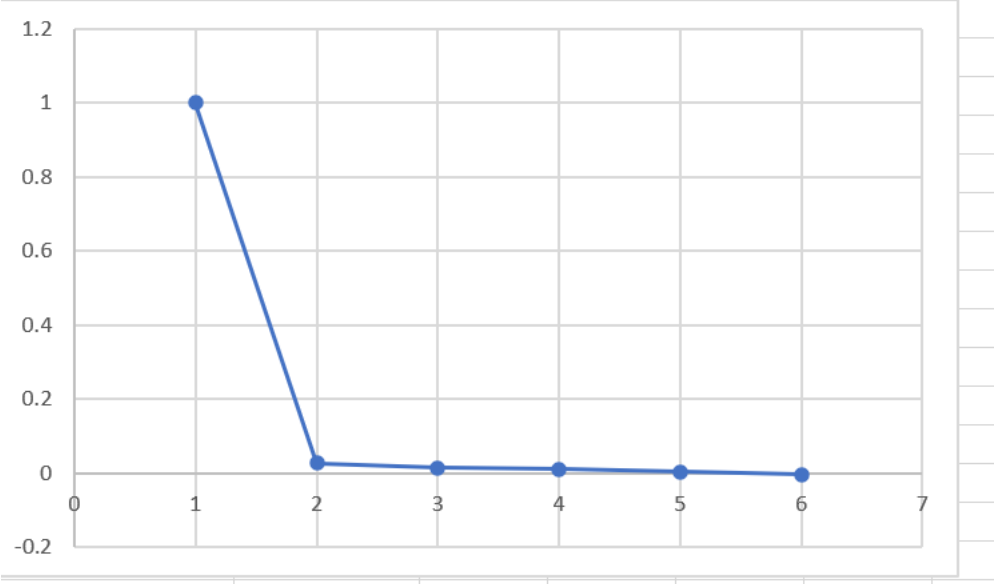
5) Finding Correlations for different scenarios

To find out the correlation of different variables with defaulter or other, the below formula has been used:

RESULT:

LABELS	CORRELATION
GENDER	1
CHILDREN	0.02636061
FAMILY MEMBERS	0.013004707
INCOME	0.010893107
REGION_RATING	0.002171149
AGE	-0.004906967

GENDER	DEFAULTERS	OTHER
MALE	1762	15411
FEMALE	2264	30559



TECH-STACK USED: This project has been done using MS Excel 2019 version due to my familiarity and friendly user interface.

INSIGHTS:

1) In the “application_data”, the columns have been reduced to 52 from 122 and the final total no. of rows is 49998 due to one duplicate being found.
   In the “previous_application”, the columns have been reduced to 27 from 37 and final rows are 50000.
2) In “application_data”,
   - CNT_CHILDREN: No significant outliers were found

- AMT_INCOME_TOTAL: Has a large value of outlier indicating some loan applicants earn high income.
- AMT_CREDIT: Has a large no. of outliers, indicating that some of the loans have been taken for huge amounts.
- AMT_ANNUITY: Has a few outliers.
- AMT_GOODS_PRICE: Has a few outliers.
- YEARS_LAST_PHONE_CHANGE: Outliers here indicate that a few customers haven't changed their phone for long time.
- DAYS_EMPLOYED: Has outliers close to 35000 days which is impossible to have, so wrong entry.
- DAYS_REGISTRATION: Has a large no. of outliers.

In "previous_application":

- AMT_ANNUITY: Huge outlier values.
- AMT_APPLICATION: Huge outlier values.
- AMT_CREDIT: Huge outlier values.
- AMT_GOODS_PRICE: Huge outlier values.
- SELLERPLACE_AREA: Huge outlier values.
- CNT_PAYMENT: Few outlier values.

3) No. of defaulters are 4026 and others are 45972.
"Unused offer" is in the minority class and "approved" is in the majority class.

4)

- 83.7% have defaulted in the case of "cash loans" and 49.5% have defaulted in the case of "revolving loans".
- More females have defaulted as compared to males.
- House Owners have defaulted less than No House Owners.
- Car Owners have defaulted more than No Car Owners.
- "Working" people have defaulted by up to 60%.
- People with "Secondary/Secondary special" level of qualification have defaulted up to 80%.

- “Unemployed” people have defaulted the most up to 25%.
- Applicants with up to 2 family members have default rate of 43%.
- Applicants having 1 child have default rate of up to 50%.
- Applicants rated “2” with respect to their city have default rate up to 70%.

5) “Gender” has the most correlation with target variable.

Result:

This project helped me in upskilling and having a much better understanding of various concepts of Statistics and Advanced MS Excel.

Drive Link:

application_data.xlsm