Name: - VARUN GANDHI (Data Analytics Trainee)
E-Mail :- varun.poseidon84@gmail.com
Contact :- +91 9465433841

Bank Loan Case Study / PROJECT – 6

DESCRIPTION :-

In this project we are going to operate on various excel sheets containing data for loan applications and also previous data records. We are going to focus on what factors influence the approval or refusal towards a loan. Our task is to use exploratory data analysis (EDA) to analyze patterns in the data and ensure that capable applications are not rejected no matter what. This is done to ensure the transparency in the system and also improves the reputation of the bank.

Our major concerns in this project will be:

- Data cleansing: This step involves preprocessing of the dataset to make it suitable for data analysis. Proceeding without cleaning the data will result in inaccurate outcomes. This includes eradicating any missing values, removing duplicates and converting data types if necessary.
- Data Analysis : This step involves analysing various factors that contribute to the approval,

cancellation, refusal or unused offer condition towards a loan application. Our major objective will be to identify the patterns that indicate if a customer will have difficulty paying their installments. This information can be further uses to make decisions on a person's loan application whether to deny, reduce or accept the loan.

PROJECT APPROACH :-

- Our approach will simply consist of performing various operations on the dataset.
- Our first step is to find the discrepancies in the dataset like missing values in rows and columns which can interfere in accuracy.
- Next, we need to perform various operations like identifying missing values and finding appropriate methods to deal with them, detecting and removing outliers and visualize the results through different graphs and charts.

TECH-STACK USED :-

INSIGHTS :-

Identifying Missing data and dealing with it appropriately	It is extremely crucial for a data analyst to handle the missing values to ensure the precision and accuracy of results. Our objective will be to identify the missing data in the dataset and decide on an appropriate method to deal with it.
Identifying Outliers and removing them	Outliers are just like exceptions in a dataset which need to be removed to ensure the accuracy of the data. Outliers can significantly impact the data analysis and also distort the results. Our task will be to detect and identify outliers in the dataset.
Analyse 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. Our task is to determine if there Is data imbalance in the loan application dataset and also calculate the ration of data imbalance using excel functions.
Perform Univariate, Segmented Univariate and Bivariate Analysis	To gain useful insights into the driving factors of the loan default, it is important to conduct various analysis on consumers and loan attributes.

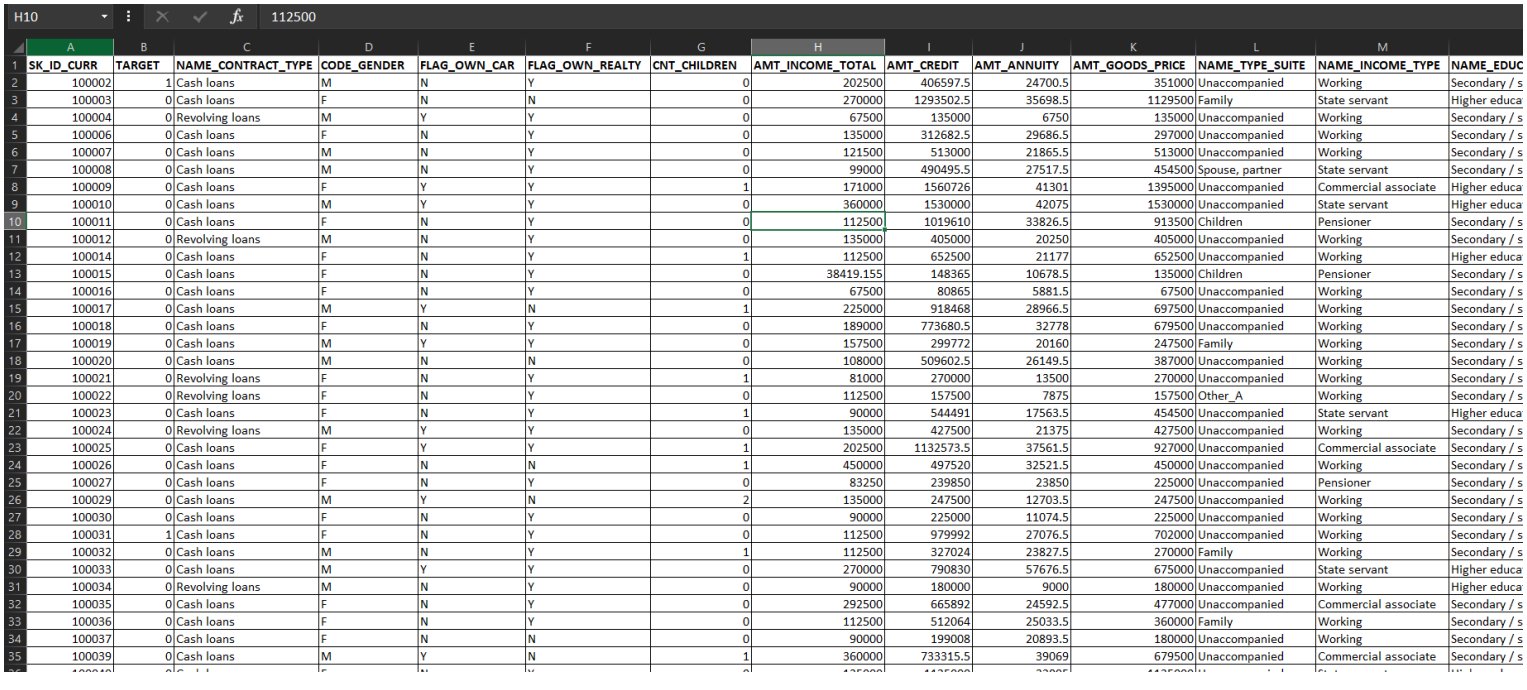
Identify top correlations for different scenarios	Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default. Our task is to segment the dataset based on different scenarios and identifying the top correlations for each segmented data.

ANALYSIS :-

PERFORMING OPERATIONS ON APPLICATION_DATA

a.) Task cleansing :

1. **Firstly we have removed all the duplicate values in correspondence with the 'SK_ID_CURR' column.**
2. **Next, we found out the columns having more than 50% null entries by using the following formula : 100-((B2/50000)*100).**
3. **And we removed the columns which exceeded more than 50% null values.**



H10 112500

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUC
2	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied	Working	Secondary / s
3	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family	State servant	Higher educa
4	100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied	Working	Secondary / s
5	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied	Working	Secondary / s
6	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied	Working	Secondary / s
7	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner	State servant	Secondary / s
8	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate	Higher educa
9	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied	State servant	Higher educa
10	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary / s
11	100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied	Working	Secondary / s
12	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied	Working	Higher educa
13	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children	Pensioner	Secondary / s
14	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied	Working	Secondary / s
15	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied	Working	Secondary / s
16	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied	Working	Secondary / s
17	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary / s
18	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / s
19	100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied	Working	Secondary / s
20	100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary / s
21	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher educa
22	100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied	Working	Secondary / s
23	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied	Commercial associate	Secondary / s
24	100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied	Working	Secondary / s
25	100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied	Pensioner	Secondary / s
26	100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied	Working	Secondary / s
27	100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied	Working	Secondary / s
28	100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompanied	Working	Secondary / s
29	100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family	Working	Secondary / s
30	100033	0	Cash loans	M	Y	Y	0	270000	790830	57676.5	675000	Unaccompanied	State servant	Higher educa
31	100034	0	Revolving loans	M	N	Y	0	90000	180000	9000	180000	Unaccompanied	Working	Higher educa
32	100035	0	Cash loans	F	N	Y	0	292500	665892	24592.5	477000	Unaccompanied	Commercial associate	Secondary / s
33	100036	0	Cash loans	F	N	Y	0	112500	512064	25033.5	360000	Family	Working	Secondary / s
34	100037	0	Cash loans	F	N	N	0	90000	199008	20893.5	180000	Unaccompanied	Working	Secondary / s
35	100039	0	Cash loans	M	Y	N	1	360000	733315.5	39069	679500	Unaccompanied	Commercial associate	Secondary / s

b.) Outlier detection/identification :

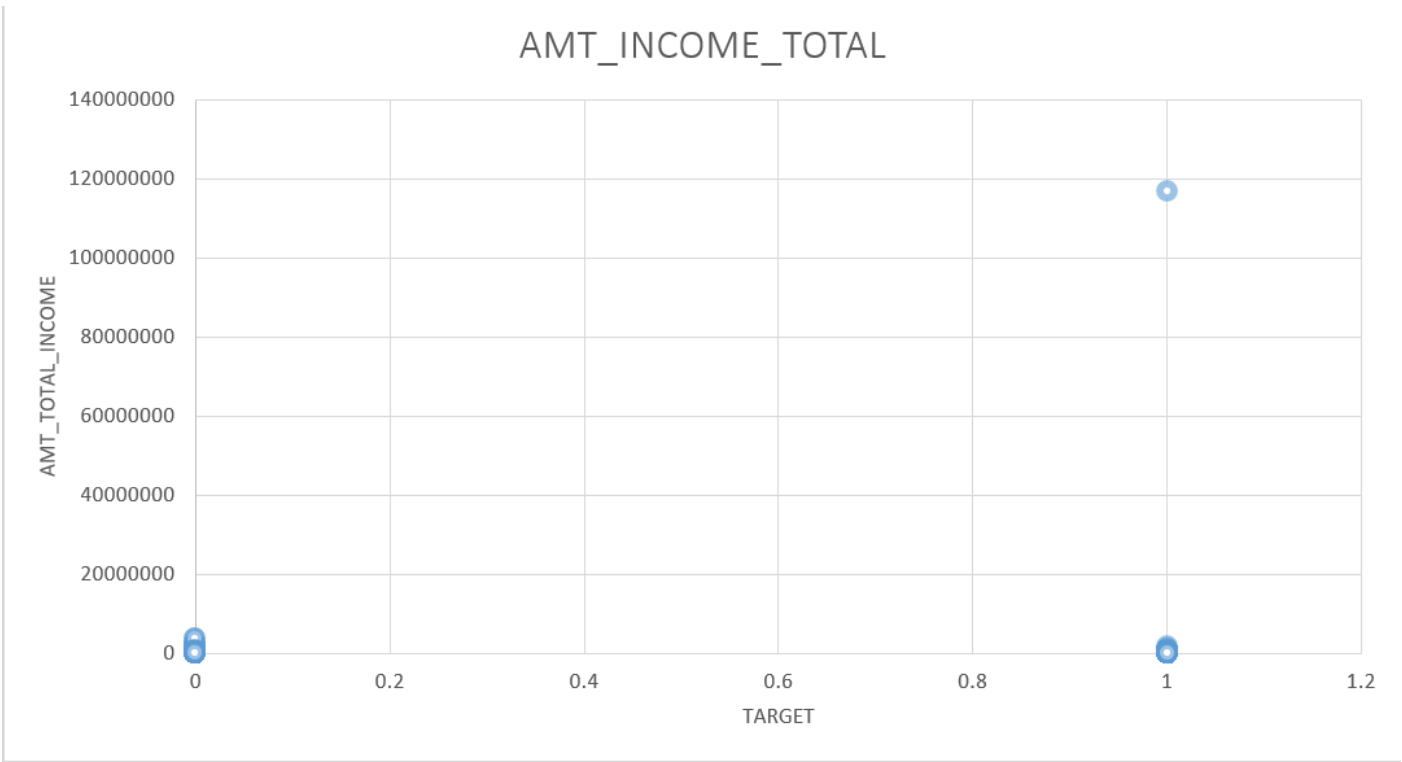
1. **Our approach is to calculate the 1st quartile, 3rd quartile, interquartile range (IQR), upper limit and the lower limit.**
2. **Next, we create a scatter plot to visualize the outlier using target and amount_total_income column.**
3. **And creating scatter plot to find the outlier using the other two columns namely target and CNT_children.**

1st quartile formula : =QUARTILE.INC(C:C, 1)
2nd quartile formula : =QUARTILE.INC(C:C, 3)

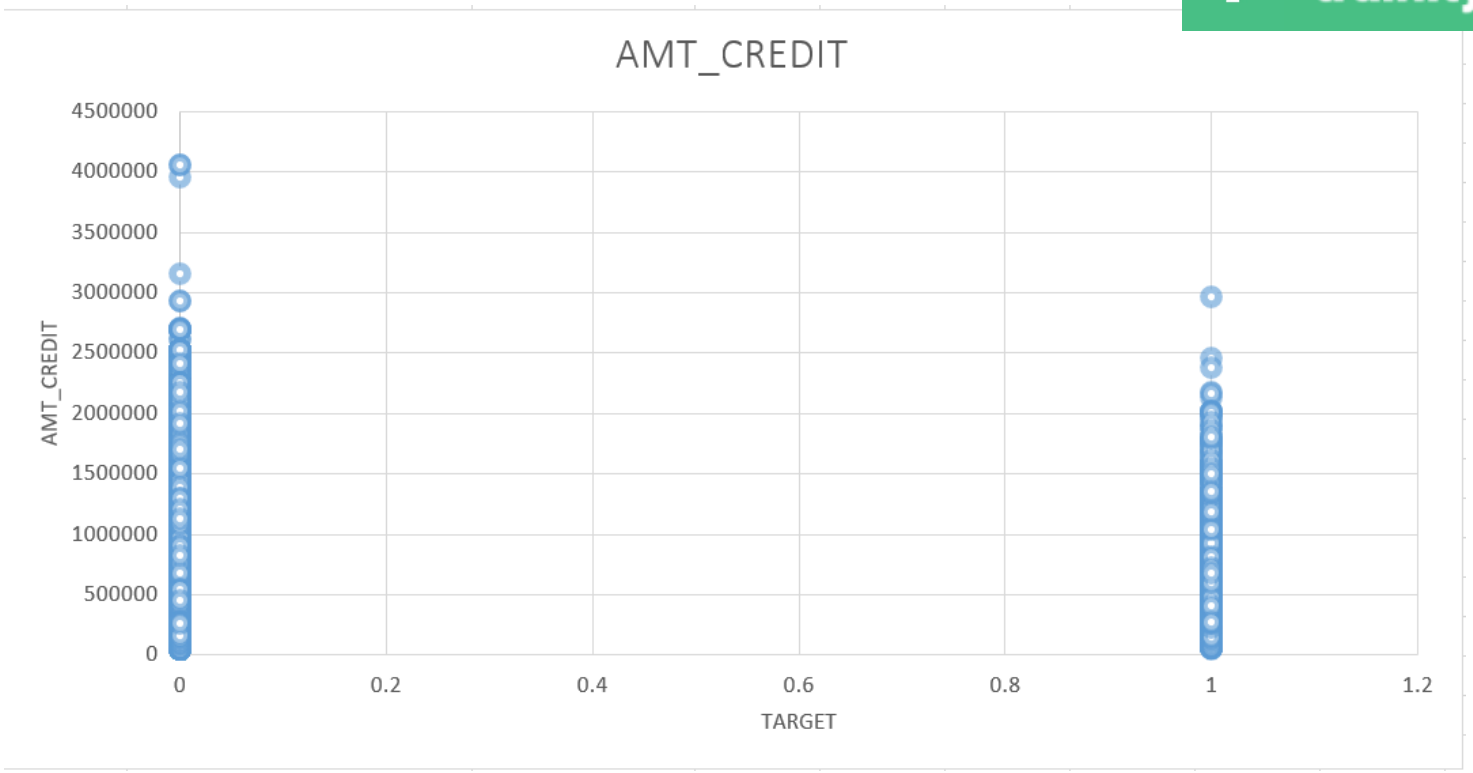
Interquartile Range formula : =K3-K2

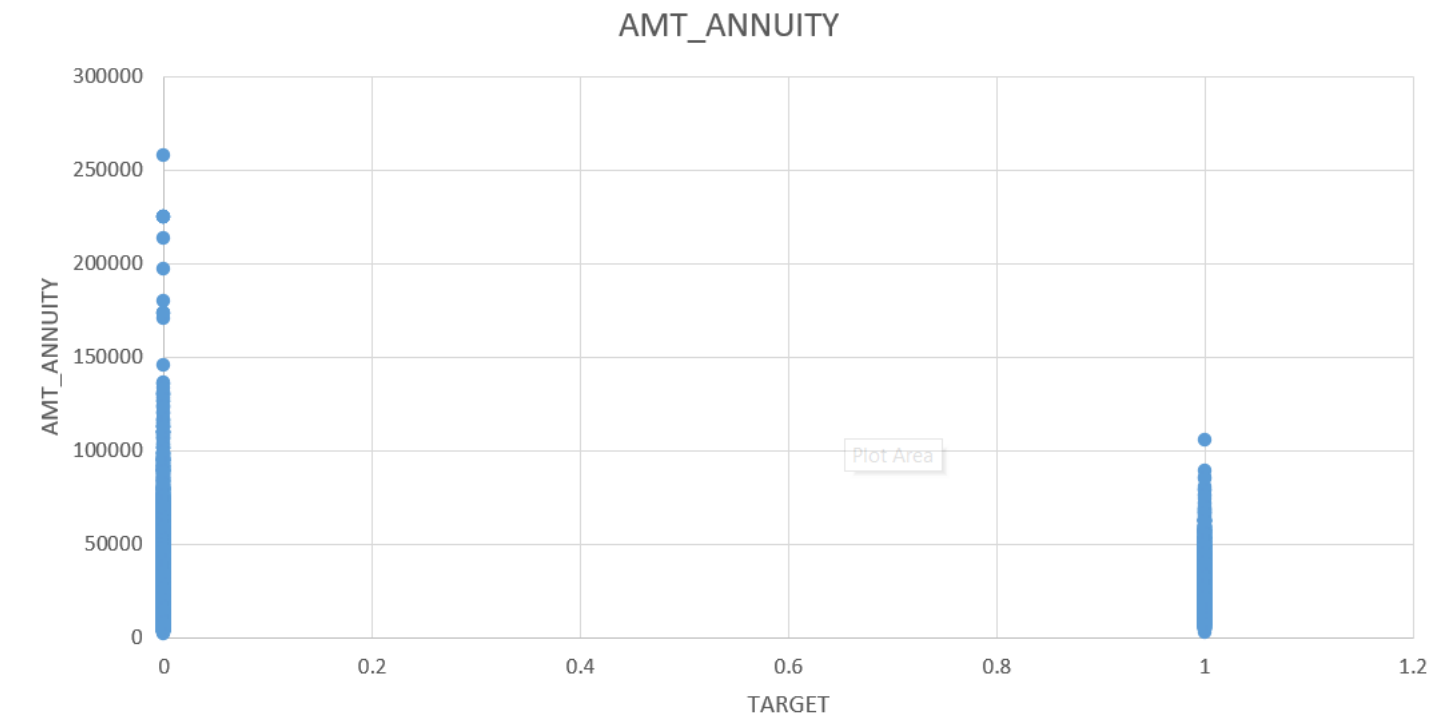
Upper Limit : =K3+(1.5*K4)
Lower Limit : =K2-(1.5*K4)

1st QUARTILE :-	112500
3rd QUARTILE :-	202500
Inter Quartile Range (IQR) :-	90000
Upper Limit :-	337500
Lower Limit :-	-22500



trainity

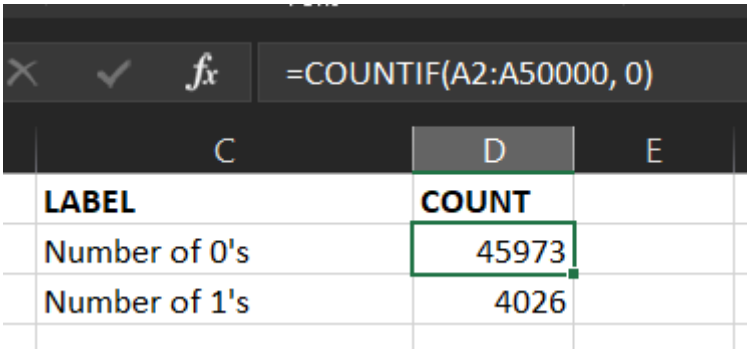

AMT_CREDIT
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
AMT_CREDIT
0
0.2
0.4
0.6
0.8
1
1.2
TARGET

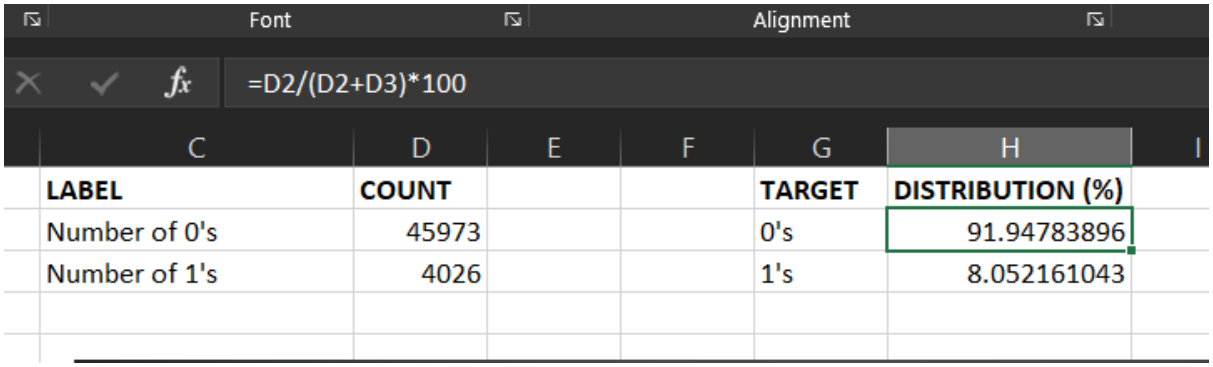

AMT_ANNUITY
300000
250000
200000
150000
100000
50000
0
AMT_ANNUITY
Plot Area
0
0.2
0.4
0.6
0.8
1
1.2
TARGET

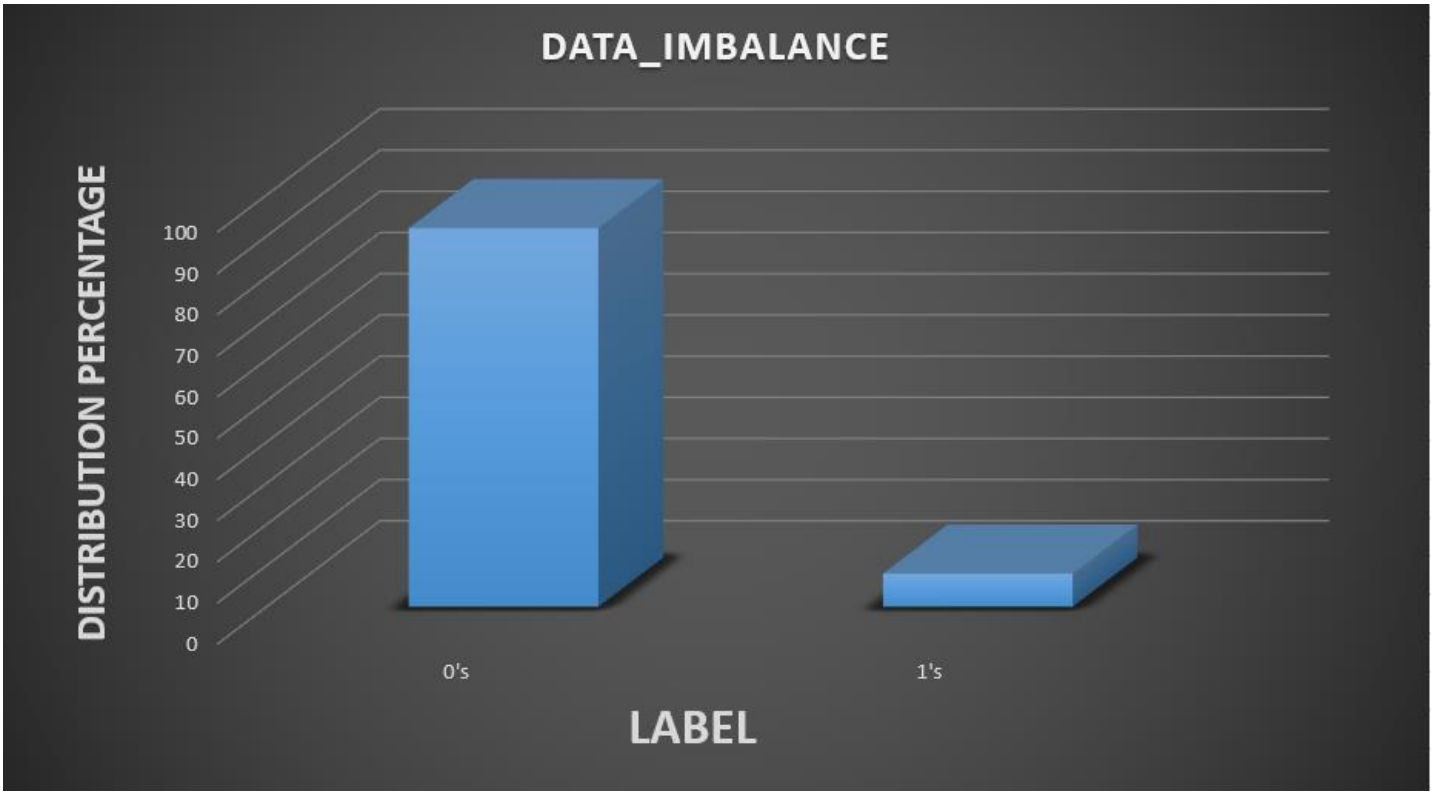
c.) **Analysing Data Imbalance :**

1.) **Our major approach here will be to count the number of 0's and 1's present in the 'target' column.**

2.) **Next, we will plot a pivot/graph to visualize the distribution of imbalanced data.**

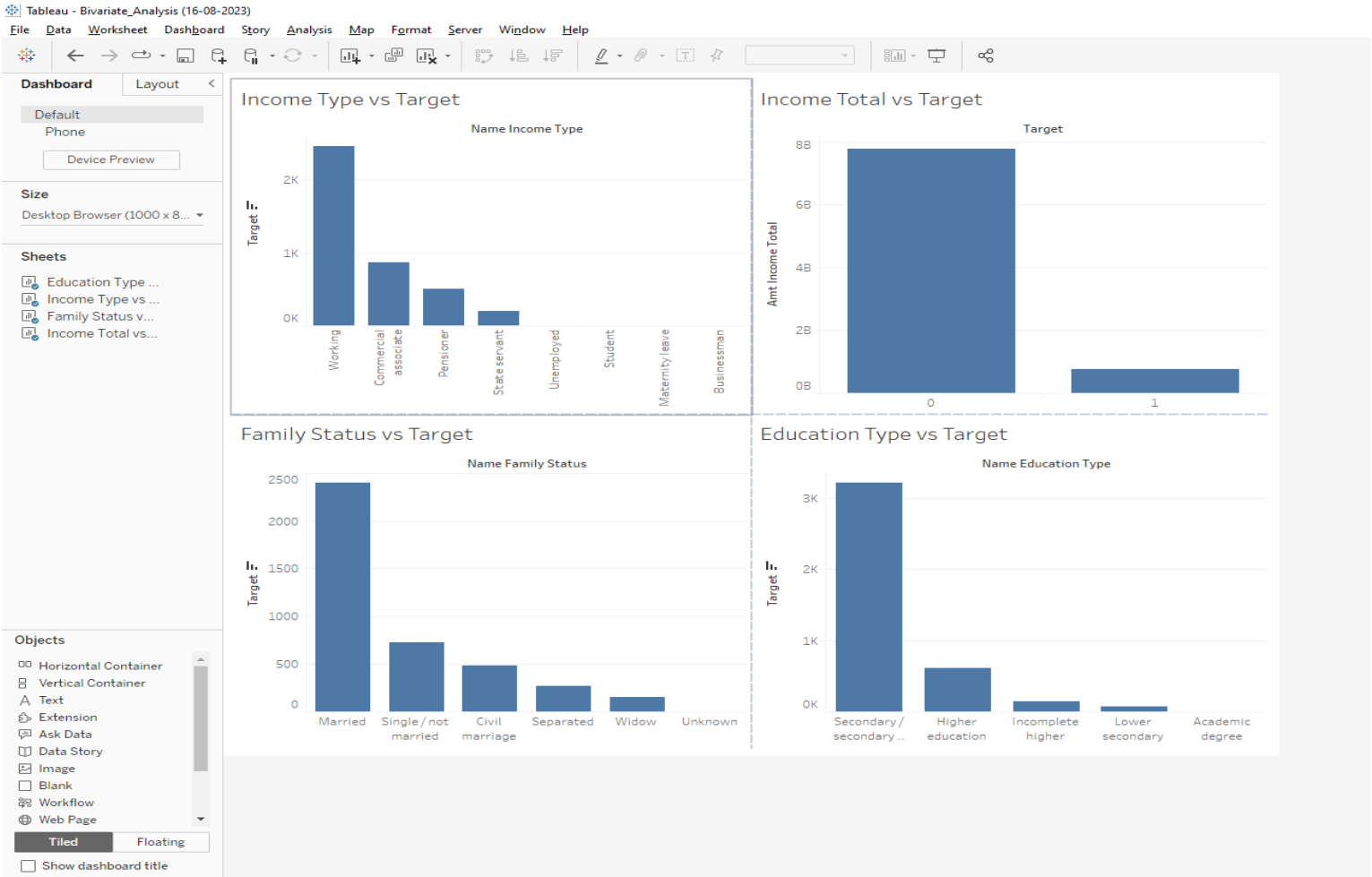




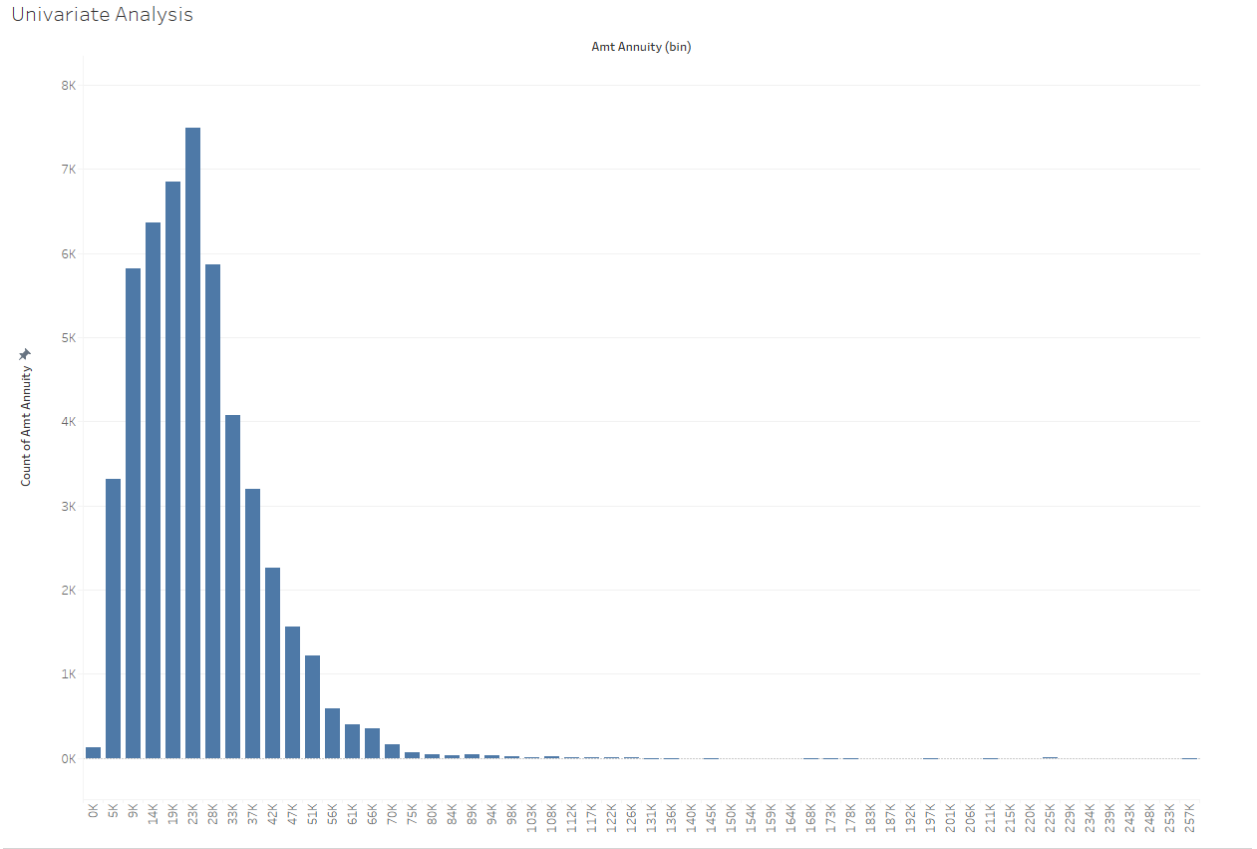


d.) Perform Univariate, Segmented Univariate and Bivariate Analysis **:**

1. Bivariate Analysis : As the name suggests, this refers to a relationship between any two variables present in the dataset. The major goal of bivariate analysis is to understand how changes in one of the variables is directly/indirectly associated with changes in another variable. We have used 'TABLEAU' for this process to make the data analysis easier. This technique further provides us with the basic trends and patterns between the two variables.
   Various relationships we have created between variables are :-
   - Education Type VS Target
   - Income Type VS Target
   - Family Status VS Target
   - Total Income VS Target



2. Univariate Analysis :- This is a statistical method which is used to analyse the distribution and characteristics of a single variable in a dataset. This involves calculating the data's different components such as mean, median, mode, range, variance, etc. We have used 'TABLEAU' for this process too to make the data analysis easier.
    - Moving on to univariate analysis : We will plot count of Amt Annuity as rows.
    - Next, we will convert a copy of Amt Annuity to bins with a bin size of '4680'.
    - Now plot the Amt Annuity (Bin) onto the column bar.



For detailed analysis, please refer to the univariate analysis workbook :- Univariate Analysis

e.) Identification of top correlations :- Major factors that influence the payments made on time are CNT_CHIILDREN, AMT_INCOME_TOTAL, AMT_CREDIT, AMT_ANNNUITY, etc.

- The approach we will follow for this is to compare the correlations between different variables with a CORREL chart.
- The chart with the rounded off values that define the correlation between the major variables :

ROUNDED OFF CORRELATION

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH
CNT_CHILDREN	1	0.01	0	0.03	0	-0.03	0.33	-0.24	0.18	-0.03
AMT_INCOME_TOTAL	0.01	1	0.07	0.08	0.07	0.03	0.02	-0.03	0.01	0
AMT_CREDIT	0	0.07	1	0.77	0.99	0.1	-0.06	-0.07	0	-0.01
AMT_ANNUITY	0.03	0.08	0.77	1	0.77	0.12	0.01	-0.11	0.03	0.01
AMT_GOODS_PRICE	0	0.07	0.99	0.77	1	0.1	-0.06	-0.07	0.01	-0.01
REGION_POPULATION_RELATIVE	-0.03	0.03	0.1	0.12	0.1	1	-0.03	0	-0.06	0
DAYS_BIRTH	0.33	0.02	-0.06	0.01	-0.06	-0.03	1	-0.61	0.33	0.27
DAYS_EMPLOYED	-0.24	-0.03	-0.07	-0.11	-0.07	0	-0.61	1	-0.2	-0.27
DAYS_REGISTRATION	0.18	0.01	0	0.03	0.01	-0.06	0.33	-0.2	1	0.1
DAYS_ID_PUBLISH	-0.03	0	-0.01	0.01	-0.01	0	0.27	-0.27	0.1	1

Link to the final excel file : Application_data :- [Application_data_excelsheet]

Now, let's move on to our net analysis which is based on the next dataset named **previous_data** which contains the details of the previous applications of the customers :-

a.) Identification of missing or inappropriate data :-

- So, first of all we need to identify the unnecessary columns having too many missing values and drop them.

  Unnecessary columns :-

  AMT_DOWN_PAYMENT
  WEEKDAY_APPR_PROCESS_START
  HOUR_APPR_PROCESS_START
  FLAG_LAST_APPL_PER_CONTRACT
  NFLAG_LAST_APPL_IN_DAY
  RATE_DOWN_PAYMENT
  RATE_INTEREST_PRIMARY
  RATE_INTEREST_PRIVILEGED
  SELLERPLACE_AREA
  NAME_SELLER_INDUSTRY
  DAYS_FIRST_DUE
  DAYS_LAST_DUE_1ST_VERSION
  DAYS_LAST_DUE
  DAYS_TERMINATION
  NFLAG_INSURED_ON_APPROVAL

- Next, we will replace all the missing values present in AMT_ANNUITY, AMT_GOODS_PRICE and NAME_TYPE_SUITE column with the following values :-

AMT_ANNUITY MEAN :	15482.59685
AMT_GOODS_PRICE MEDIAN :	104017.5
NAME_TYPE_SUITE MODE :	Unaccompanied

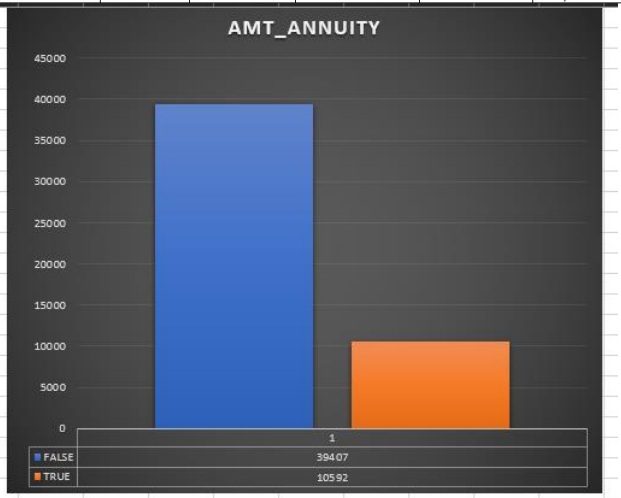
- Now, to fix the days column we need to multiply the values with -1 to convert the negative days into positive values.
- Now, that we have cleaned the dataset and also replaced the necessary values. Let's proceed with further analysis:-

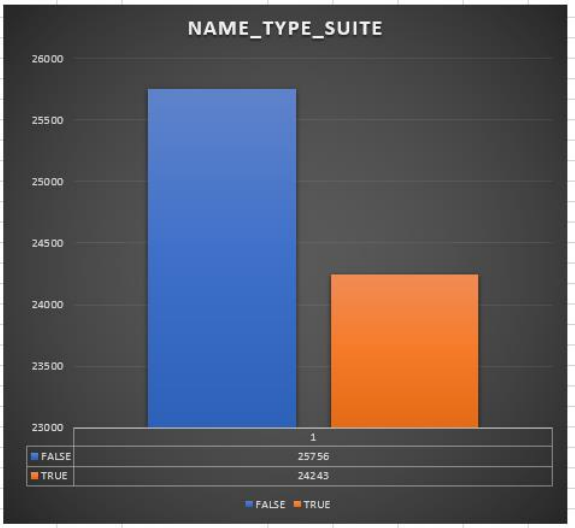
SK_ID_PREV	SK_ID_CURR	NAME_CO	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	NAME_CASH_LOAN_PURPOSE	NAME_CONTRACT_STATUS	DAYS_DECISION	DAYS_DECISION (+ve)	NAME_PAYMENT_TYPE	CODE_REJECT_REASON	NAME_TYPE_SUITE	NAME_CLIENT_TYPE	NAME_GOO
1000001	158271	Consume	6404.31	58905	65124	58905	XAP	Approved	-299	299	Cash through the bank	XAP	Unaccompanied	New	Furniture
1000009	343078	Consume	9302.85	42705	45243	42705	XAP	Approved	-488	488	Cash through the bank	XAP	Unaccompanied	Repeater	Consumer
1000016	157990	Consume	6078.15	63720	56970	63720	XAP	Approved	-717	717	Cash through the bank	XAP	Unaccompanied	Refreshed	Audio/Vide
1000055	238221	Consume	3839.4	26046	27949.5	26046	XAP	Approved	-1194	1194	Cash through the bank	XAP	Family	Repeater	Mobile
1000081	417435	Cash loan	22021.47	315000	340573.5	315000	XNA	Approved	-419	419	Cash through the bank	XAP	Family	Refreshed	XNA
1000110	390955	Consume	24885.36	379350	379350	379350	XAP	Approved	-1739	1739	XNA	XAP	Unaccompanied	New	Furniture
1000140	148166	Cash loan	22312.08	454500	536994	454500	XNA	Approved	-389	389	Cash through the bank	XAP	Unaccompanied	Repeater	XNA
1000163	331428	Consume	15775.515	268200	302787	268200	XAP	Approved	-349	349	Cash through the bank	XAP	Unaccompanied	New	Education
1000187	311379	Revolving	6750	135000	135000	135000	XAP	Approved	-316	316	XNA	XAP	Unaccompanied	Repeater	XNA
1000239	215236	Cash loan	34080.975	922500	1056447	922500	XNA	Approved	-501	501	Cash through the bank	XAP	Unaccompanied	Repeater	XNA
1000297	255234	Consume	3365.145	15975	16767	15975	XAP	Approved	-2660	2660	XNA	XAP	Unaccompanied	New	Mobile
1000338	277319	Consume	8805.375	195241.5	195241.5	195241.5	XAP	Approved	-1616	1616	Cash through the bank	XAP	Spouse, partner	Repeater	Audio/Vide
1000348	393339	Consume	16447.185	125302.5	140661	125302.5	XAP	Approved	-1842	1842	Cash through the bank	XAP	Unaccompanied	Repeater	Computers
1000371	364510	Consume	6093.81	47646	51840	47646	XAP	Approved	-414	414	Cash through the bank	XAP	Family	Repeater	Consumer
1000392	212548	Consume	14495.13	55750.5	55750.5	55750.5	XAP	Approved	-1558	1558	Cash through the bank	XAP	Spouse, partner	Repeater	Constructio
1000415	167272	Consume	4355.01	24111	21699	24111	XAP	Approved	-2649	2649	Cash through the bank	XAP	Unaccompanied	New	Mobile
1000534	383039	Cash loan	4594.005	58500	58500	58500	XNA	Approved	-322	322	Cash through the bank	XAP	Spouse, partner	Repeater	XNA
1000588	168814	Cash loan	31396.5	450000	450000	450000	XNA	Approved	-407	407	Cash through the bank	XAP	Unaccompanied	Repeater	XNA
1000609	425594	Consume	2282.4	15480	14391	15480	XAP	Approved	-2304	2304	Cash through the bank	XAP	Unaccompanied	Repeater	Mobile
1000614	372818	Consume	6712.38	58455	59296.5	58455	XAP	Approved	-1959	1959	Cash through the bank	XAP	Unaccompanied	New	Computers
1000664	393058	Consume	5930.685	128749.5	128749.5	128749.5	XAP	Approved	-753	753	Cash through the bank	XAP	Family	Repeater	Audio/Vide
1000789	233175	Consume	11315.88	113170.5	101853	113170.5	XAP	Approved	-2774	2774	Cash through the bank	XAP	Family	Repeater	Audio/Vide
1000849	363834	Revolving	9000	0	45000	104017.5	XAP	Approved	-2650	2650	XNA	XAP	Unaccompanied	Repeater	XNA
1000931	161659	Cash loan	43220.925	855000	954315	855000	XNA	Approved	-580	580	Cash through the bank	XAP	Unaccompanied	Refreshed	XNA
1000947	246513	Consume	9261.09	53955	48325.5	53955	XAP	Approved	-812	812	Cash through the bank	XAP	Family	Refreshed	Computers
1000963	155559	Consume	9007.29	81810	81810	81810	XAP	Approved	-1108	1108	Cash through the bank	XAP	Family	New	Clothing a
1001137	284758	Consume	5653.26	59490	59854.5	59490	XAP	Approved	-286	286	Cash through the bank	XAP	Unaccompanied	Repeater	Furniture
1001156	320745	Consume	7516.62	20695.5	21240	20695.5	XAP	Approved	-620	620	Cash through the bank	XAP	Family	Repeater	Consumer
1001173	312996	Cash loan	29034	900000	900000	900000	XNA	Approved	-1616	1616	Cash through the bank	XAP	Unaccompanied	Repeater	XNA
1001220	309757	Revolving	2250	45000	45000	45000	XAP	Approved	-326	326	XNA	XAP	Unaccompanied	Repeater	XNA
1001226	228787	Consume	8783.595	76455	77593.5	76455	XAP	Approved	-1711	1711	Cash through the bank	XAP	Family	Repeater	Audio/Vide
1001318	239228	Cash loan	16206.3	247500	274941	247500	XNA	Approved	-1750	1750	Cash through the bank	XAP	Unaccompanied	Repeater	XNA
1001336	256064	Consume	11523.465	86715	84483	86715	XAP	Approved	-1904	1904	Cash through the bank	XAP	Unaccompanied	New	Mobile
1001375	315080	Consume	6832.035	41215.5	23980.5	41215.5	XAP	Approved	-2919	2919	Cash through the bank	XAP	Family	New	Mobile
1001393	380192	Consume	5524.695	43105.5	47659.5	43105.5	XAP	Approved	-352	352	Cash through the bank	XAP	Unaccompanied	Repeater	Mobile
1001437	284165	Consume	4593.6	44010	39600	44010	XAP	Approved	-1268	1268	Cash through the bank	XAP	Unaccompanied	Repeater	Mobile
1001438	212386	Consume	9446.625	101250	101250	101250	XAP	Approved	-775	775	Cash through the bank	XAP	Family	Repeater	Clothing a
1001439	352767	Consume	2662.155	12555	13302	12555	XAP	Approved	-320	320	Cash through the bank	XAP	Unaccompanied	Refreshed	Mobile
1001452	321268	Consume	4463.28	24520.5	24520.5	24520.5	XAP	Approved	-1163	1163	Cash through the bank	XAP	Unaccompanied	Repeater	Auto Acces
1001514	289780	Cash loan	63011.205	900000	939204	900000	XNA	Approved	-805	805	Cash through the bank	XAP	Family	Repeater	XNA
1001527	196819	Cash loan	10492.02	90000	95940	90000	XNA	Approved	-2792	2792	Cash through the bank	XAP	Spouse, partner	Repeater	XNA
1001589	215935	Consume	8991.045	97447.5	87700.5	97447.5	XAP	Approved	-377	377	Cash through the bank	XAP	Unaccompanied	Repeater	Furniture
1001603	335361	Cash loan	19404.585	180000	243945	180000	Payments on other loans	Approved	-702	702	Cash through the bank	XAP	Unaccompanied	New	XNA
1001620	199177	Consume	15875.01	291600	349335	291600	XAP	Approved	-324	324	Cash through the bank	XAP	Unaccompanied	Repeater	Clothing a
1001703	334347	Consume	53187.03	526500	526500	526500	XAP	Approved	-407	407	Cash through the bank	XAP	Unaccompanied	Repeater	Clothing a
1001779	446537	Consume	16676.01	83016	87399	83016	XAP	Approved	-999	999	Cashless from the accour	XAP	Family	New	Audio/Vide

AMT_ANNUITY	
FALSE	39407
TRUE	10592

AMT_GOODS_PRICE	
FALSE	39254
TRUE	10745

NAME_TYPE_SUITE	
FALSE	25756
TRUE	24243





b.) Detecting Outliers in the dataset :

AMT_ANNUITY	Value
1st Quartile	7189.74
3rd Quartile	16256.16
Inter Quartile Range	9066.42
Upper Limit	29855.79
Lower Limit	0

AMT_APPLICATION	Value
1st Quartile	22045.5
3rd Quartile	180000
Inter Quartile Range	157954.5
Upper Limit	416931.75
Lower Limit	0

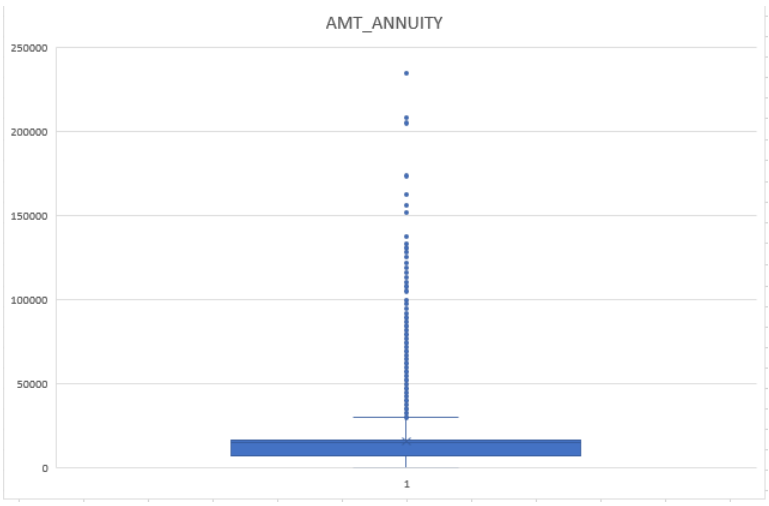
AMT_CREDIT	Value
1st Quartile	26055
3rd Quartile	198105.75
Inter Quartile Range	172050.75
Upper Limit	456181.875
Lower Limit	0

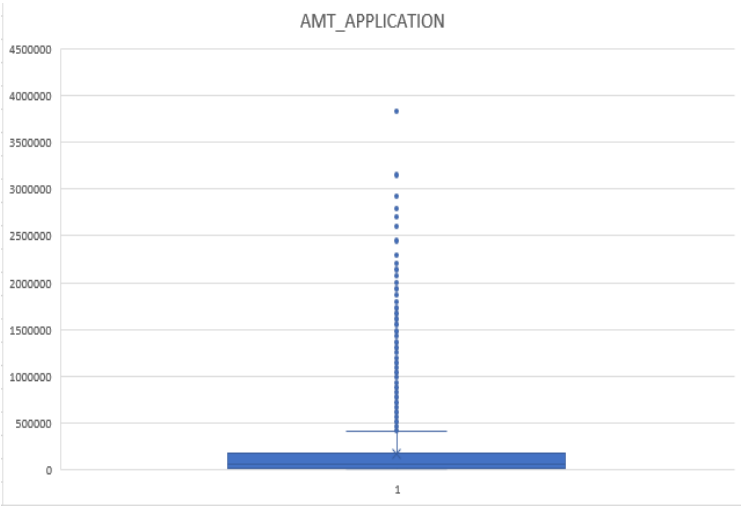
AMT_GOODS_PRICE	Value
1st Quartile	63663.75
3rd Quartile	180000
Inter Quartile Range	116336.25
Upper Limit	354504.375
Lower Limit	0

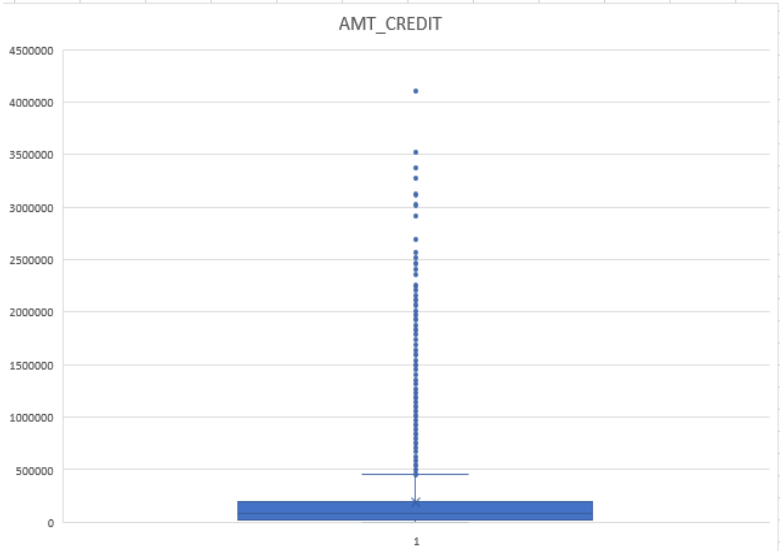
CNT_PAYMENT	Value
1st Quartile	0
3rd Quartile	14
Inter Quartile Range	14
Upper Limit	35
Lower Limit	0

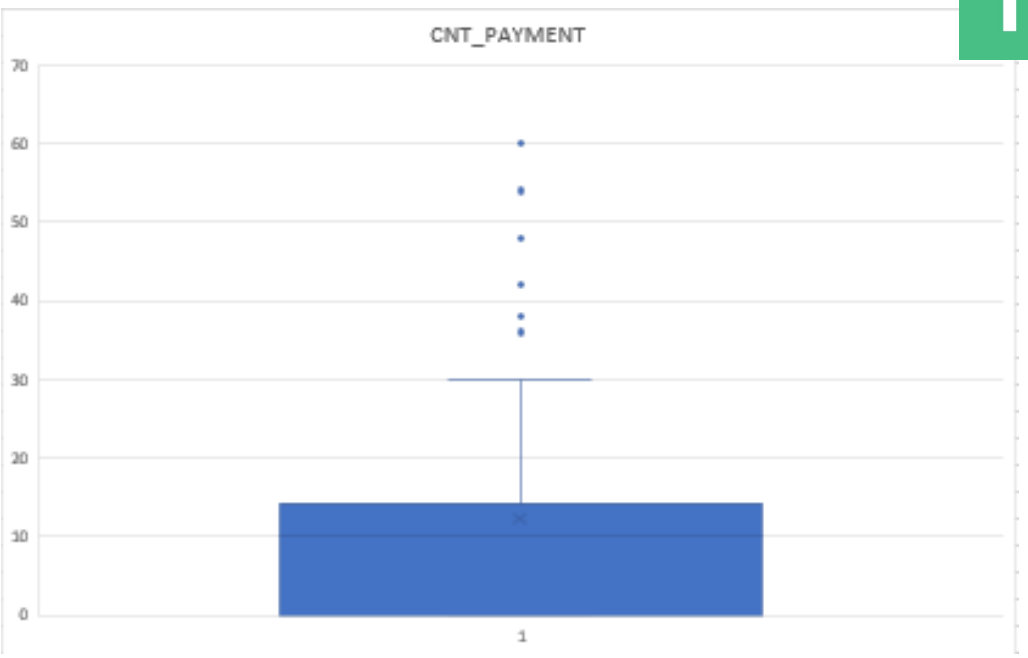
Number of Outliers in AMT_ANNUITY	Number of Outliers in AMT_APPLICATION	Number of Outliers in AMT_CREDIT	Number of Outliers in AMT_GOODS_PRICE	Number of Outliers in CNT_PAYMENT
4922	5792	5648	6569	4968

AMT_ANNUITY	Outlier (False or True)	AMT_APPLICATION	Outlier (False or True)	AMT_CREDIT	Outlier (False or True)	AMT_GOODS_PRICE	Outlier (False or True)	CNT_PAYMENT	Outlier (False or True)
6404.31	FALSE	58905	FALSE	65124	FALSE	58905	FALSE	12	FALSE
9302.85	FALSE	42705	FALSE	45243	FALSE	42705	FALSE	6	FALSE
6078.15	FALSE	63720	FALSE	56970	FALSE	63720	FALSE	12	FALSE
3839.4	FALSE	26046	FALSE	27949.5	FALSE	26046	FALSE	10	FALSE
22021.47	FALSE	315000	FALSE	340573.5	FALSE	315000	FALSE	18	FALSE
24885.36	FALSE	379350	FALSE	379350	FALSE	379350	TRUE	18	FALSE
22312.08	FALSE	454500	TRUE	536994	TRUE	454500	TRUE	42	TRUE
15775.515	FALSE	268200	FALSE	302787	FALSE	268200	FALSE	24	FALSE
6750	FALSE	135000	FALSE	135000	FALSE	135000	FALSE	0	FALSE
34080.975	TRUE	922500	TRUE	1056447	TRUE	922500	TRUE	60	TRUE
3365.145	FALSE	15975	FALSE	16767	FALSE	15975	FALSE	6	FALSE
8805.375	FALSE	195241.5	FALSE	195241.5	FALSE	195241.5	FALSE	24	FALSE
16447.185	FALSE	125302.5	FALSE	140661	FALSE	125302.5	FALSE	12	FALSE
6093.81	FALSE	47646	FALSE	51840	FALSE	47646	FALSE	10	FALSE
14495.13	FALSE	55750.5	FALSE	55750.5	FALSE	55750.5	FALSE	4	FALSE
4355.01	FALSE	24111	FALSE	21699	FALSE	24111	FALSE	6	FALSE
4594.005	FALSE	58500	FALSE	58500	FALSE	58500	FALSE	18	FALSE
31396.5	TRUE	450000	TRUE	450000	FALSE	450000	TRUE	24	FALSE
2282.4	FALSE	15480	FALSE	14391	FALSE	15480	FALSE	8	FALSE
6712.38	FALSE	58455	FALSE	59296.5	FALSE	58455	FALSE	12	FALSE
5930.685	FALSE	128749.5	FALSE	128749.5	FALSE	128749.5	FALSE	24	FALSE
11315.88	FALSE	113170.5	FALSE	101853	FALSE	113170.5	FALSE	10	FALSE
9000	FALSE	0	FALSE	45000	FALSE	104033.25	FALSE	0	FALSE
43220.925	TRUE	855000	TRUE	954315	TRUE	855000	TRUE	48	TRUE
9261.09	FALSE	53955	FALSE	48325.5	FALSE	53955	FALSE	6	FALSE
9007.29	FALSE	81810	FALSE	81810	FALSE	81810	FALSE	10	FALSE
5653.26	FALSE	59490	FALSE	59854.5	FALSE	59490	FALSE	12	FALSE
7516.62	FALSE	20695.5	FALSE	21240	FALSE	20695.5	FALSE	3	FALSE









c.) Data Imbalance :-

NAME_CONTRACT_TYPE	Count
Cash loans	20856
Consumer loans	23510
Revolving loans	5625
XNA	8

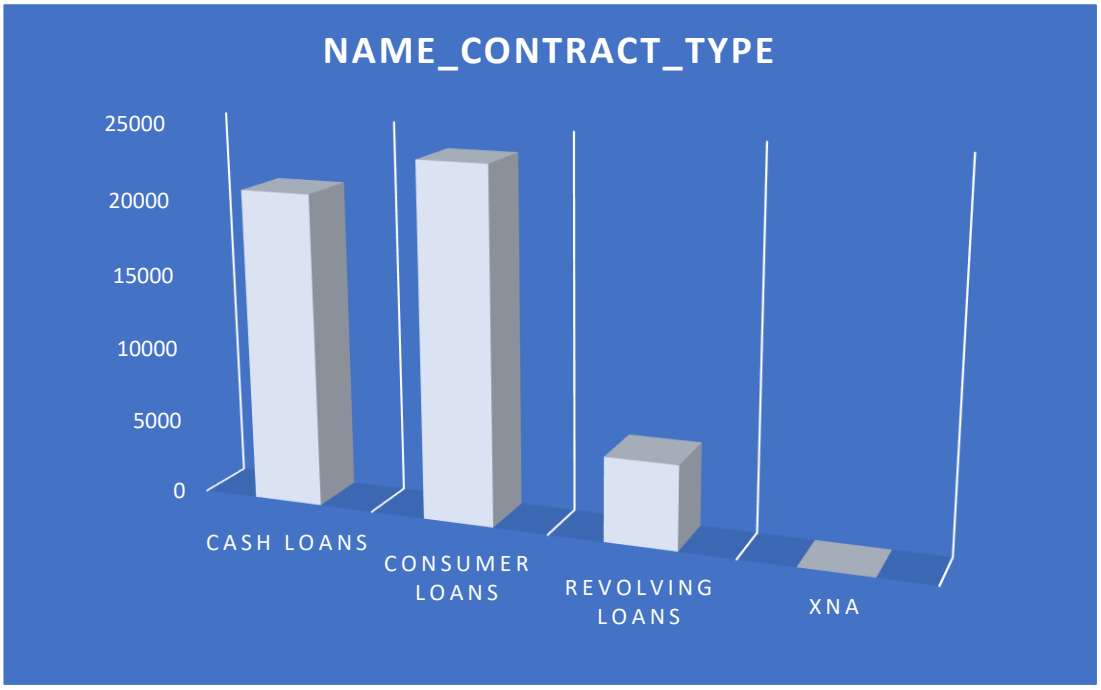
NAME_CONTRACT_STATUS	Count
Approved	31885
Canceled	8595
Refused	8660
Unused offer	859

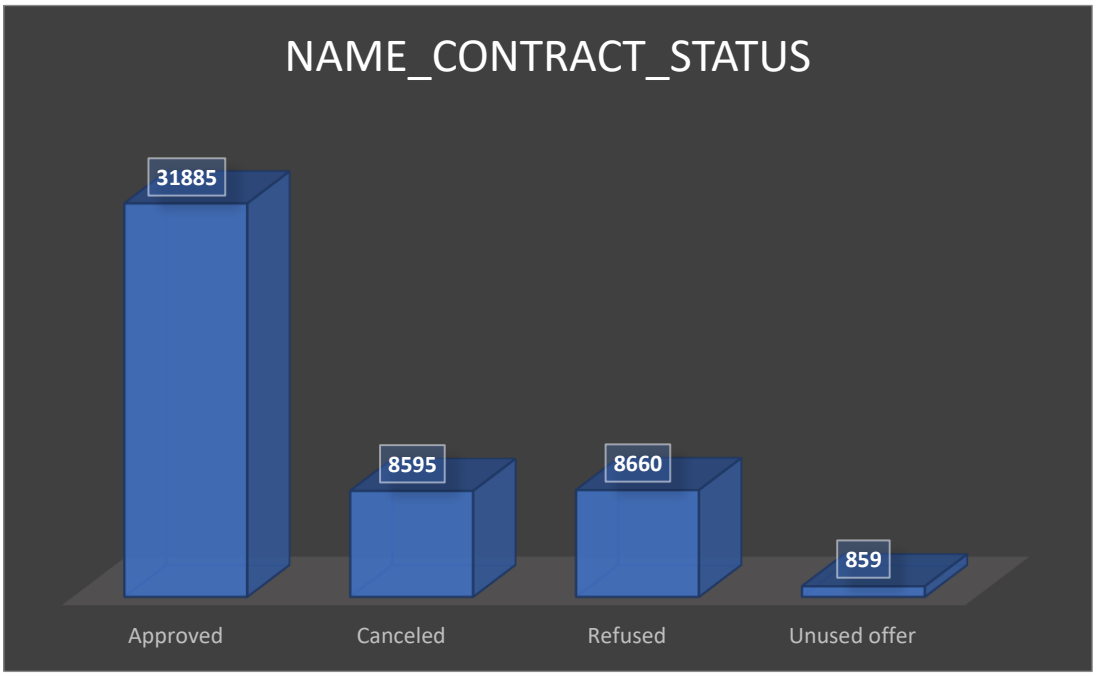
NAME_TYPE_SUITE	Count
Children	993
Family	6581
Group of people	76
Other_A	262
Other_B	551
Spouse, partner	2098
Unaccompanied	39438

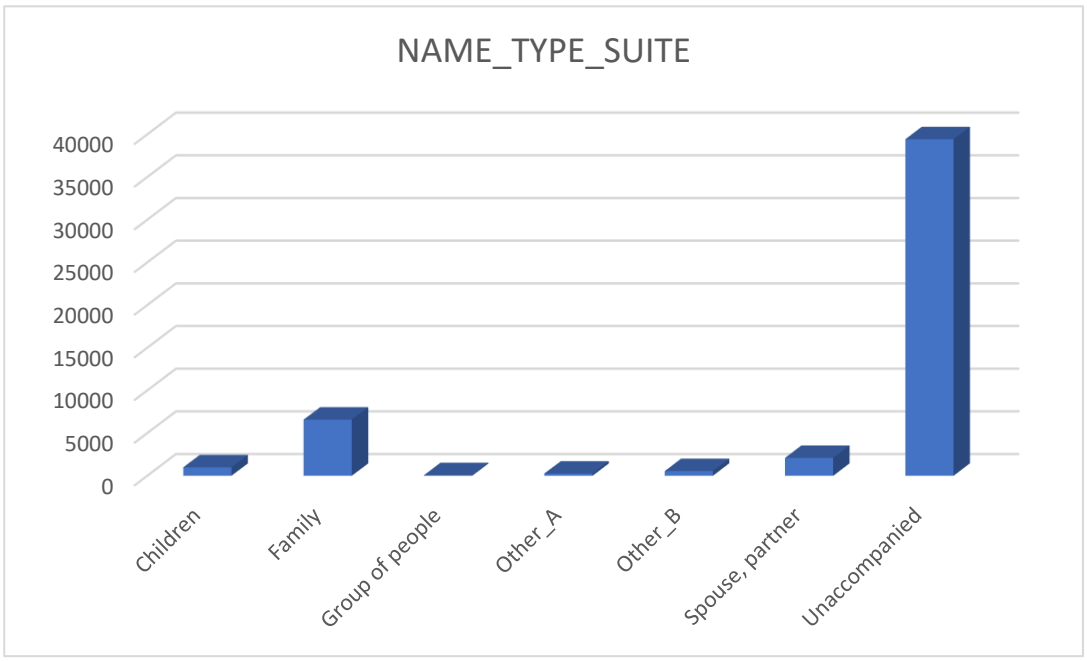
NAME_CLIENT_TYPE	Count
New	9548
Refreshed	4227
Repeater	36167
XNA	57

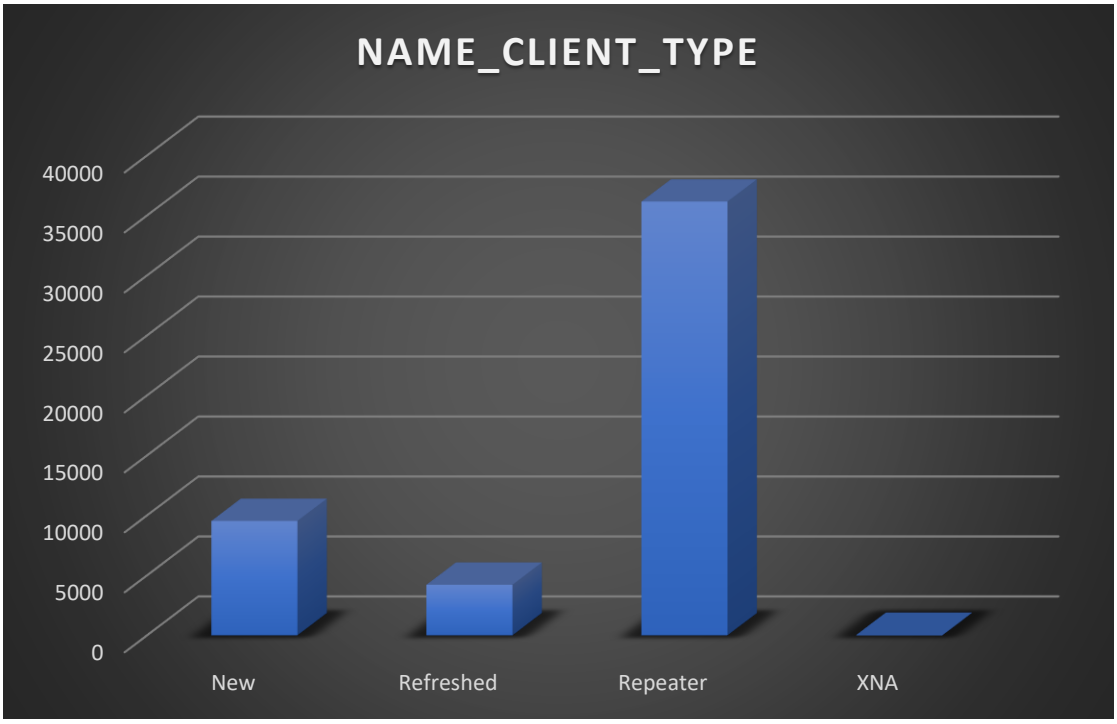
NAME_PORTFOLIO	Count
Cards	4210
Cars	14
Cash	12917
POS	22266
XNA	10592

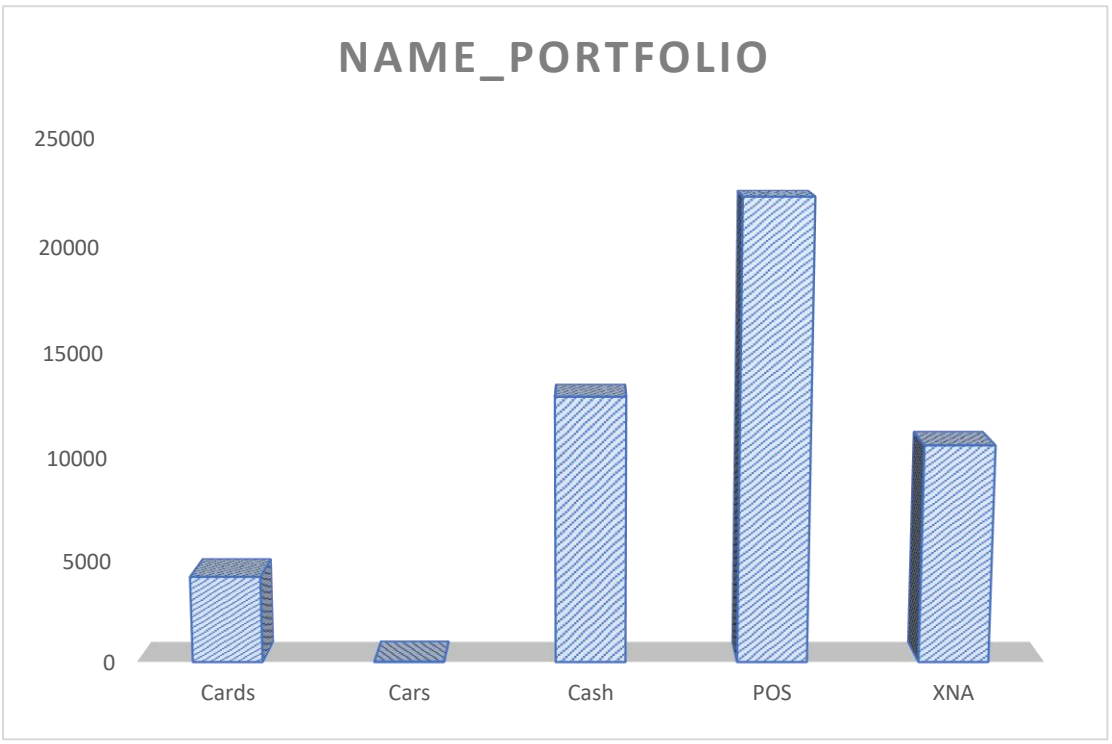
NAME_YIELD_GROUP	Count
high	10807
low_action	2953
low_normal	9858
middle	11579
XNA	14802

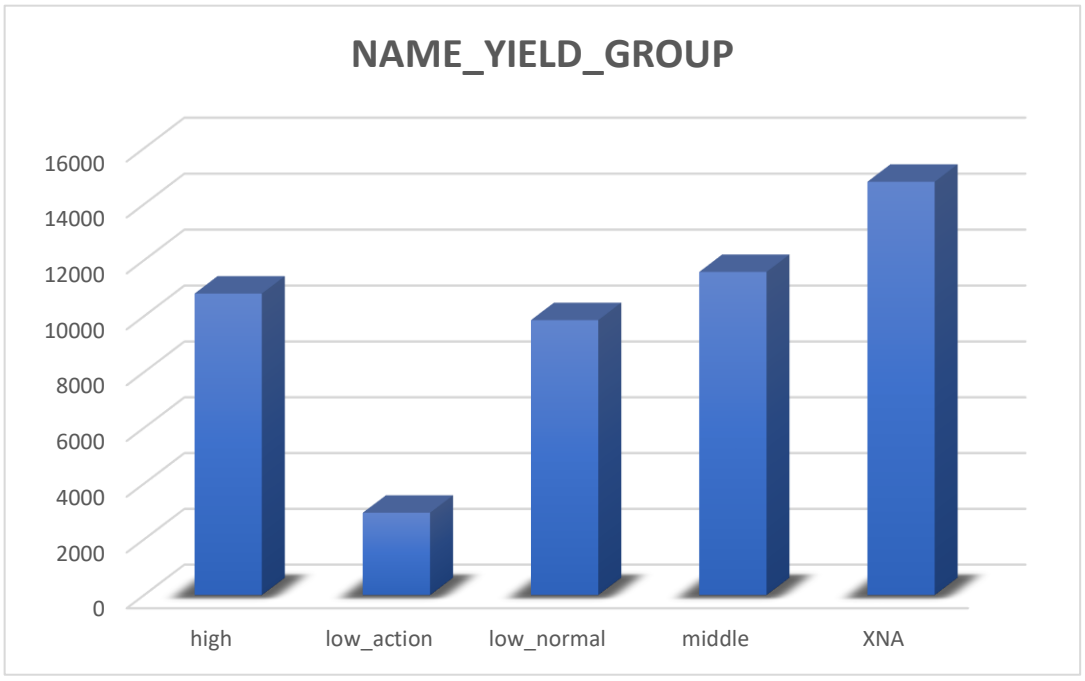

NAME_CONTRACT_TYPE
25000
20000
15000
10000
5000
0
CASH LOANS
CONSUMER LOANS
REVOLVING LOANS
XNA


NAME_CONTRACT_STATUS
31885
8595
8660
859
Approved
Canceled
Refused
Unused offer


NAME_TYPE_SUITE
40000
35000
30000
25000
20000
15000
10000
5000
0
Children
Family
Group of people
Other_A
Other_B
Spouse, partner
Unaccompanied

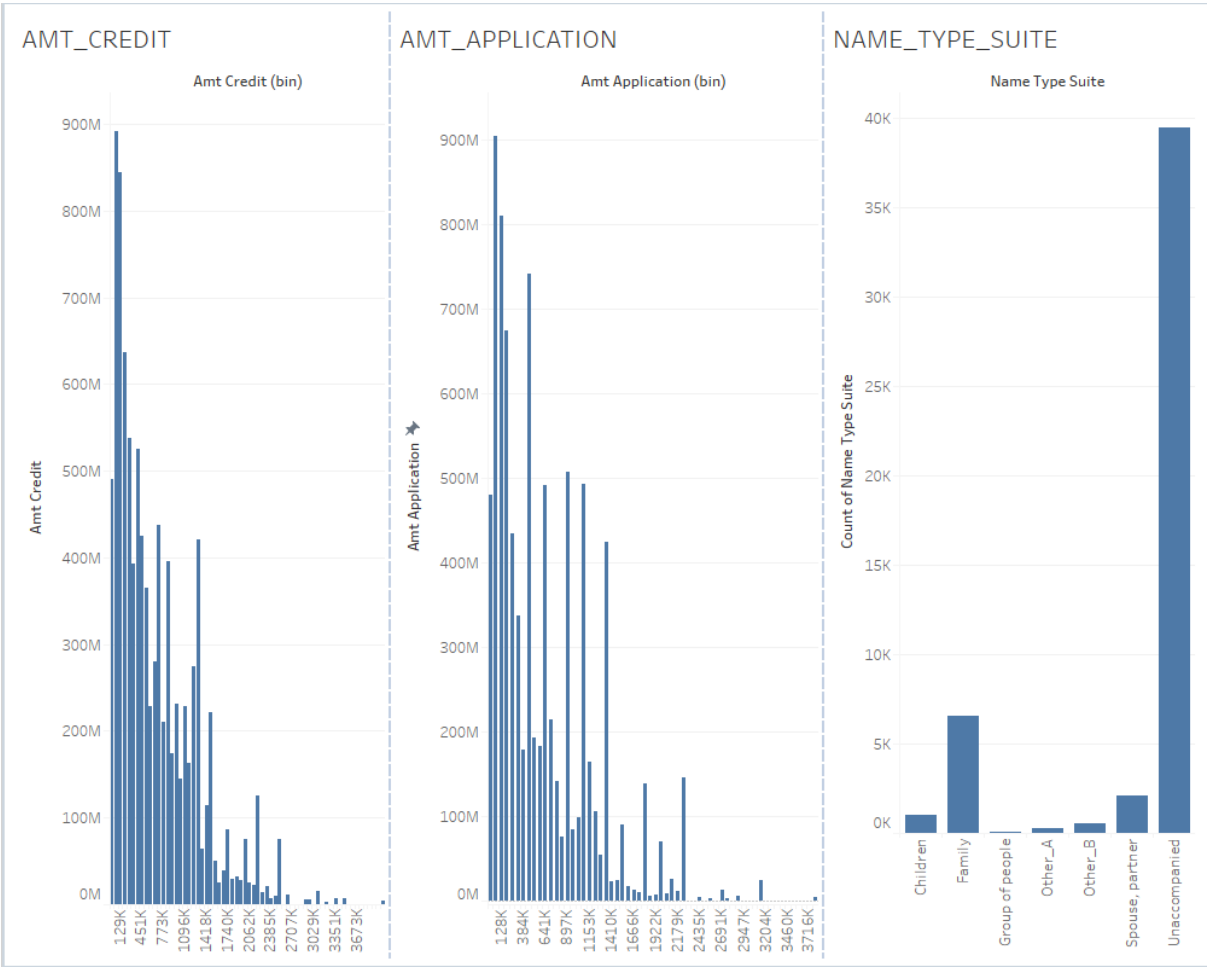

NAME_CLIENT_TYPE
40000
35000
30000
25000
20000
15000
10000
5000
0
New
Refreshed
Repeater
XNA


NAME_PORTFOLIO
25000
20000
15000
10000
5000
0
Cards
Cars
Cash
POS
XNA


NAME_YIELD_GROUP
16000
14000
12000
10000
8000
6000
4000
2000
0
high
low_action
low_normal
middle
XNA

d.) Univariate Analysis :- This is a statistical method which is used to analyse the distribution and characteristics of a single variable in a dataset. This involves calculating the data's different components such as mean, median, mode, range, variance, etc. We have used 'TABLEAU' for this process too to make the data analysis easier.

- Moving on to univariate analysis : We will plot count of Amt_Credit, Amt_Application and Name_type_suite as rows.
- Next, we will convert a copy of Amt_Credit, Amt_Application and Name_type_suite to bins with a bin size of recommended size.
- Now plot the (Bin) quantities onto the column bar.



For detailed analysis, please refer to the following tableau workbook :-

[Univariate_Analysis_Previous_Applications_dataset]

e.) Identification of top correlations :- Major factors that influence the correlation are :

AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	DAYS_DECISION	CNT_PAYMENT

- The approach we will follow for this is to compare the correlations between different variables with a CORREL chart.
- The chart with the rounded off values that define the correlation between the major variables :

	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	DAYS_DECISION	CNT_PAYMENT
AMT_ANNUITY	1	0.78	0.78	0.8	-0.25	0.39
AMT_APPLICATION	0.78	1	0.98	0.99	-0.13	0.67
AMT_CREDIT	0.78	0.98	1	0.97	-0.14	0.67
AMT_GOODS_PRICE	0.8	0.99	0.97	1	-0.19	0.66
DAYS_DECISION (+ve)	-0.25	-0.13	-0.14	-0.19	1	-0.23
CNT_PAYMENT	0.39	0.67	0.67	0.66	-0.23	1

Applications_data Excel sheet :-Applications_data_final_sheet

Previous_applications Excel sheet :- Previous_Applications_Final_Sheet

Learning Outcomes:

a. **Learnt about advanced excel techniques and tools.**
b. **Learnt about tableau functions and graph visualization.**
c. **Learnt about outliers and strategies to handle them.**
d. **Learnt about Pivot, different types of visualizations graphs and charts.**

I would like to thank TRAINITY for giving me this wonderful opportunity to learn so much from a single project. Creating a project on BANK LOAN CASE STUDY was a wonderful experience. I got to learn so many things about excel and tableau tools and useful formulas, brainstormed to find solutions whenever errors bulged in and learnt a lot about advanced excel.

The best part of the project was live implementations. I was able to implement all the excel functions and derive various results successfully in the end.