

Bank Loan Case Study

Kumar Ashutosh Kundan

Project description

The project is about finding out patterns that indicate if a customer will have difficulty paying their installments. This information can be used to make decisions such as denying the loan, reducing the amount of loan, or lending at a higher interest rate to risky applicants. We analyze this data on the following points:

- A. Identify Missing Data and Deal with it Appropriately
- B. Identify Outliers in the Dataset
- C. Analyze Data Imbalance
- D. Perform Univariate, Segmented Univariate, and Bivariate Analysis
- E. Identify Top Correlations for Different Scenarios

Missing Data

Task A: 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 COUNT, IF and ISBLANK function to get number of null values for each column.
- Then we will calculate percentage of null values for each column.
- Using TRANSPOSE function we will convert rows to columns.
- We will drop the columns which has more than or equal to 50% null values.
- We will drop irrelevant columns for doing our analysis

Formulas:

no_of_null_values: =COUNT(IF(ISBLANK(application_data!AW2:AW50000),application_data!\$A:\$A))

Transpose: =TRANSPOSE(A1:DS3)

Missing Data

Column name	no of null values	Percentage of null values
COMMONAREA_AVG	34960	70%
COMMONAREA_MODE	34960	70%
COMMONAREA_MEDI	34960	70%
NONLIVINGAPARTMENTS_AVG	34714	69%
NONLIVINGAPARTMENTS_MODE	34714	69%
NONLIVINGAPARTMENTS_MEDI	34714	69%
LIVINGAPARTMENTS_AVG	34226	68%
LIVINGAPARTMENTS_MODE	34226	68%
LIVINGAPARTMENTS_MEDI	34226	68%
FONDKAPREMONT_MODE	34191	68%
FLOORSMIN_AVG	33894	68%
FLOORSMIN_MODE	33894	68%
FLOORSMIN_MEDI	33894	68%
YEARS_BUILD_AVG	33239	66%
YEARS_BUILD_MODE	33239	66%
YEARS_BUILD_MEDI	33239	66%
OWN_CAR_AGE	32949	66%
LANDAREA_AVG	29721	59%
LANDAREA_MODE	29721	59%
LANDAREA_MEDI	29721	59%
BASEMENTAREA_AVG	29199	58%
BASEMENTAREA_MODE	29199	58%
BASEMENTAREA_MEDI	29199	58%
EXT_SOURCE_1	28172	56%
NONLIVINGAREA_AVG	27572	55%
NONLIVINGAREA_MODE	27572	55%
NONLIVINGAREA_MEDI	27572	55%
ELEVATORS_AVG	26651	53%
ELEVATORS_MODE	26651	53%
ELEVATORS_MEDI	26651	53%
WALLSMATERIAL_MODE	25459	51%
APARTMENTS_AVG	25385	51%
APARTMENTS_MODE	25385	51%
APARTMENTS_MEDI	25385	51%
ENTRANCES_AVG	25195	50%
ENTRANCES_MODE	25195	50%
ENTRANCES_MEDI	25195	50%
LIVINGAREA_AVG	25137	50%
LIVINGAREA_MODE	25137	50%
LIVINGAREA_MEDI	25137	50%
HOUSETYPE_MODE	25075	50%
FLOORSMAX_AVG	24875	50%
FLOORSMAX_MODE	24875	50%
FLOORSMAX_MEDI	24875	50%

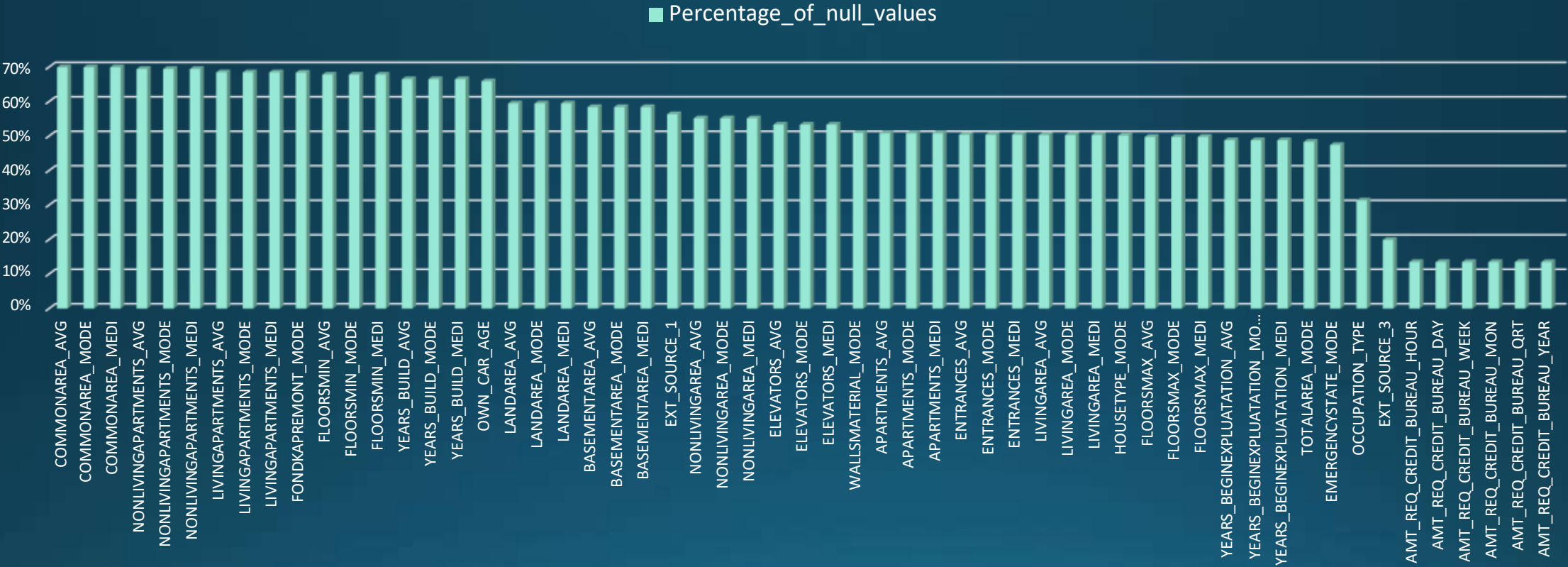
These are the columns which has null values more than or equal to 50%. These columns need to be dropped.

Missing Data

Column name	no_of_null_values	Percentage_of_null_values
FLAG_MOBIL	0	0%
FLAG_EMP_PHONE	0	0%
FLAG_WORK_PHONE	0	0%
FLAG_CONT_MOBILE	0	0%
FLAG_PHONE	0	0%
FLAG_EMAIL	0	0%
CNT_FAM_MEMBERS	1	0%
REGION_RATING_CLIENT	0	0%
REGION_RATING_CLIENT_W_CITY	0	0%
EXT_SOURCE_2	126	0%
EXT_SOURCE_3	9944	20%
YEARS_BEGINEXPLUATATION_AVG	24394	49%
YEARS_BEGINEXPLUATATION_MODE	24394	49%
YEARS_BEGINEXPLUATATION_MEDI	24394	49%
TOTALAREA_MODE	24148	48%
EMERGENCYSTATE_MODE	23698	47%
DAYS_LAST_PHONE_CHANGE	1	0%
FLAG_DOCUMENT_2	0	0%
FLAG_DOCUMENT_3	0	0%
FLAG_DOCUMENT_4	0	0%
FLAG_DOCUMENT_5	0	0%
FLAG_DOCUMENT_6	0	0%
FLAG_DOCUMENT_7	0	0%
FLAG_DOCUMENT_8	0	0%
FLAG_DOCUMENT_9	0	0%
FLAG_DOCUMENT_10	0	0%
FLAG_DOCUMENT_11	0	0%
FLAG_DOCUMENT_12	0	0%
FLAG_DOCUMENT_13	0	0%
FLAG_DOCUMENT_14	0	0%
FLAG_DOCUMENT_15	0	0%
FLAG_DOCUMENT_16	0	0%
FLAG_DOCUMENT_17	0	0%
FLAG_DOCUMENT_18	0	0%
FLAG_DOCUMENT_19	0	0%
FLAG_DOCUMENT_20	0	0%
FLAG_DOCUMENT_21	0	0%

These are the columns which have irrelevant data for analysis. These columns need to be dropped.

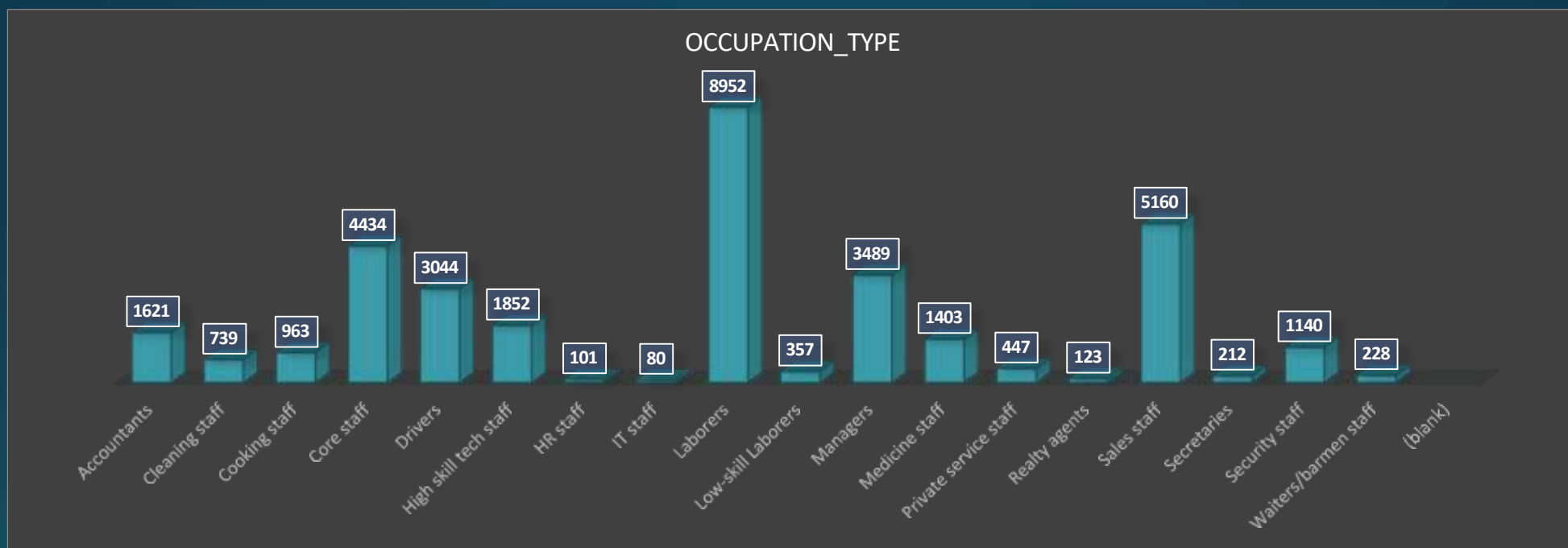
Missing Data



Missing Data

Mode Imputations:-

1. OCCUPATION_TYPE

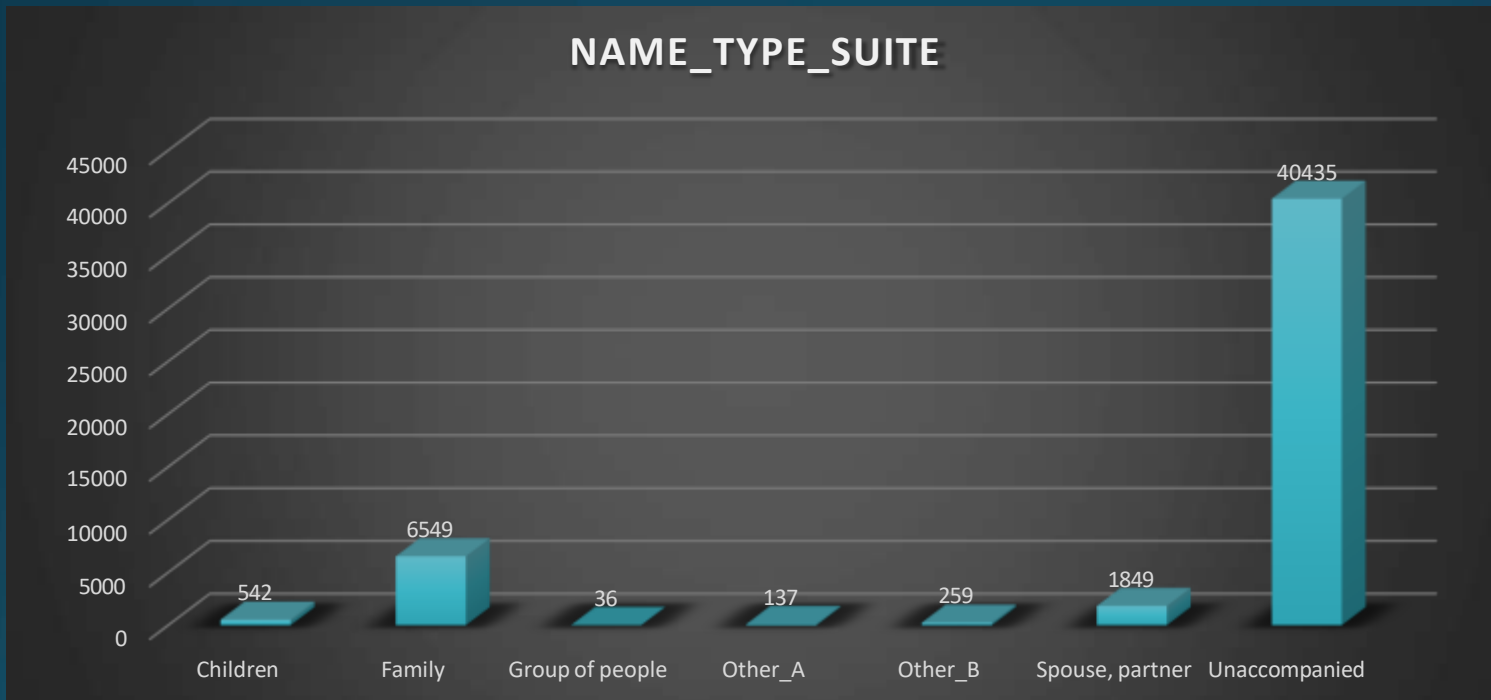


Most Occurring Variable is Laborers. We will replace blanks with 8952.

Missing Data

Mode Imputations:-

2. NAME_TYPE_SUITE

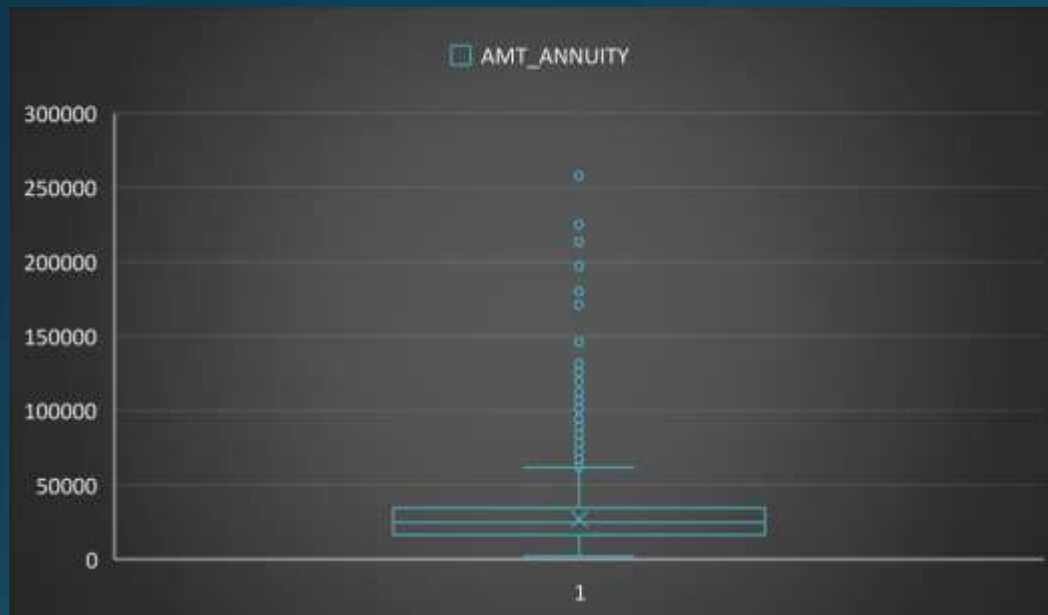


Most Occurring Variable is Unaccompanied.

Missing Data

Median Imputations:-

1. AMT_ANNUIITY

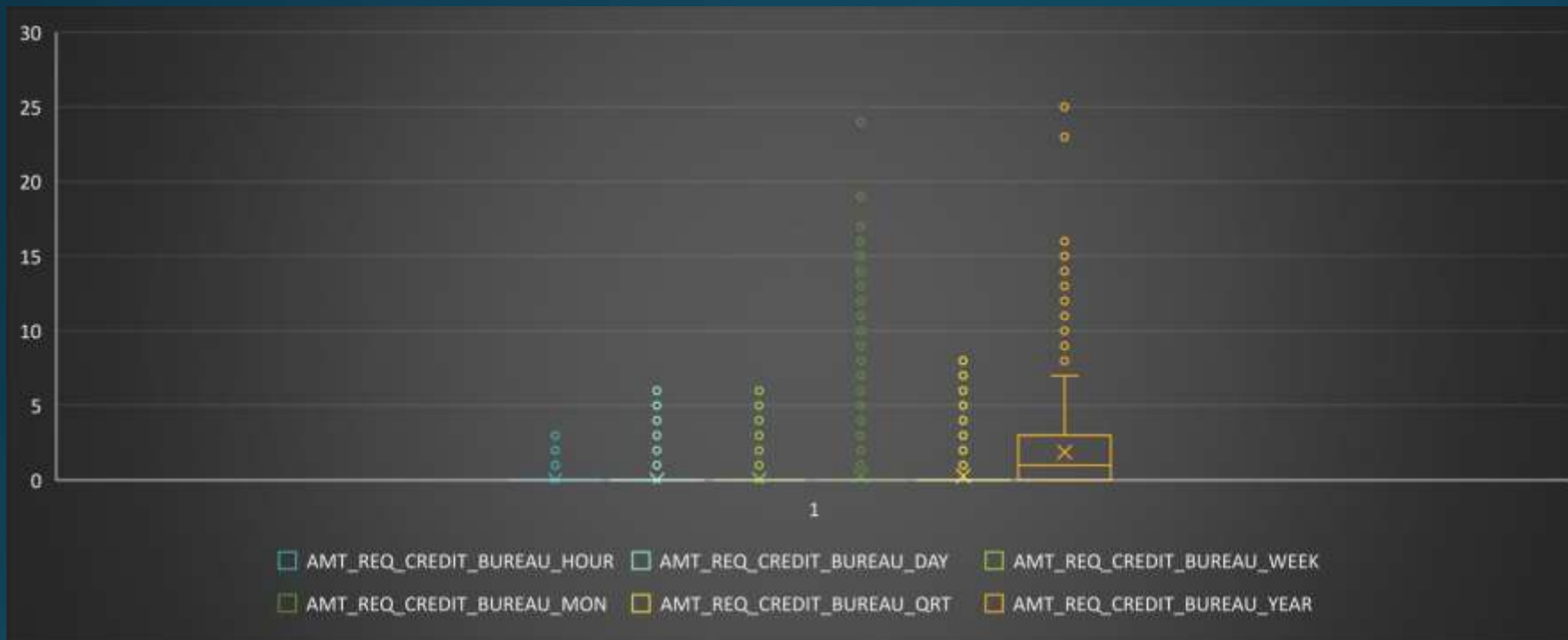


2. AMT_GOODS_PRICE



Missing Data

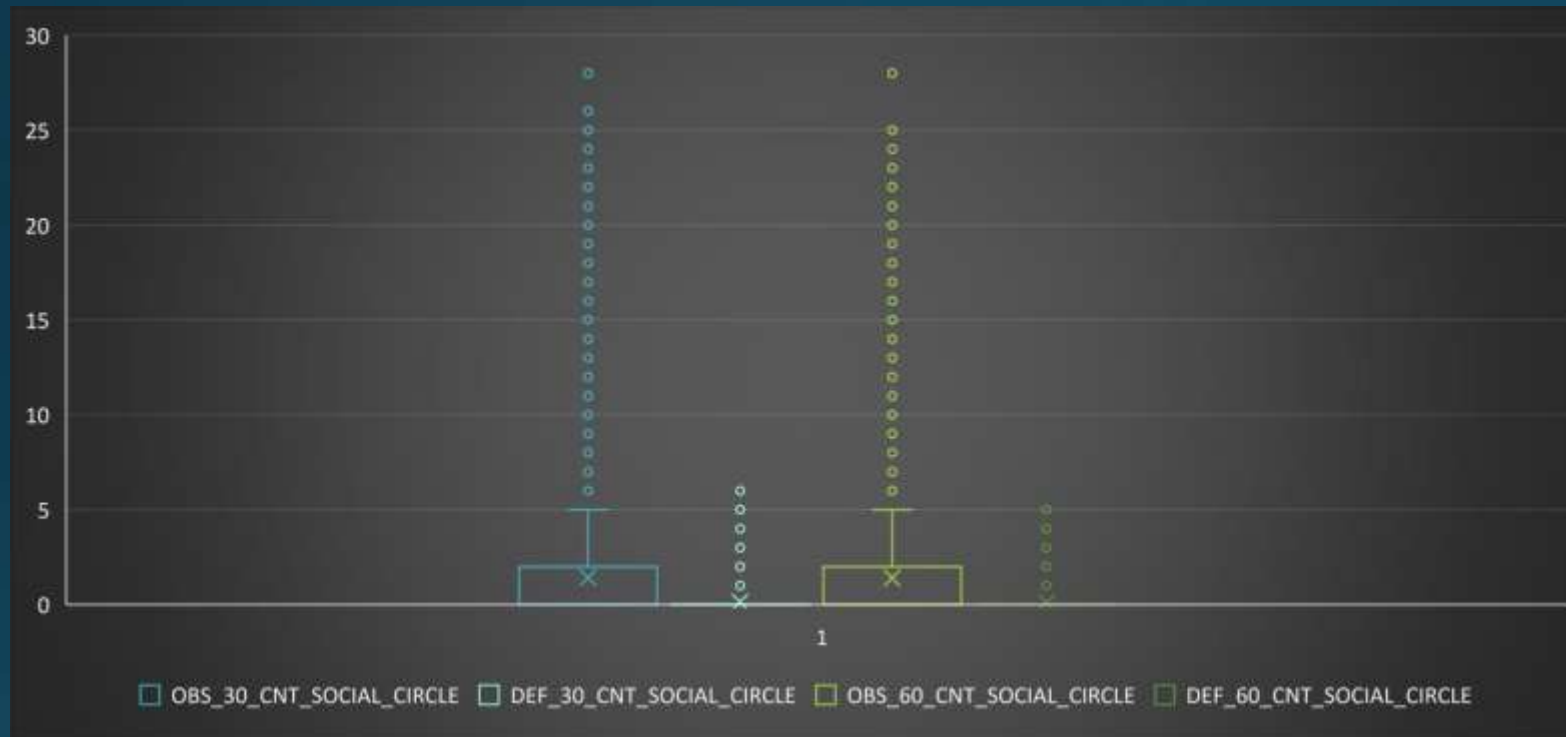
Median Imputations:-



3. AMT_REQ_CREDIT_BUREAU_HOUR
4. AMT_REQ_CREDIT_BUREAU_DAY
5. AMT_REQ_CREDIT_BUREAU_WEEK
6. AMT_REQ_CREDIT_BUREAU_MON
7. AMT_REQ_CREDIT_BUREAU_QRT
8. AMT_REQ_CREDIT_BUREAU_YEAR

Missing Data

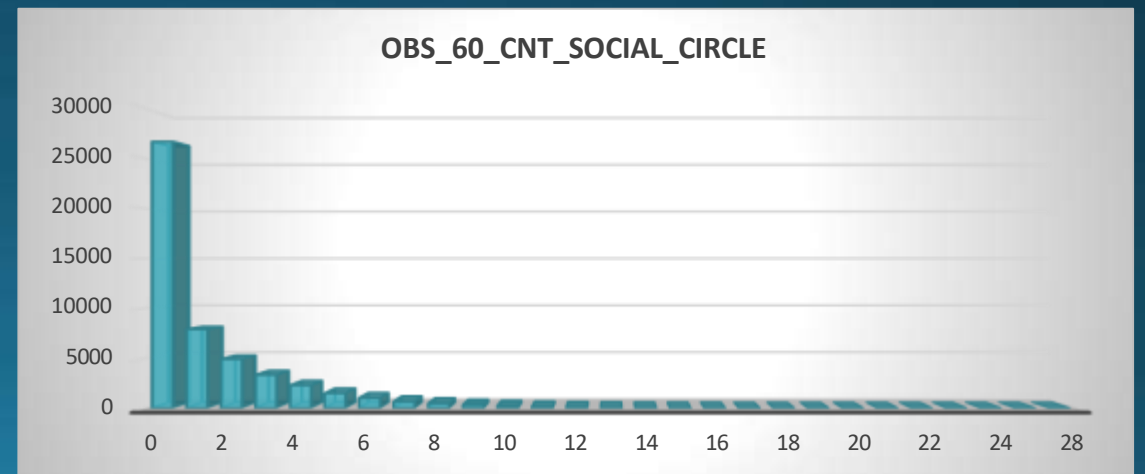
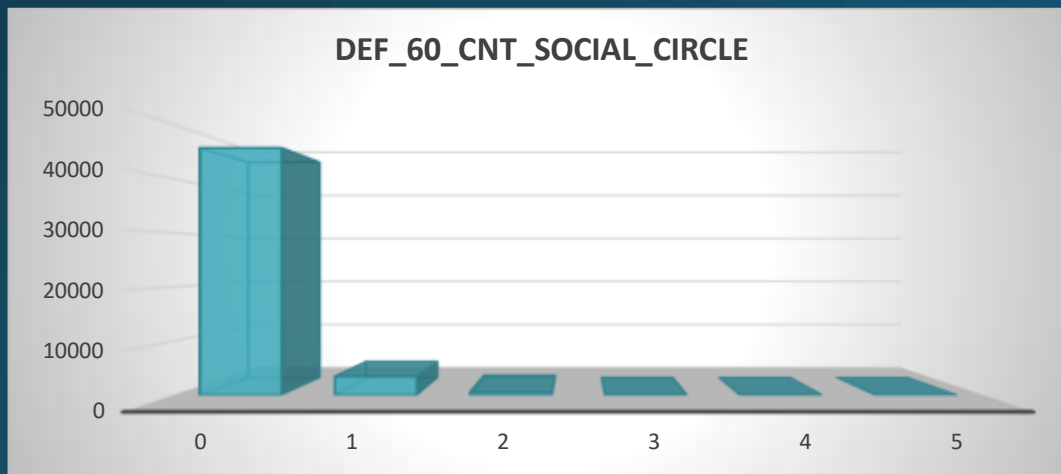
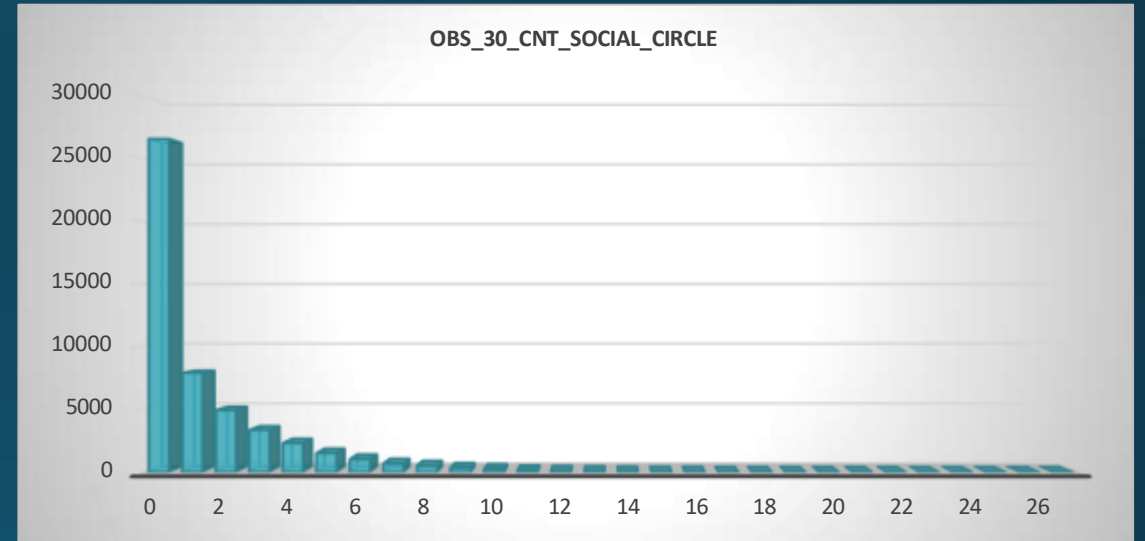
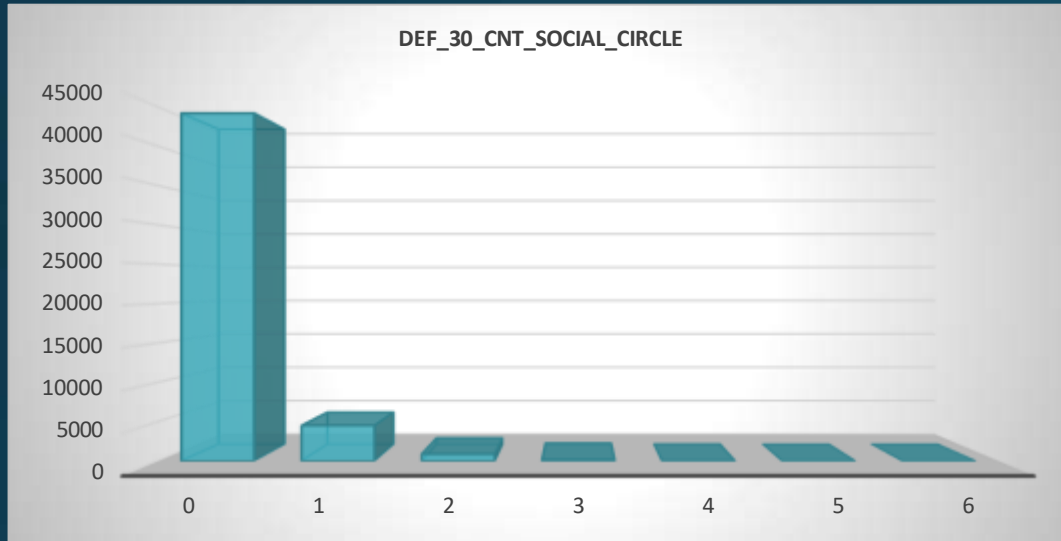
Median/Mode Imputations:-



1. DEF_30_CNT_SOCIAL_CIRCLE
2. OBS_30_CNT_SOCIAL_CIRCLE
3. DEF_60_CNT_SOCIAL_CIRCLE
4. OBS_60_CNT_SOCIAL_CIRCLE

Missing Data

Median/Mode Imputations:-



Missing Data

Previous_application datasets

Column name	no_of_null_values	Percentage_of_null_values
RATE_INTEREST_PRIMARY	49833	99.67%
RATE_INTEREST_PRIVILEGED	49833	99.67%
AMT_DOWN_PAYMENT	25197	50.40%
RATE_DOWN_PAYMENT	25197	50.40%

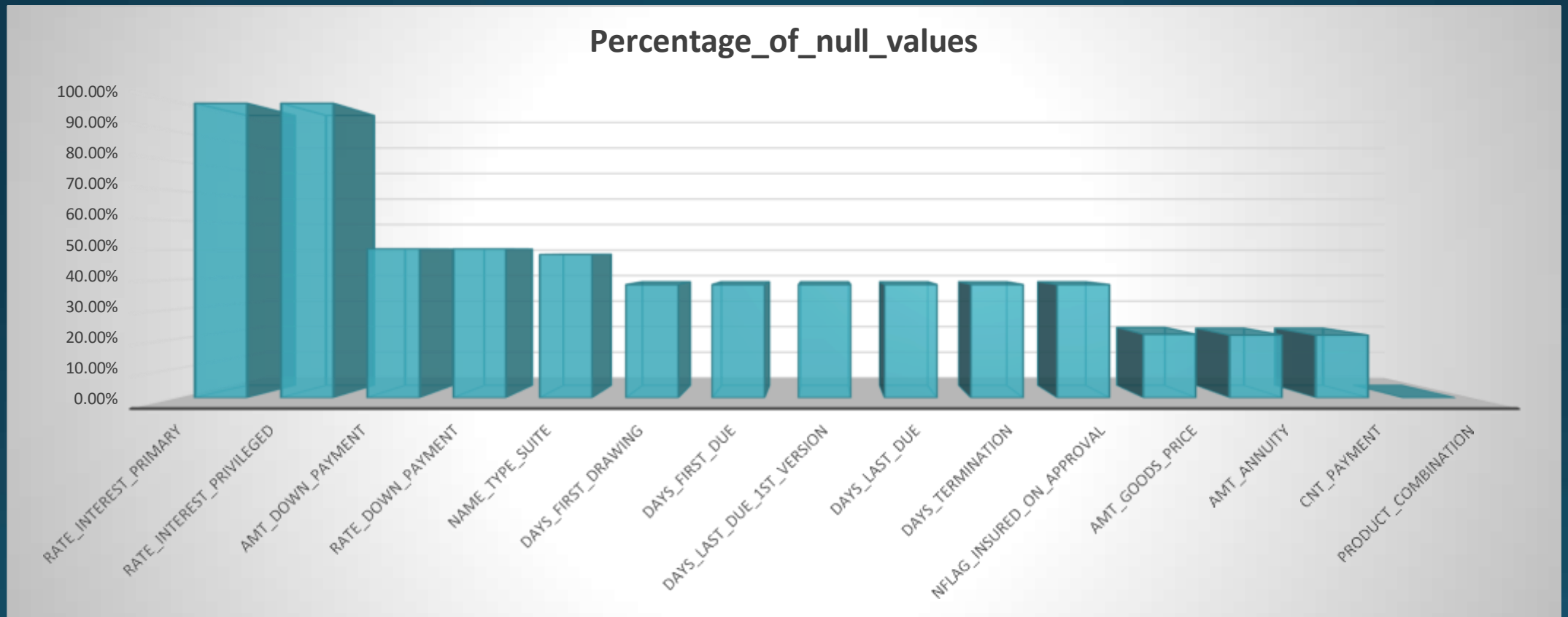
These are the columns which has null values more than or equal to 50%. These columns need to be dropped.

Column name	no_of_null_values	Percentage_of_null_values
NAME_TYPE_SUITE	24243	48.49%
PRODUCT_COMBINATION	8	0.02%
WEEKDAY_APPR_PROCESS_START	0	0.00%
HOURL_APPR_PROCESS_START	0	0.00%
FLAG_LAST_APPL_PER_CONTRACT	0	0.00%
NFLAG_LAST_APPL_IN_DAY	0	0.00%

These are the columns which have irrelevant data for analysis. These columns need to be dropped.

Missing Data

Previous_application datasets

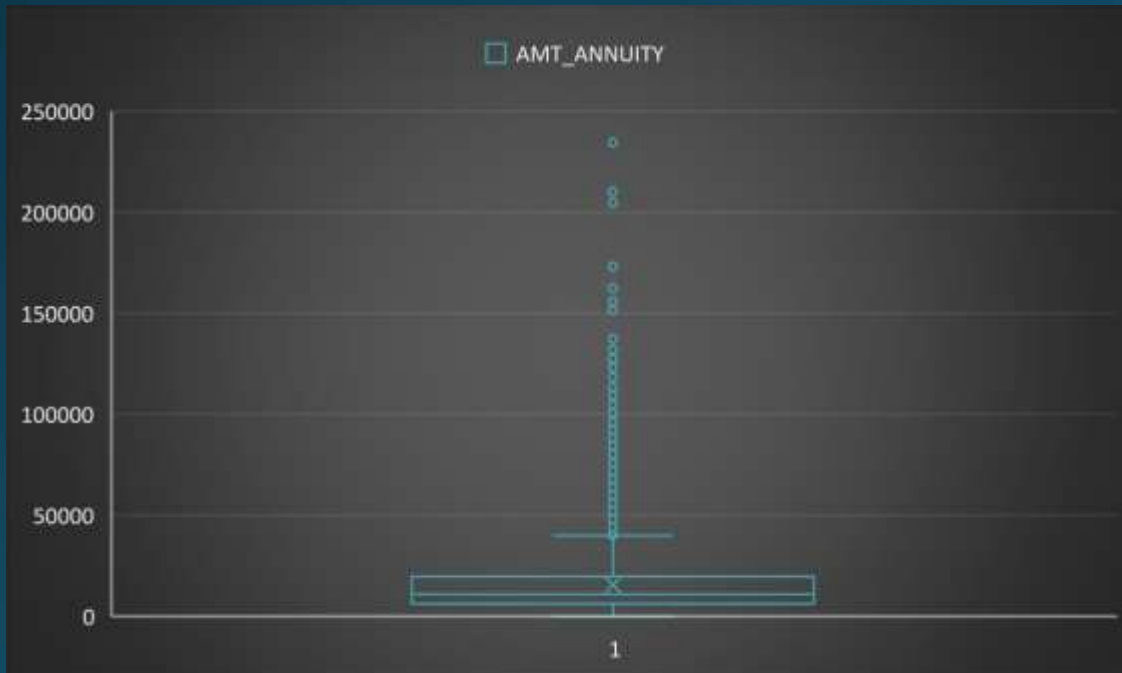


Missing Data

Previous_application datasets

Median Imputations:-

1. AMT_ANNUITY



2. AMT_GOODS_PRICE



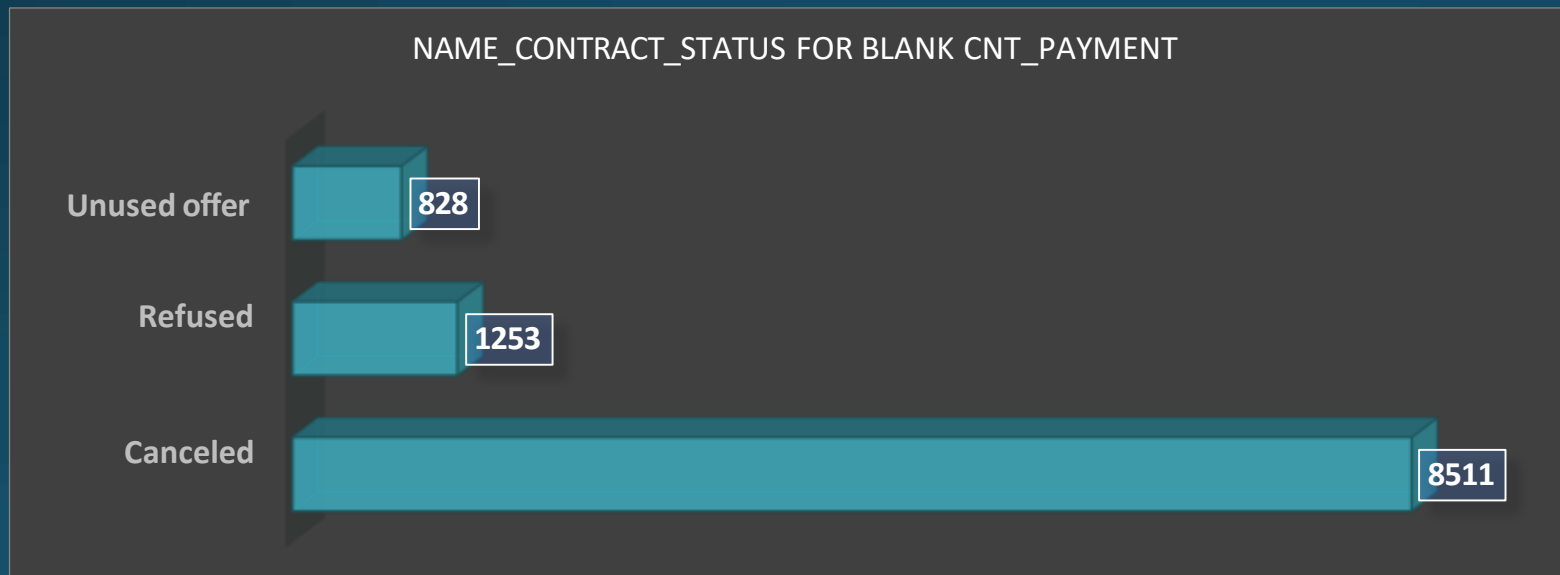
Missing Data

Previous_application datasets

Custom Imputations:-

1. CNT_PAYMENT

Most of Blank cells of cnt_payments have contract_status as canceled, refused, unused offer. So it makes more sense replacing them with 0 rather than Mean or Median.

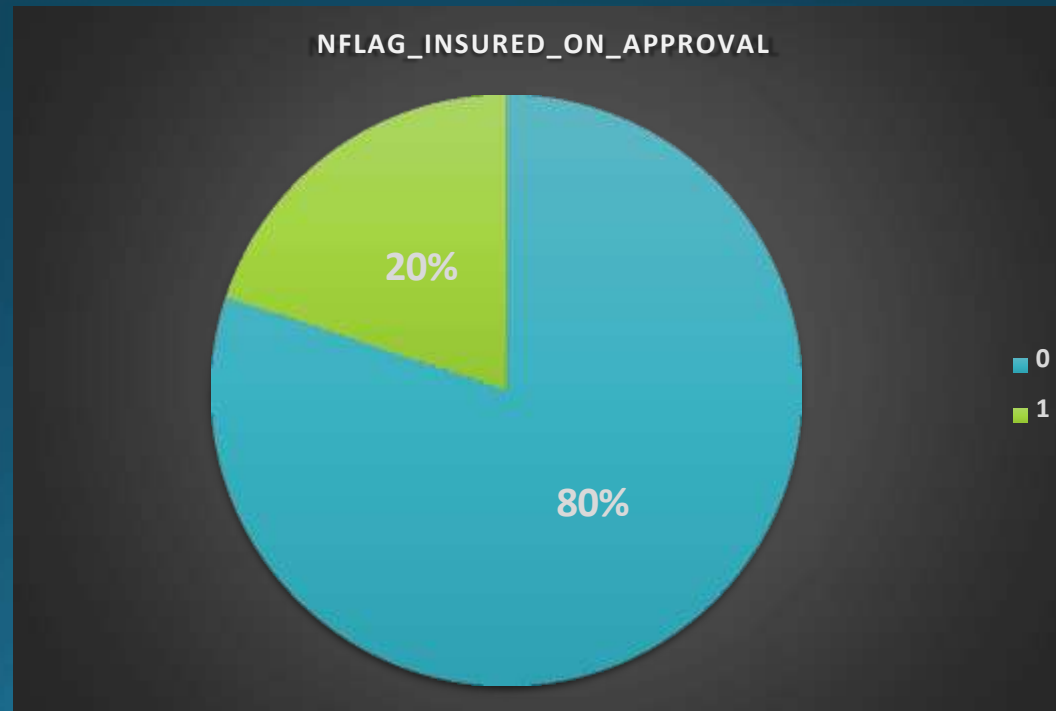


Missing Data

Previous_application datasets

Mode Imputations:-

1. NFLAG_INSURED_ON_APPROVAL



Outliers

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

- First we will select numerical columns like AMT_INCOME_TOTAL and AMT_CREDIT.
- Calculate Quartile 1, Quartile 3, IQR, Upper Limit and Lower Limit.
- We will use Box Plot to highlight the outliers.

Formulas:

Quartile 1 : =QUARTILE(A:A,1)

Quartile 3 : =QUARTILE(A:A,3)

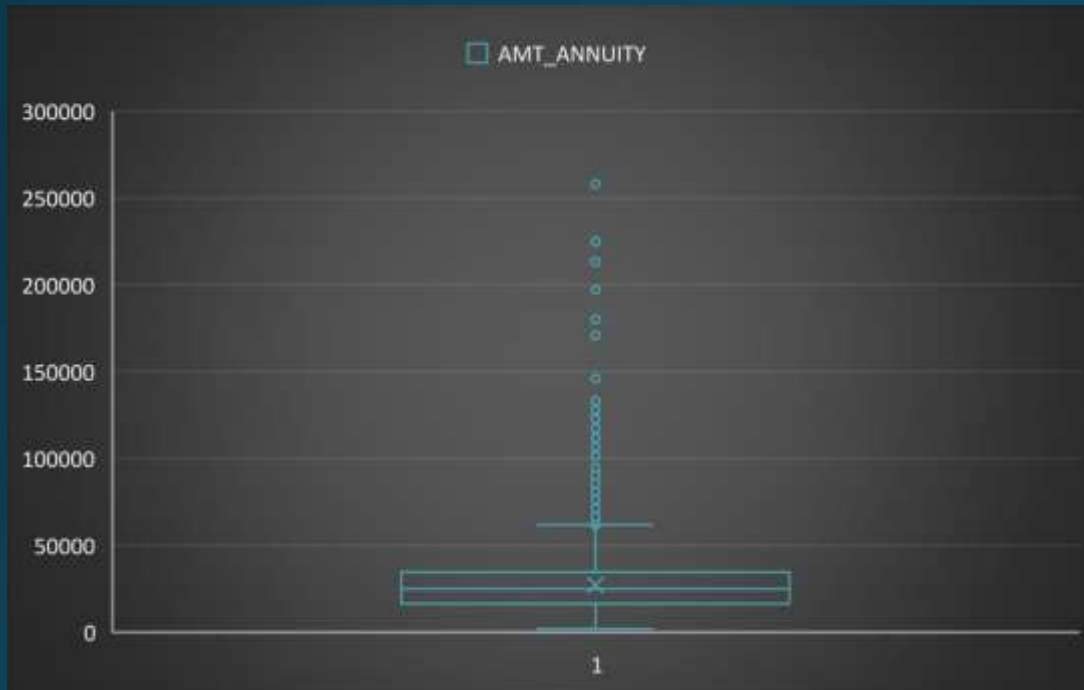
IQR = Quartile 3 - Quartile 1

Upper Limit = Quartile 3 + 1.5*IQR

Lower Limit = Quartile 1 - 1.5IQR

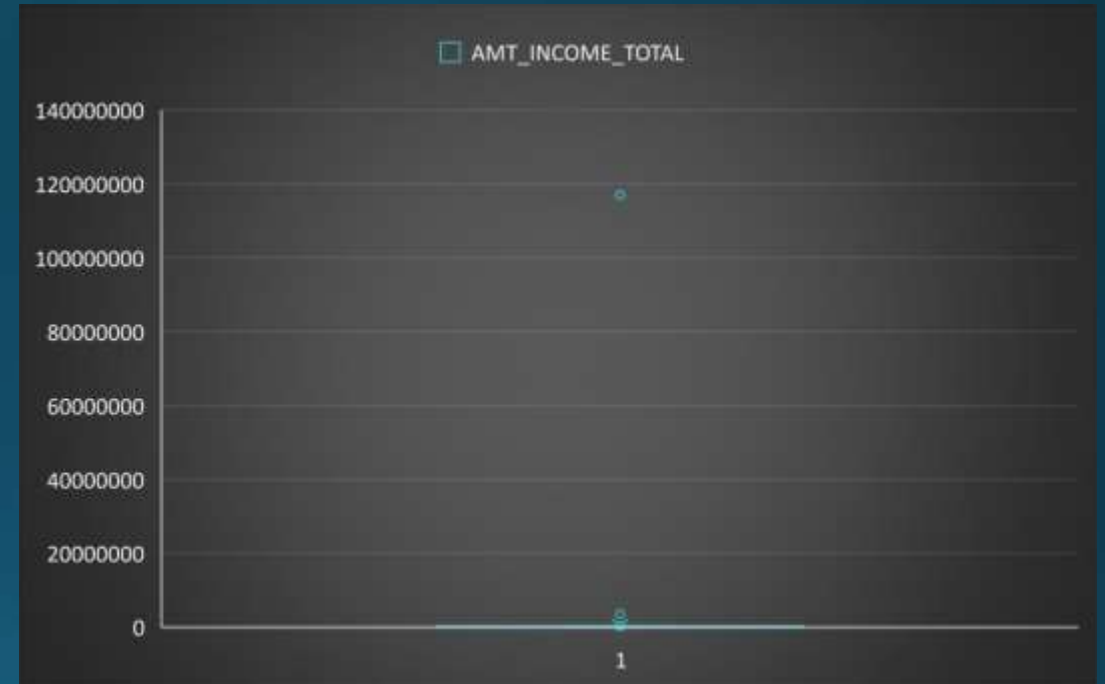
AMT_INCOME_TOTAL	
Quartile 1	112500
Quartile 3	202500
IQR	90000
Upper Limit	337500
Lower Limit	-22500

Outliers



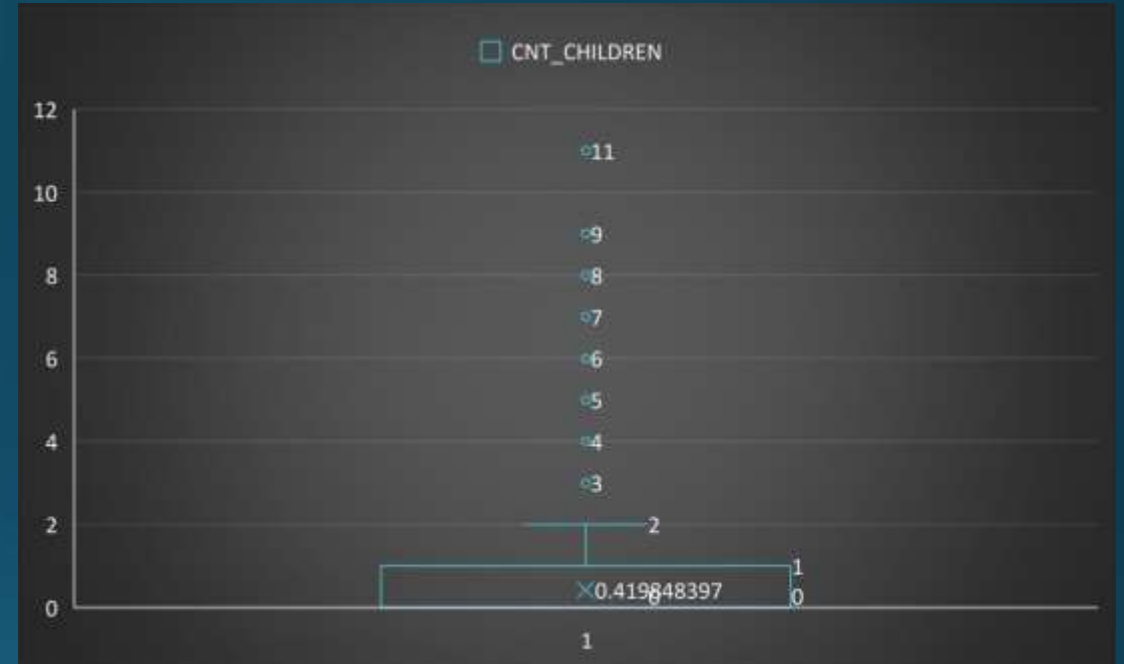
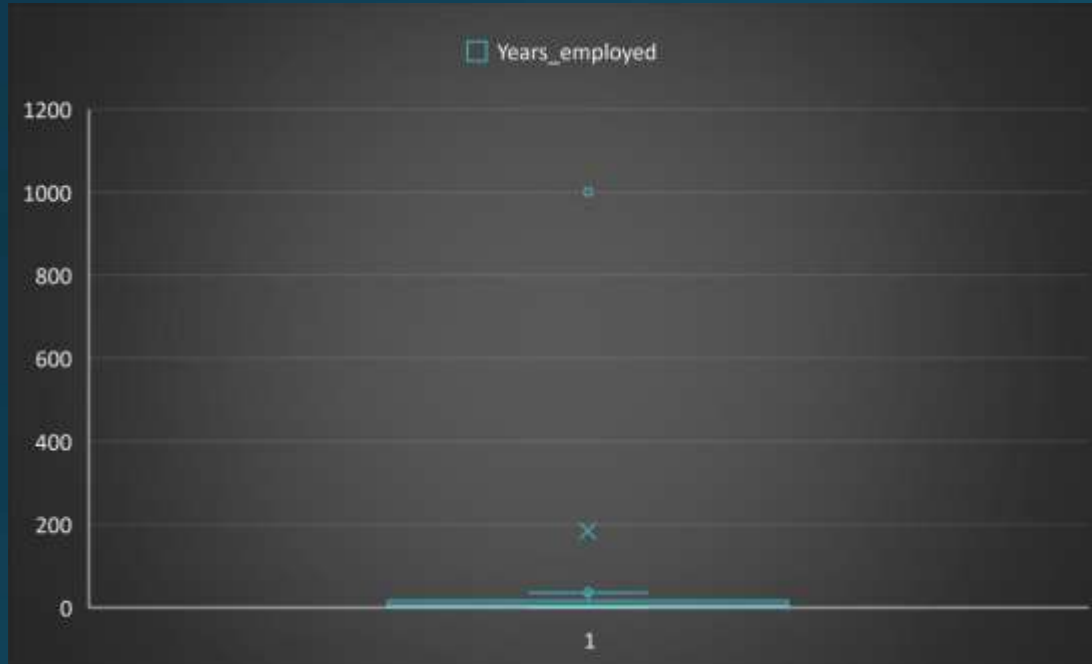
In the chart we can see there are few outliers in columns like AMT_ANNUITY and AMT_GOODS_PRICE.

Outliers



There are few outliers in columns like AMT_CREDIT and AMT_INCOME_TOTAL where amount is higher than normal. In AMT_INCOME_TOTAL one of extreme outlier is 117000000 but we will not remove because income of person varies. We will not remove outlier from AMT_CREDIT too.

Outliers



In Column Years_employed we can see people being employed for 1001 yrs which is not possible. Column CNT_CHILDREN shows people are having 11 children which is impossible in today's age

Data Imbalance

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

TARGET COLUMN

Row Labels	Count of TARGET
0	45973
1	4026
Grand Total	49999

**Almost 92% clients are loan re-payers.
8% client are Defaulters.**



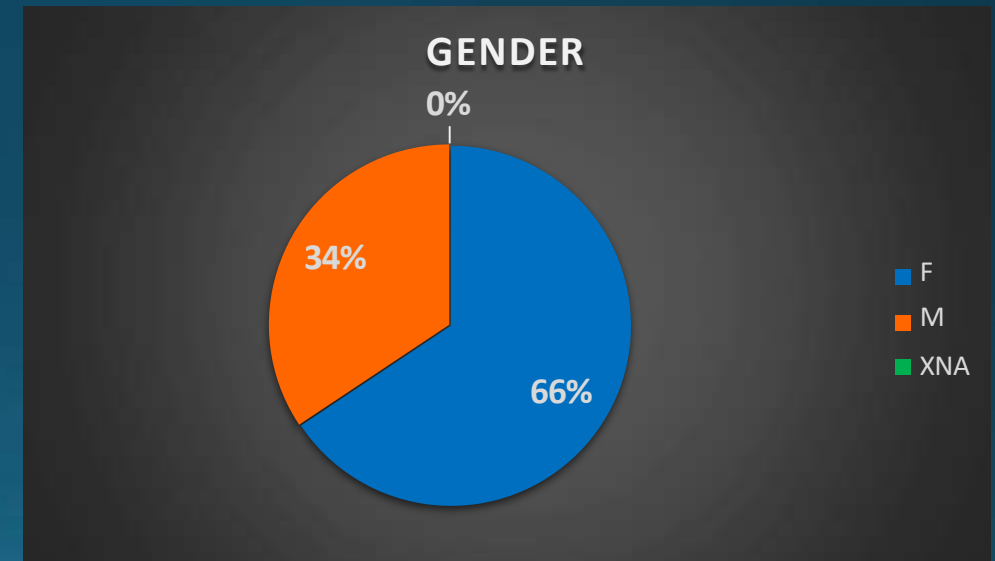
**0 – Payment on time
1 – Late Payment**

Data Imbalance

CODE_GENDER COLUMN

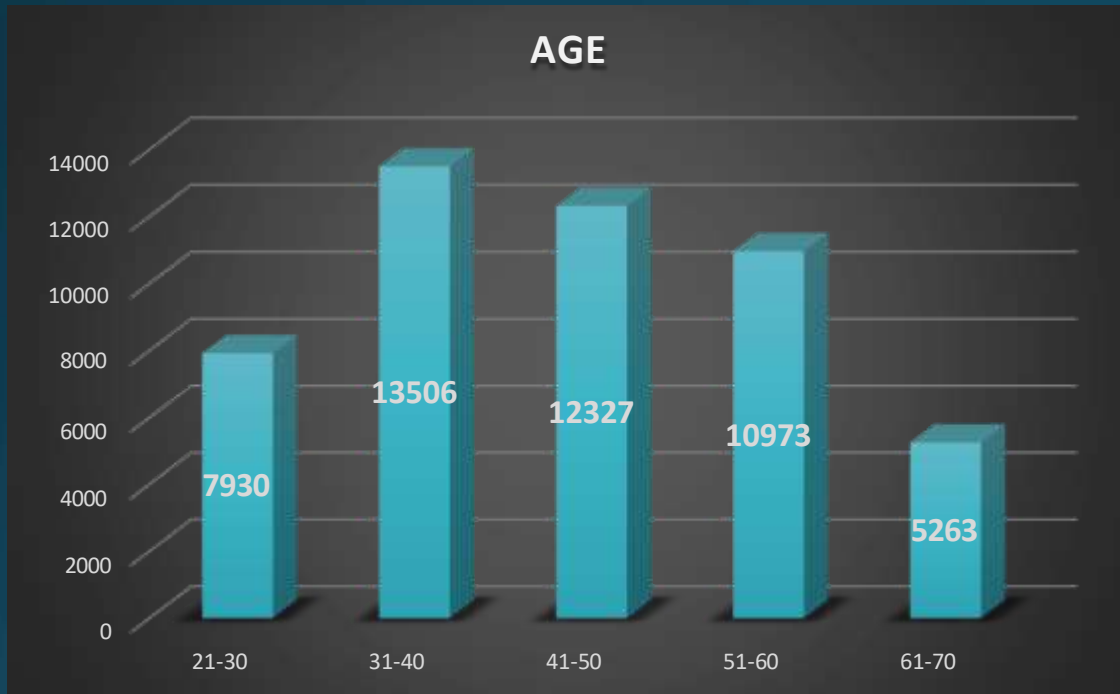
GENDER	Count of CODE_GENDER
F	32823
M	17174
XNA	2
Grand Total	49999

Almost 66% client are Female and 34% clients are Male.

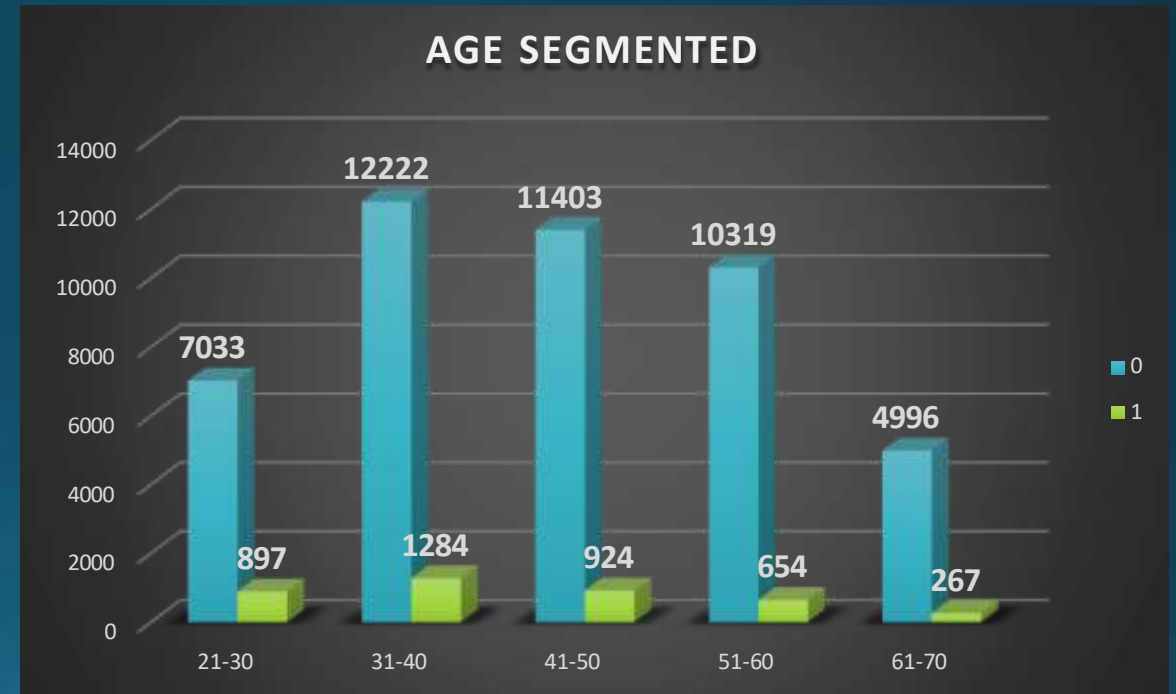


Univariate/Segmented Univariate Analysis

Task D: Perform univariate analysis to understand the distribution of individual variables, segmented univariate and bivariate analysis.

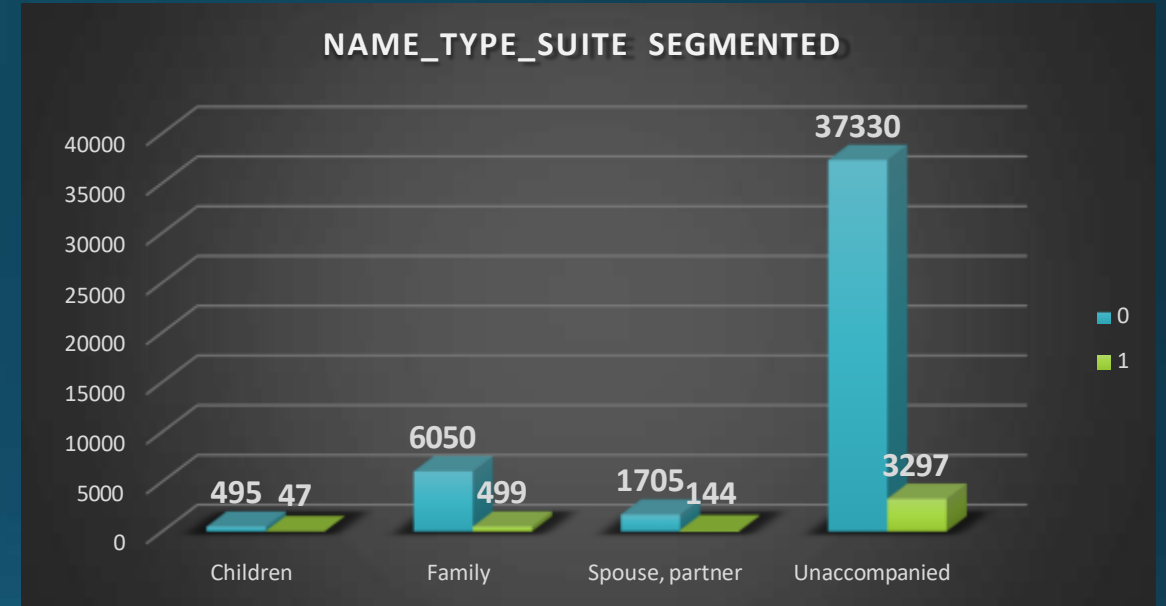
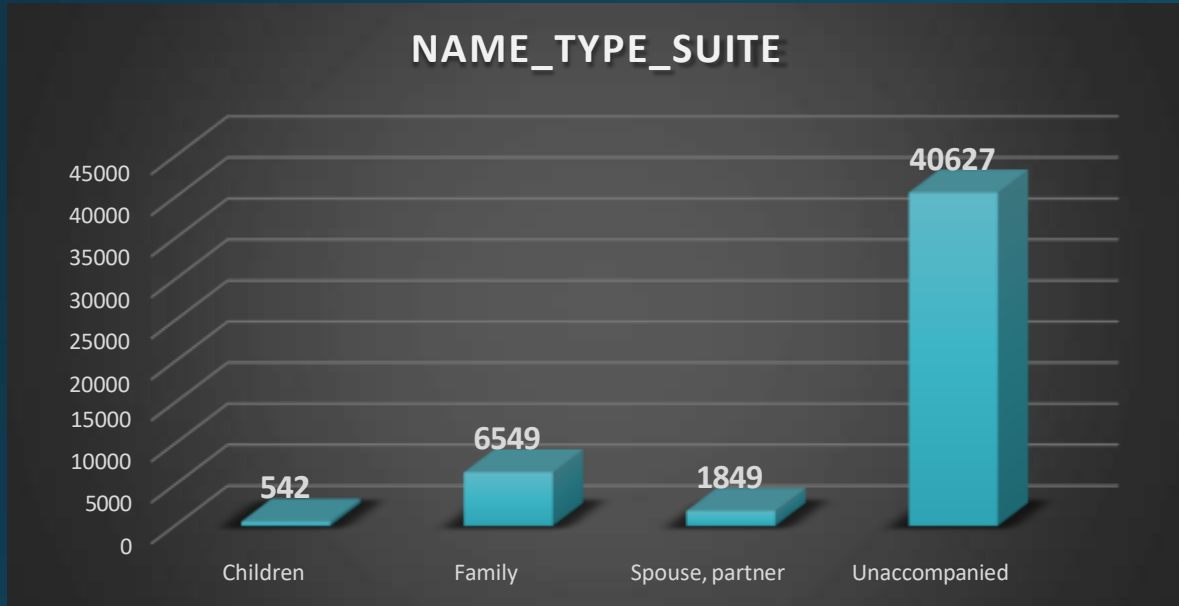


Majority of the Clients are in the age group 31-40.



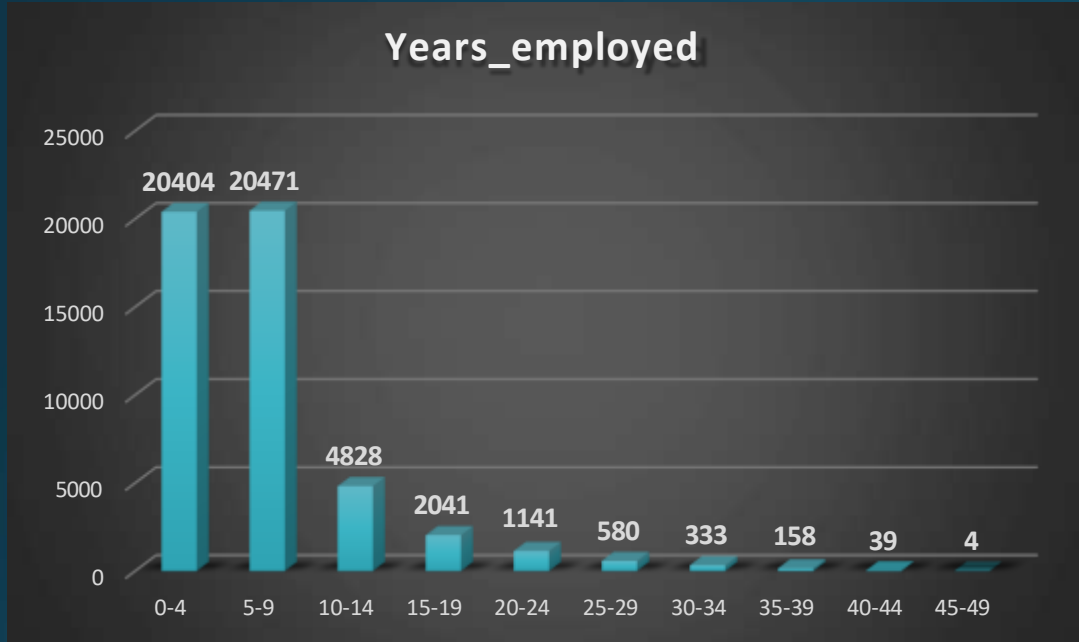
we can see as age increases , chances of defaulter decreases.

Univariate/Segmented Univariate Analysis



Majority of the Clients are Unaccompanied followed by Family.

Univariate/Segmented Univariate Analysis

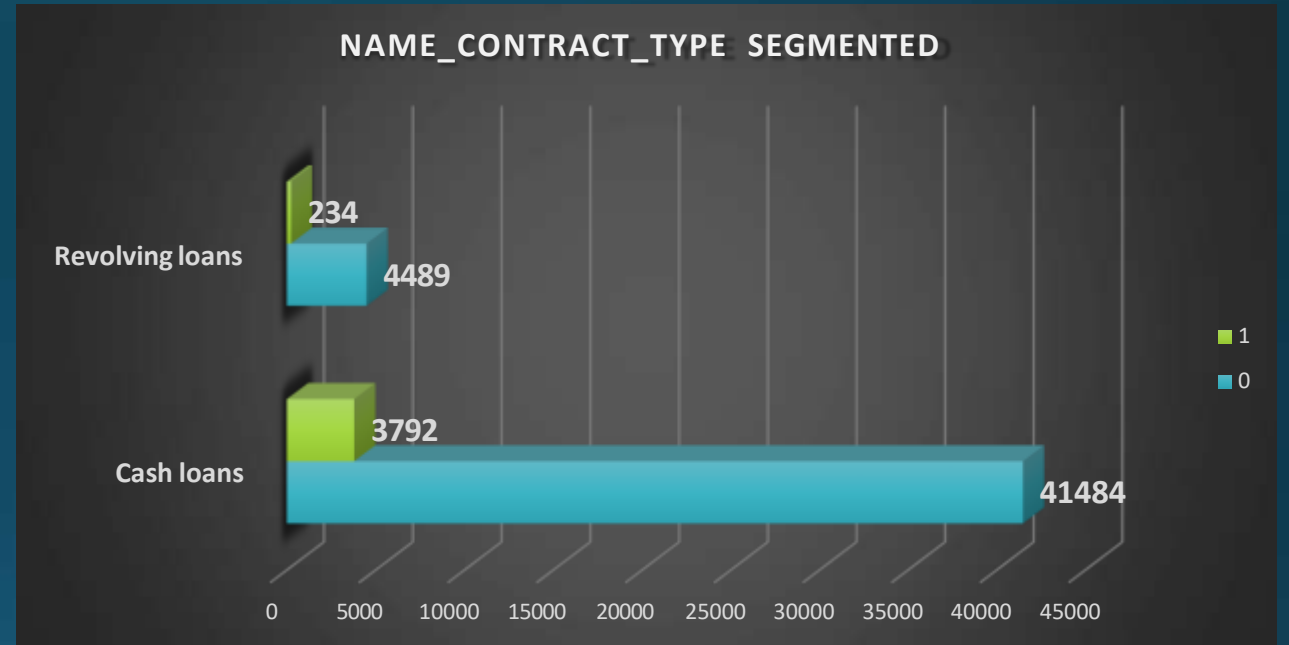
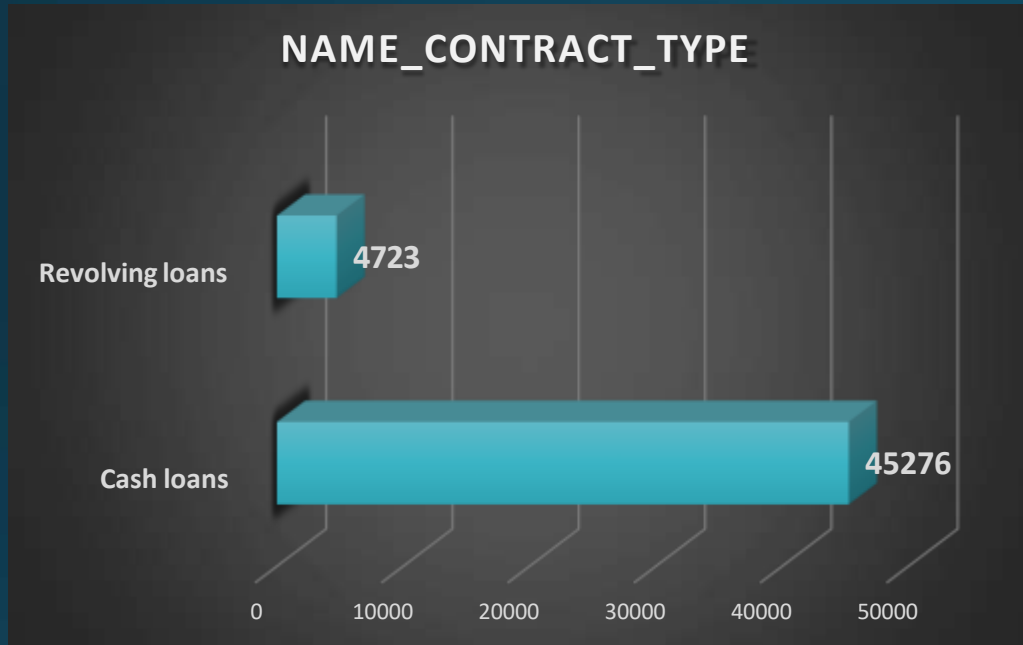


Majority of the Clients are having 0-9 years of experience.



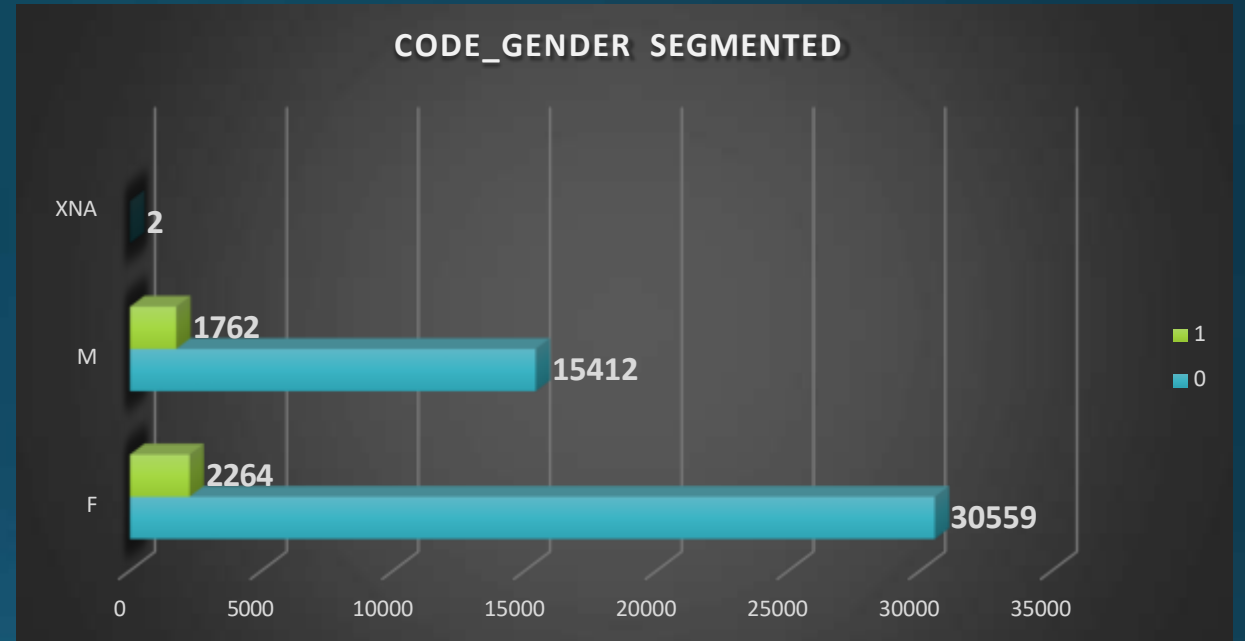
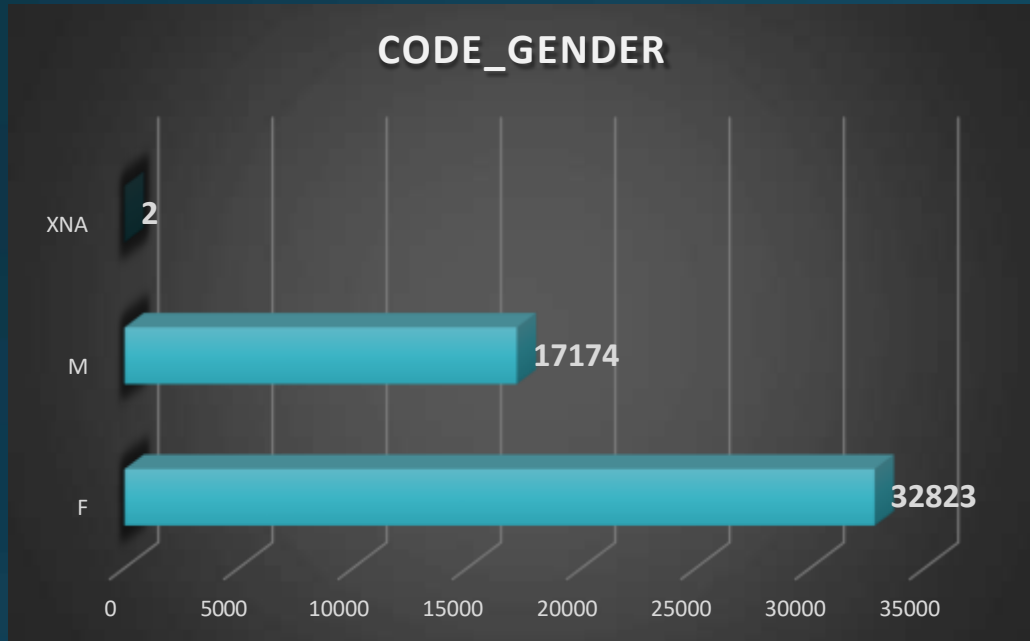
we can see as experience increases , chances of defaulting decreases.

Univariate/Segmented Univariate Analysis



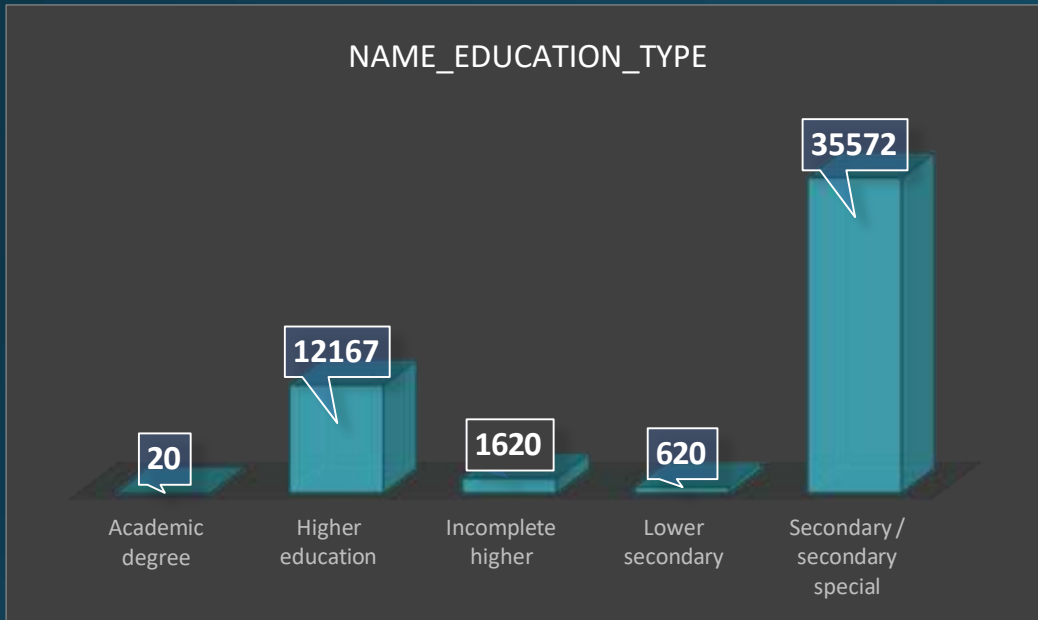
Majority of the Clients are taking Cash loans.

Univariate/Segmented Univariate Analysis

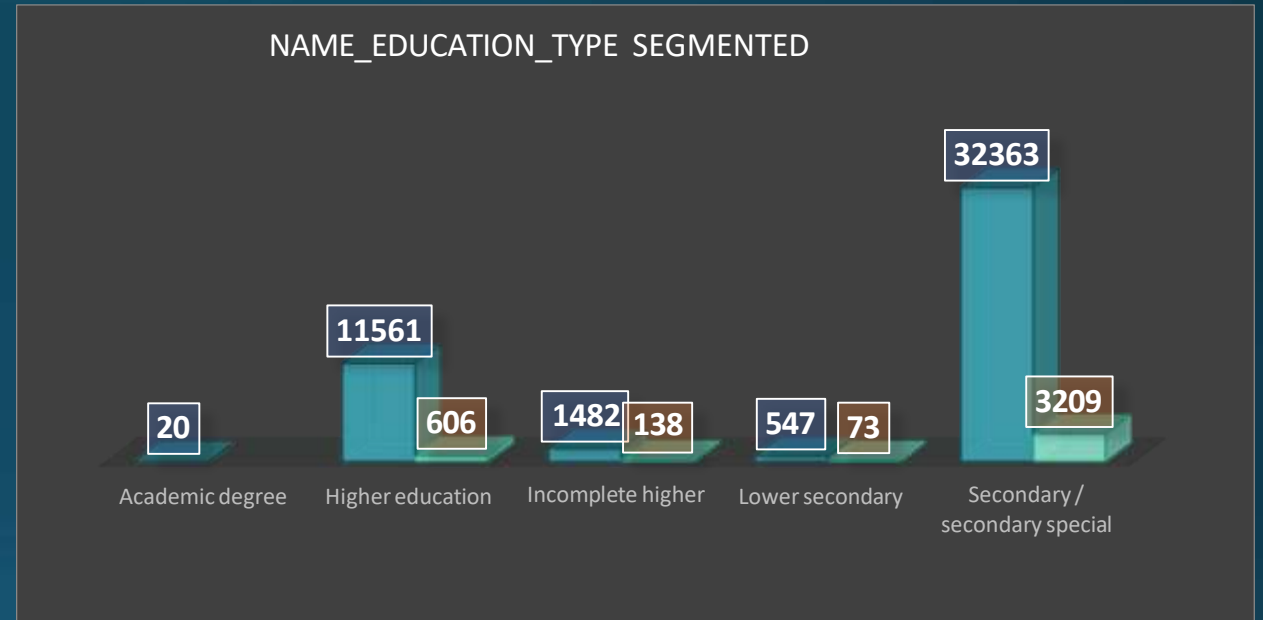


Male are less defaulters compared to Female.

Univariate/Segmented Univariate Analysis

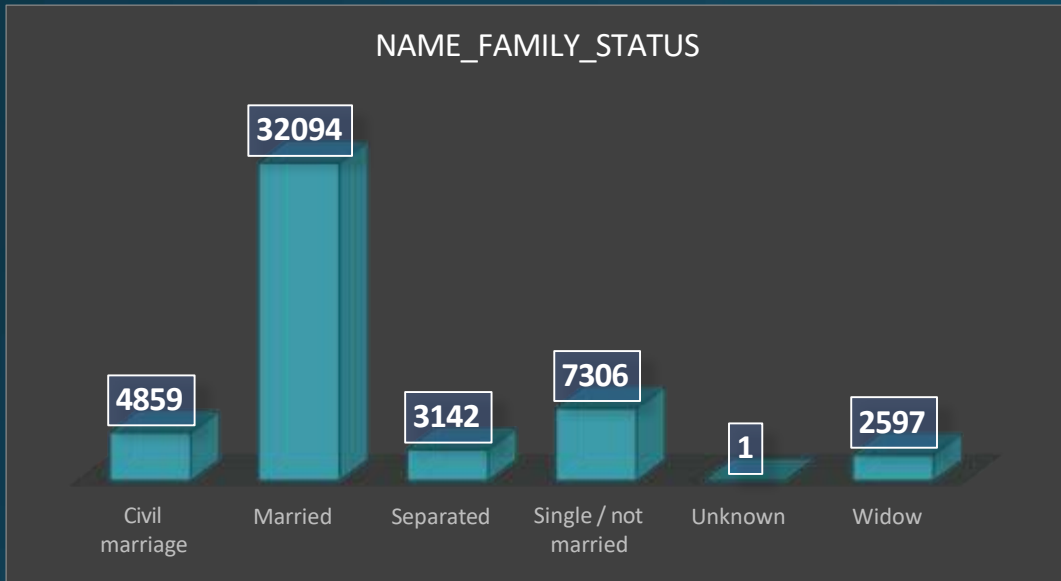


The numbers of loans taken by Clients with Secondary special Education is the highest and Academic degree is the lowest

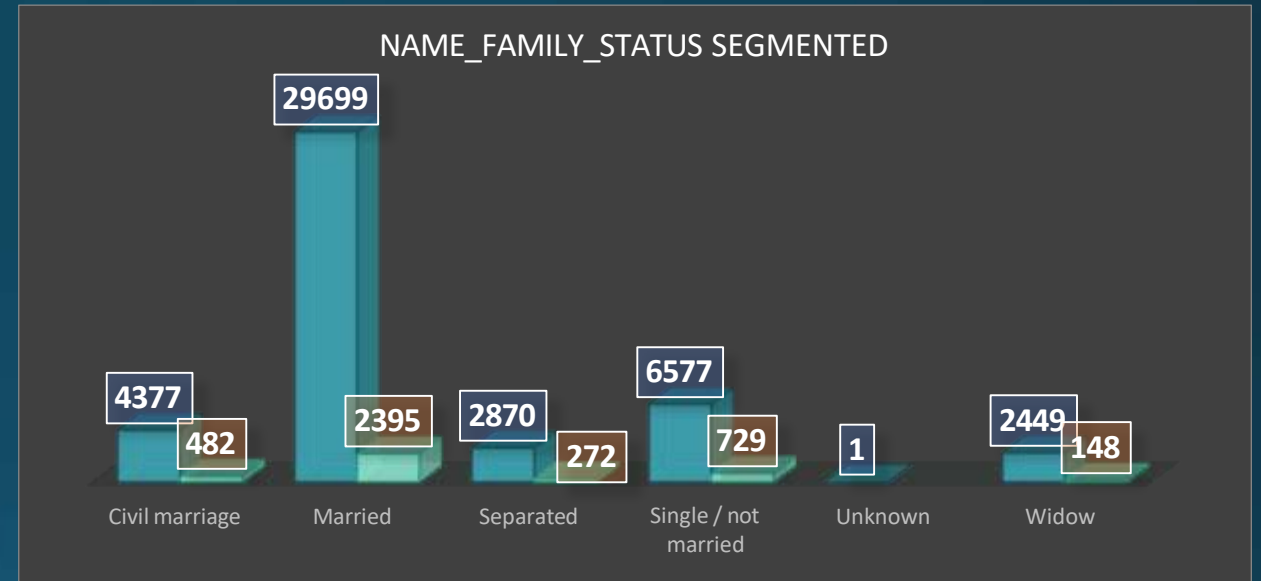


Least default: Academic degree
Highest default: Secondary special

Univariate/Segmented Univariate Analysis

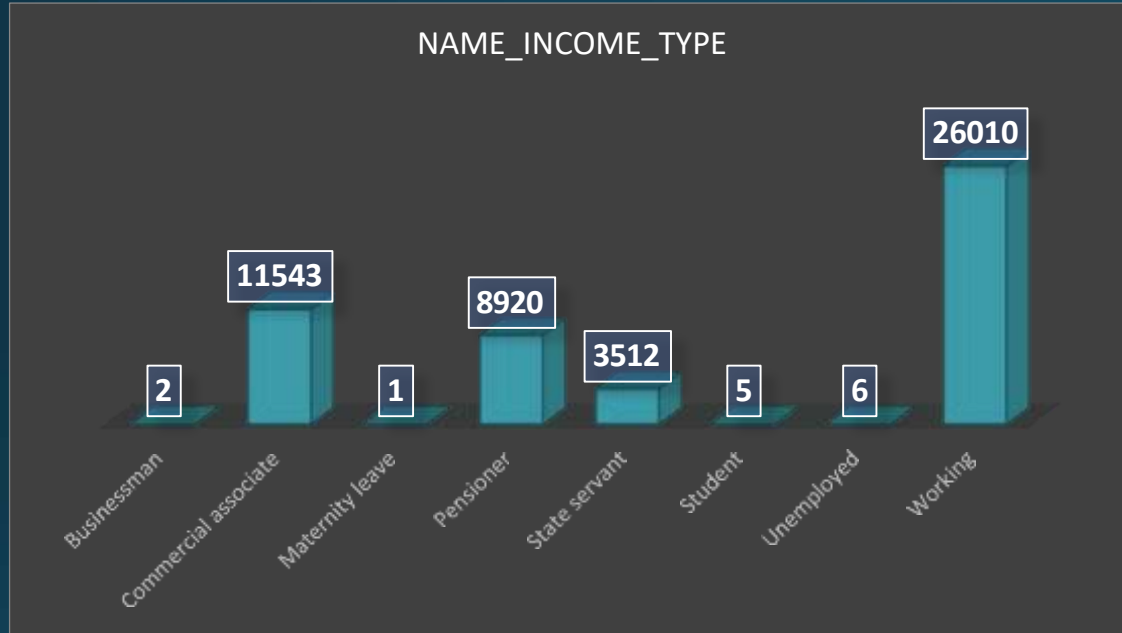


The number of loans taken by Married clients are the highest and clients who are widows are the least if we ignore unknown.

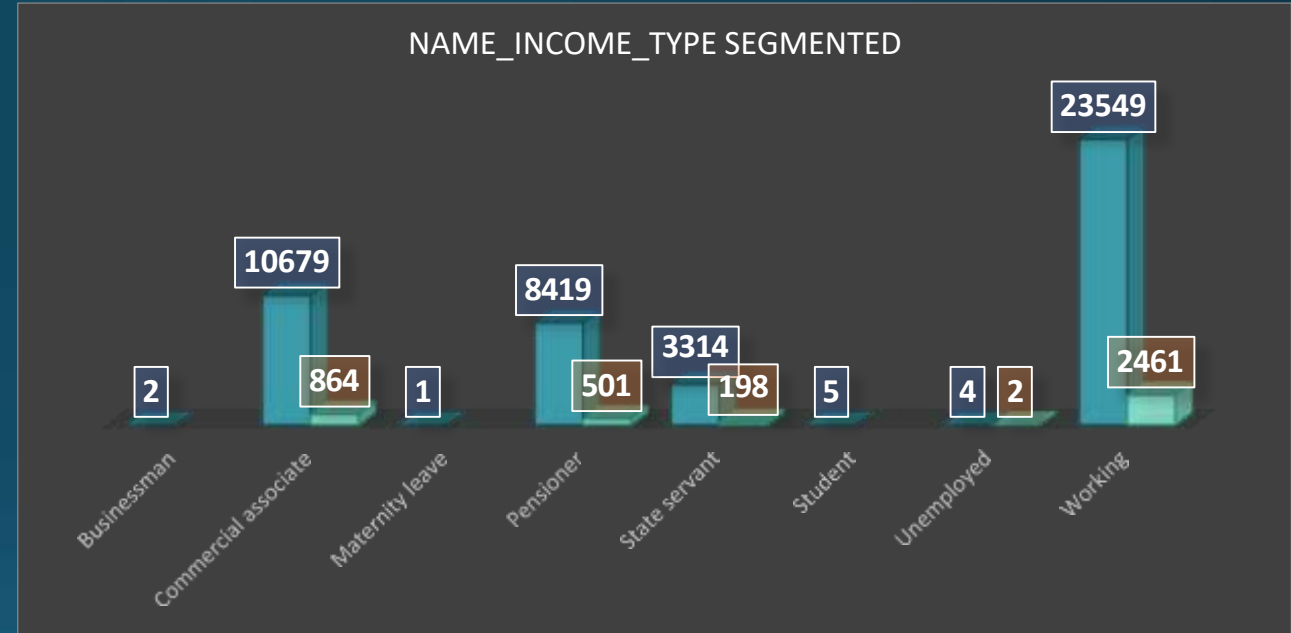


Least Defaulter: Widow
Highest Defaulter : Married

Univariate/Segmented Univariate Analysis



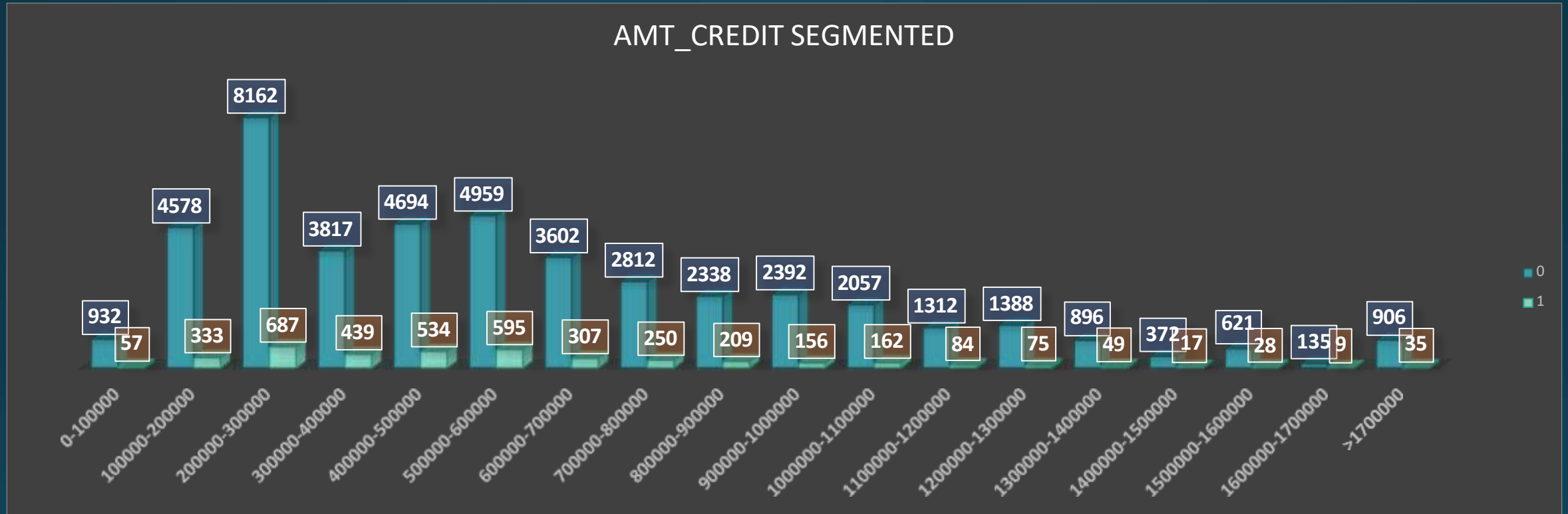
Bank target those groups whose income type is working.



Least default: Client who is Businessman or student or at Maternity leave.

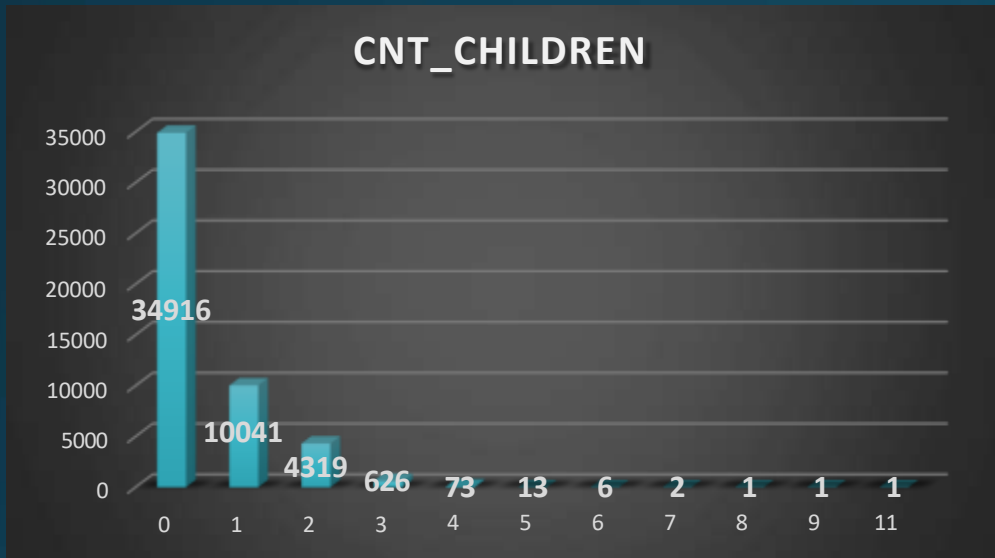
Highest default: Client who is working

Univariate/Segmented Univariate Analysis

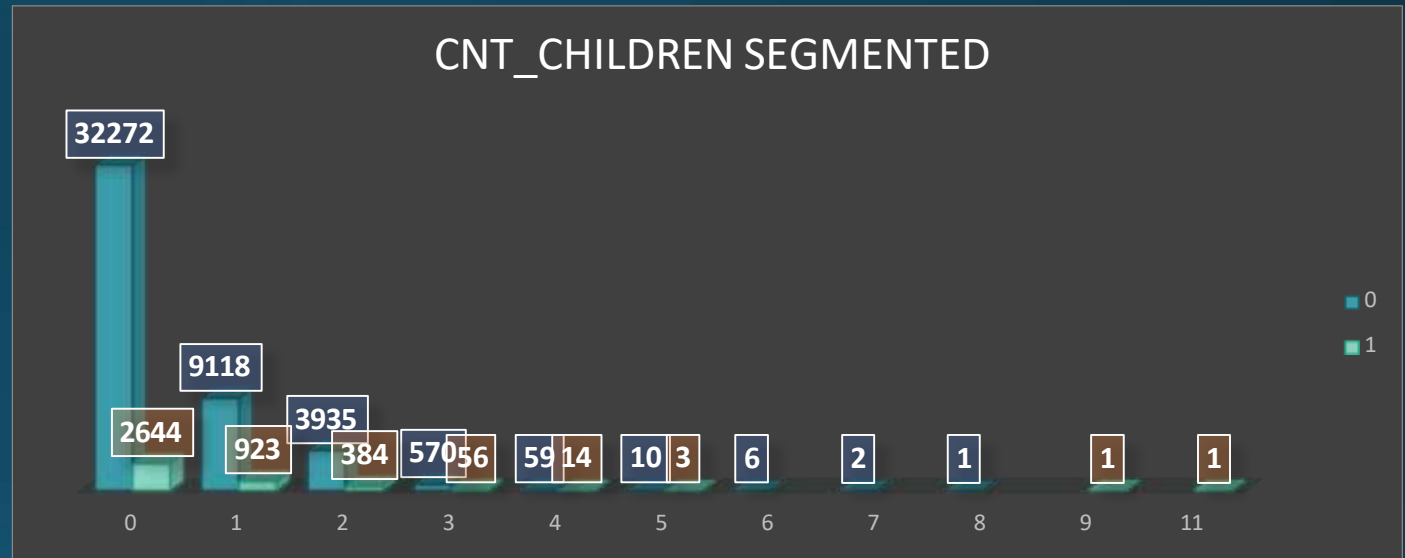


Majority of the Clients took the loan between 2L – 3L.

Univariate/Segmented Univariate Analysis

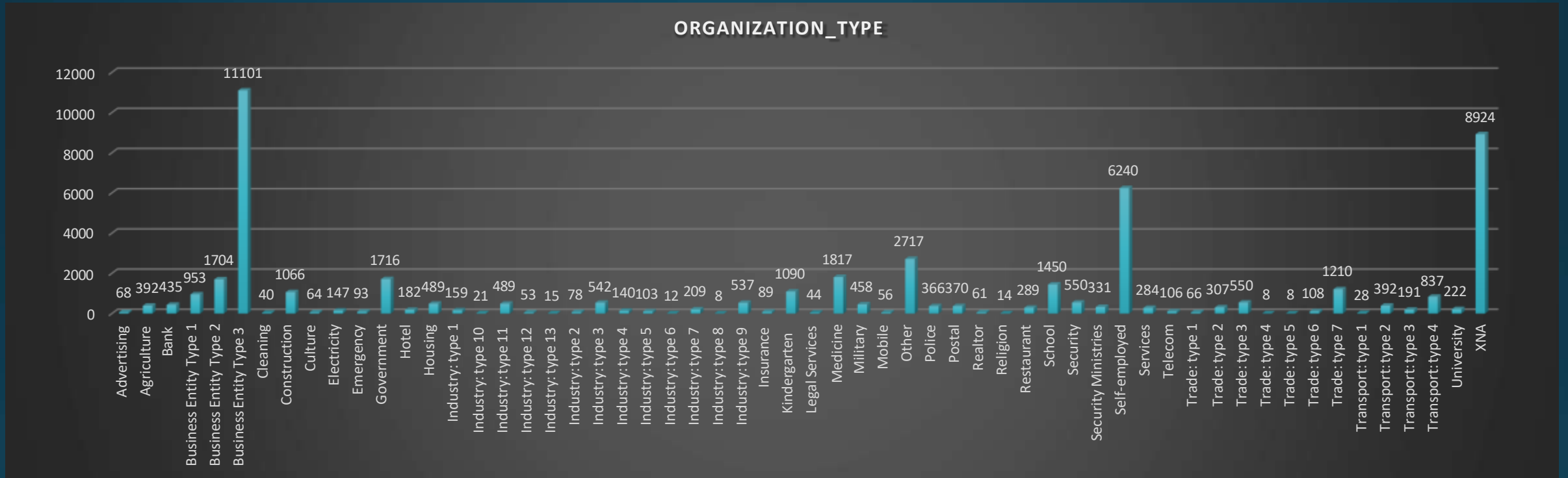


The highest number of loans are taken by Clients who does not have a child



As number of children increases, number of client who took loan decreases.

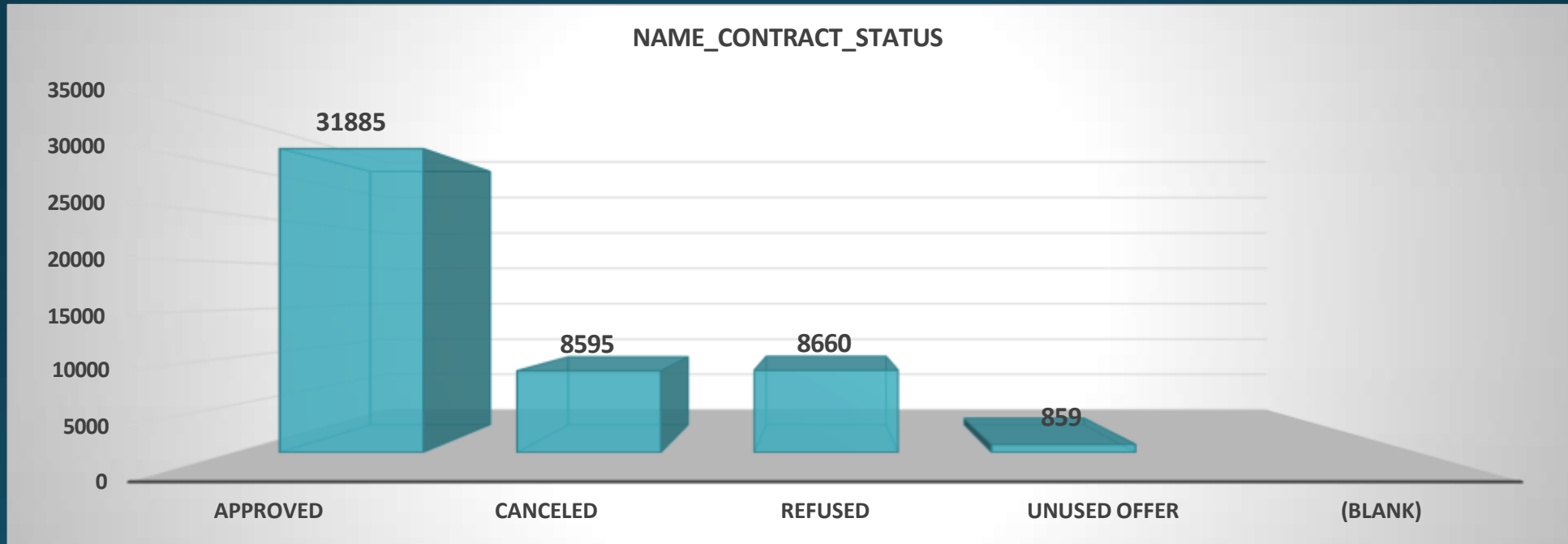
Univariate/Segmented Univariate Analysis



Clients who are working in business Entity type of Organization took the highest number of loans.

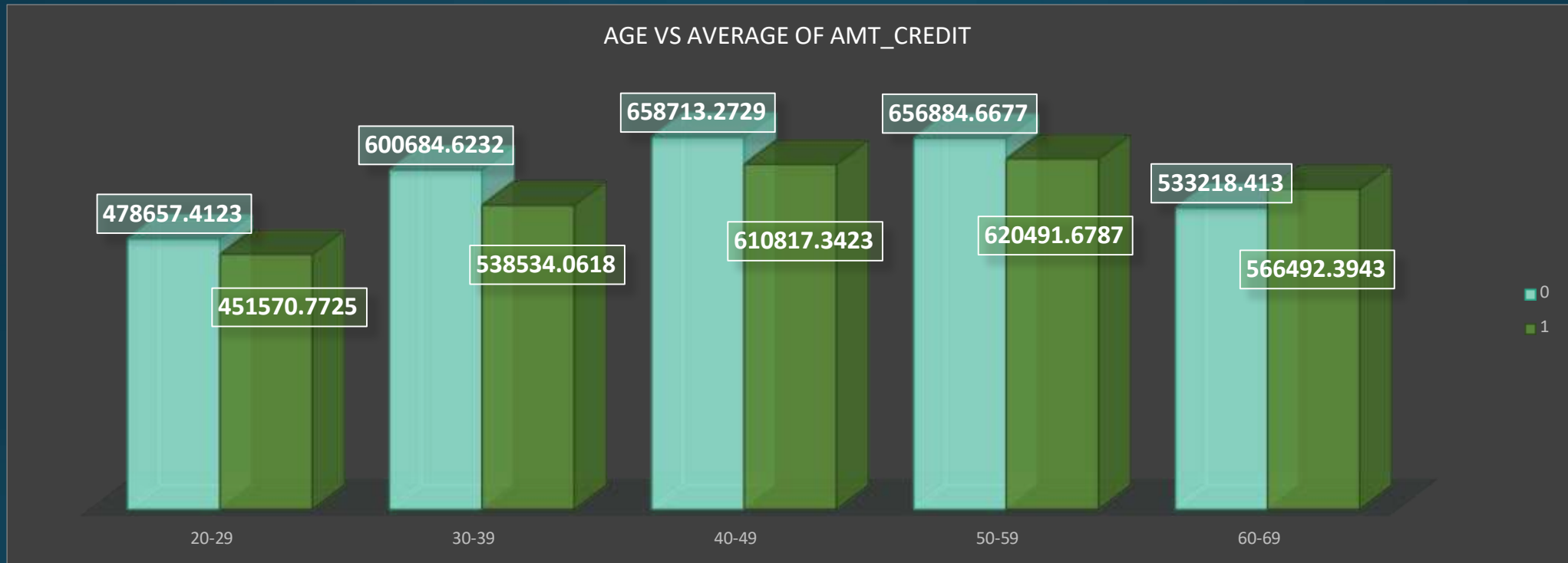
Univariate/Segmented Univariate Analysis

Previous_application datasets



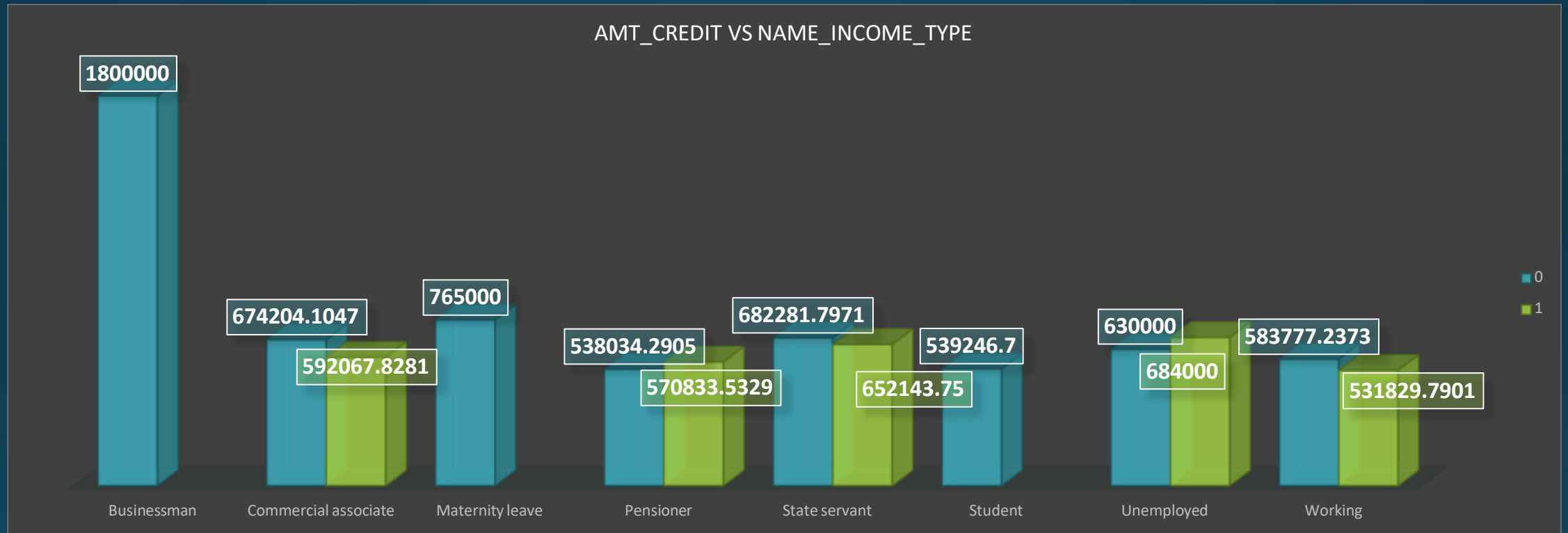
More number of Clients were approved for loans previously.

Bivariate Analysis



Age group 40-49 took the highest amount of loan but age group 50-59 are defaulter with highest amount of loan.

Bivariate Analysis



As we see Businessman took the highest amount of loan and did the payment on time. Clients who are unemployed have highest amount of loan which they didn't repay on time.

Correlations

Task E: 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.

Top Correlation Coefficients for Payment difficulties are:-

Correlation between Columns	Value
AMT_CREDIT - AMT_GOODS_PRICE	0.982267963
OBS_60_CNT_SOCIAL_CIRCLE - OBS_30_CNT_SOCIAL_CIRCLE	0.998065853
DEF_60_CNT_SOCIAL_CIRCLE - DEF_30_CNT_SOCIAL_CIRCLE	0.89051161
REG_REGION_NOT_WORK_REGION - LIVE_REGION_NOT_WORK_REGION	0.806743886
REG_CITY_NOT_WORK_CITY - LIVE_CITY_NOT_WORK_CITY	0.783754676
AMT_CREDIT - AMT_ANNUIITY	0.749665201
AMT_GOODS_PRICE - AMT_ANNUIITY	0.74950403

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EM	DAYS_REG	DAYS_ID	HOUR_AP	REG_REG	REG_REG	LIVE_REG	REG_CITY	REG_CITY	LIVE_CITY	OBS_30	DEF_30	OBS_60	DEF_60	CNT_CHILDREN
CNT_CHILDREN	1	0.010110177	0.007601905	0.029172977	-0.001079665	-0.020359154	0.2496732	-0.1893	0.15211	-0.0424	-0.0069	-0.0157	-0.0057	-0.0004	0.00175	0.04892	0.05818	0.01793	-0.0136	0.01515	-0.0185	-0.0185
AMT_INCOME_TOTAL	0.010110177	1	0.015271444	0.018004594	0.013269502	-0.006180303	0.009033662	-0.0116	-0.0096	-0.0091	0.01448	0.00059	0.00167	0.00223	-0.006	-0.0104	-0.008	-0.0113	-0.008	-0.0112	-0.0067	-0.0067
AMT_CREDIT	0.007601905	0.015271444	1	0.749665201	0.982267963	0.067775624	-0.142506035	0.01604	-0.0428	-0.0438	0.0454	0.00646	0.02354	0.0346	-0.0523	-0.0391	-0.0067	0.03347	-0.0249	0.03444	-0.029	0.029
AMT_ANNUITY	0.029172977	0.018004594	0.749665201	1	0.74950403	0.073123998	-0.008751713	-0.0796	0.02158	-0.0213	0.04489	0.03176	0.06569	0.07424	-0.0177	0.00218	0.01356	0.01382	-0.0345	0.0141	-0.0405	0.0405
AMT_GOODS_PRICE	-0.001079665	0.013269502	0.982267963	0.74950403	1	0.076635488	-0.141005898	0.02024	-0.0433	-0.0497	0.05746	0.00708	0.02502	0.03542	-0.0527	-0.044	-0.0131	0.03272	-0.0191	0.03388	-0.0206	0.0206
REGION_POPULATION_RELATIVE	-0.020359154	-0.006180303	0.067775624	0.073123998	0.076635488	1	-0.016468731	0.00774	-0.0461	-0.0051	0.15605	-0.0031	0.01917	0.05954	-0.0349	-0.0433	-0.0252	-0.0089	0.02781	-0.0071	0.02714	0.02714
DAYS_BIRTH	0.2496732	0.009033662	-0.142506035	-0.008751713	-0.141005898	-0.016468731	1	-0.5815	0.28844	0.2479	0.05789	0.03961	0.07551	0.05449	0.14911	0.22635	0.1434	-0.0112	-0.0208	-0.0126	-0.0258	0.0258
DAYS_EMPLOYED	-0.189324184	-0.011555963	0.016039571	-0.079556008	0.020235348	0.007742909	-0.581479041	1	-0.1887	-0.2301	-0.0521	-0.0353	-0.0849	-0.0723	-0.0882	-0.2463	-0.2006	0.00352	0.02986	0.00421	0.02389	-0.02389
DAYS_REGISTRATION	0.152113117	-0.009561152	-0.042844404	0.021581654	-0.043320226	-0.046130288	0.288437837	-0.1887	1	0.09029	-0.0578	0.01585	0.01639	0.01358	0.05557	0.10076	0.06982	-0.0058	0.001	-0.0059	-0.0064	0.0064
DAYS_ID_PUBLISH	-0.042360717	-0.009122006	-0.043771901	-0.02132109	-0.049723232	-0.005118563	0.247896571	-0.2301	0.09029	1	0.00552	0.02415	0.04111	0.02957	0.0641	0.08301	0.03844	-0.0273	-0.0284	-0.0262	-0.0279	0.0279
HOUR_APPR_PROCESS_START	-0.006884357	0.014482013	0.045396384	0.044891881	0.057462759	0.156049669	0.057891695	-0.0521	-0.0578	0.00552	1	0.04942	0.07615	0.06606	0.00552	0.0032	-0.0118	-0.0197	0.01767	-0.0195	0.01752	-0.01752
REG_REGION_NOT_LIVE_REGION	-0.015713279	0.000594885	0.006456715	0.031759358	0.007079035	-0.003105241	0.039614727	-0.0353	0.01585	0.02415	0.04942	1	0.5255	0.10053	0.33817	0.14759	-0.0037	-0.032	0.00849	-0.032	0.00582	-0.00582
REG_REGION_NOT_WORK_REGION	-0.005665093	0.001665752	0.023536318	0.065686571	0.025016178	0.019170075	0.075512807	-0.0849	0.01639	0.04111	0.07615	0.5255	1	0.80674	0.18375	0.22868	0.16908	-0.0321	0.00152	-0.0316	0.00493	0.00493
LIVE_REGION_NOT_WORK_REGION	-0.000389253	0.002228043	0.034604167	0.074238732	0.035424194	0.059536379	0.054493345	-0.0723	0.01358	0.02957												

Correlations

Top Correlation Coefficients for Re-payers are:-

Correlation between Columns	Value
OBS_60_CNT_SOCIAL_CIRCLE - OBS_30_CNT_SOCIAL_CIRCLE	0.998357563
AMT_GOODS_PRICE - AMT_CREDIT	0.986051701
LIVE_REGION_NOT_WORK_REGION - REG_REGION_NOT_WORK_REGION	0.861374946
DEF_60_CNT_SOCIAL_CIRCLE - DEF_30_CNT_SOCIAL_CIRCLE	0.850995792
REG_CITY_NOT_WORK_CITY - LIVE_CITY_NOT_WORK_CITY	0.825358079
AMT_ANNUITY - AMT_GOODS_PRICE	0.774006842
AMT_ANNUITY - AMT_CREDIT	0.770772818

Correlations

Top Correlation Coefficients for Re-payers are:-

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	HOURL_APPR_PROCESS_START	REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	OBS_30_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY	AMT_REQ_CREDIT_BUREAU_WEEK	AMT_REQ_CREDIT_BUREAU_MONTH	AMT_REQ_CREDIT_BUREAU_QUARTER	AMT_REQ_CREDIT_BUREAU_YEAR
CNT_CHILDREN	1	0.036319722	0.005705458	0.02638336	0.001046405	-0.024912809	0.335876269	-0.243591518	0.183072478	-0.03253722	-0.0053	-0.0104	0.0138	0.0217	0.0201	0.071	0.0679	0.0162	-0.0028	0.0163	-0.0033	0.0026	0.0012	0.0043	-0.0116	-0.0047	-0.035734
AMT_INCOME_TOTAL	0.036319722	1	0.377965752	0.451135167	0.383650216	0.181941261	0.073769425	-0.162702675	0.06893375	0.032286356	0.0854	0.0789	0.1571	0.1477	0.0099	0.0152	0.0197	-0.033	-0.032	-0.033	-0.0325	0.0081	0.0095	0.0095	0.0749	0.0158	0.031323
AMT_CREDIT	0.005705458	0.377965752	1	0.770772818	0.986051701	0.095539444	-0.051084182	-0.077367219	0.008053758	-0.00829019	0.0565	0.0278	0.0561	0.0544	-0.0214	-0.014	0.004	0.0009	-0.0135	0.0012	-0.0186	4E-05	0.0135	0.0054	0.064	0.0268	-0.031568
AMT_ANNUITY	0.02638336	0.451135167	0.770772818	1	0.774006842	0.11727925	0.009911417	-0.113005288	0.03460901	0.00942697	0.0536	0.0462	0.0825	0.0749	-0.0053	0.0016	0.0112	-0.01	-0.0197	-0.0097	-0.023	0.0101	0.0092	0.0189	0.038	0.0101	-0.004173
AMT_GOODS_PRICE	0.001046405	0.383650216	0.986051701	0.774006842	1	0.098899174	-0.048684402	-0.075069056	0.011016938	-0.00944125	0.0651	0.0304	0.0575	0.0547	-0.0204	-0.0145	0.0029	0.0006	-0.0152	0.0009	-0.0197	0.0008	0.0137	0.0058	0.0658	0.0276	-0.034352
REGION_POPULATION_RELATIVE	-0.024912809	0.181941261	0.095539444	0.11727925	0.098899174	1	-0.030435419	-0.006610653	-0.058501361	-0.00223629	0.1676	-0.0032	0.0631	0.0874	-0.0461	-0.0383	-0.0113	-0.0191	0.0089	-0.018	0.0033	-0.0031	-0.0003	0.0026	0.0707	-0.0097	0.004652
DAYS_BIRTH	0.335876269	0.073769425	-0.05108418	0.009911417	-0.048684402	-0.030435419	1	-0.615289978	0.335028046	0.270073313	0.0964	0.0604	0.0959	0.0699	0.1833	0.2361	0.1492	0.0123	0.0007	0.0123	0.0022	0.0015	0.002	-0.0024	-0.0025	-0.0215	-0.070267
DAYS_EMPLOYED	-0.243591518	-0.162702675	-0.07736722	-0.113005288	-0.075069056	-0.006610653	-0.615289978	1	-0.204370881	-0.27222439	-0.0924	-0.0364	-0.1073	-0.0956	-0.0926	-0.2541	-0.2177	0.0057	0.017	0.0055	0.0165	-0.0043	0.0016	-0.0065	-0.033	0.0146	0.044183
DAYS_REGISTRATION	0.183072478	0.06893375	0.008053758	0.03460901	0.011016938	-0.058501361	0.335028046	-0.204370881	1	0.103548902	-0.0024	0.0279	0.0347	0.0233	0.0678	0.0916	0.0612	0.011	0.0034	0.0113	0.0063	-0.0037	-0.0034	0.0007	-0.0107	0.0031	-0.02296
DAYS_ID_PUBLISH	-0.03253722	0.032286356	-0.00829019	0.00942697	-0.00944125	-0.002236288	0.270073313	-0.27222439	0.103548902	1	0.038	0.0332	0.0478	0.0338	0.0751	0.102	0.0633	-0.0119	0.0023	-0.0122	0.0026	0.0028	0.0035	-0.0047	-0.0132	-0.0246	-0.044692
HOURL_APPR_PROCESS_START	-0.005272551	0.08543156	0.056524809	0.053564989	0.065133303	0.167612161	0.09638927	-0.092357978	-0.002396446	0.037971336	1	0.0512	0.0736	0.0598	0.0197	0.0269	0.0151	-0.008	-0.0024	-0.008	-0.0061	-0.0074	0.0103	-0.0067	0.0288	-0.0005	-0.0250
REG_REGION_NOT_LIVE_REGION	-0.010383386	0.078942904	0.027812773	0.046175655	0.030367622	-0.003185217	0.060427036	-0.03641311	0.027899954	0.033228477	0.0512	1	0.4496	0.0805	0.3351	0.1426	0.0035	-0.0151	-0.0083	-0.0151	-0.0094	-0.0025	-0.0058	-0.0018	-0.0086	-0.0003	-0.019525
REG_REGION_NOT_WORK_REGION	0.013794691	0.157051351	0.05603668	0.082502425	0.057545564	0.063145413	0.095915233	-0.107331487	0.034657988	0.047811506	0.0736	0.4496	1	0.8614	0.1519	0.2368	0.1922	-0.0252	-0.0089	-0.0254	-0.0137	5E-06	0.0008	0.0033	0.0042	-0.0088	-0.02753
LIVE_REGION_NOT_WORK_REGION	0.021685073	0.147730123	0.05443061	0.074870093	0.054659311	0.087419766	0.06988551	-0.095573749	0.023280394	0.033751626	0.0598	0.0805	0.8614	1	0.0216	0.1839	0.2338	-0.0202	-0.0069	-0.0204	-0.012	0.0025	0.0029	0.0054	0.0099	-0.0124	-0.022490
REG_CITY_NOT_LIVE_CITY	0.020101944	0.009927686	-0.02137243	-0.005276721	-0.020436382	-0.046089149	0.183304735	-0.092557531	0.067811428	0.075080051	0.0197	0.3351	0.1519	0.0216	1	0.4414	0.0292	-0.0053	0.0055	-0.0055	0.0055	0.0005	8E-05	-0.0011	-0.0136	-2E-05	-0.006661
REG_CITY_NOT_WORK_CITY	0.070971057	0.015150008	-0.01400736	0.001628799	-0.01449892	-0.038253612	0.236134428	-0.254060105	0.091595217	0.102001817	0.0269	0.1426	0.2368	0.1839	0.4414	1	0.8254	-0.006	0.001	-0.006	0.0033	0.0043	-0.0002	0.0022	-0.0124	-0.0039	-0.011958
LIVE_CITY_NOT_WORK_CITY	0.067882194	0.019663673	0.00397996	0.011203272	0.002861594	-0.011278612	0.149167938	-0.217741277	0.061159259	0.063319024	0.0151	0.0035	0.1922	0.2338	0.0292	0.8254	1	-0.0052	-0.0022	-0.0051	-0.0002	0.004	-0.0012	0.0024	-0.0046	-0.0052	-0.012945
OBS_30_CNT_SOCIAL_CIRCLE	0.016180299	-0.033045993	0.000876364	-0.009992103	0.000634386	-0.01906908	0.012287026	0.005650192	0.010977833	-0.01185404	-0.008	-0.0151	-0.0252	-0.0202	-0.0053	-0.006	-0.0052	1	0.3062	0.9984	0.2292	0.0024	0.001	-0.0043	0.0082	0.0088	0.034161
DEF_30_CNT_SOCIAL_CIRCLE	-0.00282133	-0.032012977	-0.01350943	-0.019746021	-0.015155074	0.008905591	0.000683769	0.017033326	0.003448989	0.002312725	-0.0024	-0.0083	-0.0089	-0.0069	0.0055	0.001	-0.0022	0.3062	1	0.3086	0.851	-0.0044	0.0037	-0.005	0.0077	0.0054	0.014506
OBS_60_CNT_SOCIAL_CIRCLE	0.016334894	-0.03301707	0.001184762	-0.009675846	0.000856455	-0.018012695	0.01229458	0.005511276	0.011295659	-0.01215588	-0.008	-0.0151	-0.0254	-0.0204	-0.0055	-0.006	-0.0051	0.9984	0.3086	1	0.2313	0.0026	0.0009	-0.0049	0.0081	0.0087	0.034573
DEF_60_CNT_SOCIAL_CIRCLE	-0.003330304	-0.032535174	-0.01856734	-0.023010616	-0.019633991	0.003253593	0.002207122	0.016516022	0.006282428	0.002642424	-0.0061	-0.0094	-0.0137	-0.012	0.0055	0.0033	-0.0002	0.2292	0.851	0.2313	1	-0.0032	0.0028	-0.0057	0.004	0.0083	0.015204
AMT_REQ_CREDIT_BUREAU_HOUR	0.00261709	0.008122955	3.65328E-05	0.0101408	0.000827804	-0.003132124	0.001486295	-0.00429349	-0.003689166	0.002824211	-0.0074	-0.0025	5E-06	0.0025	0.0005	0.0043	0.004	0.0024	-0.0044	0.0026	-0.0032	1	0.2308	0.0121	0.0095	0.0035	0.004095
AMT_REQ_CREDIT_BUREAU_DAY	0.001198938	0.009477681	0.013486353	0.009156839	0.013665416	-0.000338841	0.001983866	0.001618187	-0.00338406	0.003514735	0.0103	-0.0058	0.0008	0.0029	8E-05	-0.0002	-0.0012	0.001	0.0037	0.0009	0.0028	0.2308	1	0.2491	-0.0007	-0.0079	-0.00085
AMT_REQ_CREDIT_BUREAU_WEEK	0.004327432	0.009487825	0.005374948	0.018909774	0.005848551	0.002644642	-0.002401164	-0.00648155	0.000859813	-0.00466588	-0.0067	-0.0018	0.0033	0.0054	-0.0011	0.0022	0.0024	-0.0043	-0.005	-0.0049	-0.0057	0.0121	0.2491	1	-0.0106	-0.0146	0.02473
AMT_REQ_CREDIT_BUREAU_MONTH	-0.011607819	0.074854679	0.063975389	0.037985476	0.065821049	0.070736631	-0.002452976	-0.032954589	-0.010724839	-0.01323263	0.0288	-0.0086	0.0042	0.0099	-0.0136	-0.0124	-0.0046	0.0082	0.0077	0.0081	0.004	0.0095	-0.0007	-0.0106	1	0.0119	0.01931
AMT_REQ_CREDIT_BUREAU_QUARTER	-0.00473083	0.015777535	0.026793294	0.010059213	0.027627409	-0.009694599	-0.021522968	0.014577401	0.003127351	-0.02458808	-0.0005	-0.0003	-0.0088	-0.0124	-2E-05	-0.0039	-0.0052	0.0088	0.0054	0.0087	0.0083	0.0035	-0.0079	-0.0146	0.0119	1	0.121744
AMT_REQ_CREDIT_BUREAU_YEAR	-0.035734888	0.031323516	-0.03156833	-0.004173747	-0.034352324	0.004652396	-0.070267716	0.044183816	-0.02296176	-0.04469288	-0.0251	-0.0195	-0.0275	-0.0225	-0.0067	-0.012	-0.0129	0.0342	0.0145	0.0346	0.0152	0.0041	-0.0009	0.0247	0.0193	0.1217	

Conclusions

- Most of the clients are loan re-payers.
- The Bank generally lends more loan to Female as compared to Male but Male are less defaulters compared to Female.
- As age and experience increases , chances of defaulter decreases.
- Most of the clients are taking cash loans.
- Educated clients tend to less defaulter compared to clients with lower education such as secondary special education so Bank should prefer clients with having such education status.
- As number of children increases, number of client who take loan decreases.
- The Bank should be more cautious when lending money to clients who are unemployed because they are the most defaulters with highest amount of credit.
- As age increases amount taken by Clients are considerably high but with higher age defaulter percentage is lower. These are least risky and more profitable for Bank.

Conclusions

We have lots of data. Few of the pivot table and charts are not supported by Google sheet. Please download the file in Excel(.XLSX) form

Google Drive Link: -

https://drive.google.com/drive/folders/1Xhxvfmntnjbb_E3Cdle-LRGo_edRq9mX?usp=sharing