Bank Loan Case Study

RAHUL M RAMCHANDANI

Project description

- Addressing the challenge of loan defaults among customers with limited credit history, this project utilizes Exploratory Data Analysis (EDA) to uncover patterns. The goal is to enhance the loan approval process, ensuring eligible applicants are not rejected while minimizing default risks.

When a customer applies for a loan, company faces two risks:

- If the applicant can repay the loan but is not approved, the company loses business.
- If the applicant cannot repay the loan and is approved, the company faces a financial loss.

When a customer applies for a loan, there are four possible outcomes:

- Approved: The company has approved the loan application.
- Cancelled: The customer cancelled the application during the approval process.
- Refused: The company rejected the loan.
- Unused Offer: The loan was approved but the customer did not use it.

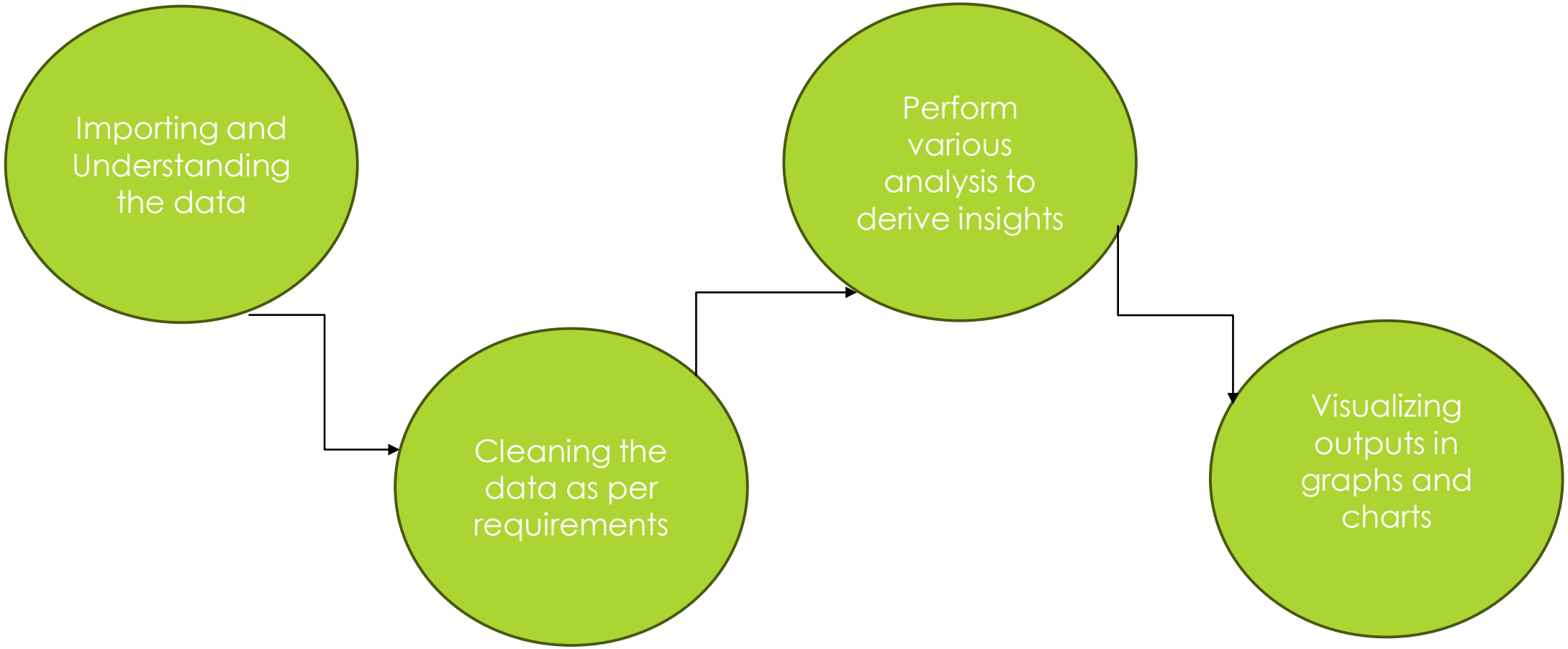
As a Data Analyst my goal in this project is to use EDA to understand how customer attributes and loan attributes influence the likelihood of default.

Tech Stack used

- **Microsoft Excel 2021**-Excel is a powerful spreadsheet software used for data organization, analysis, and visualization, facilitating efficient numerical calculations and creating dynamic charts.

- **Microsoft PowerPoint**-PowerPoint is a presentation software that enables users to create visually engaging slideshows, making it an essential tool for conveying information, ideas, and messages in a professional and compelling manner.

Approach


Importing and Understanding the data
Cleaning the data as per requirements
Perform various analysis to derive insights
Visualizing outputs in graphs and charts

Tasks

- A. Identify Missing Data and Deal with it Appropriately: Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.
- B. Identify Outliers in the Dataset: Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.
- C. Analyze Data Imbalance: Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.
- D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:
  Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.
- E. Identify Top Correlations for Different Scenarios: 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.

Task A: Identify Missing Data

Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.

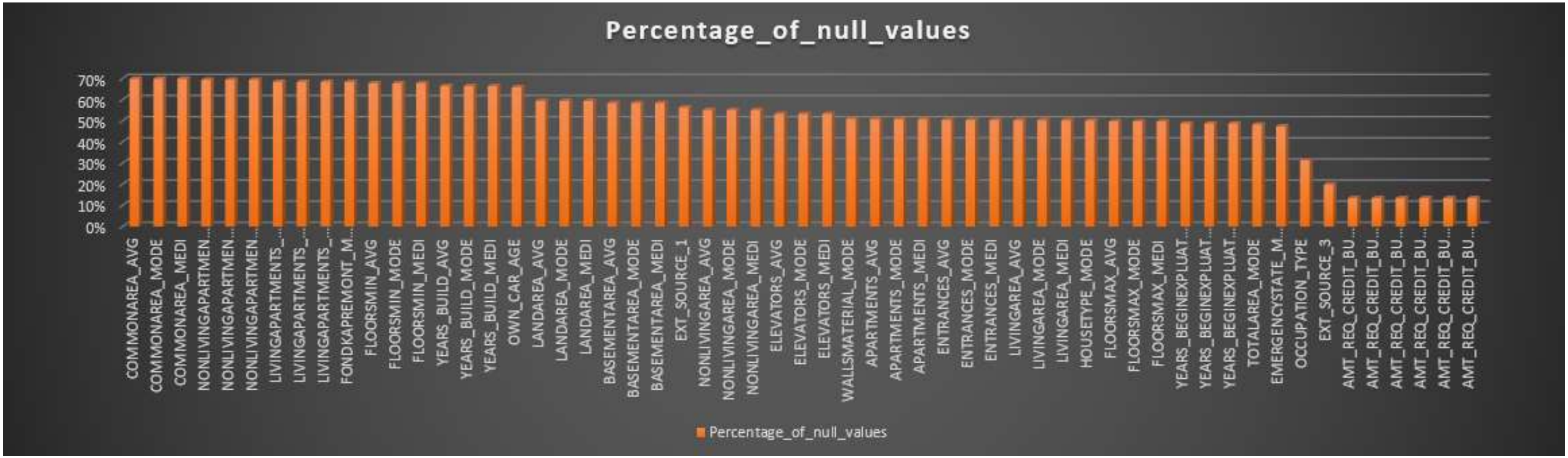
- Using COUNT, IF and ISBLANK function to get number of null values for each column.
- Then we will calculate percentage of null values for each column.
- Using Transpose function we will convert rows to columns.
- We will drop the columns which has more than or equal to 50% null values
- We will drop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

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%

- These are the columns which has null values. These columns need to be dropped. Check the full list in the Null values chart excel file

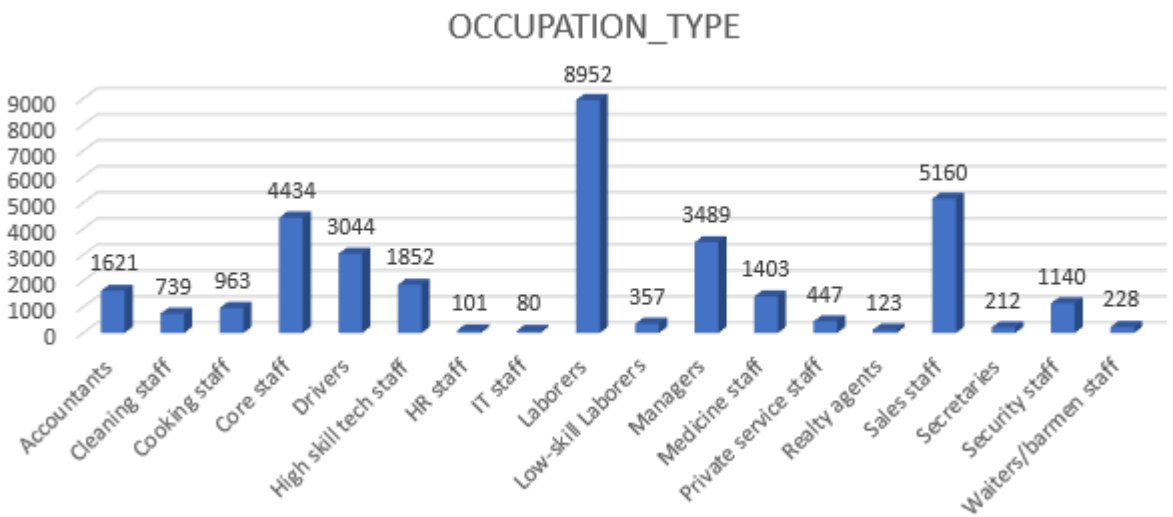
Column name	no_of_null_values	Percentage_of_null_values
FLAG_MOBIL	0	0%
FLAG_EMP_PHONE	0	0%
FLAG_WORK_PHONE	0	0%
FLAG_CONT_MOBILE	0	0%
FLAG_PHONE	0	0%
FLAG_EMAIL	0	0%
CNT_FAM_MEMBERS	1	0%
REGION_RATING_CLIENT	0	0%
REGION_RATING_CLIENT_W_CITY	0	0%
EXT_SOURCE_2	126	0%
EXT_SOURCE_3	9944	20%
YEARS_BEGINEXPLUATATION_AVG	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 contain irrelevant data for analysis and are not needed so these columns need to be dropped.

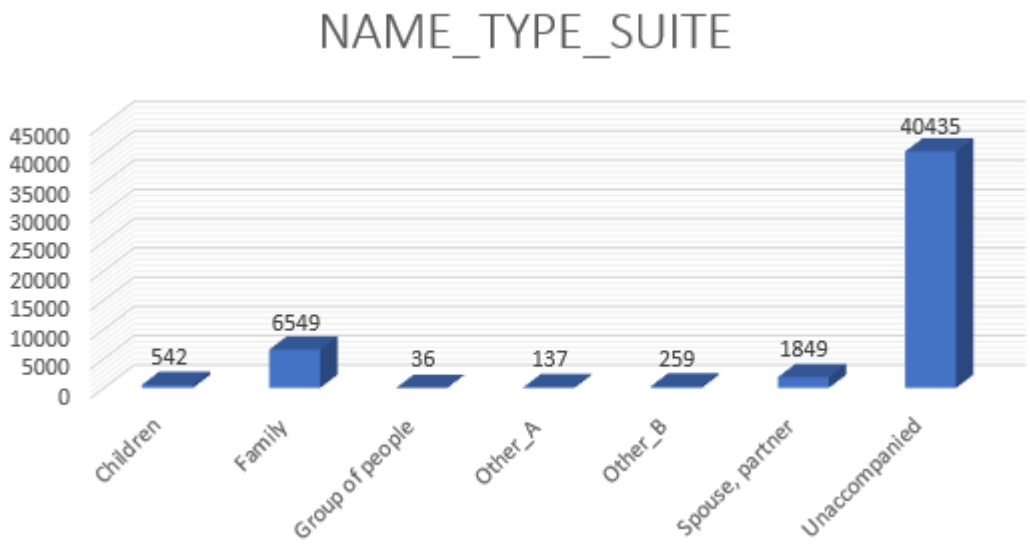

Percentage_of_null_values
70%
60%
50%
40%
30%
20%
10%
0%
COMMONAREA_AVG
COMMONAREA_MODE
COMMONAREA_MEDI
NONLIVINGAPARTMEN...
NONLIVINGAPARTMEN...
NONLIVINGAPARTMEN...
LIVINGAPARTMENTS_...
LIVINGAPARTMENTS_...
LIVINGAPARTMENTS_...
FONDKAPREMONT_M...
FLOORSMIN_AVG
FLOORSMIN_MODE
FLOORSMIN_MEDI
YEARS_BUILD_AVG
YEARS_BUILD_MODE
YEARS_BUILD_MEDI
OWN_CAR_AGE
LANDAREA_AVG
LANDAREA_MODE
LANDAREA_MEDI
BASEMENTAREA_AVG
BASEMENTAREA_MODE
BASEMENTAREA_MEDI
EXT_SOURCE_1
NONLIVINGAREA_AVG
NONLIVINGAREA_MODE
NONLIVINGAREA_MEDI
ELEVATORS_AVG
ELEVATORS_MODE
ELEVATORS_MEDI
WALLSMATERIAL_MODE
APARTMENTS_AVG
APARTMENTS_MODE
APARTMENTS_MEDI
ENTRANCES_AVG
ENTRANCES_MODE
ENTRANCES_MEDI
LIVINGAREA_AVG
LIVINGAREA_MODE
LIVINGAREA_MEDI
HOUSETYPE_MODE
FLOORSMAX_AVG
FLOORSMAX_MODE
FLOORSMAX_MEDI
YEARS_BEGINEXPLUAT...
YEARS_BEGINEXPLUAT...
YEARS_BEGINEXPLUAT...
TOTALAREA_MODE
EMERGENCYSTATE_M...
OCCUPATION_TYPE
EXT_SOURCE_3
AMT_REQ_CREDIT_BU...
AMT_REQ_CREDIT_BU...
AMT_REQ_CREDIT_BU...
AMT_REQ_CREDIT_BU...
AMT_REQ_CREDIT_BU...
AMT_REQ_CREDIT_BU...
Percentage_of_null_values

Missing Data

- Mode Imputations-
  1. OCCUPATION_TYPE

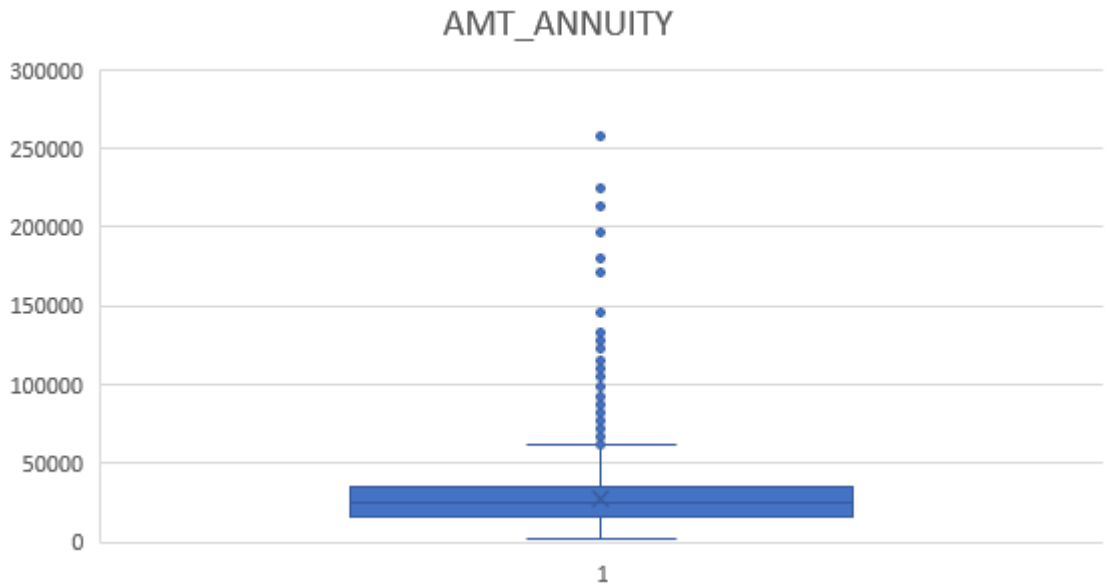


2. NAME_TYPE_SUITE

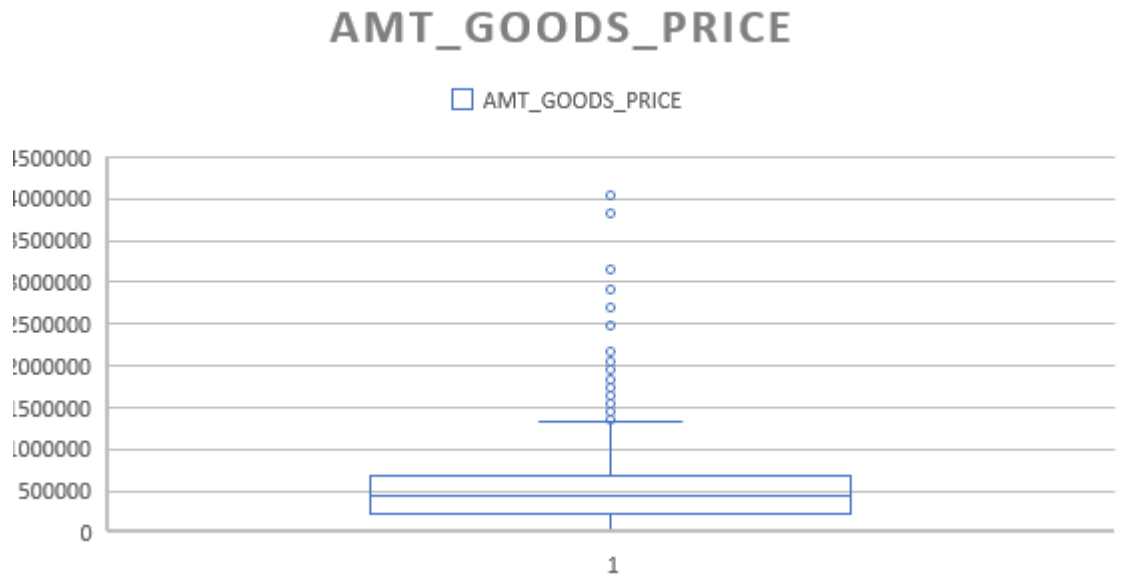


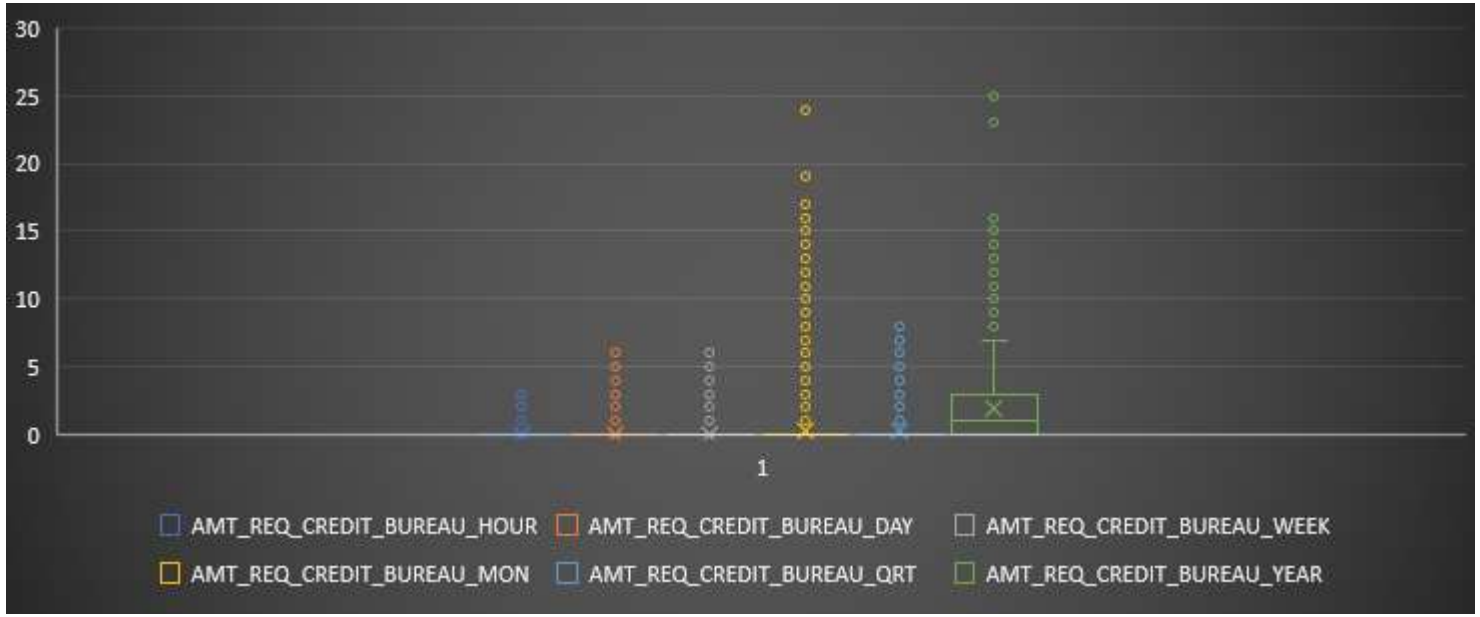
Missing Data

- Median Imputations-
  1. AMT_ANNUITY



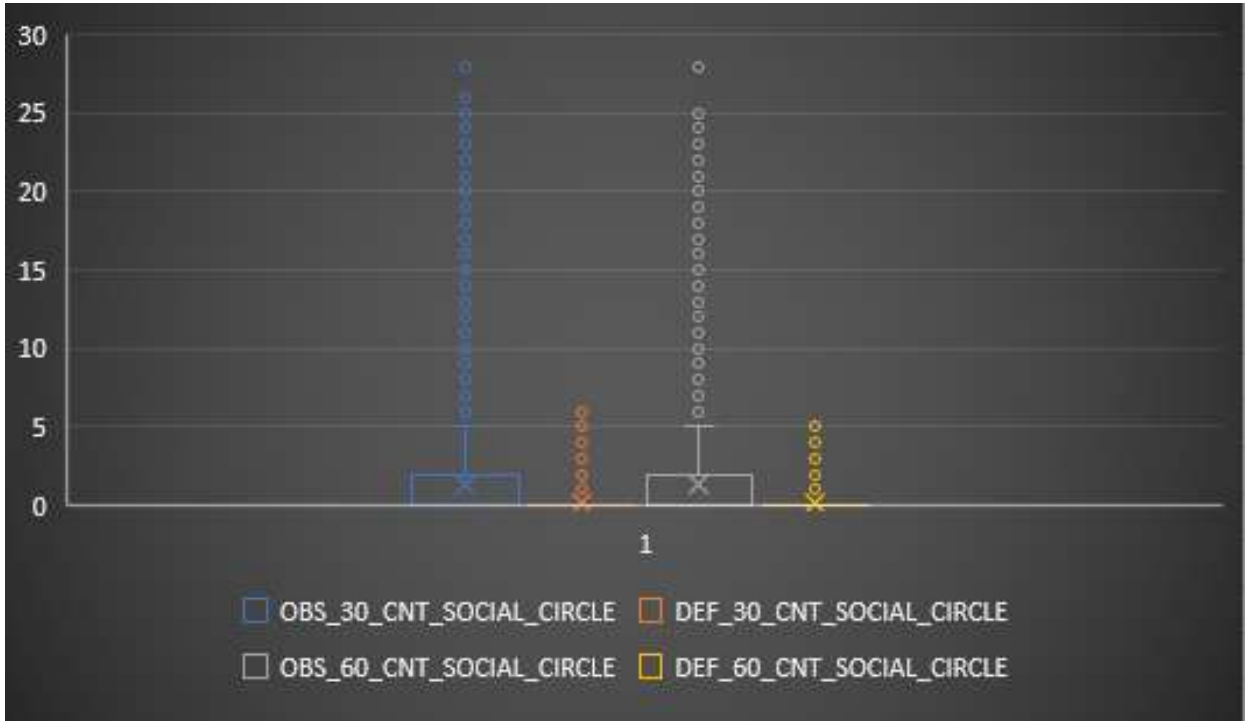
2. AMT_GOODS_PRICE





- **AMT_REQ_CREDIT_BUREAU_HOUR**
- **AMT_REQ_CREDIT_BUREAU_DAY**
- **AMT_REQ_CREDIT_BUREAU_WEEK**
- **AMT_REQ_CREDIT_BUREAU_MON**
- **AMT_REQ_CREDIT_BUREAU_QRT**

- Median/Mode Imputations



Task B: Outliers

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

Using Excel functions like QUARTILE, IQR, and conditional formatting to identify potential outliers.

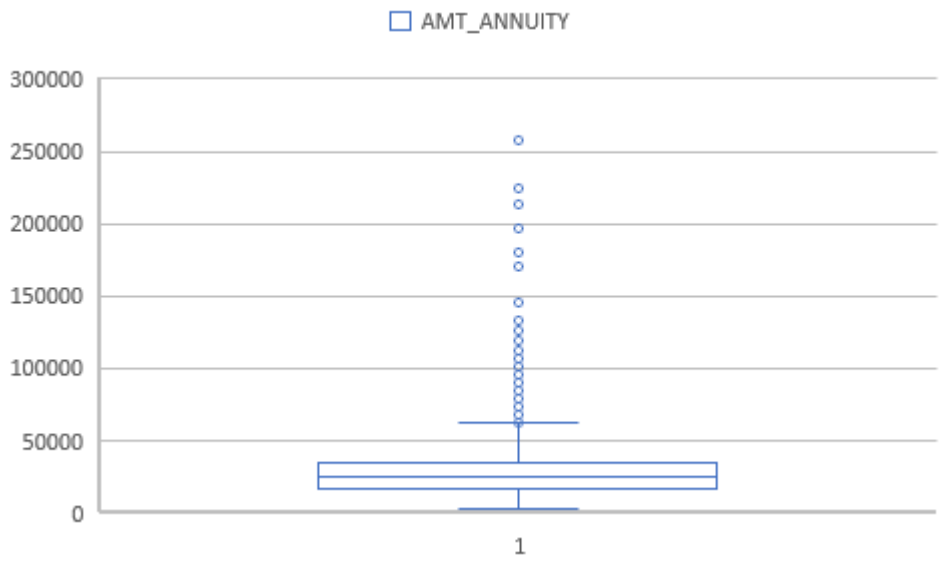
- First we will select numerical columns like AMT_INCOME_TOTAL and AMT_CREDIT
- Calculate Quartile 1, Quartile 3, IQE, Upper Limit and Lower Limit
- We will use Box Plot to highlight the Outliers

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 1-1.5*IQR

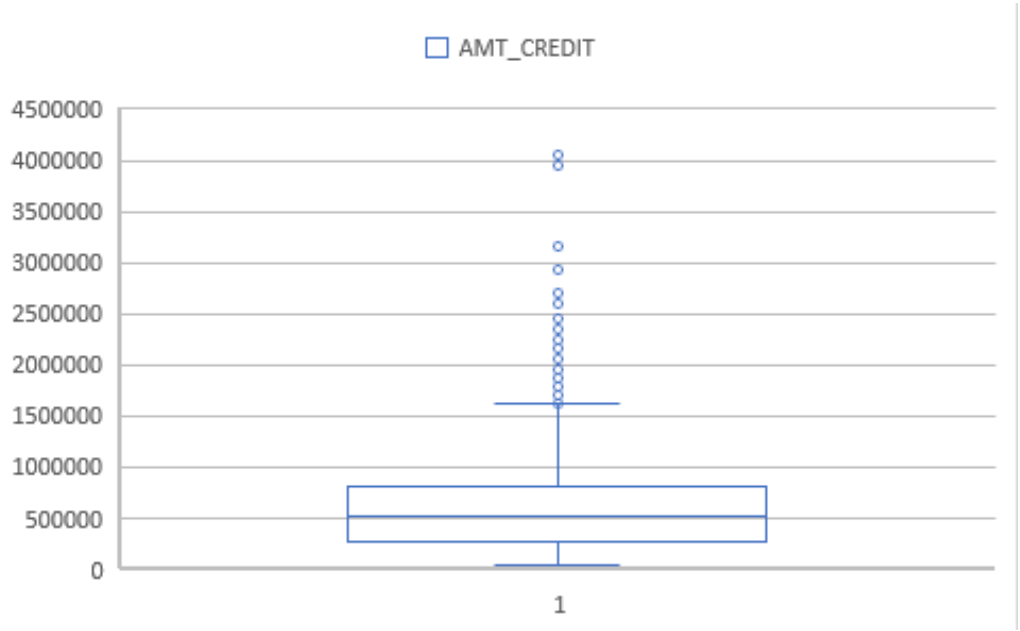
Outliers

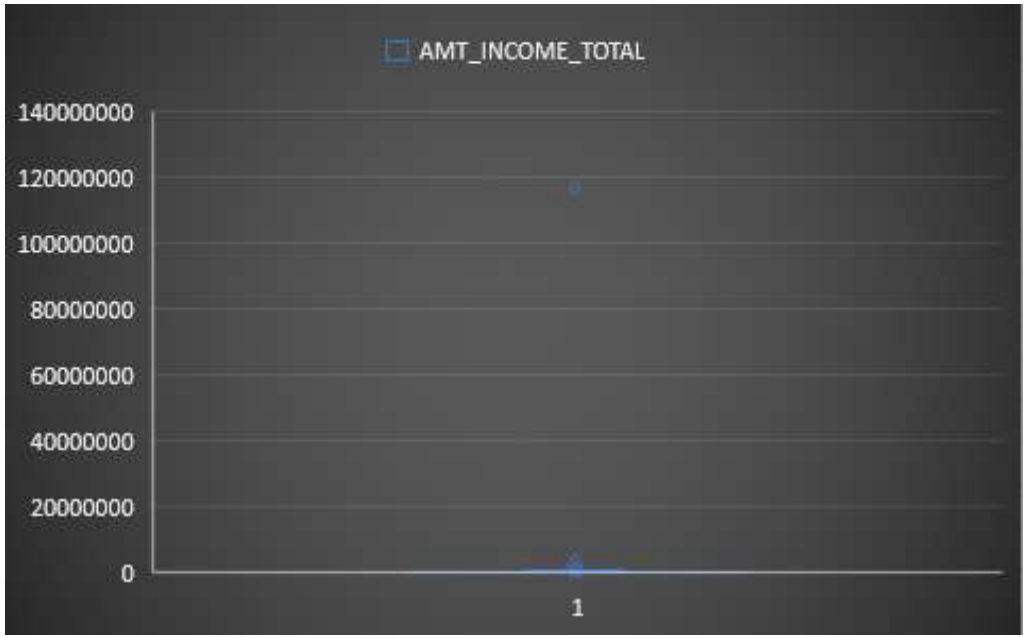
- These columns are outliers



Outliers

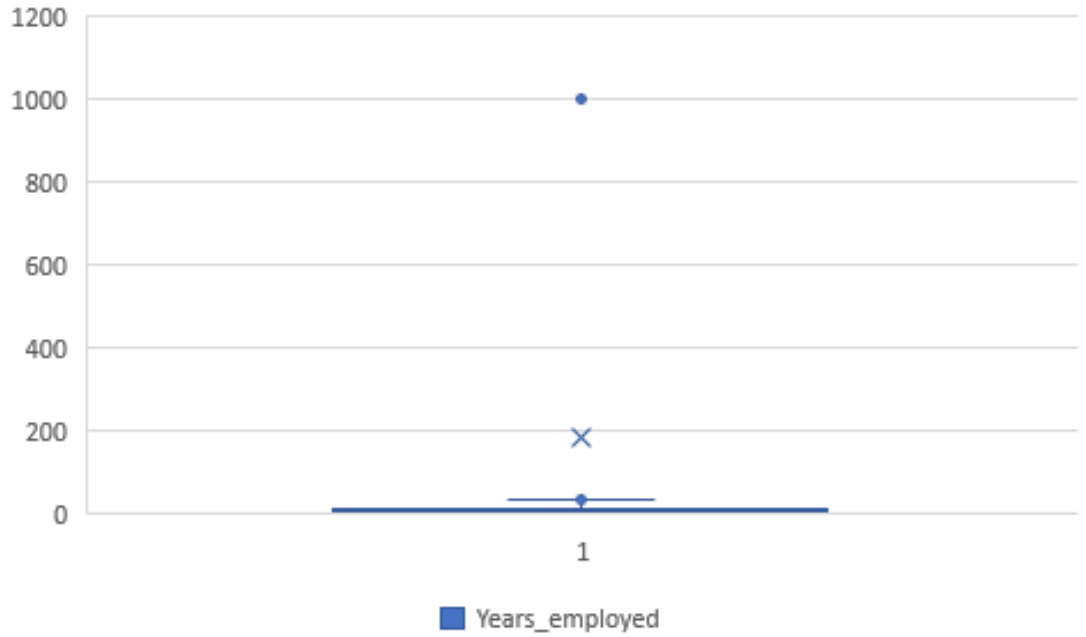
- There are few outliers in columns like AMT_CREDIT and AMT_INCOME_TOTAL whrre amount is higher than normal. In AMT_INCOME+TOTAL one of extreme outlier is 117000000 but we will not remove because income of person varies.

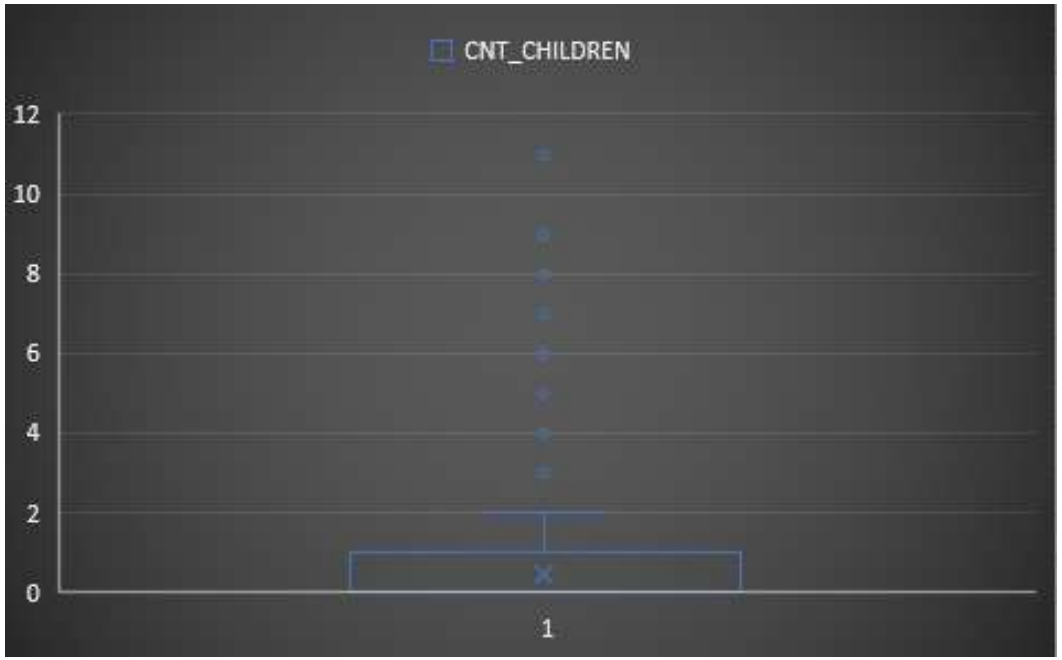




Outliers

- In columne Years_employed we can see people being employed for 1001 years which is not possible. Column CNT_CHILDREN shows people are having 11 children which is impossible





Task C: Data Imbalance

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

Using COUNTIF and SUM to calculate the proportions of each class.

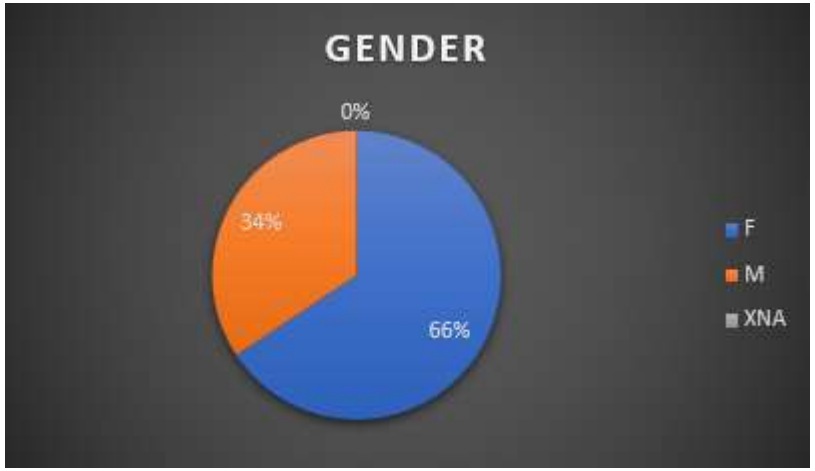
- Target Column-

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

0- Payment on time
1- Late Payment

- CODE_GENDER Column-

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



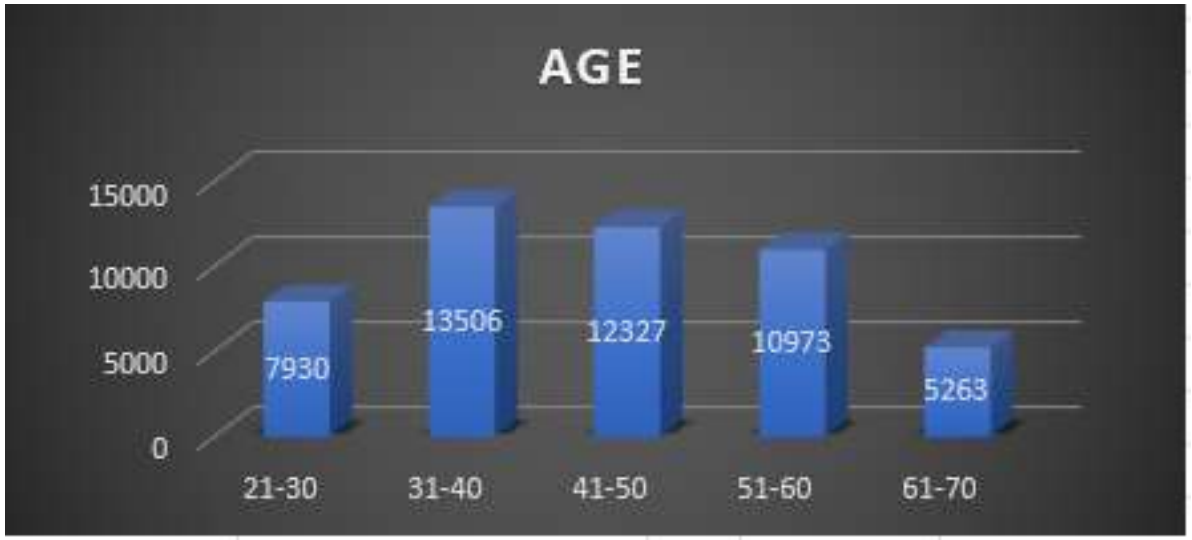
66% clients are female
34% clients are male

Task D: Univariate, Segmented Univariate and Bivariate Analysis

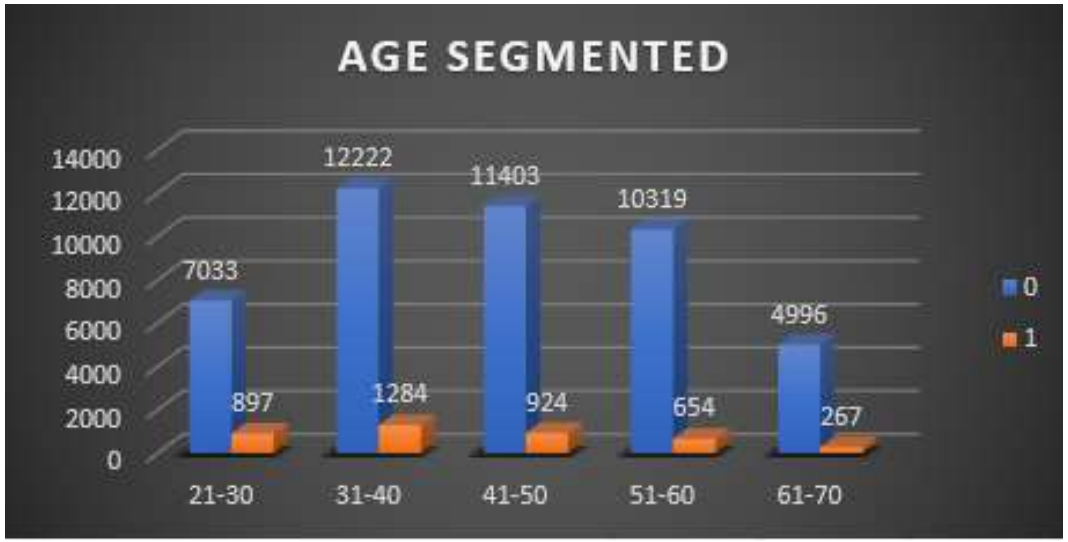
Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.

Utilize Excel functions like COUNT, AVERAGE, MEDIAN, and statistical functions for descriptive analysis. Utilize Excel features like filters, sorting, and pivot tables for segmented and bivariate analysis.

Univariate/Segmented Univariate Analysis

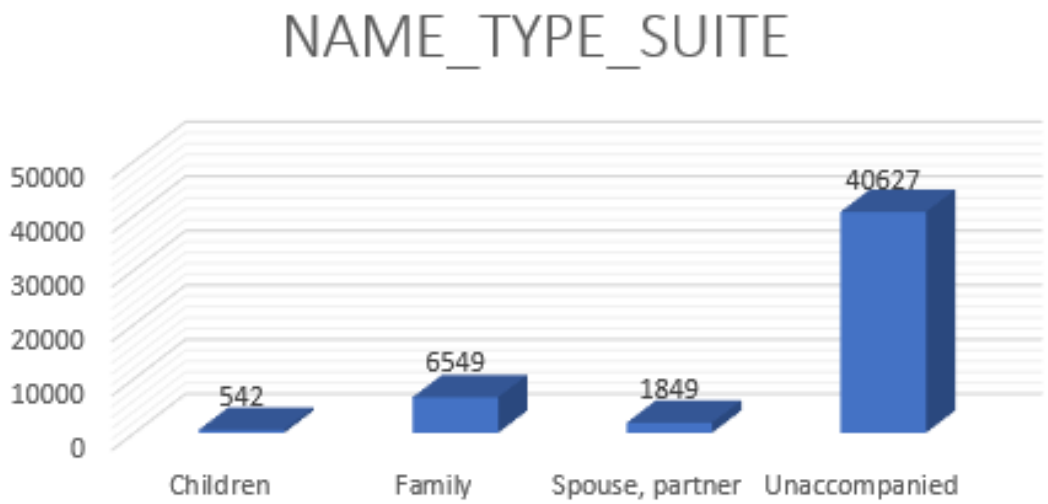


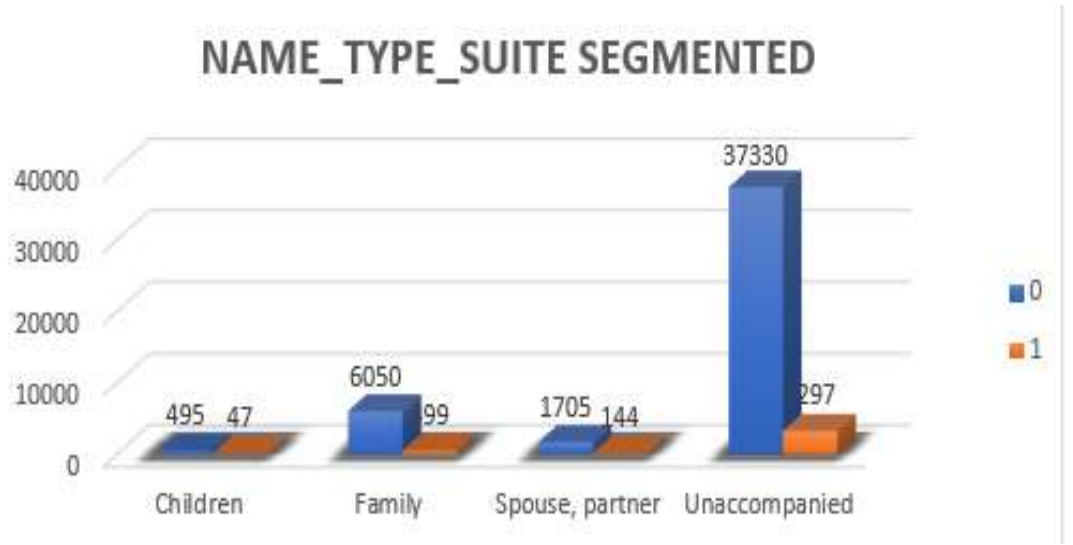
Majority of the clients are in age group 31-40



We can see as age increases chances of defaulter decreases

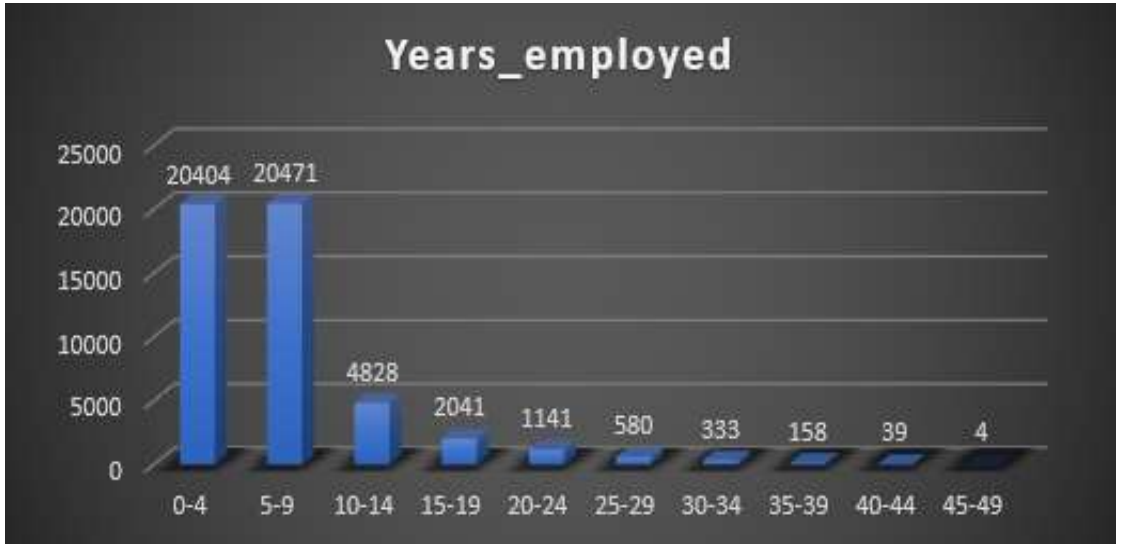
Univariate/Segmented Univariate Analysis



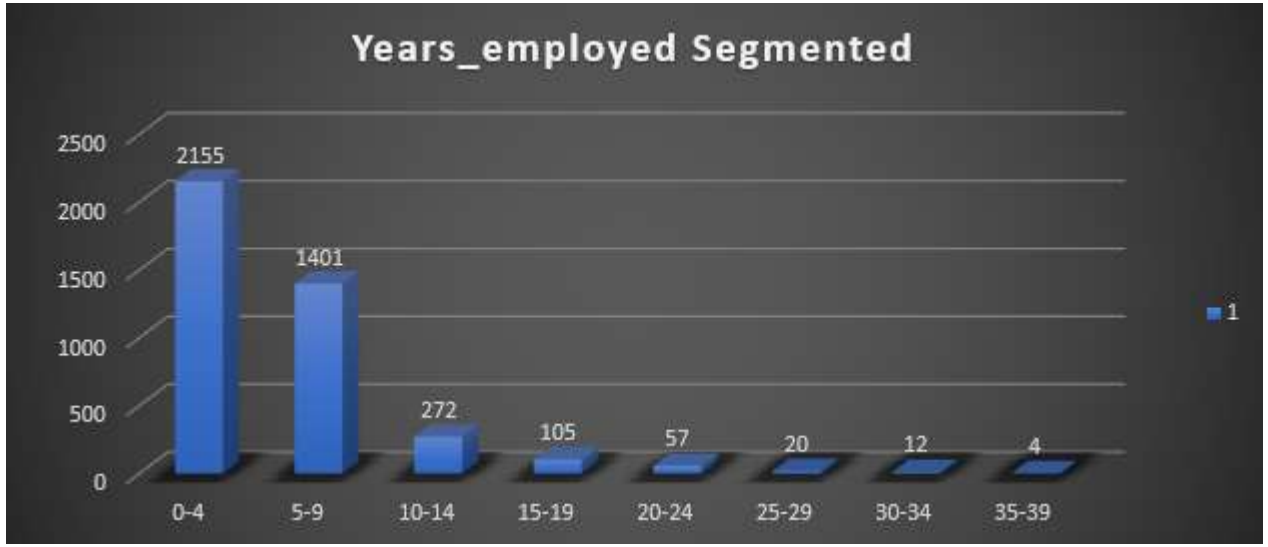


Majority of the clients Unaccompanied followed by family

Univariate/Segmented Univariate Analysis

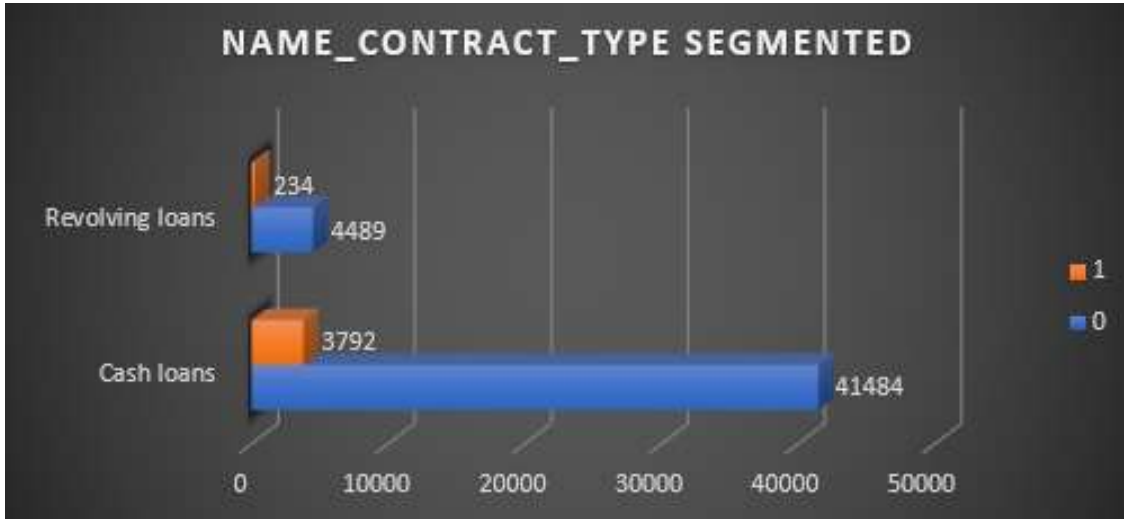


Majority of the clients are having 0-9 years of experience



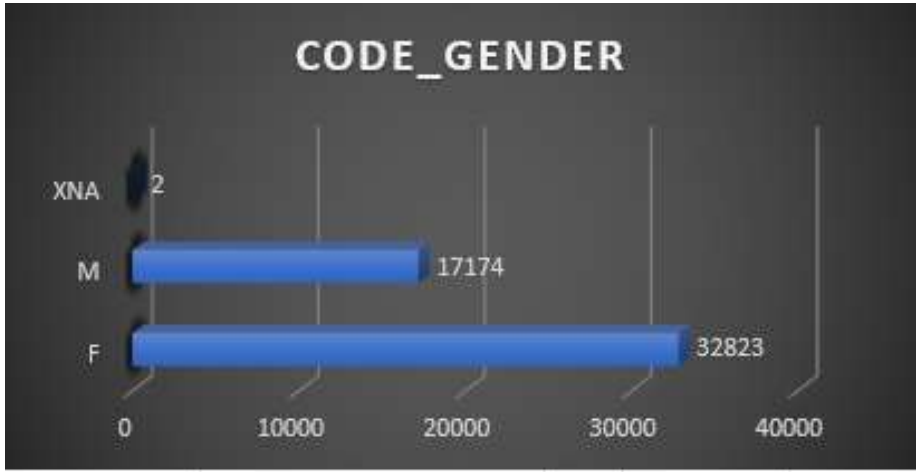
As we can see from the above figure, as experience increases, chances of defaulting decreases

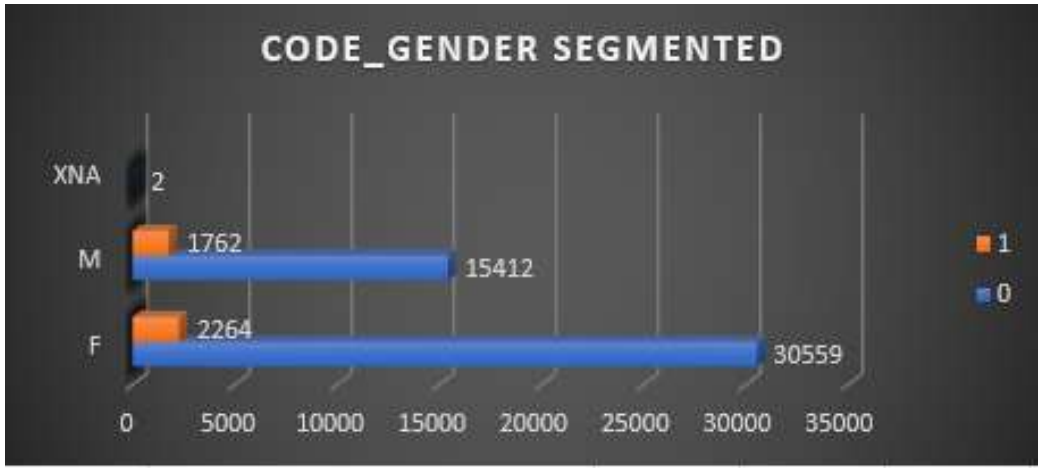
Univariate/Segmented Univariate Analysis



Majority of the clients are taking Cash loans

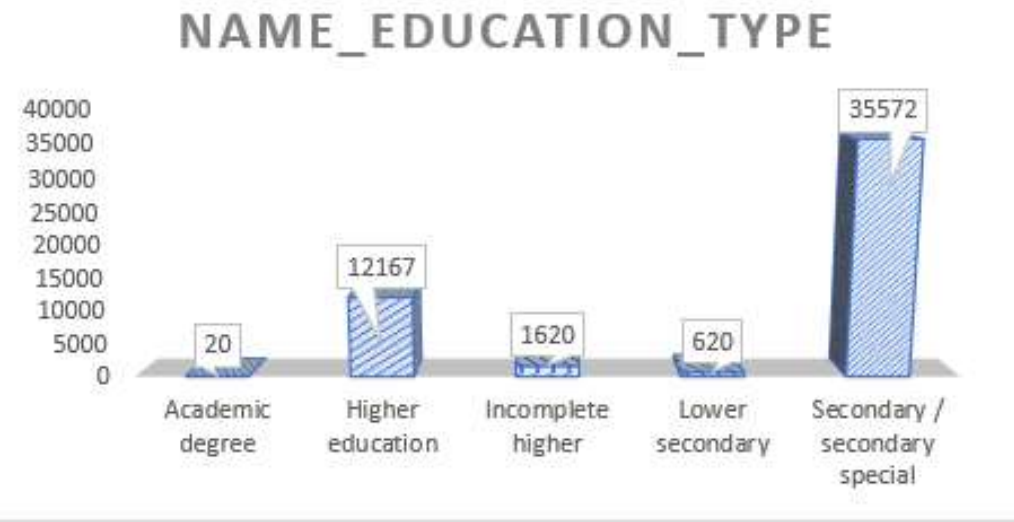
Univariate/Segmented Univariate Analysis



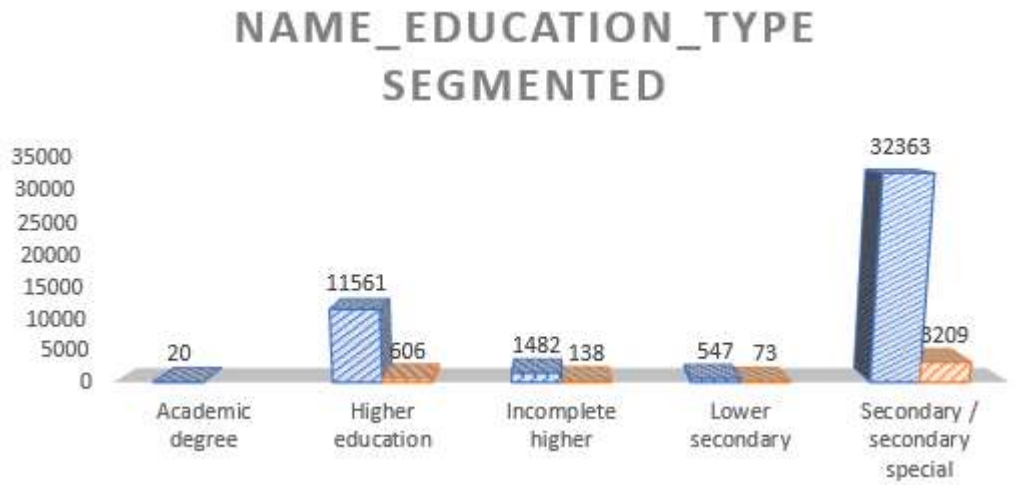


Male are less defaulters compared to females

Univariate/Segmented Univariate Analysis

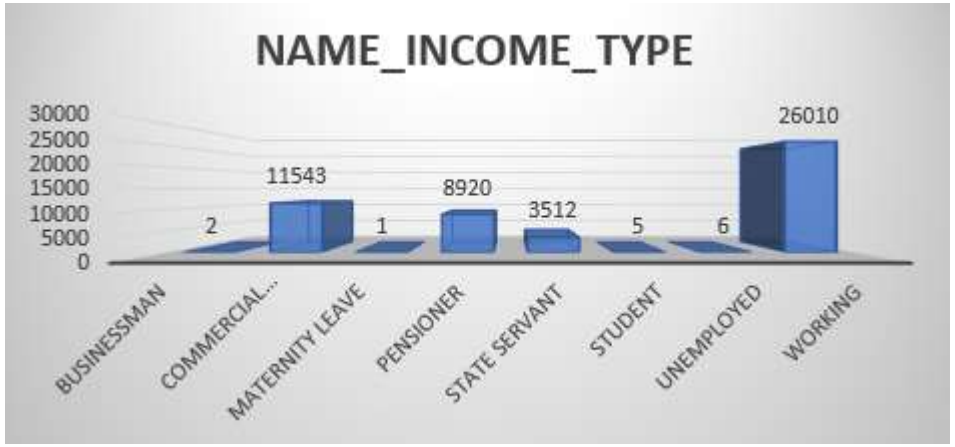


The number of loans taken by clients with Secondary/Special Education is the highest and Academic degree is the lowest

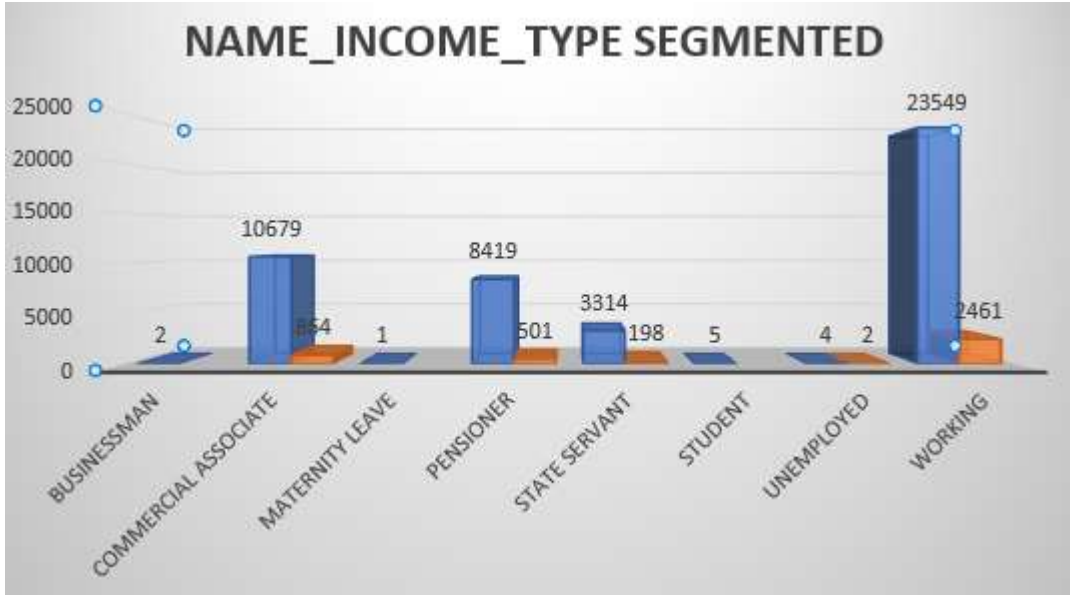


Least default: Academic Degree
Highest default: Secondary Special

Univariate/Segmented Univariate Analysis



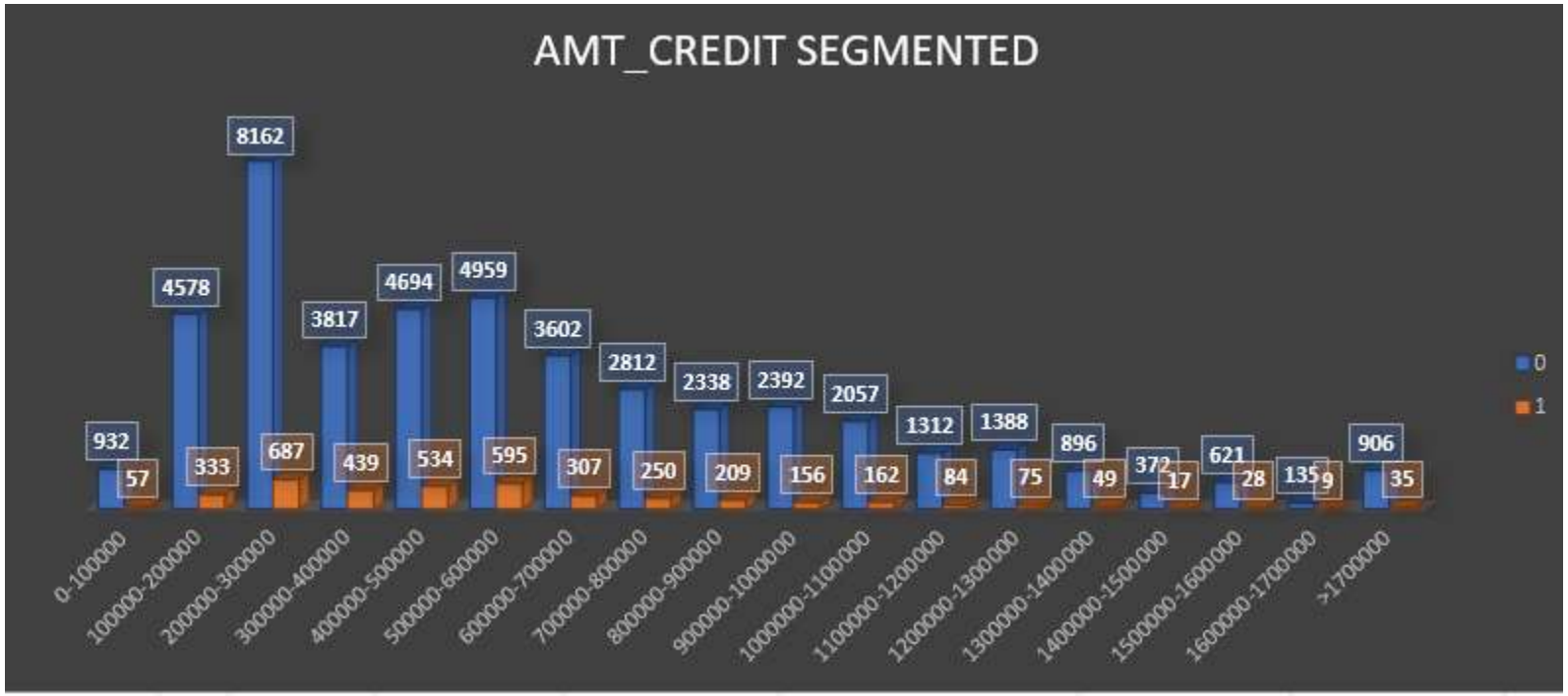
Bank target those group whose income type is working



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 loan between 2L - 3L

Univariate/Segmented Univariate Analysis

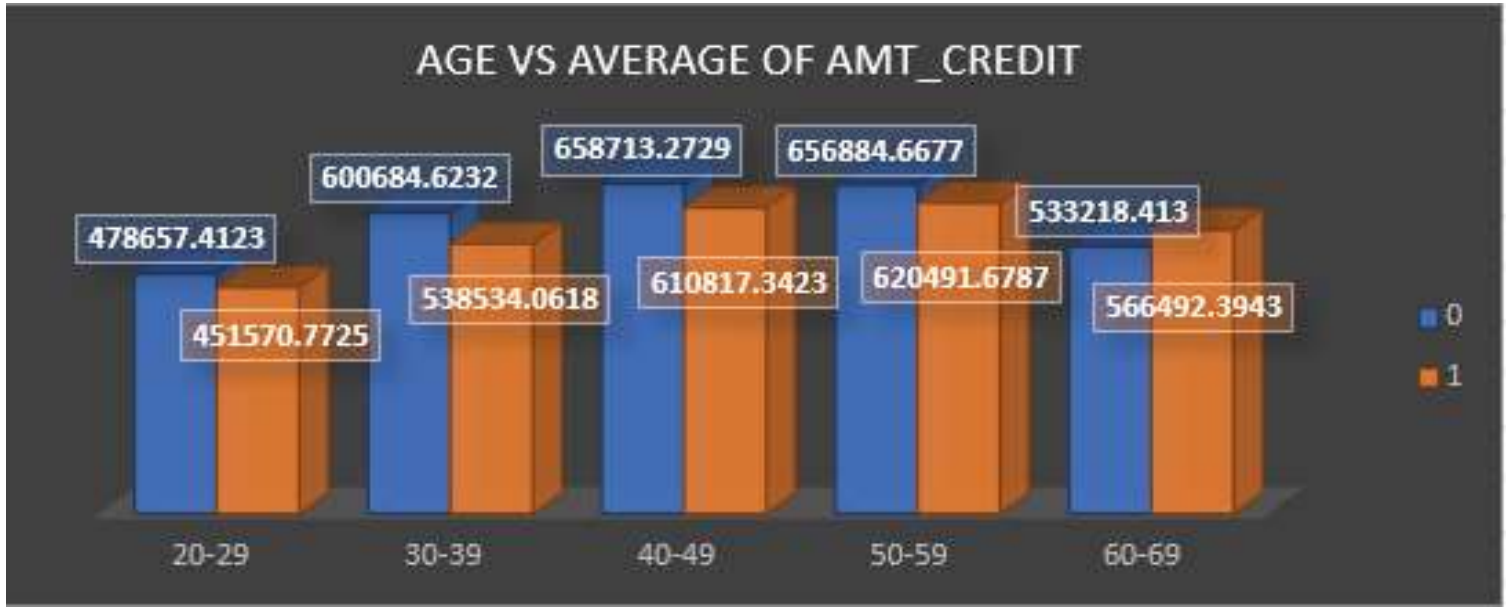
Clients who are working for Business Entity type of Organization took the highest number of loans

Univariate/Segmented Univariate Analysis

- Previous_application Dataset-

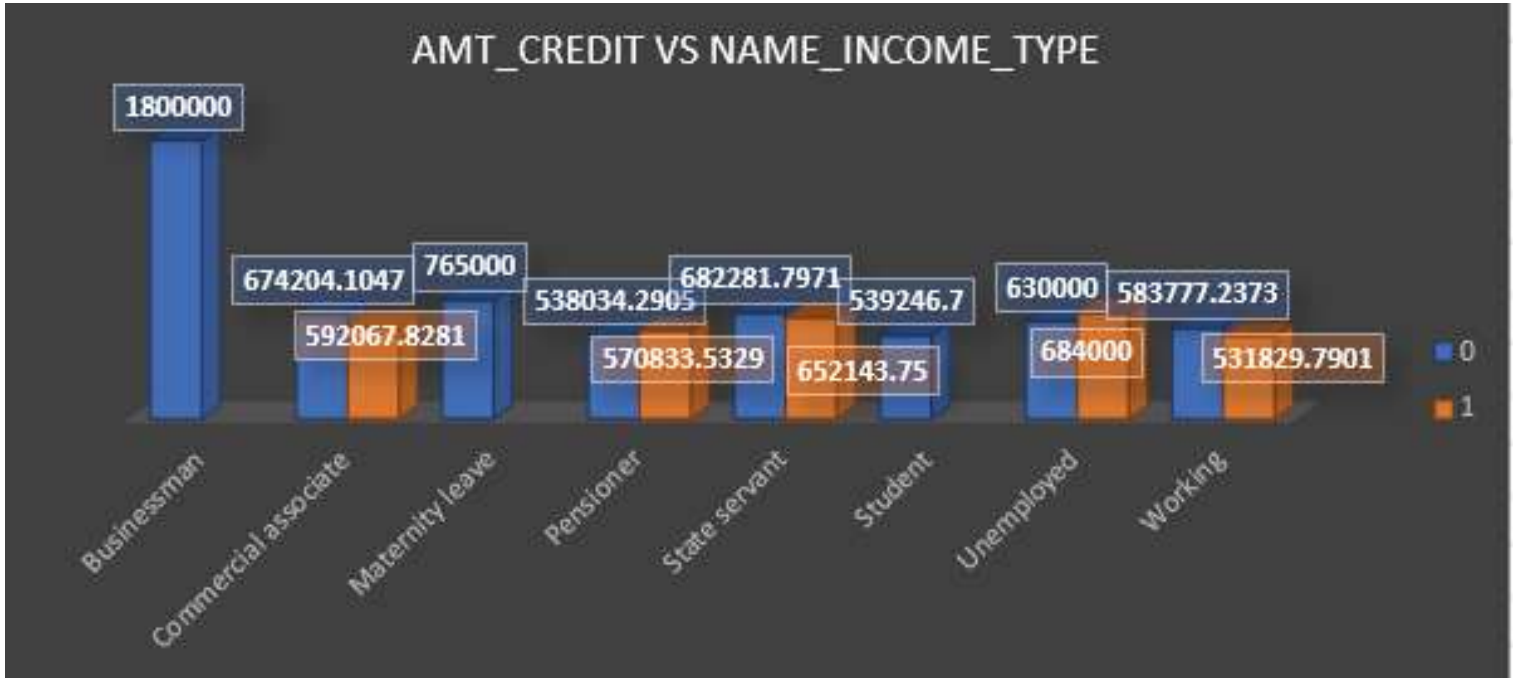
More number of clients were approved for loan previously

Bivariate Analysis



Age group 40-49 took the highest amount of loan but the 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 the highest amount of loan which they didn't repay on time

Task E: Correlations

Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.

Top Correlations Coefficients for Payment difficulties are:

Correlation between Columns	Values
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_ANNUITY	0.749665201
AMT_GOODS_PRICE - AMT_ANNUITY	0.74950403

Correlations

	CNT_CHILDRE	AMT_INCOME_	AMT_CREDI	AMT_ANNUIT	AMT_GOODS_PR	REGION_POPULATION_REL	DAYS_BIRTH	DAYS_EMPLO	DAYS_REGISTRA	DAYS_ID_PU	HOUR_A	REG_RE	REG_RE	LIVE_RE	REG_CI	REG_CI	LIVE_CI	OBS_30	DEF_30	OBS_60	DEF_60	AMT_RI	AMT_RI	AMT_RI	AMT_RI	AMT_RI	AMT_REQ_CREDIT_
CNT_CHILDREN	1	0.036319722	0.0057055	0.02638336	0.001046405	-0.024912809	0.33587627	-0.2435915	0.183072478	-0.0325372	-0.005	-0.01	0.0138	0.0217	0.0201	0.071	0.0679	0.0162	-0.003	0.0163	-0.003	0.0026	0.0012	0.0043	-0.012	-0.005	-0.035734888
AMT_INCOME_TOTAL	0.036319722	1	0.3779658	0.451135167	0.383650216	0.181341261	0.07376342	-0.1627027	0.06833375	0.0322864	0.0854	0.0789	0.1571	0.1477	0.0093	0.0152	0.0197	-0.033	-0.032	-0.033	-0.033	0.0081	0.0095	0.0095	0.0749	0.0158	0.031323516
AMT_CREDIT	0.005705458	0.377365752	1	0.770772818	0.986051701	0.095539444	-0.0510842	-0.0773672	0.008053758	-0.0082902	0.0565	0.0278	0.0561	0.0544	-0.021	-0.014	0.004	0.0009	-0.014	0.0012	-0.013	4E-05	0.0135	0.0054	0.064	0.0268	-0.031568333
AMT_ANNUITY	0.02638336	0.451135167	0.7707728	1	0.774006842	0.11727925	0.00931142	-0.1130053	0.03460901	0.009427	0.0536	0.0462	0.0825	0.0749	-0.005	0.0016	0.0112	-0.01	-0.02	-0.01	-0.023	0.0101	0.0032	0.0189	0.038	0.0101	-0.004173747
AMT_GOODS_PRICE	0.001046405	0.383650216	0.9860517	0.774006842	1	0.098839174	-0.0486644	-0.0750631	0.011016338	-0.0094413	0.0651	0.0304	0.0575	0.0547	-0.02	-0.014	0.0029	0.0006	-0.015	0.0009	-0.02	0.0008	0.0137	0.0058	0.0658	0.0276	-0.034352324
REGION_POPULATION_RELATIV	-0.02491281	0.181341261	0.0955394	0.11727925	0.098839174	1	-0.0304354	-0.0066107	-0.058501361	-0.0022363	0.1676	-0.003	0.0631	0.0874	-0.046	-0.038	-0.011	-0.019	0.0089	-0.018	0.0033	-0.003	-3E-04	0.0026	0.0707	-0.01	0.004652396
DAYS_BIRTH	0.335876263	0.073763425	-0.0510842	0.009311417	-0.048664402	-0.030435413	1	-0.61529	0.335028046	0.2700733	0.0964	0.0604	0.0959	0.0639	0.1833	0.2361	0.1492	0.0123	0.0007	0.0123	0.0022	0.0015	0.002	-0.002	-0.002	-0.022	-0.070267716
DAYS_EMPLOYED	-0.24353152	-0.162702675	-0.0773672	-0.11300529	-0.075063056	-0.006610653	-0.61529	1	-0.204370881	-0.2722244	-0.092	-0.036	-0.107	-0.096	-0.093	-0.254	-0.218	0.0057	0.017	0.0055	0.0165	-0.004	0.0016	-0.006	-0.033	0.0146	0.044183816
DAYS_REGISTRATION	0.183072478	0.06833375	0.0080538	0.03460901	0.011016338	-0.058501361	0.33502805	-0.2043709	1	0.1035483	-0.002	0.0279	0.0347	0.0233	0.0678	0.0316	0.0612	0.011	0.0034	0.0113	0.0063	-0.004	-0.003	0.0007	-0.011	0.0031	-0.02296176
DAYS_ID_PUBLISH	-0.03253722	0.032286356	-0.0082902	0.00942697	-0.009441255	-0.002236288	0.27007331	-0.2722244	0.103548302	1	0.038	0.0332	0.0478	0.0338	0.0751	0.102	0.0633	-0.012	0.0023	-0.012	0.0026	0.0028	0.0035	-0.005	-0.013	-0.025	-0.044632876
HOUR_APPR_PROCESS_START	-0.00527255	0.08543156	0.0565248	0.053564989	0.065133303	0.167612161	0.09638927	-0.092358	-0.002396446	0.0379713	1	0.0512	0.0736	0.0598	0.0197	0.0263	0.0151	-0.008	-0.002	-0.008	-0.006	-0.007	0.0103	-0.007	0.0288	-5E-04	-0.0250598
REG_REGION_NOT_LIVE_REGIO	-0.01038333	0.078342904	0.0278128	0.046175655	0.030367622	-0.003185217	0.06042704	-0.0364131	0.027839354	0.0332285	0.0512	1	0.4436	0.0805	0.3351	0.1426	0.0035	-0.015	-0.008	-0.015	-0.009	-0.002	-0.006	-0.002	-0.009	-3E-04	-0.019525847
REG_REGION_NOT_WORK_REG	0.013794691	0.157051351	0.0560969	0.082502425	0.057545564	0.063145413	0.09591523	-0.1073315	0.034657988	0.0478115	0.0736	0.4436	1	0.8614	0.1519	0.2368	0.1922	-0.025	-0.009	-0.025	-0.014	5E-06	0.0008	0.0033	0.0042	-0.009	-0.02753443
LIVE_REGION_NOT_WORK_REG	0.021685073	0.147730123	0.0544306	0.074870093	0.054659311	0.087419766	0.06388551	-0.0955737	0.023280394	0.0337516	0.0598	0.0805	0.8614	1	0.0216	0.1839	0.2338	-0.02	-0.007	-0.02	-0.012	0.0025	0.0029	0.0054	0.0039	-0.012	-0.022490817
REG_CITY_NOT_LIVE_CITY	0.020101944	0.009327686	-0.0213724	-0.00527672	-0.020436382	-0.046083149	0.18330473	-0.0925575	0.067811428	0.0750801	0.0197	0.3351	0.1519	0.0216	1	0.4414	0.0292	-0.005	0.0055	-0.006	0.0055	0.0005	8E-05	-0.001	-0.014	-2E-05	-0.006661336
REG_CITY_NOT_WORK_CITY	0.070971057	0.015150006	-0.0140074	0.001628793	-0.01443832	-0.038253612	0.23613443	-0.2540601	0.091535217	0.1020018	0.0263	0.1426	0.2368	0.1839	0.4414	1	0.8254	-0.006	0.001	-0.006	0.0033	0.0043	-2E-04	0.0022	-0.012	-0.004	-0.011356758
LIVE_CITY_NOT_WORK_CITY	0.067882194	0.013663673	0.00398	0.011203272	0.002861594	-0.011278612	0.14316734	-0.2177413	0.061153253	0.063319	0.0151	0.0035	0.1922	0.2338	0.0292	0.8254	1	-0.005	-0.002	-0.005	-2E-04	0.004	-0.001	0.0024	-0.005	-0.005	-0.012345253
OBS_30_CNT_SOCIAL_CIRCLE	0.016180293	-0.033045393	0.0008764	-0.0039321	0.000634386	-0.01906308	0.01228703	0.00565019	0.010977833	-0.011854	-0.008	-0.015	-0.025	-0.02	-0.005	-0.006	-0.005	1	0.3062	0.9984	0.2292	0.0024	0.001	-0.004	0.0082	0.0088	0.034161046
DEF_30_CNT_SOCIAL_CIRCLE	-0.00282133	-0.032012977	-0.0135094	-0.01974602	-0.015155074	0.008305531	0.00066377	0.01703333	0.003448383	0.0023127	-0.002	-0.008	-0.009	-0.007	0.0055	0.001	-0.002	0.3062	1	0.3086	0.851	-0.004	0.0037	-0.005	0.0077	0.0054	0.014506751
OBS_60_CNT_SOCIAL_CIRCLE	0.016334834	-0.03301707	0.0011848	-0.00967585	0.000856455	-0.018012635	0.01229458	0.00551128	0.011295653	-0.0121553	-0.008	-0.015	-0.025	-0.02	-0.006	-0.006	-0.005	0.9984	0.3086	1	0.2313	0.0026	0.0009	-0.005	0.0081	0.0087	0.034573624
DEF_60_CNT_SOCIAL_CIRCLE	-0.0033303	-0.032535174	-0.0185673	-0.02301062	-0.019693391	0.003253593	0.00220712	0.01651602	0.006282428	0.0026424	-0.006	-0.009	-0.014	-0.012	0.0055	0.0033	-2E-04	0.2292	0.851	0.2313	1	-0.003	0.0028	-0.006	0.004	0.0083	0.015204368
AMT_REQ_CREDIT_BUREAU_H	0.00261703	0.008122955	3.653E-05	0.0101408	0.000827804	-0.003132124	0.00148629	-0.0042935	-0.003689166	0.0028242	-0.007	-0.002	5E-06	0.0025	0.0005	0.0043	0.004	0.0024	-0.004	0.0026	-0.003	1	0.2308	0.0121	0.0095	0.0035	0.004095354
AMT_REQ_CREDIT_BUREAU_D	0.001136338	0.009477681	0.0134864	0.003156839	0.013665416	-0.000338841	0.00198387	0.00161813	-0.00338406	0.0035147	0.0103	-0.006	0.0008	0.0023	8E-05	-2E-04	-0.001	0.001	0.0037	0.0009	0.0028	0.2308	1	0.2431	-7E-04	-0.008	-0.00085677
AMT_REQ_CREDIT_BUREAU_W	0.004327432	0.009487825	0.0053743	0.018303774	0.005848551	0.002644642	-0.0024012	-0.0064816	0.000659813	-0.0046659	-0.007	-0.002	0.0033	0.0054	-0.001	0.0022	0.0024	-0.004	-0.005	-0.005	-0.006	0.0121	0.2431	1	-0.011	-0.015	0.024737117
AMT_REQ_CREDIT_BUREAU_M	-0.01160782	0.074854673	0.063376	0.037365476	0.065821043	0.070736631	-0.002453	-0.0329546	-0.010724833	-0.0132326	0.0288	-0.009	0.0042	0.0093	-0.014	-0.012	-0.005	0.0082	0.0077	0.0081	0.004	0.0095	-7E-04	-0.011	1	0.0113	0.013311173
AMT_REQ_CREDIT_BUREAU_Q	-0.00473083	0.015777535	0.0267933	0.010053213	0.027627403	-0.009634593	-0.021523	0.0145774	0.003127351	-0.0245881	-5E-04	-3E-04	-0.003	-0.012	-2E-05	-0.004	-0.005	0.0088	0.0054	0.0087	0.0083	0.0035	-0.008	-0.015	0.0113	1	0.121744813
AMT_REQ_CREDIT_BUREAU_Y	-0.03573483	0.031323516	-0.0315683	-0.00417375	-0.034352324	0.004652396	-0.0702677	0.04418382	-0.02296176	-0.0446929	-0.025	-0.02	-0.028	-0.022	-0.007	-0.012	-0.013	0.0342	0.0145	0.0346	0.0152	0.0041	-9E-04	0.0247	0.0193	0.1217	1

Correlations

Top Correlation coefficient for Re-Payers are:

Correlation between Columns	Values
OBS_60_CNT_SOCIAL_CIRCLE - OBS_30_CNT_S	0.998357563
AMT_GOODS_PRICE - AMT_CREDIT	0.986051701
LIVE_REGION_NOT_WORK_REGION - REG_RE	0.861374946
DEF_60_CNT_SOCIAL_CIRCLE - DEF_30_CNT_S	0.850995792
REG_CITY_NOT_WORK_CITY - LIVE_CITY_NOT_	0.825358079
AMT_ANNUITY - AMT_GOODS_PRICE	0.774006842
AMT_ANNUITY - AMT_CREDIT	0.770772818

Correlations

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELA	DAYS_BIRTH	DAYS_EI	DAYS_R	DAYS_ID	HOUR_A	REG_REG	REG_REG	LIVE_REG	REG_CIT	REG_CIT	LIVE_CIT	OBS_30_	DEF_30_	OBS_60_	DEF_60_	AMT_REI	AMT_REI	AMT_REI	AMT_REI	AMT_REQ_CREDIT_BU	AMT_REQ_CREDIT_BUREAU_YEA
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	-0.0126	0.01515	-0.0185	-0.0003	-0.0306	-0.0306	0.00816	-0.011520595	-0.03080113
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	-0.0011	-0.0014	-0.0022	-0.0009	-0.003749228	-0.005100984
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	0.01781	-0.0085	0.00013	0.08341	-0.019361311	-0.016459973
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	0.0374	-0.0187	0.03472	0.0713	-0.001630664	0.001564273
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	0.01526	-0.0063	0.00011	0.07891	-0.020367636	-0.023475441
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	0.00936	-0.0038	0.01206	0.0754	0.015310168	0.024023928
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	0.0249	-0.0227	-0.0097	-0.0073	-0.008783235	-0.090127316
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	-0.003	0.04948	0.02039	-0.0331	0.017875877	0.017692456
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	0.00638	-0.0015	-0.0182	-0.0015	-0.006290417	-0.025094194
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	0.01408	-0.0064	0.01954	-0.0379	-0.032671471	-0.03164306
HOUR_APPR_PROCESS_START	-0.006884357	0.01448201 3	0.045296384	0.044891881	0.05746275 9	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	-0.0331	0.00141	-0.0091	0.06634	-0.011742607	-0.037911822
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	-0.011	0.0042	0.00723	0.05155	-0.010452446	-0.033986103
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	0.30674	0.18375	0.22868	0.16908	-0.0321	0.00152	-0.0316	0.00493	0.0227	0.01115	-0.0188	-0.0159	-0.010984431	-0.029281003
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	0.06606	0.10053	0.30674	1	0.02608	0.1578	0.21787	-0.0208	-0.0061	-0.02	8.7E-05	0.03195	0.00701	-0.0322	0.04081	-0.014115297	-0.014042305
REG_CITY_NOT_LIVE_CITY	0.001745098	-0.005992314	-0.052261708	-0.017702478	-0.052733855	-0.034931305	0.149110346	-0.0882	0.05557	0.0641	0.00552	0.33817	0.18375	0.02608	1	0.4673	-0.015	-0.0499	0.00342	-0.0504	0.00258	-0.0011	-0.0191	-0.0042	-0.0356	-2.97186E-05	-0.020205505
REG_CITY_NOT_WORK_CITY	0.048916581	-0.010357192	-0.039113138	0.002176683	-0.043981018	-0.043285987	0.226350689	-0.2463	0.10076	0.08301	0.0032	0.14759	0.22868	0.1578	0.4673	1	0.78375	-0.0421	-0.0156	-0.0416	-0.0137	0.01833	-0.0053	0.00463	-0.0447	-0.049007672	-0.025982993
LIVE_CITY_NOT_WORK_CITY	0.058183771	-0.008036091	-0.006664341	0.013562938	-0.013057846	-0.025223619	0.143399639	-0.2006	0.06982	0.03844	-0.0118	-0.0037	0.16908	0.21787	-0.015	0.78375	1	-0.0241	-0.0279	-0.023	-0.0246	0.01426	0.00077	0.01057	-0.0202	-0.037724429	-0.006136593
OBS_30_CNT_SOCIAL_CIRCLE	0.01793193	-0.011280916	0.033466173	0.013819016	0.032723967	-0.008875436	-0.011150233	0.00352	-0.0058	-0.0273	-0.0197	-0.032	-0.0321	-0.0208	-0.0499	-0.0421	-0.0241	1	0.36507	0.99807	0.29795	-0.0141	-0.017	0.00584	0.01608	0.034835809	0.050517528
DEF_30_CNT_SOCIAL_CIRCLE	-0.01361871	-0.007979437	-0.024946679	-0.034545374	-0.019096612	0.027805916	-0.020838794	0.02986	0.001	-0.0284	0.01767	0.00849	0.00152	-0.0061	0.00342	-0.0156	-0.0279	0.36507	1	0.36806	0.89051	0.00273	0.01224	-0.0116	0.00809	0.020101304	0.021016648
OBS_60_CNT_SOCIAL_CIRCLE	0.015145875	-0.011211173	0.034439308	0.014098626	0.033879184	-0.007065002	-0.01257029	0.00421	-0.0059	-0.0262	-0.0195	-0.032	-0.0316	-0.02	-0.0504	-0.0416	-0.023	0.99807	0.36806	1	0.30142	-0.0136	-0.0174	0.00556	0.01698	0.036400413	0.050708513
DEF_60_CNT_SOCIAL_CIRCLE	-0.018505702	-0.006726958	-0.029007236	-0.040471029	-0.020592919	0.027142318	-0.025756651	0.02389	-0.0064	-0.0279	0.01752	0.00582	0.00493	8.7E-05	0.00258	-0.0137	-0.0246	0.29795	0.89051	0.30142	1	-0.0132	-0.0103	-0.0039	0.01303	0.025347772	0.020626159
AMT_REQ_CREDIT_BUREAU_HOUR	-0.000287596	-0.001104179	0.017806362	0.037397493	0.01526195	0.009356216	0.024898705	-0.003	0.00638	0.01408	-0.0331	-0.011	0.0227	0.03195	-0.0011	0.01833	0.01426	-0.0141	0.00273	-0.0136	-0.0132	1	0.3512	0.01932	-0.001	0.031098847	0.006159687
AMT_REQ_CREDIT_BUREAU_DAY	-0.030605254	-0.00144685	-0.008518401	-0.018688343	-0.006319208	-0.003833539	-0.02267042	0.04948	-0.0015	-0.0064	0.00141	0.0042	0.01115	0.00701	-0.0191	-0.0053	0.00077	-0.017	0.01224	-0.0174	-0.0103	0.3512	1	0.06199	-0.0163	0.025535445	0.027607009
AMT_REQ_CREDIT_BUREAU_WEEK	-0.030604048	-0.002218606	0.000125371	0.03472145	0.000114491	0.012064245	-0.009660982	0.02039	-0.0182	0.01954	-0.0091	0.00723	-0.0188	-0.0322	-0.0042	0.00463	0.01057	0.00584	-0.0116	0.00556	-0.0039	0.01932	0.06199	1	-0.001	0.011697359	0.030916965
AMT_REQ_CREDIT_BUREAU_MON	0.008160996	-0.000864018	0.083408196	0.071295225	0.078908703	0.075395596	-0.007277397	-0.0331	-0.0015	-0.0279	0.06634	0.05155	-0.0159	0.04081	-0.0356	-0.0447	-0.0202	0.01608	0.00809	0.01698	0.01303	-0.001	-0.0163	-0.001	1	0.019946401	0.038789503
AMT_REQ_CREDIT_BUREAU_QRT	-0.011520595	-0.003749228	-0.019361311	-0.001630664	-0.020367636	0.015310168	-0.008783235	0.01788	-0.0063	-0.0327	-0.0117	-0.0105	-0.011	-0.0141	-3E-05	-0.049	-0.0377	0.03484	0.0201	0.0364	0.02535	0.0311	0.02554	0.0117	0.01995	1	0.103631744
AMT_REQ_CREDIT_BUREAU_YEAR	-0.03080113	-0.005100984	-0.016459973	0.001569273	-0.023475441	0.024023928	-0.090127316	0.01769	-0.0251	-0.0816	-0.0379	-0.034	-0.0293	-0.014	-0.0202	-0.026	-0.0061	0.05052	0.02102	0.05071	0.02063	0.00616	0.02761	0.03092	0.03879	0.103631744	1

Conclusion/Insights

- 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.

Google Drive Link

- Google drive link for excel sheets are-

Bank Loan Case Study