BANK LOAN
CASE STUDY

PROJECT DESCRIPTION:

- In this project, we will apply Risk Analytics on the data provided for bank processing loan.
- The aim of the project is to identify patterns that indicate if a customer will have difficulty paying their installments.
- The information can the be used to take decisions such as denying the loan, reducing the loan amount, or lending at a higher rate of interest for risky customers who are likely to default.
- Company wants to know the key factors behind loan defaults so better decisions can be taken about loan approval.

APPROACH

Understanding the Data

To observe and grasp the dataset in order to plan for further analysis.

- previous_application.csv
- application_data.csv
- columns_description.csv

Information about previous loan application

details about current loan application

description of each columns

Data-Preprocessing

Cleaning the dataset to make it ready for analysis.

- Handling missing data
- Identifying Outliers
- Merging the datasets
- Feature Engineering

Visualization & Insights

Analyzing and visualizing the data to produce actionable insights.

TECH - STACK USED

- We have used Microsoft® Excel® 2021 MSO (Version 2310 Build 16.0.16924.20054) 64-bit for our data analysis as Excel is jam-packed with features and functions that can be used to clean, aggregate, pivot, and graph data. Also, Excel has a user-friendly visual interface that allows individuals at any level of expertise to easily learn and utilize its capabilities.

- View EXCEL file.
- Watch VIDEO PRESENTATION.

INSIGHTS

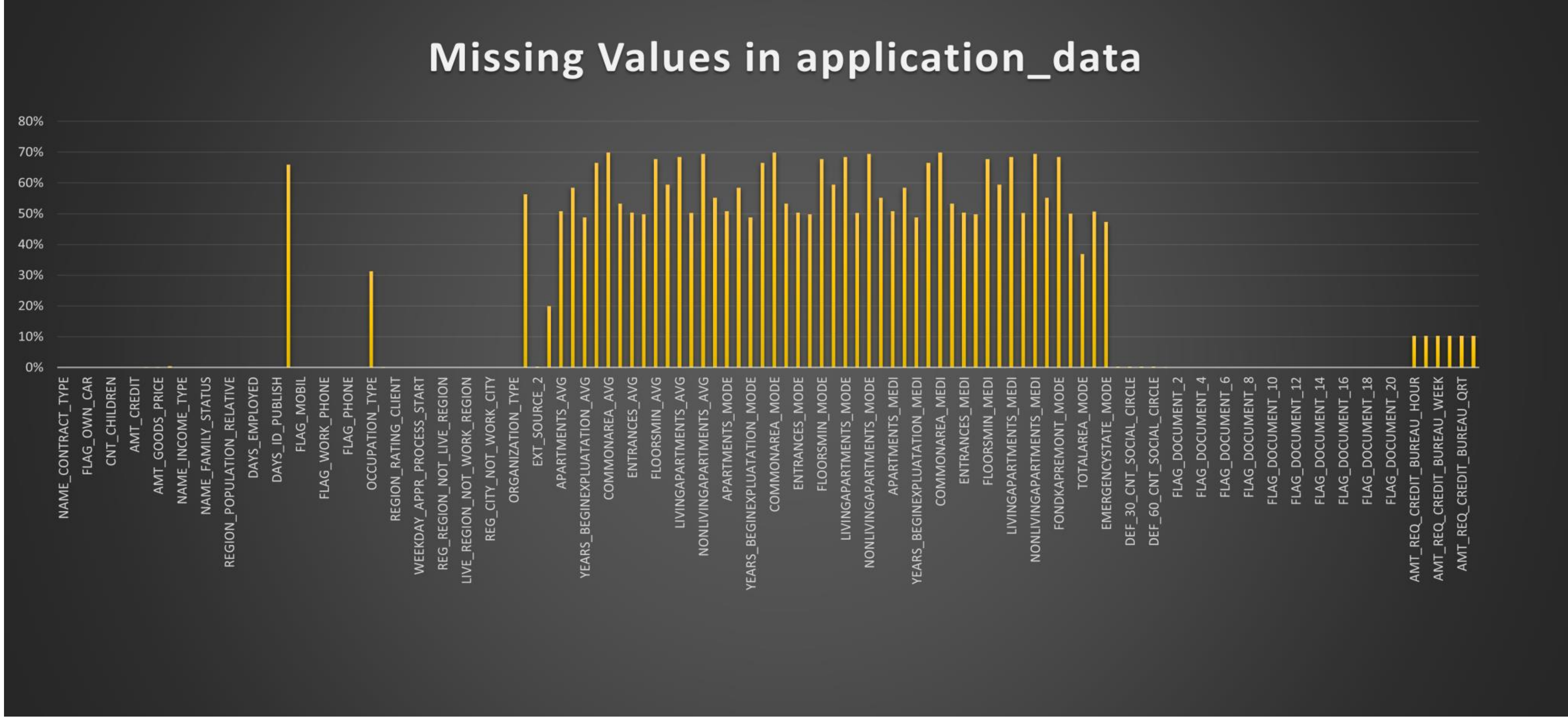
A. Identify Missing Data and Deal with it Appropriately

Finding missing data and deciding an appropriate method to deal with it.

Deleting unwanted columns and dropping columns with missing values > 40 %.

Imputing missing values and standardizing them.

Dataset
application_data.csv



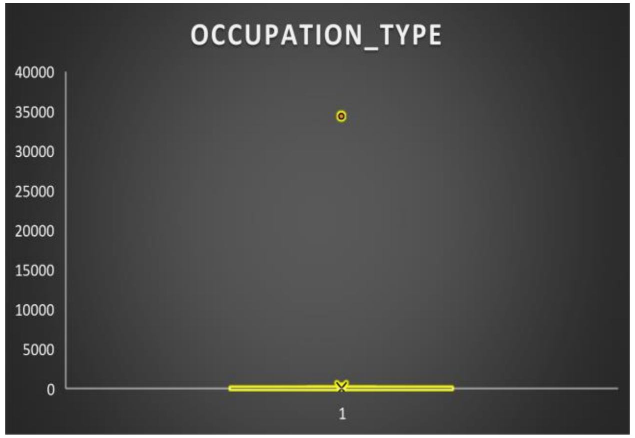
- No of rows= 49999
- No of columns= 122

- Used COUNTA() to calculate the percentage of missing values for each column.
- Visualized proportion of missing values using clustered column chart .
- We will drop the columns which has 40% or more missing values.
- Shortened the columns from 122 to 76.

Missing Values Imputation

Since the box plot is not skewed, we can impute the missing values with the mean.

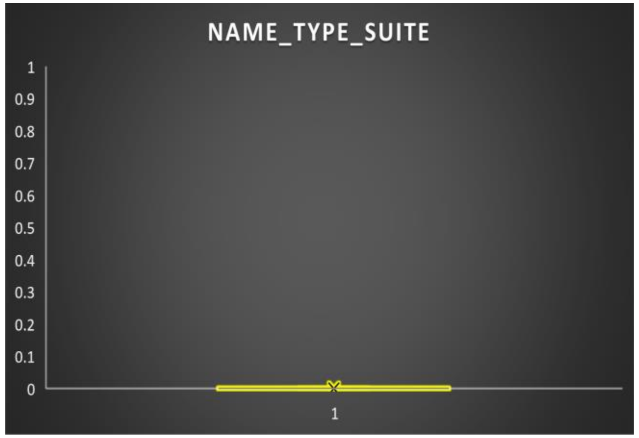
OCCUPATION_TYPE



Row Labels	Count of OCCUPATION_TYPE	Average of EXT_SOURCE_3
Accountants	1642	0.51543488
Cleaning staff	761	0.497779739
Cooking staff	976	0.498062487
Core staff	4488	0.498462842
Drivers	3085	0.505899592
High skill tech staff	1868	0.512290345
HR staff	102	0.476094895
IT staff	80	0.515983851
Laborers	24977	0.519992147
Low-skill Laborers	360	0.484077935
Managers	3525	0.496453065
Medicine staff	1424	0.507168465
Private service staff	453	0.511245123
Realty agents	125	0.479685294
Sales staff	5239	0.500298545
Secretaries	212	0.50738826
Security staff	1155	0.518082797
Waiters/barmen staff	235	0.485134139
Grand Total	50707	0.511226414

Since, **Laborers** have the most count, we replaced all missing values with it.

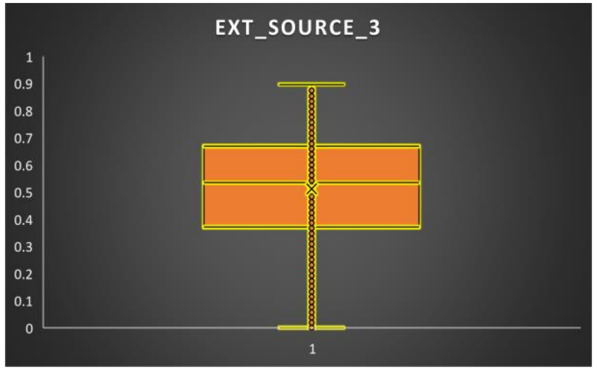
NAME_TYPE_SUITE



Row Labels	Count of NAME_TYPE_SUITE
Children	993
Family	6581
Group of people	76
Other_A	262
Other_B	551
Spouse, partner	2098
Unaccompanied	15195
(blank)	
Grand Total	25756

We replaced missing values with **Unaccompanied** since it has the highest count.

EXT_SOURCE_3



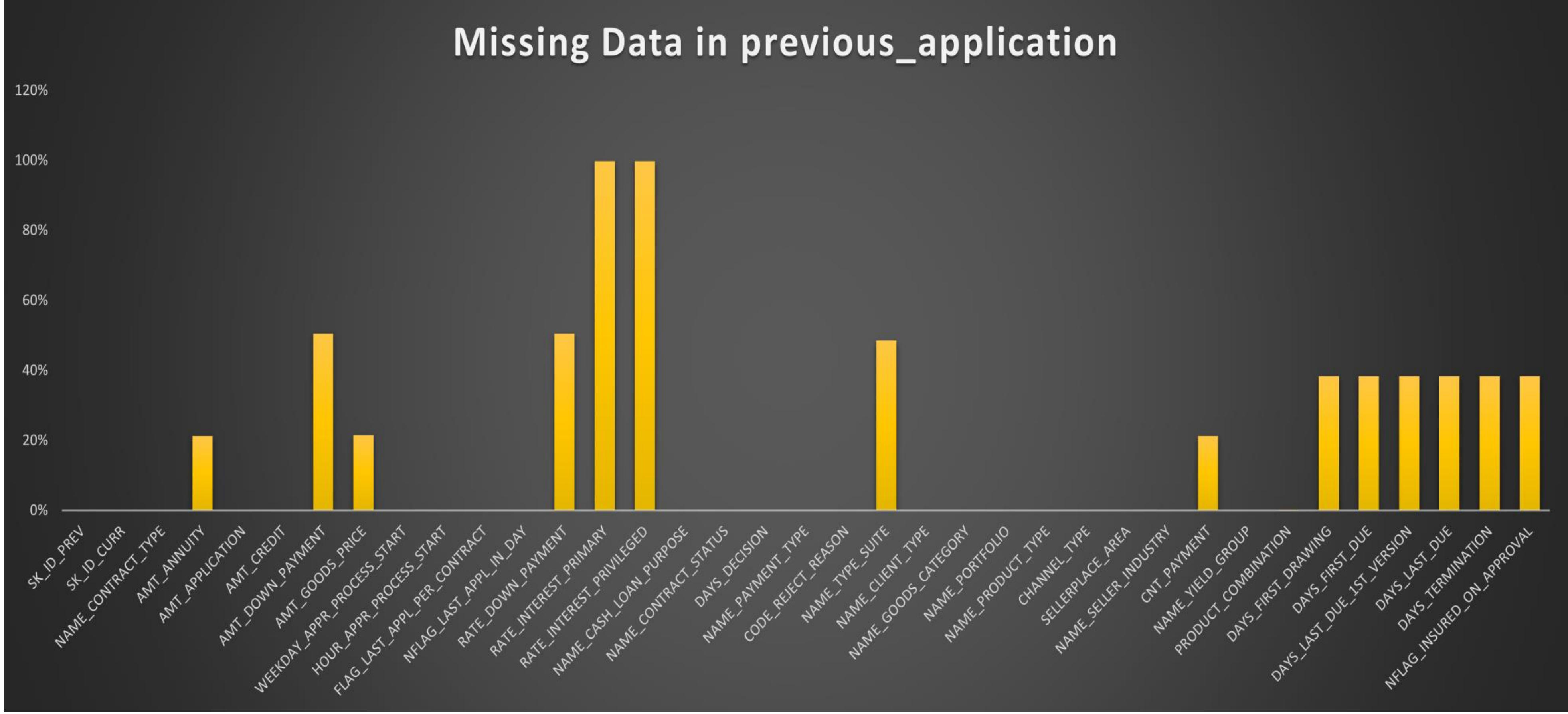
AMT_REQ_CREDIT_BUREAU_HOUR, AMT_REQ_CREDIT_BUREAU_DAY, AMT_REQ_CREDIT_BUREAU_WEEK, AMT_REQ_CREDIT_BUREAU_MON, AMT_REQ_CREDIT_BUREAU_QRT, AMT_REQ_CREDIT_BUREAU_YEAR

We replaced missing values in **EXT_SOURCE_3, 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 and AMT_REQ_CREDIT_BUREAU_YEAR** with its **mean.**

Dataset

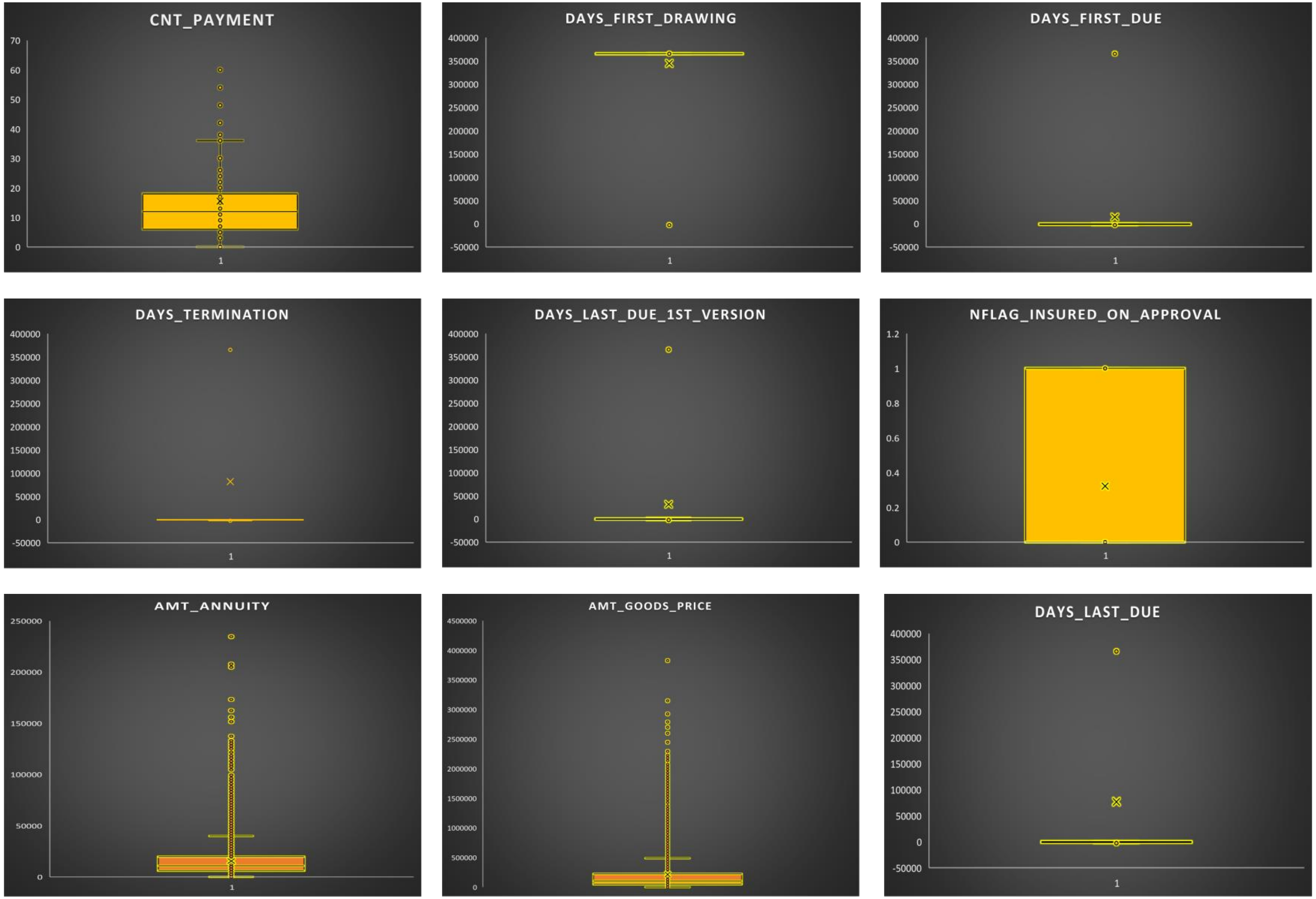
previous_data.csv

- No of rows= 49999
- No of columns= 37



- Used COUNTA() to calculate the percentage of missing values for each column.
- Visualized proportion of missing values using clustered column chart .
- We will drop the columns which has 40% or more missing values.
- Shortened the columns from 37 to 32.

Missing Values Imputation

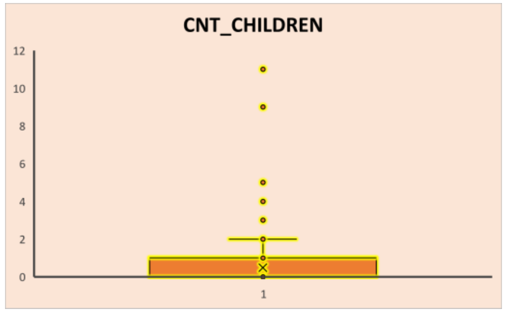


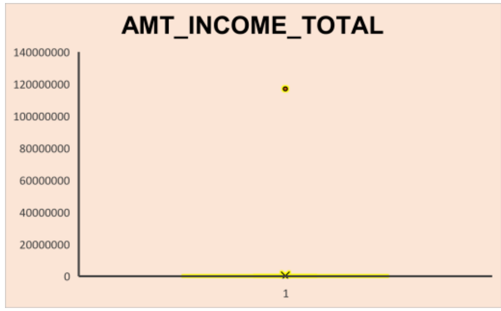
For variables which have skewedness, we replaced the missing values with median and for non-skewed, the missing values are replaced with mean.

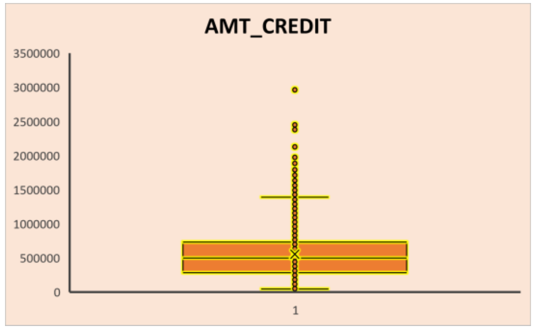
- **NFLAG_INSURED_ON_APPROVAL** is replaced with **mode** .
- **AMT_ANNUNITY,AMT_GOODS, CNT_PAYMENT** is replaced with **mean/average**.
- **DAYS_LAST_DUE_1st_VERSION,DAYS_TERMINATION,DAYS_FIRST_DRAWING,DAYS_FIRST_DUE** are replaced with **median** values.

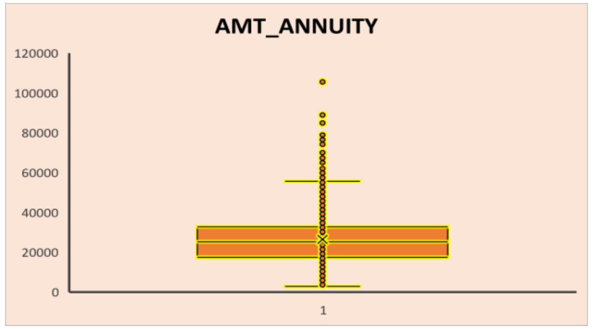
B. Identify OUTLIERS in the Dataset

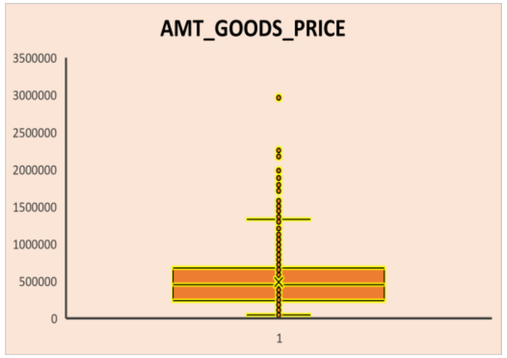
Dataset
application_data.csv

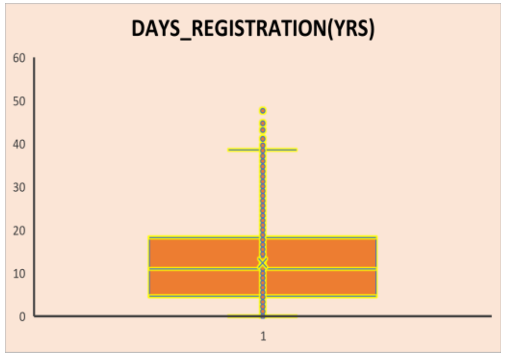


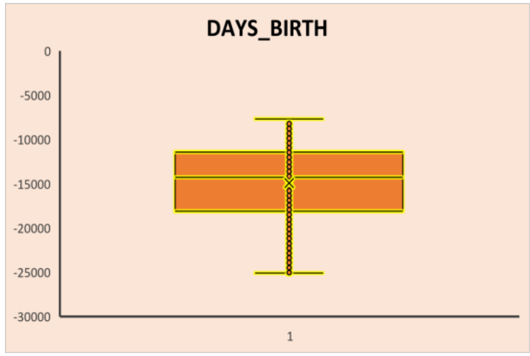


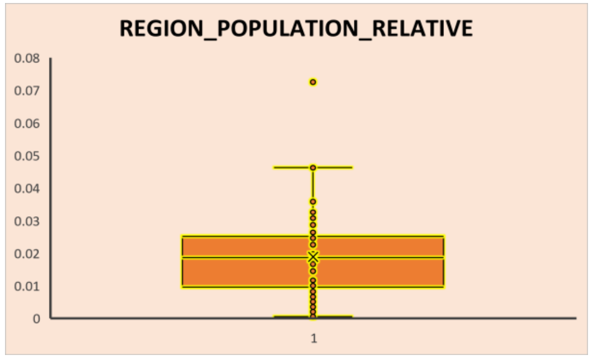


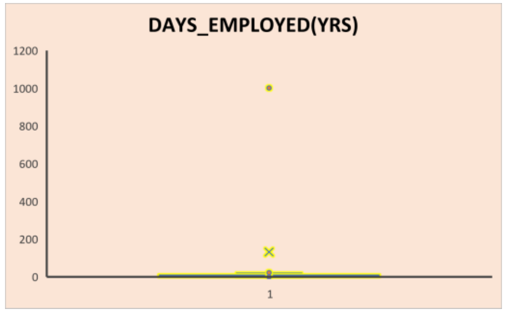


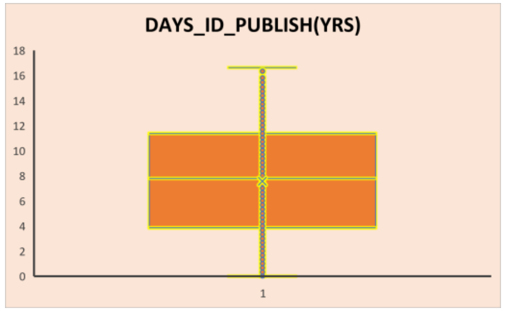






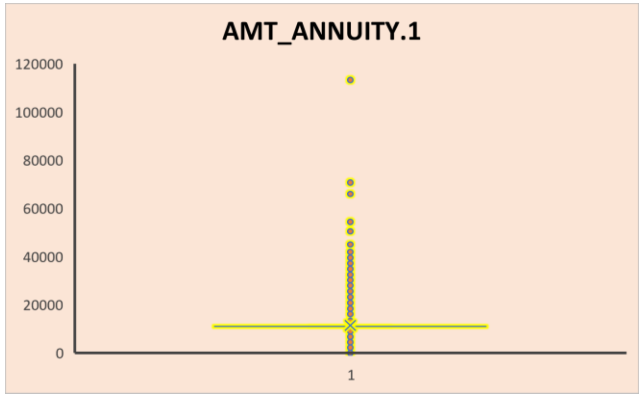


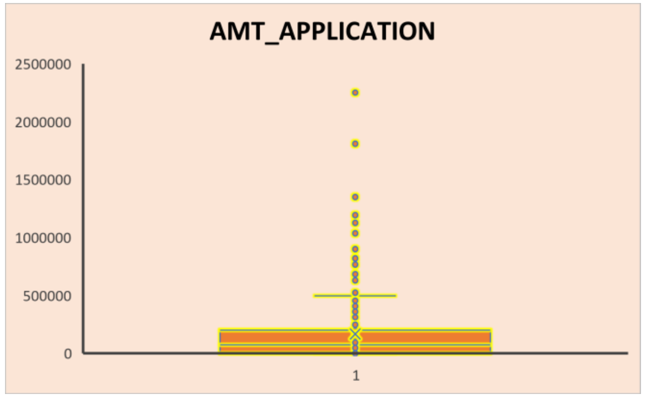


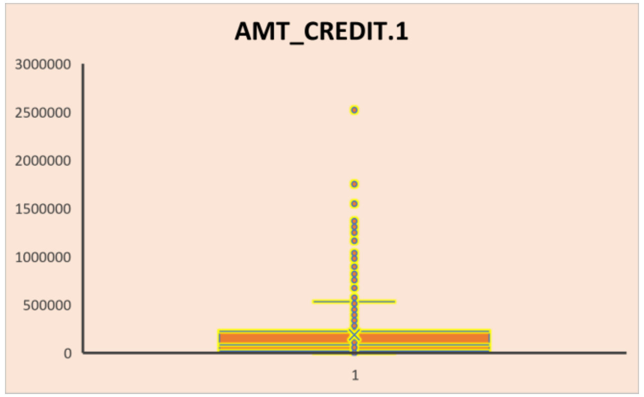


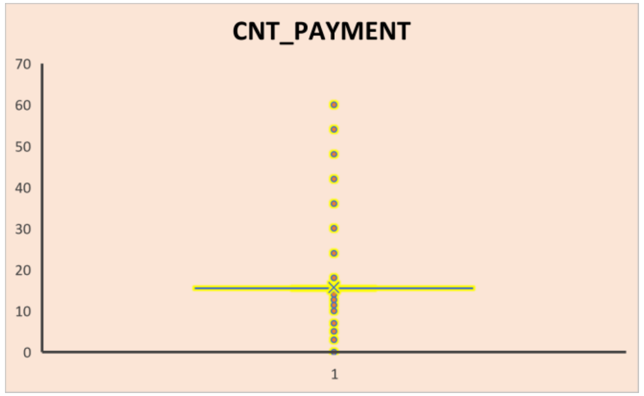
- **DAYS_REGISTRATION(YRS),DAYS_EMPLOYED(YRS)** have some outliers.
- **REGION_POPULATION_RELATIVE,DAYS_BIRTH ,DAYS_ID(YRS)** have no outlier values.
- **CNT_CHILDREN** has values **11** and **9**, which is practically not possible in today's time.
- **AMT_INCOME_TOTAL** has a outlier **117000000**.
- **AMT_CREDIT** has a outlier **2961000,2447937,2377431,2125943**, etc.
- **AMT_ANNUNITY** has a outlier **105511.5** and few more values.
- **AMT_GOODS_PRICE** has outliers **2961000** and some more values.

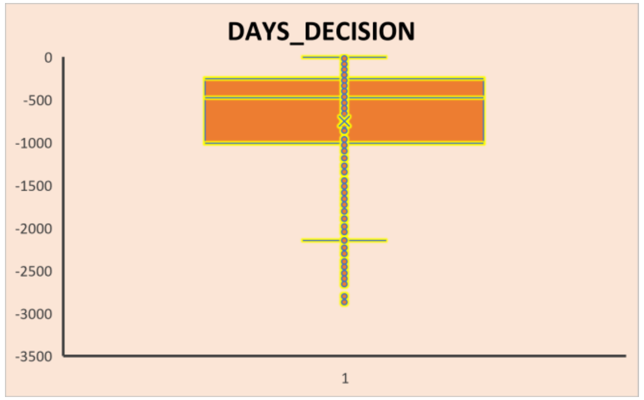
Dataset
previous_data.csv











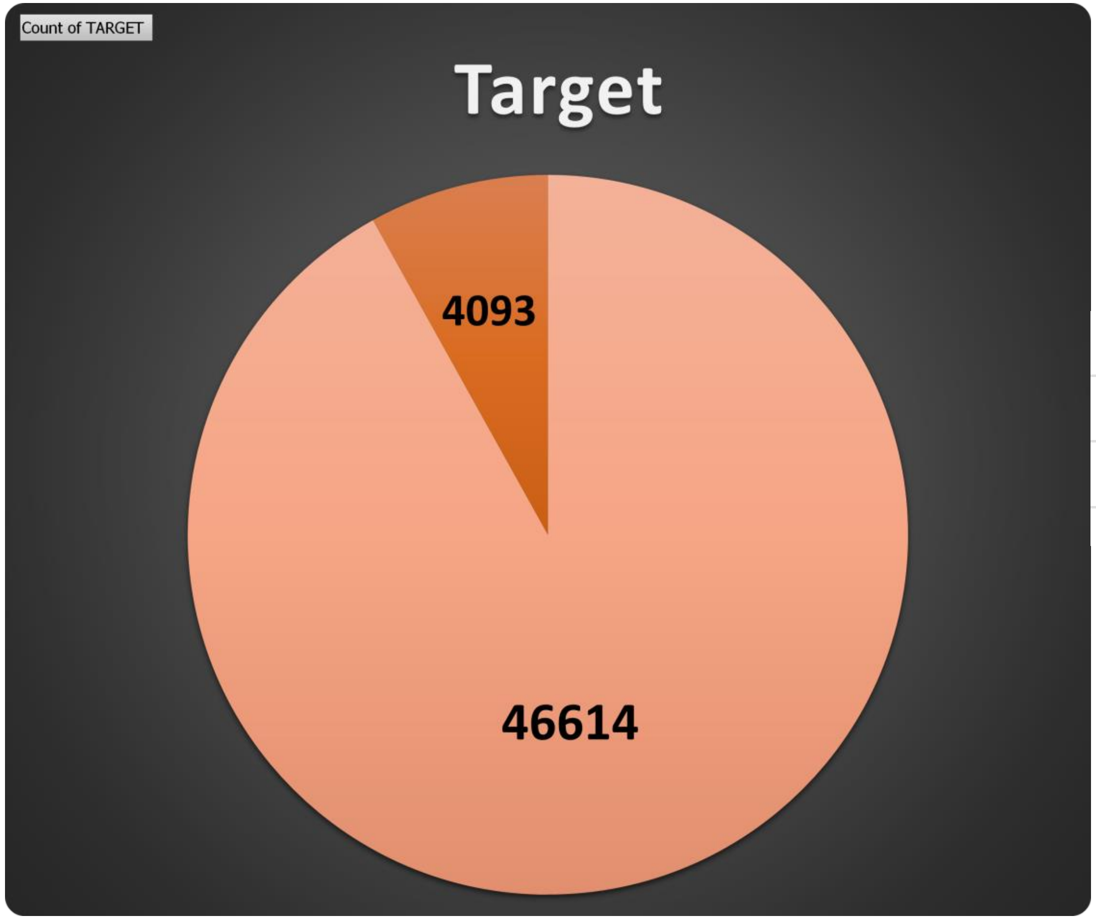
- We observe that **AMT_ANNUNITY,AMT_APPLICATION,AMT_CREDIT,AMT_GOODS_PRICE** has huge number of outliers.
- **DAYS_DECISION** also has some outliers which shows that the decisions on them were taken recently.
- **CNT_PAYMENT** has few outliers present.

Merging the Datasets

- Now the datasets, **application_data** and **previous_application** are both merged into one.

- Also, we did **feature engineering** and created **4** new columns, **DAYS_EMPLOYED(YRS), DAYS_REGISTRATION(YRS), DAYS_ID_PUBLISH(YRS), YEARS_BIRTH** by dividing the number of days by 365 and taking it's absolute value, since the column has negative values.

C. Analyze Data Imbalance



Row Labels	Count of TARGET
0	46614
1	4093
Grand Total	**50707**

	Count of 0s and 1s	Ratio	Contribution
0s	46614	11.3887	91.93%
1s	4093		8.07%

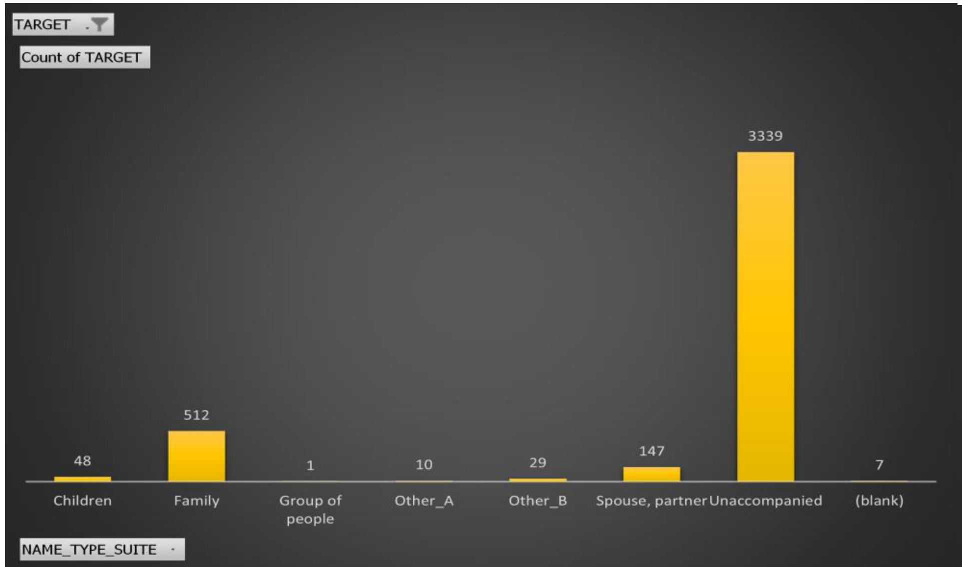
1 - Customers with payment difficulties

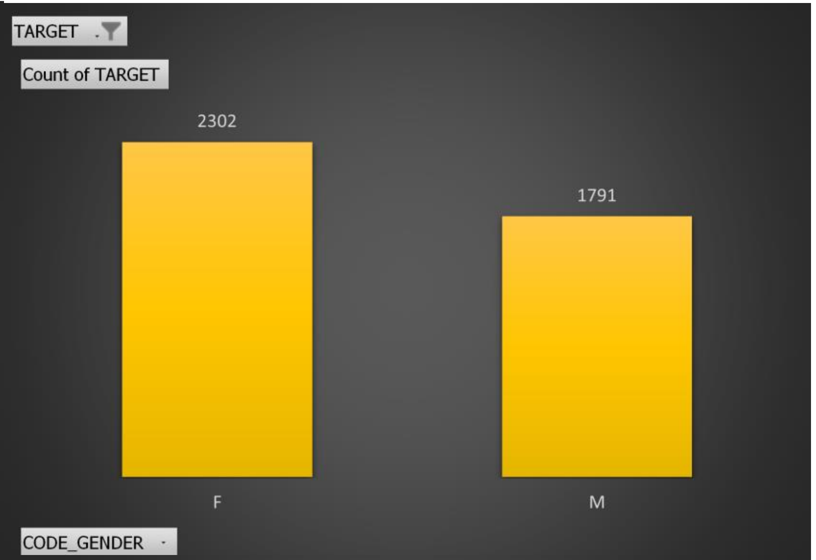
0 - All other cases.

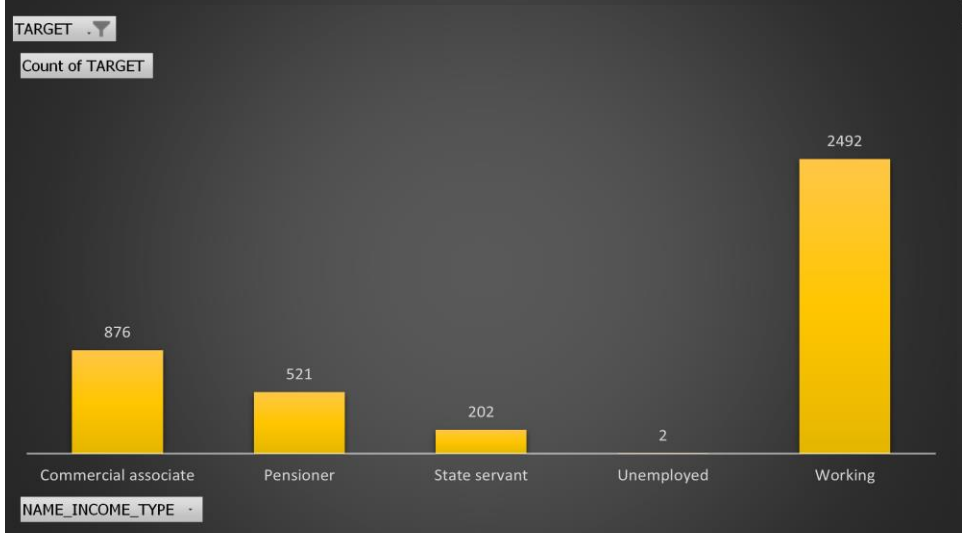
From the **pie chart**, we clearly see the imbalance in the data.

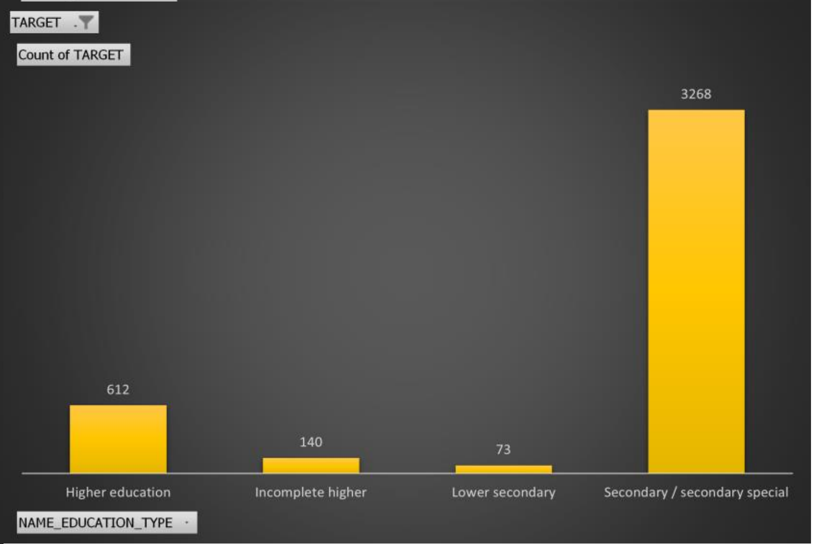
D. Perform Univariate, Segmented Univariate, and Bivariate Analysis

Univariate Analysis

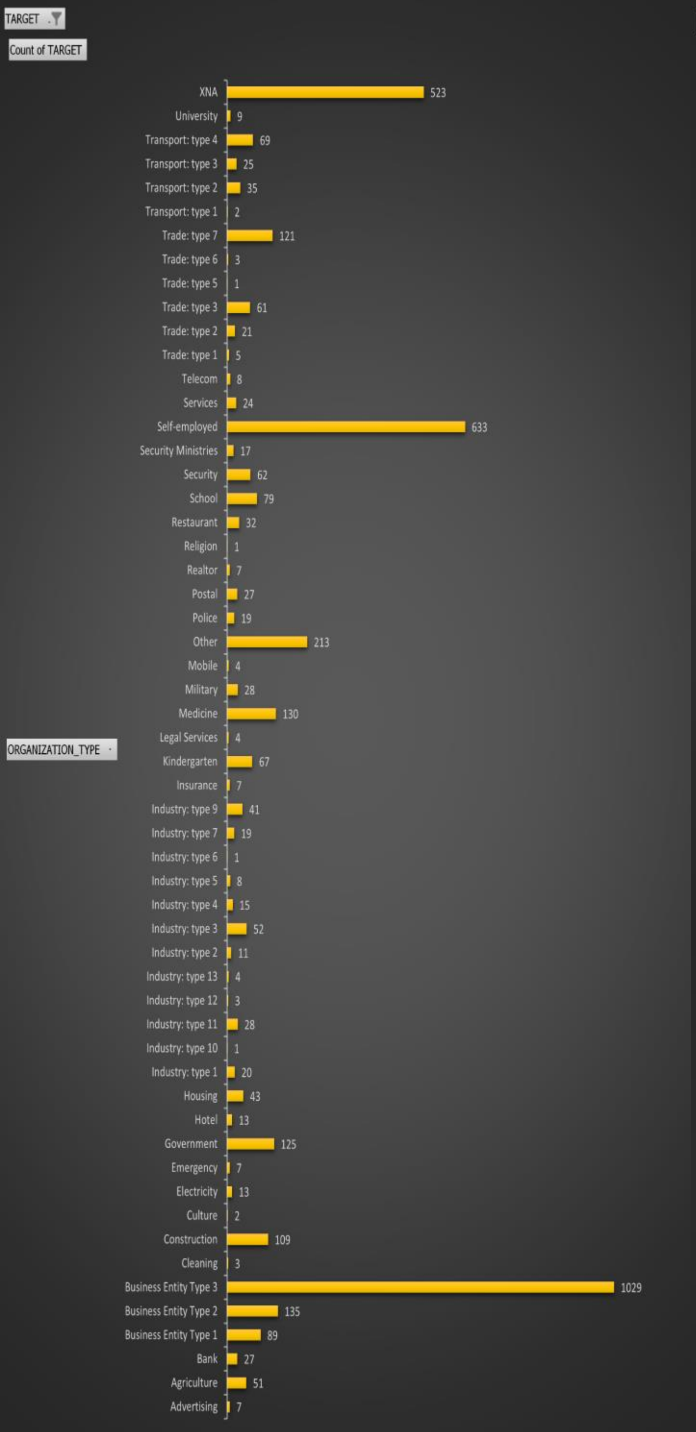


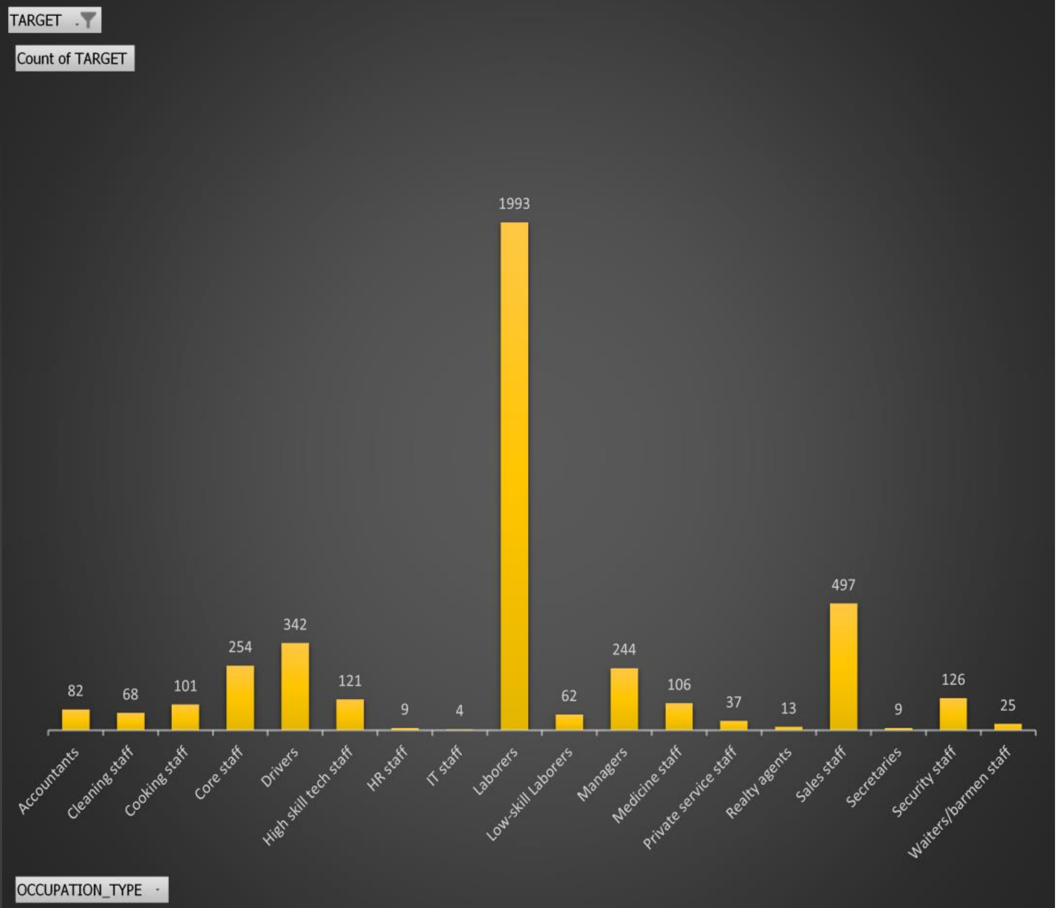




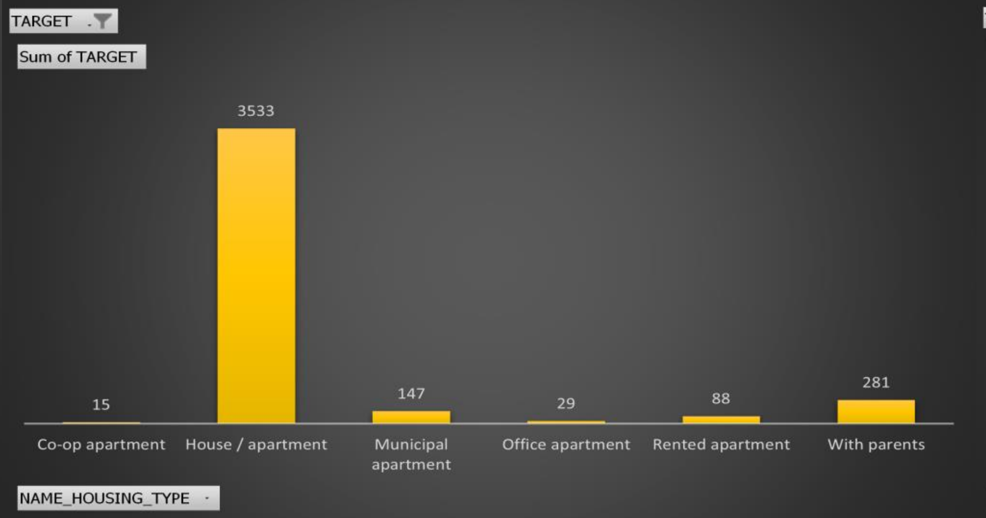


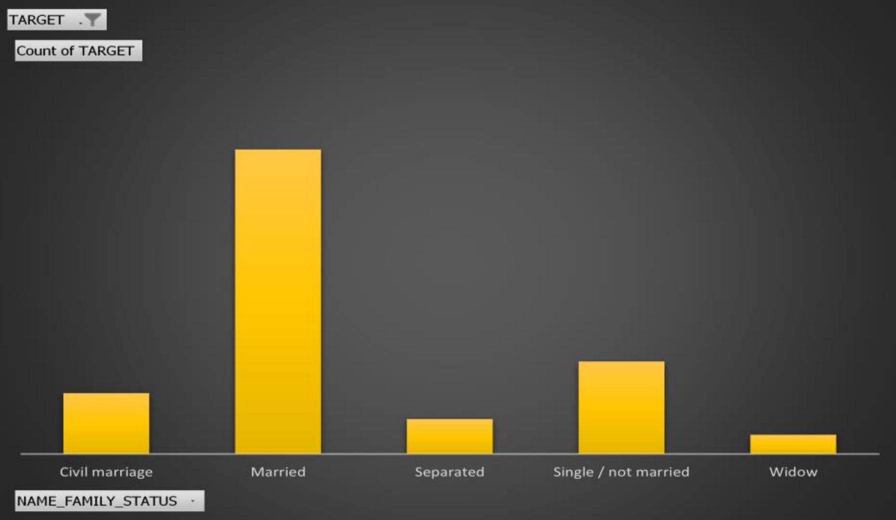
- **NAME_TYPE_SUITE** has the highest count of **Unaccompanied**.
- **Females** are more defaulters.
- Most of the defaulters are **working** professionals.
- Most customers who default have a education till **secondary, secondary/special.**

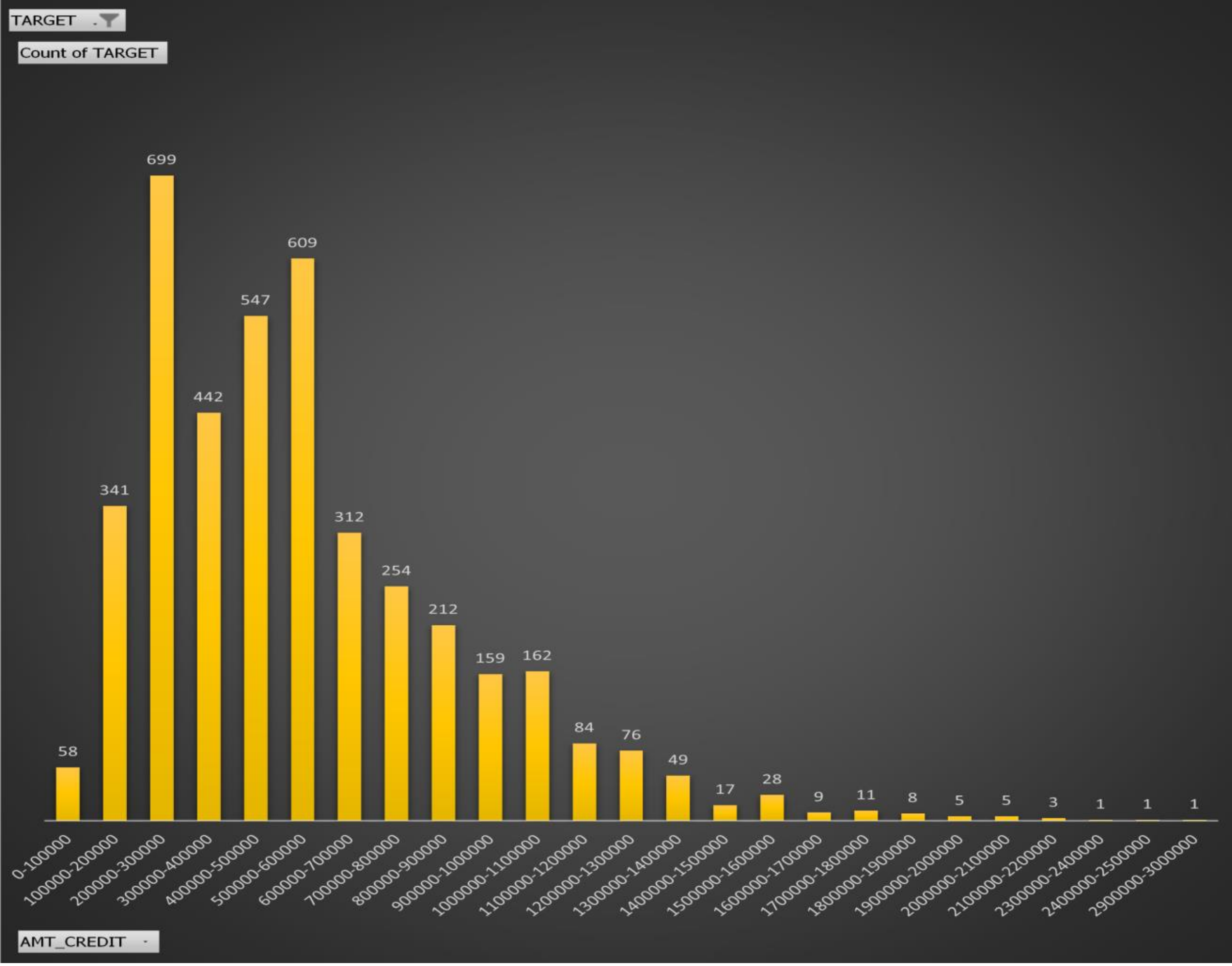


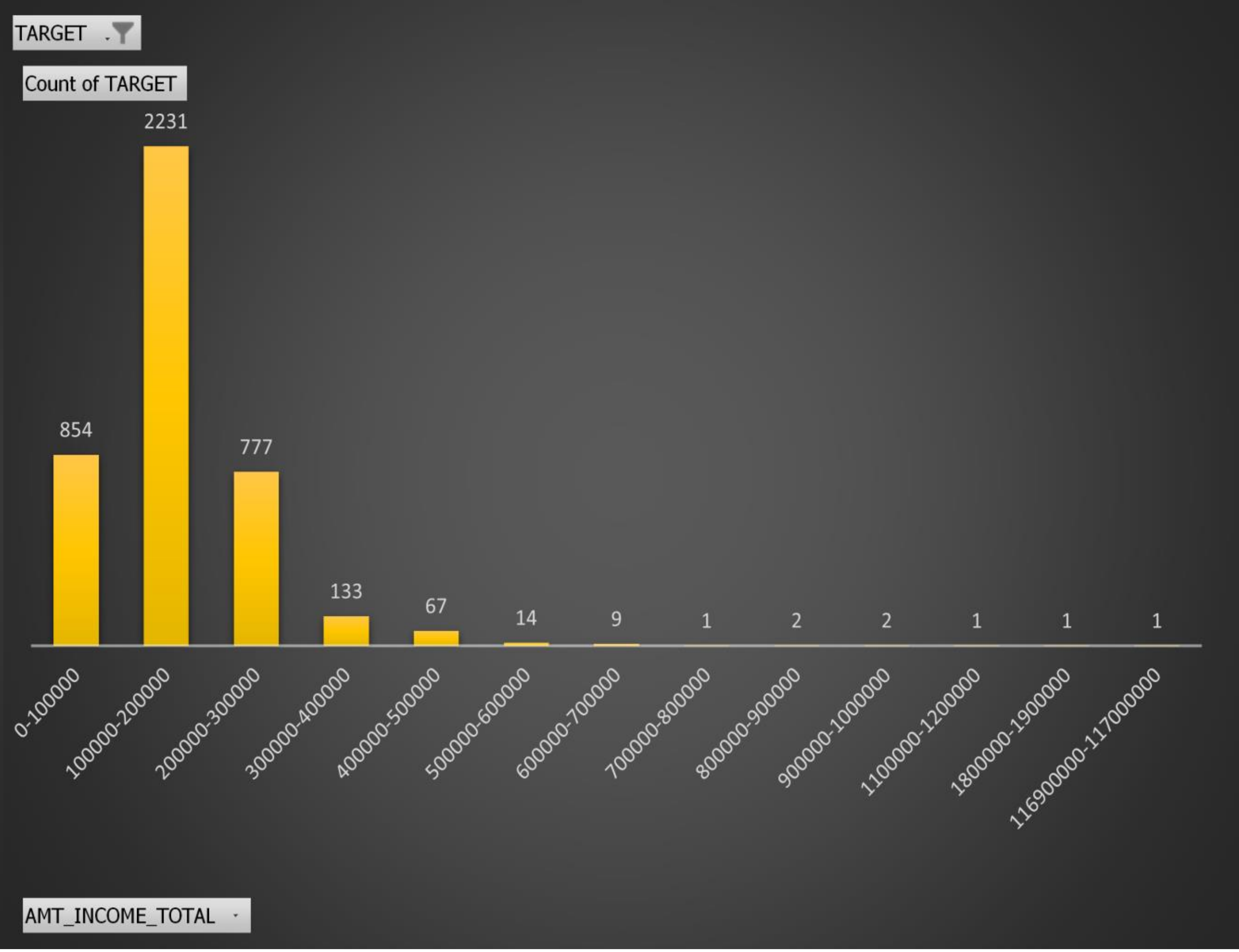


- **Business Entity Type 3** is the type of organization with most defaulters.
- A large number of defaulters are **Laborers**.
- Customers who default lives usually in a **house/apartment** and are **Married**.





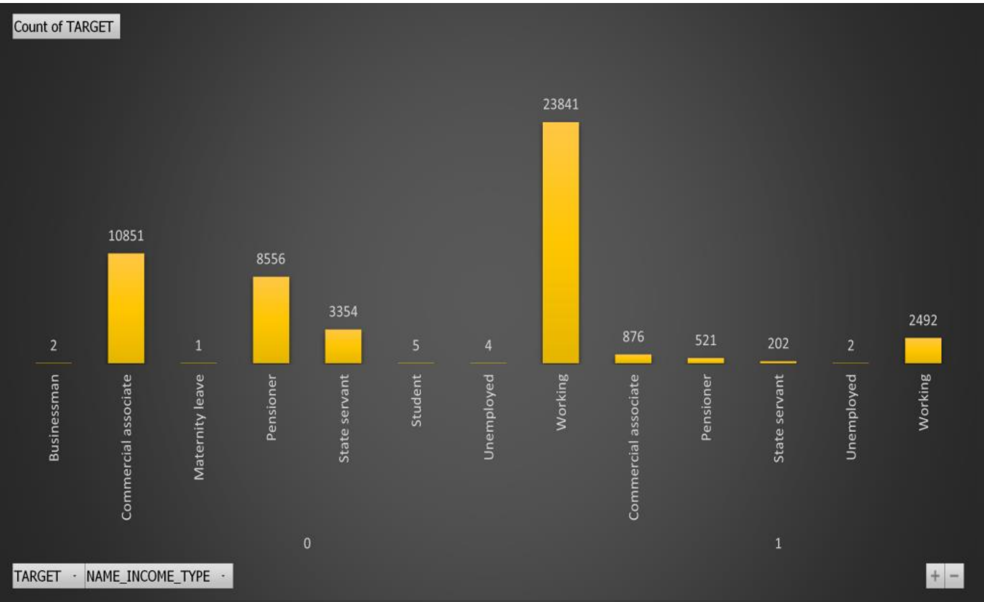


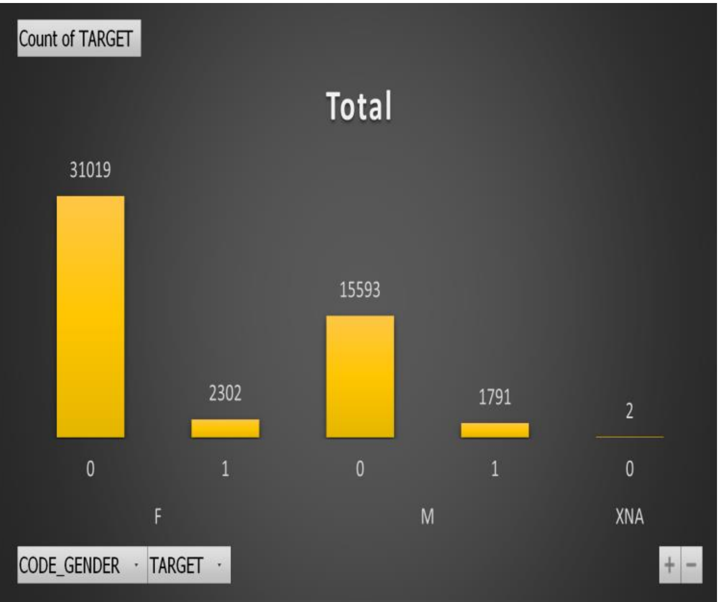


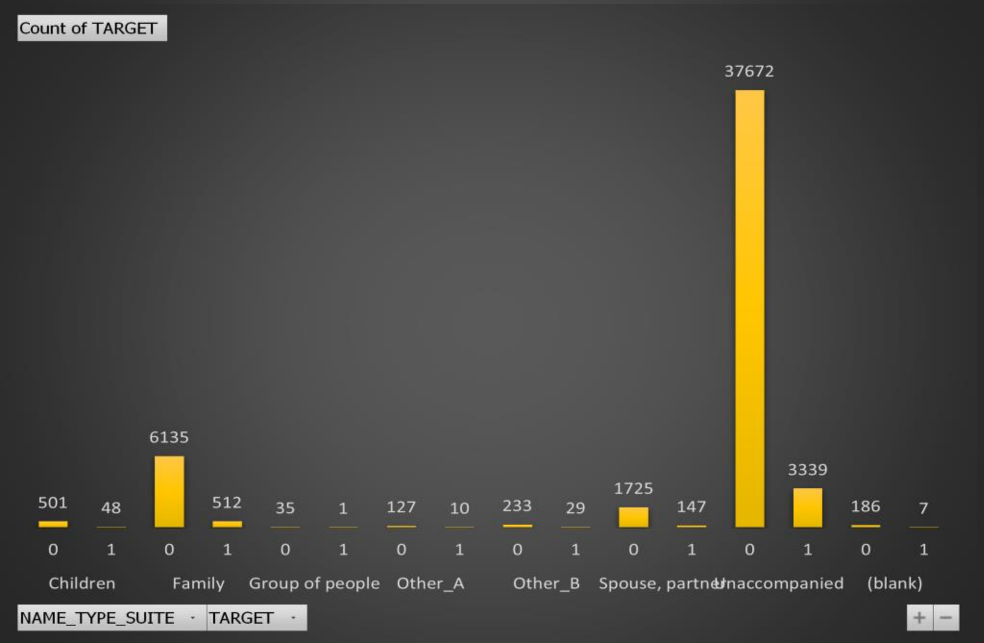
- Maximum number of defaulters gets an approved credit limit between **200000-300000**

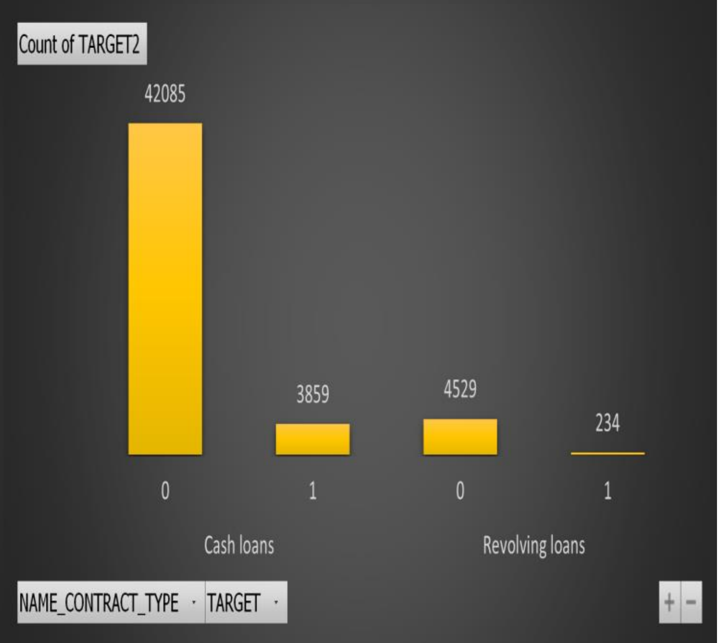
- Customers who default mostly earns between **100000-200000**

Segmented Univariate Analysis







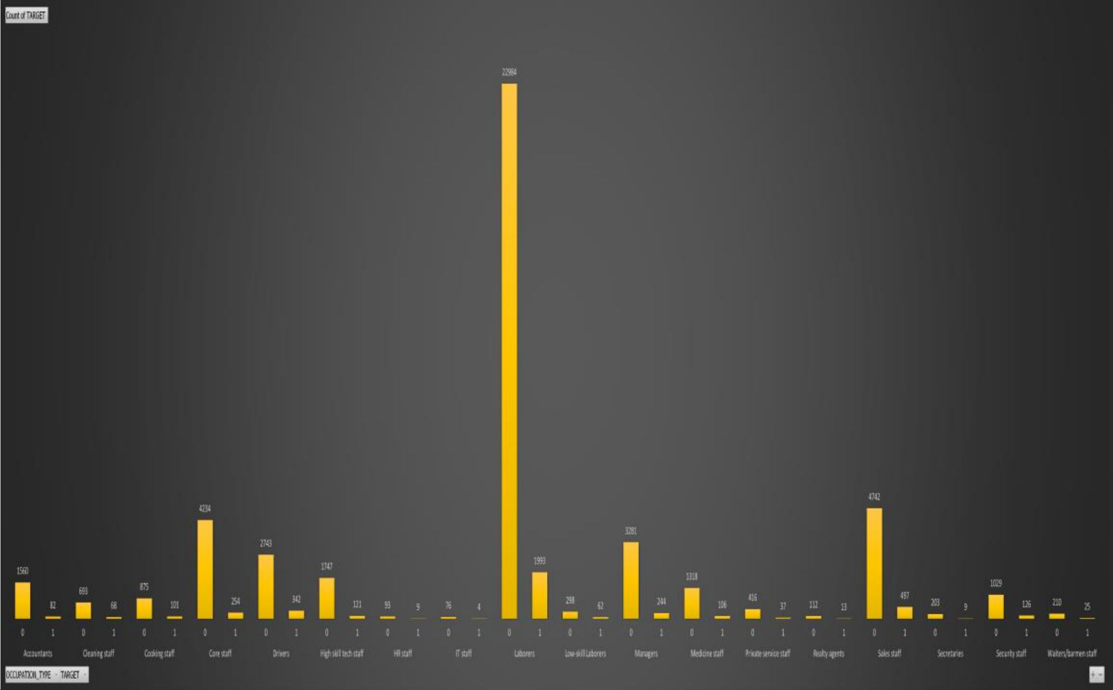


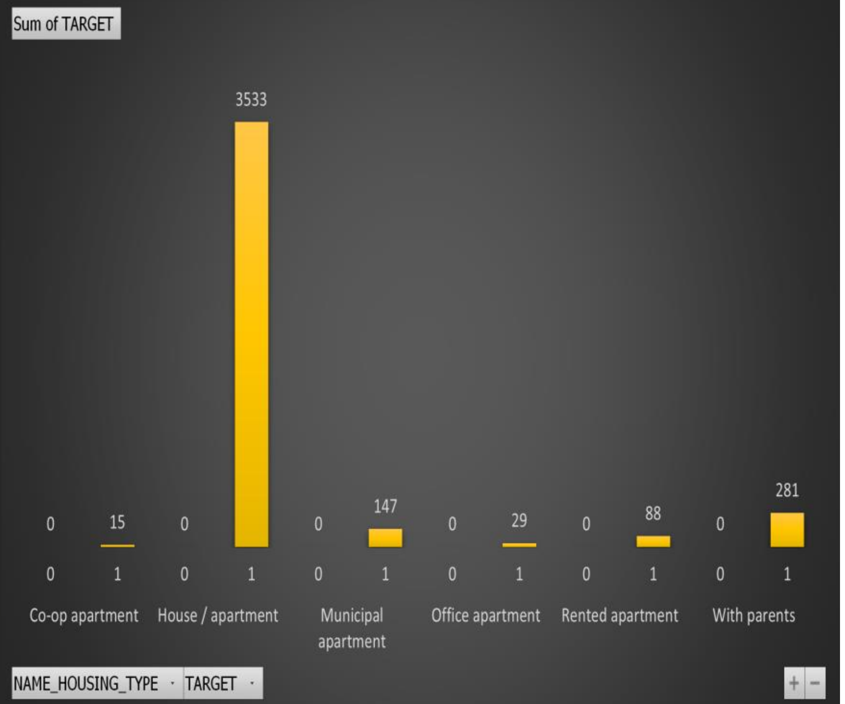
For Defaulters:

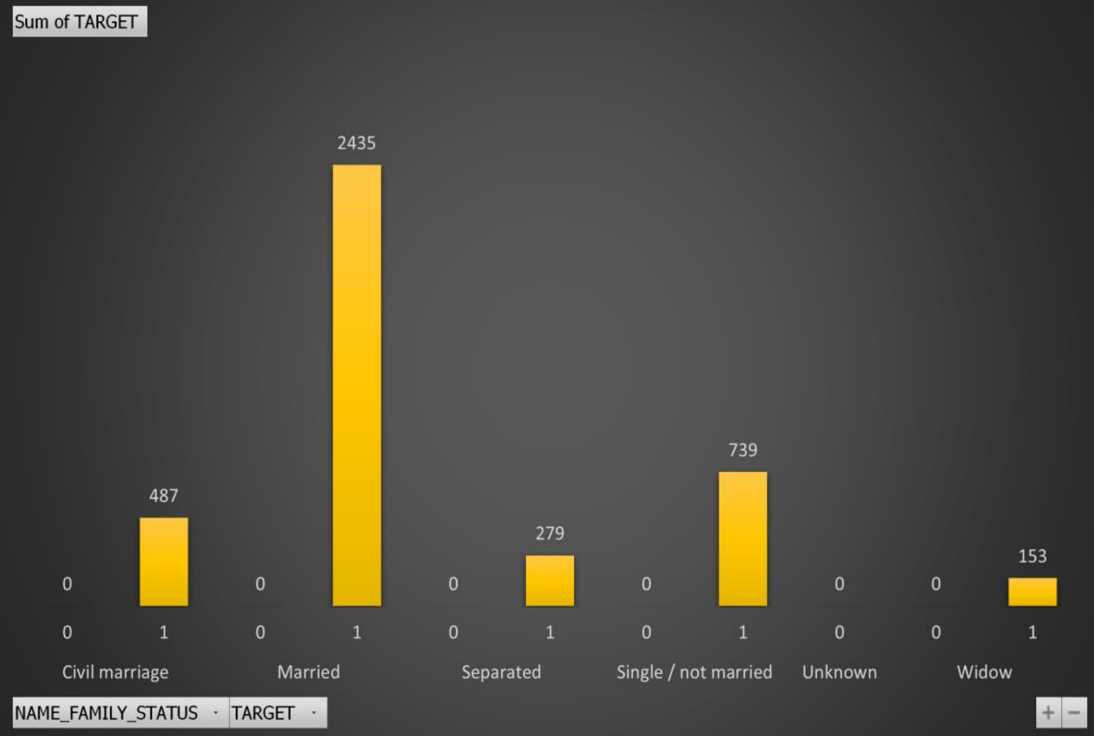
- **Working professionals** mostly default.
- **XNA** most likely defaults.

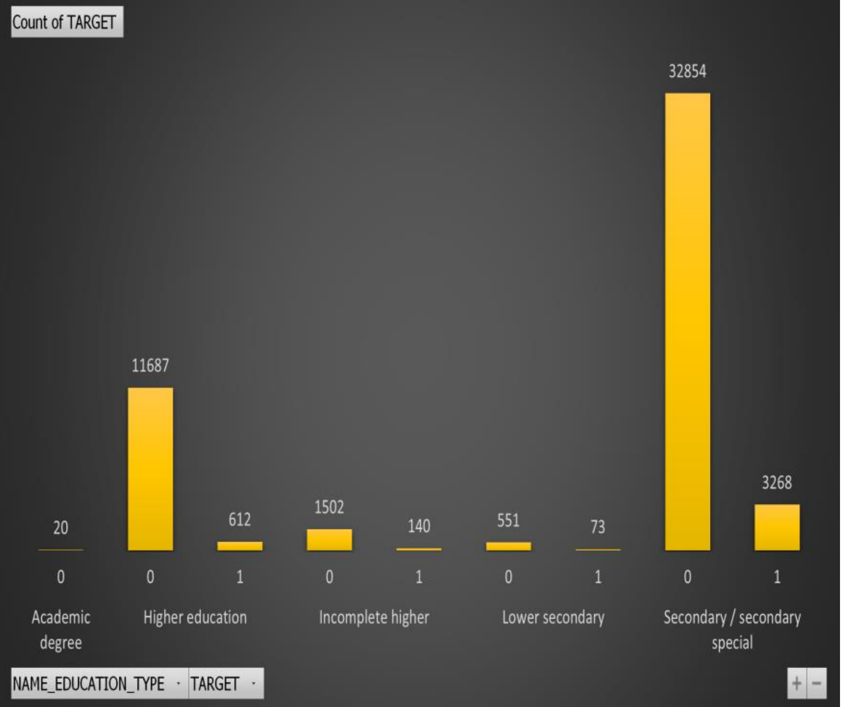
For Repayers:

- **Commercial associates** and **Pensioners** are more likely to repay.
- Both **Males** and **Females** are good repayers.
- **Cash loans** are much in number when compared to **revolving loans** are it gets repaid more.







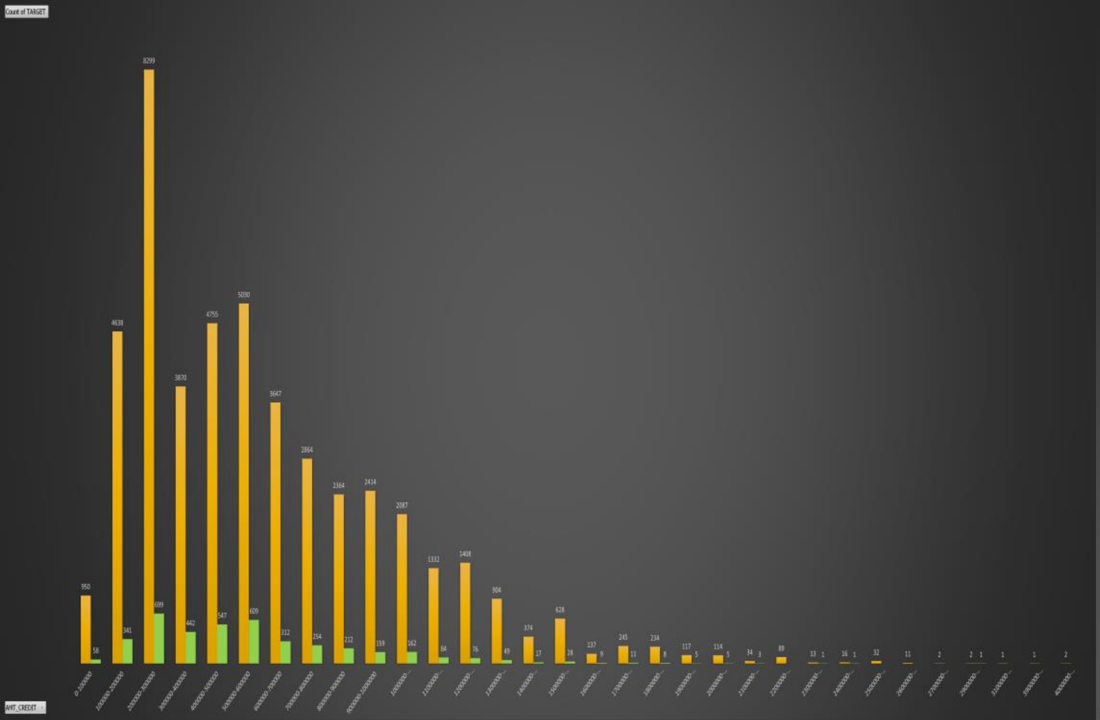


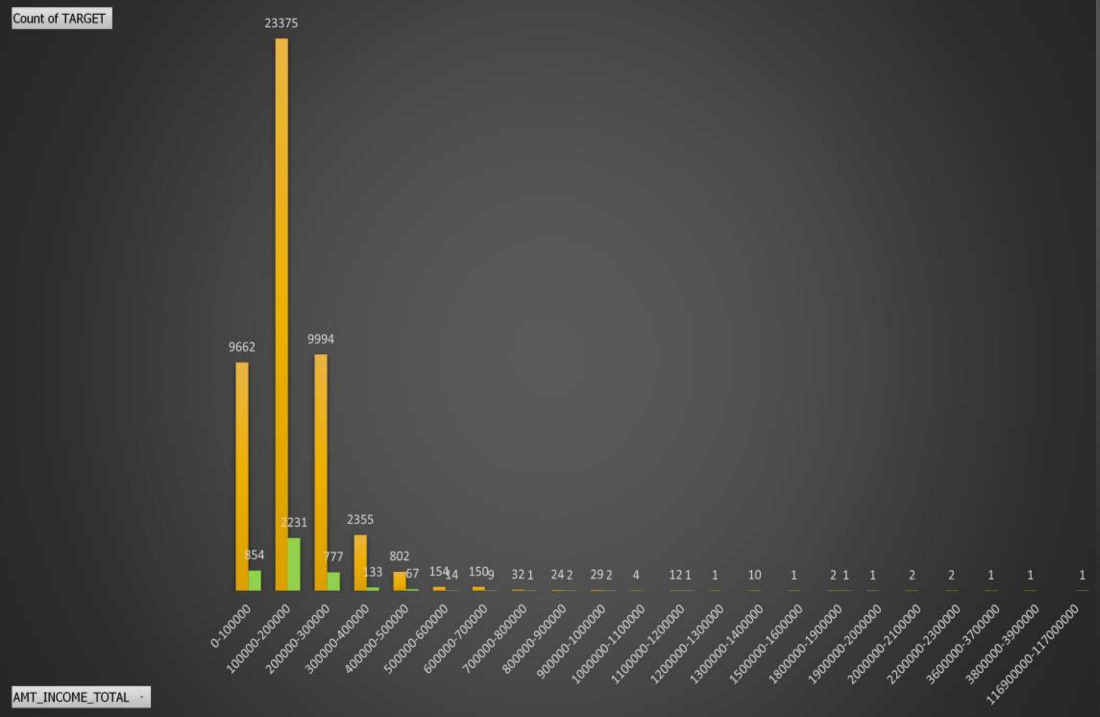
For Defaulters:

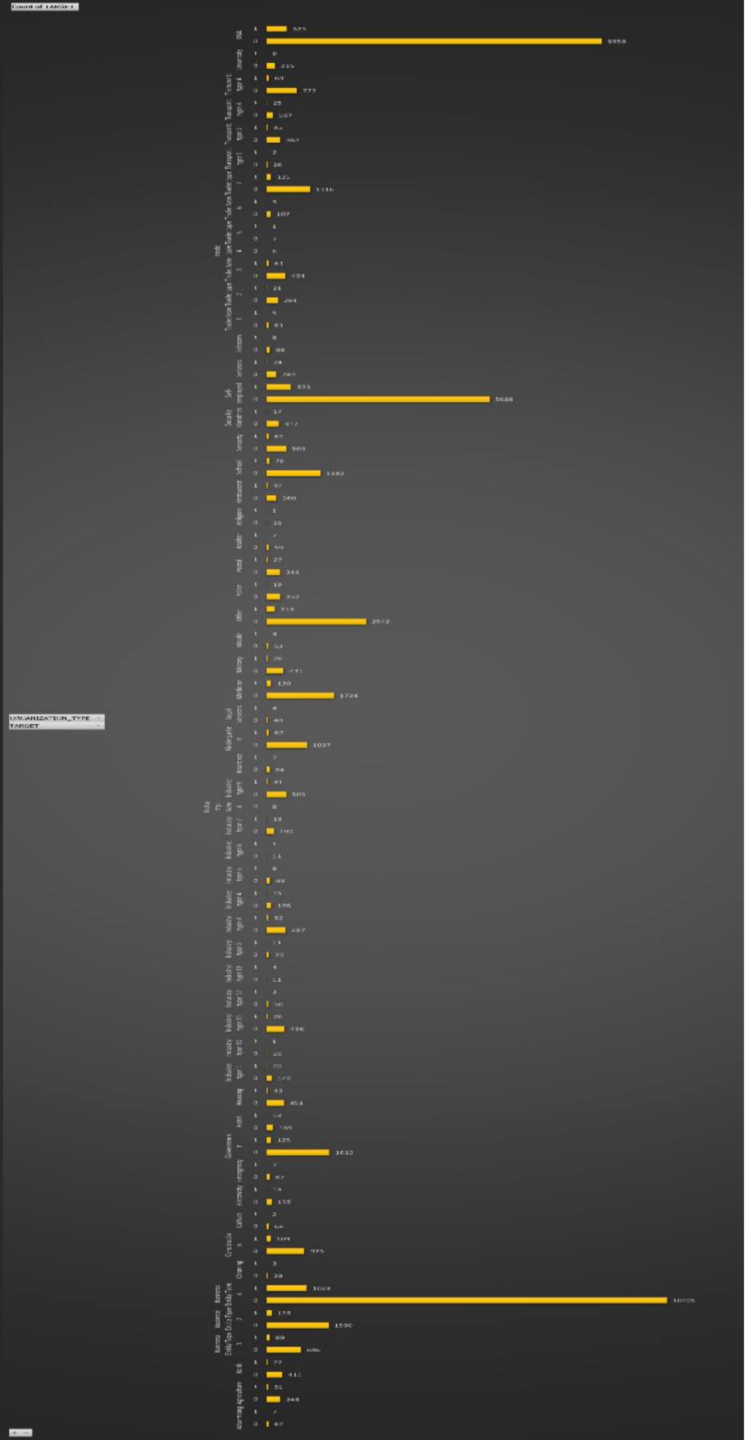
- For every **housing_type**, there are more defaulters than repayers.
- **Married** people are most likely to default than others.

For Repayers:

- Every **occupation_type** has more repayers than defaulters.
- For any **education_type**, repayers are more than defaulters.







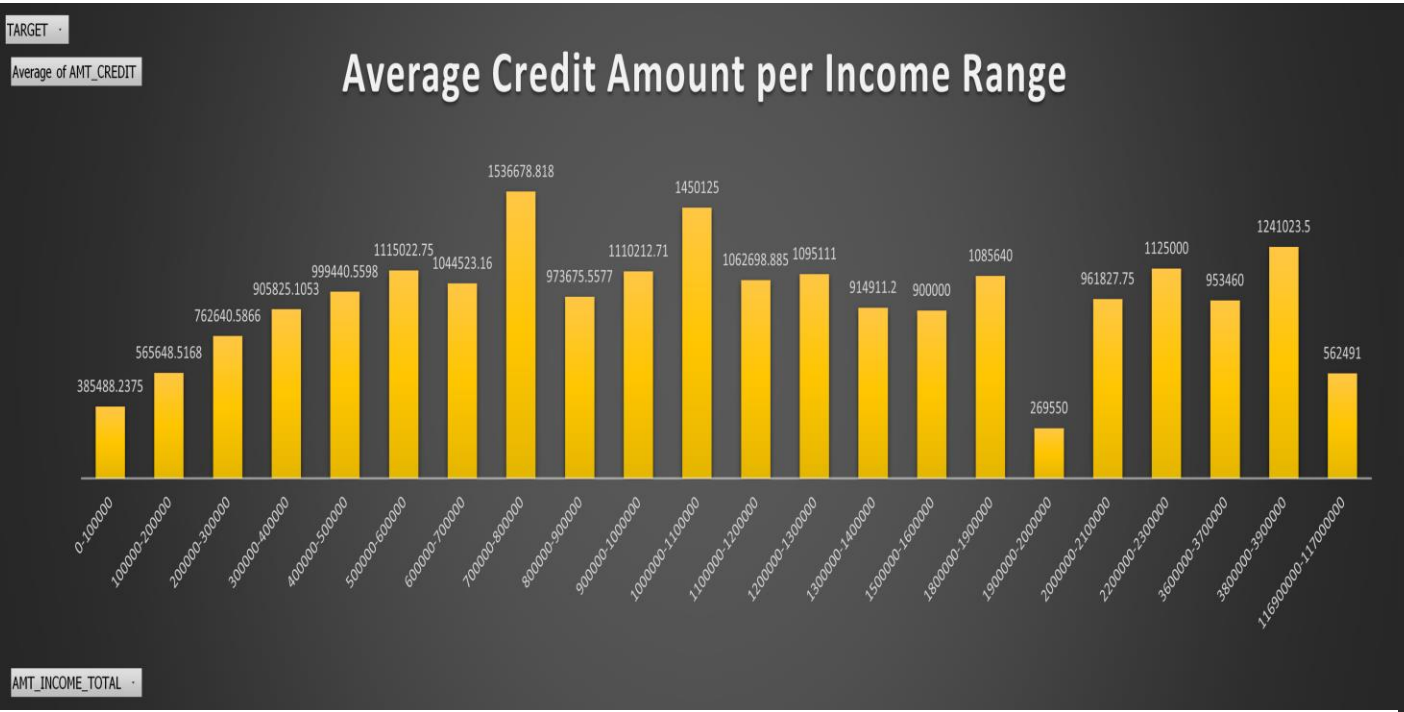
For Defaulters:

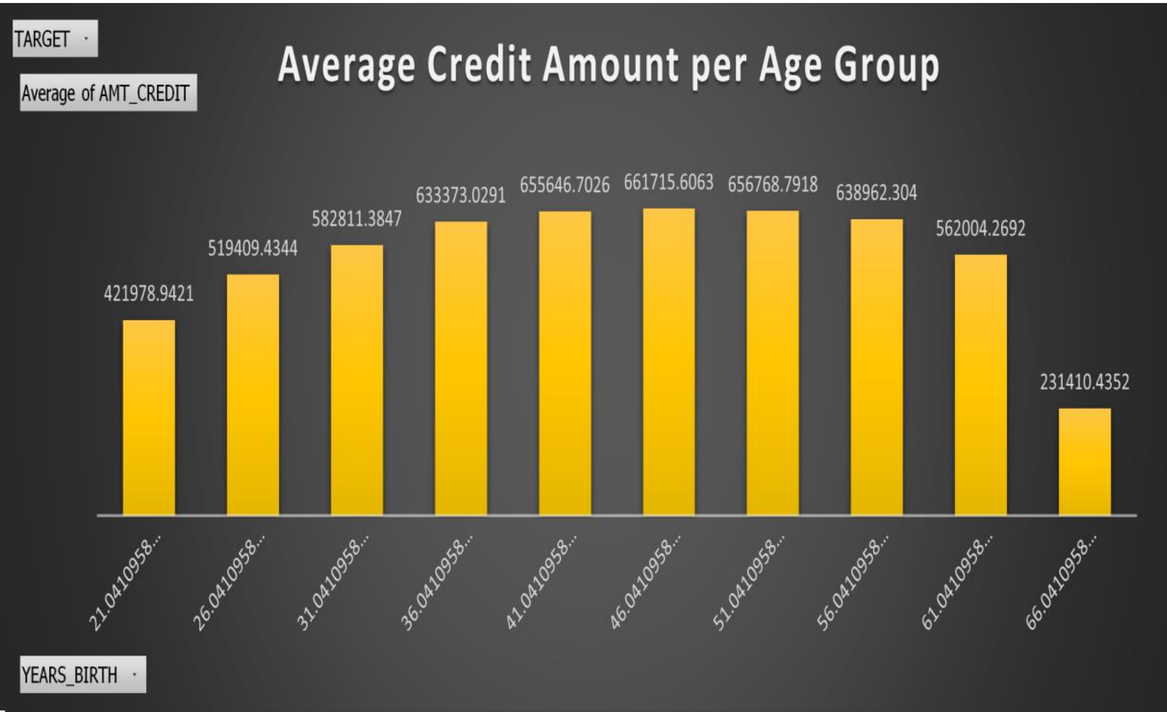
- Customers who are offered credit limit between **200000-600000** are most defaulters.
- Most defaulters are also in the range **100000-200000**.
- **Business Entitiy Type-3** has greater number of defaulters.

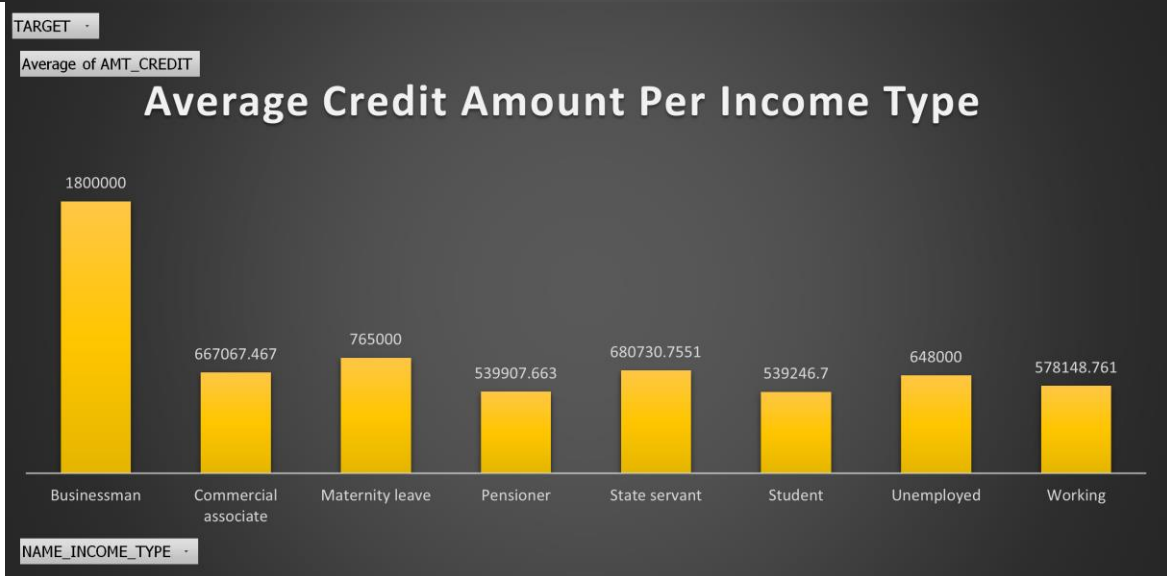
For Repayers:

- Most repayers are offered credit limit between **200000-300000**
- People earning in range **100000-200000** mostly repay.
- **Business Entity Type -3** are better repayers, followed by **XNA** and **Self-employed**.

Bivariate Analysis







- Average credit amount for customers with income range **700000-800000** is the highest followed by customers in range **1000000-1100000** and **3800000-3900000**.
- People in the age group **46-51** are offered greater credit amount followed by **51-56,41-46** and **36-41**.
- **Businessmens** gets most credit amount followed by **maternity leave** and **state servant**.

E. Identify Top Correlations for Different Scenarios :

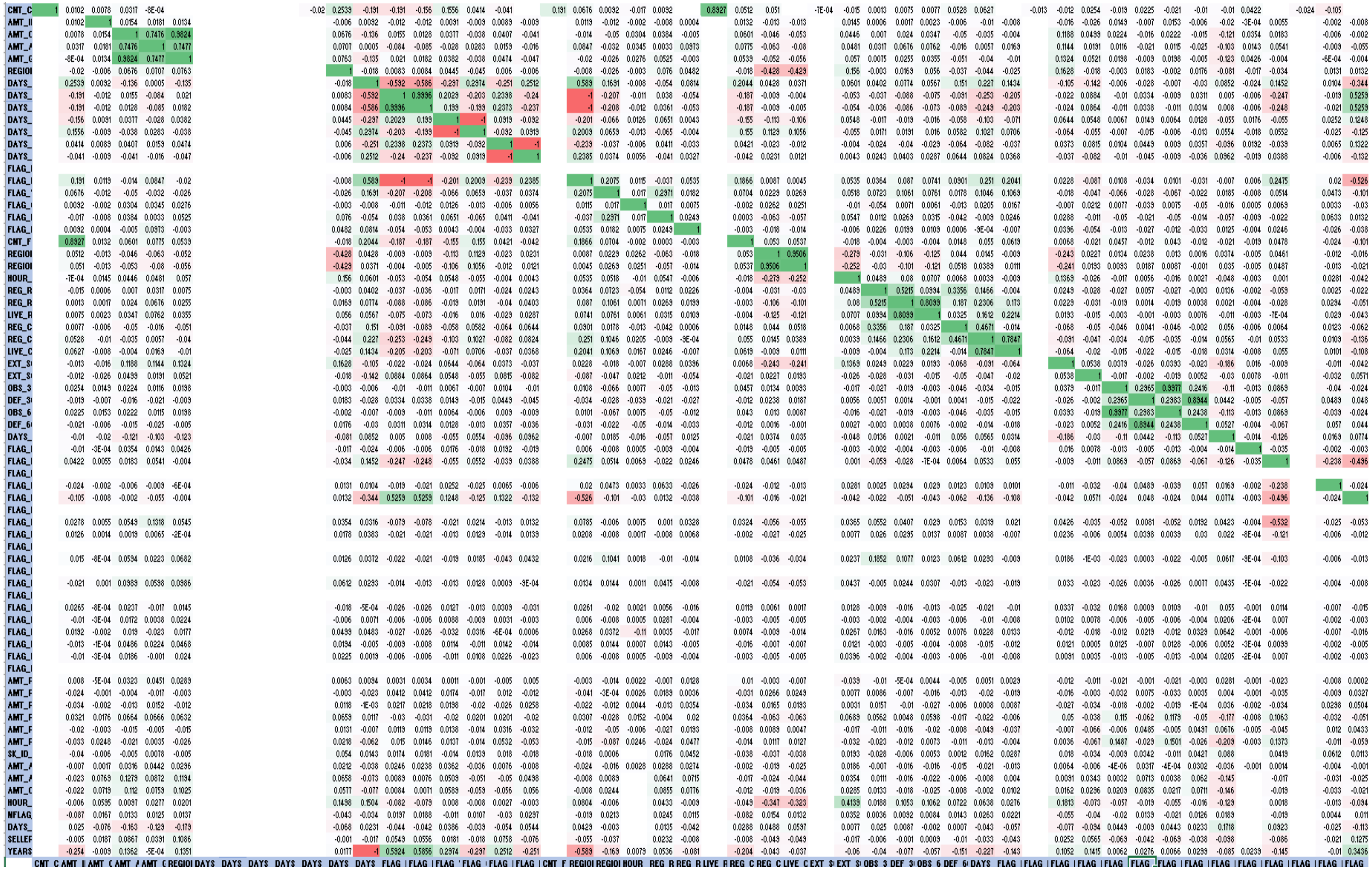
CNT_CHILDREN	1	0.036211397	0.0047218	0.02595499	0.000703586	-0.02454237	0.3368	-0.245831996	-0.243899464	-0.183349114	0.183349114	0.032356229	-0.032356229	0.002669369	0.244929469	0.054578899	-0.003853847	-0.03073248	0.0293112
AMT_INCOME_TOTAL	0.036211397	1	0.3778101	0.45129198	0.38440565	0.183491435	0.073225	-0.161594202	-0.162630969	-0.067944761	0.067944761	-0.032428854	0.032428854	0.002005182	0.162140781	-0.035307007	-0.016800969	0.002383526	0.0890271
AMT_CREDIT	0.004721827	0.377810095	1	0.77052224	0.987234083	0.096906618	-0.051662	-0.074015812	-0.076644685	-0.006382165	0.006382165	0.008827564	-0.008827564	0.003694034	0.075286058	-0.011059311	0.024234848	0.016384442	0.0121067
AMT_ANNUITY	0.025954993	0.451291983	0.7705222	1	0.775886821	0.118224801	0.009862	-0.111142264	-0.11284122	-0.033529213	0.033529213	-0.009433475	0.009433475	0.000389611	0.111982097	-0.019758929	0.022455835	0.003937147	0.0637256
AMT_GOODS_PRICE	0.000703586	0.38440565	0.9872341	0.77588682	1	0.100215727	-0.049317	-0.071852455	-0.074508224	-0.009676851	0.009676851	0.009854296	-0.009854296	0.003605937	0.073154725	0.010673399	0.021946333	0.032231806	0.0116177
REGION_POPULATION_REL/	-0.02454237	0.183491435	0.0969066	0.1182248	0.100215727	1	-0.031174	-0.00641865	-0.006276027	0.058429049	-0.058429049	0.001958195	-0.001958195	0.003434316	0.006406891	-0.015021622	-0.004861763	0.094185505	0.040755
DAYS_BIRTH	0.336800254	0.07322541	-0.051662	0.00986173	-0.049316793	-0.031174261	1	-0.623144892	-0.614967073	-0.334931261	0.334931261	-0.270141173	0.270141173	-0.00801984	0.619117077	0.175852052	-0.012764645	-0.04146094	0.0939393
DAYS_EMPLOYED(YRS)	-0.245832	-0.161594202	-0.074016	-0.11114226	-0.071852455	-0.00641865	-0.623145	1	0.999533745	0.208279891	-0.208279891	0.275076971	-0.275076971	0.002267501	-0.999738644	-0.234046478	0.016807913	0.023820104	-0.069346
DAYS_EMPLOYED	-0.24389946	-0.162630969	-0.076645	-0.11284122	-0.074508224	-0.006276027	-0.614967	0.999533745	1	0.203799264	-0.203799264	0.272789582	-0.272789582	0.002125836	-0.999742029	-0.234450964	0.016770216	0.022267672	-0.068457
DAYS_REGISTRATION(YRS)	-0.18334911	-0.067944761	-0.006382	-0.03352921	-0.009676851	0.058429049	-0.334931	0.208279891	0.203799264	1	-1	0.103696353	-0.103696353	0.000306279	-0.20600394	-0.059284102	-4.69776E-05	0.071475727	-0.031195
DAYS_REGISTRATION	0.183349114	0.067944761	0.0063822	0.03352921	0.009676851	-0.058429049	0.334931	-0.208279891	-0.203799264	-1	1	-0.103696353	0.103696353	-0.00030628	0.20600394	0.059284102	4.69776E-05	-0.07147573	0.0311953
DAYS_ID_PUBLISH(YRS)	0.032356229	-0.032428854	0.0088276	-0.00943347	0.009854296	0.001958195	-0.270141	0.275076971	0.272789582	0.103696353	-0.103696353	1	-1	0.005974594	-0.274062323	-0.052332251	0.00495265	0.031674727	-0.032638
DAYS_ID_PUBLISH	-0.03235623	0.032428854	-0.008828	0.00943347	-0.009854296	-0.001958195	0.270141	-0.275076971	-0.272789582	-0.103696353	0.103696353	-1	1	-0.00597459	0.274062323	0.052332251	-0.00495265	-0.03167473	0.0326377
FLAG_MOBIL	0.002669369	0.002005182	0.003694	0.00038961	0.003605937	0.003434316	-0.00802	0.002267501	0.002125836	0.000306279	-0.000306279	0.005974594	-0.005974594	1	-0.002196763	0.002288244	-0.000211508	0.002903183	0.001131
FLAG_EMP_PHONE	0.244929469	0.162140781	0.0752861	0.1119821	0.073154725	0.006406891	0.619117	-0.999738644	-0.999742029	-0.20600394	0.20600394	-0.274062323	0.274062323	-0.00219676	1	0.23431092	-0.016794435	-0.02309294	0.0689272
FLAG_WORK_PHONE	0.054578899	-0.035307007	-0.011059	-0.01975893	0.010673399	-0.015021622	0.175852	-0.234046478	-0.234450964	-0.059284102	0.059284102	-0.052332251	0.052332251	0.002288244	0.23431092	1	0.02255983	0.298986833	-0.009282
FLAG_CONT_MOBILE	-0.00385385	-0.016800969	0.0242348	0.02245583	0.021946333	-0.004861763	-0.012765	0.016807913	0.016770216	-4.69776E-05	4.69776E-05	0.00495265	-0.00495265	-0.00021151	-0.016794435	0.02255983	1	0.004561135	-0.011321
FLAG_PHONE	-0.03073248	0.002383526	0.0163844	0.00393715	0.032231806	0.094185505	-0.041461	0.023820104	0.022267672	0.071475727	-0.071475727	0.031674727	-0.031674727	0.002903183	-0.023092937	0.298986833	0.004561135	1	0.015961
FLAG_EMAIL	0.02931118	0.08902709	0.0121067	0.06372557	0.011617733	0.040755027	0.093939	-0.069345677	-0.068457028	-0.031195274	0.031195274	-0.032637652	0.032637652	0.001131003	0.068927217	-0.009282151	-0.011321144	0.01596097	1
CNT_FAM_MEMBERS	0.879277573	0.041510709	0.0637269	0.07746011	0.061867871	-0.023095133	0.285984	-0.235201526	-0.234165143	-0.17144856	0.17144856	0.024803995	-0.024803995	0.000792884	0.234704926	0.06786663	-0.005190071	-0.01772275	0.0250749
REGION_RATING_CLIENT	0.021135033	-0.205487013	-0.102691	-0.12968593	-0.105110251	-0.540127905	0.009272	0.040736505	0.040274752	-0.082966616	0.082966616	0.007784162	-0.007784162	0.000381039	-0.040496014	0.001789838	0.013024097	-0.08957247	-0.060589
REGION_RATING_CLIENT_W.	0.01787778	-0.220428283	-0.11199	-0.14299363	-0.113579879	-0.53776911	0.007297	0.042981579	0.042623759	-0.075089466	0.075089466	0.012590264	-0.012590264	0.000193181	-0.042797958	0.007158576	0.013157715	-0.0845444	-0.058028
HOUR_APPR_PROCESS_ST.	-0.00463528	0.085632993	0.0570547	0.05345153	0.065886738	0.167180812	0.095803	-0.09298402	-0.092337944	0.002910457	-0.002910457	-0.037067078	0.037067078	-0.00131007	0.092612817	0.03573306	0.004334557	0.060906016	0.0238885
REG_REGION_NOT_LIVE_RE	-0.01069483	0.077979116	0.0273509	0.04593748	0.02980906	-0.003265462	0.060293	-0.037602902	-0.036083868	-0.028209933	0.028209933	-0.03349117	0.03349117	0.000563131	0.036852711	0.06393329	0.001622593	0.006159729	0.0169329
REG_REGION_NOT_WORK_I	0.014254782	0.156228287	0.0557378	0.08237461	0.057051864	0.061753259	0.096078	-0.10996042	-0.107384932	-0.034347881	0.034347881	-0.047679717	0.047679717	0.001061471	0.108692242	0.068644524	-0.002507567	-0.00417624	0.0419489
LIVE_REGION_NOT_WORK_I	0.022318629	0.1471799	0.0542228	0.07476485	0.054366925	0.085747708	0.070075	-0.097732819	-0.095663095	-0.022896899	0.022896899	-0.03329426	0.03329426	0.000944514	0.096716113	0.041637908	-0.00271095	-0.00679753	0.0424528
REG_CITY_NOT_LIVE_CITY	0.021256681	0.009476139	-0.021391	-0.00501866	-0.020642351	-0.047151763	0.183868	-0.096278603	-0.092993912	-0.068597505	0.068597505	-0.076032473	0.076032473	0.001336733	0.09465966	0.054926313	-0.0027248	-0.04185609	0.0102475
REG_CITY_NOT_WORK_CITY	0.07098096	0.014992309	-0.013795	0.00181224	-0.014455014	-0.039198321	0.23581	-0.258293176	-0.254518943	-0.091738737	0.091738737	-0.102438167	0.102438167	0.002504548	0.256459965	0.122793615	-0.003436978	-0.04484911	0.0059118
LIVE_CITY_NOT_WORK_CITY	0.067343231	0.01963313	0.0044301	0.01116141	0.003261324	-0.01169509	0.148374	-0.220310991	-0.218074186	-0.060786033	0.060786033	-0.063287062	0.063287062	0.002141083	0.219241997	0.108633906	-0.00113948	-0.0273959	0.0012678
EXT_SOURCE_2	-0.01272012	0.157077669	0.136605	0.13021111	0.14374391	0.202614655	-0.080486	-0.0312324	-0.033423298	0.055056712	-0.055056712	0.041118795	-0.041118795		0.03227185	-0.013404632	-0.002553194	0.060260664	0.0229328
EXT_SOURCE_3	-0.0392841	-0.070102798	0.0312721	0.0190886	0.033582182	-0.014987317	-0.180852	0.101052509	0.097833027	0.099874814	-0.099874814	0.115626402	-0.115626402	0.000223851	-0.099459722	-0.052883011	0.011825505	-0.00982709	-0.057466
OBS_30_CNT_SOCIAL_CIRC	0.017365131	-0.02907851	-0.002627	-0.01195299	-0.002781398	-0.01549965	0.012827	0.003125772	0.003296178	-0.007414852	0.007414852	0.007175349	-0.007175349	0.00378062	-0.003200372	-0.016184481	0.008334431	-0.02693377	-0.002522
DEF_30_CNT_SOCIAL_CIRCI	-0.00176745	-0.026344225	-0.013698	-0.01990681	-0.014620866	0.010392326	0.00033	0.015301535	0.015582258	0.000881182	-0.000881182	-0.002540366	0.002540366	0.001261153	-0.015413021	-0.01311416	0.002385113	-0.02010567	-0.000838
OBS_60_CNT_SOCIAL_CIRC	0.017398598	-0.029283626	-0.002814	-0.01220893	-0.003037718	-0.015036007	0.012863	0.002902562	0.003068996	-0.008076828	0.008076828	0.007627967	-0.007627967	0.003769913	-0.002976056	-0.016603566	0.008446584	-0.02683045	-0.00193
DEF_60_CNT_SOCIAL_CIRCI	-0.0016711	-0.028854875	-0.01925	-0.02319008	-0.019459428	0.005897053	0.001538	0.014220943	0.014466155	-0.002624127	0.002624127	-0.001861673	0.001861673	0.001078446	-0.014318904	-0.010702803	0.004331893	-0.02054929	0.0002107
DAYS_LAST_PHONE_CHAN(	-0.00195343	-0.042107293	-0.059988	-0.05472146	-0.062425818	-0.033375728	0.058922	0.02512854	0.028237919	-0.042627338	0.042627338	-0.0712082	0.0712082		-0.026715756	-0.039235361	-0.024606035	-0.05433419	-0.017489
FLAG_DOCUMENT_2	0.010236993	-0.001587694	0.0095304	0.00440173	0.011113089	-0.004609052	0.002782	-0.002241583	-0.0021512	0.003140909	-0.003140909	0.004144916	-0.004144916	2.14565E-05	0.002196658	-0.002288474	0.00021154	-0.0029034	-0.001131
FLAG_DOCUMENT_3	0.049128319	-0.041980928	0.0824998	0.08019594	0.066060839	-0.076658211	0.085118	-0.203439232	-0.204622305	-0.021434684	0.021434684	-0.041783105	0.041783105	0.008647165	0.204113819	0.043983421	0.002682763	-0.00169829	0.0117738
FLAG_DOCUMENT_4	-0.00462454	0.00132534	0.0024381	0.00432767	0.002744353	0.003464077	0.000954	-0.00573673	-0.005883635	0.003269566	-0.003269566	0.002129921	-0.002129921	5.67721E-05	0.005812186	-0.006055122	0.000559717	9.95103E-05	-0.002992
FLAG_DOCUMENT_5	-0.01549775	0.006927079	-0.006273	0.00021849	-0.000781063	0.01344654	0.014754	-0.016048773	-0.01606669	0.002844175	-0.002844175	-0.000514759	0.000514759	0.000513129	0.016070488	0.033153552	0.005058944	0.069988348	0.0018843
FLAG_DOCUMENT_6	-0.13515631	-0.096593294	-0.045639	-0.06290089	-0.048732083	0.007735551	-0.348828	0.512401011	0.511972661	0.110531495	-0.110531495	0.154497545	-0.154497545	0.001253043	-0.512388591	-0.119712879	0.010485998	0.007492679	-0.039818
FLAG_DOCUMENT_7	-0.00800981	0.012851823	-0.001813	0.00433844	-0.003262052	0.005382696	-0.002113	-0.002622224	-0.002578959	0.003572998	-0.003572998	-0.00195661	0.00195661	6.43749E-05	0.002602054	-0.002977557	0.000634673	0.015309339	-0.003393
FLAG_DOCUMENT_8	0.04560631	0.162784793	0.0701294	0.11336451	0.067656948	0.07683462	0.103324	-0.10875653	-0.107977805	-0.04169034	0.04169034	-0.021593403	0.021593403	-0.01801255	0.108395871	0.024147512	-0.019485689	0.001494919	0.0229153
FLAG_DOCUMENT_9	-0.00419126	0.045079702	0.0227511	0.03147858	0.022435639	0.030339438	0.015844	-0.020796352	-0.020719681	-0.005501712	0.005501712	-0.024387866	0.024387866	0.000244034	0.020764138	0.009966263	-0.015516769	0.011447885	0.0154958
FLAG_DOCUMENT_10	-0.00266971	0.000500899	-0.004561	-0.00535188	-0.004219783	-0.003327243	-0.006723	0.009768019	0.00976805	0.002840214	-0.002840214	0.005427208	-0.005427208	2.14565E-05	-0.009767775	-0.002288474	0.00021154	-0.0029034	-0.001131
FLAG_DOCUMENT_11	-0.00121549	0.011552283	0.045415	0.01455812	0.052234485	0.018617974	0.0399	-0.028022041	-0.027284932	-0.017889489	0.017889489	-0.01690657	0.01690657	0.000270159	0.027658192	0.116192981	0.002663502	0.002817225	-0.011036
FLAG_DOCUMENT_13	0.007272938	0.054019622	0.052583	0.0255413	0.054262868	0.023758524	0.022883	-0.02330991	-0.023259779	-0.01828405	0.01828405	-0.006229624	0.006229624	0.000247799	0.023291211	0.002943622	-0.085816403	0.004009656	-0.000842
FLAG_DOCUMENT_14	-0.00625598	0.045312304	0.038947	0.03200317	0.038124876	0.03283893	0.027537	-0.023989201	-0.02365482	-0.00901864	0.00901864	-0.003891182	0.003891182	0.000243084	0.023827323	-0.006310854	-0.05158059	0.006282953	0.0067596
FLAG_DOCUMENT_15	0.011140005	0.026656722	0.0271768	0.00694247	0.028130876	0.012041649	0.007012	-0.01044196	-0.010182585	-0.0076489	0.0076489	-5.51854E-05	5.51854E-05	0.000121416	0.010314574	0.005614186	0.001197047	0.003593302	-0.0064
FLAG_DOCUMENT_16	0.015117389	0.019022588	0.0579381	0.0071392	0.054784636	0.004318508	0.024596	-0.039230903	-0.039098804	-0.028002052	0.028002052	-0.005615189	0.005615189	0.000413237	0.039175316	0.000488801	-0.027842392	0.007650621	-0.016521
FLAG_DOCUMENT_17	-0.0010715	0.002921078	-0.002007	-0.00617179	-0.002059824	0.001889593	0.012942	-0.007271852	-0.007296742	-0.003905172	0.003905172	-0.011496951	0.011496951	7.11707E-05	0.007286273	-0.000556195	-0.02994769	-0.00342262	-0.003751
FLAG_DOCUMENT_18	-0.00617536	-0.001233951	0.0282974	-0.01085074	0.027373103	0.009911959	0.041828	-0.034372417	-0.034236376	-0.023031248	0.023031248	-0.01234502	0.01234502	-0.05624345	0.034313959	0.033945456	-0.042281433	-0.00091336	-0.004172
FLAG_DOCUMENT_19	-0.00071171	0.004033153	0.0308215	0.01248233	0.030705665	-0.002766874	0.002704	-0.01151203	-0.01173442	-0.000767845	0.000767845	0.000594565	-0.000594565	0.00011357	0.011626992	0.01434697	0.001119687	0.01382006	-0.005986
FLAG_DOCUMENT_20	0.002363446	-0.003643743	0.0178193	0.00404546	0.017149263	-0.004383985	0.006339	-0.006511051	-0.00648597	-0.006558491	0.006558491	-0.003797452	0.003797452	9.10487E-05	0.006500774	-0.004211308	0.00089765	-0.00746702	0.0046793
FLAG_DOCUMENT_21	-0.00261256	-0.005881788	-0.018504	-0.02176466	-0.016849548	0.000721082	0.028133	-0.009014935	-0.008559656	-0.006269354	0.006269354	-0.014296046	0.014296046	8.58397E-05	0.008788048	0.002510868	0.000846295	0.003827309	-0.004525
AMT_REQ_CREDIT_BUREAU_	0.003683324	0.002118486	-0.002724	0.00616941	-0.002225322	-0.001898452	-0.00177	-0.00082449	-0.000759535	0.006990599	-0.006990599	-0.000888338	0.000888338	0.000302527	0.000799053	-0.008124629	-0.003592227	-0.00764963	0.0035533
AMT_REQ_CREDIT_BUREAU_	0.001320859	-0.000855967	0.0094669	0.00405119	0.009186309	-0.004152999	0.000875	0.002456473	0.002635247	0.006874263	-0.006874263	-8.25216E-05	8.25216E-05	0.000250348	-0.002540538	-0.004850269	0.002468185	0.000281358	0.0018646

We used **CORREL()** of EXCEL to calculate the correlation coefficients and then ranked them .

TOP 10 Correlations for Repayers :

- OBS_30_CNT_SOCIAL_CIRCLE-OBS_60_CNT_SOCIAL_CIRCLE
- AMT_GOODS_PRICE-AMT_CREDIT
- AMT_APPLICATION-AMT_CREDIT.1
- REGION_RATING_CLIENT_W_CITY-REGION_RATING_CLIENT
- CNT_FAM_MEMBERS-CNT_CHILDREN
- LIVE_REGION_NOT_WORK_REGION-REG_REGION_NOT_WORK_REGION
- DEF_60_CNT_SOCIAL_CIRCLE-DEF_30_CNT_SOCIAL_CIRCLE
- REG_CITY_NOT_WORK_CITY-LIVE_CITY_NOT_WORK_CITY
- AMT_ANNUITY.1-AMT_APPLICATION
- AMT_GOODS_PRICE-AMT_ANNUITY

E. Identify Top Correlations for Different Scenarios :



TOP 10 Correlations for Defaulters :

- FLAG_MOBIL-DAYS_EMPLOYED
- OBS_60_CNT_SOCIAL_CIRCLE-FLAG_DOCUMENT_6
- AMT_GOODS_PRICE-AMT_ANNUITY
- REGION_RATING_CLIENT_W_CITY-REG_CITY_NOT_WORK_CITY
- DEF_60_CNT_SOCIAL_CIRCLE-FLAG_DOCUMENT_7
- CNT_FAM_MEMBERS-AMT_INCOME_TOTAL
- LIVE_REGION_NOT_WORK_REGION-DEF_30_CNT_SOCIAL_CIRCLE
- LIVE_CITY_NOT_WORK_CITY-DAYS_LAST_PHONE_CHANGE
- AMT_ANNUITY-REGION_POPULATION_RELATIVE
- AMT_CREDIT-AMT_GOODS_PRICE

KEY INSIGHTS

1. **NAME_TYPE_SUITE:** people who are **unaccompanied** default a lot.
2. **CODE_GENDER: Females** default more than men.
3. **NAME_INCOME_TYPE:** Most of the defaulters are **working** professionals.
4. **NAME_EDUCATION_TYPE:** Most customers who default have a education till **secondary, secondary/special.**
5. **NAME_HOUSING_TYPE:** people living in **house/apartment** default a lot.
6. **NAME_FAMILY_STATUS: Married** people default more.
7. **OCCUPATION_TYPE: Laborers** are mostly the defaulters.
8. **ORGANIZATION_TYPE: Business Entity Type 3** is the type of organization with most defaulters.
9. **AMT_INCOME_TOTAL:** Customers who default mostly earns between **100000-200000.**
10. **AMT_CREDIT:** Customers who gets approved credit limit between **200000-300000** are more likely to default.
11. **NAME_CONTRACT_TYPE: Cash loans** gets repaid more than revolving loans

SUGGESTIONS

1. Since, customers who has the income range between **100000-200000** are more likely to default so, we can **decrease** the credit amount of loan taken by them to reduce risk.
2. Organizations which are **business entity type 3** are more likely to default, so, we can have **stricter** policies for them .
3. We see a increase in repayment with years of experience and age, so, we can **prioritize** senior citizens and experienced people.
4. Since customers who gets approved limit of **200000-300000** are more likely to default, we can **increase** the rate of interest on such loans .
5. Strategies should be made to offer more **cash loans** rather than revolving loans as there are higher chances of them getting repaid.

Result :

While completing the project, I learned how to work with huge datasets. It helped me understand how to merge two datasets and analyze it. The project required extensive use of EXCEL, its formulas and functions, also, discovered how to use add-ins like Data Analysis. In the dataset provided, there were huge amount of missing data, null values, outliers and the project helped me to understand how to deal with them and impute the missing values.

Thank you..!!