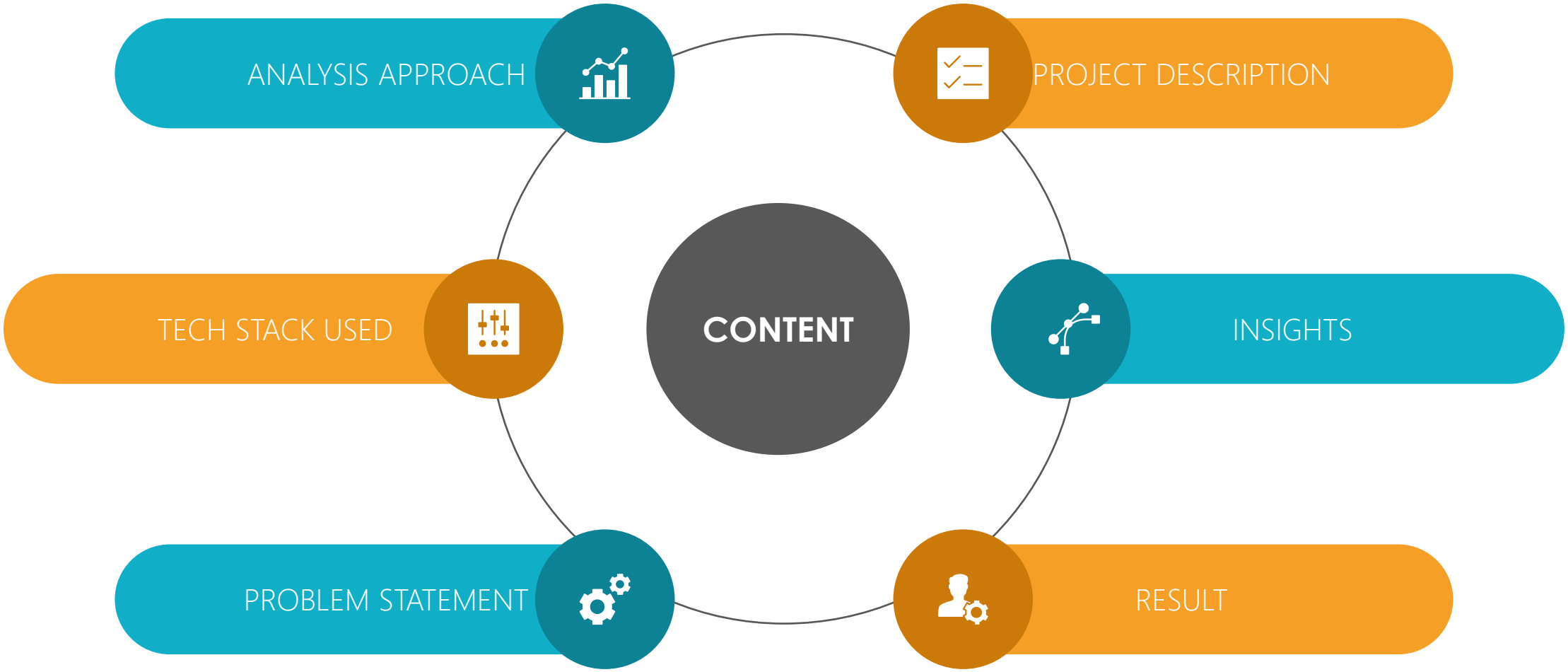
Bank Loan

Case Study

By Anjali Gupta


ANALYSIS APPROACH
PROJECT DESCRIPTION
TECH STACK USED
CONTENT
INSIGHTS
PROBLEM STATEMENT
RESULT

PROJECT DESCRIPTION

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

TECH STACK USED

To do this data analysis the tech stack used include :-

- Microsoft Excel
- Microsoft PowerPoint presentation

ANALYSIS APPROACH

We will use exploratory data analytics(EDA) to explore the data and observe the pattern correlation among the data. It is an approach to analysing the data using visual techniques. It is used to discover trends, and patterns, or to check assumptions with the help of statistical summaries and graphical representations.

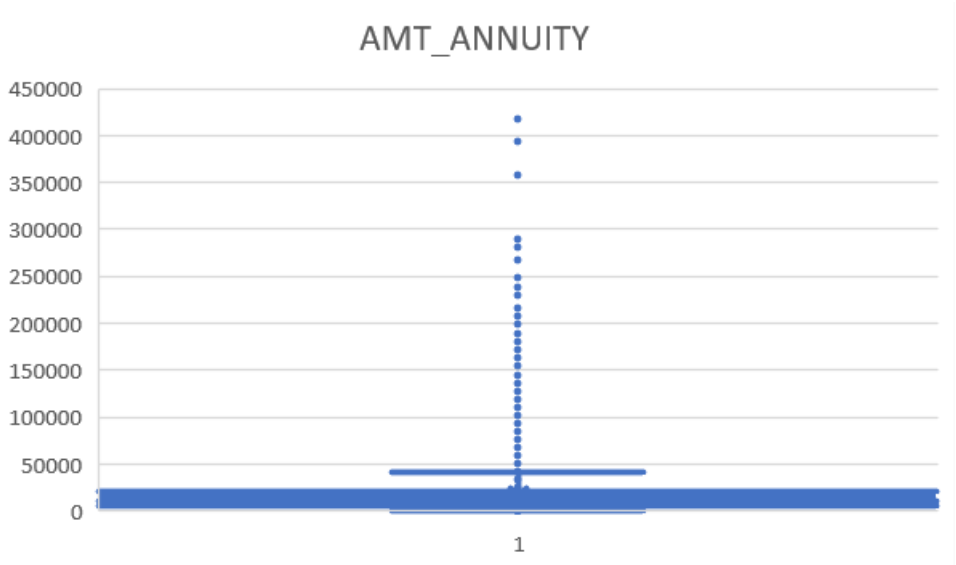
Problem Statement and Insight

❑DEALING WITH MISSING VALUES

We can deal with missing values by replacing the values with appropriate values for example we can replace the blank spaces in columns having numerical value by the mean of the column
We can replace columns having categorical values by taking the mode of the given values.
But for my dataset, I am not replacing missing values as I want the result to be accurate.

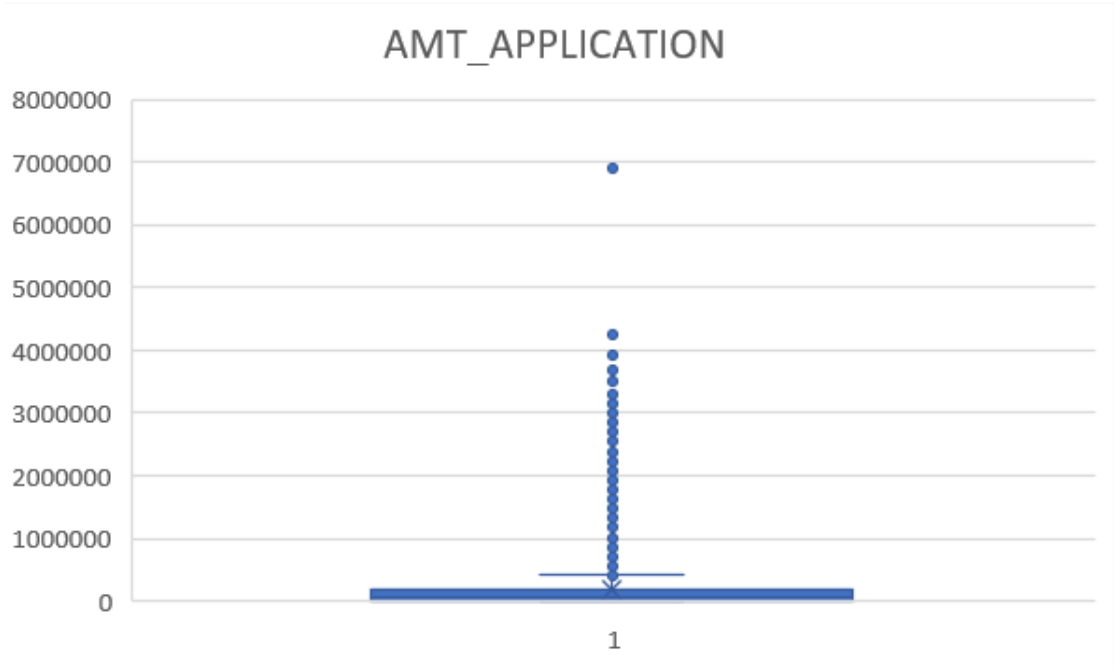
Outliers in specific columns

Outliers



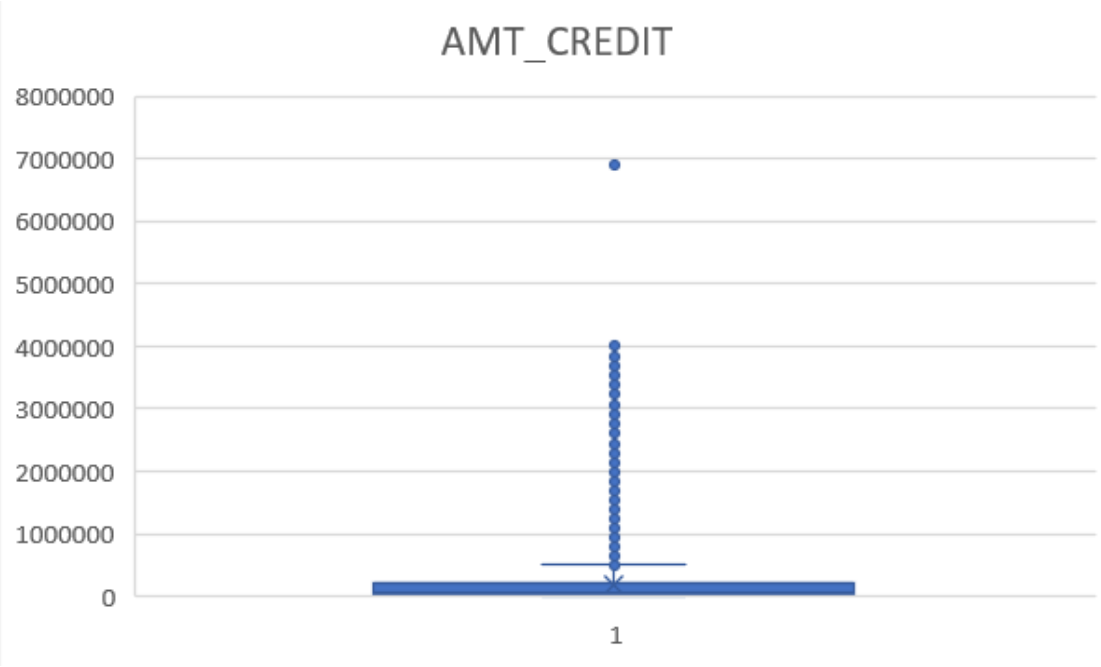
Some outliers are there in amt_annuity
As can be seen from box and whisker plot
The first quartile is very slim for amt_annuity which means that most of the data values lies in third quartile only

Outliers



There are outlier in AMT_APPLICATION also as can be seen from box and whisker plot
The first quartile is very slim as compared to third quartile therefore we can say that most of the data values lies in third quartile only

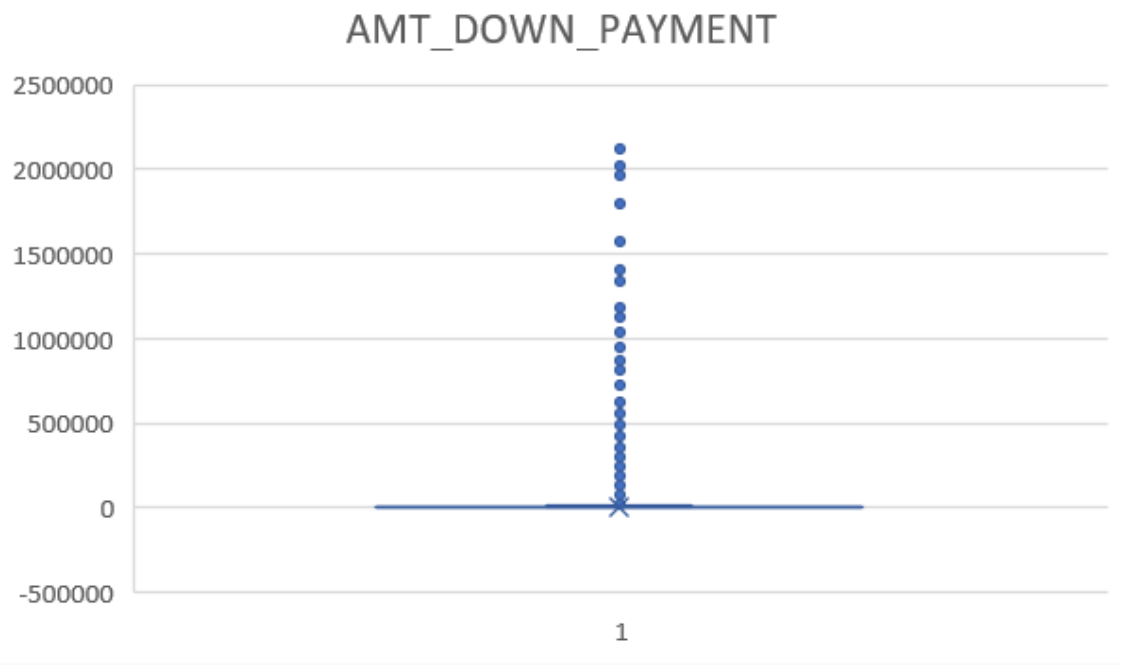
Outliers



There are outlier in AMT_CREDIT also as can be seen from box and whisker plot

The third quartile is larger than the first quartile so we can say that many data values are there in third quartile

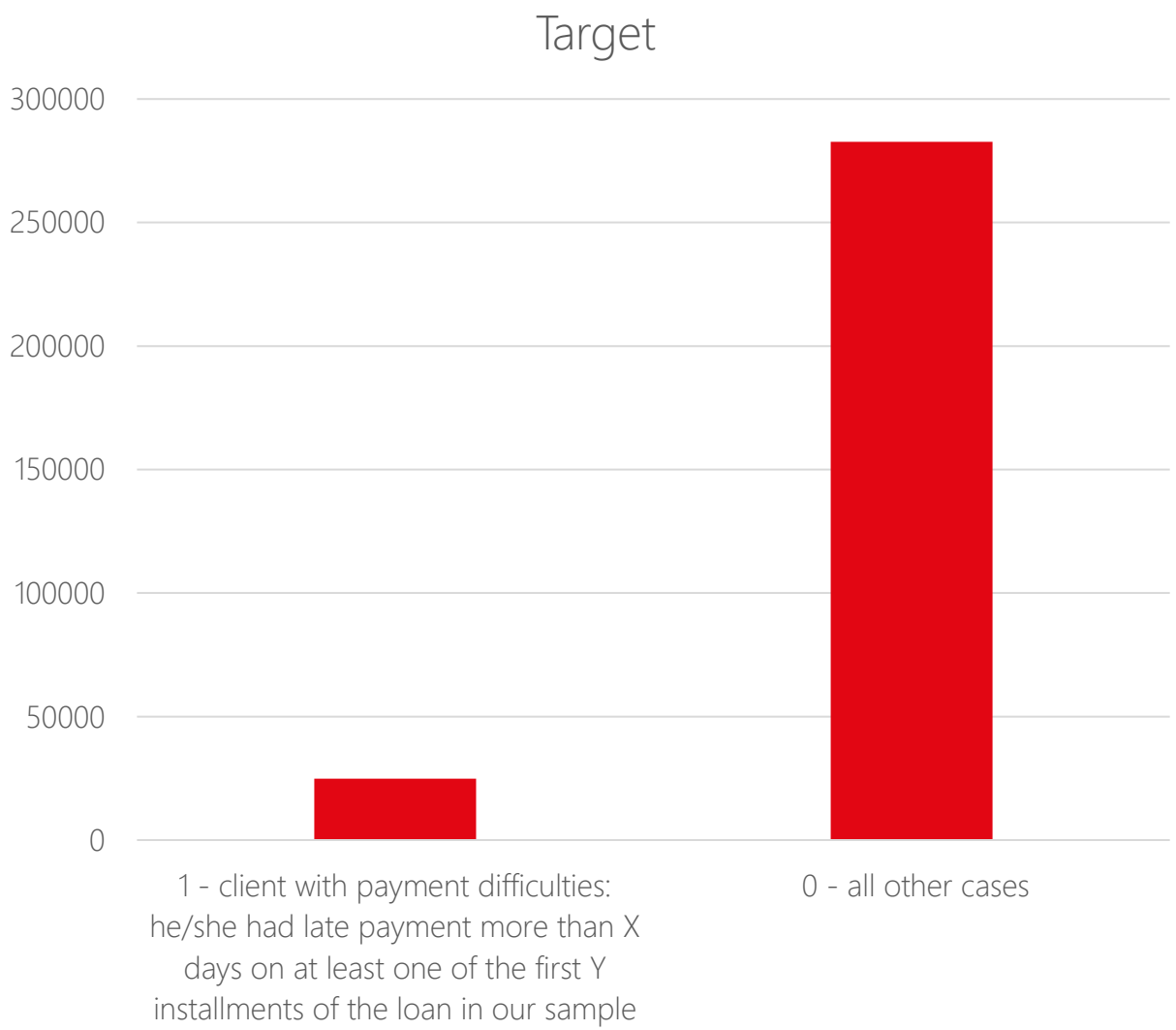
Outliers



As it can be seen that there are outliers in AMT_DOWN_PAYMENT but they are much cloaser gto each other

Data Imbalance

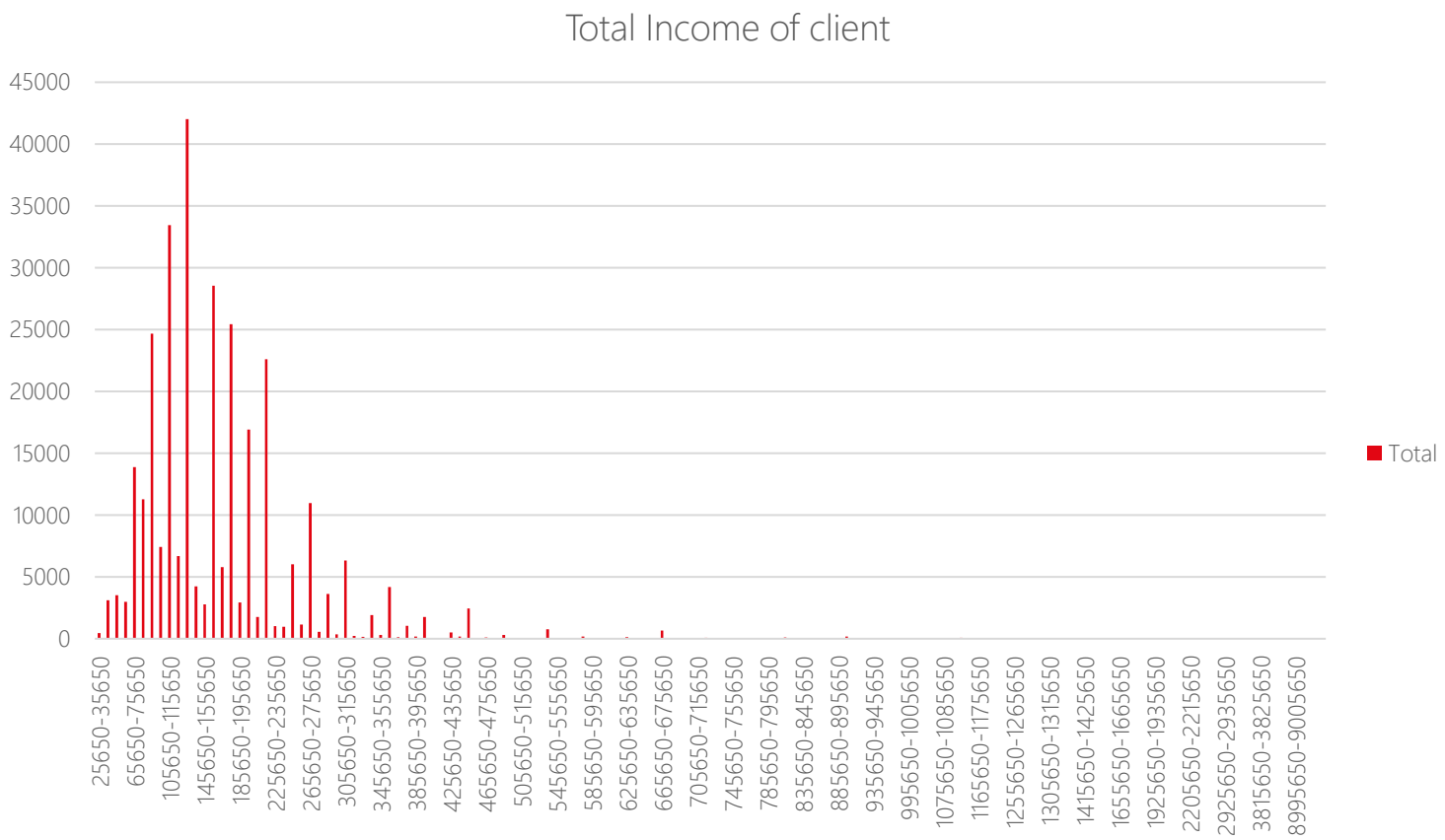
Data Imbalance



Clients with payment difficulties are much less than the one with no payment difficulties.
The percentage of people with no payment difficulty is 8.0729%
The percentage of people with payment difficulty is 91.9271%
Or we can say that the imbalance ratio is 8.781829%

Univariate Analysis

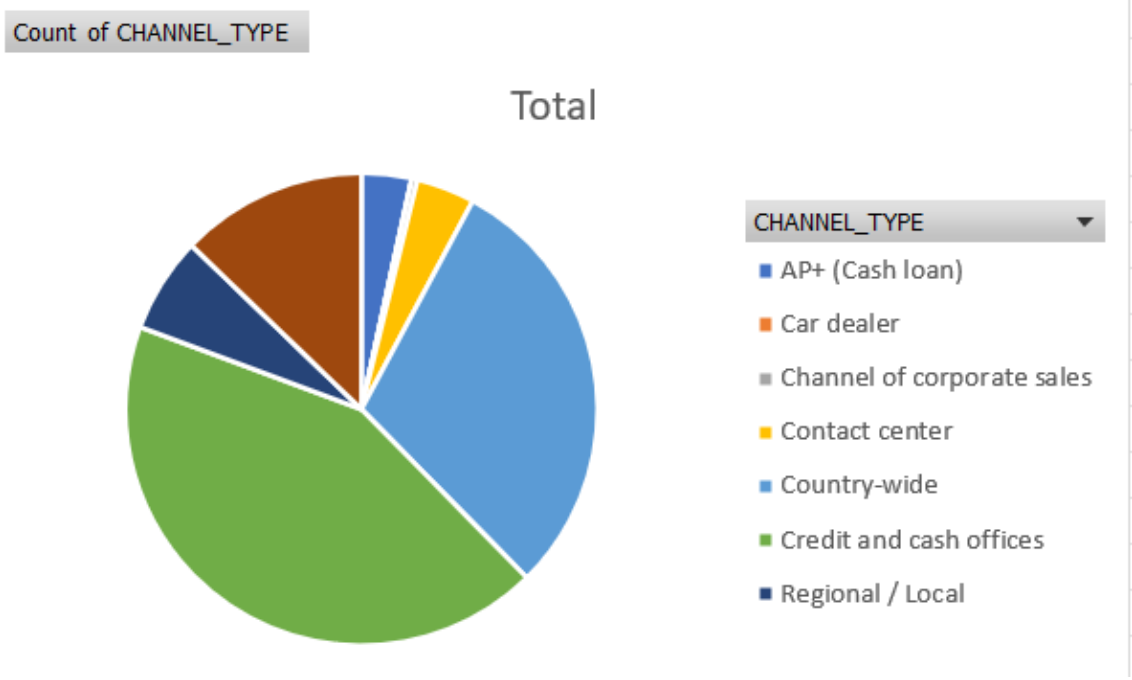
Univariate Analysis



Maximum people from income range 125650-135650 apply for loan

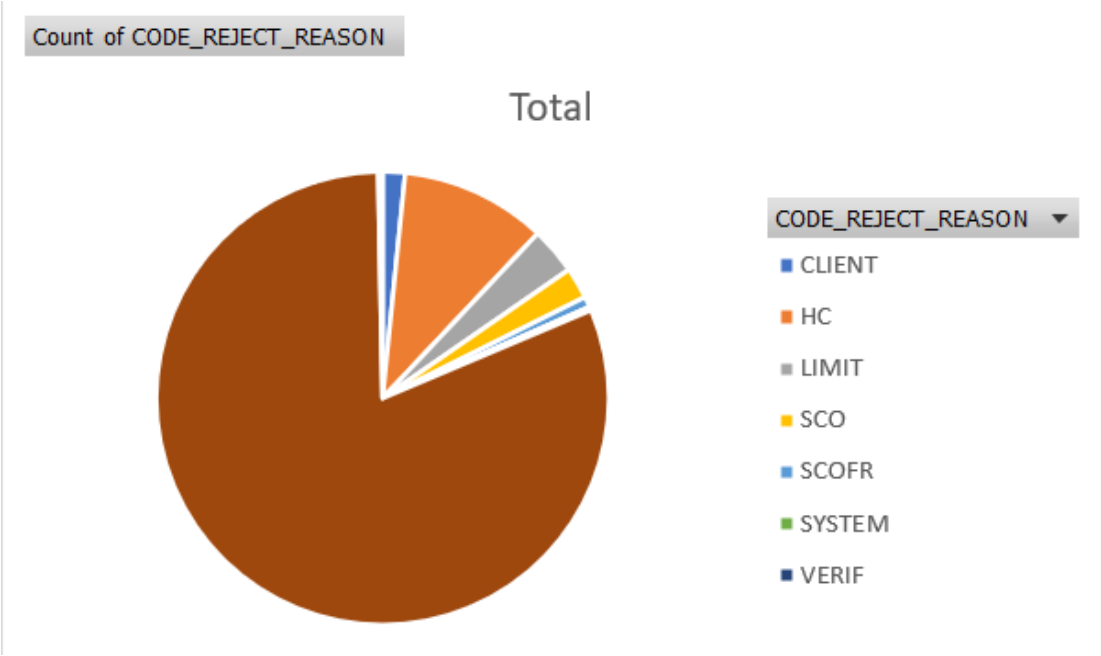
As peoples income increases they apply less for loan

Univariate Analysis



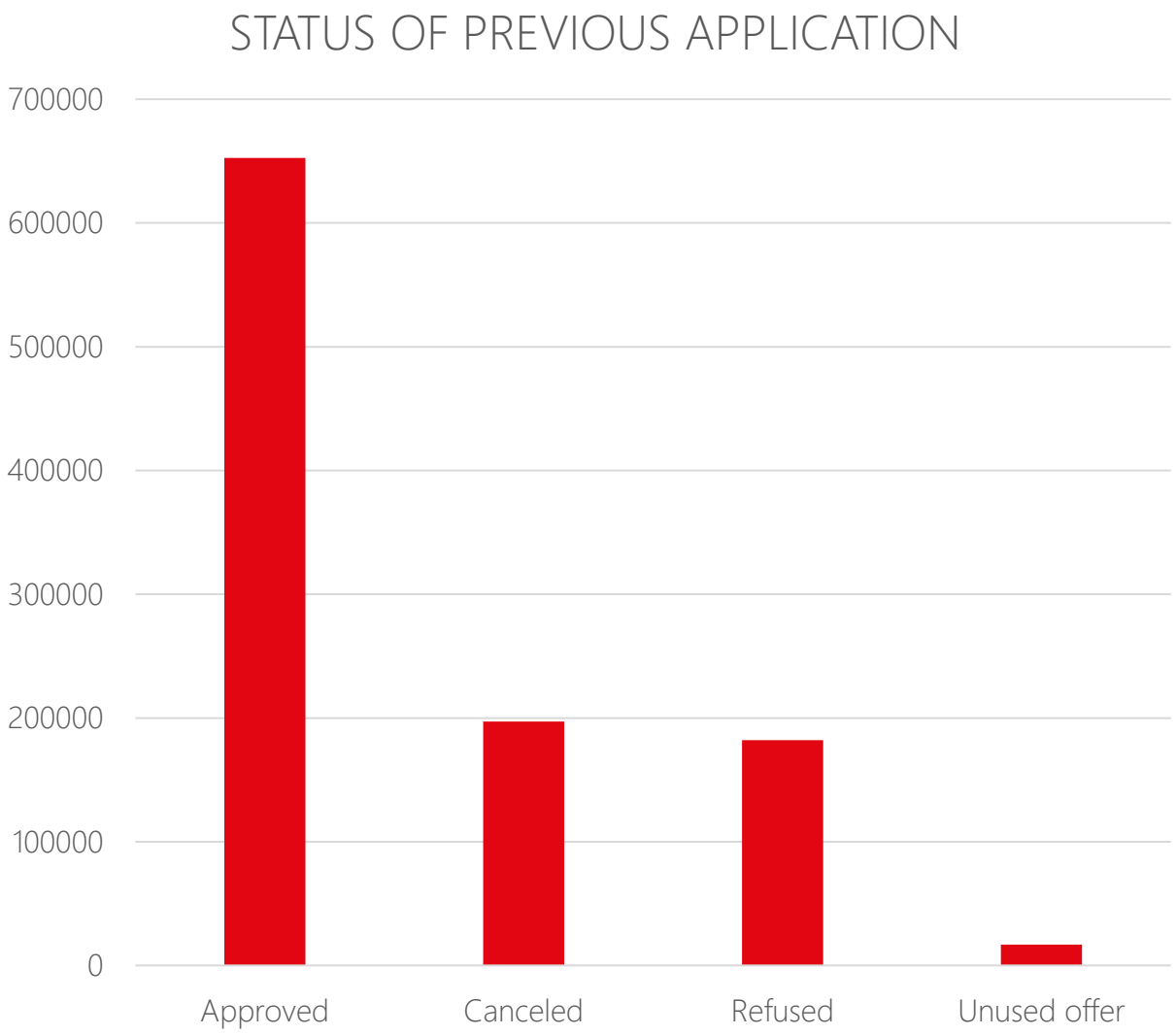
Maximum number of peoples channel type was Credit and cash offices

Univariate Analysis



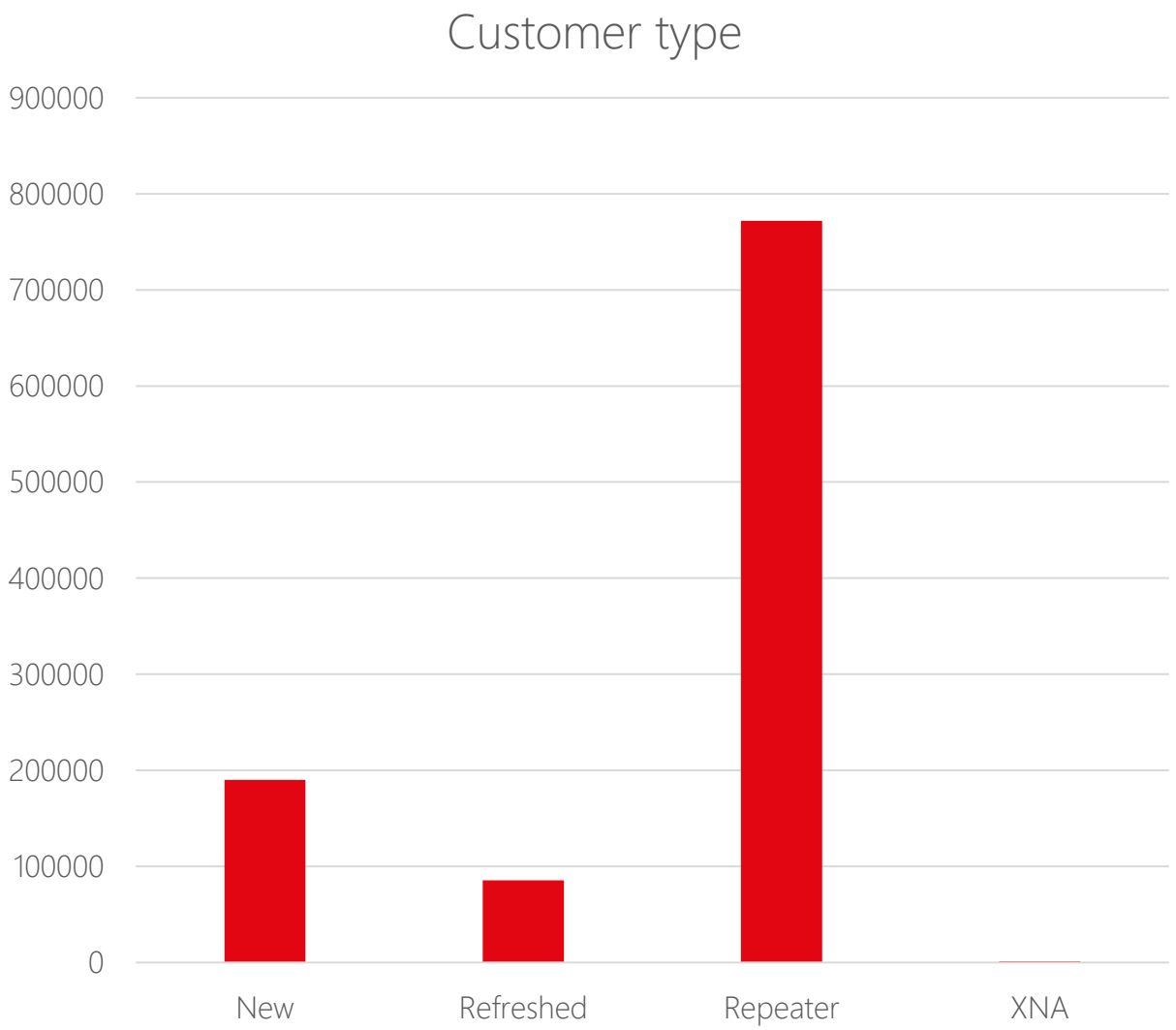
Highest amount of loans rejection reason was XAP followed by HC

Univariate Analysis



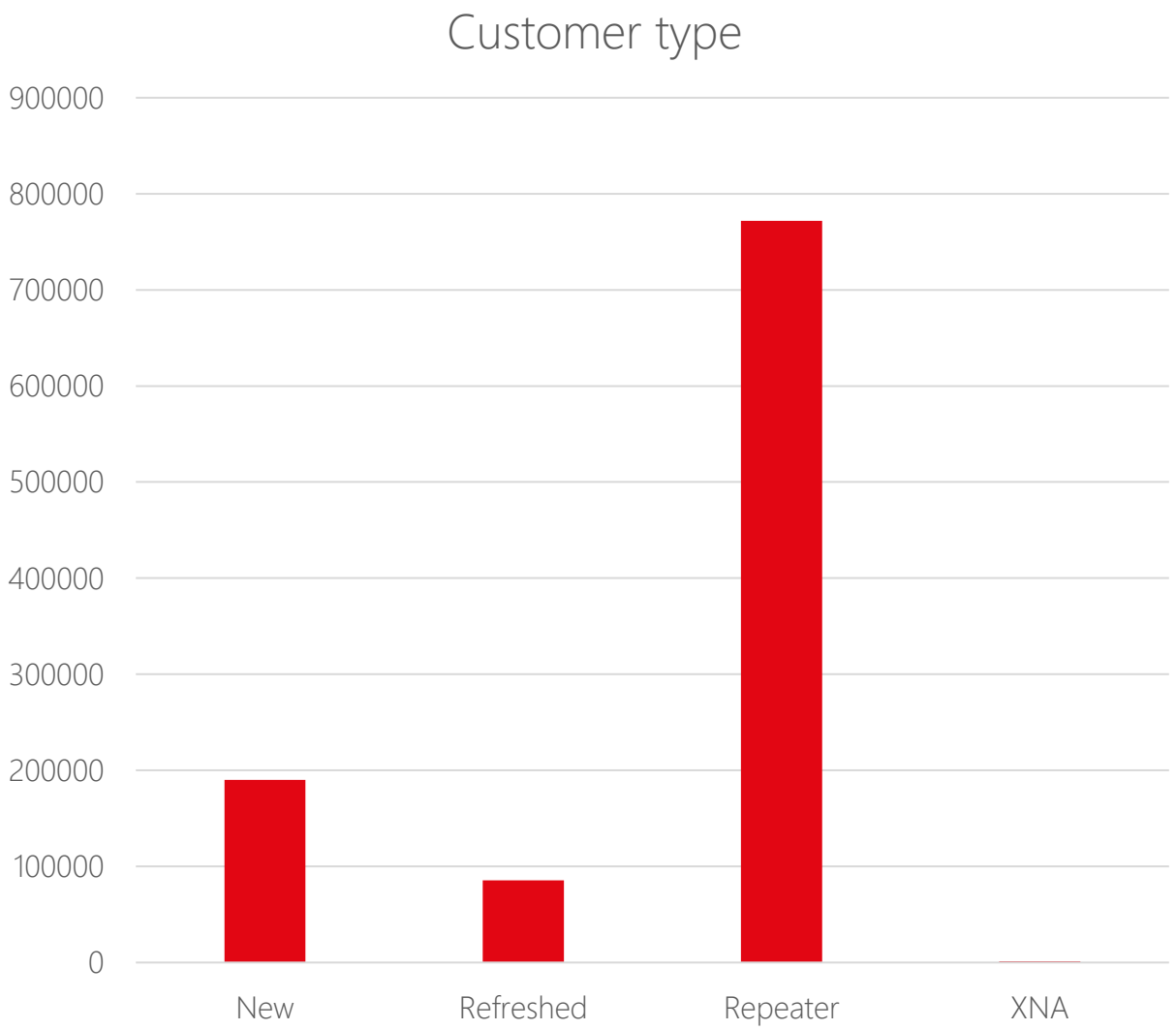
It can be seen that most of the previous applications were approved by the bank and less amount of applications were refused by the bank

Univariate Analysis



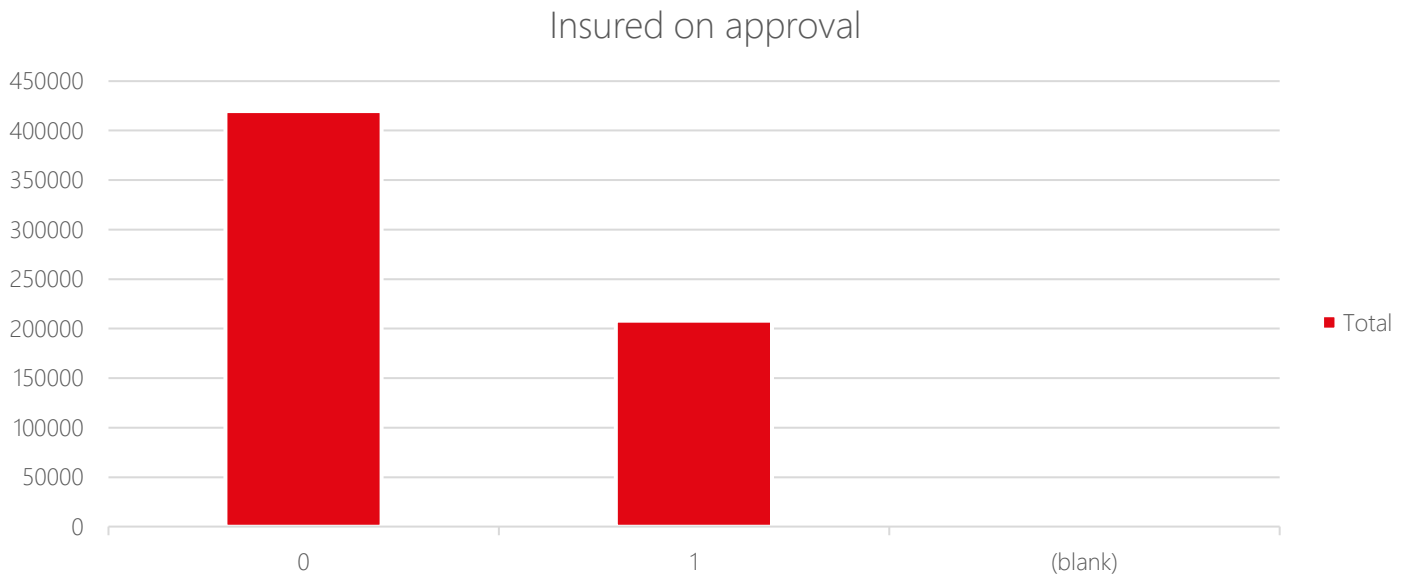
A large percentage of people were repeater whereas a comparatively less amount of new applications were there

Univariate Analysis



A large percentage of people were repeater whereas a comparatively less amount of new applications were there

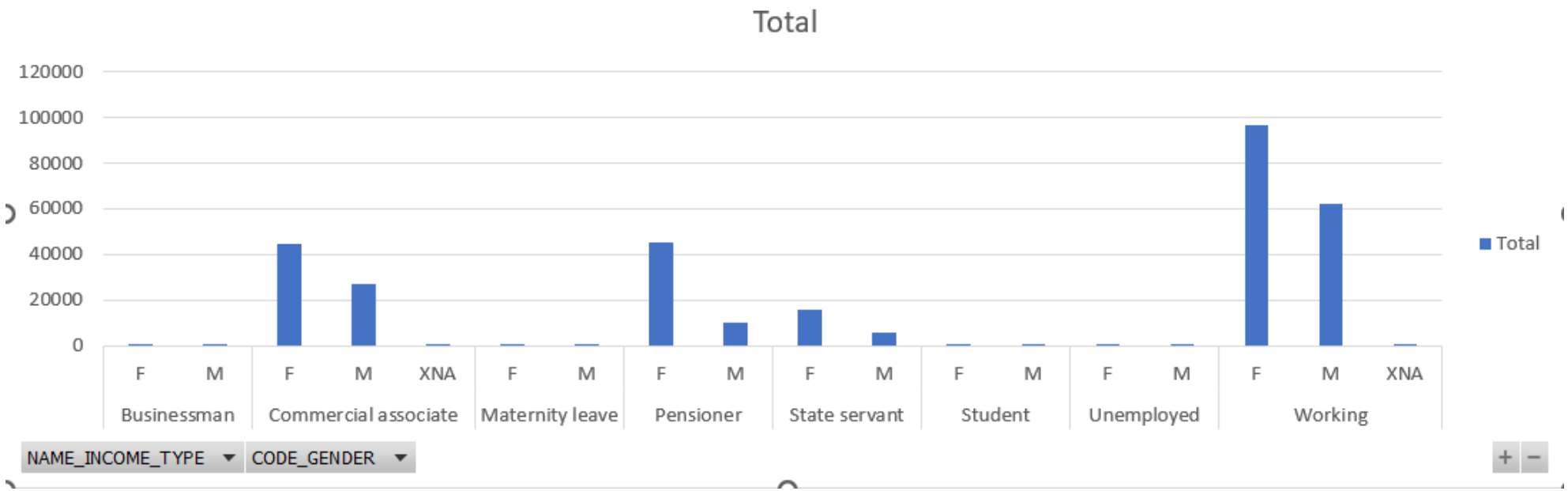
Univariate Analysis



Many clients did not request insurance during previous application

Bivariate Analysis

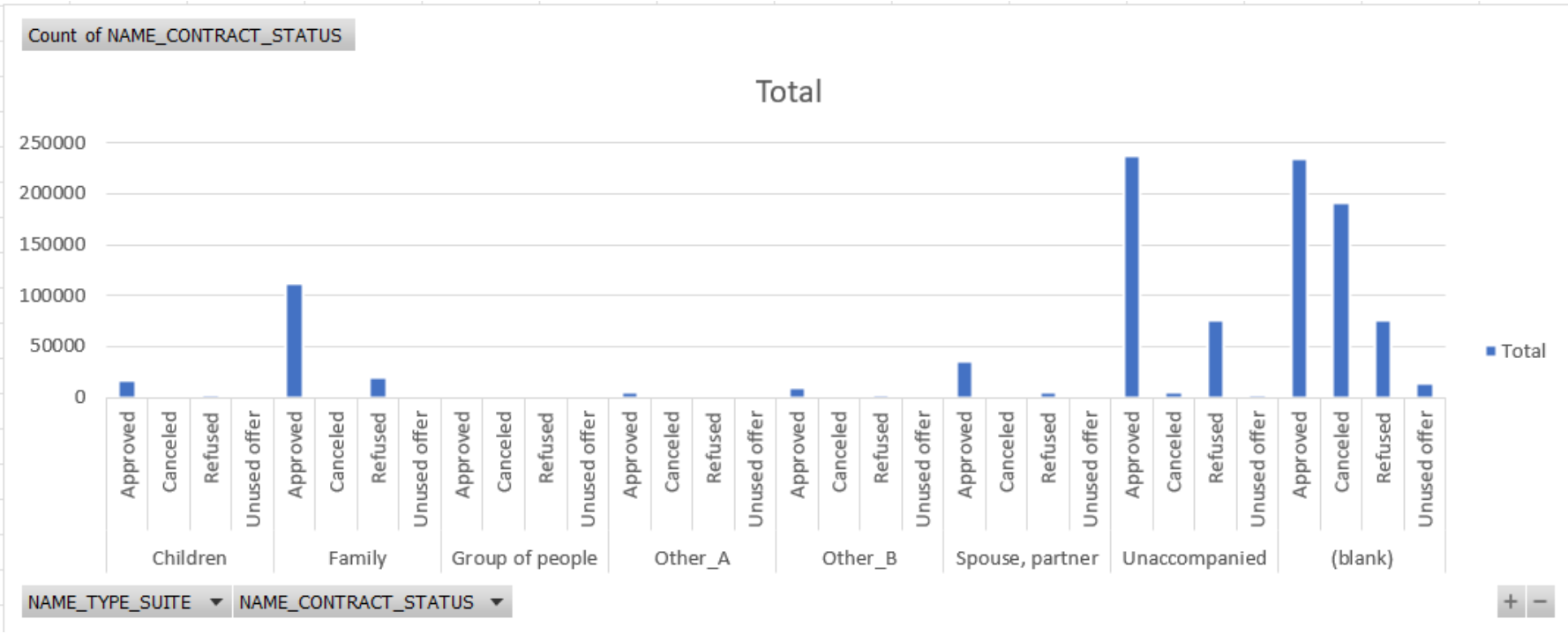
Bivariate Analysis



Maximum number of working womens gave application for loan.

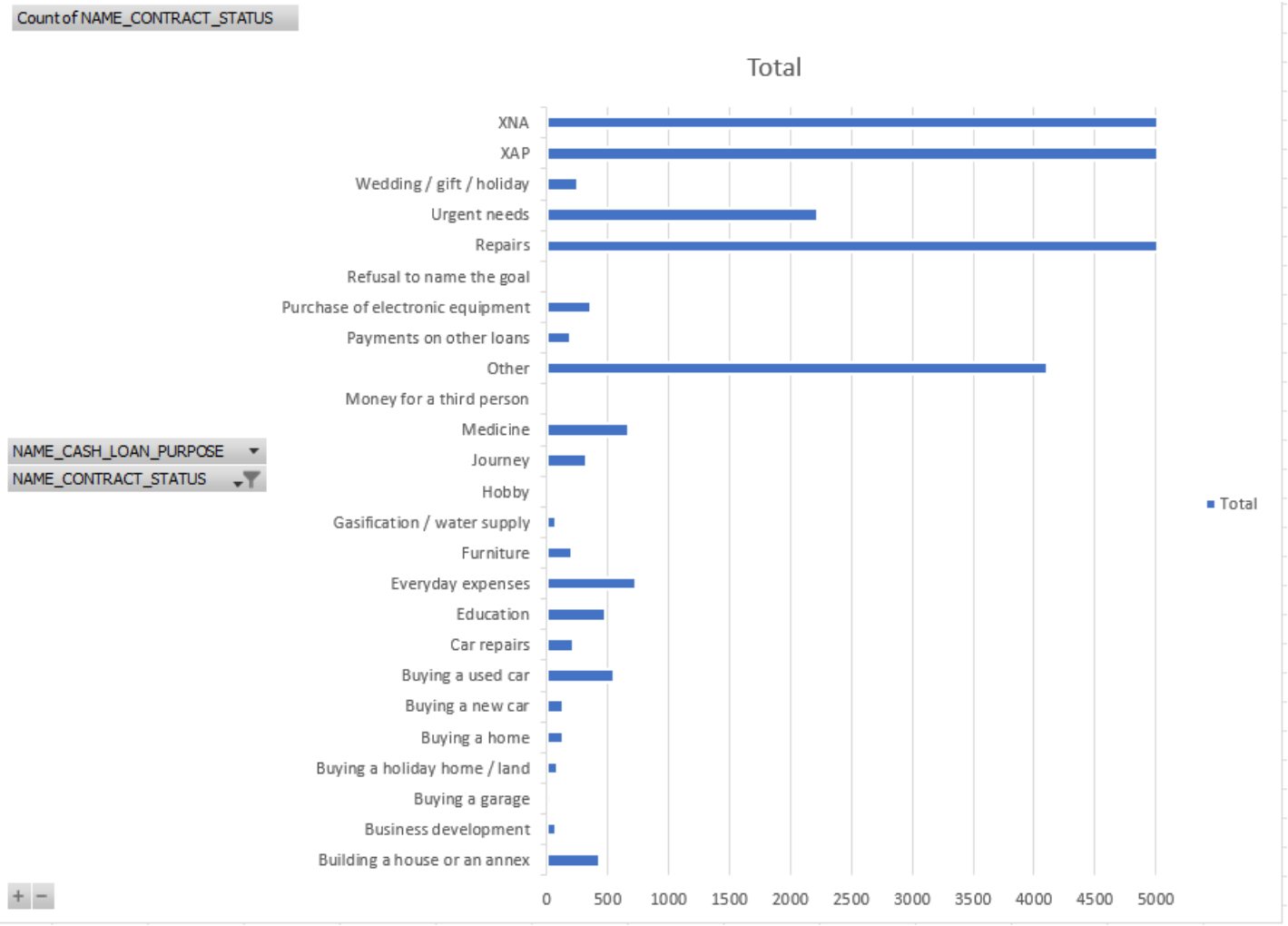
More number of female applied for loan as compared to men

Bivariate Analysis



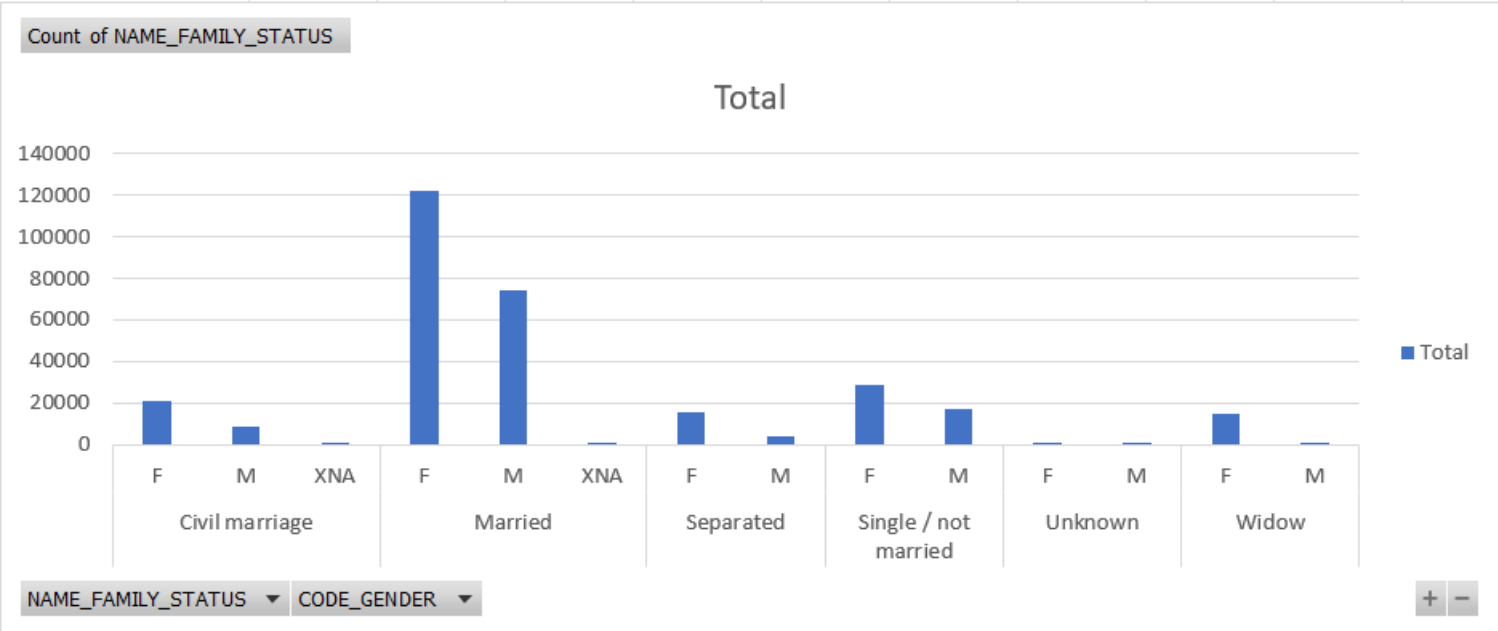
Maximum number of unaccompanied peoples loan was approved, followed by people with families

Bivariate Analysis



Maximum people who got their loan approved gave the reason for loan as XAP followed by XNA and repairs

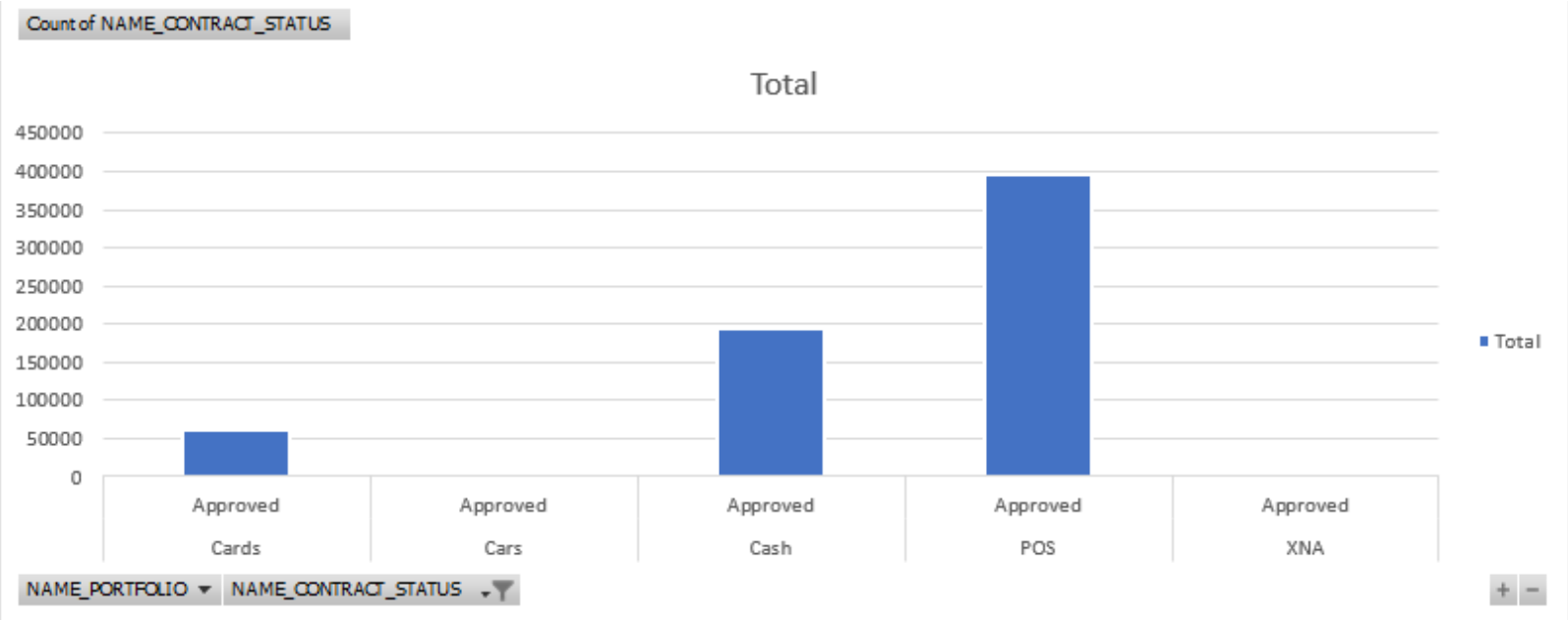
Bivariate Analysis



Maximum females with family status married applied for a bank loan

From the graph its clearly visible that more number of females applied for loan as compared to males

Bivariate Analysis



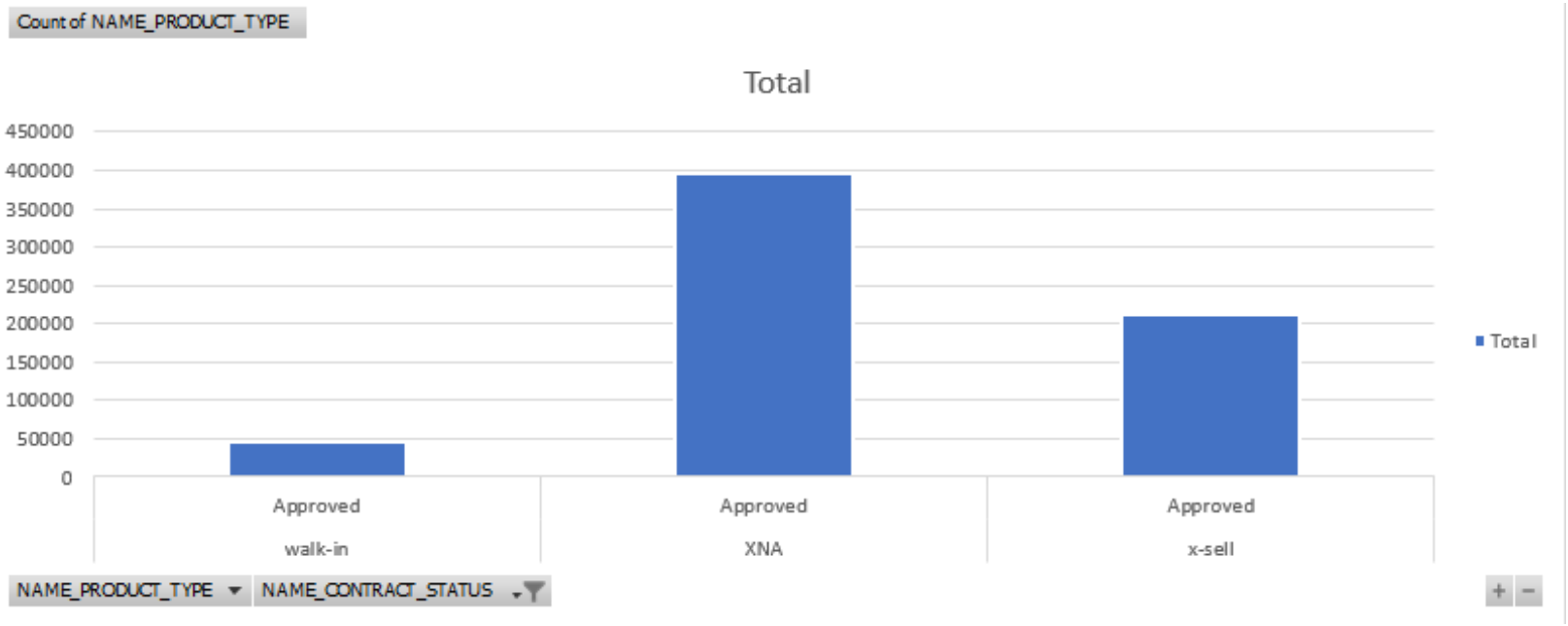
The maximum number of people whose previous application was for POS was approved

Bivariate Analysis

Most of the peoples application who choose cash through the bank as payment method for the previous application were approved

Most of the people who choose XNA as payment method were the one who cancelled their application

Bivariate Analysis



Most of the people whose previous application was x-self were approved

Target 0 Correlation

	CNT_CHILDRE	INCOME_T	MT_CREDI	NT_ANNUI	_GOODS_P	OPULATION	AYS_BIRTH	S_EMPLO	_REGISTRA	S_ID_PUBL	VN_CAR_A	FAM_MEMI	_RATING_	TING_CLIE	PR_PROCESS_START
CNT_CHILD	1														
AMT_INCO	0.027397	1													
AMT_CRED	0.003081	0.342799	1												
AMT_ANNU	0.020905	0.418953	0.771309	1											
AMT_GOOD	-0.00052	0.349462	0.98725	0.776686	1										
REGION_PO	-0.02436	0.167851	0.100604	0.120988	0.103827	1									
DAYS_BIRT	0.336966	0.062609	-0.04738	0.012263	-0.04456	-0.02524	1								
DAYS_EMP	-0.24336	-0.14125	-0.07251	-0.10642	-0.07105	-0.00696	-0.61805	1							
DAYS_REGI	0.185792	0.064937	0.013477	0.039436	0.015916	-0.05208	0.333151	-0.21019	1						
DAYS_ID_P	-0.02875	0.022896	-0.00146	0.014113	-0.00365	-0.00107	0.271314	-0.27429	0.100236	1					
OWN_CAR_	0.008381	-0.11636	-0.09496	-0.09772	-0.10403	-0.08215	0.004055	0.030451	-0.02762	0.006572	1				
CNT_FAM_	0.878571	0.034256	0.064536	0.075787	0.062814	-0.02342	0.285825	-0.23741	0.17563	-0.02046	-0.01397	1			
REGION_RA	0.022842	-0.18657	-0.10334	-0.13213	-0.10438	-0.539	0.002332	0.037851	0.075846	-0.009	0.087092	0.027872	1		
REGION_RA	0.021866	-0.20047	-0.11224	-0.14515	-0.1126	-0.5373	0.00078	0.040028	0.069357	-0.01173	0.088295	0.028657	0.950149	1	
HOUR_APP	-0.00524	0.076743	0.053619	0.053589	0.062766	0.172814	0.095916	-0.09484	-0.00804	0.03398	-0.07109	-0.01057	-0.28406	-0.26287	1

	CNT_CHILDRE	INCOME_T	MT_CREDI	MT_ANNUI	_GOODS_P	PULATION	AYS_BIRTH	S_EMPLO	_REGISTRA	S_ID_PUBL	VN_CAR_A	FAM_MEMI	_RATING_	TING_CLIE	PR_PROCESS_START
CNT_CHILD	1														
AMT_INCO	0.027397	1													
AMT_CRED	0.003081	0.342799	1												
AMT_ANNU	0.020905	0.418953	0.771309	1											
AMT_GOOI	-0.00052	0.349462	0.98725	0.776686	1										
REGION_P	-0.02436	0.167851	0.100604	0.120988	0.103827	1									
DAYS_BIRT	0.336966	0.062609	-0.04738	0.012263	-0.04456	-0.02524	1								
DAYS_EMP	-0.24336	-0.14125	-0.07251	-0.10642	-0.07105	-0.00696	-0.61805	1							
DAYS_REGI	0.185792	0.064937	0.013477	0.039436	0.015916	-0.05208	0.333151	-0.21019	1						
DAYS_ID_P	-0.02875	0.022896	-0.00146	0.014113	-0.00365	-0.00107	0.271314	-0.27429	0.100236	1					
OWN_CAR_	0.008381	-0.11636	-0.09496	-0.09772	-0.10403	-0.08215	0.004055	0.030451	-0.02762	0.006572	1				
CNT_FAM_	0.878571	0.034256	0.064536	0.075787	0.062814	-0.02342	0.285825	-0.23741	0.17563	-0.02046	-0.01397	1			
REGION_RA	0.022842	-0.18657	-0.10334	-0.13213	-0.10438	-0.539	0.002332	0.037851	0.075846	-0.009	0.087092	0.027872	1		
REGION_RA	0.021866	-0.20047	-0.11224	-0.14515	-0.1126	-0.5373	0.00078	0.040028	0.069357	-0.01173	0.088295	0.028657	0.950149	1	
HOUR_APP	-0.00524	0.076743	0.053619	0.053589	0.062766	0.172814	0.095916	-0.09484	-0.00804	0.03398	-0.07109	-0.01057	-0.28406	-0.26287	1

The pink-shaded cells represent the top 10 correlations

Top 10 Correlation

1. AMT_CREDIT and AMT_GOODS_PRICE
2. REGION_RATING_CLIENT and REGION_RATING_CLIENT_W_CITY
3. CNT_CHILDREN and CNT_FAM_MEMBERS
4. AMT_ANNUITY and AMT_GOODS_PRICE
5. AMT_CREDIT and AMT_ANNUITY
6. DAYS_BIRTH and DAYS_EMPLOYED
7. REGION_POPULATION_RELATIVE and REGION_RATING_CLIENT
8. REGION_POPULATION_RELATIVE and REGION_RATING_CLIENT_W_CITY
9. AMT_INCOME_TOTAL and AMT_ANNUITY
10. AMT_INCOME_TOTAL and AMT_GOODS_PRICE

Target 1 Correlation

	CNT_CHILDREN	T_INCOME_TO	AMT_CREDIT	AMT_ANNUITY	MT_GOODS_PRIC	OPULATION_	DAYS_BIRTH	DAYS_EMPLOYED	AYS_REGISTRATIO	YS_ID_PUBLI	OWN_CAR_AGE	T_FAM_MEMBE	ION_RATING_CLI	N_RATING_CLIENT_W	JR_APPR_PROCESS_ST
CNT_CHILDREN	1														
AMT_INCOME_TOTAL	0.004795787	1													
AMT_CREDIT	-0.00167496	0.038131435	1												
AMT_ANNUITY	0.031257119	0.046421057	0.752194735	1											
AMT_GOODS_PRICE	-0.0081117	0.037583082	0.983102519	0.752699196	1										
REGION_POPULATION_RELATIVE	-0.0319749	0.009134586	0.069161087	0.07169025	0.07604893	1									
DAYS_BIRTH	0.259108666	0.003096245	-0.135316369	-0.014303316	-0.135810334	-0.0481904	1								
DAYS_EMPLOYED	-0.19194154	-0.01497856	-0.000967744	-0.082551987	0.003586919	0.0151026	-0.575097231	1							
DAYS_REGISTRATION	0.149153857	0.000157999	-0.025854317	0.034279023	-0.025678921	-0.056222	0.289114025	-0.188928746	1						
DAYS_ID_PUBLISH	-0.0322986	-0.00421486	-0.05232898	-0.016767235	-0.056085697	-0.0155369	0.252862836	-0.226470486	0.096832619	1					
OWN_CAR_AGE	0.006231352	-0.12686074	-0.065944641	-0.081517865	-0.07806384	-0.0516302	-0.001738211	0.023442503	-0.024006707	0.0170194	1				
CNT_FAM_MEMBERS	0.885483713	0.006653677	0.05122364	0.075711476	0.04738797	-0.0301627	0.203267038	-0.1865611	0.145828292	-0.031785	-0.02735872	1			
REGION_RATING_CLIENT	0.040680482	-0.02148626	-0.059192754	-0.073783735	-0.06638988	-0.4432355	0.033927932	0.00367858	0.103855048	0.0013972	0.065625772	0.043651646	1		
REGION_RATING_CLIENT_W_CITY	0.043185374	-0.02280798	-0.071377103	-0.089290515	-0.077190843	-0.4469768	0.033631154	0.002991632	0.100285178	-0.001497	0.067554166	0.047882961	0.956637164	1	
HOUR_APPR_PROCESS_START	-0.02389902	0.013774567	0.031781954	0.03123613	0.044314656	0.1427438	0.062172285	-0.059997281	-0.032517591	0.0216609	-0.07153973	-0.027526008	-0.293908372	-0.275702529	1

	CNT_CHILDREN	IT_INCOME_TOT	AMT_CREDIT	AMT_ANNUITY	MT_GOODS_PRIC	POPULATION_	DAYS_BIRTH	DAYS_EMPLOYED	AYS_REGISTRATIO	YS_ID_PUBLI	OWN_CAR_AGE	T_FAM_MEMBE	ION_RATING_CLI	N_RATING_CLIENT_W	JR_APPR_PROCESS_START
CNT_CHILDREN	1														
AMT_INCOME_TOTAL	0.004795787	1													
AMT_CREDIT	-0.00167496	0.038131435	1												
AMT_ANNUITY	0.031257119	0.046421057	0.752194735	1											
AMT_GOODS_PRICE	-0.0081117	0.037583082	0.983102519	0.752699196	1										
REGION_POPULATION_RELATIVE	-0.0319749	0.009134586	0.069161087	0.07169025	0.07604893	1									
DAYS_BIRTH	0.259108666	0.003096245	-0.135316369	-0.014303316	-0.135810334	-0.0481904	1								
DAYS_EMPLOYED	-0.19194154	-0.01497856	-0.000967744	-0.082551987	0.003586919	0.0151026	-0.575097231	1							
DAYS_REGISTRATION	0.149153857	0.000157999	-0.025854317	0.034279023	-0.025678921	-0.056222	0.289114025	-0.188928746	1						
DAYS_ID_PUBLISH	-0.0322986	-0.00421486	-0.05232898	-0.016767235	-0.056085697	-0.0155369	0.252862836	-0.226470486	0.096832619	1					
OWN_CAR_AGE	0.006231352	-0.12686074	-0.065944641	-0.081517865	-0.07806384	-0.0516302	-0.001738211	0.023442503	-0.024006707	0.0170194	1				
CNT_FAM_MEMBERS	0.885483713	0.006653677	0.05122364	0.075711476	0.04738797	-0.0301627	0.203267038	-0.1865611	0.145828292	-0.031785	-0.02735872	1			
REGION_RATING_CLIENT	0.040680482	-0.02148626	-0.059192754	-0.073783735	-0.06638988	-0.4432355	0.033927932	0.00367858	0.103855048	0.0013972	0.065625772	0.043651646	1		
REGION_RATING_CLIENT_W_CITY	0.043185374	-0.02280798	-0.071377103	-0.089290515	-0.077190843	-0.4469768	0.033631154	0.002991632	0.100285178	-0.001497	0.067554166	0.047882961	0.956637164	1	
HOUR_APPR_PROCESS_START	-0.02389902	0.013774567	0.031781954	0.03123613	0.044314656	0.1427438	0.062172285	-0.059997281	-0.032517591	0.0216609	-0.07153973	-0.027526008	-0.293908372	-0.275702529	1

The pink-shaded cell represents the top 10 correlation

Top 10 Correlation

1. AMT_CREDIT and AMT_GOODS_PRICE
2. REGION_RATING_CLIENT and REGION_RATING_CLIENT_W_CITY
3. CNT_CHILDREN and CNT_FAM_MEMBERS
4. AMT_ANNUITY and AMT_GOODS_PRICE
5. AMT_CREDIT and AMT_ANNUITY
6. DAYS_BIRTH and DAYS_EMPLOYED
7. REGION_POPULATION_RELATIVE and REGION_RATING_CLIENT_W_CITY
8. REGION_POPULATION_RELATIVE and REGION_RATING_CLIENT
9. REGION_RATING_CLIENT and HOUR_APPR_PROCESS_START
10. DAYS_BIRTH and DAYS_REGISTRATION

RESULTS

The result of doing this project is a basic understanding of the risk analytics technique used in financial and banking services and understand how data is used to minimize the risk of losing money while lending to customers.

MICROSOFT EXCEL SHEET LINK

This slide includes the modified Excel sheet links

- Application data – https://1drv.ms/x/s!AkNUZuveZKg2gbQfqDvTcR63yO_LPA?e=Em8tTn
- Previous application- https://1drv.ms/x/s!AkNUZuveZKg2gbQgcgG5lRcfoP6dxA?e=evg0g9
- Column Description-columns_description.xlsx

Thank You