Bank Loan Case Study

Name- Mukesh Chandra Kamila

Description:

This project is all about carry out the in-depth analysis of important underlying insights of Bank Loan application dataset. Being a Data Analyst, my job was to provide actionable insights that can help bank make informed decisions. This project is about risk analytics of banks. When the bank receives a loan application, the bank must decide for loan approval based on the applicant's profile. EDA using excel was performed to draw insights.

Approach:

First, I have done data cleaning, dealt with blank cell / null values, removed duplicates and dropped unnecessary columns. Then I have done the EDA part using excel with the help of data analysis tool pack. And this analysis was done with the help of various functions, formula and tools.

Tech-Stack: Microsoft Excel 2019

Used:

Microsoft Excel 2019 was used in this project execution. The ease of access and set up with convenient user interface made it a good tool for the project.

Insights:

In this project, I learned about advanced EXCEL and statistics. And how to analyse the problem statement and the functions that I can use in EXCEL to solve the problem statement. All the questions asked has been answered or solved through Excel and statistics.

Result: In this project, I have achieved and gained knowledge how to deal and analyse the data with help of EXCEL and apply statistics logic and function. And how to interact and run different functions to get desired output from data.

Application_Data Worksheet:

https://docs.google.com/spreadsheets/d/1qZvLhmxRVbS8ZibQbo4f1mxuCyx5kXRT/edit?usp=sharing&ouid=110645243858390813184&rtpof=true&sd=true

Previous_Application Worksheet:

https://docs.google.com/spreadsheets/d/1HGPNVGBrM5D53xUlJeyFqa_vRh0vE-OF/edit?usp=sharing&ouid=110645243858390813184&rtpof=true&sd=true

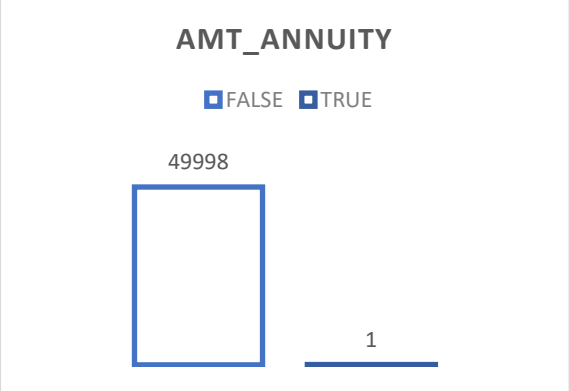
A. Identify Missing Data and Deal with it Appropriately

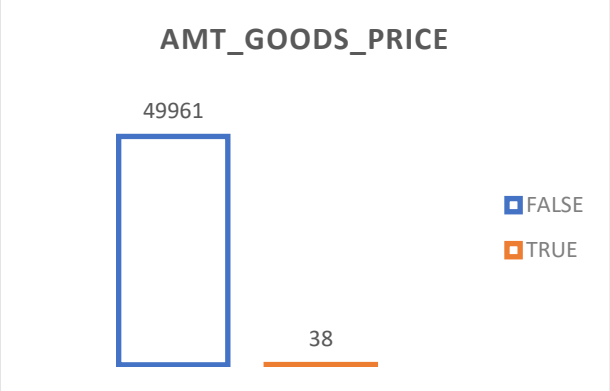
1. Dropping all the unnecessary columns in Application_Data and Previous_Application.

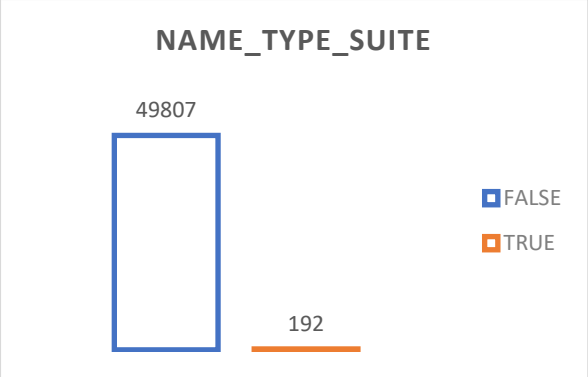
2. Dealing missing data with mean, median in numerical columns and mode in categorical columns.

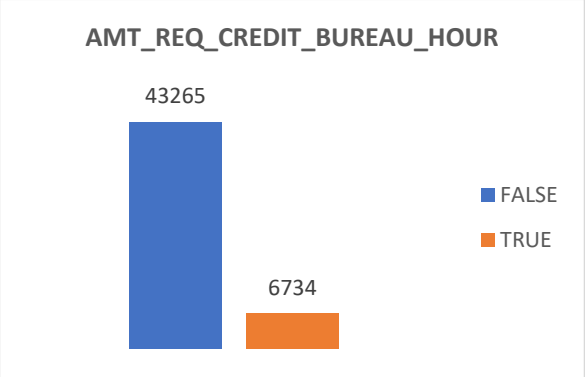
3. Converting the negative days column into positive days by multiplying (-1).

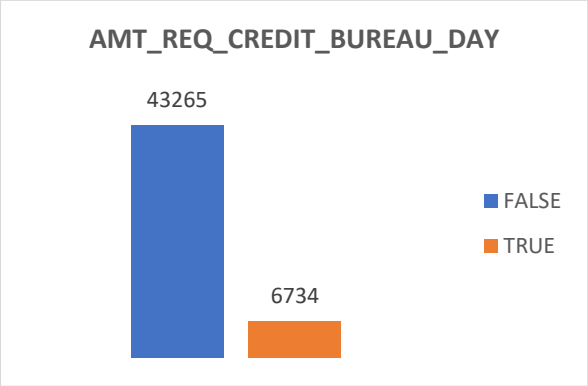
Application_Data (Missing Values):

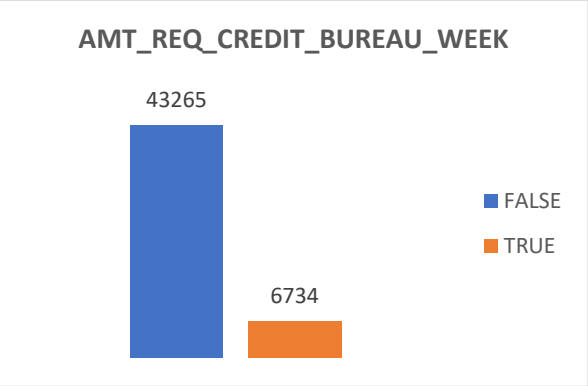


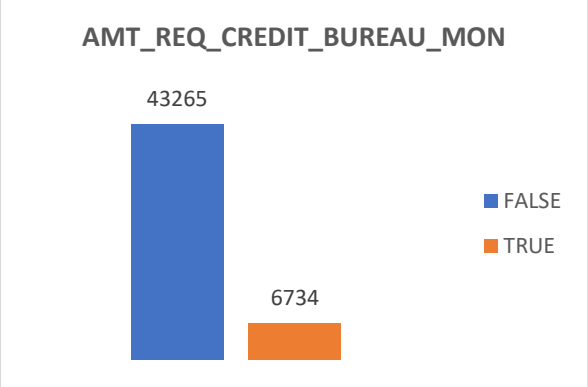


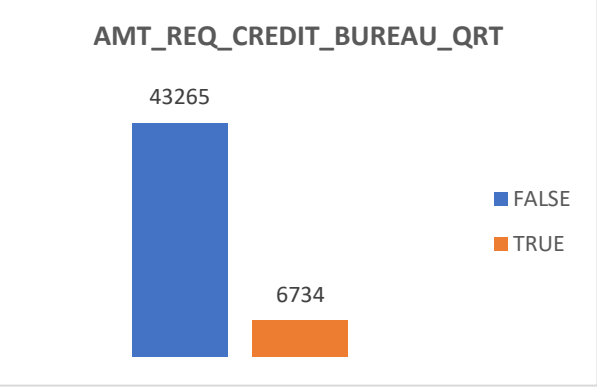

NAME_TYPE_SUITE
49807
192
FALSE
TRUE

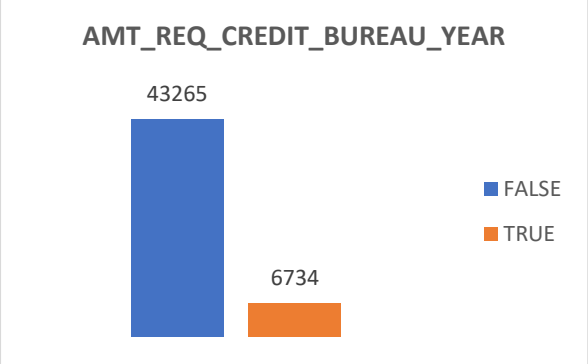

AMT_REQ_CREDIT_BUREAU_HOUR
43265
6734
FALSE
TRUE


AMT_REQ_CREDIT_BUREAU_DAY
43265
6734
FALSE
TRUE

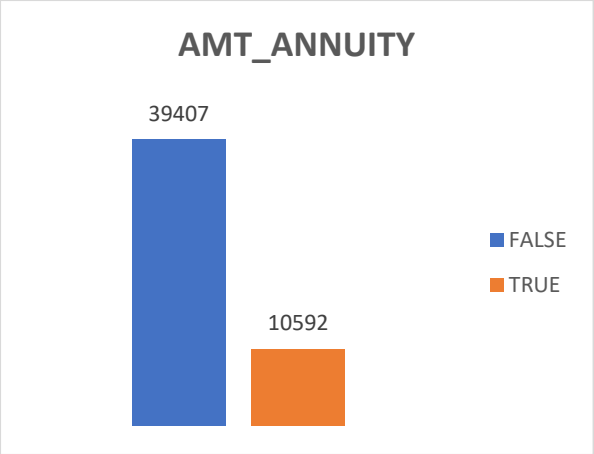

AMT_REQ_CREDIT_BUREAU_WEEK
43265
6734
FALSE
TRUE

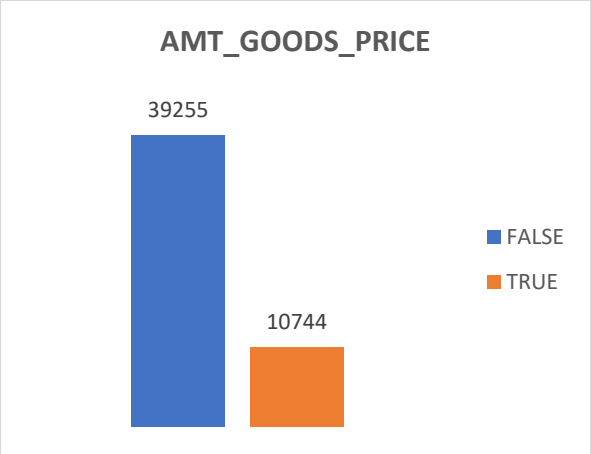

AMT_REQ_CREDIT_BUREAU_MON
43265
6734
FALSE
TRUE

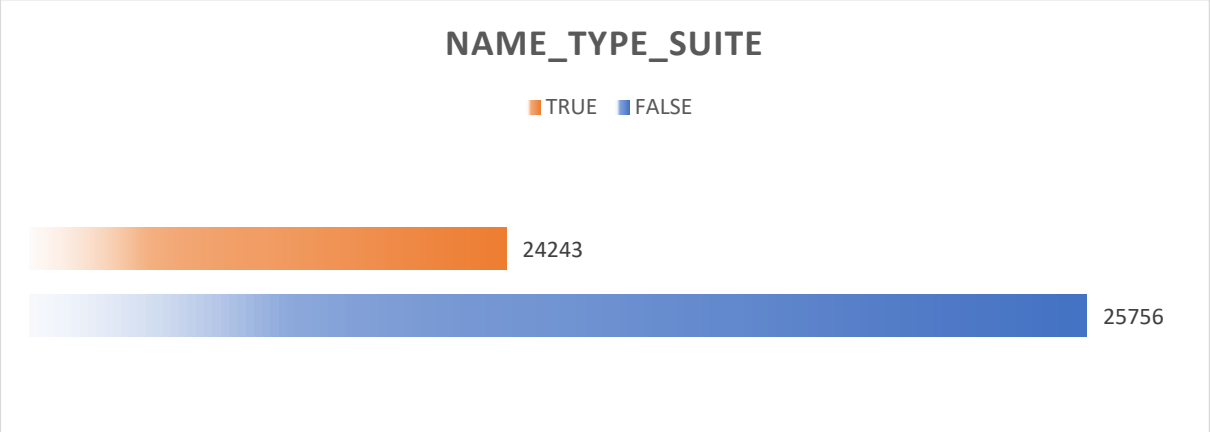

AMT_REQ_CREDIT_BUREAU_QRT
43265
6734
FALSE
TRUE


AMT_REQ_CREDIT_BUREAU_YEAR
43265
6734
FALSE
TRUE

Previous_Application(Missing Values):







B. Identify Outliers in the Dataset

Application_Data (Outliers):

AMT_INCOME_TOTAL		AMT_CREDIT		AMT_ANNUITY		AMT_GOODS_PRICE		DAYS_BIRTH		DAYS_EMPLOYED	
Q1	112500	Q1	270000	Q1	16456.5	Q1	238500	Q1	12378.5	Q1	292
Q3	202500	Q3	808650	Q3	34596	Q3	679500	Q3	19644	Q3	2786
IQR	90000	IQR	538650	IQR	18139.5	IQR	441000	IQR	7265.5	IQR	2494
UPPER LIMIT	337500	UPPER LIMIT	1616625	UPPER LIMIT	61805.25	UPPER LIMIT	1341000	UPPER LIMIT	30542.25	UPPER LIMIT	6527
LOWER LIMIT	0	LOWER LIMIT	0	LOWER LIMIT	0	LOWER LIMIT	0	LOWER LIMIT	1480.25	LOWER LIMIT	0
OUTLIERS		OUTLIERS		OUTLIERS		OUTLIERS		OUTLIERS		OUTLIERS	
TRUE	2295	TRUE	1063	TRUE	1188	TRUE	2387	TRUE	0	TRUE	11712
FALSE	47703	FALSE	48935	FALSE	48810	FALSE	47611	FALSE	49998	FALSE	38286

Previous_Application(Outliers):

AMT_ANNUITY		AMT_APPLICATION		AMT_CREDIT		AMT_GOODS_PRICE		CNT_PAYMENT	
Q1	7189.74	Q1	22045.5	Q1	26055	Q1	63663.75	Q1	0
Q3	16256.16	Q3	180000	Q3	198105.8	Q3	180000	Q3	14
IQR	9066.42	IQR	157954.5	IQR	172050.8	IQR	116336.3	IQR	14
UPPER LIMIT	29855.79	UPPER LIMIT	416931.8	UPPER LIMIT	456181.9	UPPER LIMIT	354504.4	UPPER LIMIT	35
LOWER LIMIT	0	LOWER LIMIT	0	LOWER LIMIT	0	LOWER LIMIT	0	LOWER LIMIT	0
OUTLIERS		OUTLIERS		OUTLIERS		OUTLIERS		OUTLIERS	
TRUE	4922	TRUE	5792	TRUE	5648	TRUE	6569	TRUE	4968
FALSE	45077	FALSE	44207	FALSE	44351	FALSE	43430	FALSE	45031

C. Analyze Data Imbalance

Application_Data (Data Imbalance):

With countIF formula-

NAME_CONTRACT_TYPE		NAME_TYPE_SUITE		NAME_INCOME_TYPE		NAME_HOUSING_TYPE	
Cash loans	45276	Unaccompani	40627	Working	26010	Co-op apartment	191
Revolving loa	4723	Children	542	Commercial	11543	House / apartment	44368
		Family	6549	State servan	3512	Municipal apartment	1845
CODE_GENDER		Group of peop	36	Pensioner	8920	Office apartment	427
F	32823	Other_A	137	Businessmar	2	Rented apartment	769
M	17174	Other_B	259	Maternity le	1	With parents	2399
XNA	2	Spouse, partn	1849	Student	5		
				Unemployed	6		

With pivot table-

CONTRACT TYPE	Count
Cash loans	45276
Revolving loans	4723
Grand Total	**49999**

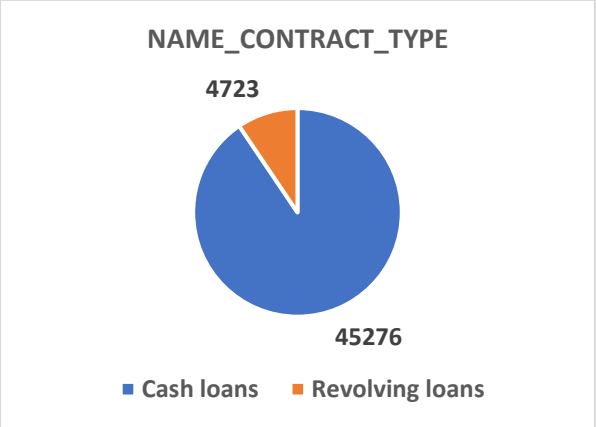
GENDER	Count
F	32823
M	17174
XNA	2
Grand Total	**49999**

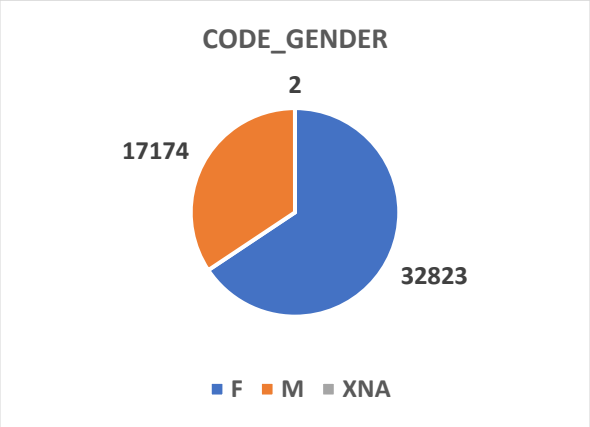
NAME TYPE SUITE	Count
Children	542
Family	6549
Group of people	36
Other_A	137
Other_B	259
Spouse, partner	1849
Unaccompanied	40627
Grand Total	**49999**

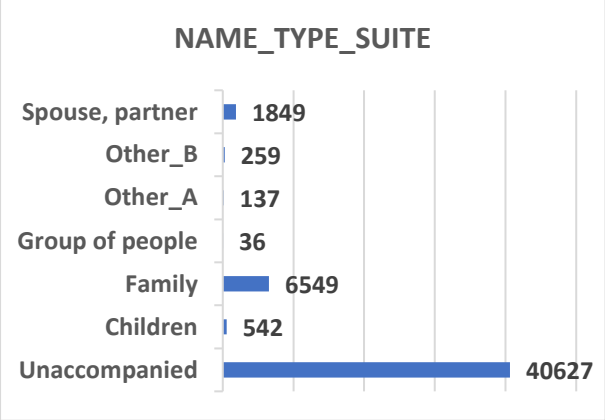
INCOME TYPE	Count
Co-op apartment	191
House / apartment	44368
Municipal apartment	1845
Office apartment	427
Rented apartment	769
With parents	2399
Grand Total	**49999**

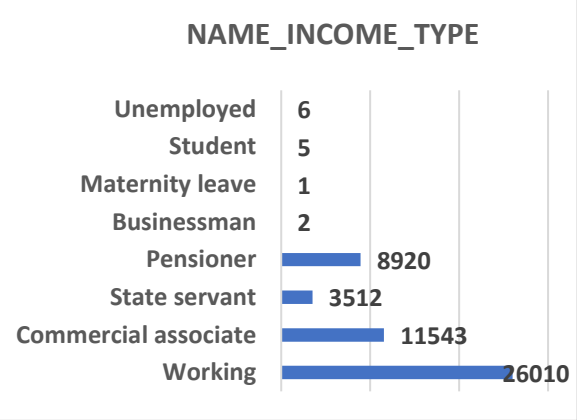
INCOME TYPE	Count
Businessman	2
Commercial associate	11543
Maternity leave	1
Pensioner	8920
State servant	3512
Student	5
Unemployed	6
Working	26010
Grand Total	**49999**

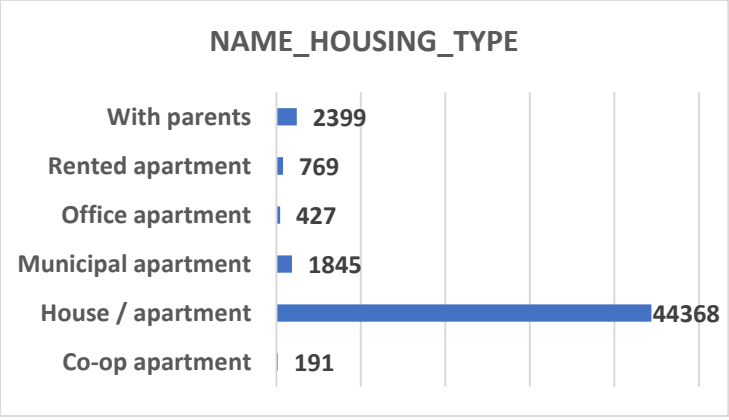
Graphs:


NAME_CONTRACT_TYPE
4723
45276
Cash loans
Revolving loans


CODE_GENDER
2
17174
32823
F
M
XNA


NAME_TYPE_SUITE
Spouse, partner 1849
Other_B 259
Other_A 137
Group of people 36
Family 6549
Children 542
Unaccompanied 40627


NAME_INCOME_TYPE
Unemployed 6
Student 5
Maternity leave 1
Businessman 2
Pensioner 8920
State servant 3512
Commercial associate 11543
Working 26010


NAME_HOUSING_TYPE
With parents 2399
Rented apartment 769
Office apartment 427
Municipal apartment 1845
House / apartment 44368
Co-op apartment 191

Previous_Application(Data Imbalance):

With countIF formula-

NAME_CONTRACT_TYPE		NAME_CONTRACT_STATUS		NAME_TYPE_SUITE		NAME_CLIENT_TYPE		NAME_PORTFOLIO		NAME_YIELD_GROUP	
Cash loans	20856	Approved	31885	Children	993	New	9548	Cards	4210	high	10807
Consumer loans	23510	Canceled	8595	Family	6581	Refreshed	4227	Cars	14	low_action	2953
Revolving loans	5625	Refused	8660	Group of peopl	76	Repeater	36167	Cash	12917	low_normal	9858
XNA	8	Unused offer	859	Other_A	262	XNA	57	POS	22266	middle	11579
				Other_B	551			XNA	10592	XNA	14802
				Spouse, partner	2098						
				Unaccompanied	39438						

With pivot table-

Contact type	Count
Cash loans	20856
Consumer loans	23510
Revolving loans	5625
XNA	8
Grand Total	**49999**

Loan status	Count
Approved	31885
Canceled	8595
Refused	8660
Unused offer	859
Grand Total	**49999**

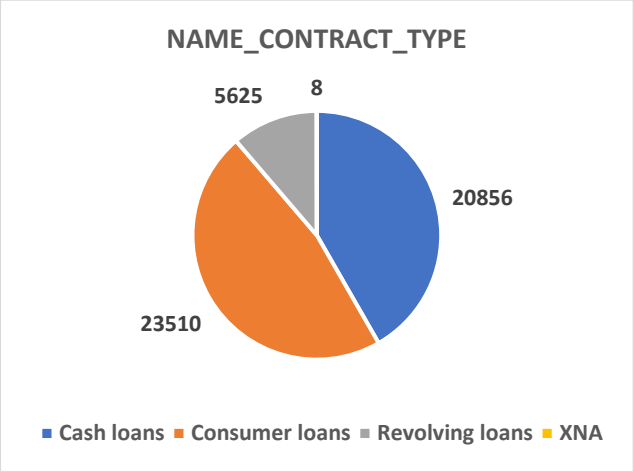
Row Labels	Count
Children	993
Family	6581
Group of people	76
Other_A	262
Other_B	551
Spouse, partner	2098
Unaccompanied	39438
Grand Total	**49999**

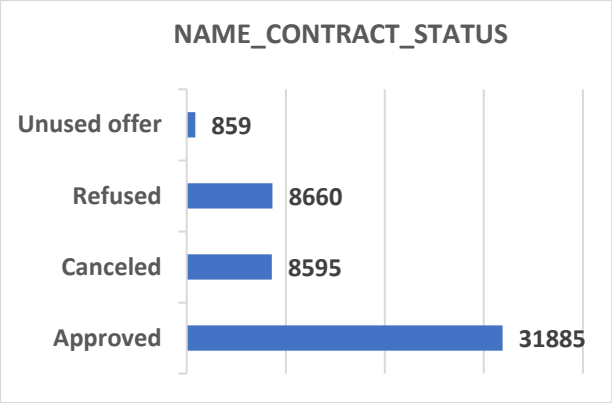
Client type	Count
New	9548
Refreshed	4227
Repeater	36167
XNA	57
Grand Total	**49999**

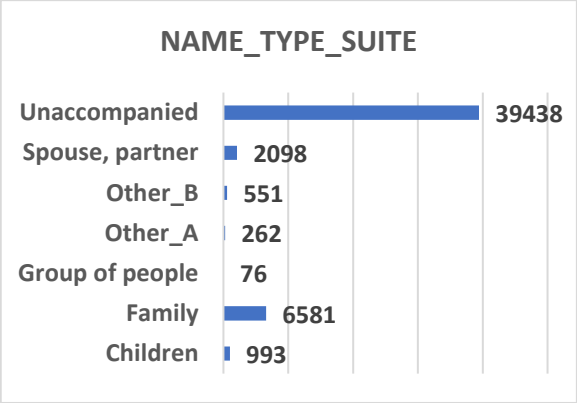
PORTFOLIO	Count
Cards	4210
Cars	14
Cash	12917
POS	22266
XNA	10592
Grand Total	**49999**

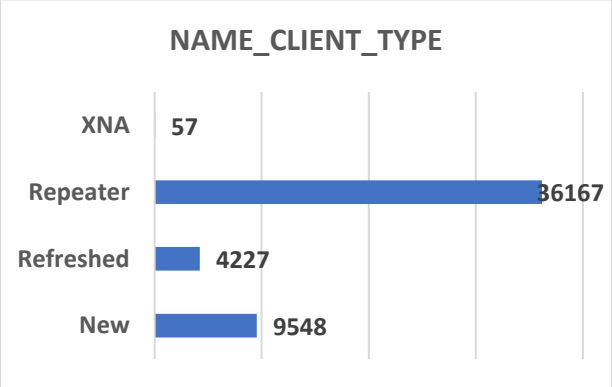
YEILD_GROUP	Count
high	10807
low_action	2953
low_normal	9858
middle	11579
XNA	14802
Grand Total	**49999**

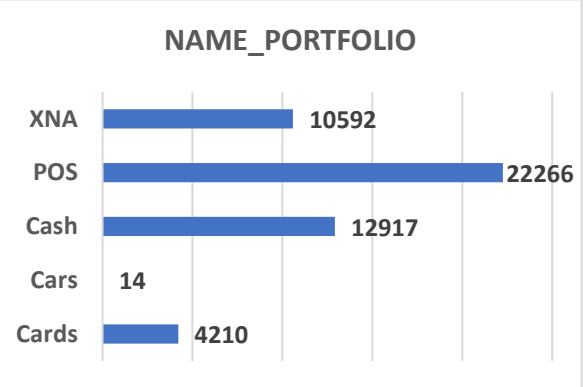
Graphs:

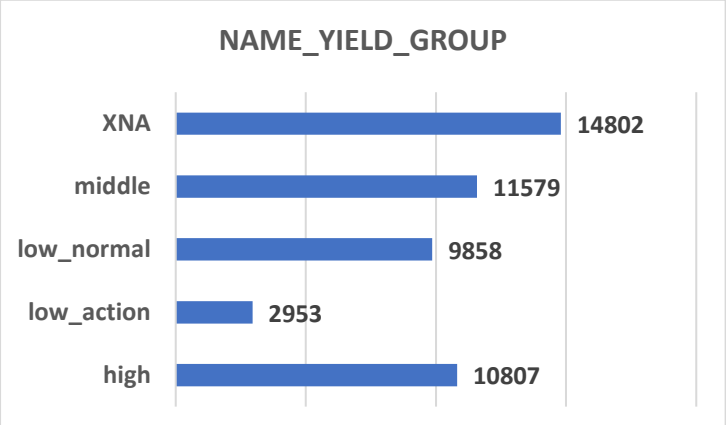

NAME_CONTRACT_TYPE
5625
8
20856
23510
Cash loans
Consumer loans
Revolving loans
XNA


NAME_CONTRACT_STATUS
Unused offer
859
Refused
8660
Canceled
8595
Approved
31885


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


NAME_CLIENT_TYPE
XNA
57
Repeater
36167
Refreshed
4227
New
9548


NAME_PORTFOLIO
XNA
10592
POS
22266
Cash
12917
Cars
14
Cards
4210


NAME_YIELD_GROUP
XNA
14802
middle
11579
low_normal
9858
low_action
2953
high
10807

D. Perform Univariate, Segmented Univariate and Bivariate Analysis

Application_Data

Univariate Analysis:

NAME_TYPE_SUITE	Count	NAME_INCOME_TYPE	Count	NAME_HOUSING_TYPE	Count
Unaccompanied	40627	Working	26010	Co-op apartment	191
Children	542	Commercial associate	11543	House / apartment	44368
Family	6549	State servant	3512	Municipal apartment	1845
Group of people	36	Pensioner	8920	Office apartment	427
Other_A	137	Businessman	2	Rented apartment	769
Other_B	259	Maternity leave	1	With parents	2399
Spouse, partner	1849	Student	5		
		Unemployed	6		

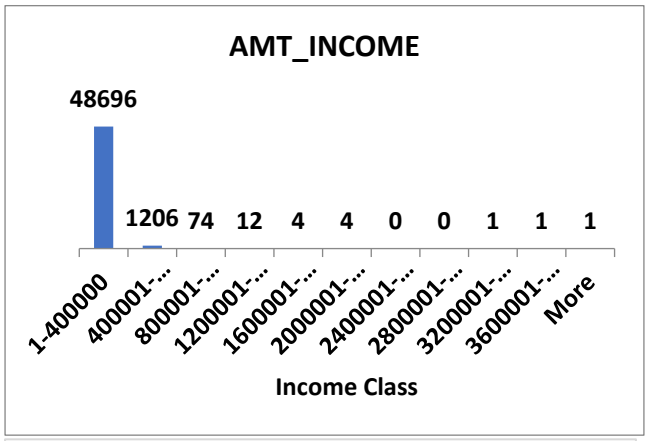
AMT_INCOME_TOTAL	
Mean	170767.5905
Standard Error	2378.391081
Median	145800
Mode	135000
Standard Deviation	531819.0951
Sample Variance	2.82832E+11
Kurtosis	46582.52582
Skewness	212.0777967
Range	116974350
Minimum	25650
Maximum	117000000
Sum	8538208758
Count	49999

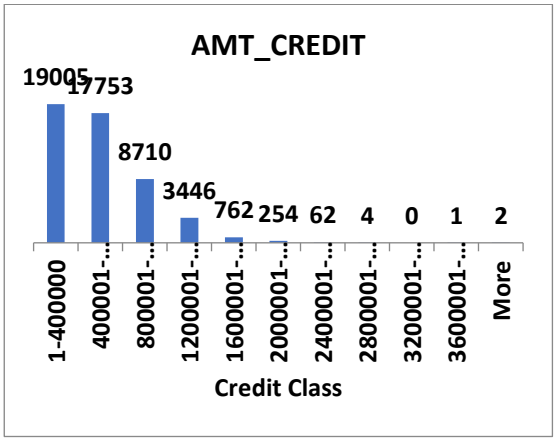
Income Class	***Frequency***
1-400000	48696
400001-800000	1206
800001-1200000	74
1200001-1600000	12
1600001-2000000	4
2000001-2400000	4
2400001-2800000	0
2800001-3200000	0
3200001-3600000	1
3600001-4000000	1
More	1

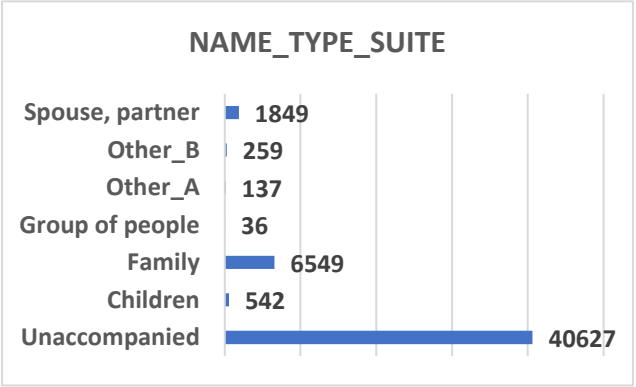
AMT_CREDIT	
Mean	599700.5815
Standard Error	1799.674528
Median	514777.5
Mode	450000
Standard Deviation	402415.4339
Sample Variance	1.61938E+11
Kurtosis	1.917459058
Skewness	1.223668739
Range	4005000
Minimum	45000
Maximum	4050000
Sum	29984429376
Count	49999

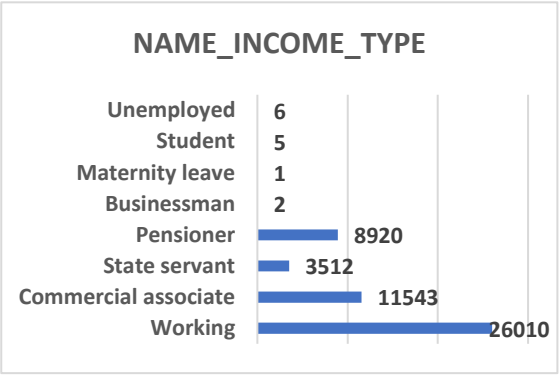
AMT_CREDIT Class	***Frequency***
1-400000	19005
400001-800000	17753
800001-1200000	8710
1200001-1600000	3446
1600001-2000000	762
2000001-2400000	254
2400001-2800000	62
2800001-3200000	4
3200001-3600000	0
3600001-4000000	1
More	2

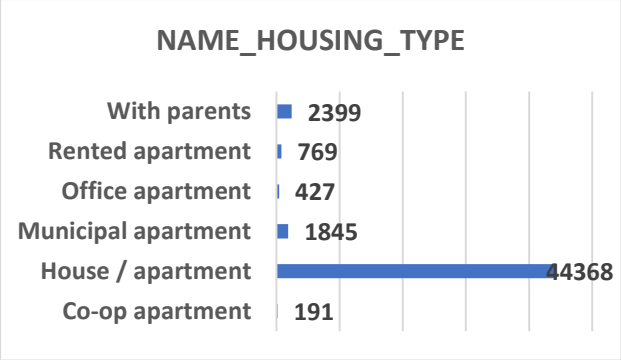
Graphs:











Bivariate Analysis:

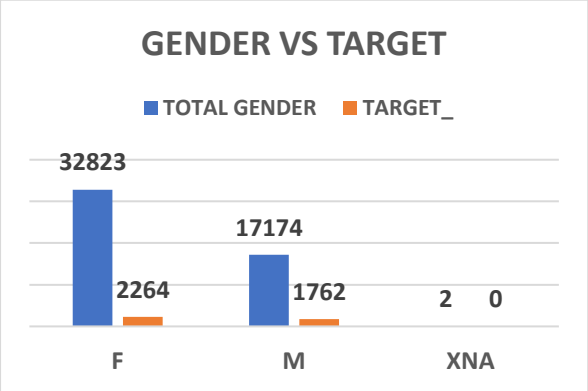
GENDER	TOTAL GENDER	TARGET_
F	32823	2264
M	17174	1762
XNA	2	0
Grand Total	**49999**	**4026**

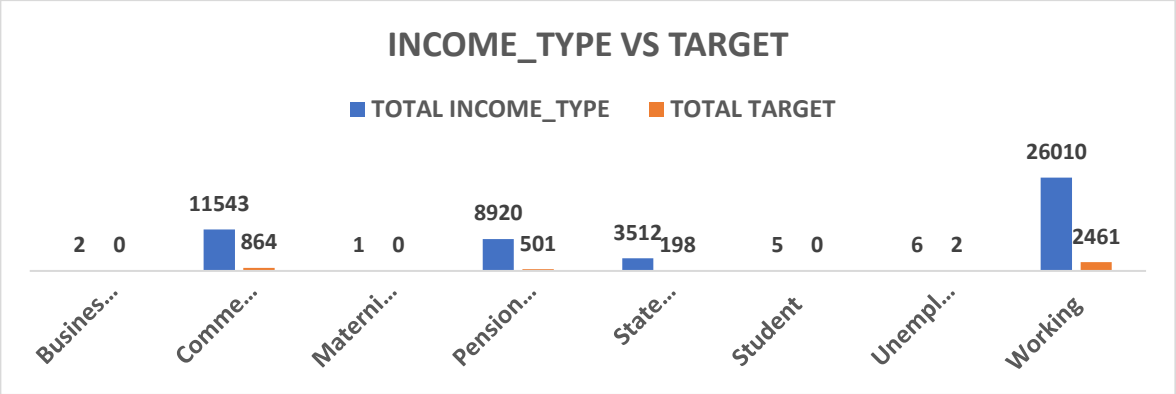
INCOME TYPE	TOTAL INCOME_TYPE	TOTAL TARGET
Businessman	2	0
Commercial associate	11543	864
Maternity leave	1	0
Pensioner	8920	501
State servant	3512	198
Student	5	0
Unemployed	6	2
Working	26010	2461
Grand Total	**49999**	**4026**

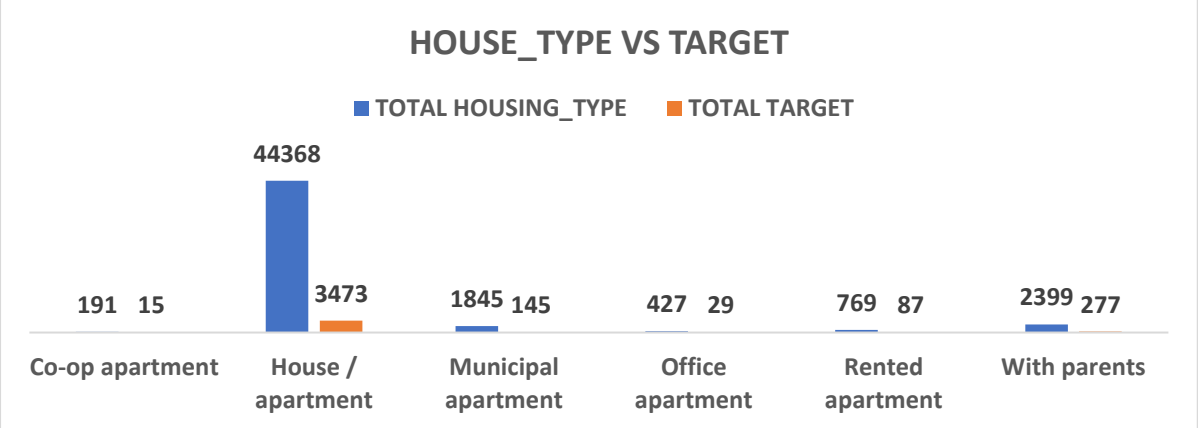
HOUSE TYPE	TOTAL HOUSING_TYPE	TOTAL TARGET
Co-op apartment	191	15
House / apartment	44368	3473
Municipal apartment	1845	145
Office apartment	427	29
Rented apartment	769	87
With parents	2399	277
Grand Total	**49999**	**4026**

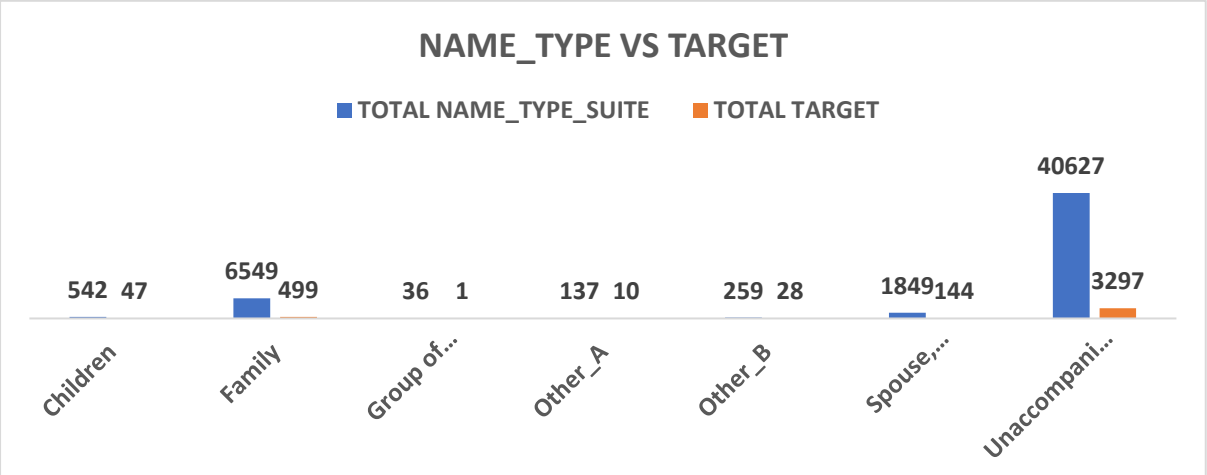
NAME TYPE	TOTAL NAME_TYPE_SUITE	TOTAL TARGET
Children	542	47
Family	6549	499
Group of people	36	1
Other_A	137	10
Other_B	259	28
Spouse, partner	1849	144
Unaccompanied	40627	3297
Grand Total	**49999**	**4026**

Graphs:


GENDER VS TARGET
TOTAL GENDER
TARGET_
32823
2264
17174
1762
2
0
F
M
XNA


INCOME_TYPE VS TARGET
TOTAL INCOME_TYPE
TOTAL TARGET
2
0
11543
864
1
0
8920
501
3512
198
5
0
6
2
26010
2461
Busines...
Comme...
Materni...
Pension...
State...
Student
Unempl...
Working


HOUSE_TYPE VS TARGET
TOTAL HOUSING_TYPE
TOTAL TARGET
191
15
44368
3473
1845
145
427
29
769
87
2399
277
Co-op apartment
House / apartment
Municipal apartment
Office apartment
Rented apartment
With parents


NAME_TYPE VS TARGET
TOTAL NAME_TYPE_SUITE
TOTAL TARGET
542
47
6549
499
36
1
137
10
259
28
1849
144
40627
3297
Children
Family
Group of...
Other_A
Other_B
Spouse,...
Unaccompani...

Previous_Application

Univariate Analysis:

AMT_APPLICATION	
Mean	168892.4546
Standard Error	1262.065087
Median	71550
Mode	0
Standard Deviation	282203.5105
Sample Variance	79638821323
Kurtosis	16.08576848
Skewness	3.471726872
Range	3826372.5
Minimum	0
Maximum	3826372.5
Sum	8444453838
Count	49999

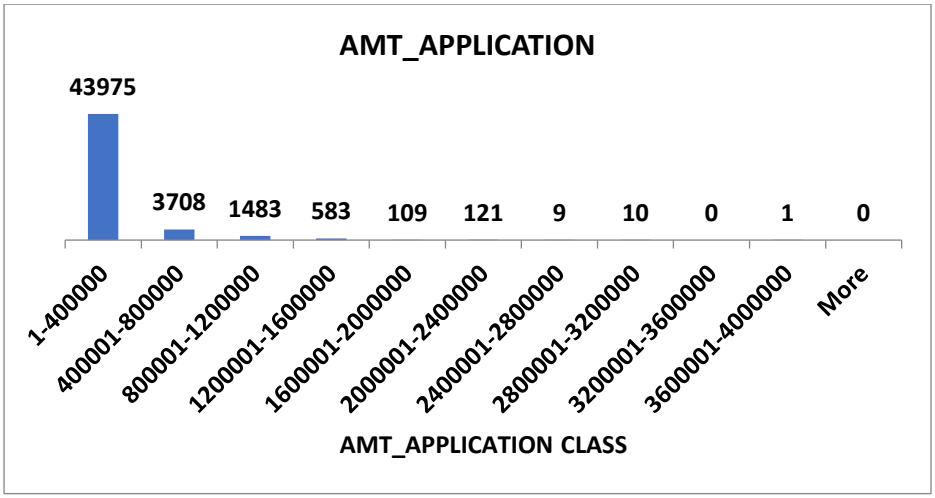
AMT_APPLication Class	***Frequency***
1-400000	43975
400001-800000	3708
800001-1200000	1483
1200001-1600000	583
1600001-2000000	109
2000001-2400000	121
2400001-2800000	9
2800001-3200000	10
3200001-3600000	0
3600001-4000000	1
More	0

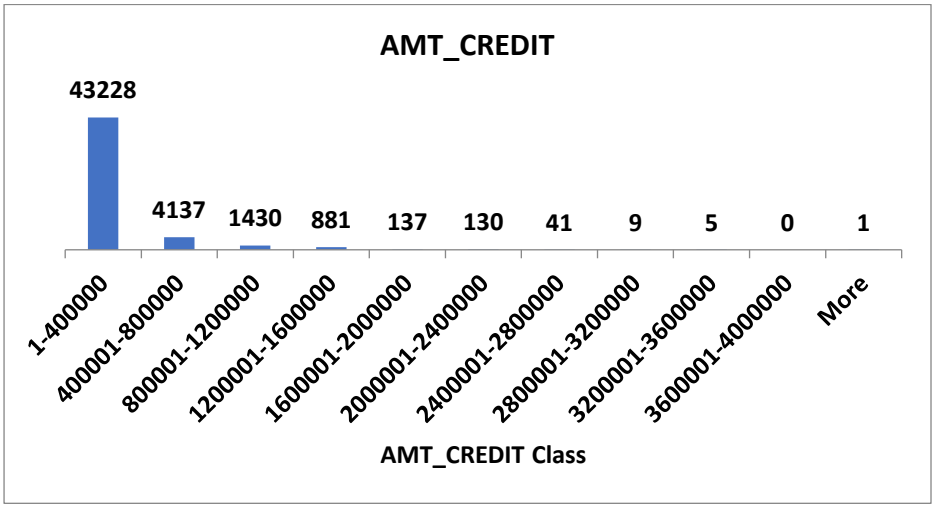
AMT_CREDIT	
Mean	188542.8855
Standard Error	1379.549679
Median	78907.5
Mode	0
Standard Deviation	308473.6014
Sample Variance	95155962744
Kurtosis	14.88061385
Skewness	3.344679263
Range	4104351
Minimum	0
Maximum	4104351
Sum	9426955730
Count	49999

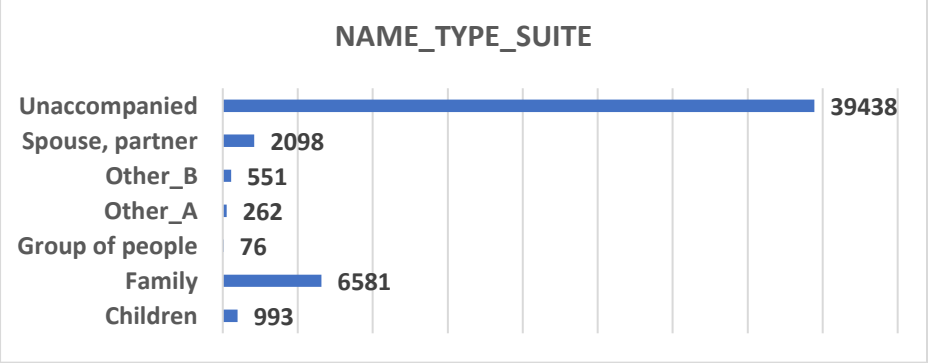
AMT_CREDIT Class	***Frequency***
1-400000	43228
400001-800000	4137
800001-1200000	1430
1200001-1600000	881
1600001-2000000	137
2000001-2400000	130
2400001-2800000	41
2800001-3200000	9
3200001-3600000	5
3600001-4000000	0
More	1

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

Graphs:







E. Identify Top Correlations for Different Scenarios

Application_Data (Correlation Matrices):

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH
AMT_INCOME_TOTAL	1							
AMT_CREDIT	0.06932	1						
AMT_ANNUITY	0.08301	0.7695	1					
AMT_GOODS_PRICE	0.06989	0.9867	0.774133	1				
DAYS_BIRTH	-0.016	0.05934	-0.00771	0.0577	1			
DAYS_EMPLOYED	-0.03151	-0.0677	-0.10871	-0.065	0.62173	1		
DAYS REGISTRATION	-0.00995	-0.0034	-0.03322	-0.006	0.33363	0.20917	1	
DAYS_ID_PUBLISH	-0.00351	0.01223	-0.00672	0.014	0.27083	0.27277	0.1043	1

Previous_Application (Correlation Matrices):

	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	DAYS_DECISION	CNT_PAYMENT
AMT_ANNUITY	1					
AMT_APPLICATION	0.775697159	1				
AMT_CREDIT	0.780417284	0.975771049	1			
AMT_GOODS_PRICE	0.80383156	0.988711762	0.97235717	1		
DAYS_DECISION	-0.245175364	-0.13399106	-0.13743121	-0.190283111	1	
CNT_PAYMENT	0.343477753	0.704327354	0.702616823	0.658715935	-0.013981989	1