

Bank Loan Case Study

Project Description:

In this case study, we will apply the techniques that we 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.

Approach:

My first approach was to understand the dataset, clean the dataset finding the blanks and missing values, inputting the missing values with the appropriate method (mean, median, mode). Then I tried to find the outliers in the dataset, there are some anomalies such as negative values which need either to be deleted or standardized. After all these I used pivot tables and basic charts to visualize the data. Moreover, insights were drawn based on my understandings.

Tech-Used:

Microsoft excel 2016 was used for creating pivot tables and charts which helped in visualizing as well as analyzing the data. And Microsoft word was used for writing the project's description.

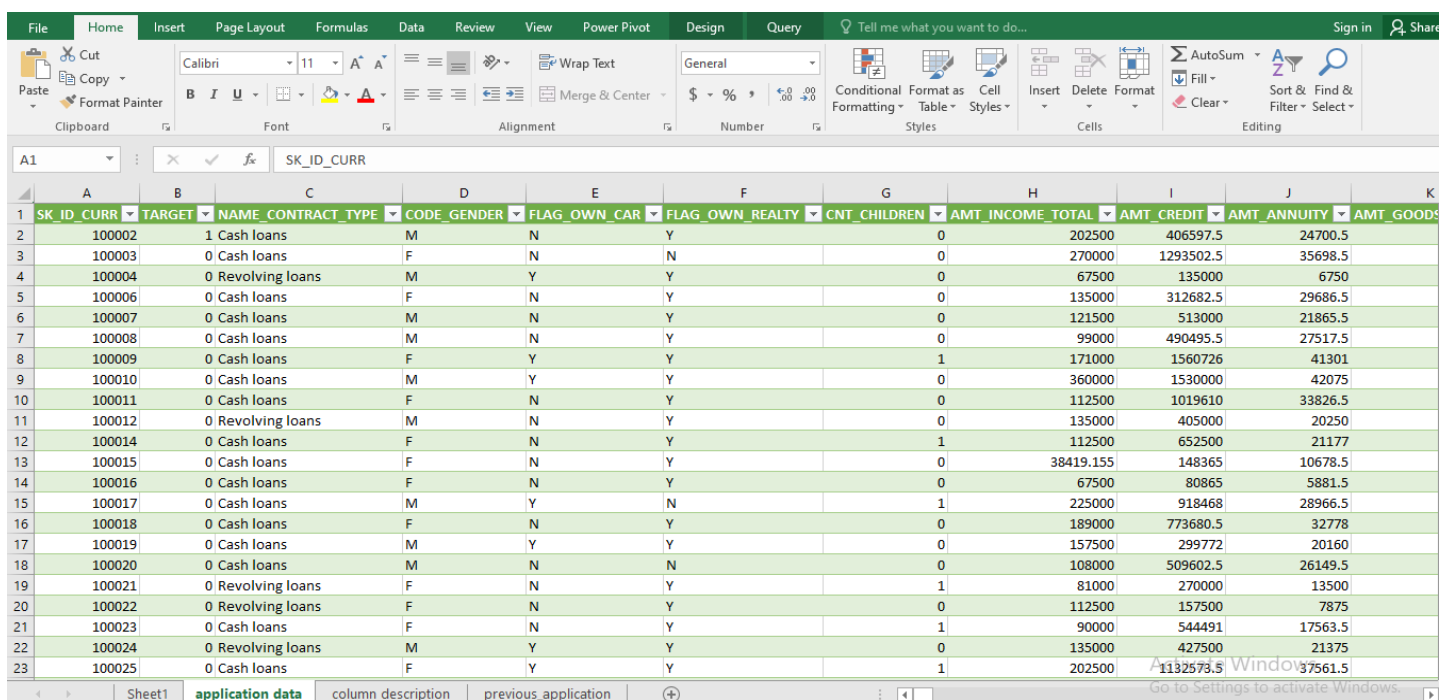
Insights:

following insights were drawn based on understanding and analysis of data:

- Cash loans have higher percentage as compared to revolving loan.
- People with lower total income are more likely to default
- The defaulters of cash loans are higher than the revolving ones.
- Most of the laborers have problems in loan payment.
- People who just got employed tends to take more loans and have greater chance to get default.

Results: Following results were obtained while doing the project:

Imported the dataset in excel:



	A	B	C	D	E	F	G	H	I	J	K
1	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
2	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	
3	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	
4	100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	
5	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	
6	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	
7	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	
8	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	
9	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	
10	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	
11	100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	
12	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	
13	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	
14	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	
15	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	
16	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	
17	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	
18	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	
19	100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	
20	100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	
21	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	
22	100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	
23	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	

File

Home

Insert

Page Layout

Formulas

Data

Review

View

Power Pivot

Design

Query

Tell me what you want to do...

Sign in

Share

Cut

Copy

Format Painter

Clipboard

Calibri

11

A

A

</

FileHomeInsertPage LayoutFormulasDataReviewViewPower PivotDesignQuery

Tell me what you want to do...

Sign inShare

CutCopyFormat Painter

Clipboard

Calibri11

BBIU

Font

Wrap Text

Alignment

General

Number

Conditional Formatting

Format as Table

Cell Styles

Styles

InsertDeleteFormat

Cells

AutoSumFillClear

Editing

Sort & Find & Filter & Select

F9

X

✓

0

	A	B	C	D	E	F	G	H	I	J
1	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKDAY_APPR_PROCESS_START	HO
2	2030495	271877	Consumer loans	1730.43	17145	17145	0	17145	SATURDAY	
3	2802425	108129	Cash loans	25188.615	607500	679671		607500	THURSDAY	
4	2523466	122040	Cash loans	15060.735	112500	136444.5		112500	TUESDAY	
5	2819243	176158	Cash loans	47041.335	450000	470790		450000	MONDAY	
6	1784265	202054	Cash loans	31924.395	337500	404055		337500	THURSDAY	
7	1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	
8	2315218	175704	Cash loans		0	0			TUESDAY	
9	1656711	296299	Cash loans		0	0			MONDAY	
10	2367563	342292	Cash loans		0	0			MONDAY	
11	2579447	334349	Cash loans		0	0			SATURDAY	
12	1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	
13	2257824	161140	Cash loans	13832.775	211500	246397.5		211500	FRIDAY	
14	2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	
15	1397919	321676	Consumer loans	7654.86	53779.5	57564	0	53779.5	SUNDAY	
16	2273188	270658	Consumer loans	9644.22	26550	27252	0	26550	SATURDAY	
17	1232483	151612	Consumer loans	21307.455	126490.5	119853	12649.5	126490.5	TUESDAY	
18	2163253	154602	Consumer loans	4187.34	26955	27297		26955	SATURDAY	
19	1285768	142748	Revolving loans	9000	180000	180000		180000	FRIDAY	
20	2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	
21	1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	
22	1506815	166490	Cash loans	25454.025	450000	491580		450000	MONDAY	
23	1182516	267782	Cash loans	20361.6	405000	451777.5		405000	SATURDAY	

Sheet1

application data

column description

previous application

Go to Settings to activate Windows.

Cleared the column if occurrences of blanks cells are greater than 50% otherwise filled them using average or median values.

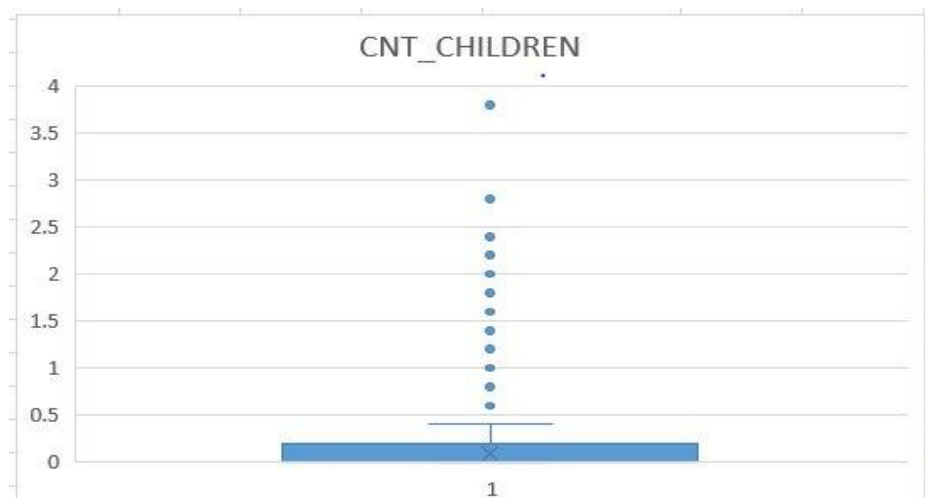
Note: 0 is considered as Non-defaulters and 1 is considered as defaulters all over the report.

Identify the outlier in the dataset.

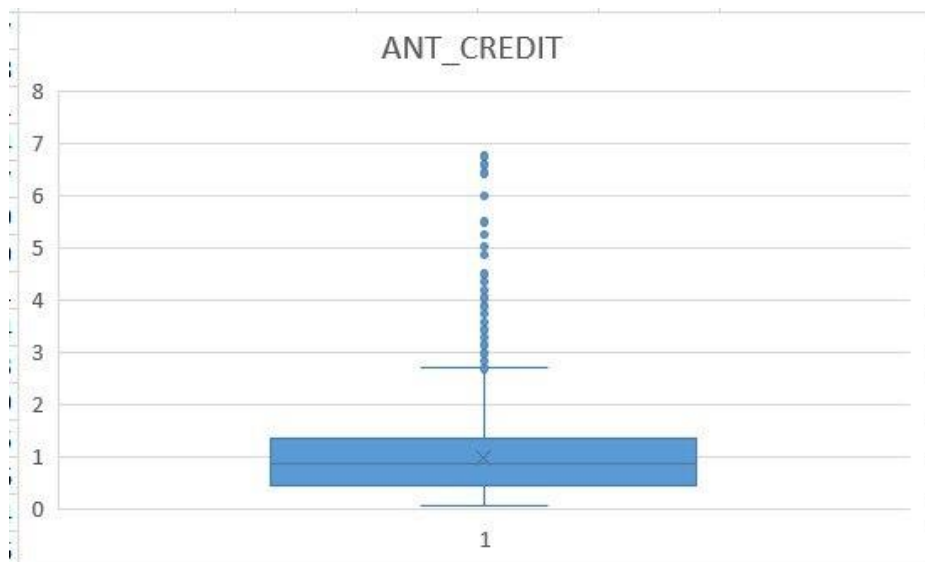
Solution:

We check numerical column-wise to check outlier and reported then on the basis of 4 variables. We used whisker graph to show outlier, the values below the whisker value (i.e.; 1) are considered outliers.

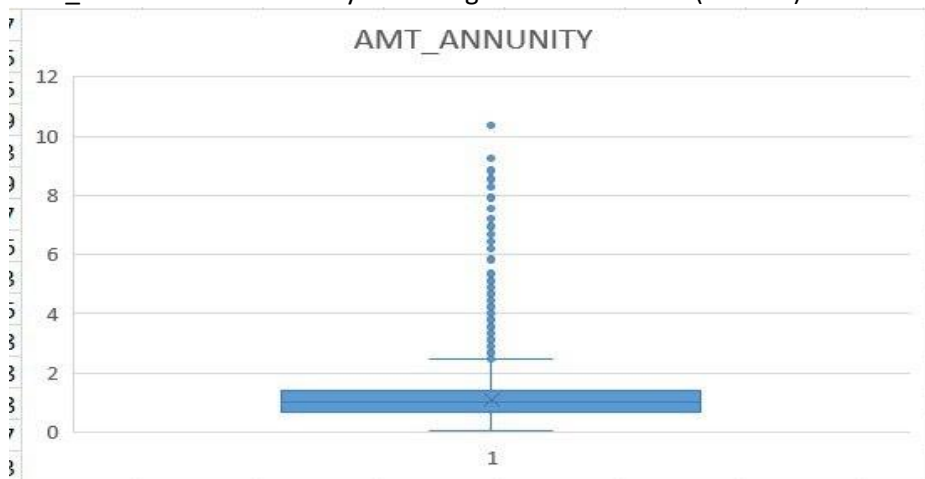
- CNT_CHILDREN: The value greater than 5 are considered as Outlier.



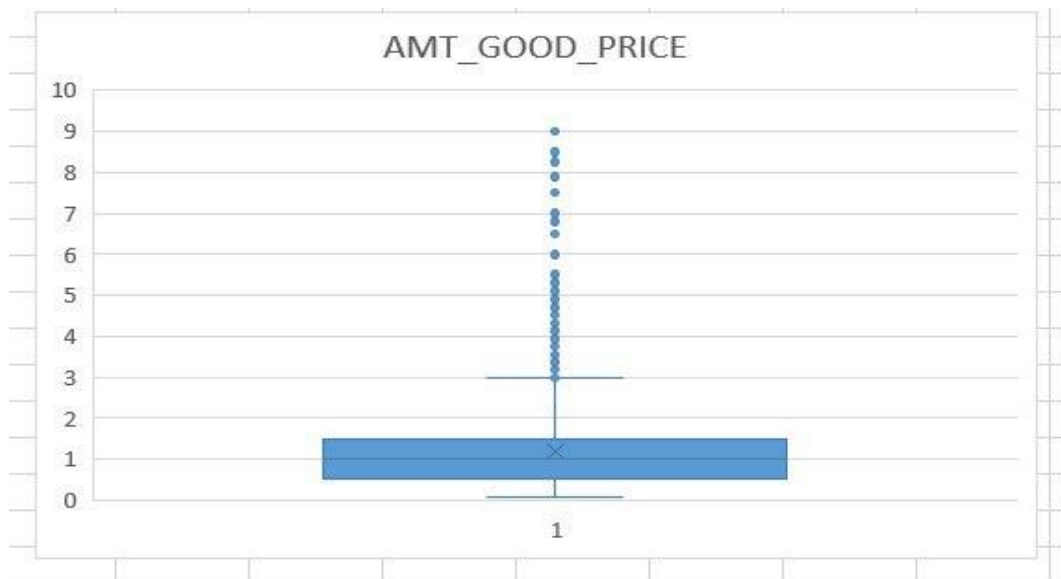
- AMT_CREDIT: The amount credited greater than 598992 (average) can be considered as an outlier.



- AMT_ANNUNITY: Loan annuity amount greater than 24903 (median) is considered to be an outlier.



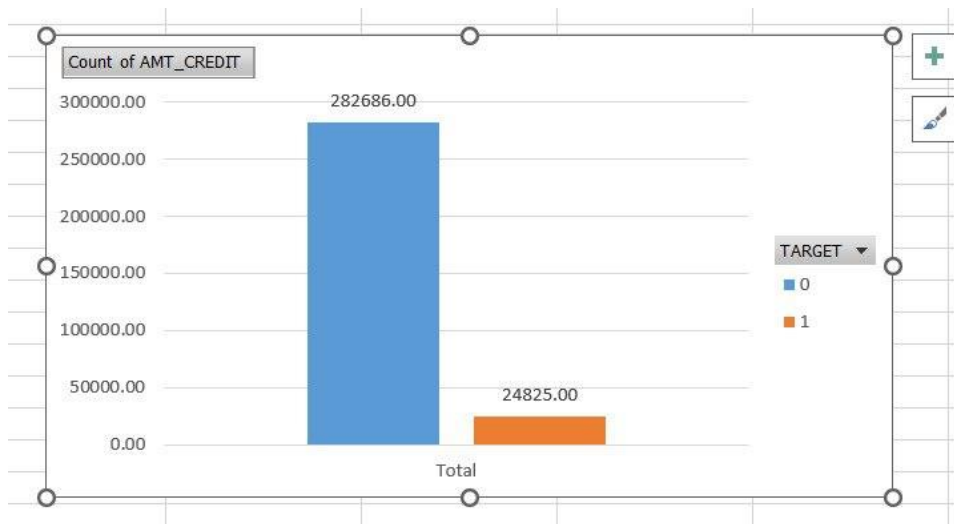
- AMT_GOODS_PRICE: Loan goods price amount greater than 450000 (median) is considered to be an outlier.



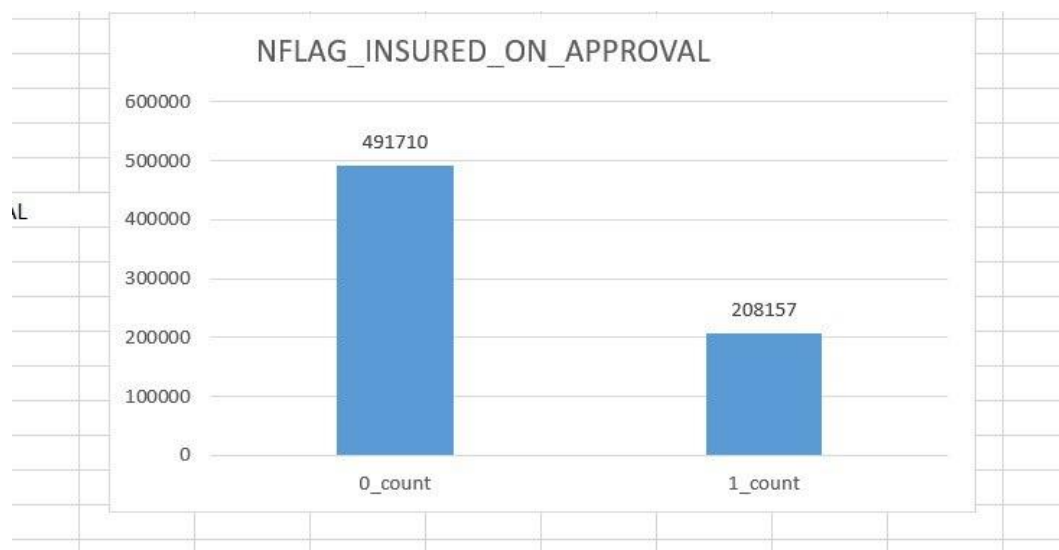
Identify if there is imbalance in the data.

Solution:

- There is also imbalance in count of amt_credit between the targets 0 and 1.



- There is imbalance in values of NFLAG_INSURED_ON_APPROVAL of count_0 and count_1.

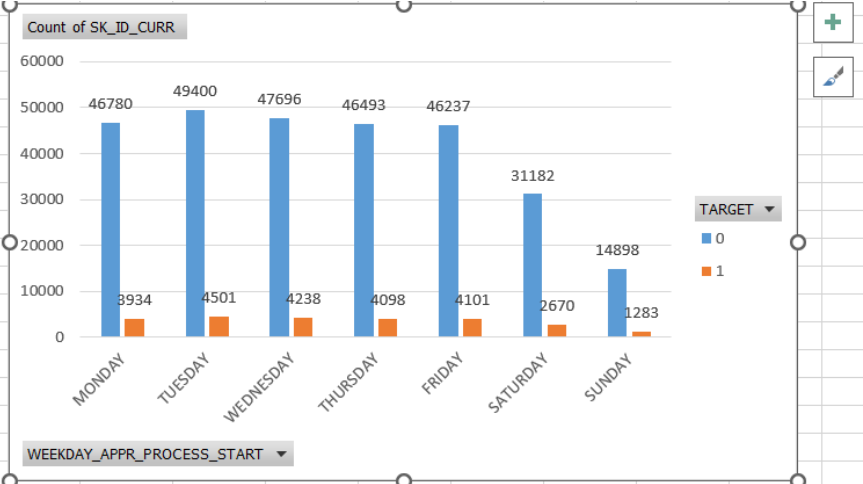


Explain the results of univariate, segmented variate and bivariate analysis.

Solution:

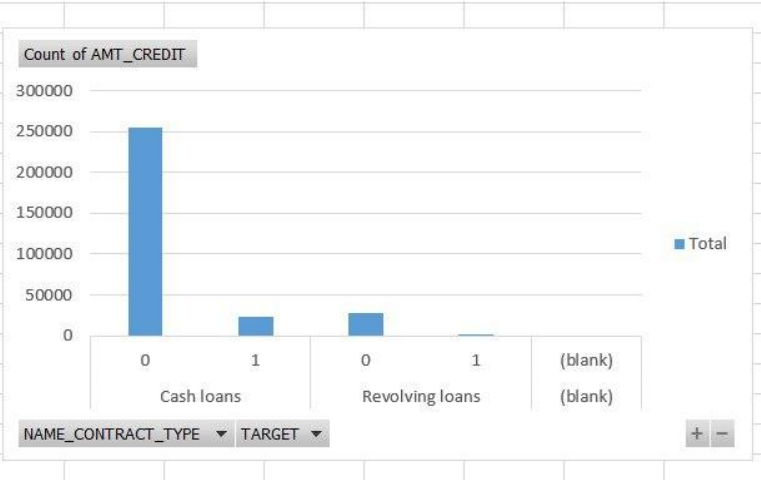
- Univariate analysis

Count of SK_ID_CURR			
Row Labels	0	1	Grand Total
MONDAY	46780	3934	50714
TUESDAY	49400	4501	53901
WEDNESDAY	47696	4238	51934
THURSDAY	46493	4098	50591
FRIDAY	46237	4101	50338
SATURDAY	31182	2670	33852
SUNDAY	14898	1283	16181
Grand Total	282686	24825	307511



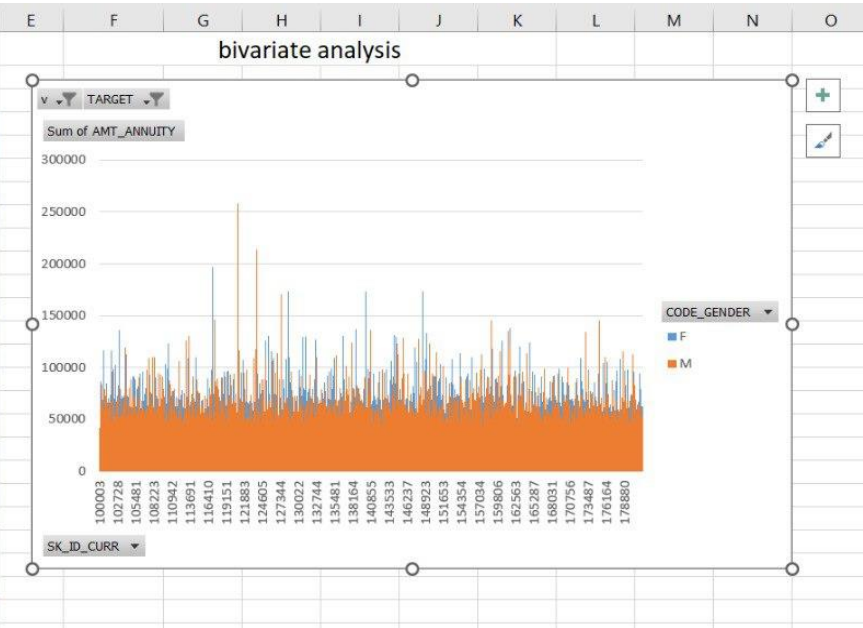
- Segmented variate analysis

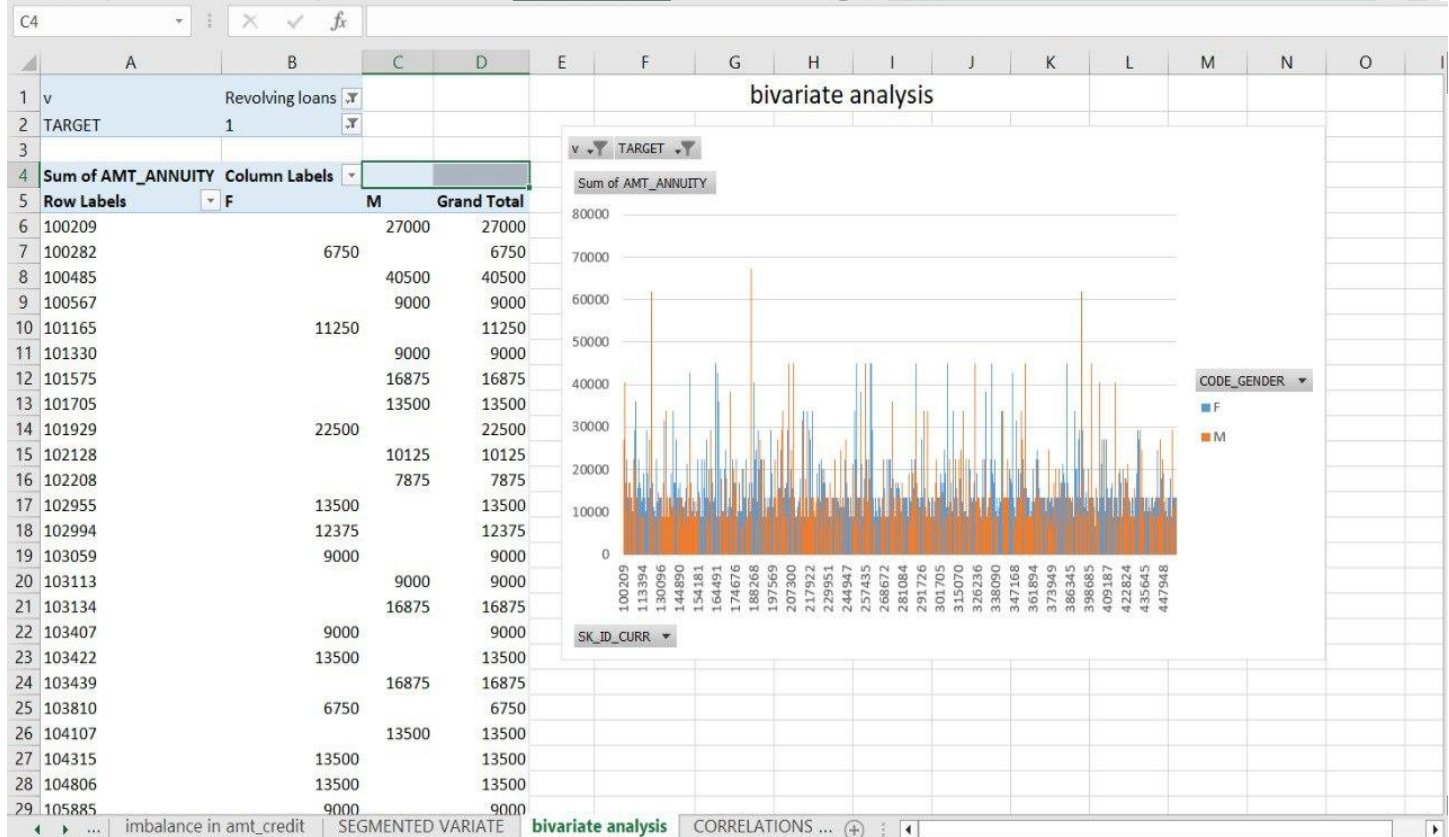
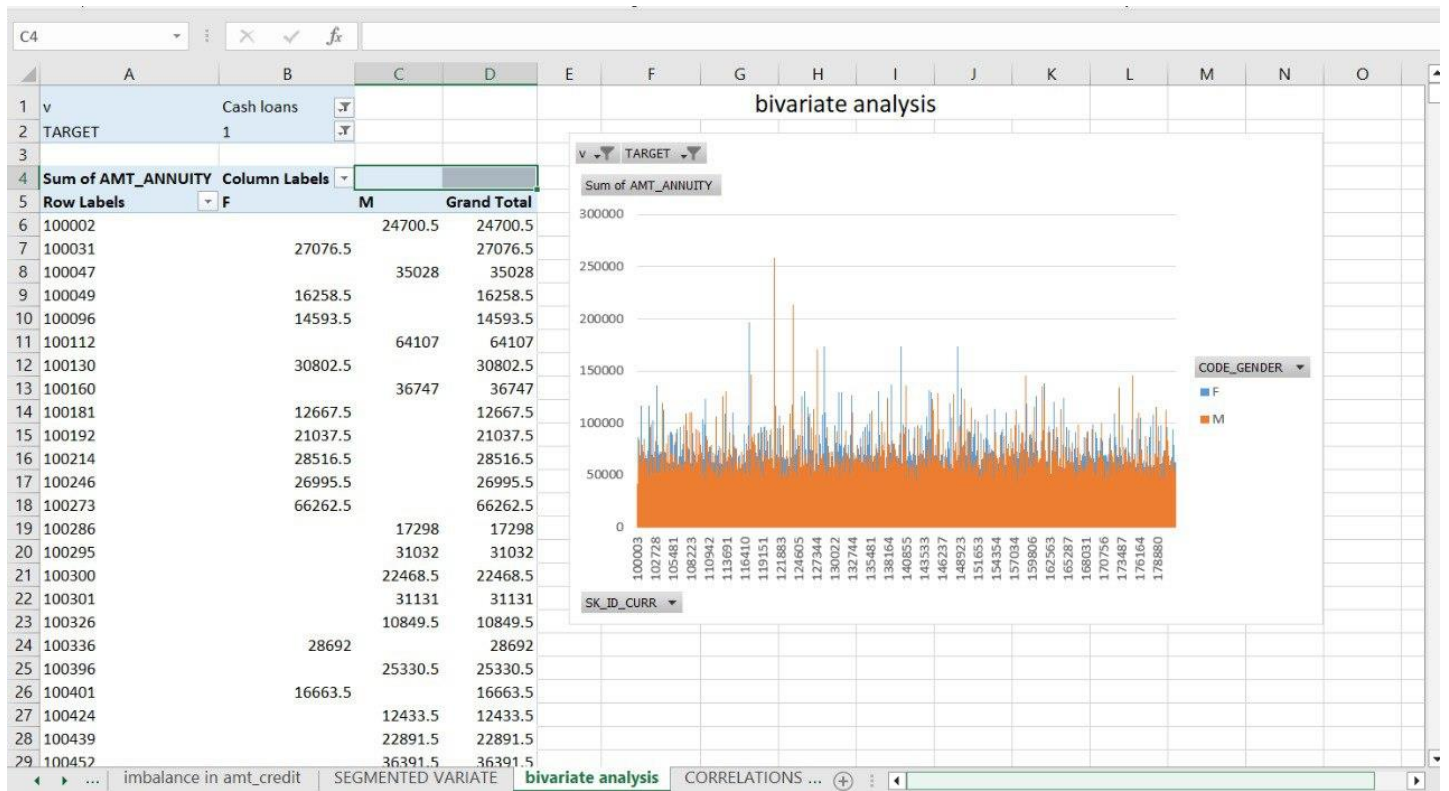
Count of AMT_CREDIT	
Row Labels	
Cash loans	278232
0	255011
1	23221
Revolving loans	29279
0	27675
1	1604
(blank)	
(blank)	
Grand Total	307511

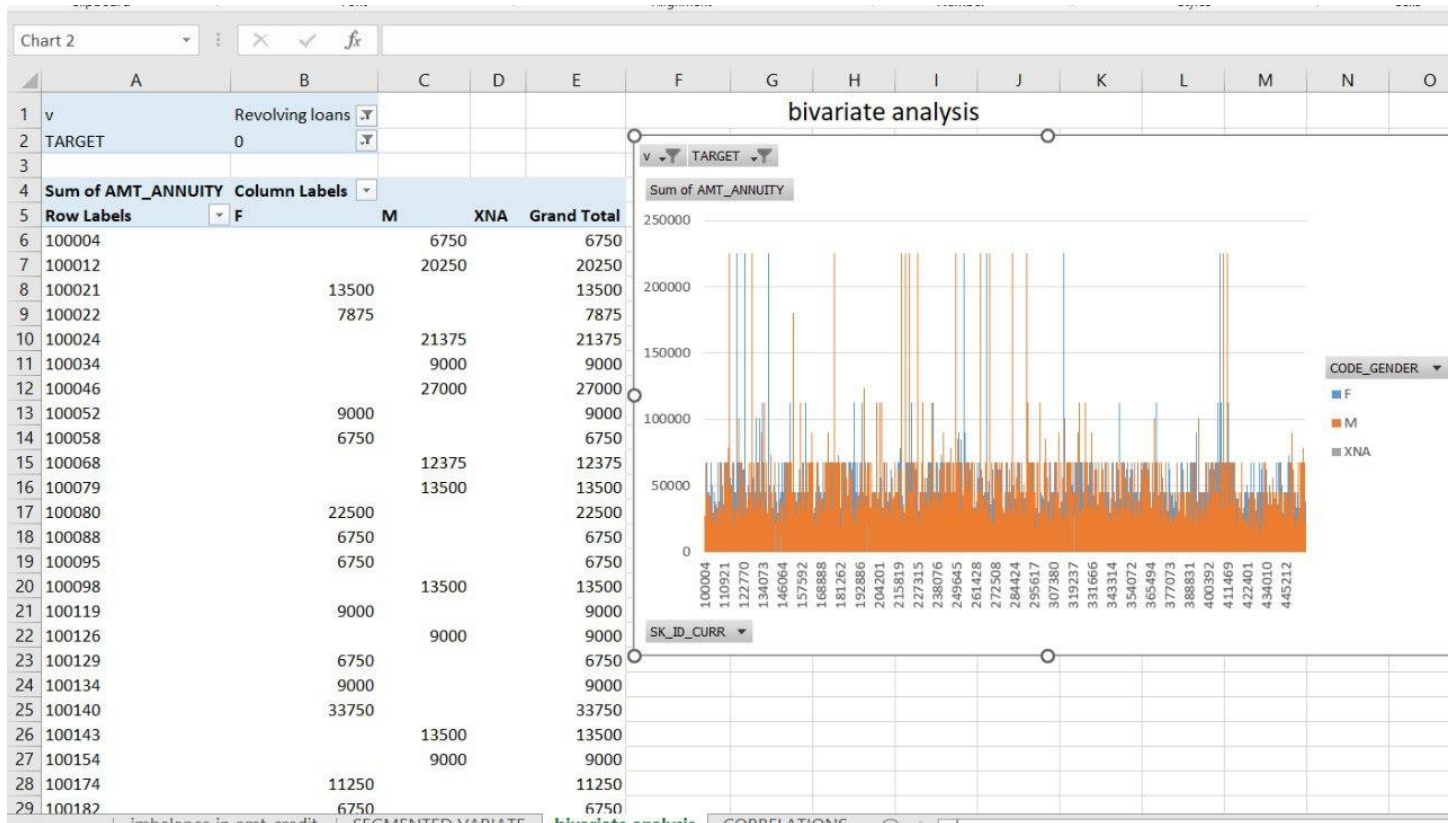


- Bivariate analysis

bivariate analysis															
1	Sum of AMT_ANNUI	Column Labels													
2	Row Labels	F	M												
3	100003	35698.5		35698.5											
4	100006	29686.5		29686.5											
5	100007		21865.5	21865.5											
6	100008		27517.5	27517.5											
7	100009	41301		41301											
8	100010		42075	42075											
9	100011	33826.5		33826.5											
10	100014	21177		21177											
11	100015	10678.5		10678.5											
12	100016	5881.5		5881.5											
13	100017		28966.5	28966.5											
14	100018	32778		32778											
15	100019		20160	20160											
16	100020		26149.5	26149.5											
17	100023	17563.5		17563.5											
18	100025	37561.5		37561.5											
19	100026	32521.5		32521.5											
20	100027	23850		23850											
21	100029		12703.5	12703.5											
22	100030	11074.5		11074.5											

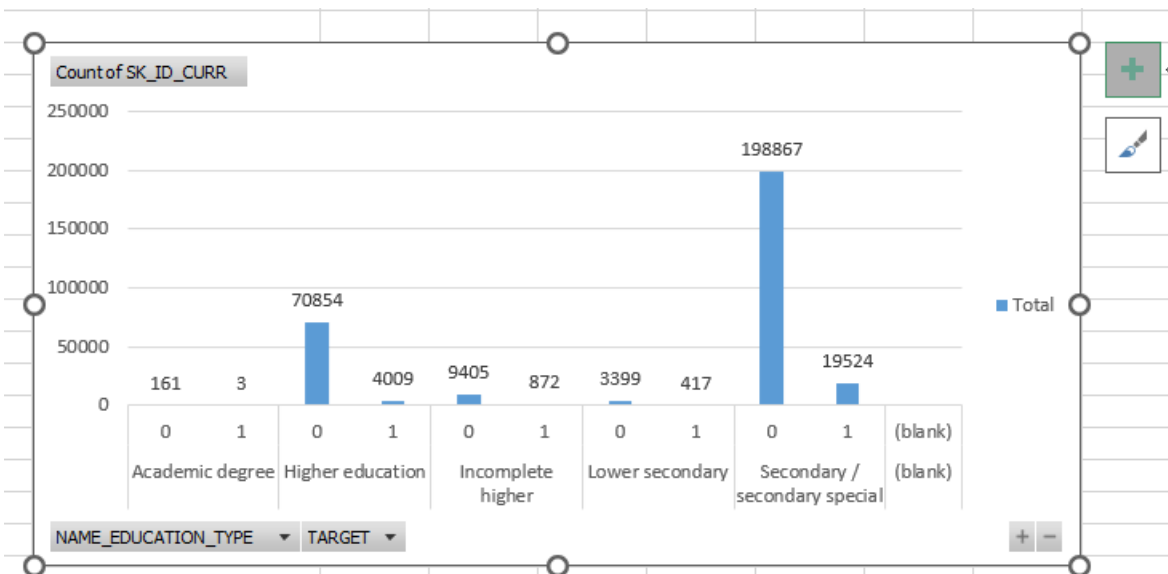


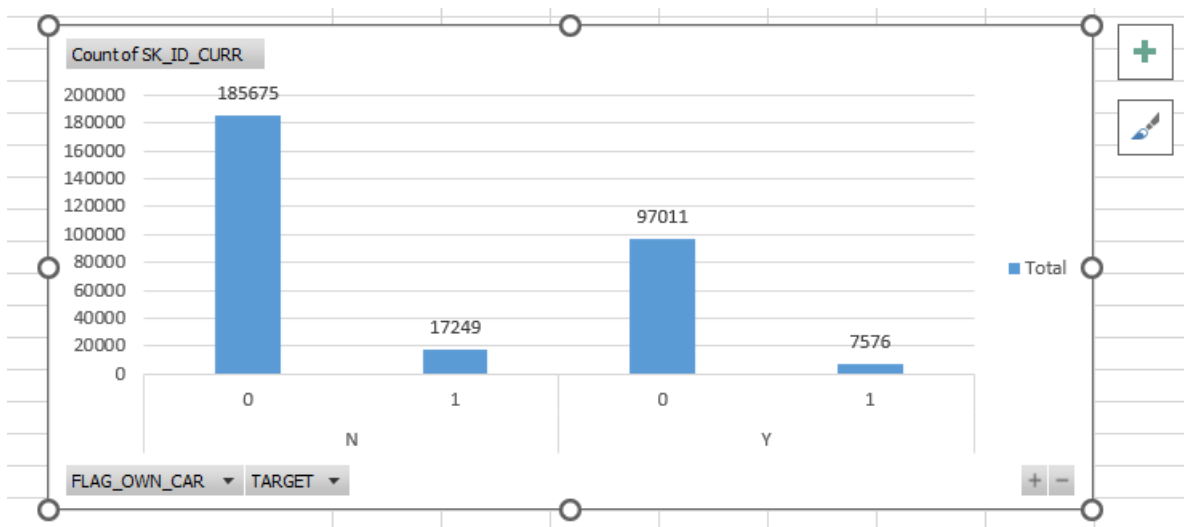
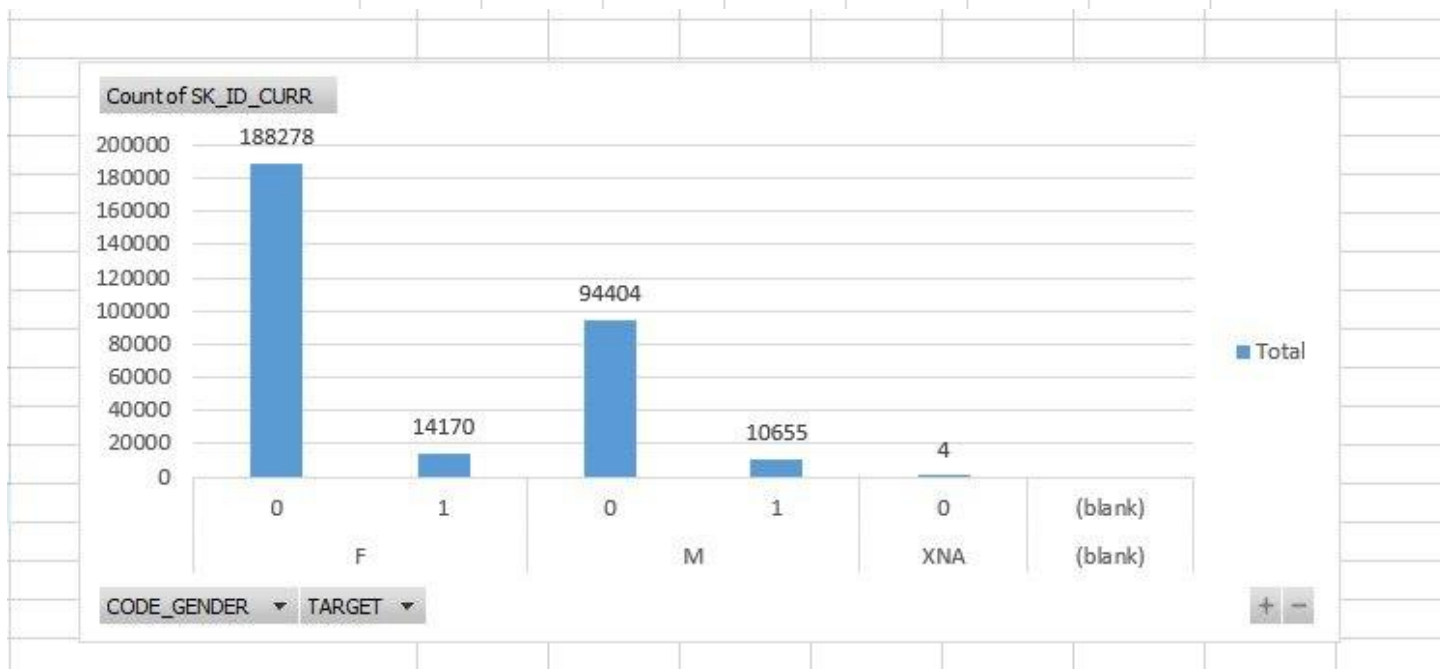
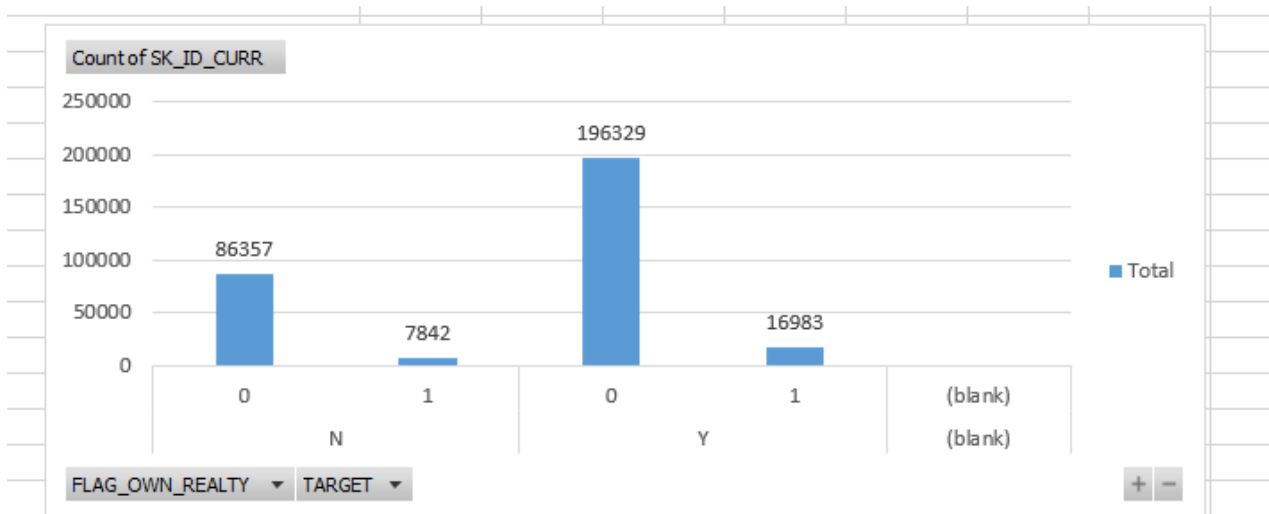


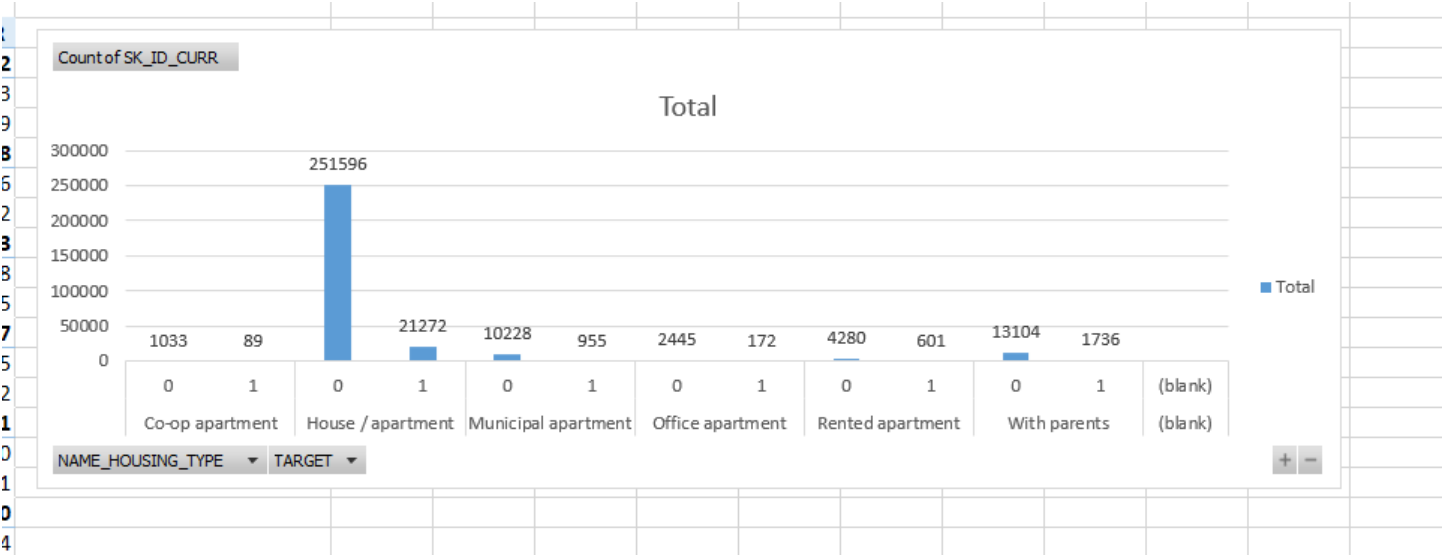
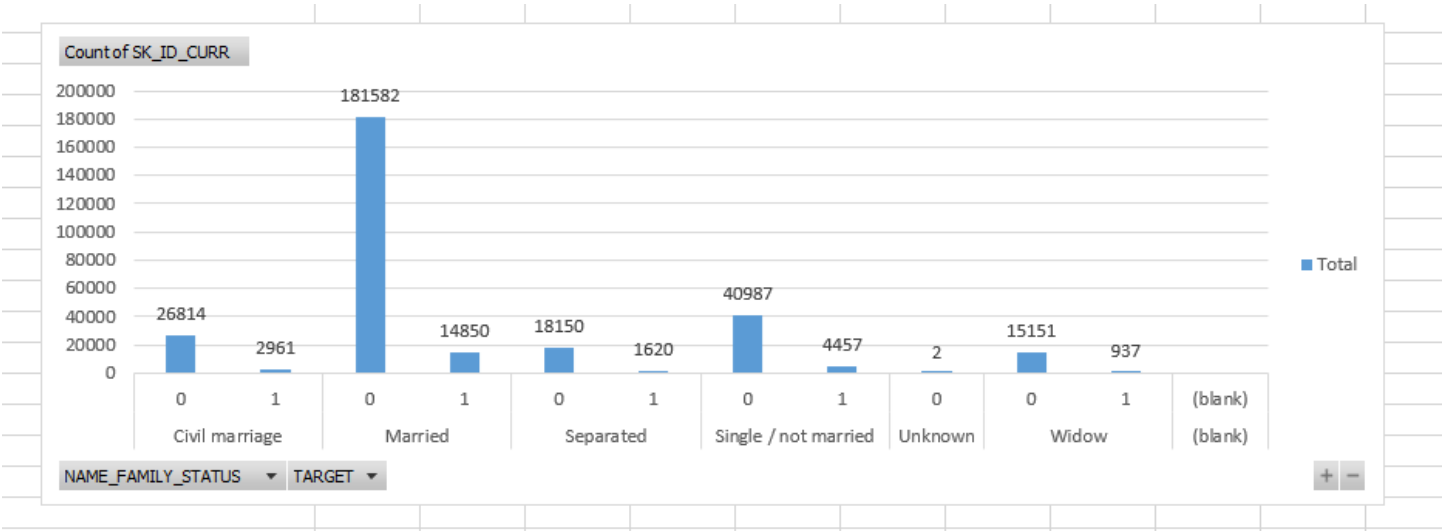


Find the top 10 correlations for Client with payment difficulties and all other case.

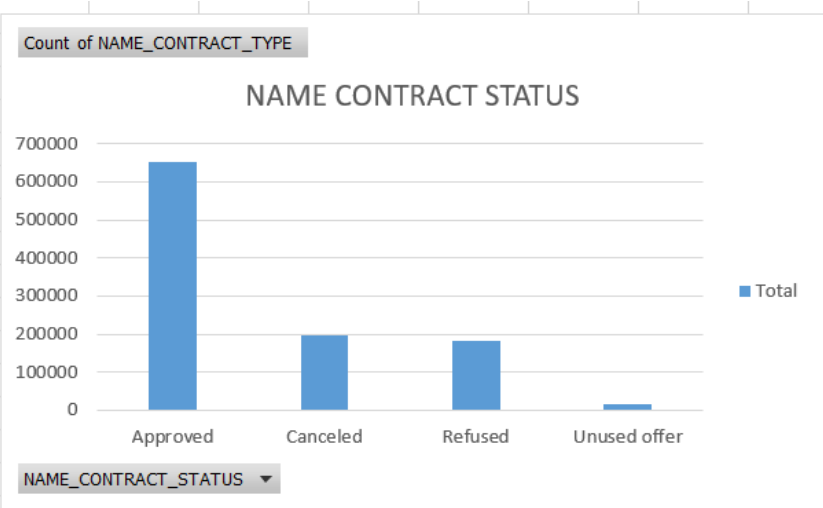
Solution:



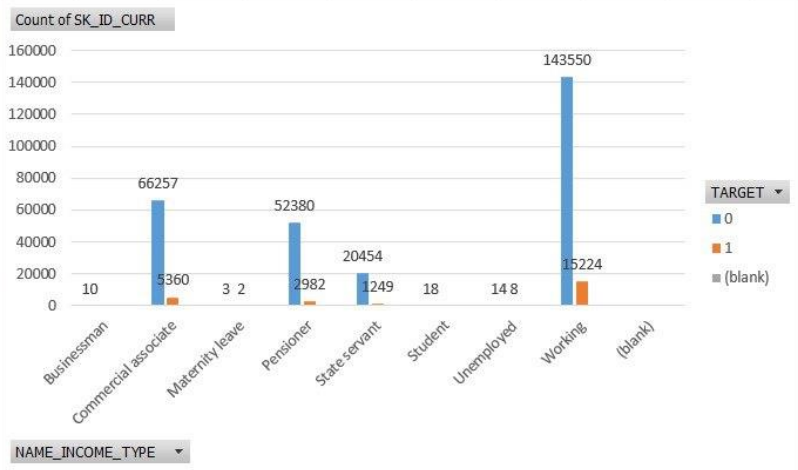




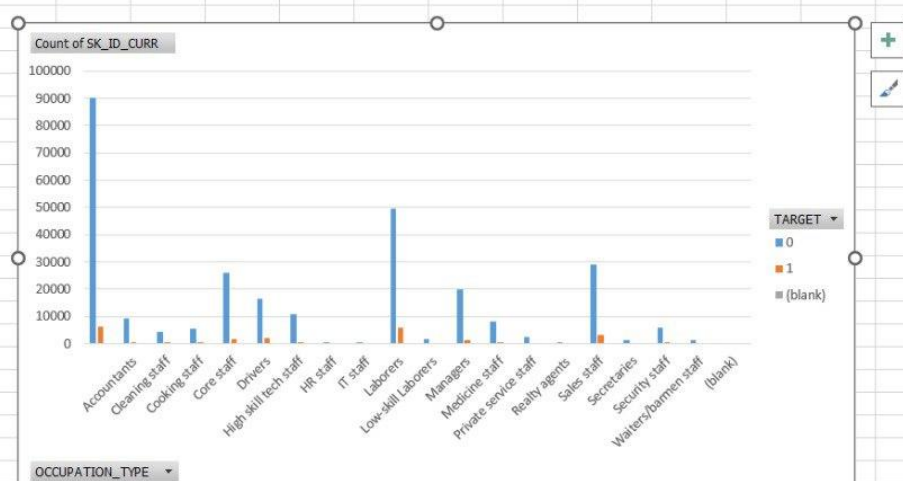
Row Labels	Count of NAME_CONTRACT_TYPE
Approved	652486
Canceled	197231
Refused	182083
Unused offer	16775
Grand Total	1048575



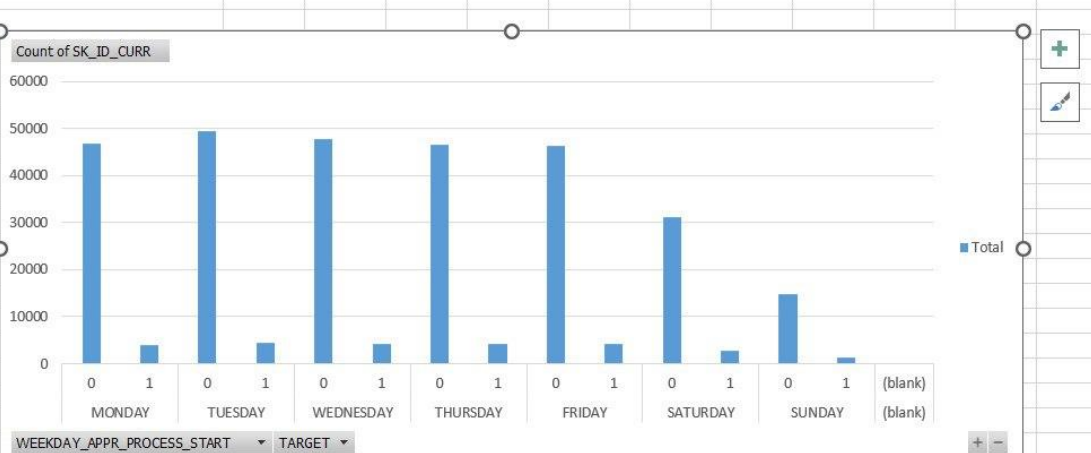
Count of SK_ID_CURR	Column Labels	
Row Labels	0	1 (blank) Grand Total
Businessman	10	10
Commercial associate	66257	5360
Maternity leave	3	2
Pensioner	52380	2982
State servant	20454	1249
Student	18	18
Unemployed	14	8
Working	143550	15224
(blank)		
Grand Total	282686	24825



Count of SK_ID_CURR	Column Labels	
Row Labels	0	1 (blank) Grand Total
Accountants	90113	6278
Cleaning staff	9339	474
Cooking staff	4206	447
Core staff	5325	621
Drivers	25832	1738
High skill tech staff	16496	2107
HR staff	10679	701
IT staff	527	36
Laborers	492	34
Low-skill Laborers	49348	5838
Managers	1734	359
Medicine staff	20043	1328
Private service staff	7965	572
Realty agents	2477	175
Sales staff	692	59
Secretaries	29010	3092
Security staff	1213	92
Walters/barmen staff	5999	722
(blank)	1196	152
Grand Total	303696	24032



Row Labels	Count of SK_ID_CURR
MONDAY	50714
0	46780
1	3934
TUESDAY	53901
0	49400
1	4501
WEDNESDAY	51934
0	47696
1	4238
THURSDAY	50591
0	46493
1	4098
FRIDAY	50338
0	46237
1	4101
SATURDAY	33852
0	31182
1	2670
SUNDAY	16181
0	14898
1	1283
(blank)	
(blank)	
Grand Total	307511



Conclusion

With the help of analysis, we came to know about portion of loan taker will repay the loan or not (defaulters) and the reasons behind being default. The people who are facing difficulty in loan payment are laborers or either who have completed secondary education or who have age above certain yrs. and the peoples who opted for cash loan.

Factor for deciding the defaulters:

- Men are at relatively higher default rate
- People who have civil marriage or who are single have greater chance to default.
- People with Lower Secondary & Secondary education.
- Clients who are either at Maternity leave OR Unemployed default a lot.
- People who live in Rating 3 has highest defaults.
- Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.
- Avoid young people who are in age group of 20-40 as they have higher probability of defaulting.
- People who have less than 5 years of employment have high default rate.
- Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.
- When the credit amount goes beyond 3lakhs, there is an increase in defaulter.
- Rating 1 clients have less chance of getting default.
- Peoples with 3 or more children have greater risk.