Drive Link for the project-
https://drive.google.com/drive/u/0/folders/1x-1--FVNwEKp7fOmcoAmgW9Iumv8tYhE

Presentation Link for the project-
https://www.loom.com/share/e451eebee6f14e70aed6fe15ed9596a6

Tableau Workbook for the project-
https://public.tableau.com/app/profile/mohd.faiz2106/viz/BankLoanCaseStudy_16813027062140/No_ofApplicationvsPaymentTypevsStatus?publish=yes

Project by –

Mohd Faiz

faiz.mohd340@gmail.com

Bank Loan Case Study

Project Description

This project will go through to find some of the insights on Bank Loan Case Study.

Project Approach

In order to find the insights, Excel and Tableau were used. Using Excel, we performed Data Cleaning and transformation first like understanding data columns, checking for missing data, checking and removing outliers, etc. After that, for Exploratory Analysis we used Tableau to get the insights we needed.

Tech Stack Used

Microsoft Excel 2021, Tableau, Microsoft Word 2021, and Google Drive.

Project Insight

1.Cleaning the data

At first, removed all the rows which were empty. Found out the number of blank cells in the particular column. After that, we find the percentage of the null values.

To find the blank values we used COUNTBLANK function in Excel.

We removed those columns whose null value percentage for that particular column is greater than **5%** for "application_data" dataset and **20**% and above for "previous_application" dataset.

A	B	C	D	E	F	G	H	I	J	K	L	M
Percentage of null values	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Total null row count	0	0	0	0	0	0	0	0	0	12	278	1292
	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	NAME_TYPE_SUITE
Total row count	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied
307511	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family
	100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied
	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied
	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied
	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner
	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied
	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied
	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children
	100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied
	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied
	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children
	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied
	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied
	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied
	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family
	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied
	100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied
	100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A
	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied
	100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied
	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied
	100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied
	100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied
	100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied
	100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied

application_data- EDA | Final Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
Percentage of null values	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
Total null row count	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	SK_ID_PREV	SK_ID_CURR	NAME_CO	AMT_APPLICATION	AMT_CREI	WEEKDAY_	HOUR_AP	FLAG_LAS	NFLAG_LA	NAME_CA	NAME_CO	DAYS_DEC	NAME_PA	CODE_REJ	NAME_CLI	NAME_GC	NAME_PO	NAME_PR	CHANNEL_	SELLERPLA	NA
Total row count	2030495	271877	Consumer	17145	17145	SATURDAY	15	Y	1	XAP	Approved	-73	Cash throu	XAP	Repeater	Mobile	POS	XNA	Country-w	35	Co
1048573	2802425	108129	Cash loans	607500	679671	THURSDAY	11	Y	1	XNA	Approved	-164	XNA	XAP	Repeater	XNA	Cash	x-sell	Contact ce	-1	XN
	2523466	122040	Cash loans	112500	136445	TUESDAY	11	Y	1	XNA	Approved	-301	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	2819243	176158	Cash loans	450000	470790	MONDAY	7	Y	1	XNA	Approved	-512	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	1784265	202054	Cash loans	337500	404055	THURSDAY	9	Y	1	Repairs	Refused	-781	Cash throu	HC	Repeater	XNA	Cash	walk-in	Credit and	-1	XN
	1383531	199383	Cash loans	315000	340574	SATURDAY	8	Y	1	Everyday e	Approved	-684	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	2315218	175704	Cash loans	0	0	TUESDAY	11	Y	1	XNA	Canceled	-14	XNA	XAP	Repeater	XNA	XNA	XNA	Credit and	-1	XN
	1656711	296299	Cash loans	0	0	MONDAY	7	Y	1	XNA	Canceled	-21	XNA	XAP	Repeater	XNA	XNA	XNA	Credit and	-1	XN
	2367563	342292	Cash loans	0	0	MONDAY	15	Y	1	XNA	Canceled	-386	XNA	XAP	Repeater	XNA	XNA	XNA	Credit and	-1	XN
	2579447	334349	Cash loans	0	0	SATURDAY	15	Y	1	XNA	Canceled	-57	XNA	XAP	Repeater	XNA	XNA	XNA	Credit and	-1	XN
	1715995	447712	Cash loans	270000	335754	FRIDAY	7	Y	1	XNA	Approved	-735	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	2257824	161140	Cash loans	211500	246398	FRIDAY	10	Y	1	XNA	Approved	-815	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	2330894	258628	Cash loans	148500	174362	TUESDAY	15	Y	1	XNA	Approved	-860	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	1397919	321676	Consumer	53779.5	57564	SUNDAY	15	Y	1	XAP	Approved	-408	Cash throu	XAP	New	Consumer	POS	XNA	Country-w	200	Co
	2273188	270658	Consumer	26550	27252	SATURDAY	10	Y	1	XAP	Approved	-726	Cash throu	XAP	New	Constructi	POS	XNA	Stone	83	Co
	1232483	151612	Consumer	126490.5	119853	TUESDAY	7	Y	1	XAP	Approved	-699	Cash throu	XAP	New	Auto Acce	POS	XNA	Regional /	130	Inc
	2163253	154602	Consumer	26955	27297	SATURDAY	12	Y	1	XAP	Approved	-1473	Cash throu	XAP	Repeater	Photo / Ci	POS	XNA	Stone	130	Co
	1285768	142748	Revolving	180000	180000	FRIDAY	13	Y	1	XAP	Approved	-336	XNA	XAP	Repeater	XNA	Cards	x-sell	AP+ (Cash	6	XN
	2393109	396305	Cash loans	180000	180000	THURSDAY	14	Y	1	XNA	Approved	-700	Cash throu	XAP	Repeater	XNA	Cash	x-sell	AP+ (Cash	6	XN
	1173070	199178	Cash loans	45000	49455	SATURDAY	16	Y	1	Everyday e	Refused	-584	XNA	HC	Repeater	XNA	Cash	walk-in	AP+ (Cash	6	XN
	1172842	302212	Cash loans	0	0	TUESDAY	9	Y	1	XNA	Refused	-239	XNA	HC	Repeater	XNA	XNA	XNA	Credit and	-1	XN
	1172937	302212	Cash loans	1129500	1277105	THURSDAY	5	Y	1	XNA	Refused	-594	Cash throu	HC	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	1555330	199353	Cash loans	0	0	SATURDAY	6	Y	1	XNA	Canceled	-202	XNA	XAP	Repeater	XNA	XNA	XNA	Credit and	-1	XN
	1543131	275707	Cash loans	229500	241920	THURSDAY	8	Y	1	XNA	Approved	-370	Cash throu	XAP	Repeater	XNA	Cash	x-sell	Credit and	-1	XN
	2536650	338725	Cash loans	369000	369000	WEDNESD.	13	Y	1	XNA	Approved	-1487	XNA	XAP	Repeater	XNA	Cash	x-sell	Country-w	-1	Co
	1676258	433469	Cash loans	247500	268083	THURSDAY	14	Y	1	XNA	Approved	-1883	XNA	XAP	Repeater	XNA	Cash	x-sell	Country-w	-1	Co
	2075578	418383	Consumer	74610	65610	MONDAY	14	Y	1	XAP	Approved	-2702	Cash throu	XAP	New	Mobile	POS	XNA	Stone	61	Co
	1583704	315664	Cash loans	0	0	WEDNESD.	15	Y	1	XNA	Refused	-430	XNA	HC	Repeater	XNA	XNA	XNA	Credit and	-1	XN

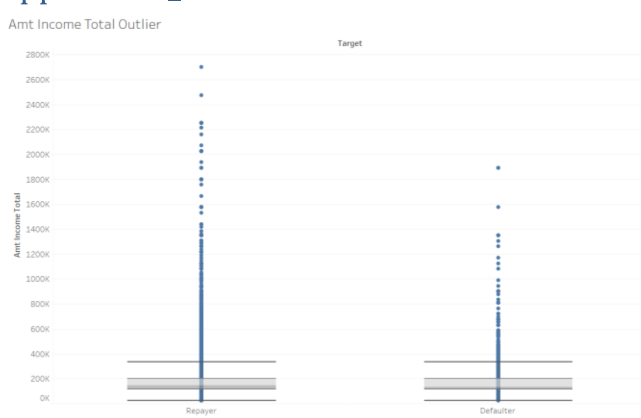
previous_application | Previous appl- Final Dataset

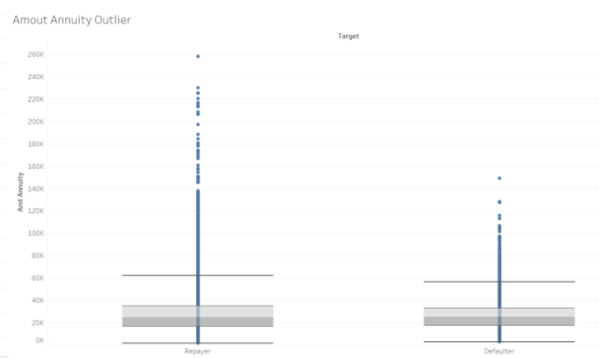
We also removed some unnecessary columns from both datasets which we were not going to use in our analysis.

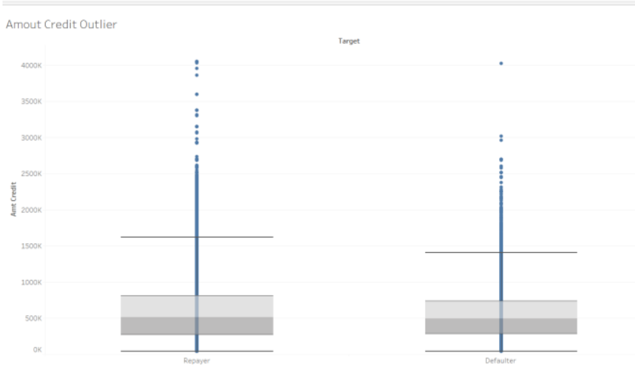
2. Finding Outliers

We box-plotted numerical columns to find out the outliers in both datasets.

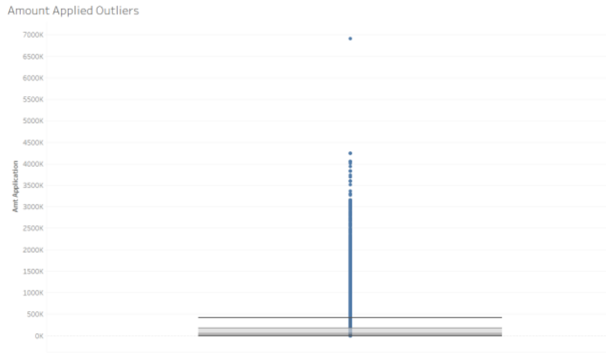
application_data dataset Outliers-

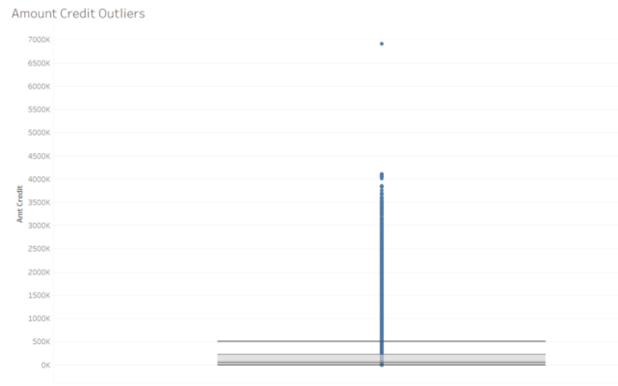






previous_application dataset Outliers-

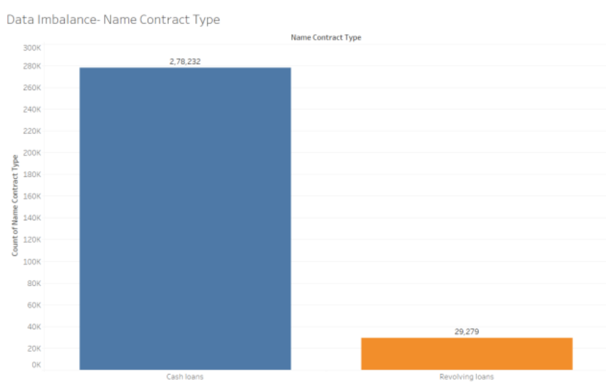


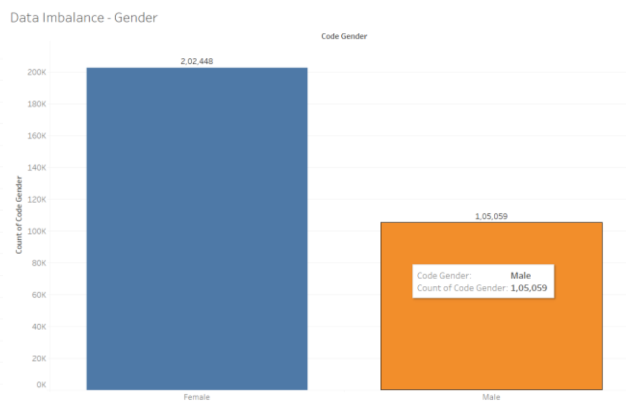


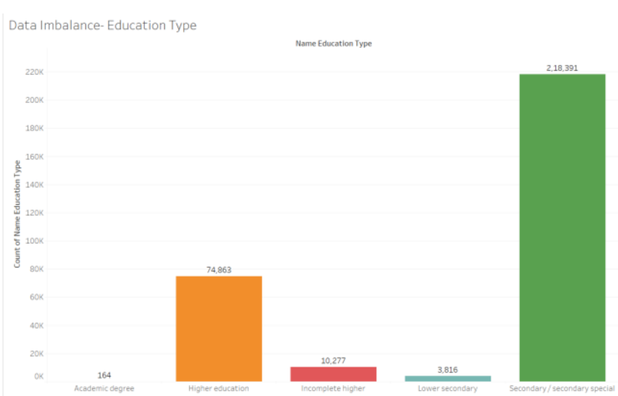
3. Data Imbalance

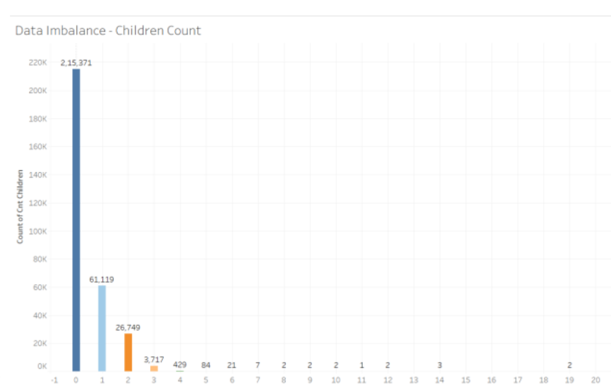
When data is distributed into unequal manner then data imbalance occurs. I plotted data imbalance using Bar/Column Charts in Tableau.

application_data Dataset data imbalance-

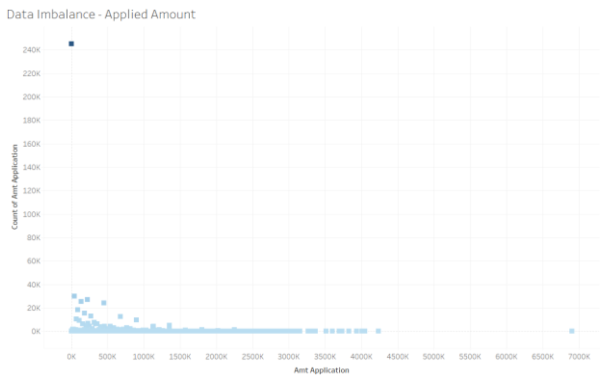


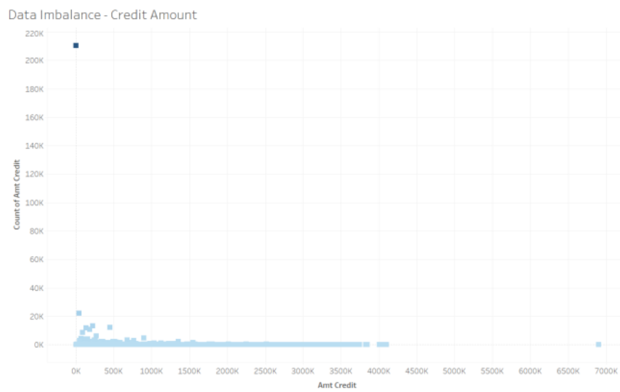


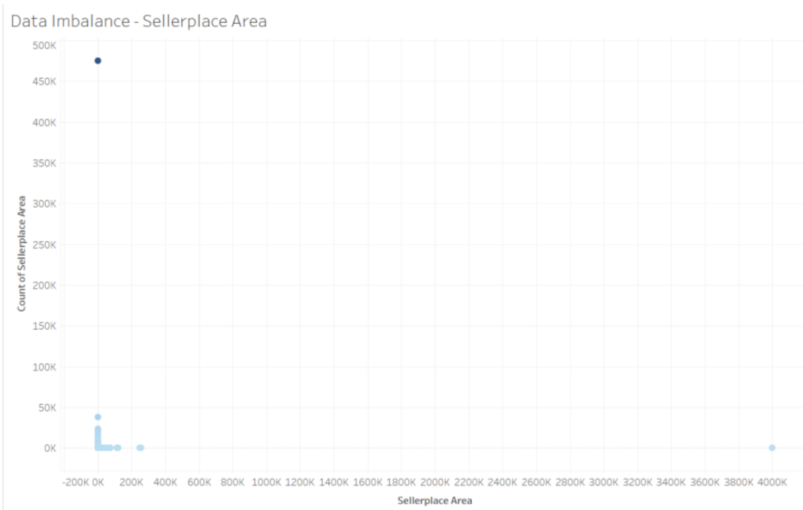




previous_application dataset data imbalance-





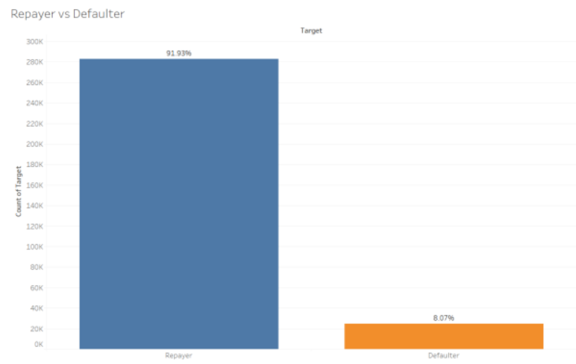


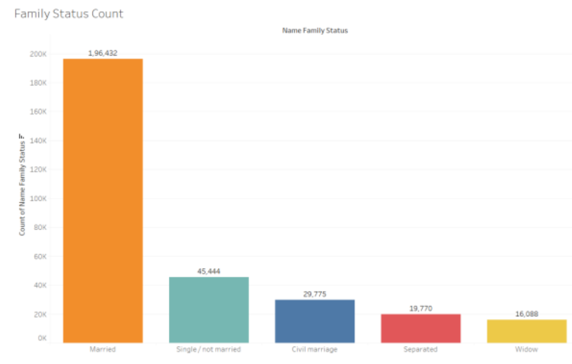
4. Univariate Analysis

application_data Dataset univariate analysis-

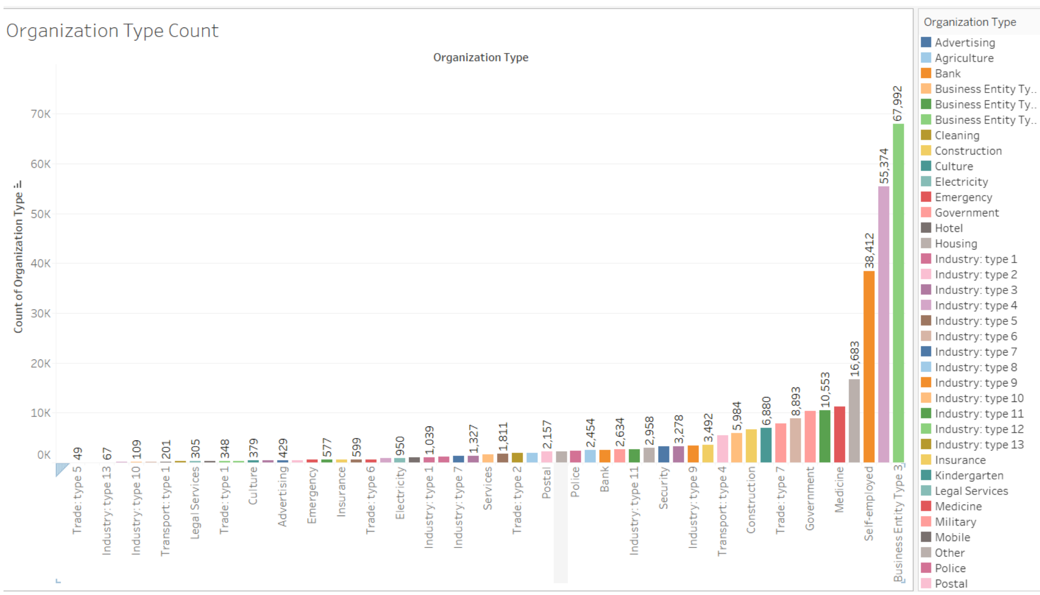
Insights-

- There are more number of Repayers than defaulters.
- Married people have requested more loans.





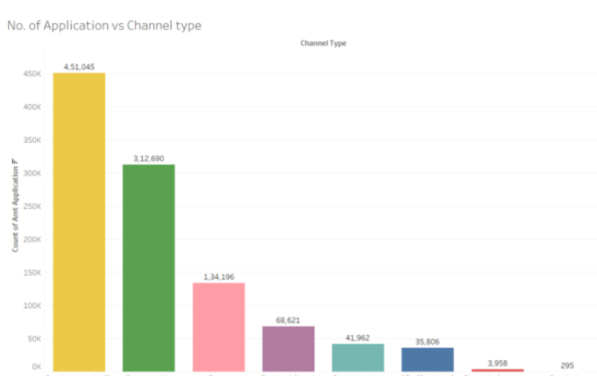
- Maximum No. of loan applications are from Business Type-3 and Industry Type-3 whereas the minimum No. of loan applications are from Trade-Type-5 and Industry Type-13.

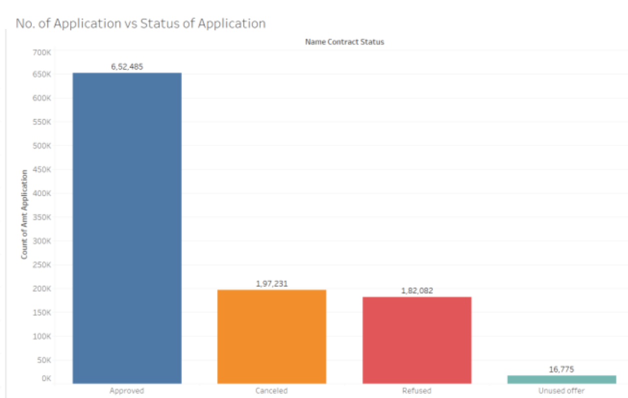


previous_application dataset univariate analysis-

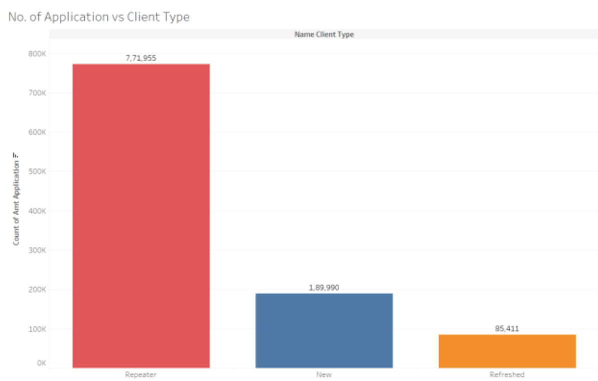
Insights-

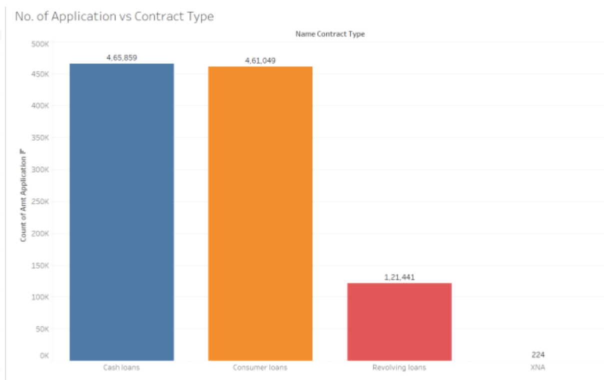
- Credit and Cash Offices have more number of loan applications than any other group.





- Repeaters are more likely to apply for loans
- Cash loans and Consumers loans are preferred loan type among all.



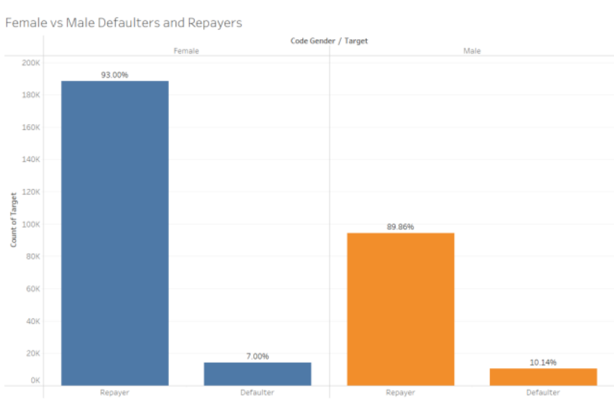


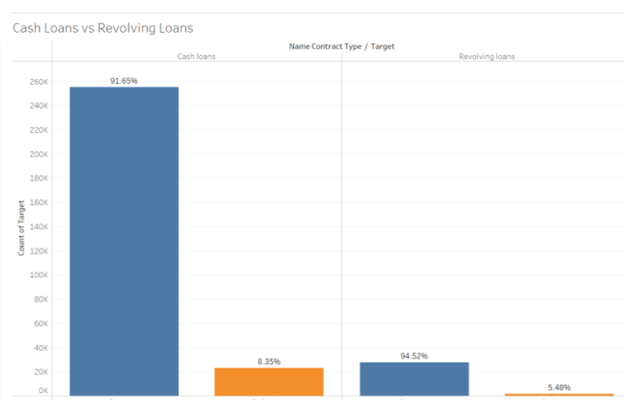
5. Bivariate Analysis

application_data Dataset bivariate analysis-

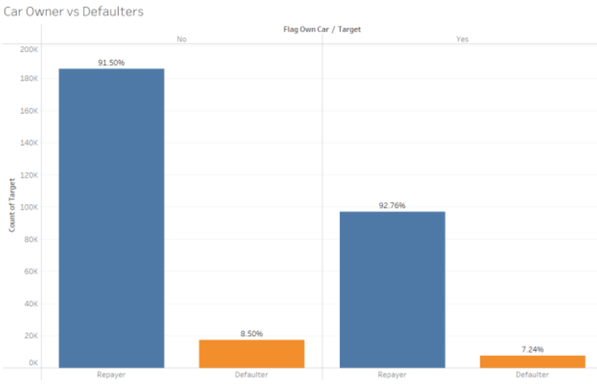
Insights-

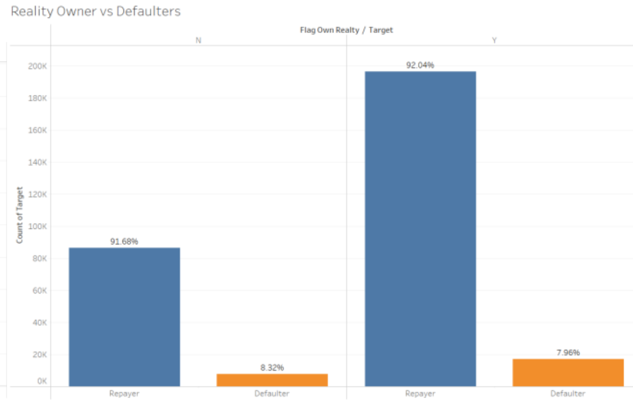
- Females took out more loans and were comparatively less likely to default than males.
- There are a greater number of cash loan applications and they are more likely to be defaulters as well.



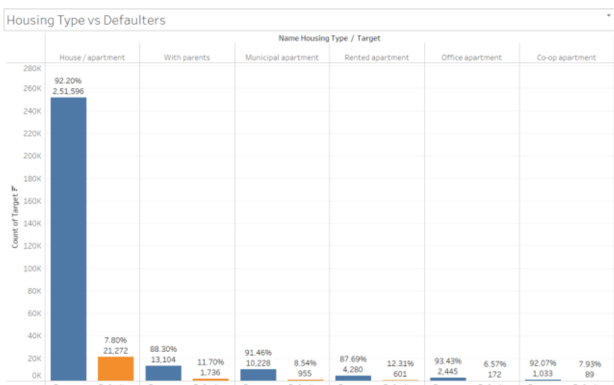


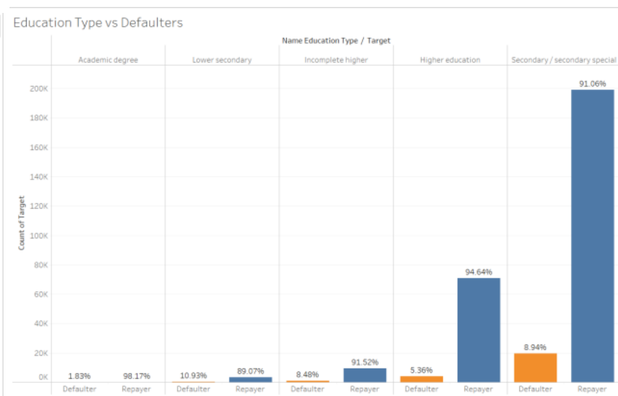
- People with no cars are much more likely to take loans and become defaulters as well.
- Reality Owners are much more likely to take loans and be good re-payers as well.



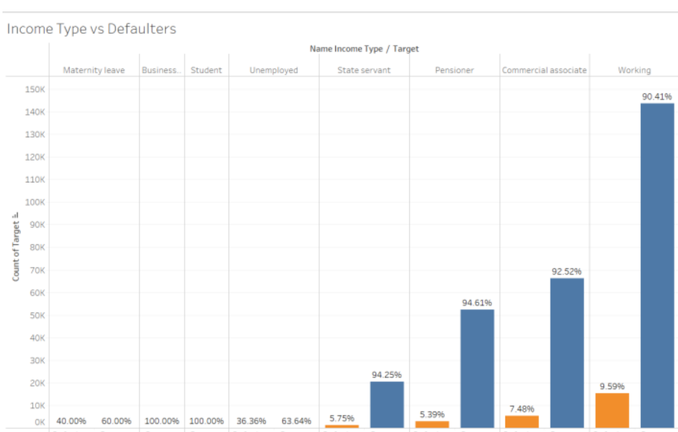


- Among all the Housing types, House/Apartments type receive more loan applications and Rented Apartment Owners are more likely to become defaulters.
- Secondary/Secondary Special has more number of applications and the lower secondary class has more defaulters.

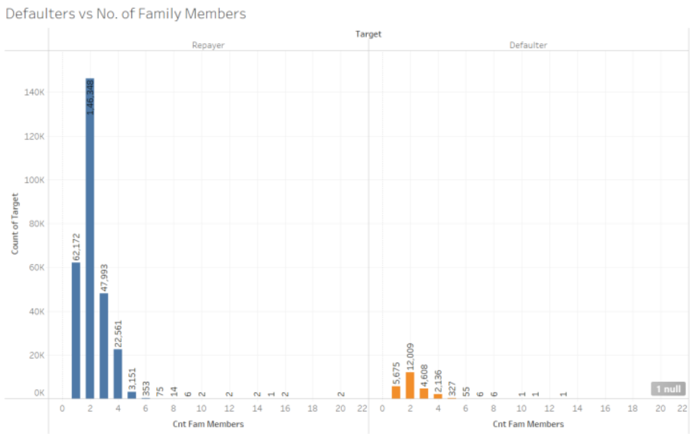


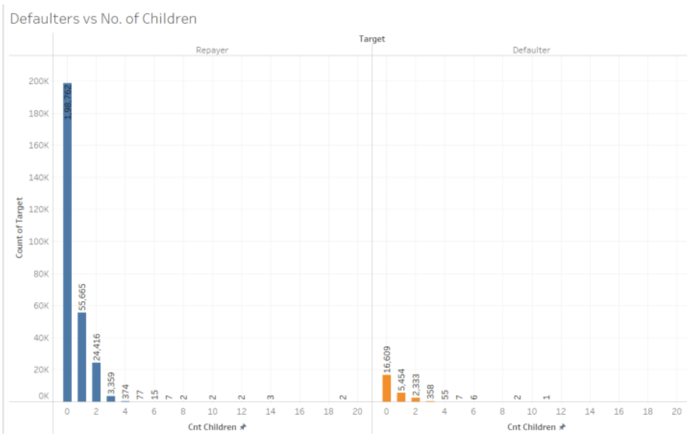


- Working-class people are more likely to take loans and maternity leave ladies are more likely to become defaulters.



- Customers with no. of family members greater than 4 are more likely to default.

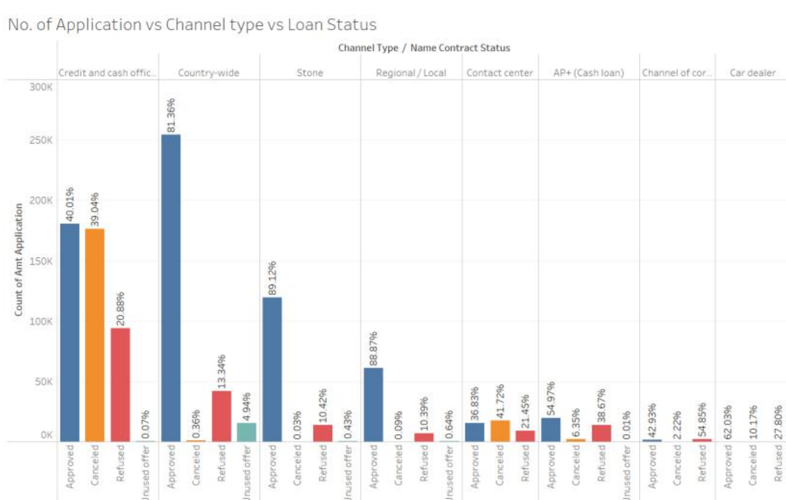




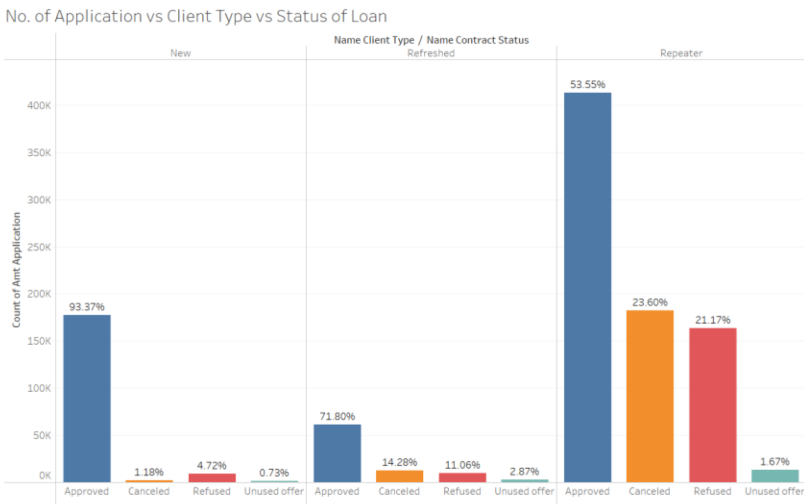
previous_application dataset bivariate analysis-

Insights-

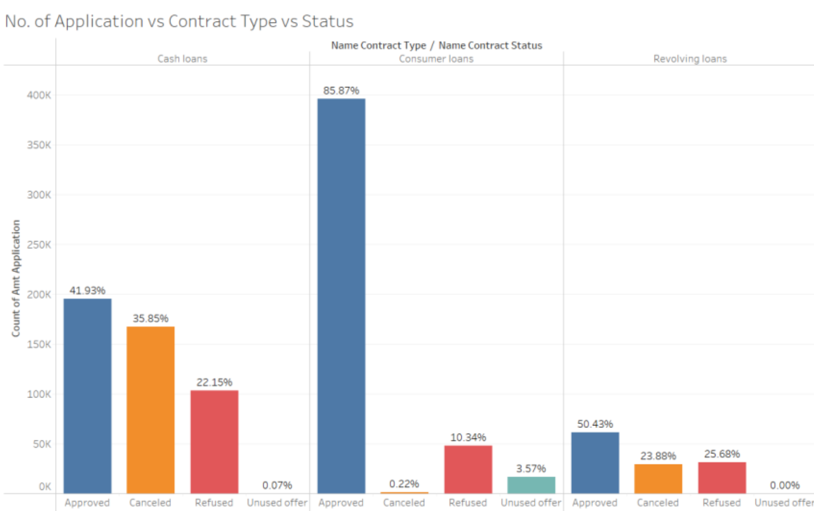
1. Country-wide and Stone are more likely to get their loan application approved whereas Cash loans and channel of correspondence applications are more likely to be refused.



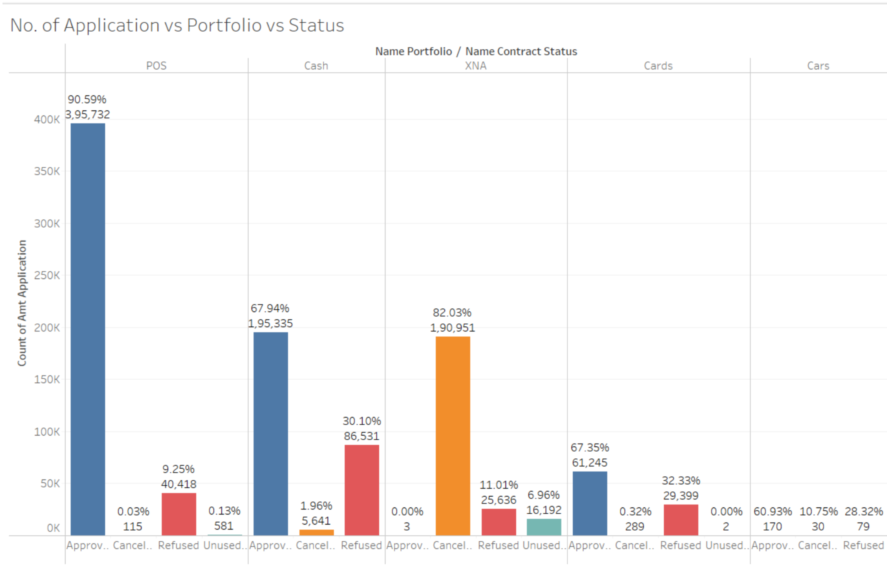
2. New applications are more likely to be accepted. Repeater's application sees the most rejections among all the categories.



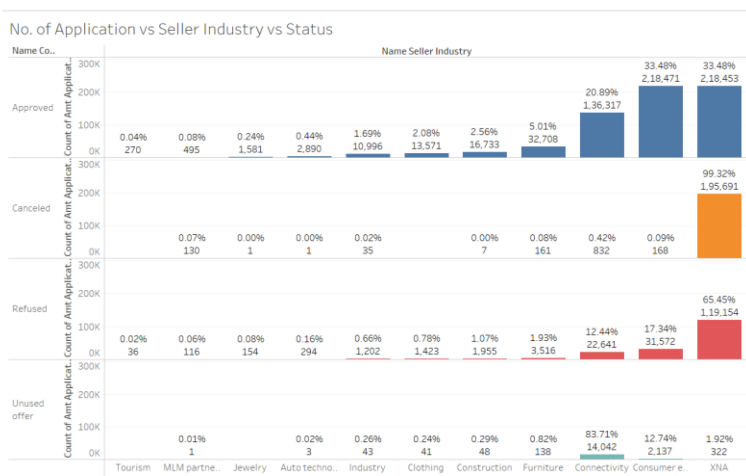
3. Consumer applications have nearly no cancellation and the greatest approval rate.



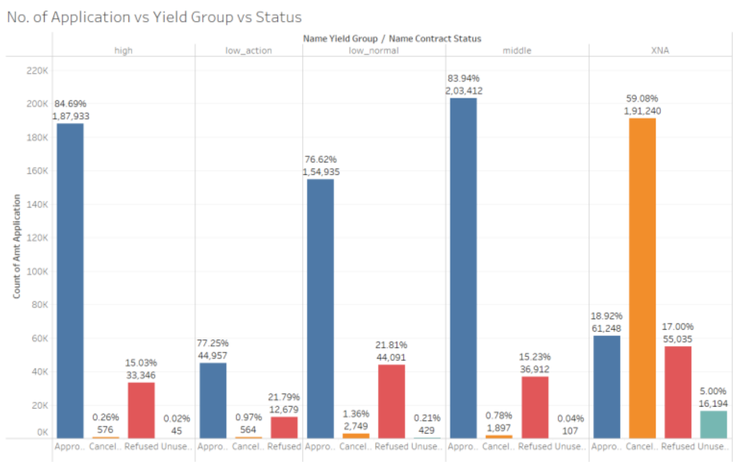
4. POS applications have the greatest approval rate and very low cancellation rate.



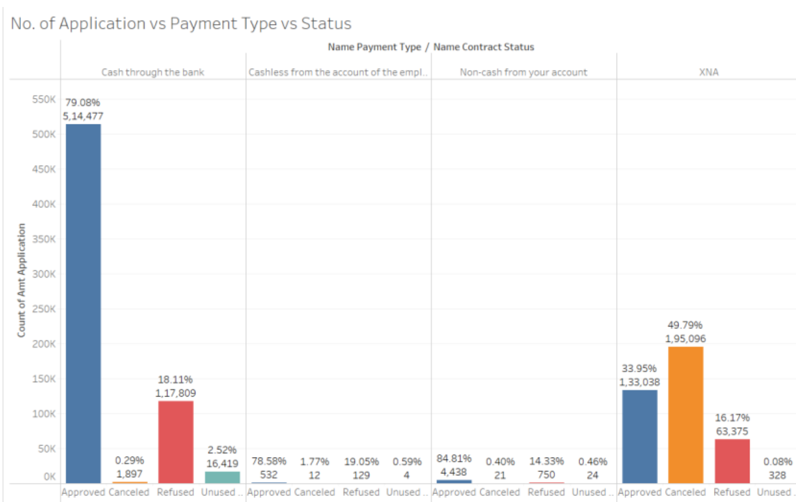
5. Consumer electronics industry have more approval rate of their application



6. High and middle class are more likely to get their application approved whereas low_normal and low_action classes have higher rejection rates among all.



7. Applications wanting cash through the bank are more likely to be approved.



6. Top Correlations

application_data Dataset-

- Family Status
- Industry Type
- Income Type
- Count of family members
- Count of Children
- Education Type
- Car Owner
- Realtor Owner

previous_application dataset-

- Channel type
- Seller Industry
- Client type
- Contract Type
- Payment Type
- Portfolio
- Yield Group

Summary of Insights

Loan Highly recommended group

- Married Clients
- More educated Clients
- Females
- House/Apartment Owners
- Working Class customers
- Customers with fewer family members (<=4)

Loan Highly risky group

- Rented Apartment Owner
- Less Educated Clients
- Car Owner
- Clients on maternity leave
- Customers with higher family members (>4)
- Unemployed customers

Project Conclusion

While analyzing the data set provided, several meaningful insights were discovered that could not have been discovered by manually searching the dataset for insights.

We could also leverage the Excel-2021 and Tableau tool and got a little more experienced in using the tools and also injecting different formulas and pivot tables and graphs to look for insights.