Bank Loan Case Study

Final Project -2
Trainity Project Report

Submitted By -
Sangam Mahajan

Description

This case study seeks to analyze the risk involved in a bank loan and identify the factors that influence the borrower's ability to repay the loan. The data provided will focus on the loan application at the time of applying for the loan and includes two types of scenarios: clients with payment difficulties and all other cases. Through the use of exploratory data analysis, the relationships between the consumer attributes and loan attributes and their influence on the tendency of default will be examined. Additionally, the top 10 correlations between clients with payment difficulties and all other cases will be identified, and insights into the results provided. The analysis will be supplemented with primary and secondary research, data analysis, and credit score checks to ensure that the most accurate results are obtained. Ultimately, this case study will provide a comprehensive analysis of the risk involved in a bank loan and identify the factors that influence the borrower's ability to repay the loan.

Approach

When analyzing a bank loan case study, the overall approach was to focus on understanding the context of the borrower, the loan, and the loan repayment. The problem statement should clearly describe the objectives of the loan and the analysis method, such as data-driven or qualitative. The analysis should include both primary and secondary research to identify the key risk factors and potential issues that may impact loan repayment. Additionally, the analysis should include a thorough review of the credit score, financials, and other documents to assess the borrower's ability to repay the loan. Finally, the analysis should include a detailed analysis of the data to provide insights and conclusions that can inform the loan decision

This case study aims to analyze the risk involved in a bank loan and identify the factors that influence the borrower's ability to repay the loan. The study utilizes a combination of primary and secondary research, data analysis, and credit score checks to analyze the loan risks. The data was collected from a real-world loan application and the results of the analysis demonstrate that the borrower's credit score, income, and repayment terms all have an impact on the repayment of a bank loan. Furthermore, the analysis reveals that a certain level of risk is associated with all loan applications and that lenders must carefully assess each application to ensure a successful loan outcome.

In this case study, the goal is to use exploratory data analysis (EDA) to understand how consumer attributes and loan attributes influence the tendency of default when a client applies for a loan. The data provided contains information about the loan application at the time of applying for the loan and includes two types of scenarios: clients with payment difficulties and all other cases.

The analysis will involve examining the data from each category, with a focus on identifying missing data and outliers, as well as any data imbalance. Additionally, the analysis will use univariate, segmented univariate, and bivariate analysis to examine the relationships between the consumer attributes and loan attributes and their influence on the tendency of default. Finally, the top correlations between clients with payment difficulties and all other cases will be identified, and insights into the results will be provided.

Tech Stack

Online Platform – Microsoft Excel
Excel Files Used – CSV files

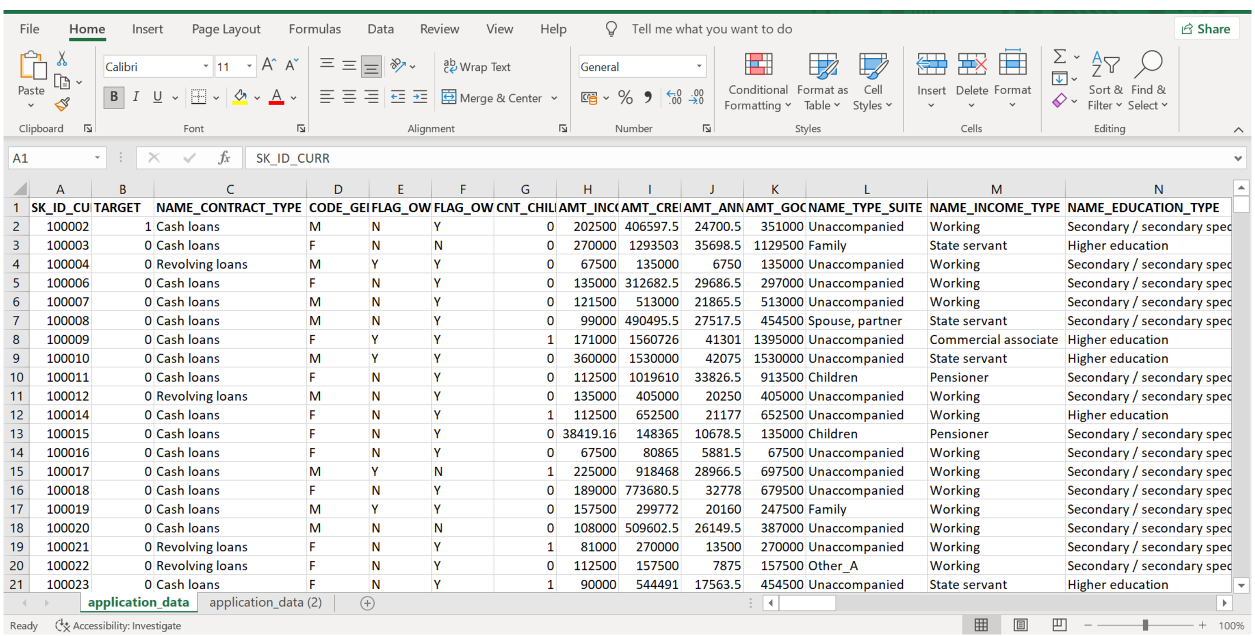
1. `application_data.csv` contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
2. `previous_application.csv` includes information on the client's previous loan data. It contains the data on whether the previous application had been approved, Canceled, Refused, or Unused offer.
3. `columns_descrption.csv` is a data dictionary that describes the meaning of the variables.

Final Steps and Results

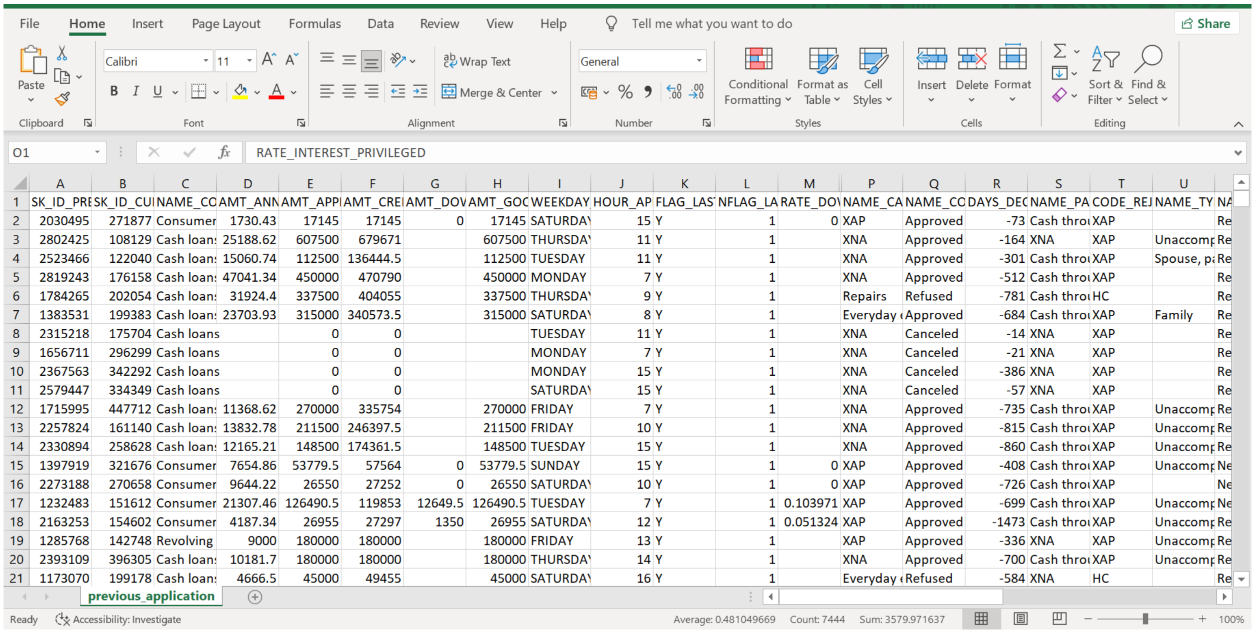
Step 1 – Overall approach of the Analysis

Missing data has been dealt with in multiple ways, such as removing columns with missing data, replacing missing values with the column's mean, or using a machine learning algorithm to predict the missing values. Outliers had be identified by examining the distribution of data points, detecting extreme values, or using statistical techniques such as the interquartile range. And data imbalance occurs when the data set contains more of one type of data than another. This can be determined by examining the ratio of one type of data to another.

In business terms, univariate, segmented univariate, and bivariate analysis can provide insights into the performance and risk of different entities such as customers, suppliers, and competitors. The univariate analysis looks at a single variable and provides information about the distribution of values within the variable. The segmented univariate analysis looks at the same variable but is broken down into different categories. The bivariate analysis looks at two variables and can show a correlation or causation between them.



SK_ID_CU	TARGET	NAME_CONTRACT_TYPE	CODE_GEI	FLAG_OW	FLAG_OW	CNT_CHILI	AMT_INC(	AMT_CRE	AMT_ANN	AMT_GOC	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE
100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied	Working	Secondary / secondary spec
100003	0	Cash loans	F	N	N	0	270000	1293503	35698.5	1129500	Family	State servant	Higher education
100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied	Working	Secondary / secondary spec
100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied	Working	Secondary / secondary spec
100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied	Working	Secondary / secondary spec
100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner	State servant	Secondary / secondary spec
100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate	Higher education
100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied	State servant	Higher education
100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary / secondary spec
100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied	Working	Secondary / secondary spec
100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied	Working	Higher education
100015	0	Cash loans	F	N	Y	0	38419.16	148365	10678.5	135000	Children	Pensioner	Secondary / secondary spec
100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied	Working	Secondary / secondary spec
100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied	Working	Secondary / secondary spec
100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied	Working	Secondary / secondary spec
100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary / secondary spec
100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / secondary spec
100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied	Working	Secondary / secondary spec
100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary / secondary spec
100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher education



	SK_ID_PRE	SK_ID_CU	NAME_CO	AMT_ANN	AMT_APPI	AMT_CREI	AMT_DOV	AMT_GOC	WEEKDAY	HOUR_AP	FLAG_LAS	NFLAG_LA	RATE_DOV	NAME_CA	NAME_CO	DAYS_DEC	NAME_PA	CODE_REJ	NAME_TY	NA
2	2030495	271877	Consumer	1730.43	17145	17145	0	17145	SATURDAY	15	Y	1	0	XAP	Approved	-73	Cash throu	XAP		Re
3	2802425	108129	Cash loans	25188.62	607500	679671		607500	THURSDAY	11	Y	1		XNA	Approved	-164	XNA	XAP	Unaccomp	Re
4	2523466	122040	Cash loans	15060.74	112500	136444.5		112500	TUESDAY	11	Y	1		XNA	Approved	-301	Cash throu	XAP	Spouse, pa	Re
5	2819243	176158	Cash loans	47041.34	450000	470790		450000	MONDAY	7	Y	1		XNA	Approved	-512	Cash throu	XAP		Re
6	1784265	202054	Cash loans	31924.4	337500	404055		337500	THURSDAY	9	Y	1		Repairs	Refused	-781	Cash throu	HC		Re
7	1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	8	Y	1		Everyday e	Approved	-684	Cash throu	XAP	Family	Re
8	2315218	175704	Cash loans		0	0			TUESDAY	11	Y	1		XNA	Canceled	-14	XNA	XAP		Re
9	1656711	296299	Cash loans		0	0			MONDAY	7	Y	1		XNA	Canceled	-21	XNA	XAP		Re
10	2367563	342292	Cash loans		0	0			MONDAY	15	Y	1		XNA	Canceled	-386	XNA	XAP		Re
11	2579447	334349	Cash loans		0	0			SATURDAY	15	Y	1		XNA	Canceled	-57	XNA	XAP		Re
12	1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	7	Y	1		XNA	Approved	-735	Cash throu	XAP	Unaccomp	Re
13	2257824	161140	Cash loans	13832.78	211500	246397.5		211500	FRIDAY	10	Y	1		XNA	Approved	-815	Cash throu	XAP	Unaccomp	Re
14	2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	15	Y	1		XNA	Approved	-860	Cash throu	XAP	Unaccomp	Re
15	1397919	321676	Consumer	7654.86	53779.5	57564	0	53779.5	SUNDAY	15	Y	1	0	XAP	Approved	-408	Cash throu	XAP	Unaccomp	Ne
16	2273188	270658	Consumer	9644.22	26550	27252	0	26550	SATURDAY	10	Y	1	0	XAP	Approved	-726	Cash throu	XAP		Ne
17	1232483	151612	Consumer	21307.46	126490.5	119853	12649.5	126490.5	TUESDAY	7	Y	1	0.103971	XAP	Approved	-699	Cash throu	XAP	Unaccomp	Ne
18	2163253	154602	Consumer	4187.34	26955	27297	1350	26955	SATURDAY	12	Y	1	0.051324	XAP	Approved	-1473	Cash throu	XAP	Unaccomp	Re
19	1285768	142748	Revolving	9000	180000	180000		180000	FRIDAY	13	Y	1		XAP	Approved	-336	XNA	XAP	Unaccomp	Re
20	2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	14	Y	1		XNA	Approved	-700	Cash throu	XAP	Unaccomp	Re
21	1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	16	Y	1		Everyday e	Refused	-584	XNA	HC		Re

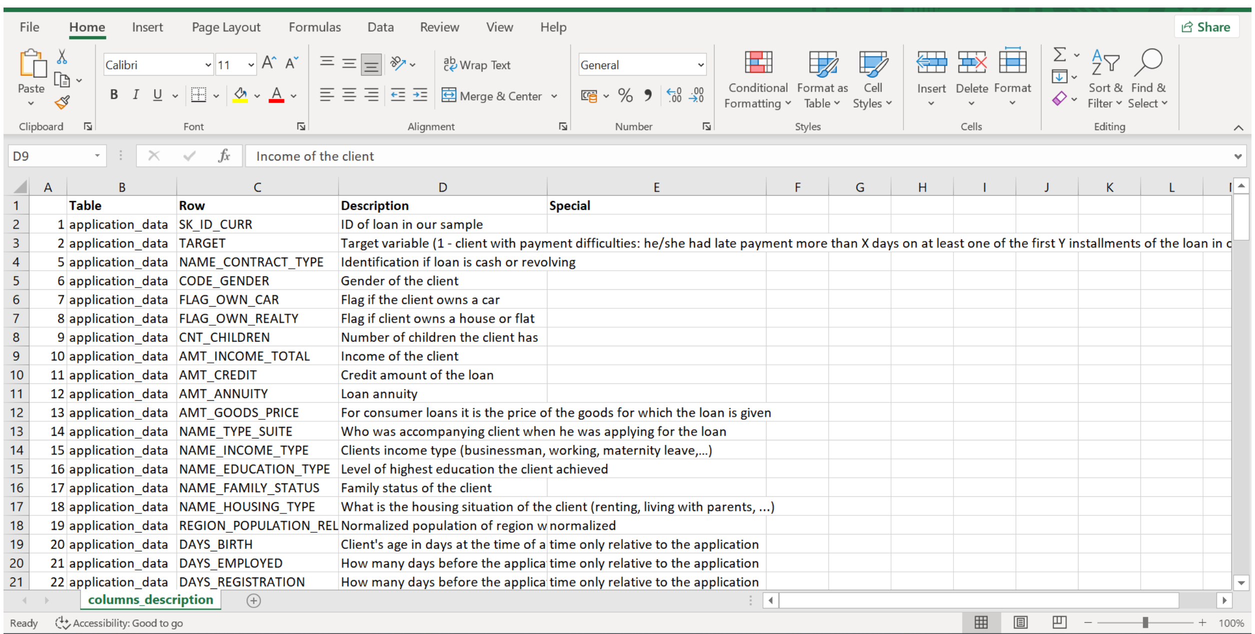


		Table	Row	Description	Special
2	1	application_data	SK_ID_CURR	ID of loan in our sample	
3	2	application_data	TARGET	Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in c	
4	5	application_data	NAME_CONTRACT_TYPE	Identification if loan is cash or revolving	
5	6	application_data	CODE_GENDER	Gender of the client	
6	7	application_data	FLAG_OWN_CAR	Flag if the client owns a car	
7	8	application_data	FLAG_OWN_REALTY	Flag if client owns a house or flat	
8	9	application_data	CNT_CHILDREN	Number of children the client has	
9	10	application_data	AMT_INCOME_TOTAL	Income of the client	
10	11	application_data	AMT_CREDIT	Credit amount of the loan	
11	12	application_data	AMT_ANNUITY	Loan annuity	
12	13	application_data	AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the loan is given	
13	14	application_data	NAME_TYPE_SUITE	Who was accompanying client when he was applying for the loan	
14	15	application_data	NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,…)	
15	16	application_data	NAME_EDUCATION_TYPE	Level of highest education the client achieved	
16	17	application_data	NAME_FAMILY_STATUS	Family status of the client	
17	18	application_data	NAME_HOUSING_TYPE	What is the housing situation of the client (renting, living with parents, ...)	
18	19	application_data	REGION_POPULATION_REL	Normalized population of region w	normalized
19	20	application_data	DAYS_BIRTH	Client's age in days at the time of a	time only relative to the application
20	21	application_data	DAYS_EMPLOYED	How many days before the applica	time only relative to the application
21	22	application_data	DAYS_REGISTRATION	How many days before the applica	time only relative to the application

Step 2 – Identify the missing data or Drop Columns

 # Cleaning the missing data

 # listing the null values columns having more than 30%

#Ctrl + F -> Leave the find what? Blank -> Tick the match entire cell content -> Look-in values -> Check the no. of cells.

Application Data has 64 columns, which has more than 30% of null values. - Removed
Previous Data has 15 columns with more than 30% of null values. - Removed

Since the 'AMT_ANNUITY' column has an outlier that is very large it will be inappropriate to fill those missing values with mean, Hence Median comes to the rescue for this and we have filled those missing banks with the median value.

A list of some unwanted columns is also removed, such as –

FLAG_MOBIL
FLAG_PHONE
REGION_RATING_CLIENT_W_CITY
FLAG_DOCUMENT_7
FLAG_DOCUMENT_13
FLAG_DOCUMENT_19
FLAG_EMP_PHONE
FLAG_EMAIL
DAYS_LAST_PHONE_CHANGE
FLAG_DOCUMENT_8
FLAG_DOCUMENT_14
FLAG_DOCUMENT_20
FLAG_WORK_PHONE
REGION_RATING_CLIENT
FLAG_DOCUMENT_2
FLAG_DOCUMENT_9
FLAG_DOCUMENT_15
FLAG_DOCUMENT_21
FLAG_CONT_MOBILE
REGION_RATING_CLIENT_W_CITY
FLAG_DOCUMENT_3
FLAG_DOCUMENT_10
FLAG_DOCUMENT_16
FLAG_EMAIL
FLAG_DOCUMENT_4

FLAG_DOCUMENT_11
FLAG_DOCUMENT_17
CNT_FAM_MEMBERS
FLAG_DOCUMENT_5
FLAG_DOCUMENT_12
FLAG_DOCUMENT_18
REGION_RATING_CLIENT
FLAG_DOCUMENT_6

Some columns where the value is mentioned as 'XNA' which means 'Not Available'. - Removed

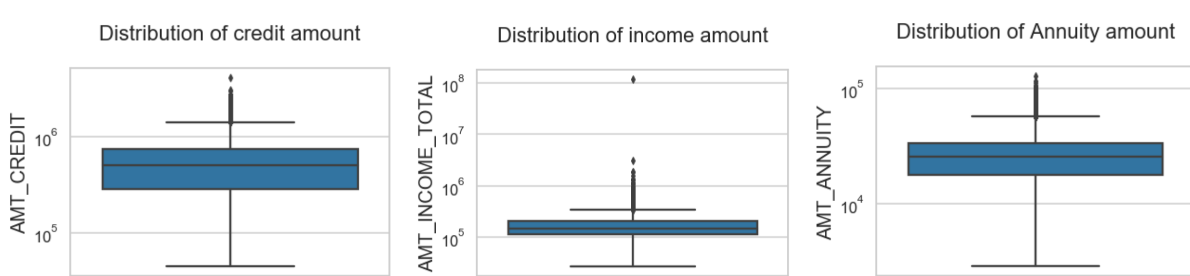
In CODE_GENDER, Female is having the majority and only 4 rows are having NA values, we can update those columns with Gender 'F' as there will be no impact on the dataset.

And for column 'ORGANIZATION_TYPE', we have a total count of 307511 rows of which 55374 rows are having 'XNA' values. Which means 18% of the column is having these values. Hence if we drop the rows of a total of 55374, will not have any major impact on our dataset.

We have also divided the dataset into two i.e. data1 - *(client with payment difficulties)* and data2 - *(all others)*

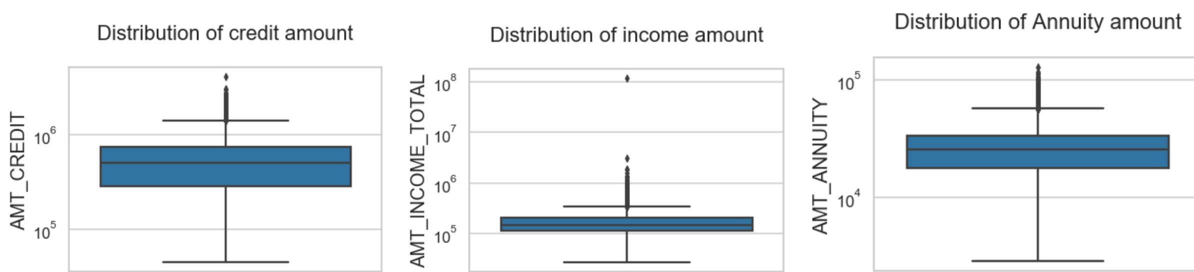
Step - 3 Find the Outliers in Data

 # For Data1 the outliers we checked by plotting a quick graph, which shows:



1. Some outliers are noticed in the annuity amount. The first quartile is bigger than the third quartile for annuity amount which means most of the annuity clients are from the first quartile.
2. Some outliers are noticed in credit amounts. The first quartile is bigger than the third quartile for credit amount which means most of the credits of clients are present in the first quartile.
3. Some outliers are noticed in income amounts. The third quartile is very slim for income amount. Most of the clients of income are present in the first quartile.

 # For Data2 the outliers we checked by plotting a quick graph, which shows:



1. Some outliers are noticed in income amount. The third quartile is very slim for income amount.
2. Some outliers are noticed in credit amounts. The first quartile is bigger than the third quartile for credit amount which means most of the credits of clients are present in the first quartile.
3. Some outliers are noticed in the annuity amount. The first quartile is bigger than the third quartile for annuity amount which means most of the annuity clients are from the first quartile.

Step - 4 Find the percentage of data imbalance

 # Calculating Imbalance percentage
 # Since the majority is data2 and the minority is data1

The Imbalance ratio is **10.51:1** (majority: minority)
Ratios of imbalance in percentage with respect to data2 and data1 data are 93.11 and 8.86.

Step - 5 Correlation

 # For data2 the correlation, shows –

1. Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
2. Income amount is inversely proportional to the number of children clients have, which means more income for fewer children clients have and vice-versa.
3. fewer children clients have in densely populated areas.
4. The income and credit amounts are higher in densely populated areas.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		CNT_CHIL	AMT_INC(	AMT_CRE	AMT_ANN	REGION_F	DAYS_BIR	DAYS_EM	DAYS_REC	DAYS_ID_	HOUR_AP	REG_REGI	REG_REGI	LIVE_REGI	REG_CITY_	REG_CITY_	LIVE_CITY_
2	CNT_CHIL	1	-0.02195	-0.02365	-0.0108	-0.03058	0.26653	0.03095	0.15552	-0.11916	-0.03016	-0.02281	-0.01548	-0.00558	0.00234	0.00749	0.0133
3	AMT_INC	-0.02195	1	0.40388	0.4722	0.11007	-0.05467	-0.06087	0.04056	-0.0367	0.0735	0.07763	0.15996	0.14828	-0.00102	-0.01386	-0.00476
4	AMT_CRE	-0.02365	0.40388	1	0.82669	0.06071	-0.16903	-0.10425	-0.01532	-0.0382	0.03692	0.01512	0.04169	0.04518	-0.04062	-0.037	-0.01119
5	AMT_ANI	-0.0108	0.4722	0.82669	1	0.06433	-0.10029	-0.07464	0.01071	-0.02735	0.03295	0.03344	0.07084	0.06905	-0.01995	-0.02409	-0.00809
6	REGION_I	-0.03058	0.11007	0.06071	0.06433	1	-0.04166	0.0009	-0.0424	-0.0103	0.13321	-0.02529	0.03245	0.05681	-0.04978	-0.03481	-0.00733
7	DAYS_BIR	0.26653	-0.05467	-0.16903	-0.10029	-0.04166	1	0.30779	0.26545	0.08333	0.0513	0.05863	0.0381	0.01279	0.16748	0.11154	0.02901
8	DAYS_EM	0.03095	-0.06087	-0.10425	-0.07464	0.0009	0.30779	1	0.12671	0.10682	0.02644	0.06544	0.08697	0.06353	0.11822	0.12595	0.06957
9	DAYS_REC	0.15552	0.04056	-0.01532	0.01071	-0.0424	0.26545	0.12671	1	0.03679	-0.02955	0.01772	0.01509	0.00772	0.03806	0.04734	0.02723
10	DAYS_ID_	-0.11916	-0.0367	-0.0382	-0.02735	-0.0103	0.08333	0.10682	0.03679	1	0.00854	0.0273	0.02082	0.00853	0.05488	0.03343	0.00148
11	HOUR_AP	-0.03016	0.0735	0.03692	0.03295	0.13321	0.0513	0.02644	-0.02955	0.00854	1	0.05174	0.06735	0.05381	0.01129	-0.00597	-0.01072
12	REG_REG	-0.02281	0.07763	0.01512	0.03344	-0.02529	0.05863	0.06544	0.01772	0.0273	0.05174	1	0.4616	0.09019	0.34232	0.14243	0.00348
13	REG_REG	-0.01548	0.15996	0.04169	0.07084	0.03245	0.0381	0.08697	0.01509	0.02082	0.06735	0.4616	1	0.86042	0.14848	0.22037	0.17847
14	LIVE_REG	-0.00558	0.14828	0.04518	0.06905	0.05681	0.01279	0.06353	0.00772	0.00853	0.05381	0.09019	0.86042	1	0.01501	0.16775	0.22087
15	REG_CITY	0.00234	-0.00102	-0.04062	-0.01995	-0.04978	0.16748	0.11822	0.03806	0.05488	0.01129	0.34232	0.14848	0.01501	1	0.44264	0.01178
16	REG_CITY	0.00749	-0.01386	-0.037	-0.02409	-0.03481	0.11154	0.12595	0.04734	0.03343	-0.00597	0.14243	0.22037	0.16775	0.44264	1	0.82083
17	LIVE_CITY	0.0133	-0.00476	-0.01119	-0.00809	-0.00733	0.02901	0.06957	0.02723	0.00148	-0.01072	0.00348	0.17847	0.22087	0.01178	0.82083	1

 # For data1 the correlation, shows –

Things are similar to data2 here, a few different points are listed below.

1. The client's permanent address does not match the contact address having fewer children and vice-versa
2. The client's permanent address does not match the work address having fewer children and vice-versa

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		CNT_CHILI	AMT_INC(	AMT_CREI	AMT_ANN	REGION_P	DAYS_BIR	DAYS_EMI	DAYS_REG	DAYS_ID_	HOUR_AP	REG_REGI	REG_REGI	LIVE_REGI	REG_CITY_	REG_CITY_	LIVE_CITY_N
2	CNT_CHILI	1	-0.03912	0.000427	0.015133	-0.02968	0.175025	0.006823	0.110854	-0.09104	-0.04034	-0.03521	-0.04085	-0.02799	-0.01607	-0.00544	0.009557
3	AMT_INC(	-0.03912	1	0.364559	0.428947	0.058005	-0.10303	-0.0538	0.011378	-0.05111	0.078779	0.075615	0.156374	0.145982	-0.00381	-0.00624	0.00423
4	AMT_CREI	0.000427	0.364559	1	0.812093	0.043545	-0.20072	-0.10761	-0.02197	-0.06514	0.024616	0.015043	0.032536	0.034861	-0.03097	-0.03288	-0.012465
5	AMT_ANN	0.015133	0.428947	0.812093	1	0.028666	-0.1002	-0.06019	0.019762	-0.04413	0.021129	0.029646	0.060363	0.059724	-0.01174	-0.01594	-0.003012
6	REGION_P	-0.02968	0.058005	0.043545	0.028666	1	-0.04444	-0.01525	-0.03349	-0.01778	0.1094	-0.0327	-0.00816	0.012602	-0.05724	-0.04476	-0.014753
7	DAYS_BIR	0.175025	-0.10303	-0.20072	-0.1002	-0.04444	1	0.25687	0.19235	0.146246	0.041994	0.04632	0.022208	0.000356	0.145884	0.096181	0.009633
8	DAYS_EMI	0.006823	-0.0538	-0.10761	-0.06019	-0.01525	0.25687	1	0.086286	0.104244	0.010328	0.069566	0.082264	0.056081	0.118869	0.139863	0.069316
9	DAYS_REG	0.110854	0.011378	-0.02197	0.019762	-0.03349	0.19235	0.086286	1	0.061563	-0.04475	0.006362	0.000896	-0.00142	0.015831	0.039204	0.026105
10	DAYS_ID_	-0.09104	-0.05111	-0.06514	-0.04413	-0.01778	0.146246	0.104244	0.061563	1	0.012709	0.02486	0.013162	0.002567	0.048184	0.015838	-0.015598
11	HOUR_AP	-0.04034	0.078779	0.024616	0.021129	0.1094	0.041994	0.010328	-0.04475	0.012709	1	0.050953	0.063877	0.0503	0.003947	0.004775	0.002319
12	REG_REGI	-0.03521	0.075615	0.015043	0.029646	-0.0327	0.04632	0.069566	0.006362	0.02486	0.050953	1	0.506747	0.068368	0.32203	0.150968	-0.013946
13	REG_REGI	-0.04085	0.156374	0.032536	0.060363	-0.00816	0.022208	0.082264	0.000896	0.013162	0.063877	0.506747	1	0.846872	0.141416	0.22437	0.181231
14	LIVE_REGI	-0.02799	0.145982	0.034861	0.059724	0.012602	0.000356	0.056081	-0.00142	0.002567	0.0503	0.068368	0.846872	1	-0.00698	0.167717	0.233975
15	REG_CITY_	-0.01607	-0.00381	-0.03097	-0.01174	-0.05724	0.145884	0.118869	0.015831	0.048184	0.003947	0.32203	0.141416	-0.00698	1	0.478266	-0.029432
16	REG_CITY_	-0.00544	-0.00624	-0.03288	-0.01594	-0.04476	0.096181	0.139863	0.039204	0.015838	0.004775	0.150968	0.22437	0.167717	0.478266	1	0.768247
17	LIVE_CITY_	0.009557	0.00423	-0.01247	-0.00301	-0.01475	0.009633	0.069316	0.026105	-0.0156	0.002319	-0.01395	0.181231	0.233975	-0.02943	0.768247	1

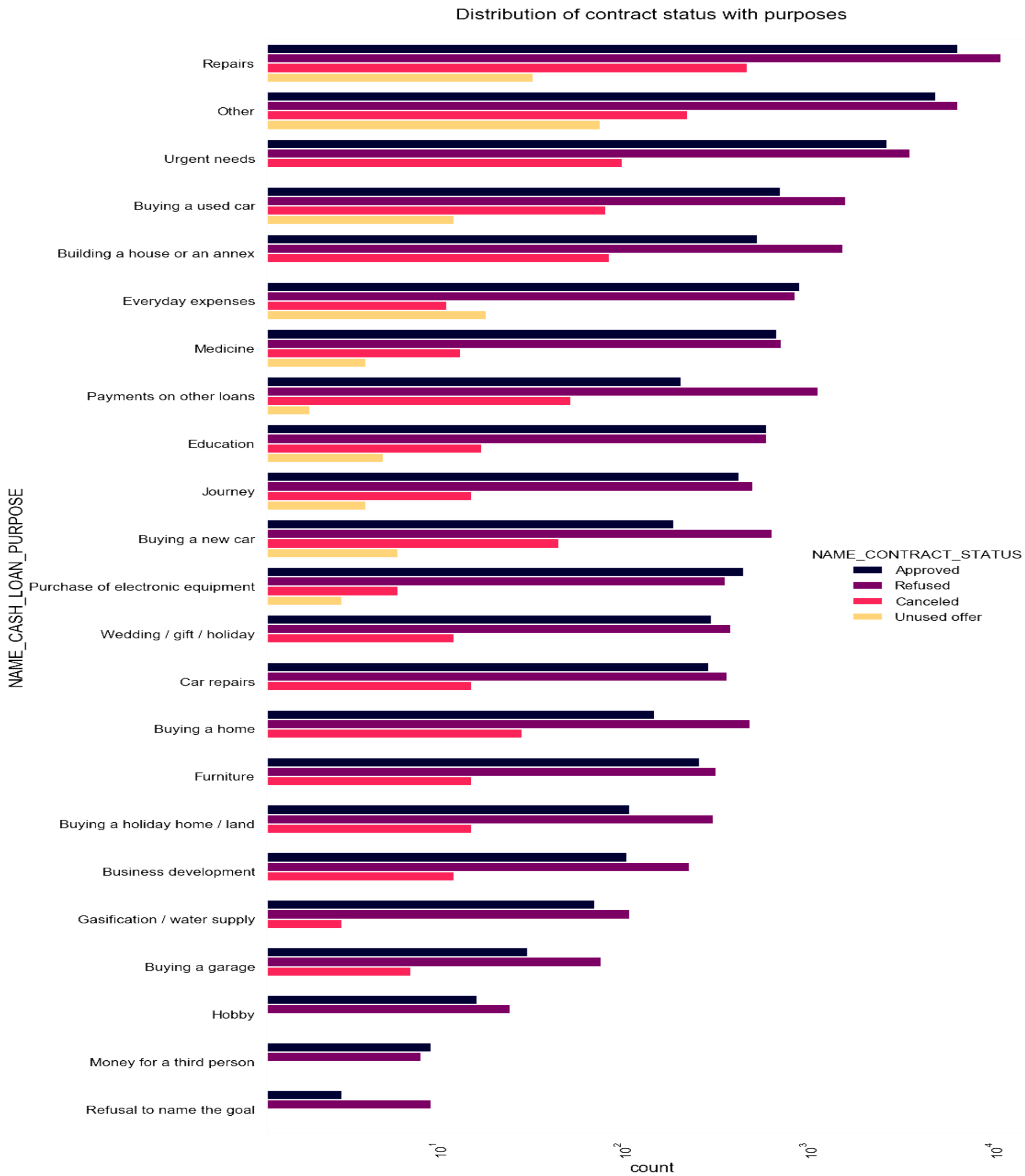
Step - 6 Results of univariate and bivariate analysis

We have merged the application and previous datasets using VLOOKUP and removed some more unwanted columns like –

SK_ID_CURR
REG_REGION_NOT_WORK_REGION
REG_CITY_NOT_WORK_CITY
HOUR_APPR_PROCESS_START_PREV
WEEKDAY_APPR_PROCESS_START
LIVE_REGION_NOT_WORK_REGION
LIVE_CITY_NOT_WORK_CITY
FLAG_LAST_APPL_PER_CONTRACT
HOUR_APPR_PROCESS_START
REG_CITY_NOT_LIVE_CITY
WEEKDAY_APPR_PROCESS_START_PREV
NFLAG_LAST_APPL_IN_DAY
REG_REGION_NOT_LIVE_REGION

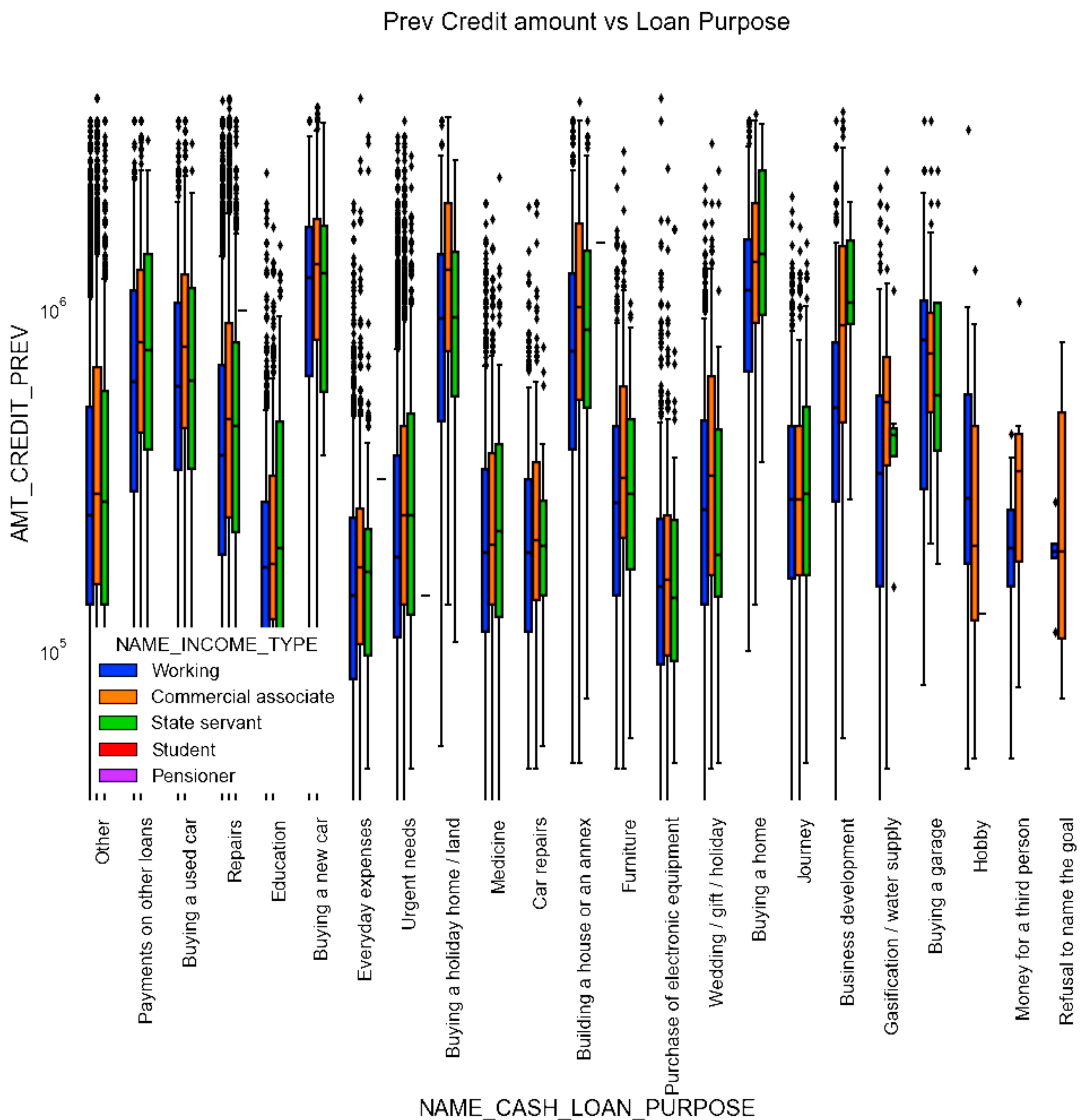
 # Univariate Analysis

1. Loan purposes with 'Repairs' are facing more difficulties in payment on time.
2. There are few places where loan payment is significantly higher and are facing difficulties. They are 'Buying a garage', 'Business development', 'Buying land',' Buying a new car' and 'Education'.
3. For education purposes we have approximately equal approves and rejections
4. Paying other loans and buying a new car is having significantly higher rejections than approves.



Bivariate Analysis

1. The credit amount for Loan purposes like 'Buying a home',' Buying land',' Buying a new car, and 'building a house' is higher.
2. Income type of state servants have a significant amount of credit applied
3. Money for the third person or a Hobby is having fewer credits applied for.



Prev Credit amount vs Loan Purpose
AMT_CREDIT_PREV
10^6
10^5
NAME_INCOME_TYPE
Working
Commercial associate
State servant
Student
Pensioner
Other
Payments on other loans
Buying a used car
Repairs
Education
Buying a new car
Everyday expenses
Urgent needs
Buying a holiday home / land
Medicine
Car repairs
Building a house or an annex
Furniture
Purchase of electronic equipment
Wedding / gift / holiday
Buying a home
Journey
Business development
Gasification / water supply
Buying a garage
Hobby
Money for a third person
Refusal to name the goal
NAME_CASH_LOAN_PURPOSE

Conclusions

1. Banks should **focus more** on contract types ‘Student’,’ pensioner’, and ‘Businessman’ with housing types other than ‘Co-op apartment’ for **successful payments.**
2. Banks should **focus less** on income type ‘Working’ as they are having the most number of **unsuccessful payments.**
3. Also with loan purposes ‘Repair’ is having a higher number of **unsuccessful payments** on time.
4. Get as many clients from the housing type ‘With parents’ as they are having the **least number of unsuccessful payments.**