

BANK LOAN CASE STUDY

Submitted By :- Shubham Saxena

Description

The objective of this case study is to provide practical insight into the application of Exploratory Data Analysis (EDA) within a genuine business context. Through this study, we'll not only implement EDA techniques covered in the module but also gain fundamental insights into risk analytics within the banking and financial sectors. By delving into this analysis, we'll grasp how data is effectively utilized to mitigate the risk associated with lending to customers, ultimately safeguarding against potential financial losses.

Business Understanding

Loan providers often face challenges when considering loan applications from individuals with either insufficient or non-existent credit histories. This creates an opportunity for certain consumers to exploit the system by defaulting on their loans. Suppose you're employed at a consumer finance company specializing in offering various types of loans to urban customers. Our task involves utilizing Exploratory Data Analysis (EDA) to examine patterns within the data. This analytical approach ensures that deserving loan applicants, capable of repaying their loans, are not unfairly rejected.

Approach

Download

I began by downloading the two CSV files and importing them into Excel for analysis.

- Previous_application.csv
- Application_data.csv

Cleaning

Due to the large size of the dataset, cleaning will be second step which will take considerable time

- Deleting unwanted columns
- Removing null values

Analysing

At last, I will initiate the exploratory data analysis (EDA) of the provided data.

Tech Stack Used

Excel

Microsoft Excel's user-friendly interface and powerful data manipulation tools make it ideal for efficient data cleaning. Its extensive range of functions simplifies error identification and correction, ensuring swift and accurate processing of large datasets.

Powerpoint

Microsoft PowerPoint offers a user-friendly interface with a plethora of creative tools, making it exceptionally easy to design visually stunning presentations.

Notepad

Notepad offers a lightweight and efficient platform for quick note-taking and text editing tasks. Its simplicity and minimalistic design ensure distraction-free writing, ideal for focusing on content.

Tasks

- A. Identify Missing Data and Deal with it Appropriately:
- B. Identify Outliers in the Dataset:
- C. Analyze Data Imbalance:
- D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:
- E. Identify Top Correlations for Different Scenarios:

Insight

Through this analysis, I gained insights into the importance of Exploratory Data Analysis (EDA) for every data analyst. Additionally, I acquired advanced Excel techniques and formulas, enhancing my confidence in data cleaning, visualization, correlation, and various analytical methods.

The dataset comprises three files: application_data.csv, containing client information at the time of application; previous_application.csv, detailing the client's previous loan data, including approval status; and columns_description.csv, serving as a data dictionary for variable definitions.

Upon observing unwanted rows and columns in the datasets, I performed data cleansing by removing unnecessary data before commencing the EDA process. The datasets under analysis are:

Downloading of Datasets

A1	X	Y	Z	AA	AB	AC
SK_ID_PREV						
1	SK_ID_PR	SK_ID_CU	NAME_CC	AMT_ANNUAL	AMT_APP	AMT_CREDIT
2	2030495	271877	Consumer	1730.43	17145	0
3	2802425	108129	Cash loan:	25188.62	607500	679671
4	2523466	122040	Cash loan:	15060.74	112500	136444.5
5	2819243	176158	Cash loan:	47041.34	450000	470790
6	1784265	202054	Cash loan:	31924.4	337500	404055
7	1383531	199383	Cash loan:	23703.93	315000	340573.5
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 loan:	11368.62	270000	335754
13	2257824	161140	Cash loan:	13832.78	211500	246397.5
14	2330894	258628	Cash loan:	12165.21	148500	174361.5
15	1397919	321676	Consumer	7654.86	53779.5	57564
16	2273188	270658	Consumer	9644.22	26550	27252
17	1232483	151612	Consumer	21307.46	126490.5	119853
18	2163253	154602	Consumer	4187.34	26955	27297
19	1285768	142748	Revolving	9000	180000	180000
20	2393109	396305	Cash loan:	10181.7	180000	180000
21	1173070	199178	Cash loan:	4666.5	45000	49455
22	1506815	166490	Cash loan:	25454.03	450000	491580
23	1182516	267782	Cash loan:	20361.6	405000	45177.5
24	1172842	302212	Cash loans	0	0	TUESDAY
25	1172937	302212	Cash loan:	39475.31	1129500	1277105
26	1555330	199353	Cash loans	0	0	SATURDAY
27	1543131	275707	Cash loan:	22619.52	229500	241920
28	2536650	338725	Cash loan:	16708.32	369000	369000
29	1676258	433469	Cash loan:	22242.83	247500	268083
30	2075578	418383	Consumer	7656.705	74610	65610
31	1583704	315664	Cash loans	0	0	WEDNESDAY
32	2138466	170289	Consumer	3286.485	32562	0
33	2191093	182450	Consumer	9789.255	100485	0
34	1471931	257846	Cash loans	0	0	MONDAY
35	1478703	393511	Cash loans	0	0	WEDNESDAY
36	1629736	255331	Revolving	10125	202500	202500
37	2027074	208000	Consumer	12065.54	56655	66411
38	2766278	178375	Consumer	7511.085	32490	38380.5
39	1151694	161591	Consumer	39111.08	214380	209497.5
40	1692577	125157	Consumer	9946.485	99319.5	109809
41	2740958	281589	Consumer	6752.25	35955	33115.5
42	2317570	258076	Consumer	4539.285	18360	21834
43	1366380	247572	Consumer loans	17950.5	0	17950.5

A1		application_data.csv - Excel																											
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
SK_ID_CURR	TARGET	NAME_CC_CODE_GEF	FLAG_OWF	FLAG_OWCNT	CHIL_AMT_INCC	AMT_CREATM	AMT_ANNAME	AMT_GOCNAME	NAME_TYP	NAME_INNAME	NAME_EDNAME	NAME_FA	NAME_HC	REGION_F	DAYS_BIR	DAYS_EMID	DAYS_REC	DAYS_ID	OWN_CAF	FLAG_MO	FLAG_EMF	FLAG_WO	FLAG_CON	FLAG_PHC	FLAG_EM	FLAG_AC	OCCUPATI		
2	100002	1	Cash loan: M	N	Y	0	202500	406597.5	24700.5	351000	Unacomp	Working	Secondary	Single / n	House / a	0.018801	-9461	-637	-3648	-2120	1	1	0	1	1	0	0	Labors	
3	100003	0	Cash loan: F	N	N	0	270000	1293503	35698.5	1129500	Family	State serv	Higher ed	Married	House / a	0.003541	-16765	-1188	-1186	-291	1	1	0	1	1	0	0	Core staff	
4	100004	0	Revolving M	Y	Y	0	67500	135000	6750	135000	Unacomp	Working	Secondary	Single / n	House / a	0.010032	-19046	-225	-4260	-2531	26	1	1	1	1	1	0	0	Labors
5	100006	0	Cash loan: F	N	Y	0	135000	312682.5	29868.5	297000	Unacomp	Working	Secondary	Civil marri	House / a	0.008019	-19005	-3039	-9833	-2437	1	1	0	1	0	0	0	Labors	
6	100007	0	Cash loan: M	N	Y	0	121500	513000	21865.5	513000	Unacomp	Working	Secondary	Single / n	House / a	0.028663	-19932	-3038	-4311	-3458	1	1	0	1	0	0	0	Core staff	
7	100008	0	Cash loan: M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, p	State serv	Secondary	Married	House / a	0.035792	-16941	-1588	-4970	-477	1	1	1	1	1	0	0	Labors	
8	100009	0	Cash loan: F	Y	Y	1	171000	1560726	41301	1395000	Unacomp	Commercial	Higher ed	Married	House / a	0.035792	-13778	-3130	-1213	-619	17	1	1	0	1	1	0	0	Accountant
9	100010	0	Cash loan: M	Y	Y	0	360000	1530000	42075	1530000	Unacomp	State serv	Higher ed	Married	House / a	0.003122	-18850	-449	-4597	-2379	8	1	1	1	1	0	0	0	Managers
10	100011	0	Cash loan: F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary	Married	House / a	0.018634	-20099	365243	-7427	-3514	1	0	0	1	0	0	0	0	
11	100012	0	Revolving M	N	Y	0	135000	405000	20250	405000	Unacomp	Working	Secondary	Single / n	House / a	0.019689	-14469	-2019	-14437	-3992	1	1	0	1	0	0	0	Labors	
12	100014	0	Cash loan: F	N	Y	1	112500	652500	21177	652500	Unacomp	Working	Higher ed	Married	House / a	0.0228	-10197	-679	-4427	-738	1	1	0	1	0	0	0	Core staff	
13	100015	0	Cash loan: F	N	Y	0	38419.16	148365	10678.5	135000	Children	Pensioner	Secondary	Married	House / a	0.015221	-20417	365243	-5246	-2512	1	0	0	1	1	0	0	0	
14	100016	0	Cash loan: F	N	Y	0	67500	80865	5881.5	67500	Unacomp	Working	Secondary	Married	House / a	0.031329	-13439	-2717	-311	-3227	1	1	1	1	1	0	0	Labors	
15	100017	0	Cash loan: M	Y	N	1	225000	918468	28966.5	697500	Unacomp	Working	Secondary	Married	House / a	0.016612	-14086	-3028	-643	-4911	23	1	1	0	1	0	0	0	Drivers
16	100018	0	Cash loan: F	N	Y	0	18900	773680.5	32778	679500	Unacomp	Working	Secondary	Married	House / a	0.010006	-14583	-203	-615	-2056	1	1	0	1	0	0	0	Labors	
17	100019	0	Cash loan: M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary	Single / n	Rented apt	0.020713	-8728	-1157	-3494	-1368	17	1	1	0	1	0	0	0	Labors
18	100020	0	Cash loan: M	N	N	0	108000	509602.5	26149.5	387000	Unacomp	Working	Secondary	Married	House / a	0.018634	-12931	-1317	-6392	-3866	1	1	0	1	0	0	0	Drivers	
19	100021	0	Revolving F	N	Y	1	81000	270000	13500	270000	Unacomp	Working	Secondary	Married	House / a	0.010966	-9776	-191	-4143	-2427	1	1	0	1	0	0	0	Labors	
20	100022	0	Revolving F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary	Widow	House / a	0.04622	-17718	-7804	-8751	-1259	1	1	0	1	1	0	0	Labors	
21	100023	0	Cash loan: F	N	Y	1	90000	544491	17563.5	454500	Unacomp	State serv	Higher ed	Single / n	House / a	0.015221	-11348	-2038	-1021	-3964	1	1	1	1	1	0	0	Core staff	
22	100024	0	Revolving M	Y	Y	0	135000	427500	21375	427500	Unacomp	Working	Secondary	Married	House / a	0.015221	-18252	-4286	-298	-1800	7	1	1	0	1	0	0	0	Labors
23	100025	0	Cash loan: F	Y	Y	1	202500	1132574	37561.5	927000	Unacomp	Commercial	Secondary	Married	House / a	0.025164	-14815	-1652	-2299	-2299	14	1	1	0	1	0	0	0	Sales staff
24	100026	0	Cash loan: F	N	N	1	450000	497520	32521.5	450000	Unacomp	Working	Secondary	Married	Rented apt	0.020713	-11146	-4306	-114	-2518	1	1	0	1	0	0	0	Sales staff	
25	100027	0	Cash loan: F	N	Y	0	83250	239850	23850	225000	Unacomp	Pensioner	Secondary	Married	House / a	0.006296	-24827	365243	-9012	-3684	1	0	0	1	1	0	0	Drivers	
26	100029	0	Cash loan: M	Y	N	2	135000	247500	12703.5	247500	Unacomp	Working	Secondary	Married	House / a	0.026392	-11286	-746	-108	-3729	7	1	1	0	1	0	0	0	Cleaning s
27	100030	0	Cash loan: F	N	Y	0	90000	225000	11074.5	225000	Unacomp	Working	Secondary	Married	House / a	0.028663	-19334	-3494	-2419	-2893	1	1	0	1	1	0	1	Cleaning s	
28	100031	1	Cash loan: F	N	Y	0	112500	979992	20765.5	702000	Unacomp	Working	Secondary	Widow	House / a	0.018029	-18724	-2628	-6573	-1827	1	1	0	1	0	0	0	Cooking st	
29	100032	0	Cash loan: M	N	Y	1	112500	327024	23827.5	270000	Family	Working	Secondary	Married	House / a	0.019101	-15948	-1234	-5782	-3153	1	1	0	1	0	0	0	Labors	
30	100033	0	Cash loan: M	Y	Y	0	270000	790830	57676.5	675000	Unacomp	State serv	Higher ed	Single / n	House / a	0.04622	-9994	-1796	-4668	-2661	1	1	0	1	0	0	1	0	
31	100034	0	Revolving M	N	Y	0	90000	180000	9000	180000	Unacomp	Working	Higher ed	Single / n	With pare	0.030755	-10341	-1010	-4799	-3015	1	1	0	1	0	0	0	Labors	
32	100035	0	Cash loan: F	N	Y	0	292500	665892	24592.5	477000	Unacomp	Commercial	Secondary	Civil marri	House / a	0.025164	-15280	-2668	-5266	-3787	1	1	0	1	0	0	1	0	
33	100036	0	Cash loan: F	N	Y	0	112500	512064	25033.5	360000	Family	Working	Secondary	Civil marri	House / a	0.008575	-11144	-1104	-7846	-2904	1	1	0	1	0	0	0	Private se	
34	100037	0	Cash loan: F	N	N	0	90000	199008	20893.5	180000	Unacomp	Working	Secondary	Civil marri	House / a	0.010032	-12974	-4404	-7123	-4464	1	1	1	1	1	0	0	Labors	
35	100039	0	Cash loan: M	Y	N	1	360000	733315.5	39069	679500	Unacomp	Commercial	Secondary	Married	House / a	0.015221	-11694	-2060	-3557	-3557	3	1	1	0	1	0	0	Drivers	
36	100040	0	Cash loan: F	N	Y	0	135000	1125000	32895	1125000	Unacomp	State serv	Higher ed	Married	House / a	0.019689	-15997	-4585	-5735	-4067	1	1	0	1	0	0	0	Core staff	
37	100041	0	Cash loan: F	N	N	0	112500	450000	44509.5	450000	Unacomp	Working	Higher ed	Married	House / a	0.008575	-12158	-1275	-6265	-2009	1	1	1	1	1	0	0	Sales staff	
38	100043	0	Cash loan: F	N	Y	2	198000	64173.5	23157	553500	Unacomp	Commercial	Secondary	Married	House / a	0.01885	-17199	-768	-63	-735	1	1	1	1	1	0	1	Private se	
39	100044	0	Cash loan: M	N	Y	0	121500	454500	15151.5	454500	Unacomp	Working	Secondary	Married	House / a	0.030755	-21077	-1288	-5474	-4270	1	1	0	1	1	0	0	Drivers	
40	100045	0	Cash loan: F	N	Y	0	99000	247275	17338.5	225000	Unacomp	Pensioner	Secondary	Married	House / a	0.006207	-23920	365243	-9817	-4969	1	0	0	1	1	0	0	Managers	
41	100046	0	Revolving M	Y	Y	0	180000	540000	27000	540000	Unacomp	Working	Higher ed	Married	House / a	0.025164	-16126	-8236	-4292	-3029	3	1	1	1	0	0	0	Labors	
42	100047	1	Cash loan: M	N	Y	0	202500	1193580	35028	855000	Unacomp	Commercial	Secondary	Civil marri	House / a	0.025164	-17482	-1262	-1182	-1029	1	1	1	1	0	0	0	Labors	
43	100048	0	Cash loan: F	N	Y	0	202500	604152	29196	540000	Unacomp	Working	Secondary	Married	House / a	0.009175	-16971	-475	-3148	-513	1	1	1	1	0	0	0	Labors	

columns_description.csv - Excel			Search	SHUBHAM SAXENA	Share
D12	A	B	C	D	
1	Column No.	Table	Row	Description	
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 our sample, 0 - all other cases)	
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_RELATIVE	Normalized population of region where client lives (higher number means the client lives in more populated region)	
19	20	application_data	DAYS_BIRTH	Client's age in days at the time of application	
20	21	application_data	DAYS_EMPLOYED	How many days before the application the person started current employment	
21	22	application_data	DAYS_REGISTRATION	How many days before the application did client change his registration	
22	23	application_data	DAYS_ID_PUBLISH	How many days before the application did client change the identity document with which he applied for the loan	
23	24	application_data	OWN_CAR_AGE	Age of client's car	
24	25	application_data	FLAG_MOBIL	Did client provide mobile phone (1=YES, 0=NO)	
25	26	application_data	FLAG_EMP_PHONE	Did client provide work phone (1=YES, 0=NO)	
26	27	application_data	FLAG_WORK_PHONE	Did client provide home phone (1=YES, 0=NO)	
27	28	application_data	FLAG_CONT_MOBILE	Was mobile phone reachable (1=YES, 0=NO)	
28	29	application_data	FLAG_PHONE	Did client provide home phone (1=YES, 0=NO)	
29	30	application_data	FLAG_EMAIL	Did client provide email (1=YES, 0=NO)	
30	31	application_data	OCCUPATION_TYPE	What kind of occupation does the client have	
31	32	application_data	CNT_FAM_MEMBERS	How many family members does client have	
32	33	application_data	REGION_RATING_CLIENT	Our rating of the region where client lives (1,2,3)	
33	34	application_data	REGION_RATING_CLIENT_W_CITY	Our rating of the region where client lives with taking city into account (1,2,3)	
34	35	application_data	WEEKDAY_APPR_PROCESS_START	On which day of the week did the client apply for the loan	
35	36	application_data	HOUR_APPR_PROCESS_START	Approximately at what hour did the client apply for the loan	
36	37	application_data	REG_REGION_NOT_LIVE_REGION	Flag if client's permanent address does not match contact address (1=different, 0=same, at region level)	
37	38	application_data	REG_REGION_NOT_WORK_REGION	Flag if client's permanent address does not match work address (1=different, 0=same, at region level)	
38	39	application_data	LIVE_REGION_NOT_WORK_REGION	Flag if client's contact address does not match work address (1=different, 0=same, at region level)	
39	40	application_data	REG_CITY_NOT_LIVE_CITY	Flag if client's permanent address does not match contact address (1=different, 0=same, at city level)	
40	41	application_data	REG_CITY_NOT_WORK_CITY	Flag if client's permanent address does not match work address (1=different, 0=same, at city level)	
41	42	application_data	LIVE_CITY_NOT_WORK_CITY	Flag if client's contact address does not match work address (1=different, 0=same, at city level)	
42	43	application_data	ORGANIZATION_TYPE	Type of organization where client works	
43	44	application_data	EXT_SOURCE_1	Normalized score from external data source	

Activate Windows
Go to Settings to activate Windows.

Tasks

A. Identify Missing Data and Deal with it Appropriately:

- Task: Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.
- Hint: Utilize Excel functions like COUNT, ISBLANK, and IF to identify missing data. Consider using functions like AVERAGE or MEDIAN for imputation or other appropriate methods available in Excel.
- Graph suggestion: Create a bar chart or column chart to visualize the proportion of missing values for each variable.

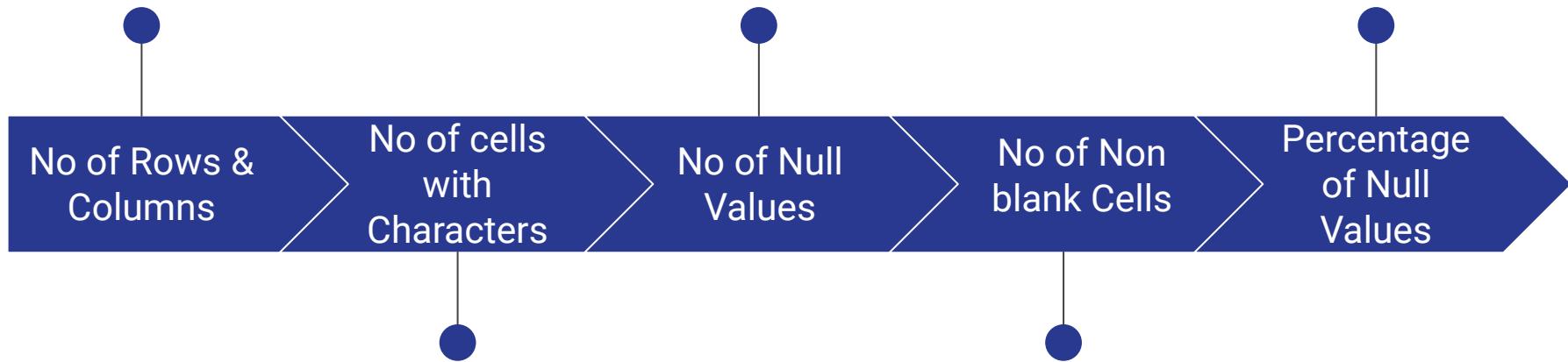
Cleaning of Datasets

Here we have used different functions like:

Syntax=COUNT(A1:A20)

Syntax=COUNTBLANK(range)

Syntax=No of Null values/No of Non-blank Cells



Syntax=COUNT(A1:A20)

Syntax=No of rows - No of Null values

A6 Condition(if null value percentage is more than 0.3 then that Coloum will be deleted)

DATA CLEANING PROCESS

PREVIOUS APPLICATION DATA

A	B	C	D	E	F	G	H	I	J	K	L	M
	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	HOUR_APPR_PROCESS_START	FLAG_LAST_APPL_PER_CONTRACT	NFLAG_LAST_APPL_I
No of Rows	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000
No of Values	50000	50000	50000	39408	50000	50000	24802	39256	50000	50000	50000	50000
No of Null Values	0	0	0	10592	0	0	25198	10744	0	0	0	0
Percentage of Null Values	0	0	0	0.268777913	0	0	1.015966454	0.273690646	0	0	0	0
Condition(if null value percentage is more than 0.3 then that Coloum will be deleted)							Deleted					
7	2030495	271877	Consumer loans	1730.43	17145	17145	0	17145	SATURDAY	15	Y	1
8	2802425	108129	Cash loans	25188.615	607500	679671		607500	THURSDAY	11	Y	1
9	2523466	122040	Cash loans	15060.735	112500	136444.5		112500	TUESDAY	11	Y	1
10	2819243	176158	Cash loans	47041.335	450000	470790		450000	MONDAY	7	Y	1
11	1784265	202054	Cash loans	31924.395	337500	404055		337500	THURSDAY	9	Y	1
12	1383531	199383	Cash loans	23703.93	315000	340573.5		315000	SATURDAY	8	Y	1
13	2315218	175704	Cash loans	0	0				TUESDAY	11	Y	1
14	1656711	296299	Cash loans	0	0				MONDAY	7	Y	1
15	2367563	342292	Cash loans	0	0				MONDAY	15	Y	1
16	2579447	334349	Cash loans	0	0				SATURDAY	15	Y	1
17	1715995	447712	Cash loans	11368.62	270000	335754		270000	FRIDAY	7	Y	1
18	2257824	161140	Cash loans	13832.775	211500	246397.5		211500	FRIDAY	10	Y	1
19	2330894	258628	Cash loans	12165.21	148500	174361.5		148500	TUESDAY	15	Y	1
20	1397919	321676	Consumer loans	7654.86	53779.5	57564	0	53779.5	SUNDAY	15	Y	1
21	2273188	270658	Consumer loans	9644.22	26550	27252	0	26550	SATURDAY	10	Y	1
22	1232483	151612	Consumer loans	21307.455	126490.5	119853	12649.5	126490.5	TUESDAY	7	Y	1
23	2163253	154602	Consumer loans	4187.34	26955	27297	1350	26955	SATURDAY	12	Y	1
24	1285768	142748	Revolving loans	9000	180000	180000		180000	FRIDAY	13	Y	1
25	2393109	396305	Cash loans	10181.7	180000	180000		180000	THURSDAY	14	Y	1
26	1173070	199178	Cash loans	4666.5	45000	49455		45000	SATURDAY	16	Y	1
27	1506815	166490	Cash loans	25454.025	450000	491580		450000	MONDAY	6	Y	1
28	1182516	267782	Cash loans	20361.6	405000	451777.5		405000	SATURDAY	4	Y	1
29	1172842	302212	Cash loans	0	0				TUESDAY	9	Y	1
30	1172937	302212	Cash loans	39475.305	1129500	1277104.5		1129500	THURSDAY	5	Y	1
31	1555330	199353	Cash loans	0	0				SATURDAY	6	Y	1
32	1543131	275707	Cash loans	22619.52	229500	241920		229500	THURSDAY	8	Y	1
33	2536650	338725	Cash loans	16708.32	369000	369000		369000	WEDNESDAY	13	Y	1
34	1676258	433469	Cash loans	22242.825	247500	268083	0	247500	THURSDAY	14	Y	1
35	2075578	418383	Consumer loans	7656.705	74610	65610	9000	74610	MONDAY	14	Y	1
36	1583704	315664	Cash loans	0	0				WEDNESDAY	15	Y	1
37	2138466	170289	Consumer loans	3286.485	32562	32562	0	32562	FRIDAY	12	Y	1
38	2191093	182450	Consumer loans	9789.255	100485	100485	0	100485	WEDNESDAY	9	Y	1
39	1471931	257846	Cash loans	0	0				MONDAY	12	Y	1
40	1478703	393511	Cash loans	0	0				WEDNESDAY	19	Y	1

Activate Windows

Go to Settings to activate Windows.



FINAL DATA

PREVIOUS_APPLICATION_DATA

C11	X	Y	f _x	202054	FINAL DATA PREVIOUS_APPLICATION_DATA																								
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z			
1		SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	WEEKDAY	APPR_PROCESS_START	HOUR	APPR_PROCESS_START	FLAG_LAST_APPL_PER_CONTRACT	NFLAG_LAST_APPL_IN_DAY	NAME_CASH_1	NAME_CASH_2	NAME_CASH_3	NAME_CASH_4	NAME_CASH_5	NAME_CASH_6	NAME_CASH_7	NAME_CASH_8	NAME_CASH_9	NAME_CASH_10	NAME_CASH_11	NAME_CASH_12	NAME_CASH_13	NAME_CASH_14	NAME_CASH_15
2	No of Rows	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000			
3	No of Values	50000	50000	50000	39408	50000	50000	39256	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000			
4	No of Null Values	0	0	0	10592	0	0	10744	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
5	Percentage of Null Values	0	0	0	0.268784734	0	0	0.273697618	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
6	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	WEEKDAY	APPR_PROCESS_START	HOUR	APPR_PROCESS_START	FLAG_LAST_APPL_PER_CONTRACT	NFLAG_LAST_APPL_IN_DAY	NAME_CASH_1	NAME_CASH_2	NAME_CASH_3	NAME_CASH_4	NAME_CASH_5	NAME_CASH_6	NAME_CASH_7	NAME_CASH_8	NAME_CASH_9	NAME_CASH_10	NAME_CASH_11	NAME_CASH_12	NAME_CASH_13	NAME_CASH_14	NAME_CASH_15	
7	2030495	271877	Consumer loans	1730.43	17145	17145	17145	SATURDAY	15			Y	1														X		
8	2802425	108129	Cash loans	25188.615	607500	679671	607500	THURSDAY	11			Y	1														X		
9	2523466	122040	Cash loans	15060.735	112500	136444.5	112500	TUESDAY	11			Y	1														X		
10	2819243	176158	Cash loans	47041.335	450000	470790	450000	MONDAY	7			Y	1														X		
11	1784265	202054	Cash loans	31924.395	337500	404055	337500	THURSDAY	9			Y	1														Rep		
12	1383531	199388	Cash loans	23703.93	315000	340573.5	315000	SATURDAY	8			Y	1														Everyday		
13	2315218	175704	Cash loans	0	0			TUESDAY	11			Y	1														X		
14	1656711	296299	Cash loans	0	0			MONDAY	7			Y	1														X		
15	2367563	342292	Cash loans	0	0			MONDAY	15			Y	1														X		
16	2579447	334349	Cash loans	0	0			SATURDAY	15			Y	1													X			
17	1715995	447712	Cash loans	11368.62	270000	335754	270000	FRIDAY	7			Y	1													X			
18	2257824	161140	Cash loans	13832.775	211500	246397.5	211500	FRIDAY	10			Y	1													X			
19	2330894	258628	Cash loans	12165.21	148500	174361.5	148500	TUESDAY	15			Y	1													X			
20	1397919	321676	Consumer loans	7654.86	53779.5	57564	53779.5	SUNDAY	15			Y	1													X			
21	2273188	270658	Consumer loans	9644.22	26550	27252	26550	SATURDAY	10			Y	1													X			
22	1232483	151612	Consumer loans	21307.455	126490.5	119853	126490.5	TUESDAY	7			Y	1													X			
23	2163253	154602	Consumer loans	4187.34	26955	27297	26955	SATURDAY	12			Y	1													X			
24	1285768	142748	Revolving loans	9000	180000	180000	180000	FRIDAY	13			Y	1													X			
25	2393109	396305	Cash loans	10181.7	180000	180000	180000	THURSDAY	14			Y	1													X			
26	1173070	199178	Cash loans	4666.5	45000	49455	45000	SATURDAY	16			Y	1													Everyday			
27	1506815	166490	Cash loans	25454.025	450000	491580	450000	MONDAY	6			Y	1													X			
28	1182516	267782	Cash loans	20361.6	405000	451777.5	405000	SATURDAY	4			Y	1													X			
29	1172842	302212	Cash loans	0	0			TUESDAY	9			Y	1													X			
30	1172937	302212	Cash loans	39475.305	1129500	1277104.5	1129500	THURSDAY	5			Y	1													X			
31	1555330	199353	Cash loans	0	0			SATURDAY	6			Y	1													X			
32	1543131	275707	Cash loans	22619.52	229500	241920	229500	THURSDAY	8			Y	1													X			
33	2536650	338725	Cash loans	16708.32	369000	369000	369000	WEDNESDAY	13			Y	1													X			
34	1676258	433469	Cash loans	22242.825	247500	268083	247500	THURSDAY	14			Y	1													X			
35	2075578	418383	Consumer loans	7656.705	74610	65610	74610	MONDAY	14			Y	1													X			
36	1583704	315664	Cash loans	0	0			WEDNESDAY	15			Y	1													X			
37	2138466	170289	Consumer loans	3286.485	32562	32562	32562	FRIDAY	12			Y	1													X			
38	2191093	182450	Consumer loans	9789.255	100485	100485	100485	WEDNESDAY	9			Y	1													X			
39	1471931	257846	Cash loans	0	0			MONDAY	12			Y	1													X			
40	1478703	393511	Cash loans	0	0			WEDNESDAY	19			Y	1													X			
41	1629736	255331	Revolving loans	10125	202500	202500	202500	SATURDAY	14			Y	1													X			
42	2027074	208000	Consumer loans	12065.535	56655	66411	56655	SATURDAY	7			Y	1													X			
43	2766278	178375	Consumer loans	7511.085	32490	38380.5	32490	MONDAY	5			Y	1													X			

DATA CLEANING PROCESS

APPLICATION DATA

B5	=B4/B3																			
A	B	C	D	E	F	G	H	I	J	N	O	P								
	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	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	NAME_FAMILY_STATUS					
No of Rows	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	
No of Values	50000	50000	50000	50000	50000	50000	50000	50000	50000	49999	49962	49808	50000	50000	50000	50000	50000	50000	50000	
No of Null Values	0	0	0	0	0	0	0	0	0	1	38	192	0	0	0	0	0	0	0	
Percentage of Null Values	0	0	0	0	0	0	0	0	0	2.0004E-05	0.000760578	0.003854802	0	0	0	0	0	0	0	
Condition(if null value percentage is more than 0.3 then that Coloum will be deleted)										DELETED										
7	100002	1	Cash loans	M	N	Y		0	202500	406597.5	24700.5	351000	Unaccompanied	Working	Secondary / secondary special	Single /				
8	100003	0	Cash loans	F	N	N		0	270000	1293502.5	35698.5	1129500	Family	State servant	Higher education	Married /				
9	100004	0	Revolving loans	M	Y	Y		0	67500	135000	6750	135000	Unaccompanied	Working	Secondary / secondary special	Single /				
10	100006	0	Cash loans	F	N	Y		0	135000	312682.5	29686.5	297000	Unaccompanied	Working	Secondary / secondary special	Civil married				
11	100007	0	Cash loans	M	N	Y		0	121500	513000	21865.5	513000	Unaccompanied	Working	Secondary / secondary special	Single /				
12	100008	0	Cash loans	M	N	Y		0	99000	490495.5	27517.5	454500	Spouse, partner	State servant	Secondary / secondary special	Married /				
13	100009	0	Cash loans	F	Y	Y		1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate	Higher education	Married /				
14	100010	0	Cash loans	M	Y	Y		0	360000	1530000	42075	1530000	Unaccompanied	State servant	Higher education	Married /				
15	100011	0	Cash loans	F	N	Y		0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary / secondary special	Married /				
16	100012	0	Revolving loans	M	N	Y		0	135000	405000	20250	405000	Unaccompanied	Working	Secondary / secondary special	Single /				
17	100014	0	Cash loans	F	N	Y		1	112500	652500	21177	652500	Unaccompanied	Working	Higher education	Married /				
18	100015	0	Cash loans	F	N	Y		0	38419.155	148365	10678.5	135000	Children	Pensioner	Secondary / secondary special	Married /				
19	100016	0	Cash loans	F	N	Y		0	67500	80865	5881.5	67500	Unaccompanied	Working	Secondary / secondary special	Married /				
20	100017	0	Cash loans	M	Y	N		1	225000	918468	28966.5	697500	Unaccompanied	Working	Secondary / secondary special	Married /				
21	100018	0	Cash loans	F	N	Y		0	189000	773680.5	32778	679500	Unaccompanied	Working	Secondary / secondary special	Married /				
22	100019	0	Cash loans	M	Y	Y		0	157500	299772	20160	247500	Family	Working	Secondary / secondary special	Single /				
23	100020	0	Cash loans	M	N	N		0	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / secondary special	Married /				
24	100021	0	Revolving loans	F	N	Y		1	81000	270000	13500	270000	Unaccompanied	Working	Secondary / secondary special	Married /				
25	100022	0	Revolving loans	F	N	Y		0	112500	157500	7875	157500	Other_A	Working	Secondary / secondary special	Widow				
26	100023	0	Cash loans	F	N	Y		1	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher education	Single /				
27	100024	0	Revolving loans	M	Y	Y		0	135000	427500	21375	427500	Unaccompanied	Working	Secondary / secondary special	Married /				
28	100025	0	Cash loans	F	Y	Y		1	202500	1132052.5	37561.5	927000	Unaccompanied	Commercial associate	Secondary / secondary special	Married /				
29	100026	0	Cash loans	F	N	N		1	450000	497520	32521.5	450000	Unaccompanied	Working	Secondary / secondary special	Married /				
30	100027	0	Cash loans	F	N	Y		0	83250	239850	23850	225000	Unaccompanied	Pensioner	Secondary / secondary special	Married /				
31	100029	0	Cash loans	M	Y	N		2	135000	247500	12703.5	247500	Unaccompanied	Working	Secondary / secondary special	Married /				
32	100030	0	Cash loans	F	N	Y		0	90000	225000	11074.5	225000	Unaccompanied	Working	Secondary / secondary special	Married /				
33	100031	1	Cash loans	F	N	Y		0	112500	979992	27076.5	702000	Unaccompanied	Working	Secondary / secondary special	Widow				
34	100032	0	Cash loans	M	N	Y		1	112500	327024	28287.5	270000	Family	Working	Secondary / secondary special	Married /				
35	100033	0	Cash loans	M	Y	Y		0	270000	790830	57676.5	675000	Unaccompanied	State servant	Higher education	Single /				
36	100034	0	Revolving loans	M	N	Y		0	90000	180000	9000	180000	Unaccompanied	Working	Higher education	Single /				
37	100035	0	Cash loans	F	N	Y		0	292500	665892	24592.5	477000	Unaccompanied	Commercial associate	Secondary / secondary special	Civil married				
38	100036	0	Cash loans	F	N	Y		0	112500	512064	25033.5	360000	Family	Working	Secondary / secondary special	Civil married				
39	100037	0	Cash loans	F	N	N		0	90000	199008	20893.5	180000	Unaccompanied	Working	Secondary / secondary special	Civil married				
40	100039	0	Cash loans	M	Y	N		1	360000	733315.5	39069	679500	Unaccompanied	Commercial associate	Secondary / secondary special	Married /				

BS4995	FINAL DATA APPLICATION DATA														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	NAME_FAMILY_STATUS	
No of Rows	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	
No of Values	50000	50000	50000	50000	50000	50000	50000	50000	50000	49962	49808	50000	50000	50000	
No of Null Values	0	0	0	0	0	0	0	0	0	38	192	0	0	0	
Percentage of Null Values	0	0	0	0	0	0	0	0	0	0.000760578	0.003854802	0	0	0	
	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	NAME_FAMILY_STATUS	
100002	1	Cash loans	M	N	Y	0	202500	406597.5	351000	Unaccompanied	Working	Secondary / secondary special	Single / not married		
100003	0	Cash loans	F	N	N	0	270000	1293502.5	1129500	Family	State servant	Higher education	Married		
100004	0	Revolving loans	M	Y	Y	0	67500	135000	135000	Unaccompanied	Working	Secondary / secondary special	Single / not married		
100006	0	Cash loans	F	N	Y	0	135000	312682.5	297000	Unaccompanied	Working	Secondary / secondary special	Civil marriage		
100007	0	Cash loans	M	N	Y	0	121500	513000	513000	Unaccompanied	Working	Secondary / secondary special	Single / not married		
100008	0	Cash loans	M	N	Y	0	99000	490495.5	454500	Spouse, partner	State servant	Secondary / secondary special	Married		
100009	0	Cash loans	F	Y	Y	1	171000	1560726	1395000	Unaccompanied	Commercial associate	Higher education	Married		
100010	0	Cash loans	M	Y	Y	0	360000	1530000	1530000	Unaccompanied	State servant	Higher education	Married		
100011	0	Cash loans	F	N	Y	0	112500	1019610	913500	Children	Pensioner	Secondary / secondary special	Married		
100012	0	Revolving loans	M	N	Y	0	135000	405000	405000	Unaccompanied	Working	Secondary / secondary special	Single / not married		
100014	0	Cash loans	F	N	Y	1	112500	652500	652500	Unaccompanied	Working	Higher education	Married		
100015	0	Cash loans	F	N	Y	0	38419.155	148365	135000	Children	Pensioner	Secondary / secondary special	Married		
100016	0	Cash loans	F	N	Y	0	67500	80865	67500	Unaccompanied	Working	Secondary / secondary special	Married		
100017	0	Cash loans	M	Y	N	1	225000	918468	697500	Unaccompanied	Working	Secondary / secondary special	Married		
100018	0	Cash loans	F	N	Y	0	189000	773680.5	679500	Unaccompanied	Working	Secondary / secondary special	Married		
100019	0	Cash loans	M	Y	Y	0	157500	299772	247500	Family	Working	Secondary / secondary special	Single / not married		
100020	0	Cash loans	M	N	N	0	108000	509602.5	387000	Unaccompanied	Working	Secondary / secondary special	Married		
100021	0	Revolving loans	F	N	Y	1	81000	270000	270000	Unaccompanied	Working	Secondary / secondary special	Married		
100022	0	Revolving loans	F	N	Y	0	112500	157500	157500	Other_A	Working	Secondary / secondary special	Widow		
100023	0	Cash loans	F	N	Y	1	90000	544491	454500	Unaccompanied	State servant	Higher education	Single / not married		
100024	0	Revolving loans	M	Y	Y	0	135000	427500	427500	Unaccompanied	Working	Secondary / secondary special	Married		
100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	927000	Unaccompanied	Commercial associate	Secondary / secondary special	Married		
100026	0	Cash loans	F	N	N	1	450000	497520	450000	Unaccompanied	Working	Secondary / secondary special	Married		
100027	0	Cash loans	F	N	Y	0	83250	239850	225000	Unaccompanied	Pensioner	Secondary / secondary special	Married		
100029	0	Cash loans	M	Y	N	2	135000	247500	247500	Unaccompanied	Working	Secondary / secondary special	Married		
100030	0	Cash loans	F	N	Y	0	90000	225000	225000	Unaccompanied	Working	Secondary / secondary special	Married		
100031	1	Cash loans	F	N	Y	0	112500	979992	702000	Unaccompanied	Working	Secondary / secondary special	Widow		
100032	0	Cash loans	M	N	Y	1	112500	327024	270000	Family	Working	Secondary / secondary special	Married		
100033	0	Cash loans	M	Y	Y	0	270000	790830	675000	Unaccompanied	State servant	Higher education	Single / not married		
100034	0	Revolving loans	M	N	Y	0	90000	180000	180000	Unaccompanied	Working	Higher education	Single / not married		
100035	0	Cash loans	F	N	Y	0	292500	665892	477000	Unaccompanied	Commercial associate	Secondary / secondary special	Civil marriage		
100036	0	Cash loans	F	N	Y	0	112500	512064	360000	Family	Working	Secondary / secondary special	Civil marriage		
100037	0	Cash loans	F	N	N	0	90000	199008	180000	Unaccompanied	Working	Secondary / secondary special	Civil marriage		
100039	0	Cash loans	M	Y	N	1	360000	733315.5	679500	Unaccompanied	Commercial associate	Secondary / secondary special	Married		
100040	0	Cash loans	F	N	Y	0	135000	1125000	1125000	Unaccompanied	State servant	Higher education	Married		
100041	0	Cash loans	F	N	N	0	112500	450000	450000	Unaccompanied	Working	Higher education	Married		
100043	0	Cash loans	F	N	Y	2	198000	641173.5	553500	Unaccompanied	Commercial associate	Secondary / secondary special	Window Married		

PREVIOUS APPLICATION DATA

Before Cleaning		After Cleaning	
No of Rows	50000	No of Rows	50000
No of Columns	37	No of Columns	26
No of Null Values	321203	No of Null Values	31936
Total No of cells	1528797	Total No of cells	1268038

APPLICATION DATA

Before Cleaning		After Cleaning	
No of Rows	50000	No of Rows	50000
No of Columns	122	No of Columns	26
No of Null Values	1488212	No of Null Values	51376
Total No of cells	4611788	Total No of cells	3398624

Tasks

B. Identify Outliers in the Dataset:

- Task: Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.
- Hint: Utilize Excel functions like QUARTILE, IQR, and conditional formatting to identify potential outliers. Consider applying thresholds or business rules to determine if the outliers are valid data points or require further investigation.
- Graph suggestion: Create box plots or scatter plots to visualize the distribution of numerical variables and highlight the outliers.

OUTLIERS IN PREVIOUS APPLICATION DATA

	AMT_APPLICATION	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
COUNT	49999	49999	39407	39255
AVG	168892.4546	188542.8855	15482.59685	215141.4173
Q3	180000	198105.75	225000	225000
Q1	22045.5	26055	49410	49410
Min	0	0	0	0
Max	3826372.5	4104351	3826372.5	3826372.5
MEDIAN	71550	78907.5	104017.5	104017.5
S.D.	282203.5105	308473.6014	302499.2745	302499.2745
IQR	157954.5	172050.75	175590	175590
UPPER BOUND	416931.75	456181.875	488385	488385
LOWER BOUND	258977.25	284131.125	312795	312795

Here are the outcomes regarding outliers in the previous_application data for the variables AMT_APPLICATION, AMT_CREDIT, AMT_ANNUITY, and AMT_GOODS_PRICE:

→ AMT_APPLICATION Outliers:

- Count: 49,999
- Outlier Bounds:
Lower Bound: 258,977.25
Upper Bound: 416,931.75

→ AMT_CREDIT Outliers:

- Count: 49,999
- Outlier Bounds:
Lower Bound: 284,131.125
Upper Bound: 456,181.875

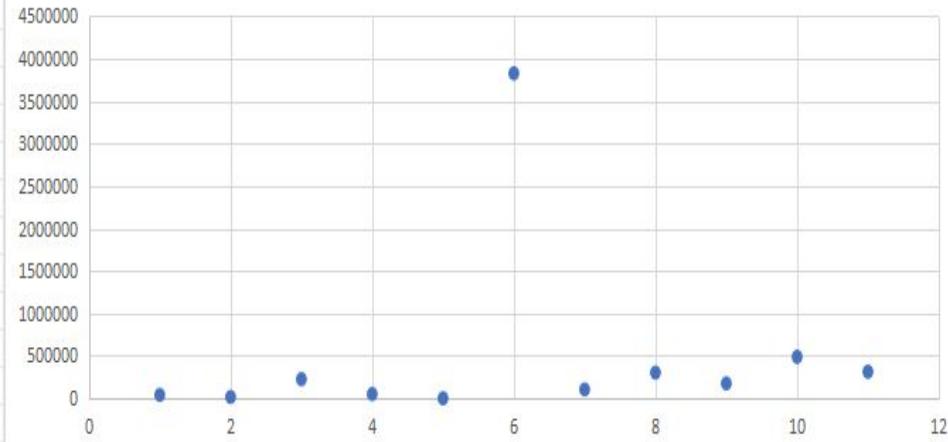
→ AMT_ANNUITY Outliers:

- Count: 39,407
- Outlier Bounds:
Lower Bound: 312,795
Upper Bound: 488,385

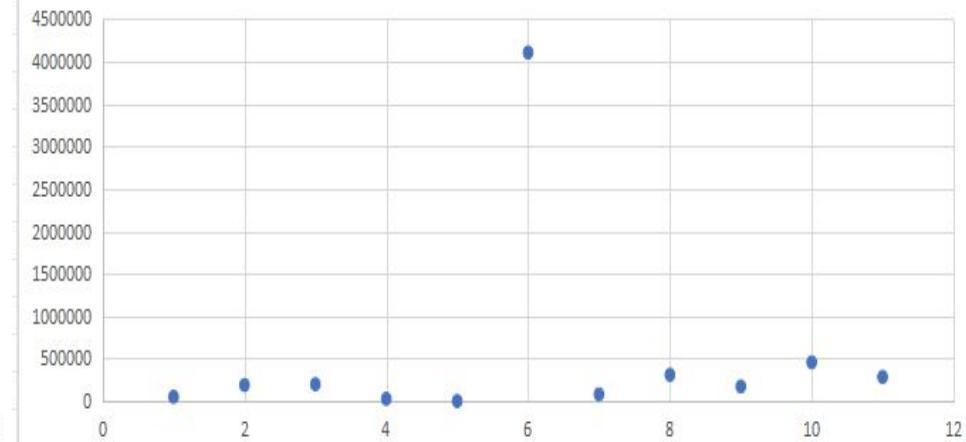
→ AMT_GOODS_PRICE Outliers:

- Count: 39,255
- Outlier Bounds:
Lower Bound: 312,795
Upper Bound: 488,385

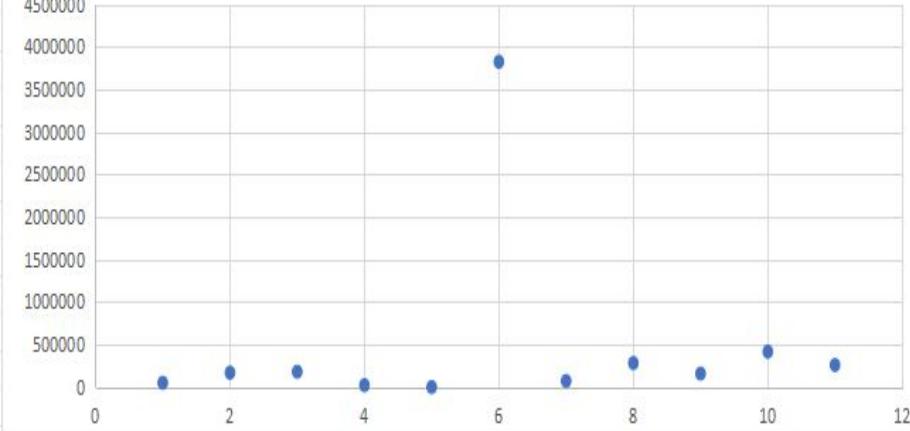
AMT_ANNUITY



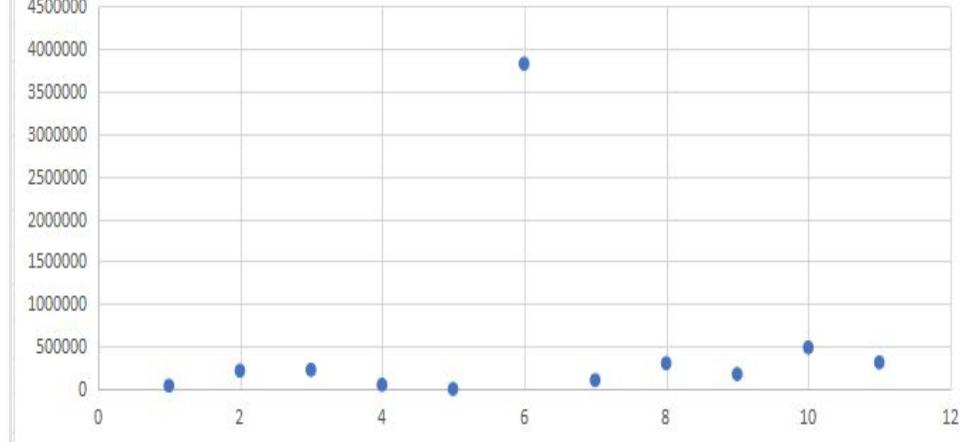
AMT_CREDIT



AMT_APPLICATION



AMT_GOODS_PRICE



OUTLIERS IN PREVIOUS_APPLICATION DATA

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_GOODS_PRICE
COUNT	49999	49999	49961
AVG	170767.5905	599700.5815	539060.0361
Q3	202500	808650	679500
Q1	112500	270000	238500
Min	25650	45000	45000
Max	117000000	4050000	4050000
MEDIAN	145800	514777.5	450000
S.D	531819.0951	402415.4339	369853.2526
IQR	90000	538650	441000
UPPER BOUND	337500	1616625	1341000
LOWER BOUND	247500	1077975	900000

Here are the outcomes regarding outliers in the previous_application data for the variables AMT_INCOME_TOTAL, AMT_CREDIT, and AMT_GOODS_PRICE:

→ AMT_INCOME_TOTAL Outliers:

- Count: 49,999
- Outlier Bounds:
Lower Bound: 247,500
Upper Bound: 337,500

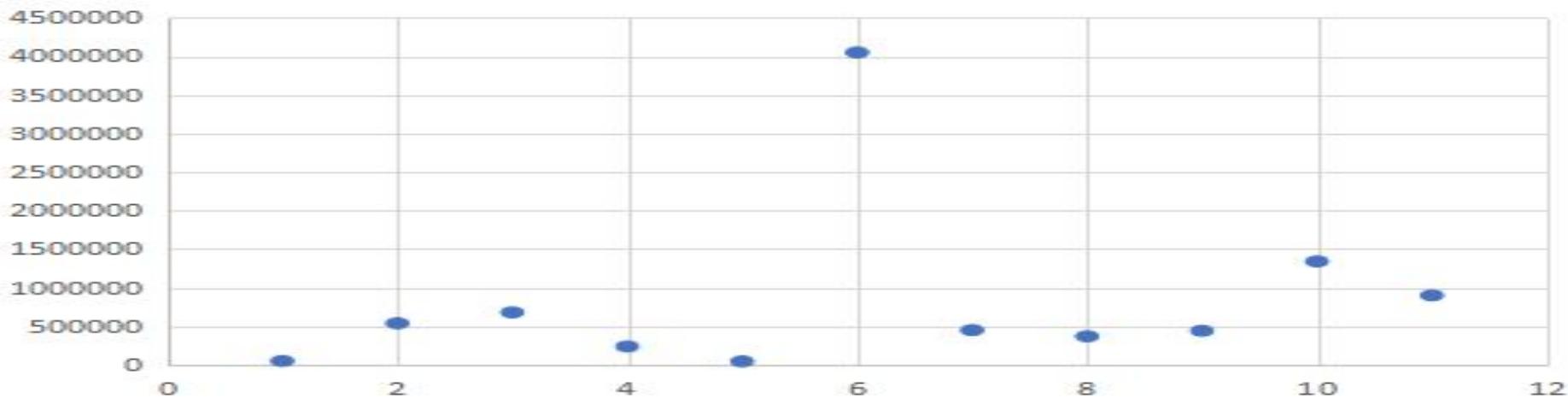
→ AMT_CREDIT Outliers:

- Count: 49,999
- Outlier Bounds:
Lower Bound: 1,077,975
Upper Bound: 1,616,625

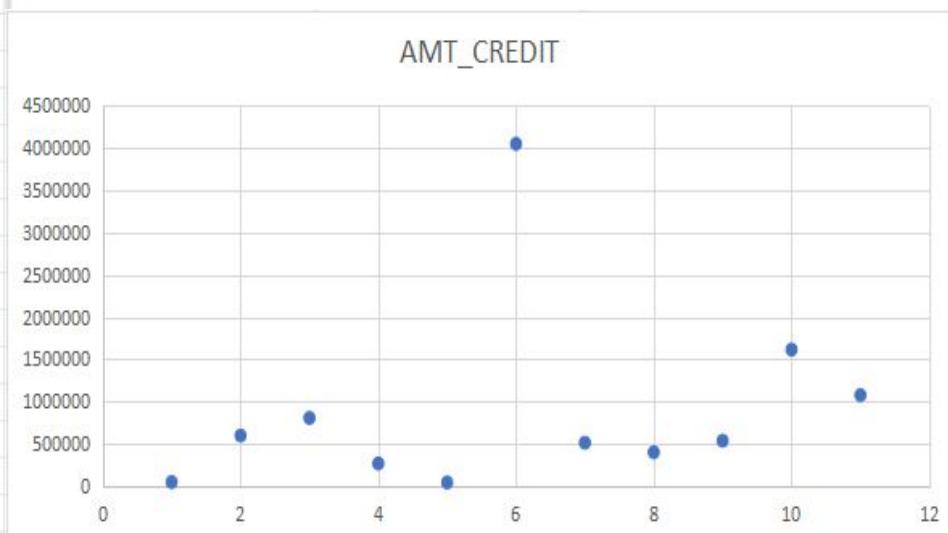
→ AMT_GOODS_PRICE Outliers:

- Count: 49,961
- Outlier Bounds:
Lower Bound: 900,000
Upper Bound: 1,341,000

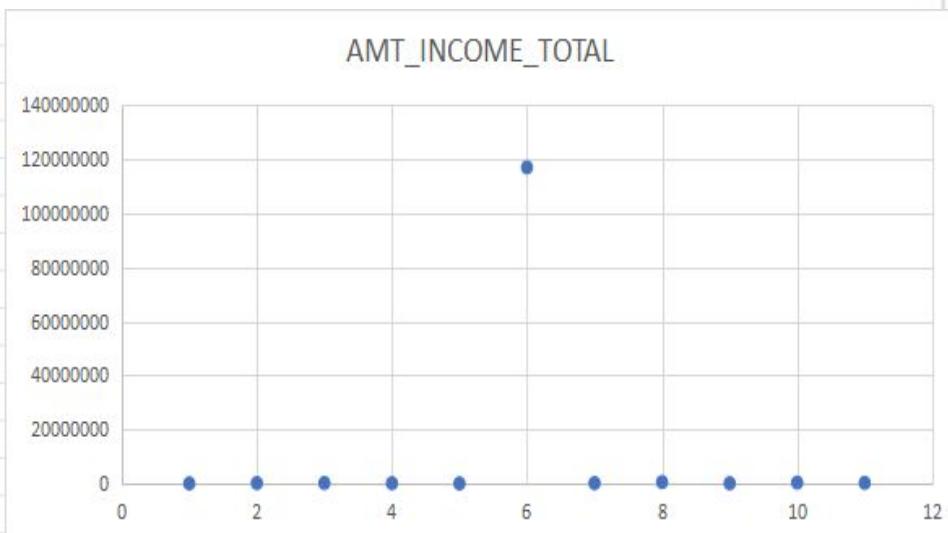
AMT_GOODS_PRICE



AMT_CREDIT



AMT_INCOME_TOTAL



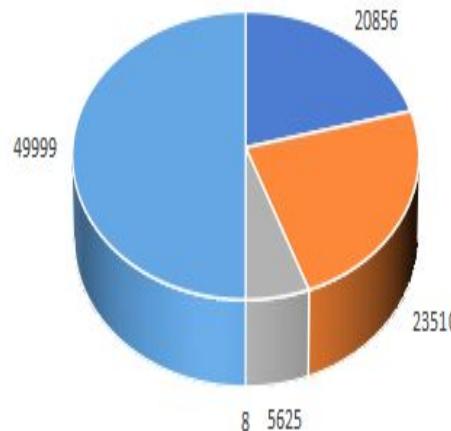
Tasks

C. Analyze Data Imbalance

- Task: Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.
- Hint: Utilize Excel functions like COUNTIF and SUM to calculate the proportions of each class. Compare the class frequencies to assess data imbalance.
- Graph suggestion: Create a pie chart or bar chart to visualize the distribution of the target variable and highlight the class imbalance.

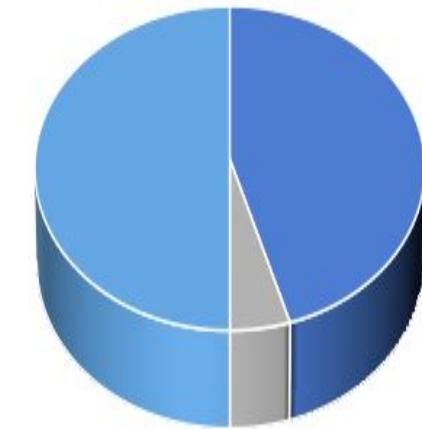
CONTRACT TYPE	PREVIOUS_APPLICATION DATA SETS USERS	APPLICATION_DATASETS USERS
CASH LOANS	20856	45276
CONSUMER LOANS	23510	0
REVOLVING LOANS	5625	4723
XNA	8	0
GRAND TOTAL	49999	49999

PREVIOUS_APPLICATION DATA SETS USERS



■ CASH LOANS ■ CONSUMER LOANS ■ REVOLVING LOANS ■ XNA ■ GRAND TOTAL

APPLICATION_DATASETS USERS

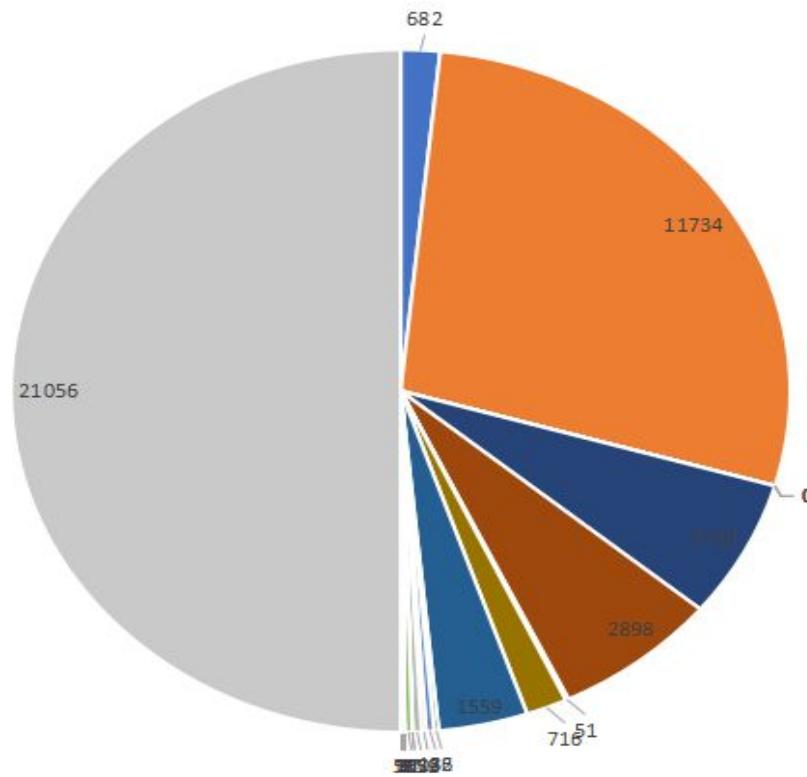


■ CASH LOANS ■ CONSUMER LOANS ■ REVOLVING LOANS ■ XNA ■ GRAND TOTAL

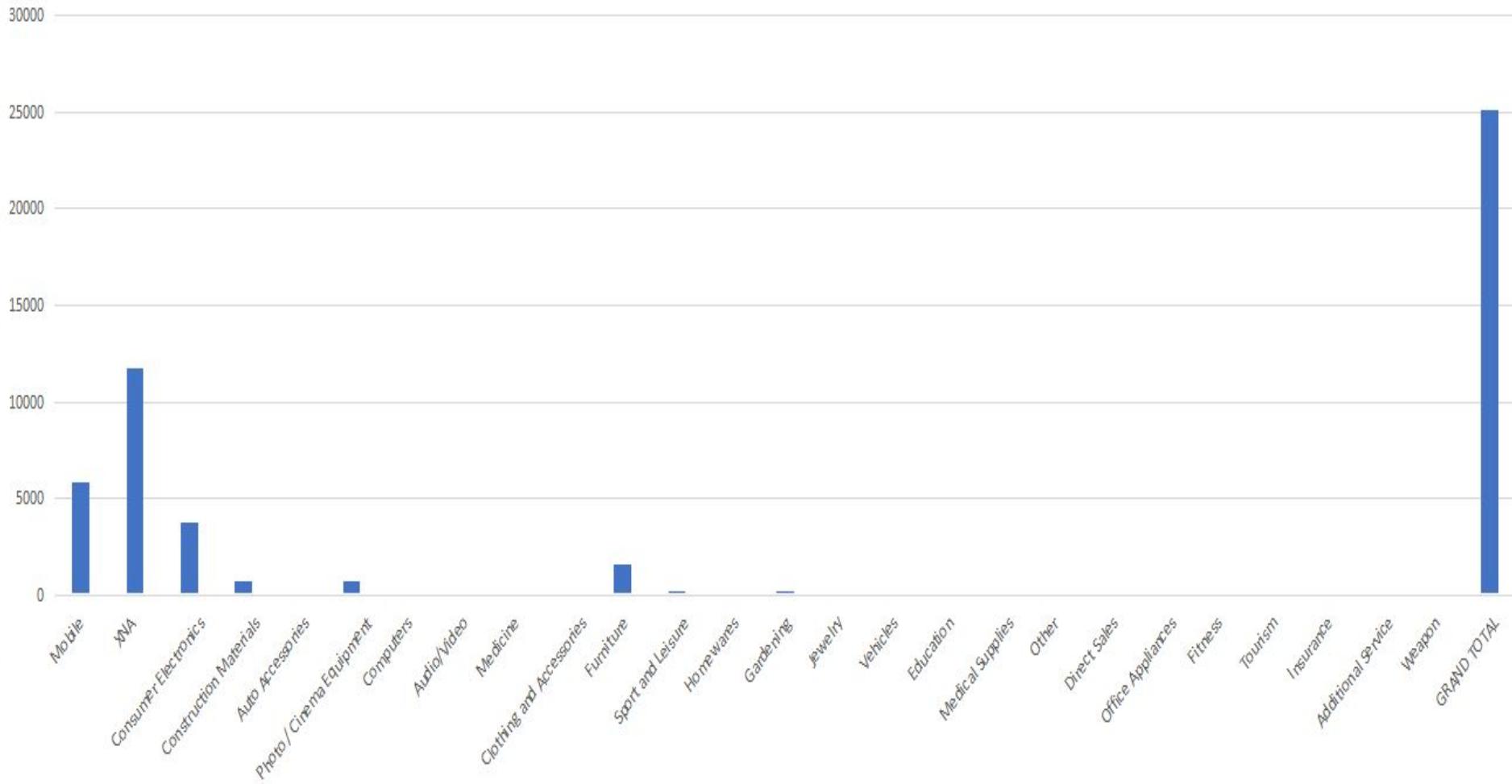
PREVIOUS APPLICATION DATA SETS USERS		
VEHICLES OWNED	Loan Approved	Loan Refused
Mobile	5881	682
XNA	11734	11734
Consumer Electronics	3722	1
Construction Materials	747	0
Auto Accessories	20	0
Photo / Cinema Equipment	704	0
Computers	3	2758
Audio/Video	2	2898
Medicine	0	51
Clothing and Accessories	0	716
Furniture	1559	1559
Sport and Leisure	82	82
Homewares	136	136
Gardening	79	79
Jewelry	152	152
Vehicles	93	7
Education	6	1
Medical Supplies	130	130
Other	65	3
Direct Sales	2	2
Office Appliances	10	0
Fitness	9	9
Tourism	5	50
Insurance	0	3
Additional Service	0	3
Weapon	2	0
GRAND TOTAL	25143	21056

Loan Refused

■ Mobile	■ XNA
■ Consumer Electronics	■ Construction Materials
■ Auto Accessories	■ Photo / Cinema Equipment
■ Computers	■ Audio/Video
■ Medicine	■ Clothing and Accessories
■ Furniture	■ Sport and Leisure
■ Homewares	■ Gardening
■ Jewelry	■ Vehicles
■ Education	■ Medical Supplies
■ Other	■ Direct Sales
■ Office Appliances	■ Fitness
■ Tourism	■ Insurance
■ Additional Service	■ Weapon
■ GRAND TOTAL	



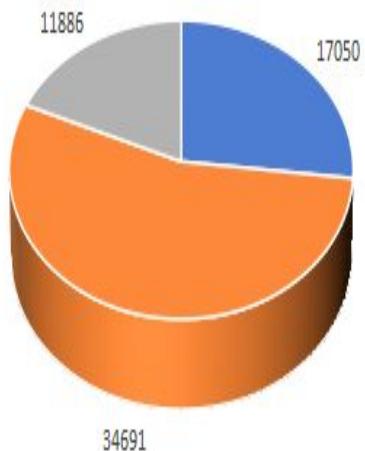
Loan Approved



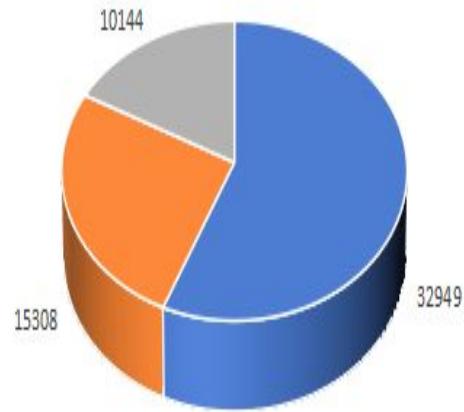
APPLICATION DATA SETS USERS

CUSTOMERS	YES	NO
CAR OWNED	17050	32949
REALTY OWNED	34691	15308
CAR AND REALTY BOTH OWNED	11886	10144

YES



NO



■ CAR OWNED ■ REALTY OWNED ■ CAR AND REALTY BOTH OWNED

■ CAR OWNED ■ REALTY OWNED ■ CAR AND REALTY BOTH OWNED

Tasks

D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:

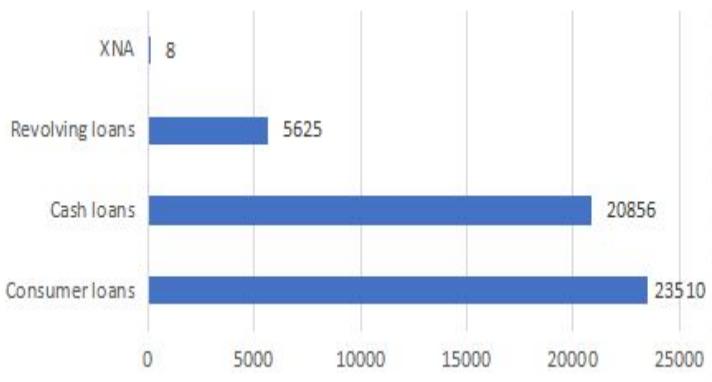
- Task: Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.
- Hint: Utilize Excel functions like COUNT, AVERAGE, MEDIAN, and statistical functions for descriptive analysis. Utilize Excel features like filters, sorting, and pivot tables for segmented and bivariate analysis.
- Graph suggestion: Create histograms, bar charts, or box plots to visualize the distributions of variables. Create stacked bar charts or grouped bar charts to compare variable distributions across different scenarios. Create scatter plots or heatmaps to visualize the relationships between variables and the target variable.

UNIVARIATE ANALYSIS

PREVIOUS_APPLICATION_DATASETS	
CONTRACT TYPE	NO OF CUSTOMERS
Consumer loans	23510
Cash loans	20856
Revolving loans	5625
XNA	8
GRAND TOTAL	49999

UNIVARIATE ANALYSIS

PREVIOUS_APPLICATION_DATASETS
NO OF CUSTOMERS

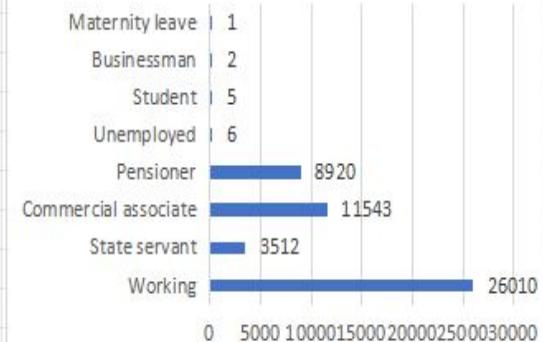


UNIVARIATE ANALYSIS

PREVIOUS_APPLICATION_DATASETS	
INCOME TYPE	NO OF CUSTOMERS
Working	26010
State servant	3512
Commercial associate	11543
Pensioner	8920
Unemployed	6
Student	5
Businessman	2
Maternity leave	1
GRAND TOTAL	49999

UNIVARIATE ANALYSIS

PREVIOUS_APPLICATION_DATABASE
NO OF CUSTOMERS



BIVARIATE ANALYSIS

PREVIOUS APPLICATION DATASETS

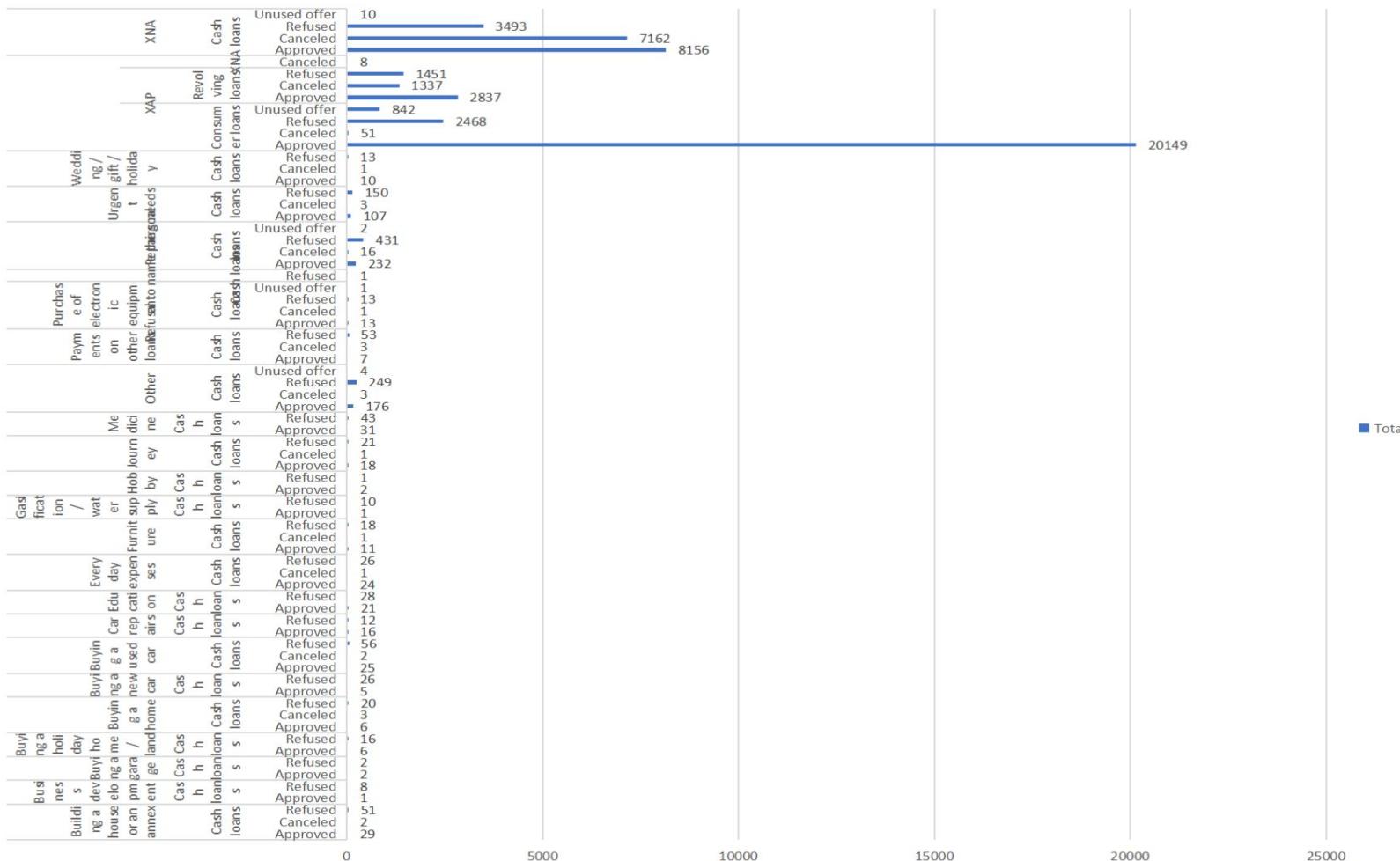
NAME_CASH_LOAN_PURPOSE	NAME_CONTRACT_TYPE	NAME_CONTRACT_STATUS	Count of NAME_CONTRACT_STATUS
Building a house or an annex	Cash loans	Approved	82
		Canceled	82
		Refused	29
Business development	Cash loans	Approved	2
		Refused	51
Buying a garage	Cash loans	Approved	9
		Refused	9
Buying a holiday home / land	Cash loans	Approved	1
		Refused	8
		Refused	9
Buying a home	Cash loans	Approved	4
		Refused	4
		Refused	2
Buying a new car	Cash loans	Approved	2
		Refused	22
		Refused	22
Buying a used car	Cash loans	Approved	6
		Refused	16
		Refused	29
Car repairs	Cash loans	Approved	29
		Refused	6
		Refused	16
		Refused	20
		Refused	29
		Refused	31
		Refused	31
		Refused	5
		Refused	26
		Approved	83
		Approved	83
		Approved	25
		Approved	2
		Approved	56
		Approved	28
		Approved	28
		Approved	16
		Refused	13

	<input type="checkbox"/> Cash loans		49
		Approved	21
		Refused	28
<input type="checkbox"/> Everyday expenses			51
	<input type="checkbox"/> Cash loans		51
		Approved	24
		Canceled	1
		Refused	26
<input type="checkbox"/> Furniture			30
	<input type="checkbox"/> Cash loans		30
		Approved	11
		Canceled	1
		Refused	18
<input type="checkbox"/> Gasification / water supply			11
	<input type="checkbox"/> Cash loans		11
		Approved	1
		Refused	10
<input type="checkbox"/> Hobby			3
	<input type="checkbox"/> Cash loans		3
		Approved	2
		Refused	1
<input type="checkbox"/> Journey			40
	<input type="checkbox"/> Cash loans		40
		Approved	18
		Canceled	1
		Refused	21
<input type="checkbox"/> Medicine			74
	<input type="checkbox"/> Cash loans		74
		Approved	31
		Refused	43
<input type="checkbox"/> Other			432
	<input type="checkbox"/> Cash loans		432
		Approved	176
		Canceled	3
		Refused	249
		Unused offer	4

■ Payments on other loans			63
	■ Cash loans		63
		Approved	7
		Canceled	3
		Refused	53
■ Purchase of electronic equipment			28
	■ Cash loans		28
		Approved	13
		Canceled	1
		Refused	13
		Unused offer	1
■ Refusal to name the goal			1
	■ Cash loans		1
		Refused	1
■ Repairs			681
	■ Cash loans		681
		Approved	232
		Canceled	16
		Refused	431
		Unused offer	2
■ Urgent needs			260
	■ Cash loans		260
		Approved	107
		Canceled	3
		Refused	150
■ Wedding / gift / holiday			24
	■ Cash loans		24
		Approved	10
		Canceled	1
		Refused	13
■ XAP			29143
	■ Consumer loans		23510
		Approved	20149
		Canceled	51
		Refused	2468
		Unused offer	842

	Revolving loans		5625
	Approved		2837
	Canceled		1337
	Refused		1451
	XNA		8
	Canceled		8
XNA			18821
	Cash loans		18821
	Approved		8156
	Canceled		7162
	Refused		3493
	Unused offer		10
Grand Total			49999

Total



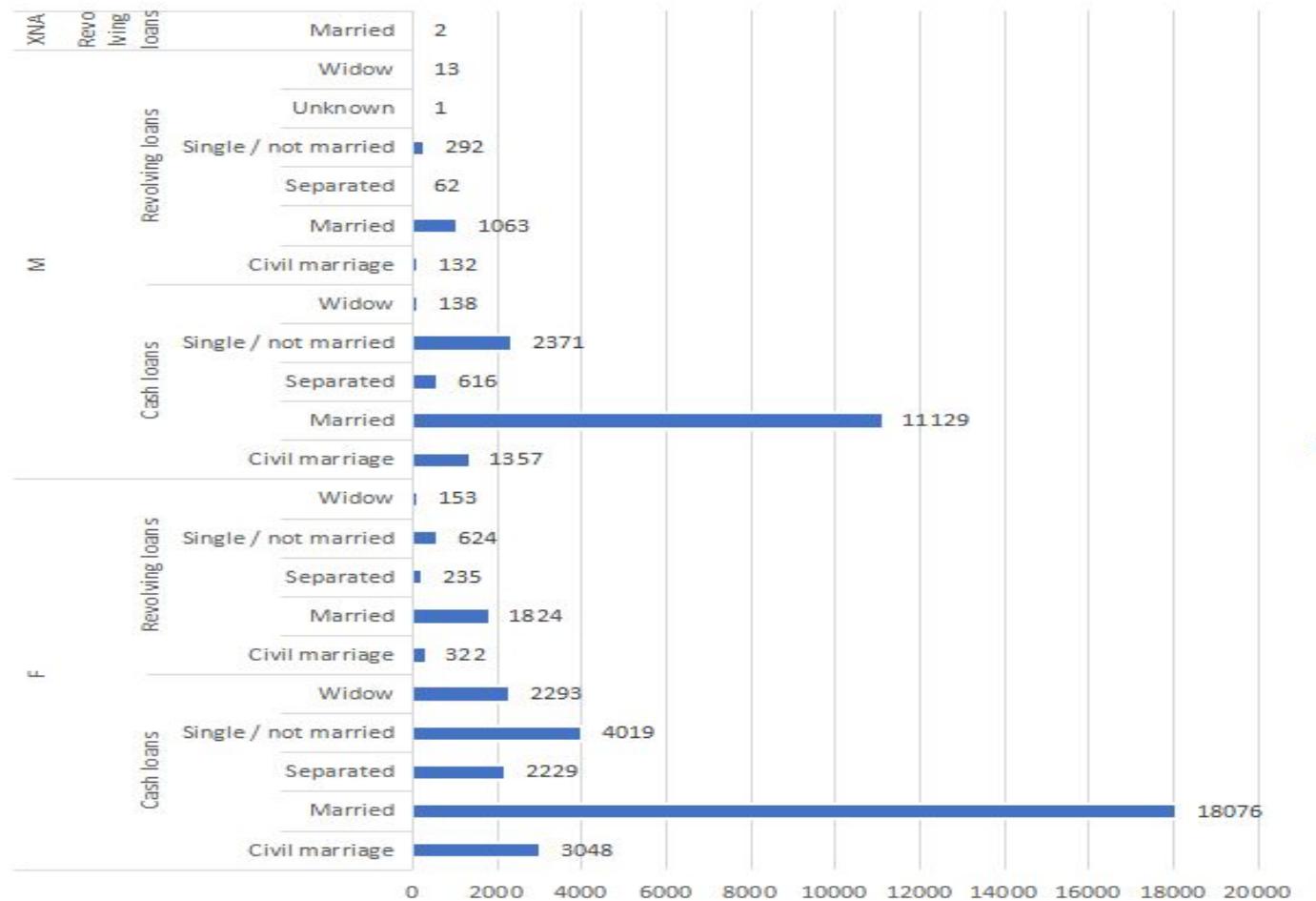
BIVARIATE ANALYSIS

PREVIOUS_APPLICATION_DATASETS			
CODE_GENDER	NAME_CONTRACT_TYPE	NAME_FAMILY_STATUS	Count of FLAG_OWN_CAR
✉ F	✉ Cash loans	Civil marriage	32823
		Married	29665
		Separated	3048
		Single / not married	18076
		Widow	2229
	✉ Revolving loans	Civil marriage	4019
		Married	2293
		Separated	3158
		Single / not married	322
		Widow	1824
		Civil marriage	235
		Married	624
		Separated	235
		Single / not married	153
		Widow	17174
✉ M	✉ Cash loans	Civil marriage	15611
		Married	1357
		Separated	11129
		Single / not married	616
		Widow	2371
	✉ Revolving loans	Civil marriage	138
		Married	1563
		Separated	132
		Single / not married	11063
		Unknown	62
		Widow	292
		Civil marriage	1
		Married	13
✉ XNA	✉ Revolving loans	Separated	2
		Single / not married	2
		Unknown	2
		Widow	49999
Grand Total		Married	49999

Count of FLAG_OWN_CAR

Total

CODE_GENDER
 NAME_CONTRACT_TYPE
 NAME_FAMILY_STATUS



Tasks

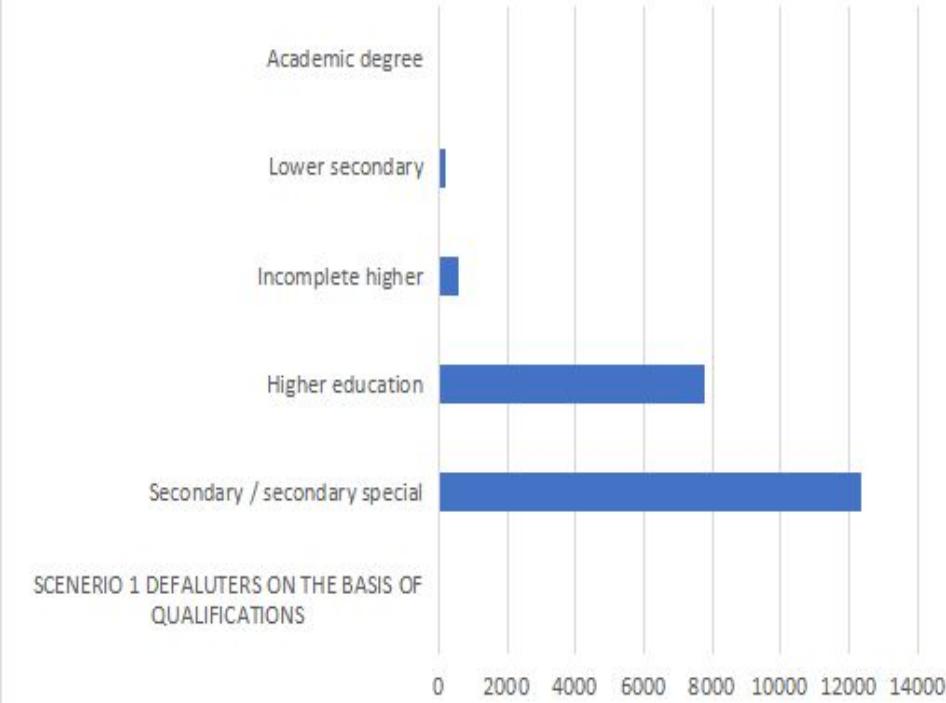
E. Identify Top Correlations for Different Scenarios:

- Task: Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.
- Hint: Utilize Excel functions like CORREL to calculate correlation coefficients between variables and the target variable within each segment. Rank the correlations to identify the top indicators of loan default for each scenario.
- Graph suggestion: Create correlation matrices or heatmaps to visualize the correlations between variables within each segment. Highlight the top correlated variables for each scenario using different colors or shading.

SCENERIO 1 DEFALUTERS ON THE BASIS OF QUALIFICATIONS

Secondary / secondary special	12333
Higher education	7793
Incomplete higher	557
Lower secondary	205
Academic degree	6
GRAND TOTAL	20894

Chart Title



SCENERIO 1 DEFALUTERS ON THE BASIS OF QUALIFICATIONS

SCENERIO 1 DEFALUTERS ON THE BASIS OF QUALIFICATIONS	Secondary / secondary special	Higher education	Incomplete higher	Lower secondary	Academic degree
Series1	12333	7793	557	205	6

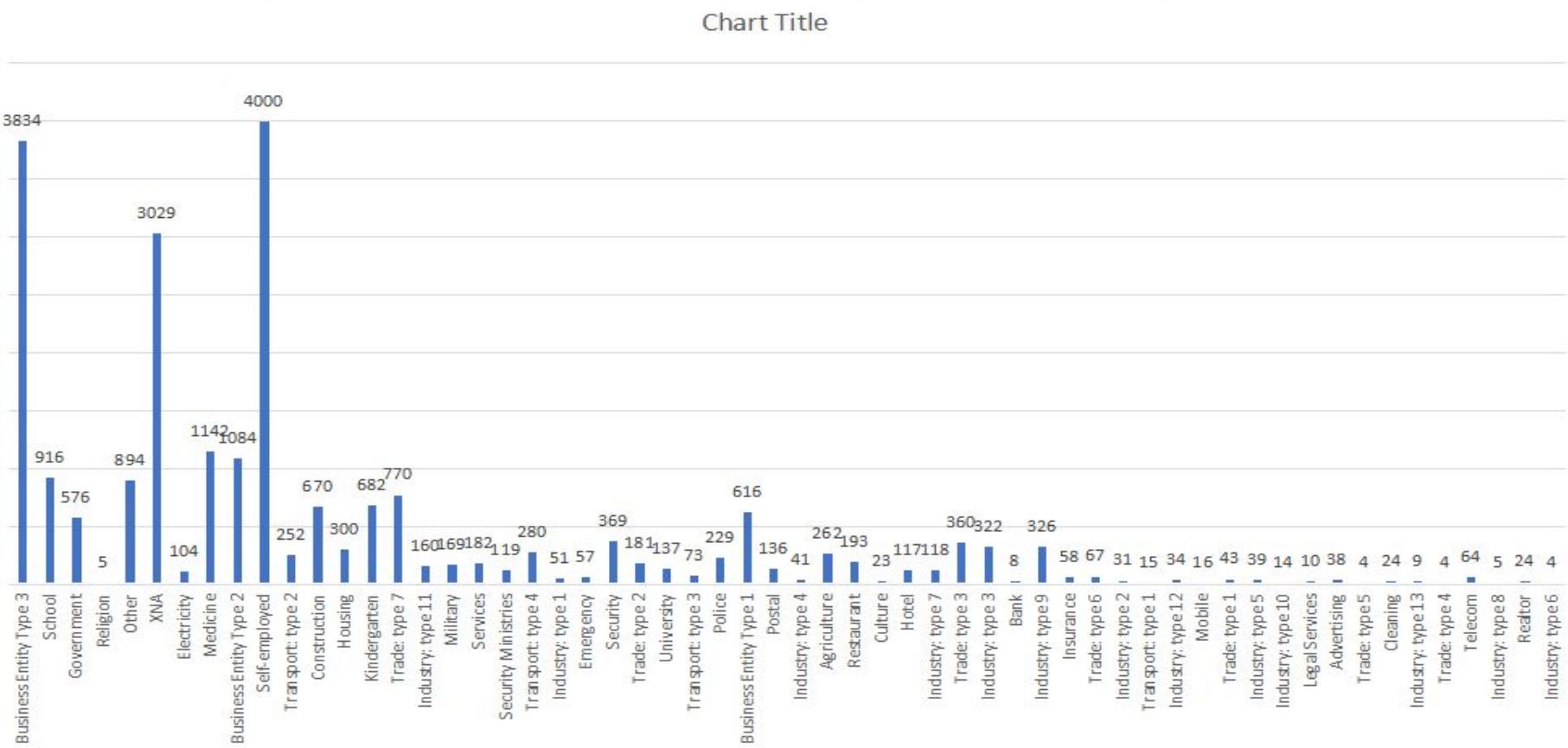
SCENARIO 2 DEFALUTERS ON THE BASIS OF ORGANIZATION TYPE

Business Entity Type 3	3834
School	916
Government	576
Religion	5
Other	894
XNA	3029
Electricity	104
Medicine	1142
Business Entity Type 2	1084
Self-employed	4000
Transport: type 2	252
Construction	670
Housing	300
Kindergarten	682
Trade: type 7	770
Industry: type 11	160
Military	169
Services	182
Security Ministries	119
Transport: type 4	280
Industry: type 1	51
Emergency	57
Security	369
Trade: type 2	181
University	137
Transport: type 3	73
Police	229
Business Entity Type 1	616
Postal	136

Industry: type 4

Agriculture	262
Restaurant	193
Culture	23
Hotel	117
Industry: type 7	118
Trade: type 3	360
Industry: type 3	322
Bank	8
Industry: type 9	326
Insurance	58
Trade: type 6	67
Industry: type 2	31
Transport: type 1	15
Industry: type 12	34
Mobile	16
Trade: type 1	43
Industry: type 5	39
Industry: type 10	14
Legal Services	10
Advertising	38
Trade: type 5	4
Cleaning	24
Industry: type 13	9
Trade: type 4	4
Telecom	64
Industry: type 8	5
Realtor	24
Industry: type 6	4
GRAND TOTAL	23290

Chart Title



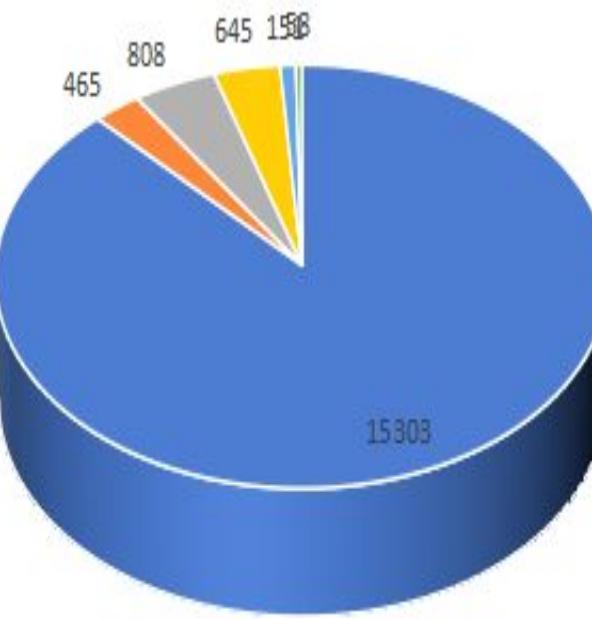
→ In this scenario, we have identified those who are self employed are the most loan defaulters in terms or rejection or cancellation.

→ For the least defaulter in organization type scenario then trade type 5 and Industry type 6 people have least no in default.

SCENARIO 3 DEFALUTERS ON THE BASIS OF HOUSING TYPE

House / apartment	15303
Rented apartment	465
With parents	808
Municipal apartment	645
Office apartment	151
Co-op apartment	58
GRAND TOTAL	17430

Chart Title

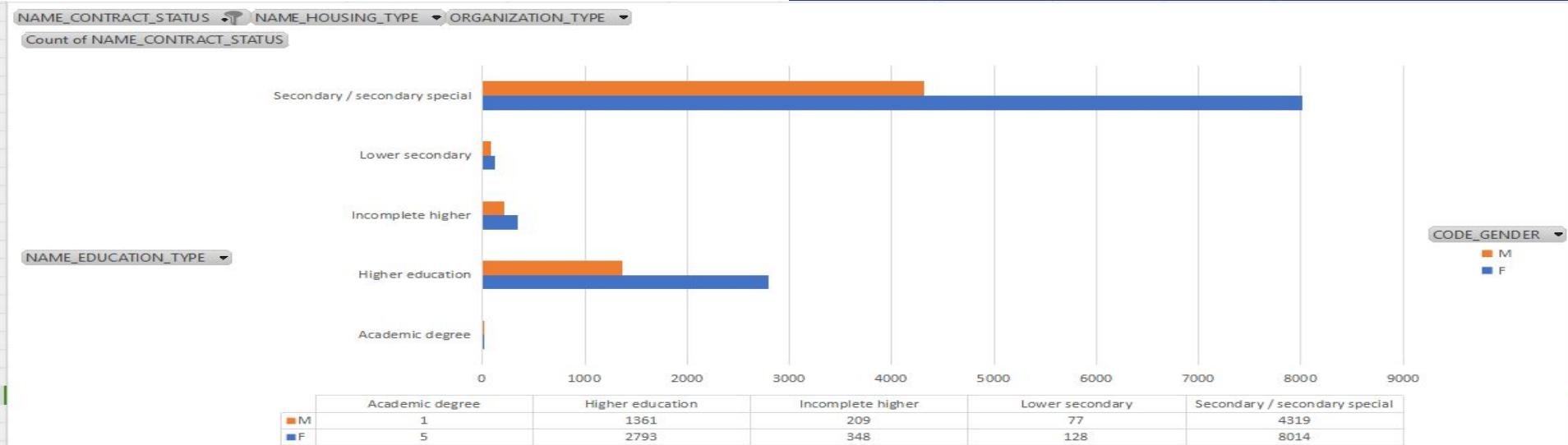


- In this scenario, we have identified those who have house/apartment as house type had their loan applications refused or canceled due to being more prone to default.
- Those who are having co-op apartment are less defaulter in the same scenario.

■ House / apartment ■ Rented apartment ■ With parents
■ Municipal apartment ■ Office apartment ■ Co-op apartment

NAME_CONTRACT_STATUS	Refused	T	
NAME_HOUSING_TYPE	(ALL)	▼	
ORGANIZATION_TYPE	(ALL)	▼	
Count of NAME_CONTRACT_STATUS			
NAME_EDUCATION_TYPE	F	M	
		Grand Total	
Academic degree	5	1	6
Higher education	2793	1361	4154
Incomplete higher	348	209	557
Lower secondary	128	77	205
Secondary / secondary special	8014	4319	12333
Grand Total	11288	5967	17255

- Here we have correlated the scenarios between "Name Contract Status, Name Housing Type, Organization Type, Name Education Type, and Gender."
- From here by using filters we can see the refused and approved loans on the basis of different scenarios like "Name Contract Status, Name Housing Type, Organization Type, Name Education Type, and Gender."



Conclusion

- This document "BANK LOAN CASE STUDY" focuses on the practical application of Exploratory Data Analysis (EDA) in the context of risk analytics within the banking and financial sectors.
- It emphasizes the importance of utilizing EDA techniques to mitigate the risk associated with lending to customers, ensuring that deserving loan applicants are not unfairly rejected.
- The tasks involved in the analysis include identifying missing data, dealing with outliers, analyzing data imbalance, performing univariate, segmented univariate, and bivariate analysis, and identifying top correlations for different scenarios.
- The document highlights the use of Microsoft Excel for efficient data cleaning and analysis, and the utilization of functions such as QUARTILE, IQR, and statistical functions for identifying missing data, outliers, and performing various types of data analysis.
- It also provides insights into segmenting the dataset based on different scenarios and identifying the top correlations for each segmented data using Excel functions.

THANK YOU

