Bank Loan Case Study: Exploratory Data Analysis Report

Introduction

Hello everyone, my name is Hritik Kumar Dutta, and I'm excited to present the findings of my project on the Bank Loan Case Study. This project was assigned to me as part of the data analytics course I'm currently pursuing with Trainity, an edtech platform dedicated to empowering learners with practical data skills.

Project Description

The objective of this project was to analyze a d

ataset containing loan applications from urban customers. Our company, specializing in financial services, faces the challenge of accurately predicting loan defaults to minimize financial losses while maximizing business opportunities. By identifying patterns and factors that influence loan default, we aim to improve our decision-making process regarding loan approvals.

Approach

In tackling this project, I followed a structured approach:

1. Data Cleaning and Preprocessing: I began by identifying missing data and outliers in the dataset, ensuring data integrity through appropriate handling techniques.

2. Exploratory Data Analysis (EDA): I conducted a thorough analysis to understand the distribution and relationships between customer attributes and loan attributes.

3. Data Imbalance Analysis: I assessed the distribution of the target variable to understand any imbalances in the dataset.

4. Correlation Analysis: I segmented the dataset based on different scenarios and identified top correlations that indicate loan default.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE
2	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied	Working	Secondary / secondary speci
3	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family	State servant	Higher education
4	100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied	Working	Secondary / secondary speci
5	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied	Working	Secondary / secondary speci
6	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied	Working	Secondary / secondary speci
7	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner	State servant	Secondary / secondary speci
8	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate	Higher education
9	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied	State servant	Higher education
10	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary / secondary speci
11	100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied	Working	Secondary / secondary speci
12	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied	Working	Higher education
13	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children	Pensioner	Secondary / secondary speci
14	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied	Working	Secondary / secondary speci
15	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied	Working	Secondary / secondary speci
16	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied	Working	Secondary / secondary speci
17	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary / secondary speci
18	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / secondary speci
19	100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied	Working	Secondary / secondary speci
20	100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary / secondary speci
21	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher education
22	100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied	Working	Secondary / secondary speci
23	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied	Commercial associate	Secondary / secondary speci
24	100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied	Working	Secondary / secondary speci
25	100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied	Pensioner	Secondary / secondary speci
26	100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied	Working	Secondary / secondary speci
27	100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied	Working	Secondary / secondary speci
28	100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompanied	Working	Secondary / secondary speci
29	100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family	Working	Secondary / secondary speci
30	100033	0	Cash loans	M	Y	Y	0	270000	790830	57676.5	675000	Unaccompanied	State servant	Higher education
31	100034	0	Revolving loans	M	N	Y	0	90000	180000	9000	180000	Unaccompanied	Working	Higher education
32	100035	0	Cash loans	F	N	Y	0	292500	665892	24592.5	477000	Unaccompanied	Commercial associate	Secondary / secondary speci
33	100036	0	Cash loans	F	N	Y	0	112500	512064	25033.5	360000	Family	Working	Secondary / secondary speci
34	100037	0	Cash loans	F	N	N	0	90000	199008	20893.5	180000	Unaccompanied	Working	Secondary / secondary speci
35	100039	0	Cash loans	M	Y	N	1	360000	733315.5	39069	679500	Unaccompanied	Commercial associate	Secondary / secondary speci
36	100040	0	Cash loans	F	N	Y	0	135000	1125000	32895	1125000	Unaccompanied	State servant	Higher education
37	100041	0	Cash loans	F	N	N	0	112500	450000	44509.5	450000	Unaccompanied	Working	Higher education

application_data (1) | columns_description | previous_application | Sheet1 | Null Value Chart | HANDLING MISSING VALUES | FULL DATA W

Tech-Stack Used

Microsoft Excel 2022: This was my primary tool for data cleaning, analysis, and visualization. I utilized functions such as `COUNT`, `ISBLANK`, `QUARTILE`, `CORREL`, and `COUNTIF`, alongside Excel's features like pivot tables, charts, and conditional formatting.

Insights

Handling Missing Data

Missing values were identified using the `ISBLANK` function and imputed with `AVERAGE` or `MEDIAN`, ensuring that the dataset remained unbiased and reliable.

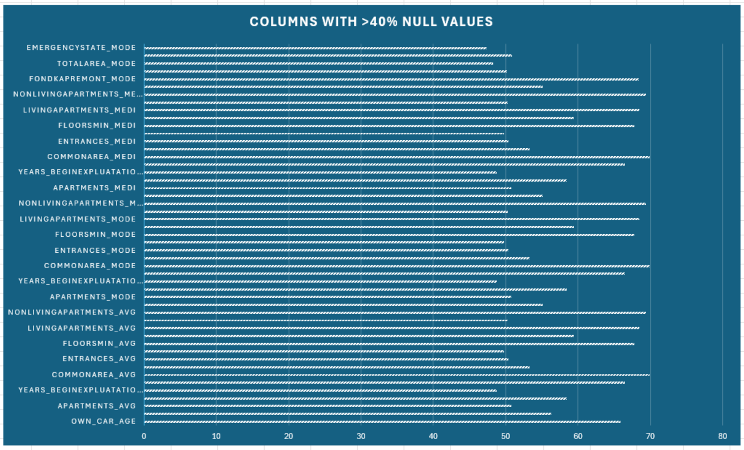
CALCULATING BLANK PERCENTAGE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Blank Count	0	0	0	0	0	0	0	0	0	1	38	192	0
2	Blank %	0	0	0	0	0	0	0	0	0	0.00200004	0.07600152	0.38400768	0
3		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
4		100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied	Working
5		100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family	State servant
6		100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied	Working
7		100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied	Working
8		100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied	Working
9		100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner	State servant
10		100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate
11		100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied	State servant
12		100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner
13		100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied	Working
14		100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied	Working
15		100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children	Pensioner
16		100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied	Working
17		100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied	Working
18		100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied	Working
19		100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working
20		100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied	Working
21		100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied	Working
22		100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A	Working
23		100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied	State servant
24		100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied	Working
25		100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied	Commercial associate
26		100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied	Working
27		100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied	Pensioner
28		100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied	Working
29		100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied	Working
30		100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompanied	Working
31		100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family	Working
32		100033	0	Cash loans	M	Y	Y	0	270000	790830	57676.5	675000	Unaccompanied	State servant
33		100034	0	Revolving loans	M	N	Y	0	90000	180000	9000	180000	Unaccompanied	Working
34		100035	0	Cash loans	F	N	Y	0	292500	665892	24592.5	477000	Unaccompanied	Commercial associate
35		100036	0	Cash loans	F	N	Y	0	112500	512064	25033.5	360000	Family	Working
36		100037	0	Cash loans	F	N	N	0	90000	199008	20893.5	180000	Unaccompanied	Working
37		100039	0	Cash loans	M	Y	N	1	360000	733315.5	39069	679500	Unaccompanied	Commercial associate

application_data (1) | columns_description | previous_application | Sheet1 | Null Value Chart | HANDLING MISSING VALUES | FULL DATA W

COLUMNS WITH > 40% BLANK VALUES

	Column1	Blank_%
27	OWN_CAR_AGE	65.90131803
47	EXT_SOURCE_1	56.3451269
50	APARTMENTS_AVG	50.77101542
51	BASEMENTAREA_AVG	58.39916798
52	YEARS_BEGINEXPLUATATION_AVG	48.78897578
53	YEARS_BUILD_AVG	66.47932959
54	COMMONAREA_AVG	69.92139843
55	ELEVATORS_AVG	53.30306606
56	ENTRANCES_AVG	50.39100782
57	FLOORSMAX_AVG	49.75099502
58	FLOORSMIN_AVG	67.78935579
59	LANDAREA_AVG	59.44318886
60	LIVINGAPARTMENTS_AVG	68.45336907
61	LIVINGAREA_AVG	50.2750055
62	NONLIVINGAPARTMENTS_AVG	69.42938859
63	NONLIVINGAREA_AVG	55.1451029
64	APARTMENTS_MODE	50.77101542
65	BASEMENTAREA_MODE	58.39916798
66	YEARS_BEGINEXPLUATATION_MODE	48.78897578
67	YEARS_BUILD_MODE	66.47932959
68	COMMONAREA_MODE	69.92139843
69	ELEVATORS_MODE	53.30306606
70	ENTRANCES_MODE	50.39100782
71	FLOORSMAX_MODE	49.75099502
72	FLOORSMIN_MODE	67.78935579
73	LANDAREA_MODE	59.44318886
74	LIVINGAPARTMENTS_MODE	68.45336907
75	LIVINGAREA_MODE	50.2750055
76	NONLIVINGAPARTMENTS_MODE	69.42938859
77	NONLIVINGAREA_MODE	55.1451029
78	APARTMENTS_MEDI	50.77101542
79	BASEMENTAREA_MEDI	58.39916798
80	YEARS_BEGINEXPLUATATION_MEDI	48.78897578
81	YEARS_BUILD_MEDI	66.47932959
82	COMMONAREA_MEDI	69.92139843
83	ELEVATORS_MEDI	53.30306606



HANDLING MISSING VALUES

	OCCUPATION_TYPE	NAME_TYPE_SUITE	AMT_ANNUITY	AMT_GOODS_PRIC	CNT_FAM_MEM	EXT_SOURCE_2	EXT_SOURCE_3	OBS_30_CNT_S	DEF_30_CNT_S	OBS_60_CNT_S	DEF_60_CN
NO. OF NULL	15654	192	1	38	1	126	9944	168	168	168	
MEDIAN			24939	450000	2	0.565585366	0.53527625	0	0	0	
5	Laborers	Unaccompanied	24700.5	351000	1	0.262948593	0.13937578	2	2	2	
6	Core staff	Family	35698.5	1129500	2	0.622245775		1	0	1	
7	Laborers	Unaccompanied	6750	135000	1	0.555912083	0.729566691	0	0	0	
8	Laborers	Unaccompanied	29686.5	297000	2	0.65044169		2	0	2	
9	Core staff	Unaccompanied	21865.5	513000	1	0.322738287		0	0	0	
10	Laborers	Spouse, partner	27517.5	454500	2	0.354224732	0.621226338	0	0	0	
11	Accountants	Unaccompanied	41301	1395000	3	0.723999852	0.492060094	1	0	1	
12	Managers	Unaccompanied	42075	1530000	2	0.714279286	0.54065445	2	0	2	
13		Children	33826.5	913500	2	0.205747288	0.751723715	1	0	1	
14	Laborers	Unaccompanied	20250	405000	1	0.746643629		2	0	2	
15	Core staff	Unaccompanied	21177	652500	3	0.651862333	0.363945239	0	0	0	
16		Children	10678.5	135000	2	0.555183162	0.652896552	0	0	0	
17	Laborers	Unaccompanied	5881.5	67500	2	0.715041819	0.176652579	0	0	0	
18	Drivers	Unaccompanied	28966.5	697500	3	0.566906613	0.77008707	0	0	0	
19	Laborers	Unaccompanied	32778	679500	2	0.642656205		0	0	0	
20	Laborers	Family	20160	247500	1	0.346633981	0.678567689	0	0	0	
21	Drivers	Unaccompanied	26149.5	387000	2	0.23637784	0.062103038	0	0	0	
22	Laborers	Unaccompanied	13500	270000	3	0.683513346		4	0	4	
23	Laborers	Other_A	7875	157500	1	0.706428403	0.556727426	8	0	8	
24	Core staff	Unaccompanied	17563.5	454500	2	0.58661714	0.477649155	0	0	0	
25	Laborers	Unaccompanied	21375	427500	2	0.113374513		0	0	0	
26	Sales staff	Unaccompanied	37561.5	927000	3	0.233766958	0.542445144	0	0	0	
27	Sales staff	Unaccompanied	32521.5	450000	3	0.457142972	0.358951229	0	0	0	
28		Unaccompanied	23850	225000	2	0.624304737	0.669056695	0	0	0	
29	Drivers	Unaccompanied	12703.5	247500	4	0.786179309	0.565607981	1	1	1	
30	Cleaning staff	Unaccompanied	11074.5	225000	2	0.651405637	0.461482391	0	0	0	
31	Cooking staff	Unaccompanied	27076.5	702000	1	0.54847716	0.190705948	10	1	10	
32	Laborers	Family	23827.5	270000	3	0.541123702	0.659405532	2	0	2	
33		Unaccompanied	57676.5	675000	1	0.68501099	0.524496446	0	0	0	
34	Laborers	Unaccompanied	9000	180000	1	0.502779038		0	0	0	
35		Unaccompanied	24592.5	477000	2	0.479987342	0.410102573	0	0	0	
36	Private service staff	Family	25033.5	360000	2	0.627300401		2	0	2	

Mode of NAME_TYPE_SUITE	
Unaccompanied	40435
Family	6549
Spouse, partner	1849
Children	542
Other_A	137
0	0
Other_B	259
Group of people	36

NOTE:

COLUMN NAME	NULL VALUES REPLACED WITH
OCCUPATION_TYPE	UNKNOWN
NAME_TYPE_SUITE	Unaccompanied
AMT_ANNUITY	24939
AMT_GOODS_PRICE	450000
CNT_FAM_MEMBERS	2
EXT_SOURCE_2	0.565585366
EXT_SOURCE_3	0.53527625
OBS_30_CNT_SOCIAL_CIRCLE	0
DEF_30_CNT_SOCIAL_CIRCLE	0
OBS_60_CNT_SOCIAL_CIRCLE	0
DEF_60_CNT_SOCIAL_CIRCLE	0
AMT_REQ_CREDIT_BUREAU_HOUR	0
AMT_REQ_CREDIT_BUREAU_DAY	0
AMT_REQ_CREDIT_BUREAU_WEEK	0
AMT_REQ_CREDIT_BUREAU_MON	0
AMT_REQ_CREDIT_BUREAU_QRT	0
AMT_REQ_CREDIT_BUREAU_YEAR	0

FULL DATA WITHOUT MISSING VALUES

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE
2	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompanied	Working	Secondary / secondary speci
3	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family	State servant	Higher education
4	100004	0	Revolving loans	M	Y	Y	0	67500	135000	6750	135000	Unaccompanied	Working	Secondary / secondary speci
5	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompanied	Working	Secondary / secondary speci
6	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompanied	Working	Secondary / secondary speci
7	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, partner	State servant	Secondary / secondary speci
8	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompanied	Commercial associate	Higher education
9	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompanied	State servant	Higher education
10	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary / secondary speci
11	100012	0	Revolving loans	M	N	Y	0	135000	405000	20250	405000	Unaccompanied	Working	Secondary / secondary speci
12	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompanied	Working	Higher education
13	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children	Pensioner	Secondary / secondary speci
14	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompanied	Working	Secondary / secondary speci
15	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompanied	Working	Secondary / secondary speci
16	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompanied	Working	Secondary / secondary speci
17	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary / secondary speci
18	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / secondary speci
19	100021	0	Revolving loans	F	N	Y	1	81000	270000	13500	270000	Unaccompanied	Working	Secondary / secondary speci
20	100022	0	Revolving loans	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary / secondary speci
21	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher education
22	100024	0	Revolving loans	M	Y	Y	0	135000	427500	21375	427500	Unaccompanied	Working	Secondary / secondary speci
23	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompanied	Commercial associate	Secondary / secondary speci
24	100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompanied	Working	Secondary / secondary speci
25	100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompanied	Pensioner	Secondary / secondary speci
26	100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompanied	Working	Secondary / secondary speci
27	100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompanied	Working	Secondary / secondary speci
28	100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompanied	Working	Secondary / secondary speci
29	100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family	Working	Secondary / secondary speci
30	100033	0	Cash loans	M	Y	Y	0	270000	790830	57676.5	675000	Unaccompanied	State servant	Higher education
31	100034	0	Revolving loans	M	N	Y	0	90000	180000	9000	180000	Unaccompanied	Working	Higher education
32	100035	0	Cash loans	F	N	Y	0	292500	665892	24592.5	477000	Unaccompanied	Commercial associate	Secondary / secondary speci
33	100036	0	Cash loans	F	N	Y	0	112500	512064	25033.5	360000	Family	Working	Secondary / secondary speci
34	100037	0	Cash loans	F	N	N	0	90000	199008	20893.5	180000	Unaccompanied	Working	Secondary / secondary speci
35	100039	0	Cash loans	M	Y	N	1	360000	733315.5	39069	679500	Unaccompanied	Commercial associate	Secondary / secondary speci
36	100040	0	Cash loans	F	N	Y	0	135000	1125000	32895	1125000	Unaccompanied	State servant	Higher education
37	100041	0	Cash loans	F	N	N	0	112500	450000	44509.5	450000	Unaccompanied	Working	Higher education

previous_application | Sheet1 | Null Value Chart | HANDLING MISSING VALUES | FULL DATA WITHOUT NULL VALUES | OUTLIERS | DATA_

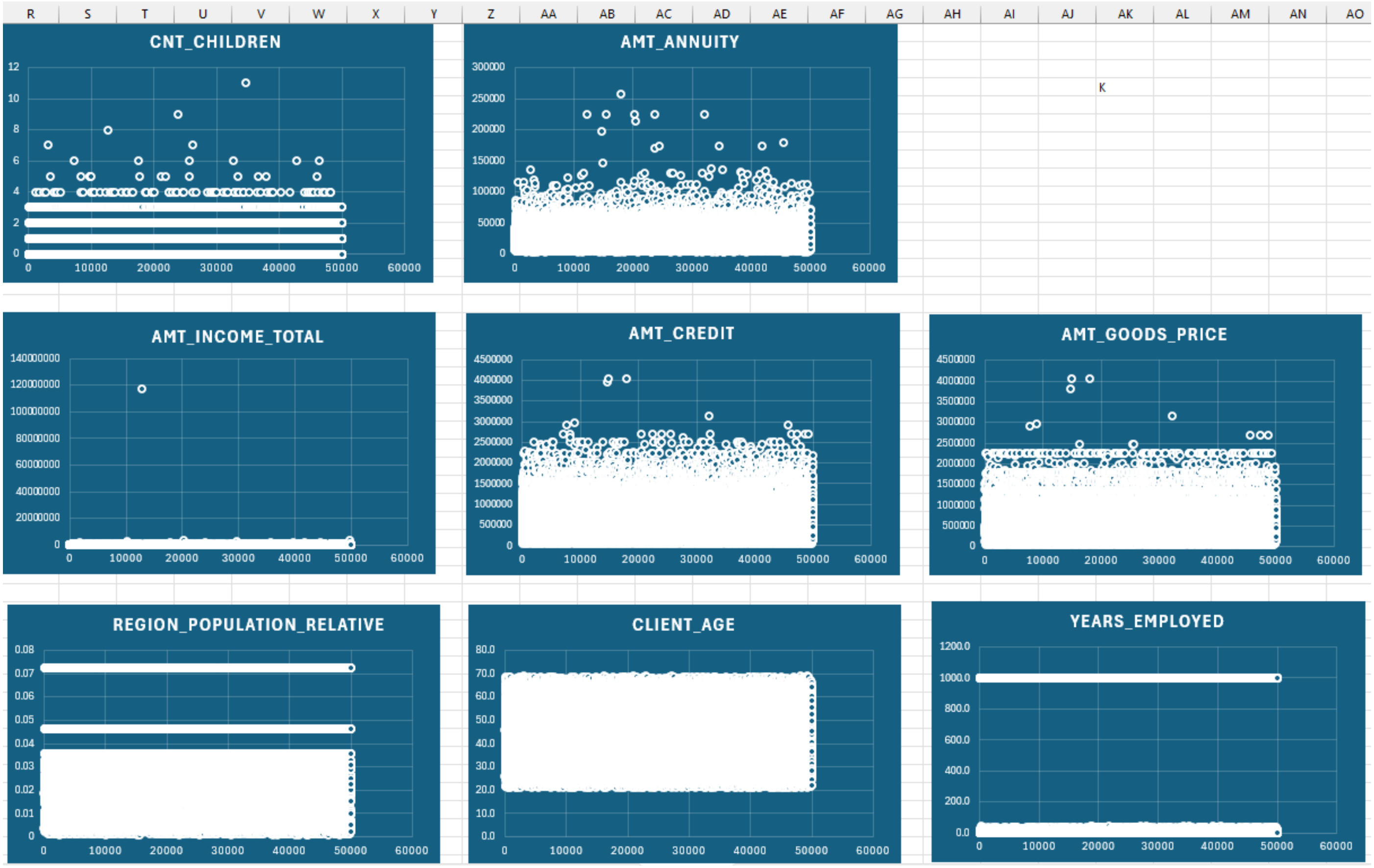
Outlier Detection

Outliers were detected using the IQR method, with box plots illustrating their presence. Validity assessments ensured that the outliers did not distort our analysis.

FINDING QUARANTILE 1, QUARANTILE 3, IQR, ETC.

K	L	M	N	O	P
COLUMN NAME	Q1	Q3	IQR	UPPER LIMIT	LOWER LIMIT
CNT_CHILDREN	0	1	1	2.5	-1.5
AMT_INCOME_TOTAL	112500	202500	90000	337500	-22500
AMT_CREDIT	270000	808650	538650	1616625	-537975
AMT_ANNUITY	16456.5	34596	18139.5	61805.25	-10752.75
AMT_GOODS_PRICE	238500	679500	441000	1341000	-423000
REGION_POPULATION_RELATIVE	0.010006	0.028663	0.018657	0.0566485	-0.0179795
client_age	33.91233	53.81918	19.90685	83.67945205	4.052054795
Years_employed	2.556164	15.66575	13.10959	35.33013699	-17.10821918
Years_Registration	5.473973	20.44795	14.97397	42.90890411	-16.9869863

OUTLIERS SCATTER PLOT

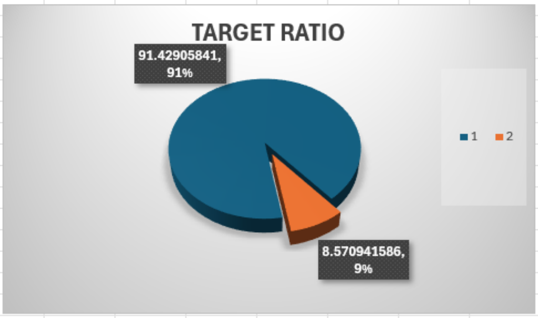


Data Imbalance

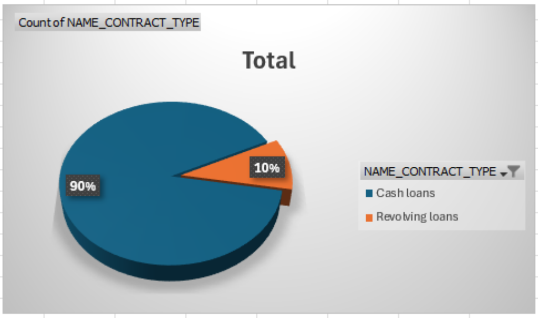
The analysis revealed a significant imbalance, with a higher number of non-defaulters than defaulters. Pie charts were utilized to visualize this imbalance, highlighting the need for balanced data for accurate predictions.

RATIO OF DEFAULTERS & NON-DEFAULETRS

Row Labels	Count of TARGET	PERCENTAGE	CUSTOMER_TYPE
0	37549	91.42905841	NON-DEFAULTERS
1	3520	8.570941586	DEFAULTERS
Grand Total	41069	100	



Row Labels	Count of NAME_CONTRACT_TYPE
Cash loans	36892
Revolving loans	4177
Grand Total	41069



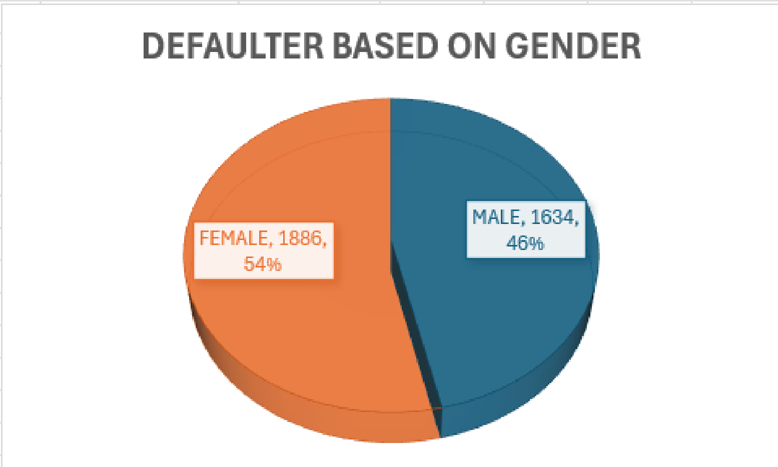
Univariate and Segmented Univariate Analysis

Univariate analysis helped identify key attributes such as income levels and credit history, which emerged as significant indicators of loan defaults. Segmented univariate analysis allowed for comparisons across different customer scenarios, revealing patterns that correlate with default likelihood.

MALE	
NON-DEFAULTER	DEFAULTER
13898	1634

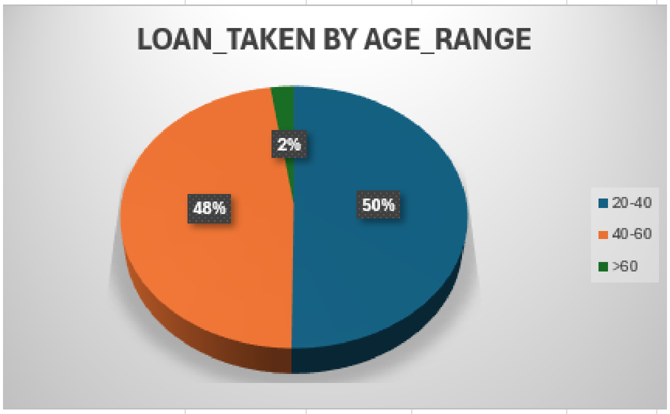
FEMALE	
NON-DEFAULTER	DEFAULTER
23649	1886

LOAN DEFAULTER	
MALE	1634
FEMALE	1886

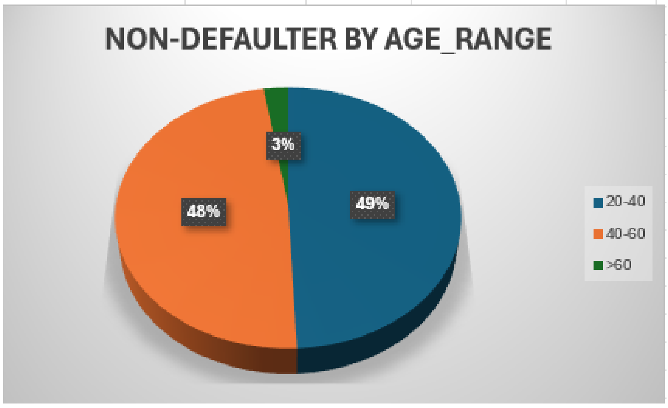


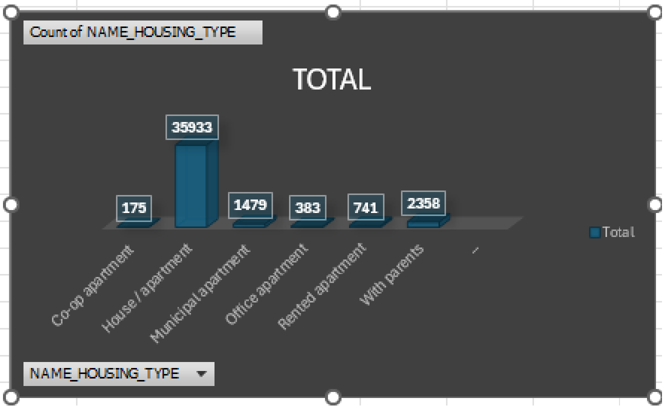
UNIVARIATE ANALYSIS

CLIENT_AGE_RANGE	LOAN_TAKEN	DEFAULTER	NON-DEFAULTER
20-40	20616	2114	18502
40-60	19472	1359	18113
>60	981	47	934



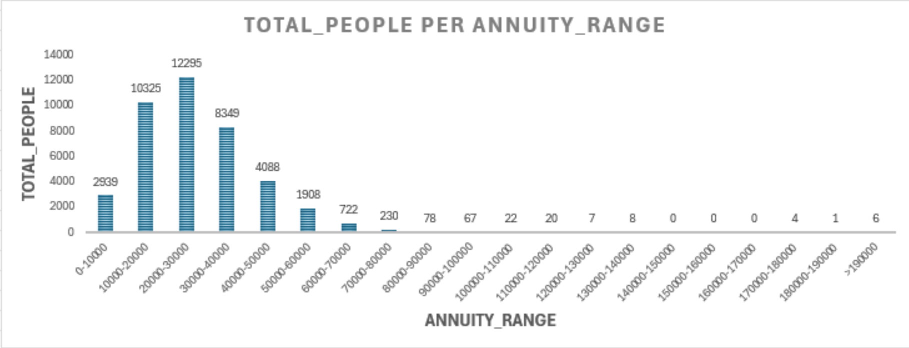
Row Labels	Count of NAME_HOUSING_TYPE
Co-op apartment	175
House / apartment	35933
Municipal apartment	1479
Office apartment	383
Rented apartment	741
With parents	2358
--	
Grand Total	**41069**

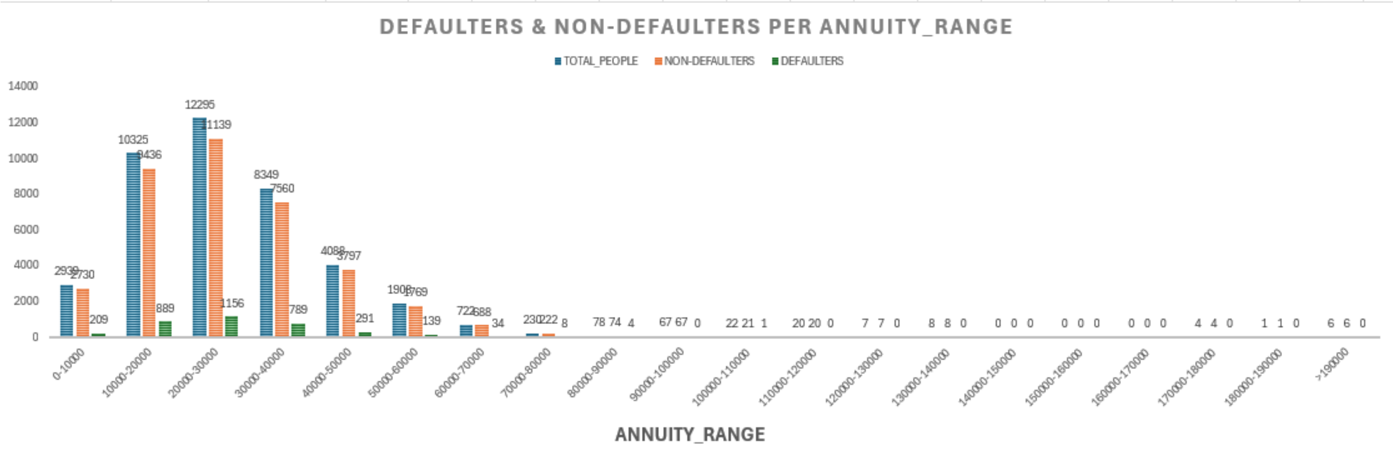




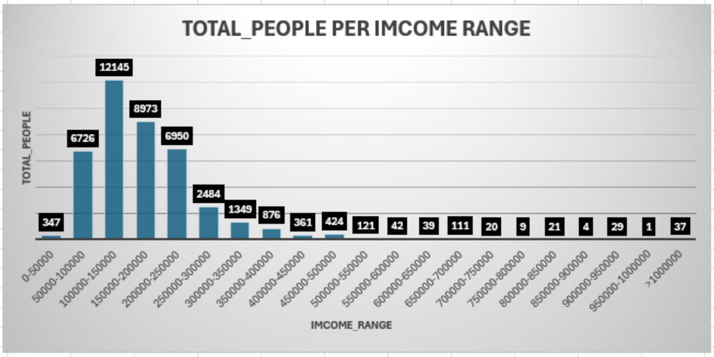
SEGMENTED UNIVARIATE ANALYSIS

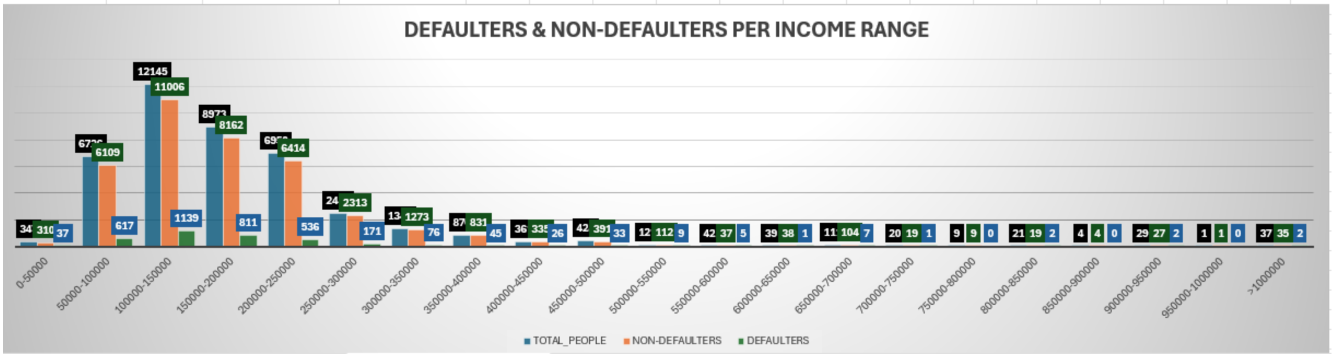
ANNUITY_RANGE	TOTAL_PEOPLE	NON-DEFAULTERS	DEFAULTERS
0-10000	2939	2730	209
10000-20000	10325	9436	889
20000-30000	12295	11139	1156
30000-40000	8349	7560	789
40000-50000	4088	3797	291
50000-60000	1908	1769	139
60000-70000	722	688	34
70000-80000	230	222	8
80000-90000	78	74	4
90000-100000	67	67	0
100000-110000	22	21	1
110000-120000	20	20	0
120000-130000	7	7	0
130000-140000	8	8	0
140000-150000	0	0	0
150000-160000	0	0	0
160000-170000	0	0	0
170000-180000	4	4	0
180000-190000	1	1	0
>190000	6	6	0



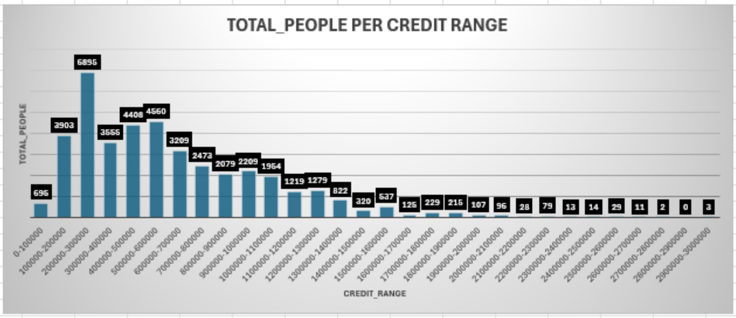


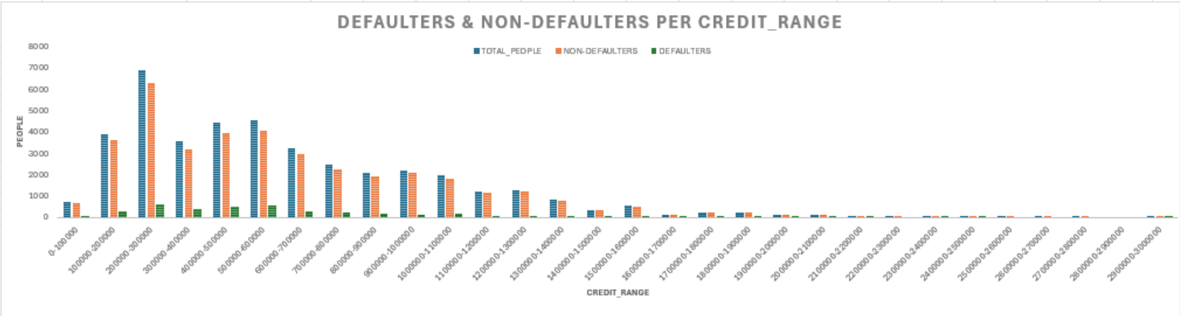
IMCOME_RANGE	TOTAL_PEOPLE	NON-DEFAULTERS	DEFAULTERS
0-50000	347	310	37
50000-100000	6726	6109	617
100000-150000	12145	11006	1139
150000-200000	8973	8162	811
200000-250000	6950	6414	536
250000-300000	2484	2313	171
300000-350000	1349	1273	76
350000-400000	876	831	45
400000-450000	361	335	26
450000-500000	424	391	33
500000-550000	121	112	9
550000-600000	42	37	5
600000-650000	39	38	1
650000-700000	111	104	7
700000-750000	20	19	1
750000-800000	9	9	0
800000-850000	21	19	2
850000-900000	4	4	0
900000-950000	29	27	2
950000-1000000	1	1	0
>1000000	37	35	2





CREDIT_RANGE	TOTAL_PEOI	NON-DEFAULTERS	DEFAULTERS
0-100000	695	648	47
100000-200000	3903	3612	291
200000-300000	6895	6302	593
300000-400000	3555	3158	397
400000-500000	4408	3927	481
500000-600000	4560	4033	527
600000-700000	3209	2943	266
700000-800000	2473	2260	213
800000-900000	2079	1912	167
900000-1000000	2209	2074	135
1000000-1100000	1954	1807	147
1100000-1200000	1219	1143	76
1200000-1300000	1279	1214	65
1300000-1400000	822	779	43
1400000-1500000	320	306	14
1500000-1600000	537	517	20
1600000-1700000	125	116	9
1700000-1800000	229	220	9
1800000-1900000	215	209	6
1900000-2000000	107	102	5
2000000-2100000	96	92	4
2100000-2200000	28	26	2
2200000-2300000	79	79	0
2300000-2400000	13	12	1
2400000-2500000	14	13	1
2500000-2600000	29	29	0
2600000-2700000	11	11	0
2700000-2800000	2	2	0
2800000-2900000	0	0	0
2900000-3000000	3	2	1





Correlation Analysis

Correlation analysis revealed top correlations using the `CORREL` function. For example, poor credit history was strongly correlated with defaults in customers with payment difficulties, providing actionable insights for risk assessment.

Top correlation of Non-Defaulters

Rank	Variable 1	Variable 2	Correlation
1	AMT_GOODS_PRICE	AMT_CREDIT	0.98635817
2	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950286525
3	CNT_CHILDREN	CNT_FAM_MEMBERS	0.893735596
4	REG_REGION_NOT_WORK_REGION -	LIVE_REGION_NOT_WORK_REGION	0.860167703
5	DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.853040752
6	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.815604978
7	REGION_RATING_CLIENT	AMT_GOODS_PRICE	0.765201743
8	AMT_ANNUITY	AMT_GOODS_PRICE	0.765201743
9	AMT_CREDIT	AMT_ANNUITY	0.760827873

Top Correlation of Defaulters

Rank	Variable 1	Variable 2	Correlation
1	AMT_GOODS_PRICE	AMT_CREDIT	0.981928143
2	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.948020808
3	CNT_FAM_MEMBERS	CNT_CHILDREN	0.895600339
4	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.891467244
5	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.805583225
6	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.773107352
7	AMT_ANNUITY	AMT_GOODS_PRICE	0.746422447
8	AMT_ANNUITY	AMT_CREDIT	0.745132112

Results

This project enhanced my understanding of the factors contributing to loan defaults. The insights gained will inform strategies to identify high-risk applicants, adjust loan offerings, and optimize interest rates, ultimately strengthening our company's financial performance. Through this project with Trainity, I've honed my data analytics skills, applying them to real-world challenges.

Drive Link

https://drive.google.com/drive/folders/1PXMUeNplewfeykrXsrqTYzbHUxlFcltm?usp=sharing