

Bank Loan Case Study

Project Description

Objective:

To analyze patterns in loan application data to identify factors influencing loan defaults.

Key Risks:

- Approving loans for customers who cannot repay.
- Rejecting loans for customers who can repay.

Goal:

Enable data-driven decisions for loan approval based on customer and loan attributes.

Approach

Data Preparation:

- **Dataset Inspection:** Downloaded and reviewed the dataset structure.

AutoSave

akriti application_data.xlsx - Saved to this PC

Search

FileHomeInsertPage LayoutFormulasDataReviewViewAutomateDeveloperHelpAcrobatPower PivotTable Design

Clipboard

Font

Alignment

Number

Styles

Cells

Editing

Add-ins

Comments

Share

Constanta

11

A

A

Wrap Text

General

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

Σ

Sort & Find

Filter & Select

Add-ins

Analyze Data

Create a PDF and Share link

Create a PDF and Share link

G21

fx

Y

	B	C	D	E	F	G	H	I	J	K	L	M	N	O																				
	B	C	D	E	F	G	H	I	J	K	L	M	N	O																				
1	0.00%	0.00%		0.00%	0.00%		0.00%		0.00%	0.00%	0.00%	0.00%	0.33%	0.00%																				
2																																		
3	SK ID	CURE	TARGE	NAME	CONTRACT TYP	CODE	GENDE	FLAG	OWN	CAR	FLAG	OWN	REALTY	CNT	CHILDRE	AMT	INCOME	TOTAL	AMT	CREDIT	AMT	ANNUITY	AMT	GOODS	PRICE	NAME	TYPE	SUM	NAME	INCOME	TYP	NAME	EDUCATION	TYP
4	1000001	1	Cash loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	202300	406597.5	24700.5	35000	Unaccompanied	Working	Secondary / secondary spec												
5	1000003	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	270000	1293502.5	35698.5	1129300	Family	State servant	Higher education												
6	1000004	0	Revolving loans	M	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	67300	132000	6730	133000	Unaccompanied	Working	Secondary / secondary spec												
7	1000006	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	133000	31068.5	20680.5	107000	Unaccompanied	Working	Secondary / secondary spec												
8	1000007	0	Cash loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	131300	533000	21865.5	533000	Unaccompanied	Working	Secondary / secondary spec												
9	1000008	0	Cash loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	99000	490495.5	27307.5	454500	Spouse, partner	State servant	Secondary / secondary spec												
10	1000009	0	Cash loans	F	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	170000	1959726	41001	1035000	Unaccompanied	Commercial associate	Higher education												
11	1000010	0	Cash loans	M	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	360000	13300000	43075	1330000	Unaccompanied	State servant	Higher education												
12	1000011	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	112300	1019600	33826.5	913500	Children	Pensioner	Secondary / secondary spec												
13	1000012	0	Revolving loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	133000	405000	10230	409000	Unaccompanied	Working	Secondary / secondary spec												
14	1000014	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	112300	615100	21277	651500	Unaccompanied	Working	Higher education												
15	1000015	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	38419.555	148565	10678.5	135000	Children	Pensioner	Secondary / secondary spec												
16	1000016	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	67300	80865	388.5	67300	Unaccompanied	Working	Secondary / secondary spec												
17	1000017	0	Cash loans	M	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	233000	918468	28966.5	697500	Unaccompanied	Working	Secondary / secondary spec												
18	1000018	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	185000	77368.5	22728	679500	Unaccompanied	Working	Secondary / secondary spec												
19	1000019	0	Cash loans	M	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	157300	299772	20060	247500	Family	Working	Secondary / secondary spec												
20	1000020	0	Cash loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	108000	509602.5	26149.5	387000	Unaccompanied	Working	Secondary / secondary spec												
21	1000021	0	Revolving loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	80000	370000	13300	270000	Unaccompanied	Working	Secondary / secondary spec												
22	1000022	0	Revolving loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	112300	157300	7875	157300	Other, A	Working	Secondary / secondary spec												
23	1000023	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	90000	544491	17563.5	454500	Unaccompanied	State servant	Higher education												
24	1000024	0	Revolving loans	M	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	133000	427300	21375	427500	Unaccompanied	Working	Secondary / secondary spec												
25	1000025	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	203500	113173.5	37596.5	927000	Unaccompanied	Commercial associate	Secondary / secondary spec												
26	1000026	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	450000	407330	31321.5	450000	Unaccompanied	Working	Secondary / secondary spec												
27	1000027	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	83500	339830	23830	235000	Unaccompanied	Pensioner	Secondary / secondary spec												
28	1000029	0	Cash loans	M	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	133000	247300	12703.5	247500	Unaccompanied	Working	Secondary / secondary spec												
29	1000030	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	90000	225000	10745.5	225000	Unaccompanied	Working	Secondary / secondary spec												
30	1000031	1	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	112300	979992	27076.5	702000	Unaccompanied	Working	Secondary / secondary spec												
31	1000032	0	Cash loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	112300	377024	23827.5	270000	Family	Working	Secondary / secondary spec												
32	1000033	0	Cash loans	M	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	270000	790830	37676.5	675000	Unaccompanied	State servant	Higher education												
33	1000034	0	Revolving loans	M	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	90000	280000	9000	180000	Unaccompanied	Working	Higher education												
34	1000035	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	292500	665832	24592.5	477000	Unaccompanied	Commercial associate	Secondary / secondary spec												
35	1000036	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	112300	512064	25033.5	360000	Family	Working	Secondary / secondary spec												
36	1000037	0	Cash loans	F	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	90000	199008	10692.5	180000	Unaccompanied	Working	Secondary / secondary spec												

<

>

application_data

Handling Missing Values

EDA

Outlier Data

Handling Outliers

Cleaned Dataset

+

-

ReadyAccessibility Insights

86%

- **Handling Missing Values:**

- Used mean/median imputation for numerical variables.

Dealing with Missing Values				
S No.	Columns Replacing Missing Values	Mean	Median	Mode
1	AMT_ANNUITY	27107.14918	24939	
2	AMT_GOODS_PRICE	539052.277	450000	
3	NAME_TYPE_SUITE			Unaccompanied
4	OCCUPATION_TYPE			Laborers
5	CNT_FAM_MEMBERS	2.158726349	2	
6	OBS_30_CNT_SOCIAL_CIRCLE	1.419396776	0	
7	DEF_30_CNT_SOCIAL_CIRCLE	0.141345654	0	
8	OBS_60_CNT_SOCIAL_CIRCLE	1.402336093	0	
9	DEF_60_CNT_SOCIAL_CIRCLE	0.09800392	0	
10	DAYS_LAST_PHONE_CHANGE	-964.3153985	-755	
11	AMT_REQ_CREDIT_BUREAU_HOUR	0.006140246	0	
12	AMT_REQ_CREDIT_BUREAU_DAY	0.00650026	0	
13	AMT_REQ_CREDIT_BUREAU_WEEK	0.028021121	0	
14	AMT_REQ_CREDIT_BUREAU_MON	0.233889356	0	
15	AMT_REQ_CREDIT_BUREAU_QRT	0.225829033	0	
16	AMT_REQ_CREDIT_BUREAU_YEAR	1.627725109	1	

- Applied mode or introduced an 'Unknown' category for categorical variables.

Name Type suite		OCCUPATION_TYPE	
Category	Count	Category	Count
Unaccompanied	40626	Laborers	24606
Family	6549	Core staff	4434
Spouse, partner	1849	Accountants	1621
Children	542	Managers	3488
Other_A	137	Drivers	3044
Other_B	259	Sales staff	5160
Group of people	36	Cleaning staff	739
Blanks	192	Cooking staff	963
		Private service staff	447
		Medicine staff	1403
		Security staff	1140
		High skill tech staff	1852
		Waiters/barmen staff	228
		Low-skill Laborers	357
		Realty agents	123
		Secretaries	212
		IT staff	80
		HR staff	101

- **Outlier Detection:**

- Identified outliers using the Interquartile Range (IQR) method.
- Adjusted outliers within calculated bounds to maintain data integrity.

	A	B	C	D	E	F	G	H
	CNT CHILDREN	AMT INCOME TOTAL	AMT CREDIT	AMT ANNUITY	AMT GOODS PRICE	DAYS BIRTH	DAYS EMPLOYED	DAYS REGISTRATION
2	0	202500	406597.5	24700.5	351000	25.92054795	1.745205479	9.994520548
3	0	270000	1293502.5	35698.5	1129500	45.93150685	3.254794521	3.249315068
4	0	67500	135000	6750	135000	52.18082192	0.616438356	11.67123288
5	0	135000	312682.5	29866.5	297000	52.06849315	8.326027397	26.93972603
6	0	121500	513000	121865.5	513000	54.60821918	8.323287671	11.8109589
7	0	99000	490495.5	27517.5	454500	46.41369863	4.350684932	13.61643836
8	1	171000	1560726	41301	1395000	37.74794521	8.575342466	3.323287671
9	0	369000	1530000	42075	1530000	51.64383562	1.230136986	12.59452055
10	0	112500	1019610	33626.5	913500	55.06578342	1000.665753	20.34794521
11	0	135000	405000	20250	405000	39.64109589	5.531506849	39.55342466
12	1	112500	652500	21177	652500	27.9369863	1.860273973	12.12876712
13	0	38419.155	148365	10678.5	135000	55.9369863	1000.665753	14.37260274
14	0	67500	80865	5881.5	67500	36.81917808	7.443835616	0.852054795
15	1	225000	918468	28966.5	697500	38.59178082	8.295890411	1.761643836
16	0	189000	773680.5	32778	679500	39.95342466	0.556164384	1.684931507
17	0	157500	299772	20160	247500	23.91232877	3.169863014	9.57260274
18	0	108000	509602.5	26149.5	387000	35.42739726	3.608219178	17.51232877
19	1	81000	270000	13500	270000	26.78356164	0.523287671	11.35068493
20	0	112500	157500	7875	157500	48.54246575	21.38082192	23.97534247
21	1	90000	544491	17963.5	454500	31.09041096	5.383561644	2.797260274
22	0	135000	427500	21375	427500	50.06547945	11.74246575	0.816438356
23	1	202500	1132573.5	37561.5	927000	40.5890411	4.526027397	6.298630137
24	1	450000	497520	32521.5	450000	30.5369863	11.79726027	0.312328767
25	0	83250	239850	23850	225000	68.01917808	1000.665753	24.69041096
26	2	135000	247500	12703.5	247500	30.92054795	2.043835616	0.295890411
27	0	90000	225000	11074.5	225000	52.96986301	9.57260274	6.62739726
28	0	112500	979992	27076.5	702000	51.29863014	7.2	18.00821918
29	1	112500	327024	23827.5	270000	43.69315068	3.380821918	15.84109589

Analysis:

- Conducted univariate, segmented univariate, and bivariate analyses.
- Assessed data imbalance and class proportions.
- Computed correlations between variables and target outcomes within segmented data.

Visualization:

- Created histograms, bar charts, pie charts, box plots, and scatter plots.

Reporting:

- Documented insights and findings.
- Linked detailed analysis in Excel sheets for reference.

Tech-Stack Used

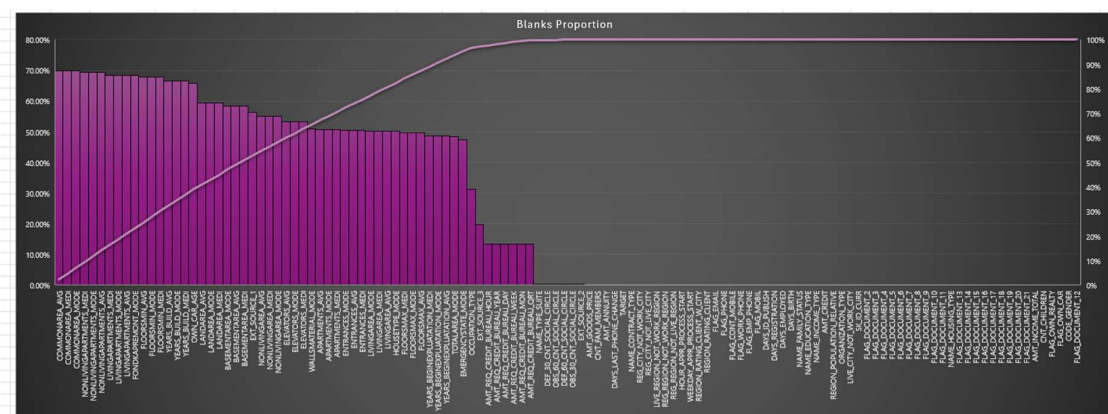
- **Microsoft Excel 2022:** Data preparation, analysis, and visualization.
- **Google Drive:** Hosted Excel files for sharing and integration.
- **Ms Word:** Designed this report.

A. Identify Missing Data and Deal with it Appropriately:

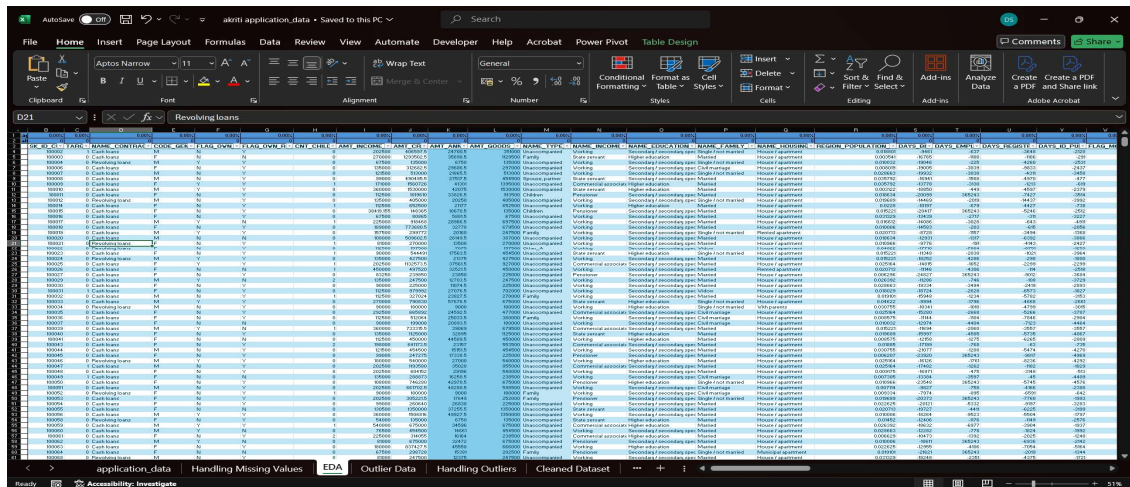
- **Identification:** Counted blank values using the COUNTBLANK() function.

A	B	C	D	E	F
Percentage Value					
Blanks	=countb				
	COUNTBLANK	Counts the number of empty cells in a specified range of cells			
	100002	1	Cash loans	M	N
	100003	0	Cash loans	F	N
	100004	0	Revolving loans	M	Y

- **Resolution:**
 - Numerical variables: Imputed using mean/median.
 - Categorical variables: Used mode or created 'Unknown' category.
- **Visualization:** Created Pareto charts showing the percentage of missing data by column.



- Columns with more than 50% missing data were deleted.



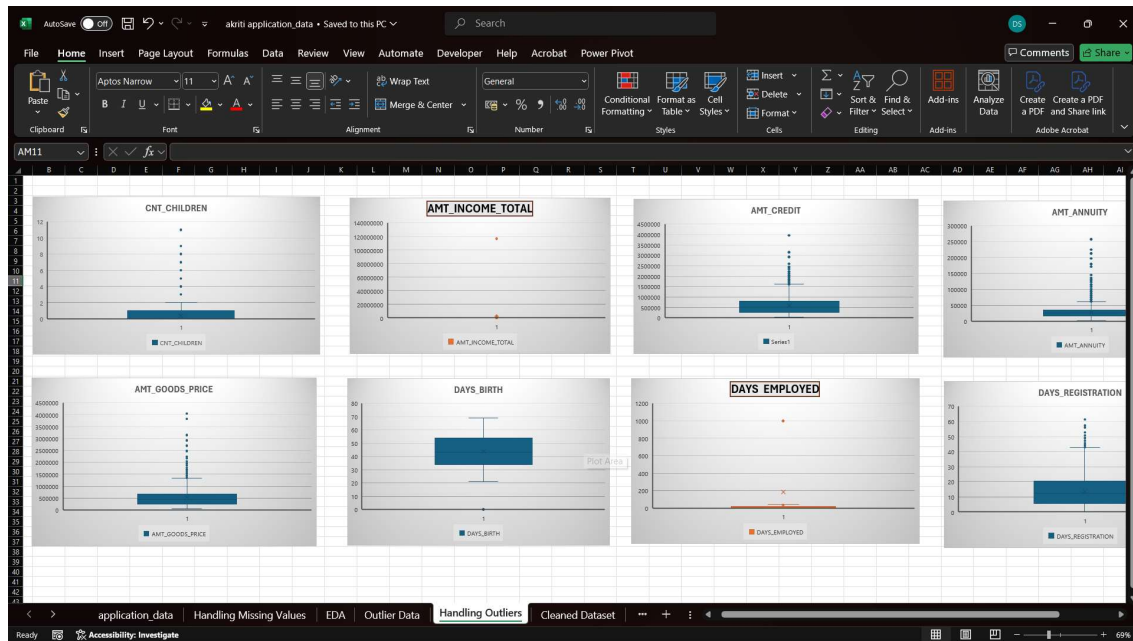
B. Identify Outliers in the Dataset:

Method:

- Calculated Q1, Q3, and IQR.
- Defined bounds: $Q1 - 1.5IQR$ and $Q3 + 1.5IQR$.
- Adjusted values beyond bounds.

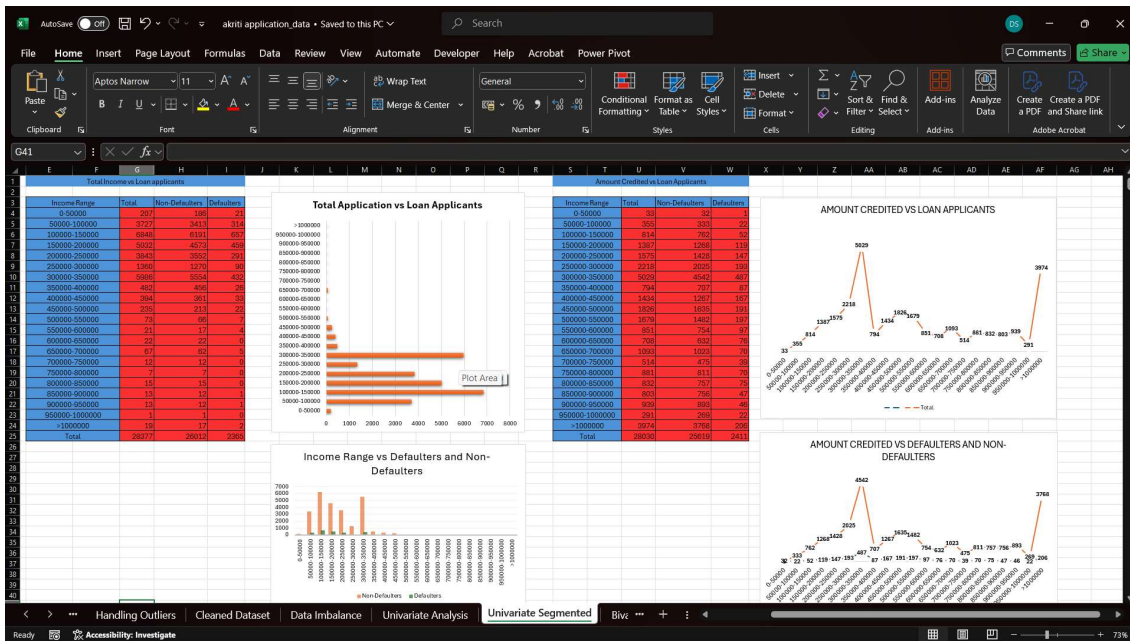
	Q1	Median(Q2)	Q3	InterQuartile Range	Lower Bound	Upper Bound	Minimum	Maximum	Mean	Mode	Range	Standard D	Variance	Sum	Count	Standard Error
CNT_CHILDREN	0	0	1	1	-1.5	2.5	0	11	0.4198484	0	11	0.7240385	0.524231818	20992	49999	0.003238031
AMT_INCOME_TOTAL	112500	145800	202500	90000	-22500	337500	25650	117000000	170767.59	135000	116974350	531819.1	2.82832E+11	8538208758	49999	2378.391081
AMT_CREDIT	270000	514777.5	808650	538650	-537975	1616625	45000	4050000	599700.58	450000	4005000	402415.43	1.61938E+11	29984429376	49999	1799.674528
AMT_ANNUITY	16456.5	24939	34596	18139.5	-10752.75	61805.25	2052	258025.5	27107.377	9000	255973.5	14562.799	212075108.9	1355341760	49999	65.12746739
AMT_GOODS_PRICE	238500	450000	679500	441000	-423000	1341000	45000	4050000	539060.04	450000	4005000	369712.68	1.36687E+11	26952462745	49999	1653.421886
DAYS_BIRTH	33.91232877	43.0986301	53.81918	19.90684932	4.052054795	83.67945205	0	68.9972603	43.895057	30.24384	68.9972603	11.950644	142.8178904	2194708.97	49999	0.053445439
DAYS_EMPLOYED	2.556164384	6.07123288	15.66575	13.10958904	-17.1082192	35.33013699	0	1000.66575	184.00088	1000.666	1000.66575	380.70657	144937.4945	9199859.893	49999	1.702588578
DAYS_REGISTRATION	5.47260274	12.3013699	20.44795	14.97534247	-16.990411	42.9109589	0	61.3479452	13.636175	0.008219	61.3479452	9.6592206	93.30054278	681795.1123	49999	0.04319778

- **Example:** A record with 1000 years of employment was identified as an error and removed.
- **Visualization:** Used box plots to identify and address outliers.



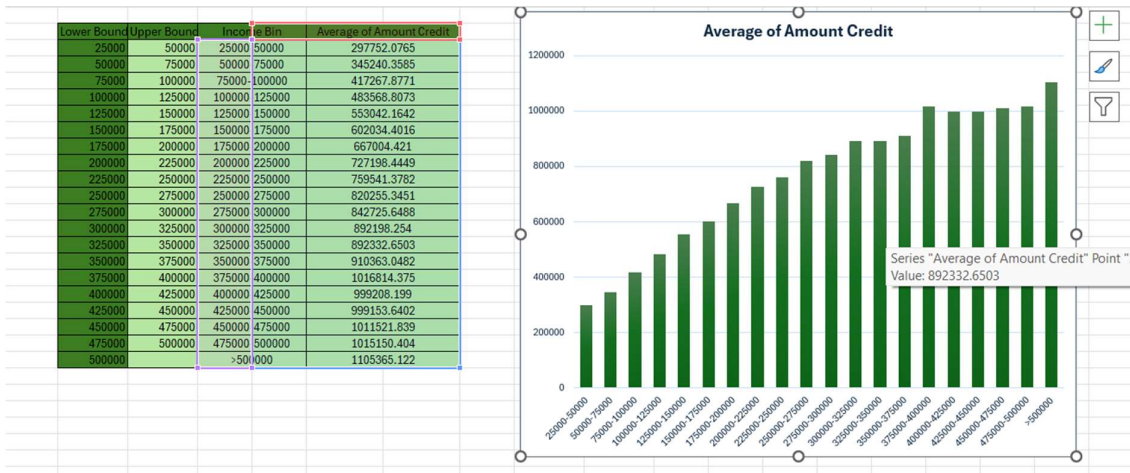
C. Analyze Data Imbalance:

- Assessed class distribution of target variables using the COUNTIF () function.
- Identified significant gaps in male-to-female loan outcomes.
- **Visualization:** Pie charts displaying class distribution.



Bivariate Analysis:

- Analyzed relationships between variables using scatter plots and pivot tables.



E. Identify Top Correlations for Different Scenarios:

Calculated correlation coefficients using the CORREL function.

TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	REGION_RATING_CLIENT
0	0	270000	1293502.5	35698.5	0.003541	-16765	-1188	-1186	1
0	0	67500	135000	6750	0.010032	-19046	-225	-4260	2
0	0	135000	312682.5	29686.5	0.008019	-19005	-3039	-9833	2
0	0	121500	513000	21865.5	0.028663	-19932	-3038	-4311	2
0	0	99000	490495.5	27517.5	0.035792	-16941	-1588	-4970	2
0	1	171000	1560726	41301	0.035792	-13778	-3130	-1213	2
0	0	360000	1530000	42075	0.003122	-18850	-449	-4597	3
0	0	112500	1019610	33826.5	0.018634	-20099	365243	-7427	2
0	0	135000	405000	20250	0.019689	-14469	-2019	-14437	2
0	1	112500	652500	21177	0.0228	-10197	-679	-4427	2
0	0	38419.155	148365	10678.5	0.015221	-20417	365243	-5246	2
0	0	67500	80865	5881.5	Plot Area 0.1329	-13439	-2717	-311	2
0	1	225000	918468	28966.5	0.016612	-14086	-3028	-643	2
0	0	189000	773680.5	32778	0.010006	-14583	-203	-615	2
0	0	157500	299772	20160	0.020713	-8728	-1157	-3494	3
0	0	108000	509602.5	26149.5	0.018634	-12931	-1317	-6392	2
0	1	81000	270000	13500	0.010966	-9776	-191	-4143	2
0	0	112500	157500	7875	0.04622	-17718	-7804	-8751	1
0	1	90000	544491	17563.5	0.015221	-11348	-2038	-1021	2
0	0	135000	427500	21375	0.015221	-18252	-4286	-298	2
0	1	202500	1132573.5	37561.5	0.025164	-14815	-1652	-2299	2
0	1	450000	497520	32521.5	0.020713	-11146	-4306	-114	3
0	0	83250	239850	23850	0.006296	-24827	365243	-9012	3
0	2	135000	247500	12703.5	0.026392	-11286	-746	-108	2
0	0	90000	235000	11073.5	0.026663	-10224	2404	-2410	2

TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	REGION_RATING_CLIENT
1	0	202500	406597.5	24700.5	0.018801	-9461	-637	-3648	2
1	0	112500	979992	27076.5	0.018029	-18724	-2628	-6573	3
1	0	202500	1193580	35028	0.025164	-17482	-1262	-1182	2
1	0	135000	288873	16258.5	0.007305	-13384	-3597	-45	3
1	0	81000	252000	14593.5	0.028663	-24794	365243	-5391	2
1	0	315000	953460	64107	0.030755	-10199	-2015	-4802	2
1	1	157500	723996	Plot Area 0.25	0.007274	-10526	-267	-387	2
1	0	292500	675000	36747	0.035792	-16667	-200	-5239	2
1	0	157500	245619	12667.5	0.022625	-17538	-7676	-774	2
1	0	111915	225000	21037.5	0.018801	-7989	-150	-2570	2
1	3	180000	540000	27000	0.02461	-15326	-1038	-782	2
1	1	202500	436032	28516.5	0.032561	-10127	-111	-1724	1
1	0	135000	495216	26995.5	0.008019	-12824	-154	-6764	2
1	0	157500	1710000	66262.5	0.004849	-23127	-9274	-817	2
1	0	73341	135000	6750	0.018801	-9157	-146	-2926	2
1	1	121500	263686.5	17298	0.014464	-17847	-448	-3595	2
1	1	225000	1019205	31032	0.072508	-11356	-602	-335	1
1	0	63000	426645	22468.5	0.018634	-12209	-1572	-6348	2
1	1	112500	571486.5	31131	0.022625	-8292	-166	-3062	2
1	0	36000	284400	10849.5	0.018209	-22078	365243	-6748	3
1	0	157500	497520	28692	0.02461	-13989	-1350	-8112	2
1	0	112500	417024	25330.5	0.025164	-13707	365243	-7809	2
1	0	202500	343683	16663.5	0.010966	-20200	365243	-973	2
1	0	112500	117162	12433.5	0.01885	-18302	-136	-4005	2
1	0	81000	213840	23801.5	0.01885	-8028	600	-8028	2

- Identified top predictors influencing loan defaults.
- Visualization:** Heatmaps showcasing correlations.

Correlation									
Correlation for applicants with no payment difficulty									
CNT_CHILDREN	1	0.036319722	0.005705458	0.02638212	-0.024912809	0.3358763	-0.243591518	0.183072478	0.021288992
AMT_INCOME_TOTAL	0.036319722	1	0.377965752	0.451135629	0.181941261	0.0737694	-0.162702675	0.06893375	-0.205031899
AMT_CREDIT	0.005705458	0.377965752	1	0.770771802	0.095539444	-0.0510842	-0.077367219	0.008053758	-0.102556478
AMT_ANNUITY	0.02638212	0.451135629	0.770771802	1	0.117280527	0.0099154	-0.113006832	0.034609087	-0.129921191
REGION_POPULATION_RELATIVE	-0.024912809	0.181941261	0.095539444	0.117280527	1	-0.0304354	-0.006610653	-0.058501361	-0.539333113
DAYS_BIRTH	0.335876269	0.073769425	-0.05108418	0.009915418	-0.030435419	1	-0.615289978	0.335028046	0.00902485
DAYS_EMPLOYED	-0.243591518	-0.162702675	-0.07736722	-0.11300683	-0.006610653	-0.61529	1	-0.204370881	0.040505636
DAYS_REGISTRATION	0.183072478	0.068933758	0.008053758	0.034609087	-0.058501361	0.335028	-0.204370881	1	0.082562812
REGION_RATING_CLIENT	0.021288992	-0.205031899	-0.10255648	-0.12992119	-0.539333113	0.0090249	0.040505636	0.082562812	1
Correlation for applicants with payment difficulty									
CNT_CHILDREN	1	0.010111286	0.003288677	0.0265292	-0.018457758	0.2573025	-0.192393824	0.155016732	0.057898672
AMT_INCOME_TOTAL	0.010111286	1	0.015252479	0.017990289	-0.006168919	0.0090437	-0.011550035	-0.009563924	-0.012841299
AMT_CREDIT	0.003288677	0.015252479	1	0.749595552	0.068025736	-0.1423525	0.016163686	-0.042907681	-0.044925355
AMT_ANNUITY	0.0265292	0.017990289	0.749595552	1	0.073307627	-0.0086125	-0.079478063	0.021544664	-0.061505596
REGION_POPULATION_RELATIVE	-0.018457758	-0.006168919	0.068025736	0.073307627	1	-0.0165709	0.007680095	-0.046104928	-0.430122181
DAYS_BIRTH	0.057898672	-0.012841299	-0.04492535	-0.06150556	-0.430122181	0.0449819	-0.009176249	0.115640782	1
DAYS_EMPLOYED	-0.192393824	-0.011550035	0.016163686	-0.07947806	0.007680095	-0.5815651	1	-0.188707428	-0.009176249
DAYS_REGISTRATION	0.155016732	-0.009563924	-0.04290768	0.021544664	-0.046104928	0.2884741	-0.188707428	1	0.115640782
REGION_RATING_CLIENT	0.057898672	-0.012841299	-0.04492535	-0.06150556	-0.430122181	0.0449819	-0.009176249	0.115640782	1

Insights

Key Insights:

1. High-income customers are less likely to default.
2. Loan amounts exceeding a certain threshold increase default risk.
3. Data imbalance in loan approvals impacts analysis.

Recommendations:

- Implement stricter checks for high-risk applicants.
 - Offer loans with adjusted terms for risky applicants (e.g., higher interest rates).
 - Improve data collection processes to minimize missing values.
-

Result

The analysis revealed actionable insights for loan approval strategies. By addressing high-risk attributes, the company can:

- Deny loans to high-risk applicants.
 - Adjust loan terms based on risk levels.
 - Develop policies for shorter loan tenures for specific customer segments.
-

Drive Link

https://drive.google.com/drive/folders/1tYHOD8kxxyXugeXr8f6fHzRcNaNcXnm_?usp=sharing