



BANK LOAN CASE STUDY



PROJECT DESCRIPTION

Objective:

The project focuses on performing EDA(Exploratory Data Analysis) to ensure that capable applicants are not rejected.

Dataset provided:

- a) **previous_application.csv**: Contains information about previous loan applications.
- b) **application_data.csv**: Provides details about the current loan applications.
- c) **columns_description.csv**: Describes the columns present in the other datasets, explaining what each column represents.



APPROACH



TECH STACK USED



MS Excel (Microsoft 365 version)

Used to analyze the dataset provided to me.

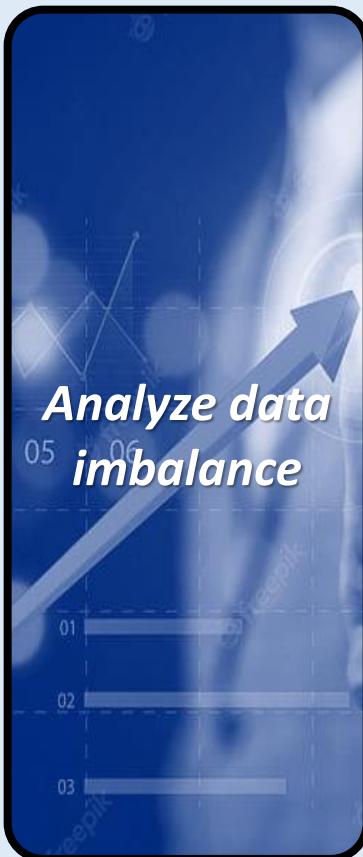
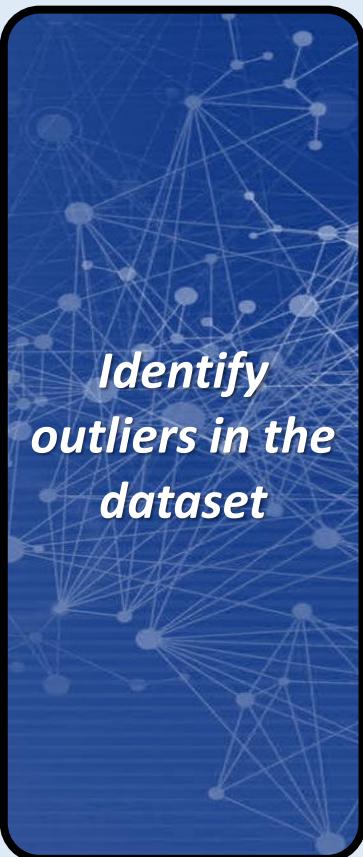


MS PowerPoint (Microsoft 365 version)

Used to create and present the presentation.

INSIGHTS

Tasks to perform:-





IDENTIFY MISSING DATA AND DEAL WITH IT APPROPRIATELY

INSIGHTS

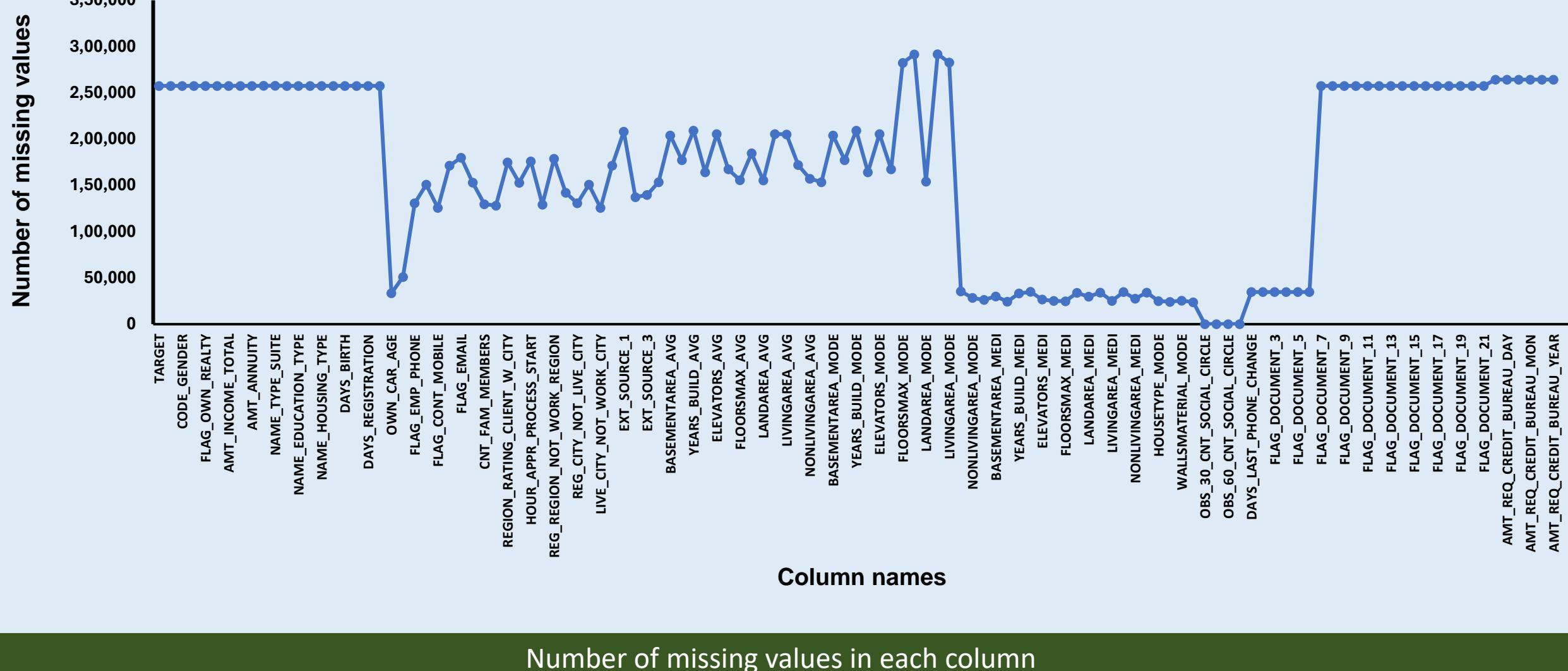
IDENTIFY MISSING DATA AND DEAL WITH IT APPROPRIATELY

- Using the **COUNTBLANK()** function first I have calculated the number of blank cells in each column.
- Removed all the rows which don't have **SK_ID_CURR**.
- Adding a row that tells if there are blank rows or not(TRUE OR FALSE).
- Removed all the columns whose output is FALSE.
- Changing the number format of each column that contains the number.



INSIGHTS

IDENTIFY MISSING DATA AND DEAL WITH IT APPROPRIATELY

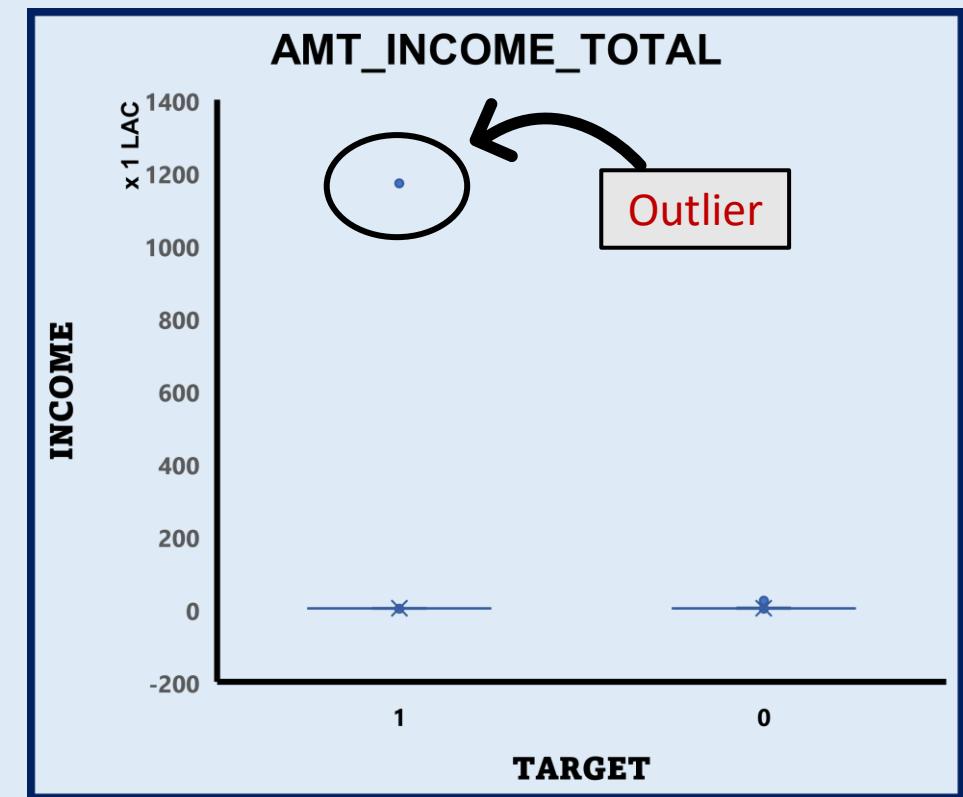


IDENTIFY OUTLIERS IN THE DATASET

INSIGHTS

IDENTIFY OUTLIERS IN THE DATASET

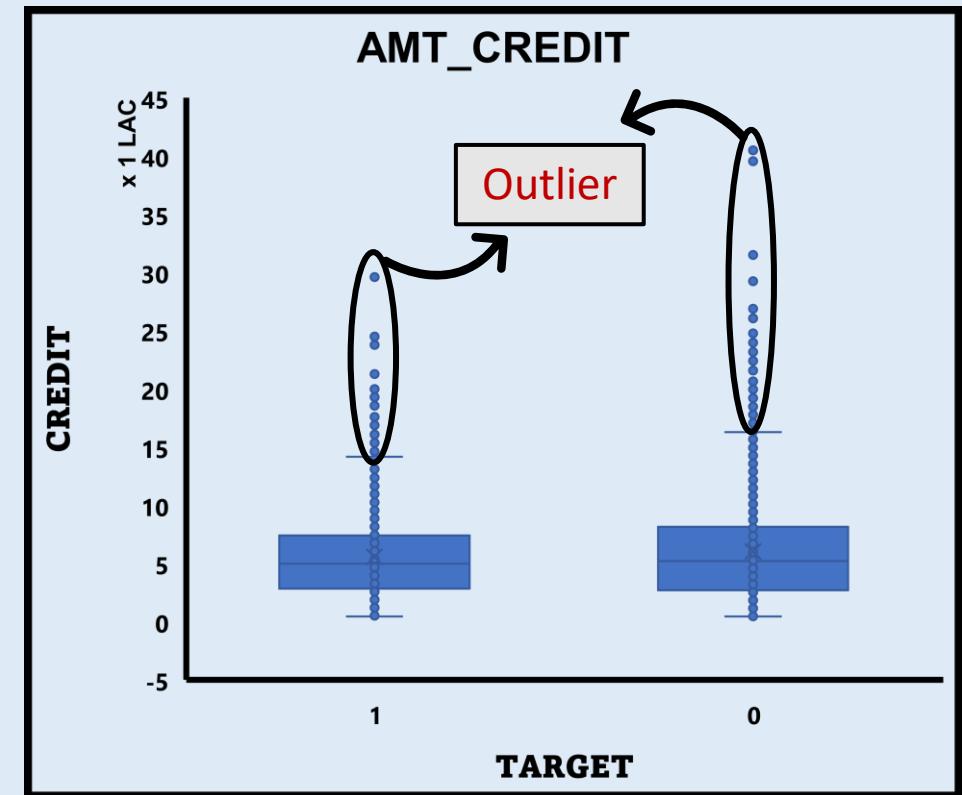
In the AMT_INCOME_TOTAL column, there is a person who has an income greater than 10cr and has payment difficulties.



INSIGHTS

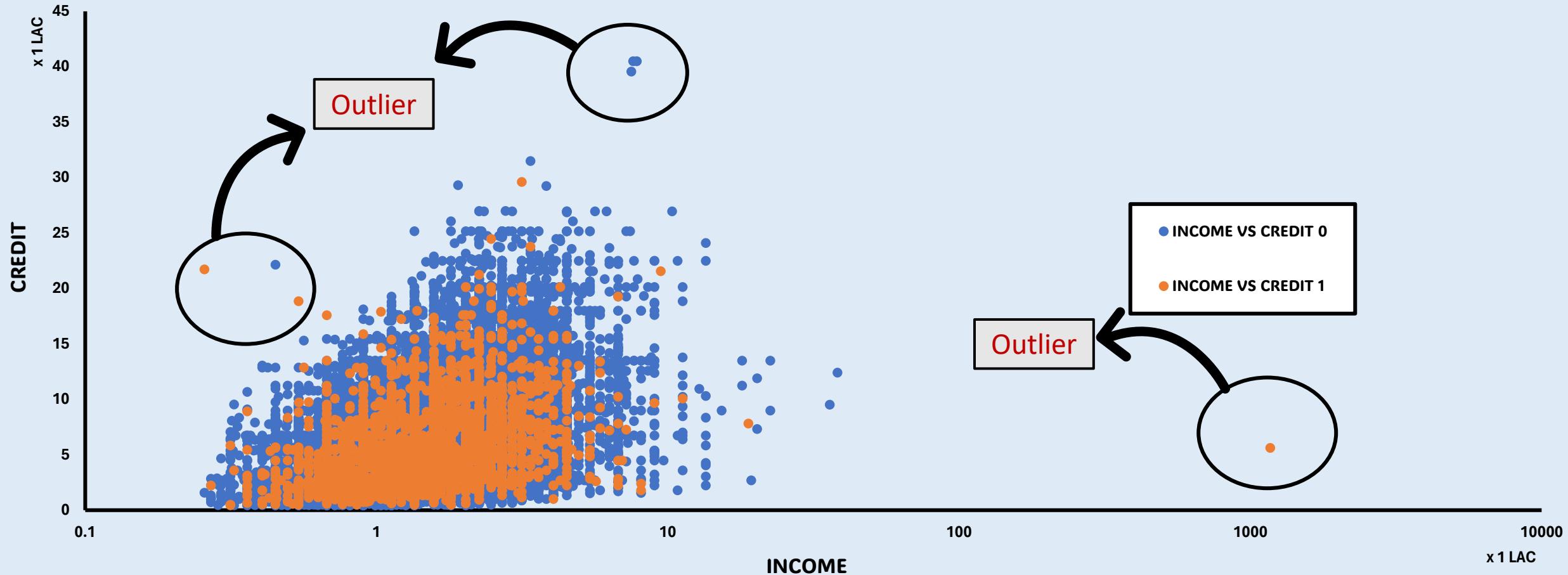
IDENTIFY OUTLIERS IN THE DATASET

It has been observed that a considerable number of individuals have obtained loans exceeding 15 lakhs in the AMT_CREDIT column, which can be considered outliers.



INSIGHTS

IDENTIFY OUTLIERS IN THE DATASET

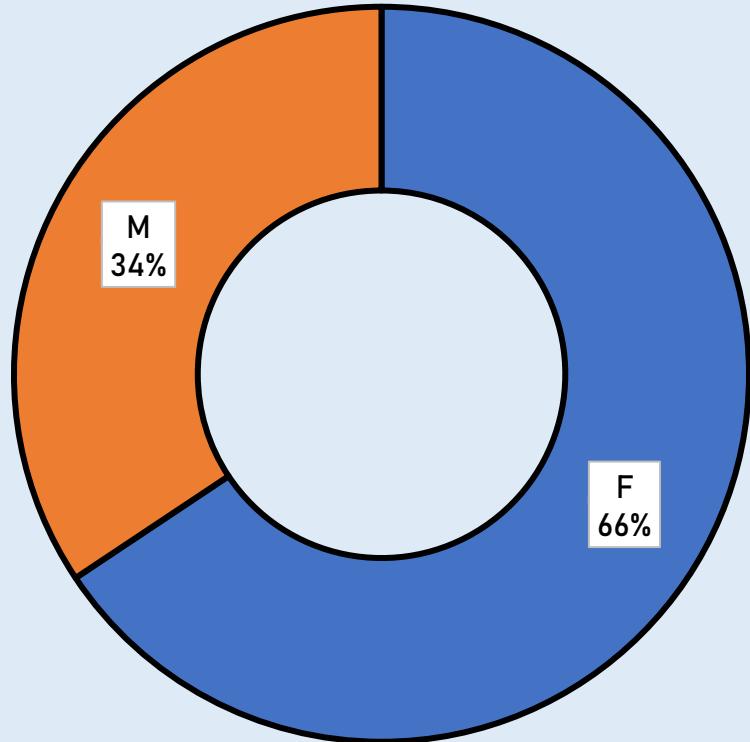


Certainly! In certain situations, individuals with exceptionally low incomes take out high amounts of credit, while in other cases, individuals with considerably high incomes have very low credit utilization. Both scenarios can be classified as outliers.

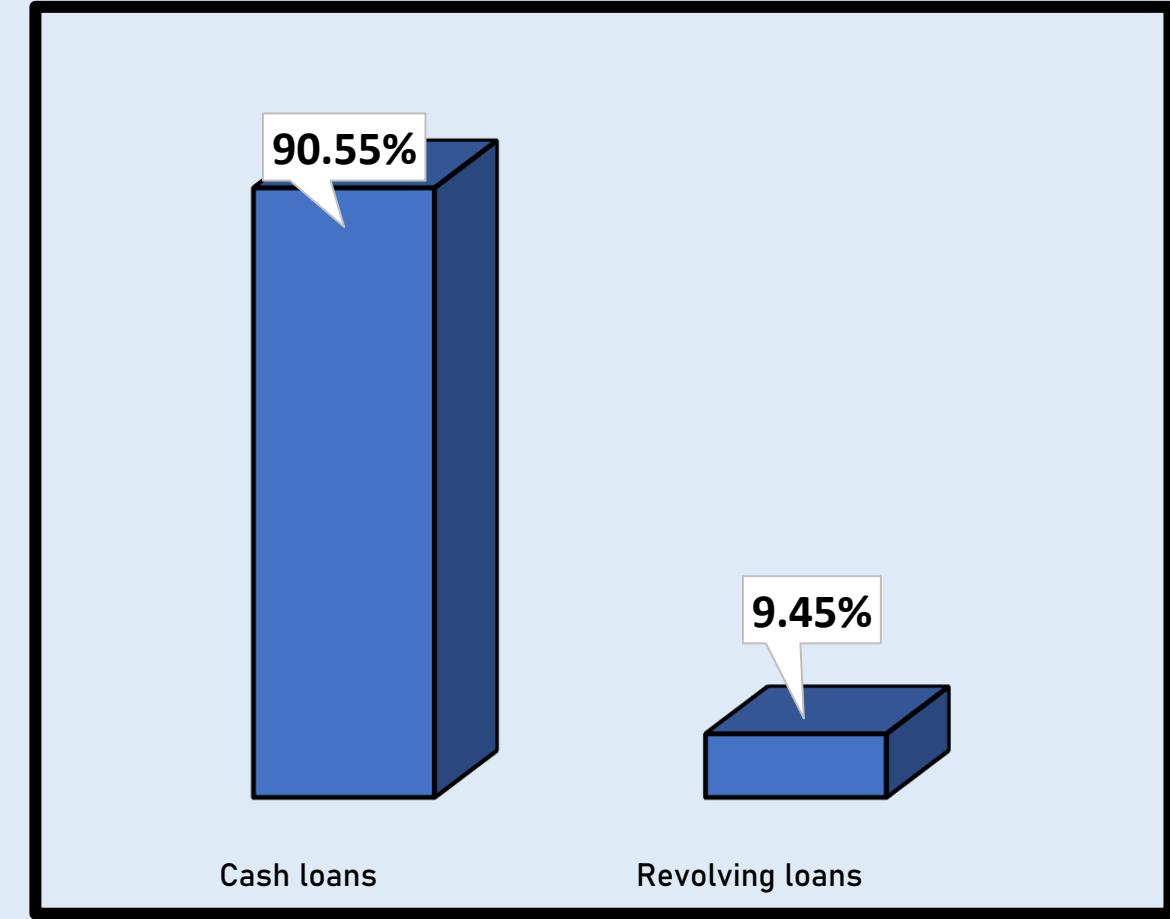
ANALYZE DATA IMBALANCE

INSIGHTS

ANALYZE DATA IMBALANCE



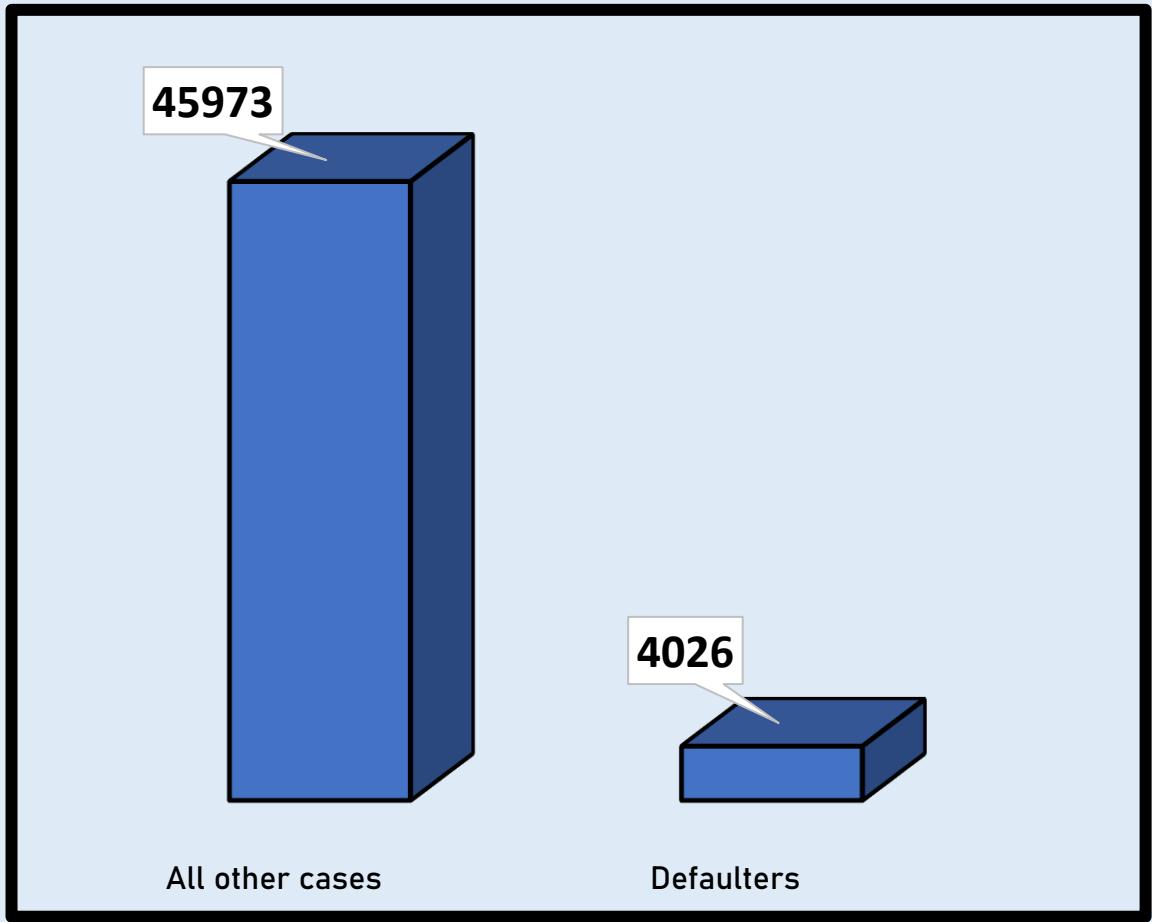
Female applicants are greater than male applicants.



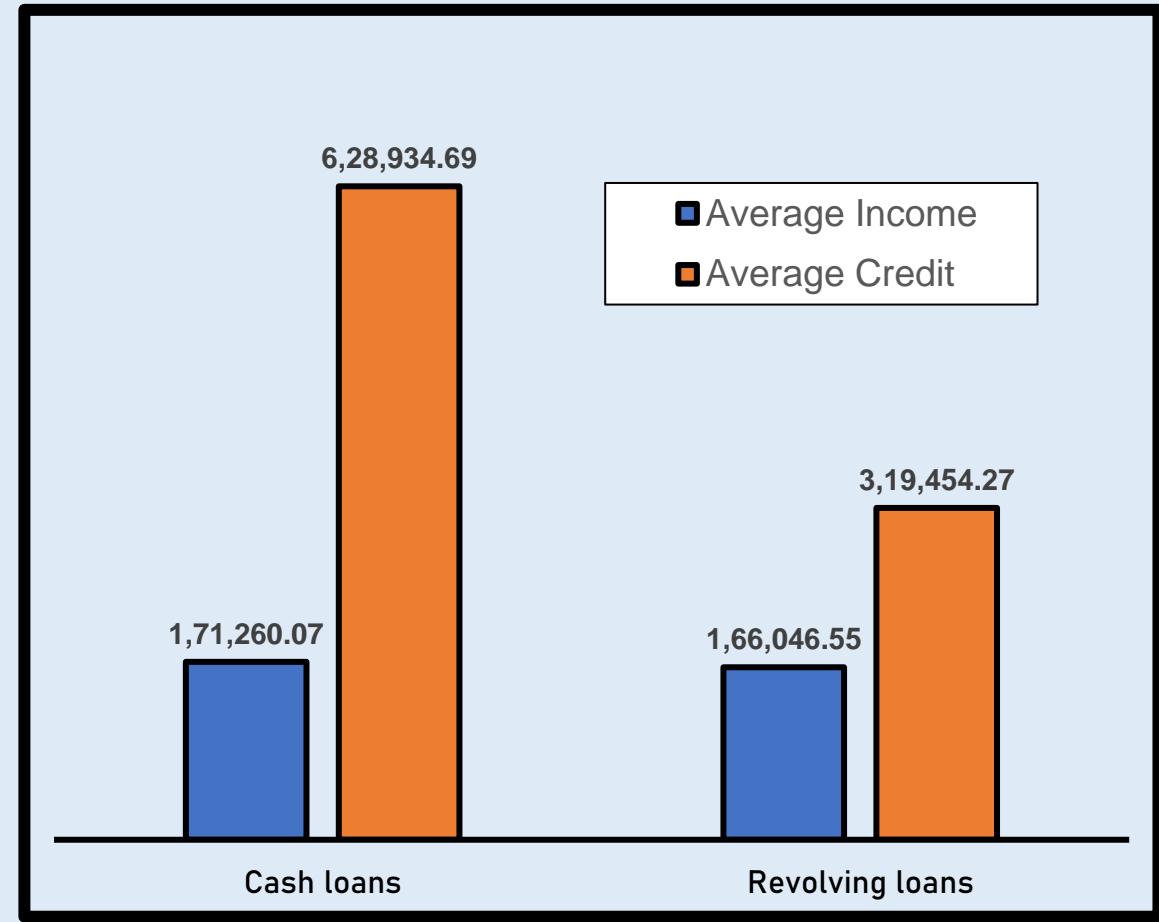
The loan taken are mostly cash loans.

INSIGHTS

ANALYZE DATA IMBALANCE



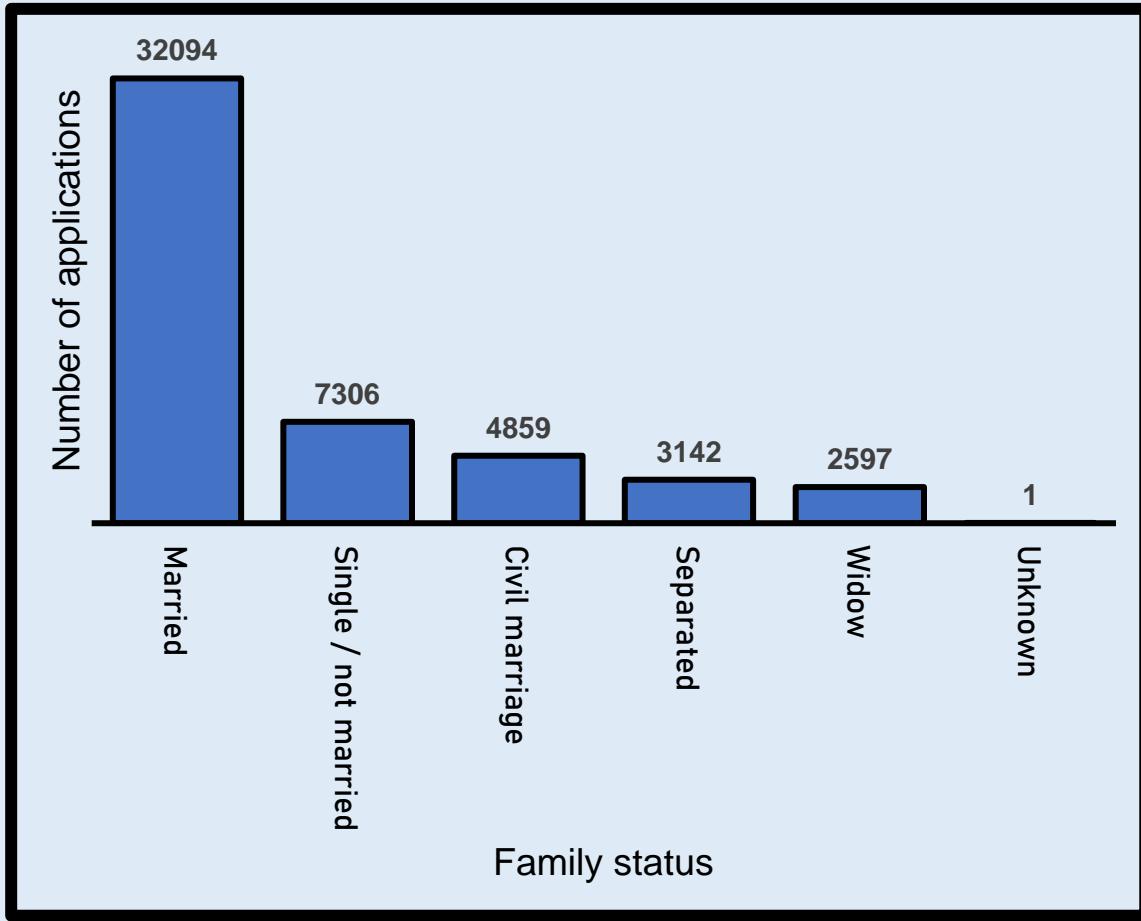
There are 4026 applicants who are having payment difficulties.



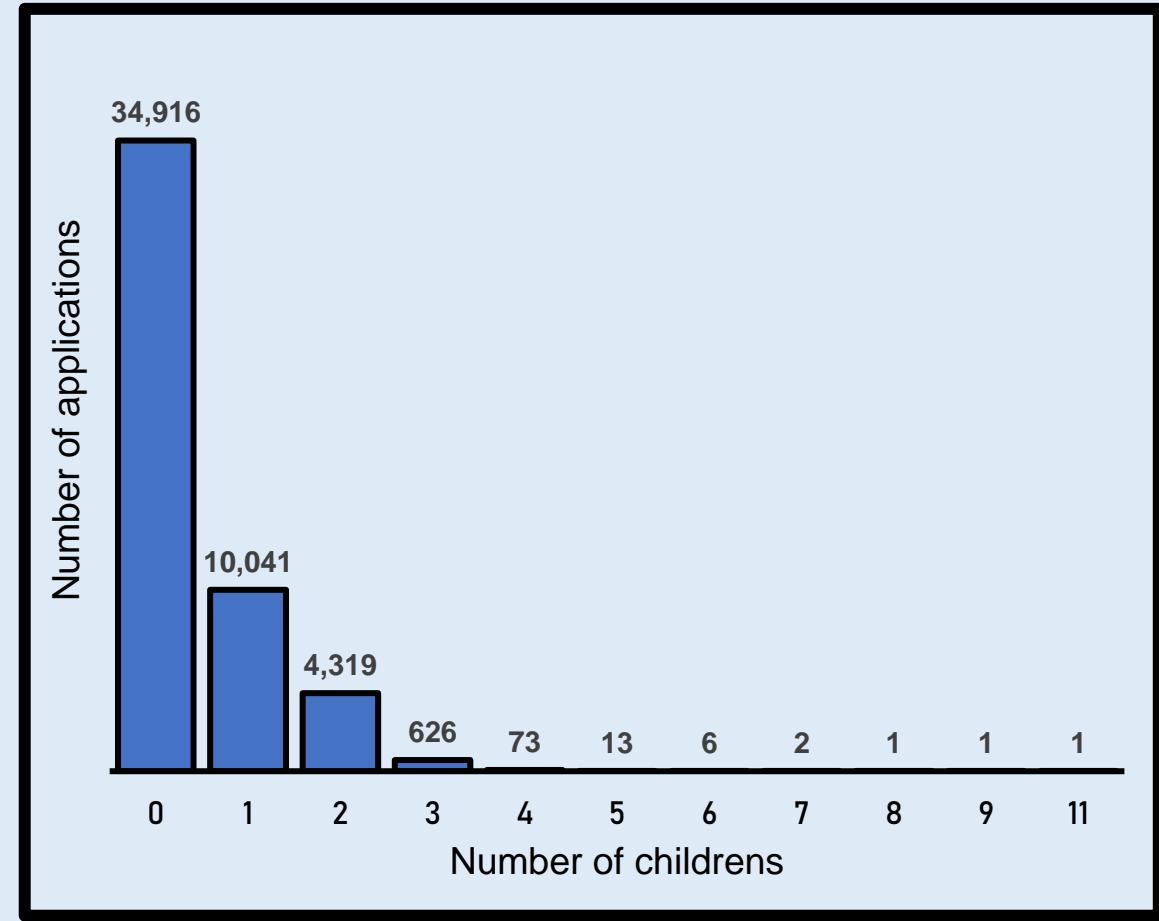
The average credit for cash loans is 2 times greater than revolving loans.

INSIGHTS

ANALYZE DATA IMBALANCE



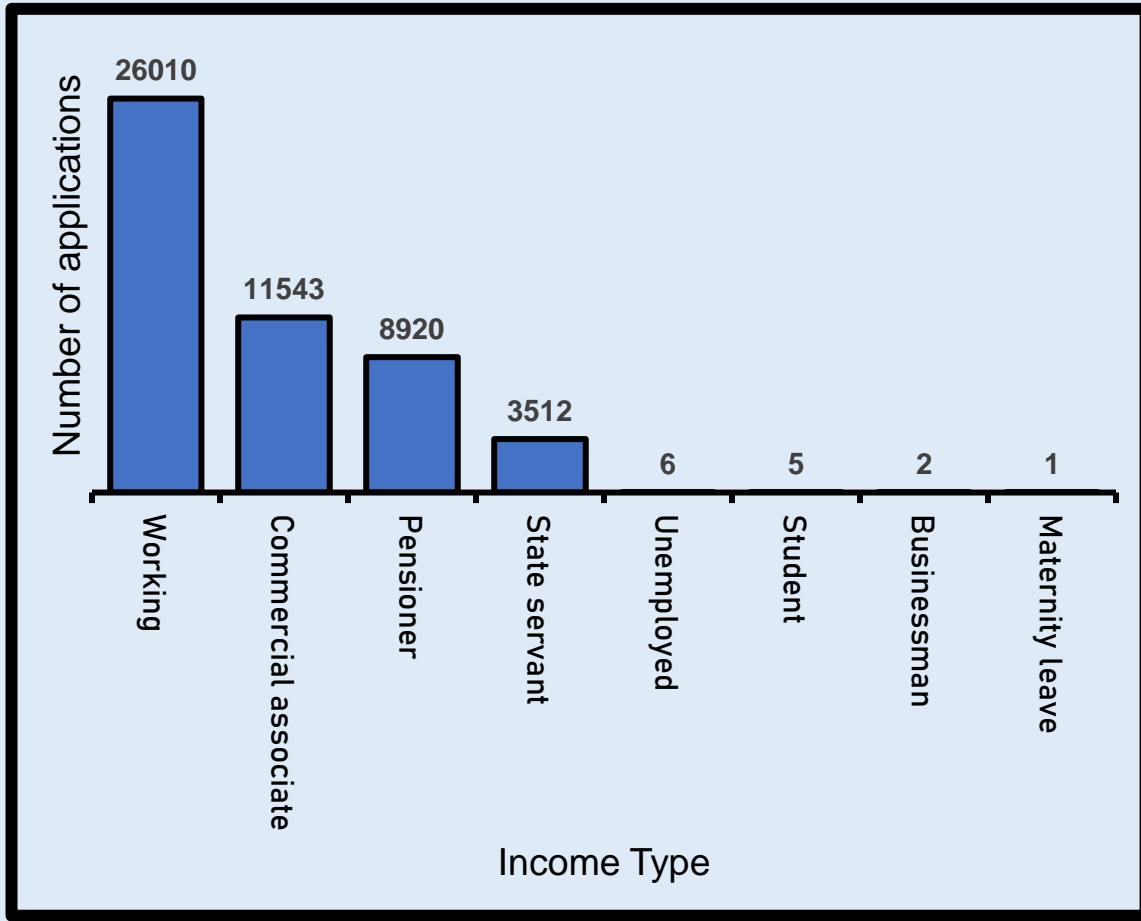
Most applications are from people who are married.



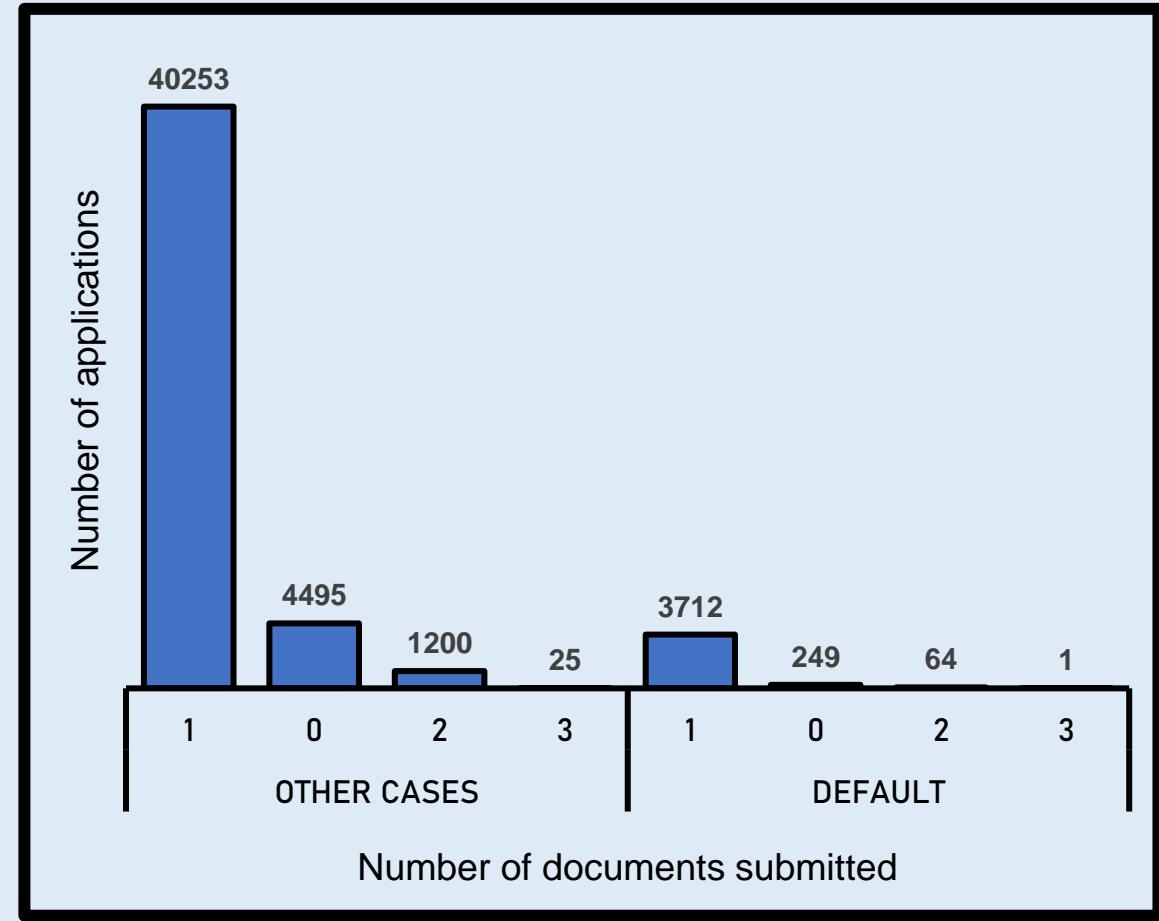
Individuals without children tend to apply for loans more frequently.

INSIGHTS

ANALYZE DATA IMBALANCE



Most of the applications are from working professionals.



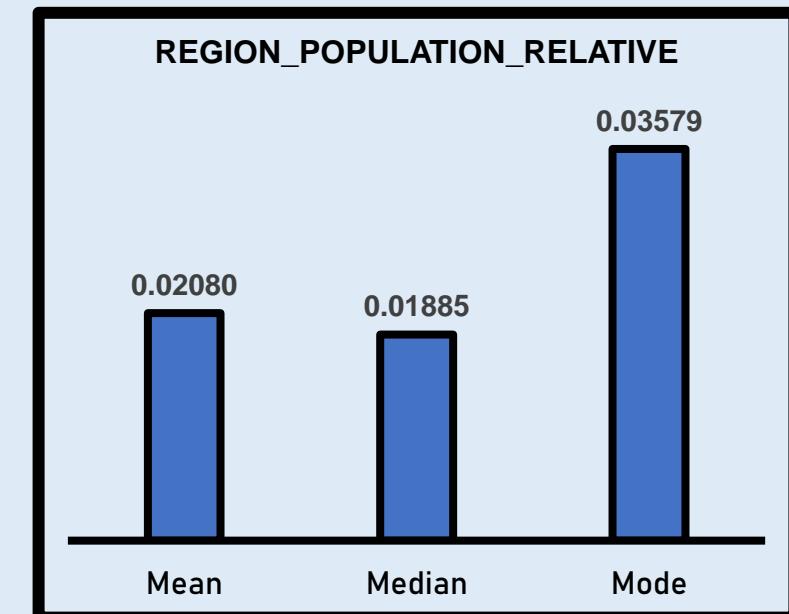
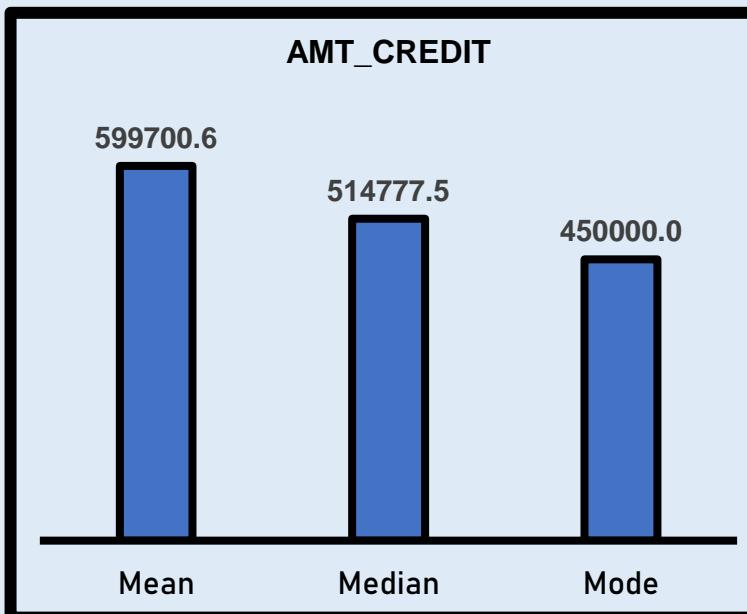
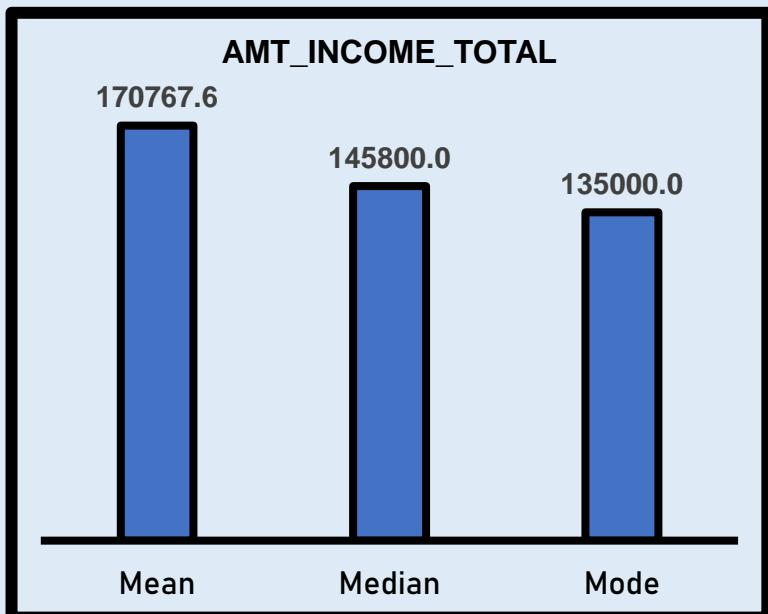
Over 85% of applications have been submitted with only one document.



PERFORM UNIVARIATE, SEGMENTED UNIVARIATE AND BIVARIATE ANALYSIS

INSIGHTS

PERFORM UNIVARIATE, SEGMENTED UNIVARIATE, AND BIVARIATE ANALYSIS



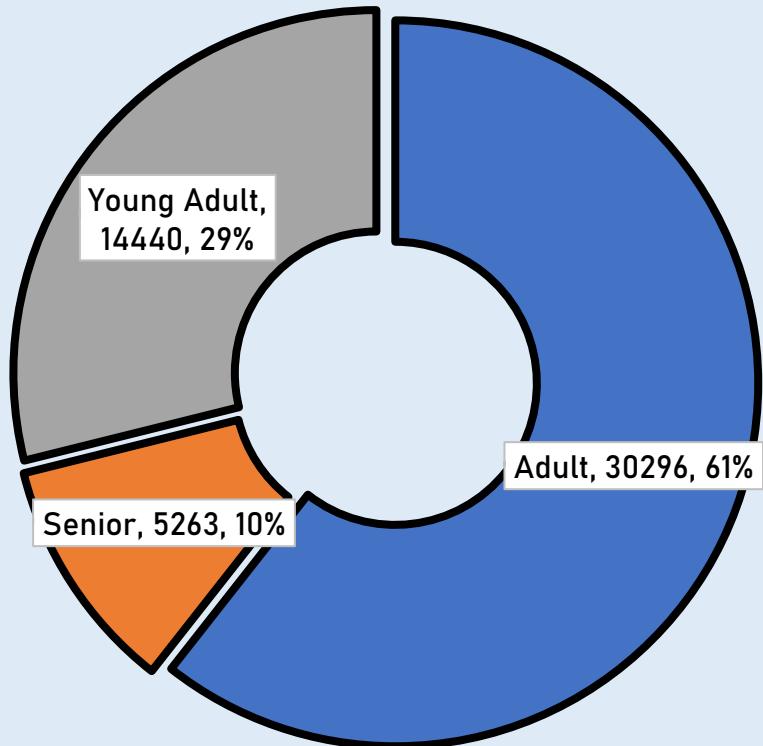
Mean	170767.6
Median	145800.0
Mode	135000.0

Mean	599700.6
Median	514777.5
Mode	450000.0

Mean	0.02080
Median	0.01885
Mode	0.03579

INSIGHTS

PERFORM UNIVARIATE, SEGMENTED UNIVARIATE, AND BIVARIATE ANALYSIS



AGE	CATEGORY	APPLICANTS
21-35	Young Adult	14440
36-60	Adult	30296
> 60	Senior	5263

SEGMENTED UNIVARIATE ANALYSIS

INSIGHTS

PERFORM UNIVARIATE, SEGMENTED UNIVARIATE, AND BIVARIATE ANALYSIS

Average of AMT_INCOME_TOTAL	Column Labels	Co-op apartment	House / apartment	Municipal apartment	Office apartment	Rented apartment	With parents
Row Labels							
Academic degree		1,62,000.00	1,82,750.00		3,51,000.00		
Higher education		1,68,959.30	2,09,864.44	1,97,414.50	2,17,770.33	2,01,170.35	1,85,443.01
Incomplete higher		1,65,214.29	1,89,166.60	1,60,714.29	1,97,156.25	1,55,283.04	1,53,675.56
Lower secondary			1,27,323.84	1,44,895.98	1,26,000.00	2,62,350.00	1,34,678.57
Secondary / secondary special		1,66,003.61	1,58,261.01	1,62,899.28	1,61,800.00	1,61,519.38	1,48,212.92

Individuals residing in office apartments and possessing an academic degree enjoy the highest income levels.

Average of AMT_CREDIT	Column Labels	Co-op apartment	House / apartment	Municipal apartment	Office apartment	Rented apartment	With parents
Row Labels							
Academic degree		2,70,000.00	8,28,062.50		5,18,562.00		
Higher education		5,31,756.58	6,96,320.83	6,92,509.04	6,70,908.79	5,57,134.51	5,79,626.82
Incomplete higher		3,26,416.50	5,95,853.32	4,45,131.00	4,17,234.94	3,90,181.26	4,73,239.64
Lower secondary			5,01,979.22	6,03,915.83	7,21,231.88	7,82,102.70	4,60,093.82
Secondary / secondary special		5,09,972.52	5,77,206.84	5,89,719.35	5,48,371.81	5,26,181.23	4,87,441.74

Individuals residing in house or apartment and possessing an academic degree takes high credit.



IDENTIFY TOP CORRELATION FOR DIFFERENT SCENARIOS

INSIGHTS

IDENTIFY TOP CORRELATIONS FOR DIFFERENT SCENARIOS

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	FLAG_MOBIL	FLAG_EMP_PHONE	FLAG_WORK_PHONE	FLAG_CONT_MOBILE	FLAG_PHONE	FLAG_EMAIL	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY
CNT_CHILDREN	1															
AMT_INCOME_TOTAL	0.010110177	1														
AMT_CREDIT	0.007601905	0.015271444	1													
REGION_POPULATION_RELATIVE	-0.020359154	-0.006180303	0.067775624	1												
DAYS_BIRTH	0.2496732	0.009033662	-0.142506035	-0.016468731	1											
DAYS_EMPLOYED	-0.189324184	-0.011555963	0.016039571	0.007742909	-0.581479041	1										
DAYS_REGISTRATION	0.152113117	-0.009561152	-0.042844404	-0.046130288	0.288437837	-0.188718437	1									
DAYS_ID_PUBLISH	-0.042360717	-0.009122006	-0.043771901	-0.005118563	0.247896571	-0.230063668	0.09029149	1								
FLAG_MOBIL	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	1							
FLAG_EMP_PHONE	0.189566508	0.011657983	-0.017405837	-0.007727337	0.584902532	-0.999901205	0.190586868	0.231379438	#DIV/0!	1						
FLAG_WORK_PHONE	0.06463311	-0.011875045	-0.052629114	-0.026921987	0.166203043	-0.205486118	0.058056566	0.036518826	#DIV/0!	0.205356525	1					
FLAG_CONT_MOBILE	0.009342338	-0.001651428	0.030642933	-0.002509474	-0.007774873	-0.012034117	-0.012464747	0.005769675	#DIV/0!	0.011928014	0.01713917	1				
FLAG_PHONE	-0.014302778	-0.008271927	0.036387061	0.078115256	-0.049104754	0.031612536	-0.065075174	-0.039134599	#DIV/0!	-0.032542305	0.29825961	0.017151262	1			
FLAG_EMAIL	0.011662997	0.000379597	-0.004036924	0.050737138	0.078036349	-0.050739487	-0.009241401	0.033361117	#DIV/0!	0.051184266	0.015201497	0.007582037	0.025468396	1		
REGION_RATING_CLIENT	0.055515557	-0.012846697	-0.045024534	-0.430032303	0.045027112	-0.009145883	0.115625377	0.025335227	#DIV/0!	0.009192186	0.024253798	0.02633495	-0.062845235	-0.022725594	1	
REGION_RATING_CLIENT_W_CITY	0.054802235	-0.01266585	-0.052954314	-0.431675881	0.038087333	-0.004137686	0.108123203	0.014431344	#DIV/0!	0.004132384	0.029153013	0.025221794	-0.055001395	-0.019500794	0.950768899	1

1	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY
2	DAY_BIRTH	FLAG_EMP_PHONE
3	FLAG_WORK_PHONE	FLAG_PHONE

TOP 3
CORRELATIONS

CORRELATIONS FOR TARGET 1(DEFAULT)

INSIGHTS

IDENTIFY TOP CORRELATIONS FOR DIFFERENT SCENARIOS

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	FLAG_MOBIL	FLAG_EMP_PHONE	FLAG_WORK_PHONE	FLAG_CONT_MOBILE	FLAG_PHONE	FLAG_EMAIL	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY
CNT_CHILDREN	1															
AMT_INCOME_TOTAL	0.036319722	1														
AMT_CREDIT	0.005705458	0.377965752	1													
REGION_POPULATION_RELATIVE	-0.024912809	0.181941261	0.095539444	1												
DAYS_BIRTH	0.335876269	0.073769425	-0.051084182	-0.030435419	1											
DAYS_EMPLOYED	-0.243591518	-0.162702675	-0.077367219	-0.006610653	-0.615289978	1										
DAYS_REGISTRATION	0.183072478	0.06893375	0.008053758	-0.058501361	0.335028046	-0.204370881	1									
DAYS_ID_PUBLISH	-0.032537221	0.032286356	-0.008290189	-0.002236288	0.270073313	-0.27222439	0.103548902	1								
FLAG_MOBIL	0.00269209	0.002009697	0.00372218	0.003461456	-0.008068695	0.002137507	-0.000304657	-0.006016013	1							
FLAG_EMP_PHONE	0.244620986	0.162219844	0.076005444	0.00674935	0.619443937	-0.999739547	0.206572521	0.273500944	-0.002208929	1						
FLAG_WORK_PHONE	0.054409837	-0.034502225	-0.011512143	-0.015101101	0.175419095	-0.234535874	0.059505721	0.050945857	0.002308233	0.234398497	1					
FLAG_CONT_MOBILE	-0.003846515	-0.016970699	0.024439133	-0.004898838	-0.012765423	0.016852654	8.39513E-05	-0.004976838	-0.00021446	-0.016876487	0.022757235	1				
FLAG_PHONE	-0.031250696	0.00273884	0.017200016	0.093910712	-0.041288098	0.020690643	-0.071428092	-0.031387579	0.002923322	-0.021521368	0.299091108	0.004592469	1			
FLAG_EMAIL	0.02832663	0.087488653	0.011859749	0.038691982	0.093733189	-0.067793808	0.033249993	0.033383068	0.001133268	0.068273194	-0.010316581	-0.011554223	0.015187351	1		
REGION_RATING_CLIENT	0.021288992	-0.205031899	-0.102556478	-0.539333113	0.00902485	0.040505636	0.082562812	-0.008097427	0.000383258	-0.040711716	0.0025479	0.013115747	-0.088733037	-0.058994634	1	
REGION_RATING_CLIENT_W_CITY	0.017873365	-0.220044862	-0.111639948	-0.536859601	0.00708431	0.042898876	0.074745932	-0.012667326	0.000192888	-0.043056438	0.007884758	0.013238981	-0.083627775	-0.05652979	0.950468157	1

1	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY
2	DAY_BIRTH	FLAG_EMP_PHONE
3	AMT_INCOME_TOTAL	AMT_CREDIT

TOP 3
CORRELATIONS



CORRELATIONS FOR TARGET 0(ALL OTHER CASES)

Thank you

EXCEL LINK: [click here](#)
VIDEO LINK: [click here](#)