

# PROJECT BANK LOAN CASE STUDY



By – Priyanshu Singh

Link to the excel file <https://docs.google.com/spreadsheets/d/1K6ok7POvOTzZVOaoxCb4xTTeRAIN2b-A/edit?usp=sharing&ouid=116598383154898349386&rtpof=true&sd=true>

# **PROJECT DESCRIPTION**

This project is based on banking business, the main objective of this project is to identify patterns that shows strong indication of payment default such indication helps banks to avoid approving loan to applicants who are likely to be a defaulter which ultimately leads to avoid making loss in their business.

## **APPROACH**

Thoroughly analyzed whole dataset and handled all missing value in each column, replaced it with statistical concept in such as mean, median and mode and further prepared dataset for analysis, testing various hypothesis using pivot table and pivot chart to derive patterns that shows strong indication of payment of default.

## **Tech-Stack Used**

Tech stack used in completing this project is Microsoft excel 365 given its versatility data cleaning, data preparation for analysis was much easier to perform and functions such as pivot table made it easier to perform univariate, segmented univariate analysis, bivariate analysis and visualization of analysis with its range of visualization options.

## **INSIGHTS**

Factors like income level and employment duration have reverse relation with payment difficulties i.e. more the income level and employment duration lesser the payment default and other factors like region rating, family size and count of children have direct relation with payment default that is more region rating, family size and count of children more the payment default and other categorial variables such as applicants with loan type cash loan, gender male, lower income, education type as lower secondary, marital status as civil marriage and single, living with parents, occupation as low skilled labor, industry type 13 and previous loan purpose and type

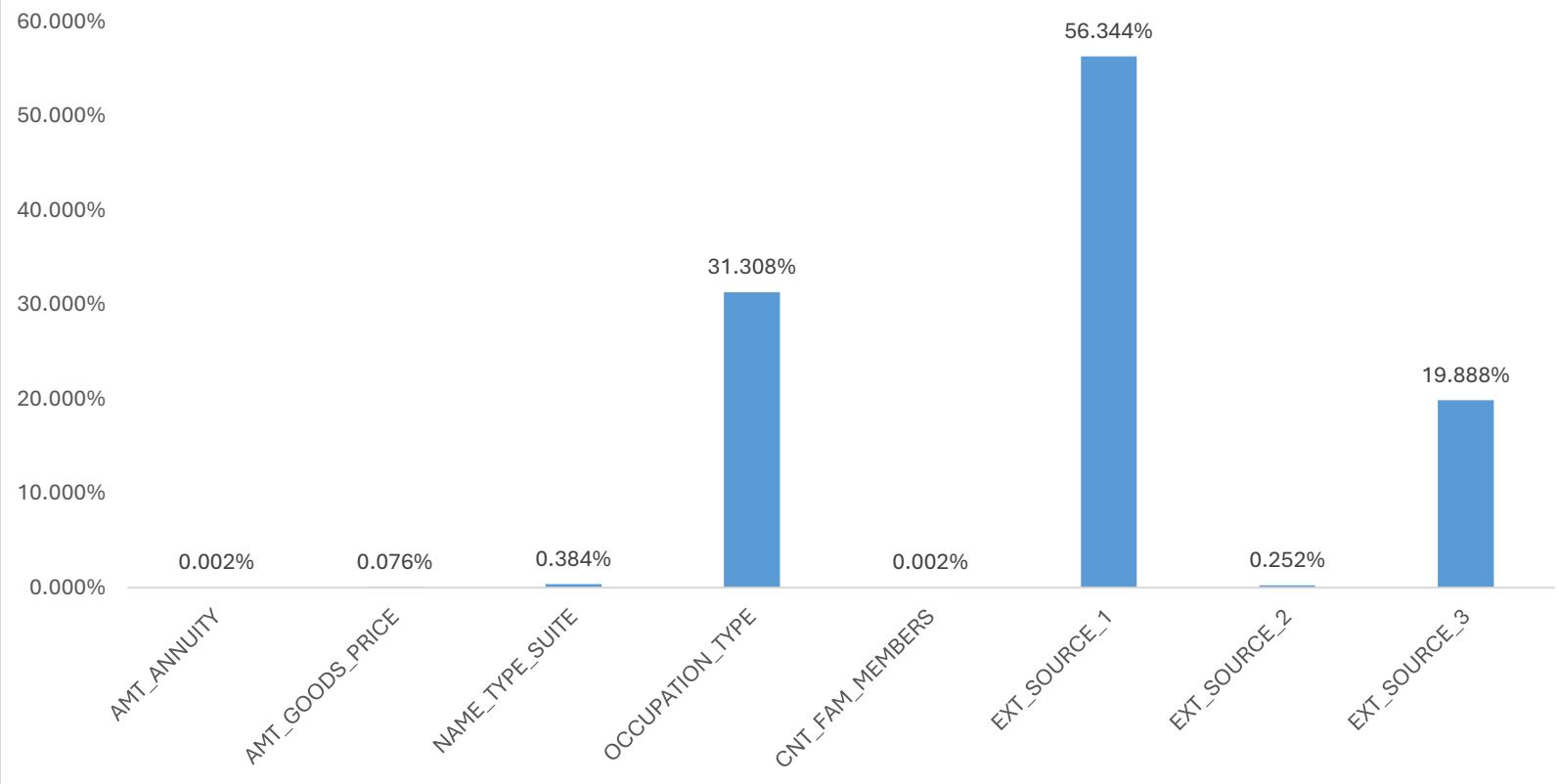
## **RESULTS**

By working on this project, I learned about how to thoroughly analyze each variable in dataset with its business context to determine which variables are important for testing various hypothesis and also learned about how exploratory data analysis (EDA) is performed on a given dataset and derive useful insights to help business.

## TASK – 1 Identify Missing Data and Deal with it Appropriately

VARIABLES	NO. OF MISSING VALUES	% OF MISSING
AMT_ANNUITY	1	0.002%
AMT_GOODS_PRICE	38	0.076%
NAME_TYPE_SUITE	192	0.384%
OWN_CAR_AGE	32950	65.900%
OCCUPATION_TYPE	15654	31.308%
CNT_FAM_MEMBERS	1	0.002%
EXT_SOURCE_1	28172	56.344%
EXT_SOURCE_2	126	0.252%
EXT_SOURCE_3	9944	19.888%

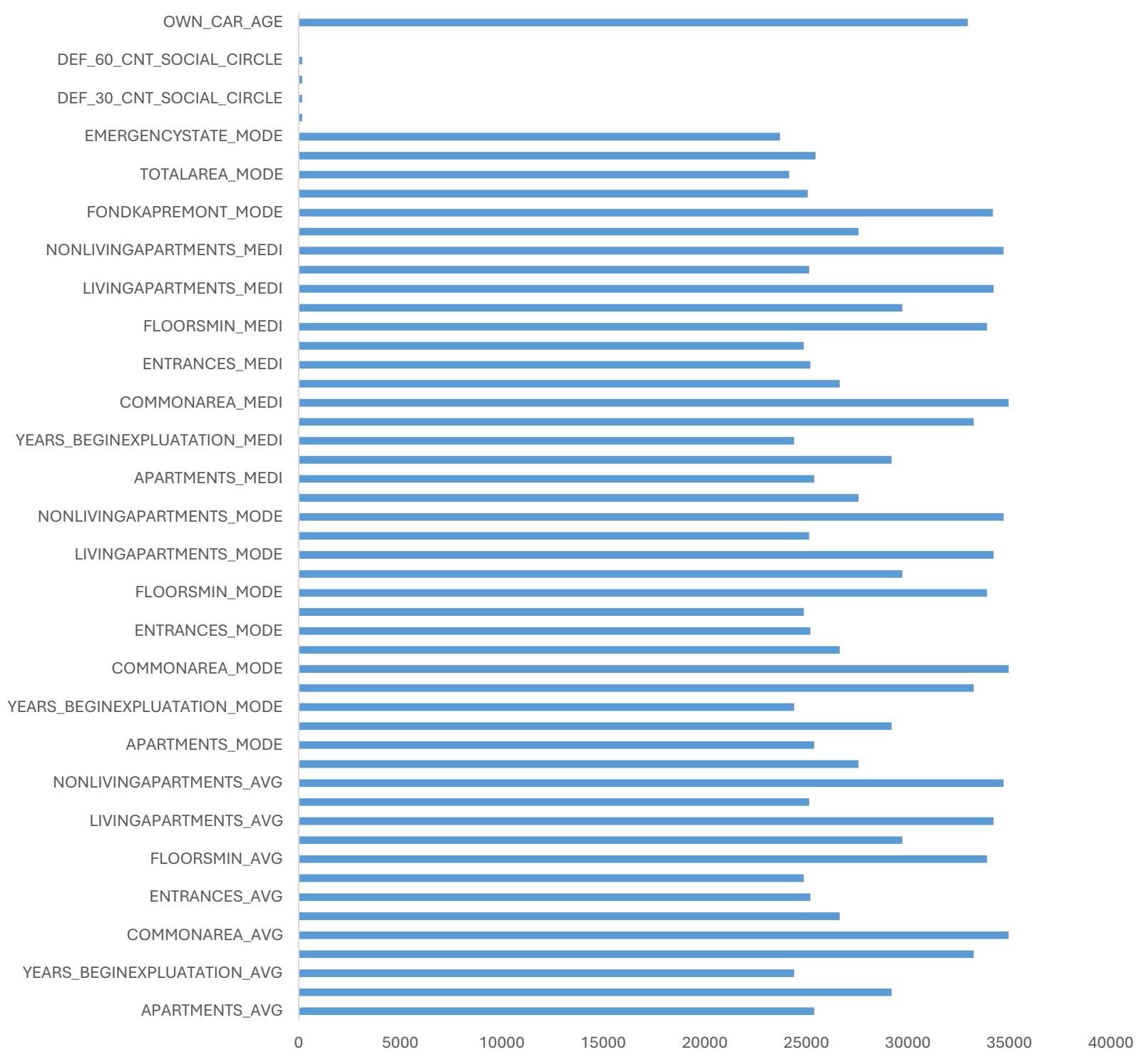
### MISSING VALUES IN SIGNIFICANT VARIABLES



- ➔ Variable “EXT\_SOURCE\_1” has the highest % of missing i.e.56.334% among all other variables.
- ➔ Variable “OCCUPATION\_TYPE” has the 2<sup>nd</sup> most % of missing value with 31.3% among all other variables.
- ➔ Variables with least % of missing values are: - AMT\_ANNUITY, AMT\_GOODS\_PRICE

ENTS_AVG	25385	50.770%	NONLIVINGAPARTMENTS_MODE	34714	69.428%
NTAREA_AVG	29199	58.398%	NONLIVINGAREA_MODE	27572	55.144%
BEGINEXPLUATATION_AVG	24394	48.788%	APARTMENTS_MEDI	25385	50.770%
BUILD_AVG	33239	66.478%	BASEMENTAREA_MEDI	29199	58.398%
ONAREA_AVG	34960	69.920%	YEARS_BEGINEXPLUATATION_MEDI	24394	48.788%
ORS_AVG	26651	53.302%	YEARS_BUILD_MEDI	33239	66.478%
CES_AVG	25195	50.390%	COMMONAREA_MEDI	34960	69.920%
MAX_AVG	24875	49.750%	ELEVATORS_MEDI	26651	53.302%
MIN_AVG	33894	67.788%	ENTRANCES_MEDI	25195	50.390%
EA_AVG	29721	59.442%	FLOORSMAX_MEDI	24875	49.750%
PARTMENTS_AVG	34226	68.452%	FLOORSMIN_MEDI	33894	67.788%
REA_AVG	25137	50.274%	LANDAREA_MEDI	29721	59.442%
LIVINGAPARTMENTS_AVG	34714	69.428%	LIVINGAPARTMENTS_MEDI	34226	68.452%
LIVINGAREA_AVG	27572	55.144%	LIVINGAREA_MEDI	25137	50.274%
ENTS_MODE	25385	50.770%	NONLIVINGAPARTMENTS_MEDI	34714	69.428%
NTAREA_MODE	29199	58.398%	NONLIVINGAREA_MEDI	27572	55.144%
BEGINEXPLUATATION_MODE	24394	48.788%	FONDKAPREMONT_MODE	34191	68.382%
BUILD_MODE	33239	66.478%	HOUSETYPE_MODE	25075	50.150%
ONAREA_MODE	34960	69.920%	TOTALAREA_MODE	24148	48.296%
ORS_MODE	26651	53.302%	WALLSMATERIAL_MODE	25459	50.918%
CES_MODE	25195	50.390%	EMERGENCYSTATE_MODE	23698	47.396%
MAX_MODE	24875	49.750%	OBS_30_CNT_SOCIAL_CIRCLE	168	0.34%
MIN_MODE	33894	67.788%	DEF_30_CNT_SOCIAL_CIRCLE	168	0.34%
EA_MODE	29721	59.442%	OBS_60_CNT_SOCIAL_CIRCLE	168	0.34%
PARTMENTS_MODE	34226	68.452%	DEF_60_CNT_SOCIAL_CIRCLE	168	0.34%
REA_MODE	25137	50.274%	DAYS_LAST_PHONE_CHANGE	1	0.0020%
<b>OBS_30_CNT_SOCIAL_CIRCLE</b>				<b>168</b>	<b>0.34%</b>
<b>DEF_30_CNT_SOCIAL_CIRCLE</b>				<b>168</b>	<b>0.34%</b>
<b>OBS_60_CNT_SOCIAL_CIRCLE</b>				<b>168</b>	<b>0.34%</b>
<b>DEF_60_CNT_SOCIAL_CIRCLE</b>				<b>168</b>	<b>0.34%</b>
<b>DAYS_LAST_PHONE_CHANGE</b>				<b>1</b>	<b>0.0020%</b>
<b>OWN_CAR_AGE</b>				<b>32950</b>	<b>65.900%</b>

## MISSING VALUES IN OTHER THAN SIGNIFICANT VARIABLES



- ➔ There is total 61 columns or variables with missing values.
- ➔ Among all 61 columns or variables COMMONAREA\_AVG has the highest % of missing value.
  - ➔ YEARS\_BEGINEXPLUATATION\_AVG has lowest % of missing value.

# TREATMENT OF MISSING VALUES HAVING MISSING VALUE LESS THAN 32%

VARIABLES	MEDIAN
AMT_ANNUITY	24939
AMT_GOODS_PRICE	450000
CNT_FAM_MEMBERS	0.0488
EXT_SOURCE_2	0.565585366
EXT_SOURCE_3	0.53527625
NAME_TYPE_SUITE	Unaccompanied
OCCUPATION_TYPE	Laborers
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

Formulas used: -

- ➔ =MEDIAN(K2:K50000) for numerical data
- ➔ =COUNTIF(K2:K50000,"Unaccompanied") most number of occurrence with 40627.
- ➔ Most number of people who came to apply for the loan application were unaccompanied.
  - ➔ Occupation with most number people is laborers i.e. 24606.
  - ➔ =MODE.MULT(BG2:BG50000) which stood to 0.

## TASK – 2 Identify Outliers in the Dataset

ME_TOT	OUTLIE	AMT_CREL	OUTLIE	AMT_ANNU	OUTLIE	AMT_GOODS_PR	OUTLIE	450000	TRUE	2517300	TRUE	72274.5	TRUE	2250000	TRUE
450000	TRUE	1971072	TRUE	62019	TRUE	1800000	TRUE	675000	TRUE	1942560	TRUE	106582.5	TRUE	1800000	TRUE
450000	TRUE	2286211.5	TRUE	116266.5	TRUE	2182500	TRUE	360000	TRUE	2517300	TRUE	69223.5	TRUE	2250000	TRUE
450000	TRUE	2125953	TRUE	81108	TRUE	1984500	TRUE	450000	TRUE	1928304	TRUE	78669	TRUE	1800000	TRUE
360000	TRUE	2250000	TRUE	116505	TRUE	2250000	TRUE	540000	TRUE	1800000	TRUE	62568	TRUE	1800000	TRUE
450000	TRUE	1965226.5	TRUE	68440.5	TRUE	1696500	TRUE	360000	TRUE	1710000	TRUE	74565	TRUE	1710000	TRUE
360000	TRUE	1872517.5	TRUE	65088	TRUE	1710000	TRUE	675000	TRUE	1800000	TRUE	62698.5	TRUE	1800000	TRUE
351000	TRUE	2085120	TRUE	72607.5	TRUE	1800000	TRUE	708000	TRUE	2250000	TRUE	225000	TRUE	2250000	TRUE
405000	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE	450000	TRUE	2022565.5	TRUE	70434	TRUE	1746000	TRUE
405000	TRUE	1928304	TRUE	79708.5	TRUE	1800000	TRUE	436500	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE
675000	TRUE	2085120	TRUE	72607.5	TRUE	1800000	TRUE	562500	TRUE	1800000	TRUE	62698.5	TRUE	1800000	TRUE
414000	TRUE	2447937	TRUE	62010	TRUE	2254500	TRUE	585000	TRUE	1800000	TRUE	62698.5	TRUE	1800000	TRUE
468000	TRUE	2356920	TRUE	119731.5	TRUE	2250000	TRUE	749331	TRUE	3956274	TRUE	197230.5	TRUE	3825000	TRUE
540000	TRUE	2245500	TRUE	83214	TRUE	2245500	TRUE	360000	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE
1350000	TRUE	2410380	TRUE	109053	TRUE	2250000	TRUE	396000	TRUE	1928304	TRUE	71590.5	TRUE	1800000	TRUE
787500	TRUE	2085120	TRUE	72607.5	TRUE	1800000	TRUE	50846.5	TRUE	4050000	TRUE	146002.5	TRUE	4050000	TRUE
720000	TRUE	1928304	TRUE	71590.5	TRUE	1800000	TRUE	382500	TRUE	2085120	TRUE	72607.5	TRUE	1800000	TRUE
360000	TRUE	1971072	TRUE	62019	TRUE	1800000	TRUE	382500	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE
426019.5	TRUE	2447937	TRUE	61879.5	TRUE	2254500	TRUE	391500	TRUE	2298843	TRUE	69831	TRUE	1984500	TRUE
360000	TRUE	1682446.5	TRUE	66721.5	TRUE	1570500	TRUE	360000	TRUE	2250000	TRUE	68220	TRUE	2250000	TRUE
450000	TRUE	2250000	TRUE	91642.5	TRUE	2250000	TRUE	360000	TRUE	1885536	TRUE	95782.5	TRUE	1800000	TRUE
405000	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE	405000	TRUE	1928304	TRUE	71590.5	TRUE	1800000	TRUE
450000	TRUE	1800000	TRUE	74416.5	TRUE	1800000	TRUE	450000	TRUE	2517300	TRUE	93420	TRUE	2250000	TRUE
1035000	TRUE	2695500	TRUE	74254.5	TRUE	2250000	TRUE	360000	TRUE	1971072	TRUE	62019	TRUE	1800000	TRUE
405000	TRUE	1762110	TRUE	67248	TRUE	1575000	TRUE	495000	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE
382500	TRUE	2925000	TRUE	73944	TRUE	2925000	TRUE	450000	TRUE	2517300	TRUE	66532.5	TRUE	2250000	TRUE
540000	TRUE	1696981.5	TRUE	93123	TRUE	1620000	TRUE	630000	TRUE	1971072	TRUE	62019	TRUE	1800000	TRUE
472500	TRUE	2606400	TRUE	79155	TRUE	2250000	TRUE	783000	TRUE	4050000	TRUE	258025.5	TRUE	4050000	TRUE
427500	TRUE	2428200	TRUE	123480	TRUE	2250000	TRUE	450000	TRUE	2250000	TRUE	116505	TRUE	2250000	TRUE
360000	TRUE	1800000	TRUE	62568	TRUE	1800000	TRUE	382500	TRUE	2085120	TRUE	72477	TRUE	1800000	TRUE
607500	TRUE	2085120	TRUE	72607.5	TRUE	1800000	TRUE	38746.5	TRUE	2250000	TRUE	95463	TRUE	2250000	TRUE
360000	TRUE	1822500	TRUE	69417	TRUE	1822500	TRUE	360000	TRUE	1762110	TRUE	71770.5	TRUE	1575000	TRUE
450000	TRUE	2517300	TRUE	72274.5	TRUE	2250000	TRUE	540000	TRUE	1800000	TRUE	62568	TRUE	1800000	TRUE
675000	TRUE	2250000	TRUE	225000	TRUE	2250000	TRUE	900000	TRUE	2250000	TRUE	112500	TRUE	2250000	TRUE
540000	TRUE	2303460	TRUE	213291	TRUE	2250000	TRUE	450000	TRUE	1885536	TRUE	95913	TRUE	1800000	TRUE
616500	TRUE	2695500	TRUE	71235	TRUE	2250000	TRUE	630000	TRUE	2366613	TRUE	71883	TRUE	2043000	TRUE
405000	TRUE	1800000	TRUE	62568	TRUE	1800000	TRUE	360000	TRUE	2517300	TRUE	69354	TRUE	2250000	TRUE
405000	TRUE	1800000	TRUE	62698.5	TRUE	1800000	TRUE	342000	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE
364500	TRUE	1800000	TRUE	62568	TRUE	1800000	TRUE	405000	TRUE	2250000	TRUE	225000	TRUE	2250000	TRUE
450000	TRUE	2695500	TRUE	71235	TRUE	2250000	TRUE	562500	TRUE	2695500	TRUE	74124	TRUE	2250000	TRUE
391500	TRUE	1800000	TRUE	74416.5	TRUE	1800000	TRUE	414000	TRUE	2250000	TRUE	68350.5	TRUE	2250000	TRUE
360000	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE	382500	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE
382500	TRUE	1665000	TRUE	114201	TRUE	1665000	TRUE	450000	TRUE	1885536	TRUE	97762.5	TRUE	1800000	TRUE
540000	TRUE	2517300	TRUE	66532.5	TRUE	2250000	TRUE	495000	TRUE	2517300	TRUE	69354	TRUE	2250000	TRUE
360000	TRUE	1735299	TRUE	75667.5	TRUE	1638000	TRUE	342000	TRUE	2447937	TRUE	61879.5	TRUE	2254500	TRUE
427500	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE	742500	TRUE	2517300	TRUE	66532.5	TRUE	2250000	TRUE
900000	TRUE	1800000	TRUE	171040.5	TRUE	1800000	TRUE	360000	TRUE	1928304	TRUE	73579.5	TRUE	1800000	TRUE
675000	TRUE	2250000	TRUE	225000	TRUE	2250000	TRUE	360000	TRUE	2300814	TRUE	66069	TRUE	2056500	TRUE
382500	TRUE	1687266	TRUE	68719.5	TRUE	1575000	TRUE	391500	TRUE	1942560	TRUE	106582.5	TRUE	1800000	TRUE
450000	TRUE	2695500	TRUE	74254.5	TRUE	2250000	TRUE	450000	TRUE	2250000	TRUE	62005.5	TRUE	2250000	TRUE
540000	TRUE	2517300	TRUE	66402	TRUE	2250000	TRUE	540000	TRUE	1971072	TRUE	62019	TRUE	1800000	TRUE
585000	TRUE	1800000	TRUE	173704.5	TRUE	1800000	TRUE	450000	TRUE	2059056	TRUE	71703	TRUE	1777500	TRUE
427500	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE	405000	TRUE	2356920	TRUE	119731.5	TRUE	2250000	TRUE
450000	TRUE	2250000	TRUE	61875	TRUE	2250000	TRUE	360000	TRUE	2032992	TRUE	70794	TRUE	1755000	TRUE
450225	TRUE	2687355	TRUE	68062.5	TRUE	2475000	TRUE	900000	TRUE	1864152	TRUE	127989	TRUE	1800000	TRUE
450000	TRUE	1885536	TRUE	97762.5	TRUE	1800000	TRUE	450000	TRUE	1928304	TRUE	73449	TRUE	1800000	TRUE
382500	TRUE	1800000	TRUE	62698.5	TRUE	1800000	TRUE	482000	TRUE	1933285.5	TRUE	79915.5	TRUE	1728000	TRUE
360000	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE	436500	TRUE	1687266	TRUE	64395	TRUE	1575000	TRUE
360000	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE	405000	TRUE	1800000	TRUE	123462	TRUE	1800000	TRUE
360000	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE	360000	TRUE	1941507	TRUE	67617	TRUE	1773000	TRUE
675000	TRUE	1928304	TRUE	79708.5	TRUE	1800000	TRUE	517500	TRUE	1800000	TRUE	62698.5	TRUE	1800000	TRUE
810000	TRUE	1800000	TRUE	62568	TRUE	1800000	TRUE	720000	TRUE	2517300	TRUE	69354	TRUE	2250000	TRUE
382500	TRUE	1971072	TRUE	68512.5	TRUE	1800000	TRUE	765000	TRUE	2250000	TRUE	112500	TRUE	2250000	TRUE
382500	TRUE	1875276	TRUE	69624	TRUE	1750500	TRUE	441000	TRUE	1750500	TRUE	114885	TRUE	1750500	TRUE
900000	TRUE	2250000	TRUE	112500	TRUE	2250000	TRUE	450000	TRUE	2517300	TRUE	66532.5	TRUE	2250000	TRUE
360000	TRUE	1692085.5	TRUE	69961.5	TRUE	1579500	TRUE	373500	TRUE	2200122	TRUE	63054	TRUE	1966500	TRUE

450000	TRUE	1800000	TRUE	1800000	TRUE	1800000	TRUE
540000	TRUE	1668555	TRUE	108018	TRUE	1575000	TRUE
540000	TRUE	1971072	TRUE	68643	TRUE	1800000	TRUE
405000	TRUE	2217456	TRUE	67360.5	TRUE	2025000	TRUE
382500	TRUE	2085120	TRUE	72607.5	TRUE	1800000	TRUE
675000	TRUE	1928304	TRUE	75582	TRUE	1800000	TRUE
360000	TRUE	2250000	TRUE	70650	TRUE	2250000	TRUE
382500	TRUE	1872517.5	TRUE	65218.5	TRUE	1710000	TRUE
540000	TRUE	2250000	TRUE	95332.5	TRUE	2250000	TRUE
360000	TRUE	1834906.5	TRUE	63909	TRUE	1584000	TRUE
1125000	TRUE	1885536	TRUE	97632	TRUE	1800000	TRUE
360000	TRUE	1823242.5	TRUE	63504	TRUE	1665000	TRUE

## QUARTILE FUNCTION

AMT_INCOME_TOTAL	
q1	112500
q3	202500
iqr	90000
Upper bound	337500
Lower bound	-22500

AMT_CREDIT	
q1	270000
q3	808650
iqr	538650
Upper bound	1616625
Lower bound	-537975

AMT_ANNUITY	
q1	16456.5
q3	34596
iqr	18139.5
Upper bound	61805.25
Lower bound	-10752.8

AMT_GOODS_PRICE	
q1	238500
q3	679500
iqr	441000
Upper bound	1341000
Lower bound	-423000

→ There is total 568 outliers across 4 significant columns.

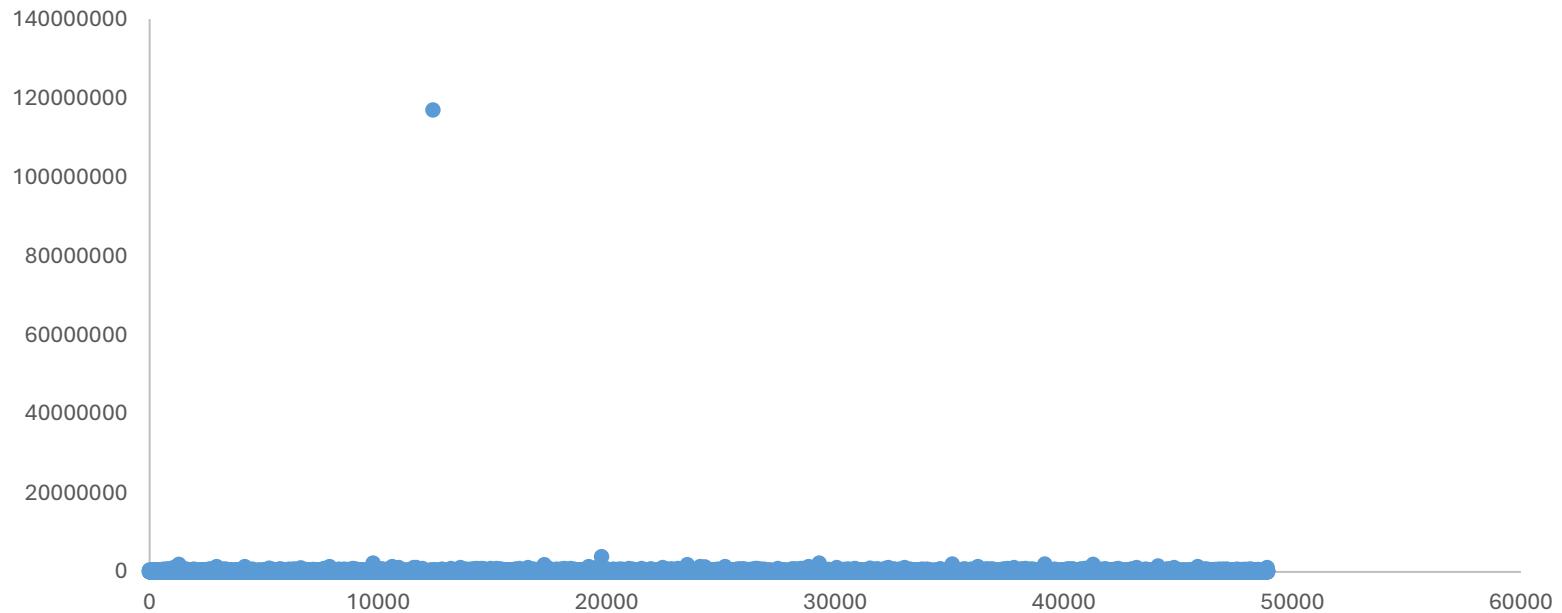
→ Each column has 142 outliers.

→ Formulas used

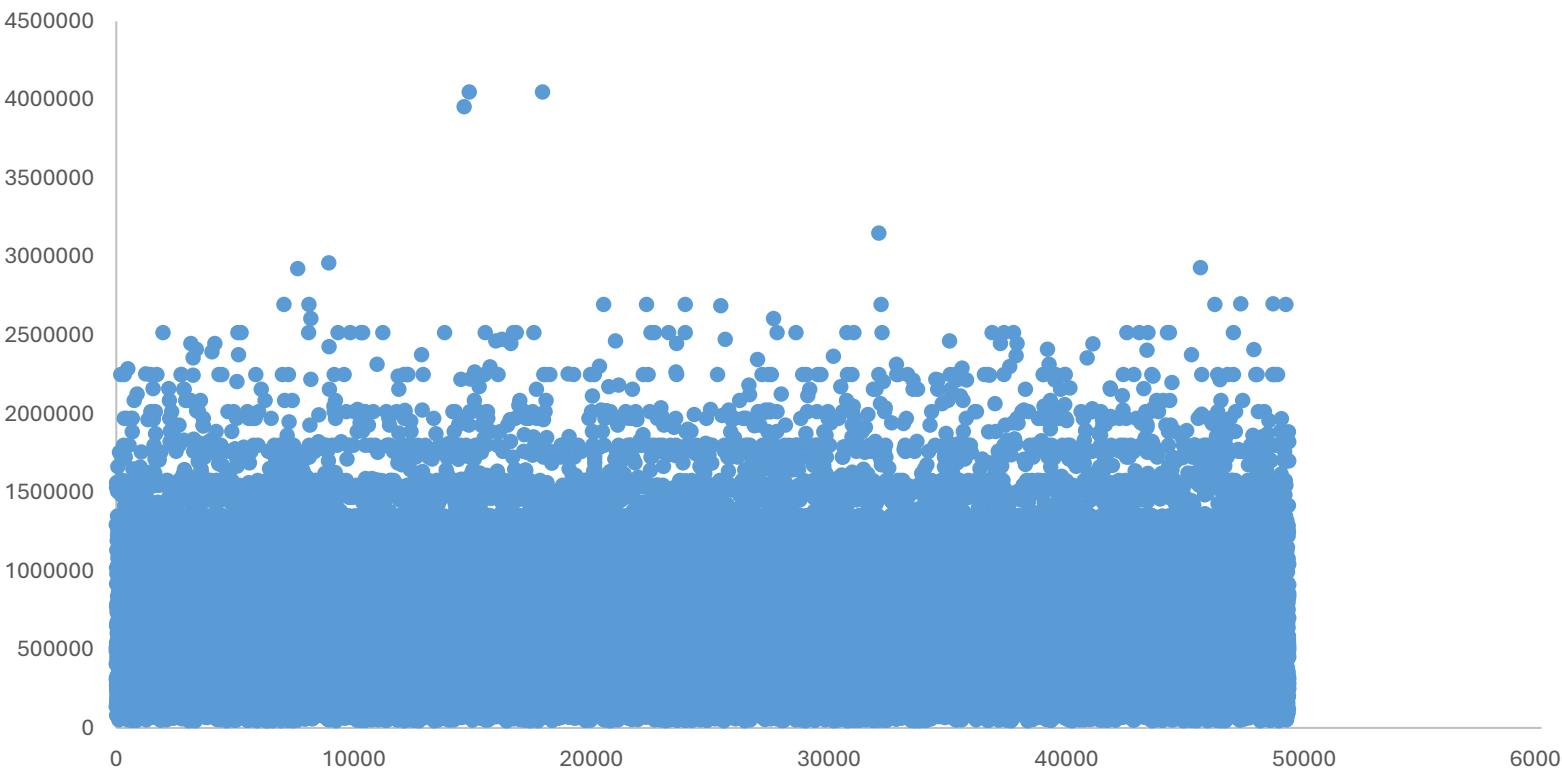
- $Q1 = QUARTILE.EXC(\$A\$2: \$A\$50000, 1)$
- $Q3 = QUARTILE.EXC(\$A\$2: \$A\$50000, 3)$ 
  - $IQR = Q3 - Q1$
- $Upper\ Bound = Q3 + (1.5 * IQR) = L4 + (1.5 * L5)$
- $Lower\ Bound = Q1 - (1.5 * IQR) = L3 - (1.5 * L5)$

# GRAPHICAL PRESENSTATION OF OUTLIERS

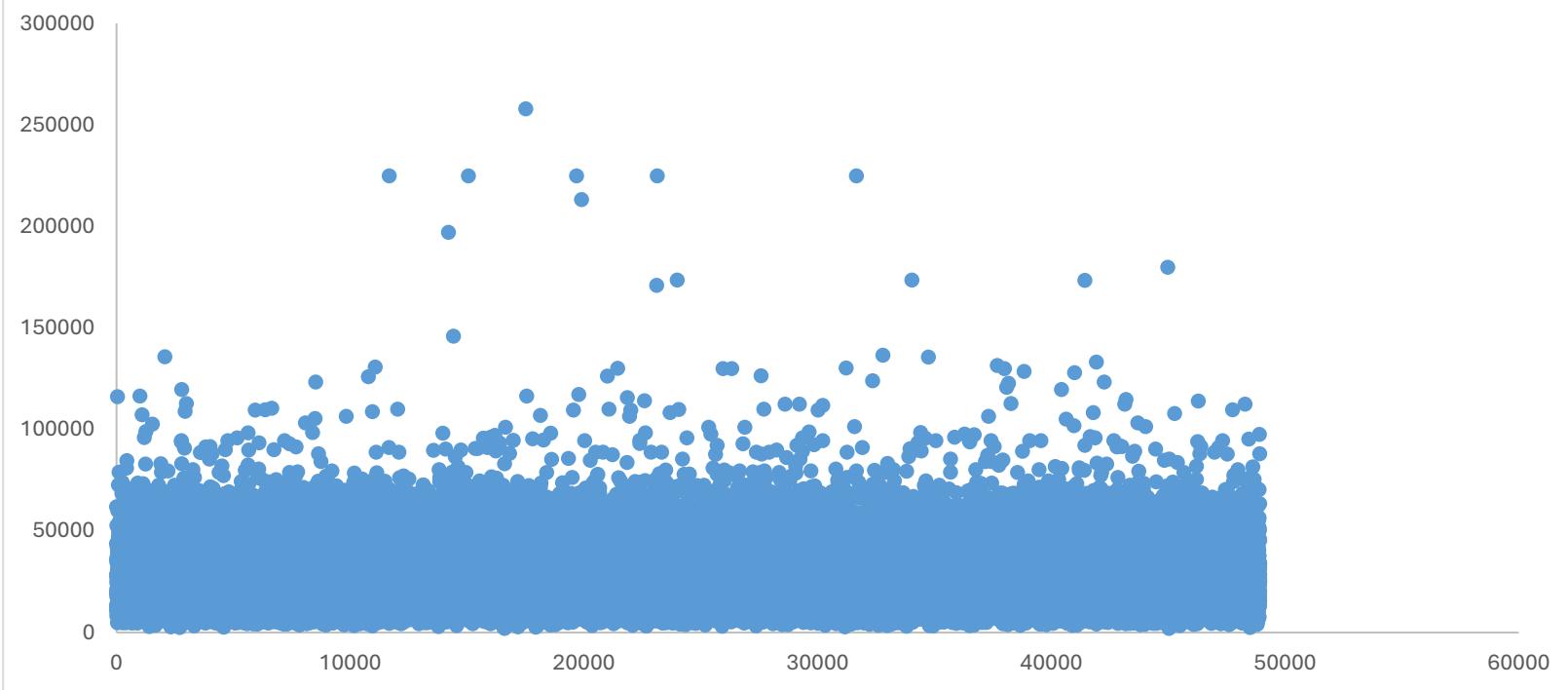
## OUTLIERS IN TOTAL INCOME COLUMN



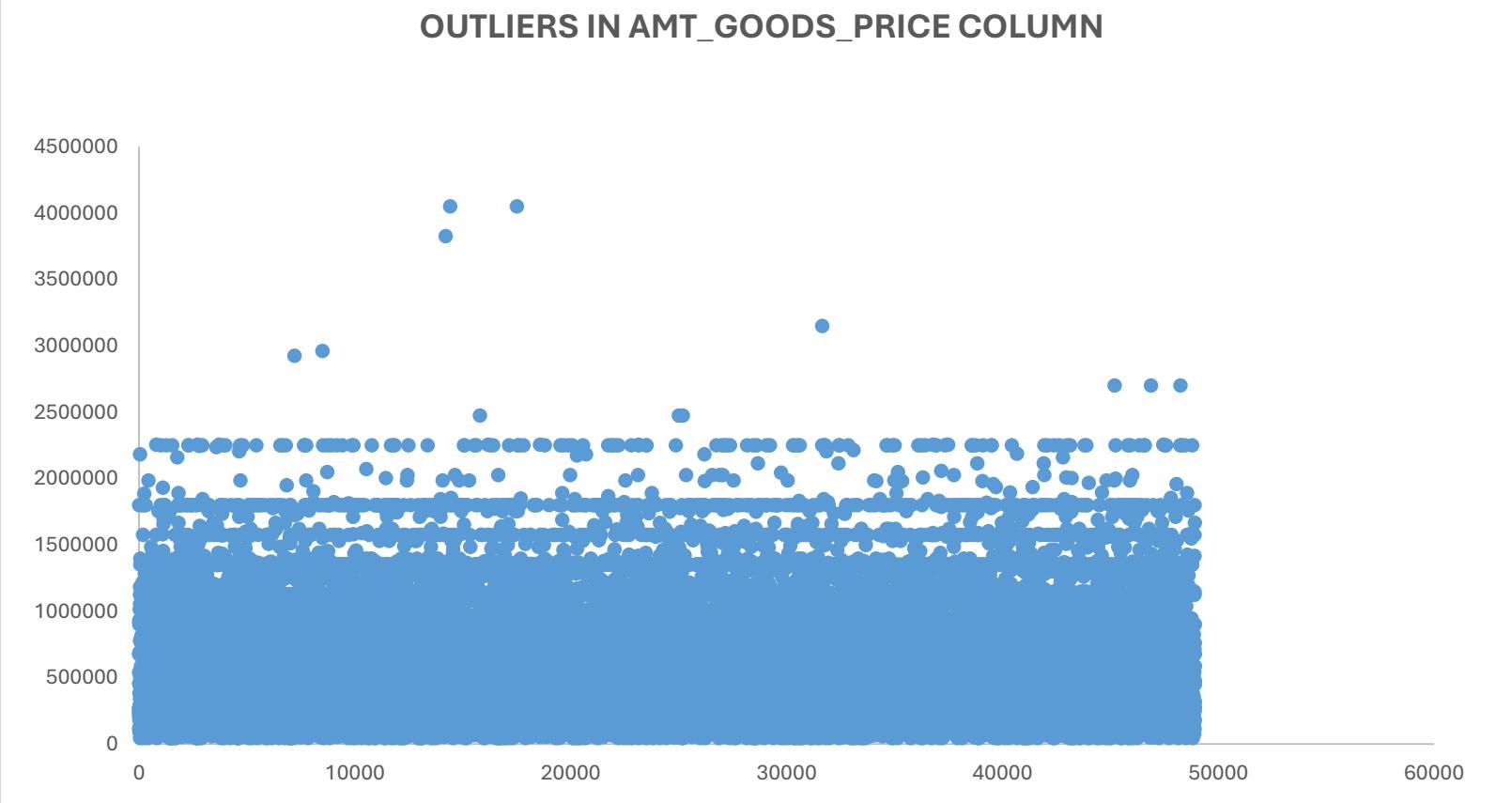
## OUTLIERS IN AMT\_CREDIT COLUMN



## OUTLIERS ANNUITY COLUMN



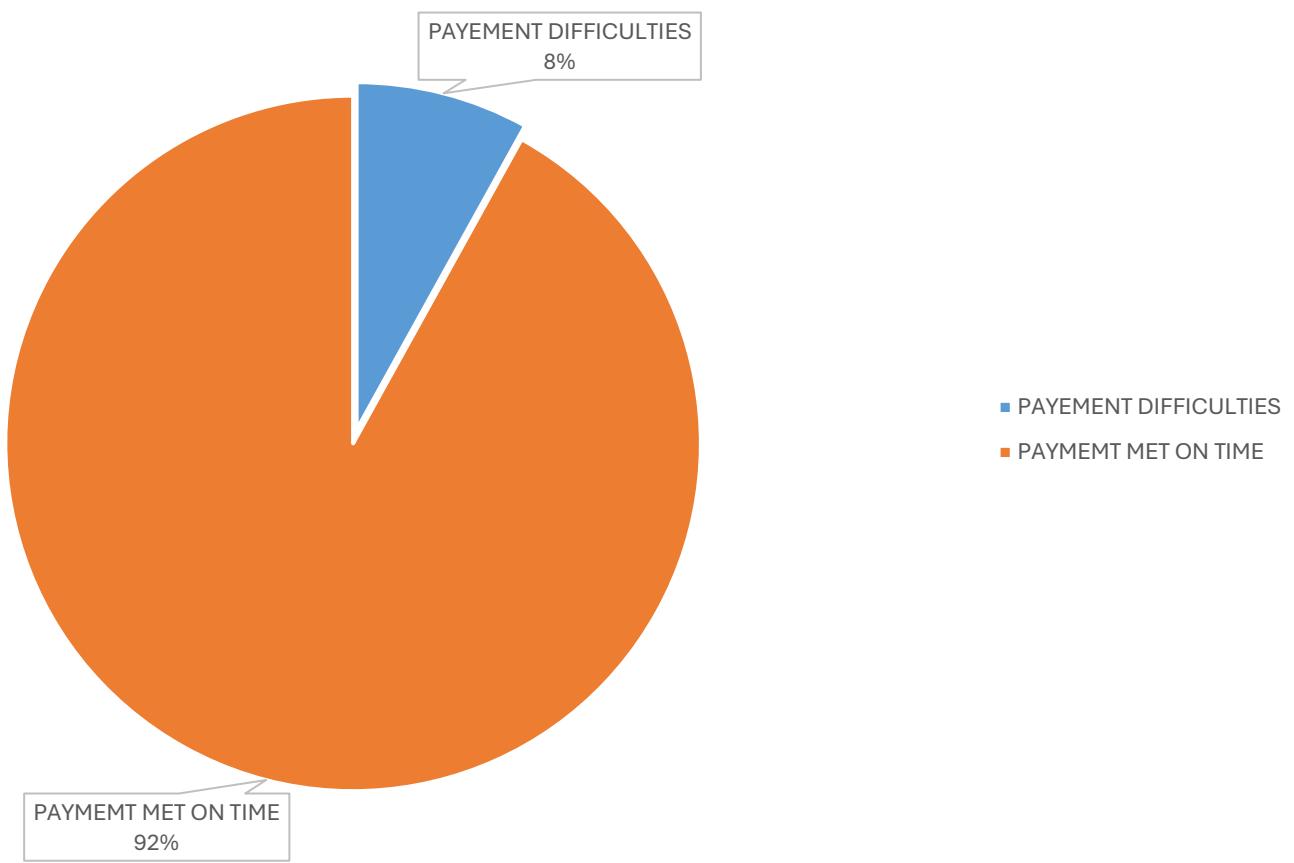
## OUTLIERS IN AMT\_GOODS\_PRICE COLUMN



## TASK 3- DATA IMBALANCE

PAYMENT STATUS	COUNT OF PAYMENT STATUS
PAYEMENT DIFFICULTIES	4026
PAYMEMT MET ON TIME	45973

DATA IMBALANCE IN TARGET VARIABLE

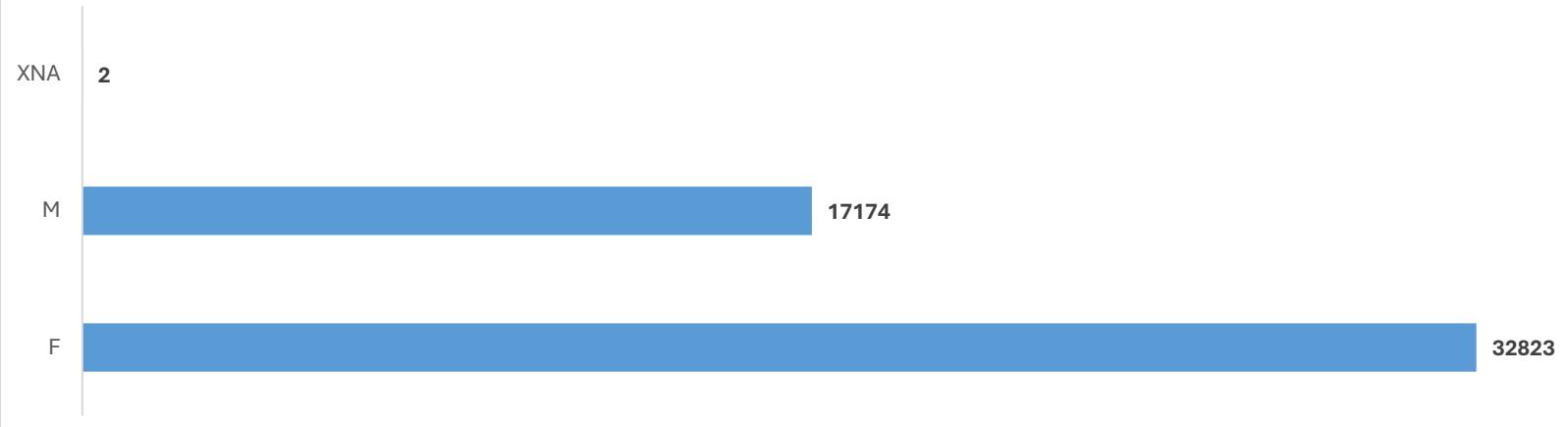


- Number of payments met on time over shadows number payment difficulties by 84%.
- While overall share of payments met on time stood to 92% and payment difficulties 8%.

## TASK 4 – Perform Univariate, Segmented Univariate, and Bivariate Analysis

GENDER TYPE	COUNT OF DIFFERENT GENDER TYPE
F	32823
M	17174
XNA	2

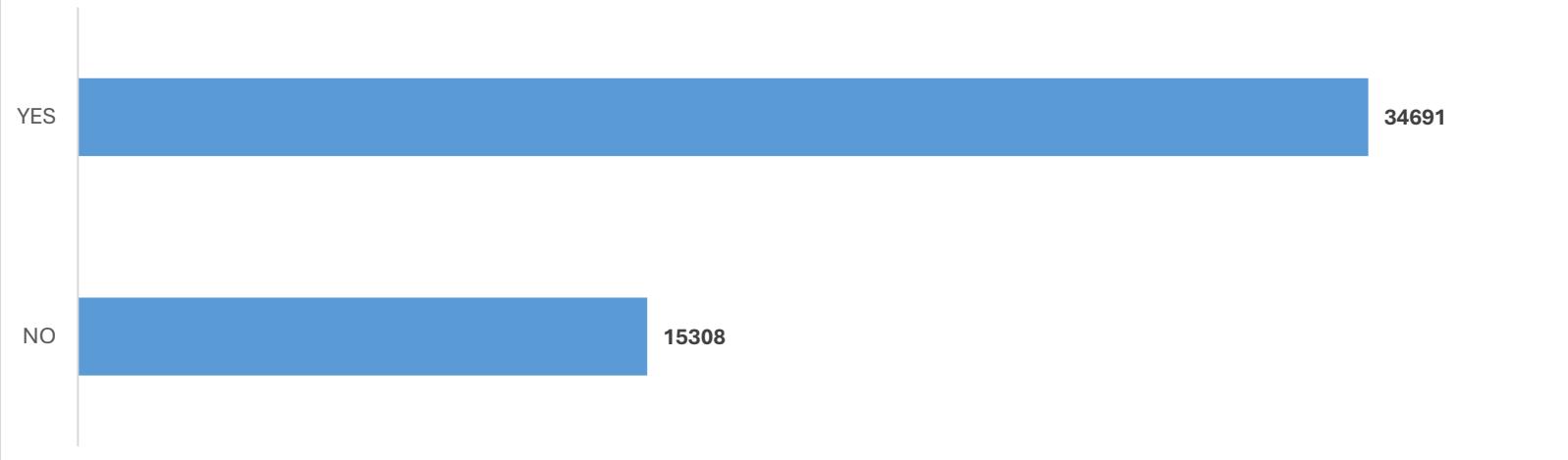
COUNT OF DIFFERENT GENDER TYPE



➔ There are more females loan applicants than males.

PROPERTY OWNER	Count of PROPERTY OWNER
NO	15308
YES	34691

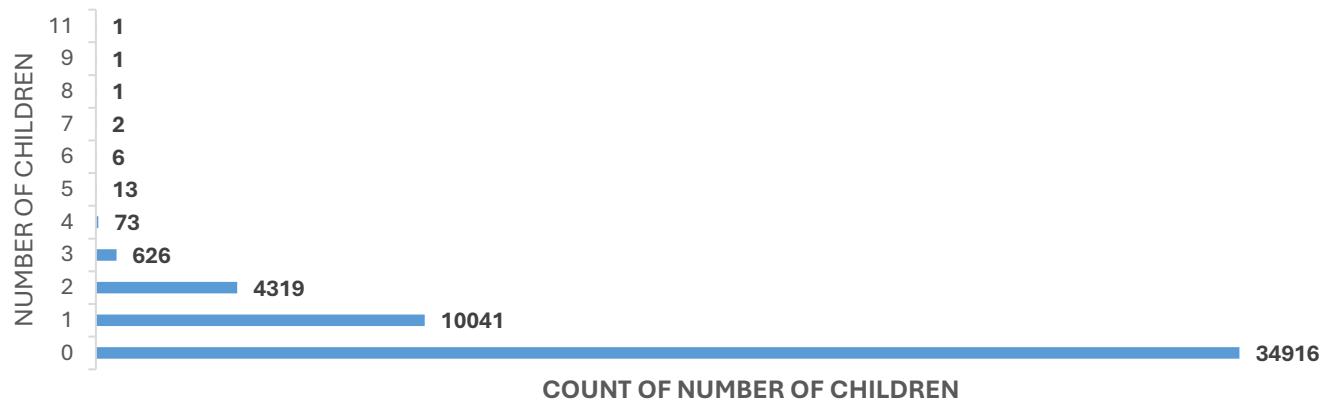
DISTRIBUTION OF PROPERTY OWNER COUNT



➔ There are a greater number of loan applicant with property to their name than applicant with no property to their name.

NUMBER OF CHILDREN	COUNT OF NUMBER OF CHILDREN
0	34916
1	10041
2	4319
3	626
4	73
5	13
6	6
7	2
8	1
9	1
11	1

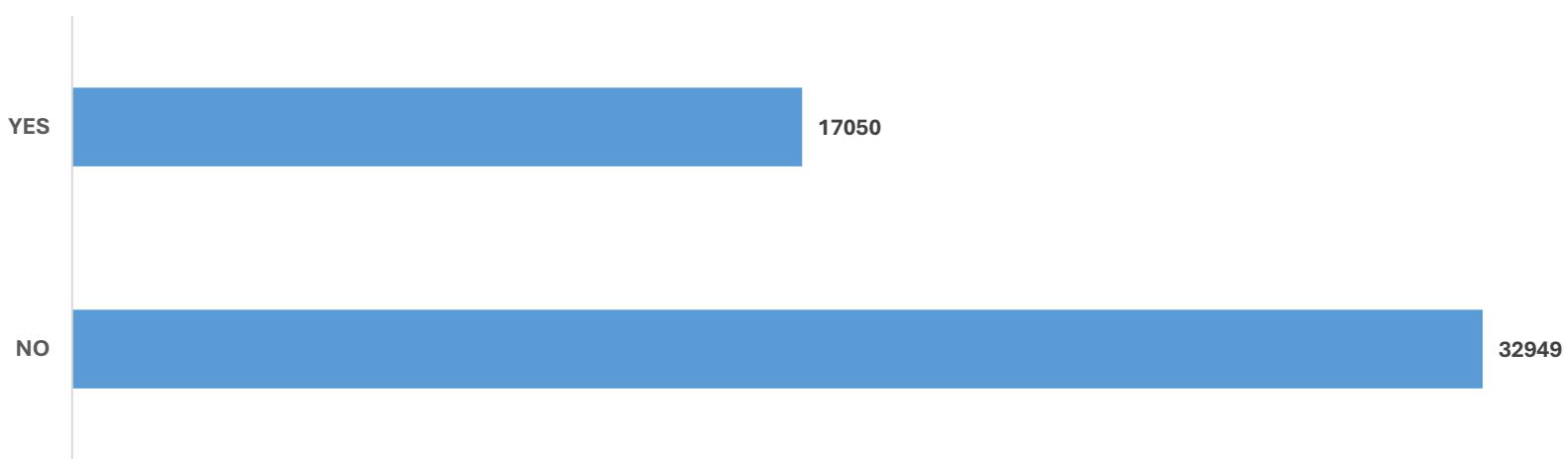
### DISTRIBUTION OF CHILDREN COUNT



➔ Most number of loan applicant have no children.

CAR OWNER	Count of CAR OWNER
NO	32949
YES	17050

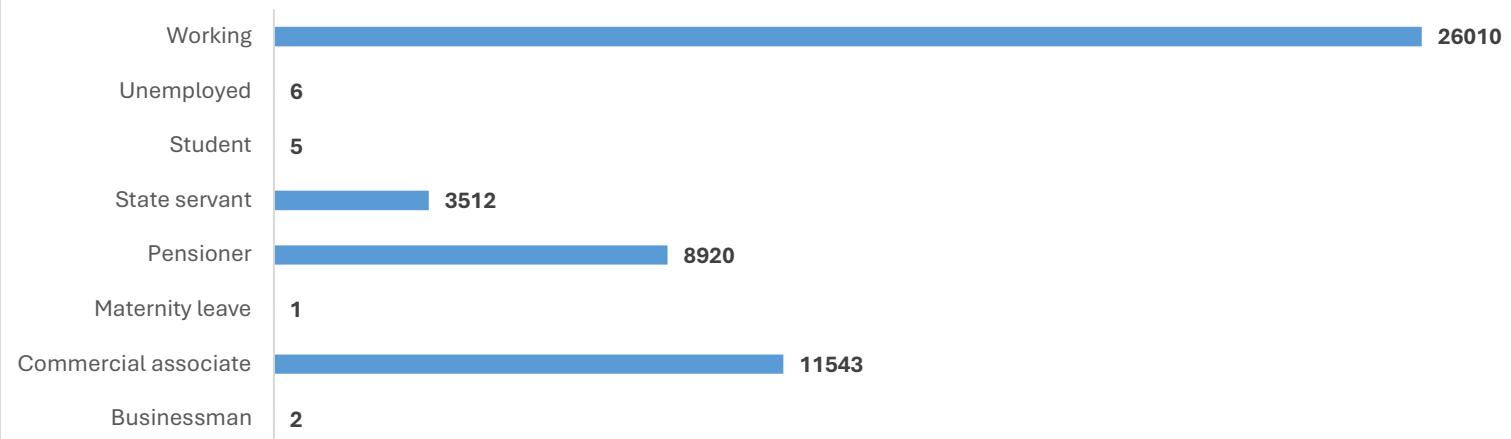
### DISTRIBUTION OF CAR OWNER



➔ Most number of loan applicants are non-car owner.

DIFFERENT INCOME TYPE	COUNT OF DIFFERENT INCOME TYPE
Businessman	2
Commercial associate	11543
Maternity leave	1
Pensioner	8920
State servant	3512
Student	5
Unemployed	6
Working	26010

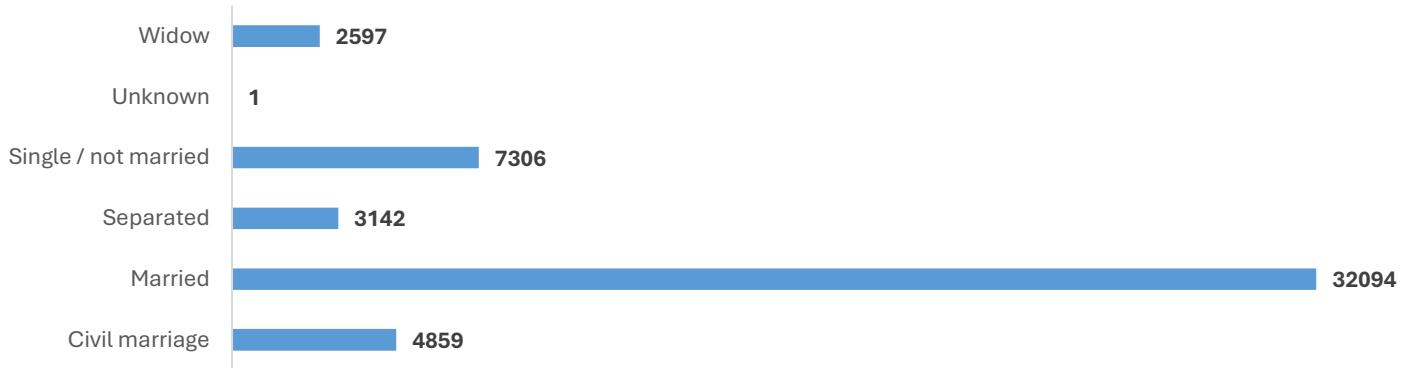
### DISTRIBUTION OF DIFFERENT INCOME TYPE



→ Most number of loan applicants' income type is working.

FAMILY STATUS	COUNT OF FAMILY STATUS
Civil marriage	4859
Married	32094
Separated	3142
Single / not married	7306
Unknown	1
Widow	2597

### DISTRIBUTION OF DIFFERENT FAMILY STATUS

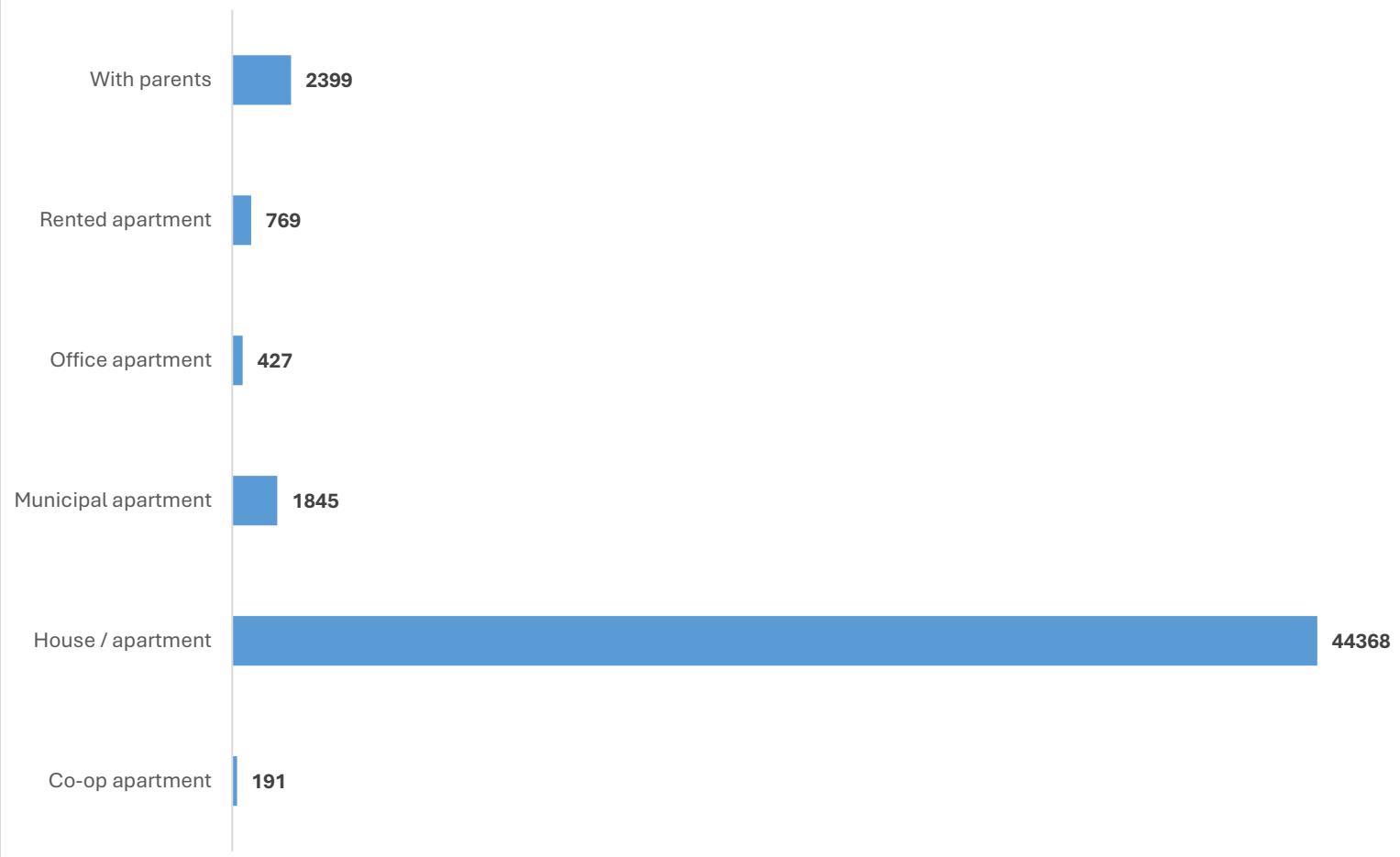


→ Most number of people who applied for the loan are married.

HOUSING TYPE	COUNT OF DIFFERENT HOUSING TYPE
Co-op apartment	191
House / apartment	44368
Municipal apartment	1845
Office apartment	427
Rented apartment	769
With parents	2399

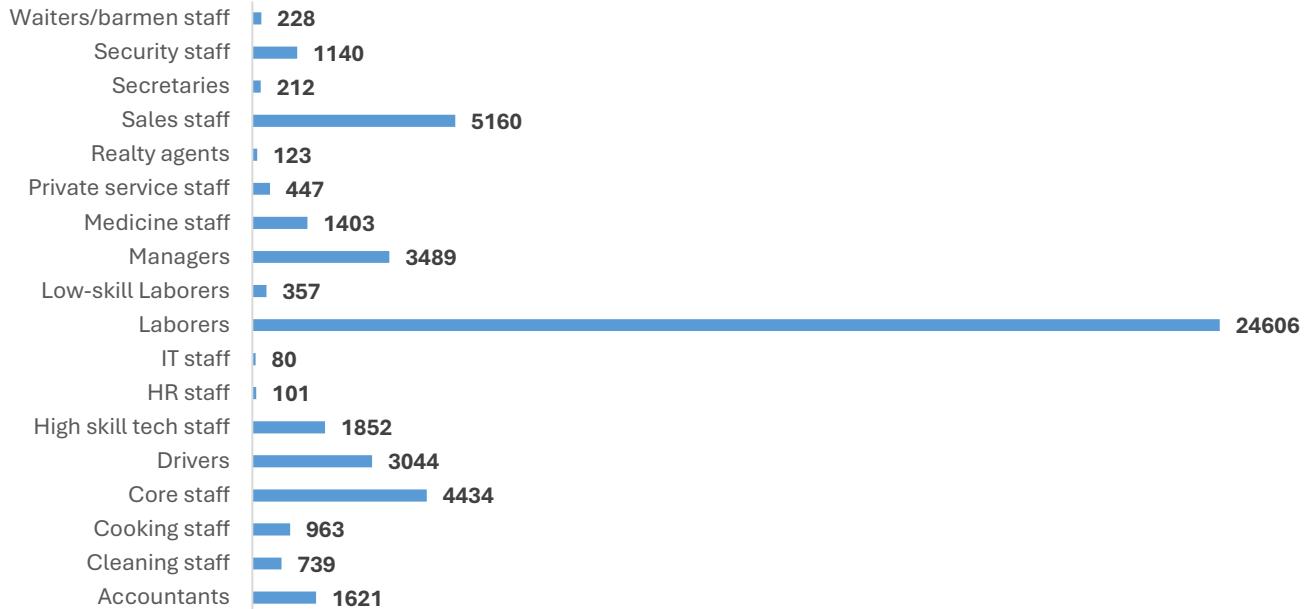
→ Most number of people who applied for the loan live in House/apartment.

### DISTRIBUTION OF DIFFERENT HOUSING TYPE



OCCUPATION_TYPE	Count of OCCUPATION_TYPE
Accountants	1621
Cleaning staff	739
Cooking staff	963
Core staff	4434
Drivers	3044
High skill tech staff	1852
HR staff	101
IT staff	80
Laborers	24606
Low-skill Laborers	357
Managers	3489
Medicine staff	1403
Private service staff	447
Realty agents	123
Sales staff	5160
Secretaries	212
Security staff	1140
Waiters/barmen staff	228

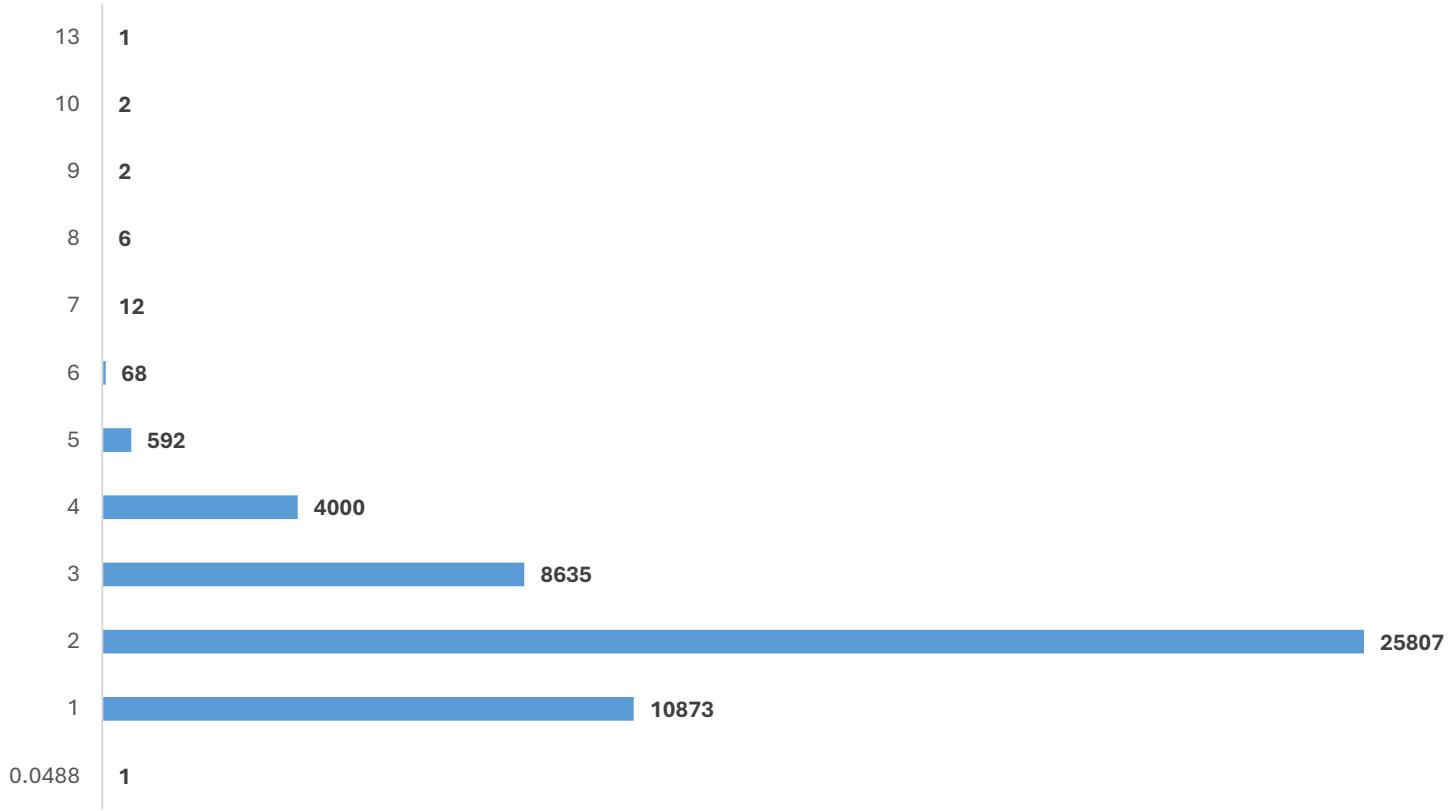
## DISTRIBUTION OF DIFFERENT OCCUPATION TYPE



- ➔ Most number of people who applied for the loan are laborers i.e. 24606.
- ➔ Least number of applicants who applied for the loan have occupation type HR STAFF.

FAMILY SIZE	Count of family size
0.0488	1
1	10873
2	25807
3	8635
4	4000
5	592
6	68
7	12
8	6
9	2
10	2
13	1

### DISTRIBUTION OF DIFFERENT FAMILY SIZE

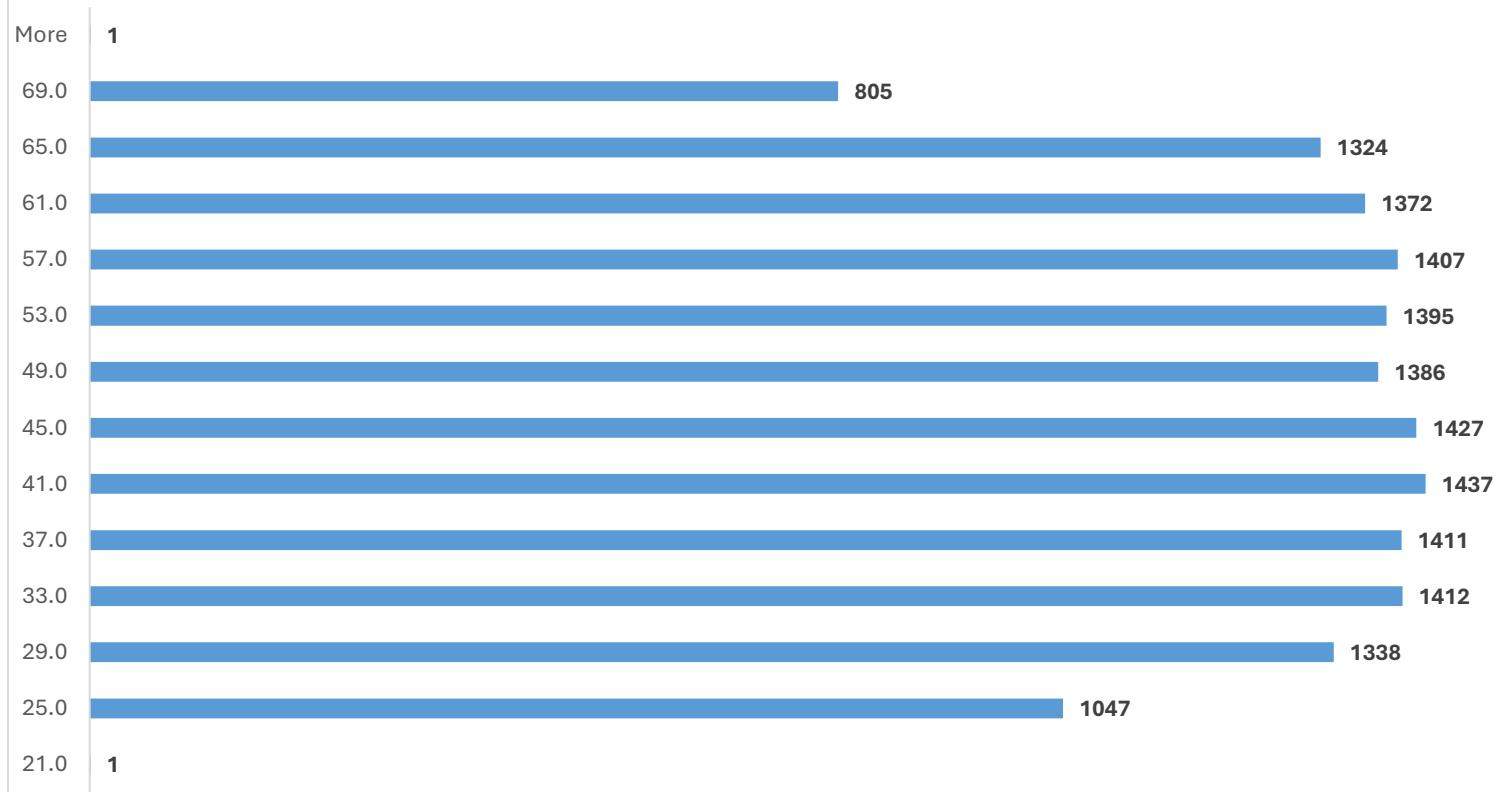


➔ Among all applicants who applied for the loan most of them have a family size of 2.

AGE	Count of applicants
21.0	1
25.0	1047
29.0	1338
33.0	1412
37.0	1411
41.0	1437
45.0	1427
49.0	1386
53.0	1395
57.0	1407
61.0	1372
65.0	1324
69.0	805
More	1

→ Most Number of applicants applied for the loan are 41 years in age.

### AGE DISTRIBUTION

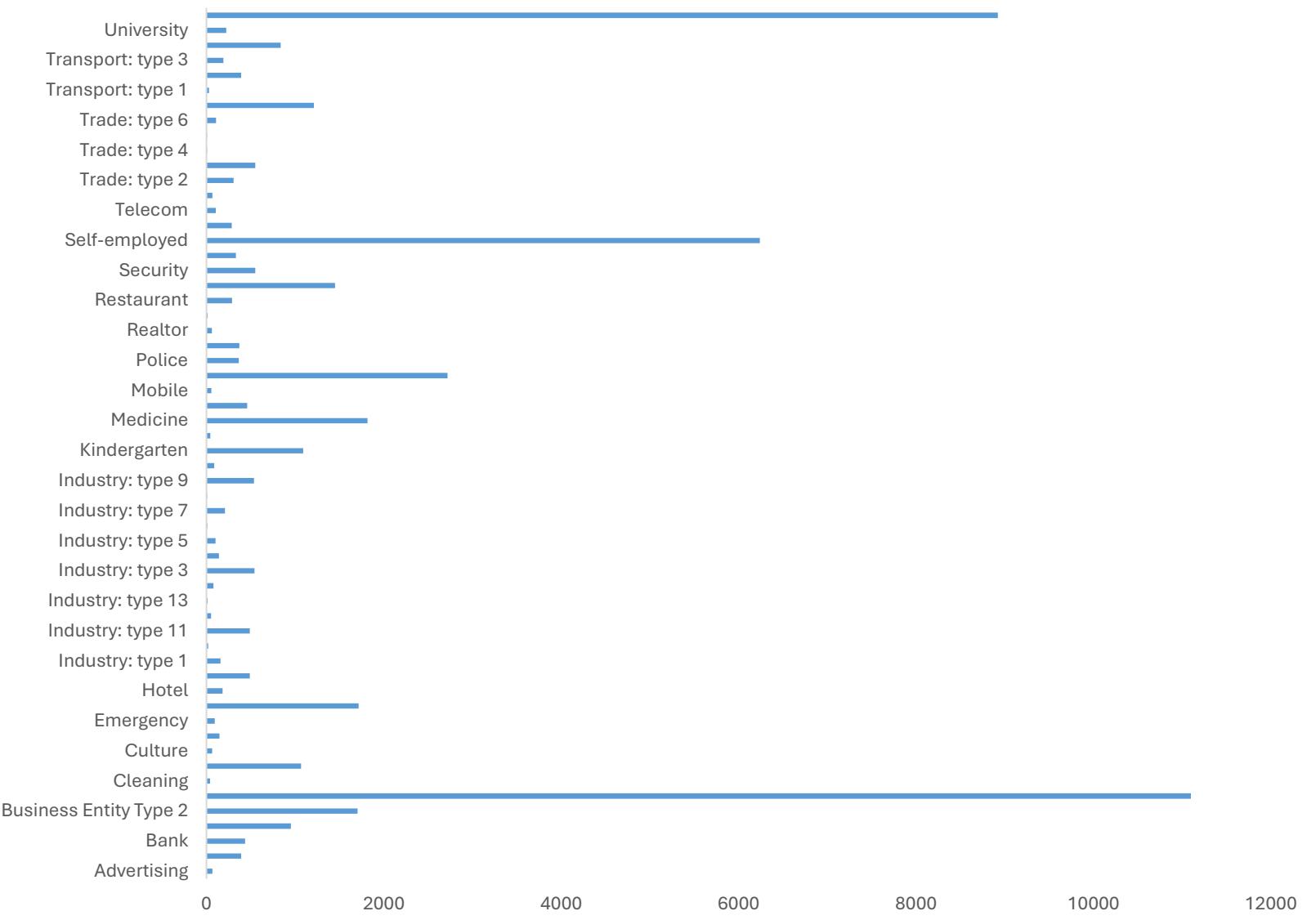


Organization Type	Count of ORGANIZATION_TYPE		
Advertising	68	Industry: type 8	8
Agriculture	392	Industry: type 9	537
Bank	435	Insurance	89
Business Entity Type 1	953	Kindergarten	1090
Business Entity Type 2	1704	Legal Services	44
Business Entity Type 3	11101	Medicine	1817
Cleaning	40	Military	458
Construction	1066	Mobile	56
Culture	64	Other	2717
Electricity	147	Police	366
Emergency	93	Postal	370
Government	1716	Realtor	61
Hotel	182	Religion	14
Housing	489	Restaurant	289
Industry: type 1	159	School	1450
Industry: type 10	21	Security	550
Industry: type 11	489	Security Ministries	331
Industry: type 12	489	Self-employed	6240
Industry: type 13	53	Services	284
Industry: type 2	15	Telecom	106
Industry: type 3	78	Trade: type 1	66
Industry: type 4	542	Trade: type 2	307
Industry: type 5	140	Trade: type 3	550
Industry: type 6	103	Trade: type 4	8
Industry: type 7	12	Trade: type 5	8
	209	Trade: type 6	108

<b>Trade: type 7</b>	<b>1210</b>
<b>Transport: type 1</b>	<b>28</b>
<b>Transport: type 2</b>	<b>392</b>
<b>Transport: type 3</b>	<b>191</b>
<b>Transport: type 4</b>	<b>837</b>
<b>University</b>	<b>222</b>
<b>XNA</b>	<b>8924</b>

# GRAPHICAL PRESENTATION OF ORANGANIZATION TYPE

## DISTRIBUTION OF ORGANIZATION TYPE

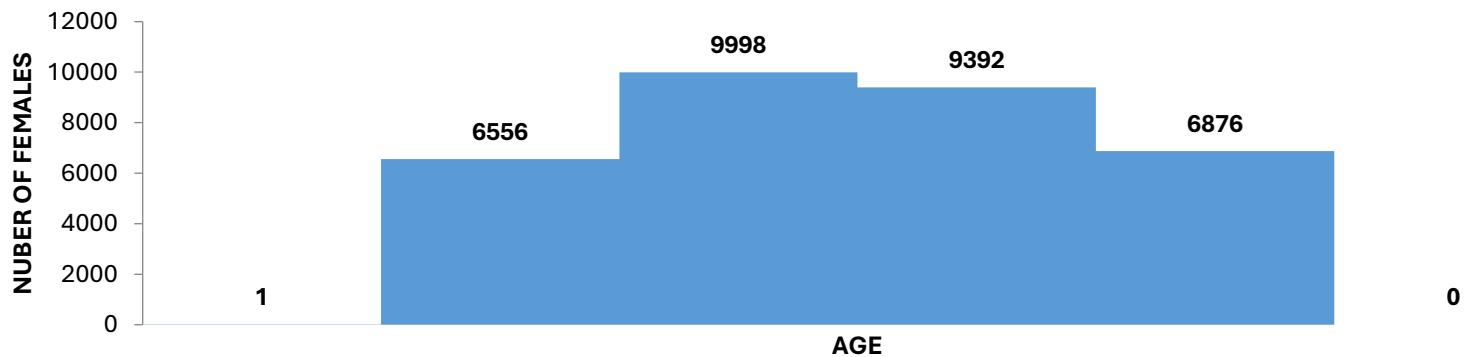


➔ Most number of people who applied for the loan belongs to “business entity type 2” organization type.

## SEGMENTED UNIVARIATE ANALYSIS

FEMALE'S AGE	Frequency
21.0	1
33.0	6556
45.0	9998
57.0	9392
69.0	6876
More	0

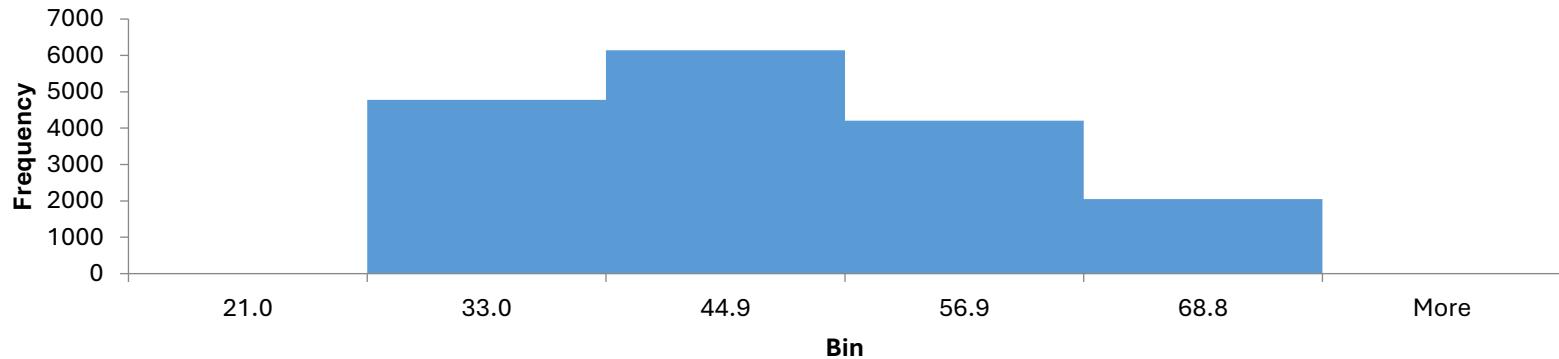
FEMALE CLIENT'S AGE DISTRIBUTION



→ Among age distribution, average age of a female who applied for the loan is 45 years.

MALE'S AGE	Count of males
21.0	1
33.0	4778
44.9	6139
56.9	4206
68.8	2049
More	1

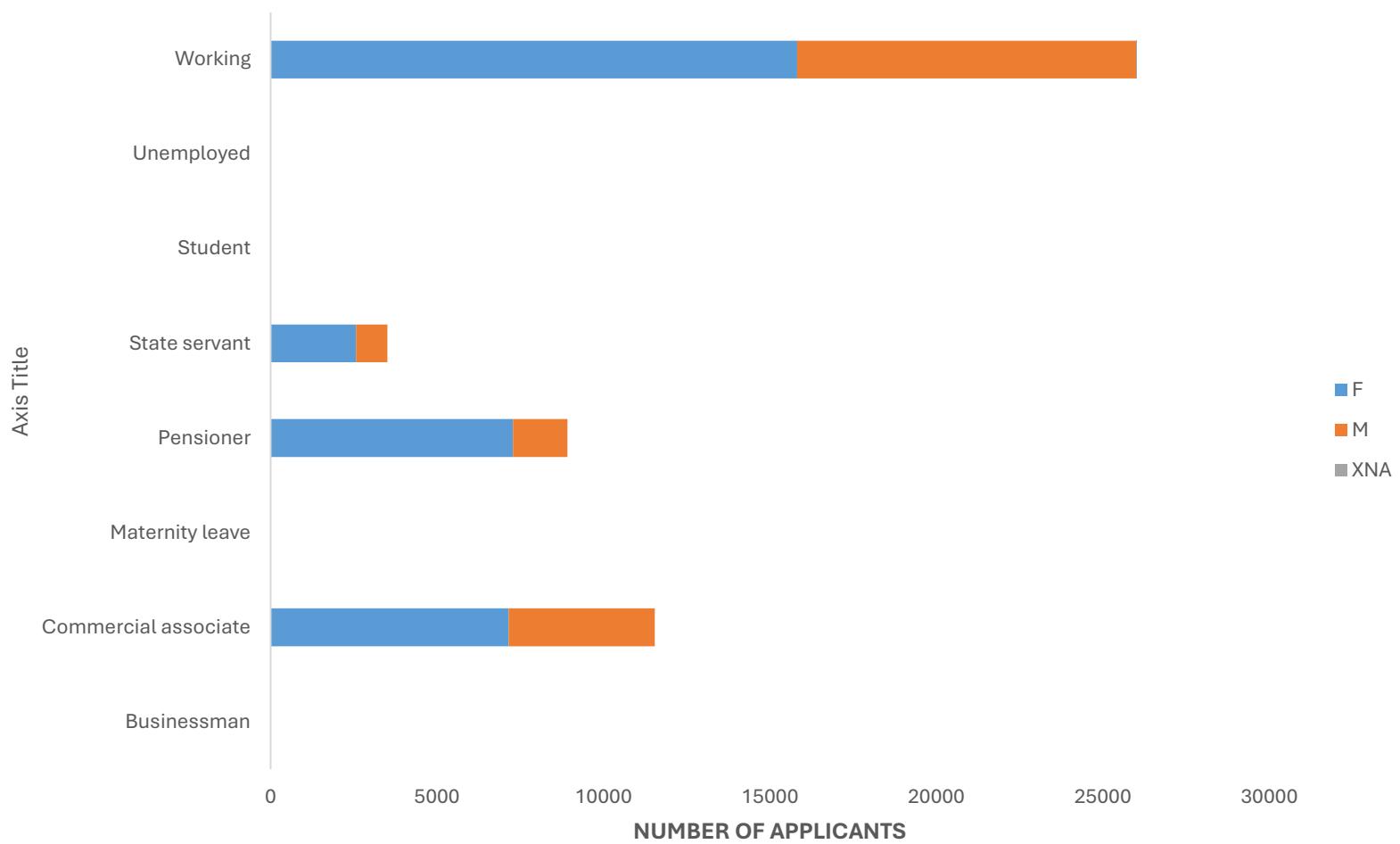
MALE CLIENT'S AGE DISTRIBUTION



→ Among age distribution, average male's age who applied for the loan is 44 years.

DIFFERENT INCOME TYPE	F	M	XNA
Businessman		2	
Commercial associate	7154	4389	
Maternity leave		1	
Pensioner	7281	1639	
State servant	2565	947	
Student	1	4	
Unemployed	5	1	
Working	15817	10191	2

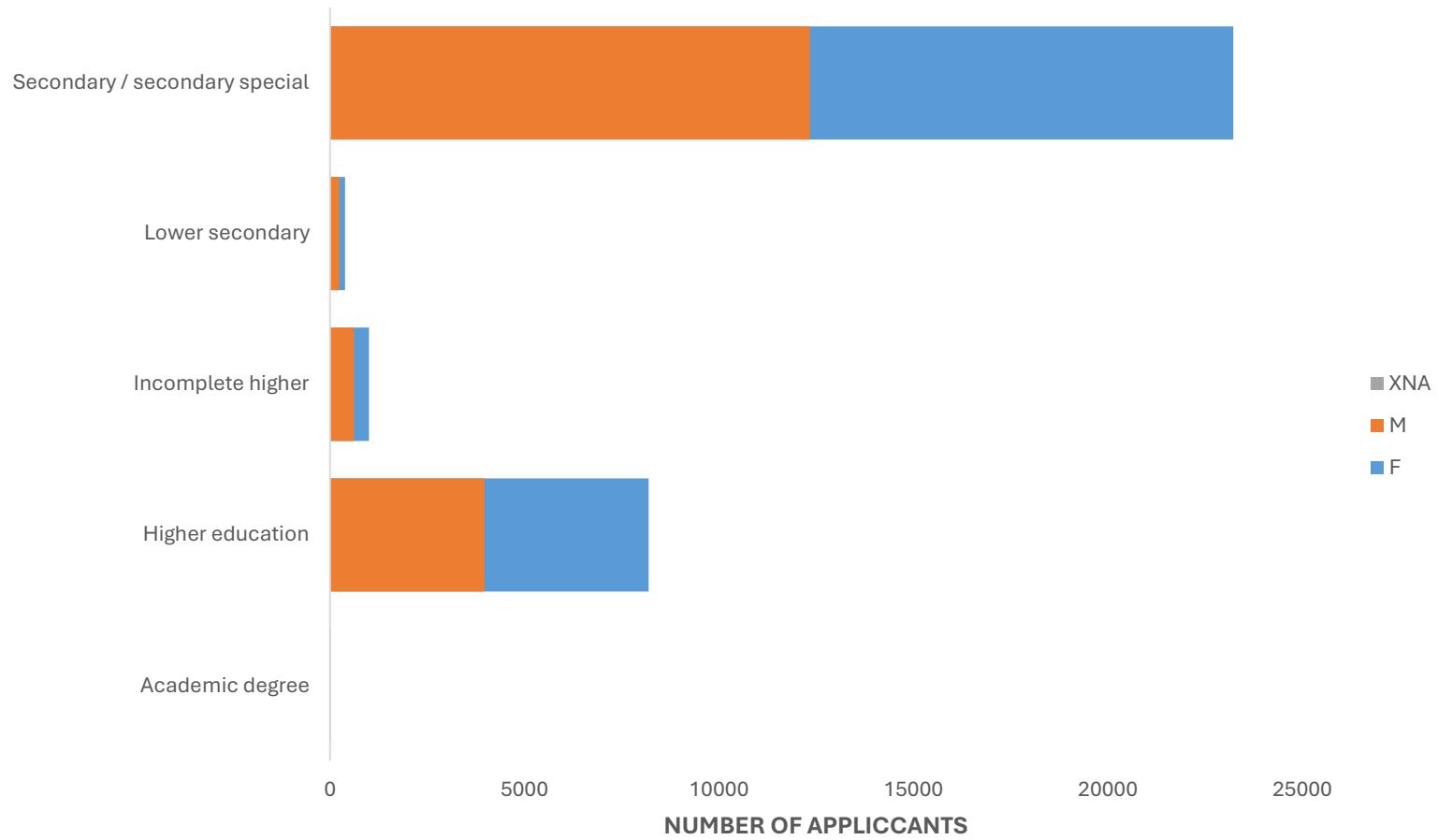
### GENDER WISE INCOME TYPE DISTRIBUTION



- ➔ Most applicant who applied for the loan have income type “Working”
- ➔ Among applicant having income type as “Working” most of them are male.

DIFFERENT EDUCATION TYPE	F	M	XNA
Academic degree	15	5	
Higher education	8194	3973	
Incomplete higher	1001	619	
Lower secondary	383	237	
Secondary / secondary special	23230	12340	2

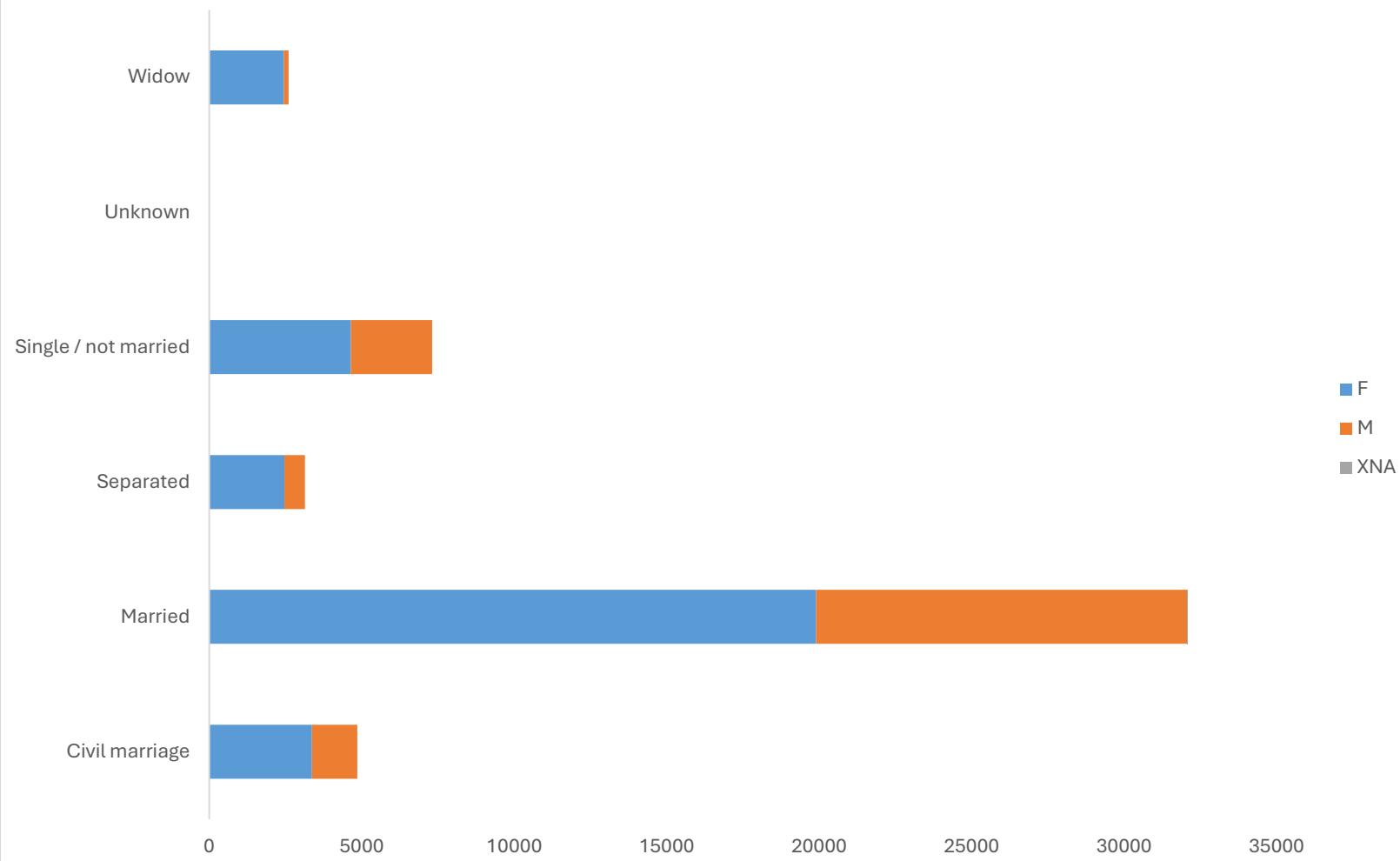
### Distribution of education type according to different gender



- ➔ Most applicant who applied for the loan have secondary education.
- ➔ Among applicant who have secondary education most of them are female.

FAMILY STATUS	F	M	XNA
Civil marriage	3370	1489	
Married	19900	12192	2
Separated	2464	678	
Single / not married	4643	2663	
Unknown		1	
Widow	2446	151	

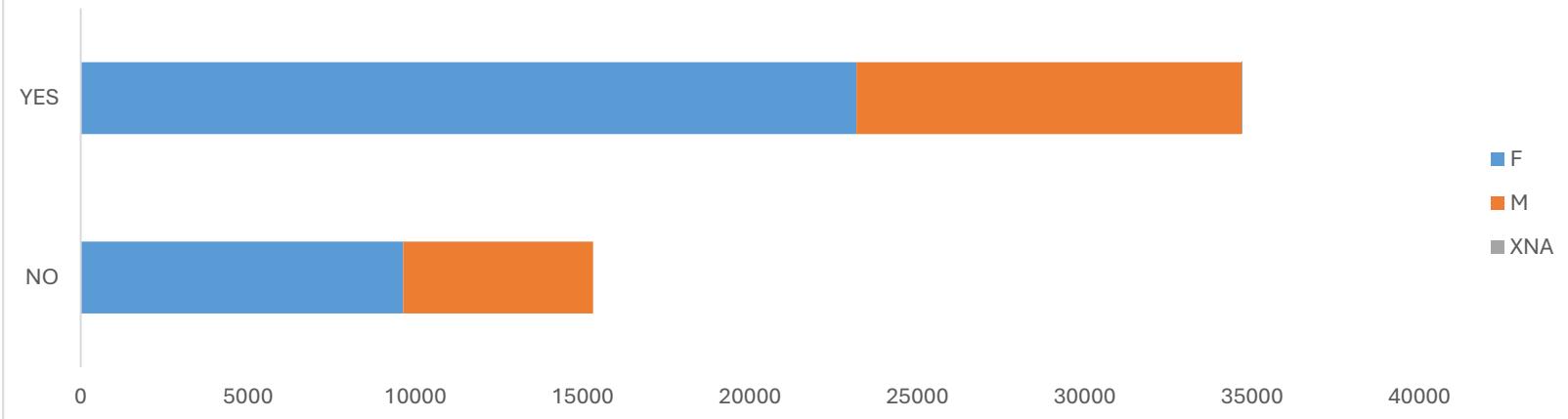
### FAMILY STATUS OF DIFFERENT GENDER



- ➔ Most applicant who applied for the loan are married
- ➔ Among applicants those who have family status as married most of them are females.

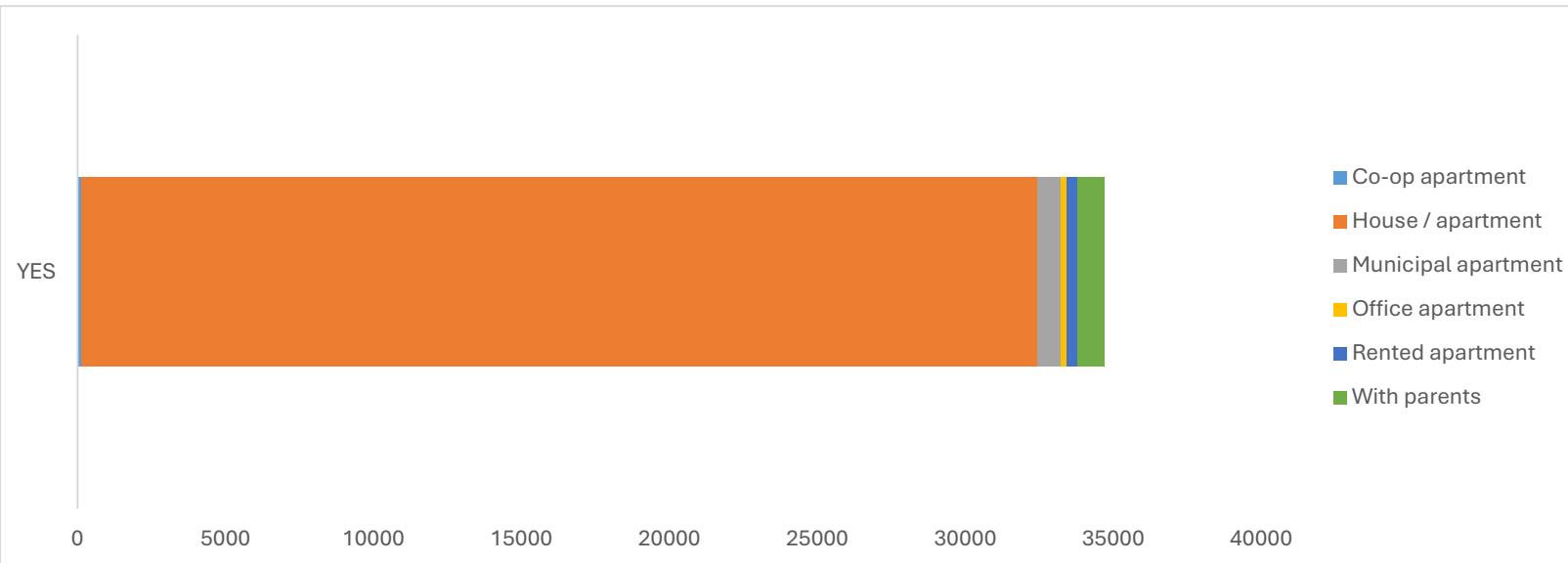
PROPERTY OWNER	F	M	XNA
NO	9640	5668	
YES	23183	11506	2

### PROPERTY OWNER ACCORDING TO DIFFERENT GENDER



- ➔ Applicant who applied for the loan most of them are property owner.
- ➔ Among property owner most of them females.

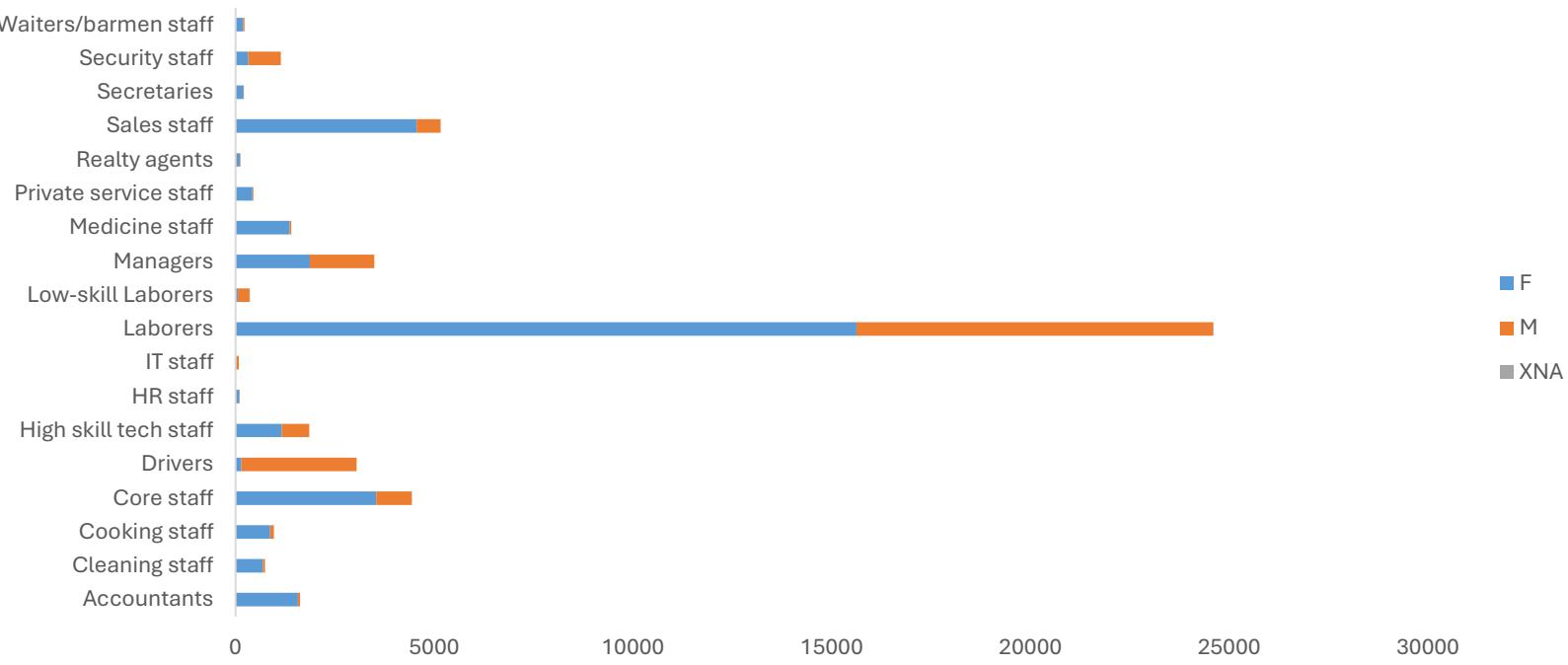
Property Owner	Co-op apartment	House / apartment	Municipal apartment	Office apartment	Rented apartment	With parents
YES	122	32325	774	219	347	904



- ➔ Most applicant who applied for the loan own house/apartment.

OCCUPATION TYPE	F	M	XNA
Accountants	1568	53	
Cleaning staff	690	49	
Cooking staff	871	92	
Core staff	3542	892	
Drivers	144	2900	
High skill tech staff	1165	687	
HR staff	97	4	
IT staff	24	56	
Laborers	15623	8982	1
Low-skill Laborers	56	300	1
Managers	1867	1622	
Medicine staff	1365	38	
Private service staff	427	20	
Realty agents	109	14	
Sales staff	4558	602	
Secretaries	201	11	
Security staff	320	820	
Waiters/barmen staff	196	32	

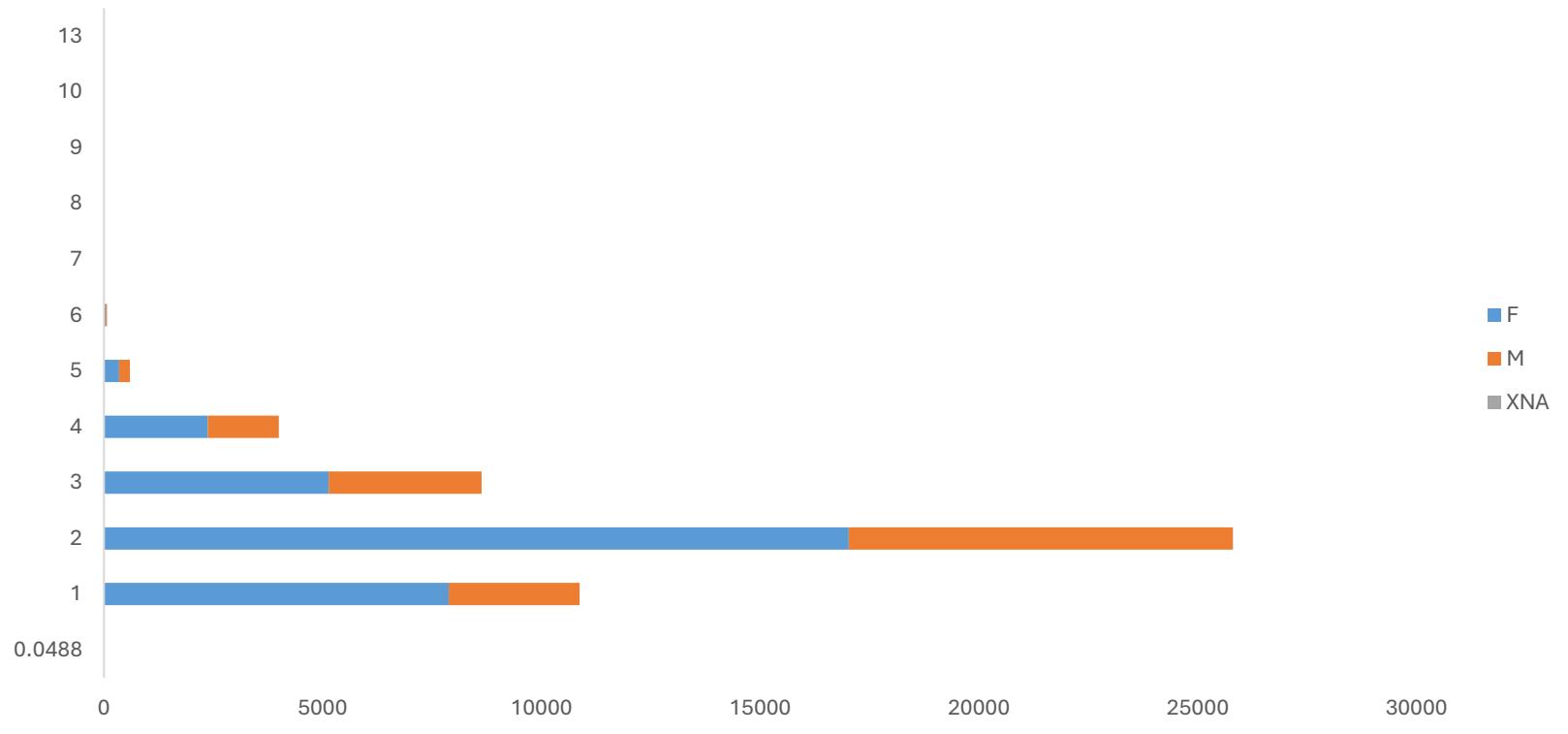
### Occupation type according various gender type



- ➔ Most number of applicants who applied for the loan are labors.
- ➔ Among labors most of them are females.

FAMILY SIZE	F	M	XNA
0.0488		1	
1	7883	2990	
2	17029	8777	1
3	5138	3497	
4	2376	1623	1
5	346	246	
6	41	27	
7	7	5	
8	1	5	
9	1	1	
10	1	1	
13		1	

Family size of various gender type



- Most applicant have family size of 2
- Among applicant with family size of 2 most of them are females.

## BIVARIATE ANALYSIS

GENDER	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME
F	7%	93%
M	10%	90%
XNA	0%	100%

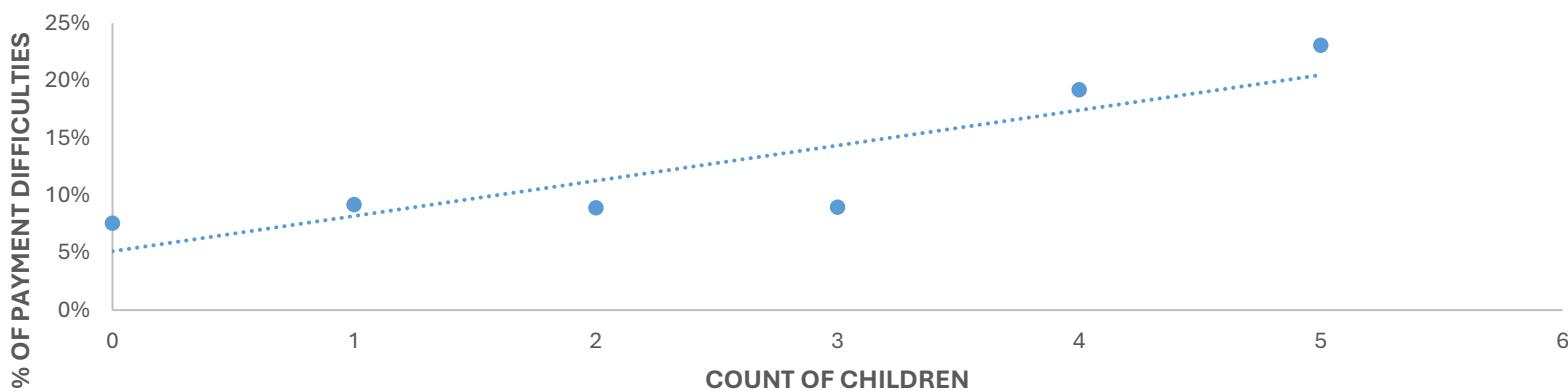
→ Most number of applicants who had faced difficulties in payment are males with 10%.

LOAN TYPE	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	PAYMENT DIFFICULTIES	PAYMEMT MET ON TIME
Cash loans	3792	41484	8%	92%
Revolving loans	234	4489	5%	95%

→ Most number of payment difficulties were related to loan type cash loans with which stood to 8%.

COUNT OF CHILDREN	NO. OF PAYEMENT DIFFICULTIES	NO. OF PAYMENT MET ON TIME	% OF PAYEMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
0	2644	32272	8%	92%
1	923	9118	9%	91%
2	384	3935	9%	91%
3	56	570	9%	91%
4	14	59	19%	81%
5	3	10	23%	77%
MORE				

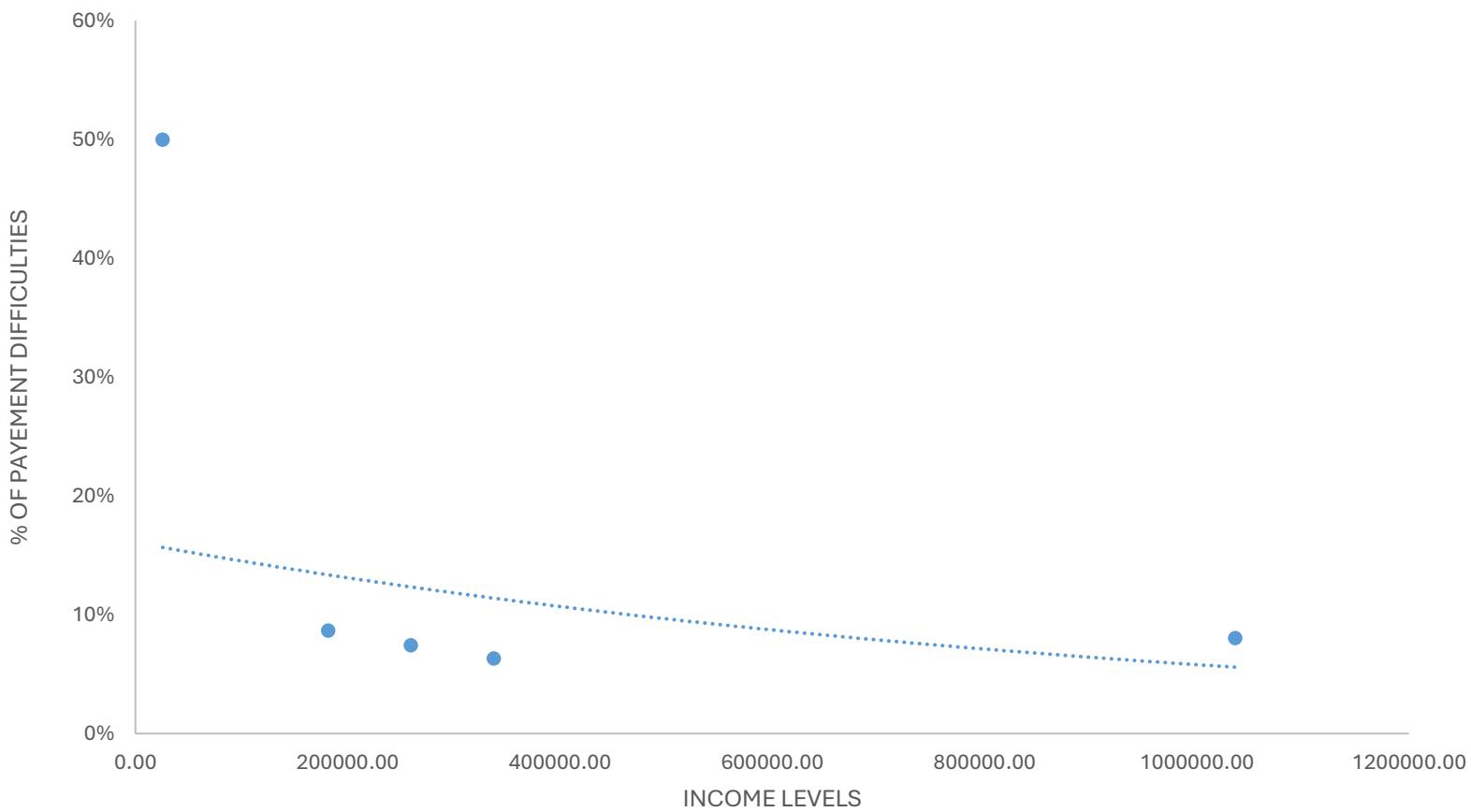
COUNT OF CHILDREN VS PAYMENT DIFFICULTIES



- With increase in count of children payment of difficulties increases.  
 → Applicant having children more than 3 tend to have more payment difficulties than those who have less than 3 children.

INCOME GROUP		NO. OF PAYEMENT DIFFICULTIES	NO. OF PAYMENT MET ON TIME	% OF PAYEMENT DIFFICULTIES	% OF PAYMENT MET ON TIME
0-	25650.00	1	1	50%	50%
25650-	1036323.00	4023	45937	8%	92%
1036323-	181616.00	2085	21944	9%	91%
181616-	259599.00	644	8020	7%	93%
259599-	337582.00	257	3798	6%	94%

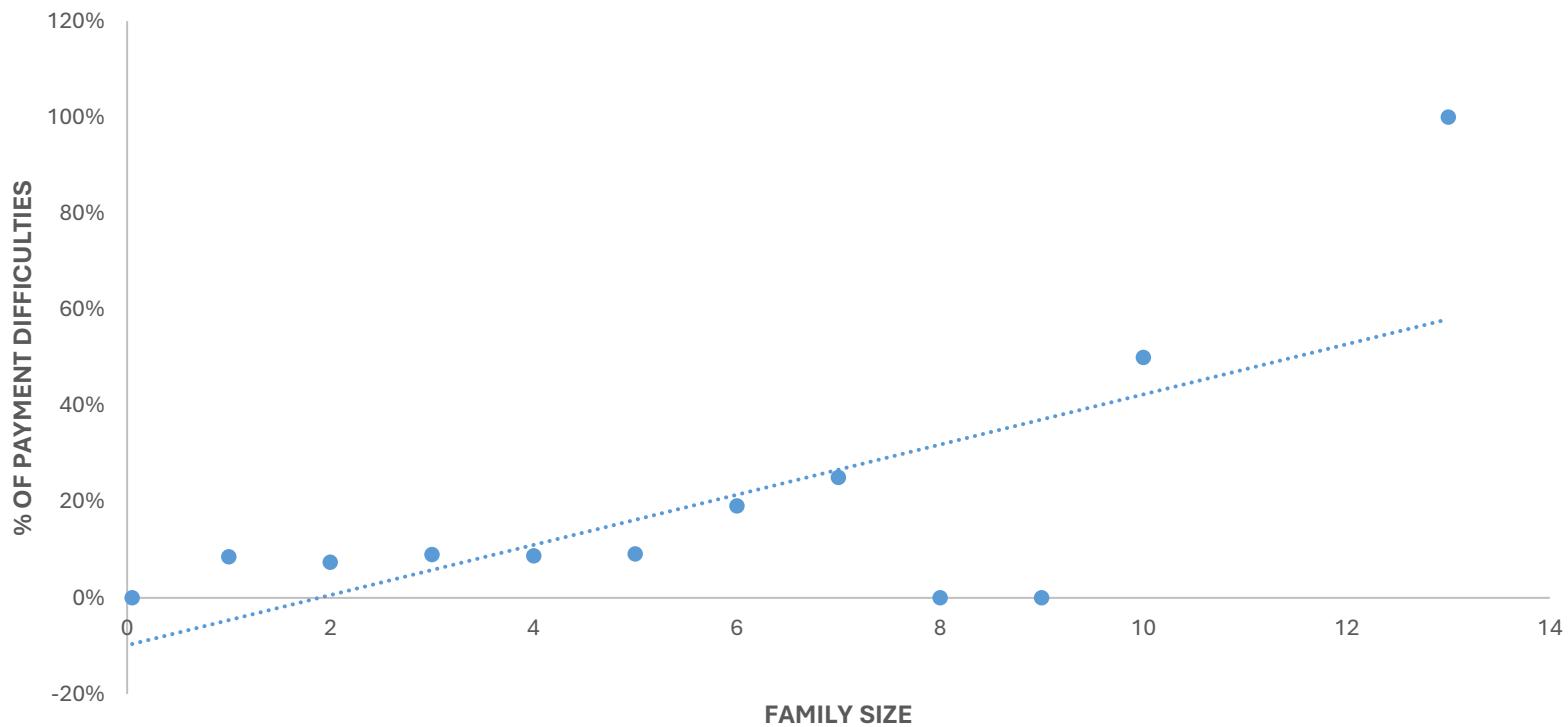
### INCOME LEVELS VS PAYMENT DIFFICULTIES



- ➔ Applicants with low income tend have more payment difficulties than those who have higher income.
- ➔ According to trendline, as the income increase, % of payment difficulties decreases.

FAMILY SIZE	PAYMENT DIFFICULTIES	PAYMENT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMENT MET ON TIME
0.0488		1	0%	100%
1	922	9951	8%	92%
2	1906	23901	7%	93%
3	777	7858	9%	91%
4	349	3651	9%	91%
5	54	538	9%	91%
6	13	55	19%	81%
7	3	9	25%	75%
8		6	0%	100%
9		2	0%	100%
10	1	1	50%	50%
13	1		100%	0%

Family size vs % of Payment difficulties



- ➔ According to the trendline, there is direct relation between family size and payment difficulties.
  - ➔ As the family size increases, % of payment difficulties also increases.
- ➔ Applicants with having family size greater than 4 tend to have more payment difficulties than those who have family size less than 3.

INCOME TYPE	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Businessman		2	0%	100%
Commercial associate	864	10679	7%	93%
Maternity leave		1	0%	100%
Pensioner	501	8419	6%	94%
State servant	198	3314	6%	94%
Student		5	0%	100%
Unemployed	2	4	33%	67%
Working	2461	23549	9%	91%

→ Applicants who are unemployed have the highest % of payment difficulties followed by Applicants who income type as working.

EDUCATION TYPE	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Academic degree		20	0%	100%
Higher education	606	11561	5%	95%
Incomplete higher	138	1482	9%	91%
Lower secondary	73	547	12%	88%
Secondary / secondary special	3209	32363	9%	91%

→ Applicants whose education is lower secondary have highest % of payment difficulties.

→ Applicants who are least educated tend to have more payment difficulties than those who are more educated.

STATUS	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Civil marriage	482	4377	10%	90%
Married	2395	29699	7%	93%
Separated	272	2870	9%	91%
Single / not married	729	6577	10%	90%
Unknown		1	0%	100%
Widow	148	2449	6%	94%

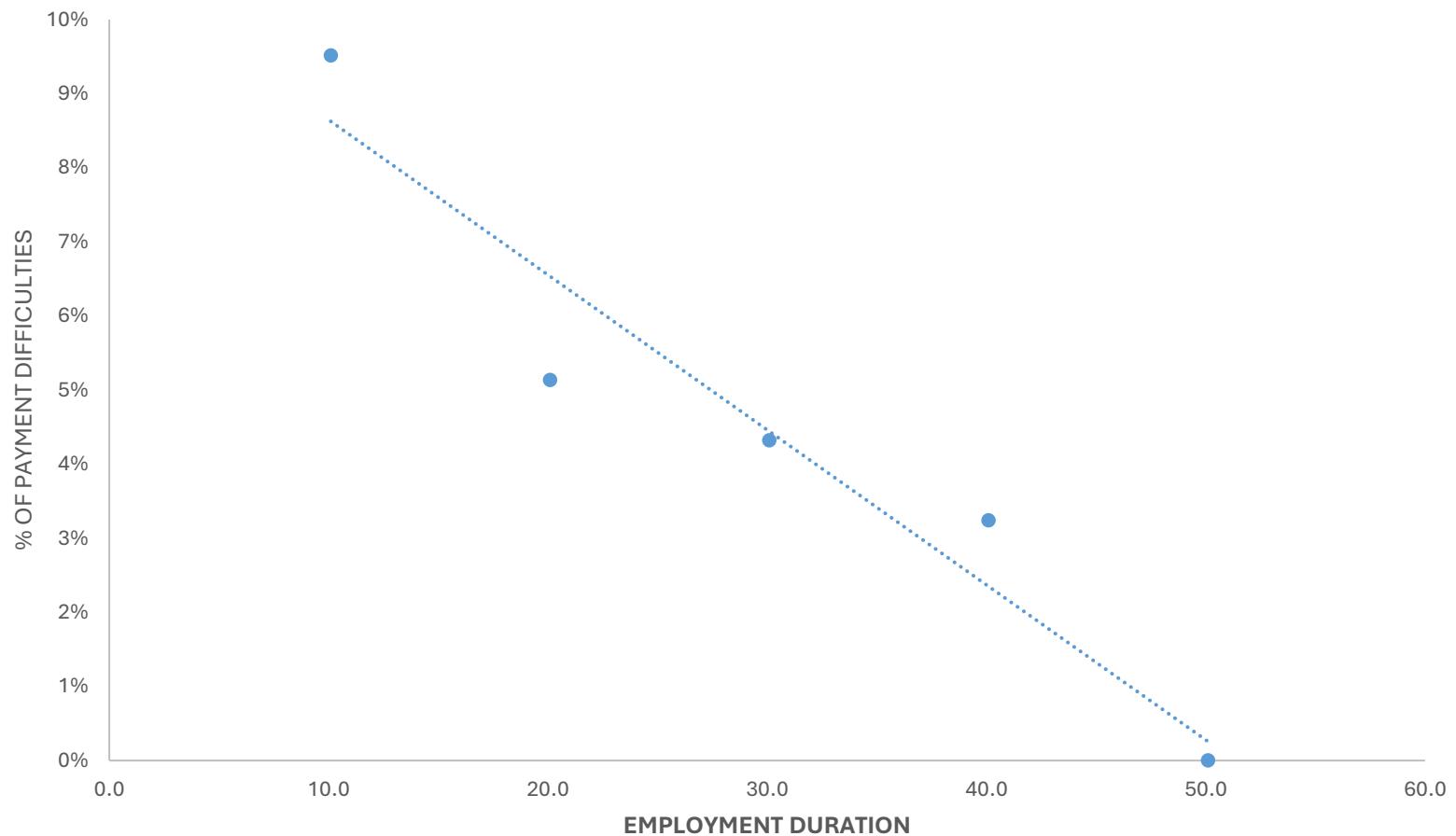
→ Those applicant whose family status is civil marriage, single or not married have the highest % of payment difficulties than others.

HOUSING TYPE	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Co-op apartment	15	176	8%	92%
House / apartment	3473	40895	8%	92%
Municipal apartment	145	1700	8%	92%
Office apartment	29	398	7%	93%
Rented apartment	87	682	11%	89%
With parents	277	2122	12%	88%

→ Applicants who live with parents and in rented apartments have highest % of payment difficulties.

EMPLOYMENT DURATION (IN YEARS) RANGE		PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMENT MET ON TIME
0.1-	10.1	3120	29676	10%	90%
10.1-	20.1	317	5855	5%	95%
20.1-	30.1	69	1528	4%	96%
30.1-	40.1	15	448	3%	97%
40.1-	50.1	0	29	0%	100%
MORE					

% OF PAYMENT DIFFICULTIES



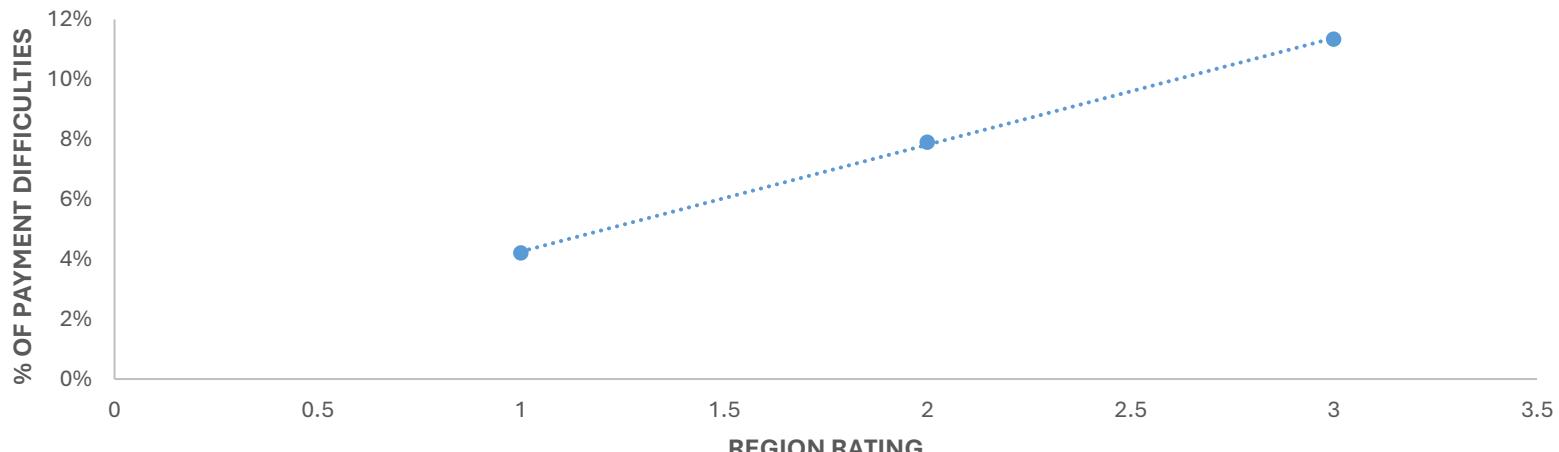
- ➔ Trendline shows inversely proportional relation between employment duration and % of payment difficulties.
  - ➔ As employment duration increases, % of payment difficulties decreases.
- ➔ Applicant with longer employment duration tend to have less % of payment difficulties than those who have shorter period of employment duration.

OCCUPATION	PAYMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Accountants	81	1540	5%	95%
Cleaning staff	68	671	9%	91%
Cooking staff	101	862	10%	90%
Core staff	250	4184	6%	94%
Drivers	338	2706	11%	89%
High skill tech staff	118	1734	6%	94%
HR staff	9	92	9%	91%
IT staff	4	76	5%	95%
Laborers	1946	22660	8%	92%
Low-skill Laborers	61	296	17%	83%
Managers	243	3246	7%	93%
Medicine staff	106	1297	8%	92%
Private service staff	37	410	8%	92%
Realty agents	13	110	11%	89%
Sales staff	492	4668	10%	90%
Secretaries	9	203	4%	96%
Security staff	125	1015	11%	89%
Waiters/barmen staff	25	203	11%	89%

- Applicants who have occupation has labor have higher % of payment difficulties than others.
- Applicants with least % of payment difficulties are those who have occupation has Secretaries.

Region Rating	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
1	220	5006	4%	96%
2	2921	34043	8%	92%
3	885	6924	11%	89%

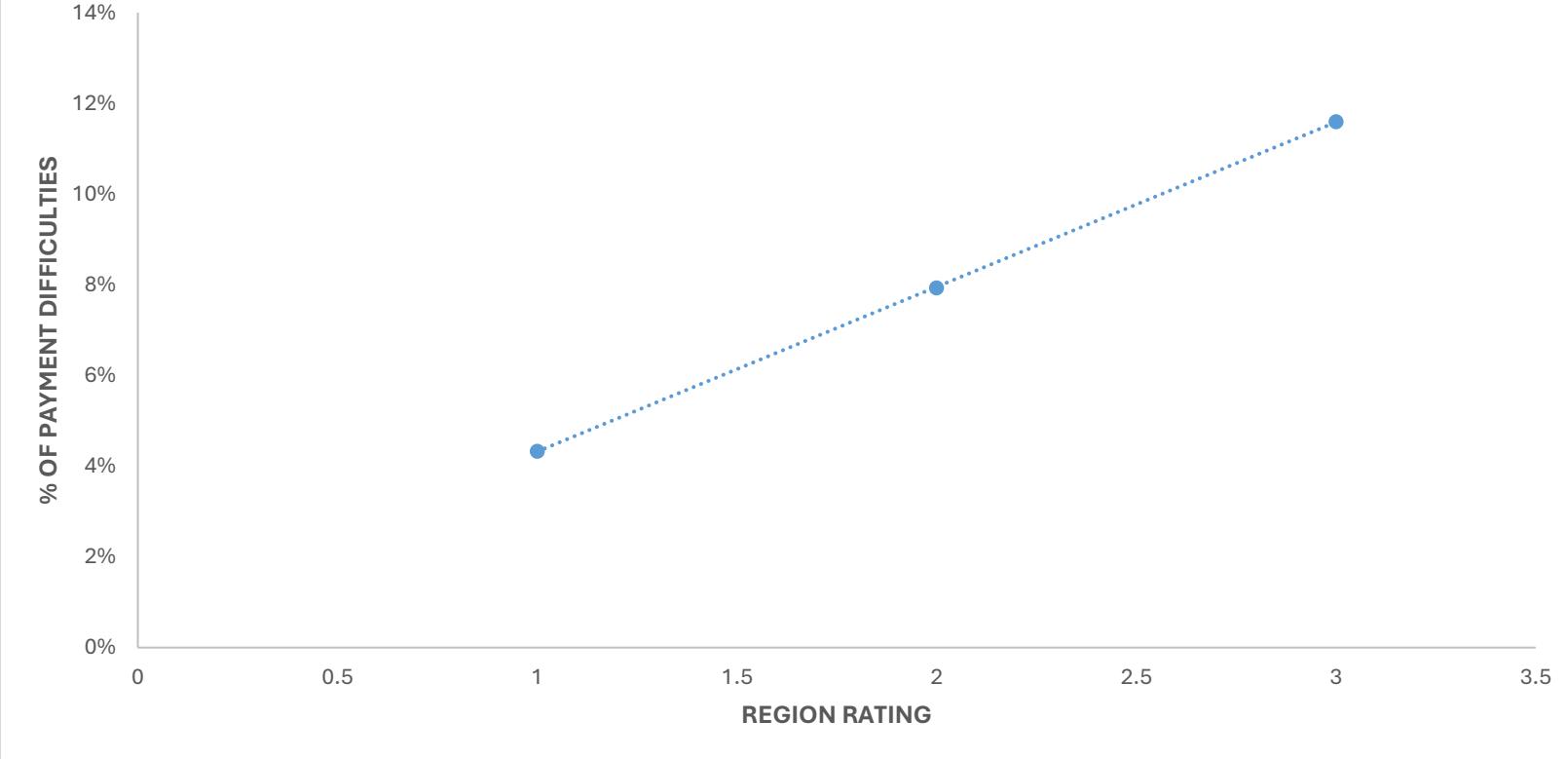
### REGION RATING VS PAYMENT DIFFICULTIES



- According to trendline there is direct relation between region without city and % of payment difficulties.
  - Higher region rating, higher the % of payment difficulties.
  - Applicants belonging to region of region rating 3 have higher % of payment difficulties.

Region Rating with city	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
1	241	5320	4%	96%
2	2962	34379	8%	92%
3	823	6274	12%	88%

Region Rating with City vs Payment difficulties



- According to trendline there is direct relation between region rating with city and % of payment difficulties.
  - Higher region rating with city, higher the % of payment difficulties.
- Applicants belonging to region of region rating with city 3 have higher % of payment difficulties.

Organization Type	Payment Difficulties	Payment Met On Time	% of Payment Difficulties	% of Payment Met On Time
Advertising	7	61	10%	90%
Agriculture	51	341	13%	87%
Bank	27	408	6%	94%
Business Entity Type 1	88	865	9%	91%
Business Entity Type 2	133	1571	8%	92%
Business Entity Type 3	1014	10087	9%	91%
Cleaning	3	37	8%	93%
Construction	108	958	10%	90%
Culture	2	62	3%	97%
Electricity	13	134	9%	91%
Emergency	7	86	8%	92%
Government	124	1592	7%	93%
Hotel	13	169	7%	93%
Housing	42	447	9%	91%
Industry: type 1	19	140	12%	88%
Industry: type 10	1	20	5%	95%
Industry: type 11	28	461	6%	94%
Industry: type 12	3	50	6%	94%
Industry: type 13	4	11	27%	73%
Industry: type 2	10	68	13%	87%
Industry: type 3	51	491	9%	91%
Industry: type 4	15	125	11%	89%
Industry: type 5	7	96	7%	93%
Industry: type 6	1	11	8%	92%
Industry: type 7	19	190	9%	91%
Industry: type 8		8	0%	100%
Industry: type 9	41	496	8%	92%
Insurance	7	82	8%	92%
Kindergarten	66	1024	6%	94%
Legal Services	4	40	9%	91%
Medicine	130	1687	7%	93%
Military	26	432	6%	94%
Mobile	4	52	7%	93%
Other	208	2509	8%	92%
Police	18	348	5%	95%
Postal	27	343	7%	93%
Realtor	7	54	11%	89%
Religion	1	13	7%	93%
Restaurant	32	257	11%	89%
School	78	1372	5%	95%
Security	62	488	11%	89%
Security Ministries	16	315	5%	95%
Self-employed	628	5612	10%	90%
Services	24	260	8%	92%
Telecom	8	98	8%	92%
Trade: type 1	5	61	8%	92%
Trade: type 2	21	286	7%	93%
Trade: type 3	60	490	11%	89%
Trade: type 4	8	8	0%	100%
Trade: type 5	1	7	13%	88%
Trade: type 6	3	105	3%	97%
Trade: type 7	120	1090	10%	90%
Transport: type 1	2	26	7%	93%
Transport: type 2	33	359	8%	92%
Transport: type 3	25	166	13%	87%
Transport: type 4	67	770	8%	92%
University	9	213	4%	96%
XNA	503	8421	6%	94%

- ➔ Applicants who belong industry type 13 have the % of payment difficulties
- ➔ Applicants who belong from industry type 8 have 0% of payment difficulty

Previous loan purpose	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Audio/Video	31	431	7%	93%
Auto Accessories	2	40	5%	95%
Clothing and Accessories	2	79	2%	98%
Computers	30	388	7%	93%
Construction Materials	5	98	5%	95%
Consumer Electronics	26	474	5%	95%
Direct Sales		2	0%	100%
Education		2	0%	100%
Fitness		1	0%	100%
Furniture	12	212	5%	95%
Gardening		11	0%	100%
Homewares	2	27	7%	93%
Jewelry	3	15	17%	83%
Medical Supplies	2	19	10%	90%
Medicine	2	3	40%	60%
Mobile	61	837	7%	93%
Office Appliances	1	8	11%	89%
Other	2	8	20%	80%
Photo / Cinema Equipment	11	107	9%	91%
Sport and Leisure	1	9	10%	90%
Tourism		3	0%	100%
Vehicles		9	0%	100%
XNA	294	2863	9%	91%

- Applicants who previously applied for the loan with purpose of medicine needs have higher % of payment difficulties.
- This observation also indicates the fact that people having emergency needs like medicines, affects applicant's ability to pay-off their current loan.
- Other purpose for which previous loan are applied have no effect on applicant's ability to pay off current loan.

Previous Loan Type	PAYEMENT DIFFICULTIES	PAYMEMT MET ON TIME	% OF PAYMENT DIFFICULTIES	% OF PAYMEMT MET ON TIME
Cash loans	226	2215	9%	91%
Consumer loans	196	2814	7%	93%
Revolving loans	65	617	10%	90%

- Applicants who previously had revolving loan have higher % of payment difficulties.

## TASK – 5 Identify Top Correlations for Different Scenarios

**TOP CORRELATION OF APPLICANTS MAKING PAYMENT ON TIME**

VARIABLES												
AMT_INCOME_TOTAL	1	0.377965752	0.451135167	0.384575912	-0.161680938	0.041464955	-0.205031899	-0.220044862	0.156172404	-0.073475401	-0.073769425	0.036319722
AMT_CREDIT		1	0.770772818	0.986999774	-0.074733443	0.064864901	-0.102556478	-0.111639948	0.136259888	0.028989838	0.051084182	0.005705458
AMT_ANNUITY			1	0.775835204	-0.111292395	0.077867095	-0.129920896	-0.143197363	0.130023853	0.018669848	-0.009911417	0.02683896
AMT_GOODS_PRICE				1	-0.072446456	0.06289989	-0.104841672	-0.113122992	0.143230576	0.031223089	0.048773297	0.001518097
DAYS_EMPLOYED					1	-0.234728771	0.040937165	0.043223355	-0.031902676	0.101710008	0.623474675	-0.245521512
CNT_FAM_MEMBERS						1	0.022207028	0.021214754	0.004603673	-0.022760738	-0.284322672	0.879213563
REGION_RATING_CLIENT							1	0.950468157	-0.294417271	0.002023497	-0.00902485	0.021288992
REGION_RATING_CLIENT_W_CITY								1	-0.289058992	0.003246357	-0.00708431	0.017873365
EXT_SOURCE_2									1	0.068881803	0.080334851	-0.013466448
EXT_SOURCE_3										1	0.179083958	-0.039263601
CLIENT'S_AGE											1	-0.335876269
CNT_CHILDREN												1
	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	DAYS_EMPLOYED	CNT_FAM_MEMBERS	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	EXT_SOURCE_2	EXT_SOURCE_3	CLIENT'S_AGE	CNT_CHILDREN

**TOP CORRELATION OF APPLICANTS HAVING PAYMENT DIFFICULTIES**

VARIABLES															
AMT_INCOME_TOTAL	1	0.0152714	0.018004594	0.013269502	-0.011758681	0.013121678			-0.012846697	-0.01266585	-0.016228127	-0.0265	-0.00903366	0.010110177	
AMT_CREDIT		1	0.749665201	0.982267963	0.018782223	0.06124869			-0.045024534	-0.052954314	0.119188409	0.04584	0.14250603	0.007601905	
AMT_ANNUITY			1	0.74950403	-0.078113894	0.075838463			-0.061578289	-0.079418668	0.113696819	0.01773	0.00875171	0.029172977	
AMT_GOODS_PRICE				1	0.023181572	0.055135807			-0.051296281	-0.056693474	0.133294869	0.04765	0.1410059	-0.001079665	
DAYS_EMPLOYED					1	-0.183362962			-0.009237108	-0.004126201	-0.016153225	0.08522	0.58824282	-0.189773227	
CNT_FAM_MEMBERS						1			0.057279521	0.057987728	0.005660625	-0.0227	-0.1991414	0.892521875	
CNT_CHILDREN							1		0.055515557	0.054802235	-0.015410537	-0.015	-0.2496732	1	
REGION_RATING_CLIENT								1	0.950768899	-0.239051675	0.0215	-0.04502711	0.055515557		
REGION_RATING_CLIENT_W_CITY									1	-0.239176892	0.01782	-0.03808733	0.054802235		
EXT_SOURCE_2										1	0.04734	0.11140724	-0.015410537		
EXT_SOURCE_3											1	0.13958798	-0.014958105		
CLIENT'S_AGE												1	-0.2496732		
CNT_CHILDREN													1		
	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	DAYS_EMPLOYED	CNT_FAM_MEMBERS	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	EXT_SOURCE_2	EXT_SOURCE_3	CLIENT'S_AGE	CNT_CHILDREN			

## TOP CORRELATIONS UNDER DIFFERENT SCENARIO

TOP CORRELATION OF APPLICANTS HAVING PAYMENT DIFFICULTIES		
VARIABLES 1	VARIABLES 2	Correlation
AMT_CREDIT	AMT_GOODS_PRICE	1
AMT_ANNUITY	AMT_GOODS_PRICE	0.771
AMT_GOODS_PRICE	AMT_CREDIT	0.987
DAYS_EMPLOYED	CLIENT'S AGE	0.623
CNT_FAM_MEMBERS	CNT_CHILDREN	0.879
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.95
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT_W_CITY	0.95
CLIENT'S AGE	DAYS_EMPLOYED	0.623
CNT_CHILDREN	CNT_FAM_MEMBERS	0.879

→ Among all other correlation between variables highest degree of positive correlation is between AMT\_CREDIT and AMT\_GOODS\_PRICE in case of applicant having payment difficulties.

TOP CORRELATION OF APPLICANTS MAKING PAYMENTS ON TIME		
VARIABLES 1	VARIABLES 2	Correlation
AMT_CREDIT	AMT_GOODS_PRICE	0.982
AMT_ANNUITY	AMT_GOODS_PRICE	0.7496
AMT_GOODS_PRICE	AMT_CREDIT	0.9822
DAYS_EMPLOYED	CLIENT'S AGE	0.588
CNT_FAM_MEMBERS	CNT_CHILDREN	0.8925
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.95
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT_W_CITY	0.95
CLIENT'S AGE	DAYS_EMPLOYED	0.588
CNT_CHILDREN	CNT_FAM_MEMBERS	0.892

→ Among all other correlation between variables highest degree of positive correlation is between AMT\_CREDIT and AMT\_GOODS\_PRICE in case of applicant meeting payment on time.