Task 6

Bank Loan Case Study

BY YASH RAJPUT
YASHRAJPUT.SOCIAL@GMAIL.COM
9175034021

Data Analytics Tasks

- Identify Missing Data and Deal with it Appropriately
- Identify Outliers in the Dataset
- Analyze Data Imbalance
- Perform Univariate, Segmented Univariate, and Bivariate Analysis
- Identify Top Correlations for Different Scenarios

Finding Null values

Firstly the percentage of null values needs to be analyzed and those columns that have more than 50% of the null data have to be dropped And those columns with less than 50% of the null data have to be replaced with mean or median or the highest occurring categorical variables

ALL THE COLUMN NAME WHICH ARE LISTED HERE NEED TO BE DROPPED DOWN AS THEY HAVE NULL VALUES GREATER THAN OR EQUAL TO 50%

Column name	Total number of null values	Percentage of null value in that column	ROUND PER
OWN_CAR_AGE	202930.00	65.99	66.00
EXT_SOURCE_1	173379.00	56.38	56.00
APARTMENTS_AVG	156061.00	50.75	51.00
BASEMENTAREA_AVG	179943.00	58.52	59.00
YEARS_BUILD_AVG	204488.00	66.50	66.00
COMMON_AREA_AVG	214865.00	69.87	70.00
ELEVATORS_AVG	163891.00	53.30	53.00
ENTRANCES_AVG	154828.00	49.70	50.00
FLOORSMAX_AVG	153021.00	49.76	50.00
FLOORSMIN_AVG	208642.00	67.85	68.00
LANDAREA_AVG	182590.00	59.38	59.00
LIVINGAPARTMENTS_AVG	210199.00	68.35	68.00
LIVINGAREA_AVG	154350.00	50.19	50.00
NONLIVINGAPARTMENTS_AVG	213514.00	69.43	69.00
NONLIVINGAREA_AVG	169682.00	55.18	55.00
APARTMENTS_MODE	156061.00	50.75	51.00
BASEMENTAREA_MODE	179943.00	58.52	59.00
YEARS_BUILD_MODE	204488.00	66.50	66.00
COMMON_AREA_MODE	214865.00	69.87	70.00
ELEVATORS_MODE	163891.00	53.30	53.00
ENTRANCES_MODE	154828.00	50.35	50.00
FLOORSMAX_MODE	153020.00	49.76	50.00
FLOORSMIN_MODE	208642.00	67.85	68.00
LANDAREA_MODE	182590.00	59.38	59.00
LIVINGAPARTMENTS_MODE	210199.00	68.35	68.00
LIVINGAREA_MODE	154350.00	50.19	50.00
NONLIVINGAPARTMENTS_MODE	213514.00	69.43	69.00
NONLIVINGAREA_MODE	169682.00	55.18	55.00
APARTMENTS_MEDIAN	156061.00	50.75	51.00
BASEMENTAREA_MEDIAN	179943.00	58.52	59.00
YEARS_BUILD_MEDIAN	204488.00	66.50	66.00
COMMON_AREA_MEDIAN	214865.00	69.87	70.00
ELEVATORS_MEDIAN	163891.00	53.30	53.00
ENTRANCES_MEDIAN	154828.00	50.35	50.00
FLOORSMAX_MEDIAN	153020.00	49.76	50.00
FLOORSMIN_MEDIAN	208642.00	67.85	68.00
LANDAREA_MEDIAN	182590.00	59.38	59.00
LIVINGAPARTMENTS_MEDIAN	210199.00	68.35	68.00
LIVINGAREA_MEDIAN	154350.00	50.19	50.00
NONLIVINGAPARTMENTS_MEDIAN	213514.00	69.43	69.00
NONLIVINGAREA_MEDIAN	169682.00	55.18	55.00
FONDKAPREMONT_MODE	210295.00	68.39	68.00
HOUSETYPE_MODE	154297.00	50.18	50.00
WALLSMATERIAL_MODE	156341.00	50.84	51.00

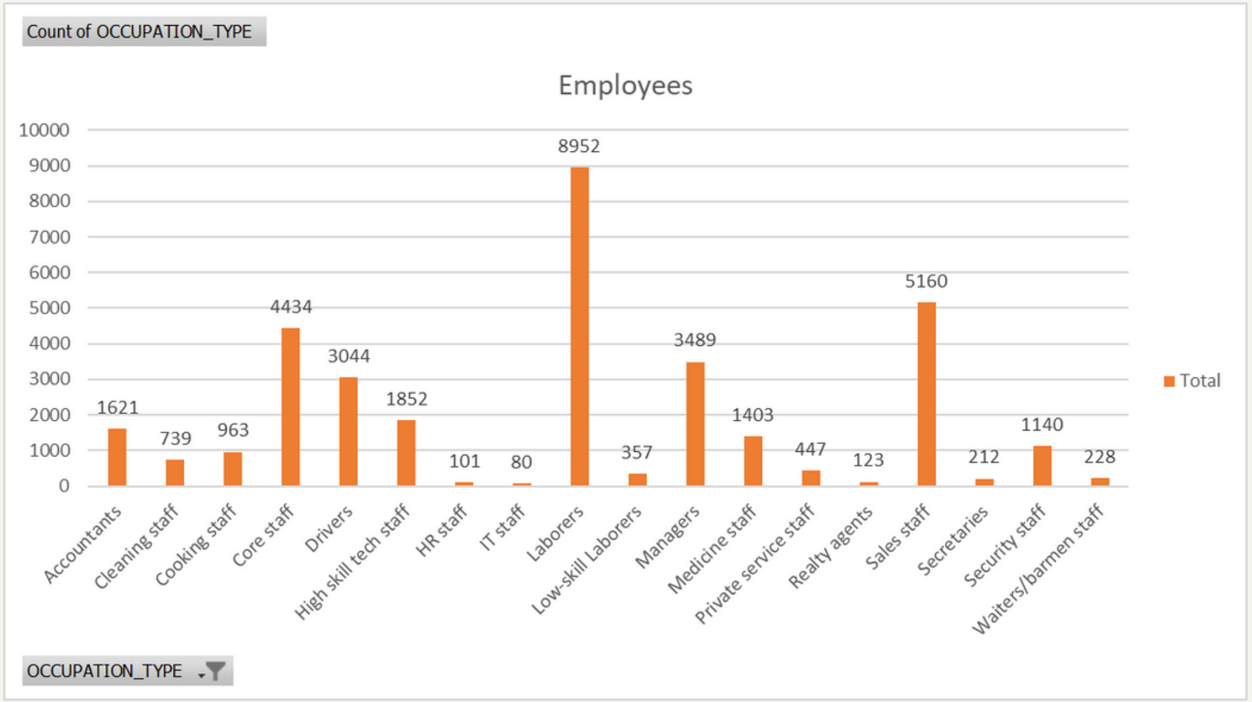
Finding Irrelevant Columns

Over here we need to understand the requirements of our analysis and need to drop columns which are irrelevant so that we don't have to deal with unnecessary data

ALL THE COLUMN NAME WHICH ARE LISTED HERE NEED TO BE DROPPED DOWN AS THEY ARE IRRELEVANT TO OUR ANALYSIS

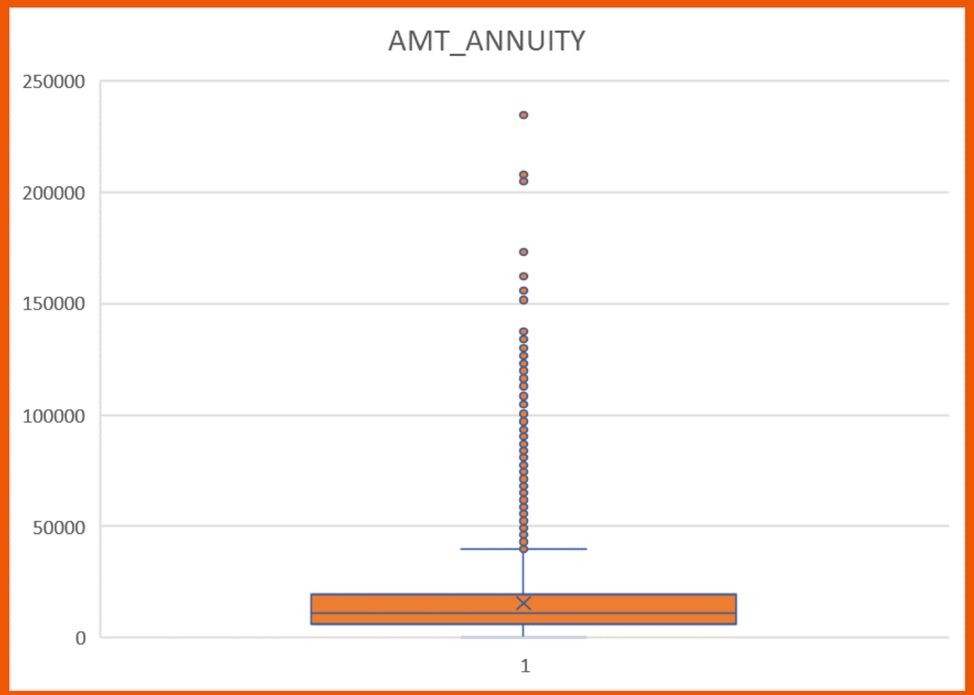
Column Name	Total number of null values	Percentage of null value in that columr	ROUND PER
FLAG_MOBIL	1	0.000325192	0
FLAG_EMPLOY_PHONE	55387	18.01138821	18
FLAG_WORK_PHONE	0	0	0
FLAG_CONT_MOBILE	0	0	0
FLAG_PHONE	0	0	0
FLAG_EMAIL	0	0	0
CNT_FAMILY_MEMBERS	2	0.000650383	0
REGION_RATING_CLENT	0	0	0
REGION_RATING_CLENT_W_CITY	0	0	0
EXT_SOURCE_3	60965	19.82530706	20
YEAR_BEGINEXPLUATATION_AVG	150008	48.78134441	49
YEAR_BEGINEXPLUATATION_MODE	150007	48.78101922	49
YEAR_BEGINEXPLUATATION_MEDIAN	150007	48.78101922	49
TOTAL_AREA_MODE	148431	48.26851722	48
EMERGENCYSTATE_MODE	145755	47.39830445	47
DAYS_LAST_PHONE_CHANGE	1	0.000325192	0
FLAG_DOC	2	0	0
FLAG_DOC	3	0	0
FLAG_DOC	4	0	0
FLAG_DOC	5	0	0
FLAG_DOC	6	0	0
FLAG_DOC	7	0	0
FLAG_DOC	8	0	0
FLAG_DOC	9	0	0
FLAG_DOC	10	0	0
FLAG_DOC	11	0	0
FLAG_DOC	12	0	0
FLAG_DOC	13	0	0
FLAG_DOC	14	0	0
FLAG_DOC	15	0	0
FLAG_DOC	16	0	0
FLAG_DOC	17	0	0
FLAG_DOC	18	0	0
FLAG_DOC	19	0	0
FLAG_DOC	20	0	0
FLAG_DOC	21	0	0

Replacing Blanks in Occupation_Type column of the Application Dataset with the highest occurring categorical variable



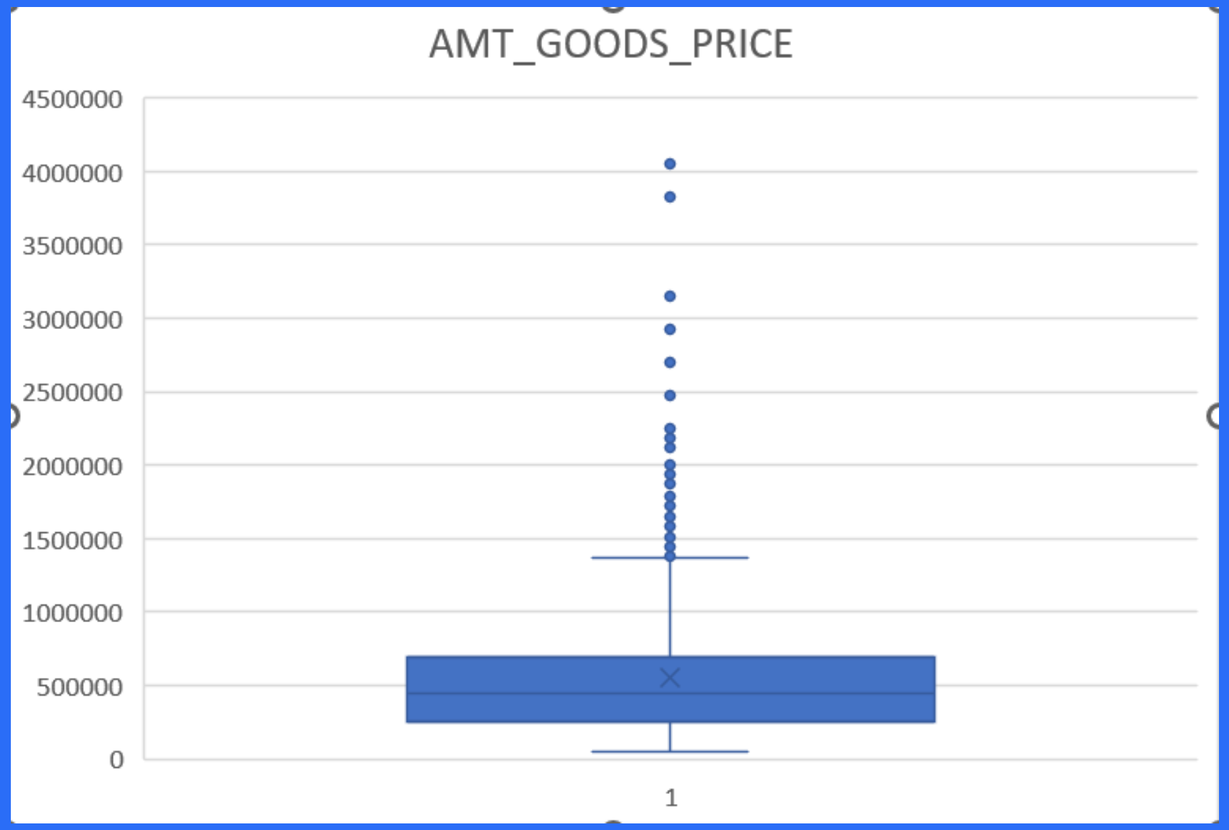
HIGHEST OCCURRING CATEGORICAL VARIABLE IS ‘LABORERS’

Replacing Blanks in AMT_ANNUITY column of the Application Dataset with the median of the AMT_ANNUITY as there exists outliers in the AMT_ANNUITY column



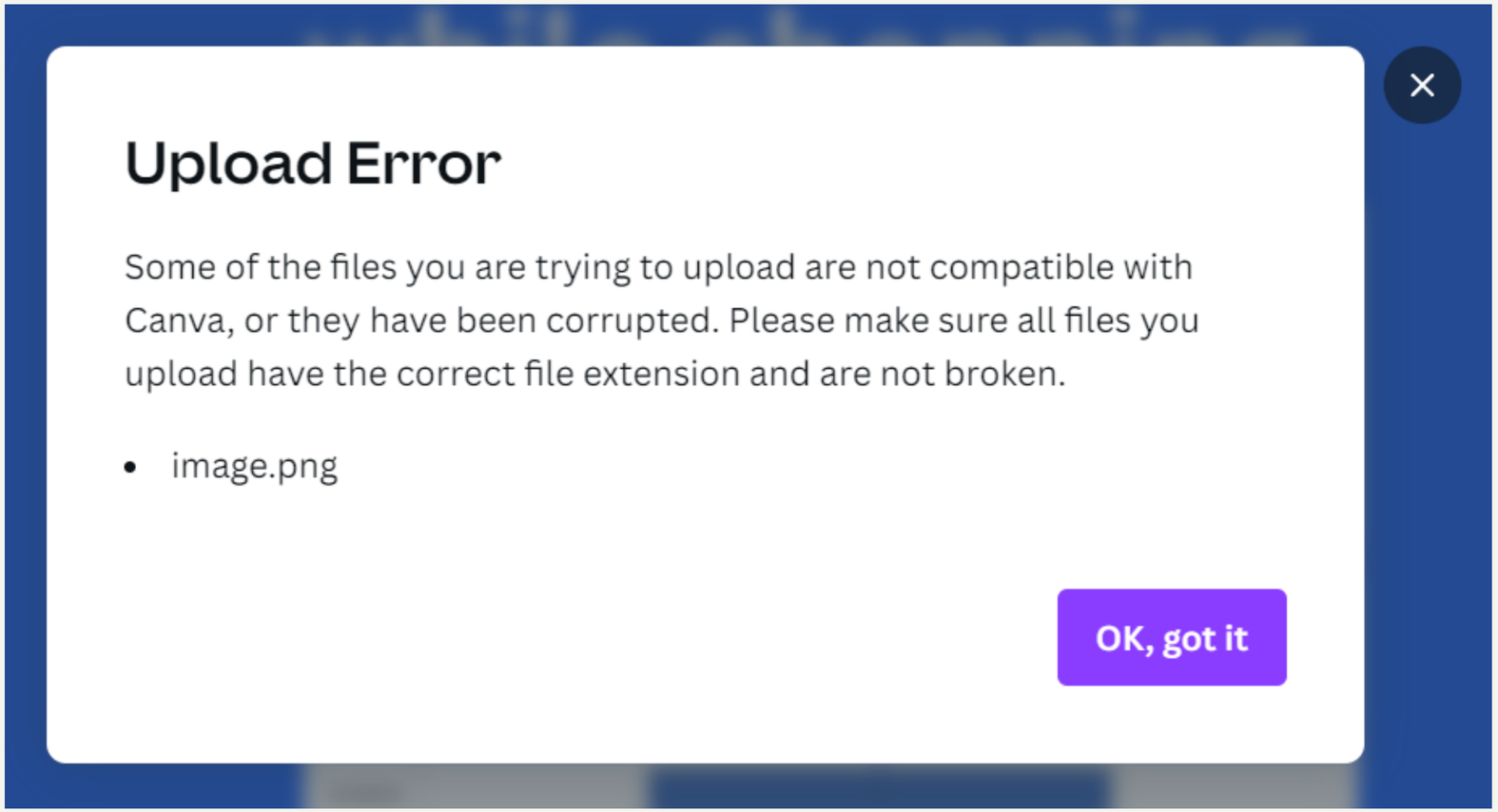
MEDIAN OF AMT_ANNUITY: 24903

Replacing Blanks in AMT_GOODS_PRICE column of the Application Dataset with the median of the AMT_GOODS_PRICE as there exists outliers in the AMT_GOODS_PRICE column



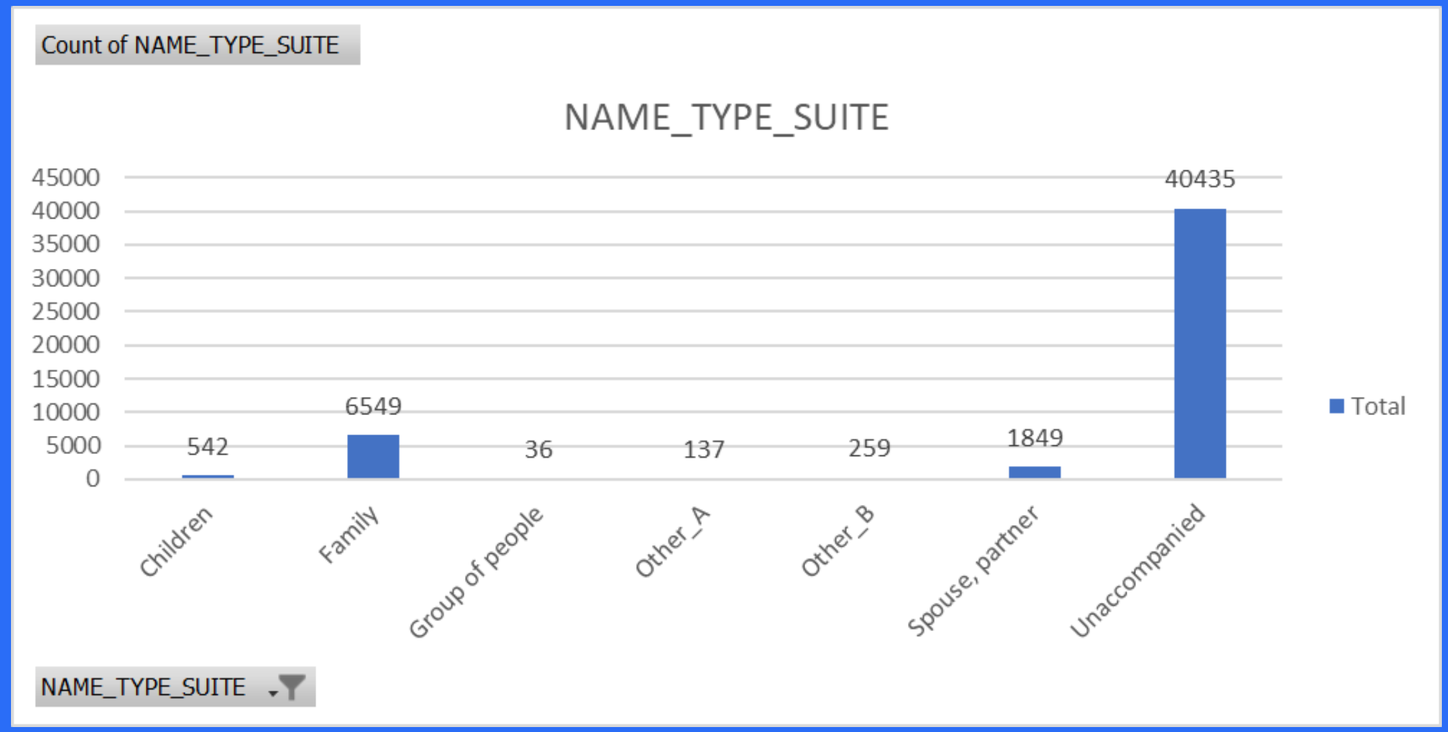
MEDIAN OF AMT_GOODS_PRICE: 450000

NOTE:



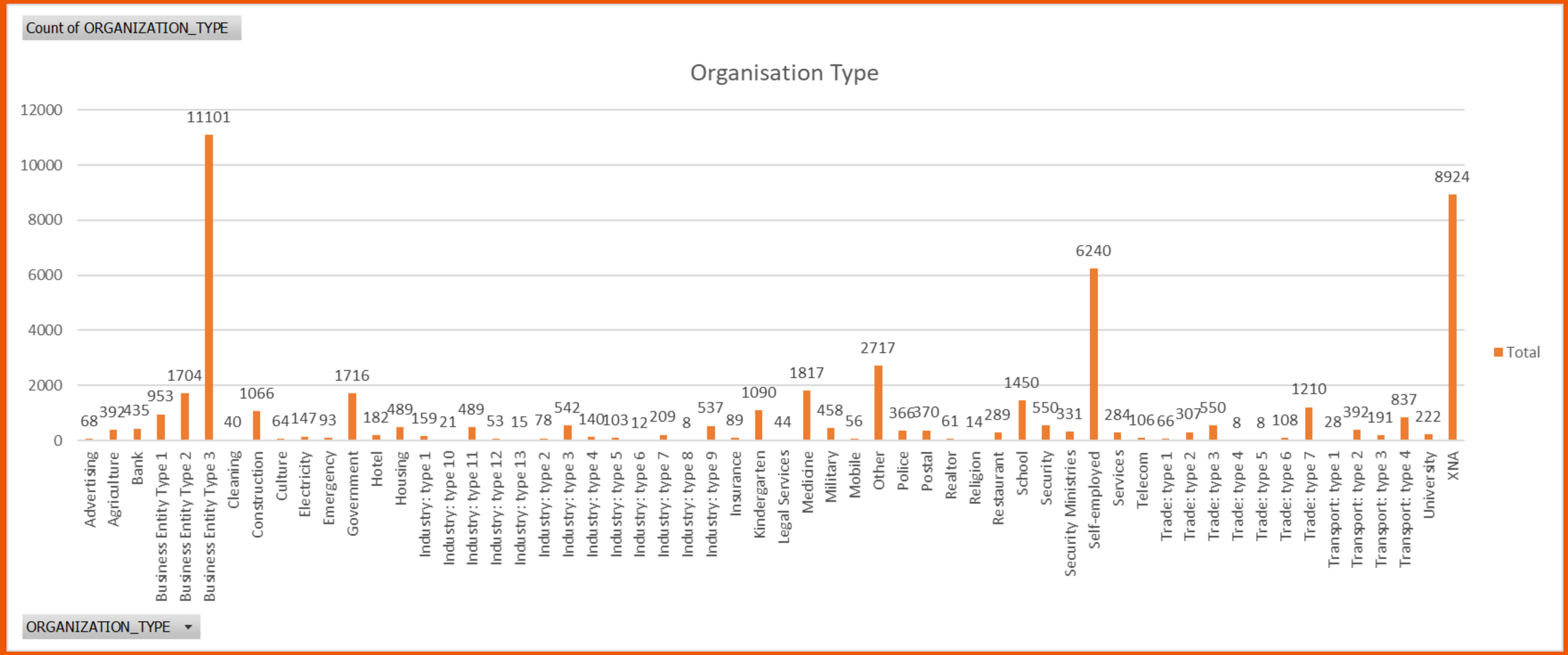
Due to few reasons, I was not able to paste the graph from Excel in canva so I had to take screenshots of the Excel Chart

Replacing Blanks in Name_Type_Suite column of the Application Dataset with the highest occurring categorical variable



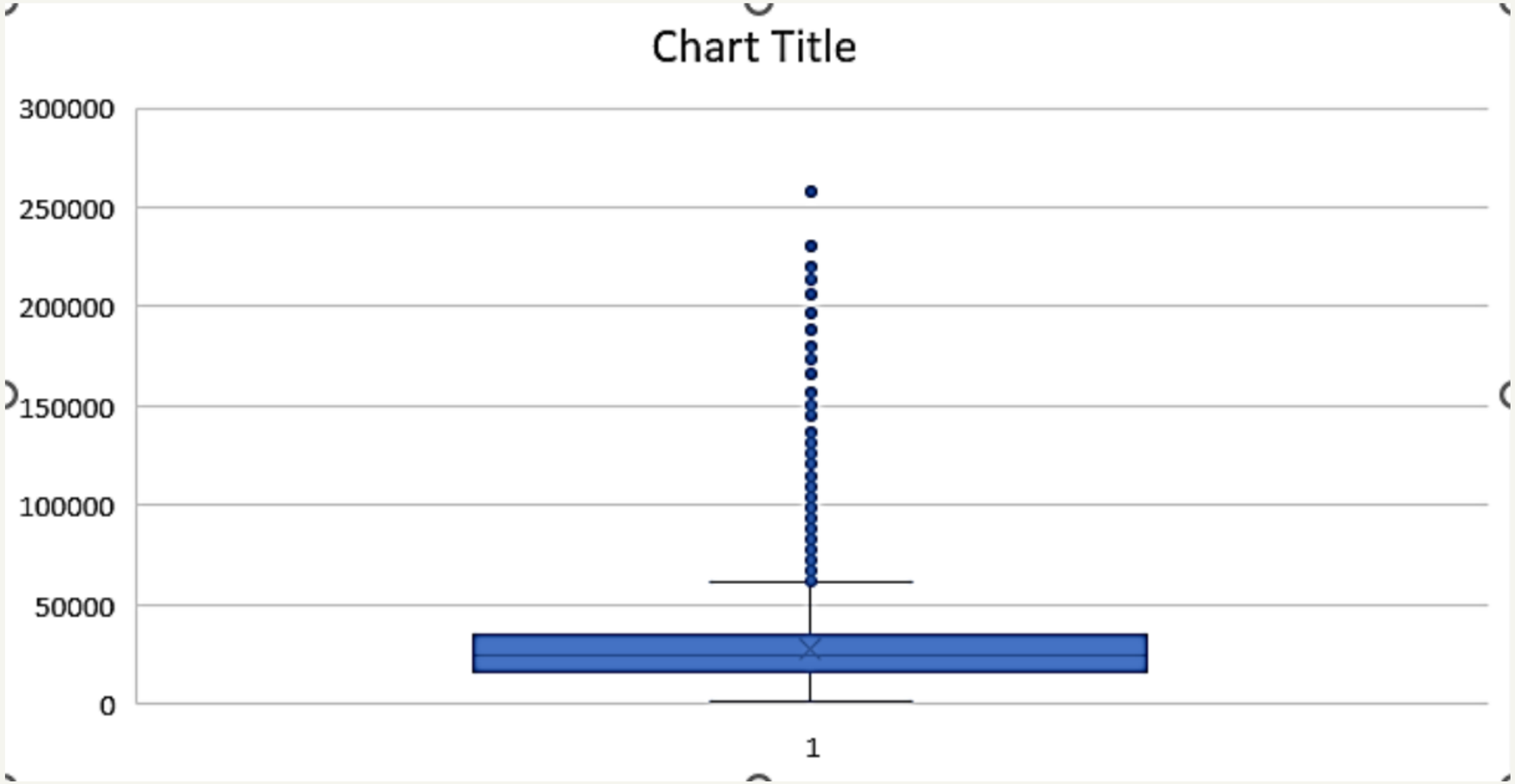
HIGHEST OCCURRING CATEGORICAL VARIABLE IS ‘UNACCOMPANIED’

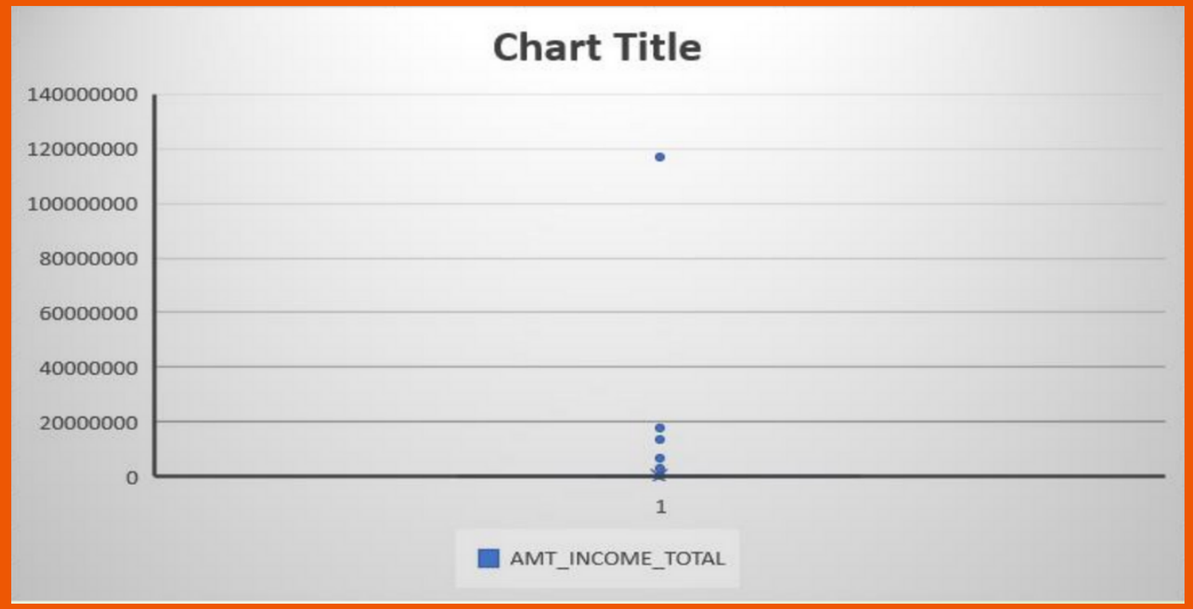
Replacing Blanks in Organization_type column of the Application Dataset with the highest occurring categorical variable



HIGHEST OCCURRING CATEGORICAL VARIABLE IS 'BUSINESS ENTITY TYPE 3

First outlier is in AMT_ANNUITY which is greater than 250000 this outlier is replaced with 24903 median of AMT_ANNUITY

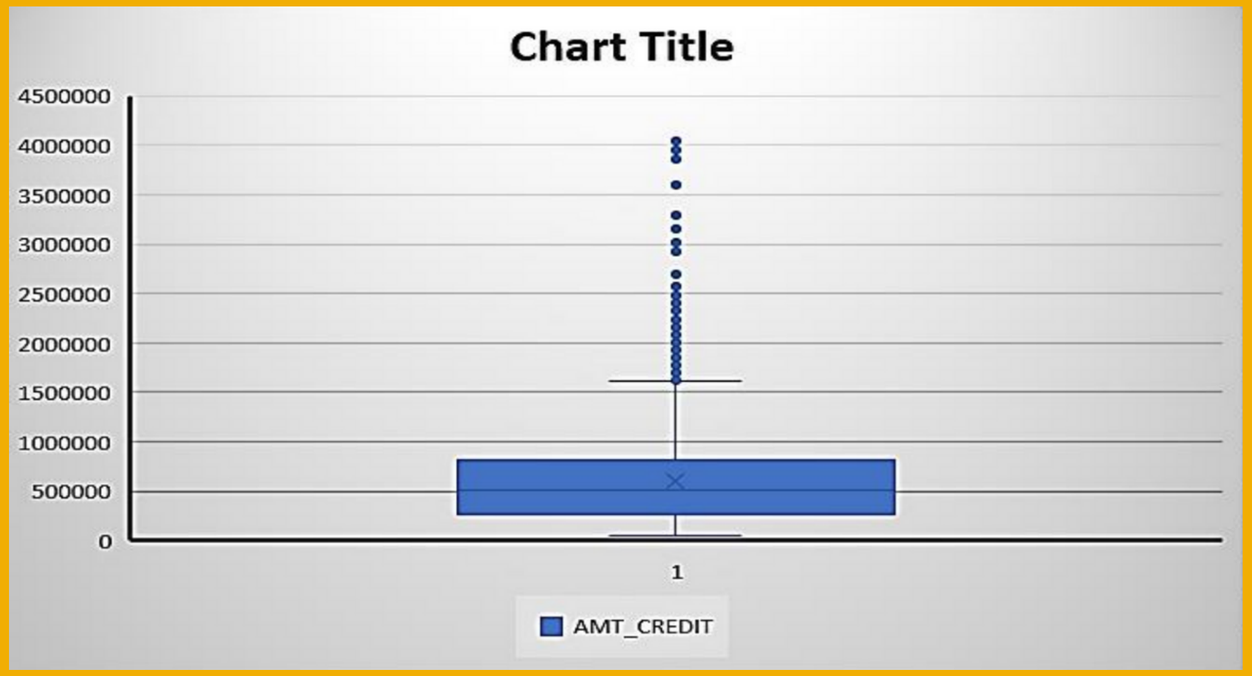




Here we can observe that there is huge difference between the 25%, 50% and 75% quartile and this is due to presence of outliers But since the amount of total income varies from person to person we will not remove the outliers

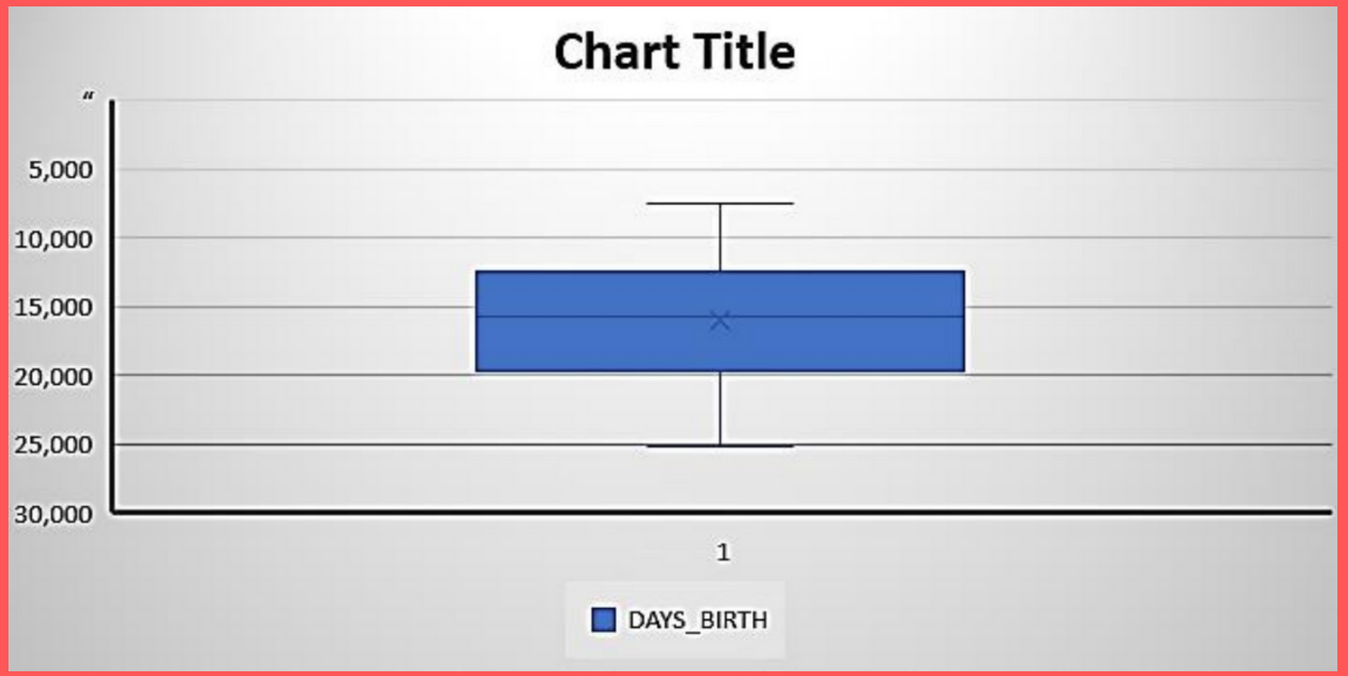
	Quartiles at AMT_INCOME_TOTAL
MIN	25650
25%	112500
50%	147150
75%	202500
MAX	117000000

outliers at extreme points i.e. max



From the chart it is clear that outliers lie in the 98% and near max side of the box plot Also there is a significant difference between the 75% quartile and the max value and this is due the presence of the outliers But since the amount of credit varies from person to person we will not remove the outliers

	AMT_CREDIT
	Quartiles at AMT_CREDIT
MIN	45000
25%	270000
50%	513531
75%	808650
MAX	4050000



As seen from the boxplot it is clear that there are no outliers The data of DAYS_BIRTH is well distributed

	DAYS_OF_BIRTH
	Quartiles at DAYS_BIRTH
MIN	25229
25%	19682
50%	15750
75%	12413
MAX	7489

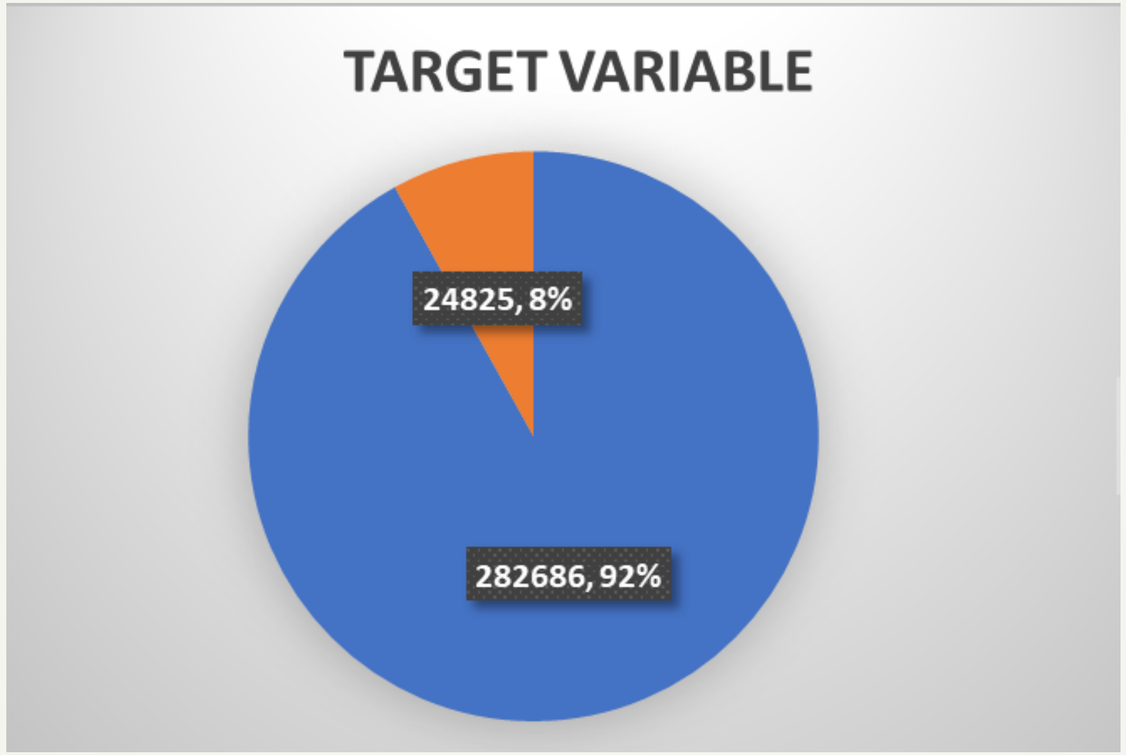
cleaned data link

TARGET VARIABLE

The Target Variable Pie chart shows that almost 92% of the total clients had no problem during payment while 8% of the clients had some or the other problem

Row	Count of Target
1	282686
0	24825
Total	307511

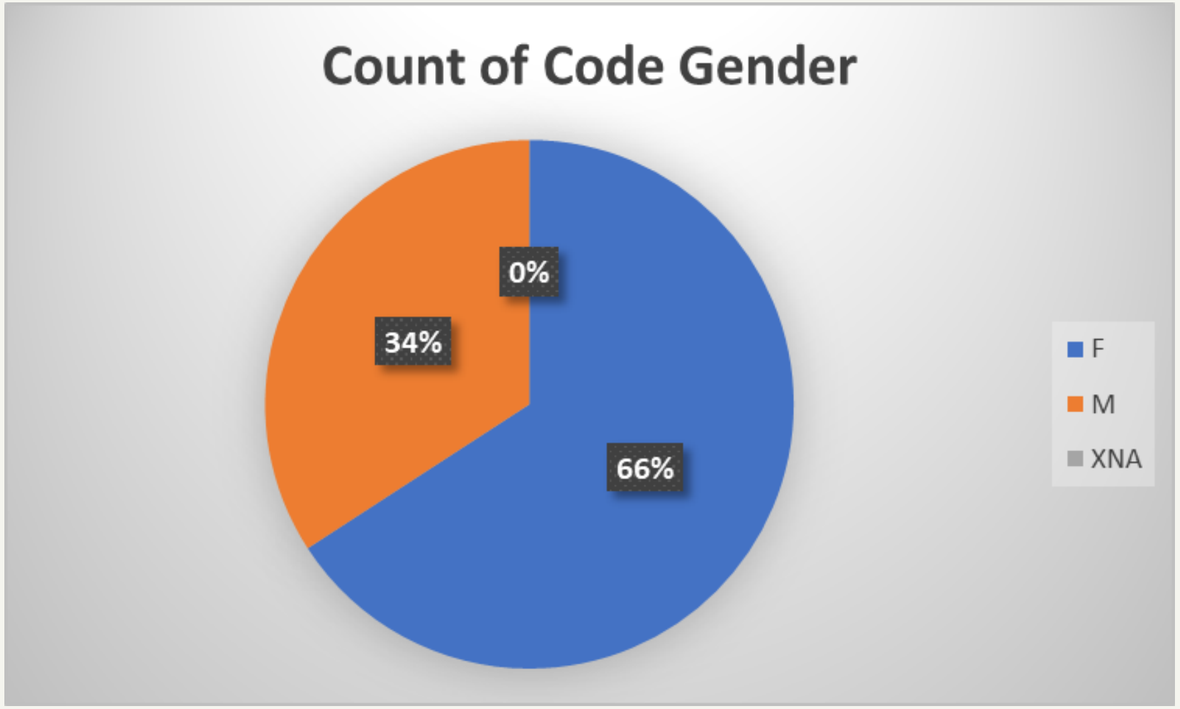
0 → No payment issues
1 → Had some payment issues



GENDER VARIABLE

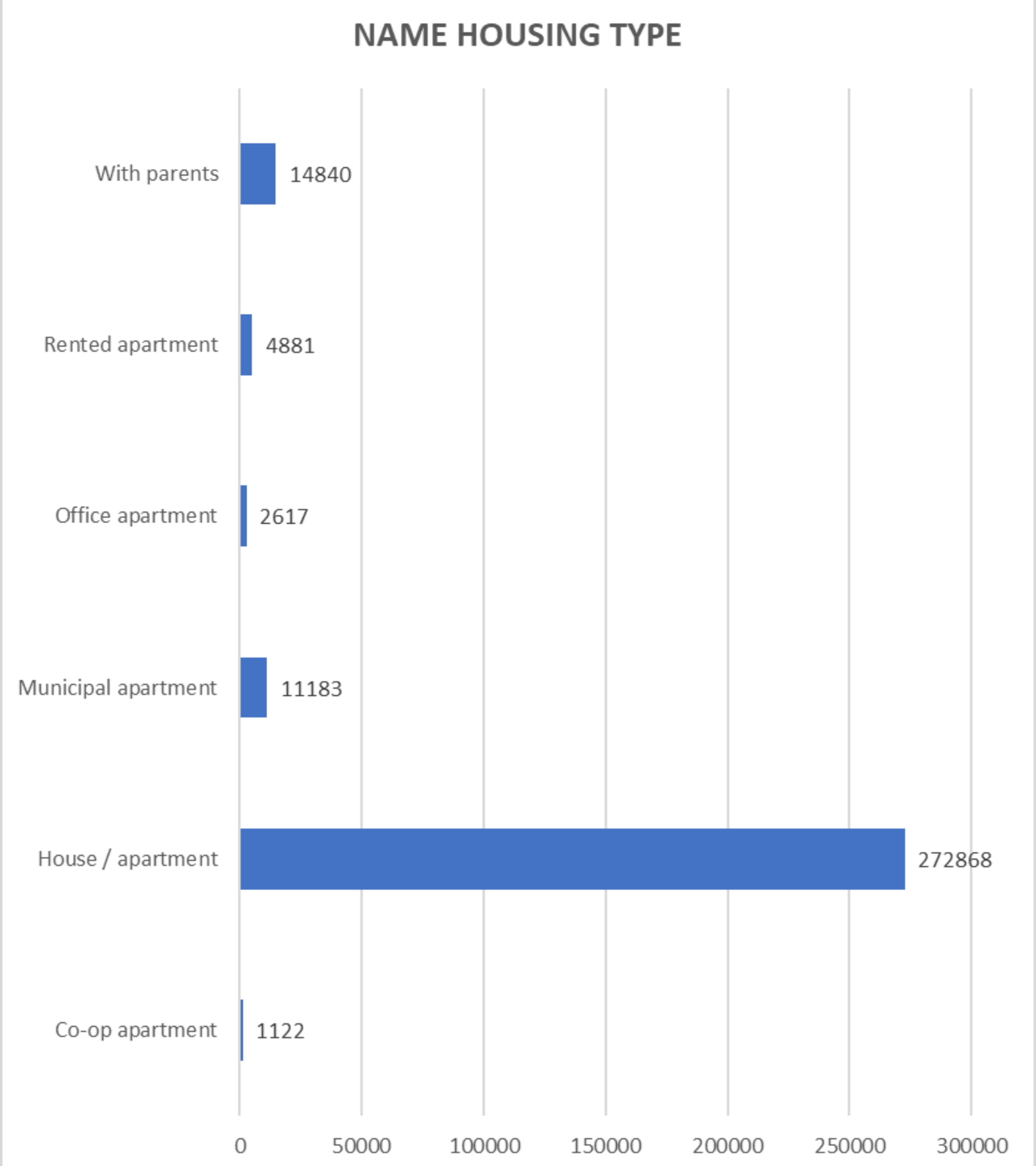
From the GENDER_VARIABLE pie chart we can infer that almost 66% of the clients are female and 34% of the clients are Male The 4 of the appicants have gender as XNA which can be ignoredwhile 8% of the clients had some or the other problem

Row	Count of Code Gender
F	202448
M	105059
XNA	4
Total	307511



NAME HOUSING TYPE

Row	Count of NAME_HOUSING_TYPE
Co-op apartment	1122
House / apartment	272868
Municipal apartment	11183
Office apartment	2617
Rented apartment	4881
With parents	14840
Total	307511

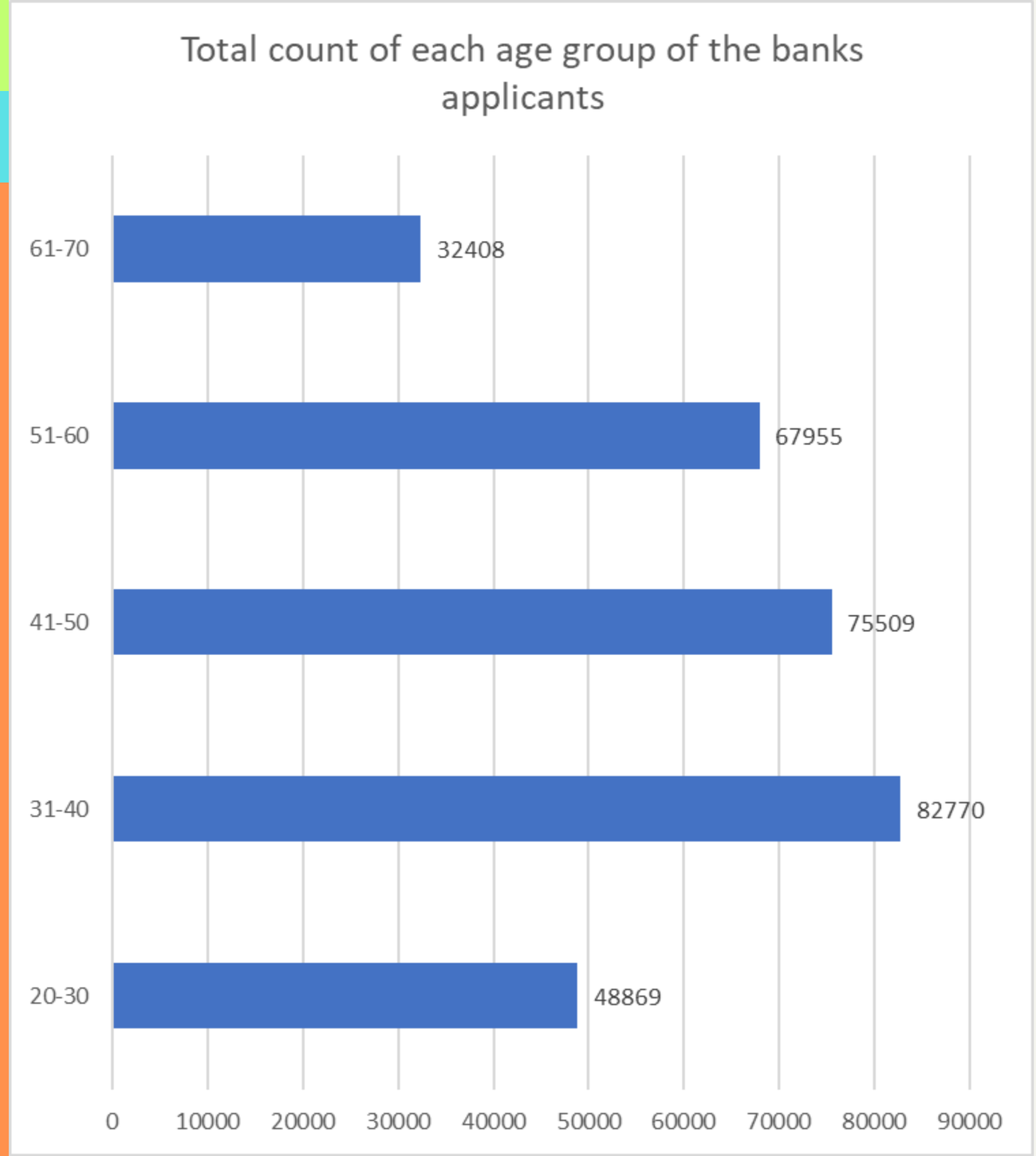


UNIVARIATE ANALYSIS

AGE GROUP

Row	Count of YEARS_BIRTH_RANGE
20-30	48869
31-40	82770
41-50	75509
51-60	67955
61-70	32408
Total	307511

From the adjacent bar plot we can infer that most of the applicants belong to the Age Group ‘31-40’

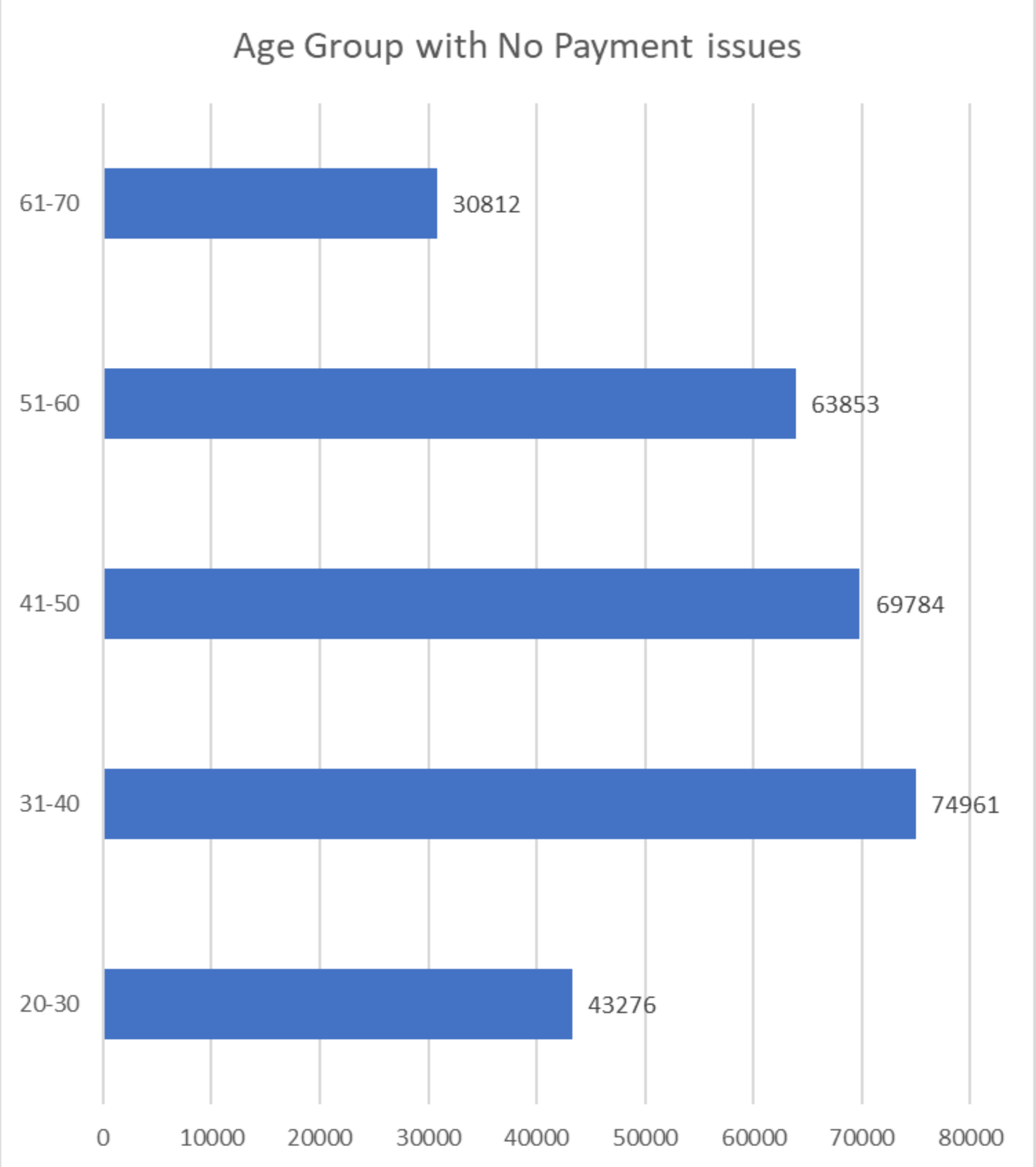


UNIVARIATE ANALYSIS

AGE GROUP

Clients Age Group with no Payment issues	
20-30	43276
31-40	74961
41-50	69784
51-60	63853
61-70	30812
Total	282686

From the adjacent Bar plot we can infer that clients/applicants in the Age Group ’31-40’ are having the highest number when it comes to doing/returning Payment to Banks

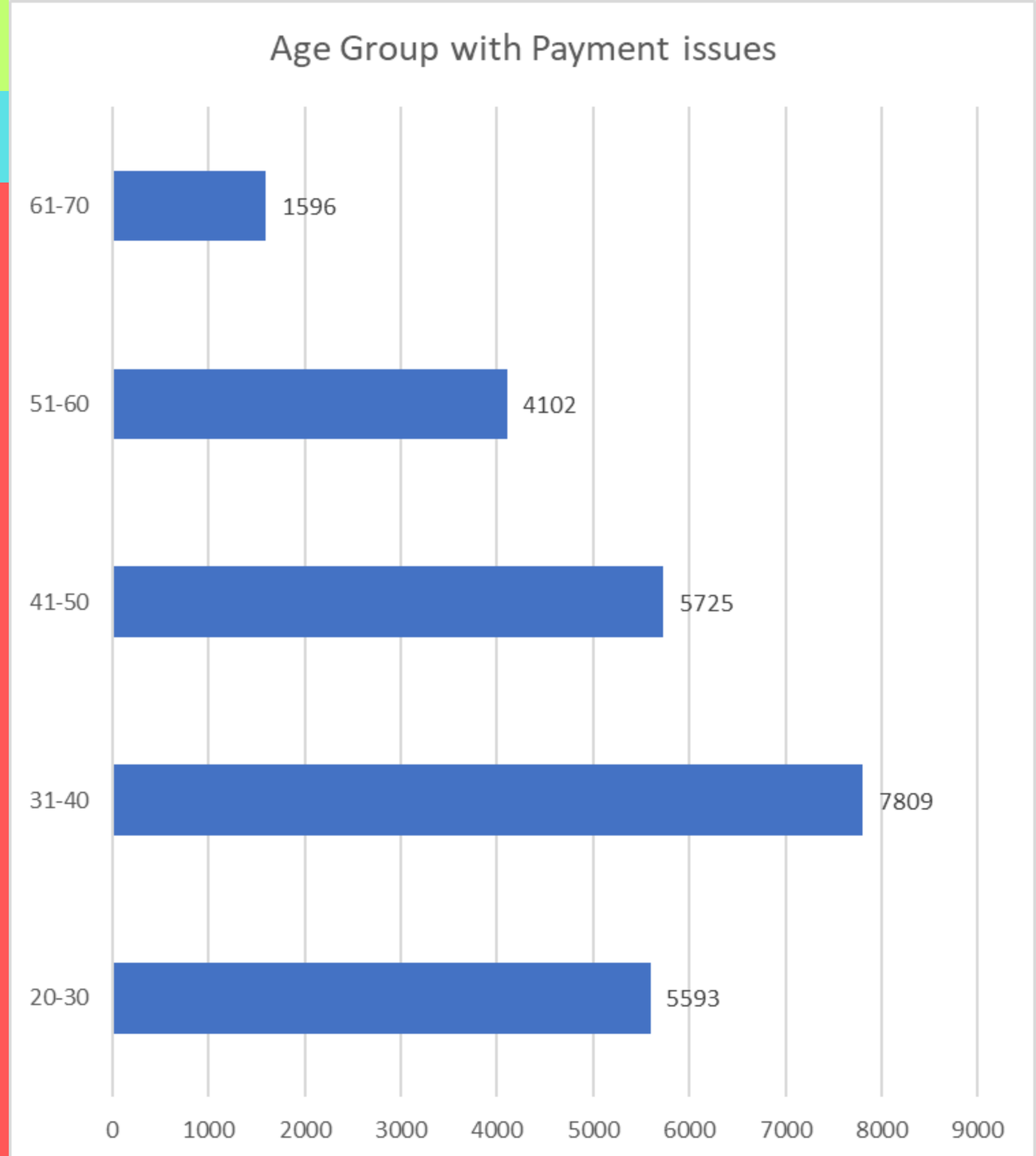


UNIVARIATE ANALYSIS

AGE GROUP

Clients Age Group with Payment issues	
20-30	5593
31-40	7809
41-50	5725
51-60	4102
61-70	1596
Total	24825

From the adjacent Bar plot we can infer that clients/applicants in the Age Group '31-40' are having the highest number of payment issues when it comes to doing/returning Payment to Banks

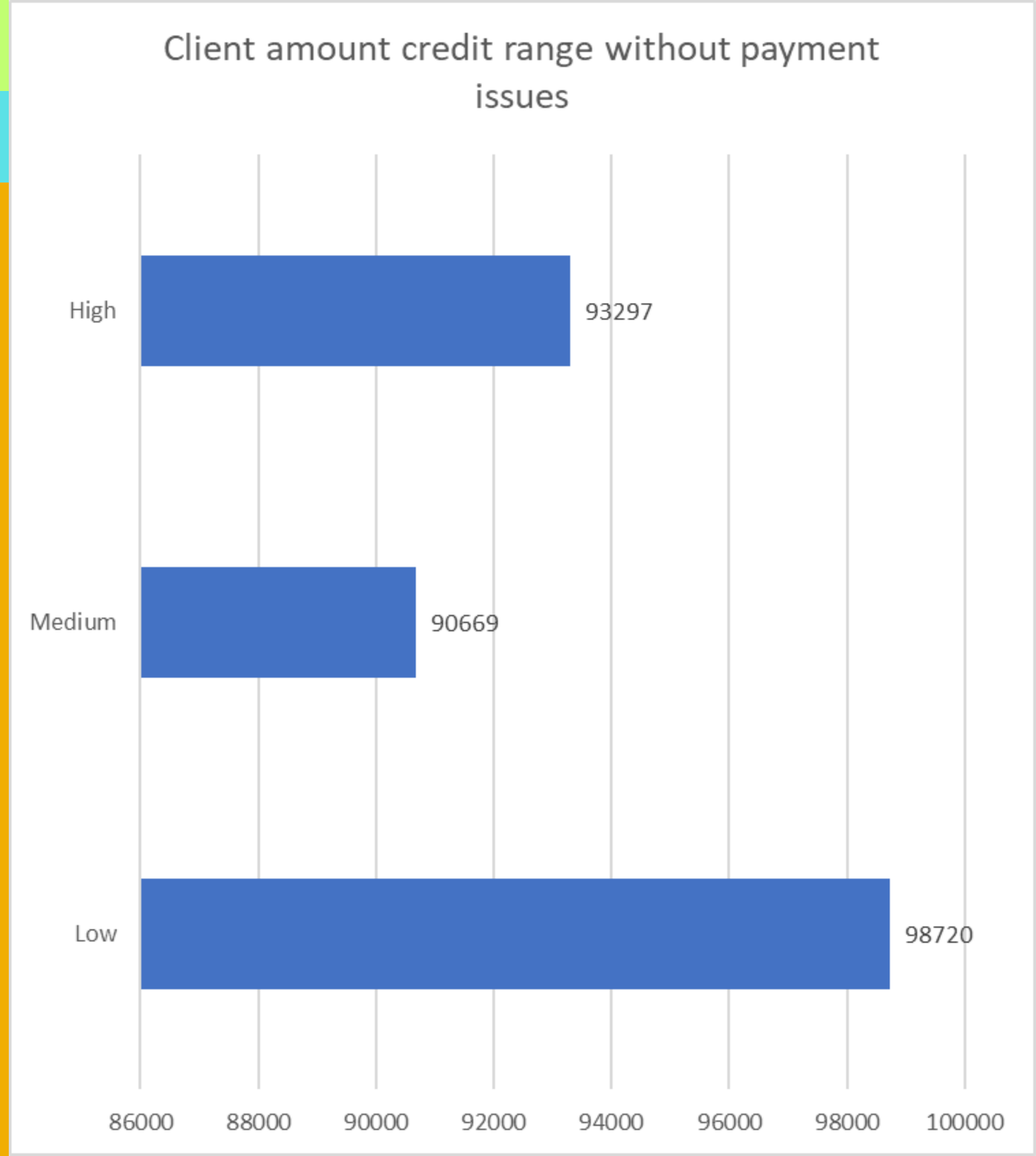


UNIVARIATE ANALYSIS

CLIENT AMOUNT CREDIT RANGE

Row	0
Low	98720
Medium	90669
High	93297
Total	282686

From the adjacent Bar plot we can infer that clients belonging to ‘Low’ income range have the highest count when it comes to clients with no payment issues

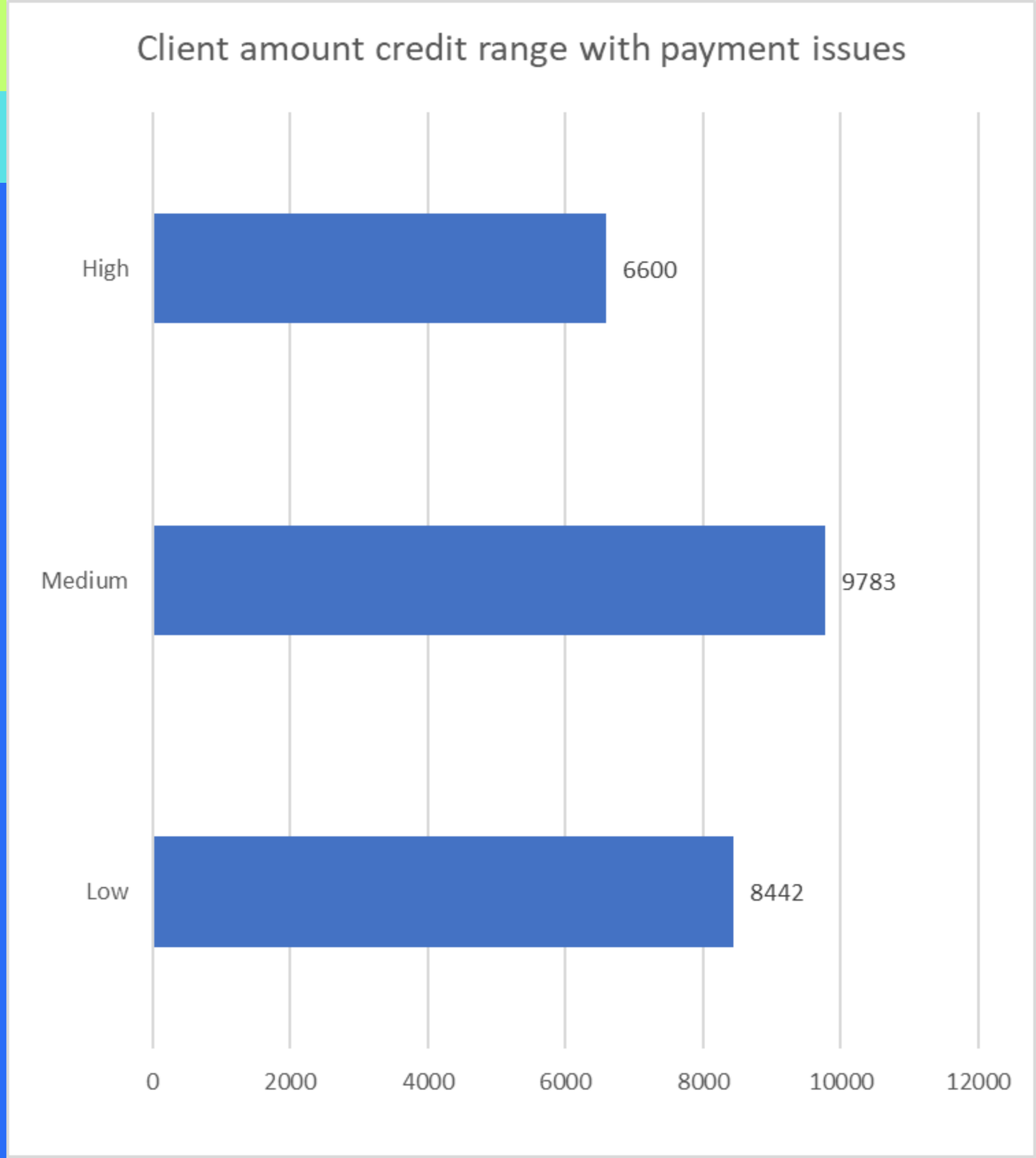


UNIVARIATE ANALYSIS

CLIENT AMOUNT CREDIT RANGE

Row	1
Low	8442
Medium	9783
High	6600
Total	24825

From the adjacent Bar plot we can infer that clients belonging to 'Medium' income range have the highest count when it comes to clients with payment issues

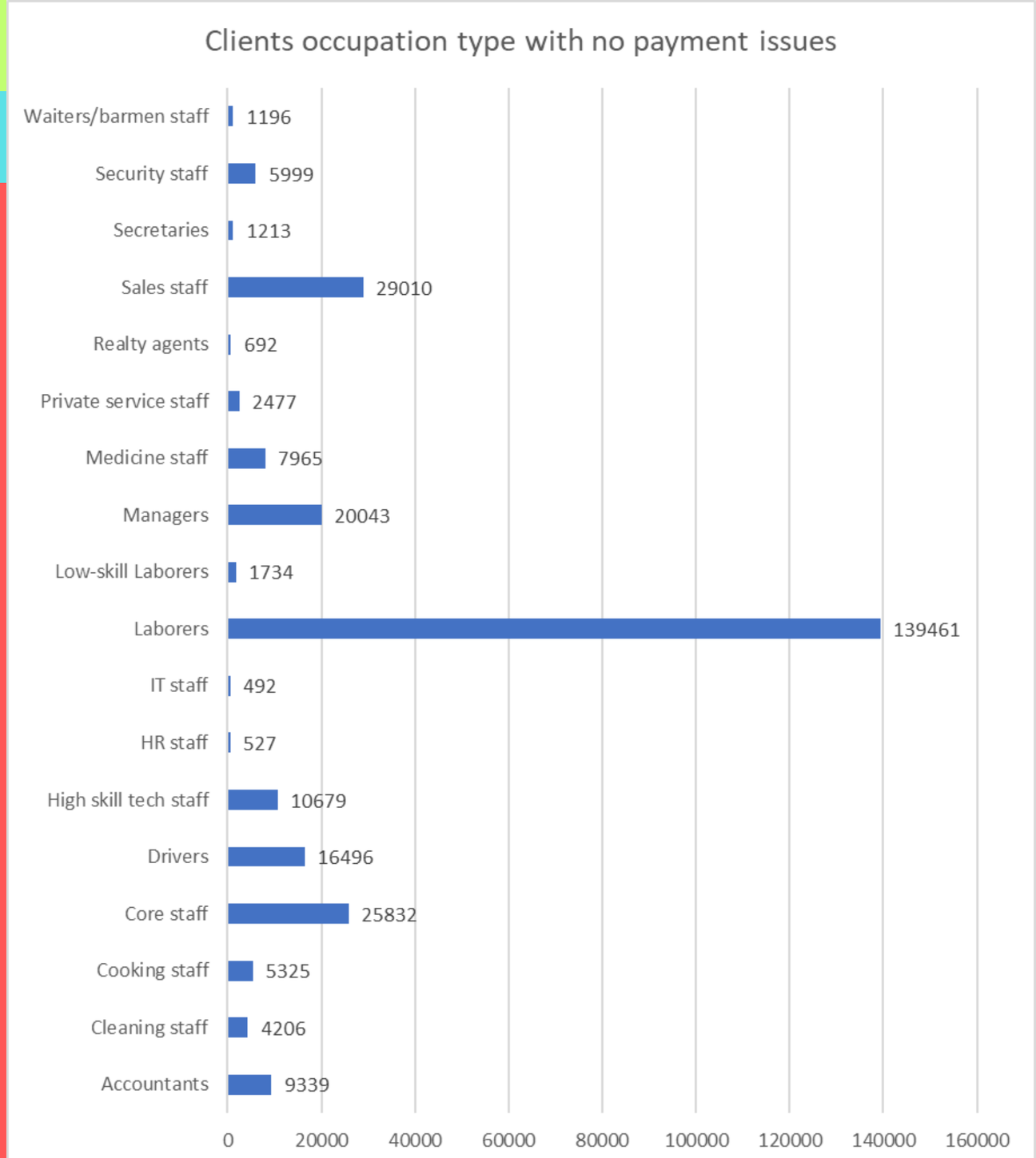


UNIVARIATE ANALYSIS

OCCUPATION TYPE

Row	0
Accountants	9339
Cleaning staff	4206
Cooking staff	5325
Core staff	25832
Drivers	16496
High skill tech staff	10679
HR staff	527
IT staff	492
Laborers	139461
Low-skill Laborers	1734
Managers	20043
Medicine staff	7965
Private service staff	2477
Realty agents	692
Sales staff	29010
Secretaries	1213
Security staff	5999
Waiters/barmen staff	1196
Total	282686

From the above bar plot we can infer that clients with occupation_type 'Laborers' have the highest number of count when it comes to clients with no payment issues

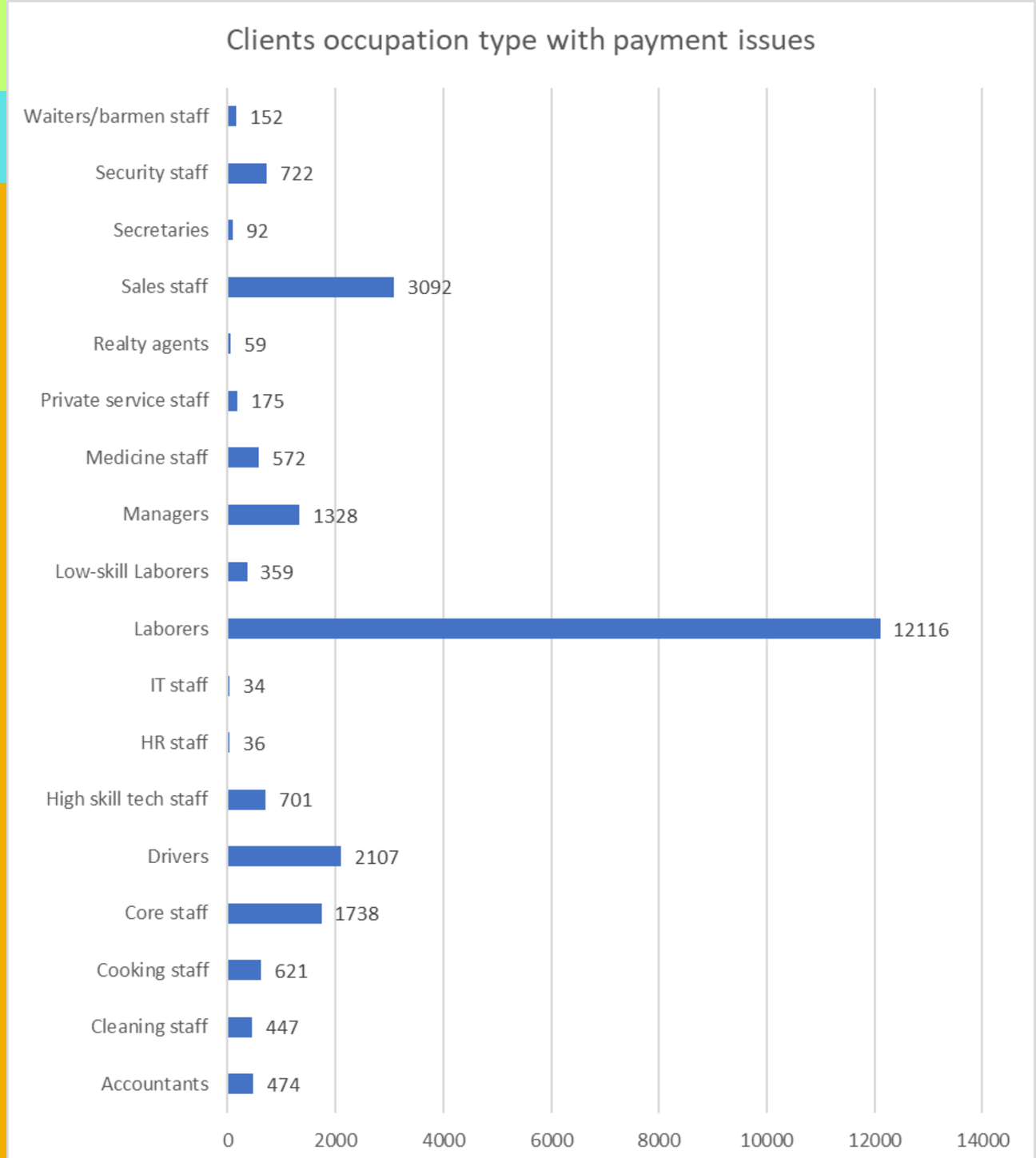


UNIVARIATE ANALYSIS

OCCUPATION TYPE

Row	1
Accountants	474
Cleaning staff	447
Cooking staff	621
Core staff	1738
Drivers	2107
High skill tech staff	701
HR staff	36
IT staff	34
Laborers	12116
Low-skill Laborers	359
Managers	1328
Medicine staff	572
Private service staff	175
Realty agents	59
Sales staff	3092
Secretaries	92
Security staff	722
Waiters/barmen staff	152
Total	24825

From the above bar plot we can infer that clients with occupation_type 'Laborers' have the highest number of count when it comes to clients with payment issues

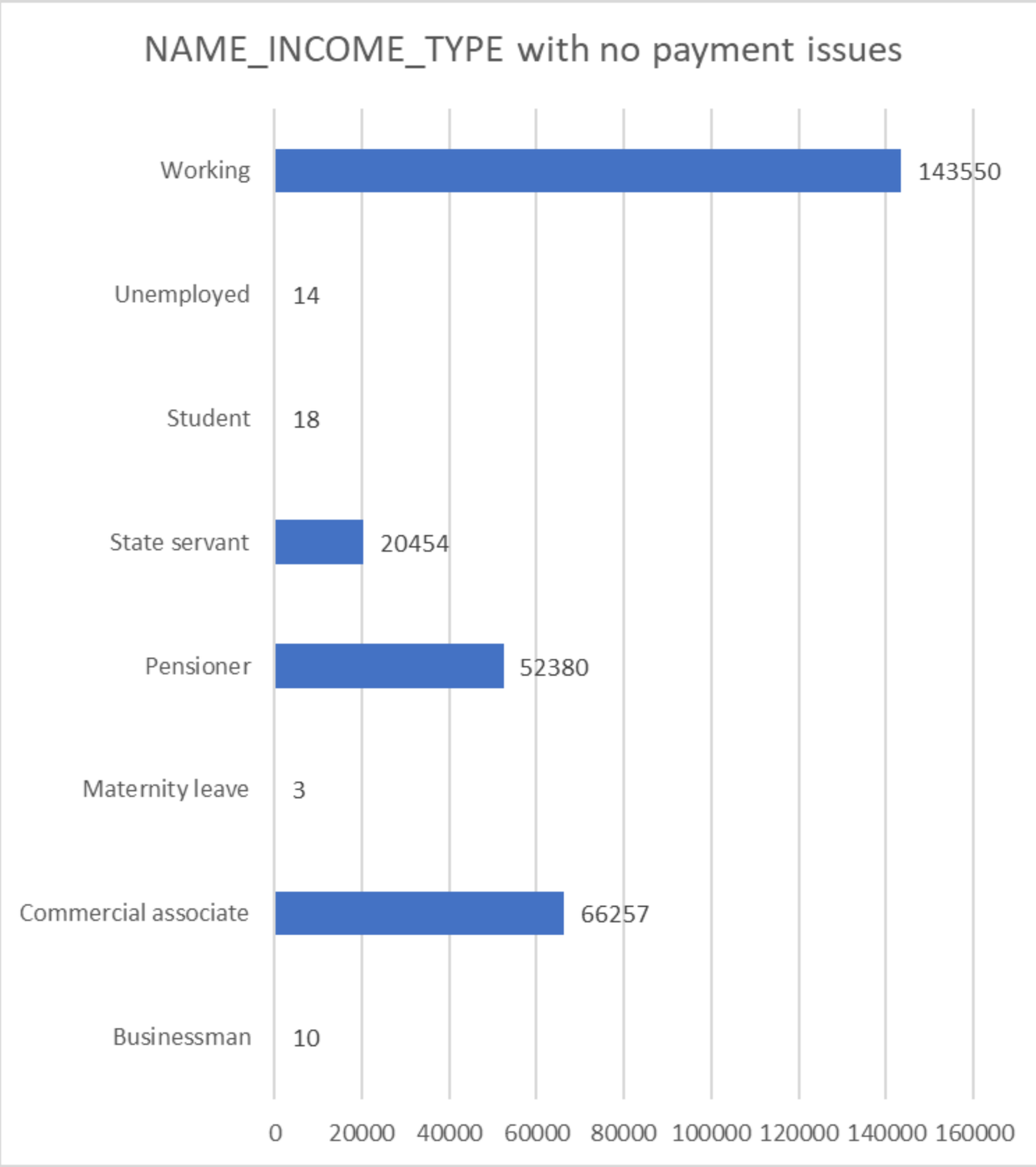


UNIVARIATE ANALYSIS

NAME INCOME TYPE

Row	0
Businessman	10
Commercial associate	66257
Maternity leave	3
Pensioner	52380
State servant	20454
Student	18
Unemployed	14
Working	143550
Total	282686

From the above Bar plot we can infer that clients having income_type as ‘WORKING’ have the highest count when it comes to clients with no payment issues

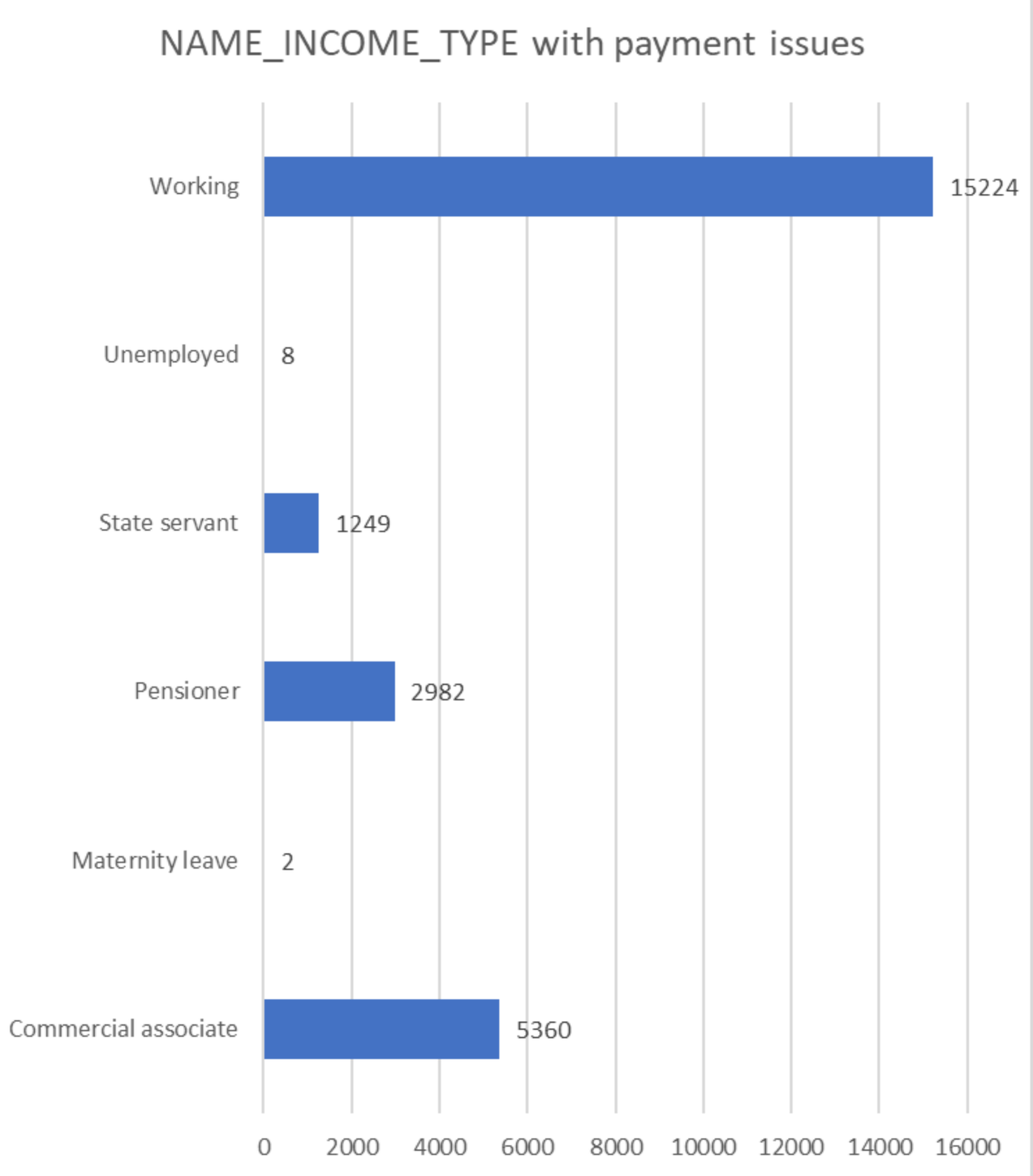


UNIVARIATE ANALYSIS

NAME INCOME TYPE

Row	1
Commercial associate	5360
Maternity leave	2
Pensioner	2982
State servant	1249
Unemployed	8
Working	15224
Total	24825

From the above Bar plot we can infer that clients having income_type as 'WORKING' have the highest count when it comes to clients with no payment issues

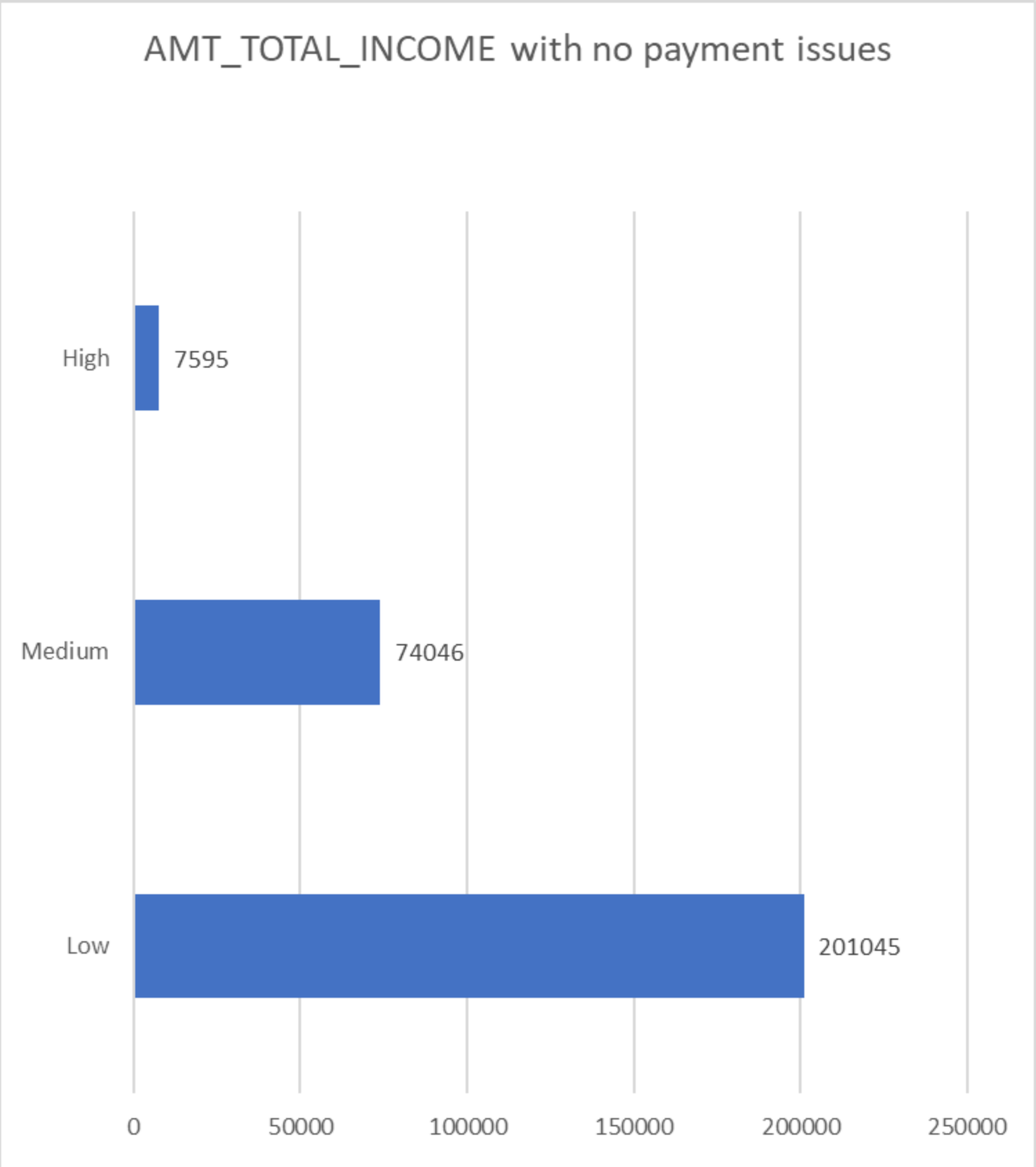


UNIVARIATE ANALYSIS

AMT_TOTAL INCOME

Row	0
Low	201045
Medium	74046
High	7595
Total	282686

From the above Bar plot we can infer that client having the total income range as 'LOW' have the highest count when it comes to clients having no payment issues

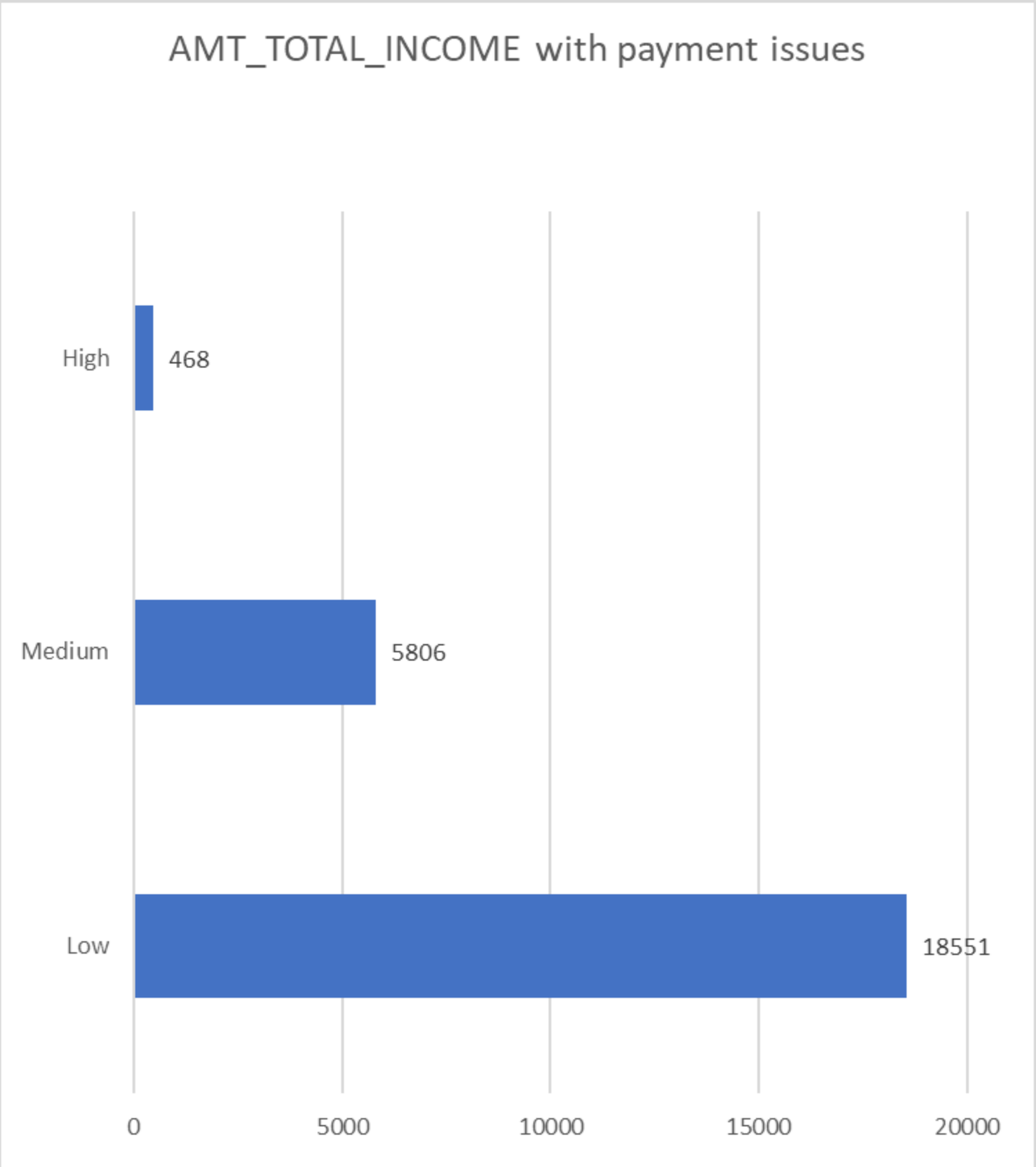


UNIVARIATE ANALYSIS

AMT_TOTAL INCOME

Row	1
Low	18551
Medium	5806
High	468
Total	24825

From the above Bar plot we can infer that client having the total income range as 'LOW' have the highest count when it comes to clients having payment issues

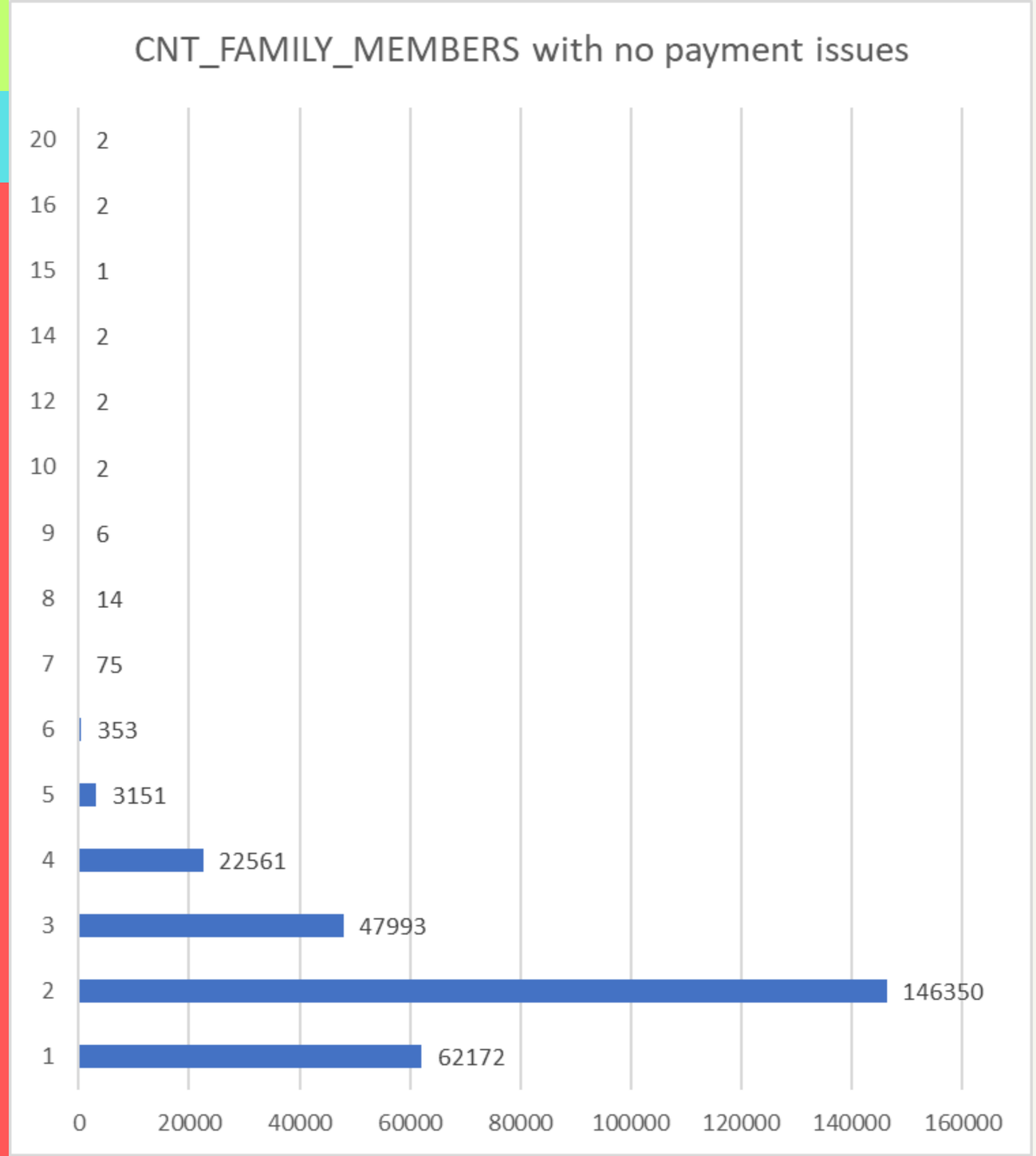


UNIVARIATE ANALYSIS

CNT_FAMILY_MEMBERS

Row	0
1	62172
2	146350
3	47993
4	22561
5	3151
6	353
7	75
8	14
9	6
10	2
12	2
14	2
15	1
16	2
20	2
Total	282686

From the above Bar plot we can infer that clients having total count of family members as 2 have the highest count when it comes to clients having no payment issues

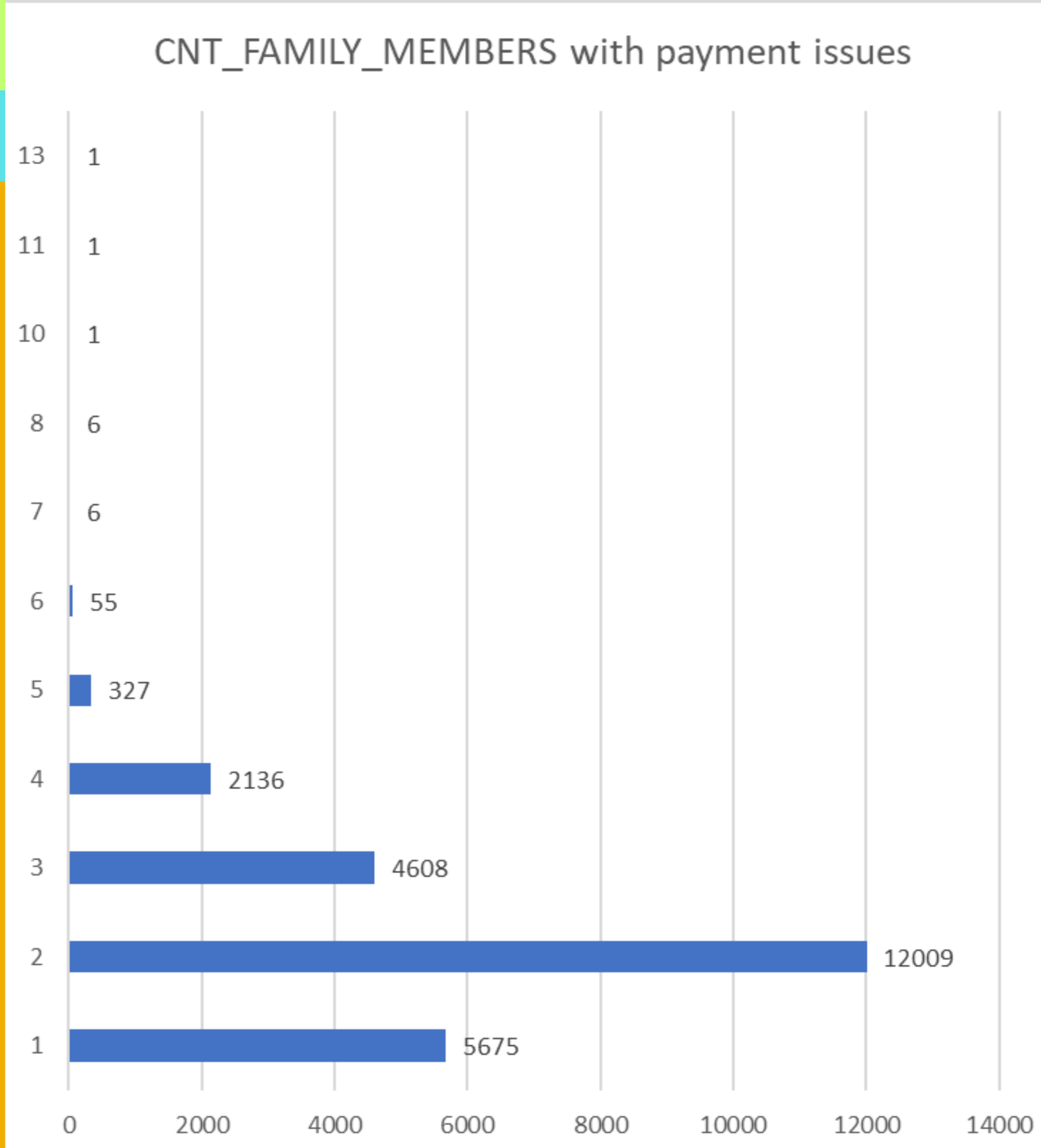


UNIVARIATE ANALYSIS

CNT_FAMILY_MEMBERS

Row	1
1	5675
2	12009
3	4608
4	2136
5	327
6	55
7	6
8	6
10	1
11	1
13	1
Total	24825

From the above Bar plot we can infer that clients having total count of family members as 2 have the highest count when it comes to clients having payment issues

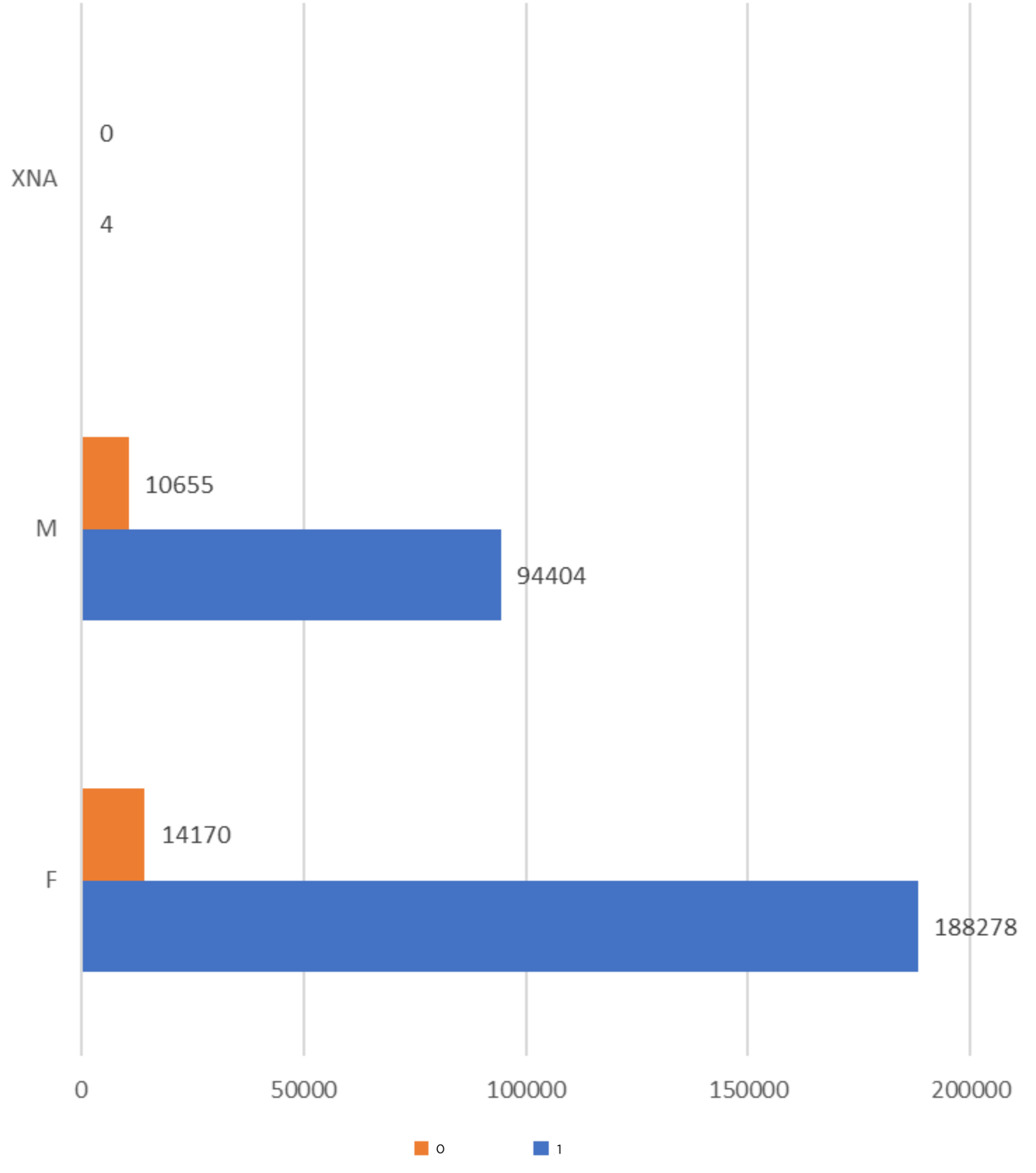


UNIVARIATE ANALYSIS FOR TARGET VARIABLE

CODE_GENDER

Row	0	1	Total
F	188278	14170	202448
M	94404	10655	105059
XNA	4		4
Total	282686	24825	

om the above Bar Plot we can infer that Clients with CODE_GENDER = 'F' have the highest number of non-defaulters i.e. 188278-14170 = 174108

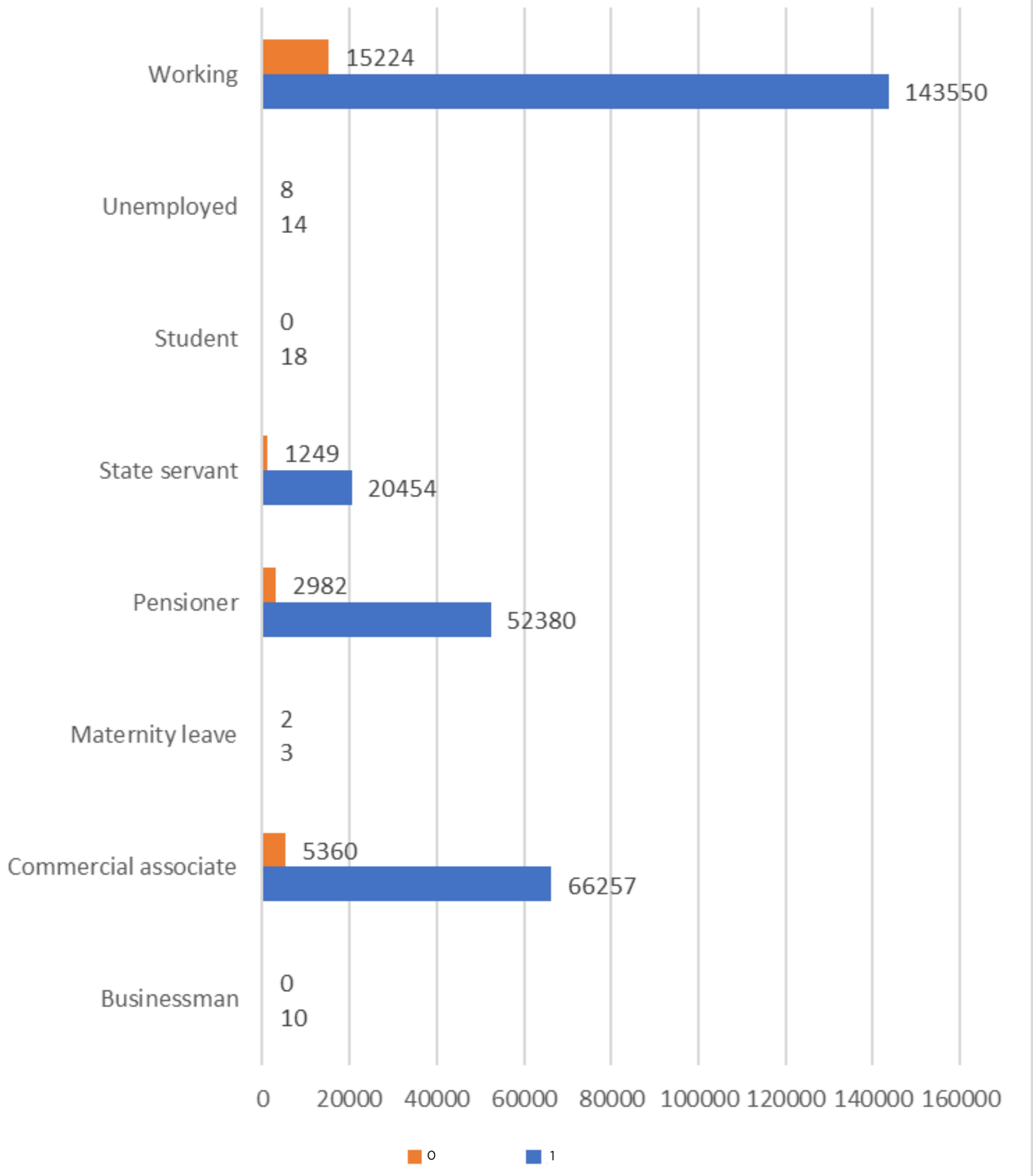


UNIVARIATE ANALYSIS FOR TARGET VARIABLE

NAME_INCOME_TYPE

Row	0	1	Total
Businessman	10	0	10
Commercial associate	66257	5360	71617
Maternity leave	3	2	5
Pensioner	52380	2982	55362
State servant	20454	1249	21703
Student	18	0	18
Unemployed	14	8	22
Working	143550	15224	158774
Total	282686	24825	307511

From the adjacent Bar Plot we can infer that clients having NAME_INCOME_TYPE = 'WORKING' having the highest count of Non-defaulters i.e. 143550-15224 = 128326

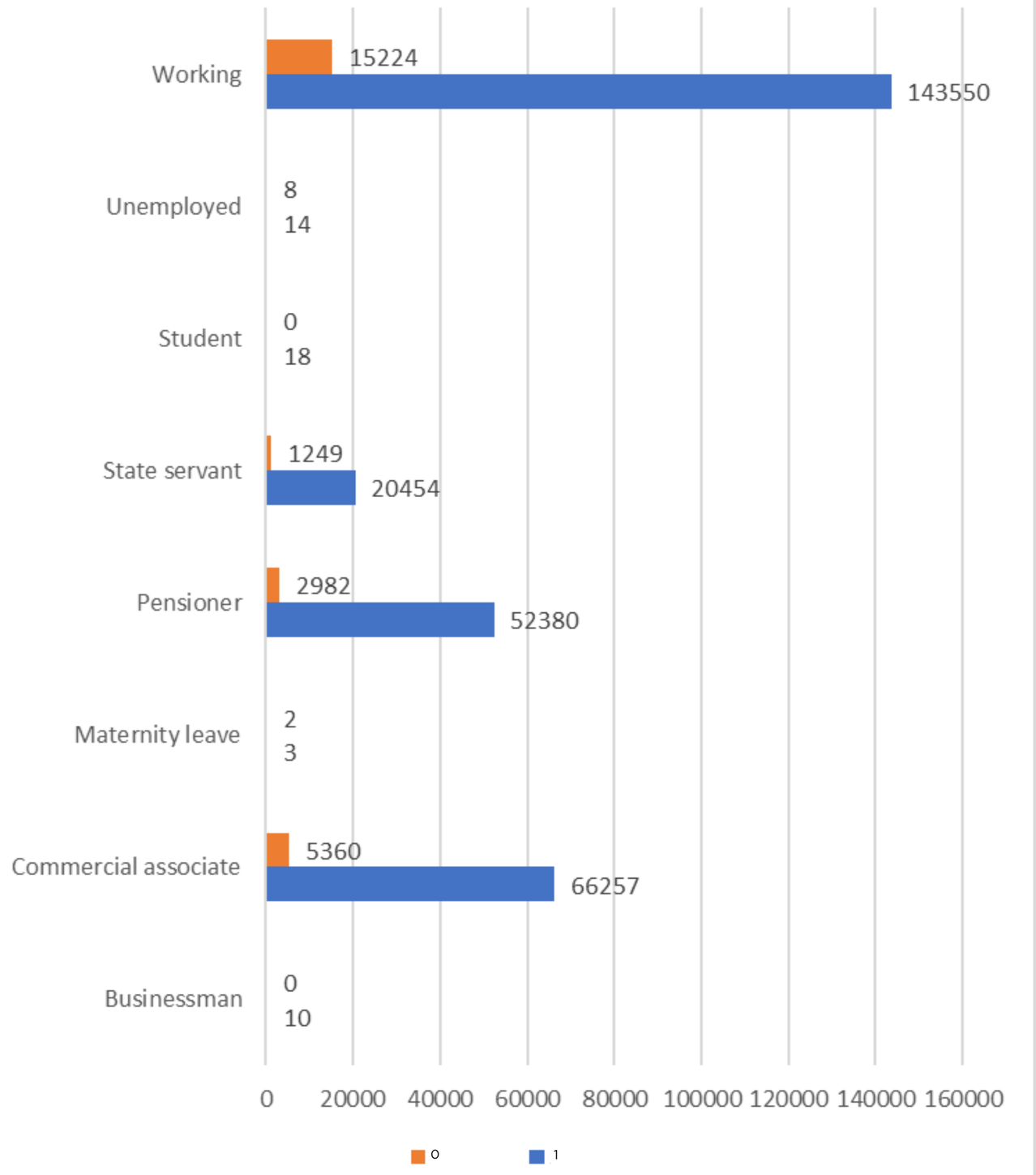


UNIVARIATE ANALYSIS FOR TARGET VARIABLE

NAME_EDUCATION_TYPE

Row	0	1	Total
Businessman	10	0	10
Commercial associate	66257	5360	71617
Maternity leave	3	2	5
Pensioner	52380	2982	55362
State servant	20454	1249	21703
Student	18	0	18
Unemployed	14	8	22
Working	143550	15224	158774
Total	282686	24825	307511

From the above Bar Plot we can infer that clients having NAME_EDUCATION_TYPE = 'SECONDARY/SECONDARY SPECIAL' have the highest count for Nondefaulters i.e. 198867-19524 = 179343

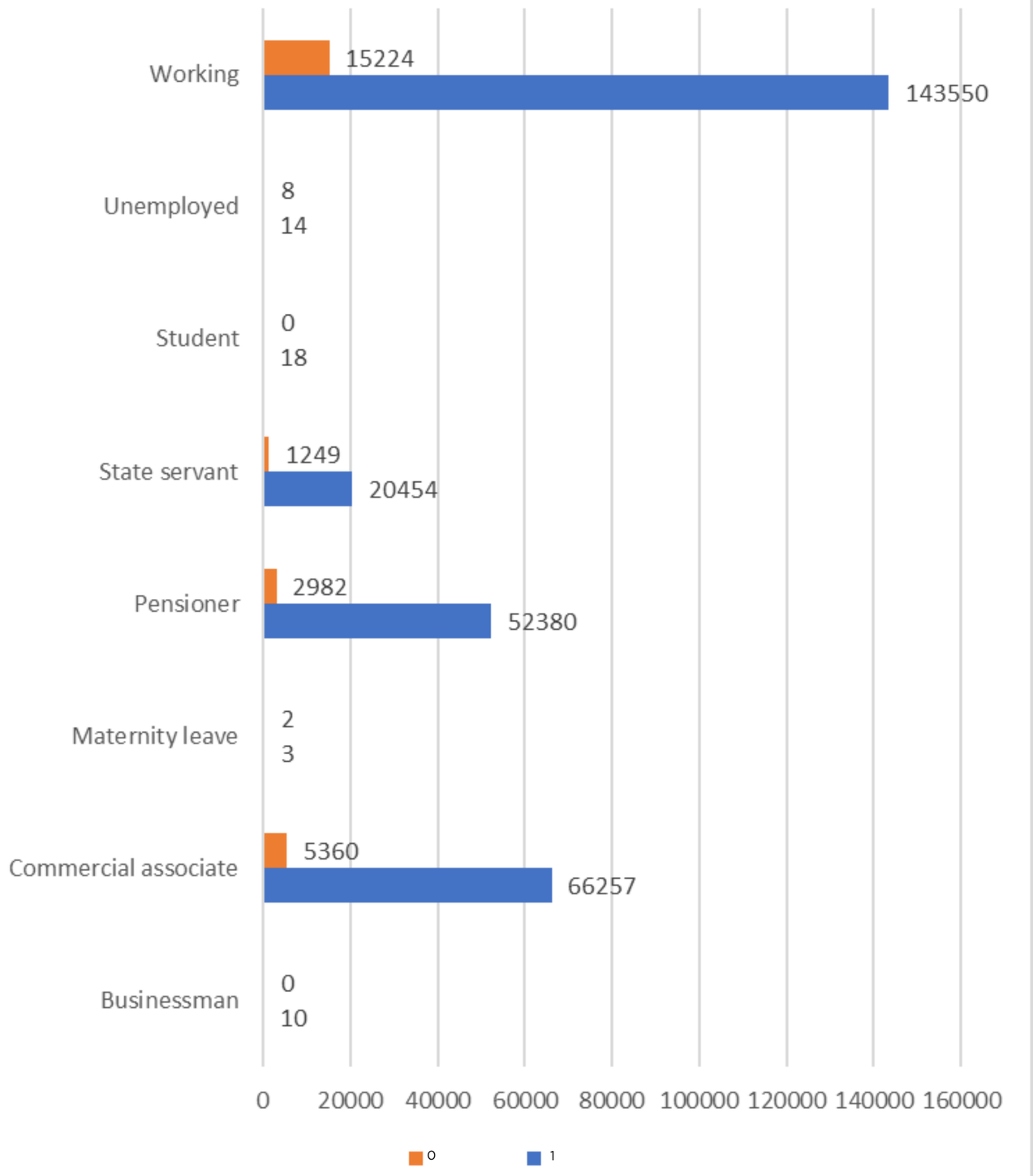


UNIVARIATE ANALYSIS FOR TARGET VARIABLE

NAME_FAMILY_STATUS

Row	0	1	Total
Civil marriage	26814	2961	29775
Married	181582	14850	196432
Separated	18150	1620	19770
Single / not married	40987	4457	45444
Unknown	2	0	2
Widow	15151	937	16088
Total	282686	24825	307511

From the adjacent Bar Plot we can infer that clients having NAME_FAMILY_STATUS = 'MARRIED' have the highest count of Nondefaulters i.e. 181582 – 14850 = 166732

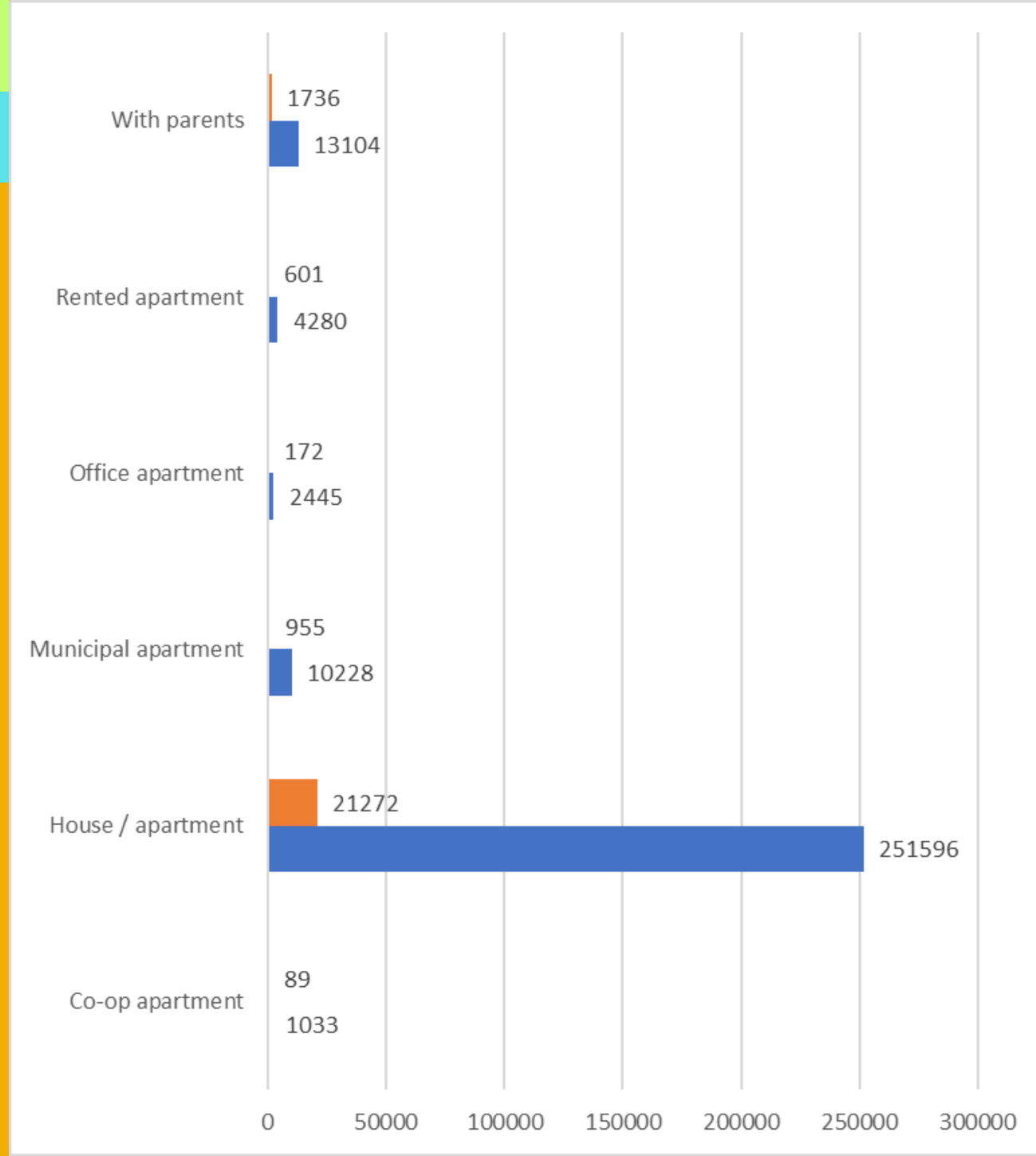


UNIVARIATE ANALYSIS FOR TARGET VARIABLE

NAME_HOUSING_TYPE

Row	0	1	Total
Co-op apartment	1033	89	1122
House / apartment	251596	21272	272868
Municipal apartment	10228	955	11183
Office apartment	2445	172	2617
Rented apartment	4280	601	4881
With parents	13104	1736	14840
Total	282686	24825	307511

From the above Bar Plot we can infer that clients having NAME_HOUSING_TYPE = 'House/Apartment' have the highest count of Non-defaulters i.e. 251596-21272 = 230324

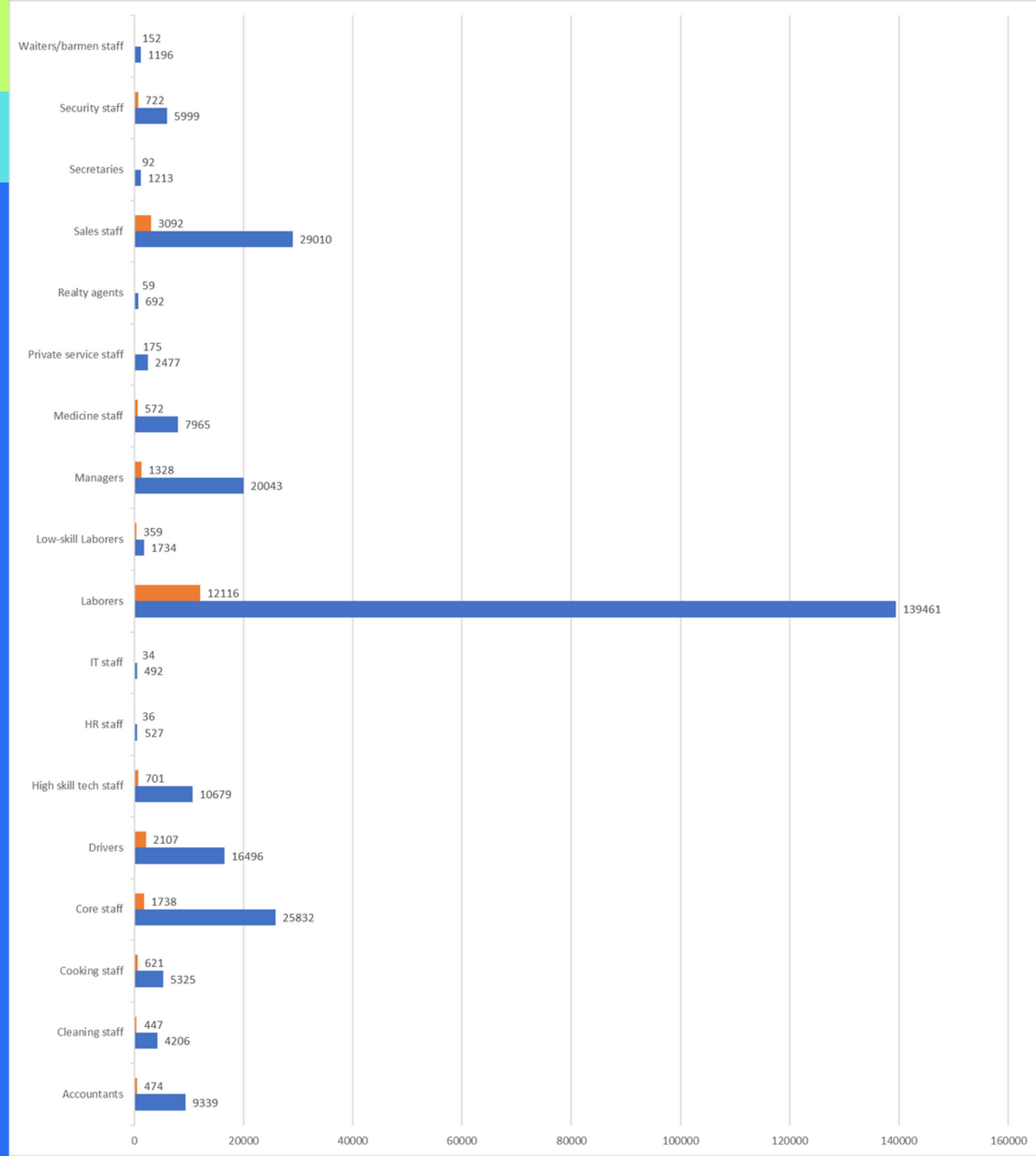


UNIVARIATE ANALYSIS FOR TARGET VARIABLE

OCCUPATION_TYPE

Row	0	1	Total
Accountants	9339	474	9813
Cleaning staff	4206	447	4653
Cooking staff	5325	621	5946
Core staff	25832	1738	27570
Drivers	16496	2107	18603
High skill tech staff	10679	701	11380
HR staff	527	36	563
IT staff	492	34	526
Laborers	139461	12116	151577
Low-skill Laborers	1734	359	2093
Managers	20043	1328	21371
Medicine staff	7965	572	8537
Private service staff	2477	175	2652
Realty agents	692	59	751
Sales staff	29010	3092	32102
Secretaries	1213	92	1305
Security staff	5999	722	6721
Waiters/barmen staff	1196	152	1348
Total	282686	24825	307511

From the adjacent Bar plot we can infer that clients having occupation_type = 'Laborers' have the highest count for Non-defaulters i.e. 139461-12116 = 127345

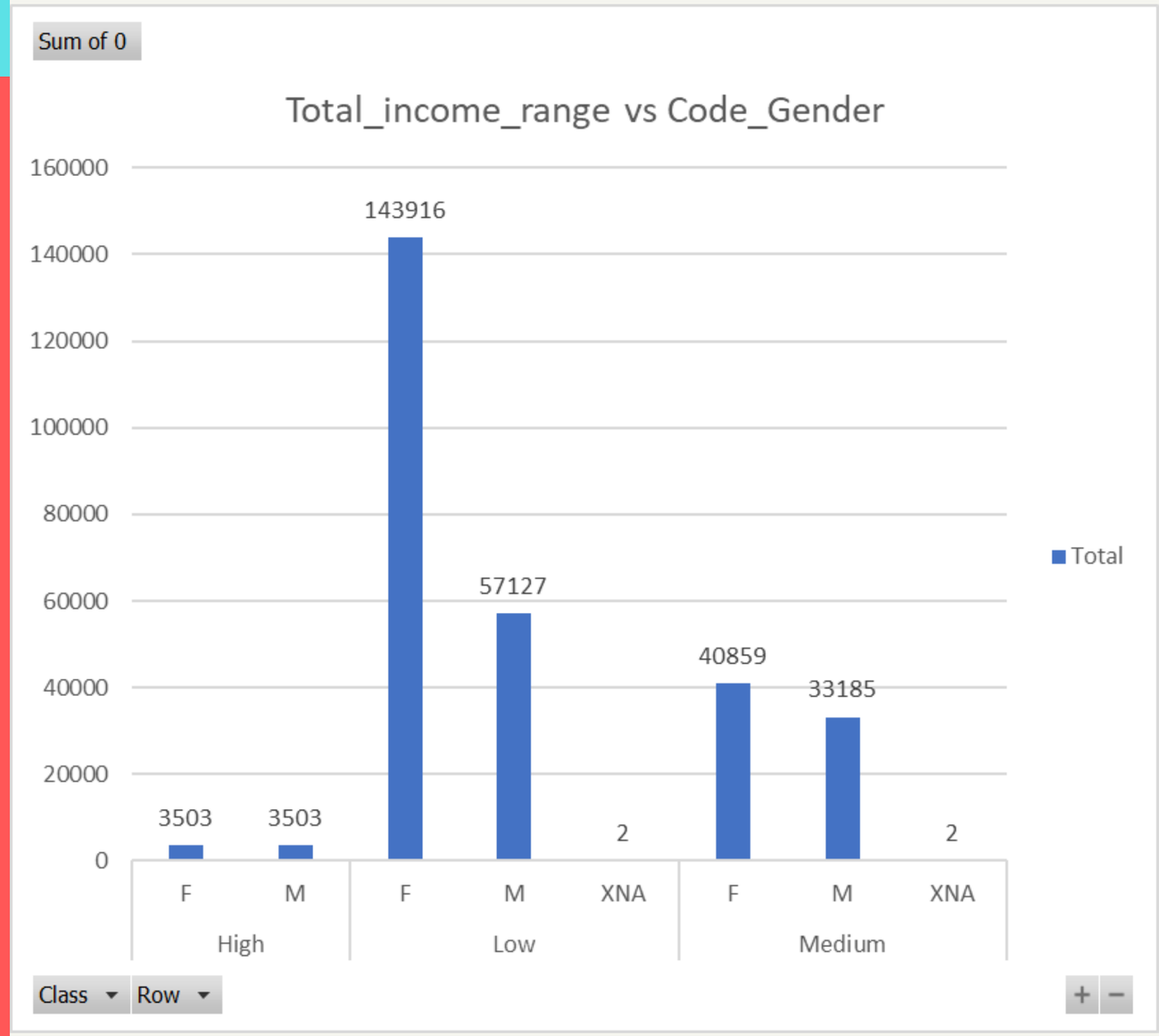


BIVARIATE ANALYSIS FOR TARGET VARIABLE

0: TOTAL_INCOME_RANGE VS CODE_GENDER

Row Labels	Sum of 0
High	**7006**
F	3503
M	3503
Low	**201045**
F	143916
M	57127
XNA	2
Medium	**74046**
F	40859
M	33185
XNA	2
Grand Total	**282097**

From the above Bar plot we can infer that Females belonging to Low income group are the highest number of clients with no payment issues

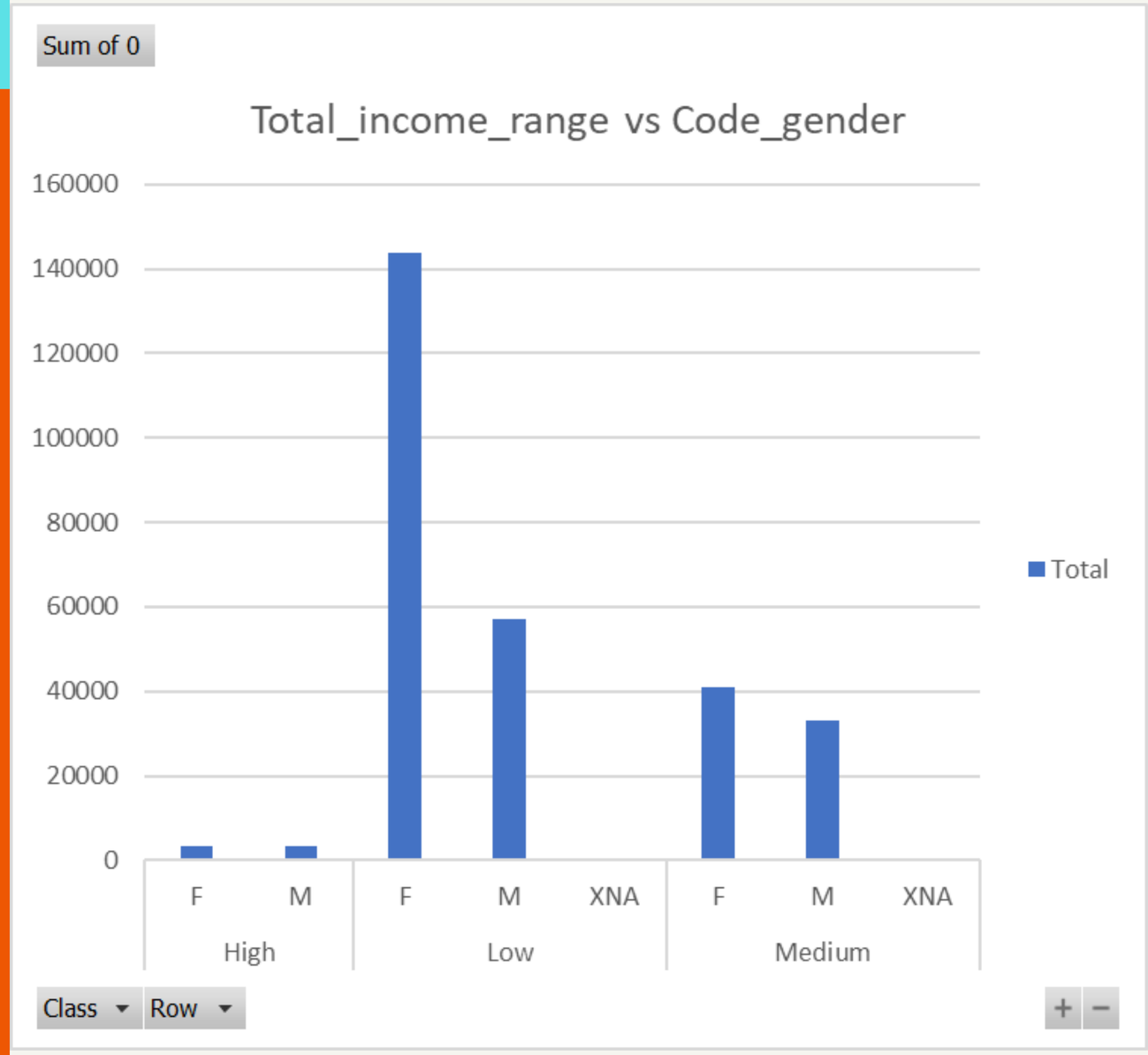


BIVARIATE ANALYSIS FOR TARGET VARIABLE

1: TOTAL_INCOME_RANGE VS CODE_GENDER

Row Labels	Sum of 0
High	**7006**
F	3503
M	3503
Low	**201045**
F	143916
M	57127
XNA	2
Medium	**74046**
F	40859
M	33185
XNA	2
Grand Total	**282097**

From the above Bar plot we can infer that Females belonging to Low income group are the highest number of clients with payment issues

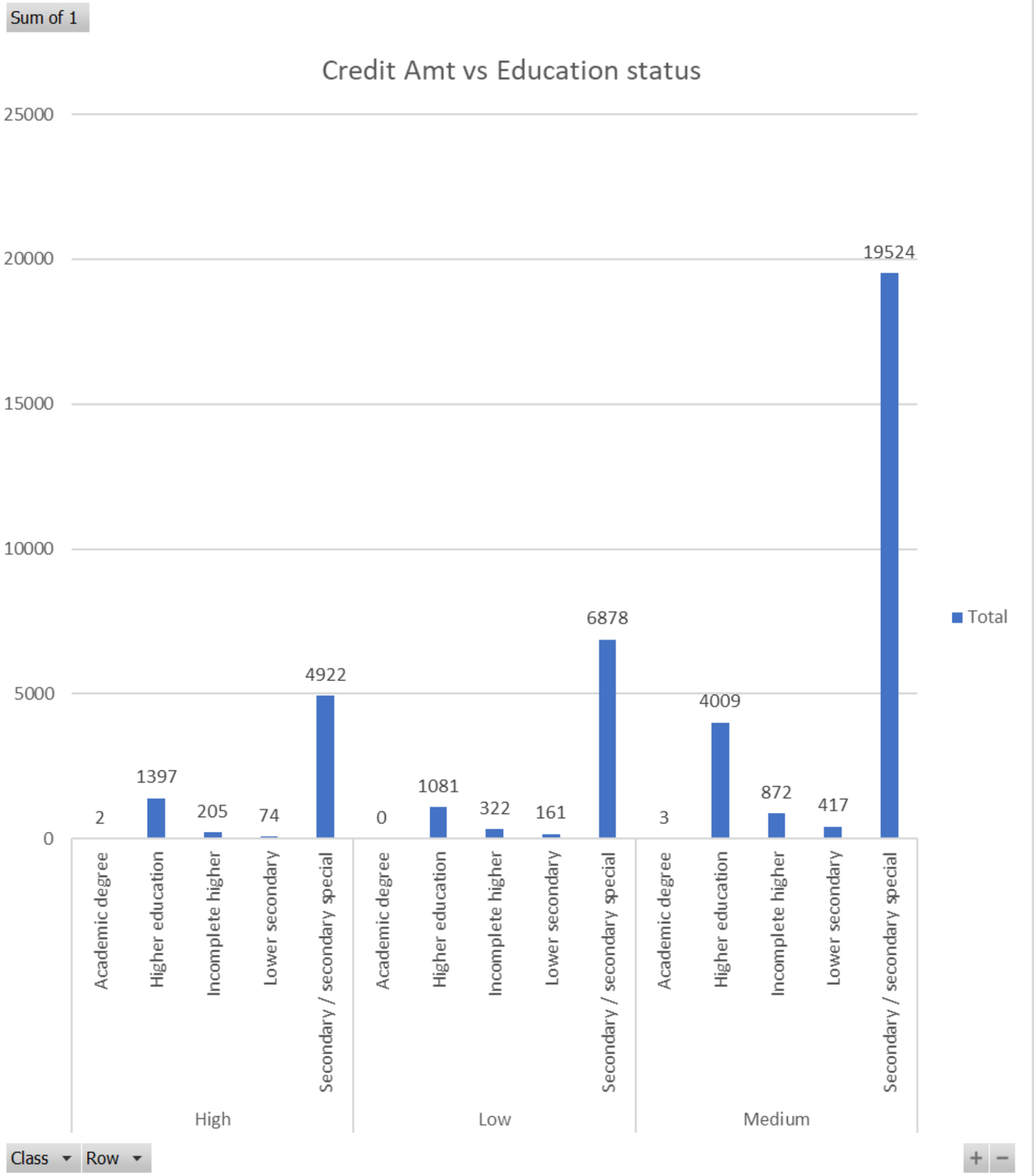


BIVARIATE ANALYSIS FOR TARGET VARIABLE

0: CREDIT AMT VS EDUCATION STATUS

Row Labels	Sum of 1
High	**6600**
Academic degree	2
Higher education	1397
Incomplete higher	205
Lower secondary	74
Secondary / secondary special	4922
Low	**8442**
Academic degree	0
Higher education	1081
Incomplete higher	322
Lower secondary	161
Secondary / secondary special	6878
Medium	**24825**
Academic degree	3
Higher education	4009
Incomplete higher	872
Lower secondary	417
Secondary / secondary special	19524
Grand Total	**39867**

From the adjacent Bar Plot we can infer that clients having credit amt range as 'Low' and education status as 'Secondary/ Secondary Special' have the highest count for clients with no payment issues

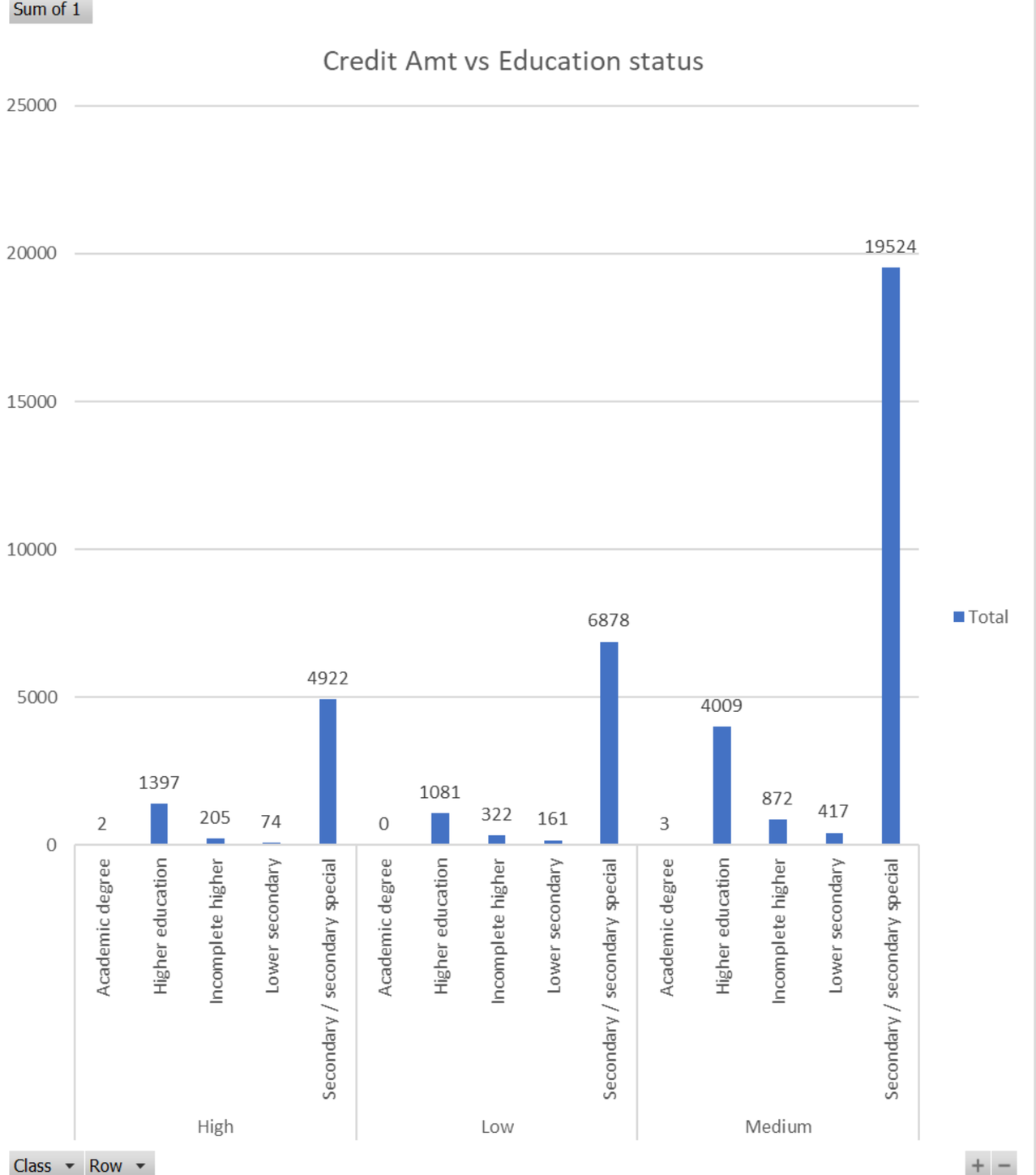


BIVARIATE ANALYSIS FOR TARGET VARIABLE

1: CREDIT AMT VS EDUCATION STATUS

Row Labels	Sum of 1
High	**6600**
Academic degree	2
Higher education	1397
Incomplete higher	205
Lower secondary	74
Secondary / secondary special	4922
Low	**8442**
Academic degree	0
Higher education	1081
Incomplete higher	322
Lower secondary	161
Secondary / secondary special	6878
Medium	**24825**
Academic degree	3
Higher education	4009
Incomplete higher	872
Lower secondary	417
Secondary / secondary special	19524
Grand Total	**39867**

From the adjacent Bar Plot we can infer that clients having credit amt range as ‘Medium’ and education status as ‘Secondary/ Secondary Special’ have the highest count for clients with payment issues

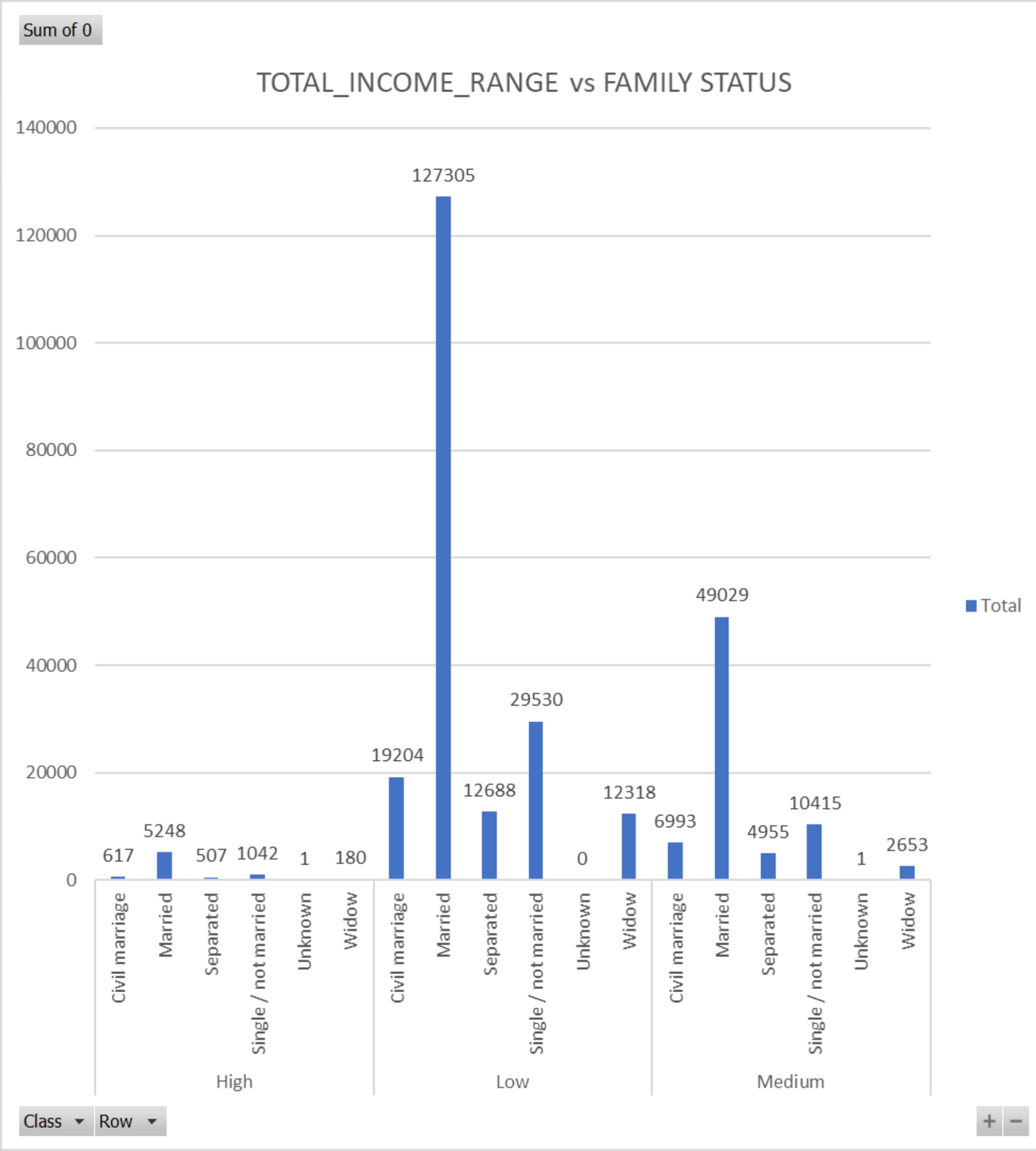


BIVARIATE ANALYSIS FOR TARGET VARIABLE

0: TOTAL INCOME VS FAMILY STATUS

Row Labels	Sum of 0
High	**7595**
Civil marriage	617
Married	5248
Separated	507
Single / not married	1042
Unknown	1
Widow	180
Low	**201045**
Civil marriage	19204
Married	127305
Separated	12688
Single / not married	29530
Unknown	0
Widow	12318
Medium	**74046**
Civil marriage	6993
Married	49029
Separated	4955
Single / not married	10415
Unknown	1
Widow	2653
Grand Total	**282686**

From the adjacent Bar plot we can infer that clients with total_income_range as 'Low' and family_status as 'Married' have the highest count for clients having no payment issues

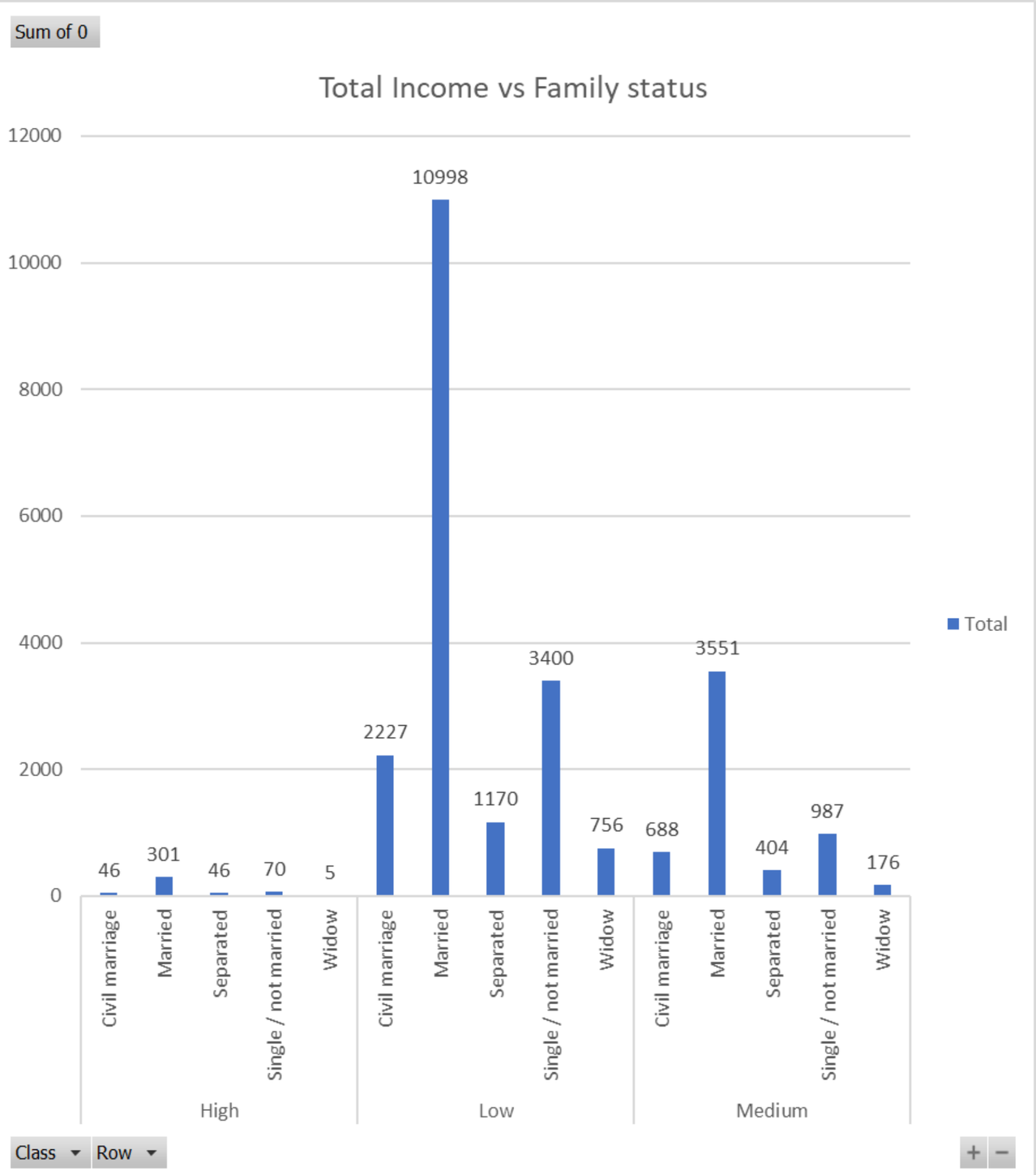


BIVARIATE ANALYSIS FOR TARGET VARIABLE

1: TOTAL INCOME VS FAMILY STATUS

Row Labels	Sum of 0
High	**468**
Civil marriage	46
Married	301
Separated	46
Single / not married	70
Widow	5
Low	**18551**
Civil marriage	2227
Married	10998
Separated	1170
Single / not married	3400
Widow	756
Medium	**5806**
Civil marriage	688
Married	3551
Separated	404
Single / not married	987
Widow	176
Grand Total	**24825**

From the adjacent Bar plot we can infer that clients with total_income_range as ‘Low’ and family_status as ‘Married’ have the highest count for clients having payment issues



Previous Application Dataset – Dropping, Imputing and analyzing Null values

The following columns of the previous application datasets need to be dropped as they are irrelevant for doing the data analysis

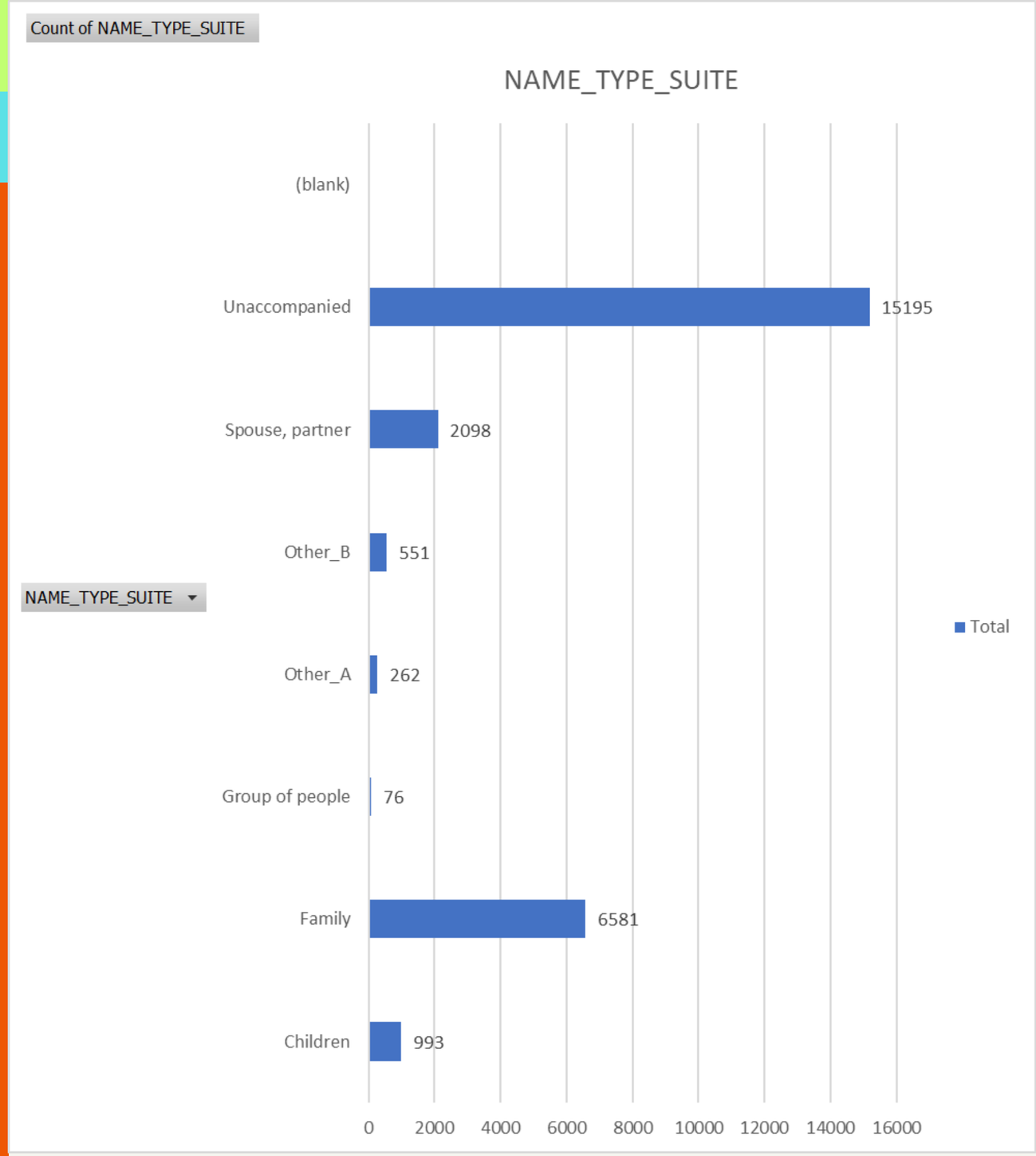
- HOUR_APPR_PROCESS_START
- WEEKDAY_APPR_PROCESS_START_PREV
- FLAG_LAST_APPL_PER_CONTRACT
- NFLAG_LAST_APPL_IN_DAY
- SK_ID_CURR
- WEEKDAY_APPR_PROCESS_START

Removing the rows with the values 'XNA' &'XAP' for the column:
NAME_TYPE_SUITE

CLEANING OF DATA

NAME_TYPE_SUITE

Replace all the Blanks with “Unaccompanied”

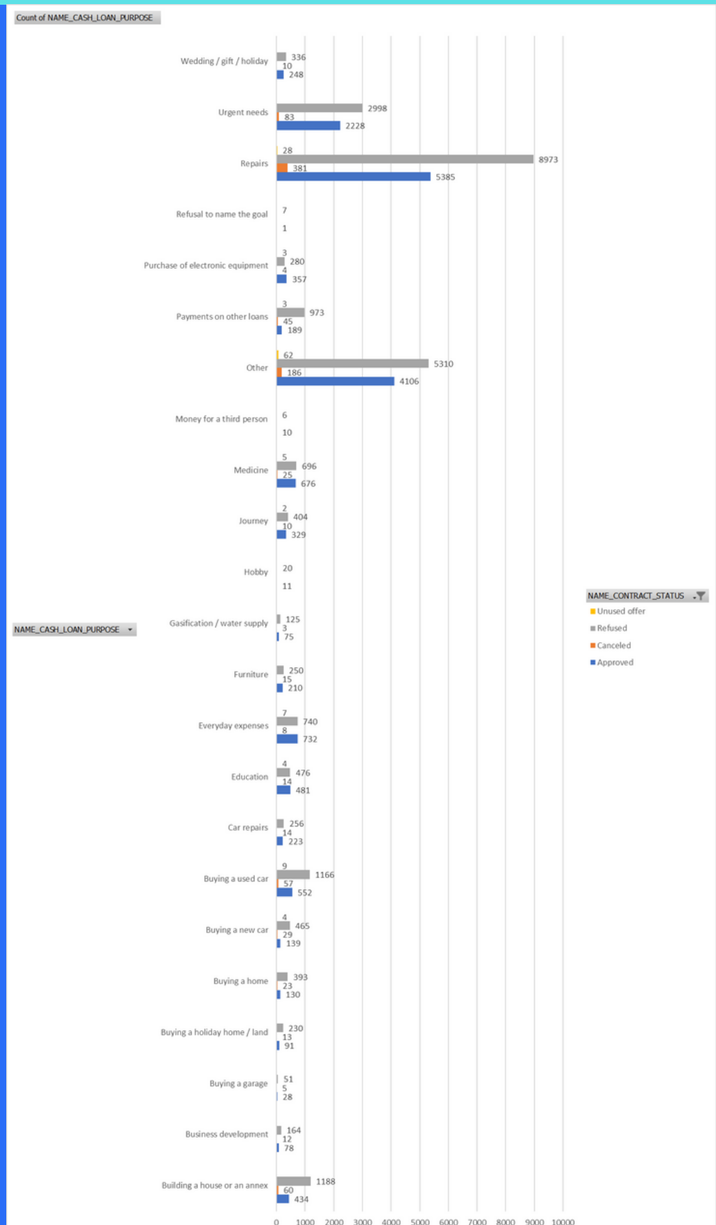


ANALYSIS

NAME_TYPE_SUITE

Count of NAME_CASH_LOAN_PURPOSE	Column Labels				
Row Labels	Approved	Canceled	Refused	Unused offer	Grand Total
Building a house or an annex	434	60	1188		1682
Business development	78	12	164		254
Buying a garage	28	5	51		84
Buying a holiday home / land	91	13	230		334
Buying a home	130	23	393		546
Buying a new car	139	29	465	4	637
Buying a used car	552	57	1166	9	1784
Car repairs	223	14	256		493
Education	481	14	476	4	975
Everyday expenses	732	8	740	7	1487
Furniture	210	15	250		475
Gasification / water supply	75	3	125		203
Hobby	11		20		31
Journey	329	10	404	2	745
Medicine	676	25	696	5	1402
Money for a third person	10		6		16
Other	4106	186	5310	62	9664
Payments on other loans	189	45	973	3	1210
Purchase of electronic equipment	357	4	280	3	644
Refusal to name the goal	1		7		8
Repairs	5385	381	8973	28	14767
Urgent needs	2228	83	2998		5309
Wedding / gift / holiday	248	10	336		594
Grand Total	**16713**	**997**	**25507**	**127**	**43344**

DISTRIBUTION



Hence the analysis are being done on both datasets Applications Dataset and Precious Applications Dataset The following conclusions were drawn from the analysis done

- THE PROPORTION/PERCENTAGE OF THE DEFAULTERS(TARGET = 1) IS AROUND 8% AND THAT OF NON-DEFAULTERS(TARGET = 0) IS AROUND 92%
- THE BANK GENERALLY LENDS MORE LOAN TO FEMALE CLIENTS AS COMPARED TO MALES CLIENTS AS THE COUNT OF FEMALE CLIENTS IN THE DEFAULTER'S LIST IS LESS THAN THAT OF MALES. STILL BANK CAN LOOK FOR MORE MALE CLIENTS IF THEIR CREDIT AMOUNT IS SATISFIED
- ALSO THE CLIENTS WHO BELONG TO WORKING CLASS TEND TO PAY THEIR LOANS ON TIME FOLLOWED BY THE CLIENTS WHO FALL UNDER COMMERCIAL ASSOCIATE
- CLIENTS HAVING EDUCATION STATUS LIKE SECONDARY/ HIGHER SECONDARY OR MORE TEND TO PAY LOAN ON TIME SO BANK CAN PREFER LENDING LOANS TO CLIENTS HAVING SUCH EDUCATION STATUS
- CLIENTS WHO FALL IN THE AGE GROUP 31-40 HAVE THE HIGHEST COUNT FOR PAYING OFF THEIR LOANS ON TIME FOLLOWED BY THE CLIENTS WHO FALL IN THE AGE GROUPS 41-60
- CLIENTS HAVING LOW CREDIT AMOUNT RANGE TEND TO PAY OFF THEIR LOANS ON TIME THAN COMPARED TO HIGH AND MEDIUM CREDIT RANGE
- CLIENTS LIVING WITH THEIR PARENTS TEND TO PAY OFF THEIR LOANS QUICKLY AS COMPARED TO OTHER HOUSING TYPE. SO BANK CAN LEND LOAN TO CLIENTS HAVING HOUSING TYPE → LIVING WITH PARENTS
- CLIENTS TAKING LOAN FOR PURCHASING NEW HOME I.E. CLIENTS TAKING HOME LOANS OR PURCHASING NEW CAR I.E. CAR LOANS AND CLIENTS WHO HAVE A INCOME TYPE AS STATE SERVANT TEND TO PAY THEIR LOANS ON TIME AND HENCE BANK SHOULD PREFER CLIENTS HAVING SUCH BACKGROUND
- THE BANK SHOULD BE MORE CAUTIOUS WHEN LENDING MONEY TO CLIENTS WITH REPAIRS PURPOSE BECAUSE THEY HAVE HIGH COUNT OF DEFAULTERS ALONG WITH HIGH COUNT OF DEFAULTERS

Thank you!