# Bank Loan Case Study

analysis performed on **Application dataset**

# Approach

The approach will be to perform the tasks in an organized way, creating separate sheets in the workbook for different tasks performed.

The project analysis follows the below mentioned steps:

Step 1: Data Cleaning (Null values removal & imputation, duplicates removal, outliers detection, irrelevant columns removal)

Step 2: Univariate Analysis

Step 3: Fraud Analysis

Step 4: Bivariate Analysis

Step 5: Correlation Analysis

# Irrelevant columns deleted

89 irrelevant columns were deleted which were not required for our analysis

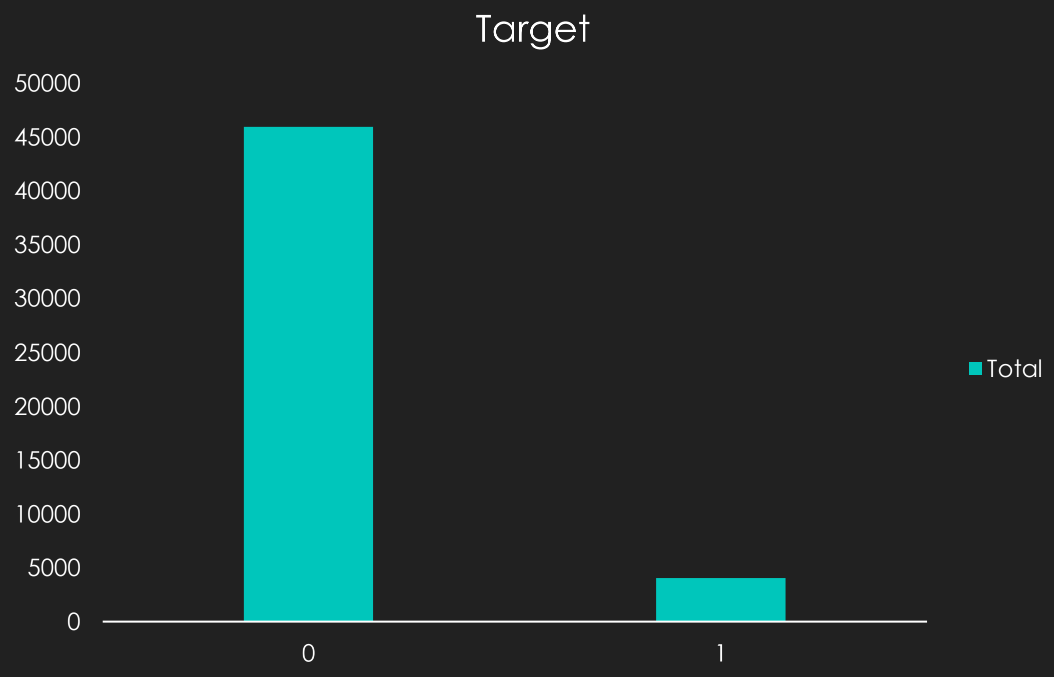
| | | |
|---|---|---|
| NAME CONTRACT TYPEa | NONLIVINGAPARTMENTS AVG | FONDKAPREMONT MODE |
| FLAG OWN CAR | NONLIVINGAREA AVG | HOUSETYPE MODE |
| CNT CHILDREN | APARTMENTS MODE | TOTALAREA MODE |
| AMT GOODS PRICE | BASEMENTAREA MODE | WALLSMATERIAL MODE |
| NAME TYPE SUITE | YEARS BEGINEXPLUATATION MODE | EMERGENCYSTATE MODE |
| NAME INCOME TYPE | YEARS BUILD MODE | OBS 30 CNT SOCIAL CIRCLE |
| REGION POPULATION RELATIVE | COMMONAREA MODE | DEF 30 CNT SOCIAL CIRCLE |
| DAYS EMPLOYED | ELEVATORS MODE | OBS 60 CNT SOCIAL CIRCLE |
| DAYS REGISTRATION | ENTRANCES MODE | DEF 60 CNT SOCIAL CIRCLE |
| OWN CAR AGE | FLOORSMAX MODE | FLAG DOCUMENT 2 |
| FLAG EMAIL | FLOORSMIN MODE | FLAG DOCUMENT 3 |
| REGION RATING CLIENT | LANDAREA MODE | FLAG DOCUMENT 4 |
| REGION RATING CLIENT W CITY | LIVINGAPARTMENTS MODE | FLAG DOCUMENT 5 |
| WEEKDAY APPR PROCESS START | LIVINGAREA MODE | FLAG DOCUMENT 6 |
| HOUR APPR PROCESS START | NONLIVINGAPARTMENTS MODE | FLAG DOCUMENT 7 |
| EXT SOURCE 1 | NONLIVINGAREA MODE | FLAG DOCUMENT 8 |
| EXT SOURCE 2 | APARTMENTS MEDI | FLAG DOCUMENT 9 |
| EXT SOURCE 3 | BASEMENTAREA MEDI | FLAG DOCUMENT 10 |
| APARTMENTS AVG | YEARS BEGINEXPLUATATION MEDI | FLAG DOCUMENT 11 |
| BASEMENTAREA AVG | YEARS BUILD MEDI | FLAG DOCUMENT 12 |
| YEARS BEGINEXPLUATATION AVG | COMMONAREA MEDI | FLAG DOCUMENT 13 |
| YEARS BUILD AVG | ELEVATORS MEDI | FLAG DOCUMENT 14 |
| COMMONAREA AVG | ENTRANCES MEDI | FLAG DOCUMENT 15 |
| ELEVATORS AVG | FLOORSMAX MEDI | FLAG DOCUMENT 16 |
| ENTRANCES AVG | FLOORSMIN MEDI | FLAG DOCUMENT 17 |
| FLOORSMAX AVG | LANDAREA MEDI | FLAG DOCUMENT 18 |
| FLOORSMIN AVG | LIVINGAPARTMENTS MEDI | FLAG DOCUMENT 19 |
| LANDAREA AVG | LIVINGAREA MEDI | FLAG DOCUMENT 20 |
| LIVINGAPARTMENTS AVG | NONLIVINGAPARTMENTS MEDI | FLAG DOCUMENT 21 |
| LIVINGAREA AVG | NONLIVINGAREA MEDI | |

# Null values analysis

| Column Name | SK_ID_CURR | TARGET | CODE_GENDER | FLAG_OWN_REALTY | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | NAME_EDUCATION_TYPE | NAME_FAMILY_STATUS | NAME_HOUSING_TYPE | DAYS_BIRTH | DAYS_ID_PUBLISH | FLAG_MOBIL | FLAG_EMP_PHONE | FLAG_WORK_PHONE | FLAG_CONT_MOBILE | FLAG_PHONE | OCCUPATION_TYPE | CNT_FAM_MEMBERS | REG_REGION_NOT_LIVE_REGION | REG_REGION_NOT_WORK_REGION | LIVE_REGION_NOT_WORK_REGION | REG_CITY_NOT_LIVE_CITY | REG_CITY_NOT_WORK_CITY | LIVE_CITY_NOT_WORK_CITY | ORGANIZATION_TYPE | DAYS_LAST_PHONE_CHANGE | AMT_REQ_CREDIT_BUREAU_HOUR | AMT_REQ_CREDIT_BUREAU_DAY | AMT_REQ_CREDIT_BUREAU_WEEK | AMT_REQ_CREDIT_BUREAU_MON | AMT_REQ_CREDIT_BUREAU_QRT | AMT_REQ_CREDIT_BUREAU_YEAR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Null Value %age | 0 | 0 | 0 | 0 | 0 | 0 | 0.002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31.3086 | 0.002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.002 | 13.4683 | 13.4683 | 13.4683 | 13.4683 | 13.4683 | 13.4683 |

Occupation_type column has been dropped due to the presence of > 30% null values

# Data Imbalance




We have data imbalance in our dataset, hence, the results would be skewed, we cannot generalize the findings of target 1 as it only represents 8% of our sample, rest 92% represents target 0.

# Feature Engineering

1. AMT_REQ_CREDIT_BUREAU_YEAR_Total – Following 6 columns were clubbed to create a new column

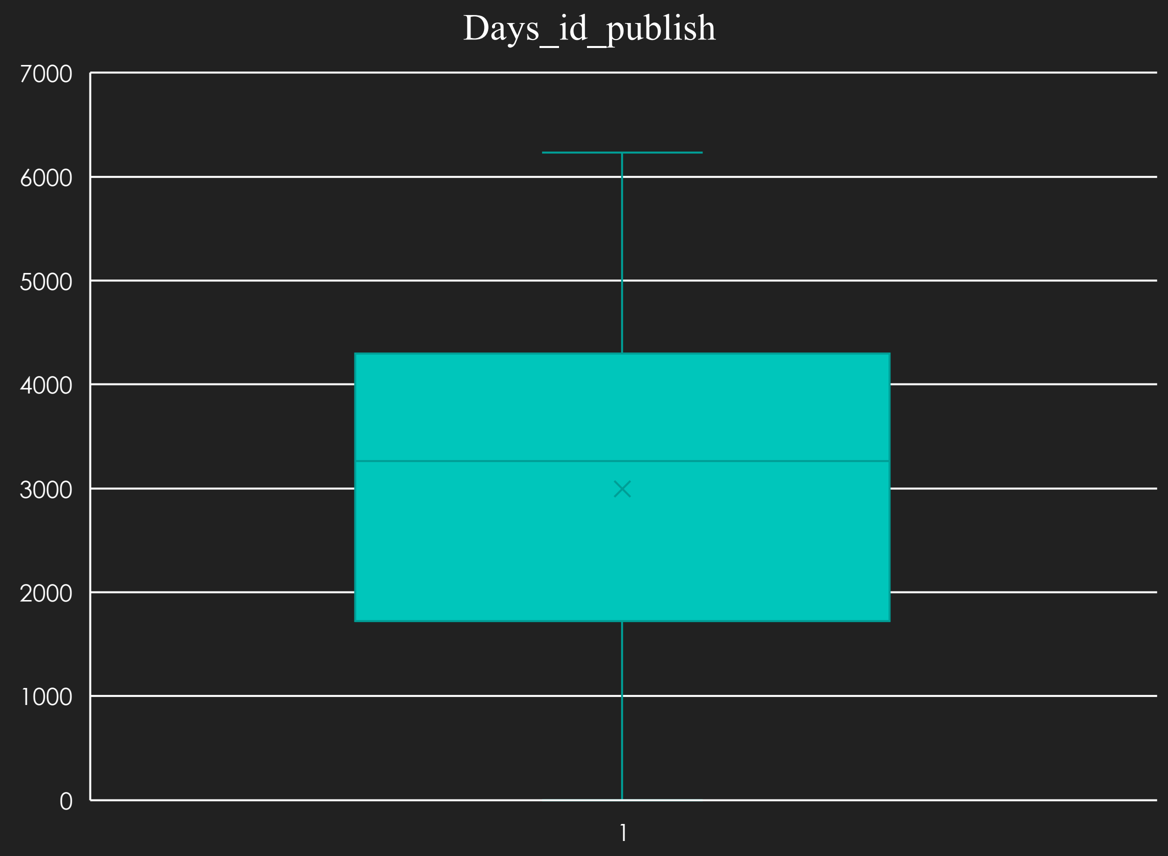
- AMT_REQ_CREDIT_BUREAU_HOUR
- AMT_REQ_CREDIT_BUREAU_DAY
- AMT_REQ_CREDIT_BUREAU_WEEK
- AMT_REQ_CREDIT_BUREAU_MON
- AMT_REQ_CREDIT_BUREAU_QRT
- AMT_REQ_CREDIT_BUREAU_YEAR

2. DAYS_ID_PUBLISH – Values were converted into positive values through absolute function.

3. Days_Birth – The column values were converted into positive year values.

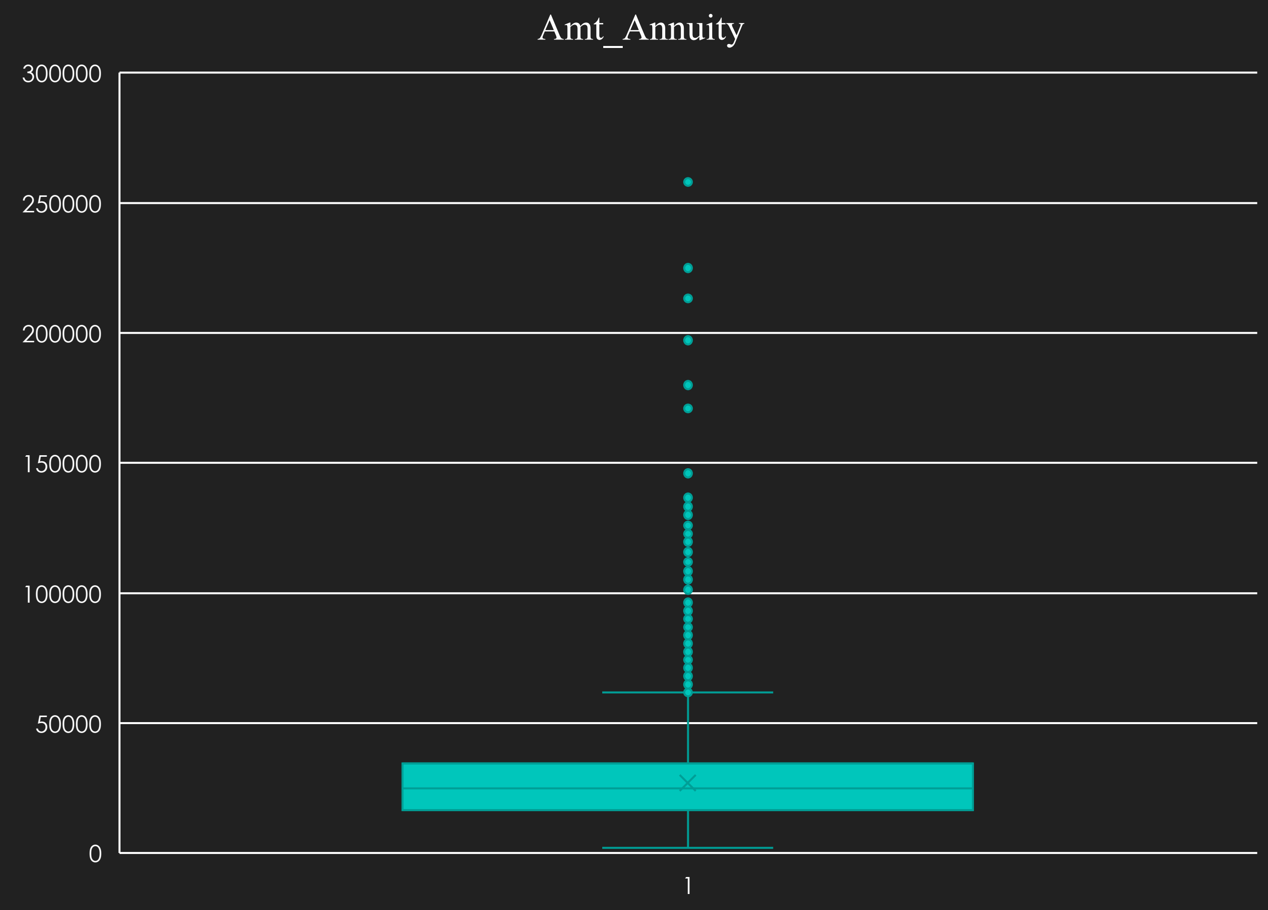
4. Days_last_phone_change - The column values were converted into positive values.

# Data distribution




The data contains no outliers and no null values.

# Data distribution

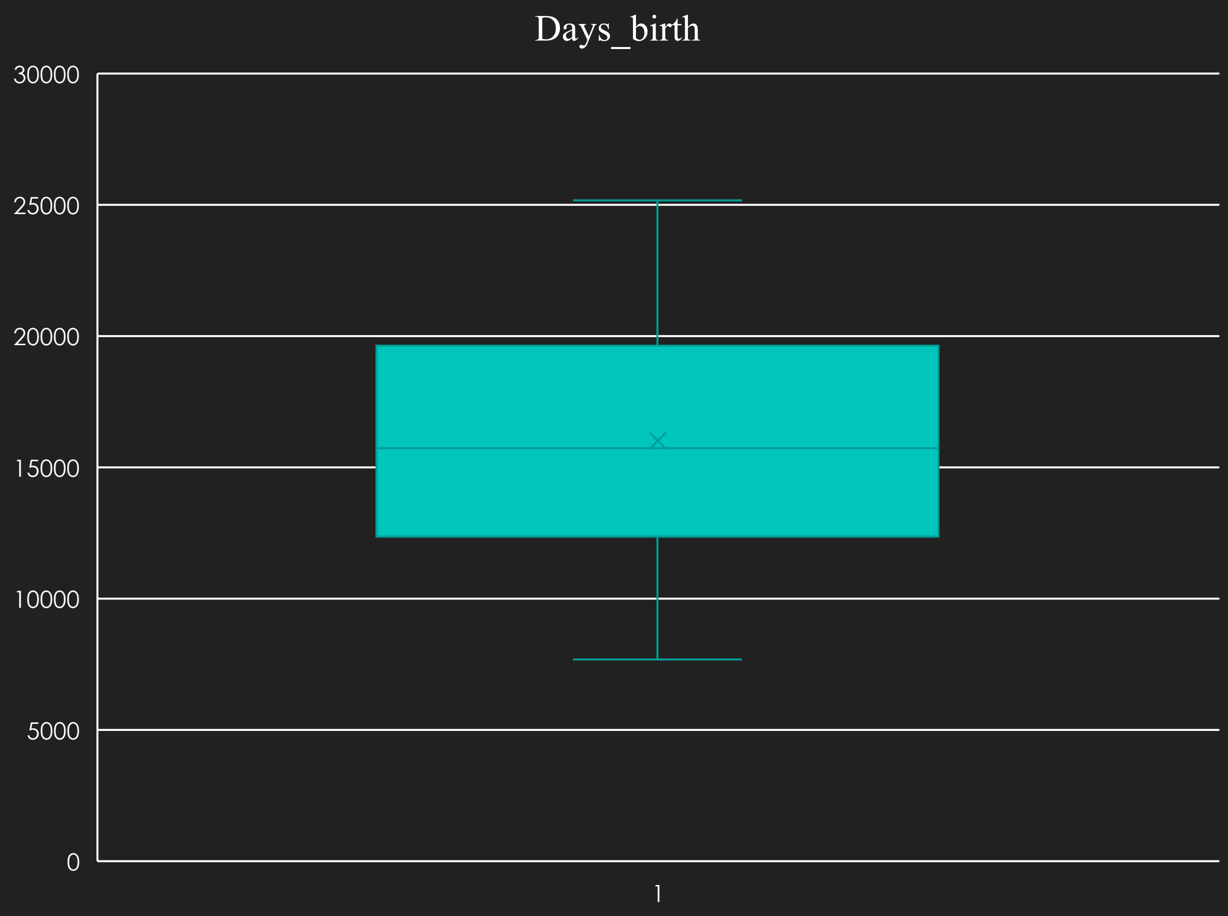



| Quartiles | AMT_ANNUITY |
|---|---|
| Q1 | 112500 |
| Q2 (median) | 146250 |
| Q3 | 202500 |
| Q4 | 117000000 |
| Q3-Q1 | 90000 |
| Lower Bound | -22500 |
| Upper Bound | 337500 |

Null values in Amt_Annuity column have been replaced with median value due to the presence of outliers in our data.
Median = 24939

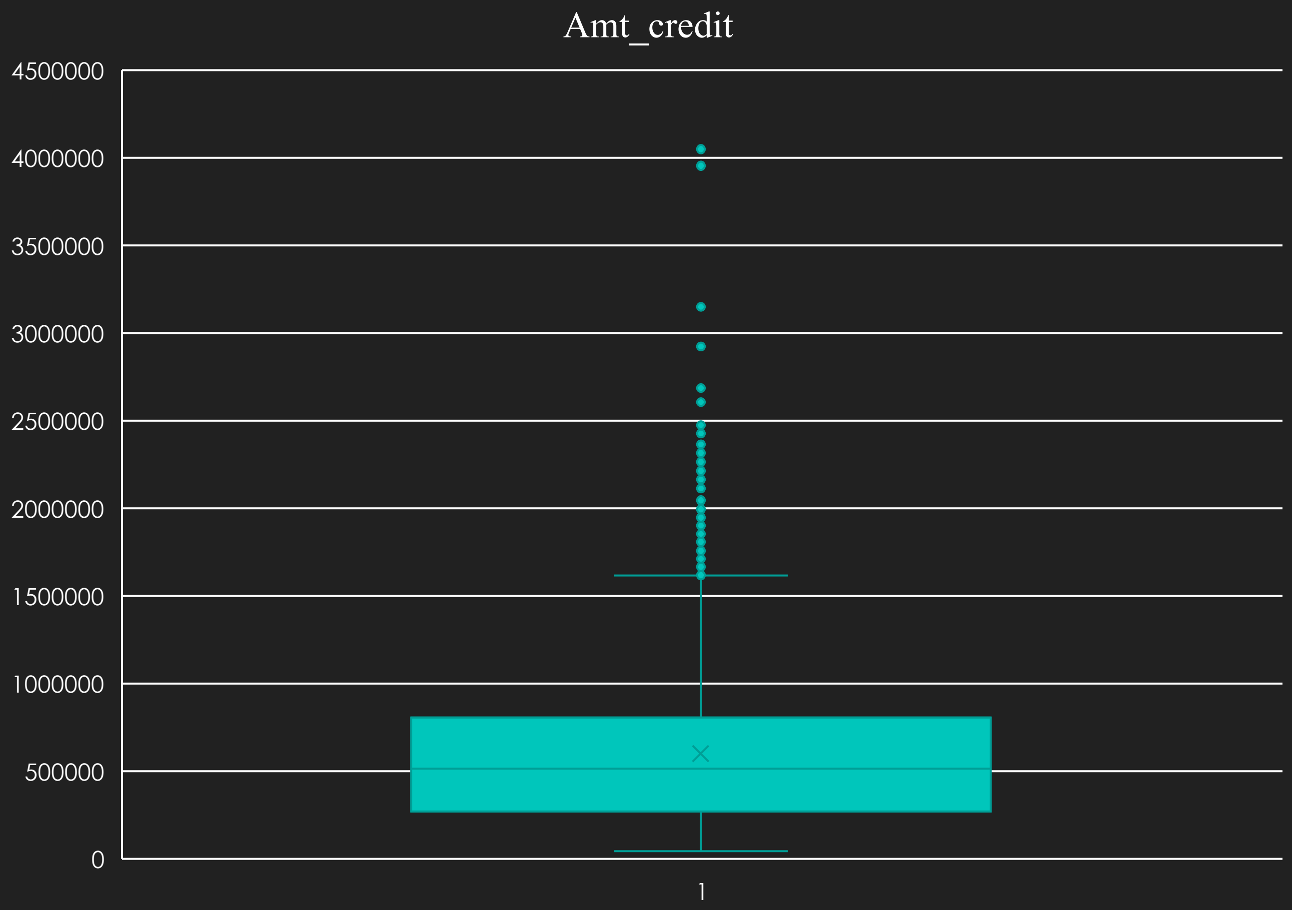
The values beyond lower and upper bound were identified as outliers. But these values were not replaced as extremes in annuity are possible.

# Data distribution




The data contains no outlier and no null value.

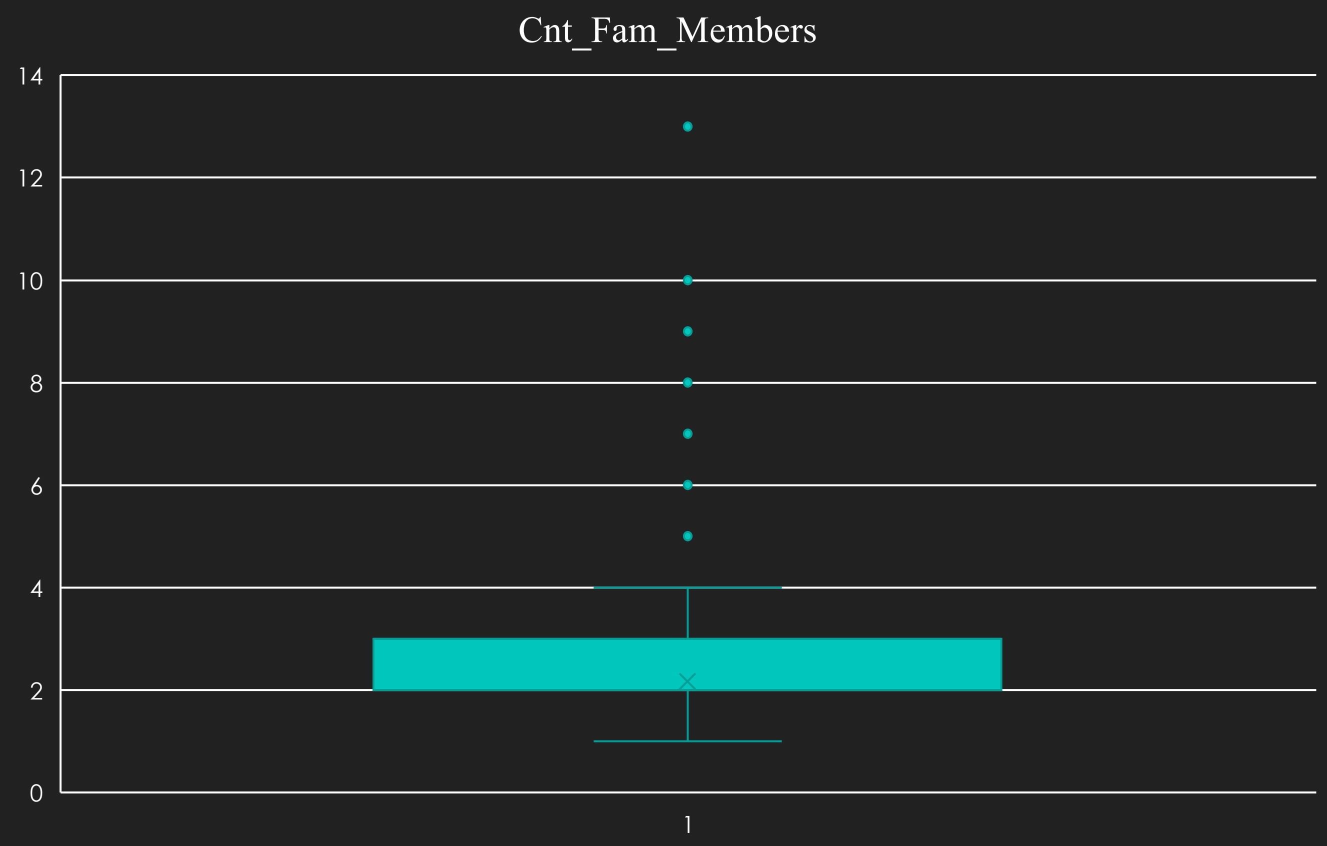
# Data distribution




The dataset contains outliers but replacing these data points is not an ideal choice as credit amount varies highly for customers.

Also, no null values were found.

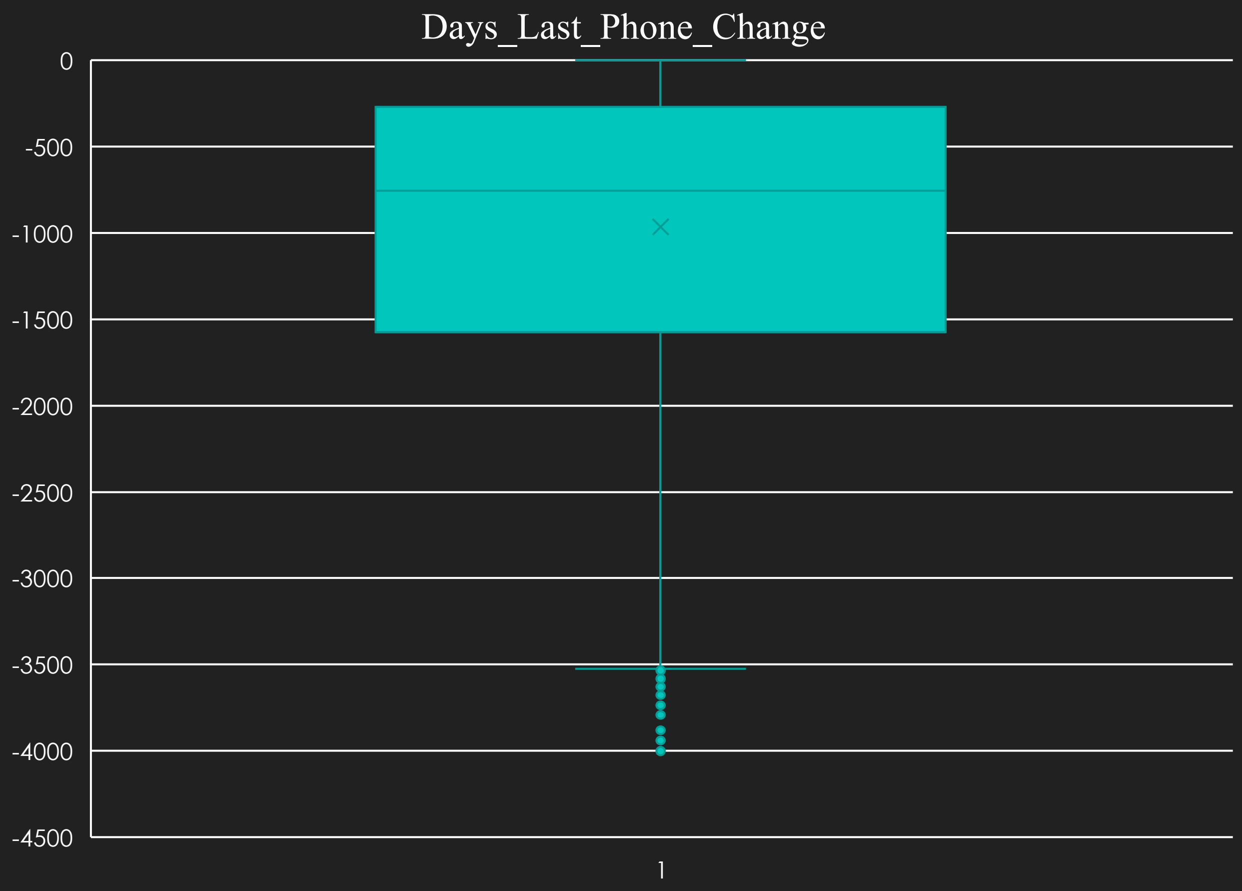
# Data distribution




Null values in Cnt_Fam_Members column have been replaced with median value due to the presence of outliers in our data.
Median = 2

But outliers have not been replaced as the highest value in the dataset is 13 which is possible. A family can have 13 members.

# Data distribution




| Quartiles | DAYS_LAST_PHONE_CHANGE |
|---|---|
| Q1 | 270 |
| Q2 (median) | 755 |
| Q3 | 1573 |
| Q4 | 4002 |
| Q3-Q1 | 1303 |
| Lower Bound | -1684.5 |
| Upper Bound | 3527.5 |

- Null values in Days_Last_Phone_Change column have been replaced with median value due to the presence of outliers in our data.
- The data points were converted into positive values with ABS function.
- Outliers have also been replaced with median value. Values beyond lower and upper bound are outliers.

# Attribute analysis

## General Analysis

bifurcation on the basis of

### Target

**0**

- CODE_GENDER
- FLAG_OWN_REALTY
- NAME_EDUCATION_TYPE
- NAME_FAMILY_STATUS
- NAME_HOUSING_TYPE
- DAYS_BIRTH
- CNT_FAM_MEMBERS
- ORGANIZATION_TYPE
- AMT_REQ_CREDIT_BUREAU_YEAR_Total

**1**

- CODE_GENDER
- FLAG_OWN_REALTY
- NAME_EDUCATION_TYPE
- NAME_FAMILY_STATUS
- NAME_HOUSING_TYPE
- DAYS_BIRTH
- CNT_FAM_MEMBERS
- ORGANIZATION_TYPE
- AMT_REQ_CREDIT_BUREAU_YEAR_Total

## Fraud detection

### Target

**0**

- DAYS_LAST_PHONECHANGE
- DAYS_ID_PUBLISH
- FLAG_CONT_MOBILE

**1**

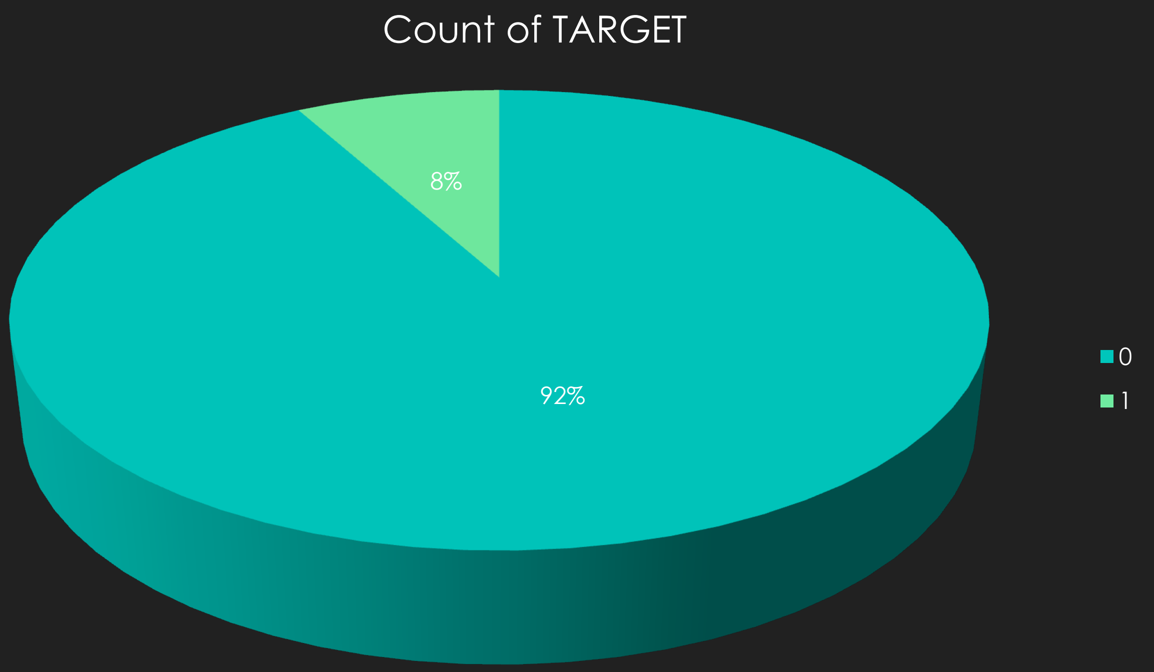
- DAYS_LAST_PHONECHANGE
- DAYS_ID_PUBLISH
- FLAG_CONT_MOBILE

# Univariate Analysis

# TARGET

| Row Labels | Count of TARGET |
|---|---|
| 0 | 45973 |
| 1 | 4026 |
| Grand Total | 49999 |

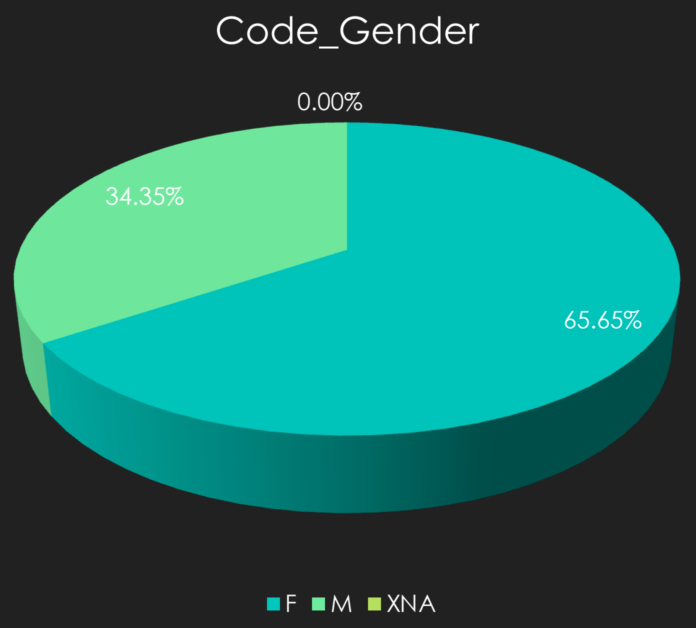
| Label | Representation |
|---|---|
| 0 | Customers with no payment difficulties |
| 1 | Customers with payment difficulties |




Most of the customers had no payment difficulties, only 8% customers made late payment of installment

# CODE_GENDER

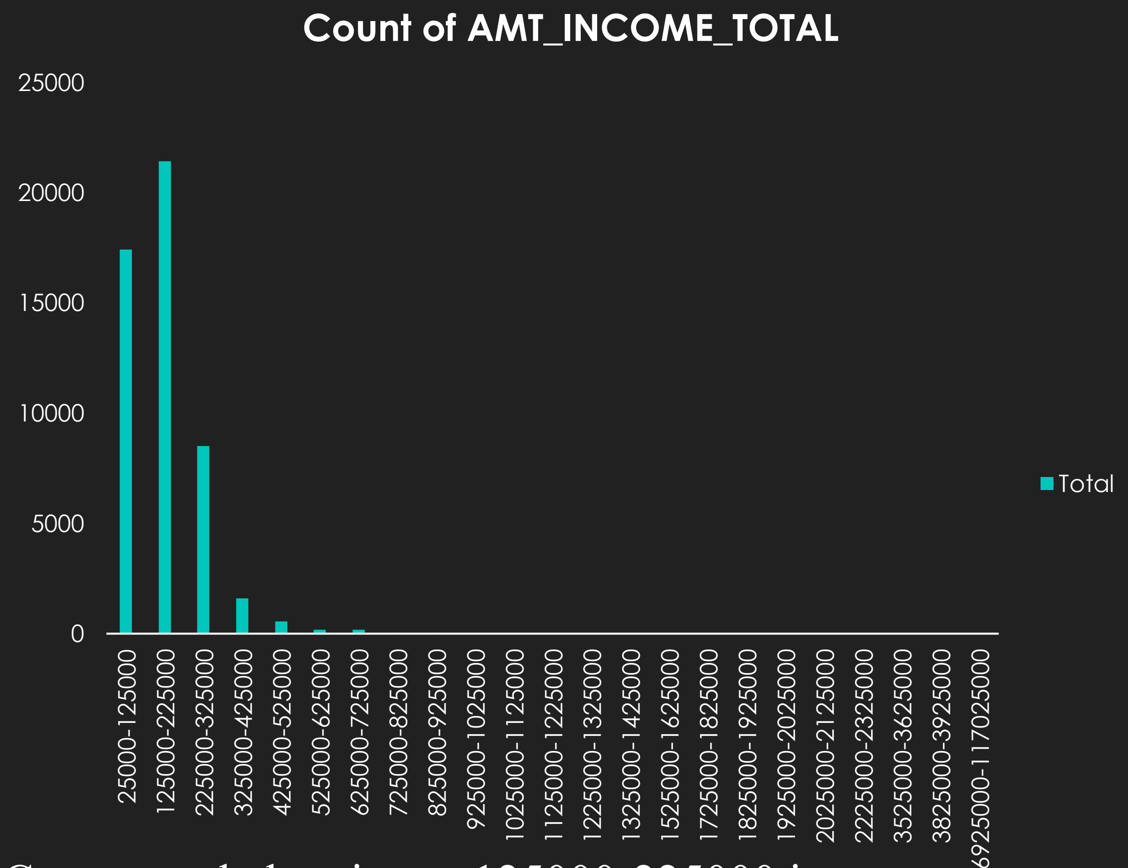
| Row Labels | Count of CODE_GENDER | Count of CODE_GENDER (%) |
|---|---|---|
| F | 32823 | 65.65% |
| M | 17174 | 34.35% |
| XNA | 2 | 0.00% |
| Grand Total | 49999 | 100.00% |




Approximately 65% of bank’s customers are females and rest are males.

# AMT_INCOME_TOTAL

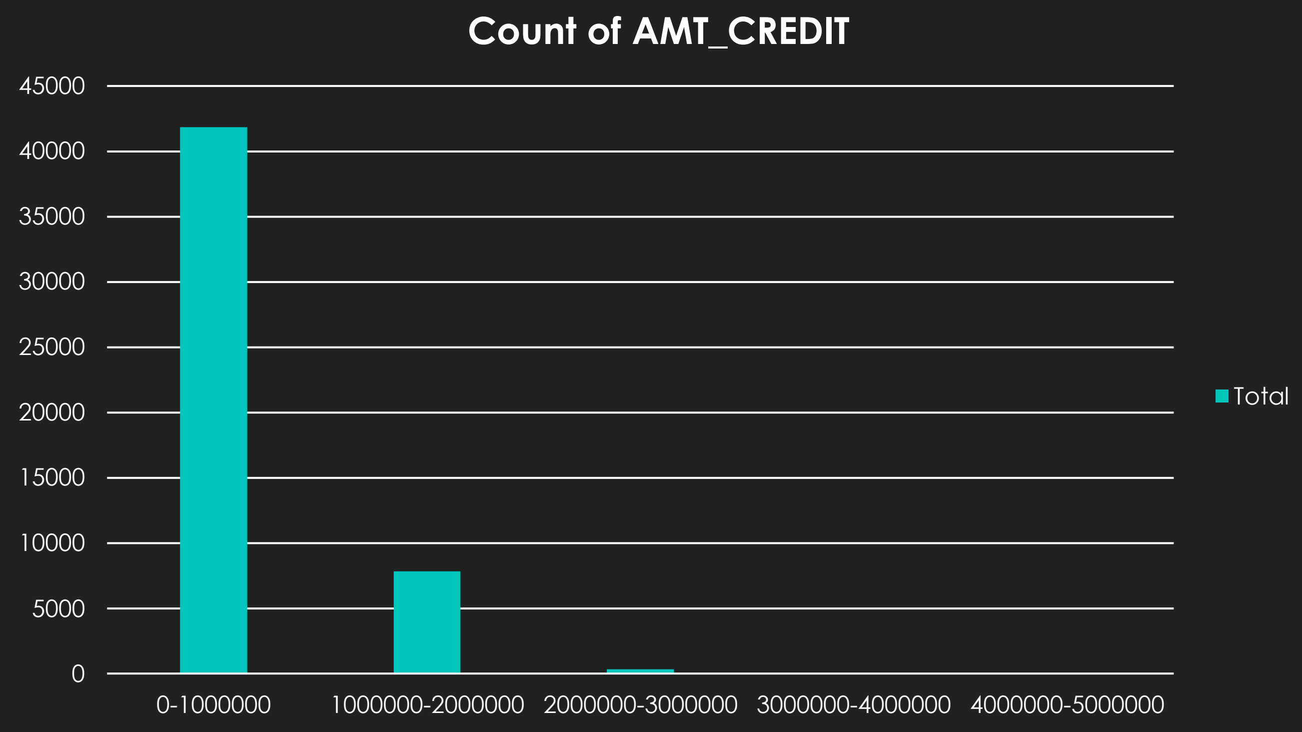
| Row Labels | Count of AMT_INCOME_TOTAL |
|---|---|
| 25000-125000 | 17440 |
| 125000-225000 | 21447 |
| 225000-325000 | 8506 |
| 325000-425000 | 1592 |
| 425000-525000 | 569 |
| 525000-625000 | 169 |
| 625000-725000 | 165 |
| 725000-825000 | 34 |
| 825000-925000 | 35 |
| 925000-1025000 | 2 |
| 1025000-1125000 | 4 |
| 1125000-1225000 | 13 |
| 1225000-1325000 | 1 |
| 1325000-1425000 | 10 |
| 1525000-1625000 | 1 |
| 1725000-1825000 | 2 |
| 1825000-1925000 | 1 |
| 1925000-2025000 | 1 |
| 2025000-2125000 | 2 |
| 2225000-2325000 | 2 |
| 3525000-3625000 | 1 |
| 3825000-3925000 | 1 |
| 116925000-117025000 | 1 |
| Grand Total | 49999 |




Customers belonging to 125000-225000 income range are highest applicants of credit and as the income increases, the less customers apply for credit.

# AMT_CREDIT

| Row Labels | Count of AMT_CREDIT |
|---|---|
| 0-1000000 | 41853 |
| 1000000-2000000 | 7823 |
| 2000000-3000000 | 319 |
| 3000000-4000000 | 2 |
| 4000000-5000000 | 2 |
| Grand Total | 49999 |

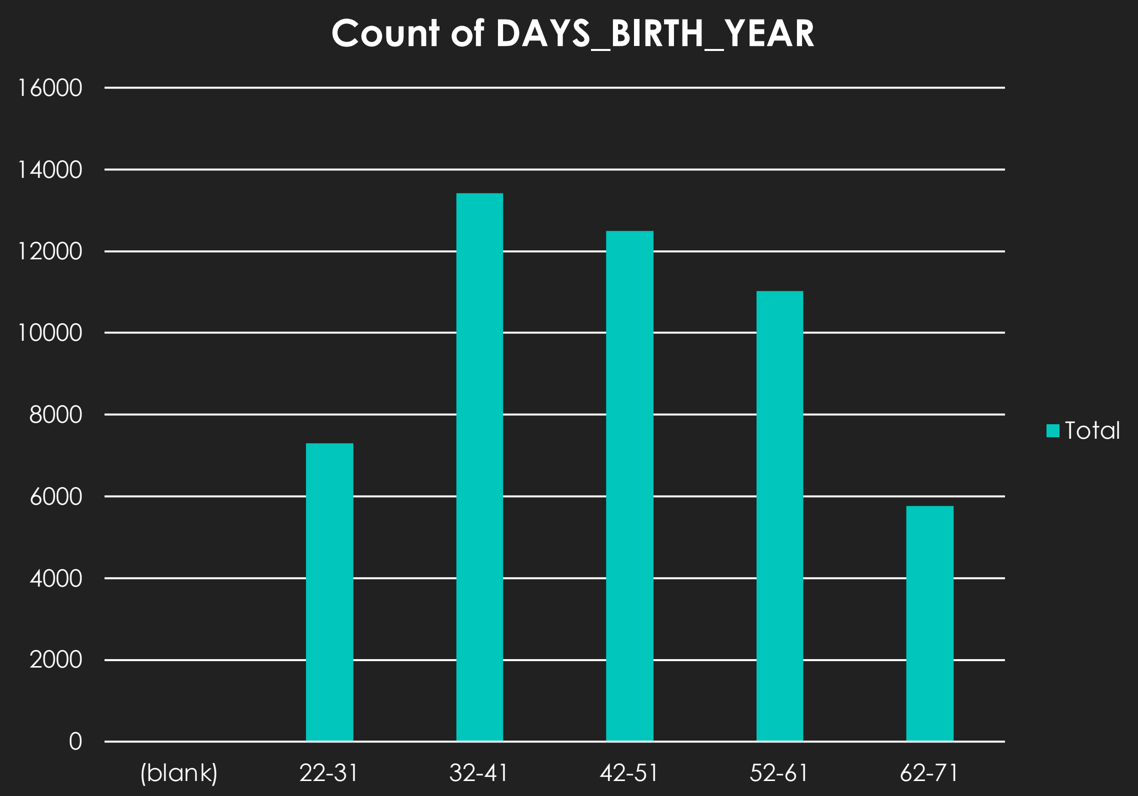



Smaller value loans of upto Rs. 100000 are mostly applied for.

# DAYS_BIRTH_YEAR

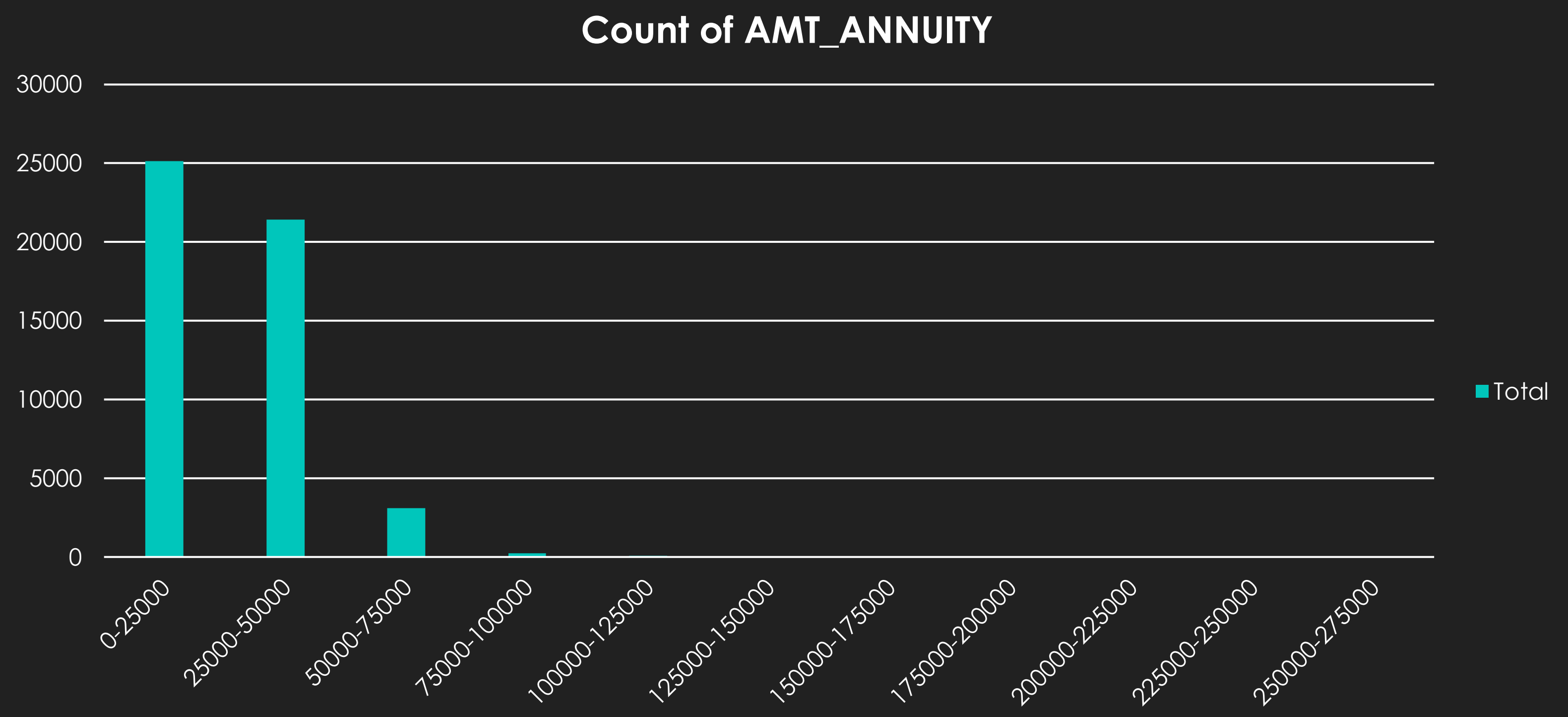
| Row Labels | Count of DAYS_BIRTH_year |
|---|---|
| 22-31 | 7302 |
| 32-41 | 13422 |
| 42-51 | 12489 |
| 52-61 | 11020 |
| 62-71 | 5766 |
| Grand Total | 49999 |

Most of the applicants belong to the 32-51 years age bracket.



# AMT_ANNUITY

| Row Labels | Count of AMT_ANNUITY |
|---|---|
| 0-25000 | 25128 |
| 25000-50000 | 21434 |
| 50000-75000 | 3107 |
| 75000-100000 | 245 |
| 100000-125000 | 55 |
| 125000-150000 | 17 |
| 150000-175000 | 4 |
| 175000-200000 | 2 |
| 200000-225000 | 1 |
| 225000-250000 | 5 |
| 250000-275000 | 1 |
| Grand Total | 49999 |



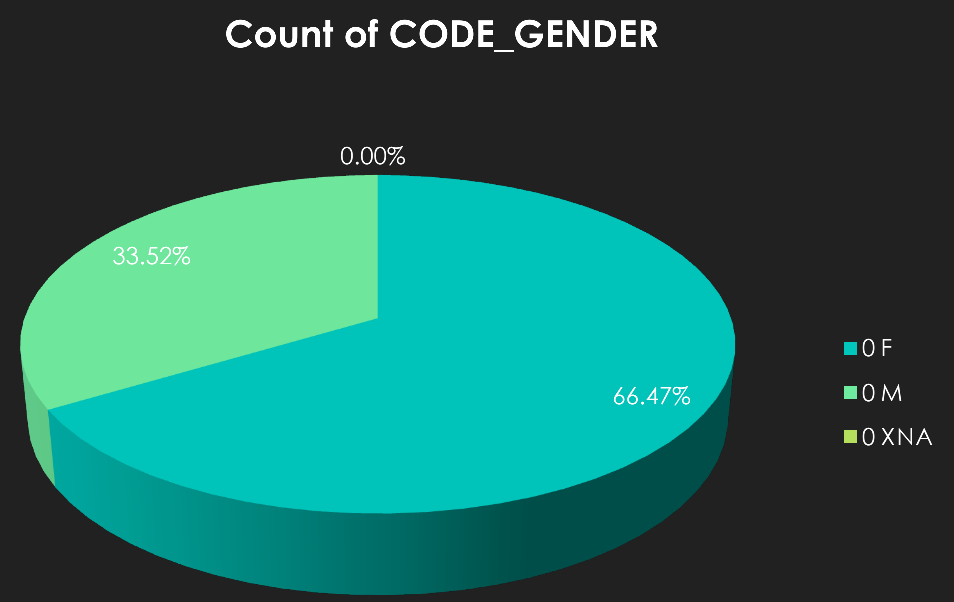

Most of the annuities are lower in value and annuities beyond Rs. 150000 are very less in number.

# Segmented Univariate Analysis

# 0 – Customers with no payment difficulties

# CODE_GENDER

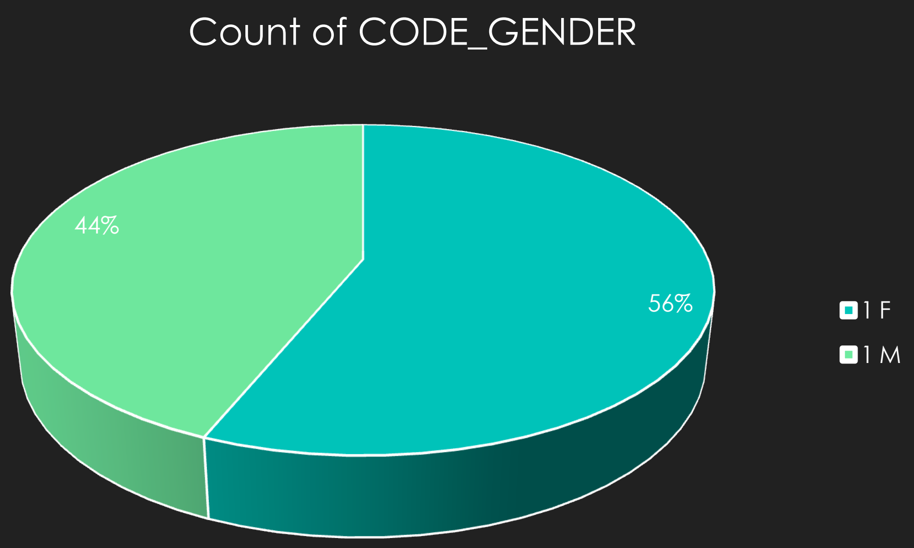
| Row Labels 0 | Count of CODE_GENDER | Count of CODE_GENDER(%) |
|---|---|---|
| F | 30559 | 66.47% |
| M | 15412 | 33.52% |
| XNA | 2 | 0.00% |
| Grand Total | 45973 | 100.00% |




Approximately 66% of customers who faced no payment difficulties were females.

# 1 – Customers with payment difficulties CODE_GENDER

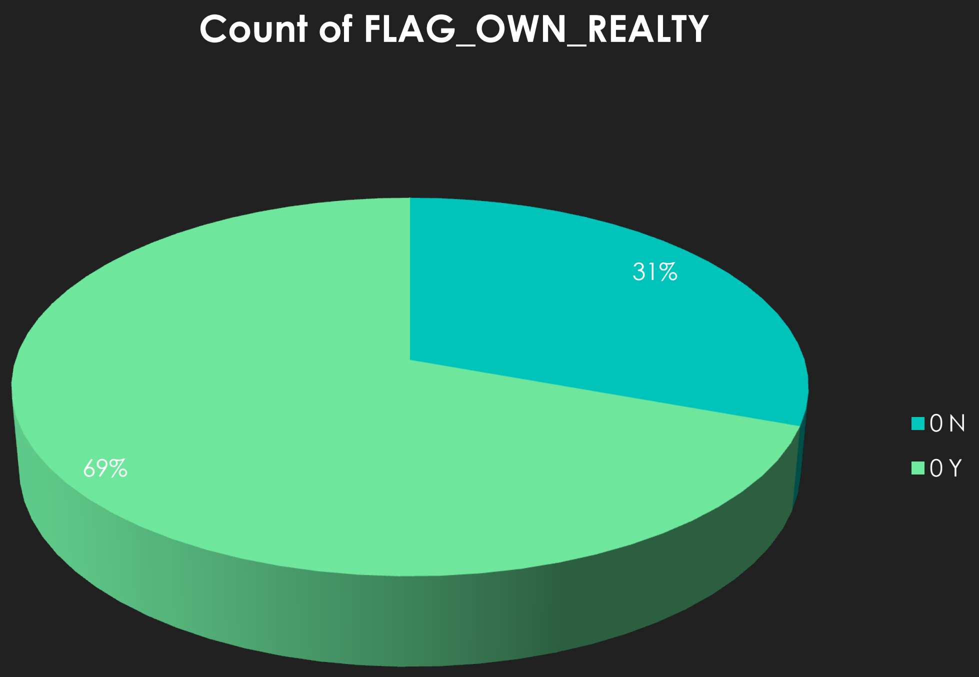
| Row Labels 1 | Count of CODE_GENDER | Count of CODE_GENDER(%) |
|---|---|---|
| F | 2264 | 56.23% |
| M | 1762 | 43.77% |
| Grand Total | 4026 | 100.00% |




Females faced more payment difficulties in comparison to males.

# 0 – Customers with no payment difficulties FLAG_OWN_REALTY

| Row Labels (0) | Count of FLAG_OWN_REALTY |
|---|---|
| N | 14034 |
| Y | 31939 |
| Grand Total | 45973 |

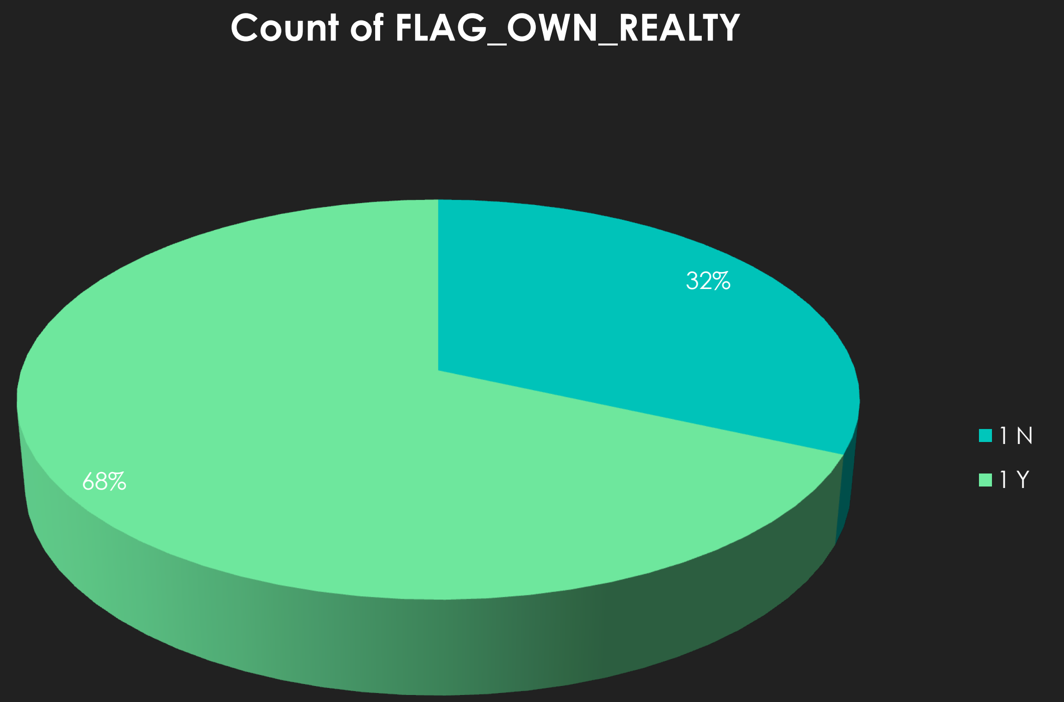



69% of customers who had no payment issues had own property

# 1 – Customers with payment difficulties FLAG_OWN_REALTY

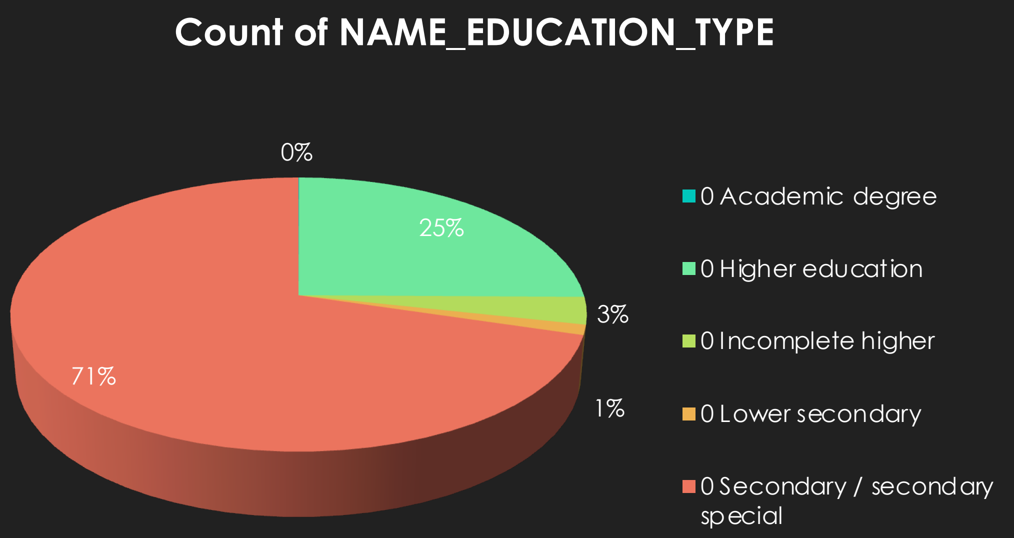
| Row Labels (1) | Count of FLAG_OWN_REALTY |
|---|---|
| N | 1274 |
| Y | 2752 |
| Grand Total | 4026 |

68% of customers who had payment difficulties had own property.



# 0 – Customers with no payment difficulties NAME_EDUCATION_TYPE

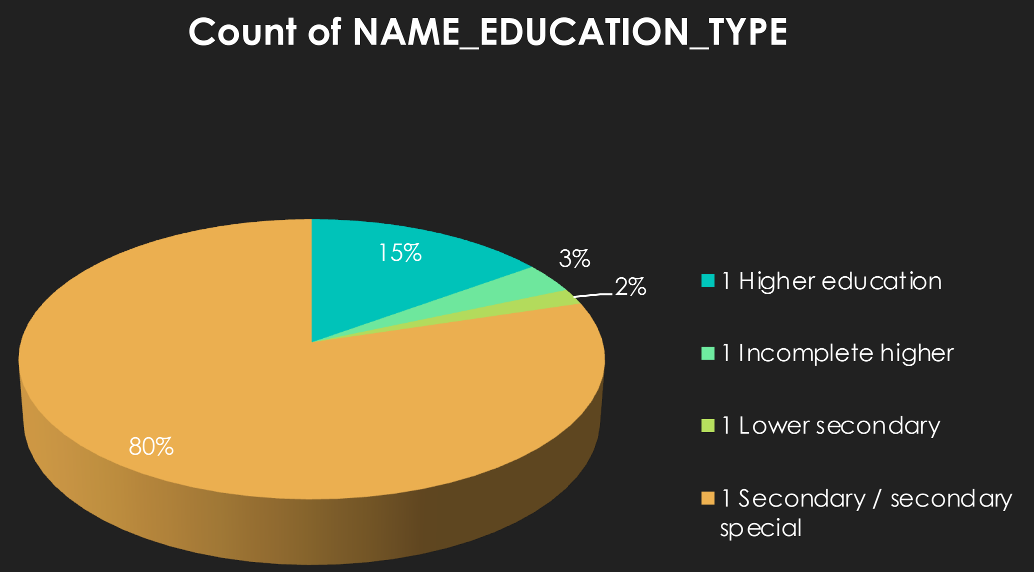
| Row Labels | Count of NAME_EDUCATION_TYPE |
|---|---|
| Academic degree | 20 |
| Higher education | 11561 |
| Incomplete higher | 1482 |
| Lower secondary | 547 |
| Secondary / secondary special | 32363 |
| Grand Total | 45973 |




71% of the customers who faced no payment difficulties had education level upto secondary/ secondary special.

# 1 – Customers with payment difficulties NAME_EDUCATION_TYPE

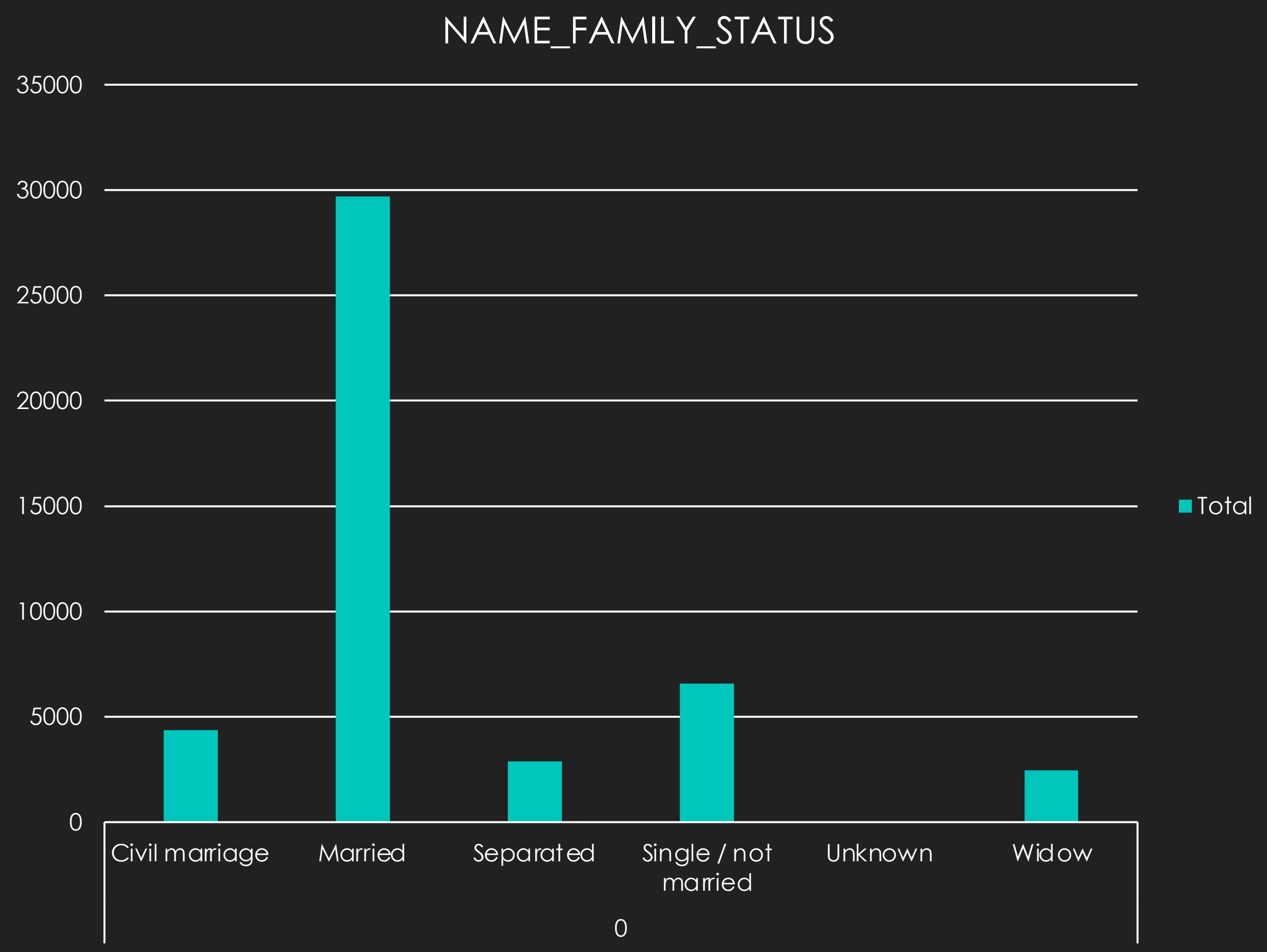
| Row Labels | Count of NAME_EDUCATION_TYPE |
|---|---|
| Higher education | 606 |
| Incomplete higher | 138 |
| Lower secondary | 73 |
| Secondary / secondary special | 3209 |
| Grand Total | 4026 |




80% of the customers who faced payment difficulties had education level upto secondary/ secondary special.

# 0 – Customers with no payment difficulties NAME_FAMILY_STATUS

| Row Labels | Count of NAME_FAMILY_STATUS |
|---|---|
| Civil marriage | 4377 |
| Married | 29699 |
| Separated | 2870 |
| Single / not married | 6577 |
| Unknown | 1 |
| Widow | 2449 |
| Grand Total | 45973 |

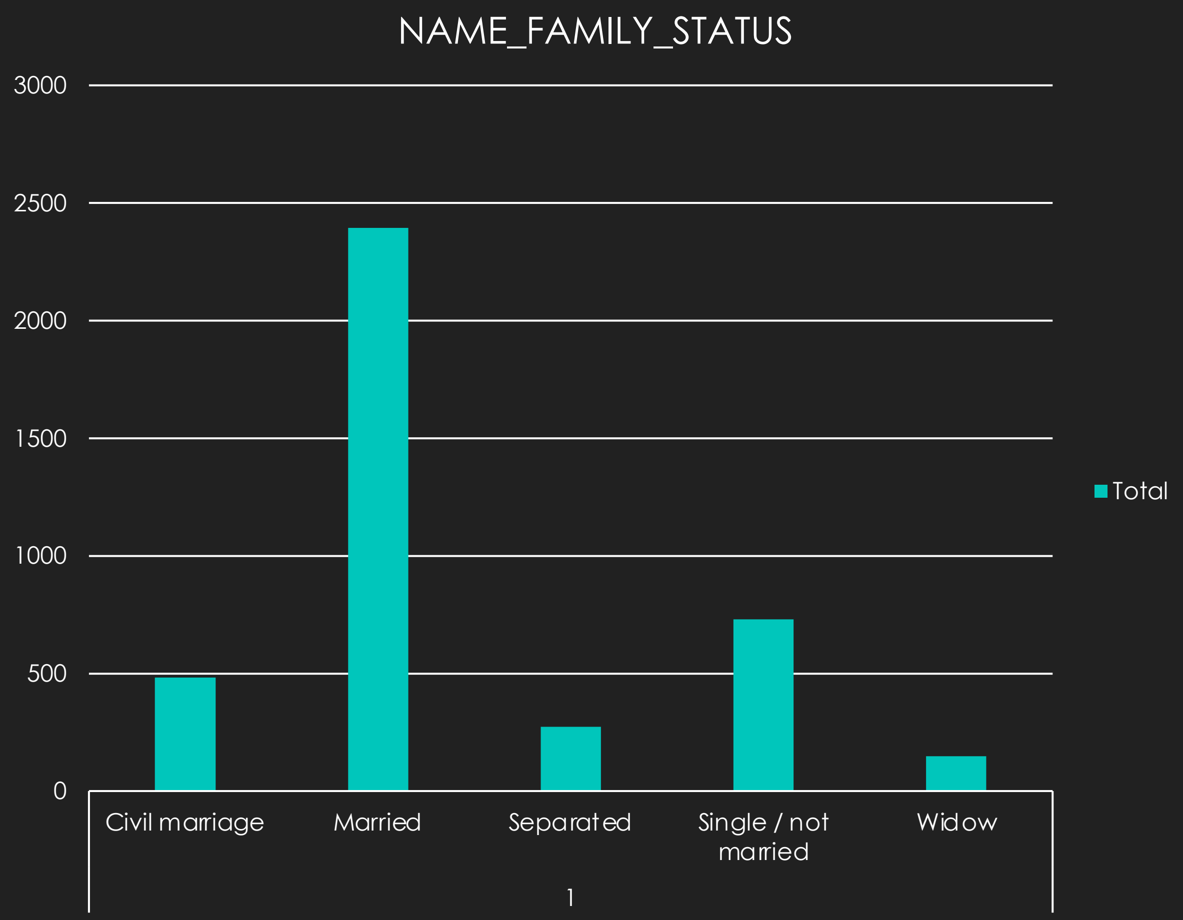



Most of the customers who did not face payment difficulties were married.

# 1 – Customers with payment difficulties NAME_FAMILY_STATUS

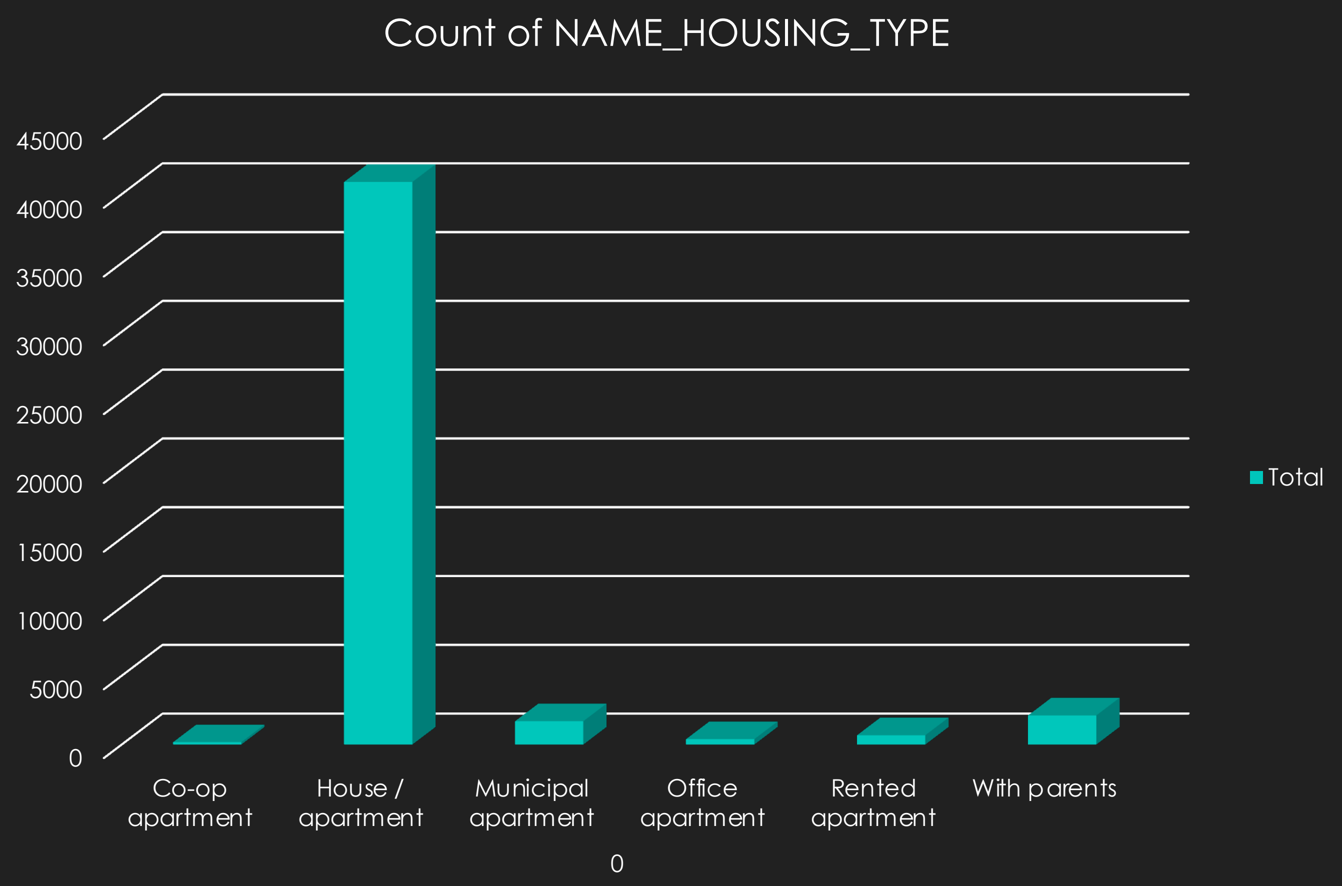
| Row Labels | Count of NAME_FAMILY_STATUS |
|---|---|
| Civil marriage | 482 |
| Married | 2395 |
| Separated | 272 |
| Single / not married | 729 |
| Widow | 148 |
| Grand Total | 4026 |

Most of the customers who faced payment difficulties were also married.



# 0 – Customers with no payment difficulties NAME_HOUSING_TYPE

| Row Labels | Count of NAME_HOUSING_TYPE |
|---|---|
| Co-op apartment | 176 |
| House / apartment | 40895 |
| Municipal apartment | 1700 |
| Office apartment | 398 |
| Rented apartment | 682 |
| With parents | 2122 |
| Grand Total | 45973 |

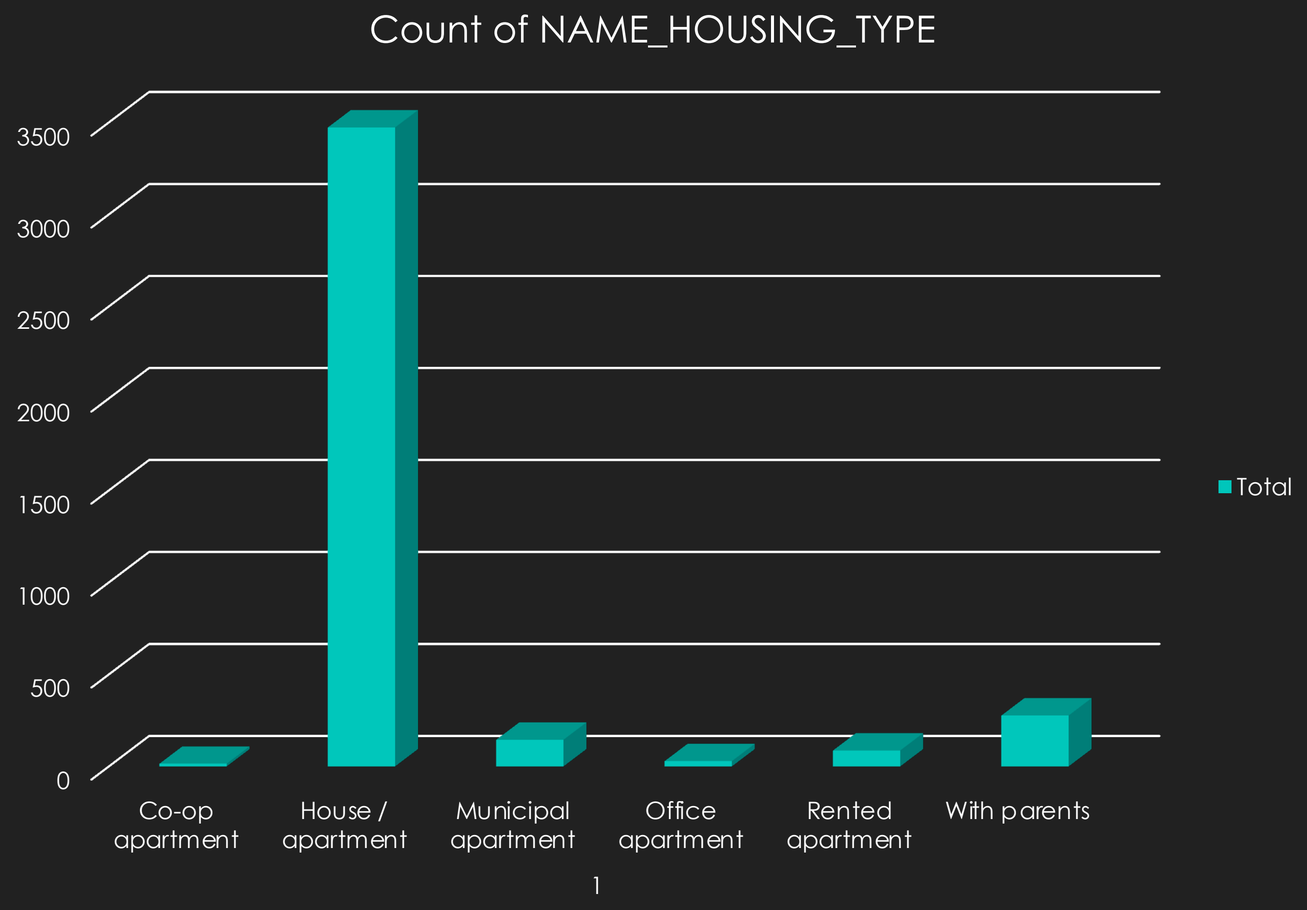



Most of the customers who faced no payment difficulties lived in own house/ apartment.

# 1 – Customers with payment difficulties NAME_HOUSING_TYPE

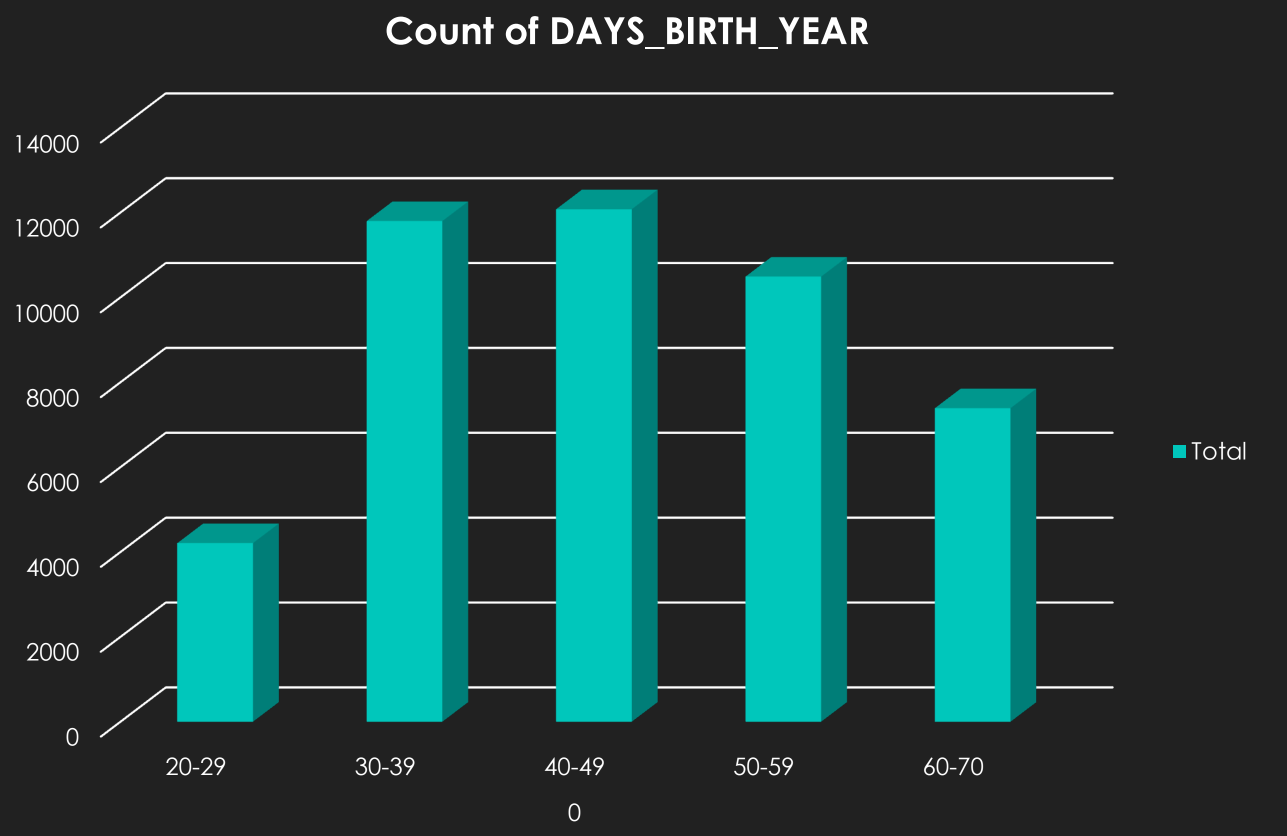
| Row Labels | Count of NAME_HOUSING_TYPE |
|---|---|
| Co-op apartment | 15 |
| House / apartment | 3473 |
| Municipal apartment | 145 |
| Office apartment | 29 |
| Rented apartment | 87 |
| With parents | 277 |
| Grand Total | 4026 |

Most of the customers who faced payment difficulties also lived in own house/ apartment.



# 0 – Customers with no payment difficulties DAYS_BIRTH_YEAR

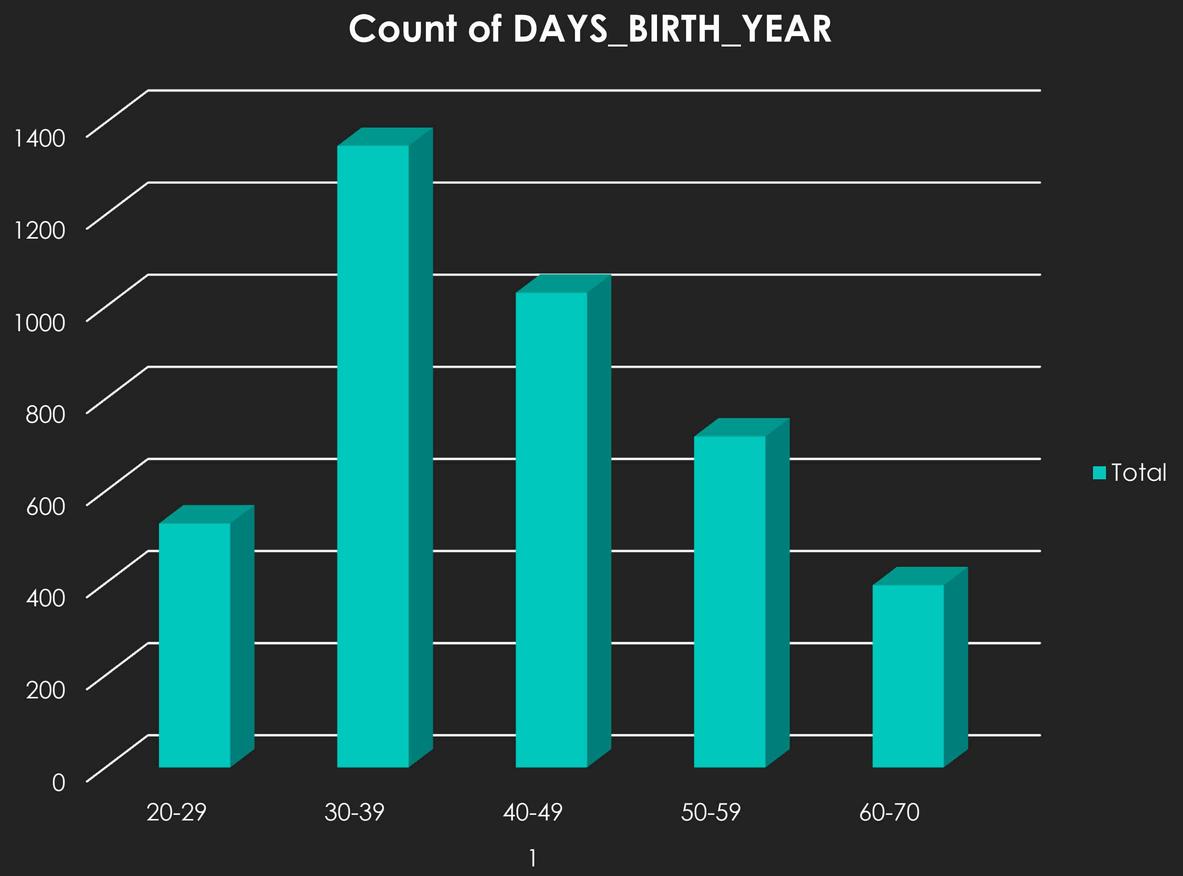
| Row Labels | Count of SK_ID_CURR |
|---|---|
| 20-29 | 4211 |
| 30-39 | 11801 |
| 40-49 | 12079 |
| 50-59 | 10491 |
| 60-70 | 7391 |
| Grand Total | 45973 |




Most of the customers with no payment difficulties belonged to 40-49 years.

# 1 – Customers with payment difficulties DAYS_BIRTH_YEAR

| Row Labels | Count of SK_ID_CURR |
|---|---|
| 20-29 | 530 |
| 30-39 | 1350 |
| 40-49 | 1031 |
| 50-59 | 719 |
| 60-70 | 396 |
| Grand Total | 4026 |

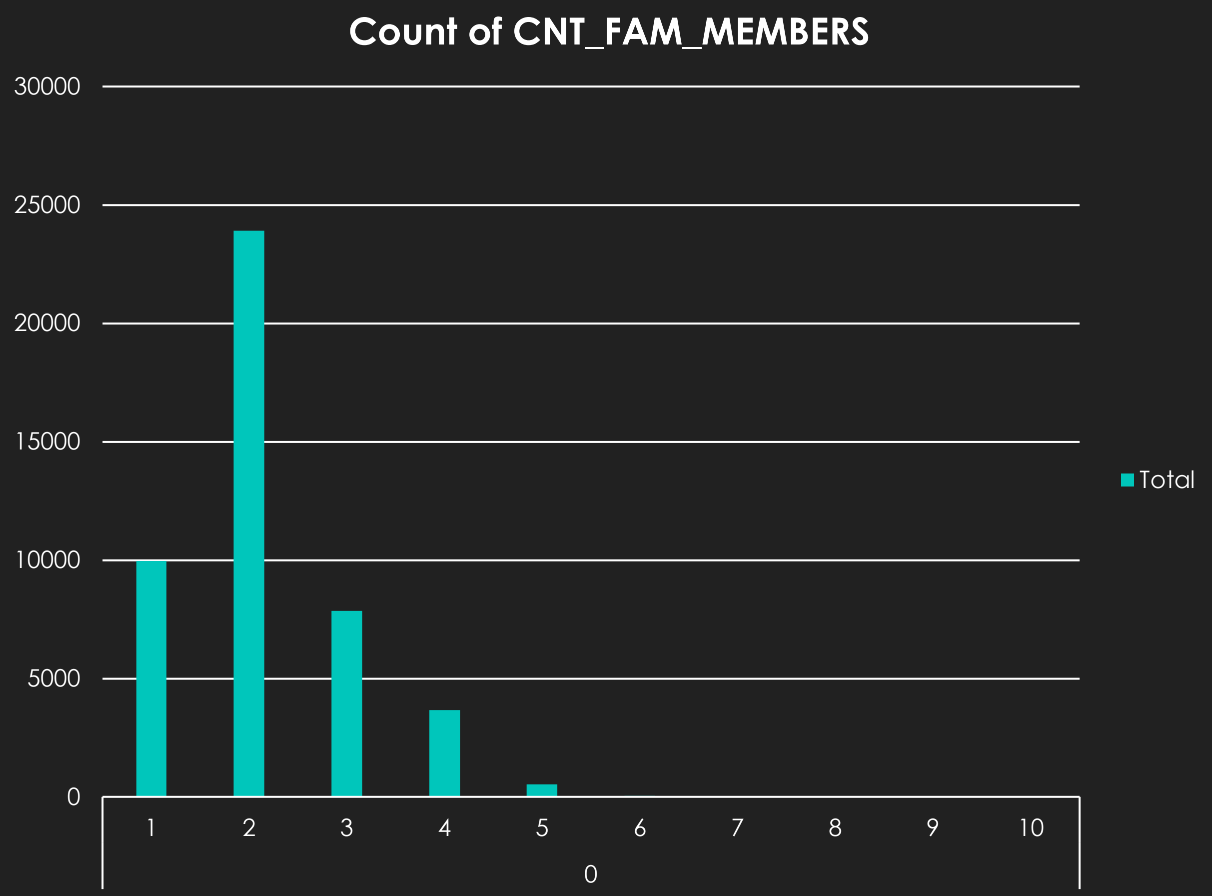



Most of the customers with payment difficulties belonged to 30-39 years.

# 0 – Customers with no payment difficulties CNT_FAM_MEMBERS

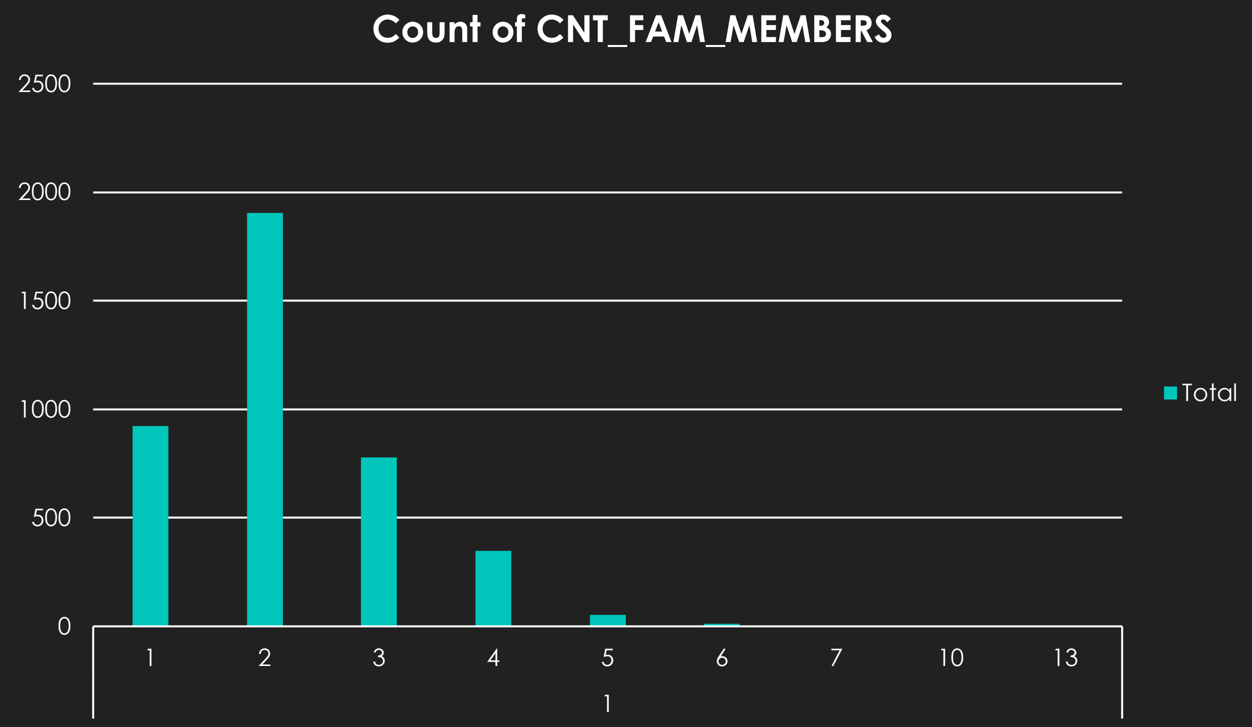
| Row Labels | Count of CNT_FAM_MEMBERS |
|---|---|
| 1 | 9951 |
| 2 | 23902 |
| 3 | 7858 |
| 4 | 3651 |
| 5 | 538 |
| 6 | 55 |
| 7 | 9 |
| 8 | 6 |
| 9 | 2 |
| 10 | 1 |
| Grand Total | 45973 |

Most of the customers with no payment difficulties had 2 members in family.



# 1 – Customers with payment difficulties CNT_FAM_MEMBERS

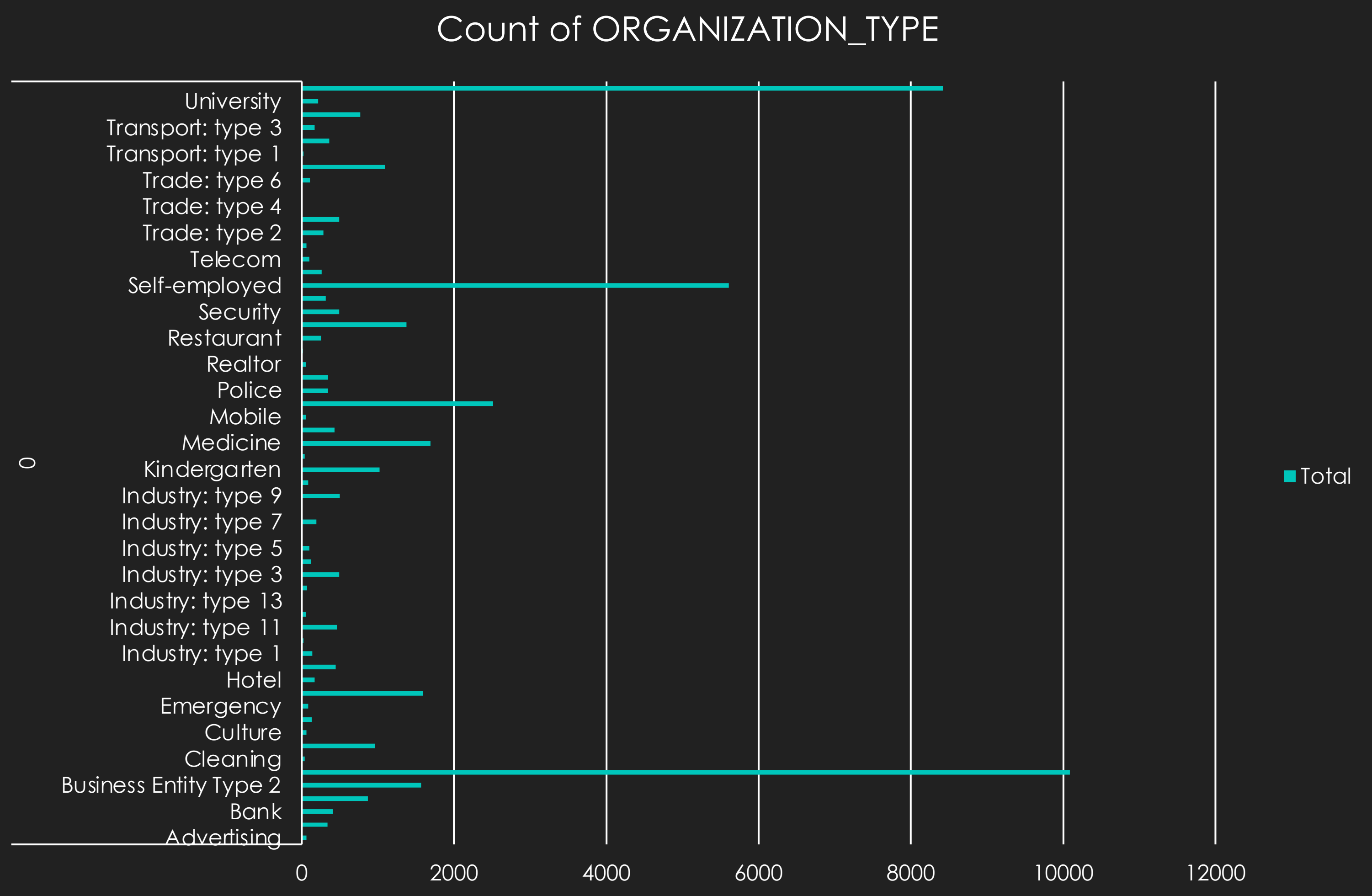
| Row Labels | Count of CNT_FAM_MEMBERS |
|---|---|
| 1 | 4026 |
| 1 | 922 |
| 2 | 1906 |
| 3 | 777 |
| 4 | 349 |
| 5 | 54 |
| 6 | 13 |
| 7 | 3 |
| 10 | 1 |
| 13 | 1 |
| Grand Total | 4026 |




Most of the customers with payment difficulties also had 2 members in family.

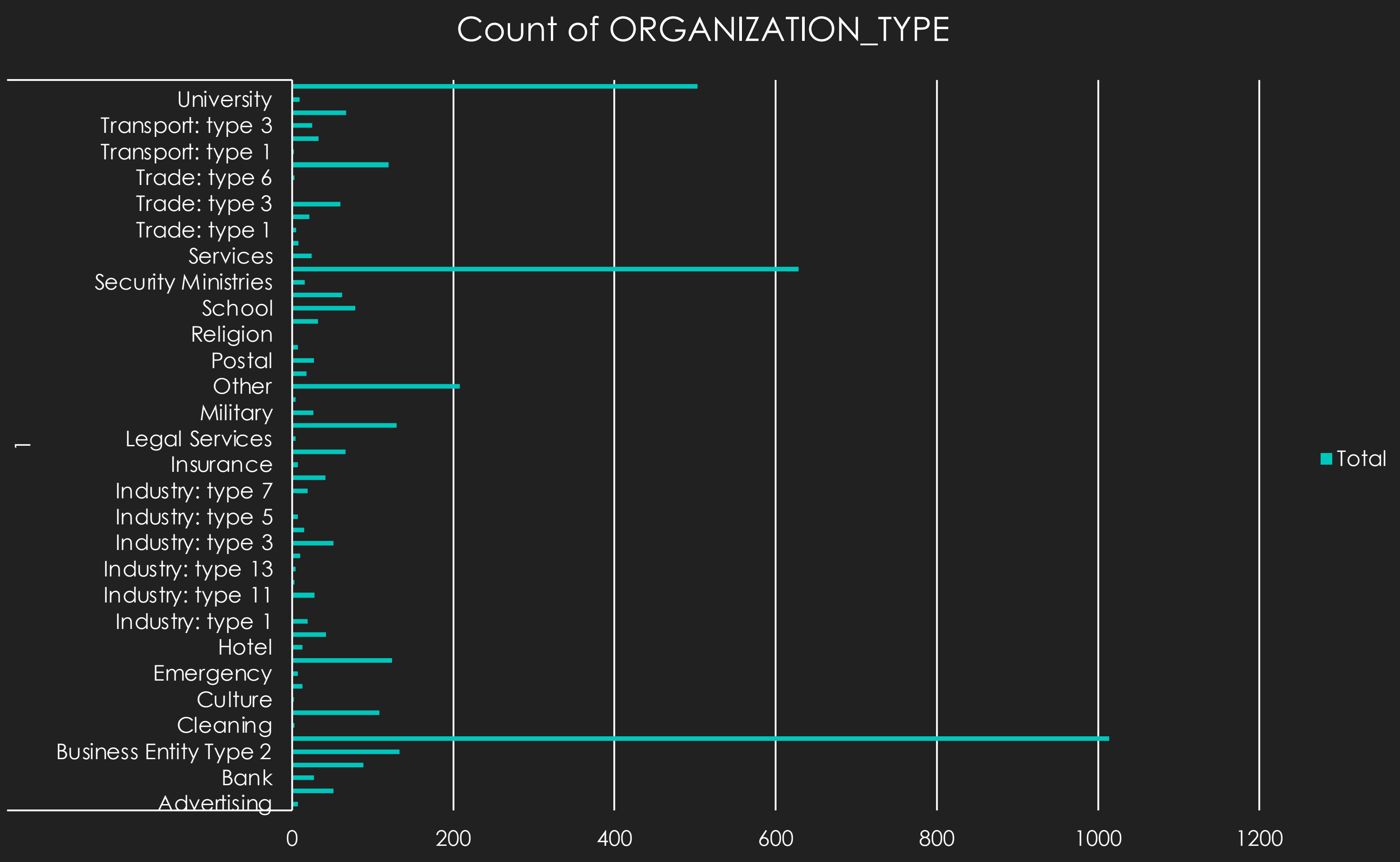
# 0 – Customers with no payment difficulties ORGANIZATION_TYPE

Most of the customers with no payment difficulties work in Business Entity type 3, and majority were also self employed.



# 1 – Customers with payment difficulties
# ORGANIZATION_TYPE

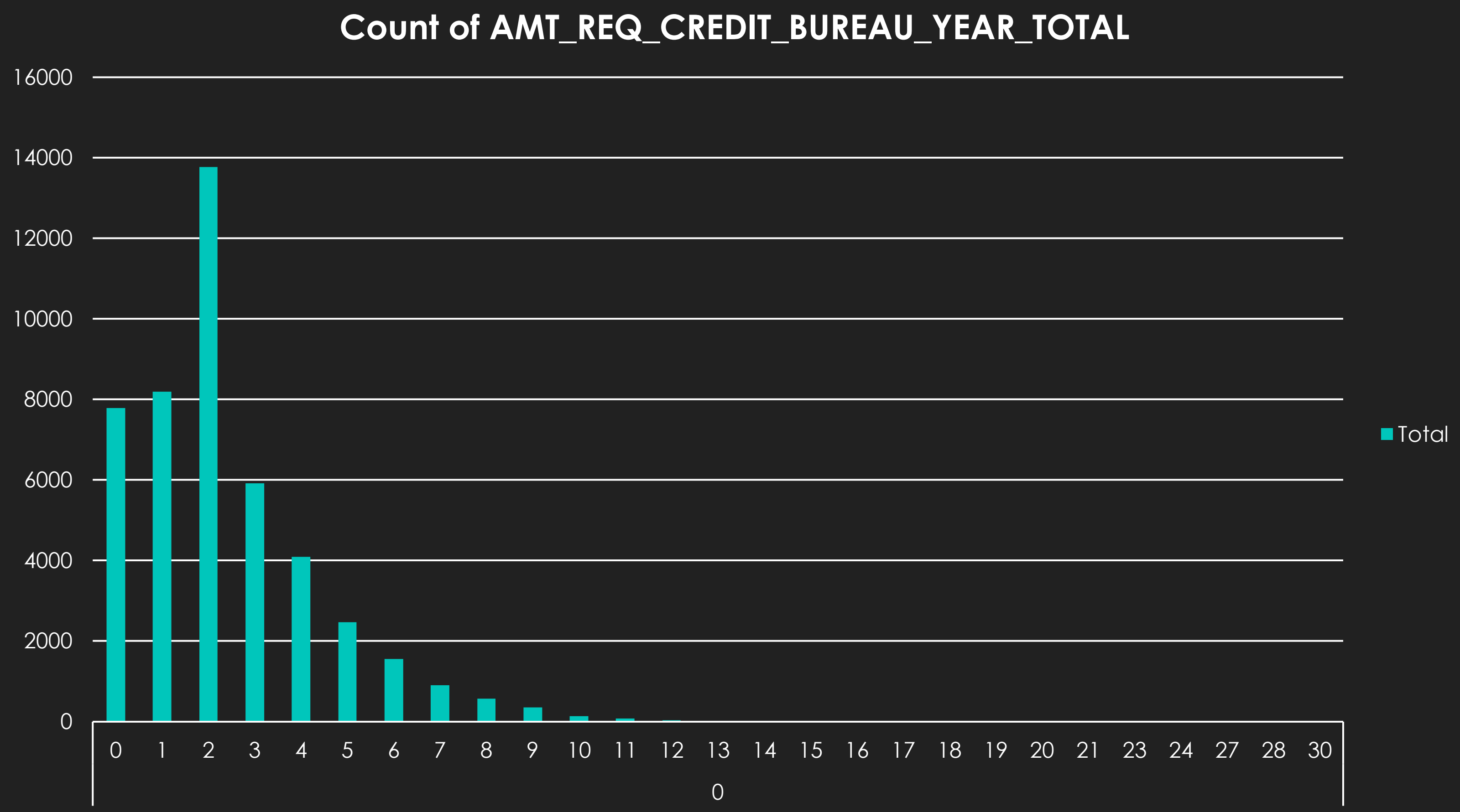
Most of the customers with payment difficulties work in Business Entity type 3, and majority were also self employed.



# 0 – Customers with no payment difficulties AMT_REQ_CREDIT_BUREAU_YEAR_TOTAL

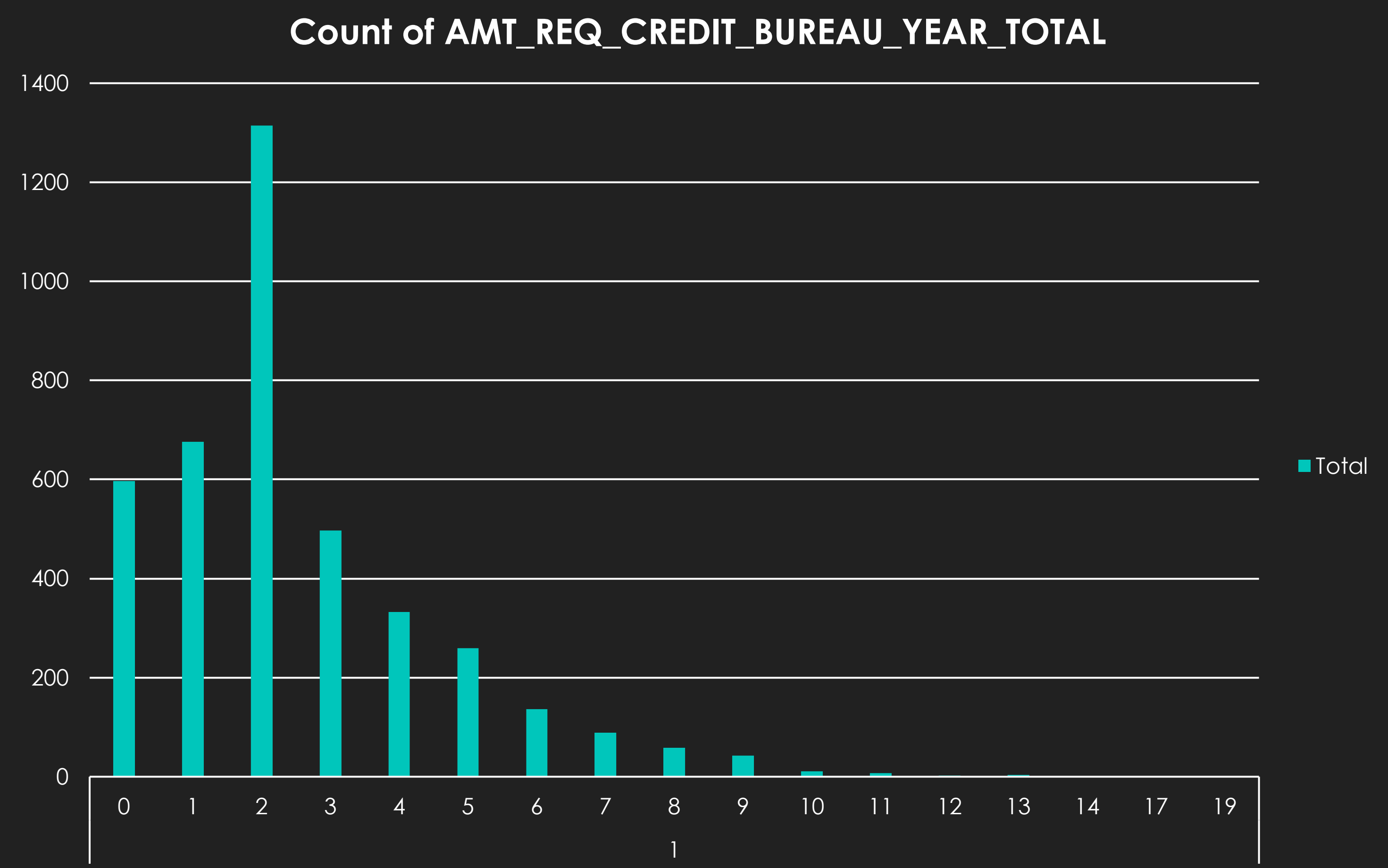
The inquiries to credit bureau twice for most of the customers with no payment difficulties.

The columns have been clubbed to sum up the count upto 1 year.



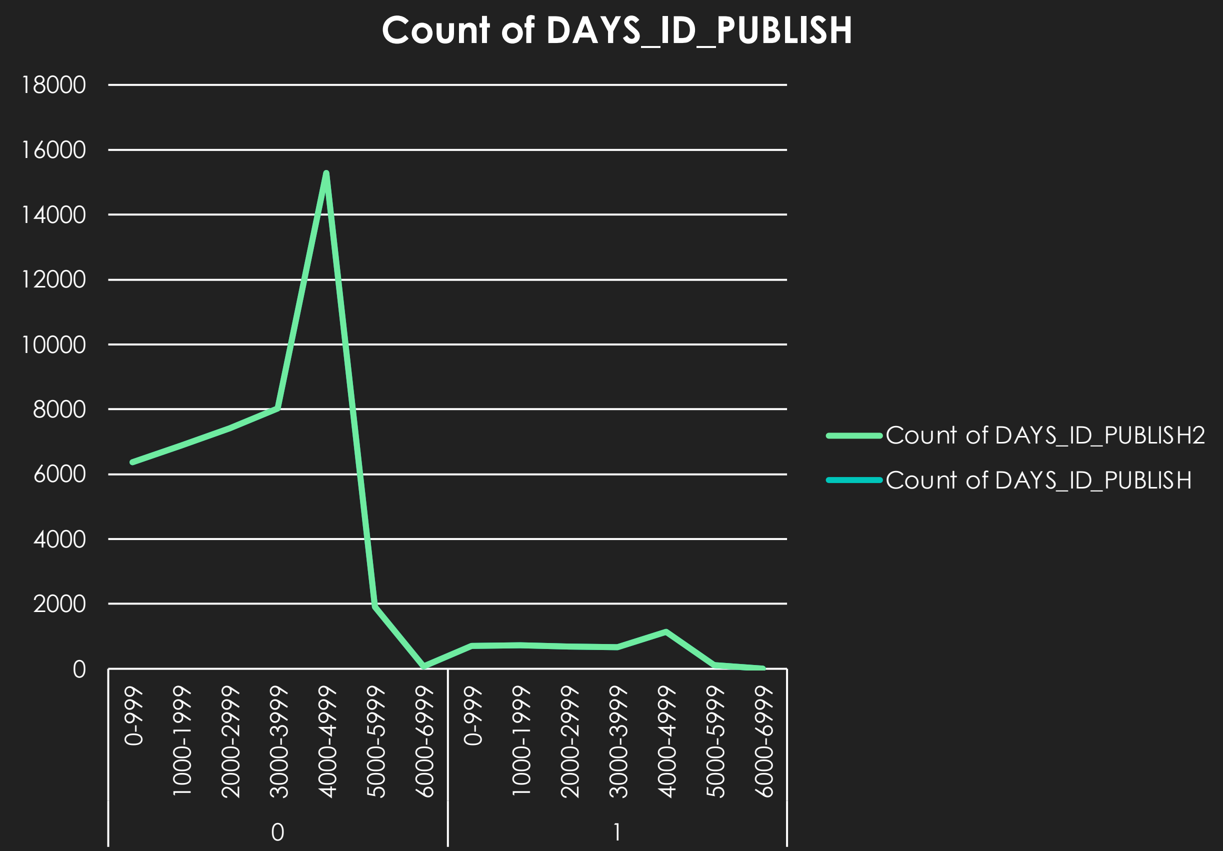
# 1 – Customers with payment difficulties AMT_REQ_CREDIT_BUREAU_YEAR_TOTAL

- The columns have been clubbed to sum up the count upto 1 year.
- The inquiries to credit bureau twice for most of the customers with payment difficulties.
- But the stats are same for customers with no payment difficulties, so it can be understood that on an average inquiries were made two times for a customer in last 1 year.



# Fraud Analysis

| Row Labels | Count of DAYS_ID_PUBLISH | Count of DAYS_ID_PUBLISH2 |
|---|---|---|
| 0 | 45973 | 91.95% |
| 0-999 | 6376 | 13.87% |
| 1000-1999 | 6887 | 14.98% |
| 2000-2999 | 7421 | 16.14% |
| 3000-3999 | 8033 | 17.47% |
| 4000-4999 | 15285 | 33.25% |
| 5000-5999 | 1902 | 4.14% |
| 6000-6999 | 69 | 0.15% |
| 1 | 4026 | 8.05% |
| 0-999 | 700 | 17.39% |
| 1000-1999 | 719 | 17.86% |
| 2000-2999 | 687 | 17.06% |
| 3000-3999 | 671 | 16.67% |
| 4000-4999 | 1132 | 28.12% |
| 5000-5999 | 112 | 2.78% |
| 6000-6999 | 5 | 0.12% |
| Grand Total | 49999 | 100.00% |

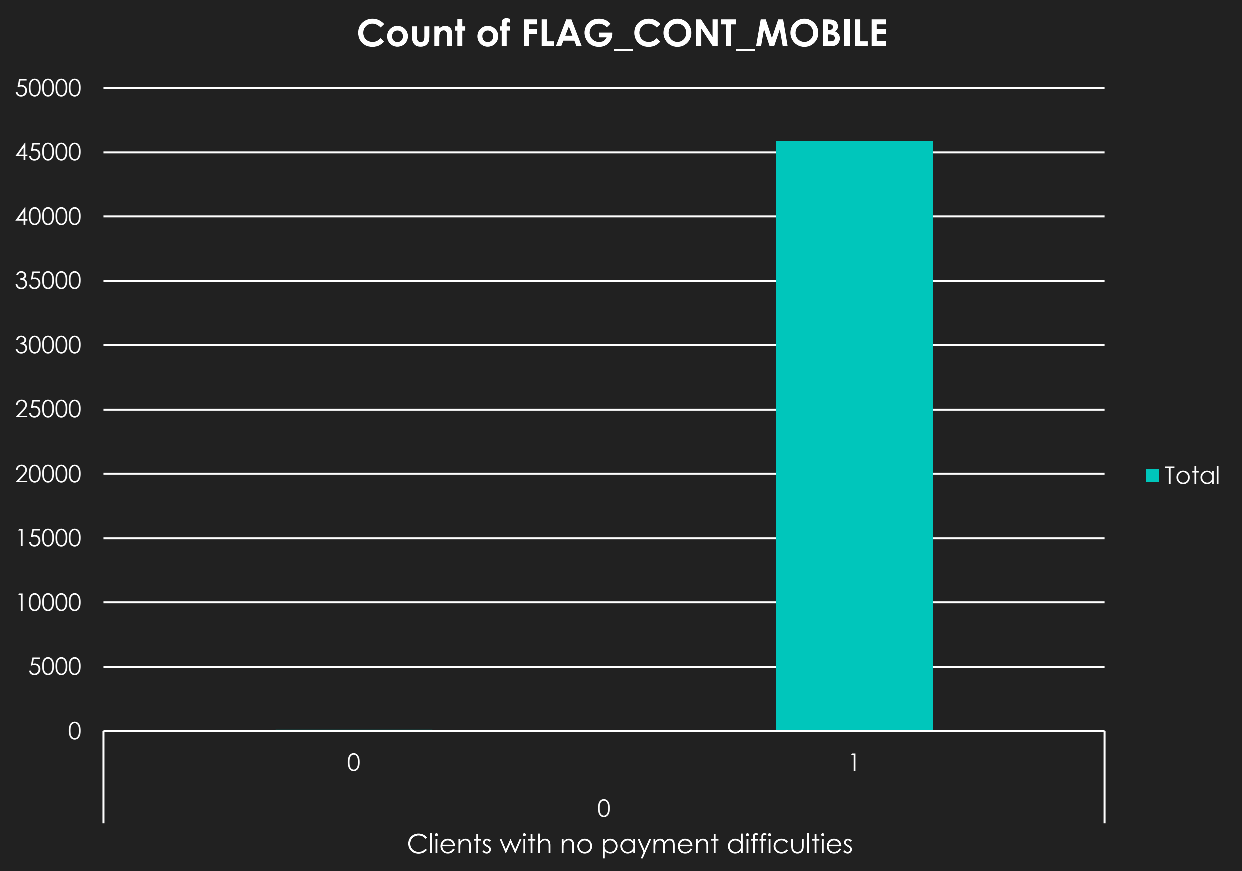



No specific trend could be found.

# Fraud Analysis

| Row Labels | Count of FLAG_CONT_MOBILE |
| --- | --- |
| 0 | 97 |
| 1 | 45876 |
| Grand Total | 45973 |

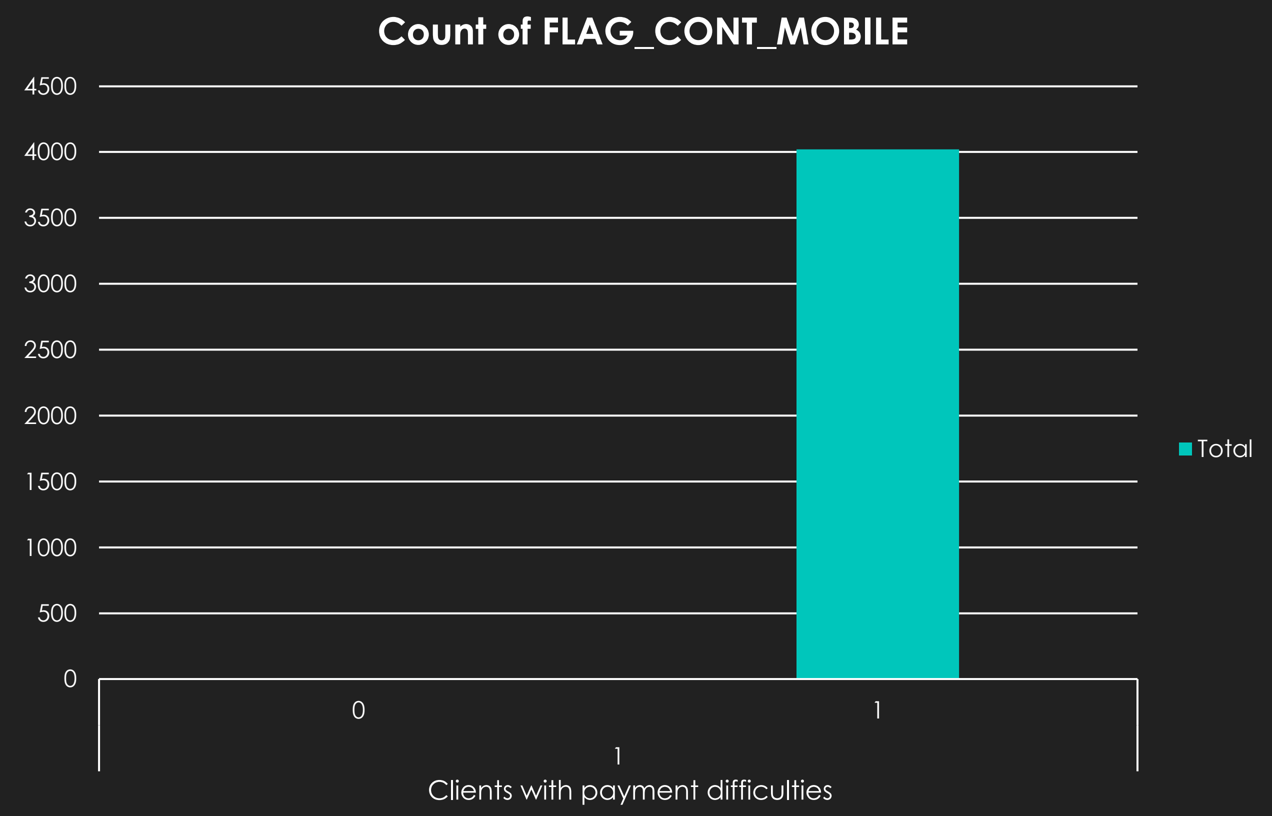
This column whether the client's phone was reachable or not.
Here, 0 represents no
1 represents yes
Majority of the clients who faced no payment difficulties had their phone reachable.



# Fraud Analysis

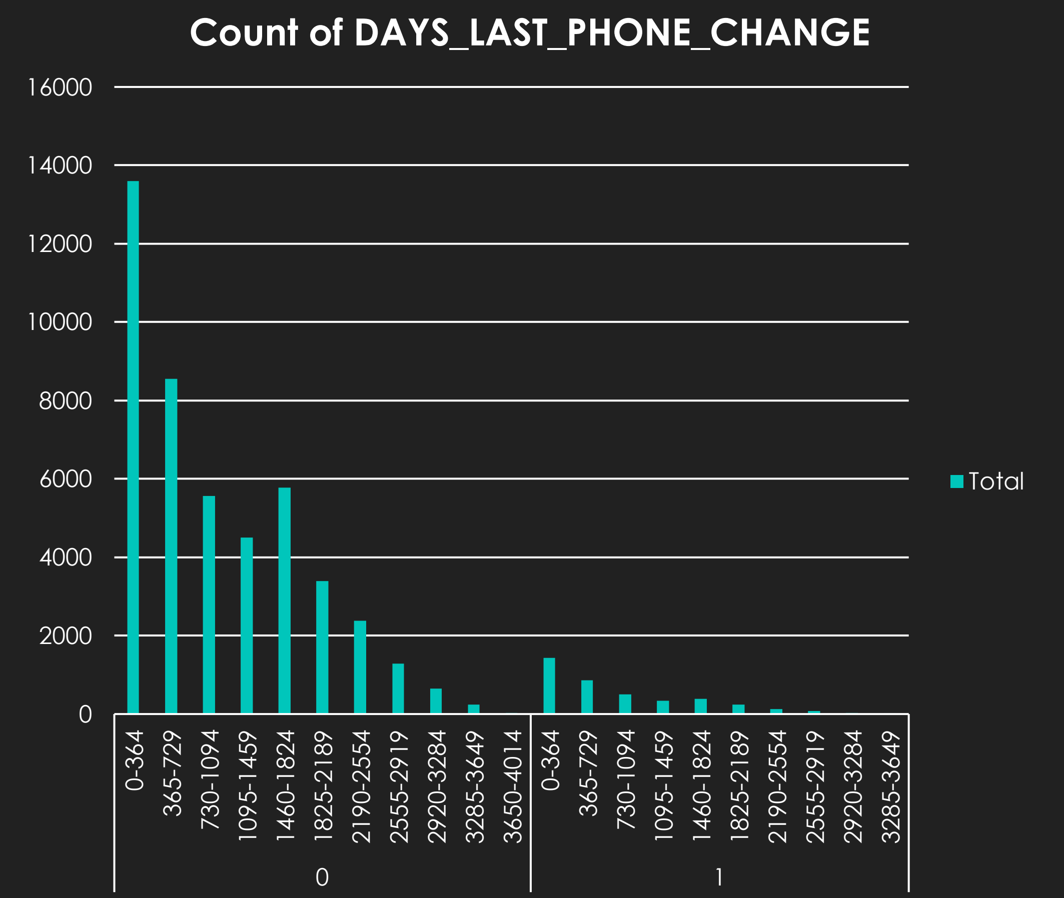
| Row Labels | Count of FLAG_CONT_MOBILE |
|---|---|
| 0 | 4 |
| 1 | 4022 |
| Grand Total | 4026 |

Majority of the clients who faced payment difficulties had their phone reachable. So, it does not seem to be a deliberate fraudulent practice of clients to bounce payments.



# Fraud Analysis

| Row Labels | Count of DAYS_LAST_PHONE_CHANGE |
|---|---|
| 0 | 45973 |
| 0-364 | 13600 |
| 365-729 | 8555 |
| 730-1094 | 5565 |
| 1095-1459 | 4509 |
| 1460-1824 | 5773 |
| 1825-2189 | 3388 |
| 2190-2554 | 2373 |
| 2555-2919 | 1293 |
| 2920-3284 | 649 |
| 3285-3649 | 238 |
| 3650-4014 | 30 |
| 1 | 4026 |
| 0-364 | 1440 |
| 365-729 | 866 |
| 730-1094 | 508 |
| 1095-1459 | 343 |
| 1460-1824 | 389 |
| 1825-2189 | 235 |
| 2190-2554 | 125 |
| 2555-2919 | 81 |
| 2920-3284 | 25 |
| 3285-3649 | 14 |
| Grand Total | 49999 |




Most of the applicants who have defaulted on payments have changed their phone in the last one year

# Correlation Matrix

| Overall | AMT INCOME TOTAL | AMT CREDIT | AMT ANNUITY | DAYS BIRTH year | DAYS ID PUBLISH | CNT FAM MEMBERS | DAYS LAST PHONE CHANGE NEW | AMT REQ CREDIT BUREAU YEAR Total |
|---|---|---|---|---|---|---|---|---|
| AMT INCOME TOTAL | 1 | | | | | | | |
| AMT CREDIT | 0.069315897 | 1 | | | | | | |
| AMT ANNUITY | 0.083008438 | 0.76949879 | 1 | | | | | |
| DAYS BIRTH year | -0.015871498 | 0.05911226 | -0.00804581 | 1 | | | | |
| DAYS ID PUBLISH | -0.003506646 | 0.01222876 | -0.006716927 | 0.271609378 | 1 | | | |
| CNT FAM MEMBERS | 0.011225511 | 0.06399716 | 0.07737959 | -0.276365674 | 0.026074278 | 1 | | |
| DAYS LAST PHONE CHANGE NEW | 0.004141658 | 0.07528529 | 0.065246635 | 0.079837875 | 0.090774609 | 0.022506978 | 1 | |
| AMT REQ CREDIT BUREAU YEAR Total | 0.007286752 | -0.00024933 | 0.015002696 | 0.060532471 | 0.038116142 | -0.024501112 | 0.111847863 | 1 |

AMT_ANNUITY and AMT_CREDIT have high positive correlation as expected while no other attribute show significant correlation.

| 0 | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | DAYS_BIRTH_year | DAYS_ID_PUBLISH | CNT_FAM_MEMBERS | DAYS_LAST_PHONE_CHANGE_NEW | AMT_REQ_CREDIT_BUREAU_YEAR_Total |
|---|---|---|---|---|---|---|---|---|
| AMT_INCOME_TOTAL | 1 | | | | | | | |
| AMT_CREDIT | 0.377965752 | 1 | | | | | | |
| AMT_ANNUITY | 0.451135167 | 0.77077282 | 1 | | | | | |
| DAYS_BIRTH_year | -0.073966771 | 0.05092406 | -0.010167705 | 1 | | | | |
| DAYS_ID_PUBLISH | -0.032286356 | 0.00829019 | -0.00942697 | 0.270944996 | 1 | | | |
| CNT_FAM_MEMBERS | 0.041599302 | 0.06487694 | 0.077892626 | -0.283548577 | 0.025054258 | 1 | | |
| DAYS_LAST_PHONE_CHANGE_NEW | 0.045978573 | 0.07048628 | 0.062421355 | 0.072071995 | 0.084246321 | 0.025018128 | 1 | |
| AMT_REQ_CREDIT_BUREAU_YEAR_Total | 0.054557569 | 0.00017615 | 0.014432023 | 0.059726216 | 0.035696206 | -0.025894464 | 0.114187737 | 1 |

Clients with no payment difficulties

There is high positive correlation between AMT_ANNUITY & AMT_CREDIT , no other substantial correlation is observed.

| 1 | AMT INCOME TOTAL | AMT CREDIT | AMT ANNUITY | DAYS BIRTH year | DAYS ID PUBLISH | CNT FAM MEMBERS | DAYS LAST PHONE CHANGE NEW | AMT REQ CREDIT BUREAU YEAR Total |
|---|---|---|---|---|---|---|---|---|
| AMT INCOME TOTAL | 1 | | | | | | | |
| AMT CREDIT | 0.015271444 | 1 | | | | | | |
| AMT ANNUITY | 0.018004594 | 0.7496652 | 1 | | | | | |
| DAYS BIRTH year | -0.008419044 | 0.14130676 | 0.007248662 | 1 | | | | |
| DAYS ID PUBLISH | 0.009122006 | 0.0437719 | 0.02132109 | 0.247697382 | 1 | | | |
| CNT FAM MEMBERS | 0.013121678 | 0.06124869 | 0.075838463 | -0.197711093 | 0.044037815 | 1 | | |
| DAYS LAST PHONE CHANGE NEW | -0.012642088 | 0.12091321 | 0.098610858 | 0.125790656 | 0.140069979 | 0.003474821 | 1 | |
| AMT REQ CREDIT BUREAU YEAR Total | -0.008196542 | -0.00024734 | 0.025387094 | 0.084832424 | 0.073252128 | -0.011741962 | 0.094573561 | 1 |

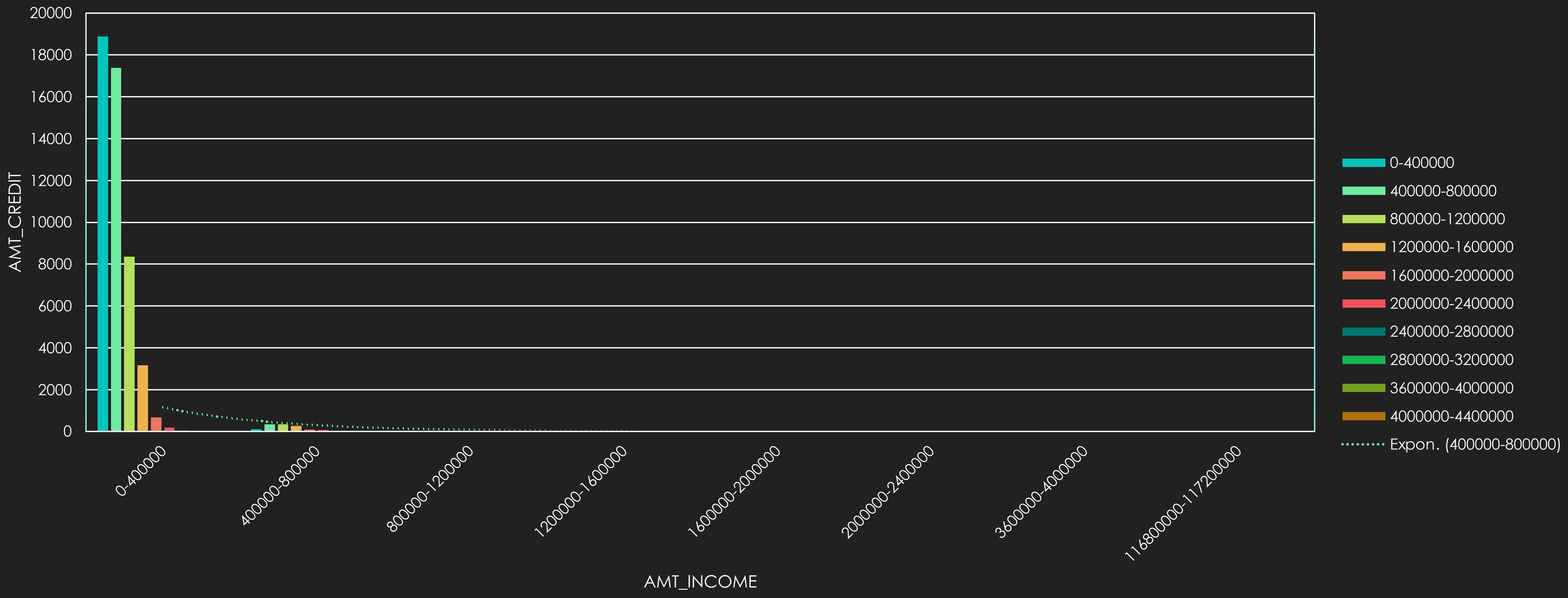
Clients with payment difficulties

As expected there is high positive correlation between AMT_CREDIT & AMT_ANNUITY, no other significant value is observed.

# Bivariate Analysis

# AMT_INCOME TO AMT_CREDIT

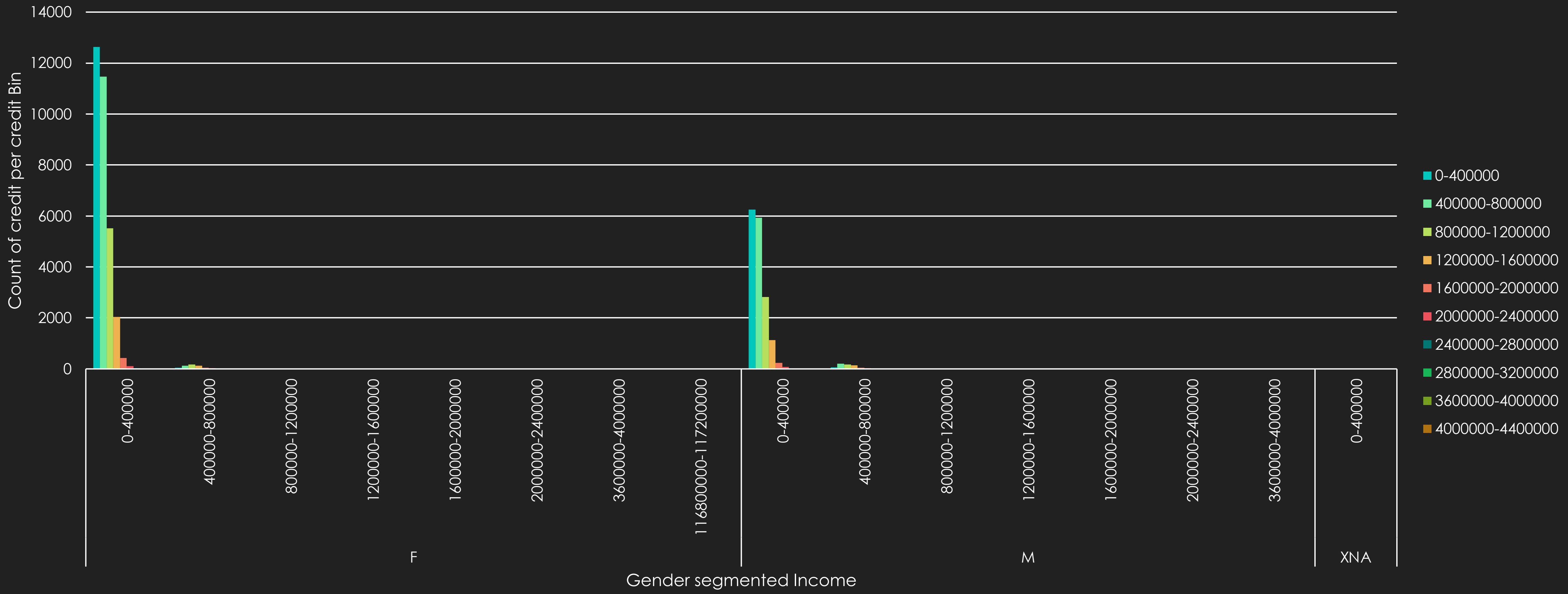
| Count of SK_ID_CURR | Column Labels | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row Labels | 0-400000 | 400000-800000 | 800000-1200000 | 1200000-1600000 | 1600000-2000000 | 2000000-2400000 | 2400000-2800000 | 2800000-3200000 | 3600000-4000000 | 4000000-4400000 | Grand Total |
| 0-400000 | 18881 | 17394 | 8351 | 3169 | 670 | 185 | 42 | 4 | | | 48696 |
| 400000-800000 | 110 | 333 | 338 | 259 | 84 | 61 | 18 | | 1 | 2 | 1206 |
| 800000-1200000 | 11 | 19 | 13 | 15 | 8 | 7 | 1 | | | | 74 |
| 1200000-1600000 | 2 | 4 | 4 | | | 1 | 1 | | | | 12 |
| 1600000-2000000 | 1 | 1 | 1 | 1 | | | | | | | 4 |
| 2000000-2400000 | | 1 | 2 | 1 | | | | | | | 4 |
| 3600000-4000000 | | | 1 | 1 | | | | | | | 2 |
| 116800000-117200000 | | 1 | | | | | | | | | 1 |
| Grand Total | 19005 | 17753 | 8710 | 3446 | 762 | 254 | 62 | 4 | 1 | 2 | 49999 |




Most of the applicants belong to the lower income range and also apply for low amount of credit. The distribution follows Poisson distribution being heavily skewed on the left.

# Gender, Credit And Income

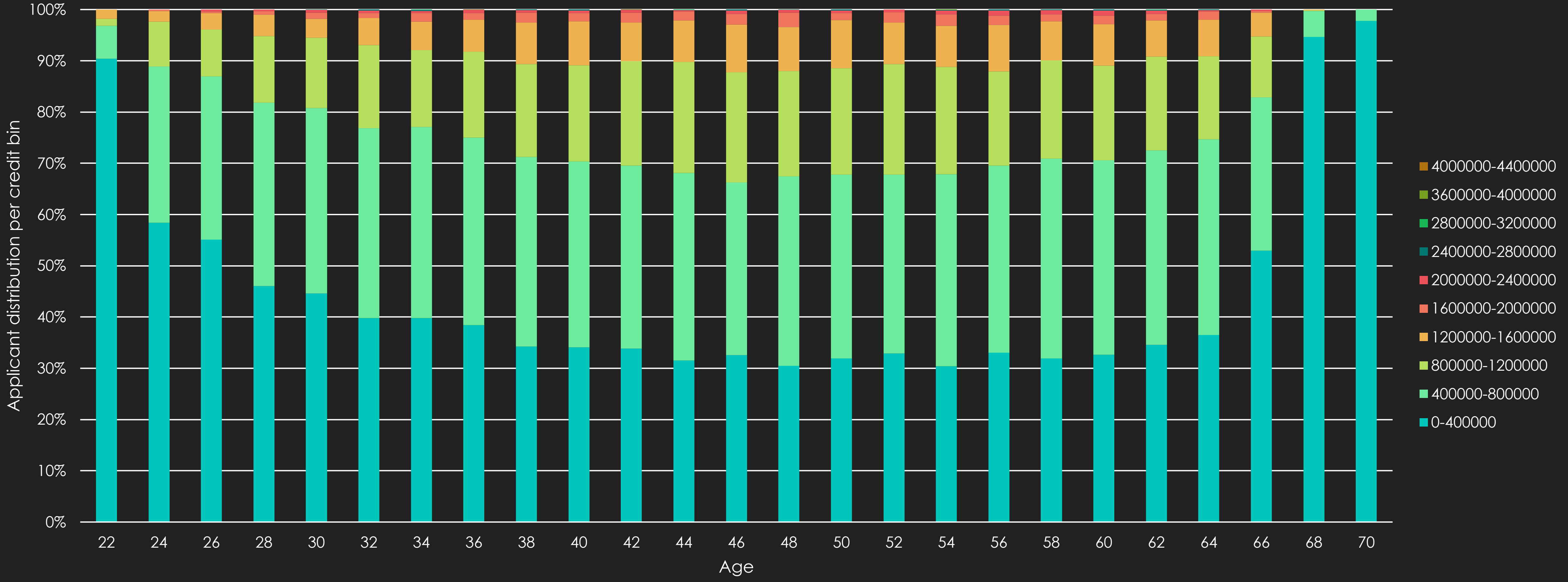
| Count of SK_ID_CURR | Column Labels | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Row Labels | 0-400000 | 400000-800000 | 800000-1200000 | 1200000-1600000 | 1600000-2000000 | 2000000-2400000 | 2400000-2800000 | 2800000-3200000 | 3600000-4000000 | 4000000-4400000 | Grand Total |
| F | 12689 | 11608 | 5700 | 2168 | 476 | 142 | 38 | 1 | 1 | | 32823 |
| 0-400000 | 12632 | 11468 | 5524 | 2043 | 432 | 114 | 26 | 1 | | | 32240 |
| 400000-800000 | 51 | 131 | 165 | 120 | 43 | 26 | 10 | | 1 | | 547 |
| 800000-1200000 | 4 | 5 | 7 | 4 | 1 | 2 | 1 | | | | 24 |
| 1200000-1600000 | 1 | 2 | 2 | | | | 1 | | | | 6 |
| 1600000-2000000 | 1 | | | 1 | | | | | | | 2 |
| 2000000-2400000 | | 1 | 1 | | | | | | | | 2 |
| 3600000-4000000 | | | 1 | | | | | | | | 1 |
| 116800000-117200000 | | 1 | | | | | | | | | 1 |
| M | 6314 | 6145 | 3010 | 1278 | 286 | 112 | 24 | 3 | | 2 | 17174 |
| 0-400000 | 6247 | 5926 | 2827 | 1126 | 238 | 71 | 16 | 3 | | | 16454 |
| 400000-800000 | 59 | 202 | 173 | 139 | 41 | 35 | 8 | | | 2 | 659 |
| 800000-1200000 | 7 | 14 | 6 | 11 | 7 | 5 | | | | | 50 |
| 1200000-1600000 | 1 | 2 | 2 | | | 1 | | | | | 6 |
| 1600000-2000000 | | 1 | 1 | | | | | | | | 2 |
| 2000000-2400000 | | | 1 | 1 | | | | | | | 2 |
| 3600000-4000000 | | | | 1 | | | | | | | 1 |
| XNA | 2 | | | | | | | | | | 2 |
| 0-400000 | 2 | | | | | | | | | | 2 |
| Grand Total | 19005 | 17753 | 8710 | 3446 | 762 | 254 | 62 | 4 | 1 | 2 | 49999 |




The highest number of applicants, both male and female belong to the lower income range and apply for the low amount of credit.

# AGE to CREDIT

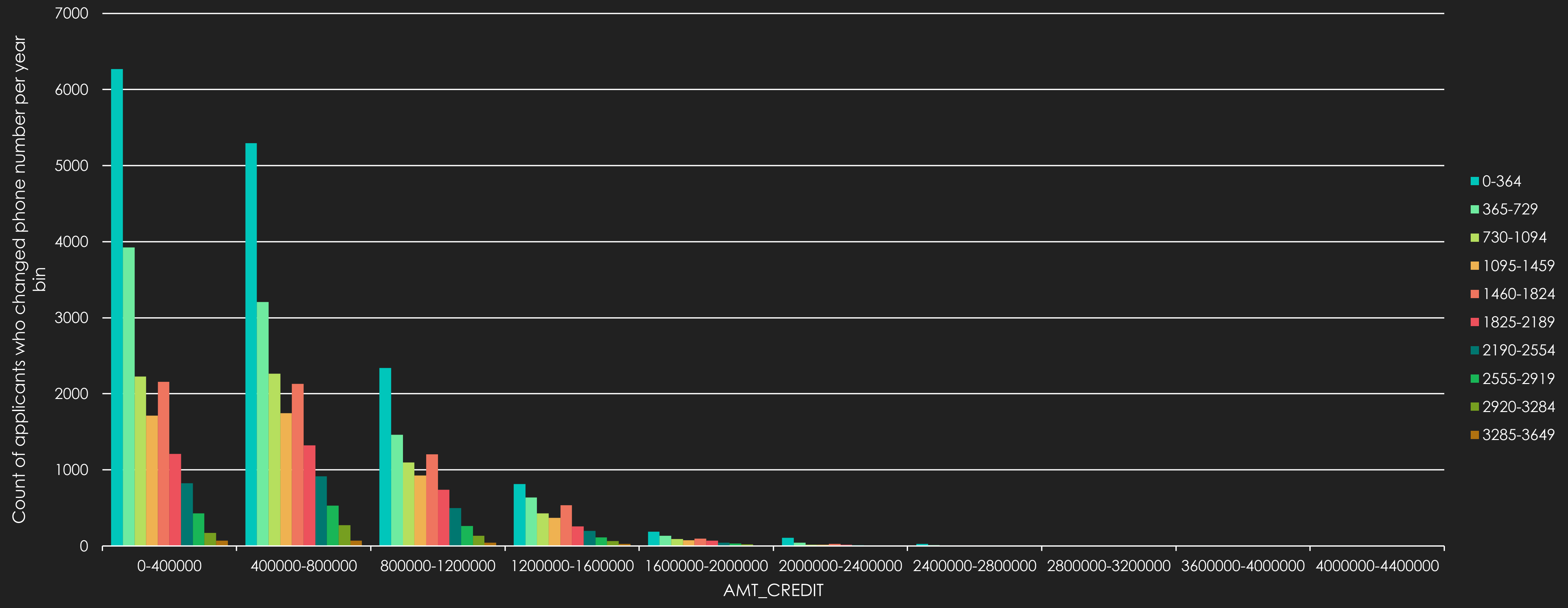
| Count of SK_ID_CURR | Column Labels | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Row Labels | 0-400000 | 400000-800000 | 800000-1200000 | 1200000-1600000 | 1600000-2000000 | 2000000-2400000 | 2400000-2800000 | 2800000-3200000 | 3600000-4000000 | 4000000-4400000 | Grand Total |
| 22 | 197 | 14 | 3 | 4 | | | | | | | 218 |
| 24 | 682 | 356 | 102 | 25 | 2 | 1 | | | | | 1168 |
| 26 | 725 | 418 | 121 | 42 | 4 | 5 | | | | | 1315 |
| 28 | 940 | 729 | 265 | 86 | 14 | 6 | | | | | 2040 |
| 30 | 1142 | 926 | 353 | 94 | 31 | 13 | 2 | | | | 2561 |
| 32 | 1071 | 997 | 435 | 143 | 25 | 13 | 6 | | | | 2690 |
| 34 | 1043 | 979 | 392 | 146 | 43 | 11 | 7 | 1 | | | 2622 |
| 36 | 951 | 905 | 416 | 155 | 31 | 15 | 1 | | | 1 | 2475 |
| 38 | 961 | 1035 | 509 | 226 | 55 | 13 | 4 | | | | 2803 |
| 40 | 966 | 1027 | 532 | 241 | 47 | 14 | 5 | | | | 2832 |
| 42 | 963 | 1014 | 580 | 214 | 55 | 15 | 1 | | | 1 | 2843 |
| 44 | 834 | 966 | 573 | 213 | 47 | 6 | 2 | 2 | | | 2643 |
| 46 | 794 | 820 | 525 | 227 | 50 | 15 | 6 | | | | 2437 |
| 48 | 718 | 870 | 485 | 202 | 66 | 12 | 2 | | | | 2355 |
| 50 | 706 | 793 | 459 | 207 | 31 | 11 | 4 | | | | 2211 |
| 52 | 706 | 751 | 464 | 173 | 43 | 12 | | | | | 2149 |
| 54 | 714 | 881 | 493 | 189 | 53 | 16 | 4 | | 1 | | 2351 |
| 56 | 757 | 835 | 421 | 206 | 44 | 21 | 5 | | | | 2289 |
| 58 | 706 | 861 | 423 | 169 | 31 | 17 | 3 | | | | 2210 |
| 60 | 660 | 766 | 374 | 164 | 34 | 18 | 5 | | | | 2021 |
| 62 | 659 | 723 | 348 | 136 | 24 | 12 | 3 | 1 | | | 1906 |
| 64 | 631 | 660 | 281 | 123 | 28 | 4 | 2 | | | | 1729 |
| 66 | 689 | 389 | 155 | 60 | 4 | 4 | | | | | 1301 |
| 68 | 660 | 35 | 1 | 1 | | | | | | | 697 |
| 70 | 130 | 3 | | | | | | | | | 133 |
| Grand Total | 19005 | 17753 | 8710 | 3446 | 762 | 254 | 62 | 4 | 1 | 2 | 49999 |




Applicants belonging to lower age groups and upper age groups above 60 years mostly apply for smaller amount of credit, while middle age applicants apply for the larger amount.

# CREDIT vs PHONE NUMBER CHANGES

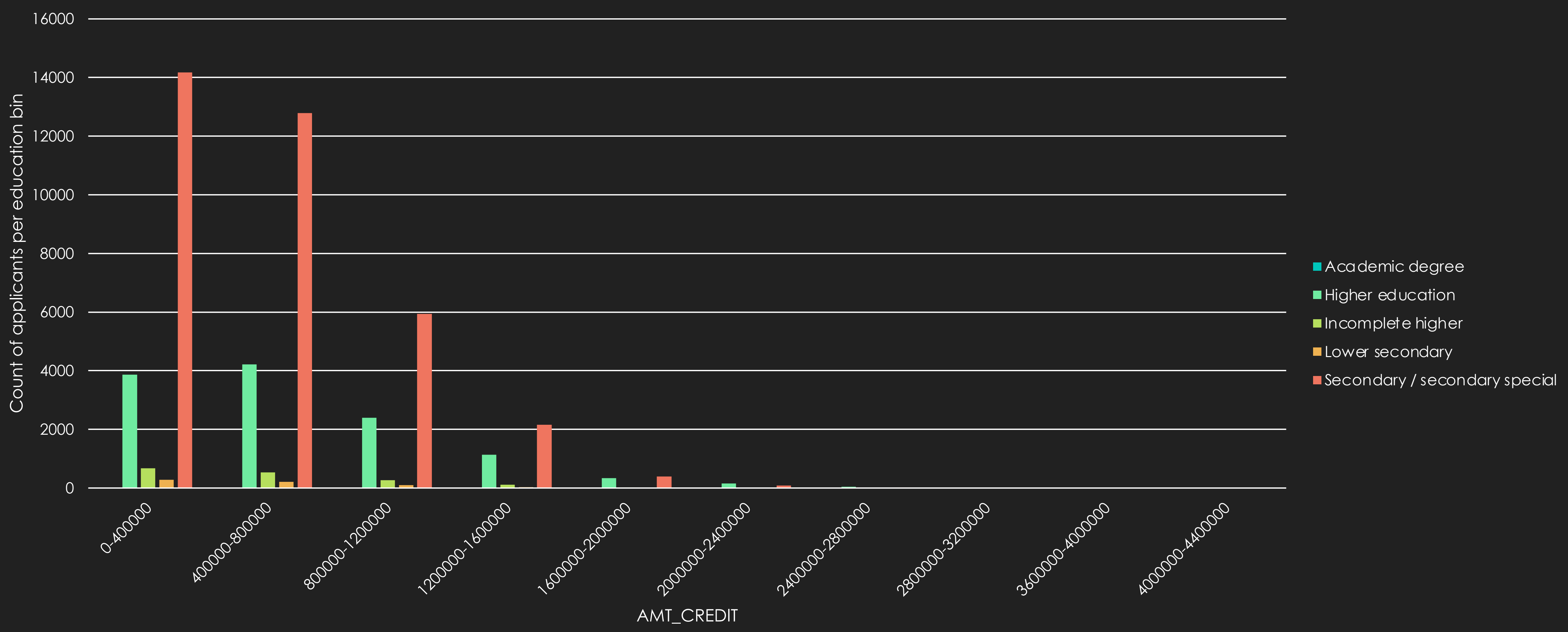
| Count of SK_ID_CURR | Column Labels | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Row Labels | 0-364 | 365-729 | 730-1094 | 1095-1459 | 1460-1824 | 1825-2189 | 2190-2554 | 2555-2919 | 2920-3284 | 3285-3649 | Grand Total |
| 0-400000 | 6269 | 3927 | 2227 | 1714 | 2160 | 1210 | 825 | 430 | 173 | 70 | 19005 |
| 400000-800000 | 5293 | 3206 | 2264 | 1746 | 2133 | 1323 | 917 | 528 | 271 | 72 | 17753 |
| 800000-1200000 | 2340 | 1463 | 1097 | 927 | 1203 | 738 | 501 | 263 | 136 | 42 | 8710 |
| 1200000-1600000 | 813 | 638 | 429 | 370 | 536 | 260 | 197 | 112 | 66 | 25 | 3446 |
| 1600000-2000000 | 186 | 136 | 92 | 76 | 99 | 71 | 45 | 32 | 20 | 5 | 762 |
| 2000000-2400000 | 106 | 41 | 19 | 16 | 26 | 18 | 11 | 7 | 5 | 5 | 254 |
| 2400000-2800000 | 29 | 9 | 7 | 3 | 4 | 3 | 2 | 2 | 3 | | 62 |
| 2800000-3200000 | 3 | | | | 1 | | | | | | 4 |
| 3600000-4000000 | | 1 | | | | | | | | | 1 |
| 4000000-4400000 | 1 | | 1 | | | | | | | | 2 |
| Grand Total | 15040 | 9421 | 6136 | 4852 | 6162 | 3623 | 2498 | 1374 | 674 | 219 | 49999 |




Highest number of applicants have changed their phone number in the last year in all the credit bins..

# CREDIT to DEGREE

| Count of SK_ID_CURR | Column Labels | | | | | |
|---|---|---|---|---|---|---|
| Row Labels | Academic degree | Higher education | Incomplete higher | Lower secondary | Secondary / secondary special | Grand Total |
| 0-400000 | 7 | 3863 | 675 | 283 | 14177 | 19005 |
| 400000-800000 | 4 | 4216 | 529 | 213 | 12791 | 17753 |
| 800000-1200000 | 3 | 2399 | 273 | 94 | 5941 | 8710 |
| 1200000-1600000 | 4 | 1142 | 118 | 24 | 2158 | 3446 |
| 1600000-2000000 | 2 | 339 | 20 | 4 | 397 | 762 |
| 2000000-2400000 | | 161 | 4 | 2 | 87 | 254 |
| 2400000-2800000 | | 42 | 1 | | 19 | 62 |
| 2800000-3200000 | | 2 | | | 2 | 4 |
| 3600000-4000000 | | 1 | | | | 1 |
| 4000000-4400000 | | 2 | | | | 2 |
| Grand Total | 20 | 12167 | 1620 | 620 | 35572 | 49999 |




Highest number of applicants in all the credit ranges have education upto secondary/ secondary special.

Video Link

https://www.loom.com/share/a95d13c8b6ef4f3fb0ad73845f5a7a88?sid=63fa5ff8-823b-4521-b237-0d3dba89a33a

Bivariate Analysis Excel Sheet

https://docs.google.com/spreadsheets/d/1aHfwWSbqmhHyGth01-OCEUVpL2zZZEcI/edit?usp=share_link&ouid=101949921485202693908&rtpof=true&sd=true

Univariate Analysis Excel Sheet

https://docs.google.com/spreadsheets/d/1bcKn2xjE_m-4x3lEyiOavscVFcbXHLA3/edit?usp=sharing&ouid=101949921485202693908&rtpof=true&sd=true