

**PROJECT 6**

**BANK LOAN  
CASE STUDY**

**TECH-STACK USED**

Microsoft Excel

**A MADHAVA VARMA**

# **ANALYSIS ON**

- A. Identify Missing Data and Deal with it Appropriately**
- B. Identify Outliers in the Dataset**
- C. Analyze Data Imbalance**
- D. Perform Univariate, Segmented Univariate, and Bivariate Analysis**
- E. Identify Top Correlations for Different Scenarios**

# A. MISSING DATA

**Objective:** As a data analyst, you come across missing data in the loan application dataset. It is essential to handle missing data effectively to ensure the accuracy of the analysis.

**Your task:** Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.

**Formulae:-**

**no\_of\_null\_values:**

=COUNT(IF(ISBLANK(application\_data!AW2:AW50000),application\_data!\$A:\$A)) Transpose:

=TRANSPOSE(A1:DS3)

# A. MISSING DATA

## RESULT:

Columns having 50% or more than 50% of null values

| Column name              | no_of_null_values | Percentage_of_null_values |
|--------------------------|-------------------|---------------------------|
| COMMONAREA_AVG           | 34960             | 70%                       |
| COMMONAREA_MODE          | 34960             | 70%                       |
| COMMONAREA_MEDI          | 34960             | 70%                       |
| NONLIVINGAPARTMENTS_AVG  | 34714             | 69%                       |
| NONLIVINGAPARTMENTS_MODE | 34714             | 69%                       |
| NONLIVINGAPARTMENTS_MEDI | 34714             | 69%                       |
| LIVINGAPARTMENTS_AVG     | 34226             | 68%                       |
| LIVINGAPARTMENTS_MODE    | 34226             | 68%                       |
| LIVINGAPARTMENTS_MEDI    | 34226             | 68%                       |
| FONDKAPREMONT_MODE       | 34191             | 68%                       |
| FLOORSMIN_AVG            | 33894             | 68%                       |
| FLOORSMIN_MODE           | 33894             | 68%                       |
| FLOORSMIN_MEDI           | 33894             | 68%                       |
| YEARS_BUILD_AVG          | 33239             | 66%                       |
| YEARS_BUILD_MODE         | 33239             | 66%                       |
| YEARS_BUILD_MEDI         | 33239             | 66%                       |
| OWN_CAR_AGE              | 32949             | 66%                       |
| LANDAREA_AVG             | 29721             | 59%                       |
| LANDAREA_MODE            | 29721             | 59%                       |
| LANDAREA_MEDI            | 29721             | 59%                       |
| BASEMENTAREA_AVG         | 29199             | 58%                       |
| BASEMENTAREA_MODE        | 29199             | 58%                       |
| BASEMENTAREA_MEDI        | 29199             | 58%                       |
| EXT_SOURCE_1             | 28172             | 56%                       |
| NONLIVINGAREA_AVG        | 27572             | 55%                       |
| NONLIVINGAREA_MODE       | 27572             | 55%                       |
| NONLIVINGAREA_MEDI       | 27572             | 55%                       |
| ELEVATORS_AVG            | 26651             | 53%                       |
| ELEVATORS_MODE           | 26651             | 53%                       |
| ELEVATORS_MEDI           | 26651             | 53%                       |
| WALLSMATERIAL_MODE       | 25459             | 51%                       |
| APARTMENTS_AVG           | 25385             | 51%                       |
| APARTMENTS_MODE          | 25385             | 51%                       |
| APARTMENTS_MEDI          | 25385             | 51%                       |
| ENTRANCES_AVG            | 25195             | 50%                       |
| ENTRANCES_MODE           | 25195             | 50%                       |
| ENTRANCES_MEDI           | 25195             | 50%                       |
| LIVINGAREA_AVG           | 25137             | 50%                       |
| LIVINGAREA_MODE          | 25137             | 50%                       |
| LIVINGAREA_MEDI          | 25137             | 50%                       |
| HOUSETYPE_MODE           | 25075             | 50%                       |
| FLOORSMAX_AVG            | 24875             | 50%                       |
| FLOORSMAX_MODE           | 24875             | 50%                       |
| FLOORSMAX_MEDI           | 24875             | 50%                       |

# A. MISSING DATA

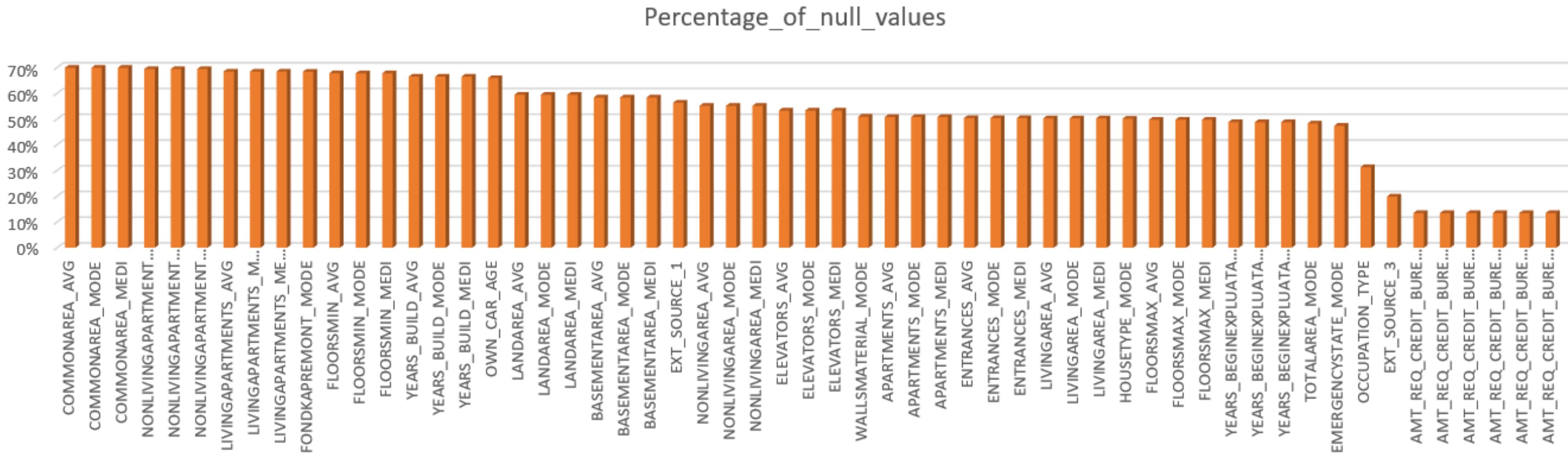
## RESULT:

Columns having irrelevant data which is not required for analysis.

| Column name                  | no_of_null_values | Percentage_of_null_values |
|------------------------------|-------------------|---------------------------|
| FLAG_MOBIL                   | 0                 | 0%                        |
| FLAG_EMP_PHONE               | 0                 | 0%                        |
| FLAG_WORK_PHONE              | 0                 | 0%                        |
| FLAG_CONT_MOBILE             | 0                 | 0%                        |
| FLAG_PHONE                   | 0                 | 0%                        |
| FLAG_EMAIL                   | 0                 | 0%                        |
| CNT_FAM_MEMBERS              | 1                 | 0%                        |
| REGION_RATING_CLIENT         | 0                 | 0%                        |
| REGION_RATING_CLIENT_W_CITY  | 0                 | 0%                        |
| EXT_SOURCE_2                 | 126               | 0%                        |
| EXT_SOURCE_3                 | 9944              | 20%                       |
| YEARS_BEGINEXPLUATATION_AVG  | 24394             | 49%                       |
| YEARS_BEGINEXPLUATATION_MODE | 24394             | 49%                       |
| YEARS_BEGINEXPLUATATION_MEDI | 24394             | 49%                       |
| TOTALAREA_MODE               | 24148             | 48%                       |
| EMERGENCYSTATE_MODE          | 23698             | 47%                       |
| DAYS_LAST_PHONE_CHANGE       | 1                 | 0%                        |
| FLAG_DOCUMENT_2              | 0                 | 0%                        |
| FLAG_DOCUMENT_3              | 0                 | 0%                        |
| FLAG_DOCUMENT_4              | 0                 | 0%                        |
| FLAG_DOCUMENT_5              | 0                 | 0%                        |
| FLAG_DOCUMENT_6              | 0                 | 0%                        |
| FLAG_DOCUMENT_7              | 0                 | 0%                        |
| FLAG_DOCUMENT_8              | 0                 | 0%                        |
| FLAG_DOCUMENT_9              | 0                 | 0%                        |
| FLAG_DOCUMENT_10             | 0                 | 0%                        |
| FLAG_DOCUMENT_11             | 0                 | 0%                        |
| FLAG_DOCUMENT_12             | 0                 | 0%                        |
| FLAG_DOCUMENT_13             | 0                 | 0%                        |
| FLAG_DOCUMENT_14             | 0                 | 0%                        |
| FLAG_DOCUMENT_15             | 0                 | 0%                        |
| FLAG_DOCUMENT_16             | 0                 | 0%                        |
| FLAG_DOCUMENT_17             | 0                 | 0%                        |
| FLAG_DOCUMENT_18             | 0                 | 0%                        |
| FLAG_DOCUMENT_19             | 0                 | 0%                        |
| FLAG_DOCUMENT_20             | 0                 | 0%                        |
| FLAG_DOCUMENT_21             | 0                 | 0%                        |

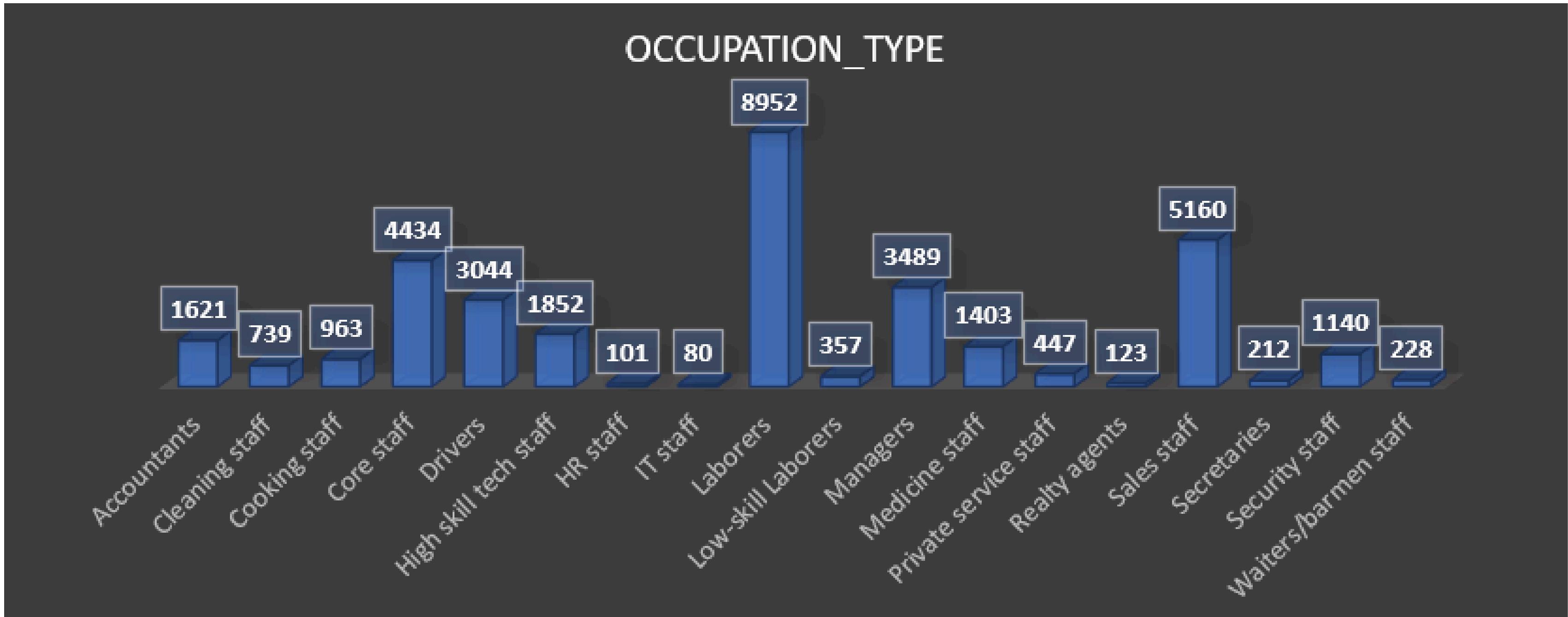
# A. MISSING DATA

**RESULT:** Bar graph showing the percentage of null values



# A. MISSING DATA

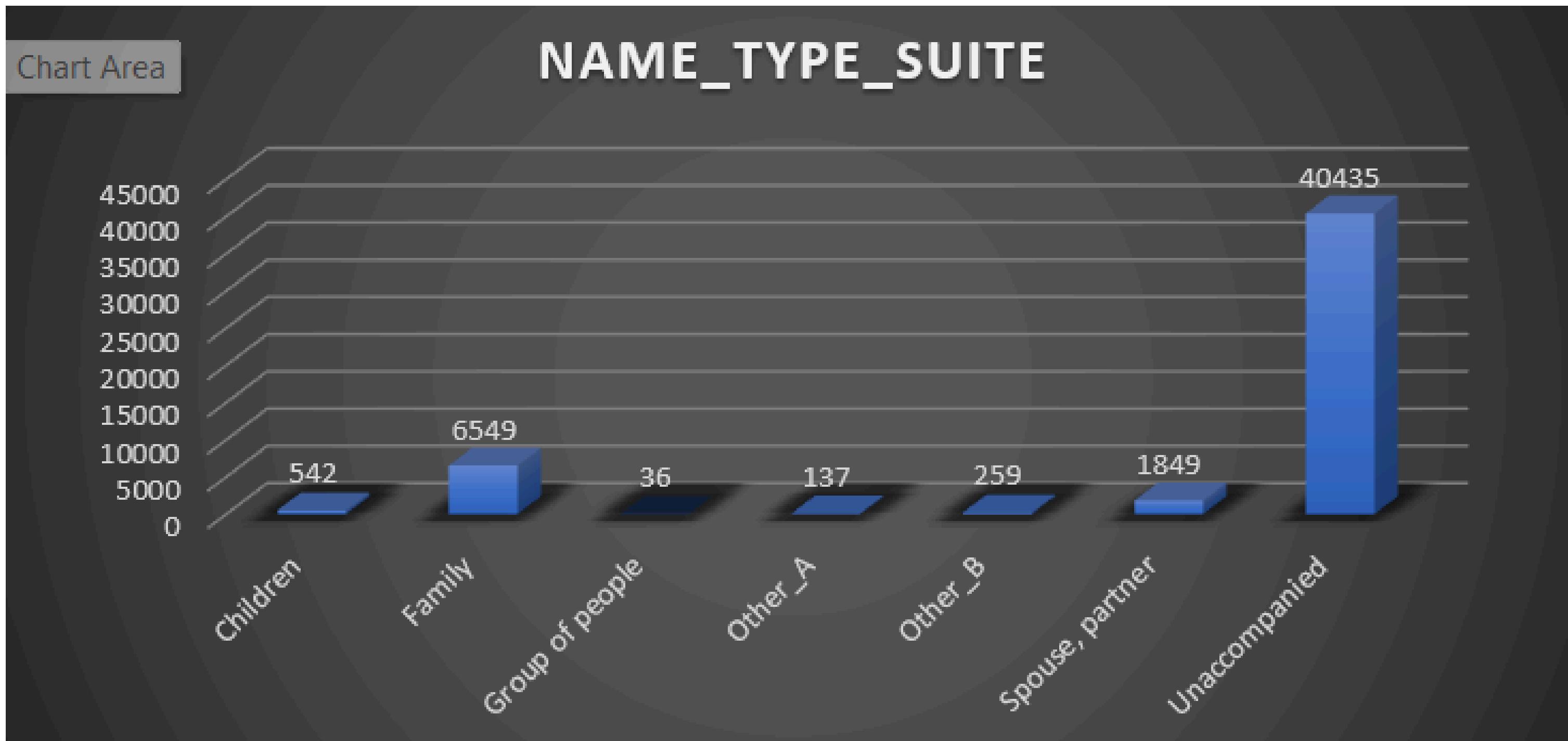
**RESULT:** Replacing Blanks in Occupation\_Type column of the Application Dataset using MODE



Highest Occurring Variable is **Laborers**.

# A. MISSING DATA

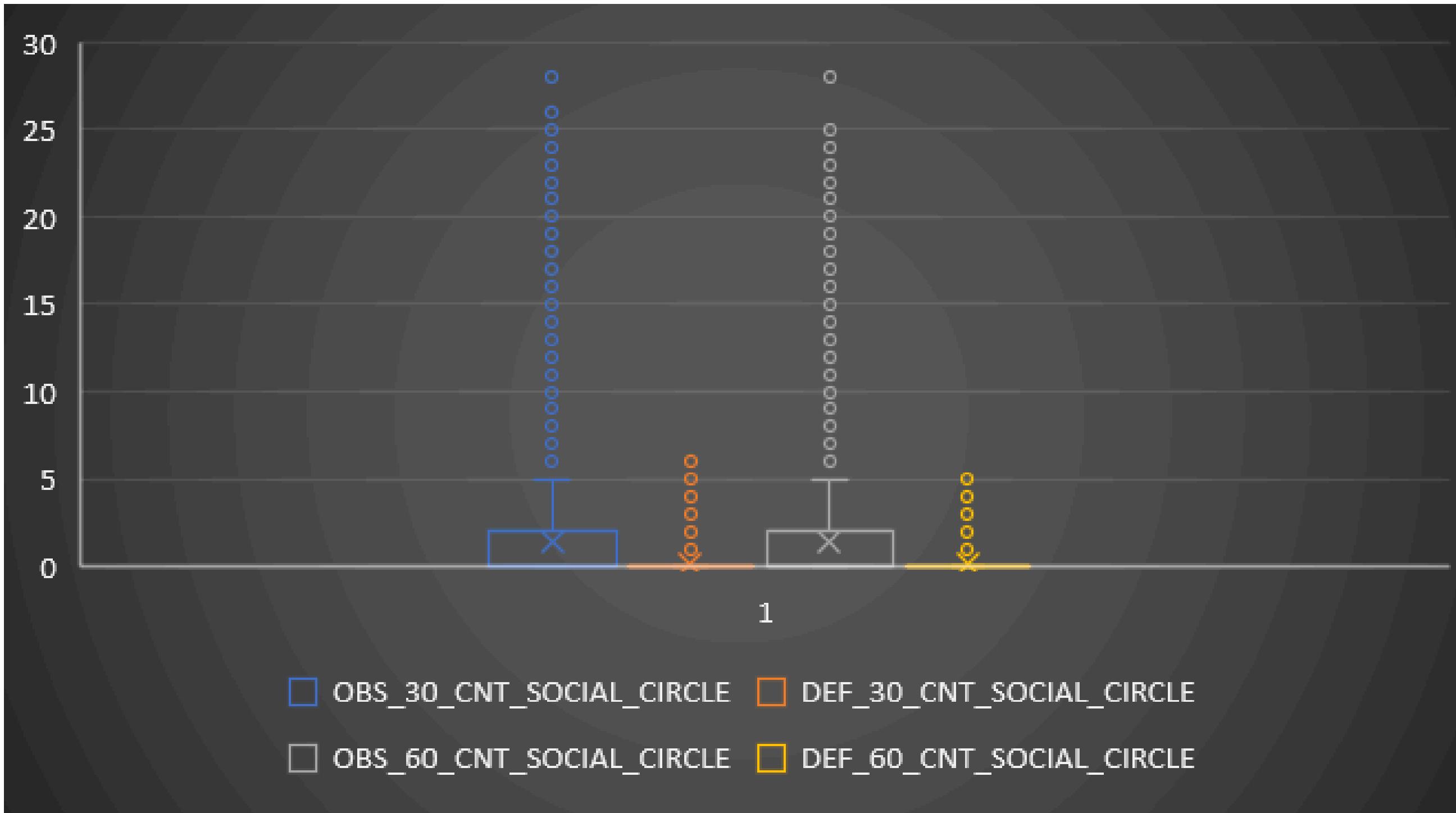
**RESULT:** Replacing Blanks in Name\_Type\_Suite column of the Application Dataset using MODE



Highest Occurring Variable is **Unacccompanied**.

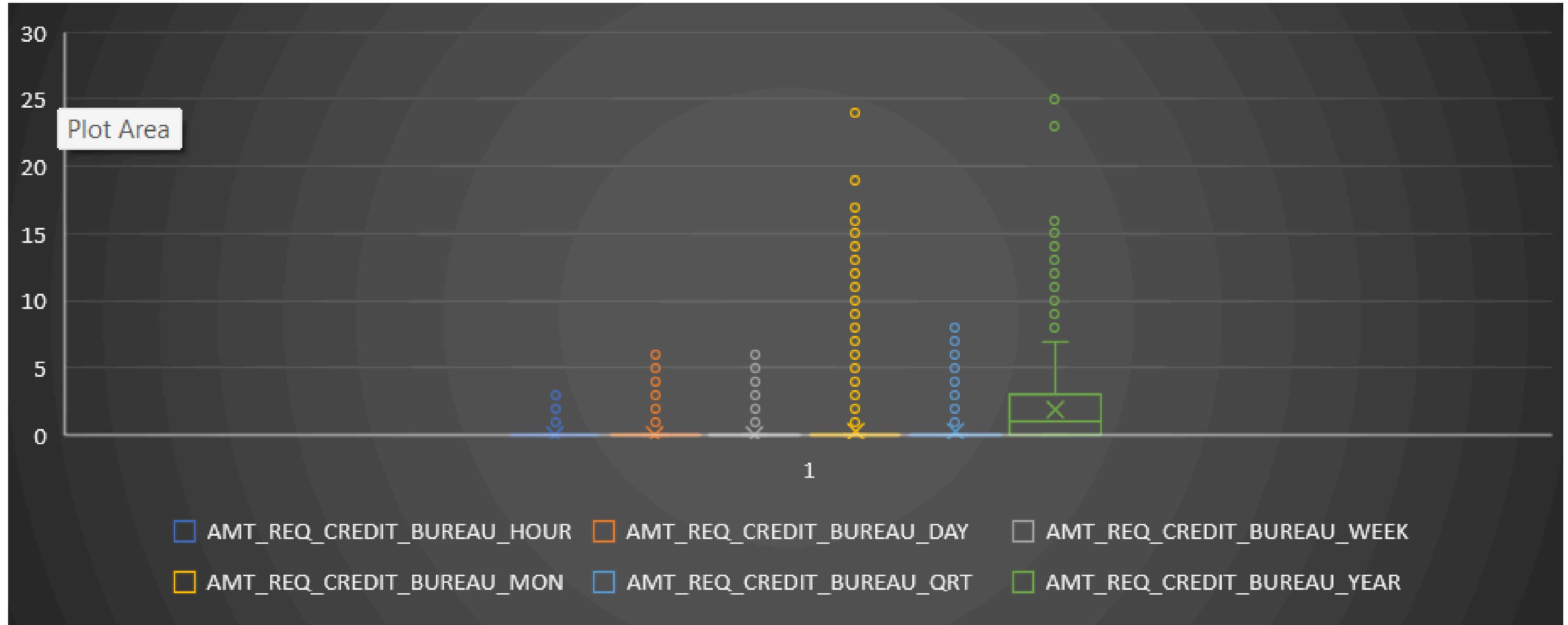
# A. MISSING DATA

**RESULT:** Replacing Blanks in other columns of the Application Dataset using MEDIAN and MODE



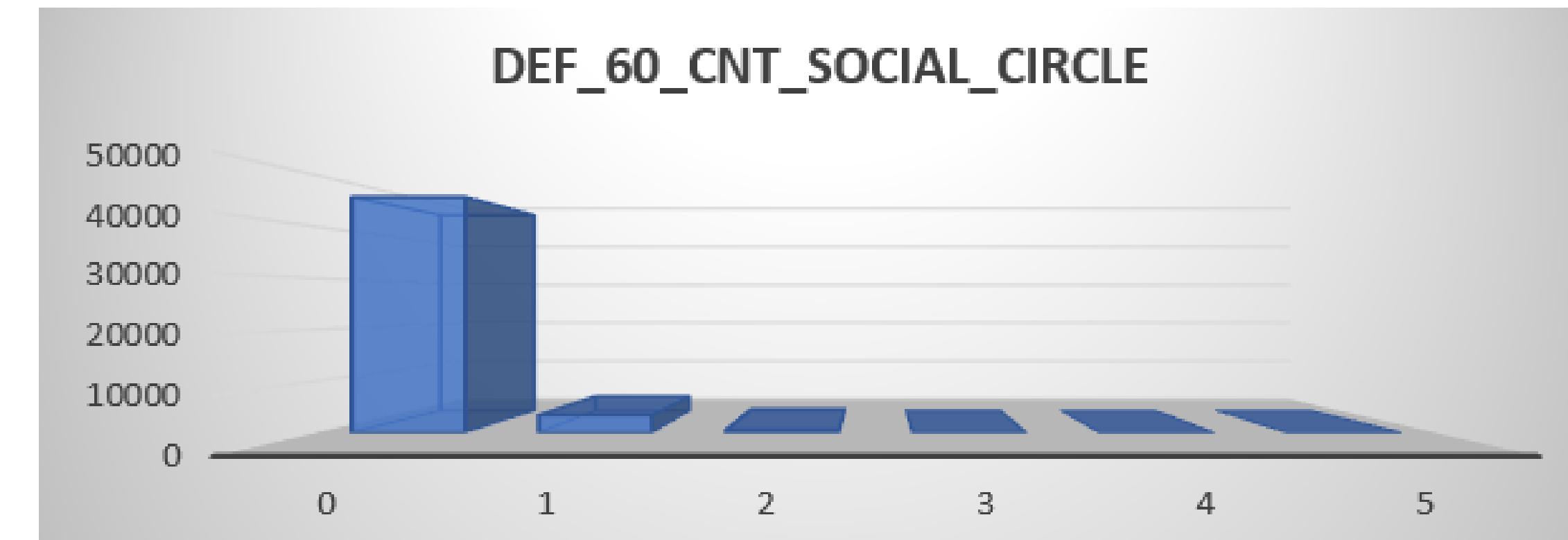
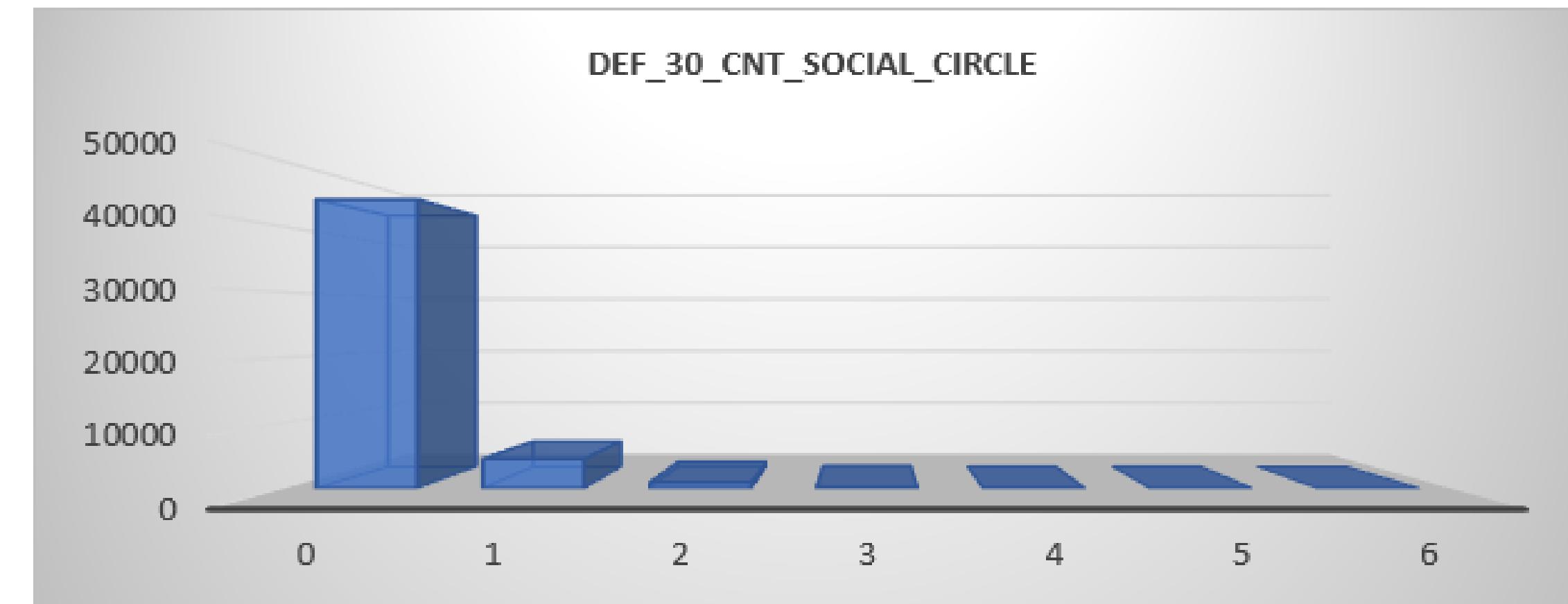
# A. MISSING DATA

**RESULT:** Replacing Blanks in other columns of the Application Dataset using MEDIAN and MODE



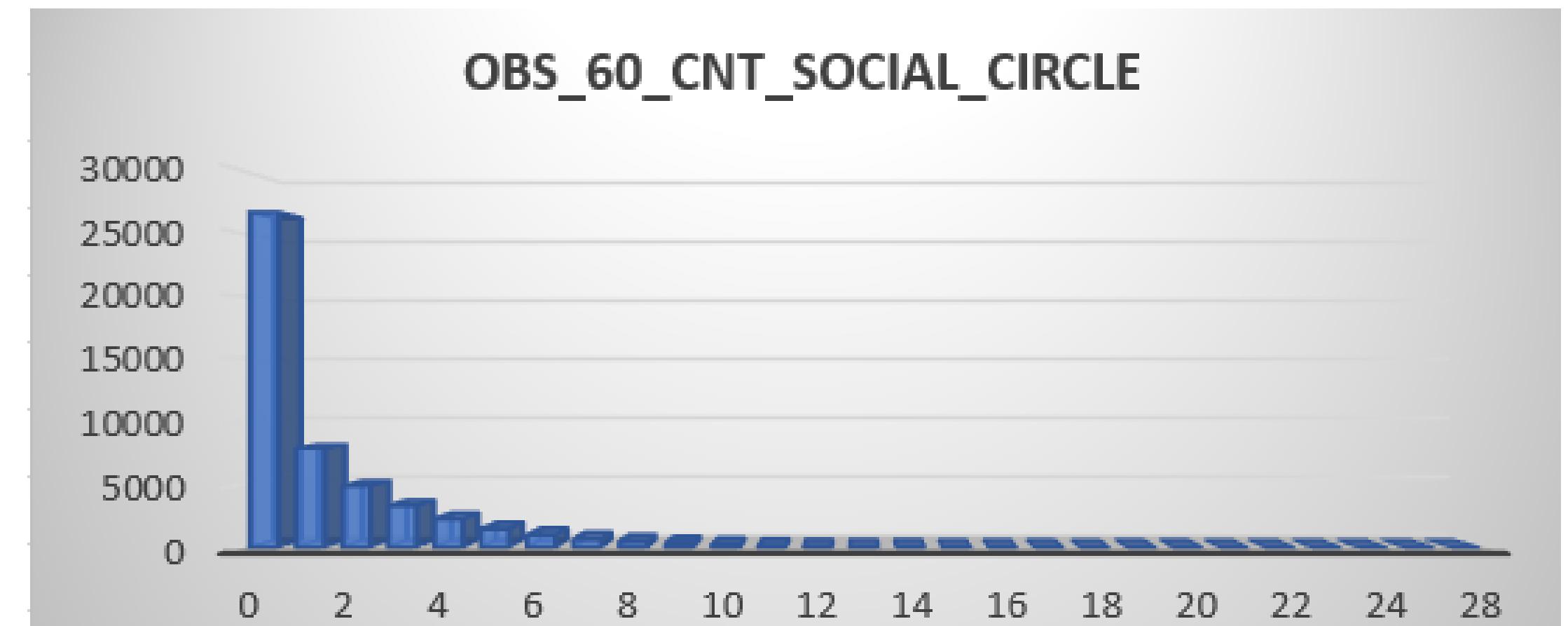
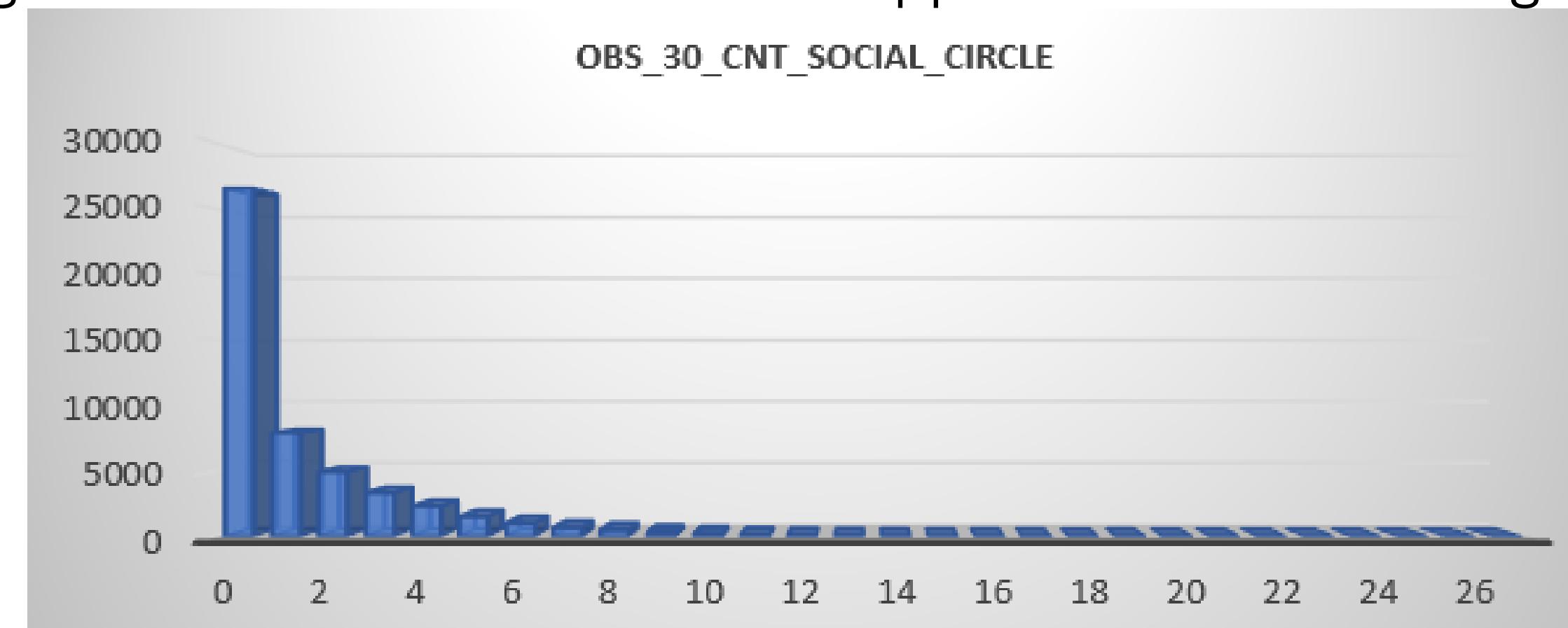
# A. MISSING DATA

**RESULT:** Replacing Blanks in other columns of the Application Dataset using MEDIAN and MODE



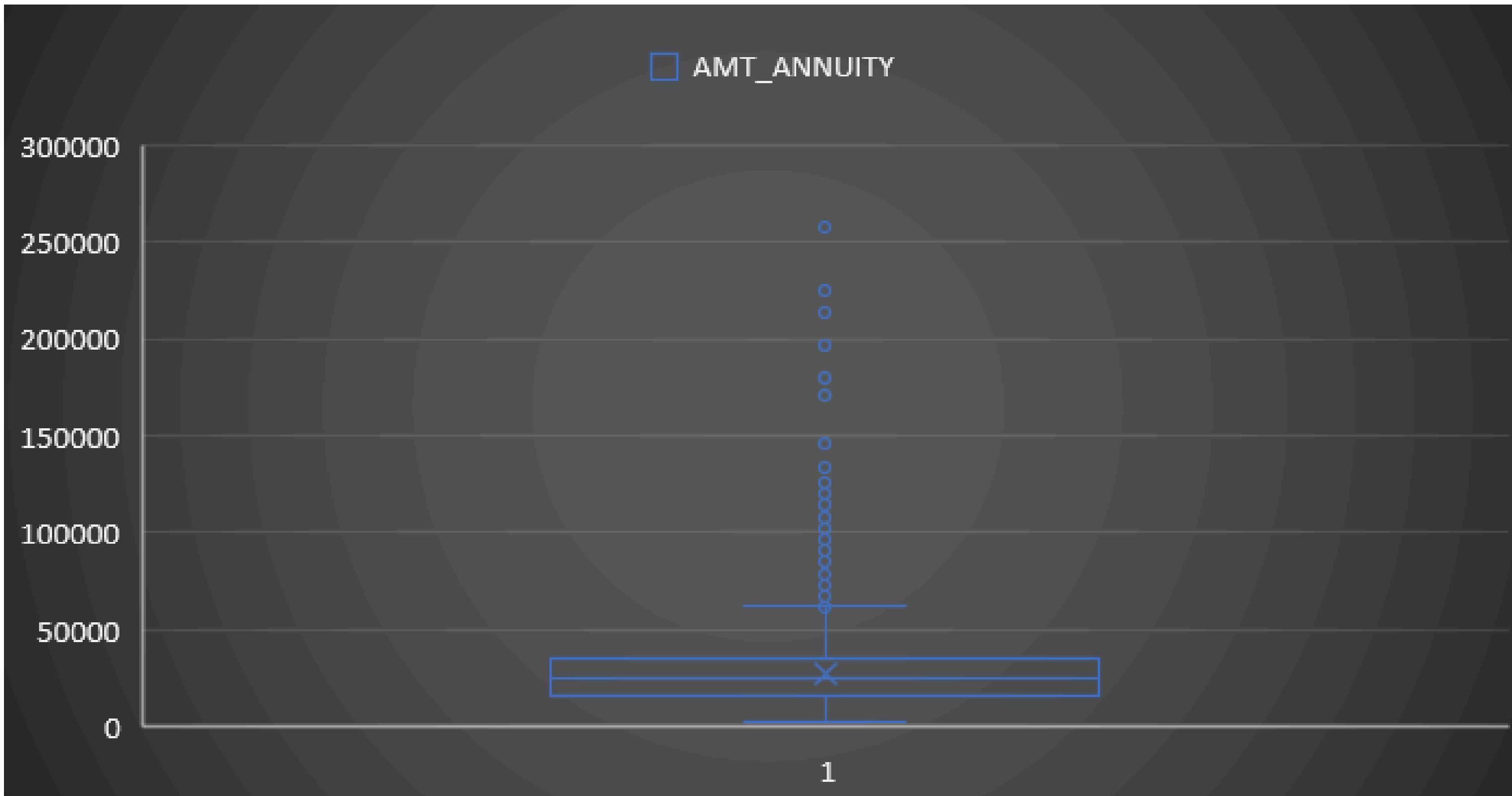
# A. MISSING DATA

**RESULT:** Replacing Blanks in other columns of the Application Dataset using MEDIAN and MODE



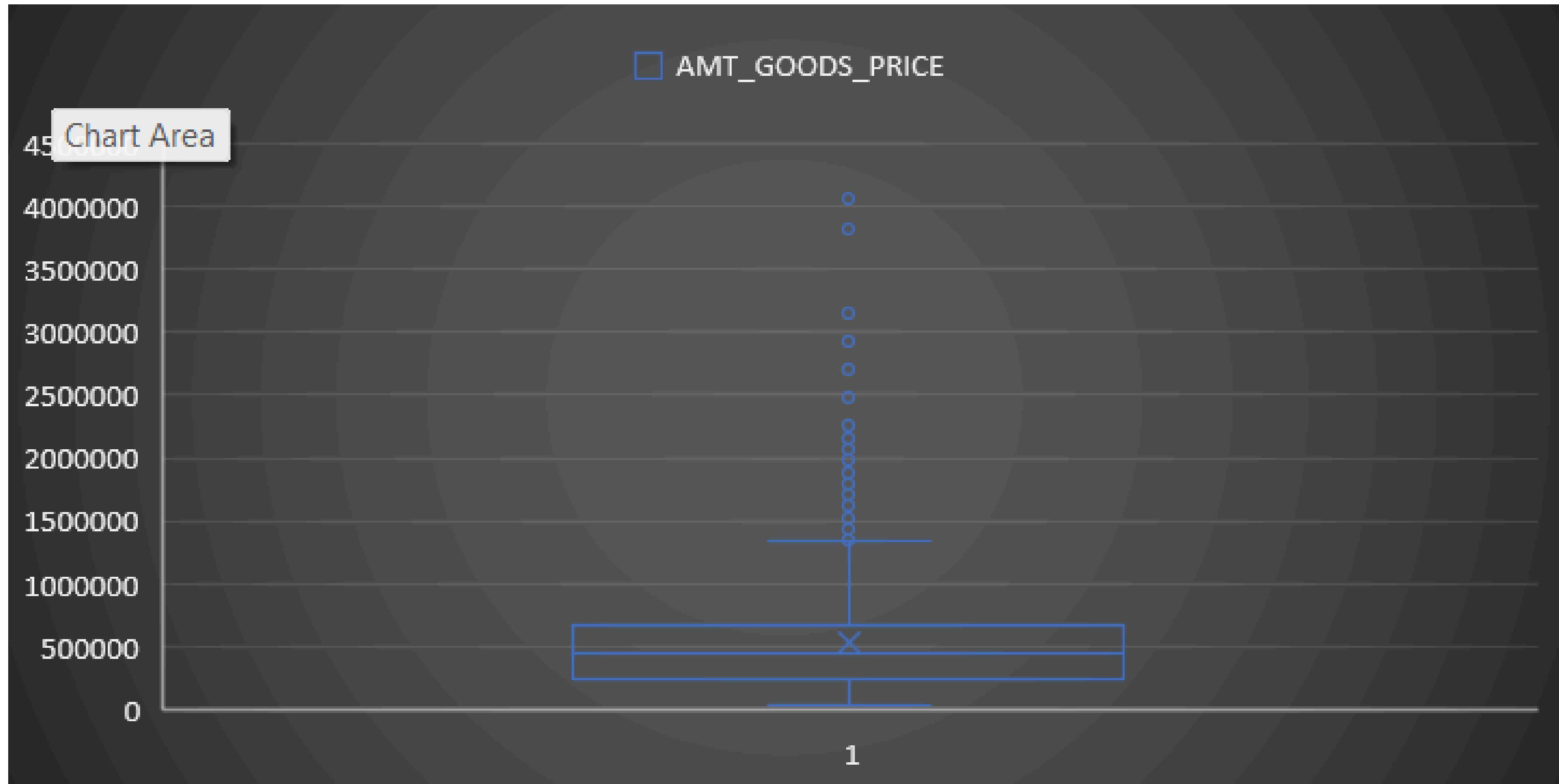
# A. MISSING DATA

**RESULT:** Replacing Blanks in AMT\_ANNUITY column of the Application Dataset with the median



# A. MISSING DATA

**RESULT:** Replacing Blanks in AMT\_GOODS\_PRICE column of the Application Dataset with the median



## B. OUTLIERS:

**Objective:** Outliers can significantly impact the analysis and distort the results. You need to identify outliers in the loan application dataset.

**Your Task:** Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.

### Formulae:

**Quartile 1** :=QUARTILE(A:A,1)

**Quartile 3** :=QUARTILE(A:A,3)

**IQR** = Quartile 3 - Quartile 1

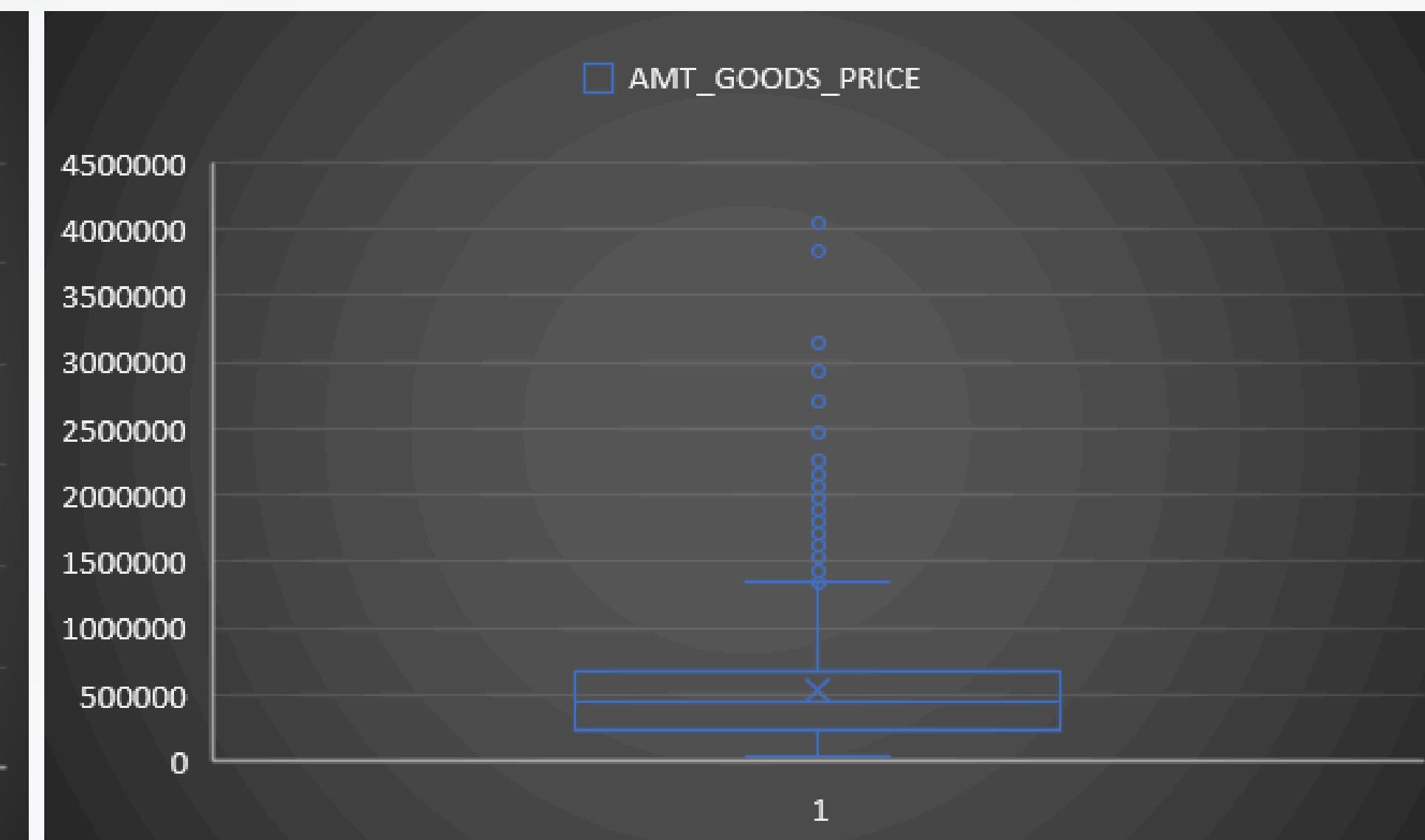
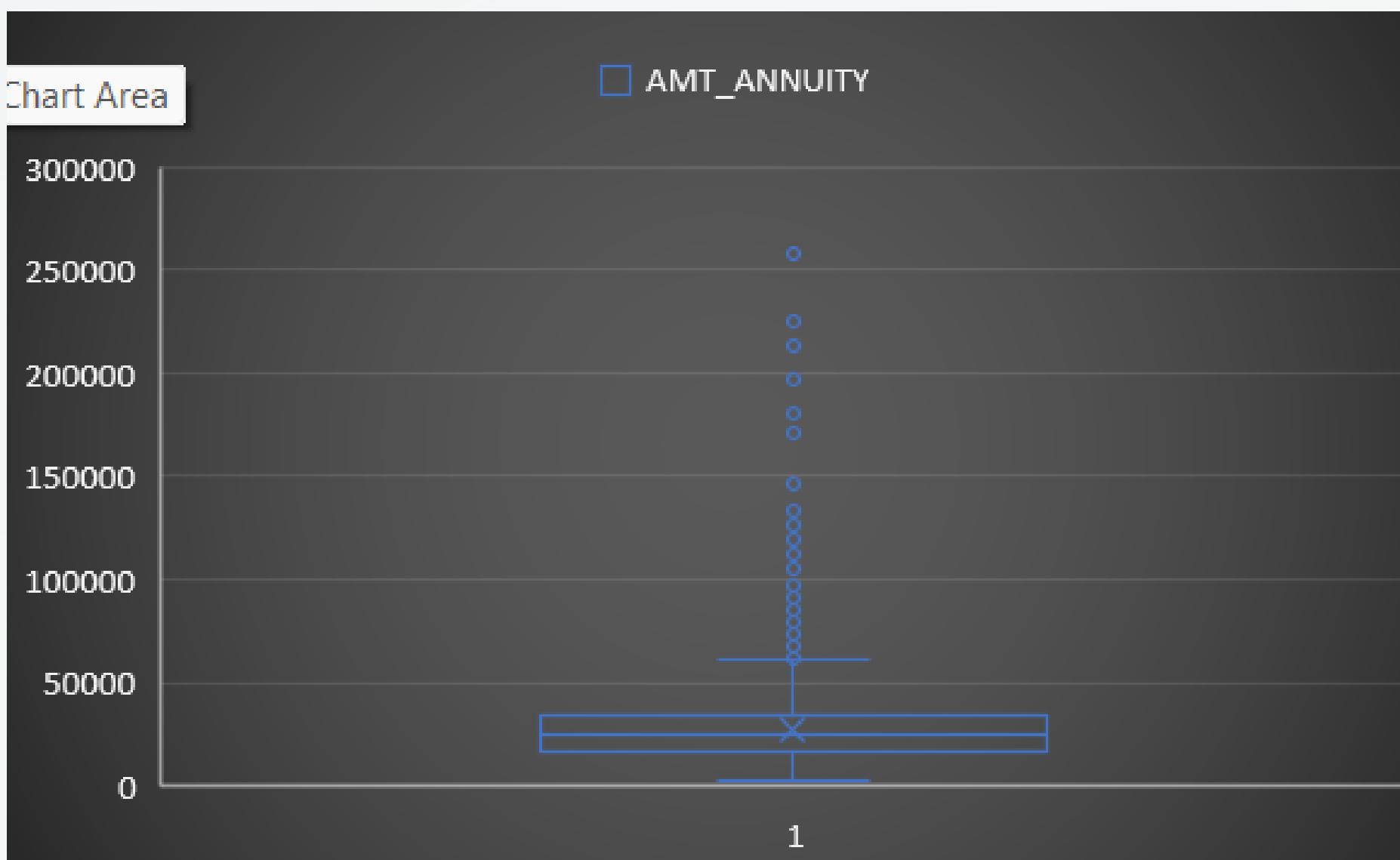
**Upper Limit** = Quartile 3 + 1.5\*IQR

**Lower Limit** = Quartile 1 - 1.5IQR

| AMT_INCOME_TOTAL |        |
|------------------|--------|
| Quartile 1       | 112500 |
| Quartile 3       | 202500 |
| IQR              | 90000  |
| Upper Limit      | 337500 |
| Lower Limit      | -22500 |

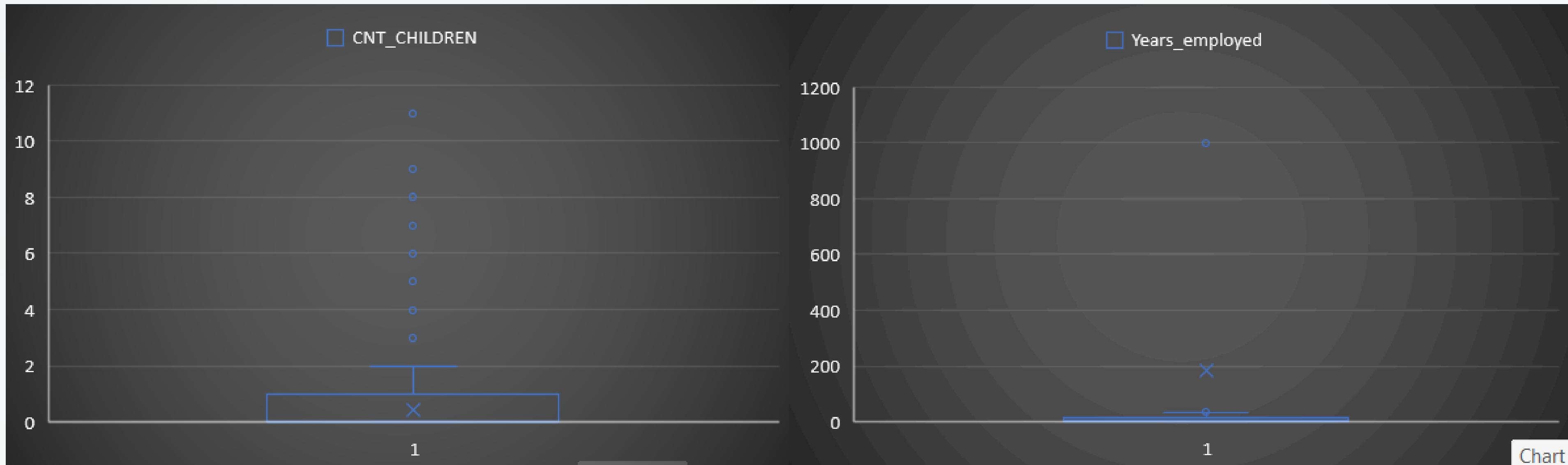
# B. OUTLIERS:

## RESULT:



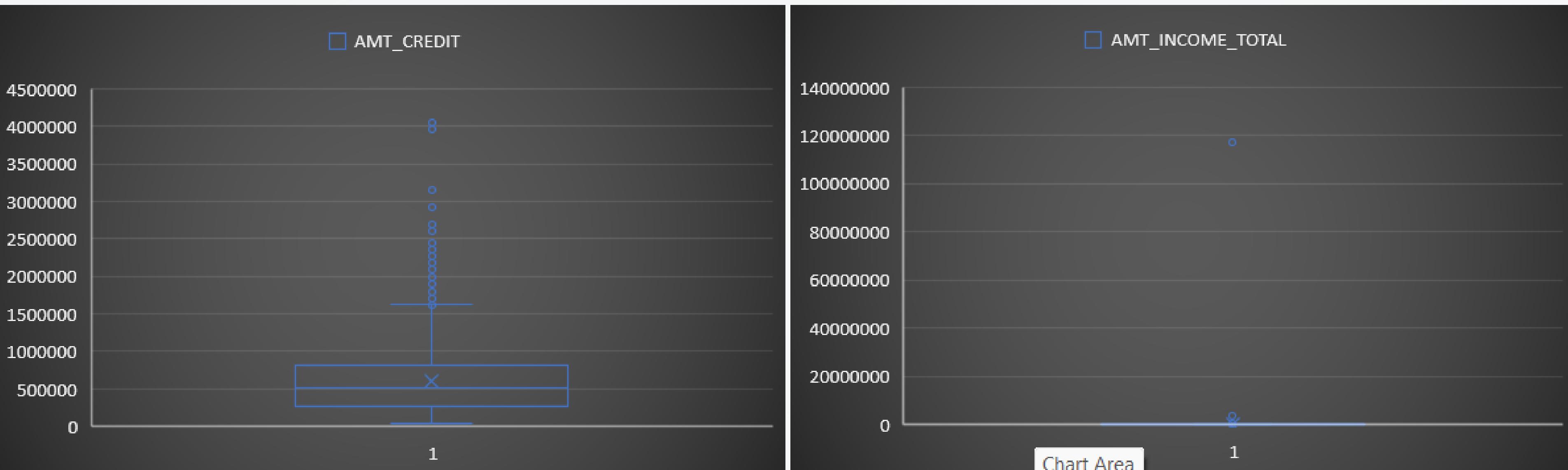
# B. OUTLIERS:

## RESULT:



# B. OUTLIERS:

## RESULT:

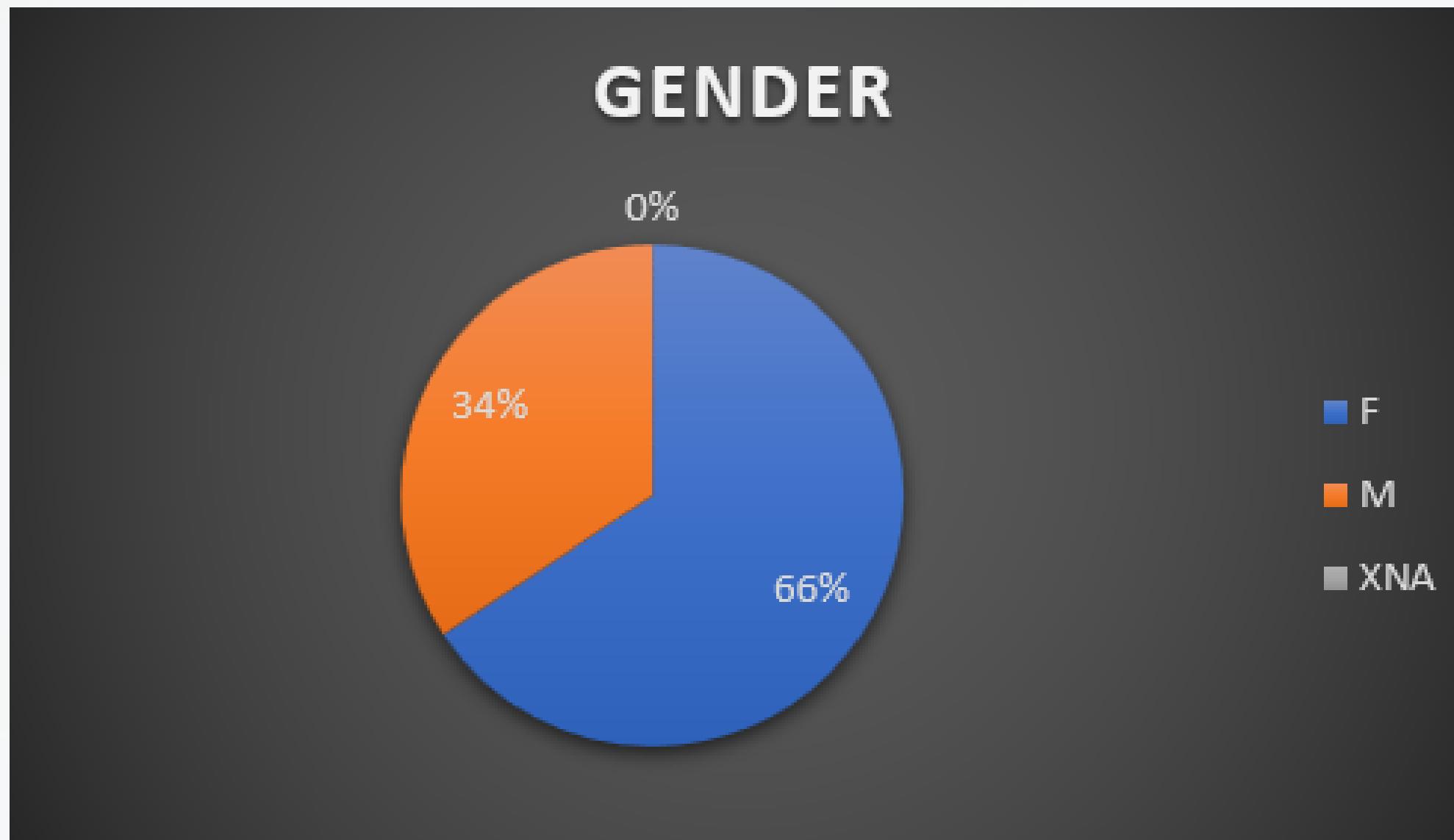


# C. DATA IMBALANCE

**Objective:** Data imbalance can affect the accuracy of the analysis, especially for binary classification problems. Understanding the data distribution is crucial for building reliable models.

**Your Task:** Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.

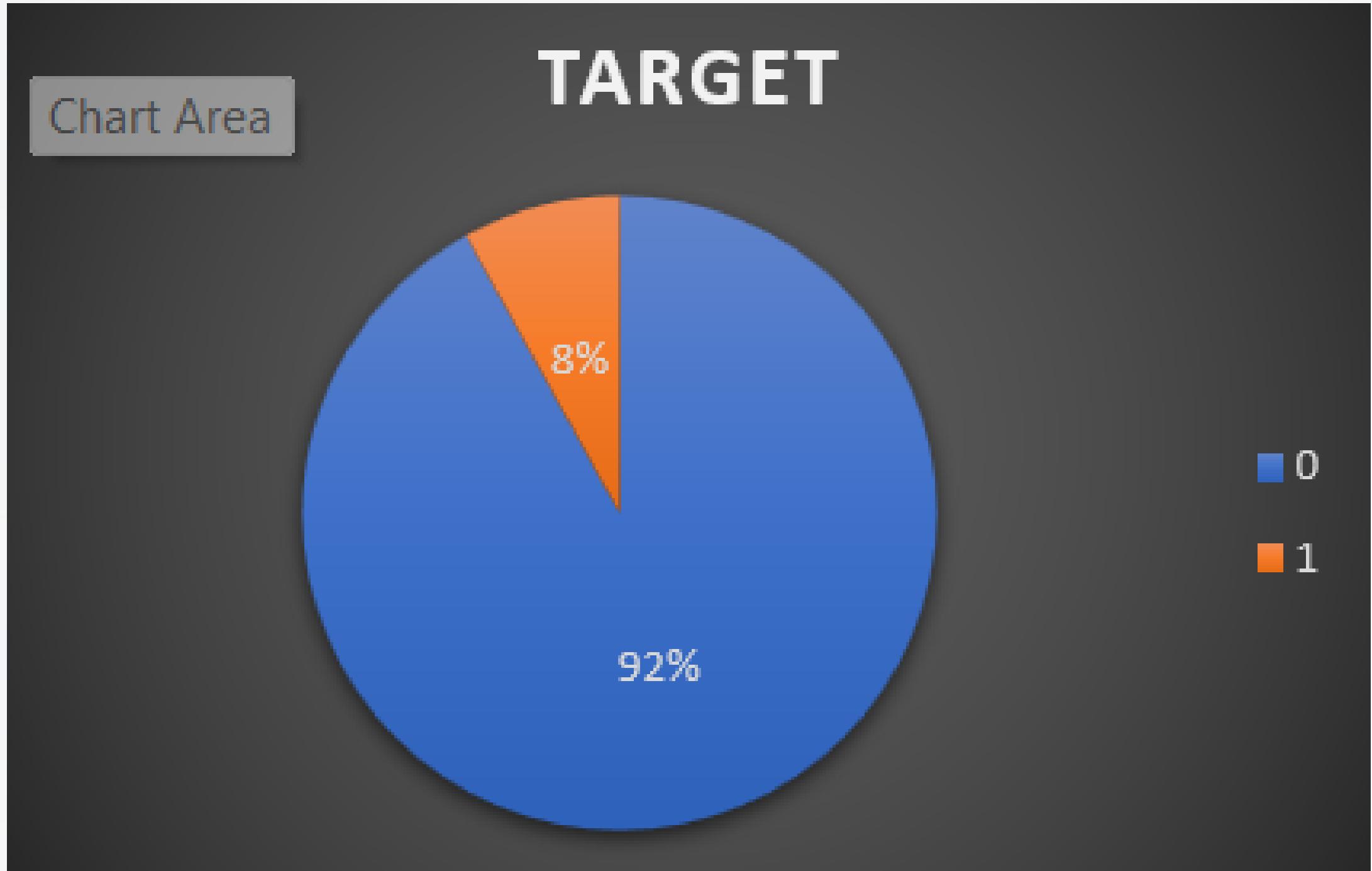
**RESULT:**



| GENDER             | Count of CODE_GENDER |
|--------------------|----------------------|
| F                  | 32823                |
| M                  | 17174                |
| XNA                | 2                    |
| <b>Grand Total</b> | <b>49999</b>         |

# C. DATA IMBALANCE

**RESULT:**



| Row Labels         | Count of TARGET |
|--------------------|-----------------|
| 0                  | 45973           |
| 1                  | 4026            |
| <b>Grand Total</b> | <b>49999</b>    |

## D. UNIVARIATE ANALYSIS:

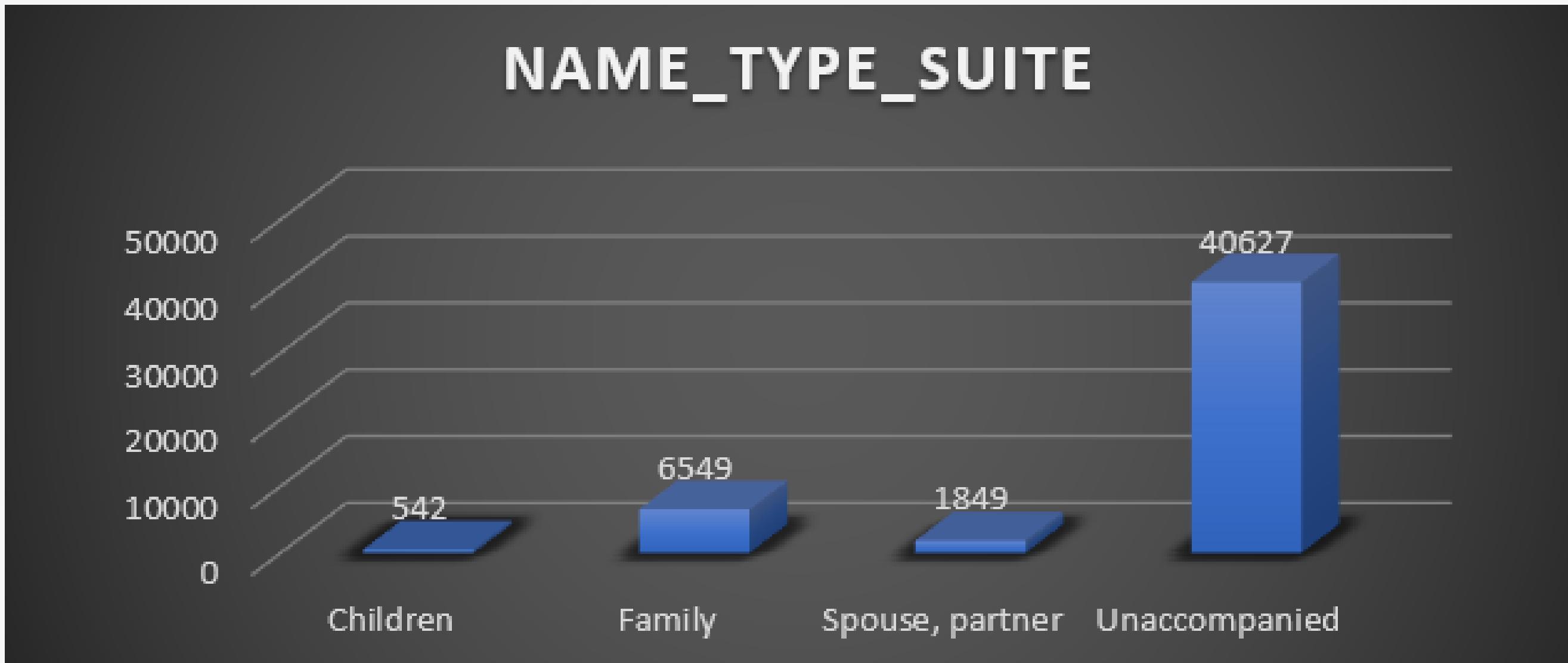
**Objective:** To gain insights into the driving factors of loan default, it is important to conduct various analyses on consumer and loan attributes.

**Your Task:** Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.

# D. UNIVARIATE ANALYSIS:

**RESULT:**

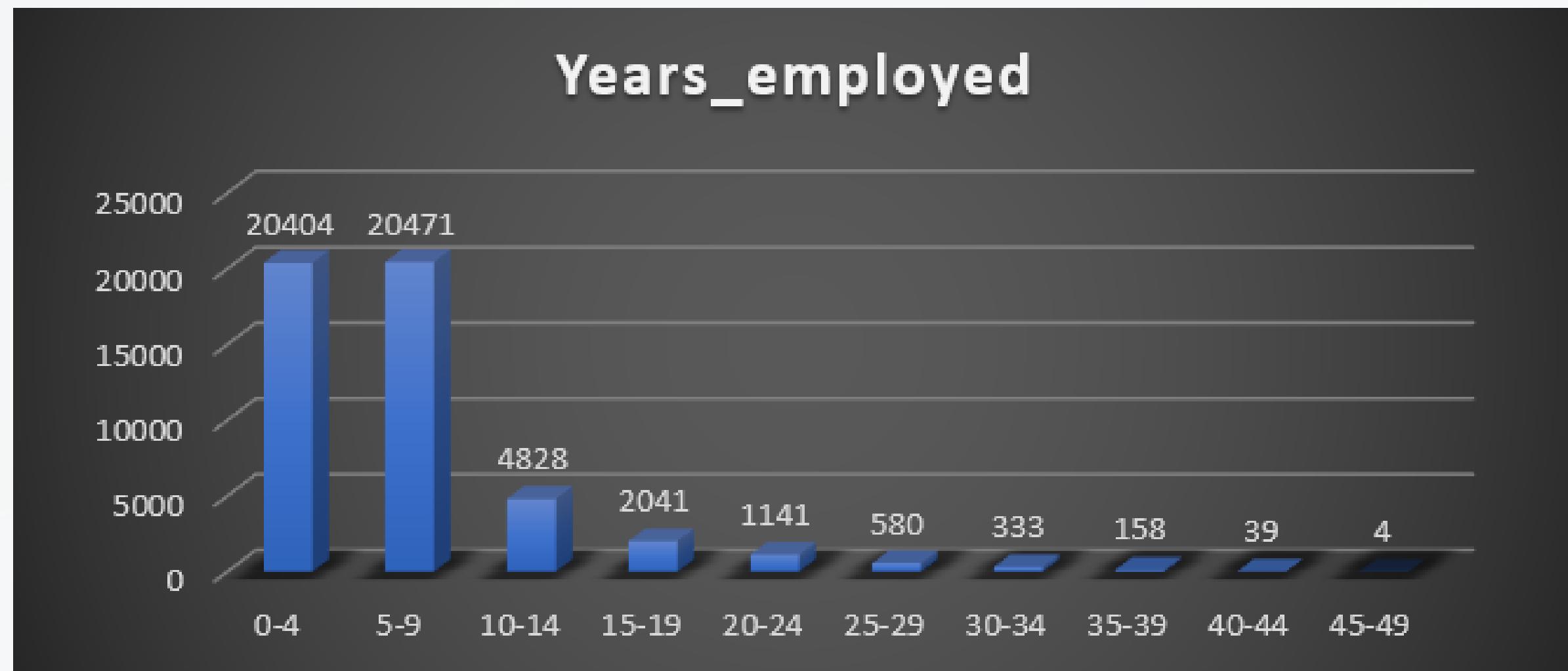
| Row Labels         | Count of NAME_TYPE_SUITE |
|--------------------|--------------------------|
| Children           | 542                      |
| Family             | 6549                     |
| Spouse, partner    | 1849                     |
| Unaccompanied      | 40627                    |
| <b>Grand Total</b> | <b>49567</b>             |



# D. UNIVARIATE ANALYSIS:

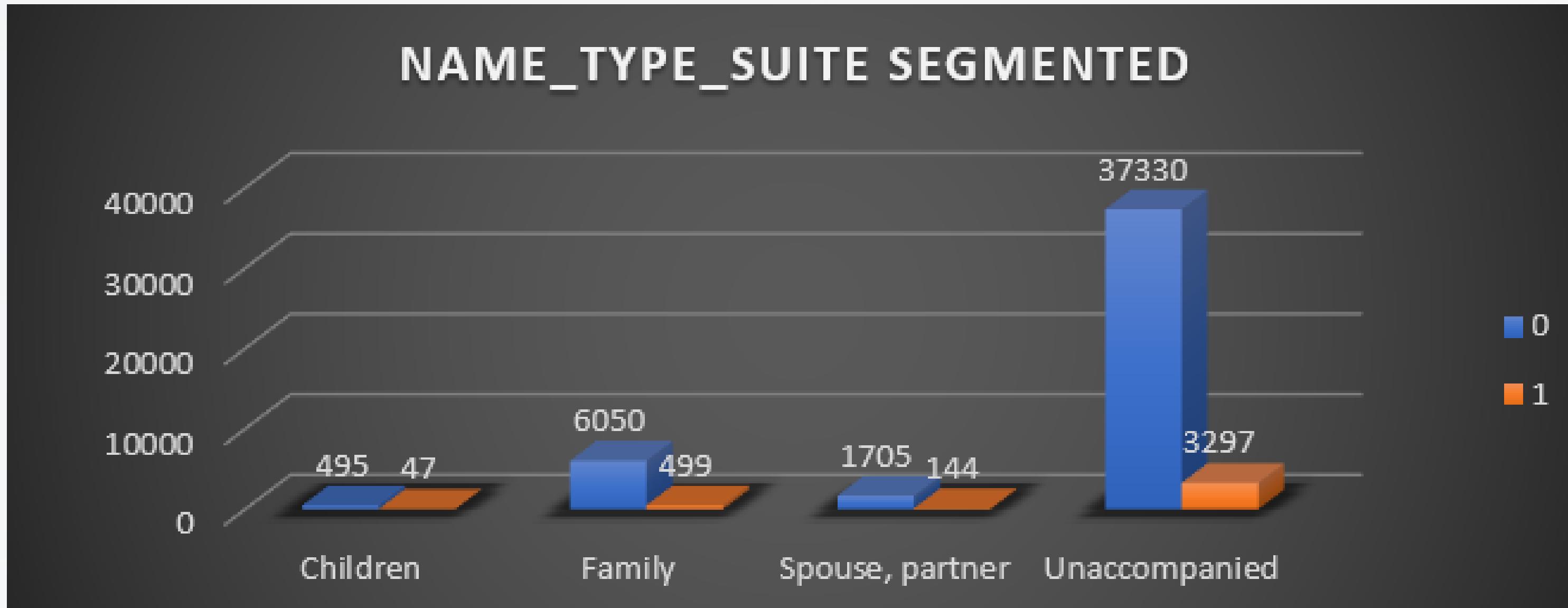
## RESULT:

| Row Labels         | Count of Years_employed |
|--------------------|-------------------------|
| 0-4                | 20404                   |
| 5-9                | 20471                   |
| 10-14              | 4828                    |
| 15-19              | 2041                    |
| 20-24              | 1141                    |
| 25-29              | 580                     |
| 30-34              | 333                     |
| 35-39              | 158                     |
| 40-44              | 39                      |
| 45-49              | 4                       |
| <b>Grand Total</b> | <b>49999</b>            |



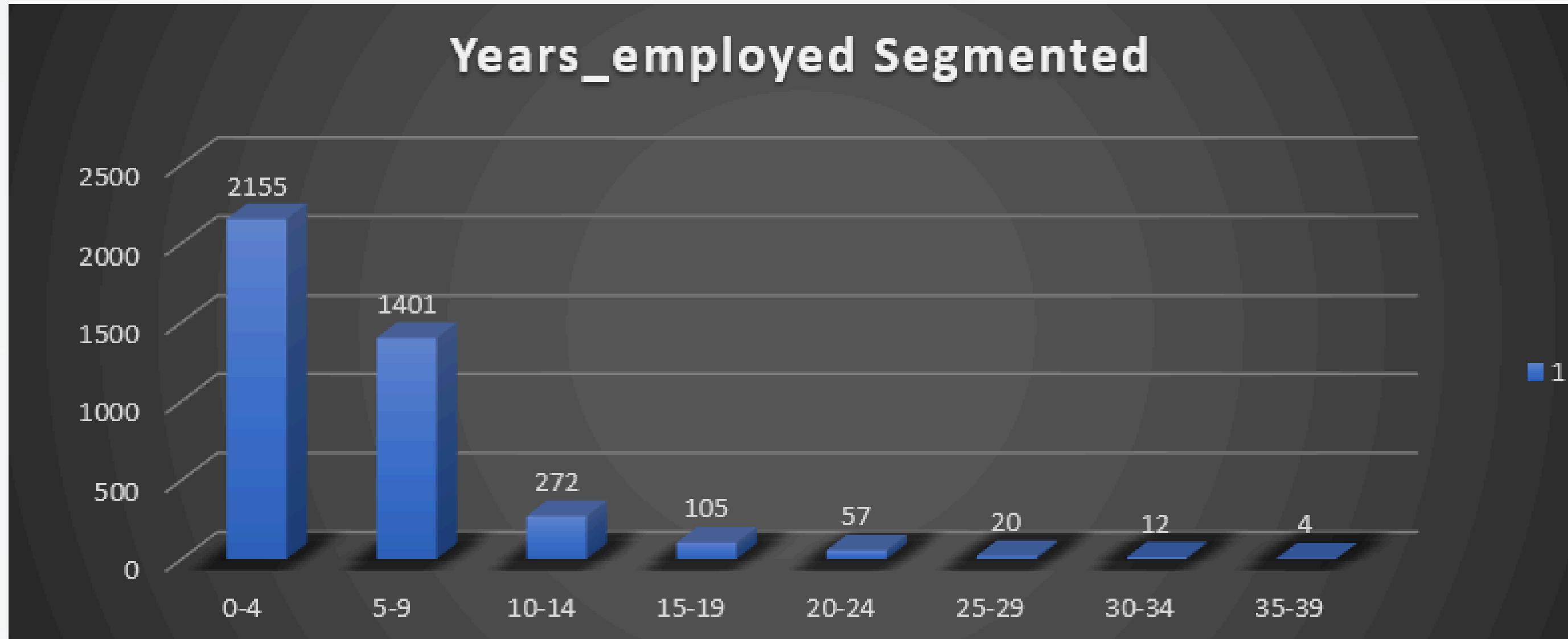
# D. UNIVARIATE ANALYSIS:

**RESULT:**



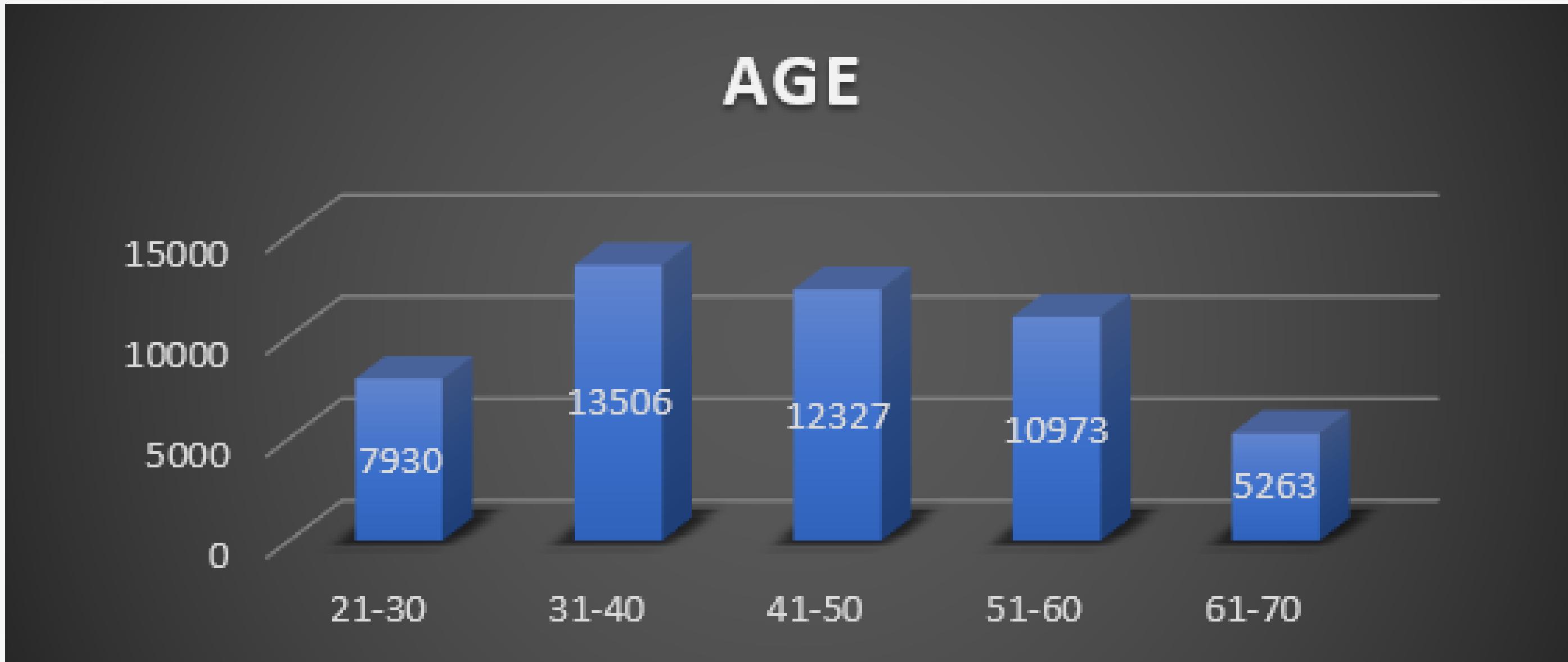
# D. UNIVARIATE ANALYSIS:

**RESULT:**



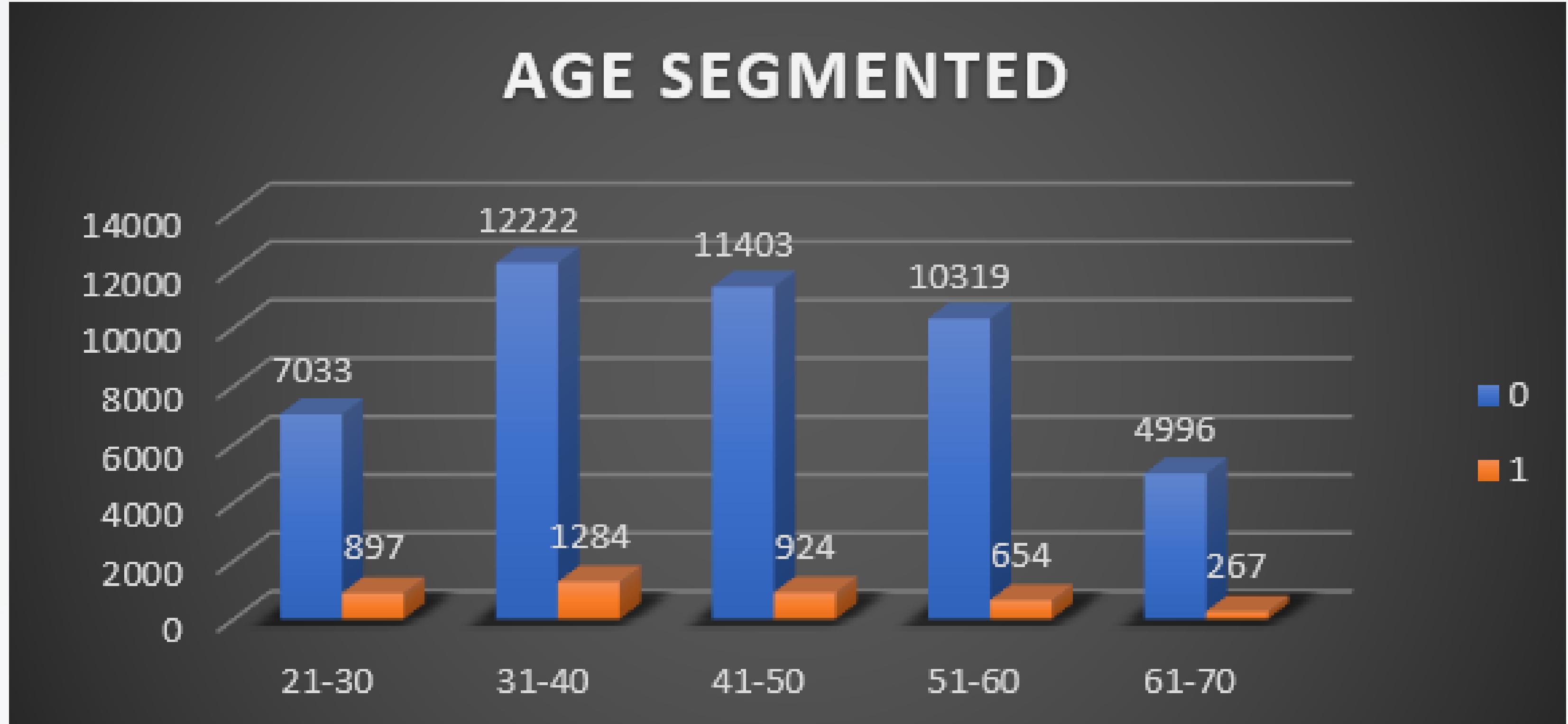
# D. UNIVARIATE ANALYSIS:

**RESULT:**



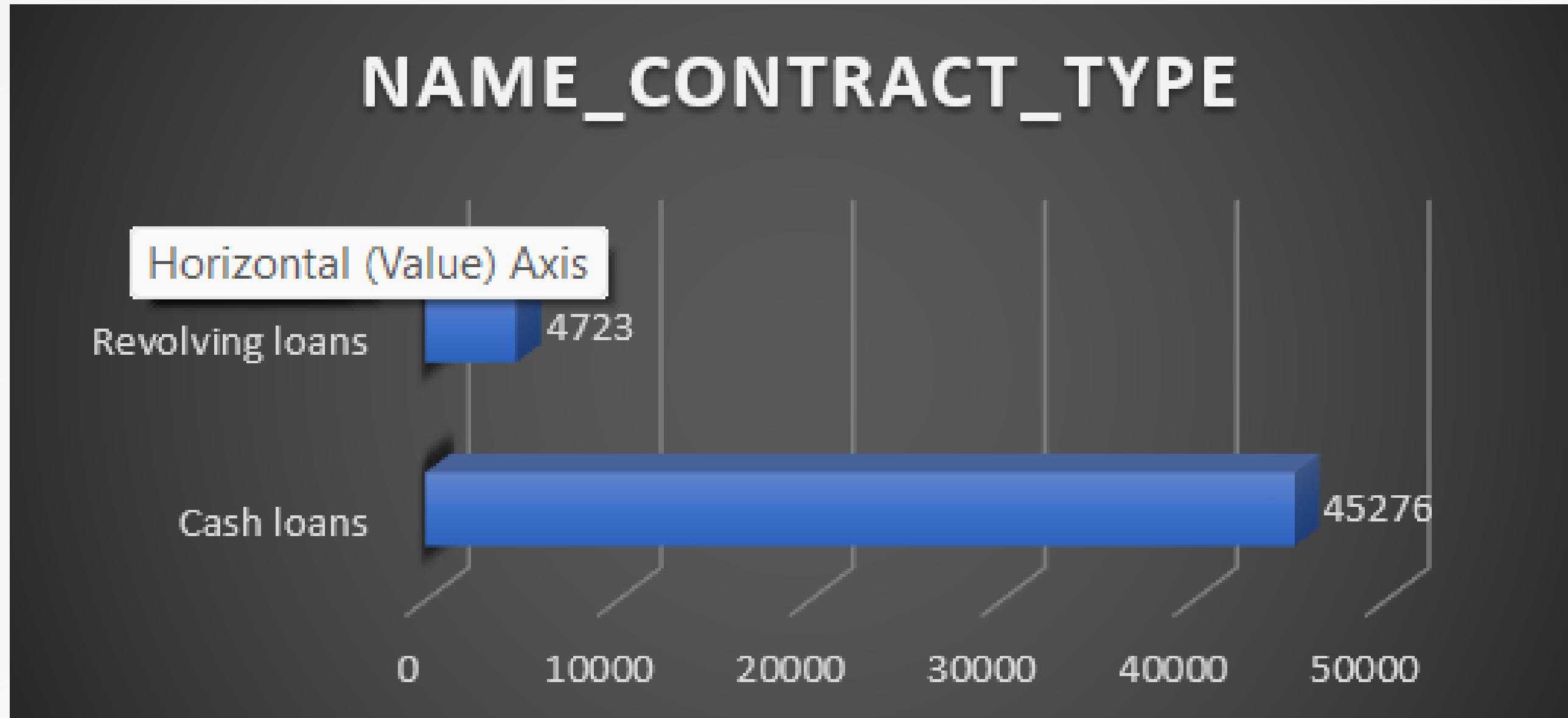
# D. UNIVARIATE ANALYSIS:

**RESULT:**



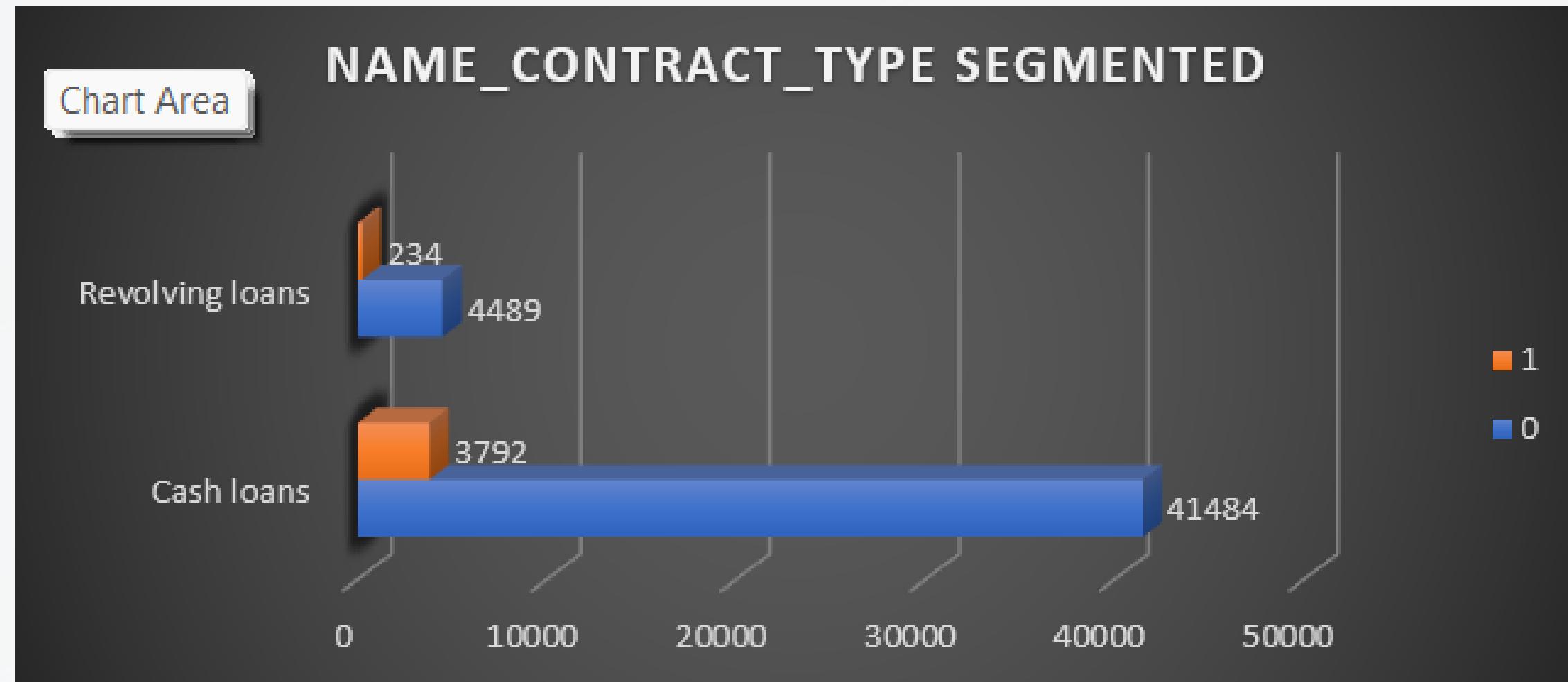
# D. UNIVARIATE ANALYSIS:

**RESULT:**



# D. UNIVARIATE ANALYSIS:

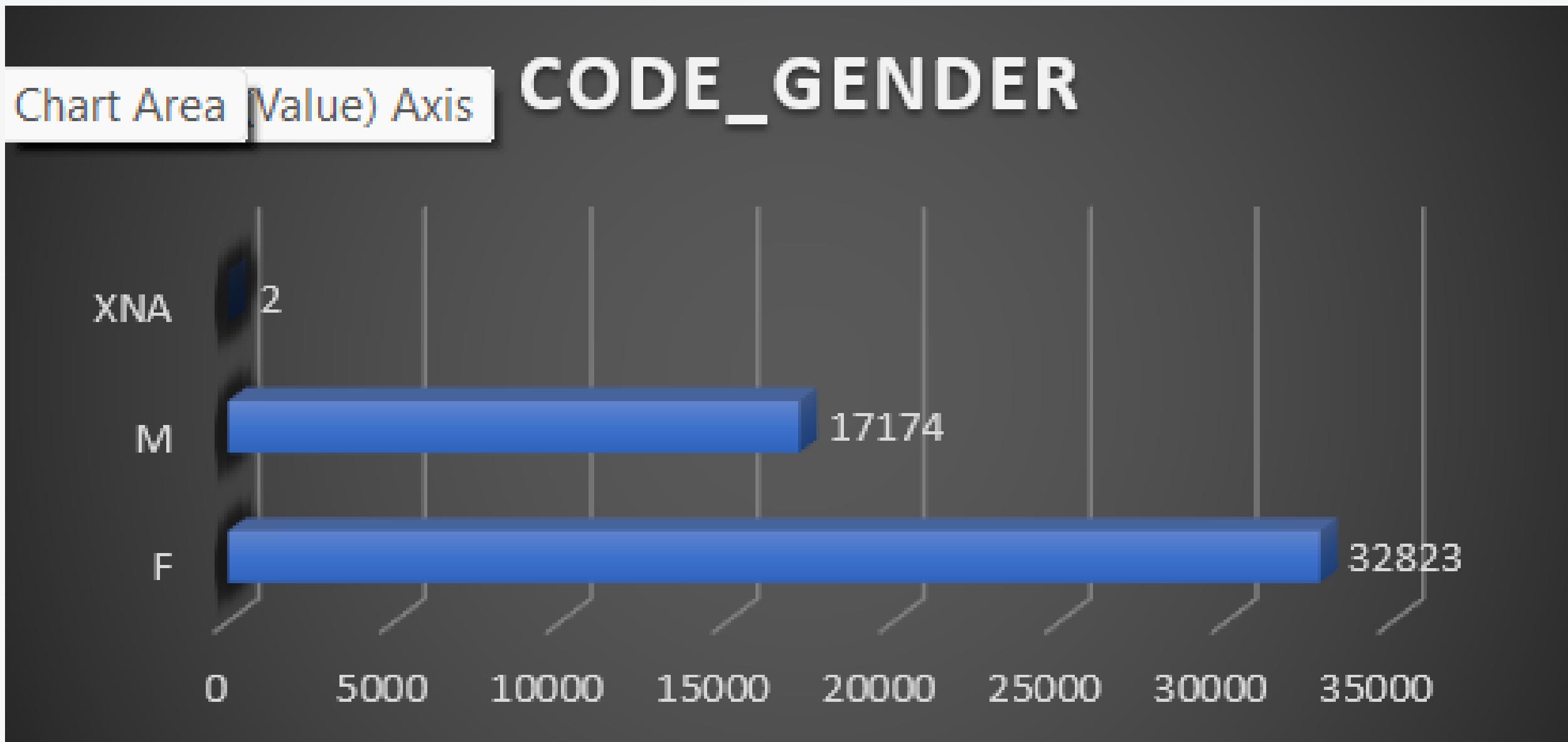
**RESULT:**



# D. UNIVARIATE ANALYSIS:

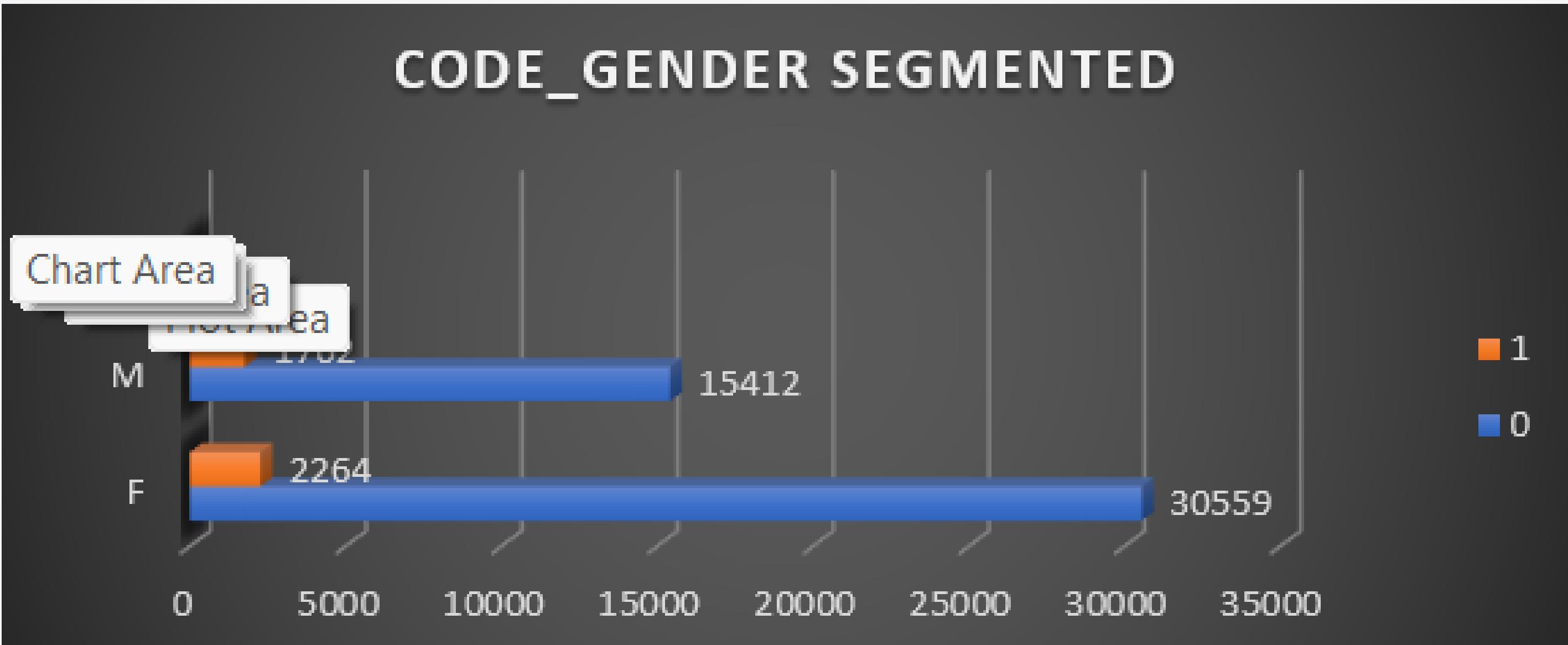
**RESULT:**

| Row Labels         | Count of CODE_GENDER |
|--------------------|----------------------|
| F                  | 32823                |
| M                  | 17174                |
| XNA                | 2                    |
| <b>Grand Total</b> | <b>49999</b>         |



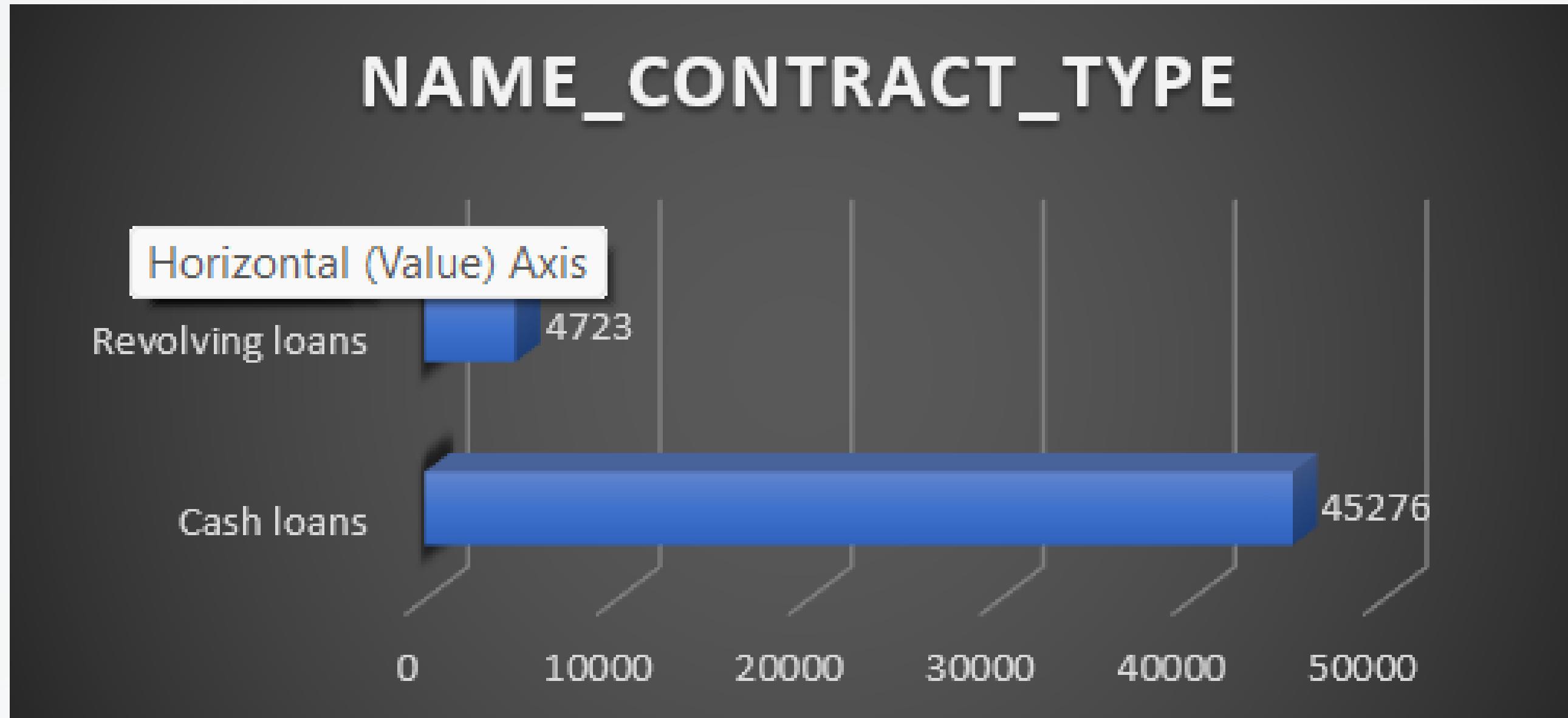
# D. UNIVARIATE ANALYSIS:

**RESULT:**



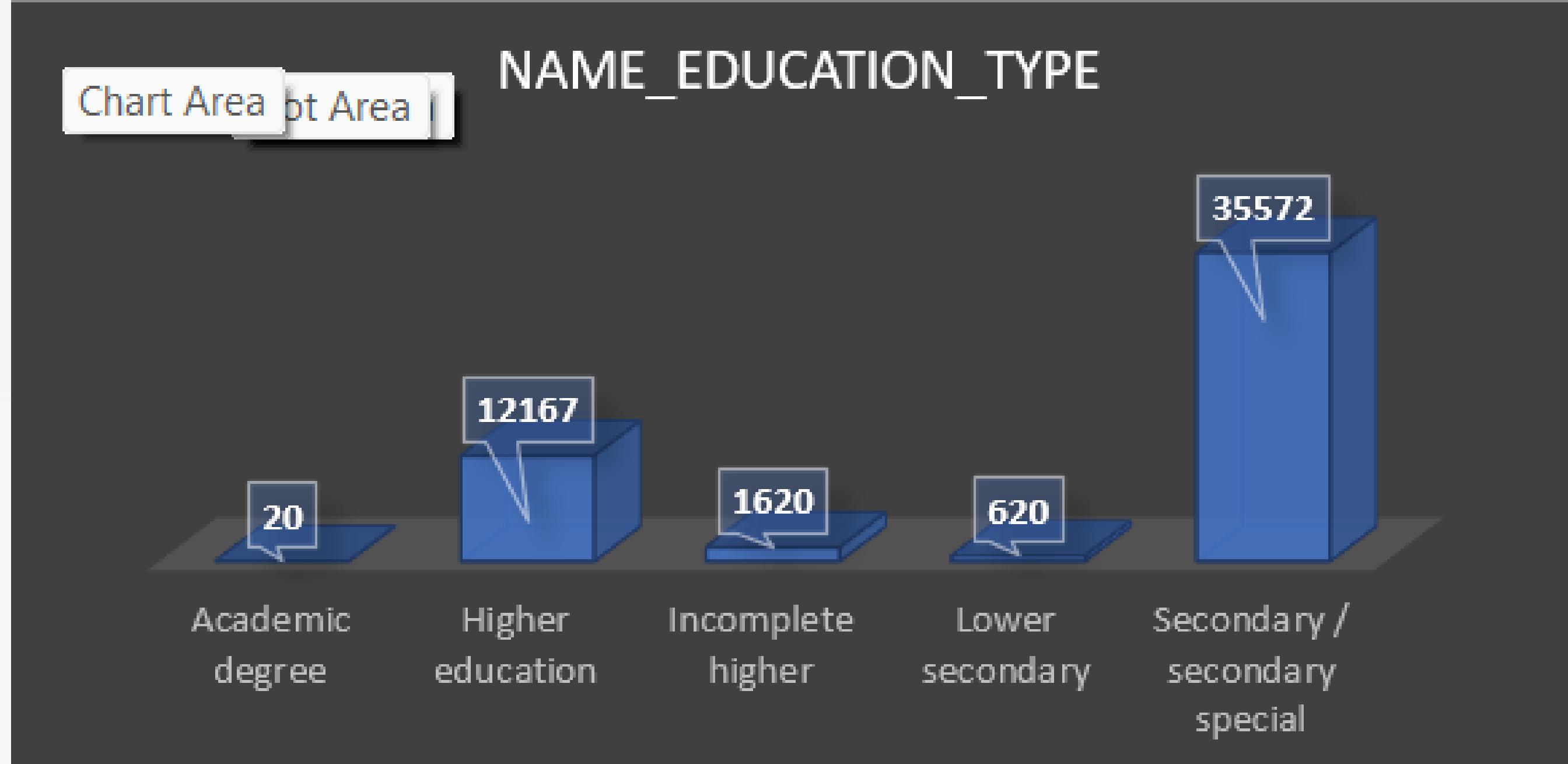
# D. UNIVARIATE ANALYSIS:

**RESULT:**



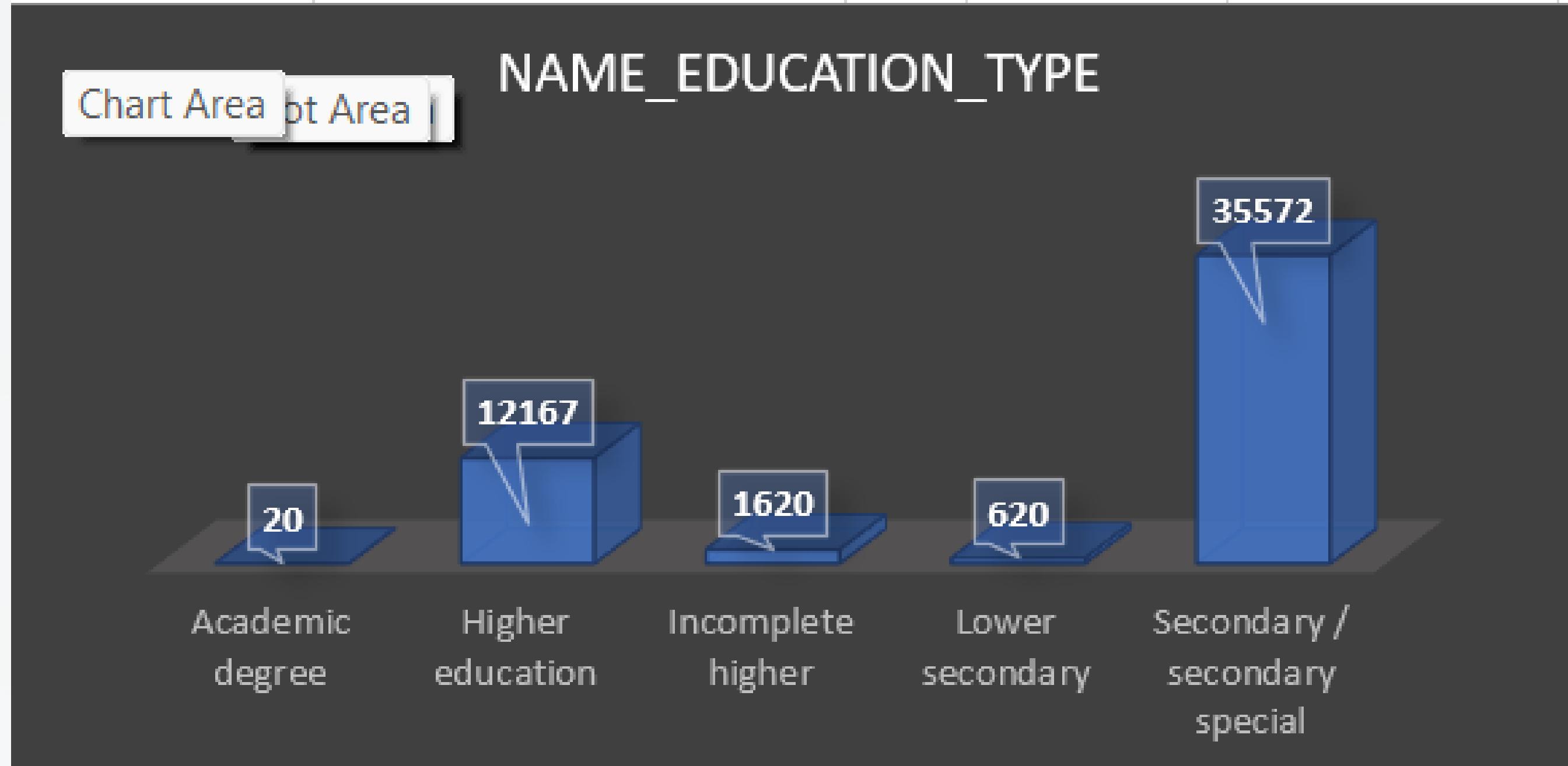
# D. UNIVARIATE ANALYSIS:

**RESULT:**



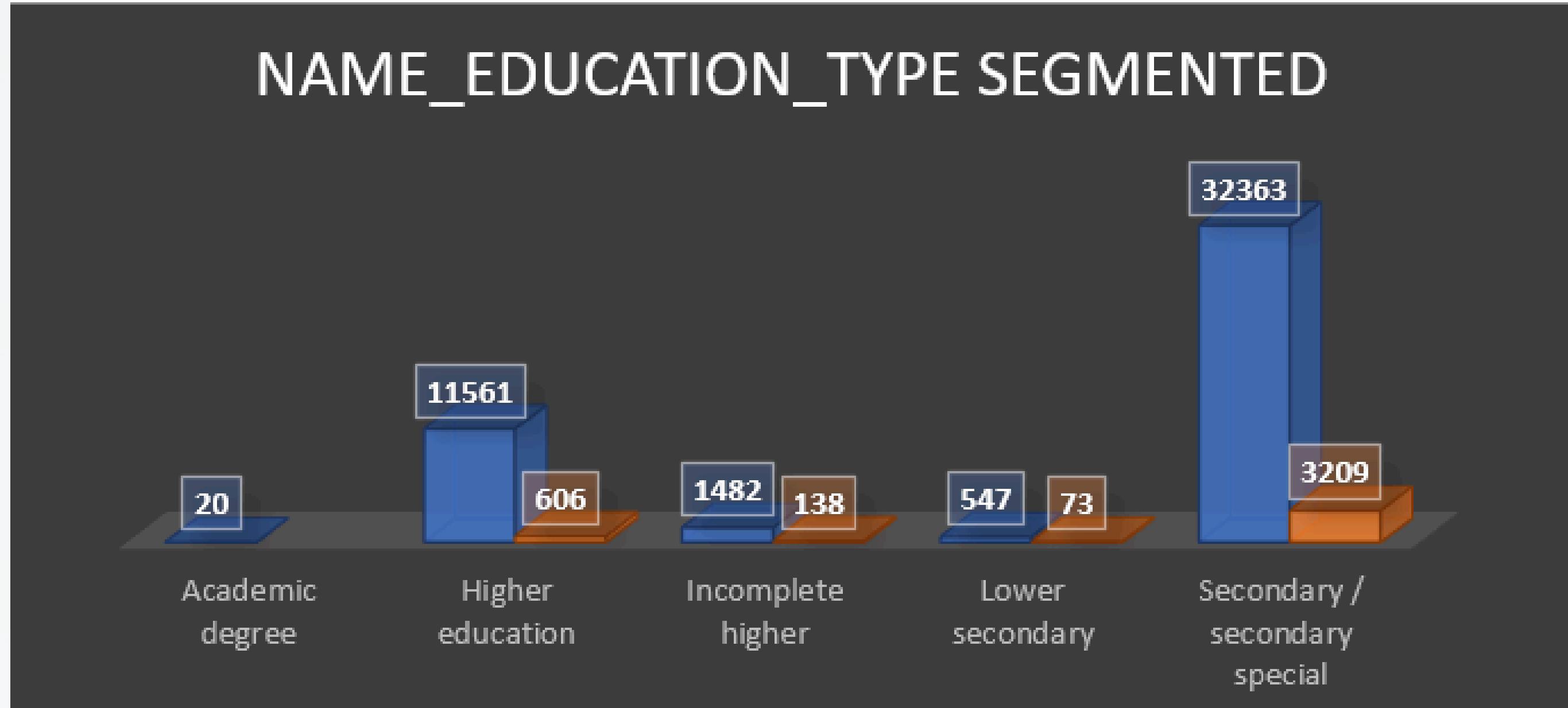
# D. UNIVARIATE ANALYSIS:

## **RESULT:**



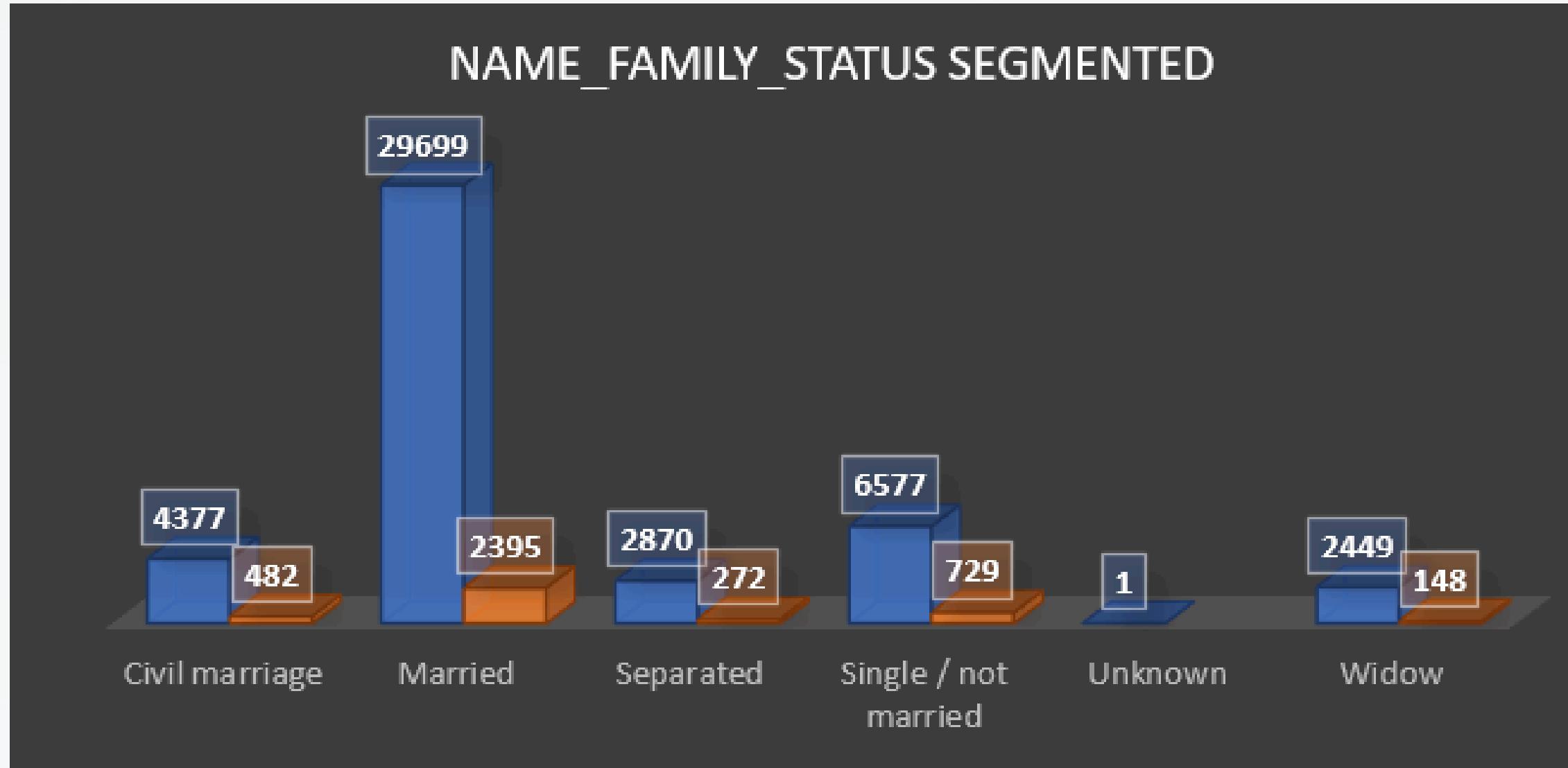
# D. UNIVARIATE ANALYSIS:

**RESULT:**



# D. UNIVARIATE ANALYSIS:

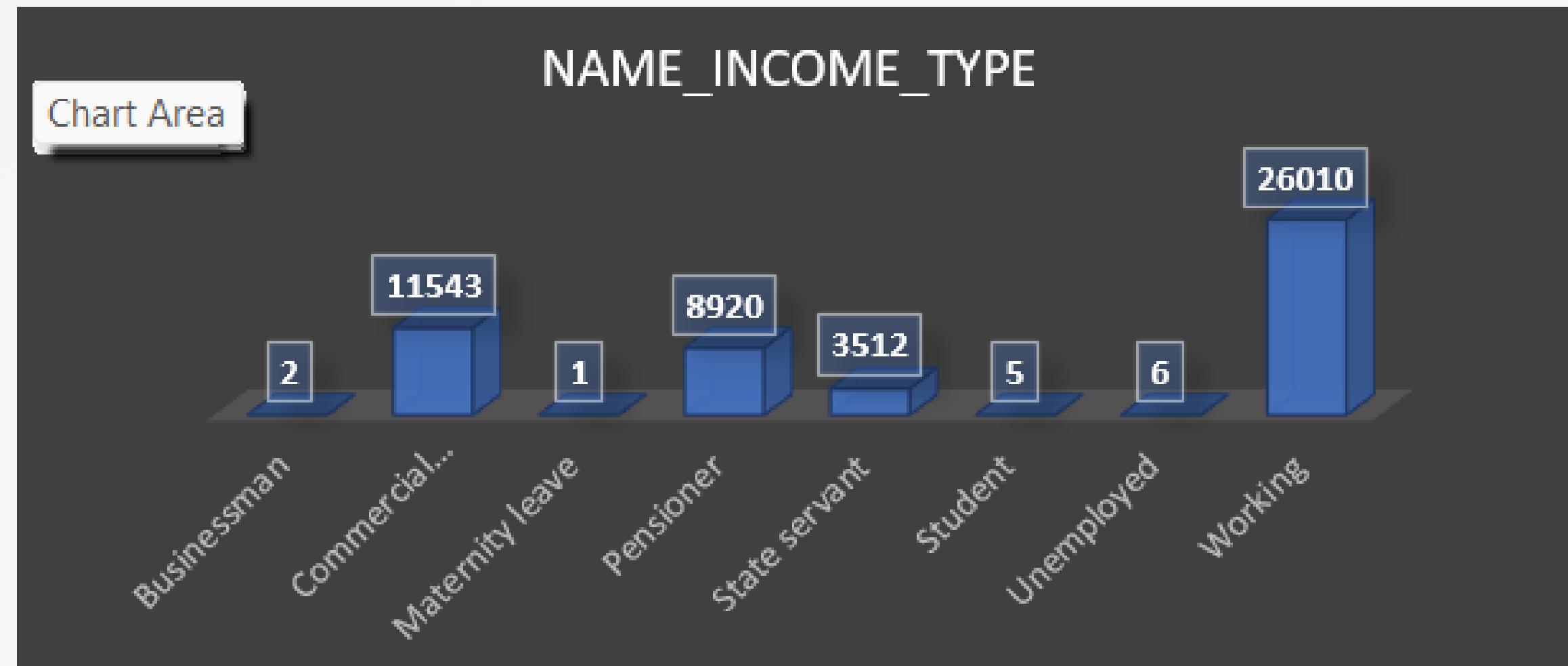
**RESULT:**



# D. UNIVARIATE ANALYSIS:

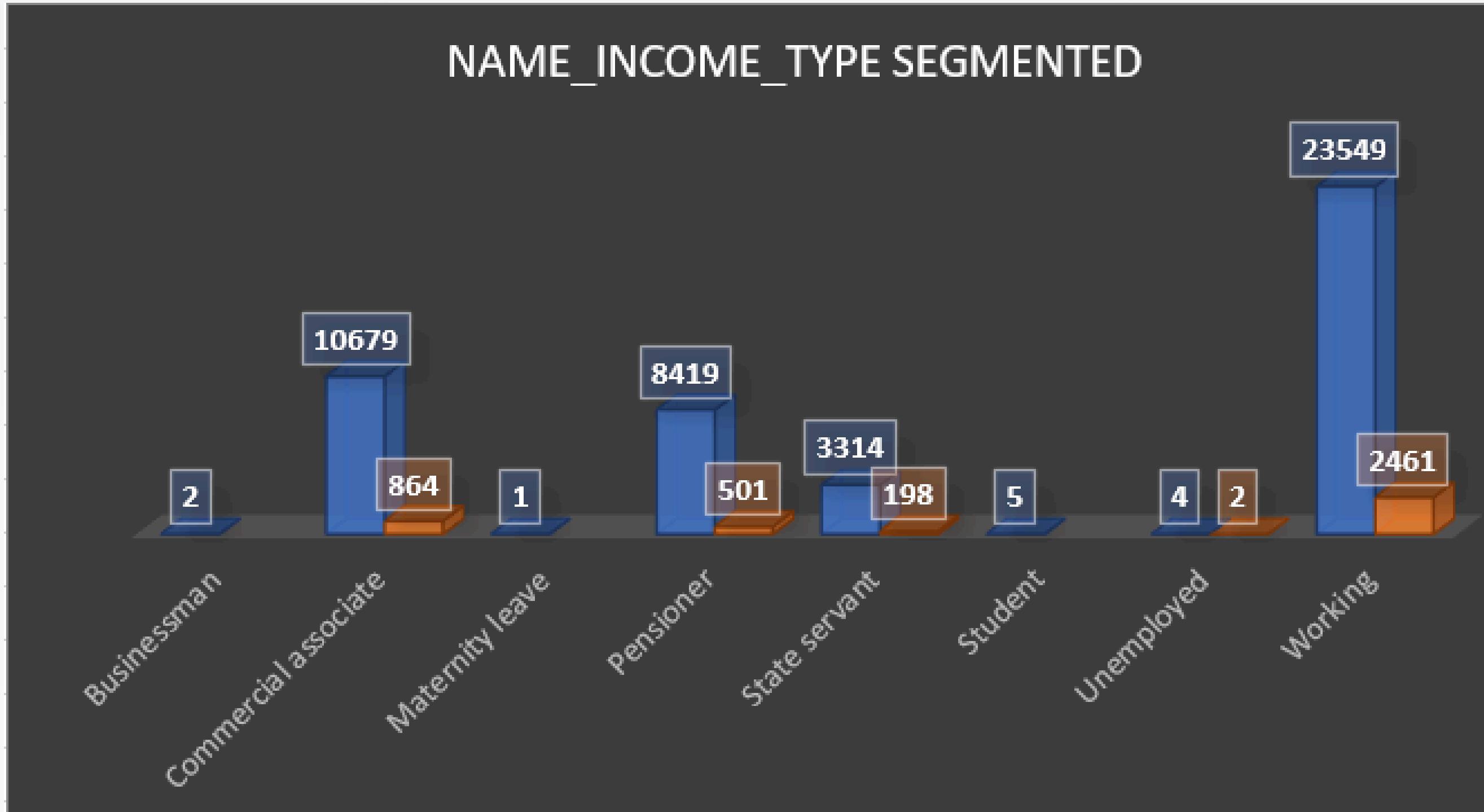
## RESULT:

| Row Labels           | Count of NAME_INCOME_TYPE |
|----------------------|---------------------------|
| Businessman          | 2                         |
| Commercial associate | 11543                     |
| Maternity leave      | 1                         |
| Pensioner            | 8920                      |
| State servant        | 3512                      |
| Student              | 5                         |
| Unemployed           | 6                         |
| Working              | 26010                     |
| <b>Grand Total</b>   | <b>49999</b>              |



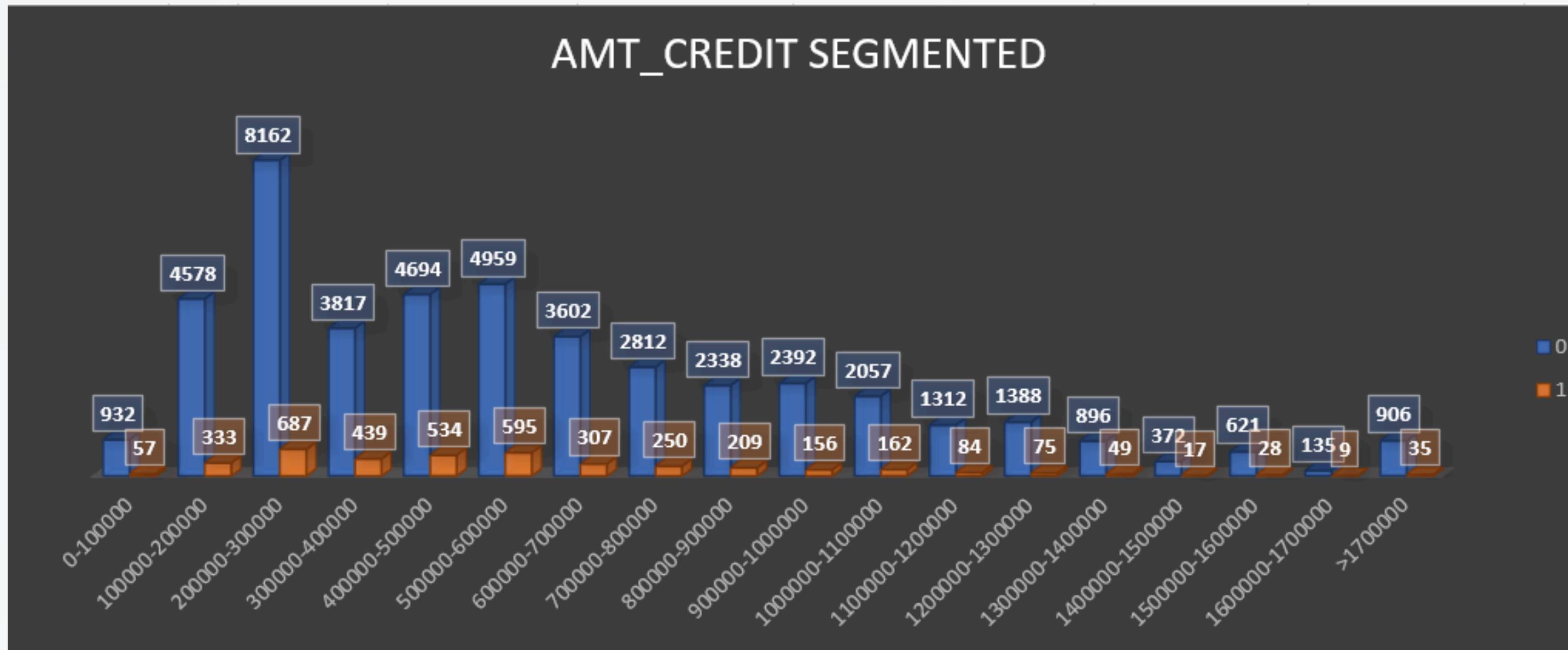
# D. UNIVARIATE ANALYSIS:

## RESULT:



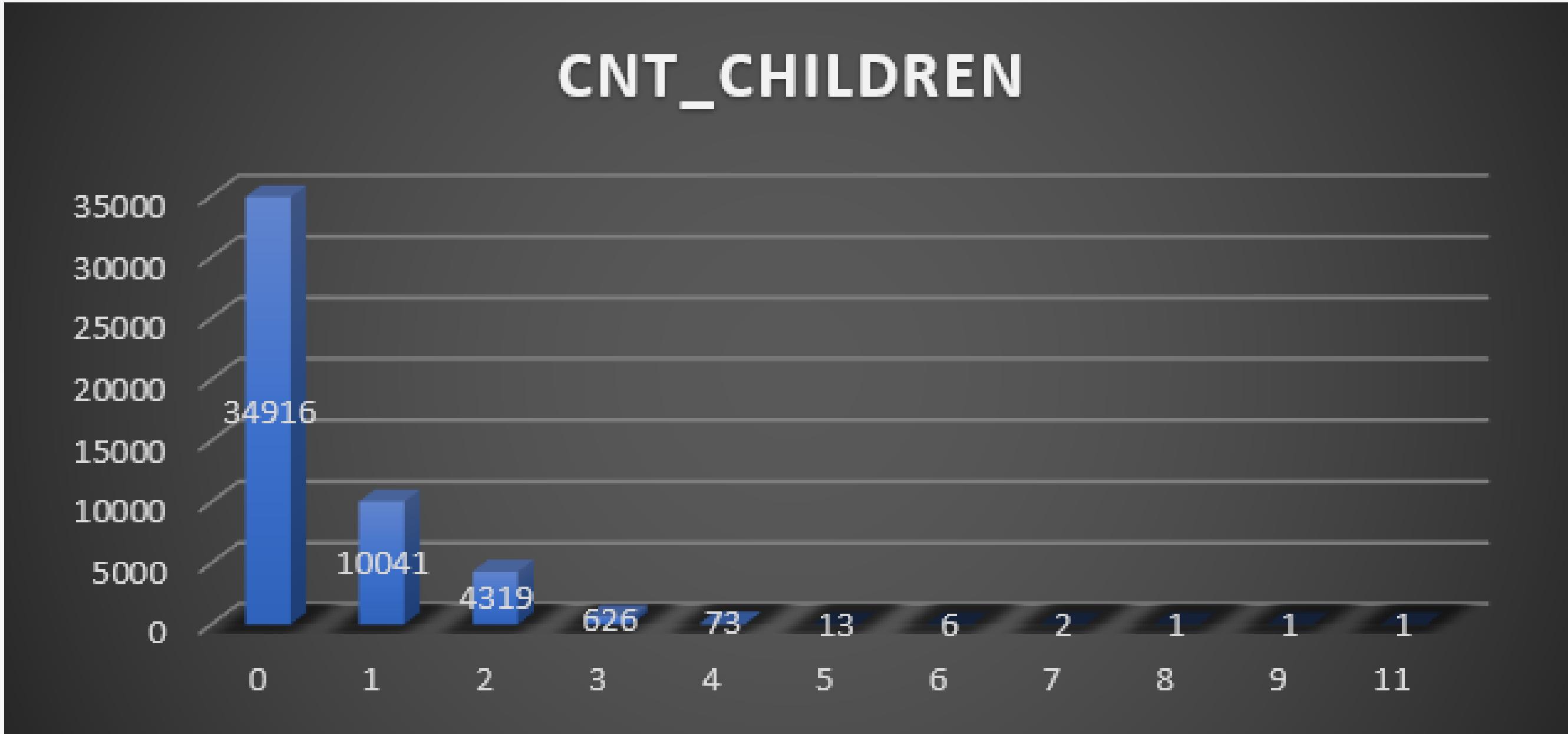
# D. UNIVARIATE ANALYSIS:

## RESULT:



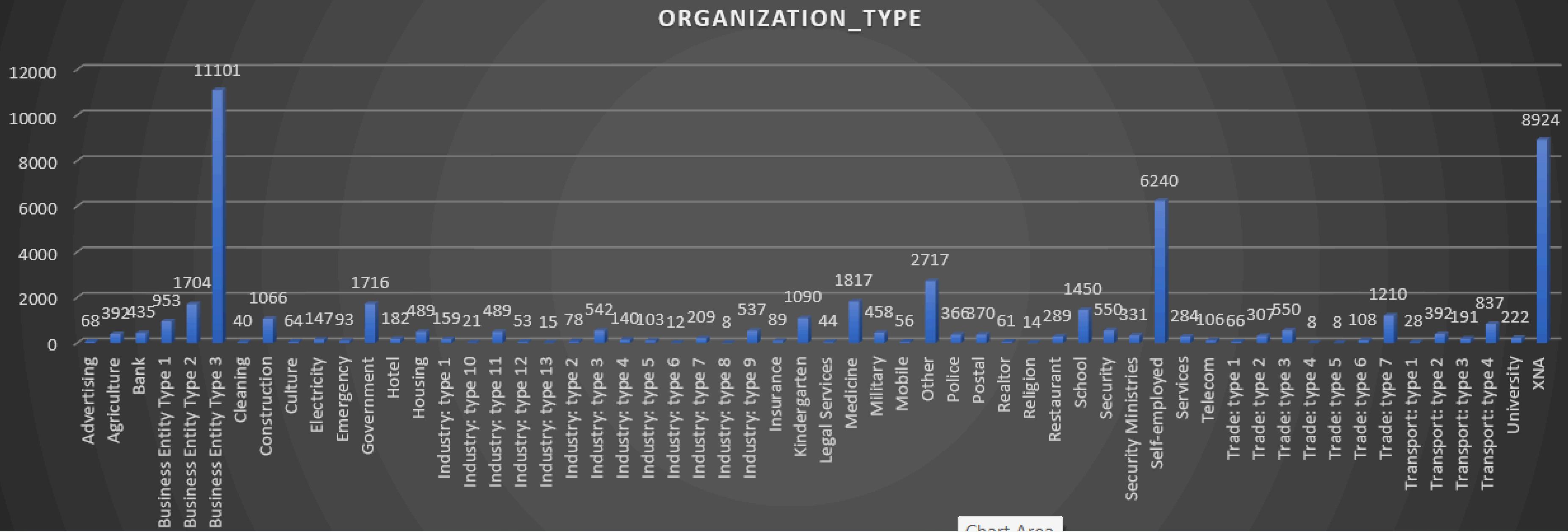
# D. UNIVARIATE ANALYSIS:

**RESULT:**



# D. UNIVARIATE ANALYSIS:

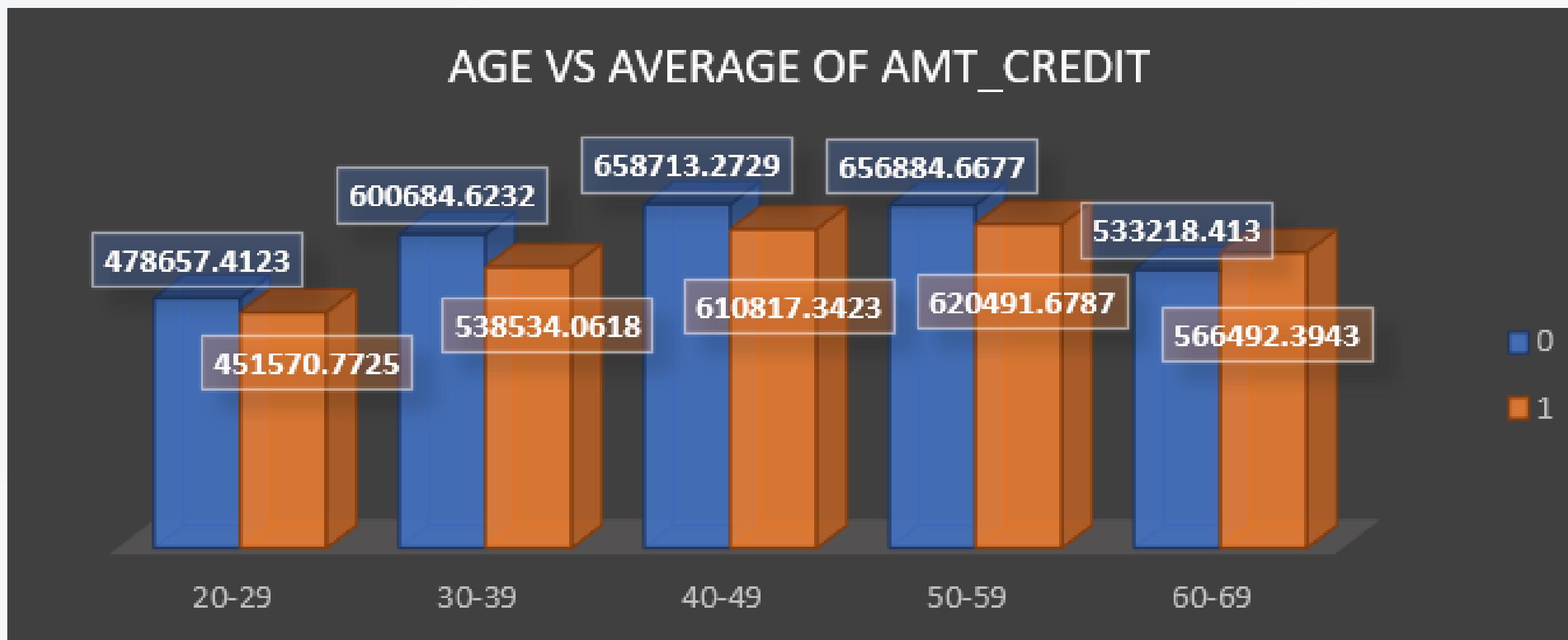
# RESULT:



# D. BIVARIATE ANALYSIS:

RESULT:

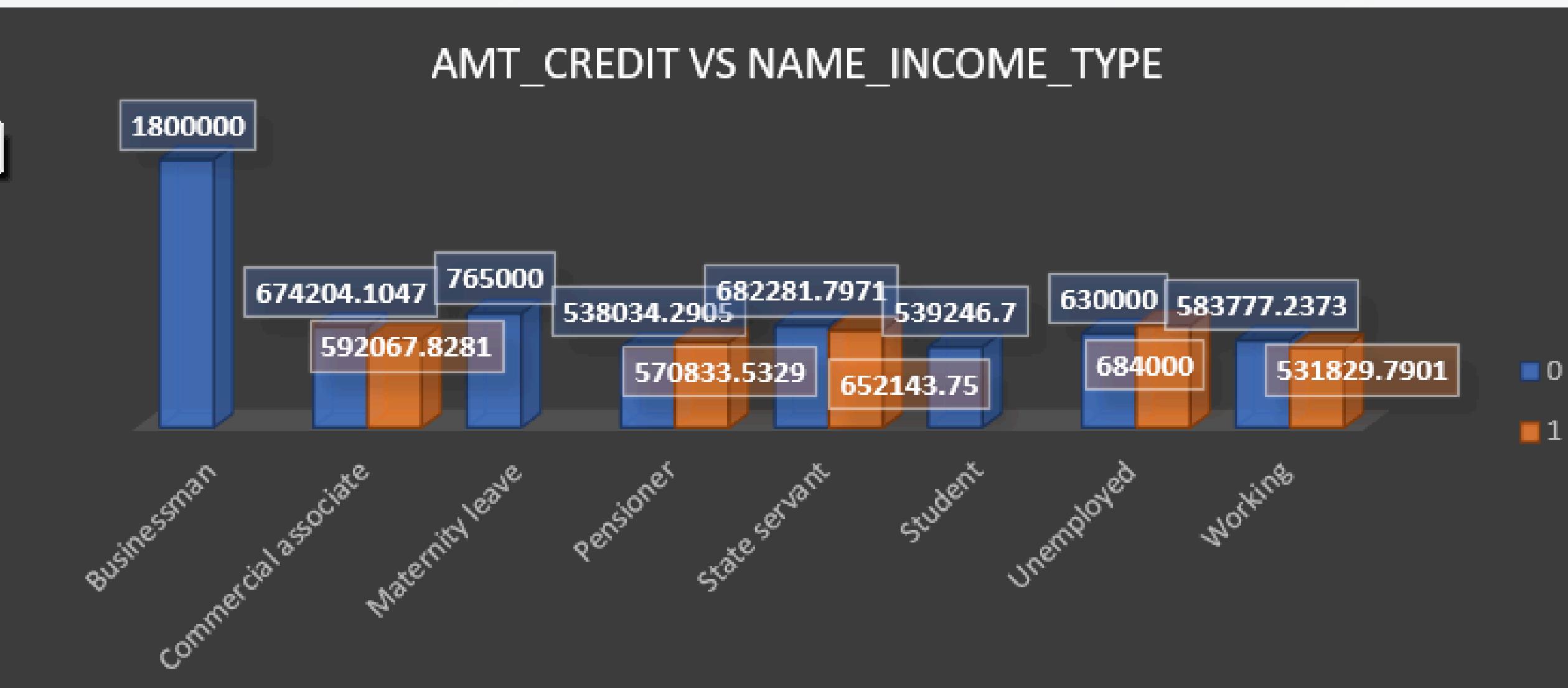
| Average of AMT_CREDIT | Column Labels |             |
|-----------------------|---------------|-------------|
| Row Labels            | 0             | 1           |
| 20-29                 | 478657.4123   | 451570.7725 |
| 30-39                 | 600684.6232   | 538534.0618 |
| 40-49                 | 658713.2729   | 610817.3423 |
| 50-59                 | 656884.6677   | 620491.6787 |
| 60-69                 | 533218.413    | 566492.3943 |
| Grand Total           | 603562.2995   | 555603.522  |



# D. BIVARIATE ANALYSIS:

## RESULT:

| NAME_INCOME_TYPE     | Average of AMT_CREDIT |                   |
|----------------------|-----------------------|-------------------|
|                      | 0                     | 1                 |
| Businessman          | 1800000               |                   |
| Commercial associate | 674204.1047           | 592067.8281       |
| Maternity leave      | 765000                |                   |
| Pensioner            | 538034.2905           | 570833.5329       |
| State servant        | 682281.7971           | 652143.75         |
| Student              | 539246.7              |                   |
| Unemployed           | 630000                | 684000            |
| Working              | 583777.2373           | 531829.7901       |
| <b>Grand Total</b>   | <b>603562.2995</b>    | <b>555603.522</b> |



## E. CORRELATIONS:

**Objective:** Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.

**Your Task:** Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.

**Result:**  
Correlation  
Coefficients for  
Payment  
difficulties are:-

| Correlation between Columns                              | Values      |
|--|-------------|
| OBS_60_CNT_SOCIAL_CIRCLE - OBS_30_CNT_SOCIAL_CIRCLE      | 0.998357563 |
| AMT_GOODS_PRICE - AMT_CREDIT                             | 0.986051701 |
| LIVE_REGION_NOT_WORK_REGION - REG_REGION_NOT_WORK_REGION | 0.861374946 |
| DEF_60_CNT_SOCIAL_CIRCLE - DEF_30_CNT_SOCIAL_CIRCLE      | 0.850995792 |
| REG_CITY_NOT_WORK_CITY - LIVE_CITY_NOT_WORK_CITY         | 0.825358079 |
| AMT_ANNUITY - AMT_GOODS_PRICE                            | 0.774006842 |
| AMT_ANNUITY - AMT_CREDIT                                 | 0.770772818 |

# E. CORRELATIONS:

**Result: Correlation Coefficients for Payment difficulties are:-**

| _INCOME_TO  | AMT_CREDIT   | AMT_ANNUITY  | AMT_GOODS_PRICE | REGION_POPULATION_RELATIV | DAYS_BIRTH   | DAYS_EMPLOYE | DAYS_REGISTRATION | DAYS_ID_PUBL | HOUR_AF      | REG_REG | REG_REG | LIVE_REG | REG_CITY | REG_CITY | LIVE_CITY | OBS_30  | DEF_30  | OBS_60  | DEF_60  | AMT_REC | AMT_REC | AMT_REC | AMT_REC | AMT_REQ_CREDI |         |          |
|-------------|--------------|--------------|-----------------|---------------------------|--------------|--------------|-------------------|--------------|--------------|---------|---------|----------|----------|----------|-----------|---------|---------|---------|---------|---------|---------|---------|---------|---------------|---------|----------|
| 0.036319722 | 0.005705458  | 0.02638396   | 0.001046405     |                           | -0.024912809 | 0.335876269  | -0.243591518      | 0.183072478  | -0.032537221 | -0.0053 | -0.0104 | 0.01379  | 0.02169  | 0.0201   | 0.07097   | 0.06788 | 0.01618 | -0.0028 | 0.01633 | -0.0033 | 0.00262 | 0.0012  | 0.00433 | -0.0116       | -0.0047 | -0.03573 |
| 1           | 0.377965752  | 0.451135167  | 0.383650216     |                           | 0.181941261  | 0.073769425  | -0.162702675      | 0.06893375   | 0.032286356  | 0.08543 | 0.07894 | 0.15705  | 0.14773  | 0.00993  | 0.01515   | 0.01966 | -0.033  | -0.032  | -0.033  | -0.0325 | 0.00812 | 0.00948 | 0.00949 | 0.07485       | 0.01578 | 0.03132  |
| 0.377965752 | 1            | 0.770772818  | 0.986051701     |                           | 0.095539444  | -0.051084182 | -0.077367219      | 0.008053758  | -0.008290189 | 0.05652 | 0.02781 | 0.0561   | 0.05443  | -0.0214  | -0.014    | 0.00398 | 0.00088 | -0.0135 | 0.00118 | -0.0186 | 3.7E-05 | 0.01349 | 0.00537 | 0.06398       | 0.02679 | -0.03156 |
| 0.451135167 | 0.770772818  | 1            | 0.774006842     |                           | 0.11727925   | 0.009911417  | -0.113005288      | 0.03460901   | 0.00942697   | 0.05356 | 0.04618 | 0.0825   | 0.07487  | -0.0053  | 0.00163   | 0.0112  | -0.01   | -0.0197 | -0.0097 | -0.023  | 0.01014 | 0.00916 | 0.01891 | 0.03799       | 0.01006 | -0.00417 |
| 0.383650216 | 0.986051701  | 0.774006842  | 1               |                           | 0.098899174  | -0.048664402 | -0.075069056      | 0.011016938  | -0.009441255 | 0.06513 | 0.03037 | 0.05755  | 0.05466  | -0.0204  | -0.0145   | 0.00286 | 0.00063 | -0.0152 | 0.00086 | -0.0197 | 0.00083 | 0.01367 | 0.00585 | 0.06582       | 0.02763 | -0.03435 |
| 0.181941261 | 0.095539444  | 0.11727925   | 0.098899174     |                           | 1            | -0.030435419 | -0.006610653      | -0.058501361 | -0.002236288 | 0.16761 | -0.0032 | 0.06315  | 0.08742  | -0.0461  | -0.0383   | -0.0113 | -0.0191 | 0.00891 | -0.018  | 0.00325 | -0.0031 | -0.0003 | 0.00264 | 0.07074       | -0.0097 | 0.00465  |
| 0.073769425 | -0.051084182 | 0.009911417  | -0.048664402    |                           | -0.030435419 | 1            | -0.615289978      | 0.335028046  | 0.270073313  | 0.09639 | 0.06043 | 0.09592  | 0.06989  | 0.1833   | 0.23613   | 0.14917 | 0.01229 | 0.00068 | 0.01229 | 0.00221 | 0.00149 | 0.00198 | -0.0024 | -0.0025       | -0.0215 | -0.07026 |
| 0.162702675 | -0.077367219 | -0.113005288 | -0.075069056    |                           | -0.006610653 | -0.615289978 | 1                 | -0.204370881 | -0.27222439  | -0.0924 | -0.0364 | -0.1073  | -0.0956  | -0.0926  | -0.2541   | -0.2177 | 0.00565 | 0.01703 | 0.00551 | 0.01652 | -0.0043 | 0.00162 | -0.0065 | -0.033        | 0.01458 | 0.04418  |
| 0.06893375  | 0.008053758  | 0.03460901   | 0.011016938     |                           | -0.058501361 | 0.335028046  | -0.204370881      | 1            | 0.103548902  | -0.0024 | 0.0279  | 0.03466  | 0.02328  | 0.06781  | 0.0916    | 0.06116 | 0.01098 | 0.00345 | 0.0113  | 0.00628 | -0.0037 | -0.0034 | 0.00066 | -0.0107       | 0.00313 | -0.0229  |
| 0.032286356 | -0.008290189 | 0.00942697   | -0.009441255    |                           | -0.002236288 | 0.270073313  | -0.27222439       | 0.103548902  | 1            | 0.03797 | 0.03323 | 0.04781  | 0.03375  | 0.07508  | 0.102     | 0.06332 | -0.0119 | 0.00231 | -0.0122 | 0.00264 | 0.00282 | 0.00351 | -0.0047 | -0.0132       | -0.0246 | -0.04469 |
| 0.08543156  | 0.056524809  | 0.053564989  | 0.065133303     |                           | 0.167612161  | 0.09638927   | -0.092357978      | -0.002396446 | 0.037971336  | 1       | 0.05119 | 0.07357  | 0.05975  | 0.01969  | 0.02693   | 0.01514 | -0.008  | -0.0024 | -0.008  | -0.0061 | -0.0074 | 0.01033 | -0.0067 | 0.02884       | -0.0005 | -0.025   |
| 0.078942904 | 0.027812773  | 0.046175655  | 0.030367622     |                           | -0.003185217 | 0.060427036  | -0.03641311       | 0.027899954  | 0.033228477  | 0.05119 | 1       | 0.44955  | 0.08045  | 0.33512  | 0.1426    | 0.00349 | -0.0151 | -0.0083 | -0.0151 | -0.0094 | -0.0025 | -0.0058 | -0.0018 | -0.0086       | -0.0003 | -0.01952 |
| 0.157051351 | 0.05609686   | 0.082502425  | 0.057545564     |                           | 0.063145413  | 0.095915233  | -0.107331487      | 0.034657988  | 0.047811506  | 0.07357 | 0.44955 | 1        | 0.86137  | 0.15193  | 0.23681   | 0.19223 | -0.0252 | -0.0089 | -0.0254 | -0.0137 | 4.9E-06 | 0.00075 | 0.00332 | 0.00424       | -0.0088 | -0.0275  |
| 0.147730123 | 0.05443061   | 0.074870093  | 0.054659311     |                           | 0.087419766  | 0.06988551   | -0.095573749      | 0.023280394  | 0.033751626  | 0.05975 | 0.08045 | 0.86137  | 1        | 0.02161  | 0.18389   | 0.23375 | -0.0202 | -0.0069 | -0.0204 | -0.012  | 0.00247 | 0.0029  | 0.00544 | 0.00993       | -0.0124 | -0.02249 |
| 0.009927686 | -0.021372433 | -0.005276721 | -0.020436382    |                           | -0.046089149 | 0.183304735  | -0.092557531      | 0.067811428  | 0.075080051  | 0.01969 | 0.33512 | 0.15193  | 0.02161  | 1        | 0.44144   | 0.02917 | -0.0053 | 0.00549 | -0.0055 | 0.00552 | 0.00048 | 8E-05   | -0.0011 | -0.0136       | -2E-05  | -0.00666 |
| 0.015150008 | -0.014007357 | 0.001628799  | -0.01449892     |                           | -0.038253612 | 0.236134428  | -0.254060105      | 0.091595217  | 0.102001817  | 0.02693 | 0.1426  | 0.23681  | 0.18389  | 0.44144  | 1         | 0.82536 | -0.006  | 0.001   | -0.006  | 0.00331 | 0.00428 | -0.0002 | 0.00218 | -0.0124       | -0.0039 | -0.01195 |
| 0.019663673 | 0.00397996   | 0.011203272  | 0.002861594     |                           | -0.011278612 | 0.149167938  | -0.217741277      | 0.061159259  | 0.063319024  | 0.01514 | 0.00349 | 0.19223  | 0.23375  | 0.02917  | 0.82536   | 1       | -0.0052 | -0.0022 | -0.0051 | -0.0002 | 0.00401 | -0.0012 | 0.00242 | -0.0046       | -0.0052 | -0.01294 |
| 0.033045993 | 0.000876364  | -0.009992103 | 0.000634386     |                           | -0.01906908  | 0.012287026  | 0.005650192       | 0.010977833  | -0.011854044 | -0.008  | -0.0151 | -0.0252  | -0.0202  | -0.0053  | -0.006    | -0.0052 | 1       | 0.30615 | 0.99836 | 0.22917 | 0.00236 | 0.00097 | -0.0043 | 0.00817       | 0.00885 | 0.03416  |
| 0.032012977 | -0.013509431 | -0.019746021 | -0.015155074    |                           | 0.008905591  | 0.000683769  | 0.017033326       | 0.003448989  | 0.002312725  | -0.0024 | -0.0083 | -0.0089  | -0.0069  | 0.00549  | 0.001     | -0.0022 | 0.30615 | 1       | 0.30856 | 0.851   | -0.0044 | 0.00369 | -0.005  | 0.00769       | 0.00535 | 0.01450  |
| 0.03301707  | 0.001184762  | -0.009675846 | 0.000856455     |                           | -0.018012695 | 0.01229458   | 0.005511276       | 0.011295659  | -0.01215588  | -0.008  | -0.0151 | -0.0254  | -0.0204  | -0.0055  | -0.006    | -0.0051 | 0.99836 | 0.30856 | 1       | 0.23128 | 0.00258 | 0.00087 | -0.0049 | 0.00813       | 0.00867 | 0.03457  |
| 0.032535174 | -0.018567338 | -0.023010616 | -0.019693991    |                           | 0.003253593  | 0.002207122  | 0.016516022       | 0.006282428  | 0.002642424  | -0.0061 | -0.0094 | -0.0137  | -0.012   | 0.00552  | 0.00331   | -0.0002 | 0.22917 |         |         |         |         |         |         |               |         |          |

# E. CORRELATIONS:

**Result: Correlation Coefficients for Re-payers are:-**

| Correlation between Columns                              | Values      |
|--|-------------|
| AMT_CREDIT - AMT_GOODS_PRICE                             | 0.982267963 |
| OBS_60_CNT_SOCIAL_CIRCLE - OBS_30_CNT_SOCIAL_CIRCLE      | 0.998065853 |
| DEF_60_CNT_SOCIAL_CIRCLE - DEF_30_CNT_SOCIAL_CIRCLE      | 0.89051161  |
| REG_REGION_NOT_WORK_REGION - LIVE_REGION_NOT_WORK_REGION | 0.806743886 |
| REG_CITY_NOT_WORK_CITY - LIVE_CITY_NOT_WORK_CITY         | 0.783754676 |
| AMT_CREDIT - AMT_ANNUITY                                 | 0.749665201 |
| AMT_GOODS_PRICE - AMT_ANNUITY                            | 0.74950403  |

| CNT_CHILDREN | AMT_INCOME_TO_AMT_CREDIT | AMT_CREDIT   | AMT_ANNUITY  | AMT_GOODS_PRI | REGION_POPULA | DAYS_BIRTH   | DAYS_EMPLOYED | DAYS_REGISTRAT | DAYS_ID_PUBLISH | HOUR_APPR_PRD | REG_REGION_NO | REG_REGION_NO | LIVE_REGION_NO | LIVE_REGION_NO | REG_CITY_NOT_L | REG_CITY_NOT_W | LIVE_CITY_NOT_W | OBS_30_CNT_SOC |
|--------------|--------------------------|--------------|--------------|---------------|---------------|--------------|---------------|----------------|-----------------|---------------|---------------|---------------|----------------|----------------|----------------|----------------|-----------------|----------------|
| 1            | 0.010110177              | 0.007601905  | 0.029172977  | -0.001079665  | -0.020359154  | 0.2496732    | -0.189324184  | 0.152113117    | -0.042360717    | -0.006884357  | -0.015713279  | -0.005665093  | -0.000389253   | 0.001745098    | 0.048916581    | 0.058183771    | 0.01793193      |                |
| 0.010110177  | 1                        | 0.015271444  | 0.018004594  | 0.013269502   | -0.006180303  | 0.009033662  | -0.01555963   | -0.009561152   | -0.009122006    | 0.014482013   | 0.000594885   | 0.001665752   | 0.002228043    | -0.005992314   | -0.010357192   | -0.008036091   | -0.011280916    |                |
| 0.007601905  | 0.015271444              | 1            | 0.749665201  | 0.982267963   | 0.067775624   | -0.142506035 | 0.016039571   | -0.042844404   | -0.043771901    | 0.045396384   | 0.006456715   | 0.023536318   | 0.034604167    | -0.052261708   | -0.039113138   | -0.006664341   | 0.033466173     |                |
| 0.029172977  | 0.018004594              | 0.749665201  | 1            | 0.74950403    | 0.073123998   | -0.008751713 | -0.079556008  | 0.021581654    | -0.02132109     | 0.044891881   | 0.031759358   | 0.065686571   | 0.074238732    | -0.017702478   | 0.002176683    | 0.013562938    | 0.013819016     |                |
| -0.001079665 | 0.013269502              | 0.982267963  | 0.74950403   | 1             | 0.076635488   | -0.141005898 | 0.020235348   | -0.043320226   | -0.049723232    | 0.057462759   | 0.007079035   | 0.025016178   | 0.035424194    | -0.052733855   | -0.04398108    | -0.013057846   | 0.032723967     |                |
| -0.020359154 | -0.006180303             | 0.067775624  | 0.073123998  | 0.076635488   | 1             | -0.016468731 | 0.007742909   | -0.046130288   | -0.005118563    | 0.156049669   | -0.003105241  | 0.019170075   | 0.059536379    | -0.034931305   | -0.043285987   | -0.025223619   | -0.008875436    |                |
| 0.2496732    | 0.009033662              | -0.142506035 | -0.008751713 | -0.141005898  | -0.016468731  | 1            | -0.581479041  | 0.288437837    | 0.247896571     | 0.057891695   | 0.039614727   | 0.075512807   | 0.054493345    | 0.149110346    | 0.226350689    | 0.143399639    | -0.01150233     |                |
| -0.189324184 | -0.011555963             | 0.016039571  | -0.079556008 | 0.020235348   | 0.007742909   | -0.581479041 | 1             | -0.188718437   | -0.230063668    | -0.052068931  | -0.035302931  | -0.084936755  | -0.0723322     | -0.088207867   | -0.24632369    | -0.200598048   | 0.003521851     |                |
| 0.152113117  | -0.009561152             | -0.042844404 | 0.021581654  | -0.043320226  | -0.046130288  | 0.288437837  | -0.188718437  | 1              | 0.09029149      | -0.057808905  | 0.015849157   | 0.016394347   | 0.013576904    | 0.05557252     | 0.100761679    | 0.069818516    | -0.005793296    |                |
| -0.042360717 | -0.009122006             | -0.043771901 | -0.02132109  | -0.049723232  | -0.005118563  | 0.247896571  | -0.230063668  | 0.09029149     | 1               | 0.005517259   | 0.024146053   | 0.04112087    | 0.029567415    | 0.064103042    | 0.083011624    | 0.038437264    | -0.027313737    |                |
| -0.006884357 | 0.014482013              | 0.045396384  | 0.044891881  | 0.057462759   | 0.156049669   | 0.057891695  | -0.052068931  | -0.057808905   | 0.005517259     | 1             | 0.049424395   | 0.076148974   | 0.066057815    | 0.005521997    | 0.003196965    | -0.0118232     | -0.01968249     |                |
| -0.015713279 | 0.000594885              | 0.006456715  | 0.031759358  | 0.007079035   | -0.003105241  | 0.039614727  | -0.035302931  | 0.015849157    | 0.024146053     | 0.049424395   | 1             | 0.525502749   | 0.100525936    | 0.338173404    | 0.147587163    | -0.003685235   | -0.031976397    |                |
| -0.005665093 | 0.001665752              | 0.023536318  | 0.065686571  | 0.025016178   | 0.019170075   | 0.075512807  | -0.084936755  | 0.016394347    | 0.04112087      | 0.076148974   | 0.525502749   | 1             | 0.806743886    | 0.183750398    | 0.228676332    | 0.169077507    | -0.032110777    |                |
| -0.000389253 | 0.002228043              | 0.034604167  | 0.074238732  | 0.035424194   | 0.059536379   | 0.054493345  | -0.0723322    | 0.013576904    | 0.029567415     | 0.066057815   | 0.100525936   | 0.806743886   | 1              | 0.026077544    | 0.157798988    | 0.217865427    | -0.020807542    |                |
| 0.001745098  | -0.005992314             | -0.052261708 | -0.017702478 | -0.052733855  | -0.034931305  | 0.149110346  | -0.088207867  | 0.05557252     | 0.064103042     | 0.005521997   | 0.338173404   | 0.183750398   | 0.026077544    | 1              | 0.4673011      | -0.015017325   | -0.049894148    |                |
| 0.048916581  | -0.010357192             | -0.039113138 | 0.002176683  | -0.04398108   | -0.043285987  | 0.226350689  | -0.24632369   | 0.100761679    | 0.083011624     | 0.003196965   | 0.147587163   | 0.228676332   | 0.157798988    | 0.4673011      | 1              | 0.783754676    | -0.042062137    |                |
| 0.058183771  | -0.008036091             | -0.006664341 | 0.013562938  | -0.013057846  | -0.025223619  | 0.143399639  | -0.200598048  | 0.069818516    | 0.038437264     | -0.0118232    | -0.003685235  | 0.169077507   | 0.217865427    | -0.015017325   | 0.783754676    | 1              | -0.024137029    |                |
| 0.01793193   | -0.011280916             | 0.033466173  | 0.013819016  | 0.032723967   | -0.008875436  | -0.011150233 | 0.003521851   | -0.005793296   | -0.027313737    | -0.01968249   | -0.031976397  | -0.032110777  | -0.020807542   | -0.049894148   | -0.042062137   | -0.024137029   | 1               |                |
| -0.01361871  | -0.007979437             | -0.024946679 | -0.034545374 | -0.019096612  | 0.027805916   | -0.020838794 | 0.029856345   | 0.000998183    | -0.028426519    | 0.017670352   | 0.008488226   | 0.001517328   | -0.006095943   | 0.003424351    | -0.015622704   | -0.027884283   | 0.365073855     |                |
| 0.01545875   | -0.011211173             | 0.034439308  | 0.014098626  | 0.033879184   | -0.007065002  | -0.01257029  | 0.004208684   | -0.005926611   | -0.026212482    | -0.019499892  | -0.031969665  | -0.031550153  | -0.019984596   | -0.050428718   | -0.041624386   | -0.022960813   | 0.998065853     |                |
| -0.018505702 | -0.006728958             | -0.029007236 | -0.040471029 | -0.020592919  | 0.027142318   | -0.025756651 | 0.023894099   | -0.006412628   | -0.027896348    | 0.01752421    | 0.005817862   | 0.004931716   | 8.67195E-05    | 0.002578282    | -0.013699572   | -0.024589896   | 0.29795102      |                |
| -0.000287596 | -0.001104179             | 0.017806362  | 0.037397493  | 0.01526195    | 0.009356216   | 0.024898705  | -0.003046463  | 0.006383728    | 0.014075827     | -0.033059219  | -0.010995025  | 0.022701249   | 0.031946125    | -0.001088254   | 0.018332705    | 0.014264686    | -0.014085065    |                |
| -0.030605254 | -0.00144685              | -0.008518401 | -0.018688343 | -0.006319208  | -0.003833539  | -0.02267042  | 0.043477616   | -0.001475079   | -0.00643299     | 0.001411229   | 0.00419991    | 0.01146392    | 0.007006175    | -0.019130488   | -0.005343937   | 0.000773297    | -0.017029194    |                |
| -0.030604048 | -0.002218606             | 0.000125371  | 0.03472145   | 0.000114      |               |              |               |                |                 |               |               |               |                |                |                |                |                 |                |

# **EXCEL SHEETS**

**The drive link for the excel sheets is:**

**[https://drive.google.com/drive/folders/1wYJKUJ7kxGirlgFVe5urVOM82pF-BGT?usp=drive\\_link](https://drive.google.com/drive/folders/1wYJKUJ7kxGirlgFVe5urVOM82pF-BGT?usp=drive_link)**

Working on this project helped me understand and use Microsoft Excel more better. It helped me gain experience on handling with graphs, charts, how to use statistics effectively and how Data Analytics is implemented using it in the real world such as bank procedures and obtain insights with the data provided as a Data Analyst. It also helped me gain experience in handling large volumes of data.

**THANK  
YOU.**

The background features a minimalist design with black wavy lines on a white surface. A large, bold, black sans-serif font displays the words "THANK" and "YOU." stacked vertically in the center. The left side of the image shows a bundle of wavy lines converging towards the center, while the right side shows a series of parallel, flowing wavy lines extending from the center towards the edge.