Welcome To Presentation

I'm Analyst, and I'll be sharing with you my point of view.

Bank Loan Case Study

ANKITA TANEJA

trainity

Table Of Contents

I
Introduction

Project Description

Strategic Perspectives

Description

As a data analyst at a finance company focused on urban loans, your task is to leverage Exploratory Data Analysis (EDA) to uncover patterns in customer and loan data. The goal is to identify high-risk applicants with insufficient credit history who may default while ensuring qualified applicants are approved, striking a balance between minimizing losses and maximizing business opportunities. The goal of this project is to use Exploratory Data Analysis (EDA) to investigate how customer and loan attributes impact the likelihood of loan default. This analysis aims to help the company minimize financial losses by identifying high-risk applicants while ensuring eligible applicants are not unfairly rejected.

Mission And Vision

OBJECTIVE

The objective of this project is to analyze and identify patterns that predict a customer's likelihood of defaulting on installments. This insight will help the company make informed decisions, such as declining loans, adjusting loan amounts, or increasing interest rates for high-risk applicants, while understanding the key factors influencing loan defaults to improve approval strategies.

Tech Stack

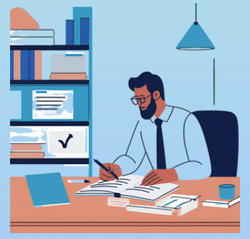
Microsoft Excel

Canva - Presentation

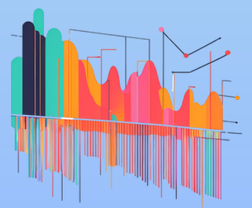
Loom - Video

Approach Used

Data Cleaning



Data Analyzing



Visualization

2
Tasks

Tasks Description

Tasks

One

Identify Missing Data and Deal with it Appropriately

Two

Identify Outliers in the Dataset

Three

Analyze Data Imbalance

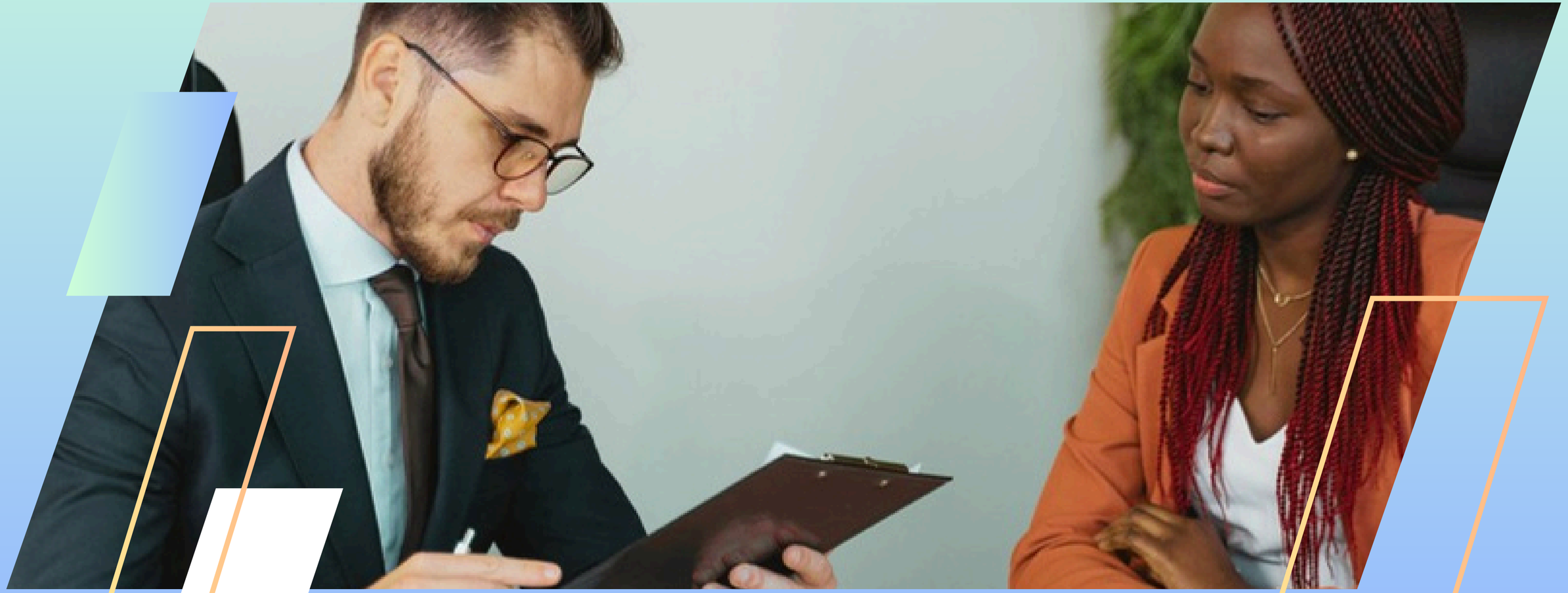
Four

Perform Univariate, Segmented Univariate, and Bivariate Analysis

Five

Identify Top Correlations for Different Scenarios

A picture is worth a thousand words



Tasks Description

Task One

Identify the missing data in the dataset & decide on an appropriate method to deal with it using Excel built-in functions & features.

Task Two

Detect & identify outliers in the dataset using Excel stats function & feature, focusing on numerical variables.

Task Three

Determine if there is data imbalance in the loan application dataset & calculate the ratio of data imbalance using Excel functions..

Task Four

To gain insights into the driving factors of loan default, it is imp. to conduct various analyses on consumer & loan attributes.

Task Five

Segment the dataset based on different scenarios & identify the top correlations for each segmented data using Excel functions.

3

Output

Outcome Results

Task 1

Identify Missing Data and Deal with it Appropriately

When handling missing data in a loan application dataset using Excel, the process can be summarized into the following steps:

1. Identify Missing Data

- Use the ISBLANK function to flag missing cells.
  - Formula: =ISBLANK(cell_reference)
  - This will return TRUE for missing cells and FALSE otherwise.
  - Formula: =BLANKCOUNT(range)
  - This will return count of blank cells.
- Use the COUNTIF or COUNT function to count missing values in each column.
  - Formula: =COUNTIF(range, "")
  - This counts all blank cells in the specified range.

2. Summarize Missing Data

- Create a summary table with column names and the count of missing values for each column.
  - Use the above COUNTIF formulas for each column.
- Calculate the proportion of missing data for each column:
  - Formula: =COUNTIF(range, "")/COUNTA(range)
  - This provides the percentage of missing values.

3. Handle Missing Data

Depending on the type of data and the proportion of missing values:

Imputation for numerical data:

Use the AVERAGE function to replace missing values with the column mean.

Formula: =IF(ISBLANK(cell_reference), AVERAGE(range), cell_reference)

Alternatively, use the MEDIAN function for skewed data.

Imputation for categorical data:

Use the MODE function to replace missing values with the most frequent category.

Formula: =IF(ISBLANK(cell_reference), MODE(range), cell_reference)

Deletion:

If a column or row has a high percentage of missing data, consider removing it, but only if it won't harm the analysis.

4. Visualize Missing Data

Create a summary chart to visualize the missing data distribution:

Highlight the summary table with column names and proportions.

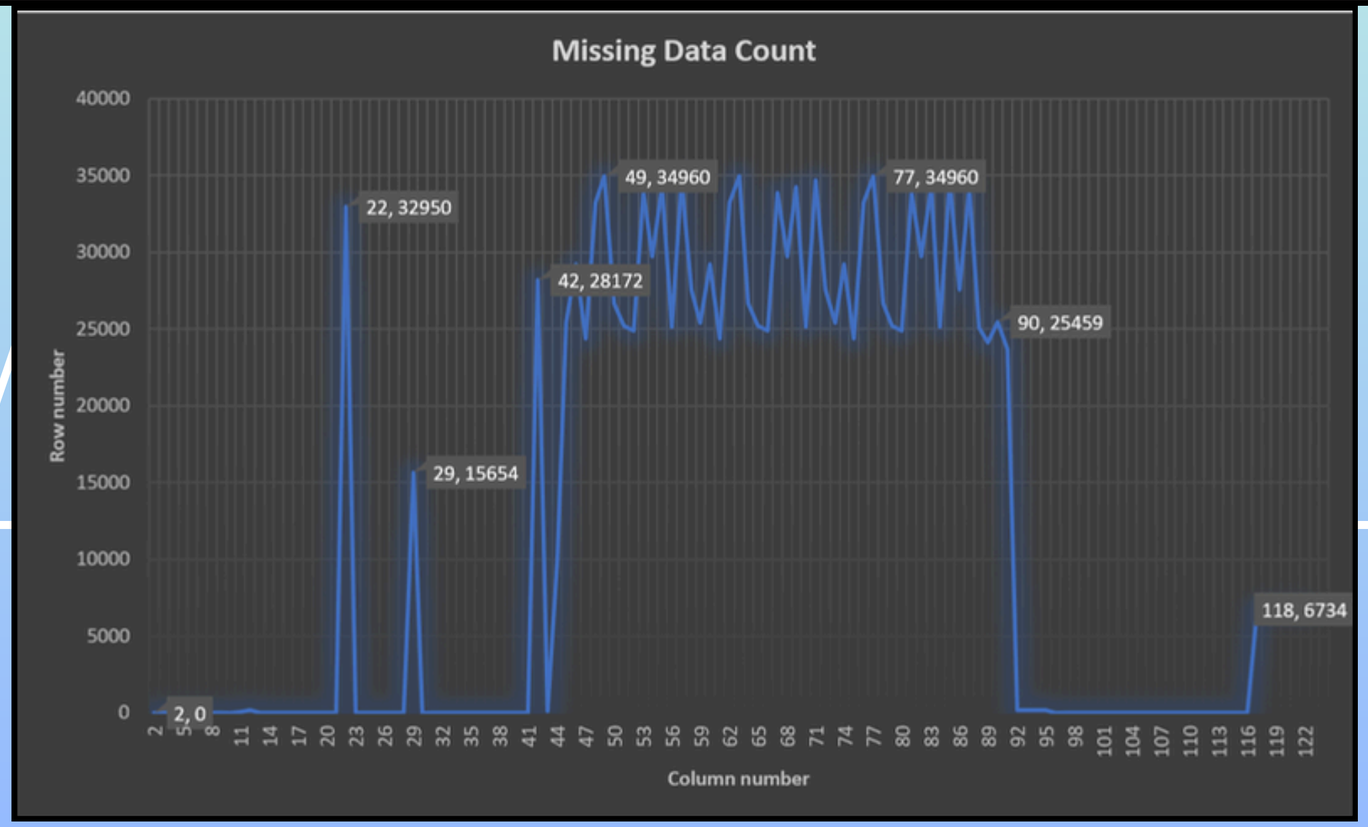
Insert a Bar Chart or Column Chart:

Go to the Insert tab, select Bar Chart or Column Chart.

Customize the chart to label axes and add a title, e.g., "Proportion of Missing Data by Variable".

By following these steps, you can identify, summarize, and handle missing data effectively in Excel.

PROPORTION
COUNTBLANK
Total Missing Values
1488212
Column with NO missing data
SK_ID_CURR, TARGET,NAME_CONTRACT_TYPE, CODE_GENDER, etc.
Column with missing data
AMT_ANNUITY, AMT_GOODS_PRICE, NAME_TYPE_SUITE etc.
Orange Color Column represents the least number of Missing Data Count. - Repaired
Red Color Column represents the number of Missing Data Count Greater 35% - Deleted



Insights

To calculate the total number of missing values for each column, use:

Excel

=COUNTBLANK(Column_Range)

Total missing values: 1488212

To calculate the Proportion:

=COUNTBLANK(Count_Range)/Rows Count

B1 =COUNTBLANK(B4:B50003) / (50000-1)

	A	B
1	PROPORTION	0.00%
2	COUNTBLANK	0

B2 =COUNTBLANK(B4:B50002)

	A	B
1	PROPORTION	0.00%
2	COUNTBLANK	0

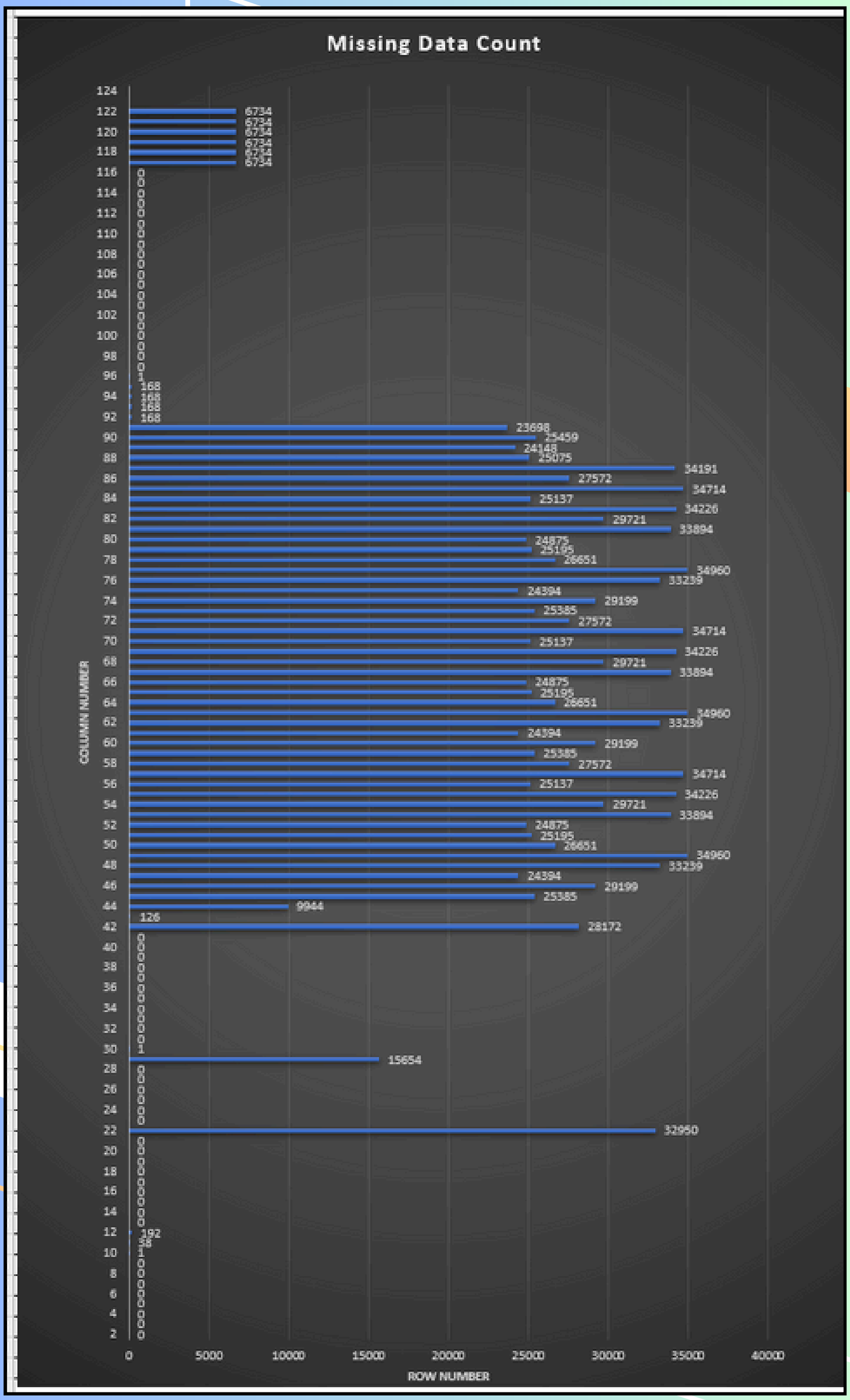
DAYS_BIRTH	DAYS BIRTH	DAYS_EMPLOYED	DAYS EMPLOYED	DAYS_REGISTRATION	DAYS REGISTERED	DAYS_ID_PU	Days ID
-9461	25.92054795	-637	1.745205479	-3648	9.994520548	-2120	5.808219178
-19046	52.18082192	-225	0.616438356	-4260	11.67123288	-2531	6.934246575
-16941	46.41369863	-1588	4.350684932	-4970	13.61643836	-477	1.306849315
-13778	37.74794521	-3130	8.575342466	-1213	3.323287671	-619	1.695890411
-18850	51.64383562	-449	1.230136986	-4597	12.59452055	-2379	6.517808219
-20099	55.06575342	365243	1000.665753	-7427	20.34794521	-3514	9.62739726
-10197	27.9369863	-679	1.860273973	-4427	12.12876712	-738	2.021917808
-20417	55.9369863	365243	1000.665753	-5246	14.37260274	-2512	6.882191781
-13439	36.81917808	-2717	7.443835616	-311	0.852054795	-3227	8.84109589
-14086	38.59178082	-3028	8.295890411	-643	1.761643836	-4911	13.45479452
-8728	23.91232877	-1157	3.169863014	-3494	9.57260274	-1368	3.747945205
-12931	35.42739726	-1317	3.608219178	-6392	17.51232877	-3866	10.59178082
-17718	48.54246575	-7804	21.38082192	-8751	23.97534247	-1259	3.449315068
-11348	31.09041096	-2038	5.583561644	-1021	2.797260274	-3964	10.86027397
-14815	40.5890411	-1652	4.526027397	-2299	6.298630137	-2299	6.298630137
-11146	30.5369863	-4306	11.79726027	-114	0.312328767	-2518	6.898630137
-24827	68.01917808	365243	1000.665753	-9012	24.69041096	-3684	10.09315068
-11286	30.92054795	-746	2.043835616	-108	0.295890411	-3729	10.21643836
-19334	52.96986301	-3494	9.57260274	-2419	6.62739726	-2893	7.926027397
-18724	51.29863014	-2628	7.2	-6573	18.00821918	-1827	5.005479452
-15948	43.69315068	-1234	3.380821918	-5782	15.84109589	-3153	8.638356164
-9994	27.38082192	-1796	4.920547945	-4668	12.7890411	-2661	7.290410959
-15280	41.8630137	-2668	7.309589041	-5266	14.42739726	-3787	10.37534247
-12974	35.54520548	-4404	12.06575342	-7123	19.51506849	-4464	12.23013699
-11694	32.03835616	-2060	5.643835616	-3557	9.745205479	-3557	9.745205479
-12158	33.30958904	-1275	3.493150685	-6265	17.16438356	-2009	5.504109589
-17199	47.12054795	-768	2.104109589	-63	0.17260274	-735	2.01369863
-21077	57.74520548	-1288	3.528767123	-5474	14.99726027	-4270	11.69863014

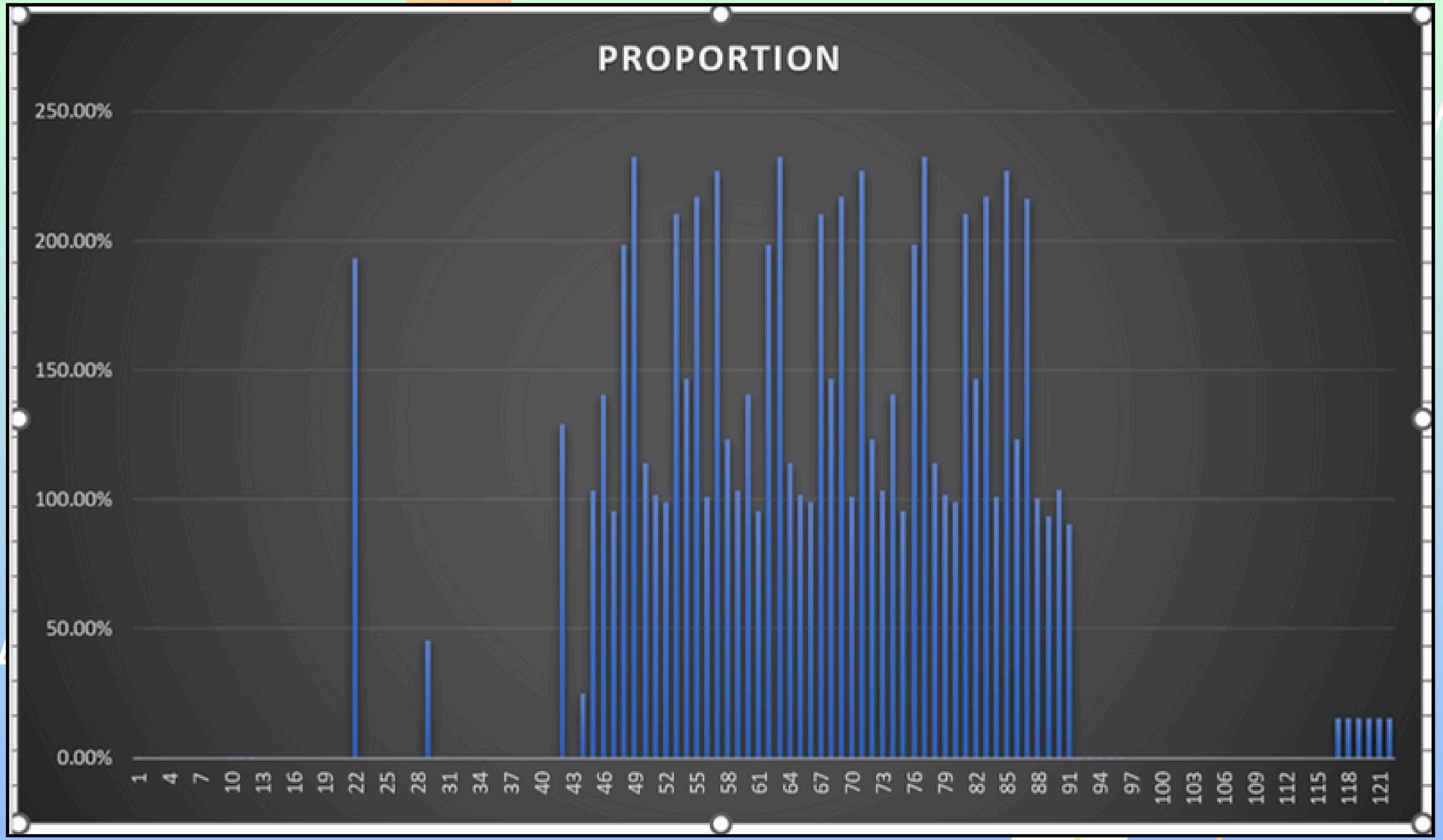
U2 =ABS(T2)/365

	S	T	U
1	DAYS BIRTH	DAYS_EMPLOYED	DAYS EMPLOYED
2	25.92054795	-637	1.745205479

COLUMN_MISSING_DATA	Average	MEDIAN
AMT_ANNUITY	27107.38	24939.00
AMT_GOODS_PRICE	539060.04	450000.00
NAME_TYPE_SUITE	unaccompanied	
CNT_FAM_MEMBERS	2.16	2.00
EXT_SOURCE_2	0.51	0.57
EXT_SOURCE_3	0.51	0.54
OBS_30_CNT_SOCIAL_CIRCLE	1.42	0.00
DEF_30_CNT_SOCIAL_CIRCLE	0.14	0.00
OBS_60_CNT_SOCIAL_CIRCLE	1.40	0.00
DEF_60_CNT_SOCIAL_CIRCLE	0.10	0.00
DAYS_LAST_PHONE_CHANGE	-964.30	-755.00
AMT_REQ_CREDIT_BUREAU_HOUR	0.01	0.00
AMT_REQ_CREDIT_BUREAU_DAY	0.01	0.00
AMT_REQ_CREDIT_BUREAU_WEEK	0.03	0.00
AMT_REQ_CREDIT_BUREAU_MON	0.27	0.00
AMT_REQ_CREDIT_BUREAU_QRT	0.26	0.00
AMT_REQ_CREDIT_BUREAU_YEAR	1.88	1.00

NAME_TYPE_SUITE	COUNT
Other_b	259
Children	542
GROUP OF PEOPLE	36
FAMILY	6549
SPOUSE, PARTNER	1849
UNACCOMPANIED	40435
Other_a	137


Missing Data Count
COLUMN NUMBER
ROW NUMBER
0
5000
10000
15000
20000
25000
30000
35000
40000
6734
168
23698
25459
24148
25075
34191
27572
34714
25137
34226
29721
33894
24875
25195
28651
34960
33239
24394
29199
25385
27572
34714
25137
34226
29721
33894
24875
25195
28651
34960
33239
24394
29199
25385
9944
126
28172
15654
32950
192
58


PROPORTION
250.00%
200.00%
150.00%
100.00%
50.00%
0.00%
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
94
97
100
103
106
109
112
115
118
121

Task 2

Identify Outliers in the Dataset

To identify and visualize outliers in a loan application dataset using Excel, follow these steps:

1. Identify Outliers

Step 1: Calculate the Interquartile Range (IQR)

- IQR Formula: IQR = Q3 - Q1
  - Use the QUARTILE function to find Q1 (25th percentile) and Q3 (75th percentile):
    - Formula for Q1: =QUARTILE(range, 1)
    - Formula for Q3: =QUARTILE(range, 3)
  - Calculate the IQR:
    - Formula: =Q3 - Q1

Step 2: Define Outlier Thresholds

- Lower Bound: Q1 - (1.5 × IQR)
- Upper Bound: Q3 + (1.5 × IQR)
  - Formula for Lower Bound: =Q1 - 1.5 * IQR
  - Formula for Upper Bound: =Q3 + 1.5 * IQR

Step 3: Identify Outliers

- Use an IF statement to flag outliers:
  - Formula: =IF(OR(cell_reference < Lower_Bound, cell_reference > Upper_Bound), "Outlier", "Normal")

2. Visualize Outliers

Box Plot

- Select the numerical data range.
- Go to the Insert tab → Insert Statistic Chart → Box and Whisker.
- Customize the box plot to display the distribution and highlight outliers.

Scatter Plot

- Use the numerical variable as the Y-axis and an identifier (e.g., row number) as the X-axis.
- Highlight outliers with a different color:
  - Add a new column indicating "Outlier" or "Normal".
  - Use Conditional Formatting to color cells or points in the scatter plot for outliers.

3. Investigate Outliers

- Thresholds: Evaluate if the outliers are valid based on business rules or domain knowledge.
  - For example, in a loan dataset, unusually high-income values could be valid for wealthy applicants.
- Treatment:
  - Retain valid data: If the outlier is a legitimate value, include it in the analysis.
  - Correct errors: Replace incorrect data with appropriate values (e.g., mean, median).
  - Exclude invalid data: Remove outliers only if they distort the analysis.

QUARTILE 1	0	0	112500	274779	16456.5
QUARTILE 3	0	1	202500	820638	34587
IRQ	0	1	90000	545859	18130.5
UL	0	2.5	337500	1639426.5	61782.75
LL	0	-1.5	-22500	-544009.5	-10739.25

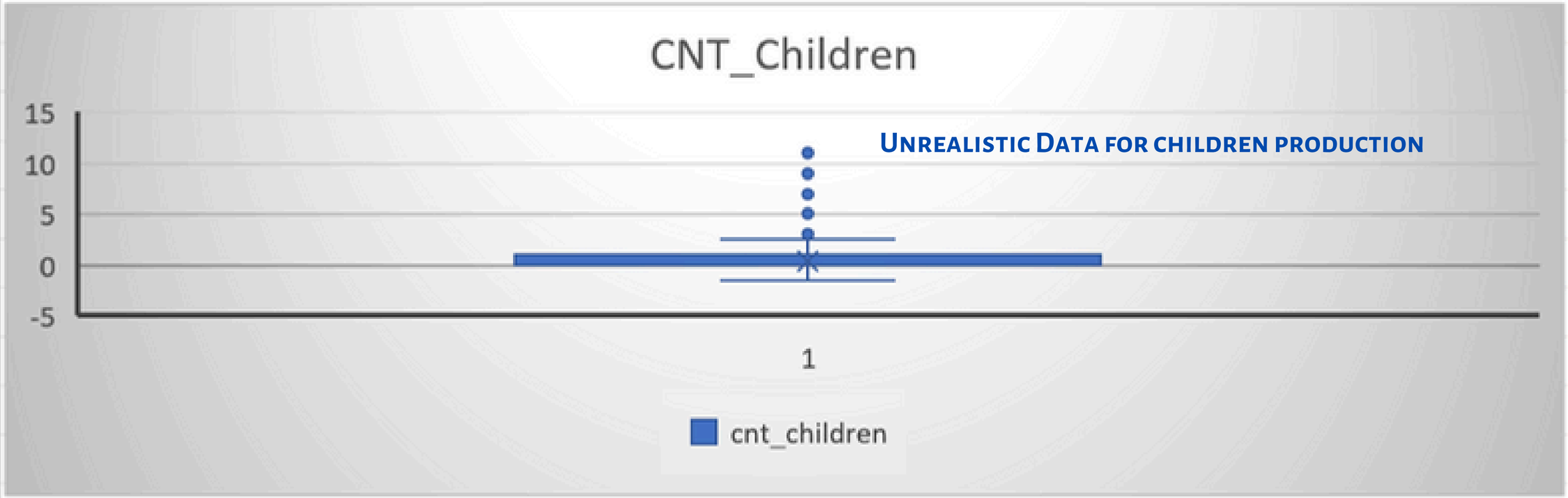
fx =QUARTILE.EXC(B7:B56341,1)

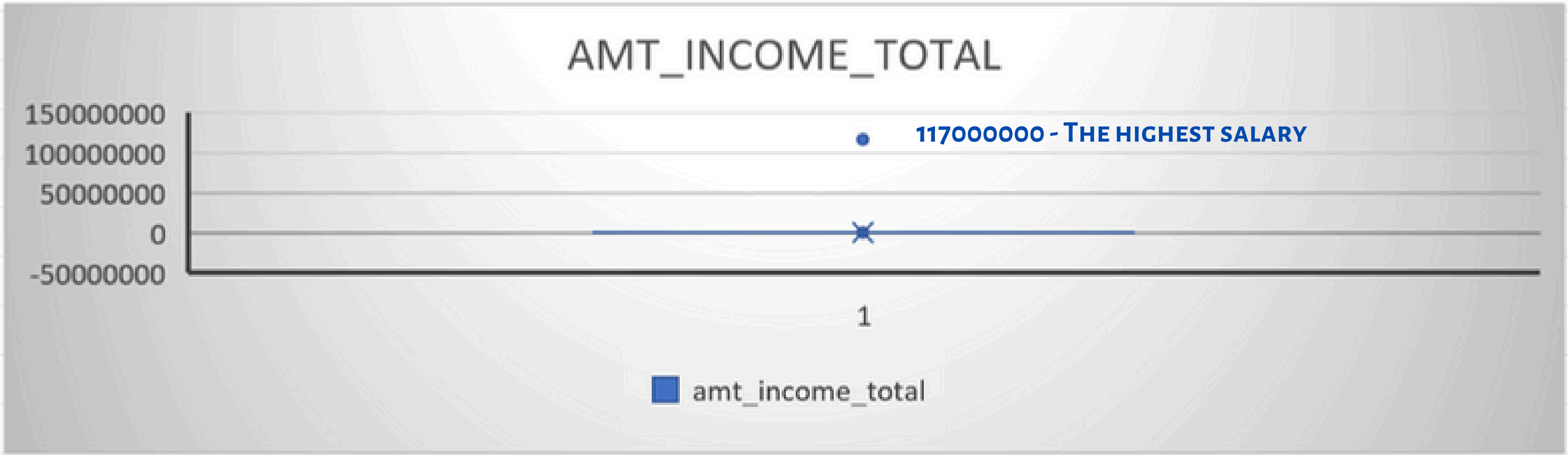
fx =QUARTILE.EXC(B7:B56341,3)

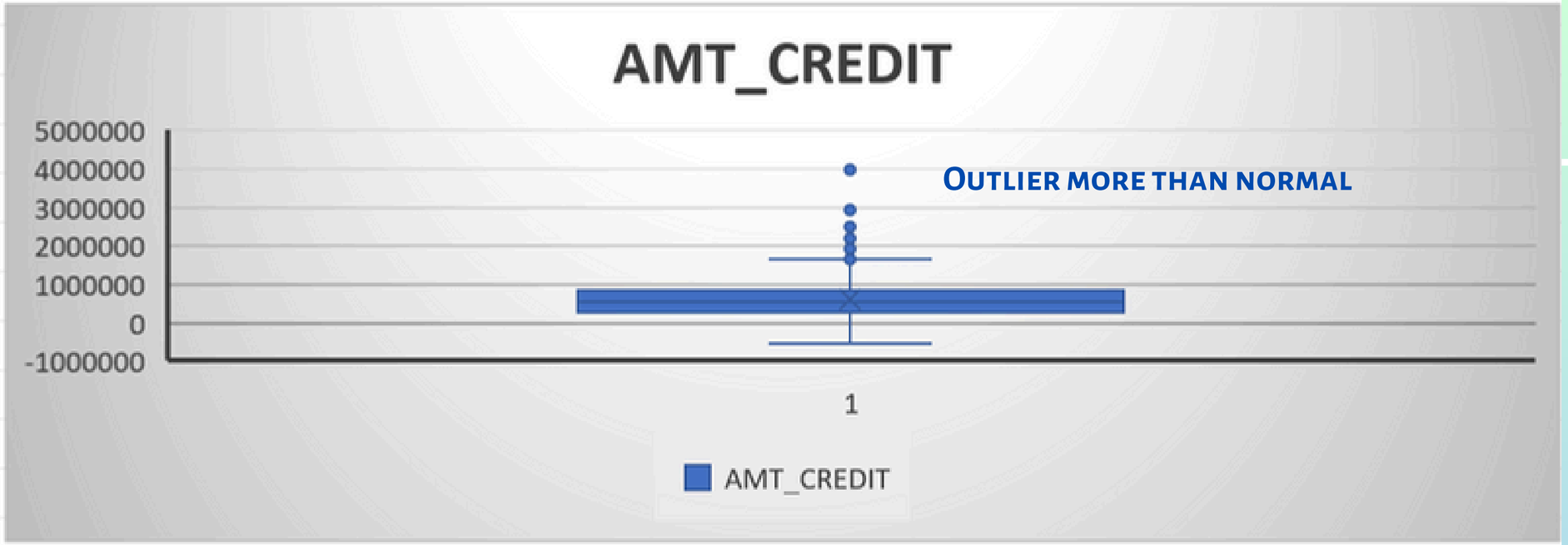
fx =B2-B1

fx =B2+(1.5*B3)

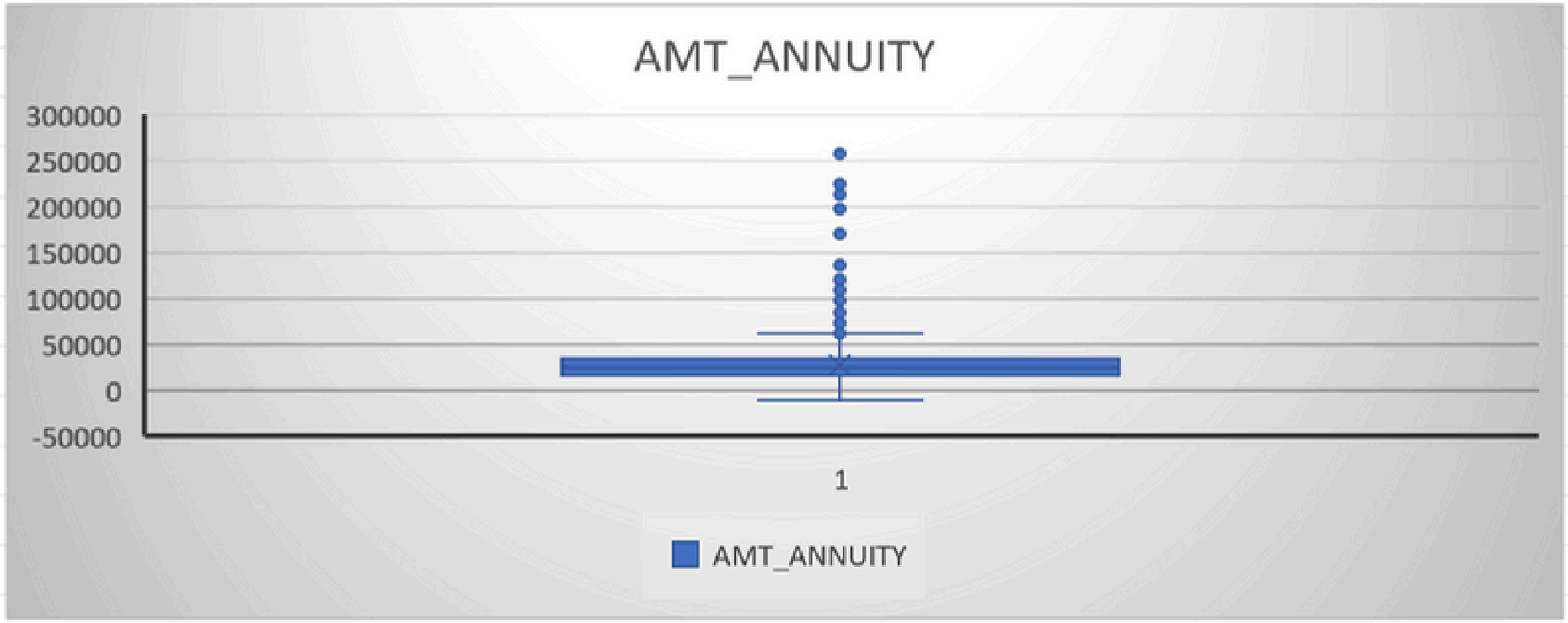
fx =B1-(1.5*B3)

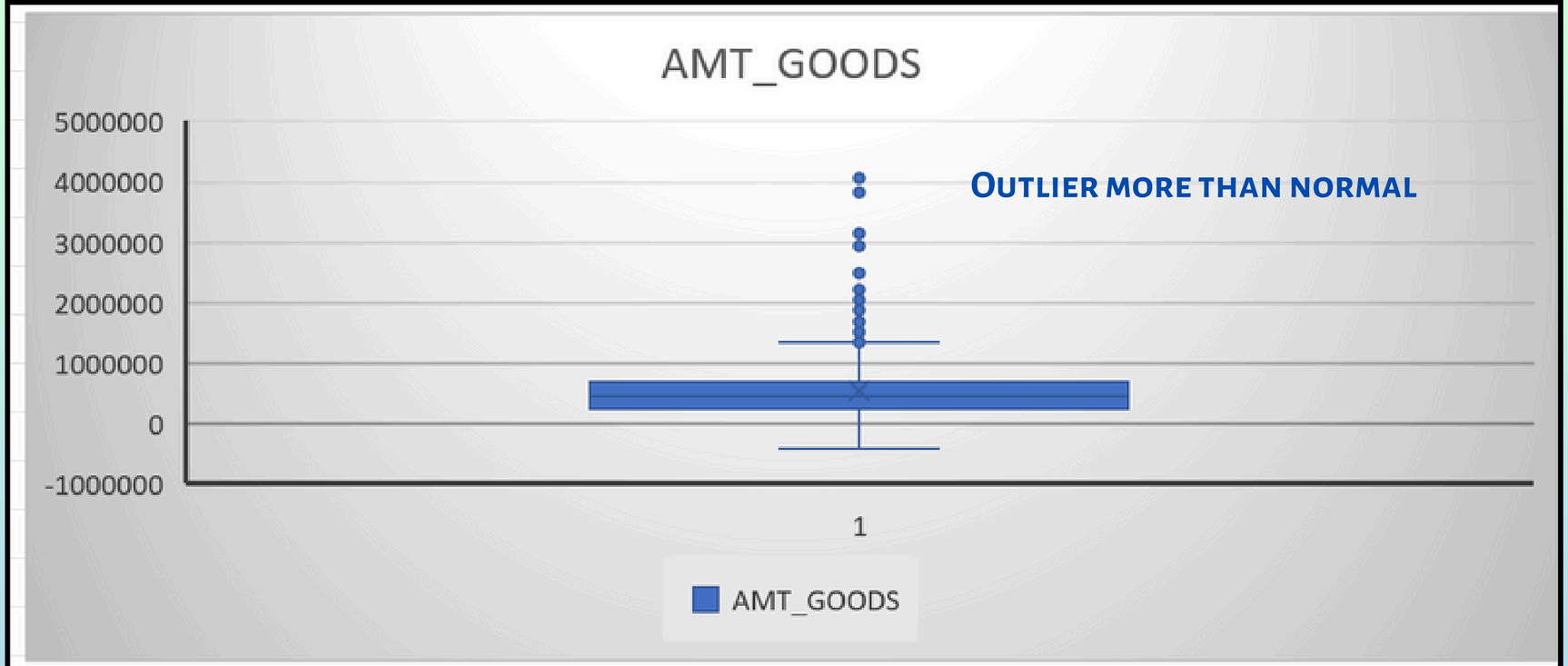

CNT_Children
15
10
5
0
-5
1
cnt_children
Unrealistic Data for children production


AMT_INCOME_TOTAL
150000000
100000000
50000000
0
-50000000
1
amt_income_total
117000000 - The highest salary

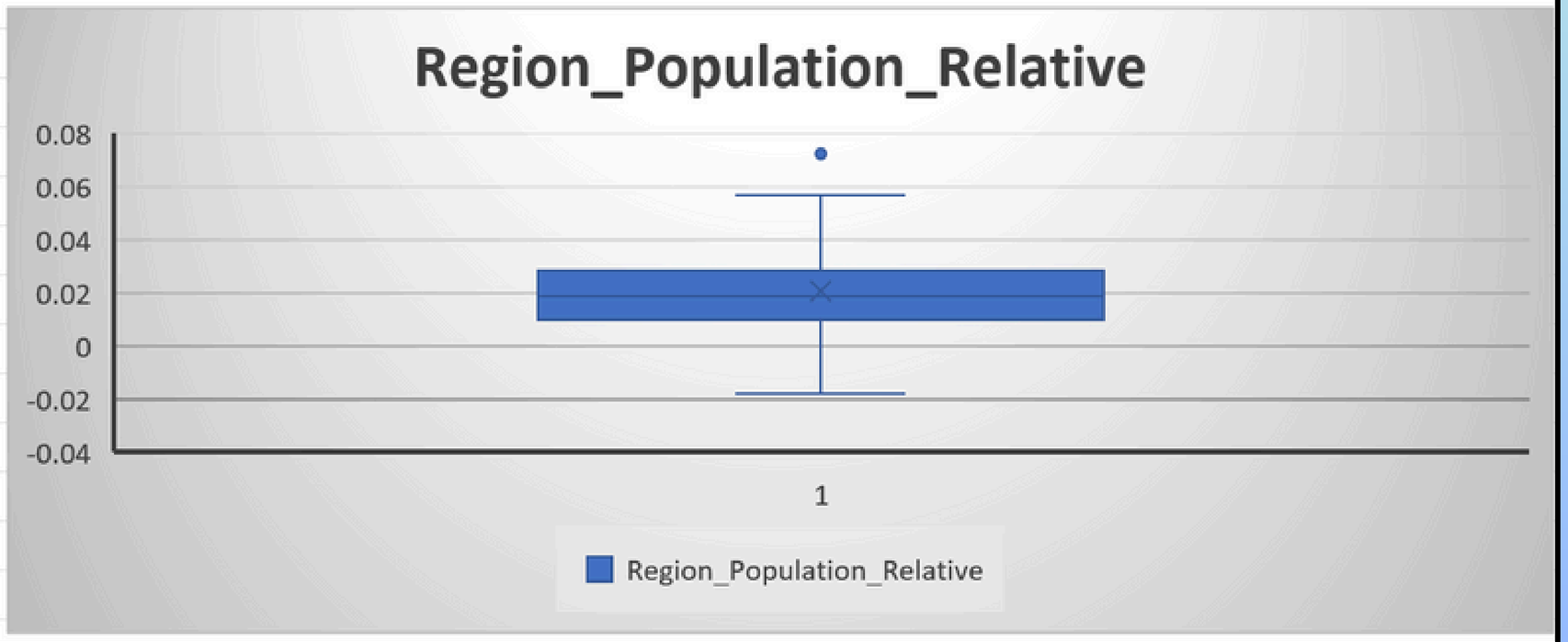


OUTLIER MORE THAN NORMAL





OUTLIER MORE THAN NORMAL



Dots in the box graph are the outliers like Years_employed have the outlier at 7100 which is not possible, amt_Goods have the abnormal outlier shown in the graph, CNT_Children can be 11 which is an outlier again, Total income cant be less than zero etc.

Task 3

Analyze Data Imbalance

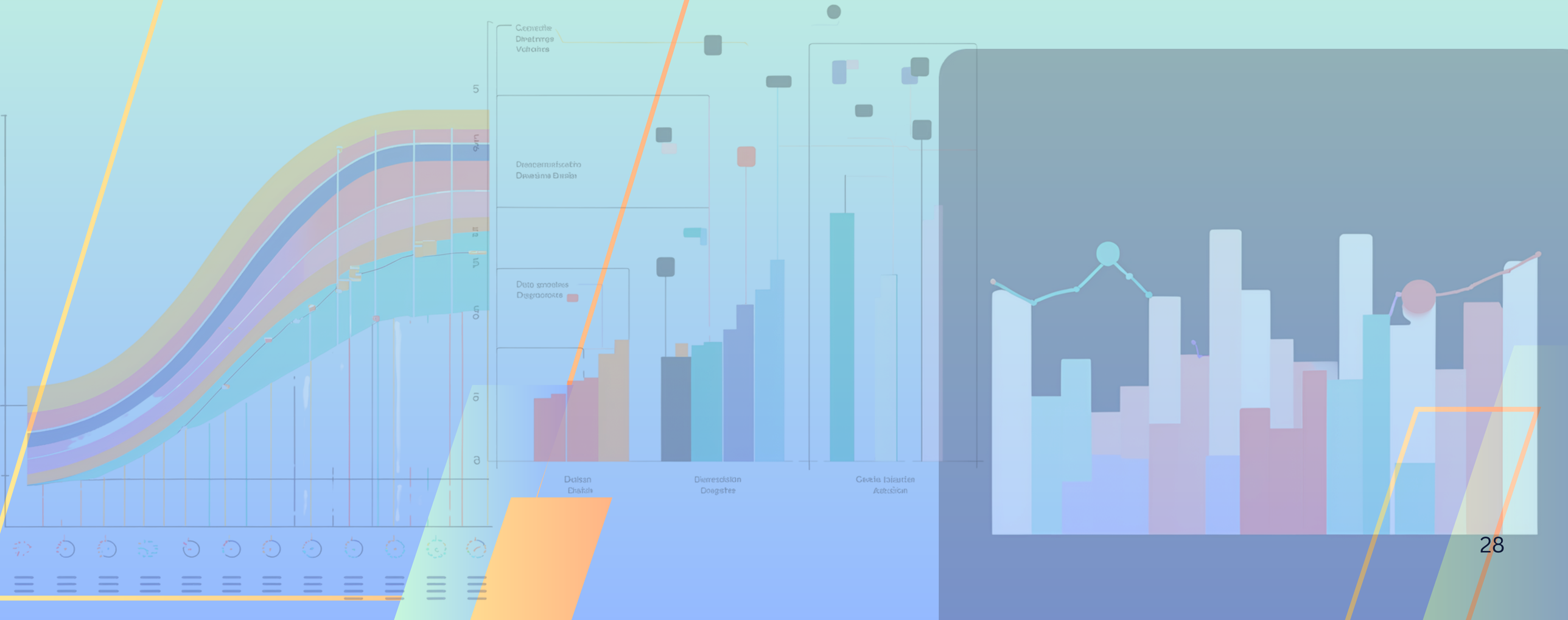
Data imbalance can significantly influence the accuracy of machine learning models, particularly in binary classification tasks. A key step in understanding this is analyzing the distribution of your target variable. This involves determining the frequency of each class and evaluating whether one class significantly outweighs the other.

Steps to Assess Data Imbalance in Excel

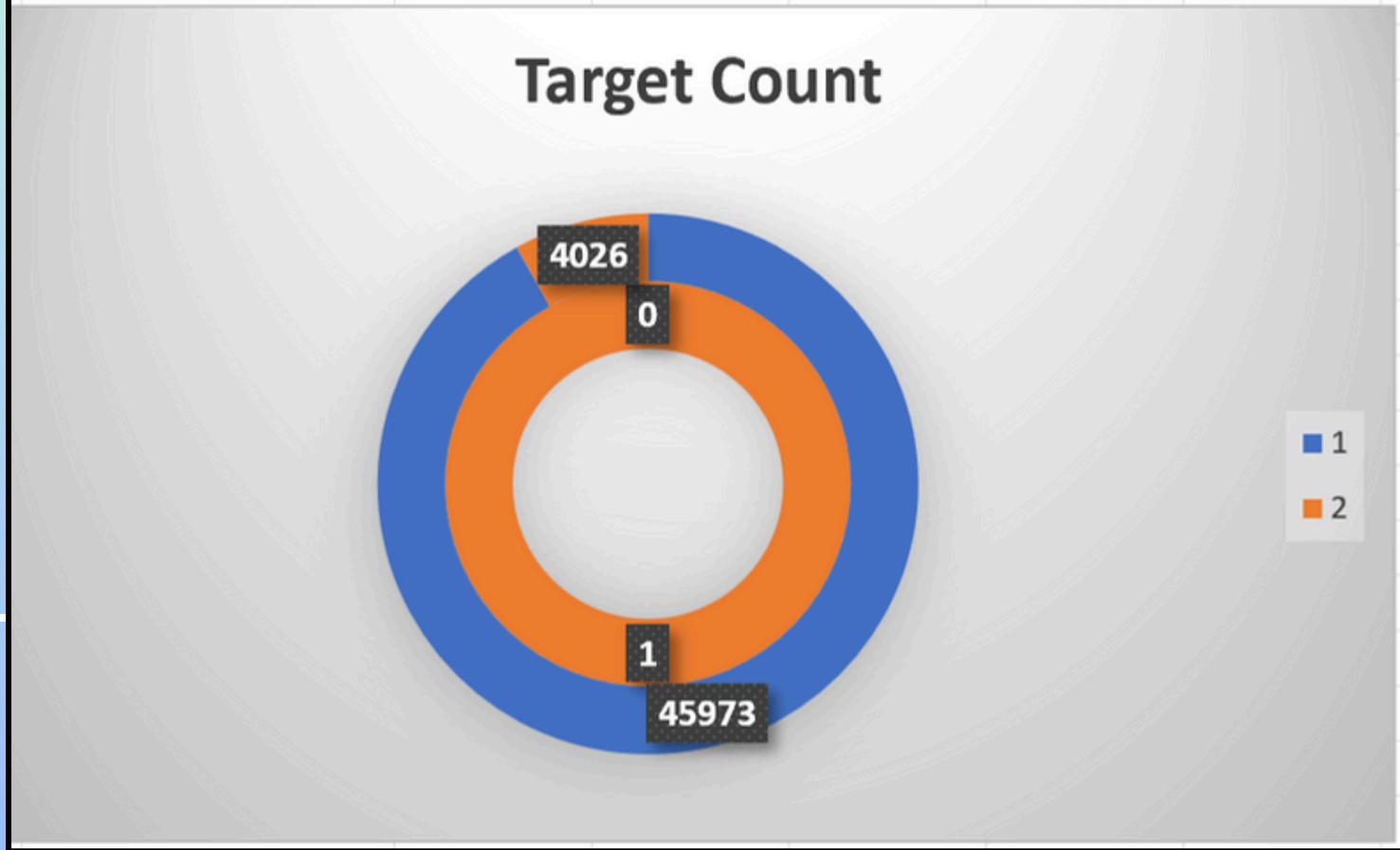
1. **Analyze Class Distribution:**
   - Use the COUNTIF function to calculate the number of instances for each class in the target variable.
     - Example: If the target variable is in column A, use:
       - =COUNTIF(A:A, "Class1") to count occurrences of Class 1.
       - =COUNTIF(A:A, "Class2") to count occurrences of Class 2.
   - Compute the total number of samples using the SUM function:
     - =SUM(COUNTIF(A:A, "Class1"), COUNTIF(A:A, "Class2")).
2. **Calculate the Imbalance Ratio:**
   - Compute the proportion of each class:
     - Proportion of Class 1: =COUNTIF(A:A, "Class1") / Total_Samples.
     - Proportion of Class 2: =COUNTIF(A:A, "Class2") / Total_Samples.
   - Determine the ratio:
     - Imbalance ratio = Proportion of Minority Class / Proportion of Majority Class.

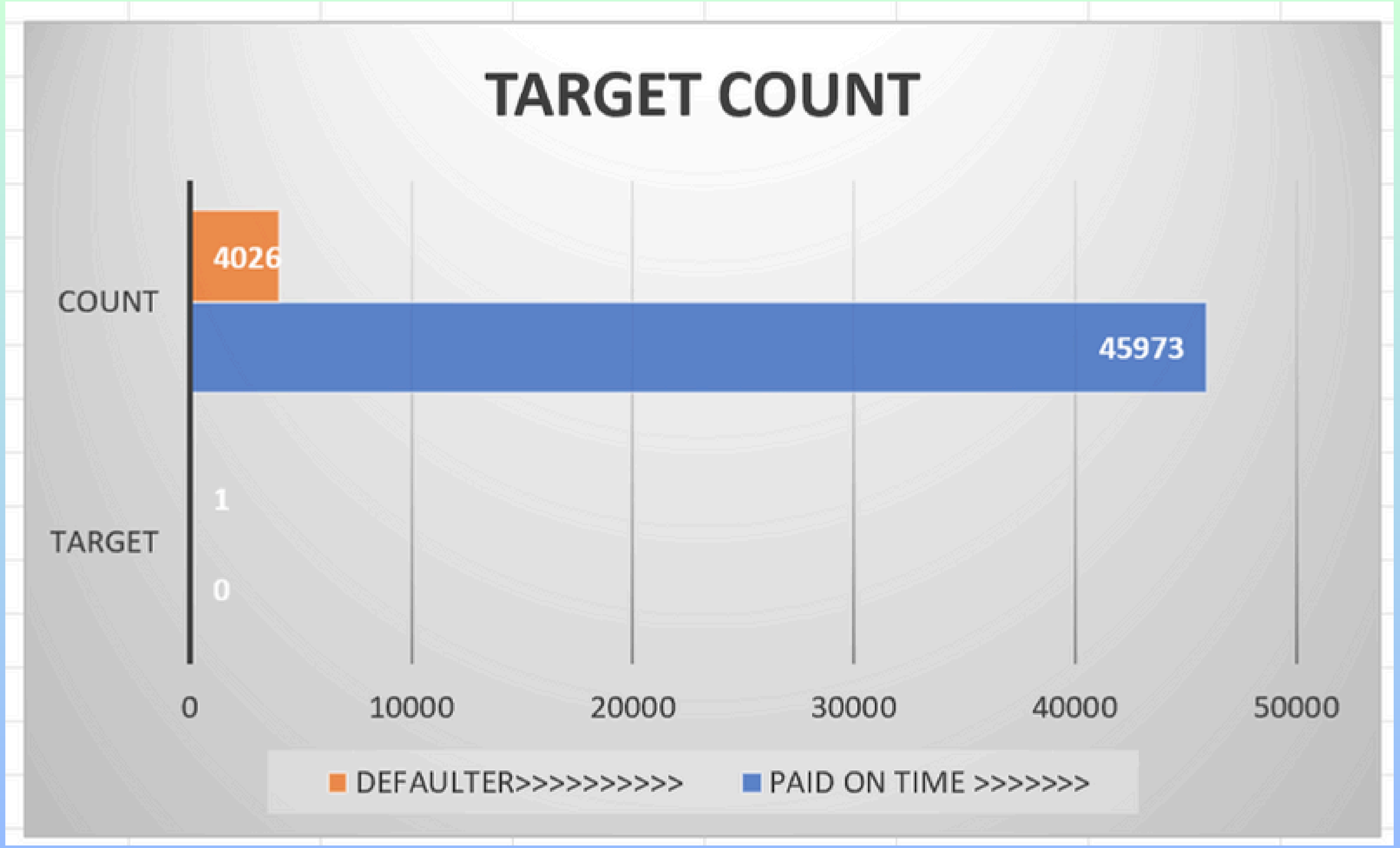
3. Visualize the Distribution:

- Create a Pie Chart or Bar Chart:
  - Select the class frequencies (e.g., count of Class 1 and Class 2).
  - Insert a pie chart or bar chart to visualize the distribution.
- Highlight the dominant class and the smaller class for better clarity.

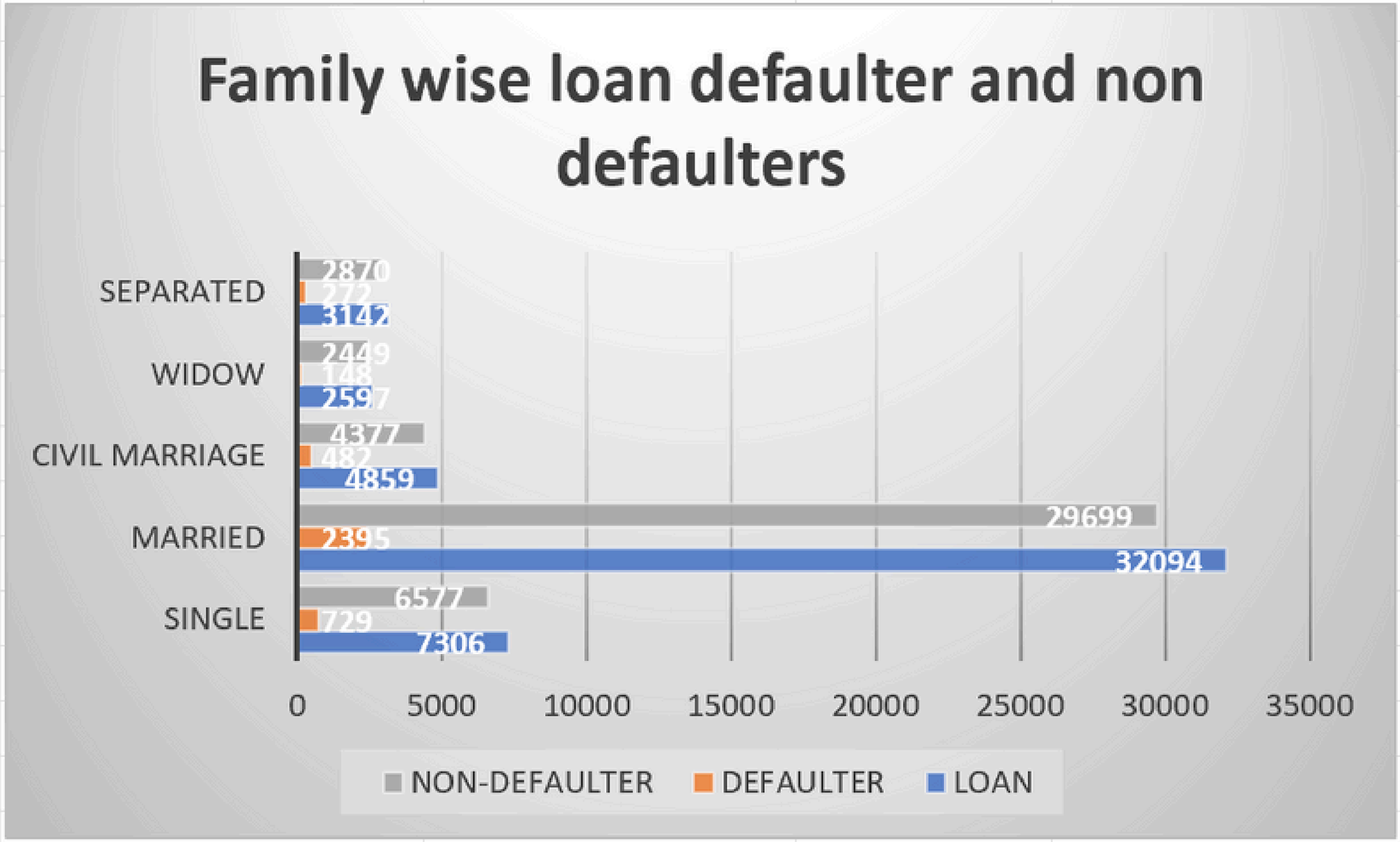


Target	count
0	45973
1	4026
total	49999
data imbalance ratio	11.41903




TARGET COUNT
COUNT
TARGET
4026
45973
1
0
0
10000
20000
30000
40000
50000
DEFAULTER>>>>>>>>>>
PAID ON TIME >>>>>>>

FAMILY STATUS	LOAN	DEFAULTER	NON-DEFAULTER
Single	7306	729	6577
Married	32094	2395	29699
Civil Marriage	4859	482	4377
Widow	2597	148	2449
Separated	3142	272	2870



Task 4

Perform Univariate, Segmented Univariate, and Bivariate Analysis

To better understand the factors contributing to loan default, it is essential to analyze consumer and loan attributes using various techniques. Here's how you can perform univariate, segmented univariate, and bivariate analyses in Excel:

1. Univariate Analysis:

Understand the distribution of individual variables.

- Steps:
  a. Use Excel statistical functions to calculate key metrics like:
    - Mean: =AVERAGE(range)
    - Median: =MEDIAN(range)
    - Standard Deviation: =STDEV.P(range)
    - Counts: =COUNT(range)
  b. Use histograms or bar charts to visualize the distribution:
    - Select the data and go to Insert > Chart > Histogram/Bar Chart.
- Purpose: Identify patterns such as central tendency, spread, and skewness in the data.

2. Segmented Univariate Analysis:

Compare variable distributions across different scenarios (e.g., defaulted vs. non-defaulted loans).

- Steps:
  a. Filter or Sort data by scenarios using Excel filters or conditions.

b. Use Pivot Tables for grouped statistics:

- Select data and insert a pivot table.
- Place the variable of interest in Rows and the segmentation (e.g., Defaulted/Non-defaulted) in Columns.
- Add statistics like counts or averages to the Values field.

- Create grouped or stacked bar charts:
  - Select Pivot Table results and use Insert > Chart > Stacked Bar/Column Chart.

- Purpose: Understand how distributions differ between groups, highlighting key contrasts.

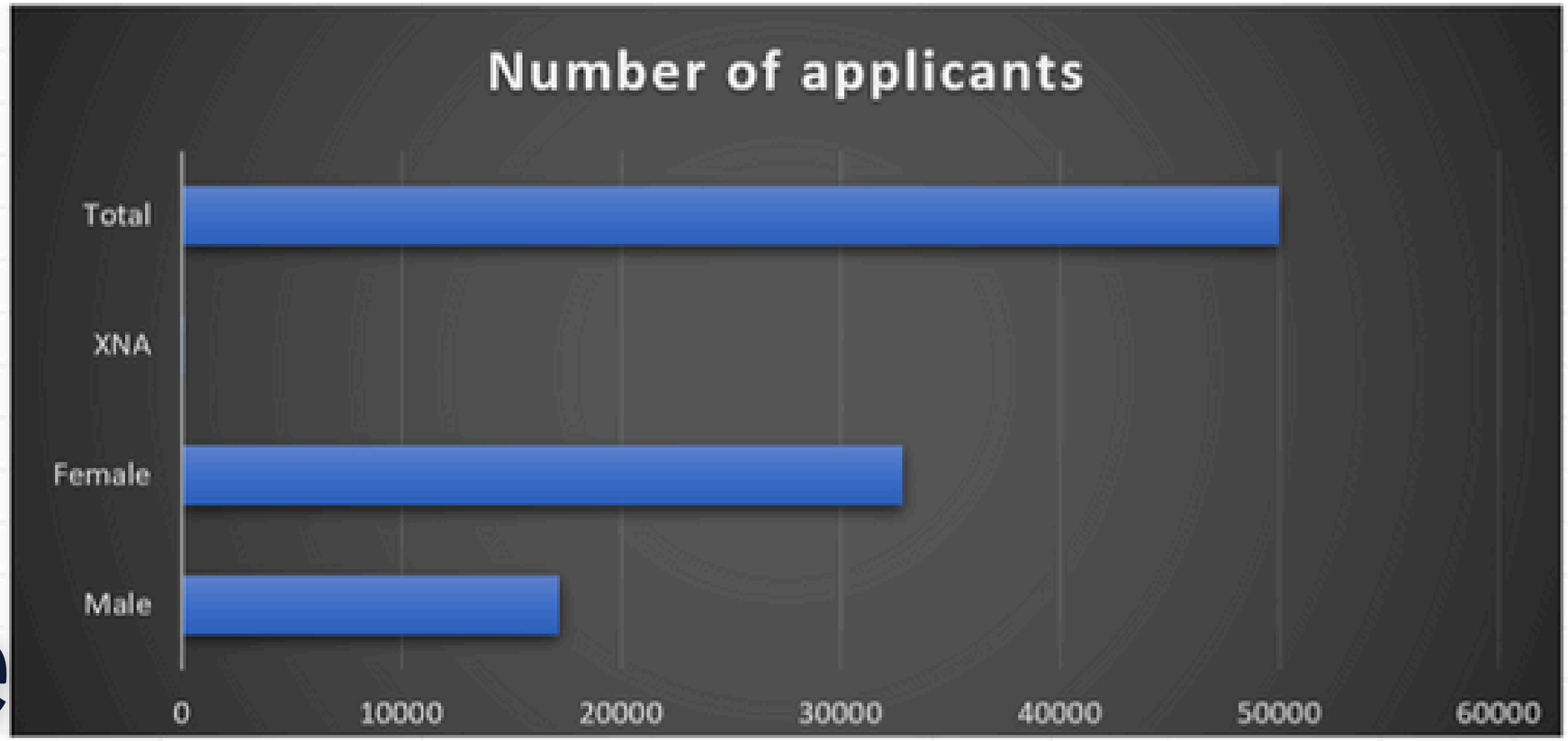
3. Bivariate Analysis:

Explore relationships between variables and the target variable (e.g., loan default).

- Steps:
  - Create scatter plots:
    - Highlight two columns of interest and use Insert > Scatter Plot.
    - Add trendlines if needed to assess relationships.
  - Generate cross-tabulations with Pivot Tables:
    - Place one variable in Rows and another in Columns, and use Counts or percentages in Values.
  - Use heatmaps to visualize correlations:
    - Calculate correlations using =CORREL(range1, range2) and color-code results using Conditional Formatting.
- Purpose: Detect patterns or trends (e.g., higher loan amounts correlate with higher default rates).

Category	Number of applicants
Male	17174
Female	32823
XNA	2
Total	49997

Univariate



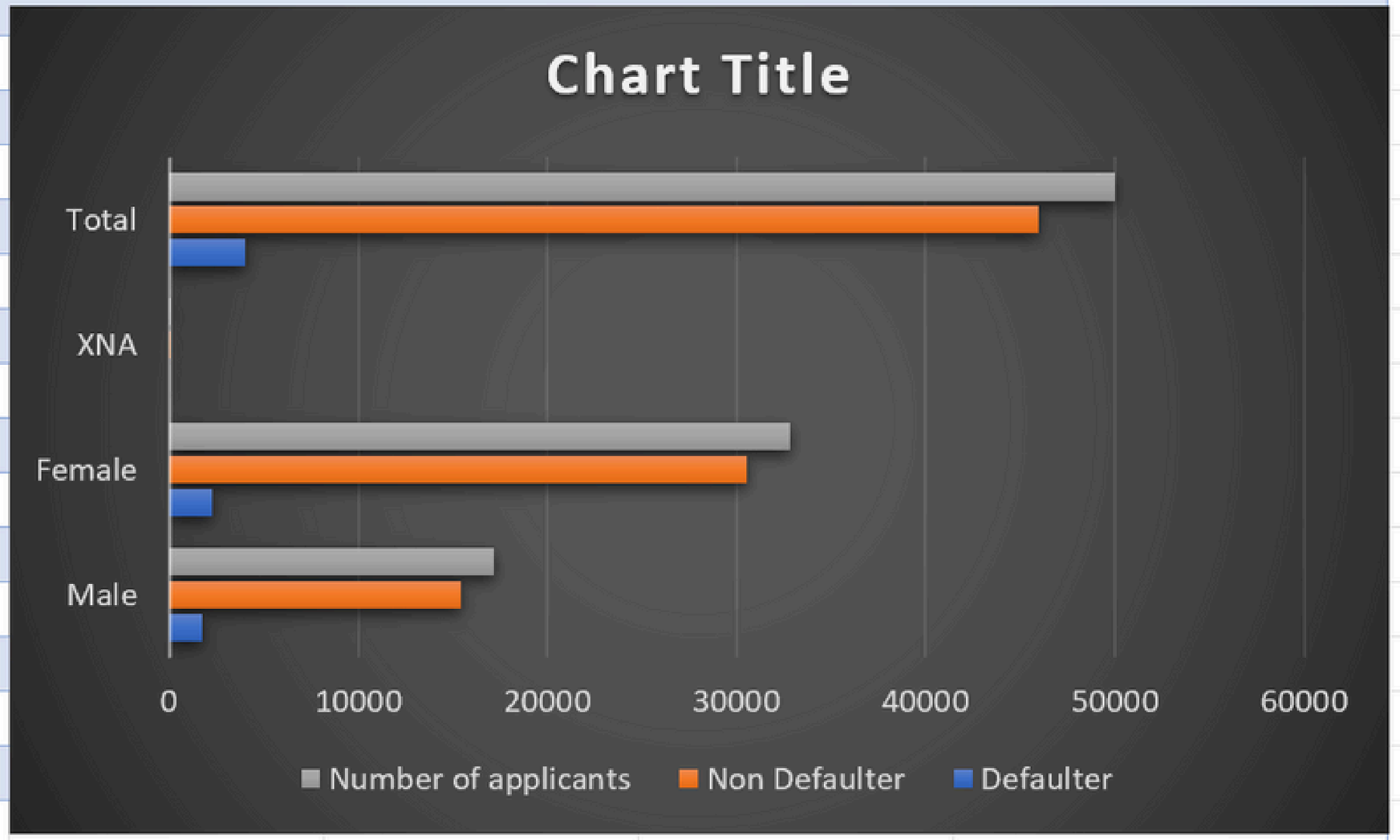
CODE_GENDER (All)

Row Labels	Sum of Number of applicants
Female	32823
Male	17174
Grand Total	**49997**

Category

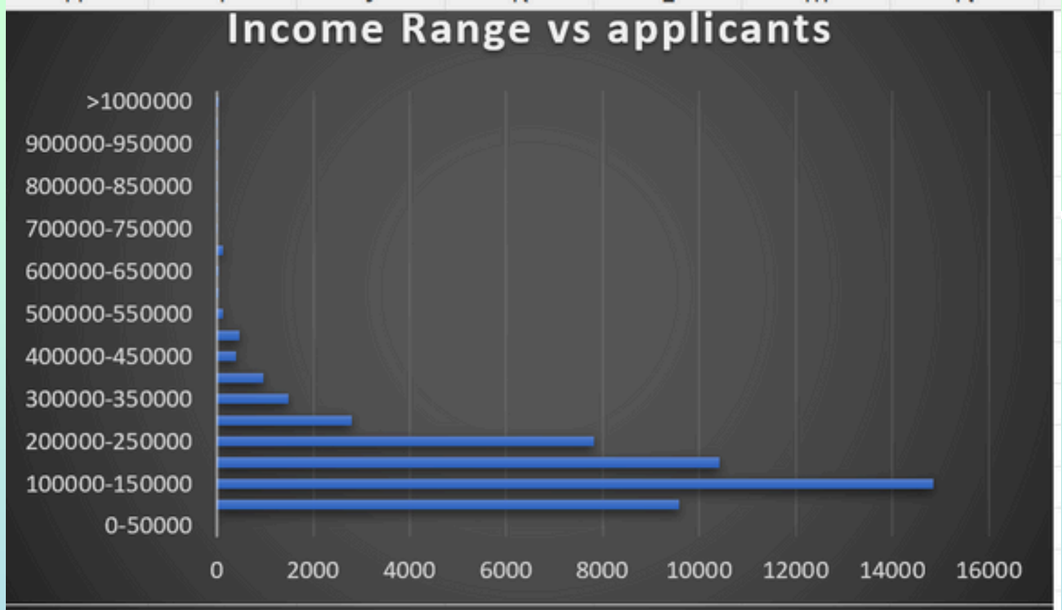
- Female
- Male
- Total
- XNA
- (blank)

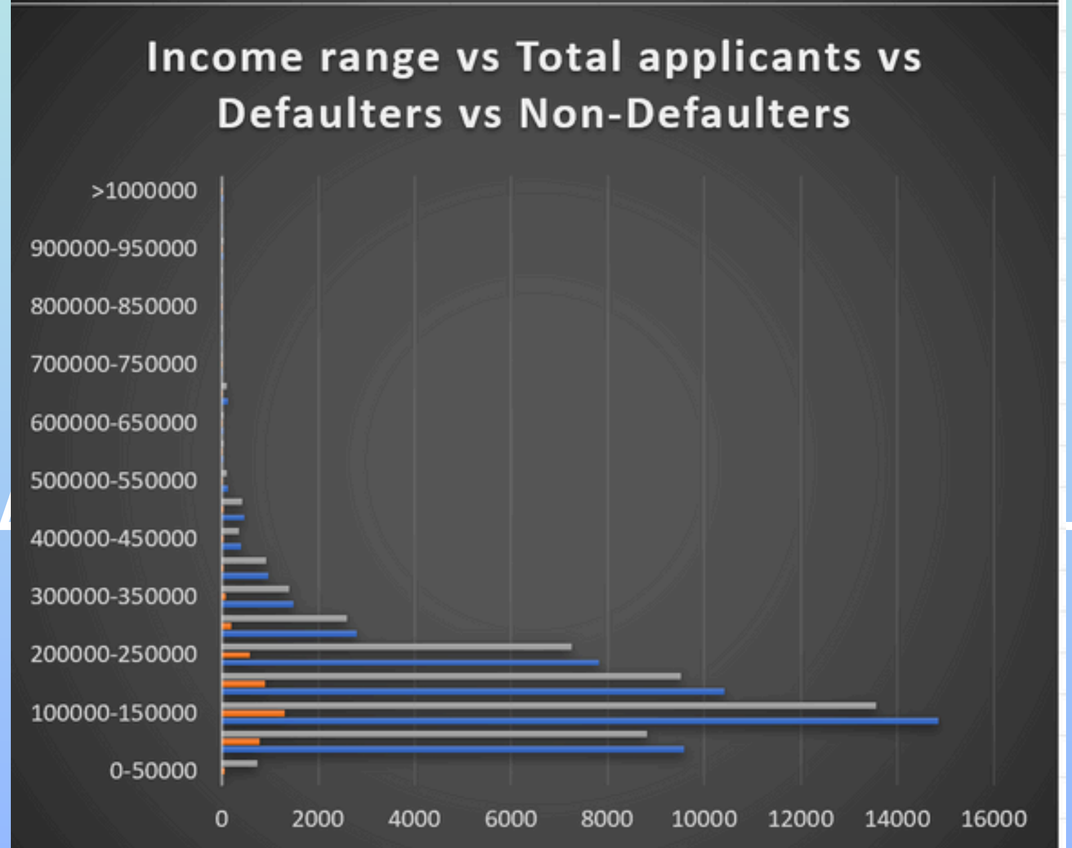
Category	Defaulter	Non Defaulter	Number of applicants
Male	1762	15412	17174
Female	2264	30559	32823
XNA	0	2	2
Total	4026	45973	49997

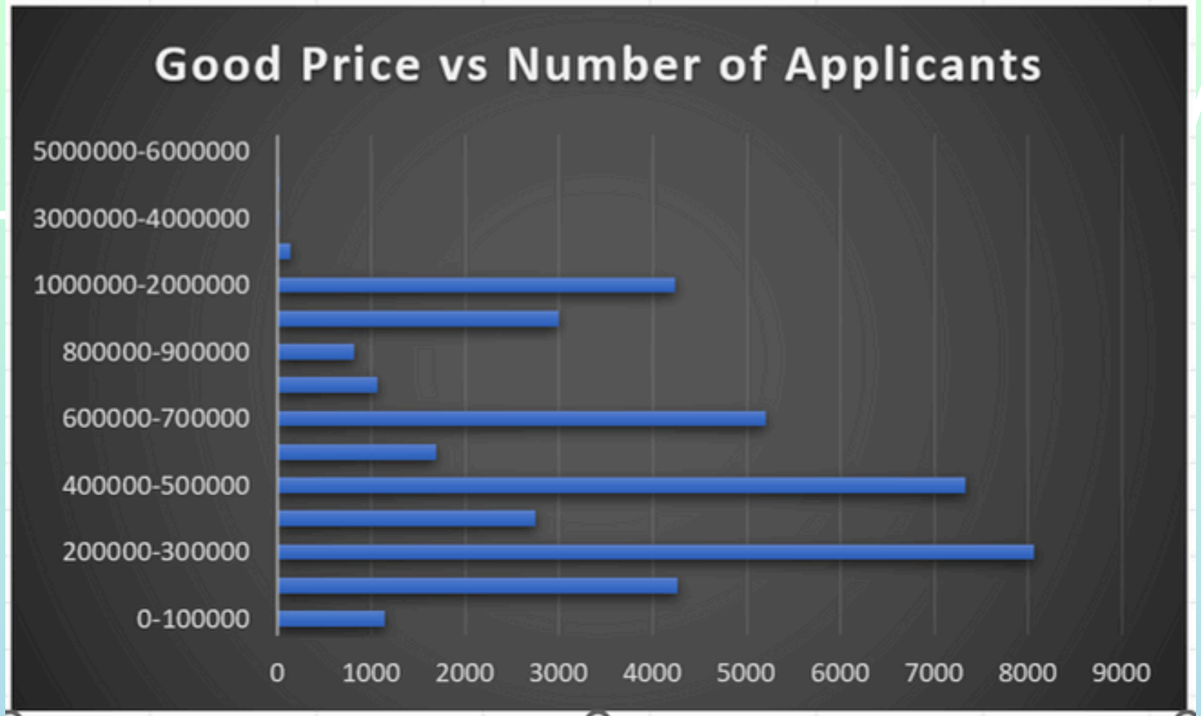


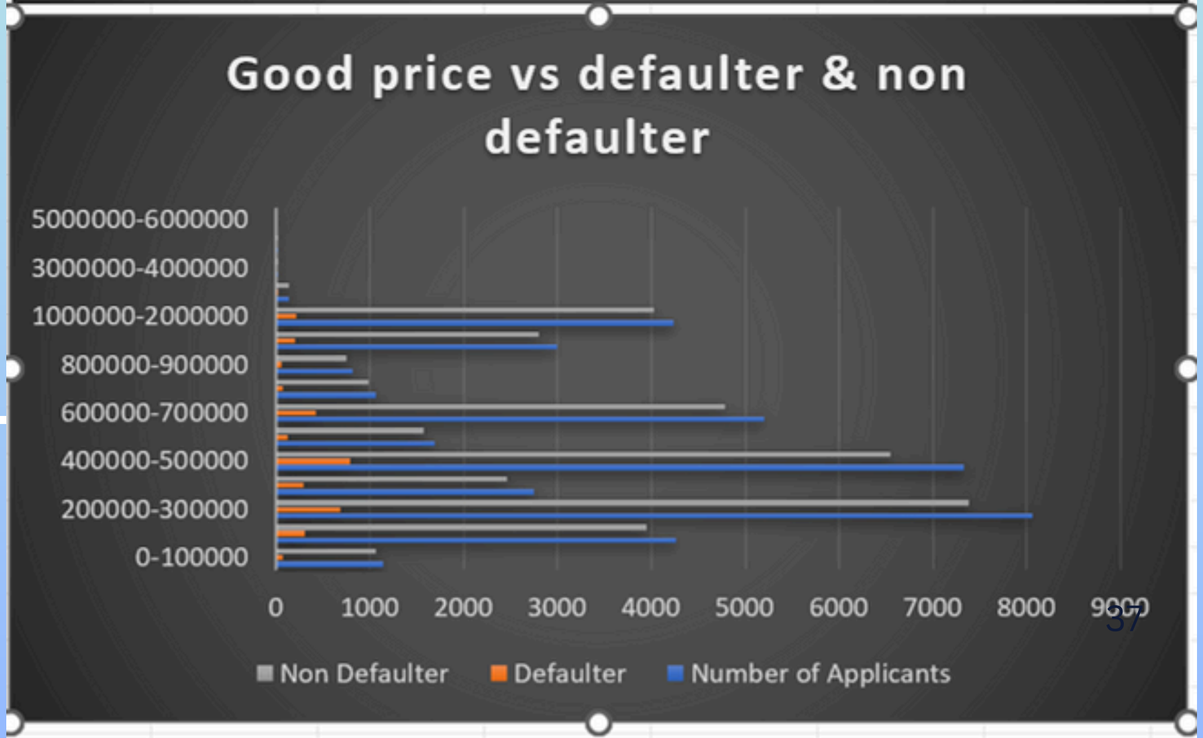
Segment Univariate

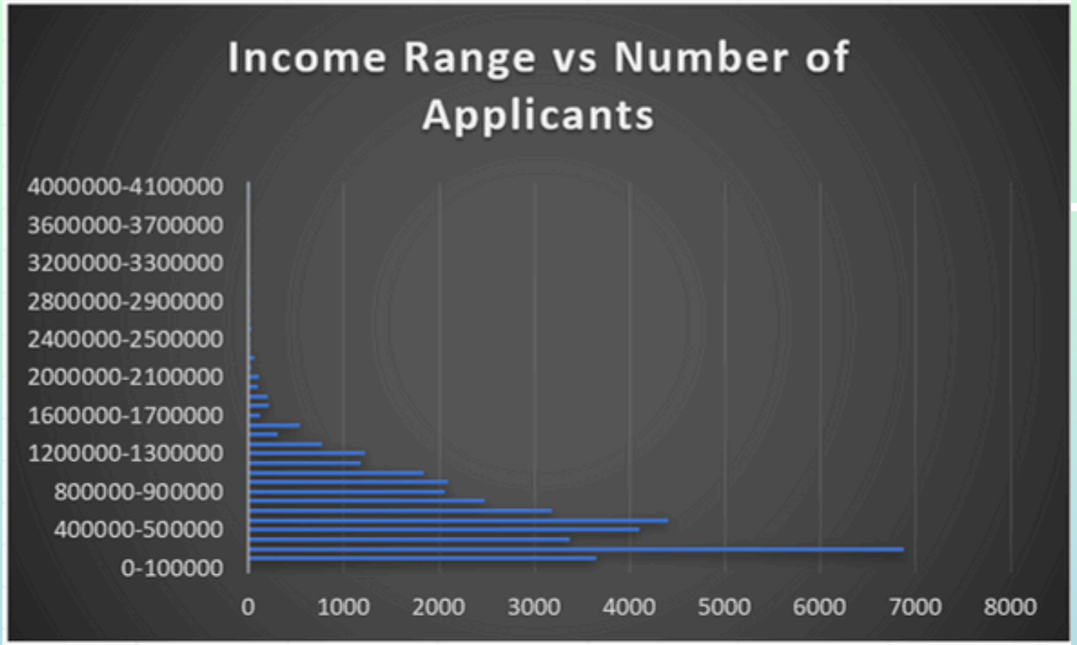
AMT_INCOME_TOTA	Income_Rang	Total Applicant	Defaulters	Non-Defaulters
202500	0-50000	0	63	741
270000	50000-100000	9588	782	8806
67500	100000-150000	14852	1298	13554
135000	150000-200000	10408	890	9518
121500	200000-250000	7818	576	7242
99000	250000-300000	2788	188	2600
171000	300000-350000	1481	83	1398
360000	350000-400000	957	48	909
112500	400000-450000	393	30	363
135000	450000-500000	456	37	419
112500	500000-550000	124	9	115
38419.155	550000-600000	43	5	38
67500	600000-650000	40	1	39
225000	650000-700000	117	8	109
189000	700000-750000	22	1	21
157500	750000-800000	11	0	11
108000	800000-850000	21	2	19
81000	850000-900000	5	0	5
112500	900000-950000	30	2	28
90000	950000-1000000	1	0	0
135000	>1000000	40	3	0
202500				
450000				
83250				
135000				

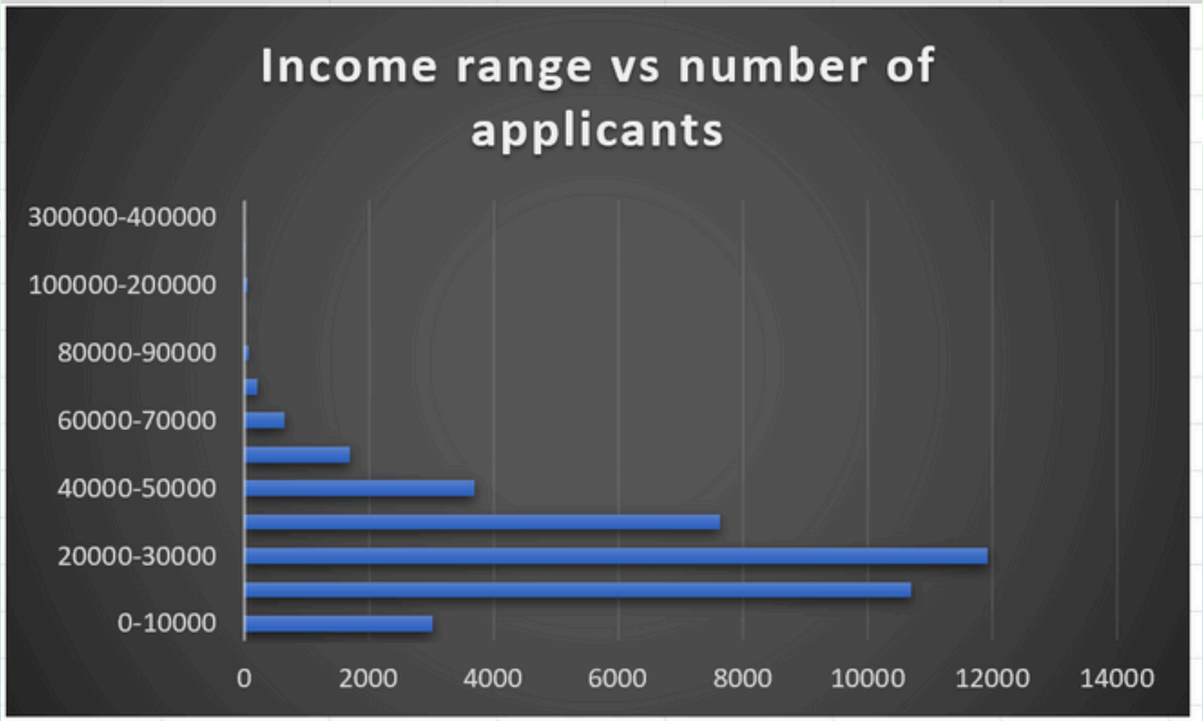

Income Range vs applicants
>1000000
900000-950000
800000-850000
700000-750000
600000-650000
500000-550000
400000-450000
300000-350000
200000-250000
100000-150000
0-50000
0
2000
4000
6000
8000
10000
12000
14000
16000

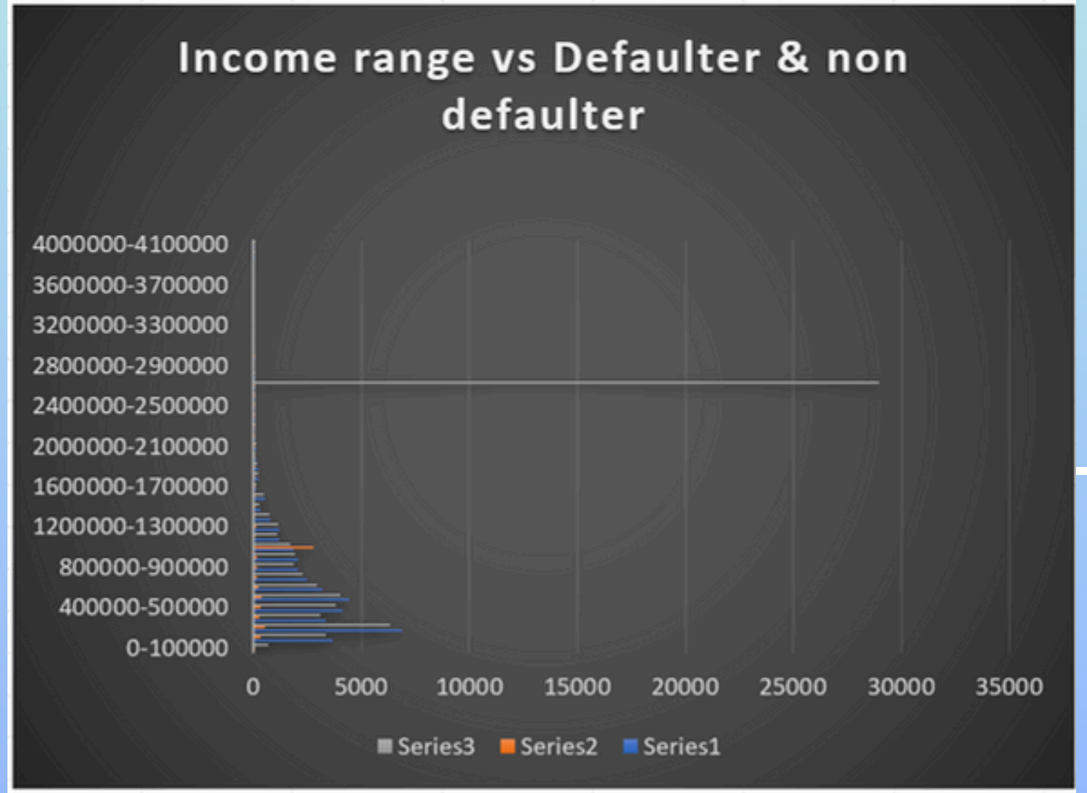

Income range vs Total applicants vs Defaulters vs Non-Defaulters
>1000000
900000-950000
800000-850000
700000-750000
600000-650000
500000-550000
400000-450000
300000-350000
200000-250000
100000-150000
0-50000
0
2000
4000
6000
8000
10000
12000
14000
16000

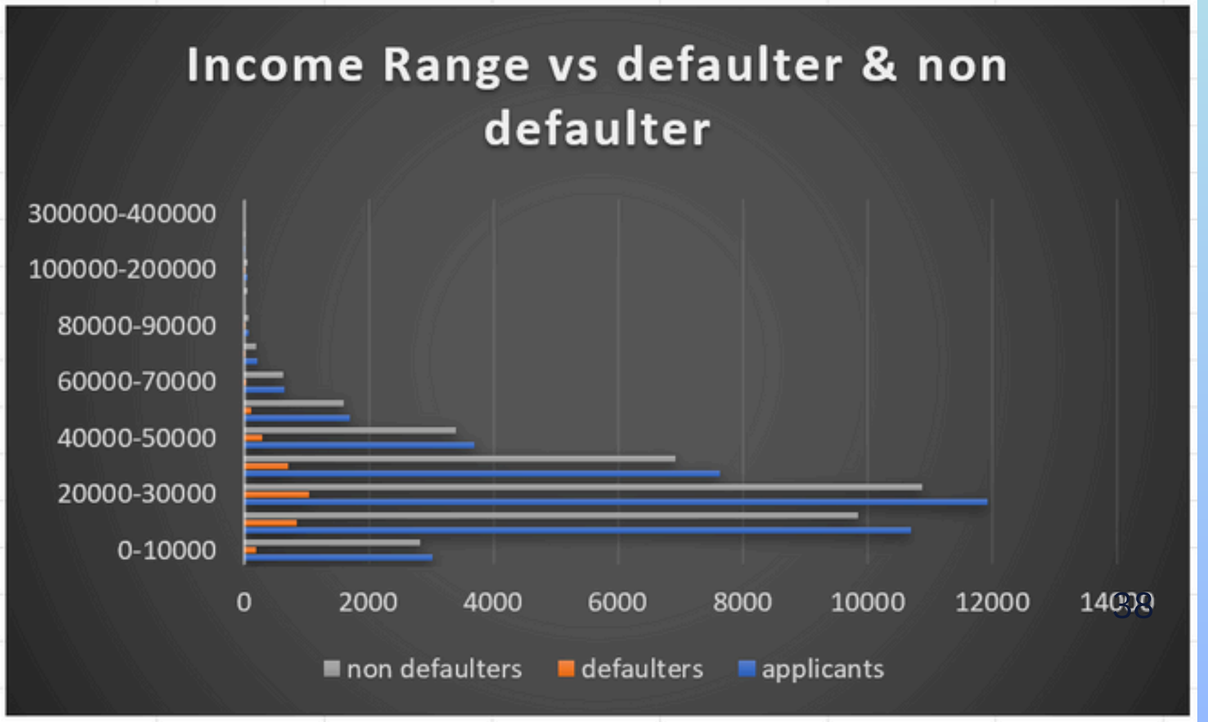

Good Price vs Number of Applicants
5000000-6000000
3000000-4000000
1000000-2000000
800000-900000
600000-700000
400000-500000
200000-300000
0-100000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000


Good price vs defaulter & non defaulter
5000000-6000000
3000000-4000000
1000000-2000000
800000-900000
600000-700000
400000-500000
200000-300000
0-100000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Non Defaulter
Defaulter
Number of Applicants


Income Range vs Number of Applicants
4000000-4100000
3600000-3700000
3200000-3300000
2800000-2900000
2400000-2500000
2000000-2100000
1600000-1700000
1200000-1300000
800000-900000
400000-500000
0-100000
0
1000
2000
3000
4000
5000
6000
7000
8000

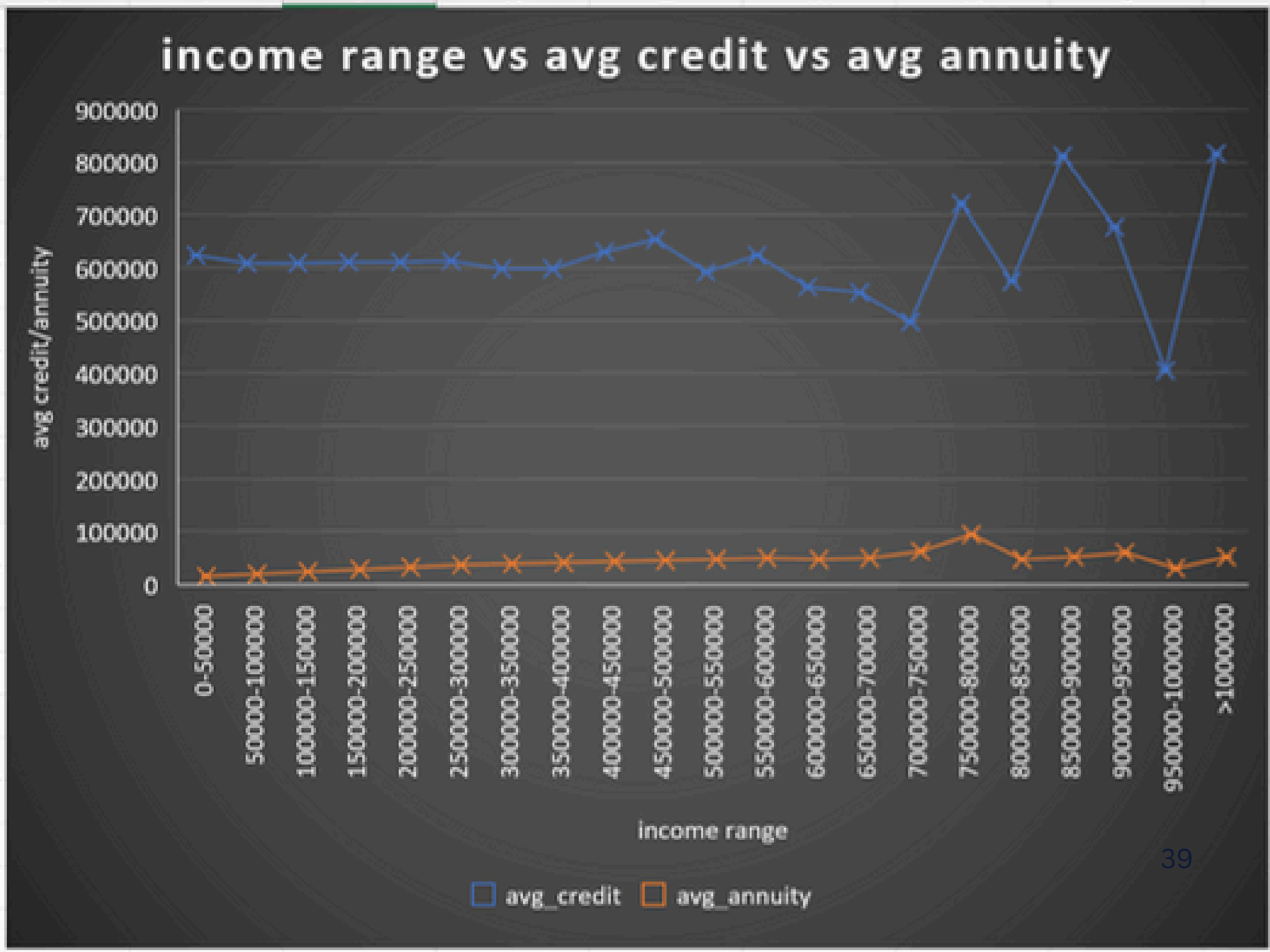

Income range vs number of applicants
300000-400000
100000-200000
80000-90000
60000-70000
40000-50000
20000-30000
0-10000
0
2000
4000
6000
8000
10000
12000
14000


Income range vs Defaulter & non defaulter
4000000-4100000
3600000-3700000
3200000-3300000
2800000-2900000
2400000-2500000
2000000-2100000
1600000-1700000
1200000-1300000
800000-900000
400000-500000
0-100000
0
5000
10000
15000
20000
25000
30000
35000
Series3
Series2
Series1


Income Range vs defaulter & non defaulter
300000-400000
100000-200000
80000-90000
60000-70000
40000-50000
20000-30000
0-10000
0
2000
4000
6000
8000
10000
12000
14000
non defaulters
defaulters
applicants

Bivariate

Income_Rang	avg_credit	avg_annuity
0-50000	623794.649	14151.99685
50000-100000	607827.981	18525.16367
100000-150000	608868.601	23811.0152
150000-200000	610981.252	28601.75194
200000-250000	609620.778	32875.69724
250000-300000	613038.438	36131.96455
300000-350000	598461.257	39045.61974
350000-400000	598587.842	41233.91502
400000-450000	629081.661	43487.16771
450000-500000	653128.563	45862.5625
500000-550000	592347.375	46678.78125
550000-600000	623037.838	48854.38235
600000-650000	564380.016	48108.65625
650000-700000	552699.574	49987.29255
700000-750000	496989	62929.18421
750000-800000	721121.143	95227.71429
800000-850000	574122.214	47424.21429
850000-900000	812246.4	51605.1
900000-950000	677430.173	59482.38462
950000-1000000	405000	30073.5
>1000000	815365.636	51568.22727



Task 5

Identify Top Correlations for Different Scenarios

To gain insights into the factors driving loan defaults, you can segment the dataset based on different scenarios (e.g., clients with payment difficulties vs. all other clients) and assess the correlations between variables and the target variable (e.g., loan default). Here's how to conduct this analysis in Excel:

1. Segmenting the Dataset

To analyze the data by scenario, first, divide your dataset into relevant segments (e.g., clients with payment difficulties vs. all other clients).

- Steps:
  a. Use Filters to segment the data based on the relevant scenario (e.g., filter out clients with payment difficulties or create a new column indicating payment difficulty).
  b. Alternatively, use IF formulas to create a new column that flags clients in different scenarios.
     - Example: =IF(PaymentStatus="Difficult", "Payment Difficulty", "No Difficulty").

2. Calculate Correlation Coefficients within Each Segment

Once the dataset is segmented, you can calculate the correlation coefficients between the variables and the target variable (e.g., loan default).

- Steps:
  a. Select the segment you want to analyze (e.g., clients with payment difficulties).

b.Use the CORREL function to calculate the correlation coefficient between the variables and the target variable:

- Example: =CORREL(A2:A100, B2:B100) where A2:A100 is the range for the independent variable (e.g., loan amount) and B2:B100 is the range for the target variable (e.g., loan default).
  - Repeat the process for all relevant variables in the segment.
  - Rank the correlation coefficients to identify the strongest correlations (use Sort or Conditional Formatting to highlight the top correlations).

3. Rank and Identify Top Indicators of Loan Default

After calculating the correlation coefficients for each variable within the segment, rank them to identify which variables have the strongest relationship with the target variable (e.g., loan default).

- Steps:
  - Create a table that lists the variables and their respective correlation coefficients for each segment.
  - Sort the correlation coefficients in descending order (highest to lowest) to identify the top indicators.
  - Highlight the top correlations using conditional formatting or color coding.

4. Visualize Correlations using a Heatmap or Correlation Matrix

A visual representation of the correlations will help quickly identify which variables are most strongly correlated with the target variable.

- Steps:
  - Use Conditional Formatting to color-code the correlation matrix:
    - Select the correlation coefficients in a table.

    - Go to Home > Conditional Formatting > Color Scales and choose a color scale.
  - Alternatively, create a Correlation Matrix by listing the variables in both rows and columns and filling the cells with the correlation coefficients.
  - Use a Heatmap to visually highlight the strongest correlations. Excel offers options to format cells with color gradients based on correlation values.

Defaulters

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	Year BIRTH	Year EMPLOYED	Year REGISTERED	Days ID
CNT_CHILDREN	100.00%	0.92%	0.54%	-0.50%	-0.14%	-2.32%	-34.28%	-24.32%	-18.14%	3.84%
AMT_INCOME_TOTAL	0.92%	100.00%	6.09%	0.00%	0.28%	2.49%	-1.64%	-2.88%	-0.85%	-0.44%
AMT_CREDIT	0.54%	6.09%	100.00%	0.33%	0.58%	9.55%	5.02%	-7.27%	-0.47%	0.10%
AMT_ANNUITY	-0.50%	0.00%	0.33%	100.00%	77.45%	0.42%	0.61%	0.45%	1.25%	0.45%
AMT_GOODS_PRICE	-0.14%	0.28%	0.58%	77.45%	100.00%	0.61%	0.40%	0.56%	0.93%	0.38%
REGION_POPULATION_RELATIVE	-2.32%	2.49%	9.55%	0.42%	0.61%	100.00%	3.18%	-0.01%	5.84%	0.19%
Year BIRTH	-34.28%	-1.64%	5.02%	0.61%	0.40%	3.18%	100.00%	61.93%	33.51%	24.53%
Year EMPLOYED	-24.32%	-2.88%	-7.27%	0.45%	0.56%	-0.01%	61.93%	100.00%	20.55%	26.28%
Year REGISTERED	-18.14%	-0.85%	-0.47%	1.25%	0.93%	5.84%	33.51%	20.55%	100.00%	8.78%
Days ID	3.84%	-0.44%	0.10%	0.45%	0.38%	0.19%	24.53%	26.28%	8.78%	100.00%

Non Defaulters

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	Year BIRTH	Year EMPLOYED	Year REGISTERED	Days ID
CNT_CHILDREN	100.00%	3.57%	0.60%	-0.15%	0.26%	-2.15%	-34.92%	-24.69%	-18.35%	3.76%
AMT_INCOME_TOTAL	3.57%	100.00%	37.28%	0.28%	0.67%	17.64%	-8.05%	-16.29%	-6.95%	-4.20%
AMT_CREDIT	0.60%	37.28%	100.00%	0.46%	0.72%	9.56%	4.28%	-7.84%	-0.82%	-0.20%
AMT_ANNUITY	-0.15%	0.28%	0.46%	100.00%	77.42%	0.31%	0.68%	0.44%	1.06%	0.63%
AMT_GOODS_PRICE	0.26%	0.67%	0.72%	77.42%	100.00%	0.49%	0.40%	0.49%	0.90%	0.51%
REGION_POPULATION_RELATIVE	-2.15%	17.64%	9.56%	0.31%	0.49%	100.00%	-0.26%	-0.26%	5.73%	-0.01%
Year BIRTH	-34.92%	-8.05%	4.28%	0.68%	0.40%	2.93%	100.00%	62.08%	33.69%	24.39%
Year EMPLOYED	-24.69%	-16.29%	-7.84%	0.44%	0.49%	-0.26%	62.08%	100.00%	20.56%	26.50%
Year REGISTERED	-18.35%	-6.95%	-0.82%	1.06%	0.90%	5.73%	33.69%	20.56%	100.00%	8.87%
Days ID	3.76%	-4.20%	-0.20%	0.63%	0.51%	-0.01%	24.39%	26.50%	8.87%	100.00%

Top Correlations

TOP CORRELATION FOR DEFAULTERS		
VARIABLE 1	VARIABLE 2	CORRELATION
AMT_ANNUITY	AMT_GOODS_PRICE	77.45%
AMT_GOODS_PRICE	AMT_ANNUITY	77.45%
Year BIRTH	Year EMPLOYED	61.93%
Year EMPLOYED	Year BIRTH	61.93%
Year REGISTERED	Year BIRTH	33.51%

TOP CORRELATIONS FOR NON-DEFAULTERS		
VARIABLE 1	VARIABLE 2	CORRELATION
AMT_ANNUITY	AMT_GOODS_PRICE	77.42%
AMT_GOODS_PRICE	AMT_ANNUITY	77.42%
Year BIRTH	Year EMPLOYED	62.08%
Year EMPLOYED	Year BIRTH	62.08%
AMT_INCOME_TOTA	AMT_CREDIT	37.28%

Borrow with caution, repay with discipline, and your financial future will shine.

Smart borrowing today leads to financial freedom tomorrow.

Link to the project

CLICK: LINK TO PROJECT DRIVE

Ankita Taneja

DATA ANALYST

trainity

Your Reviews

What can you say about My projects? Share it here!

Contact Us

Gurugram

0-7011-334-048

Venusgirlatwork@gmail.com

@ig_she_has_no_idea

www.linkedin.com/in/ankita-taneja/

Thank You!