BANK LOAN CASE STUDY

PROJECT DESCRIPTION

This project aims to analyze bank loan data to identify crucial factors affecting loan approvals and defaults. By applying data analytics, we will derive meaningful insights that can help financial institutions enhance risk assessment, minimize loan defaults, and improve decision-making in the lending process. The findings will support the development of more efficient and data-driven loan management strategies.

APPROACH

1. Data collection which was provided by trainity.
2. Data Cleaning(missing values, blanks etc).
3. Data imputation and Exploratory Data Analytics(EDA).
4. Data segmentation
5. Univariate, Segmented univariate and bivariate analysis

Using excel in-built functions.

TECHSTACK USED

Microsoft Excel 2016 was used for this project as it has many versatile inbuilt arithmetical and statistical functions.

And it is a easy to understand the best spreadsheet software.

INSIGHTS

1. **Identify Missing Data and Deal with it Appropriately:** Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.

OUTPUT-: Columns with more than 50% missing values were deleted, and columns with less than 50% missing values were corrected using average, median, and mode values to ensure the accuracy of the data.

bank loan case study 2 - Excel

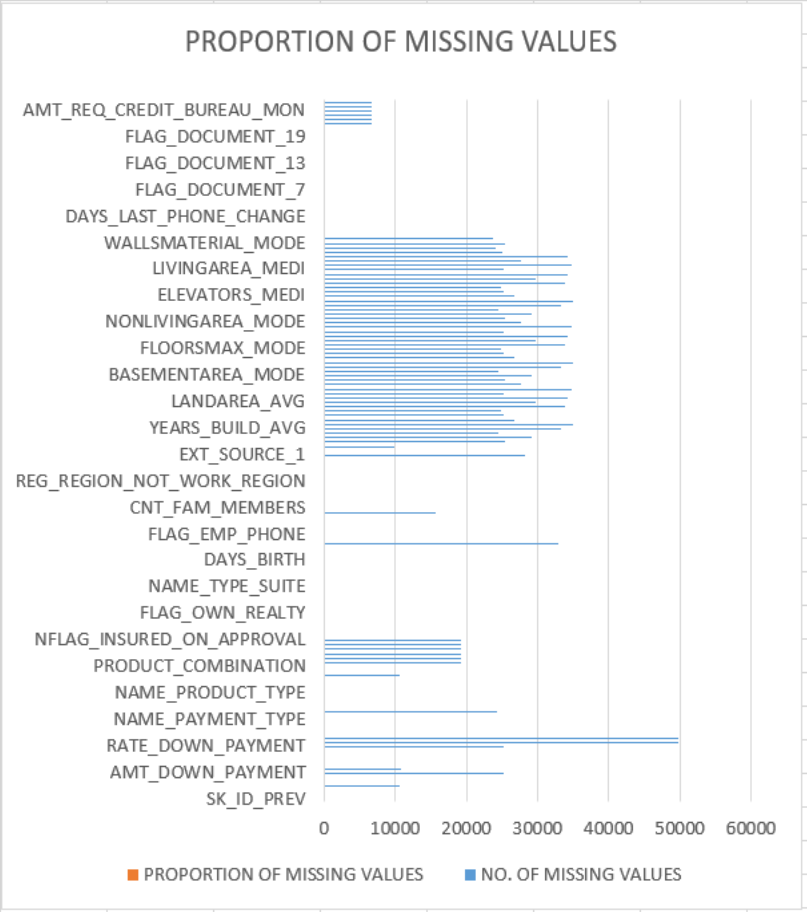
File Home Insert Page Layout Formulas Data Review View Acrobat Tell me what you want to do... Share

Paste Cut Copy Format Painter | Calibri 11 | Wrap Text Merge & Center | General | Conditional Formatting Format as Table | Normal Bad Good Neutral | Insert Delete Format | AutoSum Fill Clear | Sort & Filter Find & Select | Create a PDF

Clipboard Font Alignment Number Styles Cells Editing Adobe Acrobat

M20

	COLUMN NAME	NO. OF MISSING VALUES	PROPORTION OF MISSING VALUES
2	SK_ID_PREV	0	0
3	SK_ID_CURR	0	0
4	NAME_CONTRACT_TYPE	0	0
5	AMT_ANNUITY	10592	21.18442369
6	AMT_APPLICATION	0	0
7	AMT_CREDIT	0	0
8	AMT_DOWN_PAYMENT	25198	50.39700794
9	AMT_GOODS_PRICE	10744	21.48842977
10	WEEKDAY_APPR_PROCESS_START	0	0
11	HOUR_APPR_PROCESS_START	0	0
12	FLAG_LAST_APPL_PER_CONTRACT	0	0
13	NFLAG_LAST_APPL_IN_DAY	0	0
14	RATE_DOWN_PAYMENT	25198	50.39700794
15	RATE_INTEREST_PRIMARY	49834	99.6699934
16	RATE_INTEREST_PRIVILEGED	49834	99.6699934
17	NAME_CASH_LOAN_PURPOSE	0	0
18	NAME_CONTRACT_STATUS	0	0
19	DAYS_DECISION	0	0
20	NAME_PAYMENT_TYPE	0	0
21	CODE_REJECT_REASON	0	0
22	NAME_TYPE_SUITE	24243	48.48696974
23	NAME_CLIENT_TYPE	0	0
24	NAME_GOODS_CATEGORY	0	0
25	NAME_PORTFOLIO	0	0
26	NAME_PRODUCT_TYPE	0	0
27	CHANNEL_TYPE	0	0
28	SELLERPLACE_AREA	0	0
29	NAME_SELLER_INDUSTRY	0	0



previous_application | EDA | TASK 1 | TASK 2 | TASK2(1) | TASK 3 | TASK 5(0) | TASK 5(1) | CORRELATION

Ready 100%

INSIGHTS

2. **Identify Outliers in the Dataset:** Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.

OUTPUT-: Quartile(1,2,3), IQR were used to calculate the no. of outliers then the missing data was corrected.

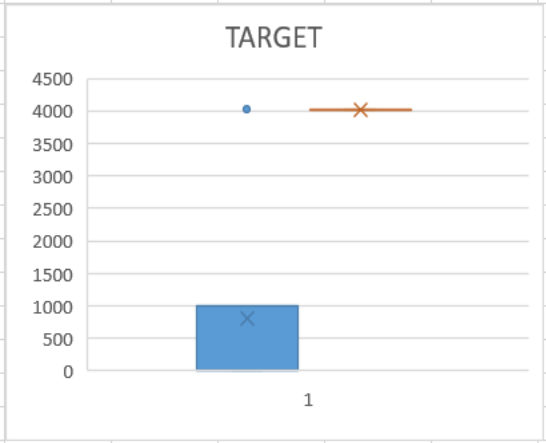
bank loan case study 2 - Excel

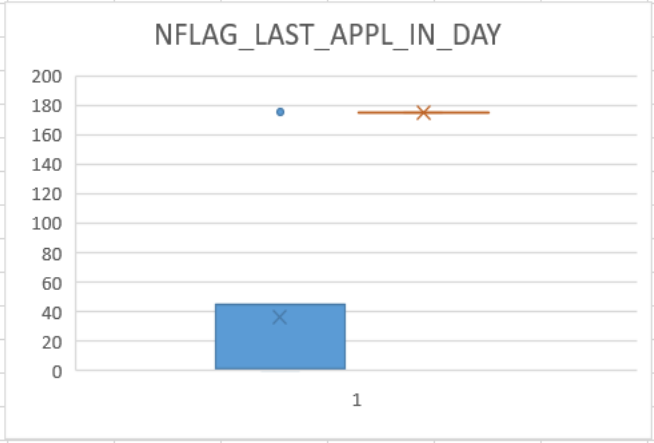
File Home Insert Page Layout Formulas Data Review View Acrobat Tell me what you want to do... Share

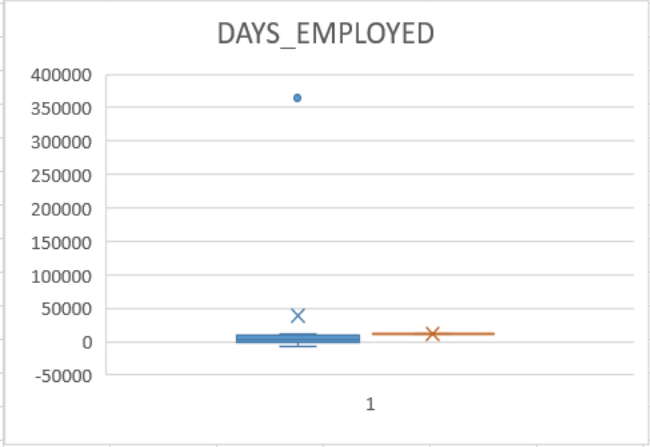
Paste Cut Copy Format Painter | Calibri 11 | Wrap Text Merge & Center | General | Conditional Formatting Format as Table | Normal Bad Good Neutral | Insert Delete Format | AutoSum Fill Clear | Sort & Filter Find & Select | Create a PDF

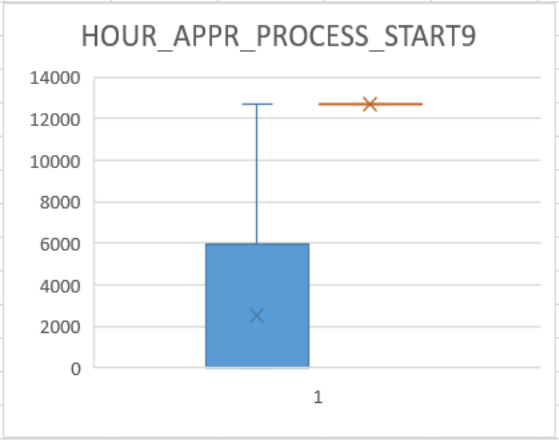
Clipboard Font Alignment Number Styles Cells Editing Adobe Acrobat

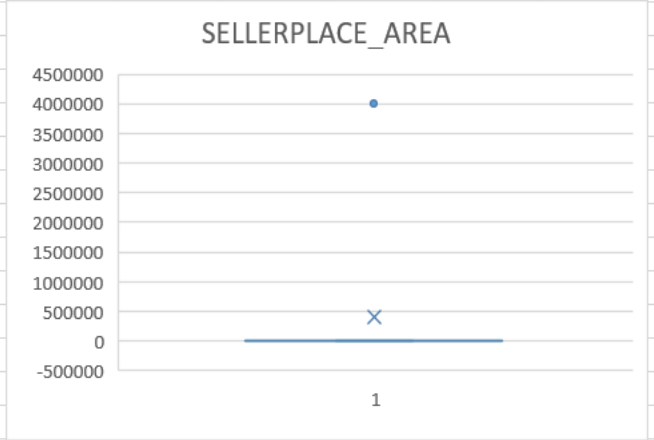
BT29

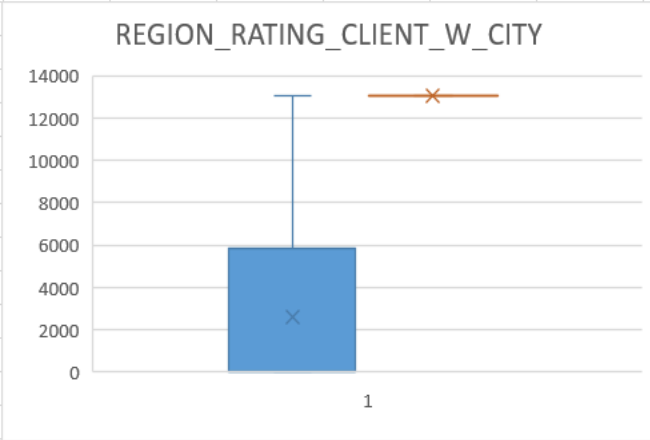












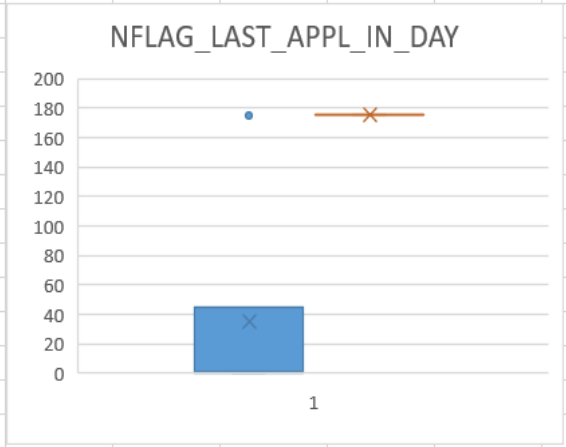
... previous_application EDA TASK 1 TASK 2 TASK2(1) TASK 3 TASK 5(0) TASK 5(1) CORRELATION

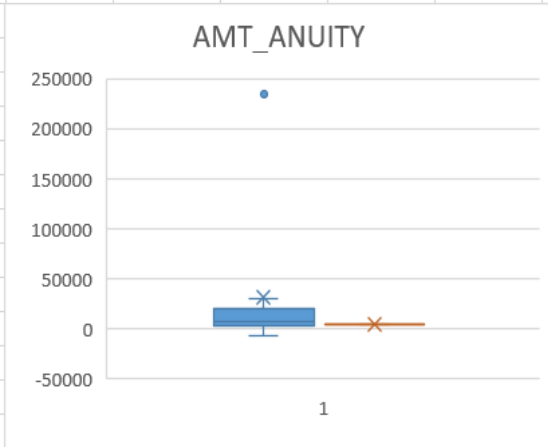
Ready 100%

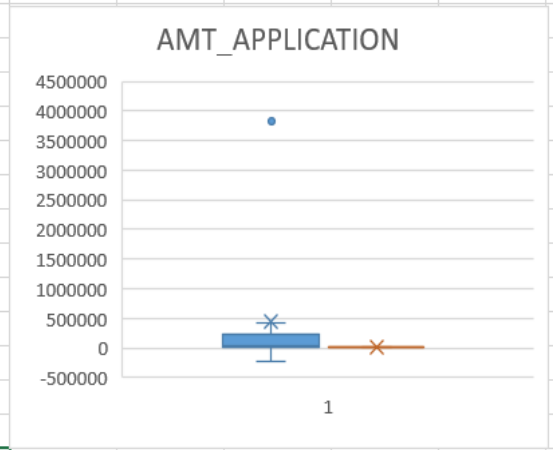
Cut Copy Format Painter Paste | Calibri 11 | Wrap Text | General | Conditional Formatting | Format as Table | Normal Bad Good Neutral | Insert Delete Format | AutoSum Fill Clear | Sort & Filter Find & Select | Create a PDF

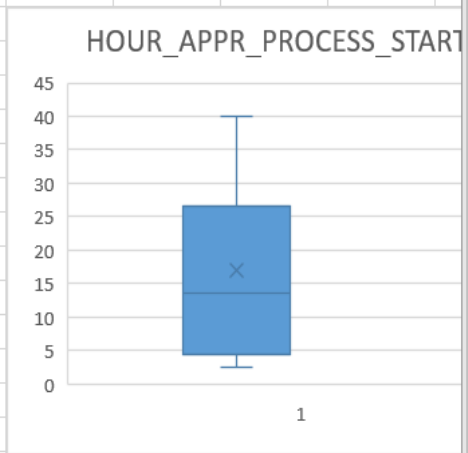
Clipboard Font Alignment Number Styles Cells Editing Adobe Acrobat

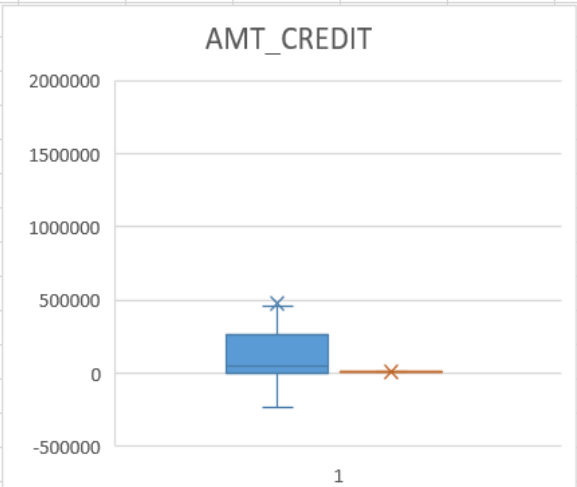
M15

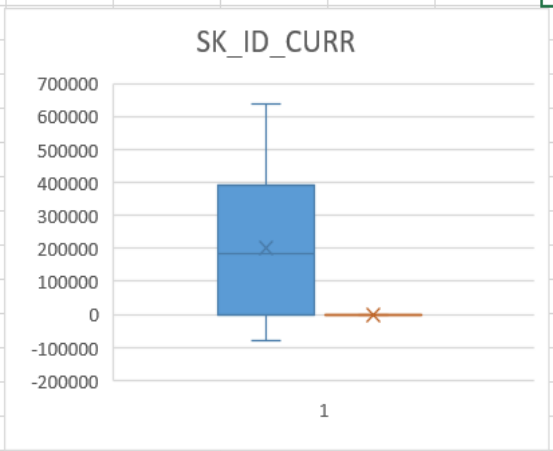


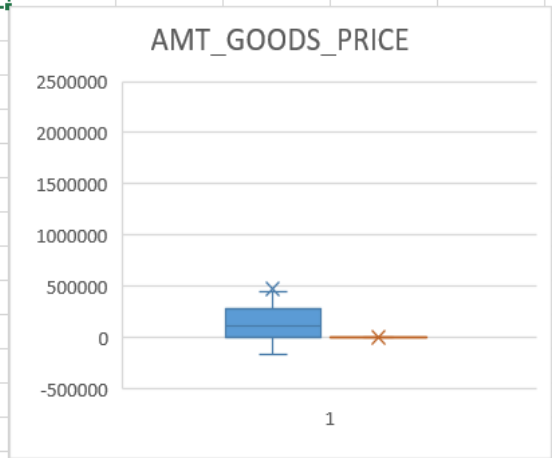


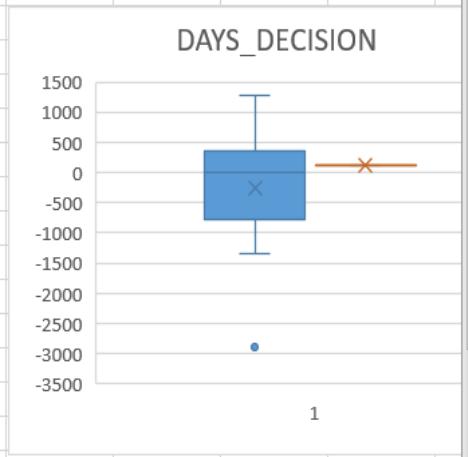












previous_application | EDA | TASK 1 | TASK 2 | TASK2(1) | TASK 3 | TASK 5(0) | TASK 5(1) | CORRELATION

Ready

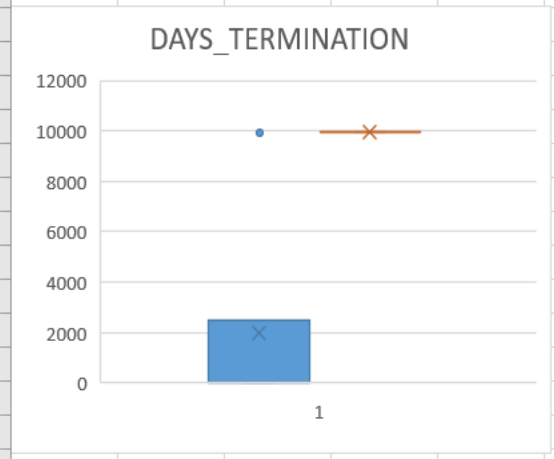
bank loan case study 2 - Excel

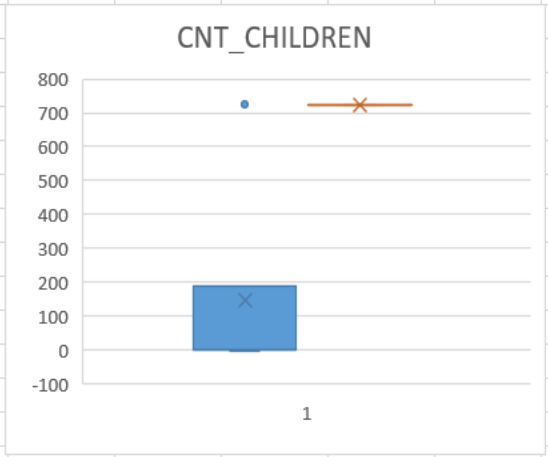
File Home Insert Page Layout Formulas Data Review View Acrobat Tell me what you want to do... Share

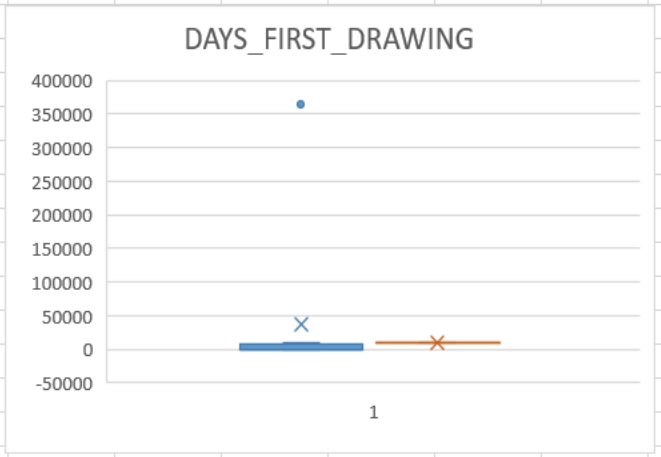
Paste Cut Copy Format Painter | Calibri 11 | Wrap Text Merge & Center | General | Conditional Formatting Format as Table | Normal Bad Good Neutral | Insert Delete Format | AutoSum Fill Clear | Sort & Filter Find & Select | Create a PDF

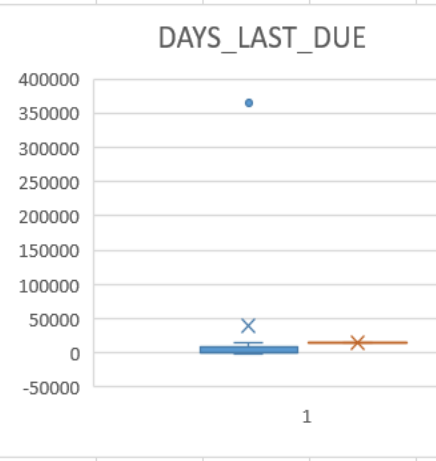
Clipboard Font Alignment Number Styles Cells Editing Adobe Acrobat

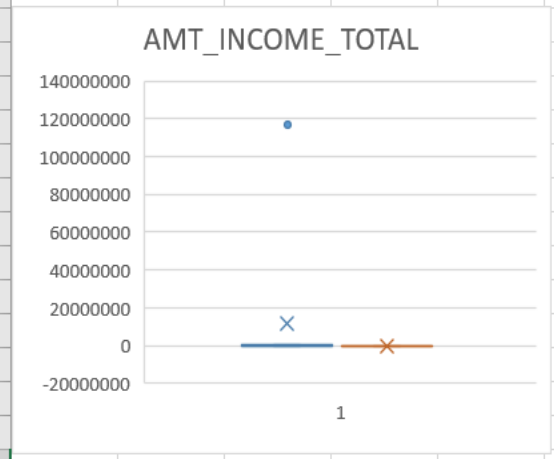
AS29

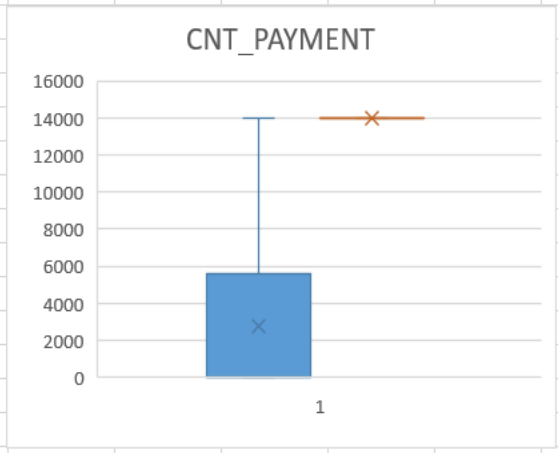


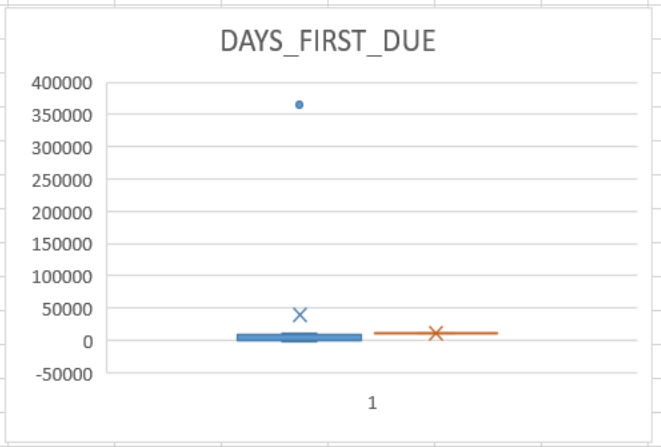


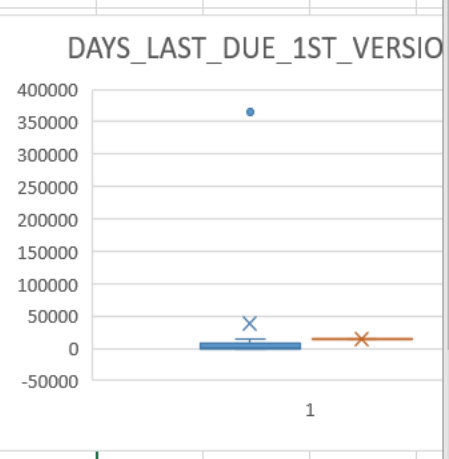












... previous_application EDA TASK 1 TASK 2 TASK2(1) TASK 3 TASK 5(0) TASK 5(1) CORRELATION

Ready 100%

INSIGHTS

3. **Analyze Data Imbalance-:**Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.

OUTPUT-: TARGET(1)(Defaulters) -: 4026

TARGET(0)(Payment on time) -: 45973

Ratio -: 11.41

bank loan case study 2 - Excel

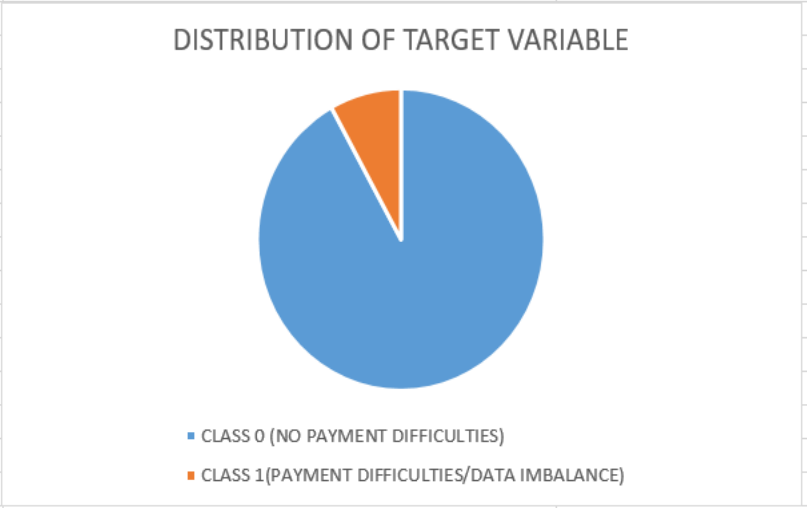
File Home Insert Page Layout Formulas Data Review View Acrobat Tell me what you want to do... Share

Paste Cut Copy Format Painter | Calibri 11 | Wrap Text Merge & Center | General | Conditional Formatting Format as Table | Normal Bad Good Neutral | Insert Delete Format | AutoSum Fill Clear | Sort & Filter Find & Select | Create a PDF

Clipboard Font Alignment Number Styles Cells Editing Adobe Acrobat

C20

	A	B	C	D
1	TARGET	CLASS 0 (NO PAYMENT DIFFICULTIES)	CLASS 1(PAYMENT DIFFICULTIES/DATA IMBALANCE)	RATIO OF DATA IMBALANCE
2	1	45973	4026	11.41902633
3	0			
4	0			
5	0			
6	0			
7	0			
8	0			
9	0			
10	0			
11	0			
12	0			
13	0			
14	0			
15	0			
16	0			
17	0			
18	0			
19	0			
20	0			
21	0			
22	0			
23	0			
24	0			
25	0			
26	0			
27	0			
28	1			
29	0			



... previous_application EDA TASK 1 TASK 2 TASK2(1) TASK 3 TASK 5(0) TASK 5(1) CORRELATION

Ready 100%

INSIGHTS

4. **Perform Univariate, Segmented Univariate, and Bivariate Analysis-:** 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.

OUTPUT-:

- There were more female applicants than male applicants.
- Majority of the applicants had family type unaccompanied and the least were a group of people.

INSIGHTS

- Most applicants were working professionals and maternity leave was the least profession among the applicants.
- Majority of the applicants had 0-5 years of employment period and the least was 45-50 years.
- Cash loans had more applicants than revolving loans.
- Applicants aged 56 and above were the highest.
- Highest Income of most of the applicants was between 125K-150K whereas the lowest income was 500K-525K.
- Female defaulters were more than male defaulters.

INSIGHTS

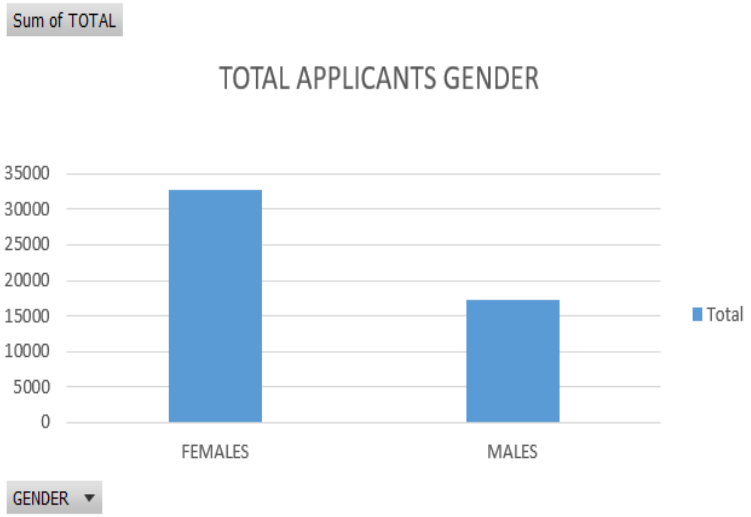
- Cash loans were more defaulted than revolving loans.
- Applicants with 0-5 years employment period defaulted their payments.
- Applicants who were working professionals defaulted the most no. of loans whereas applicants who were businessman, student and maternity leave didn’t defaulted their payments.
- Applicants who had unaccompanied family type defaulted their loans the most whereas applicants who were a group of people didn’t defaulted their payments.

INSIGHTS

- Applicants aged 20-25 years defaulted their loans the most whereas applicants aged 56 and above defaulted their loans the least.
- Majority of the defaulters had an income between 125K-150K.

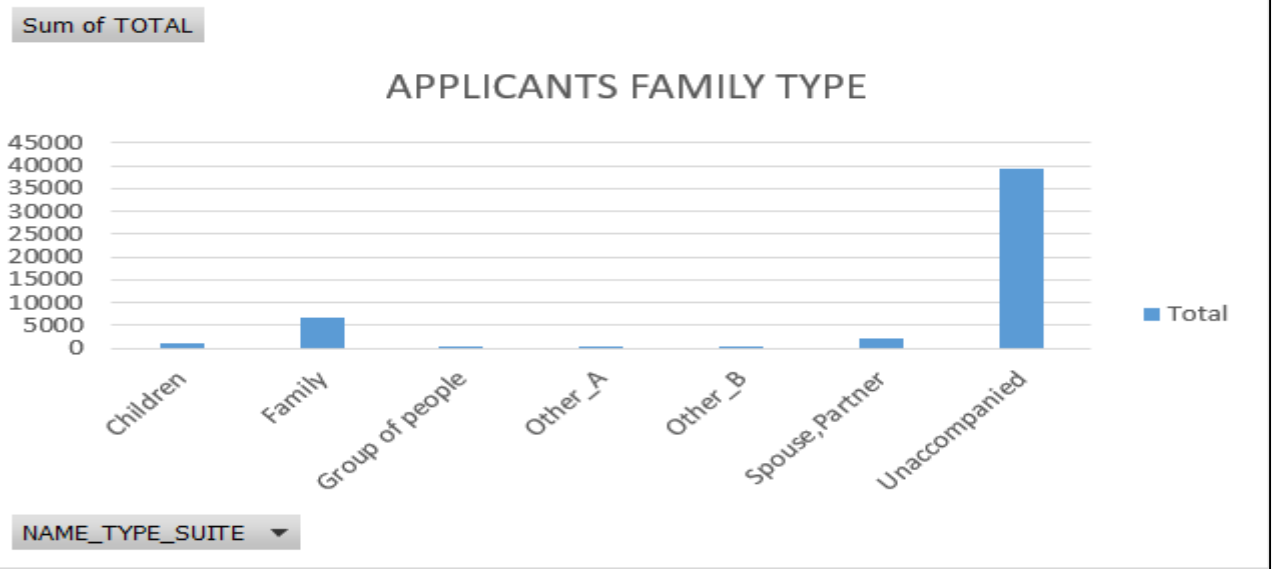
UNIVARIATE ANALYSIS

Row Labels	Sum of TOTAL
FEMALES	32823
MALES	17174
Grand Total	**49997**



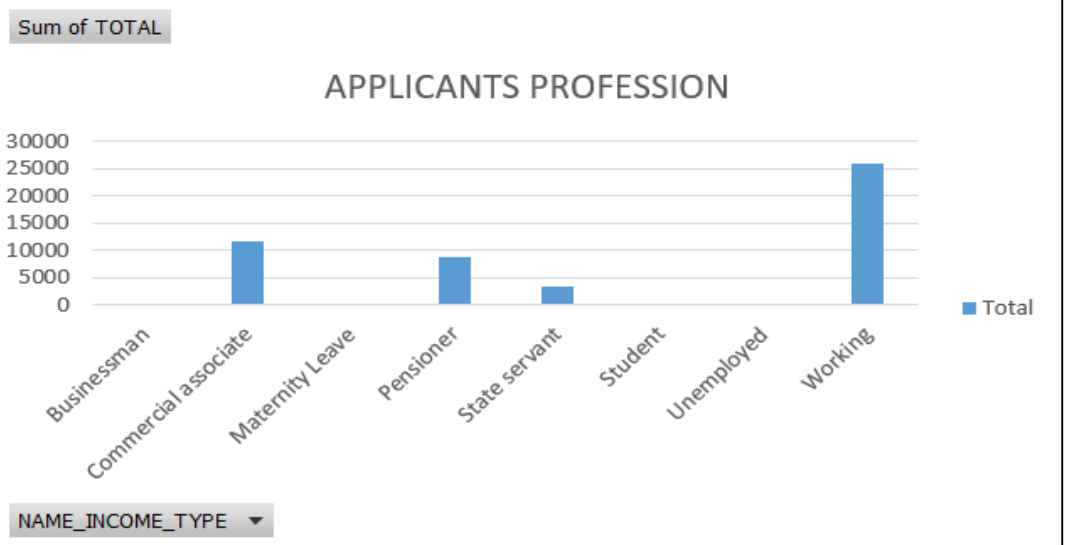
UNIVARIATE ANALYSIS

Row Labels	Sum of TOTAL
Children	993
Family	6581
Group of peopl	76
Other_A	262
Other_B	551
Spouse,Partner	2098
Unaccompanie	39438
Grand Total	**49999**



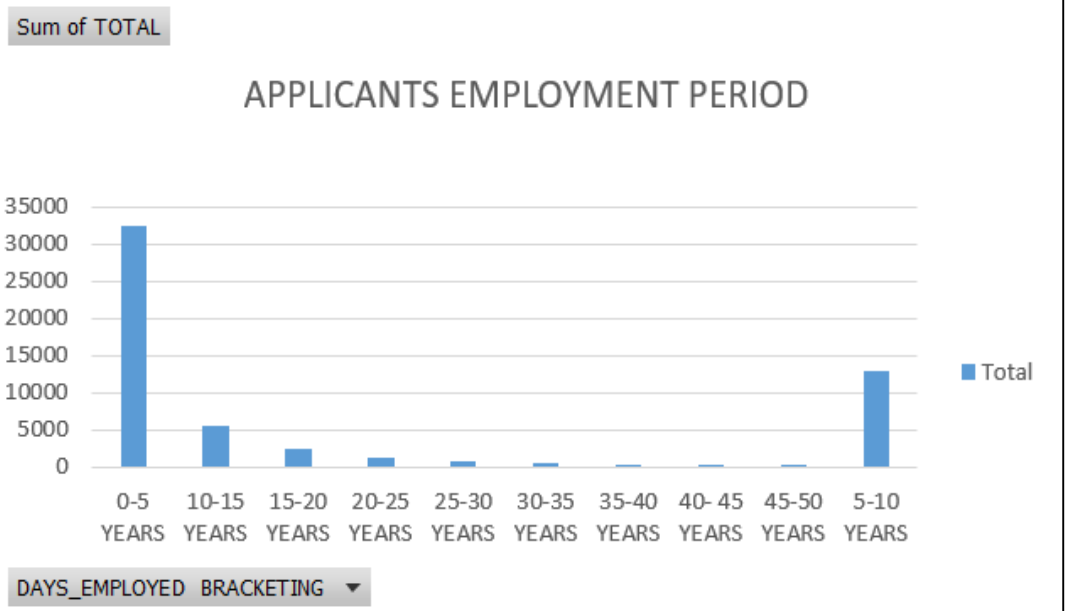
UNIVARIATE ANALYSIS

Row Labels	Sum of TOTAL
Businessman	2
Commercial as	11543
Maternity Leav	1
Pensioner	8920
State servant	3512
Student	5
Unemployed	6
Working	26010
Grand Total	**49999**



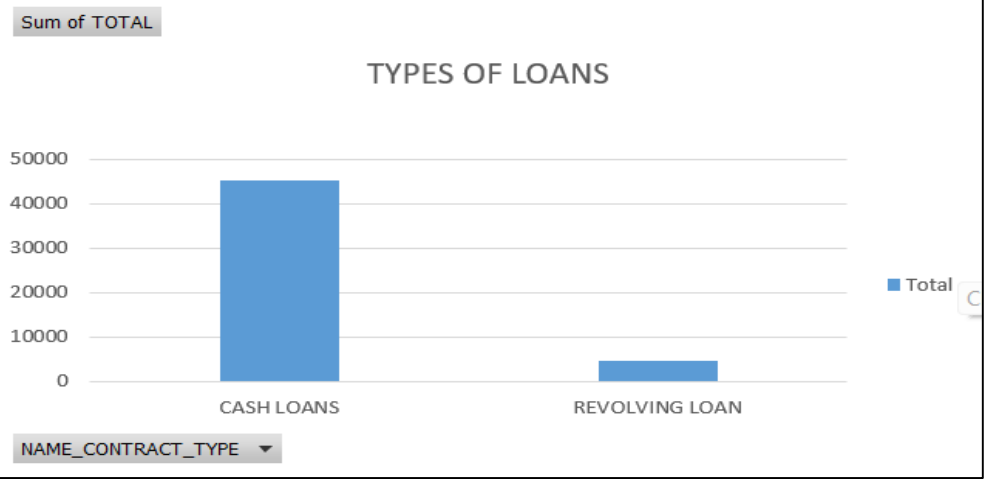
UNIVARIATE ANALYSIS

Row Labels	Sum of TOTAL
0-5 YEARS	32510
10-15 YEARS	5477
15-20 YEARS	2321
20-25 YEARS	1308
25-30 YEARS	656
30-35 YEARS	378
35-40 YEARS	174
40- 45 YEARS	40
45-50 YEARS	4
5-10 YEARS	12940
Grand Total	**55808**



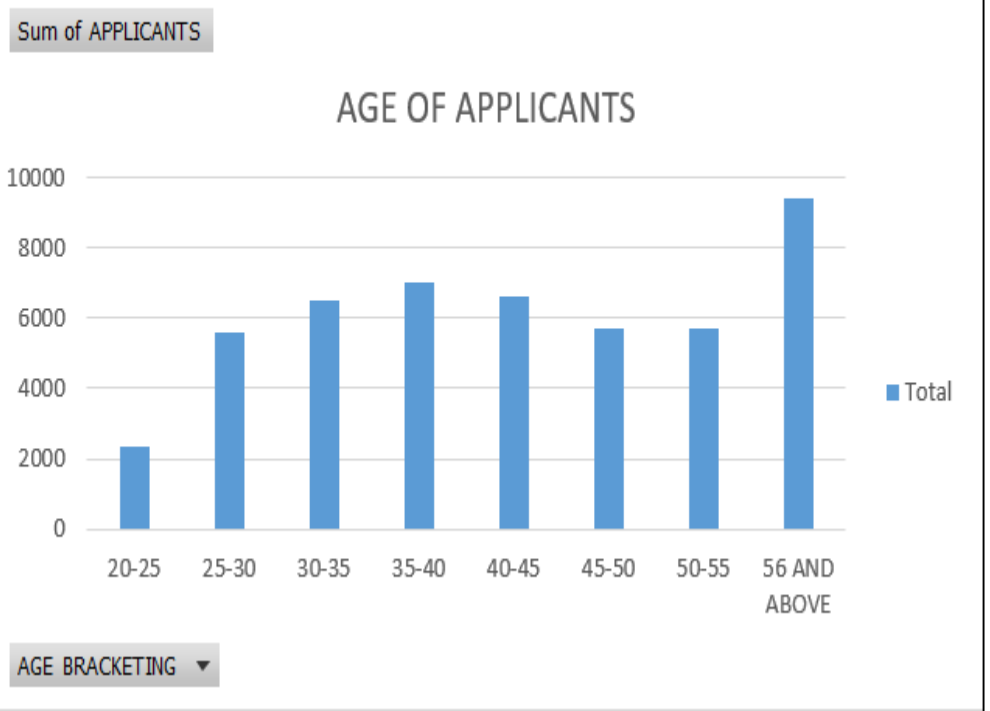
UNIVARIATE ANALYSIS

Row Labels	Sum of TOTAL
CASH LOANS	45276
REVOLVING LO	4723
Grand Total	**49999**



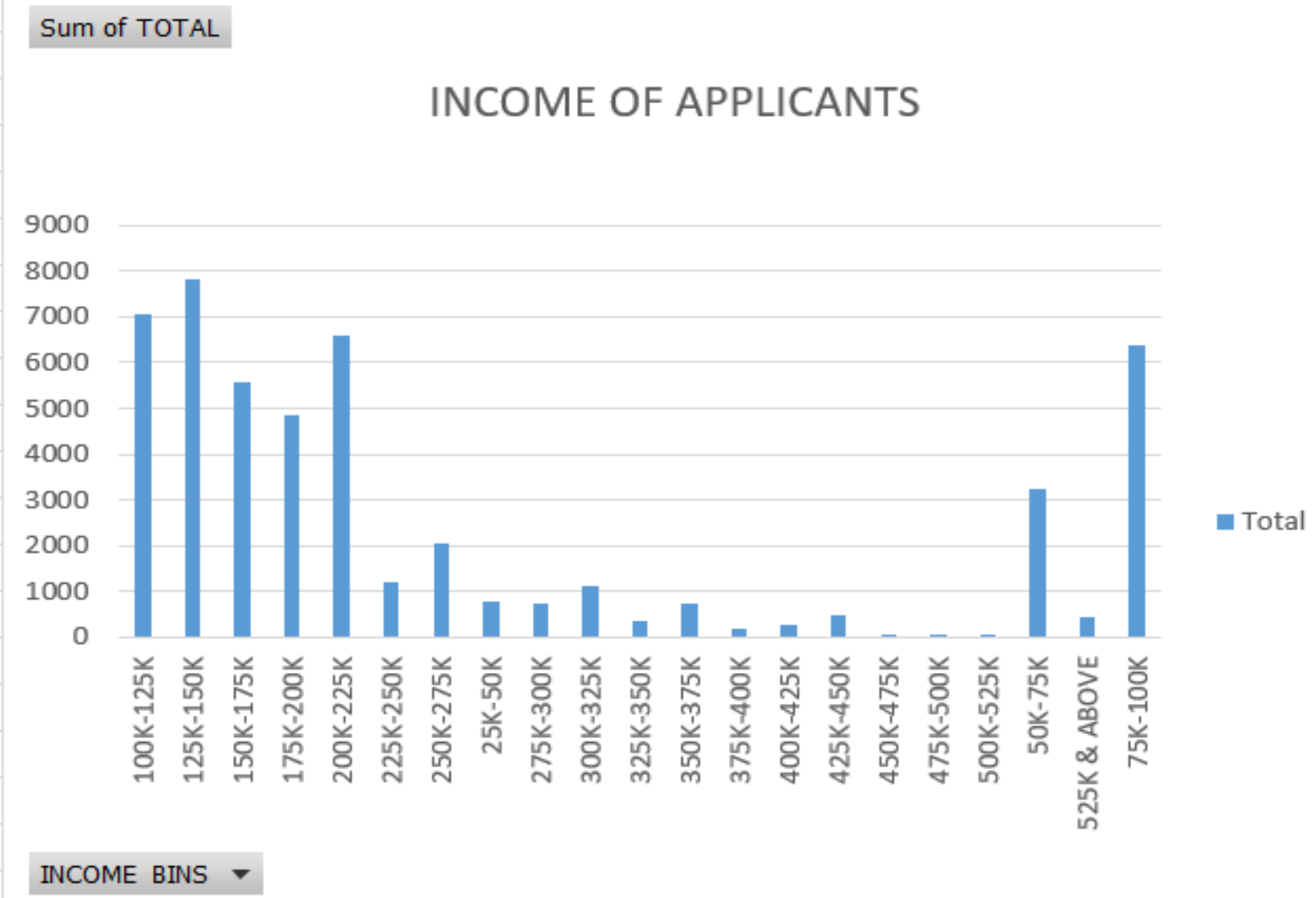
UNIVARIATE ANALYSIS

Row Labels	Sum of APPLICANTS
20-25	2366
25-30	5598
30-35	6502
35-40	7007
40-45	6608
45-50	5723
50-55	5711
56 AND ABOVE	9415
Grand Total	**48930**



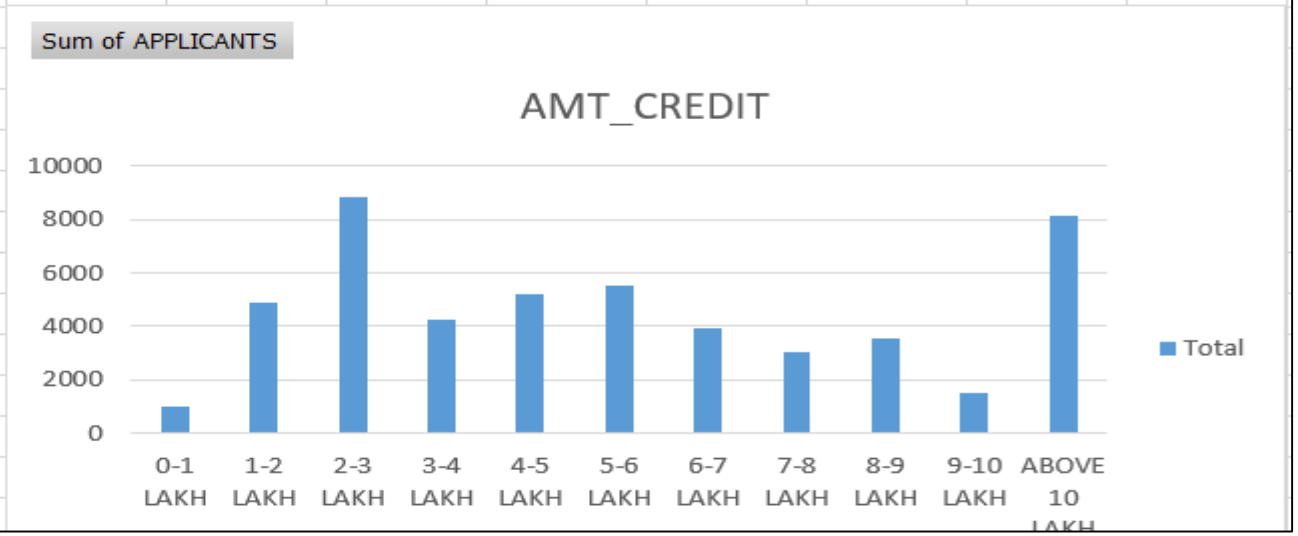
UNIVARIATE ANALYSIS

Row Labels	Sum of TOTAL
100K-125K	7048
125K-150K	7804
150K-175K	5561
175K-200K	4847
200K-225K	6612
225K-250K	1206
250K-275K	2062
25K-50K	804
275K-300K	726
300K-325K	1135
325K-350K	346
350K-375K	757
375K-400K	200
400K-425K	289
425K-450K	492
450K-475K	21
475K-500K	47
500K-525K	9
50K-75K	3226
525K & ABOVE	445
75K-100K	6362
Grand Total	**49999**



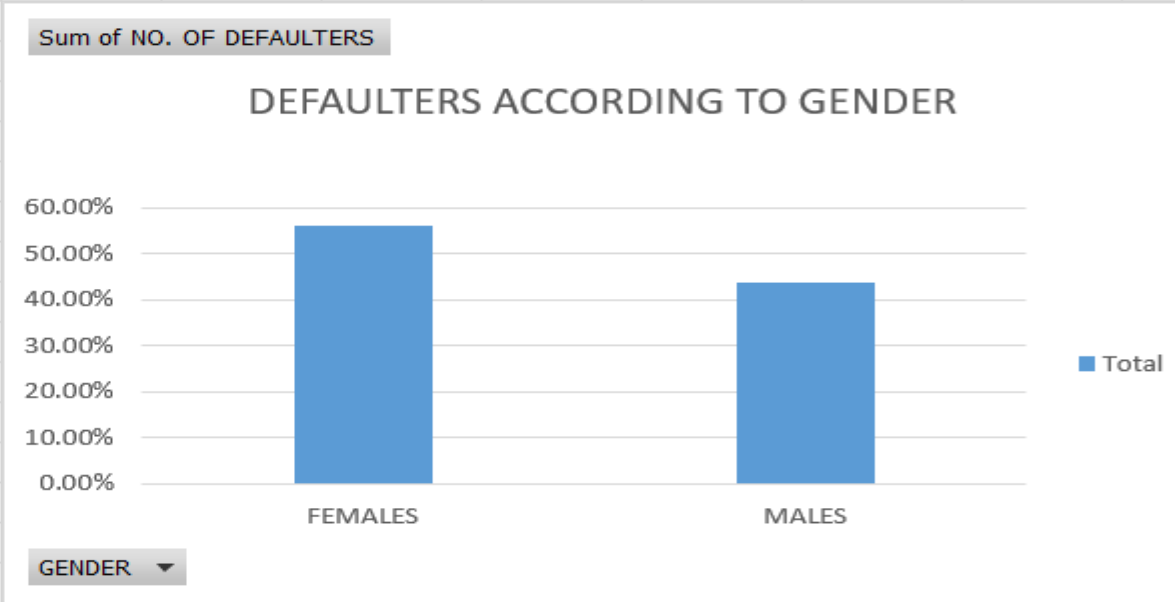
UNIVARIATE ANALYSIS

Row Labels	Sum of APPLICANTS
0-1 LAKH	989
1-2 LAKH	4911
2-3 LAKH	8849
3-4 LAKH	4256
4-5 LAKH	5228
5-6 LAKH	5554
6-7 LAKH	3909
7-8 LAKH	3062
8-9 LAKH	3571
9-10 LAKH	1524
ABOVE 10 LAKH	8146
Grand Total	**49999**



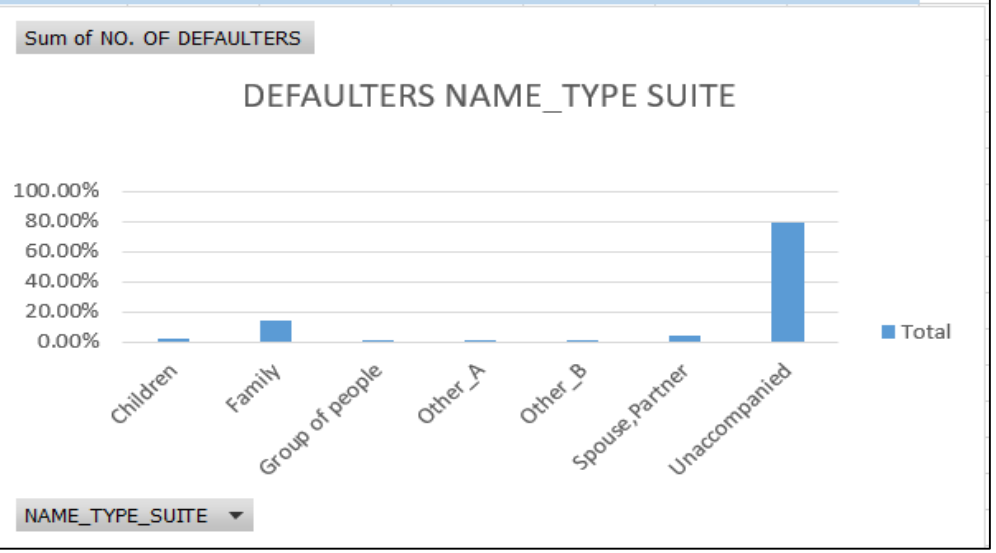
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of NO. OF DEFAULTERS
FEMALES	56.23%
MALES	43.77%
Grand Total	**100.00%**



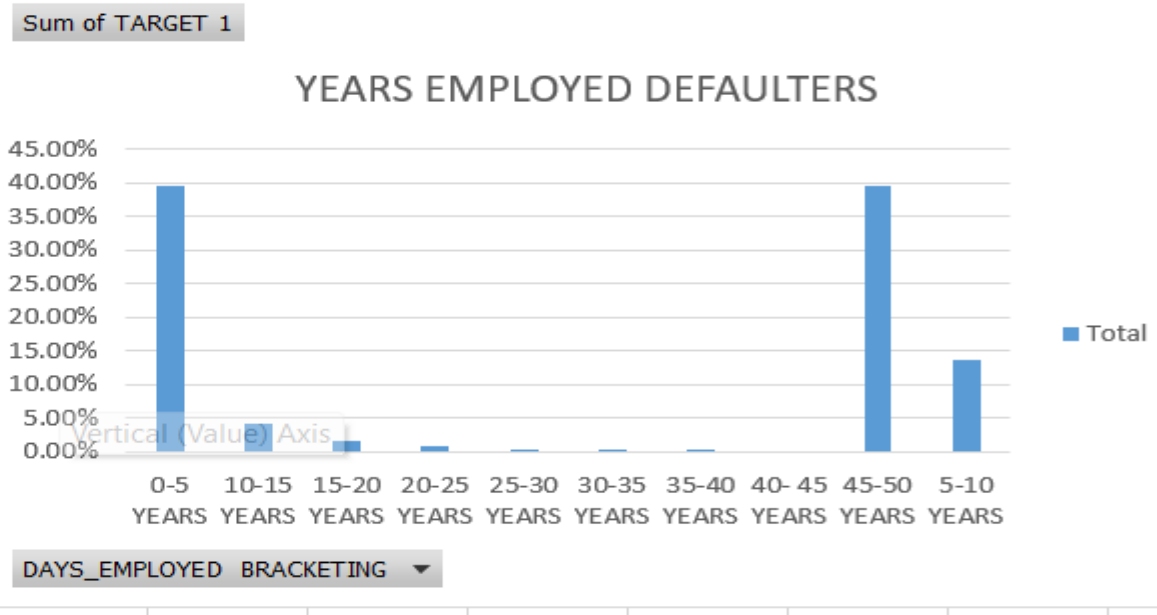
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of NO. OF DEFAULTERS
Children	1.91%
Family	13.76%
Group of peopl	0.07%
Other_A	0.47%
Other_B	1.07%
Spouse,Partner	3.78%
Unaccompanie	78.94%
Grand Total	**100.00%**



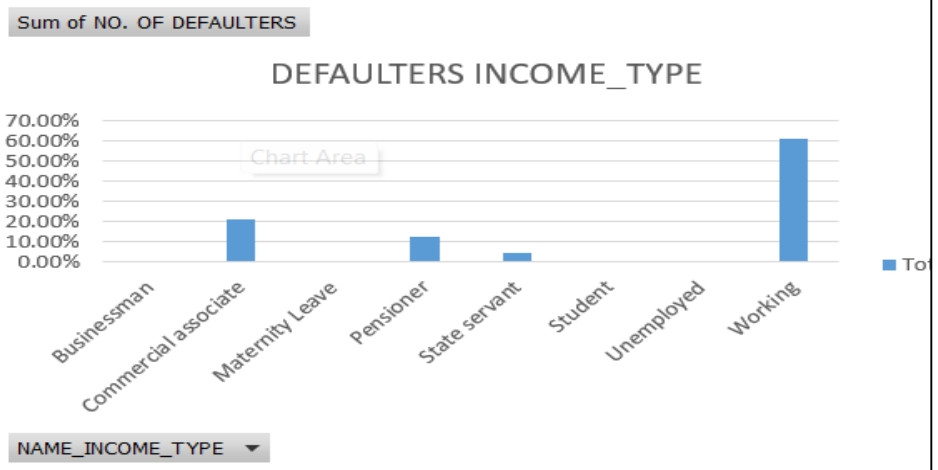
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of TARGET 1
0-5 YEARS	39.64%
10-15 YEARS	4.19%
15-20 YEARS	1.57%
20-25 YEARS	0.88%
25-30 YEARS	0.34%
30-35 YEARS	0.19%
35-40 YEARS	0.05%
40- 45 YEARS	0.00%
45-50 YEARS	39.64%
5-10 YEARS	13.51%
Grand Total	**100.00%**



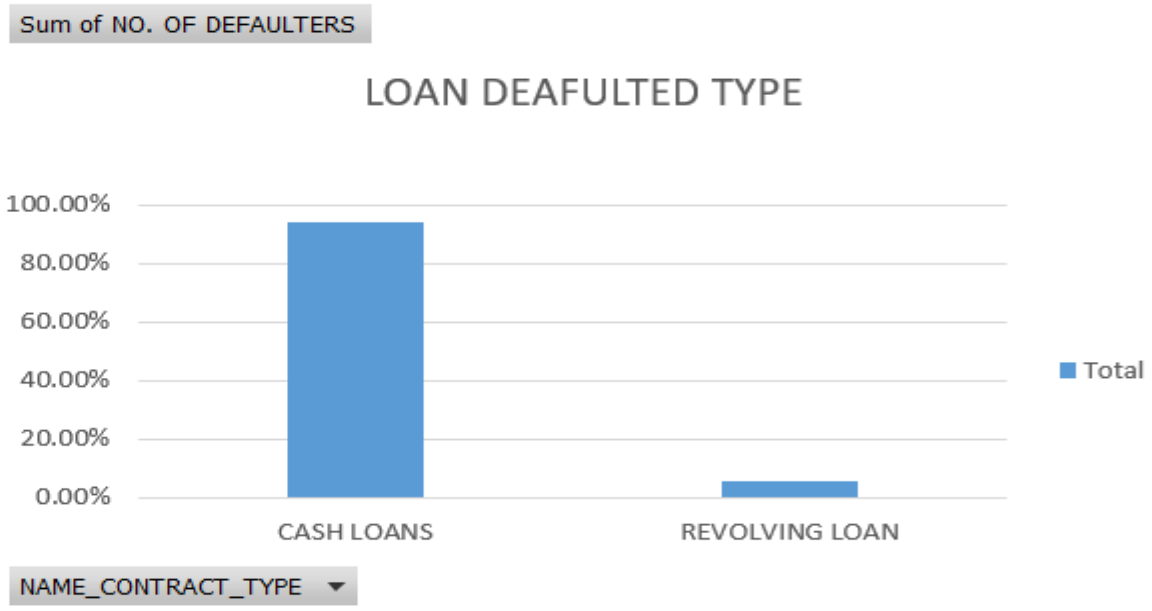
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of NO. OF DEFAULTERS
Businessman	0.00%
Commercial as	21.46%
Maternity Leav	0.00%
Pensioner	12.44%
State servant	4.92%
Student	0.00%
Unemployed	0.05%
Working	61.13%
Grand Total	**100.00%**



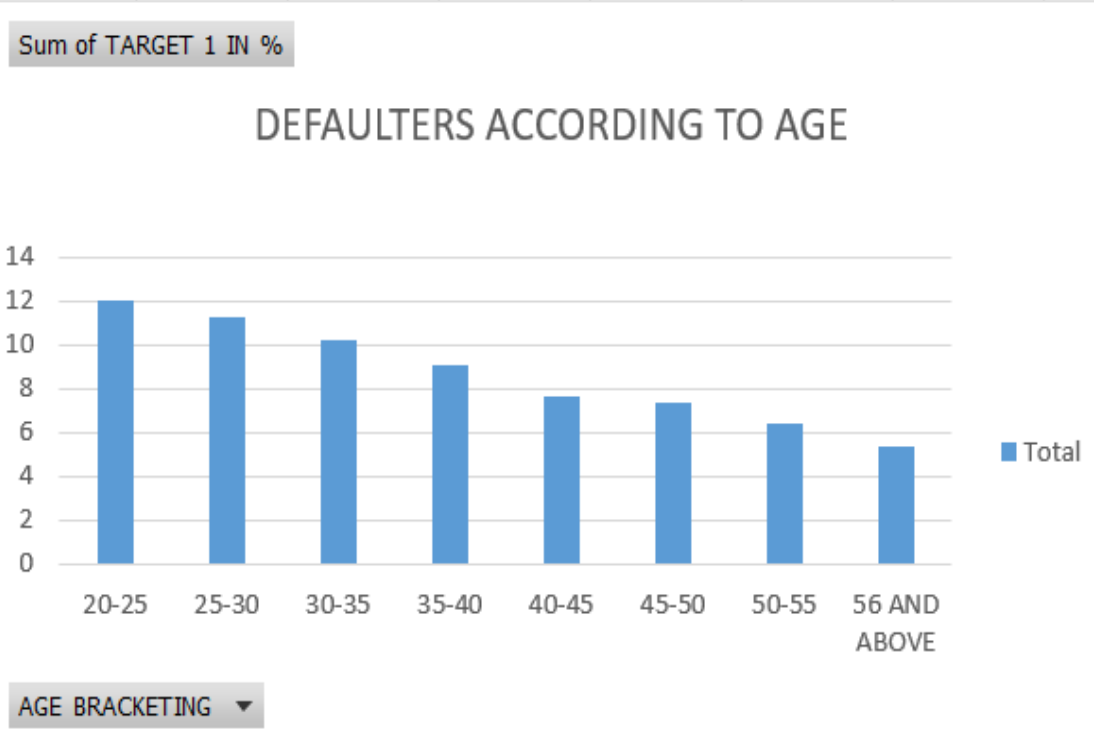
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of NO. OF DEFAULTERS
CASH LOANS	94.19%
REVOLVING LO	5.81%
Grand Total	**100.00%**



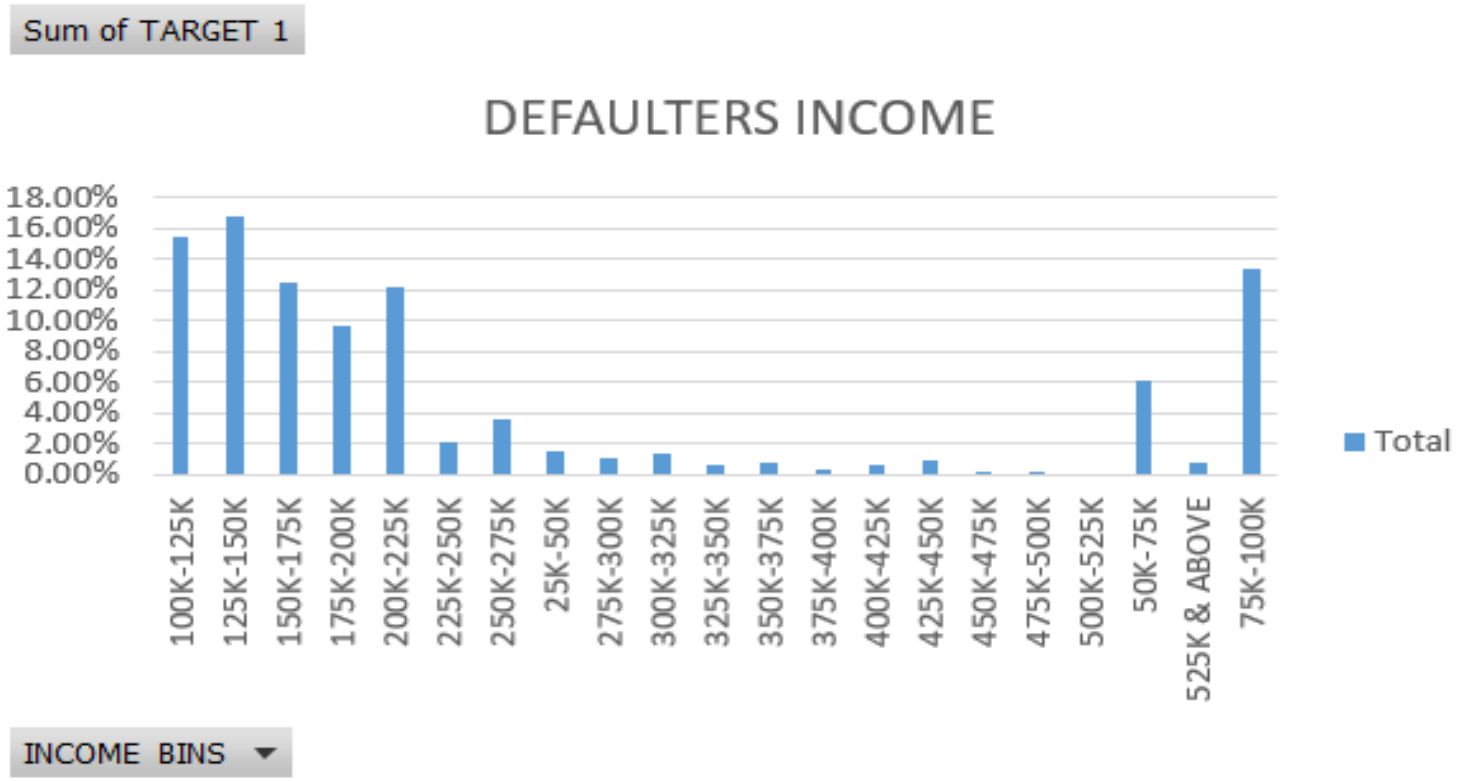
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of TARGET 1 IN %
20-25	12.00338123
25-30	11.24241456
30-35	10.25378108
35-40	9.074366129
40-45	7.65478424
45-50	7.393447376
50-55	6.417190469
56 AND ABOVE	5.303319344
Grand Total	**69.34268444**



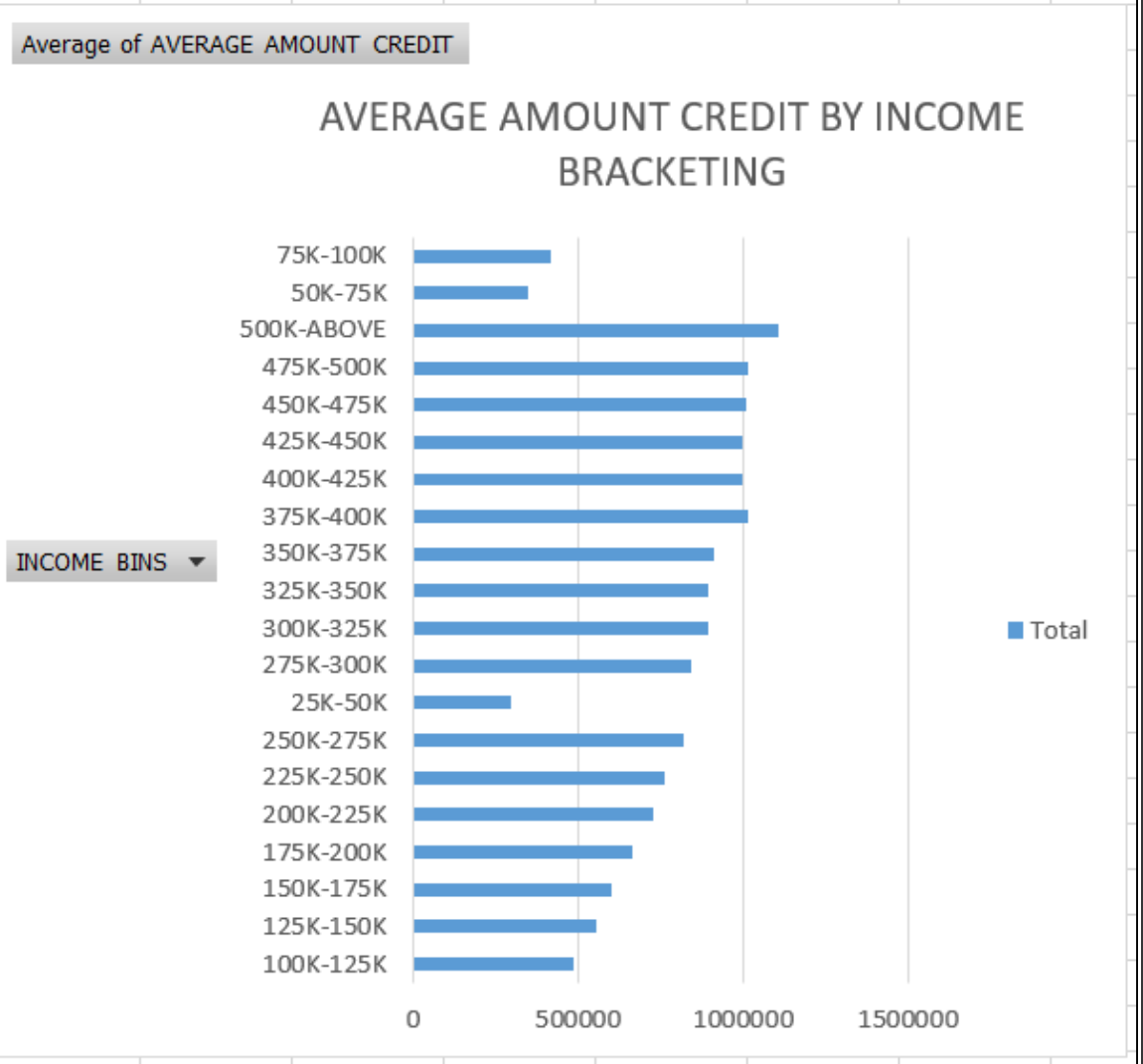
SEGMENTED UNIVARIATE ANALYSIS

Row Labels	Sum of TARGET 1
100K-125K	15.40%
125K-150K	16.84%
150K-175K	12.44%
175K-200K	9.66%
200K-225K	12.20%
225K-250K	2.11%
250K-275K	3.55%
25K-50K	1.56%
275K-300K	1.12%
300K-325K	1.47%
325K-350K	0.60%
350K-375K	0.84%
375K-400K	0.35%
400K-425K	0.65%
425K-450K	0.89%
450K-475K	0.05%
475K-500K	0.07%
500K-525K	0.00%
50K-75K	6.11%
525K & ABOVE	0.77%
75K-100K	13.31%
Grand Total	**100.00%**



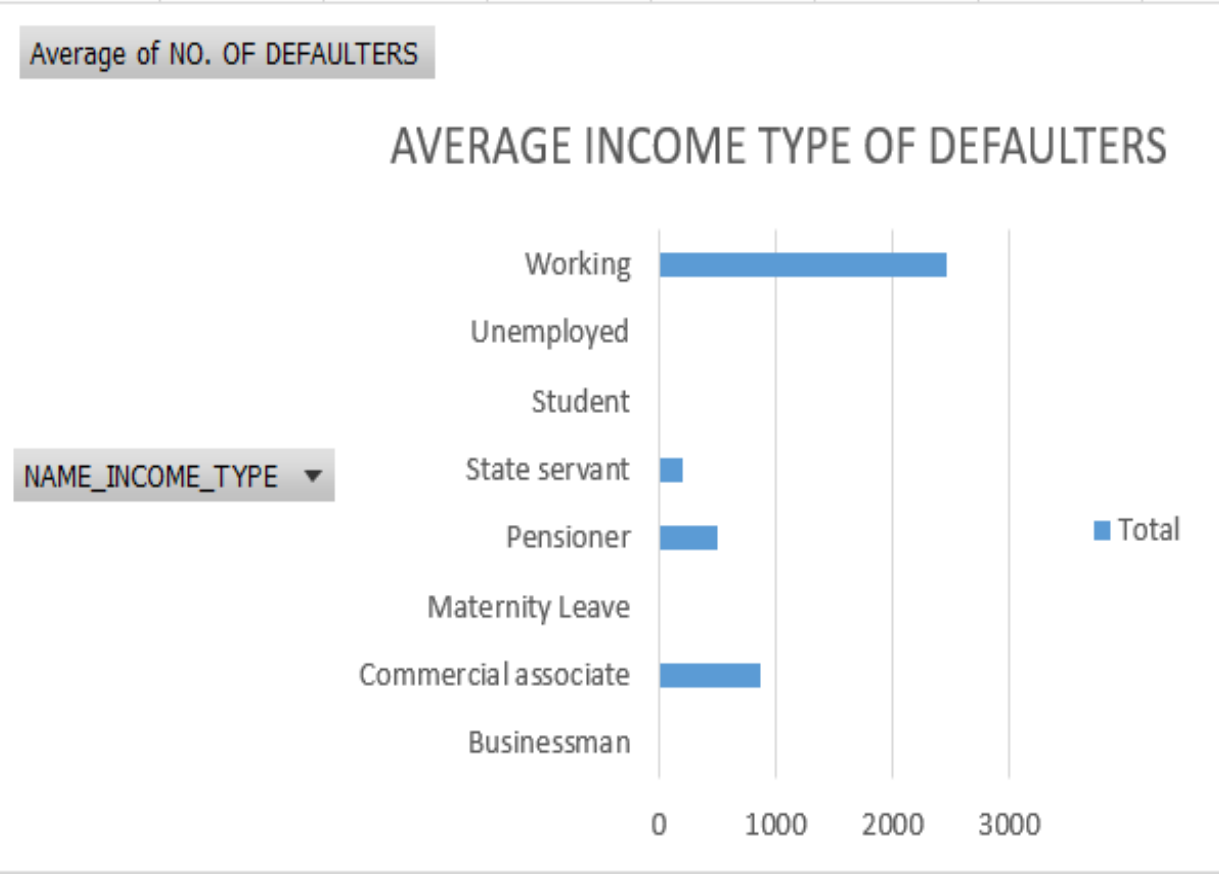
BIVARIATE ANALYSIS

Row Labels	Average of AVERAGE AMOUNT CREDIT
100K-125K	483568.8073
125K-150K	553042.1642
150K-175K	602034.4016
175K-200K	667004.421
200K-225K	727198.4449
225K-250K	759541.3782
250K-275K	820255.3451
25K-50K	297752.0765
275K-300K	842725.6488
300K-325K	892300.0718
325K-350K	892332.6503
350K-375K	910363.0482
375K-400K	1016814.375
400K-425K	999208.199
425K-450K	999153.6402
450K-475K	1011521.839
475K-500K	1015150.404
500K-ABOVE	1105365.122
50K-75K	345240.3585
75K-100K	417267.8771
Grand Total	**767892.0136**



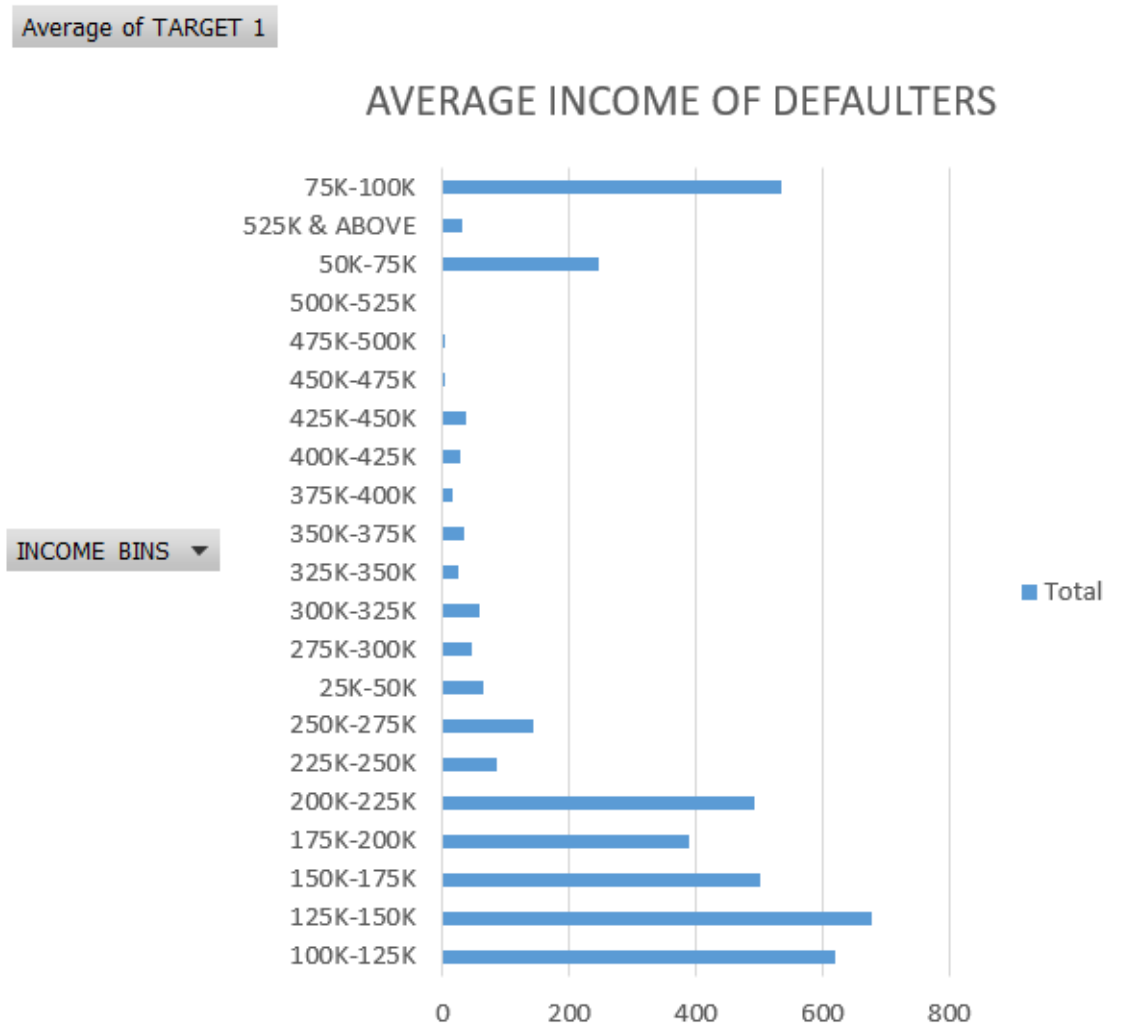
BIVARIATE ANALYSIS

Row Labels	Average of NO. OF DEFAULTERS
Businessman	0
Commercial as	864
Maternity Leav	0
Pensioner	501
State servant	198
Student	0
Unemployed	2
Working	2461
Grand Total	**503.25**



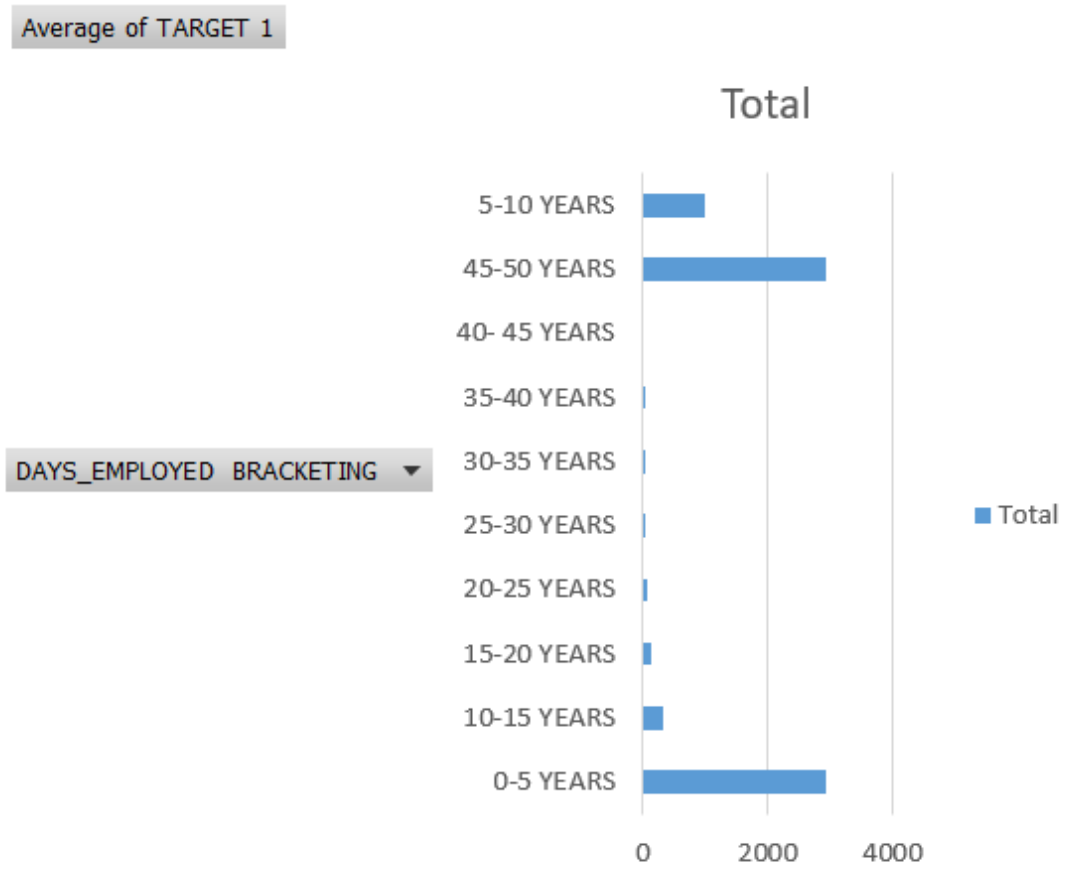
BIVARIATE ANALYSIS

Row Labels	Average of TARGET 1
100K-125K	620
125K-150K	678
150K-175K	501
175K-200K	389
200K-225K	491
225K-250K	85
250K-275K	143
25K-50K	63
275K-300K	45
300K-325K	59
325K-350K	24
350K-375K	34
375K-400K	14
400K-425K	26
425K-450K	36
450K-475K	2
475K-500K	3
500K-525K	0
50K-75K	246
525K & ABOVE	31
75K-100K	536
Grand Total	191.7142857



BIVARIATE ANALYSIS

Row Labels	Average of TARGET 1
0-5 YEARS	2934
10-15 YEARS	310
15-20 YEARS	116
20-25 YEARS	65
25-30 YEARS	25
30-35 YEARS	14
35-40 YEARS	4
40- 45 YEARS	0
45-50 YEARS	2934
5-10 YEARS	1000
Grand Total	**740.2**



INSIGHTS

5. **Identify Top Correlations for Different Scenarios:** 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.

CORRELATION FOR APPLICANTS WITH LOAN PAYMENTS MADE ON TIME

TARGET 0	AMT_CREDIT	CNT_CHILDREN	AMT_INCOME_TOTAL	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_ID_PUBLISH	REGION_RATING_CLIENT	AMT_REQ_CREDIT_BUREAU_YEAR	AMT_GO
AMT_CREDIT	1	0.004304853	-0.010141067	0.003712439	0.00578188	0.000680776	-0.00198283	-0.001172158	0.004146655	0
CNT_CHILDREN	0.004304853	1	0.036319722	-0.024912809	0.33587627	-0.243591518	-0.032537221	0.021288992	-0.035734888	0
AMT_INCOME_TOTAL	-0.01014107	0.036319722	1	0.181941261	0.07376942	-0.162702675	0.032286356	-0.205031899	0.031323516	-0
REGION_POPULATION_RELATIVE	0.003712439	-0.024912809	0.181941261	1	-0.0304354	-0.006610653	-0.002236288	-0.539333113	0.004652396	0
DAYS_BIRTH	0.005781883	0.335876269	0.073769425	-0.030435419	1	-0.615289978	0.270073313	0.00902485	-0.070267716	
DAYS_EMPLOYED	0.000680776	-0.243591518	-0.162702675	-0.006610653	-0.61529	1	-0.27222439	0.040505636	0.044183816	0
DAYS_ID_PUBLISH	-0.00198283	-0.032537221	0.032286356	-0.002236288	0.27007331	-0.27222439	1	-0.008097427	-0.044692876	-0
REGION_RATING_CLIENT	-0.00117216	0.021288992	-0.205031899	-0.539333113	0.00902485	0.040505636	-0.008097427	1	0.007004146	-0
AMT_REQ_CREDIT_BUREAU_YEAR	0.004146655	-0.035734888	0.031323516	0.004652396	0.00236758	0.044183816	-0.044692876	0.007004146	1	0
AMT_GOODS_PRICE	0.942500566	0.004978184	-0.007641498	0.004905201	-0.0702677	0.000508414	-0.002114712	-0.001540809	0.001485415	

CORRELATION FOR APPLICATION WITH LOAN PAYMENT DIFFICULTIES

TARGET 1	AMT_CREDIT	CNT_CHILDREN	AMT_INCOME_TOTAL	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_ID_PUBLISH	REGION_RATING_CLIENT	AMT_REQ_CREDIT_BUREAU_YEAR	AMT_GO
AMT_CREDIT	1	0.002802954	-0.003412694	0.002351016	0.00440945	1.18175E-06	-0.00281434	-0.002282137	0.00626982	-0
CNT_CHILDREN	0.002802954	1	0.009588558	-0.025555665	0.32926375	-0.239693041	-0.032115773	0.025913889	-0.034840358	0
AMT_INCOME_TOTAL	-0.00341269	0.009588558	1	0.029841469	0.01600277	-0.031615555	0.003506646	-0.038188511	0.004080472	-0
REGION_POPULATION_RELATIVE	0.002351016	-0.025555665	0.029841469	1	-0.0325137	-0.004101686	-0.004345136	-0.532667302	0.00538252	0
DAYS_BIRTH	0.004409453	0.329263754	0.016002774	-0.032513748	1	-0.613553972	0.270825141	0.016779196	-0.0704042	0
DAYS_EMPLOYED	1.18175E-06	-0.239693041	-0.031615555	-0.004101686	-0.613554	1	-0.270382022	0.034321673	0.041520751	
DAYS_ID_PUBLISH	-0.00281434	-0.032115773	0.003506646	-0.004345136	0.27082514	-0.270382022	1	-0.002307011	-0.047016431	
REGION_RATING_CLIENT	-0.00228214	0.025913889	-0.038188511	-0.532667302	0.0167792	0.034321673	-0.002307011	1	0.008559499	-0
AMT_REQ_CREDIT_BUREAU_YEAR	0.00626982	-0.034840358	0.004080472	0.00538252	-0.0704042	0.041520751	-0.047016431	0.008559499	1	-0
AMT_GOODS_PRICE	-0.02195155	0.013145935	-0.003557486	0.022164569	0.02800634	-0.01774067	0.00767777	-0.013930665	-0.027501305	

RESULT

Through this project, I successfully analyzed various factors influencing bank loans using data-driven insights to identify key patterns and trends. By applying statistical atechniques, I gained a deeper understanding of the factors affecting loan eligibility, such as income and credit history.
This project enhanced my ability to interpret financial data and make data-driven decisions. Additionally, it strengthened my analytical skills by allowing me to explore real-world banking scenarios . Overall, this study provided valuable insights into risk assessment in the banking sector, improving my understanding of how financial institutions evaluate loan applications.

File Link-:
https://docs.google.com/spreadsheets/d/1sa4gfGyWH9fCO_rsSSWsNL23WD0Js285/edit?usp=drive_link&ouid=114615499307515143340&rtpof=true&sd=true