BANK LOAN CASE STUDY

FINAL PROJECT-2

Name: J.Susritha

Gmail:susrithaj2102@gmail.com

Agenda

- Project Description
- Approach
- Tech-Stack Used
- Insights
- Results

Project Description.

- (Bank loan case study: Excel project) focusing on risk analytics and Exploratory Data Analysis (EDA). Learn how to clean and analyze extensive loan application datasets, identify outliers, and perform univariate and bivariate analysis using pivot tables and charts. Discover how data-driven insights can be used to reduce the risk of financial losses when lending to consumers in the banking and financial services industry. Helps in Enhancing Excel skills and gain practical knowledge in risk analytics with this immersive project.
- This case study focuses on utilizing Exploratory Data Analysis (EDA) techniques in a real-world business context, specifically in the banking industry. The objective is to demonstrate how EDA can be applied to analyze and mitigate risks associated with lending money to consumers.The project involves working with two extensive datasets: the current loan applications and the previous loan applications. These datasets contained unnecessary columns and missing data, which required initial data cleaning. Once the data was cleaned, the analysis proceeded to identify and handle outliers before conducting univariate and bivariate analysis.

Approach

- This case study has two enormous data sets: the current application and the previous application. Each included several unneeded columns that would be useless for risk assessments, as well as many blank data. I started by cleaning.
- To evaluate this enormous set of data, I first cleaned the data, located some outliers and deleted them, and then began performing univariate and bivariate analysis using pivot tables and charts

Tech-Stack Used

- Microsoft Excel 2010

Insights

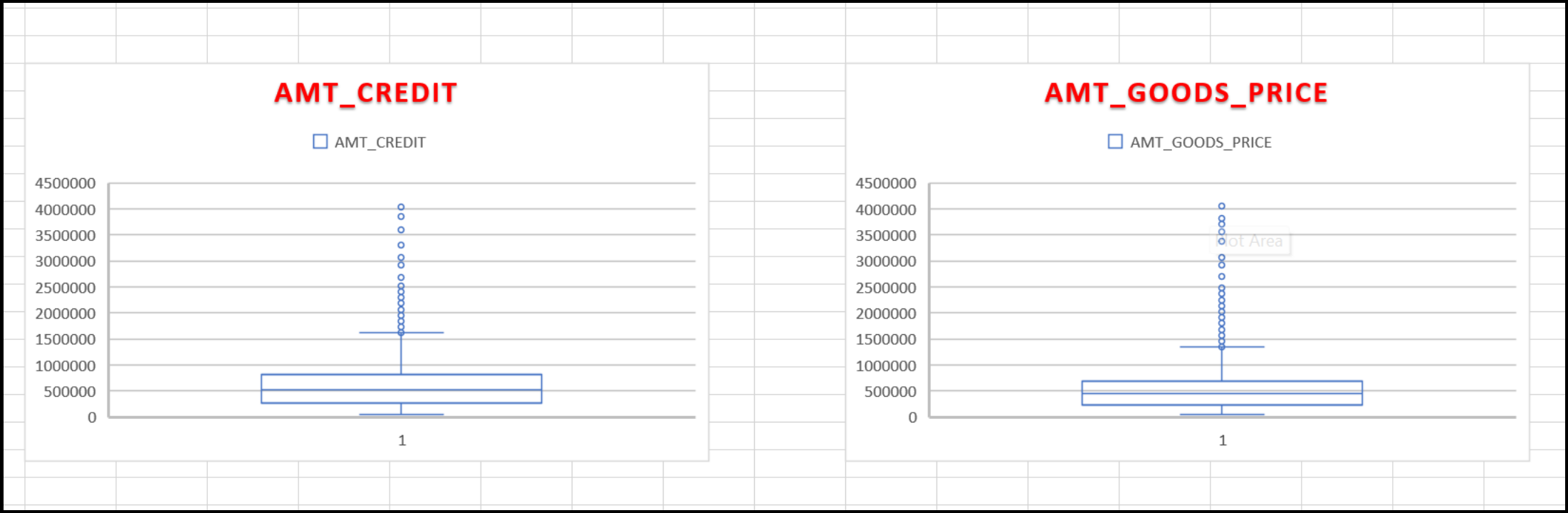
Task 1:Identify the missing data and use appropriate method to deal with it.

Steps to clean the data:

- Identify and handle missing data: Identify any missing values in your dataset and decide how to handle them. You can either delete rows or columns with missing data, fill in the missing values with averages or other suitable values, or use advanced techniques like imputation to estimate missing values.
- Remove duplicate entries: Check for duplicate rows or records in your dataset and remove them to ensure data integrity.

B. Identify Outliers in the Dataset:

Create box plots or scatter plots to visualize the distribution of numerical variables and highlight the outliers.



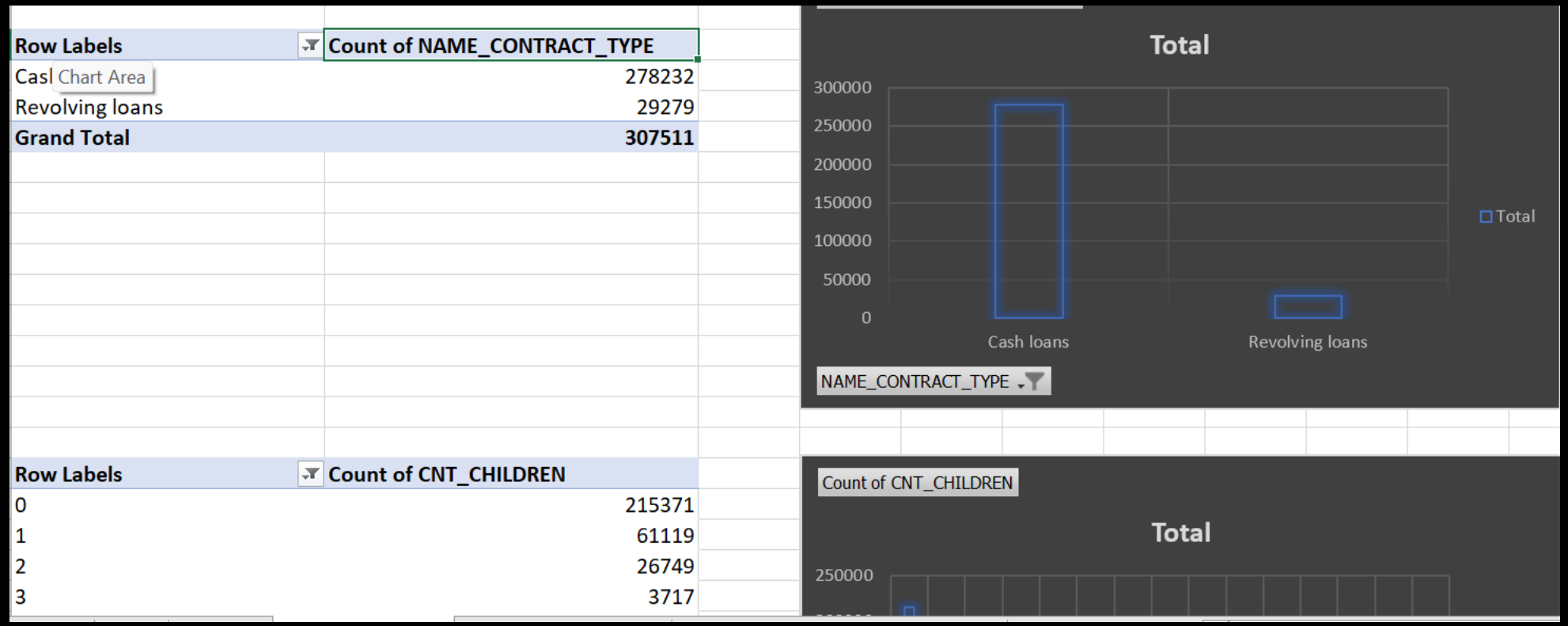
Results are in attached workbook

- Outliers can only be identified on Numeric variables. To identify outliers in Excel, follow these simple steps:

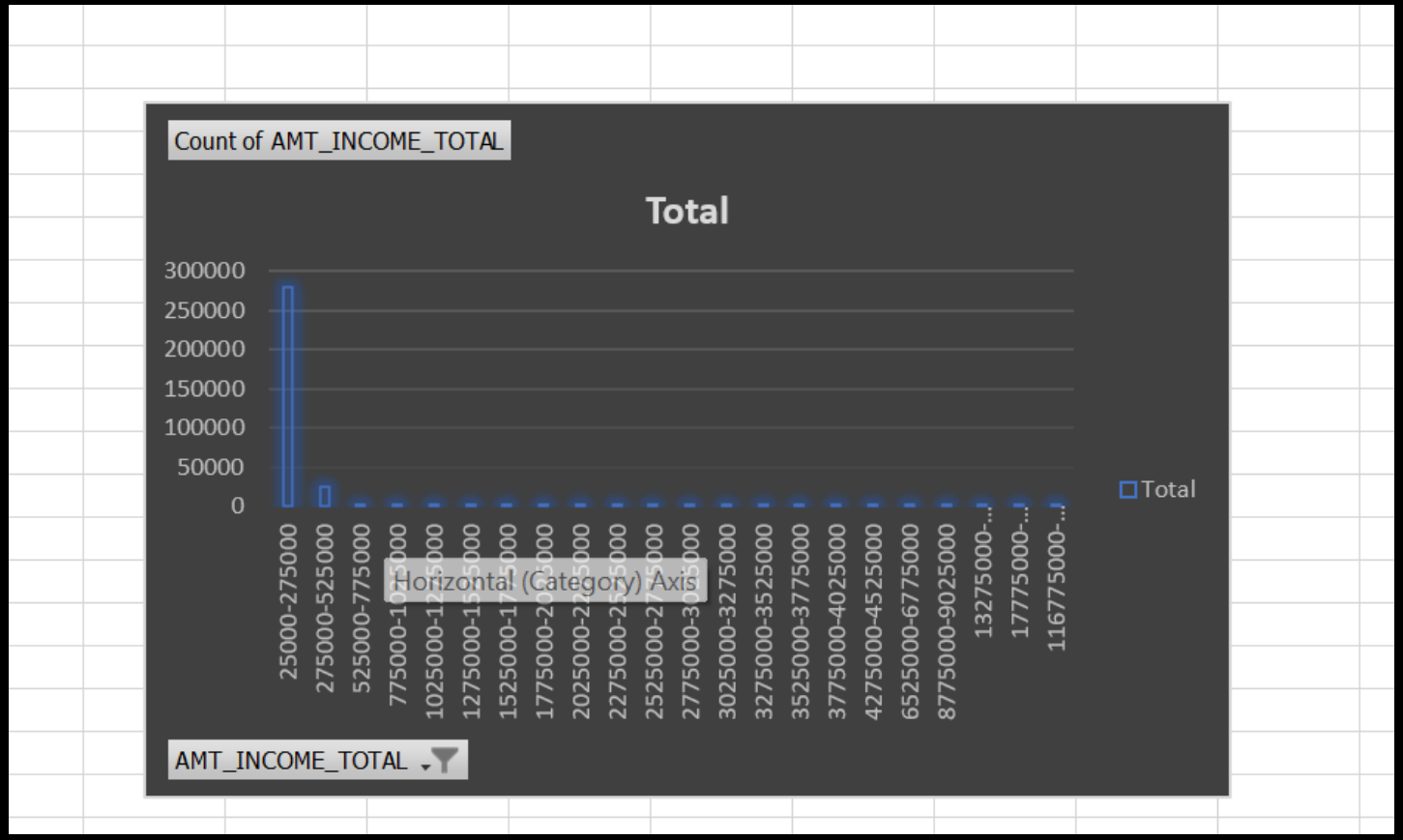
- Calculate the interquartile range (IQR) by subtracting the first quartile (25th percentile) from the third quartile (75th percentile) using the formula: “=QUARTILE.INC(range,3) – QUARTILE.INC(range,1)”.
- Determine the lower bound for outliers by subtracting 1.5 times the IQR from the first quartile using the formula: “=QUARTILE.INC(range,1) – (1.5 * IQR)”.
- Determine the upper bound for outliers by adding 1.5 times the IQR to the third quartile using the formula: “=QUARTILE.INC(range,3) + (1.5 * IQR)”.

C.Analyze Data Imbalance:

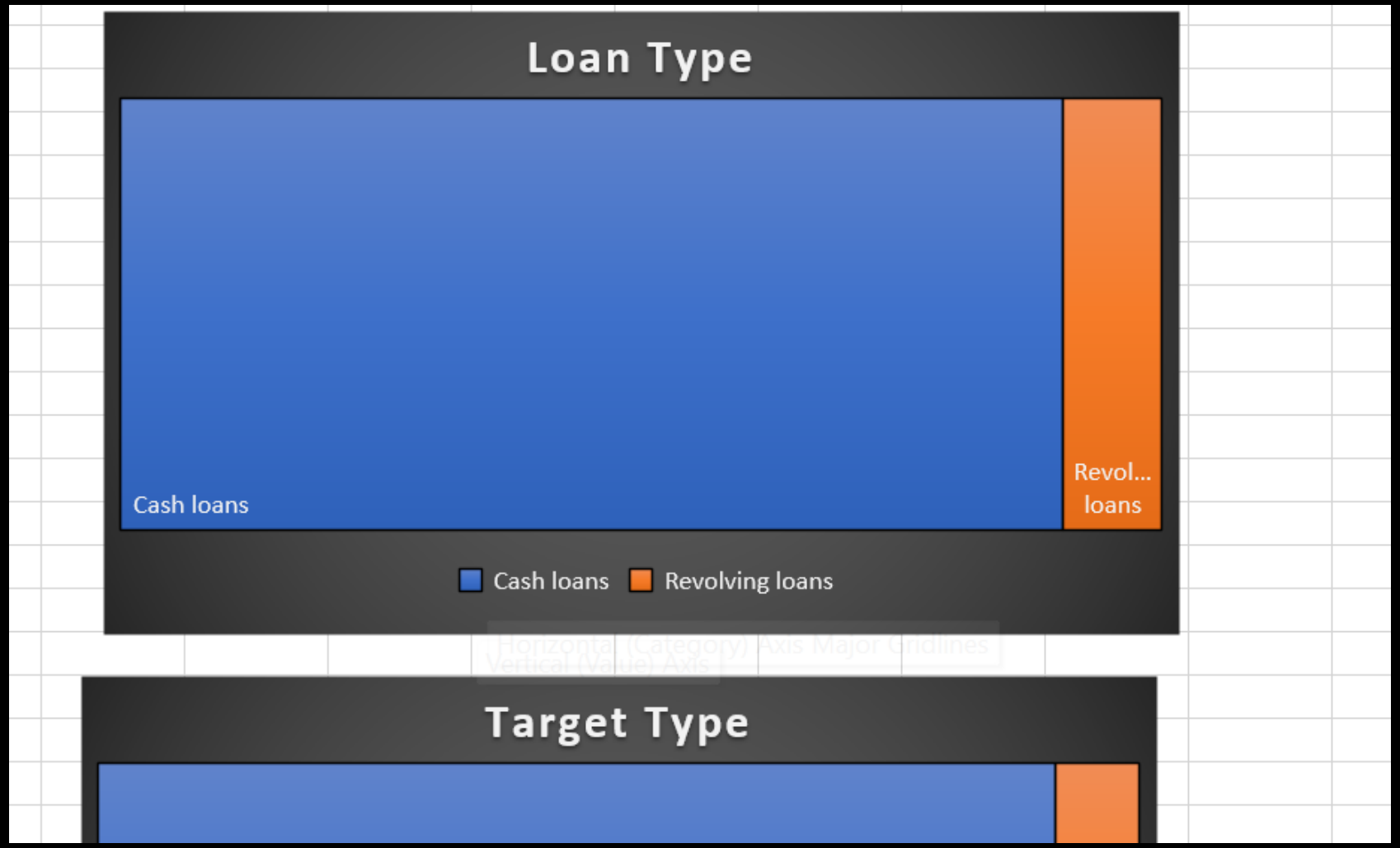
Create a pie chart or bar chart to visualize the distribution of the target variable and highlight the class imbalance.



D. Perform Univariate, Segmented Univariate, and Bivariate Analysis: Create histograms, bar charts, or box plots to visualize the distributions of variables. Create stacked bar charts or grouped bar charts to compare variable distributions across different scenarios. Create scatter plots or heatmaps to visualize the relationships between variables and the target variable.

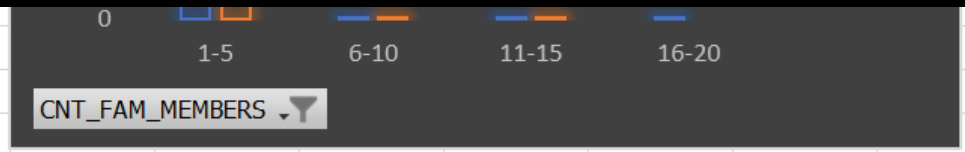


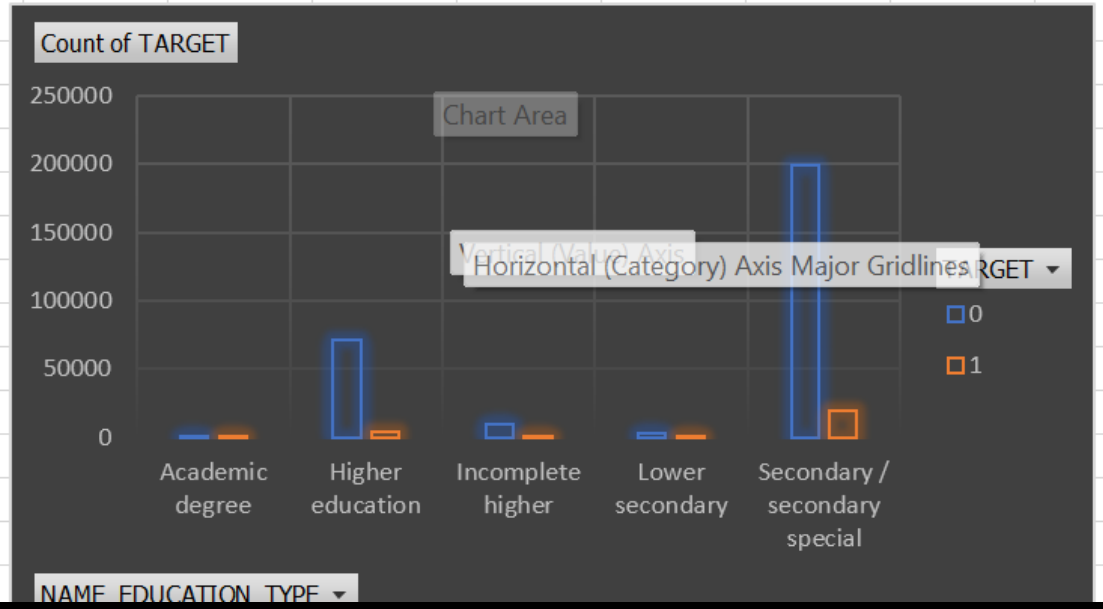
- Segmented univariate



Bivariate

Count of TARGET	Column Labels		
Row Labels	0	1	Grand Total
Academic degree	161	3	164
Higher education	70854	4009	74863
Incomplete higher	9405	872	10277
Lower secondary	3399	417	3816
Secondary / secon	198867	19524	218391
Grand Total	**282686**	**24825**	**307511**





E. Identify Top Correlations for Different Scenarios

Create correlation matrices or heatmaps to visualize the correlations between variables within each segment. Highlight the top correlated variables for each scenario using different colors or shading.

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	N_POPULATION_REL	DAYS_BIRTH	YEAR_BIRTH	
CNT_CHILDREN	1								
AMT_INCOME_T	0.004795787	1							
AMT_CREDIT	-0.001674961	0.038131435	1						
AMT_ANNUITY	0.031257119	0.046421057	0.752194735	1					
AMT_GOODS_PR	-0.008111699	0.037583082	0.983102519	0.752699196	1				
REGION_POPULA	-0.0319749	0.009134586	0.069161087	0.07169025	0.07604893	1			
DAYS_BIRTH	-0.259108666	-0.003096245	0.135316369	0.014303316	0.135810334	0.048190366	1		
YEAR_BIRTH	-0.258910307	-0.002872272	0.135318079	0.014249417	0.135743778	0.048294163	0.999679675	1	
DAYS_EMPLOYED	-0.192863828	-0.014977396	0.001930183	-0.08120712	0.006641788	0.015531849	0.582185148	0.581768792	
DAYS_REGISTRAT	-0.149153857	-0.000157999	0.025854317	-0.034279023	0.025678921	0.056222028	0.289114025	0.288807109	
DAYS_ID_PUBLISI	0.032298597	0.004214856	0.05232898	0.016767235	0.056085697	0.015536882	0.252862836	0.253036359	
FLAG_EMP_PHON	0.192468941	0.014971299	-0.000659899	0.081797315	-0.005269911	-0.015414669	-0.578519352	-0.578101479	
FLAG_WORK_PH(	0.051712264	-0.012145664	-0.059600889	-0.049844588	-0.033617477	-0.017958367	-0.155496277	-0.155447059	
FLAG_CONT_MO	0.003592443	-0.000878255	0.031770843	0.035655654	0.028656753	-0.004029068	0.011826191	0.01170198	
FLAG_PHONE	-0.024305772	-0.004478019	0.008608371	-0.013767964	0.02767849	0.071839278	0.026257308	0.026041178	
FLAG_EMAIL	0.01390223	0.009889203	0.010156814	0.078188198	0.013697265	0.028913793	-0.06254318	-0.062575111	
CNT_FAM_MEME	0.885483713	0.006653677	0.05122364	0.075711476	0.04738797	-0.030162746	-0.203267038	-0.203013983	
REGION_RATING	0.040680482	-0.021486257	-0.059192754	-0.073783735	-0.06638988	-0.443235509	-0.033927932	-0.034094362	
REGION_RATING	0.043185374	-0.022807978	-0.071377103	-0.089290515	-0.077190843	-0.446976802	-0.033631154	-0.033801742	

Results:

- **The project helped in strengthening ability to use eda in excel..**
- **The project helped in perfecting my data cleaning and data visualisation methodologies.**
- **<u>Main decisive factors:</u>**

→NAME_EDUCATION_TYPE

→NAME_INCOME_TYPE

→REGION_RATING_CLIENT

→ORGANIZATION_TYPE.

→DAYS_BIRTH

→DAYS_EMPLOYED.

→AMT_INCOME_TOTAL

→NAME_CASH_LOAN_PURPOSE: CNT_CHILDREN

CLICK HERE FOR EXCEL WORKING FILE.

THANK YOU

Name: J.Susritha
Gmail:susrithaj2102@gmail.com