



# BANK LOAN CASE STUDY

Final Project 2

**Description :** the purpose of this project is to analyze the bank loan data and identify the potential customers

**Approach:** The steps that are followed are mentioned below

- 1) Calculated the total number of rows, and the percentage of blank field.
- 2) Removed the column with error percentage greater than 30
- 3) Removed Un-necessary columns which are not required for this particular analysis

**Tech Stack used:** Microsoft Excel 2021 MSO

Reason: Excel provides easy sorting of data, large selection of formulas, provides graphs, pie charts to visualize the data and so on.

**Insight:** used EDA to analyze Customer attributes.

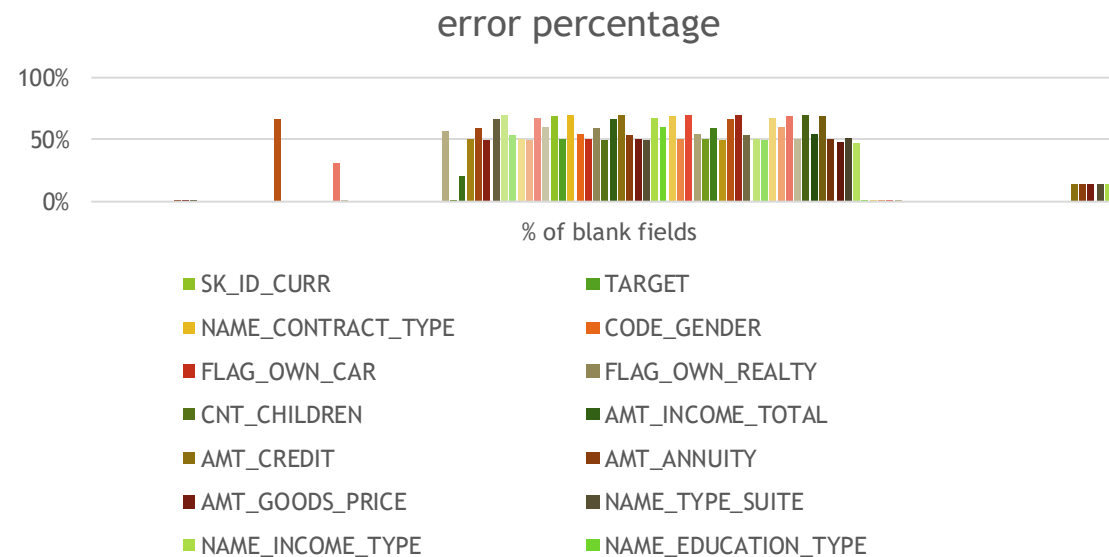
**Result :** I was able to get the data ready to draw the required results

**Drive Link :**

<https://drive.google.com/drive/folders/15QucvXbMqM8mICeQduLAA18GjERrWX8f?usp=sharing>

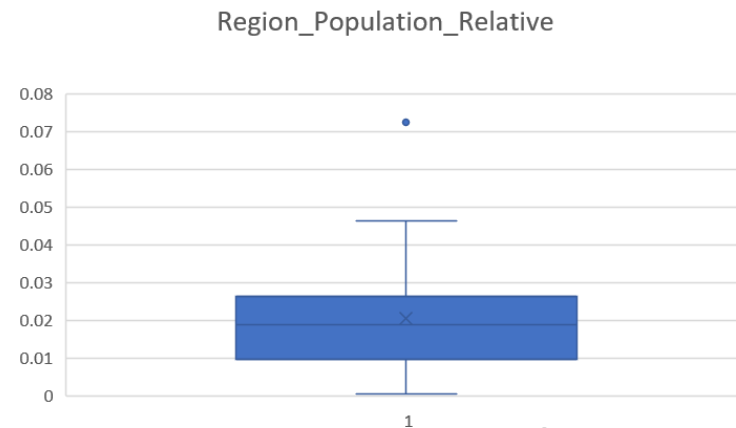
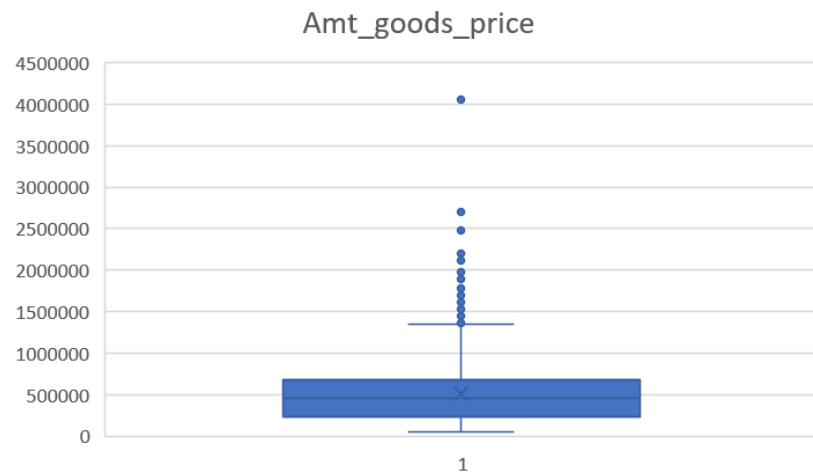
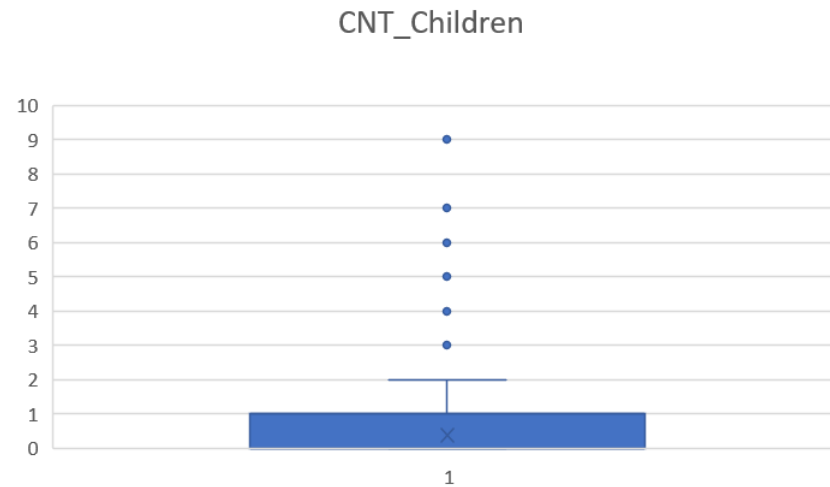
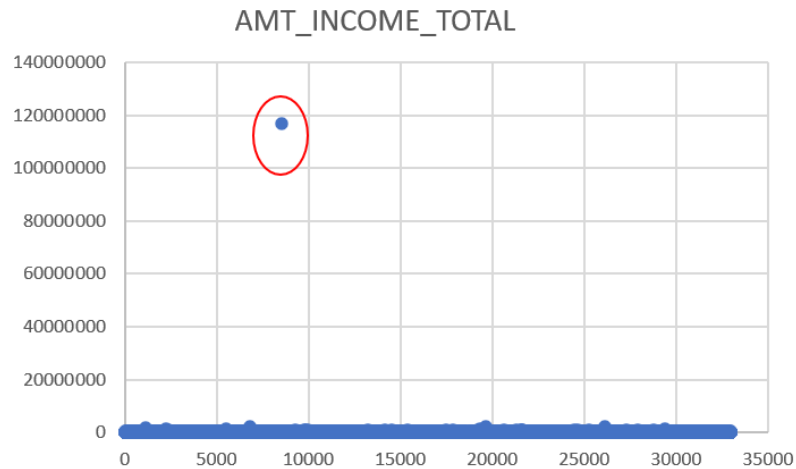
## A. Identify Missing Data and Deal with it Appropriately:

- 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.



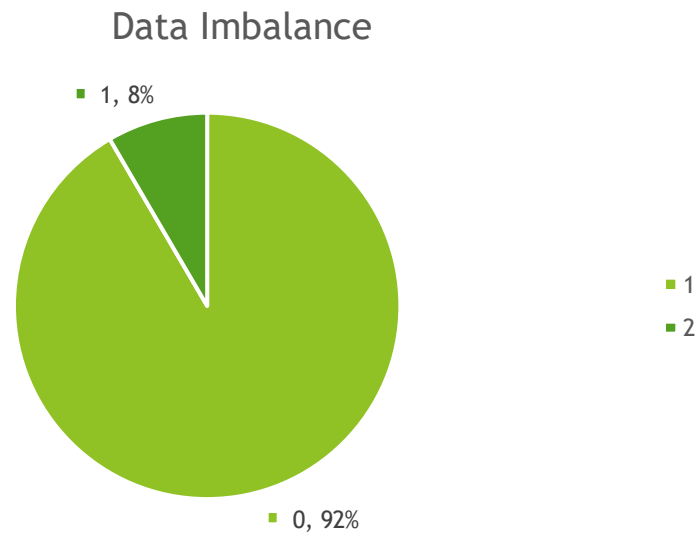
## B. Identify Outliers in the Dataset:

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



## C. Analyze Data Imbalance:

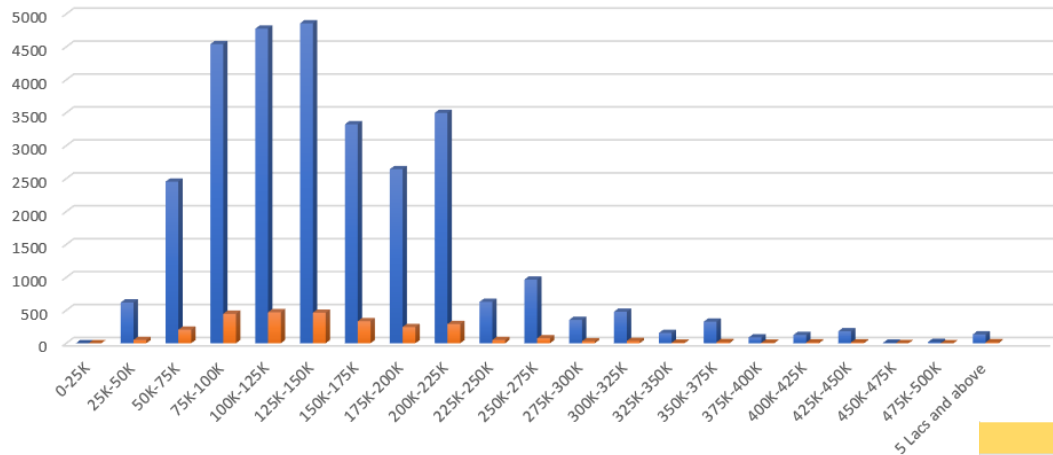
•**Task:** Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.



## D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:

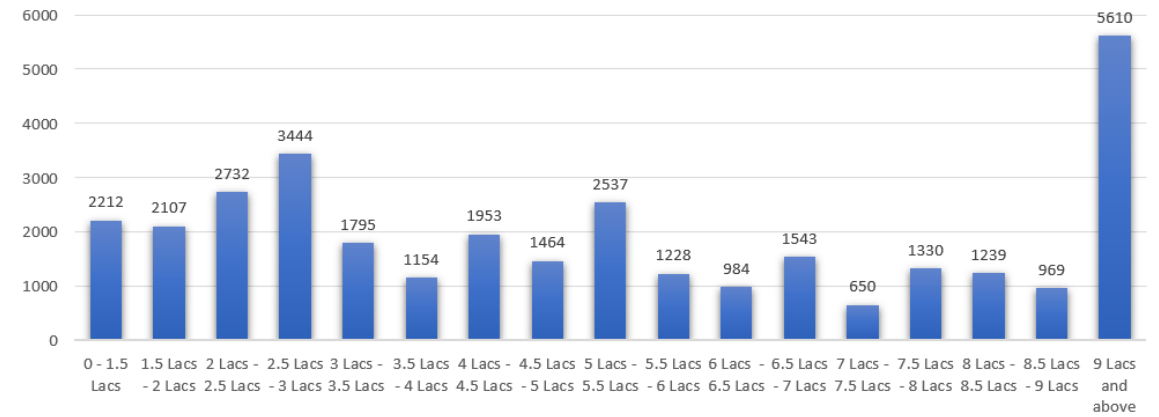
Segmented Univariate Analysis

Target Applicants : Income Bin



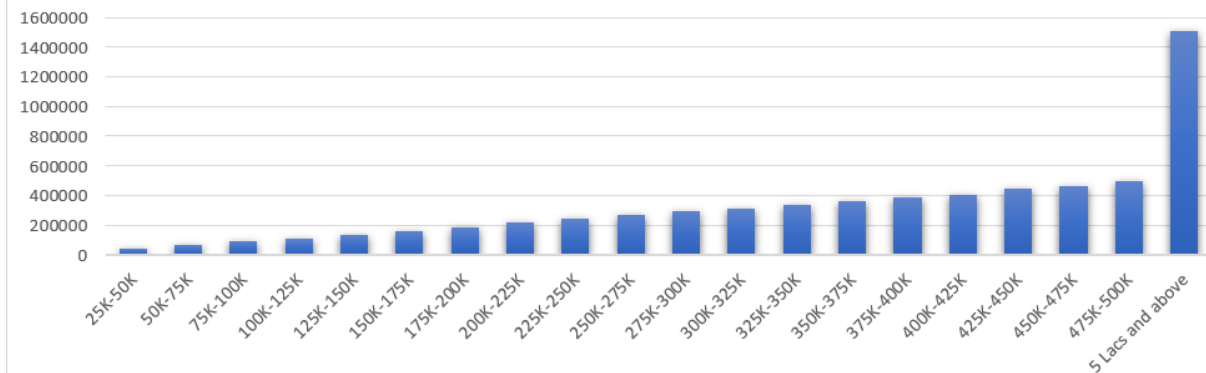
Univariate Analysis

Applicants count



Bivariate Analysis

Average



## E. Identify Top Correlations for Different Scenarios:

- 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.

	Correlation for Applicants made payment on time							
CNT_CHILDREN	1	0.006649637	-0.0005199	-0.025586336	0.351101663	-0.253632375	-0.004744871	0.023656287
AMT_INCOME_TOTAL	0.006649637	1	0.04964793	0.019220077	0.012708938	-0.022100987	0.002132565	-0.027428391
AMT_CREDIT	-0.00051995	0.049647933	1	0.096175346	-0.049140564	-0.071297877	-0.008559634	-0.100950112
REGION_POPULATION_RELATIVE	-0.025586336	0.019220077	0.09617535	1	-0.031567438	-0.001393312	-0.016472523	-0.525266782
DAYS_BIRTH(Years)	0.351101663	0.012708938	-0.0491406	-0.031567438	1	-0.640966088	0.323662747	0.0145976
DAYS_EMPLOYED (Years)	-0.253632375	-0.022100987	-0.0712979	-0.001393312	-0.640966088	1	-0.313059333	0.033552577
DAYS_ID_PUBLISH(Years)	-0.004744871	0.002132565	-0.0085596	-0.016472523	0.323662747	-0.313059333	1	0.006171201
REGION_RATING_CLIENT	0.023656287	-0.027428391	-0.1009501	-0.525266782	0.0145976	0.033552577	0.006171201	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH(Years)	DAYS_EMPLOYED (Years)	DAYS_ID_PUBLISH(Years)	REGION_RATING_CLIENT

	Correlation for Applicants with difficulties							
CNT_CHILDREN	1	0.012511658	0.00398318	-0.009900536	-0.10770393	0.010206729	0.274052344	-0.017343857
AMT_INCOME_TOTAL	0.012511658	1	0.01233019	-0.009900536	0.011538823	-0.011881657	-0.010330003	-0.011230874
AMT_CREDIT	0.003983178	0.012330187	1	0.088558546	-0.10770393	0.007059288	-0.019694243	-0.049924397
REGION_POPULATION_RELATIVE	-0.032034302	-0.009900536	0.08855855	1	-0.017900611	0.010206729	0.01072909	-0.414040533
DAYS_BIRTH(Years)	0.278593275	0.011538823	-0.10770393	-0.017900611	1	-0.612350127	0.274052344	0.045017666
DAYS_EMPLOYED (Years)	-0.209417758	-0.011881657	0.00705929	0.010206729	-0.612350127	1	-0.255454621	-0.017343857
DAYS_ID_PUBLISH(Years)	-0.022322544	-0.010330003	-0.01969424	0.01072909	0.274052344	-0.255454621	1	0.017860476
REGION_RATING_CLIENT	0.060036544	-0.011230874	-0.0499244	-0.414040533	0.045017666	-0.017343857	0.017860476	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	REGION_POPULATION_RELATIVE	DAYS_BIRTH(Years)	DAYS_EMPLOYED (Years)	DAYS_ID_PUBLISH(Years)	REGION_RATING_CLIENT