TRAINITY

PROJECT-6

**BANK LOAN CASE STUDY**

PROJECT DESCRIPTION:- In the following project I have used BANK LOAN datasets to find out insights regarding the applicants loan related information.

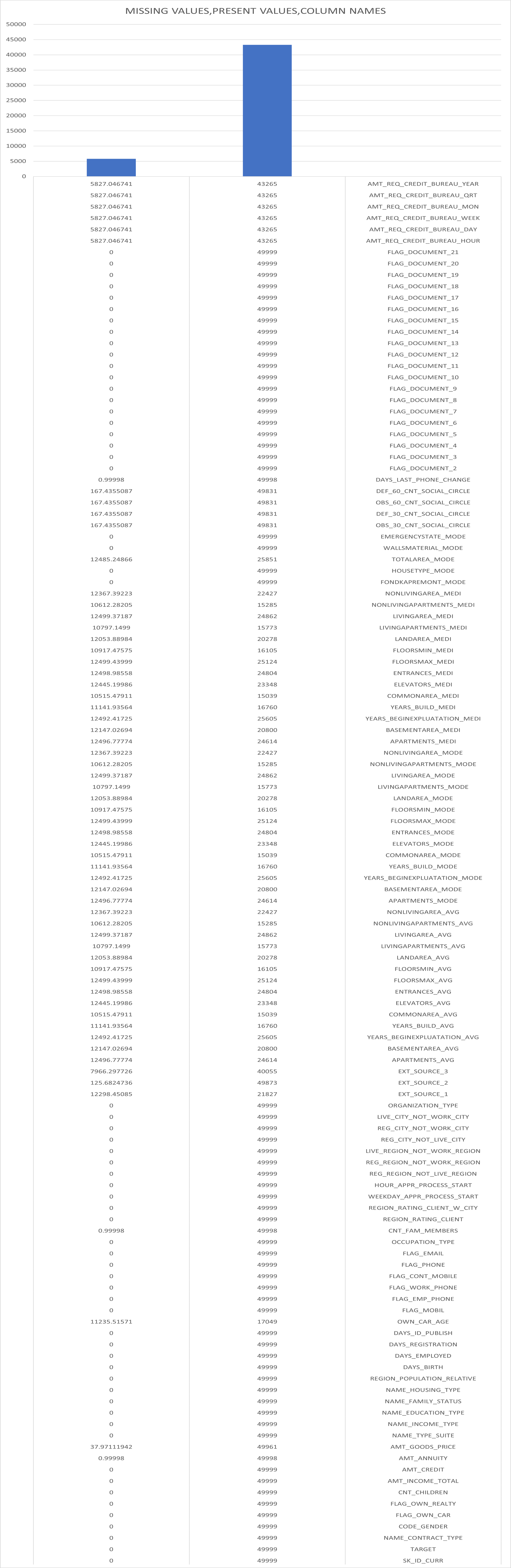
Approach:- My approach towards the project was first to understand data. Then to delete unnecessary data and fill the null rows. I have used the Microsoft Excel to analyze data and created graphs for better understanding.

Tech -Stack Used:- Microsoft Excel.

INSIGHTS:- Following are the insights which I found out by analyzing the BANK LOAN data.

1. **IDENTIFY MISSING DATA AND DEAL WITH IT APPROPRIATELY**

Following are the missing values and present values in the dataset according to column. After this I have deleted the unnecessary data and brought 122 columns up to 50 columns. I have also deleted some rows and brought 50000 rows to 43128.I have used mean, median and mode to fill the missing rows in all the necessary columns.

****

1. **Identify Outliers in the Dataset:-**

Outliers are data points that are very different from the rest in a dataset. They can happen because of mistakes or unusual events.

Following are the outliers of following columns.

1. AMT\_INCOME\_TOTAL

Here, some values that are near 120000000 which are some extreme values as compared to the data. Thus they are outliers.

1. CNT\_CHILDREN
2. YEARS\_EMPLOYED
3. YEARS\_BIRTH

1. AMT\_ANNUITY
2. AMT\_GOODS\_PRICE
3. AMT\_CREDIT
4. **ANALYSE DATA IMBALANCE**

In following table I have presented data imbalance . I have used TARGET column to represent data imbalance.

|  |  |  |  |
| --- | --- | --- | --- |
| **Row Labels** | **Count of TARGET** | Ratio | Contribution |
| 0 | 39791 | 11.92776 | 92.3 |
| 1 | 3336 |  | 7.7 |
| (blank) |  |  |  |
| **Grand Total** | **43127** |  |  |

1. **Perform Univariate, Segmented Univariate, and Bivariate Analysis.**
2. Segmented Univariate analysis.

In following chart It contains a range from 0-5 lakhs and above.

In this chart blue is represented by 0 and orange as 1. The chart showcases the 0 as successful payments and 1 as payments with difficulties.

The chart represents the applicants status of loans as per their income range.

1. Univariate analysis

The following charts represents the number of applicants as per their credit bins.

1. BIVARIATE ANALYSIS

The following charts represents the average credit amount of all the applicants for their loans as per their income bin.

1. **Identify Top Correlations for Different Scenarios**

Following are the correlations of each necessary columns with other columns.

1. Correlations of applicants with payments made on time

A screenshot of a computer screen

Description automatically generated

1. Correlations of applicants with payment difficulties

A screenshot of a computer

Description automatically generated

**RESULTS:-** Through the following project I have achieved to master Excel deep fundamentals. I have also learned critical thinking and problem solving. By completing this project I have gained confidence to find insights by analyzing the datasets in excel.