

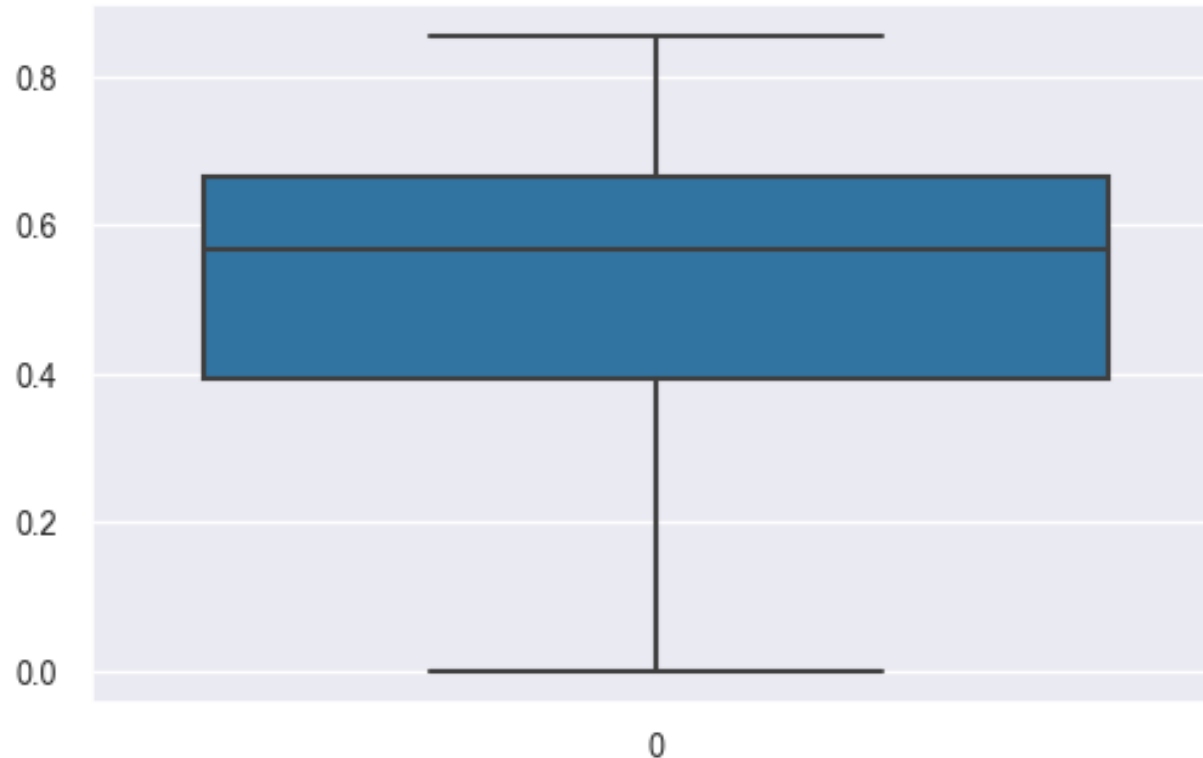
Enhancing Loan Approval Decision using EDA

- **Name :-** Pratik Pansuriya
- **Batch :-** DS C-57 June2023

Problem Statement

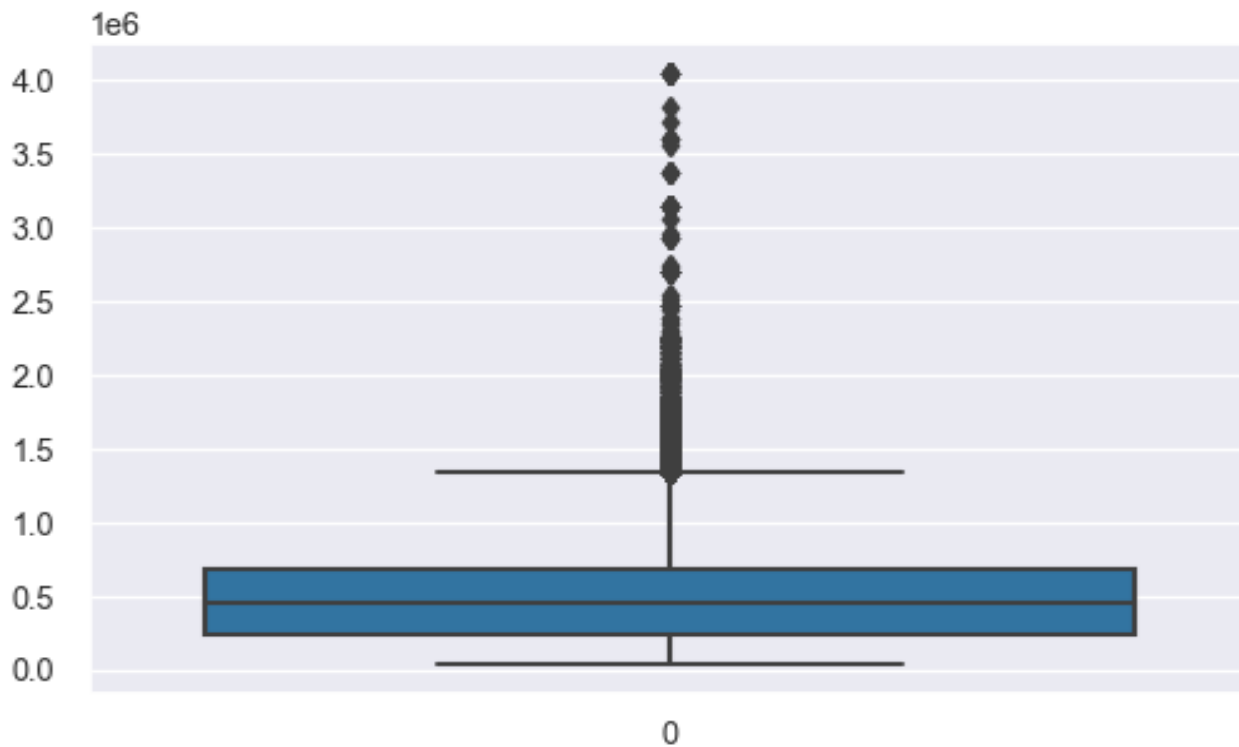
- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history.
- The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:
 1. The client with payment difficulties:
 2. All other cases:
- When a client applies for a loan, there are four types of decisions that could be taken by the client/company):
 1. Approved:
 2. Cancelled:
 3. Refused:
 4. Unused offer:

- Finding unique values with columns <15%: we observed that "EXT_SOURCE_2" and "AMT_GOODS_PRICE" are continuous variables and remaining rows are "categorical variables".
- **"EXT_SOURCE_2"---> Continuous variable**



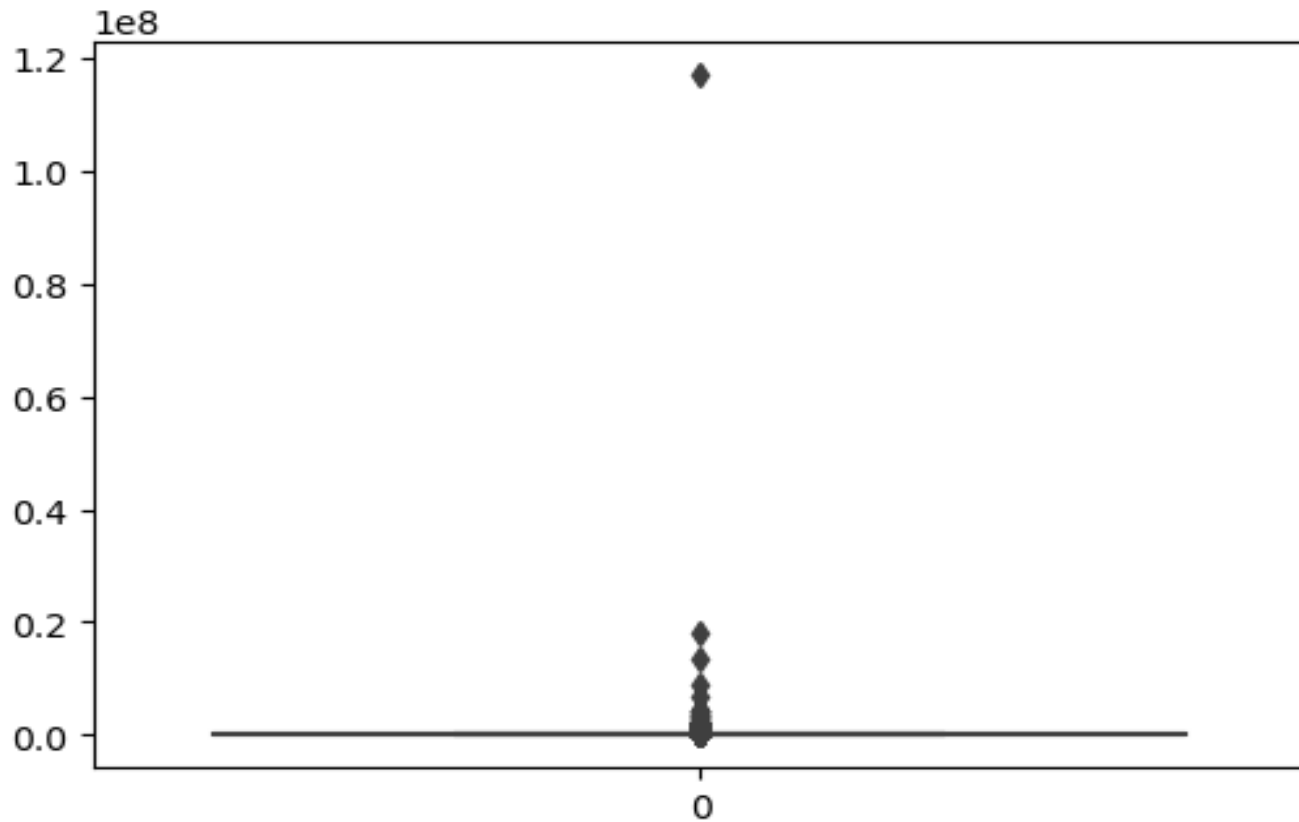
- **we observed that there is no outliers present in "EXT_SOURCE_2". data is in correctly present.**

- **"AMT_GOODS_PRICE"---> Continuous variable**



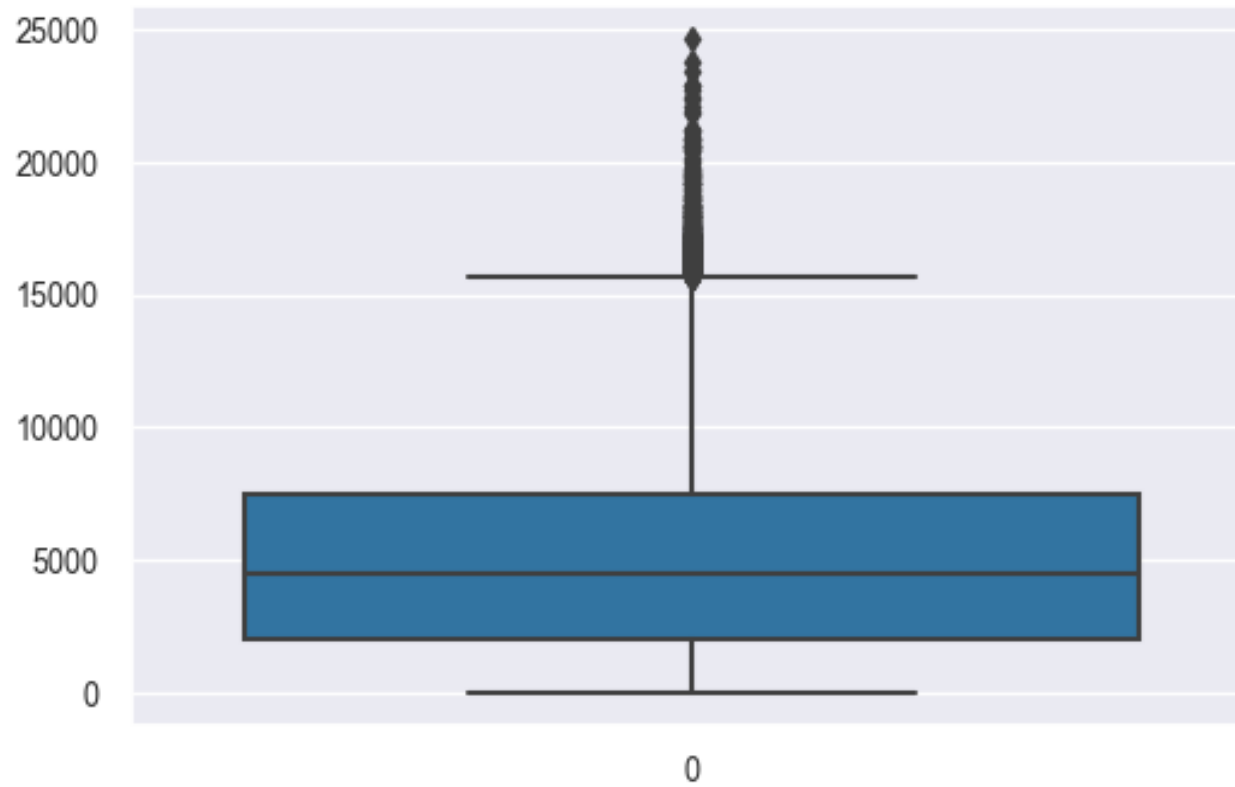
- **Outliers present in "AMT_GOODS_PRICE".**

- **Box plot for 'AMT_INCOME_TOTAL'.**

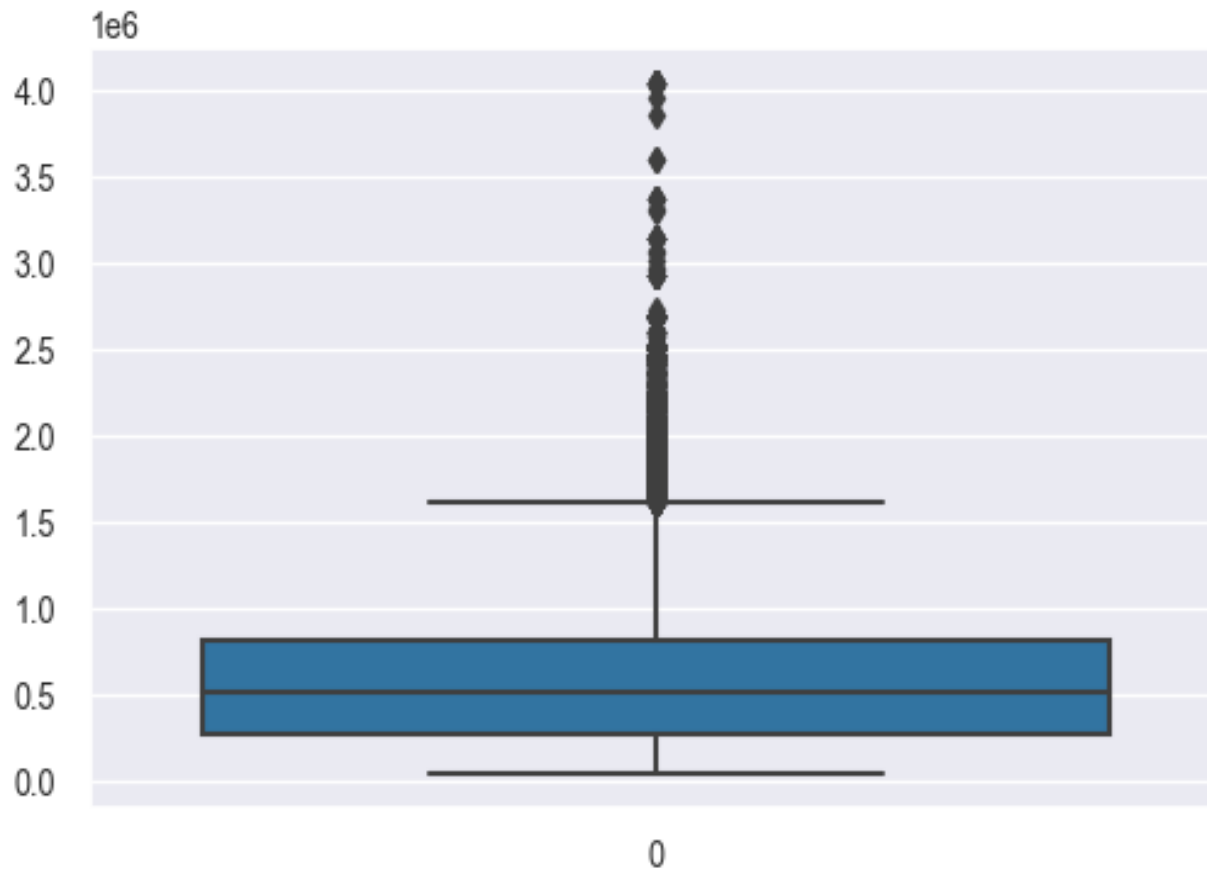


- **In "AMT_INCOME_TOTAL" only single high value data point is present as outlier.**

- **Box plot for "DAYS_REGISTRATION".**

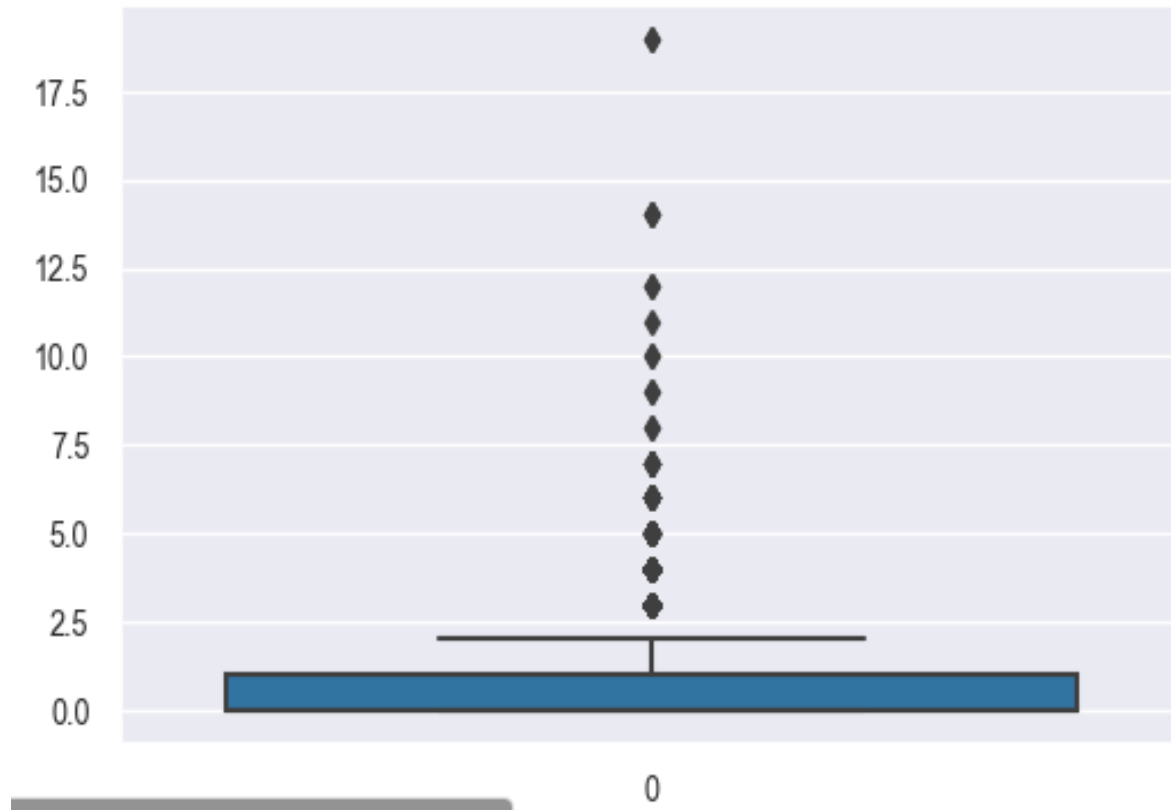


- **Box plot for 'AMT_CREDIT'.**



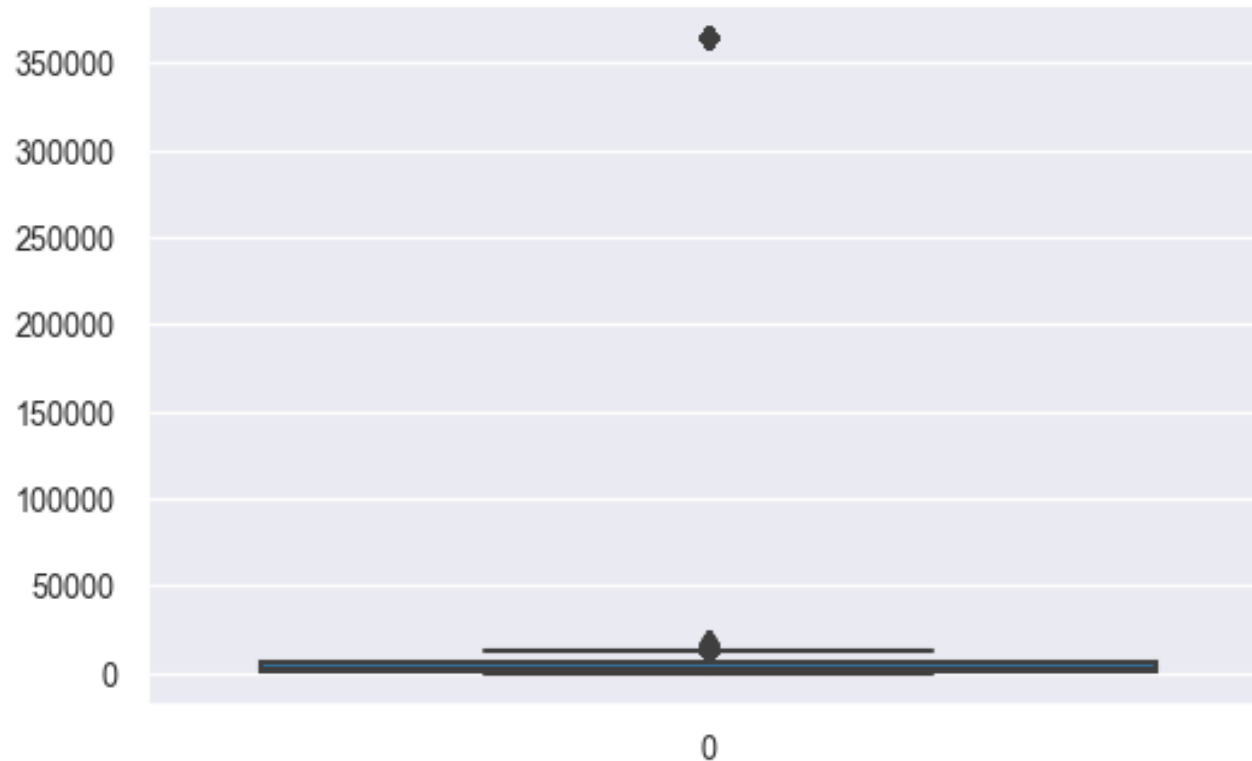
- **1st quartiles and 3rd quartile for "AMT_ANNUIITY" is moved towards first quartile.**

- **Box plot for 'CNT_CHILDREN'.**

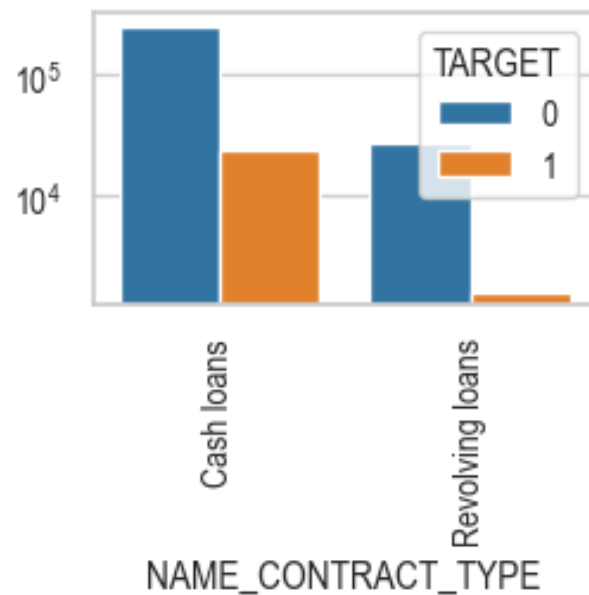
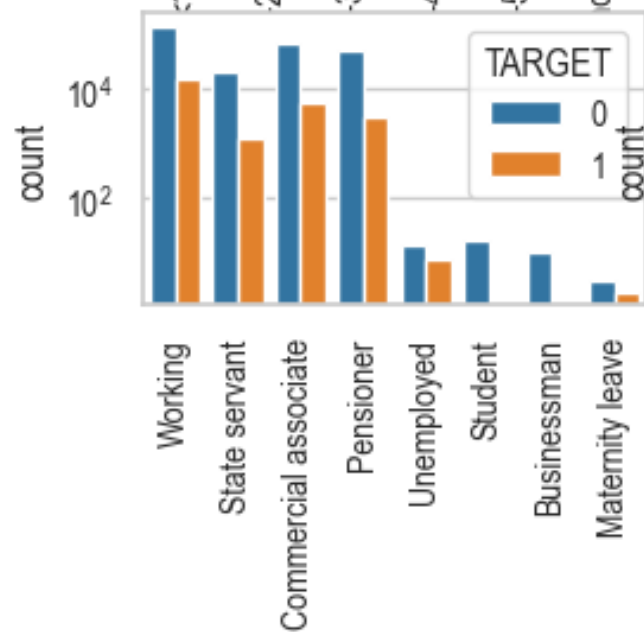
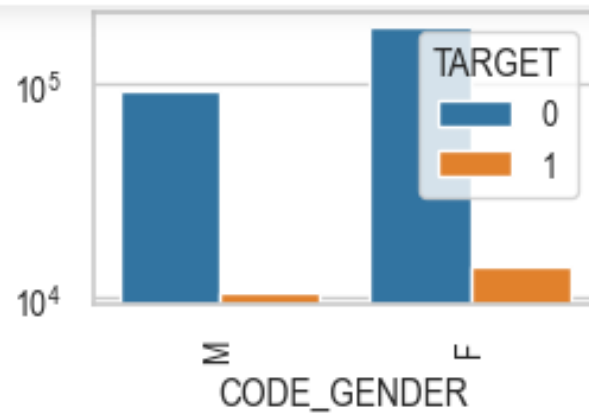
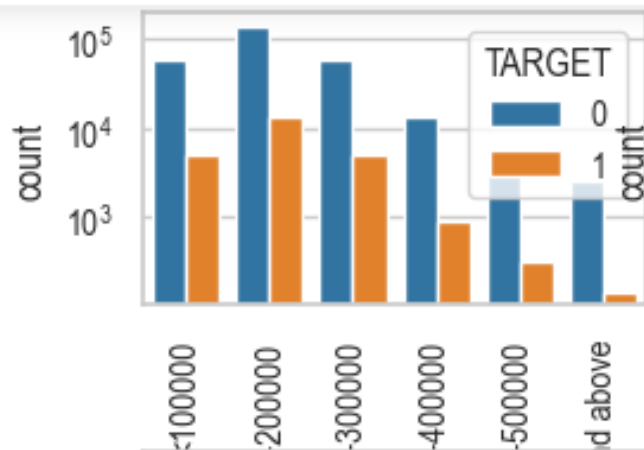


- **here we observed that most of the data are present in the 1st quartile.**

- **Box plot for 'DAYS_EMPLOYED'.**



- **1st quartiles and 3rd quartile for "DAYS_EMPLOYED" is stays first quartile.**
- **we observed From above box plots we found that numeric columns have outliers.**



❖ Observations:

- People in range 100000-200000 have high number of loan and also have high in defaulter.
- Income segment >500000 has less defaulter. CODE_GENDER:.
- The % of defaulters are more in Male than Female NAME_INCOME_TYPE:
- Student and business are higher in percentage of loan repayment.
- Working, State servant and Commercial associates are higher in default percentage.
- Maternity category is significantly higher problem in repayment. NAME_CONTRACT_TYPE
- For contract type 'Cash loans' are high in number of credits than 'Revolving loans' contract type.
- By above graph 'Revolving loans' is small amount compared to 'Cash loans'

- **Correlation**
- From the correlation analysis, we can say that the highest correlation (1.0) is between ENTRANCES_MEDI with ENTRANCES_AVG.

- **Conclusion from the Analysis:**
 1. Banks must target more on contract type 'Student' , 'Pensioner' and 'Businessman' for profitable business.
 2. Banks must focus less on income type 'Working' as it has most number of unsuccessful payments in order to get rid of financial loss for the organization.