EDA CASE STUDY
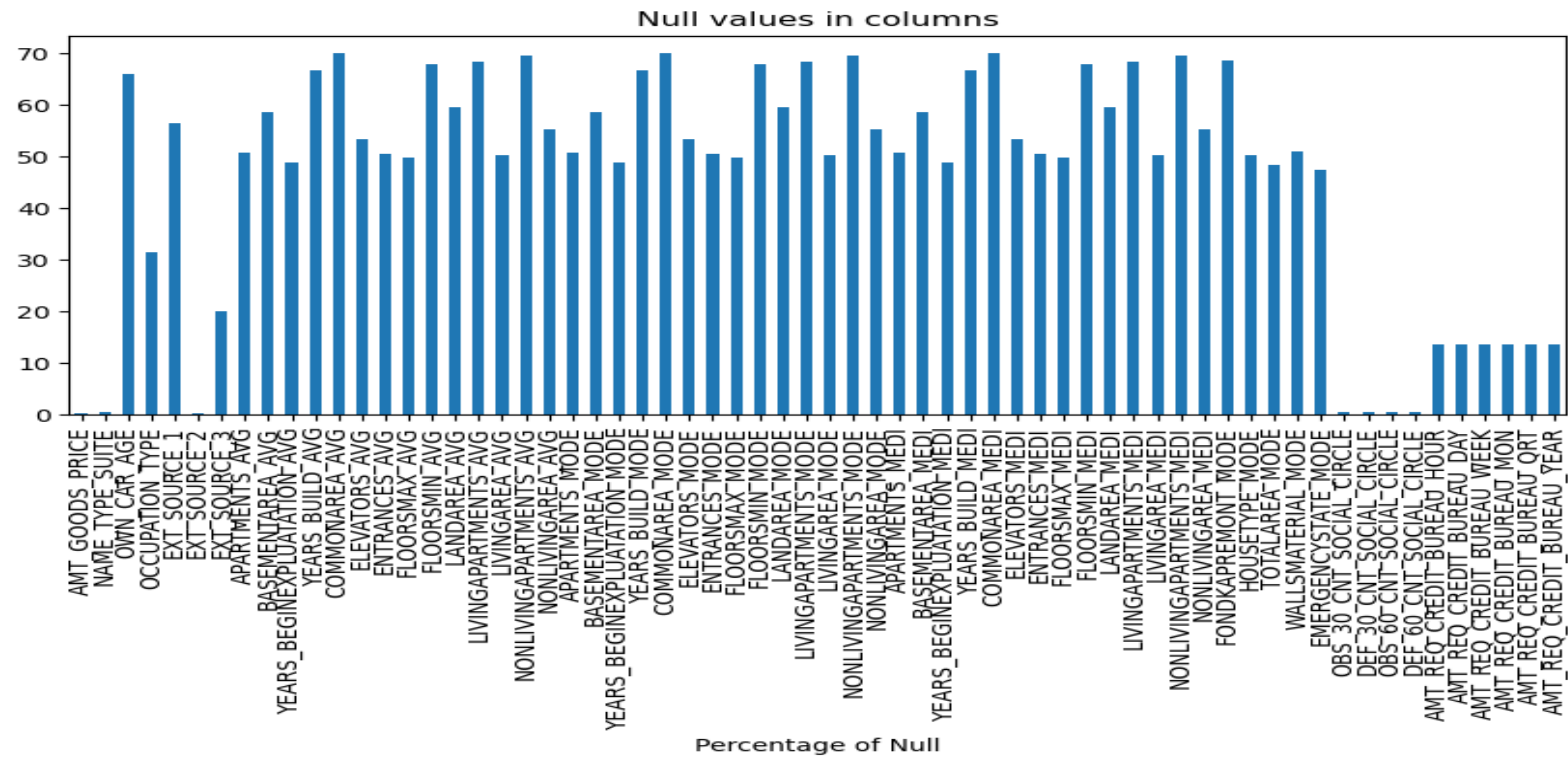
Submitted by

Gourav Bhattacharjee

About case study:
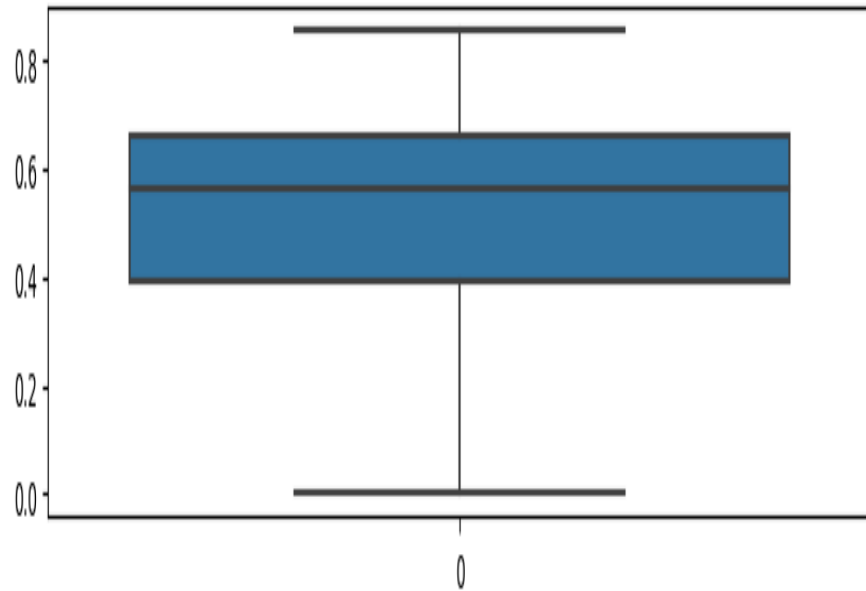
- Basic understanding of risk analytics in banking and financial service and how data is used to minimize the risk of losing money while lending to customer.
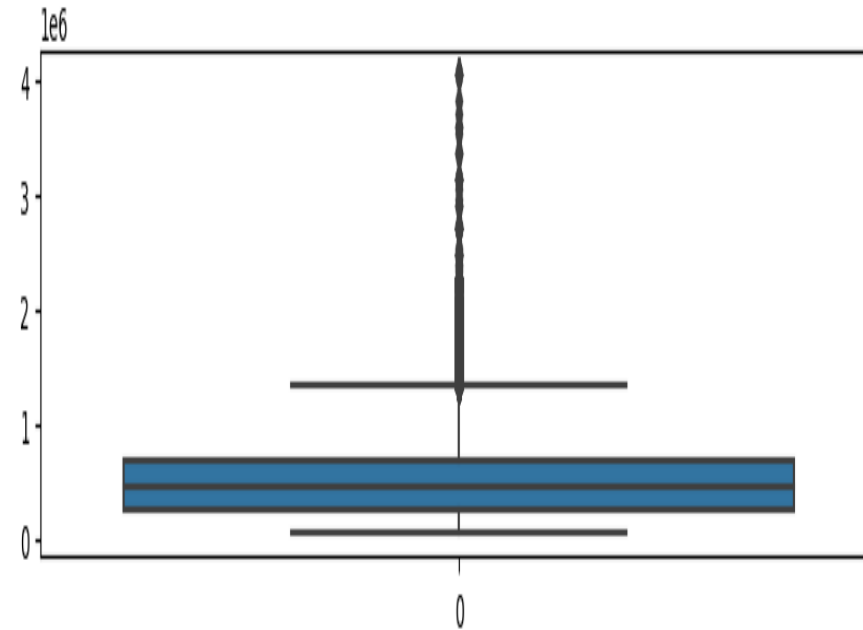
Steps:

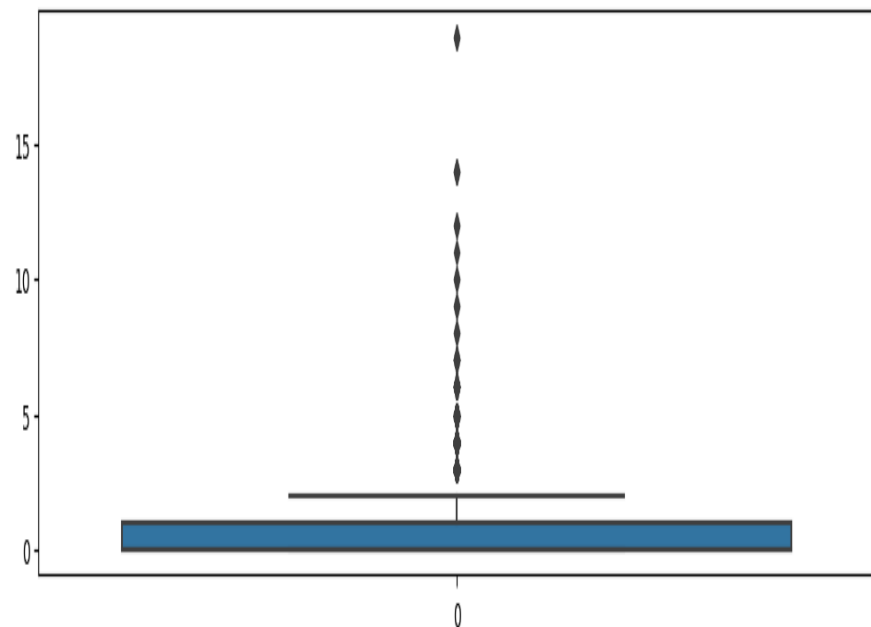Visualizing Null values of columns in graph :
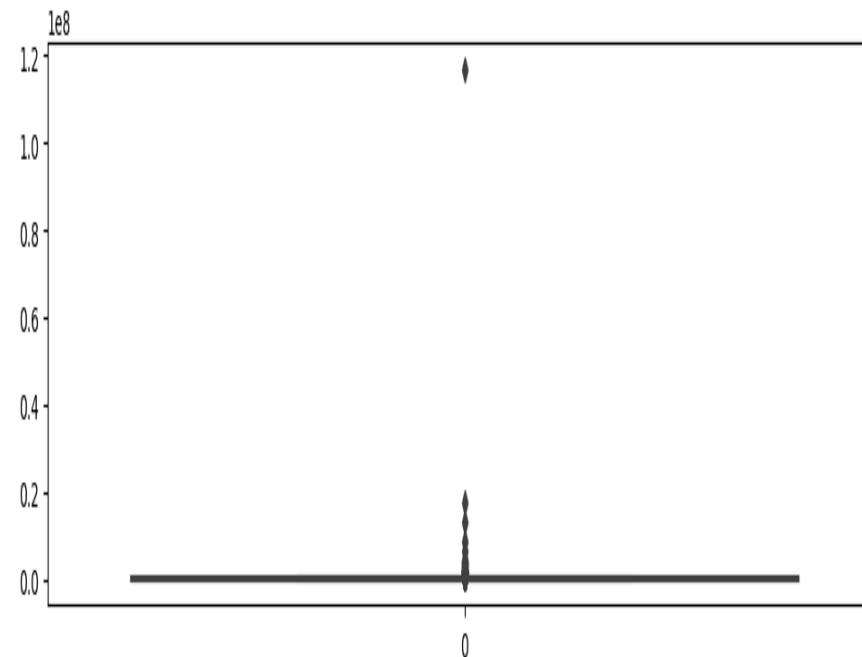
EXT_SOURCE_2

AMT_GOODS_PRICE

From Boxplots:

- ✓ For 'EXT_SOURCE_2' no outliner present. So, the data is rightly present.
- ✓ For 'AMT_GOODS_PRICE' outliner present in the data. So, need to impute with median value: 4
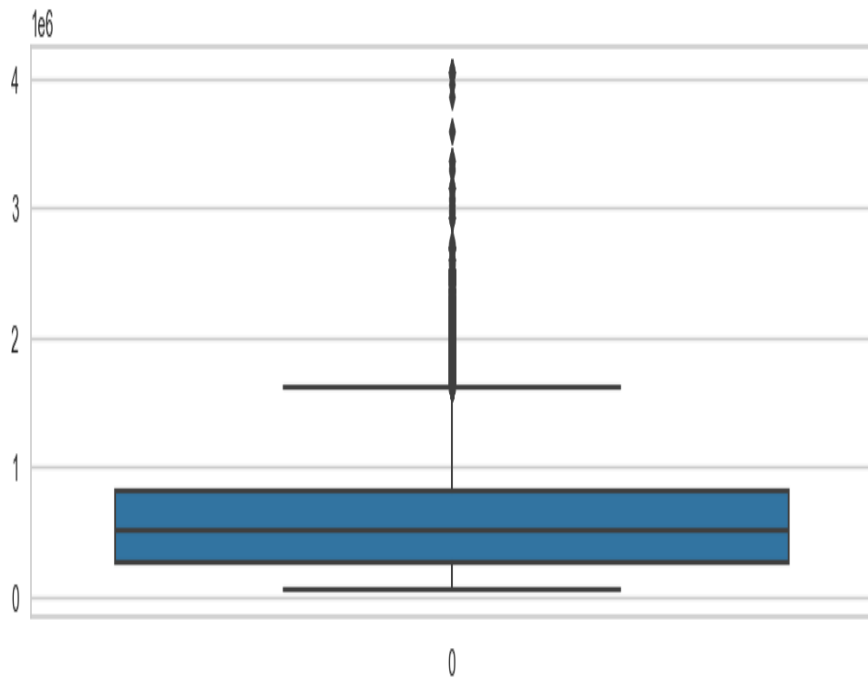
Visualizing:

CNT_CHILFREN

AMT_INCOME_TOTAL

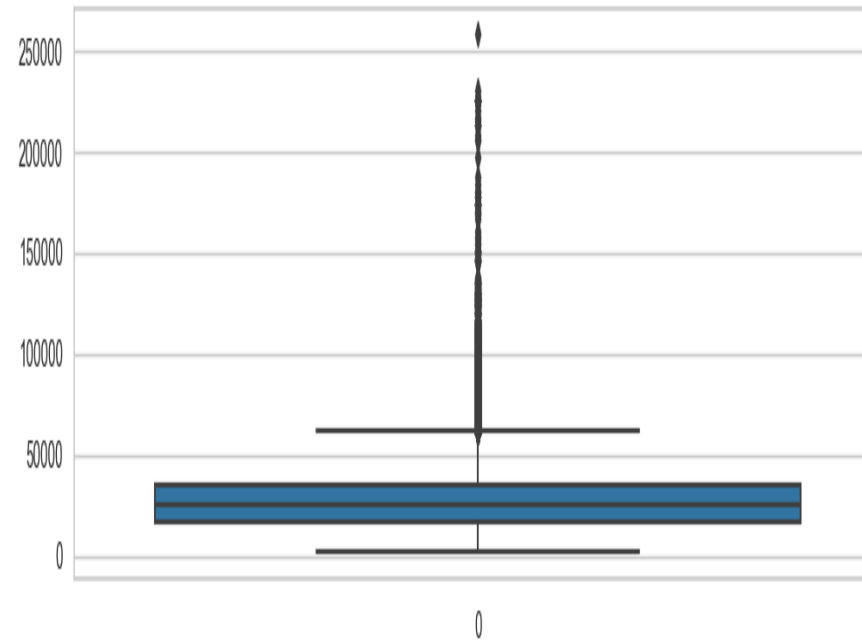
- 1st quartile is missing for CNT_CHILDREN which means most of the data are present in the 1st quartile.
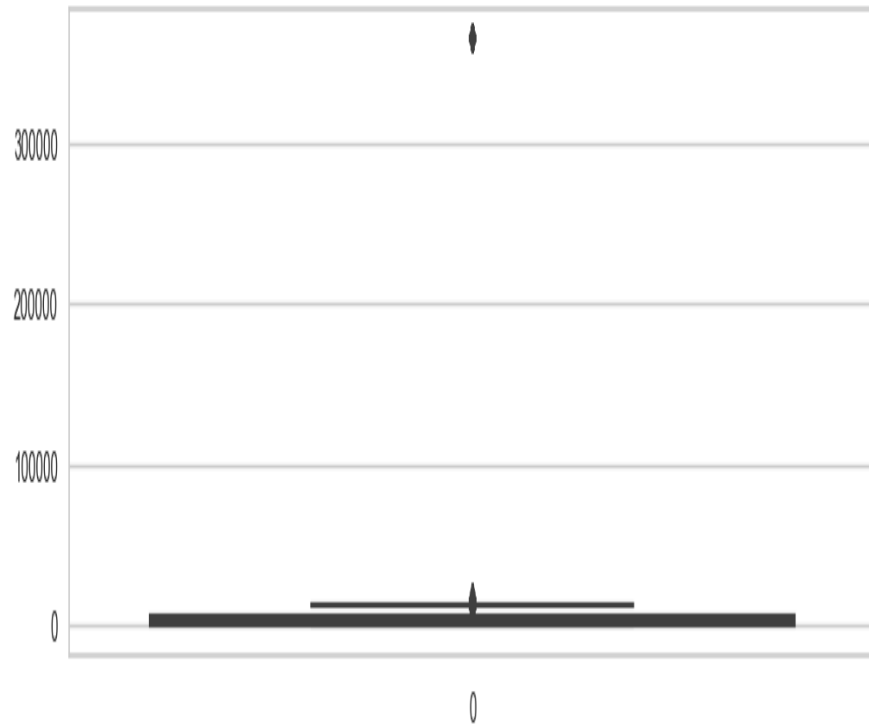- In AMT_INCOME_TOTAL only single high value data point is present as outliner.
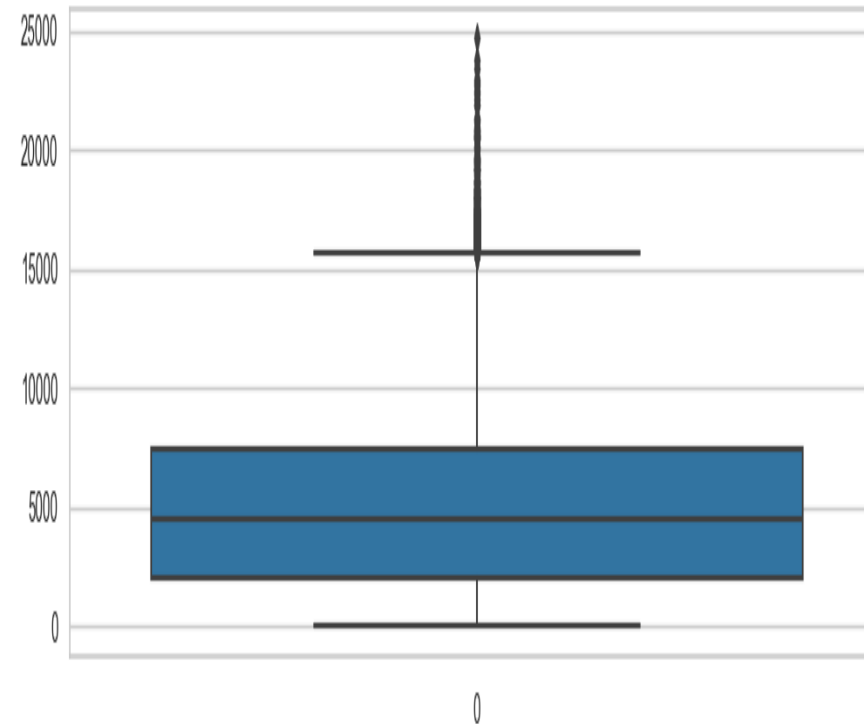
AMT_CREDIT

AMT_ANNUITY

- AMT_CREDIT has little bit more outliners.
- 1st quartiles and 3rd quartile for AMT_ANNUITY is moved towards first quartile.
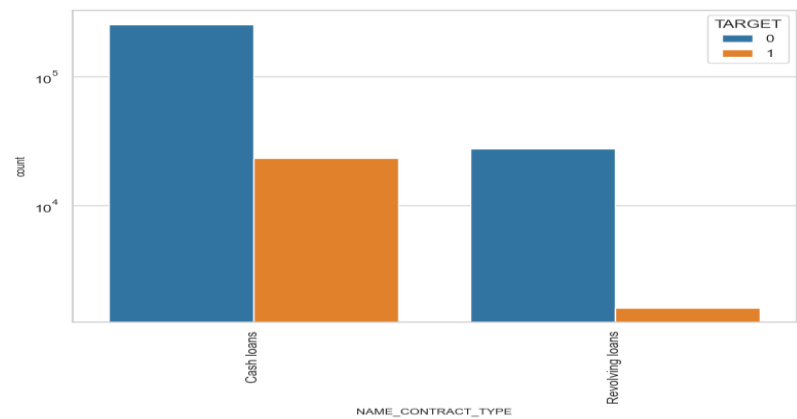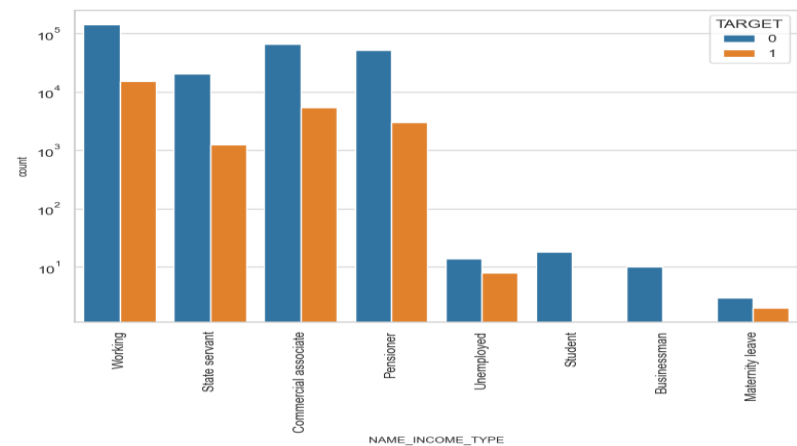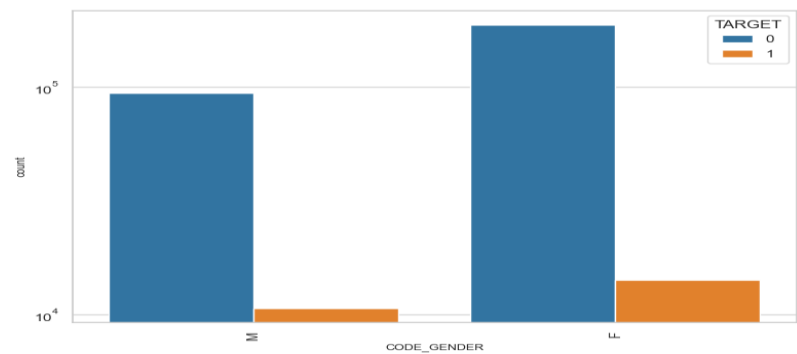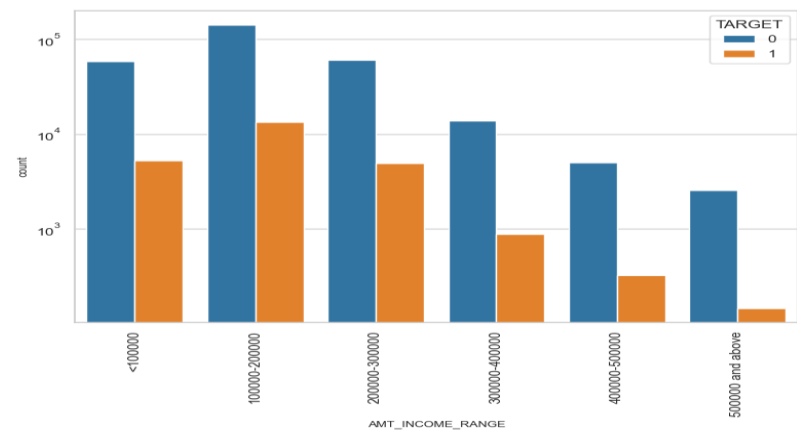
DAYS_EMPLOYED

DAYS_REGISTRATION

- 1st quartile and 3rd quartile for DAYS_EMPLOYED is stays first quartile.
- 1st quartile and 3rd quartile for DAYS_EMPLOYED is stays first quartile.

❖ **From above boxplots, we found that numeric columns have outliners.**

Univariate Analysis

Observation

AMT_INCOME_RANGE:
- 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 percentage defaulters are more in Male than Female.

NAME_INCOME_TYPE:
- Student and business are higher in percentage of loan repayment.
- Working, State servent 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 higher in number of credits than revolving loans contract type.
- By above graph revolving loans is small amount compared to cash loans.

Conclusion from the Analysis

- Bank must target more on contract type.
- Bank must focus less on income type working.

THANK YOU