# REPORT FOR THE CLEANING OF THE CREDIT_SCORE_CLASSIFICATION DATA SET

## Cleaning process:

1. After getting an overview of where each column falls between numerical and categorical/object data types, some observed problematic columns will have their dtype changed to match the data contained within.
2. Filled in missing values with appropriate placeholders and verified that there is no longer any null values.
3. Employed the use of summary statistics of the numerical data types to get insight into data and existing outliers and visualized using boxplots.
4. Resolved outliers then merged cleaned numerical values with relevant categorical equivalent to form a new cleaned dataset.

## Improvements made:

1. There were some observed problematic data within some columns of object data types which I then homogenized for readability and easy reference.

## Challenges:

1. Had some issues merging cleaned numerical data set and relevant object dtypes. Then deciding if dropping the object data types that don't have the equivalent rows in the numerical dataset won't affect the dataset.
2. Presence of more problematic column data that do not follow the common theme of the data stored within the column. Difficulty deciding if they are relevant missing values or random.

## Lessons Learned:

1. That there can still be more data that needs to be cleaned outside of dealing with general null values.
2. Trimmed data set after merging numerical data that has been cleared of outliers and relevant categorical/object data, then dropping the rest of the unmatched dataset to form the new cleaned data.