Data Quality Report

# Unnamed Columns Count (Dataset Level):

WARNING: There are 1 columns found which had no name.  
SOLUTION: Assign column names manually so as to remove confusion during data file handling.

# Duplicate Rows Percentage (Dataset Level):

INFO: There are 0.0% duplicate rows in the uploaded dataset.

# Outliers Detection (Column Level):

WARNING: There are 1 columns contain outlier values. They are as follows:

• Column Age contains 100 outlier values ranging from 58 to 58.  
  
SOLUTION: Outliers in some columns could be an issue as they might impact statistical measures such as mean and standard deviation in a wrong way so that they show influenced results. Moreover, if one plans to use the outlier affected columns in regression models, then outliers can distort the relationship between variables in the regression model and impact the model's coefficients and overall fit disproportionately. If the values in the range above are same, then it might be due to a singular value being the sole outlier but having high frequency in the column.

# Null Values Proportion (Column Level):

INFO: There are no null values in the uploaded dataset.

# Inconsistent Data Types Check (Column Level):

INFO: There are no columns which have inconsistent data types.

# Summary Statistics (Column Level):

INFO: Summary statistics for the number-related columns are as follows:

• Column Id has a non-null row count of 4997, mean of 4719, median 3579, mode 3265, minimum value 1258, maximum value 9654, standard deviation 2367, variance 5603041, range 8396, first quartile 2587, third quartile 6540, skewness 0, kurtosis 0, and has 0 na values.  
  
SOLUTION: If you find insufficient number-related columns above then please check if some columns which should be number-related have multiple data types values inside them (as mentioned the above section). Please make sure no such column is being considered as a multiple data type column.

REMARKS: If you find insufficient number-related columns above then please check if some columns which should be number-related have multiple data types values inside them (as mentioned the above section). Please make sure no such column is being considered as a multiple data type column.