Missing data is a common problem when performing statistical anlaysis, which can in turn result in bias if not resolved. In fact, according to journal [1], potential missing data that happen during the data collection process can be classified under three different categories. These include: Missing completely at random (MCAR), Missing at random (MAR), Not missing at random (NMAR). Due to the nature of the dataset, the missing values fall within the Missing completely at random (?) category, meaning that “the participants with complete data cannot be distinguished from participants with incomplete data”. A method that can be implemented to handle missing values is imputation. This approach “attempts to estimate the values of the missing data and ‘fill-in’ or impute new values”. Mean substitution is a type of imputation that we performed on the data, in which the missing values for participants were replaced by the mean of all the observed values within the same variable/feature vector. This was implemented solely on the numerical data which are namely: The child’s weight at an age around 15.5 years, the child’s height at an age around 15.5 years and the child’s total IQ score on WISC-III. Whereas, for the re

[1] <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-842X.2001.tb00294.x>