Key Takeaways:

* Charlie is an outlier for both his MathScore and EnglishScore (both scores are outside lower boundary).

A computer screen shot of a computer program

Description automatically generated

* + Since there is a correlation between all scores, decided to remove Charlie from the Analysis as we cannot determine or predict his ScienceScore.
* Both Attendance and Age has no correlation with MathScore, EnglishScore and ScienceScore. This can justify removal of Charlie when Analyzing the Scores.
* For Age Column, we capped the outliers age and impute the missing age with the median.
* For Attendance, we impute the missing value with the median.
* After removing Charlie, decided to impute the missing data with the Median for all 3 scores, since all 3 are at least moderately skewed.

Here is the dataset.

A white grid with black lines

Description automatically generated with medium confidence