**ML Assignment 1 – Statistical Measures and Hypothesis Testing**

**Students Performance in Exams - Statistical Analysis**

The goal of this analysis is to understand students' performance in exams by examining descriptive statistics, identifying and handling outliers, exploring distributions, and performing hypothesis testing.

**Dataset Link:** <https://raw.githubusercontent.com/ArchanaInsights/Datasets/main/StudentsPerformance.csv>

**Analysis Steps:-**

1. **Descriptive Statistics:**

Compute the following descriptive statistics for each of the score columns math score, reading score, and writing score:

* Mean
* Median
* Mode
* Minimum
* Maximum
* Standard Deviation

Provide a summary of these statistics and discuss any interesting observations.

1. **Identify and Handle Outliers:**
   1. Use boxplots to visually identify outliers in the math score, reading score, and writing score columns.
   2. Calculate the Interquartile Range (IQR) for each of these score columns.
   3. Determine the lower and upper bounds for outliers using the IQR method.
   4. Handle the outliers by removing them from the dataset.
2. **Check Distribution and Skewness:**
   1. Plot histograms and/or density plots for math score, reading score, and writing score to visualize their distributions.
   2. Calculate and interpret the skewness for each of these score columns.
   3. Discuss whether the distributions are skewed to the left or right and suggest any appropriate transformations if needed to address skewness.
3. **Hypothesis Testing:**
   1. **Formulate Hypotheses:**

* State the null and alternative hypotheses to test whether the mean math score differs between students who completed the test preparation course and those who did not.
  1. **Significance Level:**
* Specify the significance level (α - alpha) you will use for the hypothesis test. Common choice is α = 0.05.
  1. **t-test Test Statistic & p-value:**
* Perform a t-test to compare the means of the two groups.
* Calculate the test statistic and p-value.
* Provide the results and interpret them in the context of the hypothesis.
  1. **Decision:**
* Based on the p-value or the t-statistic, decide whether to reject or fail to reject the null hypothesis.
* Discuss the implications of your decision in relation to the context of the problem.