**Report: Analysis of Student Performance Indicators**

**1. Introduction**

* **1.1. Objective:** To analyze and visualize the relationships between various demographic and preparatory factors (gender, test preparation, lunch type) and academic performance in math, reading, and writing scores.
* **1.2. Data Source:** The analysis is based on the "Student Performance" dataset. The raw data was saved https://raw.githubusercontent.com/mosomo82/COMP\_SCI\_5530/refs/heads/main/Assignment/Assignment\_1/Q2\_Data\_Visualization/raw\_data/StudentsPerformance.csv to first loaded into a pandas DataFrame for processing.
* **1.3. Workflow Overview:** This report follows a standard data analysis workflow:
  + **Ingest:** Loading the raw data.
  + **Preprocess:** Cleaning the data, handling any missing values, and creating new features (e.g., overall\_avg) as required for analysis.
  + **Analyze & Visualize:** Generating five distinct visualizations to answer specific research questions about the dataset, followed by detailed interpretations.

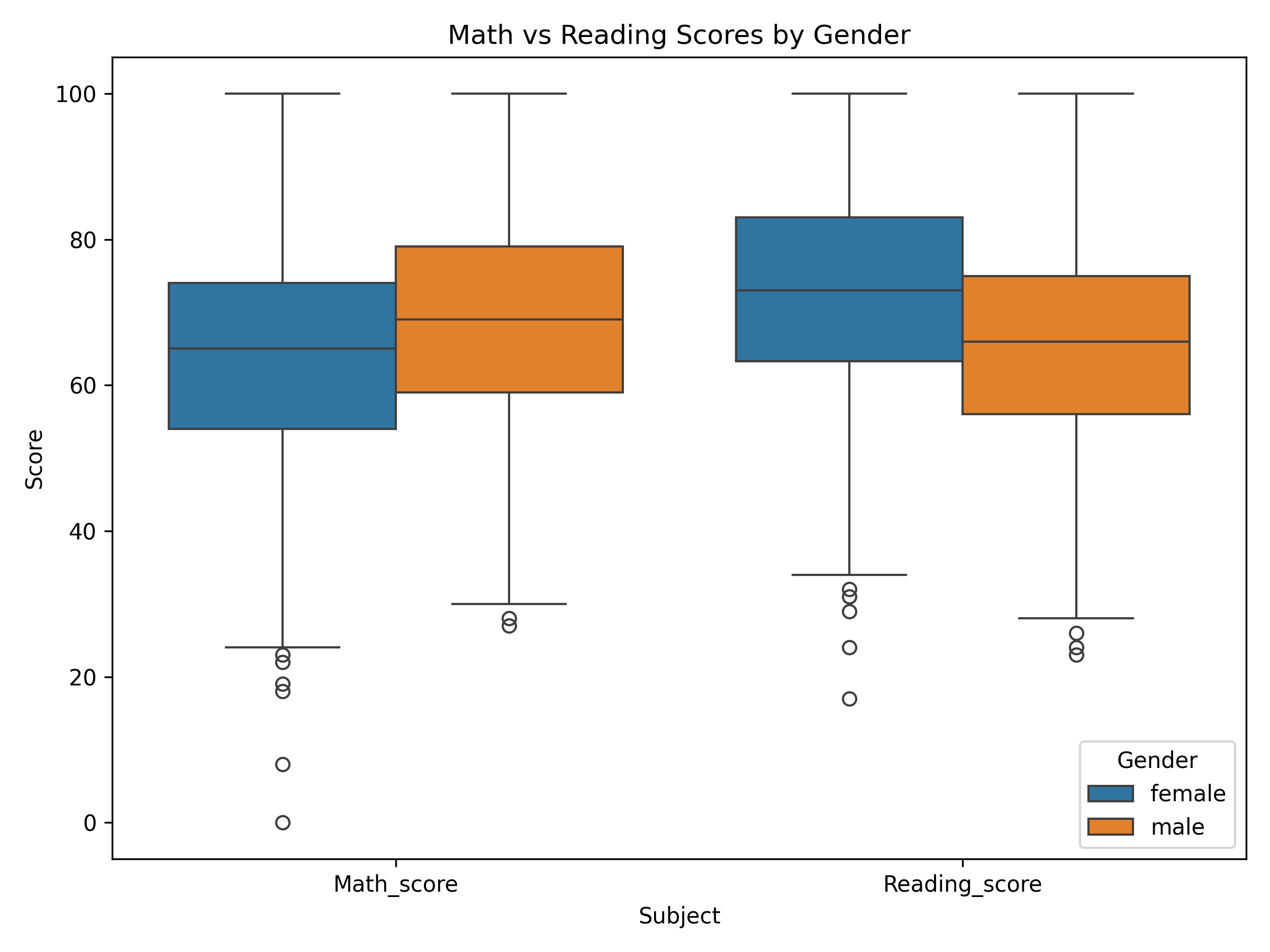
**2. Data Preprocessing**

* **2.1. Initial Data Loading & Inspection:** The dataset was loaded and examined for completeness, data types, and potential inconsistencies.
* **2.2. Handling Missing Values:** The sample dataset is large enough to drop the row with the missing values.
* **2.3. Feature Engineering:** An overall\_avg column was created by averaging the math\_score, reading\_score, and writing\_score to facilitate the analysis for Visualization 3.

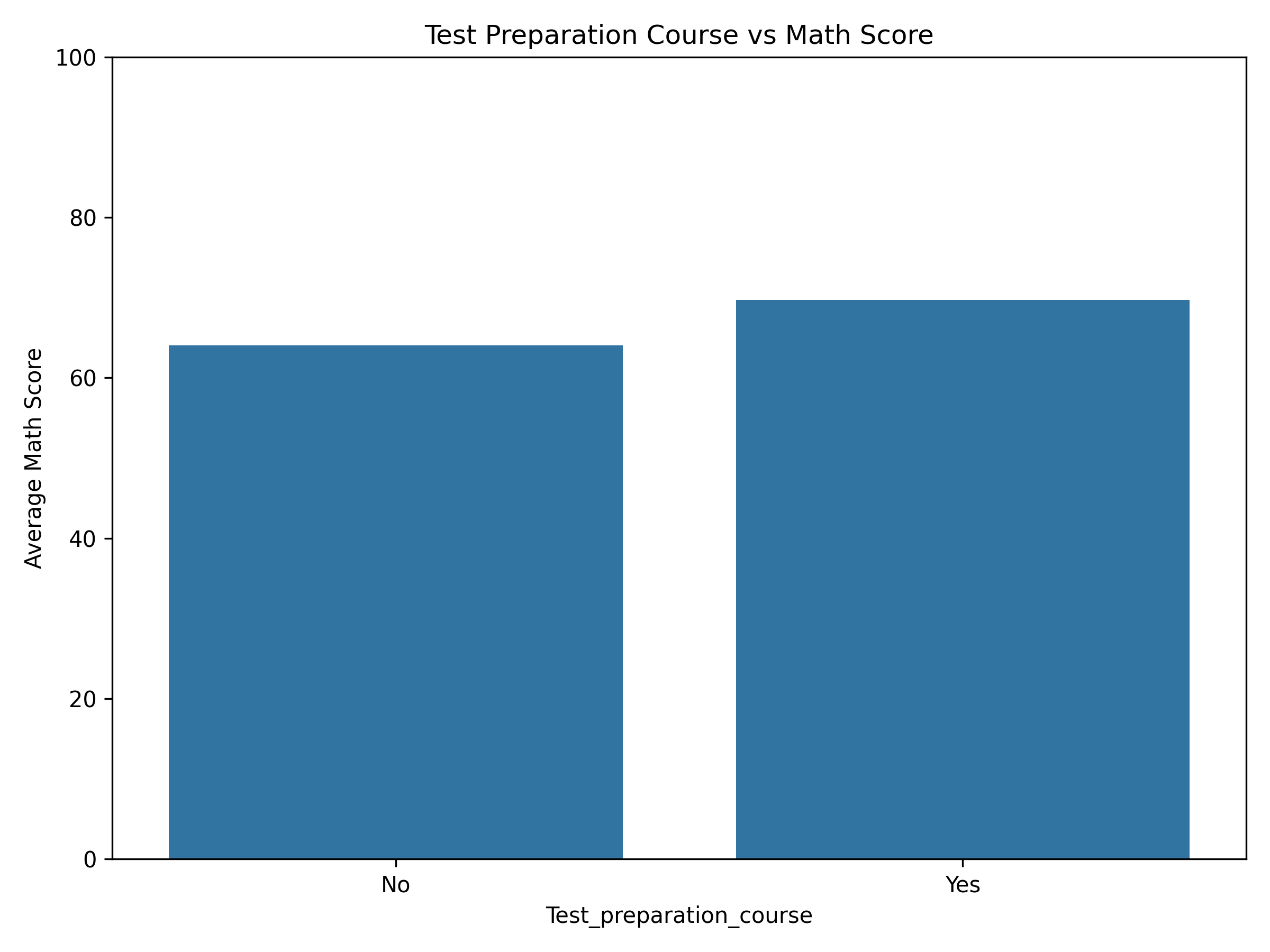
**3. Analysis and Visualizations**

This section presents the five data visualizations created to explore the student performance data. Each subsection includes the research question, the generated figure, and a detailed interpretation of the findings.

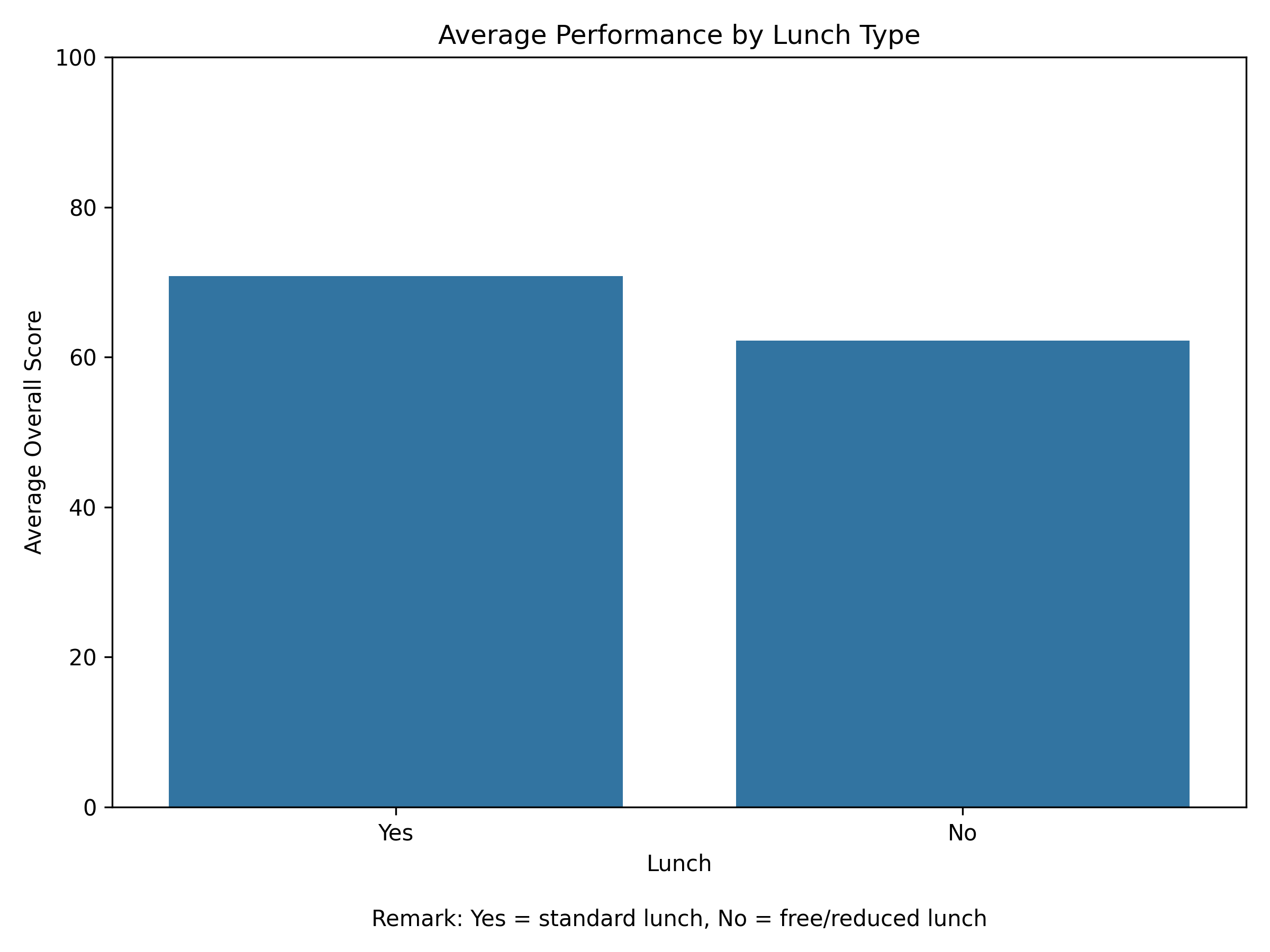
* **3.1. V1 — Gender Differences in Math vs. Reading Performance**
  + **Question:** Are there gender differences in math vs reading scores?
  + **Interpretation:** The side-by-side plots show clear differences in performance between male and females in both subjects of math and reading. Males students tend have an advantage in math. Female students tend to have a stronger advantage in reading. The gap between reading and math ability appears to be larger for females than it is for males.
  + **Figure 1:**

**

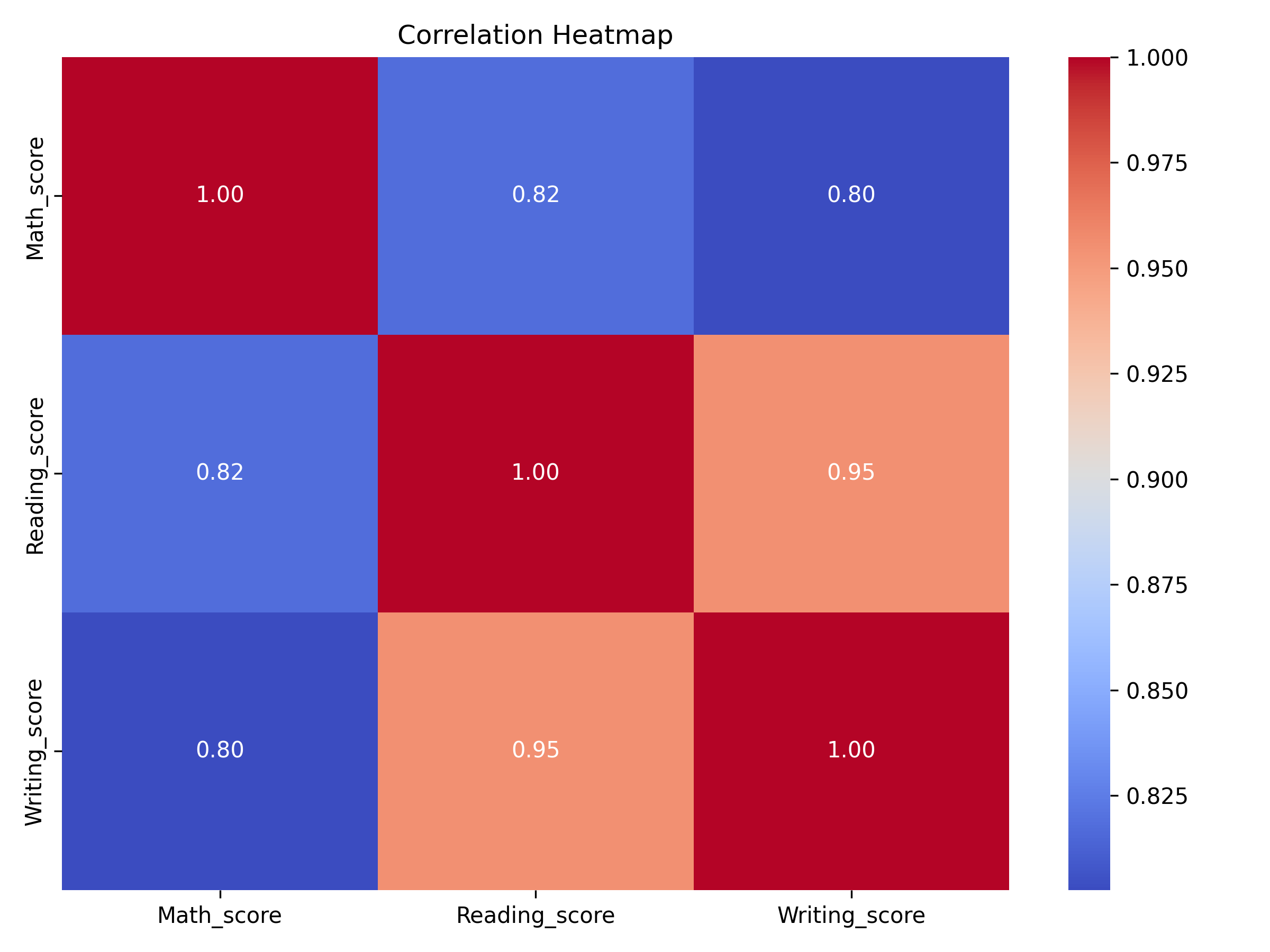
* **3.2. V2 — Impact of Test Preparation on Math Scores**
  + **Question:** Do students who completed a test preparation course score higher in math?
  + **Interpretation:** Yes, students who completed a test preparation course score higher in math on average. The difference suggests that the test preparation course has a positive impact on student’s math performance.
  + **Figure 2:**



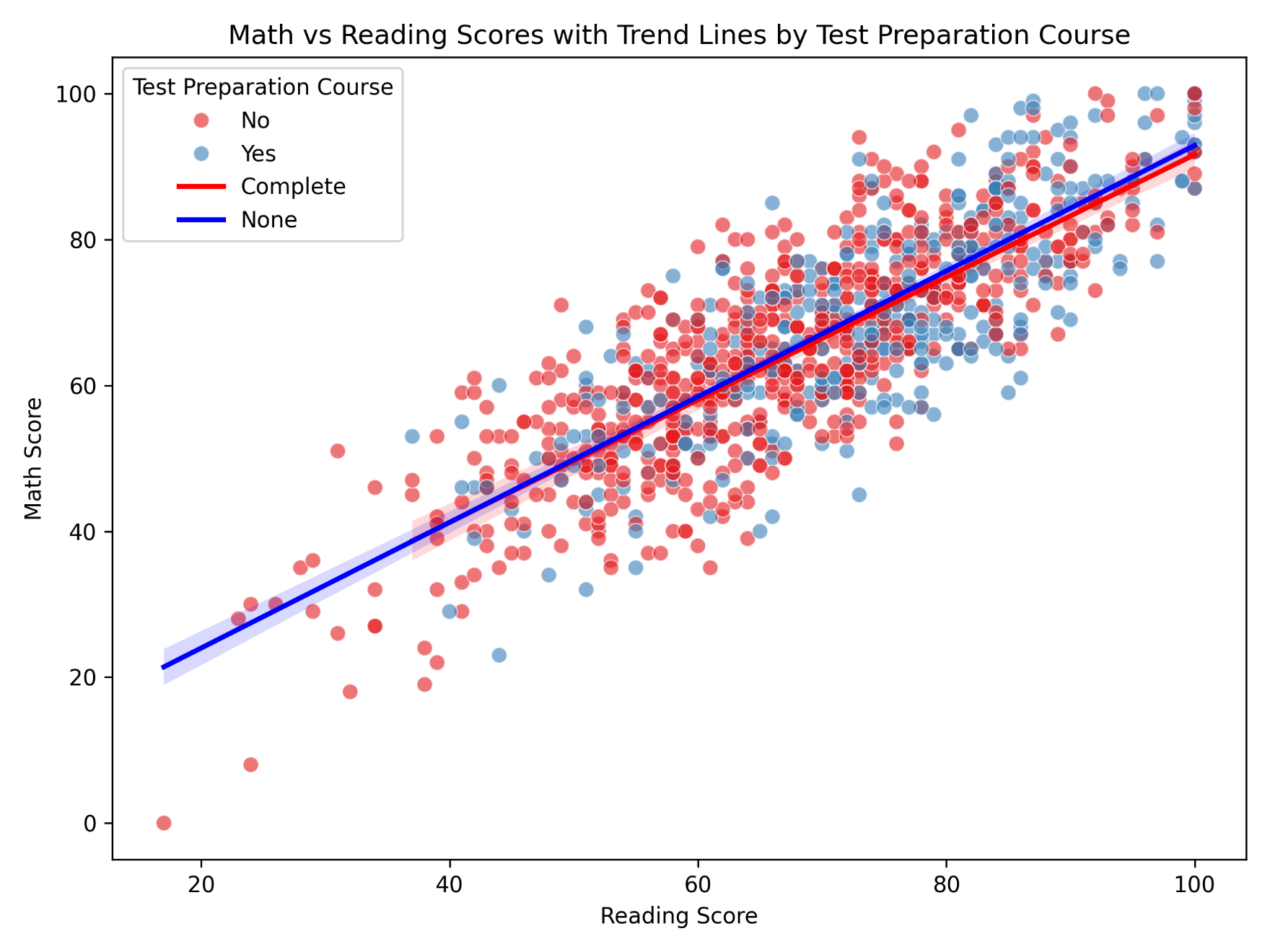
* **3.3. V3 — Lunch Type and Average Academic Performance**
  + **Question:** Does lunch type (standard vs. free/reduced) relate to overall academic outcomes?
  + **Interpretation:** Yes, the lunch type (standard labeled as Yes vs free/reduced labeled as No) is strongly related to overall academic outcomes. The big gap of almost 10 points between two groups indicates a clear and significant relationship between lunch type and academic performance, suggesting that students with access to standard lunch tend to have better academic outcomes.
  + **Figure 3**:



* **3.4. V4 — Correlation Between Subject Scores**
  + **Question:** How strongly do the three subject scores (math, reading, writing) move together?
  + **Interpretation:** Based on the correlation heat map, all three subject scores have strong positive correlations with each other. This means that as a student’s score in one subject increases, their scores in the other subjects tend to increase as well. Especially in reading and writing with correlation of .95, this is the strongest relationship compared to the other two (Math and Reading with the correlation of 0.82, and Math and Writing with the correlation of 0.85)
  + **Figure 4:**



* **3.5. V5 — Math vs. Reading Relationship by Test Preparation**
  + **Question:** How strongly are math and reading scores associated, and does the test-preparation course alter this relationship?
  + **Interpretation:** The test preparation course does not significantly alter the fundamental relationship between math and reading scores. There is only a slight increase in math performance. This suggests that for any given reading score, student who completed the test preparation course tend to have a slightly higher math score on average.
  + **Figure 5:**

****

**4. Conclusion**

* **4.1. Summary of Findings:** This analysis revealed several key insights into the student performance data. Factors such as gender, completion of a test preparation course, and lunch type all show measurable relationships with academic scores. Furthermore, the strong positive correlations between subjects suggest that student performance is often consistent across different academic areas.
* **4.2. Further Research:** Future work could explore the interaction effects between these variables or use this dataset to build a predictive model for student success.

**5. Appendix: Python Implementation**

https://raw.githubusercontent.com/mosomo82/COMP\_SCI\_5530/refs/heads/main/Assignment/Assignment\_1/Q2\_Data\_Visualization/src/Assignemnt\_1\_Data\_Visualization.ipynb