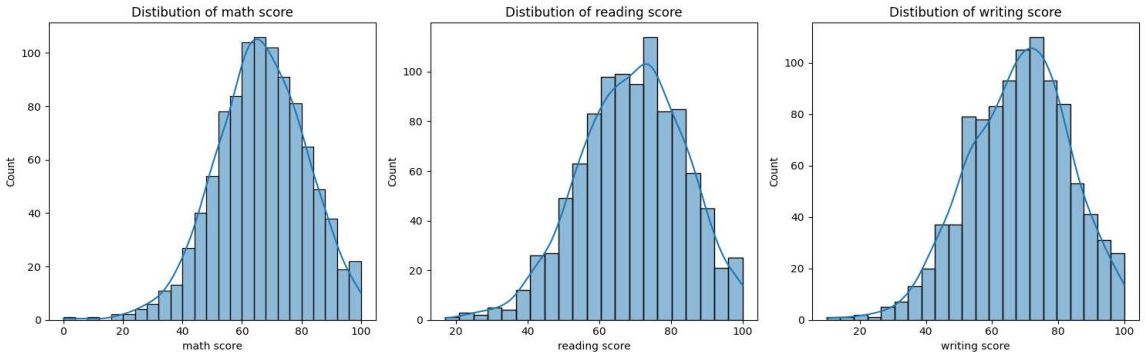
Exploratory Data Analysis Observations

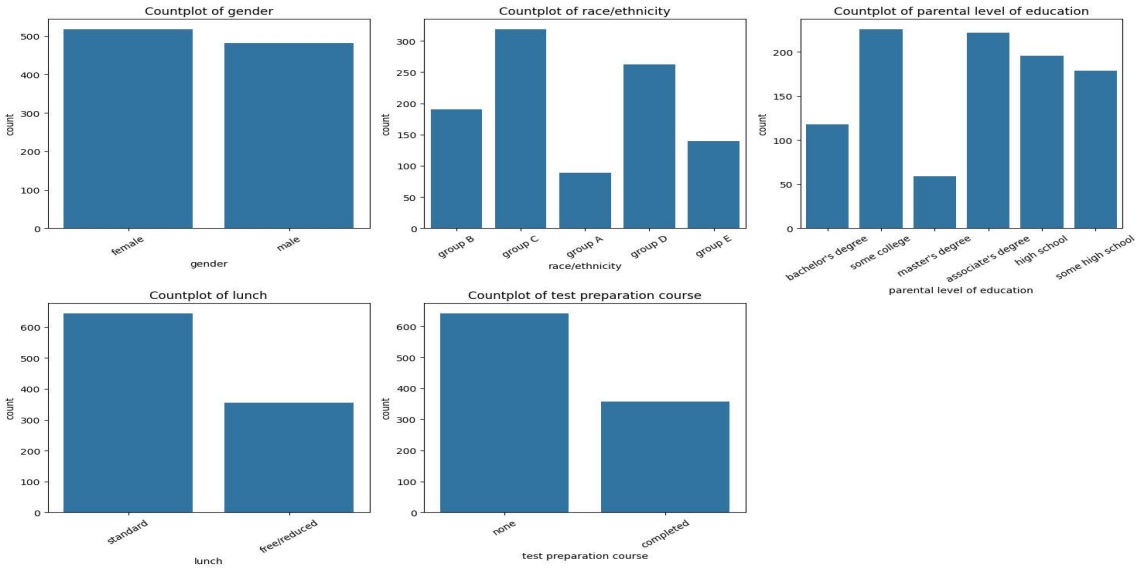
Basic Info

- The dataset contains 1000 rows and 8 columns.
- No missing data detected.
- The numerical columns include Math, Reading, and Writing scores, which range from 0 to 100.
- The categorical columns include Gender, Race/Ethnicity, Parental Education, Lunch, and Test Preparation.

Univariate Analysis (Numerical Columns)

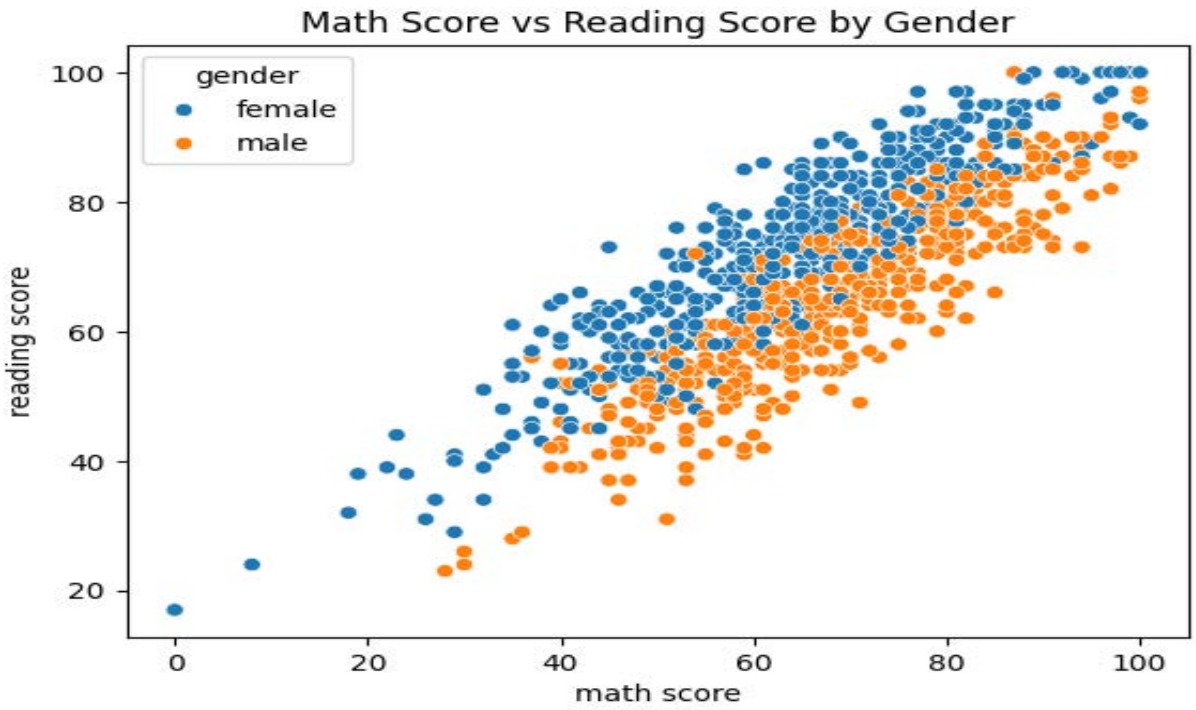


1. **Math Score:** The distribution is slightly left-skewed, with the majority of students scoring between 60 and 90. A few students scored below 30, showing potential outliers.
2. **Reading Score:** The distribution is fairly normal with most students scoring between 60 and 90.There are no significant outliers.
3. **Writing Score:** Similar to Reading, the writing scores show a slight right skew with most students scoring between 60 and 85.



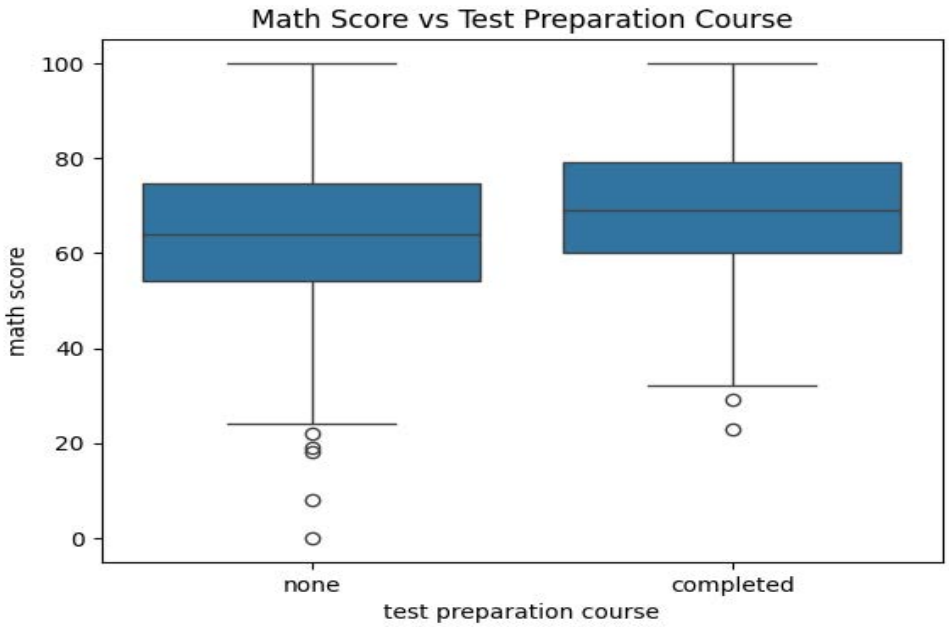
1. **Gender**: Slightly more female students than male students.
2. **Race/Ethnicity**: Group C is the most common, accounting for nearly 30% of the dataset. Groups A and E are the least frequent.
3. **Parental Level of Education**: Most students have parents with "Some College" or "Associates Degree" education level.
4. **Lunch**: The majority of students have standard lunch.
5. **Test Preparation Course**: More students have completed the test preparation course compared to those who have not completed the course.

Bivariate Analysis (Math vs Reading Score Scatterplot)



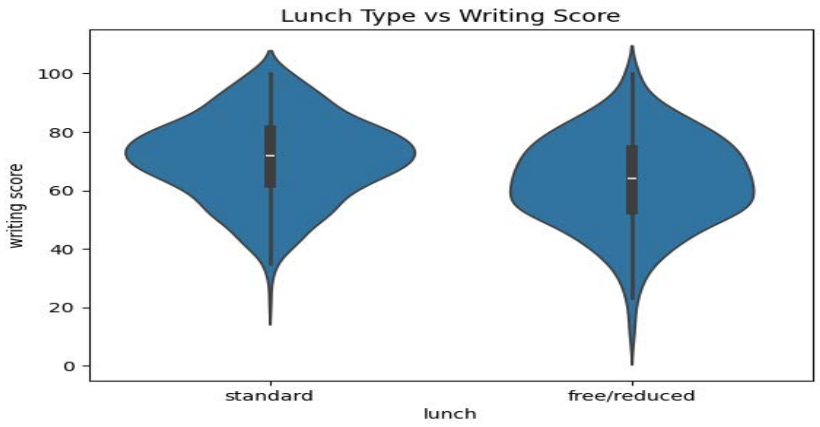
1. A strong positive relationship exists between math and reading scores. As math scores increase, reading scores tend to increase as well.
2. Female students tend to score slightly higher in reading, while male students show a slightly higher math score in general.

Bivariate Analysis (Test Preparation vs Math Score Boxplot)



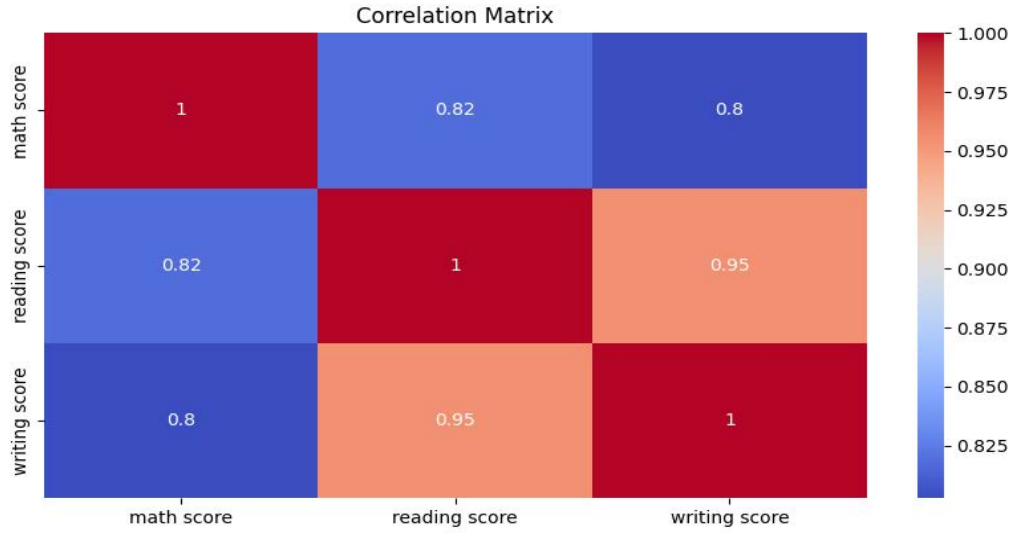
1. Students who completed the test preparation course tend to have higher math scores compared to those who did not complete the course.
2. The median math score for students who completed the test prep course is noticeably higher.

Bivariate Analysis (Lunch vs Writing Score Violinplot)



1) Students with standard lunch tend to perform better in writing compared to those with free/reduced lunch.
2) The writing scores of free/reduced lunch students have a wider spread, with some students scoring very low.

Correlation Heatmap



1) There is a strong positive correlation between math, reading, and writing scores.
2) The correlation values are above 0.8, indicating a high linear relationship.
3) No significant negative correlations are found among the features in this dataset.

Final Summary

1) Test preparation and lunch type are key factors influencing student performance, with those who completed test prep showing better scores.
2) Strong positive relationships exist between the scores of the three subjects, and females tend to outperform males in reading and writing.
3) Race/ethnicity and parental education level also show some distribution patterns but have a smaller impact on scores compared to test preparation and lunch.