Assignment 1
Principles of Data Science

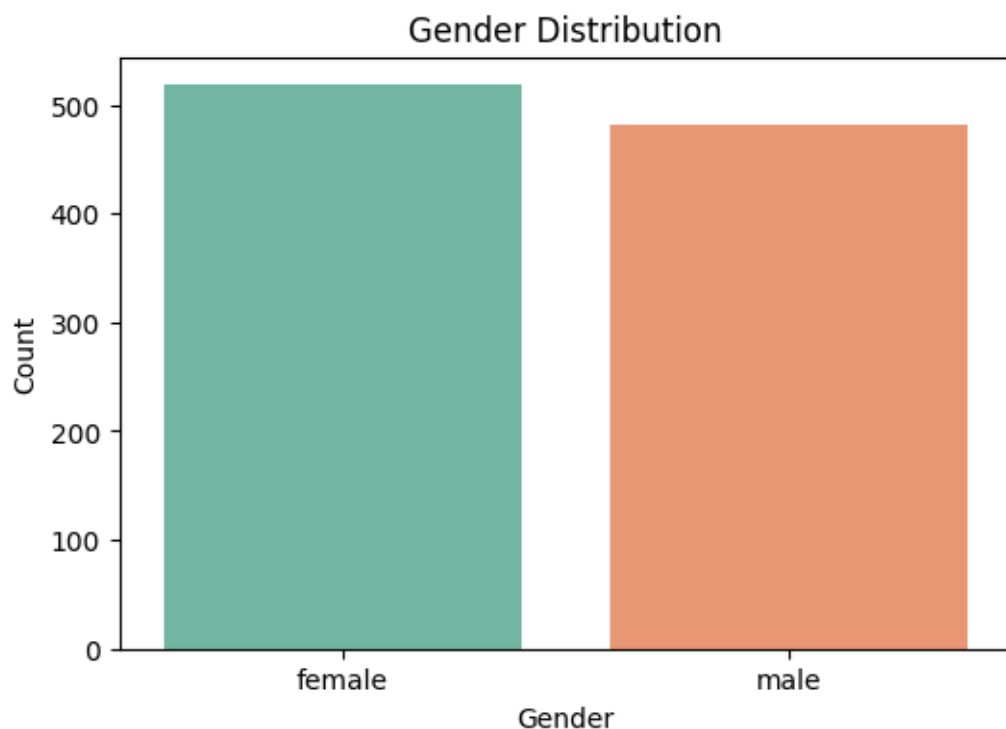**Name :** Venkata Sai Veyyagandla

Student Id : 16352979

1. Average Scores by Gender

Visualization: Bar plot of average math, reading, and writing scores by gender.

Purpose: This plot highlights the average performance of male and female students across different subjects.

Analysis: The bar plot allows for an easy comparison between genders across math, reading, and writing. By visualizing the mean scores, we can quickly identify if one gender tends to outperform the other in specific subjects. For example, if females consistently score higher in reading and writing, this could suggest gender-based trends in language-related subjects, while males might excel in math.

2. Distribution of Math Scores by Parental Level of Education

Visualization: Box plot of math scores categorized by parental level of education.

Purpose: This plot illustrates how math scores are distributed based on parental education levels.

Analysis: The box plot helps in identifying variations in student performance across different parental education levels. It provides insights into the median score, spread, and potential outliers. For instance, if students with parents holding higher degrees (e.g., master's) tend to have higher median math scores, this could indicate a potential influence of parental education on academic performance in math.


Math vs Reading Scores
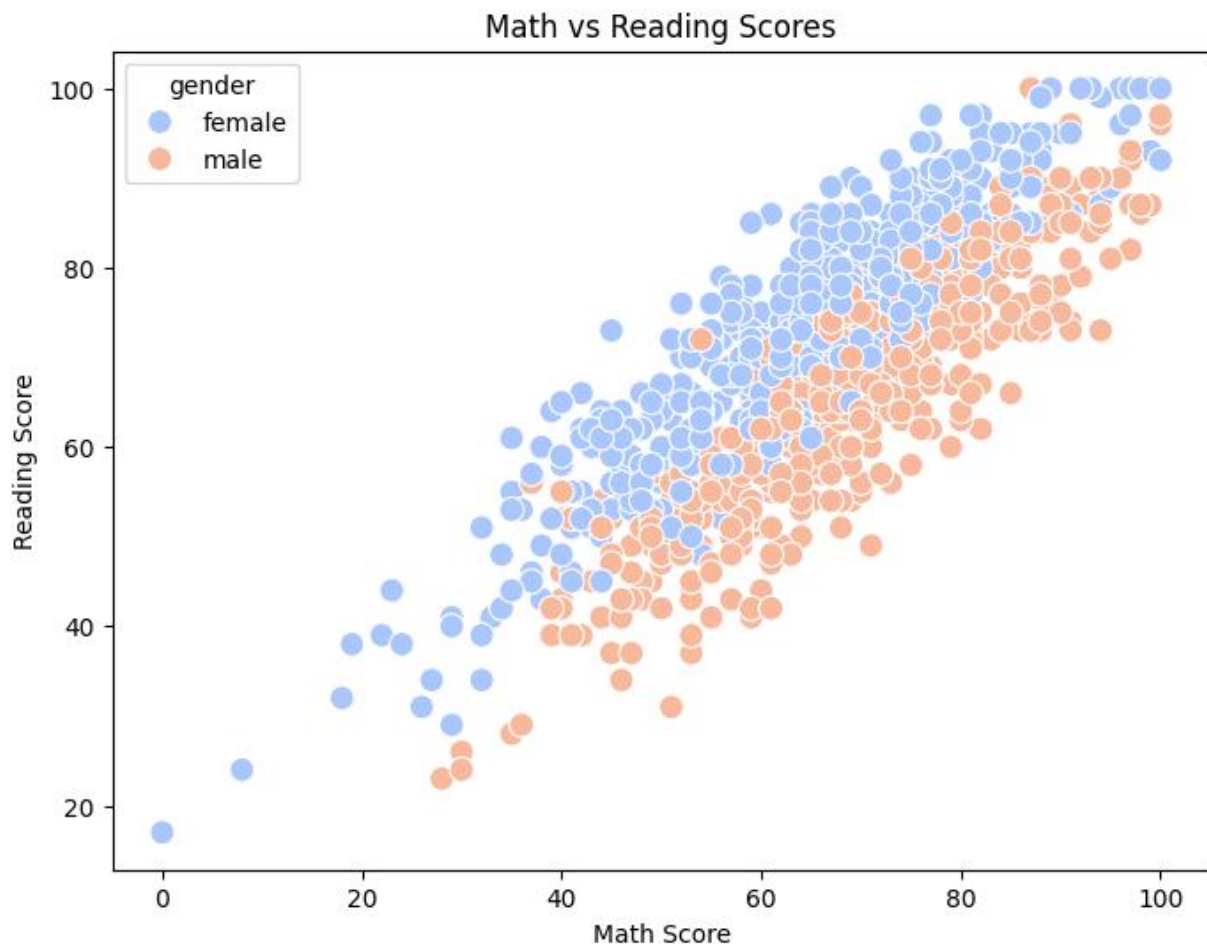
3.Distribution of Parental Level of Education

Visualization: Histogram of parental level of education.

Purpose: This histogram provides a visual representation of the frequency distribution of parental education levels in the dataset.

Analysis: The plot helps to understand how parental education is distributed among the students. The height of each bar shows the number of students whose parents fall under a specific educational category (e.g., "bachelor's degree", "some college"). The inclusion of a KDE (Kernel Density Estimate) line helps smooth out the distribution, showing trends more clearly. This visualization can reveal which parental education levels are most common in the dataset and could provide insights into potential socioeconomic influences on student performance.
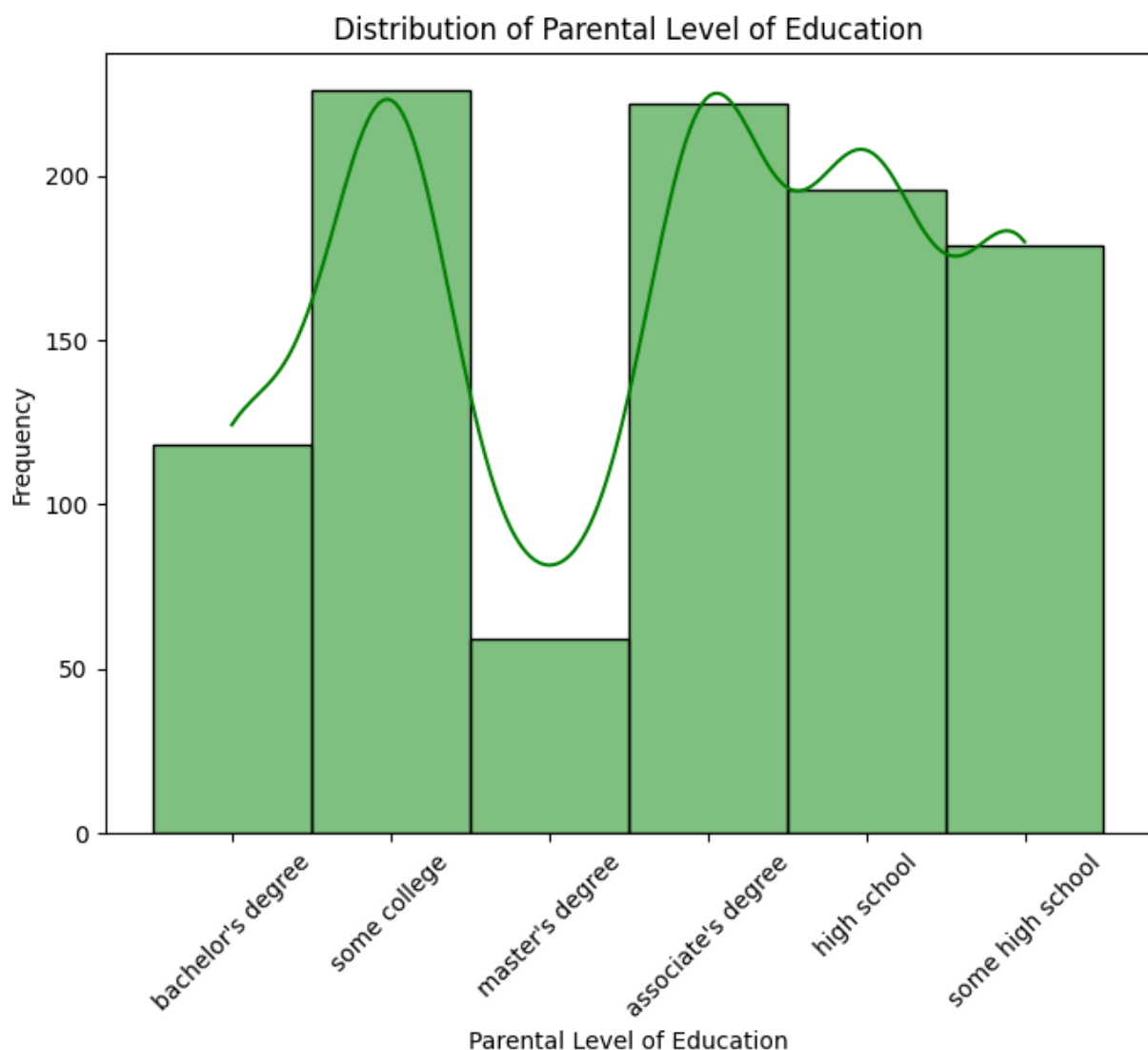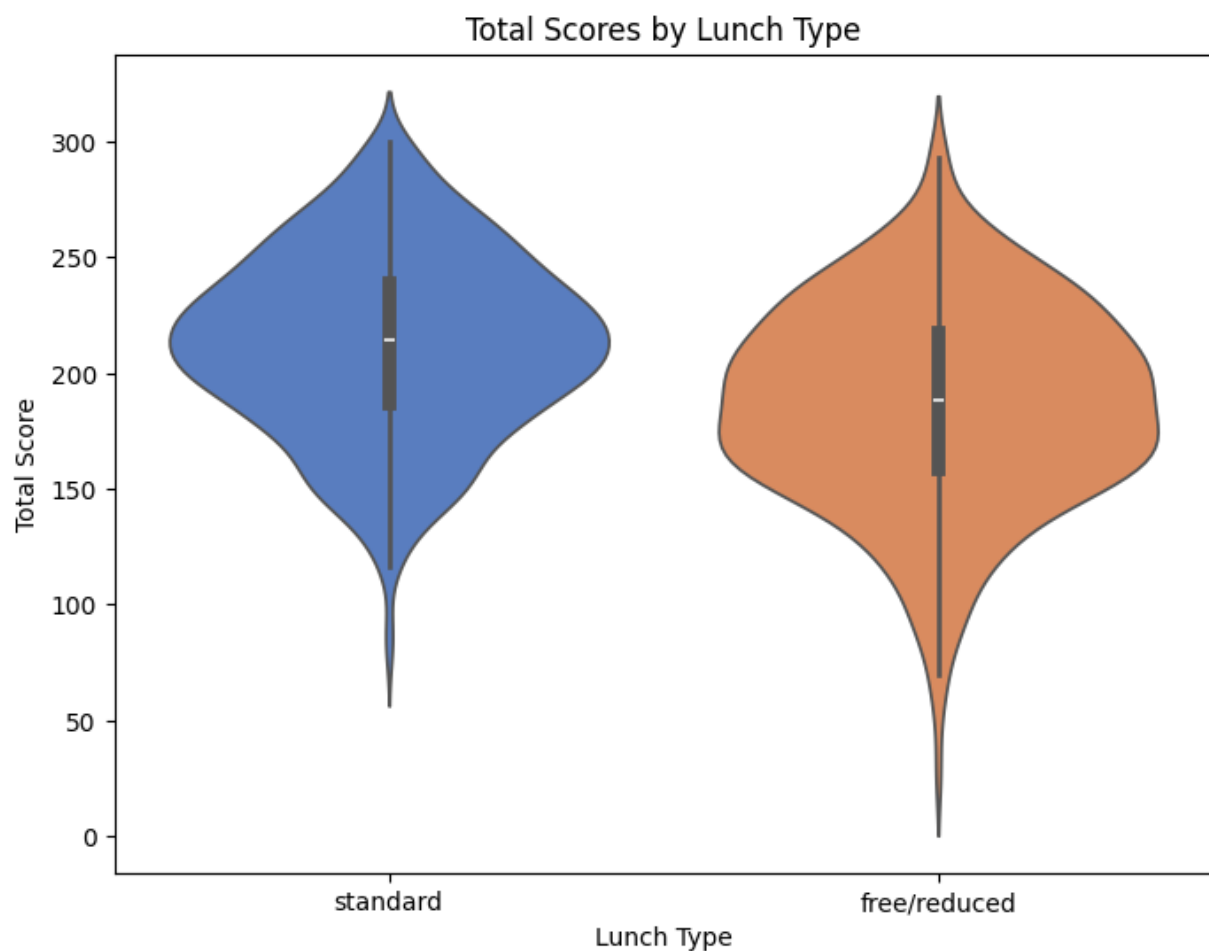
4. Total Scores by Lunch Type
Visualization: Violin plot showing total scores categorized by lunch type.
Purpose: This violin plot illustrates the distribution of total scores for students based on their lunch type (standard vs. free/reduced).
Analysis: The plot not only displays the median total score for each lunch type but also provides insights into the distribution of scores within each category. The width of the violin indicates the density of scores at different levels, showing where the majority of students' total scores fall. For example, if students on standard lunch tend to have higher total scores compared to those on free/reduced lunch, this may suggest that lunch type is associated with academic performance. This visualization helps to highlight potential disparities in student achievement based on socioeconomic factors.

5. Distribution of Math Scores by Parental Level of Education

Visualization: Box plot of math scores categorized by parental level of education.

Purpose: This plot illustrates how math scores are distributed based on parental education levels.

Analysis: The box plot helps in identifying variations in student performance across different parental education levels. It provides insights into the median score, spread, and potential outliers. For instance, if students with parents holding higher degrees (e.g., master's) tend to have higher median math scores, this could indicate a potential influence of parental education on academic performance in math.

Math Scores by Gender