**Assignment-1**

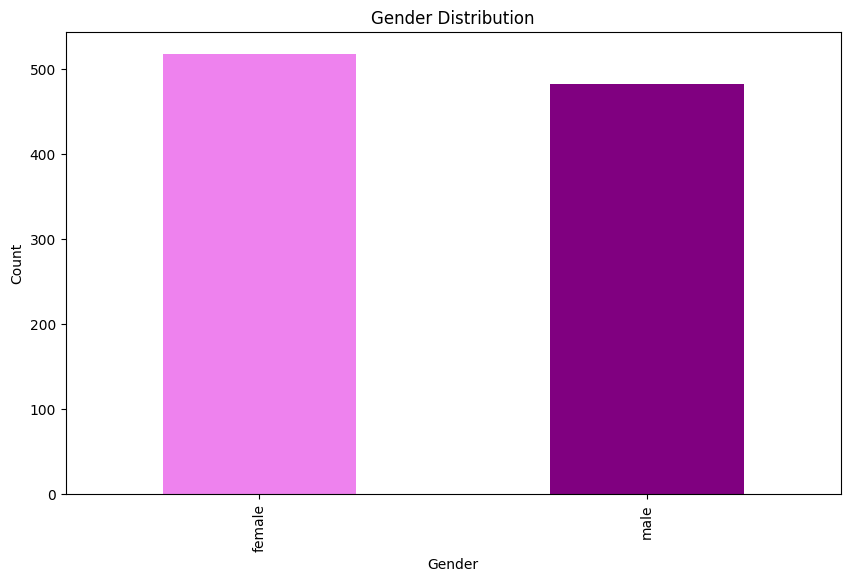
**Data Analysis of Student Performance Dataset Results**

**Vedanth goud nagapola**

**16354630**

**🡪BAR PLOT:**

The gender distribution is depicted by the bar plot, which counts the number of male and female students. It facilitates the detection of any disparity in the dataset's gender representation. The lack of further gender categories suggests that they are excluded or categorized as "Male" and "Female."



The specific data the graph shows is the number of people by gender. It appears to show that there are more females than males. However, it is important to note that the graph does not show the total number of people, so it is impossible to say for sure what proportion of the population is female or male. Additionally, the graph does not show any information about other genders, such as non-binary or transgender people.

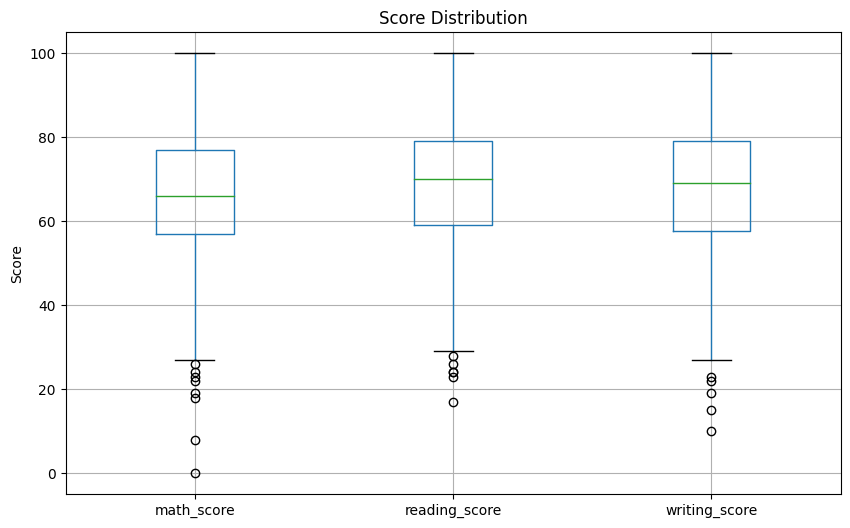
🡪**Box plot:**

The visualization technique used in the image is a box plot, which is a way to summarize the distribution of numerical data. It shows the following:

The median: The horizontal line in the middle of the box represents the median, which is the middle value when the data is ordered from least to greatest.

The interquartile range (IQR): The box encompasses the middle 50% of the data, with the bottom and top edges of the box corresponding to the first and third quartiles, respectively. The IQR is the distance between these two quartiles.

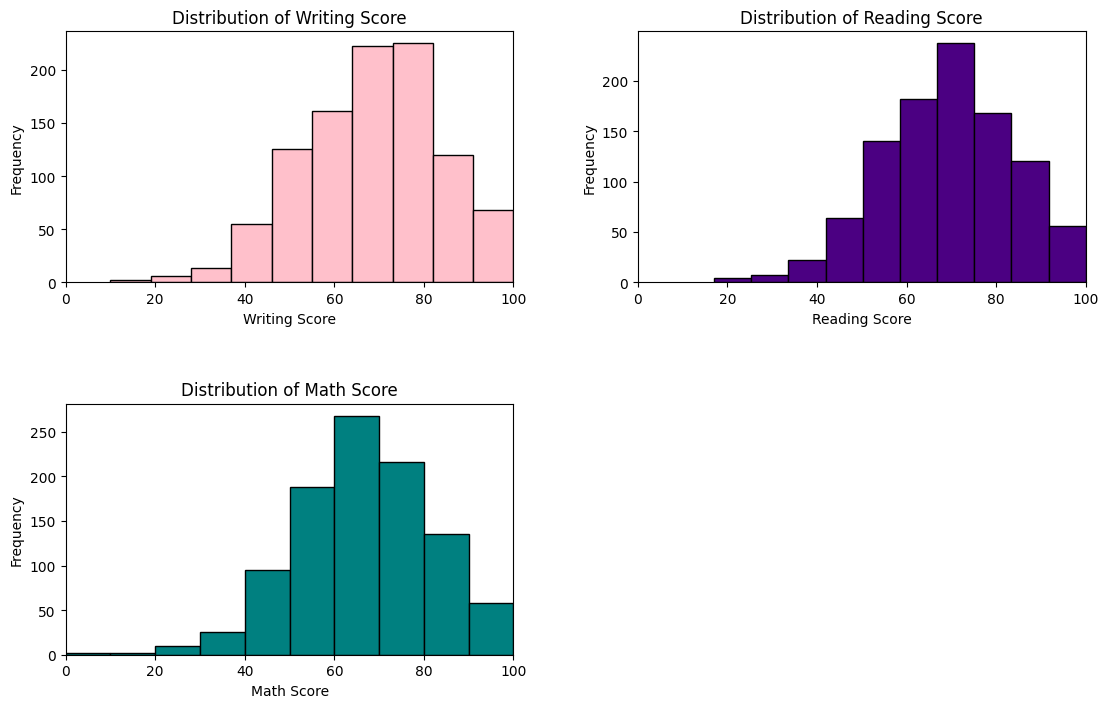
The whiskers: The lines extending from the box show the range of the remaining data, excluding outliers. Any data points that fall outside the whiskers are considered outliers and are plotted individually as circles.



Overall, the box plots provide a quick and easy way to compare the distributions of scores for different subjects. They show that students tend to score higher in math than in reading or writing, and that there is more variability in math scores than in reading or writing scores.

🡪**Histogram**

The visualization technique used in the image is a histogram. Histograms are used to represent the distribution of continuous data. They do this by dividing the range of the data into intervals (bins) and then counting the number of data points that fall into each bin. The resulting bars show the frequency of each interval.



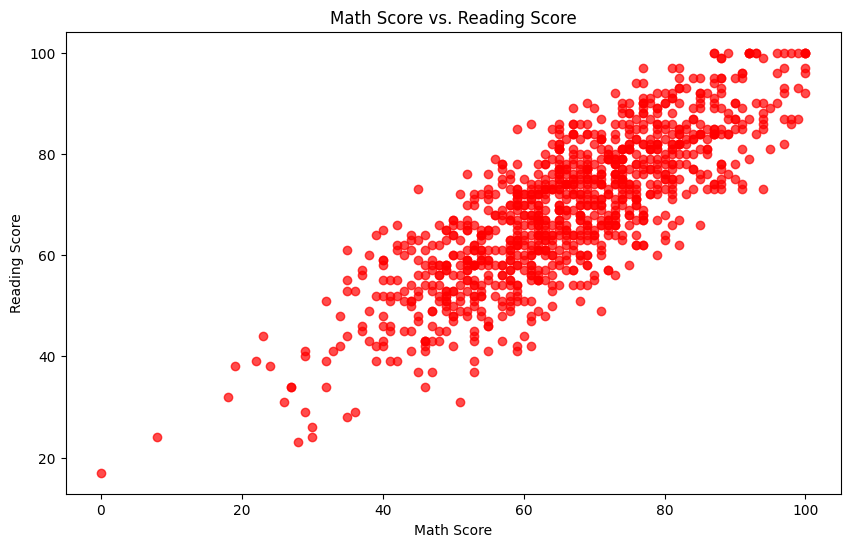
The specific data the graph shows is the distribution of reading, writing and math scores. It appears that the scores are:

* Normally distributed for all three subjects, reading, writing and math, as the bars resemble a bell curve.
* Writing scores have the widest distribution, as the bars are shorter and wider than those for reading and math. This means that writing scores are more spread out, with a greater range of scores achieved.
* Math scores have the narrowest distribution, as the bars are taller and narrower than those for reading and writing. This means that math scores are more concentrated, with a smaller range of scores achieved.

Overall, the visualization technique used in the image is effective in showing the distribution of reading, writing and math scores. However, it is important to be aware of the limitations of the data and the graph itself.

🡪**Scatter plot:**

A scatter plot is a type of graph that displays the relationship between two numerical variables. In the image, the two variables are math score and reading score. Each dot in the scatter plot represents a single student's scores on these two metrics..



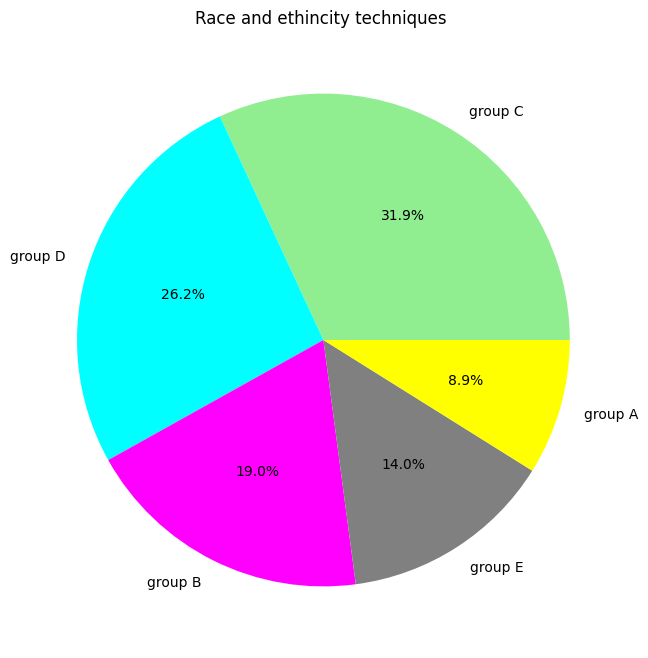
From the above scatter plot graph we can conclude the following points:

* There is a positive correlation between math score and reading score. This means that students who tend to score higher in math also tend to score higher in reading, and vice versa. This is shown by the general upward trend of the dots in the scatter plot.
* The correlation is not perfect. There are some students who score high in one subject but low in the other, as shown by the dots that are scattered away from the upward trend.
* There is a wider range of scores in math than in reading. This is shown by the wider spread of dots on the horizontal axis (math score) compared to the vertical axis (reading score).
* There may be some outliers. There are a few dots that are far away from the main trend of the data. These could be students who performed exceptionally well or poorly in one subject compared to the other.

Overall, the scatter plot provides a useful way to visualize the relationship between math score and reading score. It shows that there is a positive correlation between the two variables, but it is not a perfect correlation. There is also a wider range of scores in math than in reading, and there may be some outliers.

🡪**Pie chart:**

Pie charts are used to represent the distribution of categorical data, where each slice of the pie represents a different category and the size of the slice is proportional to the number of data points that fall into that category.



The specific data the pie chart shows is the percentage of people from five different racial or ethnic groups. It appears to show that:

* The largest group is group C, which makes up 31.9% of the population.
* Group D is the second largest group, making up 26.2% of the population.
* Group A is the third largest group, making up 19.0% of the population.
* Group B makes up 14.0% of the population.
* Group E makes up the smallest group, making up 8.9% of the population.

Overall, the pie chart provides a quick and easy way to see the relative proportions of different racial or ethnic groups in a population. However, it is important to be aware of the limitations of the data and the pie chart itself.