Assignment - 2
Principles of Data Science

Name: Surya Prakash Bandam
Student Id: 16343582

Question 2

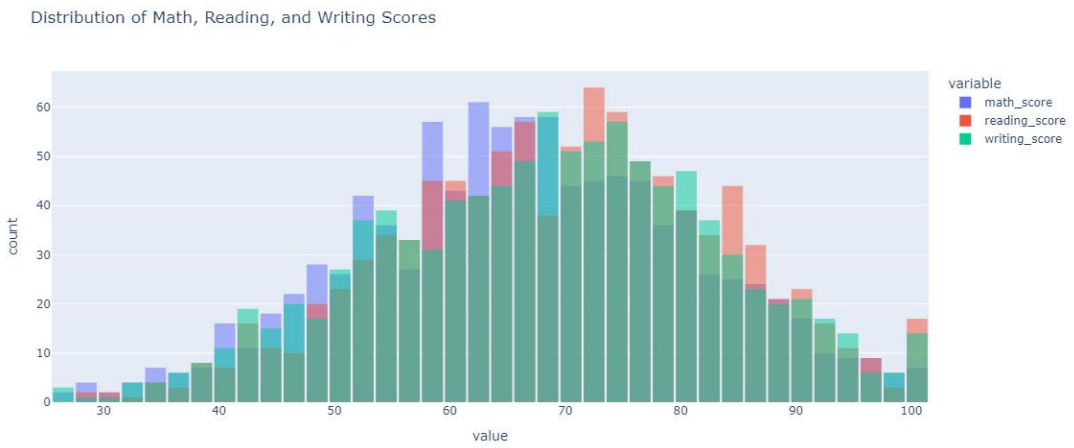
Data Visualization on Students Performance Data Set

First 10 rows of the data set

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50

Description and Interpretation of the Visualizations

1. Histogram Plot: Distribution of Scores Using Histogram



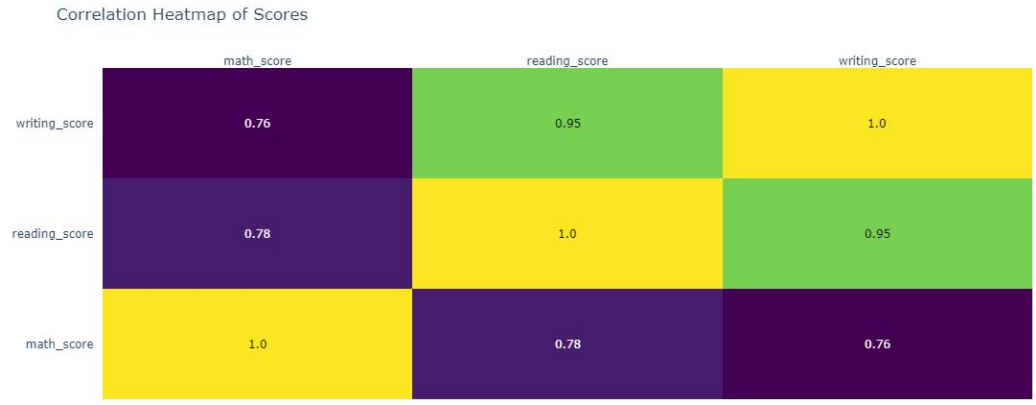
- To show the frequency distribution of student scores across different subjects (math, reading, and writing). It provides a clear visual representation of how scores are spread out across different ranges.
- Makes it simple to see where most students' scores fall, identifying common performance levels and overall trends.

- Allows for a quick comparison of how students perform across different subjects, helping to pinpoint areas of strength or weakness.

2. Box Plot:For Score Distribution by Gender

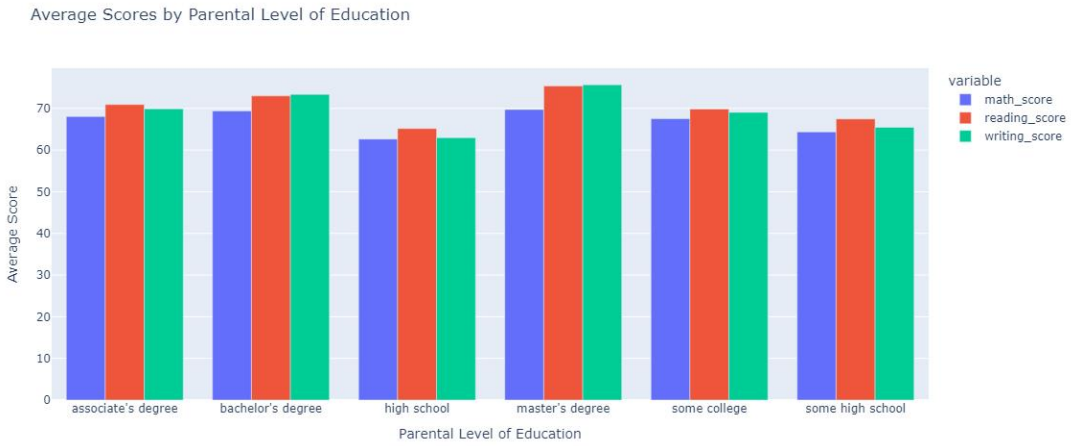
- To compare the distribution of scores between male and female students. It highlights the median, quartiles, and spread of scores within each gender group.
- Facilitates a direct comparison between male and female students' performance, showing differences in median scores and variability.
- Shows the range and spread of scores, helping to understand performance variability within each gender group and detect any systematic differences in performance.

3. Correlation Heatmap



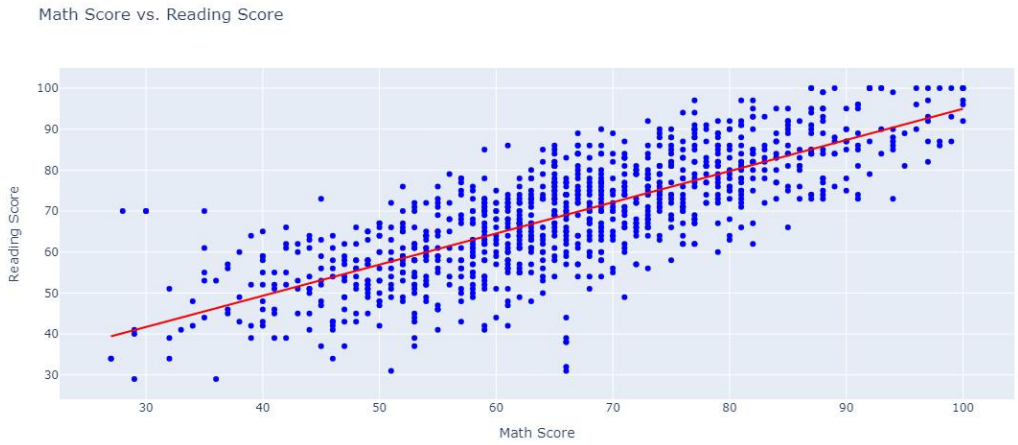
- To display the correlation coefficients between different subjects (math, reading, and writing scores), showing the strength and direction of relationships between them.
- Visualizes how strongly different subjects are related to each other, making it easier to see whether good performance in one subject predicts good performance in another.

4. Bar Plot: Average Scores by Parental Education Level



- To compare the average scores of students in different subjects based on their parental education level, illustrating the impact of parental education on student performance.
- Clearly shows how a categorical variable (parental education) affects student performance, providing insights into socio-economic influences on education.
- Simplifies the identification of patterns, such as whether higher parental education correlates with higher student scores.

5. **Scatter Plot:**



- To show the relationship between math and reading scores, with a regression line indicating the overall trend and correlation between the two subjects.
- Helps identify students who do not follow the general trend, such as those who excel in one subject but not the other, indicating unique cases worth further investigation.