Assignment1(Group of two)
CS160
Introduction to Data Science
SP2024

Working on Techniques for Analyzing Data

Instructions: Complete the following activities for this project.

1. Create a new GitHub repository named Assignment1_XXX, where XXX are your initials.
2. Using excel (to generate the result) and word documents (type answers and paste the results) work on the following questions and submit your work using **pdf** format.

Description:

This dataset contains information about exam scores of a group of students. It includes attributes such as student ID, gender, age, subject, exam score, and study hours.

Attributes:

Student ID: A unique identifier for each student.

Gender: The gender of the student (male or female).

Age: The age of the student.

Subject: The subject of the exam (e.g., Math, Science, English).

Exam Score: The score achieved by the student in the exam.

Study Hours: The number of hours the student studied for the exam.

Objective:

Perform a descriptive analysis of the student exam scores to understand factors affecting performance and identify trends.

A. **Summary Statistics:** Calculate summary statistics for exam scores and study hours (mean, median, standard deviation, etc.).

Table 1

Exam Score		*Study Hours*	
Mean	85.01111111	Mean	4.466667
Median	86	Median	4
Mode	88	Mode	4
Standard Deviation	6.896497148	Standard Deviation	1.143619
Sample Variance	47.56167291	Sample Variance	1.307865
Range	27	Range	4
Minimum	70	Minimum	2
Maximum	97	Maximum	6
IQR	10	IQR	1.75

For **Exam Score** the mean is 85.01111111, the median is 86, the mode is 88, the standard deviation is 6.896497148, the variance is 47.56167291, the range is 27, the minimum is 70, the maximum is 97, the IQR is 10.

For **Study Hours** the mean is 4.466667, the median is 4, the mode is 4, the standard deviation is 1.143619, the variance is 1.307865, the range is 4, the minimum is 2, the maximum is 6, the IQR is 1.75.

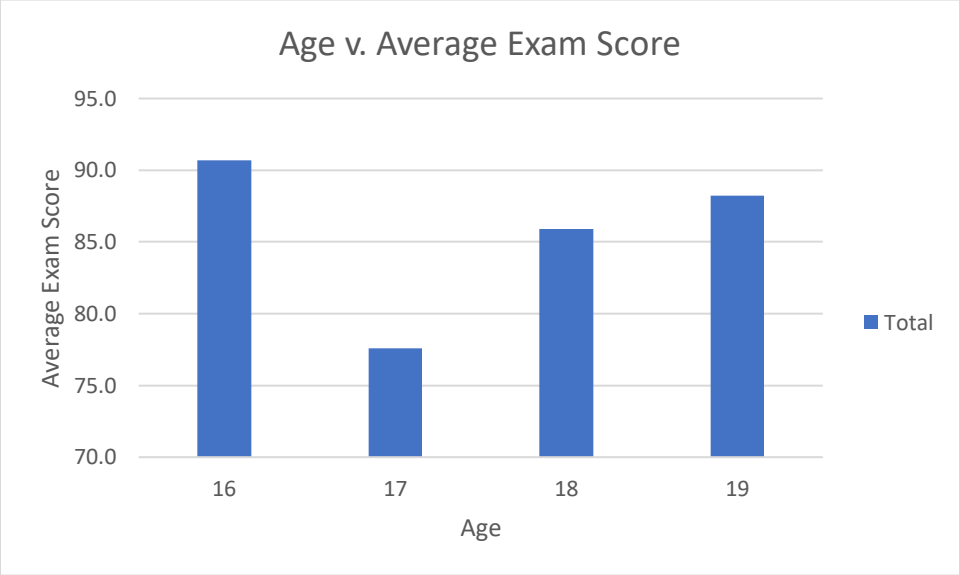
B. **Gender Analysis:** Compare average exam scores and study hours for male and female students using PivotTables or simple calculations.

Row Labels	Average of Exam Score	Average of Study Hours
Female	89	5
Male	81	4
Grand Total	**85**	**4**

On average, the females studied more than the males with an average time of 5 hours compared to the males of 4 hours. On average, the female's exam scores were higher than the males with an average exam score of 89 compared to the male's exam score of 81.

C. **Age Analysis:** Analyze how exam scores vary with age using scatter plots or trend lines.

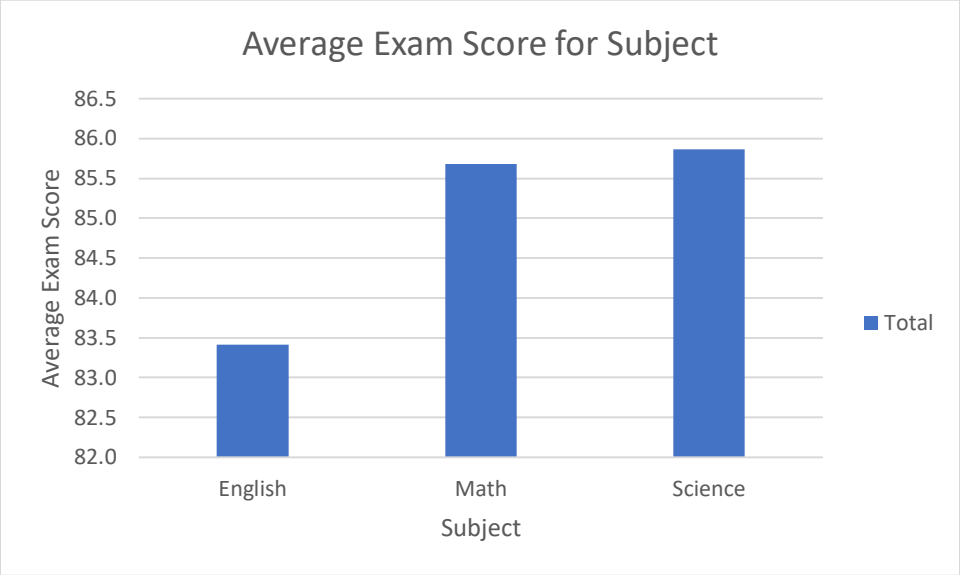
Age	Average of Exam Score
16	90.7
17	77.6
18	85.9
19	88.2
Grand Total	**85.0**



16-year-olds had the highest average exam scores with 90.7. 17-year-olds had the lowest average exam scores with 77.6. 18-year-old had an average exam score of 85.9 putting them in 3rd place. 19-year-olds had the second highest average exam scores with 88.2.

D. **Subject Analysis:** Explore average scores for each subject to identify strengths and weaknesses.

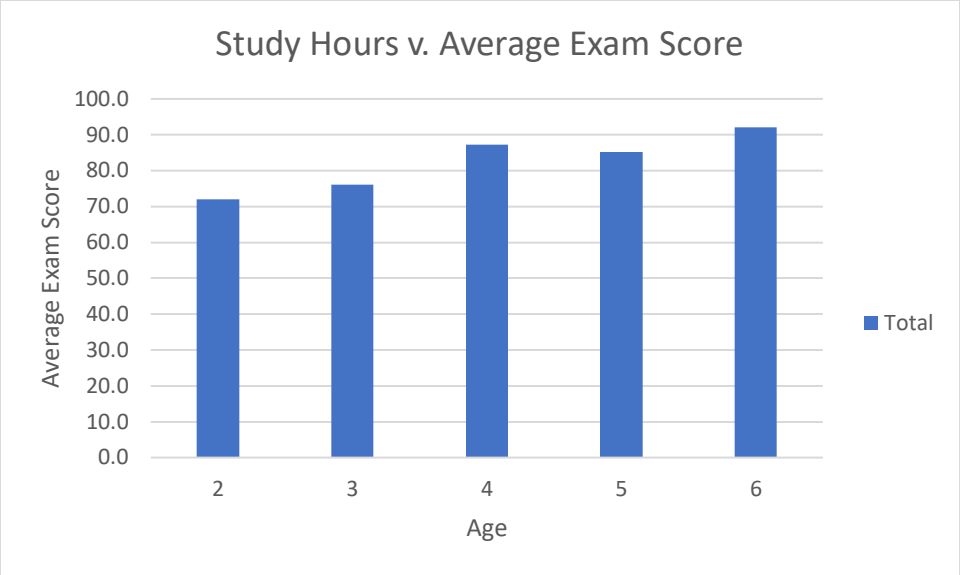
Row Labels	Average of Exam Score
English	83.4
Math	85.7
Science	85.9
Grand Total	**85.0**



The students had a subject strength in science with an average exam score of 85.9. The next highest subject was pretty much tied with science with an average exam score of 85.7. English had the lowest average exam score with a score of 83.4.

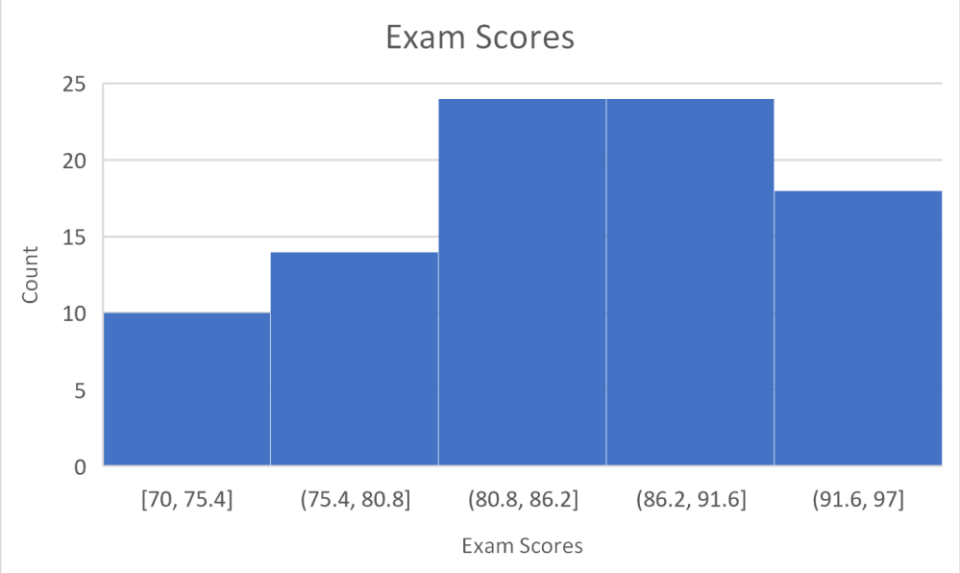
E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.

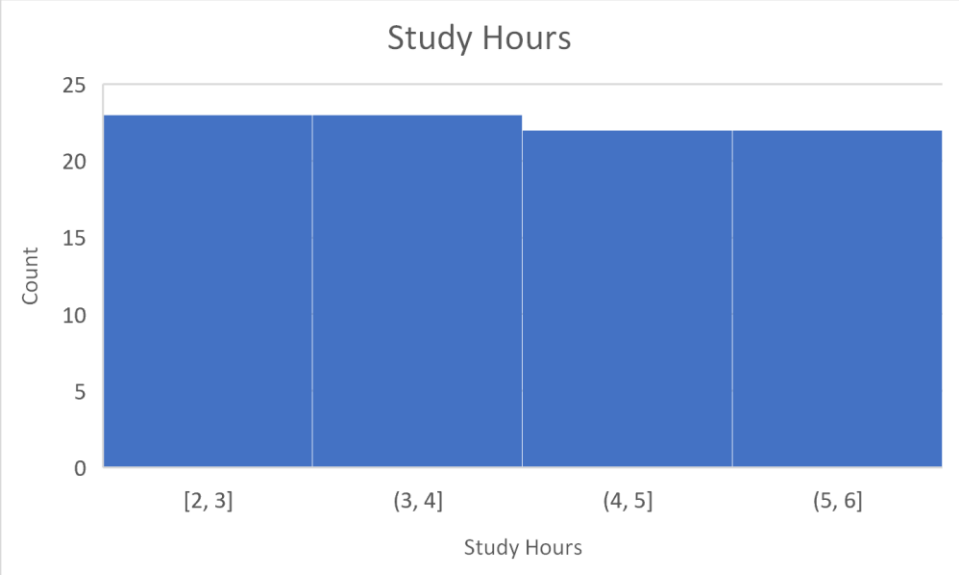
Study Hours	Average of Exam Score
2	72.0
3	76.0
4	87.2
5	85.3
6	92.0
Grand Total	**85.0**



As study hours increased so did average exam scores. A student that studied 2 hours got an average exam score of 72 while the student that studied 6 hours got an average exam score of 92.

F. **Distribution Analysis:** Create histograms to show the distribution of exam scores and study hours.





Study hours were distributed relatively evenly, while exam scores were more concentrated between 80.8 and 91.6.

G. **Top Performers:** Identify students with the highest scores and analyze their study hours, gender, and age.

Student ID	Gender	Age	Subject	Exam Score	Study Hours
90	Female	18	Science	97	6
8	Female	16	Science	96	6
18	Female	18	Science	96	6
4	Female	16	Math	95	6
38	Female	19	Math	95	6

Gender	Average of Age	Average of Exam Score	Average of Study Hours
Female	17	95.8	6
Grand Total	**17**	**95.8**	**6**

The top 5 students were all female. They had an average age of 17, with a minimum age of 16 and a maximum age of 19. They had an average exam score of 95.8, with a minimum score of 95 and a maximum score of 97. They all studied an average of 6 hours.

H. **Correlation Analysis:** Calculate the correlation between study hours and exam scores to understand their relationship.

	Exam Score	*Study Hours*
Exam Score	1	
Study Hours	0.764357716	1

The correlation between Study hours and Exam scores is 0.764357716. This shows that there is a strong positive correlation between the number of study hours and exam scores. The more you studied the higher your exam scores would likely be.

3. Provide a summary result your findings.

   The majority of students received a B on their exams with an average exam score between 80.8 and 91.6 across all subjects. Females had higher average exam scores with a score of 89 compared to the males average exam score of 81. Females also studied for the exam for 5 hours compared to the males 4 hours. 16-years-olds had the highest average exam scores with 90.7, followed by 19-year-olds with an average exam score with 88.2, next 18-year-olds had an average exam score of 85.9, this was followed by the 17-year-olds with the lowest average exam scores of 77.6. Math and science average exam scores were higher than English with both subjects having a high 85. The average exam score for English was lower with an 83. The top 5 students were all female. They had an average age of 17. They had an average exam score of 95.8. They all studied an average of 6 hours. There was a strong positive (.76) correlation between exam scores and hours studied. The more you studied the higher your exam scores would likely be.

Using the instructions provided by GitHub, create a git repository named DS160**InClassAssignment**, and push your pdf file to it. Each of you needs to submit your work.

Submission:

Paste a link to your GitHub repository in the area provided for this assignment and submit it by class time.