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.).

Exam Score		*Study Hours*	
Mean	85.01111111	Mean	4.466667
Standard Error	0.726954629	Standard Error	0.120548
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

Sum	7651	Sum	402
Count	90	Count	90
IQR	10	IQR	1.75

Exam Score: The mean is 85, Standard Error is 0.73, Median is 86, mode is 88, SD is 6.89, Sample variance is 47.56, Range is 27, Min is 70, Max is 97, sum is 7651, count is 90, IQR is 10 (Refer to table above)

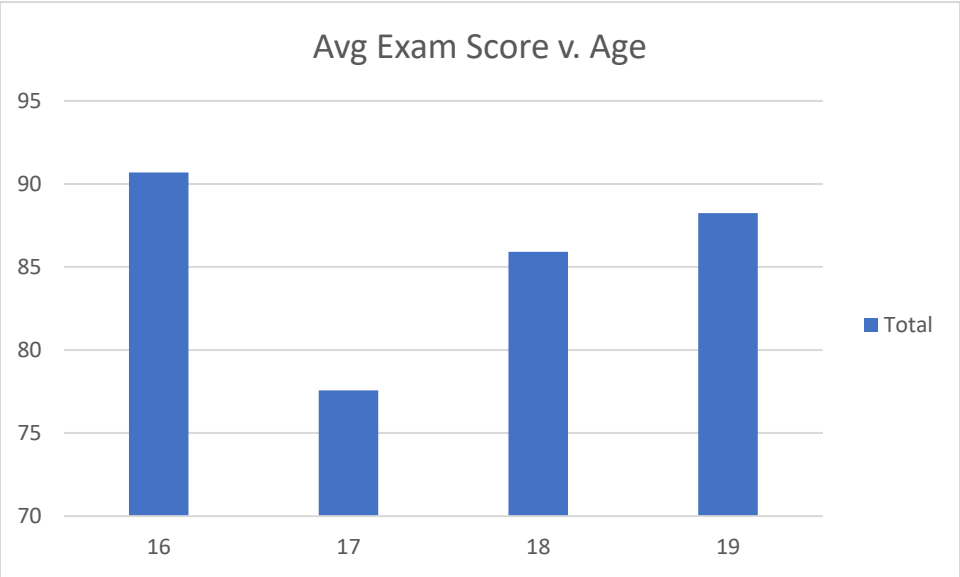
Study Hours: The mean is 4.46, SE is .12, median is 4, mode is 4, SD is 1.14, SV is 1.3, range is 4, min is 2, max is 6, sum is 402, count is 90, and IQR is 1.75 (Refer to table above)

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

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

Om average, Females studied more than males with an average time of 5 hours compared to the males 4 hours. Females also received better tests scores with an average of 89 (B+) compared to the male's score of 81 (B-)

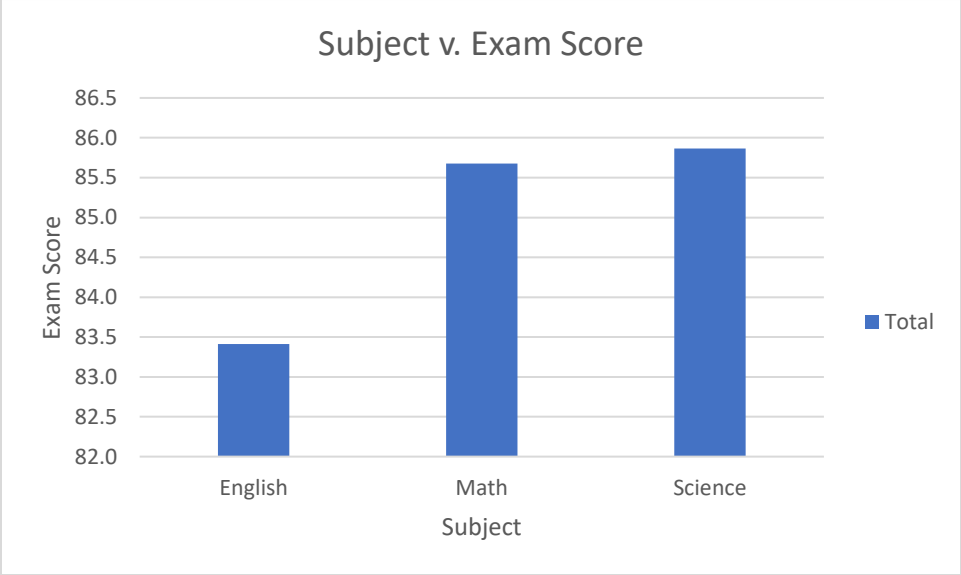
C. **Age Analysis:** Analyze how exam scores vary with age using scatter plots or trend lines.



16-year-olds tend to have the best test score, averaging an A-. Both 18 and 19 year old's average high B's on exams. However, 17-year-olds did the worst by far, averaging a C

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

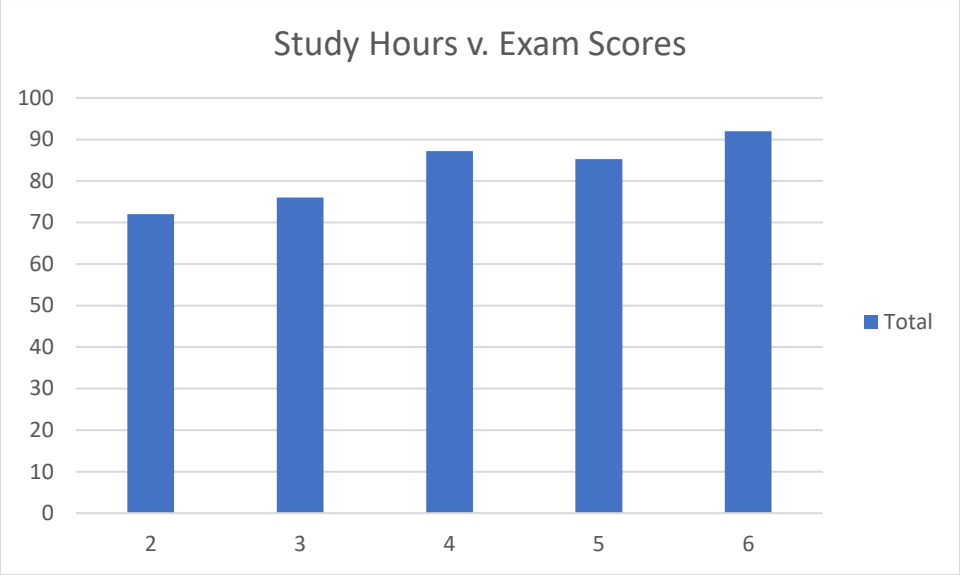
Row Labels	Average of Exam Score
English	83.4
Math	85.7
Science	85.9
Grand Total	**85.0**



The student's subject strength is Science with a high of 85.9. Followed by math as a close second with an average of 85.7. However, English is their weakest subject with an average of 83.4

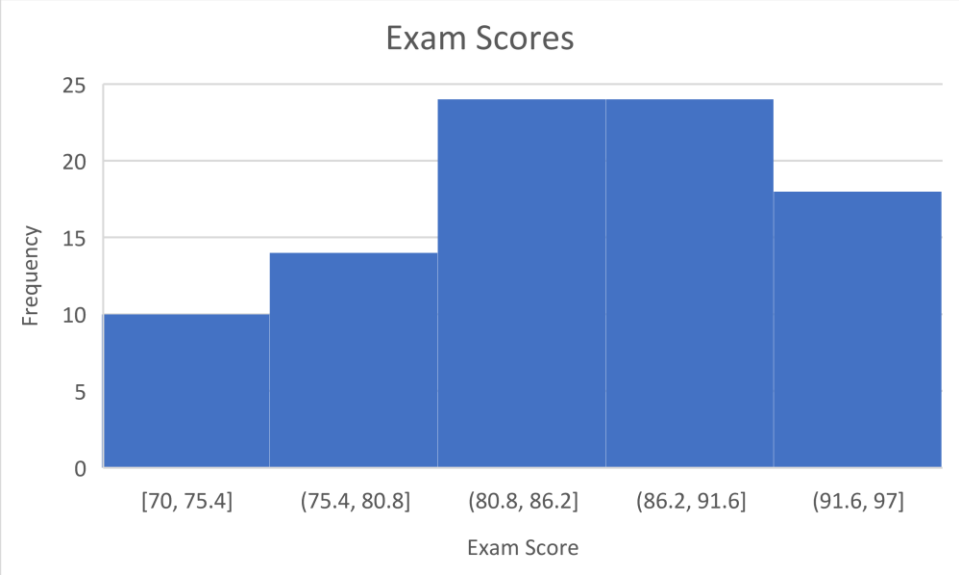
E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.

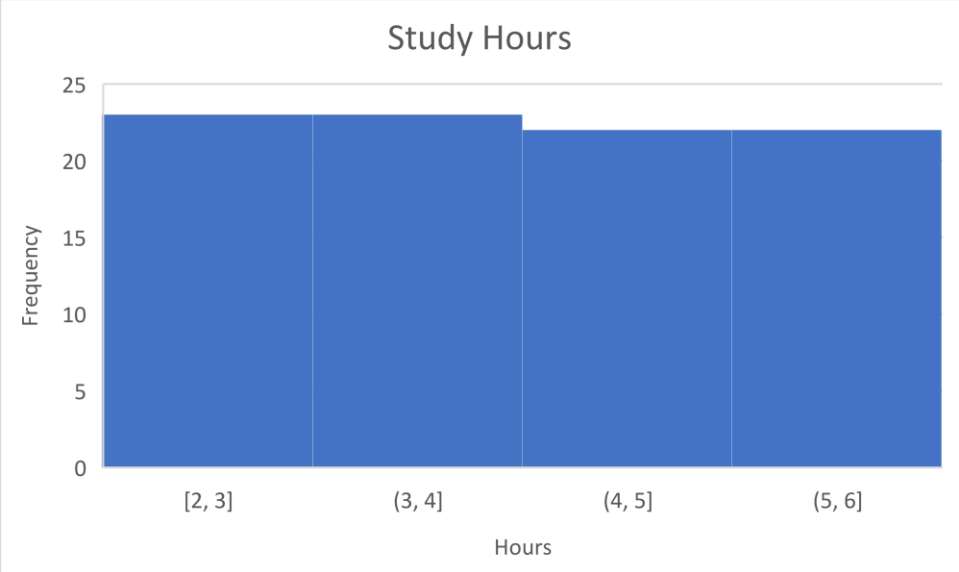
Row Labels	Average of Exam Score
2	72
3	76
4	87
5	85
6	92
Grand Total	**85**



There is high correlation between the hours the students studied and the test scores they received. It appears that the longer you studied, the better average grade the student received.

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





Study hours were relatively evenly distributed. Whereas the majority of students received a B on their exam.

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

The top five performers were all females. They also all studied 6 hours. Their ages were a bit diverse, but 4 out of 5 of the top performers were 16 or 18.

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

There is strong positive correlation between the amount of hours student studies and the exam scores they received. In other words, the more you studied, the higher your average exam score would likely be. (.76 is strong positive correlation)

3. Provide a summary result your findings.
   The majority of the class received B's on their exam scores. The class excelled best at Science and Math but struggled in English. Females performed better than males on the exams. Having an average test score eight points higher than males. There is positive correlation between study hours and test scores. Those who studied more received better test grades on average. Woman studied more than men on average, therefore, received better test scores.

4. 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.