Assignment1(Group of two)
SHAY KORHORN
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	85.0111111	Mean	4.46666667
Standard Error	0.72695463	Standard Error	0.12054806
Median	86	Median	4
Mode	88	Mode	4
Standard Deviation	6.89649715	Standard Deviation	1.14361933
Sample Variance	47.5616729	Sample Variance	1.30786517

Range	27	Range	4
Minimum	70	Minimum	2
Maximum	97	Maximum	6
Sum	7651	Sum	402
Count	90	Count	90

	Q3	Q1	IQR
Exam Score	90	80	10
Study Hours	5	3.25	1.75

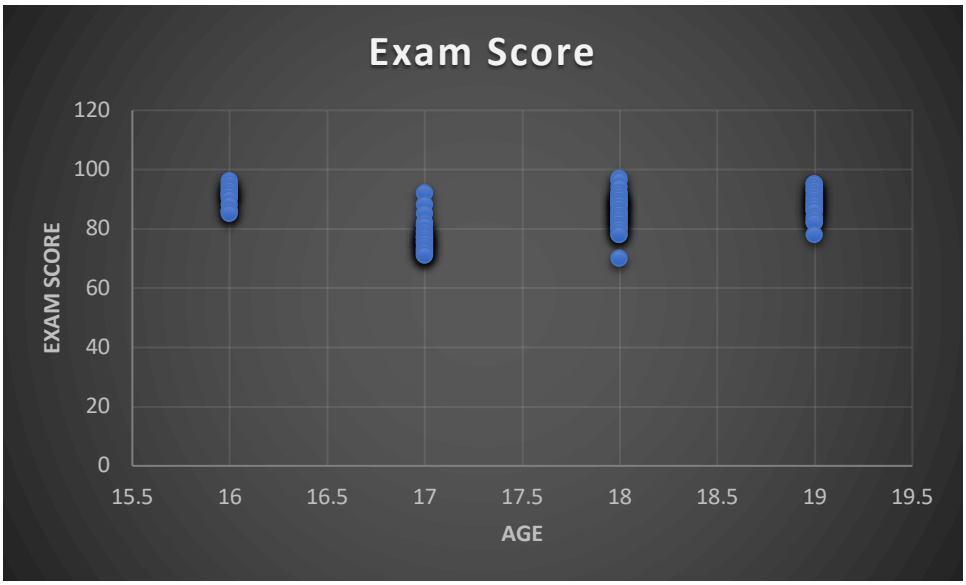
This data shows the mean, median, and standard deviation of all male and females exam scores and their study hours.

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	89.35555556	4.955555556
Male	80.66666667	3.977777778
Grand Total	**85.01111111**	**4.466666667**

This data shows the difference between males and females average exam scores in comparison with their average study hours. As shown above, females have a higher amount of average study hours than males, which reflects their higher grade than males.

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



Row Labels	Average of Exam Score
16	90.69230769
17	77.56521739
18	85.90909091
19	88.23809524
Grand Total	**85.01111111**

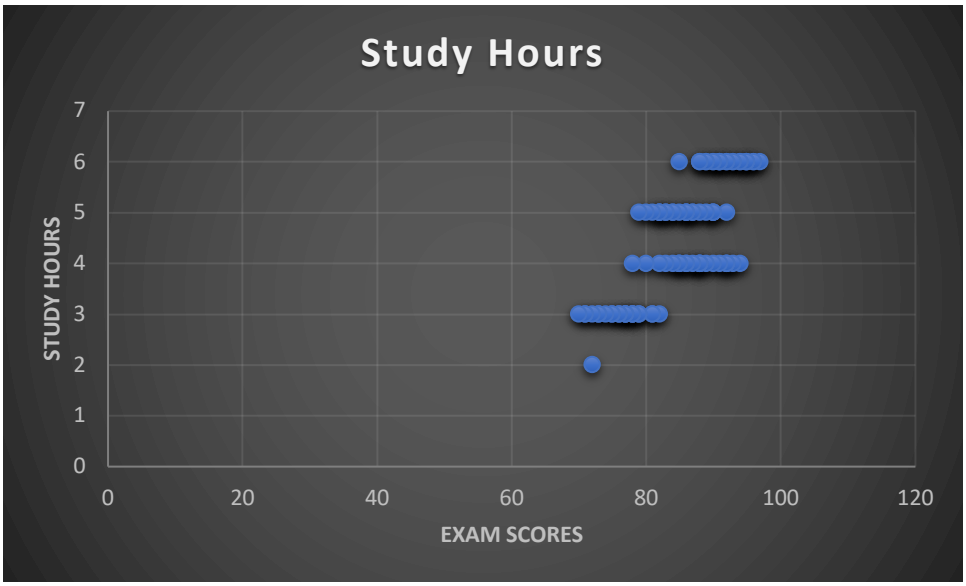
The data above shows the difference in average exam scores between different ages, such as 16, 17, 18, and 19 years old. It shows the average exam scores are the highest for 16 year olds, then for the 19 year olds, then for the 18 year olds, and finally the lowest scores with the 17 year olds.

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

Row Labels	Average of Exam Score
English	83.4137931
Math	85.67741935
Science	85.86666667
Grand Total	**85.01111111**

This data shows the average exam scores for three different subjects, English, Math, and Science. In order, students scored highest on the science portion, then on the math portion, and worst on the English portion of the exams.

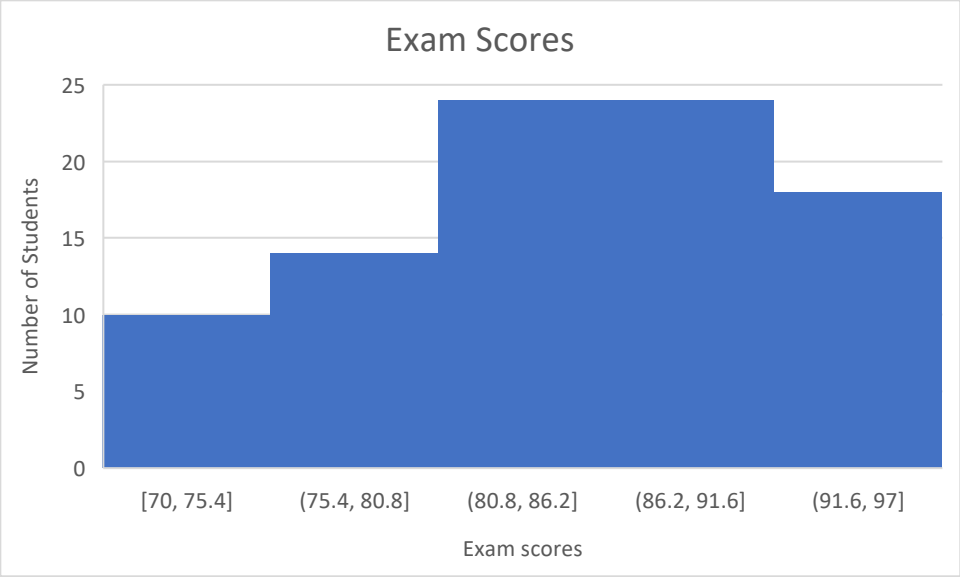
E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.

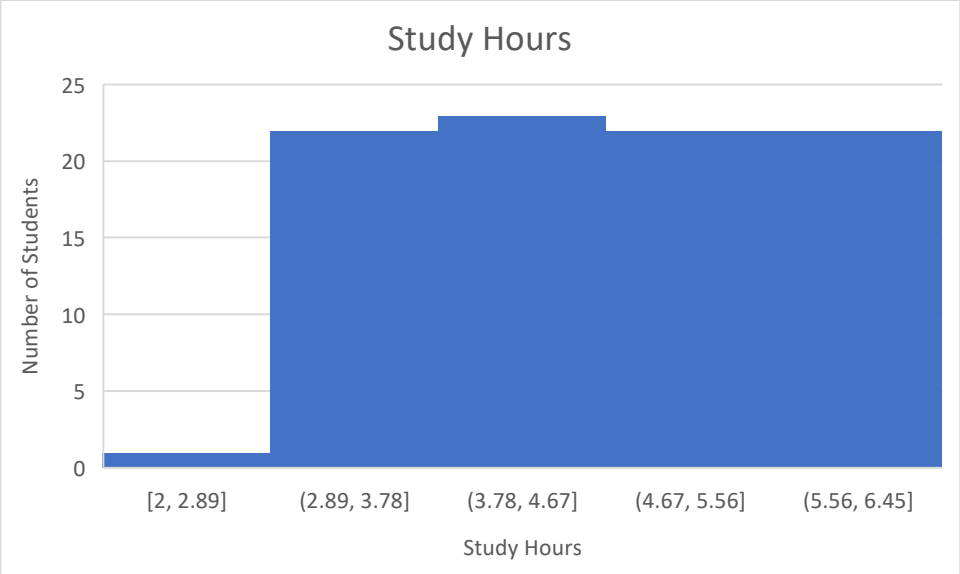


Row Labels	Average of Exam Score
2	72
3	76.04545455
4	87.2173913
5	85.27272727
6	92
Grand Total	**85.01111111**

This data can portray how study hours contribute to exam scores. On average students with more study hours can score higher on the exam scores.

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





As said before, this data above shows how students with more study hours can score higher on their exams, on average.

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

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 data above shows the top students who took the exams. As you can see they are all females due to their ability to study more, resulting in a higher grade on their exams.

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	0.76435772	1

The data shows that study hours and exam scores have a positive correlation. They have a 76% correlation.

3. Provide a summary result your findings.

To summarize these findings, we not only see that the more these students were able to study, the higher their overall grade would be on all three exams. Taking gender into account, we see that our top students or highest scorers on these three exams were all females, who are shown to on average study more than the males for these tests.

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.