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	Mean	4
Median	86	Median	4
Standard Deviation	7	Standard Deviation	1
Minimum	70	Minimum	2

Maximum	97	Maximum	6

The mean exam score is 85% while the median is 86%. The standard deviation of the exam score is 7. The minimum exam score 70% and the maximum is 97%.

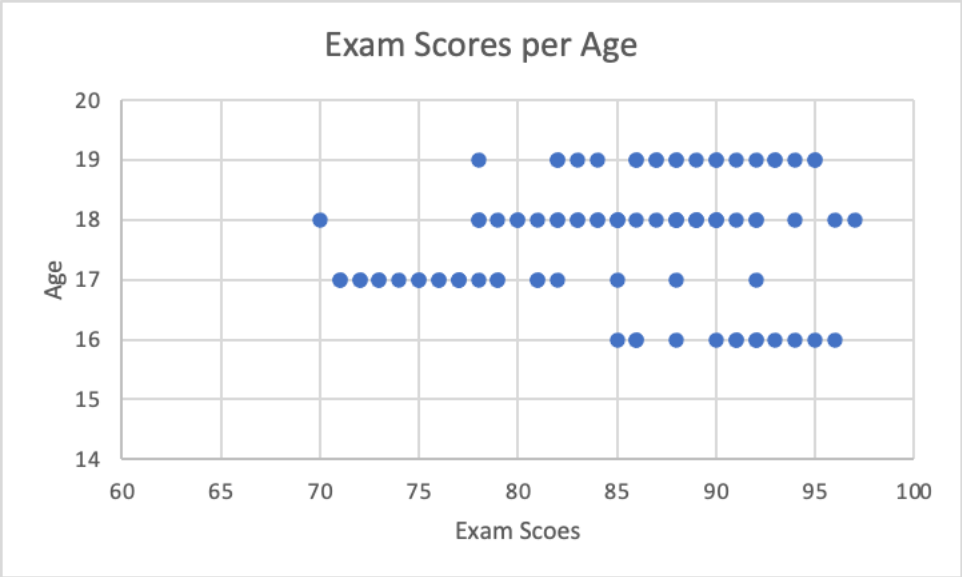
The mean and median study time is 4hrs. The standard deviation of the study time is 1. The minimum study time is 2hrs and the maximum is 6hrs.

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

    1. The average exam score for Females is 89.36% and Males is 80.67%. And the average study hours for Females is 4.96 hrs and for Males is 3.98 hrs.

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

    1. Exam scores seem to be very weakly related to age, if at all.



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

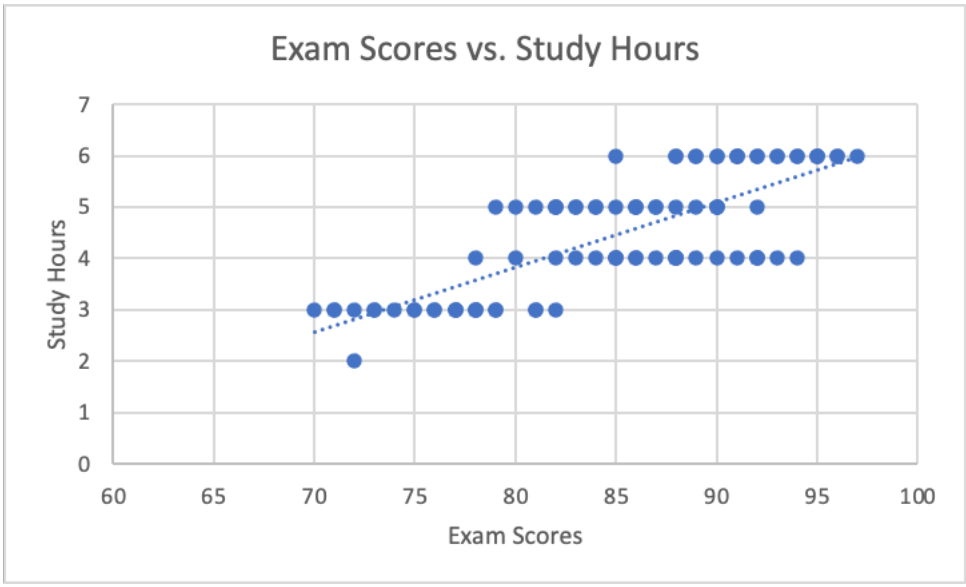
Row Labels	Average of Exam Score
Female	**89.36**
English	86.86

Math	89.73
Science	91.19
Male	**80.67**
English	80.20
Math	81.88
Science	79.79

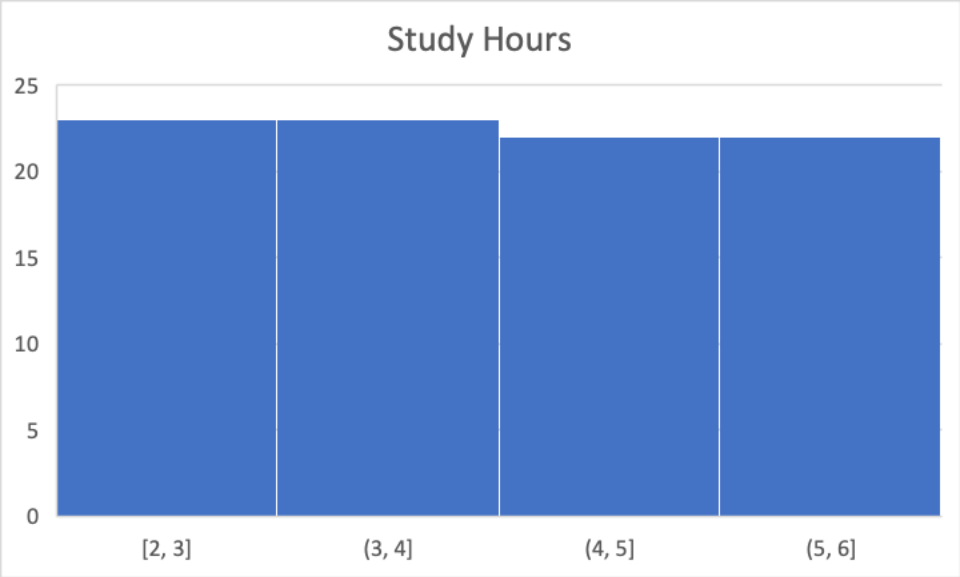
The average exam score from Females in English is 86.86%, Math is 89.73%, and Science is 91.19%.

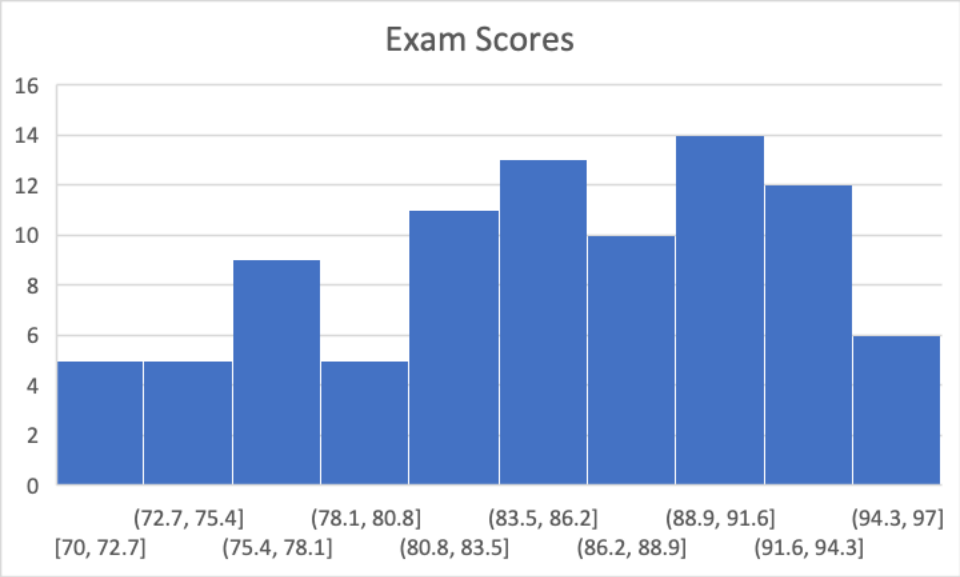
The average exam score from Males in English is 80.20%, Math is 81.88%, and Science is 79.79%.

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



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





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

Student ID	Row Labels	Age	Study Hours	Exam Score
90	Female	18	6	97
18	Female	18	6	96
8	Female	16	6	96
86	Female	19	6	95
4	Female	16	6	95

1. All of the top grades seem to come from Females who study 6hrs. And the highest scores come from people age 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	0.76435772	1

Exam Score and Study Hours have a very strong positive correlation

3. Provide a summary result your findings.

   It seems as though students who study for longer, will most likely get a higher score on the exam. Furthermore, it seems that Males tend to have lower grades overall than Females. And, age has very little to do with how a student will score on the exam.

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.