

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.).

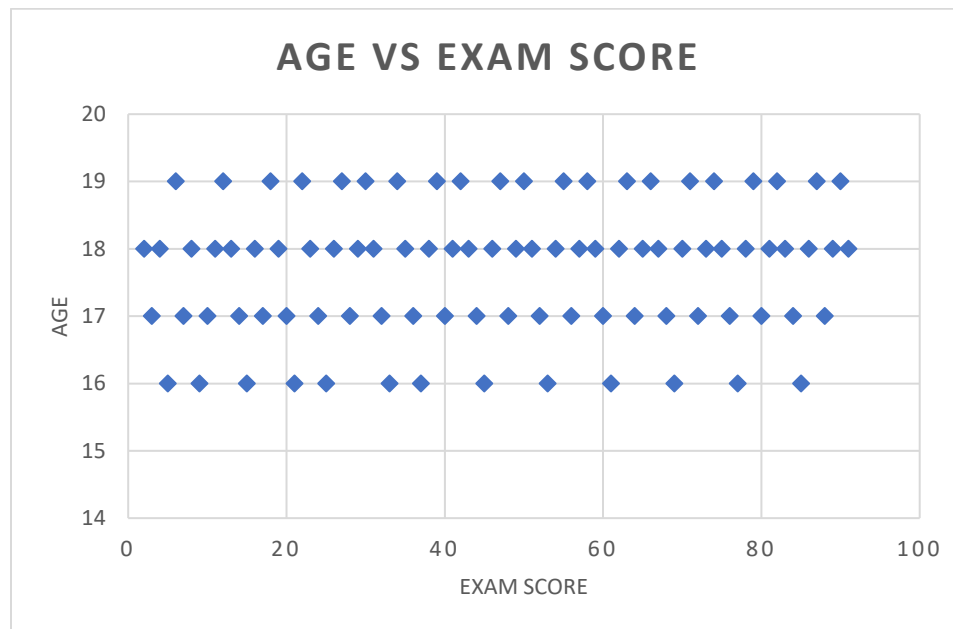
<i>Exam Score</i>		<i>Study Hours</i>	
Mean	85.01111111	Mean	4.466666667
Standard Error	0.726954629	Standard Error	0.120548062
Median	86	Median	4
Mode	88	Mode	4
Standard		Standard	
Deviation	6.896497148	Deviation	1.143619329
Range	27	Range	4
Minimum	70	Minimum	2
Maximum	97	Maximum	6

Sum	7651	Sum	402
Count	90	Count	90

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

The average exam score for females is 89.36 and average hours of study are 4.96 and the average exam score for males is 80.67 and the average of study hours is 3.98.

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



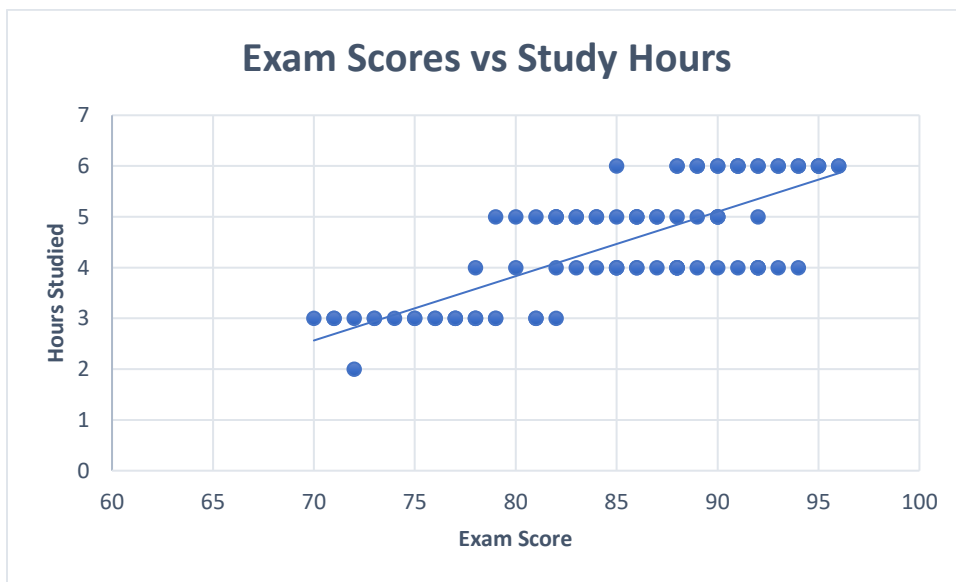
- D. **Subject Analysis:** Explore average scores for each subject to identify strengths and weaknesses.

Row Labels	Average of Exam Score	Average of Study Hours	Average of Age
English	83.41	4.28	17.66
Female	86.86	4.71	17.71
Male	80.2	3.87	17.6
Math	85.68	4.58	17.74
Female	89.73	5.07	17.47
Male	81.88	4.13	18
Science	85.87	4.53	17.67

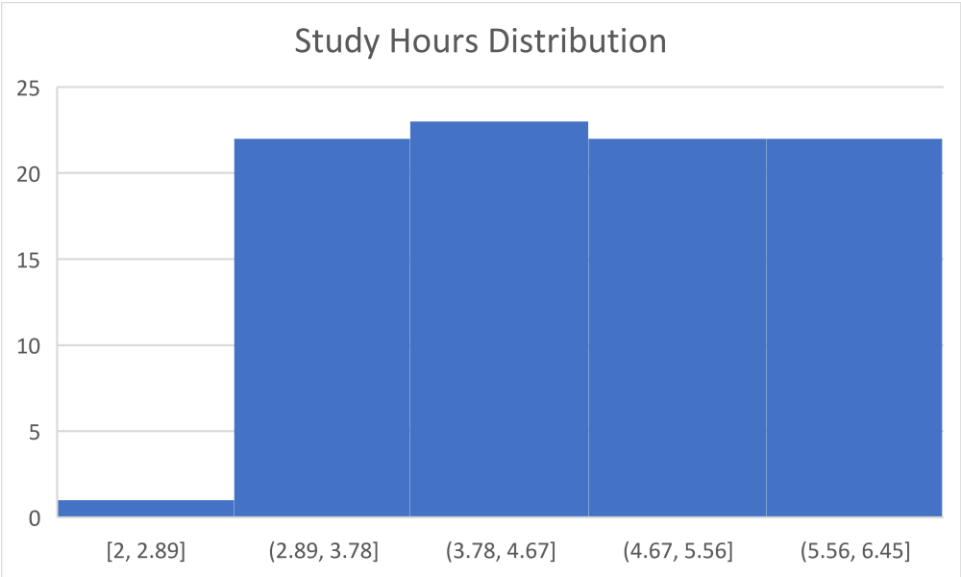
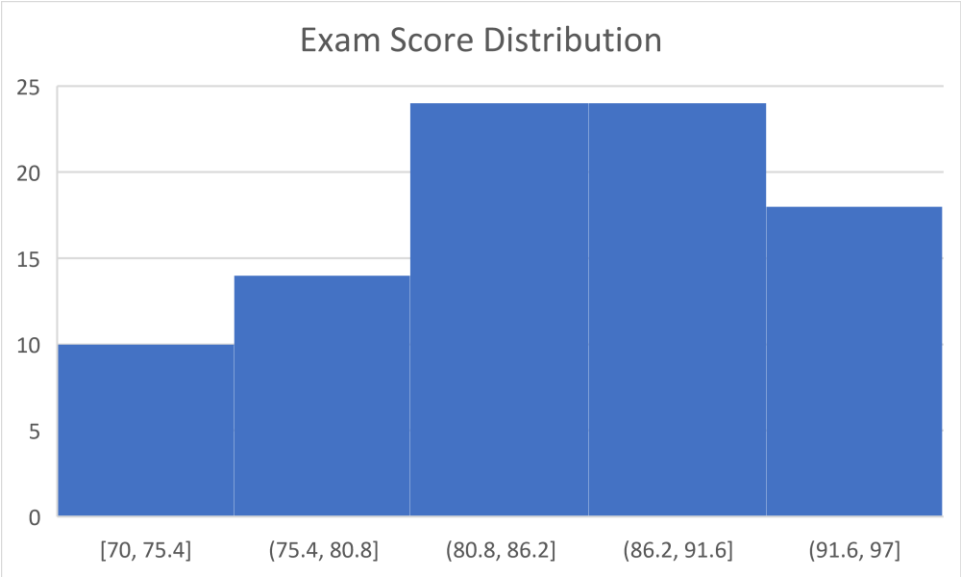
Female	91.19	5.06	17.44
Male	79.79	3.93	17.93
Grand Total	85.01	4.47	17.69

The average exam scores are as follows, English is 83.4, Math is 85.68, and Science is 85.87. However, for English the average exam score for females is 86.86 and the average score for males is 80.20, for Math the average exam score for females is 89.73 and the average score for males is 81.88, and for science the average exam score for females is 91.19 and the average exam scores for males is 79.79. This would show that solely based on exam score females tend to do better on average in all subjects compared to males. Another key factor to consider is that males on average study less than females which could be a reason for their test scores being significantly lower.

- 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.

Gender	Max of Age	Average of Study Hours
93	19	5
Female	19	5.333333333
94	19	5
Female	19	5.333333333
95	19	6

Female	19	6
96	18	6
Female	18	6
97	18	6
Female	18	6

The top five performers gender is all female, 3 of the 5 study for 6 hours on average and the other 2 study for 5.33 on average. All the top performers tend to be at the top of the age group with two of the top performers being 18 and the last 3 being 19 years old.

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

The correlation between study hours and exam scores is 0.76436. This means that there is a strong correlation between exam scores and study hours.

3. Provide a summary result your findings.

In conclusion the results found showed that there was a strong correlation between study hours and exam scores across the board. This would explain why males test scores were so much lower than females because on average females tended to study more than males meaning that they would have a higher exam score. This data set suggests that if someone wishes to get a better exam score they should spend more time studying.

4. Using the instructions provided by GitHub, create a git repository named **DS160InClassAssignment**, 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.