Alexander Kyrkjeeide

Assignment1(Individual/ Group of two)
CS160
Introduction to Data Science
Fall 2023

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	4.466666667
Standard Error	0.726954629	Standard Error	0.120548062
Median	86	Median	4
Mode	88	Mode	4
Standard Deviation	6.896497148	Standard Deviation	1.143619329
Sample Variance	47.56167291	Sample Variance	1.307865169

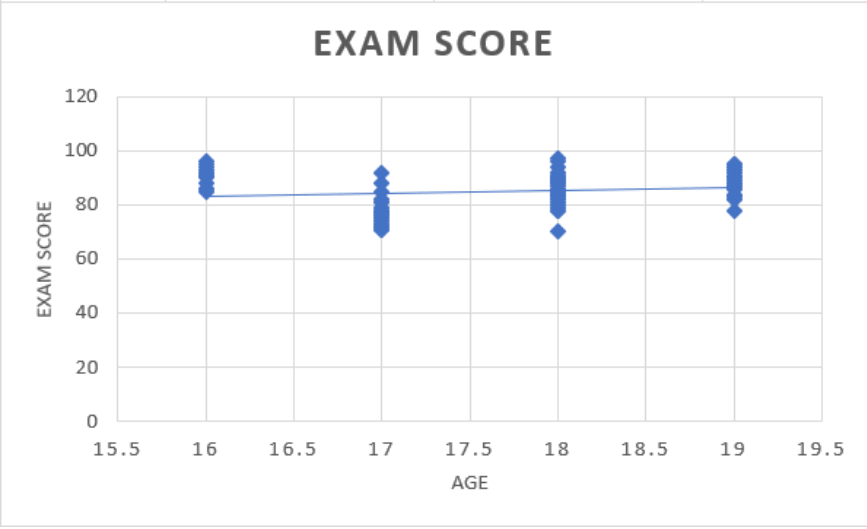
Kurtosis	-0.768538292	Kurtosis	-1.253638401
Skewness	-0.369397246	Skewness	-0.031551123
Range	27	Range	4
Minimum	70	Minimum	2

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

Upon conducting a comprehensive analysis using a PivotTable, it became evident that, on average, females dedicate approximately one additional hour to their studies in comparison to males. Furthermore, the examination results underscore a noteworthy trend where females consistently achieve significantly higher grades than males.

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



An ascending trendline is evident, indicating that as students age, their exam scores tend to improve. While this trend is apparent, it's essential to acknowledge that numerous factors may contribute to this phenomenon. However, a more comprehensive understanding of these factors would necessitate further research and

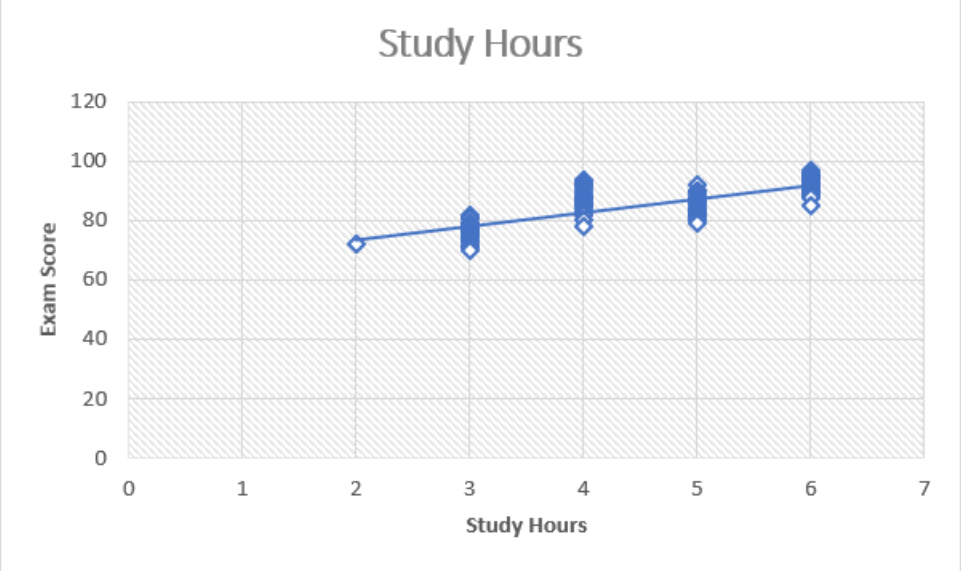
investigation.

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**

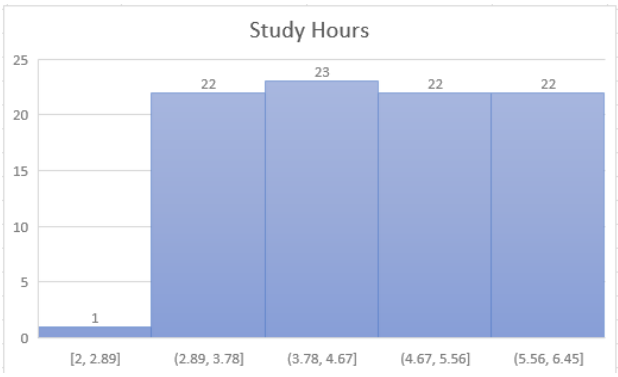
Upon a thorough examination of the pivot table, the findings highlight that English appears to be the subject where students perform least favorably, as evidenced by the lower average exam scores. Conversely, there's a nearly equal level of proficiency in the subjects of Mathematics and Science, though Science emerges as the subject where students demonstrate their greatest strength.

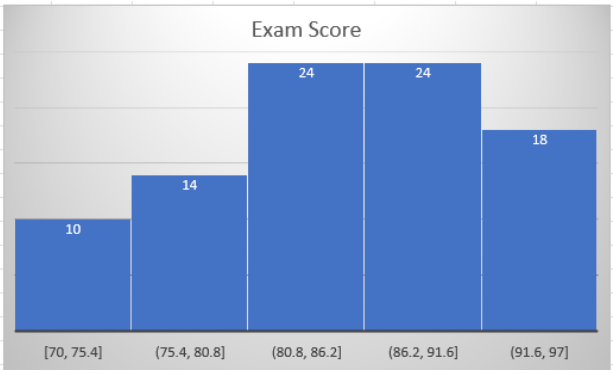
E. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.



The trending line is increasing illustrating that with more studying the higher the exam score will be.

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





Study hours exhibit a left skew as indicated by the fact that the median is lower than the mean. In contrast, exam scores display a right skew, given that the mean is greater than the median, as evident from the descriptive analysis.

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

Student ID	Gender	Age	Subject	Exam Sco	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
86	Female	19	Math	95	6
30	Female	18	Science	94	6
44	Female	16	Math	94	4
62	Female	19	Math	94	6
26	Female	19	Math	93	6
52	Female	16	English	93	4
78	Female	19	Science	93	6

Regarding gender distinctions, it's notable that none of the male students manage to secure a position within the top 10 exam score range. It's also surprising to observe that there are more students aged 16 outperforming 18-year-olds in terms of scoring, although the majority of top scorers fall within the 19-year-old bracket.

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 0.764 which shows a strong value due to the r value being larger than 0.7. This shows that the linear relationship between the two variables is strong.

3. Provide a summary result for each of the findings.

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.

Alexander Kyrkjeeide

Submission:

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