**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.

1. **Summary Statistics:** Calculate summary statistics for exam scores and study hours (mean, median, standard deviation, etc.).

|  |  |  |  |
| --- | --- | --- | --- |
| *Exam Score* |  | *Study Hours* |  |
|  |  |  |  |
| Mean | 85 | Mean | 4 |
| Standard Error | 1 | Standard Error | 0 |
| Median | 86 | Median | 4 |
| Mode | 88 | Mode | 4 |
| Standard Deviation | 7 | Standard Deviation | 1 |
| Sample Variance | 48 | Sample Variance | 1 |
| Kurtosis | -1 | Kurtosis | -1 |
| Skewness | 0 | Skewness | 0 |
| Range | 27 | Range | 4 |
| Minimum | 70 | Minimum | 2 |
| Maximum | 97 | Maximum | 6 |
| Sum | 7651 | Sum | 402 |
| Count | 90 | Count | 90 |
| Confidence Level(95.0%) | 1 | Confidence Level(95.0%) | 0 |

Most students studied for 4 hours and had the test score of 86, which is good.

1. **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 | 5 |
| Male | 81 | 4 |
| **Grand Total** | **85** | **4** |

Females have both a higher exam scores and studied more hours than males. Correlation = 1

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

The trendline isn’t increasing much therefore there isn’t a lot of correlation. Correlation = 0.15

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

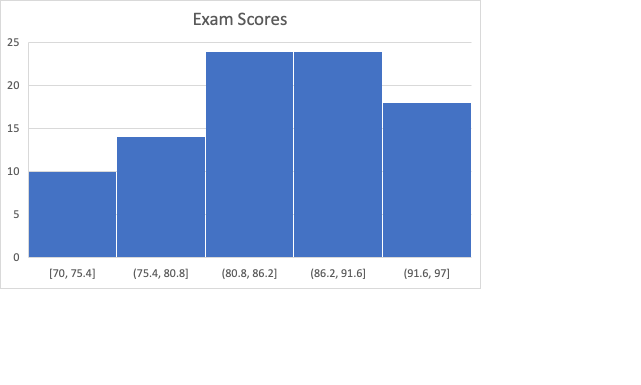
|  |  |
| --- | --- |
| **Row Labels** | **Average of Exam Score** |
| English | 83 |
| Math | 86 |
| Science | 86 |
| **Grand Total** | **85** |

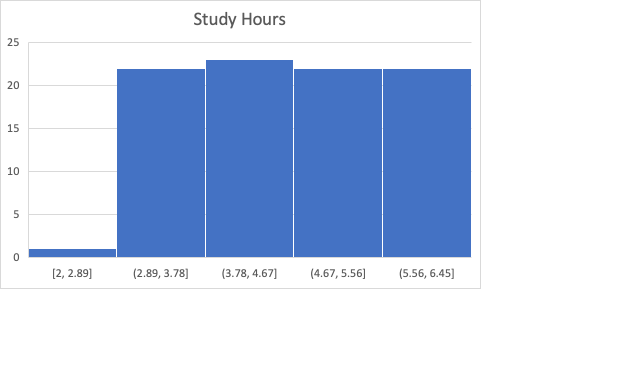
Math and Science are the strongest subject, English is the weakest subject.

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

The trendline is positive which means the more hours you study, the higher your exam score will be. Correlation = 0.76, 76%

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





Left skewed, exam scores is not uniform, study hours is more uniformed but has an outlier

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender | Age | Subject | Exam Score | Study Hours |
| Female | 18 | Science | 97 | 6 |
| Female | 16 | Science | 96 | 6 |
| Female | 18 | Science | 96 | 6 |
| Female | 16 | Math | 95 | 6 |
| Female | 19 | Math | 95 | 6 |
| Female | 19 | Math | 95 | 6 |
| Female | 18 | Science | 94 | 6 |
| Female | 16 | Math | 94 | 4 |
| Female | 19 | Math | 94 | 6 |
| Female | 19 | Math | 93 | 6 |
| Female | 16 | English | 93 | 4 |
| Female | 19 | Science | 93 | 6 |

The top 10 highest exam scores are all female and most of them studied for 6 hours.

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

Correlation = 0.76, 76%. The more time you spend studying, the higher exam score you will receive.

1. Provide a summary result for each of the findings.
2. 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.