**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 | 4.5 |
| Standard Error | 1 | Standard Error | 0.1 |
| Median | 86 | Median | 4.0 |
| Mode | 88 | Mode | 4.0 |
| Standard Deviation | 7 | Standard Deviation | 1.1 |
| Sample Variance | 48 | Sample Variance | 1.3 |
| Kurtosis | -1 | Kurtosis | -1.3 |
| Skewness | 0 | Skewness | 0.0 |
| Range | 27 | Range | 4.0 |
| Minimum | 70 | Minimum | 2.0 |
| Maximum | 97 | Maximum | 6.0 |
| Sum | 7651 | Sum | 402.0 |
| Count | 90 | Count | 90.0 |

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

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Sum of Study Hours** | **Average of Exam Score** |
| **English** | **124** | **83** |
| Female | 66 | 87 |
| Male | 58 | 80 |
| **Math** | **142** | **86** |
| Female | 76 | 90 |
| Male | 66 | 82 |
| **Science** | **136** | **86** |
| Female | 81 | 91 |
| Male | 55 | 80 |
| **Grand Total** | **402** | **85** |

Students studied the least for English which correlated to lower exam scores. Over the three classes, the females studied for longer and in return got higher test score.

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

There is a correlation of .15 meaning that exam score is not dependent on age.

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

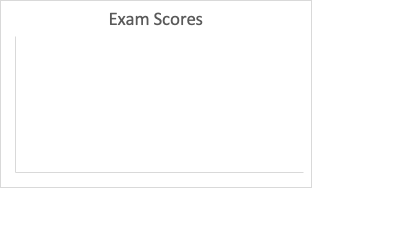
|  |  |
| --- | --- |
| **Row Labels** | **Average of Exam Score** |
| English | 83 |
| Math | 86 |
| Science | 86 |
| **Grand Total** | **85** |

Math and science are the strengths of the class and English is the weakest subject

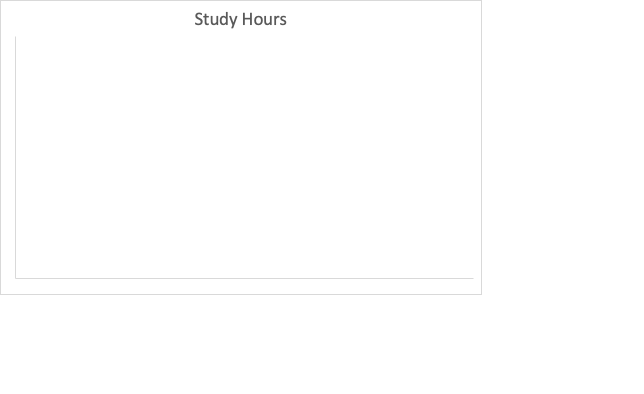
1. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.

There is a positive correlation of .76 so as study hours increase, so do the test scores.

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



The Exam scores are skewed to the left



The study hours are pretty uniform with an outlier at 2

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Student ID | Gender | Age | Subject | Exam Score | 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 |

Females take up the top ten spots for highest exam scores and all but two of them studied for six hours. There is little correlation between age and exam scores.

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

There is a strong correlation between study hours and exam scores with .76

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

For the most part, if a student studies more, they will get higher test scores.

1. 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.