**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.

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

|  |  |  |  |
| --- | --- | --- | --- |
| *Exam Score* |  | *Study Hours* |  |
|  |  |  |  |
| Mean | 85.01111 | Mean | 4.466667 |
| Standard Error | 0.726955 | Standard Error | 0.120548 |
| Median | 86 | Median | 4 |
| Mode | 88 | Mode | 4 |
| Standard Deviation | 6.896497 | Standard Deviation | 1.143619 |
| Sample Variance | 47.56167 | Sample Variance | 1.307865 |
| Range | 27 | Range | 4 |
| Minimum | 70 | Minimum | 2 |
| Maximum | 97 | Maximum | 6 |
| Sum | 7651 | Sum | 402 |
| Count | 90 | Count | 90 |
|  |  |  |  |
|  | Q3 | Q1 | IQR |
| Exam scores | 90 | 80 | 10 |
| Study Hours | 3.25 | 3.25 | 0 |

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

The female students get a higher average exam score than the male students. They also spend more hours studying, which likely correlates to the higher overall scores.

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

|  |  |
| --- | --- |
| **Row Labels** | **Average of Exam Score** |
| 16 | 90.69230769 |
| 17 | 77.56521739 |
| 18 | 85.90909091 |
| 19 | 88.23809524 |
| **Grand Total** | **85.01111111** |

The exam scores do not really change based on age, I see a higher correlation between study hours. The 16 year olds actually do the best on average compared to the other 17, 18, and 19 year olds.

1. **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** |

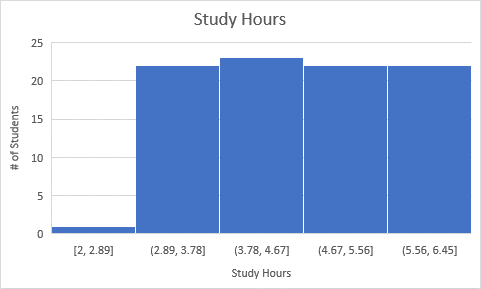
The Students overall do better in science and math compared to English. The strengths are math and science and the weakness is definitely English.

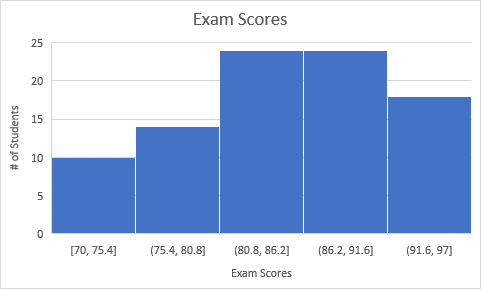
1. **Study Hours vs. Exam Score:** Create a scatter plot to visualize the relationship between study hours and exam scores.

|  |  |
| --- | --- |
| **Row Labels** | **Average of Exam Score** |
| 2 | 72 |
| 3 | 76.04545455 |
| 4 | 87.2173913 |
| 5 | 85.27272727 |
| 6 | 92 |
| **Grand Total** | **85.01111111** |
|  |  |

The average exam scores steadily increase as the amount of study hours goes up. This shows a positive correlation between study hours and exam scores.

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





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 |

The students with the highest exam scores are also students who have the largest amount of study hours, and they are also all female. The females on average get higher grades than the males.

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

|  |  |  |
| --- | --- | --- |
|  | *Exam Score* | *Study Hours* |
| Exam Score | 1 |  |
| Study Hours | 0.764358 | 1 |

The correlation is very high between study hours and Exam Scores. The correlation is a 0.76 correlation, which is EXTREMELY high.

1. Provide a summary result your findings.

From the data and analysis throughout this assignment, I can see that age does not affect exam score, however, the amount of time spent studying drastically increases the likelihood of doing well on an exam. This makes a lot of sense, if you think about it, the more time spent learning about a topic the better you will know it, and the data supports this will an extremely high positive correlation between study hours and exam 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.