**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 | 85.01111 | 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 |
| Kurtosis | -0.76854 | Kurtosis | -1.25364 |
| Skewness | -0.3694 | Skewness | -0.03155 |
| Range | 27 | Range | 4 |
| Minimum | 70 | Minimum | 2 |
| Maximum | 97 | Maximum | 6 |
| Sum | 7651 | Sum | 402 |
| Count | 90 | Count | 90 |
| Confidence Level(95.0%) | 1.444443 | Confidence Level(95.0%) | 0.239526 |

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

Based on the table we can conclude that for females, on average have one hour more than males for study hours. For females, on average they get higher scores in exams for English, Math, and Science.

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Average of Exam Score** | **Average of Study Hours** |
| **English** | **83** | **4** |
| Female | 87 | 5 |
| Male | 80 | 4 |
| **Math** | **86** | **5** |
| Female | 90 | 5 |
| Male | 82 | 4 |
| **Science** | **86** | **5** |
| Female | 91 | 5 |
| Male | 80 | 4 |
| **Grand Total** | **85** | **4** |

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

There seems to be a very weak correlation between age and exam scores. Since the correlation has a value of 0.15, we can conclude there is no correlation.

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

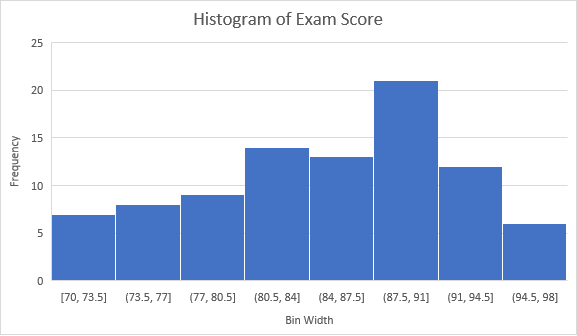
The students perform better at Math and Science which makes it a strength. The students perform slightly worse at English which makes it a weakness.

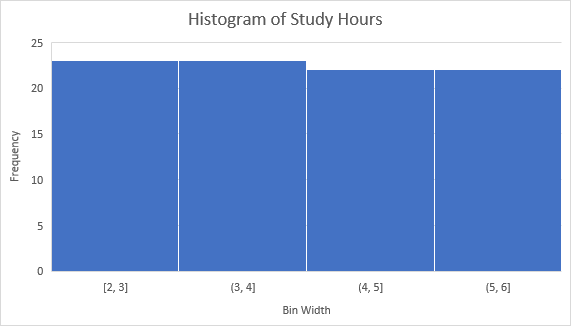
|  |  |
| --- | --- |
| **Row Labels** | **Average of Exam Score** |
| English | 83 |
| Math | 86 |
| Science | 86 |
| **Grand Total** | **85** |

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 0.76 between study hours and exam scores. This implies that there is a strong correlation between the two. This shows that the more hours a student studies, the higher exam score they can receive.

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.

The top ten students are all female and most study for 6 hours except for two students who have 4 hours. All the top ten students are between 16 to 19 years old. However, there is no correlation between age and study hours since the correlation value is 0.6 and the correlation value for age and exam hour is -0.11.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 |
| 62 | Female | 19 | Math | 94 | 6 |
| 44 | Female | 16 | Math | 94 | 4 |
| 26 | Female | 19 | Math | 93 | 6 |
| 78 | Female | 19 | Science | 93 | 6 |
| 52 | Female | 16 | English | 93 | 4 |

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

The correlation value for study hours and exam scores is 0.76. This indicates a strong positive correlation between study hours and exam scores. This shows that the more hours a student studies, the higher exam score they can 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.