**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 scores: study hours

Mean:

Median

Standard deviation

Minimum

Maximum

Interquartile range

|  |  |  |  |
| --- | --- | --- | --- |
| *Exam Score* |  | *Study Hours* |  |
|  |  |  |  |
| Mean | 85.01111 | Mean | 4.466667 |
| Median | 86 | Median | 4 |
| Mode | 88 | Mode | 4 |
| Standard Deviation | 6.896497 | Standard Deviation | 1.143619 |
| Minimum | 70 | Minimum | 2 |
| Maximum | 97 | Maximum | 6 |
| Sum | 7651 | Sum | 402 |
| Count | 90 | Count | 90 |

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** |
| **(blank)** |  |  |
| **Grand Total** | **85.01111111** | **4.466666667** |

On average, female students studied more hours and got higher exam scores.

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

|  |  |
| --- | --- |
| **age** | **Average of Exam Score** |
| 16 | 91 |
| 17 | 78 |
| 18 | 86 |
| 19 | 88 |
| (blank) |  |
| **Grand Total** | **85** |

There is a weak positive correlation between age and exam scores.

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 |
| (blank) |  |
| **Grand Total** | **85** |

The average exam score for English is below the scores for the other subjects.

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

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

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

|  |  |
| --- | --- |
| **study hours** | **Average of Exam Score** |
| 2 | 72 |
| 3 | 76 |
| 4 | 87 |
| 5 | 85 |
| 6 | 92 |
| **Grand Total** | **85** |

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

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

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Average of Age** | **Average of Study Hours** |
| 95 | 18 | 6 |
| 96 | 17 | 6 |
| 97 | 18 | 6 |
| **Grand Total** | **17.66666667** | **6** |

All of the top 6 performers were female, they all studied for 6 hours, and they were either 18 or 17 years of age.

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 |

There is a positive correlation between study hours and exam scores.

1. Provide a summary result your findings.

Analysis of the data shows that there is a strong positive correlation between the number of hours studied and exam scores. The data also shows a slight positive correlation between age and exam scores. English has lower scores on average than the other subjects suggesting that it is harder than the other subjects. All of the top performers were female which could suggest that gender is a contributing factor to 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.