**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 | 85.0111111 | Mean | 4.46666667 |
| Median | 86 | Median | 4 |
| Mode | 88 | Mode | 4 |
| Standard Deviation | 6.89649715 | Standard Deviation | 1.14361933 |
| Sample Variance | 47.5616729 | Sample Variance | 1.30786517 |
| Range | 27 | Range | 4 |
| 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 |
| **Grand Total** | **85.01111111** | **4.466666667** |

* From the table and the chart, the women had a higher average on exam scores than the men with 89.36 and they also had a higher average of study hours then the men with 4.96 hours.
* This means that the women studied more, and because of this, they had higher exam 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** |

* This shows that the 16-year-old students have the best exam scores with an average of 90.69, with the 19-year-olds following behind with an average of 88.24, then the 18-year-olds with an average of 85.91, and finally the 17-year-olds with the worst average of 77.57.

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** |

* This table and chart shows that science has the highest exam scores with math following shortly behind with a similar high score. These are the strengths for the students.
* English has a very low exam score compared to math and science, so it is considered the weakness for the students.

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

* This scatter plot shows that there is a positive correlation between exam scores and study hours. The more hours a student studies, the higher their exam score will be.

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

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

* This chart shows that to get the best average score of 92, a student needs to study for 6 hours. The next highest study hour is 4 with the next highest exam average score of 87.21and then 5 hours with an average of 85.27. With 3 study hours, an average score of 76.05 is expected, and finally, the lowest average exam score of 72 comes from studying for 2 hours.

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

* The top performer is student ID 90 who is a female, 18, studies for 6 hours, her best subject is science, and has an exam score of 97.
* The second top performer is student ID 18 who is a female, 18, studies for 6 hours, her best subject is science, and has an exam score of 96.
* The third top performer is student ID 8 who is a female, 16, studies for 6 hours, her best subject is science, and has an exam score of 96.
* The fourth top performer is student ID 86 who is a female, 19, studies for 6 hours, her best subject is math, and has an exam score of 95.
* The fifth top performer is student ID 38 who is a female, 19, studies for 6 hours, her best subject is math, and has an exam score of 95.

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 | 0.76435772 | 1 |

1. Provide a summary result your findings.

* This shows that there is a positive correlation between exam scores and study hours. The correlation is 76% which is shows a positive correlation because it is above 0.5. This means that as study hours increase, exam score will increase. So for the more hours a student studies, the higher their exam score will be. It was also clear that females performed better than males because they tended to study more hours and receive better exam scores. 16 and 19-year-olds also performed better since they studied more and received better scores. Science and math were the strengths for the students, while English was a weakness. Overall, there was a clear positive correlation between study and exam scores seen through the females and their scores and study hours being the top 5 performers.

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.

<https://github.com/morganhardin/DS160_Assignment1_MEH.git>