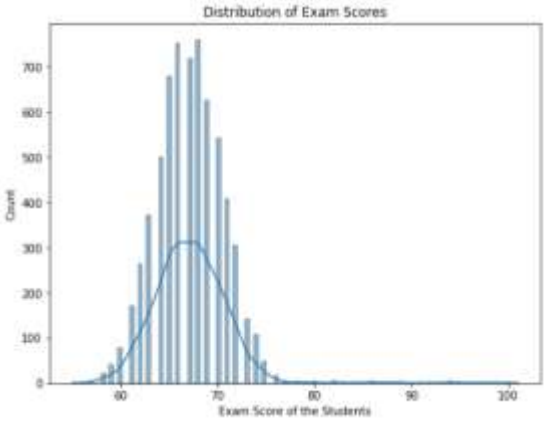
# A Data-Driven Analysis of Exam Scores and Influencing Factors

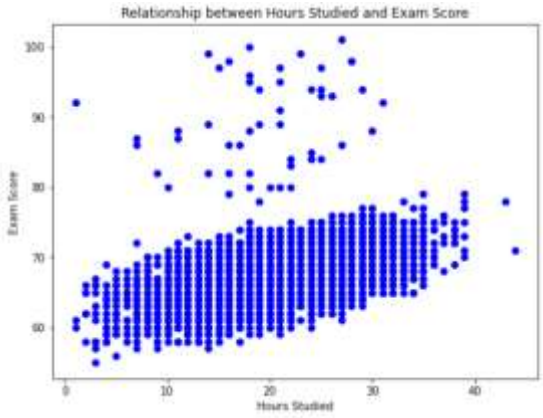
**Name**: Sriharshini Thatiparthi **Student ID**:23070133

**Email**: st24aaq@herts.ac.uk **Link**: Github

**Abstract:** Exam performance of students is examined in this report using histograms, scatter graphs, and correlation analyses. While the Kernel Density Estimate (KDE) line improves the view of score distribution, the histogram shows a concentration of scores in the 65–70 range, indicating average performance among most of the students. The scatter figure shows a positive link between test scores and hours studied, implying that more study time usually results in greater performance. Moreover, correlation studies show a closer connection between test results and attendance, therefore stressing the need of active participation in education together with study practices in determining academic performance.

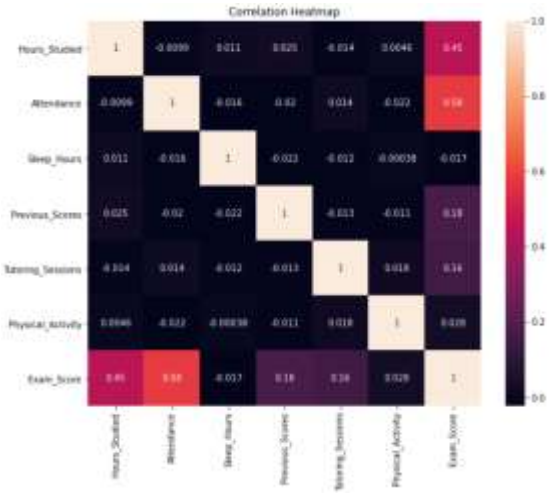



Initially, the histogram shows the exam results of the students, therefore providing information on the overall group's performance and range of scores. With a high around the 65–70 range, the histogram indicates that a noteworthy number of students fell within this range. With a normal distribution shape where scores are more concentrated at the mean, this centre clustering implies that most students performed about an average level. Additionally included in the graph is a KDE line, which offers a smooth estimate of the data distribution, thereby enabling more precisely visualisation of the form of the distribution. Though they make a smaller share of the class, the distribution seems to be somewhat slanted to the right, meaning certain students scored much better. Few students scored over 90, suggesting that quite rare really high marks are. With most students scoring passing or above-average, scores below 60 are also less common on the lower end.




The scatter plot between the exam scores and hours studied shows how students' test grades relate to the number of hours they studied. Given the points of the scatter plot usually show an upward trend, it is clear from them a positive link between study hours and test results. Although the connection does not seem to be exactly linear, this trend implies that generally students who spent more hours in studying tended to get better marks. Most students who fell between 60 and 80 on the test studied between 10 and 30 hours, which fits the centre cluster shown in the histogram. Fascinatingly, there are outliers with high scores over 80 even for students who studied

less than 10, suggesting that success may be influenced by variables other than study hours such previous knowledge, study efficiency, or learning style. Likewise, those students who studied for long stretches—more than thirty hours—did not always score very well, suggesting declining returns in performance enhancement.




Relating this to the previous plots, with a correlation value of 0.45, " Hours Studied" and "Exam Score" clearly show a positive relationship. This supports the trend shown in the scatter plot—that of students who studied more widely obtaining better test results. The modest association indicates, then, that although study hours affect performance results, they do not entirely define them. With "Exam Score", "attendance" has an even more significant positive correlation of 0.58, suggesting that consistent attendance may be more important for students' success than study hours by themselves. This implies that comprehending and remembering knowledge depends much on persistent involvement, maybe via frequent classroom exchanges, which improves results. Other variables, including "Previous Scores" and "Tutoring Sessions," exhibit lesser positive associations with test performance, suggesting that while they could have some impact, it is limited. By stressing the connectivity of variables and demonstrating that both study habits and involvement (attendance) greatly influence test results, the heatmap therefore complements the histogram and scatter plot.

Based on this analysis, attendance is considerably more important for student performance even when study hours favourably correlate with test results. According to the histogram, few students score well; most students perform around average. Though outliers indicate that elements outside of study hours, including previous knowledge and involvement, may influence performance, the scatter plot demonstrates a general tendency of better results with more study time. To improve academic performance generally, regular attendance and good study habits are very essential