

EDA Summary of Team Two

Team Two

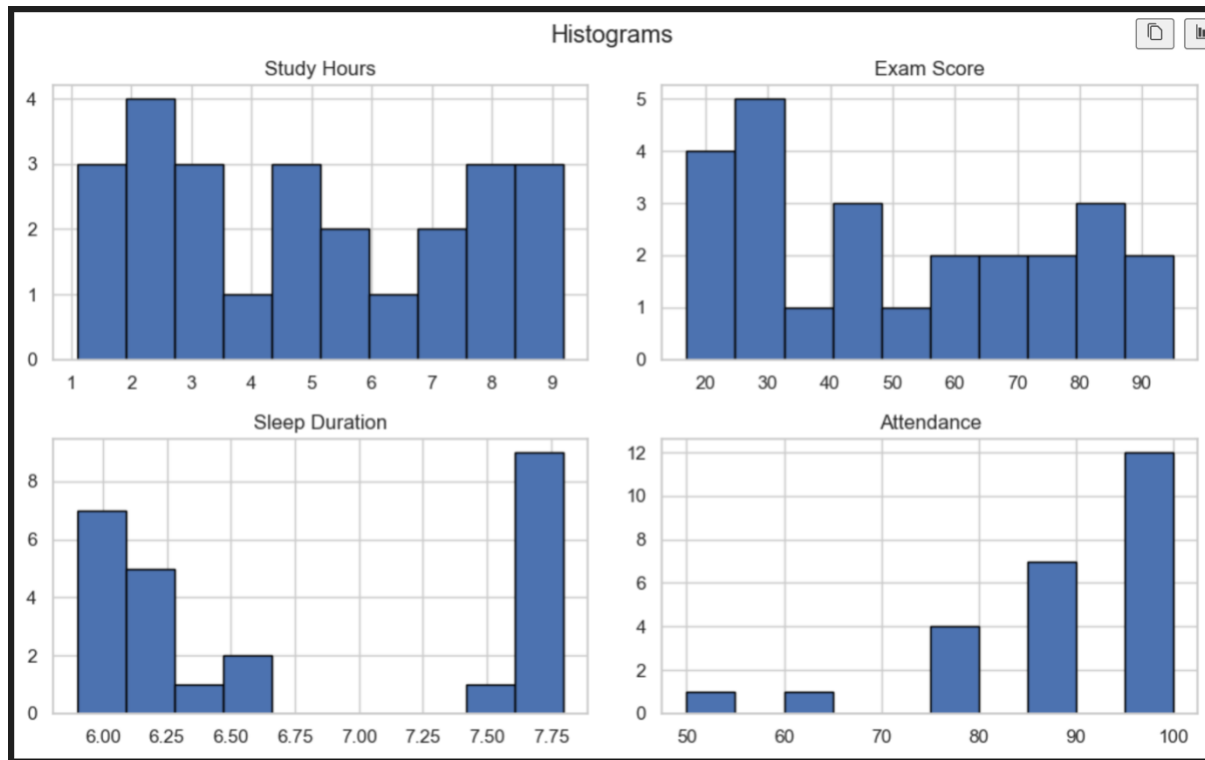
Westcliff University

Data200: Applied Statistical Analytics

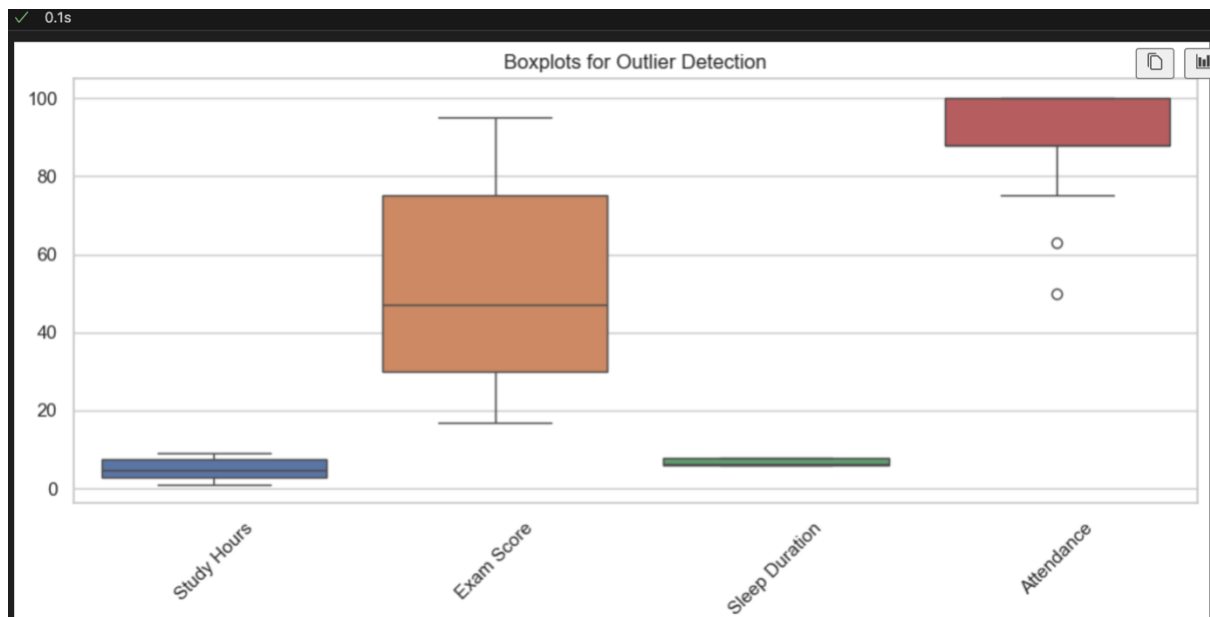
Professor Alok Khatri

EDA Summary of Team Two

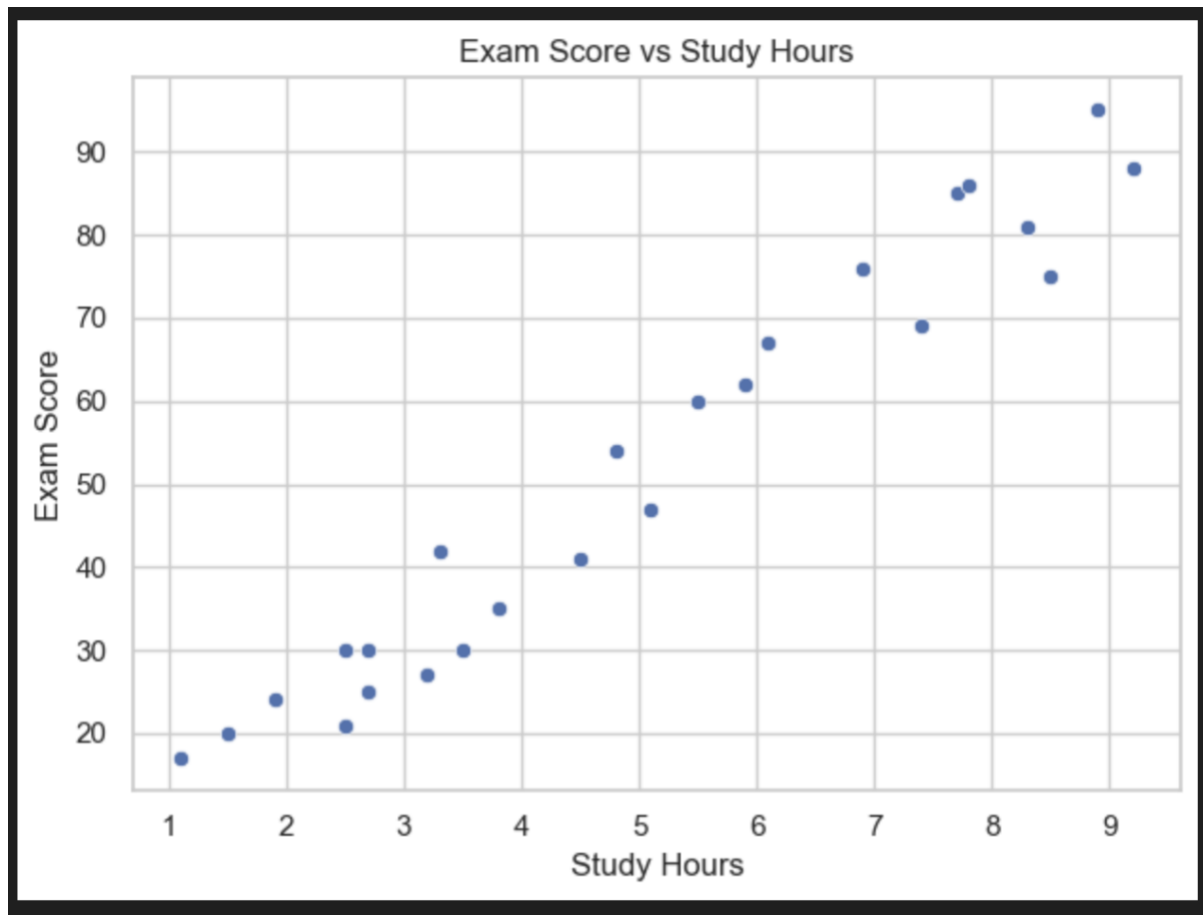
When we explored the student performance file, there were four clear findings we found.

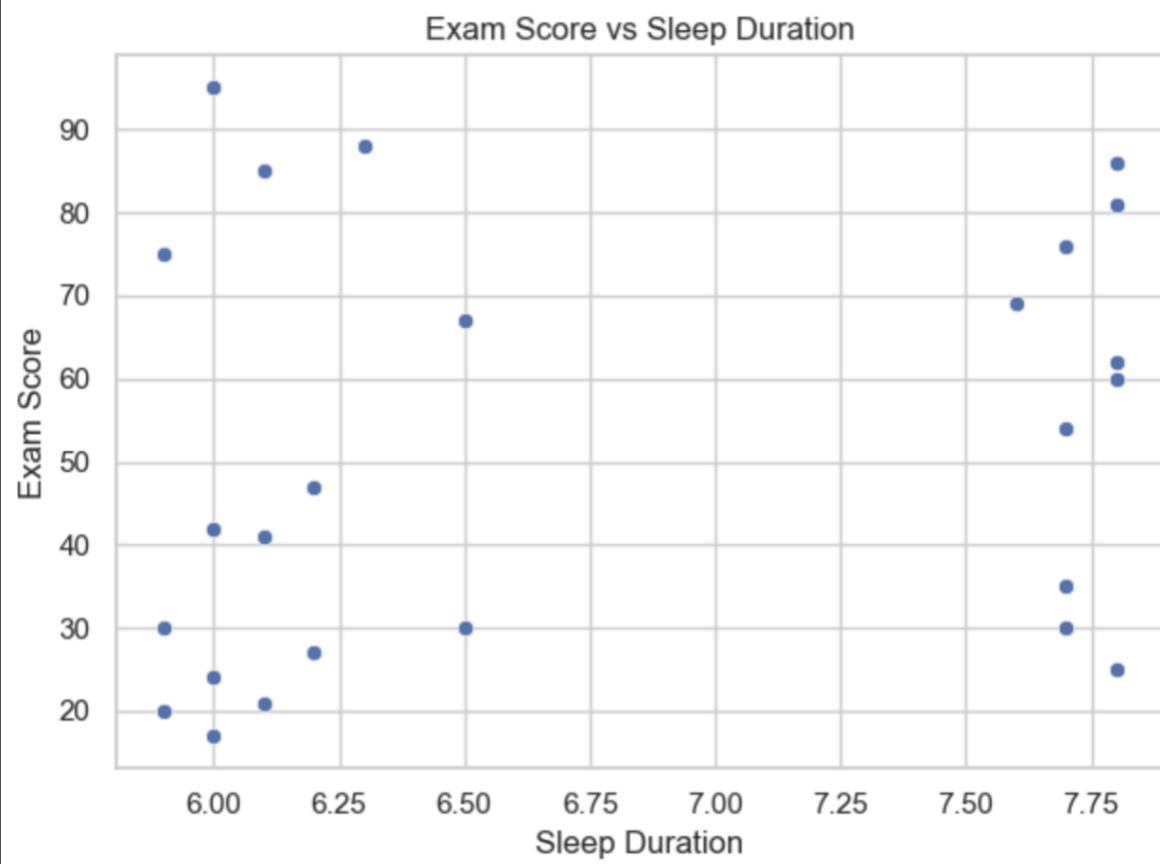


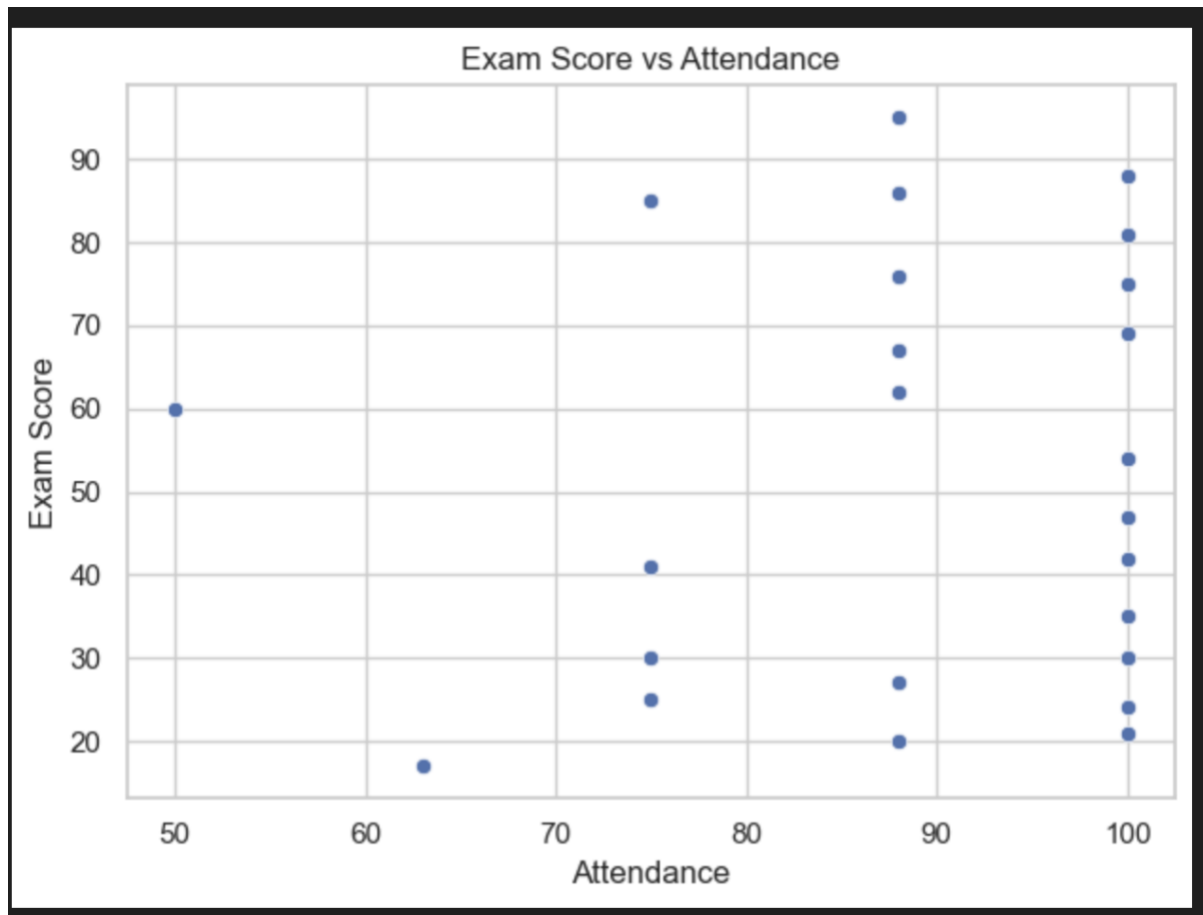
1. It's little things like that which make study time matter most. With that plot of study hours to exam scores, the dots rose in a nice, steady upward line. Those students who studied only one or two hours scored in the teens and twenties whereas students who spent seven or nine hours in their studies ended up scoring by the 70–90 range. This was confirmed with the strongest positive value for this correlation heat map. In simple words, you tend to do better, if you spend more time with the books.



2. It still counts for showing up. Students had generally high school attendance (about 89 % on average), but some students dropped to 50–70 % (the lower band). The students who are also often did sat on the lower edge of the score boxes on the box plot. Attendance versus score scatter chart outputted positively but less steep than study hours. Marks are gained from being present in class but it cannot take the place fully of self study.







3. Sleep helps but to a degree. Between six and eight hours most pupils slept. Within that range, however, scores didn't soar or plunge dramatically: Some seven hour sleepers performed well, others didn't. Sleep duration was weakly correlated and while it can not predict grades on its own, it probably interacts with other habits (long study sessions are more ordinary when one is rested). A regular sleep routine is a sensible thing, but chasing sleep isn't going to magically increase results.

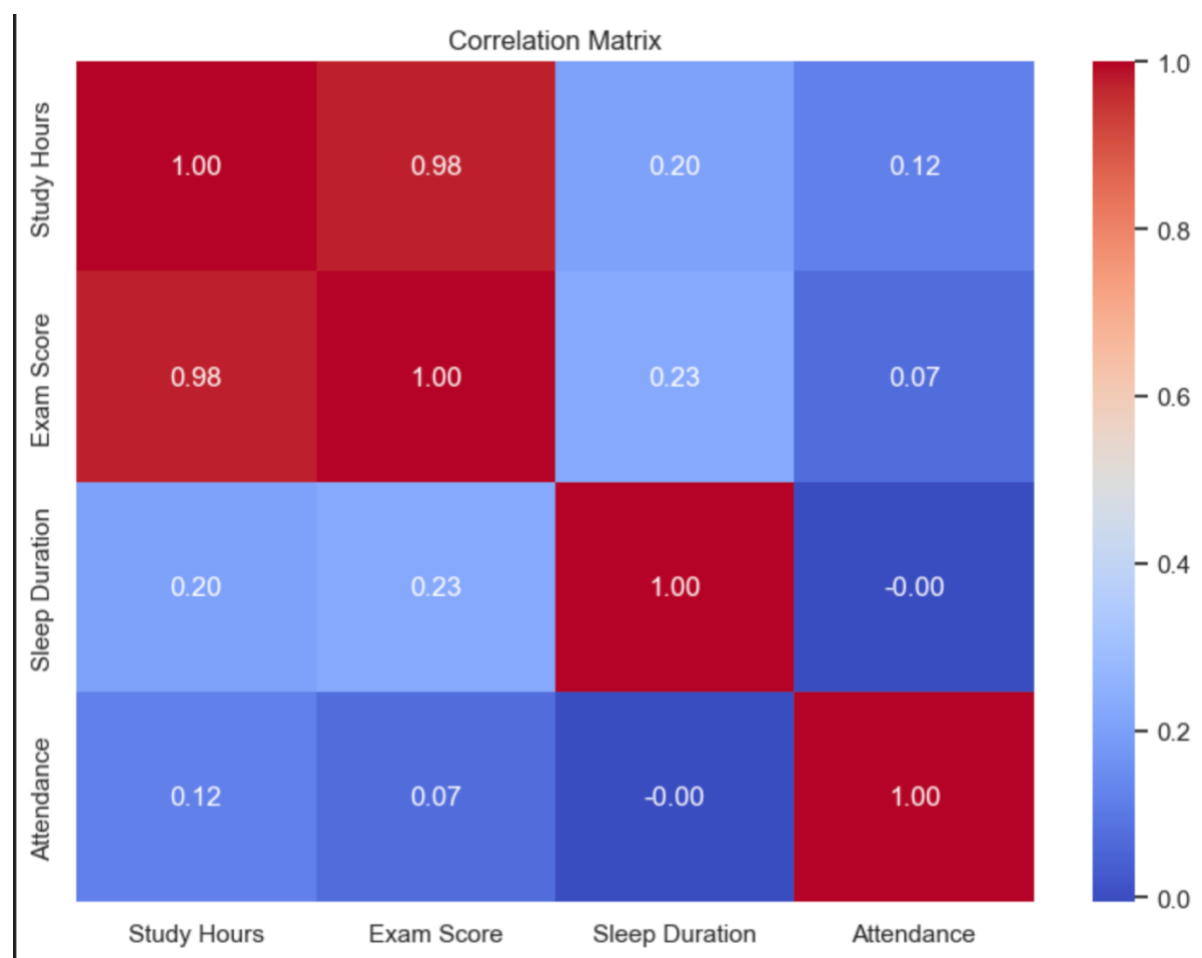
4. The data are clean and the basic premise is trustworthy. There were no missing entries, duplicate rows and just mild outliers. A few outliers of unusually low scores indeed showed up in the box plots, but they are all still in the realm of plausibility. That way we can go ahead and start to model without having to do a lot of heavy repair work. How does this information affect the students and the teachers?

Introduce consistent study hours first: one extra hour could mean the difference between a massive jump or even a stagnant score.

Early tracking of absences: engaging lessons and protecting classroom attendance are the second strongest levers.

Encourage healthy sleep (between six and eight hours): tiring students will not make best use of their study time.

The tidy dataset can be leveraged: future analytics (e.g. linear regression, ANOVA) can now focus on uncovering greater patterns instead of correcting data quality issues.



There are the two pillars of high performance, studying effectively and attending class reliably and the importance of getting at least a little sleep. Focus on these three controllable behaviours and schools can design targeted interventions and students can create practical routines for converting effort and presence into tangible academic gains.

Data were randomly selected from Kaggle and the new dataset was created merging useful data.

<https://www.kaggle.com/datasets/ahmedaliraja/attendance-sheet-data-set-for-university>

<https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

<https://www.kaggle.com/datasets/himanshunakrani/student-study-hours?resource=download>