**­­­­­­­**

**Week 2**

Ajit Shrestha

Presidential Graduate School

TECH 405: Artificial Neural Network & Deep Learning

Prof. Acharya

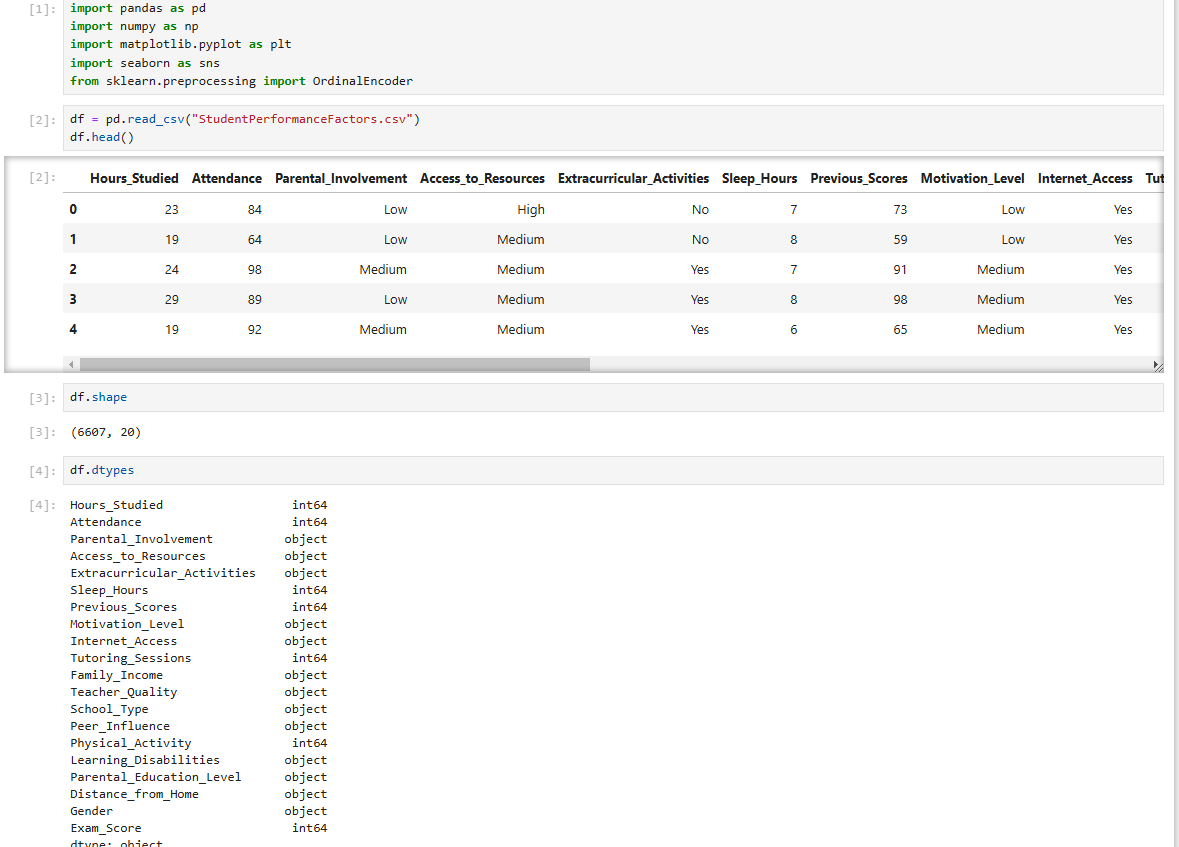
November 6, 2024

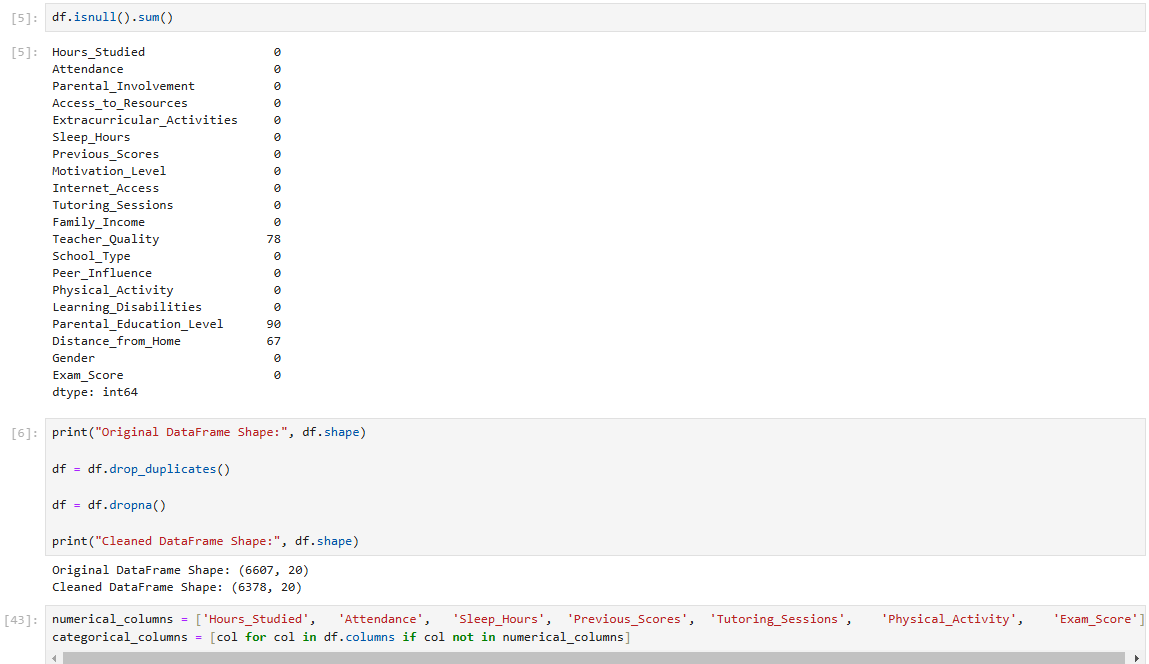
**Week 2**

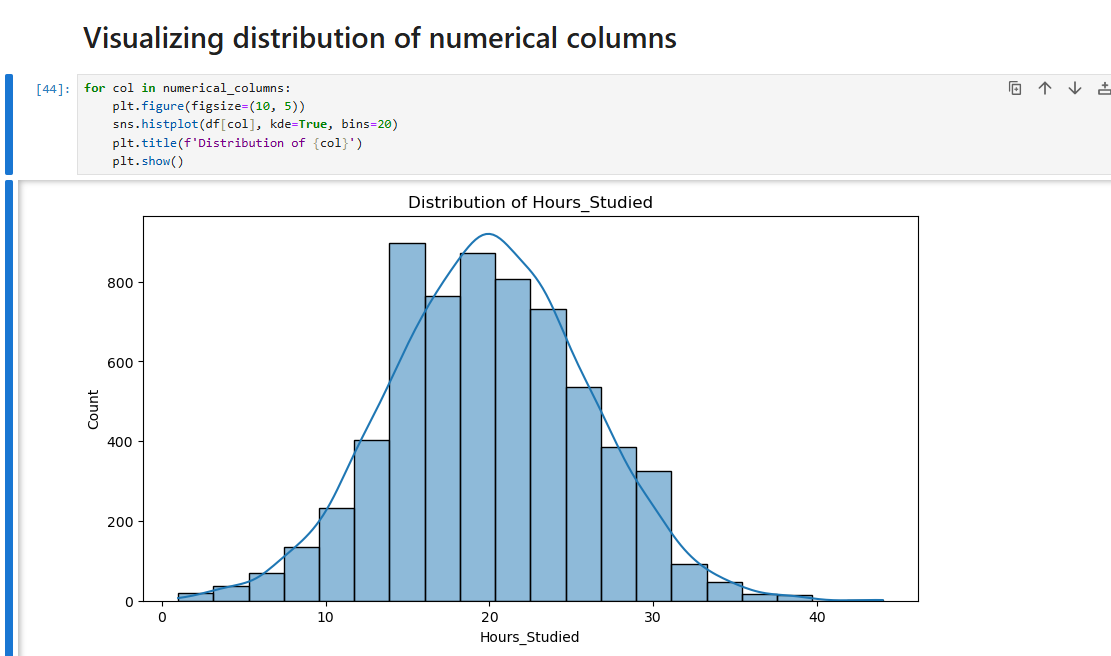
**Introduction**

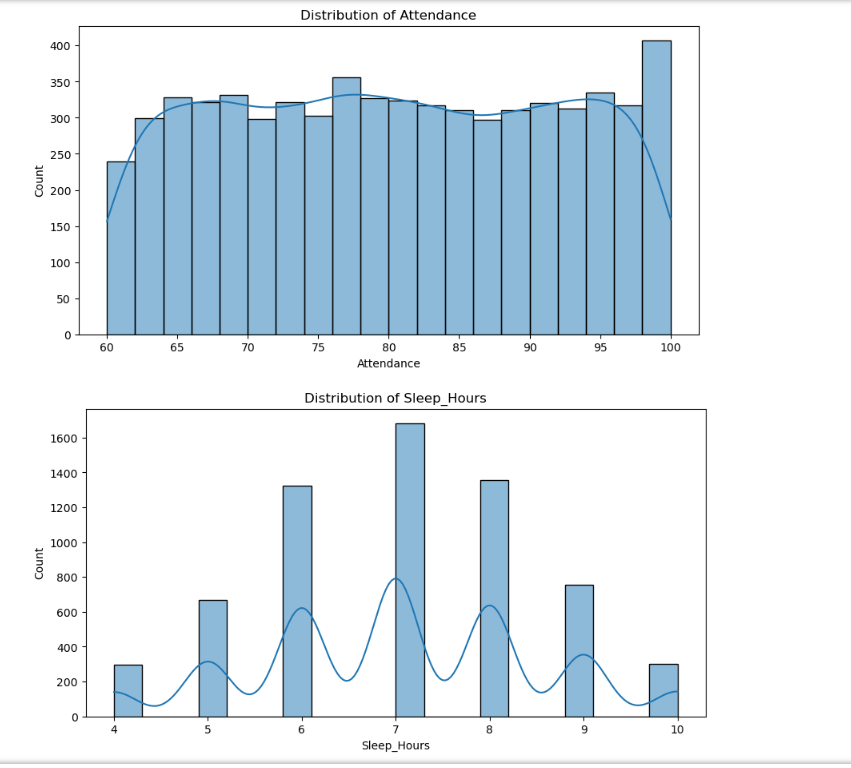
The dataset provides a thorough overview of the variables that affect students' exam scores, accounting for both environmental and personal impacts. It has a wide range of attributes and includes study habits (`Hours\_Studied`), school attendance (`Attendance`), and lifestyle attributes such as sleep patterns (`Sleep\_Hours`) and physical activity (`Physical\_Activity`). Variables such as `Parental\_Involvement`, `Family\_Income`, and `Parental\_Education\_Level` investigate parental factors giving context to the home environment, `Teacher\_Quality`, `School\_Type` and `Distance\_from\_Home` also provide insight into the educational setting. It also shows how social influences are portrayed through variables `Peer\_Influence` and `Extracurricular\_Activities` along with all of the other variables used to describe each student’s academic and living environment.

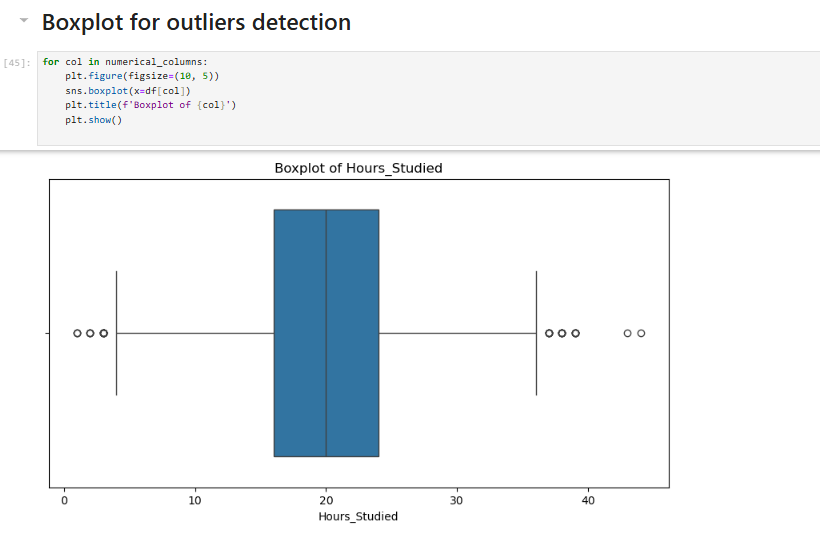
The aim for this analysis is to answer the question – what relationships exist between these factors and Exam\_Score, and what might be the most important factors contributing most to academic success. To that, I performed a thorough exploratory data analysis (EDA), with the intention of gaining insights into patterns, correlations and other relevant factors that affect performance. One of the aspects of the dataset’s scope is the possibility to assess several aspects of a student’s life that could shed light on how to use lifestyle, resources and motivation levels impact the result of the exam.

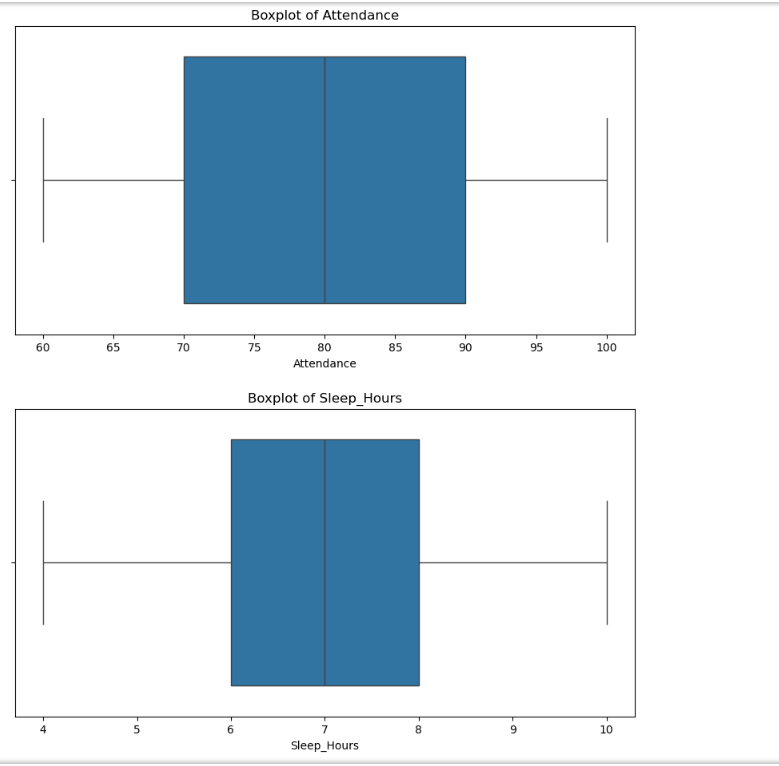
**Approach**

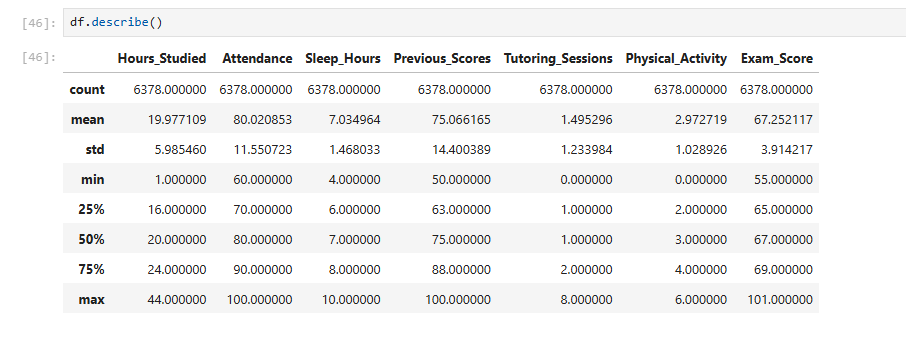


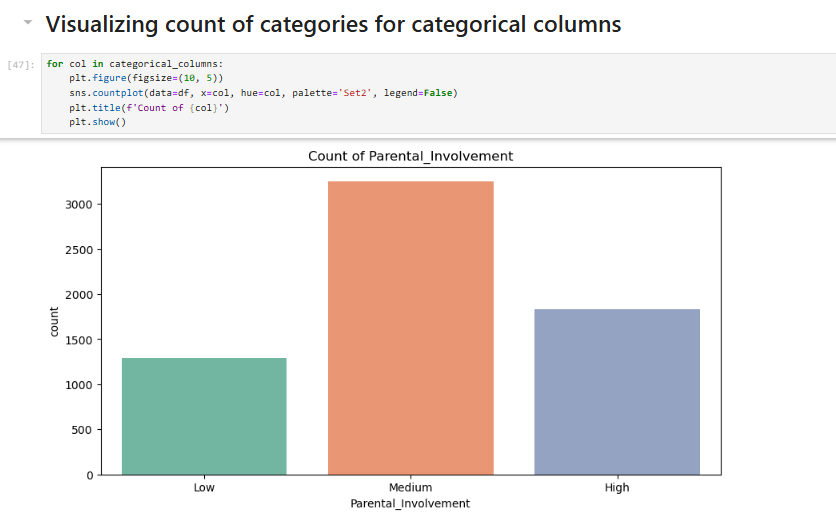


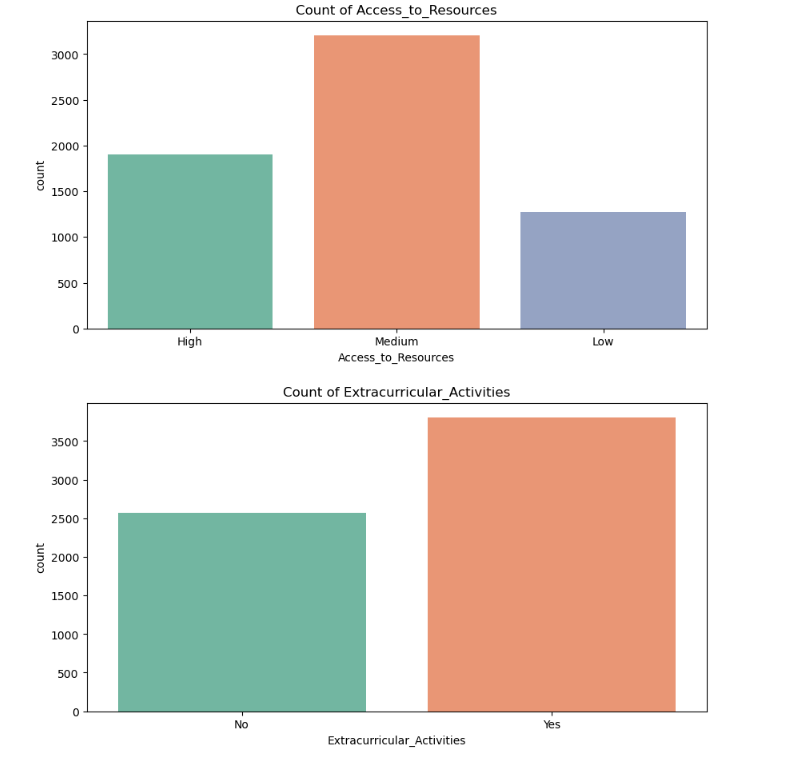


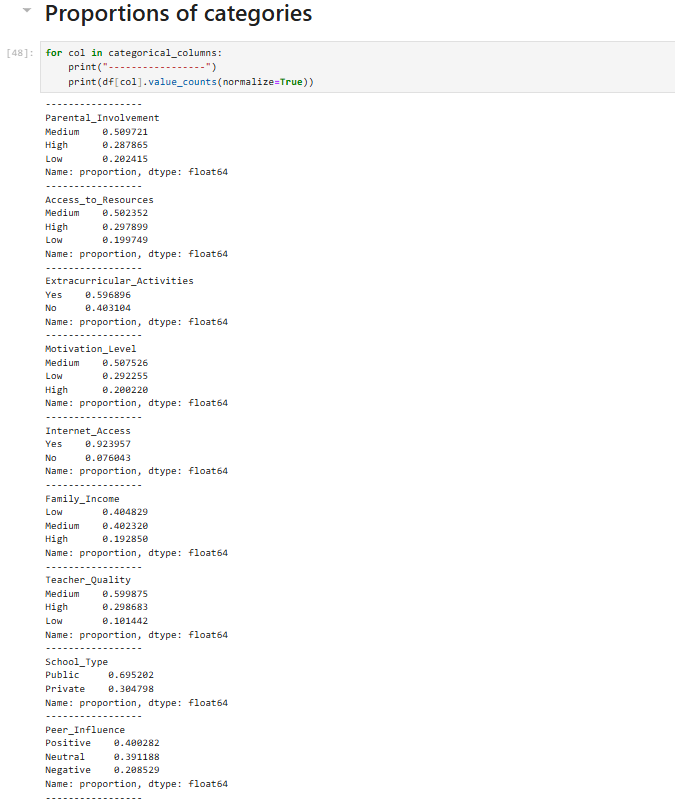


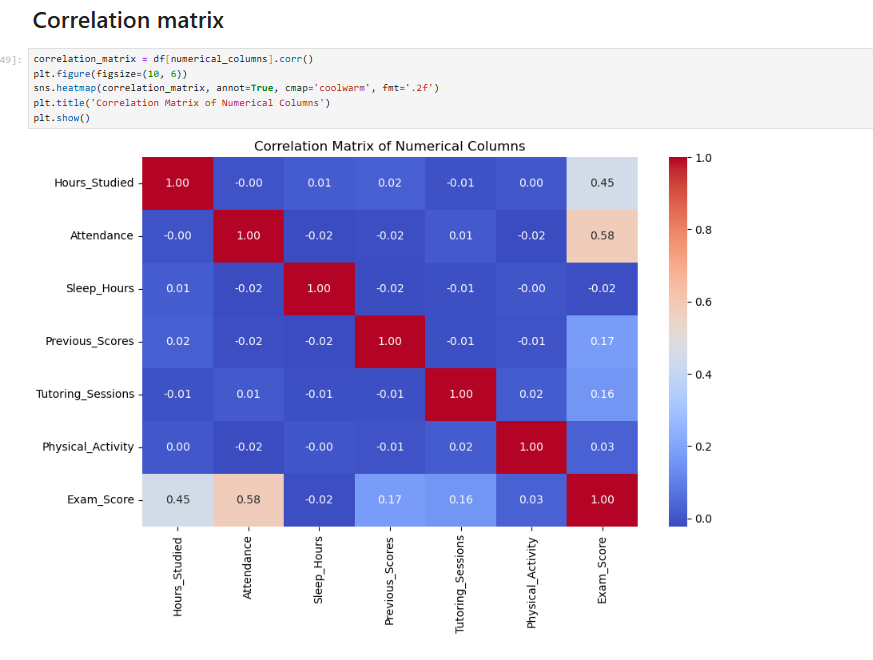


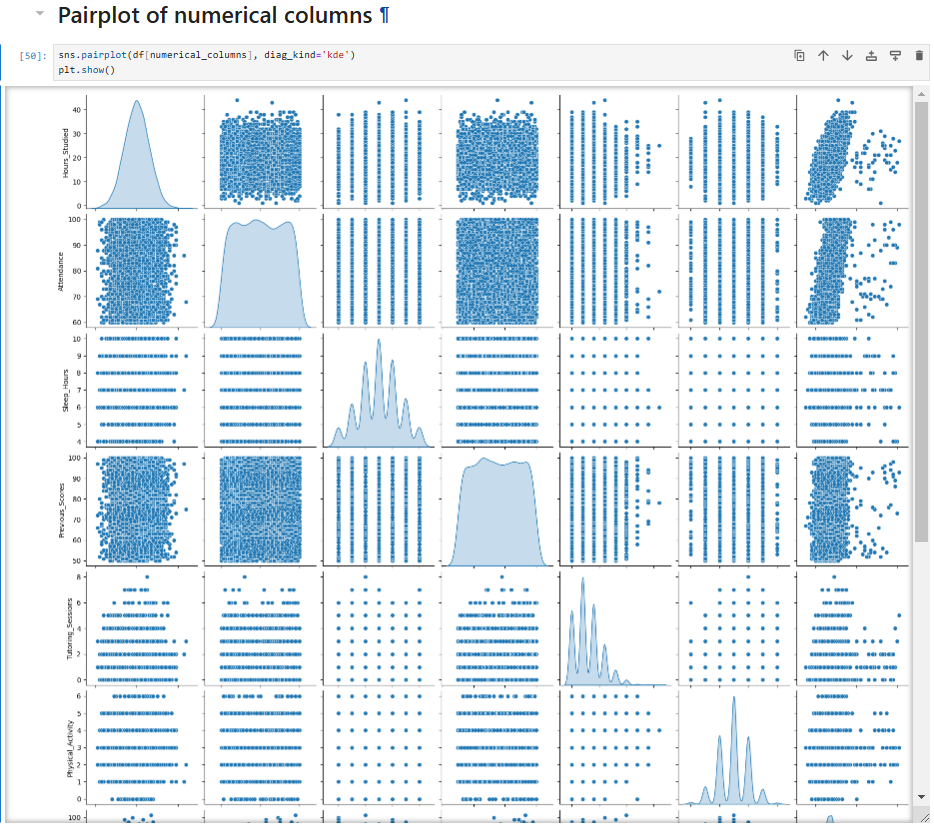


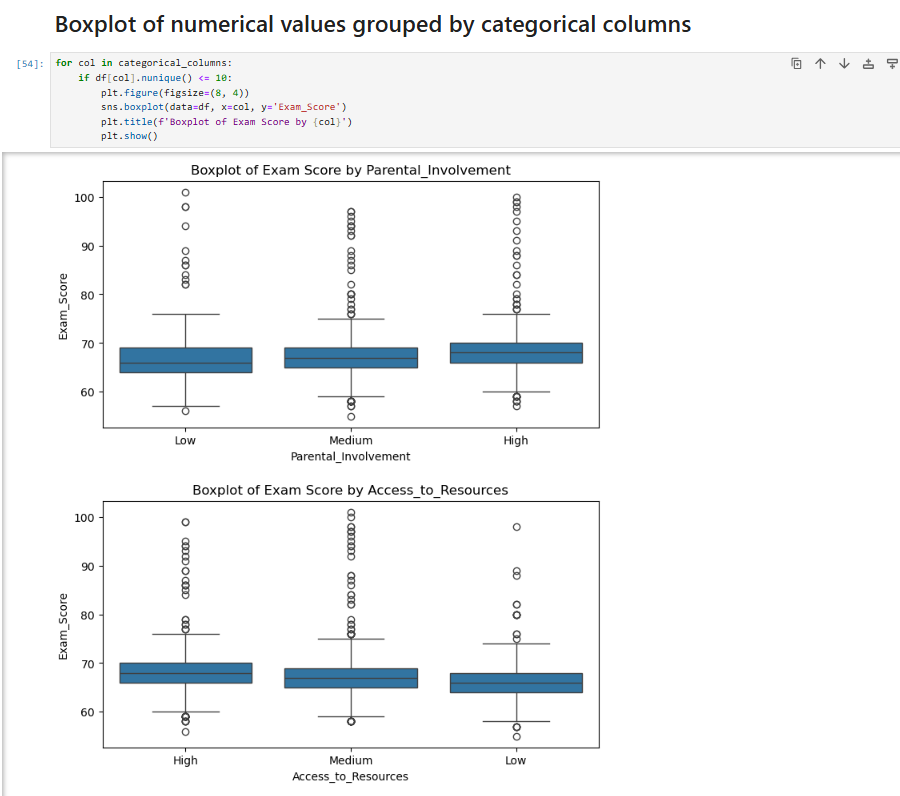


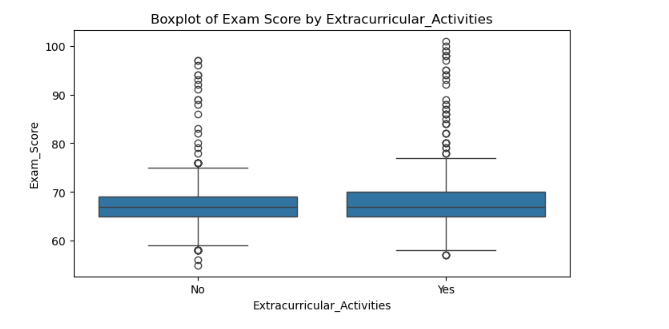


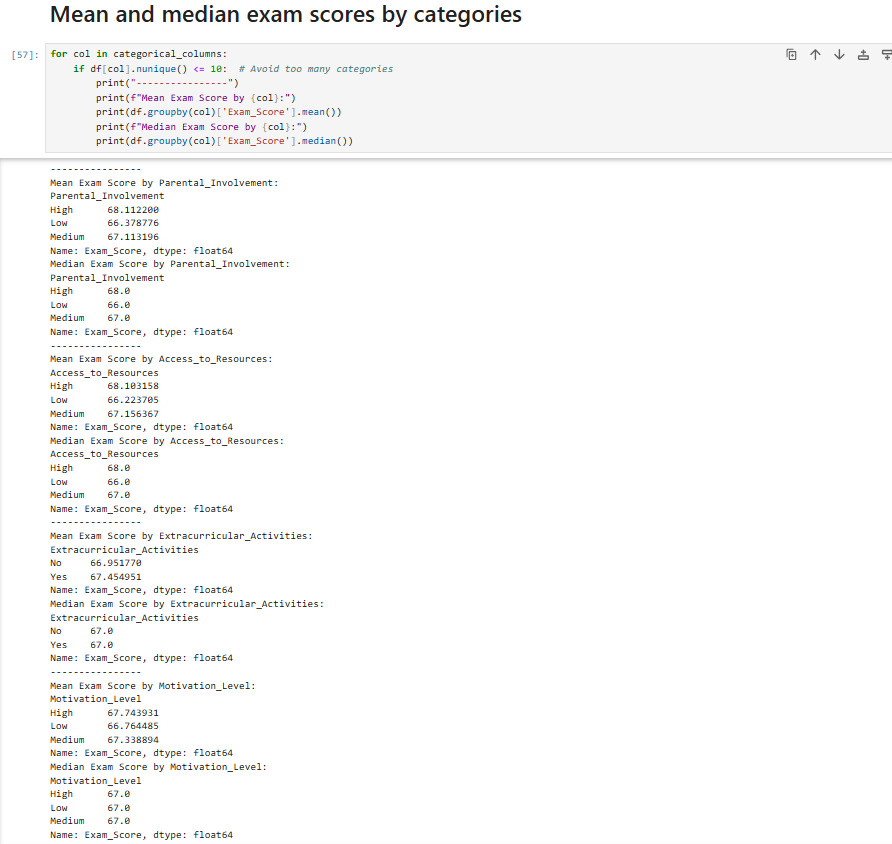












The analysis in this work primarily concentrated on conducting an initial exploration of the factors that influence student performance, especially around exam scores, through exploratory data analysis (EDA). Univariate and bivariate analysis are included in the EDA. I explored the distribution of individual variables (counts, proportions) using univariate analysis for both categorical and numerical columns. This gave a base to comprehend the degrees of distinction in engagement level, usability of resources, family foundation and lifestyle preferences amongst students.

Examining correlations between numerical variables, particularly their link to the numerical variable, `Exam\_Score`, was the main goal of bivariate analysis. With a correlation heatmap and pair plot we could bring out, how much the variables such as `Hours\_Studied`, `Attendance`, `Previous\_Scores` correlates with `Exam\_Score` in a linear manner. This one helped identify which factors may have a larger influence on academic performance, and therefore should be explored further, if you’d like to conduct more predictive modeling.  
  
**Findings**

The exploratory data analysis uncovers a couple of major lessons about what affects student performance. As we can see from the correlation heatmap `Attendance` and `Hours\_Studied` have the greatest correlations with `Exam\_Score` (0.58 and 0.48 respectively). As such, it all points to the importance of consistency with attendance and consistent study time and the resulting high probability of positive exam outcomes. Furthermore, it turns out that the other numerical variables, such as `Previous\_Scores` and `Tutoring\_Sessions`, show minimal correlations with the final `Exam\_Score`. While these factors might play a role in academic performance, they are less significant than attendance and study hours.

As most students have `Medium` levels of parental involvement, resource access and motivation, we examine this data. Over 60% of students take part in extracurricular activities, have internet access 92 percent of students are digitally connected. Interestingly, of these students, only 19 percent come from high income families and 40 percent of the students come from low-income brackets. With this distribution, we may learn something about socioeconomic factors that might lead to poorer performance by students and less access to resources.

Other significant discoveries are that `Public` school students (69%) far exceed `Private` school students, and that students nearby to school (59%) are concentrated. Relative to the students themselves, a huge majority of their parents have only attended high school, while the numbers drop below postgraduate. In such a case, parental education and income level could be immense when it comes to evaluating socio-economic determinants of academic success. These findings lay a solid foundation for further research into the combination of how these factors impact student performance so that more can be understood about which of the different attributes have the most substantial impact, and where interventions could be focused.

**Conclusion**

This analysis brings to light important factors leading to student exam performance, focusing on the interactions between academic engagement and socio-economic background. It also happens to be among the most significant predictors of exam success, namely `Attendance`, and `Hours\_Studied` which, if I may, suggest that consistent study habits and attendance is really key to academic success. This supports the hypothesis that easing development of disciplined study routines and promoting student presence can have a great effect on student ability.

In addition, Parental\_involvement and Parental\_Education\_Level are also shown to have some impact, however, this is less pronounced than direct academic engagement variables. Even today, socio-economic factors such as `Family\_Income` and `Access\_to\_Resources` still have a sizable influence, as students from less fortunate backgrounds frequently face extra obstacles to academic achievement. These findings suggest which attendance, study support, and access to resources to target in schools and in policymakers that work on engaging students from diverse backgrounds.

**GitHub Link:**

**References**

Sherpa, D. (2021, March 15). Introduction to univariate, bivariate and multivariate analysis. Medium. https://medium.com/analytics-vidhya/univariate-bivariate-and-multivariate-analysis-8b4fc3d8202c

Tandiallo, J. M. (2024, January 22). Exploratory data analysis (EDA) using Python: Basic examples about exploratory data analysis and data visualization in Python. Medium. https://medium.com/@teppan\_noodle/exploratory-data-analysis-eda-using-python-f85938cb1810