Mark Young

Data Science Programing

**Investigation of Student Performance Dataset**

**Overview**

This project investigates the *Student Performance* dataset, sourced from the UCI Machine Learning Repository, which contains data on student achievement in math and Portuguese courses from two Portuguese high schools. The dataset includes academic performance, demographic information, and behavioral factors collected via reports and questionnaires.

**Research Questions**The analysis centers on two key questions:

1. *What factors (e.g., study time, parental education) most influence student grades in math and Portuguese courses?*
2. *Are there noticeable differences in performance between male and female students for Portuguese and Math?*

**Methodology**The project utilized Python and Jupyter Notebooks for data cleaning, exploration, and analysis. Libraries like pandas, Matplotlib, and Seaborn were employed for data manipulation and visualization to uncover patterns and relationships.

**Key Findings**

* Certain factors, such as parental education level and study time, appear to significantly influence student grades.
* Notable performance differences were observed between male and female students, particularly in Portuguese courses.
* Insights can help identify students who may need additional support and predict academic outcomes.

**Future Considerations**Potential areas for further research include examining the impact of extracurricular activities, weekend alcohol consumption, and school travel time on student performance.

**Benefits**This analysis provides actionable insights for educators and policymakers to:

* Identify struggling students early and offer targeted support.
* Develop strategies to improve overall academic outcomes.