TopStudent_SelectionDataset.csv

📑 **Dataset Description: Student Performance Selection Data**

This dataset contains academic and demographic information of 25 students, with the goal of identifying and analyzing **high-performing students** based on their scores in **math, reading, and writing**.

📌 **General Overview:**

- **Total Records**: 1001 students
- **Purpose**: To analyze academic performance and identify good students based on scores, while considering factors such as gender, ethnicity, parental education, lunch type, and test preparation.

📊 **Columns and Descriptions:**

Column Name	Data Type	Description
gender	Categorical	Student's gender: male or female.
race/ethnicity	Categorical	Student's ethnic group (e.g., group A to group E). Can indicate socio-cultural background.
parental level of education	Categorical	Highest education level of the student's parent. Examples: high school, associate's degree, master's degree.
lunch	Categorical	Type of lunch program: standard or free/reduced. Acts as a proxy for economic background.
test preparation course	Categorical	Whether the student completed a test preparation course: completed or none.
math score	Numeric	Student's math exam score (0–100).
reading score	Numeric	Student's reading exam score (0–100).
writing score	Numeric	Student's writing exam score (0–100).

🎯 **Target Use Case:**

- To **cluster** or **classify** students based on academic performance.

- To **identify top-performing students** (for scholarships, rewards, support, etc.).
- To **analyze the effect of background factors** (like parental education or lunch type) on academic success.

📌 **Observations from Dataset:**

- The dataset includes both **standard** and **free/reduced** lunch categories, highlighting socioeconomic diversity.
- Students with **completed test preparation courses** often show **higher scores**.
- Students come from a mix of **parental education backgrounds**, ranging from some high school to master's degree.
- There is a mix of performance levels — from students scoring below 30 to those scoring above 90.

nsed under the [MIT License](LICENSE).