

Educational Data Analysis & Insights

Prepared By: Umesh Patidar
GitHub: <https://github.com/UmeshPatidar211>
Date: 1/1/2026

1. 🚀 Project Objective :

Clean, process, and analyze student exam score dataset.

Dataset : <https://www.kaggle.com/datasets/desalegngeb/students-exam-scores>

2. ⭐ Dataset Introduction :

The dataset provides demographic, educational and lifestyle features of students, along with academic performance scores. The primary goal is to clean & analyze how external factors influence student academic results.

3. Dataset Key Features 🤖 :

- **Gender** - Male / Female
- **EthnicGroup** - Student cultural background groups A-E
- **ParentEduc** - Education level of parents
- **LunchType** - Standard or free/reduced
- **TestPrep** - Completed / Incomplete
- **ParentMaritalStatus** - Married/Single/Divorced/Widowed
- **PracticeSport** - Regular / Sometimes / Never
- **WklyStudyHours** - Weekly study time frequency

4. Data Cleaning Process 👍 :

- ✓ Removed unnecessary columns & checked shape.
- ✓ Handled missing values with Mode/Median.
- ✓ Corrected datatypes & column inconsistencies.
- ✓ Verified no duplicate records.
- ✓ Converted categorical entries (None → incomplete).

5. Exploratory Data Analysis (EDA) 👇 :

- Ethnic Groups C & D dominate ~60% of dataset.
- Gender distribution is balanced: Female ≈ Male.
- Parent Education has the strongest score correlation.
- Sports participation is mostly 'Sometimes'.
- Math has lowest averages; Reading/Writing higher.

6. 📈 Dashboard Summary :

Feature	Observation
Parent Education	Higher education → Higher marks
Marital Status	No major impact on scores
Sports Practice	Occasional participation is highest
Ethnic Groups	C and D majority (imbalanced)
Gender	Balanced representation

7. Final Insights & Conclusion :

Parent education is the most influential factor for academic success. Marital status does not significantly impact performance, while participation in activities and study habits show trends but not definitive conclusions. Dataset is now fully cleaned and ready for modeling to build predictive systems.

