

# **Project 1: The Impact of Demographic Factors on Student Performance**

## **Context and Objectives:**

The project aims to explore how demographic factors influence student performance in standardized student tests. This dataset contains information on 202 students, focusing on their academic performance and demographic background. It includes key variables such as gender, race/ethnicity, and parental level of education, along with students' scores in math, reading, and writing. From these subject scores, total and average scores have been calculated to provide a broader view of overall academic performance. While most data is complete, a few entries are missing information on parental education or average scores. The exact source of the data is not explicitly stated, the structure and content strongly suggest that it is derived from educational performance data, similar to public datasets commonly used in academic research or standardized testing analysis, such as those found on Kaggle or the UCI Machine Learning Repository. The dataset is well-suited for educational analysis and is typically used in fields like education research, data science, and policy-making to explore how demographic factors may influence student achievement and to identify trends, disparities, or areas for intervention.

## **Key Variables:**

Gender: Binary variable (0 or 1, likely female/male), indicating the student's gender.

Race/Ethnicity: Categorical variable (Groups A, B, C, D, E), representing the student's racial/ethnic background.

Parental Level of Education: Categorical variable (associate's degree, bachelor's degree, high school, master's degree, some college, some high school), indicating the highest education level of the student's parents.

Math Score: Numerical score (0–100) for the student's performance in a math test.

Reading Score: Numerical score (0–100) for the student's performance in a reading test.

Writing Score: Numerical score (0–100) for the student's performance in a writing test.

Total Score: Numerical value, sum of math, reading, and writing scores.

Average Score: Numerical value, average of math, reading, and writing scores.

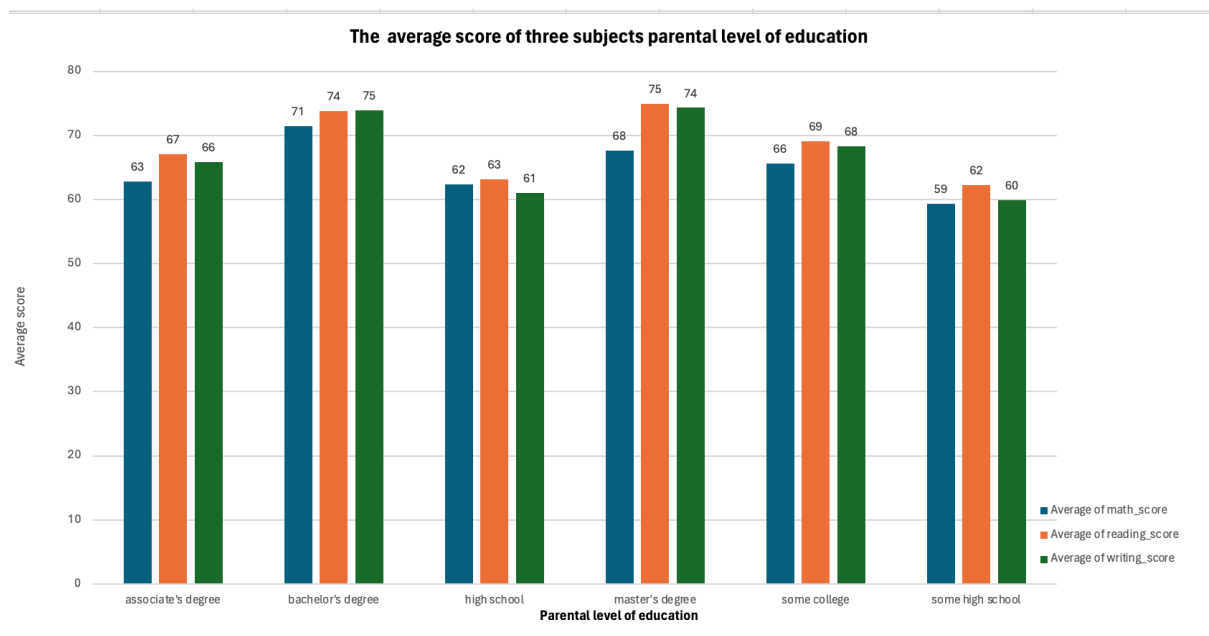
## **Missing values and Handling methods:**

Parental Level of Education: There were 3 missing values. These were filled using the mode (the most frequent value). This method is appropriate because this is categorical data, and using the most common category helps maintain the integrity of the dataset without bias.

Average Score: There were 4 missing values, which were handled by calculating the mean of the available math, reading, and writing scores. Using the average method is suitable for numerical data because it provides a central tendency that reflects the general performance of students without skewing the data too high or low.

### Analysis Finding:

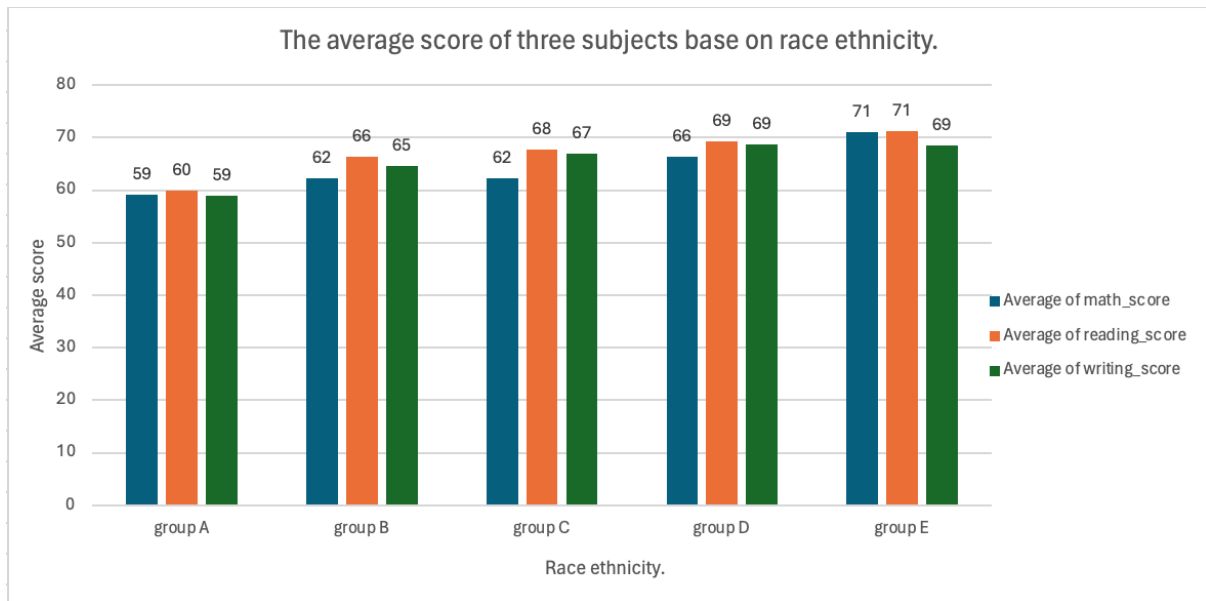
**Figure 1: The average score of three subjects based on parental level of education.**



- Students whose parents have higher education levels, such as bachelor's or master's degrees, consistently outperform those with parents who have lower education levels, with a notable gap in performance across all subjects. This insight highlights the role of parental education as a socioeconomic factor influencing academic success, suggesting that interventions targeting students from lower-educated households could help close achievement gaps.
- Students with parents holding a master's degree exhibit the strongest performance in reading and writing compared to other parental education levels, suggesting that advanced parental education particularly enhances literacy-related skills. This insight

suggests that advanced parental education (master's degree) may provide a unique advantage in literacy-based subjects, possibly due to increased exposure to reading and writing resources at home, informing targeted educational support for students from less-educated households.

**Figure 2: The average score of three subjects based on race/ethnicity.**



- Group E students demonstrate the highest academic performance across all subjects, while Group A students consistently have the lowest scores, indicating significant disparities in academic outcomes across racial/ethnic groups. This insight underscores potential systemic inequities in educational outcomes across racial/ethnic groups, suggesting a need for targeted support for underperforming groups like Group A to address achievement gaps.
- Group D students perform comparably to Group E in writing and reading, but their math scores are notably lower, highlighting subject-specific disparities within high-performing racial/ethnic groups. This insight reveals that even among high-performing groups, subject-specific disparities exist, suggesting that Group D students may benefit from targeted math interventions to align their performance with Group E, enhancing overall academic equity.

### **Conclusion and Implications:**

The analysis of the dataset reveals that higher parental education and specific racial/ethnic groups are linked to superior student performance in math, reading, and writing, with notable

gaps. These findings, supported by cleaned data and chart insights, imply a need for targeted interventions to address socioeconomic and racial disparities, such as enhanced support for students from lower-educated households and underperforming groups, alongside gender-specific strategies to balance subject performance, informing equitable educational policies.