Name: Le Chau Thanh Dat IRN: 2232300005 GitHub Portfolio Link: https://github.com/Nora-Kun		**Comment**
Criteria	**Score**	
Data Cleaning and Preparation		
Data Analysis and Insight		
GitHub Portfolio Quality		
Total		

MIS-311 - Introduction to Business Analytics

Part 1: Data Analysis and Insight

I will use Excel to perform **Exploratory Data Analysis (EDA)** on the dataset **Student Performance**

1. Data Overview

The **Student Performance** dataset provides information on students' demographic backgrounds and their academic achievements by three subjects: math, reading, and writing. It contains **202 rows** and **8 columns,** each row represents for an individual student.

Key Variables:

Race/Ethnicity (object): Different racial groups.

Parental Level of Education (object): The educational background of the student's family.

Math Score (int): Performance in math,

Reading Score (int): Performance in reading

Writing Score (int): Performance in writing skills

2. Data Cleaning

I use the **Go to** command (Ctrl + G) in Excel to to find the missing value in dataset. Results:

- 4 missing values in **average_score** column
- 3 missing values in **parental_level_of_education** column

Solutions:

- Missing values in **average_score** were recalculated using the available component scores (math, reading, and writing), as this variable is derived rather than directly observed.

- Missing values in parental level of education were retained and labeled as "**Unknown**" to avoid introducing bias through arbitrary imputation.

- Finally, I use the Remove Duplicates command to find out 3 duplicate rows and remove it. Now the dataset has 199 rows left.
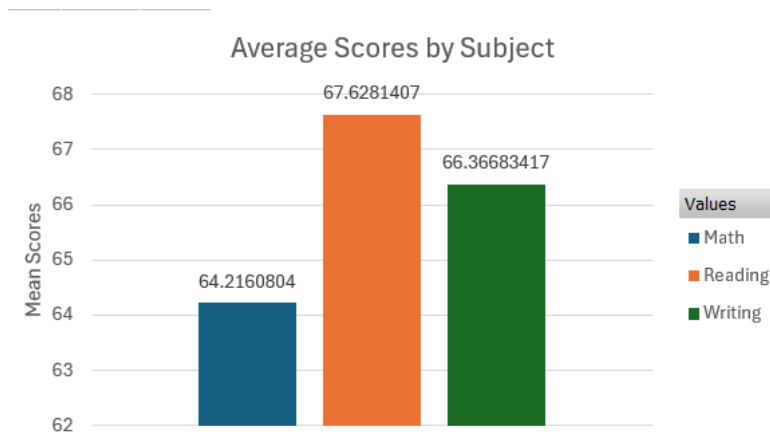
3. Descriptive Statistics

I use Data analysis tools in Excel to get **descriptive statistics** for student's score in math, reading, writing, total and average scores.
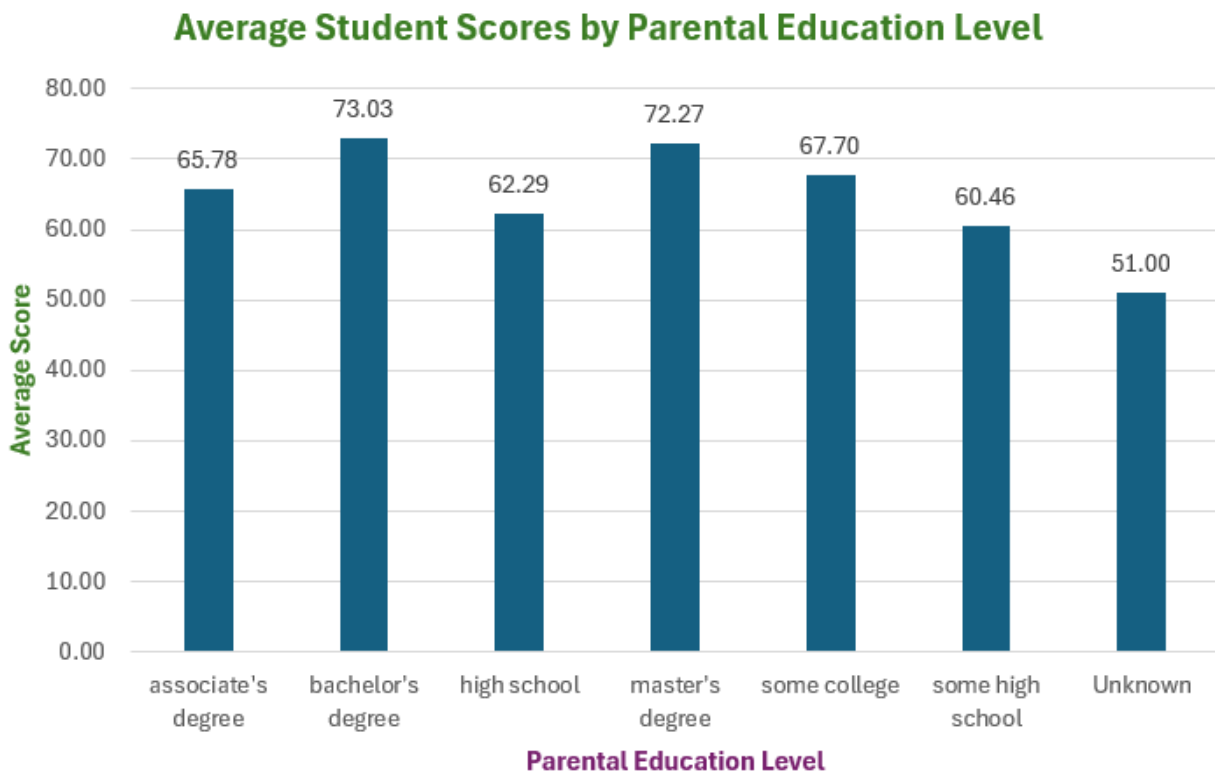
math_score			reading_score			writing_score			total_score			average_score	
Mean	64.21608		Mean	67.62814		Mean	66.36683		Mean	198.2111		Mean	66.07035
Standard E	1.127361		Standard E	1.118896		Standard E	1.152591		Standard E	3.274682		Standard E	1.091561
Median	65		Median	68		Median	68		Median	199		Median	66.33333
Mode	58		Mode	67		Mode	74		Mode	151		Mode	50.33333
Standard D	15.90339		Standard D	15.78397		Standard D	16.2593		Standard D	46.19508		Standard D	15.39836
Sample Va	252.9177		Sample Va	249.1337		Sample Va	264.3648		Sample Va	2133.986		Sample Va	237.1095
Kurtosis	0.946461		Kurtosis	-0.14052		Kurtosis	0.072614		Kurtosis	0.3414		Kurtosis	0.3414
Skewness	-0.36683		Skewness	-0.23144		Skewness	-0.33338		Skewness	-0.34874		Skewness	-0.34874
Range	100		Range	83		Range	90		Range	272		Range	90.66667
Minimum	0		Minimum	17		Minimum	10		Minimum	27		Minimum	9
Maximum	100		Maximum	100		Maximum	100		Maximum	299		Maximum	99.66667
Sum	12779		Sum	13458		Sum	13207		Sum	39444		Sum	13148
Count	199		Count	199		Count	199		Count	199		Count	199

From the result, we can see that students scored around 64.22 in math, 67.63 in reading, and 66.37 in writing, showing slightly better performance scores in reading compared to other subjects. Overall, the maximum score is 100 while the minimum is quite low, indicating the wide range of scores. Lastly, the standard deviation of about 15 to 16 shows some variation in student performance.

Key Insight:

Average Scores by Subject

Values
- ■ Math
- ■ Reading
- ■ Writing

Math: 64.2160804
Reading: 67.6281407
Writing: 66.36683417

Descriptive statistics show that students achieve higher average scores in reading and writing compared to mathematics. The lower mean score in mathematics suggests that this subject may pose greater challenges for students. This insight highlights potential gaps in quantitative skills compared to language-based subjects.

Average Student Scores by Parental Education Level

Parental Education Level	Average Score
associate's degree	65.78
bachelor's degree	73.03
high school	62.29
master's degree	72.27
some college	67.70
some high school	60.46
Unknown	51.00

Descriptive statistics indicate that students whose parents have higher levels of education tend to achieve higher average scores. A gradual increase in mean scores is observed as parental education level rises. This suggests a potential association between parental educational background and student academic performance.