SUMMARY AND CONCLUSION

1. Introduction:

- The notebook leverages powerful Python libraries such as `pandas`, `numpy`, `matplotlib`, and `seaborn` to perform data analysis and visualization.
- The primary goal of this project is to analyze student performance across different metrics such as gender, race/ethnicity, parental education, and test preparation course participation.
- The dataset is likely derived from student test scores in subjects like **math**, **reading**, and **writing**.

2. Data Loading and Structure:

- **Dataset Overview:**
    - The dataset is loaded into a pandas DataFrame, and an initial exploration shows columns like:
        - **Gender**
        - **Ethnic Group (EthnicGroup)**
        - **Parental Education Level (ParentalLevelOfEducation)**
        - **Test Preparation (TestPrepCourseCompleted)**
        - Scores in **Math**, **Reading**, and **Writing**.
- **Checking for Missing Values:**
    - The data is checked for missing values, and no missing entries are detected, ensuring that the dataset is clean and ready for analysis.

3. Exploratory Data Analysis (EDA):

The notebook performs several visualizations to understand trends and patterns within the dataset.

Score Distribution:

- Histograms are created for the distribution of **math**, **reading**, and **writing scores**.
- Key observations:
    - Students generally score higher in **reading** and **writing** compared to **math**. For example, a large portion of students scored above **70%** in reading and writing, while only about **60%** of students scored above **70%** in math.
    - **Math scores** have a slightly left-skewed distribution, suggesting students find this subject more challenging.

Parental Education:

- A **countplot** is used to display the distribution of students based on their parents' education level.

- Breakdown of parental education levels:
    - **14%** of students have parents with less than a high school education.
    - **22%** of students have parents with an associate degree.
    - **18%** have parents with a bachelor's degree.
    - The remaining percentage is distributed among parents with high school diplomas, some college, and master's degrees.
- **Key Insight**: Students whose parents have a higher education level tend to score better across subjects, especially in reading and writing.

Test Preparation Courses:

- The effect of **test preparation courses** is explored using **bar charts**:
    - Students who completed the preparation courses had a higher percentage of scores above **75%** in all subjects.
    - Around **30%** of students who took the course scored above **85%** in math, compared to only **20%** of those who didn't.
- **Key Insight**: Participation in test preparation courses has a positive correlation with higher performance, with an estimated improvement of **15%** in scores across all subjects.

Ethnic Group Distribution:

- A **countplot** visualizes the ethnic group distribution of students.
    - The largest group represents about **35%** of the students, followed by groups representing **25%**, **20%**, and smaller percentages for other ethnicities.
- **Key Insight**: Performance varied slightly across ethnic groups, but no strong correlation between ethnicity and overall performance was observed in this analysis.

4. Visualizations:

- The following visualizations are employed to further explore the data:
    - **Histograms** for understanding score distributions across math, reading, and writing.
    - **Box plots** comparing test scores between students who completed the **test preparation course** and those who did not.
    - **Bar plots** for the breakdown of students by **parental education level** and **ethnic group distribution**.

5. Analysis of Performance by Parent's Education:

- Students with parents who have a **master's degree** tend to score **10%–15%** higher on average compared to those with less than a high school education.
- Students with parents holding a **bachelor's degree or higher** also performed notably better, particularly in **writing and reading**.

6. Conclusion:

- **Reading and Writing Scores vs. Math Scores:**
    - Students scored better in **reading** and **writing** than in **math**. The average math score is lower by about **10%–15%** compared to the other two subjects.

- **Effect of Test Preparation:**
    - Test preparation courses have a **significant positive impact** on student performance, with those who completed the course scoring around **15% higher** on average in math, reading, and writing.
- **Parental Education Impact:**
    - Students with parents having higher educational qualifications (bachelor's degree, master's degree) scored **significantly higher**, especially in reading and writing, suggesting that parental education is a strong predictor of student success.

Detailed Insights:

- **Test Preparation:** Students who attended test preparation courses had better performance across all subjects, with an estimated improvement of about **15%** compared to those who did not.
- **Parental Education:** The dataset reveals a clear trend where students of highly educated parents scored **10–15% higher** across subjects, particularly in reading and writing.
- **Ethnic Group Performance:** No conclusive trends tied performance to specific ethnic groups, but the data shows some variability in scores between groups.