Executive Summary

1. **Data Loading and Setup**:
    - The notebook initializes essential libraries (Pandas, NumPy, Seaborn, Matplotlib) for data analysis and visualization. It loads a dataset titled "Student_Scores.csv" to assess students' performance metrics and sets the environment for further exploration.
    - Key checks include confirming the directory path and ensuring the data file is accessible for analysis.
2. **Data Quality Check**:
    - This section inspects the dataset for missing values and duplicate entries to ensure data integrity. There is also a verification of the dataset structure with df.describe() and df.info() to understand the numerical and categorical variables.
    - Initial findings reveal that data cleaning will be required, such as handling missing values and potential duplicates.
3. **Data Cleaning**:
    - Specific cleaning actions include removing an unnecessary "Unnamed" column and checking for redundant data. This prepares the dataset for more accurate analysis by reducing irrelevant information.
4. **Gender Distribution Analysis**:
    - This part visualizes the gender distribution within the dataset, using a bar plot to compare the counts of male and female students.
    - The observation shows a higher number of females in the dataset, suggesting a slight demographic imbalance.
5. **Impact of Parental Education on Scores**:
    - A heatmap visualization explores the relationship between parents' education levels and students' average scores in Math, Reading, and Writing.
    - The analysis indicates a positive correlation between higher parental education levels and better student performance, highlighting the influence of educational background on student outcomes.
6. **Parental Marital Status and Student Performance**:
    - This section uses a heatmap to examine if marital status affects student performance. Scores are grouped by marital status to calculate mean values.
    - Findings suggest no significant impact of parental marital status on student scores across all three subjects.
7. **Ethnic Group Distribution**:

- The notebook analyzes the ethnic group distribution within the dataset, both with a pie chart and a bar chart, to visualize the proportion and count of students from each ethnic group (A, B, C, D, E).
- This section establishes a demographic breakdown of ethnic representation without analyzing performance metrics across groups.

8. **Effect of Sports Practice on Scores**:
    - This segment examines the potential correlation between practicing sports and academic performance in Math, Reading, and Writing.
    - The findings indicate minimal or negligible impact of sports practice on academic scores, suggesting that sports involvement does not significantly influence academic outcomes.

Conclusion

The notebook provides a detailed exploration of factors potentially affecting student scores, including demographics, parental background, and extracurricular activities. While parental education levels appear influential on student performance, factors such as marital status, sports practice, and ethnicity show limited or negligible impact on scores. This analysis offers valuable insights for educational stakeholders interested in the predictors of student success.