Project Report

Title of Project:

"Analyzing Key Factors Influencing Student Exam Scores to Enhance Educational Outcomes"

Introduction:

This report presents an analysis of student exam scores from a comprehensive dataset. The objective is to identify the key factors influencing academic performance and provide actionable insights for educational institutions to enhance student outcomes. The analysis covers various dimensions such as gender distribution, the impact of parental education and marital status, score distributions, and the distribution of ethnic groups.

Business Question:

The primary business question addressed in this report is: "What are the key factors influencing student exam scores, and how can educational institutions leverage this information to enhance student performance?"

Analysis:

The analysis involved several key steps:

1. Data Cleaning and Preparation:

- The dataset was read, inspected, and cleaned. An unnecessary column was removed to ensure the data's integrity.
- Initial inspection involved checking for missing values and understanding the dataset's structure using descriptive statistics.

2. Gender Distribution:

- A count plot was created to visualize the distribution of genders in the dataset.
- The analysis revealed a nearly equal distribution of male and female students, with females slightly outnumbering males.

3. Effect of Parent's Education on Student's Score:

- The mean scores for Math, Reading, and Writing were grouped by the parent's education level.
- A heatmap was plotted to visualize the relationship between parent's education and student's scores, highlighting significant trends.

4. Effect of Parent's Marital Status on Student's Score:

- The mean scores for Math, Reading, and Writing were grouped by the parent's marital status.

- A heatmap was plotted to show the relationship between parent's marital status and student's scores.

5. Comparison of Student's Scores using Box Plots:

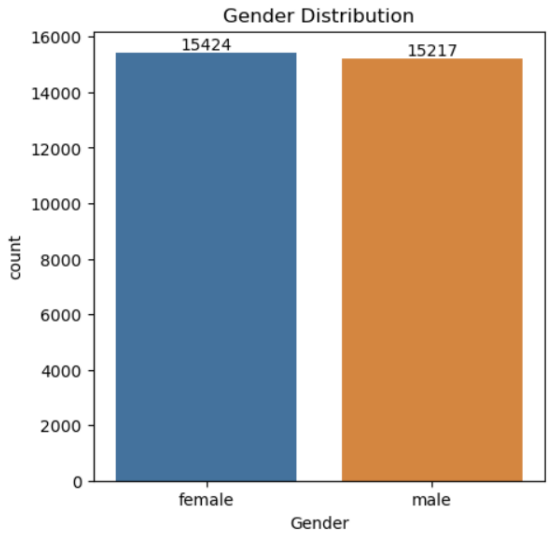
- Box plots were used to compare the distribution of scores in Math, Reading, and Writing.
- This analysis helped in identifying the spread and outliers in the scores for each subject.

6. Distribution of Ethnic Groups:

- The unique ethnic groups were identified, and their counts were calculated.
- A pie chart was plotted to show the distribution of ethnic groups in the dataset.

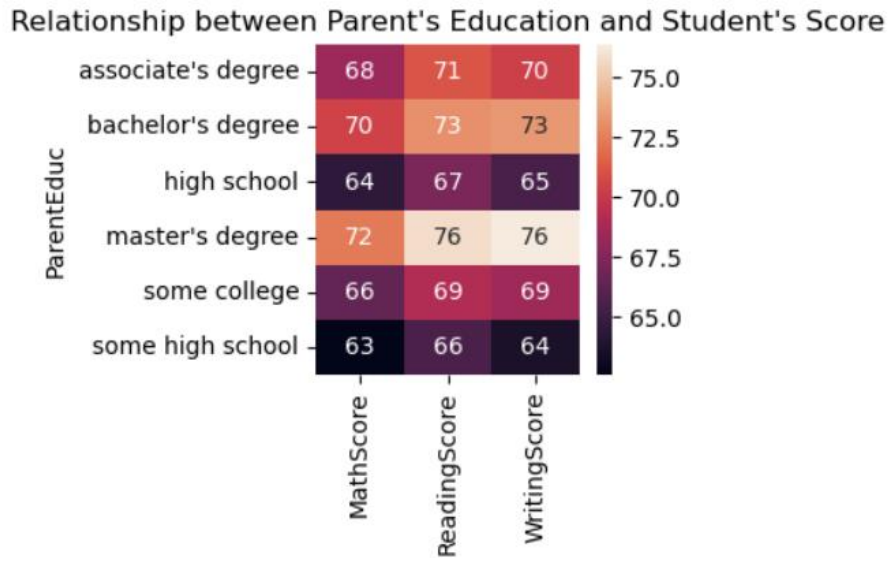
Key Findings:

1. Gender Distribution:



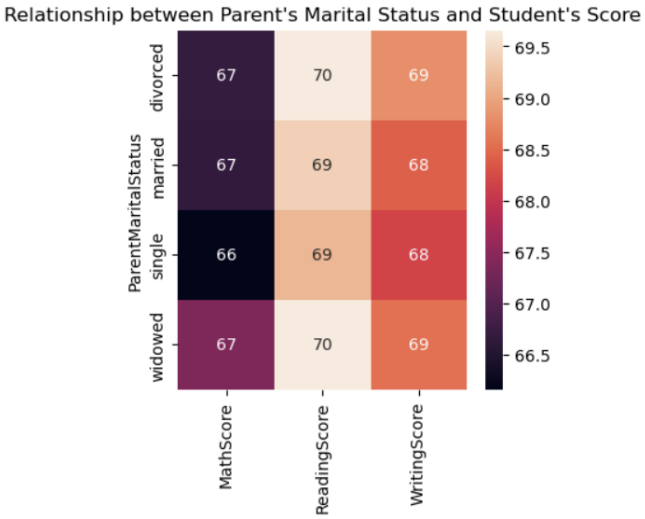
The gender distribution analysis showed that the number of female students (15424) slightly exceeds that of male students (15217).\

2. Parent's Education Impact:



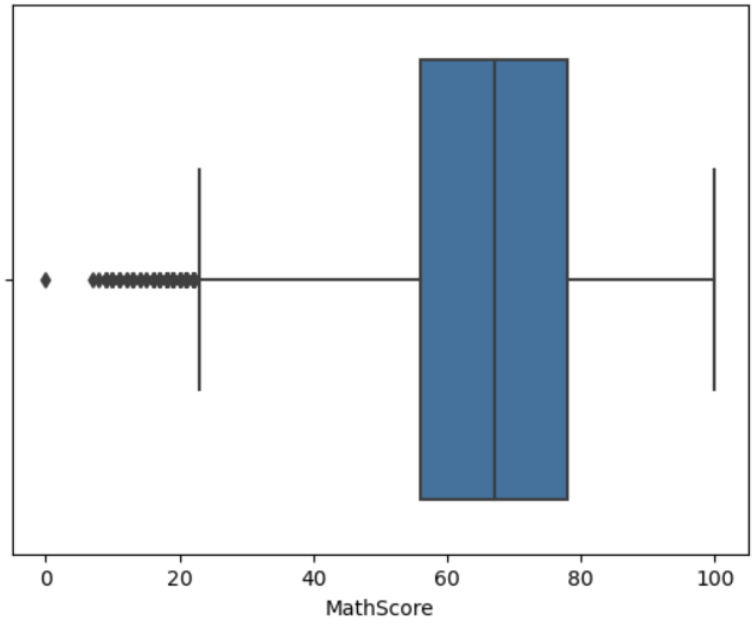
The heatmap analysis revealed a positive correlation between parental education levels and student scores. Students whose parents have higher education levels (master's or bachelor's degrees) tend to perform better in Math, Reading, and Writing.

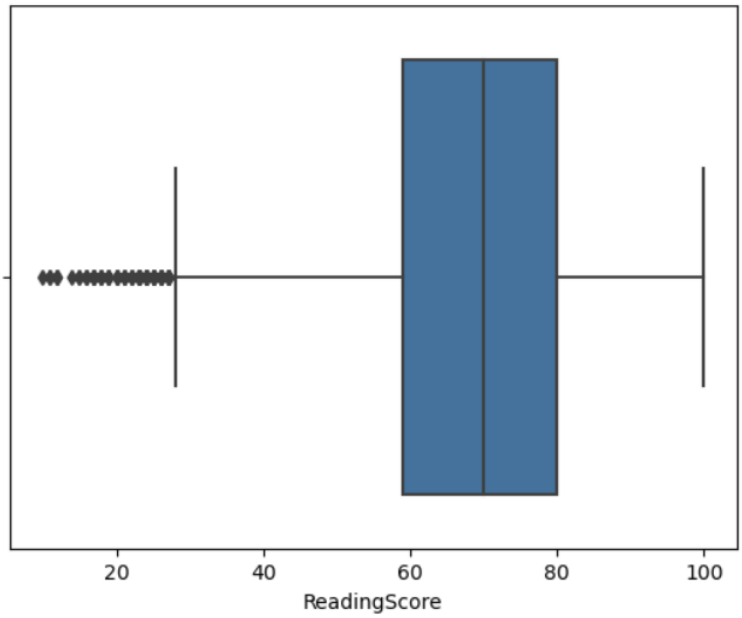
3. Parent's Marital Status Impact:

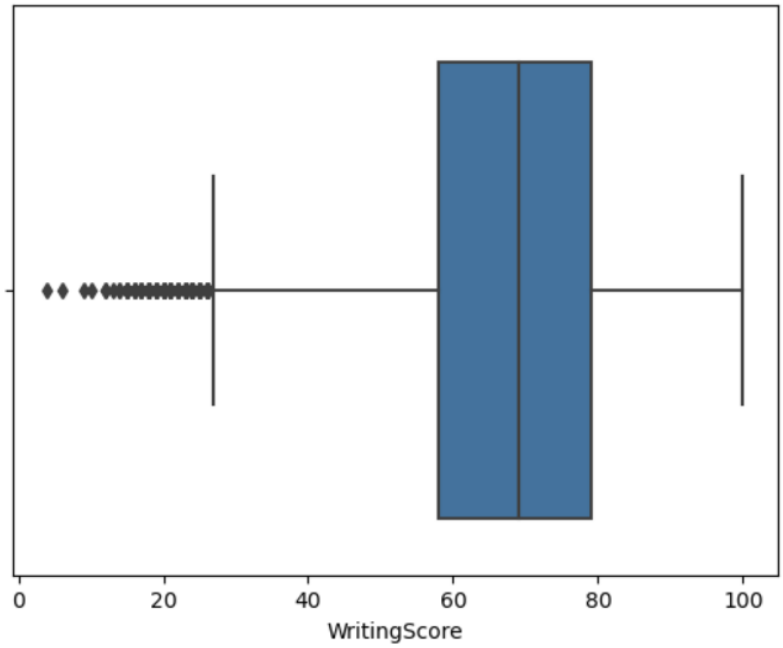


The analysis indicated that the marital status of parents has a negligible impact on student scores. All categories (divorced, married, single, widowed) show similar average scores.

4. Score Distribution:


0
20
40
60
80
100
MathScore

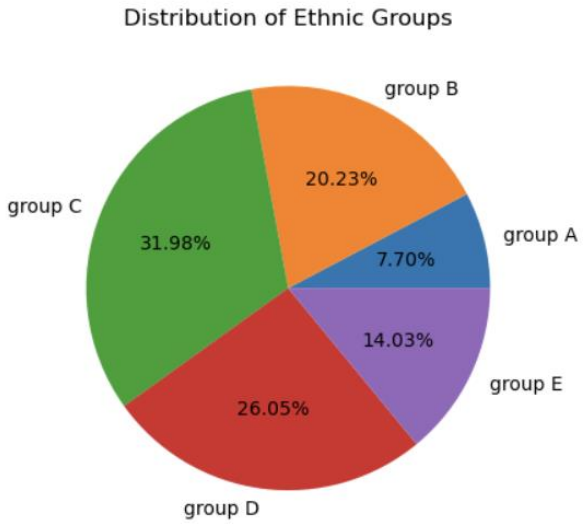

20
40
60
80
100
ReadingScore



Box plots for Math, Reading, and Writing scores highlighted that students tend to score lower in Math compared to Reading and Writing. Math scores show a wider spread with more outliers, indicating greater variability in performance.

5. Ethnic Group Distribution:

The pie chart analysis demonstrated that Group C is the largest ethnic group (31.98%), followed by Group D (26.05%), Group B (20.23%), Group E (14.03%), and Group A (7.70%).



Results:

The analysis revealed several critical insights:

1. Parental Education: Higher parental education levels correlate with better student performance across all subjects.

2. Marital Status: The marital status of parents has minimal impact on student scores.

3. Gender Distribution: There is a slight female majority in the dataset.

4. Score Variability: Students exhibit greater variability in Math scores compared to Reading and Writing.

5. Ethnic Diversity: The dataset is diverse, with significant representation from multiple ethnic groups.

Recommendations:

Based on the findings, the following recommendations are made:

1. Enhance Study Support Programs: Schools should provide additional resources and support for students to encourage more effective study habits, particularly in Math. This could include study groups, tutoring sessions, and access to online resources.

2. Parental Engagement Initiatives: Educational institutions should develop programs to engage parents, especially those with lower education levels, to support their children's academic journey.

3. Balanced Extracurricular Engagement: While encouraging participation in extracurricular activities, schools should guide students to maintain a balance between these activities and their academic responsibilities.

4. Address Socio-Economic Disparities: Schools should provide additional support to students from lower socio-economic backgrounds, ensuring they have access to the same resources and opportunities as their peers.

5. Cultural Sensitivity Programs: Implement programs that promote cultural awareness and sensitivity to ensure an inclusive environment for students from diverse ethnic backgrounds.

Conclusion:

This analysis highlights the importance of parental education and targeted support programs in enhancing student performance. The insights gained can inform policy decisions and educational strategies aimed at improving academic outcomes for all students. By addressing the identified factors and implementing the recommended interventions, educational institutions can create a

more supportive and inclusive learning environment, ultimately leading to better academic performance and student success.

References:

- Students exam scores: Extended dataset. (2023, April 14). Kaggle. https://www.kaggle.com/datasets/desalegngeb/students-exam-scores
- Obregon, A. (2024, April 27). Data Analysis with Pandas and NumPy | Medium. Medium. https://medium.com/@AlexanderObregon/how-to-use-pandas-and-numpy-for-data-analysis-in-python-0e650e9d0940
- VanderPlas, J. (n.d.). Visualization with Seaborn | Python Data Science Handbook. https://jakevdp.github.io/PythonDataScienceHandbook/04.14-visualization-with-seaborn.html
- Pyplot tutorial — Matplotlib 3.9.1 documentation. (n.d.-b). https://matplotlib.org/stable/tutorials/pyplot.html