

Student Scores Data Analysis

Introduction:

This Python project focuses on the analysis of student scores using a dataset containing various attributes such as gender, ethnic group, parent education, lunch type, test preparation, parent marital status, sports practice, birth order, number of siblings, transport means, weekly study hours, and scores in math, writing, and reading.

Dataset Loading:

The project begins with the loading of the dataset, "student_scores.csv," using the Pandas library. The dataset is then examined to identify its structure and content.

Data Cleaning:

The cleaning process involves handling missing values, removing an unnamed column, and correcting values in the "WklyStudyHours" column. The dataset's statistical summary and the count of missing values are explored to ensure data integrity.

Data Analysis:

Gender-wise Distribution:

A count plot is created to visualize the distribution of students based on gender. The analysis reveals a slight majority of females in the dataset.

Impact of Parental Education:

Grouping the data by "ParentEduc," a heatmap is generated to showcase the average scores in writing, math, and reading. The analysis indicates that students with parents holding master's degrees tend to have higher average scores compared to other educational backgrounds.

Effect of Parent Marital Status:

Another heatmap is created, grouping data by "ParentMaritalStatus," to explore its potential impact on students' scores. The analysis suggests that parent marital status has a negligible or no significant impact on student performance.

Subject-wise Score Distribution:

Boxplots are employed to visually represent the distribution of scores in math and writing. The analysis reveals that the minimum range of math scores is lower compared to scores in the other two subjects.

Distribution of Ethnic Groups:

A pie chart illustrates the distribution of students across different ethnic groups. The analysis indicates that Group C contributes the highest number of students, followed by Groups D, B, and others.

Conclusion:

This data analysis project provides valuable insights into various factors influencing student scores, such as gender distribution, parental education, marital status, and ethnic group. The visualizations and statistical analyses contribute to a better understanding of the dataset, aiding in drawing meaningful conclusions about the relationships between different variables and academic performance.