Student Performance Analysis

Documentation

TEAM MEMBERS:
KAICHARLA NAVYA
CHALUMURI URMILA
GOKIVADA PARIMALA
BORRA SRAVANI
TANALA DEVIKA

Student performance analysis

Introduction

In today's rapidly evolving educational landscape, the pursuit of academic excellence and the quest to improve student outcomes have never been more vital. With the advent of data analytics, educational institutions now possess a powerful tool to gain deeper insights into student performance, identify areas of improvement, and tailor interventions that foster meaningful growth.

This documentation presents a comprehensive analysis of student performance utilizing data analytics techniques. By harnessing the wealth of data generated within educational environments, we aim to uncover patterns, trends, and correlations that hold the potential to drive informed decision-making and targeted strategies. Through a systematic exploration of student records, attendance records, and test scores, we seek to provide educators, administrators, and stakeholders with actionable insights that can reshape teaching methodologies, student support systems, and overall educational experiences.

As we delve into the intricate interplay between data and academic performance, this documentation will elucidate the methodologies employed, the key metrics under examination, and the visual representations that facilitate a clearer understanding of complex data sets. By contextualizing findings within the broader educational landscape, we aim to catalyze meaningful discourse and inspire initiatives that lead to tangible enhancements in student outcomes.

 By embarking on this data-driven journey, we acknowledge the transformative potential of data analytics in shaping the educational landscape. Through evidence-based recommendations and a deep dive into student performance dynamics, this documentation aims to empower educators and decision-makers with the tools they need to steer education toward greater success and student empowerment.

Together, let us explore the realm of student performance analysis through the lens of data analytics, as we work collaboratively to unlock the full potential of each learner and pave the way for a brighter educational future.

Prerequisites

Before using the Student Performance Analysis web application, ensure that you have the following prerequisites installed:
- ➢ Anaconda navigator (contains Jupyter notebook, Spyder etc.)
- ➢ Jupyter notebook
- ➢ Python Libraries which should contain pandas, numpy, seaborn.

Data Collection and Preparation

Data Sources

The foundation of any data analysis is the availability of accurate and relevant data. For this analysis, we collected data from diverse sources within the educational ecosystem, including student records, assessment scores, and background information. The dataset comprises the following columns:

Student ID: A unique identifier assigned to each student within the educational system.

Gender: The gender identity of the student (e.g., male, female).

Race/Ethnicity: Categorization of the student's racial or ethnic background.

Parental Level of Education: Information about the highest level of education attained by the student's parents or guardians.

Lunch: Indicates whether the student receives a standard lunch, or if it is free/reduced.

Test Preparation Course: Describes whether the student completed a test preparation course, with options "none" or "completed".

Math Score: The score attained by the student in the math assessment.

Reading Score: The score achieved by the student in the reading assessment.

Writing Score: The score received by the student in the writing assessment.

Data Collection Process

The data collection process involved extracting information from various sources within the educational institution. Student records, including demographic details and assessment scores, were compiled into a structured dataset. This dataset formed the foundation for subsequent analysis.

Data Cleaning and Preprocessing*

Prior to analysis, the raw data underwent thorough cleaning and preprocessing to ensure its accuracy and consistency. The following steps were taken:

Handling Missing Values: Any missing data points in columns were carefully addressed. Techniques such as imputation or removal were applied based on the nature of the data and the extent of missing values.

Dealing with Outliers: Outliers in assessment scores were identified and addressed through a data-driven approach, preserving the integrity of the dataset.

Encoding Categorical Variables: Categorical variables like "Gender," "Race/Ethnicity," "Parental Level of Education," "Lunch," and "Test Preparation Course" were encoded into numerical representations for compatibility with various analysis techniques.

Data Transformation: The "Lunch" column was transformed to reflect the socioeconomic status of students through categories like "Standard" (affording standard lunch) and "Free/Reduced" (receiving free or reduced-price lunch).

Normalization: Assessment scores in the "Math," "Reading," and "Writing" columns were normalized to a common scale to facilitate meaningful comparisons.

By meticulously cleaning and preprocessing the data, we ensured that the subsequent analysis would yield accurate and reliable insights into student performance.

In the next section, we will delve into Exploratory Data Analysis (EDA) to gain initial insights into the characteristics and distribution of the prepared dataset.

Exploratory Data Analysis (EDA)

Summary Statistics

To establish a foundational understanding of the dataset, we conducted exploratory data analysis (EDA) to derive insights and identify initial patterns. The analysis began with calculating summary statistics for relevant numerical columns, including "Math Score," "Reading Score," and "Writing Score." These statistics provide a snapshot of the central tendency and dispersion of student performance in each subject:

Mean: The average score achieved in each subject.
Median: The middle value in the ordered set of scores.
Standard Deviation: A measure of the spread of scores around the mean.
Minimum and Maximum: The range within which scores vary.

Distributions and Visualizations*

Visualizations were employed to illustrate the distributions of scores across subjects and to identify potential trends. Histograms were utilized to showcase the frequency distribution of scores, allowing us to observe patterns of performance. Additionally, box plots were employed to highlight the distribution's central tendency, spread, and potential outliers for each subject.

Correlations and Relationships

To explore potential relationships between variables, correlation analysis was performed to quantify the strength and direction of associations. Correlation coefficients were calculated to measure the linear relationship between pairs of numerical variables. This helped identify whether certain factors, such as completing a test preparation course or parental education levels, might influence student performance.

Group Comparisons

EDA also encompassed group comparisons based on categorical variables. Bar charts and pie charts were employed to visualize differences in performance across categories such as "Gender," "Race/Ethnicity," and "Parental Level of Education." These visualizations provided initial insights into potential disparities in performance among different student groups.

Insights from EDA

Through exploratory data analysis, several initial insights were uncovered. These include:

Distribution of scores across subjects: Preliminary analysis revealed potential variations in performance among subjects.

Correlations between variables: Initial correlation analysis suggested relationships between completing a test preparation course and higher scores.

Group disparities: EDA indicated potential differences in performance based on gender, race/ethnicity, and parental education levels.

These insights lay the foundation for deeper analysis and provide directions for investigating patterns and trends that can contribute to our understanding of student performance dynamics.

In the subsequent sections, we will delve into the metrics and key performance indicators chosen for this analysis and the specific data analysis techniques applied to uncover meaningful patterns.
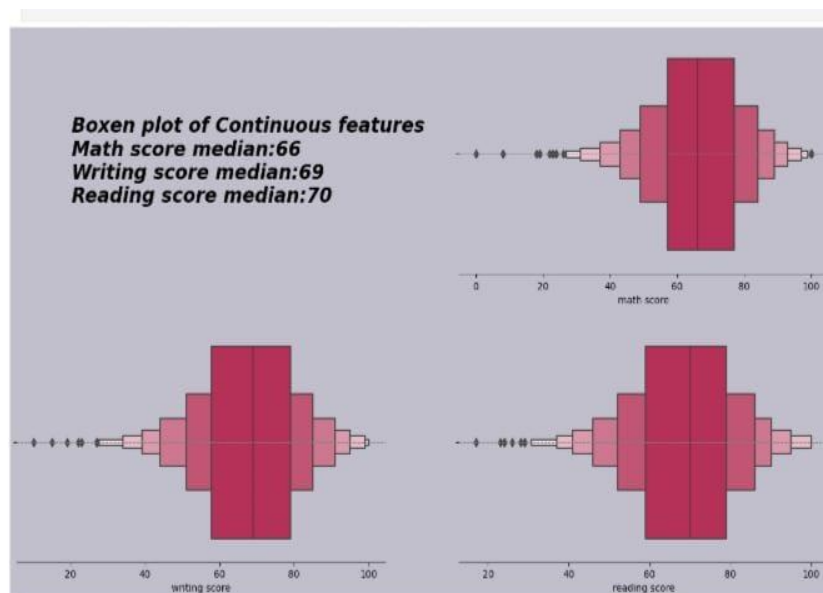
Countplot of every dataset
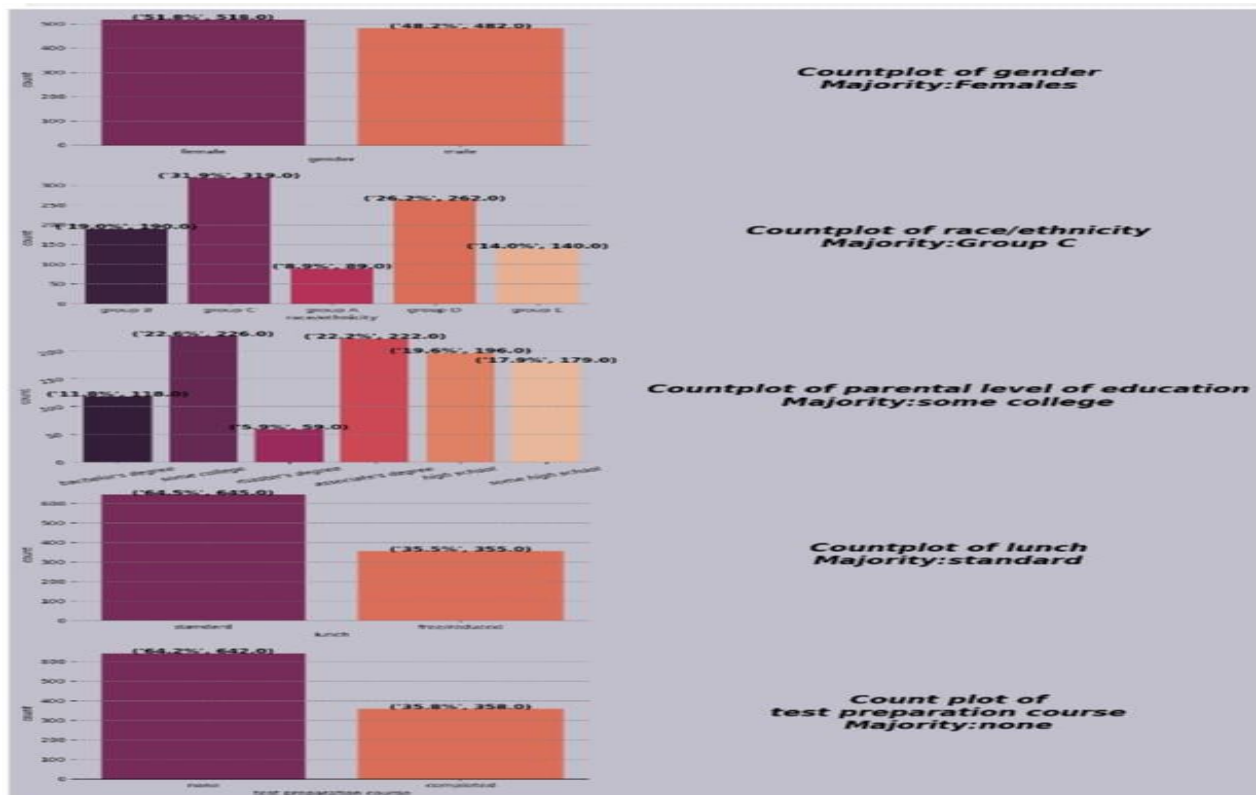
Data Analysis Techniques

Median of Each Score using Boxen Plot

Boxen plots were employed to visualize the distribution of median scores in each subject. These plots highlighted the spread and skewness of score distributions, offering insights into score disparities.

Writing Scores of Every Dataset

Visualizing writing scores across datasets illuminated trends and variations, aiding in identifying any performance patterns.

Countplots by Using Test Preparation Courses

Count plots showcased the influence of test preparation courses on performance, illustrating the impact of preparedness on assessment outcomes.

Countplots of Lunch Datasets

Count plots elucidated performance differences among students based on lunch categories, indicating potential socioeconomic impacts on academic achievement.
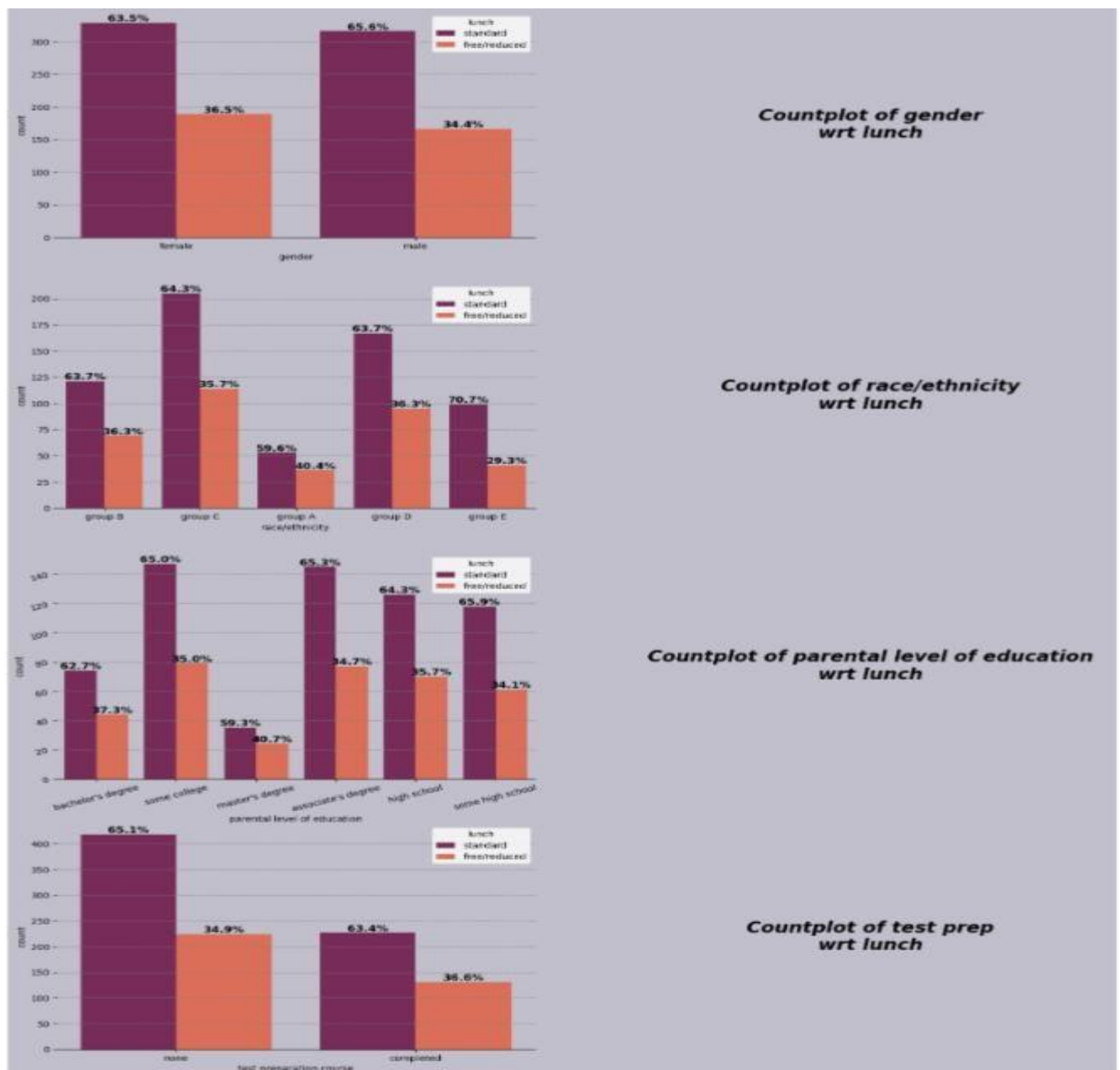
Overall Scores of Students

Visualizing overall scores provided a holistic view of student performance, allowing us to observe patterns across all assessments.

	Student Id	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	overall_score
0	101	female	group B	bachelor's degree	standard	none	72	72	74	218
1	102	female	group C	some college	standard	completed	69	90	88	247
2	103	female	group B	master's degree	standard	none	90	95	93	278
3	104	male	group A	associate's degree	free/reduced	none	47	57	44	148
4	105	male	group C	some college	standard	none	76	78	75	229

## Boxplot of Overall Scores by Using Lunch

Box plots dissected overall scores based on lunch categories, helping identify any performance variations between students with different lunch statuses.

Boxplot of Overall Scores by Using Gender

Box plots revealed overall score differences between genders, providing insights into gender-based performance dynamics.

Boxplot of Overall Scores by Using Test Preparation Course

Box plots showcased overall scores based on test preparation course completion, uncovering potential advantages for students who completed the course.
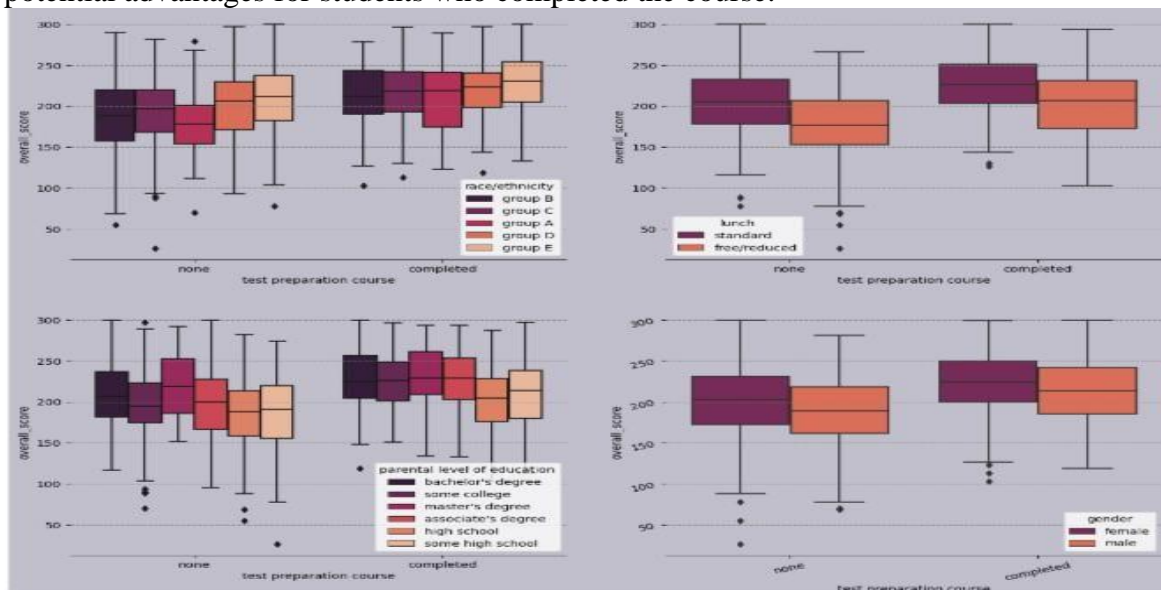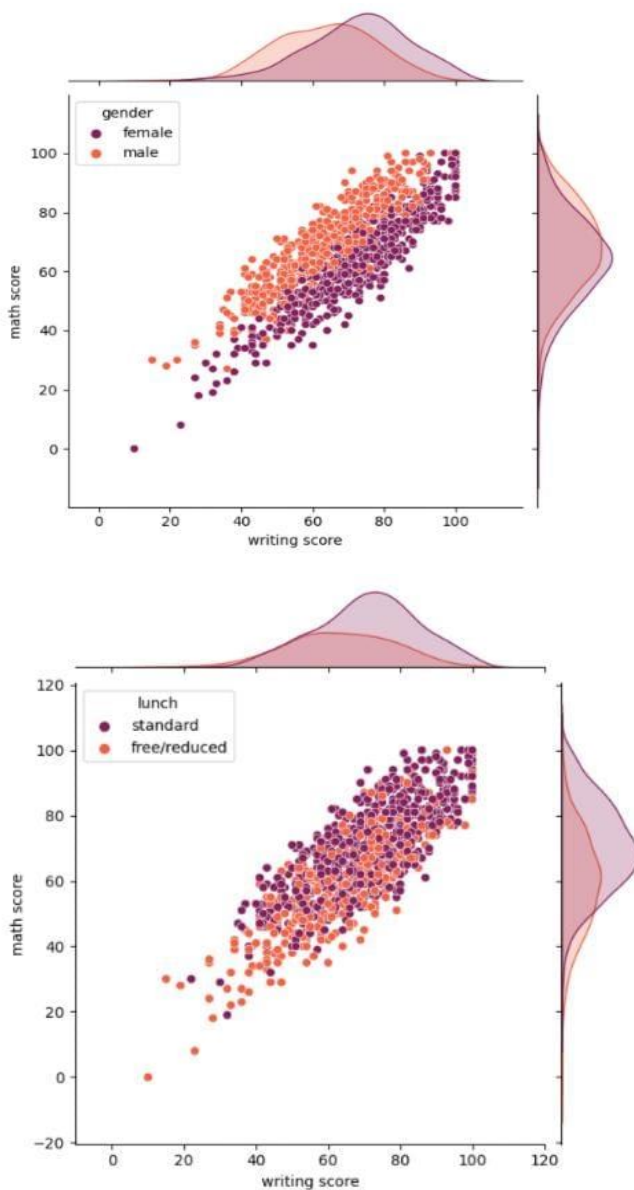
Math Score vs Writing Score by Using Lunch

The math vs writing score visualization by lunch category indicated potential lunch-related patterns in math and writing performance.

Math Score vs Reading Score by Using Gender

Visualizing math vs reading scores by gender provided insights into gender-specific performance dynamics in these subjects.

Math Score vs Reading Score by Using Lunch

Examining math vs reading scores by lunch category showcased potential performance variations influenced by lunch status.

Writing Score vs Reading Score by Using Gender

Visualizing writing vs reading scores by gender shed light on potential gender-specific performance relationships between writing and reading.

Writing Score vs Reading Score by Using Lunch

The writing vs reading score visualization by lunch category indicated potential lunch-related patterns in writing and reading performance.
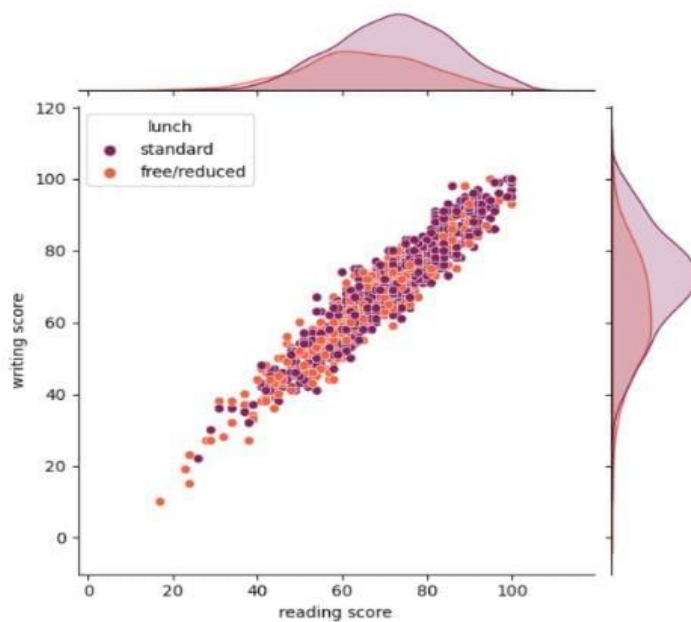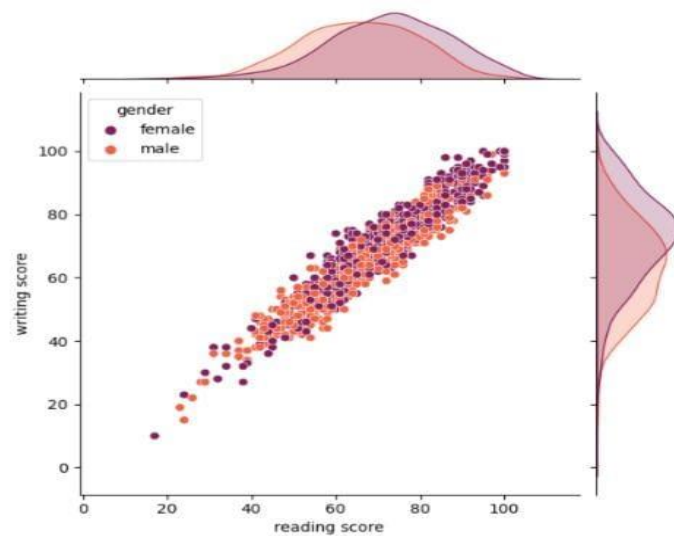
Conclusion

Incorporating data analytics into student performance analysis enriches educational insights. The analysis revealed potential influences on student performance, from socioeconomic factors to gender dynamics. These findings inform targeted strategies to enhance educational outcomes. By leveraging data-driven insights, educators and stakeholders can tailor interventions to empower students and nurture academic success.

[36]:

	Student Id	gender	lunch	test preparation course	race/ethnicity_group A	race/ethnicity_group B	race/ethnicity_group C	race/ethnicity_group D	rac
0	101	0	1	0	0	1	0	0	
1	102	0	1	1	0	0	1	0	
2	103	0	1	0	0	1	0	0	
3	104	1	0	0	1	0	0	0	
4	105	1	1	0	0	0	1	0	

ethnicity_group E	parental level of education_associate's degree	parental level of education_bachelor's degree	parental level of education_high school	parental level of education_master's degree	parental level of education_some college	parental level of education_some high school
0	0	1	0	0	0	0
0	0	0	0	0	1	0
0	0	0	0	1	0	0
0	1	0	0	0	0	0
0	0	0	0	0	1	0

Data Visualization

Visualizations play a pivotal role in conveying complex data patterns and insights effectively. They provide a lens through which data becomes accessible and actionable.