Student Performance — Exploratory Data Analysis

Presented by Riya Manoj

Introduction

What is EDA?

- Exploratory Data Analysis (EDA) is the process of analyzing datasets to summarize their main characteristics.
- It uses statistical techniques and visualization to uncover patterns ,detect anomalies and test assumptions.

Purpose of EDA

- To understand the structure and quality of data.
- To identify relationship between variables.
- To generate insights that guide further analysis or modeling.

Problem Statement

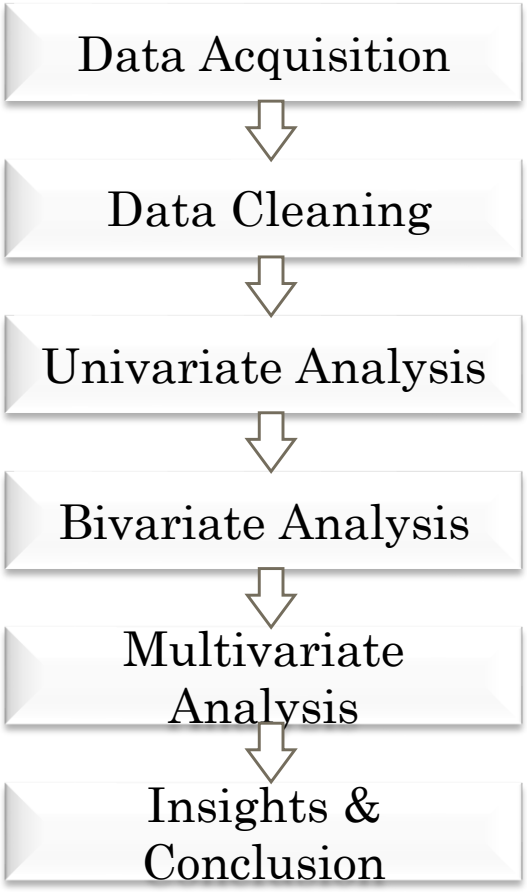
- To study student performance in subjects like Math, Reading, and Writing.
- To analyze how demographic and social factors (gender, parental education, lunch type, test preparation) influence scores.
- To provide meaningful insights that can support future predictive modeling.

About the Dataset

- - ~1000 rows, 8 columns
- - Numeric: math, reading, writing scores
- - Categorical: gender, parental education, lunch, test preparation course, race

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
female	group B	bachelor's degree	standard	none	72	72	74
female	group C	some college	standard	completed	69	90	88
female	group B	master's degree	standard	none	90	95	93
male	group A	associate's degree	free/reduced	none	47	57	44
male	group C	some college	standard	none	76	78	75

Workflow


Data Acquisition
Data Cleaning
Univariate Analysis
Bivariate Analysis
Multivariate Analysis
Insights & Conclusion

Data Cleaning

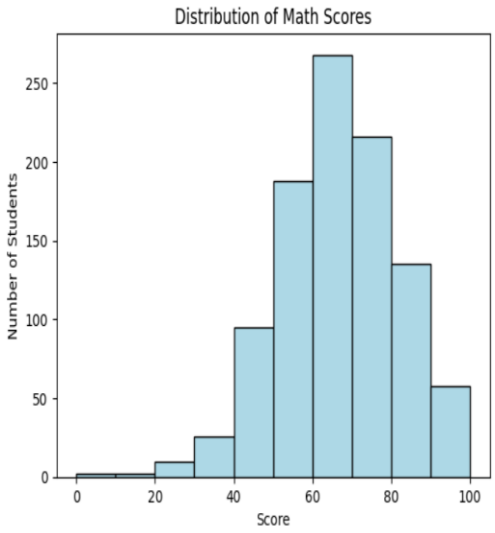
- - Verified dataset structure using info()
- - Checked for missing values none found
- - Verified data consistency (no duplicates)
- - No major cleaning required

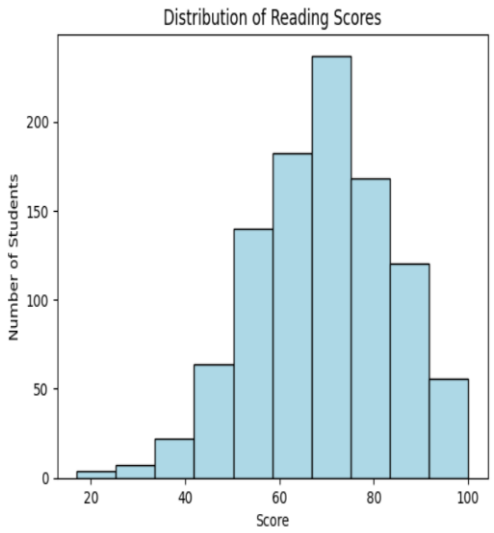
```
df.isnull().sum()
```

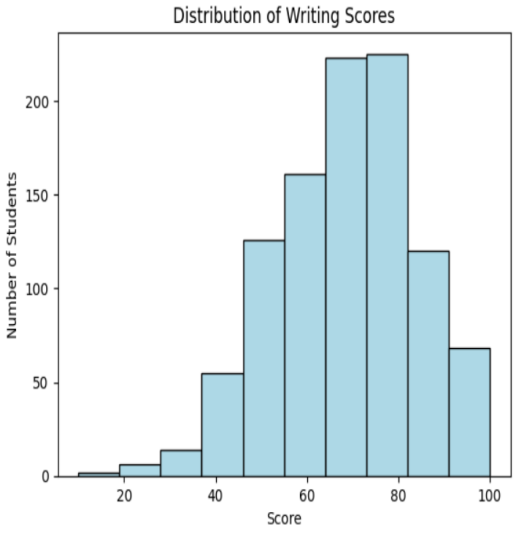
	0
gender	0
race/ethnicity	0
parental level of education	0
lunch	0
test preparation course	0
math score	0
reading score	0
writing score	0

Distribution of Scores

- - Most scores between 60–80
- - Few extreme outliers
- - Balanced distributions

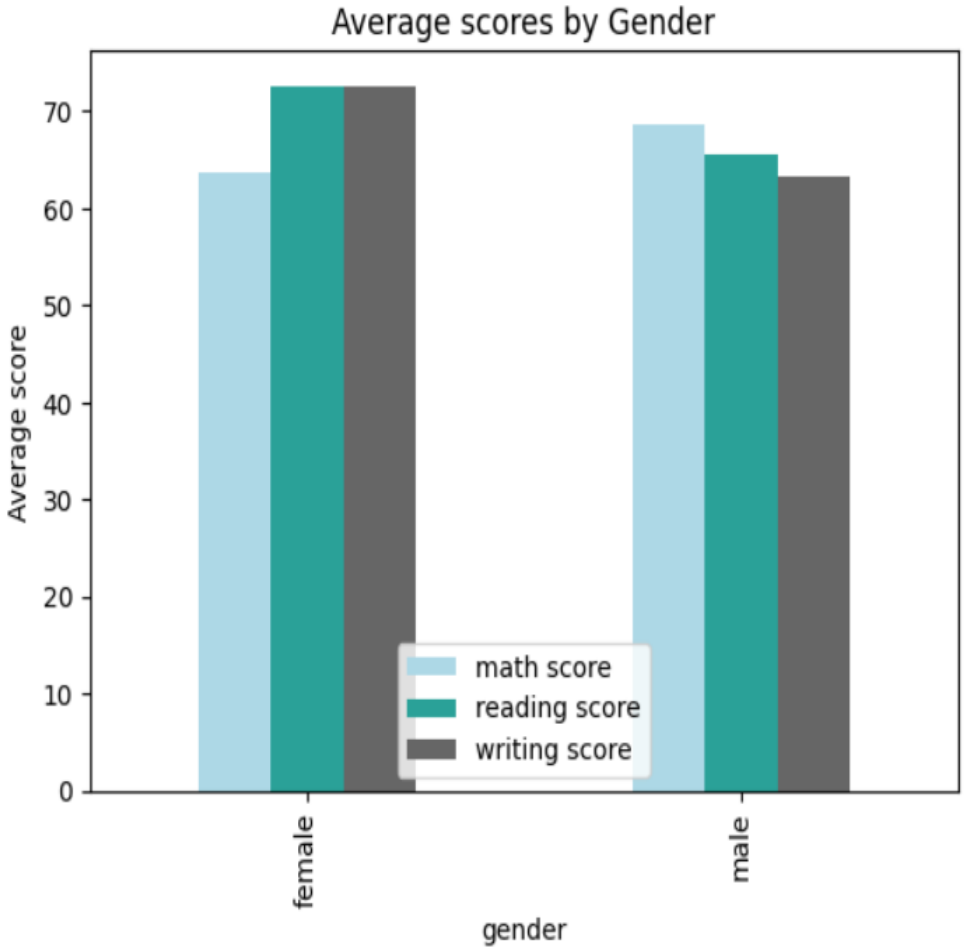






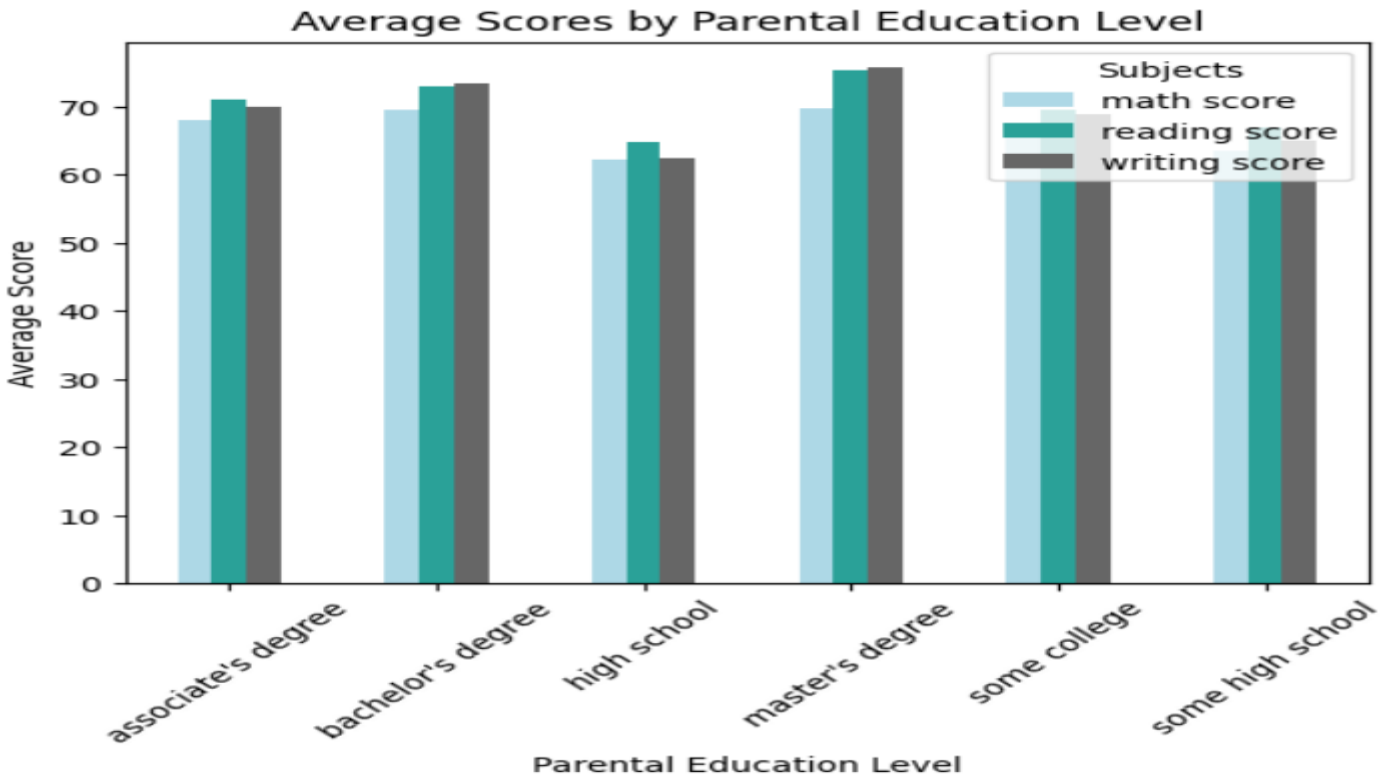
Gender and Performance

- - Females: better in reading & writing
- - Males: slightly better in math
- - Differences small



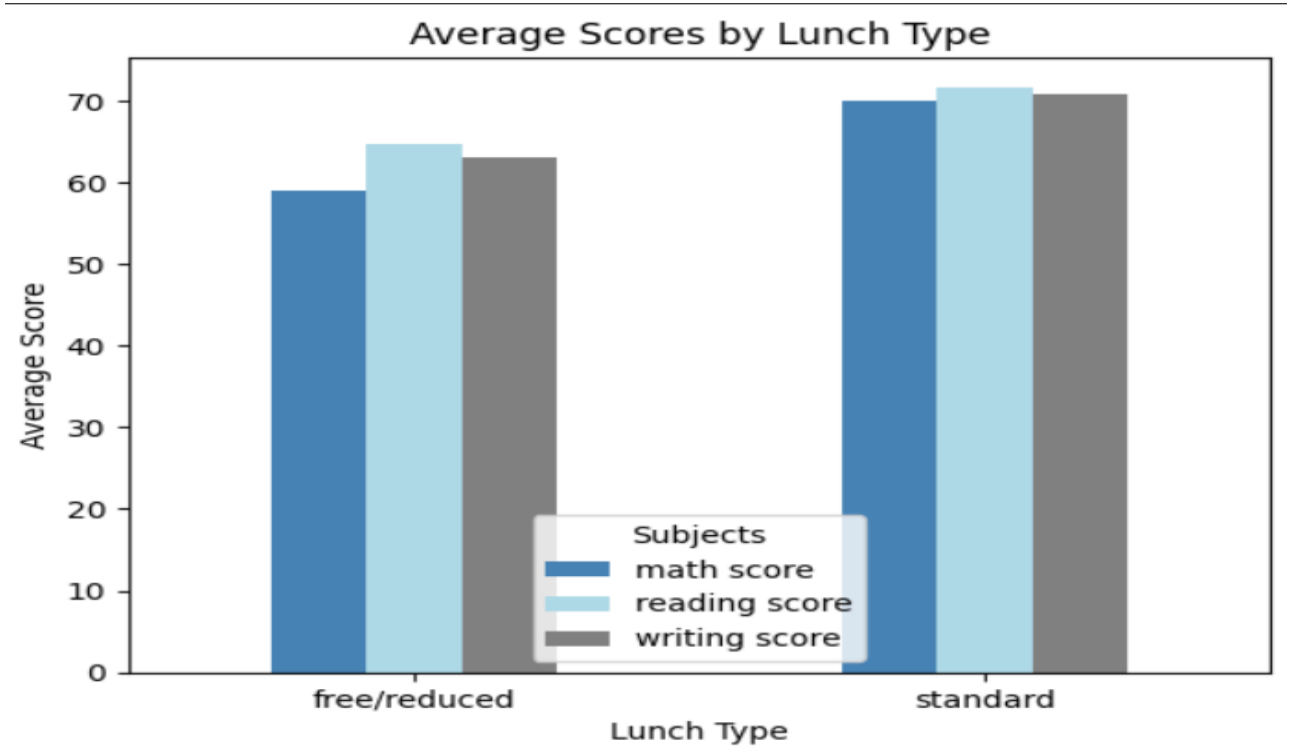
Parental Education Level

- - Higher education level → higher scores
- - Clear positive correlation



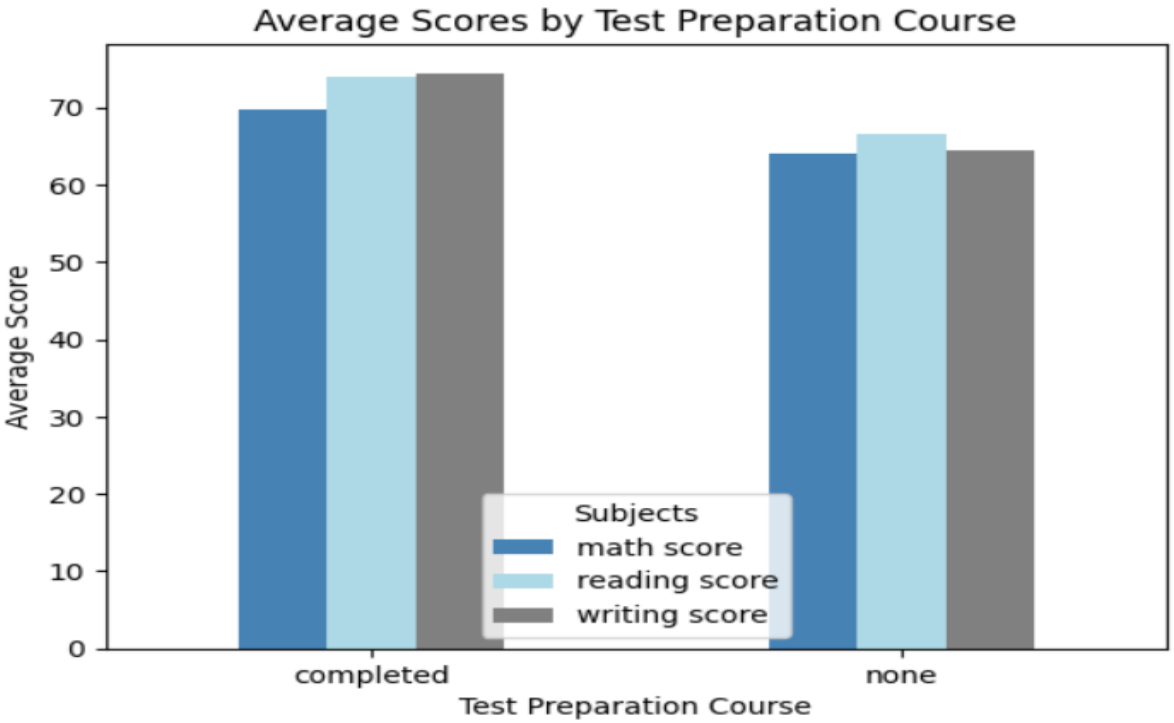
Lunch Type

- - Standard lunch → higher performance
- - Free/reduced lunch → relatively lower scores



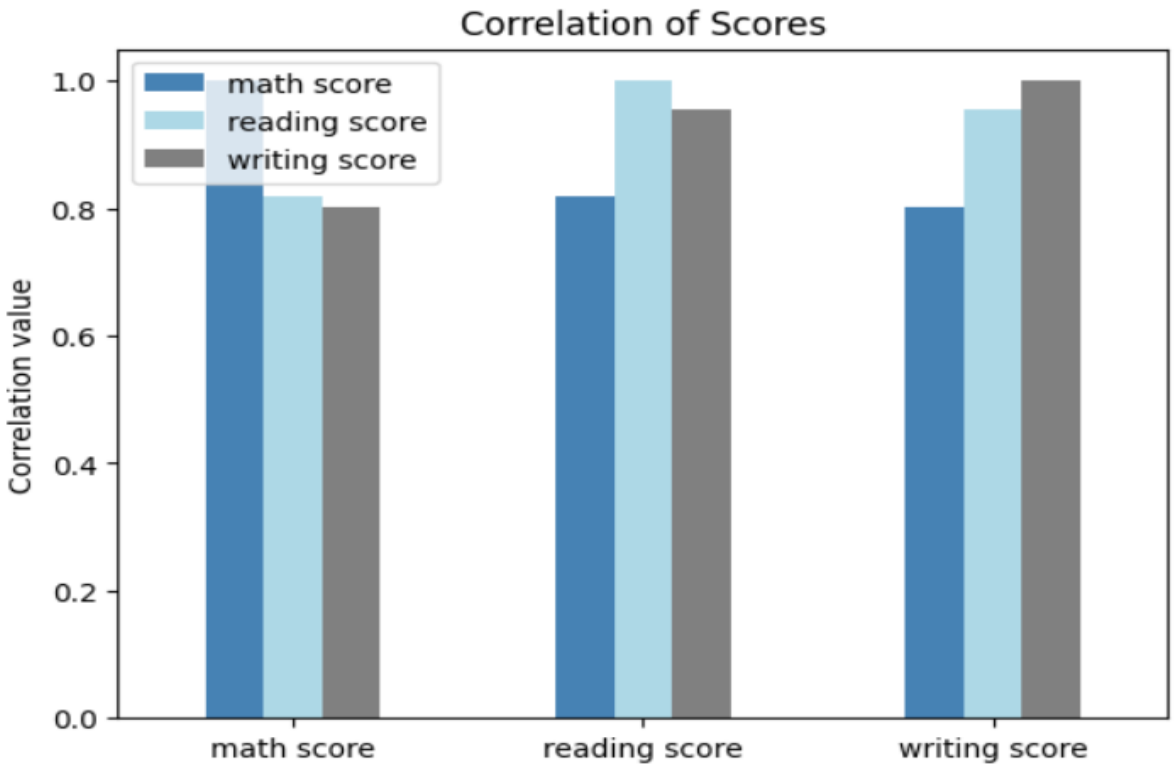
Test Preparation Course

- - Completed test prep → higher scores
- - Preparation has clear positive effect



Correlation Between Subjects

- - Math, Reading, Writing strongly correlated
- - Doing well in one subject → likely to do well in others



Key Insights

- **Test preparation course** ->Students who completed it showed consistently higher scores
- **Parental education-**>Higher parental education levels correlate with better student performance across all subjects.
- **Lunch type-**>Students with a standard lunch performed better than those with free/reduced lunch, suggesting resource or nutrition impact.
- **Gender trends-**>Females outperformed in reading & writing, while males had a slight advantage in math.
- **Score relationships-**>Strong positive correlation among math, reading, and writing scores.

Tools Used

- Python Libraries - Pandas, NumPy, Matplotlib
- Platform – Google Colab
- Dataset source -Kaggle

Conclusion

- EDA provided a clear understanding of the dataset through cleaning, exploration, and visualization
- Analysis showed that gender, parental education, lunch type, and test preparation significantly affect student performance
- Strong correlations were observed among math, reading and writing scores
- EDA proved essential for extracting insights and preparing the data for further modeling

Thank You