Student Performance

Exploratory Data Analysis

Bachir Zekhnine

Souici Mouhamed

Project Overview

Objective

Analyze student performance to understand factors influencing academic success.

Key Questions

- Scores vary by gender?
- Socioeconomic impact?
- Test preparation effect?

Methodology

- Data cleaning
- Univariate analysis
- Bivariate relationships
- Clustering analysis

The Dataset

Student Performance Dataset

Academic records with demographic and socioeconomic factors

Dimensions

Metric	Value
Records	**1000**
Features	**8**
Missing	**0**

Features

Categorical: Gender, Race, Parent Edu, Lunch, Test Prep

Numerical: Math, Reading, Writing

Data Cleaning Process

Steps

1. Missing values → **None**
2. Duplicates → **Removed**
3. Create total_score
4. Create average_score

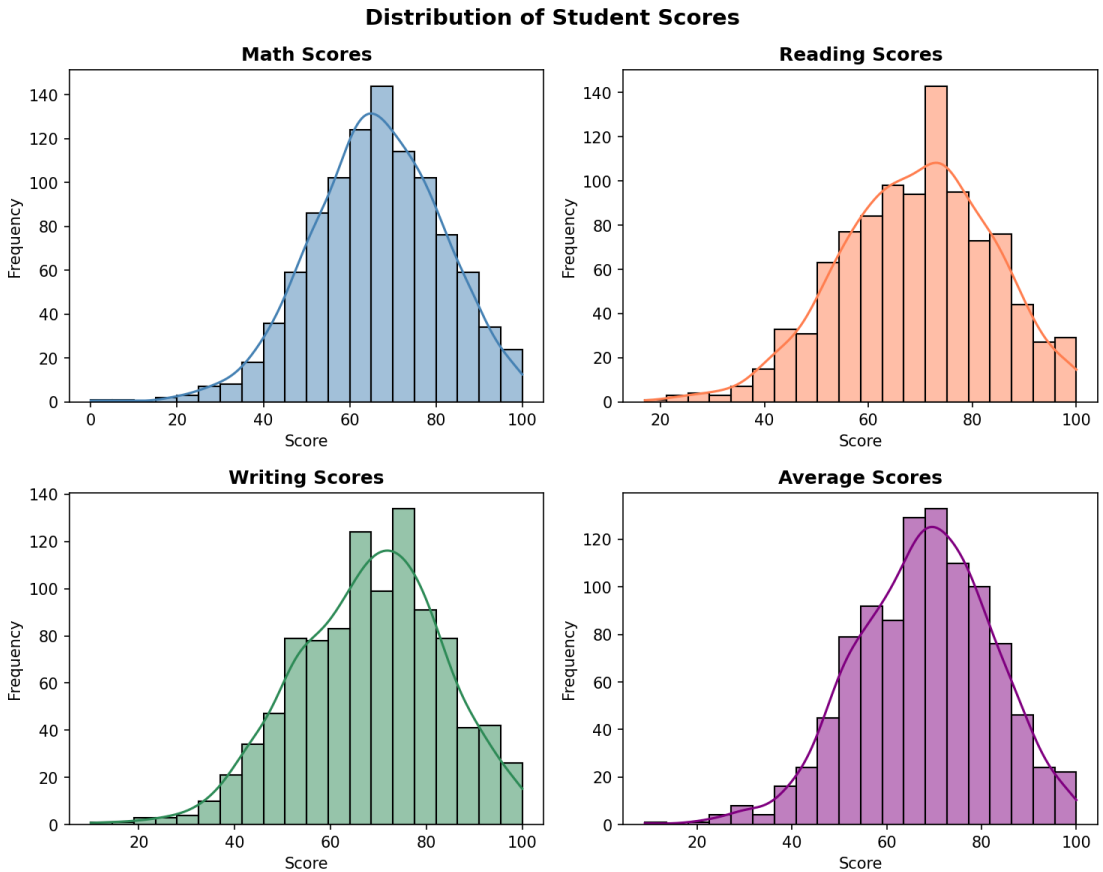
Results

- Clean dataset ready
- New features created
- 1000 records retained

Feature Engineering

$$\text{total\_score} = \text{math} + \text{reading} + \text{writing}$$
$$\text{average\_score} = \text{total\_score} / 3$$

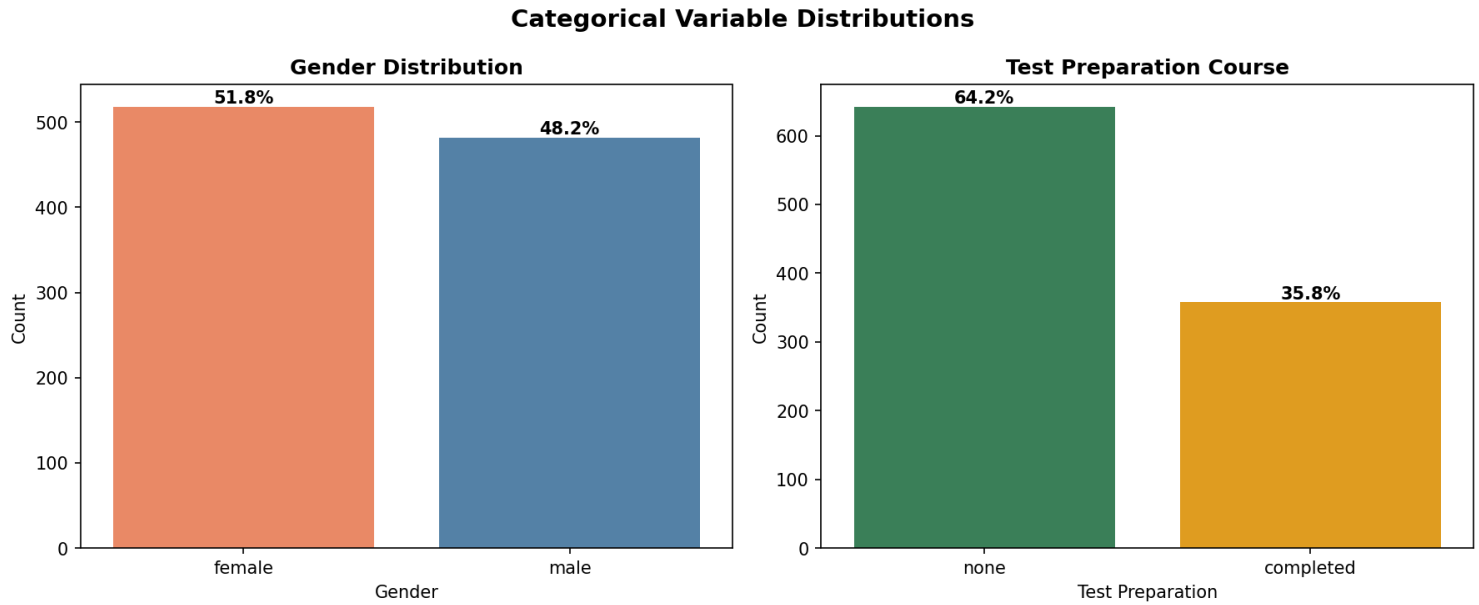
Univariate Analysis: Score Distributions



Insights

Math: 66 **Reading:** 69 **Writing:** 68 | Better in reading/writing; slightly left-skewed

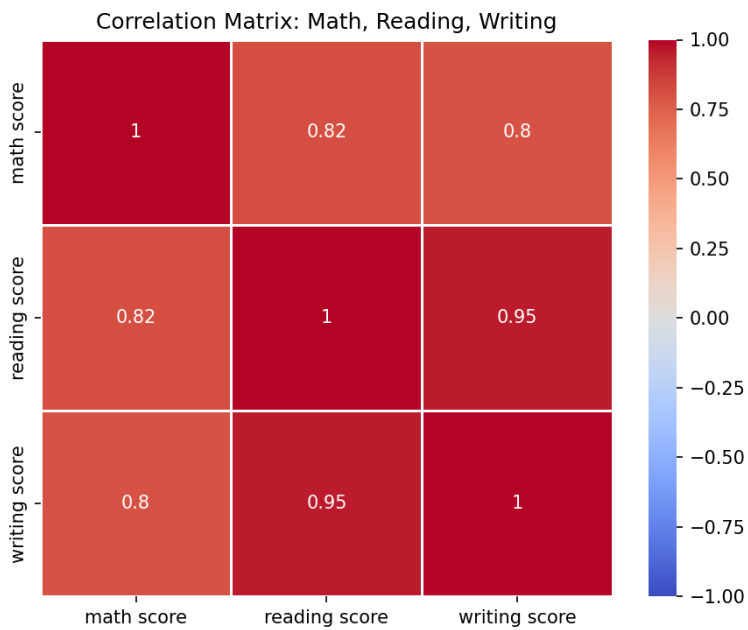
Univariate Analysis: Categorical Variables



Distribution

Gender: Female 51.8% | Male 48.2% **Test Prep:** Completed 35.8% | Not completed 64.2%

Bivariate Analysis: Correlations



Key Findings

- **Strong** (0.95): Read-Write
- **Moderate** (0.82): Math-Read
- **Moderate** (0.80): Math-Write

Interpretation

Verbal skills are highly linked. Math is more independent.

Key Insight: Test Preparation Impact

Major Finding

Test prep completion improves scores across ALL subjects

Completed

- Only **35.8%** completed
- Higher Math scores
- Higher Reading scores
- Higher Writing scores

No Course

- **64.2%** did not complete
- Lower performance
- Missed opportunity
- Need expanded access

Implication

Structured support provides **tangible benefits**. Expand access to prep programs.

Key Insight: Parental Education Correlation

🏛 Finding

Higher parental education = Better student performance

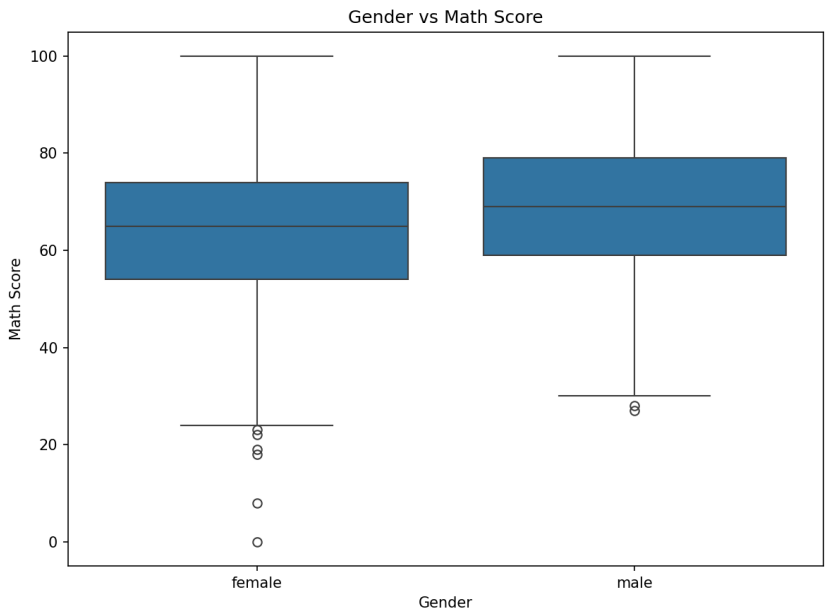
Performance Ranking

Rank	Parental Education	Performance
1	Master's degree	**Highest**
2	Bachelor's degree	High
3	Associate's degree	Above Average
4	Some college	Average
5	High school	Below Average
6	Some high school	**Lowest**

Key Trend 1: Gender Performance Gap

Finding

Gender influences subject-specific performance



Analysis

Females:

- Reading: +7.1 pts
- Writing: +9.2 pts
- Math: -5.2 pts

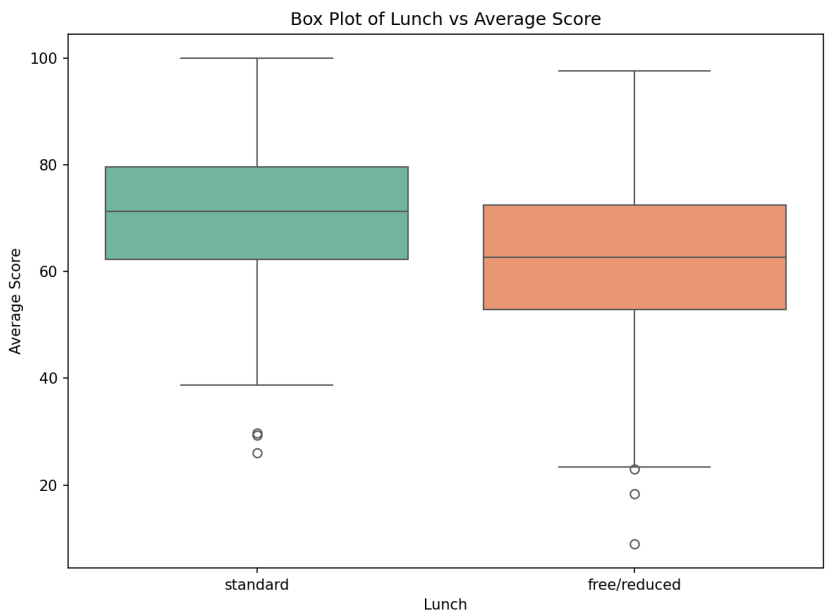
Males:

- Math: +5.2 pts
- Lower verbal scores

Key Trend 2: Socioeconomic Impact

Finding

Lunch type (SES proxy) predicts performance



Gap Analysis

Standard vs Free/Reduced:

- Average: **+11.1 pts**
- Math: +12.7 pts
- Reading: +11.5 pts
- Writing: +11.8 pts

Key Insight: Best Performing Subgroup

Top Finding

Females with Standard Lunch = Highest Performers

Cross-Group Analysis

By Gender & Lunch:

- **Female + Standard:** Highest
- Male + Standard: High
- Female + Free/Reduced: Moderate
- Male + Free/Reduced: Lowest

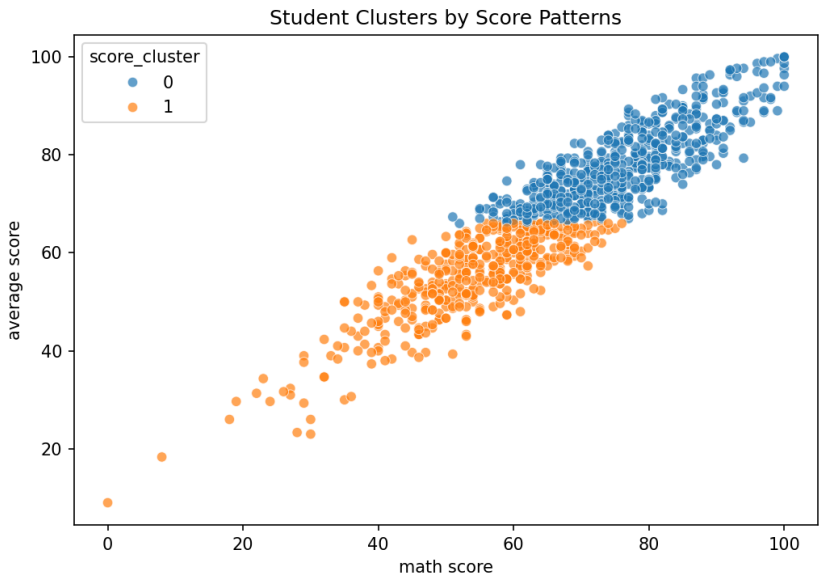
Why This Matters

- **Compound effect** of gender + SES
- Females excel in verbal skills
- Standard lunch = more resources
- Combined advantage = top scores

Implication

Male students with free/reduced lunch need the most support and intervention.

Multivariate Analysis: Student Clusters



K-Means Results

3 clusters identified:

1. High achievers (15%)
2. Average performers (70%)
3. At-risk students (15%)

Insight

Clear separation enables targeted interventions.

Cluster Analysis: Actionable Insights

Segmentation Strategy

Each cluster requires different educational approaches

High

Profile:

- Above-average
- Consistent

Action:

- Advanced programs
- Gifted courses

Avg

Profile:

- Near mean
- Stable

Action:

- Maintain support
- Monitor

Risk

Profile:

- Below-average
- Needs help

Action:

- Intervention
- Tutoring

Key Insight: Strongest Predictors of Success

Factors Ranked

What matters most for student performance?

Impact Ranking

1. **Socio-economic status** (lunch) — *Strongest*
2. **Test preparation** — *Strong positive effect*
3. **Parental education** — *Clear correlation*
4. **Race/ethnicity** — *Moderate variation*
5. **Gender** — *Modest, subject-dependent*

Critical Insight

Economic factors outweigh all others — addressing socio-economic barriers is the **top priority**.

Key Takeaways

Summary

What We Learned from This Analysis

- **Socio-economic status** is the strongest predictor
- **Test preparation** improves all scores
- **Reading & Writing** are linked (r=0.95)
- **Parental education** correlates with performance
- **Gender differences** are modest
- **Clustering** enables interventions

Bottom Line

"Address socio-economic barriers and expand test prep access for maximum impact."

Thank You!

Questions & Discussion

Contact

mouhamed.souici@univ-bba.dz

bachir.zekhnine@univ-bba.dz

github.com/kkbbmrl/EDA