

# Student Performance Analysis

## 1. Project Overview

The purpose of this project is to analyze student academic performance across three major subjects — Math, Reading, and Writing and identify key factors influencing overall achievement. Using Python-based EDA, SQL analytics, and a multi-page Power BI dashboard, this project provides data-driven insights on demographic characteristics, behavioral aspects, and subject-level weaknesses.

The outcomes of this analysis support educators and institutions in understanding performance patterns, improving learning strategies, and enhancing student success rates.

## 2. Dataset Description

The dataset consists of 1000 student records with the following features:

### **Demographic Attributes:**

- Gender
- Race/Ethnicity
- Parental Level of Education
- Lunch Type (Standard/ Free/Reduced)

### **Academic Attributes:**

- Math score
- Reading score
- Writing score
- Total score
- Average score
- Performance band (High/Low/Medium)
- Grade band (A/B/C/D/F)
- Pass indicator (Math, Reading, Writing, Overall)

**Behavioral Attribute:**

- Test preparation course (Completed/None)

**Additional Fields:**

- Total score
- Average score
- Pass/Fail Flags
- Grade Band
- Performance Band

### 3. Tools & Technology Used

- **Python**
  - Pandas, Numpy – Data Cleaning
  - Seaborn, Matplotlib – EDA Visualization
- **SQL**
  - Basic, intermediate, and analytical queries
  - Ranking, window functions, segmentation
  - Pass rate calculations
  - Demographic comparisons
- **Power BI**
  - 3-Page interactive dashboard
  - Custom DAX measures
  - KPI Cards, donut Charts, scatter plots, bar charts, histogram, slicers.

### 4. Methodology

#### 4.1 Data Cleaning (Python)

- Removed inconsistencies
- Verified data types
- Created derived features:

- total\_score, avg\_score
- pass\_math, pass\_reading, pass\_reading, passed\_all
- performance\_band
- grade\_band

## 4.2 EDA (Python)

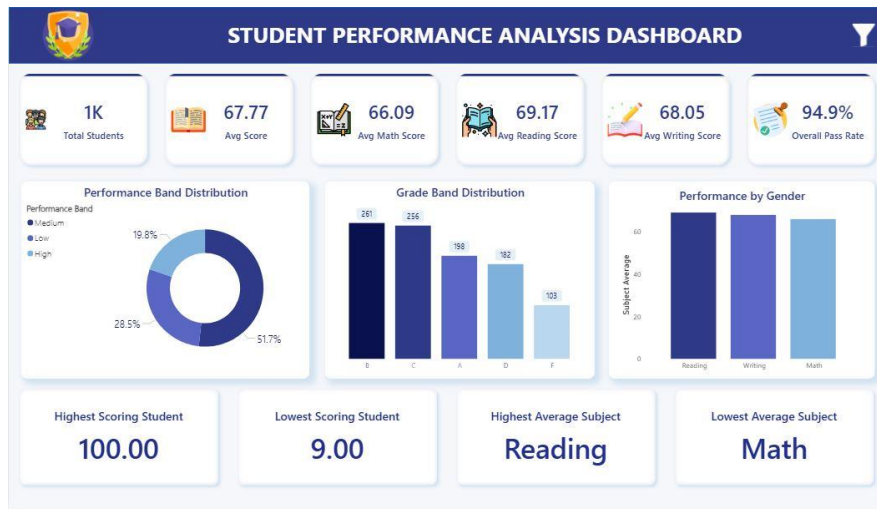
- Performed:
  - Subject core comparison
  - Correlation heatmaps
  - Pass-rate analysis
  - Performance segmentation based on avg\_score
- Key charts included:
  - Score distributions
  - Boxplots
  - Scatter plots
  - Heatmaps
  - Category wise averages

## 4.3 SQL Analysis

- Used SQL to answer:
  - Which subject students struggle to most
  - Pass rate by gender/lunch/test prep
  - High performers vs low performers
  - Demographic effects on average score
  - Ranking and window-based insights

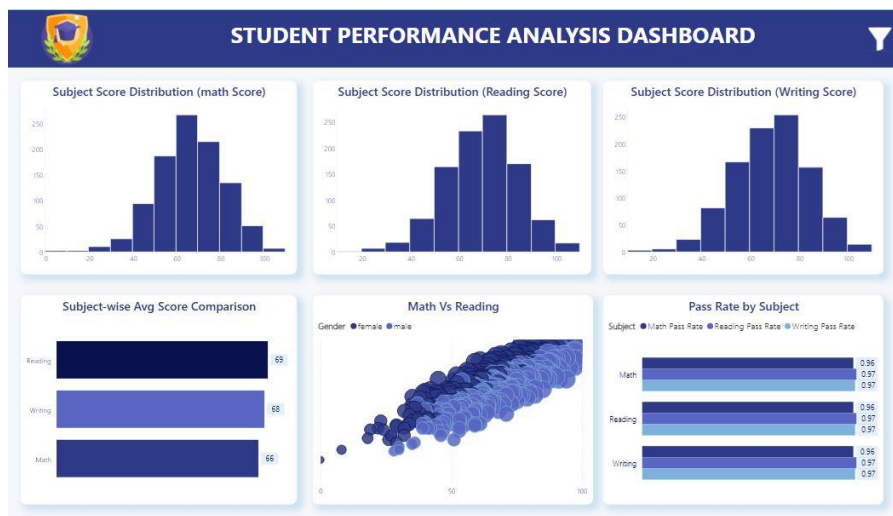
## 4.4 Power BI Dashboard Development

- Created a 3-Page interactive dashboard:
  - Page 1 – Overview Dashboard
    - KPI Cards
    - Performance band Distribution
    - Grade band analysis
    - Average score by subject
    - Cards for highest & lowest scoring subject



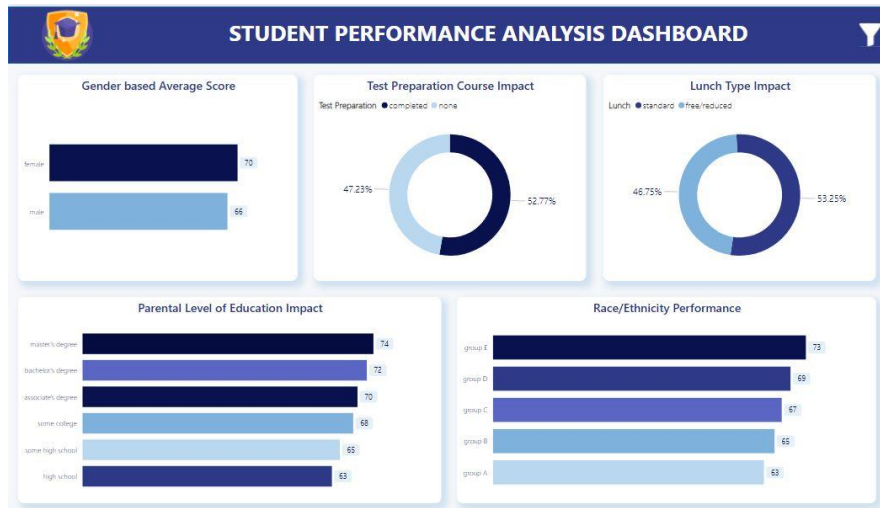
○ Page 2 – Subject-wise Analysis

- Math/Reading/Writing score distribution
- Subject comparison
- Math vs Reading scatter plot
- Subject pass rate analysis



○ Page 3 – Demographic Impact Analysis

- Gender-based scores
- Parental education influence
- Lunch Type comparison
- Test preparation effect
- Race/Ethnicity performance



## 5. Dashboard Summary & Key Insights

### 5.1 Overall Performance

- Average student score is moderate, with clear split between high and low performers.
- Reading and Writing score show a strong positive correlation, while Math varies more widely.

### 5.2 Subject-Level Insights

- Reading and Writing have the highest averages.
- Math is the most challenging subject, with lower pass rates.
- Students strong in Reading are likely also strong in Writing.

### 5.3 Demographic Insights

- Standard lunch students outperform free/reduced lunch students by 10+ points.
- Students whose parents hold Bachelor's or Master's degree show significantly higher average scores.
- Gender difference:
  - Females outperform males in Reading & Writing.
  - Math scores are relatively balanced.

### 5.4 Test Preparation Impact

- Students who completed a test preparation course score 12-15 points higher on average.
- They also show far higher pass rates across all subjects.

### 5.5 Performance Segmentation

- Higher performers (avg > 85) from a small cluster.
- Most students fall in the Medium performance band.
- Low performers show major gaps, especially in Math.

## 6. Recommendations

Based on analysis:

1. Strengthen Math Support Programs
  - Provide additional tutoring
  - Early intervention for low performers
  - Encourage participation in prep courses
2. Promote Test Preparation Courses
  - Demonstrated to significantly improve performance
  - Make prep courses mandatory for struggling students
3. Support Students from Low Socioeconomic Backgrounds
  - “Free/reduced lunch” students score lower
  - Introduce mentorship, academic support, and learning resources
4. Engage Parents
  - Higher parental education correlates with student success
  - Conduct workshops to involve low-education households
5. Focus on Reading & Writing to Improve Overall Performance
  - High correlation means improvements here can lift total performance

## 7. Conclusion

This project successfully demonstrates how data analytics can be applied to education by combining Python EDA, SQL analytics, and Power BI dashboards to uncover hidden patterns in student performance.

The insights obtained can help educators, policymakers, and institutions design more effective interventions, improve academic outcomes, and provide targeted support where students need it most.