

**BEYOND THE GRADES: AN EXCEL-DRIVEN STORY OF HARVARD  
UNIVERSITY STUDENTS' PERFORMANCE  
BY  
OYIBO OGHENENYERHOVWO DORCAS**

## **INTRODUCTION**

Harvard University is known for excellence—but behind every grade lies a story rarely told. In this analysis, I imagined walking through the quiet hallways of Harvard's academic records office, opening a spreadsheet that held the lives, efforts, and struggles of 1,000 students. Each row represented someone's late-night study sessions, someone's family background, someone's challenges, and someone's triumphs. To make the analysis relatable, the dataset is considered as if it were from Harvard University students.

This report tells that story: It shows what truly influences the performance of Harvard University students—not just grades, but human factors like gender, race, family background, and access to learning opportunities.

### **a. Objective**

To understand how the student's performance (test scores) is affected by the other variables (Gender, Ethnicity, Parental level of education, Lunch, Test preparation course).

### **b. Problem Being Addressed**

Student performance is multi-dimensional. Schools struggle to understand:

- Why some students consistently outperform others?
- How parental background affects performance?
- The impact of nutrition and preparation on test scores
- Whether race or gender disparities exist?

This analysis attempts to answer these questions using descriptive, comparative, and categorical analysis techniques in Excel.

### **c. Key Datasets and Methodologies**

Dataset from Kaggle with:

- 1000 rows
- 8 original columns + 2 appended columns (Total & Status)
- Excel tools: pivot tables, IF statements, averages, pivot charts and

## **STORY OF THE DATA**

### **a. Data Source**

Kaggle (Student Performance Dataset) (Adithya Shetty, 2025).

### **b. Data Collection Process**

Reflects academic records capturing:

- Gender
- Race/ethnicity
- Parental education
- Lunch type
- Test preparation
- Math, reading, writing scores

### **c. Data Structure**

Independent features:

- Gender: Student's gender (Male/Female)
- Race/Ethnicity: Student's ethnic group (A–E)
- Parental Education: Highest academic qualification of parents
- Lunch: Type of lunch (Standard or Free/Reduced)
- Test Preparation: Whether student completed test prep (Yes/No)
- Status: Pass or Fail

Dependent features:

- Math Score: Score out of 100 in Math
- Reading Score: Score out of 100 in Reading
- Writing Score: Score out of 100 in Writing
- Total Score: Sum of math, reading and writing score.

### **Analytical Framework Used**

1. Performance by Gender
2. Evaluation of Parental Education
3. Assessment by Race/Ethnicity
4. Count by Reading Score
5. Count by Math Score
6. Count by Writing Score
7. Evaluation of Lunch Type

8. Status Checkup (Pass/Fail)
9. Assessment of Test Preparation Course

#### **d. Feature Significance**

- Parental education strongly influences performance.
- Test prep provides a strong performance boost.
- Standard lunch correlates with higher scores.
- Gender performance differences exist.
- Race/Ethnicity groups show distinct academic patterns.

#### **e. Data Limitations**

- No socioeconomic income data
- No detailed student background information
- Race groups are coded (A–E)
- Only standardized test metrics included

#### **f. Suggestions for Improvement**

- Include attendance and participation data
- Track progress across multiple terms
- Add socio-economic background details
- Include class size or teacher influence metrics

### **DATA PREPROCESSING & SPLITTING**

#### **a. Data Cleaning**

- No duplicates
- No missing values
- Score ranges validated

#### **b. Data Transformation**

Two new fields created:

- Total Score
- Status (Pass/Fail)

#### **c. Data Splitting**

Analysis separated into:

- Gender

- Race/Ethnicity
- Parental education
- Lunch type
- Test preparation
- Score categories

#### d. Industry Context

At Harvard's standard, student performance insights guide:

- Recruitment
- Academic planning
- Program evaluation
- Equity development

#### e. Stakeholders

- University leadership
- Faculty
- Parents
- Students

#### f. Value to the Industry

- Improved student outcomes
- Better learning strategies
- Increased retention
- Enhanced university reputation

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	Total	Status
female	group B	bachelor's degree	standard	none	72	72	74	218	Pass
female	group C	some college	standard	completed	69	90	88	247	Pass
female	group B	master's degree	standard	none	90	95	93	278	Pass
male	group A	associate's degree	free/reduced	none	47	57	44	148	Fail
male	group C	some college	standard	none	76	78	75	229	Pass
female	group B	associate's degree	standard	none	71	83	78	232	Pass
female	group B	some college	standard	completed	88	95	92	275	Pass
male	group B	some college	free/reduced	none	40	43	39	122	Fail
male	group D	high school	free/reduced	completed	64	64	67	195	Pass
female	group B	high school	free/reduced	none	38	60	50	148	Fail
male	group C	associate's degree	standard	none	58	54	52	164	Pass
male	group D	associate's degree	standard	none	40	52	43	135	Fail
female	group B	high school	standard	none	65	81	73	219	Pass
male	group A	some college	standard	completed	78	72	70	220	Pass
female	group A	master's degree	standard	none	50	53	58	161	Pass
female	group C	some high school	standard	none	69	75	78	222	Pass
male	group C	high school	standard	none	88	89	86	263	Pass
female	group B	some high school	free/reduced	none	18	32	28	78	Fail
male	group C	master's degree	free/reduced	completed	46	42	46	134	Fail
female	group C	associate's degree	free/reduced	none	54	58	61	173	Pass
male	group D	high school	standard	none	66	69	63	198	Pass
female	group B	some college	free/reduced	completed	65	75	70	210	Pass
male	group D	some college	standard	none	44	54	53	151	Pass
female	group C	some high school	standard	none	69	73	73	215	Pass
male	group D	bachelor's degree	free/reduced	completed	74	71	80	225	Pass

*Image 1: Dataset Before Data Preprocessing*

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	Total	Status
female	group B	bachelor's degree	standard	none	72	72	74	218	Pass
female	group C	some college	standard	completed	69	90	88	247	Pass
female	group B	master's degree	standard	none	90	95	93	278	Pass
male	group A	associate's degree	free/reduced	none	47	57	44	148	Fail
male	group C	some college	standard	none	76	78	75	229	Pass
female	group B	associate's degree	standard	none	71	83	78	232	Pass
female	group B	some college	standard	completed	88	95	92	275	Pass
male	group B	some college	free/reduced	none	40	43	39	122	Fail
male	group D	high school	free/reduced	completed	64	64	67	195	Pass
female	group B	high school	free/reduced	none	38	60	50	148	Fail
male	group C	associate's degree	standard	none	58	54	52	164	Pass
male	group D	associate's degree	standard	none	40	52	43	135	Fail
female	group B	high school	standard	none	65	81	73	219	Pass
male	group A	some college	standard	completed	78	72	70	220	Pass
female	group A	master's degree	standard	none	50	53	58	161	Pass
female	group C	some high school	standard	none	69	75	78	222	Pass
male	group C	high school	standard	none	88	89	86	263	Pass
female	group B	some high school	free/reduced	none	18	32	28	78	Fail
male	group C	master's degree	free/reduced	completed	46	42	46	134	Fail
female	group C	associate's degree	free/reduced	none	54	58	61	173	Pass
male	group D	high school	standard	none	66	69	63	198	Pass
female	group B	some college	free/reduced	completed	65	75	70	210	Pass
male	group D	some college	standard	none	44	54	53	151	Pass
female	group C	some high school	standard	none	69	73	73	215	Pass
male	group D	bachelor's degree	free/reduced	completed	74	71	80	225	Pass

Image II: Dataset After Data Preprocessing

The following Pre-analysis, In-analysis and Post-analysis insights sections are displayed in the Costco wholesale report dashboard below:

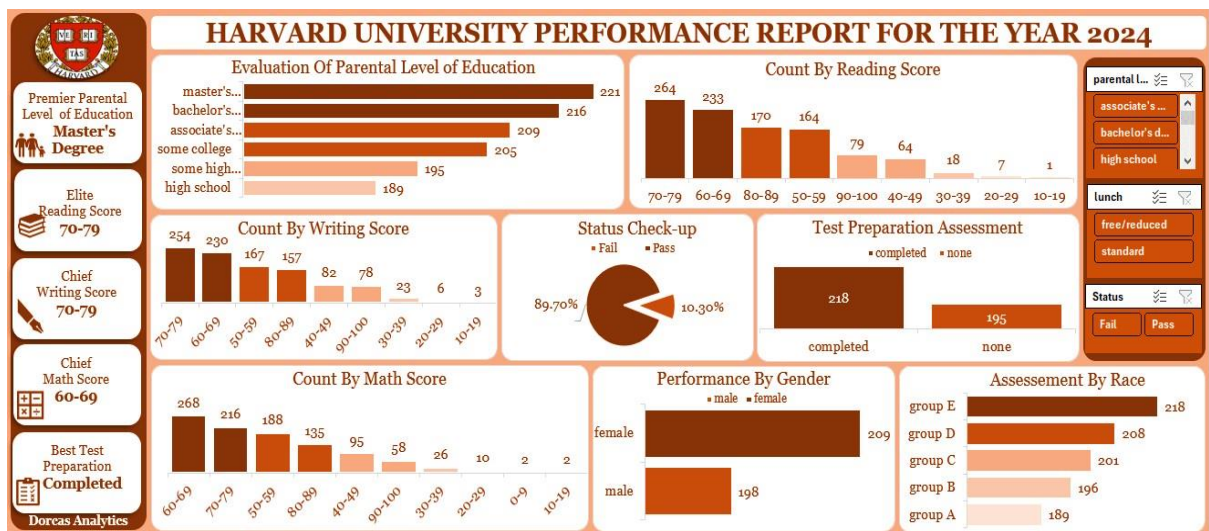


Image III: Harvard University Report Dashboard

## PRE-ANALYSIS

### a. Key Trends (Initial Observations)

- Females (209) slightly outnumber males (198) and appear to outperform them on average.
- Students with Master's-educated parents take the lead in academic performance.

- A majority (642) did not participate in the test preparation course, raising questions on accessibility.
- Race Group E consistently ranks highest, while Group A ranks lowest.
- Pass Rate is high (89.70%), with a small but important 10.30% failure rate.

#### **b. Potential Correlations**

- Higher parental education → higher scores
- Standard lunch → better performance
- Test prep → increased score levels
- Gender → reading/writing strengths
- Race groups → performance variety

#### **c. Initial Insights**

- Students in the **60–69 score range (majority)** need targeted intervention.
- High performers (70–79 range) are strong candidates for enrichment programs.
- The strong performance among females needs further investigation to understand underlying factors.

### **IN-ANALYSIS**

#### **a. Unconfirmed Insights**

- Test preparation appears to influence all score categories but needs deeper validation.
- Group E's high average suggests cultural or educational reinforcement at home.
- Students with High School-educated parents are significantly disadvantaged.
- Lunch type may be a silent predictor of academic success.

#### **b. Recommendations (During-Analysis)**

- Investigate barriers preventing 642 students from completing test preparation.
- Teachers should intensify math support, as it was the weakest subject across groups.
- Create equity programs for High-School-educated parental groups.
- Expand nutrition programs, as free/reduced lunch students consistently perform lower.

### **c. Excel Techniques Used**

- Pivot tables for categorical summarization
- IF statements for pass/fail classification
- SUM and AVERAGE functions
- Pie, Bars and Columns charts for visual patterns

## **POST-ANALYSIS & FINAL INSIGHTS**

### **a. Key Findings**

1. Parental Education is the strongest predictor of academic success:
  - Master's degree parents → highest scores (221 average)
  - Bachelor's → second (216)
  - High School only → lowest (189)
2. Test Preparation works
  - Completers → 225 average
  - Non-completers → 199 average
3. Females consistently outperform males, especially in reading and writing.
4. Race Group E and D dominate, while Group A needs strong support.
5. Standard lunch students outperform free/reduced lunch students by a wide margin.

### **b. Comparison with Early Insights**

- Initial assumptions about parental education and test preparation were confirmed.
- Gender performance differences remained consistent throughout.
- Race-based performance differences were stronger than expected.
- Nutrition emerged as a more significant factor than initially assumed.

## **DATA VISUALIZATIONS & CHARTS**

Charts used in Excel:

- Bar charts: To evaluate parental level of Education, performance by gender and test preparation.
- Column chart: to determine number of students for each grouped math, writing and reading score.



- Pie chart: To determine the percentage of pass and fail status.

## **RECOMMENDATIONS & OBSERVATIONS**

### **a. Actionable Insights**

- Expand access to test prep programs.
- Build specialized math support interventions.
- Develop equity initiatives for disadvantaged parental groups.
- Strengthen nutrition and lunch support programs.
- Track racial performance trends for targeted improvement.

### **b. Optimizations / Decisions**

- Allocate more math instructional hours.
- Implement mentorship initiatives.
- Track long-term performance by demographic groups.
- Implement meal subsidies to improve nutrition-related performance gaps.

### **c. Unexpected Outcomes**

- Lunch type stronger predictor than expected.
- Test prep yielded a surprisingly large effect.
- Female performance stronger than initial assumptions.

## **CONCLUSION**

This Excel-driven analysis of Harvard University student performance paints a clear story. Academic success is heavily influenced by parental education, preparation programs, gender dynamics, nutrition, and race/ethnicity.

With the right mix of equitable support, targeted teaching strategies, and school-wide interventions, institutions can significantly raise performance outcomes and close achievement gaps.

## **REFERENCES & APPENDICES**

1. Adithya Shetty, 2020. *Kaggle - Student performance Dataset*. Available at: <https://www.kaggle.com/datasets/adithyabshetty100/student-performance>  
(Accessed: 21st November 2025)