Bias & Equality: Performance Dataset

PHIL 123: Internet, Soc, & Phil

By: Finn Eckhart, May Phoudavong, Zachary Thill, Bridger Christie

Student_performance_data _ 2

StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1001	17	1	0	2	19.833722807854700	7	1	2	0	0	1	0	2.929195591667680	2.0
1002	18	0	0	1	15.408756055584670	0	0	1	0	0	0	0	3.042914833436380	1.0
1003	15	0	2	3	4.210569766881226	26	0	2	0	0	0	0	0.112602254461820	4.0
1004	17	1	0	3	10.028829473958200	14	0	3	1	0	0	0	2.054218139702950	3.0
1005	17	1	0	2	4.672495272973300	17	1	3	0	0	0	0	1.288061181796390	4.0
1006	18	0	0	1	8.191218545250190	0	0	1	1	0	0	0	3.084183614486390	1.0
1007	15	0	1	1	15.601680474699300	10	0	3	0	1	0	0	2.748237414891580	2.0
1008	15	1	1	4	15.424496305808100	22	1	1	1	0	0	0	1.360142712316460	4.0
1009	17	0	0	0	4.562007558047700	1	0	2	0	1	0	1	2.896819189513570	2.0
1010	16	1	0	1	18.444466363097200	0	0	3	1	0	0	0	3.573474210329770	0.0
1011	17	0	0	1	11.851363655296500	11	0	1	0	0	0	0	2.147171629185100	3.0
1012	17	0	0	1	7.598485819024029	15	0	2	0	0	0	1	1.559594519040280	4.0
1013	17	0	1	1	10.038711615617200	21	0	3	1	0	0	0	1.520077814874810	4.0
1014	17	0	1	2	12.101425068754900	21	0	4	0	1	0	0	1.751580958334080	4.0
1015	18	1	0	1	11.197810636915700	9	1	2	0	0	0	0	2.396788117124800	3.0
1016	15	0	0	2	9.728100710723560	17	1	0	0	1	0	0	1.341520716534670	4.0
1017	18	0	3	1	10.098656081788000	14	0	2	1	1	0	0	2.232175277159800	3.0
1018	18	1	0	0	3.528238208577200	16	1	2	0	0	0	0	1.384404175694030	4.0
1019	18	0	1	3	16.254658086090060	29	0	2	1	0	0	1	0.469553323379870	4.0
1020	17	0	0	1	10.835206398820300	9	0	2	0	0	1	0	2.395784094530700	3.0
1021	16	1	0	3	2.621597234094060	2	0	3	0	0	0	1	2.778411299920650	2.0
1022	15	0	0	2	15.323142031655600	25	0	1	1	0	0	0	0.346894036705015	4.0
1023	16	1	1	0	18.648879567547000	29	1	1	0	0	0	0	0.312546230525355	4.0
1024	18	1	3	4	18.946137984739820	20	0	2	1	0	0	0	1.770131876779970	4.0
1025	18	1	0	1	7.380354648223460	15	0	2	0	0	0	0	1.505155622036290	4.0
1026	16	1	0	3	2.710337471215080	5	0	4	0	0	1	0	2.977851918315740	2.0
1027	16	0	0	1	10.367992532661300	2	0	2	0	1	0	0	2.948717671911920	2.0
1028	16	1	0	3	2.252184586984410	8	0	3	0	0	1	0	2.145204720467190	3.0
1029	18	0	0	0	18.679748370252300	10	0	3	1	0	0	0	2.854803928981320	2.0
1030	18	0	0	2	3.671592547102920	20	0	3	1	0	0	0	1.519441725815140	4.0
1031	15	0	2	2	5.055317199023280	12	1	0	0	1	0	0	1.727120309522410	4.0

Source: Rabie El Kharoua’s "Student Performance Dataset" (Kaggle).

- 2,392 High School Students
- **Key Attributes:** Demographic (Ethnicity, Age), Behavioral (Study Time, Absences), and Social (Parental Support, Extracurriculars) used to predict student's final GPA

Preliminary Data Findings

Negative Factors For GPA:

Many Absences

Low Parental Support

Low Amount of Study Time Per Week

Neutral Factors For GPA:

Volunteer Work

Age

Gender

Ethnicity

Positive Factors For GPA:

High Amount of Study Time Per Week

High Parental Support

Tutoring

Extracurricular Activities - Participation in sports and music

Question:

Rather than solely focusing on GPA, we examine how variables like parental support and extracurricular activities shape how well a student will do.

How do the standardized formats used to track extracurriculars predict a student's academic success?

Methodology – Poirier’s Reading Method

The Method: We used Lindsay Poirier’s Method, applying three layers of analysis: Denotative, Connotative, and Deconstructive to find what is "missing" or "marginalized" in the data, examining the politics behind. Engaging a deconstructive practice can help prompt us to read denotatively. We asked: How is "Parental Support" measured? Is it just a survey? Does this ignore the quality of the relationship or the reasons for low support (like a parent working three jobs)?

Methodology – Koopman’s Format Anatomies

We utilize Colin Koopman’s Format Anatomies to help us investigate how the technical structure of the spreadsheet forces students into a rigid informational format. Rather than focusing on individual GPA scores, we look at the rows, columns, and categories to see how they anatomize a person into a digital subject. It is not just "cleaning" the data, it is a process and finding how the format and column header molds and creates a complex student into a predictable informational person

Denotative

- What does each column explicitly measure?
- What are the technical limits of the dataset?
- How are variables defined?

GPA → Numeric average (scale 0.0 –4.0)

Study Time → Number of hours per week (ranging from 0 to 20)

Gender → Categorical variable (0=Male,1=Female)

Parental Education → Highest completed level

Extracurricular Activities →Participation (No = 0, Yes = 1)

Deconstructive

- Are there any relevant missing variables?
- What experiences are erased?
- What categories are oversimplified?

Missing Variables:

- Mental health
- Economic instability
- Caregiving responsibilities
- Disability accommodations
- Language barriers

Connotative

- What cultural meaning is attached to these variables?
- What assumptions are embedded into these categories?

Parental Education Level→ household/education orientation stability

Extracurricular Activities → Participation shows ambition, determination, perseverance, and well-roundedness

Study Time → The number of hours per week devoted shows discipline and a work ethic

Tutoring → Access to additional academic support shows resourcefulness, willingness to seek help, and proactive learning habits

Conclusion & Impact

Ethical Implications

- Study time and consistent attendance showed the strongest positive relationship with GPA.
- Support systems such as tutoring and parental support can improve academic performance.
- Extracurricular activities may contribute to student development but do not always directly raise GPA.
- Datasets like this simplify complex student identities into measurable variables.

Key takeaways

- Data models can reinforce bias if certain social factors are weighted unfairly.
- Student identities are more complex than numerical variables.
- Predictive systems used in education should be transparent and carefully evaluated.
- Future research should examine how digital data shapes opportunities for students.