Student Performance Dataset

Presentation by Cailyn, Louise, and Lexi

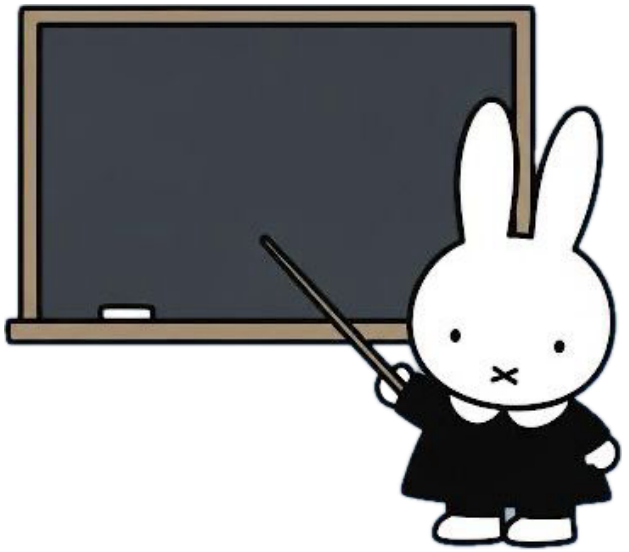
Overview

This dataset contains information regarding the academic performances of 2,392 high school students and details the demographics that may affect a student's ability to succeed. The dataset then categorizes these various demographics (detailing studying habits, parental involvement, extracurricular activities, etc.) along with students' grade point averages to model a predictive and statistical analysis of the dataset.

The classification of students' grades, based on one's GPA, are recorded into *GradeClass* and reveals a comprehensive view of the multiple factors that influence a student's academic reputation.

Initial Reading of the Dataset

A dataset that has variables which are predictive of student success. In the initial dataset, demographics such as students' age, ethnicity, and gender are categorized alongside parental involvement, education of the parents, student's volunteer hours and extracurricular activities, as well as noting the weekly studying schedules of pupils, absences and even grade point averages. However, the dataset fails to provide relevant information regarding student's non academic-focused information that directly affects one's performance. The initial reading of the dataset reveals that relative statistics have been excluded, regardless of reason, to make the data collected appear potent to the academic success of high school students.



Academic Success Factors in High School Students – Rabie El Kharoua

StudentID	Age	Gender	Ethnicity	ParentalEc	StudyTime	Absences	Tutoring	ParentalSu	Extracurric	Sports	Music	Volunteeri	GPA	GradeClass
1001	17	1	0	2	19.83372	7	1	2	0	0	1	0	2.929196	2
1002	18	0	0	1	15.40876	0	0	1	0	0	0	0	3.042915	1
1003	15	0	2	3	4.21057	26	0	2	0	0	0	0	0.112602	4
1004	17	1	0	3	10.02883	14	0	3	1	0	0	0	2.054218	3
1005	17	1	0	2	4.672495	17	1	3	0	0	0	0	1.288061	4
1006	18	0	0	1	8.191219	0	0	1	1	0	0	0	3.084184	1
1007	15	0	1	1	15.60168	10	0	3	0	1	0	0	2.748237	2
1008	15	1	1	4	15.4245	22	1	1	1	0	0	0	1.360143	4
1009	17	0	0	0	4.562008	1	0	2	0	1	0	1	2.896819	2
1010	16	1	0	1	18.44447	0	0	3	1	0	0	0	3.573474	0
1011	17	0	0	1	11.85136	11	0	1	0	0	0	0	2.147172	3
1012	17	0	0	1	7.598486	15	0	2	0	0	0	1	1.559595	4
1013	17	0	1	1	10.03871	21	0	3	1	0	0	0	1.520078	4
1014	17	0	1	2	12.10143	21	0	4	0	1	0	0	1.751581	4
1015	18	1	0	1	11.19781	9	1	2	0	0	0	0	2.396788	3
1016	15	0	0	2	9.728101	17	1	0	0	1	0	0	1.341521	4
1017	18	0	3	1	10.09866	14	0	2	1	1	0	0	2.232175	3
1018	18	1	0	0	3.528238	16	1	2	0	0	0	0	1.384404	4
1019	18	0	1	3	16.25466	29	0	2	1	0	0	1	0.469553	4
1020	17	0	0	1	10.83521	9	0	2	0	0	1	0	2.395784	3
1021	16	1	0	3	2.621597	2	0	3	0	0	0	1	2.778411	2
1022	15	0	0	2	15.32314	25	0	1	1	0	0	0	0.346894	4
1023	16	1	1	0	18.64888	29	1	1	0	0	0	0	0.312546	4
1024	18	1	3	4	18.94614	20	0	2	1	0	0	0	1.770132	4

How the Data is Measured

- **Student ID**: A unique identifier assigned to each student (1001 to 3392).
- **Age**: The age of the students ranges from 15 to 18 years.
- **Gender**: Gender of the students, where 0 represents Male and 1 represents Female.
- **Ethnicity**: The ethnicity of the students, coded as follows: 0: Caucasian, 1: African American, 2: Asian, 3: Other
- **Parental Education**: The education level of the parents, coded as follows: 0: None, 1: High School, 2: Some College, 3: Bachelor's, 4: Higher
- **Study Time Weekly**: Weekly study time in hours, ranging from 0 to 20.
- **Absences**: Number of absences during the school year, ranging from 0 to 30.
- **Tutoring**: Tutoring status, where 0 indicates No and 1 indicates Yes.
- **Parental Support**: The level of parental support, coded as follows: 0: None, 1: Low, 2: Moderate, 3: High, 4: Very High
- **Extracurricular**: Participation in extracurricular activities, where 0 indicates No and 1 indicates Yes.
- **Sports**: Participation in sports, where 0 indicates No and 1 indicates Yes.
- **Music**: Participation in music activities, where 0 indicates No and 1 indicates Yes.
- **Volunteering**: Participation in volunteering, where 0 indicates No and 1 indicates Yes.
- **GPA**: Grade Point Average on a scale from 2.0 to 4.0, influenced by study habits, parental involvement, and extracurricular activities.
- **GradeClass**: Classification of students' grades based on GPA : 'A' (GPA >= 3.5) 1: 'B' (3.0 <= GPA < 3.5) 2: 'C' (2.5 <= GPA < 3.0) 3: 'D' (2.0 <= GPA < 2.5) 4: 'F' (GPA < 2.0)

Our Research Question and Hypothesis

Research Question: Why are certain predictive variables being left out of this dataset and what are the societal implications?

- We ask this, specifically, because subjective analysis of any data is under review for inconsistencies in research, such as excluding demographics that on a case-by-case basis, effects the overall report of the research and reveal information about society and the educational system.

Hypothesis: Several demographics that are excluded from the dataset, despite being relevant to a student's academic performance, are done so to influence the initial datasets ability to appear enlightened on the educational system however this is done inefficiently by excluding said statistics.

- We chose this as our hypothesis because the data provided unfairly excludes information that is necessary to fully understand student's personal situations (ergo, their academic success)

Method 1 - Koopman's "format anatomies"

- **Microsystem**: organized around individual student units and focuses on personal characteristics.
- **Macrosystem**: demographics including gender, ethnicity, and socioeconomic status.
- No deep variables including school funding differences, educational policy, or systemic inequality.
- Heavy focus on micro-level behavior and minimal representation of macro-level structures.
- Data set is individually centered.

Method 2 - Priorier's "Reading Datasets"

Denotative:

- Student demographics, study habits, academic performance are present.
- Missing family conflict, teacher bias, peer relationships, cultural attitudes toward education, etc.
- These could be missing because they are harder to quantify.

Connotative:

- Implied meanings.
- The missing variables may imply that emotional and structural factors are considered secondary.

Deconstructive:

- Structural bias.
- Data should be measurable and numerical. Individual behavior is the primary explanatory variable.

Data Provided

The data provided in the analysis include the following,

- Extracurricular Activities (Sports, Music, Tutoring)
- Parental Support/Education
- Volunteer Hours
- Studying Habits (Weekly)
- Absences

While it projects a relative statistical analysis of high schoolers' success, it does not fully reveal the full truth or perspectives of students with specific demographics that could directly affect academic growth.

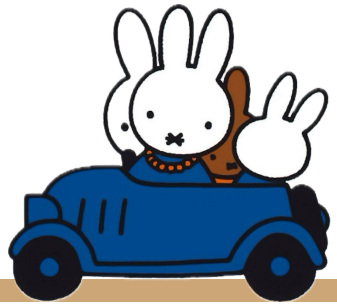
Data Excluded

These excluded examples of information did not make it into the dataset, yet their relativity to a student's academic success is pressing enough to require all the information needed to truly analysis a high schooler's academic experience.

Examples include,

- Parents' income
- Students work hours
- Number of dependents in the household (siblings, special needs dependents)
- Number of guardians in the household (Single parents, grandparents, etc.)

Information related to these demographics are crucial to statistically analyzing high school students, with such different experiences and statistics playing a role in one's academic success. For example, most high school students are required to work part-time, some even full time. Others may have younger siblings that keep them occupied from school work, and households with single parents may have a harder time juggling school, extracurricular activities, multiple dependents, etc. A student's performance is not just determined by their ability to excel or fall short, but also by outside influences that directly inhibit one's academic success rate. To analyze students without taking these and other excluded demographics into account results in an incomplete, non transparent statistical dataset.



Concluding Statements

In conclusion, while Mr. Rabie El Kharoua (owner of the original dataset) details the demographics of 2,392 high school students to categorize students' grades into classifications that provide predictive modeling and statistical analysis, this dataset fails to consider other various demographics that impact one's academic performance and use generalized scales with subjective data.

Thank you for your time!