

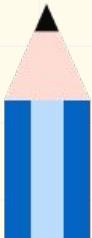
Understanding Student Performance: What Factors Contribute to Success?

This presentation explores the factors that influence consistent student performance over time using data sourced from the UCI Machine Learning Repository - Student Performance Dataset. The dataset includes information about student demographics, family background, study habits, and grades. Their approach includes analyzing the data using linear regression to identify factors influencing final exam performance and logistic regression to predict whether a student shows a consistent increase or decrease in performance.



Understanding Student Performance: What Factors Contribute to Success?

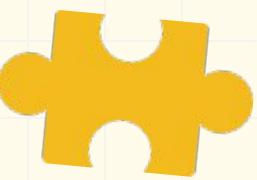
Presenters:
Sri Anirudh Parimal





Data Source and Overview

- Performance of High school students in Portuguese language subject across 3 tests in the year and factors that impact the grades.
- 650 rows x 30 features, 3 response variables
- Includes information like
 - student demographics(age,school,ambition,travel)
 - family background(parents education and job)
 - study habits(study hours,support)
 - life style(alcohol,Relationship,Ex-Curricular)
 - grades(G1,G2,G3-Marks out of 20)



Data transformations

- We modified the dataset by calculating percentile ranks and creating a "flag" variable to track performance consistency.

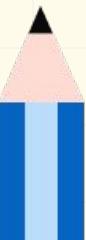
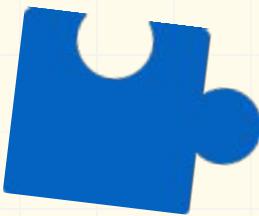
Additional Columns

Gi_perc=percentile generated from marks in Gi

Flag = 0 if G1_perc>G2_perc>G3_perc

, 1 if G3_perc>G2_perc>G1_perc

- Handled outliers from Absences with mean value.





Objective and Our Approach

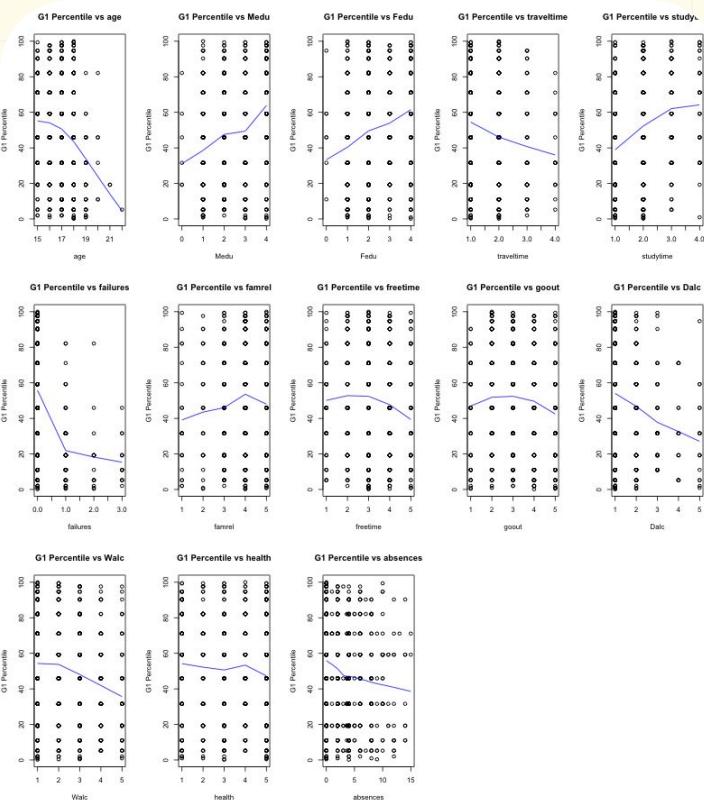
Objective : Identify factors influencing final exam performance

Approach: Linear Regression with response variable as G3_perc

Objective: Predicting whether a student shows a consistent increase or decrease in performance

Approach: Logistic Regression Model with flag column

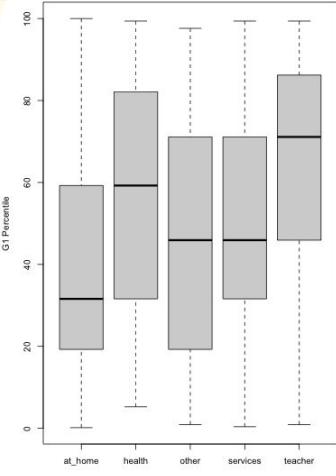
Limitations of these models: These Models are built for interpretability. Rsq2 and Acc are very low so wouldn't be used for precise predictions.



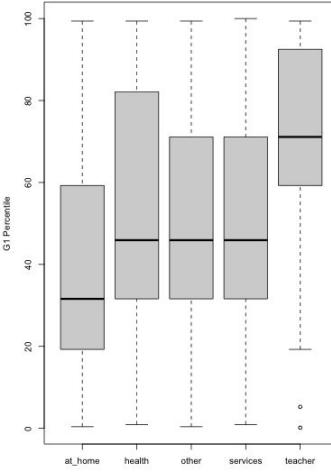
How We Analyzed the Data

Employed scatter plots for numerical variables to investigate continuous relationships and trends between student characteristics (like age, study time, etc.) and their academic performance.

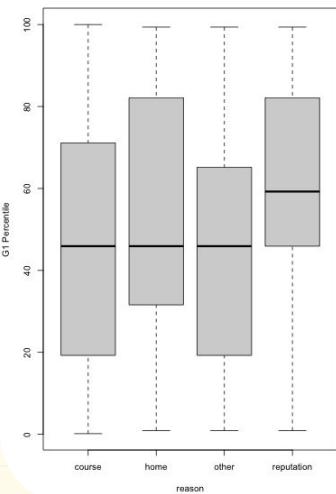
G1_perc by Mjob



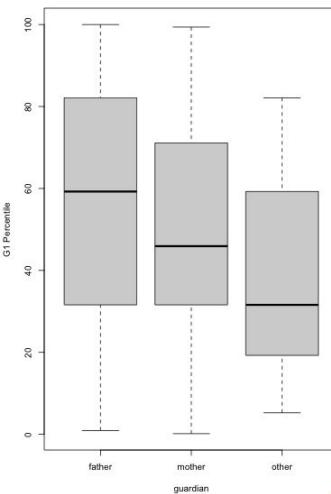
G1_perc by Fjob



G1_perc by reason

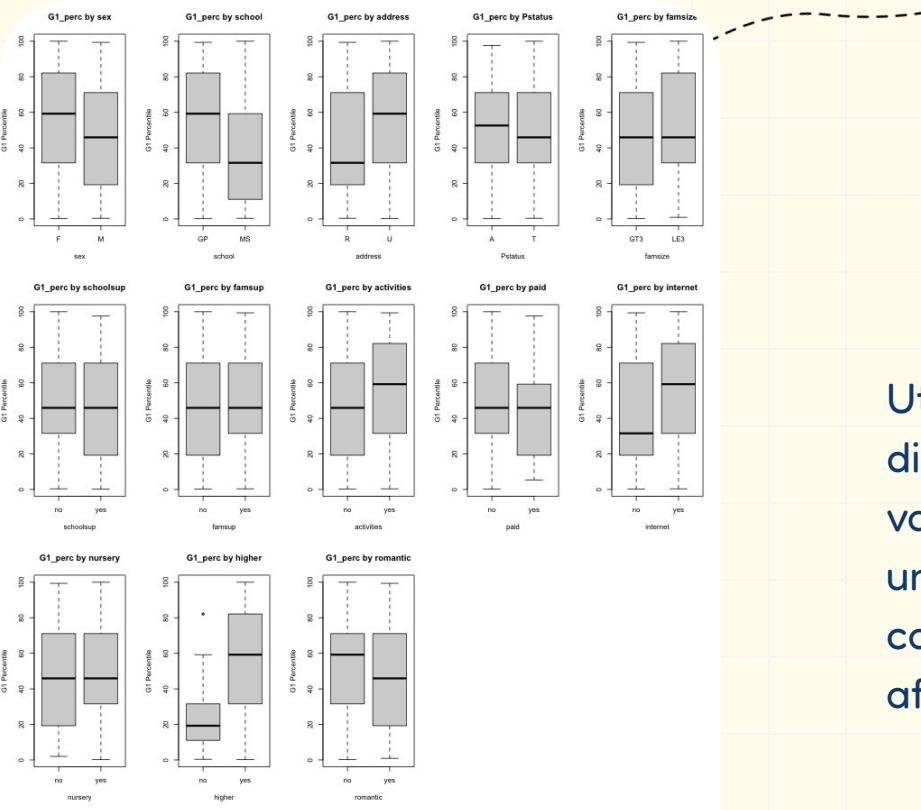


G1_perc by guardian



How We Analyzed the Data

Utilized box plots to analyze the distribution for nominal variables. This helped in understanding how different categories within these variables affect student grades.



How We Analyzed the Data

Utilized box plots to analyze the distribution and outliers for binary variables. This helped in understanding how different categories within these variables affect student grades.

Key Findings - Linear Regression

- From R-squared value around 33% of variance is explained of the G1_perc
- Students aiming for **higher education** tend to get better grades.
- Higher **maternal education** is associated with better student performance.
- Students who have **fathers working as teachers** tend to have higher performance.
- As expected, **study time** has a positive impact on G1_perc, reinforcing the value of study habits.
- Students receiving **extra educational support** have lower G1_perc. This could mean support programs are targeted at struggling students.
- Unsurprisingly, more failures lead to drastically lower G3_perc.

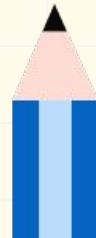
Key Findings - Logistic Regression

- Accuracy of the model is only 0.59. If I say all the values are true, then the accuracy would be 0.63. So, the model is **not accurate**. But the model is able to tell us factors that influence it.
- Students at **Mousinho da Silveira school** are about 142% more likely to show continuous improvement in performance compared to other school
- If the father's job is categorized as '**athome**', there's at least 70% increase in the likelihood of the student's performance continuously improving compared to other professions.
- Each additional year in age is associated with a 46% increase in the likelihood of continuous performance improvement.



Surprising Results

- Extracurriculars, paid classes, internet access, and even alcohol consumption didn't strongly influence performance consistency.
- Other factors might be more important or how a student engages in these activities might matter more than just participation.
- Surprisingly, students aiming for higher education and those getting extra school support also had lower odds of consistent improvement.



ON EXTRACURRICULAR PERFORMANCE



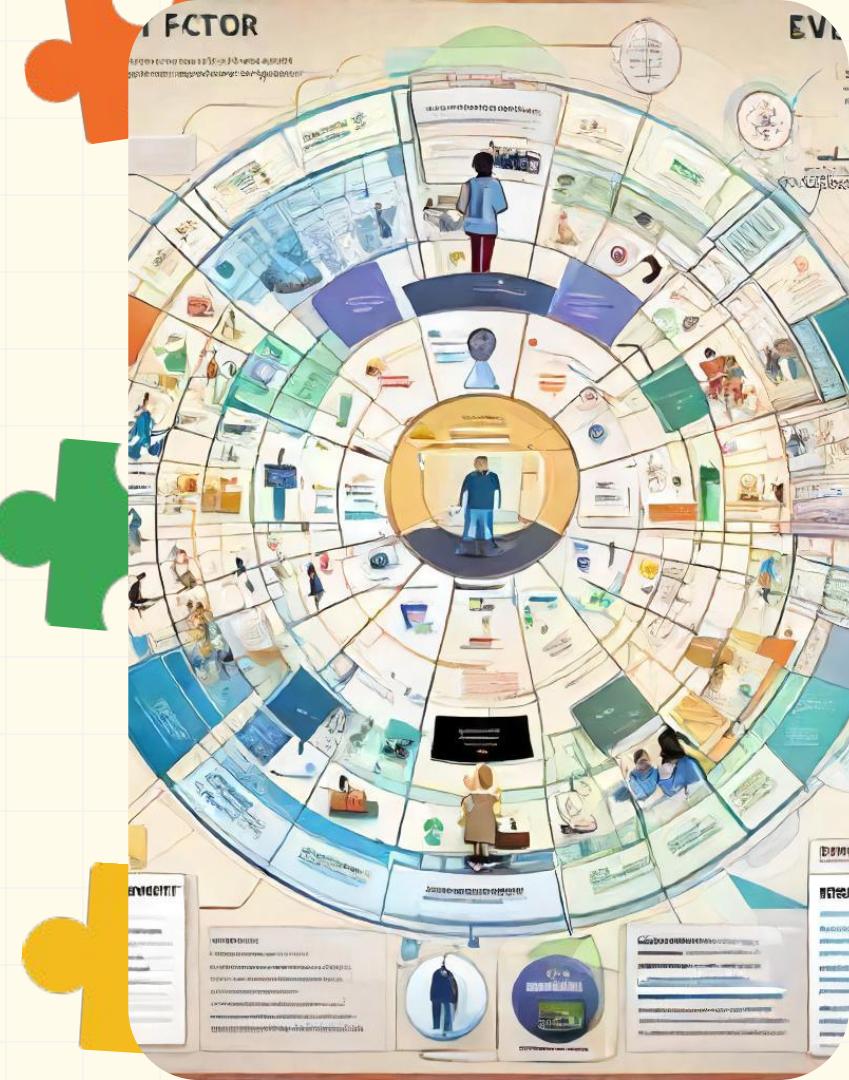
What Didn't Matter Much

Other factors might be more important or how a student engages in these activities might matter more than just participation.

Extracurriculars, paid classes, internet access, and even alcohol consumption didn't strongly influence performance consistency.

The Takeaway

- Student performance is complex, influenced by individual, familial, and schoolrelated factors.
- Our models provide insights, but more research is needed for accurate individual predictions.
- Highlight the most important factors for both consistent improvement and final exam results.



Thank you. Please feel free to ask any questions.

