



Analysis and Identification of Student Performance Factors

By your good friends Tatum Good, Suzanne Gunderson,
Brooke Proctor, and Courtney St. Onge

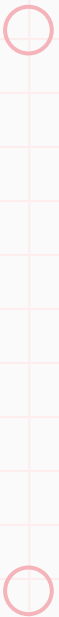


01

Intro and Background



Background




A Google Scholar search utilizing the key terms “what factors create an ideal student” yielded 4,490,000 results. After restricting the timeframe to be articles published more frequently than 2021, 17,200 results remained. A brief look into the most relevant 10 articles helped to inform the variables and analysis we did. This quick research helped us significantly build up our domain knowledge of student performance factors.



Background

“The concept of the ideal student cannot be confined to a singular model but rather focuses on creating a multifaceted framework that accommodates diverse ideals” (Esmaeeli, Arasteh, Zeinabadi, and Abasian, 2024).

“The ideal student is not only academically ambitious, capable, curious and motivated, but also an energetic and passionate member of the wider university community” (Wong, DeWitt, and Chiu, 2021).



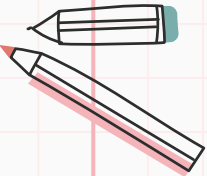
② % +

Data Overview

StudentPerformanceFactors.csv

- One CSV file
- 6,607 observations included
- Sourced from Kaggle

Description: This synthetic dataset provides a comprehensive overview of various factors affecting student performance in exams. It includes information on study habits, attendance, parental involvement, and other aspects influencing academic success.



Variables

Hours Studied

Attendance

**Parental
Involvement**

**Access to
Resources**

**Extracurricular
Activities**

Sleep Hours

**Previous
Scores**

**Motivation
Level**

Internet Access

**Tutoring
Sessions**

Family Income

Teacher Quality

School Type

Peer Influence

**Physical
Activity**

**Learning
Disabilities**

**Parental
Education Level**

**Distance from
Home**

Gender

Exam Score

Variables

Hours Studied

Attendance

**Parental
Involvement**

**Access to
Resources**

**Extracurricular
Activities**

Sleep Hours

**Previous
Scores**

**Motivation
Level**

Internet Access

**Tutoring
Sessions**

Family Income

Teacher Quality

School Type

Peer Influence

**Physical
Activity**

**Learning
Disabilities**

**Parental
Education Level**

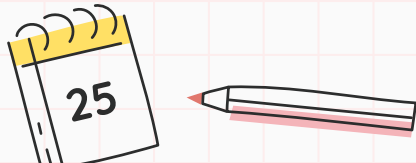
**Distance from
Home**

Gender

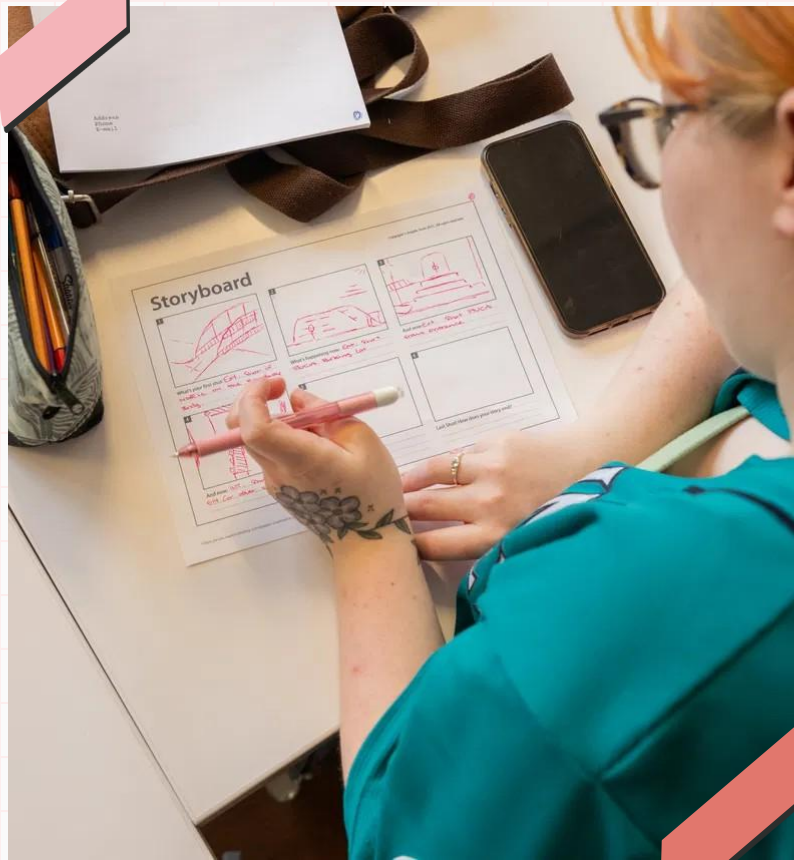
Exam Score

Project Goals

- Use Exploratory Data Analysis to identify trends and relationships within the data.
- Use machine learning models for prediction.
- Use linear regression for classifying performance levels
- Do statistical analysis to find correlations between features.



%



02

Research



Research Questions

- What relationship exists between **sleep hours** and exam scores? What outliers/leverage points can be identified in this relationship, if any?
- What is the relationship **past** and **present exam scores**?
- How do **study habits** and **attendance** impact the final exam scores? Does one of these variables impact final exams scores more than the other?
- How do **socioeconomic factors** (family income, school type, parental education level, access to resources, etc.) affect academic performance?
- What are the interactions between **motivation**, **sleep hours**, and **final exam scores**?
- What makes an **“ideal” student**? What can our criteria be for that - ideally an even mix of traits like study habits and access to resources etc.?



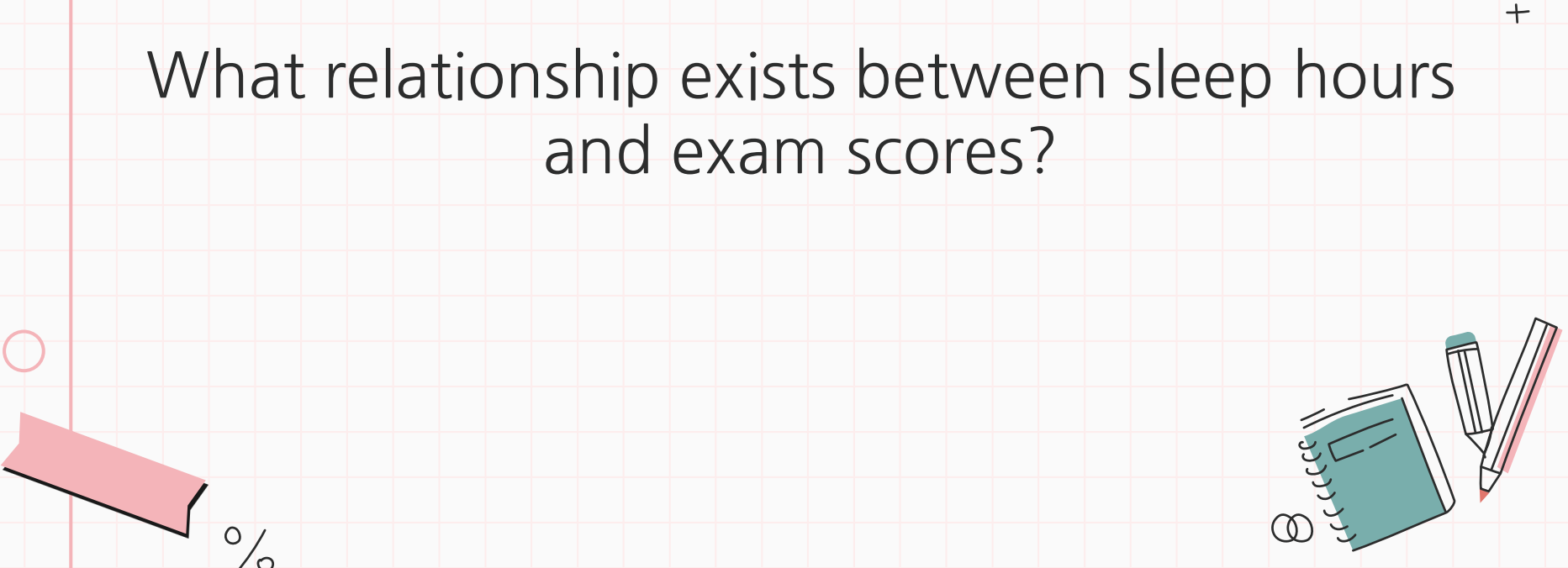
03

The Analysis

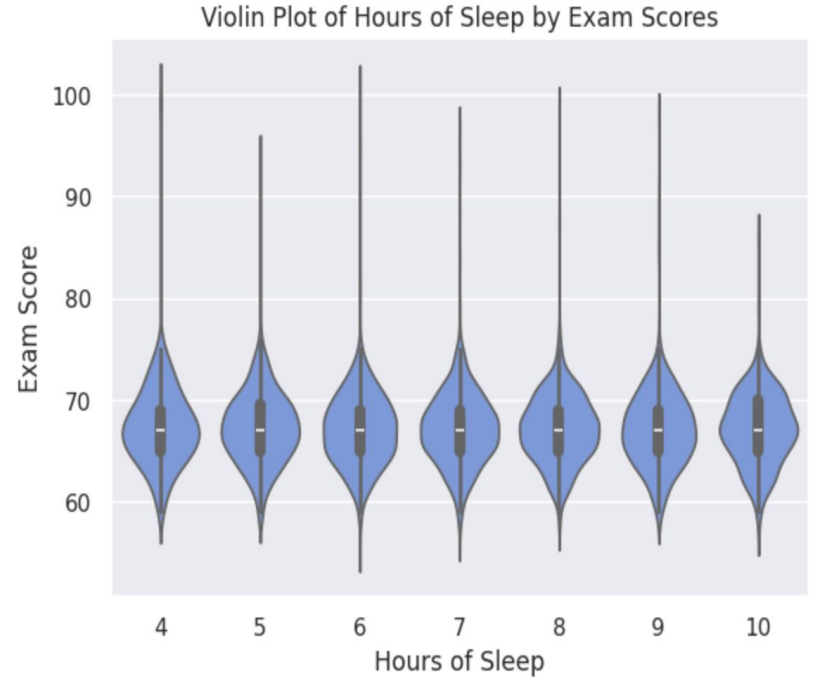
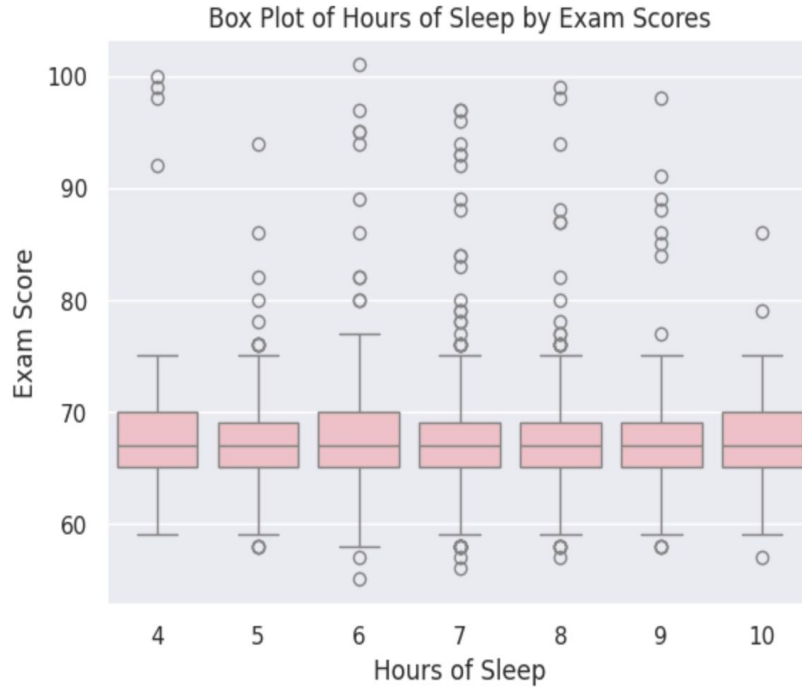


GUIDING QUESTION

What relationship exists between sleep hours
and exam scores?



Sleep & Exam Scores



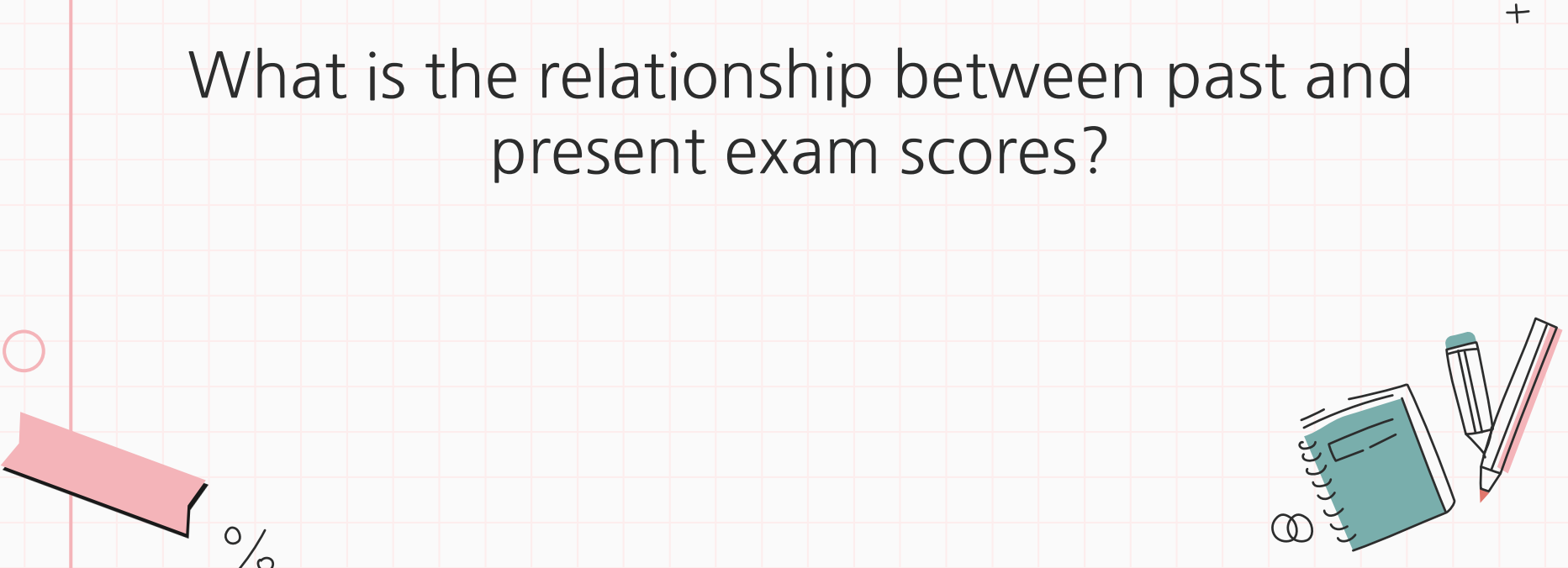
+ %





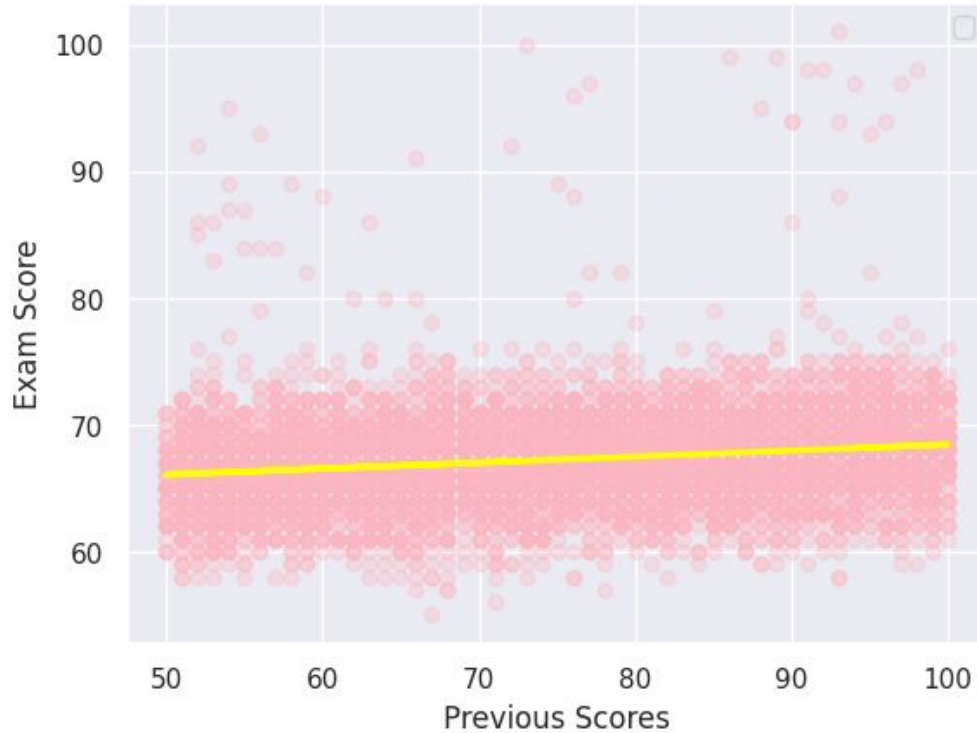
GUIDING QUESTION

What is the relationship between past and present exam scores?



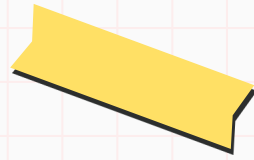
Past & Present Exam Scores

Scatter plot of student attendance and exam scores with linear regression



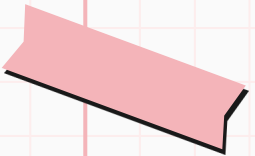


GUIDING QUESTIONS



How do study habits and attendance impact ⁺ the final exam scores?

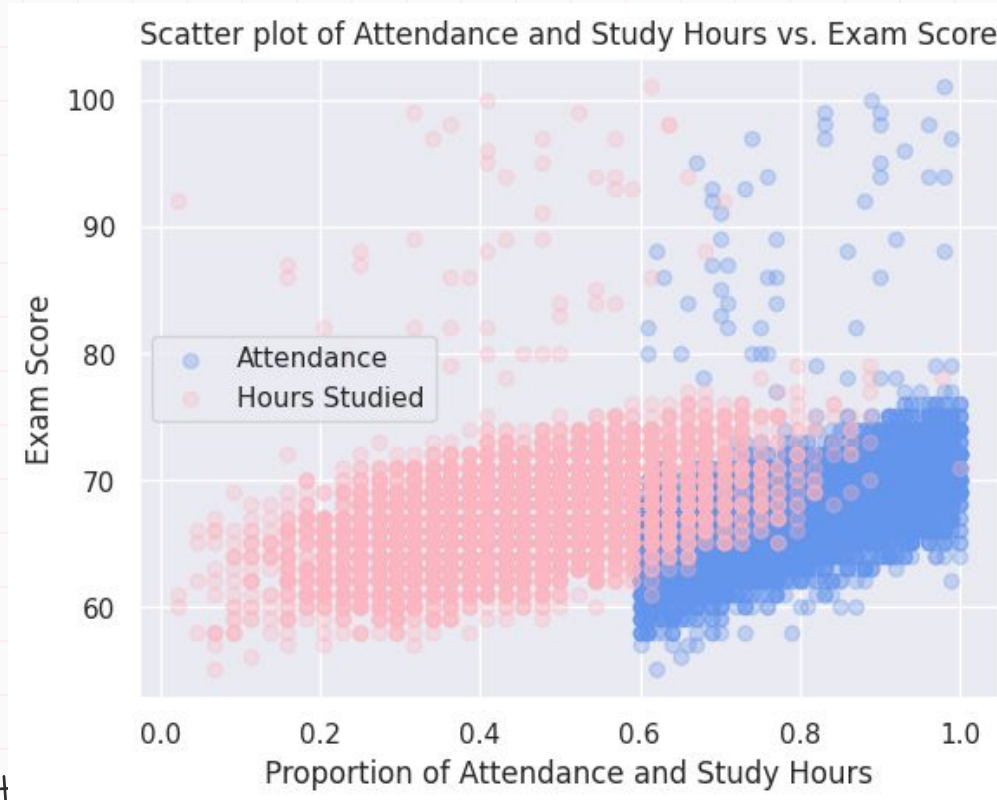
Does one of these variables impact final exams scores more than the other?



%



Study Habits, Attendance, & Exam Scores





GUIDING QUESTION

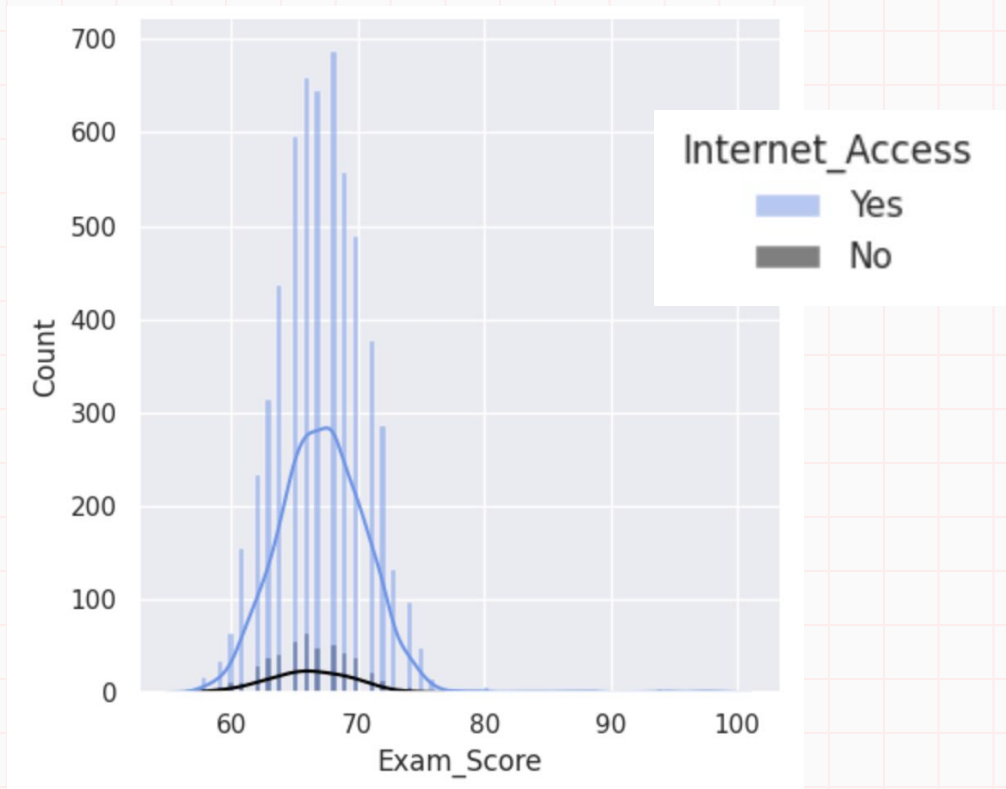
How do socioeconomic factors impact final exam scores?



+ %



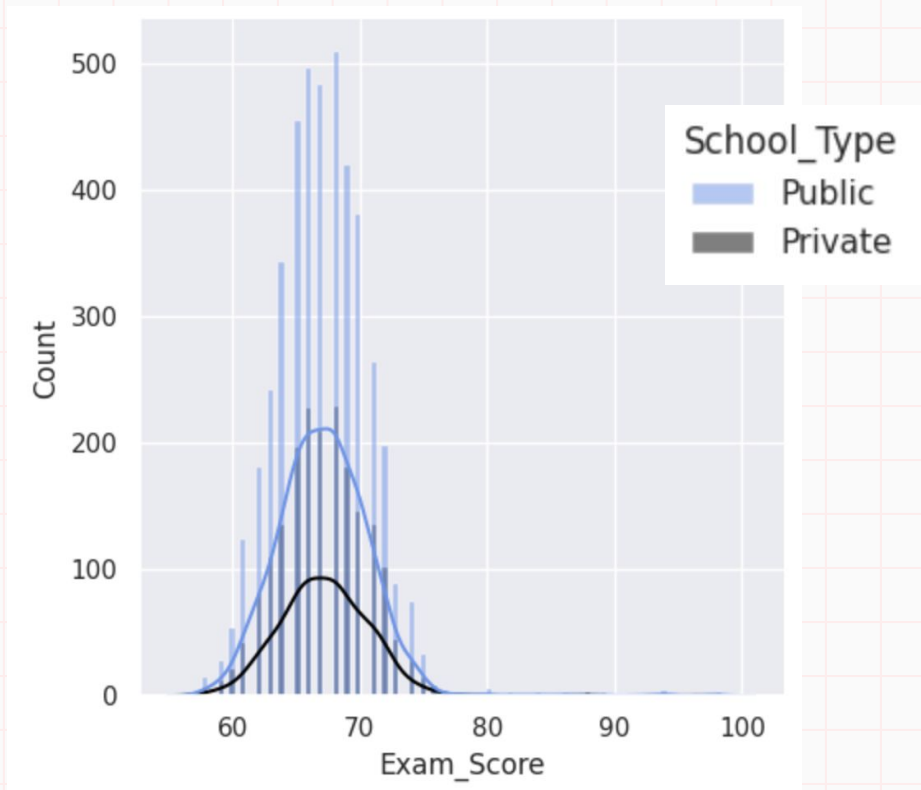
Internet Access & Exam Performance



+ %



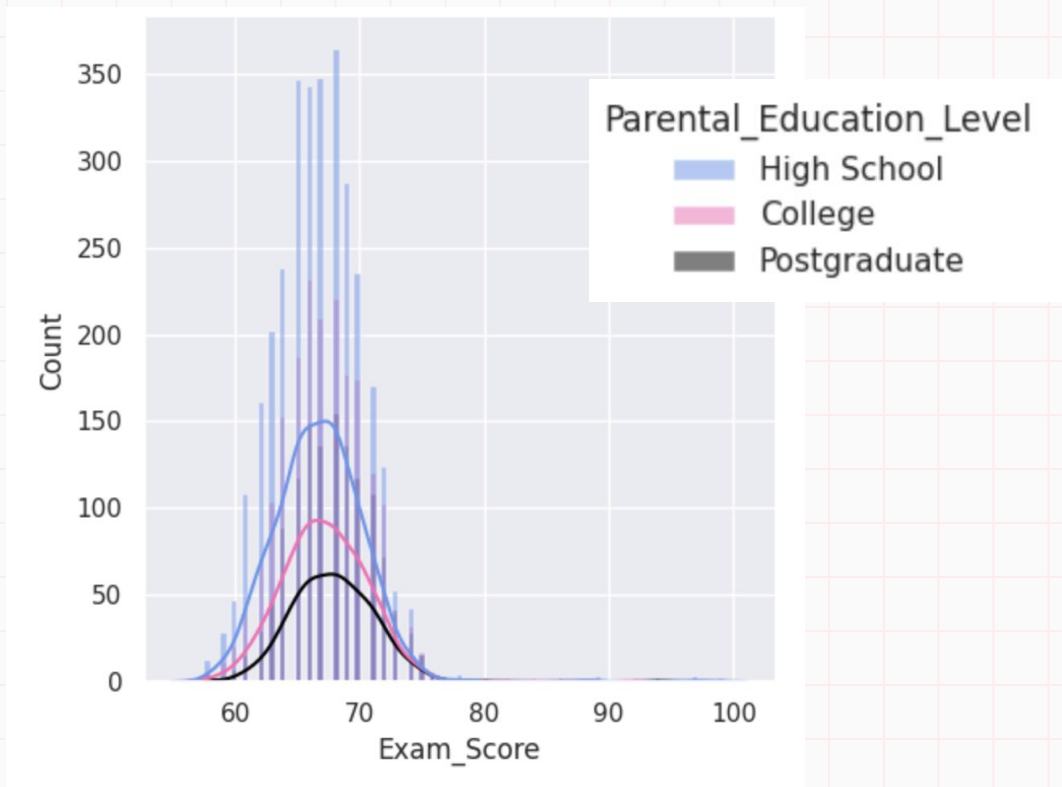
School Type & Exam Performance



+ %



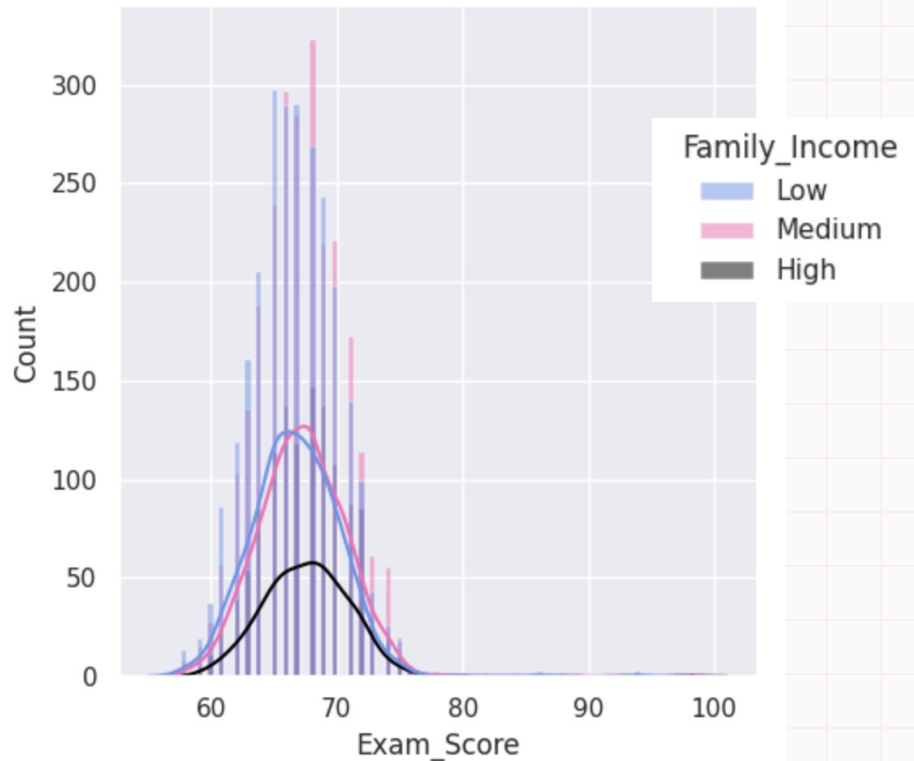
Parental Education Level & Exam Performance



+ %



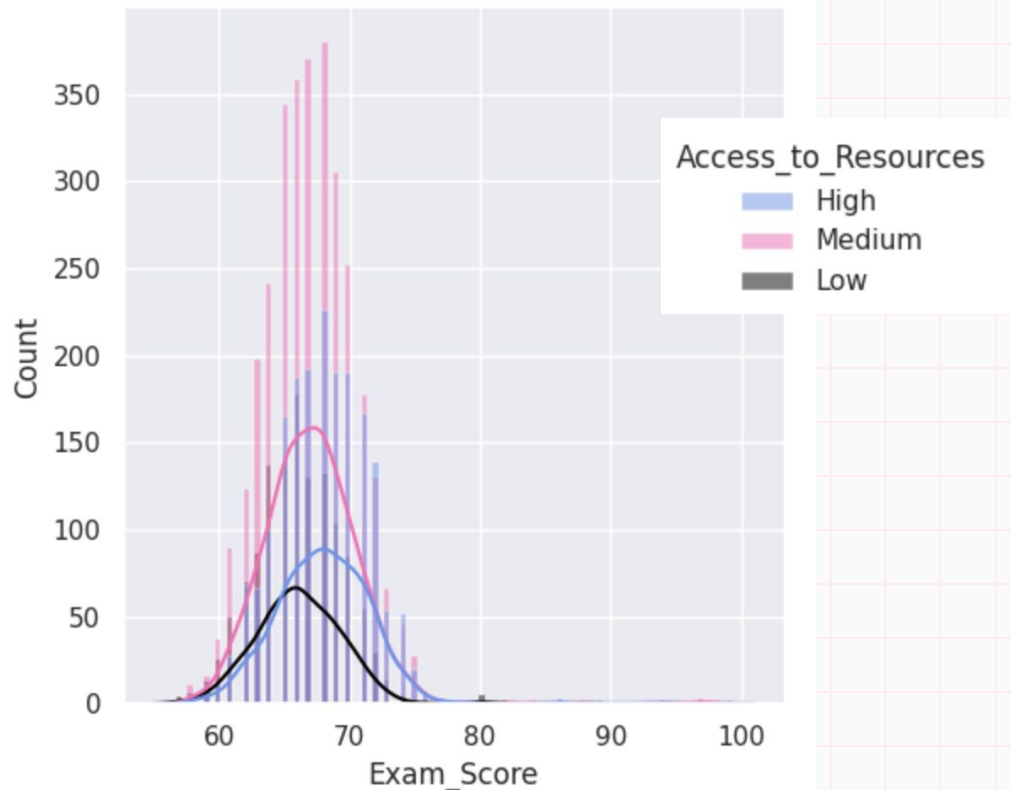
Family Income & Exam Performance



+ %



Access To Resources & Exam Performance



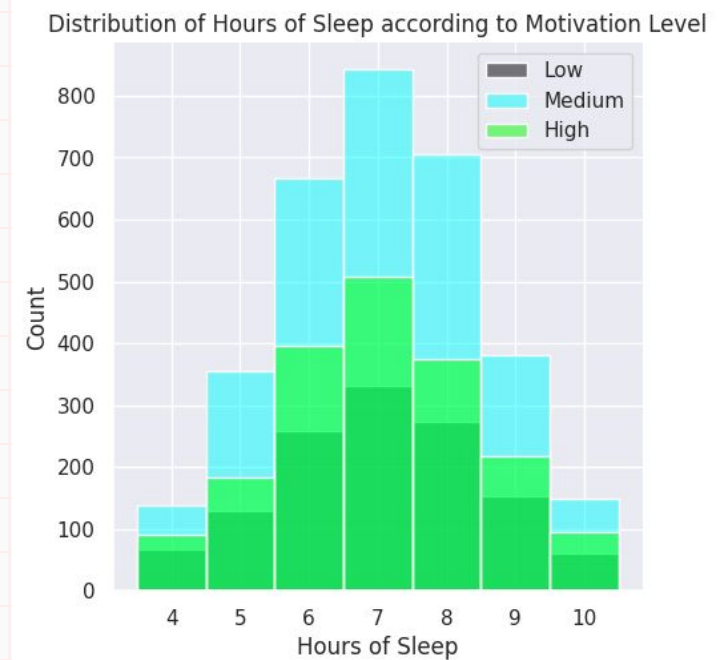
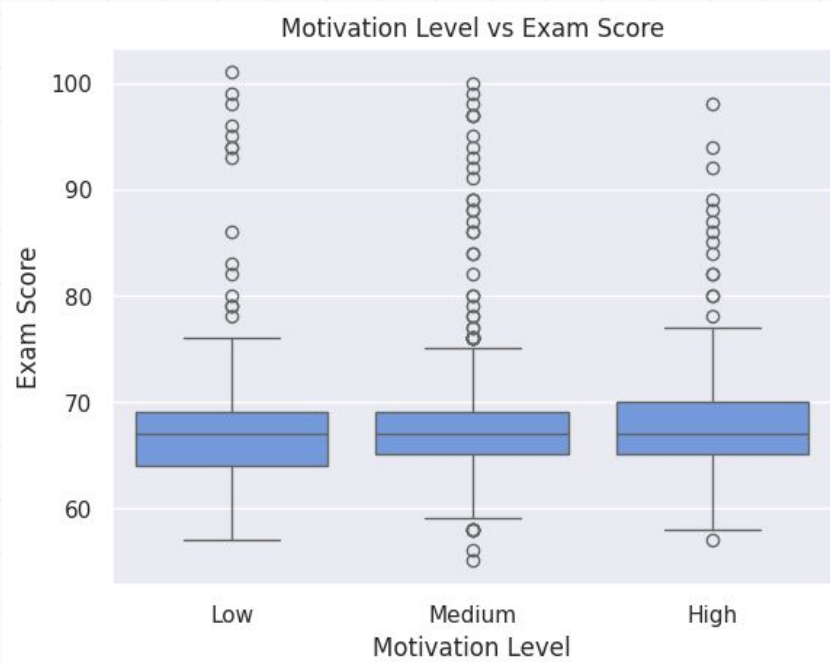


GUIDING QUESTION

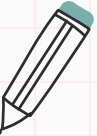
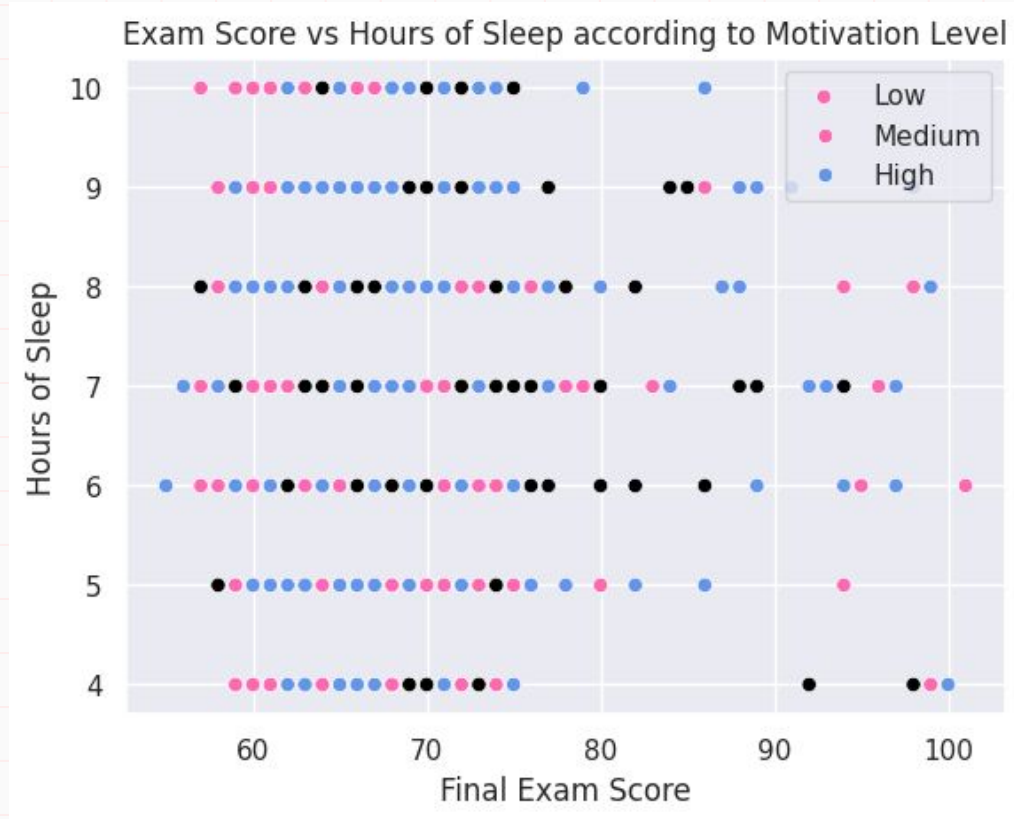
What are the interactions between motivation, sleep hours, and final exam scores



Motivation, Sleep Hours, & Final Exam Scores



Motivation, Sleep Hours, & Final Exam Scores





GUIDING QUESTIONS

What makes an “ideal” student? +


What can our criteria be for that - ideally an even mix of traits like study habits and access to resources etc.?





THE PLAN



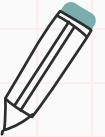
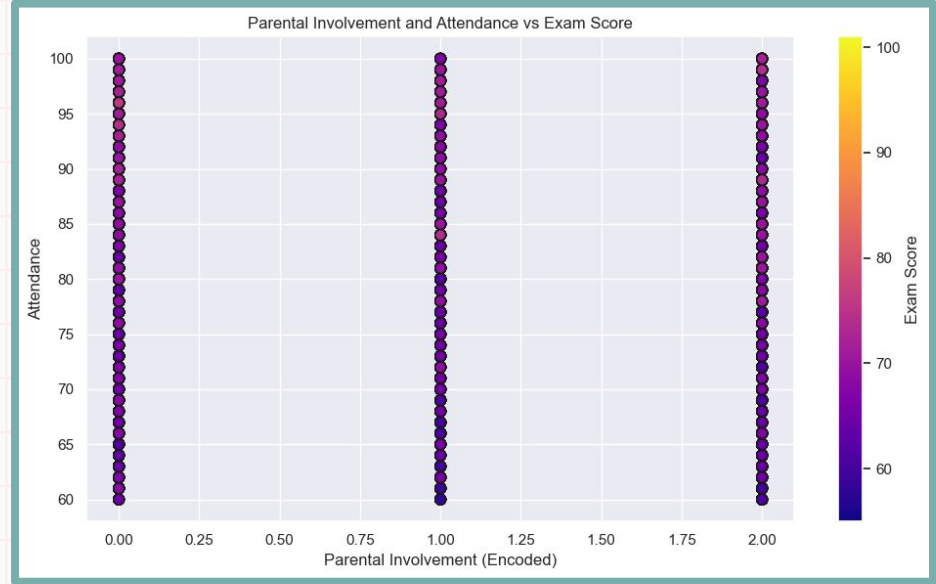
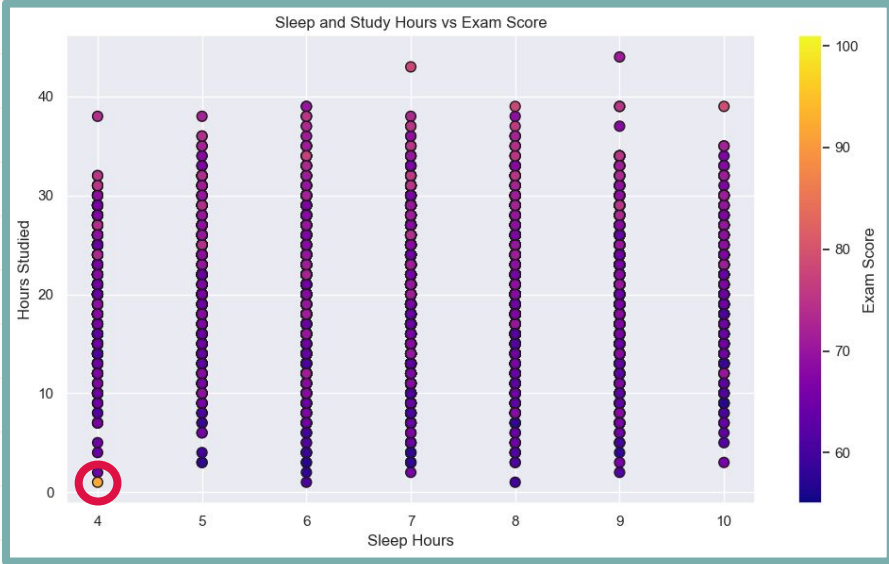
- 
- If we want to look at individual students, then we need to use a **scatter plot** to find our one “ideal” student. +
 - Try to put together variables that make sense together:
 - Sleep_hours and Study hours
 - Parental Involvement and Attendance
 - Physical Activity and Extracurricular Activities
 - Motivation Level and Study Hours
 - Tutoring Sessions and Previous Score



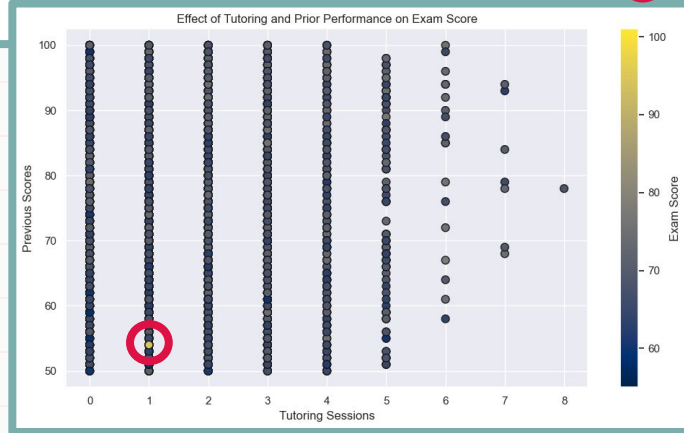
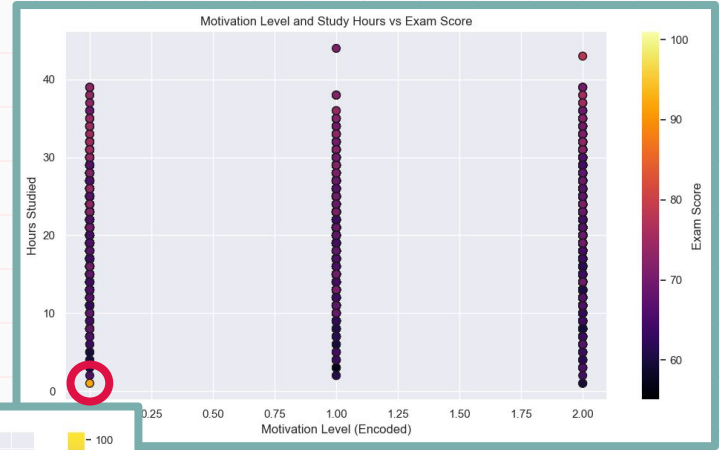
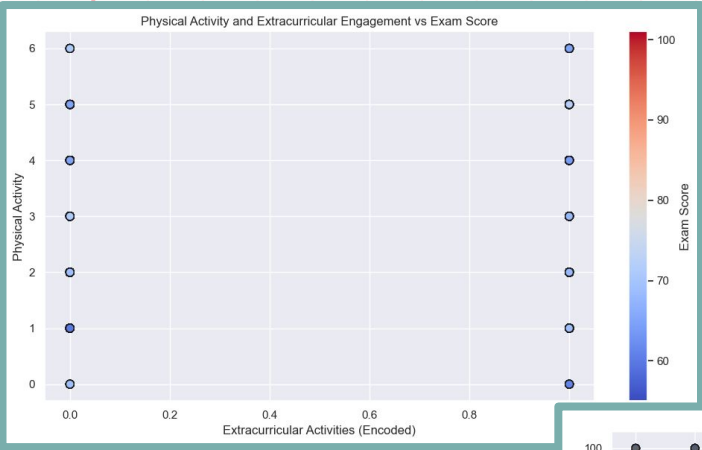
ALL IN RELATION TO EXAM SCORE



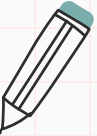
The "Ideal" Student Attempts



The "Ideal" Student Attempts



+ %



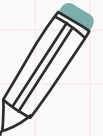
The "Ideal" Student Attempts

- Sleep hours
- Participation and Attendance
- Physical Activity and Extracurricular Activities
- Motivation Level and Study Habits
- Tutoring and Test Score

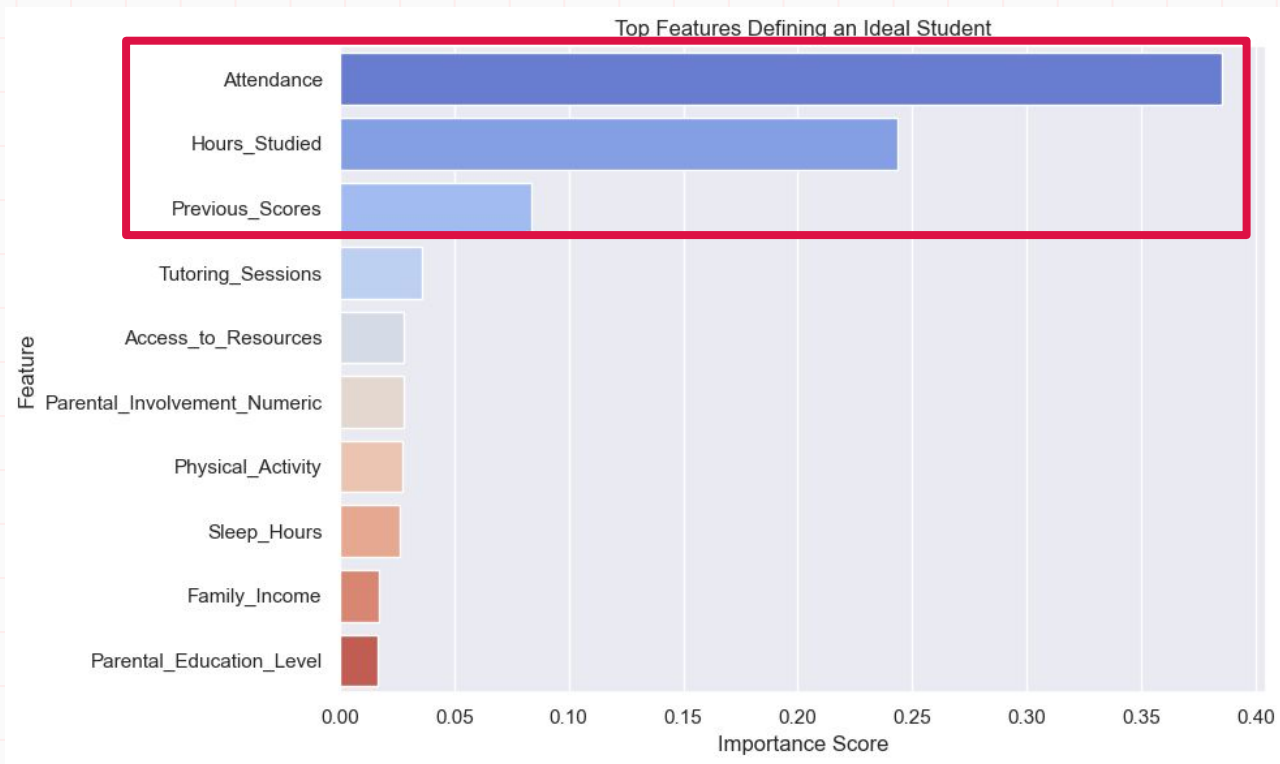
NOW WHAT?



+ %



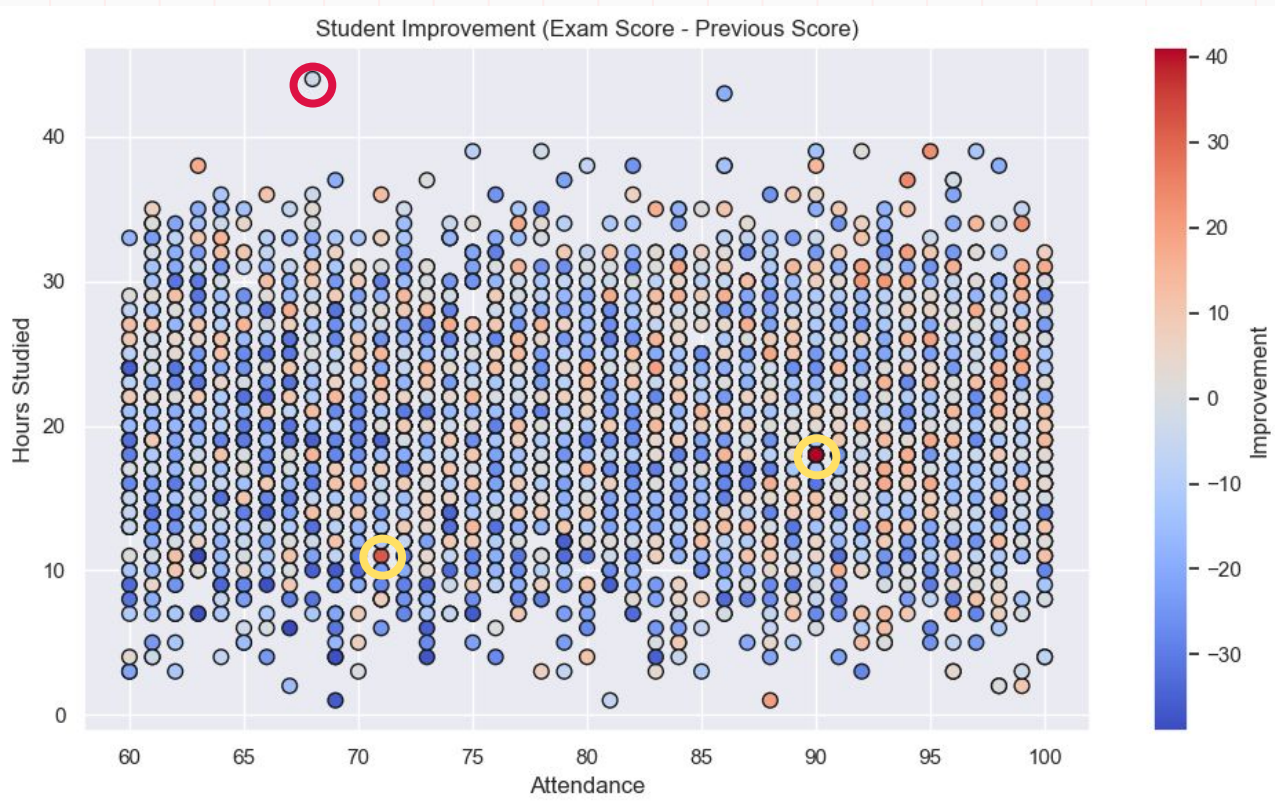
The "Ideal" Student Factors



+ %



The "Ideal" Student



+ %



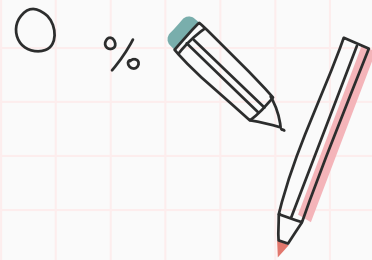
Future Work

It would be really interesting to apply the methods used within this dataset to a real world scenario. Perhaps collecting data about Willamette students and then implementing the methods of analysis used here could be interesting!



+ %





Conclusion

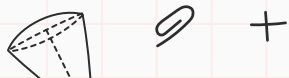
No Ideal Student

“Ideal” is a hard term to use. It is too difficult to call a student ideal or non-ideal with the data at hand. Sometimes, data is simply not representative of students on the whole.

Try Your Best!

Results vary so much from student to student that it is most important to try *your* best. Doing the maximum amount of studying does not guarantee the highest score!

Data Collection is IMPORTANT!



A red vertical line runs down the left side of the slide. Two small red circles are positioned on this line, one near the top and one near the bottom.

Thank You!

