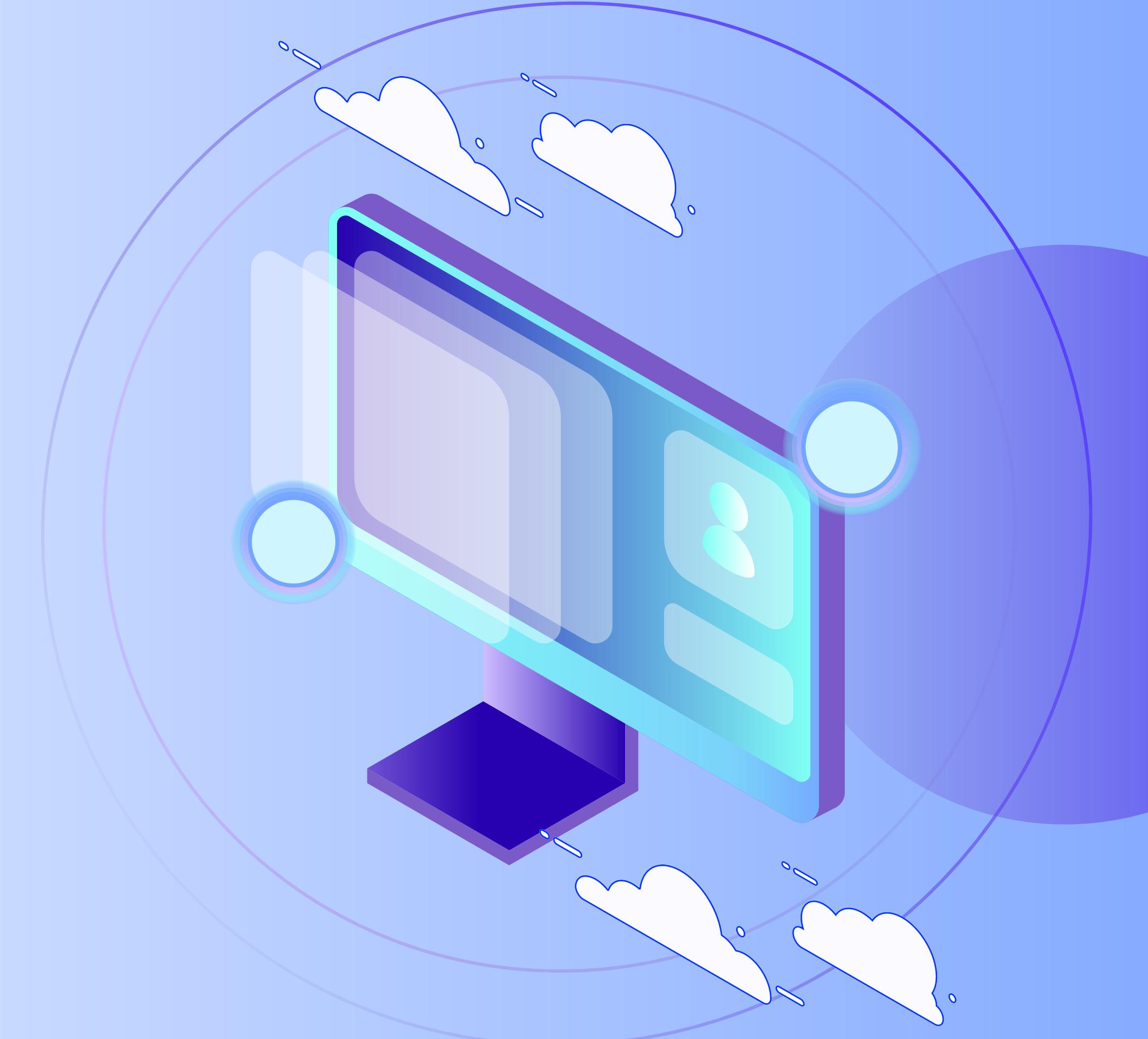


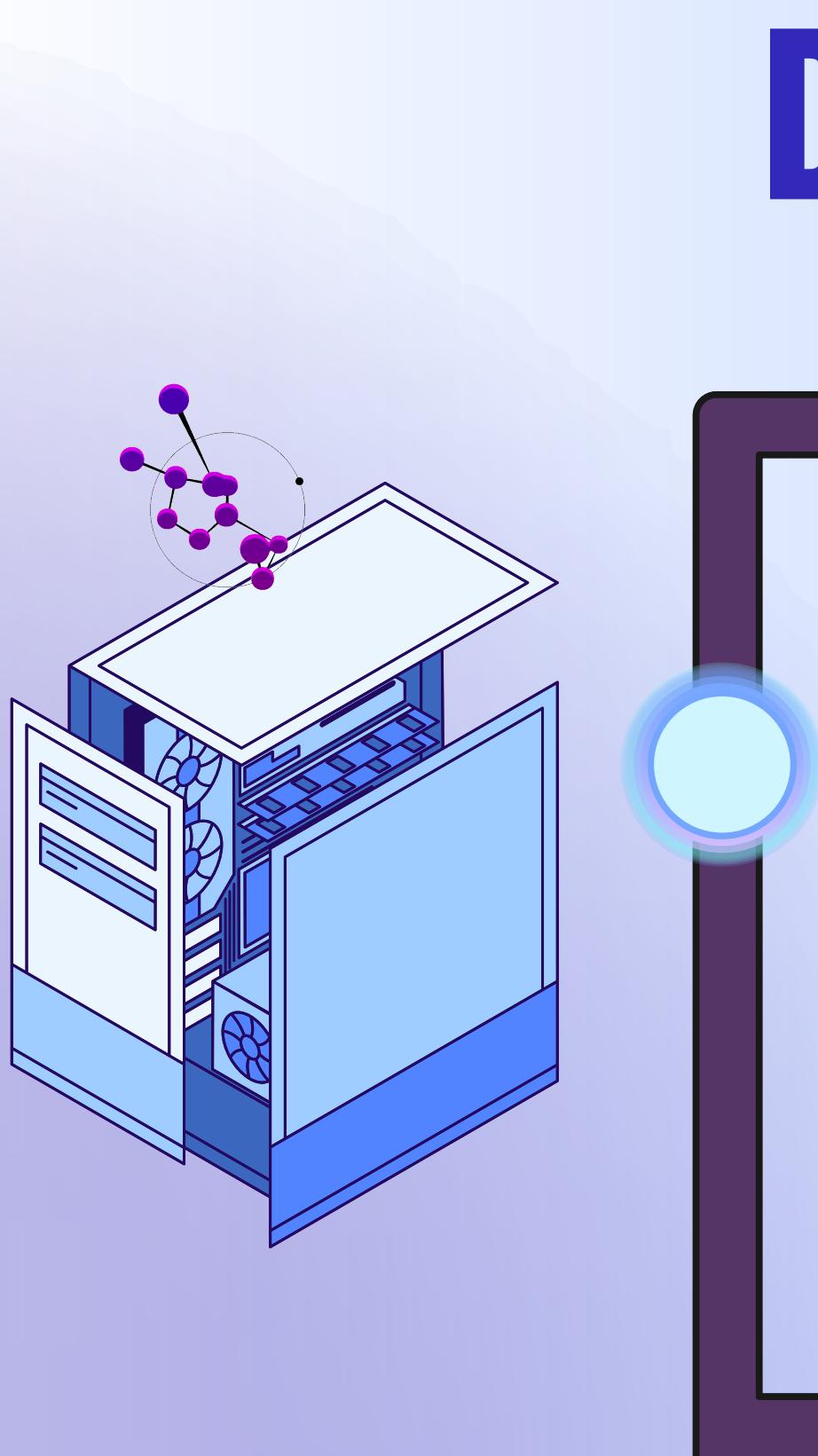


ITS

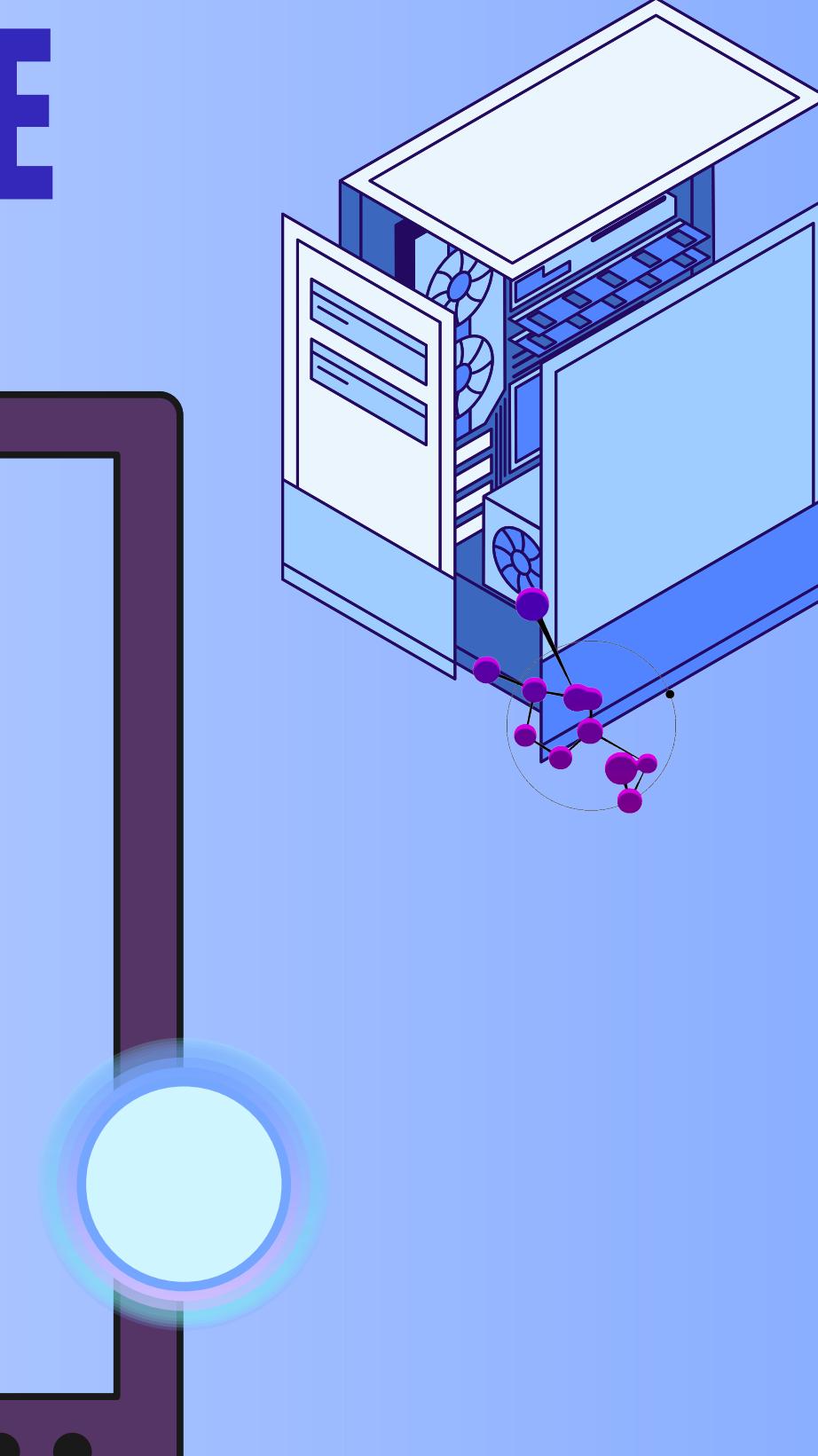
DATA SCIENCE



DEFINITION AND SCOPE



In today's fast-paced world, students are facing more pressure than ever. This project aims to answer a key question: What role does mental health play in sustaining academic success over time?



Imagine being able to predict not just a student's academic success, but also their future potential based not only on their study habits, but also on something equally crucial: their mental health.

KEY CONCEPTS



- **Academic Success Prediction:** Using performance-based features like study time, absences, and extracurricular activities to predict GPA.
- **Mental Health Insight:** Leveraging features like sleep quality, physical activity, and stress levels to evaluate a student's mental health.
- **Integrated Forecast:** Combining these two aspects to understand if a student with high academic potential is at risk of a future decline in performance due to mental health issues.



STUDENT PERF

2 DATASETS



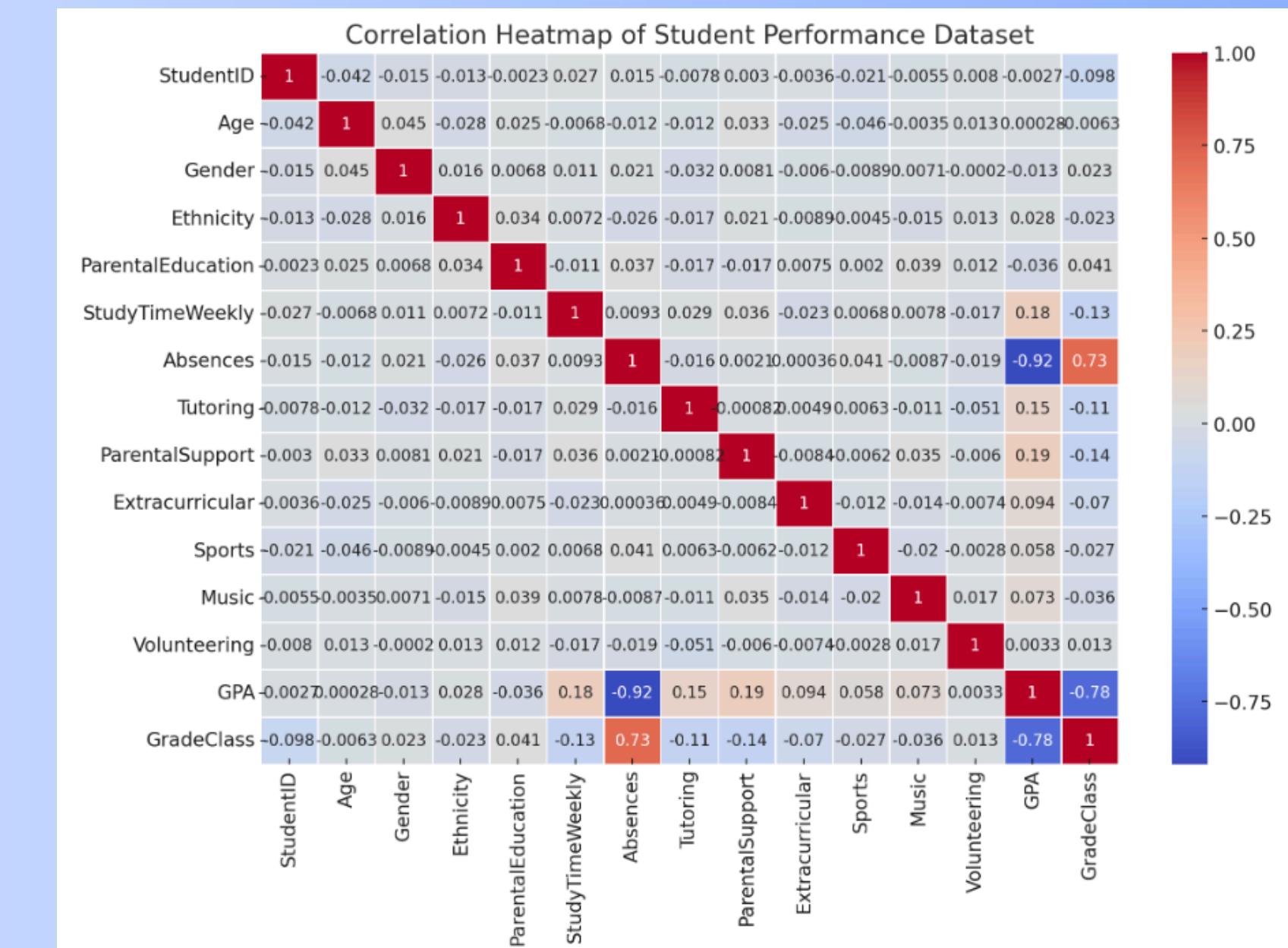
SLEEP HEALTH

- + Good reviews about the quality of the dataset.
- + No missing values and no outliers
- + Easy features to measure to predict the student performance
- + Many rows

- + Good reviews about the quality of the dataset.
- + No missing values and only 14 outliers
- + Easy feature to measure to predict the mental health
- Datas collected on adult and not on students
- The stress level is personal (not measurable)

FIRST STUDENT PER

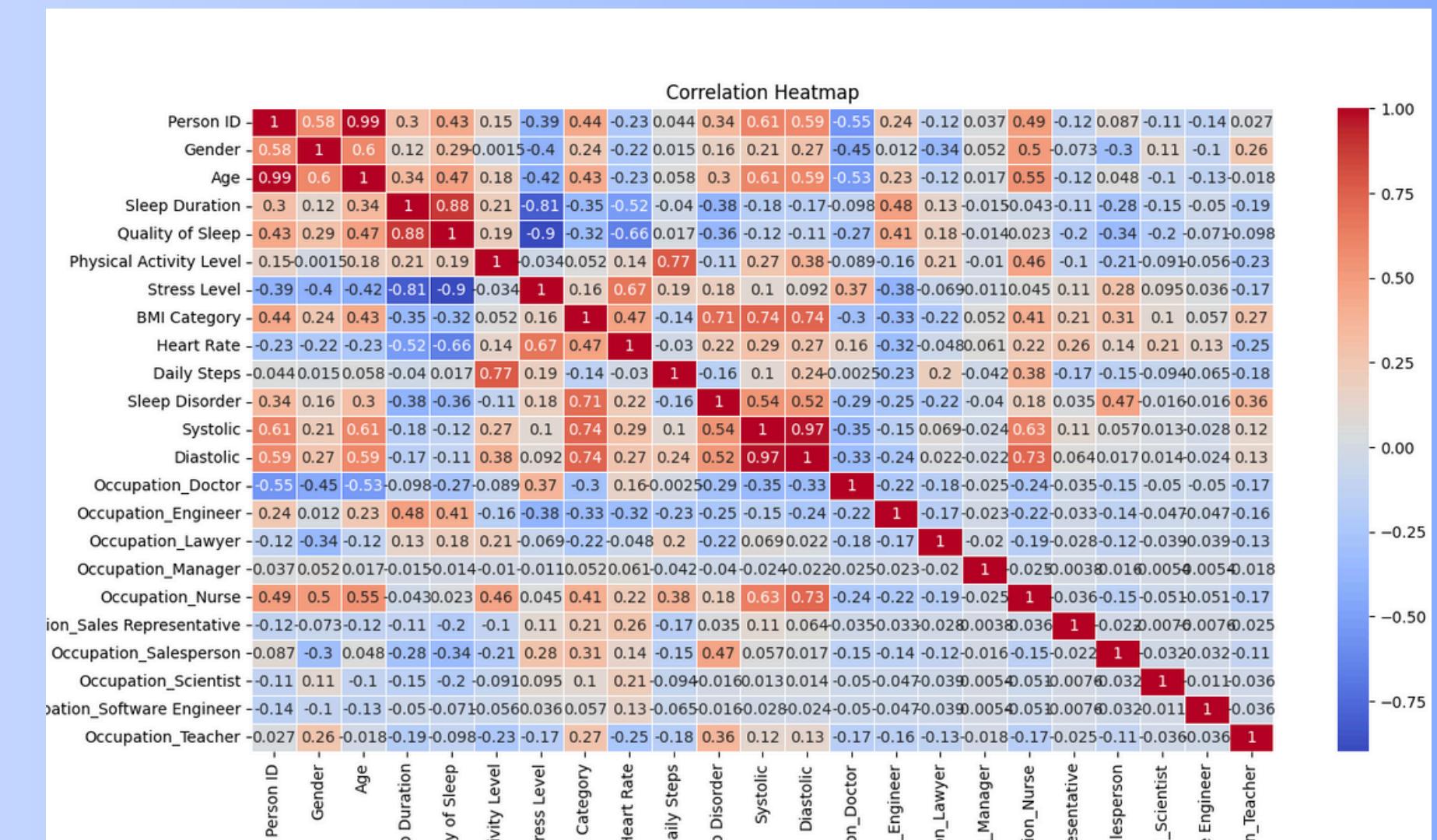
- + Good reviews about the quality of the dataset.
- + No missing values and no outliers
- + Easy features to measure to predict the student performance
- + Many rows



SECOND

SLEEP HEALTH

- + Good reviews about the quality of the dataset.
- + No missing values and only 14 outliers
- + Easy feature to measure to predict the mental health
- Datas collected on adult and not on students
- The stress level is personal (not measurable)



DATA PREPARATION



STUDENT PERFORMANCE : All the features were numerical. No missing values and no outliers so no particular preparation. Extract the best combination of feature to best predict the GPA thanks a Python code

7 features selected : 'StudentID', 'StudyTimeWeekly', 'Absences', 'Tutoring',
'ParentalSupport', 'Extracurricular', 'Sports'

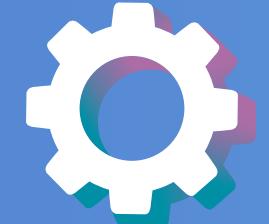


DATA PREPARATION

SLEEP HEALTH : No missing values and no outliers

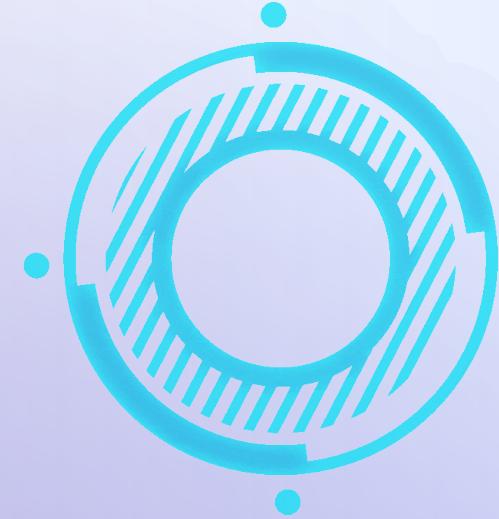
- Gender: Transformed from "Male/Female" into a binary attribute (0 for Male, 1 for Female).
- Occupation: One-Hot Encoded into separate binary columns, avoiding any ordinal relationships (e.g., "Occupation_Software Engineer," "Occupation_Doctor").
- BMI Category: Ordinarily encoded as a numerical variable (0 for Underweight, 1 for Normal, 2 for Overweight, 3 for Obese).
- Blood Pressure: Split into two columns, "Systolic" and "Diastolic," representing systolic and diastolic pressure respectively.
- Sleep Disorder: Ordinarily encoded (0 for None, 1 for Sleep Apnea, 2 for Insomnia).

7 features selected excluding Age and occupation : 'Gender', 'Sleep Duration', 'Quality of Sleep', 'Physical Activity Level', 'Heart Rate', 'Daily Steps', 'Systolic'





Liceria Tech



SELECT MODEL

Run Python code for choosing the best regression model for each dataset.

For the SLEEP HEALTH : Gradient Boosting 0.0180
0.1340 0.0372 0.9943 (MSE RMSE MAE R²)

STUDENT PERF : Linear Regression 0.0407 0.2019
0.1576 0.9507 (MSE RMSE MAE R²)

MSE (Mean Squared Error): Measures the average squared error.

RMSE (Root Mean Squared Error): Easier to interpret since it's in the same units as the target variable.

MAE (Mean Absolute Error): Shows the average absolute error.

R² (Coefficient of Determination): Indicates how much of the variance in the dependent variable is explained by the model.



THANK YOU!