

Activity 1: Dataset Design & Student Profiling

1. students.csv (50 Students)

The students.csv dataset contains 50 synthetic student records capturing academic performance, attendance, wellness, productivity, engagement, and career readiness metrics. The data is designed to be realistic and suitable for mentoring analysis and ML benchmarking.

2. Data Dictionary

Column Name	Description	Range / Scale
student_id	Unique student identifier	Categorical
age	Student age	18–25
program	Degree program	B.Tech, MBA
semester	Current semester	1–8
gpa	Academic performance	0–10
attendance	Attendance percentage	0–100
assignments_completion	Assignments completed (%)	0–100
stress_level	Self-reported stress level	1–10
sleep_hours	Average sleep per night	0–10
mental_wellbeing	Psychological wellbeing score	1–10
productivity_score	Time management & productivity	1–10
distractions	Level of distractions	1–10
career_clarity	Career goal clarity	1–10
skill_readiness	Job skill preparedness	1–10
engagement_score	Platform engagement score	0–100

3. How the Data Reflects Real Student Behavior

The dataset captures realistic correlations observed in academic environments. Students with high stress levels often show reduced productivity, poor sleep, and lower mental wellbeing. Some students demonstrate strong academic performance but lack career clarity, reflecting uncertainty despite good grades. Others exhibit low GPA but high engagement, indicating motivation to improve and actively seek mentoring support. Engagement scores vary independently to model proactive, disengaged, and struggling students realistically.