DATA DESCRIPTION – University Student Performance & Activities Dataset

Link to the dataset.

This dataset contains **1,000 student records** collected from different departments, regions, and academic years. It is intentionally **unclean** to provide hands-on learning experience in **Data Cleaning, Transformation, and Visualization** in Power BI.

Dataset Objective

The objective of this dataset is to analyze **student performance, attendance, fees, club participation, and placement outcomes** across regions and academic years.
Students will learn to clean, transform, and build dashboards that uncover insights about academic and extracurricular performance trends.

Column-Level Description

Column Name	Description	Issues Present (for Cleaning Practice)
Student_ID	Unique identifier assigned to each student	Contains **duplicates** in some rows
Name	Student's name	Extra initials added randomly (e.g., "Rahul A"), inconsistent formatting
Gender	Student gender (Male / Female)	Has values like "M", "F", "male", "Female", etc — **inconsistent text**
Age	Age of the student	Contains missing values and an **extreme outlier** (value "200")
Department	Academic department enrolled in	Inconsistent values: "CSE", "cse", "Ece", "Civil", "CIVIL", etc
Region	Region from which the student belongs	Inconsistent values: "North", "nORTH", "south", "South"
Admission_Year	Academic year of admission	Mixed formats: 2019, "2022", "20-21"

Semester_GPA	GPA scored in the latest semester	Missing values, float numbers formatted as strings sometimes
Attendance_%	Student's attendance percentage	Has values as numbers, strings like "85%", "N/A", and blanks
Clubs	Club participation (Music, Sports, Drama, Coding, etc)	Case mismatch ("Music", "music"), "None", null values
Hostel	Whether the student is staying in hostel	Values include "Yes", "No", "Y", "N", null — needs standardization
Fee_Paid	Fee paid by the student during the year	Mixed data types: numbers, strings like "50k", "50000", blanks
Fee_Due	Remaining fee after payment	Values such as numbers, blanks, and "NA" as string
Placement_Status	Placement status at the end of final year	"Placed", "placed", "Not placed", "PENDING", missing
Internship_Company	Internship company student worked with	Contains "None" as text, null values, mixed company names
Exam_Date	Date of semester examinations	Mixed date formats: YYYY-MM-DD, DD/MM/YYYY, invalid dates
Remarks	Comments on student's performance	Contains "Good", "avg", "NA", blank, inconsistent spellings