**📌 Project Statement**

**Title:**  
“Data Cleaning and Quality Assurance for Student Enrollment Records”

**Problem Statement:**  
The dataset of student records has been compromised due to errors introduced during manual data entry by a junior technician. Instead of structured values in their respective fields, multiple attributes have been merged into a single column. Furthermore, inconsistencies, missing values, and formatting errors across key identifiers (Student ID, Names, Dates, and Payments) make the dataset unreliable for analysis or reporting.

**Objective:**  
To clean, restructure, and validate the student dataset to restore data integrity, ensure consistent formatting, and create a reliable record system for decision-making.

**📖 Project Description**

This project addresses the issue of corrupted student enrollment records caused by improper data entry. The primary role of the analyst is to identify, correct, and document all critical data quality problems.

**Scope of Work:**

1. **Diagnosis of Data Quality Issues:**
   * Identify misplaced, missing, and duplicated data across Student\_ID, First\_Name, Last\_Name, Age, Gender, Course, Enrollment\_Date, and Total\_Payments.
   * Assess structural errors, including concatenated fields and inconsistent delimiters.
2. **Data Cleaning & Reconstruction:**
   * Extract and correctly assign values from Student\_ID where multiple attributes are embedded.
   * Standardize date formats.
   * Normalize age and gender fields for consistency.
   * Clean financial data by removing currency symbols and ensuring numeric formatting.
3. **Handling Missing & Incomplete Records:**
   * Separate cases where records are unrecoverable (e.g., missing both ID and name).
   * Recommend further investigation or escalation for incomplete entries (to higher-ups).
4. **Deliverables:**
   * A cleaned dataset with properly structured columns and consistent formatting.
   * A data quality report highlighting:
     + Percentage of missing values.
     + Irrecoverable errors.
     + Recommendations for improving future data capture processes.

**Expected Outcome:**

A reliable, analysis-ready dataset that reflects accurate student records, supported by clear documentation of all issues encountered and decisions made during the cleaning process.