# Student Data Analysis – Final Report

**1. Data Loading**

We started by loading the student marks dataset. The file was successfully read, and the system automatically understood the types of data (like numbers and text).  
👉 *Tip:* In real systems, it’s better to define the data types manually to avoid mistakes later.

**2. Section Data Quality Check**

While checking the **SECTION** column, we found:

* 41 students (out of 480) had missing section names.
* A few spelling mistakes, like “GAMA” instead of “GAMMA” and “SGMA” instead of “SIGMA.”

👉 *Recommendation:* Use proper validation rules or dropdown menus when entering section names to prevent such errors.

**3. Section Data Cleaning**

We corrected the spelling mistakes and filled missing section names with **ZETA**. Now, every student belongs to a valid section.  
👉 This step ensures fairness when comparing performance between sections.

**4. Subject Marks Cleaning**

Several subjects (DV, M-II, PP, BEEE, FL, FIMS) had wrong entries like **“A”**, **“AB”**, **“MP”**, **“O”**, or **“II”**. These are likely shorthand for things like **Absent** or **Malpractice**.  
👉 *Suggestion:* Keep marks and attendance/misconduct information in **separate columns** to avoid confusion.

After cleaning, all marks were converted into proper numbers.

**5. Total and Percentage Calculation**

* Each student’s total marks (out of 120) were calculated.
* Average score: **79.41**
* Range: **0 – 120**
* Converted totals into percentages for easy comparison.

👉 This gives a clear, uniform view of student performance.

**6. Grade Assignment**

Grades were given based on percentage:

* **A (80%+)** → 140 students
* **B (70–79%)** → 98 students
* **C (60–69%)** → 75 students
* **D (50–59%)** → 71 students
* **E (40–49%)** → 46 students
* **F (<40%)** → 50 students

👉 Overall, most students performed fairly well, but there are still a significant number failing.

**7. Skill Assessments**

* **Python Skills (PP subject):**
  + Excellent: 105
  + Average: 120
  + Below Average: 255
* **Data Analytics Skills (DV subject):**
  + Excellent: 91
  + Average: 192
  + Below Average: 197

👉 A large number of students need extra support in programming and analytics.

**8. Section-wise Performance**

* **Best section:** ALPHA (74.04%)
* **Weakest section:** OMEGA (61.29%)

👉 Possible reasons: differences in teachers, class sizes, or learning resources. Further investigation needed.

**9. Visualizations**

We created **9 different graphs and charts** to show:

* Section comparisons
* Grade distributions
* Skill levels
* Subject correlations
* Overall performance trends

👉 These visuals can help teachers and management quickly spot issues.

**10. Final Outcomes**

* Data was cleaned, standardized, and analyzed.
* We now have a **complete picture** of how 480 students are performing across 6 subjects and 8 sections.
* The process highlights both strengths (strong performers in some sections) and weaknesses (students struggling with Python and Analytics).