



Certified Data Analysts

Assignment 3

Student Performance Analyzer using Python

Scenario Description

You have joined a small EdTech startup as a Python intern. Your first project is to develop a "Student Performance Analyzer" tool that helps academic coordinators manage and analyze basic student data such as name, roll number, marks, and grades.

The coordinator has shared a raw CSV file (*students_raw.csv*) containing student names, roll numbers, and marks in three subjects. The data contains inconsistencies and missing values. Your task is to write a Python program that reads the data, performs necessary calculations, cleans the data, and visualizes useful insights.

Section 1: Core Python Programming

You need to simulate how the tool will behave for one student entry using core Python logic.

Tasks:

1. Write a Python script to:
 - Take student name, roll number, and mark in 3 subjects from keyboard input.
 - Calculate:
 - Total Marks
 - Percentage
 - Grade based on
 - A: $\geq 90\%$
 - B: 80–89%
 - C: 70–79%
 - D: 60–69%
 - F: $< 60\%$
2. Use appropriate data types, variables, and comments.
3. Display the result in a structured format using print () with formatting options (like sep, end).
4. Add conditional logic to:
 - Print a congratulatory message if the grade is A or B.
 - Warn the user if the grade is F.

Section 2: NumPy and Pandas Data Handling

The coordinator now asks you to analyze the full student dataset (`students_raw.csv`), which includes 50+ records. Some marks are missing or have invalid values like "N/A".

Tasks:

1. Using NumPy:

- Load student marks into NumPy arrays.
- Compute array-wise:
 - Mean marks per subject
 - Standard deviation
 - Maximum and minimum marks.
- Filter students with total marks above a threshold (e.g., 250/300).

2. Using Pandas:

- Load the CSV using `pandas.read_csv()`.
- Clean the data:
 - Replace missing values with 0 or appropriate estimates.
 - Convert data types as needed.
- Add computed columns:
 - Total
 - Percentage
 - Grade (use conditions)
- Save the cleaned DataFrame to a new CSV file *students_cleaned.csv*.

Section 3: Data Visualization with Matplotlib

Your manager wants you to create charts that summarize student performance for a report to the principal.

Task:

1. Line Plot:

- Show the trend of average marks in each subject

2. Bar Chart:

- Display the number of students per grade (A, B, C, D, F)

3. Scatter Plot:

- Plot percentage vs. total marks with color coding for grades

4. Chart Customization:

- Add appropriate titles, labels, legends.
- Save each plot as an image (.png) file.

The submission requirements:

1. Python Script / Jupyter Notebook containing:

- Section 1 - Logic-based scripts
- Section 2 - NumPy & Pandas code
- Section 3 - Matplotlib visualizations

2. Cleaned CSV file:

- students_cleaned.csv

3. Report:

- **Platform:** Google Drive
- **Format:** word and PDF
- **Page Limit:** Maximum 10 pages
- **Content:** Include your Code snippets, Output screenshots, Charts, Observations and key insights.