ZAIO Assignment 1: Student Habits and Academic Performance Analysis

Total Marks: 50

Assignment Overview

This assignment focuses on analyzing the relationship between student lifestyle habits and academic performance using object-oriented programming (OOP) principles in Python. Students will demonstrate proficiency in data cleaning, statistical analysis, visualization, and modular code design.

Learning Objectives

1. Implement OOP concepts (classes, inheritance)
2. Design reusable code modules
3. Perform exploratory data analysis (EDA)
4. Create insightful visualizations
5. Apply statistical methods to derive meaningful insights

Dataset Description

The dataset contains **1,000 student records** with these key features:

- study_time: Daily study hours (float)
- sleep: Average nightly sleep (float)
- social_media: Daily social media usage (float)
- diet_quality: Self-reported diet rating (1-5)
- mental_health: Mental health score (1-10)
- final_score: Final exam percentage (float)

Assignment Tasks

Phase 1: Data Loading & Preprocessing (10 Marks)

1. Download the dataset using the provided Kaggle API code
2. Create a DataLoader class that handles file I/O and exceptions
3. Implement a DataCleaner class to:
    - Check for missing values
    - Remove duplicates
    - Validate data ranges

Phase 2: Statistical Analysis (10 Marks)

1. Develop a StudentAnalyzer class with methods to:
    - Calculate mean/median study time by mental health tier
    - Identify correlation between sleep and exam scores
    - Detect outliers in social media usage

Phase 3: Visualization (10 Marks)

1. Build a VisualizationEngine class that generates:
    - Histogram of study time distribution
    - Scatter plot of sleep vs. final scores
    - Box plots of scores by diet quality

Phase 4: Predictive Modeling (10 Marks)

1. Create a ScorePredictor class that:
    - Uses linear regression to predict scores from habits
    - Implements error handling for invalid inputs
    - Saves models using pickle

Phase 5: Report Generation (10 Marks)

1. Exported results into a structured **Markdown/PDF** report.
2. Included key visualizations and statistical summaries.
3. Ensured proper error handling for file permissions.

Submission Requirements

1. **Python Module** (student_analysis.py) containing all classes
2. **Jupyter Notebook** (analysis.ipynb) showing execution flow
3. **PDF Report** (max 5 pages) with findings
4. **Model Files** (if any) in /models folder

Performance Level Definitions

- **Exceptional:** Exceeds all requirements with professional-quality work
- **Very Good:** Meets all requirements with minor improvements needed
- **Good:** Addresses core requirements but lacks depth
- **Satisfactory:** Partial completion with significant gaps
- **Poor:** Fails to meet basic requirements
- **Very Poor:** No meaningful attempt made

Rubric (50 Marks Total)

Criteria	Exceptional (10 marks)	Good (8 marks)	Satisfactory (5 marks)	Poor (3 marks)	Very Poor (0 marks)
Program Execution (10)	All code runs without errors. Classes interact perfectly.	Minor runtime errors (1-2). Most functionality works.	Several runtime errors but core functions work.	Many errors but some output generated.	Code fails to execute.
Data Handling (10)	Impeccable data loading/cleaning. All edge cases handled.	Robust data processing with minor issues.	Basic cleaning implemented but some gaps.	Partial data cleaning with obvious flaws.	No proper data handling.
OOP Implementation (10)	Excellent class design with inheritance. Perfect use of methods.	Good OOP structure with minor design flaws.	Basic OOP implementation but limited features.	Classes exist but poorly designed.	No proper OOP implementation.
Analysis & Visualization (10)	Professional, insightful analysis and visualizations.	Good analysis with minor visualization issues.	Basic analysis with limited insights.	Superficial analysis with poor visuals.	No analysis/visualizations.
Report Quality (10)	Well-structured, academic writing with clear insights.	Good report with minor omissions.	Basic report missing some elements.	Disorganized but contains key points.	No report submitted.