Name: ABDULLAH
RollNO: SU92-BSAIM-F24-002
Section: BSAI-3A
Subject: AI-Theory

Student Performance Analyzer - Documentation

Step 1: Class Initialization

The class 'StudentPerformanceAnalyzer' is initialized with a file path as an argument. It stores this path in the 'file_path' variable and initializes 'df' (dataframe) and 'model' as None.

Step 2: Data Loading

The 'load_data()' method reads a CSV file using pandas (pd.read_csv) and stores it in the dataframe 'df'. It prints the first few rows to confirm successful data loading.

Step 3: Data Understanding

The 'understand_data()' method provides insights into the dataset, including:
- Dataset info (data types, memory usage)
- Descriptive statistics (mean, median, std, etc.)
- Missing values count.

Step 4: Data Preprocessing

The 'preprocess_data()' method prepares the data for analysis by:
- Converting categorical variables into numeric using one-hot encoding (pd.get_dummies).
- Dropping missing values using dropna().

Step 5: Univariate Analysis

The 'univariate_analysis()' method explores single variable distributions using histograms for all columns. It helps understand data spread, skewness, and outliers.

Step 6: Bivariate Analysis

The 'bivariate_analysis()' method explores relationships between variables using a correlation heatmap. It uses seaborn's heatmap to visualize correlations among numeric columns.

Step 7: Data Splitting

The 'split_data()' method separates the dataset into features (X) and target (y), where 'math score' is the target variable. Then it splits the data into training and testing sets using train_test_split from sklearn.

Step 8: Exploratory Data Analysis (EDA)

The 'exploratory_data_analysis()' method performs deeper visual analysis including:
1. Distribution plot of target variable ('math score').

2. Boxplots for numerical features to detect outliers.
3. Countplots for categorical features to understand category frequencies.

Step 9: Execution

Finally, the script creates an instance of 'StudentPerformanceAnalyzer' using the CSV file 'StudentsPerformance.csv' and executes all methods in sequence to perform complete data analysis.