

Task 10: Python EDA + Outlier Detection

Dataset: Students Performance

Exploratory Data Analysis (EDA)

The dataset contains 1000 student records with 8 features, including demographic attributes and exam scores in Math, Reading, and Writing.

Dataset Overview

Rows: 1000

Columns: 8

Data Types

Numerical Columns: Math Score, Reading Score, Writing Score

Categorical Columns: Gender, Race/Ethnicity, Parental Level of Education, Lunch, Test Preparation Course

Missing Values

No missing values found in the dataset (0%).

Descriptive Statistics

Math Score → Mean: 66.1 | Median: 66 | Min: 0 | Max: 100

Reading Score → Mean: 69.2 | Median: 70 | Min: 17 | Max: 100

Writing Score → Mean: 68.1 | Median: 69 | Min: 10 | Max: 100

Distribution & Correlation Insights

Most students scored between 60 and 80. Reading and Writing scores show strong positive correlation, while Math shows moderate correlation with both.

Outlier Detection (IQR Method)

Math Score: Q1 = 57, Q3 = 77, IQR = 20, Lower Bound = 27

Outliers Detected: 18 students

Outlier Handling

Outliers were retained and flagged instead of removed, as low scores represent genuine student performance and are important for educational analysis.

Final Outcome

Successfully performed EDA and outlier detection, gaining meaningful insights into student academic performance.