Bootcamp Project: Student Performance Analysis

Problem Statement

You are provided with a dataset containing exam scores and demographic information of students. Your task is to perform detailed exploratory data analysis (EDA) to uncover insights about what factors affect student performance. Additionally, create a Power BI Dashboard to present your findings visually.

Dataset Details

Dataset Link: https://www.kaggle.com/datasets/spscientist/students-performance-in-exams

Columns Included:

- gender
- race/ethnicity
- parental level of education
- lunch
- test preparation course
- math score
- reading score
- writing score

Key Objectives

- Perform comprehensive EDA using Pandas.
- Answer tricky and meaningful business questions.
- Create a clean and insightful Power BI Dashboard summarizing your results.

EDA Questions (Make It Tricky)

1. Which parental education level is linked with the highest average math score?
2. Is there a significant score difference between males and females across all subjects?
3. How much does completing the test preparation course improve performance in each subject?
4. Which combination of gender, lunch type, and test preparation status produces the top 10% of scores?
5. Does lunch type have a uniform impact across all race/ethnicity groups, or does its effect vary?
6. What is the correlation between reading and writing scores? Is it stronger than math and writing?
7. Identify the top 5% performing students and analyze their demographic profiles. What patterns emerge?
8. Can we cluster students into performance categories (e.g., low, medium, high performers) using just

Pandas logic? If yes, how?

Power BI Dashboard Requirements

- At least 3 slicers/filters (Example: Gender, Lunch, Test Preparation Course).
- Visualizations:
  - Bar Charts
  - Box Plots (if possible)
  - KPI Cards showing average scores
  - Heatmap or Correlation Matrix if feasible
- Dashboard must be clean and easy to understand.

Submission Instructions

- Submit the following via Google Classroom:
  1. Jupyter Notebook or Colab File with Python code (cleaned, commented, properly structured).
  2. Power BI File (.pbix) with the final dashboard.

- Make sure:
  - Files are properly named using this format:

    YourName_StudentPerformanceAnalysis_Project3
  - Submit within the given deadline.
  - Late submissions may result in deduction of marks.