Coding Challenge: Player Performance Analytics

Scenario

You are a **Data Analyst** working for a sports club (IPL/Football). The management wants to understand **player performance trends** across multiple seasons in order to make informed decisions about team selection, contracts, and strategy. Your task is to analyze, clean, and visualize player statistics.

Note: The dataset fields are given below you need to generate the dataset with the fields give.

Dataset (Sample Columns)

- Player_Name
- Team
- Season (Year)
- Matches_Played
- Runs (Cricket) / Goals (Football)
- Strike_Rate (Cricket) / Shot_Accuracy (Football)
- Wickets (optional, cricket only)
- Assists (optional, football only)
- Rating (performance rating given by coaches, 1–10 scale)

Tasks

1. Data Cleaning & Preparation

- Handle **missing or inconsistent values** (e.g., empty goals, runs, ratings).
- Remove duplicate records.
- Standardize team and player names.
- Ensure correct **data types** (numeric vs categorical).

2. Exploratory Data Analysis (EDA)

- Find the **Top 5 performers each season** based on performance metrics (e.g., Runs × Strike Rate or Goals + Assists).

- Compare **player performance across seasons** → Identify **improving, declining, and inconsistent players**.
- **Team-wise contribution** → Which team had the most balanced performers vs most dependent on a single star?
- **Consistency analysis** → Which players have the least variation in performance across seasons?

3. Advanced Insights

- Create a **custom performance score** (weighted formula) to rank players. Example:
- Performance_Score = 0.4*Runs/Goals + 0.3*StrikeRate/Accuracy + 0.2*Assists + 0.1*Rating
- Identify **underrated players** (high performance score but low recognition in terms of media ratings).
- Find **outliers** (e.g., a player with very high strike rate but very few matches).

4. Visualization Deliverables

- **Bar Chart** → Top 10 players by performance score.
- **Line Chart** → Trend of top players across seasons.
- **Heatmap** → Correlation between metrics (Runs, Strike Rate, Assists, Rating).
- **Stacked Bar / Pie Chart** → Team contribution comparison.
- **Boxplot** → Distribution of performance scores by team.

Deliverables

1. **Cleaned dataset (CSV/Excel)** after preprocessing.
2. **Jupyter Notebook/Excel file** with data cleaning, analysis, and visualizations.
3. **Final Report (Word/PDF/PowerPoint)** containing:
    - Top performers
    - Team-wise insights
    - Consistency analysis
    - Outliers & interesting findings