

ESPN Cricket Data Summary and Analysis

Objective: Data cleaning and running some basic analysis.

Introduction:

This report provides insights based on the analysis and cleaning of cricket data extracted from ESPN. The dataset includes detailed player statistics covering various metrics, such as matches played, innings, runs scored, and strike rates. The goal of this report is to clean the data, handle missing values, and derive key cricketing insights.

Data Cleaning:

1. **Loading the Dataset:** The dataset was loaded into a pandas DataFrame. Upon inspection, the dataset included several columns with missing or irrelevant values, particularly towards the end. These columns needed to be removed for better clarity.
2. **Dropping Unnecessary Columns:**
 - Three unnamed columns at the end of the dataset were dropped as they did not contain any meaningful data.
3. **Renaming Columns:** The following columns were renamed to make them more intuitive and self-explanatory:
 - Mat → Matches
 - NO → Not_outs

- HS → Highest_inns_score
- Ave → Batting_average
- BF → Balls_faced
- SR → Batting_strike_rate

4. **Handling Missing Data:**

- Missing values in some statistical columns such as Batting_average and Batting_strike_rate were filled with appropriate values like 0 or the mean, depending on the context.
- Empty or NaN rows were inspected and either removed or imputed with meaningful values.

Data Analysis:

1. **Descriptive Statistics:** The dataset contains numerous player statistics. A brief summary of key columns:

- Matches: Number of matches played.
- Inns: Number of innings.
- Runs: Total runs scored by the player.
- Batting_average: Player's average score per innings.
- Highest_inns_score: Player's highest score in an innings.

Summary statistics (such as mean, median, min, max) for each of these metrics were calculated to get an overview of the data.

2. **Country-Wise Maximum Innings:** By grouping the data by Country, we calculated the maximum number of innings played by a player from each country. This helps identify standout players from each country who have participated in the most number of innings.

3. **Top Performers:** Using `Highest_inns_score`, the players with the highest individual scores in a single innings were identified. This metric is crucial in highlighting the most outstanding individual performances in cricket history.
4. **Consistency Metrics:**
 - **Batting_average:** By examining players with the highest averages, we identified the most consistent batsmen across various countries.
 - **Batting_strike_rate:** The strike rate reveals players who score quickly, important in limited-overs formats.

Key Findings:

1. **Top Countries by Maximum Innings Played:** By country-wise grouping, it was found that several players from major cricketing nations like Australia, India, and England consistently appear at the top in terms of innings played.
2. **Legendary Performers:** Players like Don Bradman (Australia) were highlighted with extraordinary averages, showcasing their consistency over time. Bradman's batting average of 99.94 stands out as one of the highest in cricket history.
3. **High Scoring Matches:** Several players have exceptionally high scores in individual innings, with some players surpassing 300+ runs in a single match. These performances are key to match-winning contributions in both Test and One-Day International (ODI) formats.