

Report on : Cricket Data - Statistics & Analysis

By- Biswajeet Mohanty

Contents

1. Abstract	2
2. Data collection and preparation in Jupyter Notebooks	3
1. Code	3
2. Code Summary	5
3. Cricket Dashboard in PowerBI	5
1. Player Analysis Tab	6
1. Openers	6
2. Anchors	7
3. Finishers	8
4. Fast Bowlers	8
3.2. Playing 11	9
4. Conclusion	10

1. Abstract :

The Cricket Statistics Dashboard is a data visualization tool that provides an interactive platform to analyze the performance of cricket players. The dashboard contains two tabs: Player Analysis and Playing 11. The Player Analysis tab displays the statistics of four categories of players: Openers, Anchors, Fast Bowlers, and Finishers. The statistics are presented in the form of graphs and tables, making it easy to understand and compare the performance of players. The Playing 11 tab provides an interface to select the top 11 players that need to play in forthcoming matches. The dashboard displays the statistics of the selected players and provides an analysis of their performance. The datasets used in the dashboard were refined using Jupyter Notebooks. The data was cleaned, and missing values were imputed. The dashboard is an essential tool for cricket enthusiasts, coaches, and team selectors to make informed decisions based on player performance analysis. Overall, the Cricket Statistics Dashboard provides an effective way to analyze cricket player performance and make data-driven decisions.

2. Data collection and preparation in Jupyter Notebooks

1. Code

Importing necessary Libraries :

```
import pandas as pd
import json
```

Read match results Json file :

```
with
open(r"C:\Users\KIIT\Downloads\5_CricketT20Analytics\data_collection\t20_json_files\t20_wc_match_results.json") as f:
data = json.load(f)
```

Create dataframe from match summary data :

```
df_summary = pd.DataFrame(data[0]['matchSummary'])
```

Read bowling summary json file :

```
with
open(r"C:\Users\KIIT\Downloads\5_CricketT20Analytics\data_collection\t20_json_files\t20_wc_bowling_summary.json") as f:
data = json.load(f)
```

Extract bowling summary data from all items :

```
all_records = []
for item in data:
all_records.extend(item['bowlingSummary'])
```

Create dataframe from all bowling summary records :

```
df_bowling = pd.DataFrame(all_records)
```

Read batting summary json file :

```
with
open(r"C:\Users\KIIT\Downloads\5_CricketT20Analytics\data_collection\t20_json_files\t20_wc_batting_summary.json") as f:
data = json.load(f)
```

Extract batting summary data from all items :

```
all_records = []  
for item in data:  
    all_records.extend(item["battingSummary"])
```

Create dataframe from all batting summary records :

```
df_batting = pd.DataFrame(all_records)
```

Add a column to batting dataframe indicating if player was out or not out :

```
df_batting["out/not_out"] = df_batting.dismissal.apply(lambda x : "out" if len(x) > 0  
else "not_out" )
```

Drop the dismissal column from batting Dataframe :

```
df_batting.drop(columns=["dismissal"], inplace=True)
```

Read player info json file :

```
with  
open(r"C:\Users\KIIT\Downloads\5_CricketT20Analytics\data_collection\t20_json_files\t20_wc_player_info.json") as f:  
    data = json.load(f)
```

Create dataframe from player performance data :

```
df_player_performance = pd.DataFrame(data)
```

Create a dictionary of match ids for each match :

```
match_id_dict = { }  
for id, row in df_summary.iterrows():  
    key1 = row['team1'] + ' Vs ' + row['team2']  
    key2 = row['team2'] + ' Vs ' + row['team1']  
    match_id_dict[key1] = row["scorecard"]  
    match_id_dict[key2] = row["scorecard"]
```

Add a column to batting dataframe indicating the scoreboard for each match :

```
df_batting["scoreboard"] = df_batting["match"].map(match_id_dict)
```

2.2. Code Summary :

The given code is written in Python and it is used to refine the cricket data which is in JSON format. The Pandas library is used to perform various operations on the data.

The first block of code imports the necessary libraries and reads the match results data from a JSON file using the 'json.load' function. The data is then converted into a Pandas DataFrame called 'df_summary'. Similarly, the bowling and batting summary data is also read from JSON files and converted into Pandas DataFrames called 'df_bowling' and 'df_batting', respectively. The player information data is also read from a JSON file and converted into a Pandas DataFrame called 'df_player_performance'. Next, the 'out/not_out' column is created in 'df_batting' DataFrame based on the 'dismissal' column using the 'apply' and 'lambda' functions. The 'dismissal' column is then dropped from the DataFrame.

Finally, a dictionary called 'match_id_dict' is created using the 'df_summary' DataFrame where the keys are the team names and the values are the corresponding scorecard URLs. This dictionary is used to map the 'match' column in 'df_batting' DataFrame to the corresponding scorecard URL.

In summary, this code block reads the cricket data from JSON files, converts them into Pandas DataFrames, performs some data cleaning operations, and creates a dictionary for mapping scorecard URLs to the corresponding matches. This refined data is then used in the PowerBI dashboard for analyzing the cricket player statistics.

3. Cricket Dashboard in PowerBI :

The Cricket Statistics Dashboard is a project developed using PowerBI that allows users to analyze the performance of cricket players in various categories such as openers, anchors, fast bowlers, and finishers. The dashboard has two main tabs, Player Analysis and Playing 11.

3.1. PlayerAnalysis Tab :

The Player Analysis tab is a crucial feature of the Cricket Statistics Dashboard. It provides detailed statistics of four categories of players: Openers, Anchors, Fast Bowlers, and Finishers. The statistics are presented in the form of graphs and tables, making it easy to understand and compare the performance of players.



Figure 1 : Player analysis tab in PowerBI

3.1.1. Openers :

The Openers category in the Player Analysis tab displays the performance statistics of the opening batsmen. The statistics include runs scored, batting average, strike rate, balls faced, and boundary percentage. The line graphs show the batting average, ball faced, strike rate, and boundary percentage of the top 5 players over time. These graphs allow the users to analyze the consistency of the players over time and track their progress. The bubble graph is an excellent tool to compare the batting average and strike rate of the top 5 players.



Figure 2 : Openers tab in Players Analysis Tab

3.1.2. Anchors :

The Anchors category displays the performance statistics of the middle-order batsmen who hold the innings together. The statistics include runs scored, batting average, strike rate, balls faced, and boundary percentage. The line graphs show the batting average, ball faced, strike rate, and boundary percentage of the top 5 players over time. These graphs allow the users to analyze the consistency of the players over time and track their progress. The bubble graph is an excellent tool to compare the batting average and strike rate of the top 5 players.



Figure 3 : Anchors tab in Players Analysis Tab

3.1.3. Finishers :

The Finishers category displays the performance statistics of the batsmen who finish the innings strongly. The statistics include runs scored, batting average, strike rate, balls faced, and boundary percentage. The line graphs show the batting average, ball faced, strike rate, and boundary percentage of the top players over time. These graphs allow the users to analyze the consistency of the players over time and track their progress. The bubble graph is an excellent tool to compare the batting average and strike rate of the top players.



Figure 4 : Finishers tab in Players Analysis Tab

3.1.4. Fast Bowlers :

The Fast Bowlers category displays the performance statistics of the bowlers who bowl at high speeds. The statistics include bowling average, bowling strike rate, runs conceded, economy rate, and wickets taken. The line graphs show the bowling average, bowling strike rate, and economy rate of the top players over time. These graphs allow the users to analyze the consistency of the players over time and track their progress.



3.2. Playing 11 :

In the Playing 11 section of our Cricket Statistics Dashboard, we have compiled a comprehensive list of all the players who have been a part of our analysis in the Player Analysis tab. From this list, we can select the best 11 players who are likely to perform well in the upcoming matches. We can perform filtering using the PowerBI tools whenever required, to narrow down our choices based on different performance metrics and criteria, such as runs scored, batting averages, strike rates, bowling averages, economy rates, and more. This makes it easier for us to select the best 11 players for the team, based on their recent form, fitness, and suitability for the specific match conditions. We can also sort the players based on their batting or bowling statistics such as average, strike rate, economy rate, etc. This makes it easier for us to make informed decisions while selecting the playing 11 for the next match. Overall, the Playing 11 section provides a holistic view of the performance of all players and helps us in building a strong team for the upcoming matches.



Select Your Top 11	Name	Team	playingRole	Batting Style	Batting Avg	Strike rate	bowlingStyle	Economy	Bowling Strike Rate	Bowling Average	Batting Position
<input type="checkbox"/> aaron finch	lorcan tucker	Ireland	Wicketkeeper Batter	Right hand Bat	40.80	125.45		0.00		0.00	3.00
<input type="checkbox"/> asyan afzal khan	suryakumar yadav	India	Batter	Right hand Bat	59.75	109.68	Right arm Medium, Right arm Offbreak	0.00		0.00	4.00
<input type="checkbox"/> adam zampa	ired klaassen	Netherlands	Bowler	Right hand Bat	3.33	90.91	Left arm Fast medium	0.33	22.50	1.25	10.00
<input type="checkbox"/> adil rashid	sam curran	England	Allrounder	Left hand Bat	12.00	85.71	Left arm Medium fast	0.53	10.46	0.92	7.00
<input type="checkbox"/> alif hassain	paul van meekeren	Netherlands	Bowler	Right hand Bat	45.00	166.67	Right arm Fast medium	1.45	26.91	4.09	11.00
<input type="checkbox"/> ahmed raza	michael jones	Scotland	Top order Batter	Right hand Bat	36.67	142.86	Right arm Offbreak	0.00		0.00	2.00
<input type="checkbox"/> aiden markram	shabheen shah Afridi	Pakistan	Bowler	Left hand Bat	11.00	283.33	Left arm Fast	0.87	13.73	2.00	9.00
<input type="checkbox"/> akeal hosen											
<input type="checkbox"/> alex hales											
<input type="checkbox"/> alishan sharafu											
<input type="checkbox"/> alzani joseph											
<input type="checkbox"/> andy balbirnie											
<input type="checkbox"/> anrich Nortje											
<input type="checkbox"/> arshdeep singh											
<input type="checkbox"/> aryan lakra											
<input type="checkbox"/> ashton agar											
<input type="checkbox"/> asif ali											
<input type="checkbox"/> axar patel											

Figure 5 : Finishers tab in Players Analysis Tab

4. Conclusion

In conclusion, the Cricket Statistics Dashboard is a comprehensive tool that enables cricket enthusiasts to analyze and select the top-performing players for upcoming matches. It provides an in-depth analysis of four categories of players: openers, anchors, fast bowlers, and finishers, allowing users to compare and evaluate players' performances based on different metrics. This project has a lot of potential for future improvements and developments. For instance, we can add more metrics to the existing categories or add new categories based on user feedback. We can also incorporate live data feeds and real-time match statistics to keep the dashboard updated with the latest information. Another feature that we can add is the ability to compare different players directly on the dashboard. This feature can be useful in scenarios where the user wants to compare players from different teams or different formats of the game. Additionally, we can expand the dashboard to cover other formats of cricket, such as Test cricket and One Day Internationals (ODIs). This would provide cricket enthusiasts with a more comprehensive view of player performances across different formats of the game. Finally, we can also explore the possibility of integrating machine learning algorithms to make predictions about players' future performances based on their historical data. This would be a significant addition to the dashboard and would provide users with valuable insights into players' potential for upcoming matches.

Overall, the Cricket Statistics Dashboard has the potential to be a powerful tool for cricket enthusiasts, and with future improvements and developments, it can become even more useful for fans, analysts, and coaches alike.