



Beyond Boundaries: A Comprehensive Cricket Analytics Dashboard

SHRUTI .C.S

*M.Sc (Decision and
Computing Sciences)
Coimbatore institute of
technology
Coimbatore, India*

DEEPAK RAJ .A

*M.Sc (Decision and
Computing Sciences)
Coimbatore institute of
technology
Coimbatore, India*

Dr.V.Radhamani

*Assistant Professor
Dept. of Computing (DCS)
Coimbatore institute of
technology
Coimbatore, India*

Abstract- This research delves into the intricate intricacies of cricket, employing advanced tools like Power BI and Streamlit to analyze T-20 Cricket World Cup and IPL data. It begins with web scraping data from ESPN Cricinfo and then rigorously cleans and preprocesses it. Data modeling in Power BI's DAX is central to the project, enabling nuanced insights and an interactive dashboard for match statistics, player performance, and team strategies. The Streamlit interface makes the findings accessible to a broad audience. This "Cricket Analysis Dashboard" project redefines cricket data analysis and decision-making by harmoniously merging advanced analytics with a user-friendly interface, setting a new standard in sports analytics and unlocking fresh insights in the world of cricket.

Keywords –*Streamlit, Power BI, DAX.*

I. INTRODUCTION

In the realm of sports analytics, this research delves into the intricate intricacies of cricket, employing advanced tools such as Power BI and Streamlit to dissect data from the T-20 Cricket World Cup and IPL. This paper provides a comprehensive overview of the project's methodology, underscoring its potential to transform the understanding and strategic planning within the world of cricket. The

research begins with the acquisition of data through web scraping techniques, extracting comprehensive match data from the renowned ESPN Cricinfo website. The ensuing stages involve a rigorous process of data cleaning and preprocessing, executed through Pandas, ensuring the data's integrity and reliability. Power Query is employed for data transformation, facilitating a seamless transition from raw data to a structured format amenable to thorough analysis. The core of this research lies in data modeling, a critical phase orchestrated through Power BI's DAX (Data Analysis Expressions). This phase includes the creation of sophisticated parameters and calculations that enable the extraction of nuanced insights. The integration of diverse datasets culminates in the development of an interactive and dynamic dashboard that offers a comprehensive view of match statistics, player performance, and team strategies. Streamlit, a Python framework, serves as the user interface, enabling seamless access and interaction with the Power BI dashboard. This integration makes the findings accessible to a wide audience, bridging the gap between complex analytics and end-users.

This paper highlights the significance of the "Cricket Analysis Dashboard" project as an innovative venture poised to redefine the landscape of cricket data analysis and strategic decision-

making. By harmoniously merging state-of-the-art data analytics techniques with user-friendly interfaces.

II. DATA COLLECTION AND FEATURES

For the development of the Cricket Analysis Dashboard focused on T20 cricket, the primary step is data collection through web scraping techniques. To ensure a comprehensive and reliable dataset, we extract detailed match data from the reputable source of ESPN Cricinfo. This data collection process encompasses a wide range of statistics, including player performance metrics, match outcomes, team strategies, and various other pertinent information. By sourcing data from this renowned cricket platform, we lay a solid foundation for in-depth analysis, allowing for a detailed examination of player and team performance in the T20 format. This rich dataset will be instrumental in the subsequent phases of data cleaning, transformation, and modeling, ultimately leading to the creation of an interactive and insightful Cricket Analysis Dashboard. We have used 4 different datasets for the system. The various information's availed in the columns are:

Team1: The team that is played as home team.

Team2: The team that is played as away team.

Winner: The winning team of the game.

Margin: The margin of victory.

Match_Date: The date of the match that is conducted.

Match_id: The specific id for the match.

Ground: The ground at which the game is played.

Batsman: The name of the player who is batting.

SR: The Strike rate at which the batter scores runs.

III. PROCESSING STEPS

Data cleaning is a critical step in preparing data for cricket analytics, as it ensures that the data is accurate, consistent, and suitable for analysis. In cricket analytics, you're likely dealing with various data sources, such as match statistics, player performance records, and more. Here are some common steps for data cleaning in cricket analytics:

1. Removing Null Values: Null values in the dataset can introduce biases and inaccuracies in analysis. To maintain the integrity of our data, any rows with missing values will be removed. This process ensures that our dataset remains consistent and reliable for subsequent analyses.

2. Removing Unusual Characters: Unusual characters, such as special symbols or encoding artifacts, can distort the data and lead to misinterpretation. These characters will be systematically identified and removed from the dataset. This cleansing step is essential for ensuring the uniformity and correctness of textual information.

3. Matching the Matches with Batting and Bowling: To conduct a comprehensive analysis, it is crucial to match batting and bowling data accurately. Each match's batting and bowling statistics need to be correctly aligned. This alignment guarantees that the performance metrics for both teams are consistent, enabling meaningful comparative analyses and accurate predictions.

4. Creating Unique Match IDs: Creating unique identifiers (Match IDs) for each match entry is essential for efficient data management and retrieval. These unique identifiers simplify the process of cross-referencing data across different datasets or sources. Match IDs will be generated based on a combination of match date, team names, or other unique attributes to ensure their distinctiveness.

5. Removing (c) and † Symbols: Special symbols like "(c)" denoting captaincy and "†" indicating a wicketkeeper/batsman can cause inconsistencies in data analysis. These symbols will be removed to standardize player names and roles. Standardization ensures that the dataset is uniform, enabling accurate player-specific analyses without distortions caused by these symbols.

6. Dropping Duplicate Values: Duplicate entries can skew analyses and lead to biased results. Duplicate rows, if any, will be identified and dropped from the dataset. Removing duplicates ensures that each match or player entry is unique, preventing any data redundancy and maintaining the dataset's integrity.

Match_Date	Match_Id	Team_1	Team_2	Winner	Margin	Ground
0	16-Oct-22	T20I # 1823	Namibia Sri Lanka	Namibia	55 runs	Geelong
1	16-Oct-22	T20I # 1825	Netherlands United Arab Emirates	Netherlands	3 wickets	Geelong
2	17-Oct-22	T20I # 1826	Scotland West Indies	Scotland	42 runs	Hobart
3	17-Oct-22	T20I # 1828	Ireland Zimbabwe	Zimbabwe	31 runs	Hobart
4	18-Oct-22	T20I # 1830	Namibia Netherlands	Netherlands	5 wickets	Geelong
5	18-Oct-22	T20I # 1832	Sri Lanka United Arab Emirates	Sri Lanka	79 runs	Geelong

Figure 2.1-Dataset Description

Name	Image	Team	Batting_Style	Bowling_Style	Playing_Role	Description
16 Aaron Finch	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Right hand Bat	Slow Left arm Orthodox	Top order Batter	A solidly built, aggressive batter from a coun...
17 Pat Cummins	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Right hand Bat	Right arm Fast	Bowler	Earmarked as a potential star from a young age...
18 Ashton Agar	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Left hand Bat	Slow Left arm Orthodox	Bowler	Ashton Agar initially turned almost as many he...
19 Tim David	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Right hand Bat	Right arm Offbreak	Middle order Batter	Mumbai Indians' decision to sign Tim David for...
20 Cameron Green	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Right hand Bat	Right arm Fast medium	Batting Allrounder	Contracted to Western Australia as a schoolboy...
21 Josh Hazlewood	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Left hand Bat	Right arm Fast medium	Bowler	A tall and accurate fast bowler who is unrelie...
22 Mitchell Marsh	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Right hand Bat	Right arm Medium	Allrounder	Part of one of the most well-known family name...
23 Glenn Maxwell	https://encrypted-tbn0.gstatic.com/images?q=tb...	Australia	Right hand Bat	Right arm Offbreak	Batting Allrounder	One of the fastest scorers in world cricket. G...

Figure 2.2-Player Description

Match	Bowling_Team	bowler	overs	maidens	runs	wickets	economy	Os	4s	6s	WD	NB	Match_Id	
0	Namibia vs Sri Lanka	Sri Lanka	Maheesh Theekshana	4.0	0	23	1	5.75	7	0	0	2	0	T20I # 1823
1	Namibia vs Sri Lanka	Sri Lanka	Dushmantha Chameera	4.0	0	39	1	9.75	6	3	1	2	0	T20I # 1823
2	Namibia vs Sri Lanka	Sri Lanka	Pramod Madushan	4.0	0	37	2	9.25	6	3	1	0	0	T20I # 1823
3	Namibia vs Sri Lanka	Sri Lanka	Chamika Karunaratne	4.0	0	36	1	9.00	7	3	1	1	0	T20I # 1823
4	Namibia vs Sri Lanka	Sri Lanka	Wanindu Hasaranga de Silva	4.0	0	27	1	6.75	8	1	1	0	0	T20I # 1823

Figure 3.1-Bowling Data (World cup)

Match	Team_Innings	Batting_Pos	Batsman	Dismissal	Runs	Balls	4s	6s	SR	Match_Id
0	Namibia vs Sri Lanka	Namibia	1 Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	T20I # 1823
1	Namibia vs Sri Lanka	Namibia	2 Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	T20I # 1823
2	Namibia vs Sri Lanka	Namibia	3 Jan Nicol Loftie-Eaton	c +Mendis b Karunaratne	20	12	1	2	166.66	T20I # 1823
3	Namibia vs Sri Lanka	Namibia	4 Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33	T20I # 1823
4	Namibia vs Sri Lanka	Namibia	5 Gerhard Erasmus (c)	c Gunathilaka b PWH de Silva	20	24	0	0	83.33	T20I # 1823

Figure 3.2-Bating Data(World Cup)

match	teamInnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	matchDate
0	Super Kings Vs Titans	Super Kings	1 DevonConway	b Mohammed Shami	1	6	0	0	16.66	Mar 31, 2023
1	Super Kings Vs Titans	Super Kings	2 RuturajGaikwad	c Shubman Gill b Joseph	92	50	4	9	184.00	Mar 31, 2023
2	Super Kings Vs Titans	Super Kings	3 MoeenAli	c +Saha b Rashid Khan	23	17	4	1	135.29	Mar 31, 2023
3	Super Kings Vs Titans	Super Kings	4 BenStokes	c +Saha b Rashid Khan	7	6	1	0	116.65	Mar 31, 2023
4	Super Kings Vs Titans	Super Kings	5 AmbatRayudu	b Little	12	12	0	1	100.00	Mar 31, 2023

Figure 3.3-Bating Data(IPL)

IV. STREAMLIT INTEGRATION

Incorporating the cricket data analytics project, developed using Power BI, into the Streamlit framework in Python is a crucial aspect of our presentation. Streamlit acts as the user-friendly interface bridging the gap between complex data analytics and end-users, making findings accessible to a wider audience. Our project, centered on the T-20 Cricket World Cup, began with web scraping from ESPN Cricinfo, followed by rigorous data cleaning and preprocessing to ensure data reliability. The refined dataset integrates seamlessly into Streamlit, a Python framework designed for creating interactive web applications. This integration empowers users to navigate match statistics, assess player performances, and make informed decisions, including selecting the best playing 11 from world-class players. This practical and user-friendly tool transforms analytical insights, setting a new standard in cricket data analysis and strategic decision-making, unlocking new possibilities in T-20 cricket.

V. DATA TRANSFORMATION IN POWER QUERY

- 1. Data Cleansing:** Removal of null values, duplicate rows, and unwanted characters. Handling missing data through imputation or removal strategies.
- 2. Data Standardization:** Standardizing formats, units, and naming conventions to ensure consistency. Techniques for transforming categorical data into a uniform format.
- 3. Data Enrichment:** Incorporating additional data from external sources for comprehensive

analysis.Joining, merging, and appending data tables for enriched datasets.

4. Data Aggregation: Aggregating data to different granular levels (hourly, daily, monthly) for summary analysis.Calculating key performance indicators (KPIs) using aggregation functions.

5. Data Transformation Functions: Exploration of Power Query functions (e.g., TransformColumnTypes, Group By, Merge Queries) and their applications. Custom function creation for complex data transformations.

6. Data Pivoting and Unpivoting: Techniques for pivoting data tables for wide-to-long format transformations and vice versa. Utilizing unpivot operations for better analysis of multi-dimensional data.

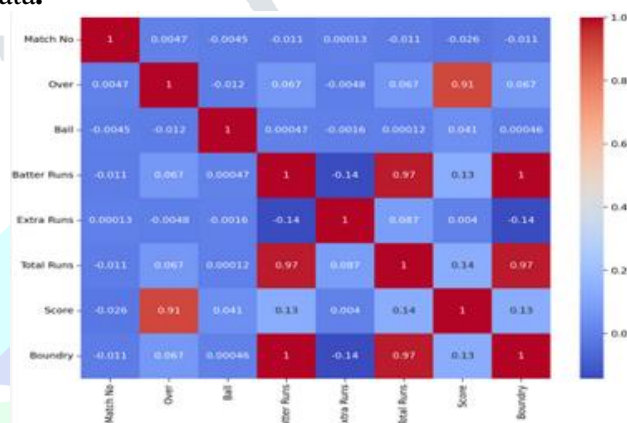


Figure 5.1-Correlation Heatmap



Figure 5.2-PairPlot.

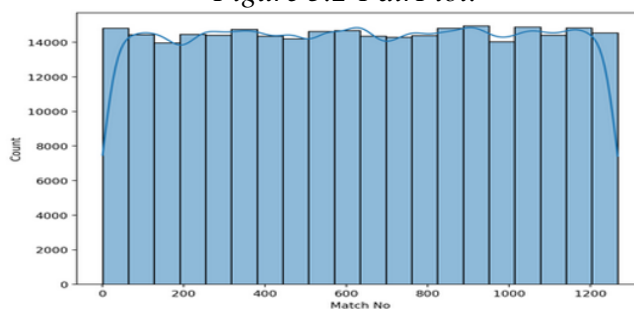


Figure 5.3-Histogram for match no. and score

VI. RESULT ANALYSIS

This section delves into the findings derived from our cricket analytics project, where we harnessed the capabilities of Power BI and Streamlit to gain in-depth insights into cricket match outcomes, player performances, and team statistics. The Power BI dashboard we developed provided a multifaceted perspective on cricket analytics, offering a comprehensive suite of visualizations, such as tables, charts, and maps. Through these visualizations, users could unravel intricate patterns, discern trends, and establish correlations in the data. Additionally, the interactive features, including filters and slicers, empowered users to tailor their exploration, enabling a granular examination of specific aspects, be it the performance of particular teams, players, time periods, or match types. This analytical endeavor unearthed a plethora of insights, spanning batting and bowling averages, match-winning strategies, and the significant impact of certain players on game outcomes.

with the data, providing a dynamic, engaging environment for cricket enthusiasts and analysts to scrutinize player statistics, team performance, and match results.

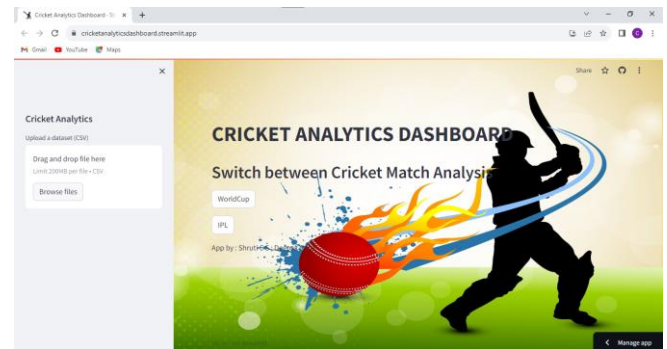


Figure 6.3 – Streamlit dashboard

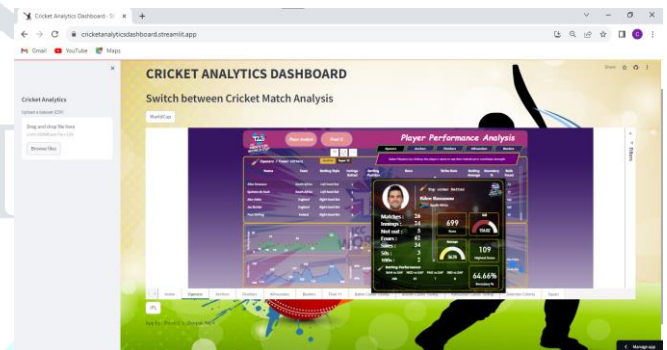


Figure 6.4 – WorldCup Analysis

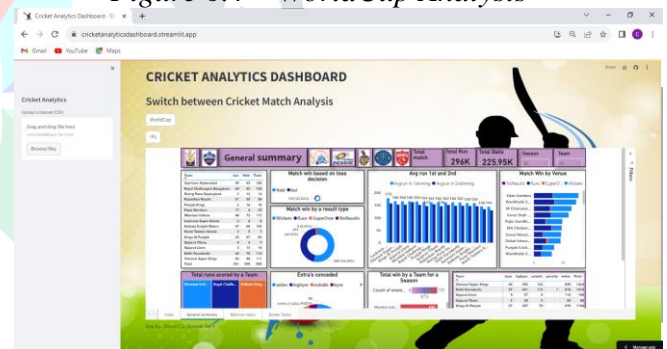


Figure 6.5 – IPL Analysis

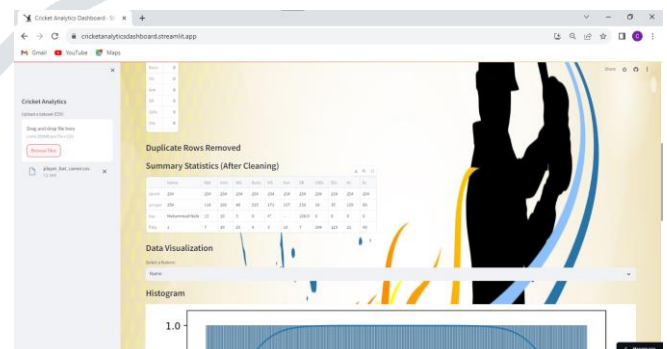


Figure 6.6 –Cricket Data Analysis



Figure 6.1 – Analytics Dashboard.



Figure 6.2 – Opener Player Analysis

On the other hand, our Streamlit web application served as a complementary platform for user-friendly data interaction. The interface of the application incorporated intuitive elements such as dropdown menus, sliders, and buttons, rendering data exploration a seamless experience. Within this environment, plots, tables, and visual representations breathed life into cricket data, facilitating rapid comprehension of critical information. The real-time data synchronization mechanism implemented in Streamlit ensured that users received prompt feedback as they interacted

When we conducted a combined analysis of the results obtained from both Power BI and Streamlit, it became evident that each tool brought unique strengths to the table. While Power BI excelled in providing an extensive range of visualizations and advanced interactivity, Streamlit added an extra layer of user-friendliness and real-time

responsiveness. This validation process further confirmed the consistency and reliability of the cricket analytics insights derived from our approach.

VII. CONCLUSION

In conclusion, the Cricket Data Analytics project, centered around the T-20 Cricket World Cup and IPL, represents a significant milestone in the realm of cricket data analysis. Leveraging the power of Power BI, this project provides an invaluable tool for comprehensively assessing the tournament's data. Not only can users delve into the intricate details of each match, but they can also make informed decisions by selecting the best playing 11 from the pool of world-class players in the competition. This project marks a transformative step in enhancing strategic decision-making in cricket, offering an interactive and data-driven approach to understanding the dynamics of T-20 cricket. It underscores the immense potential of data analytics in shaping the future of the sport and sets a new standard for cricket enthusiasts, professionals, and fans alike.

VIII. FUTURE ENHANCEMENT

1.Real-time Data Integration: Integrate real-time data feeds to provide live updates during matches. This can include ball-by-ball data, player statistics, and team performance metrics, allowing fans and analysts to follow the game closely.

2.Social Media Sentiment Analysis: Incorporate sentiment analysis from social media platforms to gauge fan reactions and opinions about players, teams, and matches. This can provide additional context to the analytics data.

3.Fan Engagement Features: Create features that engage fans, such as polls, quizzes, and fantasy league integrations. This can enhance the overall fan experience and keep them actively involved during matches.

4.Match Simulation: Develop a simulation feature that allows users to simulate match scenarios based on historical data. This can help coaches and analysts strategize for upcoming matches and simulate different game situations.

5.Video Analysis Integration: Integrate video analysis tools to the dashboard, enabling users to correlate statistical data with actual match footage. This can help in understanding player techniques, strategies, and areas of improvement.

IX. REFERENCES

- [1] Jayal A, McRobert A, Oatley G, et al. Sports analytics: analysis, visualisation and decision making in sports performance. London: Routledge. Epub ahead of print 2018.
- [2] Gudmundsson J, Horton M (2017) Spatio-temporal analysis of team sports. *ACM Comput Surv (CSUR)* 50(2):22
- [3] Legg PA, Chung DH, Parry ML, Jones MW, Long R, Griffiths IW, Chen M (2017) Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. In: *Computer graphics forum, Wiley Online Library*, vol 31, pp 1255–1264
- [4] Perin C, Vuillemot R, Stolper CD, Stasko JT, Wood J, Carpendale S (2018) State of the art of sports data visualization. In: *Computer graphics forum, Wiley Online Library*, vol 37, pp 663–686
- [5]https://stats.espncricinfo.com/ci/engine/records/team/match_results.html?id=14450;type=tournament
- [6] Sholto-Douglas R, Cook R, Wilkie M, et al. Movement demands of an elite cricket team during the big bash league in Australia. *J Sport Sci Med.* 2020;19:59–64.
- [7] Herridge R, Bishop C, Turner A. Monitoring changes in power, speed, agility and endurance in elite cricketers during the off-season. *J Strength Cond Res.* 2017;34:2285–93.
- [8] Christie C, Sheppard B, Goble D, et al. Strength and sprint time changes in response to repeated shuttles between the wickets during batting in cricket. *J Strength Cond Res.* 2019;33:3056–64.
- [9]<https://www.espncricinfo.com/series/icc-men-s-t20-world-cup-2022-23-1298134/namibia-vs-sri-lanka-1st-match-first-round-group-a-1298135/live-cricket-score>
- [10] <https://2022.t20worldcup.com/>