END TO END CRICKET DATA ANALYTICS

A MINOR PROJECT REPORT

submitted in partial fulfilment of the

requirement for the award of the degree

Of

Master of Computer Applications (MCA)

HARIPRASAD J P

23FS20MCA00017

Department of Computer Applications

Manipal University Jaipur

Jaipur, Rajasthan, 303007

Dec 2024

DEPARTMENT OF COMPUTER APPLICATION

MANIPAL UNIVERSITY JAIPUR, JAIPUR-303007 (RAJASTHAN) INDIA

15-12-2024

CERTIFICATE

This is to certify the project titled End to End Cricket Data Analytics is a record of the bonafide work completed during the period from 25.08.2024 to 15.12.2024 by Hariprasad JP (23FS20MCA00017) submitted in the partial fulfilment of the requirements for the award of the Degree of Master of Computer Applications (MCA) at the Department of Computer Applications, Manipal University Jaipur, for the academic year 2023-2025

Dr. Linesh Raja
Project Guide, Dept of Computer Application
Manipal University Jaipur

Dr. Shilpa Sharma
HOD, Dept of Computer Application
Manipal University Jaipur

DECLARATION

I hereby declare that this project report entitled End to End Cricket Data Analytics is an original and independent effort completed in fulfilment of the academic requirements for my program. The methodologies, analyses, and conclusions presented are derived from my work, adhering to academic integrity and ethical guidelines.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to the following individuals whose support and guidance were instrumental in the successful completion of this project End to End Cricket Data Analytics

First and foremost, I extend my sincere appreciation to the Dean/Director for providing the necessary resources and infrastructure essential for the execution of this project.

I am deeply indebted to the Head of the Department for their constant encouragement and support throughout this endeavour.

I extend my profound gratitude to my Project Supervisor for their invaluable guidance, patience, and expertise, which significantly contributed to the development and execution of this project.

Special thanks are due to my department guide for their insightful feedback and constructive criticism, which greatly enhanced the quality of the work presented in this project.

I would also like to acknowledge the Laboratory In charge for their assistance in providing access to the required facilities and equipment essential for conducting experiments and simulations.

Finally, I am grateful to all the faculty members whose expertise and feedback were indispensable in shaping the direction and outcomes of this project.

Their unwavering support and encouragement have been pivotal in every step of this journey, and I am truly thankful for their invaluable contributions.

Thank you

Abstract

This paper offers a comprehensive, end-to-end cricket data analytics project focused on the 2024 ICC T20 World Cup, in which it performs an in-depth analysis of player performances and uses performance to predict the best 11. The project begins with acquiring diverse datasets from trusted sources such as ESPNcricinfo, Kaggle, and Bright Data, which offers a web scraping solution for real-time data extraction, enabling the collection of a wide array of statistics, including player scores, batting averages, bowling performance, team rankings, and match results.

For data quality assurance, preprocessing and cleaning of the dataset were performed in Python and Pandas. There was a focus on dealing with missing values, removing unnecessary data, and ensuring a standardized format to maintain uniformity in different sources. The first step laid down a strong foundation for subsequent analysis that was reliable and accurate.

The next step involved advanced data transformation with Power Query in Power BI. Here, data was reshaped to give a more coherent, structured format to analyse. KPIs(Key Performance Indicators) like batting and bowling strike rates, economy rates, and boundaries hit were calculated to get deeper insights into the player's performances and the team's strategies. Utilized Power Query to join different data sources, perform operations such as grouping and filtering, and also demonstrate different transformations that have brought trends over time into relief.

The second step was using DAX, Data Analysis Expressions, in order to build dynamic models and parameters. These models include complex calculations, such as player form trends, team momentum, and match prediction models. All these enable the production of actionable insights that will be useful in making informed strategic decisions in the tournament by teams and analysts.

The project finally ends with an interactive dashboard in Power BI, that is easy to use and really beautiful. The dashboard carries heatmaps, bar charts, line graphs, and scatter plots, which means stakeholders can analyze and play with data on performances of players, strategies in teams, and match outcomes. The capability of real-time filtering and drill-downs facilitates more detailed analysis in focused areas of interest, like how a player has performed in various match conditions, or team performance across the various stages of the tournament, or comparisons between teams.

This project demonstrates an application of advanced data analytics in sports and contributes to a growing field of sports analytics by offering actionable insights into cricket performance. Combining data from multiple sources with advanced analytical techniques, it provides valuable support to coaches, players, and analysts with the potential to influence on-field performance as well as off-field strategies in future cricket tournaments.

Contents		
		Page No
Certificate		i
Declaration		vi
Acknowledgement		iii
Abstract		vi
List of Figures and Tables		vi
Chapter 1	**INTRODUCTION**	
1.1	Purpose of the Project	1
1.2	Importance of Cricket Data Analytics	1
1.3	Objectives of the Study	1
1.4	Overview of the T20 World Cup 2024	2
Chapter 2	**SURVEY OF TECHNOLOGY**	
2.1	Introduction to Data Analytics	3
2.2	Web Scraping Techniques	3
2.3	Python and Pandas for Data Processing	3
2.4	Power BI for Data Visualization	3
Chapter 3	**REQUIREMENT ANALYSIS**	
3.1	Problem Definition	4
3.2	Drawback of Existing Systems	4
3.3	Requirement Specification	4
3.4	Feasibility Study	5

List of Figures and Tables

Chapter 1: INTRODUCTION

The T20 World Cup brings together a diverse audience, with extraordinary performances, which marks the summit of international cricketing excellence. With the next edition in 2024, it will be much more innovative and competitive and, therefore, will need advanced analytical frameworks to interpret trends and come up with insights. It therefore develops a strong data pipeline in collecting, processing, modelling, and visualizing data. This study will make use of Python in pre-processing the data and PowerBI in the creation of interactivity in dashboards and will clarify patterns and dynamics significant to the game's players, analysts, and enthusiasts. Results coming from this effort will shift cricket analytics as they implement data-driven strategies and enjoy better views.

1.1 Purpose of the Project

The objective behind this effort is to design a complete cricket analytics framework that focuses on the T20 World Cup 2024. This will give users an interface to data through visual dashboards which will be significant in displaying insights regarding both player and team performances. This will ultimately enhance the decision-making of coaches, analysts, and cricket enthusiasts.

1.2 Importance of Cricket Data Analytics

Analyzing cricket data is an integral part of the improvement in the strategic and tactical decision-making processes. By using both historic and real-time data, any player or team would be able to get an idea of trends about their performance and improve weaknesses while developing a winning strategy. The technological developments, particularly in gathering and methods of analysis of cricket data, have now brought analytics in cricket to its center as an essential ingredient of contemporary cricket, most particularly the T20 World Cup

1.3 Objectives of the Study

The primary objectives of this study are:

- To collect and preprocess cricket data from various sources.
- To develop an interactive PowerBI dashboard for visualizing key insights.

- To apply data transformation techniques to enhance the accuracy and usefulness of the analysis.
- To evaluate player and team performance trends, identifying strengths and weaknesses.

1.4 Overview of the T20 World Cup 2024

The T20 World Cup 2024 is one of the most anticipated international cricket events, featuring top teams from around the globe competing in the fast-paced T20 format. The tournament is expected to showcase high levels of competition, with both established and emerging players pushing their limits. Understanding team dynamics and player performances during the event will provide valuable insights into the evolving strategies in the game.

Chapter 2: SURVEY OF TECHNOLOGY

2.1 Introduction to Data Analytics

Data analytics is the systematic computer-based examination of data to find out meaningful patterns, trends, and insights. In cricket, data analytics helps to analyze the performance of a particular player or team and find strategic patterns, thus predicting the outcome in the future. It involves the gathering of raw data, cleaning of data, and applying analytical methods to derive actionable knowledge.

2.2 Web Scraping Techniques

This involves extracting data directly from websites using an automated script. For the project, it has been using web scraping to extract match results along with the statistics of the players from ESPNcricinfo. Large sets of data are thus extracted methodically, hence providing inputs relevant for analysis and making decisions.

2.3 Python and Pandas for Data Processing

The programming language Python, in combination with the Pandas library, is heavily used for all the work involved in data processing. The Pandas library helps in cleaning and preparing the data so that unclean data is transformed into usable data for informed decision-making. In this project, Python scripts were created to handle match statistics, player performance metrics, and other relevant information to ensure consistency and accuracy before including it in the PowerBI model.

2.4 Power BI for Data Visualization

Power BI is one powerful analytical tool for business where data can be represented as interactive dashboards. Being dynamic, it makes an excellent tool for this current project because it changes as soon as the original input is changed. Among the graphics that were made possible using Power BI include: bar graphs, line graphs, and pie charts to show performances of the players and teams, the result of the matches, and the trend of the tournaments.

Chapter 3: REQUIREMENT ANALYSIS

3.1 Problem Definition

The project addresses all the issues that exist in modern packages of cricket analytics. All these issues relate to a lack of dynamic interactive visualizations, data heterogeneity, and limited analytical operations that severely restrict critical reviews of performance and strategic decisions. Combining data preprocessing, transformation, and visualization methods does provide a better balance than the traditional segregation of processes.

3.2 Drawback of Existing Systems

Limitations in current cricket analytics practices include:

- **Static Dashboards**: Most systems offer static, one-time reports, limiting real-time or on-the-fly analysis.
- **Inconsistent Data Sources**: Existing systems rely on fragmented datasets that often require manual intervention to consolidate.
- **Limited Predictive Insights**: Existing systems do not integrate predictive analytics to forecast player or team performance.
- **Lack of Interactivity**: Few systems offer a truly interactive experience, limiting user engagement and exploration of data.

3.3 Requirement Specification

Requirement Type	Details
Hardware	8 GB RAM, Intel i5 processor, SSD storage.
Software	Python 3.9, PowerBI Desktop, Pandas library.
Data Sources	ESPNcricinfo datasets, Kaggle repositories.

TABLE : 3(a)

3.4 Feasibility Study

This project is economically viable because the source tools, Python and PowerBI, are free. The tools applied are known and respected in the industry; operations have a timeline and deliverables, so the project will finish within the defined parameters.

Chapter 4: PLANNING AND SCHEDULING

4.1 Project Timeline

Week	Task
Week 1-2	Requirement Analysis & Web Scraping
Week 3-5	Data Processing, Cleaning, and EDA
Week 6-8	Dashboard Creation and Testing
Week 9	Presentation and Documentation

TABLE : 4(a)

4.2 Milestones

- **Milestone 1**: Completion of data acquisition and preprocessing.
- **Milestone 2**: Data transformation and dashboard development.
- **Milestone 3**: Finalization of report and presentation.

Chapter 5: SYSTEM DESIGN

5.1 Data Flow Diagram

The flow chart illustrates the process and components involved in managing and analyzing cricket match data. It is divided into several sections, each representing different roles and functions within the system.

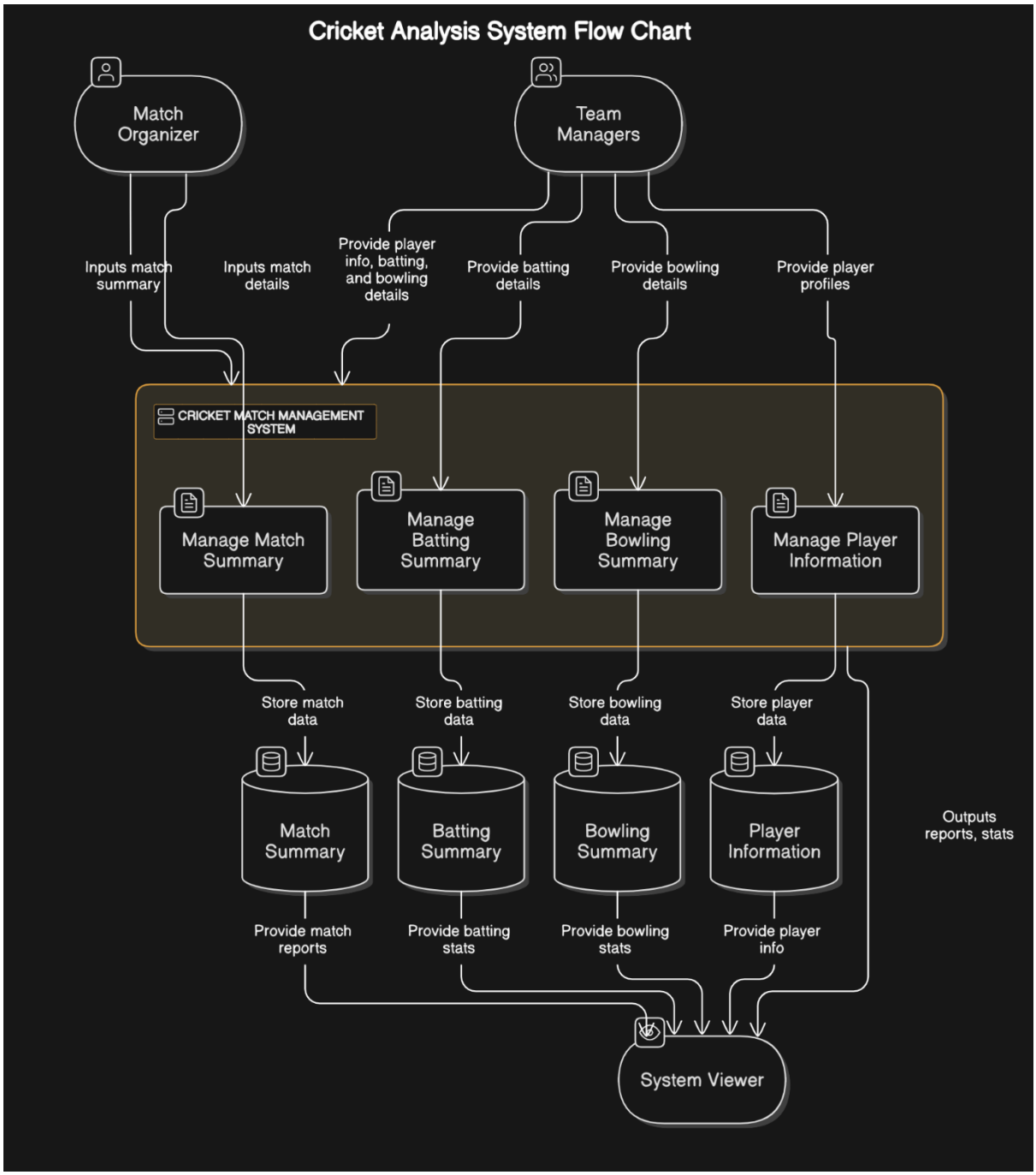


Fig 5.1

5.2 ER Diagram

The provided image depicts a database schema designed to manage and analyze cricket match data. This schema consists of four interconnected tables, each serving a specific purpose within the system.

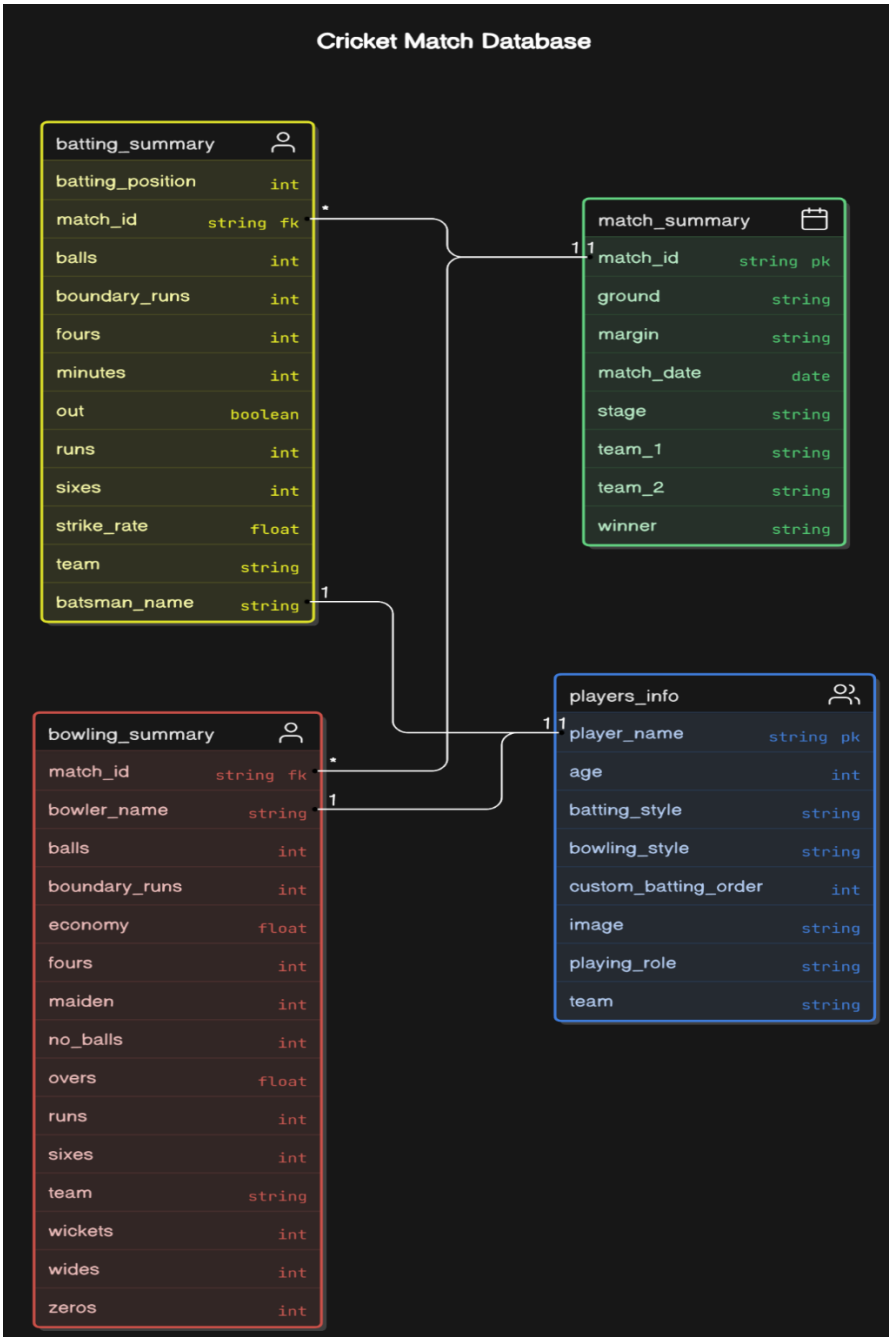


Fig 5.2

5.3 Data Dictionary and Data Model

Here we can see the detailed segment of a database schema for a T-20 cricket world cup analysis. The schema includes multiple tables designed to manage various aspects of cricket match data, specifically focusing on match summaries, batting summaries, bowling summaries, and player information.

Match_Summary

Attribute Name	Description	Data Type	Constraints
match_id	Unique identifier for the match	INTEGER	Primary Key
Ground	Location where the match is played	VARCHAR(50)	
Margin	Winning margin of the match	VARCHAR(20)	
Match_Date	Date when the match took place	DATE	
Stage	Stage of the tournament	VARCHAR(30)	
Team_1	Name of the first team	VARCHAR(50)	
Team_2	Name of the second team	VARCHAR(50)	
Winner	Name of the winning team	VARCHAR(50)	

TABLE : 5(a)

Batting_Summary

Attribute Name	Description	Data Type	Constraints
match_id	Identifier for the match	INTEGER	Foreign Key (match_id)
batsmanName	Name of the batsman	VARCHAR(50)	
battingPos	Position in the batting order	INTEGER	
balls	Number of balls faced	INTEGER	

boundary_runs	Runs scored from boundaries	INTEGER	
fours	Number of fours hit	INTEGER	
minutes	Minutes spent batting	INTEGER	
out	How the batsman got out	VARCHAR(30)	
runs	Total runs scored	INTEGER	
sixes	Number of sixes hit	INTEGER	
SR	Batting strike rate	FLOAT	
team	Name of the team	VARCHAR(50)	

TABLE : 5(b)

Bowling_Summary

Attribute Name	**Description**	**Data Type**	**Constraints**
match_id	Identifier for the match	INTEGER	Foreign Key (match_id)
bowlerName	Name of the bowler	VARCHAR(50)	
balls	Number of balls bowled	INTEGER	
boundary_runs	Runs conceded from boundaries	INTEGER	
economy	Bowling economy rate	FLOAT	
fours	Number of fours conceded	INTEGER	
maiden	Number of maiden overs bowled	INTEGER	
noBalls	Number of no balls bowled	INTEGER	

overs	Total overs bowled	FLOAT	
runs	Total runs conceded	INTEGER	
wickets	Number of wickets taken	INTEGER	
wides	Number of wides bowled	INTEGER	
zeros	Number of dot balls bowled	INTEGER	
team	Name of the team	VARCHAR(50)	

TABLE : 5(c)

Players_Info

Attribute Name	**Description**	**Data Type**	**Constraints**
name	Name of the player	VARCHAR(50)	Primary Key
age	Age of the player	INTEGER	
battingStyle	Batting style of the player	VARCHAR(30)	
bowlingStyle	Bowling style of the player	VARCHAR(30)	
Custom_Batting_Order	Custom batting order number	INTEGER	
image	URL or path to the player's image	VARCHAR(255)	
playingRole	Role of the player (e.g., batsman)	VARCHAR(30)	
team	Name of the team	VARCHAR(50)	Foreign Key (team)

TABLE : 5(d)

Relationships:

1. **match_summary** is connected to **batting_summary** and **bowling_summary** via the match_id attribute (Primary Key → Foreign Key).
2. **players_info** connects to **batting_summary** and **bowling_summary** through the team attribute.

5.4 Schema Design

The schema design outlines the structure of the databases used in the project, showing how data is organized and related within the tables for efficient querying and analysis.

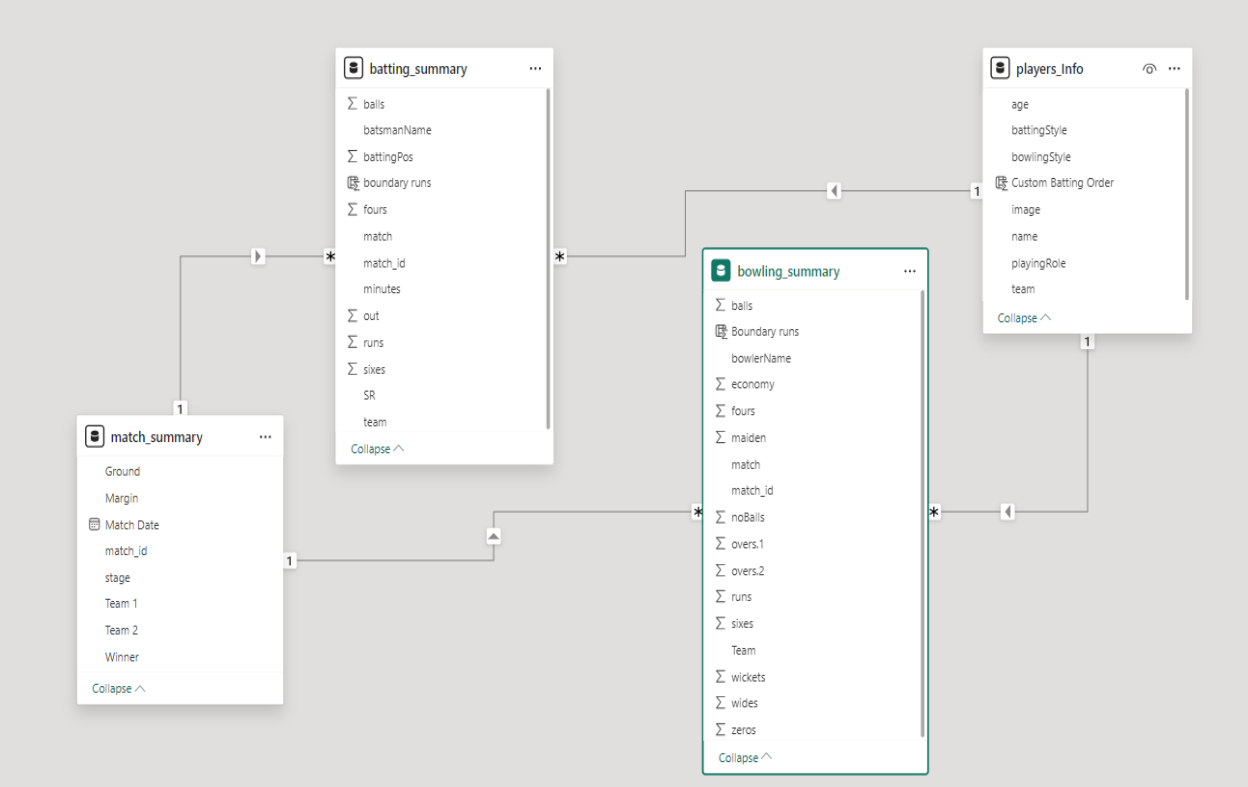


Fig 5.4(a)

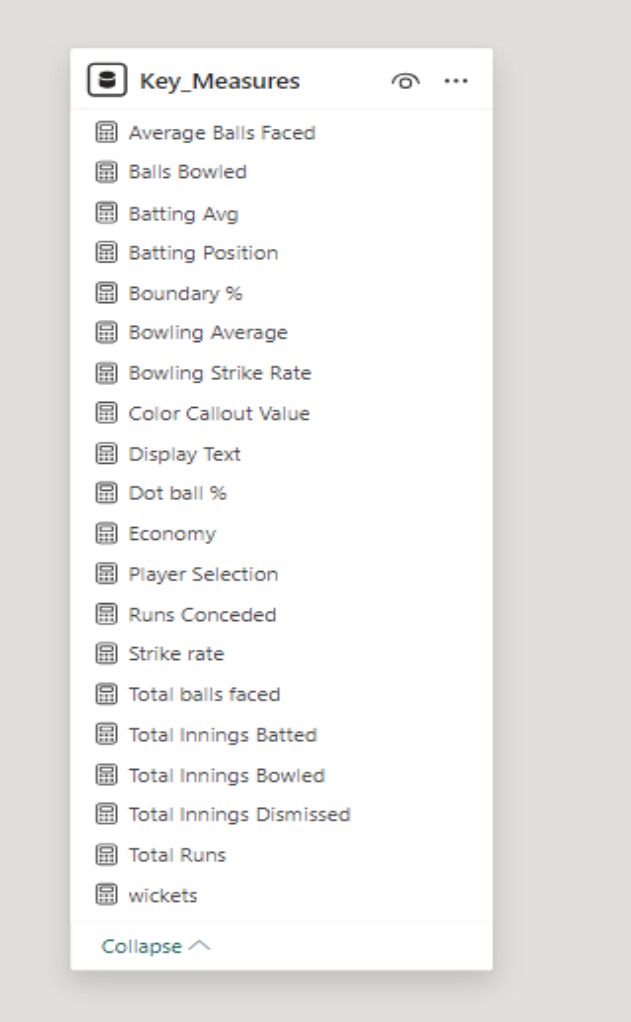

Key_Measures
Average Balls Faced
Balls Bowled
Batting Avg
Batting Position
Boundary %
Bowling Average
Bowling Strike Rate
Color Callout Value
Display Text
Dot ball %
Economy
Player Selection
Runs Conceded
Strike rate
Total balls faced
Total Innings Batted
Total Innings Bowled
Total Innings Dismissed
Total Runs
wickets
Collapse

Fig 5.4(b)

Chapter 6: IMPLEMENTATION

6.1 Web Scraping

This section outlines the approach and scripts used for web scraping to extract cricket data from websites like ESPNcricinfo.

```
/* -------------- STAGE 1 ------------ */
//------- 1.a Interaction Code ------ //

navigate('https://www.espncricinfo.com/records/tournament/team-match-results/icc-men-s-t20-world-cup-2024-15946');
collect(parse());



//------- 1.b Parser Code ------------//
//Step1: create an array to store all the records
let matchSummary = []

//Step2: Selecting all rows we need from target table
const allRows = $('table.engineTable > tbody > tr.data1');

//Step3: Looping through each rows and get the data from the cells(td)
Tabnine | Edit | Test | Expl
allRows.each((ind (parameter) element: any
    const tds = $(element).find('td');   //find the td
    matchSummary.push({
                'team1':  $(tds[0]).text(),
                'team2':  $(tds[1]).text(),
                'winner':  $(tds[2]).text(),
                'margin':  $(tds[3]).text(),
                'ground': $(tds[4]).text(),
                'matchDate': $(tds[5]).text(),
                'scorecard':   $(tds[6]).text()
    })
})


// step4: Finally returning the data
return {
  "matchSummary": matchSummary
};
```

```
/* -------------- STAGE 1 ------------ */

//------- 1.a Interaction Code ------ //
navigate('https://www.espncricinfo.com/records/tournament/team-match-results/icc-men-s-t20-world-cup-2024-15946');


let links = parse().playersLinks;
for(let i of links) {
  next_stage({url: i})
}

//------- 1.b Parser Code ------------//
let li const allRows: any
const allRows = $('table.engineTable > tbody > tr.data1');
    Tabnine | Edit | Test | Explain | Document | Ask
    allRows.each((index, element) => {
    const tds = $(element).find('td');
    const rowURL = "https://www.espncricinfo.com" +$(tds[6]).find('a').attr('href');
    links.push(rowURL);
})
return {
  'playersLinks': links
};
```

```
/* -------------- STAGE 2 ------------ */

//------- 2.a Interaction Code ------ //
navigate(input.url);
collect(parse());

//---------- 2.b Parser Code ---------//
var match = $('div').filter(function(){
    return $(this)
      .find('span > span > span').text() === String("Match Details")
}).siblings()
team1 = $(match.eq(0)).find('span > span > span').text().replace(" Innings", "")
team2 = $(match.eq(1)).find('span > span > span').text().replace(" Innings", "")
matchInfo = team1 + ' Vs ' + team2

var tables = $('div > table.ds-table');
var firstInningRows = $(tables.eq(1)).find('tbody > tr').filter(function(index, element){
  return $(this).find("td").length >= 11
})

var secondInningsRows = $(tables.eq(3)).find('tbody > tr').filter(function(index, element){
  return $(this).find("td").length >= 11
});
```

```
var bowlingSummary = []
Tabnine | Edit | Test | Explain | Document | Ask
firstInningRows.each((index, element) => {
  var tds = $(element).find('td');
  bowlingSummary.push({
        "match": matchInfo,
        "bowlingTeam": team2,
        "bowlerName": $(tds.eq(0)).find('a > span').text().replace('', ''),
        "overs": $(tds.eq(1)).text(),
        "maiden": $(tds.eq(2)).text(),
        "runs": $(tds.eq(3)).text(),
        "wickets": $(tds.eq(4)).text(),
        "economy": $(tds.eq(5)).text(),
        "0s": $(tds.eq(6)).text(),
        "4s": $(tds.eq(7)).text(),
        "6s": $(tds.eq(8)).text(),
        "wides": $(tds.eq(9)).text(),
        "noBalls": $(tds.eq(10)).text()
  });
});
```

```
Tabnine | Edit | Test | Explain | Document | Ask
secondInningsRows.each((index, element) => {
  var tds = $(element).find('td');
   bowlingSummary.push({
        "match": matchInfo,
        "bowlingTeam": team1,
        "bowlerName": $(tds.eq(0)).find('a > span').text().replace('', ''),
        "overs": $(tds.eq(1)).text(),
        "maiden": $(tds.eq(2)).text(),
        "runs": $(tds.eq(3)).text(),
        "wickets": $(tds.eq(4)).text(),
        "economy": $(tds.eq(5)).text(),
        "0s": $(tds.eq(6)).text(),
        "4s": $(tds.eq(7)).text(),
        "6s": $(tds.eq(8)).text(),
        "wides": $(tds.eq(9)).text(),
        "noBalls": $(tds.eq(10)).text()
  });
});

return {"bowlingSummary": bowlingSummary}
```

```
/* -------------- STAGE 1 ------------ */

//------- 1.a Interaction Code ------ //
navigate('https://www.espncricinfo.com/records/tournament/team-match-results/icc-men-s-t20-world-cup-2024-15946');


let links = parse().matchSummaryLinks;
for(let i of links) {
  next_stage({url: i})
}

//------- 1.b Parser Code ------------//
let links = []
const allRows = $('table.engineTable > tbody > tr.data1');
 Tabnine | Edit | Test | Explain | Document | Ask
allRows.each((index, element) => {
  const tds = $(element).find('td');
  const rowURL = "https://www.espncricinfo.com" +$(tds[6]).find('a').attr('href');
  links.push(rowURL);
 })
return {
  'matchSummaryLinks': links
};
```

```
/* -------------- STAGE 2 ------------ */

//------- 2.a Interaction Code ------ //
navigate(input.url);
collect(parse());

//------- 2.b Parser Code ------------//
var match = $('div').filter(function(){
    return $(this)
      .find('span > span > span').text() === String("Match Details")
}).siblings()
team1 = $(match.eq(0)).find('span > span > span').text().replace(" Innings", "")
team2 = $(match.eq(1)).find('span > span > span').text().replace(" Innings", "")
matchInfo = team1 +  ' Vs ' + team2

var tables = $('div > table.ci-scorecard-table');
var firstInningRows = $(tables.eq(0)).find('tbody > tr').filter(function(index, element){
  return $(this).find("td").length >= 8
})

var secondInningsRows = $(tables.eq(1)).find('tbody > tr').filter(function(index, element){
  return $(this).find("td").length >= 8
});
```

```
var battingSummary = []
Tabnine | Edit | Test | Explain | Document | Ask
firstInningRows.each((index, element) => {
  var tds = $(element).find('td');
  battingSummary.push({
        "match": matchInfo,
        "teamInnings": team1,
        "battingPos": index+1,
        "batsmanName": $(tds.eq(0)).find('a > span > span').text().replace('', ''),
        "dismissal": $(tds.eq(1)).find('span > span').text(),
        "runs": $(tds.eq(2)).find('strong').text(),
        "balls": $(tds.eq(3)).text(),
        "4s": $(tds.eq(5)).text(),
        "6s": $(tds.eq(6)).text(),
        "SR": $(tds.eq(7)).text()
  });
});
Tabnine | Edit | Test | Explain | Document | Ask
secondInningsRows.each((index, element) => {
  var tds = $(element).find('td');
   battingSummary.push({
        "match": matchInfo,
        "teamInnings": team2,
        "battingPos": index+1,
        "batsmanName": $(tds.eq(0)).find('a > span > span').text().replace('', ''),
        "dismissal": $(tds.eq(1)).find('span > span').text(),
        "runs": $(tds.eq(2)).find('strong').text(),
        "balls": $(tds.eq(3)).text(),
        "4s": $(tds.eq(5)).text(),
        "6s": $(tds.eq(6)).text(),
        "SR": $(tds.eq(7)).text()
  });
});

return {"battingSummary": battingSummary}
```

Fig 6.1

6.2 Data Preprocessing

A detailed explanation of how raw data was cleaned and transformed using Python and Pandas, ensuring that it is ready for analysis and visualization. This includes handling missing values, data normalization, and feature engineering.

[1]:
```
import pandas as pd
import json
from collections import defaultdict
```

Match Details

[2]:
```
with open('T20_Data/MatchSummary.json') as f:
    data = json.load(f)
df_match = pd.DataFrame(data[0]['matchSummary'])
df_match.head()
```

[2]:

	Team 1	Team 2	Winner	Margin	Ground	Match Date	match_id
0	India	South Africa	India	7 runs	Bridgetown	Jun 29, 2024	T20I # 2729
1	England	India	India	68 runs	Providence	Jun 27, 2024	T20I # 2724
2	Afghanistan	South Africa	South Africa	9 wickets	Tarouba	Jun 26, 2024	T20I # 2723
3	Afghanistan	Bangladesh	Afghanistan	8 runs	Kingstown	Jun 24, 2024	T20I # 2722
4	Australia	India	India	24 runs	Gros Islet	Jun 24, 2024	T20I # 2721

[3]:
```
df_match.shape
```

[3]:
```
(52, 7)
```

[4]:
```
df_match.to_csv('T20_Data_csv/match_summary.csv', index = False)
```

Batting Summary

[5]:
```
with open('T20_Data/BattingSum.json') as f:
    data = json.load(f)

    all_records = []

    for rec in data:
        all_records.extend(rec['battingSummary'])

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

[5]:

	match	teamInnings	battingPos	batsmanName	out/not_out	R	B	M	4s	6s	SR	match_id
0	India vs South Africa	India	1	Rohit Sharma	out	9	5	7	2	0	180.00	T20I # 2729
1	India vs South Africa	India	2	Virat Kohli	out	76	59	87	6	2	128.81	T20I # 2729
2	India vs South Africa	India	3	Rishabh Pant	out	0	2	2	0	0	0.00	T20I # 2729
3	India vs South Africa	India	4	Suryakumar Yadav	out	3	4	9	0	0	75.00	T20I # 2729
4	India vs South Africa	India	5	Axar Patel	out	47	31	38	1	4	151.61	T20I # 2729

[6]:
```
df_batting.to_csv('T20_Data_csv/batting_summary.csv', index = False)
```

Bowling Summary

```
[7]: with open('T20_Data/BowlingSummary.json') as f:
         data = json.load(f)
         all_records = []

         for rec in data:
             all_records.extend(rec['bowlingSummary'])

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

[7]:

	match	bowlingTeam	bowlerName	O	M	R	W	ECON	0s	4s	6s	WD	NB	match_id
0	India vs South Africa	South Africa	Marco Jansen	4.0	0	49	1.0	12.25	4	5	2	1	1	T20I # 2729
1	India vs South Africa	South Africa	Keshav Maharaj	3.0	0	23	2.0	7.66	10	3	1	0	0	T20I # 2729
2	India vs South Africa	South Africa	Kagiso Rabada	4.0	0	36	1.0	9.00	8	2	2	2	0	T20I # 2729
3	India vs South Africa	South Africa	Aiden Markram	2.0	0	16	NaN	8.00	1	0	1	0	0	T20I # 2729
4	India vs South Africa	South Africa	Anrich Nortje	4.0	0	26	2.0	6.50	7	2	0	2	0	T20I # 2729

```
[8]: df_bowling.to_csv('T20_Data_csv/bowling_summary.csv', index = False)
```

Player Info

```
[9]: with open('T20_Data/Player_Info.json') as f:
         data = json.load(f)
     df_info = pd.DataFrame(data)
     df_info.head()
```

[9]:

	team	name	role	age	batting_style	bowling_style	image_url
0	Afghanistan	Azmatullah Omarzai	Allrounder	24y 38d	Right hand Bat	Right arm Medium fast	https://images.icc-cricket.com/image/upload/t_...
1	Afghanistan	Fareed Ahmad	Bowler	29y 265d	Left hand Bat	Left arm Fast medium	https://images.icc-cricket.com/image/upload/t_...
2	Afghanistan	Fazalhaq Farooqi	Bowler	23y 222d	Right hand Bat	Left arm Fast medium	https://images.icc-cricket.com/image/upload/t_...
3	Afghanistan	Gulbadin Naib	Batting Allrounder	32y 332d	Right hand Bat	Right arm Medium fast	https://images.icc-cricket.com/image/upload/t_...
4	Afghanistan	Hazratullah Zazai	Opening Batter	26y 39d	Left hand Bat	Slow Left arm Orthodox	https://images.icc-cricket.com/image/upload/t_...

```
[10]: df_info.to_csv('T20_Data_csv/Player_Info.csv', index = False)
```

Fig 6.2

6.3 Building Dynamic Models and parameters

The image provides an organized list of various cricket statistics measures along with their respective descriptions, DAX formulas, and the table names from which the data is derived. This table is essential for understanding and calculating cricket performance metrics dynamically using DAX (Data Analysis Expressions) in tools like Power BI.

Measures:

Sno	Measures	Description / Purpose	DAX FORMULA	TABLE
1	Total Runs	Total number of runs scored by the batsman	Total Runs = SUM(batting_summary[runs])	batting_summary
2	Total Innings Batted	Total number of innings a batsman got a chance to bat	Total Innings Batted = COUNT(batting_summary[match_id])	batting_summary
3	Total Innings Dismissed	To find the number of innings batsman got out	SUM(batting_summary[out])	batting_summary
4	**Batting Average**	**Average runs scored in an innings**	**Batting Avg = DIVIDE([Total Runs],[Total Innings Dismissed],0)**	**batting_summary**
5	Total balls Faced	Total number of balls faced by the batsman	total balls faced = SUM(batting_summary[balls])	batting_summary
6	**Strike Rate**	**No of runs scored per 100 balls**	**Strike rate = DIVIDE([Total Runs],[total balls faced],0)*100**	**batting_summary**
7	Batting Position	Batting position of a player	Batting Position = ROUNDUP(AVERAGE(batting_summary[batting_pos]),0)	batting_summary
8	Boundary %	Percentage of boundaries scored by the Batsman	Boundary % = DIVIDE(SUM(batting_summary[Boundary runs]),[Total Runs],0)	batting_summary
9	Avg. balls Faced	Average balls faced by the batter in an innings	AVERAGE(batting_summary[balls])	batting_summary
10	Wickets	Total number of wickets taken by a bowler	wickets = SUM(bowling_summary[wickets])	bowling_summary
11	balls Bowled	Total number of balls bowled by the bowler	balls Bowled = SUM(bowling_summary[balls])	bowling_summary
12	Runs Conceded	Total runs conceded by the bowler	Runs Conceded = SUM(bowling_summary[runs])	bowling_summary
13	**Bowling Economy**	**Average number of runs conceded in an over**	**Economy = DIVIDE([Runs Conceded], ([balls Bowled]/6),0)**	**bowling_summary**
14	**Bowling Strike Rate**	**Number of balls bowled per wicket**	**Bowling Strike Rate = DIVIDE([balls Bowled], [wickets],0)**	**bowling_summary**
15	Bowling Average	No. of runs allowed per wicket	Bowling Average = DIVIDE([Runs Conceded],[wickets],0)	bowling_summary
16	Total Innings Bowled	Total number of innings bowled by a bowler	Total Innings Bowled = DISTINCTCOUNT(bowling_summary[match_id])	bowling_summary
17	Dot Ball %	Percentage of dot balls bowled by a bowler	Dot ball % = DIVIDE(SUM(bowling_summary[zeros]), SUM(bowling_summary[balls]),0)	bowling_summary
18	Player Selection	To understand if a player is selected or not	Player Selection = if(ISFILTERED(players_Info[name]),"1","0")	
19	Display Text	To display a text of no player is selected	Display Text = if([Player Selection] = "1", " ","Select Player(s) by clicking the player's name to see their individual or combined strength.")	
20	Color Callout Value	To display a value only when a player is selected	Color Callout Value = if([Player Selection]="0", "#D0CF1D","#1D1D2E")	

Calculated Columns

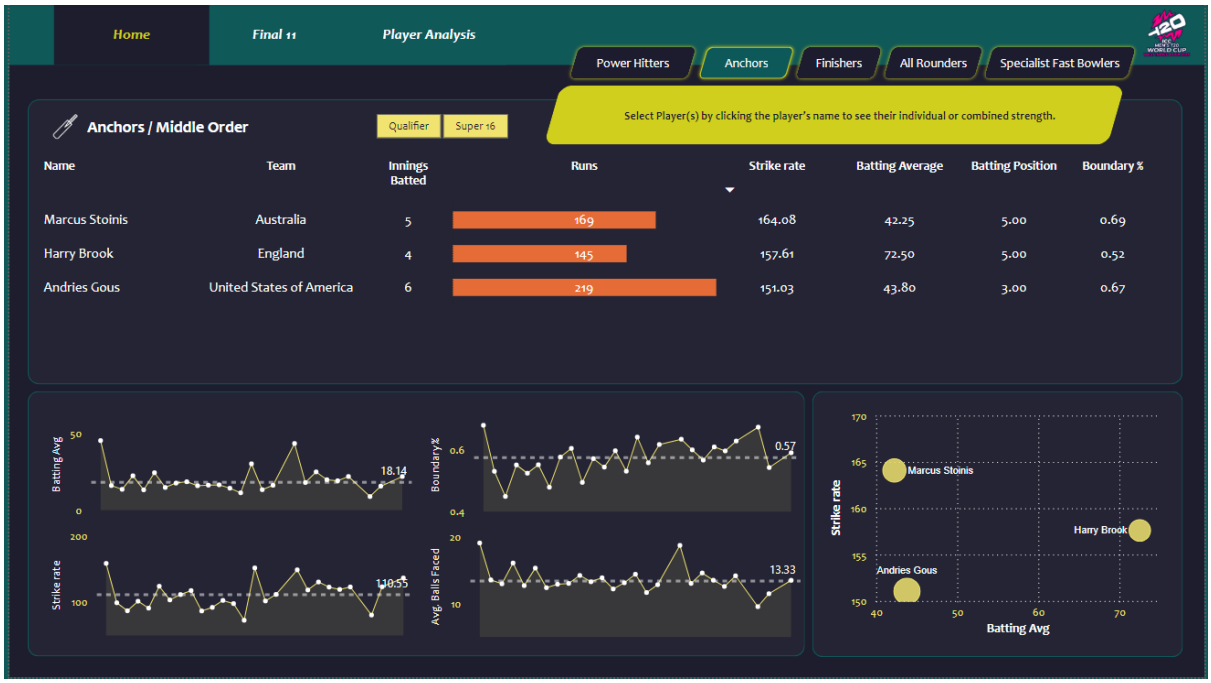
Sno.	Calculated Column Name	Description / Purpose	DAX formula	Table
1	boundary runs	to find the total number of runs scored by hitting fours and sixes	boundary runs = batting_summary[fours]*4 + batting_summary[sixes]*6	batting_summary
2	Boundary runs bowling	to find the total number of runs conceded by bowlers in boundaries	Boundary runs = bowling_summary[fours]*4 +bowling_summary[Sixes]*6	bowling_summary

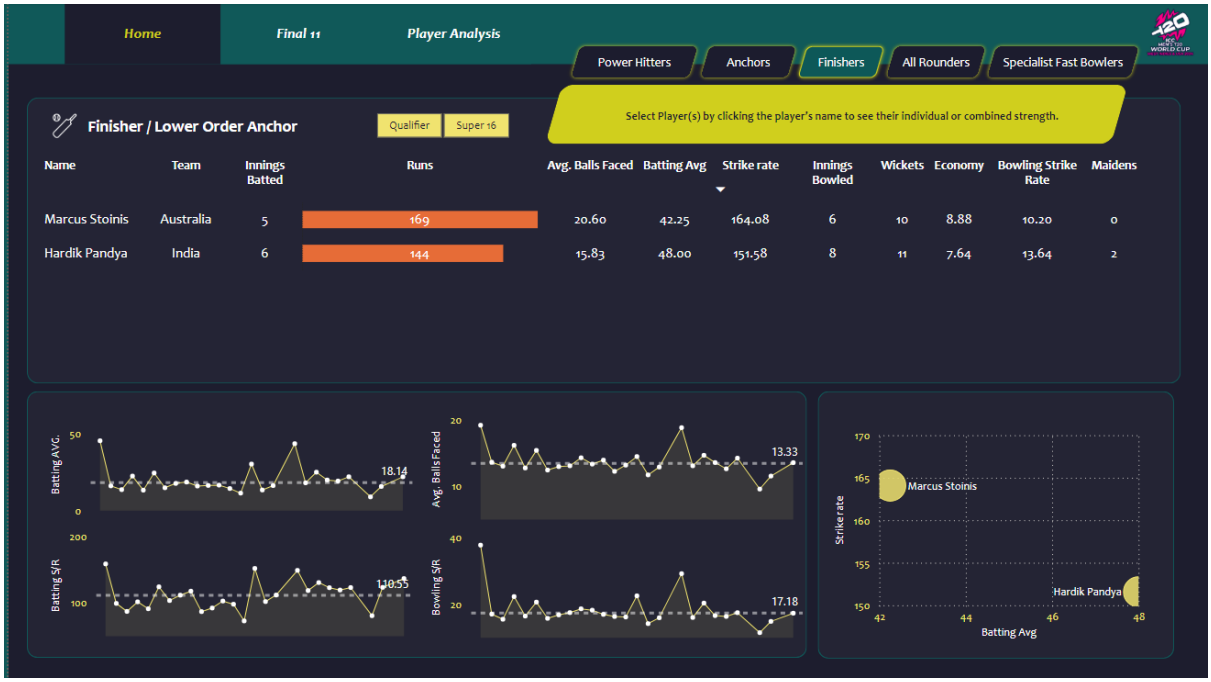
3 Custom Batting Order	To assign the batting order to potential final 11	SWITCH(TRUE(), players_Info[name] = "Rohit Sharma",1, players_Info[name] = "Rahmanullah Gurbaz",2, players_Info[name] = "Nicholas Pooran",3, players_Info[name] = "Suryakumar Yadav",4, players_Info[name] = "Marcus Stoinis" ,5, players_Info[name] = "Hardik Pandya" ,6, players_Info[name] = "Axar Patel" ,7, players_Info[name] = "Rashid Khan" ,8, players_Info[name] = "Jasprit Bumrah" ,9, players_Info[name] = "Arshdeep Singh" ,10, players_Info[name] = "Fazalhaq Farooqi" ,11, players_Info[name] = "Anrich Nortje" ,12)	players_Info

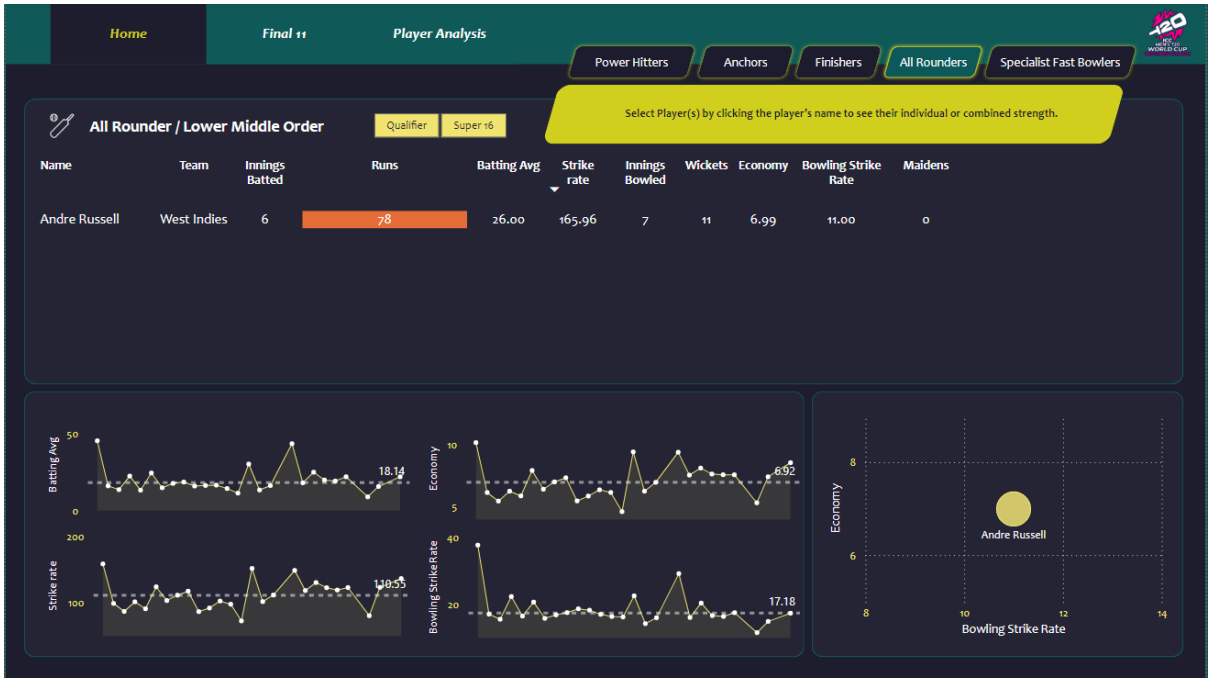
Fig 6.3

Chapter 7: SCREENSHOTS OF PROJECT OUTPUTS

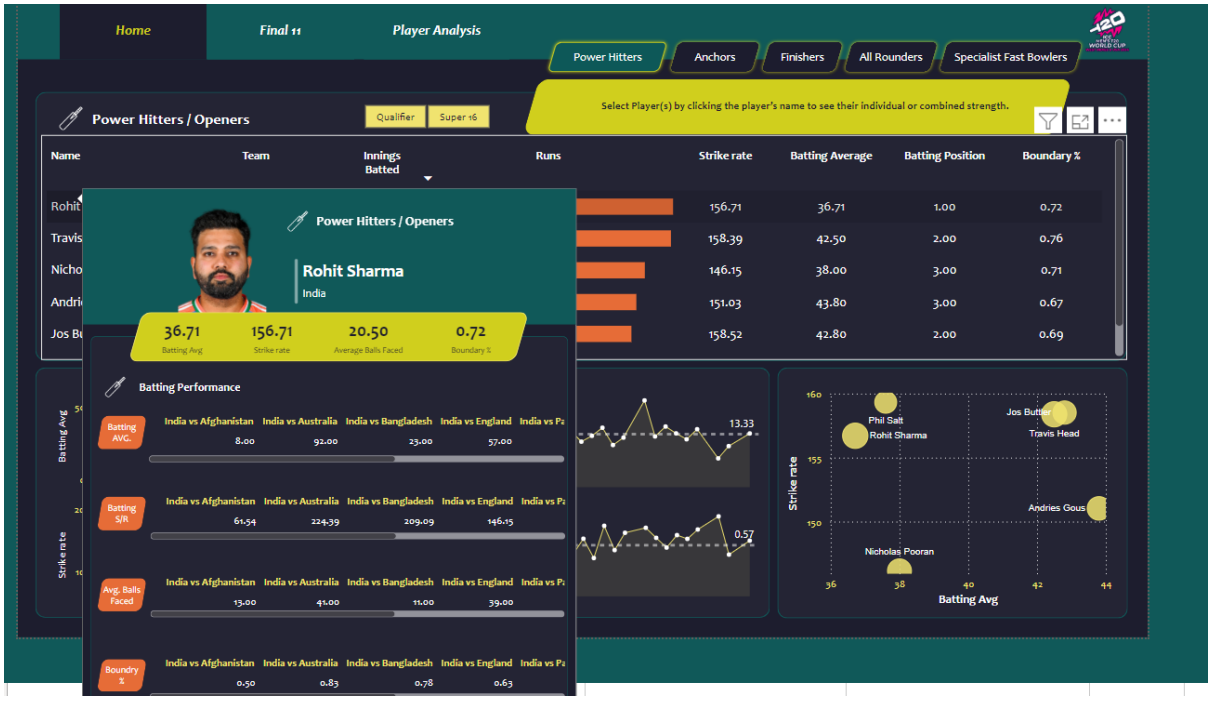
This chapter contains screenshots of the PowerBI dashboard and visualizations, showcasing the results and how insights into player and team performances were effectively presented.



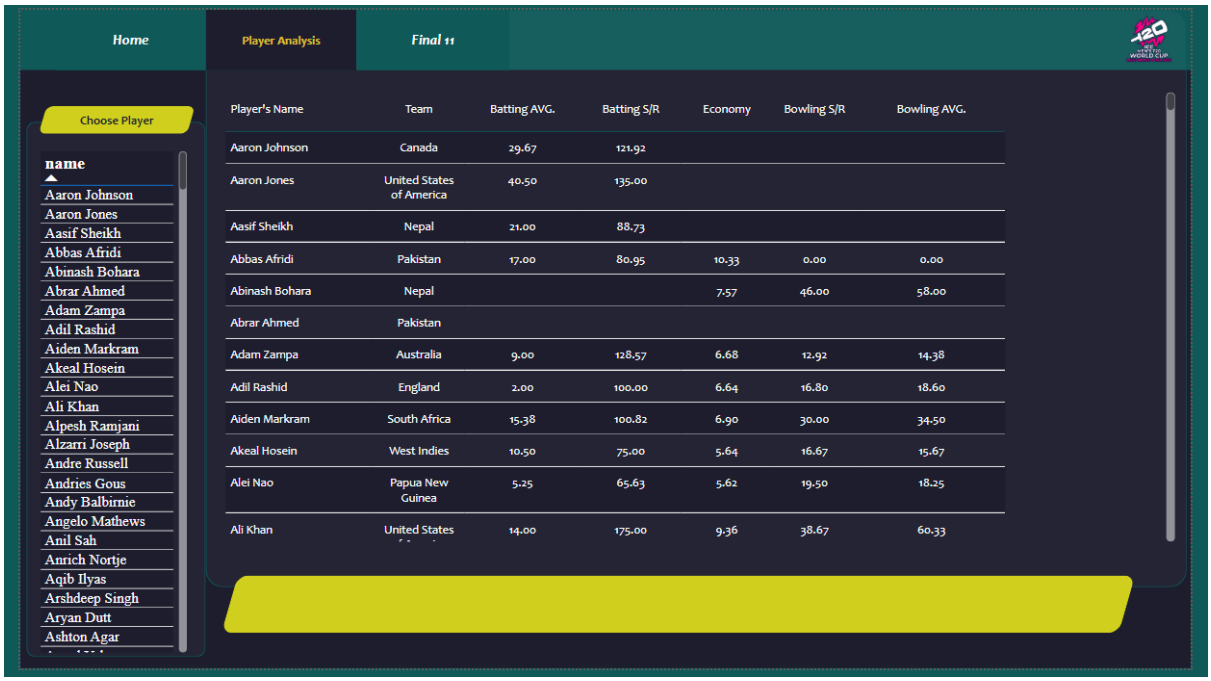

Home
Final 11
Player Analysis
Power Hitters
Anchors
Finishers
All Rounders
Specialist Fast Bowlers
Finisher / Lower Order Anchor
Qualifier
Super 16
Select Player(s) by clicking the player's name to see their individual or combined strength.
Name Team Innings Batted Runs Avg. Balls Faced Batting Avg Strike rate Innings Bowled Wickets Economy Bowling Strike Rate Maidens
Marcus Stoinis Australia 5 169 20.60 42.25 164.08 6 10 8.88 10.20 0
Hardik Pandya India 6 144 15.83 48.00 151.58 8 11 7.64 13.64 2
Batting AVG.
Avg. Balls Faced
Batting SR
Bowling SR
Strike rate
Batting Avg
Marcus Stoinis
Hardik Pandya


Home
Final 11
Player Analysis
Power Hitters
Anchors
Finishers
All Rounders
Specialist Fast Bowlers
All Rounder / Lower Middle Order
Qualifier
Super 16
Select Player(s) by clicking the player's name to see their individual or combined strength.
Name Team Innings Batted Runs Batting Avg Strike rate Innings Bowled Wickets Economy Bowling Strike Rate Maidens
Andre Russell West Indies 6 78 26.00 165.96 7 11 6.99 11.00 0
Batting Avg
Economy
Strike rate
Bowling Strike Rate
Andre Russell


Home
Final 11
Player Analysis
Power Hitters
Anchors
Finishers
All Rounders
Specialist Fast Bowlers
Specialist Fast Bowlers / Tail End
Qualifier
Super 16
Select Player(s) by clicking the player's name to see their individual or combined strength.
Name
Team
Innings Bowled
Runs Conceded
Wickets
Economy
Bowling Average
Bowling Strike Rate
Dot ball %
Maidens
Fazalhaq Farooqi
Afghanistan
8
160
17
6.32
9.41
8.94
0.57
0
Anrich Nortje
South Africa
9
201
15
5.74
13.40
14.00
0.52
0
Jasprit Bumrah
India
8
124
15
4.18
8.27
11.87
0.62
2
Rashid Khan
Afghanistan
8
179
14
6.17
12.79
12.43
0.47
0
Kagiso Rabada
South Africa
9
195
13
6.29
15.00
14.31
0.52
2
Bowling Average
19.82
Economy
6.92
Bowling Strike Rate
17.18
Dot ball%
0.47
Andre Russell
Kuldeep Yadav
Fazalhaq Farooqi
Naveen-ul-Haq
Kagiso Rabada
Rashid Khan
Anrich Nortje
Roston Chase
Jasprit Bumrah
Bowling Strike Rate


Home
Final 11
Player Analysis
Power Hitters
Anchors
Finishers
All Rounders
Specialist Fast Bowlers
Power Hitters / Openers
Qualifier
Super 16
Select Player(s) by clicking the player's name to see their individual or combined strength.
Name
Team
Innings Batted
Runs
Strike rate
Batting Average
Batting Position
Boundary %
Rohit
Travis
Nicho
Andri
Jos Bu
156.71
36.71
1.00
0.72
158.39
42.50
2.00
0.76
146.15
38.00
3.00
0.71
151.03
43.80
3.00
0.67
158.52
42.80
2.00
0.69
Power Hitters / Openers
Rohit Sharma
India
36.71
Batting Avg
156.71
Strike rate
20.50
Average Balls Faced
0.72
Boundary %
Batting Performance
Batting AVG.
Batting S/R
Avg. Balls Faced
Boundry %
India vs Afghanistan
India vs Australia
India vs Bangladesh
India vs England
India vs Pa
8.00
92.00
23.00
57.00
61.54
224.39
209.09
146.15
13.00
41.00
11.00
39.00
0.50
0.83
0.78
0.63
13.33
0.57
Phil Salt
Rohit Sharma
Jos Buttler
Travis Head
Andries Gous
Nicholas Pooran
Strike rate
Batting Avg


Home
Final 11
Player Analysis
Select your Final 11
Search
Andre Russell
Hardik Pandya
Marcus Stoinis
Rohit Sharma
Anrich Nortje
Arshdeep Singh
Axar Patel
Fazalhaq Farooqi
Jasprit Bumrah
Nicholas Pooran
Rahmanullah Gur...
Rashid Khan
Suryakumar Yadav
Player's Name | Team | Batting AVG. | Batting S/R | Economy | Bowling S/R | Bowling AVG. | Sum of Custom Batting Order | wickets
Rohit Sharma | India | 36.71 | 156.71 | | | | 1 |
Rahmanullah Gurbaz | Afghanistan | 35.13 | 124.34 | | | | 2 |
Nicholas Pooran | West Indies | 38.00 | 146.15 | | | | 3 |
Suryakumar Yadav | India | 28.43 | 135.37 | | | | 4 |
Marcus Stoinis | Australia | 42.25 | 164.08 | 8.88 | 10.20 | 15.10 | 5 | 10
Hardik Pandya | India | 48.00 | 151.58 | 7.64 | 13.64 | 17.36 | 6 | 11
Axar Patel | India | 23.00 | 139.39 | 7.86 | 14.67 | 19.22 | 7 | 9
Rashid Khan | Afghanistan | 11.40 | 123.91 | 6.17 | 12.43 | 12.79 | 8 | 14
Jasprit Bumrah | India | 0.00 | 0.00 | 4.18 | 11.87 | 8.27 | 9 | 15
Arshdeep Singh | India | 12.00 | 75.00 | 7.17 | 10.59 | 12.65 | 10 | 17
Fazalhaq Farooqi | Afghanistan | | | 6.32 | 8.94 | 9.41 | 11 | 17
Anrich Nortje | South Africa | 0.00 | 100.00 | 5.74 | 14.00 | 13.40 | 12 | 15
Team Performance
31.30 Batting Avg
141.04 Strike rate
17.31 Average Balls Faced
12.91 Bowling Avg.
11.83 Bowling S/R
6.54 Economy
0.51 Dot ball %


Home
Player Analysis
Final 11
Choose Player
name
Aaron Johnson
Aaron Jones
Aasif Sheikh
Abbas Afridi
Abinash Bohara
Abrar Ahmed
Adam Zampa
Adil Rashid
Aiden Markram
Akeal Hosein
Alei Nao
Ali Khan
Alpesh Ramjani
Alzarri Joseph
Andre Russell
Andries Gous
Andy Balbirnie
Angelo Mathews
Anil Sah
Anrich Nortje
Aqib Ilyas
Arshdeep Singh
Aryan Dutt
Ashton Agar
Player's Name | Team | Batting AVG. | Batting S/R | Economy | Bowling S/R | Bowling AVG.
Aaron Johnson | Canada | 29.67 | 121.92 | | | |
Aaron Jones | United States of America | 40.50 | 135.00 | | | |
Aasif Sheikh | Nepal | 21.00 | 88.73 | | | |
Abbas Afridi | Pakistan | 17.00 | 80.95 | 10.33 | 0.00 | 0.00
Abinash Bohara | Nepal | | | 7.57 | 46.00 | 58.00
Abrar Ahmed | Pakistan | | | | | |
Adam Zampa | Australia | 9.00 | 128.57 | 6.68 | 12.92 | 14.38
Adil Rashid | England | 2.00 | 100.00 | 6.64 | 16.80 | 18.60
Aiden Markram | South Africa | 15.38 | 100.82 | 6.90 | 30.00 | 34.50
Akeal Hosein | West Indies | 10.50 | 75.00 | 5.64 | 16.67 | 15.67
Alei Nao | Papua New Guinea | 5.25 | 65.63 | 5.62 | 19.50 | 18.25
Ali Khan | United States | 14.00 | 175.00 | 9.36 | 38.67 | 60.33


Power Hitters / Openers
Aaron Johnson
Afghanistan
18.14
Batting Avg
110.55
Strike rate
13.33
Average Balls Faced
0.57
Boundary %
Batting Performance
Batting AVG.
Afghanistan vs Australia
15.31
Afghanistan vs Bangladesh
14.00
Afghanistan vs New Zealand
13.50
Batting S/R
Afghanistan vs Australia
106.52
Afghanistan vs Bangladesh
86.73
Afghanistan vs New Zealand
105.37
Avg. Balls Faced
Afghanistan vs Australia
12.78
Afghanistan vs Bangladesh
13.29
Afghanistan vs New Zealand
11.39
Boundry %
Afghanistan vs Australia
0.56
Afghanistan vs Bangladesh
0.48
Afghanistan vs New Zealand
0.56

Fig 7.1

Chapter 8: TESTING

8.1 Types of Testing Conducted

- **Unit Testing**: Ensured that every script and every function behaves as it should.
- **Integration Testing**: Verified data sent between modules is correctly and seamlessly integrated with PowerBI.
- **User Acceptance Testing (UAT)**: Conducted to ensure the dashboard meets the user requirements and provides intuitive insights.

Test Cases:

1. Data Ingestion Test Cases

I. **Verify Data Import**:

- **Test Case**: Ensure that raw data is correctly imported into the Python environment from various sources (e.g., CSV files, databases).
- **Failed Case**: Certain CSV files failed to import due to incorrect file encoding.

II. **Validate Data Integrity**:

- **Test Case**: Check that the imported data maintains integrity (e.g., no missing values in critical columns).
- **Failed Case**: Found few data's with few missing values and a few with some unwanted characters.

2. Data Processing and Transformation Test Cases

I. **Verify Data Cleaning**:

- **Test Case**: Ensure that data cleaning scripts remove or correct erroneous data without losing critical information.

II. **Validate Data Transformation**:

- **Test Case**: Verify that data transformation processes (e.g., aggregations, calculations) are performed correctly.

III. **Check Data Enrichment**:

- **Test Case**: Ensure that additional data sources are correctly integrated and enrich the existing dataset.

3. Data Export and Loading Test Cases

I. **Verify Data Export**:

- **Test Case**: Ensure that processed data is correctly exported to the desired format (e.g., CSV, Excel) for loading into Power BI.

II. **Validate Data Loading**:

- **Test Case**: Ensure that Power BI correctly loads the exported data without errors.

4. Data Visualization and Reporting Test Cases

I. **Verify Dashboard Accuracy**:

- **Test Case**: Ensure that Power BI dashboards accurately reflect the underlying data.

II. **Validate Interactive Features**:

- **Test Case**: Check that interactive features (e.g., filters, slicers) on Power BI dashboards work correctly.

III. **Ensure Visualization Clarity**:

- **Test Case**: Verify that visualizations are clear, well-labeled, and easy to understand.

5. Performance Test Cases

I. **Check Data Processing Speed**:

- **Test Case**: Ensure that Python scripts process data within acceptable time limits.

II. **Verify Dashboard Performance**:

- **Test Case**: Ensure that Power BI dashboards load and update efficiently, even with large datasets.

6. Security Test Cases

I. **Ensure Data Confidentiality**:

- **Test Case**: Verify that sensitive data is securely stored and not exposed in visualizations.

8.2 Testing Results and Feedback

The results of testing indicated that the system performed efficiently with no major problem at any point during testing phases. Testees were also of the opinion that it is an efficient dashboard through which one gets insight into things, though some proposed adding more interactivity, like filtering.

Chapter 9: LIMITATIONS AND FUTURE SCOPE

9.1 Limitations of the Current Analysis

Although the system is good for insight into the current cricket performance, it has limitations that restrict its overall efficacy.

1. **Data Dependency on External Sources**: The system depends mostly on external sources like ESPNcricinfo and Kaggle, which rarely update data; in such cases, the information they provide cannot be relied upon at times.
2. **Limited Real-Time Data Integration**: The current system is not capable of adding real-time match data and therefore cannot perform real-time match analysis. It thereby fails to provide insight throughout active tournaments.
3. **Scalability Concerns**: As data increases, particularly with additional tournaments and player data, there could be a drop in performance unless optimized techniques are used when dealing with large volumes of data.
4. **Lack of Predictive Analytics**: The system does after-match analysis but never attempts to integrate predictive models that predict on outcomes, player performance, and even team strategies based on historical data.
5. **Simple Visualization**: PowerBI supports interactivity for dashboards but is very far from the perfect visualisations that should be provided-such as predictive trend line or sentiment analysis.

9.2 Suggestions for Future Work

There are areas wherein improvement and potential can be noted for the cricket analytics system to make it more efficient and functional.

1. **Real-Time Analytics**:
    - **Live Data Integration**: Including real-time data streams, for example, scores and the statistics of the players of the match, would provide live analysis of matches with current insights.

- **Live Dashboards**: The live dashboards on PowerBI can potentially depict actual performance trends during a game, thus allowing immediate changes in strategy.

2. **Predictive Analytics**:
    - **Player Performance Prediction**: Using machine learning algorithms such as linear regression, decision trees, or neural networks, the system can predict further performance in light of past trends, environment, and player conditions.
    - **Match Outcome Prediction**: The algorithms of machine learning can predict the results of the future matches based on team performance, venue information, and condition of players.
3. **Advanced Data Visualizations**:
    - **Heat Maps and Geospatial Analysis**: Using advanced visualization tools to generate heat maps or geographical analysis could give insights into player movements, scoring patterns, and team strategies.
    - **Advanced Charts**: Implementing more complex charts such as radar charts for player comparisons or Sankey diagrams for team strategies could enrich the user experience.
4. **Sentiment Analysis**:
    - **Fan Sentiment Analysis**: Scraping social media platforms and analyzing fan sentiment could provide insights into public opinion, helping analysts understand how external factors (like fan pressure) may influence team strategies.
    - **Player Sentiment**: Incorporating sentiment analysis of player interviews, press conferences, and social media posts could gauge player morale and mental state, which can be crucial for match predictions.
5. **Incorporating Wearable Data**:
    - **Player Fitness Monitoring**: Incorporating data from wearable devices (such as fitness trackers) could provide a deeper analysis of player fitness levels, fatigue, and recovery, influencing performance predictions.

- **Biomechanical Analysis**: Including data on players' movements, shot technique, and bowling action could help assess the physical efficiency and injury risk, contributing to player management strategies.

6. **Extended Tournaments and Data Sources**:
   - **Cross-Tournament Analysis**: Expanding the scope to include data from other tournaments, like IPL, BBL, and international T20 series, would allow for a broader analysis of player performance across different conditions and formats.
   - **Incorporation of Additional Data**: Incorporating more detailed data, such as player injuries, weather conditions, and pitch reports, would make the analysis more comprehensive.
7. **Cloud-Based Solution**:
   - **Cloud Integration**: Moving the project to the cloud (e.g., AWS, Azure) would make the system more scalable, allowing for faster data processing and real-time analytics on larger datasets. This would also help in providing easier access to dashboards and reports globally.
8. **Interactive Decision Support System (DSS)**:
   - **Incorporation of AI-Based Recommendations**: Creating an AI-based decision support system where coaches and team strategists can input real-time match conditions and get tactical recommendations would be highly beneficial. For example, recommending the best bowling attack or optimal batting order based on the current match data.

By implementing these future enhancements, the cricket analytics system can be elevated from a basic performance tracking tool to a dynamic, real-time, predictive analytics platform that can significantly assist teams, analysts, and cricket enthusiasts in making informed decisions.

Chapter 10: REFERENCES

ICC Men's T20 World Cup, 2024 team match results Records | ESPNcricinfo

ICC Men T20 World Cup 2024 Full Scorecards & Stats