**An Project Report**

**on**

**“Cricket Uncovered”**

**By**

**Aum Vashi (2203396160383)**

**Preet Desai (2203396160060)**

**Kirtan Chauhan (2203396160047)**

**Harikrushna Chauhan (2203396160046)**

**(5th sem Computer Engineering Department)**

**Internal Guide**

**Prof. Tirthesh Gajjar**

**&**

**Prof. Raviraj Pandya**

**(Computer Engineering Department)**

**Submitted To**



****

**PARUL INSTITUTE OF ENGINEERING & TECHNOLOGY(DIPLOMA STUDIES)**

**LIMDA,VAGHODIA,VADODARA**

**CERTIFICATE**

This is to Certify that UDP report by **AUM VASHI (2203396160383), PREET DESAI (2203396160060), KIRTAN CHAUHAN (2203396160047)** and **HARIKRUSHNA CHAUHAN (2203396160046)** of Computer Engineering Department of **PARUL INSTITUTE OF ENGINEERING & TECHNOLOGY (DIPLOMA STUDIES), LIMDA** is the record of work carried out by them under our supervision and guidance. The work submitted has in our opinion reached a level required for being accepted for examination.

|  |  |
| --- | --- |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **PRINCIPAL**  **Dr. Ruchi Shrivastava**  Principal  PIET(DS), LIMDA | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **HOD**  **Prof. Hetal Bhaidasana**  Department of Computer Engineering  PIET(DS), LIMDA |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **GUIDE**  **Prof. Tirthesh Gajjar**  Department of Computer Engineering  PIET(DS), LIMDA | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **GUIDE**  **Prof. Raviraj Pandya**  Department of Computer Engineering  PIET(DS), LIMDA |

**ACKNOWLEDGEMENT**

We are very thankful to God whose blessings helped us a lot to make our effort successful. Thanks to our Parents who inspired us to go ahead in our work. We also like to thank our Principal **Dr. Ruchi Shrivastava** and Head of Department **Prof. Hetal Bhaidasna** for the moral support. It is our pleasure to take this opportunity to thank all those who directly or indirectly helped us in our Project. We are extremely grateful to our internal Guide **Prof. Tirthesh Gajjar** and **Prof. Raviraj Pandya** for there excellent guidance, valuable suggestion, and encouragement to carry out this project, valuable suggestions and guidance to bring out the best from this project. Finally, we would like to express our deep sense of gratitude to all our classmates and many other friends of us who have supported and enriched us by sharing their ideas, and sorting out our doubts having discussion with us. We thank one and all.

**ABSTRACT**

In today’s data-driven world, sports analytics plays a pivotal role in enhancing both performance and audience engagement. "Cricket Uncovered" is a data analysis project aimed at providing deep insights into cricket player performances throughout a tournament. This project utilizes data extracted from ESPN using Bright Data and processed using various tools like Pandas, Power Query, and Power BI. The primary objective of the project is to showcase the best players in different roles—opener, middle-order batsman, finisher, bowler—and determine the best 11 players of the tournament through advanced logic and key performance indicators such as strike rate, batting average, and economy rate.

The data is first cleaned and prepared using Python’s Pandas library, ensuring accuracy and consistency. Power Query then transforms the dataset, structuring it for detailed analysis. Data Analysis Expressions (DAX) are employed to model the data, allowing for the calculation of critical statistics and player rankings. These insights are then visualized using Power BI, which offers dynamic, interactive dashboards. The final output is a website created using HTML and CSS, where users can explore the dashboard and analyze player performance across multiple roles.

The project goes beyond basic statistics by focusing on role-specific contributions and providing a holistic comparison of players. This allows for a more nuanced understanding of how players perform under different conditions. The flexibility of Power BI’s filters enhances the user experience, enabling cricket fans, analysts, and coaches to interact with the data and derive meaningful insights effortlessly.

"Cricket Uncovered" lays the foundation for further developments, including the integration of predictive models and real-time data analysis, thereby providing a powerful tool for cricket analytics in the future.

**TABLE OF CONTENTS**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No.** |  | **Chapter Name** | | | **Page No.** |  |
|  |  | | | |  |  |
| **1.** | **INTRODUCTION** | | | |  |  |
|  | 1.1 | Introduction | | | 1 |  |
|  | 1.2 | Existing System | | | 2 |  |
|  | 1.3 | Proposed Work | | | 3 |  |
|  | 1.4 | Advantages | | | 4 |  |
|  | 1.5 | Scope | | | 5 |  |
|  |  | | | |  |  |
| **2.** | **SYSTEM REQUIREMENT STUDY** | | | |  |  |
|  | 2.1 | Feasibility Study | | | 6 |  |
|  | 2.2 | Requirements of Project | | | 7 |  |
|  | 2.3 | Tools & Technology | | | 8 |  |
|  |  | | | |  |  |
| **3.** | **DIAGRAMS** | | | |  |  |
|  | 3.1 | ER Diagram | | | 9 |  |
|  | 3.2 | Use case Diagrams | | | 11 |  |
|  | 3.3 | DFD Diagrams | | | 13 |  |
|  | 3.4 | Activity Diagrams | | | 14 |  |
|  | 3.5 | Class Diagram | | | 16 |  |
|  |  | |  |
| **4.** | **IMPLEMENTATION** | | | |  |  |
|  | 4.1 | Data Extraction | | | 17 |  |
|  | 4.2 | Data Cleaning | | | 17 |  |
|  | 4.3 | Data Transformation | | | 17 |  |
|  | 4.4 | Data Modeling | | | 18 |  |
|  | 4.5 | Data Visualization | | | 18 |  |
|  | 4.6 | Website Development | | | 18 |  |
| **5.** | **CONCLUSION** | | | | 21 |  |
| **6.** | **FUTURE SCOPE** | | | |  |  |
|  | 6.1 | Predictive Analytics | | | 22 |  |
|  | 6.2 | Real-Time Data Integration | | | 22 |  |
|  | 6.3 | Expansion to Multiple Leagues | | | 22 |  |
|  | 6.4 | Enhanced User Interface | | | 22 |  |
| **7.** | **BIBLIOGRAPHY** | | | | 23 |  |

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| **1** | **Symbols of E-R Diagram** | 9 |
| **2** | **E-R Diagram** | 10 |
| **3** | **Symbols of Use Case Diagram** | 11 |
| **4** | **Use Case Diagram** | 12 |
| **5** | **DFD Diagram** | 13 |
| **6** | **Symbols of Activity Diagram** | 14 |
| **7** | **Activity Diagram** | 15 |
| **8** | **Class Diagram** | 16 |
| **9** | **Dashboard 1** | 19 |
| **10** | **Dashboard 2** | 19 |
| **11** | **Dashboard 3** | 20 |
| **12** | **Dashboard 4** | 20 |

**LIST OF FIGURES**

**Chapter-1 INTRODUCTION**

**1.1 Introduction**

Cricket is not just a sport but an emotion for millions across the globe. In this data-driven era, analyzing player performance has become a key factor in decision-making, whether for team selection, strategy formation, or improving individual skills. The vast amount of statistical data generated from each match—ranging from runs scored, balls faced, strike rates, bowling economy, and fielding performance—holds immense potential for deriving actionable insights. However, interpreting this data manually or through traditional means can be cumbersome, time-consuming, and prone to human error.

The "Cricket Uncovered" project aims to resolve these challenges by using modern data analysis techniques to deliver a comprehensive, dynamic, and interactive dashboard. This project provides insights into the performance of players across different roles—such as openers, middle-order batsmen, finishers, and bowlers—and selects the best 11 players of the tournament. The dashboard offers detailed visualizations and comparative statistics, enabling users to explore data and make informed conclusions about player performance.

By leveraging various technologies, such as Pandas for data cleaning, Power Query for data transformation, and Power BI for data visualization, "Cricket Uncovered" showcases how data can enhance the way we understand cricket. The dashboard is designed to be user-friendly and is accessible through a website created using HTML and CSS, allowing cricket enthusiasts, coaches, and analysts to interact with the data effortlessly.

In this fast-evolving digital landscape, cricket analytics is more critical than ever. Our project not only focuses on analyzing past performance but also lays the groundwork for future advancements, such as predictive modeling, real-time analytics, and integration of additional performance parameters.

**1.2 Existing System**

The traditional system of cricket data analysis has several limitations. Cricket boards, coaches, and enthusiasts primarily rely on static tables, charts, and lists provided by popular sports websites like ESPN Cricinfo. While these platforms do offer access to a vast repository of cricket data, their approach is largely limited to individual player statistics or match summaries, rather than a holistic analysis that compares players across roles or positions.

Existing systems often involve manual interpretation of data or simplistic automated rankings based on basic parameters such as total runs or wickets. For example, most systems calculate the "best" batsman or bowler based on aggregates without considering nuanced factors like strike rates, pressure scenarios, or the role-specific contributions of players (i.e., opener vs. middle-order). This leaves a gap in evaluating player effectiveness within specific contexts, which is crucial in modern-day cricket.

Moreover, there is a lack of dynamic, customizable tools that allow users to filter and explore data based on various conditions. While ESPN and other cricket platforms offer raw data, they fall short in delivering advanced analytical insights. Furthermore, these platforms do not integrate data visualization techniques that enable users to visualize performance trends over time or compare players interactively.

**1.3 Proposed Work**

To overcome the limitations of the existing systems, "Cricket Uncovered" introduces a comprehensive data analysis and visualization solution. By automating the data extraction process using Bright Data, our project ensures that a large dataset from ESPN can be regularly and efficiently pulled without manual intervention. This raw data is then cleaned using Python’s Pandas library, ensuring it is free from duplicates, errors, and inconsistencies.

Power Query, integrated with Power BI, transforms the cleaned data into a structured format suitable for analysis. With DAX (Data Analysis Expressions), we model the data to compute key performance indicators such as strike rate, batting average, and bowling economy. These metrics are then applied across different player roles (e.g., opener, middle-order, finisher, bowler) to derive insights and compare performances.

Finally, Power BI's advanced visualization capabilities allow us to present this data in an intuitive, interactive dashboard. The proposed work culminates in a web-based interface, where users can navigate the dashboard, filter data, and extract insights without needing any technical expertise.

**1.4 Advantages**

The "Cricket Uncovered" project offers several advantages over traditional systems:

**Automation:** Automated data extraction and cleaning reduce human intervention, errors, and delays, ensuring timely updates.

**Comprehensive Analysis:** The system offers role-specific performance analysis, allowing users to evaluate players based on their specific contributions (e.g., opener vs. finisher) rather than just aggregated statistics.

**Customizable Filters:** Users can explore the data through multiple filters (e.g., role, strike rate, average) to compare players dynamically.

**Interactive Visualization:** Power BI's advanced visual tools allow for an interactive user experience, where insights can be gained through dynamic charts and graphs.

**Accessible Interface:** The web-based interface ensures the dashboard is accessible to all, whether it’s for a casual fan or a team analyst.

**1.5 Scope**

The scope of this project is not limited to just one tournament or dataset. It has the potential to expand globally by incorporating data from multiple cricket leagues, including IPL, BBL, and international matches. Moreover, as technology advances, the system can evolve to include machine learning models that provide predictive analytics, such as forecasting a player’s future performance or team success based on past trends. Furthermore, with the integration of real-time data feeds, this system could provide live analytics during matches, adding a whole new dimension to the cricket-watching experience.

**Chapter: 2 SYSTEM REQUIREMENT STUDY**

**2.1 Feasibility Study**

**Technical Feasibility:**

From a technical standpoint, the tools and technologies selected for the "Cricket Uncovered" project are robust, scalable, and widely used in the data analytics domain. Bright Data offers an efficient way to scrape data from ESPN, while Python (Pandas) is an industry-standard for data cleaning and manipulation. Power Query, integrated with Power BI, is an excellent tool for data transformation, and DAX allows for complex data modeling. HTML and CSS are sufficient for creating a simple yet effective front-end interface.

**Economic Feasibility:**

This project is economically feasible, as it only requires access to open-source tools (Python, Pandas) and Power BI, which offers free versions for educational purposes. No additional costs are associated with hardware beyond a standard personal computer capable of running these tools. Additionally, using Bright Data for data extraction eliminates the need for expensive proprietary APIs for accessing cricket data.

**Operational Feasibility:**

From an operational perspective, the system is designed to be user-friendly and accessible. Power BI's intuitive interface ensures that users with minimal technical expertise can explore the data and extract insights. Furthermore, the web-based dashboard allows users to access it from anywhere, making it a versatile tool for coaches, analysts, and cricket fans alike.

**2.2 Requirements of Project**

**Hardware Requirements:**

Processor: Intel i5 or equivalent (or higher)

RAM: 8 GB or more for efficient data processing

Storage: At least 500 GB SSD to store datasets and project files

Internet: High-speed connection for data scraping and real-time dashboard updates

**Software Requirements:**

Operating System: Windows 10 or higher

Programming Languages: Python (for data extraction and cleaning)

Tools: Pandas, Power Query, DAX, Power BI, HTML, CSS

Other: Bright Data account for data extraction

**2.3 Tools & Technology**

**Bright Data:** Web scraping tool used to extract data from the ESPN website. It ensures that up-to-date and accurate match and player statistics are collected for analysis.

**Pandas:** Python's data manipulation library is used for cleaning the raw data. Pandas handles tasks like removing duplicates, filling missing values, and converting data types, making the dataset ready for transformation.

**Power Query:** Integrated with Power BI, Power Query handles the transformation of the cleaned data. It allows for filtering, sorting, and merging datasets, making the data structured and ready for analysis.

**DAX (Data Analysis Expressions):** DAX is used in Power BI for data modeling. It enables the creation of custom calculated fields such as player strike rates, averages, and rankings, helping to derive meaningful insights from raw data.

**Power BI:** Power BI is a powerful visualization tool used to create an interactive dashboard. It allows users to explore the data dynamically, compare players, and drill down into specific performance metrics.

**HTML/CSS:** These are used to build the front-end website where the Power BI dashboard is embedded. The website provides a seamless, accessible platform for users to interact with the data.

**Chapter: 3 DIAGRAM’S**

**3.1 ER Diagram**

The Entity-Relationship (ER) diagram illustrates the relationships between entities such as Player, Match, and Performance Metrics. It demonstrates how each entity interacts with the others and how data is connected throughout the system.

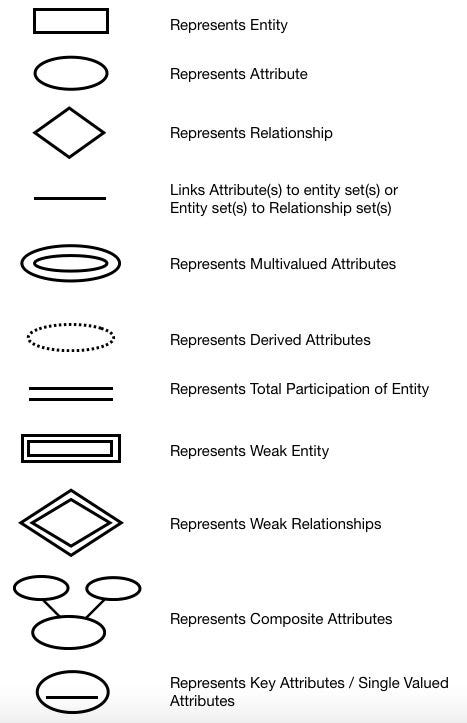


Figure 1 : Symbols of E-R Diagram

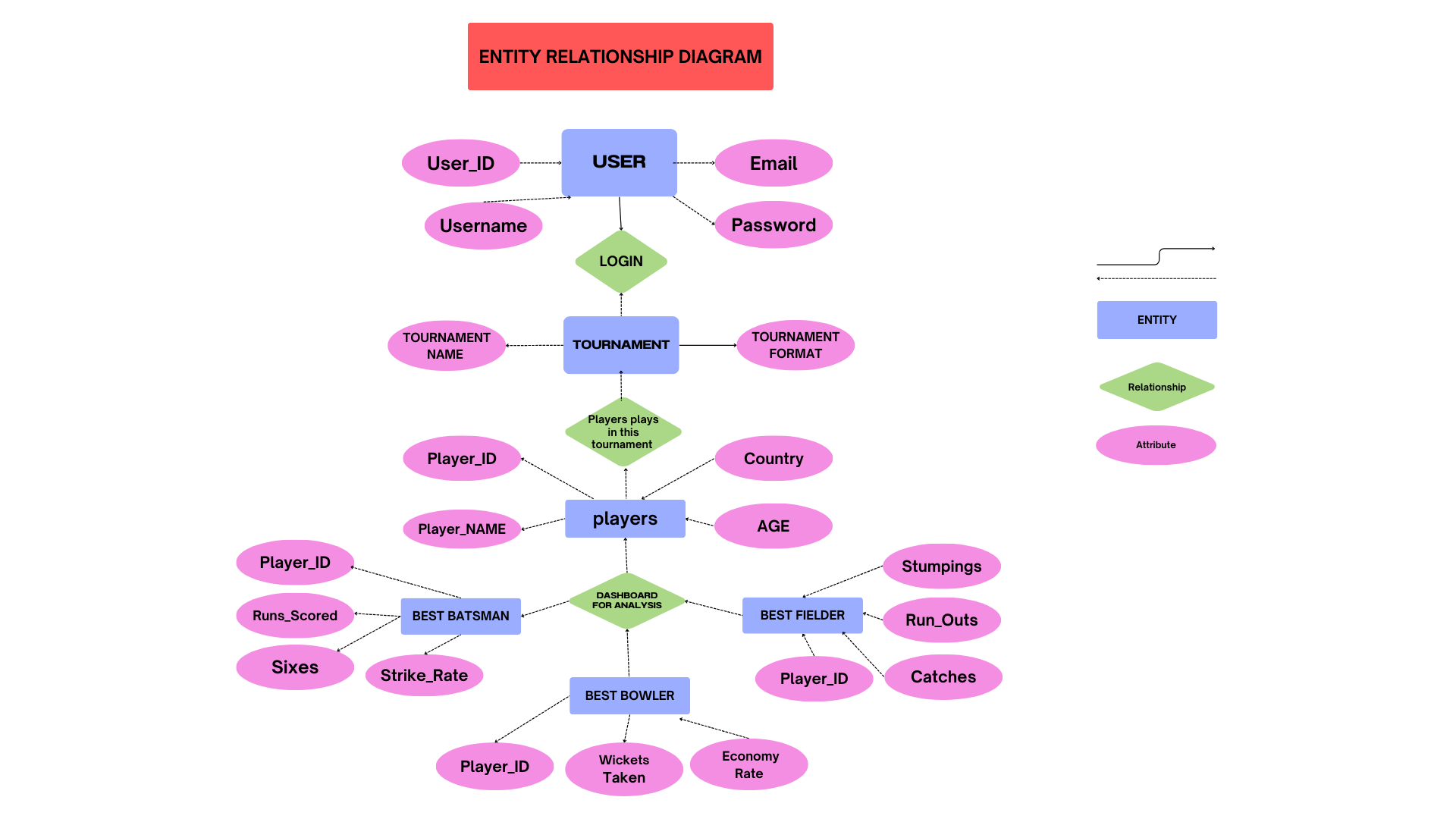


Figure 2 : E-R Diagram

**3.2 Use Case Diagrams**

Use case diagrams represent the different user interactions within the system, such as a user accessing the dashboard to view player statistics or interact with filters to drill down into specific data points.

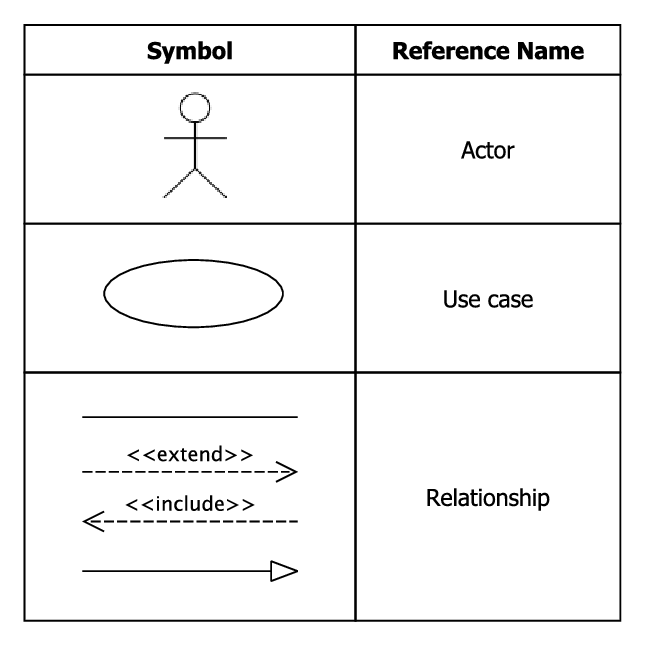


Figure 3 : Symbols of Use Case Diagram

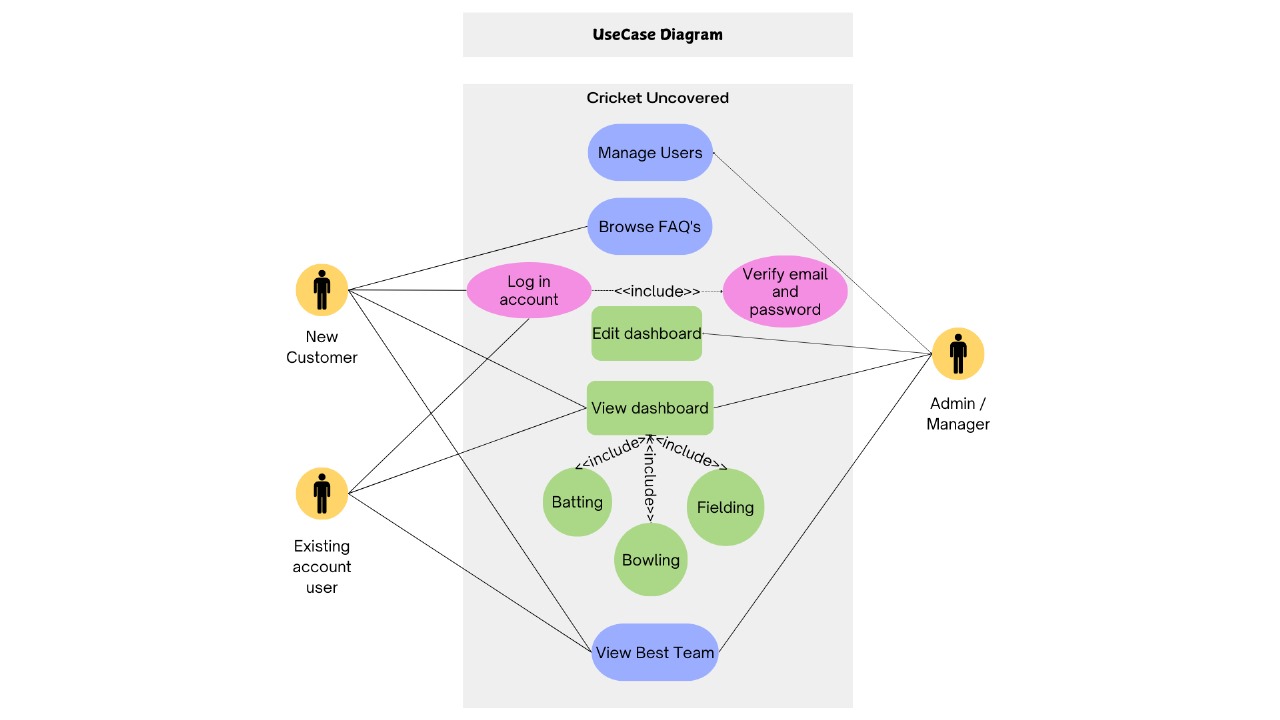


Figure 4 : Use Case Diagram

**3.3 DFD Diagrams**

Data Flow Diagrams (DFD) represent the flow of data between different components of the system, from data extraction using Bright Data to its final visualization in Power BI.

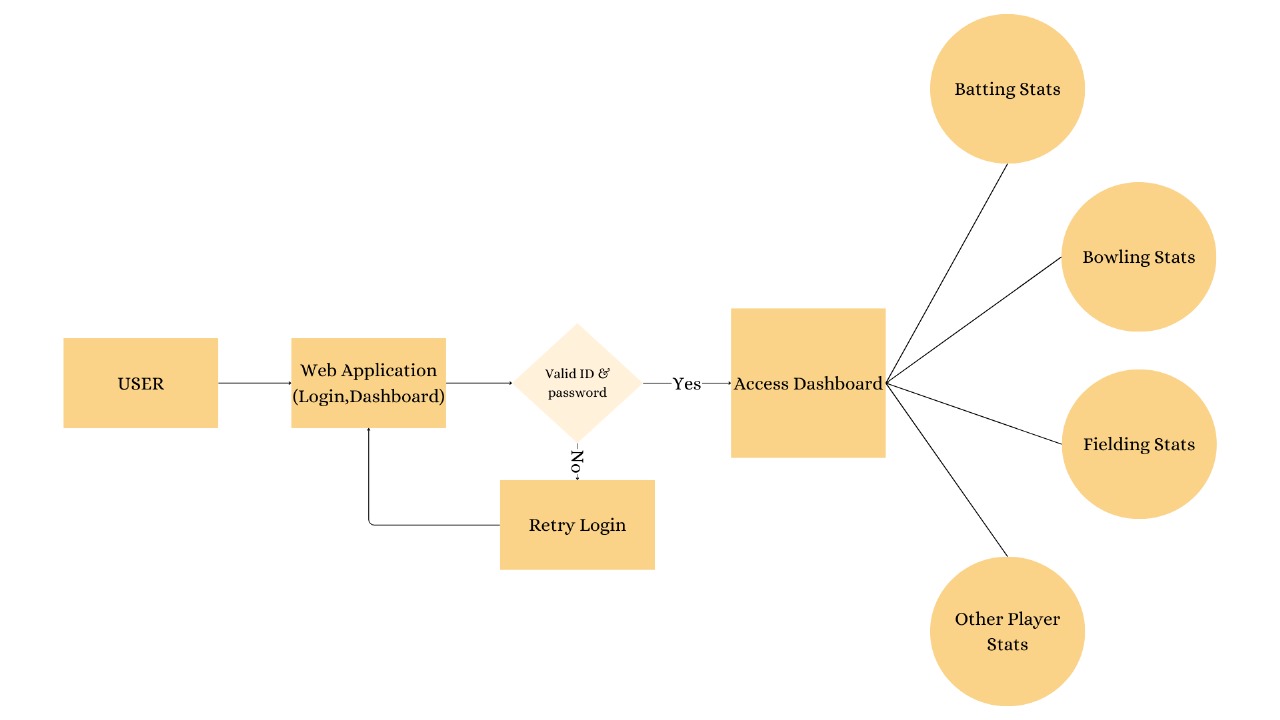


Figure 5 : DFD Diagram

**3.4 Activity Diagrams**

Activity diagrams depict the sequence of activities involved in the workflow, starting from data extraction to dashboard presentation.

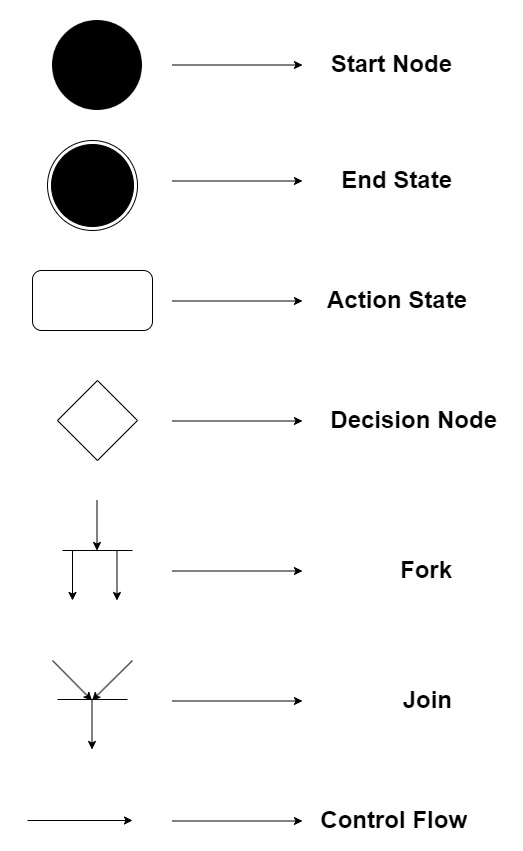


Figure 6: Symbols of Activity Diagram

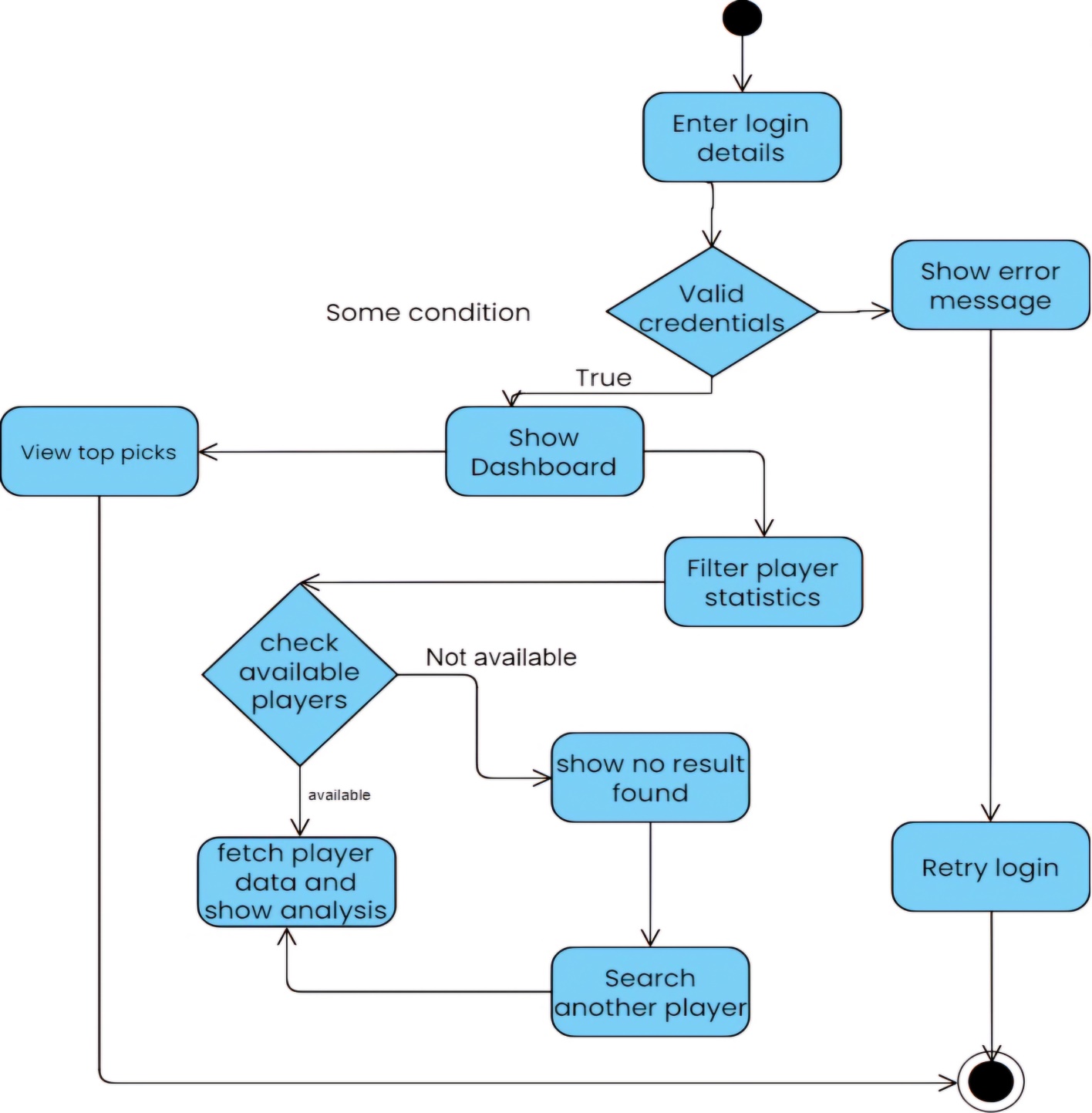


Figure 7 : Activity Diagram

**3.5 Class Diagram**

The class diagram shows the system's structure by illustrating its classes, attributes, and the relationships between them, such as the connection between Player and Statistics.

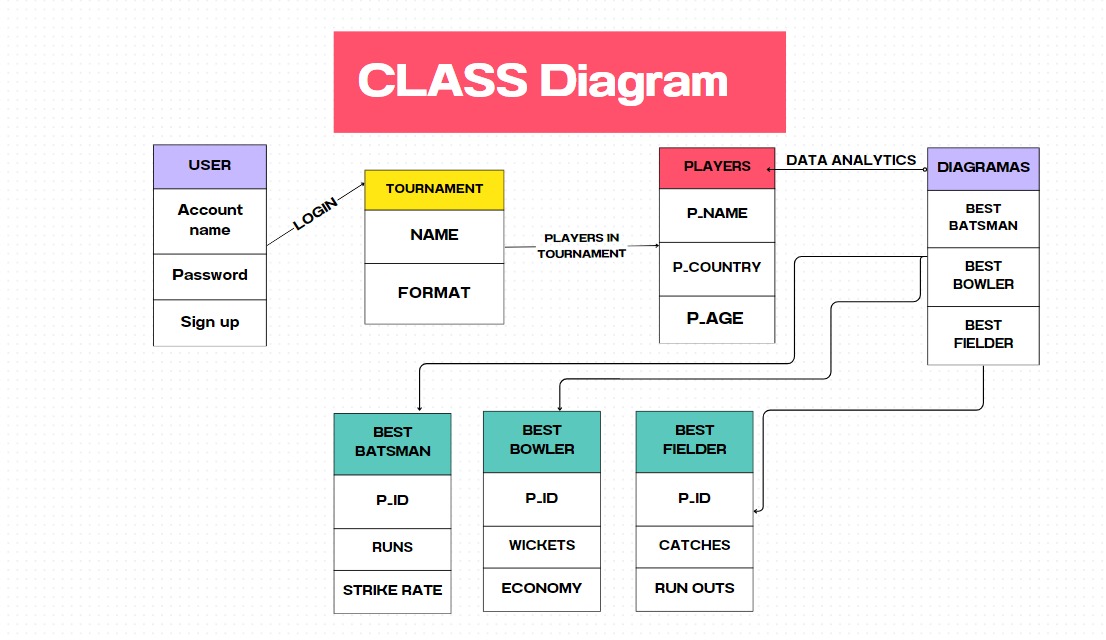


Figure 8 : Class Diagram

**Chapter: 4 IMPLEMENTATION**

The implementation phase of the "Cricket Uncovered" project is a multi-step process involving data extraction, cleaning, transformation, modeling, visualization, and website development. Each step was carefully executed using various tools and technologies, ensuring the project achieved its objectives of showcasing the best cricket players in a tournament through interactive dashboards.

**4.1 Data Extraction**

The project begins by extracting cricket data from ESPN’s website using Bright Data, a web scraping tool. The extraction focuses on collecting relevant data such as match results, individual player performance metrics (runs, strike rates, economy rates, etc.), and role-specific performance data (opener, middle-order batsman, bowler, etc.). Bright Data automates the scraping process, fetching and storing data in a structured format, which ensures that we have a large, reliable dataset to work with.

**4.2 Data Cleaning**

Once the data is extracted, it undergoes a thorough cleaning process using Python and the Pandas library. This step involves removing duplicate entries, handling missing data, and ensuring data consistency. For example, player names are standardized, data types are corrected, and match statistics are aligned across different roles. Pandas also enables the removal of irrelevant columns or features that are not useful for analysis, ensuring that the dataset is concise and focused.

**4.3 Data Transformation**

With a clean dataset in hand, the next phase is data transformation. Power Query is utilized here, which is integrated with Power BI. Power Query helps filter, sort, and merge data, ensuring that the dataset is structured for detailed analysis. This step allows us to convert the raw data into a more usable format, grouping players by role, summarizing statistics, and preparing the data for modeling.

**4.4 Data Modeling**

Data modeling is performed using DAX (Data Analysis Expressions) within Power BI. DAX enables the creation of calculated columns and measures, such as strike rate, batting average, and bowling economy. These metrics are crucial for determining the best players across different roles. For instance, we calculate strike rates for batsmen, which helps identify the most efficient run scorers. Similarly, bowling averages and economy rates are calculated for bowlers. These metrics are then combined with other performance factors, such as pressure situations and consistency, to select the best opener, middle-order player, finisher, and bowler.

**4.5 Data Visualization**

The processed data is visualized using Power BI. The interactive dashboard is created, featuring various charts, graphs, and filters that allow users to explore player performance in depth. Visualizations such as bar charts, pie charts, and line graphs are employed to compare players, display key statistics, and showcase role-specific contributions. Power BI’s dynamic filtering capabilities enable users to drill down into the data, analyzing specific matches, players, or teams.

**4.6 Website Development**

The final stage involves embedding the Power BI dashboard into a website. This is achieved using HTML and CSS, creating a user-friendly interface for the dashboard. The website is designed to be simple yet effective, allowing users to interact with the dashboard without any technical barriers. The website ensures that the data is accessible to cricket fans, analysts, and coaches alike

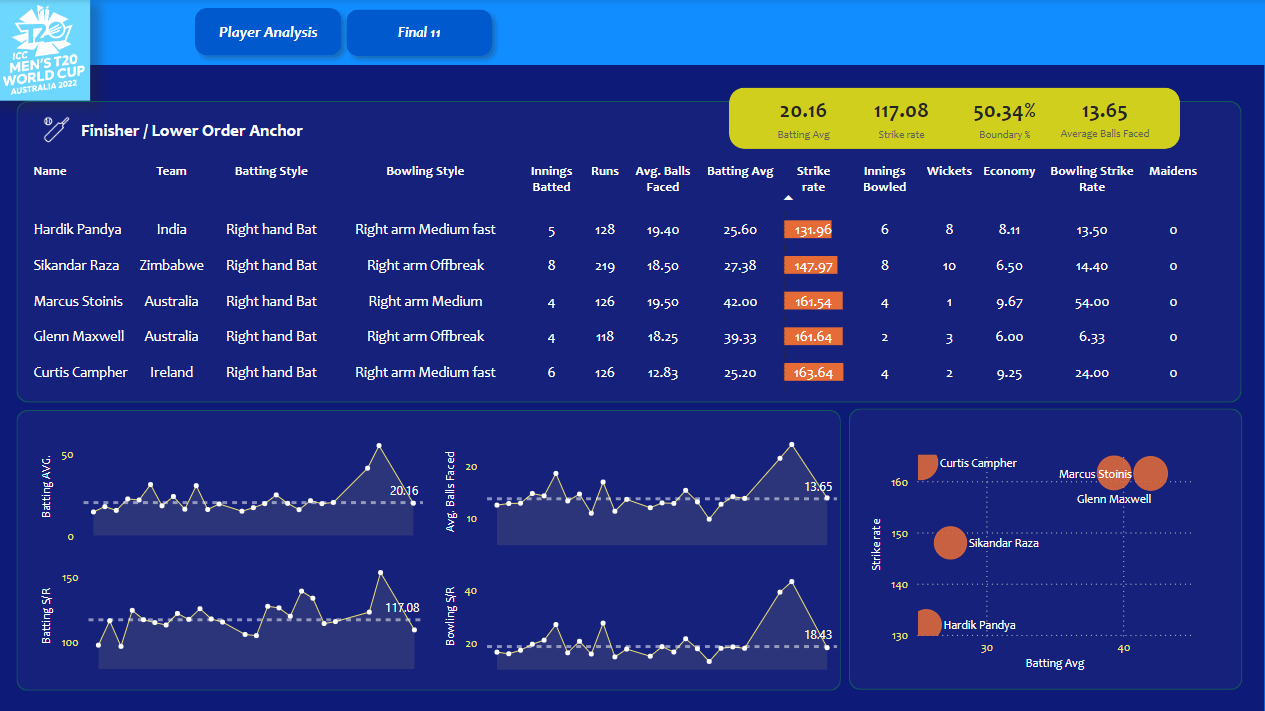


Figure 9 : Dashboard 1



Figure 10 : Dashboard 2



Figure 11 : Dashboard 3



Figure 12 : Dashboard 4

**Chapter: 5 CONCLUSION**

The "Cricket Uncovered" project demonstrates the potential of modern data analysis techniques in the world of cricket. By leveraging tools like Bright Data, Pandas, Power Query, and Power BI, we have successfully developed an interactive dashboard that provides deep insights into player performance. The project goes beyond traditional cricket statistics by focusing on role-specific contributions, helping users identify the best players in various positions, such as opener, middle-order batsman, finisher, and bowler.

Through the dynamic capabilities of Power BI, users can explore the data interactively, comparing players based on key performance indicators like strike rate, batting average, and bowling economy. The web-based interface ensures that the dashboard is accessible to a wide audience, from casual cricket fans to professional analysts and coaches.

This project highlights the growing importance of data in sports and demonstrates how technology can be used to enhance the understanding and appreciation of cricket. "Cricket Uncovered" not only provides a comprehensive analysis of player performance but also serves as a foundation for future developments in the field of sports analytics.

**Chapter: 6 FUTURE SCOPE**

The future scope of the "Cricket Uncovered" project is broad, with several opportunities for expansion and enhancement. While the current system provides a detailed analysis of player performance based on historical data, future iterations of the project could introduce advanced features and capabilities:

**6.1 Predictive Analytics**

One of the most promising areas for future development is the incorporation of predictive analytics. By applying machine learning algorithms to the existing data, the project could forecast future player performance, helping teams and analysts make strategic decisions based on predictive models. For example, we could predict which players are likely to perform well in upcoming matches based on their historical performance and current form.

**6.2 Real-Time Data Integration**

Currently, the project relies on historical data extracted from ESPN. However, the system could be extended to include real-time data integration, allowing for live updates during ongoing matches. By incorporating APIs or real-time scraping, the dashboard could provide live analytics, helping coaches and teams make in-game decisions based on current match situations.

**6.3 Expansion to Multiple Leagues**

While the current project focuses on a single tournament, future versions could incorporate data from multiple cricket leagues and tournaments, such as the Indian Premier League (IPL), Big Bash League (BBL), and international matches. This would allow users to compare player performance across different leagues, providing a more comprehensive view of their abilities.

**6.4 Enhanced User Interface**

In the future, the web interface could be enhanced with additional features, such as player profiles, match highlights, and detailed reports. User authentication and personalized dashboards could be introduced, allowing users to save their preferences, track specific players, or receive notifications when new data is available.

**Chapter: 7 BIBLIOGRAPHY**

ESPN Cricket Data:

ODI:

[**https://www.espncricinfo.com/records/tournament/team-match-results/icc-cricket-world-cup-2023-24-15338**](https://www.espncricinfo.com/records/tournament/team-match-results/icc-cricket-world-cup-2023-24-15338)

T20:

[**https://stats.espncricinfo.com/ci/engine/records/team/match\_results.html?id=14450;type=tournament**](https://stats.espncricinfo.com/ci/engine/records/team/match_results.html?id=14450;type=tournament)