Cricket WC - Team Selection

Submitted in partial fulfillment of the requirements of

PG-DIPLOMA IN BIG DATA ANALYSIS

By

Dhawal Dattani	230310125005
Mohd Yusuf	230310125010
Sudhanshu Prajapati	230310125019

Project Guide

Sh. Sunil Kumar

(Senior Project Engineer, C-DAC, ACTS, New Delhi)

CENTER FOR DEVELOPMENT OF ADVANCED COMPUTING

New Delhi.

March 2023 – September 2023

TABLE OF CONTENTS

CERTIFICATE

This is to certify that the project entitled "**Cricket WC Team Selection**" is a teamwork work of "**Dhawal Dattani (230310125005), Mohd Yusuf (230310125010), Sudhanshu Prajapati (230310125019).**" Submitted to **CDAC, ACTS, New Delhi** in partial fulfillment of the requirement for the PG-Diploma in Big Data Analysis.

Mr. Sunil Kumar

Senior Project Engineer

C-DAC, ACTS, New Delhi

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Dhawal Dattani)
230310125005

(Mohd Yusuf)
230310125010

(Sudhanshu Prajapati)
230310125019

Date: 04-09-2023

ABSTRACT

This project delved into analyzing the ODI performances of players from the Indian Cricket Team, starting with data exploration of ESPNCricinfo dataset. Utilizing Python and Bright data, web scraping techniques were employed to collect extensive player performance data. Subsequent data cleaning processes were carried out using Google Colab and pandas, addressing missing values and standardizing formats. The project culminated in a comprehensive Power BI dashboard, showcasing insights via visualizations such as line charts, bar graphs, and scatter plots. This dashboard provides a valuable resource for informed decision-making, offering a deep understanding of player metrics and correlations, thus enabling data-driven strategies for enhancing the team's ODI performance.

1. INTRODUCTION

In the dynamic realm of international cricket, the process of team selection for a significant event such as the Cricket World Cup assumes paramount importance. It resembles the meticulous assembly of a puzzle, necessitating the judicious selection of the most proficient components to construct a triumphant picture. This undertaking amalgamates astute strategic planning, a meticulous analysis of player statistics and performance metrics, and a comprehensive understanding of how individual players harmonize within the collective unit.

The Cricket World Cup, often likened to the Olympics in the cricketing world, stands as the pinnacle tournament within the sport. Consequently, the composition of the chosen team assumes a position of utmost significance. To achieve this objective, a comprehensive evaluation of players' proficiency in both batting and bowling disciplines is imperative. Our aspiration is to construct a team capable of excelling in all facets of the game.

The aim is to forge a cohesive unit, akin to a finely-tuned machine, to perform optimally on the grandest stage of international cricket. This endeavor carries profound significance, and it commences with the judicious selection of the most suitable players.

2. DESCRIPTION

The selection of a Cricket World Cup team involves a meticulous process of evaluating players based on their recent tournament performance, skills encompassing batting and bowling, while aiming for a balanced composition that covers all facets of the game. This endeavor is greatly facilitated by the utilization of cutting-edge technology. Python, for instance, plays a pivotal role in automating data collection and analysis, while tools like Google Colab facilitate efficient web scraping to gather crucial statistics and insights from platforms such as ESPNCricinfo.

Once the data is acquired, it undergoes rigorous scrutiny and refinement using Pandas and MS Excel, ensuring that it is accurate and comprehensive. This step is vital in making informed decisions about player selection. Furthermore, Power BI is employed for data visualization, enabling the selection committee to gain actionable insights and assess player performance trends effectively.

The final squad is strategically tailored to execute game plans and tactics, with backups in place to address potential contingencies, all in pursuit of assembling a harmonious and competitive team capable of contending for the World Cup title. This synergy between cricketing expertise and technological prowess enhances the precision and effectiveness of the team selection process.

3. SCOPE

The meticulous process of selecting a Cricket World Cup team is augmented by the strategic utilization of technology, predominantly Python and dainty Bright Data for web scraping, Pandas and MS Excel for data cleaning, and Power BI for visualization. These technological tools play a pivotal role in enhancing the efficiency and accuracy of the player evaluation and selection process.

Python enable the committee to efficiently gather extensive data on player performance from various sources, providing a comprehensive foundation for decision-making. The data collected is then meticulously cleaned and organized using Pandas and MS Excel, ensuring that statistical insights are based on accurate and reliable information.

Power BI serves as a powerful tool for visualizing data trends, enabling the committee to discern patterns and make informed decisions regarding player selection. This integration of technology ensures that the final squad is strategically tailored to execute game plans and tactics effectively, with contingency plans in place.

In summary, the incorporation of Python, Pandas, MS Excel, and Power BI enhances the selection process, aiding in the assembly of a harmonious and competitive team capable of contending for the coveted World Cup title.

4. SYSTEM REQUIREMENTS

4.1   Hardware Requirements:

- Processor: Intel Core i3 and above

- Disk Space 5 GB Minimum

- RAM: 2 GB or more

4.2   Software Requirements:

- Internet: Web Browser, Operating System (Windows 7 or above).

- Internet Connection: (any).

- Dashboard/Visualization – Power BI

4.3   Technologies Used:

- Web Scraping– Python, Bright Data

- Data Cleaning – Python & Google Colab

- Dashboard/Visualization – Power BI.

5. SYSTEM ANALYSIS

The system analysis is provided on basis on which requirement is fulfilled.

5.1 Openers

Parameter	Description	Criteria
Batting Average	Average Runs scored in an innings	>40
Strike Rate	Runs scored per 100 balls	>90
Innings Batted	Total Innings Batted	>20
Boundary %	Percentage of runs scored in boundaries	>50
Batting Position	Order at which Batter arrives to pitch	<3

5.2 Middle Order

Parameter	Description	Criteria
Batting Average	Average Runs scored in an innings	>35
Strike Rate	Runs scored per 100 balls	>85
Innings Batted	Total Innings Batted	>35
Boundary %	Percentage of runs scored in boundaries	>40
Batting Position	Order at which Batter arrives to pitch	>2

5.3 All Rounders

Parameter	Description	Criteria
Batting Average	Average Runs scored in an innings	>20
Strike Rate	Runs scored per 100 balls	>100
Innings Batted	Total Innings Batted	>5
Boundary %	Percentage of runs scored in boundaries	>40
Batting Position	Order at which Batter arrives to pitch	>5
Innings Bowled	Total Innings Bowled	>30

5.4 Spinner

Parameter	Description	Criteria
Innings Bowled	Total innings bowled	>50
Bowling Economy	Average runs per over	<6
Bowling Strike Rate	Average number of balls before a wicket	<36
Bowling Style	Bowling Style of player	Spin
Bowling Average	Runs allowed per wicket	<30

5.5 Fast Bowlers

Parameter	Description	Criteria
Innings Bowled	Total innings bowled	>20
Bowling Economy	Average runs per over	<5.7
Bowling Strike Rate	Average number of balls before a wicket	<32
Bowling Style	Bowling Style of player	Fast
Bowling Average	Runs allowed per wicket	<26

6. DATA FLOW DIAGRAM

We commence with Data Exploration, delving into ESPNCricinfo's dataset, scrutinizing ODI performances of Indian Cricket Team players. The focus is on comprehending data structures, identifying pertinent features, and drawing insightful conclusions, analyzing batting averages, bowling averages, strike rates, and wicket counts.

The subsequent step involves Web Scraping, utilizing Python and Bright data to amass an extensive ODI performance dataset. This necessitates meticulous HTML table examination for data collection. Following data acquisition, our journey advances to Data Cleaning, ensuring the dataset's readiness for analysis. This involves managing missing data, resolving inconsistencies, and standardizing formats, facilitated through Google Colab and pandas.

The pinnacle of our effort manifests in the Dashboard/Visualization. Where, a dynamic Power BI dashboard showcases crucial insights, from individual player performances to metric comparisons, fostering data-driven decision-making for cricket team selection and performance analysis.
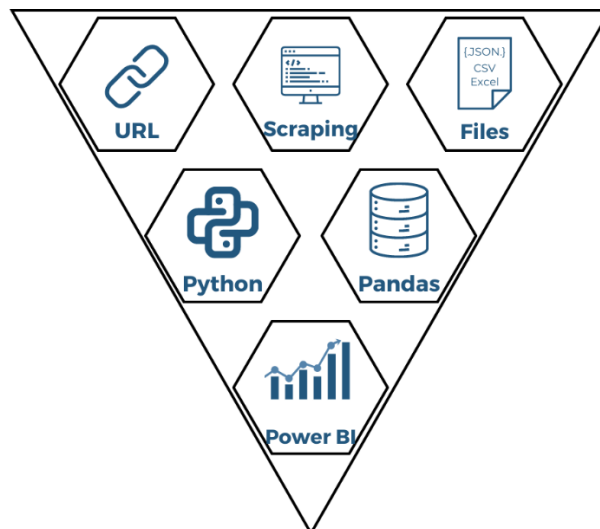

Figure - 6.1 Data Flow Diagram

7. ACTIVITY FLOW DIAGRAM

7.1 Data Exploration:

We began your project by exploring data from ESPNCricinfo, focusing on the ODI performances of players from the Indian Cricket Team. This exploration likely involved understanding the structure of the data, identifying relevant features, and gaining insights into player performances. We looked at various statistics, such as batting averages, bowling averages, strike rates, wicket counts, and more.

7.2 Web Scraping:

For gathering the ODI performance data, we utilized web scraping techniques. This involved using Python & Bright data, in which we had to inspect through the HTML tables for the data. We were able to collect a substantial amount of data on player performances

Figure - 7.2.1

7.3   Data Cleaning:

After collecting the data, we performed data cleaning to prepare it for analysis. This step might have included handling missing values, dealing with inconsistencies or errors in the data, standardizing formats, and ensuring the data is in a structured and usable form. Google Colab, along with the pandas library, was your chosen platform for performing these data cleaning operations.

Figure - 7.3.1

7.4   Dashboard/Visualization:

With the cleaned data, you proceeded to create a dashboard using Power BI. This dashboard likely displayed key insights and visualizations derived from the data. We could have included visualizations like line charts for individual player performances, bar charts for comparing different metrics, scatter plots for assessing correlations, and more. Our goal was to create a comprehensive dashboard that would allow you to make informed decisions based on the data.

Figure - 7.4.1

8. SCREENSHOTS OF PROJECT

Below are the screenshots of the dashboard that depicts various functions and selection criteria's that are mentioned above in 5.1, 5.2, 5.3, 5.4 and 5.5.
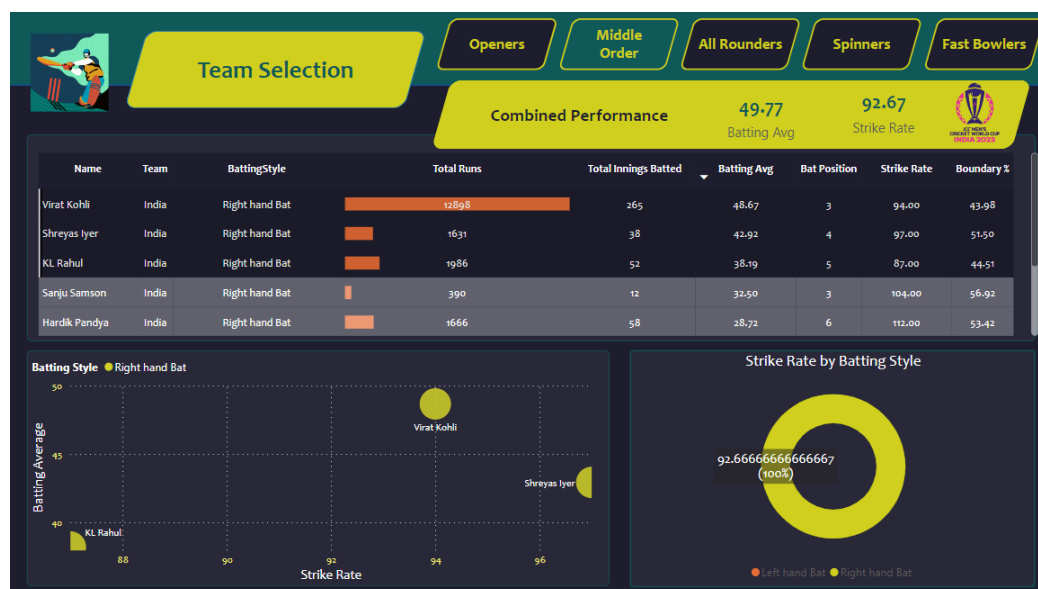

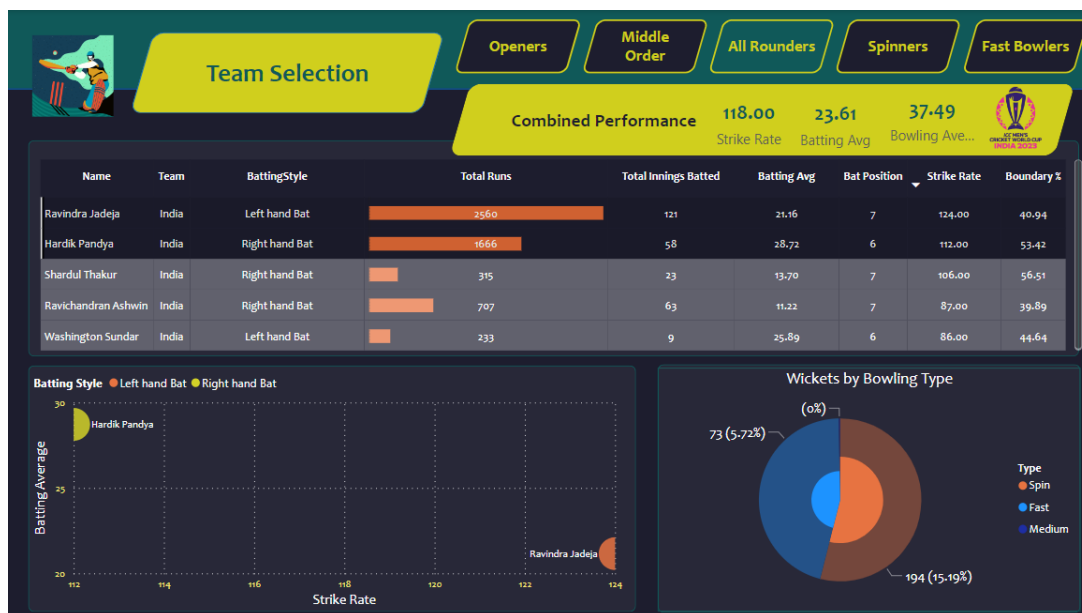Figure - 8.1 – Opener Selection

Figure - 8.2 – Middle Order Selection

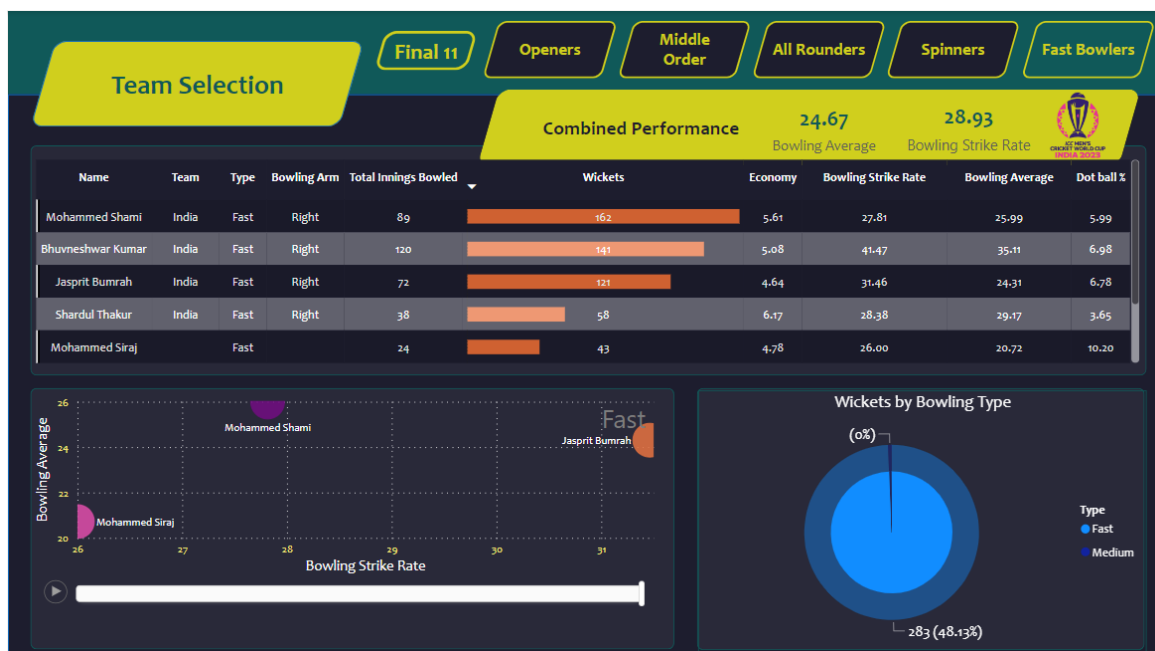Figure - 8.3 – All Rounders

Figure - 8.4 – Spinner

Figure - 8.5 – Fast Bowlers

Figure - 8.6 – Final 11

9. CONCLUSION

We embarked on a comprehensive project involving ODI performance analysis of Indian Cricket Team players. Leveraging web scraping techniques with Python, you collected performance data from ESPNCricinfo website. Subsequently, we performed data cleaning and preprocessing using Google Colab and pandas to ensure the data's quality and consistency. The cleaned data was then transformed into a CSV format.

This project showcased our ability to navigate the entire data analysis, from data collection and cleaning to visualization and decision-making. It demonstrated our proficiency in using a combination of programming tools and data visualization platforms to derive actionable insights from cricket performance data.

REFERENCES

- ESPNCricinfo - https://www.espncricinfo.com/records

- Google Images- https://www.google.com/

- Python - Data Wrangling with Python – Tips and Tools to Make Your Life Easier by Jacqueline Kazil, Katharine Jarmul

- Web Scraping – https://www.geeksforgeeks.org/python-web-scraping-tutorial/

- Power BI - https://www.youtube.com/

- Pandas - https://towardsdatascience.com/

- Literature Review -

  - Player selection in cricket based on similarity of playing conditions by Walid Mohammad, Sadia Sharmin (May 2020)

  - Intelligent Cricket Team Selection by Predicting Individual Players' Performance using Efficient Machine Learning Technique by Chetan Kapadiya, Ankit Shah, Kinjal Adhvaryu, Pratik Barot (February 2020)

  - Selection of Indian Cricket Team in ODI using Integer Optimization by Prachi Agrawal, Talari Ganesh (December 2019)