IPL BATSMEN EFFICENCY : BASED ON BATTING ORDER.

BDM163 MSC BUSINESS ANALYTICS BUSINESS PROJECT

ASTON UNIVERSITY

ASTON BUSINESS SCHOOL

SEPTEMBER 2023

KAPIL DAGADE

STUDENT NO: 220246725 CANDIDATE NO: 766203

ACKNOWLEDGMENT:

I extend my heartfelt gratitude to Dr. Gary Simpson. His unwavering support and insightful feedback have been instrumental throughout the course of this project. I am deeply appreciative of his genuine guidance.

I would also like to express my profound thanks to the MSc Business Analytics faculty at Aston University. Their expert instruction has equipped me with invaluable skills and methodologies that not only aided in the completion of this report but will also serve me well in my future endeavours.

Lastly, my gratitude extends to my family. Their steadfast support and encouragement have been the bedrock of my academic journey. I am eternally grateful for their sacrifices and belief in me.

DECLARATION:

I declare that I have personally prepared this report and that it has not in whole or in part been submitted for any other degree or qualification. Nor has it appeared in whole or in part in any textbook, journal or any other document previously published or produced for any purpose. The work described here is my/our own, carried out personally unless otherwise stated. All sources of information, including quotations, are acknowledged by means of reference, both in the final reference section and at the point where they occur in the text.

EXECUTIVE SUMMARY:

In our current analysis, we benchmark the efficiency of batsmen across roles in the Indian Premier League (IPL), T20 cricket tournament using data envelopment analysis. We focused on three batting positions: openers, top middle order, and lower middle order. Key metrics include runs scored, boundaries, traditional metrics like balls played and strike rate, and advanced metrics such as four strike rate and average balls per inning etc. The dataset captures performances from various batsmen over the IPL seasons 2018-22 and franchises.

The hypothesis tests , the test With a p-value 0.05 rejected the null hypothesis to identify marked differences in batting order performances. Jos Buttler, Lokesh Rahul, Prithvi Shaw, Devon Conway, Matthew Wade, Manan Vohra, Parthiv Patel, and Brendon McCullum were the benchmark players for openers. In the top middle-order position, benchmark batsmen were Bhanuka Rajapaksa, Sunil Narine, Liam Livingstone, Colin Munro, Rishabh Pant, Rajat Patidar, Suryakumar Yadav, and AB de Villiers. This selection was based on inputs such as innings and balls played, with outputs including boundaries, average, and strike rate. For the lower middle-order position, benchmark batsmen included Andre Russell, Rowman Powell, and Krishnappa Gowtham. Using the Variable Returns to Scale (VRS) Model and correlation analysis, we gauge batsmen efficiency and relationships between metrics. These insights hold substantial value for IPL franchises and players, guiding better team strategies and on-field performances.

Keywords: Performance Efficiency; Batting Roles; Batsmen Benchmarking; Variable Returns to Scale (VRS) Model; Correlation Analysis; Team-related Metrics; Individual Performance Metrics; IPL Landscape

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 CRICKET HISTORY AND TYPES

Cricket is a sport with a long history, going all the way back to the 16th century. Over the years, cricket has undergone significant transformations, giving rise to various formats. It has evolved from long-drawn Test matches to One Day Internationals (ODIs), initially featuring 60 overs but eventually settling on 50 overs per side and eventually to the fast-paced most popular Twenty20 (T20) format.

1.1.1 T20 CRICKET

T20 cricket, also known as Twenty20, was introduced in 2003 and initially contested between English and Welsh domestic sides in the Twenty20 Cup (Manage & Scariano, 2013).

In T20 cricket, each team bats for a maximum of 20 overs, which usually results in the game being completed in about three hours. This shorter format has been designed to be more viewer-friendly compared to traditional Test and One-Day International (ODI) formats. Perera, H., & Swartz, T. (2013). Resource estimation in T20 cricket. D Test matches have a duration of up to five days, while One Day Internationals (ODIs) can occupy an entire day. This extended timeframe can be less accessible to casual fans or individuals with limited time to spare. In contrast, Twenty20 (T20) cricket has garnered immense popularity due to its rapid pace and condensed duration. It is often considered more engaging and thrilling for spectators, offering a dynamic and fast-paced cricketing experience.

The inaugural international T20 match was played between New Zealand and Australia, 2005, didn't take long for the cricketing world to embrace the fast-paced and thrilling format. Two years later, In 2007, the first-ever ICC T20 World Cup took place. This tournament marked the beginning of international cricket's exciting Twenty20 format. It was held in South Africa from September 11 to September 24, 2007. Twelve teams participated, including the ten Test-playing nations and the finalists of the 2007 WCL Division One tournament, Kenya and Scotland. India emerged as the champions, defeating Pakistan in an exciting final match. This inaugural T20 World Cup was a significant moment in cricket history. The tournament witnessed extraordinary performances, breathtaking sixes, and nail-biting finishes that became the trademark of T20 cricket (Anon, 2023).

India's victory in the 2007 T20 World Cup was remarkable due to several factors. The team, led by a young MS Dhoni and coached by Lalchand Rajput, had played just one T20I before the tournament, making them relative newcomers to the format.Espn.(n.d).

The memorable match in the 2007 T20 World Cup final showcased the intense rivalry between India and Pakistan and ended with India as deserving champions, securing their first major silverware in decades. India defied the odds and emerged victorious, marking a historic moment. The triumph of the 2007 ICC T20 World Cup paved the path for the swift expansion of T20 leagues worldwide, including the likes of the Indian Premier League (IPL), Big Bash League (BBL), and the Caribbean Premier League (CPL). These leagues have now become familiar names in households and have nurtured a fresh wave of cricketing sensations who shine in the briefest version of the sport (Anon, 2007).

1.1.2 INDIAN PREMIERE LEAGUE

The T20 World cup had a significant impact on cricket, leading to the birth of the Indian Premier League (IPL) and the evolution of players in the T20 format. Espn.2023

In 2008, a big change happened in the world of cricket when the Indian Premier League (IPL) started. The IPL was the idea of the Board of Control for Cricket in India (BCCI), and it mixed sports with entertainment. This made the T20 format of cricket very popular, and it attracted cricket players from different countries. Because of this, the IPL gained many fans, both in India and around the world. People could watch international players in action every year, and this made more and more people like the IPL (Ghai & Zipp, 2020).

The first IPL match set the tone for the future. The first IPL match was between the Kolkata Knight Riders (KKR) and the Royal Challengers Bangalore (RCB) at the M Chinn swamy Stadium in Bangalore. In that match, a foreign player from Kolkata scored 158 runs in just 73 balls. He hit 13 sixes and 10 fours during his innings. This amazing performance was just the beginning of something great.

It took many years for someone to break this record. Eventually, a foreign batsman from the West Indies scored 175 runs in 66 balls for the Royal Challengers Bangalore. He also scored the fastest century ever in any format of cricket. To this day, his record remains unbroken. Espn.2023. This internationalization has made the IPL a global sporting spectacle, watched by millions worldwide (Ghai & Zipp, 2020).

The Indian Premier League (IPL) garnered millions of viewers, not only boosting its fanbase but also attracting substantial sponsorship deals, brand advertisements, and increased revenue. Indian Premier League (IPL) brought financial stability to a diverse group of cricketers, transforming the economic landscape of cricket itself (Karhadkar, 2018).

Major companies like DLF, PepsiCo, Vivo, and Tata have sponsored the IPL, significantly contributing to its financial success. For instance, in the first year, DLF provided substantial sponsorship, leading to high profits due to increased viewership. With more and more people watching IPL matches on TV and online, companies are willing to pay a lot to advertise during the games. So, in simple terms, the IPL has become a super valuable cricket league by getting companies to sponsor them, through ticket sales and viewership.

Figure 1: Valuation of Sporting Leagues

In recent years, the IPL has made a ton of money. They got a massive $6 billion deal for broadcasting their matches and are now worth around $10.9 billion. This makes them one of the most valuable sports leagues globally, right after the NFL and ahead of Major League Baseball.

Now, where does all this money go?

Much of the IPL's revenue is distributed among the participating teams.

A significant portion of the IPL revenue goes to the teams themselves. Teams receive a share of the central pool of revenue generated by broadcasting rights, digital partnerships, and other league-wide sources. This revenue is divided among the teams, with popular teams getting a larger share and Sponsorship Deals. The distribution is not uniform and depends on various factors.

Broadcasting Rights: IPL's broadcasting rights are sold to television networks and digital platforms. A substantial portion of the overall revenue generated from broadcasting rights is distributed among the teams.

Moreover, Teams with larger fan bases secure lucrative sponsorship deals from companies looking to gain visibility through the tournament. For instances, sponsorship, these deals can include jersey sponsorships, helmet sponsorships, and more.

Ticket Sales: Revenue is generated from ticket sales for the matches. The teams earn a significant share of this revenue from home matches. Stadiums can hold tens of thousands of fans, and ticket sales contribute substantially to team earnings.

Prize Money: Teams that perform well during the IPL season are rewarded with prize money. The champion, runner-up, and other high-performing teams receive substantial cash prizes, providing an additional incentive for teams to excel.

Merchandise Sales: Teams sell merchandise such as jerseys, hats, scarves, and other fan gear. These sales contribute to the team's revenue, and fans often purchase these items to show their support for their favourite teams.

So, IPL teams make money from various sources like TV deals, sponsorships, ticket sales, merchandise, and more. The better the team and the more popular it is, the more money it can earn. This system keeps the competition exciting and profitable for all the teams involved (Arpika, 2023). With its franchise-based structure, the Indian Premier League (IPL) has emerged as a prominent T20 cricket tournament that has garnered global attention (Manage & Scariano, 2013).

However, the revenue generation, sponsorship deals, and fan engagement of a sports team are intricately tied to the team's performance, which, in turn, is heavily reliant on the individual performances of its players. The excellence of each player directly influences the overall success of the team, as their collective efficiency and synergy as a unit play a pivotal role in determining the team's future prospects. A team's performance can significantly improve when they have a combination of skilled and efficient players working together effectively.

Chennai Super Kings and Mumbai Indians, tend to receive a larger share of the revenue due to their immense popularity and unparalleled achievements in the Indian Premier League (IPL) (Arpika. 2023).

As the epitome of excellence in the league, both franchises have each secured a remarkable five IPL championships, setting the gold standard for success in the tournament. This on-field dominance has not only elevated their status but also significantly amplified their mass appeal, thereby drawing legions of devoted supporters and further enriching their financial offers. As a result, their player performance, enhancing both their brand value and their financial position within the league.

This clearly tells the story of players performance as the base of all glory and earning. Cricket, as a sport, holds significant importance globally due to its entertaining nature. The performance of each individual player within a team is crucial for the team's success. The economic factors associated with cricket have also amplified the need for statistical analysis of player performance and decision-making processes based on such analyses (Asif et al., 2019).

In conclusion, the performance of cricket players is at the heart of the success of the team and the economic implications of the sport.

The assessment of cricketers' performance has been a focal point of many research endeavours. Lemmer (2004) took a pioneering step by quantifying the consistency of cricket batsmen, introducing a consistency coefficient based on the adjusted coefficient of variation (Davis et al., 2015). Furthered the discourse by presenting an innovative metric, the "expected run differential," specifically tailored for Twenty20 cricket. While traditional and single-metric analyses have added substantial value, they occasionally fall short in predicting and forecasting future performances. It's here that the significance of efficiency comes into play. Efficient metrics not only gauge past and present performance but also set the stage for more accurate future predictions. Murali (2020) emphasized that embracing advanced techniques like data envelopment analysis can offer a holistic, nuanced, and, importantly, efficient perspective on player performance.

Data envelopment analysis (DEA) is a non-parametric method used to evaluate the efficiency of decision-making units (DMUs) based on multiple inputs and outputs. In the context of cricket, DMUs can be individual players or teams, and the inputs and outputs can be various performance metrics such as runs scored, wickets taken, strike rate, economy rate, and so on (Murali, 2020).

The advantage of using DEA in cricket performance analysis is that it provides a holistic view of a player's performance, taking into account various aspects of the game. Traditional methods might focus on one or two metrics, potentially overlooking other significant contributions a player might make. For instance, a batsman might not score many runs in a particular match but might have played a crucial role in building partnerships or rotating the strike, which might not be captured by traditional metrics. DEA, with its multi-input and multi-output approach, can capture such nuances.

Moreover, DEA allows for benchmarking players against a set of efficient players, helping coaches and selectors identify areas of improvement for individual players. It also aids in making informed decisions regarding team selection, player roles, and strategies.

As a result, the study of this paper will be focused on efficiency analysis of batsmen at different batting position for IPL data using Data envelopment analysis. The Indian Premier League (IPL) is distinguished from other cricket leagues by its franchise-based structure. During auctions, teams engage in heated bidding wars for players, exacerbating the IPL's commercial importance. This one-of-a-kind structure not only emphasises the game's economic dimensions, but it also emphasises the significance of accurate player selection and peak performance. With major investments at stake and each player's strategic position being critical, the need for player performance analysis becomes even more evident (Deep, Patwardhan, & Singh, 2016).

1.2 AIM OF THE PROJECT

Maximizing Player Performance in the Unique IPL Landscape by Benchmarking Batsmen Efficiency by Batting Position. Examine the performance efficiency of batsmen across different batting positions. Establish benchmark players for each batting role based on their efficiency scores. Offer actionable insights to franchises on which batsmen excel in their respective batting positions and provide guidelines on improving the efficiency of underperforming players. Using cost efficient technique.

Value Proposition & Example:

Identifying and benchmarking the most efficient batsmen for various batting positions enables franchises to make nuanced decisions that optimize player performance. For instance, if a middle-order batsman is identified as inefficient in comparison to the benchmark player, coaches can delve into specifics tailor team strategy to address these areas. This approach ensures each player is primed to contribute maximally, aligning with the IPL's distinctive demands. The study stands as a strategic linchpin for IPL franchises, directing data-driven decisions during auctions, in training, and in on-field strategies .

2 LITERATURE REVIEW

The review of literature goes extensively into the complexities of Data Envelopment Analysis (DEA) and its specialised applications in the realm of cricket.

We begin by studying the fundamental notions of classical DEA before progressing through its progression, which is defined by the incorporation of increasingly complex analytical meth odologies.The focus then shifts to the unique application of DEA in football, highlighting how this analytical tool has been skillfully used to uncover insights and influence decisions in this popular sport.This investigation is supplemented by our examination of Exploratory Data Ana lysis (EDA) in cricket, emphasising its critical role in data interpretation.As we progress throu gh these issues, we critically evaluate the existing literature, drawing together significant tak eaways and concluding with a concise overview of the current state of knowledge.

2.1 DATA ENVELOPMENT ANALYSIS IN CRICKET

Chaudhary et al. (2019) solved this issue by offering a novel technique for strategically selecting Indian Cricket Team players that harnesses the strengths of Data Envelopment Analysis (DEA). While traditional methods are trustworthy, they typically overlook the complexities that data-driven methodologies can disclose. The ability of DEA to analyse the relative efficiency of comparable Decision-Making Units (DMUs) justifies its usage. These DMUs corresponded to particular players in cricket. The versatility of DEA in using several inputs to produce multiple outputs provides a holistic evaluation that incorporates the multifaceted nature of player performance in cricket. The writer used an output-oriented DEA model primarily. This orientation aimed to maximise efficiency scores generated from several performance measurements, so providing a clear image of each player's contribution on the pitch. The article methodically collected efficiency scores for both batsmen and bowlers based on specific performance indicators. Outputs for hitters included metrics such as runs scored, strike rate, and batting average, providing a three-dimensional view of their batting ability. Bowlers, on the other hand, were evaluated using indicators like as maiden overs, wickets taken, and economy rate, which demonstrate their ability to limit the opposition and take important wickets. The CCR (Charnes, Cooper, and Rhodes), DEA model served as the study's foundation. This model, which is known for its simplicity and resilience, was perfectly suited. It not only facilitated a clear evaluation of players but also allowed for easy interpretation of results, thereby bridging the gap between complex analytics and actionable insights. The alignment of the writers methodology with the actual team selection by the BCCI for the India-England test series was perhaps the most compelling validation of Chaudhary et al.'s (2019) methodology. This convergence highlighted the model's accuracy in determining player efficiency, using in a new era of data-driven team selection techniques.

Similarly, The Amin & Sharma (2014) leaned heavily on Data Envelopment Analysis (DEA), for its adeptness in handling multiple inputs and outputs. Real data from the 2011 IPL was the foundation upon which their DEA model was built. Key metrics in focus for batsmen included Average (AVG) and Strike Rate (S/R), foundational pillars in the evaluation of a batsman's progress. An output orientation was the chosen path, aligning with the primary objective of maximizing player performance outputs given their inputs. By introducing a systematic, data-driven approach, Amin & Sharma aimed to minimize subjective biases in team selections. The IPL's unique franchise-based structure, where players are auctioned, further underscored the need for such an objective analysis. The commercial dimension of the IPL, as highlighted by Deep, Patvardhan, & Singh (2016), accentuates the importance of player performance analysis in team selection and strategizing.

However, Singh's (2011) investigation of the Indian Premier League (IPL) provides a balanced viewpoint . This paper's strength is its comprehensive approach, combining economic and sports performance in one model which helped us to answer the relationship of efficiency scores and how well teams did in the league and technical factors.

Writers says, Historically the analytical frameworks utilised for sports teams, particularly in well-known team sports in the United States, relied significantly on stochastic productivity and efficiency measuring methodologies. However, according to the writer these assumptions underlying these strategies, however, can be restrictive. As a result, the author uses Data Envelopment Analysis (DEA) as a tool of choice in this context.

The author used both constant returns to scale (CRS) and variable returns to scale (VRS) DEA models throughout the 2009 IPL season. While the CRS model provides a comprehensive assessment of technical efficiency, the VRS model focuses on a more localised pure technical efficiency examination. These models used total expenses as the major input, which included everything from player salary to incidental charges. Outputs were carefully selected, taking into account points awarded, net run rate, profit, and income. Correlation was important in this selection since it provided low repetition among output variables and strong connection with the primary input.

The study's findings were distinct. Efficiency models included the Chennai Super Kings, Delhi Daredevils, and Rajasthan Royals. In contrast, huge spenders like Royal Challenger Bangalore did not always match their spending in terms of efficiency. Surprisingly, the on-field champions of 2009, the Deccan Chargers, struggled with economic efficiency due to high input prices. Singh also emphasised the complex impact of individual player performance on team efficiency, implying that both great players and consistent performers contribute to a club's financial and on-field success.

While Singh's analysis explains IPL's efficiency factors, it is not without limitations. The use of data from a single season may fail to portray the league's dynamic character. Furthermore, while DEA provides a complete perspective, its outcomes could differ from standard success criteria due to its dependency on the inputs and outputs chosen.

2.2 DATA ENVELOPMENT ANALYSIS IN CRICKET WITH ADDITIONAL ANALYSIS.

Gweshe and Durbach (2013) discussed the significance of efficiency in limited overs cricket, particularly given the number of overs limits. Traditional measurements, such as batsmen's runs and bowlers' wickets, are frequently employed to evaluate performance. However, the authors suggest that success in limited overs cricket is not just about scoring runs or taking wickets, but also about doing so efficiently, using as few balls as possible.

The authors assess the efficiency with the players who turned inputs (balls faced or bowled) into performance outputs using Data Envelopment Analysis (DEA) and Stochastic Multicriteria Acceptability Analysis (SMAA). Unwanted inputs, such as runs conceded by bowlers, were employed in the paper. According to the author, efficiency can be improved by either decreasing input or undesired output levels or increasing desirable output levels.

Furthermore, they incorporated non-discretionary inputs, including a test-status component to determine how well-resourced a team is. Teams from test-playing countries typically have greater facilities, training, and financial backing than their counterparts from non-test-playing countries. This element contributes to accounting for these inherent benefits, providing a more complete picture of efficiency accounting for these inherent advantages, offering a more holistic view of efficiency.

The study's data included 140 players from 14 teams, with specific emphasis on batting and bowling performance. Batting study included 77 players (51 big nations and 26 minor nations) who faced more than 100 balls. Bowling analysis includes 76 players (47 major nations and 29 small nations) who bowled over 100 balls, with 21 players appearing in both datasets. Notably, the study's DEA formulation used the Variable Returns to Scale (VRS) technique, ensuring a thorough efficiency assessment.

The author present a unique perspective on player performance, emphasising the importance of efficiency over absolute performance measurements. Their analytical approach, which incorporates both DEA and SMAA, provides significant insights into how athletes optimise their limited game resources, challenging standard ideas and providing a new lens for sports analytics.

In their study, Adhikari et al.'s (2018) went beyond just looking at efficiency as important factor. They also looked at how steady and reliable the team members were when picking the best team using the consistency factor. The efficiency is measured using a modified DEA model, while consistency is assessed using a semi-variance approach with a common reference frame. These measures are aggregated using Shannon's entropy concept, and a novel value index is introduced to determine player importance. The proposed method is applied to select the all-time best one-day international Cricket XI team for a specific time period. The advantages of the proposed measures over existing methods are explained, and a comparative analysis is performed with teams announced by the ICC in 2011 and the BBC in 2015.

Whereas, Yin, Ye and Shah (2023) employed the DEA Super-SBM model to evaluate the player's efficiency in batting, bowling, and fielding departments of all three formats. The authors discuss the limitations of the radial DEA model based on CCR in accounting for the effect of laziness on productivity. To overcome this limitation, they used the SBM (Slack-Based Measure) and super-efficiency models. The super-efficiency SBM model removes the most efficient DMU from the evaluation set and compares the remaining DMUs' performance to it, providing a fair approach for efficiency measurement. The authors used the data of international cricketers (1877-2019).Compared to the traditional parameters, the proposed study indices are more accurate and comprehensive in nature.

2.3 DATA ENVELOPMENT ANALYSIS IN FOOTBALL

Despite football's international appeal, there is a significant lack in scientific study concentrating on its statistical and economic components, particularly when viewed through the perspective of DEA. The groundbreaking study by Tomek and Pelloneová (2023) addresses this gap by focusing on players from two major European leagues: the Czech Fortuna:Liga and the Danish 3F Superliga. Their study encompassed several seasons, from 2015/16 to 2019/20, and provided a complete picture of player performance over time.

An input-oriented DEA model was used with cluster analysis in their analysis. They used the CCR model, which was developed in 1978 by Charnes, Cooper, and Rhodes as one of the basic DEA models. This model focuses on efficiency evaluation by considering constant returns to scale, making it particularly suitable for player evaluation when performance impact may be stable across multiple input scales.

The input variable for the DEA model was the player's market value, which is an important economic indicator in football. Their careful selection of output factors, customised to various player positions, ensured a comprehensive review.

The correlation of output variables was a major consideration for this decision. Three requirements were emphasised by the researchers: minimal correlation within the output variables to avoid redundancy, a high correlation of the output variables with the input variable (market value) to ensure relevance, and an optimal number of variables in relation to the number of evaluated players to maintain analytical robustness.

For instance, forwards were assessed primarily based on offensive metrics like goals and assists, while defenders were through defensive statistics. The strength of their approach lies in the robustness of the DEA model, which allows for the assignment of weights to each input and output variable, ensuring that each player's efficiency is maximized based on the available data. Additionally, the cluster analysis provided further granularity by categorizing players into compact clusters based on their individual characteristics. Forwards, for example, were evaluated mostly on offensive measures such as goals and assists, whilst defenders were evaluated primarily on defensive statistics. The resilience of the DEA model, which allows for the assignment of weights to each input and output variable, ensures that each player's efficiency is maximised based on the available data, is the strength of their method. Furthermore, the cluster analysis added granularity by grouping participants into small clusters based on their particular features.

Arsu (2021) examined the efficiency of 10 elite football clubs from Europe's big-five leagues, including Spain's La Liga, England's Premier League, Italy's Serie A, Germany's Bundesliga, and France's Ligue 1, over five seasons. The study utilized the Bi-objective multi-criteria data envelopment analysis (BiO-MCDEA) to analyze multiple inputs such as social media followers, average viewership, and club market value, against outputs like the UEFA club score and total club revenues. Clubs like Arsenal, Paris Saint-Germain, and Juventus emerged as the most efficient. The research was input-oriented, focusing on minimizing inputs for given outputs, especially in the context of sponsorships. It also incorporated a variable return to scale DEA model, recognizing that returns on sponsorship investments might vary based on the project's scale. The choice of a 5-season analysis ensured a comprehensive and reliable assessment, eliminating potential anomalies from a single season's data.

2.4 EXPLORATORY DATA ANALYSIS IN CRICKET

Kanungo and Tulasi's (2019) paper delves into IPL team and batsman performances through data visualization. Analysing 696 IPL matches from 2008-2018, they employ tables and bar charts to highlight key metrics such as "Maximum man of the matches," "Maximum centuries scored by batsmen," and the "top 10 batsmen with maximum runs." The authors also explore toss decisions, using horizontal bar plots to depict teams' choices post-toss. They emphasize the correlation between high scorers and the choice to bat first, showcasing the toss's strategic implications in cricket matches.

Thomson, Perera, and Swartz (2021) took a new approach to assessing cricket players' performance by highlighting game environment. The histogram, which is produced from all second innings balls in the combined ODI/Twenty20 dataset, is a key component of their research. This histogram, similar to a bar chart, displays the frequency of various game circumstances, highlighting the many challenges teams encounter when attempting to score. Using this as a foundation, they created a visualisation for batsmen's second innings, resulting in the "clutch batting" measure. For bowlers, a comparable "clutch bowling" statistic was developed. Their research, which included 395 ODI and 625 T20 matches from the IPL and BBL between April 2015 and October 2019, showed insights that went beyond typical statistics. The "clutch batting" value captured performance in difficult situations by calculating the area between the "Contextual Batting Function" curve and a baseline ("par line"). The study shed light on complex player performances in various game conditions by using Steve Smith's "clutch batting" value of 0.46 as an example.

2.5 CRITICAL ANALYSIS AND FINDINGS FROM LITERATURE REVIEW

Chaudhary et al. (2019): Introduced a data-driven approach for team selection using Data Envelopment Analysis (DEA). Their method aligns with actual BCCI team selection, but there's a potential trade-off between simplicity and capturing player nuances.

Amin & Sharma: Used DEA to make player evaluations objective, including non-traditional metrics. The paper's concern is its focus on output orientation and limited generalizability due to a specific dataset.

Singh: Combined economic factors with sports results, showing that spending doesn't always equate to efficiency. The limitation is the use of data from just one IPL season.

Gweshe and Durbach (2013): Their study includes a test-status feature for fair team comparisons and analyses a diverse set of players, ensuring unbiased results.

Adhikari et al. (2018): Introduced a semi-variance method to measure player consistency. The method is more suited for longer formats than T20 cricket.

Tomek and Pelloneová (2023): Analysed football using DEA, covering multiple seasons and leagues. Their strength is position-specific evaluation and use of the CCR model.

Kanungo and Tulasi (2019): Presented findings with tables and bar charts, exploring the strategic depth of cricket through toss decisions.

Thomson, Perera, and Swartz (2021): Introduced "clutch batting" and "clutch bowling" metrics, using visual aids like histograms and covered diverse match types.

This summary provides a quick overview of each paper's main points and findings.

2.6 LEARNING FROM THE LITERATURE.

Data Envelopment Analysis (DEA) is a useful tool to evaluate cricket performance. It helps selectors decide the best players for teams by analyzing individual performance. Coaches can also use DEA to monitor and aid player improvement. Additionally, it provides a means to compare the overall efficiency of different teams, offering insights into team strategies. The versatility of DEA allows it to be applied to various datasets, such as those from the IPL.

There are two main methods for using DEA: the CCR Model and the BCC Model. The CCR Model is common, but it might miss some performance nuances. On the other hand, the BCC Model is particularly suited for sports, as it considers the variability in performance under different circumstances. Although it's a bit more complex than the CCR model, modern technology has made it straightforward to use.

One of the most valuable aspects of DEA is benchmarking, which identifies the top players or teams and sets a performance standard for others. Depending on the analysis goal, DEA can focus on either minimizing inputs, like the number of balls played, or maximizing outputs, such as runs scored. To make the results more accessible and understandable, it's beneficial to present them visually, using charts or graphs. In essence, DEA offers a comprehensive way to understand and enhance cricket performance using various models tailored to specific needs.

2.7 Conclusion from literature review

There appears to be a considerable gap in the existing body of work on Data Envelopment Analysis (DEA) applied to sports performance, particularly in cricket. While DEA is a reliable non-parametric method for assessing the effectiveness of decision-making units, its use in cricket has not been adapted to account for the complex responsibilities of players dependent on their batting positions. Especially in shorter formats like T20s, when the game dynamics change quickly, a batsman's position and expectations are significantly dependent on their batting order. Openers are tasked with utilising powerplay overs, middle-order batters are tasked with innings stabilisation, and finishers are expected to increase the run rate as the innings progresses. Each of these roles has its own set of performance criteria. For example, while an opener's average is important, a finisher's strike rate takes centre stage. Failure to separate data based on these batting positions in the current literature might lead to skewed evaluations, perhaps misrepresenting a player's efficiency in their intended function. This disparity highlights the need for a more segmented and position-specific approach to DEA in cricket, ensuring that players are evaluated in the context of their on-field tasks.

Research questions:

1. What effect does batting order have on runs scored?
2. What effect does batting order on strike rate?

3 RESEARCH METHODOLOGY

3.1 DATA

Data collection -This was the foundational step of our research. The initial plan was to get the dataset directly from the Indian Premier League's (IPL) official website. However, difficulties developed as the official website of the IPL restricted data collection.

The data collected was a secondary data set, as its information was gathered for different purpose but now it is being used for different purpose. The dataset was obtained as a substitution from another website, bigbashboard.com. This dataset contains information on 274 IPL players throughout the course of five seasons, from 2018 to 2022.

Data validation: It was essential to ensure the accuracy and dependability of the data obtained from bigbashboard.com. The data was verified with the official IPL website to rectify this. Runs, strike rate, balls, boundaries, and other critical variables were manually confirmed. A random sample of players was picked for this validation process, and their statistics were analysed. This step ensured that the statistics gathered from the secondary source were in alignment with the primary and official data source.

3.1.1 DATA COLLECTION METHOD AND SOFTWARE

Data for each IPL season from 2018 to 2022 was obtained individually from bigbashboard.com. Individual datasets were required because the website split data seasonally rather than a combined dataset covering the entire 2018-2022 period. Excel made data extraction easier. For each year, the "From Web" option in Excel's "Data" tab, where we pasted the website's URL into the designated field to obtain and download the data in XLSX format.

DATA TAB ▷ DATA TOOLS ▷ FROM WEB ▷ URL ▷ DATA CONVERTED TO XLSX

Figure 2: Data Collection through Excel

Consolidation of Data: The collected data was organised into different sheets for each IPL season. To produce a combined list, the "names" columns from each sheet (from 2018 to 2022) were copied and pasted sequentially onto another sheet labelled "Data 2018-2022", beginning with the most recent season and going to the oldest. Then, Excel's duplication removal feature was used to confirm the uniqueness of player names and avoid repetitions, resulting in a combined list of 274 distinct players. As a, result for the seasons 2018-2022, a unified list of player names was developed with deleting all the names that were repeated from every season.

Variable Alignment: Before combining the data, we created different columns for each variable, such as Innings and Runs, for each year between 2018 and 2022. To avoid confusion, we created five separate columns for things like innings and labelled them from the most recent 2022 season down to 2018. We used Excel's VLOOKUP feature to exactly associate each player's name with their corresponding statistics for a certain year because the data was distributed over multiple years.

The SUM function in Excel was used to combine variables like as Runs, Innings, Balls, 100s, 50s, 4s, 6s, outs, and 0s to create a consolidated dataset for the years 2018-2022. This made it easier to compute aggregate and average numbers for each player across the five IPL seasons. Following this organised approach carefully, a comprehensive dataset covering the performances of 274 IPL players from 2018 to 2022 was collected (espn, 2023).

3.1.2 DATA DESCRIPTION

This dataset is a valuable resource for conducting in-depth studies and gaining profound insights into player performances in the Indian Premier League (IPL) over a five-year period." The five-year dataset from 2018 to 2022 was chosen on purpose to provide a more thorough and up-to-date perspective. Because this dataset contains both experienced and beginners, our research is more relevant to the present IPL landscape. Previous research, such as Singh's (2011), relied on single-year datasets to quantify player efficiency, but their breadth was limited. We chose a five-year sample to widen our analysis and include a broader spectrum of player experiences. This span covers the period between two mega auctions in which all players in the IPL franchises are auctioned off, with the exception of the five players retained by each team's management. The most recent IPL auction was in 2022, with the preceding one in 2018. Using this dataset, we may compare player performance between two mega auctions and get new insights into the season that follows the last mega auction. This method provides a holistic perspective of player dynamics in the IPL, taking both established stars and growing potential into account."

Inn (Innings): The number of times a player has come out to bat. It may differ from the total number of matches if the player did not bat in some games.

NO (Not Out): The number of times a player was still batting at the end of an inning. This could be due to their side completing the allotted overs or reaching the target score before the player was dismissed.

Runs: This is a player's Total score over all innings over five seasons. It is the key metric for determining a batsman's contribution to the team.

HS (Highest Score): This is the player's best individual performance in a single inning. It demonstrates a player's maximum potential.

Ave (Average): An important metric for batsmen, the average represents a player's regular performance per innings. A higher average indicates more consistent performance.

Balls: The total number of deliveries faced by a player. It gives context to the runs scored, indicating whether they were completed quickly or over a longer period of time.

Strike Rate (SR): A measure of a player's scoring rate. A higher strike rate suggests aggressive batting, while a lower rate might indicate a more defensive approach.

Ave (Average): An important metric for batsmen, the average represents a player's regular performance per innings. A higher average indicates more consistent performance.

Balls: The total number of deliveries faced by a player. It gives context to the runs scored, indicating whether they were completed quickly or over a longer period of time.

Strike Rate (SR): A measure of a player's scoring rate. A higher strike rate suggests aggressive batting, while a lower rate might indicate a more defensive approach.

Cricket's milestones are 100 (centuries) and 50 (half centuries). A century or half-century is a known accomplishment that demonstrates a player's ability to play long innings.

30+ (Thirty Plus): This statistic emphasises consistency. Scores of 30 or more runs are frequently considered as exceptional.

0 (Zeroes): Represents the number of times a player was dismissed without scoring. It can be an indicator of a player's form or consistency.

4s (Fours) & 6s (Sixes): These are the primary ways to score boundaries. A higher count indicates a player's ability to clear the field.

Bpb, Bp4, & Bp6: These metrics delve into a player's efficiency and frequency in hitting boundaries. They provide insights into a player's style and strengths.

PrB, Pr4, & Pr6: These percentages give a clearer idea of a player's approach to scoring runs, whether they rely on boundaries or rushing between the wickets. These include the percentage of runs scored in boundaries as Prb, the percentage of runs scored in sixes as Pr6, and the percentage of runs scored in fours as Pr4.

BBP: This metric offers a perspective on a player's aggressiveness and ability to dominate bowlers by frequently hitting boundaries.

RpB & RpNB: These provide a breakdown of a player's scoring patterns, highlighting their dependency on boundaries versus other scoring methods.

RpI & BpI: These averages give insights into a player's typical contribution in each innings and their patience or aggression levels. RpI stands for runs per innings whereas balls stands Balls per innings.

Basra: This recognizes standout performances, highlighting players who score rapidly even in short appearances.

ABP (Average Batting Position): This indicates the role or responsibility of a player in the batting lineup. Top-order players often face more deliveries and are expected to set the game's pace, while middle or lower-order players might play more finishing roles.

3.2 DATA VISUALISATION

Data visualisation is the technique of presenting information and data using graphical components such as charts, graphs, and maps. These tools not only make complex data accessible, but they also show relationships, differences, and connections within the data, allowing viewers to clearly discover trends, anomalies, and underlying patterns on and data. By using visual elements like charts, graphs and maps, data visualisation tools provide an accessible way to see and understand trends, outliers and patterns in data.

3.2.1 VISUALISATION THAT WILL BE USED IN THE DATA SET

1.Histogram

A histogram is like a bar chart that groups numbers into ranges. The bottom of the chart shows these number ranges, and the height of each bar shows how many times numbers in that range appear in our data. Sometimes, our data isn't exact and can be a bit "fuzzy" (Viertl, 2006). Histograms help us see these fuzzy details better. To make a histogram, we break our data into parts called bins. There's no one right way to pick the size of these bins, but it's important they're not too small or big. One key thing to remember is that the size of the bar area (not just how tall it is) tells us how many times numbers in that bin show up (Anon., n.d.).

Types of histograms and description of the histogram:

Normal Histogram: This is like a bell-shaped chart where most of the data gathers in the middle. Both sides of the middle point usually have matching data. Bimodal Distribution Histogram: This chart has two main peaks. It's like having two bell-shaped charts combined. Each peak can be looked at like its own normal histogram. Right Skewed Histogram: In this chart, most of the data is on the left, with less on the right. It leans more to the left.

Left Skewed Histogram: Here, most of the data is on the right, with less on the left. It leans more to the right.

Histograms will be used in our dataset to get insight into the distribution patterns of various metrics. We can easily spot any skewness and precisely analyse the qualities and tendencies within our measurements by visualising this data. This method will help us comprehend and lead our later analyses. A skewness value greater than 1 or less than -1 indicates high skewness. A skewness value between 0.5 and 1 (or between -0.5 and -1) indicates moderate skewness. A skewness value between -0.5 and 0.5 indicates approximately symmetric distribution. Although, to determine the skewness of our dataset, we'll utilize the skew() function provided by the pandas library. This method offers a straightforward way to assess data asymmetry.

Histogram helps in analysing the outliers, to analyse the potential outlier through histogram we will use the interquartile range (IQR). The interquartile range (IQR) is a measure of statistical dispersion, or in simpler terms, it represents the spread of the middle 50% of the data.

The IQR is calculated as:

$$IQR=Q3-Q1$$

$$IQR=Q3-Q1$$

Where:

Q3 = third quartile (the 75th percentile)

Q1 = first quartile (the 25th percentile)

The IQR can also be used to identify outliers. A common method is:

Lower Limit: Q1−1.5 × IQRQ1−1.5×IQR

Upper Limit: Q3+1.5 × IQRQ3+1.5×IQR

To identify potential outliers within our dataset, we'll utilize the quantile function from the pandas library. This function aids in determining the data's quartiles, which can then be used to calculate the Interquartile Range (IQR). Data points that fall outside the bounds of the IQR are typically considered "**outliers**".

The list of histograms that we are going to create to get insights are as follow:

1. RUNS SCORED

2. BALLS PLAYED

3. STRIKE RATE

4. BALLS FACED

5. BATTING AVERAGE

1. **RUNS**

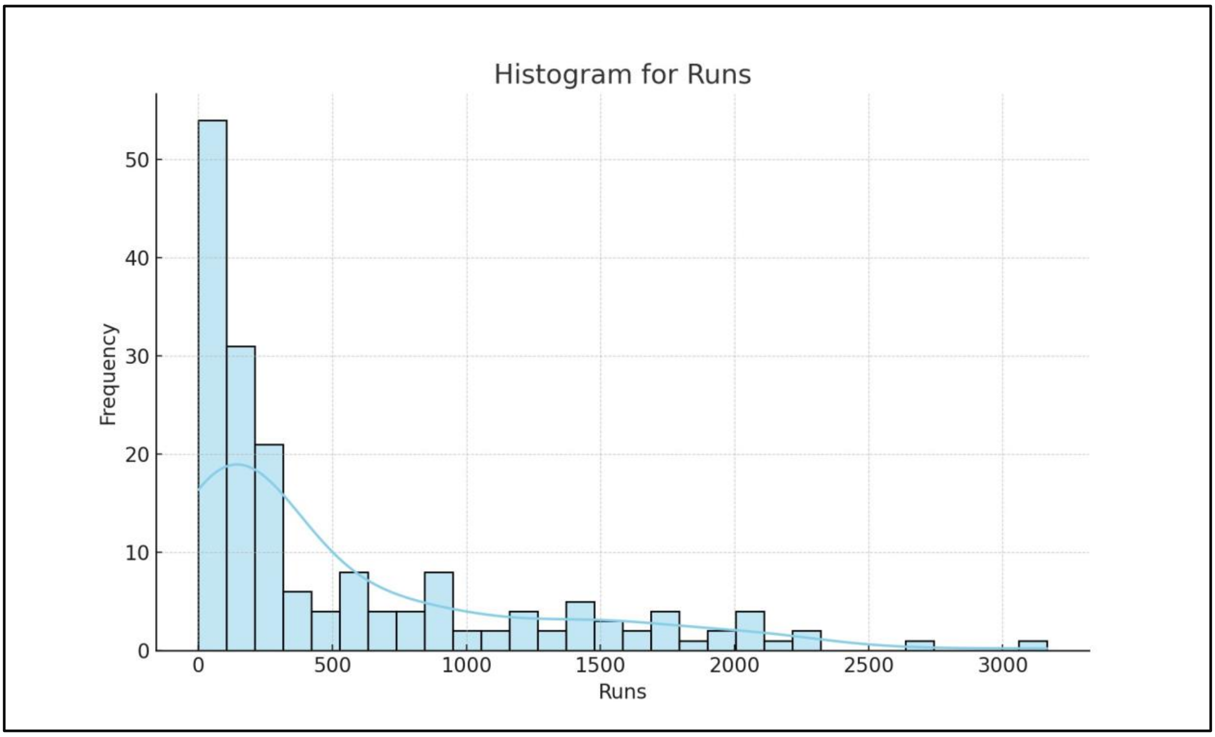


Figure 3: Histogram for Runs

Analysis: The "Runs" column has a right-skewed distribution with a skewness value of 1.71, indicating that while many players score average runs, a few consistently score highly. This skewness can be attributed to factors like batting position, opportunities, or inherent talent.

Potential Outliers: Players like Virat Kohli, Jos Buttler, and Lokesh Rahul have significantly higher runs. These outliers, being the main scorers, showcase their importance to their teams. Being an outlier here highlights their exceptional run-scoring ability.

The outliers in this section contained all the ones who are openers hence we can find a relation that top scorers are openers amongst the openers.

2. BALLS FACED.

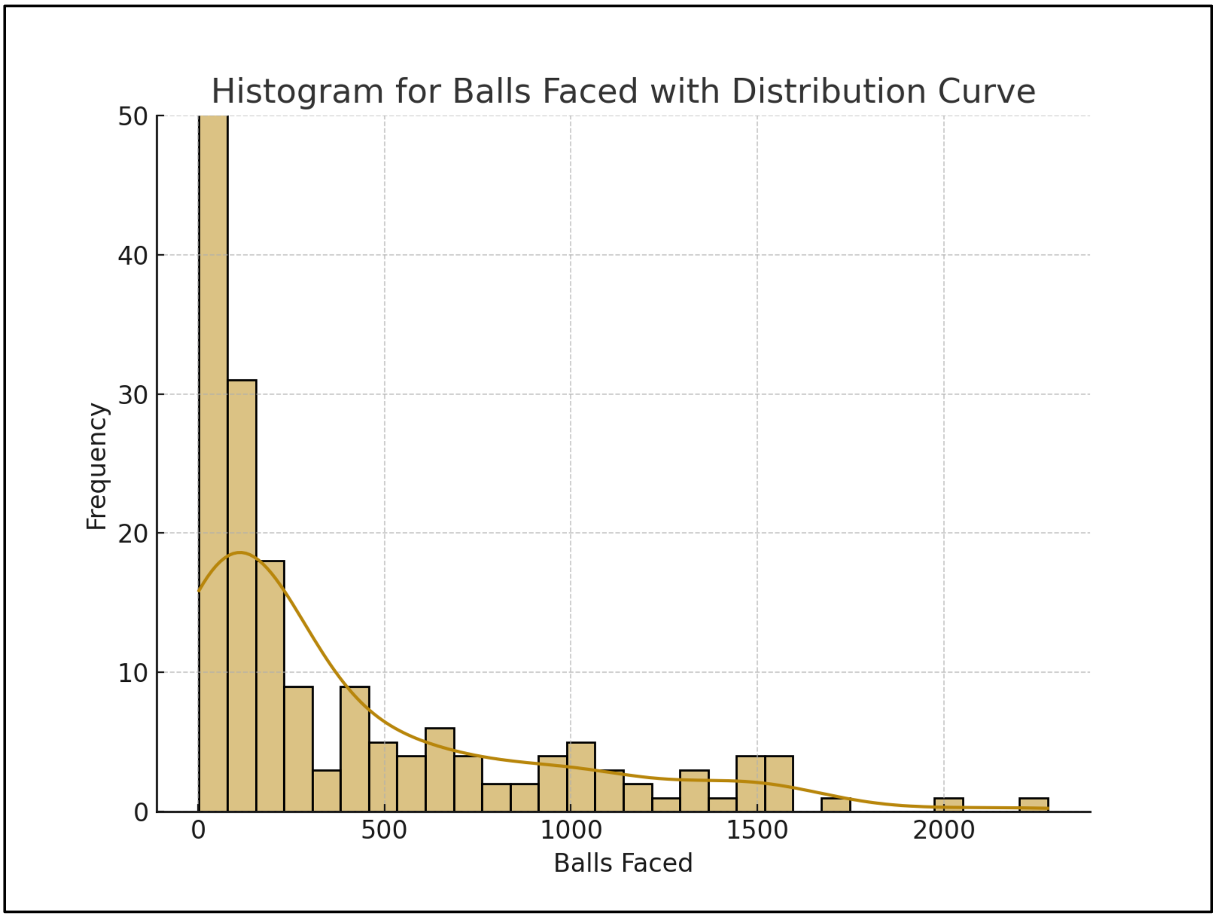


Figure 4: Histogram for Balls Faced with Distribution Curve

Observation: The "Balls Faced" histogram is right-skewed, supported by a skewness value of 1.55.

Descriptive Evaluation: Most players face a typical number of balls, but a few face significantly more. This suggests that while many have average innings durations, only a select few consistently play longer innings, reflecting their importance to their teams.

Potential Outliers: Notable players like Jos Buttler, Lokesh Rahul, and Virat Kohli have faced significantly more balls than others.

Analysis: These outliers are mainly top-order batsmen, indicating their role in anchoring innings for their teams. Facing more balls in T20 often corresponds to the roles of openers or top-order batsmen. Being an outlier here emphasizes a player's significance and reliability to the team.

3. STRIKE RATE

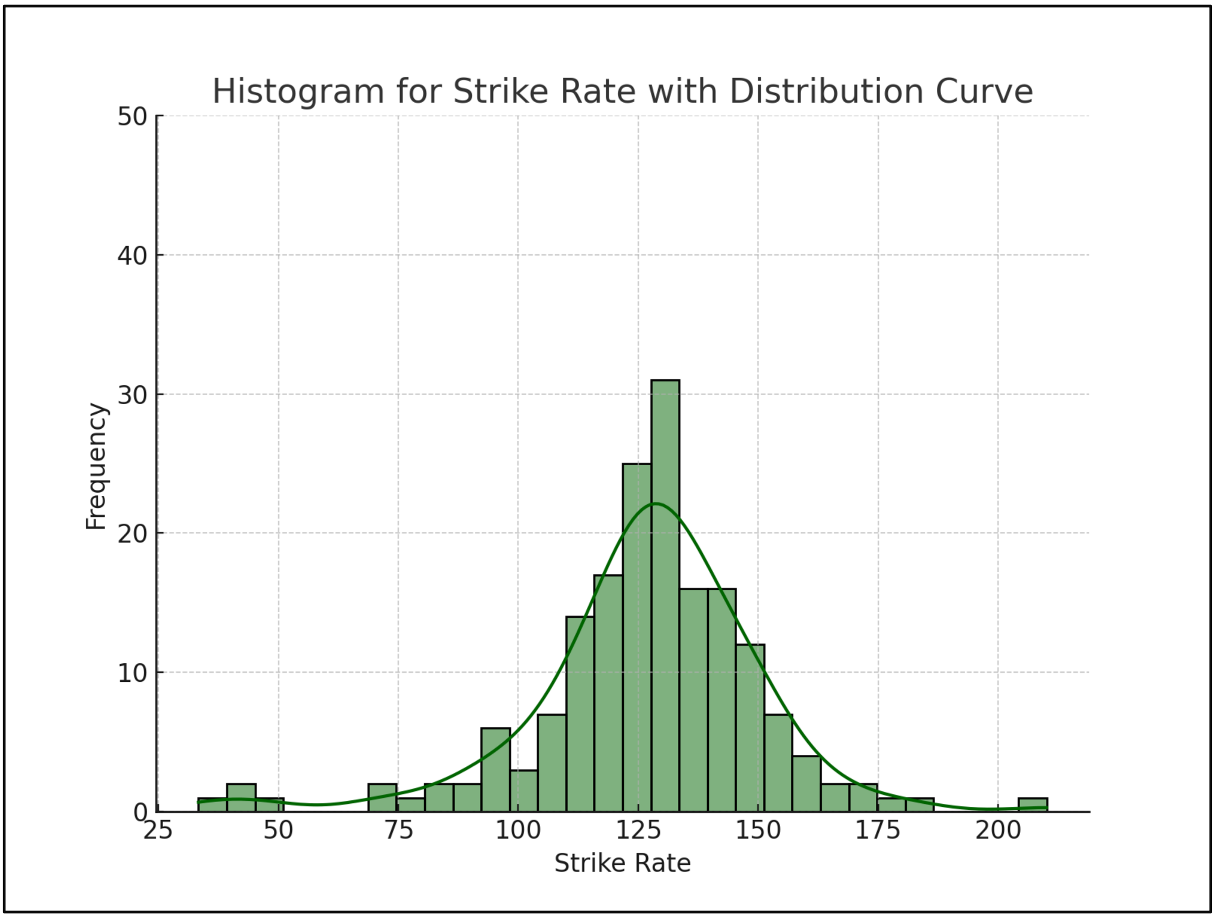


Figure 5: Histogram for Strike Rate with Distribution Curve

Observation: The "Strike Rate" histogram is left-skewed, confirmed by a skewness value of -0.85.

Analysis: Most players have similar strike rates, but a few have significantly lower rates.

Potential Outliers: Notable outliers in strike rates include players like Andre Russell, Tim David, and others.

Analysis: These outliers consist of both aggressive batsmen and bowlers with limited batting chances. The presence of bowlers can skew the analysis, as they're not primarily known for their batting.

It is critical to consider the amount of balls encountered by each player while analysing the strike rate.

.4. **AVERAGE**

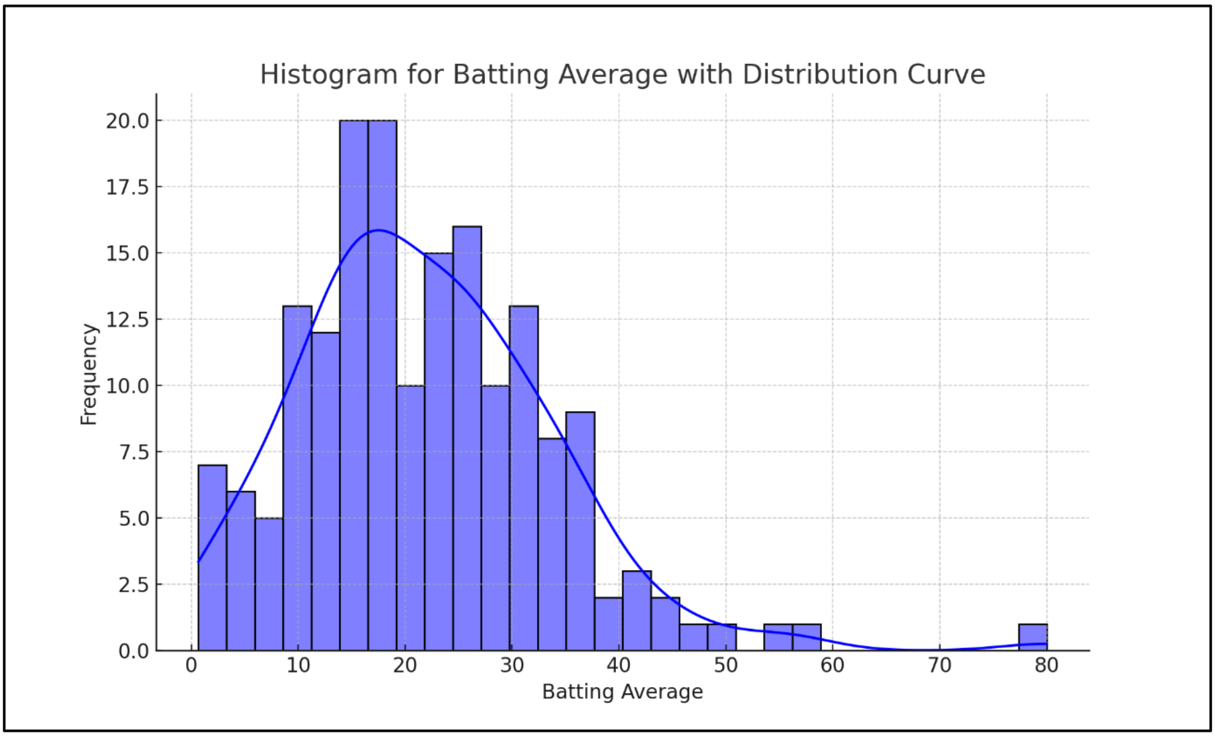


Figure 6: Histogram for Batting Average with Distribution Curve

Histogram & Skewness:

Observation: The "Average" histogram displays a rightward shift, with a skewness value of 1.57. Outliers in the batting averages include notable players like Virat Kohli, KL Rahul, and others.

Analysis: Most players have moderate batting averages due to the aggressive nature of T20 cricket. The outliers are top-order batsmen known for consistent performances across seasons. Their high averages underline their cricketing prowess.

Critical Analysis: While T20 emphasizes strike rate, a high average provides stability to the team. For holistic player assessment, consider combining average with other metrics like strike rate and total runs.

3.3 DATA SEGMENTATION

After looking at the histogram and distribution of runs other metrics we can say that its appropriate to segment our data set.

To evaluate the performance of batsmen at different batting positions, the dataset must be segmented based on their respective batting positions. We may accomplish this by utilising the dataset's ABP (Average Batting Position) measure. This method allows for a more precise study of each batsman's contribution at various points of the game. The segmented data is referred from https://en.wikipedia.org/ as no literature mentioned about bating position. Moreover, did little variations in the from the referred site.

The dataset will be segmented for more detailed analysis-

- Batting Positions 1-2: Opener
- Batting Positions 3-4 in the Top Middle Order
- Positions 5-6 in the lower middle order of the batting order
- All-Rounder Bowler: Batting Positions 7-8

This classification allows us to assess each batsman's performance depending on their individual function and location in the lineup.

3.4 Bar plot.

A bar plot, also known as a bar chart or bar graph, is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. A bar plot or bar graph is used to compare discrete groups. Different graph types might differ in group comparison due to differences in underlying graph schemas. In the study by Zhao and Gaschler (2021), they examined the efficiency of bar graphs in comparison to dot graphs and tally charts for representing discrete groups. The results of their study suggested that tally charts were more efficient for group comparison compared to bar graphs. However, when bar and dot graphs were paired together, they showed no mixing costs, indicating that they might be interchangeable in certain contexts.

In our dataset, we will use bar graphs to evaluate the performance metrics across various batting positions." Bar graphs excel at depicting comparisons between distinct categories. In the context of batting, one axis will identify certain batting positions, such as opener, middle-order, or tail-ender, while the other axis will measure performance indicators, such as average runs scored, strike rate, or boundaries hit. We may see the variations and trends in batting performance between positions by using a bar graph. This aids in recognising the strengths and limitations of specific positions, but also provides insights into strategic decisions.

3.4.1 Bar plot Visualisation.

Average runs as per batting order:

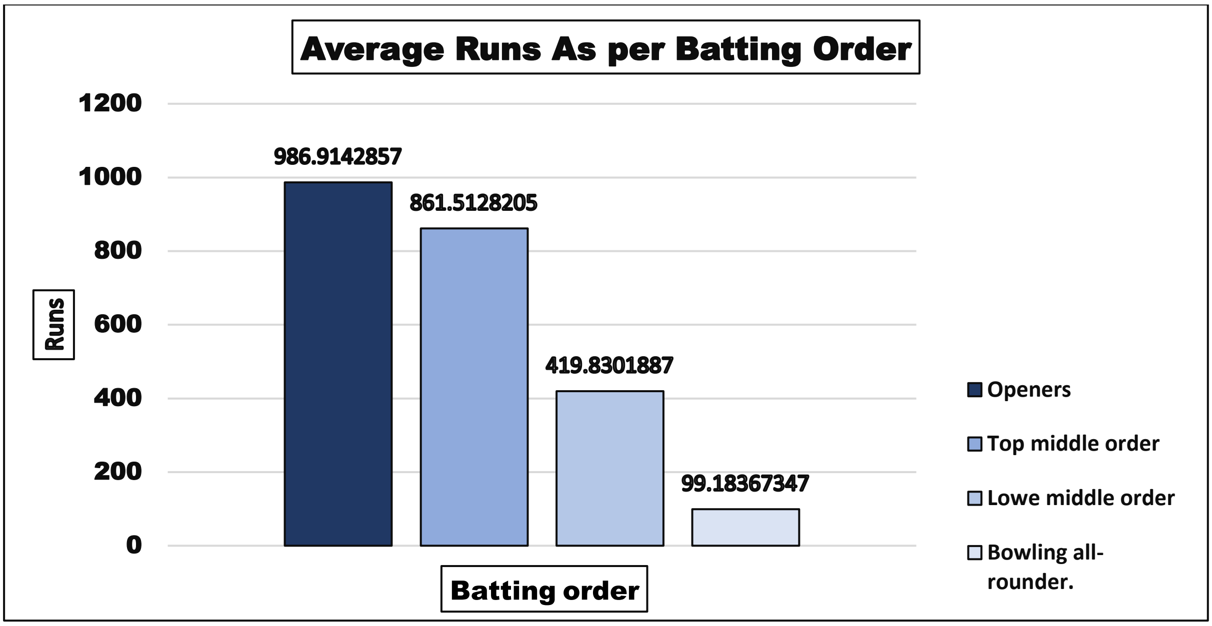


Figure 7: Average Runs as per batting order

The chart illustrates disparity in average number of runs scored across various batting positions. Notably, openers have amassed a remarkable 987 average number of runs, which is approximately ten times higher than the 99 average number of runs achieved by bowling all-rounders. Furthermore, the contrast between the top middle order and the lower middle order is substantial, with the top middle order contributing around 861 average number of runs, more than twice the 420 runs recorded by the lower middle order. This big gap in average number of runs scored highlights how where a player bats in the order can greatly affect the number of runs, they make.

Average strike rate by batting position:

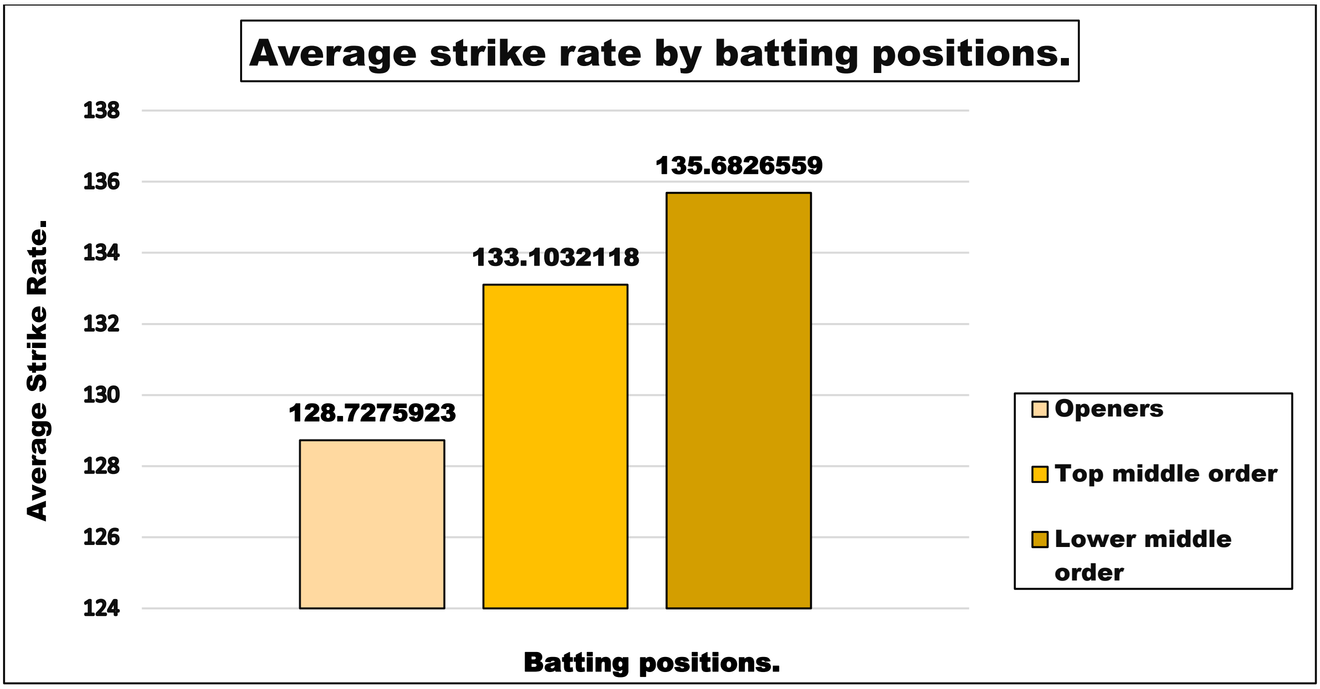


Figure 8: Average Strike Rate by batting position

The above figure provides valuable insights into how batting positions influence strike rates, shedding light on the strategic nuances of cricket. It showcases the evolving roles of players as they progress through the batting order, from providing stability at the start to adopting a more aggressive approach as the innings progresses. The lower middle order has an average strike rate approximately 7% higher than that of the openers. This percentage difference underscores the contrasting roles and approaches of these two batting groups

Openers (Batting Position 1): The players who start the game have an average score rate of around -128.727. This means they usually play more carefully to set up a strong beginning for the team.

Top Middle Order (Batting Positions 2-4): The next set of players have an average score rate of about -133.103. They try to balance between scoring and not getting out quickly, building on the start.

Lower Middle Order (Batting Positions 5-7): Players in the lower middle order score the fastest, with an average rate of around -135.682. They step up when it's time to speed things up or recover from a tough situation.

To highlight the variation in strike rates among batsmen at different batting positions more effectively, we will examine the strike rates of the top 10 batsmen in each position. By focusing on the top 10 batsmen in each position, will help us gain more comprehensive picture of how different batting roles influence strike rates and overall team dynamics.

Average strike rate of top 10 batsmen by batting position:

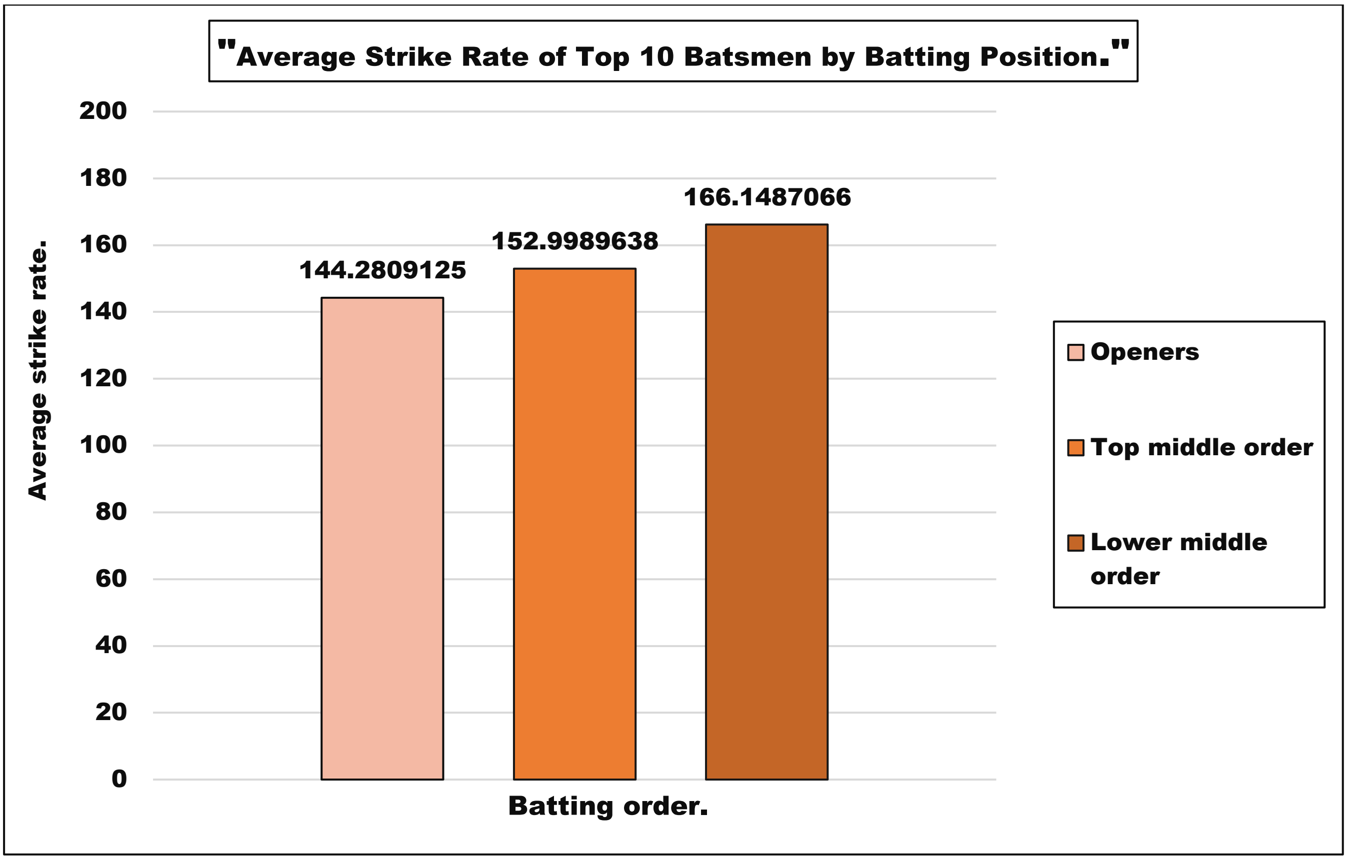


Figure 9: Average Strike Rate of top 10 Batsmen by batting position

When we look at the top 10 batsmen with highest strike rate in each position, we find something interesting.

The lower middle order batsmen have a higher average strike rate, meaning they tend to score runs faster. Their average strike rate is around 166.14. They play aggressively in the middle of the innings to increase the team's score. On the other hand, openers have a lower average strike rate, about 144.2. They play more cautiously at the start of the game to make sure the team gets off to a steady beginning. There's big difference of more than 22 units in strike rate between these two groups. shows how their roles in the game differ. The lower middle order accelerates the scoring, while the openers focus on stability. The top middle order's strike rate is approximately 9 units higher than that of the openers, highlighting their ability to maintain a balanced approach. However, it is about 13 units less than the striking rate of the lower middle order.

Balls faced as per batting order:

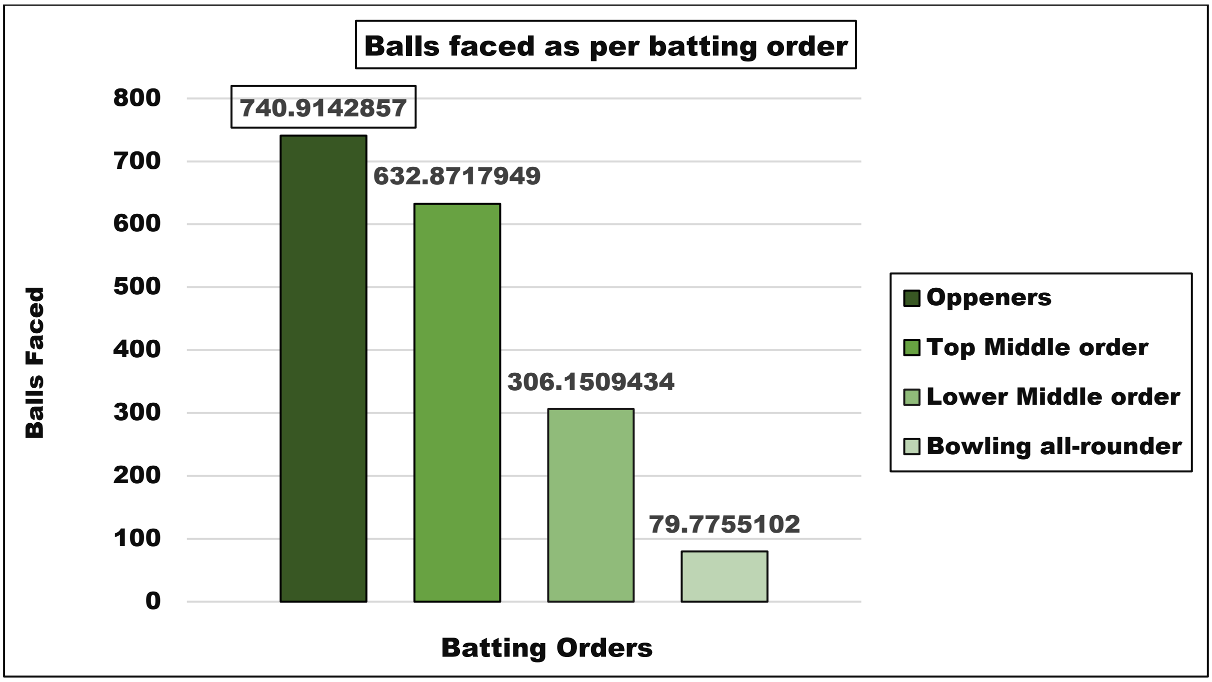


Figure 10: Balls faced as per batting order

The graph reflects a clear contrast in the average number of balls faced by players in different batting positions. Openers and top middle-order batsmen faced significantly more balls compared to the lower middle order and bowling all-rounders. Specifically, openers faced a substantial 741 average number of balls, while top middle-order players encountered 633 average number of balls. In contrast, the lower middle-order batters faced just 306 balls, and bowling all-rounders faced the least, with only 80 average number of balls.

This data emphasizes that openers and top middle-order players had more opportunities to score runs because they faced many deliveries, whereas the lower middle order and bowling all-rounders had fewer chances due to their lower ball-faced counts.

Percentage of runs in sixes by batting position:

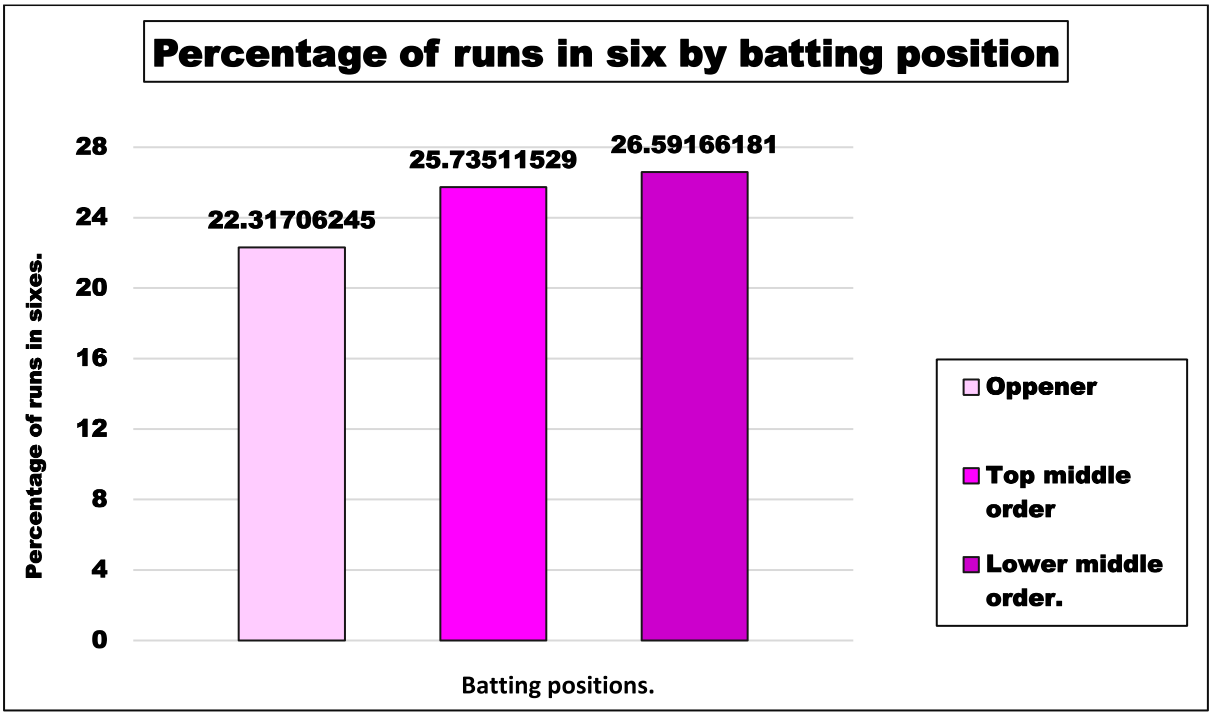


Figure 11: Percentage of runs in six by batting position

Examining the chart above, we can clearly observe a trend in the percentage of runs scored in sixes as the batting order shifts.

Specifically, the top order batsmen contribute an average of 22.317% of their runs via sixes, while the middle-order batsmen increase this figure to 26% on average. Remarkably, the lower-middle order surpasses both of these groups by scoring an impressive 27% of their runs through sixes on average. This analysis highlights how the way runs are scored changes as we move through the batting order, showing that the lower-middle order tends to hit a lot more sixes.

To further understand how the batting order impacts the percentage of runs scored as sixes and show variations, we'll take a closer look at a chart that shows the percentage of runs scored as sixes for the top 10 batsmen. To further understand how the batting order impacts the percentage of runs scored as sixes and show variations, we'll take a closer look at a chart that shows the percentage of runs scored as sixes for the top 10 batsmen.

Percentage of Runs in Sixes by Batting Position for Top 10 Batsmen:

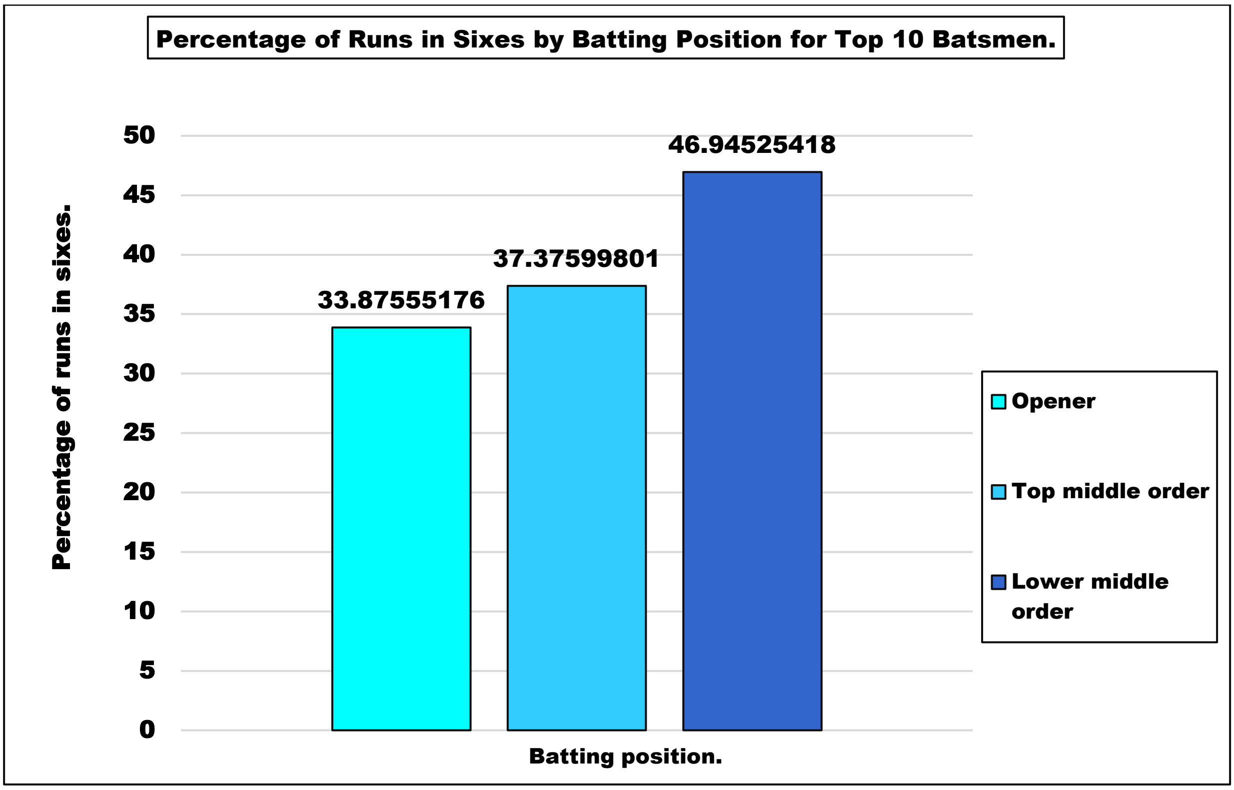


Figure 12: Percentage of Runs in sixes by batting position for top 10 batsmen

Looking at the chart above, we notice a clear trend among the top 10 batsmen across different batting positions.

The lower-order batsmen stand out by scoring a significant 47% of their runs through sixes. In contrast, the top-order batsmen achieve approximately 34% of their runs in sixes, which is notably 13% less than their counterparts in the lower-middle order.

Moving on to the top middle order, we observe an average of 37% of runs coming from sixes. This figure falls approximately 10% short of the lower-middle order but surpasses the openers by 3%.

In summary, this data paints a vivid picture of how different batting positions influence the percentage of runs scored as sixes among the top 10 batsmen. The lower-middle order emerges as the most proficient in this aspect, while the top order and top middle order demonstrate their own distinct patterns.

3.5 HYPOTHESIS TESTING

Hypothesis testing is a method of using data to choose between two concepts or assumptions about a broader population. We begin with a simple guess (called the null hypothesis) and another guess to compare it to (the alternative hypothesis). We then examine the data to determine which guess is most likely correct. In layman's words, it's like seeing if what we see fits what we expect and then judging whether our guess is correct (Dennis et al., 2019).

We'll use hypothesis testing to find out the effect of specific metrics on the batting order. Our major goal is to see if these indicators differ considerably among batting positions. To accomplish this, we will test two hypotheses:

The null hypothesis (H0): states that there is not a significant difference in the measure between batting orders 'a' and 'b'.

Alternative Hypothesis (Ha): The measure differs significantly across these batting orders.

The p-value will determine our selection. If the p-value is less than 0.05, showing that there is less than a 5% probability that the observed results occurred by chance alone, we will reject the null hypothesis, indicating a significant difference at a 95% confidence level.

3.5.1 Independent T – test

The t-test is a statistical method used to compare the means of two groups. When conducting the t-test, it is assumed that the data is measured on an interval or ratio scale, obtained through simple random sampling, and is normally distributed. Additionally, an appropriate sample size is essential for the validity of the test. One of the primary assumptions is the homogeneity of variances, which means that the variances of the two groups being compared should be equal. However, it's worth noting that there are versions of the t-test, such as Welch's t-test, that are designed to be used when the assumption of equal variances is not met (Delacre, Lakens, & Leys, 2017).

The t-test is a technique that allows us to determine whether two groups are truly different based on their average results. We will use the t-test as a statistical hypothesis testing approach for our study to confirm differences in several batting metrics across two different batting orders. This decision is influenced by insights gained from multiple visualisations we've created. We intend to use the t-test to determine how different batting order performances differ, as well as how their metrics and analytics differ. This will help us understand the differences between the two batting orders.

3.5.2 HYPOTHESIS TESTING 1

Hypothesis testing for average runs:

Looking at the chart of average runs as per batting order we can see that there is substantial difference in runs between the Openers, top middle order compared to lower middle order, although very less difference openers between to validate that we will conduct a hypothesis test of independent t sample test between

1. Openers and lower middle order.

2. Openers and top middle order.

1.**<u>Openers and lower middle order</u>**:

Formulation of hypothesis

Ho - The average runs scored by Openers and Lower Middle Order are the same.

Ha -The average runs scored by Openers and Lower Middle Order are significantly different.

Results of hypothesis testing 1:

Group Statistics

	Batting position 2	N	Mean	Std. Deviation	Std. Error Mean
Runs	1	35	986.91	882.102	149.102
	2	53	419.83	410.158	56.339

Figure 13: Hypothesis Testing Group Statistics

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Runs	Equal variances assumed	32.499	<.001	4.069	86	<.001	<.001	567.084	139.352	290.062	844.106
	Equal variances not assumed			3.558	43.818	<.001	<.001	567.084	159.392	245.814	888.354

Figure 14: Independent Samples Test

1. Formulating the Hypothesis:

Null Hypothesis (H0): There is no significant difference in the average runs scored by the Openers and the lower middle order.

Alternative Hypothesis (H1): There is a significant difference in the average runs scored by the Openers and the lower middle order.

2. Levene's Test for Equality of Variances:

Before conducting the t-test, it's common to check if the variances of the two groups are equal using Levene's test.

F-value: 32.499 Significance (p-value): <.001

Since the p-value is less than 0.05, we reject the null hypothesis of Levene's test, suggesting that the variances of the two groups are significantly different. Therefore, we should not assume equal variances for the t-test.

3. T-test for Equality of Means:

t-value (Equal variances not assumed): 3.558 Degrees of Freedom (df): 43.818 Two-Sided p-value: <.001

The two-sided p-value is less than 0.05, which means we reject the null hypothesis of the t-test. This suggests that there is a statistically significant difference in the average runs scored by the Openers and the lower middle order.

4. Confidence Interval:

The 95% confidence interval for the difference in means (assuming unequal variances) ranges from 245.814 to 888.354. Since this interval does not include zero, it further supports the conclusion that there is a significant difference between the two groups.

Explanation:

Based on the results of the independent samples t-test, we can conclude that there is a statistically significant difference in the average runs scored by the Openers and the lower middle order. Specifically, the Openers score, on average, 567.084 runs more than the lower middle order. This difference is statistically significant, as indicated by the very low p-value.

In the context of the game, this suggests that the Openers tend to perform better in terms of scoring runs compared to the lower middle order.

2. Openers Vs top Middle order

Group Statistics

	Batting position 2	N	Mean	Std. Deviation	Std. Error Mean
Runs	1	35	986.91	882.102	149.102
	2	39	861.51	690.694	110.600

Figure 15: Group Statistics

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Runs	Equal variances assumed	2.812	.098	.684	72	.248	.496	125.401	183.219	-239.840	490.643
	Equal variances not assumed			.675	64.293	.251	.502	125.401	185.644	-245.433	496.236

Figure 16: Independent Samples Test

1. Formulating the Hypothesis:

Null Hypothesis (H0): There is no significant difference in the average runs scored by the Openers and the top middle order.

Alternative Hypothesis (H1): There is a significant difference in the average runs scored by the Openers and the top middle order.

2. Levene's Test for Equality of Variances:

Before conducting the t-test, it's common to check if the variances of the two groups are equal using Levene's test.

F-value: 2.812 Significance (p-value): 0.098

Since the p-value is greater than 0.05, we fail to reject the null hypothesis of Levene's test, suggesting that the variances of the two groups are not significantly different. Therefore, we can assume equal variances for the t-test.

3. T-test for Equality of Means:

t-value (Equal variances assumed): 0.684 Degrees of Freedom (df): 72 Two-Sided p-value: 0.496

The two-sided p-value is greater than 0.05, which means we fail to reject the null hypothesis of the t-test. This suggests that there is no statistically significant difference in the average runs scored by the Openers and the top middle order.

4. **Confidence Interval:**

The 95% confidence interval for the difference in means ranges from -239.840 to 490.643. Since this interval includes zero, it further supports the conclusion that there is no significant difference between the two groups.

Explanation:

Based on the results of the independent samples t-test, we can conclude that there is no statistically significant difference in the average runs scored by the Openers and the top middle order. This means that, statistically speaking, the two groups perform similarly in terms of scoring runs.

3.5.3 HYPOTHESIS TESTING 2.

By examining the charts depicting the average strike rate of the top 10 batsmen, it becomes evident that there exists a significant disparity in strike rates among various batting order of lower middle order. To substantiate these visual observations, we intend to conduct an independent sample t-test, aiming to statistically validate the observed differences.

1.**Openers lower and lower middle order:**

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
SR	Equal variances assumed	3.123	.081	-2.000	86	.024	.049	-8.065676550	4.0329598309	-16.08293535	-.0484177496
	Equal variances not assumed			-2.157	85.879	.017	.034	-8.065676550	3.7397968562	-15.50029552	-.6310575822

Figure 17: Hypothesis Test 2 Independent Samples Test

1. Formulating the Hypothesis:

Null Hypothesis (H0): There is no significant difference in the strike rate (SR) of the Openers and the lower middle order.

Alternative Hypothesis (H1): There is a significant difference in the strike rate (SR) of the Openers and the lower middle order.

2. Levene's Test for Equality of Variances:

Before conducting the t-test, it's common to check if the variances of the two groups are equal using Levene's test.

F-value: 3.123 Significance (p-value): 0.081

Since the p-value is greater than 0.05, we fail to reject the null hypothesis of Levene's test, suggesting that the variances of the two groups are not significantly different. Therefore, we can assume equal variances for the t-test.

3. T-test for Equality of Means:

t-value (Equal variances assumed): -2.000 Degrees of Freedom (df): 86 Two-Sided p-value: 0.049

The two-sided p-value is slightly less than 0.05, which means we reject the null hypothesis of the t-test when assuming equal variances. This suggests that there is a statistically significant difference in the strike rate (SR) of the Openers and the lower middle order.

4. Confidence Interval:

The 95% confidence interval for the difference in means (assuming equal variances) ranges from -16.083 to -0.048. Since this interval does not include zero, it further supports the conclusion that there is a significant difference between the two groups.

Explanation:

Based on the results of the independent samples t-test, we can conclude that there is a statistically significant difference in the strike rate (SR) between the Openers and the lower middle order. Specifically, the Openers have a strike rate that is, on average, 8.066 units lower than the lower middle order. This difference is statistically significant, as indicated by **the p-value being less than 0.05**.

3.6 RESEARCH QUESTION ANSWERS

Q1 What effect does batting order have on runs scored?

The batting order does have a significant effect on the runs scored. Specifically, the Openers tend to score significantly more runs on average compared to the lower middle order. The difference in average runs is 567.084, and this difference is statistically significant with a very low p-value (<.001).

In simpler terms, players who bat in the opening positions tend to score more runs than those who bat in the lower middle order positions. This suggests that the batting order plays a crucial role in determining the performance of the batsmen in terms of runs scored.

On the other hand through bar plot and hypothesis testing we can confirm that Openers and the top middle order batsmen have comparable performances.

The Openers significantly outperform the lower middle order in terms of runs scored. This indicates that the batting order plays a crucial role in determining the performance of the batsmen, with Openers generally being more successful in accumulating runs than the lower middle order.

Q2 What effects does batting order have on strike rate?

It is evident that there's a significant disparity in the strike rates of openers compared to the lower middle order. However, the difference between the strike rates of the top middle order and the lower middle order is marginal, as indicated by the charts.

Looking at the charts and hypothesis testing we can confirm that there is significant difference among the various stats and we can proceed for our final section data envelopment analysis

We observed through chart analysis and hypothesis testing that the upper middle order not only scores more runs but also has a higher strike rate. This emphasises the uniqueness of their performance in comparison to other batting positions.

3.7 DATA ENVELOPMENT ANALYSIS

3.7.1 SELECTION OF INPUT AND OUTPUT USING CORRELATION ANALYSIS

Correlation analysis is a technique for determining the level and direction of a linear relationship between two numerical variables. The correlation coefficient, which can range from -1 to 1, measures the strength of the association. A positive correlation means that as one variable increases, so does the other, and vice versa. A negative correlation means that as one variable rises, the other falls. A correlation coefficient around 1 or -1 denotes a strong association, whereas a coefficient near 0 denotes a weak relationship. (Saleh, Salem, & Ela strikebd, 2020).

Correlation Analysis in Data Envelopment Analysis (DEA):

Correlation analysis in DEA is a technique used to understand the relationship between multiple variables or factors in a system. By employing DEA, the efficiency and effectiveness of these interactions can be quantified, helping to pinpoint areas of improvement and potential synergies (Lv, Zhao, & Liu, 2022).

We will use correlation analysis to discover the links between various variables in our dataset, such as batting average and strike rate. By determining the nature of these correlations—whether positive or negative—we can make more strategic decisions about what variables to prioritise as inputs or outputs. Gaining insight into these connections not only broadens our awareness of how different measures influence one another, but it also enriches our modelling approach. Our goal is to refine and improve the precision of our decision-making by utilising the power of correlation analysis.

3.7.2 CONDUCTING CORRELATION ANALYSIS

To execute this analysis, we'll utilize Python's renowned pandas' library, supplemented with NumPy. The correlation among metrics will be computed using the formula:

correlation matrix = data. Orr()

3.7.3 VISUALISATION OF CORRELATION

Correlation analysis through visualization is like using charts and graphics to understand how different pieces of data relate to each other. Imagine you have a lot of numbers in a table, this method turns those numbers into a visual picture, like a web or network, making it easier to see patterns. Freeman et al. (2022).

We will be using visualisation to find out the connections between various cricket metrics. By transforming this data into visual representations, it becomes simpler to identify patterns. For instance, we can swiftly determine the relationship between a batsman's average runs and the number of games they've participated. This method provides a more immediate and clear way to understand these relationships.

Scatter plot:

A scatter plot is like a map for data points. It shows how two things relate by placing dots on a chart. For instance, if you want to see if taller people tend to weigh more, you'd put height on one side and weight on the other. If the dots form a line going up, it means as one thing increases, so does the other. If the dots are all over the place, it means there's no clear connection between the two (based on Watanabe & Mizukami, 2018). .

We'll use scatter plots to show how two cricket metrics relate. Consider, making a graph of a batsman's average runs against the number of matches they've played. The dots making a rising path from the bottom left to the top right suggest that batsmen score better as they play more matches. On the other hand, if the dots are scattered with no clear direction, it suggests that there isn't a clear relationship between the two variables. In similar ways our real data can be assed, by the logic of upward and downward trend of the dots.

This section dives into the critical process of selecting the best inputs and outputs for our efficiency analysis model. The choice is not random; it is based on two fundamental pillars: Results of Correlation Analysis: Before finalising the inputs and outputs, we performed a comprehensive correlation analysis. This study aided in recognising of connections between numerous potential components, highlighting which move in sync and which are independent.

We can verify that our model is not only statistically robust but also free of redundancy by understanding these connections.

Cricket Understanding: While statistical analysis is effective, it may not capture the complexity of the game. As a result, our choice is also influenced by a strong understanding of cricket. This knowledge guarantees that the inputs and outputs selected are not only statistically significant, but also contextually relevant to the sport, capturing the essence of what genuinely influences a batsman's efficiency. By combining correlation analysis insights with the complexities of cricket knowledge, we hope to build a model that is both data-driven and contextually grounded, ensuring its relevance and accuracy.

Correlation for openers:

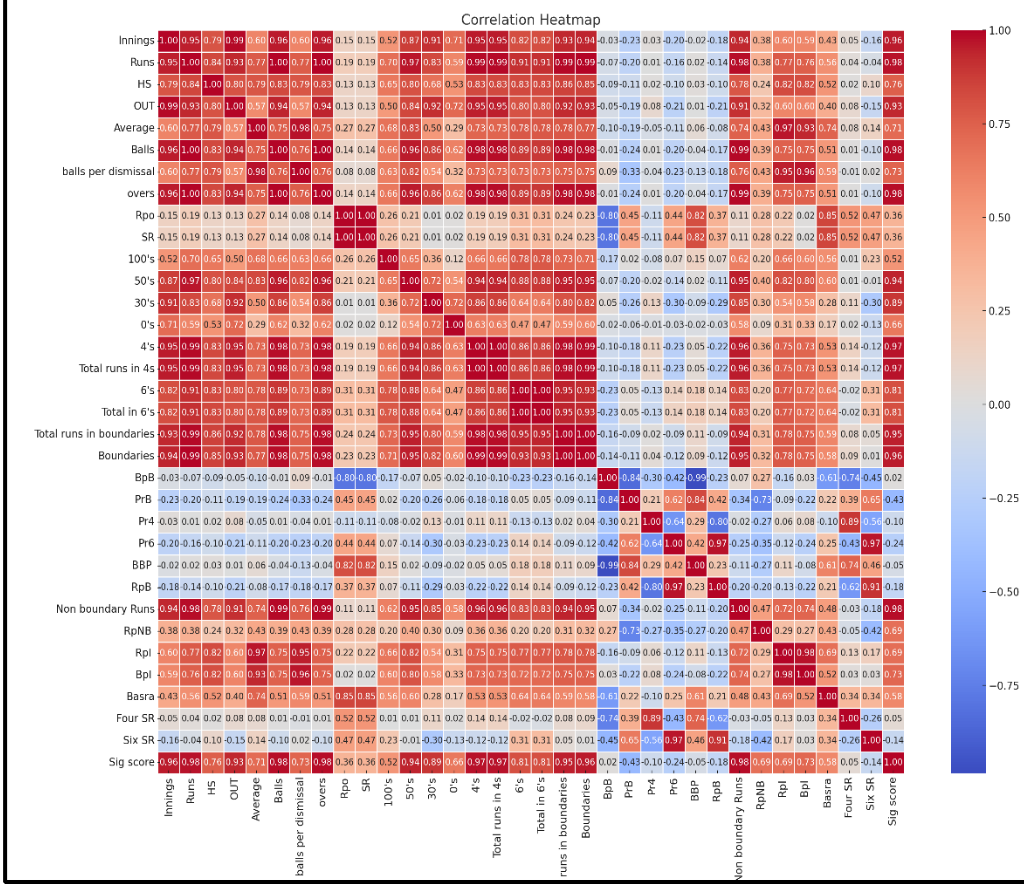


Figure 18: Correlation Matrix for Openers

The above chart shows visualisation of correlation of Opening batsmen. To determine our inputs and outputs we will find the correlation of variable who are positively correlated.

From our visualisation we can see that Variables that the variables RpI(Runs per innings) and BpI are highly significant correlated. With a correlation of 0.98, while positive correlation of 0.59 can be found out between RPI and OUT. Similarly, Four strike rate and runs per over have a positive correlation of 0.52. To understand it better we will use scatter plot to show the visualisation.

The Pearson correlation coefficient between "RpI" (Runs per Innings) and "OUT" (Number of times a player got out) is approximately 0.5960.596.

This value indicates a moderate positive linear correlation between the two variables. In other words, players who tend to score more runs per innings generally also have a higher number of innings in which they got out. This is logical as players who play more innings have more opportunities both to score runs and to get out.

Correlation visualisation:

Balls per innings vs Runs per innings.

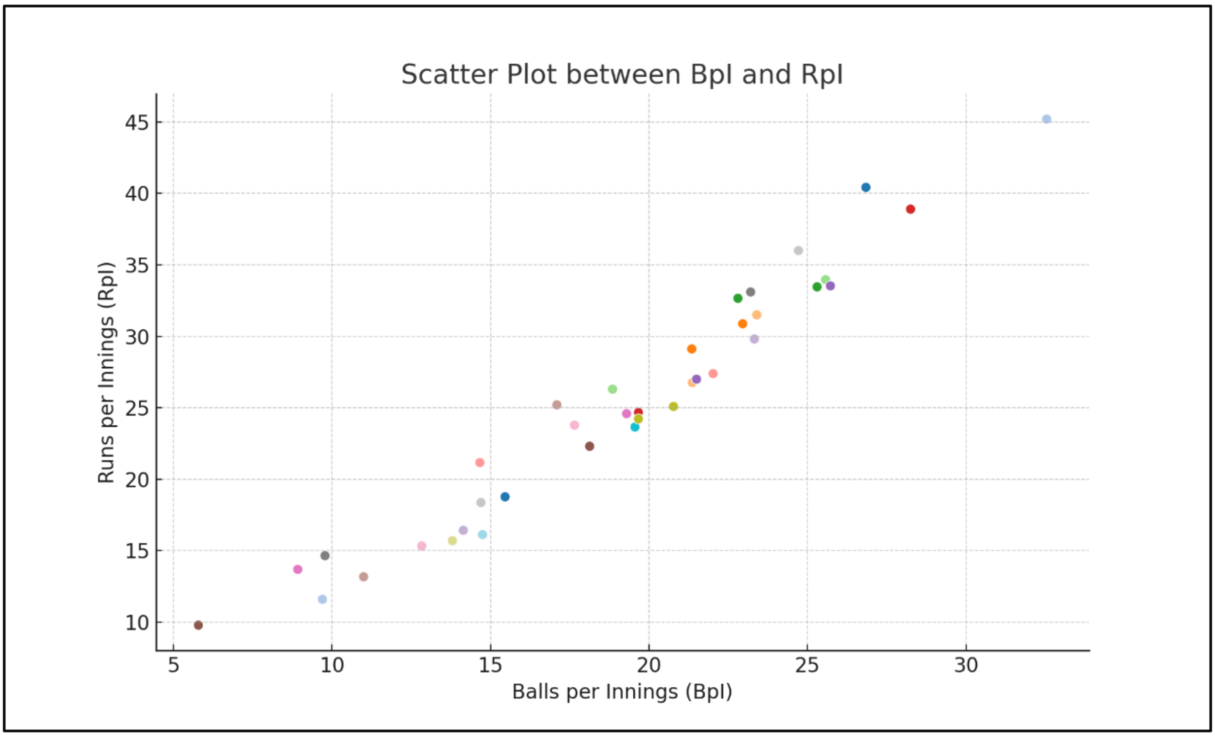


Figure 19: Scatter Plot between BPL & RPL

Openers in cricket are expected to provide a solid start, face a significant number of balls, and lay the foundation for the innings. They often face the new ball, and once they get set, they have the opportunity to play longer innings and score more runs. This is reflected in the positive correlation between the number of balls played and the runs scored per innings.

The scatter plot we visualized earlier does support this observation. Openers or top-order batsmen generally have a higher number of balls faced per innings and, consequently, a higher runs per innings value.

Scatter plot for Runs per over vs Four strike rate

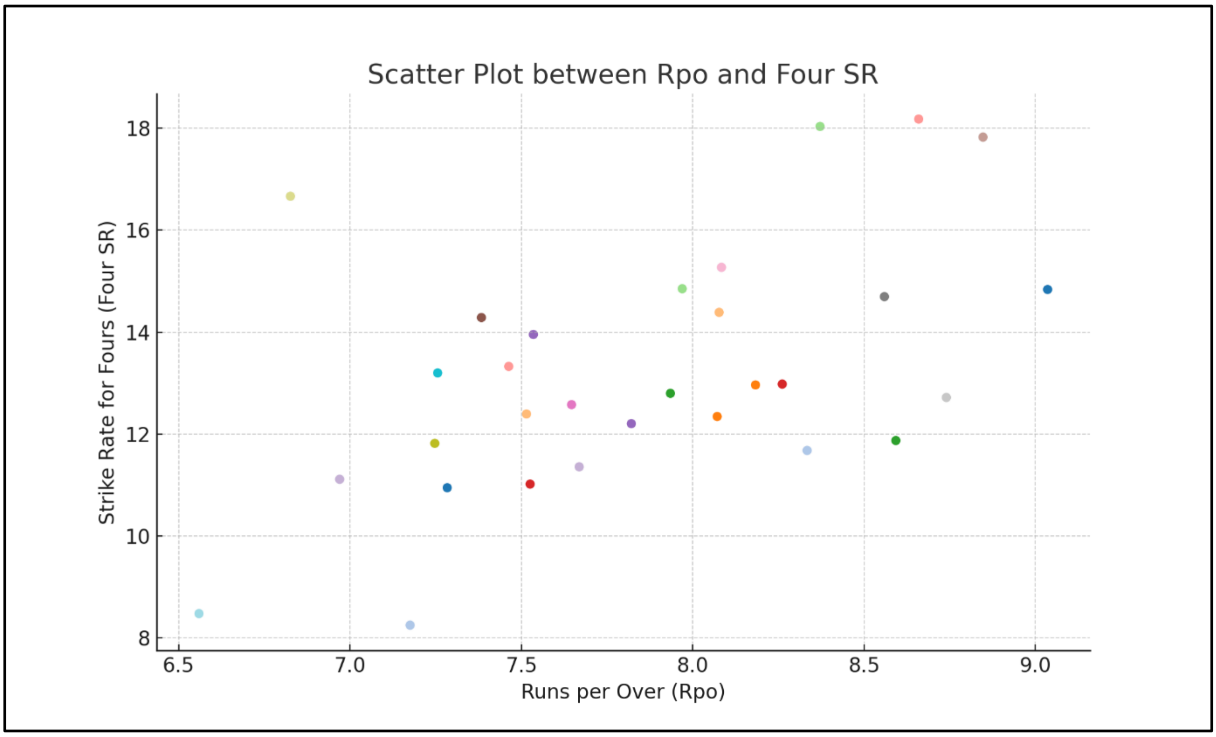


Figure 20: Scatter Plot for RPO and Four SR

From the correlation chart, it's evident that there's a positive association between Rpo and Four SR. This suggests that as players or teams increase their strike rate for fours, they tend to score at a faster overall rate per over. In simpler terms, players who excel at hitting fours more frequently contribute to accelerating the team's overall scoring rate.

This observation aligns with the dynamics of the game. Consistently finding the boundary not only bolsters the scoreboard but also exerts pressure on the opposition, forcing them to adjust their field placements and bowling strategies. A high Four SR often translates to a higher Rpo, as boundaries significantly contribute to the total runs scored in an over. Also, previous visualised openers tend to score more runs in four compared to others .

Selection of Input and outputs for openers

Given the goal of maximizing the output metrics (Four SR, RpO, RpI) while maintaining a steady input (BpI, OUT), it's evident that we're looking for a model that emphasizes batting aggressiveness without compromising on the available balls. Therefore, an output-oriented approach would be apt. This approach would prioritize enhancing the scoring rate and boundary-hitting capabilities while ensuring that the player utilizes the available deliveries efficiently.

In essence, by focusing on these carefully selected metrics, we aim to strike a balance between aggressive batting and judicious utilization of deliveries, ensuring optimal performance on the cricket field.

TOP MIDDLE ORDER

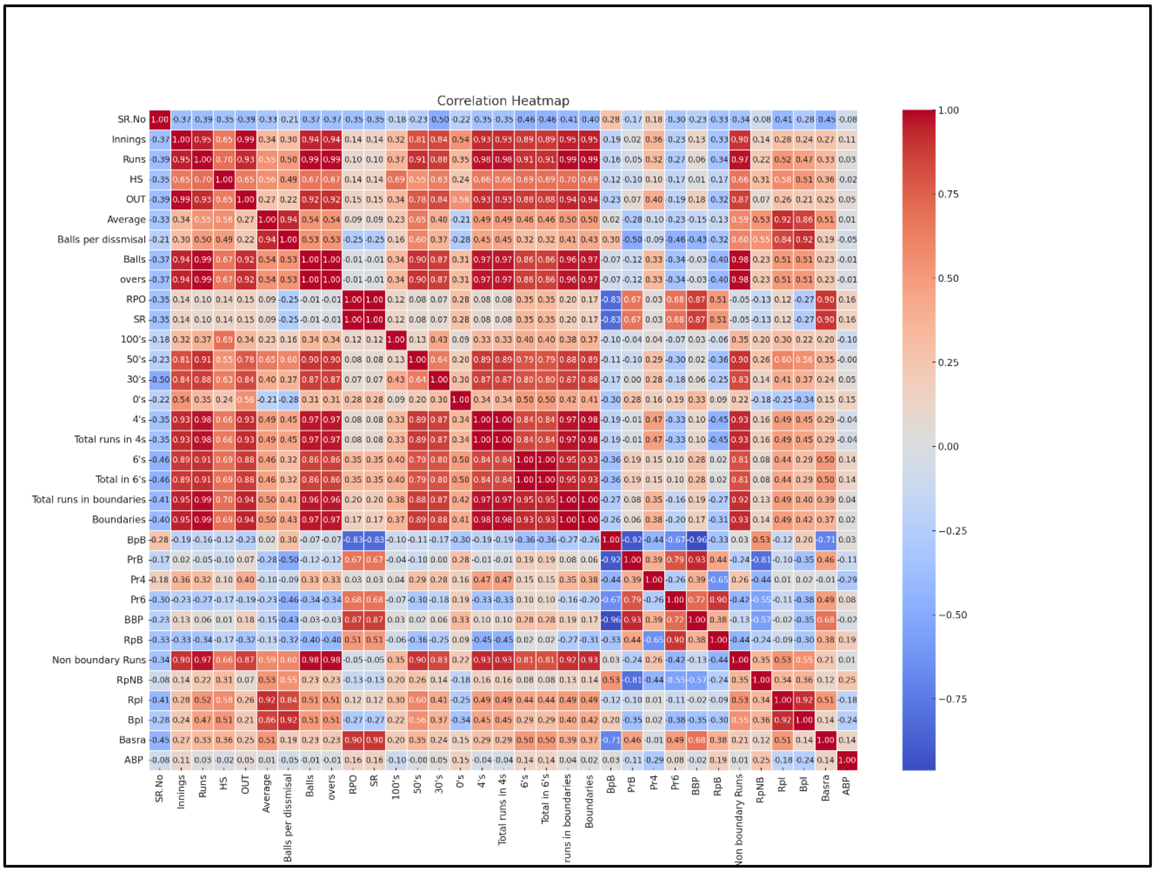


Figure 21: Correlation Matrix for Middle Order

From the heatmap, it's evident that there's a moderate positive correlation between "Balls" and "Average", with a coefficient of approximately 0.54. This suggests that players who have faced more balls tend to have a higher batting average, indicating consistent performance.

Additionally, there's a pronounced positive correlation between "Boundaries" and "Innings". This relationship suggests that as players participate in more innings, the number of boundaries they hit also increases. This could be attributed to experience and adaptability; as players play more innings, they become more accustomed to various bowling styles and conditions, leading them to hit more boundaries. We will validate this with the help of scatter plot.

Correlation visualisation

1. **Average vs balls**

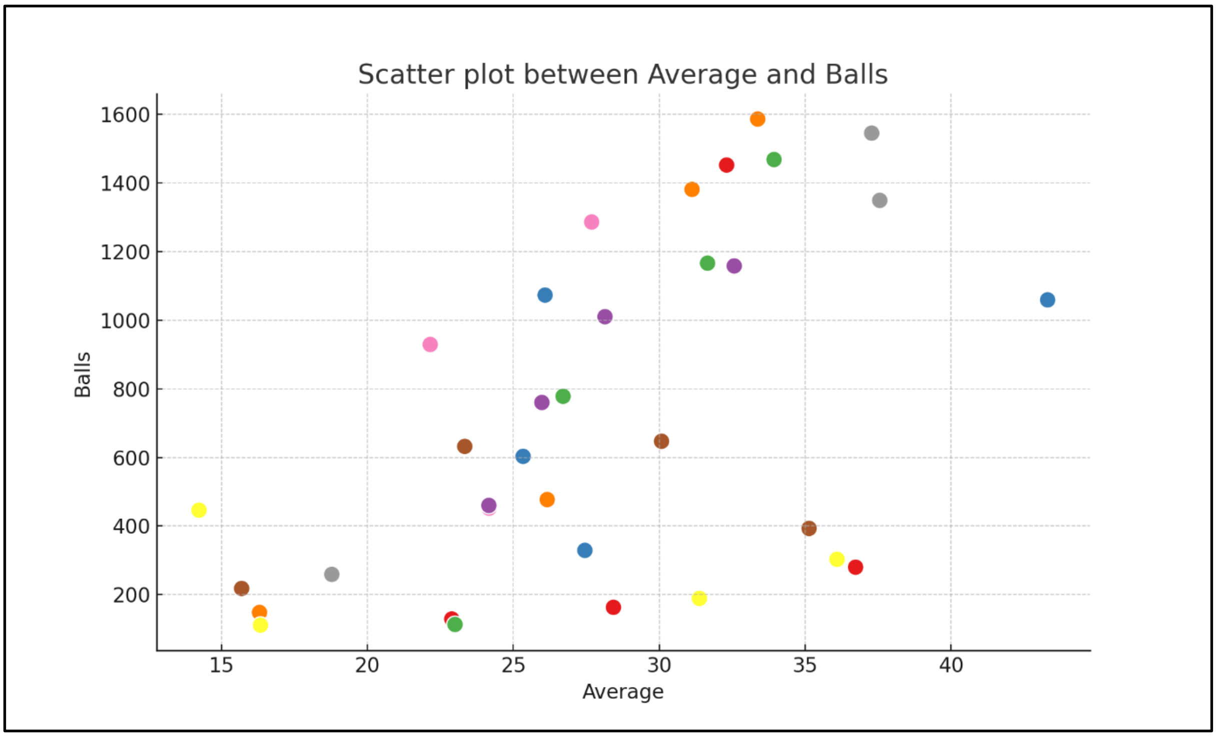


Figure 22: Scatter Plot between Average & Balls

The distribution of the scatter plot shows a significant clustering trend rather than a definite linear inclination. This suggests the presence of a possible positive association, although not a very strong one. Furthermore, the data reveals a specific trend among top middle-order batsmen: their consistency in scoring increases as the number of balls they face increases. Their higher batting averages reflect this, highlighting the necessity of patience and strategic play in cricket, particularly for those in crucial batting positions. After falling a wicket of openers the middle order tend play slow and accelerate their innings in the second half of their innings.

Boundaries vs innings.

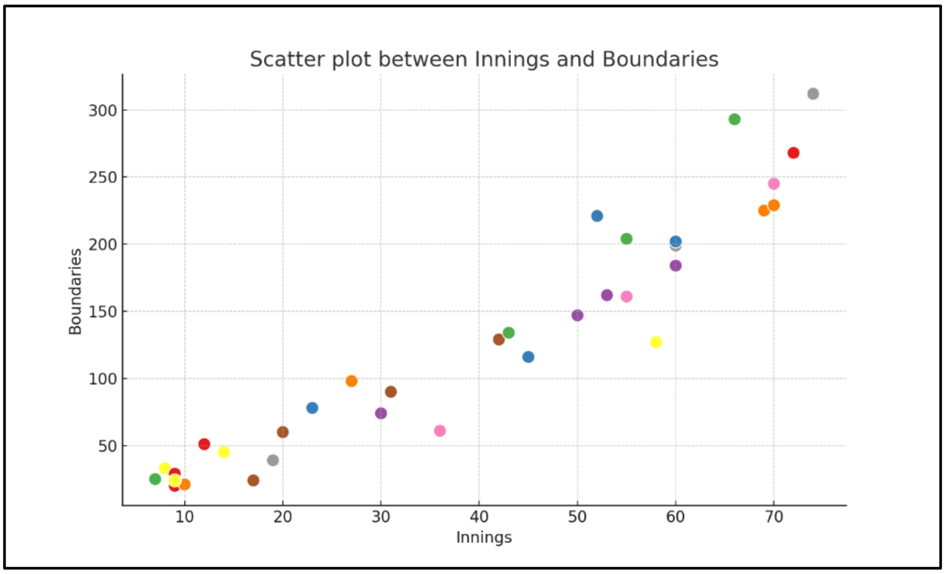


Figure 23: Scatter Plot between Innings and Boundries

The plot displays a clear upward trend, indicating that as the number of innings played by a batsman increases, so does the total number of boundaries they hit. This relationship is expected since more innings provide more opportunities to score boundaries.

Furthermore, the number of points along the trend line suggests that consistent players, or those who play matches on a regular basis, take use of their opportunities by hitting a substantial number of boundaries. It demonstrates these players' skill and adaptability: as they gain experience by playing more innings, they become more suited to locate scoring possibilities, resulting in more boundaries.

Selection of input and output variable

Based on the correlation study, we identified "Innings" and "Balls" as essential input variables for evaluating top middle-order batsmen, with the primary goal of maximising outputs like "Average" and "Boundaries". Furthermore, the inclusion of "Strike Rate" as a variable is validated by previous visual representations that showed that middle-order batsmen, particularly those in positions 3 and 4, frequently had a higher strike rate than openers.

The reason behind choosing these specific factors is based on the normal features of batsmen in these positions. Batsmen batting in positions 3 and 4 are typically expected to be consistent and patient but also capable of shifting gears and displaying aggression when necessary. As a result of focusing on these criteria, we can properly assess and harness the potential of top performers.

Correlation for Lower middle order.

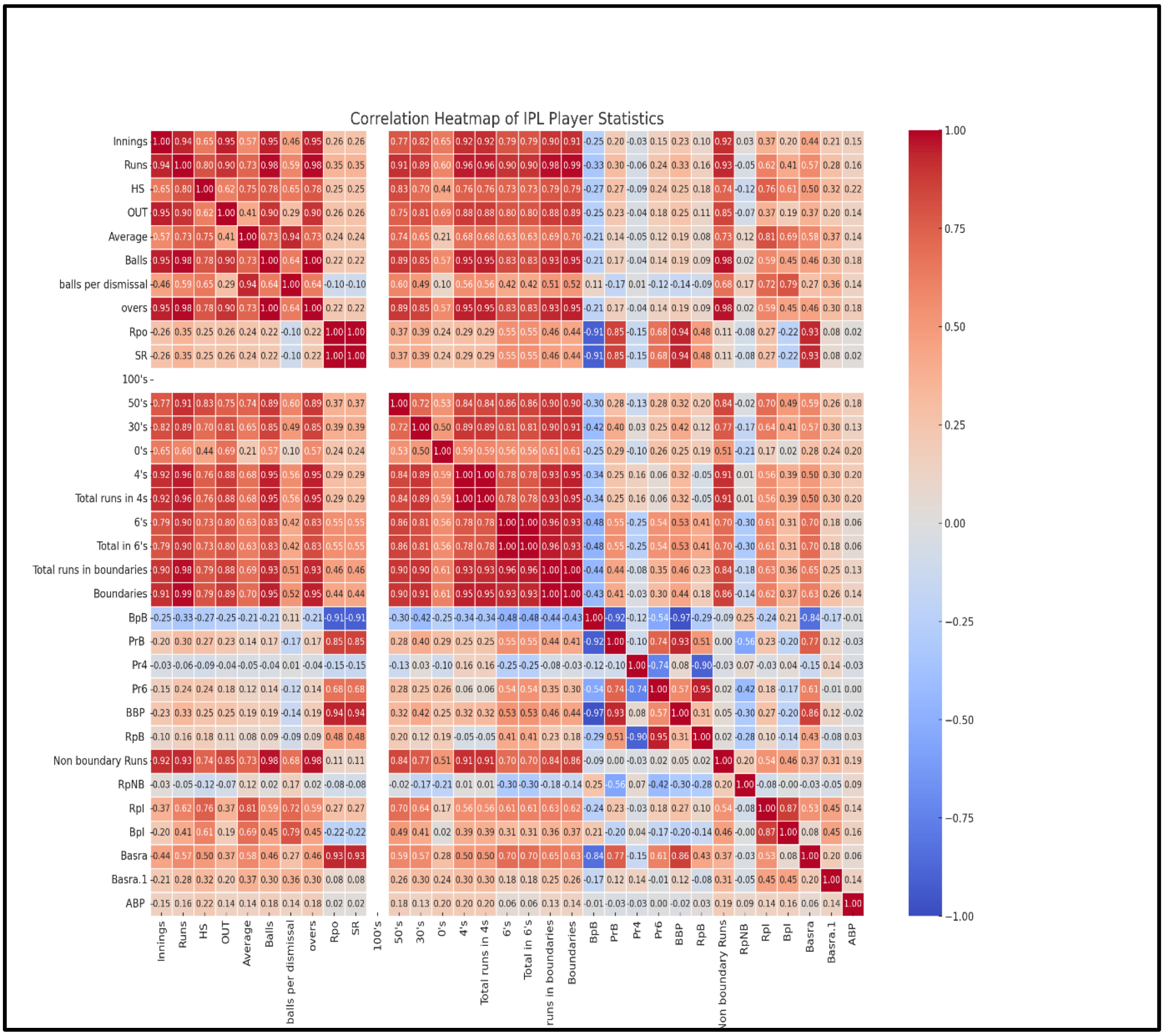


Figure 24: Correlation for Lower Middle Order

The graphic shows a small positive association between the number of balls faced and the proportion of runs scored through sixes (Pr6), as demonstrated by a correlation coefficient of 0.14. This small connect corresponds to realistic cricketing conditions, particularly in the context of lower-order batsmen, who frequently do not get the opportunity to face a significant number of balls and just score more in sixes with the balls they face. To further investigate we will be looking a the scatter plot.

Scatter plot visualisation

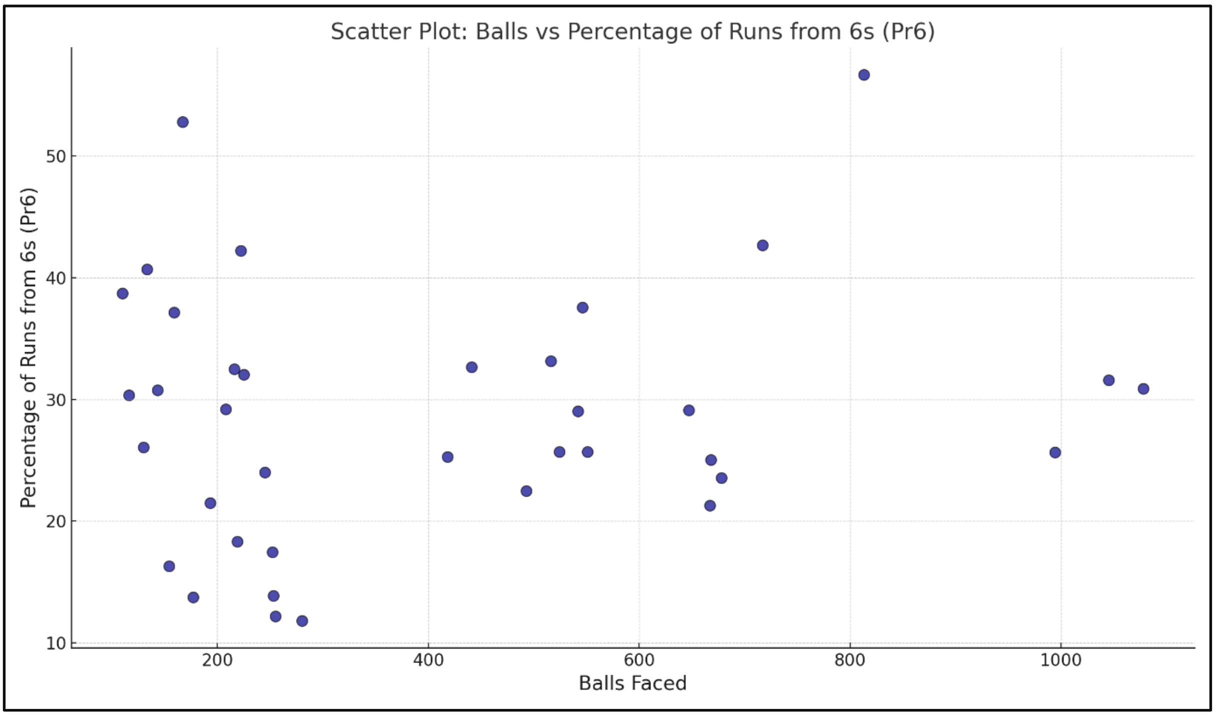


Figure 25: Scatter Plot Balls vs Percentage of Runs from 6s

From the plot we can visualise that A greater number of balls faced doesn't necessarily translate to a higher proportion of runs coming from sixes. This indicates the multifaceted nature of the game, where a batsman's role can vary based on match situations, team strategy, or personal strengths. A cluster of players who haven't faced many balls but have a high percentage of runs from sixes can be identified as "power hitters". These players might be entrusted with the role of providing a late-innings surge or chasing down targets in high-pressure situations. Conversely, those who've faced numerous balls but have a lower percentage of runs from sixes are likely the "innings builders". They ensure stability, rotate the strike, and set the stage for the power hitters to accelerate towards the end. And might be inefficient player at this position. Although, there are few exceptions who've both faced a considerable number of balls and maintain a high six-hitting percentage. Such players blend the roles of an anchor and an aggressor, making them invaluable assets to their teams.

Selection of Input and Output Variables- We have chosen "Balls" as our independent variable and will be using "Percentage of Runs Scored from Sixes" in conjunction with "Boundaries" as our dependent variables. The primary objective is to optimize and maximize this output, thereby leveraging the relationship between the balls faced and the combined impact on scoring, particularly through boundaries and sixes.

3.7.4 SELECTION OF RETURNS TO SCALE.

In this section we will be selecting returns to scale

Variable returns to scale.

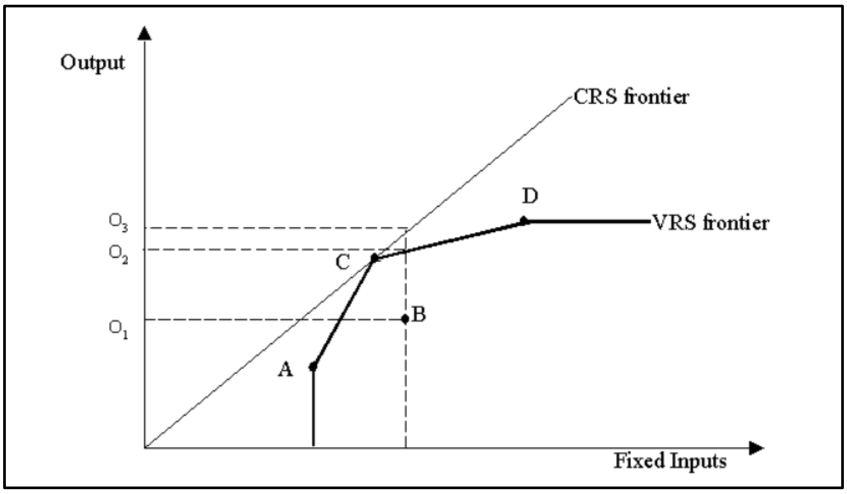


Figure 26: Variable Return to Scale

The BCC model, which uses variable returns to scale developed by Banker, Charnes, and Cooper, was a pivotal advancement from the original CCR model in DEA analysis. This innovative approach, as highlighted by Banker & Thrall (1992), was designed to understand that organizations (or DMUs) might be efficient at distinct scales or magnitudes. In essence, the BCC model recognized that productivity could fluctuate across different levels, yet each unit could still be deemed efficient within its respective context. This insight was integral to the BCC model, aiming to factor in the varying conditions that might influence productivity in different settings. As previously pointed out, the efficiency of varied production methods can be impacted by the operational scale of these units, a notion that can be traced back to the foundational contributions by Banker, Charnes, & Cooper (1984).

In this paper we will employing BCC model with variable return to scale, Using the CCR model in cricket would mean expecting a consistent and proportional performance increase with increased effort or input. In contrast, the BCC model would account for the variable nature of sports, where performance can be influenced by numerous unpredictable factors. For instance, VRS, acknowledges that a batsman's scoring rate might vary based on various factors, and there's no fixed proportionality between the balls faced and runs scored.

3.7.5 FORMULATION FOR OUTPUT-ORIENTED DEA MODEL

$$
\begin{aligned}
&\text{Max } \alpha \\
&\text{s.t.} \\
&\quad \sum_{j=1}^{n} \lambda_j X_{ij} \le X_{ij_0} \qquad ; i = 1, \ldots, m \\
&\quad \sum_{j=1}^{n} \lambda_j Y_{rj} \ge \alpha Y_{rj_0} \qquad ; r = 1, \ldots, s \\
&\quad \sum_{j=1}^{n} \lambda_j = 1. \\
&\quad \lambda_j \ge 0 \qquad ; j = 1, \ldots, n
\end{aligned}
$$

The output-oriented DEA model is presented above with input variables (X1, ..., Xm) and output variables (Y1, ..., Ys) with n decision making units (j = 1, ..., n).

θ is tis the efficiency score of the DMU under evaluation.

λ is a vector of weights assigned to each DMU.

The constraint ∑j=1nλj=1 ensures variable returns to scale.

Slacks and weights(Banker, Charnes, & Cooper (1984)), Charnes, Cooper, & Rhodes, 1978)

Slacks: Slacks represent the potential for improvement within a Decision-Making Unit (DMU) in Data Envelopment Analysis (DEA). They signify how much a DMU can enhance its outputs without altering its input levels or reduce its inputs without affecting its outputs. Slacks serve as indicators of underutilized resources or excess inputs within inefficient DMUs. Positive slack in an input suggests that the DMU is not fully utilizing that resource. Conversely, positive slack in an output implies that the DMU could produce more of that output without increasing its inputs. In the context of DEA, a DMU is considered efficient if it has no slacks (slack values are all zero) since this indicates optimal resource utilization. However, positive slacks suggest areas where improvements can be made.

Weights: Weights are coefficients assigned to each input and output in the DEA model. These weights are derived through the optimization process of DEA and convey the relative importance of each input and output when assessing the efficiency of a DMU. Weights play a crucial role in calculating the efficiency score of each DMU, representing how much emphasis is placed on each input and output in the efficiency assessment. A DMU with a higher efficiency score is using its inputs more effectively to generate its outputs. The weights provide insights into the contribution of each input and output to this efficiency score. Inputs or outputs with higher weights are deemed more influential in determining the overall efficiency.

3.7.6 EFFICIENCY EVALUATION IN PIM DEA

To evaluate the efficiency of players we have used PIM's DEAsoft-V3 which software offers a robust platform for evaluating efficiency and productivity across various organizational units. Beyond these capabilities, it also aids in setting performance benchmarks, defining targets, and enabling effective performance management. Papagapiou, Mingers, & Thanassoulis, (1997). The evaluated results will be shown through bar plot visualisation. In our analysis, we will first identify and evaluate a select group of benchmark efficient players who consistently demonstrate high levels of performance. We will evaluate then provide actionable recommendations and guidance to help inefficient players enhance their performance and become more efficient.

4 RESULTS AND ANALYSIS

4.1 DEA ANALYSIS RESULTS

In this section we will be evaluating the results of the analysis. The section is divided is divided in to three parts –

- Efficiency evaluation of Opener Batsmen.
- Efficiency evaluation of Top middle order Batsmen.
- Efficiency evaluation of Lower middle Batsmen.

4.1.1 EFFICIENCY EVALUATION OF OPENERS

Efficiency scores chart.

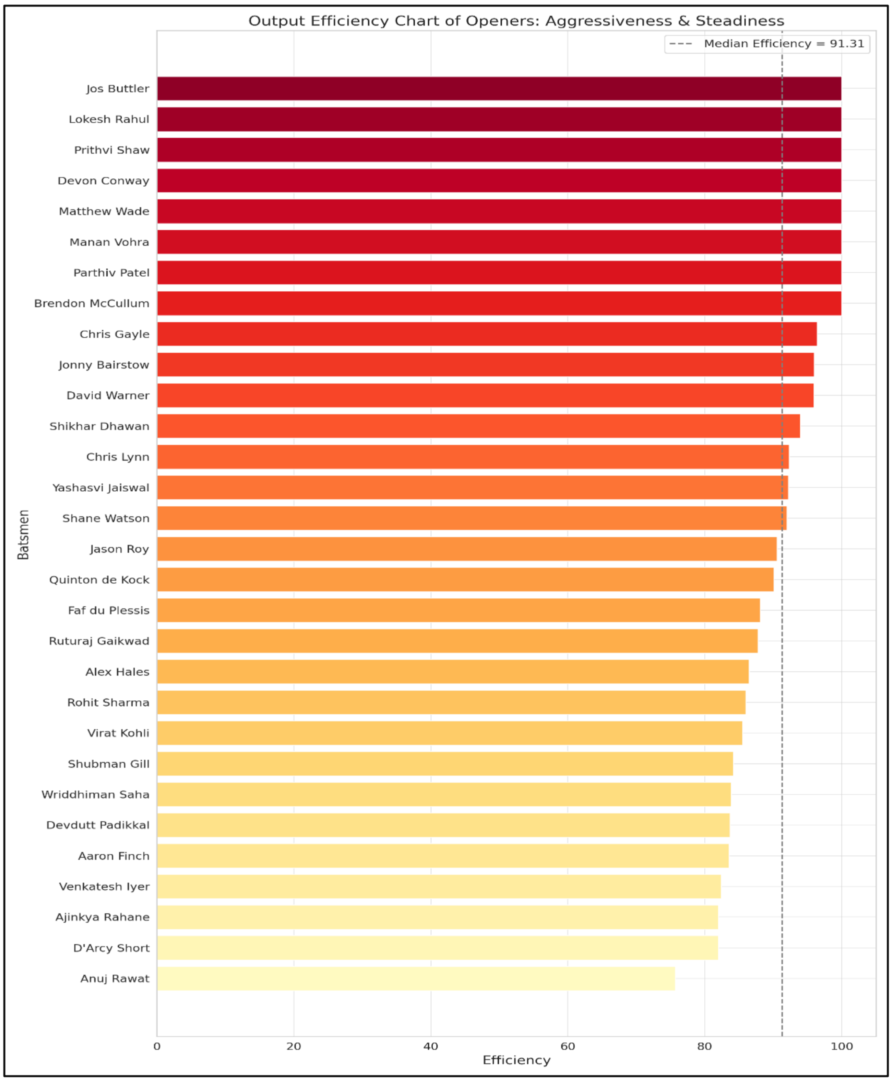


Figure 27: Efficiency scores chart for Openers

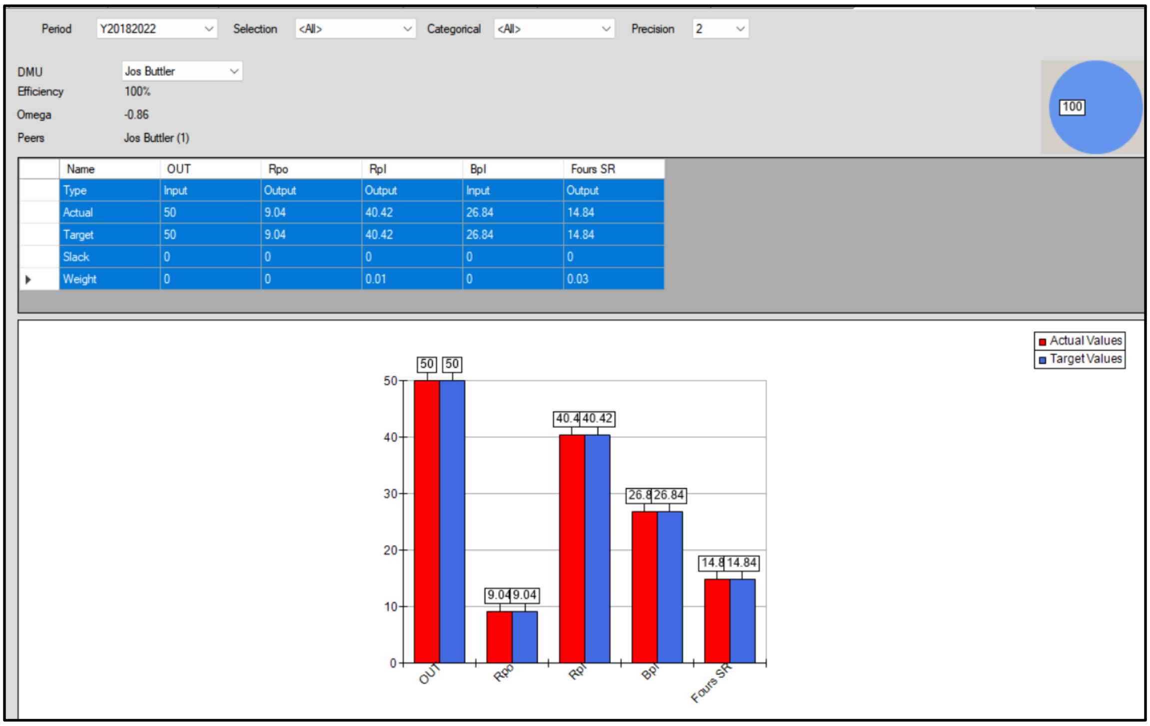
EFFICENCY SCORES OF TOP ORDER BATSMEN

The above chart shows the efficiency scores of the openers in percentage when the inputs OUT and BPI (balls per innings) and output Runs per over, Runs per innings and Four SR was used. The chart vividly presents the efficiency scores of cricket openers, expressed as percentages. These scores were derived by factoring in both inputs and outputs associated with each player's performance: Inputs: Number of Outs (OUT), Balls Per Innings (BPI). Outputs: Runs Per Over, Runs Per Innings, Four Strike Rate (Four SR)

From the visual representation, the colour intensity provides a clear distinction: Deep Red Bars: Signify the top-performing players, reflecting the highest efficiency based on the given parameters. Pale Yellow Bars: Indicate players with lower efficiency scores, suggesting areas where there might be room for improvement or different playing styles. This subtle use of colour aids in the instant identification of the most efficient players, ensuring a thorough grasp of each opener's performance in comparison to their colleagues.

EFFICIENCY ANALYSIS OF OPENERS.

1. **Jos Buttler.**



Name	OUT	Rpo	RpI	BpI	Fours SR
Type	Input	Output	Output	Input	Output
Actual	50	9.04	40.42	26.84	14.84
Target	50	9.04	40.42	26.84	14.84
Slack	0	0	0	0	0
Weight	0	0	0.01	0	0.03

Figure 28: Efficiency Analysis of Jos Buttler

Efficiency: Jos Buttler achieved an efficiency score of 100%, indicating that he operates on the efficient frontier based on the metrics considered.

Slacks: All slack values for Jos Buttler are zero, confirming that there's no excess or shortage in any of the inputs or outputs. This aligns with his status on the efficient frontier.

Weights: Weights assigned in the DEA indicate the relative importance of each metric:

'RPI' (Runs Per Innings): 0.01

'FOUR STRIKE RATE': 0.03

The zero weights for the other metrics imply they didn't contribute to Jos Buttler's efficiency score in this DEA model.

KL RAHUL

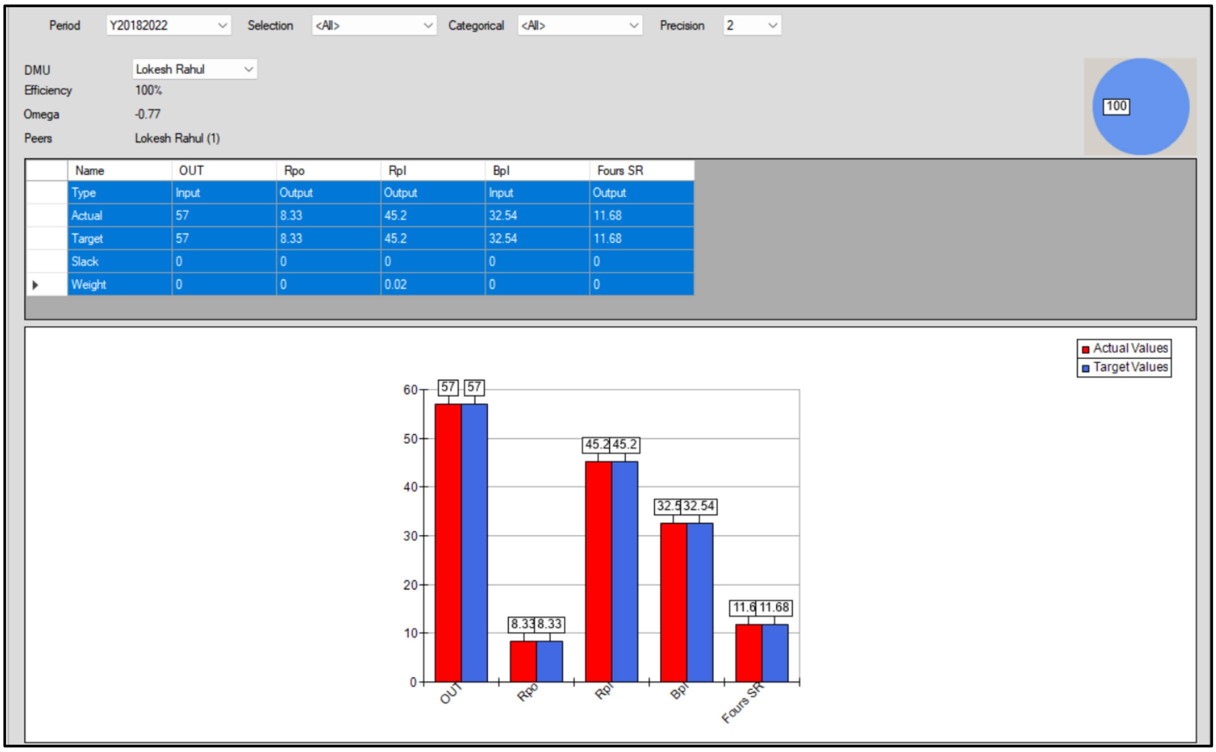


Figure 29: Efficiency Analysis of KL Rahul

Efficiency: KL Rahul achieved an efficiency score of 100%, indicating optimal performance based on the metrics analyzed.

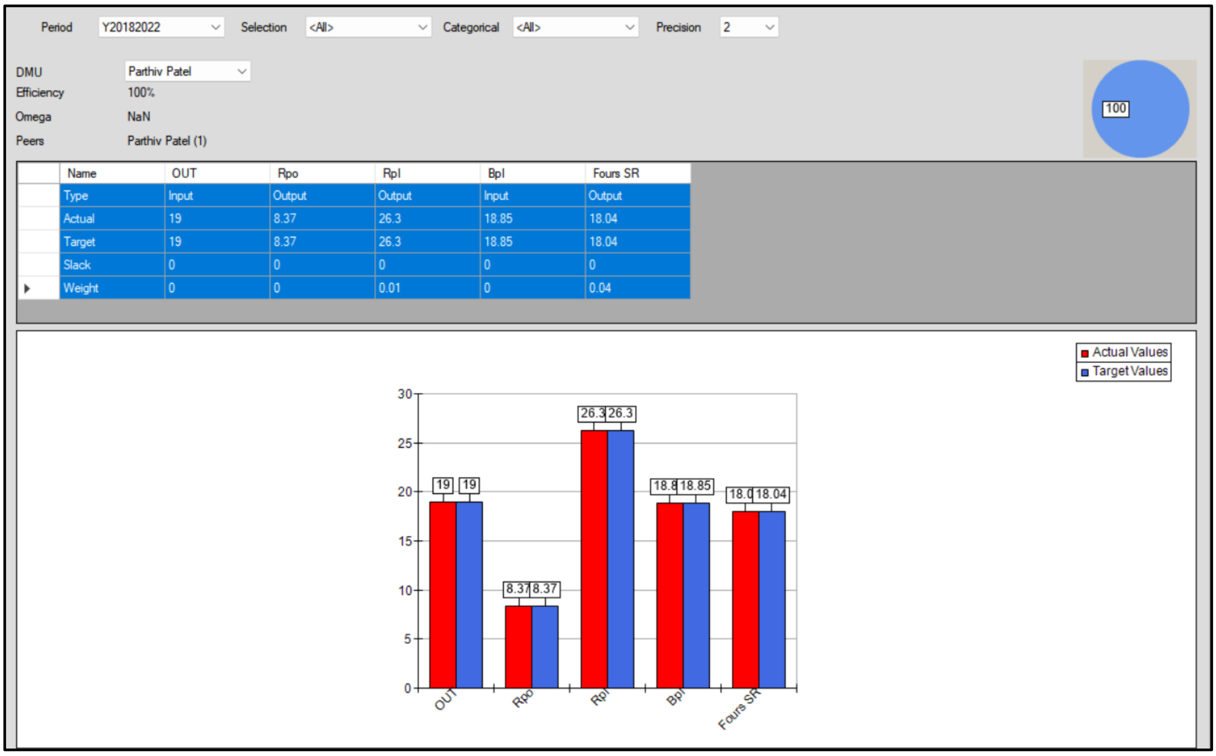
Slacks: All slack values for KL Rahul are zero, suggesting no deviation from the target values in any of the inputs or outputs.

Weights: In the DEA evaluation:

The weight for 'RPI' (Runs Per Innings) is 0.02.

All other metrics have a weight of zero, indicating they did not influence KL Rahul's efficiency score in this DEA model.

PARTHIV PATEL



Name	OUT	Rpo	Rpl	Bpl	Fours SR
Type	Input	Output	Output	Input	Output
Actual	19	8.37	26.3	18.85	18.04
Target	19	8.37	26.3	18.85	18.04
Slack	0	0	0	0	0
Weight	0	0	0.01	0	0.04

Figure 30: Efficiency Analysis of Parthiv Patel

Efficiency: Parthiv Patel has an efficiency score of 100%, signifying optimal performance based on the considered metrics.

Slacks: All slack values for Parthiv Patel are zero, indicating no discrepancies from the target values across all inputs and outputs.

Weights: In the DEA evaluation for Parthiv Patel:

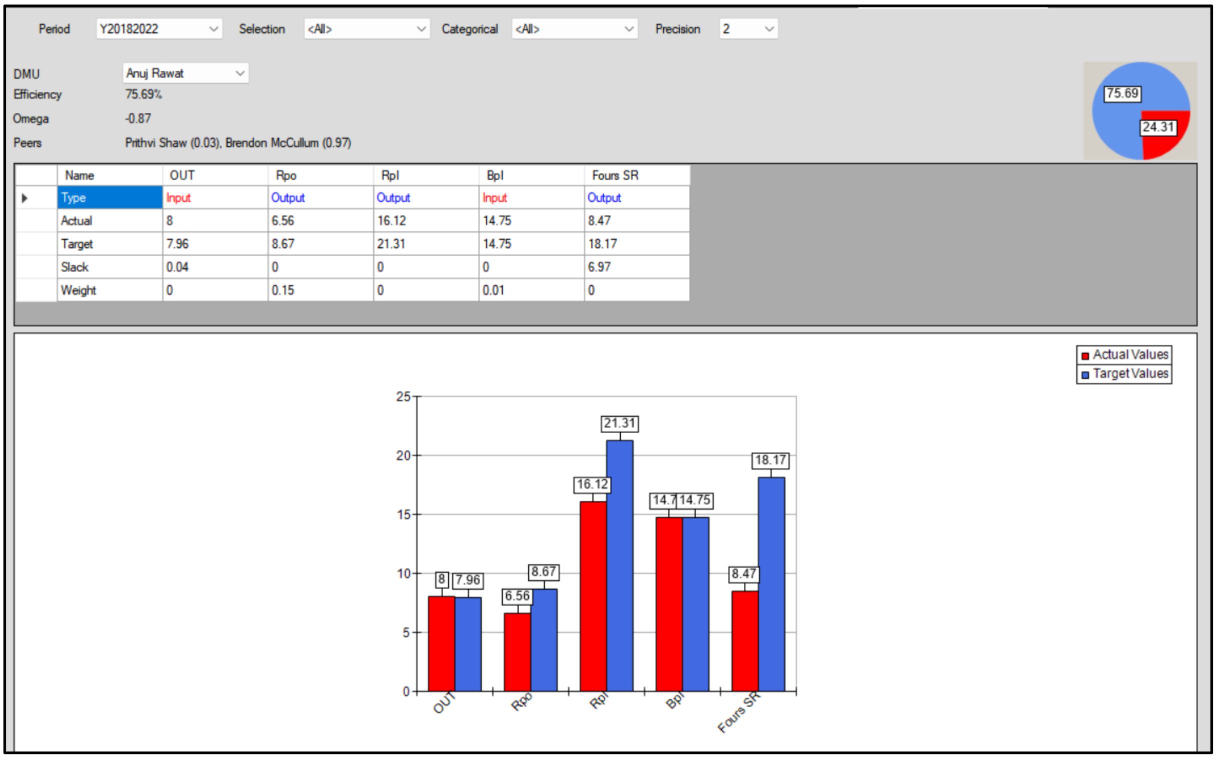
The weight assigned to 'RPI' (Runs Per Innings) is 0.01.

The weight for 'FOUR STRIKE RATE' is 0.04.

All other metrics have a weight of zero, suggesting they did not influence Parthiv Patel's efficiency score in this DEA model.

Inefficient player

ANUJ RAWAT



Name	OUT	Rpo	Rpl	Bpl	Fours SR
Type	Input	Output	Output	Input	Output
Actual	8	6.56	16.12	14.75	8.47
Target	7.96	8.67	21.31	14.75	18.17
Slack	0.04	0	0	0	6.97
Weight	0	0.15	0	0.01	0

Figure 31: : Efficiency Analysis of Anuj Rawat

Anuj Rawat's performance metrics, when compared to the target benchmarks, indicate inefficiencies in several areas. Here's a breakdown:

OUT (Output): Although the actual output slightly exceeds the target, the difference is marginal. The weight assigned to this metric is zero, suggesting that it may not be a primary factor for consideration.

RPO (Output): The actual output for RPO is below the target by a significant margin. Given a weight of 0.15, this indicates that improvements in this area could have a substantial impact on overall performance.

RPI (Output): The RPI output also falls short of the target. However, as the weight is zero, it might not be a primary focus area.

BPI (Input): BPI meets the target exactly, indicating that this particular metric is on par with the benchmark. With a very low weight of 0.01, it holds minimal influence on the overall performance.

FOUR STRIKE RATE (Output): The FOUR STRIKE RATE output is notably below the target, with a significant slack of 6.97. However, its weight of zero suggests that it might not be a primary factor for consideration.

In comparison to peers such as Prithvi Shaw and McCullum, Anuj Rawat's performance metrics highlight areas of inefficiency, particularly in the RPO output. Considering the weights and slacks, it is evident that there are specific metrics where focused improvement could lead to better alignment with target benchmarks.

4.1.2 TOP MIDDLE ORDER

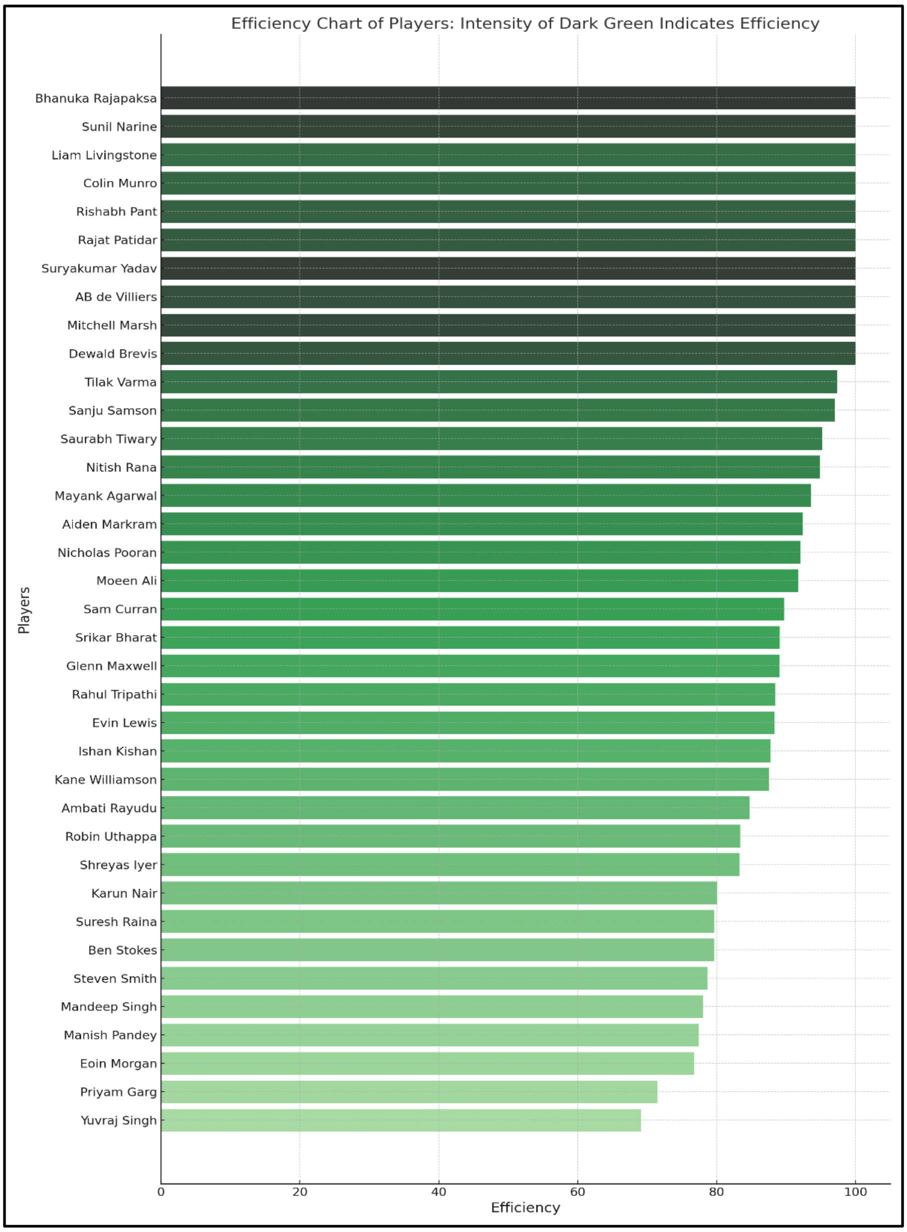


Figure 32: Efficiency Chart of Middle Order Players

Liam living stone

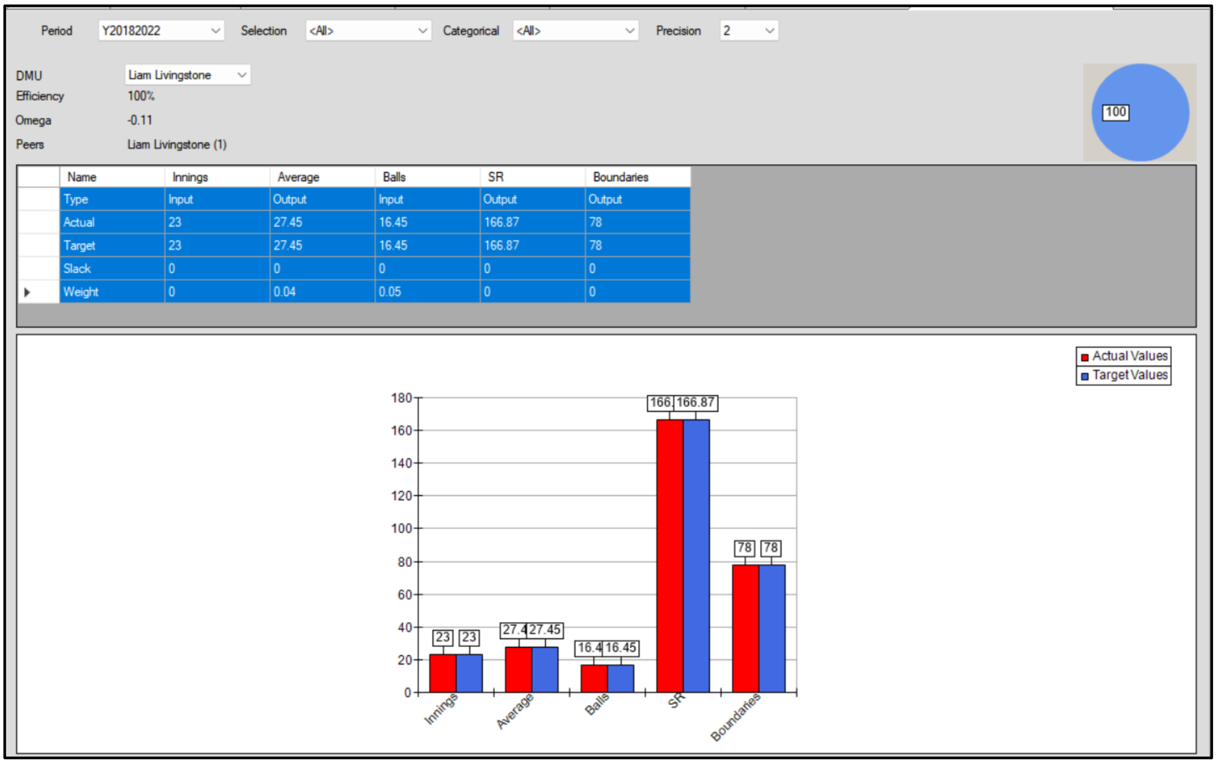


Figure 33: Efficiency Analysis of Liam Livingstone

Livingstone has played an average of 23 innings, which aligns perfectly with the target. This metric provides insight into his consistent presence in the team lineup. He faces an average of 27.45 balls per innings, matching the target, which quantifies his exposure at the crease and serves as a foundation for understanding his output metrics. The more balls a player faces, the more opportunities they have to impact the game. Livingstone's average score per innings stands at 16.45, which is on par with the target, highlighting his consistency in scoring. His strike rate is an impressive 166.87, meeting the target precisely, showcasing his aggressive approach. Livingstone has hit 78 boundaries, which matches the target, indicating his ability to find the fence regularly. There's no slack for the inning average, and the weight is 0. This suggests that while the number of innings played is essential, it might not be the primary focus in the DEA model for Livingstone. With no slack and a weight of 0.04, this metric emphasizes the importance of Livingstone's ability to face a consistent number of balls and get set at the crease. The absence of slack and a weight of 0.05 underscores the significance of Livingstone's consistent scoring in the model. Both metrics have no slack and a weight of 0, suggesting that while Livingstone meets the targets in these areas, they might not be the primary focus in the model. This could hint at the model's emphasis on other aspects of his game.

Inefficient – Sanju samson

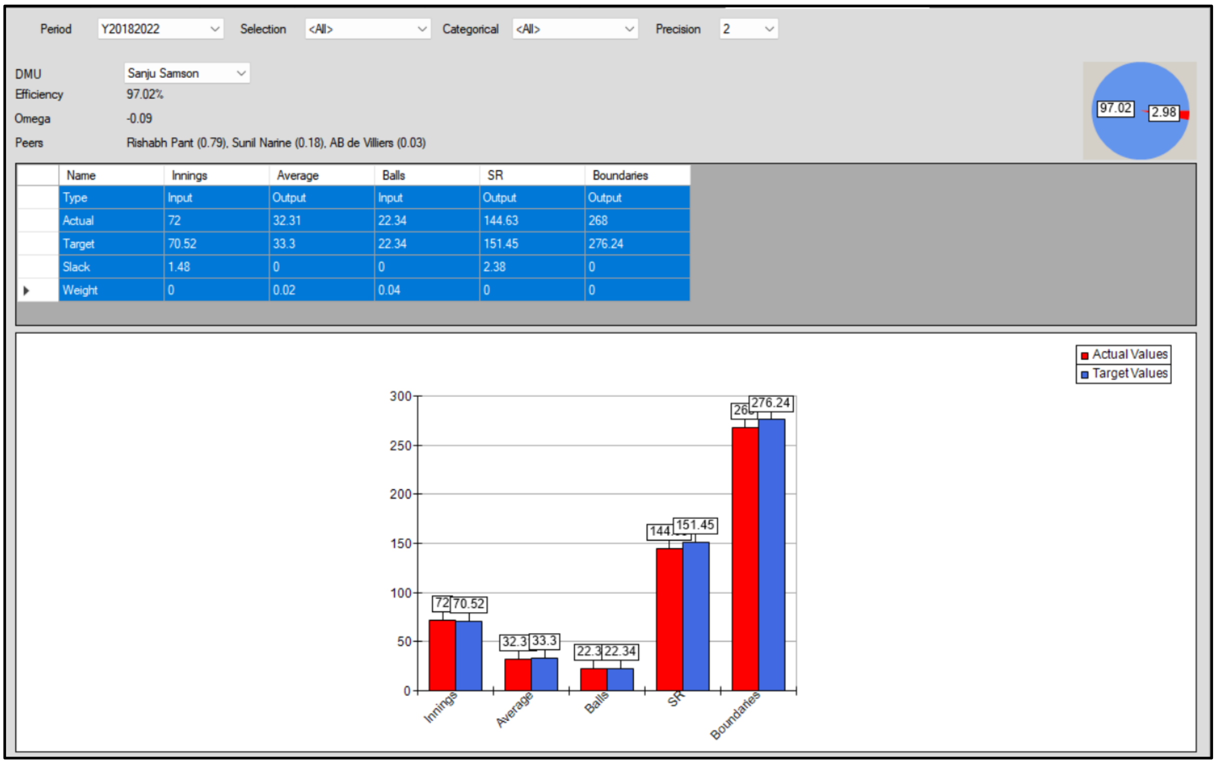


Figure 34: Efficiency Analysis of Sanju Samson

Sanju Samson's actual inning average is 72, slightly higher than the target of 70.52. This indicates that Samson has been consistently present in the lineup. The actual average balls faced by Samson is 32.31, which is marginally below the target of 33.3. This suggests that he is close to meeting the expected number of balls faced per innings. Samson's actual strike rate stands at 144.63, which is below the target of 151.45, indicating a potential area for improvement in his scoring rate. Samson has hit 268 boundaries, which is slightly below the target of 276.24, hinting at a potential to increase his boundary-scoring ability. The slack for inning average is 1.48, indicating a minor surplus compared to the target. The weight of 0 suggests that the inning average might not be the primary focus in the DEA model. There's no slack for average balls faced, indicating that Samson's performance aligns closely with the target. The weight of 0.02 emphasizes the importance of this metric. The slack of 2.38 for strike rate points to a slight gap between his current performance and the target. The weight of 0.04 underscores the significance of the strike rate in the model. There's no slack for boundaries, suggesting that Samson's boundary-hitting ability is close to the target. The weight of 0 might indicate that the sheer number of boundaries isn't as crucial as other metrics in the model.

4.1.3 LOWER MIDDLE ORDER

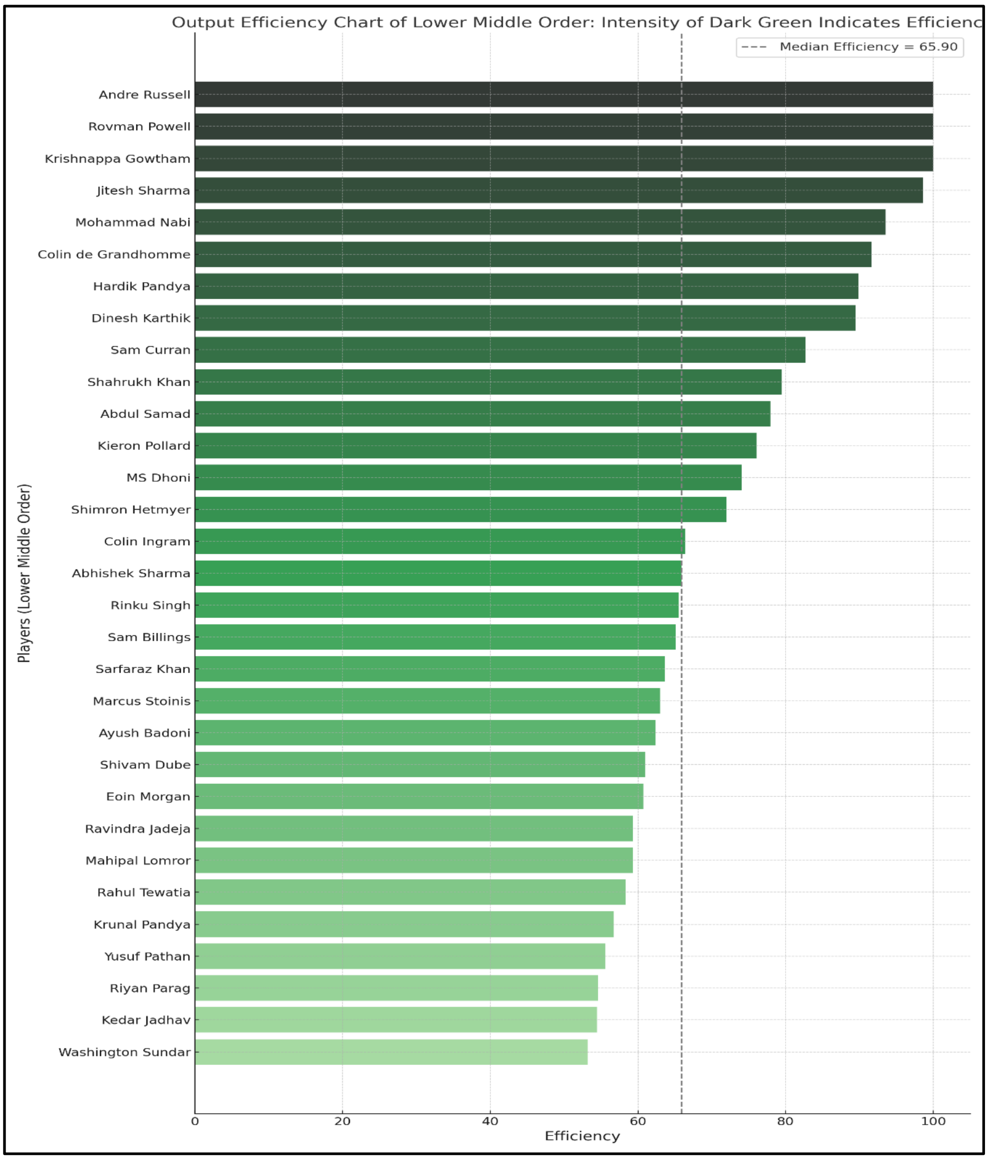


Figure 35: Efficiency chart of Lower Middle Order

The chart elegantly depicts the efficiency of lower middle order cricketers, with the metrics derived from specific inputs and outputs. The input parameter considered is the number of balls played, while the outputs encompass the boundaries scored and the runs per 6 balls (Pr6).

From the visual representation, it's evident how colour intensity acts as a quick gauge of performance.

Deep Green Bars: These represent players at the pinnacle of efficiency, excelling in maximizing runs within limited balls. Pale Green Bars: These highlight players whose efficiency metrics are comparatively lower. It suggests potential areas of improvement or different playing styles

This color-coded approach provides an immediate, intuitive understanding of each cricketer's proficiency in the lower middle order, allowing for easy comparison relative to their contemporaries.

Andre Russell

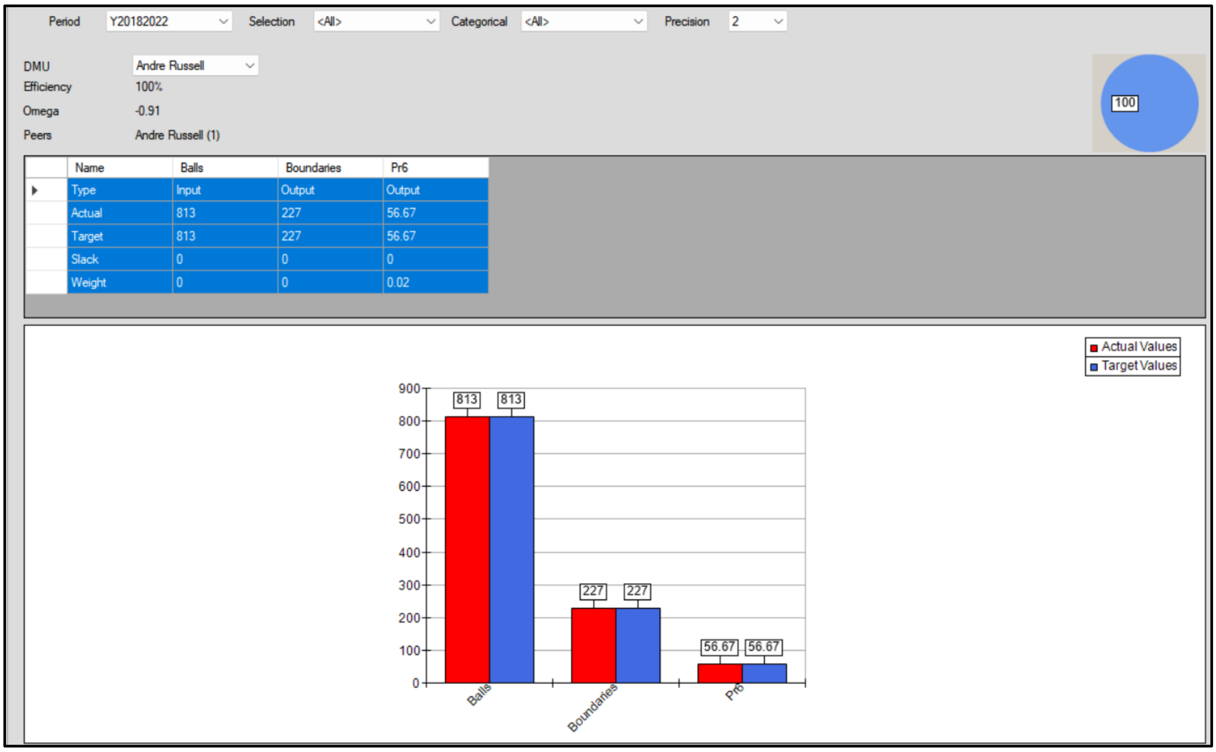


Figure 36: Efficiency Analysis of Andre Russell

Andre Russell's Performance Metrics: Andre Russell is 100 % efficient with all the actual variables meet the target variables.

Slacks: The slack values for all three metrics (Balls, Boundaries, Pr6) are zero. This indicates that there is no excess or shortage in any of the inputs or outputs concerning the target values. In other words, Russell is achieving the desired level of performance in all of these metrics.

Weights: The weights assigned to the metrics are provided. It appears that only the "Pr6" metric is given a weight, which is 0.2. The weights represent the relative importance of each metric in the DEA analysis. Since the weights for "Balls" and "Boundaries" are both zero, it implies that these metrics didn't contribute to Russell's efficiency score in this DEA model. However, the "Pr6" metric is considered and weighted with a value of 0.2.

Krishnappa Gowtham

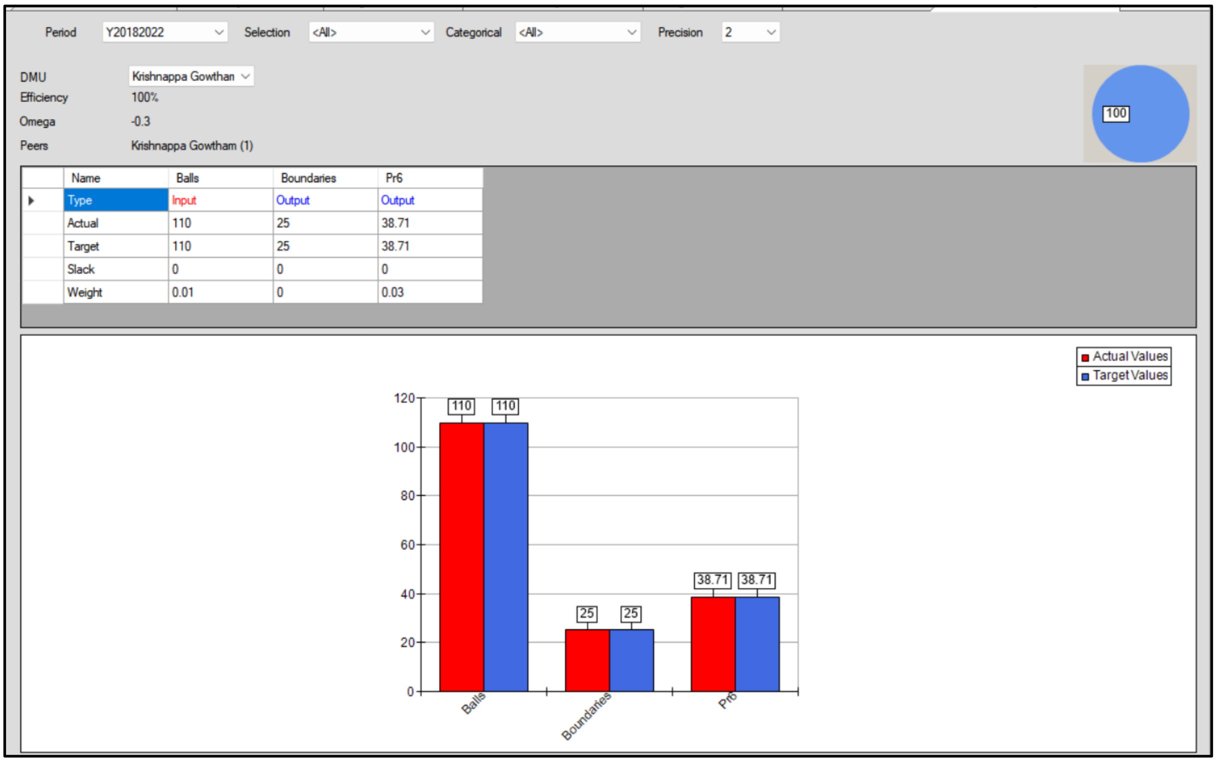


Figure 37: Efficiency analysis of Krishnappa Gowtham

Performance Metrics: Similarly, Krishnappa Gowtham is 100 % efficient, which means all actual variables are meeting target variable.

Slacks: The slack values for all three metrics (Balls, Boundaries, Pr6) are zero. This indicates that there is no excess or shortage in any of the inputs or outputs concerning the target values. In other words, Krishnappa Gowtham is achieving the desired level of performance in all of these metrics.

Weights: The weights assigned to the metrics are provided. It appears that only the "Balls" and "Pr6" metrics are given non-zero weights. The weight for "Balls" is 0.1, and the weight for "Pr6" is 0.3. These weights represent the relative importance of each metric in the DEA analysis.

Krishnappa Gowtham is operating efficiently with respect to the "Balls" and "Pr6" metrics, achieving the target outputs for both without any excess or shortage in inputs. The "Boundaries" metric does not contribute to Krishnappa Gowtham's efficiency score in this particular analysis due to its zero weight. The efficiency score 100 operating at maximum efficiency based on the weighted metrics considered in this DEA model.

Sam curran

Sam Curran has faced 225 balls. The target set for him is also 225 balls. This shows that Curran, similar to Gowtham, has been at the crease for a considerable duration, reflecting his all-rounder role where he contributes both with the bat and ball. He has hit 48 boundaries. The set target is 58.04 boundaries. Curran falls short of the target, suggesting there's room to enhance his boundary-hitting capability to match the efficiency of his peers. 32.05% of his boundaries are sixes. The set target is 41.65%. Curran's percentage is below the target, indicating he could benefit from elevating his power-hitting game to clear the boundary more frequently. The weight associated with balls faced is 0.01, emphasizing the significance of his involvement in the game. The weight for boundaries stands at 0.02, which underscores the importance of boundary-hitting in T20 cricket. Interestingly, the weight for the percentage of boundaries as sixes is zero, suggesting that in this context, just hitting boundaries (be it fours or sixes) is more vital than the nature of the boundary.

5 DISCUSSION AND FINDINGS

5.1 EFFICIENT PLAYERS FOR OPENERS

Discussion

In this section, we delve into the key findings and implications of our cricket analysis project. We address the significance of our results, their alignment with existing knowledge, and the broader impact on the sport. Additionally, we highlight the limitations of our study and suggest directions for future research.

1. Batting Position Efficiency Analysis

Our study focused on assessing the efficiency of cricket batsmen in different batting positions. The efficiency scores, as determined by the Data Envelopment Analysis (DEA) model, provided valuable insights into player performance. The 100% efficiency achieved by top-order batsmen like Jos Buttler and Lokesh Rahul highlights their proficiency as openers. Their aggressive yet consistent batting styles, reflected in high Runs Per Over (RPO) and average runs, underscore their effectiveness in setting the tone for their teams.

2. Weight Assignments and Significance

The weight assignments in the DEA model played a pivotal role in evaluating player efficiency. For instance, the higher weight assigned to 'FOUR STRIKE RATE' for Jos Buttler emphasized its importance in his efficiency assessment. This finding aligns with the cricketing wisdom that aggressive openers often dictate the tempo of the game, with boundary-hitting prowess being a key contributor.

3. Slack Values and Optimal Performance

The zero slack values across all metrics for the analyzed players indicate optimal performance, closely aligned with target values. This demonstrates that these players excel in their respective roles, leaving no room for improvement in the measured parameters. These findings provide valuable benchmarks for teams and selectors when evaluating player efficiency.

4. Implications for Team Strategy

Efficient players not only contribute significantly to their teams but also influence overall team strategies. Our analysis of Parthiv Patel, who excels in both aggressive and steady roles, suggests that teams can benefit from versatile players who can adapt to different match situations. Patel's ability to accelerate the run rate or anchor the innings provides valuable options for team strategists.

5.2 EFFICIENT TOP MIDDLE ORDER BATSMEN

Interestingly, the non-zero weights for average balls faced and average score, as opposed to zero weights for innings, strike rate, and boundaries, reveal the model's priorities. It emphasises the importance of consistency and the ability to get set in the crease over just aggressive scoring.

Livingstone's 100% efficiency stands out in the world of T20 cricket, when many players alternate between explosive performances and slumps. It demonstrates his adaptability, consistency, and ability to meet the varying demands of various match circumstances.

Finally, Liam Livingstone's performance serves as a gold standard for T20 cricket efficiency. His continuous ability to meet targets across multiple measures demonstrates his great expertise, versatility, and important commitment to the team.

5.3 EFFICIENT PLAYERS FOR LOWER ORDER

1.Russell

While Russell's boundary count is impressive, the weightage provided to the Pr6 metric illuminates what sets him apart: his extraordinary ability to hit sixes. In the high-octane environment of T20 cricket, sixes can shift momentum, demoralize bowlers, and change the complexion of a match within a few deliveries. This is where Russell excels and why his six-hitting prowess is weighted more significantly than just boundaries.

However, the zero weights for balls faced and total boundaries could raise questions. Does this mean these aspects are not vital for a player like Russell? Not necessarily. It likely signifies that in the context of this Data Envelopment Analysis (DEA), the distinctive attribute that showcases Russell's efficiency best is his six-hitting capability. This doesn't diminish the importance of other metrics but rather places emphasis on what truly sets him apart in the T20 format.

In conclusion, while boundaries and balls faced are integral components of any T20 batsman's arsenal, for a player like Andre Russell, it's the sheer volume of sixes that defines his efficiency and impact on a match. This analysis suggests that Russell's ability to consistently clear the boundary is his most valuable trait, making him one of the most formidable T20 players in the world.

2.Krishnappa Gowtham

Krishnappa Gowtham's metrics portray him as a balanced player who can contribute effectively with the bat. His boundary count, coupled with a high percentage of sixes, underscores his role as an impact player who can accelerate when needed. The weights associated with each metric provide a nuanced understanding of his gameplay. While the balls he faces are important, it's his six-hitting prowess that stands out as a defining trait. This ability to hit sixes, especially in the latter stages of an innings, can be a game-changer in T20 cricket.

In conclusion, considering the actual values aligning perfectly with the target values, Gowtham exemplifies efficiency in his role. His capability to mix boundary-hitting with the knack for clearing the fence makes him a valuable asset in T20 cricket. The weights in the DEA further accentuate the significance of his six-hitting ability, marking him as a player who can change the course of a game in a few deliveries.

6 CONCLUSION

6.1 SUMMARY.

- A significant disparity was observed in the statistical means of runs and strike rates between the lower middle order and openers in cricket batting. Openers exhibited higher runs, balls played, and batting average, while the lower middle order displayed a higher strike rate and a greater percentage of runs scored through sixes.
- Utilizing correlation analysis, the most relevant input and output variables were selected for further analysis.
- The Variable Returns to Scale (VRS) model was employed to identify efficient batsmen across all possible batting orders. The choice of output orientation in the VRS model played a crucial role in determining efficiency.
- For the opener's position, benchmark batsmen identified included Jos Buttler, Lokesh Rahul, Prithvi Shaw, Devon Conway, Matthew Wade, Manan Vohra, Parthiv Patel, and Brendon McCullum.
- In the top middle-order position, benchmark batsmen were Bhanuka Rajapaksa, Sunil Narine, Liam Livingstone, Colin Munro, Rishabh Pant, Rajat Patidar, Suryakumar Yadav, and AB de Villiers. This selection was based on inputs such as innings and balls played, with outputs including boundaries, average, and strike rate.
- For the lower middle-order position, benchmark batsmen included Andre Russell, Rovman Powell, and Krishnappa Gowtham. The analysis considered inputs like balls faced and boundaries, with outputs focused on Pr6 (possibly indicating the number of sixes hit).
- Weighting factors played a pivotal role in determining a batsman's efficiency, emphasizing the importance of specific performance metrics in the assessment of cricket batting performance.

6.2 LIMITATION

The availability of relevant literature in the context of cricket was noticeably low, making acquiring full insights difficult. Furthermore, the dataset employed for analysis was severely limited, owing mostly to the batting position segmentation. This segmentation restricted the sample even further, lowering the amount of data accessible for investigation.

Furthermore, natural elements such as weather, pitch qualities, or external influences, which might have a substantial impact on player performance, were not included in the dataset. The lack of these natural factors in the dataset limits the depth of research and the ability to make broader conclusions.

It should be noted that the dataset's segmentation made it less suited for comparisons with other models or the use of machine learning techniques. The data's specialised nature, focusing on batting positions, limited its applicability for larger analytical applications. As a result, the dataset's utility was limited primarily to the setting of batting position analysis.

6.3 FUTURE WORK

The dissertation has emphasised the critical need of position-specific analysis of batsmen, particularly in the fast-paced game of cricket. While the current analysis gives useful insights, there is clearly room for improvement by incorporating other datasets, such as ball-by-ball statistics and data capturing critical match moments. These enhancements will greatly widen and deepen the scope and depth of efficiency analyses.

We may acquire a more thorough knowledge of a batsman's performance by incorporating specific ball-by-ball statistics, such as their ability to adjust to different match scenarios and their decision-making under pressure. This expanded dataset will allow for a more comprehensive assessment of efficiency and effectiveness in T20 cricket.

Furthermore, the existing model must be viewed as a basic framework that may be used as a starting point for future assessments. The model can evolve and adapt to incorporate multiple performance metrics, player qualities, and contextual elements as more broad and diverse datasets become available. This process allows for continuing improvement in the analysis's accuracy and relevance.

Another factor to consider is determining which batting position draws different types of fans. Understanding fan preferences for certain batting positions can help with fan engagement and marketing strategies. Certain batting stances, for example, may be more appealing to fans looking for spectacular sixes, but others may prefer consistent and methodical play. This analysis can be beneficial.

Peer evaluation is still an important factor to consider in sports-related research since collaborative efforts with other academics, coaches, analysts, and fans can provide valuable input, validate findings, and provide varied views. This all-inclusive approach guarantees that the analysis is thorough, adheres to industry standards and best practises, and is well received by the cricket community at large.

In summary, expanding data sources, constant model development, collaborative peer evaluation, and analysing fan preferences for different batting positions are critical steps in enhancing batters efficiency analyses in T20 cricket. This comprehensive strategy not only delivers actionable insights for players and teams, but it also improves the whole fan experience, building a stronger bond between the sport and its fans."

7 REFERENCE

Amin, G.R. and Sharma, S. kumar (2012). Cricket team selection using data envelopment analysis. European Journal of Sport Science, 14(sup1), pp.S369–S376. doi:https://doi.org/10.1080/17461391.2012.705333.

Aripaka, R. (2023). IPL: How India became home to the biggest, baddest cricketing league in the world. The Economic Times. [online] 30 Apr. Available at: https://economictimes.indiatimes.com/news/sports/ipl-how-india-became-home-to-the-biggest-baddest-cricketing-league-in-the-world/articleshow/99885321.cms.

Arsu, T. (2021). Investigation into the efficiencies of European football clubs with bi-objective multi-criteria data envelopment analysis. Decision Making: Applications in Management and Engineering, 4(2), pp.106–125. doi:https://doi.org/10.31181/dmame210402106a.

Asif, M. and McHale, I.G. (2016). In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model. International Journal of Forecasting, 32(1), pp.34–43. doi:https://doi.org/10.1016/j.ijforecast.2015.02.005.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, [online] 30(9), pp.1078–1092. Available at: https://www.jstor.org/stable/2631725.

Banker, R.D. and Thrall, R.M. (1992). Estimation of returns to scale using data envelopment analysis. European Journal of Operational Research, 62(1), pp.74–84. doi:https://doi.org/10.1016/0377-2217(92)90178-c.

Chaudhary, R., Bharadwaj, S. and Lakra, S. (2018). A DEA Model for Selection of Indian Cricket Team Players. [online] Available at: https://www.researchgate.net/publication/332755063_A_DEA_Model_for_Selection_of_Indian_Cricket_Team_Players/references.

Davis, J., Perera, H. and Swartz, T.B. (2015). Player evaluation in Twenty20 cricket. Journal of Sports Analytics, [online] 1(1), pp.19–31. doi:https://doi.org/10.3233/jsa-150002.

Delacre, M., Lakens, D. and Leys, C. (2017). Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. International Review of Social Psychology, 30(1), p.92. doi:https://doi.org/10.5334/irsp.82.

Dennis , J., Shields , B. and Hattersley, A. (2019). Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. The Lancet Diabetes & Endocrinology, [online] 7(6), pp.442–451. doi:https://doi.org/10.1016/S2213-8587(19)30087-7.

Dennis, B., Ponciano, J.M., Taper, M.L. and Lele, S.R. (2019). Errors in Statistical Inference Under Model Misspecification: Evidence, Hypothesis Testing, and AIC. Frontiers in Ecology and Evolution, 7. doi:https://doi.org/10.3389/fevo.2019.00372.

ESPNcricinfo. (2022). 'We made history when nobody was expecting us to win' - Lalchand Rajput looks back at 2007 T20 World Cup triumph. [online] Available at: https://www.espncricinfo.com/story/india-s-2007-t20-world-cup-win-lalchand-rajput-we-created-history-1336348 [Accessed 28 Sep. 2023].

ESPNcricinfo. (n.d.). ICC World Twenty20 | 2007/08 ICC World Twenty20 | Live Score, Schedule, News. [online] Available at: https://www.espncricinfo.com/series/icc-world-twenty20-2007-08-286109 [Accessed 28 Sep. 2023].

ESPNcricinfo. (n.d.). KKR vs RCB Scorecard 2007/08 | Cricket Scorecard. [online] Available at: https://www.espncricinfo.com/series/indian-premier-league-2007-08-313494/royal-challengers-bangalore-vs-kolkata-knight-riders-1st-match-335982/full-scorecard.

Freeman, D., Lambe, S., Kabir, T., Petit, A., Rosebrock, L., Yu, L.-M., Dudley, R., Chapman, K., Morrison, A., O'Regan, E., Aynsworth, C., Jones, J., Murphy, E., Powling, R., Galal, U., Grabey, J., Rovira, A., Martin, J., Hollis, C. and Clark, D.M. (2022). Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. The Lancet Psychiatry, [online] 0(0). doi:https://doi.org/10.1016/S2215-0366(22)00060-8.

Ghai, K. and Zipp, S. (2020). Governance in Indian cricket: Examining the Board of Control for Cricket in India through the good governance framework. Sport in Society, 24(5), pp.1–16. doi:https://doi.org/10.1080/17430437.2020.1819598.

Gweshe, T. and Durbach, I.N. (2013). An analysis of the efficiency of player performance at the 2011 Cricket World Cup. Orion. [online] Available at: https://open.uct.ac.za/handle/11427/25856 [Accessed 28 Sep. 2023].

India win dramatic Twenty20 final. (2007). news.bbc.co.uk. [online] 24 Sep. Available at: http://news.bbc.co.uk/sport1/hi/cricket/7009035.stm [Accessed 28 Sep. 2023].

Kanungo, V. and B, T. (2019). Data visualization and toss related analysis of IPL teams and batsmen performances. International Journal of Electrical and Computer Engineering (IJECE), 9(5), p.4423. doi:https://doi.org/10.11591/ijece.v9i5.pp4423-4432.

Karhadkar, A. (2018). How did IPL become a success story? The Hindu. [online] 26 May. Available at: https://www.thehindu.com/sport/cricket/how-did-ipl-become-a-success-story/article24001697.ece [Accessed 28 Sep. 2023].

Lv, S., Zhao, S. and Liu, H. (2022). Research on the Coupling between the Double Cycle Mode and Technological Innovation Systems: Empirical Evidence from Data Envelopment Analysis and Coupled Coordination. Systems, [online] 10(3), p.62. doi:https://doi.org/10.3390/systems10030062.

Manage, A.B.W. and Scariano, S.M. (2013). An Introductory Application of Principal Components to Cricket Data. Journal of Statistics Education, 21(3). doi:https://doi.org/10.1080/10691898.2013.11889689.

Papagapiou, A., Mingers, J. and Thanassoulis, E. (1997). Would you buy a used car with DEA? OR Insight, 10(1), pp.13–19. doi:https://doi.org/10.1057/ori.1997.3.

Perera, H.P. and Swartz, T.B. (2012). Resource estimation in T20 cricket. IMA Journal of Management Mathematics, 24(3), pp.337–347. doi:https://doi.org/10.1093/imaman/dpr020.

psycnet.apa.org. (2017). APA PsycNet. [online] Available at: https://psycnet.apa.org/record/2017-30365-001 [Accessed 28 Sep. 2023].

Saleh, M.M., Salem, K.F.M. and Elabd, A.B. (2020). Definition of selection criterion using correlation and path coefficient analysis in rice (Oryza sativa L.) genotypes. Bulletin of the National Research Centre, 44(1). doi:https://doi.org/10.1186/s42269-020-00403-y.

Singh, S. (2011). Measuring the Performance of Teams in the Indian Premier League . [online] Available at: https://www.researchgate.net/publication/268400474_Measuring_the_Performance_of_Teams_in_the_Indian_Premier_League/figures?lo=1.

Thomson, J., Perera, H. and Swartz, T.B. (2021). Contextual batting and bowling in limited overs cricket. South African Statistical Journal, 55(1), pp.73–86. doi:https://doi.org/10.37920/sasj.2021.55.1.6.

Watanabe, T. and Mizukami, H. (2018). Effectiveness of Tactile Scatter Plots: Comparison of Non-visual Data Representations. Lecture Notes in Computer Science, pp.628–635. doi:https://doi.org/10.1007/978-3-319-94277-3_97.

Wikipedia. (2023). 2007 ICC World Twenty20. [online] Available at: https://en.wikipedia.org/wiki/2007_ICC_World_Twenty20#cite_note-india_win- [Accessed 28 Sep. 2023].

Wolfgang, K. (2006). On the Analysis of Fuzzy Life Times and Quality of Life Data. Springer eBooks, pp.446–451. doi:https://doi.org/10.1007/0-387-26023-4_30.

Yin, W., Ye, Z. and Shah, W.U.H. (2023). Indices Development for Player's Performance Evaluation through the Super-SBM Approach in Each Department for All Three Formats of Cricket. Sustainability, 15(4), p.3201. doi:https://doi.org/10.3390/su15043201.

Zhao, F. and Gaschler, R. (2021). Best Graph Type to Compare Discrete Groups: Bar, Dot, and Tally. Frontiers in Psychology, 12. doi:https://doi.org/10.3389/fpsyg.2021.775721.