Classification and Bayes Probability Approach for Evaluating the Performance of T-20 Batsmen

Mohib Ullah Khan[a], Qamruz Zaman[*b],Mujeeb Hussain[c], Sofia[d], Syed Habib Shah[e],Neelam[f]

[a,b,f]Department of Statistics, University of Peshawar, Pakistan. [c]Government College Peshawar, Pakistan. [d]College of Home Economics,University of Peshawar, Pakistan. [e]Institute of Numerical Sciences Kohat University of Science

[*]**Email:** cricsportsresearchgroup@gmail.com

Abstract: This thesis presents a comprehensive analysis of the performance dynamics of top-ranked T20I batsmen in the ICC rankings. By considering factors such as player probabilities against specific teams, physical attributes like height and weight, and the implications of Bayesian probabilities, a multifaceted understanding of player achievements and their associations has been revealed. Through individual player probability examinations, distinctive insights emerged, such as Babar Azam's varying probabilities against different teams, highlighting the diverse challenges faced. Aiden Kyle Markram's strong performances against Sri Lanka underscored the opponent's influence on achievements. The study identified players excluded due to low probabilities against top-ranked teams, emphasizing strict "Man of the Match" criteria and the need for consistent excellence. Nicholas Pooran's mixed performance outcomes demonstrated the limitation of singular matches for assessing overall prowess. The investigation of height and weight correlation among T20I batsmen unveiled a lack of statistical significance, challenging assumptions. The classification tree visually portrayed attributes and performance across continents, affirming the influence of regional factors on player characteristics and achievements. This research underscores the complexity of player success and calls for a holistic evaluation of performances in diverse contexts.

Keywords: Performance Dynamics, Bayesian Probabilities, Player Attributes Statistical Significance, Regional Influence, Chi-Square test, Odd Ratio

1. Introduction

The passage discusses the role of statistics in various fields, with a particular focus on its application in sports, especially cricket. It explains how statistics, a branch of mathematics, is used to collect, analyze, interpret, and present data, helping in decision-making and identifying patterns across diverse fields like economics, healthcare, and sports. Statistics in sports has expanded beyond traditional reliance on physiology and medicine, incorporating economics, sociology, and law to assess performance. In cricket, statistical methods like time series, linear regression, and various distributions (e.g., Poisson, Gamma) are widely used to evaluate player and team performance. The discipline of sports statistics relies on techniques such as least squares, maximum likelihood, and Bayesian methods for parameter estimation and simulation models. Cricket is highlighted as a game with vast amounts of data, earning it the nickname "the game of statisticians' delight." The sport generates numerous statistics, which help make decisions during games and evaluate player performance. The sport is globally popular, particularly in countries under the International Cricket Council (ICC). Major international cricket tournaments are organized by the ICC, and the game is played in three formats: Test matches, One-Day Internationals, and

Twenty20 (T-20). Cricket also has a significant financial market, with legal and illegal betting valued at billions of dollars, primarily driven by the Asian markets. The importance of statistical analysis in this domain is evident, though limited research has been conducted on cricket gambling. Despite being a traditional sport, cricket presents unique opportunities for research and statistical analysis, adding to its cultural and economic significance.

1.1 The Genesis of Twenty20 Cricket

In response to declining spectator attendance, the England and Wales Cricket Board (ECB) introduced the Twenty20 cricket format in 2003, spearheaded by marketing manager Stuart Robertson. This new format replaced the 50-over Benson and Hedges Cup and shortened games to 20 overs per side, making matches more accessible by reducing the time commitment to about three hours. The first Twenty20 Cup in England was a huge success, capturing the attention of many fans with its fast pace and thrilling nature. Though initially intended as a domestic initiative, Twenty20 cricket quickly spread globally, with Australia, South Africa, and other cricket-playing nations adopting the format. Despite initial resistance, particularly from the Board of Control for Cricket in India (BCCI), the Twenty20 format was eventually incorporated into international cricket. The first international Twenty20 match was played between Australia and New Zealand in February 2005, followed by other countries hosting their inaugural matches. India was the last of the test-playing nations to embrace Twenty20, playing their first match in December 2006 against South Africa. The format's rapid rise marked a significant shift in cricket's global landscape.

1.2 ICC WT20

The ICC World Twenty20 tournament is held biennially, except when it coincides with the ICC Cricket World Cup, in which case it occurs the year before. The inaugural event in 2007, held in South Africa, saw India claim victory over Pakistan in the final. The tournament included two Associate teams chosen from the 2007 ICC World Cricket League Division One. In December 2007, a qualifying tournament was introduced using a 20-over format, with six teams competing for two spots in the 2009 World Twenty20 and a $250,000 prize.Subsequent tournaments saw Pakistan win the 2009 edition in England, followed by England's victory over Australia in 2010, and the West Indies' triumph over Sri Lanka in 2012. The 2014 ICC World Twenty20, held in Bangladesh, saw Sri Lanka defeat India in the final, while the West Indies won again in 2016 by beating England in India. The growing popularity of T20 cricket has also led to the rise of domestic leagues, such as the Pakistan Super League (PSL), which started in 2015 with five teams and expanded to six teams by its third season. The PSL features top international players and is known for its competitive nature in both batting and bowling.

1.3 Physiological responses to batting in cricket

Shah et al. (2017) discussed various statistics and metrics used to evaluate cricket players' performances across different aspects of the game, focusing on key performance indicators such as batting, bowling, fielding, and all-round performance. For batting performance, five significant measures are highlighted: the highest individual score (HS), batting average, strike rate (SR), number of fours (4's), and number of sixes (6's). The highest individual score refers to the most runs a batsman score in a single match. Batting average is calculated by dividing total runs scored by the number of times the batsman was out. Strike rate is determined by dividing runs scored by the number of balls faced. Additionally, the number of fours and sixes hit by a batsman is tracked as 4's and 6's, respectively.

1.4 Batsman

A batsman's primary role in cricket is to score runs and defend their wicket. They need to accumulate runs by finding gaps in the field and hitting boundaries, while also protecting their stumps from the bowler's deliveries. To succeed, a batsman must have technical skills, such as good footwork and shot selection, along with the ability to adapt to different playing conditions and bowlers. Mental strength and concentration are also crucial, especially during long innings or high-pressure situations. Additionally, a batsman plays according to the team's strategy, whether it involves building a steady total or chasing a target aggressively.

1.4.1 Scoring Runs

The main responsibility of a batsman is to score runs by hitting the ball delivered by the bowler. The batsman scores run by running between the wickets or hitting boundaries (Four runs are scored when ball hits the ground

before hitting or passing the boundary and six runs if it clears the boundary on the full).

1.4.2 Defending the Wicket
Another crucial aspect of a batsman's role is to protect their wicket. The wicket consists of three stumps and two bails, and the batsman's primary goal is not to get out. A batsman can be dismissed in various ways, such as being bowled (the ball hitting the stumps and dislodging the bails), caught by a fielder, leg before wicket (LBW), stumped by the wicketkeeper, hit-wicket, or run out.

1.4.3 Technique and Footwork
Batsmen employ various techniques and footwork to play different types of deliveries. They need to judge the line (the direction of the ball concerning the stumps) and length (the point where the ball bounces) of each delivery to decide how to play the shot.

1.4.4 Batting Order
Batsmen in a cricket team are usually arranged in a batting order, which determines the sequence in which they bat. The top-order batsmen are typically more specialized in building partnerships and scoring quickly, while the lower-order batsmen might focus more on batting with the tail-enders and adding crucial runs.

1.4.5 Opening Batsmen
The opening batsmen are the first two players in the batting order. They face the new ball and play a critical role in setting the tone for the innings. They need to negotiate the early swing and bounce of the ball, often against some of the best bowlers of the opposing team.

1.4.6 Middle-order Batsmen
The middle-order batsmen follow the openers and are responsible for stabilizing the innings and building partnerships. They often face the second new ball and play a crucial role in setting up a competitive total.

1.4.7 Tail-end Batsmen
The tail-end batsmen are the last few players in the batting order who are generally not as skilled with the bat compared to the specialist batsmen. However, they can still contribute valuable runs to extend the team's total.

1.4.8 Batting Styles
Batsmen have different styles of play, such as aggressive, defensive, or a combination of both. Some players are known for their power hitting, while others rely on their technique and patience to accumulate runs.

1.5 Significance of the study
Many works have been done on the performance of the batsman, but nothing has been done on the conditions (e.g., player of the match, average score, height, and weight). The purpose of the study is to measure the posterior probability by which we forecast for the future through the use of Bayes' theorem. The posterior probability is a type of conditional probability that is constructed by updating the prior probability using data described by the likelihood. Also, check the association between height and weight through the Chi-square test.

1.6 Research Aim and Objectives
1. To manipulate the Bayes probabilities of the performance of top-ranked batsmen in T-20I cricket.
2. To study the effect of height and weight on the performance of top-ranked batsmen in T-20I cricket.
3. To assess the efficiency of the batters using classification trees.

2. Literature Review
Quazi et al. (2022) introduced a statistical modeling approach to predict outcomes of upcoming cricket tournaments. The model uses survey sampling techniques to analyze team selection patterns based on individual players' performance history. Bayesian priors assess players' past performances against various opponents, including all ICC-recognized nations. A case study focused on the ICC World Cup 2023, and simulation results were thoroughly examined. To validate the model, the 2020 Indian Premier League season was used, where the

model accurately predicted the winning probabilities for the participating teams.

Nawaz et al. (2012) explored the use of quality control charts as a novel approach to evaluating cricket player performance. While traditional metrics like mean ratings, strike rates, and mean scores per dismissal have been widely used, this study introduces individual and moving range control charts to assess the performance of two prominent batsmen. By applying sensitizing criteria, the study identifies non-random and artificial variations in performance, allowing analysts to investigate the causes of out-of-control scores. The findings aim to help detect and correct these deviations, leading to improved player performance.

Maddox et al. (2022) introduced two innovative Bayesian models for predicting the likelihood of a home team winning during a game. The first model uses a prior that adjusts based on factors like lead differential and time elapsed. The second, an improved version, combines the Bayesian estimator with a time-weighted pre-game success probability. When compared to existing methods, these new approaches showed better accuracy in both estimation and prediction. Their practical application was demonstrated using the 2016 NCAA Division 1 Championship game, highlighting their effectiveness.

Kapadiya et al. (2020) emphasized the importance of selecting the best players for a team to ensure overall success, particularly in cricket. The selection process involves considering various factors such as individual performance, ground conditions, weather forecasts, and the opposition's strengths and weaknesses. The study introduces a machine learning model designed to predict player performance by leveraging historical data on team and individual achievements. Unlike previous models, this new approach incorporates features related to ground conditions and weather, aiming to enhance team selection and improve overall team performance.

Koulis et al. (2014) explores the use of a batsman-specific hidden Markov model (HMM) to assess batting efficiency in one-day international (ODI) cricket. This model integrates various dynamics of batting performance and uses parallel sampling to identify hidden states. Performance is evaluated through reliability analysis and metrics such as availability, reliability, failure rate, and mean time to failure. By categorizing states as "up" or "down," the model provides forecasts of a batsman's performance, adaptable to different game contexts. It effectively handles unforeseen elements and recent rule changes, with data from 20 batters representing different countries demonstrating its practical application.

Premkumar et al. (2020) focus on improving cricket player rankings by incorporating new criteria and key performance indicators, moving beyond traditional methods like the ICC rating system. The study introduces a dynamic methodology that generates factor scores for players on a match-by-match basis using factor analysis. This approach provides more nuanced rankings for batsmen and bowlers in One Day International (ODI) cricket for the 2015 calendar year, addressing limitations of previous static ranking systems.

Wickramasinghe (2020) emphasizes the importance of player classification in cricket for assigning roles and responsibilities. The study aims to categorize all-rounders into four distinct groups within One Day International (ODI) cricket and predict potential new all-rounders. Utilizing a dataset of 177 players and ten performance metrics, the study applies three artificial intelligence classifiers: Naive Bayes (NB), k-nearest neighbors (kNN), and Random Forest (RF). The results show that Random Forest achieves significantly higher prediction accuracy, reaching 99.4%, outperforming the other classifiers.

Santos-Fernandez et al. (2019) review the growing application of Bayesian methods in sports analytics, emphasizing their strengths in handling complex problems, providing probability estimates and forecasts, and integrating various data sources. The increasing use of Bayesian approaches is linked to the expansion of sports data and advancements in Bayesian software. The review highlights recent developments in sports such as basketball, football, and baseball, noting that research predominantly focuses on modeling game outcomes and analyzing athlete performance. Additionally, the volume of Bayesian method publications from 2013 to 2018 matches the total from the previous three decades, indicating a significant rise in their use in sports analytics.

Purkayastha (2017) focuses on enhancing the accuracy of cricket match prediction models. The study emphasizes the role of feature engineering and attribute selection in improving classifier performance. Three classifiers—k-NN, Naive Bayes, and Random Forest—were used, with Random Forest achieving the highest accuracy at 84%. Naive Bayes followed with an accuracy of 83%, and k-NN was third with 82%.

Shah (2017) examines the importance of context in evaluating cricket players' performances for effective squad selection. Traditional metrics like batting and bowling averages often fail to reflect the quality of performances, especially in challenging situations. The study proposes a method that considers both the bowler's average and the batsman's average against each bowler to better quantify performance. By aggregating individual performances and creating comprehensive performance indices for both batsmen and bowlers, the study offers a more nuanced

approach to ranking players.

Wickramasinghe and Indika Pradeep (2014)
This study focuses on forecasting batsmen's performance in Test cricket using a three-stage hierarchical linear model. The model incorporates player, team, and match series features and accounts for factors like handedness and team rank. It is based on longitudinal data collected over five years and accurately predicts player performance.

Manage et al. (2013)
This paper addresses the challenge of analyzing cricket players' performance with interconnected criteria. It proposes a new, transparent method for evaluating player performance in the T20 World Cup 2012, which can be adapted for other team sports, aiming to improve clarity in performance assessment.

Lemmer et al. (2011)
The study introduces a method to compare batsmen's scores by adjusting for playing conditions. It demonstrates that traditional averages are inadequate for assessing performance in varying conditions and provides an adjusted run total method using data from the Indian Premier League 2009.
  Sharp et al. (2011). This paper explores methods to calculate cricket player performance for team composition in the financially lucrative Twenty20 format. It uses performance metrics to inform team selection and demonstrates its approach with data from the 2007 Twenty20 World Cup.
  Scarf et al. (2011). This research uses negative binomial distributions to analyze partnership and innings scores in Test cricket. It includes run rate as a covariate to evaluate batting strategies and match outcomes, helping captains decide between aggressive and defensive strategies.**Taliep et al. (2010).** The study investigates the link between upper body muscle strength and batting performance in cricket. It finds a correlation between strength and hitting distance but no significant relationship with batting average or strike rate in One Day or T20 matches.
  Saikia et al. (2010). This research assesses the performance of all-rounders in the Indian Premier League using strike rate and economy rate. It classifies all-rounders into categories and develops a Bayesian classifier to predict future performance, aiding team management in player selection.**Saikia et al. (2010).**The study examines the impact of winning the toss and playing at home on T20 match outcomes. It uses binomial logistic regression to analyze data from international T20 matches, considering the effect of match time and location.
  Lemmer (2004). This research develops a comprehensive measure to evaluate cricket batsmen by combining batting average, strike rate, and consistency. It introduces a consistency coefficient to assess batsman performance and creates a classification system for ranking players.**Preston et al. (2000).**The study uses dynamic programming to optimize batting strategies in limited-overs cricket. It distinguishes between target-setting and chasing strategies and applies a statistical framework to analyze batting behavior in English county-level matches.
  Asif et al. (2019). This paper introduces a generalized non-linear forecasting model (GNLM) to predict remaining runs in cricket innings. It focuses on estimating second innings runs in T20 matches and provides real-time run difference estimates, with applications to target revision models.

3. Methodology
3.1 Universe of the Study
The research investigation focuses on evaluating the performance of top-ranked batsmen in T20 International (T20I) cricket using a Bayesian approach. It also examines the relationship between height and weight through a chi-square test. Additionally, the study evaluates classification models based on various factors using a probability approach.

3.2 Source of Data
Data for the study was collected from ESPN Cricinfo, MRFTyers (ICC Ranking), and Cricbuzz up to April 10, 2022. The analysis focused on the top-ranked T20 International batsmen based on the International Cricket Council (ICC) rankings, which are freely available on these websites

3.3 Bayesian Analysis
Bayes' theorem, introduced by Thomas Bayes in the 18th century, calculates conditional probability based on additional information. It is widely used in banking and finance for assessing lending risks. Bayesian statistics,

grounded in Bayes' theorem, involves updating our understanding of statistical parameters by combining prior knowledge with observed data. This process results in a posterior distribution, which can be used for predictions and continuous improvement of knowledge.

3.4 Objectives of Bayesian Analysis
Bayesian analysis is useful when dealing with extensive and varied data, allowing for precise insights. It integrates prior probabilities with observed data to produce an updated posterior probability distribution. This updated information is then used to make well-informed decisions with improved understanding.

3.5 Bayesian Estimation
Bayesian estimation provides an effective method for modeling real-world problems by incorporating prior information. Unlike classical methods, it treats parameters as random variables, offering a more flexible and comprehensive analysis. This approach helps in determining the best estimates for unknown parameters, improving the accuracy of the estimation.The Bayes probability theorem is expressed by the following,

$$P(A/B) = \frac{P(A).P(B/A)}{P(B)} \tag{1}$$

A represent the first event, B represent the second event. If there are multiple conditions, the extended Bayes theorem is applicable.

$$P(Ai/B) = \frac{P(Ai).P(B/Ai)}{\sum P(Ai).P(B/Ai)} \tag{2}$$

(OR)

$$\text{Bayes probability} = \frac{(\text{Prior probability}).\,(\text{Conditional probability})}{\sum(\text{Prior probability}).\,(\text{Conditional probability})} \tag{3}$$

Where, $P(Ai/B)$ = Bayes probability or Posterior probability, $P(Ai)$ = Prior probability, $P(B/Ai)$= Conditional probability, $\sum P(Ai).P(B/Ai)$ = Sum of Prior and Conditional probability, The performance of the Bayes theorem in this research is evaluated as follows:

$$P \text{ (Name of player)}/\text{(player of the match)} = \frac{P(\frac{\text{no.of matches against country}}{\text{(total no of matches)}}).P(\frac{\text{player of the match}}{\text{no.of matches} \geq 30})}{\sum P(\frac{\text{no.of matches against country}}{\text{(total no of matches)}}).P(\frac{\text{player of the match}}{\text{no.of matches} \geq 30})} \tag{4}$$

P (Name of player)/(player of the match) =Bayes probability OR Posterior Probability (5)

$$P(\frac{\text{no.of matches against country}}{\text{(total no of matches)}}) = \text{Prior probability} \tag{6}$$

$$P(\frac{\text{player of the match}}{\text{no.of matches} \geq 30}) = \text{Conditional probability} \tag{7}$$

$$\sum P(\frac{\text{no.of matches against country}}{\text{(total no of matches)}}).P(\frac{\text{player of the match}}{\text{no.of matches} \geq 30}) = \text{Summation of PriorandConditional probability} \tag{8}$$

(OR)

$$P(\text{Name of player})/\text{(player of the match)} = \frac{\text{Joint probability}}{\text{Sum of Joint probability}} \tag{9}$$

3.6 Chi-square test
The chi-square test is frequently used to analyze data in a contingency table, displaying the frequency distribution of two categorical variables. This test compares observed frequencies with expected frequencies to determine if the variables are independent. It is a widely recognized method for hypothesis testing with discrete data, helping assess whether a relationship exists between two variables or if they are independent (Snedecor et al., 1989).

3.7 Chi-square goodness of fit test
The chi-square test relies on two main conditions: (a) independence among measures, and (b) reasonably sized theoretical frequencies. It is used to assess whether observed categorical data aligns with a specified distribution, such as binomial, Poisson, or normal. The specific distribution is less important than its fundamental properties. In a goodness of fit test, the chi-square test statistic is calculated to determine if the observed data matches the expected distribution.

$$\chi^2 = \sum (\frac{(Oi - Ei)^2}{Ei}) \tag{10}$$

Where, $\chi^2$is the chi-square test statistic,Σ denotes the sum of the calculations across all categories,Oi represents the

observed frequency in each category, Ei represents the expected frequency in each category. Now, the average height of the "n" batsmen is calculated and denoted by,

$$Average\ height = \frac{H1+H2+H3+ \text{.........}+Hn}{n} \qquad (11)$$

Similarly, the average weight of "n" players is calculated and denoted by,

$$Average\ height = \frac{W1+W2+W3+ \text{.........}+Wn}{n}$$ Average height=H1+H2+H3+ ………+Hnn

(11)

Similarly, the average weight of "n" players is calculated and denoted by,

Average height=W1+W2+W3+ ………+Wnn   (12)

Table 3.1: 2×2 contingency table

Cochran (1954) researched how to determine if success probabilities are equal for different treatments in contingency tables. He recommended comparing these probabilities by aggregating the total number of successes for one treatment across all tables.

Height record	Weight record		Total
	Less than Average	Greater than Average	
Less than Average	a	b	a+b
Greater than Average	c	d	c+d
Total	a+c	b+d	a+b+c+d

3.7 Classification

Classification involves assigning objects to predefined groups based on their attributes and is widely used in fields like business, science, industry, and medicine. It is a key data mining approach for predicting group membership of data instances. Various classification methods include decision trees, Bayesian networks, k-nearest neighbors, case-based reasoning, genetic algorithms, and fuzzy logic. Neural networks are particularly useful for estimating posterior probabilities, which aids in creating classification rules and performing statistical analyses. Classical statistical classification methods, such as discriminant analysis, are based on Bayesian decision theory.

4. Results and Discussion

This segment summarizes the analysis of data collected from 1,756 top-ranked T20I batsmen. Bayesian probabilities for the top 30 batsmen were evaluated using Excel and SPSS Version 25.0. The results were visually represented through bar graphs, cross tabulations, pie charts, and descriptive statistics. The analysis also explored the relationship between height and weight among these players.

Table 4.1: Summary of the number of players, teams, and matches played in each continent

Continent	No; of players	No; of teams	Sum of matches played
Asia	12	06	631
Australia	05	02	341
Africa	04	01	154
Europe	07	03	524
North America	02	01	106

The table provides a summary of the distribution of players, teams, and matches across different continents. It details that out of the 30 top-ranked T20I batsmen, there are 12 players from Asia, participating in 6 teams and

having played a total of 631 matches. Australia has 5 players across 2 teams, with 341 matches played. Africa contributes 4 players from 1 team, who have participated in 154 matches. Europe has 7 players spread across 3 teams, totaling 524 matches. Finally, North America has 2 players in 1 team, with 106 matches played. Overall, these 30 players are involved in 13 teams and havecollectively played 1,756 matches.
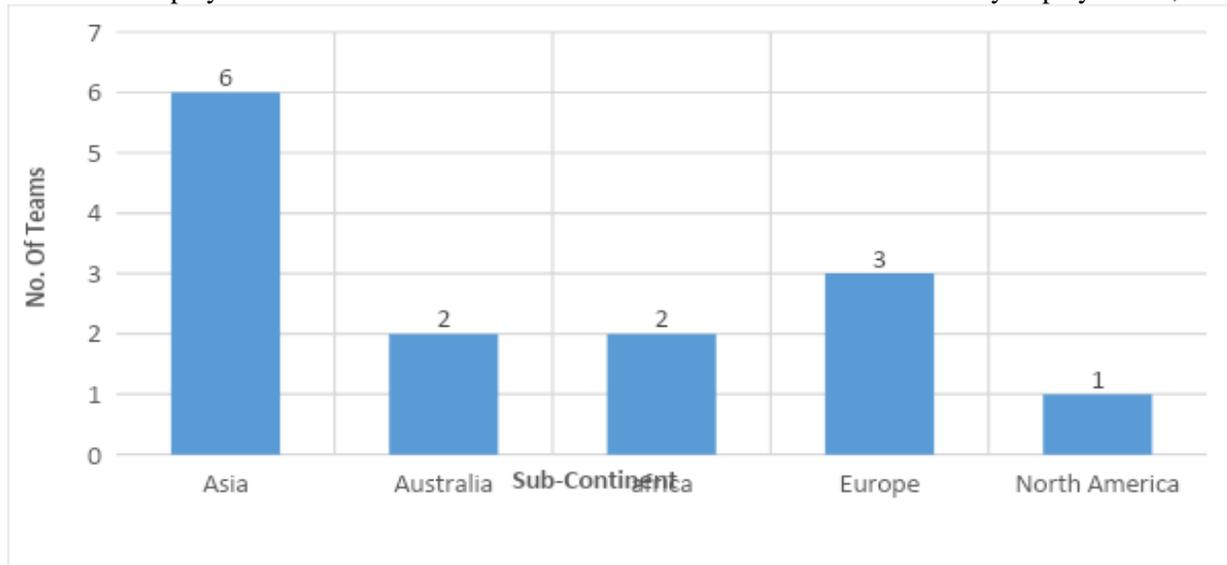
Figure 4.1: Bar Chart of the Number of teams in each Continent

The given bar graph illustrates the distribution of top 30 ICC ranked players across continents. Out of the total 13 teams considered, Asia is represented by six teams, Australia and Africa each have two teams, Europe has three teams, and North America has one team.
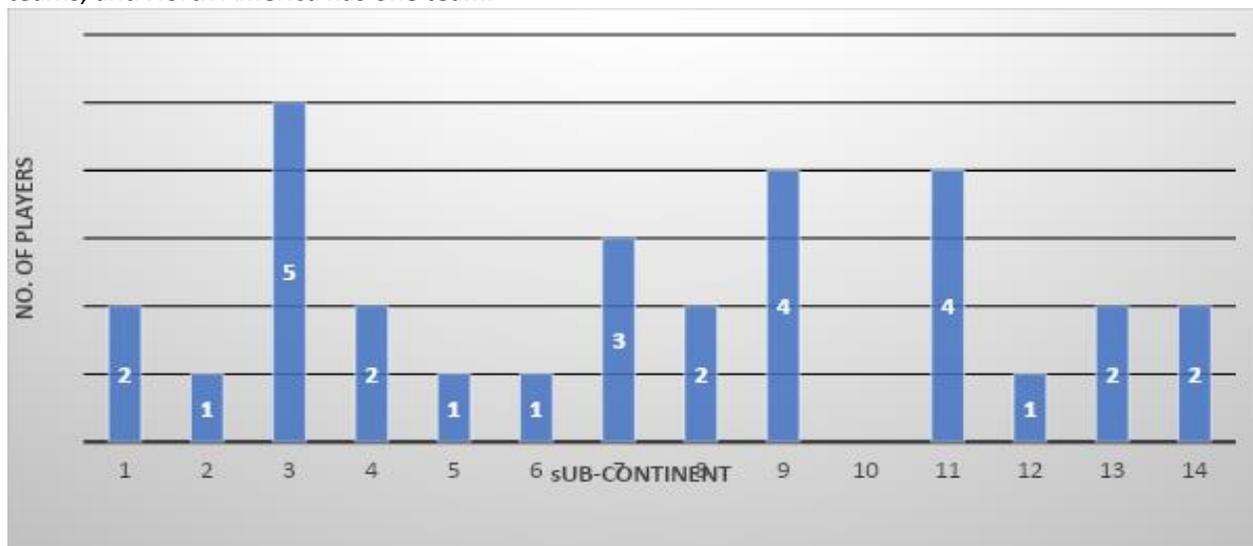
Figure 4.2: Bar Chart of the Number of players in each Country

The top 30 ICC batsmen are distributed across various continents as follows: Asia has Afghanistan and Pakistan with 2 players each, and Bangladesh, Sri Lanka, and the UAE with 1 each. Australia has 3 players, New Zealand has 2, and South Africa leads Africa with 4 players. Europe features England with 4 players, Scotland with 2, and Ireland with 1. The West Indies represents North America with 2 players.
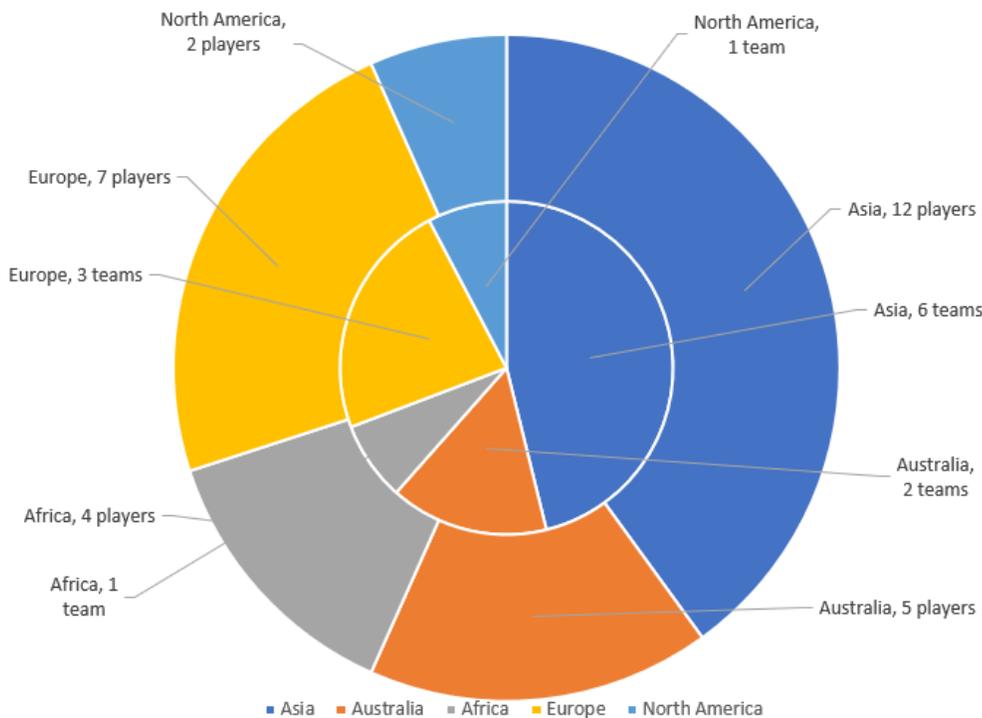
Figure 4.3: Pie chart of the No; of team and No; of players in each Continent
The provided pie chart illustrates a dual representation, wherein the inner section displays data based on team-wise distribution, while the outer section represents data based on the number of top ICC ranked players.

Table 4.2: Ranking distribution for each continent

Continent	Ranking			Total
	Rank ≤10	10<Rank≤20	20<Rank≤30	
Asia	04	05	03	12
Australia	03		02	05
Africa	02	01	01	04
Europe	01	03	03	07
North America		01	01	02
Total	10	10	10	30

The table presents a summary of player rankings across continents. From Asia, which has 12 players, 4 are in the top 10, 5 are ranked 11-20, and 3 are ranked 21-30. Australia, with 5 players, has 3 in the top 10 and 2 ranked 21-30, but none in the 11-20 range. Africa, with 4 players, includes 2 in the top 10, 1 in the 11-20 range, and 1 in the 21-30 range. Europe, with 7 players, has 1 in the top 10, 3 in the 11-20 range, and 3 in the 21-30 range. North America, with 2 players, has none in the top 10, 1 in the 11-20 range, and 1 in the 21-30 range. Overall, there are 10 players ranked in each of the top 10, 11-20, and 21-30 ranges, totaling 30 players. This data provides a clear view of the distribution of player rankings across different continents.
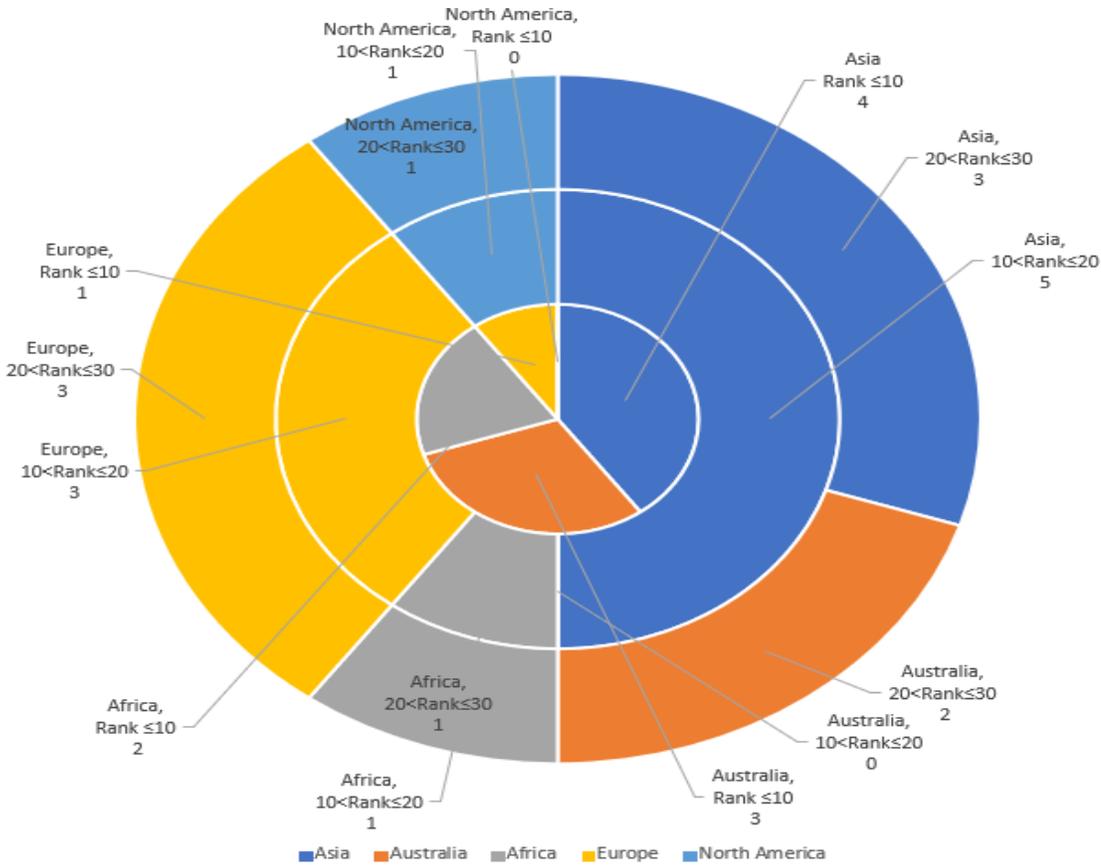
Figure 4.4: Pie chart of the Ranking distribution for each continent

The pie chart illustrates the distribution of player rankings across continents. In the top 10 rankings, Asia leads with 4 out of 12players, followed by Australia with 3 out of 5, and Africa with 2 out of 4. Europe has only 1 out of 7 players in the top 10, while North America has none. For rankings between the 10th and 20th positions, Asia has 5 players, while Africa and North America each have 1. Europe has 3 players in this range, and Australia has none. In the 20th to 30th rankings, Asia and Europe each have 3 players, Australia has 2, and Africa and North America each have 1 player.

Table 4.3: Number of matches against major countries of the players, getting the score ≥ 30 and achieving the man of the match award

Babar Azam (Pak)			
Country	**No. of Matches Against**	**Man, of the Match**	**No. of Matches > =30**
Bangladesh	5	1	1
New Zealand	7	1	4
South Africa	10	1	5
West Indies	17	1	7
Zimbabwe	6	1	4
Total	45	5	21
Aiden Kyle Markram (RSA)			
Sri Lanka	6	1	1
West Indies	4	1	2
Total	10	2	3
Muhammad Rizwan (Pak)			
New Zealand	5	1	2
South Africa	10	2	2

West Indies	7	1	4
Zimbabwe	9	1	3
Total	31	5	11
Dawid Johannes Malan (ENG)			
Australia	6	1	3
South Africa	6	3	4
Sri Lanka	4	1	1
Total	16	4	8
Devon Philip Conway (NZ)			
Australia	5	1	3
Bangladesh	3	1	1
Total	8	2	4
Aaron James Finch (AUS)			
Bangladesh	2	1	2
England	13	2	9
New Zealand	8	1	4
Pakistan	14	1	6
West Indies	8	1	3
Zimbabwe	2	1	1
Total	47	7	25
Hendrik Erasmus van der Dussen (RSA)			
England	7	1	5
Sri Lanka	5	1	2
Total	12	2	7
Martin James Guptill (NZ)			
Australia	13	2	6
Pakistan	19	2	7
South Africa	10	3	5
Zimbabwe	5	2	4
Total	47	9	22
Kannaur Lokesh Rahul (IND)			
New Zealand	8	1	5
West Indies	9	1	3
Total	17	2	8
Quinton de Kock (RSA)			
Australia	13	1	6
Pakistan	4	1	3
Sri Lanka	11	1	3
Total	28	3	12

Number of "Man of the Match" Awards and Scores ≥ 30 Runs by Player

Man of the Match Awards:

```
20 |████████████████████████████████████████  Virat Kohli (17)
15 |                                   Joseph Buttler (7)
10 |████████████████████         Rohit Sharma (9)
 5 |██████    Evin Lewis (6)   Nicholas Pooran (1)
 0 +-------------------------------------------
   Rohit Sharma  Nicholas Pooran   Virat Kohli   Joseph Buttler   Evin Lewis
```

Scores ≥ 30 Runs:
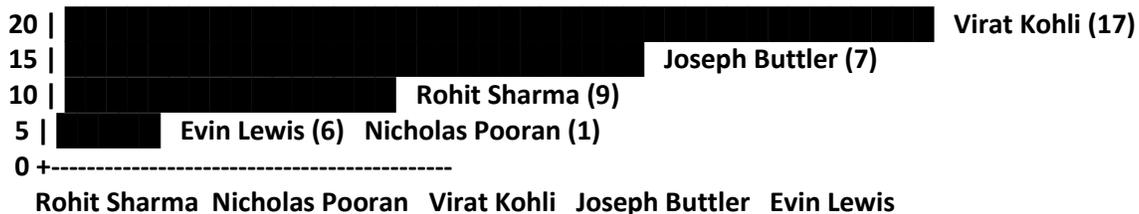
```
35 | ████████████████████████████████ Virat Kohli (32)
30 | ████████████████████████ Rohit Sharma (31)
25 | ████████████████ Joseph Buttler (21)
20 | ██████████ Evin Lewis (12)
15 | ████ Nicholas Pooran (1)
 0 +-------------------------------------------
    Rohit Sharma  Nicholas Pooran  Virat Kohli  Joseph Buttler  Evin Lewis
```
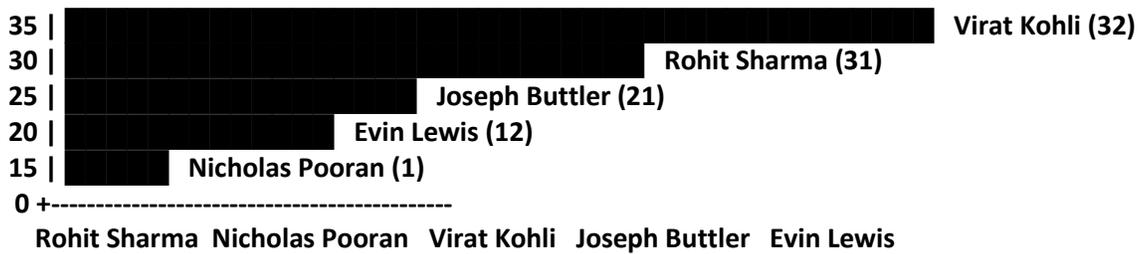
The table outlines the performances of various cricketers against top teams. Babar Azam from Pakistan played 45 matches, securing 5 'Man of the Match' awards and scoring 30 or more in 21 instances. Aiden Kyle Markram of South Africa participated in 10 matches, earning 2 'Man of the Match' accolades and scoring 30 or more 3 times. Muhammad Rizwan, also from Pakistan, featured in 31 matches, achieving 3 'Man of the Match' awards and scoring 30 or higher 4 times. Dawid Johannes Malan from England played 16 matches, winning 4 'Man of the Match' awards and scoring 30 or more 7 times. Devon Philip Conway from New Zealand took part in 8 matches, receiving 2 'Man of the Match' titles and scoring 30 or more 4 times. Players such as Pathum Nissanka Silva (Sri Lanka), Muhammad Waseem (UAE), Hazratullah Zazai (Afghanistan), Henry George Munsey (Scotland), Reeza Hendricks (South Africa), and Mohammad Naim Sheikh (Bangladesh) did not achieve significant performances and have been excluded from further competition.

Table 4.4: Bayes Probability of the top 30 batsmen along with their average height and average weight

S. No	Player Name	Actual height & Avg Height	Actual weight & Avg Weight	AFG	AUS	BAN	ENG	IND	NZ	PAK	RSA	SL	WI	ZIM
1.	Babar Azam (Pak)	5.11<5.82	67<75.06	0	0	**0.394**	0	0	0.138	0	0.157	0	0.192	0.118
2.	Aiden Kyle Markram (RSA)	6.1>5.82	75<75.06	0	0	0	0	0	0	0	0	**0.75**	0.25	0
3.	Muhammad Rizwan (Pak)	5.8=5.82	65<75.06	0	0	0	0	0	0.204	0	**0.326**	0	0.143	0.245
4.	Dawid Johannes Malan (ENG)	6.0>5.82	72<75.06		0.171	0		0	0.1	0	**0.385**	0.343	0	
5.	Devon Philip Conway (NZ)	5.11<5.82	77>75.06	0	0.357	**0.64**	0	0		0			0	

No.		Name													
6.		Aaron James Finch (AUS)	5.9>5.82	90>75.06			0.077	**0.224**	0	0.155	0.181	0	0	0.206	0.155
7.		Hendrik Erasmus van der Dussen (RSA)	6.2>5.82	75<75.06	0	0		0.36	0		0		**0.64**	0	0
8.	8	Martin James Guptill (NZ)	6.2>5.82	90>75.06	0	0.214	0	0	0	0.268	**0.296**	0	0		0.123
9.	10	Kannaur Lokesh Rahul (IND)	6.0>5.82	75<75.06	0	0	0	0		0.286	0		0		**0.536**
10.		Quinton de Kock (RSA)	5.7<5.82	66<75.06	0	0.302	0	0	0	0	0.186		**0.512**		
11.		Rohit Gurunath Sharma (IND)	5.9>5.82	72<75.06	0	0	0.192	0.096		0.074	0	**0.303**	0.209	0.089	0
12.		Nicholas Pooran (WI)	5.8<5.82	75<75.06			0	**1**	0	0	0	0	0	0	
13.		Virat Kohli (IND)	5.9>5.82	70<75.06	0.083	0.264	0	0		0	0.175	**0.28**	0.117	0.083	0
14.		Joseph Charles Buttler (ENG)	6.0>5.82	73<75.06	0	0.148	0		0.209	0	0		0.149	**0.493**	0
15.		Jason Jonathan Roy (ENG)	6.0>5.82	75<75.06	0	0	0.117		0	**0.47**	0.412	0	0	0	
16.		Shreyas Santosh Iyer (IND)	5.11<5.82	65<75.06			0	0	0	**0.504**		0	0.495	0	
17.		Richard Douglas Berrington (Scotland	5.10<5.82	80>75.06	0		**0.444**		0	0	0				0

No.	Player	Height	Weight											
	)													
18.	Paul Robert Stirling (IRE)	5.9>5.82	78>75.06	0.113	0	0	0	0		0	0	0	**0.279**	0.127
19.	Glenn James Maxwell (AUS)	6.0>5.82	73<75.06		0		0.278	0.257	0	0	0		**0.463**	0	0
20.	Mitchell Ross Marsh (AUS)	6.4>5.82	90>75.06		0		0.298	0	**0.522**	0	0	0	0.179	
21.	Eoin Joseph Gerard Morgan (ENG)	5.9>5.82	65<75.06	0	0	0		0.203	0.072	0.323	**0.401**	0	0	
22.	Suryakumar Ashok Yadav (IND)	5.11<5.82	75<75.06	0			0.25		**0.5**	0		0	0.25	
23.	Evin Lewis (WI)	6.2>5.82	85>75.06	0	0.108	**0.324**	0	0.243		0.216	0.108	0		
24.	Mohammad Nabi (AFG)	5.9>5.82	82>75.06		0.15	0	0	0	0	0	0	0	**0.30**

The table presents data on the top 30 ICC-ranked batsmen, detailing their average height and weight, along with their performance probabilities against various teams. The average height of these players is 5.82 feet, and their average weight is 75.06 kg.

4.1 Babar Azam (Pakistan) shows the highest Bayes probability of 0.394 against Bangladesh, having secured the 'Man of the Match' title in 5 matches. His probabilities are lower against New Zealand (0.138) and Zimbabwe (0.118), and he also has zero probabilities against Afghanistan, Australia, India, England, and Sri Lanka, indicating less impressive performances against these teams.

4.2 Aiden Kyle Markram (South Africa) has a notable Bayes probability of 0.75 against Sri Lanka, where he earned the 'Man of the Match' title in one out of six matches. His probability against the West Indies is 0.25, with one 'Man of the Match' award from two high-scoring innings. He has zero probabilities against other top teams, reflecting that his performances did not meet the criteria for 'Man of the Match' against these opponents.

4.3 Nicholas Pooran (West Indies) participated in 57 matches against associated teams and scored over 30 runs frequently but was awarded 'Man of the Match' only once, with a perfect Bayes probability of 1 against Bangladesh. His performance did not meet the criteria for the award in matches against other associated teams.

4.4 Richard Douglas Berrington (Scotland) played 74 matches against associated teams, achieving a Bayes probability of 0.444 against Bangladesh, where he won the 'Man of the Match' award. Despite this, he did not secure the title in matches against other associated teams.

Players such as Pathum Nissanka Silva (Sri Lanka), Hazratullah Zazai (Afghanistan), Henry George Munsey (Scotland), Reeza Hendricks (South Africa), and Mohammad Naim Sheikh (Bangladesh) did not achieve a Bayes probability greater than zero against top-ranked ICC teams, resulting in their exclusion from further competition. Muhammad Waseem (UAE) is omitted from discussions due to the absence of matches against top 10 ICC teams but has shown notable performances against other teams.

Table 4.5: Maximum Probabilities of top 30 Batsmen of getting the score ≥ 30 and achieving the man of the match award

S. No	Player Name	country	Bayes Probability
1.	Babar Azam (Pak)	Bangladesh	0.394
2.	Aiden Kyle Markram (RSA)	Sri Lanka	0.75
3.	Muhammad Rizwan (Pak)	South Africa	0.326
4.	Dawid Johannes Malan (ENG)	South Africa	0.385
5.	Devon Philip Conway (NZ)	Bangladesh	0.64
6.	Aaron James Finch (AUS)	England	0.224
7.	Hendrik Erasmus van der Dussen (RSA)	Sri Lanka	0.64
8.	Martin James Guptill (NZ)	South Africa	0.296
9.	Kannaur Lokesh Rahul (IND)	West Indies	0.536
10.	Quinton de Kock (RSA)	Sri Lanka	0.512
11.	Rohit Gurunath Sharma (IND)	South Africa	0.303
12.	Nicholas Pooran (WI)	Bangladesh	1
13.	Virat Kohli (IND)	South Africa	0.28
14.	Joseph Charles Buttler (ENG)	Sri Lanka	0.493
15.	Jason Jonathan Roy (ENG)	New Zealand	0.47
16.	Shreyas Santosh Iyer (IND)	New Zealand	0.504
17.	Richard Douglas Berrington (Scotland)	Bangladesh	0.444
18.	Paul Robert Stirling (IRE)	West Indies	0.279
19.	Glenn James Maxwell (AUS)	Sri Lanka	0.463
20.	Mitchell Ross Marsh (AUS)	New Zealand	0.522
21.	Eoin Joseph Gerard Morgan (ENG)	South Africa	0.401
22.	Suryakumar Ashok Yadav (IND)	New Zealand	0.5
23.	Evin Lewis (WI)	Bangladesh	0.324
24.	Mohammad Nabi (AFG)	Zimbabwe	0.30

The table showcases the highest Bayes probabilities among top-ranked T20I batsmen. Aiden Kyle Markram of South Africa stands out with a Bayes probability of 0.75 against Sri Lanka, achieved in six matches where he scored over 30 runs and won the 'Man of the Match' title. Devon Philip Conway from New Zealand has a notable Bayes probability of 0.64 against Bangladesh, with three matches where he scored over 30 runs and earned 'Man of the Match' recognition. Hendrik Erasmus van der Dussen, also from South Africa, has a Bayes probability of 0.64 against Sri Lanka, achieving 'Man of the Match' in five matches by scoring over 30 runs in both innings. Nicholas Pooran of the West Indies participated in 57 matches against associated teams, frequently scoring over 30 runs but secured the 'Man of the Match' award only once against Bangladesh, where he had a perfect Bayes probability of 1. Despite this exceptional performance, he did not achieve the same results against other associated teams. Aaron James Finch from Australia has a relatively low Bayes probability of 0.22 against England, despite scoring over 30 runs in 9 of 13 innings and receiving the 'Man of the Match' title twice. His performance against other top-ranked teams is less impressive.

Table 4.6: Height and Weight Crosstabulation

Height record	Weight record	Total

	Less than Average	Greater than Average	
Less than Average	7	2	9
Greater than Average	12	9	21
Total	19	11	30

The chi-square test examined the relationship between height and weight among T20I batsmen, yielding a chi-square value of 1.155 with 1 degree of freedom and a P-value of 0.282. Since the P-value exceeds the significance level of 0.05, there is no statistically significant association between height and weight. According to the data, among the 30 players, 7 have both below-average height and weight, 2 have below-average height but above-average weight, 12 have above-average height but below-average weight, and 9 have both above-average height and weight.
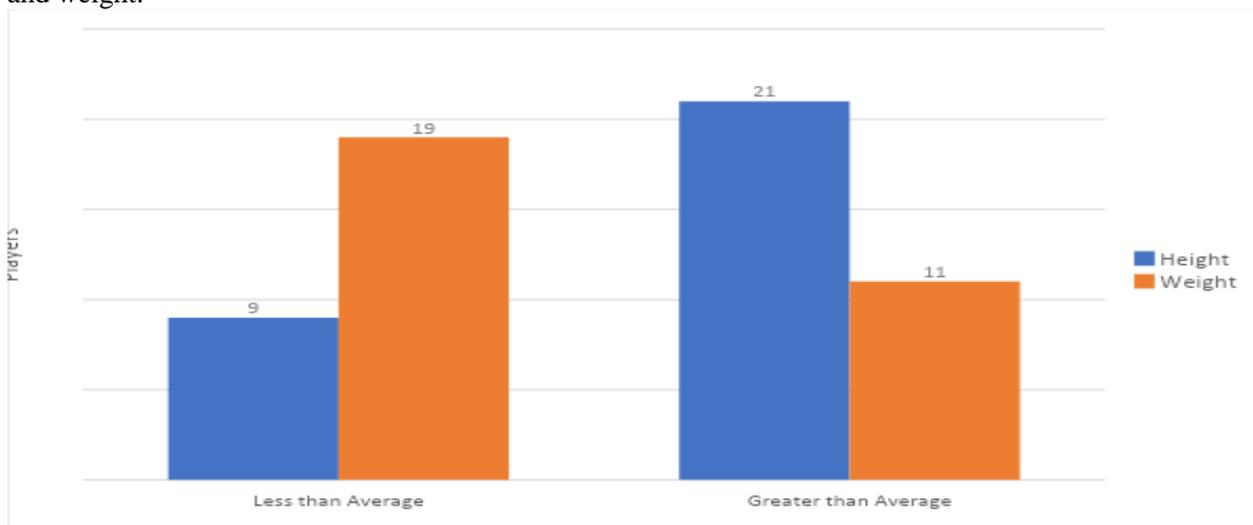

Figure 4.5: Distribution of the players along with height and weight

The provided bar graph displays how the heights and weights of the top 30 ICC ranked players are distributed. Among these players, 9 have heights below the average, while 21 players have heights above the average. Additionally, 19 players weight less than the average, while 11 players weight more than the average. The classification tree illustrates the average Bayes probabilities and physical statistics for the top 30 batsmen by continent. Asian batsmen have an average Bayes probability of 0.0735, with an average height of 5 feet 6 inches and a weight of 73.83 kg. Australian players show a higher average Bayes probability of 0.096, with an average height of 5 feet 9 inches and a weight of 84 kg. In Africa, the average Bayes probability is 0.085, with players averaging 5 feet 9 inches in height and 72 kg in weight. Similar average Bayes probabilities and physical characteristics are observed in European and North American players.

5. Conclusion
The analysis of top-ranked T20I batsmen in the ICC rankings provided several important insights into player performance dynamics. It was observed that Babar Azam demonstrated a high Bayes probability against Bangladesh, but his performance against other top-ranked teams was comparatively lower, illustrating how player challenges can vary based on opponents. Similarly, Aiden Kyle Markram's notable performances against Sri Lanka emphasized the influence of specific opponents on player success. The study also highlighted that players with Bayes probabilities below zero against top teams were excluded from further analysis, stressing the necessity of consistent excellence to achieve recognition such as the "Man of the Match" title. Nicholas Pooran's perfect Bayes probability against Bangladesh, coupled with his limited number of "Man of the Match" awards, underscored the limitation of relying on single matches to assess overall player prowess. Additionally, the analysis revealed no statistically significant correlation between height and weight and player performance, challenging any assumptions of a direct link between physical attributes and success. Finally, the classification tree analysis illustrated that

player characteristics and performance probabilities differ across continents, reflecting the significant regional and contextual variations in player profiles and achievements.

Reference

Ashraf, J., Bhatti, S., & Talpur, S. Predicting the Best Team Players of Pakistan Super League   using   Machine Learning Algorithms. *International Journal of Computer Applications*,   **(2021),** *975*, 8887.

Asif, M., & McHale, I. G. A generalized non-linear forecasting model for limited overs   international cricket. *International Journal of Forecasting*, **(2019),** *35*(2), 634-640

Albert, J., & Koning, R. H. (Eds.). *Statistical thinking in sports*. CRC Press **(2007),**

Cochran, W. G. Some methods for strengthening the common χ 2 tests. *Biometrics*,   **(1954),** *10*(4), 417-451.

Duda, R. O., & Hart, P. E. *Pattern classification and scene analysis,* **(1973),** (Vol. 3, pp. 731-   739). New York: Wiley.

Gupta, A. India and the IPL: Cricket's globalized empire. *The Round Table*, **(2009),** *98*(401),   201-211.

Kapadiya, C., Shah, A., Adhvaryu, K., & Barot, P. Intelligent cricket team selection by   predicting   individual players' performance using efficient machine learning   technique. *Int. J. Eng. Adv. Technol*, **(2020),** *9*(3), 3406-3409.

Kampakis, S., & Thomas, W. Using machine learning to predict the outcome of English county   twenty   over cricket matches. *arXiv preprint arXiv:* **(2015),** *1511.05837*.

Koulis, T., Muthukumarana, S., & Briercliffe, C. D. A Bayesian stochastic model for batting   performance evaluation in one-day cricket. *Journal of Quantitative Analysis in Sports*,   **(2014),** *10*(1), 1-13.

Kaluarachchi, A., & Aparna, S. V. Cric AI: A classification-based tool to predict the outcome in   ODI   cricket. *In 2010 Fifth International Conference on Information and Automation for   Sustainability 17-19 Dec.* (**2010),** (IEEE): Colombo, Sri Lanka, 250–255. doi:10.1109/ICIAFS.2010.5715668

Kimber, A. C., & Hansford, A. R. A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **(1993),** *156*(3), 443-455.

Lemmer, H. H. The single match approach to strike rate adjustments in batting performance   measures   in cricket. *Journal of sports science & medicine*, **(2011),** *10*(4), 630.

Lewis, A. J. Towards fairer measures of player performance in one-day cricket. *Journal of the   Operational Research Society*, **(2005),** *56*, 804-815.

Lemmer, H. H. A measure for the batting performance of cricket players. *South African Journal   for   Research   in Sport, Physical Education and Recreation*, **(2004),** *26*(1), 55-64.

Lewis, D., & Burke, C. J. The use and misuse of the chi-square test. *Psychological bulletin*,   **(1949),** *46*(6), 433.

Maddox, J. T., Sides, R., & Harvill, J. L. Bayesian estimation of in-game home team win probability   for   college basketball. *Journal of Quantitative Analysis in Sports*,   **(2022),** *18*(3), 201-213.

Maryam BiBi, Bayesian Analysis of the Flexible Lomax Distribution under a Uniform and   Jeffery   Prior. M.phil. Thesis, University of Peshawar, Peshawar, KP, (**2022**).

Manage, A., Scariano, S., & Hallum, C. Performance analysis of t20-world cup cricket 2012. *Sri   Lankan   Journal of Applied Statistics*, **(2013),** *14*(1).

Mail, D. Meet the man who invented Twenty20 cricket—The man missing out on millions,   **(2008).**

Nawaz, T., & Aleem, M. Analysis of Batting Performance in Cricket using Individual and   Moving   Range (MR) Control Charts, **(2012).**

Premkumar, P., Chakrabarty, J. B., & Chowdhury, S. Key performance indicators for factor   score   based ranking in One Day International cricket. *IIMB Management Review*,   **(2020),** *32*(1), 85-95.

Purkayastha, V. *Predicting outcome of cricket matches using classification learners* (Doctoral   dissertation, Dublin, National College of Ireland), **(2017).**

Perera, H., Davis, J., & Swartz, T. B. Optimal lineups in Twenty20 cricket. *Journal of Statistical Computation and Simulation*, **(2016),** *86*(14), 2888-2900.

Preston, I., & Thomas, J. Batting strategy in limited overs cricket. *Journal of the Royal   Statistical Society Series D: The Statistician*, **(2000),** *49*(1), 95-106.

Phyu, T. N. Survey of classification techniques in data mining. In *Proceedings of the   international multiconference of engineers and computer scientists* **(2009),** (Vol. 1, No. 5,   pp. 727-731). Citeseer.

Quazi, M., Clifford, J., & Datta, P. Predicting Cricket Outcomes using Bayesian Priors.   **(2022),** *arXiv   preprint*

arXiv:2203.10706.

Richard, M. D., & Lippmann, R. P. Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural computation*, **(1991),** *3*(4), 461-483.

Santos-Fernandez, E., Wu, P., & Mengersen, K. L. Bayesian statistics meets sports: a comprehensive review. *Journal of Quantitative Analysis in Sports*, **(2019),** *15*(4), 289- 312.

Shah, S., Hazarika, P. J., & Hazarika, J. A study on performance of cricket players using factor analysis approach. *International Journal of Advanced Research in Computer Science* **(2017),** *8*(3), 656-660.

Shah, P. New performance measure in Cricket. *ISOR Journal of Sports and Physical Education*, **(2017),** *4*(3), 28-30.

Saikia, H., Bhattacharjee, D., & Radhakrishnan, U. K. A new model for player selection in cricket. *International Journal of Performance Analysis in Sport*, **(2016),** *16*(1), 373-388

Sharp, G. D., Brettenny, W. J., Gonsalves, J. W., Lourens, M., & Stretch, R. A. Integer optimisation for the selection of a Twenty20 cricket team. *Journal of the Operational Research Society*, **(2011),** *62*(9), 1688-1694.

Scarf, P., Shi, X., & Akhtar, S. On the distribution of runs scored and batting strategy in test cricket. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **(2011),** *174*(2), 471-497.

Saikia, H., & Bhattacharjee, D. On the Effect of Home Team Advantage and Winning the Toss in the Outcome in T20 International Cricket Matches. *Assam University Journal of Science and Technology*, **(2010),** *6*(2), 88-93.

Saikia, H., & Bhattacharjee, D. A Bayesian classification model for predicting the performance of all-rounders in the Indian premier league, **(2010),** *Available at SSRN 1622060.*

Saikia, H., & Bhattacharjee, D. On the effect of home team advantage and winning the toss in the outcome in T20 international cricket matches. *Assam University Journal of Science and Technology*, **(2010),** *6*(2), 88-93.

Snedecor, G. W., & Cochran, W. G. Statistical Methods, eight editions. *Iowa state University press, Ames, Iowa*, **(1989),** *1191*(2).

Taliep, M. S., Prim, S. K., & Gray, J. Upper body muscle strength and batting performance in cricket batsmen. *The Journal of Strength & Conditioning Research*, **(2010),** *24*(12), 3484-3487.

Van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G. & Yau, C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, **(2021),** *1*(1), 1.

ICC Men's T20 World Cup [Online] 2021/22. Available from ESPN cricinfo**. https://www.espncricinfo.com/series/icc-men-s-t20-world-cup-2021-22-1267897 (accessed 9/02/2023)** Men's T-20I player rankings [Online]. Available from ICC CRICKET. https://www.icc-cricket.com/homepage **(accessed 9/02/2023)** . ICC player ranking: Men's ranking: Top 100 batsmen [Online]. Available from Cricbuzz. https://www.cricbuzz.com/cricket-stats/icc-rankings/men/batting (accessed 9/02/2023)