



Exploring the Strength of Batting Partnerships in T-20 Cricket

Muhammad Irshad^a, Qamruz Zaman^{*b}, Mujeeb Hussain^c, Sofia^d, Jawad Ullah

^{a,b}Department of Statistics, University of Peshawar, Pakistan. ^cGovernment College Peshawar, Pakistan.

^dCollege of Home Economics, University of Peshawar, Pakistan

*Email: cricsportsresearchgroup@gmail.com

Abstract: This study examines the batting partnerships of three major cricket teams—Pakistan, India, and Australia—using K-Means clustering to categorize player pairs based on their performance. Bar charts and frequency tables highlight the top partnerships for each team, with key pairings such as Mohammad Rizwan and Babar Azam for Pakistan, Shikhar Dhawan and Rohit Sharma for India, and David Warner and Aaron Finch for Australia. K-Means clustering further categorizes partnerships into performance clusters, providing insights into each team's strengths and weaknesses. The analysis reveals that India has the strongest partnerships, with more consistent high-performing clusters, while Pakistan and Australia have a wider range of both strong and weak partnerships

Keywords: T-20 cricket, Batting partnership, K-Means Clustering, Bar Charts, Frequency table

1. Introduction

Cricket is a sport played between two teams, each comprising eleven players, on a field. One team bats to score runs, while the other team bowls and fields to prevent runs and take wickets. The game has different formats: T-20 cricket with 20 overs per side, one day cricket with 50 overs per side, and Test cricket played over five days with a maximum of 90 overs per day. Each format offers its own challenges and strategies for both batting and bowling teams (Dey, 2017). In England, during the late sixteenth century, various ball games gained popularity, including cricket, stoolball, baseball, stickball, trap ball, and others (Underdown 2000). The international cricket match took place between the USA and Canada from September 24–26, 1844, at Bloomingdale Park in Manhattan. Nevertheless, there are contentions surrounding the classification of the match as international, as the players were primarily selected from two clubs representing their respective nations (Williamson, 2007). Twenty20 cricket was introduced in 2003 in England through a domestic tournament called the Twenty20 Cup. The first international Twenty20 cricket match was played on June 13, 2005, between England and Australia at the Rose Bowl in Hampshire (Saikia H. B., 2012).

Cricket is special game because it's one of the few sports where statisticians have been involved since the beginning. While in other sports like soccer or tennis, we now see performance stats on TV, cricket has always had them. That's why many statisticians and data scientists use different tools to understand cricket better and solve its challenges such as adjusting targets in matches affected by rain, evaluating team and player performances, picking the right team, predicting match results, planning match schedules, and making decisions in many other aspects of the game (Saikia H. B., 2019).

In this study, descriptive statistics and bar charts were employed to identify the strongest player pairs based on their partnership performance. Descriptive statistics provided a comprehensive summary of the data, while bar charts

visually represented the strength of these partnerships, facilitating a clear comparison. To further analyze and categorize the player pairs, K-means clustering was utilized. This method grouped similar player pairs according to their partnership scores, enabling the identification of clusters with comparable performance metrics.

2 Literature Review

Bhattacharjee et al. (2018) highlight the significance of batting and bowling as core skills in cricket. Unlike bowling, which involves a solo performance, batting occurs in partnerships of two batsmen. When a batsman is out, the partnership ends, and a new one begins with the incoming batsman. Traditionally, batting performance measures focus on individual statistics rather than the dynamics of these partnerships. This research introduces a novel metric designed to assess the performance of batting partnerships. This metric not only considers the runs scored by the partnership but also the context in which these runs are achieved. The study uses data from the 2016 Twenty20 World Cup held in India to quantify this model. By analyzing partnership scores across multiple matches for a team, coaches can identify partnerships that excel under pressure, aiding in strategic decisions such as optimizing the batting order and selecting the most effective opening partners for limited-overs cricket.

Valero et al. (2013) explore the role of synergies in batting partnerships and their impact on performance in cricket. The study compares the effectiveness of opening batsmen when paired with their "synergistic" partners versus alternative partners. For Test cricket, the analysis uses a bivariate performance statistic that combines strike rate and batting average, evaluated through a nonparametric binomial test. In one-day cricket, a univariate statistic based on runs scored relative to resources used is assessed using a single sample t-test. The findings of their statistical analyses suggest that the perceived benefits of synergies in opening partnerships may be more of a sporting myth than a substantiated advantage.

Kachoyan and West (2014) build on how an ensemble of batsmen approximates a memoryless distribution by scaling batting averages and considering mixed exponential distributions. The paper shows that a mixed exponential model can closely align with survival data for male Test cricketers, suggesting that an individual batsman's survival function might be viewed as varying around a "true" memoryless function for similarly skilled players. The paper then examines whether these principles can be applied to predict batting partnerships or team scores. While predicting team scores is challenging due to the rarity of identical team line-ups, modeling batting partnerships is more feasible with available data. The study derives expected performance of batting pairs using the memoryless property and compares these predictions with historical data, finding good agreement in the mean but significant variance in the actual performance, with both under- and over-performing pairs.

3 Methodology

3.1 Data

We collected data on batting partnerships for Pakistan, India, and Australia from the official websites ESPN Cricinfo and Howstat, covering the period from 2005 to 2023.

We utilize descriptive statistics and bar charts to identify the strongest player pairs based on their batting partnerships. By analyzing the frequency and performance metrics of various player combinations, we aimed to highlight the most effective partnerships for each team. This approach provided a clear visualization of how often specific pairs contribute to their team's success. Employing K-means clustering to group similar player pairs based on their partnership scores. This statistical method enabled us to categorize partnerships into distinct clusters according to their performance, offering insights into the relative strengths and weaknesses of different pairs. By combining these methodologies, the study effectively illustrated both the top-performing player combinations and the broader patterns of partnership effectiveness across the teams.

3.2 Bar chart

Bar charts are a powerful visual tool used to display and compare the frequency or magnitude of different categories. In data analysis, bar charts offer a clear and straightforward way to represent categorical data, allowing for easy comparison between various groups or variables. Each bar in the chart represents a category, with its length or height corresponding to the value or frequency of that category. This visual representation makes it simple to identify trends, patterns, and outliers within the data. Bar charts are particularly useful when dealing with discrete data points, as they provide an immediate visual comparison that aids in understanding the relative significance of each category in the dataset (Dadhich et al. 2021).

3.3 Frequency table

A frequency table is a fundamental statistical tool used to organize and summarize data by displaying the number of occurrences, or frequency, of each unique value or range of values within a dataset. It provides a clear and concise way to view the distribution of data, making it easier to identify patterns, trends, and outliers. The table typically consists of two columns: one for the distinct values or intervals and another for the corresponding frequencies. By condensing data into a frequency table, researchers can quickly assess the central tendencies and variability within the dataset, facilitating more effective analysis and interpretation (Gries, 2014).

3.4 K-means clustering

K-means clustering is an unsupervised machine learning technique used to partition a dataset into K distinct clusters based on feature similarities. The algorithm works by initializing K cluster centroids randomly, then iteratively assigning each data point to the nearest centroid and updating the centroids based on the mean of the assigned points. This process continues until the centroids stabilize or a predetermined number of iterations is reached. K-means aims to minimize the within-cluster variance while maximizing the between-cluster variance, effectively grouping similar data points together. Despite its simplicity and efficiency, K-means requires specifying the number of clusters in advance and can be sensitive to the initial placement of centroids, which may affect the final clustering outcome (Ahmad et al. 2007).

4 Analysis

Fig 1 shows, the most effective partnership for Pakistan is between Mohammad Rizwan and Babar Azam, with a frequency of 46, highlighting their exceptional synergy and consistency. Following closely is the pair of Babar Azam and Fakhar Zaman, who have a frequency of 34, indicating a strong and productive collaboration. Umar Akmal and Shahid Afridi also form a notable partnership with a frequency of 21, suggesting a successful dynamic when they bating together. Mohammad Hafeez and Umar Akmal, with a frequency of 19, show a respectable level of effectiveness, though slightly less prominent compared to the top partnerships. These pairings reflect key combinations that have contributed significantly to Pakistan's performance.

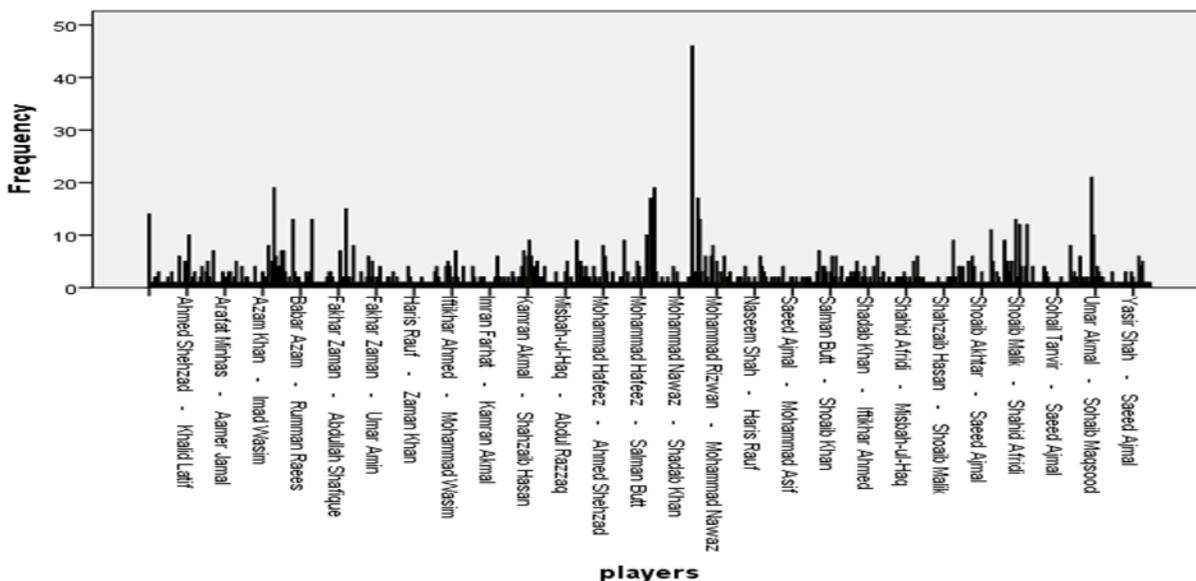


Fig 1: Bar chart of player partnership for team Pakistan

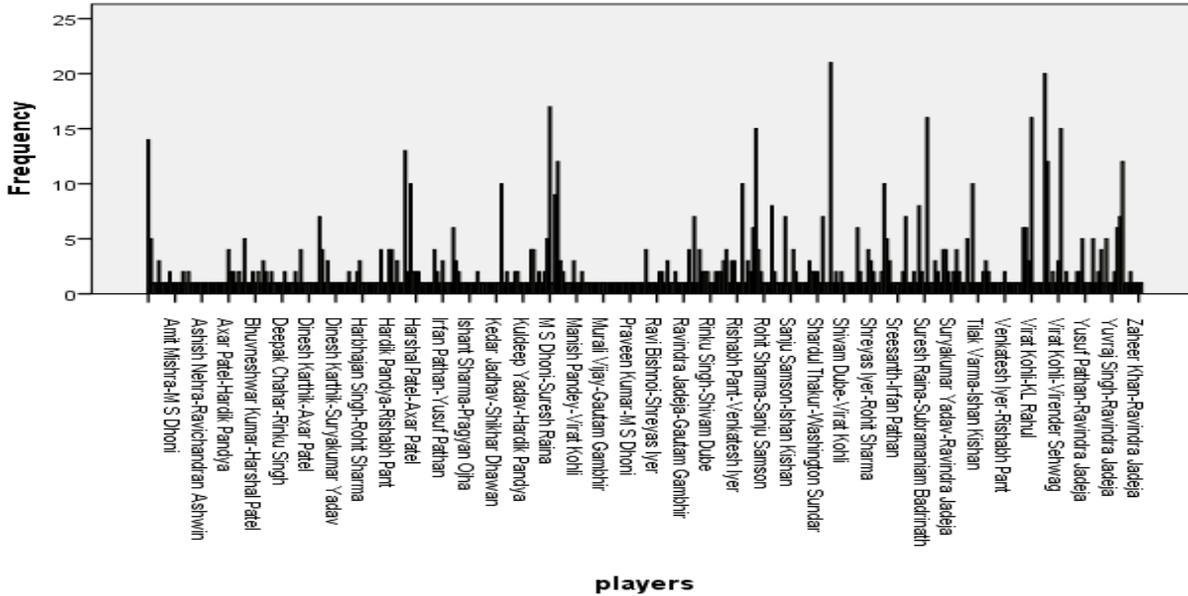


Fig 2: Bar chart of player partnership for team India

Fig 2 shows, the most successful partnership is between Shikhar Dhawan and Rohit Sharma, with a frequency of 21, underscoring their reliable and productive collaboration. Close behind is the pair of Virat Kohli and Rohit Sharma, with a frequency of 20, indicating a strong and consistent partnership. The partnership of M S Dhoni and Suresh Raina, with a frequency of 17, also demonstrates effective teamwork, though less frequent compared to the top pairs. These partnerships highlight the key combinations that have been instrumental in India's performance over time (also see table 1).

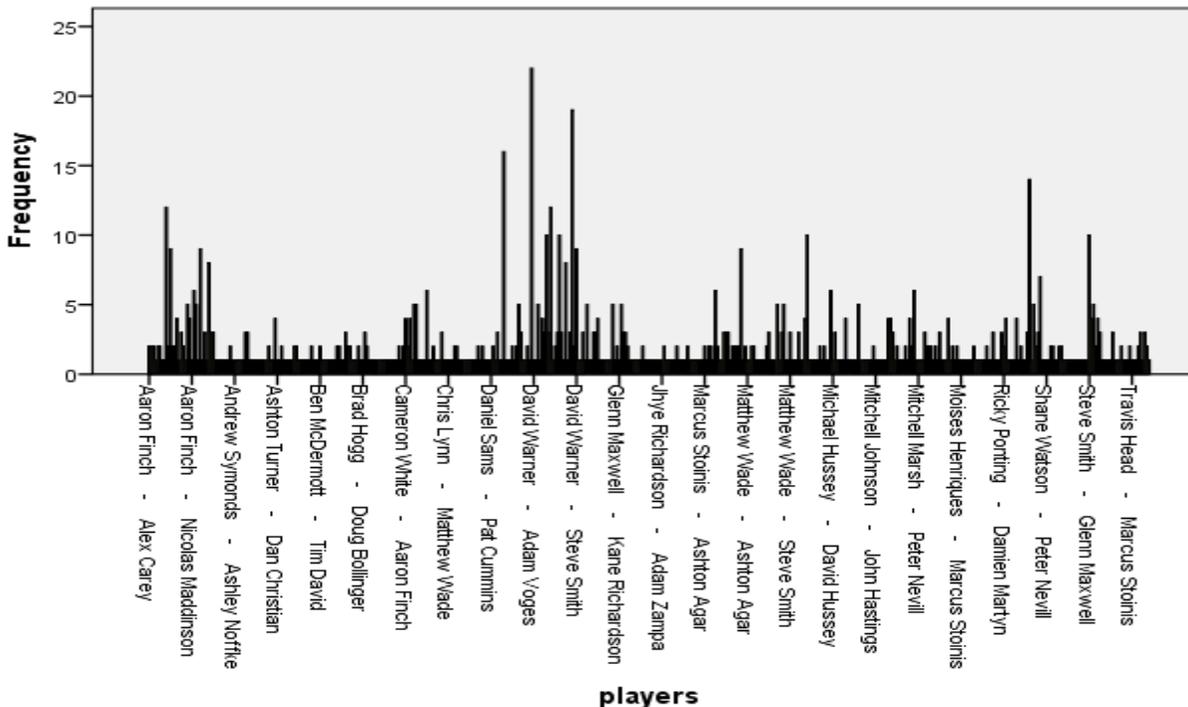


Fig 3: Bar chart of player partnership for team Australia

Fig 3 shows, the most prominent partnership is between David Warner and Aaron Finch, with a frequency of 22, showcasing their successful and consistent pairing. The duo of David Warner and Shane Watson follows with a

frequency of 19, indicating another highly effective combination. Additionally, David Hussey and Cameron White, with a frequency of 16, represent a notable partnership, albeit with slightly less frequency. The partnership of Shane Watson and David Warner, with a frequency of 14, still reflects a good level of effectiveness. These pairings illustrate the key collaborations that have significantly contributed to Australia's cricketing success (also see table 1).

Table 1: Frequency table for top 10 player pairs partnership

Pakistan player	Frequency	India player	Frequency	Australia player	Frequency
Mohammad Rizwan - Babar Azam	46	Shikhar Dhawan- Rohit Sharma	21	David Warner - Aaron Finch	22
Umar Akmal - Shahid Afridi	21	Virat Kohli-Rohit Sharma	20	David Warner - Shane Watson	19
Babar Azam - Fakhar Zaman	34	M S Dhoni-Suresh Raina	17	David Hussey - Cameron White	16
Mohammad Hafeez - Umar Akmal	19	Suresh Raina-Virat Kohli	16	Shane Watson - David Warner	14
Mohammad Hafeez - Shoaib Malik	17	Virat Kohli-KL Rahul	16	David Warner - Glenn Maxwell	12
Mohammad Rizwan - Fakhar Zaman	17	Rohit Sharma-KL Rahul	15	Aaron Finch - David Warner	12
Shoaib Malik - Umar Akmal	15	Virender Sehwag- Gautam Gambhir	15	David Warner - Michael Clarke	10
Babar Azam - Mohammad Hafeez	13	Hardik Pandya- Suryakumar Yadav	13	David Warner - David Hussey	10
Babar Azam - Shoaib Malik	13	M S Dhoni-Yuvraj Singh	12	Michael Clarke - David Hussey	10

Table 2: K-mean clustering of bating partnership for team Pakistan

Clusters	Number of Cases in each Cluster	Partnership
1	372.000	37.61
2	949.000	6.96
3	1.000	197.00
4	63.000	86.17

5	9.000	133.56
---	-------	--------

The above table shows that K-Means clustering of batting partnerships for Team Pakistan reveals five distinct clusters based on the performance of player pairs. Cluster 1, with 372 player pairs, has an average partnership of 37.61 runs, while Cluster 2, containing 949 pairs, shows weaker partnerships with an average of 6.96 runs. Cluster 3 represents a standout pair with 197 runs, making it an outlier. Cluster 4, consisting of 63 pairs, has a stronger average partnership of 86.17 runs, and Cluster 5, with only 9 pairs, shows the highest average partnership of 133.56 runs, representing the most successful pairs.

Table 3: K-mean clustering of bating partnership for team India

Clusters	Number of Cases in each Cluster	Partnership
1	18.000	119.00
2	63.000	80.00
3	2.000	165.00
4	250.000	40.00
5	529.000	5.73

The above table shows that K-Means clustering of batting partnerships for Team India. Player pairs into five distinct clusters based on their partnership performance. Cluster 1, with 18 pairs, has a high average of 119 runs, indicating strong partnerships. Cluster 2, consisting of 63 pairs, shows a good average partnership of 80 runs. Cluster 3 is a small cluster with only 2 pairs, but these partnerships are particularly successful, averaging 165 runs. Cluster 4, containing 250 pairs, has an average partnership of 40 runs, representing moderate performance. Lastly, Cluster 5, the largest with 529 pairs, has the weakest partnerships, averaging just 5.73 runs per pair.

Table 4: K-mean clustering of bating partnership for team Australia

Clusters	Number of Cases in each Cluster	Partnership
1	616.000	5.58
2	62.000	72.40
3	20.000	108.80
4	249.000	33.96
5	1.000	161.00

The above table shows K-Means clustering of batting partnerships for Team Australia divides the player pairs into five clusters based on their performance. Cluster 1, the largest with 616 pairs, has the weakest partnerships, averaging only 5.58 runs. Cluster 2, consisting of 62 pairs, shows a solid average partnership of 72.40 runs. Cluster 3 includes 20 pairs with a strong average of 108.80 runs, indicating successful partnerships. Cluster 4, with 249 pairs, averages 33.96 runs, representing moderate performance. Lastly, Cluster 5 is an outlier with only 1 pair achieving a remarkable 161 runs.

On the basis of partnerships, India emerges as the strongest team. It has fewer weak partnerships, and its high-performing clusters (Clusters 1 and 3) are both substantial and consistent. Pakistan has a wider range, including some very strong partnerships, but the large number of weak pairs (949 pairs with 6.96 runs) drags down its overall performance. Australia shows more weak partnerships compared to India and Pakistan but has a few high-performing pairs.

5 Conclusion

The comparative analysis of batting partnerships across various cricket teams reveals insightful patterns in terms of consistency and overall performance. India emerges as the best team in terms of consistency and partnership strength, with fewer weak pairs and more robust high-scoring clusters. Pakistan, while having standout partnerships, suffers from a large number of weak-performing pairs, which lowers its overall performance. Australia, similarly, has a high number of weak partnerships but includes a few strong pairs that contribute to its success. This study highlights the importance of strong and consistent partnerships in team performance and provides a statistical framework for assessing team effectiveness in cricket.

References

- Ahmad, A., & Dey, L. (2007). A k-mean clustering algorithm for mixed numeric and categorical data. *Data & Knowledge Engineering*, 63(2), 503-527.
- Bhattacharjee, D., Lemmer, H. H., Saikia, H., & Mukherjee, D. (2018). Measuring performance of batting partners in limited overs cricket. *South African Journal for Research in Sport, Physical Education and Recreation*, 40(3), 1-12.
- Dadhich, K., Daggubati, S. C., & Sreevalsan-Nair, J. (2021, February). BarChartAnalyzer: Digitizing Images of Bar Charts. In *IMPROVE* (pp. 17-28).
- Dey, P., Ganguly, M., & Roy, S. (2017). Network centrality based team formation: A case study on T-20 cricket. *Applied Computing and Informatics*, 161-168.
- Gries, S. T. (2014). Frequency tables: tests, effect sizes, and explorations. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 365-389.
- Kachoyan, B., & West, (2014) M. PREDICTING THE EXPECTED BATTING PARTNERSHIPS IN CRICKET.
- Saikia, H., Bhattacharjee, D., & Mukherjee, D. (2019). Cricket performance management: Mathematical formulation and analytics. Springer Singapore.
- Underdown, D. "Start of play: Cricket and culture in eighteenth-century England." 2000: 5-7.
- Valero, J., & Swartz, T. B. (2013). An investigation of synergy between batsmen in opening partnerships. *Sri Lankan Journal of Applied Statistics*, 13.
- Williamson, M. (2007). The oldest international contest of them all. *The Cricketer Annual*, 22-28.