Identifying Top Performers in the Durham Junior Cricket League (Under - 13) by Applying Unsupervised Learning

Mohammad Jamshaid Iqbal

Grey College

Project Supervisor: Dr. Peter G.P. Swift

A report presented for the degree of

Master of Data Science: Social Analytics

2023

Abstract

This study uses unsupervised learning techniques to identify the top-performing players in the Durham Junior Cricket League, specifically the Under-13s. Given how vast and diverse the data is in sports, it has become extremely critical for the coaching staff to identify top talents, talents that are not in top performing teams and the key performance metrics that distinguish them from others. Hence, this study aims to use two types of data: ball-by-ball data and separate scorecards for batting and bowling. The study will apply two techniques of dimensional reduction separately, one is Principal Component Analysis (PCA) and the other is autoencoders, followed by applying K-Means and Hierarchical Clustering to recognize patterns and cluster players according to their performances. The study will address three aspects: identifying key performance metrics, clustering players according to their performances using unsupervised learning, and highlighting the correlation among various performance metrics. The study's result shows different clusters of players according to their performance, in addition to which performance metric contributes more significantly to top-performing players. In addition, the correlation of different metrics is provided to give an understanding of player performance dynamics. The study provides a framework for Durham Cricket County to extract their top-performing players, not just for Under-13s but other levels.

Table of Contents

1. Introduction 5
2. Literature Review 6
   2.1. Use of Traditional Methods for Analysis 7
   2.2. Use of Supervised Learning for Analysis 9
   2.3. Use of Unsupervised Learning for Analysis 11
   2.4. Combination of Supervised and Unsupervised Learning 13
   2.5. Machine Learning Techniques Used in Other Sports 14
   2.6. Gaps in Current Literature 15
3. Methodology And Data 17
   3.1. Data Collection 17
   3.2. Data Cleaning and Preprocessing 20
      3.2.1. Data Cleaning 20
      3.2.2. Data Preprocessing 27
      3.2.3. Feature Scaling 35
   3.3. Theoretical Framework and Its Application 36
      3.3.1. Principal Component Analysis (PCA) 37
      3.3.2. Autoencoders 43
      3.3.3. KMeans++ Clustering 48
      3.3.4. Hierarchical Clustering 53
4. Results 55
5. Discussion 61
6. Conclusion 63

Bibliography 65

List of Figures, Tables, and Equations 68
   List of Figures in Main Content 68
   List of Figures in Appendix A 69
   List of Figures in Appendix B 69

List of Tables in Main Content 70
List of Tables in Appendix A 70
List of Tables in Appendix C 70
List of Equations 71
Appendix 73
Appendix A 73
Appendix B 81
Appendix C 103

1. Introduction

In recent years, the amount of data in sports has been growing at an exponential rate which has gained a lot attention towards using sports analysis to provide the ability to team management to assess, enhance and identify player performance, develop team strategies, and improve game understanding. This study aims to use the application of unsupervised learning techniques to identify top-performing players in the County Durham Junior Cricket League, specifically the under-13s. The study's main objective is to use advanced analytical methods to investigate hidden data patterns, providing a framework to identify talent in the Under-13 cricket clubs.

The Junior Cricket League at Durham, like other youth leagues in different sports, provides an important platform for the development of young talents and provides them with the opportunity to display their skill to the coaching staff and selectors who as a result can select these players to higher levels of competitions. Furthermore, it will also help cricket at grassroot level as identifying players at an incredibly young age would give more time to develop them further and hence providing senior cricket team with strong players. However, the traditional methods of assessing players' performance are subjective and they use basic statistical measures such as runs scored by a batter, strike rate, average, wickets taken by a bowler or economy rate. But not all skilful players will have the right balance of statistics above, some might be good finishers while others might be hard-hitting, some might be consistent while others might be able to hold onto a struggling team, some might have a high number of wickets while others are big wicket takers. In short, there will always be a subjective eye of a coach or selector when recommending a player. Hence, unsupervised learning techniques such as Principal Component Analysis (PCA), K-Means Clustering and Hierarchical Clustering provide a powerful alternative that uses the complete data of a player and enhances conventional assessment methods.

The study will use two primary data sources: detailed ball-by-ball data for each match and complete scorecards for batting and bowling. These datasets provide a complete foundation for analysis having every kind of performance metric such as runs scored, number of balls faced, whether a batter was out or not out at the end of the innings, wicket taken, strike rates, and economy rates.

Given how multidimensional the data is, the first techniques that will be applied are autoencoder and PCA separately, which will reduce the dimensionality of the data and identify the key performance metrics. This is a key step to be applied as it will make the data simple while also retaining the information of the data. After the application of autoencoder and PCA, the study employs K-Means and Hierarchical Clustering to categorize the players into distinct clusters. Both clustering methods will help identify the top-performing players in a differentiated cluster of their peers.

There are three main questions that the study will be answering:

1. What key performance metrics distinguish top-performing players in the Junior Cricket League?
2. Are unsupervised learning techniques such as K-Means and Hierarchical Clustering able to cluster top-performing players from their peers?
3. Are there correlations between different performance metrics, and what can be comprehended from these correlations?

The study's aim, through answering these questions, is to contribute to sports analytics by demonstrating how to utilize unsupervised learning techniques in evaluating junior cricket league players' performance. The study is set to provide a framework of valuable insights for the coaching staff to make decisions using comprehensive data and developing young cricket talent accordingly. Moreover, the methodology and analytical framework used in this study can be used in other cricket leagues too, highlighting the broader applicability of unsupervised learning in higher leagues.

2. Literature Review

In the last few decades analysis in sports has evolved drastically; this has happened due to vast amount of data with respect to player performance being available and with that advanced statistical methods and machine learning techniques being used to evaluate the data and

identifying talent. This section of the literature review will discuss various forms of methodologies and theoretical frameworks used to assess players' and teams' performance, particularly focusing on cricket. The literature review is divided into five sections:

1. Use of traditional methods for analysis
2. Use of Supervised Learning for analysis
3. Use of Unsupervised Learning for analysis
4. Combination of Supervised and Unsupervised Learning
5. Machine Learning techniques used in other sports.
6. Theoretical Framework for the study

By reviewing a wide range of literature, this section aims to provide a detailed understanding of the methodologies, tools, metrics, and frameworks used in cricket, further highlighting its strengths and limitations. An interesting fact in existing literatures are that most of the research has been done using supervised learning methods to analyse either players' or teams' performance and extraordinarily little literature has been developed in using unsupervised learning methods to analyse performances. Most of the unsupervised learning that has been done to analyse performances has been done in combination with supervised learning.

2.1. Use of Traditional Methods for Analysis

When it comes to analysing players in cricket, the coaching staff or selectors have traditionally relied on statistics such as the batting average and strike rates for batters and number of wickets taken, economy rate for bowlers. One cannot deny that these statistics are important, but they do not provide the complete picture of a player's performance, especially when it comes to formats such as T20 cricket as there are other aspects that come into player such as how fast a batter can score, how long can the batter stay in crease, is the bowler able to bowl more dot balls or give more extras. Researchers have worked on different statistical approaches using these statistics to better understand a player's performance.

One such approach is a popular one in Indian Premier League (IPL), that is the Most Valuable Player Index (MVPI). MVPI takes account of statistics such as runs scored, strike rates, wickets

taken, and economy rates to provide a complete picture of how a player has contributed to the team. Even though it takes an integrated approach, the drawback of MVPI is that the players who have played for a long time, their MVPI score will be more hence not considering the player's recent form.

Kimber and Hansford (1993) discussed the limitations of the traditional batting average and proposed a nonparametric method in which runs scored in used to assess batting performance. They mention how in the traditional batting average the treatment of 'not out' scores at the end of an innings is considered as completed innings which, as a result, skews the average. To provide a more accurate representation of a player's performance they suggested using a Product Limit (PL) estimator from survival analysis. Furthermore, their paper discusses the geometric distribution of cricket scores and the effect it has on statistical analysis. According to their research, the geometric model does not fit accurately with the empirical data, hence causing inconsistencies in the traditional batting average. They suggested using an alternative approach i.e., using nonparametric average which is not dependent on the geometric distribution. Their research performs detailed goodness-of-fit tests on individual player data to confirm the limitation of the geometric distribution assumption. They suggest using the combination of averages and distribution summaries for a much more detailed and accurate assessment of batting performances.

Moreover, when it comes to analysing player performances using traditional methods, using batting averages and bowling economy rates neglects the context of runs scored and wickets taken. Lewis (2005) suggested a methodology which uses Duckworth/Lewis (D/L), often used in matches affected by rain, to create a more balanced performance measure. The framework that Lewis used was net contribution, which is the comparison of actual runs to expected runs using D/L model, and resource averages, which calculates runs per unit of resource used. Using this framework on data from the 2003 Victoria Bitter Series final, Lewis highlighted that these new variables could provide a more balanced evaluation of player performance by keeping the match context in consideration.

The above research has shown how the traditional analysis of performances using strike rate, batting average, bowling average, economy rate etc. can be enhanced into a broader statistical analysis to measure player performance. Yet, these methods fail to take the current form of the players and the evolving nature of the game. While these methods are comprehensive, they fail to consider the game's change of pace with the introduction of T20 cricket, where a more dynamic approach is needed to analyse the performances. Given how fast the game is changing, one must consider the playing style, and strategies also change with time, hence there is a need to use advanced methods to analyse player performances. The above methods have given foundation for cricket analytics, but their limitation in handling the complex and dynamic modern data is why there is a need to adopt more advanced techniques. This is where authors who have used supervised and unsupervised learning to tackle these limitations and provide more comprehensive models to analyse player performances.

2.2. Use of Supervised Learning for Analysis

In the world of cricket analysis, there is a huge amount of study done on performance analysis or outcome of matches using supervised learning methods. There has been a great advancement in analysing player performance by using labelled datasets that, with the help of supervised learning methods, predict either the outcome of cricket matches or classification of performance metrics. Some methods the researchers used include regression analysis, classification algorithms, and neural networks.

Saikia and Bhattacharjee (2012) have used Artificial Neural Networks (ANN) to predict the performance of bowlers in the Indian Premier League (IPL). The authors showed how ANNs (Artificial Neural Networks) can be used to model and predict bowlers' performances, exemplifying that the model can demonstrate the non-linear relationship between different performance metrics which might get overlooked by the traditional methods. The ANN model uses the historical data to train and predict the future performances of the bowlers with great accuracy, hence assisting coaches and selectors for team selection and strategy.

Similarly, Lemmer (2004) in his paper discussed using a comprehensive way to measure batting performance of cricket players, this formula included average, strike rate, and consistency.

Lemmer wanted to address the failure of traditional methods to consider the variability of Batman's performance and how quickly they can score. Hence, Lemmer introduced a measure of performance which used batting average with strike rate and consistency coefficient (CC), which is further derived from the adjusted coefficient of variation (ACV) to adjust biases against low not-out scores and inflated standard deviations. Furthermore, to give importance to recent performances, Lemmer used exponentially weight average (EWA). Using the datasets of ODI and Test players, Lemmer incorporated bootstrap techniques to classify the player into ten classes utilizing their consistency and batting performance measure (BP). Hence, Lemmer's paper gives a much broader perspective to the performance measure that can be used to analyse batting performance hence giving a complete picture of a player's true ability.

While classification of players is one of the most used machine learning algorithms, Deep, Patvardhan, and Singh (2016) introduced a 'Deep Performance Index' (DPI) to rank IPL T20 cricketers using Recursive Feature Elimination (RFE). The authors, like previous ones, acknowledge the fact that the traditional methods do not provide a complete picture of a player hence they proposed the DPI with the usage of RFE algorithm which helps identify and weigh significant performance metrics for batters and bowlers. RFE provides a comprehensive analysis to evaluate a player's performance using multiple factors and simultaneously reduces the weight of metrics that are less critical. Hence, applying DPI and RFE together, as suggested by author, will help identify key performance metrics and a more accurate way to rank players according to their performance.

Tharoor and Dhanya (2022) used multiple data science concepts, Exploratory Data Analysis (EDA) and visualization and classification, to analyse the performance of Indian Cricket Team (ICT) in test cricket. The paper's primary focus was on the Test Cricket format where it covered matches from 2004 to early 2022 and applied preprocessing techniques to provide a solid analysis of the different performance metrics. They then applied two machine learning techniques (Random Forest and Support Vector Machine) and compared them. The Random Forest algorithm provided a 75% accuracy model, which was more than what Support Vector Machine provided. Then the study applied the Apriori algorithm which showed significant patterns, for example the impact of overs played on match outcomes and how critical a coin toss is. The analysis, again

using and enhancing traditional methods, provided a comprehensive view of how effective machine learning algorithms are for analysing a team and player's performance.

Bayesian approaches have also played a critical part in using machine learning techniques to handle the uncertainty and variability of performance metrics. For example, Saikia and Bhattacharjee (2011) used a Bayesian classification model to predict the performance of all-rounders in the IPL (Indian Premier League). The model used probabilistic relationships between different performance metrics, giving a broader view of a player's ability to perform. The strength of the model is that its prediction can be updated as new data is input into it, hence the model is adaptive and can make changes according to the dynamic nature of cricket.

The above models have provided approaches that have enhanced the robustness of predictions. For example, regression models can be used to predict how a batsman or bowler will perform if certain performance metrics act as an input feature, classification algorithms will categorize players according to different performance levels using previous data and random forest can use multiple models to improve prediction accuracy and reduce overfitting leading to a more stable and accurate performance assessment and outcome predictions.

In conclusion, supervised learning techniques, being one of the most used techniques to analyse performance, have assisted in advancing the performance analysis in cricket. They use labelled datasets and comprehensive algorithms to untangle complex relationships in the data and provide a much more accurate analysis of performance than the traditional methods. In the next section, we will discuss the role of unsupervised learning techniques in cricket which will help us understand how to uncover already hidden patterns and groupings in the data.

2.3. Use of Unsupervised Learning for Analysis

There is not a lot of literature done on unsupervised learning, most of the unsupervised learning that has been done by researchers has been done in combination with supervised learning which we will discuss in the next section. But in the little study that is available on unsupervised learning, the authors have used it to open new avenues of research in performance analysis through

identifying patterns and grouping unlabelled data. The two techniques used by the researchers are PCA and K-Means Clustering.

PCA is an unsupervised learning technique which reduces the dimension of a dataset with many variables to a smaller set but retains most of the information of the complete dataset. Manage and Scariano (2013) demonstrated how PCA can be applied in cricket to evaluate performances of players in the IPL 2012 season. The study's primary purpose is to introduce PCA in cricket and how one can utilize it to manage multicollinearity and reduce data dimensionality. The study used six performance metrics for batting and four for bowling and applied PCA on it which resulted in capturing major variances in player performance, about 71% of variance was explained for batters and 65% for bowlers. Then the result of analysis was compared to ESPN's rankings, which concluded that PCA can capture more detailed performance insights while using multiple variables simultaneously. Hence, this validated PCA's efficacy when it came to player's performance analysis.

Clustering is another unsupervised learning technique which can use different algorithms to group data according to how similar they are. Bose et al. (2021) applied a few clustering methods to evaluate the player performance in T-20 cricket. These methods included K-Means, Fuzzy C-Means, Agglomerative Clustering, and Gaussian Mixture Models. The study's aim was to classify players into categories like batters, bowlers, and all-rounders. Based on silhouette scores, the study found that Agglomerative Clustering was the most effective method to classify players and then it used auto-encoders, a deep learning unsupervised learning technique, to reduce the dimensions and refine the results. The performance metrics that the study used were like average runs, strike rate, wickets taken and economy rate. These metrics clustered players in distinct groups and recommended the best playing eleven according to the pitch's condition. The study also identified limitations and areas where future research could be done, it suggested using field metrics, usage of limited data of fresh players, weather conditions and toss results.

In summary, unsupervised learning techniques have further enhanced the ability to uncover hidden patterns and correlations in a player's performance data. Using methods such as PCA, clustering, auto-encoders that help understand player performance much better and make

strategical decisions regarding team selection and strategy. The next section will investigate research that has used the combination of supervised and unsupervised learning techniques, which has helped researchers make a comprehensive model for analysing player performance.

2.4. Combination of Supervised and Unsupervised Learning

As mentioned above, most of the unsupervised learning studies that have been done in analysing player's performance in cricket have been in combination with supervised learning. This integration has developed a more holistic and comprehensive model to analyse the player's performance. This approach uses the strengths of both methodologies, hence allowing a much greater perspective and predictions.

One such example is the use of Category-based Deep Performance Index (CDPI) by Prakash (2016a) where the author uses clustering method to group players based on their batting positions or bowling roles. In the study RFE was used to extract relevant features and assign weightage to develop CDPI. The result showed a much more comprehensive ranking compared to those given by traditional metrics such as MVPI.

Another example is the use of Deep Player Performance Index (DPPI) by Deep Prakash and Verma (2022). DPPI uses K-Means clustering in combination with supervised learning algorithms such as Random Forest to analyse player's performance in T20 cricket. The research first clustered players according to their roles and then applied Random Forest to predict player's performance within the assigned clusters. Hence, DPPI's result showed a more accurate representation of measuring player's capabilities in each context. Using Random Forest in combination with clustering increases the model's prediction accuracy hence allowing analysts to identify the key performance indicators (KPI) that are most relevant for different player roles.

In addition to the above, the combination of both the machine learning techniques has helped predict match outcomes and team recommendations. Jayanth et al. (2018) in their research titled 'A Team Recommendation System and Outcome Prediction for the Game of Cricket' illustrated frameworks that could predict outcomes of a match and at the same time recommend team for it as well. They used Support Vector Machine (SVM) models with varying kernels such as linear,

polynomial, and radial basis function (RBF) and non-linear SVM with RBF kernel. Furthermore, the study used K-Means clustering methods using historical data, such bowling average, batting strike rate etc., to recommend players in the team for the next match. To validate their models, the authors used 2011 Cricket World Cup data.

In conclusion, using supervised and unsupervised learning together has provided a broader area to work on when analysing a player's performance. Using the strengths of both learning techniques, an analyst can produce a more comprehensive model which can incorporate the complexity of player's performance metrics. Before moving onto the limitations and gaps of current literature, this study will also review the use of machine learning techniques in other sports.

2.5. Machine Learning Techniques Used in Other Sports

Machine learning techniques have been used beyond the realms of cricket, and they have been used effectively in sports such as football, cycling and American football. The methodologies below provide value contribution to sports analytics and can be adapted when used for evaluation for player's performance.

In football, Spencer et al. (2016) used data from the 2012-2014 Australian Football League (AFL) seasons and applied K-Means clustering technique to analyse performances which resulted in eight distinct team profiles. Random Forest was used to predict match outcomes according to significance of 'team.' Furthermore, the study found that 'longitudinal' changes in profile of the teams, for example the change of coach, can affect the classification of the team over time.

In track cycling, statistical methods, such as Kolmogorov-Smirnov test, and machine learning techniques, such as K-Means clustering and the Naïve Bayes Classifier, were used by Ofoghi et al. (2013) to analyse the omnium performances. Key performance indicators were identified predicted the outcomes of races, assisting coaches and athletes to develop their training and competition strategies.

Similarly, Shelly et al. (2020) in their research used K-Means clustering to create training groups for elite American football student athletes based on their game demands by using wearable

devices to extract data and optimize the training routines. As done in cricket, the methodology clustered the athletes into groups/categories that represented their game rather than positions.

In summary, the methods of machine learning used and their application in other sports provides a valuable direction that analysis of player performance in cricket can take. The next section will identify the gaps in current literature and provide how our study will deal with them.

2.6. Gaps in Current Literature

The literature on analysing performances in cricket has transformed and evolved remarkably in past few decades, i.e. from using traditional methods to analyse performances to using supervised and unsupervised learning. Even though these literatures provide a variety of ways to analyse performance, it leaves a very crucial aspect which is identifying high performing junior players through performance analysis as outlined in my research. The gaps help us identify the areas where our work will be concentrated on and develop models which will be effective and applicable to junior cricket leagues

One of the major gaps in the above literatures was how little literature is available in applying unsupervised learning techniques to junior cricket league. Most of the literature that is available are focused on senior cricket leagues hence leaving a gap on how to apply these techniques to junior cricket. For example, Deep Prakash and Verma (2022) applied PCA and K-Means clustering to analyse performances of professional T20 cricket players and teams, but it fails to identify on how these methods can be applied to evaluate player performance in junior cricket leagues.

Furthermore, the dynamics, context, and performance metrics for junior cricket leagues are completely different to professional cricket leagues hence the traditional and advanced methods may not be directly applicable to junior cricket leagues. Our study does aim to adapt one of the indices of the existing literature to measure player's performance metric in junior cricket league. The application of the index will be further discussed in detail in the methodology section.

Moreover, there is extremely little research on how the factors such as team dynamics, match situations, and opposition strength can be crucial in influencing player's performance. Saikia and Bhattacharjee (2011) did consider some of these factors and used Bayesian approach on them

whereas Prakash (2016a) has introduced weightage systems to be used according to team's strength and match situations but still a comprehensive approach needs to be developed to consider wider context specially when it comes to junior cricket league. One of the approaches that my study will be using is assigning weights to teams according to their league position. The application of weightage system will be discussed further in the methodology section.

The existing literature has used a lot of past data and related ranking to develop their methodology, this kind of ranking is rarely available for junior players. Mostly, players in junior cricket league are not ranked the way professional players are ranked. Hence, this takes out the option of applying supervised machine learning to classify the high performing players as supervised machine learning uses labelled datasets to validate the models.

The gap in studying the performance of junior players also adds on the factor of whether the categories of players (batter, bowlers, wicketkeeper, all-rounder) and their ideal performance metrics in the above literature are like what we will find out in junior player's performance metrics. Unfortunately, the data we are using for the junior players are not assigned as batter, bowlers, or all-rounders. Therefore, will the typical performance metrics of each category be as a threshold to consider someone in junior team from that category?

As mentioned above, there has been extremely small number of studies done in using unsupervised learning only to cluster players according to their performances. The studies that have been done, provide an extremely detailed version on how the methods are applied but again the question arises will the methods be accurate enough for junior cricket players? Our study aims to apply unsupervised learning only to cluster the high performing players according to the player's performance metrics.

In conclusion, for the past few decades there has been a significant amount of study done using data-driven performance analysis in cricket, but certain gaps remain which are needed to be addressed to make the methods applicable in junior cricket. By focusing of unsupervised learning techniques, using derived performance metrics of junior players, and involving contextual factors, our study will provide an enhanced approach in evaluating the junior cricket league player's

performance. Once these gaps are addressed and our methods are applied successfully, we aim to identify young talent to be promoted to higher leagues.

3. Methodology And Data

The study extracted data from Durham Cricket County's Junior League website (https://dcbjuniorlge.play-cricket.com/home). As previously mentioned, the focus of this study will be the Under-13 Leagues as these players will be promoted to higher tier competition and furthermore, in addition to helping grassroot level cricket it will also help senior cricket teams in the future as good junior players would eventually be selected to senior team. The study will be focusing on last year's season i.e. 2023. This is because when the study commenced, the 2024 season was yet to begin and by the end of the study the 2024 season does still have few matches to play. Therefore, 2024 season might not be able to give a comprehensive analysis or would have had some sort of bias if used.

Two types of data are extracted from the website – one is the complete scorecard of batting and bowling performances; the other is ball by ball data. The reason of extracting two types of data is to compare the result of clustering and checking if adding in additional metric makes any kind of change to clustering.

Once the data is collected from the website, the data is pre-processed and cleaned to apply derived performance metric formulas to it. The derived performance metrics would give a holistic view of player's performance. Then we will apply PCA and Autoencoders separately. After having the reduced dimension datasets, KMeans++ and Hierarchical Clustering was applied. The cluster with high performing players were then ranked using a composite score on scaled derived performance metric.

3.1. Data Collection

The major aspect of this study was to extract the under-13 Junior Cricket League 2023 season data from the Durham Cricket County's Junior Cricket League website. To extract the data, the software that was used was Python and the libraries used in Python were *Numpy, Pandas,*

Selenium and *BeautifulSoup*. As mentioned above, there were two types of data that we would be extracting:

1. Batting and Bowling Scorecard
2. Ball By Ball Data

As the website was accessed through Selenium, the leagues in the season 2023 were extracted and collected in a separate file. The following are the leagues and the number of teams playing in that league:

League Name	Number of Teams
Central U-13 League Division One	6
Central U-13 League Division Two	8
Northeast U-13 League Division One	5
Northeast U-13 League Division Two	8
Southeast U-13 League	11
Southwest U-13 League	10
Tyneside U-13 League Division One	7
Tyneside U-13 League Division Two	8

Table 3.1: List of Leagues and Number of Teams in Each League

The teams in each league can be viewed in Appendix A Table 1.

The scorecards for the teams in each league were accessed using *BeautifulSoup* and *Selenium.* The scorecard page had two kind of data – one tab was the complete Batting and Bowling Scorecard for both the teams and the other tab was the Ball-by-Ball Data. The metrics extracted can be seen in Appendix A Table 2.

For the batting scorecard 4,543 data points were extracted. This meant that each batter had multiple entries because they played multiple matches. Similarly, for the bowling scorecard 3,457 data points were extracted. Whereas for the ball by ball, 38,416 data points were extracted. The ball-by-ball data, as the name states, had the data for each ball played in the leagues above.

To make sure that records were not being duplicated, a list of matches played was being updated. For example, if 'Team A vs Team B' was recorded using Team A's scorecard it is then kept in a separate list of 'Match Recorded'. When Team B's scorecard is accessed, the Python script will

ignore the match 'Team A vs Team B' as it was already in the list of matches that were recorded. Hence, this helped us avoid the duplication of data points.

As the batting scorecard had the names of all the players that were involved in the match, whether they batted in it or not. The scorecard was used to make a separate *csv* file for all the teams which recorded the names of the players. Hence, we created a team sheet for each team. The team sheet did have a problem with names of the players – there were duplicate entries for some players because those players had two styles of name. One was that they had their complete first and last name and the other style was they had the initial of the first name and complete last name. For example, when the team sheets were made, in 'Brandon CC, Durham Under 13' the team sheet had 'T Burnham' and 'Thom Burnham'. Both players, we assume, are the same players but their names were recorded differently for different matches. Hence, an iterative operation was run through the team sheet file which did the following:

1. Collect the list of all the players in the team (including duplicate ones).
2. Iterate through the players, create a current name variable which splits a player's name into two parts as a list for example 'T Burnham' would be split into *['T', 'Burnham'].* The first part of the list is the first name, and the second part of the list is the last name.
3. If the first part of the list is greater than 1, that means we have the complete name of the player. We create a separate list and append this complete name into it.
4. If the first part of the list is not greater than 1, then we create a separate list which is for players having initials as first name.
5. Then we iterate through the initial list and the complete name list. If the last name of the initial list and the last name of complete name list is equal and the first element of initial list i.e. the initial of the first name, and the first element of the complete name list is equal that means we are looking at the player with similar names but different style.
6. We extract the complete name and put it in a separate list which is for players with complete name.
7. We repeat this process with each player name in the Team Sheet file and can remove the duplicates effectively and make a new Team Sheet file for the team.

3.2. Data Cleaning and Preprocessing

As the data was extracted from the website, it is important to check whether there is consistency in the datapoints or not. Furthermore, from the extracted performance metrics, a derived performance metrics system will be developed which is like the development of formulas done by Prakash (2016). After calculating these performance metrics, an exploratory analysis was performed on the data to check the distribution of the data and its correlation and then accordingly the data was transformed and scaled. This section details the data cleaning process for all the 3 scorecards and then preprocessing of the clean data which will be used for analysis.

3.2.1. Data Cleaning

The data cleaning process will involve making the data consistent by removing '*NaN*' or '*NA*' values, making sure the team names are identical for each data point, applying an iterative process to make sure that the names of the players are consistent, calculating thresholds to filter out players, making a complete season scorecard and then using that scorecard to apply the derived performance metrics formulas.

Two processes that are common in all the scorecards are revising the name of 'Tudhoe CC Under 13' and correcting the name of players. In starting matches the team 'Tudhoe CC Under 13' was also known as 'Tudhoe CC Under 13 B', to make sure that all instances containing 'Tudhoe CC Under 13' and 'Tudhoe CC Under 13 B' were consistent, their names were changed, where needed to be, to 'Tudhoe CC Under 13'.

One of the major problems that exists in the raw scorecards was the name of the players. The scorecards contained inconsistency with names of the players such as mentioned above i.e. 'T Burnham' will also be 'Thom Burnham' in the Batting Scorecard. Hence, we extract all the unique team names from the scorecards and open their relevant Team Sheet file in the Python Script. All the players in the Team Sheet file are extracted to a list and a dictionary is created with Team Names as keys and players with complete names as values.

We then iterate through each team and select the rows that are related to the team. One by one, we extract the player's name from each row and compare it with the relevant team's team sheet

in the dictionary we created earlier. The comparing of the names is like what we did before when creating Team Sheets but with a minor change. First, we check whether the player's name is exactly like the name on the Team Sheet, if yes then we iterate to next batter. If not, we select the player's last name and find the equivalent last name in Team Sheet simultaneously making sure that the first name initial is the same. Once found, we replace the player's name with the name on Team Sheet. In this way we can have consistent names of all the batters in the scorecards.

Batting Scorecard:

The batting scorecard had some batters who did not play in matches hence they had '*NaN'* value against them. Therefore, these batters were dropped from the batting scorecard. This reduced the data points to 2,982. The bowling side for batting scorecard was not necessary anymore hence that column was also dropped from the Batting Scorecard.

Moving on, we take out the average runs scored in each innings/match by the batters. We have done this because in the Junior League, player is do not get a chance to make a fifty or hundred. We had to create a metric which was considered as a batter having a 'big inning.' A new column, 'Above Average' was added to the Batting Scorecard. If the batter scored equal or more than the average score in an innings/match, a '1' will be added to the 'Above Average' column for that match. The average score came out to be 8 runs. So, any batter who scored more than 8 runs in an innings/match, the column 'Above Average' would have a '1' listed against the batter for that very match.

As we move ahead to calculating over own performance metrics, we first collated each batter's complete statistics. A player's statistics data frame was made which contained the following columns:

Column Name	Description
Team	The name of the team the batter belongs to
Player	The name of the batter
Total Matches Played	Total matches played by the batter in 2023 season

Total Runs	Total runs scored by the batter in 2023 season
Total Balls	Total balls faced by the batter in 2023 season
Total 4s	Total 4s scored by the batter in 2023 season
Total 6s	Total 6s scored by the batter in 2023 season
Fours To Sixes Ratio	$\frac{Total\ 4s}{Total\ 6s}$
TRWB	Total runs scored without boundary
Balls Without Boundary	Total number of balls faced without scoring a boundary
Batting Average	Season batting average
SR	Season strike rate
Total Times Above Average	Total number of times the batter scored above average in a match during the 2023 season

Table 3.2: Description of player statistics for the 2023 season (Batting Scorecards)

Once the new data frame was made, we had the statistics of 937 unique batters. Then we calculated the average of 'Total Balls' which came out to be 42 (7 overs). We selected the batters who played 42 or more balls because this would separate the batters who have played a substantial number of matches. Players who batted for less than 42 balls will not provide a reliable measure of the performance metrics and hence might skew their abilities and the whole dataset as well. After filtering those who have played more than 42 balls, we were left with 339 batters. We added another layer of filter i.e. the batters should have played 4 or more matches. 4 was selected as number of matches because some of the leagues only have 5 to 6 teams in them hence consistent players in that league must have played a minimum of 4 matches. Players who have played 4 or more matches also gives a holistic view of their contribution to the team and finally, these players provide a more robust sample size hence statistics will be more reliable. After applying the filter, we were left with 269 batters.

Batting Scorecard (Ball by Ball):

In addition to fixing the names of the players, there were multiple steps taken to clean the data for preprocessing. Given that ball by ball data had the information related to both, batters, and bowlers, we decided to make a scorecard type data frame before applying further cleaning processes. This meant that the ball-by-ball data was separated into two further datasets – one was a batting scorecard, and the other was bowling scorecard.

We dropped the columns 'Bowling Team' and 'Bowler' as that was irrelevant to the batting scorecard and its analysis. While scraping the data from the website, it was found that there are missing names of some of the batters hence we created our own '*NA NA'* value to fill in the first and last name of these batters. Our first step in cleaning was to drop these rows and this reduced our data to 35,125 balls.

One of the fundamental issues faced during scraping of the data from the website was the inconsistency of different webpages on the website. We had to create multiple '*try'* and *'except'* codes to tackle this problem, yet we faced inconsistency in gathering the 'Run Type' data. There were several rows which had 'Run Type' ending in 'Play,' this would have caused issues when making our own scorecard as we had to use 'Run Type' to collate the data of whether the batter scored while running between wickets or through a boundary or was dismissed. Hence, we were able to remove the 'Play' text from these rows. As we are concentrating on batter's data, we did not need the runs which were scored through extras. Hence, we dropped all such rows leaving the dataset with 29,980 balls.

We renamed some of the 'Run Type' to the following:

- If the 'Runs' column has 0 in it, then the 'Run Type' will be 'dot ball'
- If the 'Runs' column had 1, 2, 3 or 5 in it then the 'Run Type' will be 'rbw' which stands for running between wickets
- If the 'Runs' column had 4 or 6 in it, then 'Run Type' will be 'boundary'

Now the dataset was ready to be converted into a Batting Scorecard for each match, which we did by grouping the data by 'Match,' 'Team' and 'Batsman.' It was not a traditional Batting Scorecard as we added in our statistics in it. The scorecard had the following columns:

Column Name	**Description**
Match	The teams playing in the match
Team	The team which the batter belongs to
Batter	Name of the batter
Runs	Runs scored by the batter in the match
Balls	Number of balls faced by the batter in the match
TDB	Number of dot balls faced by the batter in the match

Fours	Number of fours scored by the batter in the match
Sixes	Number of sixes scored by the batter in the match
Total Boundaries	Sum of Fours and Sixes
Runs Scored (Boundaries)	Runs scored by the batter through boundaries in the match
Runs Scored (RBW)	Runs scored by the batter through running between wickets in the match
O/NO	If the batter were out, the column would have 1 in it. If the batter was not out, then the column would have 0 in it.
Strike Rate	Number of runs per 100 balls

Table 3.3: Customised Batting Scorecard for ball-by-ball data

Hence, we had the statistics of 2,461 batters including duplicates. Just like what we did in the Batting Scorecard, we took out the average runs scored in each innings and added a column 'Above Average' to record if the batter scored runs above average in the innings or not. The average runs scored again came out to be 8. Moving ahead with collating each batter's statistics, we used the same player statistics in Table 3.2 but with some additional statistics which are given below:

Column Name	Description
Runs with Boundaries	Total runs scored through boundaries by the batter in the 2023 season
Total Dot Balls	Total dot balls faced by the batter in the 2023 season
Total Dismissed	Total number of times the batter was dismissed during the 2023 season
Total Not Dismissed	Total number of times the batter was not out during the 2023 season

Table 3.4: Additional batting statistics for ball-by-ball analysis

After collating the data, we had the data for 804 unique batters. Again, like we did in Batting Scorecard, we take out the average number of balls faced by the batters during the season which came out to be 38. Then we selected those batters who played more than or equal to 38 balls (6.2 overs) and were left with 300 unique batters. And the last filter that we applied is batter should have played 4 or more, leaving 222 batters in the dataset.

Bowling Scorecard:

The bowling scorecard did not have any missing values, the 'Batting Side' column was dropped from the bowling scorecard as it was of no use. Few more columns were added to the bowling scorecard, those were '3W', '4W' and '5W'. These columns stated whether the selected bowler

took 3 wickets, 4 wickets or 5 wickets in the match. Like batting scorecard, player statistics were calculated for the season which will be used in the formulas of derived performance metrics:

Column Name	Description
Total Matches Played	Number of matches played by the bowler in 2023 season
Total Overs Bowled	Number of overs balled by the bowler in 2023 season
Total Balls Bowled	Number of balls bowled, including extras, by the bowler in 2023 season
Total Maidens	Number of maiden overs bowled by the bowler in 2023 season
Average Balls Bowled per Match (ABBM)	Average number of bowls bowled by the bowler per match
WTS	Strike rate of the bowler for the 2023 season
Total Wickets	Number of wickets taken by the bowler in 2023 season
Total Runs Conceded	Number of runs conceded by the bowler in 2023 season
Economy Rate	Economy rate of the bowler for 2023 season
Total Extras	Number of wides + no balls + byes bowled in 2023 season
Total 3Ws	Number of times bowler takes 3 wickets in a match
Total 4Ws	Number of times bowler takes 4 wickets in a match
Total 5Ws	Number of times bowler take 5 or more wickets in a match
Total Matches with 3W+	Sum of Total 3Ws, Total 4Ws and Total 5Ws
Total Wickets (3W+)	Sum of 3, 4 and 5 wickets taken
Consistent Score (CS)	Bowling Average

Table 3.5: Description of player statistics in 2023 season (Bowling Scorecards)

After calculating the statistics for each bowler, we had 948 unique bowlers in our new data frame. Next, we calculate the average overs that were bowled throughout the 2023 season. The goal was to select bowlers who have bowled at least average number of overs in the season. The reason is like the reason we filtered out players in batting scorecard, to have more reliable measure of performance metrics of the bowlers. The number of bowlers were reduced to 405. We also wanted to make sure that the bowlers played substantial numbers of matches only couple of matches. Hence, we selected all the bowlers who played less than 4 matches leaving us with 390 bowlers.

Bowling Scorecard (Ball by Ball):

The ball-by-ball bowling scorecard data cleaning process was same as ball-by-ball batting scorecard cleaning process, with just a few changes in the data that is to be extracted from the complete dataset. After dropping the rows which contained bowler's name as 'NA NA' we were

left with 35,317 balls. In the bowling scorecard, we will consider all kinds of 'Run Type' i.e. runs, dismissed, boundary, wides, no balls, byes. This is because it adds to different statistics of a bowler. The wides, no balls and byes were renamed 'extras' as the 'Run Type,' whereas the rest of the 'Run Types' remained like what they were renamed in the Batting Scorecard.

Just like the ball-by-ball batting scorecard, we made our own customised bowling scorecard for each match. The following were the columns we used:

Column Name	Description
Team	The team the bowler belongs to
Bowler	The name of the bowler
Balls Bowled with Extras	Total number of balls bowled including extras in the match
Extras	Total number of extras bowled in the match
Runs	Total runs conceded in the match
Wickets	Total wickets taken in the match
Dot Balls	Total number of dot balls bowled in the match
Boundaries	Total number of boundaries conceded in the match
Balls Bowled without Extras	Total number of balls bowled excluding extras in the match
Overs Bowled	Total number of overs bowled in the match
3W	If the bowler took 3 wickets in the match, the column would have 1 in it
4W	If the bowler took 4 wickets in the match, the column would have 1 in it
5W	If the bowler took 5 wickets in the match, the column would have 1 in it

Table 3.6: Customised Bowling Scorecard for ball-by-ball data

After creating a bowling scorecard for each match, a new data frame was created which captured the bowler's statistics for the season. This included all the columns from the Bowling Scorecard's player statistics table except for a couple more that were added which were related to the number of boundaries the bowler conceded and the number of dot balls the bowler bowled:

Column Name	Description
Total Dot Balls	Total number of dot balls bowled by the bowler in the 2023 season
Total Boundaries	Total number of boundaries conceded by the bowler in the 2023 season

Table 3.7: Additional player statistics that were added to Bowling Scorecard's player statistics. These will be used to calculate derived performance metrics of bowlers from ball-by-ball dataset

We were able to get the statistics of 1,043 unique bowlers. These were further reduced to 405 after we calculated the average number of overs bowled by bowlers in 2023 season (which came out to be 5) and selected the bowlers who bowled equal to or more than the average. Then the

number of matches played filter was also applied i.e. the bowlers should have played at least 4 matches. In the end we were left with 308 unique bowlers.

3.2.2. Data Preprocessing

Before we proceed to data preprocessing, we would first define the derived performance metrics which will help us cluster players at the end of our analysis. The derived performance metrics are used so that we get a holistic picture of the player's performance. Rather than concentrating on a particular statistic, a combination of statistics will give a complete view of the performance of a player. The derived performance metrics can be viewed in Appendix A Table 2.

Batting Scorecard:

As derived performance metrics for all the batters have been calculated, now exploratory data analysis is done to view the distribution and correlation between the performance metrics. The following are the descriptive statistics for the batters from Batting Scorecard:

	BFI	BI	FSS	CS	RBWI	BII	FSRatio
mean	15.90895	5.462884	0.627683	8.076134	0.326719	0.446522	0.065732
std	5.965744	4.492395	0.306734	7.685751	0.146066	0.283776	0.05544
min	6	0	0.0769	0.0514	-0.0256	0	0
25%	11	1.7143	0.3864	2.0101	0.2154	0.2222	0.02
50%	14.6	4	0.6167	5.6008	0.3176	0.4444	0.05
75%	19.5	8.25	0.8511	11.6932	0.4176	0.6667	0.1
max	40.1429	24.5	1.7083	37.8225	1.1224	1	0.25

Table 3.8: Descriptive statistics of derived performance metrics (Batting Scorecard)

From the above we can see how spread out the derived performance metrics are, this gives us an indication of scaling the data so that the influence of each metrics is similar. The histogram below confirms that the data needs to be transformed before making any analysis as majority of the derived performance metrics are skewed to the right. The histograms were made using *matplotlib.pyplot* and *seaborn* library in Python:

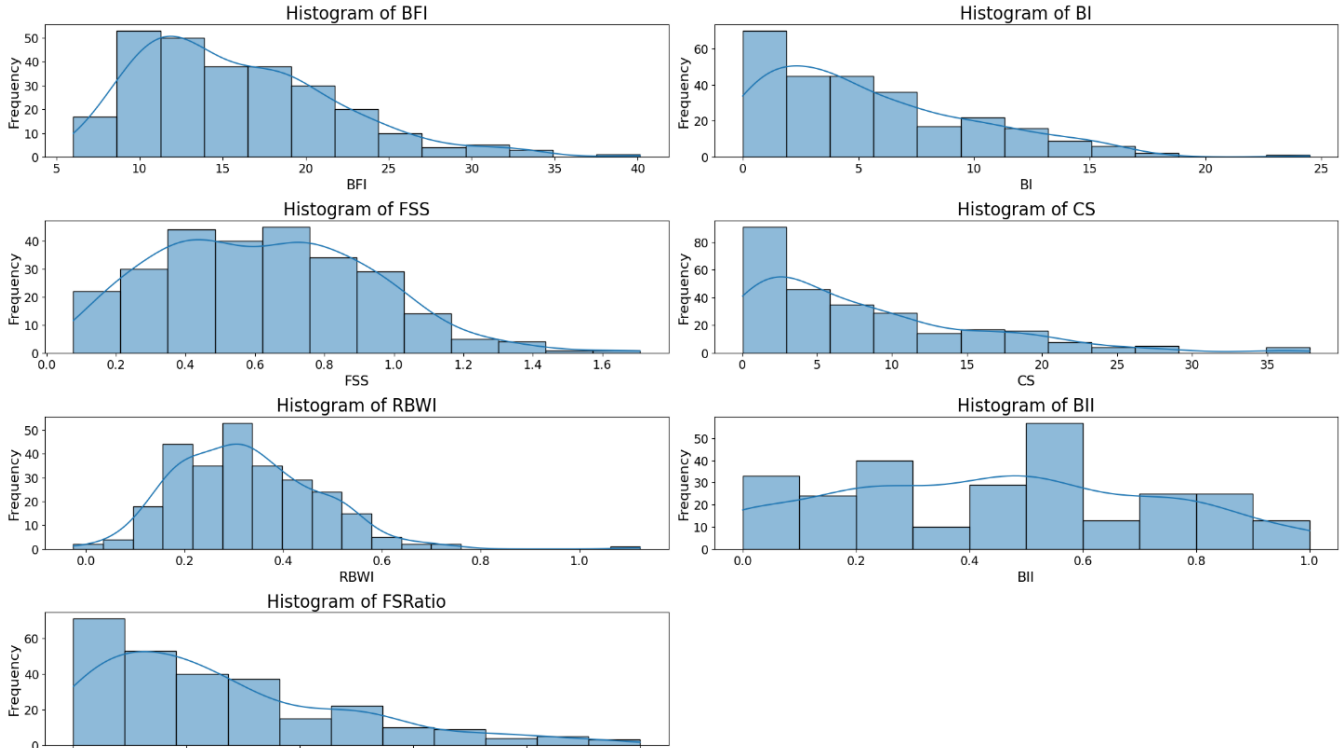


Figure 3.1: Histogram of derived performance metrics (Batting Scorecard)

It is also important to check the correlation between the metrics to confirm that reducing the dimensionality will help in analysing the data furthermore, it also provides insights on the

relationship between different metrics. Below are two plots which show the correlation between each derived performance metrics:

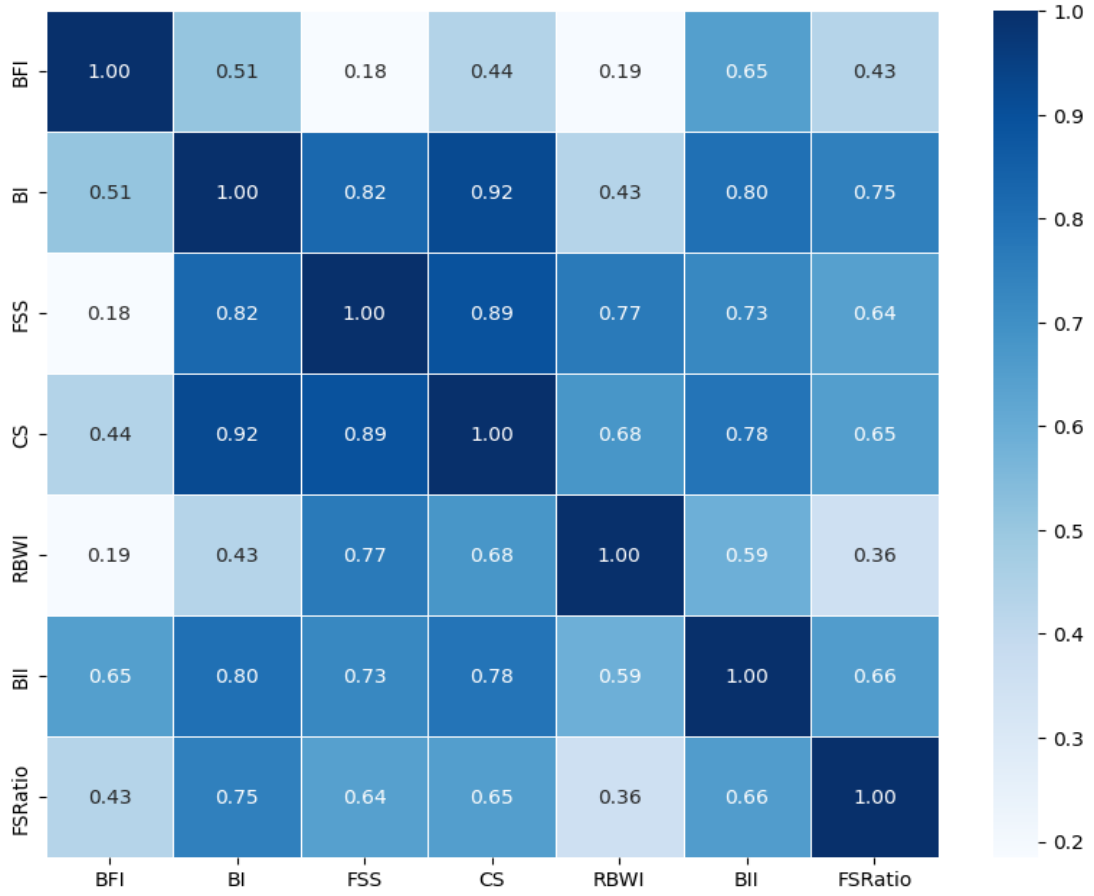


Figure 3.2: Correlation matrix of derived performance metrics (Batting Scorecard)

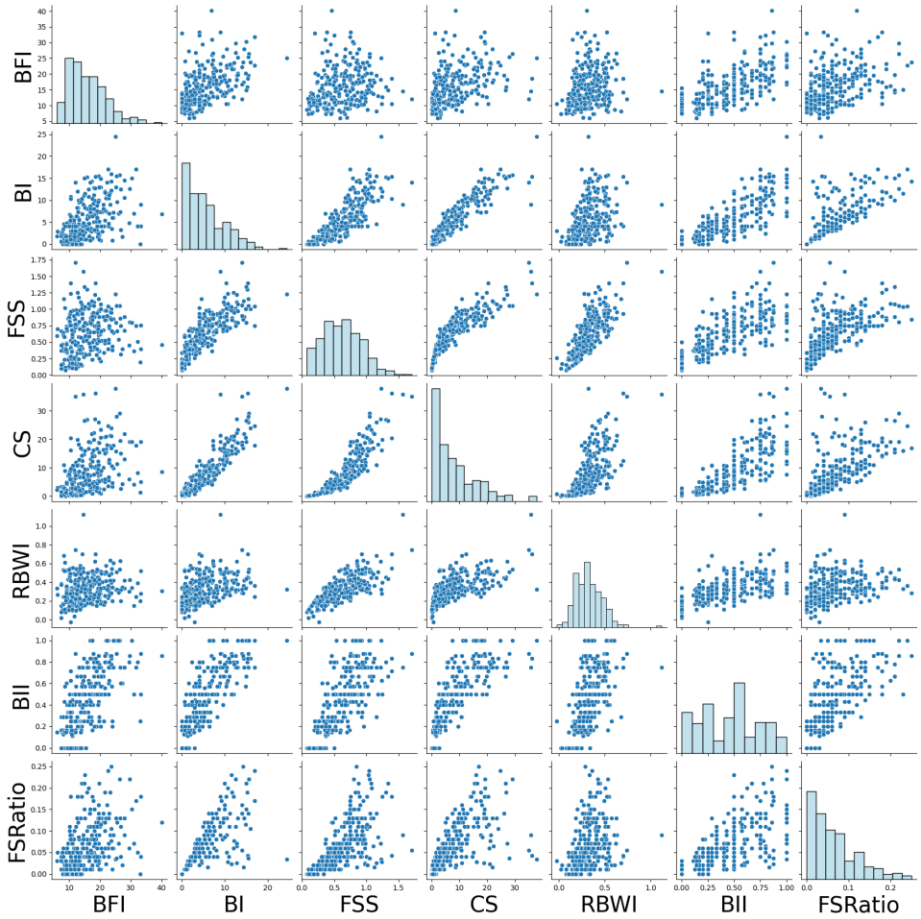


Figure 3.3: Pair plot of derived performance metrics (Batting Scorecard)

Figure 3.2 and Figure 3.3 shows strong correlations between several metrics for example, Boundary Index (BI) is strong correlated to Fast Scorer Score (FSS), Consistent Score (CS), Big Innings Index (BII) and Fours to Sixes Ration (FSRatio) – this does make sense as scoring runs quickly (having a high FSS) requires batters to hit boundaries hence the strong positive correlation. The main point of the above graphs is that given the strong correlation between some metrics, it is justifiable to perform dimension reduction before analysis as strong correlated metrics convey similar information regarding the data, it can reduce the curse of dimensionality (more about this in theoretical framework) and simplify the models further with reduced number of features.

Batting Scorecard (Ball by Ball):

The following is the descriptive statistics for the Batting Scorecard extracted from ball-by-ball data:

	BFI	BI	FSS	CS	RBWI	BII	FSRatio	DBI	FS
count	222	222	222	222	222	222	222	222	222
mean	14.50619	5.168083	0.660552	8.246518	0.350727	0.42505	0.061633	8.559837	0.308495
std	5.541785	4.326201	0.332823	7.795435	0.152077	0.287561	0.05409	3.778372	0.212493
min	5.4286	0	0.0208	0.0035	0.0208	0	0	2	0
25%	10.4	1.6	0.4042	1.88615	0.233425	0.2	0.02	6	0.1667
50%	13.20835	4	0.66155	5.5538	0.3426	0.4143	0.05	8	0.2857
75%	17.65003	7.916675	0.90525	12.45287	0.46945	0.6667	0.09	10	0.5
max	42.4	17.2	1.7011	38.1254	0.75	1	0.25	29.8	0.8333

Table 3.9: Descriptive statistics of derived performance metrics (Batting Scorecard - Ball by Ball)

Just like the Batting Scorecard, we see a lot of variances in the derived performance metrics which means scaling the data before starting our analysis. Furthermore, in some of the histograms of these derived performance metrics (see Appendix A Figure 1) right skewness is observed, which indicates in performing transformation and the correlation matrix (see Appendix A Figure 2) and pair plot (see Appendix A Figure 3) also indicates that most of the metrics have a strong correlation so performing dimensionality reduction would be better for clustering analysis.

Bowling Scorecard:

After performing calculations on bowler's statistics, we perform the same steps that we did for the batter's statistics i.e. exploratory data analysis. The following is the descriptive statistics for the Bowling Scorecard:

	WTS	CS	ES	BBI	BWTS	SPS	MI	EI	AVMS
count	390	390	390	390	390	390	390	390	390
mean	0.051442	19.03132	4.287605	14.35006	0.043085	0.576734	0.111513	1.168764	0.215202
std	0.031162	17.7329	1.561571	2.568622	0.088741	0.345761	0.106822	1.099905	0.198266
min	0	0	0.8	8.4857	0	0	0	0	0
25%	0.0304	8	3.1601	12.5	0	0.3333	0	0.4	0.0952
50%	0.04665	13.6667	4.1914	14	0	0.5714	0.09305	0.8571	0.15855
75%	0.0682	22.3333	5.151075	16.3214	0	0.75	0.1667	1.6	0.269
max	0.2361	99	10.4878	21.75	0.6	3	0.5	9	1.6789

Table 3.10: Descriptive statistics of derived performance metrics (Bowling Scorecard)

Again, like the Batting Scorecards, there is a lot of variances in the derived performance metrics hence indicating to scale the data so that we have a more balanced influence of all the metrics. To check data transformation was needed or not, histograms of all the metrics were plotted:

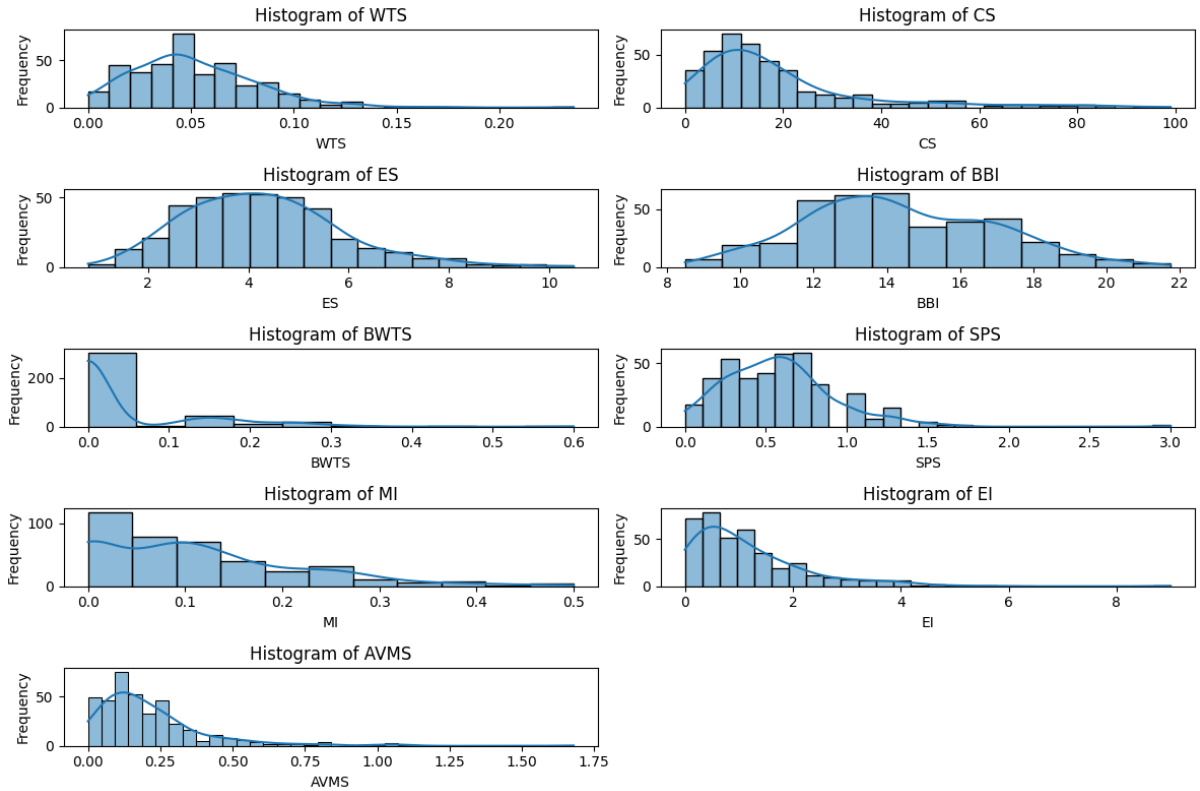


Figure 3.4: Histogram of derived performance metrics (Bowling Scorecard)

Figure 3.4 shows most of the metrics are skewed to the right, hence it is better to perform transformation on all the metrics to reduce the effect of skewness when performing the analysis.

Given there are 9 performance metrics, it is better to reduce the dimensions to perform our analysis, but it is important to check the correlation between the derived metrics to see if there is any kind of relationship between the derived metrics:

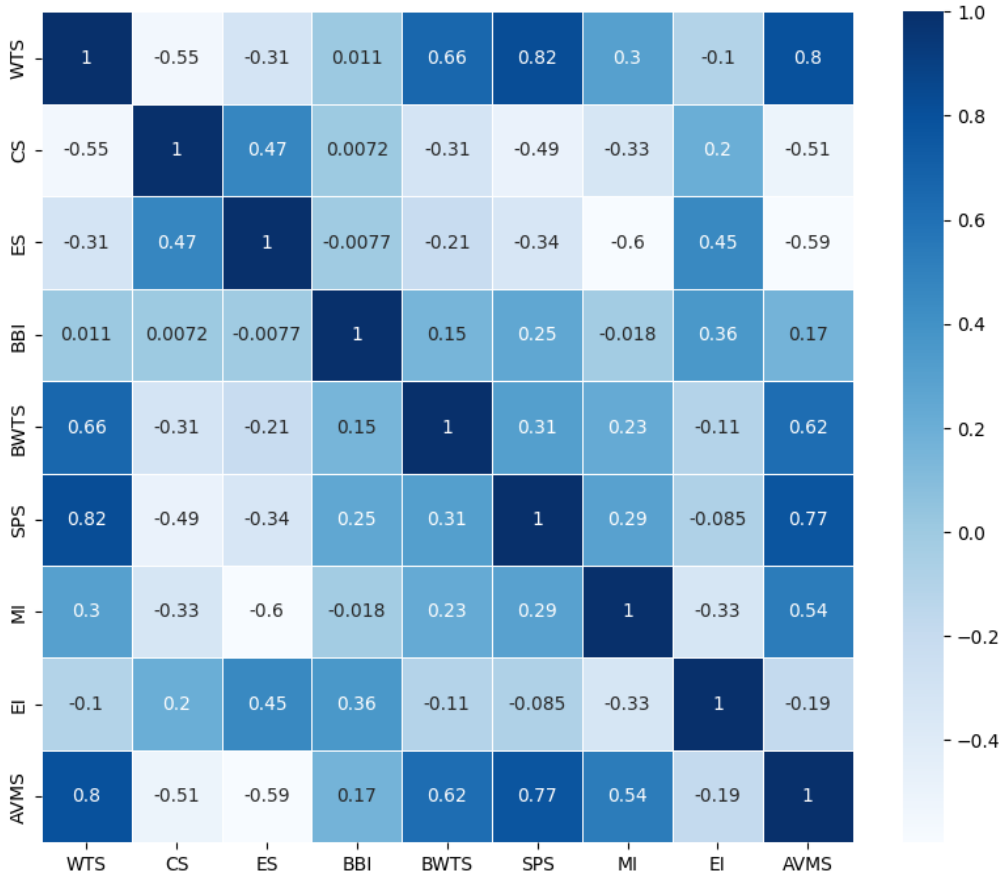


Figure 3.5: Correlation matrix of derived performance metrics (Bowling Scorecard)

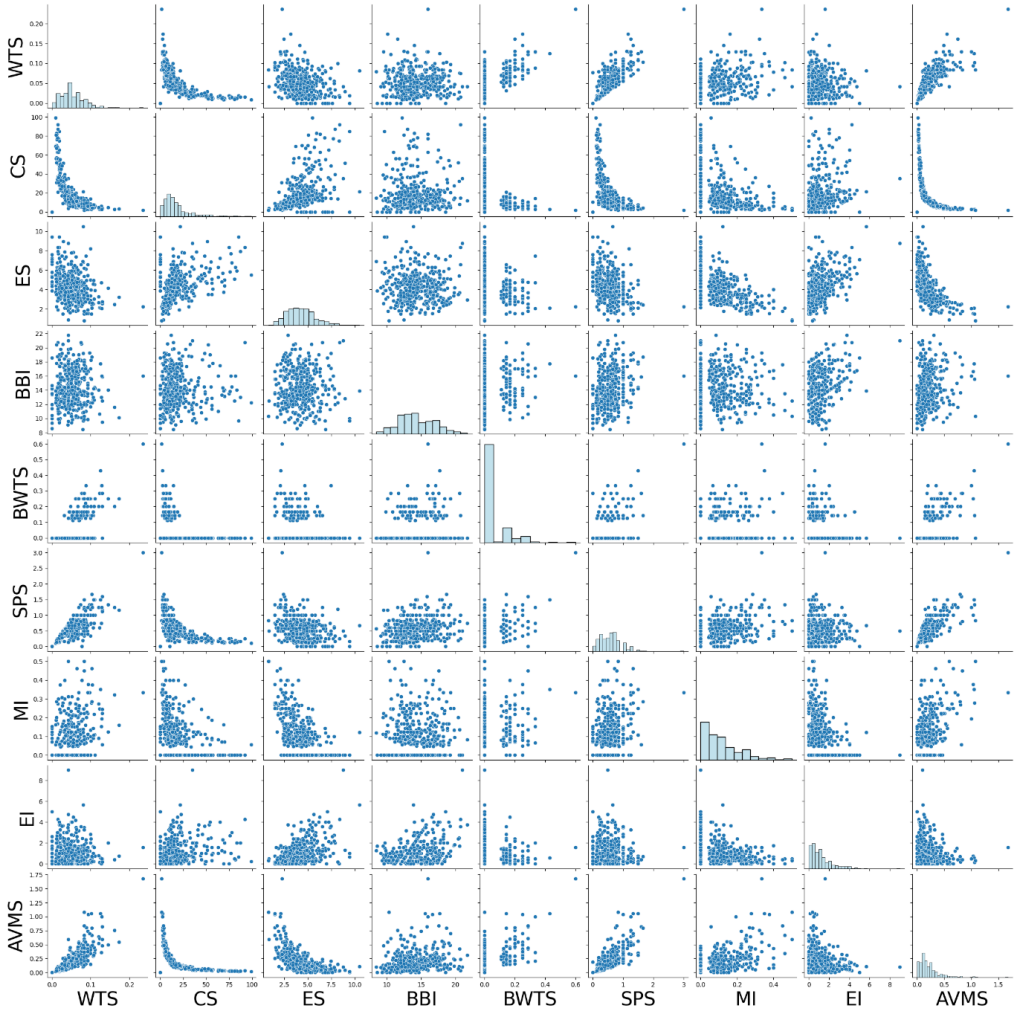


Figure 3.6: Pair plot of derived performance metrics (Bowling Scorecard)

Like Batting Scorecard, in the above plots we can see some strong correlation between metrics. This justifies the usage of dimensionality reduction techniques on the Bowling Scorecard derived performance metrics. Furthermore, there are some metrics which has a non-linear relationship such as Consistent Scorer (CS) with Wicket Taker Score (WTS) or CS with Short Performance Score (SPS) etc. this also indicates that if we use autoencoders the dimensions will be reduced effectively as neural networks are designed to capture these non-linearities which will lead to better clustering analysis as well.

Bowling Scorecard (Ball by Ball):

As we have done with previous scorecards, we performed exploratory data analysis to check whether the scorecard needed scaling and transforming before the clustering analysis. In

addition, the exploratory data analysis will also give a check on whether this data needs to be its dimensions reduced or not (ideally would need to reduce the dimension given the amount of derived performance statistics there are).

	ES	WTS	BBI	CS	BWTS	SPS	DBI	EI	AVMS	BI
count	308	308	308	308	308	308	308	308	308	308
mean	0.859164	0.05548	13.99374	19.05884	0.038824	0.58426	0.596172	1.000575	1.115483	0.501225
std	0.284416	0.035532	2.796685	16.56027	0.085343	0.39374	0.101411	0.637607	1.111168	0.288276
min	0.135593	0	8.333333	0	0	0	0.333333	0	0	0
25%	0.672421	0.029851	12	8.75	0	0.333333	0.534884	0.536559	0.445509	0.319787
50%	0.826837	0.051503	13.70833	14.5	0	0.535714	0.6	0.894937	0.827658	0.444444
75%	1.019914	0.075118	15.7625	25	0	0.783333	0.662416	1.307692	1.454989	0.63173
max	1.909091	0.265625	24.5	119	0.4	3.666667	0.848485	4.459459	10.33333	1.71875

Table 3.11: Descriptive statistics of derived performance metrics (Bowling Scorecard Ball by Ball data)

Figure 3.11 confirms that data needs scaling and like previous scorecards, histogram (see Appendix A Figure 4), the correlation matrix (see Appendix A Figure 5) and pair plots of derived performance metrics (see Appendix A Figure 6) also confirm that the ball-by-ball Bowling Scorecard also needs its feature to be transformed.

3.2.3. Feature Scaling:

Using logarithmic transformation of data, the data's skewness will be better and then data will be scaled to a minimum value of 0 and maximum value of 1. To do this, we used Python's *'Numpy'* and *'Sklearn'* libraries.

After scaling the data, the following are the descriptive statistics of each scorecard:

	BFI	BI	FSS	CS	RBWI	BII	FSRatio
count	269	269	269	269	269	269	269
mean	0.464832	0.491266	0.428867	0.493219	0.388996	0.50418	0.27945
std	0.192875	0.245368	0.203726	0.252881	0.137822	0.28931	0.227497
min	0	0	0	0	0	0	0
25%	0.304322	0.308315	0.273926	0.291462	0.283895	0.28948	0.088744
50%	0.452454	0.496943	0.440561	0.509041	0.387608	0.53047	0.218649
75%	0.606679	0.686892	0.587371	0.690227	0.481577	0.736994	0.427125
max	1	1	1	1	1	1	1

Table 3.12: Scaled derived performance metrics of Batting Scorecard

	BFI	BI	FSS	CS	RBWI	BII	FSRatio	DBI	FS

count	222	222	222	222	222	222	222	222	222
mean	0.430517	0.528745	0.479465	0.499112	0.507967	0.481316	0.26243	0.468176	0.42192
std	0.177494	0.279529	0.206641	0.257004	0.207527	0.29509	0.222577	0.158139	0.268642
min	0	0	0	0	0	0	0	0	0
25%	0.299972	0.329325	0.327705	0.288225	0.351017	0.263034	0.088744	0.363819	0.254372
50%	0.415284	0.554707	0.500645	0.51226	0.508361	0.500014	0.218649	0.47173	0.414611
75%	0.557725	0.754041	0.641293	0.708568	0.675847	0.736994	0.386198	0.557895	0.668954
max	1	1	1	1	1	1	1	1	1

Table 3.13: Scaled derived performance metrics of Batting Scorecard (Ball by Ball)

	WTS	**CS**	**ES**	**BBI**	**BWTS**	**SPS**	**MI**	**EI**	**AVMS**
count	390	390	390	390	390	390	390	390	390
mean	0.234635	0.575683	0.557899	0.53416	0.082843	0.312471	0.249953	0.291735	0.186655
std	0.137773	0.193543	0.161202	0.192769	0.167264	0.150561	0.228202	0.190103	0.144399
min	0	0	0	0	0	0	0	0	0
25%	0.141286	0.477121	0.451984	0.403421	0	0.207501	0	0.146128	0.092284
50%	0.215108	0.583166	0.571469	0.523863	0	0.326025	0.219427	0.268835	0.149349
75%	0.31126	0.683988	0.662984	0.688352	0	0.403677	0.380253	0.414973	0.241757
max	1	1	1	1	1	1	1	1	1

Table 3.14: Scaled derived performance metrics of Bowling Scorecard

	ES	**WTS**	**BBI**	**CS**	**BWTS**	**SPS**	**DBI**	**EI**	**AVMS**	**BI**
count	308	308	308	308	308	308	308	308	308	308
mean	0.511905	0.226873	0.454626	0.550252	0.104157	0.281104	0.544512	0.382526	0.270323	0.38906
std	0.160645	0.140436	0.184375	0.204286	0.226557	0.148918	0.196804	0.171604	0.167378	0.183091
min	0	0	0	0	0	0	0	0	0	0
25%	0.411527	0.124865	0.32968	0.47567	0	0.186753	0.430913	0.253068	0.151769	0.277423
50%	0.505409	0.213191	0.452516	0.5725	0	0.278313	0.558097	0.376574	0.248393	0.367661
75%	0.612213	0.307474	0.582588	0.680544	0	0.375521	0.675239	0.492679	0.369939	0.489557
max	1	1	1	1	1	1	1	1	1	1

Table 3.15: Scaled derived performance metrics of Bowling Scorecard (Ball by Ball)

3.3. Theoretical Framework and Its Application

This study will be using various performance metrics for batters and bowlers separately to analyse their performance and cluster them according to their metrics. But using a lot of performance metrics also brings in the curse of dimensionality. The curse of dimensionality was first mentioned by Bellman (1966) with respect to dynamic programming, but the concept is applied majorly across other fields also. The curse of dimensionality is as the number of features increases, there is an exponential increase in volume of space as well. In result, the data points become sparse due to increase in space therefore several affects take into place:

1. The distance between the data points contains less information as it becomes harder to distinguish between the points that are close and the points that are in distance. This effects the performance of algorithms like clustering as they rely majorly on distance metrics.
2. There is a chance that the model overfits as number of features increase. This in result fits the noise and disrupts any kind of pattern formed. The model will perform well when applied on the training data but when new data is inserted its performance will be unexpected.
3. Finally, visualising the data points becomes a challenge. Anything beyond three dimensions becomes difficult to interpret for humans.

Hence, reducing the dimensions of the data is important as the study will be handling data with high-dimension and some features are highly correlated with each other. The study will be using two techniques of reducing the dimensions of the data – one is the Principal Component Analysis (PCA), and the other is Autoencoders. The goal of these dimension reduction techniques is that it will reduce the number of features while maintaining the structure and information provided by the data. The reduction in dimensionality also helps in boosting clustering because it clusters using the feature space rather than the data space as mentioned by Mousavi et al. (2019). After the data's dimensionality is reduced, then different clustering algorithms can be applied on the reduced. This study will be applying two algorithms, KMeans++ Clustering and Hierarchical Clustering with Ward's Linkage

3.3.1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a technique which reduces the dimension of the data into a new coordinate system. The coordinate system is also known as the principal component. The first principal component captures the maximum variance of the data, which means it retains the maximum information of the data. Then the components after it retains more information resulting in cumulative variance to increase as the components increase in addition to remaining orthogonal to previous components as discussed by Joliffe and Morgan (1992). There are several steps involved for the original data to be reduced into the new space, which are as follows:

- **Step 1: Standardisation:**

  One of the most important steps, and the first one, is to standardise the data matrix $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{n*p}}$, where n is the number of samples and p is the number of features. The goal of standardisation is to make sure that there is equal contribution of each feature in the analysis:

$$\boldsymbol{X}_{std} = \frac{X - \mu}{\sigma}$$

Equation 1: Standard deviation

  μ is the mean vector of the numerical data, and the σ is the standard deviation of the vector. It is important to standardize the data because PCA is sensitive to the scales of the input features.

- **Step 2: Covariance Matrix Calculation**

  To understand the relation between the features of the data, the covariance matrix $\boldsymbol{C} \in \mathbb{R}^{\boldsymbol{p*p}}$ is calculated. The following is the covariance matrix:

$$C = \frac{1}{n-1} X_{std}^{T} X_{std}$$

Equation 2: Covariance matrix

  The matrix above provides the covariance for each pair of features. The diagonal elements represent the variance of each feature whereas the elements in off-diagonal represents the covariance for each pair of features.

- **Step 3: Eigen Decomposition**

  To take out the principal components, eigen decomposition is performed on the above calculated covariance matrix. This step includes calculating the eigenvalues λ_i and eigenvectors v_i of C.

$$Cv_i = \lambda_i v_i$$

Equation 3: Eigen decomposition

  These eigenvectors correspond to the largest eigenvalues which is the direction of maximum variance in the data. The eigenvectors are how the new coordinate system will be constructed, and data projected on it.

- **Step 4: Data Projection**

  The last step consists of the standardised data being projected on the selected eigenvectors which will gives a reduced-dimensionality data matrix Z. This is calculated in the following way:

$$Z = X_{std}W$$

Equation 4: Data projection

  W represents the matrix of selected eigenvectors. Z consists of data, which is displayed in the new coordinate system, with reduced dimensions. The use of PCA is effective when more than 80% variance of the data can be captured in few dimensions, this helps in the analysis process such a clustering and assists in visualising the data in reduced dimensions.

When applying PCA to our datasets, the first step is to find the cumulative variance of all the components. The cumulative variance will help us choose the number of components we would need to use when performing our analysis. The threshold of cumulative variance in this study is 80% or more (0.80 or more), therefore as soon as the cumulative variance passes the 80% mark, we select the number of principal components that was needed to pass the 80% mark. Using '*Sklearn'* library in Python we applied PCA to our dataset.

Batting Scorecards:

When applying PCA to both of our Batting Scorecards, the following are the cumulative variance:

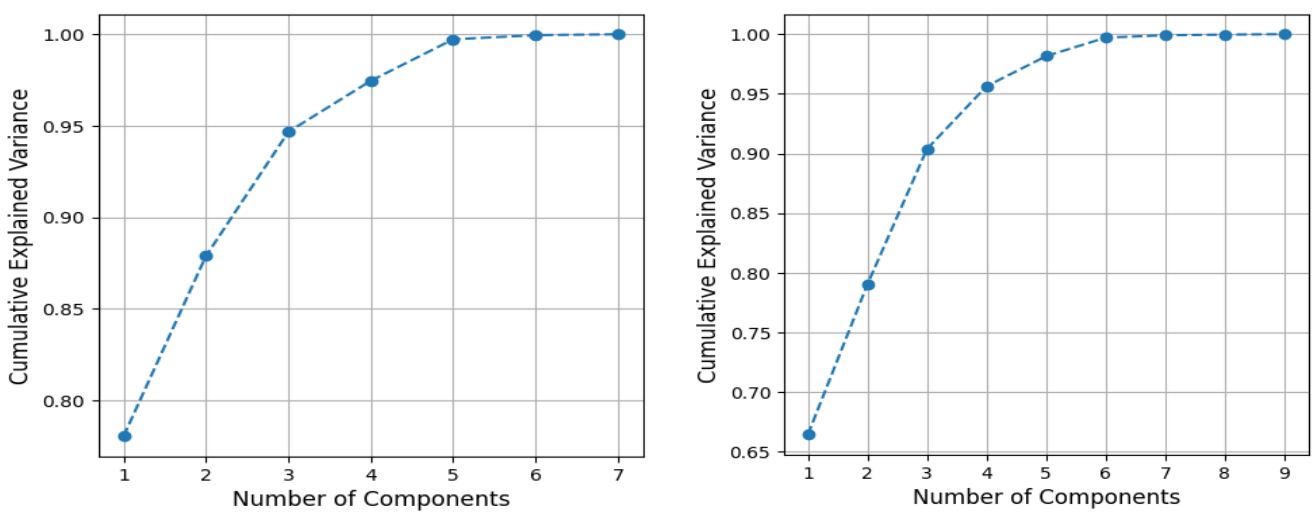


Figure 3.7: Cumulative variance of Batting Scorecard (left) and Batting Scorecard ball by ball (right)

Figure 3.7 shows that both the scorecards require different number of principal components to explain 80% of the data's variance. For the Batting Scorecard, we would need only 2 principal components whereas the ball-by-ball Batting Scorecard would require 3 principal components. This means that the Batting Scorecard's dataset has been reduced to 2 dimensions, while maintaining most of the information and to 3 dimensions for the ball-by-ball's Batting Scorecard's dataset.

It is important to find out which derived performance metric influences each principal components because this will give us insights on which derived performance metrics, in combination, made the clusters to form and characterise the players. Hence, we checked the loading vectors of each derived performance metrics on each dimension. We plotted heatmap which helped us determine the derived performance metrics influence on different principal components.

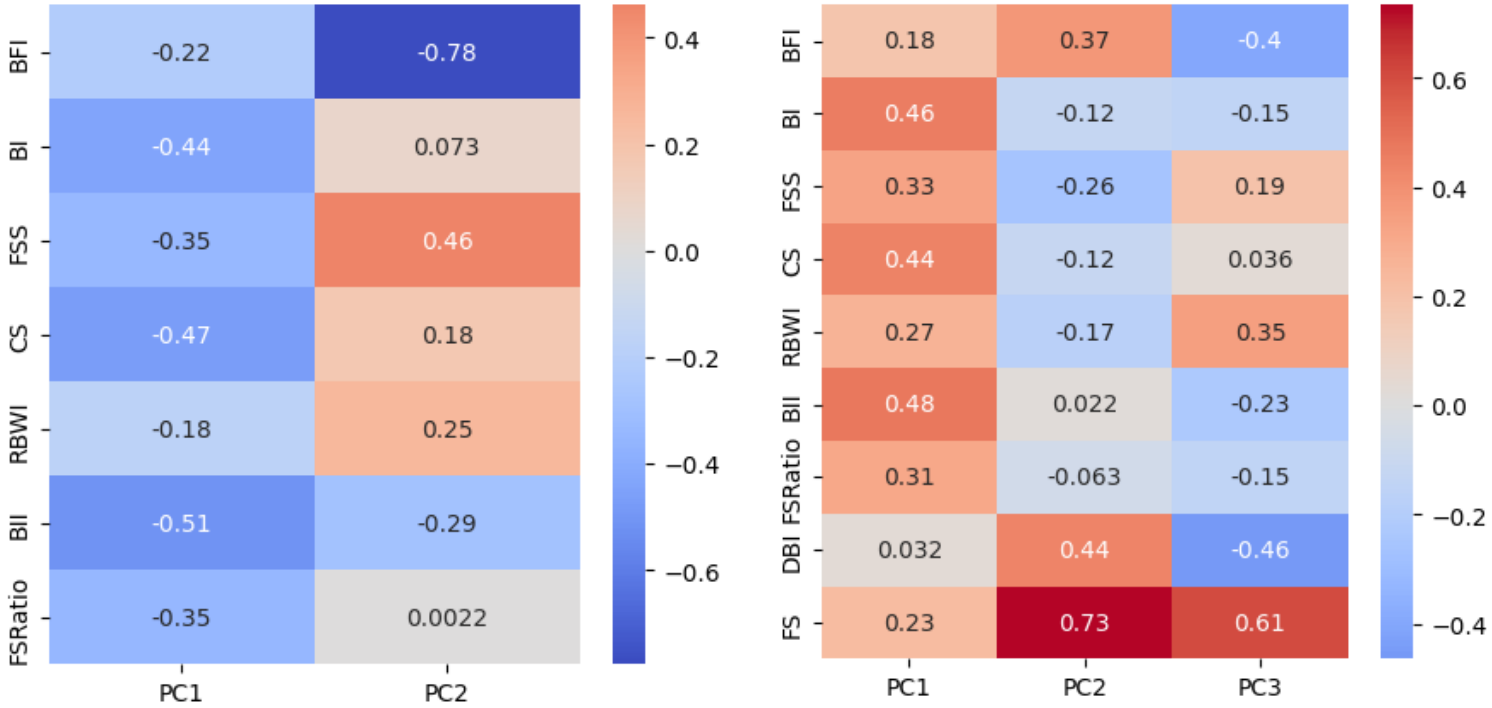


Figure 3.8: Loading vectors using heatmap to see which derived performance metric influences which principal components. On the left is the heatmap for Batting Scorecard and the right is the heat map for ball-by-ball data's Batting Scorecard

Figure 3.8 shows that the first principal component of both the Batting Scorecards is majorly influenced by the Boundary Index (BI), Consistent Scorer (CS), Big Innings Index (BII) and mildly

dominated by Balls Faced Index (BFI), Fast Scorer Score (FSS) and Fours to Sixes Ratio (FSS), the only difference being that in the Batting Scorecard it influences negatively with Principal Component 1 (PC1) which means as these values increases the PC1 would decrease significantly whereas in ball-by-ball's Batting Scorecard it is the opposite i.e. these variables have a positive influence on PC1. One can say that PC1 would separate players who are hard hitters i.e. scores most of the runs through boundaries.

The PC2 in Batting Scorecard haves a higher loading of Balls Faced Index (BFI) negatively and FSS positively – this can be interpreted as batters on this would ideally be separated on how many balls they faced and how quickly they scored facing those balls. Whereas, in ball-by-ball's Batting Scorecard PC2 is heavily influenced by Finisher Score (FS) and Dot Balls Index (DBI) and mildly by BFI and FSS.

For PC3 in ball-by-balls' Batting Scorecard most of the data is influenced by FS positively and DBI and BFI negatively. Whereas there is a mild loading of Running Between Wickets Index (RBWI) positively and BII negatively. As mentioned above, all these values of loading vector suggests that if the cluster is performed in one of the PCs, it will be due to those derived performance metrics who have high or mildly high loading vectors in that very PC.

Bowling Scorecards:

When applying PCA to both of our bowling scorecards, the following are the cumulative variance explained:

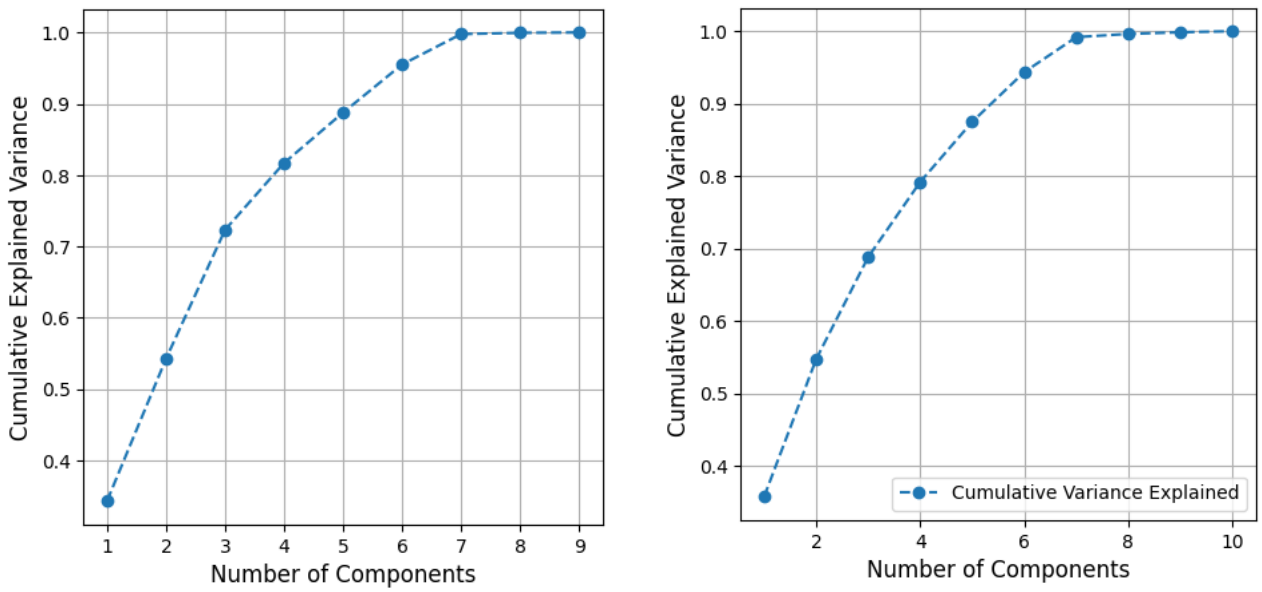


Figure 3.9:Cumulative variance of Bowling Scorecard (left) and Bowling Scorecard ball by ball (right)

As seen from the above diagrams, for the Bowling Scorecard at PC4 can explain more than 80% of the data and for the ball-by-ball's Bowling Scorecard crosses the 80% threshold at PC5. Hence, dimensions have been reduced to 4 for the Bowling Scorecard and to 5 for the ball-by-ball's Bowling Scorecard.

Figure 3.10 shows the influence of derived performance metrics on the principal component dimensions which will help us define the characteristic of the clusters once formed in these dimensions.

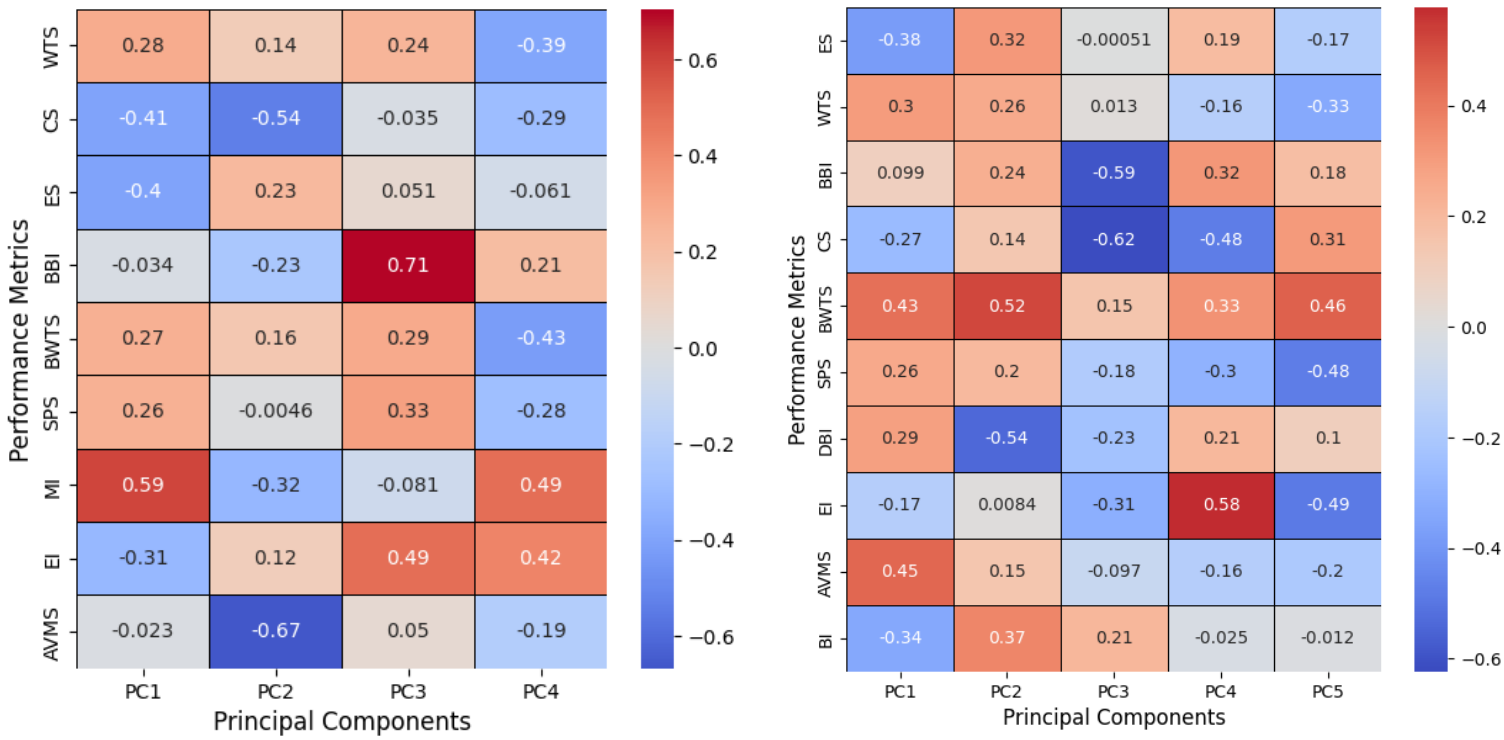


Figure 3.10: On the left is the loading vector heatmap of Bowling Scorecard and on the right is loading vector heat map of ball-by-ball Bowling Scorecard's loading vector heat map. These heat map shows which derived performance metric influences which dimensions

3.3.2. Autoencoders

Autoencoders, as discussed by Hinto and Salakhutdinov (2006), are a type of neural networks which has a network architecture of an encoder, a bottleneck also known as latent space, and a decoder. The encoder is provided the input data which it compresses into a lower-dimensional representation in the latent space, and then the decoder reconstructs the input using the lower-dimensional representation. The activation function used in this study is Leaky ReLU.

The Leaky ReLU, ReLU standing for Rectified Linear Unit, brings in a small slope for inputs that are negative, which solves the issue of "dying ReLU" in which there are chances of neuron becoming inactive and not learn anymore. The Leaky ReLU is as follows:

$$LeRLU(x) = \max(\alpha * x, x)$$

Equation 5: Calculation of LeakyReLU

Over here the α is a small positive constant usually around 0.01. This α assists in rectifying the negative inputs. The above function keeps the network in a learning phase even if the input to a

neuron is negative, in other words the neurons do not become inactive during the training process.

As mentioned above, the encoder compresses the input data $X \in \mathbb{R}^{n*p}$ into a lower-dimensional representation $h \in \mathbb{R}^{n*d}$. p is the original dimension and d is the reduced dimension. Using the Leaky ReLU function, the following is the way transformation done:

$$h = LeRLU(XW_e + b_e)$$

Equation 6: Lower dimensional transformation of original data by encoder

Where, W_e is the weight matrix and b_e is the bias vector. As mentioned above, Leaky ReLU makes sure that the network is in a continuous learning phase even if some inputs are negative.

Once the lower-dimensional representation is in the latent space, the decoder then reconstructs the original data

$$\hat{X} = LeRLU(hW_d + b_d)$$

Equation 7: Reconstruction of the lower-dimensional data to original data by the decoder

In the above equation, W_d is the weight matrix of the decoder and b_d is the bias vector. The main goal of the decoder is to reconstruct the input data as closely as possible in a reduced dimension.

During the training process, the model is trained by minimizing the reconstruction error. The reconstruction error is usually calculated using the mean squared error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left\|X_i - \hat{X}_i\right\|^2$$

Equation 8: Reconstruction error which is calculated using mean squared error formula

The MSE is the measurement of the average squared different of the original input X_i and the reconstructed output $\hat{X}_i$. The goal of minimizing the error is to make sure that maximum amount of information is retained when the data's dimensionality is reduced. Using '*Tensor'/'Keras'* library in Python we applied the autoencoding to our datasets.

Batting Scorecards:

Before applying autoencoders to our dataset, it is important to take out the optimal number of dimensions that data need to be reduced to. This will be the dimension having the lowest reconstruction error. Figure 3.11 show the reconstruction error and its relationship with different number of dimensions for both batting scorecards:

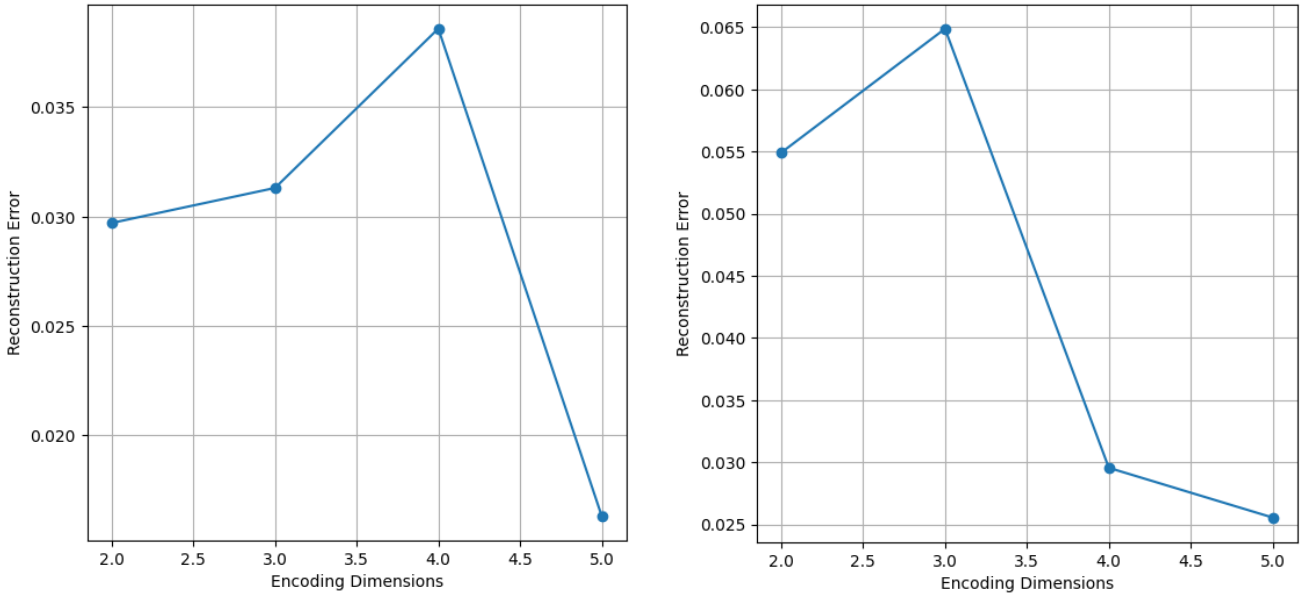


Figure 3.11: Relationship between reconstruction error and number of encoded dimensions. On the left is for the Batting Scorecard and on the right for ball-by-ball Batting Scorecard.

Figure 3.11 show that the lowest reconstruction error for the Batting Scorecards is when the encoding dimensions is 5. At these dimensions, the autoencoder will retain majority of information in the data when reducing the dimensions.

For the autoencoder, we used 7 layers including the input and output layers for Batting Scorecard and 9 layers for ball-by-ball Batting Scorecard, with each layer having LeakyReLU as its activation function. Figure 3.12 gives a representation of how the layers are like for the Batting Scorecard:

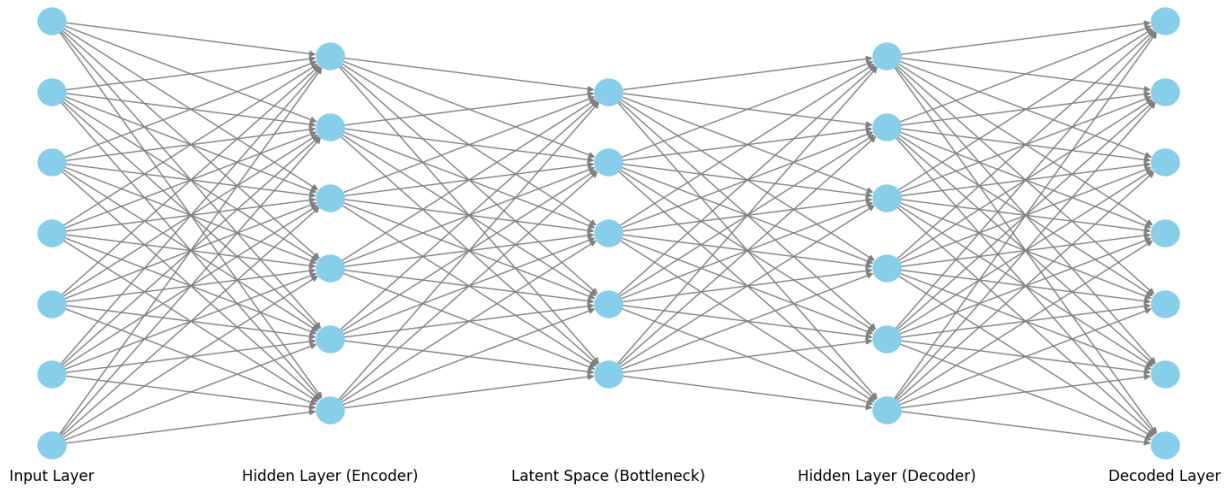


Figure 3.12: Autoencoder for the Batting Scorecard

When doing our clustering analysis on these encoded dimensions, it is important to know which derived performance metric is influencing the encoded dimension. Figure 3.13 below shows the influence of each derived performance metric for both batting scorecards:

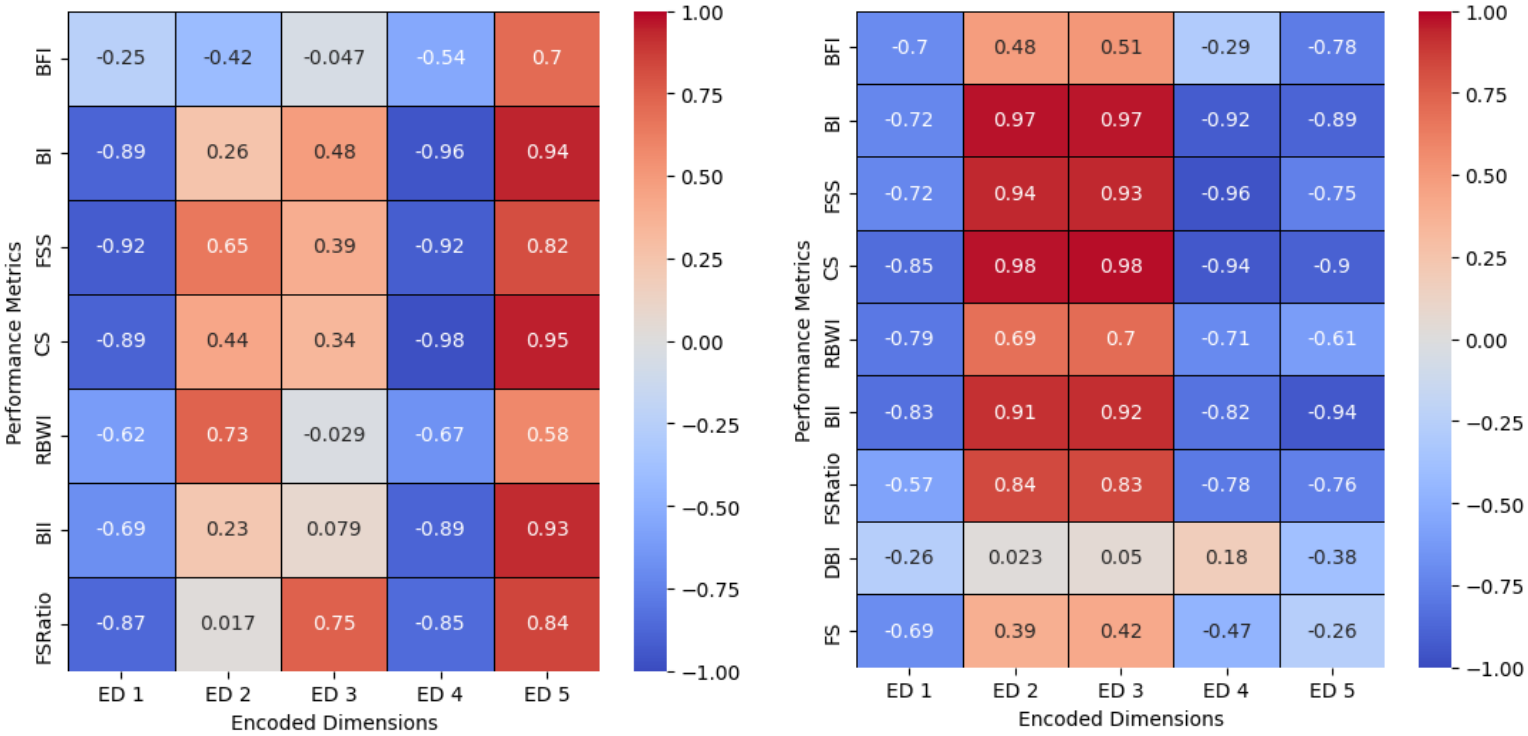


Figure 3.13: Derived performance metric influence on each encoded dimension for Batting Scorecard (left) and ball-by-ball Batting Scorecard (right)

The explanation of the above heat map is like the PCA loading vector's heat map. Each encoded dimension shows which derived performance metric influences the data points in it if any clusters are formed. More of this will be mentioned after clustering analysis.

Bowling Scorecard:

As done with batting scorecard, we perform the same steps with both bowling scorecards. First, we search for the optimal number of reduced encoded dimensions by analysing the reconstruction error. Figure 3.14 give details for both bowling scorecard's reconstruction error in relation to encoded dimensions:

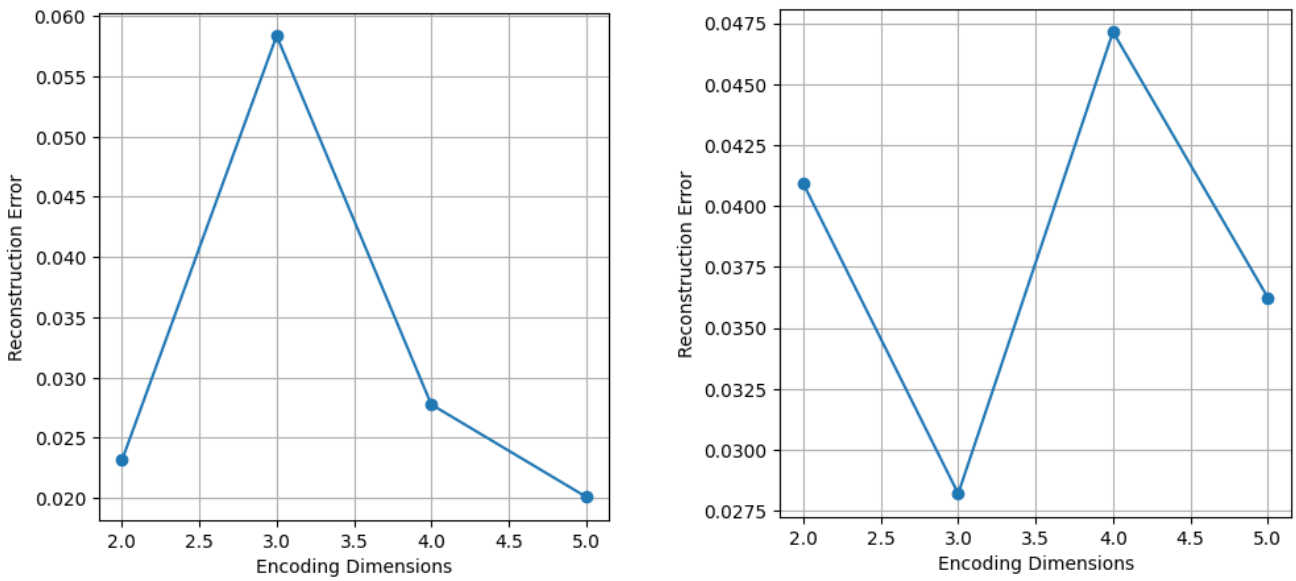


Figure 3.14: Relationship between reconstruction error and number of encoded dimensions. On the left is for the Bowling Scorecard and on the right for ball-by-ball Bowling Scorecard

Figure 3.14 shows the reconstruction error is at minimum for bowling scorecard when the derived performance metrics are encoded to 5 dimensions and for ball-by-ball Bowling Scorecard it is at 3 dimensions. Both the scorecards had autoencoders with 9 layers, like ball-by-ball Batting Scorecard. Figure 3.15

shows the influence of derived performance metrics on each encoded dimension:

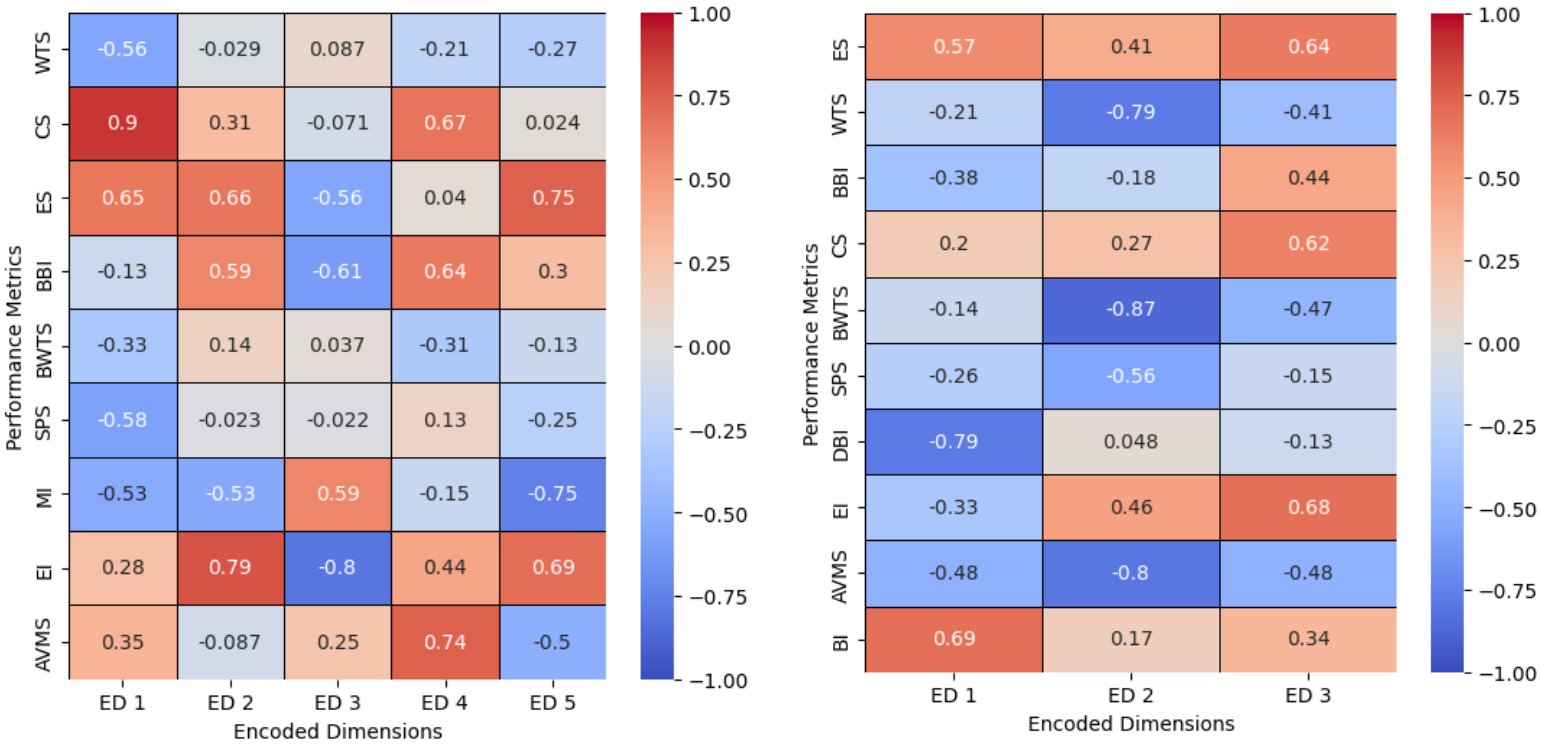


Figure 3.15: Derived performance metric influence on each encoded dimension for Bowling Scorecard (left) and ball-by-ball Bowling Scorecard (right)

3.3.3. KMeans++ Clustering

The most common use of clustering method is KMeans Clustering, but this study will be using a much more enhanced version of KMeans Clustering i.e. KMeans++ clustering. This is due to the algorithm's optimisation in selecting the initial cluster centroid which in result leads to a more accurate distinct cluster. The process of clustering in KMeans++ algorithm is minimizing sum of squared distances between data points and their respective cluster centroids as mentioned by Arthur and Vassilvitskii (2007) in their research. By doing so, the dataset is separated into k clusters.

The clusters in KMeans are created using within-cluster sum of squares (WCSS), which is defined as:

$$\emptyset(C,\mu) = \sum_{i=1}^{k} \sum_{x_j \in C_i} \left\| x_j - \mu_i \right\|^2$$

Equation 9: Within-cluster sum of squares (WCSS)

The main objective of KMeans algorithm is to minimize WCSS to form clusters. In Equation 9, C_i are the data points in the ith cluster, μ_i is the centroid of cluster C_i and x_j is the individual data point. The WCSS measures how similar the cluster is within, smaller WCSS means the points in the cluster are like each other.

One of the major aspects of KMeans++ algorithm is the selection of initial centroid. In KMeans the first centroid is chosen at random and then succeeding centroids are selected using probability that is proportional to the squared distance from the nearest already selected centroid. The following equation is used for it:

$$p(x) = \frac{D(x)^2}{\sum_{x' \in X} D(x')^2}$$

Equation 10: Probability formula for selection of centroid

In Equation 10 the $D(x)$ is distance of point x to the nearest centroid. The goal is for the initial centroids to be spread away effectively which can lead to better and more compact clusters.

Then the algorithm assigns each data point to the closest centroid, after which it recalculates the centroids. The new centroid is calculated by averaging all the points assigned to the cluster using the following formula:

$$\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j$$

Equation 11: Equation to calculate the new centroid

The process iterates until there is no meaningful change in the centroids which means that clusters have been created.

For the KMeans++ algorithm to work, it must be given the number of clusters. To find the optimal number of clusters, this study will use two methods. The reason of using two methods is to validate the optimal number of clusters to be used. The first method is the Elbow Method, and the second method is the Silhouette Score Method.

The Elbow Method is used by plotting WCSS, as calculated using Equation 9, against the number of clusters. The term elbow comes from the point where the decreasing rate of WCSS is sharply

reduced, hence that point is what is to be considered as having optimal number of clusters. After this point, the rate of decrease is extremely slow hence indicating that the optimal number of clusters have been achieved.

The second method to find the optimal number of clusters for KMeans++ is the Silhouette Score. It is a measure of how each data point is suitably fitted in its assigned cluster. The Silhouette Score is calculated in the following way:

$$SS = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Equation 12: Silhouette Score

$a(i)$ is the intra-cluster distance i.e. the average distance between point i and the other data points in the same cluster. Whereas $b(i)$ is the minimum inter-cluster distance i.e. the minimum average distance between point i and all the data points in the other cluster. The higher the silhouette score, much better defined the clusters will be hence we will choose the optimal number of clusters which will have the highest silhouette score.

Batting Scorecards (PCA and Autoencoder):

As mentioned above, before applying KMeans++ clustering algorithms it is important to identify how many clusters do we go ahead with. Hence, Figure 3.16 show the plots of the Elbow Method, using '*Sklearn*' library in Python, when applied to the Principal Components from PCA and encoded data from autoencoders:

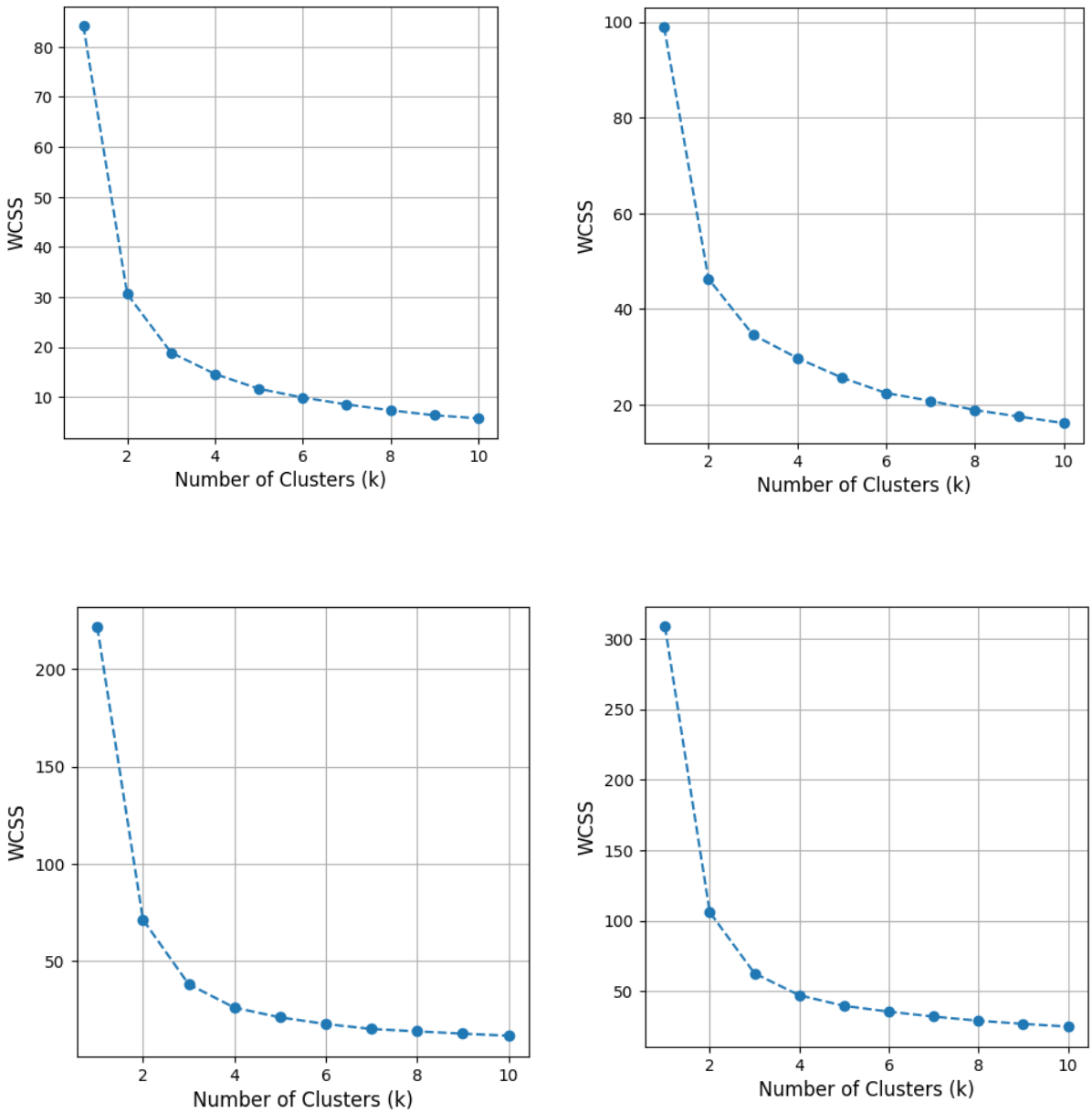


Figure 3.16: Finding the optimal cluster of batting scorecards using elbow method. Batting Scorecard (PCA) on the top left, Batting Scorecard (Autoencoder) on the bottom left, ball-by-ball Batting Scorecard (PCA) on top right and ball-by-ball Batting Scorecard (Autoencoder) on the bottom right.

The above plots show that all the Batting Scorecards have optimal number of cluster either 2 or 3 as can be seen an 'elbow' being formed in those coordinates. It is not noticeably clear on whether to choose 2 or 3 as optimal clusters, hence the use of Silhouette Score will ensure us on how many optimal numbers of clusters should be taken in KMeans++ Clustering. Figure 3.17 shows the plots for all Batting Scorecard's silhouette score:

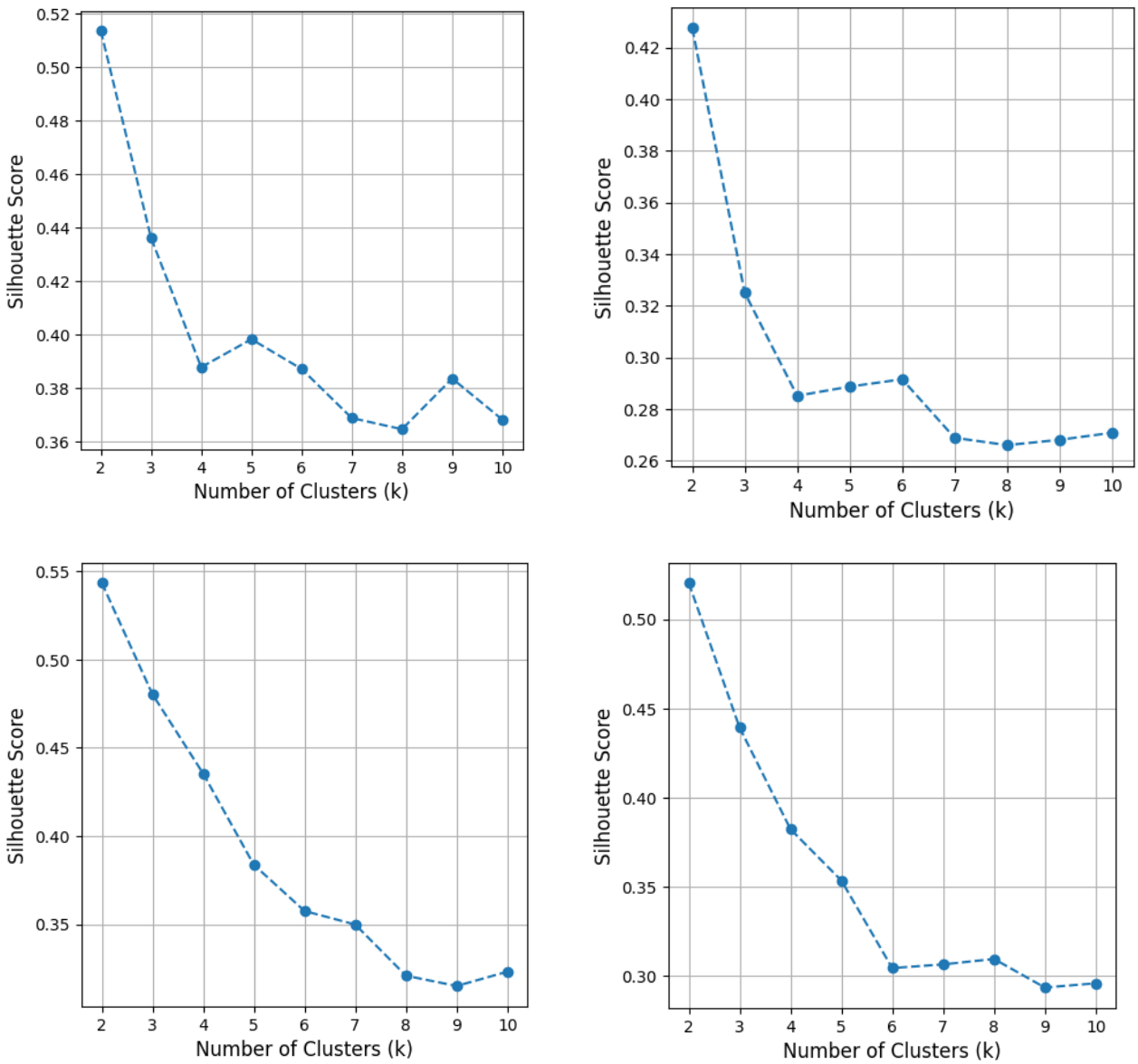


Figure 3.17: Silhouette Scores to find the optimal number of clusters for Batting Scorecard (PCA) on the top left, Batting Scorecard (Autoencoder) on the bottom left, ball-by-ball Batting Scorecard (PCA) on top right and ball-by-ball Batting Scorecard (Autoencoder) on the bottom right

The Silhouette Scores gives a confirmation that the optimal number of clusters to be used for all the Batting Scorecard is 2. We move ahead with applying 2 clusters to our KMeans++ algorithm, again using '*Sklearn*' library. The clusters were plotted in a pair plot to identify which principal components and encoded dimensions are separating the clusters. The cluster made on principal component dataset can be viewed in Appendix B Figure 1 and clusters made on the encoded dimensions can be viewed in Appendix B Figure 2 for Batting Scorecard and Appendix B Figure 3 for ball-by-ball Batting Scorecard.

Bowling Scorecards:

Applying the same methods to the Bowling Scorecards, we first find the optimal number of clusters using the Elbow Method and Silhouette Score and then we use those optimal number of clusters to in the KMeans algorithm. The elbow point and Silhouette Score can be viewed on Appendix B Figure 4 and Appendix B Figure 5 respectively. From the Elbow Method it is not clear what the optimal number of clusters is hence using the highest Silhouette Score we can observe that the principal component of Bowling Scorecard has optimal number of clusters as 3, the principal components of ball-by-ball Bowling Scorecard has optimal number of clusters of 4 and both encoded dimension scorecards has optimal number of clusters as 2. Like Batting Scorecards, these clusters were applied to KMeans algorithm, and the plot of the results can be seen in Appendix B Figure 6, Appendix B Figure 7, Appendix B Figure 8, and Appendix B Figure 9. In detail discussion about the characteristics of these clusters will be done in the Results section.

3.3.4. Hierarchical Clustering

As discussed by Bridges (1966), Murtagh and Contreras (2012), hierarchical clustering is tree-like clustering method, also known as a dendrogram. The study will use agglomerative hierarchical clustering with Ward's linkage.

The agglomerative hierarchical clustering method starts with each data point considered as an individual cluster. Then two clusters that have the smallest increase of the within-cluster variance merges together. This process goes on until we are left with one cluster.

The clustering process will happen with the use of Ward's linkage. Ward's linkage is the method which checks on minimizing the increase in variance whenever clusters are merged. The following method is used:

$$d_{Ward}(A, B) = \frac{|A| * |B|}{|A| + |B|} \|\mu_A - \mu_B\|^2$$

Equation 13: Ward's linkage

A and B are two clusters - μ_A and μ_B are there centroids respectively and $|A|$ and $|B|$ are how big these clusters are. Ward's linkage main objective is to produce clusters of same size and shape.

Then the clusters can be visualised using a dendrogram which shows where and how are clusters merged. Dendrogram's distance between the clusters can help select the optimal number of clusters.

To choose the optimal number of clusters using dendrogram, we select the largest vertical distances between horizontal lines that connect clusters. The large vertical distances indicate a natural separation between clusters. Then we draw a horizontal line when we have selected the largest vertical distance, this can be known as threshold. The optimal number of clusters will be the vertical lines individually separated by this line.

Batting Scorecard:

The method to find the optimal number of clusters is simple; we plot a dendrogram, find the largest vertical distance, cut it horizontally and see how many branches or optimal number of clusters are coming from that horizontal line. We used '*Scipy*' library to create ward linkage and using that Ward linkage we made the dendrogram whose figures can be viewed in Appendix B. The following table summarises the dendrogram:

Scorecard Name	**Vertical Distance where cut was made**	**Number of Clusters formed**	**Figure Number**
Batting Scorecard (PCA)	4.5	2	Appendix B Figure 10
Ball-by-Ball Batting Scorecard (PCA)	4.5	2	Appendix B Figure 11
Batting Scorecard with Encoded Dimensions	6	3	Appendix B Figure 12
Ball-by-Ball Batting Scorecard with Encoded Dimensions	12	2	Appendix B Figure 13

Table 3.16: Summary of dendrograms plotted and clusters formed of all the Batting Scorecards

From the above table, we can see that other than the Batting Scorecard with Encoded Dimensions, all other Batting Scorecards' dendrogram is giving optimal number of clusters as two. Cluster formed in different principal components can be seen in Appendix B Figure 18 and Appendix B Figure 19. For clusters formed in encoded dimensions it can be seen in Appendix B Figure 20 and Appendix B Figure 21. The characteristics of these clusters will be discussed in detail in the Results section.

Bowling Scorecard:

The summary of dendrogram for Bowling Scorecards is given below:

Scorecard Name	Vertical Distance where cut was made	Number of Clusters formed	Figure Number
Bowling Scorecard (PCA)	4	3	Appendix B Figure 14
Ball-by-Ball Bowling Scorecard (PCA)	4	4	Appendix B Figure 15
Bowling Scorecard with Encoded Dimensions	10	2	Appendix B Figure 16
Ball-by-Ball Bowling Scorecard with Encoded Dimensions	6	3	Appendix B Figure 17

Table 3.17: Summary of dendrograms plotted and clusters formed of all the Bowling Scorecards

The dendrograms of Bowling Scorecards show interesting clustering summary, other than Bowling Scorecard (PCA) and ball-by-ball Bowling Scorecard with Encoded Dimensions, no other Bowling Scorecard is forming same number of clusters. The clusters are plotted in different principal components in Appendix B Figure 22 and Appendix B Figure 23 and for the encoded dimensions the plots can be seen in Appendix B Figure 24 and Appendix B Figure 25. What characteristics each cluster upholds will be discussed in detail in the next section.

4. Results

This section will identify the characteristics of clusters formed on different principal components and encoded dimensions. The characteristics of clusters will define what kind of players are in that cluster. Furthermore, using the above derived performance metric heatmaps, this section will also find which derived performance metrics are key in clustering the data. At the end, a composite ranking system will be developed using the scaled data which will rank the players in the high performing clusters.

Batting Scorecards:

As seen from the above section and in Appendix B Figure 1, Appendix B Figure 2, Appendix B Figure 3 each Batting Scorecard had 2 clusters when KMeans++ is performed on the datasets. At

first, we will look at what derived performance metrics influenced in creating these clusters for all the Batting Scorecards.

Looking at the Batting Scorecard (PCA) KMeans++ Clustering Appendix B Figure 1, we see clearly that the data is clustered due to the Principal Component 1 (PC1), hence it will be interesting to see PC1's derived performance metric heatmap. From Figure 3.2, we can see that PC1 is heavily influenced by BI, CS and BII and mildly influenced by FSS and FSRatio. This means that the cluster that is formed have separated the players who consistently scores a lot of boundaries – and due to them scoring a lot of boundaries they are able to have a high FSS and FSRatio as well. Hence, these derived performance metrics are the ones that create these clusters. In the two clusters created, the following are the means summary of these clusters:

KCluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio
0	0.548336	0.666733	0.570069	0.684264	0.462938	0.707607	0.418363
1	0.361147	0.273395	0.25354	0.256005	0.297184	0.251591	0.106965

Table 4.1: KMeans++ clusters mean summary for Batting Scorecard (PCA)

From Table 4.1 we can differentiate the high performing cluster from the low performing cluster. Cluster 0 is the cluster which contains the player that perform well compared to players in Cluster 1 as the mean of most of the derived performance metric is high for Cluster 0 than Cluster 1. Out of the 269 batters we had, 149 batters were assigned to Cluster 0.

The result of hierarchical clustering was also similar, with PC1 creating the clusters, as can be seen in Appendix B Figure 18. The means summary of the clusters is in Appendix C Table 1. The only minor difference was that 143 batters were allotted to the high performing cluster and 126 to the low performing one.

Moving on towards, ball-by-ball Batting Scorecard (PCA), it also shows in Appendix B Figure 1 that PC1 is the principal component that is creating the clusters. PC1 in this dataset is also heavily influenced by the same derived metrics as the ones in Batting Scorecard (PCA). Looking at the means summary for this dataset, we get the following:

KCluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio	DBI	FS
0	0.511704	0.734251	0.629072	0.70205	0.63436	0.700468	0.404555	0.480976	0.514042
1	0.336719	0.291316	0.306619	0.264649	0.36194	0.228121	0.098228	0.453388	0.315488

Table 4.2: KMeans++ clusters mean summary for ball-by-ball Batting Scorecard (PCA)

As was the case with Batting Scorecard, this means summary also shows that Cluster 0 is the high performing cluster compared to Cluster 1. The number of players assigned to Cluster 0 are 119 and to Cluster 1 are 103.

Again, like K-Means++, hierarchical clustering on the ball-by-ball Batting Scorecard (PCA) produced the same result as shown in Appendix B Figure 19. The only difference was the high performing cluster had 105 batters in it whereas the other cluster had 117. The means summary table can be seen Appendix C Table 2.

Similar procedure is followed for the Batting Scorecard with encoded dimensions. Separate clusters are created in Encoded Dimension (ED) 1, ED4 and ED5 as seen in Appendix B Figure 2. The derived performance metrics that influences these dimensions are the same as above i.e. BI, FSS, CS, FSRatio and BII. Appendix C Table 3 gives means summary of the clusters show that Cluster 0 having higher means in derived performance metric than Cluster 1. Hence, Cluster 0 contains high performing players. The number of players in Cluster 0 are 147 and in Cluster 1 are 122 – which is like KMeans++ on Batting Scorecard (PCA).

For the hierarchical clustering on the encoded dimensions of Batting Scorecard, it created 3 clusters, and these 3 clusters were differentiated by ED4 and ED5 unlike KMeans++ which also had ED1 to differentiate the clusters. The only difference between the derived performance metrics influencing ED1 from ED4 and ED5 was the BII. BII has a mild influence in ED1 and high influence in ED4 and ED5. Hence, the hierarchical algorithm separated the players who had a lot of big innings. The mean summary of the clusters can be seen in Appendix C Table 4. Cluster 1 is the high performing cluster compared to other 2 clusters. Cluster 1 has 77 batters, cluster 2 has 67 batters and cluster 3 has 125.

When it comes to applying KMeans++ algorithm on ball-by-ball Batting Scorecard with encoded dimensions, separate clusters were created in all the EDs as can be seen in Appendix B Figure 3. It is not surprising because all the EDs are influenced by the same derived metrics as above. Taking out the means summary of both clusters, from the Appendix C Table 5 we can see that Cluster 0 is the one with having higher means in derived performance metrics than Cluster 1. And just like ball-by-ball Batting Scorecard (PCA), Cluster 0 has 117 players and Cluster 1 has 105

players. Hierarchical clustering had a similar result, but the cluster was much more sophisticated. Cluster 1, having high performing batters, had 77 batters in it whereas cluster 2 had 150.

To rank the batters in the clusters, a composite score on their scaled data was calculated using the following formula:

$$CS = \sum_{i=1}^{n} w_i * x_i$$

Equation 14: Composite score calculation to rank the players

Where n is the number of total metrics, x_i is the value of the ith metric, and w_i is the weight assigned to the ith metric. The composite score of each player in the high performing cluster is calculated and they are ranked according to their composite score, the player having the highest composite score being the first and the lowest one being the last.

For the Batting Scorecards, the top 5 players are given below, and a complete list of ranked players are given in Appendix C Table 6 to Appendix C Table 13.

	Batting Scorecard			
	PCA			
	Kmeans++		**Hierarchical**	
Ranking	**Player**	**Team**	**Player**	**Team**
1	C Dyer	Littletown CC Under 13	C Dyer	Littletown CC Under 13
2	Oliver Jackson	Castle Eden CC Under 13	Oliver Jackson	Castle Eden CC Under 13
3	Alfie Blenkiron	Bishop Auckland CC Under 13	Alfie Blenkiron	Bishop Auckland CC Under 13
4	C McCarthy	Dawdon Welfare CC Under 13	C McCarthy	Dawdon Welfare CC Under 13
5	Manas Chengalavala	Durham City CC Under 13 Blue	Manas Chengalavala	Durham City CC Under 13 Blue
	Ball-by-Ball Batting Scorecard			
	PCA			
	Kmeans++		**Hierarchical**	
Ranking	**Player**	**Team**	**Player**	**Team**
1	C Dyer	Littletown CC Under 13	C Dyer	Littletown CC Under 13
2	Oliver Jackson	Castle Eden CC Under 13	Oliver Jackson	Castle Eden CC Under 13
3	O Henderson	Chester Le Street CC Under 13	O Henderson	Chester Le Street CC Under 13
4	Oliver Jackson	Durham City CC Under 13 Blue	Oliver Jackson	Durham City CC Under 13 Blue
5	K Russell	Easington C.W. CC Under 13	K Russell	Easington C.W. CC Under 13

Table 4.3: Ranking of player after applying composite score formula in high performing cluster

Bowling Scorecard:

The procedure of analysing the characteristics of the clusters is the same as above.

When KMeans++ algorithm was applied to Bowling Scorecard (PCA), it can be observed from Appendix B Figure 6 that PC1 defines the 2 of the clusters and PC2 separates the 3rd cluster from the other two. From Figure 3.10 PC1 is majorly influenced by Consistent Score (CS), Economy Score (ES) and Maiden Index (MI). This means that the 2 clusters that are separated in PC1 are dependent on how often a bowler bowls a maiden over and how many runs the bowler concede. Whereas the 3rd cluster which is separated because of PC2 is influenced not only by CS but also by Average Match Score i.e. having a high PC2 score means having a low AVMS. Hence the 3rd cluster has a low AVMS compared to other 2 clusters. When taking out the means summary of all 3 clusters (see Appendix C Table 14), we find Cluster 0 as the high performing one as the mean of derived performance metrics is much better (not higher as some derived performance metric having a lower score means a better score such as ES) than the other 2 clusters. And as expected, the AVMS of cluster 0 and cluster 2 are the same hence cluster 1 is the cluster created by PC2. The number of bowlers in Cluster 0 are 151, Cluster 1 has 15 and Cluster 2 has 224. The hierarchical clustering of this data is also similar, 3 clusters were created, and 2 clusters separated in PC1. Cluster 1 is the high performing one with a better means summary (see Appendix C Table 15). Cluster 1 has 154 bowlers; Cluster 2 has 15 and Cluster 3 has 221.

Moving on, applying KMeans++ algorithm to Bowling Scorecard with encoded dimensions, we can see from Appendix B Figure 8, clusters are created in ED2, ED3 and a bit of ED5. Looking at Figure 3.15, we can see that ED2 and ED3 are heavily influenced by ES, Balls Bowled Index (BBI), Extras Index (EI) and MI. Whereas as ED5 is additionally slightly influenced by AVMS. Hence, we can see that clusters are majorly made on the type of ball the bowlers bowls, whether they are bowling extras or not and are the amount of runs they concede. Moving on towards mean summary of clusters, all the derived performance metrics have the same mean except ES, MI and EI which is lower in Cluster 0 this means Cluster 0 is the high performing cluster. Cluster 0 had 207 bowlers allotted to it and Cluster 1 had 183. Likewise, is the case for hierarchical clustering on the same encoded data, 2 clusters were formed on the same EDs hence having the same characteristics (see Appendix B Figure 24). Cluster 2, the high performing cluster, had 219 bowlers and Cluster 1 had 171 bowlers.

Looking at ball-by-ball Bowling Scorecard (PCA), when applied KMeans++ algorithm to it, the data did not form any kind of clear cluster in any of the principal components as can be seen in Appendix B Figure 7. But in hierarchical clustering a bit of clarity is shown in PC1 with the separation of clusters as can be seen in Appendix B Figure 23. When looking at the loading vector heat map of PC1, see Figure 3.10, we can see that PC1 is majorly influenced by ES, WTS, Big Wicket Taking Score (BWTS), AVMS and Boundary Index (BI). This means that players are being separated in PC1 with respect to the amount of score they concede specially through boundaries and the number of wickets they take, specially more than 3 wickets in a match. After calculating the means summary of each cluster, it could be seen Appendix C Table 16 that players in cluster 1 are differentiated from cluster 2 mainly because of BWTS and WTS. Hence, wicket taking ability of a player. Cluster 1 has 58 bowlers, cluster 2 has 21, cluster 3 has 138 and cluster 4 has 93.

When using encoded dimension dataset of ball-by-ball Bowling Scorecard, after applying KMeans++ algorithm, we can observe (see Appendix B Figure 9) that ED2 is the dimension which clearly separates the dataset into 2 clusters, whereas ED3 does not do it clearly but does some form of separation. Looking at the heatmap of derived performance metric influence on different EDs (see Figure 3.15), we observe the ED2 is heavily influenced by BWTS, WTS and AVMS. This particularly means that the clusters are formed due to the wicket taking ability of the bowler. While ED3 brings in slight differentiated clusters also, the performance metric that influences ED3 is EI, CS and ES which means that players are separated on ED3 according to how many runs the bowler concede.

After performing mean summary on clusters, we see from Appendix C Table 17 that Cluster 1 is the high performing cluster majorly because of having better means for most of the derived performance metrics specially BWTS. Meaning, players in these clusters are more likely to take 3 wickets or more in a match. Cluster 1 has 54 players, whereas cluster 0 has 254.

Moving on, when hierarchical cluster analysis is performed in this dataset, 3 clusters are created, and they are clearly separated in ED3 (see Appendix B Figure 25). This means that players in these clusters are majorly separated because of the amount of runs they have conceded. After calculating the mean summary of each cluster, we observe cluster 1 (see Appendix C Table 18) to

be the high performing one due to its derived performance metrics having a better mean compared to other clusters. There are 54 bowlers in cluster 1, 115 in cluster 2 and 139 in cluster 3.

With all the high performing clusters above, the same composite ranking formula, Equation 14, was applied to the scaled data of the players. Then the players were ranked according to their composite score with highest score being top ranked and the lowest score being bottom ranked. The complete ranking table of all the high performing clusters can be seen from

5. Discussion

The analysis above used dimensions reduction techniques such as PCA and Autoencoder and applied KMeans++ and Hierarchical algorithms to the reduced dimension dataset to create clusters of players in Junior Cricket League according to derived performance metrics. In Batting Scorecards, derived performance metrics like Boundary Index (BI), Big Innings Index (BII), Finisher Scorer (FS), Fours to Sixes Ratio (FSRatio) and Consistent Scorer (CS) influenced the clusters that were formed. This meant that players who were in the high-performing clusters had the ability to score more boundaries consistently while also scoring above average every match. Likewise, for the Bowling Scorecard, derived performance metrics like Consistent Scorer (CS), Economy Score (ES), Maiden Index (MI), Extras Index (EI), Big Wicket Taking Score (BWTS) are ones that influence in creating clusters and separating top performing players. This meant that these players had the ability to consistently concede less runs and at the same time had better ability to take 3 or more wickets in a match.

Using the clusters that were created, a composite score was calculated which created ranking among the players in the high performing cluster. The composite score was calculated using weights and the scaled derived performance metric.

The finding from this study can help Durham Cricket County scout players effectively and more broadly rather than relying on human judgement. It can help the selectors develop players and later be selected into the senior cricket team. The above derived performance metrics can help selectors and coaches to develop training sessions which particularly focuses on these areas.

Even with the cluster that did not perform well, training sessions could be developed to improve their skills to keep them intact in the game. And those players who are in high performing clusters can further strengthen their abilities helping them in further player development.

Furthermore, the ranking system can also help in selecting teams with selectors adjusting the weighting systems according to their needs for example, if they want players in a team who bowls more dot balls, they can increase the weights of derived performance metrics related to dot balls.

Even with such valuable insights, the study did have its limitations. One of the limitations is that the player's performance was more focused on the batting or bowling ability while ignoring the third aspect of the game i.e. fielding. Fielding is as important as batting and bowling (as the saying goes 'catches win matches'). The study could not take the fielding statistics of the players due to format in which the website was (which was another limitation).

Moreover, the study above can only be used in grassroot levels where the captaincy of a player or the ability to make partnership is not influential to team performance. Whereas in senior cricket, these metrics in addition to others match situation related metrics are important to perform well consistently.

Last but not the least, there were a lot of matches which did not have ball-by-ball data, hence the ball-by-ball data (even though it is most detailed one for analysis) is not as complete as the simple scorecard data. Therefore, this was the reason some of the players might have different ranking when compared with the simple scorecard.

For future research, it will be better for the cricket county to start encouraging scorers to record matches ball-by-ball as it will provide a much more detailed data. In addition, an API should be setup which could extract all kind of data easily as there is a chance that web scraping might miss out crucial data such as the fielding data which was not able to scrap from the website.

Once, the above analysis is used for few seasons, it would be better to develop advanced machine learning techniques such as Random Forest or Neural Network to provide more insights into player performance metrics.

Future studies should incorporate on how to include match condition related data such as partnerships, captaincy decisions, fielding positions etc. and how can these impact individual performances and overall team performances.

6. Conclusion

The aim of the study was to identify top-performing players in Durham Junior Cricket League, specifically under-13, by using unsupervised learning techniques such as PCA and Autoencoders to reduce the dimensions of the data and form clusters using KMeans++ and Hierarchical algorithms. The study was able to cluster players according to the derived performance metrics, which was calculated using the statistics of the players. The result of the study showed that metrics related to boundaries such as Boundary Index (BI), Fours to Sixes Ratio (FSRatio) are key in identifying batting talents. Whereas metrics related to wickets and runs conceded such as Big Wicket Taking Score (BWTS) and Economy Score (ES) are key in identifying bowling talents. The derived performance metrics gave a holistic view of player's ability compared to traditional performance metrics.

The result provides the ability to identify talent and develop players according to the performance metric that the coaches and selectors want to train on. Furthermore, the study helps in not only clustering players but identifying metrics on which the trainings can be prepared, and strategy could be developed for player development.

However, the study does come with limitations such as the study is focused on the bowling and batting ability of players, and it tends to not take in factor the fielding aspect. Moreover, the future research can incorporate not only the fielding aspect but also other aspect which are related to match conditions. The cricket county, while it develops an API system for its website, should encourage scorers to start recording ball-by-ball data more consistently to provide a broader view of player performance. In future, supervised learning methods can be incorporated with the unsupervised ones to enhance the ability to assess player performance.

In conclusion, the study provides a foundation on how unsupervised learning can be used in grassroot levels of sports and opens an avenue to consider future studies across other level of the sport.

Bibliography

Arthur, David, and Sergei Vassilvitskii. 2007. "K-Means++: The Advantages of Careful Seeding." In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms.*

Bellman, Richard. 1966. "Dynamic Programming." *Science* 153(3731): 34–37. https://www.science.org.

Bose, Avijit, Sannoy Mitra, Souham Ghosh, Raima Ghosh, Tiyash Patra, and Satyajit Chakrabarti. 2021. "Unsupervised Learning Based Evaluation of Player Performances." *Innovations in Systems and Software Engineering* 17(2): 121–30. doi:10.1007/s11334-020-00374-3.

Bridges, Cecil C. 1966. "Hierarchical Cluster Analysis." *Pyschological Reports* 18(3): 851–54.

Deep, C., C. Patvardhan, and Sushobhit Singh. 2016. "A New Machine Learning Based Deep Performance Index for Ranking IPL T20 Cricketers." *International Journal of Computer Applications* 137(10): 42–49. doi:10.5120/ijca2016908903.

Deep Prakash, C., and Sanjay Verma. 2022. "A New In-Form and Role-Based Deep Player Performance Index for Player Evaluation in T20 Cricket." *Decision Analytics Journal* 2. doi:10.1016/j.dajour.2022.100025.

Hinto, G.E, and R.R Salakhutdinov. 2006. "Reducing the Dimensionality of Data with Neural Networks." *Science* 313(5786): 504–7. doi:10.1126/science.1129198.

Jayanth, Sandesh Bananki, Akas Anthony, Gududuru Abhilasha, Noorni Shaik, and Gowri Srinivasa. 2018. "A Team Recommendation System and Outcome Prediction for the Game of Cricket." *Journal of Sports Analytics* 4(4): 263–73. doi:10.3233/jsa-170196.

Joliffe, I.T., and Bjt Morgan. 1992. "Principal Component Analysis and Exploratory Factor Analysis." *Statistical Methods in Medical Research* 1(1): 69–95.

Kimber, Alan C, and Alan R Hansford. 1993. 156 Source: Journal of the Royal Statistical Society. Series A (Statistics in Society *a Statistical Analysis of Batting in Cricket*.

Lemmer, Hermanus H. 2004. "A Measure for the Batting Performance of Cricket Players." *South African Journal for Research in Sport, Physical Education and Recreation* 26(1): 55–64.

Lewis, A. J. 2005. "Towards Fairer Measures of Player Performance in One-Day Cricket." *Journal of the Operational Research Society* 56(7): 804–15. doi:10.1057/palgrave.jors.2601876.

Manage, Ananda B.W., and Stephen M. Scariano. 2013. "An Introductory Application of Principal Components to Cricket Data." *Journal of Statistics Education* 21(3). doi:10.1080/10691898.2013.11889689.

Mousavi, S. Mostafa, Weiqiang Zhu, William Ellsworth, and Gregory Beroza. 2019. “Unsupervised Clustering of Seismic Signals Using Deep Convolutional Autoencoders.” *IEEE Geoscience and Remote Sensing Letters* 16(11): 1693–97. doi:10.1109/LGRS.2019.2909218.

Murtagh, Fionn, and Pedro Contreras. 2012. “Algorithms for Hierarchical Clustering: An Overview.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2(1): 86–97. doi:10.1002/widm.53.

Ofoghi, Bahadorreza, John Zeleznikow, Dan Dwyer, and Clare Macmahon. 2013. “Modelling and Analysing Track Cycling Omnium Performances Using Statistical and Machine Learning Techniques.” *Journal of Sports Sciences* 31(9): 954–62. doi:10.1080/02640414.2012.757344.

Prakash, Chellapilla Deep. 2016a. “A New Category Based Deep Performance Index Using Machine Learning for Ranking IPL Cricketers.” *International Journal of Electronics, Electrical and Computational System*. https://www.researchgate.net/publication/310240177.

Prakash, Chellapilla Deep. 2016b. *A New Category Based Deep Performance Index Using Machine Learning for Ranking IPL Cricketers*. https://www.researchgate.net/publication/310240177.

Saikia, Hemanta, and Dibyojyoti Bhattacharjee. 2011. “On Classification of All-Rounders of the Indian Premier League (IPL): A Bayesian Approach.” *Vikalpa* 36(4): 51–66. https://journals.sagepub.com/doi/abs/10.1177/0256090920110404 (July 28, 2024).

Saikia, Hemanta, Dibyojyoti Bhattacharjee, and Hermanus H. Lemmer. 2012. “Predicting the Performance of Bowlers in IPL: An Application of Artificial Neural Network.” *International Journal of Performance Analysis in Sport* 12(1): 75–89. doi:10.1080/24748668.2012.11868584.

Shelly, Zachary, Reuben F Burch, Wenmeng Tian, Lesley Strawderman, Anthony Piroli, and Corey Bichey. 2020. “Using K-Means Clustering to Create Training Groups for Elite American Football Student-Athletes Based on Game Demands.” *International Journal of Kinesiology & Sports Science* 8(2). doi:10.7575/aiac.ijkss.v.8n.2p.47.

Spencer, Bart, John Zeleznikow, Samuel J Robertson, Bartholomew Spencer, Stuart Morgan, and Sam Robertson. 2016. “CLUSTERING TEAM PROFILES IN THE AUSTRALIAN FOOTBALL LEAGUE USING PERFORMANCE INDICATORS.” In *Proceedings of the 13th Australasian Conference on Mathematics and Computers in Sports*, Melbourne, Australia. https://www.researchgate.net/publication/303786094.

Tharoor, Vishnusai Viswajith, and N. M. Dhanya. 2022. “Performance of Indian Cricket Team in Test Cricket: A Comprehensive Data Science Analysis.” In *Proceedings of the 2022 International Conference on Electronic Systems and Intelligent Computing, ICESIC 2022*,

Institute of Electrical and Electronics Engineers Inc., 128–33. doi:10.1109/ICESIC53714.2022.9783492.

List of Figures, Tables, and Equations

List of Figures in Main Content

FIGURE 3.1: HISTOGRAM OF DERIVED PERFORMANCE METRICS (BATTING SCORECARD) 28

FIGURE 3.2: CORRELATION MATRIX OF DERIVED PERFORMANCE METRICS (BATTING SCORECARD) 29

FIGURE 3.3: PAIR PLOT OF DERIVED PERFORMANCE METRICS (BATTING SCORECARD) 30

FIGURE 3.4: HISTOGRAM OF DERIVED PERFORMANCE METRICS (BOWLING SCORECARD) 32

FIGURE 3.5: CORRELATION MATRIX OF DERIVED PERFORMANCE METRICS (BOWLING SCORECARD) 33

FIGURE 3.6: PAIR PLOT OF DERIVED PERFORMANCE METRICS (BOWLING SCORECARD) 34

FIGURE 3.7: CUMULATIVE VARIANCE OF BATTING SCORECARD (LEFT) AND BATTING SCORECARD BALL BY BALL (RIGHT) 39

FIGURE 3.8: LOADING VECTORS USING HEATMAP TO SEE WHICH DERIVED PERFORMANCE METRIC INFLUENCES WHICH PRINCIPAL COMPONENTS. ON THE LEFT IS THE HEATMAP FOR BATTING SCORECARD AND THE RIGHT IS THE HEAT MAP FOR BALL-BY-BALL DATA'S BATTING SCORECARD 40

FIGURE 3.9:CUMULATIVE VARIANCE OF BOWLING SCORECARD (LEFT) AND BOWLING SCORECARD BALL BY BALL (RIGHT) 42

FIGURE 3.10: ON THE LEFT IS THE LOADING VECTOR HEATMAP OF BOWLING SCORECARD AND ON THE RIGHT IS LOADING VECTOR HEAT MAP OF BALL-BY-BALL BOWLING SCORECARD'S LOADING VECTOR HEAT MAP. THESE HEAT MAP SHOWS WHICH DERIVED PERFORMANCE METRIC INFLUENCES WHICH DIMENSIONS 43

FIGURE 3.11: RELATIONSHIP BETWEEN RECONSTRUCTION ERROR AND NUMBER OF ENCODED DIMENSIONS. ON THE LEFT IS FOR THE BATTING SCORECARD AND ON THE RIGHT FOR BALL-BY-BALL BATTING SCORECARD. 45

FIGURE 3.12: AUTOENCODER FOR THE BATTING SCORECARD 46

FIGURE 3.13: DERIVED PERFORMANCE METRIC INFLUENCE ON EACH ENCODED DIMENSION FOR BATTING SCORECARD (LEFT) AND BALL-BY-BALL BATTING SCORECARD (RIGHT) 46

FIGURE 3.14: RELATIONSHIP BETWEEN RECONSTRUCTION ERROR AND NUMBER OF ENCODED DIMENSIONS. ON THE LEFT IS FOR THE BOWLING SCORECARD AND ON THE RIGHT FOR BALL-BY-BALL BOWLING SCORECARD 47

FIGURE 3.15: DERIVED PERFORMANCE METRIC INFLUENCE ON EACH ENCODED DIMENSION FOR BOWLING SCORECARD (LEFT) AND BALL-BY-BALL BOWLING SCORECARD (RIGHT) 48

FIGURE 3.16: FINDING THE OPTIMAL CLUSTER OF BATTING SCORECARDS USING ELBOW METHOD. BATTING SCORECARD (PCA) ON THE TOP LEFT, BATTING SCORECARD (AUTOENCODER) ON THE BOTTOM LEFT, BALL-BY-BALL BATTING SCORECARD (PCA) ON TOP RIGHT AND BALL-BY-BALL BATTING SCORECARD (AUTOENCODER) ON THE BOTTOM RIGHT. 51

FIGURE 3.17: SILHOUETTE SCORES TO FIND THE OPTIMAL NUMBER OF CLUSTERS FOR BATTING SCORECARD (PCA) ON THE TOP LEFT, BATTING SCORECARD (AUTOENCODER) ON THE BOTTOM LEFT, BALL-BY-BALL BATTING SCORECARD (PCA) ON TOP RIGHT AND BALL-BY-BALL BATTING SCORECARD (AUTOENCODER) ON THE BOTTOM RIGHT 52

List of Figures in Appendix A

Appendix A Figure 1: Histogram of derived performance metrics (Batting Scorecard Ball by Ball) 78

Appendix A Figure 2: Correlation matrix of derived performance metrics (Batting Scorecard Ball by Ball Data) 78

Appendix A Figure 3: Pair plot of derived performance metrics (Batting Scorecard Ball by Ball data)..... 79

Appendix A Figure 4: Histogram of derived performance metrics (Bowling Scorecard Ball by Ball data) 79

Appendix A Figure 5: Correlation matrix of derived performance matrix (Bowling Scorecard ball by ball data) 80

Appendix A Figure 6: Pair plot of derived performance metrics (Bowling Scorecard Ball by Ball data) 80

List of Figures in Appendix B

Appendix B Figure 1: KMeans Clustering on Principal Components of Batting Scorecards. On the left is the Batting Scorecard and on the right is the ball-by-ball Batting Scorecard..... 81

Appendix B Figure 2: KMeans Clustering of encoded dimension of Batting Scorecard..... 81

Appendix B Figure 3: KMeans Clustering of encoded dimension of ball-by-ball Batting Scorecard 82

Appendix B Figure 4: Plot of WCSS against Number of Clusters (k) to use elbow method to extract optimal number of clusters. On the top left is Bowling Scorecard (PCA), bottom left is Bowling Scorecard (Encoded), top right is ball-by-ball Bowling Scorecard (PCA) and bottom right is ball-by-ball Bowling Scorecard (Encoded)..... 83

Appendix B Figure 5: Silhouette Score plot for Bowling Scorecard (PCA) on top left, Bowling Scorecard (Autoencoder) on bottom left, ball-by-ball Bowling Scorecard (PCA) on top right and ball-by-ball Bowling Scorecard (Autoencoder) on bottom right 84

Appendix B Figure 6: KMeans Clustering plot for Bowling Scorecard (PCA) 85

Appendix B Figure 7: KMeans Clustering plot for ball-by-ball Bowling Scorecard (PCA)..... 86

Appendix B Figure 8: KMeans Clustering plot for encoded dimensions of Bowling Scorecards..... 87

Appendix B Figure 9: KMeans Clustering plot of encoded dimensions of ball-by-ball Bowling Scorecard 88

Appendix B Figure 10: Dendrogram of Batting Scorecard (PCA)..... 89

Appendix B Figure 11: Dendrogram of ball-by-ball Batting Scorecard (PCA) 90

Appendix B Figure 12: Dendrogram of Encoded Dimensions of Batting Scorecard..... 91

Appendix B Figure 13: Dendrogram of Encoded Dimensions of ball-by-ball Batting Scorecard 92

Appendix B Figure 14: Dendrogram of Bowling Scorecard (PCA) 93

Appendix B Figure 15: Dendrogram of ball by ball Bowling Scorecard (PCA)..... 94

Appendix B Figure 16: Dendrogram of Encoded Dimension of Bowling Scorecard..... 95

Appendix B Figure 17: Dendrogram of Encoded Dimension of ball by ball Bowling Scorecard 96

Appendix B Figure 18: Hierarchical Clustering of Batting Scorecard (PCA) 97

Appendix B Figure 19: Hierarchical Clustering of ball-by-ball Batting Scorecard (PCA) 97

Appendix B Figure 20: Hierarchical Clustering of Encoded Dimension of Batting Scorecard..... 98

Appendix B Figure 21: Hierarchical Clustering of Encoded Dimension of ball-by-ball Batting Scorecard 99

Appendix B Figure 22: Hierarchical Clustering of Bowling Scorecard (PCA) 100

Appendix B Figure 23: Hierarchical Clustering of ball-by-ball Bowling Scorecard (PCA)..... 101

Appendix B Figure 24: Hierarchical Clustering of Encoded Dimension of Bowling Scorecard..... 102

Appendix B Figure 25: Hierarchical Clustering of Encoded Dimension of ball-by-ball Bowling Scorecard 102

List of Tables in Main Content

Table 3.1: List of Leagues and Number of Teams in Each League 18
Table 3.2: Description of player statistics for the 2023 season (Batting Scorecards) 22
Table 3.3: Customised Batting Scorecard for ball-by-ball data 24
Table 3.4: Additional batting statistics for ball-by-ball analysis 24
Table 3.5: Description of player statistics in 2023 season (Bowling Scorecards) 25
Table 3.6: Customised Bowling Scorecard for ball-by-ball data 26
Table 3.7: Additional player statistics that were added to Bowling Scorecard's player statistics. These will be used to calculate derived performance metrics of bowlers from ball-by-ball dataset 26
Table 3.8: Descriptive statistics of derived performance metrics (Batting Scorecard) 27
Table 3.9: Descriptive statistics of derived performance metrics (Batting Scorecard - Ball by Ball) 31
Table 3.10: Descriptive statistics of derived performance metrics (Bowling Scorecard) 31
Table 3.11: Descriptive statistics of derived performance metrics (Bowling Scorecard Ball by Ball data) 35
Table 3.12: Scaled derived performance metrics of Batting Scorecard 35
Table 3.13: Scaled derived performance metrics of Batting Scorecard (Ball By Ball) 36
Table 3.14: Scaled derived performance metrics of Bowling Scorecard 36
Table 3.15: Scaled derived performance metrics of Bowling Scorecard (Ball by Ball) 36
Table 3.16: Summary of dendrograms plotted and clusters formed of all the Batting Scorecards 54
Table 3.17: Summary of dendrograms plotted and clusters formed of all the Bowling Scorecards 55
Table 4.1: KMeans++ clusters means summary for Batting Scorecard (PCA) 56
Table 4.2: KMeans++ clusters mean summary for ball-by-ball Batting Scorecard (PCA) 56
Table 4.3: Ranking of player after applying composite score formula in high performing cluster 58

List of Tables in Appendix A

Appendix A Table 1: List of teams in each league 73
Appendix A Table 2: Metrics extracted from the scorecard page of each team 75
Appendix A Table 3: The derived performance metrics is calculated using the performance metrics from the scorecard. These performance metrics will be used for cluster analysis 76

List of Tables in Appendix C

Appendix C Table 1: The means summary of derived performance metrics cluster-wise for the Batting Scorecard (PCA) 103
Appendix C Table 2: Means summary of derived performance metrics cluster-wise for the ball-by-ball Batting Scorecard (PCA) - Hierarchical Clustering 103
Appendix C Table 3: Means summary of derived performance metrics cluster-wise for Batting Scorecard with encoded dimensions - Kmeans++ Clustering 103
Appendix C Table 4: Means summary of derived performance metrics cluster-wise for Batting Scorecard with encoded dimensions - Hierarchical Clustering 103
Appendix C Table 5: Means summary of derived performance metrics cluster-wise for ball-by-ball Batting Scorecard with encoded dimensions - KMeans++ clustering 103
Appendix C Table 6: Ranked Players in high performing cluster for ball-by-ball Batting Scorecard with encoded dimensions - K-Means++ 103
Appendix C Table 7: Ranked Players in high performing cluster for ball-by-ball Batting Scorecard with encoded dimensions - Hierarchical Clustering 106

APPENDIX C TABLE 8: RANKED PLAYERS IN HIGH PERFORMING CLUSTER FOR BALL-BY-BALL BATTING SCORECARD (PCA) - KMEANS 108
APPENDIX C TABLE 9: RANKED PLAYERS IN HIGH PERFORMING CLUSTER FOR BALL-BY-BALL BATTING SCORECARD (PCA) – HIERARCHICAL CLUSTERING 110
APPENDIX C TABLE 10: RANKED PLAYERS IN HIGH PERFORMING CLUSTER FOR BATTING SCORECARD WITH ENCODED DIMENSIONS – KMEANS++ CLUSTERING 113
APPENDIX C TABLE 11: RANKED PLAYERS IN HIGH PERFORMING CLUSTER FOR BATTING SCORECARD WITH ENCODED DIMENSIONS – HIERARCHICAL CLUSTERING 117
APPENDIX C TABLE 12: RANKED PLAYERS IN HIGH PERFORMING CLUSTER FOR BATTING SCORECARD (PCA) – KMEANS++ CLUSTERING 118
APPENDIX C TABLE 13: RANKED PLAYERS IN HIGH PERFORMING CLUSTER FOR BATTING SCORECARD (PCA) – HIERARCHICAL CLUSTERING 122
APPENDIX C TABLE 14: KMEANS++ CLUSTERS MEAN SUMMARY FOR BOWLING SCORECARD (PCA) 126
APPENDIX C TABLE 15: HIERARCHICAL CLUSTERS MEAN SUMMARY FOR BOWLING SCORECARD (PCA) 126
APPENDIX C TABLE 16: HIERARCHICAL CLUSTERS MEAN SUMMARY FOR BALL-BY-BALL BOWLING SCORECARD (PCA) 126
APPENDIX C TABLE 17: KMEANS++ CLUSTERS MEAN SUMMARY FOR BALL-BY-BALL BOWLING SCORECARD WITH ENCODED DIMENSIONS 126
APPENDIX C TABLE 18: HIERACHICAL CLUSTERS MEAN SUMMARY FOR BALL-BY-BALL BOWLING SCORECARD WITH ENCODED DIMENSIONS 126
APPENDIX C TABLE 19: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BOWLING SCORECARD (PCA) USING KMEANS++ CLUSTERING 126
APPENDIX C TABLE 20: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BOWLING SCORECARD (PCA) USING HIERARCHICAL CLUSTERING 130
APPENDIX C TABLE 21: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BOWLING SCORECARD WITH ENCODED DIMENSIONS USING KMEANS++ CLUSTERING 134
APPENDIX C TABLE 22: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BOWLING SCORECARD WITH ENCODED DIMENSIONS USING HIERARCHICAL CLUSTERING 139
APPENDIX C TABLE 23: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BALL-BY-BALL BOWLING SCORECARD (PCA) USING HIERARCHICAL CLUSTERING 144
APPENDIX C TABLE 24: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BALL-BY-BALL BOWLING SCORECARD WITH ENCODED DIMENSIONS USING KMEANS++ CLUSTERING 145
APPENDIX C TABLE 25: PLAYER RANKING IN HIGH PERFORMANCE CLUSTERS OF BALL-BY-BALL BOWLING SCORECARD WITH ENCODED DIMENSIONS USING HIERARCHICAL CLUSTERING 147

List of Equations

EQUATION 1: STANDARD DEVIATION 38
EQUATION 2: COVARIANCE MATRIX 38
EQUATION 3: EIGEN DECOMPOSITION 38
EQUATION 4: DATA PROJECTION 39
EQUATION 5: CALCULATION OF LEAKYRELU 43
EQUATION 6: LOWER DIMENSIONAL TRANSFORMATION OF ORIGINAL DATA BY ENCODER 44
EQUATION 7: RECONSTRUCTION OF THE LOWER-DIMENSIONAL DATA TO ORIGINAL DATA BY THE DECODER 44
EQUATION 8: RECONSTRUCTION ERROR WHICH IS CALCULATED USING MEAN SQUARED ERROR FORMULA 44
EQUATION 9: WITHIN-CLUSTER SUM OF SQUARES (WCSS) 48
EQUATION 10: PROBABILITY FORMULA FOR SELECTION OF CENTROID 49
EQUATION 11: EQUATION TO CALCULATE THE NEW CENTROID 49

EQUATION 12: SILHOUETTE SCORE **50**
EQUATION 13: WARD'S LINKAGE **53**
EQUATION 14: COMPOSITE SCORE CALCULATION TO RANK THE PLAYERS **58**

Appendix

Appendix A

The Python code for this report can be found using this link:

https://github.com/jamshaid2610/analysis-jcl/ or

https://drive.google.com/drive/folders/1bv_c40h5AVxfsc66EFq5oufXmK4FXM03?usp=sharing

Appendix A Table 1: List of teams in each league

Team Name	League Name
Brandon CC, Durham Under 13	Southwest U-13 League
Raby Castle CC Under 13	
Ushaw Moor CC Under 13	
Bishop Auckland CC Under 13	
Tudhoe CC Under 13	
Hunwick CC Under 13	
Langley Park & Bearpark CC Under 13'S	
Tudhoe CC Under 13 B	
Willington CC Under 13	
Crook CC Under 13	
Durham City CC Under 13 Blue	Central U-13 League Division One
Sacriston CC Under 13 - Red	
Chester Le Street CC Under 13	
Hetton Lyons CC Under 13 Green	
Washington CC, Durham Washington U13 Nova	
Durham City CC Under 13 Gold	
Littletown CC Under 13	
Sunderland CC Under 13	
Burnmoor CC Under 13	
Sacriston CC Under 13 - Black	
Philadelphia CC Under 13	
Eppleton CC Under 13	
Hetton Lyons CC Under 13 Gold	
Washington CC, Durham Washington U 13 Scotia	
Whitburn CC Under 13 A	

Gateshead Fell CC Under 13 A	Northeast U-13 League Division One
Felling CC Under 13	
Whitburn CC Under 13 B	
South Shields CC Under 13 A	
Boldon Ca CC Under 13	Northeast U-13 League Division Two
Gateshead Fell CC Under 13 B	
Marsden CC, Tyne & Wear Under 13	
Wearmouth & Yav CC Under 13	
Hylton CCc Under 13	
Bill Quay CC Under 13	
South Shields CC Under 13 B	
Boldon CC Under 13	
Dawdon Welfare CC Under 13	Southeast U-13 League
Seaham Park CC Under 13	
Seaham Harbour CC Under 13 Yellow	
Easington C.W. CC Under 13	
Castle Eden CC Under 13	
Murton CC Under 13	
Silksworth CC Under 13	
Seaham Park CC Under 13 B	
Seaham Harbour CC Under 13 Purple	
Ryhope CC Under 13	
Peterlee CC Under 13	
Shotley Bridge CC Under 13	Tyneside U-13 League Division One
South Northumberland CC Under 13 Green	
Greenside CC Under 13	
Whickham CC Under 13	
Ryton CC Under 13 Rockets	
Lintz CC Under 13	
Ryton CC Under 13 Raiders	
Leadgate CC Under 13	Tyneside U-13 League Division Two
Annfield Plain CC Under 13	
Lanchester CC Under 13	
Consett CC Under 13	
Blaydon CC Under 13	
Burnopfield CC Under 13	
Beamish And E. Stanley CC Under 13	
Burnhope CC Under 13	

Appendix A Table 2: Metrics extracted from the scorecard page of each team

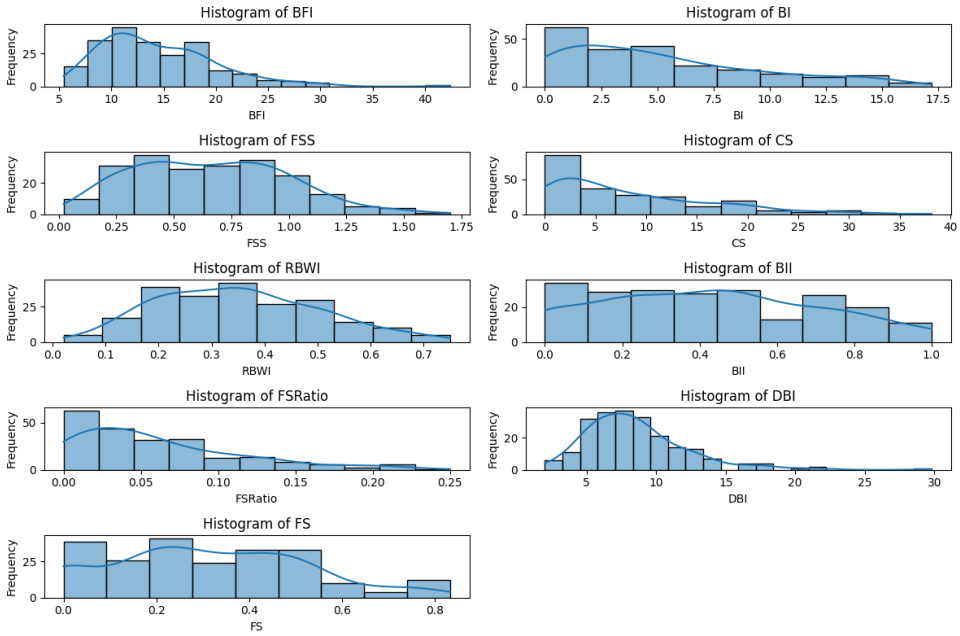
Type of Data	Metrics	Metric Definition
Batting Scorecard	Match Between the Teams	This stated what match was being recorded
	Batting Side	The batting team whose data is being recorded
	Bowling Side	The bowling team is added as a validation purpose to make sure correct match's data point is being recorded
	Batter Name	The name of the batter
	Runs	Runs scored by the batter in the match
	Balls	Number of balls played by the batter in the match
	4s	Number of fours scored by the batter in the match
	6s	Number of sixes scored by the batter in the match
	SR	Strike Rate of the Batsman for the match
	League	Which league the match belonged to
Bowling Scorecard	Match Between the Teams	This stated what match was being recorded
	Batting Side	The batting team is added as a validation purpose to make sure correct match's data point is being recorded
	Bowling Side	The bowling team whose data is being recorded
	Bowler Name	The name of the bowler
	Overs	Number of overs the bowler bowled for the match
	Maidens	Number of maiden overs bowled by the bowler for the match
	Runs	Number of runs conceded by the bowler for the match
	Wickets	Number of Wickets taken by the bowler for the match
	Wides	Number of wides bowled by the bowler for the match
	No Balls	Number of no balls bowled by the bowler for the match
	Econ	The economy rate of the bowler for the match
	League	Which league the match belonged to
Ball By Ball Data	Match Between the Teams	This states what match is being recorded
	Batting Team	The batting team's name
	Bowling Team	The bowling team's name
	Overs	Which over is being recorded
	Batter	The name of the batter
	Bowler	The name of the bowler
	Runs	Runs scored on the specific ball being recorded
	Run Type	The type of the run being recorded i.e. 'dismissed' if it is a wicket, 'run/runs' if it is any type of score, 'nb/w/nb+w/b' if it is an extra
	League	Which league the match belonged to

Appendix A Table 3: The derived performance metrics is calculated using the performance metrics from the scorecard. These performance metrics will be used for cluster analysis

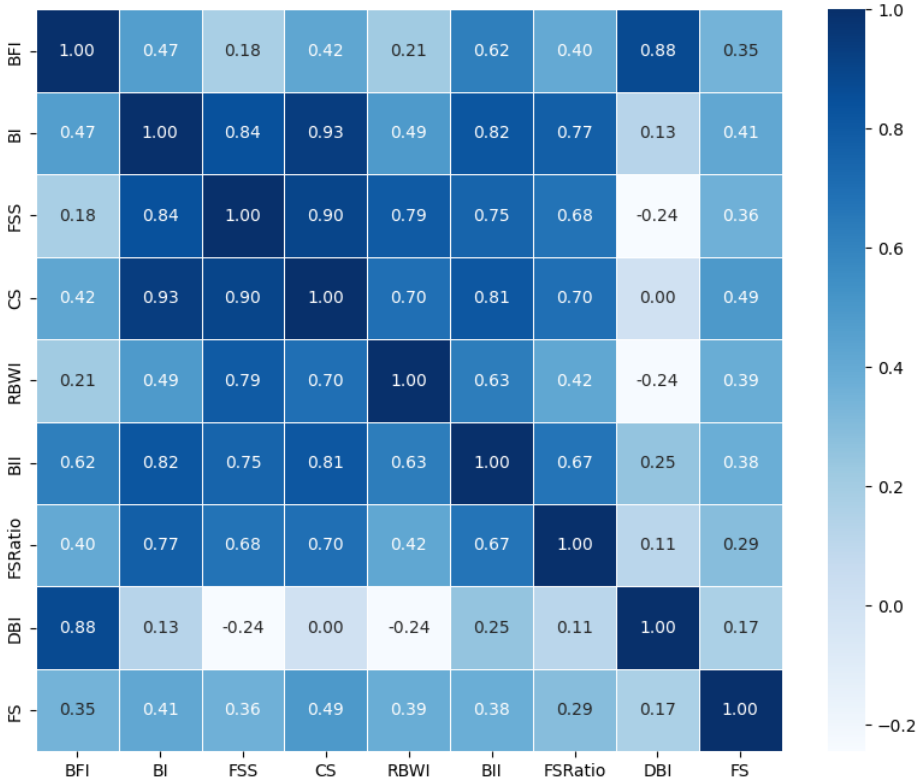
Derived Performance Metrics	Formula	Description
Batting Scorecard		
Balls Faced Index (BFI)	$\frac{Total\ Balls}{Total\ Matches\ Played}$	The BFI considers of how many balls did a batter face per match. The higher the index, the more balls a batter faced in a match meaning that the batter spent a substantial amount of time batting
Boundary Index (BI)	$\frac{Total\ 4s + Total\ 6s}{Total\ Matches\ Played}$	The BI calculates how many boundaries did a batter score per match. Higher index means that the batter scores a lot of boundaries in a match
FSRatio	$Fours\ to\ Sixes\ Ratio$	The FSRatio indicates what shots do a batter prefer. A high FSRatio indicates that the batter prefers ground shots whereas low FSRatio indicates that they prefer lofted shots
Fast Score Scorer (FSS)	$\frac{Total\ Runs}{Total\ Balls}$	The FSS indicates how fast a batter scores. A higher score means that the batter scores quickly
Consistent Scorer (CS)	$SR * Batting\ Average$	The CS indicates, as the name suggests, how consistent is the batter performance. A high score indicates that a batter has a high strike rate and average hence performance good consistently
Running Between the Wicket Index (RBWI)	$\frac{TRWB}{Balls\ without\ Boundary}$	RBWI indicates how often the batter converted dot balls to runs
Big Innings Index (BII)	$\frac{Times\ Above\ Average}{Total\ Matches\ Played}$	BII indicates how often the batter has performed above average. The higher the index the more the batter performs above average
Dot Balls Index (DBI)	$\frac{Total\ Dot\ Balls}{Total\ Matches\ Played}$	DBI indicates how often the batter faces a dot ball in a match. High DBI would mean that the batter faces a lot of dot balls in a match
Finisher Score (FS)	$\frac{Total\ Not\ Dismissed}{Total\ Matches\ Played}$	FS indicates how often does a batter goes not out, a higher score would mean that the batter can end a match going not out

Bowling Scorecard		
Economy Score (ES)	$\frac{Total\ Runs\ Conceded}{Total\ Overs\ Bowled}$	The ES indicates how expensive a bowler is i.e. how many runs does a bowler conceded per over bowled
Wicket Taking Score (WTS)	$\frac{Total\ Wickets\ Taken}{Total\ Balls\ Bowled - Total\ Extras}$	WTS indicates how likely a bowler is to take wickets per legal bowl bowled. The higher the values, the more wicket taking ability bowler has
Balls Bowled Index (BBI)	$\frac{Total\ Balls\ Bowled}{Total\ Matches\ Played}$	BBI indicates how many bowls does the bowler ball per match
Consistent Scorer (CS)	$\frac{Total\ Runs\ Conceded}{Total\ Wickets\ Taken}$	CS indicates the bowling average of the bowler i.e. how many runs does the bowler concede per wicket taken. Lower CS value means that the bowler concedes less runs or takes more wickets
Big Wicket Taker Score (BWTS)	$\frac{Total\ Matches\ with\ 3W+}{Total\ Matches\ Played}$	BWTS indicates the ability of the bowler to take 3 or more wickets
Short Performance Score (SPS)	$\frac{Total\ Wickets - Total\ Wickets\ (3W+)}{Total\ Matches - Total\ Matches\ with\ 3W+}$	SPS indicates the ability of the bowler to take less than 3 wickets, the higher the score the more likely bowler can take less than 3 wickets
Maiden Index (MI)	$\frac{Total\ Maidens}{Total\ Overs\ Bowled}$	MI indicates the ability of the bowler to bowl maiden overs
Extras Index (EI)	$\frac{Total\ Extras}{Total\ Matches\ Played}$	EI indicates the ability of the bower to bowl extras in a match
Average Match Score (AVMS)	$\left(\frac{1}{ES}\right) * (WTS) * (ABBM)$	AVMS indicates the average score for a match that a bowler achieves, this means that if the bowler concedes more runs, then AVMS decreases, if the bowler takes more wickets the AVMS increases.
Dot Balls Index (DBI)	$\frac{Total\ Dot\ Balls}{Total\ Balls\ Bowled\ without\ Extras}$	DBI indicates how often the bowler balls a dot ball. The higher the index, the higher the bowler can ball a dot ball
Boundary Index (BI)	$\frac{Total\ Boundaries}{Total\ Overs}$	BI indicates the ability of the bowler to concede a boundary. Higher positive value means that the bowler concedes a lot of boundaries.

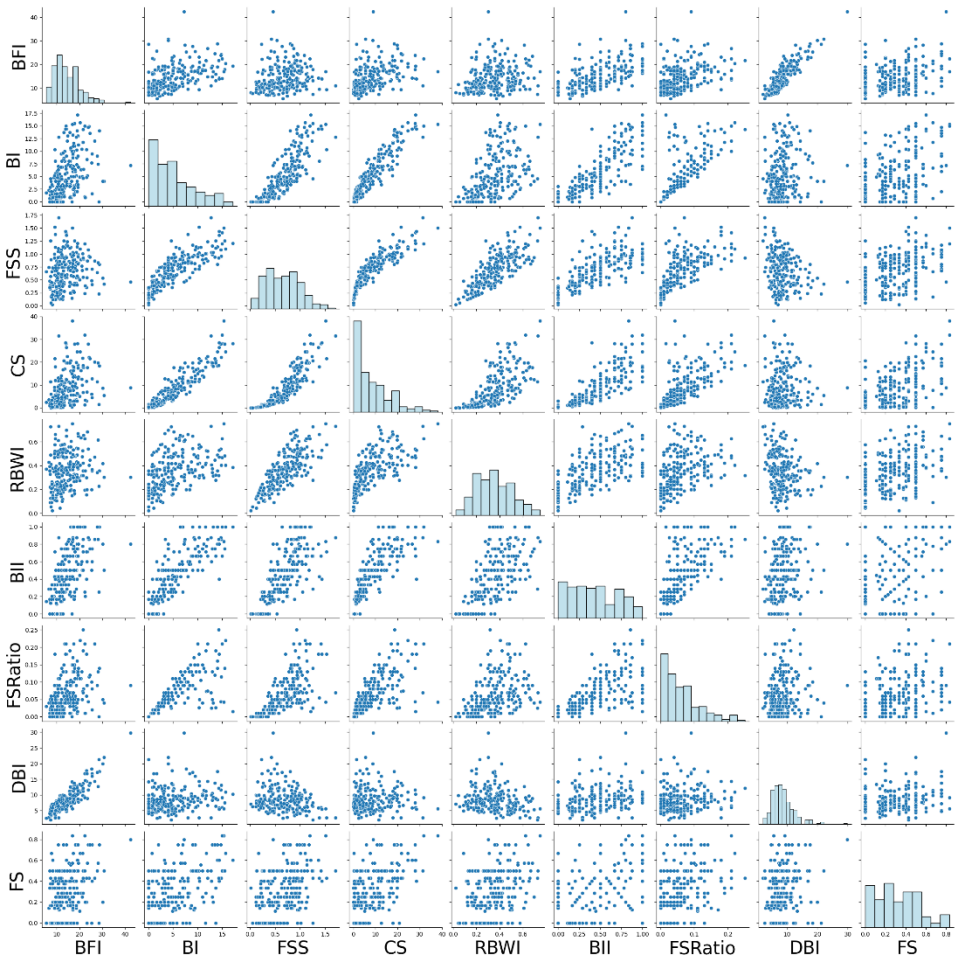
Appendix A Figure 1: Histogram of derived performance metrics (Batting Scorecard Ball by Ball)



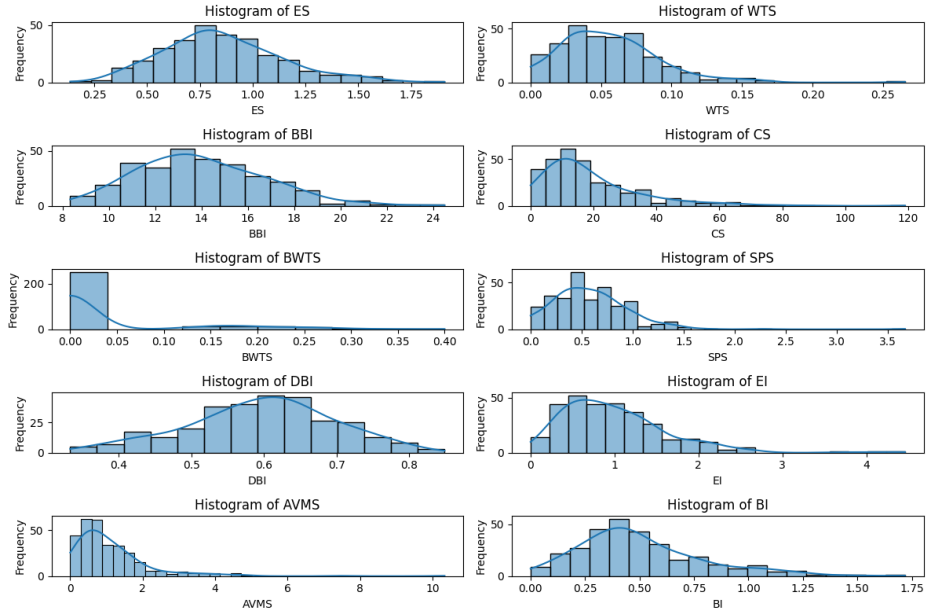
Appendix A Figure 2: Correlation matrix of derived performance metrics (Batting Scorecard Ball by Ball Data)



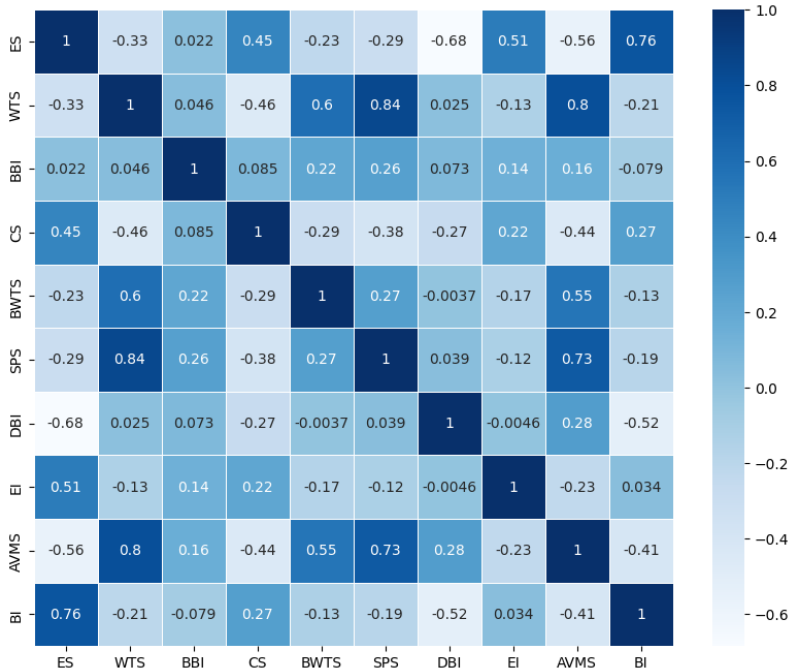
Appendix A Figure 3: Pair plot of derived performance metrics (Batting Scorecard Ball by Ball data)



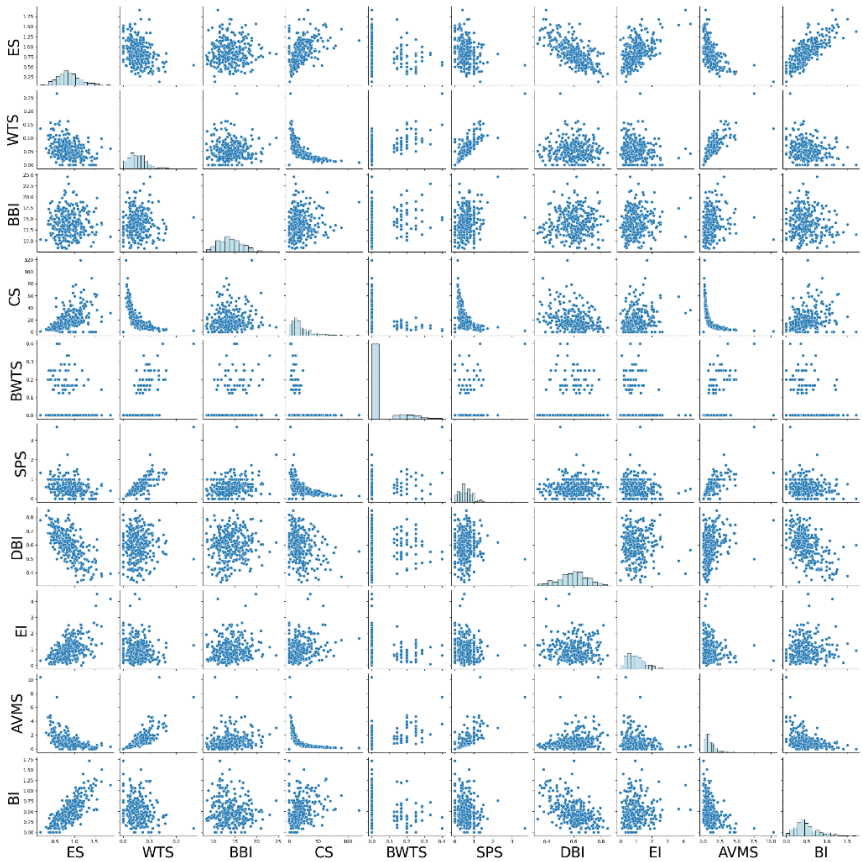
Appendix A Figure 4: Histogram of derived performance metrics (Bowling Scorecard Ball by Ball data)



Appendix A Figure 5: Correlation matrix of derived performance matrix (Bowling Scorecard ball by ball data)

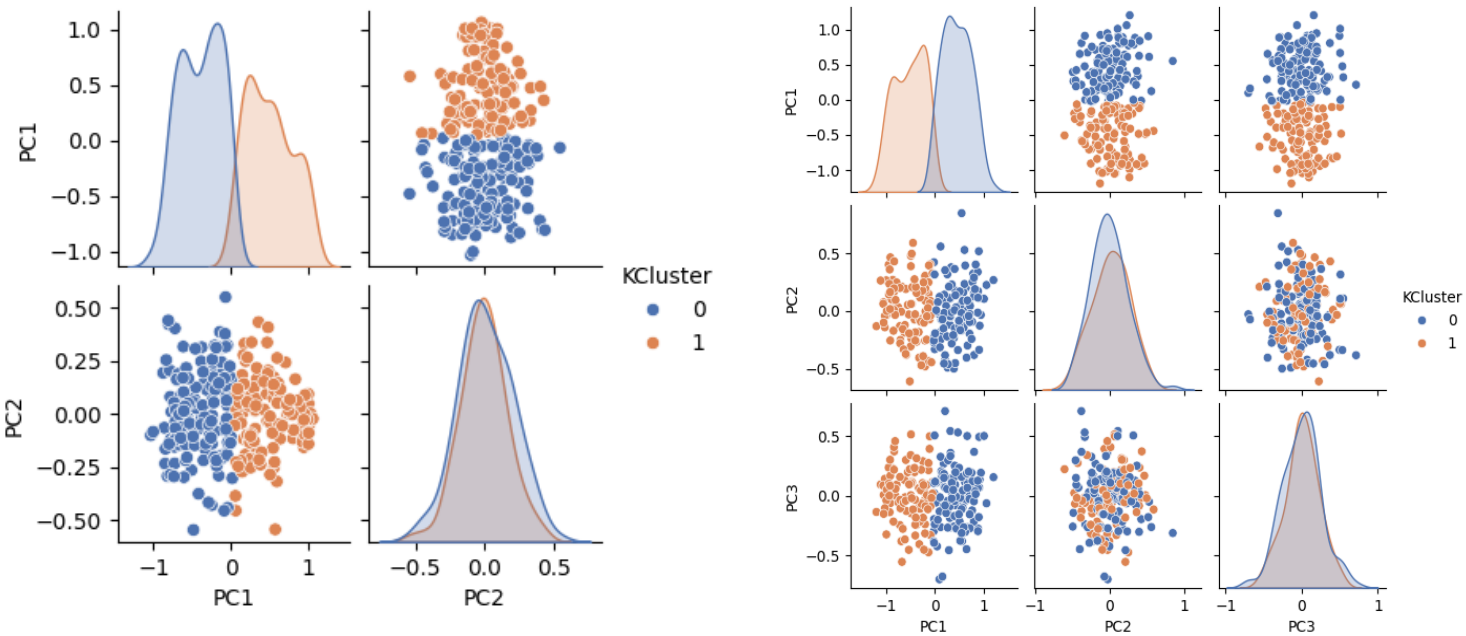


Appendix A Figure 6: Pair plot of derived performance metrics (Bowling Scorecard Ball by Ball data)

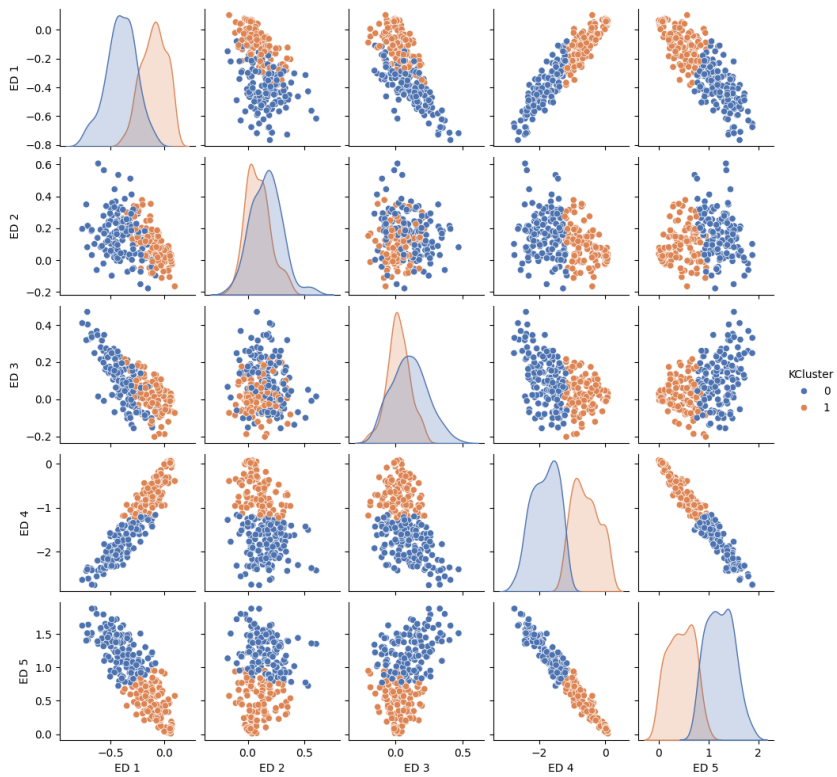


Appendix B

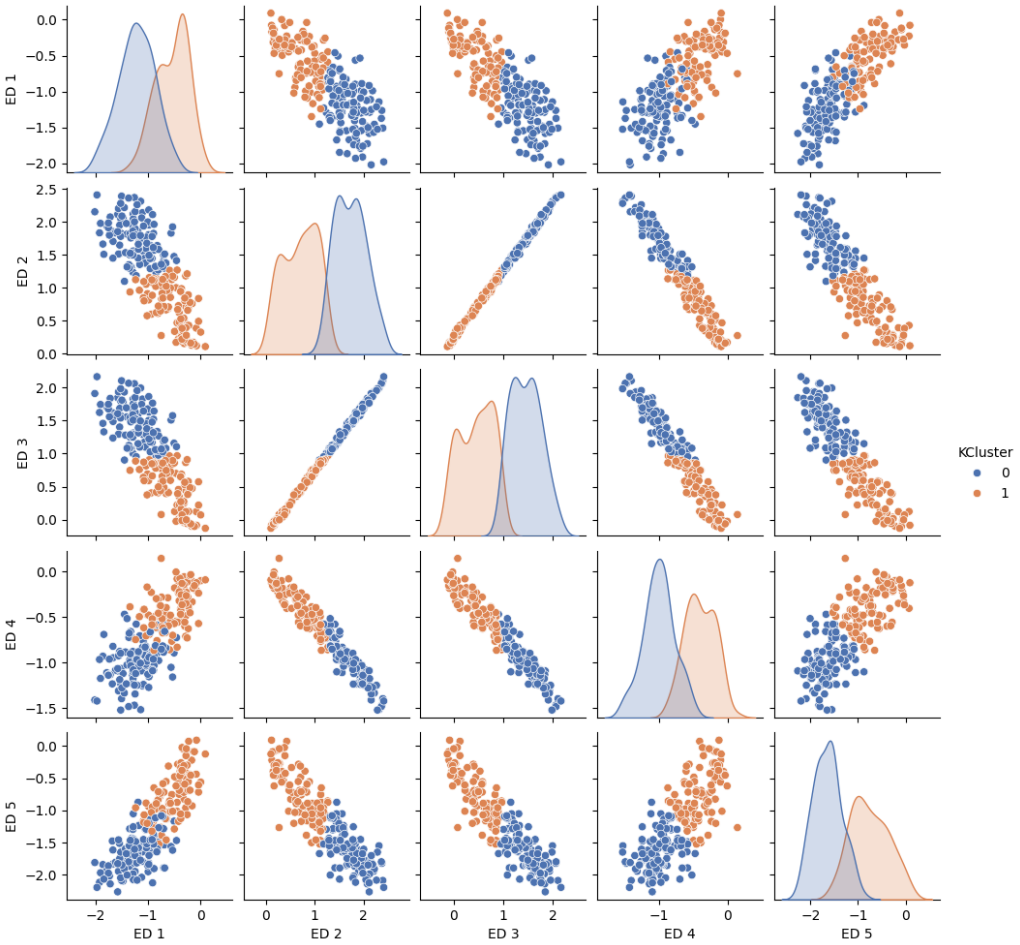
Appendix B Figure 1: KMeans Clustering on Principal Components of Batting Scorecards. On the left is the Batting Scorecard and on the right is the ball-by-ball Batting Scorecard



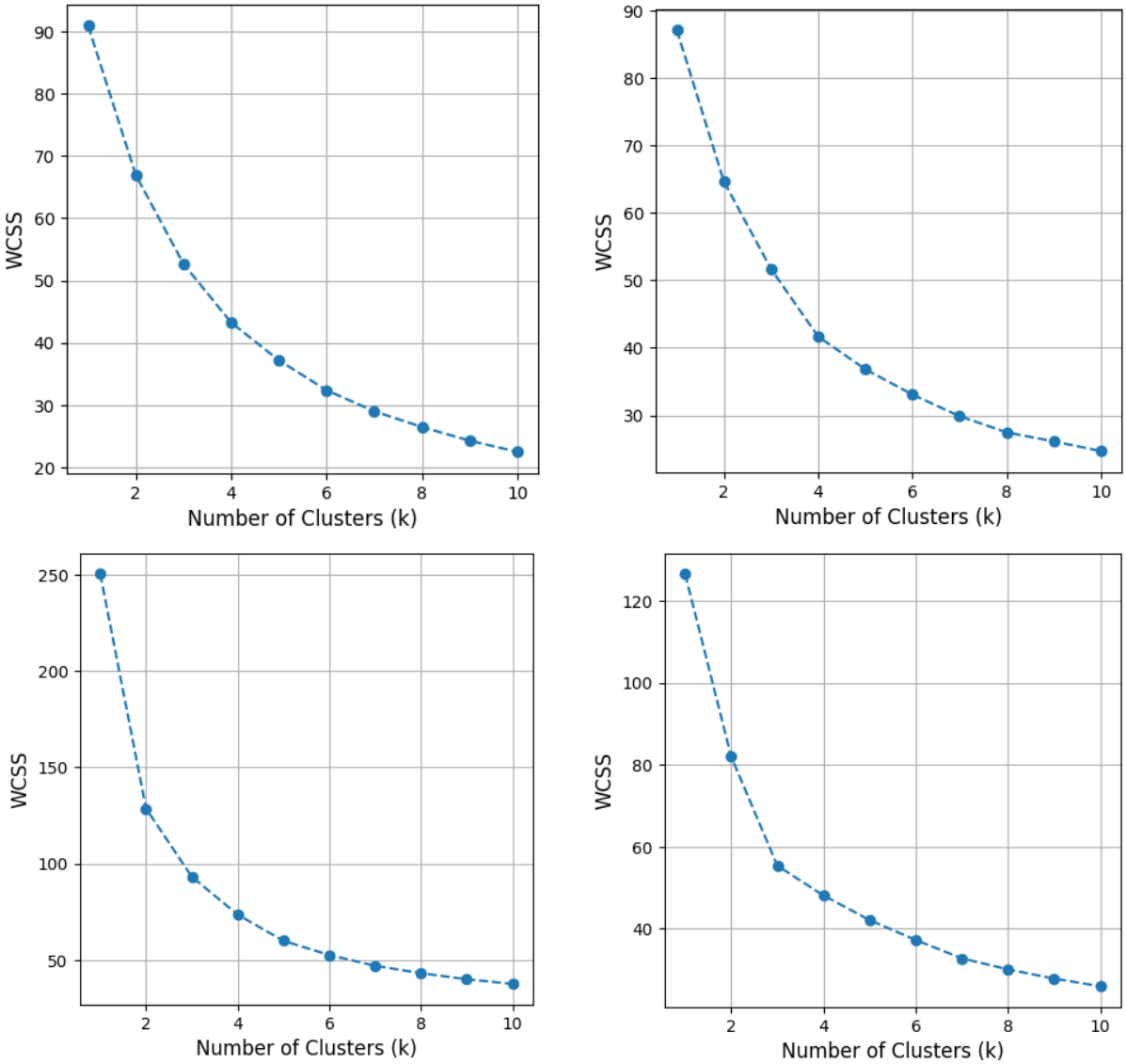
Appendix B Figure 2: KMeans Clustering of encoded dimension of Batting Scorecard



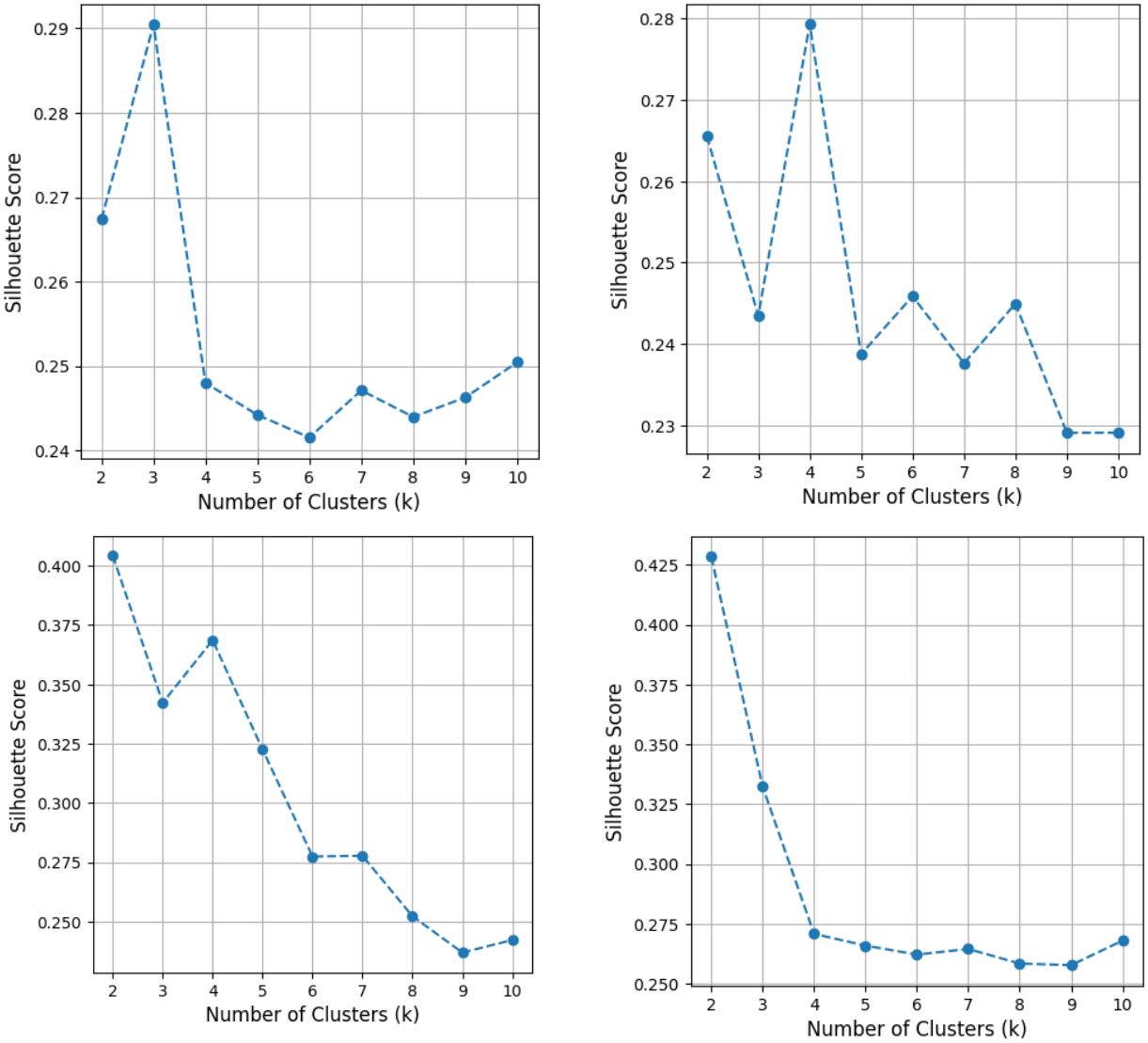
Appendix B Figure 3: KMeans Clustering of encoded dimension of ball-by-ball Batting Scorecard



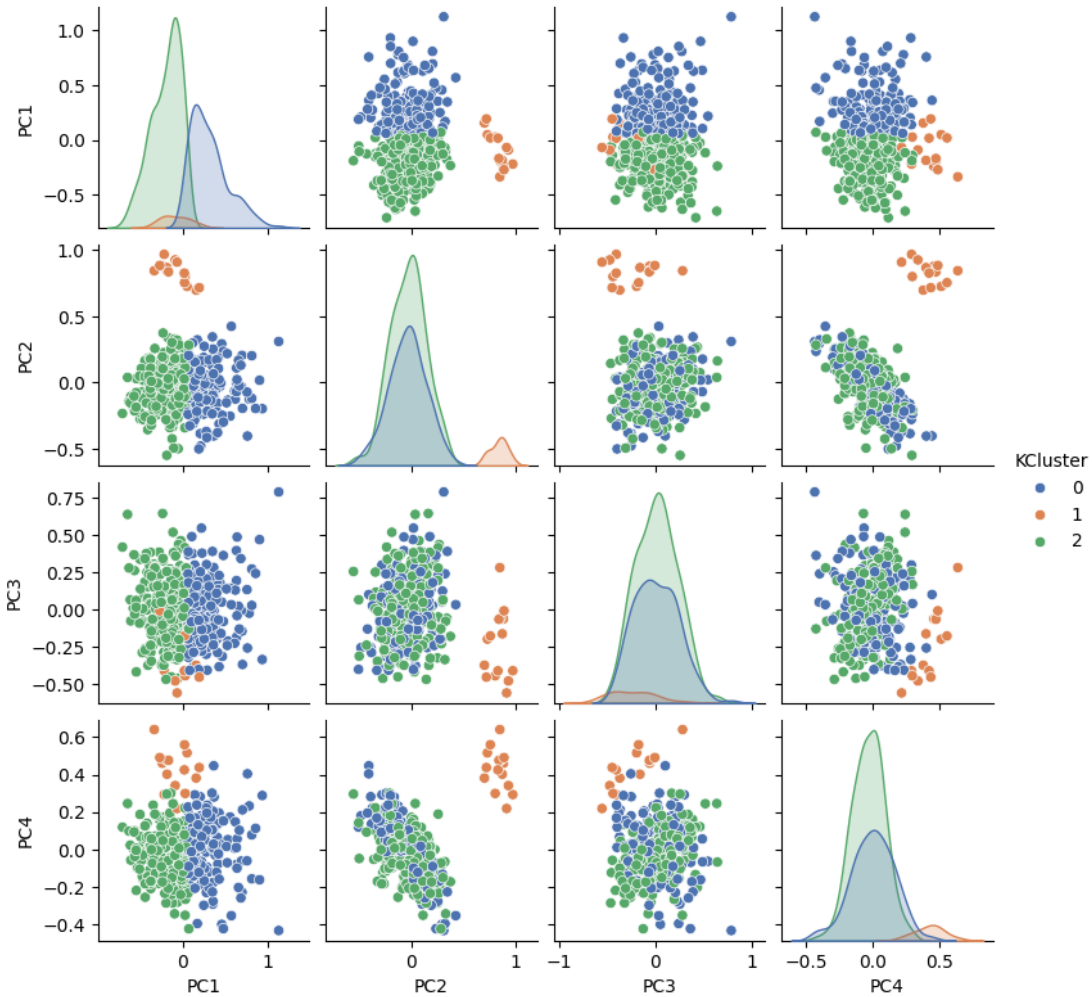
Appendix B Figure 4: Plot of WCSS against Number of Clusters (k) to use elbow method to extract optimal number of clusters. On the top left is Bowling Scorecard (PCA), bottom left is Bowling Scorecard (Encoded), top right is ball-by-ball Bowling Scorecard (PCA), and bottom right is ball-by-ball Bowling Scorecard (Encoded)



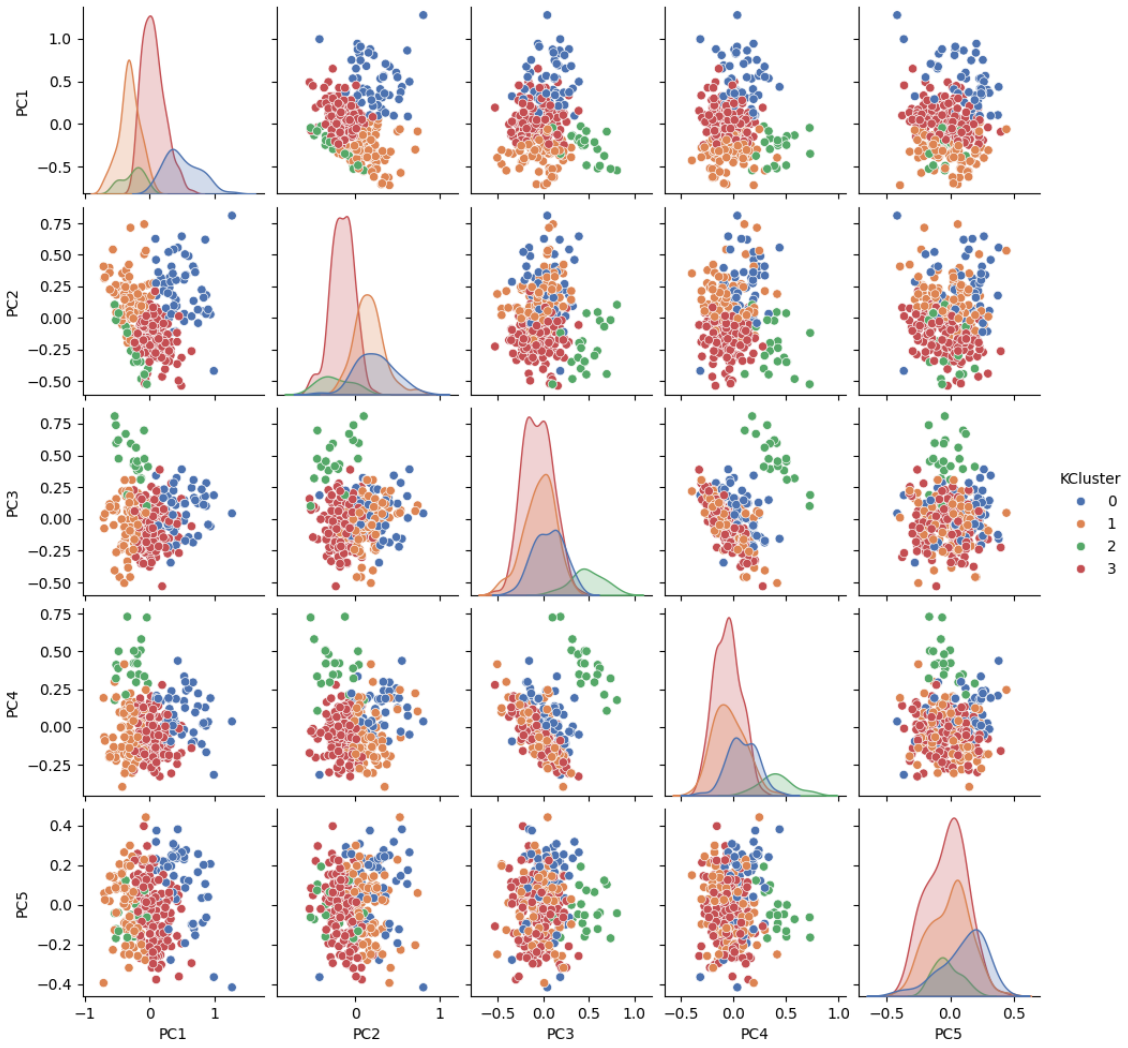
Appendix B Figure 5: Silhouette Score plot for Bowling Scorecard (PCA) on top left, Bowling Scorecard (Autoencoder) on bottom left, ball-by-ball Bowling Scorecard (PCA) on top right and ball-by-ball Bowling Scorecard (Autoencoder) on bottom right



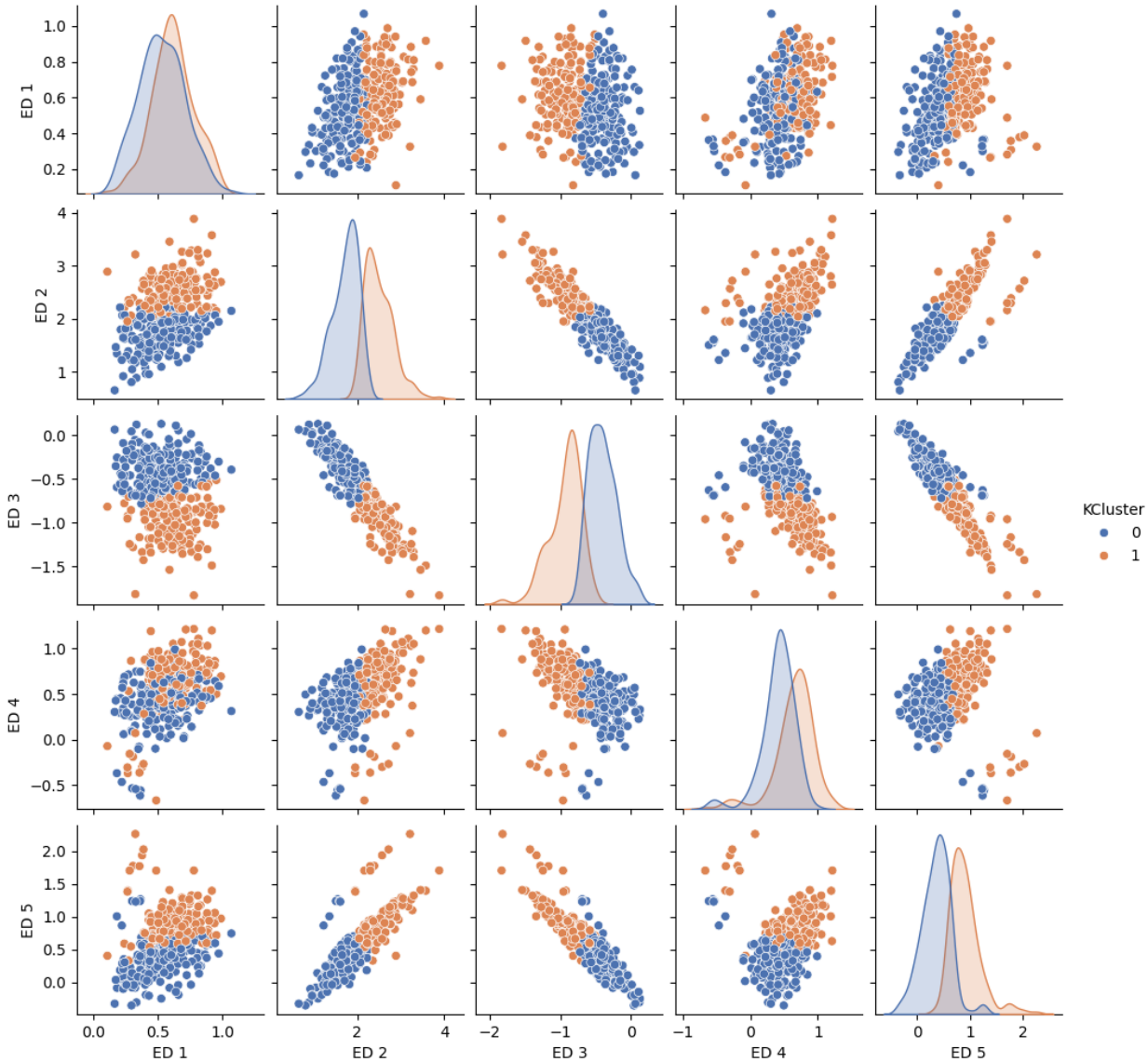
Appendix B Figure 6: KMeans Clustering plot for Bowling Scorecard (PCA)



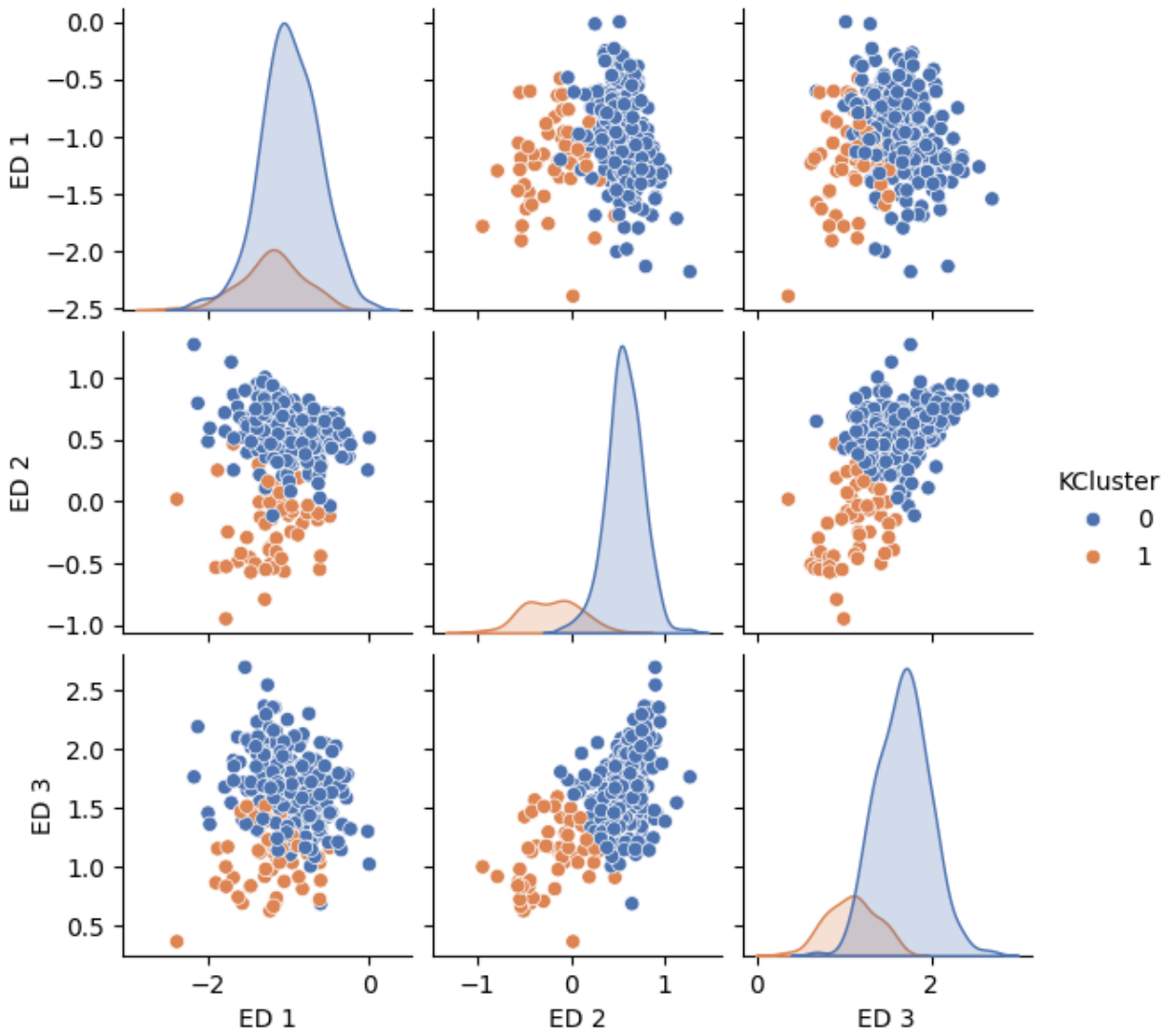
Appendix B Figure 7: KMeans Clustering plot for ball-by-ball Bowling Scorecard (PCA)



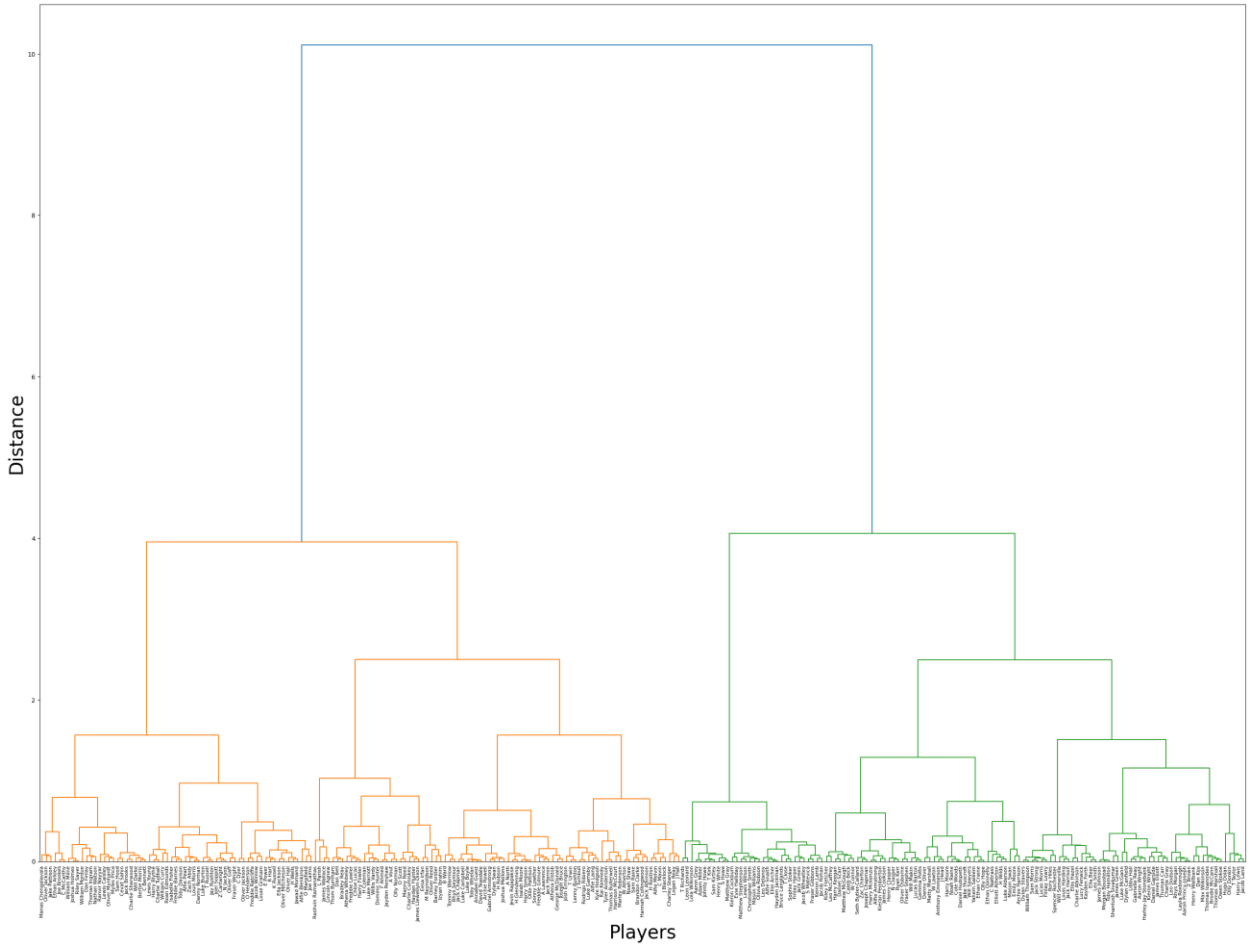
Appendix B Figure 8: KMeans Clustering plot for encoded dimensions of Bowling Scorecards



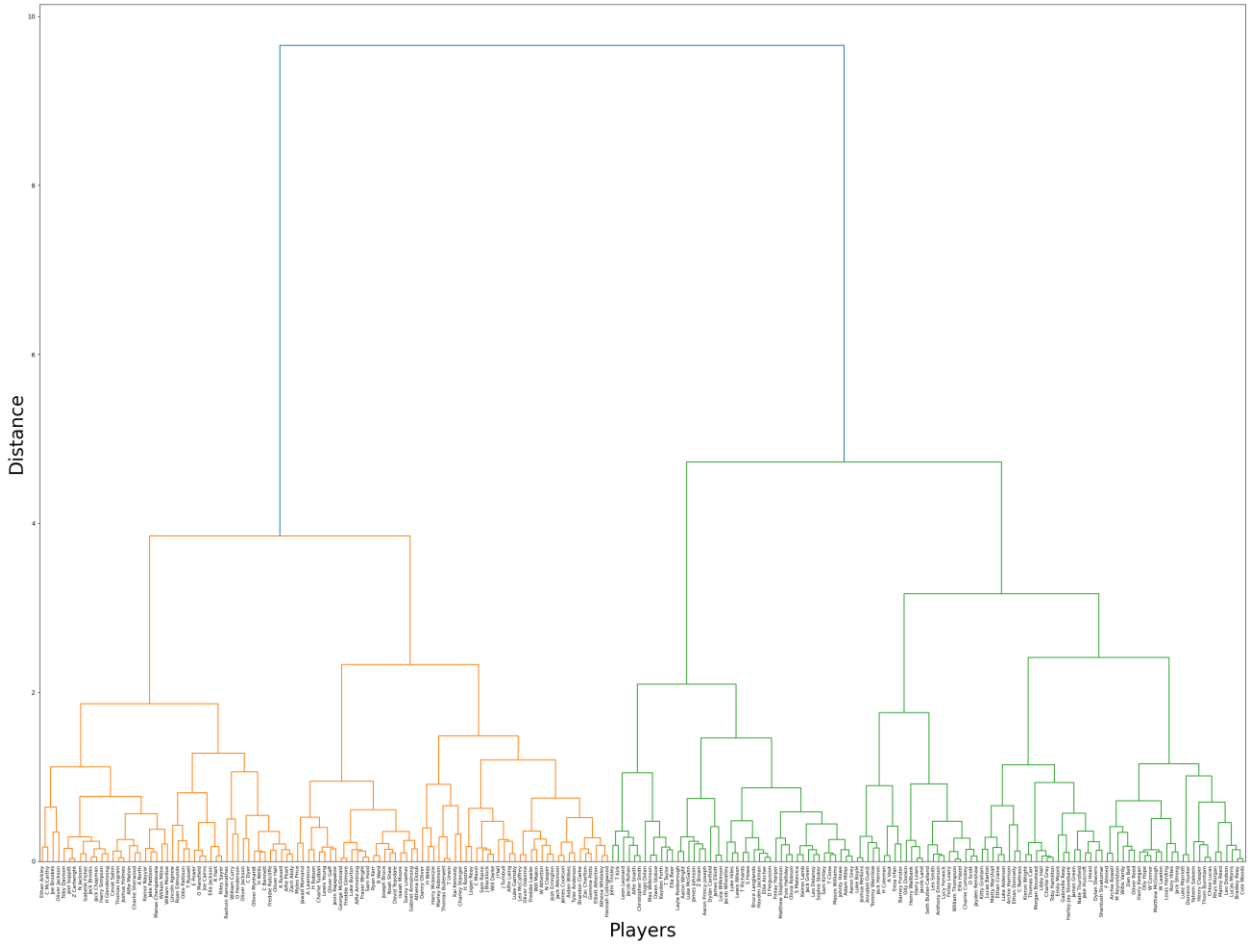
Appendix B Figure 9: KMeans Clustering plot of encoded dimensions of ball-by-ball Bowling Scorecard



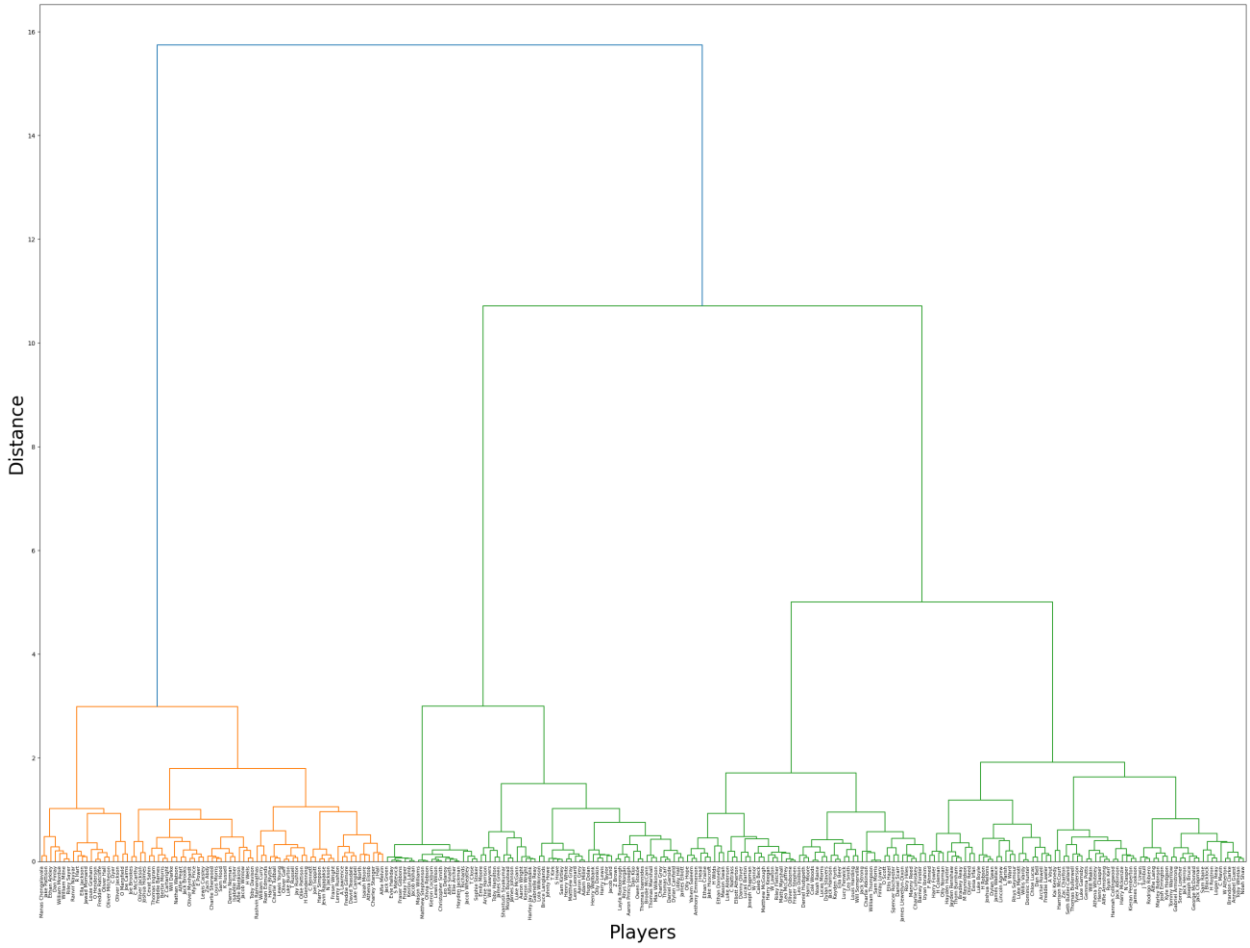
Appendix B Figure 10: Dendrogram of Batting Scorecard (PCA)



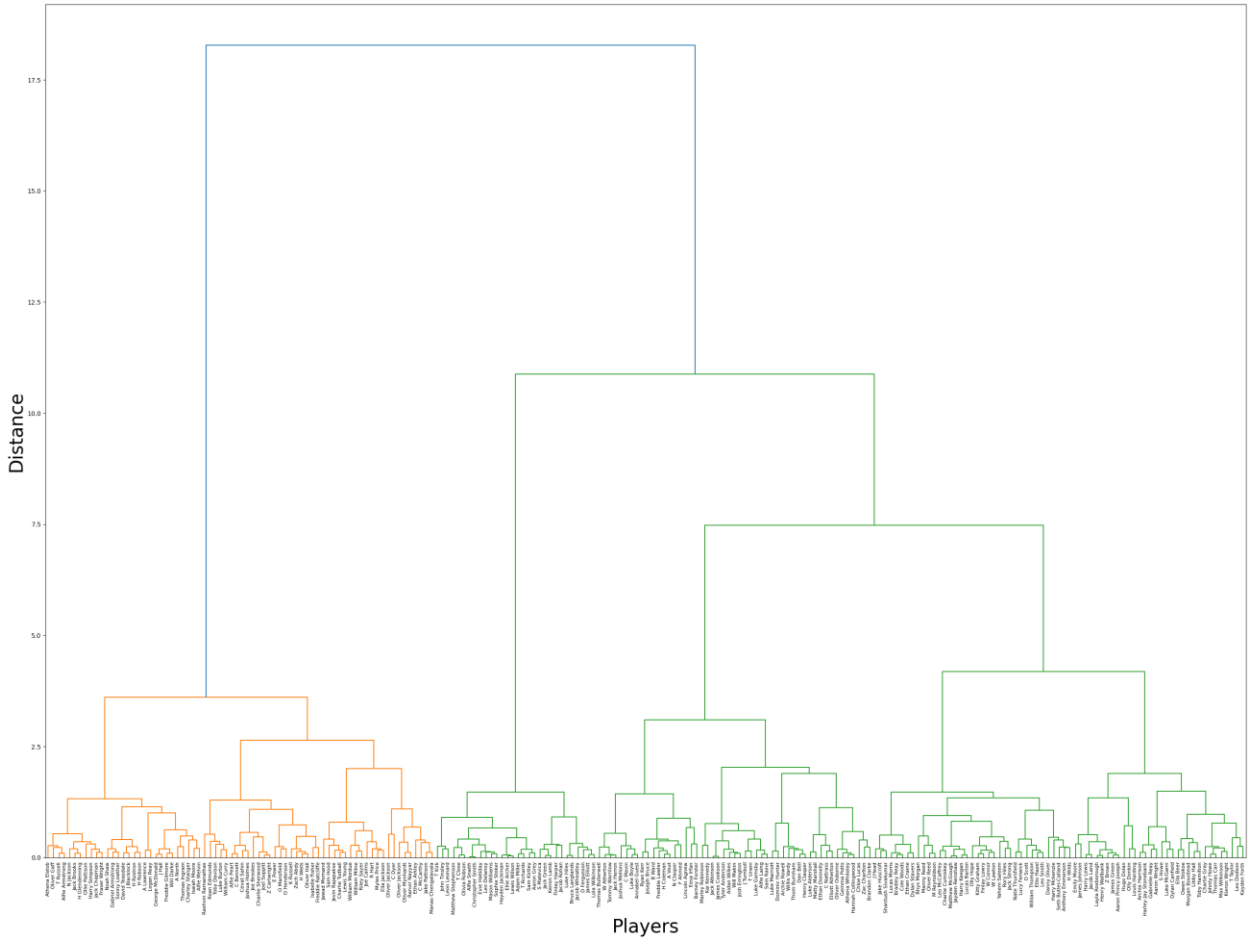
Appendix B Figure 11: Dendrogram of ball-by-ball Batting Scorecard (PCA)



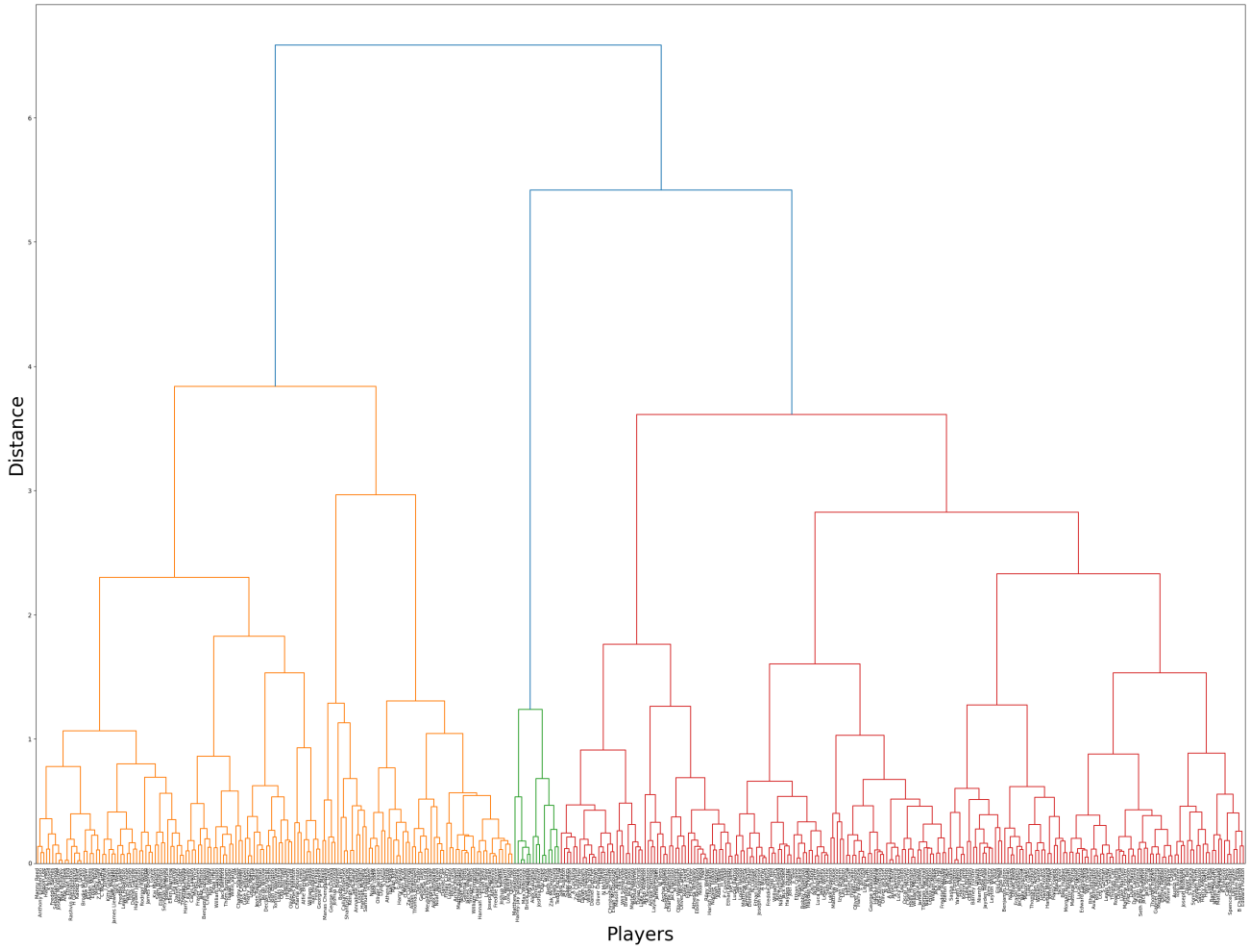
Appendix B Figure 12: Dendrogram of Encoded Dimensions of Batting Scorecard



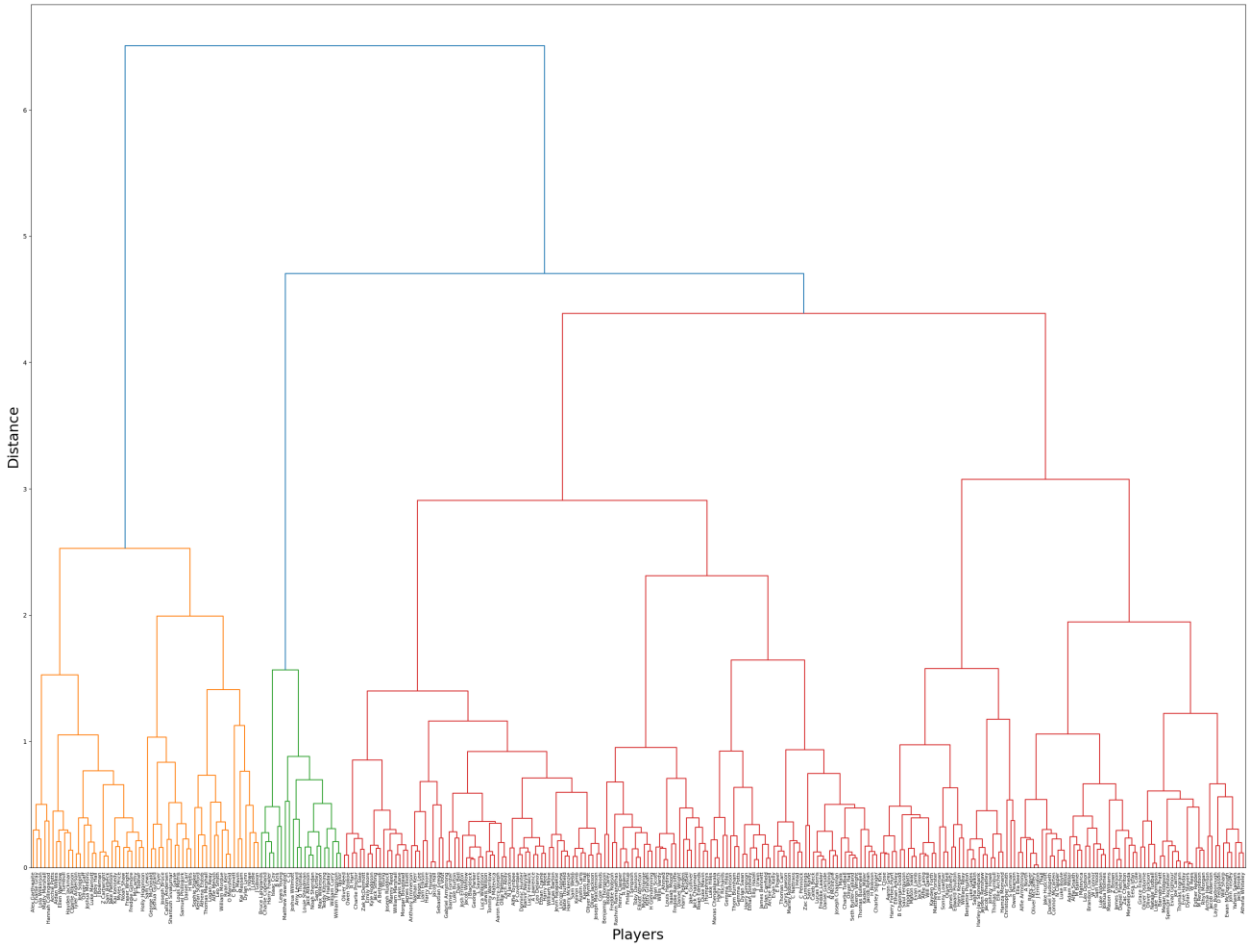
Appendix B Figure 13: Dendrogram of Encoded Dimensions of ball-by-ball Batting Scorecard



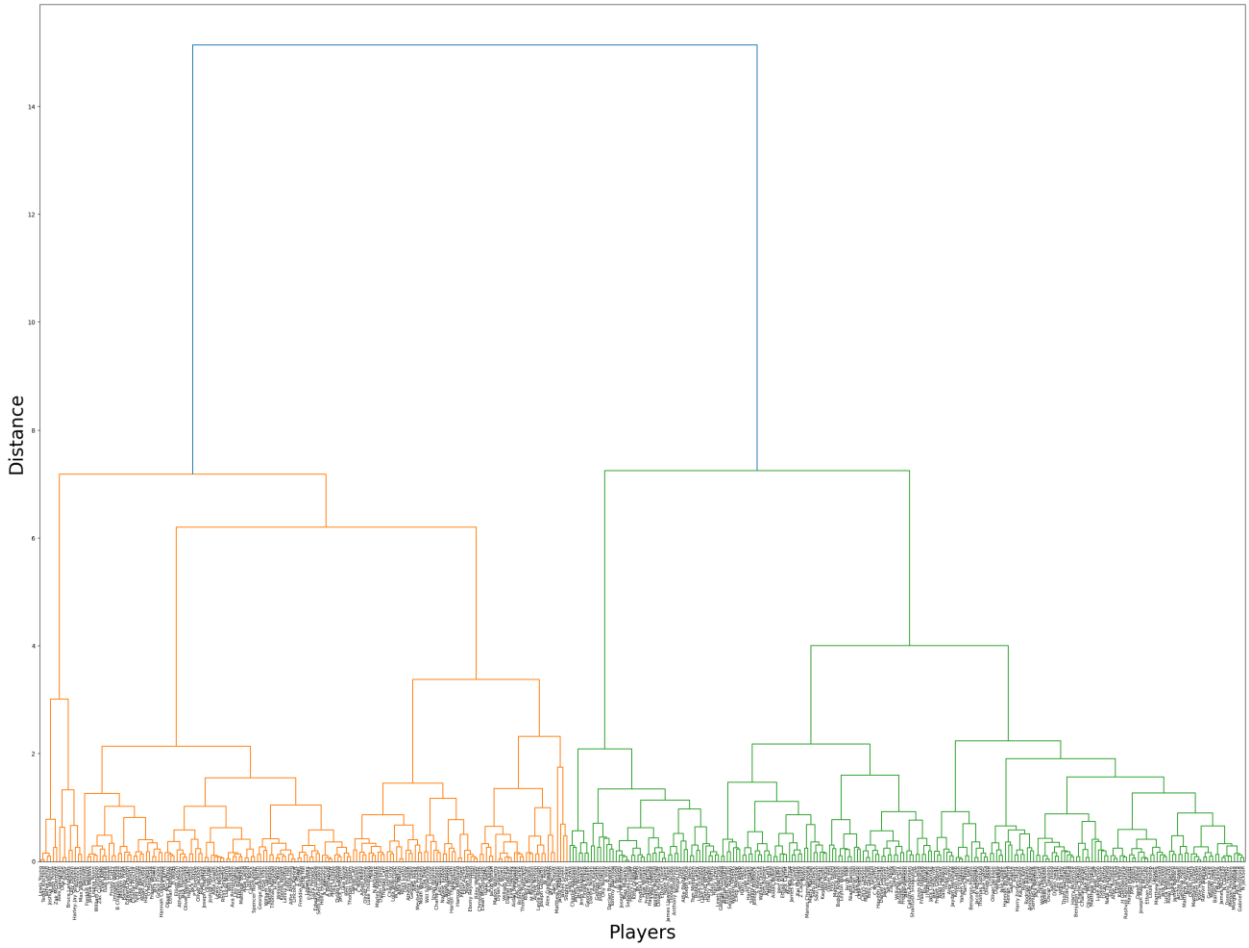
Appendix B Figure 14: Dendrogram of Bowling Scorecard (PCA)



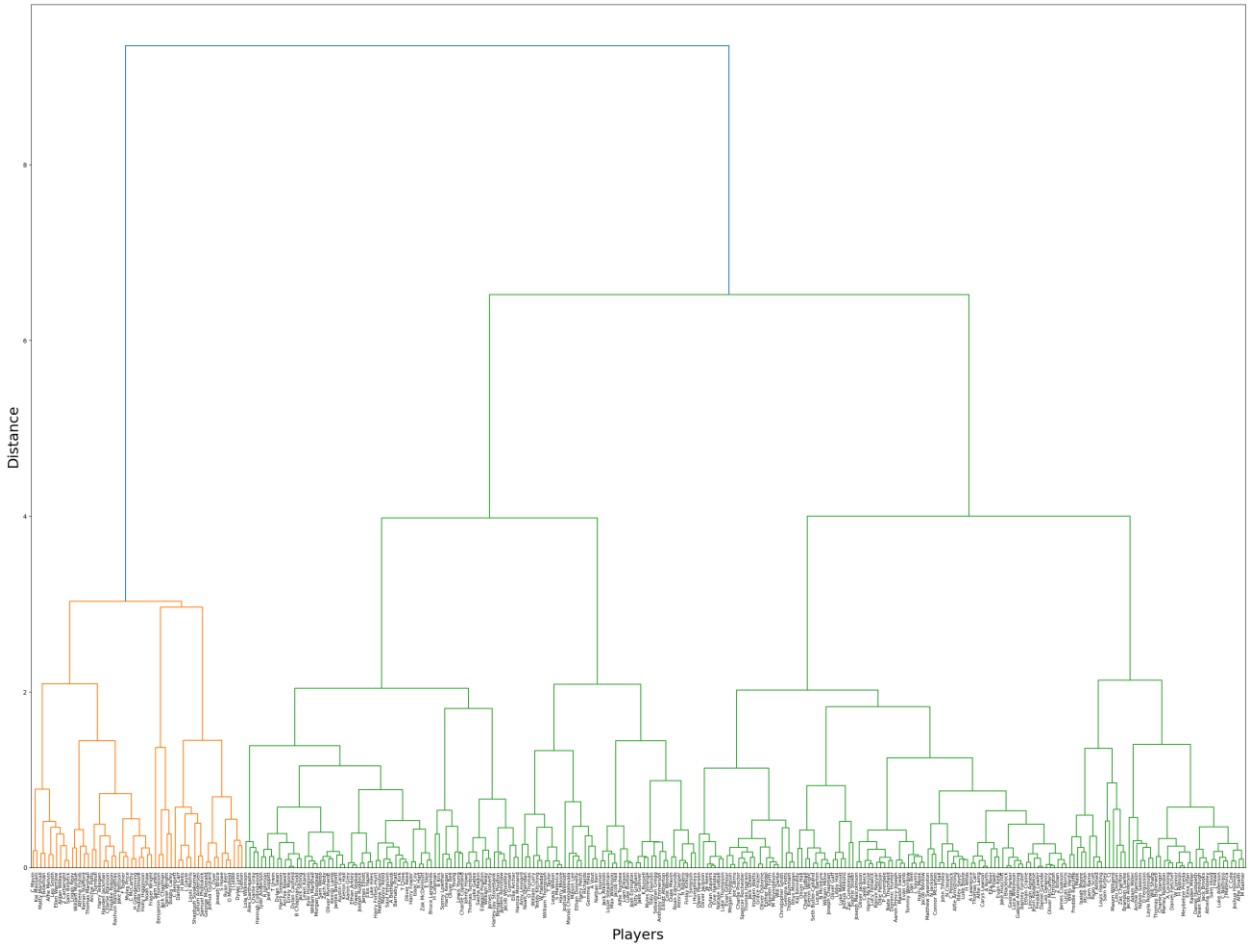
Appendix B Figure 15: Dendrogram of ball-by-ball Bowling Scorecard (PCA)



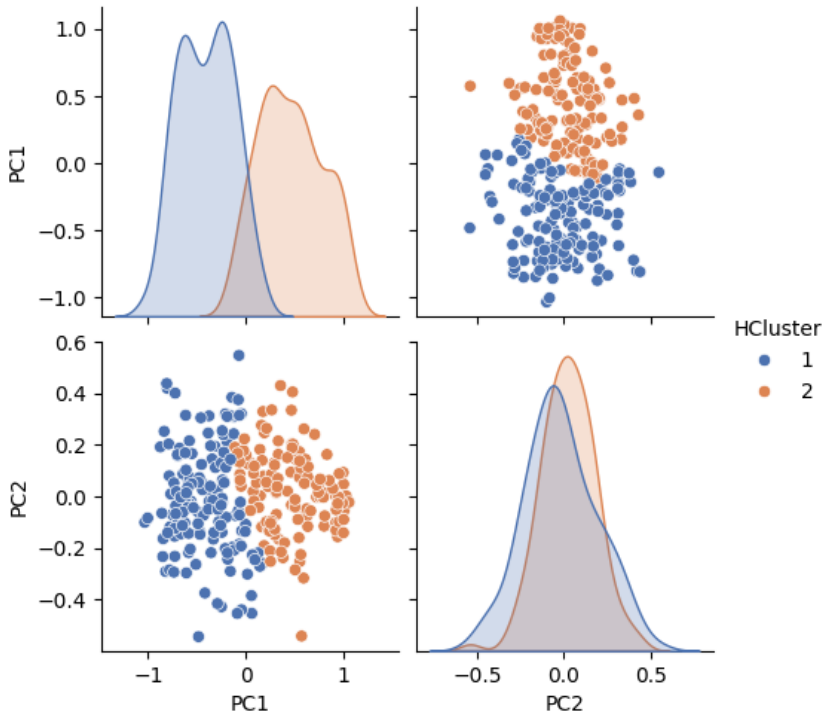
Appendix B Figure 16: Dendrogram of Encoded Dimension of Bowling Scorecard



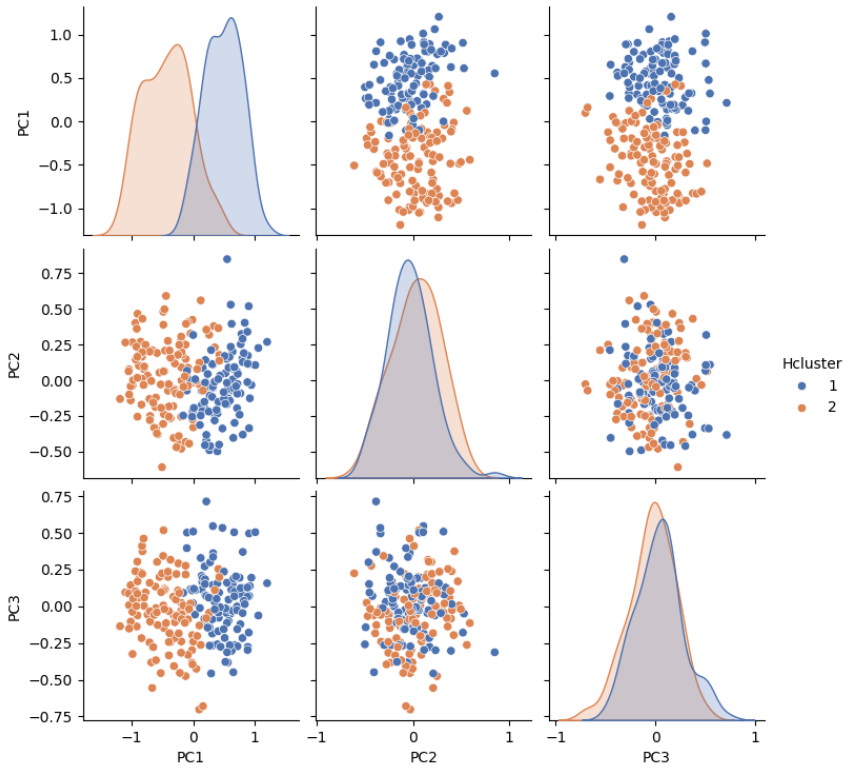
Appendix B Figure 17: Dendrogram of Encoded Dimension of ball-by-ball Bowling Scorecard



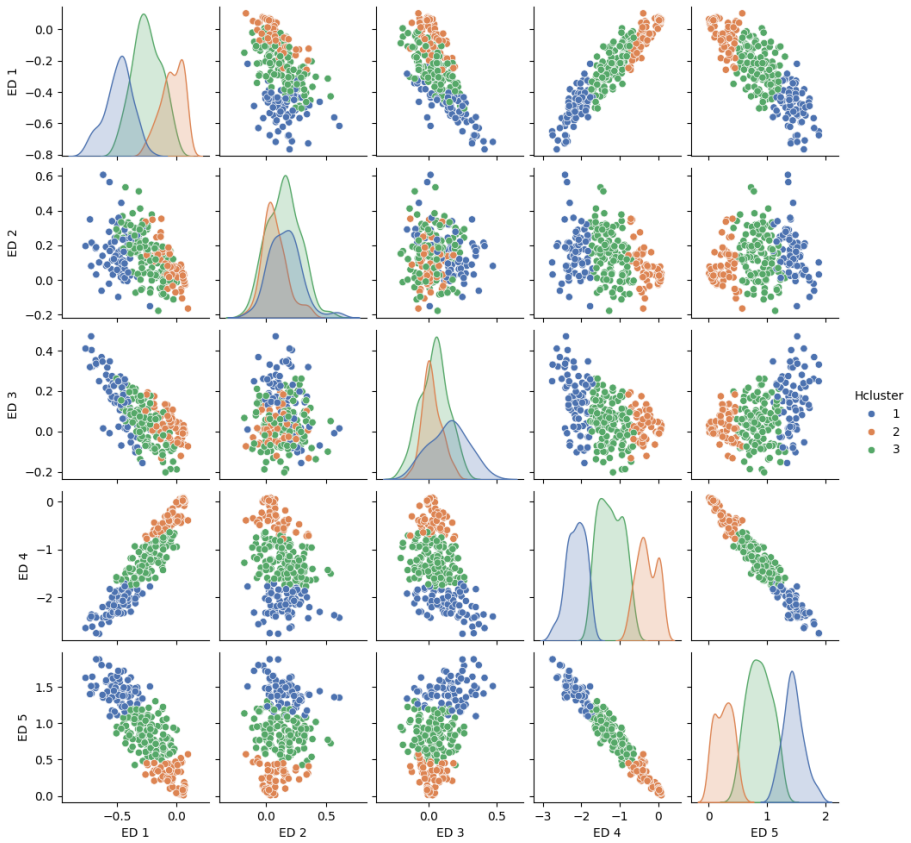
Appendix B Figure 18: Hierarchical Clustering of Batting Scorecard (PCA)



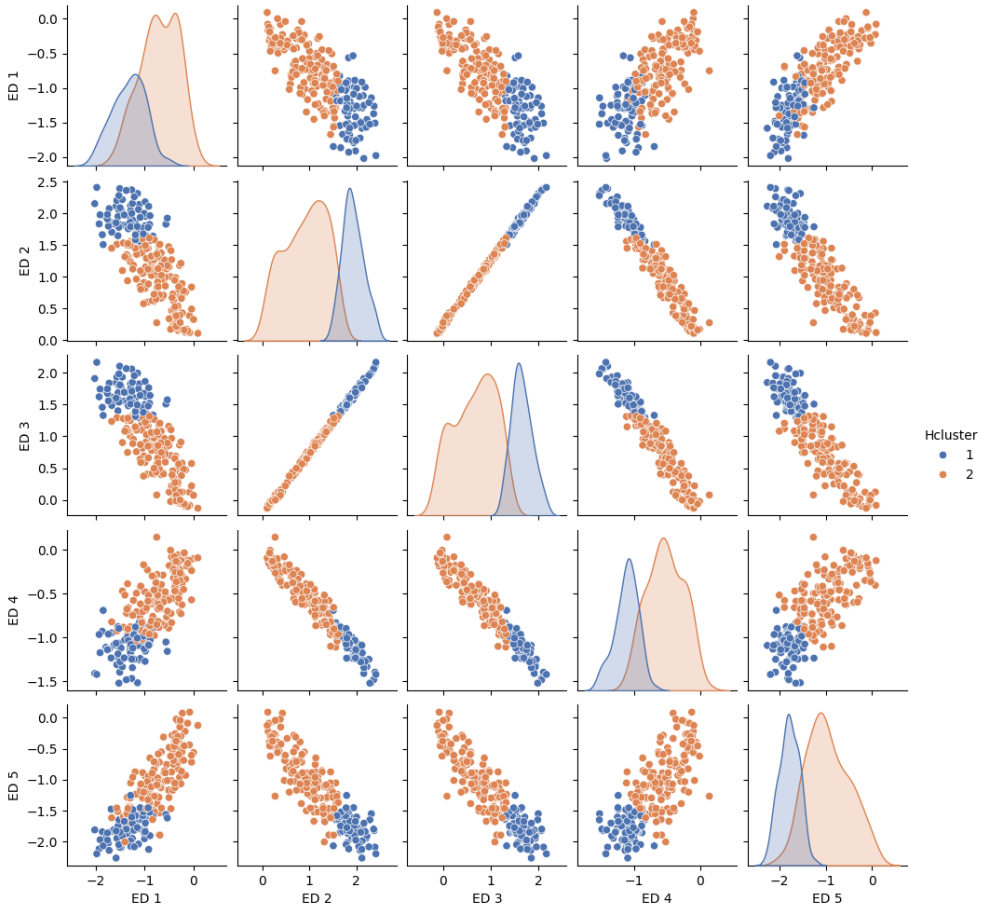
Appendix B Figure 19: Hierarchical Clustering of ball-by-ball Batting Scorecard (PCA)



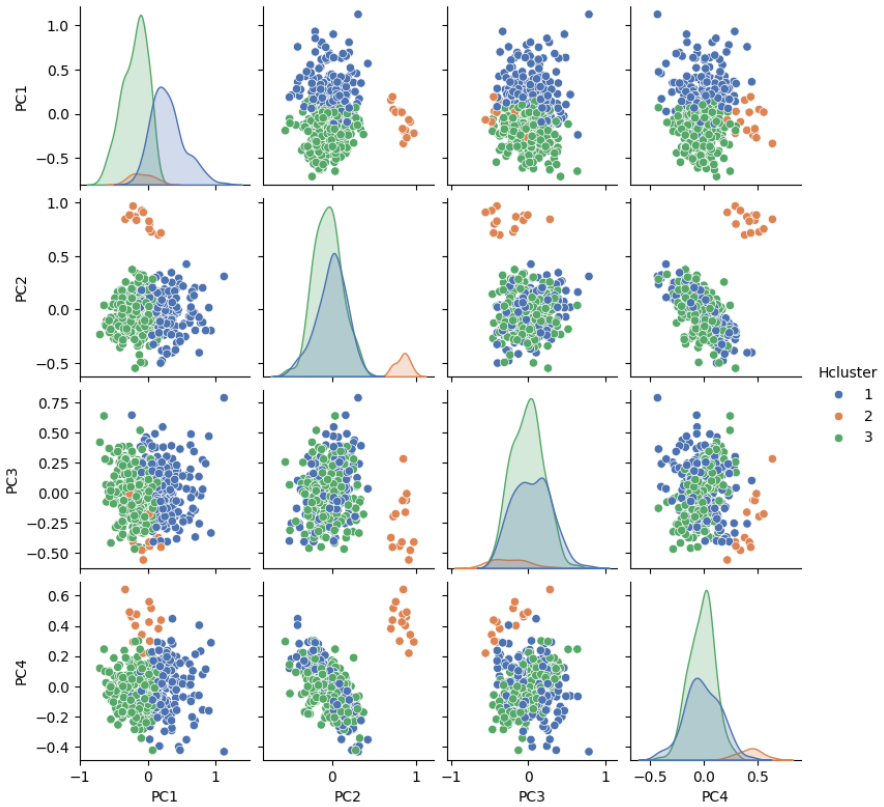
Appendix B Figure 20: Hierarchical Clustering of Encoded Dimension of Batting Scorecard



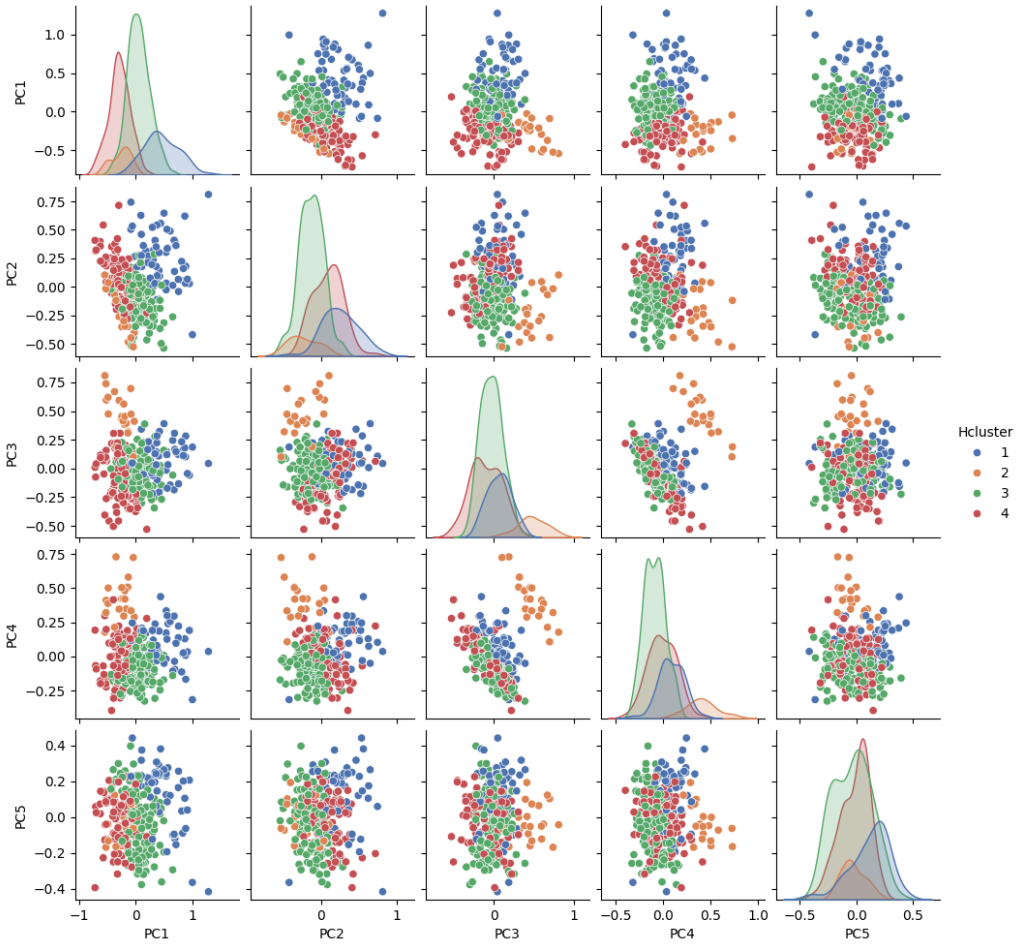
Appendix B Figure 21: Hierarchical Clustering of Encoded Dimension of ball-by-ball Batting Scorecard



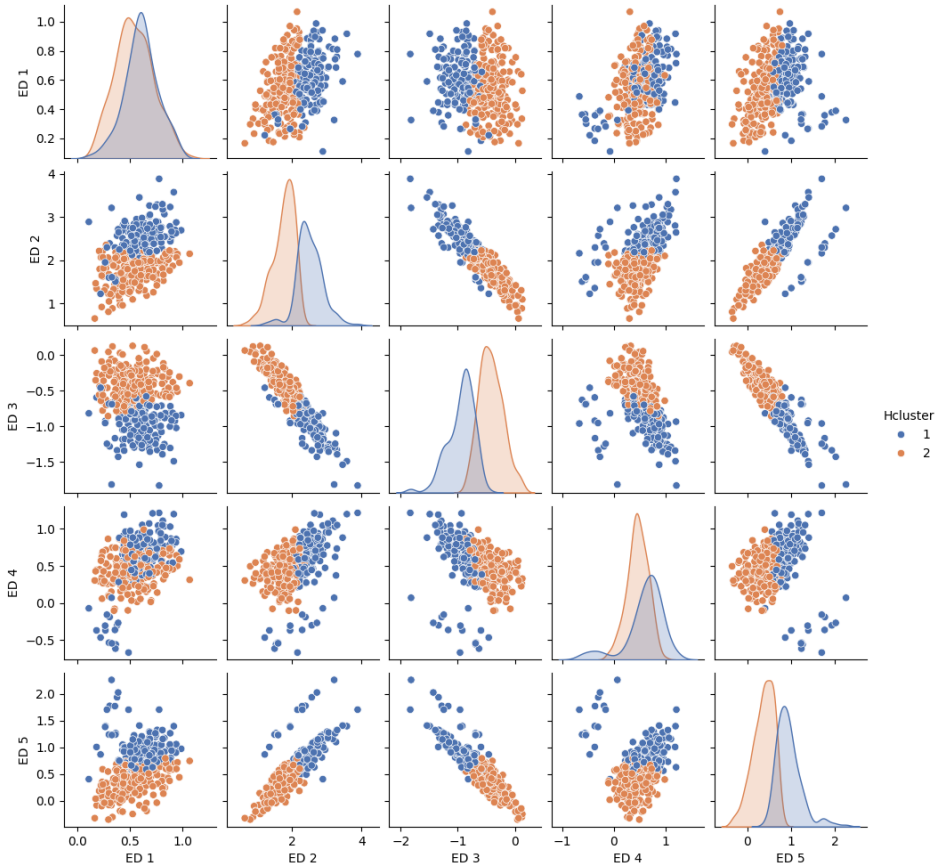
Appendix B Figure 22: Hierarchical Clustering of Bowling Scorecard (PCA)



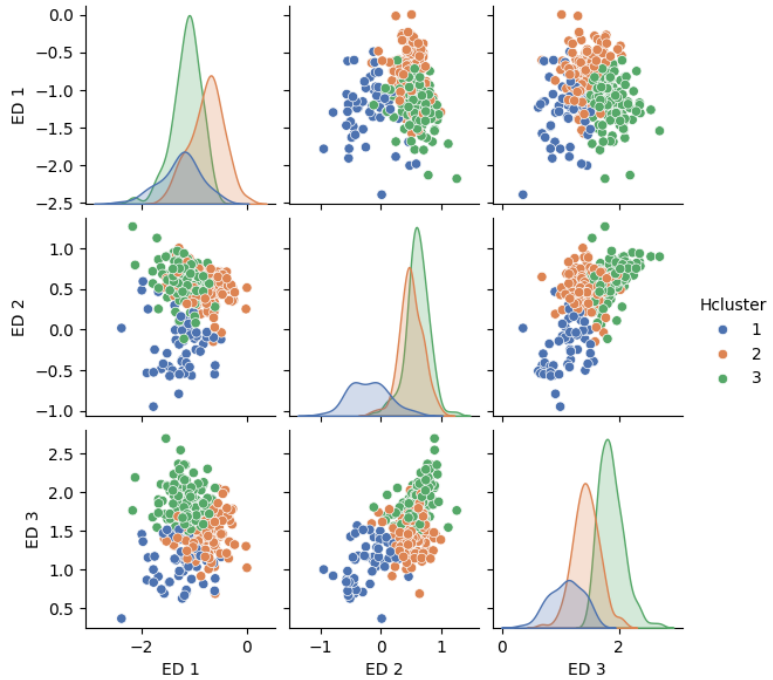
Appendix B Figure 23: Hierarchical Clustering of ball-by-ball Bowling Scorecard (PCA)



Appendix B Figure 24: Hierarchical Clustering of Encoded Dimension of Bowling Scorecard



Appendix B Figure 25: Hierarchical Clustering of Encoded Dimension of ball-by-ball Bowling Scorecard



Appendix C

Appendix C Table 1: The means summary of derived performance metrics cluster-wise for the Batting Scorecard (PCA)

KCluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio
0	0.548336	0.666733	0.570069	0.684264	0.462938	0.707607	0.418363
1	0.361147	0.273395	0.25354	0.256005	0.297184	0.251591	0.106965

Appendix C Table 2: Means summary of derived performance metrics cluster-wise for the ball-by-ball Batting Scorecard (PCA) - Hierarchical Clustering

Hcluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio	DBI	FS
1	0.486698	0.747409	0.654109	0.718105	0.650469	0.703854	0.41423	0.450383	0.529634
2	0.380099	0.332508	0.322734	0.302579	0.380081	0.281602	0.126199	0.484144	0.325253

Appendix C Table 3: Means summary of derived performance metrics cluster-wise for Batting Scorecard with encoded dimensions - Kmeans++ Clustering

KCluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio
0	0.551166	0.672221	0.571371	0.686939	0.460749	0.703951	0.422253
1	0.360806	0.273229	0.25716	0.259802	0.302539	0.263472	0.107383

Appendix C Table 4: Means summary of derived performance metrics cluster-wise for Batting Scorecard with encoded dimensions - Hierarchical Clustering

Hcluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio
1	0.60455	0.764234	0.639712	0.785263	0.481768	0.801193	0.538466
2	0.301949	0.163444	0.182584	0.157895	0.262915	0.127463	0.05487
3	0.46607	0.49883	0.430993	0.493054	0.399428	0.52314	0.24027

Appendix C Table 5: Means summary of derived performance metrics cluster-wise for ball-by-ball Batting Scorecard with encoded dimensions - KMeans++ clustering

KCluster	BFI	BI	FSS	CS	RBWI	BII	FSRatio	DBI	FS
0	0.510611	0.740751	0.632623	0.705634	0.634168	0.703895	0.409953	0.47895	0.510193
1	0.341269	0.29251	0.308804	0.268986	0.367343	0.233299	0.098047	0.45617	0.323558

Appendix C Table 6: Ranked Players in high performing cluster for ball-by-ball Batting Scorecard with encoded dimensions - K-Means++

Rank	Batter	Team	Composite Score	Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	5.79906	55	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	4.146324
2	Oliver Jackson	Castle Eden CC Under 13	5.593234	56	Isaiah Moore	Durham City CC Under 13 Blue	4.108977

3	O Henderson	Chester Le Street CC Under 13	5.455711	57	Alfie Melvin	Greenside CC Under 13	4.058258
4	Oliver Jackson	Durham City CC Under 13 Blue	5.338377	58	Luke Burton	Annfield Plain CC Under 13	4.036009
5	K Russell	Easington C.W. CC Under 13	5.291774	59	Jack Herron	Ryhope CC Under 13	4.030158
6	Oliver Hall	Hylton CCC Under 13	5.236307	60	Luke Boyle	Annfield Plain CC Under 13	4.018693
7	C Bennett	Littletown CC Under 13	5.204027	61	Oliver Gaff	Philadelphia CC Under 13	3.983437
8	Manas Chengalavala	Durham City CC Under 13 Blue	5.168781	62	Freddie Gilmore	Seaham Park CC Under 13	3.951182
9	Oliver Mynhardt	Burnmoor CC Under 13	5.151896	63	Athena Elstob	Seaham Harbour CC Under 13 Purple	3.949136
10	Joe Cairns	Lintz CC Under 13	5.132396	64	George McDonald	Burnmoor CC Under 13	3.945973
11	Alfie Peart	Castle Eden CC Under 13	5.104288	65	Joshua Mellons	Shotley Bridge CC Under 13	3.945846
12	O Mansfield	Ushaw Moor CC Under 13	5.08687	66	Marley Robinson	Consett CC Under 13	3.919863
13	Jake Pattison	Seaham Harbour CC Under 13 Yellow	5.069048	67	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	3.888812
14	Ranvir Nayyar	Durham City CC Under 13 Blue	5.039553	68	Sonny Lowther	Hetton Lyons CC Under 13 Green	3.869288
15	E Power	Willington CC Under 13	5.000981	69	H Coleman	Durham City CC Under 13 Gold	3.865713
16	Zach Abdy	Dawdon Welfare CC Under 13	4.943262	70	Charley Stanger	Greenside CC Under 13	3.858793
17	Freddie Ratcliffe	Greenside CC Under 13	4.894922	71	Logan Reay	Eppleton CC Under 13	3.83955
18	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	4.871188	72	Lincoln Agnew	Silksworth CC Under 13	3.831224
19	William Curry	Brandon CC, Durham Under 13	4.827849	73	Will Darke	Ryhope CC Under 13	3.820562
20	C McCarthy	Dawdon Welfare CC Under 13	4.791206	74	A Vout	Hetton Lyons CC Under 13 Green	3.789774
21	H Wells	Littletown CC Under 13	4.771811	75	A Lawrence	Tudhoe CC Under 13	3.754772
22	Roan Edmunds	Sacriston CC Under 13 - Black	4.732562	76	T Urwin	Easington C.W. CC Under 13	3.752164
23	Cevat Sahin	Dawdon Welfare CC Under 13	4.710403	77	Eesa Irfan	South Shields CC Under 13 B	3.751026
24	Riley Sayer	Shotley Bridge CC Under 13	4.710204	78	Thomas Bullerwell	Blaydon CC Under 13	3.730603
25	Joshua Holmes	Shotley Bridge CC Under 13	4.643411	79	David Teasdale	Seaham Park CC Under 13	3.7013
26	R Hart	Dawdon Welfare CC Under 13	4.642625	80	H Robson	Castle Eden CC Under 13	3.698383
27	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	4.63007	81	Noah Shaw	Beamish and E. Stanley CC Under 13	3.696106
28	Ella Jackson	Hetton Lyons CC Under 13 Green	4.627399	82	J Blacklock	Durham City CC Under 13 Gold	3.662795
29	William Mordue	Sunderland CC Under 13	4.613258	83	J Hall	Easington C.W. CC Under 13	3.66029
30	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	4.607554	84	W Atherton	Easington C.W. CC Under 13	3.653075
31	N Jackson	Littletown CC Under 13	4.564013	85	C Mavin	Durham City CC Under 13 Blue	3.633221

32	Z Cartwright	Littletown CC Under 13	4.551028	86	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.594964
33	Isabelle Fisher	Seaham Park CC Under 13	4.550467	87	B Ward	Durham City CC Under 13 Gold	3.590511
34	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	4.548395	88	Will Makin	Seaham Park CC Under 13	3.561397
35	Thomas Ingram	Lintz CC Under 13	4.524659	89	Ryan Kerr	Consett CC Under 13	3.55365
36	William Milne	Murton CC Under 13	4.511954	90	Josh Errington	Hetton Lyons CC Under 13 Green	3.531195
37	Joe Brookes	Raby Castle CC Under 13	4.511769	91	Dominic Hunter	Ryhope CC Under 13	3.493061
38	Jack Chapman	Burnmoor CC Under 13	4.485298	92	Sam Raine	Murton CC Under 13	3.486991
39	Jack Brooks	Burnhope CC Under 13	4.462898	93	Luke Marriott	Eppleton CC Under 13	3.363335
40	H Glendenning	Easington C.W. CC Under 13	4.445375	94	Archie Rowell	Ryton CC Under 13 Rockets	3.332866
41	Jevin Rajapakse	Sunderland CC Under 13	4.444549	95	F Ahmed	Durham City CC Under 13 Blue	3.32973
42	Toby Davison	Annfield Plain CC Under 13	4.431941	96	J Turnbull	Easington C.W. CC Under 13	3.31714
43	Harry Simpson	Chester Le Street CC Under 13	4.404106	97	Aidan Wilkes	Ushaw Moor CC Under 13	3.310777
44	T Russell	Seaham Park CC Under 13 B	4.394604	98	H Rostron	Willington CC Under 13	3.292087
45	A North	Gateshead Fell CC Under 13 A	4.385868	99	Joseph Bruce	Seaham Park CC Under 13	3.269571
46	Sam Hood	Washington CC, Durham Washington U13 Nova	4.366842	100	Jack Atkinson	Burnopfield CC Under 13	3.262642
47	Charlie Tudball	Castle Eden CC Under 13	4.344231	101	James Cookson	Burnopfield CC Under 13	3.232877
48	Charlie Sherwood	Annfield Plain CC Under 13	4.289613	102	N Clasper	Tudhoe CC Under 13	3.219151
49	Myles Price	Dawdon Welfare CC Under 13	4.287672	103	Alfie Laing	Dawdon Welfare CC Under 13	3.217343
50	Jawad Momand	Peterlee CC Under 13	4.267557	104	Kai Kennedy	Dawdon Welfare CC Under 13	3.21163
51	Alfie Armstrong	Shotley Bridge CC Under 13	4.232573	105	Willis Vardy	Murton CC Under 13	3.207296
52	Annabel Guest	Burnmoor CC Under 13	4.208769	106	Luke Gamsby	Greenside CC Under 13	3.186384
53	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.172148	107	Oliver Osborne	Peterlee CC Under 13	3.177851
54	Lewis Young	Lintz CC Under 13	4.155209	108	Tyler Anderson	Tudhoe CC Under 13	3.176838
109	Brandon Clarke	Greenside CC Under 13	3.164862				
110	Zac Charlton	Marsden CC, Tyne & Wear Under 13	3.129114				
111	Dan Bell	Beamish and E. Stanley CC Under 13	3.071295				
112	Hannah Collingwood	Shotley Bridge CC Under 13	3.012692				
113	Gemma Potts	Sunderland CC Under 13	2.991539				
114	Chloe Lucas	Seaham Harbour CC Under 13 Purple	2.940111				

115	Thom Burnham	Brandon CC, Durham Under 13	2.920571				
116	Athena Whiteley	Ryhope CC Under 13	2.917159				
117	Henry Clasper	Tudhoe CC Under 13	2.856105				

Appendix C Table 7: Ranked Players in high performing cluster for ball-by-ball Batting Scorecard with encoded dimensions - Hierarchical Clustering

Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	5.79906
2	Oliver Jackson	Castle Eden CC Under 13	5.593234
3	O Henderson	Chester Le Street CC Under 13	5.455711
4	Oliver Jackson	Durham City CC Under 13 Blue	5.338377
5	K Russell	Easington C.W. CC Under 13	5.291774
6	Oliver Hall	Hylton CCC Under 13	5.236307
7	C Bennett	Littletown CC Under 13	5.204027
8	Manas Chengalavala	Durham City CC Under 13 Blue	5.168781
9	Oliver Mynhardt	Burnmoor CC Under 13	5.151896
10	Joe Cairns	Lintz CC Under 13	5.132396
11	Alfie Peart	Castle Eden CC Under 13	5.104288
12	O Mansfield	Ushaw Moor CC Under 13	5.08687
13	Jake Pattison	Seaham Harbour CC Under 13 Yellow	5.069048
14	Ranvir Nayyar	Durham City CC Under 13 Blue	5.039553
15	E Power	Willington CC Under 13	5.000981
16	Zach Abdy	Dawdon Welfare CC Under 13	4.943262
17	Freddie Ratcliffe	Greenside CC Under 13	4.894922
18	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	4.871188
19	William Curry	Brandon CC, Durham Under 13	4.827849
20	C McCarthy	Dawdon Welfare CC Under 13	4.791206
21	H Wells	Littletown CC Under 13	4.771811
22	Roan Edmunds	Sacriston CC Under 13 - Black	4.732562
23	Cevat Sahin	Dawdon Welfare CC Under 13	4.710403
24	Riley Sayer	Shotley Bridge CC Under 13	4.710204
25	Joshua Holmes	Shotley Bridge CC Under 13	4.643411
26	R Hart	Dawdon Welfare CC Under 13	4.642625
27	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	4.63007
28	Ella Jackson	Hetton Lyons CC Under 13 Green	4.627399
29	William Mordue	Sunderland CC Under 13	4.613258
30	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	4.607554
31	N Jackson	Littletown CC Under 13	4.564013
32	Z Cartwright	Littletown CC Under 13	4.551028
33	Isabelle Fisher	Seaham Park CC Under 13	4.550467

34	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	4.548395
35	Thomas Ingram	Lintz CC Under 13	4.524659
36	William Milne	Murton CC Under 13	4.511954
37	Joe Brookes	Raby Castle CC Under 13	4.511769
38	Jack Chapman	Burnmoor CC Under 13	4.485298
39	Jack Brooks	Burnhope CC Under 13	4.462898
40	H Glendenning	Easington C.W. CC Under 13	4.445375
41	Jevin Rajapakse	Sunderland CC Under 13	4.444549
42	Toby Davison	Annfield Plain CC Under 13	4.431941
43	Harry Simpson	Chester Le Street CC Under 13	4.404106
44	T Russell	Seaham Park CC Under 13 B	4.394604
45	A North	Gateshead Fell CC Under 13 A	4.385868
46	Sam Hood	Washington CC, Durham Washington U13 Nova	4.366842
47	Charlie Tudball	Castle Eden CC Under 13	4.344231
48	Charlie Sherwood	Annfield Plain CC Under 13	4.289613
49	Myles Price	Dawdon Welfare CC Under 13	4.287672
50	Jawad Momand	Peterlee CC Under 13	4.267557
51	Alfie Armstrong	Shotley Bridge CC Under 13	4.232573
52	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.172148
53	Lewis Young	Lintz CC Under 13	4.155209
54	Isaiah Moore	Durham City CC Under 13 Blue	4.108977
55	Alfie Melvin	Greenside CC Under 13	4.058258
56	Luke Burton	Annfield Plain CC Under 13	4.036009
57	Oliver Gaff	Philadelphia CC Under 13	3.983437
58	Freddie Gilmore	Seaham Park CC Under 13	3.951182
59	Athena Elstob	Seaham Harbour CC Under 13 Purple	3.949136
60	George McDonald	Burnmoor CC Under 13	3.945973
61	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	3.888812
62	Sonny Lowther	Hetton Lyons CC Under 13 Green	3.869288
63	Charley Stanger	Greenside CC Under 13	3.858793
64	Logan Reay	Eppleton CC Under 13	3.83955
65	Will Darke	Ryhope CC Under 13	3.820562
66	A Lawrence	Tudhoe CC Under 13	3.754772
67	David Teasdale	Seaham Park CC Under 13	3.7013
68	H Robson	Castle Eden CC Under 13	3.698383
69	Noah Shaw	Beamish and E. Stanley CC Under 13	3.696106
70	J Blacklock	Durham City CC Under 13 Gold	3.662795
71	J Hall	Easington C.W. CC Under 13	3.66029
72	H Rostron	Willington CC Under 13	3.292087

Appendix C Table 8: Ranked Players in high performing cluster for ball-by-ball Batting Scorecard (PCA) - KMeans

Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	5.79906
2	Oliver Jackson	Castle Eden CC Under 13	5.593234
3	O Henderson	Chester Le Street CC Under 13	5.455711
4	Oliver Jackson	Durham City CC Under 13 Blue	5.338377
5	K Russell	Easington C.W. CC Under 13	5.291774
6	Oliver Hall	Hylton CCC Under 13	5.236307
7	C Bennett	Littletown CC Under 13	5.204027
8	Manas Chengalavala	Durham City CC Under 13 Blue	5.168781
9	Oliver Mynhardt	Burnmoor CC Under 13	5.151896
10	Joe Cairns	Lintz CC Under 13	5.132396
11	Alfie Peart	Castle Eden CC Under 13	5.104288
12	O Mansfield	Ushaw Moor CC Under 13	5.08687
13	Jake Pattison	Seaham Harbour CC Under 13 Yellow	5.069048
14	Ranvir Nayyar	Durham City CC Under 13 Blue	5.039553
15	E Power	Willington CC Under 13	5.000981
16	Zach Abdy	Dawdon Welfare CC Under 13	4.943262
17	Freddie Ratcliffe	Greenside CC Under 13	4.894922
18	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	4.871188
19	William Curry	Brandon CC, Durham Under 13	4.827849
20	C McCarthy	Dawdon Welfare CC Under 13	4.791206
21	H Wells	Littletown CC Under 13	4.771811
22	Roan Edmunds	Sacriston CC Under 13 - Black	4.732562
23	Cevat Sahin	Dawdon Welfare CC Under 13	4.710403
24	Riley Sayer	Shotley Bridge CC Under 13	4.710204
25	Joshua Holmes	Shotley Bridge CC Under 13	4.643411
26	R Hart	Dawdon Welfare CC Under 13	4.642625
27	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	4.63007
28	Ella Jackson	Hetton Lyons CC Under 13 Green	4.627399
29	William Mordue	Sunderland CC Under 13	4.613258
30	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	4.607554
31	N Jackson	Littletown CC Under 13	4.564013
32	Z Cartwright	Littletown CC Under 13	4.551028
33	Isabelle Fisher	Seaham Park CC Under 13	4.550467
34	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	4.548395
35	Thomas Ingram	Lintz CC Under 13	4.524659
36	William Milne	Murton CC Under 13	4.511954
37	Joe Brookes	Raby Castle CC Under 13	4.511769
38	Jack Chapman	Burnmoor CC Under 13	4.485298
39	Jack Brooks	Burnhope CC Under 13	4.462898
40	H Glendenning	Easington C.W. CC Under 13	4.445375

41	Jevin Rajapakse	Sunderland CC Under 13	4.444549
42	Toby Davison	Annfield Plain CC Under 13	4.431941
43	Harry Simpson	Chester Le Street CC Under 13	4.404106
44	T Russell	Seaham Park CC Under 13 B	4.394604
45	A North	Gateshead Fell CC Under 13 A	4.385868
46	Sam Hood	Washington CC, Durham Washington U13 Nova	4.366842
47	Charlie Tudball	Castle Eden CC Under 13	4.344231
48	Charlie Sherwood	Annfield Plain CC Under 13	4.289613
49	Myles Price	Dawdon Welfare CC Under 13	4.287672
50	Jawad Momand	Peterlee CC Under 13	4.267557
51	Alfie Armstrong	Shotley Bridge CC Under 13	4.232573
52	Annabel Guest	Burnmoor CC Under 13	4.208769
53	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.172148
54	Lewis Young	Lintz CC Under 13	4.155209
55	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	4.146324
56	Isaiah Moore	Durham City CC Under 13 Blue	4.108977
57	Alfie Melvin	Greenside CC Under 13	4.058258
58	Luke Burton	Annfield Plain CC Under 13	4.036009
59	Jack Herron	Ryhope CC Under 13	4.030158
60	Luke Boyle	Annfield Plain CC Under 13	4.018693
61	Oliver Gaff	Philadelphia CC Under 13	3.983437
62	Freddie Gilmore	Seaham Park CC Under 13	3.951182
63	Athena Elstob	Seaham Harbour CC Under 13 Purple	3.949136
64	George McDonald	Burnmoor CC Under 13	3.945973
65	Joshua Mellons	Shotley Bridge CC Under 13	3.945846
66	Marley Robinson	Consett CC Under 13	3.919863
67	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	3.888812
68	Sonny Lowther	Hetton Lyons CC Under 13 Green	3.869288
69	H Coleman	Durham City CC Under 13 Gold	3.865713
70	Charley Stanger	Greenside CC Under 13	3.858793
71	Logan Reay	Eppleton CC Under 13	3.83955
72	Lincoln Agnew	Silksworth CC Under 13	3.831224
73	Will Darke	Ryhope CC Under 13	3.820562
74	A Vout	Hetton Lyons CC Under 13 Green	3.789774
75	A Lawrence	Tudhoe CC Under 13	3.754772
76	T Urwin	Easington C.W. CC Under 13	3.752164
77	Eesa Irfan	South Shields CC Under 13 B	3.751026
78	Thomas Bullerwell	Blaydon CC Under 13	3.730603
79	David Teasdale	Seaham Park CC Under 13	3.7013
80	H Robson	Castle Eden CC Under 13	3.698383
81	Noah Shaw	Beamish and E. Stanley CC Under 13	3.696106
82	J Blacklock	Durham City CC Under 13 Gold	3.662795
83	J Hall	Easington C.W. CC Under 13	3.66029

84	W Atherton	Easington C.W. CC Under 13	3.653075
85	C Mavin	Durham City CC Under 13 Blue	3.633221
86	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.594964
87	B Ward	Durham City CC Under 13 Gold	3.590511
88	Will Makin	Seaham Park CC Under 13	3.561397
89	Ryan Kerr	Consett CC Under 13	3.55365
90	Josh Errington	Hetton Lyons CC Under 13 Green	3.531195
91	Dominic Hunter	Ryhope CC Under 13	3.493061
92	Sam Raine	Murton CC Under 13	3.486991
93	Luke Marriott	Eppleton CC Under 13	3.363335
94	Archie Rowell	Ryton CC Under 13 Rockets	3.332866
95	F Ahmed	Durham City CC Under 13 Blue	3.32973
96	J Turnbull	Easington C.W. CC Under 13	3.31714
97	Aidan Wilkes	Ushaw Moor CC Under 13	3.310777
98	Barney Forster	Willington CC Under 13	3.292423
99	H Rostron	Willington CC Under 13	3.292087
100	Joseph Bruce	Seaham Park CC Under 13	3.269571
101	Jack Atkinson	Burnopfield CC Under 13	3.262642
102	Danny Oliver	Tudhoe CC Under 13	3.25329
103	James Cookson	Burnopfield CC Under 13	3.232877
104	N Clasper	Tudhoe CC Under 13	3.219151
105	Alfie Laing	Dawdon Welfare CC Under 13	3.217343
106	Kai Kennedy	Dawdon Welfare CC Under 13	3.21163
107	Willis Vardy	Murton CC Under 13	3.207296
108	Luke Gamsby	Greenside CC Under 13	3.186384
109	Oliver Osborne	Peterlee CC Under 13	3.177851
110	Tyler Anderson	Tudhoe CC Under 13	3.176838
111	Brandon Clarke	Greenside CC Under 13	3.164862
112	Zac Charlton	Marsden CC, Tyne & Wear Under 13	3.129114
113	Dan Bell	Beamish and E. Stanley CC Under 13	3.071295
114	Hannah Collingwood	Shotley Bridge CC Under 13	3.012692
115	Gemma Potts	Sunderland CC Under 13	2.991539
116	Chloe Lucas	Seaham Harbour CC Under 13 Purple	2.940111
117	Thom Burnham	Brandon CC, Durham Under 13	2.920571
118	Athena Whiteley	Ryhope CC Under 13	2.917159
119	Henry Clasper	Tudhoe CC Under 13	2.856105

Appendix C Table 9: Ranked Players in high performing cluster for ball-by-ball Batting Scorecard (PCA) – Hierarchical Clustering

Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	5.79906
2	Oliver Jackson	Castle Eden CC Under 13	5.593234

3	O Henderson	Chester Le Street CC Under 13	5.455711
4	Oliver Jackson	Durham City CC Under 13 Blue	5.338377
5	K Russell	Easington C.W. CC Under 13	5.291774
6	Oliver Hall	Hylton CCC Under 13	5.236307
7	C Bennett	Littletown CC Under 13	5.204027
8	Manas Chengalavala	Durham City CC Under 13 Blue	5.168781
9	Oliver Mynhardt	Burnmoor CC Under 13	5.151896
10	Joe Cairns	Lintz CC Under 13	5.132396
11	Alfie Peart	Castle Eden CC Under 13	5.104288
12	O Mansfield	Ushaw Moor CC Under 13	5.08687
13	Jake Pattison	Seaham Harbour CC Under 13 Yellow	5.069048
14	Ranvir Nayyar	Durham City CC Under 13 Blue	5.039553
15	E Power	Willington CC Under 13	5.000981
16	Zach Abdy	Dawdon Welfare CC Under 13	4.943262
17	Freddie Ratcliffe	Greenside CC Under 13	4.894922
18	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	4.871188
19	William Curry	Brandon CC, Durham Under 13	4.827849
20	C McCarthy	Dawdon Welfare CC Under 13	4.791206
21	H Wells	Littletown CC Under 13	4.771811
22	Roan Edmunds	Sacriston CC Under 13 - Black	4.732562
23	Cevat Sahin	Dawdon Welfare CC Under 13	4.710403
24	Riley Sayer	Shotley Bridge CC Under 13	4.710204
25	Joshua Holmes	Shotley Bridge CC Under 13	4.643411
26	R Hart	Dawdon Welfare CC Under 13	4.642625
27	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	4.63007
28	Ella Jackson	Hetton Lyons CC Under 13 Green	4.627399
29	William Mordue	Sunderland CC Under 13	4.613258
30	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	4.607554
31	N Jackson	Littletown CC Under 13	4.564013
32	Z Cartwright	Littletown CC Under 13	4.551028
33	Isabelle Fisher	Seaham Park CC Under 13	4.550467
34	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	4.548395
35	Thomas Ingram	Lintz CC Under 13	4.524659
36	William Milne	Murton CC Under 13	4.511954
37	Joe Brookes	Raby Castle CC Under 13	4.511769
38	Jack Chapman	Burnmoor CC Under 13	4.485298
39	Jack Brooks	Burnhope CC Under 13	4.462898
40	H Glendenning	Easington C.W. CC Under 13	4.445375
41	Jevin Rajapakse	Sunderland CC Under 13	4.444549
42	Toby Davison	Annfield Plain CC Under 13	4.431941
43	Harry Simpson	Chester Le Street CC Under 13	4.404106
44	T Russell	Seaham Park CC Under 13 B	4.394604
45	A North	Gateshead Fell CC Under 13 A	4.385868

46	Sam Hood	Washington CC, Durham Washington U13 Nova	4.366842
47	Charlie Tudball	Castle Eden CC Under 13	4.344231
48	Charlie Sherwood	Annfield Plain CC Under 13	4.289613
49	Myles Price	Dawdon Welfare CC Under 13	4.287672
50	Jawad Momand	Peterlee CC Under 13	4.267557
51	Alfie Armstrong	Shotley Bridge CC Under 13	4.232573
52	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.172148
53	Lewis Young	Lintz CC Under 13	4.155209
54	Isaiah Moore	Durham City CC Under 13 Blue	4.108977
55	Alfie Melvin	Greenside CC Under 13	4.058258
56	Luke Burton	Annfield Plain CC Under 13	4.036009
57	Oliver Gaff	Philadelphia CC Under 13	3.983437
58	Freddie Gilmore	Seaham Park CC Under 13	3.951182
59	Athena Elstob	Seaham Harbour CC Under 13 Purple	3.949136
60	George McDonald	Burnmoor CC Under 13	3.945973
61	Marley Robinson	Consett CC Under 13	3.919863
62	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	3.888812
63	Sonny Lowther	Hetton Lyons CC Under 13 Green	3.869288
64	Charley Stanger	Greenside CC Under 13	3.858793
65	Logan Reay	Eppleton CC Under 13	3.83955
66	Lincoln Agnew	Silksworth CC Under 13	3.831224
67	Will Darke	Ryhope CC Under 13	3.820562
68	A Lawrence	Tudhoe CC Under 13	3.754772
69	T Urwin	Easington C.W. CC Under 13	3.752164
70	Thomas Bullerwell	Blaydon CC Under 13	3.730603
71	David Teasdale	Seaham Park CC Under 13	3.7013
72	H Robson	Castle Eden CC Under 13	3.698383
73	Noah Shaw	Beamish and E. Stanley CC Under 13	3.696106
74	J Blacklock	Durham City CC Under 13 Gold	3.662795
75	J Hall	Easington C.W. CC Under 13	3.66029
76	W Atherton	Easington C.W. CC Under 13	3.653075
77	C Mavin	Durham City CC Under 13 Blue	3.633221
78	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.594964
79	B Ward	Durham City CC Under 13 Gold	3.590511
80	Will Makin	Seaham Park CC Under 13	3.561397
81	Ryan Kerr	Consett CC Under 13	3.55365
82	Josh Errington	Hetton Lyons CC Under 13 Green	3.531195
83	Sam Raine	Murton CC Under 13	3.486991
84	J Turnbull	Easington C.W. CC Under 13	3.31714
85	Aidan Wilkes	Ushaw Moor CC Under 13	3.310777
86	H Rostron	Willington CC Under 13	3.292087
87	Joseph Bruce	Seaham Park CC Under 13	3.269571
88	Jack Atkinson	Burnopfield CC Under 13	3.262642

89	Danny Oliver	Tudhoe CC Under 13	3.25329
90	James Cookson	Burnopfield CC Under 13	3.232877
91	N Clasper	Tudhoe CC Under 13	3.219151
92	Alfie Laing	Dawdon Welfare CC Under 13	3.217343
93	Kai Kennedy	Dawdon Welfare CC Under 13	3.21163
94	Luke Gamsby	Greenside CC Under 13	3.186384
95	Oliver Osborne	Peterlee CC Under 13	3.177851
96	Tyler Anderson	Tudhoe CC Under 13	3.176838
97	Brandon Clarke	Greenside CC Under 13	3.164862
98	Zac Charlton	Marsden CC, Tyne & Wear Under 13	3.129114
99	Hannah Collingwood	Shotley Bridge CC Under 13	3.012692
100	Gemma Potts	Sunderland CC Under 13	2.991539
101	Leo McCaffrey	Marsden CC, Tyne & Wear Under 13	2.983847
102	Harry Mckenzie	Seaham Park CC Under 13	2.965501
103	Athena Whiteley	Ryhope CC Under 13	2.917159
104	H Wilds	Durham City CC Under 13 Blue	2.721211
105	Elliott Atherton	Seaham Park CC Under 13	2.695397

Appendix C Table 10: Ranked Players in high performing cluster for Batting Scorecard with encoded dimensions – KMeans++ Clustering

Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	4.077043
2	Oliver Jackson	Castle Eden CC Under 13	4.066798
3	Alfie Blenkiron	Bishop Auckland CC Under 13	3.977343
4	C McCarthy	Dawdon Welfare CC Under 13	3.951857
5	Manas Chengalavala	Durham City CC Under 13 Blue	3.930869
6	Jake Pattison	Seaham Harbour CC Under 13 Yellow	3.860599
7	Oliver Jackson	Durham City CC Under 13 Blue	3.813972
8	Joe Brookes	Raby Castle CC Under 13	3.783425
9	Lewis Caisley	Boldon CA CC Under 13	3.734425
10	Ranvir Nayyar	Durham City CC Under 13 Blue	3.728376
11	Joe Cairns	Lintz CC Under 13	3.709377
12	Oliver Mitcheson	Leadgate CC Under 13	3.693733
13	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	3.685854
14	O Henderson	Chester Le Street CC Under 13	3.664301
15	O Mansfield	Ushaw Moor CC Under 13	3.653313
16	E Power	Willington CC Under 13	3.643273
17	K Russell	Easington C.W. CC Under 13	3.600448
18	Oliver Mynhardt	Burnmoor CC Under 13	3.596992
19	Joshua Holmes	Shotley Bridge CC Under 13	3.585596
20	Ella Jackson	Hetton Lyons CC Under 13 Green	3.58273

21	Jacob Winlow	Sunderland CC Under 13	3.522963
22	Oliver Hall	Hylton CCC Under 13	3.515869
23	Riley Sayer	Shotley Bridge CC Under 13	3.510556
24	Jawad Momand	Peterlee CC Under 13	3.50671
25	Myles Price	Dawdon Welfare CC Under 13	3.503789
26	Sam Hood	Washington CC, Durham Washington U13 Nova	3.497047
27	Louie Graham	Consett CC Under 13	3.49404
28	William Milne	Murton CC Under 13	3.489933
29	Freddie Ratcliffe	Greenside CC Under 13	3.481576
30	R Hart	Dawdon Welfare CC Under 13	3.474571
31	Alfie Peart	Castle Eden CC Under 13	3.472458
32	William Mordue	Sunderland CC Under 13	3.457763
33	C Bennett	Littletown CC Under 13	3.440715
34	H Wells	Littletown CC Under 13	3.433387
35	Zach Abdy	Dawdon Welfare CC Under 13	3.405766
36	Bertie Morris	Raby Castle CC Under 13	3.395647
37	Danny Norman	Felling CC Under 13	3.392454
38	Cevat Sahin	Dawdon Welfare CC Under 13	3.378342
39	Dan Finlay	Lanchester CC Under 13	3.375848
40	Isabelle Fisher	Seaham Park CC Under 13	3.365507
41	Jack Brooks	Burnhope CC Under 13	3.36458
42	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.361389
43	Thomas Ingram	Lintz CC Under 13	3.35134
44	N Jackson	Littletown CC Under 13	3.349631
45	Louis Morris	Raby Castle CC Under 13	3.341727
46	Freddie Barnes	Whitburn CC Under 13 A	3.327774
47	Charlie Sherwood	Annfield Plain CC Under 13	3.322743
48	Toby Davison	Annfield Plain CC Under 13	3.290844
49	Nathan Watson	Lanchester CC Under 13	3.273427
50	Lewis Young	Lintz CC Under 13	3.254938
51	Roan Edmunds	Sacriston CC Under 13 - Black	3.250297
52	Harry Blythe	Felling CC Under 13	3.239904
53	Jevin Rajapakse	Sunderland CC Under 13	3.232057
54	Will Darke	Ryhope CC Under 13	3.214921
55	Charlie Tudball	Castle Eden CC Under 13	3.209803
56	Z Cartwright	Littletown CC Under 13	3.20309
57	Fraser Wright	Seaham Harbour CC Under 13 Yellow	3.184467
58	Charley Stanger	Greenside CC Under 13	3.173992
59	A North	Gateshead Fell CC Under 13 A	3.158199
60	Logan Reay	Eppleton CC Under 13	3.151769
61	Harry Simpson	Chester Le Street CC Under 13	3.12335
62	Oliver Gaff	Philadelphia CC Under 13	3.095298
63	H Glendenning	Easington C.W. CC Under 13	3.083213

64	William Curry	Brandon CC, Durham Under 13	3.071823
65	Isaiah Moore	Durham City CC Under 13 Blue	3.030955
66	T Russell	Seaham Park CC Under 13 B	3.021618
67	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	3.000951
68	Luke Burton	Annfield Plain CC Under 13	2.994725
69	Jay Rushton	Lanchester CC Under 13	2.990193
70	J Hall	Easington C.W. CC Under 13	2.952087
71	Athena Elstob	Seaham Harbour CC Under 13 Purple	2.941014
72	George McDonald	Burnmoor CC Under 13	2.927454
73	Alfie Melvin	Greenside CC Under 13	2.898774
74	Freddie Gilmore	Seaham Park CC Under 13	2.886725
75	L Parish	Whitburn CC Under 13 B	2.872752
76	A Lawrence	Tudhoe CC Under 13	2.868016
77	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	2.865725
78	C Mavin	Durham City CC Under 13 Blue	2.843647
79	Joseph Bruce	Seaham Park CC Under 13	2.838463
80	H Rostron	Willington CC Under 13	2.827762
81	J Blacklock	Durham City CC Under 13 Gold	2.811342
82	Rhys Seymour	Leadgate CC Under 13	2.79419
83	Jack Herron	Ryhope CC Under 13	2.775634
84	B Ward	Durham City CC Under 13 Gold	2.763032
85	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	2.74531
86	H Robson	Castle Eden CC Under 13	2.734489
87	Brandon Clarke	Greenside CC Under 13	2.728303
88	Jack Chapman	Burnmoor CC Under 13	2.723792
89	Sonny Lowther	Hetton Lyons CC Under 13 Green	2.716595
90	Lenny Rushton	Lanchester CC Under 13	2.702346
91	Luke Campbell	Leadgate CC Under 13	2.693413
92	Sam Raine	Murton CC Under 13	2.663891
93	T Urwin	Easington C.W. CC Under 13	2.661683
94	Toby Whelan	Whitburn CC Under 13 A	2.652111
95	J Turnbull	Easington C.W. CC Under 13	2.642341
96	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	2.637521
97	Noah Shaw	Beamish and E. Stanley CC Under 13	2.625416
98	Rodrigo Ribiero	Annfield Plain CC Under 13	2.604535
99	W Atherton	Easington C.W. CC Under 13	2.59711
100	Luke Boyle	Annfield Plain CC Under 13	2.596887
101	Annabel Guest	Burnmoor CC Under 13	2.594775
102	Josh Errington	Hetton Lyons CC Under 13 Green	2.5821
103	Kyle Hodgson	Murton CC Under 13	2.579843
104	Dylan Swan	Blaydon CC Under 13	2.577006
105	Harrison McCourt	Lanchester CC Under 13	2.576199
106	David Teasdale	Seaham Park CC Under 13	2.575938

107	Hannah Collingwood	Shotley Bridge CC Under 13	2.57541
108	Lincoln Agnew	Silksworth CC Under 13	2.548926
109	Archie Rowell	Ryton CC Under 13 Rockets	2.541255
110	Jack Atkinson	Burnopfield CC Under 13	2.52091
111	Joshua Mellons	Shotley Bridge CC Under 13	2.508703
112	Kieran Pemberton	Hylton CCC Under 13	2.495405
113	James Wallace	Marsden CC, Tyne & Wear Under 13	2.485283
114	Luke Marriott	Eppleton CC Under 13	2.482827
115	Kai Kennedy	Dawdon Welfare CC Under 13	2.45973
116	Dominic Hunter	Ryhope CC Under 13	2.455079
117	Alfie Laing	Dawdon Welfare CC Under 13	2.452601
118	Luke Gamsby	Greenside CC Under 13	2.451436
119	Marley Robinson	Consett CC Under 13	2.419394
120	Willis Vardy	Murton CC Under 13	2.390286
121	Alfie Armstrong	Shotley Bridge CC Under 13	2.37959
122	Thomas Bullerwell	Blaydon CC Under 13	2.366572
123	H Coleman	Durham City CC Under 13 Gold	2.364472
124	James Cookson	Burnopfield CC Under 13	2.364319
125	Harry Mckenzie	Seaham Park CC Under 13	2.340608
126	Tyler Anderson	Tudhoe CC Under 13	2.337755
127	Zac Charlton	Marsden CC, Tyne & Wear Under 13	2.329036
128	A Vout	Hetton Lyons CC Under 13 Green	2.321773
129	Chloe Lucas	Seaham Harbour CC Under 13 Purple	2.280878
130	Freddie Lawler	Langley Park & Bearpark CC Under 13's	2.274551
131	Seth Butler-Calland	Consett CC Under 13	2.273706
132	Ryan Kerr	Consett CC Under 13	2.271795
133	N Clasper	Tudhoe CC Under 13	2.248218
134	Harry Fowler	Blaydon CC Under 13	2.237009
135	Athena Whiteley	Ryhope CC Under 13	2.234999
136	Henry Clasper	Tudhoe CC Under 13	2.216527
137	Jayden Renshaw	Castle Eden CC Under 13	2.209307
138	Thom Burnham	Brandon CC, Durham Under 13	2.200375
139	Will Makin	Seaham Park CC Under 13	2.195668
140	Gemma Potts	Sunderland CC Under 13	2.184896
141	F Ahmed	Durham City CC Under 13 Blue	2.177984
142	M Reynoldson	Burnhope CC Under 13	2.177272
143	Bradley Reay	Burnhope CC Under 13	2.126116
144	Dan Bell	Beamish and E. Stanley CC Under 13	2.073619
145	Eesa Irfan	South Shields CC Under 13 B	2.058309
146	W Connor	Gateshead Fell CC Under 13 A	2.0161
147	Hayden Hunter	Consett CC Under 13	1.928554

Appendix C Table 11: Ranked Players in high performing cluster for Batting Scorecard with encoded dimensions – Hierarchical Clustering

Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	4.077043
2	Oliver Jackson	Castle Eden CC Under 13	4.066798
3	Alfie Blenkiron	Bishop Auckland CC Under 13	3.977343
4	C McCarthy	Dawdon Welfare CC Under 13	3.951857
5	Manas Chengalavala	Durham City CC Under 13 Blue	3.930869
6	Jake Pattison	Seaham Harbour CC Under 13 Yellow	3.860599
7	Oliver Jackson	Durham City CC Under 13 Blue	3.813972
8	Joe Brookes	Raby Castle CC Under 13	3.783425
9	Lewis Caisley	Boldon CA CC Under 13	3.734425
10	Ranvir Nayyar	Durham City CC Under 13 Blue	3.728376
11	Joe Cairns	Lintz CC Under 13	3.709377
12	Oliver Mitcheson	Leadgate CC Under 13	3.693733
13	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	3.685854
14	O Henderson	Chester Le Street CC Under 13	3.664301
15	O Mansfield	Ushaw Moor CC Under 13	3.653313
16	E Power	Willington CC Under 13	3.643273
17	K Russell	Easington C.W. CC Under 13	3.600448
18	Oliver Mynhardt	Burnmoor CC Under 13	3.596992
19	Joshua Holmes	Shotley Bridge CC Under 13	3.585596
20	Ella Jackson	Hetton Lyons CC Under 13 Green	3.58273
21	Jacob Winlow	Sunderland CC Under 13	3.522963
22	Oliver Hall	Hylton CCC Under 13	3.515869
23	Riley Sayer	Shotley Bridge CC Under 13	3.510556
24	Jawad Momand	Peterlee CC Under 13	3.50671
25	Myles Price	Dawdon Welfare CC Under 13	3.503789
26	Sam Hood	Washington CC, Durham Washington U13 Nova	3.497047
27	Louie Graham	Consett CC Under 13	3.49404
28	William Milne	Murton CC Under 13	3.489933
29	Freddie Ratcliffe	Greenside CC Under 13	3.481576
30	R Hart	Dawdon Welfare CC Under 13	3.474571
31	Alfie Peart	Castle Eden CC Under 13	3.472458
32	William Mordue	Sunderland CC Under 13	3.457763
33	C Bennett	Littletown CC Under 13	3.440715
34	H Wells	Littletown CC Under 13	3.433387
35	Zach Abdy	Dawdon Welfare CC Under 13	3.405766
36	Bertie Morris	Raby Castle CC Under 13	3.395647
37	Danny Norman	Felling CC Under 13	3.392454
38	Cevat Sahin	Dawdon Welfare CC Under 13	3.378342
39	Dan Finlay	Lanchester CC Under 13	3.375848

40	Isabelle Fisher	Seaham Park CC Under 13	3.365507
41	Jack Brooks	Burnhope CC Under 13	3.36458
42	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.361389
43	Thomas Ingram	Lintz CC Under 13	3.35134
44	N Jackson	Littletown CC Under 13	3.349631
45	Louis Morris	Raby Castle CC Under 13	3.341727
46	Freddie Barnes	Whitburn CC Under 13 A	3.327774
47	Charlie Sherwood	Annfield Plain CC Under 13	3.322743
48	Toby Davison	Annfield Plain CC Under 13	3.290844
49	Nathan Watson	Lanchester CC Under 13	3.273427
50	Lewis Young	Lintz CC Under 13	3.254938
51	Roan Edmunds	Sacriston CC Under 13 - Black	3.250297
52	Harry Blythe	Felling CC Under 13	3.239904
53	Jevin Rajapakse	Sunderland CC Under 13	3.232057
54	Will Darke	Ryhope CC Under 13	3.214921
55	Charlie Tudball	Castle Eden CC Under 13	3.209803
56	Z Cartwright	Littletown CC Under 13	3.20309
57	Fraser Wright	Seaham Harbour CC Under 13 Yellow	3.184467
58	Charley Stanger	Greenside CC Under 13	3.173992
59	A North	Gateshead Fell CC Under 13 A	3.158199
60	Harry Simpson	Chester Le Street CC Under 13	3.12335
61	Oliver Gaff	Philadelphia CC Under 13	3.095298
62	H Glendenning	Easington C.W. CC Under 13	3.083213
63	William Curry	Brandon CC, Durham Under 13	3.071823
64	Isaiah Moore	Durham City CC Under 13 Blue	3.030955
65	T Russell	Seaham Park CC Under 13 B	3.021618
66	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	3.000951
67	Luke Burton	Annfield Plain CC Under 13	2.994725
68	Jay Rushton	Lanchester CC Under 13	2.990193
69	J Hall	Easington C.W. CC Under 13	2.952087
70	Athena Elstob	Seaham Harbour CC Under 13 Purple	2.941014
71	Alfie Melvin	Greenside CC Under 13	2.898774
72	Freddie Gilmore	Seaham Park CC Under 13	2.886725
73	A Lawrence	Tudhoe CC Under 13	2.868016
74	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	2.865725
75	Lenny Rushton	Lanchester CC Under 13	2.702346
76	Luke Campbell	Leadgate CC Under 13	2.693413
77	David Teasdale	Seaham Park CC Under 13	2.575938

Appendix C Table 12: Ranked Players in high performing cluster for Batting Scorecard (PCA) – KMeans++ Clustering

Rank	Batter	Team	Composite Score

1	C Dyer	Littletown CC Under 13	4.077043
2	Oliver Jackson	Castle Eden CC Under 13	4.066798
3	Alfie Blenkiron	Bishop Auckland CC Under 13	3.977343
4	C McCarthy	Dawdon Welfare CC Under 13	3.951857
5	Manas Chengalavala	Durham City CC Under 13 Blue	3.930869
6	Jake Pattison	Seaham Harbour CC Under 13 Yellow	3.860599
7	Oliver Jackson	Durham City CC Under 13 Blue	3.813972
8	Joe Brookes	Raby Castle CC Under 13	3.783425
9	Lewis Caisley	Boldon CA CC Under 13	3.734425
10	Ranvir Nayyar	Durham City CC Under 13 Blue	3.728376
11	Joe Cairns	Lintz CC Under 13	3.709377
12	Oliver Mitcheson	Leadgate CC Under 13	3.693733
13	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	3.685854
14	O Henderson	Chester Le Street CC Under 13	3.664301
15	O Mansfield	Ushaw Moor CC Under 13	3.653313
16	E Power	Willington CC Under 13	3.643273
17	K Russell	Easington C.W. CC Under 13	3.600448
18	Oliver Mynhardt	Burnmoor CC Under 13	3.596992
19	Joshua Holmes	Shotley Bridge CC Under 13	3.585596
20	Ella Jackson	Hetton Lyons CC Under 13 Green	3.58273
21	Jacob Winlow	Sunderland CC Under 13	3.522963
22	Oliver Hall	Hylton CCC Under 13	3.515869
23	Riley Sayer	Shotley Bridge CC Under 13	3.510556
24	Jawad Momand	Peterlee CC Under 13	3.50671
25	Myles Price	Dawdon Welfare CC Under 13	3.503789
26	Sam Hood	Washington CC, Durham Washington U13 Nova	3.497047
27	Louie Graham	Consett CC Under 13	3.49404
28	William Milne	Murton CC Under 13	3.489933
29	Freddie Ratcliffe	Greenside CC Under 13	3.481576
30	R Hart	Dawdon Welfare CC Under 13	3.474571
31	Alfie Peart	Castle Eden CC Under 13	3.472458
32	William Mordue	Sunderland CC Under 13	3.457763
33	C Bennett	Littletown CC Under 13	3.440715
34	H Wells	Littletown CC Under 13	3.433387
35	Zach Abdy	Dawdon Welfare CC Under 13	3.405766
36	Bertie Morris	Raby Castle CC Under 13	3.395647
37	Danny Norman	Felling CC Under 13	3.392454
38	Cevat Sahin	Dawdon Welfare CC Under 13	3.378342
39	Dan Finlay	Lanchester CC Under 13	3.375848
40	Isabelle Fisher	Seaham Park CC Under 13	3.365507
41	Jack Brooks	Burnhope CC Under 13	3.36458
42	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.361389
43	Thomas Ingram	Lintz CC Under 13	3.35134

44	N Jackson	Littletown CC Under 13	3.349631
45	Louis Morris	Raby Castle CC Under 13	3.341727
46	Freddie Barnes	Whitburn CC Under 13 A	3.327774
47	Charlie Sherwood	Annfield Plain CC Under 13	3.322743
48	Toby Davison	Annfield Plain CC Under 13	3.290844
49	Nathan Watson	Lanchester CC Under 13	3.273427
50	Lewis Young	Lintz CC Under 13	3.254938
51	Roan Edmunds	Sacriston CC Under 13 - Black	3.250297
52	Harry Blythe	Felling CC Under 13	3.239904
53	Jevin Rajapakse	Sunderland CC Under 13	3.232057
54	Will Darke	Ryhope CC Under 13	3.214921
55	Charlie Tudball	Castle Eden CC Under 13	3.209803
56	Z Cartwright	Littletown CC Under 13	3.20309
57	Fraser Wright	Seaham Harbour CC Under 13 Yellow	3.184467
58	Charley Stanger	Greenside CC Under 13	3.173992
59	A North	Gateshead Fell CC Under 13 A	3.158199
60	Logan Reay	Eppleton CC Under 13	3.151769
61	Harry Simpson	Chester Le Street CC Under 13	3.12335
62	Oliver Gaff	Philadelphia CC Under 13	3.095298
63	H Glendenning	Easington C.W. CC Under 13	3.083213
64	William Curry	Brandon CC, Durham Under 13	3.071823
65	Isaiah Moore	Durham City CC Under 13 Blue	3.030955
66	T Russell	Seaham Park CC Under 13 B	3.021618
67	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	3.000951
68	Luke Burton	Annfield Plain CC Under 13	2.994725
69	Jay Rushton	Lanchester CC Under 13	2.990193
70	J Hall	Easington C.W. CC Under 13	2.952087
71	Athena Elstob	Seaham Harbour CC Under 13 Purple	2.941014
72	George McDonald	Burnmoor CC Under 13	2.927454
73	Alfie Melvin	Greenside CC Under 13	2.898774
74	Freddie Gilmore	Seaham Park CC Under 13	2.886725
75	L Parish	Whitburn CC Under 13 B	2.872752
76	A Lawrence	Tudhoe CC Under 13	2.868016
77	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	2.865725
78	C Mavin	Durham City CC Under 13 Blue	2.843647
79	Joseph Bruce	Seaham Park CC Under 13	2.838463
80	H Rostron	Willington CC Under 13	2.827762
81	J Blacklock	Durham City CC Under 13 Gold	2.811342
82	Rhys Seymour	Leadgate CC Under 13	2.79419
83	Jack Herron	Ryhope CC Under 13	2.775634
84	B Ward	Durham City CC Under 13 Gold	2.763032
85	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	2.74531
86	H Robson	Castle Eden CC Under 13	2.734489

87	Brandon Clarke	Greenside CC Under 13	2.728303
88	Jack Chapman	Burnmoor CC Under 13	2.723792
89	Sonny Lowther	Hetton Lyons CC Under 13 Green	2.716595
90	Lenny Rushton	Lanchester CC Under 13	2.702346
91	Luke Campbell	Leadgate CC Under 13	2.693413
92	Sam Raine	Murton CC Under 13	2.663891
93	T Urwin	Easington C.W. CC Under 13	2.661683
94	Toby Whelan	Whitburn CC Under 13 A	2.652111
95	J Turnbull	Easington C.W. CC Under 13	2.642341
96	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	2.637521
97	Noah Shaw	Beamish and E. Stanley CC Under 13	2.625416
98	Rodrigo Ribiero	Annfield Plain CC Under 13	2.604535
99	W Atherton	Easington C.W. CC Under 13	2.59711
100	Luke Boyle	Annfield Plain CC Under 13	2.596887
101	Annabel Guest	Burnmoor CC Under 13	2.594775
102	Josh Errington	Hetton Lyons CC Under 13 Green	2.5821
103	Kyle Hodgson	Murton CC Under 13	2.579843
104	Dylan Swan	Blaydon CC Under 13	2.577006
105	Harrison McCourt	Lanchester CC Under 13	2.576199
106	David Teasdale	Seaham Park CC Under 13	2.575938
107	Hannah Collingwood	Shotley Bridge CC Under 13	2.57541
108	Lincoln Agnew	Silksworth CC Under 13	2.548926
109	Archie Rowell	Ryton CC Under 13 Rockets	2.541255
110	Jack Atkinson	Burnopfield CC Under 13	2.52091
111	Joshua Mellons	Shotley Bridge CC Under 13	2.508703
112	Kieran Pemberton	Hylton CCC Under 13	2.495405
113	James Wallace	Marsden CC, Tyne & Wear Under 13	2.485283
114	Luke Marriott	Eppleton CC Under 13	2.482827
115	Kai Kennedy	Dawdon Welfare CC Under 13	2.45973
116	Dominic Hunter	Ryhope CC Under 13	2.455079
117	Alfie Laing	Dawdon Welfare CC Under 13	2.452601
118	Luke Gamsby	Greenside CC Under 13	2.451436
119	Marley Robinson	Consett CC Under 13	2.419394
120	Willis Vardy	Murton CC Under 13	2.390286
121	Alfie Armstrong	Shotley Bridge CC Under 13	2.37959
122	Thomas Bullerwell	Blaydon CC Under 13	2.366572
123	H Coleman	Durham City CC Under 13 Gold	2.364472
124	James Cookson	Burnopfield CC Under 13	2.364319
125	Harry Mckenzie	Seaham Park CC Under 13	2.340608
126	Tyler Anderson	Tudhoe CC Under 13	2.337755
127	Zac Charlton	Marsden CC, Tyne & Wear Under 13	2.329036
128	A Vout	Hetton Lyons CC Under 13 Green	2.321773
129	Joseph Chapman	Greenside CC Under 13	2.304622

130	Olly Norman	Felling CC Under 13	2.294395
131	Chloe Lucas	Seaham Harbour CC Under 13 Purple	2.280878
132	Freddie Lawler	Langley Park & Bearpark CC Under 13's	2.274551
133	Seth Butler-Calland	Consett CC Under 13	2.273706
134	Ryan Kerr	Consett CC Under 13	2.271795
135	N Clasper	Tudhoe CC Under 13	2.248218
136	Harry Fowler	Blaydon CC Under 13	2.237009
137	Athena Whiteley	Ryhope CC Under 13	2.234999
138	Henry Clasper	Tudhoe CC Under 13	2.216527
139	Jayden Renshaw	Castle Eden CC Under 13	2.209307
140	Thom Burnham	Brandon CC, Durham Under 13	2.200375
141	Will Makin	Seaham Park CC Under 13	2.195668
142	Gemma Potts	Sunderland CC Under 13	2.184896
143	F Ahmed	Durham City CC Under 13 Blue	2.177984
144	M Reynoldson	Burnhope CC Under 13	2.177272
145	Bradley Reay	Burnhope CC Under 13	2.126116
146	Dan Bell	Beamish and E. Stanley CC Under 13	2.073619
147	Lucas Barton	Ushaw Moor CC Under 13	2.073327
148	Eesa Irfan	South Shields CC Under 13 B	2.058309
149	Oliver Reed	Langley Park & Bearpark CC Under 13's	2.000858

Appendix C Table 13: Ranked Players in high performing cluster for Batting Scorecard (PCA) – Hierarchical Clustering

Rank	Batter	Team	Composite Score
1	C Dyer	Littletown CC Under 13	4.077043
2	Oliver Jackson	Castle Eden CC Under 13	4.066798
3	Alfie Blenkiron	Bishop Auckland CC Under 13	3.977343
4	C McCarthy	Dawdon Welfare CC Under 13	3.951857
5	Manas Chengalavala	Durham City CC Under 13 Blue	3.930869
6	Jake Pattison	Seaham Harbour CC Under 13 Yellow	3.860599
7	Oliver Jackson	Durham City CC Under 13 Blue	3.813972
8	Joe Brookes	Raby Castle CC Under 13	3.783425
9	Lewis Caisley	Boldon CA CC Under 13	3.734425
10	Ranvir Nayyar	Durham City CC Under 13 Blue	3.728376
11	Joe Cairns	Lintz CC Under 13	3.709377
12	Oliver Mitcheson	Leadgate CC Under 13	3.693733
13	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	3.685854
14	O Henderson	Chester Le Street CC Under 13	3.664301
15	O Mansfield	Ushaw Moor CC Under 13	3.653313
16	E Power	Willington CC Under 13	3.643273
17	K Russell	Easington C.W. CC Under 13	3.600448
18	Oliver Mynhardt	Burnmoor CC Under 13	3.596992

19	Joshua Holmes	Shotley Bridge CC Under 13	3.585596
20	Ella Jackson	Hetton Lyons CC Under 13 Green	3.58273
21	Jacob Winlow	Sunderland CC Under 13	3.522963
22	Oliver Hall	Hylton CCC Under 13	3.515869
23	Riley Sayer	Shotley Bridge CC Under 13	3.510556
24	Jawad Momand	Peterlee CC Under 13	3.50671
25	Myles Price	Dawdon Welfare CC Under 13	3.503789
26	Sam Hood	Washington CC, Durham Washington U13 Nova	3.497047
27	Louie Graham	Consett CC Under 13	3.49404
28	William Milne	Murton CC Under 13	3.489933
29	Freddie Ratcliffe	Greenside CC Under 13	3.481576
30	R Hart	Dawdon Welfare CC Under 13	3.474571
31	Alfie Peart	Castle Eden CC Under 13	3.472458
32	William Mordue	Sunderland CC Under 13	3.457763
33	C Bennett	Littletown CC Under 13	3.440715
34	H Wells	Littletown CC Under 13	3.433387
35	Zach Abdy	Dawdon Welfare CC Under 13	3.405766
36	Bertie Morris	Raby Castle CC Under 13	3.395647
37	Danny Norman	Felling CC Under 13	3.392454
38	Cevat Sahin	Dawdon Welfare CC Under 13	3.378342
39	Dan Finlay	Lanchester CC Under 13	3.375848
40	Isabelle Fisher	Seaham Park CC Under 13	3.365507
41	Jack Brooks	Burnhope CC Under 13	3.36458
42	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.361389
43	Thomas Ingram	Lintz CC Under 13	3.35134
44	N Jackson	Littletown CC Under 13	3.349631
45	Louis Morris	Raby Castle CC Under 13	3.341727
46	Freddie Barnes	Whitburn CC Under 13 A	3.327774
47	Charlie Sherwood	Annfield Plain CC Under 13	3.322743
48	Toby Davison	Annfield Plain CC Under 13	3.290844
49	Nathan Watson	Lanchester CC Under 13	3.273427
50	Lewis Young	Lintz CC Under 13	3.254938
51	Roan Edmunds	Sacriston CC Under 13 - Black	3.250297
52	Harry Blythe	Felling CC Under 13	3.239904
53	Jevin Rajapakse	Sunderland CC Under 13	3.232057
54	Will Darke	Ryhope CC Under 13	3.214921
55	Charlie Tudball	Castle Eden CC Under 13	3.209803
56	Z Cartwright	Littletown CC Under 13	3.20309
57	Fraser Wright	Seaham Harbour CC Under 13 Yellow	3.184467
58	Charley Stanger	Greenside CC Under 13	3.173992
59	A North	Gateshead Fell CC Under 13 A	3.158199
60	Logan Reay	Eppleton CC Under 13	3.151769
61	Harry Simpson	Chester Le Street CC Under 13	3.12335

62	Oliver Gaff	Philadelphia CC Under 13	3.095298
63	H Glendenning	Easington C.W. CC Under 13	3.083213
64	William Curry	Brandon CC, Durham Under 13	3.071823
65	Isaiah Moore	Durham City CC Under 13 Blue	3.030955
66	T Russell	Seaham Park CC Under 13 B	3.021618
67	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	3.000951
68	Luke Burton	Annfield Plain CC Under 13	2.994725
69	Jay Rushton	Lanchester CC Under 13	2.990193
70	J Hall	Easington C.W. CC Under 13	2.952087
71	Athena Elstob	Seaham Harbour CC Under 13 Purple	2.941014
72	George McDonald	Burnmoor CC Under 13	2.927454
73	Alfie Melvin	Greenside CC Under 13	2.898774
74	Freddie Gilmore	Seaham Park CC Under 13	2.886725
75	L Parish	Whitburn CC Under 13 B	2.872752
76	A Lawrence	Tudhoe CC Under 13	2.868016
77	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	2.865725
78	C Mavin	Durham City CC Under 13 Blue	2.843647
79	Joseph Bruce	Seaham Park CC Under 13	2.838463
80	H Rostron	Willington CC Under 13	2.827762
81	J Blacklock	Durham City CC Under 13 Gold	2.811342
82	Rhys Seymour	Leadgate CC Under 13	2.79419
83	Jack Herron	Ryhope CC Under 13	2.775634
84	B Ward	Durham City CC Under 13 Gold	2.763032
85	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	2.74531
86	H Robson	Castle Eden CC Under 13	2.734489
87	Brandon Clarke	Greenside CC Under 13	2.728303
88	Jack Chapman	Burnmoor CC Under 13	2.723792
89	Sonny Lowther	Hetton Lyons CC Under 13 Green	2.716595
90	Lenny Rushton	Lanchester CC Under 13	2.702346
91	Luke Campbell	Leadgate CC Under 13	2.693413
92	Sam Raine	Murton CC Under 13	2.663891
93	T Urwin	Easington C.W. CC Under 13	2.661683
94	Toby Whelan	Whitburn CC Under 13 A	2.652111
95	J Turnbull	Easington C.W. CC Under 13	2.642341
96	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	2.637521
97	Noah Shaw	Beamish and E. Stanley CC Under 13	2.625416
98	Rodrigo Ribiero	Annfield Plain CC Under 13	2.604535
99	W Atherton	Easington C.W. CC Under 13	2.59711
100	Luke Boyle	Annfield Plain CC Under 13	2.596887
101	Annabel Guest	Burnmoor CC Under 13	2.594775
102	Josh Errington	Hetton Lyons CC Under 13 Green	2.5821
103	Kyle Hodgson	Murton CC Under 13	2.579843
104	Dylan Swan	Blaydon CC Under 13	2.577006

105	Harrison McCourt	Lanchester CC Under 13	2.576199
106	David Teasdale	Seaham Park CC Under 13	2.575938
107	Hannah Collingwood	Shotley Bridge CC Under 13	2.57541
108	Lincoln Agnew	Silksworth CC Under 13	2.548926
109	Archie Rowell	Ryton CC Under 13 Rockets	2.541255
110	Jack Atkinson	Burnopfield CC Under 13	2.52091
111	Joshua Mellons	Shotley Bridge CC Under 13	2.508703
112	James Wallace	Marsden CC, Tyne & Wear Under 13	2.485283
113	Luke Marriott	Eppleton CC Under 13	2.482827
114	Kai Kennedy	Dawdon Welfare CC Under 13	2.45973
115	Dominic Hunter	Ryhope CC Under 13	2.455079
116	Alfie Laing	Dawdon Welfare CC Under 13	2.452601
117	Luke Gamsby	Greenside CC Under 13	2.451436
118	Marley Robinson	Consett CC Under 13	2.419394
119	Willis Vardy	Murton CC Under 13	2.390286
120	Thomas Bullerwell	Blaydon CC Under 13	2.366572
121	H Coleman	Durham City CC Under 13 Gold	2.364472
122	Tyler Anderson	Tudhoe CC Under 13	2.337755
123	A Vout	Hetton Lyons CC Under 13 Green	2.321773
124	Olly Norman	Felling CC Under 13	2.294395
125	Chloe Lucas	Seaham Harbour CC Under 13 Purple	2.280878
126	Freddie Lawler	Langley Park & Bearpark CC Under 13's	2.274551
127	Harry Fowler	Blaydon CC Under 13	2.237009
128	Athena Whiteley	Ryhope CC Under 13	2.234999
129	Jayden Renshaw	Castle Eden CC Under 13	2.209307
130	Thom Burnham	Brandon CC, Durham Under 13	2.200375
131	F Ahmed	Durham City CC Under 13 Blue	2.177984
132	M Reynoldson	Burnhope CC Under 13	2.177272
133	Bradley Reay	Burnhope CC Under 13	2.126116
134	Dan Bell	Beamish and E. Stanley CC Under 13	2.073619
135	Eesa Irfan	South Shields CC Under 13 B	2.058309
136	Oliver Reed	Langley Park & Bearpark CC Under 13's	2.000858
137	Hayden Hunter	Consett CC Under 13	1.928554
138	James Llewellyn-Dixon	Consett CC Under 13	1.903246
139	Izyan Ibrahim	Wearmouth & YAV CC Under 13	1.872243
140	Barney Forster	Willington CC Under 13	1.867227
141	Charlie Cumiskey	Silksworth CC Under 13	1.830954
142	Manny Reed	Tudhoe CC Under 13	1.76409
143	O Scott	Tudhoe CC Under 13	1.70099

Appendix C Table 14: KMeans++ Clusters mean summary for Bowling Scorecard (PCA)

KCluster	WTS	CS	ES	BBI	BWTS	SPS	MI	EI	AVMS
0	0.321317	0.472083	0.427108	0.532284	0.16359	0.392924	0.449479	0.201132	0.691258
1	0	0	0.727483	0.412871	0	0	0.107816	0.341241	0
2	0.191914	0.684071	0.634711	0.543546	0.033959	0.279162	0.124969	0.349496	0.70451

Appendix C Table 15: Hierarchical Clusters mean summary for Bowling Scorecard (PCA)

Hcluster	WTS	CS	ES	BBI	BWTS	SPS	MI	EI	AVMS
1	0.334208	0.46858	0.443477	0.544411	0.185154	0.399698	0.419574	0.227713	0.678283
2	0	0	0.727483	0.412871	0	0	0.107816	0.341241	0
3	0.181174	0.68939	0.626122	0.535249	0.017172	0.272897	0.141402	0.332987	0.713731

Appendix C Table 16: Hierarchical Clusters mean summary for ball-by-ball Bowling Scorecard (PCA)

HCluster	ES	WTS	BBI	CS	BWTS	SPS	DBI	EI	AVMS	BI
1	0.41827	0.385793	0.53128	0.447125	0.546268	0.345332	0.564046	0.30768	0.471707	0.328189
2	0.615009	0	0.333667	0	0	0	0.52466	0.443607	0	0.492513
3	0.440998	0.233421	0.420647	0.566858	0	0.322268	0.61678	0.340495	0.293202	0.326657
4	0.650714	0.169415	0.483822	0.714535	0.004267	0.244328	0.431131	0.476878	0.172312	0.494919

Appendix C Table 17: KMeans++ Clusters mean summary for ball-by-ball Bowling Scorecard with encoded dimensions

KCluster	ES	WTS	BBI	CS	BWTS	SPS	DBI	EI	AVMS	BI
0	0.544178	0.188799	0.447782	0.580123	0.021448	0.260602	0.531141	0.405827	0.219296	0.411661
1	0.360104	0.405962	0.486818	0.40975	0.493198	0.377543	0.607409	0.272925	0.510339	0.282754

Appendix C Table 18: Hierachical Clusters mean summary for ball-by-ball Bowling Scorecard with encoded dimensions

Hcluster	ES	WTS	BBI	CS	BWTS	SPS	DBI	EI	AVMS	BI
1	0.348949	0.398914	0.498963	0.408373	0.483164	0.377295	0.626162	0.277816	0.515763	0.266821
2	0.533939	0.189227	0.330699	0.512058	0.025669	0.237911	0.462298	0.308969	0.201968	0.436278
3	0.556983	0.191183	0.53993	0.63697	0.021853	0.279471	0.580811	0.484062	0.231525	0.397484

Appendix C Table 19: Player Ranking in high performance clusters of Bowling Scorecard (PCA) using KMeans++ Clustering

Rank	Bowler	Team	Composite Score
1	C Bennett	Littletown CC Under 13	5.864507
2	Hannah McNally	Leadgate CC Under 13	4.94551
3	Keiran Stafford	Hylton CCC Under 13	4.89761

4	Athena Elstob	Seaham Harbour CC Under 13 Purple	4.674263
5	Sam Raine	Murton CC Under 13	4.528458
6	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.495707
7	S Howe	Seaham Park CC Under 13 B	4.470842
8	William Davison	Washington CC, Durham Washington U 13 Scotia	4.407679
9	Dan Finlay	Lanchester CC Under 13	4.386155
10	Callum Andersen	Boldon CA CC Under 13	4.364122
11	Noah Workman	Whickham CC Under 13	4.360842
12	George Tilley	Lanchester CC Under 13	4.351277
13	Eesa Irfan	South Shields CC Under 13 B	4.324049
14	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.310757
15	Zac Cummings	Eppleton CC Under 13	4.305254
16	Alfie Reay	Blaydon CC Under 13	4.277568
17	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.273253
18	O Mansfield	Ushaw Moor CC Under 13	4.271922
19	Annabelle Tiplady	Raby Castle CC Under 13	4.204387
20	C Cole	Dawdon Welfare CC Under 13	4.129889
21	Charley Stanger	Greenside CC Under 13	4.119355
22	Harry Simpson	Chester Le Street CC Under 13	4.118556
23	Grace Frazer	Burnhope CC Under 13	4.118495
24	Isah Mollik	Wearmouth & YAV CC Under 13	4.091898
25	Evan Kelly	Seaham Park CC Under 13	4.078345
26	Thom Burnham	Brandon CC, Durham Under 13	4.071944
27	Harry Forbes-Hunt	Brandon CC, Durham Under 13	4.070043
28	Oliver Gaff	Philadelphia CC Under 13	4.055114
29	Dylan Lynn	Brandon CC, Durham Under 13	4.038898
30	Lenny Rushton	Lanchester CC Under 13	4.035272
31	Fraser Stephen	Blaydon CC Under 13	4.013473
32	Kai Kennedy	Dawdon Welfare CC Under 13	4.007938
33	George McDonald	Burnmoor CC Under 13	4.006976
34	J Gibson	Littletown CC Under 13	3.991759
35	Hannah Collingwood	Shotley Bridge CC Under 13	3.988985
36	Myles Price	Dawdon Welfare CC Under 13	3.985348
37	Ryan Scott	Boldon CA CC Under 13	3.966658
38	H Robson	Castle Eden CC Under 13	3.965187
39	Ryan Kerr	Consett CC Under 13	3.961826
40	William Shields	Whitburn CC Under 13 A	3.957879
41	Bobby Carey	Durham City CC Under 13 Blue	3.942442
42	C Mavin	Durham City CC Under 13 Blue	3.926368
43	H Walker	Brandon CC, Durham Under 13	3.922017
44	Isaiah Moore	Durham City CC Under 13 Blue	3.91992
45	L Gayle	Dawdon Welfare CC Under 13	3.916618
46	Jase Liddle	Annfield Plain CC Under 13	3.91526

47	Charlie Tudball	Castle Eden CC Under 13	3.912273
48	Willis Vardy	Murton CC Under 13	3.897535
49	Louis Harding	Silksworth CC Under 13	3.893029
50	Kieron Wright	Murton CC Under 13	3.892599
51	Ray Wright	Brandon CC, Durham Under 13	3.869889
52	William Thompson	Beamish and E. Stanley CC Under 13	3.850425
53	Louie Graham	Consett CC Under 13	3.838353
54	Logan Reay	Eppleton CC Under 13	3.825136
55	Joshua Gustard	Dawdon Welfare CC Under 13	3.821338
56	Manas Chengalavala	Durham City CC Under 13 Blue	3.811518
57	James Bennett	Annfield Plain CC Under 13	3.811489
58	Archie Rowell	Ryton CC Under 13 Rockets	3.79963
59	Imogen Watts	Castle Eden CC Under 13	3.790065
60	Noah Shaw	Beamish and E. Stanley CC Under 13	3.788977
61	Oliver Jackson	Durham City CC Under 13 Blue	3.786965
62	Jack Benford	Leadgate CC Under 13	3.78696
63	Benjamin Thompson	Boldon CA CC Under 13	3.781443
64	Harry Mckenzie	Seaham Park CC Under 13	3.780102
65	F Todner	Seaham Park CC Under 13	3.7796
66	Finlay Curry	Chester Le Street CC Under 13	3.775869
67	Sam Webb	Raby Castle CC Under 13	3.764392
68	Declan Buchanan	Sacriston CC Under 13 - Red	3.743529
69	F Ahmed	Durham City CC Under 13 Blue	3.74197
70	Fraser Wright	Seaham Harbour CC Under 13 Yellow	3.736725
71	Noah Mcvittie	Annfield Plain CC Under 13	3.736041
72	T Urwin	Easington C.W. CC Under 13	3.697193
73	Luke Burns	Washington CC, Durham Washington U 13 Scotia	3.677837
74	Carys Lawson	Hetton Lyons CC Under 13 Gold	3.677776
75	Matthew Burdon	Chester Le Street CC Under 13	3.673394
76	Rodrigo Ribiero	Annfield Plain CC Under 13	3.665013
77	Tom Middleton	Whitburn CC Under 13 A	3.662063
78	Kieron Hall	Seaham Park CC Under 13	3.640633
79	Olly Norman	Felling CC Under 13	3.631495
80	Jack Chapman	Burnmoor CC Under 13	3.614535
81	C McCarthy	Dawdon Welfare CC Under 13	3.594469
82	H Wells	Littletown CC Under 13	3.565152
83	Daniel Garside	Shotley Bridge CC Under 13	3.558786
84	Z Cartwright	Littletown CC Under 13	3.55261
85	Oliver Harris	Hunwick CC Under 13	3.544797
86	Hayden Jackman	Annfield Plain CC Under 13	3.539104
87	Lewis Kane	Lintz CC Under 13	3.498402
88	William Curry	Brandon CC, Durham Under 13	3.464539
89	C Dyer	Littletown CC Under 13	3.438359

90	Jack Heckels	Sacriston CC Under 13 - Red	3.432867
91	J Turnbull	Easington C.W. CC Under 13	3.425726
92	Jack Webb	Sunderland CC Under 13	3.420071
93	James Llewellyn-Dixon	Consett CC Under 13	3.412616
94	Oliver Mitcheson	Leadgate CC Under 13	3.412033
95	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	3.411331
96	H Glendenning	Easington C.W. CC Under 13	3.400954
97	Oliver Mynhardt	Burnmoor CC Under 13	3.397359
98	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.389873
99	Myles Strange	Tudhoe CC Under 13	3.381091
100	J Humphreys	Littletown CC Under 13	3.380939
101	Dan Bell	Beamish and E. Stanley CC Under 13	3.380771
102	Luke Campbell	Leadgate CC Under 13	3.37157
103	Danny Norman	Felling CC Under 13	3.369478
104	Huw Morgan	Whitburn CC Under 13 A	3.36567
105	Kai Massam	Ryton CC Under 13 Rockets	3.36554
106	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	3.360953
107	George Smith	Whitburn CC Under 13 A	3.347118
108	Kitty Graham	Willington CC Under 13	3.346324
109	James Johnson	Lintz CC Under 13	3.343947
110	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.329597
111	Jake Gulliver	Silksworth CC Under 13	3.315693
112	Charlie Atkinson	Sacriston CC Under 13 - Black	3.299877
113	Thomas Jacobs	Bishop Auckland CC Under 13	3.298414
114	Kaden Miller	Tudhoe CC Under 13	3.274089
115	Manny Reed	Tudhoe CC Under 13	3.2739
116	Louis Morris	Raby Castle CC Under 13	3.255928
117	Harry Tellegen	Raby Castle CC Under 13	3.250295
118	Elliott Atherton	Seaham Park CC Under 13	3.250146
119	Josh Errington	Hetton Lyons CC Under 13 Green	3.245851
120	Ethan Hind	Leadgate CC Under 13	3.245499
121	Aidan Wilkes	Ushaw Moor CC Under 13	3.214309
122	Toby Davison	Annfield Plain CC Under 13	3.203637
123	Alfie Blenkiron	Bishop Auckland CC Under 13	3.198135
124	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.198065
125	Brodie McCann	Felling CC Under 13	3.174278
126	Sam Lewis	Burnmoor CC Under 13	3.15537
127	Matty Neal	Whitburn CC Under 13 A	3.155268
128	Joseph Bruce	Seaham Park CC Under 13	3.143389
129	Finlay Kirkup	Raby Castle CC Under 13	3.13578
130	Lewis Wilson	Philadelphia CC Under 13	3.11603
131	Sam Irvine	Bishop Auckland CC Under 13	3.114531
132	James Wallace	Marsden CC, Tyne & Wear Under 13	3.095443

133	Jacob Winlow	Sunderland CC Under 13	3.087472
134	Seyshan Ragulan	Sunderland CC Under 13	3.070204
135	Henry Clasper	Tudhoe CC Under 13	3.068899
136	Jevin Rajapakse	Sunderland CC Under 13	3.047607
137	Freddie Robson	Burnmoor CC Under 13	3.040262
138	Lewis Robertson	Sacriston CC Under 13 - Red	3.029274
139	Anthony Emmerson	Murton CC Under 13	3.011354
140	Mason Lee	Ushaw Moor CC Under 13	3.004654
141	Georgie Franklin	Gateshead Fell CC Under 13 B	2.995463
142	Ethan Craine	Langley Park & Bearpark CC Under 13's	2.968132
143	Noah Raine	Consett CC Under 13	2.940184
144	Harry Lewis	Sunderland CC Under 13	2.932292
145	Freddie Barnes	Whitburn CC Under 13 A	2.906231
146	K Russell	Easington C.W. CC Under 13	2.88763
147	Lily Burrell	Hetton Lyons CC Under 13 Gold	2.859701
148	Joseph McAllister	Hunwick CC Under 13	2.83102
149	Sebastian Gregg	Marsden CC, Tyne & Wear Under 13	2.715709
150	Max Mullarky	Bishop Auckland CC Under 13	2.689531
151	Daniel Wright	Burnmoor CC Under 13	2.639355

Appendix C Table 20: Player Ranking in high performance clusters of Bowling Scorecard (PCA) using Hierarchical Clustering

Rank	Bowler	Team	Composite Score
1	C Bennett	Littletown CC Under 13	5.864507
2	Hannah McNally	Leadgate CC Under 13	4.94551
3	Keiran Stafford	Hylton CCC Under 13	4.89761
4	Sophie Stoker	Seaham Harbour CC Under 13 Purple	4.701663
5	Athena Elstob	Seaham Harbour CC Under 13 Purple	4.674263
6	Alex Garthwaite	Washington CC, Durham Washington U13 Nova	4.62834
7	Sam Raine	Murton CC Under 13	4.528458
8	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.495707
9	S Howe	Seaham Park CC Under 13 B	4.470842
10	William Davison	Washington CC, Durham Washington U 13 Scotia	4.407679
11	Dan Finlay	Lanchester CC Under 13	4.386155
12	Aidan Newbould	Leadgate CC Under 13	4.370568
13	Callum Andersen	Boldon CA CC Under 13	4.364122
14	Noah Workman	Whickham CC Under 13	4.360842
15	George Tilley	Lanchester CC Under 13	4.351277
16	Eesa Irfan	South Shields CC Under 13 B	4.324049
17	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.310757
18	Zac Cummings	Eppleton CC Under 13	4.305254
19	Alfie Reay	Blaydon CC Under 13	4.277568

20	Charlie Gray	Burnhope CC Under 13	4.276566
21	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.273253
22	O Mansfield	Ushaw Moor CC Under 13	4.271922
23	Annabelle Tiplady	Raby Castle CC Under 13	4.204387
24	Meybeline Pineda	Wearmouth & YAV CC Under 13	4.193942
25	Thomas Bullerwell	Blaydon CC Under 13	4.184049
26	C Cole	Dawdon Welfare CC Under 13	4.129889
27	Charley Stanger	Greenside CC Under 13	4.119355
28	Harry Simpson	Chester Le Street CC Under 13	4.118556
29	Isah Mollik	Wearmouth & YAV CC Under 13	4.091898
30	Evan Kelly	Seaham Park CC Under 13	4.078345
31	Lola Wilkinson	Seaham Harbour CC Under 13 Purple	4.071984
32	Thom Burnham	Brandon CC, Durham Under 13	4.071944
33	Harry Forbes-Hunt	Brandon CC, Durham Under 13	4.070043
34	W Westgarth	Bishop Auckland CC Under 13	4.040719
35	Dylan Lynn	Brandon CC, Durham Under 13	4.038898
36	Fraser Stephen	Blaydon CC Under 13	4.013473
37	Kai Kennedy	Dawdon Welfare CC Under 13	4.007938
38	George McDonald	Burnmoor CC Under 13	4.006976
39	J Gibson	Littletown CC Under 13	3.991759
40	Hannah Collingwood	Shotley Bridge CC Under 13	3.988985
41	Myles Price	Dawdon Welfare CC Under 13	3.985348
42	Daniel Carr	Annfield Plain CC Under 13	3.967367
43	Ryan Scott	Boldon CA CC Under 13	3.966658
44	H Robson	Castle Eden CC Under 13	3.965187
45	Ryan Kerr	Consett CC Under 13	3.961826
46	William Shields	Whitburn CC Under 13 A	3.957879
47	Bobby Carey	Durham City CC Under 13 Blue	3.942442
48	C Mavin	Durham City CC Under 13 Blue	3.926368
49	H Walker	Brandon CC, Durham Under 13	3.922017
50	Isaiah Moore	Durham City CC Under 13 Blue	3.91992
51	L Gayle	Dawdon Welfare CC Under 13	3.916618
52	Jase Liddle	Annfield Plain CC Under 13	3.91526
53	Charlie Tudball	Castle Eden CC Under 13	3.912273
54	Willis Vardy	Murton CC Under 13	3.897535
55	Rhys Curtis	Washington CC, Durham Washington U13 Nova	3.896784
56	Louis Harding	Silksworth CC Under 13	3.893029
57	Alfie Melvin	Greenside CC Under 13	3.889868
58	Ray Wright	Brandon CC, Durham Under 13	3.869889
59	William Thompson	Beamish and E. Stanley CC Under 13	3.850425
60	Freddie Gilmore	Seaham Park CC Under 13	3.849852
61	Louie Graham	Consett CC Under 13	3.838353
62	Logan Reay	Eppleton CC Under 13	3.825136

63	Joshua Gustard	Dawdon Welfare CC Under 13	3.821338
64	Manas Chengalavala	Durham City CC Under 13 Blue	3.811518
65	James Bennett	Annfield Plain CC Under 13	3.811489
66	Archie Rowell	Ryton CC Under 13 Rockets	3.79963
67	Imogen Watts	Castle Eden CC Under 13	3.790065
68	Noah Shaw	Beamish and E. Stanley CC Under 13	3.788977
69	Oliver Jackson	Durham City CC Under 13 Blue	3.786965
70	Jack Benford	Leadgate CC Under 13	3.78696
71	Benjamin Thompson	Boldon CA CC Under 13	3.781443
72	F Todner	Seaham Park CC Under 13	3.7796
73	Finlay Curry	Chester Le Street CC Under 13	3.775869
74	Sam Webb	Raby Castle CC Under 13	3.764392
75	Joshua Mellons	Shotley Bridge CC Under 13	3.757274
76	Declan Buchanan	Sacriston CC Under 13 - Red	3.743529
77	Fraser Wright	Seaham Harbour CC Under 13 Yellow	3.736725
78	Noah Mcvittie	Annfield Plain CC Under 13	3.736041
79	E Power	Willington CC Under 13	3.71191
80	T Urwin	Easington C.W. CC Under 13	3.697193
81	Carys Lawson	Hetton Lyons CC Under 13 Gold	3.677776
82	Matthew Burdon	Chester Le Street CC Under 13	3.673394
83	Rodrigo Ribiero	Annfield Plain CC Under 13	3.665013
84	Tom Middleton	Whitburn CC Under 13 A	3.662063
85	Kieron Hall	Seaham Park CC Under 13	3.640633
86	Olly Norman	Felling CC Under 13	3.631495
87	Joseph Chapman	Greenside CC Under 13	3.616652
88	Jack Chapman	Burnmoor CC Under 13	3.614535
89	C McCarthy	Dawdon Welfare CC Under 13	3.594469
90	Oliver Hall	Hylton CCC Under 13	3.572334
91	H Wells	Littletown CC Under 13	3.565152
92	Daniel Garside	Shotley Bridge CC Under 13	3.558786
93	Louisa Nunn	Washington CC, Durham Washington U 13 Scotia	3.558191
94	Z Cartwright	Littletown CC Under 13	3.55261
95	Hayden Jackman	Annfield Plain CC Under 13	3.539104
96	Lewis Kane	Lintz CC Under 13	3.498402
97	Gemma Potts	Sunderland CC Under 13	3.476155
98	C Dyer	Littletown CC Under 13	3.438359
99	J Turnbull	Easington C.W. CC Under 13	3.425726
100	Jack Webb	Sunderland CC Under 13	3.420071
101	James Llewellyn-Dixon	Consett CC Under 13	3.412616
102	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	3.411331
103	H Glendenning	Easington C.W. CC Under 13	3.400954
104	Oliver Mynhardt	Burnmoor CC Under 13	3.397359
105	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.389873

106	Myles Strange	Tudhoe CC Under 13	3.381091
107	J Humphreys	Littletown CC Under 13	3.380939
108	Luke Campbell	Leadgate CC Under 13	3.37157
109	Danny Norman	Felling CC Under 13	3.369478
110	Huw Morgan	Whitburn CC Under 13 A	3.36567
111	Kai Massam	Ryton CC Under 13 Rockets	3.36554
112	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	3.360953
113	George Smith	Whitburn CC Under 13 A	3.347118
114	Kitty Graham	Willington CC Under 13	3.346324
115	James Johnson	Lintz CC Under 13	3.343947
116	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.329597
117	Jake Gulliver	Silksworth CC Under 13	3.315693
118	Emily Moore	Castle Eden CC Under 13	3.309421
119	Charlie Atkinson	Sacriston CC Under 13 - Black	3.299877
120	Thomas Jacobs	Bishop Auckland CC Under 13	3.298414
121	Alfie Toynbee	Sacriston CC Under 13 - Black	3.283611
122	Kaden Miller	Tudhoe CC Under 13	3.274089
123	Manny Reed	Tudhoe CC Under 13	3.2739
124	Louis Morris	Raby Castle CC Under 13	3.255928
125	Harry Tellegen	Raby Castle CC Under 13	3.250295
126	Elliott Atherton	Seaham Park CC Under 13	3.250146
127	Josh Errington	Hetton Lyons CC Under 13 Green	3.245851
128	Ethan Hind	Leadgate CC Under 13	3.245499
129	Aidan Wilkes	Ushaw Moor CC Under 13	3.214309
130	Toby Davison	Annfield Plain CC Under 13	3.203637
131	Alfie Blenkiron	Bishop Auckland CC Under 13	3.198135
132	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.198065
133	Brodie McCann	Felling CC Under 13	3.174278
134	Sam Lewis	Burnmoor CC Under 13	3.15537
135	Matty Neal	Whitburn CC Under 13 A	3.155268
136	Joseph Bruce	Seaham Park CC Under 13	3.143389
137	Finlay Kirkup	Raby Castle CC Under 13	3.13578
138	James Wallace	Marsden CC, Tyne & Wear Under 13	3.095443
139	Jacob Winlow	Sunderland CC Under 13	3.087472
140	Seyshan Ragulan	Sunderland CC Under 13	3.070204
141	Henry Clasper	Tudhoe CC Under 13	3.068899
142	Jevin Rajapakse	Sunderland CC Under 13	3.047607
143	Lewis Robertson	Sacriston CC Under 13 - Red	3.029274
144	Anthony Emmerson	Murton CC Under 13	3.011354
145	Mason Lee	Ushaw Moor CC Under 13	3.004654
146	Georgie Franklin	Gateshead Fell CC Under 13 B	2.995463
147	Harry Lewis	Sunderland CC Under 13	2.932292
148	Freddie Barnes	Whitburn CC Under 13 A	2.906231

149	K Russell	Easington C.W. CC Under 13	2.88763
150	Lily Burrell	Hetton Lyons CC Under 13 Gold	2.859701
151	Joseph McAllister	Hunwick CC Under 13	2.83102
152	Sebastian Gregg	Marsden CC, Tyne & Wear Under 13	2.715709
153	Max Mullarky	Bishop Auckland CC Under 13	2.689531
154	Daniel Wright	Burnmoor CC Under 13	2.639355

Appendix C Table 21: Player Ranking in high performance clusters of Bowling Scorecard with encoded dimensions using KMeans++ Clustering

Rank	Bowler	Team	Composite Score
1	Hannah McNally	Leadgate CC Under 13	4.94551
2	Sam Raine	Murton CC Under 13	4.528458
3	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.495707
4	Callum Andersen	Boldon CA CC Under 13	4.364122
5	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.310757
6	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.273253
7	Annabelle Tiplady	Raby Castle CC Under 13	4.204387
8	C Cole	Dawdon Welfare CC Under 13	4.129889
9	Charley Stanger	Greenside CC Under 13	4.119355
10	Harry Simpson	Chester Le Street CC Under 13	4.118556
11	Harry Forbes-Hunt	Brandon CC, Durham Under 13	4.070043
12	Oliver Gaff	Philadelphia CC Under 13	4.055114
13	Fraser Stephen	Blaydon CC Under 13	4.013473
14	Kai Kennedy	Dawdon Welfare CC Under 13	4.007938
15	George McDonald	Burnmoor CC Under 13	4.006976
16	J Gibson	Littletown CC Under 13	3.991759
17	Myles Price	Dawdon Welfare CC Under 13	3.985348
18	Ryan Kerr	Consett CC Under 13	3.961826
19	William Shields	Whitburn CC Under 13 A	3.957879
20	Bobby Carey	Durham City CC Under 13 Blue	3.942442
21	C Mavin	Durham City CC Under 13 Blue	3.926368
22	H Walker	Brandon CC, Durham Under 13	3.922017
23	Isaiah Moore	Durham City CC Under 13 Blue	3.91992
24	L Gayle	Dawdon Welfare CC Under 13	3.916618
25	Jase Liddle	Annfield Plain CC Under 13	3.91526
26	Charlie Tudball	Castle Eden CC Under 13	3.912273
27	Willis Vardy	Murton CC Under 13	3.897535
28	Louis Harding	Silksworth CC Under 13	3.893029
29	Ray Wright	Brandon CC, Durham Under 13	3.869889
30	Luke Alderson	Seaham Park CC Under 13	3.852315
31	Matthew Stokoe	Shotley Bridge CC Under 13	3.850501

32	Louie Graham	Consett CC Under 13	3.838353
33	Joshua Gustard	Dawdon Welfare CC Under 13	3.821338
34	Manas Chengalavala	Durham City CC Under 13 Blue	3.811518
35	James Bennett	Annfield Plain CC Under 13	3.811489
36	Noah Shaw	Beamish and E. Stanley CC Under 13	3.788977
37	Oliver Jackson	Durham City CC Under 13 Blue	3.786965
38	Benjamin Thompson	Boldon CA CC Under 13	3.781443
39	Harry Mckenzie	Seaham Park CC Under 13	3.780102
40	F Todner	Seaham Park CC Under 13	3.7796
41	Sam Webb	Raby Castle CC Under 13	3.764392
42	Declan Buchanan	Sacriston CC Under 13 - Red	3.743529
43	Noah Mcvittie	Annfield Plain CC Under 13	3.736041
44	T Urwin	Easington C.W. CC Under 13	3.697193
45	Luke Burns	Washington CC, Durham Washington U 13 Scotia	3.677837
46	Carys Lawson	Hetton Lyons CC Under 13 Gold	3.677776
47	Matthew Burdon	Chester Le Street CC Under 13	3.673394
48	Rodrigo Ribiero	Annfield Plain CC Under 13	3.665013
49	Tom Middleton	Whitburn CC Under 13 A	3.662063
50	Kieron Hall	Seaham Park CC Under 13	3.640633
51	Olly Norman	Felling CC Under 13	3.631495
52	Jack Chapman	Burnmoor CC Under 13	3.614535
53	C McCarthy	Dawdon Welfare CC Under 13	3.594469
54	E Slee	Seaham Park CC Under 13 B	3.573349
55	Luke Burton	Annfield Plain CC Under 13	3.573307
56	H Wells	Littletown CC Under 13	3.565152
57	Leo Smith	Silksworth CC Under 13	3.561606
58	Daniel Garside	Shotley Bridge CC Under 13	3.558786
59	Z Cartwright	Littletown CC Under 13	3.55261
60	Oliver Harris	Hunwick CC Under 13	3.544797
61	Hayden Jackman	Annfield Plain CC Under 13	3.539104
62	Amy Leonard	Burnhope CC Under 13	3.53842
63	Dylan Swan	Blaydon CC Under 13	3.518194
64	Nate Thursfield	Burnmoor CC Under 13	3.511756
65	Lewis Kane	Lintz CC Under 13	3.498402
66	Jack Carter	Hunwick CC Under 13	3.466309
67	William Curry	Brandon CC, Durham Under 13	3.464539
68	N Jackson	Littletown CC Under 13	3.452586
69	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.447292
70	Harry Meek	Langley Park & Bearpark CC Under 13's	3.444172
71	C Dyer	Littletown CC Under 13	3.438359
72	Jack Heckels	Sacriston CC Under 13 - Red	3.432867
73	Luke Marriott	Eppleton CC Under 13	3.431923
74	Indie Ash	Crook CC Under 13	3.431089

75	J Turnbull	Easington C.W. CC Under 13	3.425726
76	N Clasper	Tudhoe CC Under 13	3.42216
77	Jack Webb	Sunderland CC Under 13	3.420071
78	Sam Hood	Washington CC, Durham Washington U13 Nova	3.41856
79	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	3.416443
80	James Llewellyn-Dixon	Consett CC Under 13	3.412616
81	Oliver Mitcheson	Leadgate CC Under 13	3.412033
82	Matthew Forster	Lintz CC Under 13	3.411788
83	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	3.411331
84	H Glendenning	Easington C.W. CC Under 13	3.400954
85	Oliver Mynhardt	Burnmoor CC Under 13	3.397359
86	Jake Bailey	Hunwick CC Under 13	3.390075
87	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.389873
88	Myles Strange	Tudhoe CC Under 13	3.381091
89	J Humphreys	Littletown CC Under 13	3.380939
90	Dan Bell	Beamish and E. Stanley CC Under 13	3.380771
91	Luke Campbell	Leadgate CC Under 13	3.37157
92	Danny Norman	Felling CC Under 13	3.369478
93	Huw Morgan	Whitburn CC Under 13 A	3.36567
94	Kai Massam	Ryton CC Under 13 Rockets	3.36554
95	Hayden Hunter	Consett CC Under 13	3.364361
96	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	3.360953
97	Bobby Anderson	Sacriston CC Under 13 - Red	3.358181
98	Louie Heslop	Raby Castle CC Under 13	3.348866
99	George Smith	Whitburn CC Under 13 A	3.347118
100	Kitty Graham	Willington CC Under 13	3.346324
101	James Johnson	Lintz CC Under 13	3.343947
102	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	3.341509
103	Luca Wilkinson	Philadelphia CC Under 13	3.339864
104	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.329597
105	Kyle Hodgson	Murton CC Under 13	3.323439
106	L Sayers	Willington CC Under 13	3.322294
107	Jake Gulliver	Silksworth CC Under 13	3.315693
108	Jayden Renshaw	Castle Eden CC Under 13	3.311734
109	T Close	Willington CC Under 13	3.311636
110	Matty Marshall	Seaham Harbour CC Under 13 Yellow	3.310598
111	Matthew Bailey	Hunwick CC Under 13	3.309646
112	Emily Moore	Castle Eden CC Under 13	3.309421
113	Ranvir Nayyar	Durham City CC Under 13 Blue	3.306042
114	Saul Ferguson	Seaham Harbour CC Under 13 Yellow	3.304314
115	Charlie Atkinson	Sacriston CC Under 13 - Black	3.299877
116	Alex Storey	Beamish and E. Stanley CC Under 13	3.298524
117	Thomas Jacobs	Bishop Auckland CC Under 13	3.298414

118	Sasha Malkin	Ryhope CC Under 13	3.286033
119	Lucas Barton	Ushaw Moor CC Under 13	3.285828
120	Alfie Toynbee	Sacriston CC Under 13 - Black	3.283611
121	Kaden Miller	Tudhoe CC Under 13	3.274089
122	Manny Reed	Tudhoe CC Under 13	3.2739
123	Esther Reddel	Sunderland CC Under 13	3.262982
124	Louis Morris	Raby Castle CC Under 13	3.255928
125	Harry Tellegen	Raby Castle CC Under 13	3.250295
126	Elliott Atherton	Seaham Park CC Under 13	3.250146
127	Josh Errington	Hetton Lyons CC Under 13 Green	3.245851
128	A Murray	Bishop Auckland CC Under 13	3.245504
129	Ethan Hind	Leadgate CC Under 13	3.245499
130	Joseph Malcolmson	Dawdon Welfare CC Under 13	3.219399
131	Aidan Wilkes	Ushaw Moor CC Under 13	3.214309
132	Jack Strong	Peterlee CC Under 13	3.209202
133	Toby Davison	Annfield Plain CC Under 13	3.203637
134	Alfie Blenkiron	Bishop Auckland CC Under 13	3.198135
135	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.198065
136	Ethan Donnelly	Ushaw Moor CC Under 13	3.197851
137	R Hart	Dawdon Welfare CC Under 13	3.197701
138	Leyton Wilson	Hunwick CC Under 13	3.192246
139	Brodie McCann	Felling CC Under 13	3.174278
140	William Marshall	Annfield Plain CC Under 13	3.170166
141	Kayden Forth	Eppleton CC Under 13	3.162834
142	Sam Lewis	Burnmoor CC Under 13	3.15537
143	Matty Neal	Whitburn CC Under 13 A	3.155268
144	Niamh Barron	Hunwick CC Under 13	3.153386
145	J Bowman	Easington C.W. CC Under 13	3.152656
146	Oliver Reed	Langley Park & Bearpark CC Under 13's	3.143405
147	Joseph Bruce	Seaham Park CC Under 13	3.143389
148	Finlay Kirkup	Raby Castle CC Under 13	3.13578
149	Jacob Hall	Murton CC Under 13	3.132785
150	Lewis Wilson	Philadelphia CC Under 13	3.11603
151	Sam Irvine	Bishop Auckland CC Under 13	3.114531
152	Benjamin Poulton	Ryhope CC Under 13	3.111983
153	Ted McKie	Sacriston CC Under 13 - Red	3.099125
154	Joseph Ruddock	Marsden CC, Tyne & Wear Under 13	3.098374
155	James Wallace	Marsden CC, Tyne & Wear Under 13	3.095443
156	Jacob Winlow	Sunderland CC Under 13	3.087472
157	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	3.082622
158	Barney Forster	Willington CC Under 13	3.075683
159	Luke Gates	Ryhope CC Under 13	3.075049
160	Seyshan Ragulan	Sunderland CC Under 13	3.070204

161	Henry Clasper	Tudhoe CC Under 13	3.068899
162	Niall Gulliver	Silksworth CC Under 13	3.068604
163	Duke Mavin	Durham City CC Under 13 Blue	3.061325
164	Jevin Rajapakse	Sunderland CC Under 13	3.047607
165	George Cave	Ryton CC Under 13 Raiders	3.041736
166	J Blacklock	Durham City CC Under 13 Gold	3.040372
167	Freddie Robson	Burnmoor CC Under 13	3.040262
168	Lewis Robertson	Sacriston CC Under 13 - Red	3.029274
169	Anthony Emmerson	Murton CC Under 13	3.011354
170	Morgan Boustead	Willington CC Under 13	3.01044
171	Zach Abdy	Dawdon Welfare CC Under 13	3.005727
172	Mason Lee	Ushaw Moor CC Under 13	3.004654
173	Georgie Franklin	Gateshead Fell CC Under 13 B	2.995463
174	Ellis Hazel	Ushaw Moor CC Under 13	2.980796
175	Ethan Craine	Langley Park & Bearpark CC Under 13's	2.968132
176	Roan Edmunds	Sacriston CC Under 13 - Black	2.944701
177	Noah Raine	Consett CC Under 13	2.940184
178	Unsure	Leadgate CC Under 13	2.937419
179	Harry Lewis	Sunderland CC Under 13	2.932292
180	Mason Swan	Seaham Park CC Under 13	2.918037
181	Owen Stokoe	Ryhope CC Under 13	2.913412
182	Freddie Barnes	Whitburn CC Under 13 A	2.906231
183	Fraser Gibbons	Hylton CCC Under 13	2.899329
184	Jake Huscroft	Langley Park & Bearpark CC Under 13's	2.893974
185	K Russell	Easington C.W. CC Under 13	2.88763
186	Luke Hiles	Langley Park & Bearpark CC Under 13's	2.861429
187	Lily Burrell	Hetton Lyons CC Under 13 Gold	2.859701
188	Joseph McAllister	Hunwick CC Under 13	2.83102
189	Sonny Lowther	Hetton Lyons CC Under 13 Green	2.804835
190	Harry Petch	Dawdon Welfare CC Under 13	2.787697
191	Harry Keegan	Hetton Lyons CC Under 13 Green	2.765727
192	Austin Caris	Philadelphia CC Under 13	2.737125
193	Sebastian Gregg	Marsden CC, Tyne & Wear Under 13	2.715709
194	William Dixon	Sacriston CC Under 13 - Red	2.70952
195	A Vout	Hetton Lyons CC Under 13 Green	2.708832
196	Max Mullarky	Bishop Auckland CC Under 13	2.689531
197	Daniel Wright	Burnmoor CC Under 13	2.639355
198	Tom Bulmer	Sacriston CC Under 13 - Red	2.637893
199	Hayden O'Keefe	Bishop Auckland CC Under 13	2.614458
200	S Matevica	Tudhoe CC Under 13	2.605045
201	Oliver Bell	Peterlee CC Under 13	2.546352
202	Teddy Kitching	Langley Park & Bearpark CC Under 13's	1.40216
203	Zak McDermott	Lintz CC Under 13	1.354288

204	Beth Poole	Raby Castle CC Under 13	1.352985
205	Lewis Young	Lintz CC Under 13	1.335147
206	Jamie Lemin	Blaydon CC Under 13	0.981481
207	Joshua Willmott	Philadelphia CC Under 13	0.938097

Appendix C Table 22: Player Ranking in high performance clusters of Bowling Scorecard with encoded dimensions using Hierarchical Clustering

Rank	Bowler	Team	Composite Score
1	Hannah McNally	Leadgate CC Under 13	4.94551
2	Keiran Stafford	Hylton CCC Under 13	4.89761
3	Sam Raine	Murton CC Under 13	4.528458
4	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.495707
5	Callum Andersen	Boldon CA CC Under 13	4.364122
6	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.310757
7	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.273253
8	Annabelle Tiplady	Raby Castle CC Under 13	4.204387
9	C Cole	Dawdon Welfare CC Under 13	4.129889
10	Charley Stanger	Greenside CC Under 13	4.119355
11	Thom Burnham	Brandon CC, Durham Under 13	4.071944
12	Harry Forbes-Hunt	Brandon CC, Durham Under 13	4.070043
13	Oliver Gaff	Philadelphia CC Under 13	4.055114
14	Dylan Lynn	Brandon CC, Durham Under 13	4.038898
15	Fraser Stephen	Blaydon CC Under 13	4.013473
16	Kai Kennedy	Dawdon Welfare CC Under 13	4.007938
17	George McDonald	Burnmoor CC Under 13	4.006976
18	J Gibson	Littletown CC Under 13	3.991759
19	Myles Price	Dawdon Welfare CC Under 13	3.985348
20	Ryan Scott	Boldon CA CC Under 13	3.966658
21	Ryan Kerr	Consett CC Under 13	3.961826
22	William Shields	Whitburn CC Under 13 A	3.957879
23	Bobby Carey	Durham City CC Under 13 Blue	3.942442
24	C Mavin	Durham City CC Under 13 Blue	3.926368
25	H Walker	Brandon CC, Durham Under 13	3.922017
26	Isaiah Moore	Durham City CC Under 13 Blue	3.91992
27	L Gayle	Dawdon Welfare CC Under 13	3.916618
28	Jase Liddle	Annfield Plain CC Under 13	3.91526
29	Charlie Tudball	Castle Eden CC Under 13	3.912273
30	Willis Vardy	Murton CC Under 13	3.897535
31	Louis Harding	Silksworth CC Under 13	3.893029
32	Ray Wright	Brandon CC, Durham Under 13	3.869889
33	Luke Alderson	Seaham Park CC Under 13	3.852315

34	Matthew Stokoe	Shotley Bridge CC Under 13	3.850501
35	Louie Graham	Consett CC Under 13	3.838353
36	Joshua Gustard	Dawdon Welfare CC Under 13	3.821338
37	Manas Chengalavala	Durham City CC Under 13 Blue	3.811518
38	James Bennett	Annfield Plain CC Under 13	3.811489
39	Archie Rowell	Ryton CC Under 13 Rockets	3.79963
40	Noah Shaw	Beamish and E. Stanley CC Under 13	3.788977
41	Oliver Jackson	Durham City CC Under 13 Blue	3.786965
42	Jack Benford	Leadgate CC Under 13	3.78696
43	Benjamin Thompson	Boldon CA CC Under 13	3.781443
44	Harry Mckenzie	Seaham Park CC Under 13	3.780102
45	F Todner	Seaham Park CC Under 13	3.7796
46	Sam Webb	Raby Castle CC Under 13	3.764392
47	Declan Buchanan	Sacriston CC Under 13 - Red	3.743529
48	F Ahmed	Durham City CC Under 13 Blue	3.74197
49	T Urwin	Easington C.W. CC Under 13	3.697193
50	Luke Burns	Washington CC, Durham Washington U 13 Scotia	3.677837
51	Carys Lawson	Hetton Lyons CC Under 13 Gold	3.677776
52	Matthew Burdon	Chester Le Street CC Under 13	3.673394
53	Rodrigo Ribiero	Annfield Plain CC Under 13	3.665013
54	Tom Middleton	Whitburn CC Under 13 A	3.662063
55	Kieron Hall	Seaham Park CC Under 13	3.640633
56	Olly Norman	Felling CC Under 13	3.631495
57	James Cookson	Burnopfield CC Under 13	3.616986
58	Jack Chapman	Burnmoor CC Under 13	3.614535
59	C McCarthy	Dawdon Welfare CC Under 13	3.594469
60	Lucas Copping	Crook CC Under 13	3.592825
61	E Slee	Seaham Park CC Under 13 B	3.573349
62	Luke Burton	Annfield Plain CC Under 13	3.573307
63	H Wells	Littletown CC Under 13	3.565152
64	Yahem Saleem	Eppleton CC Under 13	3.565123
65	Leo Smith	Silksworth CC Under 13	3.561606
66	Daniel Garside	Shotley Bridge CC Under 13	3.558786
67	Z Cartwright	Littletown CC Under 13	3.55261
68	J Errington	Willington CC Under 13	3.545954
69	Oliver Harris	Hunwick CC Under 13	3.544797
70	Keiron Lamb	Philadelphia CC Under 13	3.542441
71	Hayden Jackman	Annfield Plain CC Under 13	3.539104
72	Amy Leonard	Burnhope CC Under 13	3.53842
73	Dylan Swan	Blaydon CC Under 13	3.518194
74	Nate Thursfield	Burnmoor CC Under 13	3.511756
75	Lewis Kane	Lintz CC Under 13	3.498402
76	Oscar Davison	Consett CC Under 13	3.496709

77	Gemma Potts	Sunderland CC Under 13	3.476155
78	Jack Carter	Hunwick CC Under 13	3.466309
79	William Curry	Brandon CC, Durham Under 13	3.464539
80	James McDevitt	Ryton CC Under 13 Raiders	3.45808
81	N Jackson	Littletown CC Under 13	3.452586
82	William Mordue	Sunderland CC Under 13	3.448494
83	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.447292
84	Harry Meek	Langley Park & Bearpark CC Under 13's	3.444172
85	C Dyer	Littletown CC Under 13	3.438359
86	Jack Heckels	Sacriston CC Under 13 - Red	3.432867
87	Indie Ash	Crook CC Under 13	3.431089
88	J Turnbull	Easington C.W. CC Under 13	3.425726
89	N Clasper	Tudhoe CC Under 13	3.42216
90	Jack Webb	Sunderland CC Under 13	3.420071
91	Sam Hood	Washington CC, Durham Washington U13 Nova	3.41856
92	Ollie Pattison	Seaham Harbour CC Under 13 Yellow	3.416443
93	James Llewellyn-Dixon	Consett CC Under 13	3.412616
94	Oliver Mitcheson	Leadgate CC Under 13	3.412033
95	Matthew Forster	Lintz CC Under 13	3.411788
96	Tommy Warrilow	Langley Park & Bearpark CC Under 13's	3.411331
97	H Glendenning	Easington C.W. CC Under 13	3.400954
98	Oliver Mynhardt	Burnmoor CC Under 13	3.397359
99	Jake Bailey	Hunwick CC Under 13	3.390075
100	Freddie Lawler	Langley Park & Bearpark CC Under 13's	3.389873
101	Dominic Hunter	Ryhope CC Under 13	3.388322
102	Megan Hodgson	Sunderland CC Under 13	3.381548
103	Myles Strange	Tudhoe CC Under 13	3.381091
104	J Humphreys	Littletown CC Under 13	3.380939
105	Dan Bell	Beamish and E. Stanley CC Under 13	3.380771
106	Luke Campbell	Leadgate CC Under 13	3.37157
107	Danny Norman	Felling CC Under 13	3.369478
108	Huw Morgan	Whitburn CC Under 13 A	3.36567
109	Kai Massam	Ryton CC Under 13 Rockets	3.36554
110	Hayden Hunter	Consett CC Under 13	3.364361
111	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	3.360953
112	Bobby Anderson	Sacriston CC Under 13 - Red	3.358181
113	Louie Heslop	Raby Castle CC Under 13	3.348866
114	George Smith	Whitburn CC Under 13 A	3.347118
115	Kitty Graham	Willington CC Under 13	3.346324
116	James Johnson	Lintz CC Under 13	3.343947
117	Gabriel Armstrong	Marsden CC, Tyne & Wear Under 13	3.341509
118	Luca Wilkinson	Philadelphia CC Under 13	3.339864
119	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.329597

120	Kyle Hodgson	Murton CC Under 13	3.323439
121	L Sayers	Willington CC Under 13	3.322294
122	Jake Gulliver	Silksworth CC Under 13	3.315693
123	Jayden Renshaw	Castle Eden CC Under 13	3.311734
124	T Close	Willington CC Under 13	3.311636
125	Matty Marshall	Seaham Harbour CC Under 13 Yellow	3.310598
126	Matthew Bailey	Hunwick CC Under 13	3.309646
127	Emily Moore	Castle Eden CC Under 13	3.309421
128	Ranvir Nayyar	Durham City CC Under 13 Blue	3.306042
129	Saul Ferguson	Seaham Harbour CC Under 13 Yellow	3.304314
130	Charlie Atkinson	Sacriston CC Under 13 - Black	3.299877
131	Alex Storey	Beamish and E. Stanley CC Under 13	3.298524
132	Thomas Jacobs	Bishop Auckland CC Under 13	3.298414
133	Sasha Malkin	Ryhope CC Under 13	3.286033
134	Lucas Barton	Ushaw Moor CC Under 13	3.285828
135	Alfie Toynbee	Sacriston CC Under 13 - Black	3.283611
136	H Coleman	Durham City CC Under 13 Gold	3.275446
137	Kaden Miller	Tudhoe CC Under 13	3.274089
138	Manny Reed	Tudhoe CC Under 13	3.2739
139	Esther Reddel	Sunderland CC Under 13	3.262982
140	Louis Morris	Raby Castle CC Under 13	3.255928
141	Harry Tellegen	Raby Castle CC Under 13	3.250295
142	Elliott Atherton	Seaham Park CC Under 13	3.250146
143	Josh Errington	Hetton Lyons CC Under 13 Green	3.245851
144	A Murray	Bishop Auckland CC Under 13	3.245504
145	Ethan Hind	Leadgate CC Under 13	3.245499
146	Joseph Malcolmson	Dawdon Welfare CC Under 13	3.219399
147	Aidan Wilkes	Ushaw Moor CC Under 13	3.214309
148	Jack Strong	Peterlee CC Under 13	3.209202
149	Toby Davison	Annfield Plain CC Under 13	3.203637
150	Alfie Blenkiron	Bishop Auckland CC Under 13	3.198135
151	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.198065
152	Ethan Donnelly	Ushaw Moor CC Under 13	3.197851
153	R Hart	Dawdon Welfare CC Under 13	3.197701
154	Leyton Wilson	Hunwick CC Under 13	3.192246
155	Brodie McCann	Felling CC Under 13	3.174278
156	William Marshall	Annfield Plain CC Under 13	3.170166
157	Kayden Forth	Eppleton CC Under 13	3.162834
158	Jacob Wheatley	Peterlee CC Under 13	3.160127
159	Sam Lewis	Burnmoor CC Under 13	3.15537
160	Matty Neal	Whitburn CC Under 13 A	3.155268
161	Niamh Barron	Hunwick CC Under 13	3.153386
162	J Bowman	Easington C.W. CC Under 13	3.152656

163	Oliver Reed	Langley Park & Bearpark CC Under 13's	3.143405
164	Joseph Bruce	Seaham Park CC Under 13	3.143389
165	Finlay Kirkup	Raby Castle CC Under 13	3.13578
166	Thomas Turnbull	Ryhope CC Under 13	3.135471
167	Jacob Hall	Murton CC Under 13	3.132785
168	Lewis Wilson	Philadelphia CC Under 13	3.11603
169	Sam Irvine	Bishop Auckland CC Under 13	3.114531
170	Benjamin Poulton	Ryhope CC Under 13	3.111983
171	Ted McKie	Sacriston CC Under 13 - Red	3.099125
172	Joseph Ruddock	Marsden CC, Tyne & Wear Under 13	3.098374
173	James Wallace	Marsden CC, Tyne & Wear Under 13	3.095443
174	Jacob Winlow	Sunderland CC Under 13	3.087472
175	Ethan Arkley	Seaham Harbour CC Under 13 Yellow	3.082622
176	Barney Forster	Willington CC Under 13	3.075683
177	Luke Gates	Ryhope CC Under 13	3.075049
178	Seyshan Ragulan	Sunderland CC Under 13	3.070204
179	Henry Clasper	Tudhoe CC Under 13	3.068899
180	Niall Gulliver	Silksworth CC Under 13	3.068604
181	Duke Mavin	Durham City CC Under 13 Blue	3.061325
182	Jevin Rajapakse	Sunderland CC Under 13	3.047607
183	George Cave	Ryton CC Under 13 Raiders	3.041736
184	J Blacklock	Durham City CC Under 13 Gold	3.040372
185	Freddie Robson	Burnmoor CC Under 13	3.040262
186	Lewis Robertson	Sacriston CC Under 13 - Red	3.029274
187	Anthony Emmerson	Murton CC Under 13	3.011354
188	Morgan Boustead	Willington CC Under 13	3.01044
189	Zach Abdy	Dawdon Welfare CC Under 13	3.005727
190	Mason Lee	Ushaw Moor CC Under 13	3.004654
191	Georgie Franklin	Gateshead Fell CC Under 13 B	2.995463
192	Ellis Hazel	Ushaw Moor CC Under 13	2.980796
193	Ethan Craine	Langley Park & Bearpark CC Under 13's	2.968132
194	Roan Edmunds	Sacriston CC Under 13 - Black	2.944701
195	Noah Raine	Consett CC Under 13	2.940184
196	Unsure	Leadgate CC Under 13	2.937419
197	Harry Lewis	Sunderland CC Under 13	2.932292
198	Mason Swan	Seaham Park CC Under 13	2.918037
199	Owen Stokoe	Ryhope CC Under 13	2.913412
200	Freddie Barnes	Whitburn CC Under 13 A	2.906231
201	Fraser Gibbons	Hylton CCC Under 13	2.899329
202	Jake Huscroft	Langley Park & Bearpark CC Under 13's	2.893974
203	K Russell	Easington C.W. CC Under 13	2.88763
204	Luke Hiles	Langley Park & Bearpark CC Under 13's	2.861429
205	Lily Burrell	Hetton Lyons CC Under 13 Gold	2.859701

206	Joseph McAllister	Hunwick CC Under 13	2.83102
207	Sonny Lowther	Hetton Lyons CC Under 13 Green	2.804835
208	Harry Petch	Dawdon Welfare CC Under 13	2.787697
209	Harry Keegan	Hetton Lyons CC Under 13 Green	2.765727
210	Austin Caris	Philadelphia CC Under 13	2.737125
211	Sebastian Gregg	Marsden CC, Tyne & Wear Under 13	2.715709
212	William Dixon	Sacriston CC Under 13 - Red	2.70952
213	A Vout	Hetton Lyons CC Under 13 Green	2.708832
214	Max Mullarky	Bishop Auckland CC Under 13	2.689531
215	Daniel Wright	Burnmoor CC Under 13	2.639355
216	Tom Bulmer	Sacriston CC Under 13 - Red	2.637893
217	Hayden O'Keefe	Bishop Auckland CC Under 13	2.614458
218	S Matevica	Tudhoe CC Under 13	2.605045
219	Oliver Bell	Peterlee CC Under 13	2.546352

Appendix C Table 23: Player Ranking in high performance clusters of ball-by-ball Bowling Scorecard (PCA) using Hierarchical Clustering

Rank	Bowler	Team	Composite Score
1	C Bennett	Littletown CC Under 13	5.933566
2	Keiran Stafford	Hylton CCC Under 13	5.45731
3	Athena Elstob	Seaham Harbour CC Under 13 Purple	5.333594
4	Thomas Marshall	Philadelphia CC Under 13	5.213014
5	Sophie Stoker	Seaham Harbour CC Under 13 Purple	5.174099
6	Dylan Lynn	Brandon CC, Durham Under 13	5.067583
7	Logan Reay	Eppleton CC Under 13	5.019718
8	S Howe	Seaham Park CC Under 13 B	4.991163
9	J Gibson	Littletown CC Under 13	4.961257
10	Alex Garthwaite	Washington CC, Durham Washington U13 Nova	4.905242
11	Lola Wilkinson	Seaham Harbour CC Under 13 Purple	4.71872
12	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.705054
13	Alfie Melvin	Greenside CC Under 13	4.659939
14	Luke Marriott	Eppleton CC Under 13	4.608565
15	William Mordue	Sunderland CC Under 13	4.596894
16	Louis Morris	Raby Castle CC Under 13	4.569865
17	Charlie Gray	Burnhope CC Under 13	4.558365
18	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.527809
19	J Liddle	Annfield Plain CC Under 13	4.524336
20	Libby Hall	Seaham Harbour CC Under 13 Purple	4.498087
21	O Mansfield	Ushaw Moor CC Under 13	4.471258
22	Ryan Kerr	Consett CC Under 13	4.427195
23	C Mavin	Durham City CC Under 13 Blue	4.418809

24	Noah Shaw	Beamish and E. Stanley CC Under 13	4.373983
25	L Gayle	Dawdon Welfare CC Under 13	4.368369
26	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.368267
27	Callum Andersen	Boldon CA CC Under 13	4.366276
28	Joshua Mellons	Shotley Bridge CC Under 13	4.362121
29	Kai Kennedy	Dawdon Welfare CC Under 13	4.331752
30	Oliver Gaff	Philadelphia CC Under 13	4.311767
31	Joseph Bruce	Seaham Park CC Under 13	4.296744
32	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.292822
33	Hannah Collingwood	Shotley Bridge CC Under 13	4.279789
34	Myles Price	Dawdon Welfare CC Under 13	4.183592
35	Freddie Gilmore	Seaham Park CC Under 13	4.170852
36	Leo Smith	Silksworth CC Under 13	4.14204
37	Archie Potts	Boldon CA CC Under 13	4.128625
38	H Wells	Littletown CC Under 13	4.100979
39	George McDonald	Burnmoor CC Under 13	4.095373
40	Kai Massam	Ryton CC Under 13 Rockets	4.077959
41	F Todner	Seaham Park CC Under 13	4.054238
42	Tyler Anderson	Tudhoe CC Under 13	4.038821
43	Joshua Gustard	Dawdon Welfare CC Under 13	4.030763
44	C McCarthy	Dawdon Welfare CC Under 13	4.010298
45	Daniel Carr	Annfield Plain CC Under 13	3.995337
46	Z Cartwright	Littletown CC Under 13	3.953268
47	Matty Marshall	Seaham Harbour CC Under 13 Yellow	3.945639
48	J Bowman	Easington C.W. CC Under 13	3.86222
49	C Dyer	Littletown CC Under 13	3.79555
50	J Turnbull	Easington C.W. CC Under 13	3.791629
51	Aidan Wilkes	Ushaw Moor CC Under 13	3.688321
52	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.669281
53	Holly Higginson	Shotley Bridge CC Under 13	3.662095
54	Hayden Jackman	Annfield Plain CC Under 13	3.646019
55	Charlie Atkinson	Sacriston CC Under 13 - Black	3.631706
56	Joel Soppitt	Seaham Harbour CC Under 13 Yellow	3.615104
57	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.593669
58	Sam Lewis	Burnmoor CC Under 13	3.473315

Appendix C Table 24: Player Ranking in high performance clusters of ball-by-ball Bowling Scorecard with encoded dimensions using KMeans++ Clustering

Rank	Bowler	Team	Composite Score
1	C Bennett	Littletown CC Under 13	5.933566
2	Keiran Stafford	Hylton CCC Under 13	5.45731

3	Athena Elstob	Seaham Harbour CC Under 13 Purple	5.333594
4	Thomas Marshall	Philadelphia CC Under 13	5.213014
5	Dylan Lynn	Brandon CC, Durham Under 13	5.067583
6	Logan Reay	Eppleton CC Under 13	5.019718
7	S Howe	Seaham Park CC Under 13 B	4.991163
8	J Gibson	Littletown CC Under 13	4.961257
9	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.705054
10	Alfie Melvin	Greenside CC Under 13	4.659939
11	William Mordue	Sunderland CC Under 13	4.596894
12	Louis Morris	Raby Castle CC Under 13	4.569865
13	Charlie Gray	Burnhope CC Under 13	4.558365
14	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.527809
15	J Liddle	Annfield Plain CC Under 13	4.524336
16	O Mansfield	Ushaw Moor CC Under 13	4.471258
17	Ryan Kerr	Consett CC Under 13	4.427195
18	C Mavin	Durham City CC Under 13 Blue	4.418809
19	Noah Shaw	Beamish and E. Stanley CC Under 13	4.373983
20	L Gayle	Dawdon Welfare CC Under 13	4.368369
21	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.368267
22	Callum Andersen	Boldon CA CC Under 13	4.366276
23	Joseph Bruce	Seaham Park CC Under 13	4.296744
24	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.292822
25	Myles Price	Dawdon Welfare CC Under 13	4.183592
26	Freddie Gilmore	Seaham Park CC Under 13	4.170852
27	Leo Smith	Silksworth CC Under 13	4.14204
28	Archie Potts	Boldon CA CC Under 13	4.128625
29	H Wells	Littletown CC Under 13	4.100979
30	George McDonald	Burnmoor CC Under 13	4.095373
31	Kai Massam	Ryton CC Under 13 Rockets	4.077959
32	F Todner	Seaham Park CC Under 13	4.054238
33	Joshua Gustard	Dawdon Welfare CC Under 13	4.030763
34	C McCarthy	Dawdon Welfare CC Under 13	4.010298
35	Daniel Carr	Annfield Plain CC Under 13	3.995337
36	Z Cartwright	Littletown CC Under 13	3.953268
37	Matty Marshall	Seaham Harbour CC Under 13 Yellow	3.945639
38	Bobby Carey	Durham City CC Under 13 Blue	3.88924
39	J Bowman	Easington C.W. CC Under 13	3.86222
40	C Dyer	Littletown CC Under 13	3.79555
41	J Turnbull	Easington C.W. CC Under 13	3.791629
42	Aidan Wilkes	Ushaw Moor CC Under 13	3.688321
43	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.669281
44	Holly Higginson	Shotley Bridge CC Under 13	3.662095
45	Hayden Jackman	Annfield Plain CC Under 13	3.646019

46	Charlie Atkinson	Sacriston CC Under 13 - Black	3.631706
47	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.593669
48	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	3.573315
49	Gemma Hiles	Langley Park & Bearpark CC Under 13's	3.532195
50	Sam Lewis	Burnmoor CC Under 13	3.473315
51	Harry Tellegen	Raby Castle CC Under 13	3.387026
52	Freddie Robson	Burnmoor CC Under 13	3.386393
53	T Richards	Tudhoe CC Under 13	3.236404
54	Daniel Wright	Burnmoor CC Under 13	2.956785

Appendix C Table 25: Player Ranking in high performance clusters of ball-by-ball Bowling Scorecard with encoded dimensions using Hierarchical Clustering

Rank	Bowler	Team	Composite Score
1	C Bennett	Littletown CC Under 13	5.933566
2	Keiran Stafford	Hylton CCC Under 13	5.45731
3	Athena Elstob	Seaham Harbour CC Under 13 Purple	5.333594
4	Thomas Marshall	Philadelphia CC Under 13	5.213014
5	Dylan Lynn	Brandon CC, Durham Under 13	5.067583
6	Logan Reay	Eppleton CC Under 13	5.019718
7	S Howe	Seaham Park CC Under 13 B	4.991163
8	J Gibson	Littletown CC Under 13	4.961257
9	Jake Pattison	Seaham Harbour CC Under 13 Yellow	4.705054
10	Alfie Melvin	Greenside CC Under 13	4.659939
11	William Mordue	Sunderland CC Under 13	4.596894
12	Louis Morris	Raby Castle CC Under 13	4.569865
13	Fraser Wright	Seaham Harbour CC Under 13 Yellow	4.527809
14	J Liddle	Annfield Plain CC Under 13	4.524336
15	O Mansfield	Ushaw Moor CC Under 13	4.471258
16	Ryan Kerr	Consett CC Under 13	4.427195
17	C Mavin	Durham City CC Under 13 Blue	4.418809
18	Noah Shaw	Beamish and E. Stanley CC Under 13	4.373983
19	L Gayle	Dawdon Welfare CC Under 13	4.368369
20	Shantush Sivakumar	Wearmouth & YAV CC Under 13	4.368267
21	Callum Andersen	Boldon CA CC Under 13	4.366276
22	Joseph Bruce	Seaham Park CC Under 13	4.296744
23	Samuel Philliskirk	Wearmouth & YAV CC Under 13	4.292822
24	Myles Price	Dawdon Welfare CC Under 13	4.183592
25	Freddie Gilmore	Seaham Park CC Under 13	4.170852
26	Leo Smith	Silksworth CC Under 13	4.14204
27	Archie Potts	Boldon CA CC Under 13	4.128625
28	H Wells	Littletown CC Under 13	4.100979

29	George McDonald	Burnmoor CC Under 13	4.095373
30	Kai Massam	Ryton CC Under 13 Rockets	4.077959
31	F Todner	Seaham Park CC Under 13	4.054238
32	Joshua Gustard	Dawdon Welfare CC Under 13	4.030763
33	C McCarthy	Dawdon Welfare CC Under 13	4.010298
34	Daniel Carr	Annfield Plain CC Under 13	3.995337
35	Z Cartwright	Littletown CC Under 13	3.953268
36	Matty Marshall	Seaham Harbour CC Under 13 Yellow	3.945639
37	Bobby Carey	Durham City CC Under 13 Blue	3.88924
38	Benjamin Thompson	Boldon CA CC Under 13	3.881473
39	J Bowman	Easington C.W. CC Under 13	3.86222
40	C Dyer	Littletown CC Under 13	3.79555
41	J Turnbull	Easington C.W. CC Under 13	3.791629
42	H Glendenning	Easington C.W. CC Under 13	3.751257
43	Aidan Wilkes	Ushaw Moor CC Under 13	3.688321
44	Sam Birbeck	Hetton Lyons CC Under 13 Green	3.669281
45	Holly Higginson	Shotley Bridge CC Under 13	3.662095
46	Hayden Jackman	Annfield Plain CC Under 13	3.646019
47	Charlie Atkinson	Sacriston CC Under 13 - Black	3.631706
48	Ellen Warrilow	Langley Park & Bearpark CC Under 13's	3.593669
49	Rashvin Ramanathas	Wearmouth & YAV CC Under 13	3.573315
50	Sam Lewis	Burnmoor CC Under 13	3.473315
51	Jack Chapman	Burnmoor CC Under 13	3.446828
52	Harry Tellegen	Raby Castle CC Under 13	3.387026
53	Freddie Robson	Burnmoor CC Under 13	3.386393
54	Daniel Wright	Burnmoor CC Under 13	2.956785