Ranking System of IPL 2018 Players

Asutosh College, Department of Statistics

University Reg No: 012-1111-0836-20

University Roll No: 203012-21-0109

College Roll No: 922

Subject: Statistics Honours

Paper: DSE B2

DECLARATION

I**, Souraj Chakraborty**, a student of B.Sc. Semester-6, Statistics Honours, of University of Calcutta, **Registration no-012-1111-0836-20, Roll no- 203012-21-0109**, hereby declare that I have done this piece of project work entitled "**A Ranking System of IPL Players of 2018**" under the supervision of **Ms. Oindrila Bose** (Faculty Member , Department of Statistics, Asutosh College) as a part of B.Sc. Sem-6 examination according to the syllabus paper DSE-B2. I further declare that the piece of project work has not been published elsewhere for any degree or diploma or taken from any published project.

Signature:

____

CONTENTS:

Summary:

The object of this research was to develop a rating system based on the overall performance of a player in a particular season of limited-overs cricket. Regarded as one of the toughest t20 leagues all over over world, the Indian Premier League or IPL served as a firsthand source of required data. Firstly, we devised a unique formula based on our subjective assessment of the game and the individual performances of the players. This was done with the help of analytic hierarchy process, otherwise known as AHP rating.

Principal Component Analysis is widely used in applied multivariate data analysis, and this article shows how to motivate student interest in this topic using cricket sports data. Here, principal component analysis is successfully used to rank the cricket batsmen and bowlers who played in the 2018 Indian Premier League (IPL) competition. In particular, the first principal component is seen to explain a substantial portion of the variation in a linear combination of some commonly used measures of cricket prowess. This application provides an excellent, elementary introduction to the topic of principal component analysis.

The results thus revealed prove to be satisfactory and useful for those who intend to indulge further on the subject.

Introduction

The growth of sport analytics and the need for meaningful sport related statistics has emerged in recent decades due to the large volume of monetary resources that is increasingly being invested in a single player or team. The rise in player salaries and salary caps over the last 25 years provide ample evidence of the growth of sport analytics, with investors, franchises, clubs and other stakeholders wanting to determine the true value of their investment. For example, in the National Football League (NFL) there has been an increase of approximately 950% in player salaries since the 1980’s, and an increase of 288% in salary cap since 1994. With global sports revenue estimated to grow by US$145.3 billion over the 2010-2015 period and winning teams earning significantly larger revenue than that of losing teams, there is a strong incentive for managers and coaching staff of sport teams to succeed. Given the large investment of resources and the stakes involved, coaches and managerial staff cannot solely rely on subjective views and personal beliefs to make team and player selection decisions. Solutions must be augmented with objective approaches by implementing analytical techniques.

It is important to distinguish sport analytics from collecting quantitative data. Quantitative data collection, in sport, is the measurement and storage of the behaviours or actions of a team or a player, while analytics is the use of data to inform decision makers. An early example of data collection within sports dates back to the 1850’s with the publication of cricket averages in magazines. Although the collection and recording of numerical data within sports has been conducted for quite some time, the application of quantitative and statistical methods to this data is still in its infancy.

Due to the nature of human contest, sport lends itself to fluctuations and discrepancies in game outcomes, this in turn generates spectator interest. This outcome volatility is predominately due to variation in performance between individual players and teams. Therefore coaches, managers, fans, media and other interested parties utilise analytical approaches to understand the root of this variation, and handle and reduce its effect in order to produce ‘better’, more consistent results. Moreover these analytical techniques allow the user to rank and rate player and team performances. In general, sport rating systems provide an objective evaluation of a team or individual based on prior performances, and are implemented for player comparisons, improving the player/ team selection process and betting purposes.

The explosion in the sporting industry in terms of popularity and revenue is evident in cricket. Cricket has seen a huge global growth in revenue in recent years, and transformed into a sporting juggernaut due to the advent of T20 cricket. This is a relatively new short form format, where teams each face up to 20 overs. A match typically concludes in three hours, which increases spectator appeal.

Indian Premier League is a T20 League based in India, where different franchises represent different teams. In this project, we would try to introduce the data set of the batsmen and bowlers of Indian Premier League (2018) and we would try to introduce a ranking system for batsmen and bowlers and try to rank them according to their ratings. We would use PCA (Principal Component Analysis) to find the ratings of the players. Now, the players are ranked according to their ratings.

Descriptive analysis of the provided cricket statistics:

1. **Batting Performances**:
   - Kane Williamson tops the batting chart with the highest runs (735) and an impressive average of 52.5. He has been consistent with 8 half-centuries and a strike rate of 142.44.
   - Rishabh Pant has had a breakthrough season with 684 runs at an outstanding strike rate of 173.6. He has a century and five fifties to his name.
   - KL Rahul has also been consistent with 659 runs and a high average of 54.91. He scored six fifties but missed out on centuries.
   - Shane Watson and Ambati Rayudu have been key contributors to their team, with Watson scoring two centuries, and Rayudu scoring one century and three fifties.
2. **All-Round Performances:**
   - Andre Russell stands out as a valuable all-rounder with 316 runs at a strike rate of 184.79 and also contributing with the ball, picking up 11 wickets.
   - Hardik Pandya has been crucial for his team, scoring 260 runs at a strike rate of 133.33 and taking 18 wickets with the ball.
   - Shakib Al Hasan and Krunal Pandya have also contributed significantly with both bat and ball.
3. **Bowling Performances:**
   - Andrew Tye leads the bowling charts with 24 wickets at an average of 18.66 and an economy rate of 8.00.
   - Rashid Khan, despite playing a few more matches, has been economical with an economy rate of 6.73 and has taken 21 wickets.
   - Siddarth Kaul, Umesh Yadav, and Trent Boult have been key pacers, taking 20 or more wickets each.
4. **Spin Bowling:**
   - Sunil Narine and Kuldeep Yadav have been the main spinners for their respective teams. Narine has taken 17 wickets, and Kuldeep has 17 wickets with a 4-wicket haul to his name.
5. **Emerging Players:**
   - Mayank Markande and Shreyas Gopal have shown promise as young spinners, taking 15 and 11 wickets, respectively.
   - Shivam Mavi and Prasidh Krishna have impressed as fast bowlers, picking up 5 and 10 wickets, respectively.
6. **Captaincy Impact:**
   - Kane Williamson and Rishabh Pant have led from the front, being the top scorers for their teams.
   - Andrew Tye and Rashid Khan have been successful in leading their bowling attacks.

7. **Impact of All-Rounders:**
   - Andre Russell and Hardik Pandya have been game-changers with their all-round performances, making significant contributions with both bat and ball.
8. **High Strike Rate:**
   - Sunil Narine and Shane Watson stand out with their high strike rates of 189.89 and 154.59, respectively.
9. **Economical Bowling:**
   - Rashid Khan, Kuldeep Yadav, and Sunil Narine have been the most economical bowlers, making it difficult for batsmen to score against them.

Overall, this descriptive analysis provides an overview of the standout performances in batting and bowling across different players. It highlights the key performers, emerging talents, and the impact of captaincy and all-rounders in the league.

Methodology

At the beginning, let us introduce some terminologies and variables used in our further computations and analysis for ranking the cricketers.

1) Terminologies related to Batsmen

While considering batsmen, the goal in limited-overs cricket, like Twenty20, is to score as many as runs as possible using as few balls as possible. On the other hand, building longer innings instead of trying to score runs for each ball is the key goal in test cricket. For a batting analysis, there is set of widely-recognized variables that can be used to measure the quality of each batsman. These variables, which are mentioned below, are commonly used by cricket commentators and sports authorities, and are also shown in scoreboards to describe player profiles.

a) **Matches** : Total number of matches where a player features in the playing 11 in the IPL 2017 season. More number of matches mean chances to score more runs, which in turn indicate stronger performance.

b) **Innings** : Total number of innings indicate the number of times a player gets a chance to bat in the IPL 2017 season. This factor does not consider the matches where the batsmen featured in the starting lineup but did not get a chance to bat.

c) **Runs**: The total number of runs scored by a player in the IPL 2017 season. Higher values indicate stronger performance.

d) **Balls Faced** : The total number of balls faced by a batsmen throughout the 2017 IPL season. This provides a batsmen opportunities to score more runs, but higher value of this variable also reduces the strike rate of player.

e) **Not-Outs** : The total number of times a player remains till the end of the batting innings or run chase without being dismissed in the IPL 2017 season. Higher values increases batting average, which in turn, indicates stronger performance.

f) **Batting Average (Ave):** The total number of runs a batsman has scored divided by the total number of times he has been called out in the IPL 2017 season. Higher values indicate stronger performance. However, for a batsman with several "not out" cases, this number overestimates the batsman, which is a weakness in this measure, and this is why it should not be used as the only variable for batting performance analysis.

g) **Batting Strike Rate (SR)**: The Batting Strike Rate is defined as the number of runs scored per 100 balls faced by a batsman in the IPL 2017 season. Again, similar to batting average, higher values indicate stronger performance. An aggressive batting style is always helpful in shorter versions of the game. However, a high strike rate accompanying a low batting average is not desirable.

h) **Centuries** : Total number of times a batsman scores 100 or more runs in the 2017 IPL season. Higher value indicates very strong performance.

i) **Half Centuries** : Total number of times a batsman scores 50 or more runs in the 2017 IPL season. Just like in case of centuries, higher value indicates better performance.

j) **Total Boundaries** : The total number of boundaries (fours = four runs, six= six runs) made in the IPL 2012 season by a batsman. Higher values indicate stronger performance. Scoring boundaries is a great way to increase the number of runs without wasting resources, and it helps increase the batting average and strike rate.

k) **Percentage Boundary** : As suggested by the name, this is the number of boundaries (fours and sixes) hit by a batsman per 100 balls. Higher value indicates stronger performance.

Other variables, like batting position (batting order), winning the toss, and the so-called called "home-field advantage" might also be considered, but incorporating those here would be beyond the scope and objective of this article.

2) Terminologies related to Bowlers :

Turning now to bowling performance, we use a set of widely-recognized variables to measure the quality of each bowler. These variables are the ones widely used by cricket announcers to describe player performance and are likewise shown in scoreboards to describe player profiles.

a) **Total Overs** : The total number of overs a bowler has bowled during the whole course of the IPL 2017 season. Most of the time, more number of overs mean that the player is of premium quality and the go-to man for the team. This also provides more opportunities to take wickets.

b) **Wickets** : The number of wickets taken by a bowler. There are ten possible wickets for an innings and there should be at least five bowlers, each of whom can bowl a maximum of four overs. A bowler's goal is to take the maximum number of wickets from the overs that he bowls, so taking a large number of wickets is one performance measure for bowlers. However, like the total number of runs statistic for a batsman, the number of wickets taken is not sufficient to measure the quality of a bowler. The goal of a bowler is to get the maximum number of wickets by using a minimum number of balls while simultaneously conceding a minimum number of runs. Higher values indicate better performance.

c) **Bowling Average** : The average number of runs conceded per wicket. Here, lower values are preferred since a bowler's goal is to concede the minimum number of runs while simultaneously earning the maximum number of wickets.

d) **Strike Rate** : The average number of balls bowled per wicket taken. Lower values are preferred since a bowler should try to bowl the minimum number of balls per wicket.

e) **Economy Rate** : The average number of runs conceded per over. Lower values are preferred since this is the run-rate against a specific bowler for a batting team. Therefore, the bowler's aim is to keep this measure as small as possible.

f) **Total Boundaries conceded**: Total number of boundaries (fours or sixes) conceded by a bowler throughout the length of the IPL 2017 season. Lower values indicate better performance for this variable.

g) **Total Runs Conceded** : Total runs that were conceded by the bowler in IPL 2017 season. Lower values suggest stronger performance.

h) **4-Wicket (4W) and 5-Wicket (5W) Haul** : The total number of 4-wicket and 5-wicket hauls (4wickets/5 wickets taken in a match) taken by a player in the IPL 2017 season. Higher values generally indicate stronger performance. Sometimes, it also depends on the situation of the match when that particular player is bowling.

As with batsmen, other variables, like the significance of the wickets taken (wickets taken at the front end of the batting order are usually harder to get than those towards the end) might also be considered.
Now coming to the main part, we will discuss here how we created a ranking system for ranking the cricketers of IPL 2017. It should be mentioned here that we have taken those batsmen under our consideration who played at least 5 matches in the whole season of IPL 2017. In case of bowlers too, we have ranked those players only who bowled in at least 5 matches, and took at least 1 wicket in the whole season.

Method I:

From the number of relevant variables defined above, after **subjective discussion**, we come to the conclusion that some variables deserve better recognition compared to others. Particularly so, in this form of the game. Therefore, we decided to choose three variables each, from batting and bowling perspective, which, according to us is most influential for the result of the game. Players who excel in these features of cricket are better suited as a T20 player and prove to be key performers to ensure their team's success.

- ## Most relevant Batting variables:-

❖ Runs

When we talk about about a good batsmen, the first thing to prove our point of view, we state how many he scored in the last tournament or the series. One doesn't need to be a cricket expert to say a batsman is in good form if he scored a lot of runs in his last tournament. Whether those runs came very fast or slow is of little relevance in case of test matches. But in limited overs format, one must take that into account.

So, the total number of runs scored by a player in this season will be one of the key factors that will influence his ranking as well as his team's performance.

❖ Strike Rate

In Test cricket, a batsman's strike rate is of secondary relevance to his ability to score runs without getting out. This means a Test batsman's most important statistic is generally considered to be his batting average, rather than his strike rate.

In limited overs cricket, strike rates are of considerably more importance. Since each team only faces a limited number of balls in an innings, the faster a batsman scores, the more runs his team will be able to accumulate. Strike rates of over 150 are becoming common in Twenty20 cricket. Strike rate is probably considered by most as the key factor in a batsman in one day cricket. Accordingly, the batsmen with the higher strike rate, especially in Twenty20 matches, are more valued than those with a lesser strike rate.

❖ Percentage Boundary

Another factor that can be considered as important as the previous two is the recorded percentage boundary of a batsman. This is essentially the number of boundaries (fours and sixes) hit by a batsman per hundred balls he faced. Especially in short and fast paced format like T20, the boundary percentage of the whole team is often regarded as the key to success. As a batsman, a higher boundary percentage not only ensures a very high strike rate, it also affects the mindset of opposition bowlers and toy with their concentration.

From our minimal understanding of the T20 version of cricket and after referring to various similar works, we decided subjectively to apply the weightage to the variables as given below -

Runs - 0.33; Strike Rate - 0.37; Percentage Boundary - 0.30

And the weights are now combined in **product-weighted measure** to rate the batsmen in the following way ~

Raw rating formula for batsmen : $(\text{Runs}^{0.33})\times(\text{Strike Rate}^{0.37})\times(\text{Percentage Boundary}^{0.30})$

Actual rating formula for batsmen : (Raw Rating)×(Total no. of batsmen under study)/ (Sum of the Raw Ratings)

According to this Actual rating all the batsmen are ranked . In this method, a batsman with higher rating is a better player, so he will be ranked lower.

Now, moving on the bowling department, we repeated the same process and decided to focus on three most relevant variables that boosts up the ranking of a bowler in T20 format.

- ## Most Relevant Bowling Variables :-

❖ Economy Rate

Anyone who has followed T20 cricket for some time will surely regard low economy rate as one of, if not the most necessary traits of a T20 bowler. At some times, a bowler who has a lower economy prove to be more vital than another bowler who has taken more wickets. Surely, we have seen too many matches to believe otherwise.

Suppose a bowler took 3 wickets but gave away 35 runs in his three overs while another one gave only 14 runs in his three. We often see that the captain opts for the bowler who has a lower economy rate because keeping the batsmen quite is a valuable trick to have up your sleeve in crunch situations. This is because a bowler with higher economy rate can take 2 wickets but give away more than 10 valuable runs while defending a moderate score. But a bowler with a lower economy will hold them tight, which will increase pressure on the batsmen, in turn, yielding wickets.

❖ Bowling Strike Rate

Even if it suffers a somewhat decrease in importance in T20 matches, the ability to take wickets is still a vital trick to have up your sleeve. Bowling strike rate is the average number balls bowled by a bowler to take a wicket. A lower bowling strike rate means that a bowler is taking quick wickets which results in the exposure of the opponent team's batting order. We often see that bowlers with a lower bowling strike rate often have a lower economy rate as well. This is because the batsmen don't want to have their wickets taken by this bowler and are playing with additional caution, thereby choking their free flowing scoring rate.

❖ Bowling Average

The bowling average measures the no of runs conceded per wicket. In T20 format, it is also very much necessary for a bowler to take wickets regularly besides defending runs. Because often a situation occurs in a match (especially in Batting Powerplay), when there are such fielding restrictions that a bowler is compelled to search for wickets, because the situation doesn’t allow him to do defensive bowling or to focus on restricting the flow of runs. So in such situations a bowler’s wicket taking capability becomes an important factor, which is measured by Bowling Average.

Again, from our minimal understanding of the T20 version of cricket and after referring to various similar works, we decided subjectively to apply the weightage to the bowling variables as given below -

Economy Rate - 0.37; Bowling strike rate - 0.33; Bowling Average - 0.30

And the weights are now combined in **product weighted measure** to rate the bowlers in the following way ~

Raw rating formula for bowlers: **(Bowling SR$^{0.33}$)×(Economy Rate$^{0.37}$)×(Bowling Average$^{0.30}$)**

Actual rating formula for bowlers : **(Raw Rating)×(Total no. of bowlers under study)/(Sum of the Raw Ratings)**

According to this Actual rating all the bowlers are ranked . In this method, a bowler with lower rating is a better player, so he will be ranked lower.

Method II

Ranking the Cricketers with the help of Principal Component Analysis (PCA) :-

- **THEORY OF PCA** :
  When working on a complex science project with a lot of data where each example is described by many characteristics, you may want to **visualize the data**. In fact, visualization in 1D, 2D or 3D is easy, but if you want to visualize your data composed of 100 characteristics, you won't see anything in 100D. So you have to **reduce the dimension** and place your data in a space with a dimension equal to or smaller than 3D. This can be done using Principal Component Analysis (PCA). Another very good use of PCA is to **speed up the training process** of your machine learning algorithm. The more features you have, the more complex your machine learning algorithm is and the more parameters it needs to learn. Thus, the longer the computation time will be. Now, it seems obvious that if you reduce the size of your data, the time needed to learn your algorithm will decrease considerably. **But, you may wonder if we reduce the number of variables, then we will lose a lot of information from our data and the result will not be accurate at all.** This is the main challenge of the PCA algorithm and we will see in this article how to ensure good accuracy.

Maths Behind Principal Component Analysis

In the principal component analysis algorithm, the objective is to find the k vectors on which to project the data in order to minimize the projection error. **These k vectors will be the k directions on which to project the data**. Here, k corresponds to your final dimension: if you want to look at your data in a 2D dimensional space, then k will be equal to 2.

How do we find these k vectors ?

Let's call **A the matrix which describes our data**. PCA involves transforming interdependent variables (called "correlated" in statistics) into new variables that are uncorrelated with each other. These variables are called principal components and will describe the information conveyed by the data. We need to look at the covariance matrix because **covariance** is a measure of the joint variability of two random variables. But why covariance? Let's think about how we can learn from the data. You look at the mean and how the variables are far from or close to the mean. This is essentially covariance: the deviation from the mean. So in the PCA, we have to calculate the covariance of our data matrix and look for the **directions or vectors** that collect the most information so that we can keep the PI and get rid of the rest. But it's not as simple as that and we're going to look at it step by step to see how we do it.

1. Standardization of the data:

The aim of this step is to **standardize the range of the continuous initial variables** so that each one of them contributes equally to the analysis.

More specifically, the reason why it is critical to perform standardization prior to PCA, is that the latter is quite sensitive regarding the variances of the initial variables. That is, if there are large differences between the ranges of initial variables, those variables with larger ranges will dominate over those with small ranges (For example, a variable that ranges between 0 and 100 will dominate over a variable that ranges between 0 and 1), which will lead to biased results. So, transforming the data to comparable scales can prevent this problem.

Mathematically, this can be done by subtracting the mean and dividing by the standard deviation for each value of each variable. Recall that Ais the matrix which describes our data, thus each row is an example and each column is a feature. Let's set **n equal to the number of features** et **m the number of examples**. Thus the matrix A is a matrix **mxn.**

$$A^i_{Standard} = \frac{A - mean}{standard\ deviation}\ \text{where } A^i \text{ is one feature of A}$$

$$\text{and } A_{\text{standard}} \text{ is the new data matrix}$$

2. Covariance matrix and diagonalisation

The covariance matrix is given by the following formula:

$$cov(\mathrm{A_{standard}}) = \Sigma = \frac{1}{m}\sum_{i=1}^{n}(A^i_{Standard})(A^i_{Standard})^T = A^T A$$

Then, we need to diagonalize the covariance matrix. We will call S the diagonal matrix. U and V will be the transformation matrices. Thus we have:

$$\Sigma = USV \text{ with } U = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ u^{(1)} & u^{(2)} & \cdots & u^{(n)} \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} \text{ with } U = V^{-1} \text{ and } u^{(i)} \text{ the eigen vectors of A}$$

If you have heard about PCA you may have heard about SVD. SVD is for Singular Value Decomposition and it is what we are applying here on A. The SVD theory states that it exist the following decomposition for the matrix A:

$$A = US'V^T$$

where *U* and *V* are **orthogonal matrices** with orthonormal **eigenvectors** chosen from AA^T and $A^T\boldsymbol{A}$ respectively. ***S'*** is a diagonal matrix with *r* elements equal to the root of the positive eigenvalues of XX^T or $X^T X$. The diagonal elements are composed of singular values.

$$S' \\ \begin{pmatrix} \sqrt{\lambda_1} & & & & & \\ & \sqrt{\lambda_2} & & & & \\ & & \dots & & & \\ & & & \sqrt{\lambda_r} & & \\ & & & & \dots & \\ & & & & & 0 \end{pmatrix} \equiv \begin{pmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \dots & & & \\ & & & \sigma_r & & \\ & & & & \dots & \\ & & & & & 0 \end{pmatrix}$$

And we have ***S'²* = *S*.** The next step is to reorganize the eigenvectors to produce U and V. To standardize the solution, we order the eigenvectors such that vectors with higher eigenvalues come before those with smaller values.

$$\begin{array}{ccc} \sigma_1 > \sigma_2 > \dots > \sigma_m \end{array}$$

$$\begin{pmatrix} | & & | \\ u_1 & \dots & u_m \\ & & \end{pmatrix}$$

Comparing to eigen decomposition, SVD works on non-square matrices. *U* and *V* are invertible for any matrix in SVD.

3. Dimensionality reduction

The idea of dimensionality reduction is to keep k eigen vectors which describe the best the data. Thus, we have:

$$A \begin{bmatrix} \boldsymbol{v}_1 & \cdot & \cdot & \boldsymbol{v}_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \cdot & \cdot & \boldsymbol{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \sigma_r \end{bmatrix}$$

Then we add null space, which is already orthogonal to first r v's and u's to go from reduced to full SVD.

$$A\begin{bmatrix} v_1 & \cdot\cdot & v_r & \cdot\cdot & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \cdot\cdot & u_r & \cdot\cdot & u_m \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

4. How to choose the k eigen vectors:

There is a formula we need to apply to keep a certain percentage of the total variance. Let's say we want to keep **99%** of the variance then we have the following:

$$\frac{\sum_{i=1}^{k} \sigma_i}{\sum_{i=1}^{n} \sigma_i} \geq 0.99$$

In R programming, you can perform Principal Component Analysis (PCA) using the `prcomp` function from the `stats` package. PCA is a popular technique used for dimensionality reduction and identifying patterns or trends in data.

Here's a step-by-step guide on how to perform PCA in R:

Step 1: Load the required libraries.

R Copy code

```R
# Load the required library
library(stats)
```

Step 2: Prepare your data. Ensure that your data is in a suitable format (e.g., a data frame) and numeric. If your data contains categorical variables, you may need to preprocess it (e.g., one-hot encoding) before performing PCA.

Step 3: Perform PCA using `prcomp`.

```
# Assuming your data is in a data frame called 'data'
# You can specify the variables that you want to include in the PCA as the formula
# Here, we will use all columns in 'data' for PCA
    pca_result <- prcomp(data, scale = TRUE)  # 'scale = TRUE' is used to
    standardize the variables
```

R Copy code

```r
# Assuming your data is in a data frame called 'data'
# You can specify the variables that you want to include in the PCA as the f
# Here, we will use all columns in 'data' for PCA
pca_result <- prcomp(data, scale = TRUE)  # 'scale = TRUE' is used to standa
```

Step 4: Explore the results. After performing PCA, you can access various aspects of the PCA results, such as the principal components, eigenvalues, and variance explained by each component.

R

```r
    # Access the principal components (scores)
pca_scores <- pca_result$x

# Access the loadings (coefficients) for each original variable in the principal
    components
pca_loadings <- pca_result$rotation

# Access the standard deviations (square root of eigenvalues)
pca_sd <- pca_result$sdev

# Access the proportion of variance explained by each component
    pca_var_exp <- pca_result$sdev^2 / sum(pca_result$sdev^2)
```

```R
# Access the principal components (scores)
pca_scores <- pca_result$x

# Access the loadings (coefficients) for each original variable in the princ
pca_loadings <- pca_result$rotation

# Access the standard deviations (square root of eigenvalues)
pca_sd <- pca_result$sdev

# Access the proportion of variance explained by each component
pca_var_exp <- pca_result$sdev^2 / sum(pca_result$sdev^2)
```

Step 5: Visualize the results. You can visualize the PCA results to gain insights into the data.

# Plotting the proportion of variance explained by each component
plot(pca_var_exp, type = "b", xlab = "Principal Component", ylab = "Proportion of Variance Explained", main = "Scree Plot")

```R
# Plotting the proportion of variance explained by each component
plot(pca_var_exp, type = "b", xlab = "Principal Component", ylab = "Proporti
```

This plot, known as the scree plot, helps you determine the number of principal components to retain based on the proportion of variance explained by each component.

Please note that PCA can be sensitive to the scaling of the variables, so it's common to scale the data (as done with `scale = TRUE` in `prcomp`) before performing PCA. Additionally, PCA assumes that the data is linear and continuous, so it may not be appropriate for all types of data.

- ## Ranking Batsmen using the Principal Component Analysis:

This analysis includes the **Runs, Batting Average** (**Ave**), **Batting Strike Rate** (**SR**), **Fours**, **Sixes**, and **Half century** variables, for all batsmen who have played at least five innings in the 2018 IPL season. The number of innings (always plural) played for a batsman is the number of games in which he actually bats; however, in limited-overs cricket, the game could conclude before a batsman ever gets to bat, which would not count as an innings for that particular batsman. Values for each of these variables were collected together into a $(6 \times 1)^T$ column vector of the form (**Runs**, **Ave**, **Batting SR**, **Fours**, **Sixes**, **Half century**)^T for each of the batsmen. These we call the *batting vectors*. Once data have been obtained, the (6×6) sample correlation matrix associated with the sample batting vectors is calculated for the correlation structure inherent in these batting variables. Since these variables are measured on very different scales, they must be standardized before PCA analysis. However, the process of finding the principal components by using the standardized variables is equivalent to finding principal components by using the correlation matrix instead of the covariance matrix.

Once the correlation matrix for the batting vectors is obtained (**Using 'R' Software**), we calculate the ordered eigenvalues correspond to the correlation matrix, and the total variability attributed to each.

Next, we make the scree plot diagram by plotting the ordered eigenvalues sequentially and look for an 'elbow bend ' in the plot. All these further calculations are performed using **Minitab software(Version 21.3.0)**

According to the theory of PCA, both from the **scree plot** and the table of **ordered eigenvalues with total variability attributed to them**, we look for the highest eigenvalue which explains most of the total variability. The components of the corresponding eigenvector are used as the weights for rating formula. In the rating formula we combine the batting variables with their corresponding weights in a linear combination.

Formula : Runs*0.459 + Batting Avg*0.416 + Strike Rate*0.25 + Fours*0.433 + Sixes*0.429 + Half Centuries*0.427

- ## Ranking Bowlers using the Principal Component Analysis

This analysis includes the **Wickets, Bowling Average, Economy Rate and Bowling Strike Rate** variables, for all bowlers who have played at least five innings and have taken at least one wicket in the 2018 IPL season. Values for each of these variables were collected together into a $(4 \ x \ 1)^T$ column vector of the form (**Wickets, Bowling Average, Economy Rate and Bowling Strike Rate**)^T for each of the bowlers. These we call the *bowling vectors*. Once data have been obtained, the $(4 \ x \ 4)$ sample correlation matrix associated with the sample bowling vectors is calculated for the correlation structure inherent in these bowling variables. Since these variables are measured on very different scales, they must be standardized before PCA analysis. However, the process of finding the principal components by using the standardized variables is equivalent to finding principal components by using the correlation matrix instead of the covariance matrix.

Once the correlation matrix for the bowling vectors is obtained (**Using 'R' Software**), we calculate the ordered eigenvalues correspond to the correlation matrix, and the total variability attributed to each.

Next, we make the scree plot diagram by plotting the ordered eigenvalues sequentially and look for an 'elbow bend ' in the plot. All these further calculations are performed using **Minitab software**.

According to the theory of PCA, both from the **scree plot** and the table of **ordered eigenvalues with total variability attributed to them**, we look for the highest eigenvalue which explains most of the total variability. The components of the corresponding eigenvector are used as the weights for rating formula. In the rating formula we combine the bowling variables with their corresponding weights in a linear combination.

Formula : Wickets*(-0.477) + Bowling Avg* 0.59 + Economy Rate*0.35 + Strike Rate*0.549

Calculation and Results:

Table 1 : Ranking of Batsmen by Method I

POS	PLAYER	Runs	4s	6s	Total Boundaries	BF	Percentage Boundaries	SR	Raw Ranking	Actual Ranking	Rank
1	Kane Williamson	735	64	28	92	516	17.82945736	142.44	131.2381461	1.6970443	7
2	Rishabh Pant	684	68	37	105	394	26.64974619	173.6	155.5642622	2.011606018	1
3	KL Rahul	659	66	32	98	416	23.55769231	158.41	143.1500718	1.851077759	2
4	Ambati Rayudu	602	53	34	87	402	21.64179104	149.75	132.6601639	1.715432454	5
5	Shane Watson	555	44	35	79	359	22.00557103	154.59	131.3329182	1.698269801	6
6	Jos Buttler	548	52	21	73	353	20.67988669	155.24	128.5681442	1.662518428	8
7	Virat Kohli	530	52	18	70	381	18.37270341	139.1	117.840342	1.523796905	12
8	Suryakumar Yadav	512	61	16	77	384	20.05208333	133.33	117.7416866	1.522521189	13
9	Dinesh Karthik	498	49	16	65	337	19.28783383	147.77	119.7891935	1.54899756	10
10	Shikhar Dhawan	497	59	14	73	363	20.11019284	136.91	117.8429191	1.523830229	11
11	Chris Lynn	491	56	18	74	377	19.62864721	130.23	114.3844763	1.479109004	15
12	AB de Villiers	480	39	30	69	275	25.09090909	174.54	136.1964121	1.761159783	4
13	MS Dhoni	455	24	30	54	302	17.8807947	150.66	114.4772142	1.4803082	14
14	Suresh Raina	445	46	12	58	336	17.26190476	132.44	107.2093128	1.386326755	18
15	Sanju Samson	441	30	19	49	320	15.3125	137.81	104.6435432	1.353148713	19
16	Shreyas Iyer	411	29	21	50	310	16.12903226	132.58	102.3686933	1.323732563	21
17	Evin Lewis	382	32	24	56	276	20.28985507	138.4	108.7639437	1.406429733	17
18	Ajinkya Rahane	370	39	5	44	313	14.05750799	118.21	90.94129433	1.175964533	29
19	Chris Gayle	368	30	27	57	252	22.61904762	146.03	113.2180436	1.464025829	16
20	Sunil Narine	357	40	23	63	188	33.5106383	189.89	138.9892132	1.797273575	3
21	Robin Uthappa	351	30	21	51	265	19.24528302	132.45	102.4254464	1.324466439	20
22	Andre Russell	316	17	31	48	171	28.07017544	184.79	125.3260372	1.620594648	9
23	Nitish Rana	304	26	14	40	232	17.24137931	131.03	94.1341182	1.217251032	25
24	Karun Nair	301	23	13	36	221	16.28959276	136.19	93.56937811	1.209948361	26
25	Rohit Sharma	286	25	12	37	215	17.20930233	133.02	92.72119907	1.198980533	28
26	Manish Pandey	284	22	5	27	246	10.97560976	115.44	76.70002174	0.991810221	42
27	Ishan Kishan	275	22	17	39	184	21.19565217	149.45	101.7225677	1.315377495	22
28	Yusuf Pathan	260	22	11	33	200	16.5	130	87.97219505	1.137571024	31
29	Hardik Pandya	260	20	11	31	195	15.8974359	133.33	87.8137745	1.135522483	32
30	Mandeep Singh	252	16	11	27	186	14.51612903	135.48	85.07681912	1.100130833	33
31	Prithvi Shaw	245	27	10	37	160	23.125	153.12	101.417089	1.31142734	23
32	Shakib Al Hasan	239	26	5	31	197	15.73604061	121.31	82.22079858	1.063199548	38
33	Krunal Pandya	228	22	10	32	157	20.38216561	145.22	93.50683929	1.20913967	27
34	Rahul Tripathi	226	18	8	26	167	15.56886228	135.32	83.77931091	1.083352716	34
35	Vijay Shankar	212	11	11	22	148	14.86486486	143.24	82.6195679	1.068356046	36
36	Shubman Gill	203	22	5	27	139	19.42446043	146.04	88.88591689	1.149386388	30
37	Quinton de Kock	201	20	8	28	162	17.28395062	124.07	80.53864403	1.041447558	41
38	Ben Stokes	196	13	6	19	161	11.80124224	121.73	70.73265194	0.914645988	49
39	Glenn Maxwell	169	14	9	23	120	19.16666667	140.83	82.22119904	1.063204726	37
40	Faf du Plessis	162	17	6	23	129	17.82945736	125.58	76.0469796	0.983365714	43

41	Parthiv Patel	153	20	4	24	109	22.01834862	140.36	82.84449183	1.071264544	35
42	Alex Hales	148	13	6	19	118	16.10169492	125.42	71.55558866	0.925287407	48
43	Dwayne Bravo	141	8	10	18	91	19.78021978	154.94	80.99760082	1.047382342	40
44	Aaron Finch	134	6	8	14	100	14	134	68.04743659	0.87992339	54
45	Kieron Pollard	133	10	7	17	100	17	133	71.75147364	0.927820401	47
46	Colin de Grandhomme	131	4	10	14	84	16.66666667	155.95	75.27638291	0.973401106	45
47	Brendon McCullum	127	16	6	22	88	25	144.31	81.76678244	1.057328652	39
48	Krishnappa Gowtham	126	9	9	18	64	28.125	196.87	94.77558407	1.225545846	24
49	Wriddhiman Saha	122	17	1	18	102	17.64705882	119.6	67.80488451	0.876786942	56
50	Mayank Agarwal	120	9	5	14	94	14.89361702	127.65	65.65345246	0.848966711	60
51	Jason Roy	120	9	7	16	94	17.0212766	127.65	68.33688099	0.883666204	53
52	D'Arcy Short	115	11	5	16	99	16.16161616	116.16	64.06881139	0.828475671	61
53	Sam Billings	108	8	5	13	78	16.66666667	138.46	67.58875789	0.873992202	57
54	Ravichandran Ashwin	102	7	5	12	71	16.90140845	143.66	67.51949647	0.873096581	58
55	Marcus Stoinis	99	6	4	10	76	13.15789474	130.26	59.81300665	0.773443736	62
56	Ben Cutting	96	5	8	13	58	22.4137931	165.51	75.90477248	0.98152683	44
57	Ravindra Jadeja	89	3	4	7	74	9.459459459	120.27	50.78281877	0.656674113	67
58	Deepak Hooda	87	2	3	5	81	6.172839506	107.4	42.52638172	0.549909884	74
59	Gautam Gambhir	85	8	1	9	88	10.22727273	96.59	47.21294575	0.61051198	69
60	Axar Patel	80	3	4	7	69	10.14492754	115.94	49.39246491	0.638695406	68
61	Moeen Ali	77	4	6	10	46	21.73913043	167.39	70.22544635	0.908087298	50
62	Yuvraj Singh	65	6	2	8	73	10.95890411	89.04	42.80929263	0.553568213	72
63	Washington Sundar	65	5	4	9	38	23.68421053	171.05	68.68384834	0.888152849	52
64	Colin Munro	63	7	4	11	41	26.82926829	153.65	67.82379378	0.877031458	55
65	Harshal Patel	60	1	6	7	33	21.21212121	181.81	66.19477593	0.855966581	59
66	Rashid Khan	59	3	6	9	31	29.03225806	190.32	73.56218953	0.951234822	46
67	Tim Southee	52	5	1	6	46	13.04347826	113.04	45.77148724	0.591872438	71
68	Shreevats Goswami	52	6	1	7	40	17.5	130	52.64432461	0.680745299	65
69	Sarfaraz Khan	51	7	1	8	41	19.51219512	124.39	53.16926053	0.687533261	64
70	Deepak Chahar	50	1	4	5	29	17.24137931	172.41	57.4327763	0.742664908	63
71	Shreyas Gopal	50	5	0	5	45	11.11111111	111.11	42.78741144	0.553285267	73
72	Rahul Tewatia	50	5	1	6	43	13.95348837	116.27	46.58949239	0.602450087	70
73	Jaydev Unadkat	49	6	1	7	38	18.42105263	128.94	52.26397962	0.675827047	66
74	Manoj Tiwary	47	4	1	5	44	11.36363636	106.81	41.59430768	0.537857208	75
75	Stuart Binny	44	2	2	4	39	10.25641026	112.82	40.27363723	0.520779579	76
76	JP Duminy	36	3	1	4	40	10	90	34.40721758	0.444920735	79
77	Andrew Tye	32	2	1	3	38	7.894736842	84.21	30.08052053	0.38897209	81
78	Harbhajan Singh	29	3	1	4	36	11.11111111	80.55	31.7367783	0.410389207	80
79	Piyush Chawla	27	1	1	2	34	5.882352941	79.41	25.47822616	0.329459687	85
80	Mohammed Siraj	25	2	1	3	22	13.63636364	113.63	36.49742187	0.471949228	78
81	Tom Curran	23	3	0	3	28	10.71428571	82.14	29.29168744	0.378771666	82
82	Mayank Markande	21	2	0	2	24	8.333333333	87.5	26.98490995	0.348942659	84
83	Chris Woakes	17	1	1	2	19	10.52631579	89.47	27.21771826	0.35195311	83
84	Mitchell Johnson	16	2	0	2	11	18.18181818	145.45	37.62312504	0.486505729	77
85	Shardul Thakur	15	3	0	3	5	60	300	68.87955852	0.890683583	51
86	Jofra Archer	15	2	0	2	21	9.523809524	71.42	23.31667765	0.30150864	86

87	Bhuvneshwar Kumar	13	1	0	1	16	6.25	81.25	20.55890398	0.265847788	88
88	Shivam Mavi	13	1	0	1	15	6.666666667	86.66	21.46677686	0.277587519	87

Analysis:

In Table 1, all the batsmen were listed in a descending order of their total runs scored in the IPL season 2018. If we look at the ranking of batsmen by Method I which is performed in the Table 1, we can observe some noteworthy points.

Kane Williamson who had scored the most run in that season is placed at no. 7 in our ranking, whereas Rishav Pant, the second highest scorer of that season is moved to the top in our rank list. Surprisingly, Sunil Narine, who was at no 20 according to the total runs scored, has come up to the place 3 in our ranking.

In search of the reasons behind these, the key point which has come up is that, not only the runs scored, but also the strike rate as well as the ability to hit boundaries have played a great role in our raking system. It can be easily observed from Table 1, that Kane Williamson had a quite weak strike rate for T20 format, whereas Rishav Pant and company had better strike rate than him. If we look at Sunil Narine's case, he had an excellent strike rate as well as his percentage boundary score is also higher than the other two batsmen stated above, which indicates that his ability to hit boundaries is also better than others, which helped him to come up in a good place of our ranking. The same happens for ABD and Andre Russell too, though they were at no. 12 and at no. 22 according to the total runs scored, but their extraordinary strike rate and percentage boundary score pushed them up at no. 4 and at no. 8 in our rank list.

From all these evidences, it is quite clear that the strike rate and the ability to hit boundaries are more important factors in T20 format, than to score a long slow innings.

Table 2 : Ranking of Bowlers by Method I

POS	PLAYER	Avg	SR	Econ	Raw Rating	Actual Rating	Rank
1	Andrew Tye	18.66	14	8	12.40620745	0.696734734	4
2	Rashid Khan	21.8	19.42	6.73	13.58390909	0.76287466	12
3	Siddarth Kaul	26.04	18.85	8.28	15.31850087	0.860289631	25
4	Umesh Yadav	20.9	15.95	7.86	13.31255009	0.747635092	9
5	Trent Boult	25.88	17.55	8.84	15.29988158	0.859243969	24
6	Hardik Pandya	21.16	14.22	8.92	13.48175908	0.757137898	10
7	Sunil Narine	27.47	21.52	7.65	15.79247787	0.886908261	30
8	Kuldeep Yadav	24.58	18.11	8.14	14.76449789	0.829176729	23
9	Jasprit Bumrah	21.88	19.05	6.88	13.62345498	0.765095564	13
10	Shardul Thakur	26.93	17.5	9.23	15.71802921	0.882727211	29
11	Mayank Markande	24.53	17.6	8.36	14.76198527	0.82903562	22
12	Jofra Archer	21.66	15.53	8.36	13.64584515	0.766353	14
13	Shakib Al Hasan	32.57	24.42	8	17.61751664	0.989402751	41
14	Dwayne Bravo	38.07	22.92	9.96	19.60657074	1.10110837	45
15	Piyush Chawla	29.42	21	8.4	16.55421166	0.929687361	37
16	Mujeeb Ur Rahman	20.64	17.71	6.99	13.14563556	0.738261144	8
17	Mitchell McClenaghan	23.71	17.14	8.3	14.44647632	0.811316584	18
18	Andre Russell	27.3	17.46	9.38	15.86494034	0.890977766	32
19	Krunal Pandya	23.66	20.08	7.07	14.33512978	0.805063343	17
20	Yuzvendra Chahal	30.25	25	7.26	16.75268283	0.940833536	38
21	Sandeep Sharma	27.75	22	7.56	15.88661772	0.892195171	34
22	Amit Mishra	22	18.5	7.13	13.69419793	0.7690685	15
23	Ravindra Jadeja	27.54	22.36	7.39	15.80204582	0.887445599	31
24	Krishnappa Gowtham	28.36	21.81	7.8	16.13032956	0.90588207	35
25	Jaydev Unadkat	44.18	27.45	9.65	21.50610662	1.20778663	51
26	Shreyas Gopal	21.45	16.9	7.61	13.51273079	0.758877275	11
27	Mohammed Siraj	33.36	22.36	8.95	17.96668813	1.009012282	43
28	Ankit Rajpoot	20.27	14.27	8.52	13.10022292	0.735710762	7
29	Lungi Ngidi	14.18	14.18	6	10.31506545	0.579295842	3
30	Ravichandran Ashwin	41	30.4	8.09	20.3762562	1.144334037	49
31	Deepak Chahar	27.8	22.9	7.28	15.88355672	0.892023265	33
32	Prasidh Krishna	26	16.8	9.28	15.37576804	0.863505766	26
33	Bhuvneshwar Kumar	39.33	30.77	7.66	19.79994129	1.111968093	47
34	Ben Laughlin	23.44	14	10.04	14.44943973	0.811483009	19
35	Ben Stokes	37.87	27.75	8.18	19.3858166	1.088710779	44
36	Chris Woakes	23.75	13.75	10.36	14.58888444	0.819314248	21

37	Harbhajan Singh	38.57	27.28	8.48	19.64309865	1.103159783	46
38	Mohit Sharma	46	25.42	10.85	22.16380308	1.244722976	52
39	Mustafizur Rahman	32.85	23.57	8.36	17.74416622	0.996515413	42
40	Harshal Patel	23.85	15	9.54	14.58107451	0.818875641	20
41	Shane Watson	41.83	28	8.96	20.71869863	1.163565662	50
42	Rahul Tewatia	28.83	22	7.86	16.30268809	0.915561754	36
43	Imran Tahir	31.33	20.66	9.09	17.27626114	0.970237783	39
44	Tom Curran	19.66	10.16	11.6	13.0076677	0.730512845	6
45	Glenn Maxwell	26.4	19.2	8.25	15.4546276	0.867934532	28
46	Shivam Mavi	54	33.6	9.64	24.40723553	1.370714526	57
47	Tim Southee	52.2	34.8	9	23.82824352	1.338198236	56
48	Ish Sodhi	27	27.6	5.86	15.45365613	0.867879974	27
49	Nitish Rana	11	9.25	7.13	8.848867729	0.496953927	1
50	Dhawal Kulkarni	47	29.75	9.47	22.34238595	1.25475222	53
51	Liam Plunkett	56.25	37.5	9	24.97680049	1.402701392	59
52	Washington Sundar	48	30	9.6	22.66014797	1.27259779	54
53	Karn Sharma	22.25	14.25	9.36	13.94215716	0.782993933	16
54	Barinder Sran	57.25	33	10.4	25.3947875	1.426175614	60
55	Avesh Khan	51	28.5	10.73	23.64241051	1.327761822	55
56	Axar Patel	72.66	52	8.38	29.25970939	1.643230289	63
57	Marcus Stoinis	40	22	10.9	20.29856264	1.139970753	48
58	Shahbaz Nadeem	53.66	30	10.73	24.41558372	1.371183362	58
59	Moeen Ali	32.33	26.33	7.36	17.47331419	0.981304317	40
60	Colin de Grandhomme	64.5	45	8.6	27.17643409	1.526233191	62
61	Ben Cutting	84	51	9.88	32.27319882	1.812468371	64
62	Mitchell Johnson	108	63	10.28	37.86578463	2.126548948	65
63	Yusuf Pathan	14	12	7	10.29566722	0.578206434	2
64	Vijay Shankar	58	30	11.6	25.72337021	1.444628875	61
65	D'Arcy Short	19	18	6.33	12.42754601	0.697933112	5
				Total	1157.403879	65	

Analysis :

Table 2, all the bowlers were listed in a descending order of the no of wickets taken by them in the whole season. According to the table 2, Andrew Tye was at no. 1 in terms of no. of wickets taken, but in our rank list i.e. in Table 2, he is placed at no. 4. Another striking fact is that the most of the top ten wicket takers are not in top 10 ranks in our ranking system. And to our surprise, players like Yusuf Pathan, Nitish Rana, Darcy Short, who were very far behind in the wickets column made a fantastic recovery to feature within top 10 of our rank list. This is mostly due to the fact that they bowled fewer overs with very low economy rate and even managed to take a respectable number of wickets, thus, maintaining a wonderful bowling strike rate.

Here one point should be noted that, while ranking the bowlers with Method I, we didn't take into account the variable 'Wickets'. We considered only the variables 'Bowling Average', 'Economy Rate' and 'Bowling Strike Rate'. Due to this fact, Lungi Ngidi, despite taking less no. of wickets than the players in top 10, he is placed at rank 3 in our list because his Economy rate is very good, considering T20 format. His bowling average and bowling strike rate is also much better than the top 10 wicket takers of that IPL season. Whereas players like Jasprit Bumrah, Siddharth Kaul, though being some of the highest wicket takers of that season, has been moved down to the 13th and 25th place in our ranking due to weaker bowling average and comparatively high bowling strike rate.

In spite of knowing the facts that the economy rate, bowling strike rate are very much important factors in T20 format, but there may be some questions that may arise whether we should take the variable 'Wickets' into our account, besides the other factors, while preparing a ranking method, because it is also an important factor in some situation like batting powerplay, slog overs etc. The fact that most of the top ten wicket takers could not take place in our top ten ranking by Method I, which strengthens the above stated question. We have tried to improve this in our next ranking method.

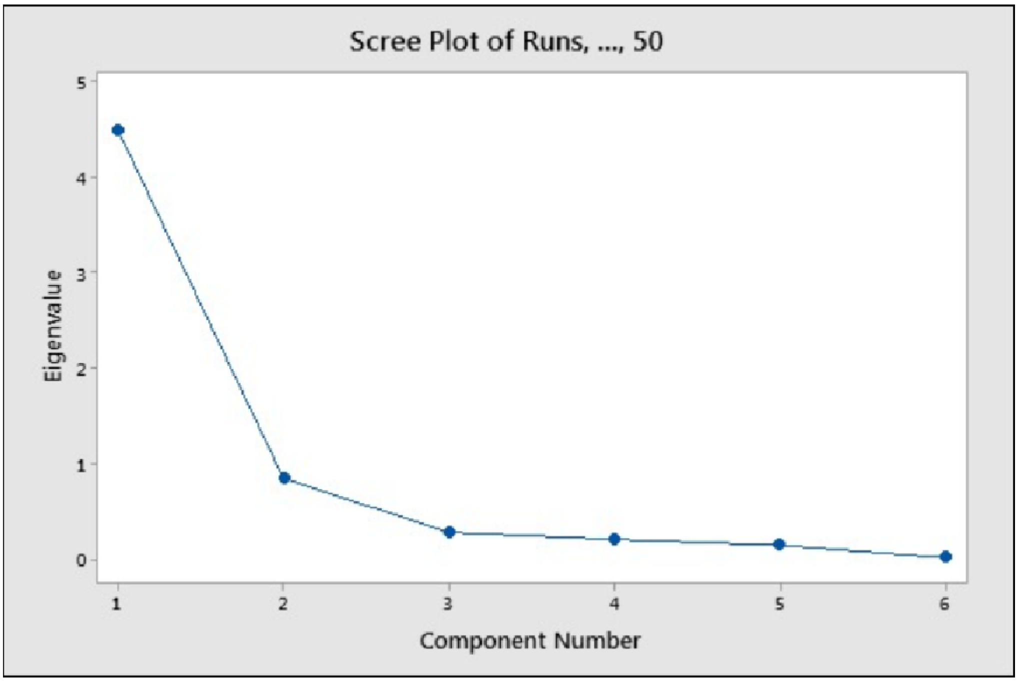
Correlation Matrix of Batsmen :-

Correlation Matrix	Runs	Avg	Strike Rate	Fours	Sixes	Half Centuries
Runs	1	0.7652637	0.2830077	0.9634222	0.8936323	0.8882752
Avg	0.7652737	1	0.3666226	0.6875607	0.731419	0.7265134
Strike Rate	0.2830077	0.3666226	1	0.2439821	0.4040558	0.2334416
Fours	0.9634222	0.6875607	0.2439821	1	0.7912434	0.8652765

Ordered Eigen Values and Corresponding Percentages of Total Variability for Batsmen :-

Eigen Values	4.5004	0.8482	0.2818	0.2078	0.1476	0.0142
Total Variabilty	75	14.1	4.7	3.5	2.5	0.2

Batsman scree plot :-



Eigen Values and Eigen Vector pairs for the Correlation Matrix of Batsmen :-

Eigen values	4.5004	0.8482	0.2818	0.2078	0.01476	0.0142
Variable	PC1	PC2	PC3	PC4	PC5	PC6
Runs	0.459	-0.166	-0.195	0.106	-0.23	-0.812
Average	0.416	-0.008	0.86	-0.085	-0.266	0.094
Strike Rate	0.25	0.903	-0.139	-0.316	0	-0.051
Fours	0.433	-0.244	-0.443	-0.24	-0.491	0.508
Sixes	0.429	0.166	-0.079	0.808	0.245	0.264
Half Centuries	0.427	-0.264	-0.019	-0.414	0.758	0.031

Table 3 : Ranking of Batsmen by PCA

POS	PLAYER	Runs	Avg	SR	4s	6s	50	Rating	Rank
1	Kane Williamson	2.84957	1.90959	0.246828	2.525958	1.924318	4.040615	5.808664	1
2	Rishabh Pant	2.573102	1.917258	1.213153	2.747191	2.863934	2.27159	5.670052	2
3	KL Rahul	2.437578	2.077605	0.742085	2.636574	2.341925	2.861265	5.536736	3
4	Ambati Rayudu	2.128584	1.247289	0.473524	1.917566	2.550729	1.092239	4.005228	6
5	Shane Watson	1.8738	1.013043	0.62362	1.41979	2.655131	0.502564	3.40582	8
6	Jos Buttler	1.835853	2.069936	0.643778	1.862257	1.193505	2.27159	4.153035	5
7	Virat Kohli	1.738276	1.608417	0.143249	1.862257	0.8803	1.681914	3.404966	9
8	Suryakumar Yadav	1.640699	0.799015	-0.03569	2.360033	0.671497	1.681914	3.104693	10
9	Dinesh Karthik	1.564806	1.721357	0.412121	1.696332	0.671497	0.502564	2.774539	12
10	Shikhar Dhawan	1.559385	0.914744	0.075334	2.249416	0.462693	1.681914	3.005794	11
11	Chris Lynn	1.526859	0.531306	-0.13182	2.083491	0.8803	1.092239	2.635082	13
12	AB de Villiers	1.467229	1.967454	1.242304	1.143249	2.133122	2.861265	4.434391	4
13	MS Dhoni	1.331705	3.536062	0.501744	0.313623	2.133122	1.092239	3.724985	7
14	Suresh Raina	1.277496	0.83457	-0.06329	1.530407	0.253889	1.681914	2.407492	15
15	Sanju Samson	1.255812	0.445556	0.103244	0.645473	0.984702	1.092239	1.955893	18
16	Shreyas Iyer	1.093183	0.854091	-0.05895	0.590165	1.193505	1.681914	2.328069	16
17	Evin Lewis	0.935976	0.297758	0.121541	0.75609	1.506711	0.502564	1.772226	19
18	Ajinkya Rahane	0.870925	0.233619	-0.50459	1.143249	-0.47692	-0.08711	0.624024	27
19	Chris Gayle	0.860083	1.099491	0.35816	0.645473	1.819916	1.092239	2.468326	14
20	Sunil Narine	0.800452	-0.19513	1.718333	1.198557	1.402309	0.502564	2.050976	17
21	Robin Uthappa	0.767927	-0.22163	-0.06298	0.645473	1.193505	-0.08711	0.998845	21
22	Andre Russell	0.578193	0.251746	1.560174	-0.07354	2.237523	-0.08711	1.651021	20

23	Nitish Rana	0.513142	-0.12054	-0.10702	0.42424	0.462693	-0.08711	0.50363	28
24	Karun Nair	0.496879	-0.00202	0.053005	0.258315	0.358291	0.502564	0.720631	25
25	Rohit Sharma	0.415565	-0.08917	-0.0453	0.368932	0.253889	0.502564	0.625587	26
26	Manish Pandey	0.404723	0.048872	-0.59049	0.203007	-0.47692	1.092239	0.408165	29
27	Ishan Kishan	0.355935	-0.1533	0.46422	0.203007	0.775898	0.502564	0.851012	22
28	Yusuf Pathan	0.27462	0.2629	-0.13896	0.203007	0.149488	-0.67679	0.063722	33
29	Hardik Pandya	0.27462	0.2629	-0.03569	0.09239	0.149488	-0.08711	0.293434	30
30	Mandeep Singh	0.231253	0.006346	0.030987	-0.12884	0.149488	-0.67679	-0.16412	37
31	Prithvi Shaw	0.193306	0.147172	0.578033	0.479548	0.045086	0.502564	0.73604	23
32	Shakib Al Hasan	0.16078	-0.23627	-0.40845	0.42424	-0.47692	-0.67679	-0.43649	41
33	Krunal Pandya	0.10115	-0.16097	0.333041	0.203007	0.045086	-0.67679	-0.11902	36
34	Rahul Tripathi	0.090308	7.11E-05	0.026025	-0.01823	-0.16372	-0.08711	-0.06734	35
35	Vijay Shankar	0.014415	1.944448	0.271638	-0.40539	0.149488	-0.08711	0.734818	24
36	Shubman Gill	-0.03437	0.607994	0.35847	0.203007	-0.47692	-0.08711	0.172871	31
37	Quinton de Kock	-0.04522	0.000768	-0.32286	0.09239	-0.16372	-0.08711	-0.16858	38
38	Ben Stokes	-0.07232	-0.61203	-0.39542	-0.29477	-0.37252	-0.67679	-0.96309	52
39	Glenn Maxwell	-0.21869	-0.7689	0.196899	-0.23946	-0.05932	-0.67679	-0.78913	50
40	Faf du Plessis	-0.25663	0.5083	-0.27603	-0.07354	-0.37252	-0.08711	-0.2042	40
41	Parthiv Patel	-0.30542	0.382811	0.182324	0.09239	-0.58133	-0.08711	-0.18194	39
42	Alex Hales	-0.33253	-0.0313	-0.28099	-0.29477	-0.37252	-0.67679	-0.81233	51
43	Dwayne Bravo	-0.37047	0.70699	0.634474	-0.57131	0.045086	-0.08711	0.017448	34
44	Aaron Finch	-0.40842	-0.58275	-0.01491	-0.68193	-0.16372	-0.67679	-1.08811	55
45	Kieron Pollard	-0.41384	-0.42589	-0.04592	-0.46069	-0.26812	-0.08711	-0.7303	46
46	Colin de Grandhomme	-0.42468	0.076061	0.665796	-0.79254	0.045086	-0.67679	-0.60966	45
47	Brendon McCullum	-0.44637	-0.27531	0.30482	-0.12884	-0.37252	-0.67679	-0.74779	47
48	Krishnappa Gowtham	-0.45179	-0.77447	1.934795	-0.516	-0.05932	-0.67679	-0.58372	44
49	Wriddhiman Saha	-0.47347	-0.68733	-0.46148	-0.07354	-0.89453	-0.67679	-1.3232	61
50	Mayank Agarwal	-0.48431	-0.9139	-0.21183	-0.516	-0.47692	-0.67679	-1.37246	62

51	Jason Roy	-0.48431	0.340982	-0.21183	-0.516	-0.26812	-0.08711	-0.50906	42
52	D'Arcy Short	-0.51142	-0.60576	-0.56816	-0.40539	-0.47692	-0.67679	-1.2979	58
53	Sam Billings	-0.54936	-0.80933	0.123402	-0.57131	-0.47692	-0.08711	-1.04716	54
54	Ravichandran Ashwin	-0.58189	-0.86162	0.284663	-0.62662	-0.47692	-0.67679	-1.31927	60
55	Marcus Stoinis	-0.59815	-0.02503	-0.13089	-0.68193	-0.58133	-0.67679	-1.15134	57
56	Ben Cutting	-0.61442	-0.07731	0.962268	-0.73724	-0.16372	-0.67679	-0.75206	48
57	Ravindra Jadeja	-0.65236	-0.50955	-0.4407	-0.84785	-0.58133	-0.67679	-1.52708	67
58	Deepak Hooda	-0.6632	-0.23417	-0.83982	-0.90316	-0.68573	-0.67679	-1.58602	68
59	Gautam Gambhir	-0.67405	-0.56532	-1.17506	-0.57131	-0.89453	-0.08711	-1.50665	65
60	Axar Patel	-0.70115	-0.82118	-0.57498	-0.84785	-0.58133	-0.67679	-1.71268	70
61	Moeen Ali	-0.71741	-0.40846	1.02057	-0.79254	-0.37252	-0.08711	-0.78425	49
62	Yuvraj Singh	-0.78246	-0.99547	-1.4092	-0.68193	-0.79013	-0.67679	-2.04879	77
63	Washington Sundar	-0.78246	-0.24045	1.134073	-0.73724	-0.58133	-0.67679	-1.03326	53
64	Colin Munro	-0.79331	-0.87207	0.594469	-0.62662	-0.58133	-0.67679	-1.388	63
65	Harshal Patel	-0.80957	2.432459	1.467759	-0.95847	-0.37252	-0.67679	0.143434	32
66	Rashid Khan	-0.81499	-0.92785	1.731668	-0.84785	-0.37252	-0.67679	-1.14307	56
67	Tim Southee	-0.85294	0.062118	-0.66492	-0.73724	-0.89453	-0.67679	-1.52385	66
68	Shreevats Goswami	-0.85294	-0.54232	-0.13896	-0.68193	-0.89453	-0.67679	-1.61986	69
69	Sarfaraz Khan	-0.85836	-1.03939	-0.31293	-0.62662	-0.89453	-0.67679	-1.84867	74
70	Deepak Chahar	-0.86378	-0.58903	1.176249	-0.95847	-0.58133	-0.67679	-1.30084	59
71	Shreyas Gopal	-0.86378	-0.58903	-0.72477	-0.73724	-0.99893	-0.67679	-1.85945	75
72	Rahul Tewatia	-0.86378	-0.58903	-0.56475	-0.73724	-0.89453	-0.67679	-1.77466	72
73	Jaydev Unadkat	-0.8692	-0.89648	-0.17183	-0.68193	-0.89453	-0.67679	-1.78287	73
74	Manoj Tiwary	-0.88004	-0.65874	-0.85812	-0.79254	-0.89453	-0.67679	-1.90842	76
75	Stuart Binny	-0.8963	-1.137	-0.67174	-0.90316	-0.79013	-0.67679	-2.07135	79
76	JP Duminy	-0.93967	0.759277	-1.37942	-0.84785	-0.89453	-0.67679	-1.50017	64
77	Andrew Tye	-0.96136	-1.37891	-1.55898	-0.90316	-0.89453	-0.67679	-2.46844	84
78	Harbhajan Singh	-0.97762	-1.07704	-1.67249	-0.84785	-0.89453	-0.67679	-2.35476	80
79	Piyush Chawla	-0.98846	-1.27991	-1.70784	-0.95847	-0.89453	-0.67679	-2.50087	85
80	Mohammed Siraj	-0.9993	-0.87905	-0.64662	-0.90316	-0.89453	-0.67679	-2.04983	78
81	Tom Curran	-1.01014	-1.21647	-1.62318	-0.84785	-0.99893	-0.67679	-2.46015	83
82	Mayank Markande	-1.02099	-1.01848	-1.45695	-0.90316	-0.99893	-0.67679	-2.36516	81
83	Chris Woakes	-1.04267	-1.15791	-1.39586	-0.95847	-0.89453	-0.67679	-2.397	82

84	Mitchell Johnson	-1.04809	-0.63504	0.340173	-0.90316	-0.99893	-0.67679	-1.76881	71
85	Shardul Thakur	-1.05351	-0.70476	5.133031	-0.84785	-0.99893	-0.67679	-0.57813	43
86	Jofra Archer	-1.05351	-1.54135	-1.95562	-0.90316	-0.99893	-0.67679	-2.72227	88
87	Bhuvneshwar Kumar	-1.06435	-1.29734	-1.65078	-0.95847	-0.99893	-0.67679	-2.57347	86
88	Shivam Mavi	-1.06435	-1.44863	-1.483	-0.95847	-0.99893	-0.67679	-2.59446	87

Analysis :

In the second method of ranking the batsmen [performed in Table 3, i.e. ranking by the help of Principal Component Analysis, we have included two more factors under our consideration, which are 'Batting Average', and 'Half Century'. In a shorter format like T20, a half century worths like a century in bigger format, which may change the result of a match. Also, unlike the previous method, here we have chosen the weights which are assigned to different variables used in ranking, in an objective way.

Unlike the previous ranking, here, Kane Williamson has secured his place at the top followed by Rishav Pant and KL Rahul. Sunil Narine has been moved to the 17th position in this method.

Actually, in this method, the 'batting average' and the 'half centuries' have been assigned sufficient amount of weights, which reflects on the ranking. So the batsmen like Kane Williamson, Virat Kohli, who like to build their innings and then accelerate have been ranked higher, whereas the players like Sunil Narine who loves to play big and risky shots from the beginning of the innings, fails to score long fifties and gets out more no. of times, failed to make a significant rank jump.

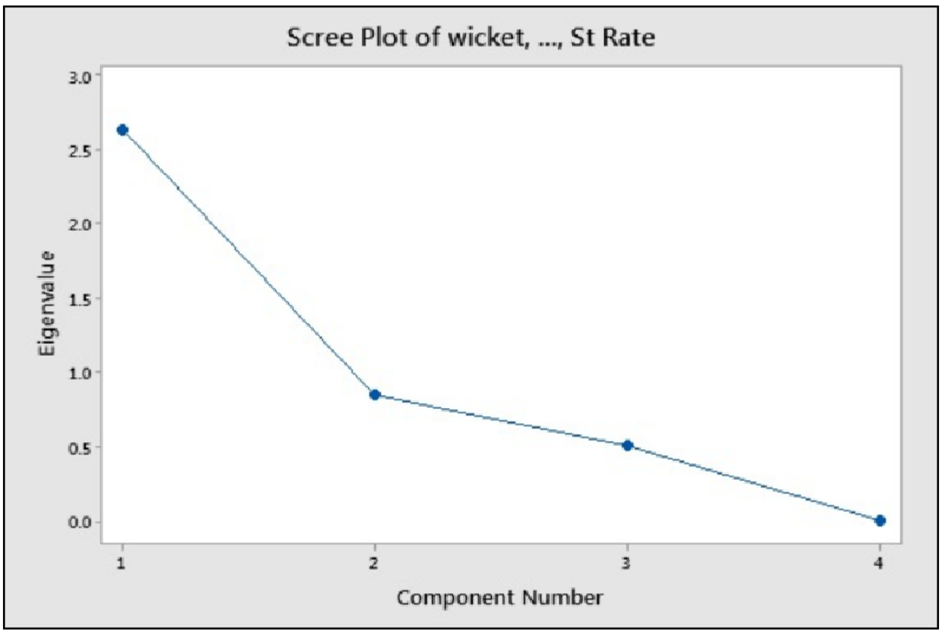
Correlation matrix for bowlers:-

Correlation Matrix	Wickets	Bowling Avg	Economy Rate	Strike Rate
Wickets	1	-0.5173652	-0.3314173	-0.4782758
Bowling Avg	-0.5173652	1	0.4992704	0.9575183
Economy Rate	-0.3314173	0.4992704	1	0.250687
Strike Rate	-0.4782758	0.9575183	0.250687	1

Ordered Eigen Values and Corresponding Percentages of Total Variability for Bowlers:-

Eigen Values	2.6366	0.8493	0.5085	0.0056
Total Variability	65.9	21.2	12.7	0.1

Bowler scree plot :-



Eigen Values and Eigen Vector pairs for the Correlation Matrix of Bowlers:-

Eigen Values	2.6366	0.8493	0.5085	0.0056
Variables	PC1	PC2	PC3	PC4
Wickets	-0.477	-0.139	-0.867	-0.018
Bowling Average	0.59	-0.194	-0.308	0.721
Economy Rate	0.35	0.856	-0.326	-0.195
Strike Rate	0.549	-0.458	-0.215	-0.665

Table 4 : Ranking of Bowlers by PCA

POS	PLAYER	Wkts	Avg	Econ	SR	Rating	Rank
1	Andrew Tye	2.577115	-0.9074	-0.44995	-0.96657	-2.45277	1
2	Rashid Khan	2.047012	-0.72474	-1.41848	-0.41743	-2.12966	2
3	Siddarth Kaul	2.047012	-0.47809	-0.23641	-0.47518	-1.60212	8
4	Umesh Yadav	1.870311	-0.77709	-0.55671	-0.769	-1.96765	4
5	Trent Boult	1.516909	-0.4874	0.190656	-0.60689	-1.27758	13
6	Hardik Pandya	1.516909	-0.76197	0.251666	-0.94428	-1.60345	7
7	Sunil Narine	1.340208	-0.39491	-0.71687	-0.20466	-1.23554	14
8	Kuldeep Yadav	1.340208	-0.56302	-0.34318	-0.55015	-1.39361	11
9	Jasprit Bumrah	1.340208	-0.72008	-1.30409	-0.45491	-1.77031	5
10	Shardul Thakur	1.163508	-0.42632	0.488079	-0.61196	-0.97166	21
11	Mayank Markande	0.986807	-0.56593	-0.1754	-0.60182	-1.1964	16
12	Jofra Archer	0.986807	-0.73288	-0.1754	-0.81155	-1.41004	10
13	Shakib Al Hasan	0.810106	-0.09823	-0.44995	0.089159	-0.55291	30
14	Dwayne Bravo	0.810106	0.221714	1.044793	-0.06282	0.075582	41
15	Piyush Chawla	0.810106	-0.28147	-0.1449	-0.25735	-0.74449	24
16	Mujeeb Ur Rahman	0.810106	-0.79222	-1.2202	-0.59068	-1.60518	6
17	Mitchell McClenaghan	0.810106	-0.61363	-0.22116	-0.64843	-1.18186	17
18	Andre Russell	0.633405	-0.40479	0.602472	-0.61601	-0.66829	26
19	Krunal Pandya	0.456704	-0.61654	-1.15919	-0.35056	-1.17978	18
20	Yuzvendra Chahal	0.456704	-0.23319	-1.01429	0.147923	-0.62922	29
21	Sandeep Sharma	0.456704	-0.37862	-0.7855	-0.15603	-0.80182	22
22	Amit Mishra	0.456704	-0.7131	-1.11343	-0.51064	-1.30862	12
23	Ravindra Jadeja	0.280003	-0.39083	-0.91515	-0.11955	-0.75009	23
24	Krishnappa Gowtham	0.280003	-0.34313	-0.60247	-0.17528	-0.6431	28
25	Jaydev Unadkat	0.280003	0.577143	0.80838	0.39615	0.707373	49
26	Shreyas Gopal	0.280003	-0.7451	-0.74737	-0.67275	-1.20409	15
27	Mohammed Siraj	0.280003	-0.05227	0.274544	-0.11955	-0.13395	40
28	Ankit Rajpoot	0.280003	-0.81374	-0.05338	-0.93921	-1.14798	19
29	Lungi Ngidi	0.280003	-1.16801	-1.97519	-0.94833	-2.03464	3
30	Ravichandran Ashwin	0.103302	0.392157	-0.38131	0.695036	0.430213	47
31	Deepak Chahar	0.103302	-0.37571	-0.99904	-0.06484	-0.6562	27
32	Prasidh Krishna	0.103302	-0.48042	0.52621	-0.68288	-0.52345	31
33	Bhuvneshwar Kumar	-0.0734	0.29501	-0.70924	0.732523	0.362989	45
34	Ben Laughlin	-0.0734	-0.62934	1.105803	-0.96657	-0.47991	32
35	Ben Stokes	-0.2501	0.21008	-0.31268	0.426545	0.367982	46
36	Chris Woakes	-0.2501	-0.6113	1.349843	-0.9919	-0.31348	35
37	Harbhajan Singh	-0.4268	0.2508	-0.08389	0.378926	0.530226	48

38	Mohit Sharma	-0.4268	0.683015	1.723528	0.190476	1.314369	52
39	Mustafizur Rahman	-0.4268	-0.08194	-0.1754	0.00304	0.095516	42
40	Harshal Patel	-0.4268	-0.60549	0.724492	-0.86525	-0.3751	34
41	Shane Watson	-0.6035	0.44044	0.28217	0.451875	0.894569	50
42	Rahul Tewatia	-0.6035	-0.31579	-0.55671	-0.15603	-0.17896	38
43	Imran Tahir	-0.6035	-0.17036	0.381311	-0.29179	0.160621	43
44	Tom Curran	-0.6035	-0.84923	2.295495	-1.35562	-0.15399	39
45	Glenn Maxwell	-0.7802	-0.45715	-0.25929	-0.43972	-0.22972	37
46	Shivam Mavi	-0.7802	1.148388	0.800754	1.019251	1.889539	57
47	Tim Southee	-0.7802	1.043679	0.312675	1.140832	1.723681	55
48	Ish Sodhi	-0.7802	-0.42225	-2.08196	0.411348	-0.37982	33
49	Nitish Rana	-0.9569	-1.35299	-1.11343	-1.44782	-1.52638	9
50	Dhawal Kulkarni	-0.9569	0.741187	0.671108	0.62918	1.474051	53
51	Liam Plunkett	-0.9569	1.279274	0.312675	1.414388	2.09715	58
52	Washington Sundar	-0.9569	0.799358	0.770249	0.654509	1.556977	54
53	Karn Sharma	-0.9569	-0.69856	0.58722	-0.94124	-0.26692	36
54	Barinder Sran	-0.9569	1.337446	1.380348	0.958461	2.254853	60
55	Avesh Khan	-0.9569	0.973873	1.632013	0.502533	1.878124	56
56	Axar Patel	-1.1336	2.233871	-0.16015	2.883488	3.385695	63
57	Marcus Stoinis	-1.1336	0.333985	1.761659	-0.15603	1.268702	51
58	Shahbaz Nadeem	-1.1336	1.12861	1.632013	0.654509	2.137139	59
59	Moeen Ali	-1.1336	-0.11219	-0.93803	0.282675	0.301416	44
60	Colin de Grandhomme	-1.31031	1.75919	0.007626	2.174267	2.85928	62
61	Ben Cutting	-1.31031	2.893537	0.983784	2.782171	4.203939	64
62	Mitchell Johnson	-1.31031	4.289656	1.288833	3.997977	5.801894	65
63	Yusuf Pathan	-1.48701	-1.17848	-1.21257	-1.1692	-1.05229	20
64	Vijay Shankar	-1.48701	1.381075	2.295495	0.654509	2.686885	61
65	D'Arcy Short	-1.48701	-0.88762	-1.72353	-0.5613	-0.72578	25

Analysis :

Similar to the third table, we have used one more variable named 'Wickets', in our second method of ranking the bowlers. Here also, like the case of batsmen, the weights assigned to the different variables used in the formula for ranking, has been chosen objectively with the help of PCA. This ranking is done in Table 4.

If we look at the Table 4, we can now observe that many of the top 10 wicket takers secured their place in top 10 ranking. This is due to the fact that in this method, the variable 'Wickets' has been given a lot more importance by assigning a sufficient weight to this variable. Its reflection falls on the ranking. Here Andrew Tye has secured his place at the top, followed by Rashid Khan. The interesting part is, Lungi Ngidi still manages to hold on to the third rank in this method as well. This is a huge evidence that in t20 format, economy rate is as important as taking wickets. The bowlers who managed to keep a balance between the two variables have been more successful and in higher ranks.

To measure the extent to which the two sets of ranks are associated with each other, we calculated the correlation coefficients between the two sets of ranks for both the cases of Batsmen and Bowlers. For Batsmen, the obtained Rank Correlation Coefficient between the two ranks by Method I and Method II is 0.9472258, which concludes that the two sets of ranks are highly positively correlated. For Bowlers, the obtained Rank Correlation Coefficient between the two ranks by Method I and Method II is 0.9357185, which leads to the same conclusion that there is a high degree of agreement between the two sets of ranks.

Conclusion:

Quantifying athletic performance (performance analysis) is a challenging task in any sport. It is especially important in competitive sports impacted by player auctions or trades which, by their nature, usually involve organizations spending large monetary sums with the hope of future, competitive advantages. The basis of these transactions lies in past player performance, and there are typically several indicators or aspects available to measure the various contributions of prized athletes.

Interested candidates in future can also use our work as a foundation to figure out the auction results of the following years based on previous years' performance. They can also work on how a particular individual's effect on the overall results of the team.

Unfortunately, these indicators are generally highly correlated with one another, making it difficult to judge overall player performance. There is no doubt that currently available player ranking procedures are opaque, so there is a need for new and transparent methods of performance analysis.

Although expert opinion can be quite valuable; it is also very subjective. Here, we have worked on two different methods to rank the individual performances based on important variables. On one of these methods, we tried to find the ranks using weightage in a traditional subjective way. We referred to previous thesis papers on the same topic for finding the weightage. Secondly, we have demonstrated a simple method using principal component analysis that can be directly applied to correlated, multivariate data. Using 2018 Indian Premier League (IPL 2018) data, we have shown how to rank batsmen and bowlers based on their contributions to their teams during this competitive season. The simplicity and straight-forwardness of this technique make it very appealing as an introduction to the topic of principal component analysis. Moreover, this real classroom example is accessible to upper undergraduate and graduate students having just a basic understanding of multivariate statistics. As a result, the motivation and relevance of principal component analysis becomes immediately apparent to them, and real learning soon follows.

ACKNOWLEDGEMENT

I am indebted to a number of people for helping me in the preparation of this project.

Firstly, I would like to mention Dr. Apurba Roy, Vice-Principal, Asutosh College, University of Calcutta, without whose help I couldn't have been a part of this prestigious college.

I owe a deep debt of gratitude to my supervisor Ms. Oindrila Bose for necessary guidance for the presentation of this dissertation, valuable comments and suggestions. I am extremely grateful to her for giving me the necessary stimulus, support and valuable time.

Special thanks to Dr. Dhiman Dutta , Head of the Department of Statistics, Asutosh College.

I am greatly indebted to Dr. Shirsendu Mukherjee, Dr. Sankha Bhattacharya, Dr. Parthasarathi Bera (Faculty members) who often took pains and stood by me in adverse circumstances. Without their encouragement and inspiration, it was not possible for me to complete this project.

Finally, my earnest thanks go to my friends who were always beside me when I needed them without any excuses and made these three years worthwhile.

This project is not only a mere project. It is the memories spent with the whole department which has created a mutual understanding among us. There are many emotions related to this piece of work,
especially respect and duty towards teachers and vice versa; educational attachment with my friends; social
attachment with my college.

REFERENCES:

https://www.iplt20.com/stats/2018/most-runs

https://www.iplt20.com/stats/2018/most-wickets

http://jse.amstat.org/v21n3/scariano.pdf

https://researcharchive.vuw.ac.nz/xmlui/handle/10063/5296

www.google.com

www.youtube.com

Fundamental of Statistics (Volume 1) – by A.M.Goon,

M.K.Gupta, B. Dasgupta [8th edition, 2008]

An Introduction to Multivariate Statistical Analysis – by T.W. Anderson [3rd

edition] Aspects of Multivariate Statistical Theory – by R. J. Muirhead

Appendix:

POS	PLAYER	Mat	Inns	NO	Runs	HS	Avg	BF	SR	100	50	4s	6s
1	Kane Williamson	17	17	3	735	84	52.5	516	142.44	0	8	64	28

2	Rishabh Pant	14	14	1	684	128*	52.61	394	173.6	1	5	68	37
3	KL Rahul	14	14	2	659	95*	54.91	416	158.41	0	6	66	32
4	Ambati Rayudu	16	16	2	602	100*	43	402	149.75	1	3	53	34
5	Shane Watson	15	15	1	555	117*	39.64	359	154.59	2	2	44	35
6	Jos Buttler	13	13	3	548	95*	54.8	353	155.24	0	5	52	21
7	Virat Kohli	14	14	3	530	92*	48.18	381	139.1	0	4	52	18
8	Suryakumar Yadav	14	14	0	512	72	36.57	384	133.33	0	4	61	16
9	Dinesh Karthik	16	16	6	498	52	49.8	337	147.77	0	2	49	16
10	Shikhar Dhawan	16	16	3	497	92*	38.23	363	136.91	0	4	59	14
11	Chris Lynn	16	16	1	491	74	32.73	377	130.23	0	3	56	18
12	AB de Villiers	12	11	2	480	90*	53.33	275	174.54	0	6	39	30
13	MS Dhoni	16	15	9	455	79*	75.83	302	150.66	0	3	24	30
14	Suresh Raina	15	15	3	445	75*	37.08	336	132.44	0	4	46	12
15	Sanju Samson	15	15	1	441	92*	31.5	320	137.81	0	3	30	19
16	Shreyas Iyer	14	14	3	411	93*	37.36	310	132.58	0	4	29	21
17	Evin Lewis	13	13	0	382	65	29.38	276	138.4	0	2	32	24
18	Ajinkya Rahane	15	14	1	370	65*	28.46	313	118.21	0	1	39	5
19	Chris Gayle	11	11	2	368	104*	40.88	252	146.03	1	3	30	27
20	Sunil Narine	16	16	0	357	75	22.31	188	189.89	0	2	40	23
21	Robin Uthappa	16	16	0	351	54	21.93	265	132.45	0	1	30	21
22	Andre Russell	16	14	3	316	88*	28.72	171	184.79	0	1	17	31
23	Nitish Rana	15	15	2	304	59	23.38	232	131.03	0	1	26	14
24	Karun Nair	13	12	0	301	54	25.08	221	136.19	0	2	23	13
25	Rohit Sharma	14	14	2	286	94	23.83	215	133.02	0	2	25	12
26	Manish Pandey	15	13	2	284	62*	25.81	246	115.44	0	3	22	5
27	Ishan Kishan	14	12	0	275	62	22.91	184	149.45	0	2	22	17
28	Yusuf Pathan	15	13	4	260	45*	28.88	200	130	0	0	22	11
29	Hardik Pandya	13	13	4	260	50	28.88	195	133.33	0	1	20	11
30	Mandeep Singh	14	13	3	252	47*	25.2	186	135.48	0	0	16	11
31	Prithvi Shaw	9	9	0	245	65	27.22	160	153.12	0	2	27	10
32	Shakib Al Hasan	17	13	2	239	35	21.72	197	121.31	0	0	26	5
33	Krunal Pandya	14	13	3	228	41*	22.8	157	145.22	0	0	22	10
34	Rahul Tripathi	12	12	3	226	80*	25.11	167	135.32	0	1	18	8
35	Vijay Shankar	13	11	7	212	54*	53	148	143.24	0	1	11	11
36	Shubman Gill	13	11	5	203	57*	33.83	139	146.04	0	1	22	5
37	Quinton de Kock	8	8	0	201	53	25.12	162	124.07	0	1	20	8
38	Ben Stokes	13	13	1	196	45	16.33	161	121.73	0	0	13	6
39	Glenn Maxwell	12	12	0	169	47	14.08	120	140.83	0	0	14	9
40	Faf du Plessis	6	6	1	162	67*	32.4	129	125.58	0	1	17	6

41	Parthiv Patel	6	6	1	153	53	30.6	109	140.36	0	1	20	4
42	Alex Hales	6	6	0	148	45	24.66	118	125.42	0	0	13	6
43	Dwayne Bravo	16	10	6	141	68	35.25	91	154.94	0	1	8	10
44	Aaron Finch	10	9	1	134	46	16.75	100	134	0	0	6	8
45	Kieron Pollard	9	8	1	133	50	19	100	133	0	1	10	7
46	Colin de Grandhomme	9	8	3	131	40	26.2	84	155.95	0	0	4	10

47	Brendon McCullum	6	6	0	127	43	21.16	88	144.31	0	0	16	6
48	Krishnappa Gowtham	15	13	4	126	33*	14	64	196.87	0	0	9	9
49	Wriddhiman Saha	11	10	2	122	35	15.25	102	119.6	0	0	17	1
50	Mayank Agarwal	11	11	1	120	30	12	94	127.65	0	0	9	5
51	Jason Roy	5	5	1	120	91*	30	94	127.65	0	1	9	7
52	D'Arcy Short	7	7	0	115	44	16.42	99	116.16	0	0	11	5
53	Sam Billings	10	8	0	108	56	13.5	78	138.46	0	1	8	5
54	Ravichandran Ashwin	14	9	1	102	45	12.75	71	143.66	0	0	7	5
55	Marcus Stoinis	7	7	3	99	29*	24.75	76	130.26	0	0	6	4
56	Ben Cutting	9	6	2	96	37	24	58	165.51	0	0	5	8
57	Ravindra Jadeja	16	10	5	89	27*	17.8	74	120.27	0	0	3	4
58	Deepak Hooda	9	8	4	87	32*	21.75	81	107.4	0	0	2	3
59	Gautam Gambhir	6	5	0	85	55	17	88	96.59	0	1	8	1
60	Axar Patel	9	8	2	80	19	13.33	69	115.94	0	0	3	4
61	Moeen Ali	5	4	0	77	65	19.25	46	167.39	0	1	4	6
62	Yuvraj Singh	8	6	0	65	20	10.83	73	89.04	0	0	6	2
63	Washington Sundar	7	6	3	65	35	21.66	38	171.05	0	0	5	4
64	Colin Munro	5	5	0	63	33	12.6	41	153.65	0	0	7	4
65	Harshal Patel	5	2	1	60	36*	60	33	181.81	0	0	1	6
66	Rashid Khan	17	7	2	59	34*	11.8	31	190.32	0	0	3	6
67	Tim Southee	8	4	2	52	36*	26	46	113.04	0	0	5	1
68	Shreevats Goswami	6	3	0	52	35	17.33	40	130	0	0	6	1
69	Sarfaraz Khan	7	6	1	51	22*	10.2	41	124.39	0	0	7	1
70	Deepak Chahar	12	4	1	50	39	16.66	29	172.41	0	0	1	4
71	Shreyas Gopal	11	4	1	50	24	16.66	45	111.11	0	0	5	0
72	Rahul Tewatia	8	5	2	50	24	16.66	43	116.27	0	0	5	1
73	Jaydev Unadkat	15	7	3	49	26	12.25	38	128.94	0	0	6	1
74	Manoj Tiwary	5	4	1	47	35	15.66	44	106.81	0	0	4	1
75	Stuart Binny	7	5	0	44	22	8.8	39	112.82	0	0	2	2
76	JP Duminy	6	4	3	36	23	36	40	90	0	0	3	1
77	Andrew Tye	14	8	2	32	14	5.33	38	84.21	0	0	2	1

78	Harbhajan Singh	13	3	0	29	19	9.66	36	80.55	0	0	3	1
79	Piyush Chawla	15	7	3	27	12	6.75	34	79.41	0	0	1	1
80	Mohammed Siraj	11	4	2	25	14	12.5	22	113.63	0	0	2	1
81	Tom Curran	5	4	1	23	18	7.66	28	82.14	0	0	3	0
82	Mayank Markande	14	6	4	21	7*	10.5	24	87.5	0	0	2	0
83	Chris Woakes	5	4	2	17	11	8.5	19	89.47	0	0	1	1
84	Mitchell Johnson	6	2	2	16	12*	16	11	145.45	0	0	2	0
85	Shardul Thakur	13	1	1	15	15*	15	5	300	0	0	3	0
86	Jofra Archer	10	8	3	15	8	3	21	71.42	0	0	2	0
87	Bhuvneshwar Kumar	12	4	2	13	7	6.5	16	81.25	0	0	1	0
88	Shivam Mavi	9	4	1	13	7	4.33	15	86.66	0	0	1	0

POS	PLAYER	Mat	Inns	Ov	Runs	Wkts	Avg	Econ	SR	4w	5w
1	Andrew Tye	14	14	56	448	24	18.66	8	14	3	0
2	Rashid Khan	17	17	68	458	21	21.8	6.73	19.42	0	0
3	Siddarth Kaul	17	17	66	547	21	26.04	8.28	18.85	0	0
4	Umesh Yadav	14	14	53.1	418	20	20.9	7.86	15.95	0	0
5	Trent Boult	14	14	52.4	466	18	25.88	8.84	17.55	0	0
6	Hardik Pandya	13	13	42.4	381	18	21.16	8.92	14.22	0	0
7	Sunil Narine	16	16	61	467	17	27.47	7.65	21.52	0	0
8	Kuldeep Yadav	16	16	51.2	418	17	24.58	8.14	18.11	1	0
9	Jasprit Bumrah	14	14	54	372	17	21.88	6.88	19.05	0	0
10	Shardul Thakur	13	13	46.4	431	16	26.93	9.23	17.5	0	0
11	Mayank Markande	14	14	44	368	15	24.53	8.36	17.6	1	0
12	Jofra Archer	10	10	38.5	325	15	21.66	8.36	15.53	0	0
13	Shakib Al Hasan	17	17	57	456	14	32.57	8	24.42	0	0
14	Dwayne Bravo	16	16	53.3	533	14	38.07	9.96	22.92	0	0
15	Piyush Chawla	15	15	49	412	14	29.42	8.4	21	0	0
16	Mujeeb Ur Rahman	11	11	41.2	289	14	20.64	6.99	17.71	0	0
17	Mitchell McClenaghan	11	11	40	332	14	23.71	8.3	17.14	0	0
18	Andre Russell	16	15	37.5	355	13	27.3	9.38	17.46	0	0
19	Krunal Pandya	14	13	40.1	284	12	23.66	7.07	20.08	0	0
20	Yuzvendra Chahal	14	14	50	363	12	30.25	7.26	25	0	0
21	Sandeep Sharma	12	12	44	333	12	27.75	7.56	22	0	0
22	Amit Mishra	10	10	37	264	12	22	7.13	18.5	0	0
23	Ravindra Jadeja	16	14	41	303	11	27.54	7.39	22.36	0	0
24	Krishnappa Gowtham	15	15	40	312	11	28.36	7.8	21.81	0	0
25	Jaydev Unadkat	15	15	50.2	486	11	44.18	9.65	27.45	0	0
26	Shreyas Gopal	11	10	31	236	11	21.45	7.61	16.9	1	0
27	Mohammed Siraj	11	11	41	367	11	33.36	8.95	22.36	0	0

28	Ankit Rajpoot	8	8	26.1	223	11	20.27	8.52	14.27	0	1
29	Lungi Ngidi	7	7	26	156	11	14.18	6	14.18	1	0
30	Ravichandran Ashwin	14	14	50.4	410	10	41	8.09	30.4	0	0
31	Deepak Chahar	12	12	38.1	278	10	27.8	7.28	22.9	0	0
32	Prasidh Krishna	7	7	28	260	10	26	9.28	16.8	1	0
33	Bhuvneshwar Kumar	12	12	46.1	354	9	39.33	7.66	30.77	0	0
34	Ben Laughlin	7	7	21	211	9	23.44	10.04	14	0	0
35	Ben Stokes	13	12	37	303	8	37.87	8.18	27.75	0	0
36	Chris Woakes	5	5	18.2	190	8	23.75	10.36	13.75	0	0
37	Harbhajan Singh	13	12	31.5	270	7	38.57	8.48	27.28	0	0
38	Mohit Sharma	9	9	29.4	322	7	46	10.85	25.42	0	0
39	Mustafizur Rahman	7	7	27.3	230	7	32.85	8.36	23.57	0	0
40	Harshal Patel	5	5	17.3	167	7	23.85	9.54	15	0	0
41	Shane Watson	15	11	28	251	6	41.83	8.96	28	0	0
42	Rahul Tewatia	8	8	22	173	6	28.83	7.86	22	0	0
43	Imran Tahir	6	6	20.4	188	6	31.33	9.09	20.66	0	0
44	Tom Curran	5	5	10.1	118	6	19.66	11.6	10.16	0	0
45	Glenn Maxwell	12	10	16	132	5	26.4	8.25	19.2	0	0
46	Shivam Mavi	9	9	28	270	5	54	9.64	33.6	0	0
47	Tim Southee	8	8	29	261	5	52.2	9	34.8	0	0
48	Ish Sodhi	6	6	23	135	5	27	5.86	27.6	0	0
49	Nitish Rana	15	5	6.1	44	4	11	7.13	9.25	0	0
50	Dhawal Kulkarni	8	8	19.5	188	4	47	9.47	29.75	0	0
51	Liam Plunkett	7	7	25	225	4	56.25	9	37.5	0	0
52	Washington Sundar	7	7	20	192	4	48	9.6	30	0	0
53	Karn Sharma	6	5	9.3	89	4	22.25	9.36	14.25	0	0
54	Barinder Sran	6	6	22	229	4	57.25	10.4	33	0	0
55	Avesh Khan	6	6	19	204	4	51	10.73	28.5	0	0
56	Axar Patel	9	8	26	218	3	72.66	8.38	52	0	0
57	Marcus Stoinis	7	6	11	120	3	40	10.9	22	0	0
58	Shahbaz Nadeem	6	6	15	161	3	53.66	10.73	30	0	0
59	Moeen Ali	5	5	13.1	97	3	32.33	7.36	26.33	0	0
60	Colin de Grandhomme	9	7	15	129	2	64.5	8.6	45	0	0
61	Ben Cutting	9	7	17	168	2	84	9.88	51	0	0
62	Mitchell Johnson	6	6	21	216	2	108	10.28	63	0	0
63	Yusuf Pathan	15	1	2	14	1	14	7	12	0	0
64	Vijay Shankar	13	4	5	58	1	58	11.6	30	0	0
65	D'Arcy Short	7	2	3	19	1	19	6.33	18	0	0