**R\_CODE:**

*library(readxl)*

*library(dplyr)*

*library(stats)*

*dataset\_path <- "C:/Users/adars/Downloads/NBA\_Player\_Stats.xlsx"*

*nba\_stats <- read\_excel(dataset\_path)*

*head(nba\_stats)*

*variable\_names <- colnames(nba\_stats)*

*# Create a data frame with descriptions for each variable*

*variable\_descriptions <- data.frame(*

*Initial = variable\_names,*

*FullName = c("Rank", "Player", "Position", "Age", "Team", "Games", "Games Started",*

*"Minutes Played", "Field Goals", "Field Goal Attempts", "Field Goal Percentage",*

*"3-Point Field Goals", "3-Point Field Goal Attempts", "3-Point Field Goal Percentage",*

*"2-Point Field Goals", "2-Point Field Goal Attempts", "2-Point Field Goal Percentage",*

*"Effective Field Goal Percentage", "Free Throws", "Free Throw Attempts", "Free Throw Percentage",*

*"Offensive Rebounds", "Defensive Rebounds", "Total Rebounds", "Assists",*

*"Steals", "Blocks", "Turnovers", "Personal Fouls", "Points", "Season", "Most Valuable Player"),*

*Description = c("A unique identifier for each row/player.",*

*"The name of the player.",*

*"The playing position of the player.",*

*"The age of the player.",*

*"The NBA team the player played for in that season.",*

*"The number of games the player appeared in.",*

*"The number of games the player started.",*

*"The total minutes the player played.",*

*"The total number of field goals made.",*

*"The total number of field goal attempts.",*

*"The percentage of field goal attempts made.",*

*"The total number of 3-point field goals made.",*

*"The total number of 3-point field goal attempts.",*

*"The percentage of 3-point field goal attempts made.",*

*"The total number of 2-point field goals made.",*

*"The total number of 2-point field goal attempts.",*

*"The percentage of 2-point field goal attempts made.",*

*"Adjusted field goal percentage accounting for 3-pointers.",*

*"The total number of free throws made.",*

*"The total number of free throw attempts.",*

*"The percentage of free throw attempts made.",*

*"The total number of offensive rebounds.",*

*"The total number of defensive rebounds.",*

*"The total number of rebounds.",*

*"The total number of assists.",*

*"The total number of steals.",*

*"The total number of blocks.",*

*"The total number of turnovers.",*

*"The total number of personal fouls.",*

*"The total number of points scored.",*

*"The NBA season.",*

*"Boolean indicating if the player was MVP that season.")*

*)*

*# variable descriptions*

*print(variable\_descriptions)*

*# missing values*

*missing\_values <- sapply(nba\_stats, function(x) sum(is.na(x)))*

*print(missing\_values)*

*#Removal of missing values*

*nba\_stats\_clean <- na.omit(nba\_stats)*

*# Adjust column selection as per my dataset structure*

*numerical\_columns <- nba\_stats\_clean %>% select(-c(Rk, Player, Pos, Tm, Season, MVP))*

*# Normalize the numerical data*

*nba\_stats\_normalized <- scale(numerical\_columns)*

*# Checking the structure of the normalized data*

*str(nba\_stats\_normalized)*

*# Perform PCA*

*pca\_result <- prcomp(nba\_stats\_normalized, center = FALSE, scale. = FALSE)*

*# Summary of PCA results*

*summary(pca\_result)*

*# PCA Plot*

*plot(pca\_result$sdev^2 / sum(pca\_result$sdev^2), xlab = "Principal Component", ylab = "Variance Explained", type = 'b')*

*# from the PCA plot i had selected 3 PCA result*

*library(cluster)*

*pca\_scores <- as.data.frame(pca\_result$x[, 1:3])*

*# Print the first few rows of pca\_scores to see PCA scores and cluster assignments*

*print(head(pca\_scores))*

*min\_pc1 <- min(pca\_scores$PC1)*

*print(min\_pc1)*

*max\_pc1 <- max(pca\_scores$PC1)*

*print(max\_pc1)*

*min\_pc2 <- min(pca\_scores$PC2)*

*print(min\_pc2)*

*max\_pc2 <- max(pca\_scores$PC2)*

*print(max\_pc2)*

*min\_pc3 <- min(pca\_scores$PC3)*

*print(min\_pc3)*

*max\_pc3 <- max(pca\_scores$PC3)*

*print(max\_pc3)*

*kmeans\_result <- kmeans(pca\_scores, centers = 5) # Example: 5 clusters*

*# Print the summary of kmeans clustering results*

*print(kmeans\_result)*

*#summary function for a concise overview*

*summary(kmeans\_result)*

*# Adding cluster assignments to PCA scores*

*pca\_scores$cluster <- as.factor(kmeans\_result$cluster)*

*library(ggplot2)*

*# Plot with clusters*

*ggplot(pca\_scores, aes(x = PC1, y = PC2, color = cluster)) +*

*geom\_point() +*

*ggtitle("NBA Player Clusters Based on PCA")*

**OUTPUT:**

|  |
| --- |
| > library(readxl)  > library(dplyr)  > library(stats)  > dataset\_path <- "C:/Users/adars/Downloads/NBA\_Player\_Stats.xlsx"  > nba\_stats <- read\_excel(dataset\_path)  > head(nba\_stats)  # A tibble: 6 × 32  Rk Player Pos Age Tm G GS MP FG FGA `FG%` `3P` `3PA` `3P%` `2P` `2PA`  *<dbl>* *<chr>* *<chr>* *<dbl>* *<chr>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1 Mahmoud … PG 28 SAC 31 0 17.1 3.3 8.8 0.377 0.2 1 0.161 3.2 7.8  2 2 Tariq Ab… SG 23 SAC 59 16 16.3 2.4 6.1 0.403 0.1 0.3 0.211 2.4 5.7  3 3 Shareef … SF 21 VAN 82 82 36 8 16.4 0.485 0.3 0.6 0.412 7.7 15.8  4 4 Cory Ale… PG 24 TOT 60 22 21.6 2.9 6.7 0.428 1.1 2.9 0.375 1.8 3.7  5 4 Cory Ale… PG 24 SAS 37 3 13.5 1.6 3.9 0.414 0.5 1.7 0.313 1.1 2.2  6 4 Cory Ale… PG 24 DEN 23 19 34.7 4.8 11.1 0.435 2 4.9 0.411 2.8 6.2  # ℹ 16 more variables: `2P%` <dbl>, `eFG%` <dbl>, FT <dbl>, FTA <dbl>, `FT%` <dbl>, ORB <dbl>,  # DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>, TOV <dbl>, PF <dbl>, PTS <dbl>,  # Season <chr>, MVP <lgl>  > variable\_names <- colnames(nba\_stats)  > # Create a data frame with descriptions for each variable  > variable\_descriptions <- data.frame(  + Initial = variable\_names,  + FullName = c("Rank", "Player", "Position", "Age", "Team", "Games", "Games Started",  + "Minutes Played", "Field Goals", "Field Goal Attempts", "Field Goal Percentage",  + "3-Point Field Goals", "3-Point Field Goal Attempts", "3-Point Field Goal Percentage",  + "2-Point Field Goals", "2-Point Field Goal Attempts", "2-Point Field Goal Percentage",  + "Effective Field Goal Percentage", "Free Throws", "Free Throw Attempts", "Free Throw Percentage",  + "Offensive Rebounds", "Defensive Rebounds", "Total Rebounds", "Assists",  + "Steals", "Blocks", "Turnovers", "Personal Fouls", "Points", "Season", "Most Valuable Player"),  + Description = c("A unique identifier for each row/player.",  + "The name of the player.",  + "The playing position of the player.",  + "The age of the player.",  + "The NBA team the player played for in that season.",  + "The number of games the player appeared in.",  + "The number of games the player started.",  + "The total minutes the player played.",  + "The total number of field goals made.",  + "The total number of field goal attempts.",  + "The percentage of field goal attempts made.",  + "The total number of 3-point field goals made.",  + "The total number of 3-point field goal attempts.",  + "The percentage of 3-point field goal attempts made.",  + "The total number of 2-point field goals made.",  + "The total number of 2-point field goal attempts.",  + "The percentage of 2-point field goal attempts made.",  + "Adjusted field goal percentage accounting for 3-pointers.",  + "The total number of free throws made.",  + "The total number of free throw attempts.",  + "The percentage of free throw attempts made.",  + "The total number of offensive rebounds.",  + "The total number of defensive rebounds.",  + "The total number of rebounds.",  + "The total number of assists.",  + "The total number of steals.",  + "The total number of blocks.",  + "The total number of turnovers.",  + "The total number of personal fouls.",  + "The total number of points scored.",  + "The NBA season.",  + "Boolean indicating if the player was MVP that season.")  + )  > # variable descriptions  > print(variable\_descriptions)  Initial FullName Description  1 Rk Rank A unique identifier for each row/player.  2 Player Player The name of the player.  3 Pos Position The playing position of the player.  4 Age Age The age of the player.  5 Tm Team The NBA team the player played for in that season.  6 G Games The number of games the player appeared in.  7 GS Games Started The number of games the player started.  8 MP Minutes Played The total minutes the player played.  9 FG Field Goals The total number of field goals made.  10 FGA Field Goal Attempts The total number of field goal attempts.  11 FG% Field Goal Percentage The percentage of field goal attempts made.  12 3P 3-Point Field Goals The total number of 3-point field goals made.  13 3PA 3-Point Field Goal Attempts The total number of 3-point field goal attempts.  14 3P% 3-Point Field Goal Percentage The percentage of 3-point field goal attempts made.  15 2P 2-Point Field Goals The total number of 2-point field goals made.  16 2PA 2-Point Field Goal Attempts The total number of 2-point field goal attempts.  17 2P% 2-Point Field Goal Percentage The percentage of 2-point field goal attempts made.  18 eFG% Effective Field Goal Percentage Adjusted field goal percentage accounting for 3-pointers.  19 FT Free Throws The total number of free throws made.  20 FTA Free Throw Attempts The total number of free throw attempts.  21 FT% Free Throw Percentage The percentage of free throw attempts made.  22 ORB Offensive Rebounds The total number of offensive rebounds.  23 DRB Defensive Rebounds The total number of defensive rebounds.  24 TRB Total Rebounds The total number of rebounds.  25 AST Assists The total number of assists.  26 STL Steals The total number of steals.  27 BLK Blocks The total number of blocks.  28 TOV Turnovers The total number of turnovers.  29 PF Personal Fouls The total number of personal fouls.  30 PTS Points The total number of points scored.  31 Season Season The NBA season.  32 MVP Most Valuable Player Boolean indicating if the player was MVP that season.  > # missing values  > missing\_values <- sapply(nba\_stats, function(x) sum(is.na(x)))  > print(missing\_values)  Rk Player Pos Age Tm G GS MP FG FGA FG% 3P 3PA 3P%  0 0 0 0 0 0 0 0 0 0 88 0 0 2198  2P 2PA 2P% eFG% FT FTA FT% ORB DRB TRB AST STL BLK TOV  0 0 154 88 0 0 749 0 0 0 0 0 0 0  PF PTS Season MVP  0 0 0 0  > #Removal of missing values  > nba\_stats\_clean <- na.omit(nba\_stats)  > # Adjust column selection as per my dataset structure  > numerical\_columns <- nba\_stats\_clean %>% select(-c(Rk, Player, Pos, Tm, Season, MVP))  > # Normalize the numerical data  > nba\_stats\_normalized <- scale(numerical\_columns)  > # Checking the structure of the normalized data  > str(nba\_stats\_normalized)  num [1:11962, 1:26] 0.308 -0.872 -1.344 -0.636 -0.636 ...  - attr(\*, "dimnames")=List of 2  ..$ : NULL  ..$ : chr [1:26] "Age" "G" "GS" "MP" ...  - attr(\*, "scaled:center")= Named num [1:26] 26.7 50.04 24.37 21.41 3.27 ...  ..- attr(\*, "names")= chr [1:26] "Age" "G" "GS" "MP" ...  - attr(\*, "scaled:scale")= Named num [1:26] 4.24 24.23 27.83 9.42 2.12 ...  ..- attr(\*, "names")= chr [1:26] "Age" "G" "GS" "MP" ...  > # Perform PCA  > pca\_result <- prcomp(nba\_stats\_normalized, center = FALSE, scale. = FALSE)  > # Summary of PCA results  > summary(pca\_result)  Importance of components:  PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10  Standard deviation 3.5274 1.9134 1.53028 1.09102 1.00901 0.95900 0.92082 0.81912 0.77004 0.64256  Proportion of Variance 0.4786 0.1408 0.09007 0.04578 0.03916 0.03537 0.03261 0.02581 0.02281 0.01588  Cumulative Proportion 0.4786 0.6194 0.70945 0.75523 0.79439 0.82976 0.86238 0.88818 0.91099 0.92687  PC11 PC12 PC13 PC14 PC15 PC16 PC17 PC18 PC19  Standard deviation 0.62494 0.58207 0.56903 0.51160 0.45744 0.37516 0.32282 0.27061 0.16400  Proportion of Variance 0.01502 0.01303 0.01245 0.01007 0.00805 0.00541 0.00401 0.00282 0.00103  Cumulative Proportion 0.94189 0.95492 0.96737 0.97744 0.98549 0.99090 0.99491 0.99773 0.99876  PC20 PC21 PC22 PC23 PC24 PC25 PC26  Standard deviation 0.11572 0.1025 0.08736 0.01664 0.01603 0.01017 0.008028  Proportion of Variance 0.00052 0.0004 0.00029 0.00001 0.00001 0.00000 0.000000  Cumulative Proportion 0.99928 0.9997 0.99997 0.99998 0.99999 1.00000 1.000000  > # PCA Plot  > plot(pca\_result$sdev^2 / sum(pca\_result$sdev^2), xlab = "Principal Component", ylab = "Variance Explained", type = 'b')  > # from the PCA plot i had selected 3 PCA result  A line graph with a line  Description automatically generated with medium confidence  > library(cluster)  > pca\_scores <- as.data.frame(pca\_result$x[, 1:3])  > # Print the first few rows of pca\_scores to see PCA scores and cluster assignments  > print(head(pca\_scores))  PC1 PC2 PC3  1 2.14496584 -1.1442190 -1.4302328  2 1.69609787 0.3870488 -1.3554549  3 -7.97630014 1.3701878 -1.7618373  4 0.04602544 -1.5508379 0.6504907  5 2.67225276 -0.4776822 0.4209221  6 -3.27594115 -3.1065426 0.3469486  > min\_pc1 <- min(pca\_scores$PC1)  > print(min\_pc1)  [1] -12.88522  > max\_pc1 <- max(pca\_scores$PC1)  > print(max\_pc1)  [1] 8.007224  > min\_pc2 <- min(pca\_scores$PC2)  > print(min\_pc2)  [1] -6.630489  > max\_pc2 <- max(pca\_scores$PC2)  > print(max\_pc1)  [1] 8.007224  > min\_pc3 <- min(pca\_scores$PC3)  > print(min\_pc1)  [1] -12.88522  > max\_pc3 <- max(pca\_scores$PC3)  > print(max\_pc1)  [1] 8.007224  > kmeans\_result <- kmeans(pca\_scores, centers = 5) # Example: 5 clusters  > # Print the summary of kmeans clustering results  > print(kmeans\_result)  K-means clustering with 5 clusters of sizes 1408, 2866, 3926, 2392, 1370  Cluster means:  PC1 PC2 PC3  1 -3.600897 3.1517490 -0.4991260  2 -1.238311 -1.2782402 0.4504695  3 1.815920 0.5938145 0.8279672  4 4.064369 -0.3746724 -1.3132149  5 -6.008914 -1.6126465 -0.5092449  Clustering vector:  [1] 4 3 5 2 3 2 5 2 4 2 2 2 2 3 3 3 3 4 3 2 4 1 2 4 3 1 3 3 3 3 4 3 2 2 2 5 3 2 2 4 3 3 1 5 4 4 3  [48] 2 3 2 4 4 2 3 3 3 3 4 4 3 3 1 3 1 1 2 5 3 2 3 1 2 1 3 3 5 1 2 4 2 4 4 4 4 3 3 3 3 3 1 2 3 1 3  [95] 2 3 3 4 1 3 2 1 2 3 2 1 3 5 2 3 1 3 3 2 4 4 4 2 2 2 2 3 5 3 4 3 1 4 3 4 3 5 3 1 3 2 4 3 3 4 1  [142] 4 1 4 1 1 1 2 3 1 3 4 1 3 1 4 4 5 4 3 5 5 2 2 3 4 2 1 1 4 3 4 5 1 4 3 2 1 5 1 2 4 4 4 1 5 2 2  [189] 5 2 5 2 4 2 2 3 2 1 5 1 5 3 1 3 3 5 5 2 3 2 1 4 4 4 2 4 4 3 3 2 4 2 5 2 1 5 1 2 5 1 3 4 3 3 2  [236] 3 1 3 3 3 3 3 4 4 1 3 1 5 2 3 3 2 3 4 2 3 1 2 3 2 4 2 4 3 1 1 4 4 1 4 1 2 4 1 5 2 2 2 3 3 2 3  [283] 2 2 2 4 5 1 4 2 3 3 2 2 1 2 3 3 3 3 4 5 4 5 5 3 4 4 4 1 1 5 3 4 1 3 3 5 2 4 2 1 5 2 2 3 4 3 1  [330] 1 4 2 2 3 3 4 3 4 4 3 1 1 1 3 1 5 4 1 3 4 3 5 5 5 5 2 4 4 2 5 5 2 2 3 5 1 3 3 3 4 1 4 4 4 2 4  [377] 4 3 1 1 1 1 1 1 2 5 3 3 4 1 5 1 1 1 3 4 3 2 3 4 5 4 5 4 4 3 4 4 3 3 3 4 4 1 3 4 3 3 3 2 2 2 1  [424] 1 3 4 1 4 4 2 5 2 5 2 2 2 3 4 3 3 3 4 2 4 3 1 3 1 3 3 5 2 2 4 4 3 4 4 4 4 1 5 2 5 3 2 4 2 4 5  [471] 1 3 4 3 4 3 1 1 1 4 4 4 4 4 1 2 5 2 3 2 2 3 4 2 5 3 3 1 2 3 4 3 3 3 3 3 2 4 2 3 3 3 2 4 4 4 3  [518] 1 4 2 1 2 4 2 1 3 2 2 2 2 1 4 4 1 3 5 2 1 4 3 3 1 4 3 3 4 3 1 1 3 4 1 4 3 1 5 5 2 2 3 4 3 2 4  [565] 1 2 4 3 4 4 5 4 4 3 4 2 3 5 1 3 3 2 5 3 3 2 2 4 3 1 4 2 2 4 4 4 4 5 2 5 4 3 3 1 4 4 5 2 4 2 4  [612] 3 5 3 1 4 3 3 4 1 2 4 1 2 2 3 2 2 5 4 1 5 5 5 1 2 5 1 2 4 3 2 1 1 4 3 3 4 5 4 4 3 2 2 2 3 2 2  [659] 3 3 1 2 2 2 4 2 2 3 4 3 4 4 1 1 1 1 4 3 3 3 2 3 4 4 5 2 3 4 3 4 3 2 2 2 5 5 1 4 2 1 1 3 3 1 4  [706] 1 3 4 4 4 2 4 5 2 5 1 5 4 3 4 1 3 4 2 3 2 1 5 3 4 3 3 4 1 4 4 1 1 5 1 2 5 2 2 3 3 2 2 2 3 3 5  [753] 3 4 3 1 3 1 3 4 3 1 3 1 5 5 4 4 5 3 1 4 2 3 1 3 4 5 3 4 3 4 3 3 4 2 1 2 1 1 2 1 3 1 1 2 5 4 5  [800] 4 2 5 2 2 2 3 3 3 5 2 3 3 4 1 3 4 2 5 2 2 3 3 4 4 4 4 3 4 3 3 1 1 5 3 3 3 3 4 1 2 4 5 3 4 4 3  [847] 3 1 1 1 2 4 2 2 5 5 1 1 4 3 3 4 4 3 3 2 1 3 1 2 4 4 2 3 2 4 4 3 3 3 3 3 3 4 3 3 3 3 3 3 4 5 1  [894] 3 3 3 1 2 2 4 3 3 4 3 1 3 3 3 5 2 4 3 3 5 3 5 3 1 1 2 1 4 2 3 4 3 1 3 2 1 3 3 5 2 2 4 3 1 3 2  [941] 4 1 4 4 5 1 2 2 3 5 1 2 5 2 5 3 2 5 3 3 4 5 2 3 3 1 4 4 4 2 1 3 2 4 4 3 4 5 4 4 4 4 1 4 3 5 2  [988] 2 5 5 2 1 1 4 2 4 2 4 3 3  [ reached getOption("max.print") -- omitted 10962 entries ]  Within cluster sum of squares by cluster:  [1] 11835.67 11516.10 19645.97 10810.32 10187.31  (between\_SS / total\_SS = 71.0 %)  Available components:  [1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss"  [7] "size" "iter" "ifault"  > #summary function for a concise overview  > summary(kmeans\_result)  Length Class Mode  cluster 11962 -none- numeric  centers 15 -none- numeric  totss 1 -none- numeric  withinss 5 -none- numeric  tot.withinss 1 -none- numeric  betweenss 1 -none- numeric  size 5 -none- numeric  iter 1 -none- numeric  ifault 1 -none- numeric  > # Adding cluster assignments to PCA scores  > pca\_scores$cluster <- as.factor(kmeans\_result$cluster)  > library(ggplot2)  > # Plot with clusters  > ggplot(pca\_scores, aes(x = PC1, y = PC2, color = cluster)) +  + geom\_point() +  + ggtitle("NBA Player Clusters Based on PCA")  > |
|  |
| |  | | --- | | > | |

**A diagram of a number of dots

Description automatically generated with medium confidence**

**OBSERVATION:**

**Cluster Distribution**

**- Clusters:** The players are divided into five distinct clusters, each represented by a different color. The clusters are spread out over the PCA space, suggesting variability in the data that PCA has captured and that K-means has used to group the players.

- **Cluster 1 (Red):** This cluster appears to be centered around the middle of PC1 and the upper half of PC2, suggesting these players share a set of characteristics that are average in PC1's aspect and higher in PC2's aspect compared to the other players.

**- Cluster 2 (Green**): Players in this cluster are centered towards the lower end of both PC1 and PC2, indicating these players share characteristics that are lower in both aspects captured by these principal components.

**- Cluster 3 (Blue):** This cluster is spread towards the left of PC1 and the middle of PC2, suggesting these players have a lower score in the aspect represented by PC1 and an average score in the aspect represented by PC2.

**- Cluster 4 (Cyan):** Players here are located in the lower half of PC1 and slightly higher in PC2. They might represent players with lower scores in PC1's aspect and slightly above-average scores in PC2's aspect.

**- Cluster 5 (Magenta):** This cluster has players with scores around the middle to lower end of PC1 and the lower end of PC2. They may share characteristics that are average to lower in the aspects represented by both PC1 and PC2.

**Interpretation of Clusters**

**- Performance Grouping**: Each cluster might represent a group of players with similar performance metrics. For example, one cluster might consist of high scorers, another of defensive specialists, and another of all-around players. The exact interpretation would require a closer look at the loadings of PC1 and PC2 to understand which performance metrics they most strongly represent.

**Conclusion for NBA Player Clusters**

Based on PCA and K-means clustering, I have identified five distinct groups of NBA players, each with unique characteristics:

**Conclusion on NBA Player Clusters Based on PCA**

The PCA and K-means clustering analysis has identified five distinct groups of NBA players. These groups are characterized by their underlying performance metrics, which are captured by their loadings on the principal components. Each cluster can be tentatively described as follows:

**1. Cluster 1 (Versatile Contributors):** Players in this group may be all-around contributors who perform well across various aspects of the game without necessarily being superstars. They likely have a good balance between offensive and defensive skills.

**2. Cluster 2 (Specialized Role Players):** This cluster might consist of players who excel in specific roles. They may be three-point shooters, defensive specialists, or players who contribute significantly to certain aspects of the game, such as rebounding or assists.

**3. Cluster 3 (Potential Stars):** Athletes in this group could be emerging stars or players who have the potential to significantly impact the game in multiple areas. They might be characterized by higher athleticism or the ability to make plays that change the course of a game.

**4. Cluster 4 (Veteran Presence):** Players in this cluster may be veterans who provide experience and stability to their teams. Their performance might not be as high in terms of raw statistics, but they offer intangibles that don't always show up in traditional metrics.

**5. Cluster 5 (Developing Talents):** This group likely includes younger or less experienced players who are still developing their skills. They might have lower overall statistics but possess the potential for growth.