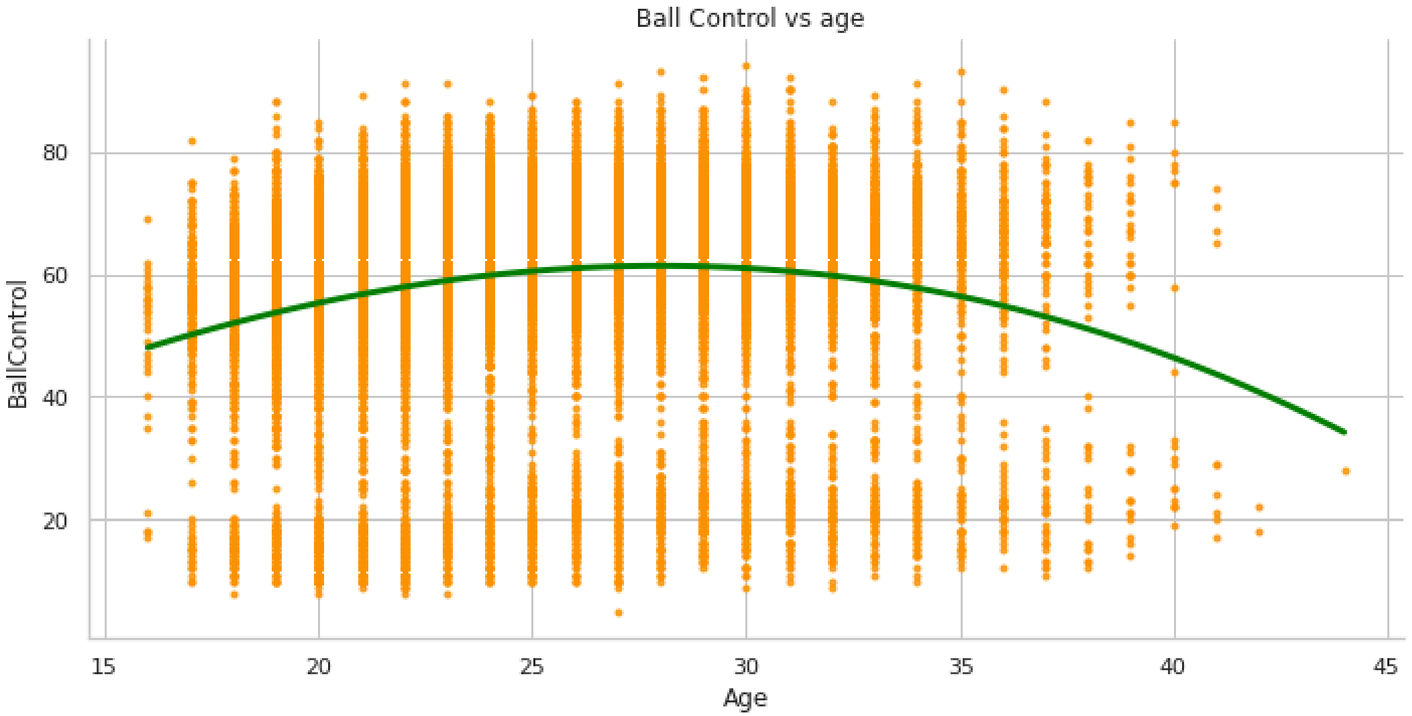
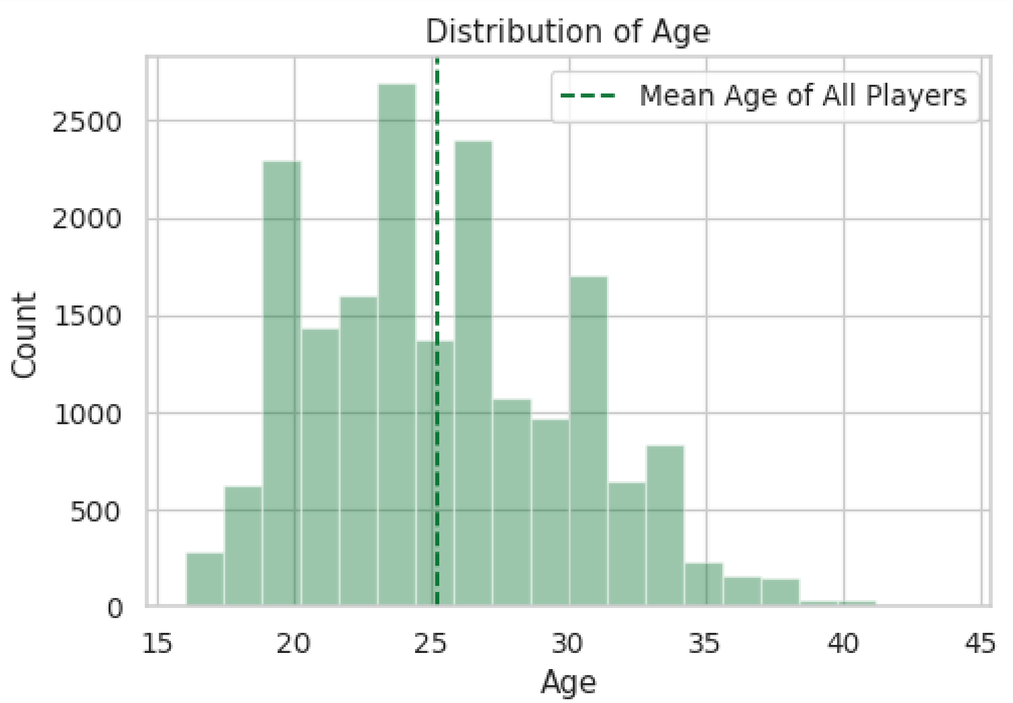
Does the Age of the Player Affect on his Ball Control Performance?

Figure :

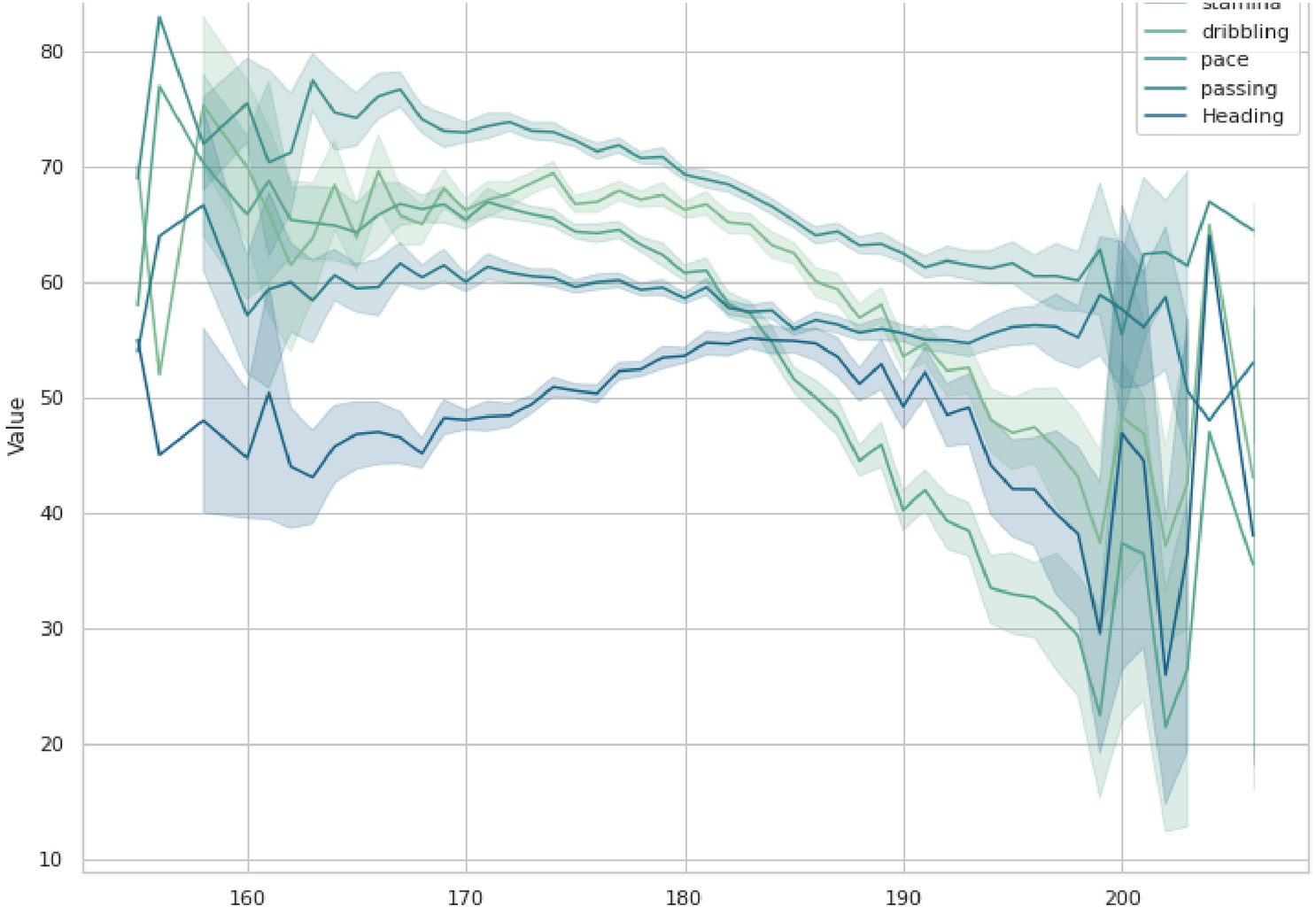




- So We can deduce that the age has an effect on the Player's Ball Control

- While the Age is increasing, the Ball Control decreases.

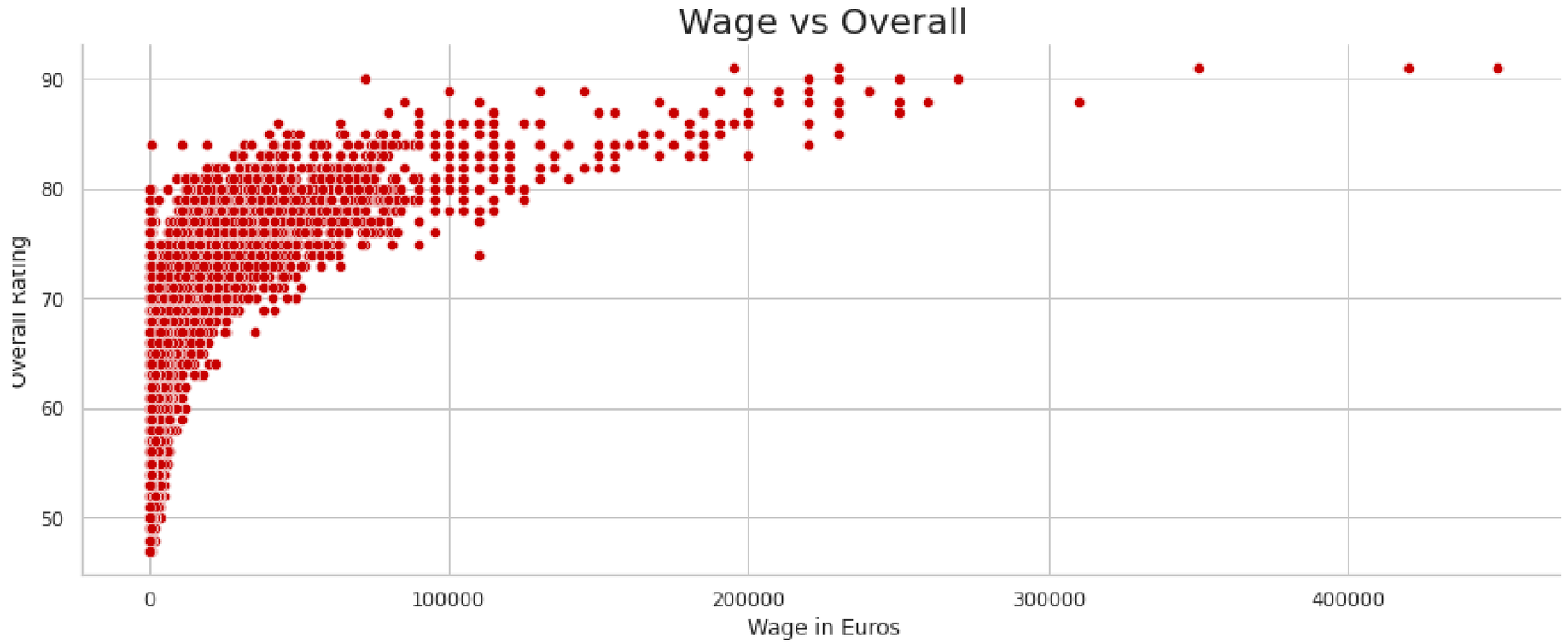
How Height affects different factors like stamina, dribbling, pace, passing and HeadingAccuracy ?



- As height increases, features like stamina, dribbling, pace, passing decreases.

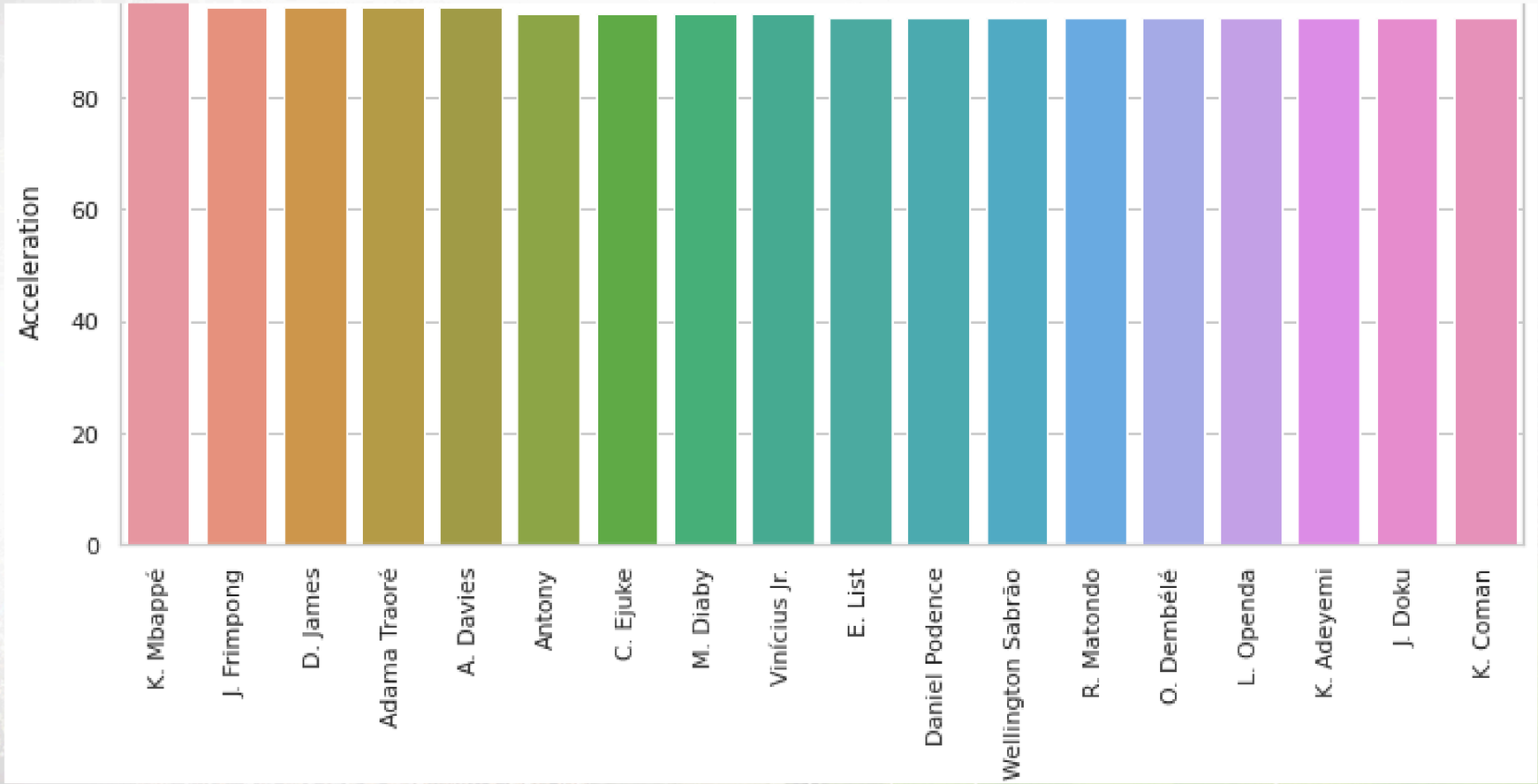
- As height increases, features like Heading increase.

Show if there is a relation between Wage and Overall of the Players

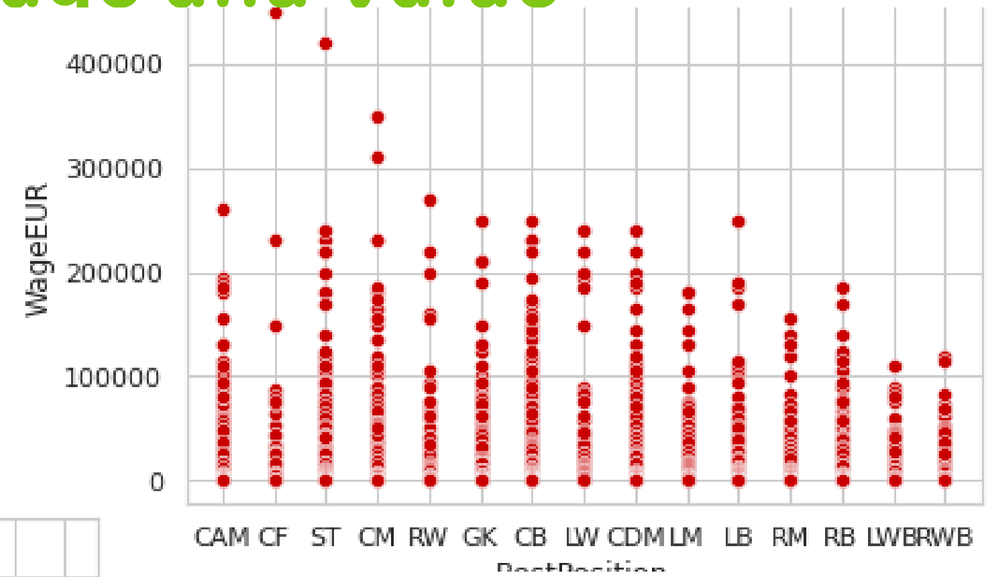


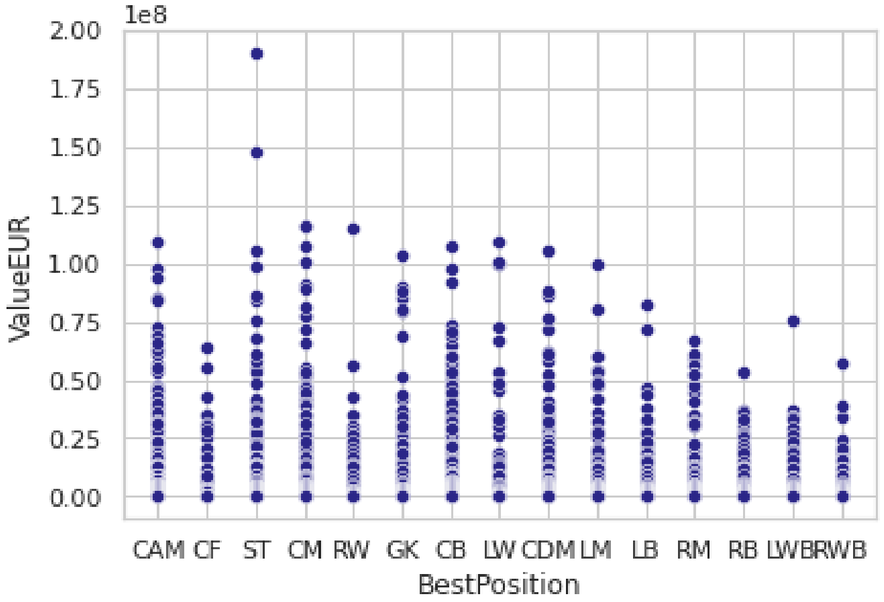
As the Overall Rating Increase, the Wage of the Player Increases too.

Show the top Fastest Players



Determine if their is a relation between the Position of the Player and his Wage and Value

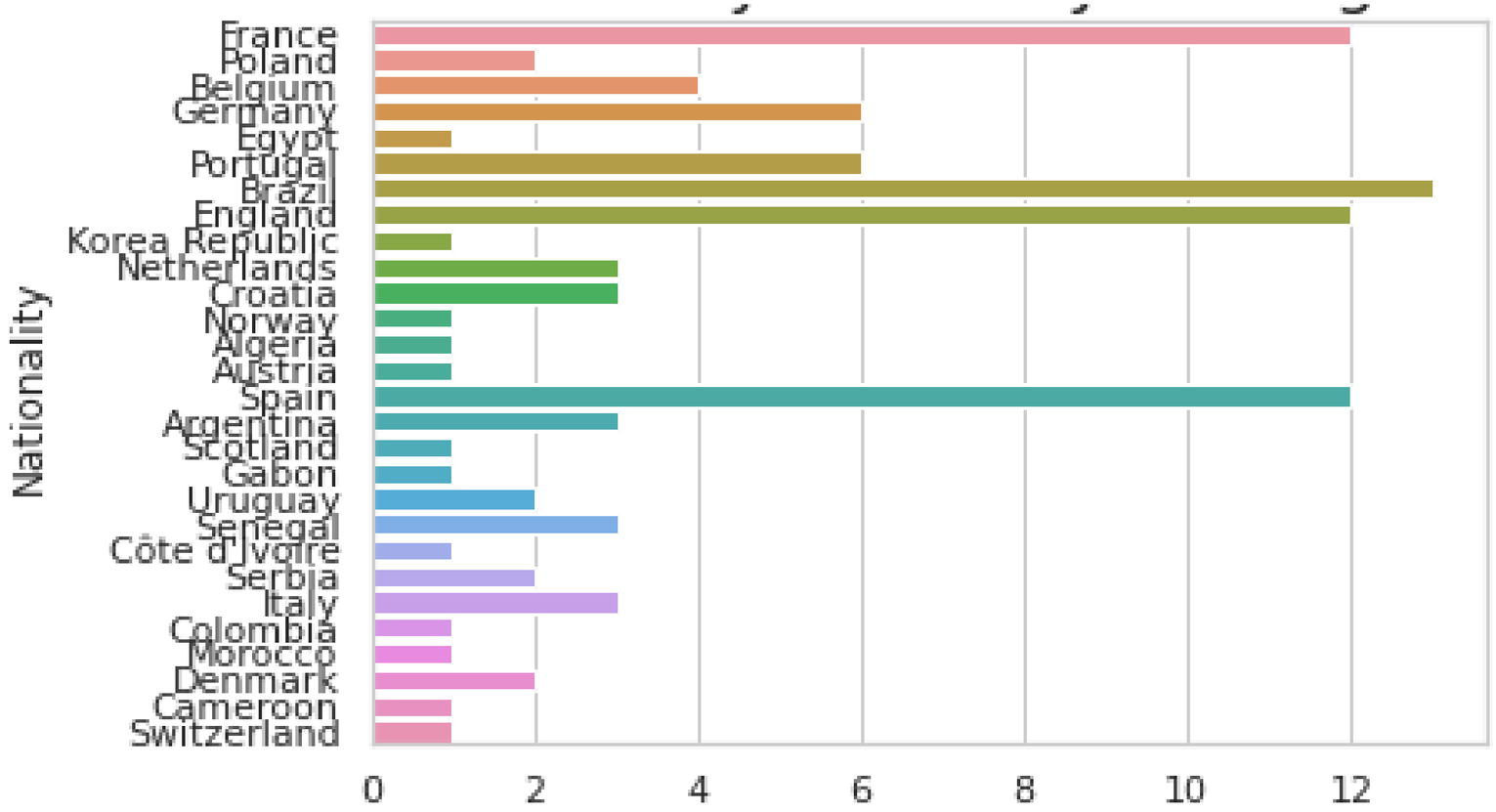




- So we can see that the Players in Positions LM, RM, RB, LWB, RWB got the lowest Wages.

- And the Players With Positions LB, RB, LWB, RWB, CF, RW have the lowest Values.

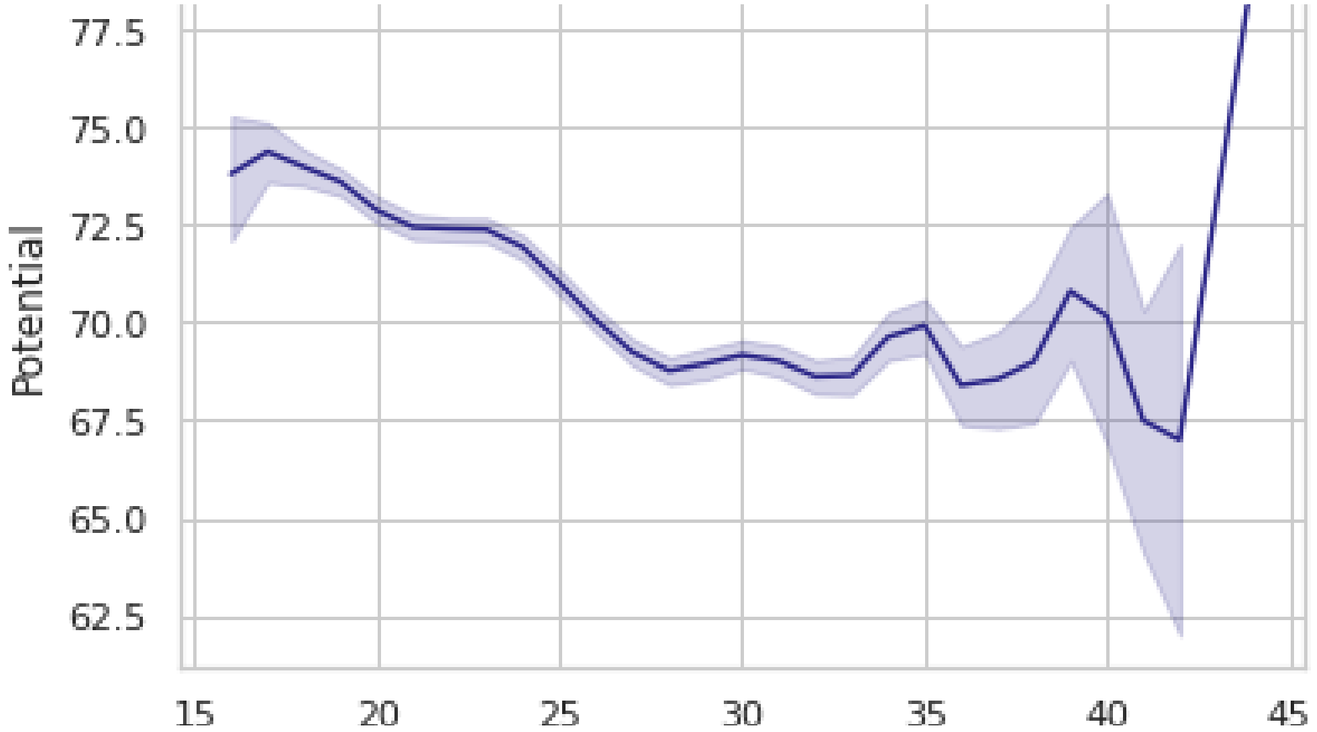
See the Nationality of the Players that got the highest Wages



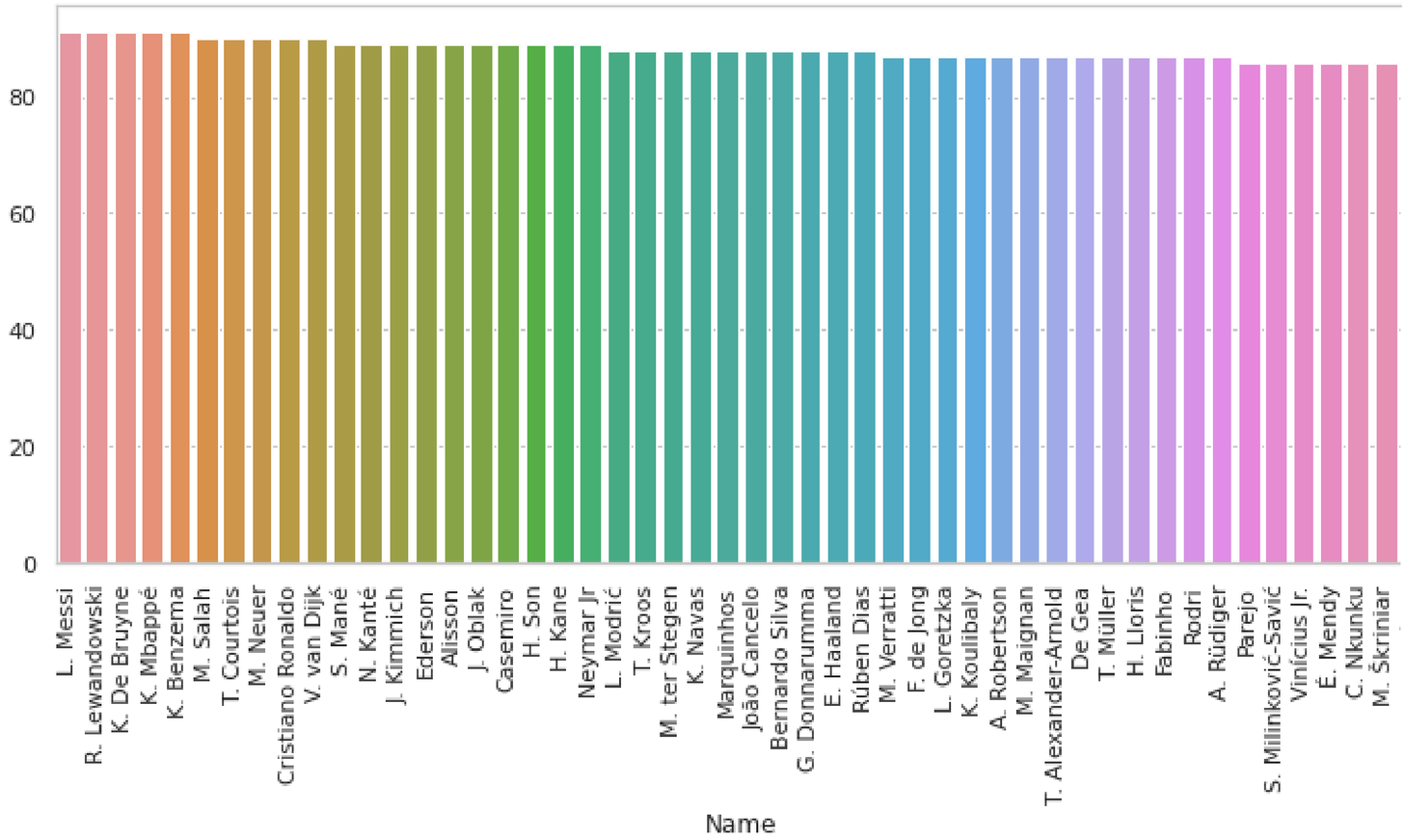
So we can deduce that the Players that got the Maximum Wage are from Brazil, France, England and Spain.

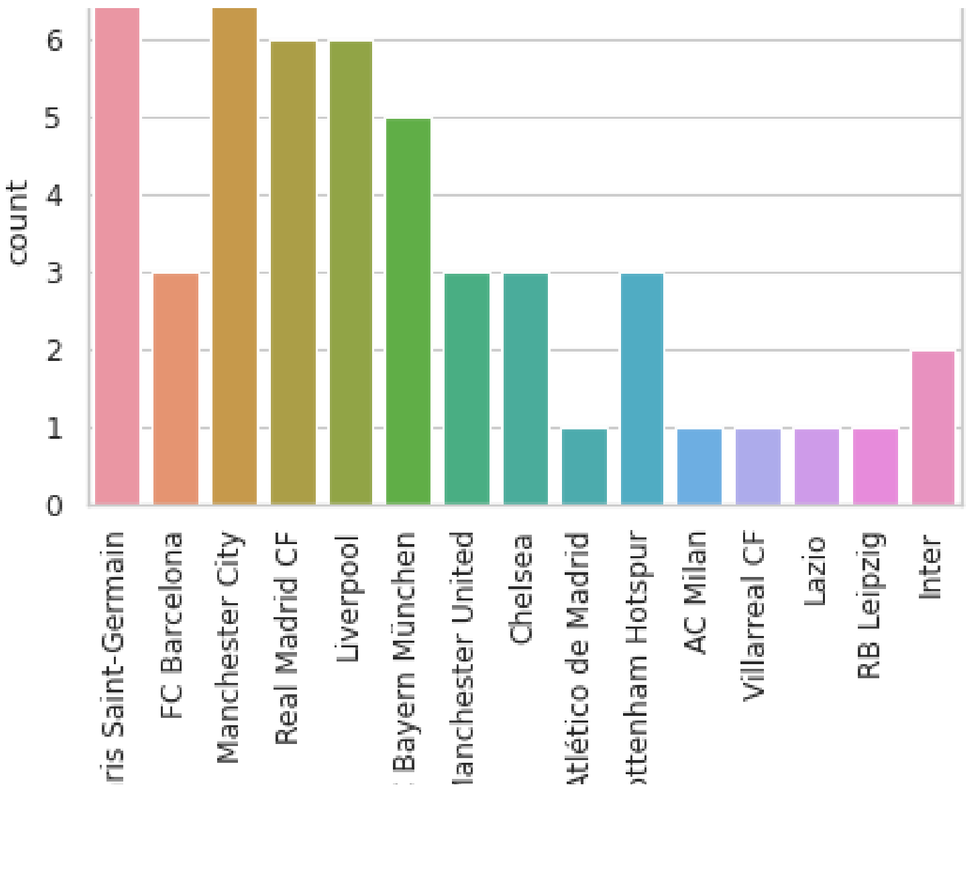
Show the effect of the Age on the Potential of the Players

While the Age Increases the Potential of the Player Decreases.



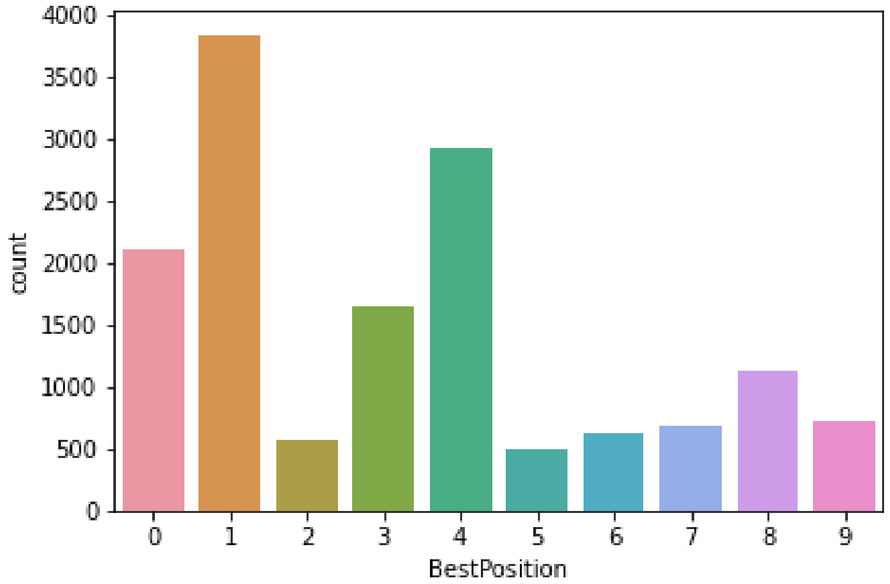
View the Top 50 Players and their Clubs





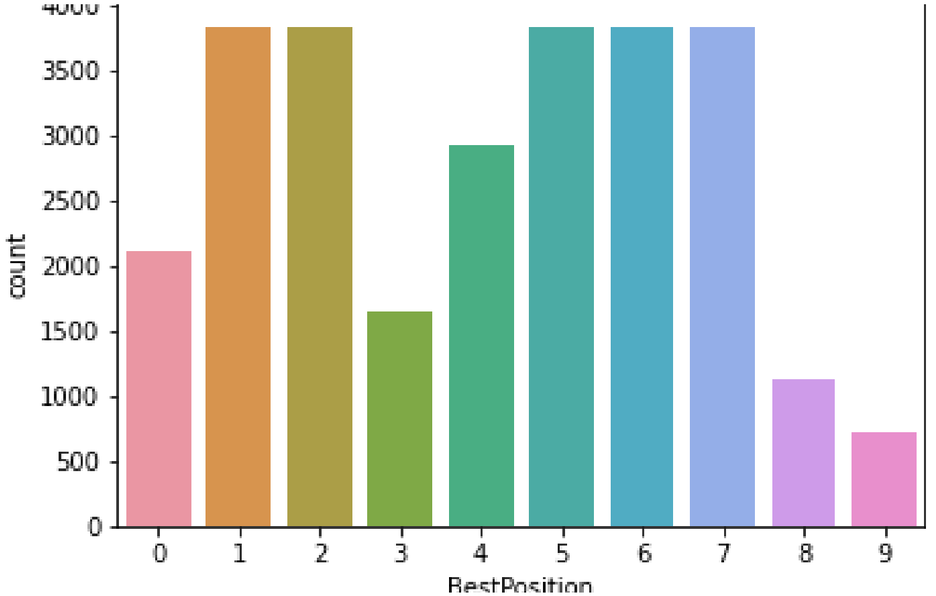
- Paris Saint-Germainand Machester City have the maximum top Players numbers
- Liverpool and Real Madrid have the second Maximum top Players numbers.

Handle the Imbalanced Data



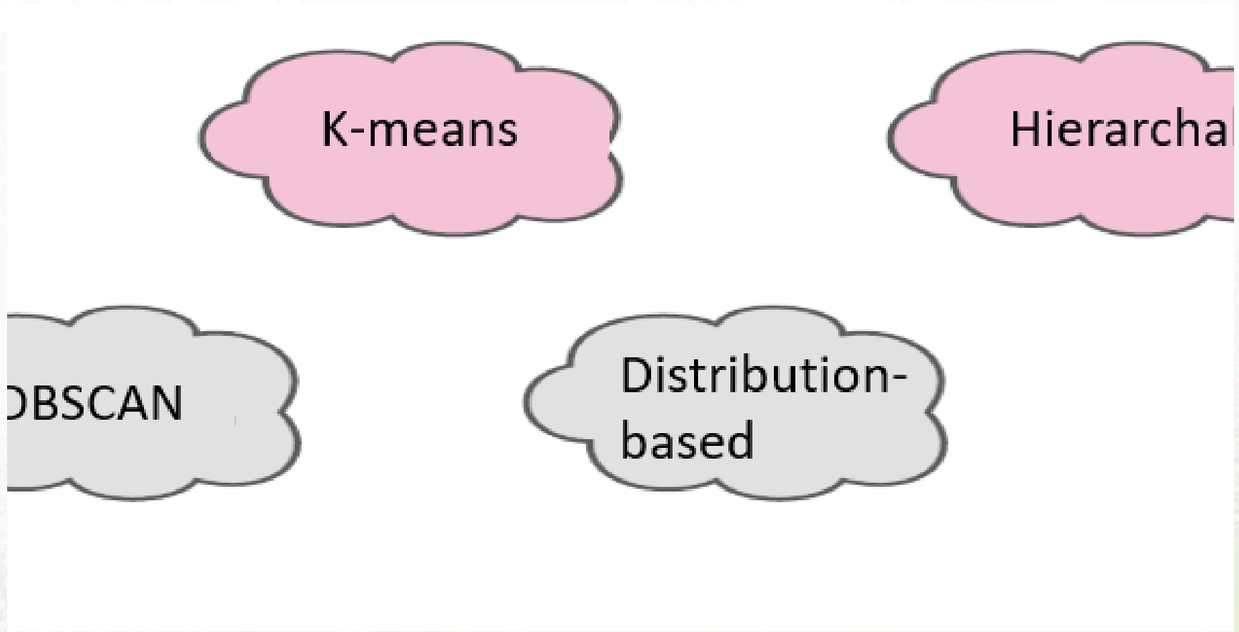
As We can see Here the Data is Imbalanced so we need to fix this issue.

Used the Over Sampling method to Balance the classes 2, 5, 6, 7 So the model would
not be biased.

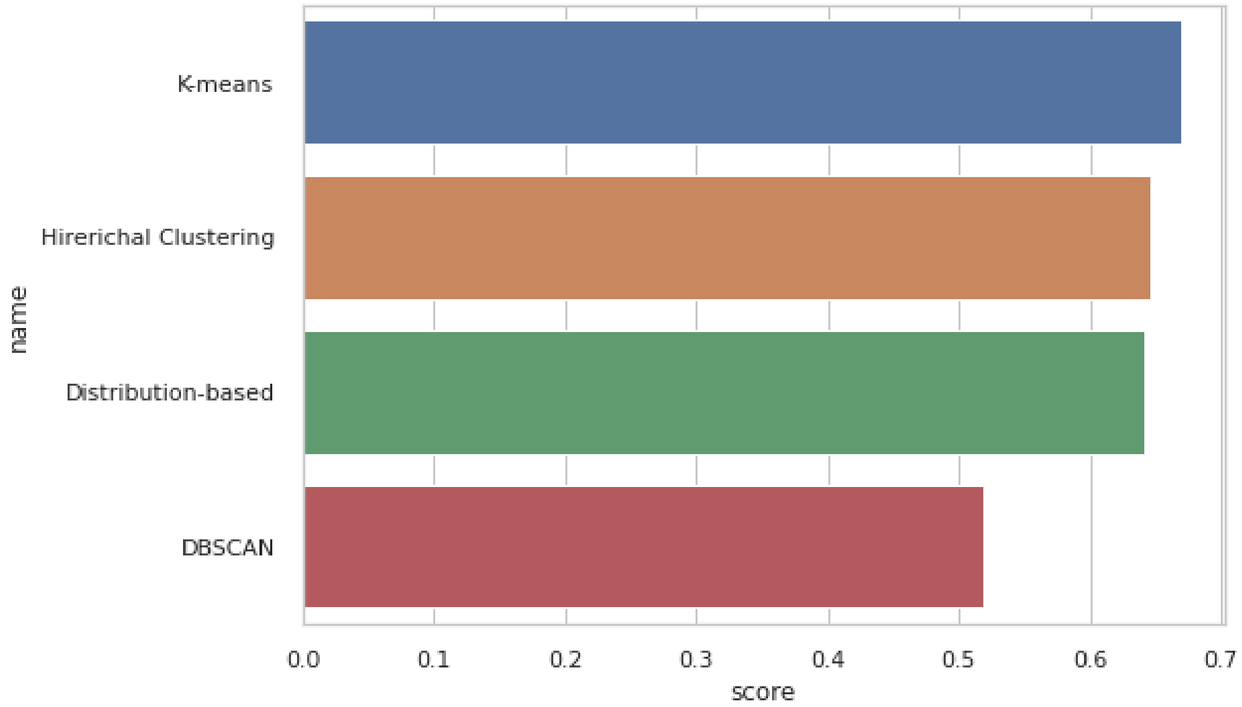


B. Group the Players in Clusters (with Overall > 86)

Models Used:



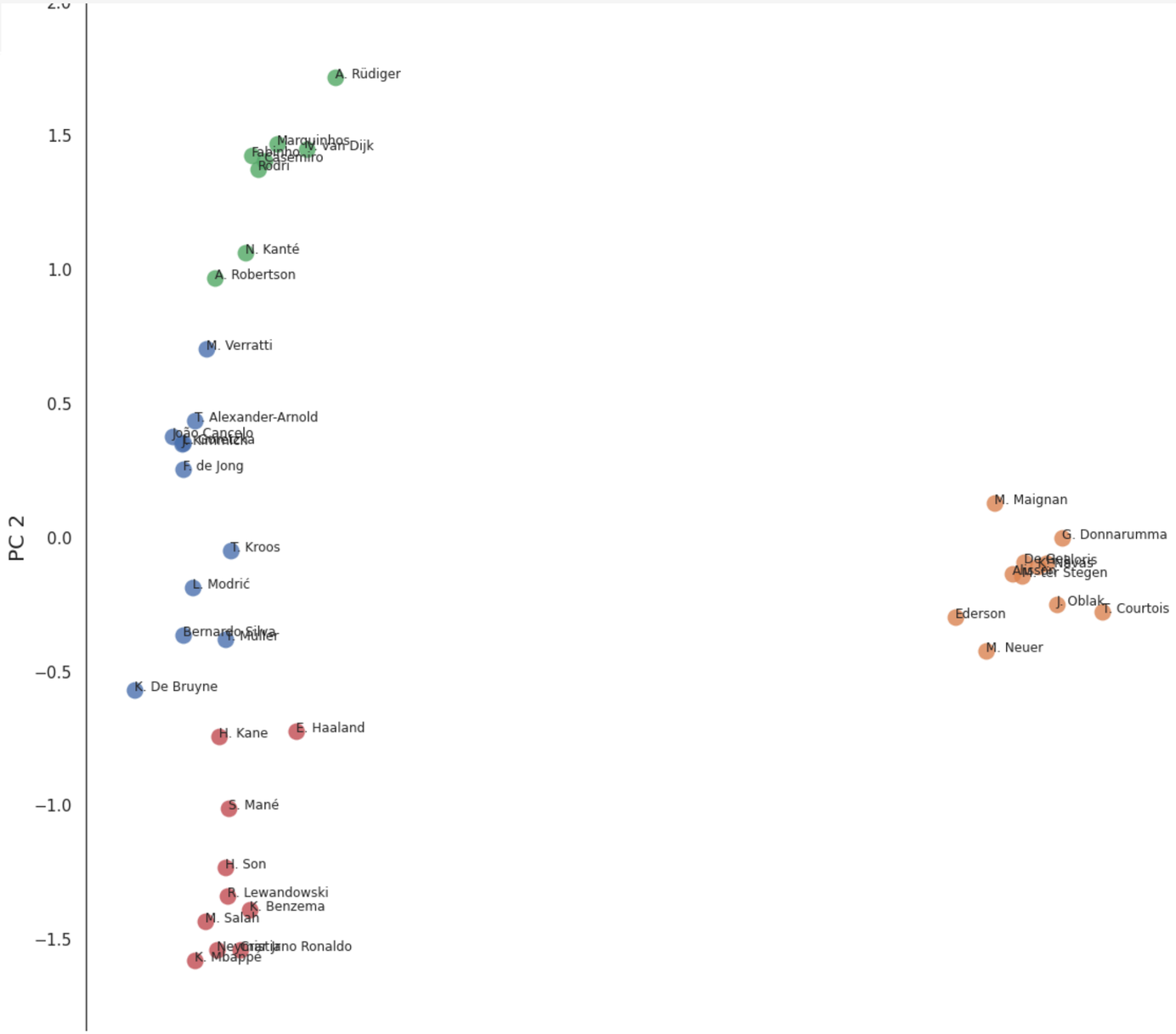
Comparing the 4 Algorithms based on the Silhouette Score



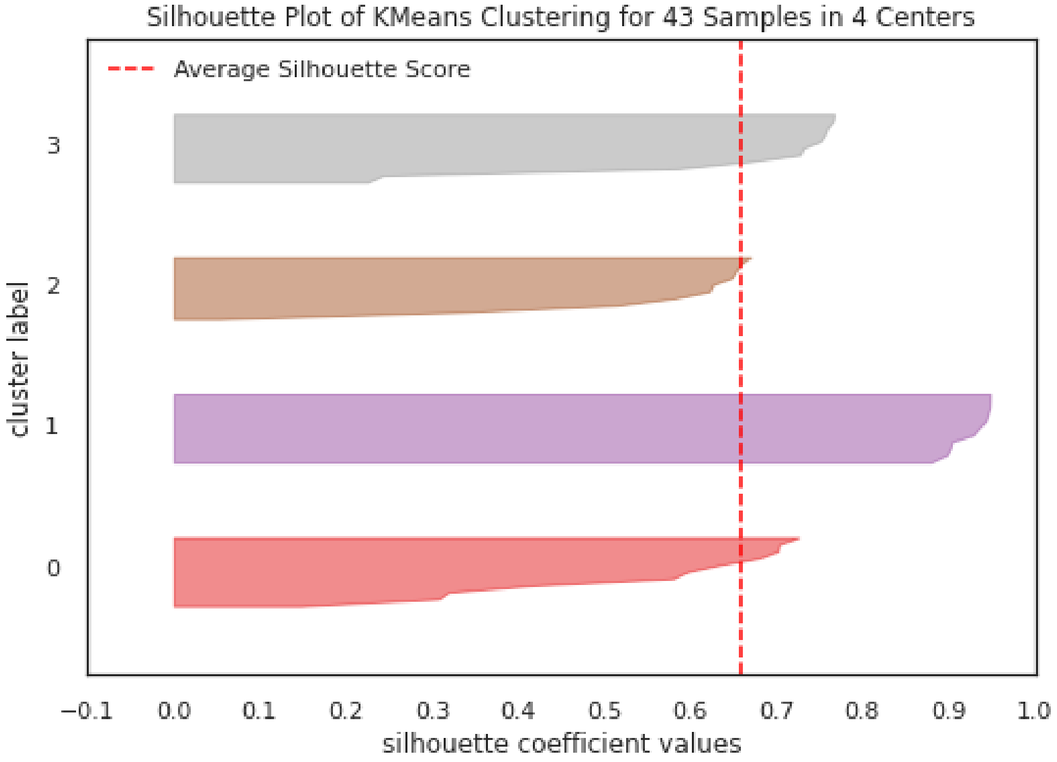
name	scor
K-means	0.66946
Hirerichal Clustering	0.64675
DBSCAN	0.51851
Distribution-based	0.64106

So We Can Say that the Hierarchal Clustering and the K-means Algorithms are the best 2 Algorithms for that problem.

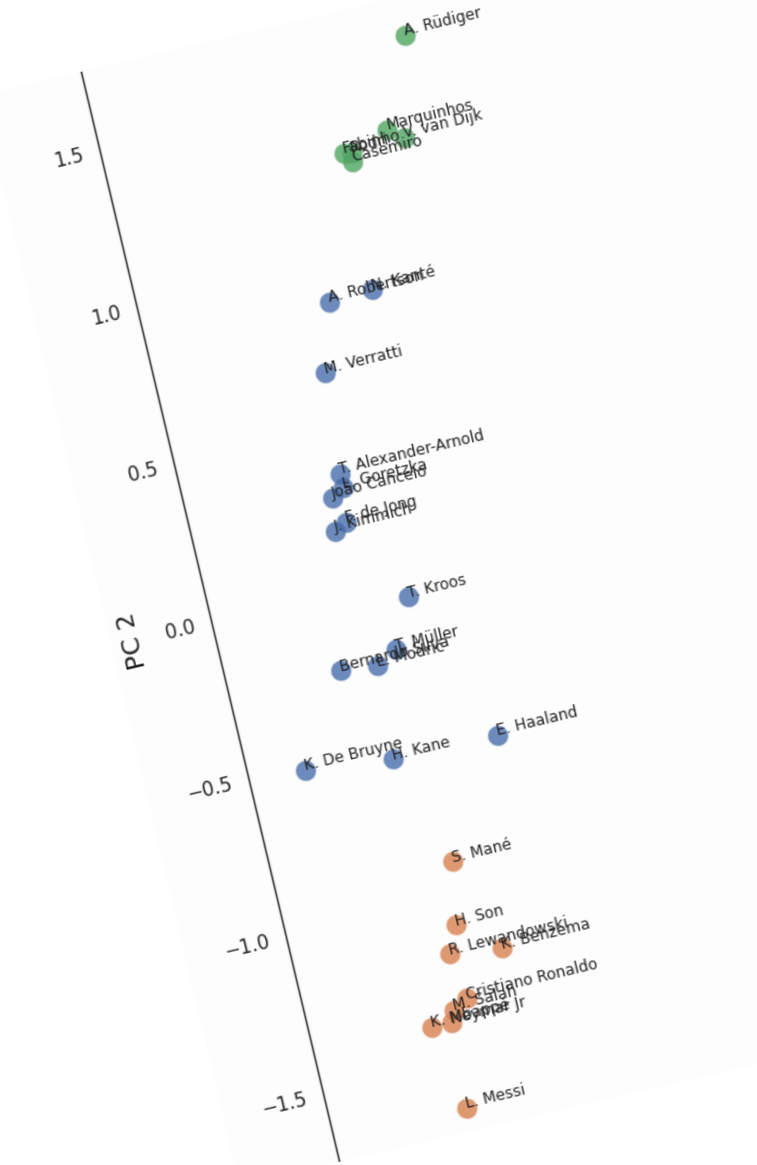
K-means

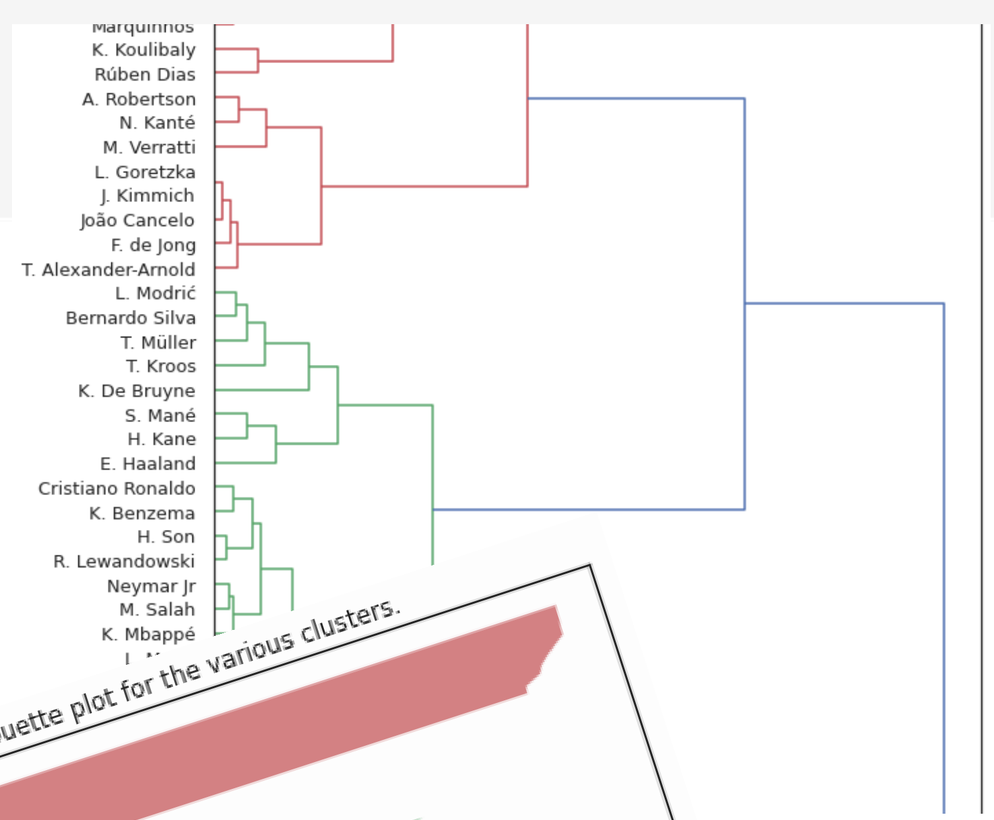

A. Rüdiger
Marquinhos
V. van Dijk
Fabinho
Casemiro
Rodri
N. Kanté
A. Robertson
M. Verratti
T. Alexander-Arnold
João Cancelo
F. de Jong
T. Kroos
L. Modrić
Bernardo Silva
T. Müller
K. De Bruyne
H. Kane
E. Haaland
S. Mané
H. Son
R. Lewandowski
K. Benzema
M. Salah
Cristiano Ronaldo
K. Mbappé
PC 2
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

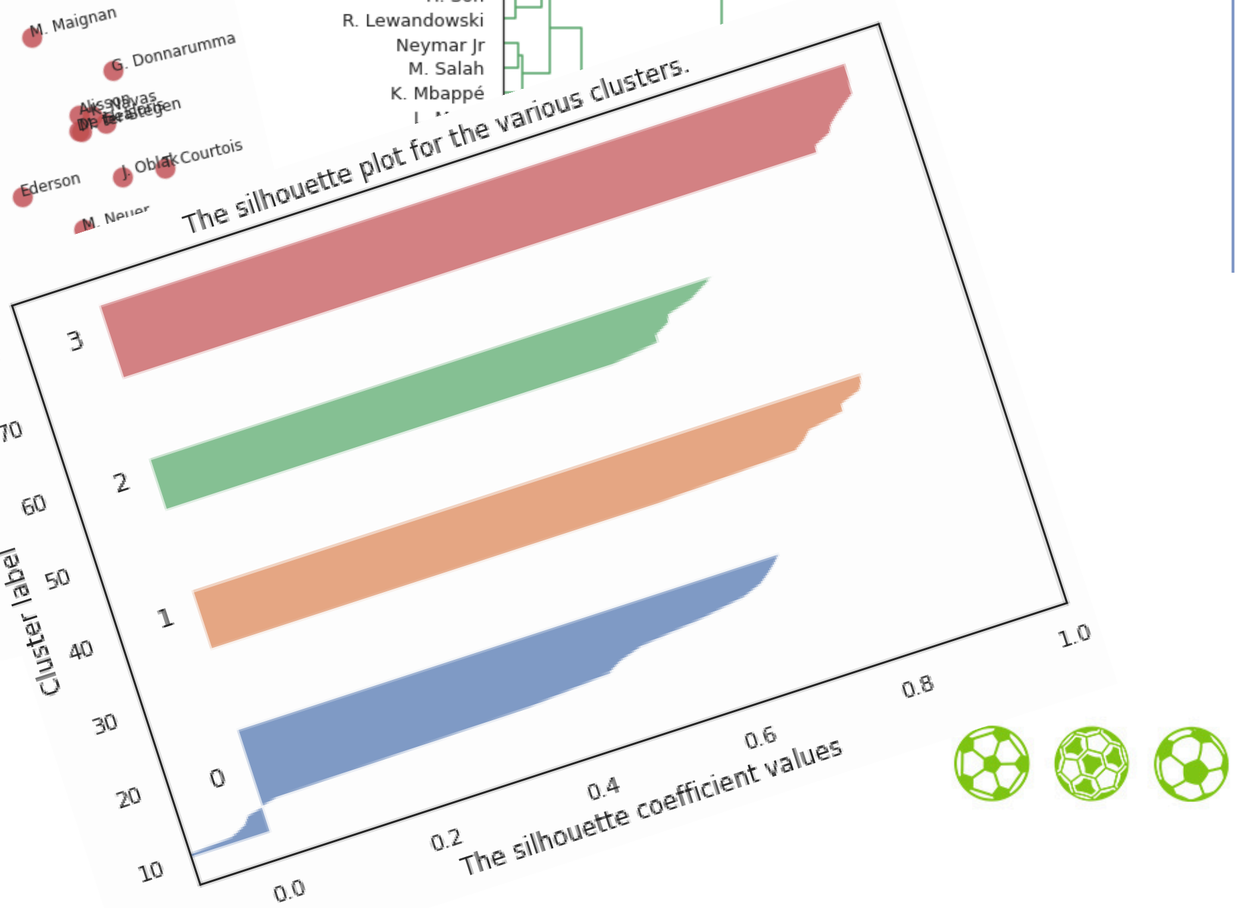
M. Maignan
G. Donnarumma
M. ter Stegen
J. Oblak
T. Courtois
Ederson
M. Neuer


Silhouette Plot of KMeans Clustering for 43 Samples in 4 Centers
Average Silhouette Score
cluster label
3
2
1
0
−0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
silhouette coefficient values

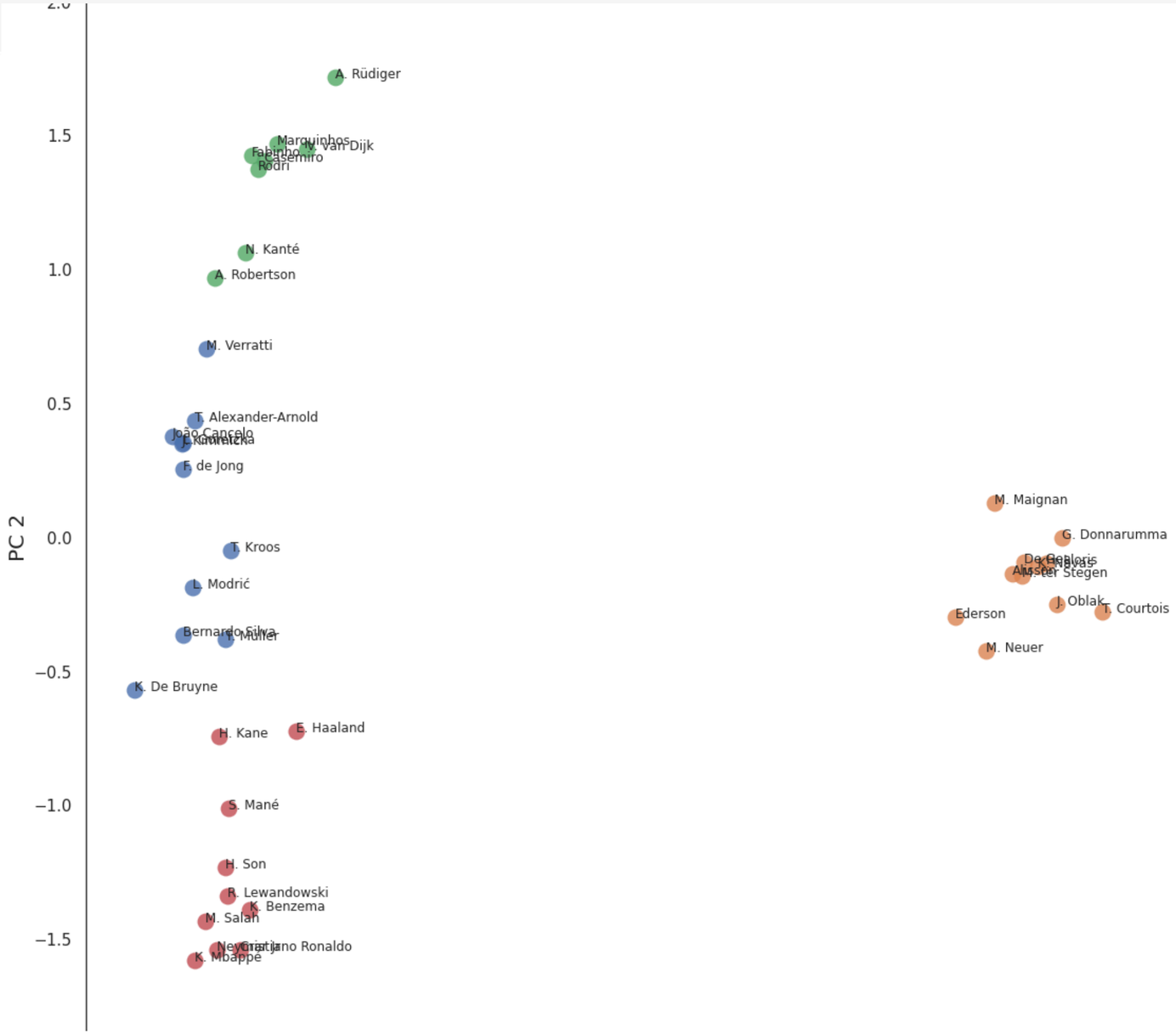
Hierarchal Clustering



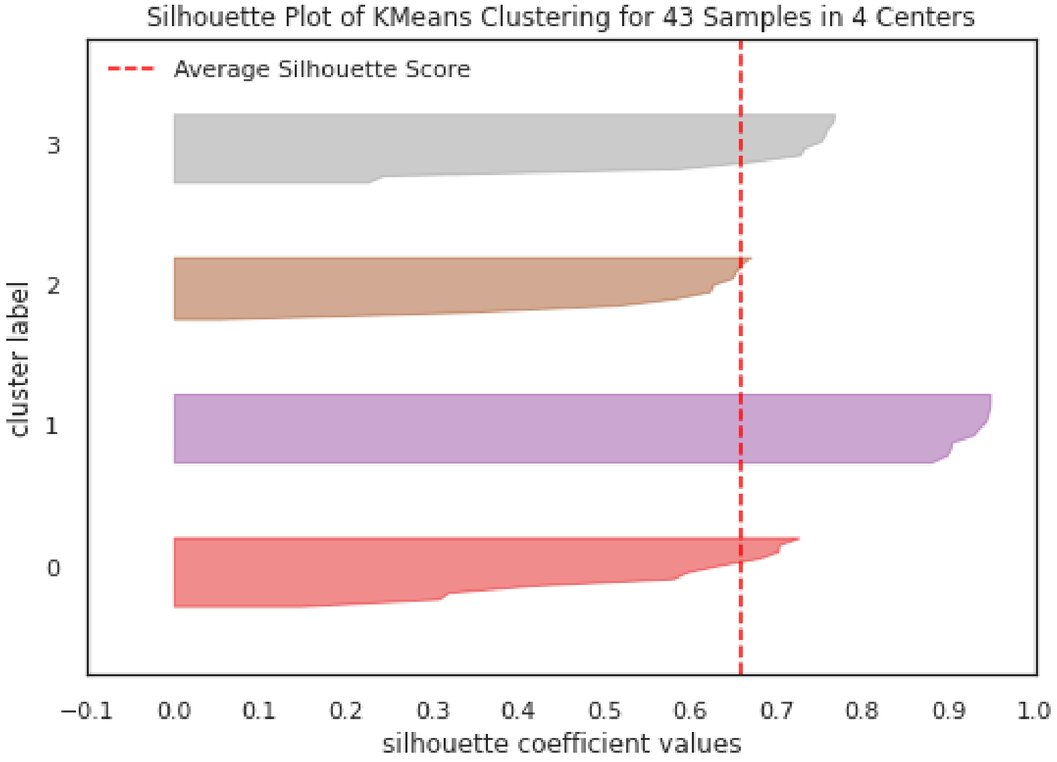




K-means


A. Rüdiger
Marquinhos
V. van Dijk
Fabinho
Casemiro
Rodri
N. Kanté
A. Robertson
M. Verratti
T. Alexander-Arnold
João Cancelo
F. de Jong
T. Kroos
L. Modrić
Bernardo Silva
T. Müller
K. De Bruyne
H. Kane
E. Haaland
S. Mané
H. Son
R. Lewandowski
K. Benzema
M. Salah
Cristiano Ronaldo
K. Mbappé
PC 2
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

M. Maignan
G. Donnarumma
M. ter Stegen
J. Oblak
T. Courtois
Ederson
M. Neuer


Silhouette Plot of KMeans Clustering for 43 Samples in 4 Centers
Average Silhouette Score
cluster label
3
2
1
0
−0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
silhouette coefficient values