REAL-WORLD ANALYTICS PROJECT (RWAP)

FOOTBALL PLAYER PERFORMANCE ANALYSIS

GROUP 3

Ishaan Agarwal(045021)
Parth Narula (045039)
Rahul Gohri (045044)
Sahibe aalam(045002)

DATASET

Football Player Performance Metrics- The dataset contains detailed information about football players, including their positions, ratings, and various attributes.

- Number of Entries: 25,000
- Number of Columns: 110
- Data Sections: Numerical Attributes (48 columns), Categorical Attributes (17 columns)

KEY METRICS

Player Positions

Categorizes player roles on the field.

Overall & Potential Ratings

Current and future performance levels

Age

Player Age

Attributes

Key skills like pace, shooting, passing, dribbling, defending, and physicality.

FIFA Version

The specific game edition the data is from.

OBJECTIVE OF ANALYTICAL DASHBOARD

- Develop a comprehensive analytical dashboard to analyze football player performance metrics
- Utilize Descriptive Statistics to summarize the dataset
- Applied Inferential Statistics to draw conclusions
- Implement Regression Analysis to predict player value or wage based on performance metrics and physical attributes.
- Enhance data-driven decision-making by providing actionable insights through statistical analysis and predictive modeling.

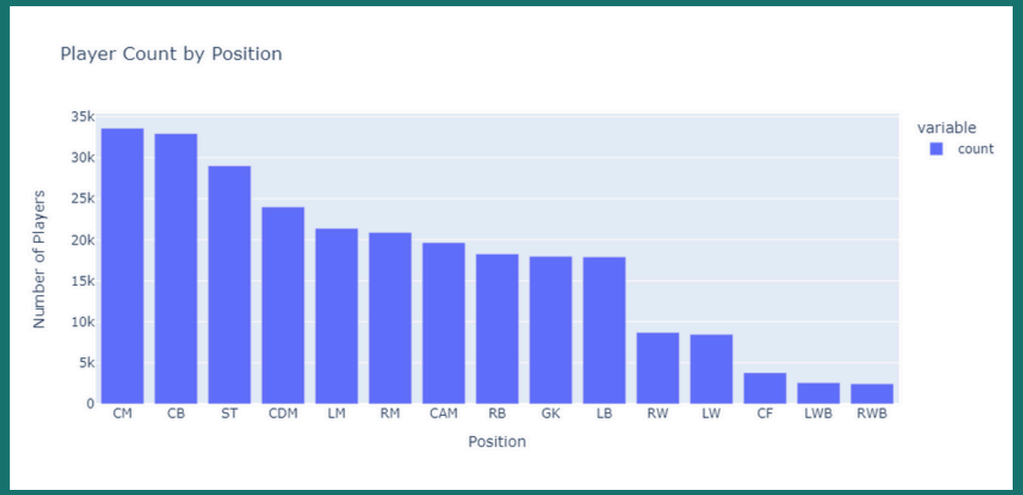
OBJECTIVE OF ML MODEL

Primary Objectives:

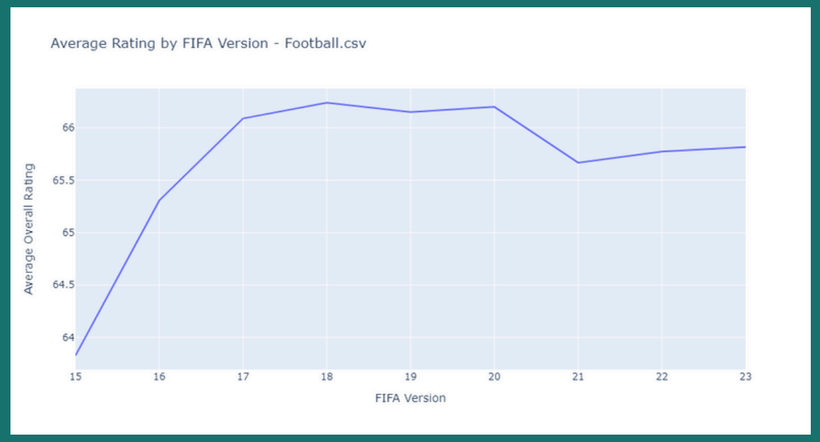
- Identify distinct player archetypes: To uncover inherent groupings within the dataset based on player attributes, leading to the creation of player clusters or segments.
- Analyze player characteristics: To understand the key attributes that differentiate player clusters and identify potential correlations between player traits and performance.
- Evaluate clustering algorithms: To compare the effectiveness of K-Means and DBSCAN in segmenting players based on the given dataset and identify the most suitable algorithm for the task.

Secondary Objectives:

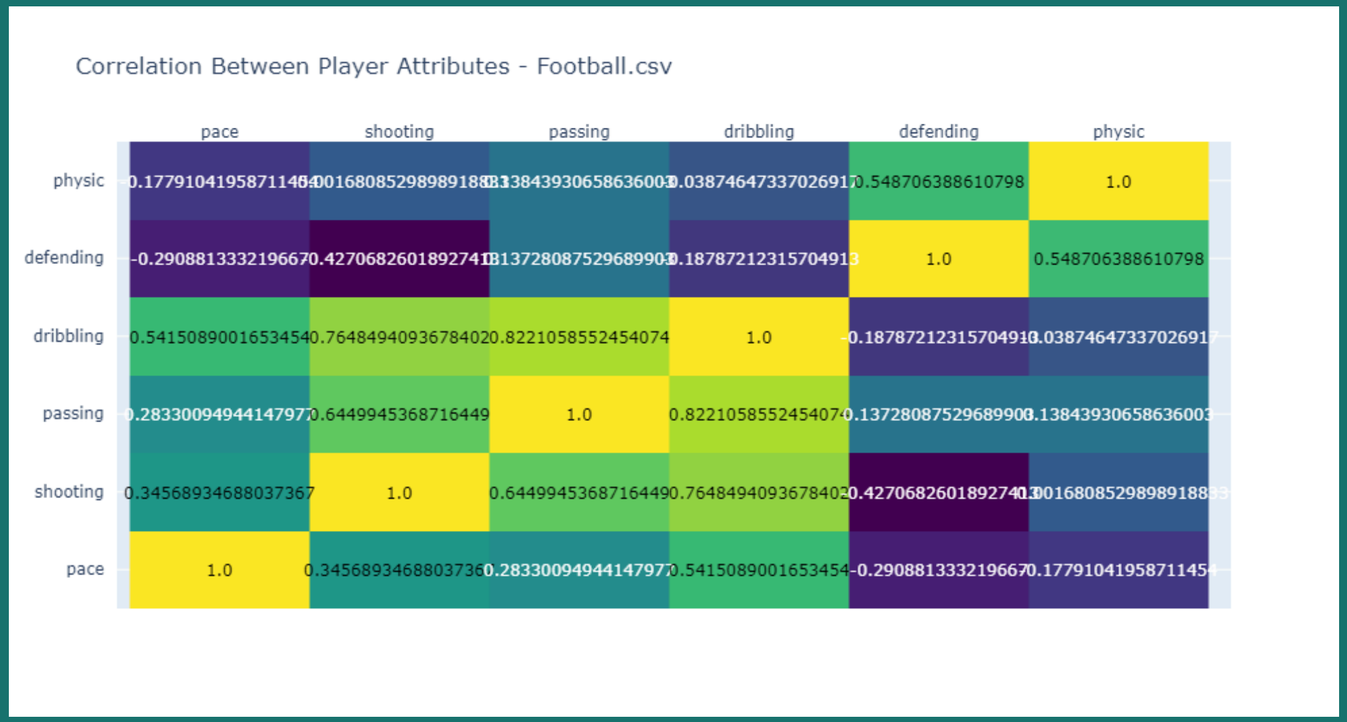
- Optimize clustering parameters: To fine-tune the parameters of the selected clustering algorithm (e.g., number of clusters for K-Means, epsilon and minPts for DBSCAN) to improve clustering results.
- Visualize cluster distributions: To create visual representations of the clusters to gain insights into player distributions and potential overlaps.
- Explore potential applications: To identify potential use cases for the clustering results, such as player scouting, team building, or performance analysis.


Player Count by Position
Number of Players
35k
30k
25k
20k
15k
10k
5k
0
CM
CB
ST
CDM
LM
RM
CAM
RB
GK
LB
RW
LW
CF
LWB
RWB
Position
variable
count

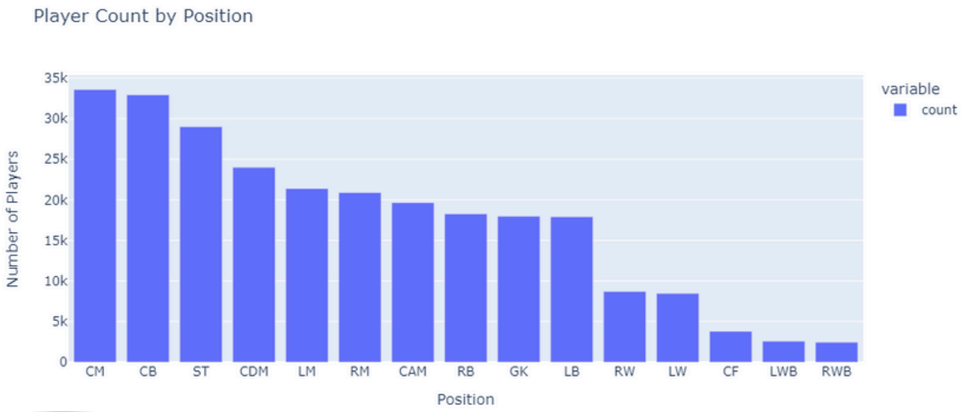

Potential vs. Overall Rating - Football.csv
Potential Rating
90
80
70
60
50
40
40
50
60
70
80
90
Overall Rating
player_positions
CF
LW, LM
RM, LM, RW
ST
GK
CM, LW
ST, CF
CM, CDM
LM
LM, RM
CDM, RB, CM
CB
LM, CAM
RM, RW
CM
ST, CF, CAM
CAM, LW

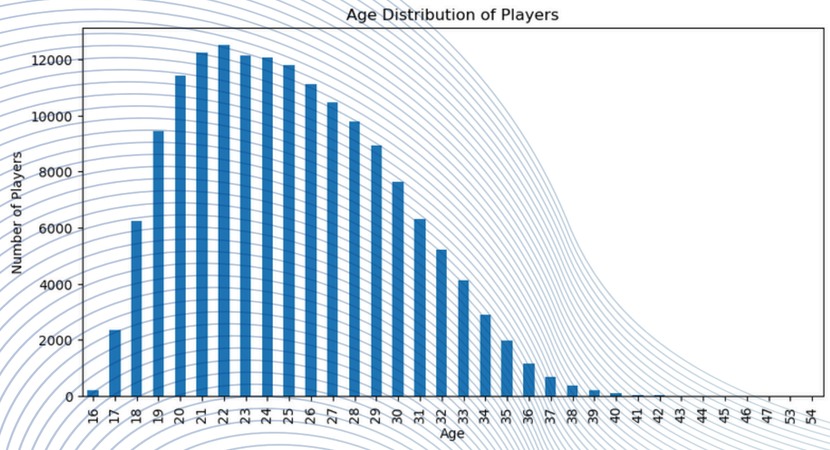

Average Rating by FIFA Version - Football.csv
Average Overall Rating
66
65.5
65
64.5
64
15
16
17
18
19
20
21
22
23
FIFA Version


Top Players by Overall Rating
Player Name
Cristiano Ronaldo
L. Messi
0
100
200
300
400
500
600
Overall Rating
potential
95
94.5
94
93.5
93

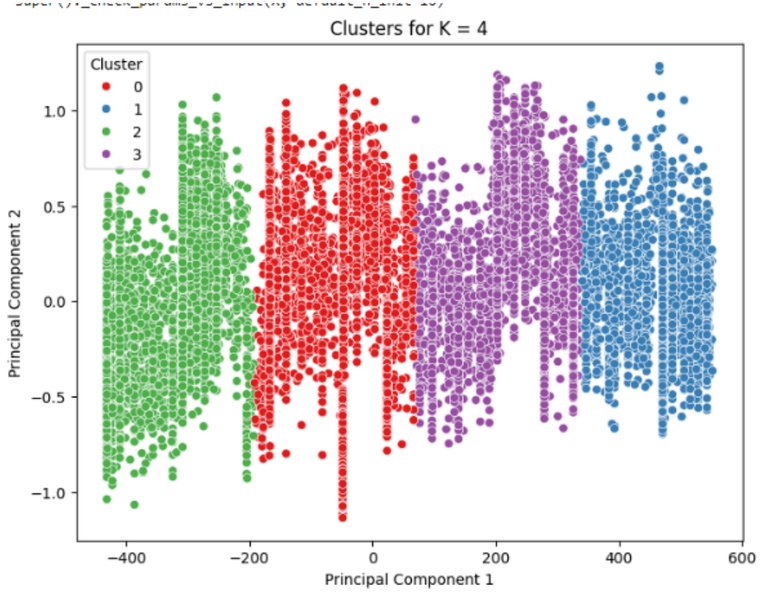

Correlation Between Player Attributes - Football.csv
pace
shooting
passing
dribbling
defending
physic
physic
defending
dribbling
passing
shooting
pace
1.0
0.548706388610798
0.5415089001653454
0.7648494093678402
0.6449945368716449
0.3456893468803736
-0.2908813332196670
0.4270682601892741
0.8221058552454074
0.8221058552454074
0.1372808752968990
0.2833009494414797

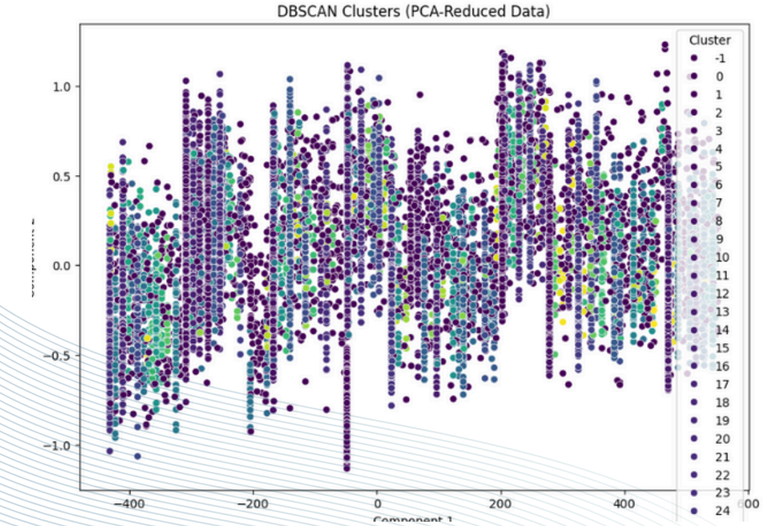
OVERVIEW ON THE ANALYSIS: ANALYTICAL DASHBOARD





- Central Midfielders (CM) lead due to their vital game control, with Center Backs (CB) and Strikers (ST) emphasizing defense and scoring balance.
- Correlation Between Player Attributes: Strong positive correlations between dribbling, passing, and shooting
- Age Distribution of Players: Right-skewed, peak frequency in 21-23 age range. Shows the number of players in each age.
- Prediction Model: To predict a player's overall rating based on their other attributes. Here's an interpretation of the metrics: Linear Regression with high R-squared value (0.9298) and low Mean Squared Error (0.0014).





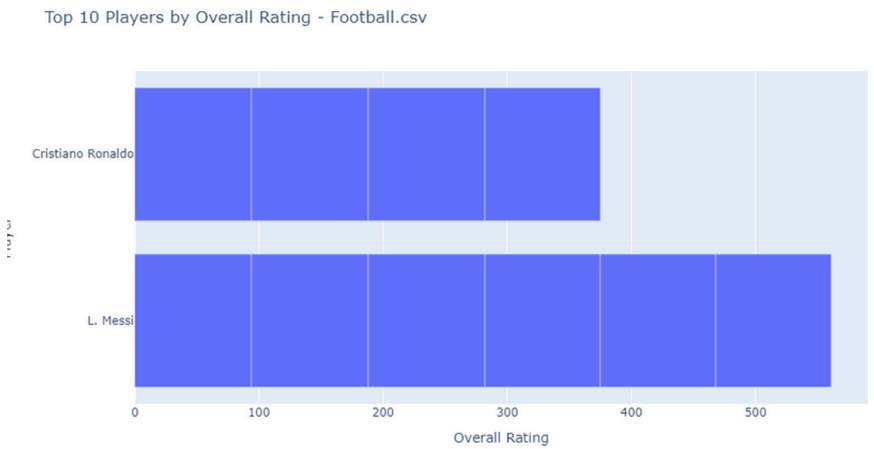
OVERVIEW ON THE ANALYSIS: ML MODEL

K=4 seems to offer a good balance between the Silhouette and Davies-Bouldin scores. However, further exploration and visualization might be necessary to make a definitive decision on the basis of

The Davies-Bouldin Score is 0.36261, suggesting relatively good cluster separation. Inertia is 81625560.0. While inertia is lower for higher K values, it's difficult to determine if 4 clusters is optimal without additional analysis.

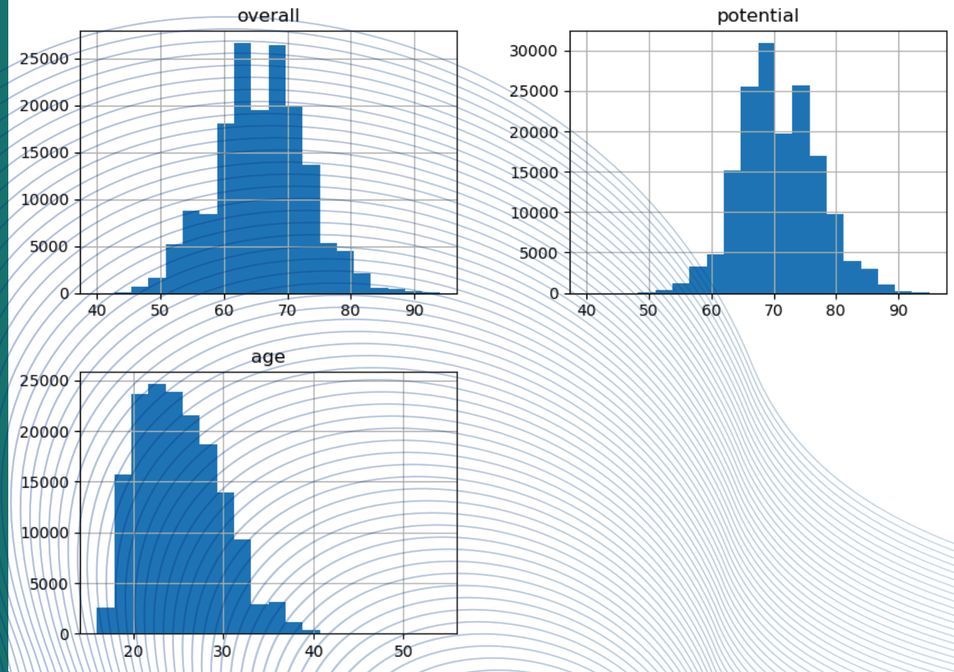
DBSCAN:

Impact of Eps on Clustering: As the eps value increases, the number of clusters also increases, while the number of noise points decreases. This is expected behavior, as a larger eps value implies a larger neighborhood, leading to the inclusion of more points in clusters and consequently fewer outliers.



OVERVIEW ON THE ANALYSIS: DASHBOARD

- Players with high potential ratings tend to have high overall ratings
- Different player positions have distinct distributions in the Potential vs. Overall Rating scatter plot
- Top players like Messi and Ronaldo maintain high overall and potential ratings across multiple years
- Age is a significant factor in player performance, with a decline in player count beyond mid-30s
- Strong positive correlations between attributes like dribbling, passing, and shooting



OBSERVATIONS FROM THE ANALYSIS: ML MODEL

Feature Importance: The selected features, including pace, shooting, passing, dribbling, defending, physical attributes, and player positions, were crucial in defining player profiles and enabling effective clustering.

Clustering Algorithms: Both K-Means and DBSCAN were employed to identify player clusters. K-Means provided a clear partitioning of data into predefined clusters, while DBSCAN was more suitable for discovering arbitrary shaped clusters and handling noise.

Cluster Evaluation: Metrics like Silhouette Score and Davies-Bouldin Index were utilized to assess the quality of the clusters. However, visual inspection of cluster distributions was equally important for gaining insights.

Computational Efficiency: The choice of clustering algorithm and the number of data points impacted computational resources. DBSCAN, in particular, can be computationally intensive for large datasets.

Managerial Insights: Analytical Dashboard

- Invest in top talent and provide a supportive environment for sustained performance
- Focus on developing younger players to replace aging players
- Utilize data-driven insights to inform training and recruitment strategies
- Balance between defensive strength and pace is crucial for team success
- Cross-training players in multiple positions can maximize their impact on the team
- Continuously bring in and develop fresh talent through youth academies and scouting systems

Managerial Insights: ML Model

- Understanding Player Profiles and Building Effective Teams By identifying distinct player archetypes through clustering, football clubs can gain valuable insights to optimize team composition and performance.
- **Targeted Recruitment:**
- Identify player profiles that align with specific team needs, such as a creative midfielder or a defensive stalwart. Focus recruitment efforts on players who fit these identified archetypes to strengthen the squad. Positional Flexibility:
- Analyze player profiles within clusters to identify players with multiple skill sets, enabling them to play in different positions. This flexibility can enhance tactical adaptability and provide depth to the squad. Performance Evaluation:
- Benchmark players against others within their cluster to assess individual performance relative to peers. Identify areas for improvement or potential development based on cluster analysis.
- **Injury Management:**
- Analyze the impact of injuries on cluster composition and team performance. Develop strategies to mitigate the effects of injuries by identifying suitable replacements within existing player profiles.

Managerial Insights: ML Model

Youth Development:

- Identify young players with potential to fit specific player archetypes. Develop tailored training programs to enhance their skills and accelerate their development.

Tactical Implications Formation Optimization:

- Analyze player clusters to identify optimal formations that maximize the strengths of the team. Experiment with different formations based on opponent strengths and weaknesses.

Player Substitution:

- Use cluster analysis to determine suitable substitutes based on the current game situation and the characteristics of the players on the field.

Set Piece Strategies:

- Identify players with specific skills (e.g., crossing, heading, free-kick accuracy) to design effective set piece routines. By leveraging the insights gained from player clustering, football clubs can make data-driven decisions to improve team performance, optimize player utilization, and ultimately achieve greater success.

THANK YOU