Visual Analytics for Soccer Player Profiling: An Interactive Dashboard Using Dimensionality Reduction and Clustering

DAVIDE MAZZATENTA 2140428 AND DAVIDE FORTUNATO 1936575

University of Sapienza, Rome, Italy

Compiled March 27, 2025

This project presents an interactive dashboard for soccer player analysis, utilizing dimensionality reduction and clustering technique to uncover patterns in player attributes from EA Sports FC 25 datasets. The system enables data analysis through coordinated visualizations, including scatterplots, radar charts, and time-series trends. Unlike rigid classification methods, our approach allows flexible filtering and interactive comparisons, providing deeper insights into player roles, skill evolution, and similarities.

1. INTRODUCTION

In recent years, soccer analytics has gained significant traction, with data-driven methodologies playing an important role in player scouting, performance evaluation, and tactical decision-making. Traditional scouting techniques, which primarily relied on subjective assessments and in-game observations, are increasingly being supplemented with computational tools that extract insights from large datasets. This shift has enabled clubs, analysts, and coaches to make more informed decisions about player recruitment and development.

Our project aims to contribute to this evolving landscape by developing an interactive dashboard that facilitates player analysis through data visualization and clustering techniques. Our work leverages a combination of t-SNE for dimensionality reduction and k-means++ clustering to group players based on their attribute similarities. By utilizing data from EA Sports FC 25, we enable users to explore player trends, analyze attributes dynamically and identify emerging talent across different leagues and positions, through customizable filters and comparative analytics.

This dashboard is intended for a diverse range of users, including professional football scouts, analysts, coaches, and even enthusiasts interested in player evaluation, which can make informed recruitment decisions, find insights in performance trends or explore the dataset interactively to gain a deeper understanding of player similarities.

2. RELATED WORKS

Recent advances in soccer analytics have emphasized diverse methodologies for player scouting, spanning predictive simulations to data-driven clustering. Cao et al. [1] introduced Team-Scouter, a framework that simulates tactical scenarios to predict player-team compatibility. Their approach combines agent-based modeling and reinforcement learning to simulate how a player's behavior (e.g., positional awareness, pass selection) adapts to a team's tactical system (e.g., high-pressing vs. counterattacking styles). By generating heatmaps of player movements and quantifying compatibility scores, Team-Scouter identifies recruits likely to thrive in specific formations or roles. For example, it might predict how a midfielder's tendency to drift wide would synergize with a team that prioritizes overlapping fullbacks. While Team-Scouter excels in forecasting tactical fit, its computational intensity and reliance on predefined scenarios limit real-time exploration. In contrast, our work adopts a statistical paradigm, leveraging t-SNE for nonlinear dimensionality reduction and k-means clustering to categorize players based on intrinsic attribute similarities (e.g., pace, dribbling, defensive awareness). By visualizing clusters through interactive scatterplots—augmented with radar charts for multi-attribute comparisons and time-series trends for longitudinal analysis—our dashboard prioritizes attribute-driven discovery. This enables coaches to dynamically explore hybrid roles (e.g., a forward with defender-like tackling) or identify undervalued players whose profiles deviate from rigid tactical templates. Unlike Team-Scouter's predictive focus, our system emphasizes flexible, user-guided analysis, empowering stakeholders to iteratively refine queries without relying on simulated outcomes.

The growing recognition of non-traditional data sources has further enriched this domain. Cotta et al. [2] validated the FIFA video game dataset as a viable proxy for soccer analytics, analyzing tactical phenomena like FC Barcelona's tiki-taka midfield using curated attributes (e.g., dribbling, vision). Building on this, Soto-Valero [3] advanced the use of FIFA data by applying Gaussian mixture clustering to categorize 7,705 European players into four roles (defenders, midfielders, forwards, goalkeepers). Their work identified dribbling as the most discriminative attribute and utilized PCA for dimensionality reduction, establishing a statistical foundation for role-based player analysis. However, their static, model-driven approach lacks interactive exploration tools, limiting real-time insights into player similarities or attribute evolution.

Our project bridges these gaps by operationalizing the FC 25 game dataset for dynamic, user-driven analysis. While Soto-Valero's clustering relies on predefined roles and retrospective PCA, our system employs t-SNE projections to reveal emergent player clusters and k-means to accommodate fluid positional roles in modern soccer. Unlike Soto-Valero's rigid Gaussian mixture classification into four groups, our dashboard enables coaches to interactively filter players by roles (e.g., defenders, wingers), leagues, and age, while brushing the scatterplot to explore subclusters or hybrid player profiles. For instance, coaches could filter midfielders with high finishing attributes to identify "utility players" who blur traditional positional boundaries, or examine how dribbling trends vary across leagues. Furthermore, we extend prior work by integrating time-series visualizations to track attribute evolution (e.g., defensive awareness over seasons) and radar charts to compare player profiles, addressing Soto-Valero's limitation of static analysis through real-time, multi-criteria exploration.

3. DATASET

For this study, we integrated two complementary datasets. The primary dataset originates from EA Sports FC 25 player statistics scraped from SoFIFA [4], the authoritative community platform for FIFA game data. This dataset [6], hosted on GitHub, contains 18,735 players characterized by 76 attributes including technical skills, physical metrics, and demographic information. The second dataset used is the EA Sports FC 24 Complete Player Dataset from Kaggle [5], which includes data from FIFA 15 to EA Sports FC 24 with 109 attributes per player. We combined these datasets to gain better insights into the evolution of football players over time, as visualized in the line chart.

Fig. 1. Example of player attributes ratings in FC 25 game

A. Data pre-processing

Given the large size of the main dataset, we performed several pre-processing steps. First, we conducted an exploratory data analysis to examine the distribution of players based on their overall rating, which ranges from 45 to 91. The distribution followed a normal pattern. To reduce the dataset size while retaining relevant information, we filtered out players with an overall rating below 65, as lower-rated players are generally less known. This filtering resulted in a dataset containing slightly more than half of the original players.

Next, we analyzed the dataset's features to determine which attributes were necessary for our study. We removed 15 columns related to IDs and other non-informative attributes. We then selected 29 key attributes related to player performance, including shooting, defending, dribbling, passing, pace, and physicality, for dimensionality reduction and clustering using k-means++.

Handling missing data was another crucial step. Sparse data, such as null or NaN values, were either dropped or replaced using column-wise mean values, depending on the attribute type. Finally, we applied standard scaling to normalize the selected features, ensuring compatibility with our analytical techniques.

After pre-processing, we obtained a refined dataset containing 10878 players with 61 features respecting the Angelini-Santucci index, with AS = 663558.

4. VISUALIZATION TECHNIQUES

In our project we used a total of 6 visualizations. There are no standing alone visualization, each of them interacts with the others by applying filters or showing features. The dashboard is composed by 6 cards, one for each visualization, and each of them has an header with the title and tools to interact with. The visualizations are: Scatterplot, Soccer Field, Player Information, Bar Chart, Radar Chart, Line Chart.

A. Scatterplot

The scatterplot visualization is a dynamic, interactive, and data exploration tool designed to represent player distributions based on t-SNE dimensionality reduction and k-means++ clustering. It maps high dimensional players data into a two-dimensional space, where proximity indicates similarity in playing style, attributes and performances. The scatterplot serves as the central hub, integrates filter mechanism and links other visualizations to provide a comprehensive scouting and analytical framework. Players are color-coded by cluster, using a standard `colorbrewer2` qualitative palette, categorizing them into distinct playing styles: defensive, technical, offensive, athletic players and goalkeepers.

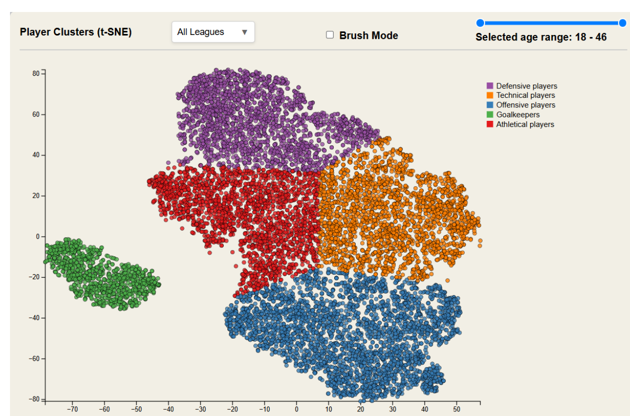

Fig. 2. Scatterplot clusters visualization

The visualization automatically adjusts based on the dimensions of its container, ensuring a fluid display and the axes are defined based on the original data extent, ensuring a stable reference framework even when new filters are applied. Hovering a player with the mouse displays some details about him like name, position, age and overall rating. Clicking on a player triggers updates in linked visualizations, such as bar chart, radar chart, line chart and player information section. If compare mode is active, users can select a second player, but only within the same role classification, so goalkeepers cannot be compared with outfield players. A pie chart displays dynamically the distributions of clusters according to filters and brushing mode and a legend updates the number of total players visualized.

In the header of the scatterplot there are 3 filters:

- **League-based filter** that contains all the possible league in FC25 and enable users to filter on this criteria;
- **Age filter** implemented as a double slider that helps you in finding players with desired age;
- **Brush mode checkbox** that if enabled allow users to select multiple players, even from different clusters, which updates only the bar chart with inherent information.

B. Soccer Field Line-Up

The soccer field line-up is implemented with an image of a soccer field in background and circles in each possible role present in FC25. We used a qualitative color brewer palette for each major role, goalkeeper, defenders, midfielders and attackers.

By clicking a role on the pitch it will be highlighted and we filter players on the scatterplot by showing only those players with main role as the one selected. This type of filtering is cumulative up to 3 roles, so by clicking another role we filter by both of them. In the soccer field header there is a reset button which reset each kind of filter in each visualization.

C. Player Information

The player information visualization displays a player profile with an image and some attributes. Among these we find the overall rating, the positions he can play, the club he is playing at the moment with its logo, personal info and play styles. Player image and club logo are displayed only if they are present in the dataset, otherwise there is a placeholder image. At the bottom of the profile, we can find clickable names of the nearest and most similar players to him. By clicking one of them the visualization will update.

D. Bar Chart: Player and Clusters

This bar chart visualization is designed to compare a selected player attributes value against the average values of the cluster he belongs to. By clicking on a player of the scatterplot, the bar chart shows the top 10, over 34, attributes with the highest cluster averages. The Y-axis is always in a scale from 0 to 100, while the X-axis displays the names of 10 attributes that changes dynamically based on the clusters. The colors chosen for the bars are a 2 color palette from `Coolors` website. If the user has brushed mode checked, the bar chart does not display a single player's attributes, instead, it calculates and displays the top 15 highest average attributes values of the brushed players.

A checkbox in the header allows to calculate averages values only on visible players of the scatterplot granting a more detailed comparison between group of players. A legend permits to distinguish between individual and the cluster/brushed averages values. A tooltip appears when hovering a bar with the mouse showing the precise value of player and cluster. This chart helps in understanding how a player compares to similar players in the dataset, allows to highlight key attributes that excels or underperforms, and permits analysis on brushed subsets of players chosen by the user.

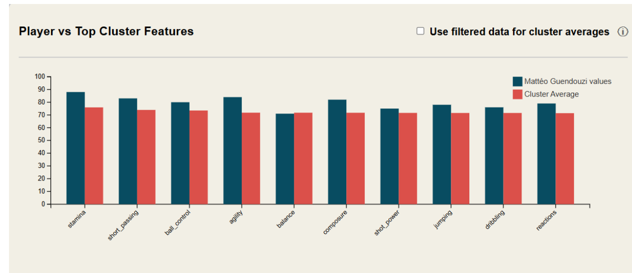

Fig. 3. Bar Chart player and cluster visualization

E. Radar Chart: player attributes

The radar chart has a structured and dynamic design, displaying multiple axes and a grid starting from a center point. Each axis represents a distinct player attribute that in case of goalkeepers is a pentagon with labels like diving, handling, kicking, positioning and reflexes and in case of other players the it is a hexagon with labels like shooting, passing, dribbling, defending, pace, or physics. The grid provides reference levels for performance values from 0 to 100 in 5 steps.

We interact with the radar chart by clicking a player on the scatterplot that visualize its shape according to the type of player with its aggregated statistics(each of the is the mean of different numerical attributes). Each player's data is illustrated through a colored, semi-transparent polygon, forming a unique shape that highlights their strengths and weaknesses. The color palette used in this case is a `d3.js` standard palette called `schemeTableau10`. If we hover the mouse cursor on a player's shape, an interactive tooltip appears, revealing the player's name and the corresponding polygon expands in opacity and border thickness, emphasizing the data.

In the header of the card there are 2 possible tools that can be used. The first one is a "compare mode" checkbox, which if checked, permits to click on another player of the scatterplot to watch its radar chart and comparing the two. There is a restriction if you want to compare different types of player like goalkeepers with others, because they have different attributes and so shapes. The second one is a slider from 0 to 5 that shows the chart of at most 5 nearest players, pointing out the fact that, according to the k-means++ algorithm, the players are well clustered and similar in their attributes if they are near.

F. Line Chart: player trends

The line chart is designed to track and analyze a selected player's performance and financial evolution over multiple FIFA versions, from 2015 to 2025. This visualization enables users to observe trends, fluctuations, and player development over time. For this visualization we used the second dataset that contains players from last versions of FIFA games.

A drop down menu in the header of the card allows users to switch between 3 different metrics. When a new metric is selected, the current chart is cleared and redrawn dynamically. The X-axis always represents time in years, and it on the available data for the selected player in the datasets. The Y-axis changes whenever a new metric is selected and will have different scales. Those metrics are:

- **Wage**, the default one, that in most of the cases contains value formatted in the dataset as thousands(K) so has a scale of thousands of euros;

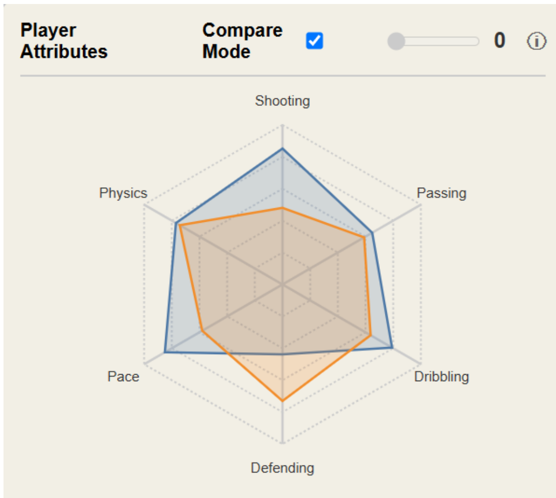

Fig. 4. Radar Chart compare visualization

- **Value**, that contains value formatted in the dataset as millions(M) so has a scale of millions of euros;
- **Statistics**, that have a scale from 0 to 100. This chart automatically adapts to the player's position, so if it is a goalkeeper the attributes displayed are diving, handling, kicking, positioning, and reflexes, while for all other players, the attributes include pace, dribbling, physics, shooting, passing, and defending. .

The visualization is further enhanced by color coding explained in the legend. When displaying wage or value, the line appears in light blue. For statistics, each attribute is represented by a distinct color-coded line that follows the same qualitative palette used in the scatter plot, improving readability. There is also a tooltip when hovering with the mouse that highlights the respective line, and each point on the line displays a label with the attribute name and its corresponding value. Additionally, a trend explainer has been implemented to emphasizes significant patterns in a player's performance attributes. This section dynamically displays characteristics that notably increases or decreases over seasons, making it easier to identify key developments in a player's career.

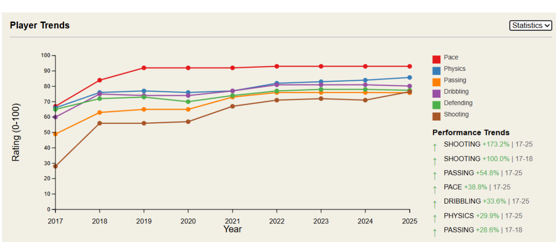

Fig. 5. Line Chart statistics visualization

5. CASE STUDIES AND INSIGHTS

Our dashboard supports data-driven decision-making in scouting, coaching, and financial planning through interactive visualizations. The following case studies highlight its use in talent identification, squad analysis, and market intelligence, demonstrating how it quantifies player attributes, enables objective comparisons, and uncovers value opportunities. These insights help optimize recruitment, tactical decisions, and transfer strategies.

A. Talent Scouting for Position-Specific Player Profiles

A scouting team was tasked with identifying a young defensive midfielder using the visual analytics dashboard. The process began by filtering players aged 17-23 from European leagues known for developing future stars such as Eredivise or Primera Liga. This selection initially revealed several hundred candidates. The t-SNE visualization automatically grouped these players into distinct clusters based on their playing characteristics, with our focus drawn to the defensive-oriented cluster that emerged organically from the data. Using the field position filter, we narrowed our analysis specifically to central midfield roles, which reduced the candidate pool to a manageable selection of defensive-minded players.

The dashboard's comparative tools then enabled detailed evaluation of promising candidates and in particular, the bar chart visualization allowed us to compare key defensive metrics between selected players and their respective cluster averages. This revealed one particular candidate whose defensive attributes consistently appeared in the top percentile when benchmarked against his cluster peers. The radar chart provided further insight by enabling side-by-side comparison of six fundamental characteristics between two final candidates, highlighting one player's superior balance of defensive and technical qualities.

Temporal analysis through line charts showed this player's consistent performance improvement and steady market value growth over multiple seasons, despite his team performance was not brilliant. Crucially, all attribute insights were derived exclusively through comparative visualizations - either against cluster averages or through direct player comparisons - as the dashboard intentionally surfaces relative rather than absolute metrics. This approach successfully identified a young midfielder whose defensive capabilities and development trajectory suggested significant untapped potential, therefore representing what the scouting team was asked.

B. Opponent Profiling and Match Strategy Adjustment

In preparation for an upcoming match against a team from a lesser-known league, the coaching staff used the dashboard's Brush Mode to explore the opponent's player profiles. By drawing different boxes in the t-SNE scatterplot, they examined how various groups of players compared in terms of top 15 attributes displayed in the bar chart.

Initially, a broad selection of the opponent's midfielders was analyzed, revealing unexpectedly high physical and pace attributes, suggesting a style favoring aggressive pressing and rapid transitions. Refining the selection to wide players, the staff noticed strong dribbling and crossing abilities, indicating a tendency to create chances from the flanks. Another trial focused on their defensive line, which exhibited relatively low physicality and aerial ability, hinting at potential vulnerabilities against direct play.

Based on these insights, the team adapted its match strategy accordingly. Defensively, they prioritized compactness in midfield to counteract the opponent's high-intensity transitions. Offensively, they sought to exploit crossing situations and targeted aerial duels against the weaker defensive line. Additionally, recognizing the opponent's reliance on wing play, fullbacks were instructed to double-mark wide players and limit their ability to deliver dangerous crosses.

Through an iterative approach, leveraging Brush Mode and bar chart insights, the coaching staff refined their tactical plan, ensuring their squad was well-prepared to counter the strengths and exploit the weaknesses of an unfamiliar opponent.

C. Comparing Player Profiles Across Different Leagues for Market Intelligence

A team of soccer analysts working for a club's recruitment department aimed to evaluate how players from different leagues compare in terms of physicality and technical skills. The goal was to identify leagues that produce players with a specific profile suited to their playing philosophy.

Using the scatterplot, they applied a league-based filter to isolate players from different competitions, such as the Bundesliga, Premier League, and Serie A. Observing how clusters formed within each league helped them detect structural differences in player characteristics. For example, Bundesliga players appeared more evenly distributed across all the clusters, while Premier League players showed a quite concentration in the physical cluster.

The analysts then used the radar chart to compare midfielders from different leagues, confirming that Serie A players had higher passing and defending, whereas Bundesliga players in average tend to have more pace aligning with a play style focused more in pressing and ball recovery. The bar chart reinforced these findings by revealing the top 10 most distinctive attributes in each league.

To further refine their analysis, they used the line chart to track how players from different leagues evolved over time, focusing on those who transferred to new environments. This helped assess whether certain leagues develop specific skill sets that persist when players move abroad.

These insights guided the club's scouting strategy, allowing them to prioritize leagues aligned with their tactical model and identify undervalued markets for potential transfers.

6. CONCLUSIONS AND FUTURE WORK

Our case studies demonstrated the dashboard's practical utility in talent scouting, opponent analysis, and market intelligence. For instance, the system successfully identified undervalued players, highlighted tactical weaknesses in opposing teams, and revealed league-specific player profiles. Unlike static classification methods, our approach accommodates the fluid nature of modern soccer roles, enabling users to discover hybrid player types and track performance trends over time.

The integration of multiple visualizations ensures a comprehensive understanding of player attributes, while interactive features like brushing and comparison modes facilitate deeper exploration. By bridging the gap between data-driven analytics and intuitive visual interfaces, this project contributes to the growing field of soccer analytics, offering a tool that enhances decision-making in recruitment, tactical planning, and player development.

While the current dashboard provides a robust foundation for player analysis, several enhancements could further improve its utility:

- **Expanded Data Integration**, incorporating real-time performance metrics from match-tracking systems (e.g., Opta or StatsBomb) could enrich the dataset, enabling more nuanced analysis of on-field contributions beyond static attributes.
- **Advanced Clustering Techniques**, exploring hierarchical clustering or density-based algorithms (e.g., DBSCAN) might reveal finer-grained player subgroups, particularly for hybrid roles like "false nines" or "wing-backs."
- **Predictive Analytics**, integrating machine learning models to forecast player development or transfer market value could provide forward-looking insights, aiding long-term recruitment strategies.
- **Multi-Team Comparison**, extending the dashboard to compare squad compositions across leagues or seasons could help identify systemic strengths or gaps in team rosters.
- **Line-Up simulator**, implementing a tool that generates optimal line-ups based on inputs like budget, player strength, and tactical preferences, balancing squad quality, constraints, and strategy to suggest the best possible starting XI.

REFERENCES

1. A. Cao, X. Xie, R. Zhang, Y. Tian, M. Fan, H. Zhang, and Y. Wu, *TeamScouter: Simulative Visual Analytics of Soccer Player Scouting*, IEEE Transactions on Visualization and Computer Graphics, vol. 31, no. 1, pp. 1-11, 2025.
2. COTTA, Leonardo, et al. *Using fifa soccer video game data for soccer analytics.* In: Workshop on large scale sports analytics. 2016. p. 1-4.
3. C. Soto-Valero, *A Gaussian mixture clustering model for characterizing football players using the EA Sports' FIFA video game system*, Revista Internacional de Ciencias del Deporte, vol. 13, no. 49, pp. 244-259, 2017.
4. *SoFIFA website*, https://sofifa.com.
5. *EA Sports FC 24 Complete Player Dataset*, 2023. Available: https://www.kaggle.com/datasets/stefanoleone992/ea-sports-fc-24-complete-player-dataset.
6. *SoFIFA Web Scraped Dataset*, 2024. Available: https://github.com/prashantghimire/sofifa-web-scraper.