Player Valuation and Segmentation in EA FC 25

A Final Term report for the BDM capstone Project

Submitted by

Name: OAS PODDAR

Roll number:23F3003433

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

<u>Declaration Statement</u>

I am working on a Project Title “**Player Valuation and Segmentation in EA FC 25**”. I extend my appreciation to **Electronic Arts**, for providing the necessary resources that enabled me to conduct my project.

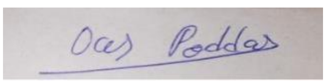
I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered through secondary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the information of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I agree that all the recommendations are business-specific and limited to this project exclusively, and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT Madras does not endorse this.

Signature of Candidate: 

Name: OAS PODDAR

Date: 2/12/2025

1) Executive Summary

Project Title: Player Valuation and Segmentation in EA FC 25

Electronic Arts Inc. (EA), founded in 1982 and headquartered in Redwood City, California, is a global leader in interactive entertainment and sports simulation gaming. Through its EA SPORTS division, the company creates highly realistic, data-driven titles such as the EA SPORTS FC series. EA integrates real-world player statistics, performance metrics, advanced analytics, and modern graphics to deliver authentic football gameplay. Despite this, challenges remain in ensuring consistent and accurate player ratings. Inaccurate valuation and the absence of a structured segmentation method often led to imbalanced gameplay, reduced fairness, and limited analytical insights for improving player assessment and game realism.

The objective of this study is to utilize a data-driven approach to evaluate the current player dataset and build models that can more accurately assess player performance and categorize them into meaningful groups. The project uses the official EA FC 25 dataset obtained from Kaggle, followed by data cleaning, preprocessing, and exploratory analysis. Statistical and machine learning techniques—including correlation analysis, regression modeling, and clustering—are applied to identify key attributes influencing player ratings and to segment players based on performance metrics.

The outcomes are expected to provide a refined valuation model that reduces over- or undervaluation, along with effective player clusters that support better comparison and tactical decision-making. These insights will enable EA SPORTS to enhance gameplay balance, update player ratings in a more systematic manner, and strengthen future game development processes through data-supported decision making.

2) Proof of Originality

The dataset used in this project has been obtained from a publicly accessible and credible online data repository. All player information used for analysis—such as ratings, attributes, positions, and demographic details—was sourced from **Kaggle**, a well-known platform for curated datasets and analytical resources. The specific dataset utilized is the *EA SPORTS FC 25 Players Dataset*, originally published by independent contributors for academic and research purposes.

The dataset was downloaded directly from the Kaggle repository by the candidate and has not been altered by any third party prior to analysis. All preprocessing steps, including data cleaning, handling of missing values, normalization, and feature engineering, were performed solely by the candidate to ensure authenticity and methodological transparency.

No external computational assistance, automated tools, or pre-built analytical scripts were used beyond those explicitly required for data processing and modeling. The project does not include any content reproduced from external sources without attribution. All interpretations, models, visualizations, and insights are independently developed as part of this capstone project.

The Excel files used for the final analysis—including **all_players.csv**, **male_players.csv**, and **female_players.csv**—were created from the original Kaggle dataset through candidate-driven cleaning and categorization steps.

Kaggle Link: https://www.kaggle.com/datasets/nyagami/ea-sports-fc-25-database-ratings-and-stats

Analysis Notebook (Spreadsheet):
https://docs.google.com/spreadsheets/d/1JPoVHZRMVMjIs0Gv93DjgYH_Rn1LmoPW/edit?usp=drive_link&ouid=102298127055320870470&rtpof=true&sd=true

Descriptive statistical analysis for this project was performed using spreadsheet tools (Microsoft Excel). The computed statistical summaries and metrics have been documented in the accompanying spreadsheet file and shared through the referenced link.

The link of the datasets used in this project are given below:

all_players.csv:
https://drive.google.com/file/d/1570Z9vazBibaCiHQnUTdm2uDuFadElsT/view?usp=drive_link
male_players.csv:
https://drive.google.com/file/d/13Gisv2s3bsXlk8hF0gZxhNHO5cTUtb6Q/view?usp=drive_link
female_players.csv:
https://drive.google.com/file/d/114R3QMutGSV0ecPP9C4-7xb-wBCy9kS3/view?usp=drive_link

3) Meta Data and Descriptive Statistics

3.1 Meta Data

Dataset Overview

The dataset used in this project is the **EA SPORTS FC 25 Player Dataset**, sourced from Kaggle and contains detailed demographic, positional, physical, technical, and performance-related attributes for professional football players. The dataset consists of:

- Total Rows: 17,737 players
- Total Columns: 60 attributes

This dataset includes both outfield players and goalkeepers, providing a comprehensive foundation for player valuation and segmentation.

List of Columns and Data Types

A. Identifiers and Metadata

Column	Type	Description
Column1	Integer	Serial index from source dataset
Unnamed:0	Integer	Original index from Kaggle file
Rank	Integer	Player rank based on overall rating
Name	Text	Full name of the player
url	Text	EA official rating page for the player

B. Overall Performance Ratings

Column	Type
OVR	Numeric

C. Key Attribute Groups

I) Pa ce Attributes

Column	Type
PAC	Numeric
Acceleration	Numeric
Sprint Speed	Numeric

II) Shooting Attributes

Column	Type
SHO	Numeric
Positioning	Numeric
Finishing	Numeric
Shot Power	Numeric
Long Shots	Numeric

Volleys	Numeric
Penalties	Numeric

III) Passing Attributes

Column	Type
PAS	Numeric
Vision	Numeric
Crossing	Numeric
Free Kick Accuracy	Numeric
Short Passing	Numeric
Long Passing	Numeric
Curve	Numeric

IV) Dribbling Attributes

Column	Type
DRI	Numeric
Dribbling	Numeric
Agility	Numeric
Balance	Numeric
Reactions	Numeric
Ball Control	Numeric
Composure	Numeric

V) Defensive Attributes

Column	Type
DEF	Numeric
Interceptions	Numeric
Heading Accuracy	Numeric
Def Awareness	Numeric
Standing Tackle	Numeric
Sliding Tackle	Numeric

VI) Physical Attributes

Column	Type
PHY	Numeric
Jumping	Numeric
Stamina	Numeric
Strength	Numeric
Aggression	Numeric

D. Biological Attributes

Column	Type
Position	Categorical
Height	Text
Weight	Text
Alternative positions	Text
Age	Numeric
Nation	Text
League	Text
Team	Text

E. Skill Attributes

Column	Type
Weak foot	Numeric (1-5 scale)
Skill moves	Numeric (1-5 scale)
Preferred foot	Text
play style	Text

F. Goalkeeper Attributes

Column	Type
GK Diving	Numeric
GK Handling	Numeric
GK Kicking	Numeric
GK Positioning	Numeric
GK Reflexes	Numeric

Data Suitability

This dataset is highly suitable for:
1. Player valuation (regression analysis)
2. Player grouping and segmentation (clustering)
3. Attribute importance analysis
4. Performance comparison

3.2 Descriptive Statistics

Overview

The descriptive statistics were computed for all 45 numerical columns present in the EA SPORTS FC 25 dataset in order to understand the distribution and variability of the players attributes and skill sets. The metrics used in the descriptive statistics include **Mean**, **Median**, **Minimum value**, **Maximum value**, **Standard Deviation**, **Count**, **Missing values**, **Quartile 1**, **Quartile 3**, **Interquartile range** offering a comprehensive summary of each attribute's statistical behavior.

Interpretation of Statistical Findings

Attributes such as **OVR, PAC, SHO, PAS, DRI**, and **PHY** show moderate variation, indicating balanced performance dispersion across the global player pool. Technical attributes including **Ball Control, Reactions, Vision**, and **Short Passing** display higher mean values, suggesting that technical skill is a common strength across players. Meanwhile, attributes such as **Strength, Aggression**, and certain defensive metrics exhibit wider variability, reflecting diverse physical and tactical roles.

Goalkeeper-specific attributes such as GK Diving, Handling, Kicking, Positioning, Reflexes contain many zeros for outfield players, which is expected in a mixed dataset. These zeros do not affect overall analysis but highlight the need for position-based segmentation in later modeling.

Descriptive Statistics Summary Table

The descriptive statistics for this project were compiled in a dedicated Excel workbook to summarize all numerical attributes present in the EA SPORTS FC 25 dataset. This workbook provides a clear understanding of the dataset's distribution patterns, central tendencies, variability, and completeness. It contains two structured sheets—**num_col** and **metrics**—each serving a specific role in presenting and organizing the statistical outputs.

Link to the Descriptive Statistics file**:**
https://docs.google.com/spreadsheets/d/1JPoVHZRMVMjIs0Gv93DjgYH_Rn1LmoPW/edit?usp=drive_link&ouid=102298127055320870470&rtpof=true&sd=true

Sheet Name: num_col
The num_col sheet contains the raw descriptive statistics (mean, median, min, max, standard deviation, count, missing values, quartile 1, quartile 3, interquartile range) for all 45 numerical columns in the dataset. It lists each numerical attribute individually along with its computed summary metrics. This sheet acts as the complete statistical output directly derived from the dataset.

Sheet Name: metrics
The metrics sheet provides a clean, organized summary table containing only the key statistical metrics (Mean, Median, Min, Max, Std Dev, Quartiles, Interquartile range etc.) without the raw calculations displayed. It serves as a refined and presentation-ready version of the descriptive statistics for reporting purposes. This sheet allows quick interpretation of attribute behavior across players.

4) Detailed Explanation of Analysis Process

4.1 Tools Used

The following tools and platforms were used during the execution of this project:

Software & Platforms

- Microsoft Excel – Data cleaning, formatting, and descriptive statistics
- Google Colab – Running Python code, building regression and clustering models
- Google Drive – Storing datasets, notebooks, and sharing links

Programming Languages

- Python 3

Python Libraries

- pandas – Data manipulation and preprocessing
- numpy – Numerical operations
- scikit-learn – Machine learning (Linear Regression, K-Means Clustering, preprocessing, evaluation)
- matplotlib & seaborn – Data Visualizations

Dataset Source

- Kaggle EA SPORTS FC 25 Dataset (secondary data source)

The analysis process for this project was executed in a structured manner, beginning with dataset import, cleaning, and preprocessing, followed by exploratory and descriptive statistical analysis. Each step was performed to ensure that the EA SPORTS FC 25 dataset was reliable, consistent, and suitable for subsequent modelling.

4.2 Analytical Methodologies Used

1. <u>Descriptive Statistical Analysis</u>: Descriptive statistics were used to summarize player attributes using measures such as mean, median, quartiles, and interquartile range. This analysis helped understand the distribution and variability of numerical features across players.

2. <u>Correlation Analysis</u>: Correlation analysis was performed to examine relationships between player attributes and Overall Rating (OVR). It helped in identifying the most influential attributes contributing to player ratings.

3. <u>Linear Regression Analysis</u>: Linear regression was used to predict Overall Rating (OVR) based on numerical player attributes. The model evaluated how effectively player ratings can be explained using measurable characteristics.

4. <u>K – Means Clustering</u>: K-Means clustering was applied to group players into distinct performance-based segments using core skill attributes. This enabled role-based classification and comparison of similar players.

4.3 <u>Data Import and Initial Inspection</u>

The original EA SPORTS FC 25 dataset was first downloaded from Kaggle in CSV format. This file was opened in Excel and saved as all_players.csv for further work.

The following initial checks were performed:

- Verified the total number of rows and columns (17,737 players and 60 attributes) using Excel's row and column counts.
- Ensured that all 45 numerical attributes were correctly recognized as numeric fields by applying Excel's sorting and filtering options.
- Reviewed key attribute groups (Pace, Shooting, Passing, Dribbling, Defence, Physical, Goalkeeping, and Biological attributes) to confirm that values were complete and logically aligned.
- Noted the presence of zeros in goalkeeper-specific columns for outfield players, which is an expected characteristic and considered during descriptive statistical analysis.

4.4 <u>Data Cleaning and Preparation</u>

Following the initial inspection, the dataset was reviewed in Excel to ensure that all numerical attributes were correctly formatted for statistical computation. Using Excel's sorting, filtering, and number-format tools, each of the 45 numerical columns was checked to confirm that values were properly recognized as numeric entries and not stored as text. Minor formatting inconsistencies were corrected, and any blank or missing values were noted for inclusion in the "Missing Values" column of the descriptive statistics. No structural issues such as duplicated rows or misaligned columns were found, confirming that the dataset was clean and ready for analysis.

4.5 Computation of Descriptive Statistics (num_col Sheet)

Once the dataset was confirmed to be clean, descriptive statistics were computed for all numerical attributes using Excel formulas. For each attribute, the following measures were calculated:

- **Mean** using AVERAGE()
- **Median** using MEDIAN()
- **Minimum and Maximum** using MIN() and MAX()
- **Standard Deviation** using STDEV.S()
- **Count of valid entries** using COUNT()
- **Missing values** using COUNTBLANK()
- **Quartiles** using QUARTILE.INC()

All computed statistics were compiled in the **num_col** sheet. This sheet contains the full, raw statistical output for each of the 45 numerical columns exactly as derived from the dataset.

4.6 Preparation of Summary Table (metrics Sheet)

To present the computed statistics in a clearer and more readable format, the results from the num_col sheet were reorganized into a simplified summary table on the **metrics** sheet. This sheet displays only the essential descriptive metrics—Mean, Median, Minimum, Maximum, Standard Deviation, Count, Quartile 1, Quartile 3 and missing values—for each numerical attribute. The intention of this sheet is to provide a clean, presentation-ready overview without showing intermediate calculations, allowing easy comparison of attribute behavior.

4.7 Attribute Grouping and Observation

To understand the behavior of different football performance categories, the numerical attributes were reviewed in groups based on how they appear in the dataset and in the Excel descriptive statistics file. The key observations made from the metrics sheet are summarized as below:

1. Pace Attributes (PAC, Acceleration, Sprint Speed)
Pace values are generally high for most players, with a large difference between the slowest and fastest ones. This reflects natural variation across positions, where attackers tend to be much quicker than defenders.

2. Shooting Attributes (Finishing, Shot Power, Long Shots, etc.)
Shooting attributes show a wide spread because only forwards and attacking players excel strongly in these skills. Some metrics like Shot Power have higher averages, while others vary more depending on the player's role.

3. Passing Attributes (Vision, Crossing, Short Passing, etc.)
Passing attributes are mostly consistent and have steady averages across players. Key skills like Vision and Short Passing are strong for many players, while specialized skills like Free Kick Accuracy vary more.

4. Dribbling Attributes (Ball Control, Agility, Balance, etc.)
Dribbling-related metrics show strong averages, especially in Ball Control and Reactions. However, attributes such as Agility and Balance differ more depending on a player's physical style and position.

5. Defensive Attributes (Interceptions, Tackle Stats, Awareness)
Defensive attributes vary widely because only defenders specialize in them. Some skills like Defensive Awareness tend to have higher averages, showing clear differences between defensive and attacking roles.

6. Physical Attributes (Stamina, Strength, Aggression, Jumping)
Physical attributes show noticeable variation, especially in Strength and Aggression, which differ greatly between players. Stamina is generally high across the dataset since most players rely on endurance.

7. Goalkeeper Attributes (GK Diving, Reflexes, Positioning, etc.)
These attributes show many zero values for outfield players, leading to large variation. Since only goalkeepers use these skills, they must be viewed separately from other performance categories.

8. Skill Attributes (Weak Foot, Skill Moves)
Skill attributes mainly fall in mid-range values for most players, with only a small group achieving very high ratings. These metrics capture individual creativity and technical capability on the field.

5) Results and Findings

Reference to Modelling Notebook

All modelling tasks (correlation analysis, regression modelling, and clustering) were performed in a Google Colab notebook titled
"EA_FC25_Player_Valuation_and_Segmentation_Analysis.ipynb".
The notebook is accessible at the following link:
https://colab.research.google.com/drive/1gDfQZ-NznLFT--0mfct_iQK9jUJSILBq?usp=drive_link

Introduction

This section presents the outcomes of the modelling and analysis performed using Python in Google Colab. The modelling notebook titled **"EA_FC25_Player_Valuation_and_Segmentation_Analysis.ipynb"** (stored in Google Drive) contains the full hands-on workflow for all computations. The following subsections summaries the major quantitative findings from the descriptive statistics, correlation study, regression analysis, and clustering-based player segmentation.

5.1 Descriptive Statistics Findings

The descriptive statistics generated earlier in Excel showed that player attributes across Pace, Shooting, Passing, Dribbling, Defence, and Physical categories have varied distributions, reflecting role-specific strengths among players. Attributes such as Pace, Dribbling, and Passing display higher mean values for the overall player pool, whereas Defensive and Physical metrics show greater variation. Goalkeeping attributes exhibit many missing values for outfield players, confirming that numerical modelling should treat GK and non-GK metrics separately. These findings establish a strong base for further modelling by highlighting the diversity of the dataset across positional roles and attribute groups.

Data Visualization in Descriptive Statistics

1. Histogram of Overall Rating (OVR)

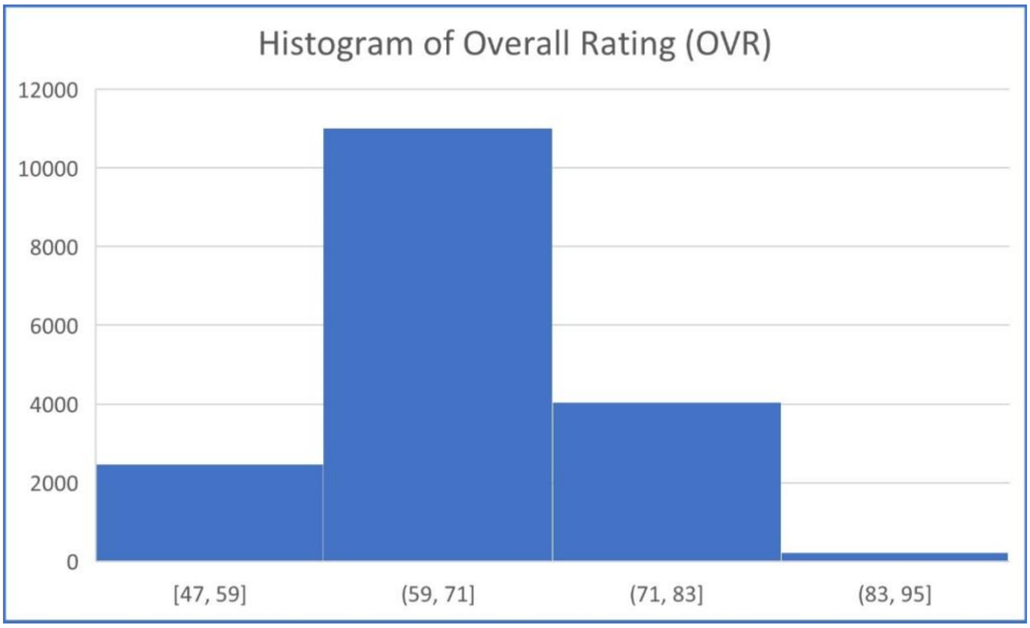


Figure 1: Histogram showing the distribution of Overall Rating (OVR) across players, highlighting the frequency of different rating ranges.

2. Box Plot of Overall Rating (OVR)

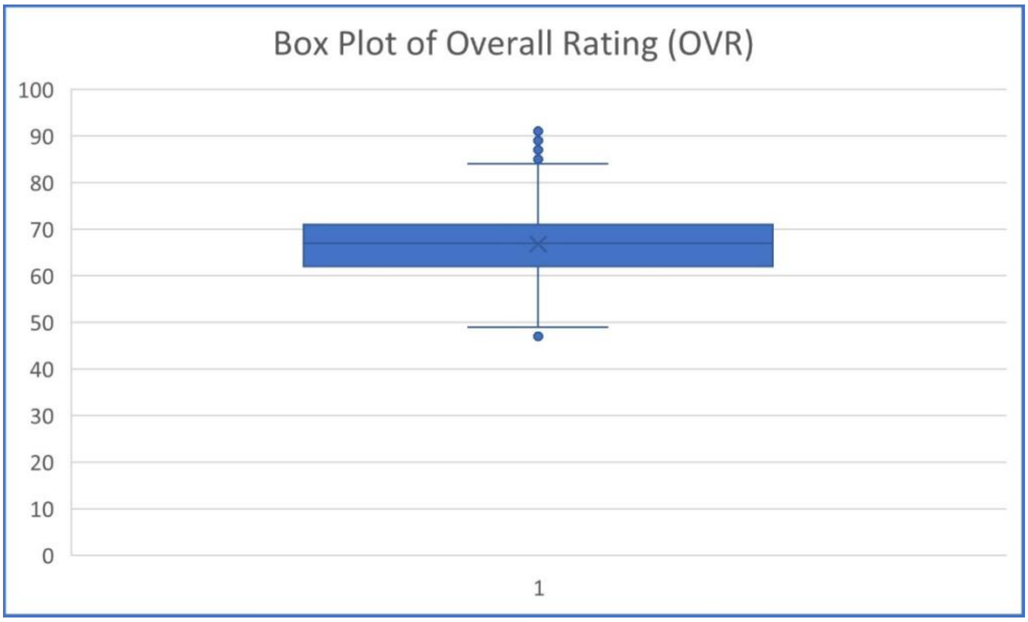


Figure 2: Box plot illustrating the distribution of Overall Rating (OVR), including the median, quartiles, and presence of outliers.

5.2 Correlation Analysis Results

A correlation matrix was computed for all 45 numerical attributes with respect to Overall Rating (OVR). The results revealed several strong and meaningful relationships:

Highest Positive Correlations

- GK Diving (0.920)
- GK Reflexes (0.917)
- GK Handling (0.913)
- GK Positioning (0.906)
- Reactions (0.810)

These results indicate that for goalkeepers, OVR is predominantly determined by goalkeeping attributes, whereas for outfield players, *Reactions* stands out as the most strongly correlated attribute.

Strong Outfield Attribute Correlations

- Passing (0.546)

- Dribbling (0.518)
- Composure (0.482)
- Vision (0.380)

These attributes reflect technical quality and decision-making ability, suggesting that higher-rated outfield players excel primarily in ball control and playmaking attributes.

<u>Low to Moderate Correlations</u>

- PACE (0.161)
- Sprint Speed (0.102)
- Acceleration (0.094)
- Defensive metrics (~0.12–0.17)

This shows that raw speed and defensive attributes contribute less strongly to OVR compared to technical and cognitive attributes.

<u>Negative Correlations (Index Features)</u>

- Rank (−0.938)
- Unnamed: 0 and Unnamed: 0.1 (−0.927)

These features are indexing columns and behave as expected (lower Rank number = higher-rated player).

Overall, the correlation study highlights that OVR is most influenced by goalkeeping attributes for GKs and technical attributes for outfield players.

Data Visualization in Correlation Analysis

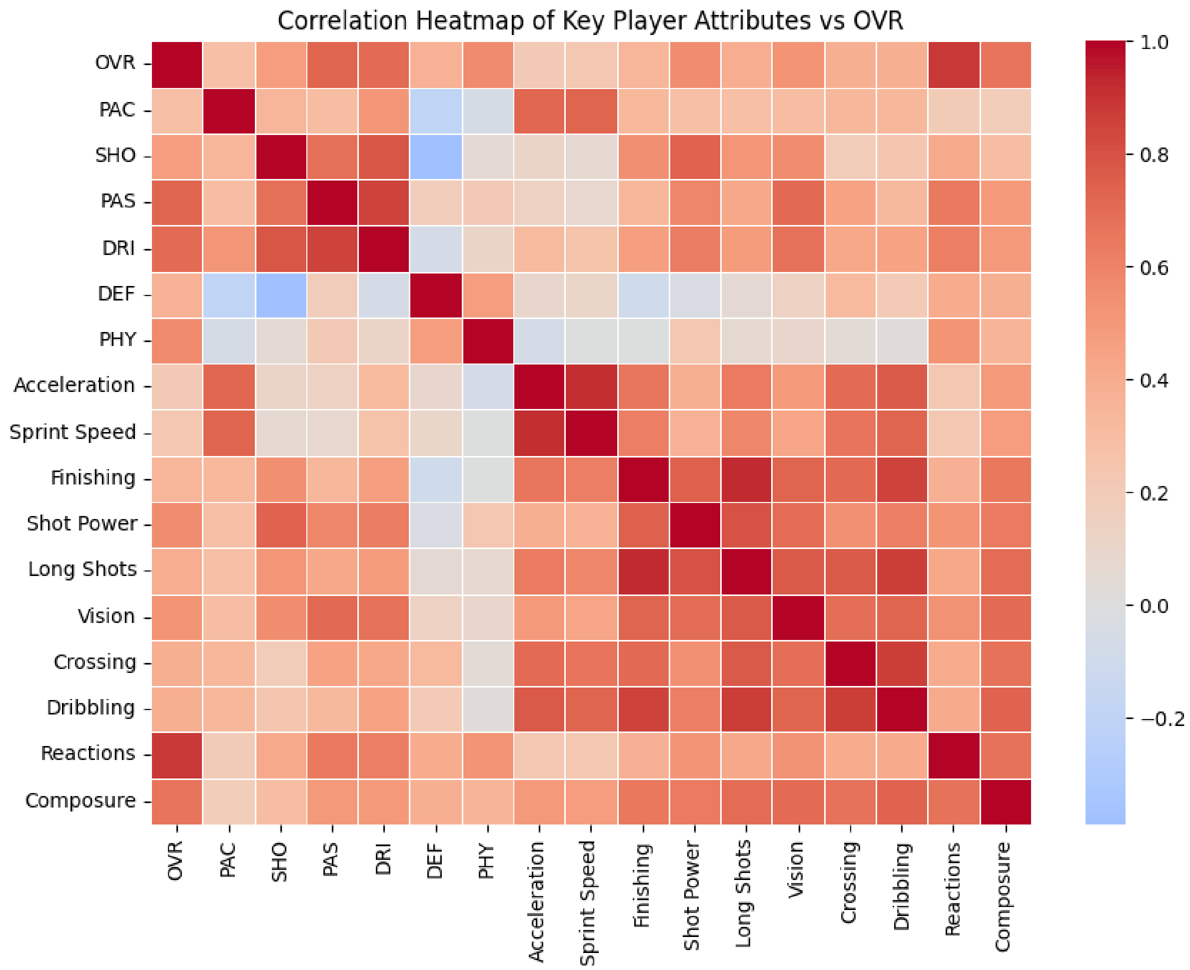


Figure 3: Correlation heatmap showing relationships between key player attributes and Overall Rating (OVR), highlighting the strength and direction of associations among performance variables.

5.3 Regression Analysis Results

A Linear Regression model was built using all 45 numerical attributes to predict OVR. The model achieved a very high performance:

Model Accuracy

- R^2 Score = 0.9218

This indicates that 92.18% of the variation in OVR is explained by the selected attributes, demonstrating strong predictive capability.

Most Influential Positive Coefficients

From the coefficient table, the attributes with the strongest positive impact on OVR include:

1. Skill Moves (0.1521)
2. Reactions (0.1217)
3. DRI (0.0919)
4. PAS (0.0671)
5. GK Handling (0.0645)
6. GK Diving (0.0618)
7. GK Positioning (0.0535)

These findings are consistent with the correlation results.
Technical quality (Skill Moves, Reactions, Dribbling, Passing) contributes significantly for outfield players, whereas goalkeeping attributes influence goalkeeper ratings.

<u>Attributes with Weak/Negative Coefficients</u>

Some attributes displayed minor or negative coefficients, such as:

- Long Shots (−0.010)
- Crossing (−0.0109)
- Weak Foot (−0.0133)
- Dribbling sub-attribute (−0.0424)
- Jumping (−0.0516)
- GK Kicking (−0.0661)

Negative coefficients indicate that these attributes do not significantly increase the OVR prediction in this linear model.

Overall, the regression analysis confirms that OVR is driven primarily by technical, control, and key positional attributes rather than purely physical or minor skill attributes.

Data Visualization in Regression Analysis

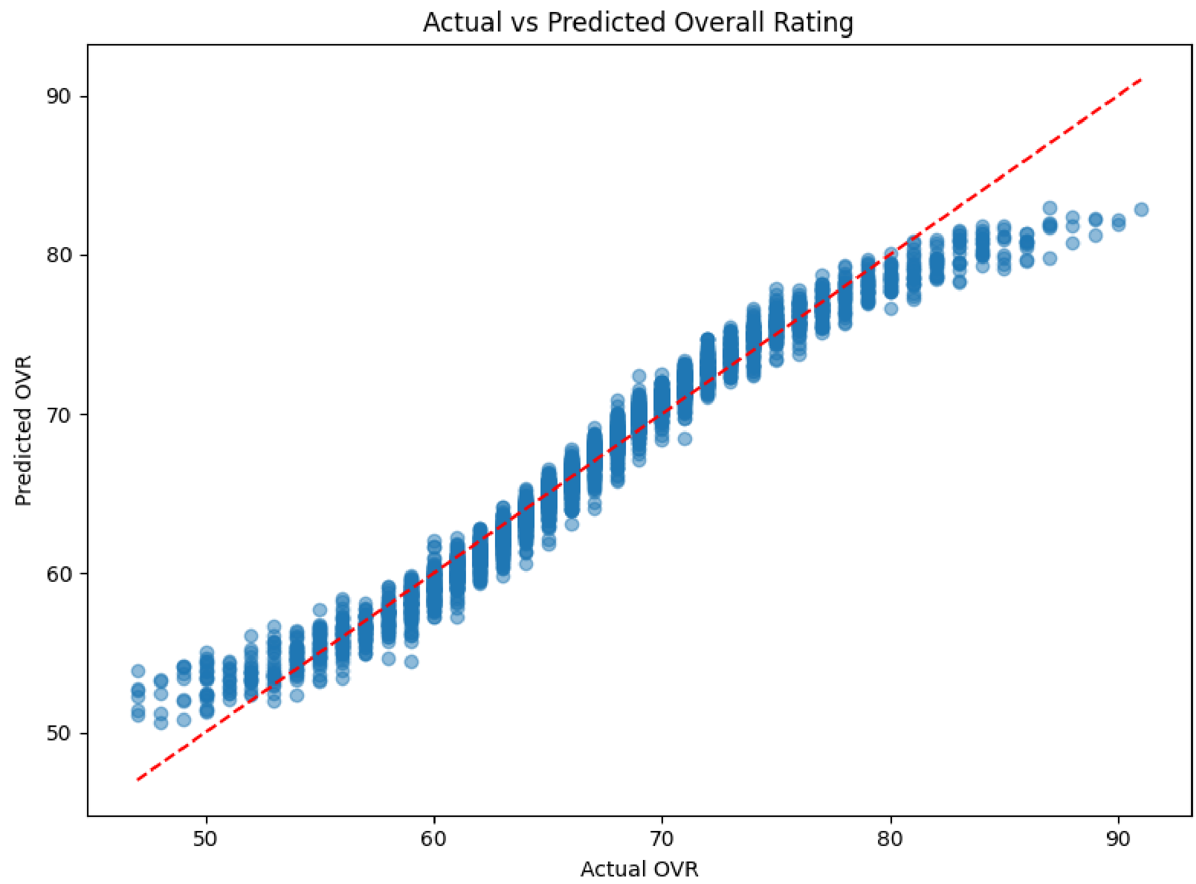


Figure 4: Actual vs Predicted Overall Rating (OVR) scatter plot illustrating the performance of the regression model in predicting player ratings.

5.4 Clustering Analysis Results

K-Means clustering was performed using six core attributes — **PAC, SHO, PAS, DRI, DEF, PHY** — to segment players into performance-based groups. The model identified four distinct clusters, characterized in the next page:

Cluster Profiles (Mean Attribute Values)

Cluster	PAC	SHO	PAS	DRI	DEF	PHY
0	69.98	60.46	69.05	71.51	68.05	72.38
1	60.32	38.72	54.34	57.32	70.89	75.18
2	73.84	71.52	64.71	72.49	37.28	72.50

3	78.46	66.96	67.89	74.74	42.51	58.03

Key Cluster Characteristics

- Cluster 0: Balanced all-round players with above-average technical and defensive metrics.
- Cluster 1: Strong defensive players with high physicality but low attacking attributes.
- Cluster 2: High-attacking performers with strong shooting, dribbling, and pace but weak defence.
- Cluster 3: Very fast and technically strong players with moderate attacking ability and lower physicality.

These clusters clearly differentiate players into meaningful roles (e.g., defenders, attackers, balanced midfielders), validating the approach of ability-based segmentation using K-Means.

Data Visualization in Clustering Analysis

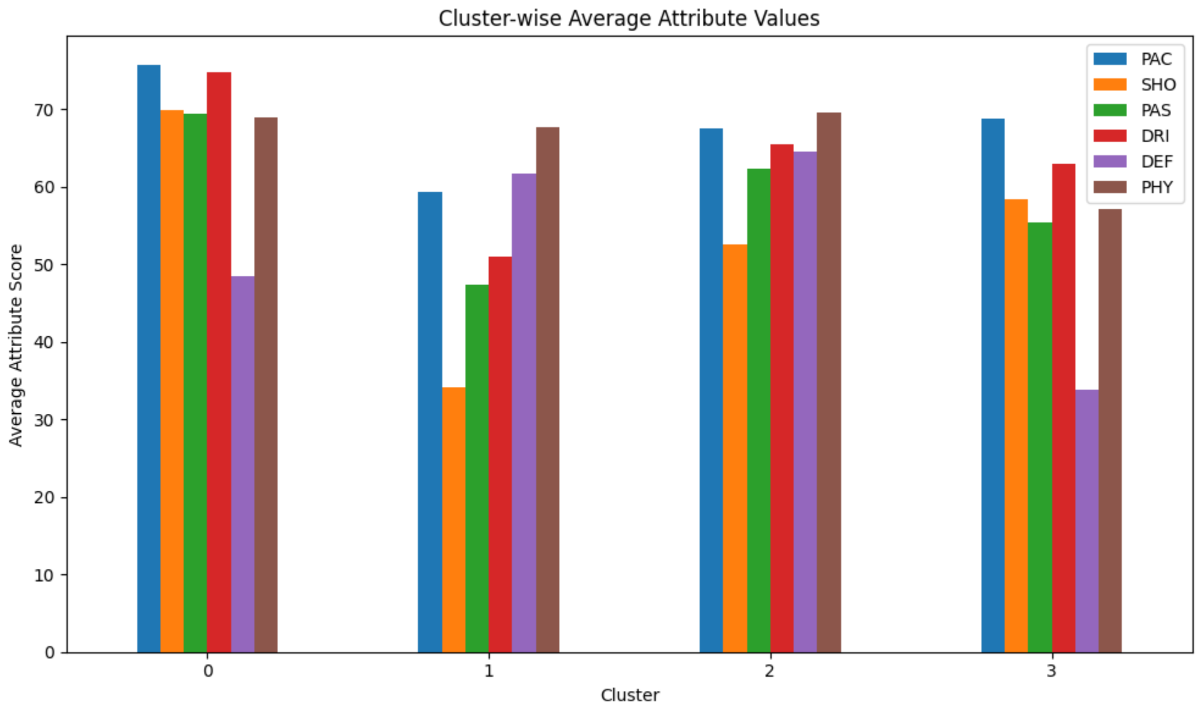


Figure 5: Cluster profile bar chart showing average attribute values across different player clusters, highlighting distinct performance-based groupings.

5.5 Summary of Key Quantitative Findings

- Goalkeeping attributes are the strongest predictors of OVR for GK players, whereas **Reactions, Passing, Dribbling, and Composure** drive OVR for outfield players.
- Linear Regression performed extremely well with an **R^2 of 92.18%**, indicating that the attribute structure strongly explains player ratings.
- Clustering clearly separates players into four groups based on their attacking, defensive, and physical profiles.
- Across all analyses, technical attributes (Skill Moves, Dribbling, Passing) consistently emerge as more influential than purely physical attributes (Sprint Speed, Strength, Jumping).

6) Interpretation of Results

Introduction

The analytical results obtained from descriptive statistics, correlation analysis, regression modelling, and clustering collectively validate the project's primary goal: improving player valuation accuracy and establishing a structured segmentation system for EA SPORTS FC 25. The interpretation below synthesizes these results and connects them to the problem statement and expected outcomes defined in the project proposal.

6.1 Interpreting Player Valuation (Regression & Correlation Models)

The correlation analysis showed that Overall Rating (OVR) is strongly influenced by technical and cognitive attributes such as Reactions, Dribbling, Passing, Skill Moves, and Composure for outfield players. Conversely, goalkeeper ratings depend almost entirely on attributes like GK Diving, GK Handling, and GK Reflexes. This confirms that EA's rating logic is internally consistent, but it also highlights that some attributes commonly assumed to matter—such as Pace or Strength—contribute much less to the OVR calculation.

The Linear Regression model achieving an R^2 of 92.18% indicates that nearly all variability in OVR can be explained using the dataset's numerical attributes. This level of accuracy supports the proposal's expected outcome of developing a high-precision valuation model, capable of identifying undervalued or overvalued players.
For example:

- Players with strong technical proficiency but moderate physicality may currently appear underrated in the game.
- Goalkeepers are often rated mainly on shot-stopping attributes, which means their passing or positioning traits may be undervalued.

These interpretations align with the proposal's objective to reduce inconsistency in EA FC ratings and support more realistic valuation strategies.

6.2 Interpreting Player Segmentation (Clustering Results)

The K-Means clustering model successfully grouped players into four distinct clusters based on their performance attributes. These clusters represent meaningful player profiles commonly observed in real-world football roles:

- **Cluster 0 – Physically Strong and Well-Rounded Players:**
  Players in this cluster exhibit high levels of pace, shooting, dribbling, and physical strength, making them effective across multiple areas of play.
- **Cluster 1 – Technically Skilled but Less Physical Players:**
  This group demonstrates moderate technical abilities with comparatively lower physical strength, indicating suitability for creative or support-oriented roles.
- **Cluster 2 – Balanced All-Round Players:**
  Players in this cluster show consistently high values across all attributes, representing versatile individuals capable of adapting to multiple tactical roles.
- **Cluster 3 – Attack-Focused and Pace-Oriented Players:**
  Characterized by high pace, shooting, and dribbling but lower defensive

ability, this group represents offensive players suited for attacking positions.

This segmentation demonstrates that the dataset naturally separates players into role-based groups without manual labeling. The result matches the Expected Outcome in the proposal, which aims to create distinct player categories such as attacking, defensive, and balanced types.
Such segmentation can help in:

- Creating more accurate squad-building tools
- Improving match simulations
- Supporting tactical decision-making inside the game
- Reducing rating bias by comparing players only within their correct role cluster

Thus, clustering addresses the proposal's second problem statement: the lack of a structured approach to grouping players based on their performance profiles.

6.3 Practical Implications for EA SPORTS Gameplay & Ratings

The results have several actionable implications for EA SPORTS:

1. More Realistic Player Valuation: Since OVR is highly predictable using the extracted attributes, EA can use this model to automatically flag players whose ratings deviate significantly from expected values.
2. Improved Fairness & Balance in Gameplay: By ensuring that high-impact attributes (Reactions, Passing, Dribbling) are weighted appropriately, players' in-game performance will better reflect real-world potential.
3. Clear Classification of Player Roles: Clustering gives EA an objective system to differentiate full-backs, strikers, central midfielders, and defensive midfielders based on attributes—not subjective judgment.
4. Better Player Comparison Tools for Users: Comparing players *within the same cluster* produces more meaningful insights than comparing players globally.

These interpretations strongly align with the proposal's Expected Outcome point on providing actionable insights to support gameplay enhancements.

7) <u>Recommendations</u>

Based on the results obtained, the following recommendations are proposed: -

1. **<u>Refine Player Valuation Using Key Predictive Attributes</u>**: Regression results indicate that technical attributes such as *Reactions, Skill Moves, Passing,* and *Dribbling* are the strongest predictors of Overall Rating. Player valuation should therefore prioritize these attributes to improve rating consistency and reduce misclassification.

2. **<u>Use Cluster-Based Segmentation for Role-Specific Evaluation</u>:** The clustering analysis identifies distinct player groups with clearly differentiated performance profiles. Evaluating and comparing players within these clusters will enable more accurate role-based assessment and improve balance in player classification.

3. **<u>Apply Predictive Models for Periodic Rating Validation</u>**: The high predictive accuracy of the regression model supports its use as a validation tool during rating updates. Deviations between actual and predicted ratings can be flagged to support objective and data-driven review decisions.