SCHOOL OF MANAGEMENT

REAL WORLD ANALYTICS PROJECT

Title: FIFA Player Data Insights: An Interactive Approach with Dashboards and Machine Learning Models

GROUP 2

AGATHIYAN K (045005)

HRITIKA GUPTA (045020)

MOHIT VAIDYA (045032)

NANDINI ARYA (045034)

This dataset contains detailed information about football players from a FIFA series, encompassing various attributes related to player performance, club affiliations, and personal characteristics. The data spans multiple versions of the game, tracking updates and changes over time. It includes metrics related to player skills, physical attributes, and playing positions across different leagues and clubs.

DATA SOURCE

Kaggle
https://www.kaggle.com/datasets/sabir0000/male-football-players-data

ABOUT DATA

25000 X 110
Categorical : 23
Non categorical : 86

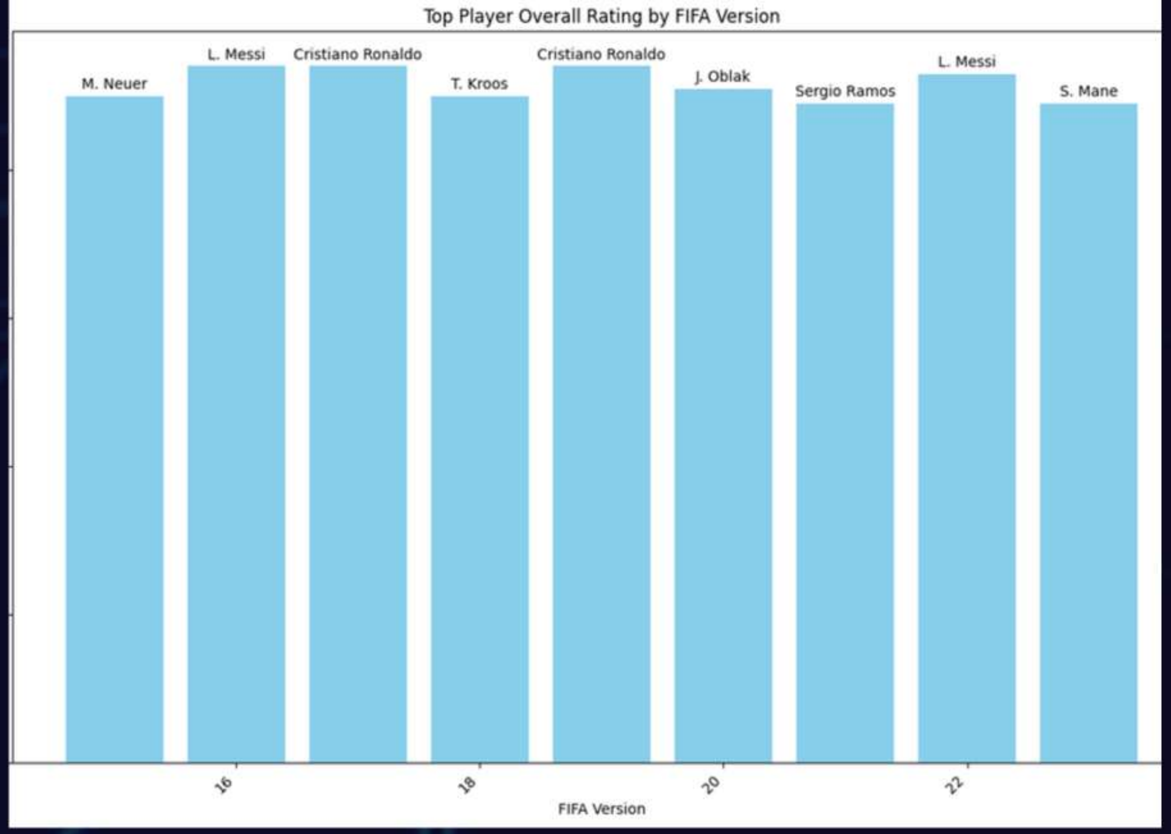
KEY COLUMN ATTRIBUTES

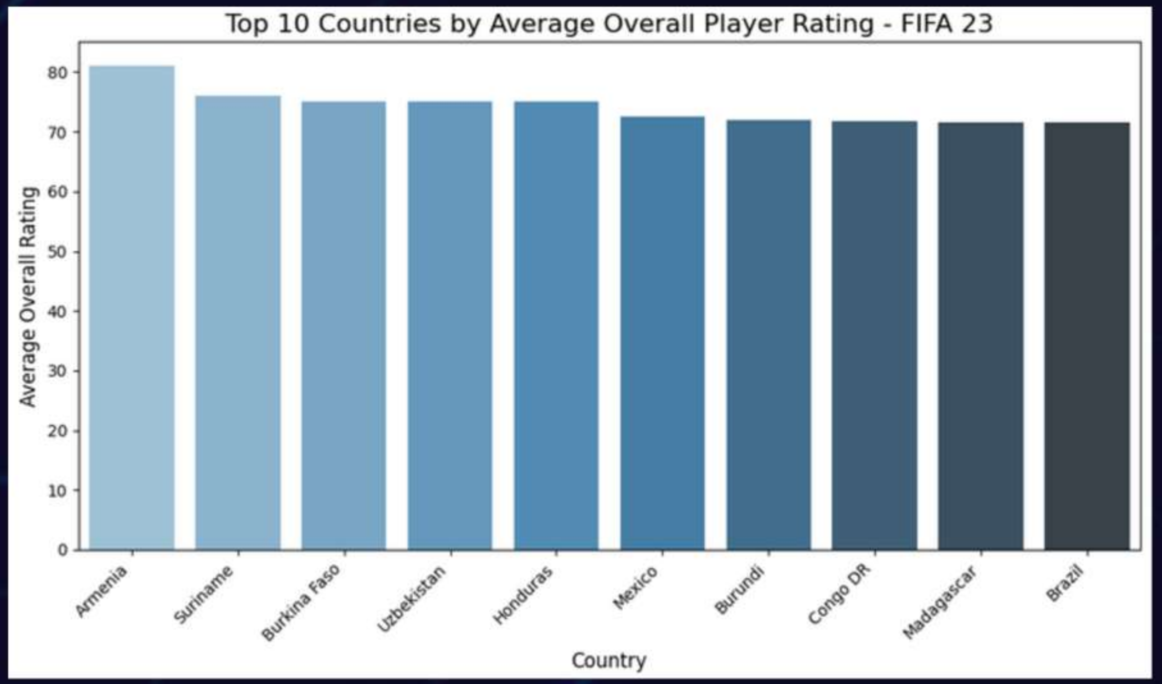
- Basic Information
- Club and League Information
- Player Attributes and Performance Metrics
- Detailed Skill Attributes
- Movement and Physical Attributes
- Mentality and Defending Attributes
- Goalkeeping Attributes
- Position-Specific Attributes

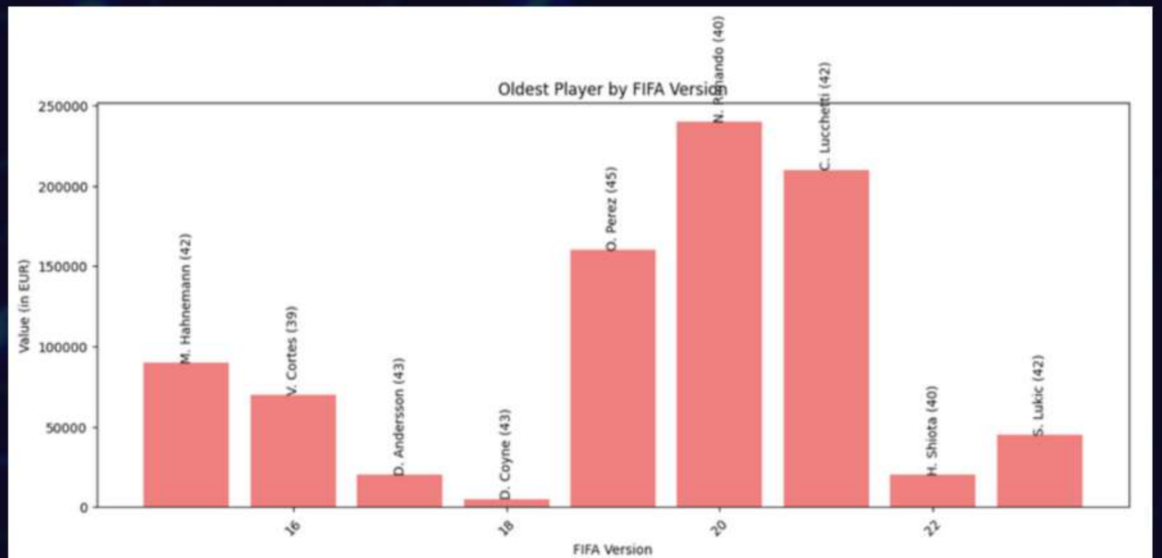
ABOUT THE DATA SET

ISSUE WITH THE DATA

- 23% of the variables have null values.and 6 columns have more than 50% values missing.
- All rows have at least 1 missing value.
- Special Characters in Names
- Numeric Values with Additional Symbols







FORE SCHOOL OF MANAGEMENT

OBJECTIVES OF THE ANALYSIS: ANALYTICAL DASHBOARD

01 Evaluate Wage Disparities Among Top 3 Footballing Nations

02 Assess Correlation Between Player Rankings and Market Value

03 Investigate Mean Pace Variations Across Player Positions

04 Examine Differences in Player Value and Wage Across Positions

05 Evaluate Influence of Key Variables on Player Valuation

06 Evaluate Impact of Variables on Overall Player Rating

07 Identify Top Player by Position Based on Wages

08 Compare Players Based on Tags/Skillsets:

09 Predicting the value of a player based on their career stats

10 Comparing the player stats based on their Position and Clubs

SCHOOL OF MANAGEMENT

OBJECTIVES OF THE ANALYSIS:

ML MODEL

Perform Feature Reduction:

Conduct correlation analysis between columns to reduce the number of factors, streamlining the clustering process by eliminating redundant features.

Optimize Cluster Selection:

Determine the optimal number of clusters by evaluating Davies-Bouldin (DB) and Silhouette (SS) scores, ensuring a well-balanced intra-cluster cohesion and inter-cluster separation.

In-Depth Cluster Profiling:

Analyze and profile each cluster based on key parameters, offering detailed insights into the distinct characteristics and roles of the players.

Cluster Naming and Classification:

Assign meaningful names to each cluster, reflecting their dominant traits and providing a clear understanding of their defining features.

OVERVIEW OF THE ANALYTICAL DASHBOARD

The Dashboard is a web-based tool for analyzing and visualizing player data from different FIFA game versions, allowing users to explore statistics, compare performance, and predict player values interactively.

KEY COMPONENTS

01 Refine data analysis with options for FIFA version, player position, league, club, wage range, and player tags.

02 Includes bar charts, treemaps, and scatter plots to represent player data effectively.

03 Displays detailed metrics such as player value, potential, age, and other relevant attributes.

04 Incorporates a predictive model to estimate player values based on selected attributes.

STREAMLIT LINK

Functionality

01 Used ANOVA and Tukey HSD to test and identify differences in player attributes across various groups.

02 Conducts Pearson and Spearman correlations to measure the strength and direction of relationships between variables like player value and overall rating, helping identify key predictors.

03 Integrates predictive modeling to estimate player values based on selected attributes, enhancing insights and accuracy.

04 Uses linear regression to predict player values from multiple attributes, with hypothesis testing to assess the significance of regression coefficients and improve accuracy.

OVERVIEW OF THE ML MODEL

Initially, we identify the optimal score and DB (Davies-Bouldin) score to assess model performance. Following this, we apply a clustering algorithm to segment the data, creating distinct profiles based on the specialisation.

01 The model performs clustering to group players into distinct categories based on various attributes such as age, position, and skill metrics

02 The selection of clusters is guided by evaluating both the Davies-Bouldin (DB) score and Silhouette (SS) score, to have optimal cluster

03 Correlation analysis identifies and eliminates redundant features, focusing on the most impactful variables to improve model efficiency

04 Clusters are analyzed and profiled based on key parameters to understand player characteristics, informing decisions on team composition and player acquisitions

OBSERVATION FOR ANALYTICAL DASHBOARD

Value Trends:

- Portugal (FIFA 17) and Egypt (FIFA 18) had high maximum average player values, €5.16M and €17.45M, respectively.
- Armenia showed substantial growth, with its maximum value increasing from €6M (FIFA 15) to €16M (FIFA 23).
- Burundi (FIFA 15) had the lowest minimum average value at €35K, followed by Ethiopia (FIFA 23) and Bermuda (FIFA 18) at €70K.

Ranking:

- Armenia (FIFA 23) has an average rating of 81, outperforming other listed countries.
- Brazil’s rating (72) is consistent with teams like Burundi and Congo DR.

Top Players (FIFA 23):

- Sadio Mané, Neymar Jr., and Ederson are top players.
- Manchester City dominates with Ederson, Bernardo Silva, and Ruben Dias.
- Top players come from Senegal, Brazil, and Portugal across various positions.

Youngest and Oldest Players:

- **Youngest:** Market values increased significantly, e.g., Gavi (FIFA 22) valued at €2.1M.
- **Oldest:** Players (aged 39-45) are goalkeepers, with decreasing market values as they age.

Player Compensation:

- Messi (FIFA 16, 22) and Ronaldo (FIFA 17, 19) consistently top player wages.
- Wages vary widely, e.g., Messi (FIFA 16) at €550K vs. Mané (FIFA 23) at €145K.

Inferential Statistics:

- **Wages Comparison:** No significant difference in average wages among players from top 3 nations Armenia, Suriname, and Burkina Faso.
- **Player Rankings vs. Market Value:** Significant correlation between player rankings and market value.
- **Pace Differences:** Significant differences in mean pace across player positions, except for similar pace attributes between CAM vs. ST, CF vs. LB, CF vs. LWB, CF vs. RB, CF vs. RWB, LM vs. RM, LW vs. RW, LW vs. RM, LWB vs. RB, LWB vs. RWB, RB vs. RWB, and RM vs. RWB.
- Significant differences in mean value across different player positions.

Regression Analysis:

- **Key Variables:** Wage, international reputation, and skill moves have a significant positive impact on player valuation.
- **Positional Influence:** Positions like GK, CF, and CDM show a strong positive effect on player value, while LB, RB, and RM are not statistically significant.
- **Overall Rating:** Independent variables such as age, potential, value_eur, wage_eur, preferred_foot, skill_moves, work_rate, and weight_kg significantly affect the overall player rating.

FORE SCHOOL OF MANAGEMENT

OBSERVATIONS FROM ML MODEL

CLUSTER 0: THE BALANCED DEFENDERS 2016

- Age: Primarily 26 years, players in their prime.
- Attributes: Slightly higher heading accuracy (58), balanced defending (61), low goalkeeping attributes.
- Physical Traits: Average height (180 cm), "Normal" body type.
- Representation: Strong presence in Serie A, predominantly English, right-footed.
- Position: Center Backs (CB) with a "Medium/Medium" work rate, contributing to both defense and playmaking
- Predicted Value (EUR): 16,912,814.16.

CLUSTER 1: THE SEASONED PLAYMAKERS 6486

- Age: Predominantly 24 years old, balancing youth and maturity.
- Attributes: Moderate heading accuracy (55), solid defending (63), low goalkeeping skills.
- Physical Traits: Average height (180 cm), "Normal" body type.
- Representation: Mostly from Belgium and the Super League, with a preference for the right foot.
- Position: Primarily Right Wingers (RW) with a "Medium/Medium" work rate.
- Predicted Value (EUR): 17,155,064.42

CLUSTER 2: THE EMERGING GOALKEEPERS 6471

- Age: Mainly 20 years, indicating young talents.
- Attributes: Stronger heading accuracy (64), respectable defending (62), developing goalkeeping skills.
- Physical Traits: Taller (185 cm), "Normal (185+)" body type.
- Representation: Liga Professional, predominantly Brazilian, left-footed.
- Position: Primarily goalkeepers or versatile defenders.
- Predicted Value (EUR): 19,234,580.86

CLUSTER 3: THE TACTICAL MASTERS 5796

- Age: Predominantly 20 years, with high potential.
- Attributes: Focus on technical skills, lower defending (28), diverse goalkeeping attributes.
- Physical Traits: Slightly shorter (175 cm), "Normal" body type.
- Representation: Liga Professional, mainly Uruguayan, left-footed.
- Position: Central Defensive Midfielders (CDM) with a "Medium/Medium" work rate, excelling in tactical roles.
- Predicted Value (EUR): 18,363,582.68

FORE SCHOOL OF MANAGEMENT

MANAGERIAL INSIGHTS - ANALYTICS DASHBOARD

Goalkeepers generally tend to survive longer in professional football than other players

Smaller nations like Burundi and Ethiopia consistently have low player valuations, highlighting the financial gap in global football.

Lower-tier leagues, such as the Premier Division and League Two, have the lowest average player ratings

The high valuation of players in key tactical positions (GK, CAM, CM) highlights the importance of building a strong team core.

Identifying undervalued players in emerging markets and lower leagues can be a strategic move for clubs aiming to compete with top-tier teams.

Player ratings and values fluctuate across FIFA versions, reflecting changes in player performance and market perceptions. Monitoring these trends can guide long-term player investments.

Players in lower-rated leagues may be undervalued gems, offering potential for transfer to higher leagues at a lower initial cost.

FORE SCHOOL OF MANAGEMENT

MANAGERIAL INSIGHTS - ML MODELS

01 Top clubs consistently secure high-value players, maintaining their competitive edge across multiple FIFA versions.

02 Player wages, especially for key positions, have significantly increased, reflecting growing financial demands in football.

03 Central positions are crucial, with top players often occupying CM, CB, and ST roles, reinforcing a strong team core.

04 Younger stars like Haaland are becoming more prominent, indicating a shift towards future potential over established players.

05 The rise of versatile roles like RWB shows teams' adaptation to evolving tactics and the need for adaptable players.

06 Managers are confident in investing in young players who play in defending positions, while tending to have wider options for important positions like CF and CDM

FUE SCHOOL OF MANAGEMENT

THANK YOU