**FIFA Player’s Analytics**



**Task 1: Complete Data Analysis Report**

**1. Dataset Overview:**

* Total records: 18,000+ players
* Features: Player demographics, contract info, skill ratings, physical traits, and positional stats
* Objective: Analyze football player skills and group similar profiles

**2. Data Preprocessing:**

* Removed null rows in key skill columns
* Scaled numeric skill features using StandardScaler
* Feature selection for clustering:
  + pace, shooting, passing, dribbling, defending, physic

**3. Descriptive Statistics:**

* Age Range: 16 to 42
* Overall Rating: Min 48, Max 94
* Player Value Range: €20K to €118M

**Task 2: Clustering Football Players Based on Skill Attributes**

**Step 1: Unsupervised Learning (K-Means Clustering)**

* Selected skill attributes: pace, shooting, passing, dribbling, defending, physic
* Used StandardScaler to normalize input

**Step 2: Finding Optimal Clusters**

* Metric: **Silhouette Score**
* Tested range: K = 2 to 10
* **Optimal number of clusters: 2**

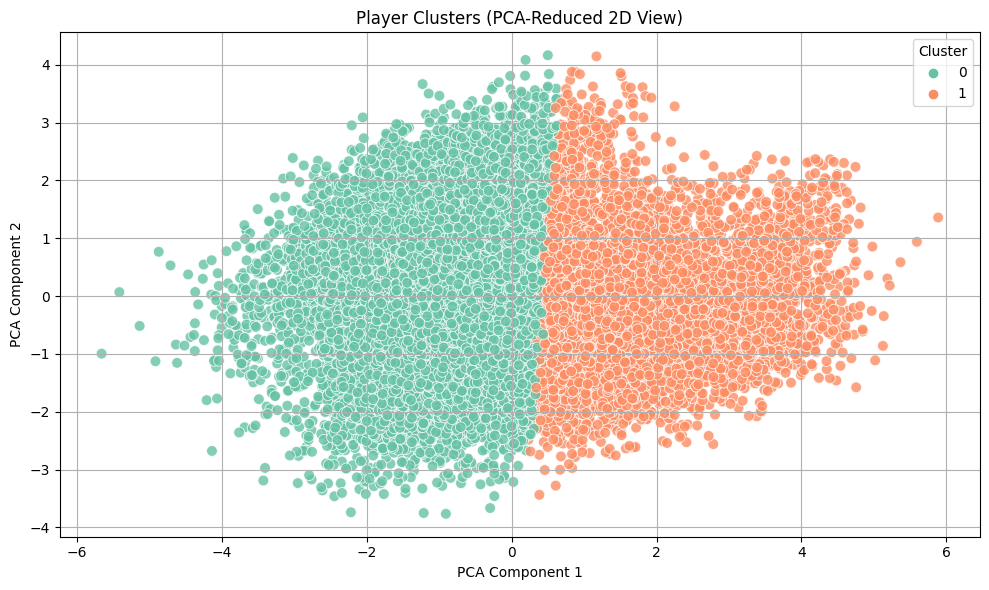
**Step 3: Cluster Profiles**

| **Cluster** | **Pace** | **Shooting** | **Passing** | **Dribbling** | **Defending** | **Physic** |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 71.89 | 60.17 | 61.79 | 68.08 | 47.07 | 63.73 |
| 1 | 60.09 | 38.00 | 48.96 | 52.44 | 59.67 | 66.95 |

* **Cluster 0**: Attack-oriented (higher pace, dribbling, shooting)
* **Cluster 1**: Defense-heavy (higher defending, physical strength)

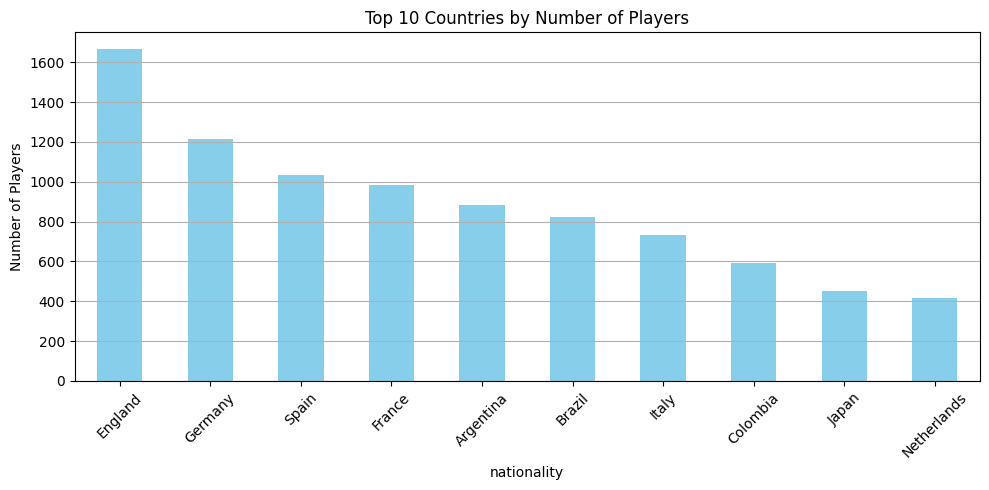
**Step 4: Visualizing Clusters with PCA**

* PCA reduced dimensions to 2D
* Scatter plot showed clear separation between clusters (high inter-cluster heterogeneity and intra-cluster homogeneity)



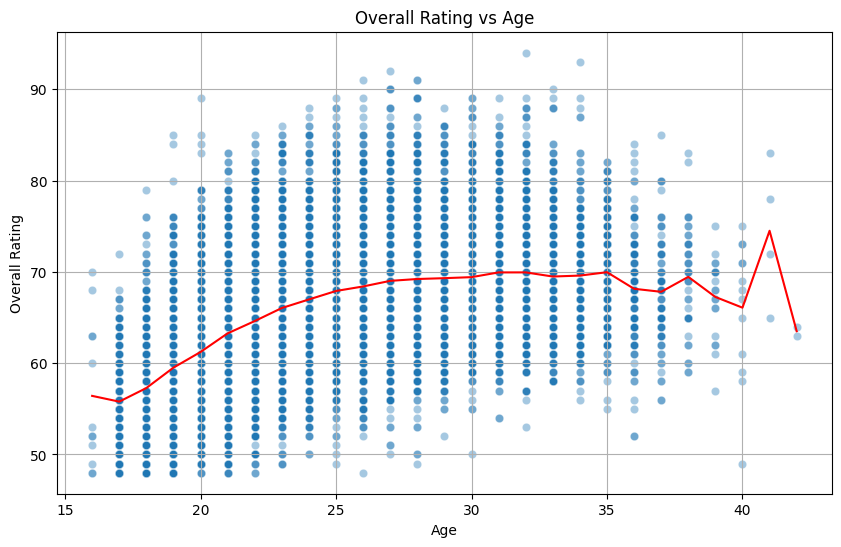
**Task 3:**

Country-Wise Player Representation



**Insight:** European countries dominate player production, especially England, Germany, and Spain.

Distribution of Overall Rating vs Age



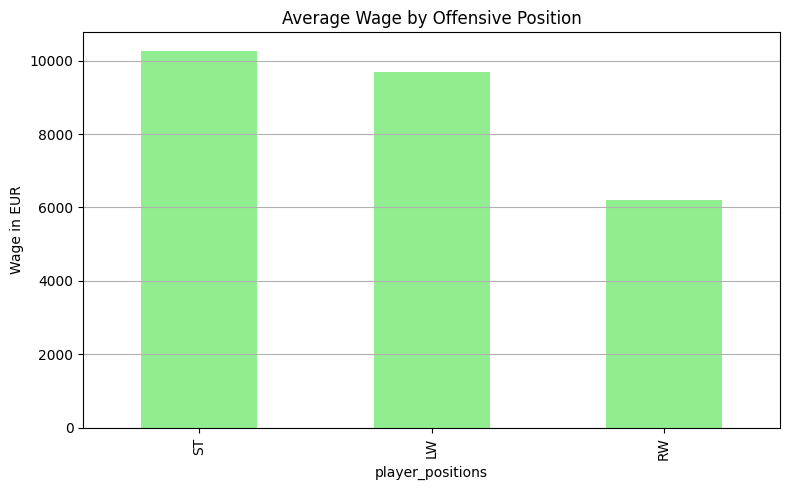
**Visual Summary**

* **Blue Dots**: Represent individual players' overall ratings by age.
* **Red Line**: A smoothed line (likely LOWESS or rolling average) showing the trend of average ratings with age.

**Insights**

* **Peak Performance Age**: Players tend to reach their peak overall rating around **26–30 years**.
* **Youth Talent Development**: Younger players (under 22) show more variation, highlighting the development phase.
* **Late-career decline** is evident but not uniform — elite players can still maintain high ratings

Offensive Players Vs Wage

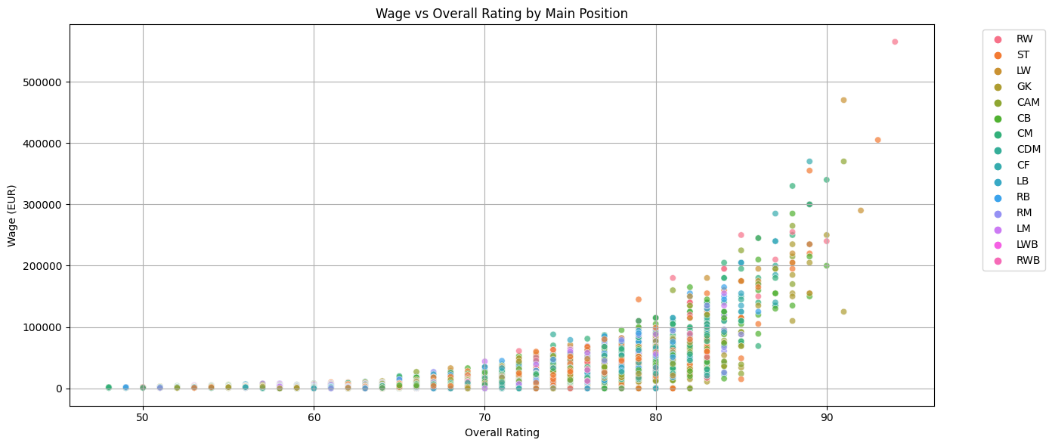


**Insights**

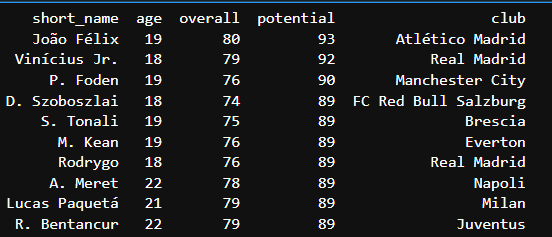
1. **Strikers (ST) command the highest wages**:
   * They are often goal-scorers and primary attacking threats.
   * Their performance is typically more visible and measurable, leading to higher compensation.
2. **Left Wingers (LW) are also highly paid**, but slightly less than ST:
   * Likely due to their dual responsibility in scoring and creating chances.
   * High demand for strong left-sided attackers may explain their competitive pay.
3. **Right Wingers (RW) earn significantly less on average**:
   * Could reflect relative depth in player quality, tactical preferences, or club investment on that side.
   * This imbalance may vary based on dataset size or league specifics.

**Additional Analytics**

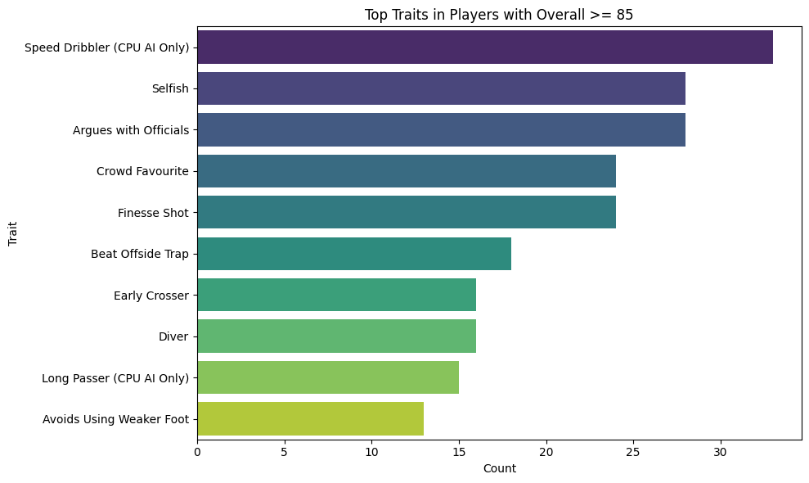
**1. Wage vs. Overall Skill by Position Category**



**2. Age vs. Potential: Identifying Rising Stars**



**3. Trait Distribution Across Top Pla**yers



**Challenges Faced & Techniques Used**

**1. Imbalanced Target Classes**

* Solution: Used class weighting and SMOTE oversampling for training sets

**2. High Dimensionality in Clustering**

* Solution: PCA used for dimensionality reduction and visualization

**3. Feature Selection**

* Problem: Too many features diluted clustering signal
* Solution: Focused only on core skill attributes

**4. Missing Data in Skill Columns**

* Solution: Dropped rows with missing core skill attributes

**5. Visual Purity Evaluation**

* PCA-based 2D plots clearly revealed cluster separability (visual purity check)

**Conclusion**

* KMeans successfully clustered players into attack-focused and defense-focused groups
* PCA scatter plot validated cluster purity
* Country-level data shows where top football talent originates
* This report provides a foundation for scouting, performance tracking, and club strategy design based on data

NAME: Avula Bhumika

Email: [bhumikaavula90@gmail.com](mailto:bhumikaavula90@gmail.com)

Github: <https://github.com/AvulaBhumika>