

Description of the Football Player Dataset Analysis for Potential Advanced Statistical Techniques

Grp 14 : MESKI - BEN YAALA - LABIDI

1. Dataset Overview

The dataset we selected from [Kaggle](#) contains comprehensive information about football players, clubs, competitions, games, and transfer events. It features a mix of numerical and categorical variables, facilitating a range of analyses from descriptive statistics to advanced techniques.

Key Files in the Dataset:

- **players.csv** (23 columns): Offers player data and market values, allowing tracking player statistics and analysing performance trends.
- **transfers.csv** (10 columns): Details player transfers, useful for analysing transfer trends, market values, and player movements.
- **game_lineups.csv** (10 columns): Provides team lineups, aiding studies on player impacts especially captains.
- **competitions.csv** (11 columns): Information on competitions to rank player and club performances across different tournaments.
- **games.csv** (23 columns): Contains match details, supporting analysis of outcomes.
- **game_events.csv** (10 columns): Records in-game events for detailed match analysis .
- **appearances.csv** (13 columns): Tracks player appearances and statistics which helps monitor player contributions and evaluate their performances.
- **clubs.csv** (17 columns): Information about clubs, beneficial for clubs performance and valuation analysis.
- **club_games.csv** (11 columns): : Details matches played by clubs, including results and player performances which facilitates performance analysis of clubs in specific competitions
- **player_valuations.csv** (5 columns): Records market value assessments over time which helps track fluctuations in player values and market trends.
- **sample_data**: Contains test data to understand dataset structure.

These files collectively provide extensive insights for analysing player and club performance in football supporting a variety of statistical analyses and research applications.

2. Potential Analytical Techniques

- **Principal Component Analysis (PCA):**
Reduces dataset dimensionality while preserving variance. For example, PCA can analyse metrics like shooting accuracy and passing success to identify key performance drivers.
- **Canonical Correlation Analysis (CCA):**
Explores relationships between two sets of variables , making it perfect for analysing connections between performance metrics (e.g., goals, assists) and fitness indicators (e.g., injuries, playing time). It can also help reveal how these factors influence market value, suggesting that consistent performance correlates with higher valuation offering valuable insights for player valuation in the transfer market.
- **Factorial Correspondence Analysis (FCA):**
Visualizes relationships between categorical variables. For instance, it can highlight how midfielders from specific leagues contribute more assists than others.
- **Clustering Techniques:**
Groups players based on performance metrics and characteristics. Techniques like k-means can classify players into tiers (elite, average, underperformers), aiding strategic decisions in recruitment and training based on identified player archetypes.

3. Key Benefits of the Dataset

- **Diverse Data:** A mix of quantitative and categorical variables supports various analyses.
- **Performance vs. Market Value:** Enables exploration of how success influences player valuation.
- **Comparative Studies:** Facilitates comparisons across leagues and clubs.
- **Research Versatility:** Suitable for analyses ranging from basic statistics to advanced techniques.

Conclusion :

The football player dataset from Kaggle presents significant opportunities for statistical analysis. Techniques like PCA, CCA, FCA, and clustering enhance understanding of player performance and market dynamics. Insights gained can inform strategic decisions in player recruitment and team development, making this dataset invaluable for football analytics.