

ACADEMIA DE STUDII ECONOMICE BUCUREȘTI
FACULTATEA DE CIBERNETICĂ STATISTICĂ ȘI
INFORMATICĂ ECONOMICĂ



PROJECT

*Unveiling Performance Patterns in Football Teams
through Data Analysis*

Titular disciplină:

Prof. univ.dr **Vințe Claudiu**

Realizat de:

Zaremba Daniel,

Spunei Artur &

Rață Ștefania

București, 2022

Contents

Introduction.....	3
Principal Component Analysis (PCA)	3
HCA (Hierarchical Cluster Analysis)	3
Description of the variables	4
PCA	5
PC1: "Offensive and Possession Efficiency" Component	5
PC2: "Discipline and Defensive" Component	5
HCA.....	6
Cluster 1: "Offensive Excellence"	6
Cluster 2: "Defensive Focus and Discipline"	6
Cluster 3: "Balanced Performance"	6
Cluster 4: "Defensive Strategy with Disciplinary Actions"	6
Cluster 5: "Miscellaneous Group"	6
Summary	7
Data Source	7

Introduction

In the realm of sports, understanding the nuanced dynamics of football team performance is crucial for both enthusiasts and analysts. This project embarks on a journey to delve into the intricacies of various football teams by analyzing key performance variables. The dataset encompasses essential metrics such as goals, shots, yellow and red cards, possession percentage, passing accuracy, aerial duels won, and overall team rating. To extract meaningful insights, we leverage two robust data analysis methods: Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA).

Principal Component Analysis (PCA)

Principal Component Analysis is a potent method employed for dimensionality reduction, an imperative step when dealing with datasets rich in variables. The motivation behind opting for PCA lies in its ability to discern the pivotal components that wield significant influence over the overall variance within the data. By transforming the original variables into uncorrelated principal components, PCA unravels the essence of team performance, spotlighting the key contributors that shape the competitive landscape.

Motivation: The dataset has multiple variables, and PCA can help identify the most significant variables that contribute the most to the variance in the data.

Explain that PCA transforms the original variables into a new set of uncorrelated variables (principal components) that capture the maximum variance in the data.

HCA (Hierarchical Cluster Analysis)

Complementing the dimensionality reduction prowess of PCA, Hierarchical Cluster Analysis focuses on revealing the interconnected narratives among football teams. Motivated by the desire to identify underlying patterns and similarities, HCA creates a hierarchical structure of clusters. This method enables us to categorize teams based on their performance variables, thereby unraveling latent relationships and clustering teams with similar playing styles or strategies.

Motivation: To identify patterns or similarities between football teams based on their performance variables.

HCA creates a hierarchy of clusters, making it useful for understanding the relationships between different teams.

Description of the variables

The dataset aligns with the specific research questions or objectives of the analysis. The research aims to investigate the performance trends of football teams in a particular league or tournament.

Link: <https://www.kaggle.com/datasets/varpit94/football-teams-rankings-stats>

1. Team:

Serves as the identifier for each team in the dataset.

2. Tournament:

Provides context to the performance metrics by specifying the competition.

3. Goals:

Fundamental metric reflecting offensive prowess and goal-scoring efficiency.

4. Shots:

Indicates the team's offensive activity and shooting accuracy.

5. Yellow Cards:

Reflects the team's discipline and potential vulnerabilities in aggressive play.

6. Red Cards:

Signifies severe disciplinary issues and potential impact on team performance.

7. Possession%:

Illustrates a team's control over the game and playing style.

8. Pass%:

Reflects the team's ability to maintain possession and execute strategic plays.

9. Aerials Won:

Indicates strength in aerial challenges and ability to dominate in the air.

10. Rating:

Provides a consolidated measure of the team's overall performance.

PCA

Interpreting Principal Components (PCA) involves understanding the patterns of high factor loadings for each variable within a given component. Factor loadings indicate the strength and direction of the relationship between the original variables and the principal component. Let's elaborate on the provided interpretation for PC1 and PC2:

PC1: "Offensive and Possession Efficiency" Component

Strong Negative Loadings:

"Goals," "Shots pg," "Possession%," and "Pass%."

Interpretation: Teams with higher values in PC1 are characterized by lower goal-scoring rates, fewer shots per game, reduced possession percentage, and lower pass completion. This suggests an emphasis on defensive efficiency and a more cautious offensive approach.

Strong Positive Loadings:

"AerialsWon"

Interpretation: Higher success in aerial duels is a distinguishing feature of teams with elevated PC1 values. This suggests that, despite a more conservative offensive strategy, these teams excel in aerial challenges, emphasizing a robust defensive and aerial presence.

PC2: "Discipline and Defensive" Component

Positive Loadings:

"Tournament," "yellow_cards," "red_cards," and "Rating."

Interpretation: Teams with higher values in PC2 are associated with specific tournaments, indicating a tournament-specific influence. Additionally, higher counts of yellow and red cards imply a more disciplined yet potentially defensive playing style. The positive loading for "Rating" suggests that teams in this component might be rated higher overall.

Negative Loadings:

"Goals," "Shots pg," "Possession%," and "Pass%."

Interpretation: Lower goal-scoring rates, fewer shots per game, reduced possession percentage, and lower pass completion are indicative of a defensive orientation. Teams with higher PC2 values prioritize defensive stability over offensive aggressiveness.

HCA

Cluster 1: "Offensive Excellence"

Characteristics: Higher values in "Goals," "Shots pg," "Possession%," and "Pass%."

Interpretation: Teams in this cluster excel in offensive performance, showcasing a possession-oriented style of play. An exemplar for this cluster could be Barcelona, renowned for their emphasis on ball possession and attacking prowess.

Cluster 2: "Defensive Focus and Discipline"

Characteristics: Higher values in "Tournament," "yellow_cards," "red_cards," and "Rating."

Interpretation: Teams in this cluster prioritize defensive play and discipline. This may include teams with strong defensive strategies, potentially emphasizing physical play. The higher rating suggests a well-rounded and disciplined approach to the game.

Cluster 3: "Balanced Performance"

Characteristics: A mix of moderate offensive and defensive characteristics.

Interpretation: Teams in this cluster strike a balance between offensive and defensive aspects of the game. They may not excel in extreme metrics but exhibit a well-rounded performance.

Cluster 4: "Defensive Strategy with Disciplinary Actions"

Characteristics: Low offensive performance and high disciplinary actions.

Interpretation: Teams in this cluster rely on disciplined defensive strategies and set pieces for scoring. The emphasis is on a well-organized defense, and scoring may come from strategic situations.

Cluster 5: "Miscellaneous Group"

Characteristics: Varying characteristics, representing a diverse group.

Interpretation: Teams in this cluster exhibit a wide range of playing styles and performance metrics. This cluster is a mix of teams with diverse strategies and approaches to the game.

Summary

A complex picture has been revealed through the use of Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) to analyze football team performance. PCA spotlighted crucial components, displaying offensive and possession efficiency (PC1) and a combination of discipline and defensive methods (PC2). When combined with HCA clusters, these elements highlight the variety of football playing styles.

In the end, winning teams combine offensive skill, defensive tenacity, and strategic discipline. The combination of PCA and HCA in this study creates a foundation for a better comprehension of the elements that characterize football success and serves as a launchpad for more research into the complex realm of team performance.

Data Source

<https://www.kaggle.com/datasets/varpit94/football-teams-rankings-stats>

Seminars & courses on online.ase.ro