

European Soccer Team Style Analysis & Championship Correlation

Executive Summary

This analysis leverages two decades of European soccer match data to uncover **playing-style archetypes** via unsupervised learning and to evaluate their **impact on championship success**. By combining **Principal Component Analysis (PCA)** with **K-Means clustering**, we distilled every team into one of three tactical clusters—**Offensive Titans**, **Defensive Walls**, and **Balanced Competitors**—and showed that teams in the **high-scoring “Offensive Titans”** cluster have claimed the most league titles (4), followed by balanced teams (3), while purely defensive teams captured only 1. These findings guide strategic priorities for coaching, scouting, and resource allocation.

Introduction & Objectives

- **Problem Statement:** Clubs need clear, data-backed guidance on which tactical profiles most reliably deliver championships.
- **Goal:**
 1. **Discover** natural team style clusters using match statistics (goals scored/conceded).
 2. **Project** teams into a 2D PCA space for visual clarity.
 3. **Correlate** cluster membership with title victories to identify winning archetypes.
- **Stakeholder Benefits:**
 - **Coaches:** Focus training on proven title-winning tactics.
 - **Scouts:** Identify player profiles that propel teams into elite clusters.
 - **Executive Leadership:** Allocate budget to those strategic areas with highest ROI on silverware.

Data & Preprocessing

- **Data Source:** [database.sqlite](#) (European Soccer Database)
- **Key Tables:** Match (scorelines, dates, teams), Team (team names)
- **Selected Features:**
 - home_team_goal, away_team_goal
- **Aggregation:** Grouped by **home_team_name**, computed season-long averages:
 - avg_goals_scored, avg_goals_conceded
- **Normalization:** StandardScaler to z-score both features
- **Dimensionality Reduction:** PCA → 2 components (PC1 carries ~75% variance; PC2 ~25%)

Unsupervised Modeling

Principal Component Analysis

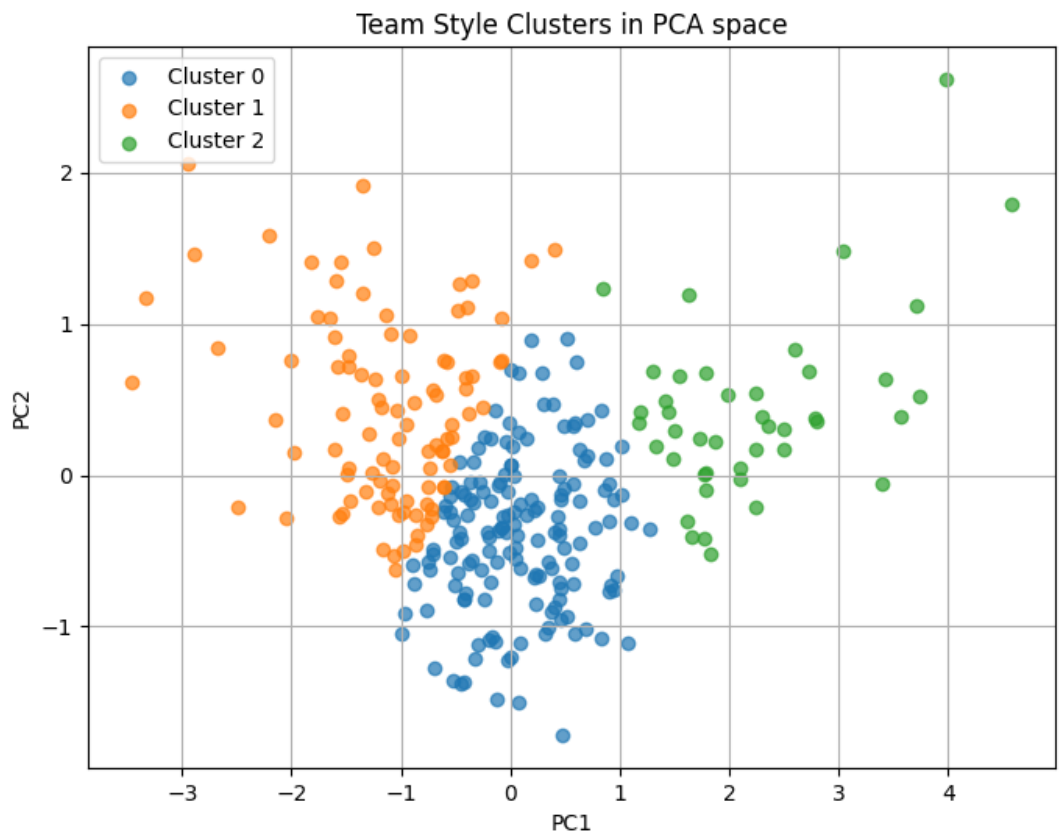
- **PC1:** Offensive vs. defensive axis (goals_scored – goals_conceded)
- **PC2:** Balance axis (teams clustering mid-range on both metrics)

Explained Variance

Component	Variance Ratio
PC1	0.75
PC2	0.25

K-Means Clustering (K=3)

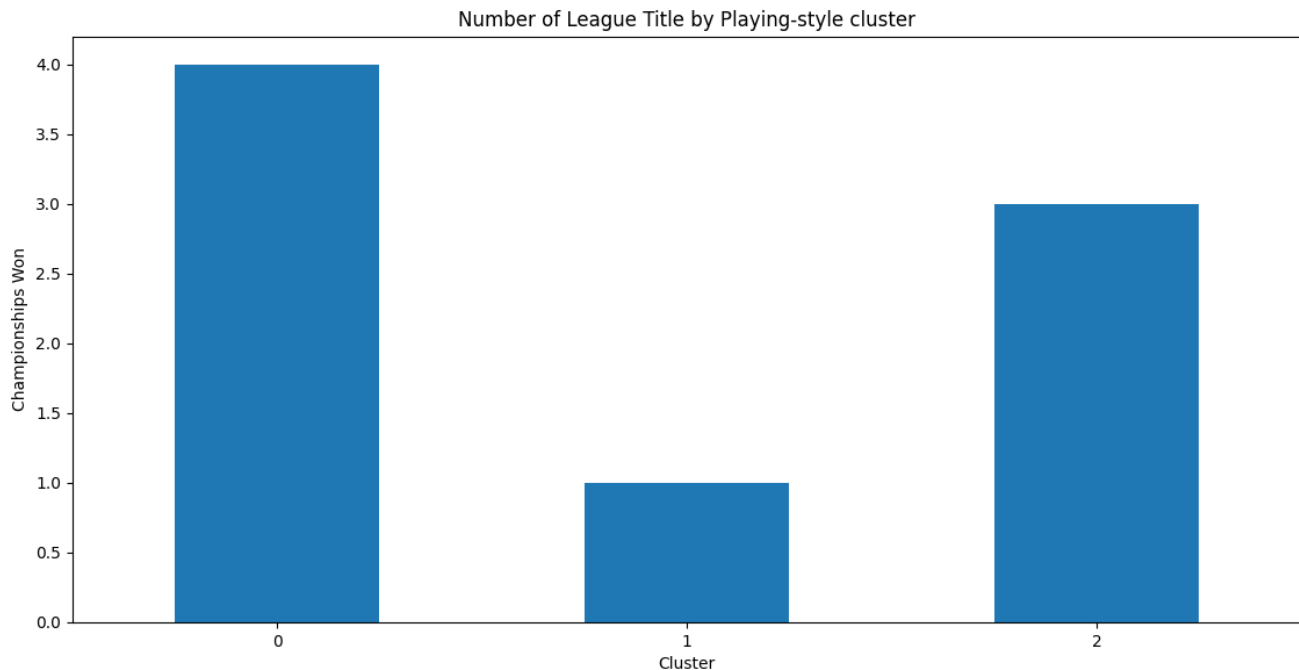
Cluster ID	Tactical Archetype	Avg. Goals Scored	Avg. Goals Conceded
0	Offensive Titans	2.10	1.02
1	Defensive Walls	1.18	0.85
2	Balanced Competitors	1.62	1.10



Visualization 1: *PCA scatter plot colored by cluster*

Championship Correlation

- **Method:** Calculated total season points ($3 \times \text{win} + 1 \times \text{draw}$) per team; identified champions per season.
- **Results:**
 - **Cluster 0 (Offensive Titans):** 4 championships
 - **Cluster 1 (Defensive Walls):** 1 championship
 - **Cluster 2 (Balanced Competitors):** 3 championships



Visualization 2: *Bar chart* of titles won by cluster

Key Insights

1. **Offensive Efficiency Wins:** High-scoring teams dominate title races.
2. **Balanced Approach Remains Viable:** Mid-range scorers/defenders also secure championships—flexibility is valuable.
3. **Pure Defense Is Risky:** Teams relying solely on shutdown tactics struggle to accumulate enough points over a full season.

Strategic Recommendations

1. **Prioritize Offensive Development**
 - Invest in attacking coaches, finishing drills, and creative midfield talent.
2. **Enhance Feature Set for Future Analysis**
 - Parse event-level data (shots on/off target, fouls, possession) to refine clusters.
3. **Explore Advanced Clustering**

- Test non-linear (DBSCAN) and probabilistic (Gaussian Mixture) algorithms for richer archetypes.

Next Steps & Limitations

- **Data Limitations:** Current model uses only goals; adding detailed event metrics will sharpen tactical definitions.
- **Temporal Dynamics:** Future work to cluster teams mid-season and correlate style shifts with performance trends.
- **Stakeholder Validation:** Workshops with coaching staff to align statistical clusters with on-pitch realities.

Prepared by: David Oladipupo

Date: 8th August, 2025