Assessing UEFA Euro 2024 Team Playstyles using Cluster Analysis

Applied Research Project (UD7004)

Course: Performance Analysis in Football MSc

Name: Connor Hayes

Student ID: 39639

Submission: July 2025

Supervisor: Leah Johnstone

Abstract

The purpose of this project is to combine the research methods of Ghezzi (2024) and Trower et al. (2023) to objectively categorise and measure team playstyles in football. I employed Gaussian Mixture Models (GMMs) to cluster playstyles from UEFA Euro 2024 using StatsBomb 360 data. I analysed 25 features including match phases (in-possession, out-of-possession, and in transition) and metrics that minimise team-quality bias. Four playstyles were identified: high-pressing possession-based teams with short central passing, possession-based teams favouring attacking wide play, low block counterattacking teams, and direct low-block teams. The findings reveal that teams who adopt high-press and counterpress strategies alongside short, accurate passes in central midfield and attacking thirds, score the best KPI values both in-possession and out-of-possession. Whereas the least effective playstyle was a low block counterattacking approach. I found a statistically significant relationship ($p=0.02$) between playstyle and goals conceded per 90, indicating that playstyle has a significant impact on defensive performance but less of an impact on attacking performance. This project advances performance analysis in football by providing a reliable method for objectively assessing playstyles and their effectiveness. This offers practical applications for opposition analysis, managerial decision-making, and longitudinal performance reports.

Table of Contents

Introduction

In football, tactics are the coordinated actions performed by players in response to dynamically changing match situations (Low, 2020). Teams adopt tactical approaches based on various contextual factors such as the competition or league, team and opponent quality, or game state (González-Ródenas, 2020; Plakias et al., 2023). In addition, clubs and national teams are often influenced by underlying philosophies or managerial beliefs (Partington, 2024; Wenger, 2021). Together, these tactical approaches and overarching philosophies form a team's style of play, or playstyle – defined by Plakias et al. (2023) as a team's distinctive and repeated patterns observed over numerous matches.

While much of the existing literature on data analysis in football focuses on evaluating individual player performance (Lucey et al., 2014; Decroos et al., 2019; Fernández et al., 2021; Power et al., 2017; Link & de Lorenzo, 2016), fewer studies have investigated team-level playstyles and how they relate to success. However, in recent years, the increasing availability of tracking data has allowed practitioners to quantify and measure team playstyles through concepts such as match phases (Ghezzi, 2024). Measuring playstyles in relation to key performance indicators (KPIs) such as goals scored or expected goals (xG) allows analysts to evaluate the effectiveness of different approaches – providing valuable insights for match analysis, opposition analysis, player recruitment, and managerial decision-making.

The aim of this project is to quantify and categorise team playstyles at UEFA Euro 2024 using StatsBomb 360 data and to assess the relationship between these styles and KPIs. This project builds on the current literature, incorporating Gaussian Mixture Models (GMMs) to analyse team-level playstyles using performance data. By linking playstyle classification to tournament performance, I aim to benchmark successful playstyles and provide a data-driven approach to evaluating tactical decisions, with potential applications in coaching, scouting, and performance analysis.

In the following sections, I will first review and evaluate the relevant literature on football playstyles. This will be followed by the methodology section, which outlines the data sources and analysis techniques used. I will then present the results of the analysis, discussing the key findings and their practical implications. Finally, I will reflect on the limitations of the project, give thought to future research, and conclude with a summary of the results.

Literature Review

The existing literature has been categorised into three main areas. First, I will outline the football concepts supported by academic literature that are relevant to this project. Second, I will examine studies that measure match phases. Third, I will review data-driven studies that classify and evaluate team playstyles. Throughout, I will assess the reliability and the validity of both the football concepts, and the data analysis methods used. By comparing these sources, I aim to identify current knowledge gaps in the literature, and justify the contribution of this research.

Football Concepts

Due to a lack of consistency and scientific validation in defining football concepts, it is important to clarify the key ideas used in this project. There are several playstyles that frequently appear in both academic literature and professional reports, including possession-based (FIFA, 2022), counterattack (FA, 2021), high-press (FIFA, 2021), and direct play (Kempe, 2014). While these styles may overlap, and teams may adapt them to suit their context, they provide a useful foundation for evaluating team-level behaviours.

A key concept used by coaches and analysts to help implement these playstyles is match phases. Match phases are used to segment football matches based on the location of the ball, the team in possession, and the duration of that possession - allowing coaches to deploy instructions and review performance specific to each phase. There is currently no set framework for defining match phases; different practitioners may use a different number of phases and have different definitions for each phase.

Measuring Match Phases

The mass of data collection in football through tracking or hybrid techniques has enabled practitioners to more reliably identify and measure match phases. Hewitt (2016) proposed an early framework for measuring playstyles by segmenting matches into five "moments of play". These include: established offence, transition to offence, established defence, transition to defence, and set pieces. These phases were manually assigned to Opta event data from the 2013/14 English Premier League season. Hewitt (2016) then analysed the frequency and effectiveness of events, ranking team performance in each phase.

Although Hewitt's (2016) addition of match phases offered valuable tactical context, the lack of statistical modeling to test relationships between performance and playstyle limits the validity of the research. Moreover, the absence of tracking data reduces the reliability and repeatability of the manual classifications. Despite its limitations, this research provides a strong foundation for classifying match phases, which becomes more practical with the increased availability of tracking data.

More recently, Ghezzi (2024) utilised StatsBomb 360 data to classify the proportion of events recorded in 12 open-play match phases, as well as attacking and defensive set-pieces. Working with FC Lugano, Ghezzi developed an algorithm that automatically assigns event data to one of the 14 match phases. This allowed FC Lugano to consistently measure and review their own and their opponent's tactical behaviours over multiple games. The framework included in-possession phases (Build-up, Progression, Finishing), out-of-possession phases (Low Block, Low Press, Mid Block, Mid Press, High Block), and transition phases (Consolidate, Counterattack, Regroup, Counterpress).

By combining event and tracking data, Ghezzi's framework significantly improved the reliability and validity of match phase classification. The automated system enabled efficient post-match and opposition analysis by providing an objective overview of the proportion of events in each phase. These insights can be combined with performance metrics to measure performance in relation to match phases, offering clubs objective tactical insights.

Player-Level Clustering

In terms of methodology, the most relevant study to this project is by Trower et al. (2023), who used StatsBomb event data to categorise Women's Super League (WSL) players into 11 clusters. They achieved this by first applying Principal Component Analysis (PCA) to reduce the dimensionality of the dataset based on 25 player-level features, followed by a Bayesian Gaussian Mixture Model (BGMM) to group players with at least 1,000 minutes played across five seasons. The features included event-based statistics such as number of dribbles and pass distance, as well as location-based metrics such as number of passes in the middle third. All statistics were normalised per 90 minutes and standardised using z-scores.

A strength of the Trower et al. (2023) study is the use of PCA to reduce the dimensionality to 16 features, preserving 95% of the data's variance, while minimising noise and multicollinearity. Also, GMM is highly effective for capturing the multidimensional nature of performance data in football. By accounting for variance in cluster shape, size, and distribution, GMM identified "hybrid players" – those who are probabilistically assigned to multiple clusters. Utilising BGMM also allowed the number of clusters to be inferred rather than pre-defined, which is useful for problems where the optimal cluster count may vary or be difficult to determine.

However, one of the clusters identified, named 'Wildcards', was removed because the players were weakly associated with each other and lacked a clear common function. In addition, there appeared to be strong overlap between clusters such as "Interception masters" and "Intercepting strikers", highlighting the subjectivity involved in interpreting and labelling clusters. Also, despite the use of location-based features in the data, the analysis does not link player actions to specific match phases. For example, while the "Interception masters" cluster described out-of-possession behaviours, no inference can be made regarding how these players perform in-possession. Integrating different match phases, such as the build-up phase or defensive transition, would provide greater tactical context to the player profiles.

The motivation for the Trower et al. (2023) study stems from Soccerment (2022), who originally clustered player profiles from Europe's top five men's leagues based on playing style and functions. Soccerment (2022) analysed over 1,300 outfield players using UMAP - a non-linear dimensionality reduction method - followed by BGMM. Using Opta event data, the dataset included 28 features that were normalised to represent playing style rather than performance. To minimise player-quality bias, performance metrics such as xG and expected assists (xA) were excluded from the analysis while defensive and passing actions were scaled relative number of touches. The data included average in-possession player coordinates and categorised actions across five game phases. Soccerment (2022) iterated BGMM to identify 5 macro, and 13 sub clusters. The resulting sub clusters represent player functions, such as "Build-Up Initiator" and "Chance Creator". Players were also assigned probabilistic memberships to allow for the identification of hybrid players who belong to multiple clusters.

The use of proportional bar charts to visualise cluster memberships for individual players is highly effective. This provides a clear overview of a player's role flexibility, offering scouts and analysts an empirical measure of their ability to adapt to different roles. Soccerment's (2022) use of UMAP also offers an effective alternative to PCA, as it can capture complex, multidimensional relationships between features. However, UMAP is more sensitive to hyperparameters and produces outputs that are less interpretable than PCA (Chang, 2025).

Despite the integration of Opta's five predefined game phases, these phases do not account for ball location or possession duration. For example, Opta's "defensive phase" aggregates all defensive actions regardless of whether a team is in a high-press or a low block. As a result, there is no connection to player abilities in attacking or defensive transition. The average player coordinates are based on in-possession data only and therefore offer no insight into defensive positioning or out-of-possession behaviours. Like Trower et al. (2023), the absence of context-specific match phases limits the ability to fully describe a player's playstyle or function.

Team-Level Playstyle Clustering

Moving on to a team-level focus, Gollan et al. (2018) clustered English Premier League teams based on their performance across Hewitt's (2016) five match phases. Employing k-means clustering on Opta data from the 2015/16 season, Gollan et al. (2018) identified three cluster types representing teams strong in established defence, dominant in transition, and strong in established offence and set-pieces. The study highlighted that team's dominant in transitions tended to rank higher in the league.

The use of k-means clustering has limitations, as this approach assumes equal-sized clusters with no overlap between them, which is not suitable for performance data in football. Furthermore, due to the inclusion of performance metrics such as goals and chances created, Plakias et al. (2023) notes that Gollan et al. (2018) categorised teams based on the phases in which they excelled, rather than providing an unbiased categorisation of playstyles that minimises the influence of team quality. The research is also constrained by the lack of context-specific match phases, as no location data is available within the event data.

More recently, Gleeson (2024) investigated the variation in Premier League playstyles across matches and within games. Gleeson (2024) incorporated a mixture of tracking and event data, including On-Ball-Value (OBV) and Balance Score, while considering the influence of game state by segmenting games into time periods. To measure the variance in playstyle, Gleeson (2024) employed Gower's distance to measure similarity across team performance vectors. Hierarchical clustering with Ward's method of clustering was then used to classify the different styles. Gleeson (2024) found that 4 clusters produced the most accurate and interpretable results, these were labelled: Positive, Direct, Passive, and Balanced playstyles. The results suggest that Positive and Balanced were the most common playstyles observed. However, a balanced playstyle returned the worst outcomes, especially when identified in the 2nd half of a game. Lower quality teams also tended to have less variation in tactical behaviour.

The research builds on the existing literature with the addition of game state as a factor in identifying and measuring tactical behaviours. The inclusion of game state introduces an empirical approach to applying game theory in tactical decision-making with potential implications on opposition analysis. However, Gleeson (2024) notes that a larger sample size is needed to determine a causal relationship in the findings. Furthermore, while Gleeson (2024) does try to mitigate team strength bias, the data still favours high-quality teams, as they tend to have more shots, passes, and actions than weaker teams. Once more, despite the use of geometric variables, specific match phases are not recorded in the data.

Knowledge Gaps and Rationale for This Project

The existing literature highlights the potential of using match phases to cluster teams based on their playstyle. Currently, there are inconsistent definitions of match phases with frameworks varying between studies. Ghezzi (2024) has produced the most detailed and reliable framework for measuring match phases using StatsBomb 360 data. The use of clustering has been successfully applied to group players based on their individual functions (Trower et al., 2023; Soccerment, 2022), but these profiles could be enhanced with the inclusion of specific match phases in the feature space. Additionally, few studies have successfully explored how team playstyles relate to performance outcomes. This is because the studies often incorporate performance metrics such as goals and xG in the data used to measure playstyles - which results in a bias to high-quality teams. Furthermore, the studies by Ghezzi (2024), Trower et al. (2023), and Gleeson (2024) were published through StatsBomb industry conferences and may not be peer-reviewed. Therefore, these sources should be classed as industry research rather than academic literature. Highlighting the lack of academic validation on this topic.

To address the gaps in the literature, I will apply the methodology of Trower et al. (2023) to team-level playstyles and incorporate an adapted version of Ghezzi's (2024) match phases framework to classify and measure different styles. I will apply GMM to aggregated event data and location-based metrics gathered from StatsBomb UEFA Euro 2024 360 data. The resulting clusters will be analysed in relation to key performance indicators such as goals per 90 and xG per 90. This approach offers a reliable method for classifying and evaluating team playstyles, with implications for both academic research and applied performance analysis.

Methodology

Research Design

This project employed a quantitative research design using secondary data from StatsBomb. It aims to objectively measure and categorise team playstyles from UEFA Euro 2024, and in turn evaluate the effectiveness of the styles identified by comparing them with Key Performance Indicators. A quantitative approach is appropriate given the numerical nature of the data. All 24 teams from the tournament are measured and compared, reflecting a between-subjects research design. The data analysis can be accessed in a Jupyter Notebook uploaded to my StatsBomb360 repository on my GitHub.

Data Source

StatsBomb is an industry-leading data provider in football, collecting over 3,400 events per match. Offering free data via their Python package statsbombpy, they provide detailed records of both in-possession and out-of-possession actions, including pitch coordinates for all events. StatsBomb 360 data adds further contextual information by capturing the locations of all visible attacking and defensive players within the camera frame at the moment of each event (Figure 1). In this project, I analysed StatsBomb event and 360 data for all 51 matches played at the UEFA Euro 2024 tournament.

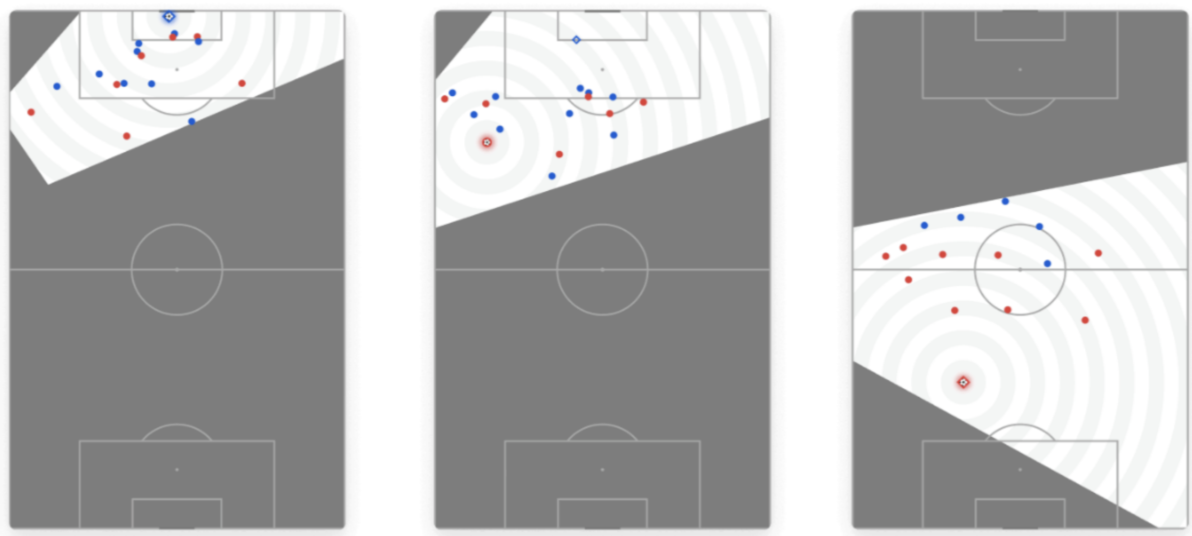


Figure 1: *Example of StatsBomb 360 data freeze-frames (Source: StatsBomb, 2025)*

Data Pre-processing

Having loaded the relevant Python libraries and the StatsBomb data, I created a function to segment the pitch into defensive, midfield, and attacking thirds, in addition to eight specific zones – merging the wide areas and half-spaces to aggregate the data (Figure 2). Dividing the pitch into zones is common practice among coaches and analysts, with academic research demonstrating that playstyle is closely linked to the areas of the pitch where a team performs actions (Brooks et al., 2016; Amatria et al., 2019). For simplicity, I did not incorporate set-play metrics in my feature set as I wanted to focus on teams' style of play characteristics in open-play.

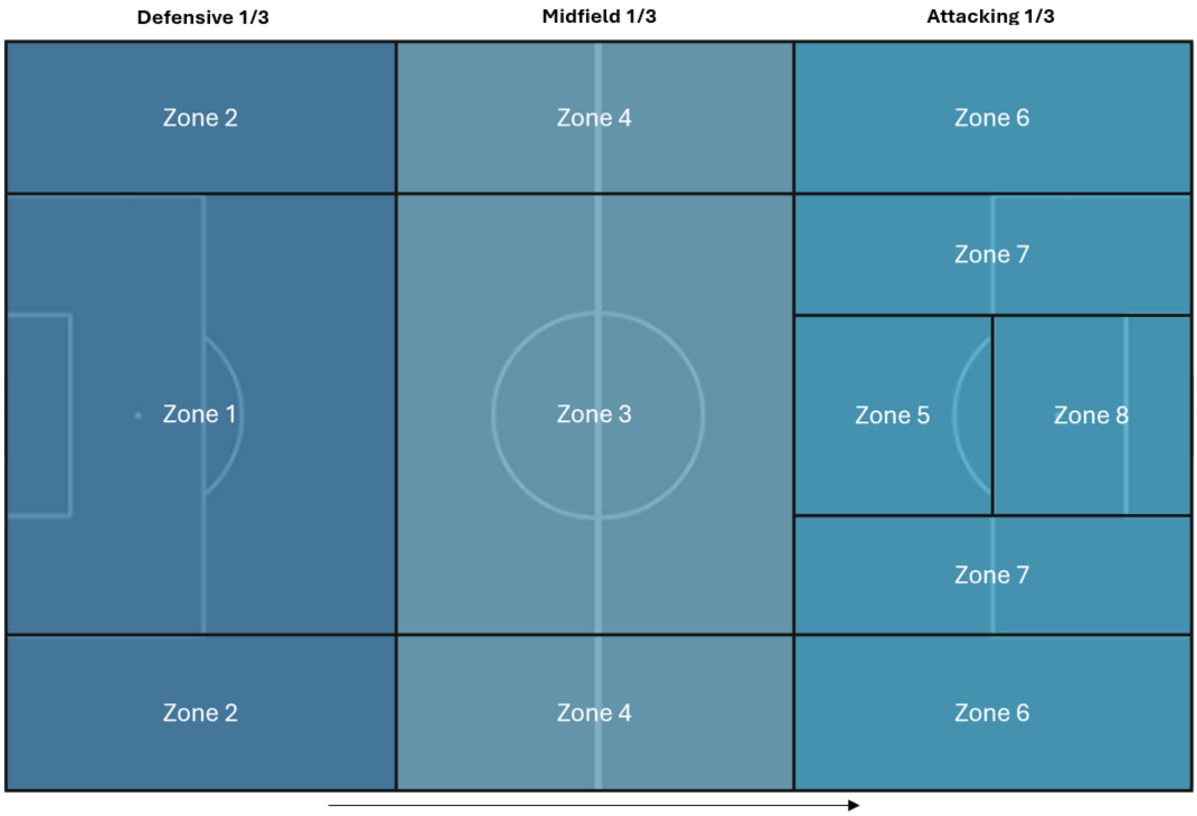


Figure 2: *Pitch zones for identifying in-possession match phases and attacking style*

In total, 25 variables were measured to characterise team playstyles, with an additional four KPIs also measured to evaluate the resulting clusters (Table 1). I looped through each of the 51 games from the tournament, producing a dataset that contains statistics for each team, per match – resulting in 102 total rows of data. I then calculated the average values for each team, giving a final 24 rows of data representing one row per team. All features were represented per 90 minutes or expressed as percentages and subsequently standardised using z-scores.

In-Possession Features

To characterise playstyle in-possession, I used the pitch zones function to measure which zones of the pitch are used most often to progress possession and generate attacking opportunities. Furthermore, I measured carries per 90, median pass length per 1/3, pass direction %, and aerial pass % to help describe how teams move the ball. I measured counterattacks per 90 using the play_pattern variable in the StatsBomb match data, recording events marked as "From Counter". Finally, I measured in-possession performance with xG per 90 and goals per 90.

Variable	Phase	Definition
Goals per 90	In-Possession KPI	Total number of goals scored per 90 minutes.
xG per 90	In-Possession KPI	Total expected goals (xG) generated per 90 minutes, based on StatsBomb's model that calculates the probability of a shot resulting in a goal.
Counterattacks per 90	In-Possession Transition	Total number of Counterattacks per 90 minutes, denoted by a possession started with an open play turnover outside the counter-attacking team's final third, moved at least 75% direct towards goal (as measured by StatsBomb's possession chain metrics), and travelled at least 18 yards towards goal.
Carries per 90	In-Possession	Total number of times a player moves the ball at their feet while maintaining control (includes movement while standing still) per 90 minutes.
Passes per 90	In-Possession	Total number of passes completed between teammates per 90 minutes.
Possession %	In-Possession	The proportion of a team's total passes relative to the combined total passes of both teams.
Pass Accuracy %	In-Possession	The percentage of a team's successful passes out of their total passes.
Aerial pass %	In-Possession	The percentage of a team's aerial passes (ball goes above shoulder level at peak height) out of their total passes.
Forward pass %	In-Possession	The percentage of a team's forward passes (angle of pass is >=-0.79 and <=0.79) out of their total passes.
Sideways pass %	In-Possession	The percentage of a team's sideways passes (angle of pass is >0.79 and <=2.36 or <-0.79 and >=-2.36) out of their total passes.
Backward pass %	In-Possession	The percentage of a team's backwards passes (angle of pass is >2.36 and <-2.36) out of their total passes.
Median pass length (Def 1/3)	In-Possession	The median pass length of passes into their Defensive 1/3.
Median pass length (Mid 1/3)	In-Possession	The median pass length of passes into their Midfield 1/3.
Median pass length (Att 1/3)	In-Possession	The median pass length of passes into their Attacking 1/3.
Passes (Def 1/3 central)	In-Possession	The proportion of a team's passes that are played into the central zone in their Defensive 1/3 (Zone 1).
Passes (Def 1/3 wide)	In-Possession	The proportion of a team's passes that are played into the wide zones in their Defensive 1/3 (Zone 2).
Passes (Mid 1/3 central)	In-Possession	The proportion of a team's passes that are played into the central zone in their Midfield 1/3 (Zone 3).
Passes (Mid 1/3 wide)	In-Possession	The proportion of a team's passes that are played into the wide zones in their Midfield 1/3 (Zone 4).
Passes (Att 1/3 central low)	In-Possession	The proportion of a team's passes that are played into the central low zone in their Attacking 1/3 (Zone 5).
Passes (Att 1/3 wide)	In-Possession	The proportion of a team's passes that are played into the wide zones in their Attacking 1/3 (Zone 6).
Passes (Att 1/3 half-space)	In-Possession	The proportion of a team's passes that are played into the half space zones in their Attacking 1/3 (Zone 7).
Passes (Att 1/3 central high)	In-Possession	The proportion of a team's passes that are played into the central high zone in their Attacking 1/3 (Zone 8).
Goals conceded per 90	Out-of-Possession KPI	Total number of goals conceded per 90 minutes.
xGA per 90	Out-of-Possession KPI	Total expected goals (xG) conceded per 90 minutes, based on StatsBomb's model that calculates the probability of a shot resulting in a goal.
Counterpress %	Out-of-Possession Transition	The percentage of opposition open play turnovers followed by a team performing a pressing action (pressure, dribbled past, 50-50, duel, block, interception, or foul committed) recorded within 5 seconds. Proportionate to the total out-of-possession and defensive transition phases.
Regroup %	Out-of-Possession Transition	The percentage of out-of-possession and defensive transition phases where a team is defending an opposition counterattack.
High-Press %	Out-of-Possession	A team is out-of-possession and has at least four players in their Attacking 1/3, proportionate to the total out-of-possession and defensive transition phases.
Mid Block %	Out-of-Possession	A team is out-of-possession and has at least four players in their Midfield 1/3, proportionate to the total out-of-possession and defensive transition phases.
Low Block %	Out-of-Possession	A team is out-of-possession and has at least six players in their Defensive 1/3, proportionate to the total out-of-possession and defensive transition phases.

Table 1: *Full feature set including match phase and definitions*

Out-of-Possession Features
The pitch zones function is also used to describe out-of-possession playstyle by measuring events that are classed as low block, mid block, or high-press. Like Ghezzi (2024), I used StatsBomb 360 data to classify these phases based on the number of out-of-possession players visible within each third. I measured defensive transitions with the proportion of counterpress and regroup scenarios identified. Similar to counterattacks, counterpresses are already identified within the StatsBomb match data, and regroup situations are when a team is defending an opposition counterattack. Out-of-possession performance is also measured using xGA per 90 and goals conceded per 90.

Hierarchical Clustering Approach with Gaussian Mixture Models (GMMs)
GMM is a soft clustering technique that accounts for uncertainty in cluster membership, making it suitable for identifying teams with overlapping or similar behaviours (Trower et al., 2023). The model assumes that the data is generated from a mixture of Gaussian distributions and uses the Expectation-Maximisation (EM) algorithm to estimate the parameters of these distributions. The algorithm iteratively performs two steps: the expectation step calculates the probability that each team belongs to each cluster, and the maximisation step updates the model parameters – including the mean, covariance, and weight of the clusters, based on these probabilities (McLachlan, 2000).

I applied GMM to the 25 standardised features, testing cluster counts from 1 to 9. Since my dataset is relatively small, and football performance data tends to be highly correlated, I used full covariance matrices with minimal regularisation to maintain model stability while capturing correlated tactical behaviours. The optimal number of clusters was determined using model selection criteria including the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) (Schwarz, 1978; Akaike, 2003), which balance model fit with complexity to avoid overfitting. I also tested the optimal number of clusters using the silhouette score, which measures how separated the resulting clusters are (Rousseeuw, 1987). After identifying the optimal cluster count, I implemented a hierarchical clustering approach, like Soccerment (2022), to identify more specific tactical behaviours within sub-clusters. First, I established the macro-clusters using the BIC/AIC and silhouette scores and then applied clustering within each macro-cluster to identify the sub-groups.

Results

Cluster Selection

The results of the AIC and BIC selection criteria clearly show an elbow point at 2 clusters (Figure 3). Therefore, according to the AIC and BIC tests, the optimal number of clusters to provide a balance between model fit and complexity is 2. The Silhouette score also supports a cluster count of 2, with a score of 0.325. However, the GMM clusters produced hard assignments for all teams, with cluster classification probabilities of either 0 or 1.

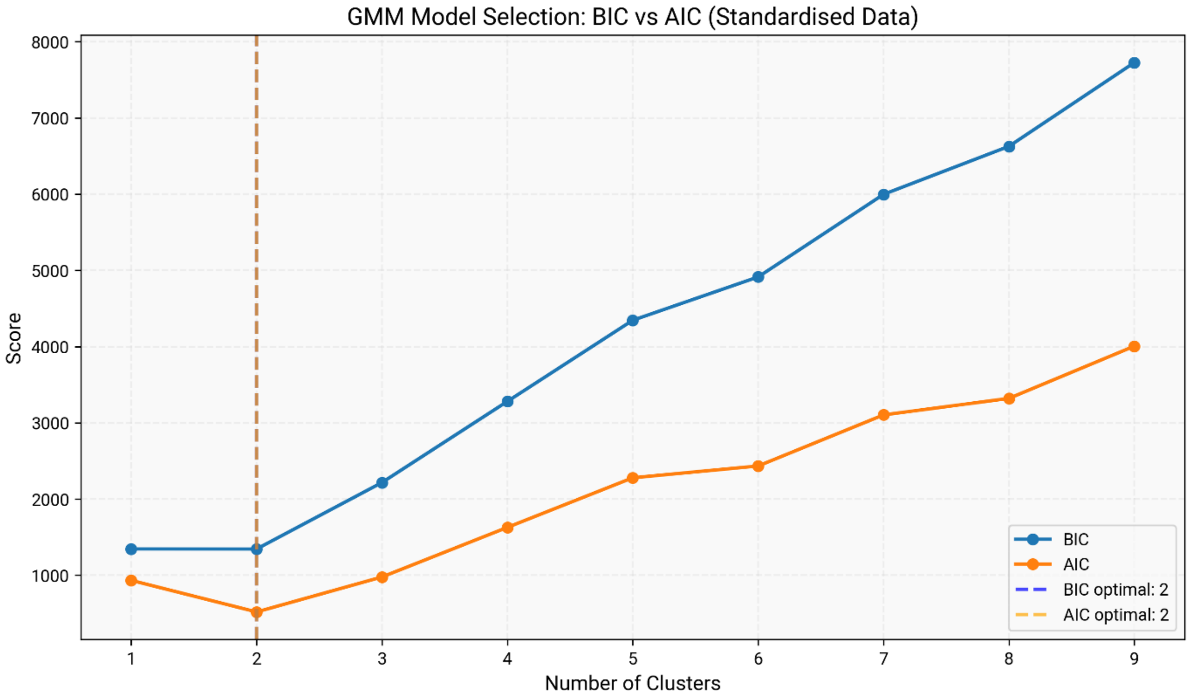


Figure 3: *Graph showing cluster selection based on AIC and BIC criteria*

Cluster Results

Two macro-clusters were identified, with a further four sub-clusters found using a hierarchical approach. In this process, both macro-clusters were clustered separately, resulting in each macro-cluster being divided into 2 sub-clusters. The final 4 clusters produced an overall silhouette score of 0.156. The clusters contained 6, 10, 5, and 3 teams respectively. Table 2 presents the team assignments across all four clusters, alongside KPI metrics. Figure 4 displays the separation between the teams and clusters in a two-dimensional PCA plot, while Figure 5 shows the relationship between clusters and the standardised features.

Cluster	team	Goals per 90	xG per 90	Goals conceded per 90	xGA per 90
0	Austria	1.43	1.45	1.18	1.34
0	Croatia	0.61	1.82	1.88	1.21
0	Denmark	0.48	0.93	0.96	1.12
0	Germany	2.08	1.6	0.48	0.73
0	Italy	0.71	0.78	0.96	0.98
0	Spain	1.85	1.39	0.38	0.82
1	Belgium	0.48	1.04	0.23	0.94
1	England	1.51	1.23	1.07	1.07
1	France	0.91	1.71	0.83	1.08
1	Netherlands	1.43	1.17	0.94	0.93
1	Portugal	1.44	2.41	1.29	1.63
1	Serbia	0.32	0.65	0.64	0.85
1	Slovakia	0.9	0.83	1.1	1.39
1	Switzerland	1.91	1.41	1.43	1.43
1	Turkey	1.5	1.22	1.14	1.53
1	Ukraine	0.63	0.95	1.27	0.8
2	Albania	0.93	0.68	1.27	1.34
2	Georgia	0.95	1	1.91	2.48
2	Hungary	0.61	1.04	1.59	1.19
2	Poland	0.96	1.13	1.92	1.6
2	Scotland	0.32	0.32	2.25	1.21
3	Czech Republic	0.95	1.45	1.26	1.34
3	Romania	0.96	0.9	1.44	1.36
3	Slovenia	0.48	1.2	1.02	1.63

Table 2: *Table showing team cluster assignment and KPIs*

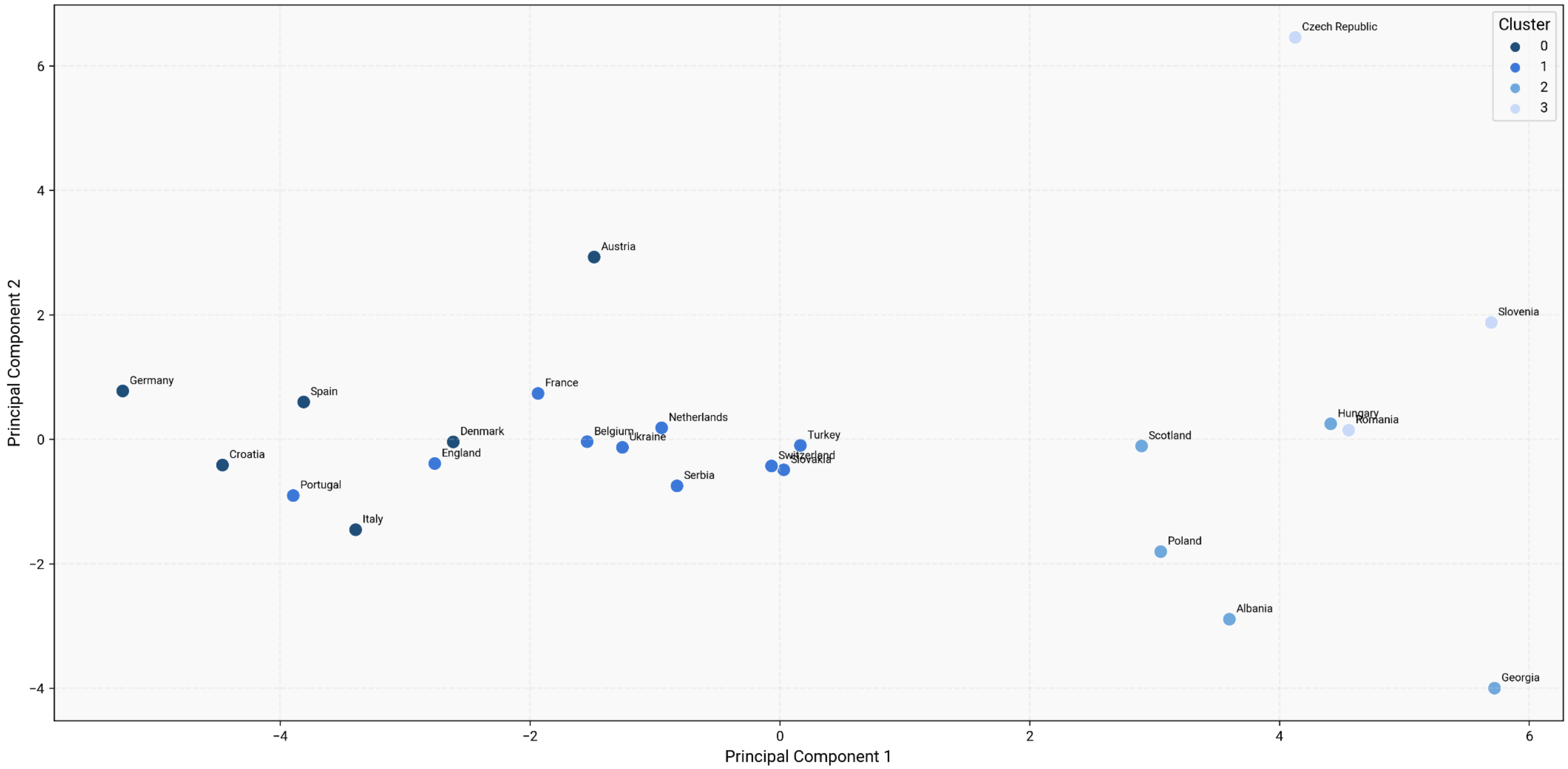


Figure 4: *Team cluster assignments in a two-dimensional PCA plot*

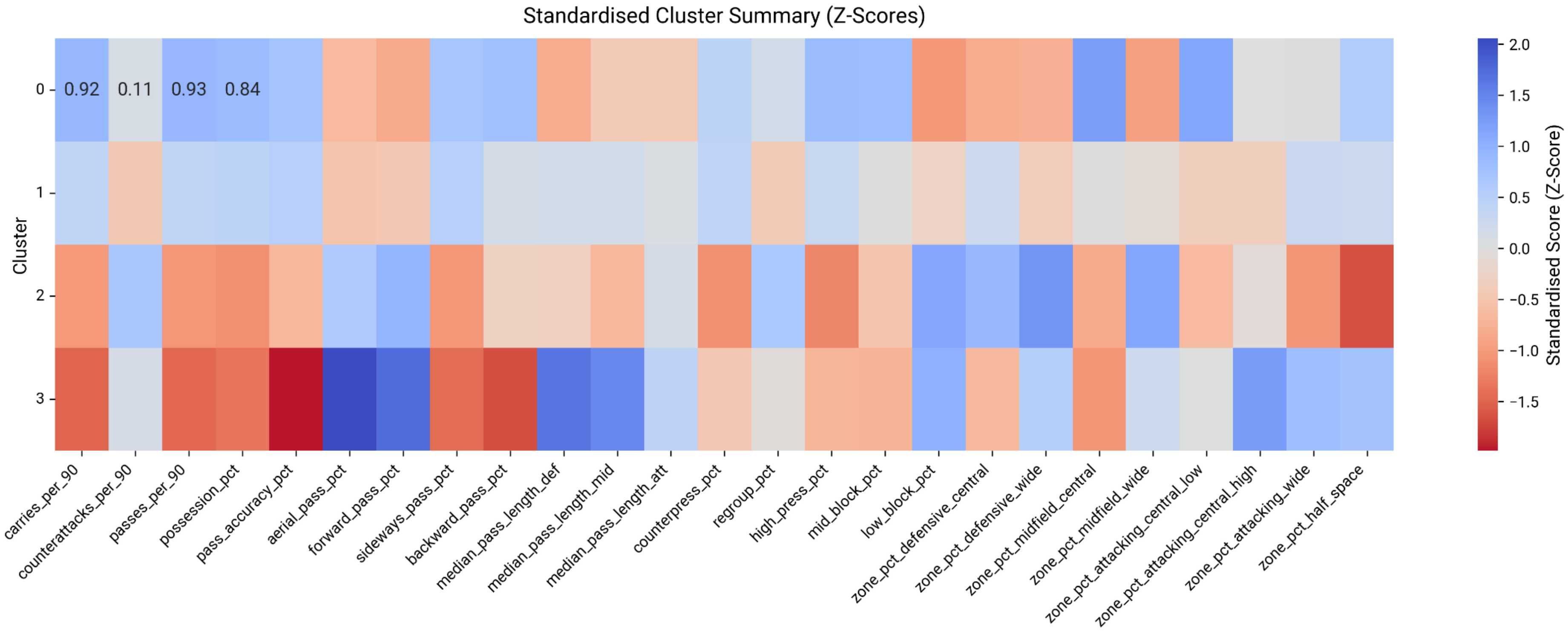
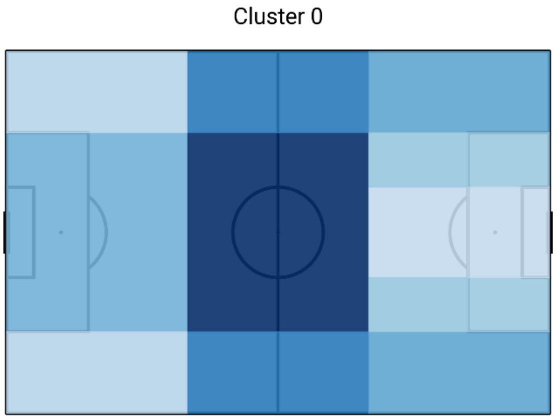
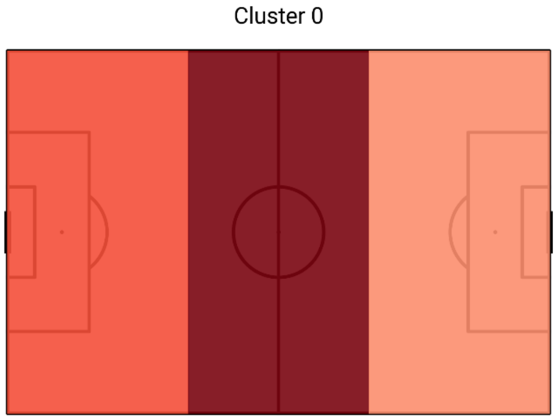


Figure 5: *Heatmap showing cluster assignments and standardised feature performance*

Cluster 0

Cluster 0 contains Austria, Croatia, Denmark, Germany, Italy, Spain. Teams in this cluster dominate possession, with the highest passes per 90 and shortest median pass lengths out of all the clusters (Figure 9). On average, teams in this cluster record the highest proportion of passes into the midfield central and attacking central low zones (Figure 7) – indicating controlled possession in midfield. Teams in this cluster record a high proportion of passes into the half-spaces and the out-of-possession metrics suggest frequent high-pressing (Figure 8) and counterpressing.





Figures 7 & 8: *Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 0*

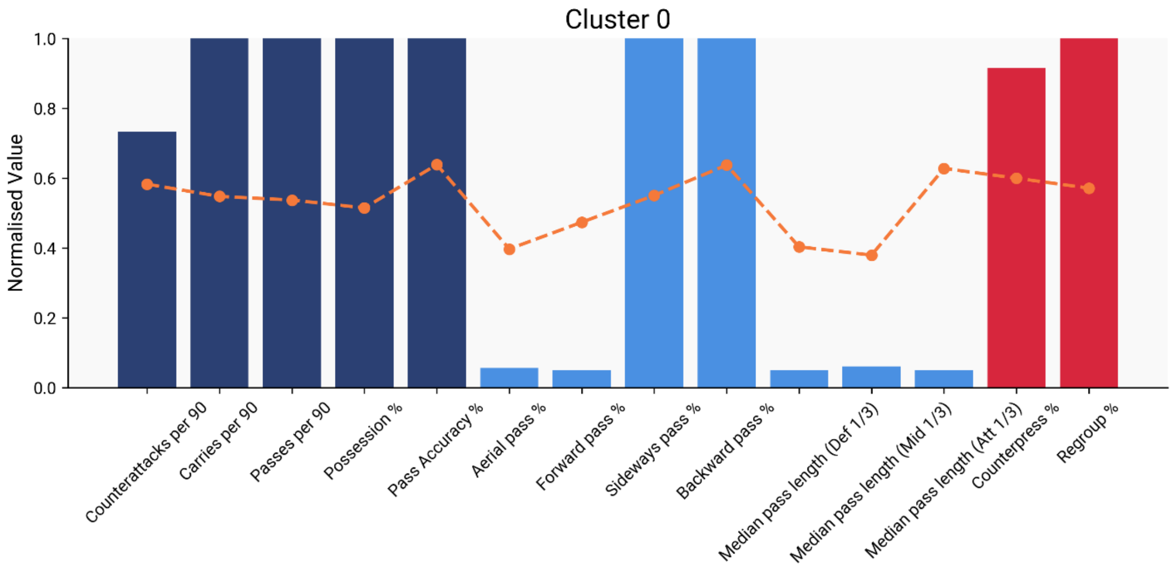
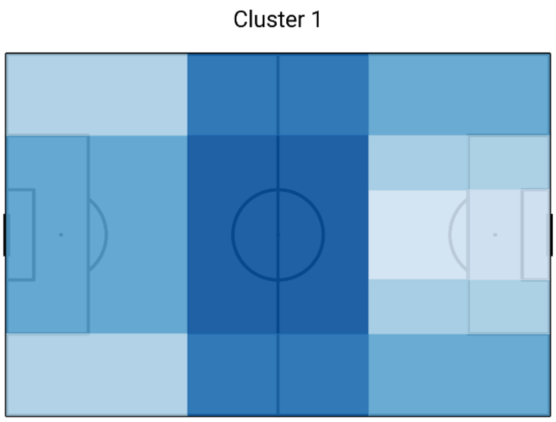
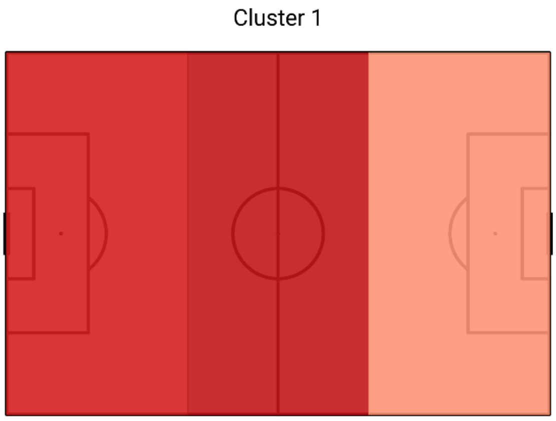


Figure 9: *Chart showing cluster 0 performance with normalised metrics*

Cluster 1

Cluster 1 is the largest cluster with Belgium, England, France, Netherlands, Portugal, Serbia, Slovakia, Switzerland, Turkey, and Ukraine. This cluster is also characterised by teams who are possession-based, with high passes and carries per 90, and low aerial pass % (Figure 12). Teams in cluster 1 also show high levels of counterpress and high-press scenarios (Figure 11). However, in contrast to cluster 0 teams, teams in cluster 1 record more passes into the attacking wide areas (Figure 10) and tend to have higher median pass lengths in the attacking third – suggesting a more direct, wide-play approach.





Figures 10 & 11: *Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 1*

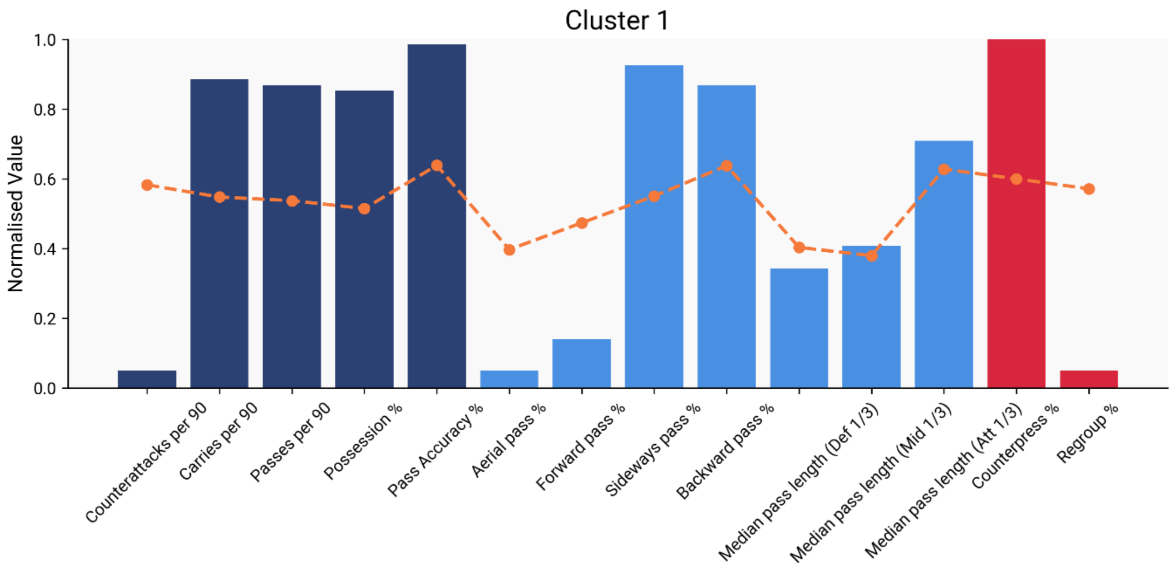
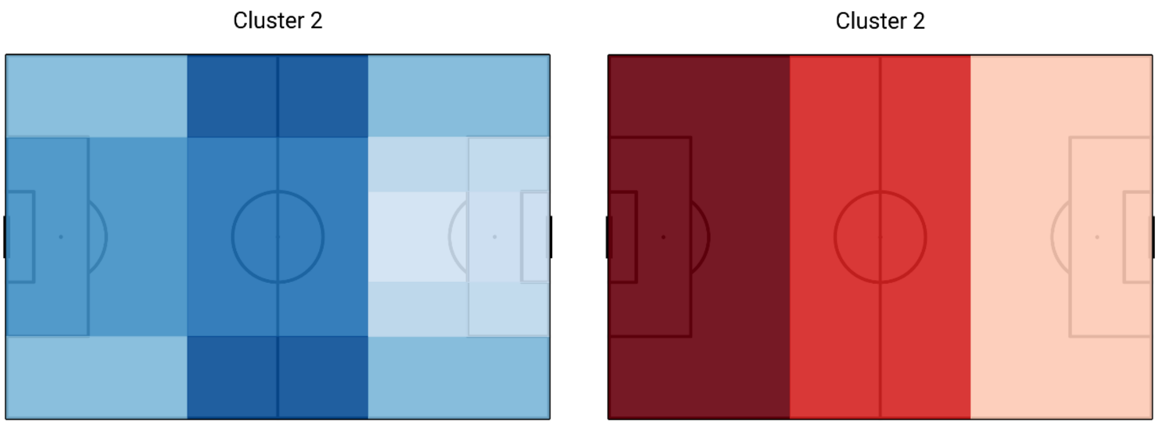


Figure 12: *Chart showing cluster 1 performance with normalised metrics*

Cluster 2

Cluster 2 is represented by Albania, Georgia, Hungary, Poland, and Scotland. These teams record the most counterattacks per 90 (Figure 15). This coincides with a more direct playstyle, with low passes per 90 and high forward pass %. In addition, they record their highest proportion of passes into the midfield wide zones (Figure 13). For out-of-possession, cluster 2 teams record the highest proportion of low block scenarios (Figure 14), and the lowest proportion of high-press and counterpress scenarios. This suggests a counterattacking playstyle where teams defend deep and wait for opportunities in attacking transition.



Figures 13 & 14: *Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 2*

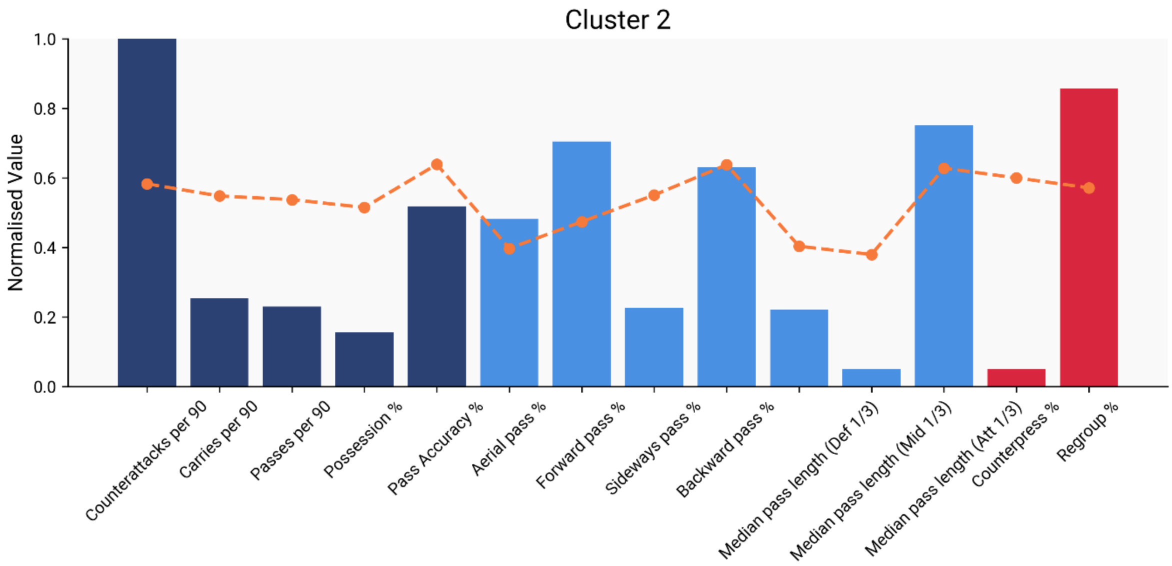
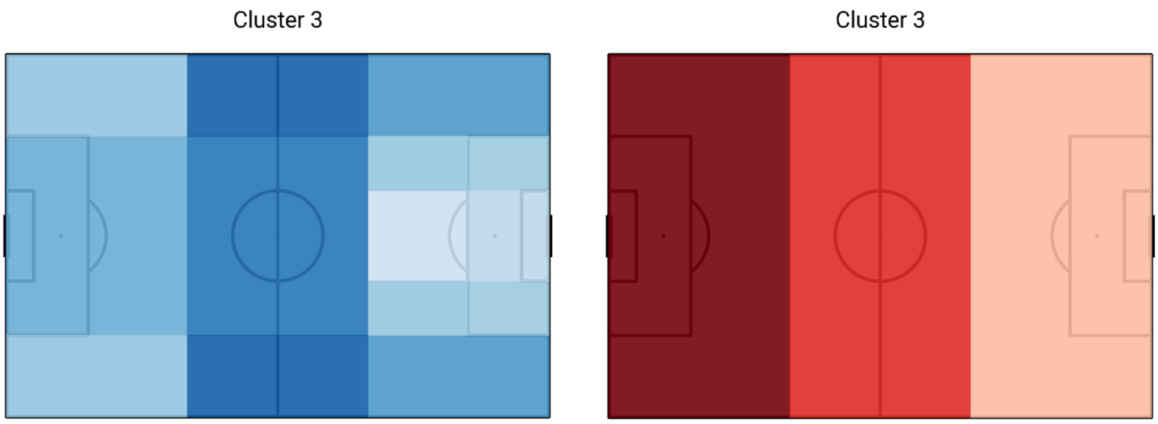


Figure 15: *Chart showing cluster 2 performance with normalised metrics*

Cluster 3

Cluster 3 is the smallest cluster with Czech Republic, Romania, and Slovenia. These teams are characterised by the most direct style of play with the lowest passes per 90 and the highest percentage of aerial and forward passes (Figure 18), and the highest median pass lengths across all three thirds. They also record the highest proportion of passes into the attacking central high zone, and scores relatively high for passes into the attacking wide and half space zones (Figure 16). Like cluster 2, teams in cluster 3 tend to defend in a low block (Figure 17), but with more counterpress moments.



Figures 16 & 17: *Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 3*

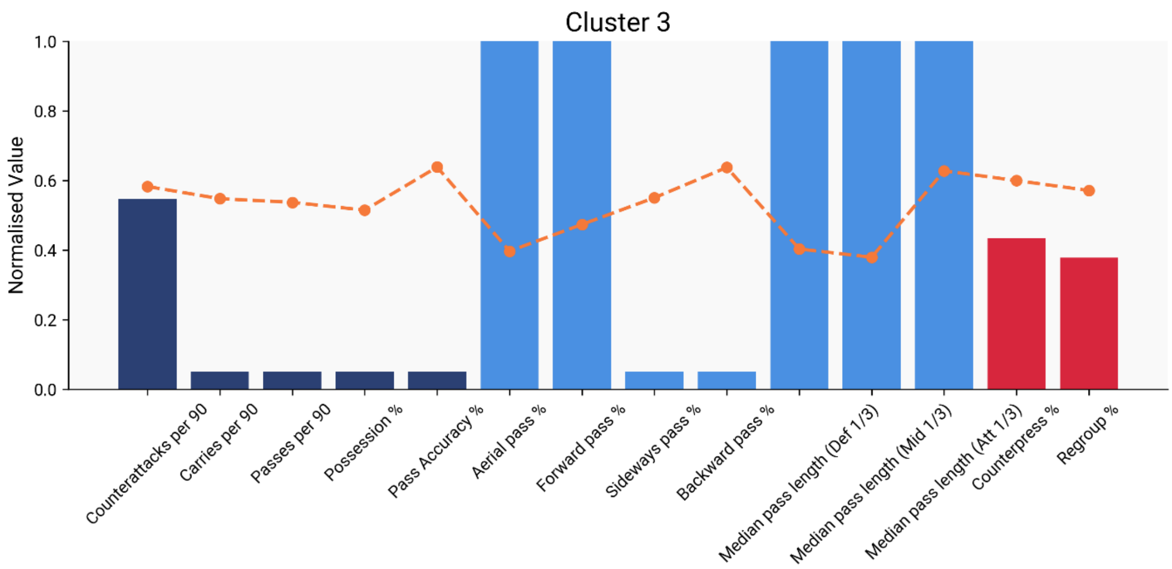


Figure 18: *Chart showing cluster 3 performance with normalised metrics*

Cluster Classification and Performance

Cluster 0 contains the highest performing teams with 0.22 average goal difference per 90 and 0.29 average xG difference per 90 (Figure 19) – making this playstyle the most effective during UEFA Euro 2024. Cluster 1 ranks second with 0.11 average goal difference per 90, and 0.10 xG difference per 90. Cluster 2 is the worst performing group, conceding 1.79 goals per 90 on average and recording a goal difference of -1.03 per 90. Cluster 3 improves on these scores but still records a negative goal difference of -0.44 per 90.

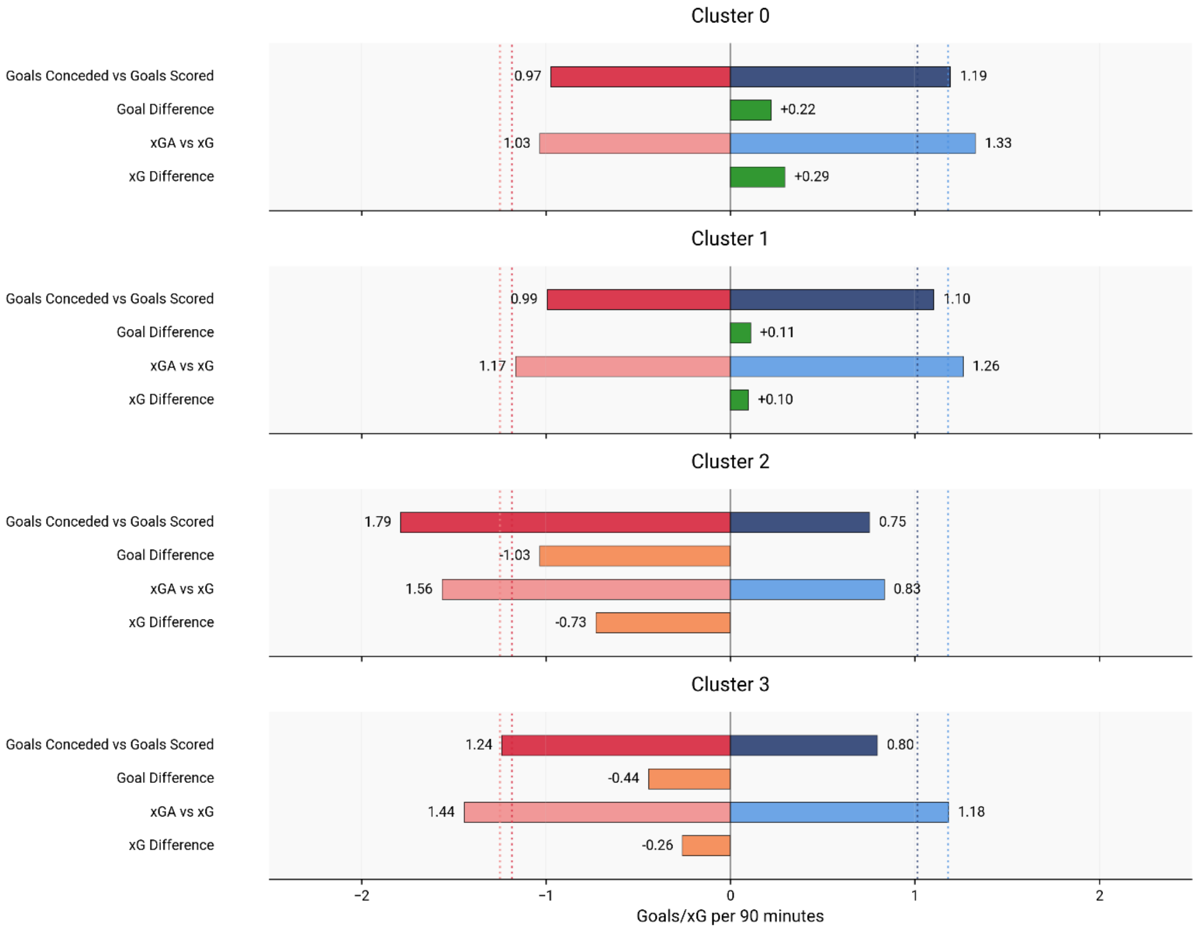


Figure 19: *Chart showing cluster classification and performance output*

I applied a Kruskal-Wallis test to measure the significance between cluster classification and KPI output, based on the team-level data in Table 2. The test revealed a statistically significant difference in goals conceded per 90 between clusters ($p = 0.02$). This suggests that playstyle had a significant impact on the number of goals conceded at UEFA Euro 2024. However, there were no significant differences for goals per 90, xG per 90, or xGA per 90. This implies that playstyle may not have a meaningful impact on these metrics.

Discussion

Findings

The aim of this research was to quantify and categorise team playstyles at UEFA Euro 2024 and assess the relationships between clusters and KPIs. The four clusters identified include possession-based high-pressing teams (cluster 0), possession-based counterpressing teams with a wide attacking focus (cluster 1), low block counterattacking teams (cluster 2), and low block direct possession teams (cluster 3). When averaging each cluster's performance data, it is revealed that cluster 0 is the most effective playstyle, while cluster 2 is the least effective. A Kruskal-Wallis test revealed a statistically significant difference in goals conceded per 90 between clusters.

A key driver of the effectiveness of teams in cluster 0 is that they record the lowest proportion of defensive actions in their defensive third. This is achieved by controlling possession in their midfield and attacking thirds, which forces their opponents to defend deeper. Naturally, these teams recorded the shortest median pass length in all three thirds. Shorter passes in central areas allows teams to counterpress more effectively when they lose the ball, as the players are more closely positioned around the transition moment (FIFA, 2023).

My findings are consistent with Ghezzi's (2024) from the Swiss Super League 2022/23 season. According to Ghezzi (2024), St Gallen recorded the least number of events in a low block, the highest amount of counterpress moments, and the highest number of in-possession events in the attacking third. However, despite St Gallen scoring the second most goals in the season, they finished 6th in the league. The league winners, BSC Young Boys, ranked highly for high-pressing and in-possession events in the attacking third, but they also ranked highly in the build-up and progression phases. This indicates that controlling possession across all three thirds creates more goal-scoring opportunities but also allows for more defensive solidity.

The resulting clusters identified support the existing literature that possession-based playstyles are positively correlated with creating goal-scoring opportunities. Research from Plakias (2024) suggests that teams who utilise short passing and occupy central areas have a higher probability of winning, whereas teams who frequently cross the ball have a lower probability of winning. This is evident in my results, where the possession-based clusters convert more goals, and have higher xG. Cluster 3 contains the highest proportion of passes into attacking wide areas and the highest median pass length in all thirds and ranks low on goals scored per 90. Furthermore, Plakias (2024) identifies that high-pressing significantly increases a team's chances of winning. My findings support this as the clusters demonstrate a positive correlation between high-press scenarios with goal and xG differences.

Interestingly, cluster 2 ranks the highest in terms of counterattacks per 90 and the lowest in terms of goals and xG per 90. This contradicts research from Plakias (2023), which identified swift counterattacks as being more effective in creating goal-scoring opportunities compared to less direct possession sequences. However, other factors may impact the effectiveness counterattacks. For example, cluster 2 records the lowest proportion of counterpress and high-press scenarios and ranks first for low-block. Additionally, teams in cluster 2 have low passes per 90 and poor pass accuracy %. This indicates that, despite teams in cluster 2 counterattacking more frequently, they are mostly defending within their defensive third and rarely pressing the opposition high up the pitch. This signals that when they win the ball, these teams are too far from the opponent's goal to create high quality chances. Their poor passing metrics suggest that they are unable to control possession and progress the ball up the pitch which would allow them to operate closer to the opponent's goal. In contrast, cluster 0 record fewer counterattacks per 90, but they significantly outperform cluster 1 in their passing and pressing metrics, which enables them to produce high xG counterattack scenarios. This means a high-pressing approach out-of-possession is effective for reducing chances conceded but also results in more opportunities to win possession in advanced areas where counterattacks can be more dangerous. On the other hand, a more pragmatic low-block approach forces teams to defend deeper and rely on longer, less effective transitions that are easier for opponents to defend against.

Significance and Implications

The Kruskal-Wallis test results reveal that among the four KPIs measured, goals conceded per 90 minutes showed the only statistically significant relationship with cluster classification. This finding suggests that playstyle significantly influences defensive performance, while having less impact on team's ability to score goals. A potential reason for this is because defensive structures are more systematic compared to attacking situations which involve more creativity and are less predictable. Based on these results, coaches should prioritise a clear defensive game model as the foundation of their playstyle, as it provides more reliable performance outcomes.

These results offer coaches and analysts an objective overview of playstyle effectiveness from UEFA Euro 2024. Most notably, there is a clear link between high-pressing teams and success. Teams with a higher proportion of passes into central midfield, central attacking zones, and the half spaces rank higher in performance, but this is only when they also complete a high volume of short and accurate passes. Therefore, when constructing a playstyle, a manager or director of football should look to combine high-pressing and short passing in central areas, as this approach consistently outperforms low-block, direct and counterattacking styles. As a result, clubs should prioritise recruiting and developing players with strong short passing and pressing capabilities.

My analysis contributes to data-driven methods for evaluating playstyles in football by clustering event and match phase data, while separating performance outcomes and minimising team-quality bias, offering a reliable method for measuring tactical behaviours. With this approach, football clubs and national teams can assess their playstyle against successful benchmarks such as cluster 0 and identify areas for improvement which inform training periodisation and recruitment strategies.

Limitations

The data used in this project raises several limitations. Although StatsBomb are a quality data provider in football, Ghezzi (2024) notes that 10.5% of the total event data used in his study did not have the corresponding 360 freeze-frames. Unlike Ghezzi (2024), I did not incorporate a workaround for this issue, and I only analysed events containing freeze-frame data. Therefore, the events with missing freeze-frame data could not be classified into defensive phases – reducing the accuracy and reliability of the results. Also, the freeze-frame data does not capture players outside of the camera shot, which can affect the accuracy of out-of-possession results. Another issue is that the limited sample of 24 teams caused overfitting. The GMM assigned each team to a cluster with 100% certainty, rather than producing probabilistic classifications of overlapping styles which would be expected from the data. A smaller sample size also reduces the statistical significance of the results, with only goals conceded per 90 producing a statistically significant difference between clusters. Another limitation is that by averaging the data between teams in each cluster, the data for cluster 3, which has only 3 teams, will have a higher variance than the data for cluster 1, which has 10 teams – this may skew the results of smaller clusters giving more unreliable results.

In terms of the methodology, I have used a simplified version of Ghezzi's (2024) match phases that captures more detailed location data in-possession, but my model produces a less comprehensive assignment of overall match phases across the tournament. As a result, the model may have missed complex tactical behaviours within the data that more detailed match phases analysis would capture. Another limitation of the methodology is that unlike Gleeson (2024), my analysis does not account for tactical variation in a team's playstyle – instead the teams analysed in this research are categorised into rigid styles. As Gleeson (2024) highlights, tactics often vary depending on game state or opposition. For example, it is unrealistic that a team employing a low block strategy would maintain this if they were losing or if they are playing the lowest ranked team in the competition.

Future Research

Future research should utilise season long data from multiple leagues to generate a robust dataset of 100+ clubs to uncover more complex tactical behaviours and prevent overfitting. The feature set could be enhanced with the incorporation of full tracking data that captures comprehensive match phase analysis and more detailed metrics in each phase such as player positioning (and opponent positioning) in the build-up phase. Integrating off-the-ball events such as FIFA's offering to receive and defensive pressure metrics (FIFA, 2022) could also capture more complex patterns within tactical behaviours. The incorporation of game state and opposition variables may also provide more insight to playstyle variation. Measuring set-pieces may also uncover any correlation or variation between set-piece behaviours and overall playstyle.

Conclusion

The aim of this project was to develop an objective, data-driven approach to categorising and measuring playstyles in football. I employed GMM to cluster team playstyles from UEFA Euro 2024 based on match phases and non-team-quality biased metrics generated from StatsBomb data. I then assessed the effectiveness of the resulting clusters by comparing cluster classification and KPI outcomes such as goals and expected goals per 90.

Analysis of 25 features produced four distinctive clusters representing different playstyles at UEFA Euro 2024. The findings suggest that a combination of high-pressing and a high volume of short and accurate passes in central areas in the midfield and attacking thirds was the most effective playstyle. On the other hand, a counterattacking low block approach resulted in the worst performance outcomes, suggesting that counterattacks are more effective when combined with a high-press style. A statistically significant relationship was observed between playstyle and goals conceded per 90. This indicates that playstyle has a significant influence on defensive performance, while having a lesser impact on attacking performance.

Based on these findings, football clubs, national teams, and coaches should prioritise high-press and counterpress strategies as the foundation of their playstyle while playing frequent short and accurate passes in central midfield and attacking areas. In addition, the methodological approach of this research provides analysts with a valuable framework to objectively categorise and measure playstyles for use in opposition analysis, managerial decision-making, or the production of longitudinal performance reports.

Future research should incorporate a larger dataset from multiple leagues to uncover more playstyles and prevent GMM overfitting. The feature set could be enhanced further with the inclusion of full tracking data in match phases, off-the-ball metrics such as offers or pressures, and set-piece analysis. Furthermore, introducing game state and opposition variables would provide more context to the playstyles analysed, and potentially reveal optimal tactics to counter each playstyle.

References

Akaike, H. (2003). A new look at the statistical model identification. *IEEE transactions on automatic control, 19(6)*, 716-723.

Amatria, M. M. (2019). Analysis of successful offensive play patterns by the Spanish soccer team. . *Journal of human kinetics, 69*, 191.

Brooks, J. K. (2016). Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 9(5)*, 338-349.

Chang, Y. (2025). A Survey: Potential Dimensionality Reduction Methods. *arXiv preprint arXiv:2502.11036*.

FA. (2021, December 21). *FA Cup Counter Attacking Case Study 2020/21*. Retrieved from England Football Learning: https://community.thefa.com/coaching/b/insights-analysis-blogs/posts/the-importance-of-transition-fa-cup-2020-21-counter-attacking-case-study

Fernández, J. B. (2019). Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. . *13th MIT Sloan Sports Analytics Conference (Vol. 2)*.

FIFA. (2021, November 8). *High press*. Retrieved from FIFA Training Centre: https://www.fifatrainingcentre.com/en/game/game-analysis/out-of-possession/high-press.php

FIFA. (2022, October 19). *Part 2.1: Possession-based vs defensive styles of play.* Retrieved from FIFA Training Centre.: https://www.fifatrainingcentre.com/en/environment/fifa-research/high-performance/possession-analysis/part-2-1-possession-based-v-defensive-styles-of-play-new.php

FIFA. (2022, November 24). *The FIFA Football Language*. Retrieved from FIFA Training Centre: https://www.fifatrainingcentre.com/en/game/performance-analysis/football-language-analysis/the-fifa-football-language.php

FIFA. (2023, July 27). *Germany: Ruthless counter-press & connection in possession*. Retrieved from FIFA Training Centre: https://www.fifatrainingcentre.com/en/game/tournaments/fifa-womens-world-cup/2023/team-analyses/germany_ruthless_counter_press_complemented_by_build_up_structure.php

Gagliardi, A. (2022). *The Clustering Project: A New Framework for Smarter Scouting.* . Soccerment Research.

Ghezzi, P. &. (2024). Match phases in practice. *StatsBomb*.

Gleeson, M. (2024). Modelling Team Play Style Using Tracking Data to Evaluate the Effectiveness of Variance in Tactical Behaviour. *StatsBomb*.

Gollan, S. F. (2018). Characterising game styles in the English Premier League using the “moments of play” framework. . *International Journal of Performance Analysis in Sport, 18(6)*, 998-1009.

González-Ródenas, J. A.-M. (2020). The effect of contextual variables on the attacking style of play in professional soccer. *Journal of Human Sport and Exercise, 16(2)*, 399-410.

Hewitt, A. G. (2016). Game style in soccer: what is it and can we quantify it? *International Journal of Performance Analysis in Sport, 16(1)*, 355-372.

Jolliffe, I. a. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, 374(2065)*, 20150202.

Kempe, M. V. (2014). Possession vs. direct play: evaluating tactical behavior in elite soccer. . *International Journal of Sports Science, 4(6A)*, 35-41.

Low, B. C. (2020). A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine, 50*, 343-385.

McLachlan, G. a. (2000). *Finite mixture models.* John Wiley & Sons.

Pappalardo, L. C. (2019). PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. . *ACM Transactions on Intelligent Systems and Technology (TIST), 10(5)*, 1-27.

Partington, M. a. (2024). A deconstruction of coaching philosophy. *Sports Coaching Review*, 1-23.

Plakias, S. M. (2023). Identifying soccer teams' styles of play: a scoping and critical review. *Journal of Functional Morphology and Kinesiology, 8(2)*, 39.

Plakias, S. T. (2024). Tactical situations and playing styles as key performance indicators in soccer. *Journal of Functional Morphology and Kinesiology, 9(2)*, 88.

Power, P. R. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. . *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1605-1613.

Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics, 20*, 53-65.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Stanojevic, R. a. (2016). Towards data-driven football player assessment. . *IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 167-172.

StatsBomb. (2025). *StatsBomb Products*. Retrieved from https://www.hudl.com/en_gb/products/statsbomb

Trower, M., Graham, N., Cottrell, N., & Hengster, Y. (2023). Clustering women's football players: Identifying functional patterns for performance optimisation. *StatsBomb*.

Wenger, A. (2021). FIFA Technical Advisory Panel: The Future of Football. Zurich.

Y. Zhao, A. K. (2019). Regularized Gaussian Mixture Model for High-Dimensional Clustering. *IEEE Transactions on Cybernetics, vol. 49, no. 10*, 3677-3688.

Appendix

Figure 1: Example of StatsBomb 360 data freeze-frames (Source: StatsBomb, 2025)

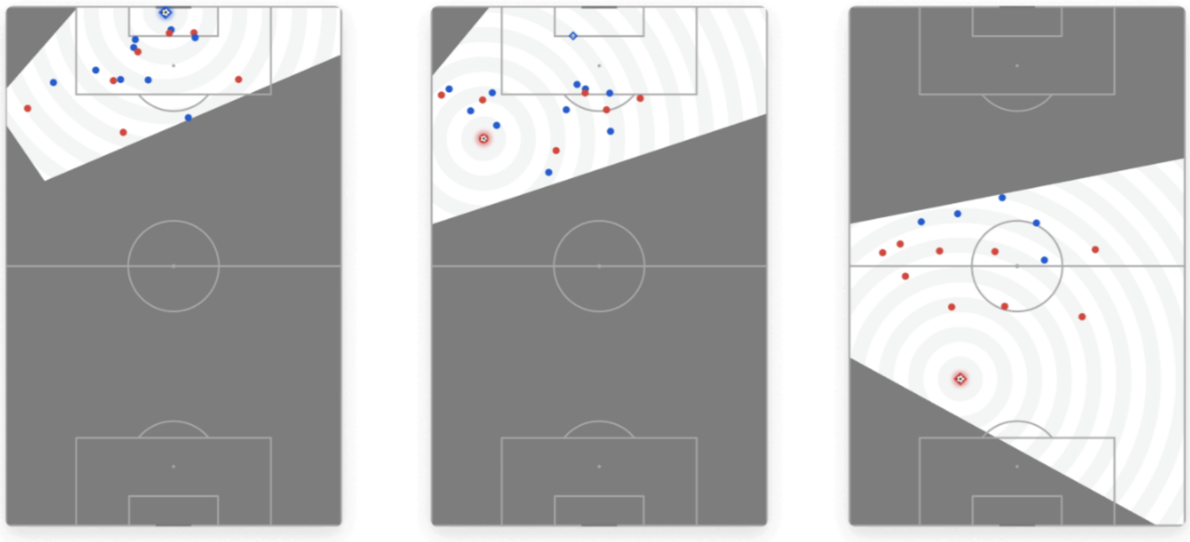


Figure 2: Pitch zones for identifying in-possession match phases and attacking style

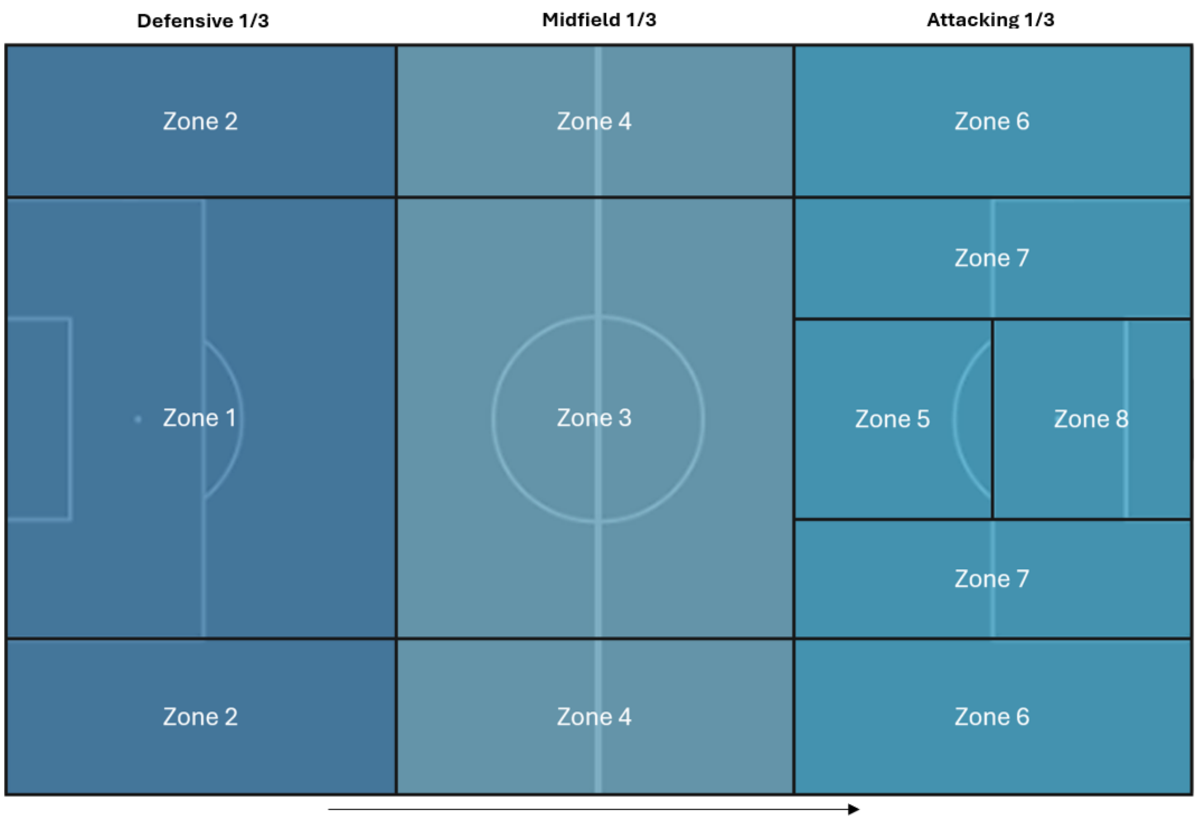


Table 1: Full feature set including match phase and definitions

Variable	Phase	Definition
Goals per 90	In-Possession KPI	Total number of goals scored per 90 minutes.
xG per 90	In-Possession KPI	Total expected goals (xG) generated per 90 minutes, based on StatsBomb's model that calculates the probability of a shot resulting in a goal.
Counterattacks per 90	In-Possession Transition	Total number of Counterattacks per 90 minutes, denoted by a possession started with an open play turnover outside the counter-attacking team's final third, moved at least 75% direct towards goal (as measured by StatsBomb's possession chain metrics), and travelled at least 18 yards towards goal.
Carries per 90	In-Possession	Total number of times a player moves the ball at their feet while maintaining control (includes movement while standing still) per 90 minutes.
Passes per 90	In-Possession	Total number of passes completed between teammates per 90 minutes.
Possession %	In-Possession	The proportion of a team's total passes relative to the combined total passes of both teams.
Pass Accuracy %	In-Possession	The percentage of a team's successful passes out of their total passes.
Aerial pass %	In-Possession	The percentage of a team's aerial passes (ball goes above shoulder level at peak height) out of their total passes.
Forward pass %	In-Possession	The percentage of a team's forward passes (angle of pass is >=-0.79 and <=0.79) out of their total passes.
Sideways pass %	In-Possession	The percentage of a team's sideways passes (angle of pass is >0.79 and <=2.36 or <-0.79 and >=-2.36) out of their total passes.
Backward pass %	In-Possession	The percentage of a team's backwards passes (angle of pass is >2.36 and <-2.36) out of their total passes.
Median pass length (Def 1/3)	In-Possession	The median pass length of passes into their Defensive 1/3.
Median pass length (Mid 1/3)	In-Possession	The median pass length of passes into their Midfield 1/3.
Median pass length (Att 1/3)	In-Possession	The median pass length of passes into their Attacking 1/3.
Passes (Def 1/3 central)	In-Possession	The proportion of a team's passes that are played into the central zone in their Defensive 1/3 (Zone 1).
Passes (Def 1/3 wide)	In-Possession	The proportion of a team's passes that are played into the wide zones in their Defensive 1/3 (Zone 2).
Passes (Mid 1/3 central)	In-Possession	The proportion of a team's passes that are played into the central zone in their Midfield 1/3 (Zone 3).
Passes (Mid 1/3 wide)	In-Possession	The proportion of a team's passes that are played into the wide zones in their Midfield 1/3 (Zone 4).
Passes (Att 1/3 central low)	In-Possession	The proportion of a team's passes that are played into the central low zone in their Attacking 1/3 (Zone 5).
Passes (Att 1/3 wide)	In-Possession	The proportion of a team's passes that are played into the wide zones in their Attacking 1/3 (Zone 6).
Passes (Att 1/3 half-space)	In-Possession	The proportion of a team's passes that are played into the half space zones in their Attacking 1/3 (Zone 7).
Passes (Att 1/3 central high)	In-Possession	The proportion of a team's passes that are played into the central high zone in their Attacking 1/3 (Zone 8).
Goals conceded per 90	Out-of-Possession KPI	Total number of goals conceded per 90 minutes.
xGA per 90	Out-of-Possession KPI	Total expected goals (xG) conceded per 90 minutes, based on StatsBomb's model that calculates the probability of a shot resulting in a goal.
Counterpress %	Out-of-Possession Transition	The percentage of opposition open play turnovers followed by a team performing a pressing action (pressure, dribbled past, 50-50, duel, block, interception, or foul committed) recorded within 5 seconds. Proportionate to the total out-of-possession and defensive transition phases.
Regroup %	Out-of-Possession Transition	The percentage of out-of-possession and defensive transition phases where a team is defending an opposition counterattack.
High-Press %	Out-of-Possession	A team is out-of-possession and has at least four players in their Attacking 1/3, proportionate to the total out-of-possession and defensive transition phases.
Mid Block %	Out-of-Possession	A team is out-of-possession and has at least four players in their Midfield 1/3, proportionate to the total out-of-possession and defensive transition phases.
Low Block %	Out-of-Possession	A team is out-of-possession and has at least six players in their Defensive 1/3, proportionate to the total out-of-possession and defensive transition phases.

Figure 3: Graph showing cluster selection based on AIC and BIC criteria

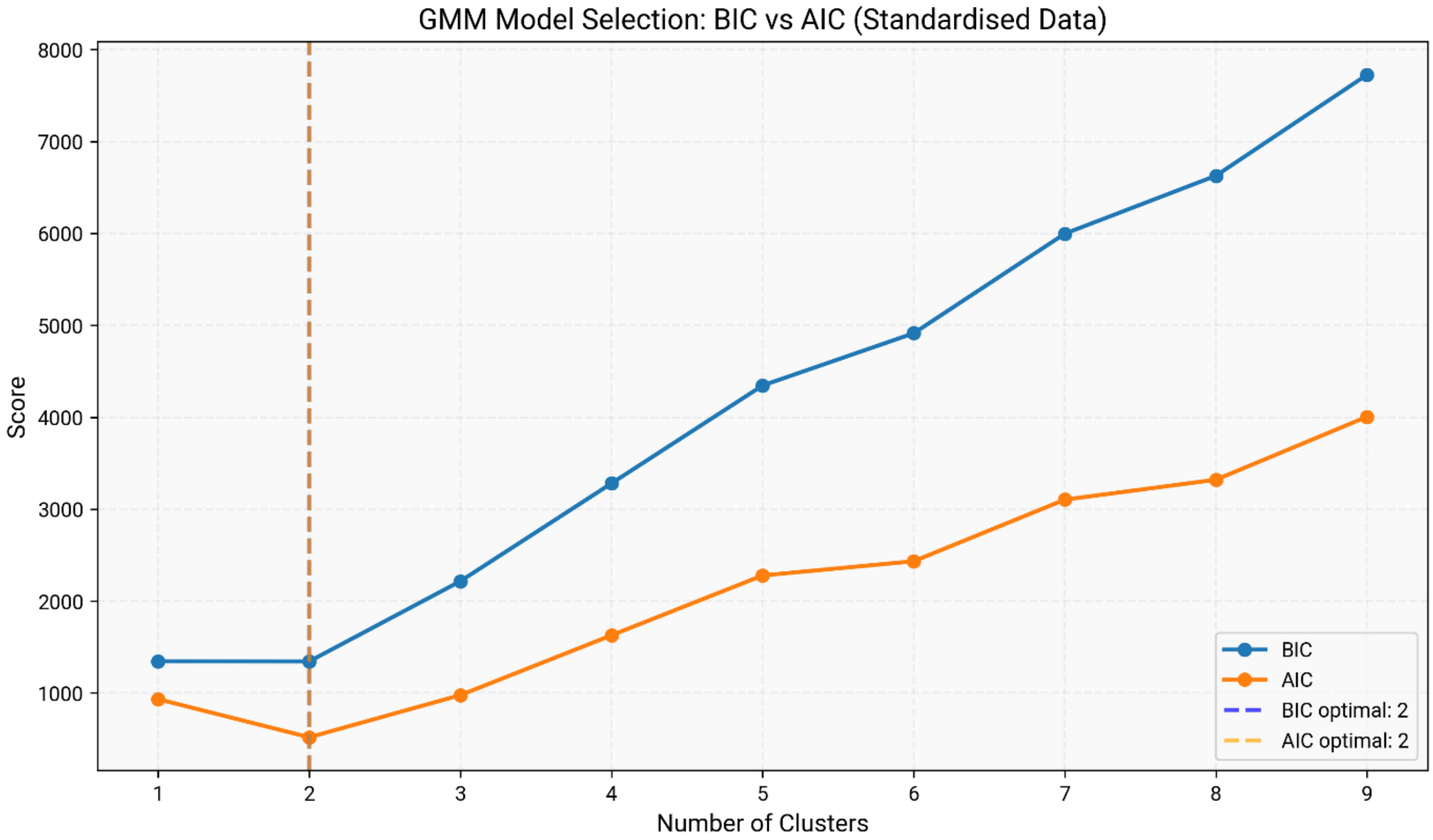


Table 2: Table showing team cluster assignment and KPIs

Cluster	team	Goals per 90	xG per 90	Goals conceded per 90	xGA per 90
0	Austria	1.43	1.45	1.18	1.34
0	Croatia	0.61	1.82	1.88	1.21
0	Denmark	0.48	0.93	0.96	1.12
0	Germany	2.08	1.6	0.48	0.73
0	Italy	0.71	0.78	0.96	0.98
0	Spain	1.85	1.39	0.38	0.82
1	Belgium	0.48	1.04	0.23	0.94
1	England	1.51	1.23	1.07	1.07
1	France	0.91	1.71	0.83	1.08
1	Netherlands	1.43	1.17	0.94	0.93
1	Portugal	1.44	2.41	1.29	1.63
1	Serbia	0.32	0.65	0.64	0.85
1	Slovakia	0.9	0.83	1.1	1.39
1	Switzerland	1.91	1.41	1.43	1.43
1	Turkey	1.5	1.22	1.14	1.53
1	Ukraine	0.63	0.95	1.27	0.8
2	Albania	0.93	0.68	1.27	1.34
2	Georgia	0.95	1	1.91	2.48
2	Hungary	0.61	1.04	1.59	1.19
2	Poland	0.96	1.13	1.92	1.6
2	Scotland	0.32	0.32	2.25	1.21
3	Czech Republic	0.95	1.45	1.26	1.34
3	Romania	0.96	0.9	1.44	1.36

Figure 4: Team cluster assignments in a two-dimensional PCA plot

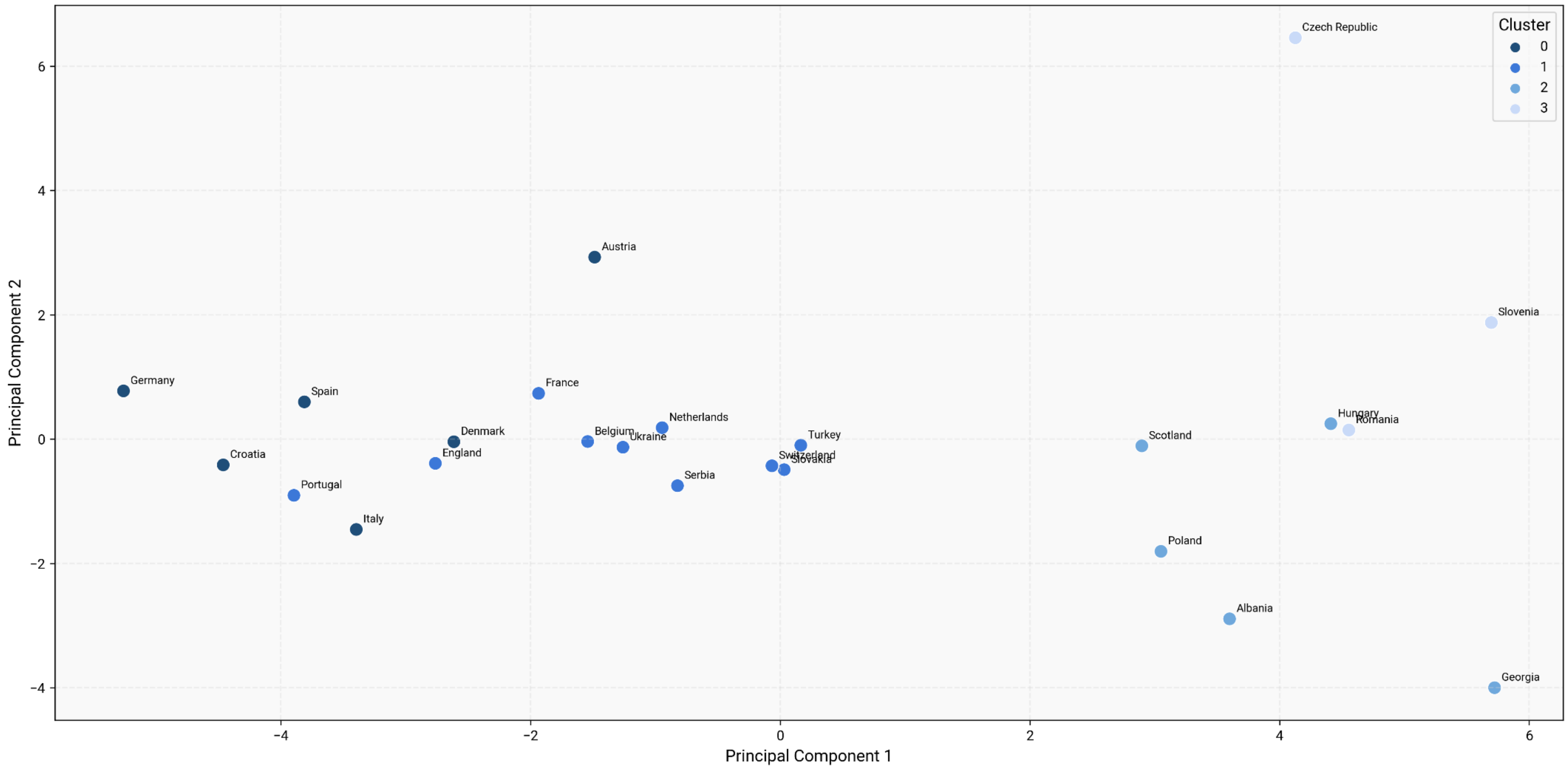
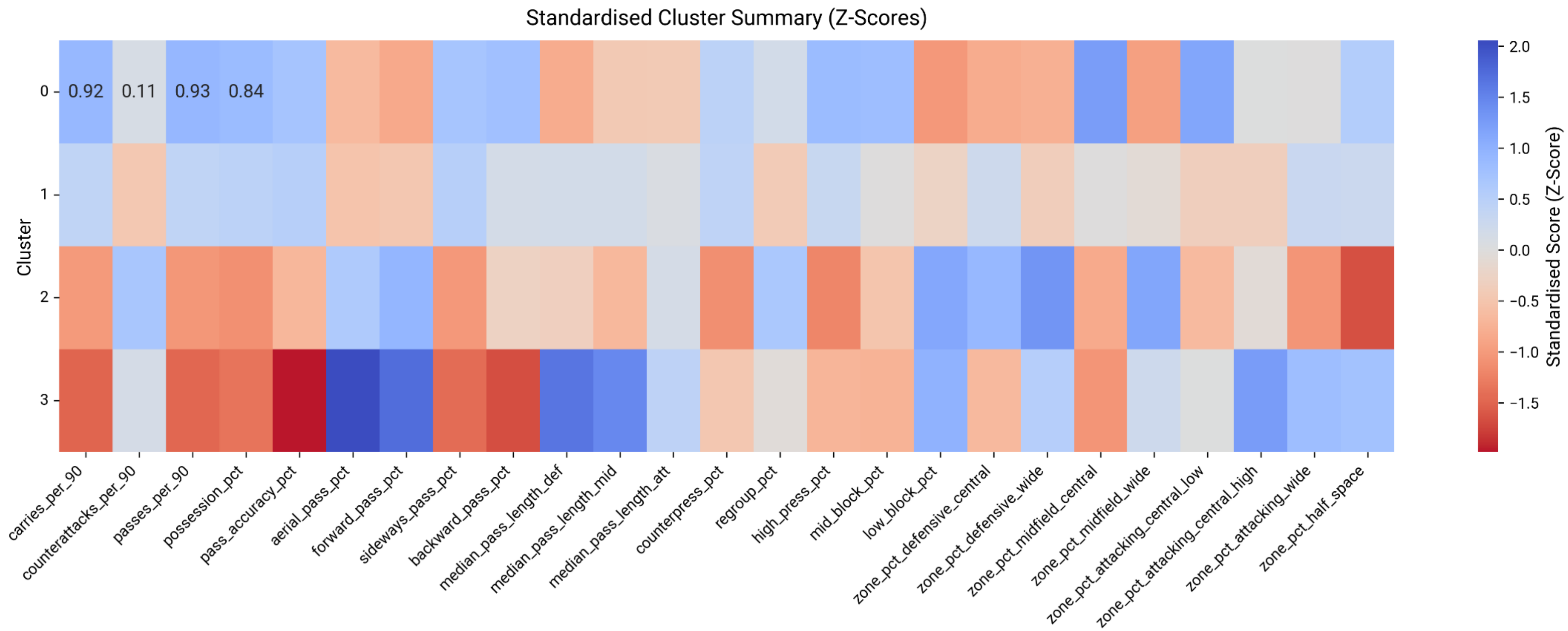
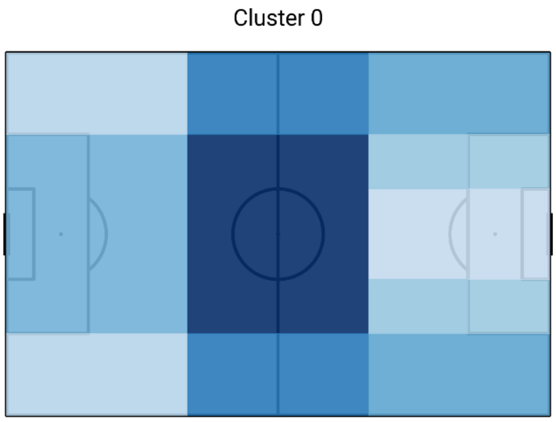


Figure 5: Heatmap showing cluster assignments and standardised feature performance



Figures 7 & 8: Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 0



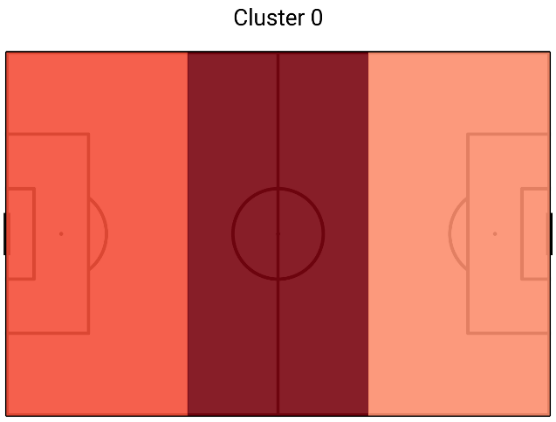
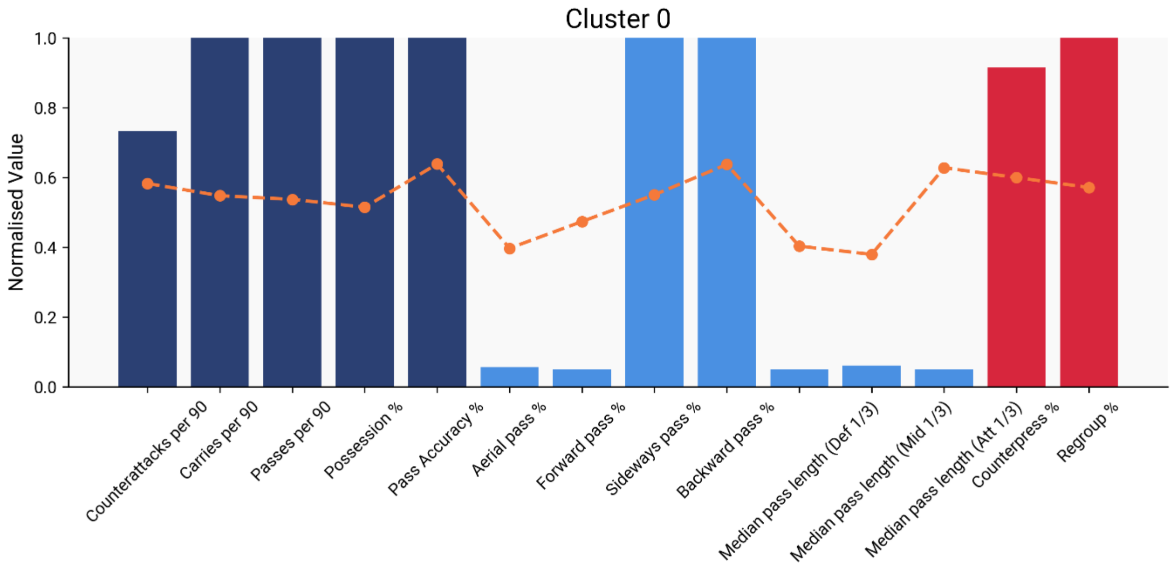
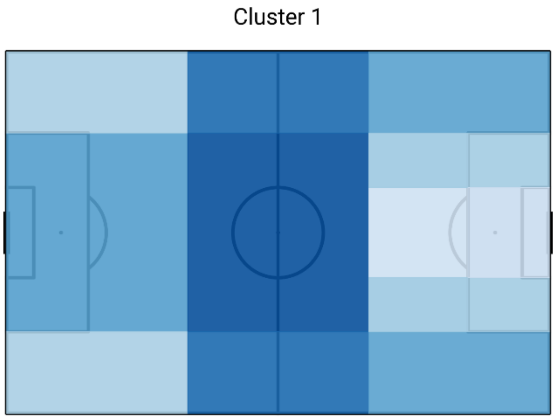


Figure 9: Chart showing cluster 0 performance with normalised metrics



Figures 10 & 11: Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 1



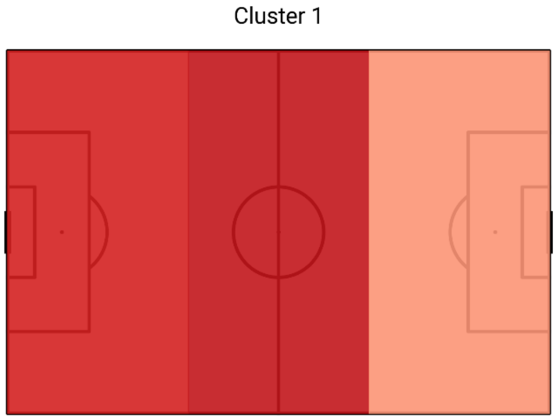
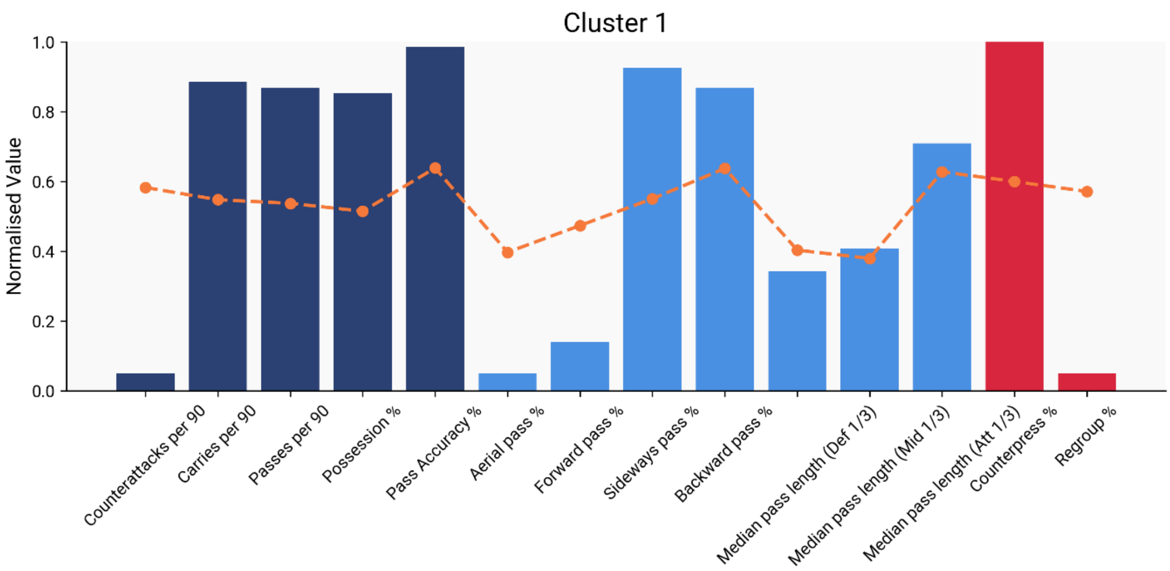
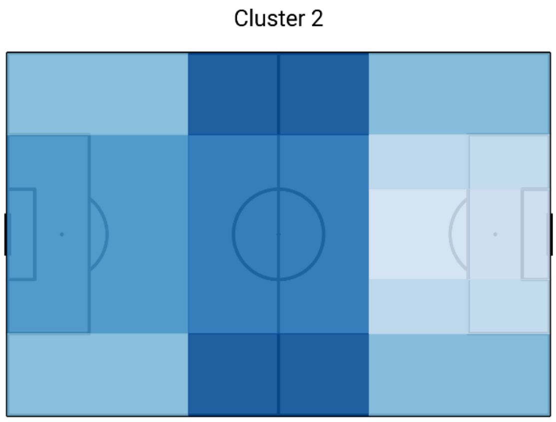


Figure 12: Chart showing cluster 1 performance with normalised metrics



Figures 13 & 14: Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 2



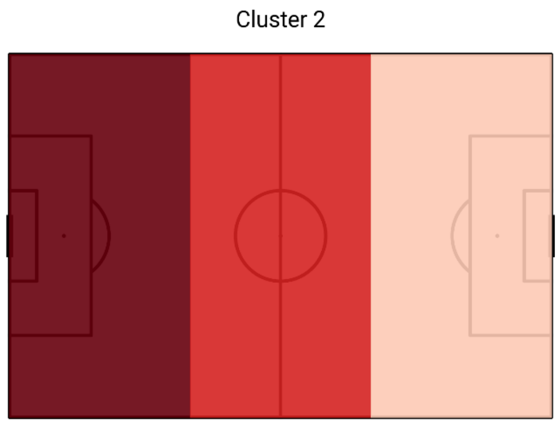


Figure 15: Chart showing cluster 2 performance with normalised metrics

Figures 16 & 17: Heatmaps showing the proportion of passes into each zone and the proportion of defensive phases from teams in cluster 3

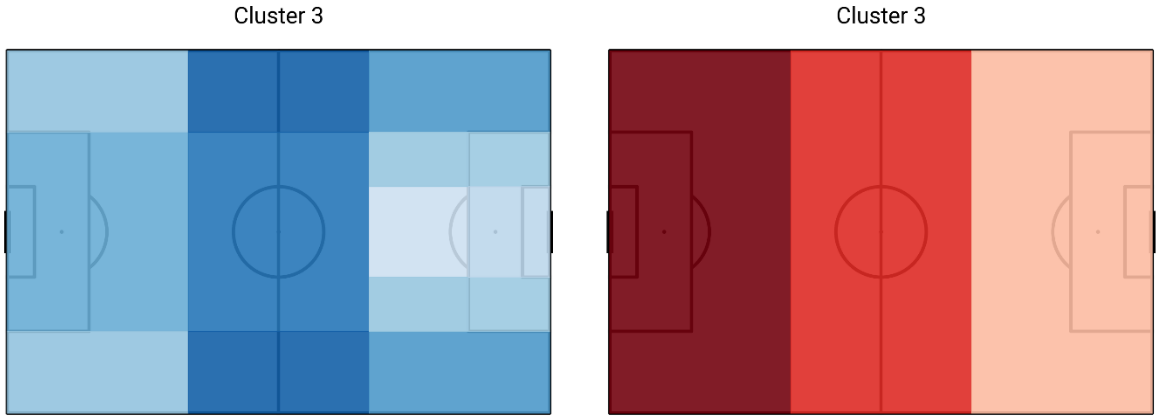


Figure 18: Chart showing cluster 3 performance with normalised metrics

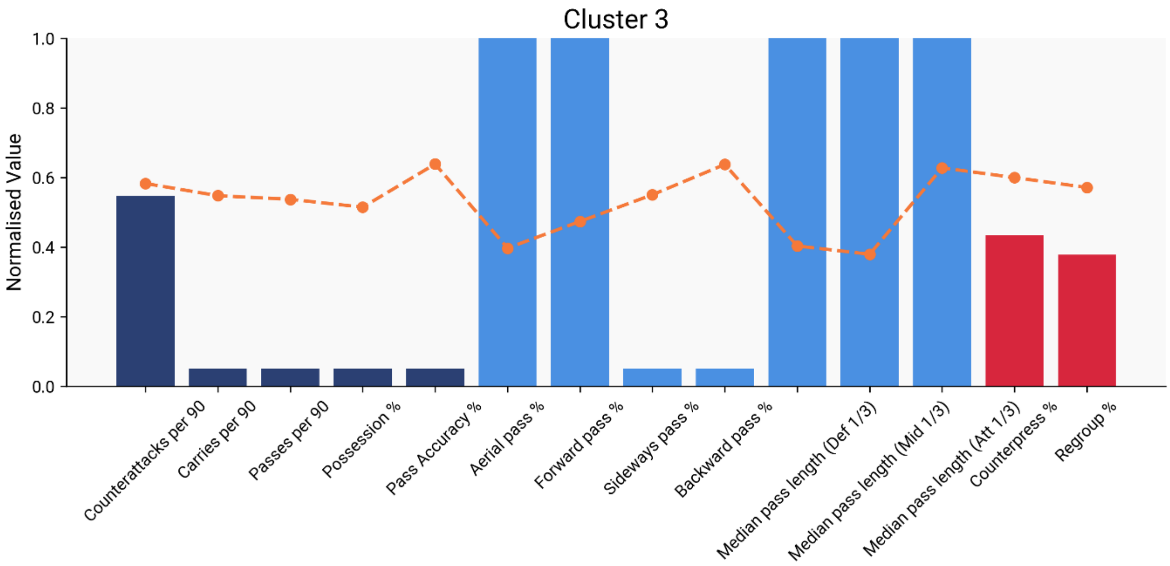


Figure 19: Chart showing cluster classification and performance output