Statsbomb Football Data Analysis 101 Based on Jupyter Notebook

Moureen Lyn

December 31, 2025

Abstract

This paper serves as a simple starting point for the beginners on how to utilize the free data provided by Statsbomb to analyze the football match. The programming language selected is Jupyter Notebook due to the easier operation and visualization.

The paper includes the author's examination across nine modules: 1, Basic introduction on the football data and the differences among various providers 2, Overview of the data providers and how to obtain the data, including the location of the repository of the free data. 3, Structure of Statsbomb free data, and different methods of reorganization and key pitfalls in the analysis. 4-7, The examples of data analysis on shot comparison, pass evaluation, action assessment and player physics diagnosis respectively. 8, The purpose to conduct data analysis and the methodology to improve the quality control 9. The future plan and desired accomplishments

As a beginner as well, the author shared own journey of learning, in the hope that the readers can ramp up quickly and generate more ideas.

Dedication

I would like to dedicate this paper to my favorite player Luka Modrić, who inspires me to conduct this set of studies to identify any meaningful quantitative characteristics to explain the intelligence, quality, tactics even strategy of the football, therefore provide at least a supplemental way to improve it. This most popular sport should not be considered only as the scores and odds of a pure entertainment business or an emotional outlet. Let's build a better framework of science and engineering as a domain that benefits from structured analysis, methodological discipline, and continuous improvement.

Acknowledgments

All references will be included in the paper, I would like to sincerely appreciate the work and education that have been conducted by David Sumpter(Author of Soccermatics), Andrew Rowlinson (Author of mplsoccer) and Jan Van Haaren for his soccer analytics review from 2020-2024.

Module 1 Basic Introduction of the Football Data

Football data refers to structured information collected during matches and training that can be used to analyze player performance, team tactics, and game dynamics. Today, the football team analysts can access event data, tracking data, and physical data — each offering unique insights into the game. The data could be divided into three most important and used categories[1]:

Event data is related to the action on the ball such as pass, shot, dribble, tackle, interception etc. The data includes the action type, time stamp, the players involved (maybe from both teams), the description of the actions, the positions of the action and the outcomes of the action. The purpose for this type of data is to understand the player's decision/action on the ball and the team tactical breakdown. Since it is very focused on the ball, the team space control and the off ball actions from the players are most likely ignored. The commercial providers include **StatsBomb**, **Opta**, **Wyscout**, **ReSpo.Vision** etc. **Statsbomb** also provides some improvement called 360 data, that includes several extra players captured around the player with on the ball action, but the information related to extra players was unclear and uncomplete.

Tracking data is related to the position of the ball and all players. The data includes the spatial information of all players and the ball alongside the time stamp. The tracking data is high frequency data; it contains millions of data points compared to event data with only thousands of data points. Most tracking data is not directly related to the actions happened on the ball, however, the existence of them could explain how the space is occupied and controlled; how the team is coordinated for off ball action such as the intensity and rhyme of defense and offense. But the author agrees, most of the off ball movement might not generate the immediate impact, the thorough analysis of them may provide less immediate value for understanding the match. The commercial providers include **Second Spectrum**, **Tracab**, **SkillCorner, Metrica Sports.**

Physical data is related to the athlete output of the player. The data is used to measure how the players move with speed, acceleration, distance covered, therefore any indication on fatigue and metabolism etc. The data is normally obtained by special wear on devices. The track data could be converted to the physical data through some assumptions.

It can be concluded that combining both event data and tracking data provides the insights for the action evaluations and spatial controls of both teams.

Module 2 Access the data

Most of the commercial providers have the details listed on their website regarding their data. Table 1 is a list to summarize the general comparison, the pricing is just an estimation, but at least provide some ball park understanding on the cost associated. Moreover, most data providers operate with B2B business mode, the individual data analyst might not have opportunity to obtain their desired block of data.

Table 1 Common Football Data Provider and Estimated Pricing

Provider	Data Type	Estimated Pricing (internet searching)
Opta (Stats Perform)	Event, tracking, physical	$10,000–$100,000+/year
StatsBomb	Event (rich), freeze frames	$5k–$50k+/year
Second Spectrum	Tracking, contextual events	$100,000+/year
SkillCorner	Broadcast-derived tracking	$20,000–$80,000/year
Tracab	Optical tracking, physical	$50,000–$150,000+/year
ReSpo.Vision	3D tracking from video	Custom pricing
Metrica Sports	Tracking, event, video sync	€5k–€50k+/year
Wyscout (Hudl)	Event, video	$1,000–$10,000/year
InStat	Event, video, physical	$500–$5,000/year
Sportmonks	Event, odds, fixtures	$49–$299/month
Enetpulse	Event, live scores	$1,000–$10,000/year

Based on this list, the beginner is probably overwhelmed since it is hard to decide who to start and also what to start? Fortunately, there are some free data providers, I sincerely believe they are best starting points for anybody who has the interest to at least understand what they will deal with and what they will expect to learn.

Table 2 summarizes all the free data providers and their repositories.

Table 2 Free Football Data Repository

Source	Data Type	Repository Link
StatsBomb Open Data	Event data with freeze frames	github.com/statsbomb/open-data
Metrica Sports Sample	Tracking + event + video sync	github.com/metrica-sports/sample-data
DataHub.io	Global football statistics	datahub.io/blog/football-data
FootyStats	Team and player performance stats	footystats.org/download-stats-csv

FBref (Opta-powered)	Match and player stats	fbref.com/en
Football-Data.co.uk	Historical match results & odds	football-data.co.uk
Kaggle Datasets	Mixed (event, match, ratings)	kaggle.com/datasets

Among them, StatsBomb Open Data definitely stands out due to the quality (depth and breadth of the matches) and quantity (numbers and coverages of matches) that they provide, especially relate to the matches directly. Most of the data are the raw annotated events of the matches, which I prefer over all the free data providers. It is highly recommended to the beginners to start with their free data.

The StatsBomb data is located in **github**, actually **github** is a good platform for all other football related information. A simple search based on "football" generated the repositories related as shown in Table 3.

Table 3 Github Football Related Repository

Repository	Owner	Link
open-data	StatsBomb	github.com/statsbomb/open-data
sample-data	Metrica Sports	github.com/metrica-sports/sample-data
socceraction	Pieter Robberechts	github.com/ML-KULeuven/socceraction
openfootball	OpenFootball	github.com/openfootball
football_analytics	Edd Webster	github.com/eddwebster/football_analytics

Upon this point as the beginner, one should understand what types of data are provided for the football matches, who are the data providers and where to find the free data to initiate the study. In the next modules, the focus is to explain what types of information are included in the Statsbomb free data, and what analysis would be conducted to deepen the understanding of the football match.

Module 3 Statsbomb Free Data Information and How to Organize Them

StatsBomb Free Data is organized into a directory structure that reflects the hierarchy of competitions and seasons, then matches, finally events and lineups of the matches. To find the exact event and lineup of a certain match, one first uses the competitions.json file to identify the **competition ID** based on the available competitions and the **season ID** based on their corresponding seasons. Next locates the match data stored in the matches folder with the **match ID**, where each subfolder is named after a competition ID and a corresponding season ID. Finally for an individual match, the detailed event data and player lineups are stored in the events and lineups folders respectively, with each file named by its **match ID**. Additionally, for selected matches, StatsBomb provides enhanced spatial context through three-sixty data — freeze-frame snapshots of player positions at key moments — also organized by **match ID**. So most important identifier is the match ID for this set of data. One pitfall is that the season ID is not totally following chronological order as normal people would think, it starts 2017/2018 Europe season as season 1 since that is the first year that they transformed from a blog/consultancy into a data provider. The **2017/2018** season was the first full season they officially collected and sold granular event data. Therefore, in their internal database, that season was assigned the first index.

One then should decide where to store the data, there are two ways to handle them. First is to save them in the local drive, a Jupyter Notebook **download_statsbomb_jsons (match_ids, save_dir)** could be built to download the batches of event data and saved to the desired directory in the hard drive. The match ID (list) and the saved directory(str) are the two parameters required. The base_url for this batch operation is https://raw.githubusercontent.com/statsbomb/open-data/master/data/events/. Second, a parser **Sbopen** provided by **mplsoccer** library could be used, which parses the data based on urls of the event and line up with a certain match ID. An example of code includes the following: from mplsoccer import Sbopen, parser = Sbopen(), df, related, freeze, tactics = parser.event(match_id); lineup = parser.lineup(match_id)

A short summary is provided on what information is included in the event data. All other information could be obtained based on open data description[2].

The original data structure of StatsBomb Data and the data parsed by the Sbopen are somehow different as shown in the

Table 4. Original StatsBomb data of both events and lineup are heavily nested, it is recommended to flatten it by using pd.json_normalize before the analysis.

Table 4 Original Data Structure and Sbopen Parser Structure

Aspect	Original Flattened Data	Sbopen Parser (mplsoccer)
Column Count	~120 columns	~60–80 columns (varies by match and parser version)
Naming Convention	Nested keys with dot notation (e.g., pass.recipient.id)	Flattened and renamed (e.g., recipient_id)
Location Fields	Stored as lists (e.g., location, pass.end_location)	Split into x, y, end_x, end_y
Freeze Frame	Stored as nested JSON under shot.freeze_frame	Parsed into separate DataFrame (freeze_frame_360)
Event Types	type.id, type.name	type_name only (cleaned)
Player & Team Info	player.id, player.name, team.id, team.name	Renamed to player_id, player_name, etc.
Outcome & Technique Fields	Multiple nested fields (e.g., pass.outcome.name, shot.technique.name)	Flattened and standardized
Boolean Flags	Mixed types (under_pressure, counterpress, pass.switch)	Cleaned to consistent True/False
Missing Data Handling	Manual null checks required	Pre-cleaned with consistent NaN or None values
Additional Parsing	Requires manual flattening and renaming	Auto-parsed into multiple DataFrames: events, freeze_frame_360, lineups, etc.

The below are the columns shared in both data frames. Though naming conventions may differ slightly (dots/original data vs underscores/Sbopen):

- id, index, period, timestamp, minute, second, possession, duration
- type_id, type_name
- possession_team_id, possession_team_name

- play_pattern_id, play_pattern_name
- team_id, team_name
- tactics_formation
- player_id, player_name
- position_id, position_name
- pass_recipient_id, pass_recipient_name
- pass_length, pass_angle, pass_height_id, pass_height_name
- under_pressure, counterpress, off_camera
- foul_won_defensive, foul_won_advantage, foul_committed_advantage, foul_committed_penalty, foul_won_penalty
- pass_switch, pass_backheel, pass_cross, pass_shot_assist, pass_assisted_shot_id, pass_goal_assist
- shot_statsbomb_xg, shot_key_pass_id, shot_first_time, shot_redirect, shot_deflected
- block_deflection, block_offensive
- aerial_won, pass_deflected, pass_miscommunication
- ball_recovery_offensive, ball_recovery_recovery_failure
- dribble_overrun
- technique_id, technique_name, body_part_id, body_part_name
- substitution_replacement_id, substitution_replacement_name
- foul_committed_card_id, foul_committed_card_name
- bad_behaviour_card_id, bad_behaviour_card_name
- foul_committed_offensive
- goalkeeper_position_id, goalkeeper_position_name

The Table 5 explains the detailed differences between two data frames.

Table 5 Difference Comparison between Original data and Sbopen data

Category	Original Flattened	Sbopen Parser
Naming Style	Dot notation (e.g., pass.recipient.id)	Underscore (e.g., pass_recipient_id)
Location Fields	location, pass.end_location, shot.end_location (as lists)	x, y, end_x, end_y, end_z (split into columns)
Match Reference	No match_id in event rows	Includes match_id
Freeze Frame	shot.freeze_frame (nested JSON)	Parsed separately into freeze_frame_360
Carry Events	Includes carry.end_location	Not present in Sbopen list

Goalkeeper Actions	Includes detailed fields like goalkeeper.end_location, goalkeeper.type.id, goalkeeper.technique.name, etc.	Only includes goalkeeper_position_id, goalkeeper_position_name
Shot Details	Includes shot.technique.name, shot.body_part.name, shot.type.name, shot.outcome.name	Not listed in Sbopen columns
Duel, Interception, Clearance	Includes duel.type.name, interception.outcome.name, clearance.aerial_won	Not present in Sbopen list
Outcome Fields	Action-specific (e.g., pass.outcome.name, dribble.outcome.name)	Unified as outcome_name
Subtypes	Action-specific (e.g., pass.type.name, shot.type.name)	Unified as sub_type_name
Additional Flags	Includes shot.aerial_won, goalkeeper.body_part.name, foul_committed.type.name	Not present in Sbopen list

The readers need to decide which way they select to organize the data. **Sbopen** seems cleaner and simpler to follow, however, the parser needs to connect the data online. If the localized saved data is where the reader wants to work with, a customized parser may be a way to go, or at least, the reader needs to keep in mind they are different, whatever codes working with data parsed by **Sbopen** might not work for the original data.

First pitfall is the time related information on timestamp, minutes and seconds. It is well known that the football match is divided into the periods, could be up to 5 including final penalty shootout. However different as to normal time series events, the later period normally starts earlier based on time stamp/ minutes and seconds. The reason is the 2nd period, 3rd extra time or 4th extra time always starts 45 minutes, 90 minutes and 105 minutes respectively, while 1st period, 2nd period or 3rd extra time normally ends after 45 minutes, 90 minutes and 105 minutes due to the extra minutes added. If the readers are interested in plotting the whole time periods, extra attention must be paid on how to allocate the correct events corresponding to period rather than solely depending on timestamp.

Second problem is related to the players' names. The event data frame uses the players' formal names, most of them are quite different from the names who are known to the public. The lineup contains the nicknames of the players, those are well known names. It is recommended to build a look up directory to replace all player's names in the event data frame with the nick names at the beginning.

Third issue is on the possession team and team. It is quite common that for one set of events some of the possession team and team are different. This is due to the granularity of Statsbomb data including extensive defensive actions such as pressure, etc. The readers may double check if the situation makes sense for their own tactical analysis.

Module 4 Data Analysis - Shot

Before moving into the real examples of the data analysis, the author highly recommends reviewing David Sumpter's educational website **soccermatics**[3] first, it is one of the best free education tool for anybody who is interested in learning the football analytics. The YouTube channel of **Friends of Tracking**[4] contains a lot of insight as well, but the owners seemed to stop updating for some time. Soccermatics teaches how to conduct the analysis based on quite a few of data providers, but only Statsbomb's data is focused in this paper, in addition to the original applications based on Statsbomb data, some concepts based on other data providers are also reworked to apply to Statsbomb data.

Naturally, since most important event of a football match is shot. The first module of the analysis is on the shot. The Table 6 shows the information could be generated based on Statsbomb data.

Table 6 Shot Analysis Generated by Statsbomb Data

Analysis Type	What to generate	Value for improvement
Shot quantity and quality	Number of shot, who are the player, conversion rate to score, finishing efficiency based on Xg and score	Reveal who are the major shooters and how efficient they are
Shot location and player's technique	Player's favorite location and technique	The offense team to improve the chances by providing the situation or the defense team to reduce the chances by blocking the situation
Team attacking pattern and shot creation	Reveal what are the preferred pattern and spaces to generate the shot	The offense team to improve the chances by generating the pattern and space or the defense team to avoid the pattern and block the space.

Pressure & defensive context	Use freeze frames to measure defensive proximity and goalkeeper positioning	Adds context to shot quality by showing defensive pressure and keeper positioning
Set-Piece effectiveness	Isolate penalties, free kicks, and corners to evaluate conversion rates	Assesses team strength in dead-ball situations, often decisive in close matches

The work flow include loading data, filter the shot event, extract the contextual feature, aggregate and summarize based on player level, match level, team level, visualize the shot map and conduct the advanced analysis such as pressure analysis and set pieces evaluation, finally export and report. The time series of data could be generated based on the period and substitutions. The modular structures are recommended, then various situation could be combined, exported and reported.

An example of a shot map comparison is shown in Figure 1. This is match 18241, the 2014 UEFA final between Real Madrid and Atletico Madrid. More advanced study could be conducted to go deep understanding. An obvious pattern could be identified that Real Madrid shot more from the left and central, there were much less shots from the right. Also comparing to Atletico Madrid, Real Madrid shot more from outside of the penalty area.

Real Madrid (red) and Atlético Madrid (blue) shots withouth final penalty

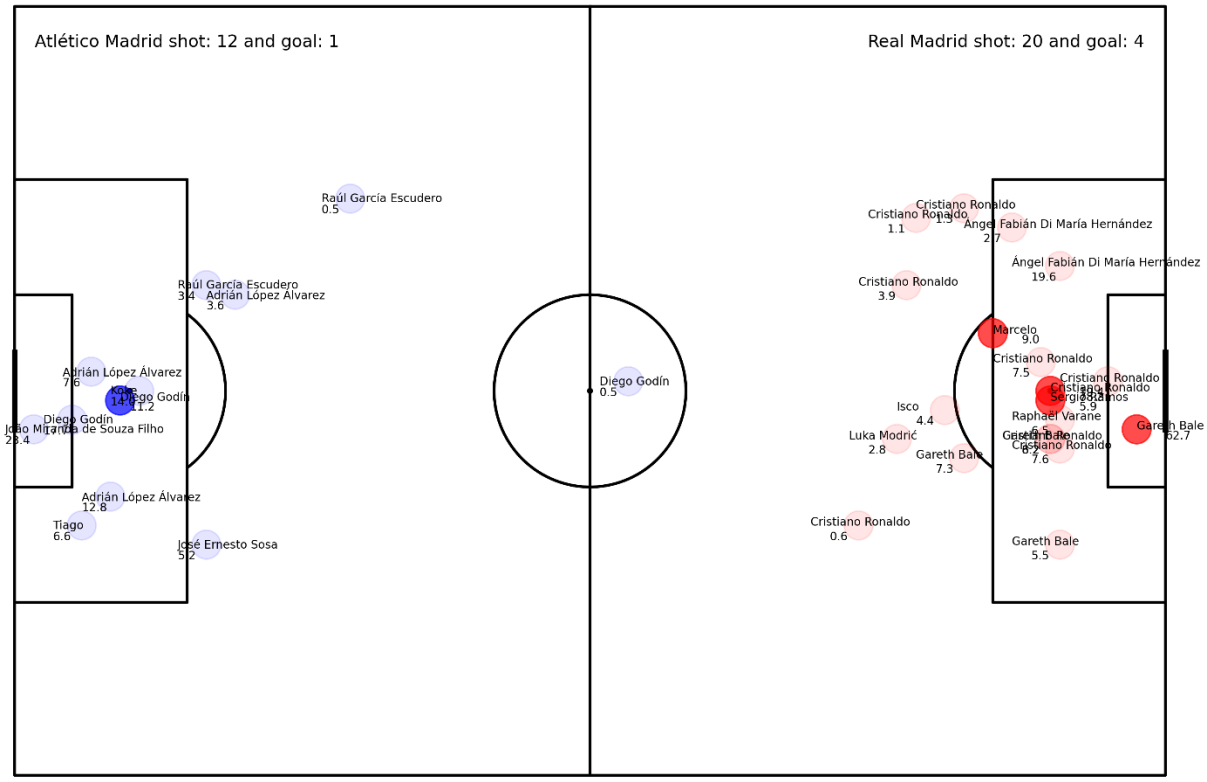


Figure 1 Shot Map of Match 18241 (number specifies Xg*100)

Another example of the shot map of same two teams is shown in Figure 2. This is match 18243, the 2016 UEFA final between Real Madrid and Atletico Madrid. This time, Real Madrid shot more from the right compared to 2 years ago, they still preferred to shoot from the outside of the penalty box maybe due to the extensive blocking from the opposite team. However it seemed the efficiency to shoot from the outside of the penality box was really not impressive. Contrary to the normal instinction, is it appropriate to suggest avoiding shot from the outside of the box, instead doing best to pass or carry into the panelty box, even it might result in losing the possession. If shot from the outside of the penality box would still be the perferred way due to the expected heavy blocking, then the team should specifically train to improve the converation rate.

Real Madrid (red) and Atlético Madrid (blue) shots withouth final penalty

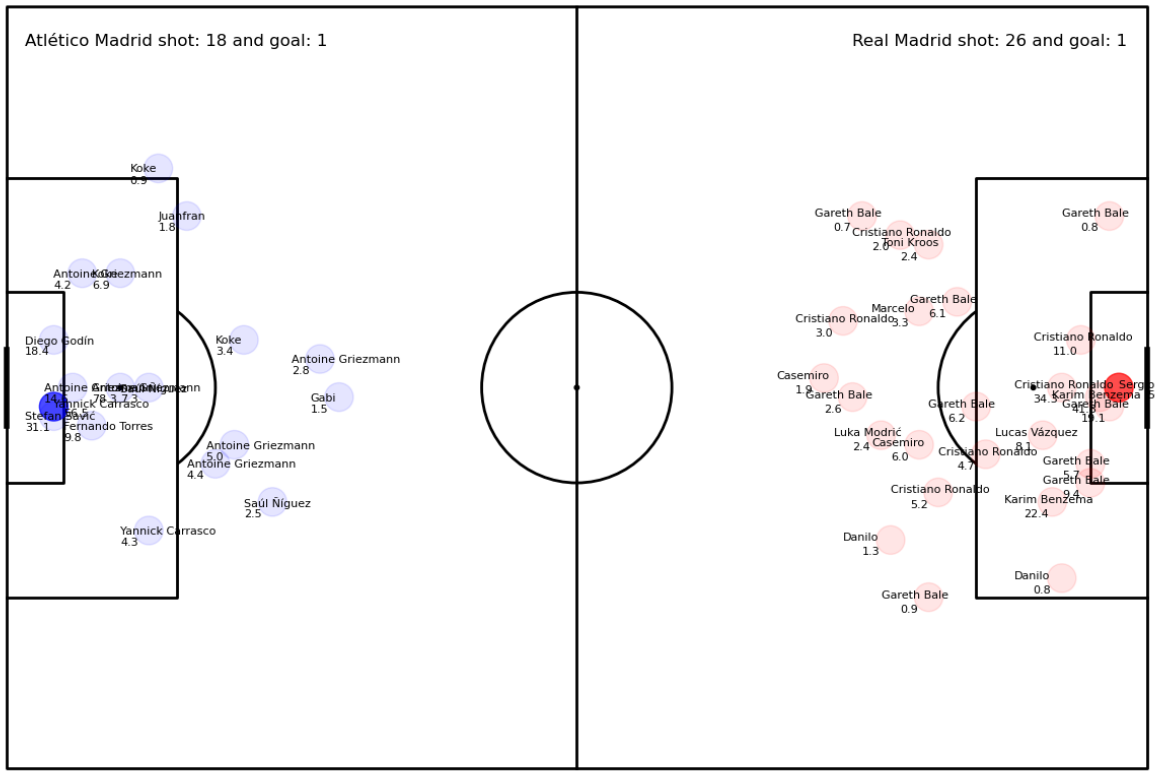


Figure 2 Shot Map of Match 18243 (number specifies Xg*100)

Module 5 Data Analysis – Passing

Obviously, passing is the major event in the football match, in my opinion, passing is the most beautiful portion of the match since it truly demonstrate the intelligence and quality of the team, while other events such as dribble and carry displayed more on the individual excellence.

There are several categories could be analyzed on the pass by the data such as the volume and distribution, accuracy and outcome, network and connectivity, spatial and progression, the comparison of two teams and the profile on the players could be generated as well.

The examples are provided below to provide some starting points for the readers

1. **Passing network and passing comprehensive analysis for match**

Passing networks analysis probably is most noted analysis[56], most of the analysis describes the network shape based on the nodes(average position of the players) and edges(the connection, the frequency and the direction of between the players), it provides some information such as team shape and structure, which explains team bias and the player connection, which explains the strong hub and same time, if over reliance on one player to cause the targeted weakness.

An example is shown in Figure 3, this is for match 18241, the 2014 UEFA final between Real Madrid and Atletico Madrid. The passing networks were generated before 1st substitution, after 1st substitution and after final substitution. The analysis indicated that before the first institution, Real Madrid formed a relatively dense structure and overall the team didn't move too front, there were several passing hubs in Real Madrid, from left side, Ramos was the major hub, and from the right side, both Modric and Carvajal acted as the hub. The team seemed balanced with no obvious overload on any direction. After the first substitution, the team structure became much open also moved forward, and there was obvious bias on left, Modric and Isco all moved left to support the network, again the hubs players were Ramos, Modric and Carvajal, but would that be a problem to leave that much space on the right? After the final substitution, Real Madrid's shape packed again and also moved forward one more time, and they were very heavily connected at the front, it indicted they finally over powered the opponent team and didn't need to pass backward that much. The team also balanced back with no obvious bias on any direction, Isco moved to right side to support the network compared to previous formation. The major hub players were Modric, Marcelo, De Maria and Isco. Although the figure provides a clear description on the formation shape and interconnection between the team members. The analysis doesn't really explain the location and direction of the pass very well.

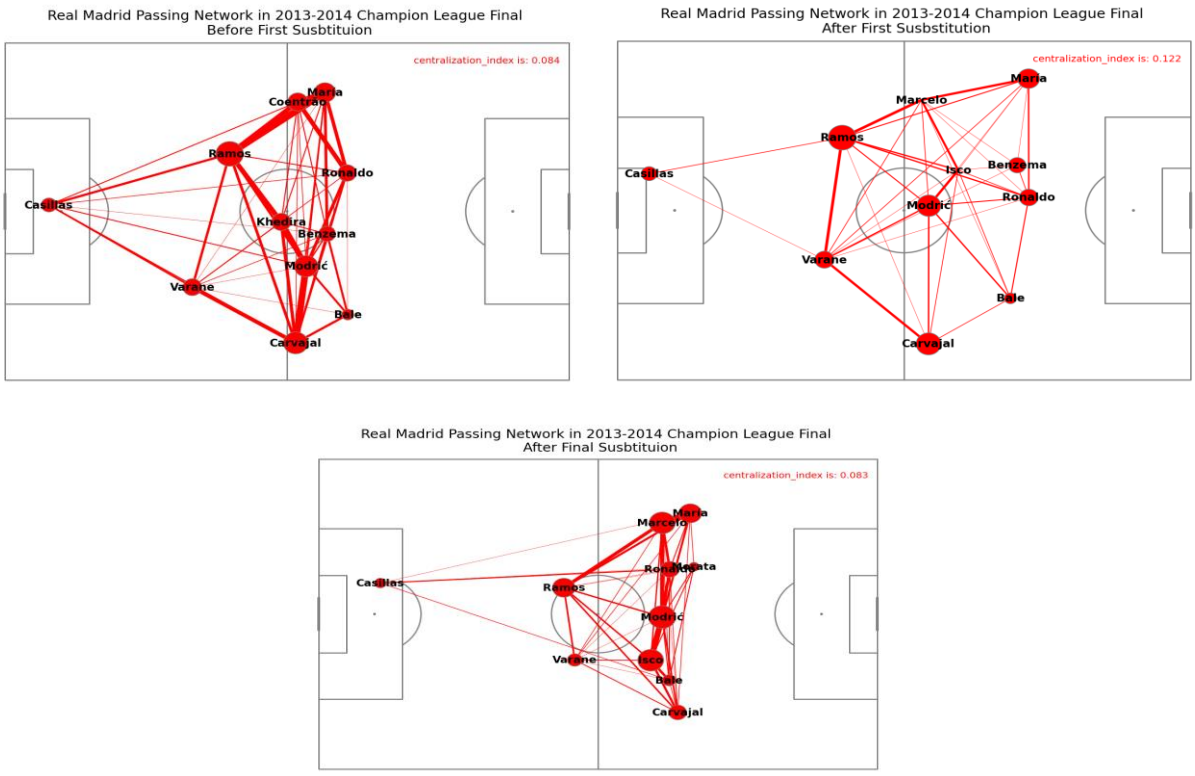


Figure 3 Match 18241 Passing Network of Real Madrid at different time zones

More detailed analysis could be conducted to include the true information of the location and direction of the pass. Figure 4 show the hybrid of passing network and passing in a different way for both teams in same match.

Similar to the previous analysis, Real Madrid's dense team formation was maintained in the first half but they moved forward obviously after 20 mins in the play, the structure became more open in the second half, the majority of the team moved almost in the opposite half after 58 mins until the end. Real Madrid maintained very close passing links within the team members, there were multiple pairs of pass links, it could be concluded they worked as a well-planned group to move the ball around.

On the contrary, the passing networks of Atlético Madrid seemed less organized, except for the first half, they turned to much occupy certain area and leave a lot of open space especially at two flanks for Real Madrid to explore. Real Madrid maintained their general organized shape and evenly distributed in the field in a much better way. Another obvious difference was the passing links, Atlético Madrid showed much less close link between their players, their passing among the teams seemed more random, not very well planned.

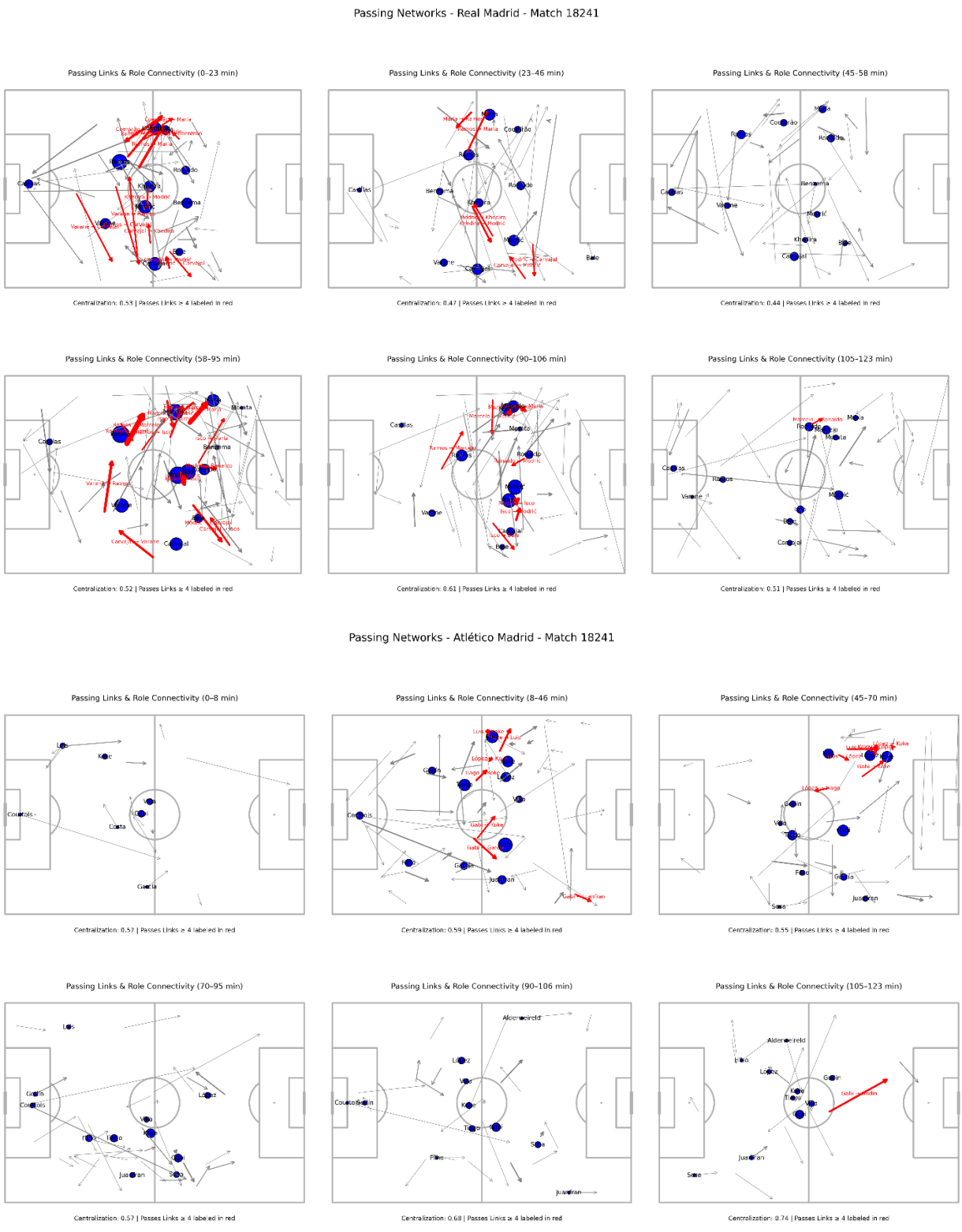


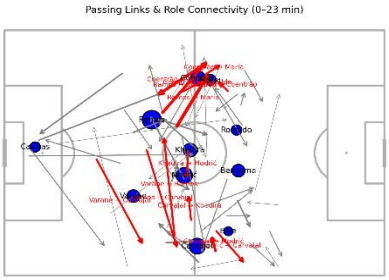
Figure 4 Passing Network and Passing at different time zones

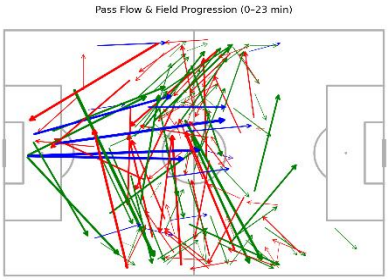
Further evaluation on the passing of these two teams could be demonstrated below **Error! Reference source not found.** and Figure 6. In addition to passing network, actual passing direction, heatmap and failed pass points were all included to provide the comprehensive information. From the set of the charts, Real Madrid definitely demonstrated better passing quality, the medium and long range diagonal and horizontal passes seemed very powerful to move the ball around to open the flank spaces for them. They expanded as much as they could and used long and mid range pass to control the field. The obvious bias on the left started from second half until the first extra time. On the contrary, Atlético Madrid tended to only occupy certain areas, they used more vertical short pass rather than moving the ball horizontally, there was no obvious bias, maybe a little on the right side.

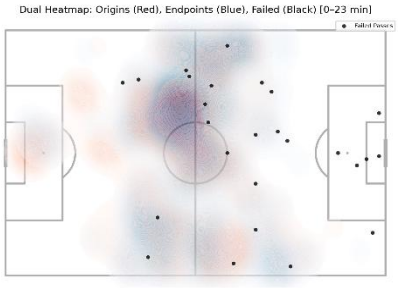
From the failed pass and heat map, it could be concluded that Atlético Madrid defended well when Real Madrid heavily loaded on the left in the second half. From a very basic tactic view, should Atlético Madrid use long range diagonal pass to move the ball quickly to the right side to take the advantage of Real Madrid's left overload? But they didn't, they just fought head to head with Real Madrid on the same side. There were some moments in the second half 45-70 minutes, they worked on the right weak side of Real Madrid, but just playing short pass might provide some buffer for Real Madrid to defend, maybe they lost the confidence on the tactical experiment, then they moved back to fight with Real Madrid's left again.

Finally going down to the extra period, it seemed Atlético Madrid collapsed without forming a reasonable network, while Real Madrid's continuingly explored the left side to generate the fruit for the final victory. And same time, their players moved a little back to right positionally to defend better.

Tactical Visuals - Real Madrid - Match 18241 - Interval 0–23 min



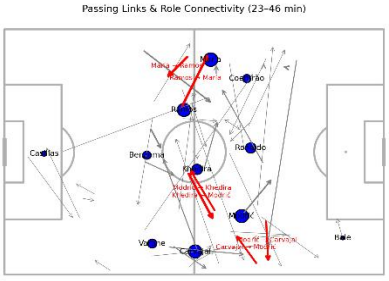


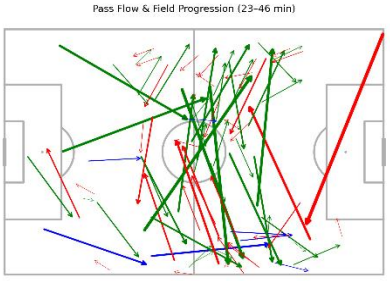


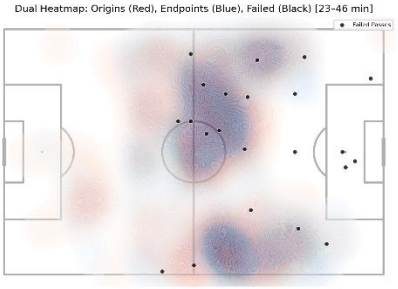
Centralization: 0.53 | Passes Links ≥ 4 labeled in red

Total Passes: 149 | Avg Length: 21.1m | Directional Ratio: -0.21

Tactical Visuals - Real Madrid - Match 18241 - Interval 23–46 min



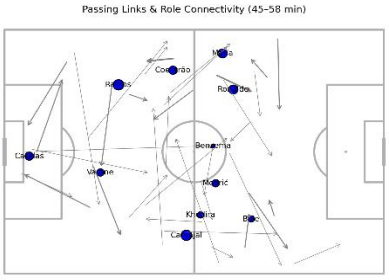


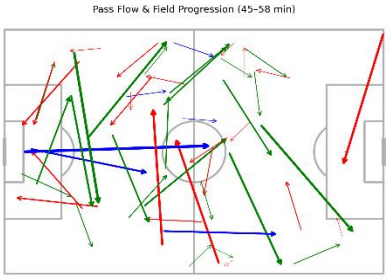


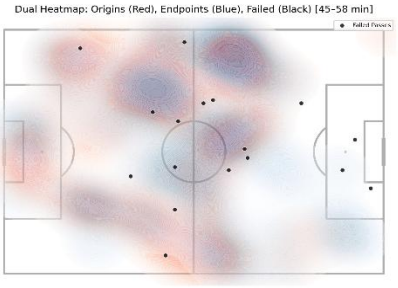
Centralization: 0.47 | Passes Links ≥ 4 labeled in red

Total Passes: 88 | Avg Length: 21.5m | Directional Ratio: -0.07

Tactical Visuals - Real Madrid - Match 18241 - Interval 45–58 min



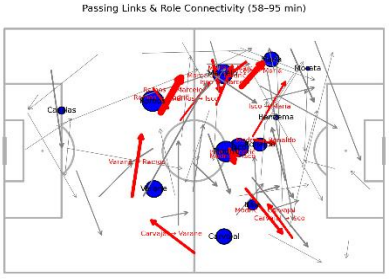


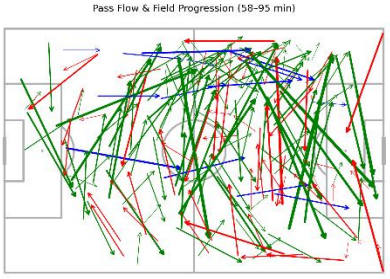


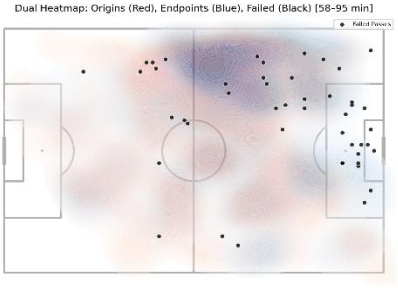
Centralization: 0.44 | Passes Links ≥ 4 labeled in red

Total Passes: 53 | Avg Length: 23.4m | Directional Ratio: 0.41

Tactical Visuals - Real Madrid - Match 18241 - Interval 58–95 min







Centralization: 0.52 | Passes Links ≥ 4 labeled in red

Total Passes: 201 | Avg Length: 21.6m | Directional Ratio: -0.32

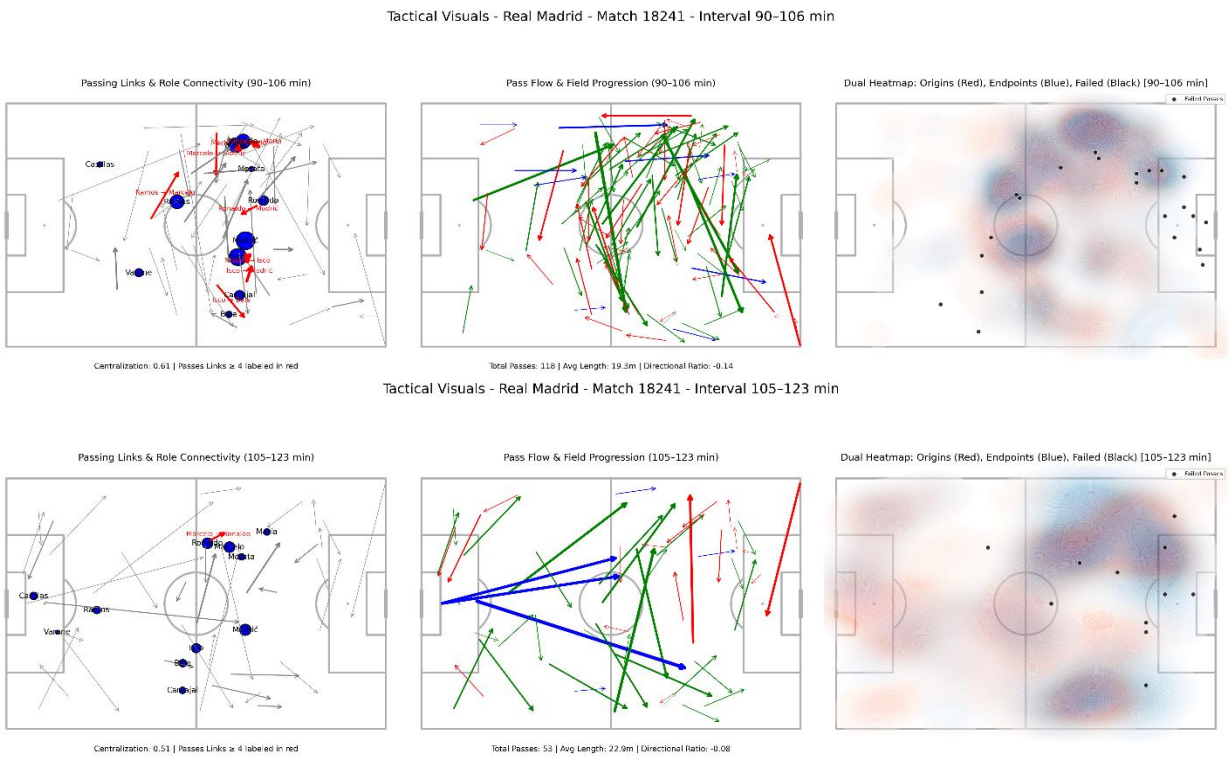
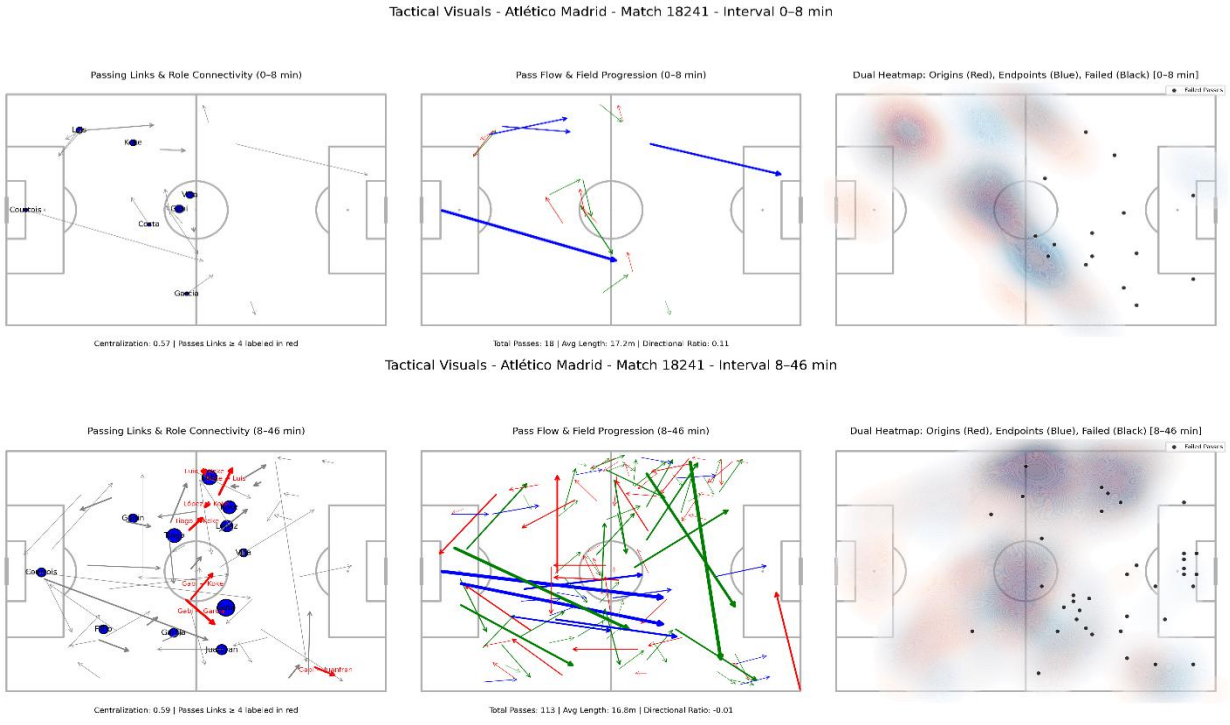


Figure 5 Comprehensive Tactical Visual for Real Madrid in Match 18241



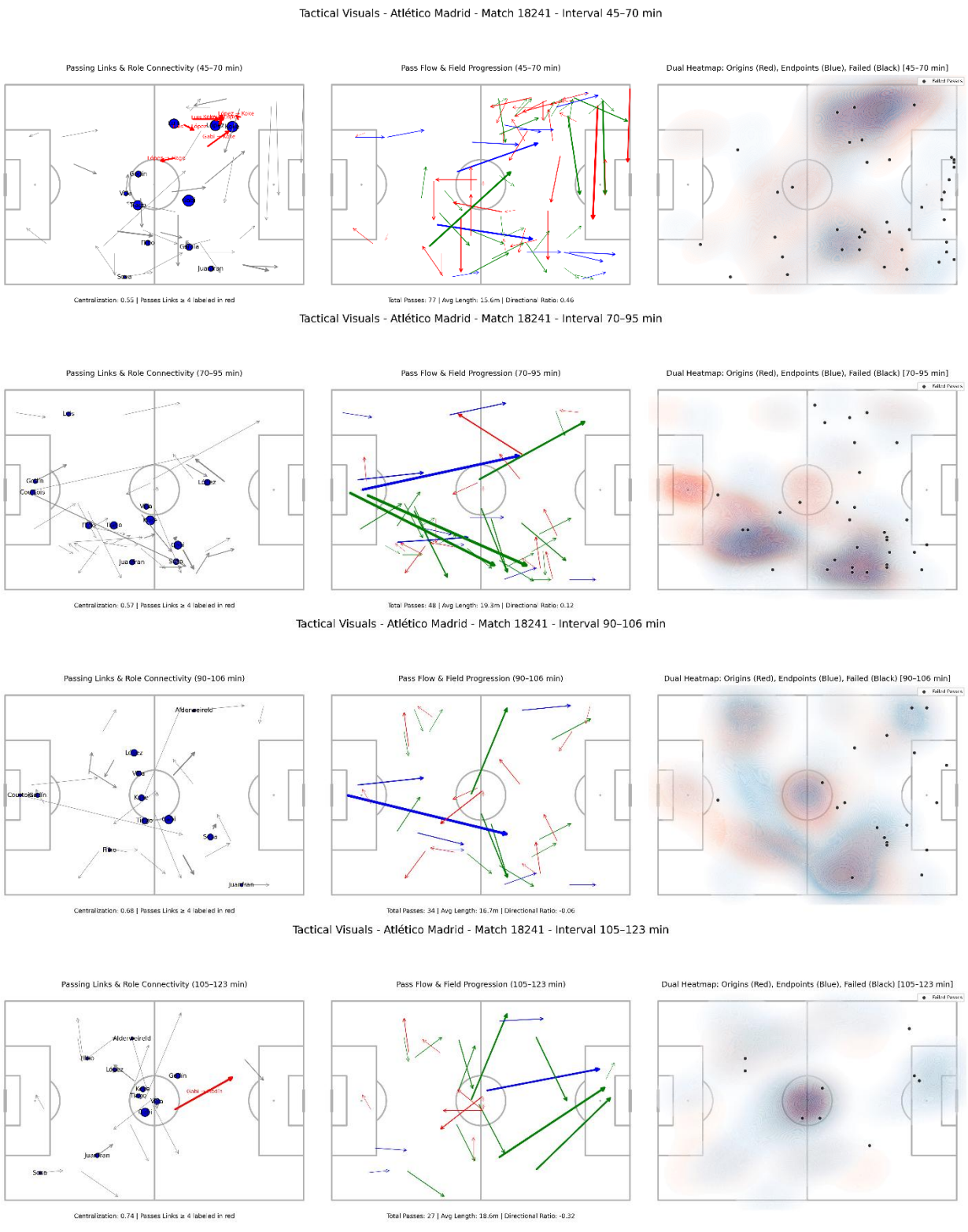


Figure 6 Comprehensive Tactical Visual for Atlético Madrid in Match 18241

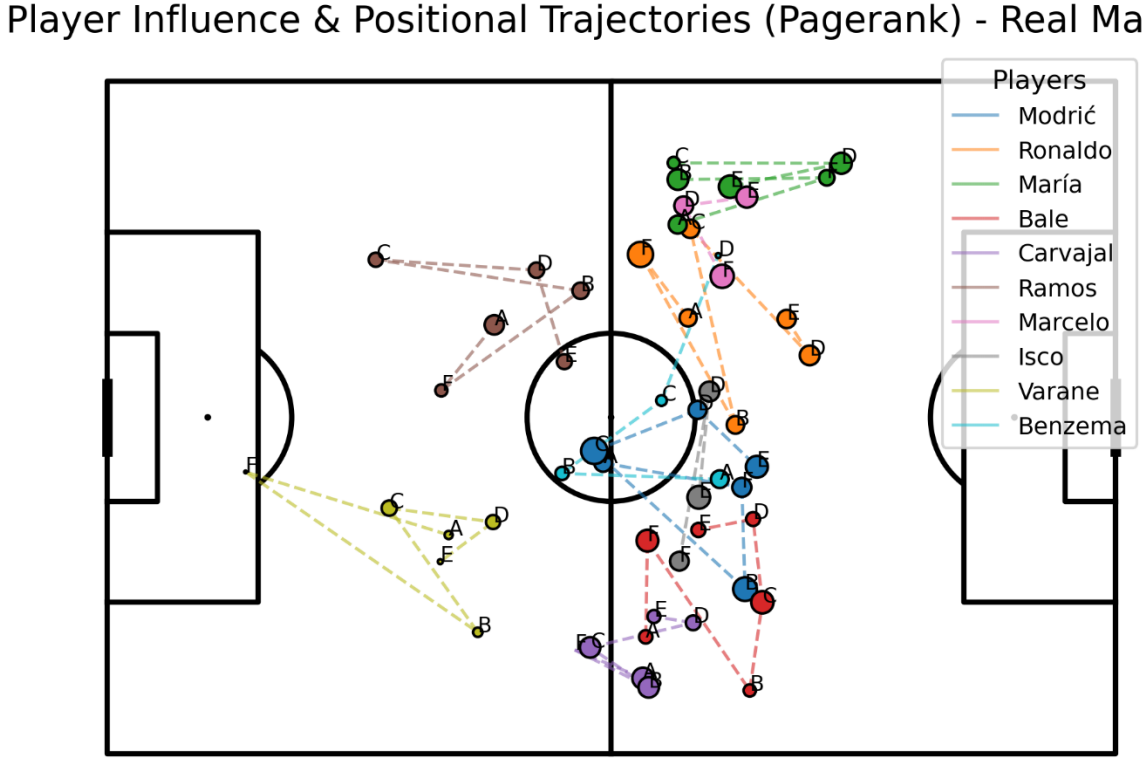
The interesting thing is to compare this match to match 18243, the 2016 UEFA final between Real Madrid and Atletico Madrid. Due to the limitation of the paper, no figures are shown. It was obvious that Atletico Madrid both passing network and passing quality improved dramatically. They maintained their formation structure very well, occupied more space and forward, had more passing links compared to Real Madrid in the first half. In the second half, Real Madrid utilized their tool to over load left again, but rather than head to head fight, Atletico Madrid responded with more diagonal mid range pass to explore Real Madrid's right side.

After they realized this problem, during the extra time, Real Madrid overloaded the right side, this round, Atletico Madrid decided to go head to head fight, and they controlled the field pretty well with all horizontal and diagonal mid/long range pass. Actually for quite some time during this match, compared to two years ago, Atletico Madrid over powered Real Madrid based on the quality of passing network and passing.

There was no surprise that the game went down to penalty shoot out.

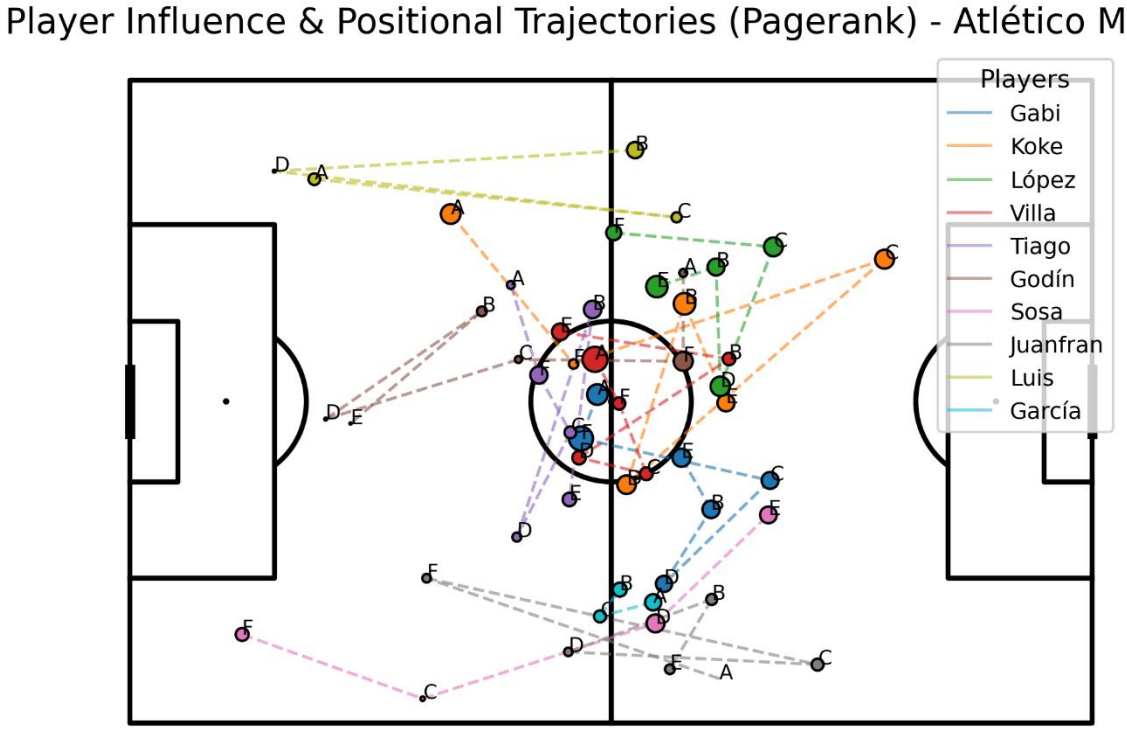
The quantitative metrics analysis could also be conducted on the passing network as well[7],[8]. For an example, the tactical metrics including the following: 'total_passes', 'network_length', 'network_width', 'positional_length_std', 'positional_width_std', 'central_point_x', 'central_point_y', 'decentralization_index', 'forward_movement', 'lateral_movement', 'directional_ratio', 'avg_pass_length', 'pass_length_std' and the results can be compared between two teams how they use the pitch, move the ball and balance the risk and control. Moreover the graph metrics on the passing network including clustering_coefficient, 'shortest_path_length', 'largest_eigenvalue', 'algebraic_connectivity', 'centrality_dispersion', 'max_centrality' can be compared between two teams how they connect the structure, and how resilient, efficient, and balanced the passing structures are.

In addition to the team performance comparison, the metrics can be set up to evaluate the player's performance, a concept of pagerank was used to understand the construction of the hubs in passing networks. In match 18241, both team's top 10 players based on pagerank and their average positions in various periods were shown in Figure 7 and Figure 8 as an example. Based on the charts, in match 18241, from the Real Madrid, both Ronaldo and De Maria built the major passing work at the left side while Bale and Carvajal built the major passing work at the right side. Modrić was the top hub in the center, he worked a little more on right side. On the contrary, 4 of top 5 players in Atletico Madrid heavily occupied in the middle, only top hub Gabi stayed more on the right. It clearly demonstrated that Real Madrid expanded more and controlled more spaces, especially on flank to play wingers.



Interval Legend: A: 0–23,B: 23–46,C: 45–58,D: 58–95,E: 90–106,F: 105–123

Figure 7 Pagerank of Real Madrid Players and Their Position in Match 18241



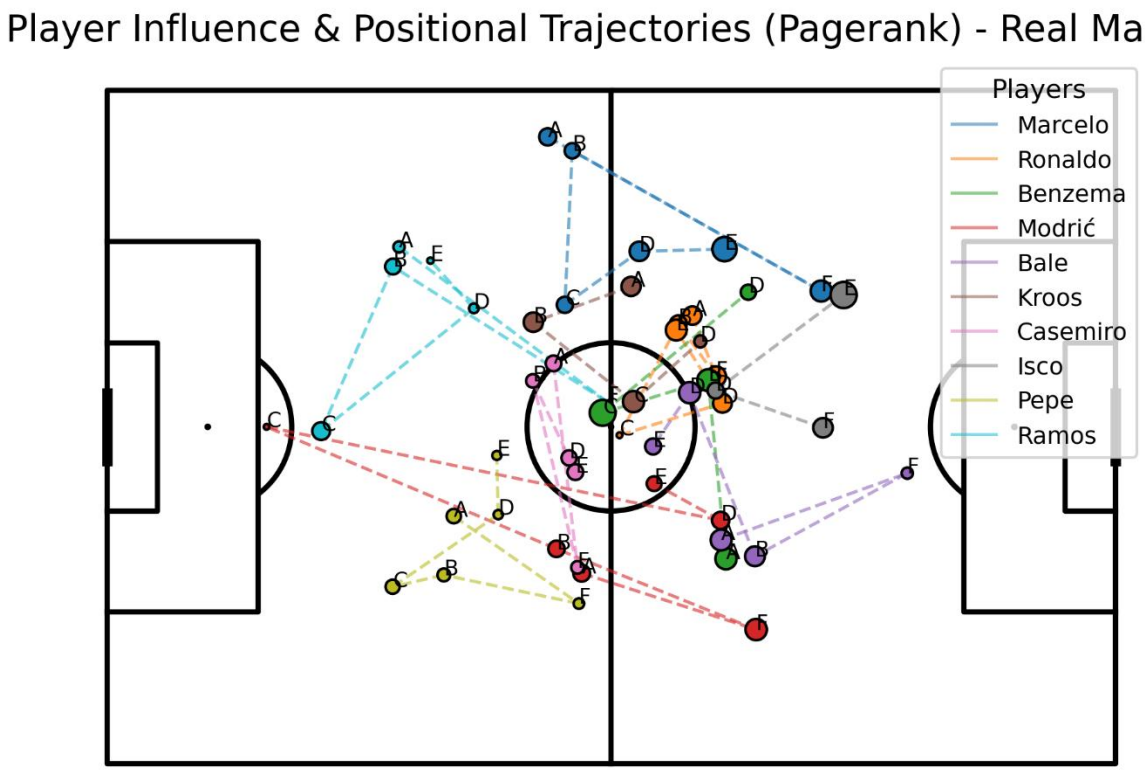
Interval Legend: A: 0–8,B: 8–46,C: 45–70,D: 70–95,E: 90–106,F: 105–123

Figure 8 Pagerank of Atletico Madrid Players and Their Position in Match 18241

As for the match 18243, the 2016 UEFA final between Real Madrid and Atletico Madrid. Both team's top 10 players based on pagerank and their average positions in various periods were shown. It was similar Real Madrid still maintained the strategy to expand to both flanks, Marcelo on the left, Ronaldo and Benzema in the middle, Modrić and Bale on the right. The only concern was Marcelo as top hub and also stayed so high might cause some distraction on the defense side compared to match 18241. As for the Atletico Madrid, they maintained the strategy of strong passing network in the middle with Gabi, Koke and Griezmann, but they built strong network on the left with Luis and Carrasco.

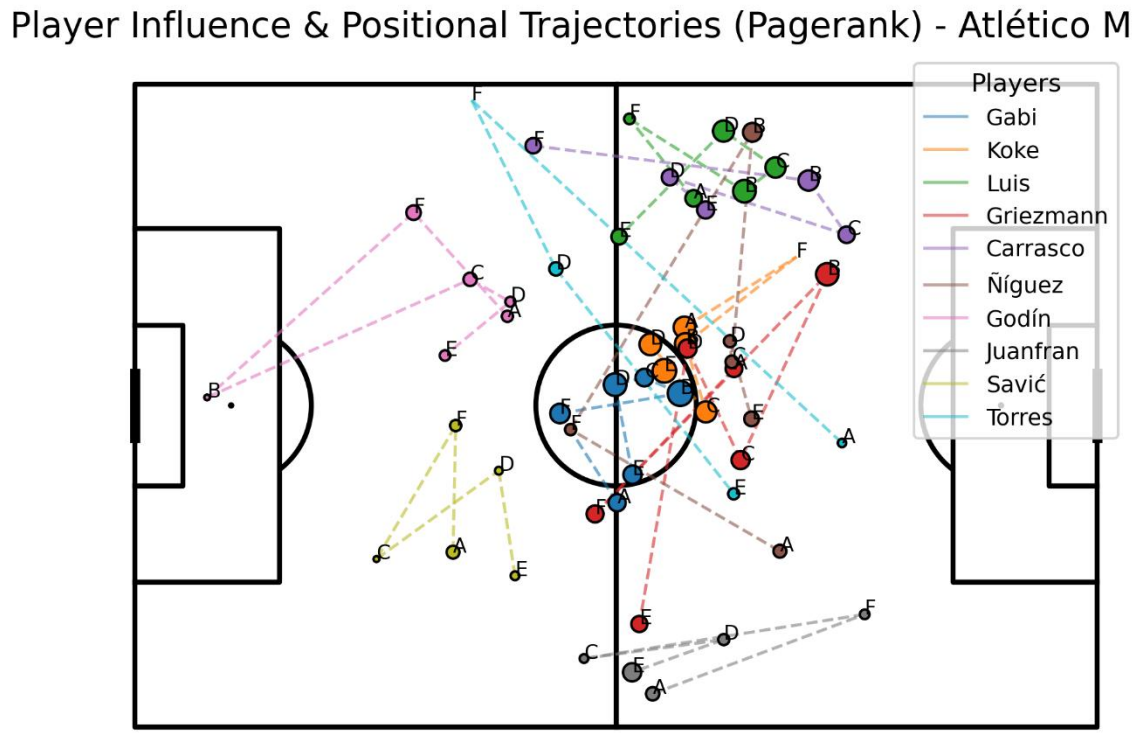
It could be concluded that Real Madrid really enjoyed their strategy to expand the team to utilize the flank and winger, also overloaded left to gain the advantage. In both matches, the forwards team of Real Madrid really served as the connection hubs, while in Atletico Madrid 18243 match, their forward didn't move to top hubs. In this situation, in match 18241, Atletico Madrid generated a relatively poor network, while in match 18243, even their network significantly improved and they seemed used the correct strategy to fight Real Madrid's left overload, but their forward seemed less efficiently connected.

On the other side, the better strategy based on these two matches, rather than head to head fighting on the same side, the opposite team should explore the weak (right) side of Real Madrid.



Interval Legend: A: 0–23, B: 23–46,C: 45–51,D: 51–93,E: 90–105,F: 105–122

Figure 9 Pagerank of Real Madrid Players and Their Position in Match 18243

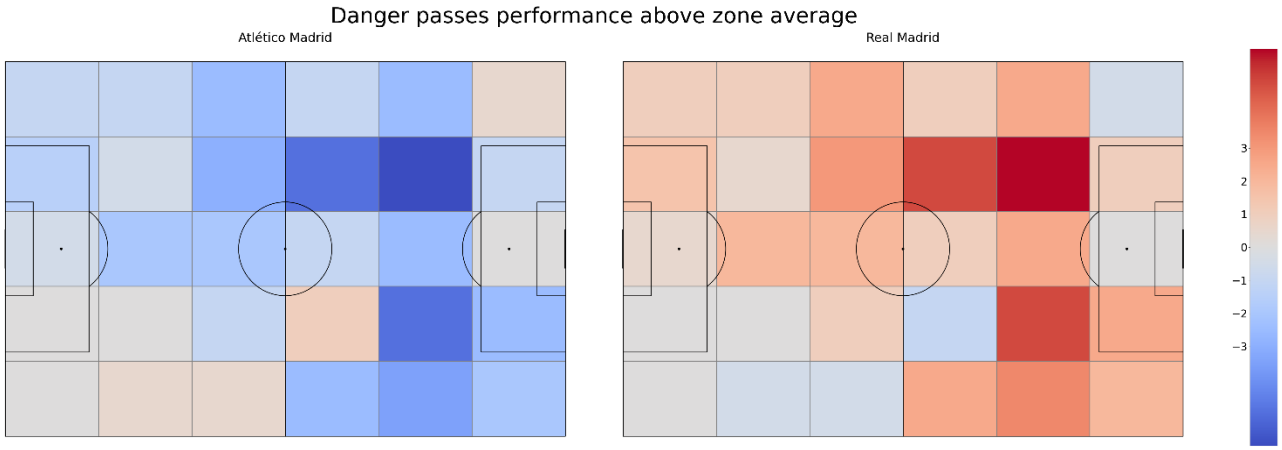


Interval Legend: A: 0–45,B: 45–46,C: 45–69,D: 69–93,E: 90–105,F: 105–122

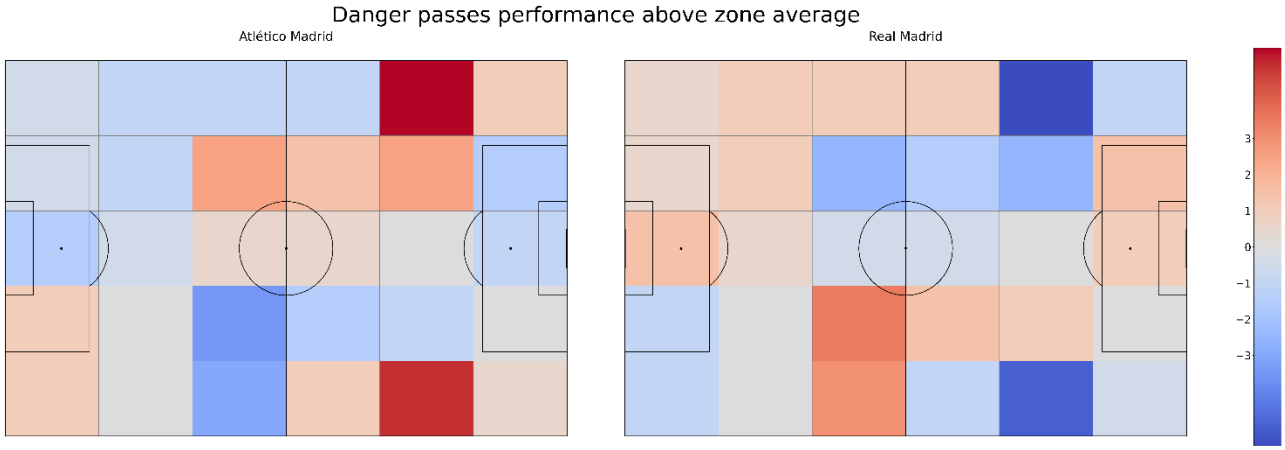
Figure 10 Pagerank of Atletico Madrid Players and Their Position in Match 18241

2. Danger pass comparison and its relationship with shot and goal

Another project was conducted to compare the danger pass for all teams. The purpose of the pass was not just to possess the ball and open the space, it has to convert to the danger pass, the pass that resulted in the shot, hopefully the goal eventually. To understand where the danger pass was created and which team had the advantage, Figure 11 compared both teams in match 18241 and 18243, this demonstrated again Real Madrid overwhelmed Atletico Madrid in match 18241 from so many zones, while in match 18243 Atletico Madrid at least was comparable to Real Madrid, also it showed Real Madrid had some protection problem of both flanks, especially on the right side. Here the danger pass was defined as the pass occurred 120 seconds before a shot.



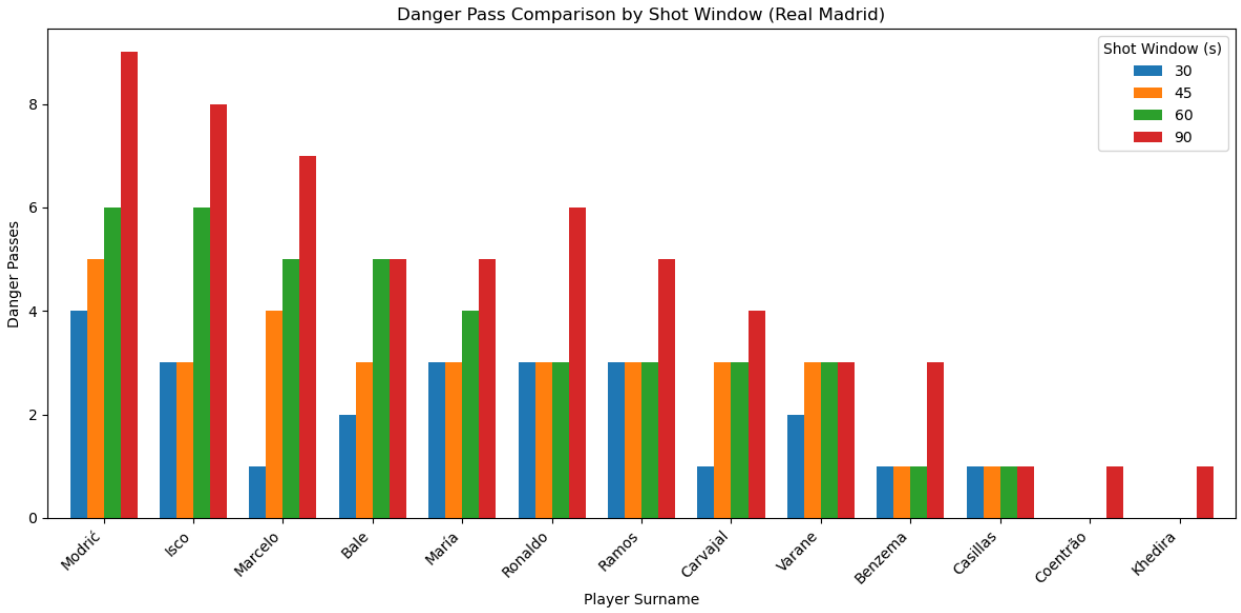
(a) Match 18241



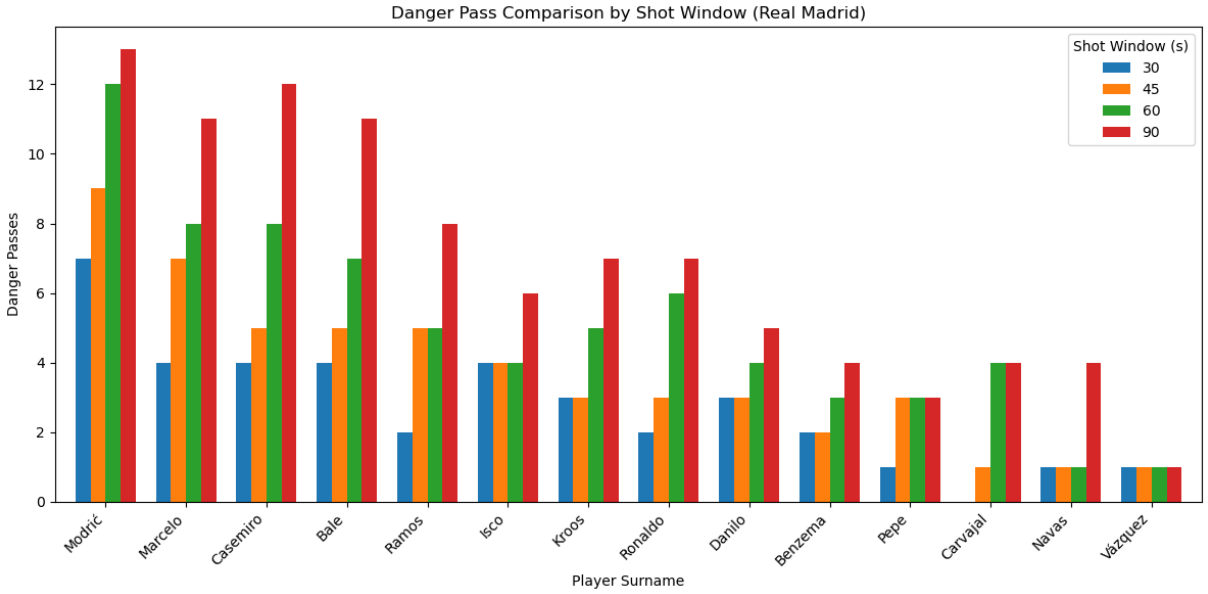
(b) Match 18243

Figure 11 Danger Pass Zones Comparison

The danger pass concept could be used to understand who were the major players that helped to build the attaching powder as demonstrated in Figure 12. For an examples, in match 18241, top 3 players were Modrić, Isco and Marcelo and in match 18243, top 3 players were Modrić, Marcelo and Casemiro.



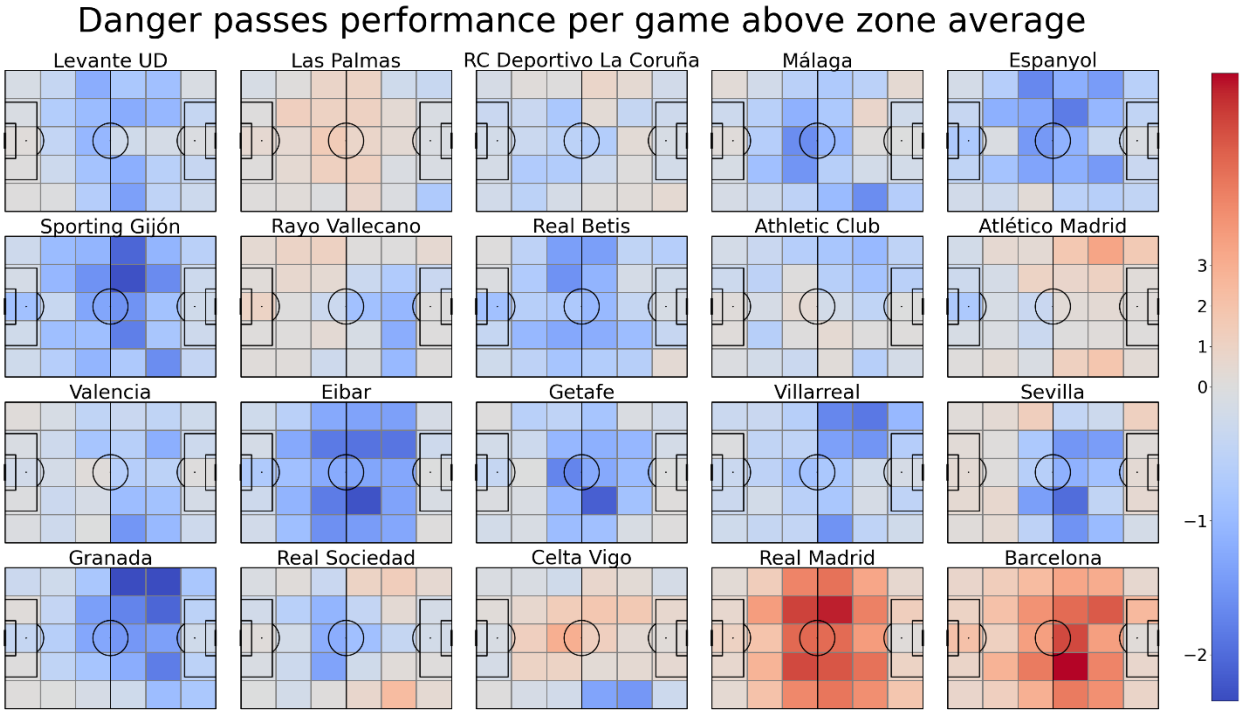
(a) Match 18241 Real Madrid



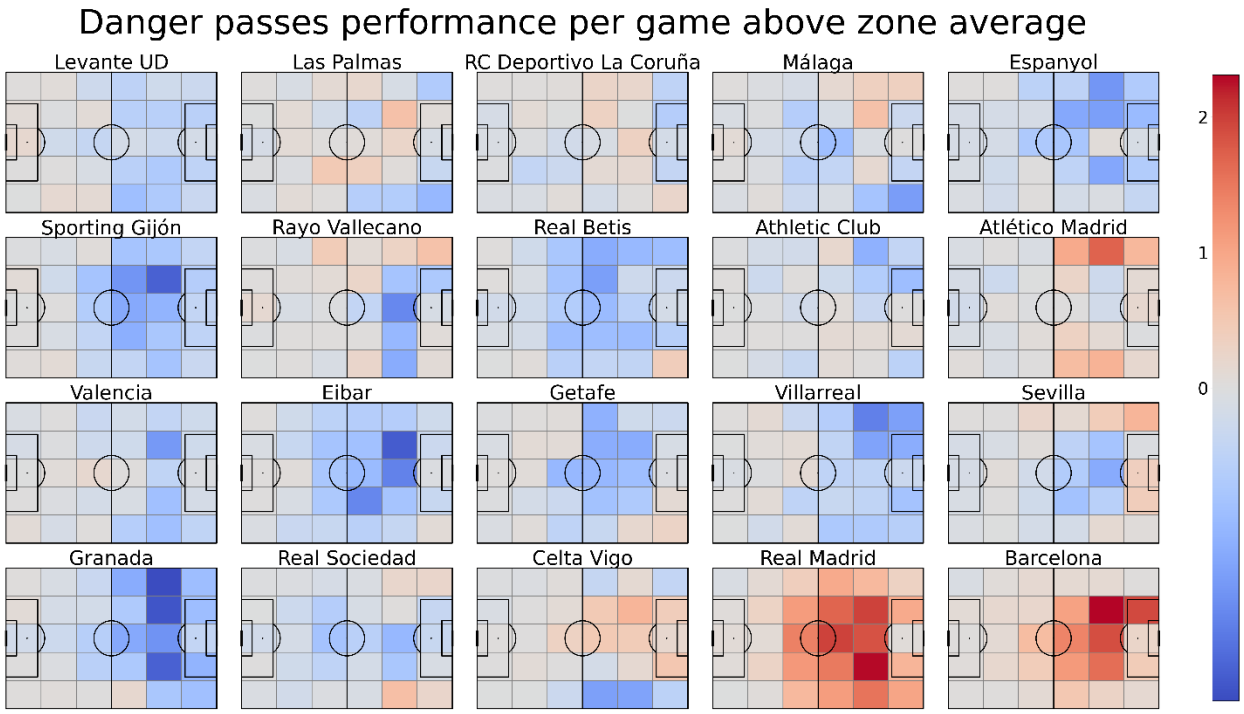
(b) Match 18243 Real Madrid

Figure 12 The players to contribute to danger pass

The danger pass concept could also be used to determine the full season results, La Liga season 2015-16 had full set of 380 matches of information, this season was used as an example. As shown in Figure 13, Real Madrid and Barcelona were two teams obviously standing out on their attacking powders. Here the danger pass was defined as the pass occurred 120 seconds or 15 seconds before a shot. It showed for 120 seconds window, Real Madrid and Barcelona had similar zone performance, while the window dropped to 15 seconds, Real Madrid still had very broad zones in the opposite half while Barcelon focused on the center area of the opposite half.



(a) Danger pass defined as 120 seconds before the shot



(b) Danger pass defined as 15 seconds before the shot

Figure 13 Danger pass performance for La Liga 15/16 season

The danger pass didn't necessarily mean winning the games. It was correlated better with shot, as comparison, the correlation of the pass with the shot was much weaker as shown in Figure 14. This was obvious, since a lot of passes were not effective to deliver the shot. Barcelona generated most of passes among all the team, but the counts of danger pass were a little less than Real Madrid, which indicated their style of possession rather than moving forward to generate shot. This could be further validated by Figure 15, Real Madrid had similar slope as the rest of team while Barcelona had lower slope to indicate less danger pass produced with same pass, but they generated more passes, the danger pass number could eventually be compensated.

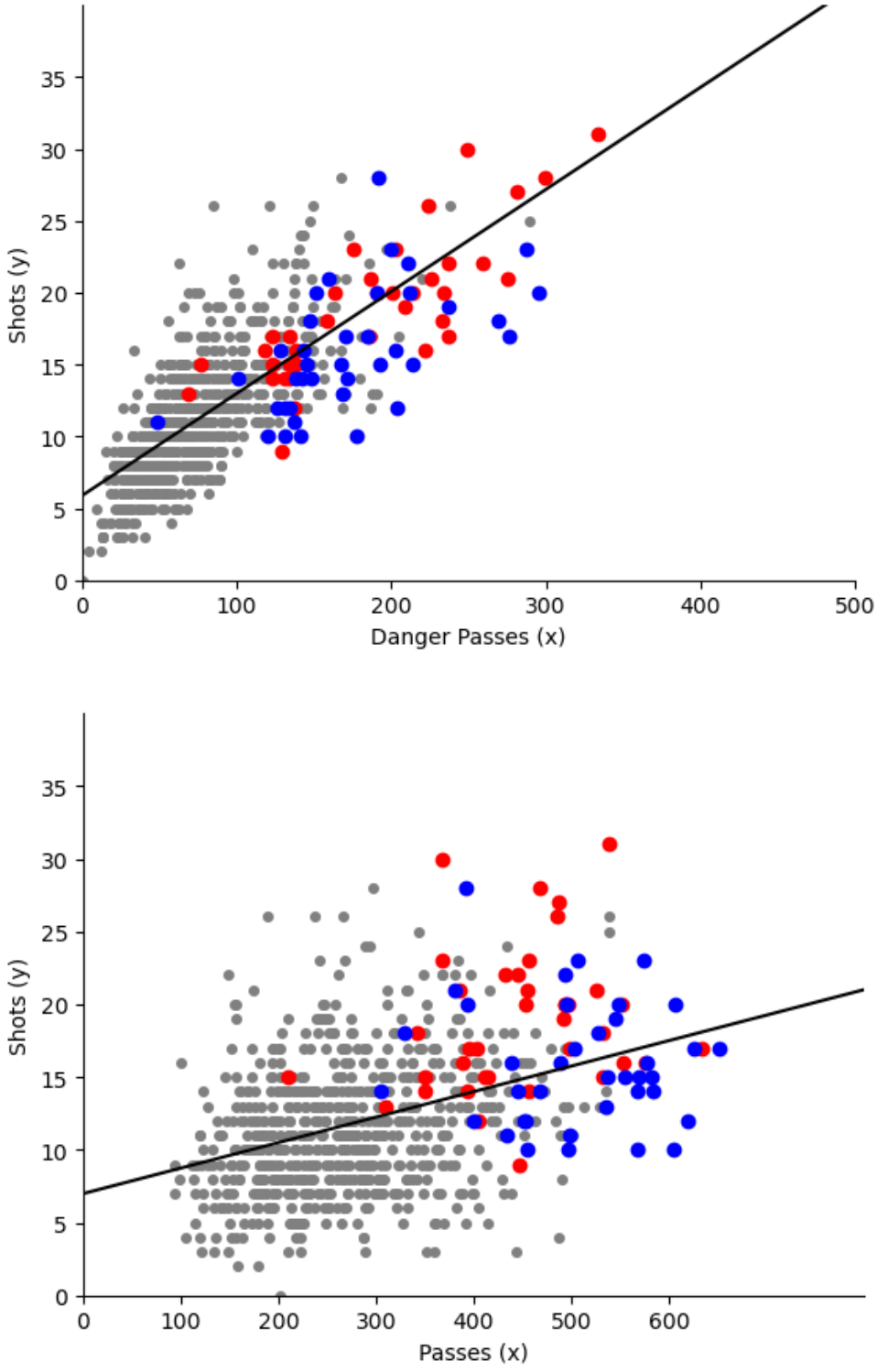


Figure 14 The relationship between shot vs. danger pass or pass Red: Real Madrid; Blue: Barcelona; Grey: All other

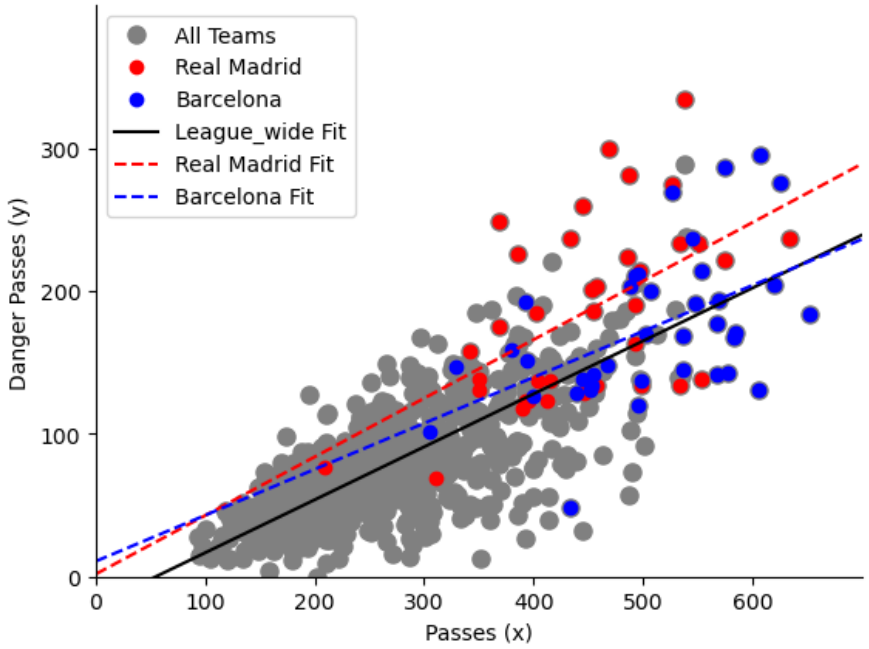


Figure 15 The relationship between the danger pass and pass

Next the relationship between the goal and danger pass was compared in Figure 16 , it was obvious Real Madrid showed several outliners for the goals even with very high danger pass counts and Barcelona seemed more consistent. This might be the reason that cost Real Madrid the championship of season 15-16.

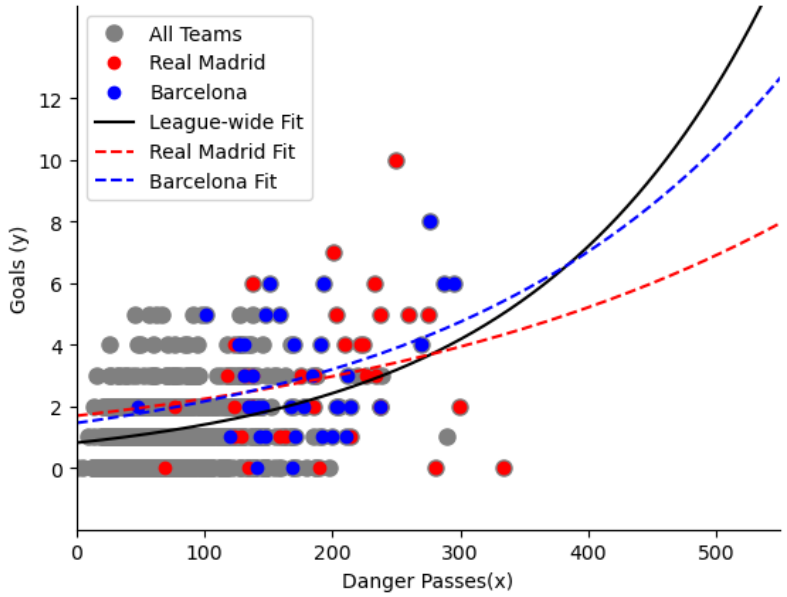
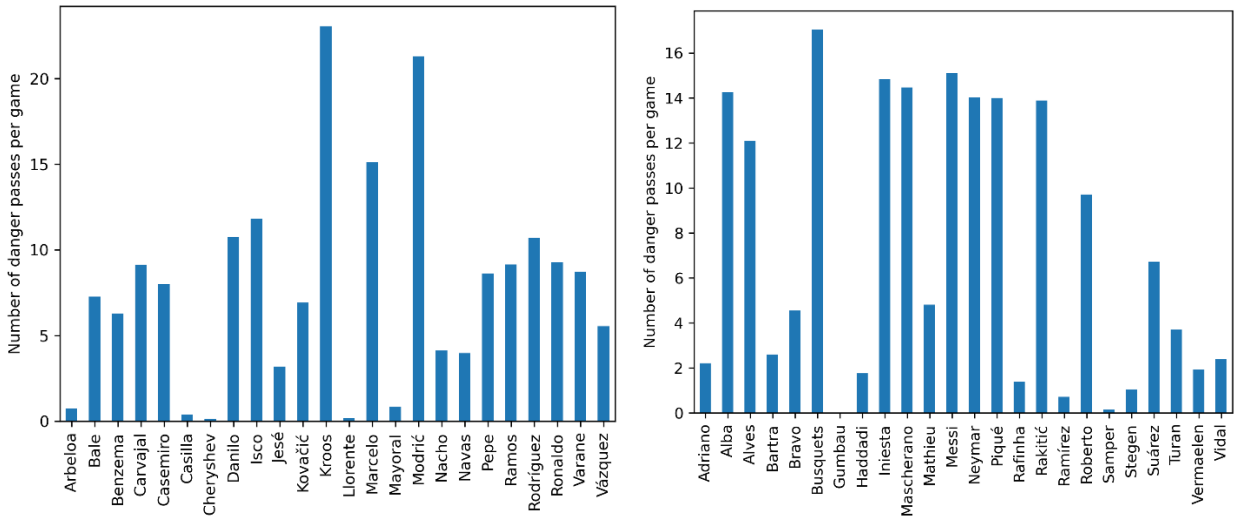
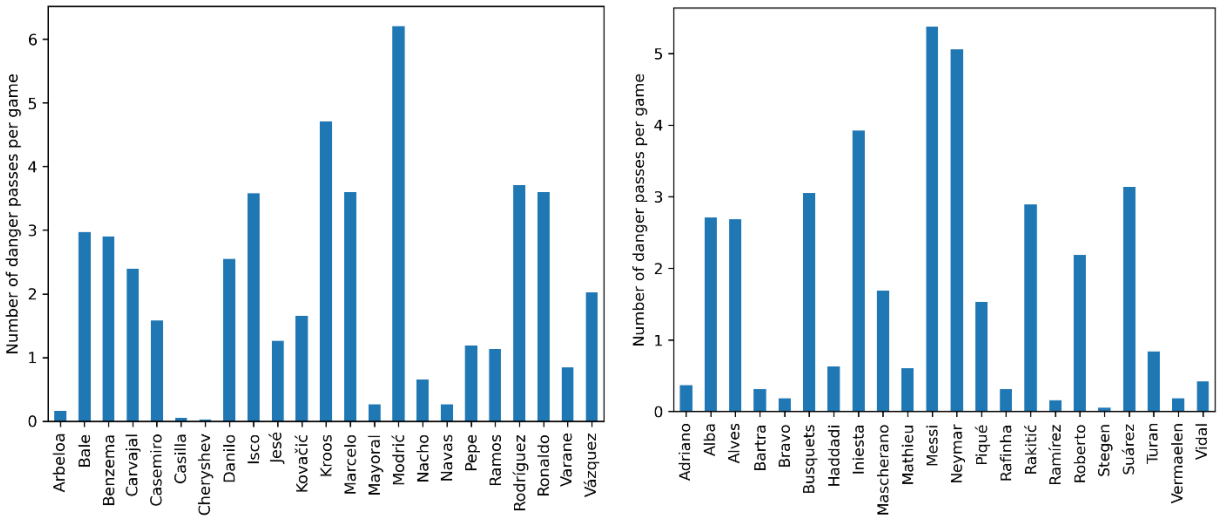


Figure 17 The relationship between danger pass and goal

The player evaluation could be conducted as well. Both Real Madrid and Barcelona's player's involvement in the danger pass were summarized in Figure 18 . How to define the danger pass is an arbitrary selection, as shown in previous study, two numbers were selected, 120 seconds or 15 seconds before a shot to be considered as build up and execution stages. It showed the obvious difference trends on the players' position and style for two teams. More players participated in the danger pass contribution in Barcelona based on 120 seconds window, 8 players contributed over 14 passes per games, while Real Madrid only had 3 players to contribute over 15, although their numbers were higher than Barcelona; This indicated the hierarchy of Real Madrid to be as a model of creator/finisher/distributor while Barcelona acted as a system load with complicated operation. Based on 15 seconds window, if 3 passes was considered as cut off, then both teams had 8 players, but the interesting thing was top contributors from Barcelona were two forwards, while top contributors from Real Madrid were two mid fielders. Not quite sure if that was the reason why Barcelona shot could be more consistent to convert to goal since the forwards helped each other to create the space and adjust for the position through danger pass. On the other side, Real Madrid's forward and Barcelona's mid fielder contributed decently based on 15 seconds window, but since the championship was defined only by 1 point difference, the winning team had to really control quality to the best, which most likely was decided by shot conversion rate. Barcelona type of style might bring the benefit in a league level competition, however it could bring the negative effect as discussed next.



(a) Danger pass defined as 120 seconds before the shot



(b) Danger pass defined as 15 seconds before the shot

Figure 18 The players involvement in danger pass for Real Madrid (left) and Barcelon (right)

As mentioned above, the set of danger passes data could quantify the structural differences between the two teams' possession models. From the build up stages, 120 seconds before a shot, **FC Barcelona:** exhibited a "flat" distribution of involvement, where central defenders (Piqué), fullbacks (Alba), and forwards (Messi) shared nearly equal possession loads (~14–17 danger passes/game). This indicates a high "System Maintenance Cost," where the team relies on a complex, multi-node retention network to progress the ball. While this topology minimizes variance—ideal for domestic consistency as the defensive weapon to win over low block team in the domestic league—it requires the entire XI to function in unison, creating vulnerability against high-intensity European pressing structures. **Real Madrid:** Displayed a hierarchical, midfield-centric distribution peaked around two nodes: Kroos and Modrić (~22 passes/game). Forward players (Ronaldo, Benzema) showed significantly lower involvement in this phase, indicating a system engineered to bypass the first line of pressure efficiently, preserving forward energy for the finishing phase.

Isolating the final 15 seconds of attacking moves revealed the root cause of Barcelona's possible fragility in knockout competitions. **FC Barcelona "Forward" Dependency:** In the final third, Barcelona's midfield contribution effectively reduced. The data shows that creative responsibility was heavily centralized on the forward Messi and Neymar, with Messi bearing the dual burden of deep progression and final-third creation. The statistical disappearance of the midfield (Rakitić/Iniesta) in this window suggests a lack of systemic redundancy; This might help to bring more efficient goal conversion however resulted in a danger that if the forward line was neutralized, the team lacked alternative threat vectors. **Real Madrid's Distributed Threat:** Conversely, Real Madrid's 15-second profile demonstrated high redundancy. Creative output was distributed across on more players who

significantly diluted the defense. Notably, Luka Modrić recorded the highest volume of danger passes (~6.2/game), outpacing the forwards. This "Polycentric Attack" meant that neutralizing the primary goalscorer (Ronaldo) did not neutralize the creative system.

A comparative analysis of rotation players ("B Team") highlights a critical disparity in squad engineering. Real Madrid's rotation options (e.g., James Rodríguez, Isco) registered danger pass volumes (≈ 3.6/game) comparable to, and in some cases exceeding, starters from . This implies a modular system where personnel changes did not degrade tactical output. In contrast, Barcelona's rotation options showed a steep drop-off in creative output compared to the starters, confirming a "Single Point of Failure" architecture

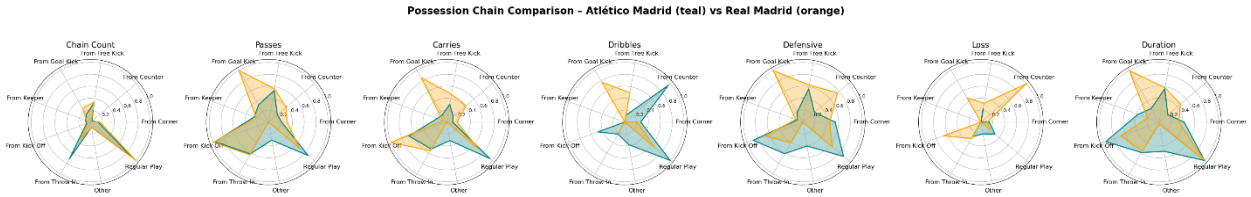
Module 6 Data Analysis – Valuing Action

The shot analysis is focused on the end product efficiency since the shot is directly connected to the goal, that decides the winning. Passing analysis discloses the team connectivity, structure, possession dynamics and style. Now let's understand how all other actions were conducted and influenced the play.

(1) Possession Chain and Shot Chain analysis

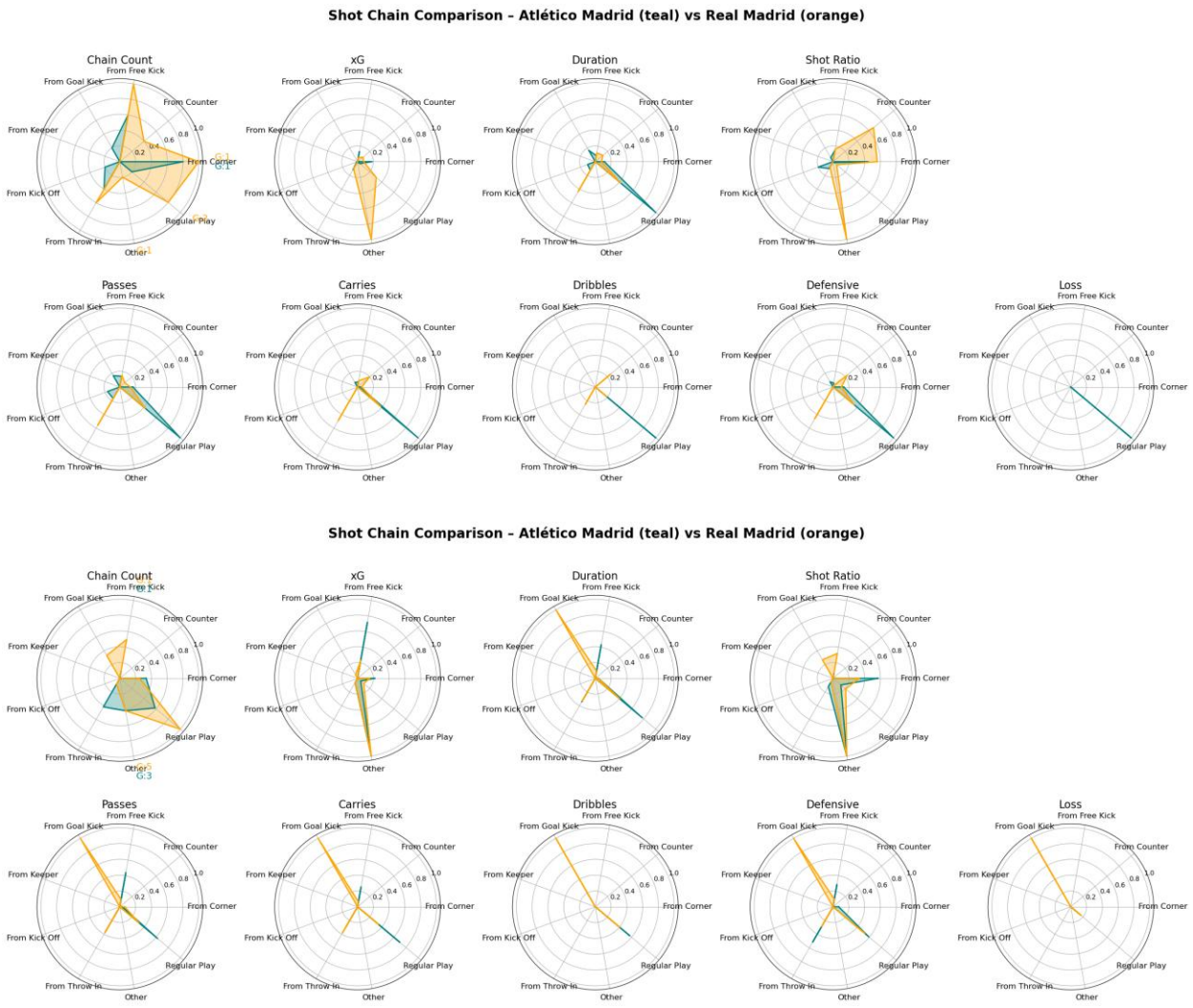
Both Match 18241 and 18243 were studied again, both possession chain and shot chains were compared between two teams in both matches. The analysis of the possession chain is shown in Figure 19. It was once showed Real Marid dominated Atlético Madrid for most of play patterns in match 18241, 2 years later, Atlético Madrid reversed the dominance significantly in carry, dribble, defense and reducing the loss in match 18243.

(a) Match 18241



(b) Match 18243

Figure 19 Possession Chain Analysis between Real Madrid and Atletico Madrid



(a) Match 18243

Figure 20 Shot Chain Analysis between Real Madrid and Atletico Madrid

However based on the shot chain analysis, it indicated the change of the dominance in possession didn't translate well in the shot chain. Atlético Madrid seemed less efficient to convert their possession chain to shot chain, while Real Madrid was better. With limited data, it demonstrated Real Madrid defended similar as Atlético Madrid in the shot chain during the regular play in Match 18243, while Real Atlético Madrid defended better in the shot chain during the regular play in Match 18241. So overall impression was even Atlético Madrid possessed better and played with good strategy, but they were less efficient to shot and Real Madrid defended similar as them, all those factors resulted in an equivalent game.

(2) Time series analysis

Time analysis of the actions in the both matches were conducted. The goals for the studies were to evaluate the team dynamics, capture the swing of the moment, link the possible cause and effect. The analysis are shown in Figure 21 and Figure 22.

In match 18241, Real Madrid really controlled the game after second half, they probably should end the game earlier, but they didn't convert well during 2nd half, the part of the reasoning was discussed in the shot analysis and shot chain analysis. In the extra time, Atlético Madrid really collapsed and their loss was inevitable then.

2 years later, in match 18243, Real Madrid only had limited time to possess well, but clearly even with much less possession, Real Madrid still managed to shot more than Atlético Madrid, also Real Madrid seems always had better physical in last half in the extra time, but none of those converted to a goal for Real Madrid either. In this match the problem for Atlético Madrid they could convert their possession into shot, while Real Madrid could convert their shot to goal. At the end, Real Madrid was a little better to overtake Atlético Madrid on the penalty shootout, the poorer physical in the second half of the extra time from Atlético Madrid might forecast this results.

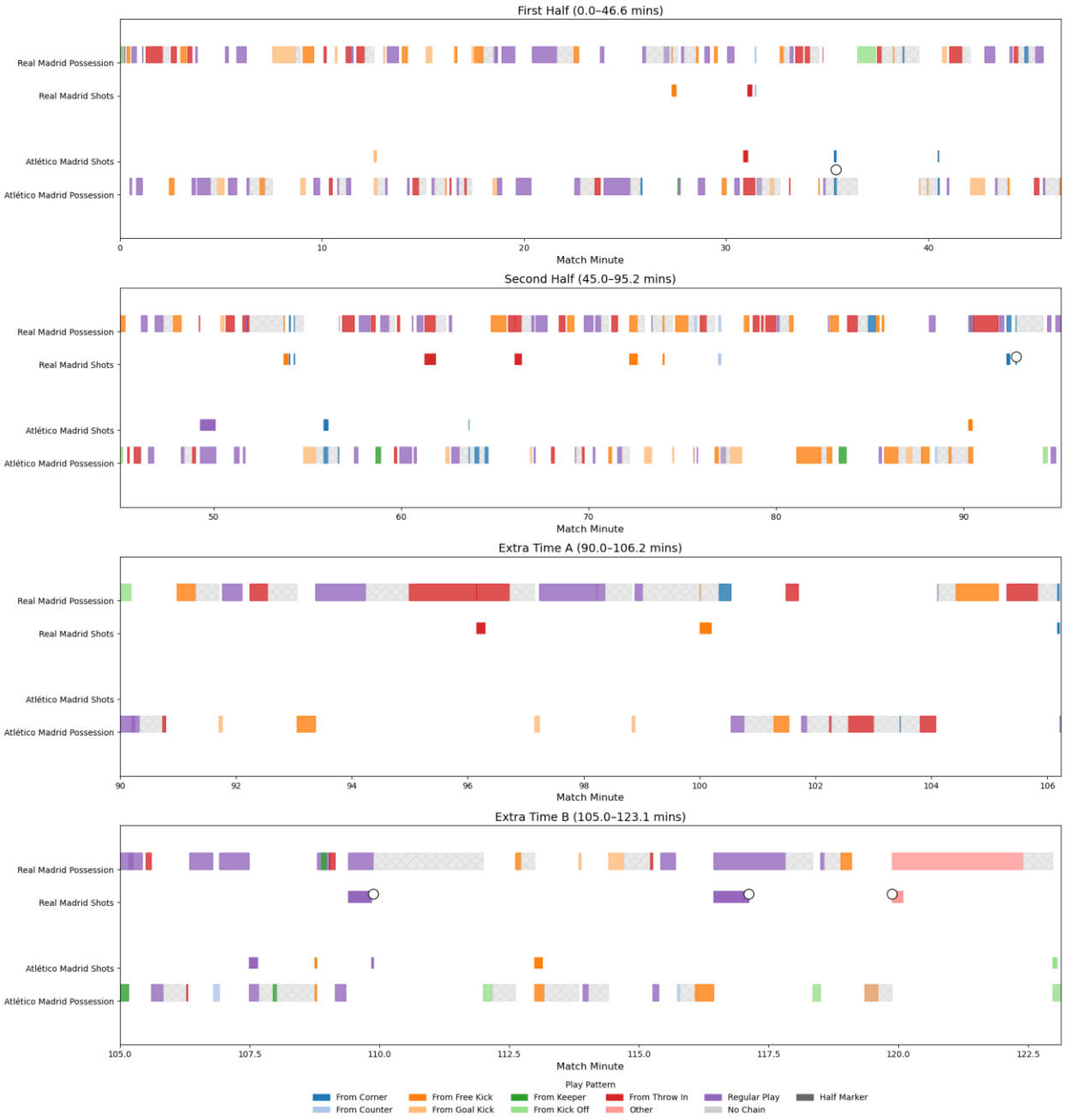


Figure 21 Time series analysis for Match 18241

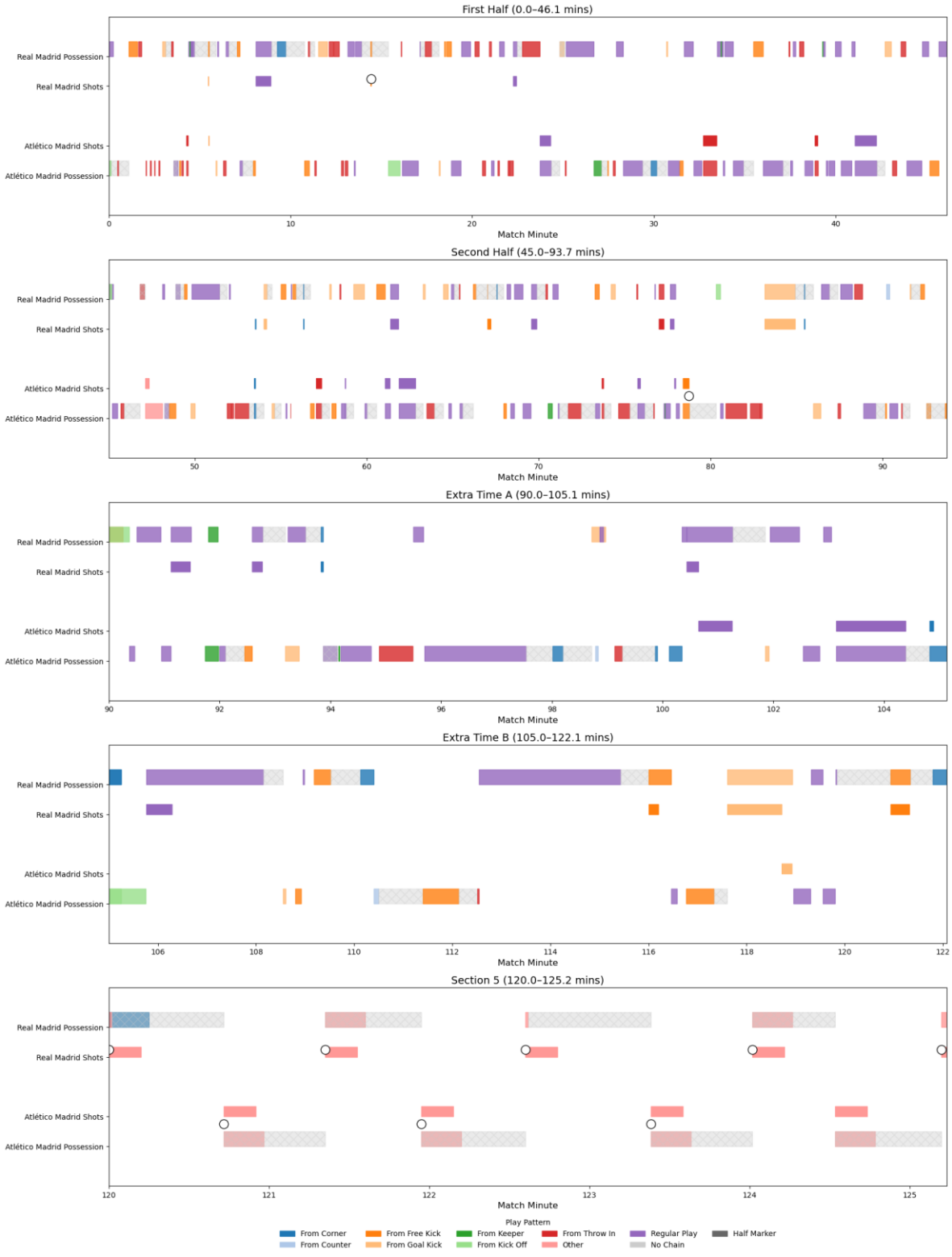


Figure 22 Time series analysis for Match 18241

(3) Shot Chain and player

The shot chain analysis could be assembled with the players. With this analysis, the styles and the actions of the players could be identified. An example of Real Madrid's goal though regular play in Match 18241 was shown in Figure 23. It concluded with a typical style of Real Madrid, started from the flank then penetrated into the central, when lost the possession, pressed and got it back as soon as possible, and moved to the other side of flank, then penetrated to central again. All happened very quickly, almost 40 events occurred in 40 seconds.

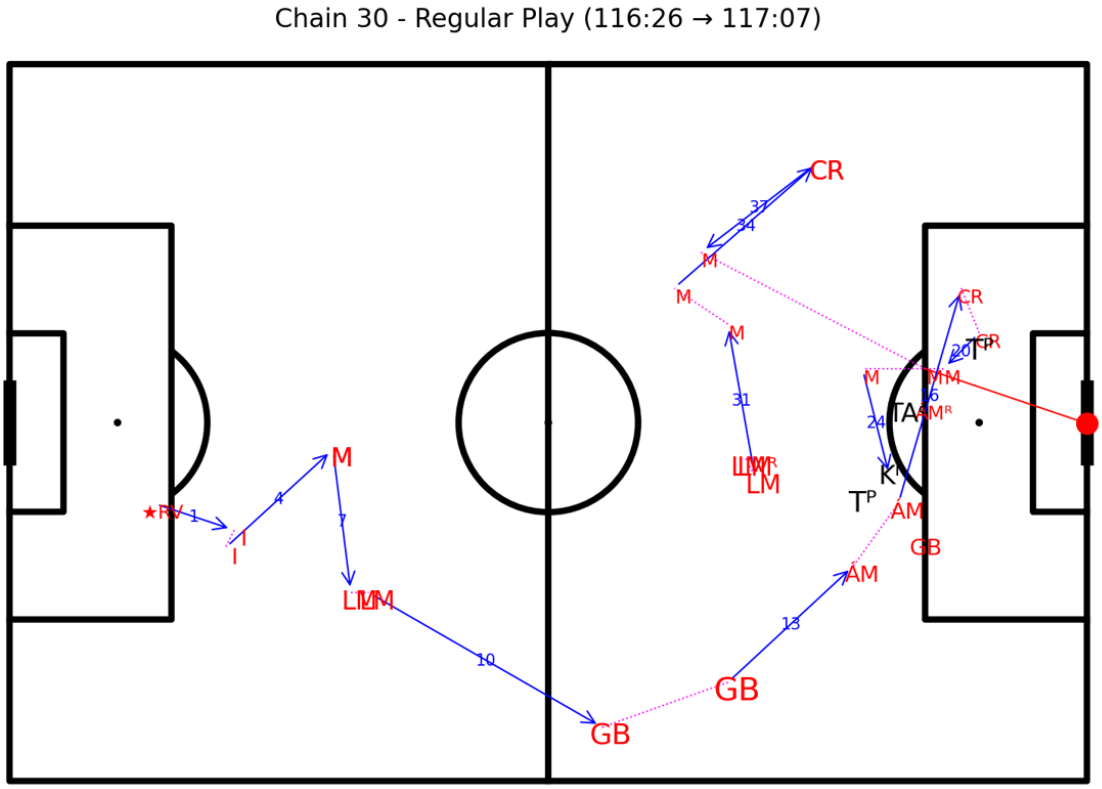


Figure 23 An example of a shot chain of Real Madrid resulted in a goal

As said before, the Statsbomb data was not track data, no full information for all off ball actions, however some of the defensive actions were captured in the possession chain to at least explain some efforts for the defense. The author set some basic rules to quantitatively evaluate the participation of the players in the possession chains, and the extra weight was given to those chains turned to the shot chains or goal chains. The scores on both attacking and defending involvement for the match 18241 were summarized in Figure 24.

Team_name	Player	Position	Possession Events	Shot Events	Goal Events	Possession Only	Shot Only	Score	Substitute	Sub Entry Minute	Replacement	Substitution Type
Real Madrid	Luka Modrić	Right Center Midfield	278	44	7	234	37	34.3				
Real Madrid	Ángel Di María	Left Center Midfield	295	45	0	250	45	34				
Real Madrid	Sergio Ramos	Left Center Back	248	37	2	211	35	29.1				
Real Madrid	Cristiano Ronaldo	Left Wing	196	46	9	150	37	26.9				
Atlético Madrid	Gabi	Right Defensive Midfield	202	37	4	165	33	25.1				
Real Madrid	Daniel Carvajal	Right Back	225	12	1	213	11	24				
Real Madrid	Marcelo	Left Back	162	33	12	129	21	23.1	Yes	58	Fábio Coentrão	Tactical
Atlético Madrid	Koke	Left Midfield	182	35	0	147	35	21.7				
Real Madrid	Gareth Bale	Right Wing	153	34	6	119	28	20.5				
Real Madrid	Isco	Center Defensive Midfiel	157	33	3	124	30	19.9	Yes	58	Sami Khedira	Tactical
Atlético Madrid	Adrián López	Right Center Forward	140	42	0	98	42	18.2	Yes	8	Diego Costa	Injury
Real Madrid	Raphaël Varane	Right Center Back	132	13	1	119	12	14.8				
Atlético Madrid	Tiago	Left Defensive Midfield	102	10	1	92	9	11.5				
Atlético Madrid	Filipe Luis	Left Back	85	20	0	65	20	10.5				
Atlético Madrid	David Villa	Left Center Forward	95	6	1	89	5	10.4				
Real Madrid	Fábio Coentrão	Left Back	102	1	0	101	1	10.3				
Atlético Madrid	Juanfran	Right Back	79	10	2	69	8	9.5				
Real Madrid	Sami Khedira	Center Defensive Midfiel	89	0	0	89	0	8.9				
Real Madrid	Karim Benzema	Center Forward	80	6	0	74	6	8.6				
Atlético Madrid	Raúl García	Right Midfield	71	10	0	61	10	8.1				
Atlético Madrid	Diego Godín	Left Center Back	59	14	2	45	12	7.9				
Real Madrid	Álvaro Morata	Center Forward	45	12	4	33	8	6.9	Yes	78	Karim Benzema	Tactical
Real Madrid	Iker Casillas	Goalkeeper	54	9	0	45	9	6.3				
Atlético Madrid	João Miranda de Souza F	Right Center Back	53	4	0	49	4	5.7				
Atlético Madrid	José Sosa	Right Midfield	47	1	0	46	1	4.8	Yes	65	Raúl García	Tactical
Atlético Madrid	Thibaut Courtois	Goalkeeper	42	2	0	40	2	4.4				
Atlético Madrid	Toby Alderweireld	Left Back	16	1	0	15	1	1.7	Yes	82	Filipe Luis	Injury
Atlético Madrid	Diego Costa	Right Center Forward	12	0	0	12	0	1.2				

(a) Attacking Scores

Team_name	Player	Position	Possession Events	Shot Events	Goal Events	Possession Only	Shot Only	Score	Substitute	Sub Entry Minute	Replacement	Substitution Type
Atlético Madrid	Tiago	Left Defensive Midfield	58	11	2	47	9	7.5				
Atlético Madrid	Gabi	Right Defensive Midfield	66	9	0	57	9	7.5				
Atlético Madrid	David Villa	Left Center Forward	52	8	1	44	7	6.3				
Atlético Madrid	Juanfran	Right Back	40	8	2	32	6	5.4				
Atlético Madrid	Koke	Left Midfield	46	5	1	41	4	5.4				
Real Madrid	Daniel Carvajal	Right Back	42	7	0	35	7	4.9				
Real Madrid	Sergio Ramos	Left Center Back	36	10	0	26	10	4.6				
Real Madrid	Sami Khedira	Center Defensive Midfiel	31	7	2	24	5	4.4				
Atlético Madrid	Raúl García	Right Midfield	35	7	0	28	7	4.2				
Real Madrid	Luka Modrić	Right Center Midfield	33	9	0	24	9	4.2				
Atlético Madrid	José Sosa	Right Midfield	40	2	0	38	2	4.2	Yes	65	Raúl García	Tactical
Atlético Madrid	Diego Godín	Left Center Back	32	6	1	26	5	4.1				
Atlético Madrid	Toby Alderweireld	Left Back	26	8	2	18	6	4	Yes	82	Filipe Luis	Injury
Atlético Madrid	Adrián López	Right Center Forward	36	4	0	32	4	4	Yes	8	Diego Costa	Injury
Atlético Madrid	João Miranda de Souza F	Right Center Back	34	5	0	29	5	3.9				
Real Madrid	Raphaël Varane	Right Center Back	23	5	2	18	3	3.4				
Real Madrid	Isco	Center Defensive Midfiel	23	2	0	21	2	2.5	Yes	58	Sami Khedira	Tactical
Real Madrid	Fábio Coentrão	Left Back	20	4	0	16	4	2.4				
Real Madrid	Marcelo	Left Back	21	1	0	20	1	2.2	Yes	58	Fábio Coentrão	Tactical
Real Madrid	Ángel Di María	Left Center Midfield	22	0	0	22	0	2.2				
Atlético Madrid	Filipe Luis	Left Back	18	1	0	17	1	1.9				
Real Madrid	Cristiano Ronaldo	Left Wing	11	2	0	9	2	1.3				
Real Madrid	Gareth Bale	Right Wing	11	2	0	9	2	1.3				
Real Madrid	Karim Benzema	Center Forward	11	1	0	10	1	1.2				
Real Madrid	Álvaro Morata	Center Forward	12	0	0	12	0	1.2	Yes	78	Karim Benzema	Tactical
Atlético Madrid	Thibaut Courtois	Goalkeeper	9	0	0	9	0	0.9				
Real Madrid	Iker Casillas	Goalkeeper	1	0	0	1	0	0.1				
Atlético Madrid	Diego Costa	Right Center Forward	1	0	0	1	0	0.1				

(b) Defending Scores

Figure 24 Player Scores in Match 18241

The scores were based on the involvements, the weights were given if the involvements resulted in the shot or goal. This was a very primary study, but at least demonstrated the overwhelming strength from Real Madrid on the attacking and the better strength from Atlético

Madrid on the defending. And who were the major contributors for those strengths. Another interesting thing was to evaluate the effect of the substitution, the replacement of two players from Real Madrid at 58 minutes really added more attacking strength without losing too much on defending strength.

The results of match 18243 were demonstrated in Figure 25. It was really obvious that Atlético Madrid significantly improved their attacking strength compared to two years ago, probably equivalent to Real Madrid while Real Madrid defended better than two years ago, also equivalent to Atlético Madrid. Regarding the substitution, it seemed both teams' replacements were meaningful adjustments as well.

Team_name	Player	Position	Possession Events	Shot Events	Goal Events	Possession Only	Shot Only	Score	Substitute	Sub Entry Minute	Replacement	Substitution Type
Atlético Madrid	Gabi	Right Defensive Midfield	309	71	6	238	65	39.8				
Atlético Madrid	Koke	Left Midfield	324	65	0	259	65	38.9				
Atlético Madrid	Filipe Luis	Left Back	255	49	0	206	49	30.4				
Atlético Madrid	Antoine Griezmann	Left Center Forward	219	57	4	162	53	28.8				
Real Madrid	Luka Modrić	Right Center Midfield	211	53	0	158	53	26.4				
Real Madrid	Gareth Bale	Right Wing	183	65	3	118	62	25.7				
Real Madrid	Marcelo	Left Back	198	55	1	143	54	25.6				
Real Madrid	Casemiro	Center Defensive Midfiel	175	59	0	116	59	23.4				
Atlético Madrid	Juanfran	Right Back	158	47	8	111	39	22.9				
Real Madrid	Cristiano Ronaldo	Left Wing	160	49	1	111	48	21.2				
Atlético Madrid	Yannick Carrasco	Left Defensive Midfield	140	38	2	102	36	18.4	Yes	45	Augusto Fernán	Tactical
Atlético Madrid	Saúl Ñíguez	Right Midfield	141	38	1	103	37	18.2				
Real Madrid	Isco	Left Center Midfield	108	64	0	44	64	17.2	Yes	71	Toni Kroos	Tactical
Real Madrid	Sergio Ramos	Left Center Back	125	32	3	93	29	16.6				
Real Madrid	Danilo	Right Back	115	51	0	64	51	16.6	Yes	51	Daniel Carvajal	Tactical
Atlético Madrid	Diego Godín	Left Center Back	136	24	1	112	23	16.3				
Real Madrid	Toni Kroos	Left Center Midfield	125	16	1	109	15	14.4				
Real Madrid	Karim Benzema	Center Forward	125	13	0	112	13	13.8				
Real Madrid	Pepe	Right Center Back	115	21	0	94	21	13.6				
Atlético Madrid	Augusto Fernández	Left Defensive Midfield	111	15	0	96	15	12.6				
Atlético Madrid	Stefan Savić	Right Center Back	112	12	0	100	12	12.4				
Real Madrid	Daniel Carvajal	Right Back	110	5	0	105	5	11.5				
Real Madrid	Lucas Vázquez	Center Forward	68	35	1	33	34	10.6	Yes	76	Karim Benzema	Tactical
Real Madrid	Keylor Navas	Goalkeeper	64	7	0	57	7	7.1				
Atlético Madrid	Fernando Torres	Right Center Forward	49	7	0	42	7	5.6				
Atlético Madrid	Jan Oblak	Goalkeeper	39	3	0	36	3	4.2				
Atlético Madrid	Lucas Hernandez	Left Back	15	0	0	15	0	1.5	Yes	108	Filipe Luis	Tactical
Atlético Madrid	Thomas Partey	Left Defensive Midfield	8	0	0	8	0	0.8	Yes	115	Koke	Tactical

(a) Attacking Scores

Team_name	Player	Position	Possession Events	Shot Events	Goal Events	Possession Only	Shot Only	Score	Substitute	Sub Entry Minute	Replacement	Substitution Type	
Real Madrid	Casemiro	Center Defensive Midfield	57	8	1	49	7	6.8					
Atlético Madrid	Gabi	Right Defensive Midfield	43	10	0	33	10	5.3					
Atlético Madrid	Saúl Ñíguez	Right Midfield	34	16	0	18	16	5					
Real Madrid	Marcelo	Left Back	34	8	1	26	7	4.5					
Atlético Madrid	Diego Godín	Left Center Back	36	9	0	27	9	4.5					
Real Madrid	Luka Modrić	Right Center Midfield	37	7	0	30	7	4.4					
Atlético Madrid	Yannick Carrasco	Left Defensive Midfield	27	15	0	12	15	4.2	Yes	45	Augusto Fernán	Tactical	
Real Madrid	Toni Kroos	Left Center Midfield	33	9	0	24	9	4.2					
Atlético Madrid	Stefan Savić	Right Center Back	33	7	0	26	7	4					
Atlético Madrid	Juanfran	Right Back	28	9	0	19	9	3.7					
Real Madrid	Cristiano Ronaldo	Left Wing	30	7	0	23	7	3.7					
Atlético Madrid	Koke	Left Midfield	27	8	0	19	8	3.5					
Atlético Madrid	Fernando Torres	Right Center Forward	25	8	0	17	8	3.3					
Real Madrid	Danilo	Right Back	27	5	0	22	5	3.2	Yes	51	Daniel Carvajal	Tactical	
Real Madrid	Gareth Bale	Right Wing	28	4	0	24	4	3.2					
Real Madrid	Pepe	Right Center Back	27	4	0	23	4	3.1					
Real Madrid	Sergio Ramos	Left Center Back	25	5	0	20	5	3					
Atlético Madrid	Filipe Luis	Left Back	22	7	0	15	7	2.9					
Atlético Madrid	Antoine Griezmann	Left Center Forward	25	3	0	22	3	2.8					
Real Madrid	Karim Benzema	Center Forward	26	2	0	24	2	2.8					
Real Madrid	Lucas Vázquez	Right Wing	25	2	0	23	2	2.7	Yes	76	Karim Benzema	Tactical	
Real Madrid	Isco	Left Center Midfield	20	2	0	18	2	2.2	Yes	71	Toni Kroos	Tactical	
Real Madrid	Daniel Carvajal	Right Back	10	2	0	8	2	1.2					
Atlético Madrid	Augusto Fernández	Left Defensive Midfield	11	1	0	10	1	1.2					
Atlético Madrid	Lucas Hernandez	Left Back	7	4	0	3	4	1.1	Yes	108	Filipe Luis	Tactical	
Atlético Madrid	Thomas Partey	Left Defensive Midfield	7	2	0	5	2	0.9	Yes	115	Koke	Tactical	
Real Madrid	Keylor Navas	Goalkeeper	7	0	0	7	0	0.7					
Atlético Madrid	Jan Oblak	Goalkeeper	1	1	0	0	1	0.2					
No defending events found for team: Atlético Madrid			0	0	0	0	0	0					

(b) Defending scores

Figure 25 Player Scores in Match 18243

(4) Position Based Expected Threat (xT)

The analysis of the possession chain and shot chain disclosed the involvement of the actions by the players, one might argue some actions were more valuable than other actions. So the approach was utilized by Karun Singh[9] to reward more to the actions that put the ball in the position that carried more threat.

Similar studies were conducted to both matches of 18241 and 18243 as well, for the single match, convergence check didn't really behave as expected, eventually to simplify, 10 moves were used arbitrarily to demonstrate the application as shown in Figure 26.

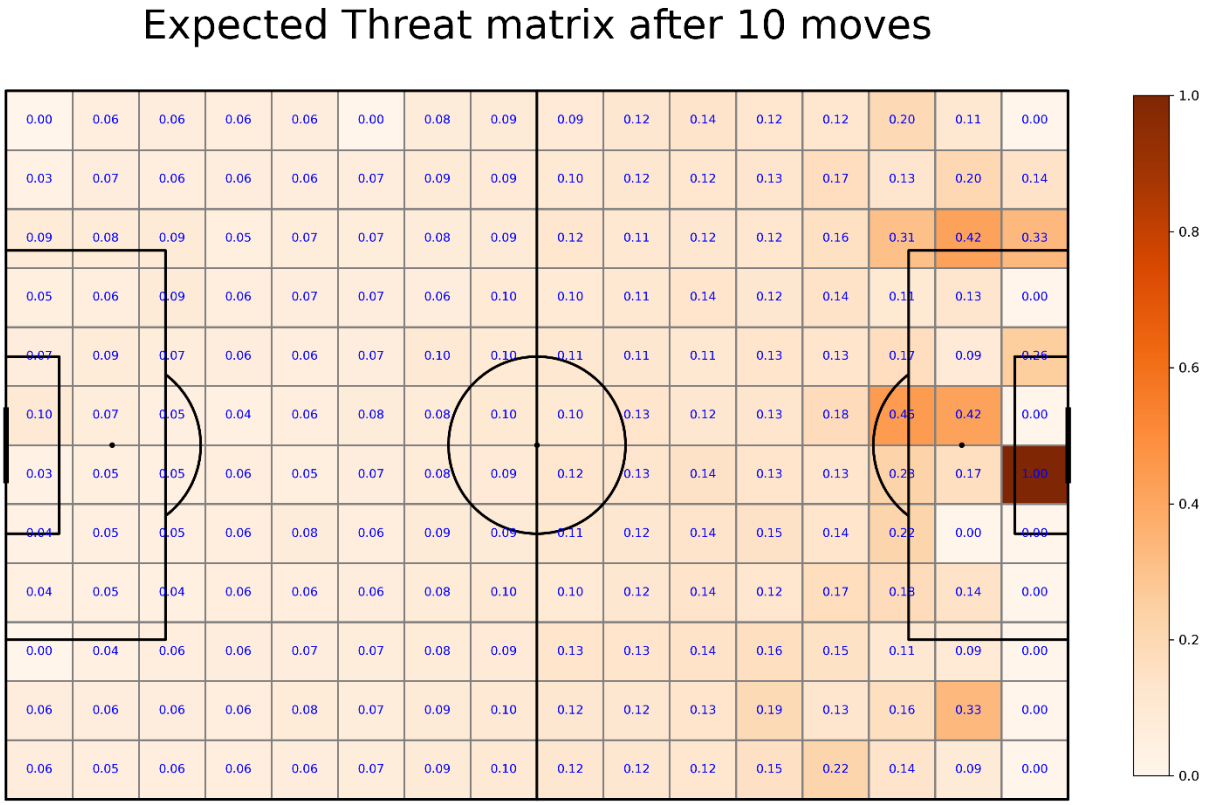


Figure 26 Match 18241 Expected Threat Matrix

Based on the expected threat, the attribution of the players for both teams could be summarized in **Error! Not a valid bookmark self-reference.**. By comparing this summary to the summary based on simple involvements studies as shown in **Figure 24 Player Scores in Match 18241Figure** 24, there were not too many difference, based on simple involvements attacking scores, the top 5 from Real Madrid were Modrić, Di Maria, Ramos, Christino Ronaldo and Carvajal, while based on expected threat studies, the top 5 from Real Madrid were Di Maria, Modrić, Marcelo, Carvajal, and Ramos. For Atlectico Madrid, based on

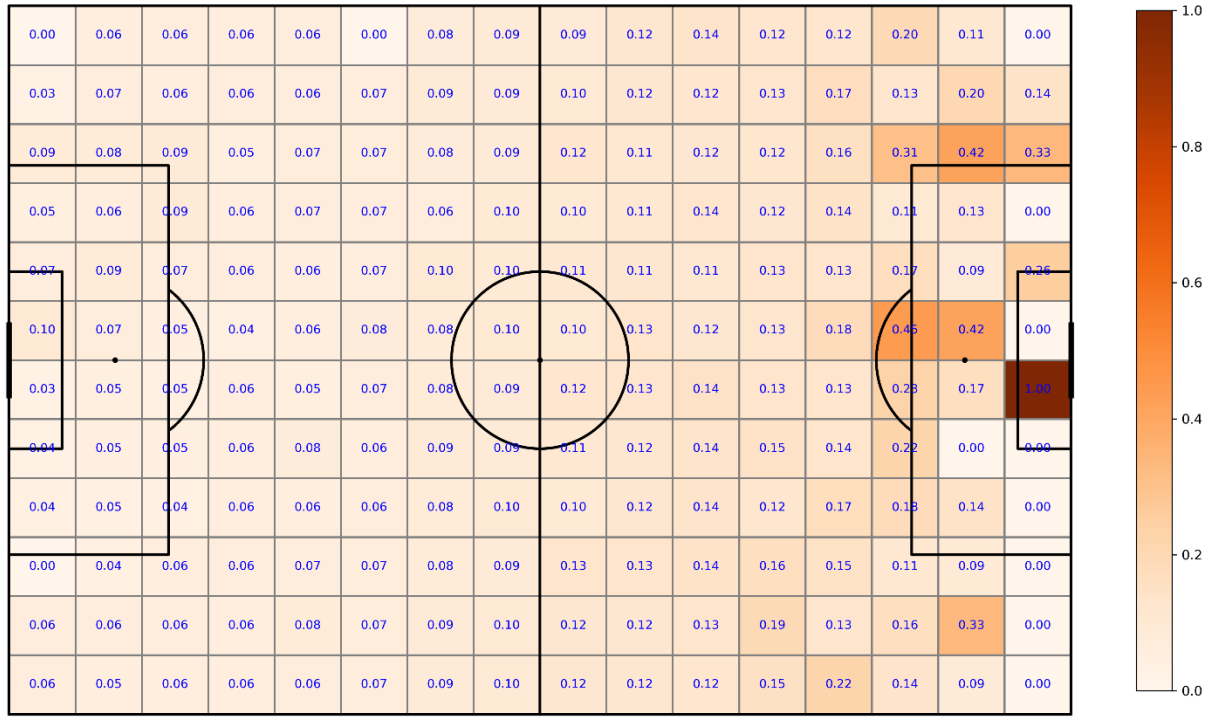


Figure 26

Table 7 Expected Threat Contribution from the Players in Match 18241

player_name	xT_added	team_name	position_name	is_substitute
Ángel Di María	2.846	Real Madrid	Left Center Midfield	FALSE
Luka Modrić	2.729	Real Madrid	Right Center Midfield	FALSE
Marcelo	1.998	Real Madrid	Left Back	TRUE
Daniel Carvajal	1.945	Real Madrid	Right Back	FALSE
Gabi	1.901	Atlético Madrid	Right Defensive Midfield	FALSE

Sergio Ramos	1.780	Real Madrid	Left Center Back	FALSE
Juanfran	1.433	Atlético Madrid	Right Back	FALSE
Isco	1.296	Real Madrid	Center Defensive Midfield	TRUE
Cristiano Ronaldo	1.217	Real Madrid	Left Wing	FALSE
Gareth Bale	1.136	Real Madrid	Right Wing	FALSE
Raphaël Varane	1.128	Real Madrid	Right Center Back	FALSE
Adrián López	1.041	Atlético Madrid	Right Center Forward	TRUE
Koke	1.038	Atlético Madrid	Left Midfield	FALSE
Iker Casillas	0.913	Real Madrid	Goalkeeper	FALSE
Thibaut Courtois	0.730	Atlético Madrid	Goalkeeper	FALSE
Tiago	0.623	Atlético Madrid	Left Defensive Midfield	FALSE
Filipe Luís	0.620	Atlético Madrid	Left Back	FALSE
Fábio Coentrão	0.401	Real Madrid	Left Back	FALSE
Álvaro Morata	0.368	Real Madrid	Center Forward	TRUE
João Miranda de Souza Filho	0.366	Atlético Madrid	Right Center Back	FALSE
José Sosa	0.317	Atlético Madrid	Right Midfield	TRUE
Sami Khedira	0.304	Real Madrid	Center Defensive Midfield	FALSE
Raúl García	0.267	Atlético Madrid	Right Midfield	FALSE
Toby Alderweireld	0.215	Atlético Madrid	Left Back	TRUE
Diego Godin	0.214	Atlético Madrid	Left Center Back	FALSE
Karim Benzema	0.187	Real Madrid	Center Forward	FALSE
David Villa	0.140	Atlético Madrid	Left Center Forward	FALSE
Diego Costa	0.022	Atlético Madrid	Right Center Forward	FALSE

The same studies were conducted on the match 18243, by comparing this summary as shown in

Table 8 to the summary based on simple involvements studies as shown in Figure 24 Player Scores in Match 18241Figure 25, similarly no too many difference detected. Based on simple involvements attacking scores, the top 5 from Real Madrid were Modrić, Bale, Marcelo, Casemiro and Christino Ronaldo, while based on expected threat studies, the top 5 from Real Madrid were Bale, Marcelo, Modrić, Casemiro, and Vázquez. For Atlectico Madrid, based on

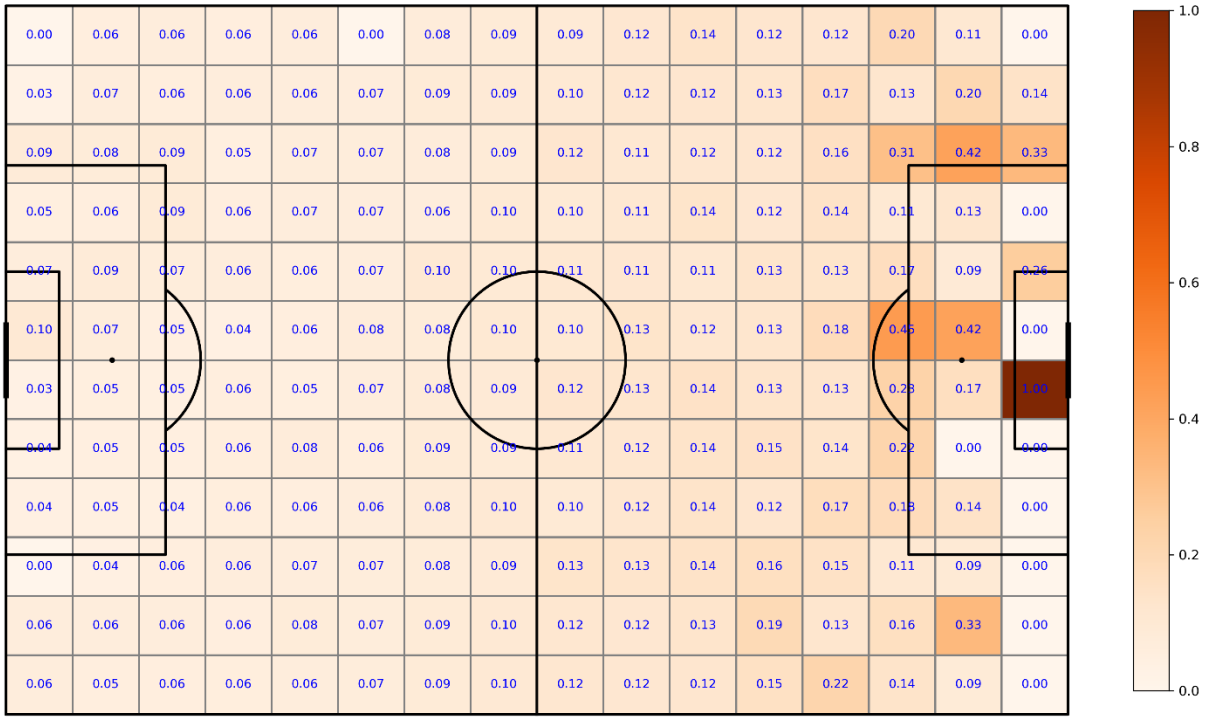


Figure 26In both cases, as the major shot and goal involvers in Real Madrid, Christino Ronaldo was replaced by somebody else in the expected threat studies, the possible reason could be the simple involvement studies gave the heavy weights on the events related to both shot chain and goal chains.

Similar to simple involvement attacking scores, the expected threat metrics could be used to judge the decision of the substitution, it indicted in Match 18241, the substitutions in Real Madrid significnatly improved the expected threat, while in Match 18243, the substitution in both Atlectico Madrid and Real Madrid greatly added their expected threat.

Table 8 Expected Threat Contribution from the Players in Match 18243

player_name	xT_added	team_name	position_name	is_substitute
Filipe Luís	1.919	Atlético Madrid	Left Back	FALSE
Gabi	1.659	Atlético Madrid	Right Defensive Midfield	FALSE
Yannick Carrasco	1.501	Atlético Madrid	Left Defensive Midfield	TRUE
Koke	1.453	Atlético Madrid	Left Midfield	FALSE
Gareth Bale	1.351	Real Madrid	Right Wing	FALSE
Marcelo	1.119	Real Madrid	Left Back	FALSE
Luka Modrić	1.073	Real Madrid	Right Center Midfield	FALSE
Juanfran	1.041	Atlético Madrid	Right Back	FALSE
Casemiro	0.974	Real Madrid	Center Defensive Midfield	FALSE

Lucas Vázquez	0.952	Real Madrid	Right Wing	TRUE
Isco	0.933	Real Madrid	Left Center Midfield	TRUE
Antoine Griezmann	0.848	Atlético Madrid	Left Center Forward	FALSE
Diego Godín	0.687	Atlético Madrid	Left Center Back	FALSE
Keylor Navas	0.658	Real Madrid	Goalkeeper	FALSE
Danilo	0.647	Real Madrid	Right Back	TRUE
Sergio Ramos	0.565	Real Madrid	Left Center Back	FALSE
Toni Kroos	0.555	Real Madrid	Left Center Midfield	FALSE
Saúl Ñíguez	0.539	Atlético Madrid	Right Midfield	FALSE
Augusto Fernández	0.519	Atlético Madrid	Left Defensive Midfield	FALSE
Karim Benzema	0.496	Real Madrid	Center Forward	FALSE
Stefan Savić	0.484	Atlético Madrid	Right Center Back	FALSE
Cristiano Ronaldo	0.454	Real Madrid	Left Wing	FALSE
Pepe	0.395	Real Madrid	Right Center Back	FALSE
Jan Oblak	0.281	Atlético Madrid	Goalkeeper	FALSE
Daniel Carvajal	0.244	Real Madrid	Right Back	FALSE
Fernando Torres	0.065	Atlético Madrid	Right Center Forward	FALSE
Lucas Hernández	0.051	Atlético Madrid	Left Back	TRUE
Thomas Partey	0.025	Atlético Madrid	Left Defensive Midfield	TRUE

(5) Voronoi diagram of the shot

All the above studies demonstrated the attacking actions due to the nature of the Statsbomb event data. Starting from (5) and (6), the defense behavior will be explained by combining the Statsbomb 360 data. A paper on the concept of expected disruption could be referenced to understand the work related to the defense[10]. Statsbomb 360 data was collected by a way called freeze_frame, in this frame, for a particular event, all visible players' position were recorded, however only the relationship between the event player and other players were described, as true for teammate, or false as opponent, no exact identifications were provided on who were those players. Also the data were not true track data, all other players outside this event freeze_frame were not included. At least, part of the defense actions on the events could be better understood by appending Statsbomb 360 data into the event data.

First project was conducted to draw the Voronoi diagram of the shots. There were no free Statsbomb 360 data for match 18241 and 18243. The author decided to use Croatia World Cup 2022 matches for the investigation. Also not all events had matching 360 data. When combined the event data and 360 data, and filtered for the study, there was possibility that some 360 data didn't exist. It is encouraged to read how Voronoi diagram was used to evaluate the players' position[11], the purpose of the paper was to show how Statsbomb data could be used to conduct something similar. As explained, Voronoi diagram might not be the best tool to explain why some of shot could be not converted to the goal, but at least to provide a starting point for the further evaluation. The Voronoi diagrams of Brazil and Croatia were demonstrated in

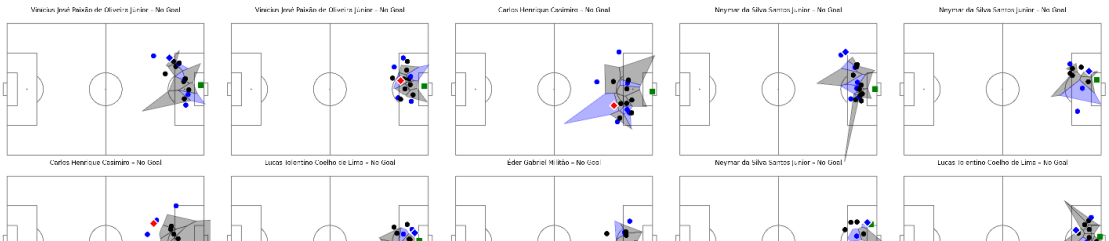


Figure 27 and

Figure 28. Generally speaking, both team defended relatively well, neither team gave the opponent teams much space to score.

1

10

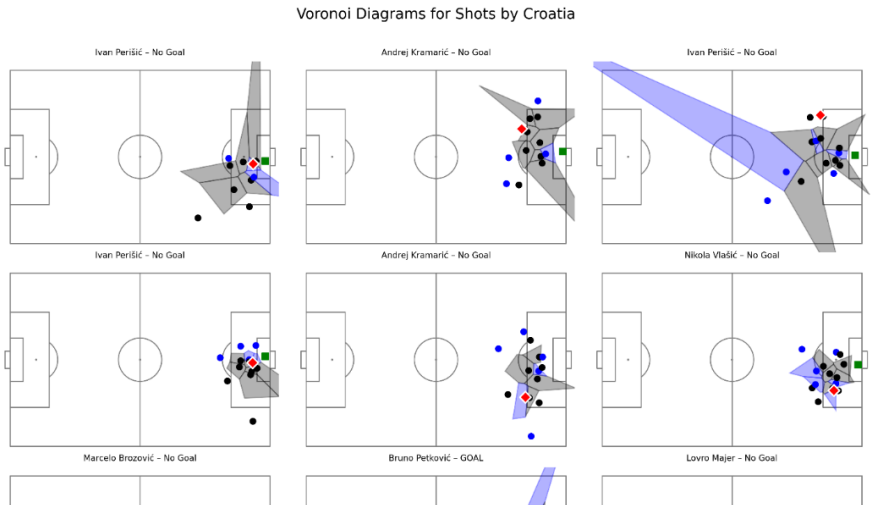
8

19

23

27

Figure 27 Voronoi Diagram for Shots By Brazil



12

22 25

Figure 28 Voronoi Diagram for Shots By Croatia

Based on the Vonoroi diagram, the defending action could be further evaluated quantitatively as shown in the

Table 9. Defense area was used to evaluate the quality for both teams. An arbitrary number of max 75 was selected as the threshold, there were total 10 shots met the threshold, among them, 7 were assigned to Brazil, which indicated the extraordinary shooting capability of Brazil team over powered the defending action from Croatia team, however only 1 turned to be goal, it was largely contributed by the excellent action by the goal keeper of Croatia team. On other side, 3 shots generated by the Croatia team belonged to the group of low defense area, also explained the better coordination of Brazil defending effort. Finally the victory was with Croatia team, it was another evidence the winning didn't always follow the seemed better team. The quality of the team requires continuous improvement (not just satisfied with better) to secure the guaranteed success. For an example, for Brazil team. if there could be a way for them to recognize the uniqueness of the goal keeper of the Croatia team during the period 1 and 2 (4 shots with very low defensive areas were saved), should they practice a different way for shot, therefore, 2 other shots in period 3 and 4 with low defensive area might score? For Croatia team, the problem was their shot on target was too low, 11.1% comparing to 52.6% from Brazil team, should they adjust their shooting position or shooting technique to improve the on target rate? It was true, they won this match, but at next Semi final match, their shot on target rate was still low at 14.3% while the opponent team Argentina was at 71.4%. The quality of the team requires continuous improvement, especially on many details (not just satisfied with better) to secure the guaranteed success.

Table 9 Voronoi quantitative analysis

Shot	match_id	team	shooter	period	play_pattern	is_goal	on_target	shot_outcome	shooter_area	nearest_defender	defensive_area
0	3869420	Brazil	Vinícius	1	From Free Kick	FALSE	TRUE	Saved	NaN	3.3	60.5
1	3869420	Croatia	Ivan Perišić	1	From Throw In	FALSE	FALSE	Off T	21.8	1.8	401.6
2	3869420	Croatia	Andrej Kramarić	1	Regular Play	FALSE	FALSE	Blocked	NaN	2.9	159.4
3	3869420	Brazil	Vinícius	1	Regular Play	FALSE	FALSE	Blocked	28.2	2.2	133
4	3869420	Brazil	Casemiro	1	Regular Play	FALSE	FALSE	Blocked	255.2	4.2	85.2
5	3869420	Croatia	Ivan Perišić	1	Regular Play	FALSE	FALSE	Off T	NaN	1.4	0
6	3869420	Brazil	Neymar	1	From Free Kick	FALSE	TRUE	Saved	NaN	10.4	0
7	3869420	Croatia	Ivan Perišić	2	From Corner	FALSE	FALSE	Wayward	3.3	0.4	398.3
8	3869420	Brazil	Neymar	2	From Goal Kick	FALSE	TRUE	Saved	NaN	0.8	68.5
9	3869420	Brazil	Casemiro	2	From Free Kick	FALSE	FALSE	Blocked	NaN	10.9	0
10	3869420	Brazil	Lucas Paquetá	2	From Throw In	FALSE	TRUE	Saved	NaN	5.1	40.2
11	3869420	Brazil	Éder Militão	2	From Corner	FALSE	FALSE	Blocked	17.4	1.8	149.8
12	3869420	Croatia	Andrej Kramarić	2	From Free Kick	FALSE	FALSE	Blocked	81.8	1.8	36
13	3869420	Brazil	Neymar	2	Regular Play	FALSE	TRUE	Saved	12	0.3	105.6
14	3869420	Brazil	Lucas Paquetá	2	From Throw In	FALSE	TRUE	Saved	NaN	2.4	177
15	3869420	Brazil	Richarlison	2	Regular Play	FALSE	FALSE	Off T	39.8	2.5	164.9
16	3869420	Croatia	Nikola Vlašić	2	From Throw In	FALSE	FALSE	Blocked	27.6	1.3	88
17	3869420	Brazil	Éder Militão	2	Regular Play	FALSE	FALSE	Blocked	41.8	1.4	314.6
18	3869420	Brazil	Éder Militão	2	Regular Play	FALSE	FALSE	Blocked	21.2	0.8	228.3
19	3869420	Brazil	Antony	2	From Free Kick	FALSE	TRUE	Saved	5059.9	1.7	66.5
20	3869420	Brazil	Pedro Guilherme	3	From Corner	FALSE	FALSE	Wayward	41	0.6	5914.4
21	3869420	Brazil	Pedro Guilherme	3	From Free Kick	FALSE	TRUE	Saved	22.2	1	102.4
22	3869420	Croatia	Marcelo Brozović	3	Regular Play	FALSE	FALSE	Off T	NaN	3.8	31.6
23	3869420	Brazil	Danilo	3	From Throw In	FALSE	FALSE	Off T	NaN	2.8	40.1
24	3869420	Brazil	Neymar	3	From Throw In	TRUE	TRUE	Goal	NaN	2.4	37.3
25	3869420	Croatia	Bruno Petković	4	Regular Play	TRUE	TRUE	Goal	27.3	4.2	71.1
26	3869420	Croatia	Lovro Majer	4	From Free Kick	FALSE	FALSE	Off T	288.1	4	159.8
27	3869420	Brazil	Casemiro	4	From Free Kick	FALSE	TRUE	Saved	NaN	3.9	33.1

(6) Pitch Control of the shot

In last part of this module, the pitch control concept was briefly investigated as well by applying to Statsbomb data. Voronoi diagram produces a quick snapshot on the spatial dominance. But it didn't take into the consideration of the player's difference on the speed, direction and acceleration. The investigation was first introduced by William Spearman[12].

Statsbomb 360 data is very limited on the track data, therefore it might not make sense to produce the convincible pitch control investigation. But the study was conducted nevertheless to demonstrate the understanding of the value of actions might still be squeezed with very limited information. To simplify, only 6 events in a shot sequence (5 plus shot) were studied. An example of the shot sequence from Croatia team by Bruno Petković to equalize the match was showed Figure 29. In addition to final control, the pass from Orsić and the ball receipt of Petković seemed to indicate Brazil team didn't control the pitch very well, thus provided the enough space for the easy passing and ball receiving to result in final shot.

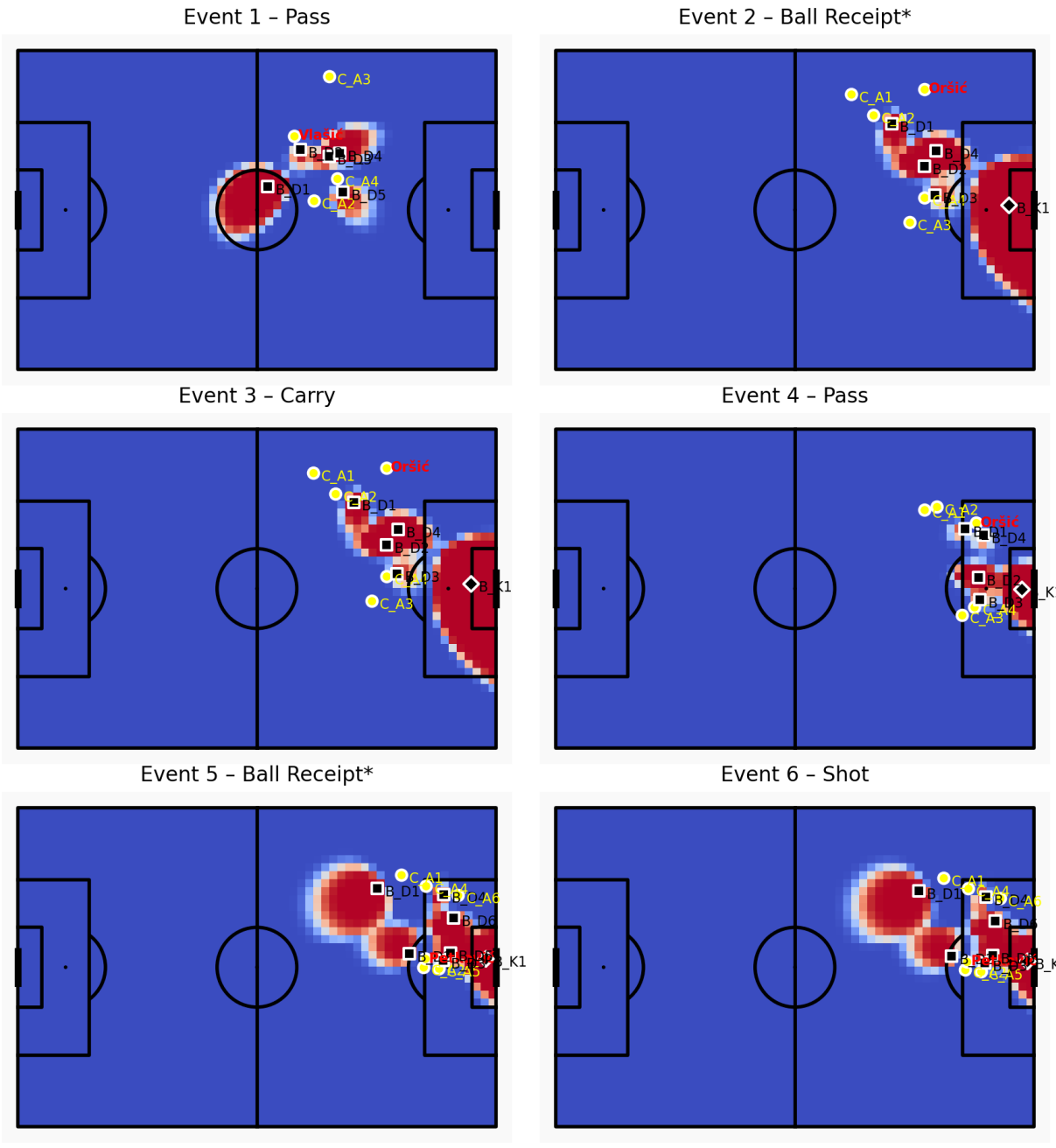


Figure 29 An example of the pitch control of the shot sequence

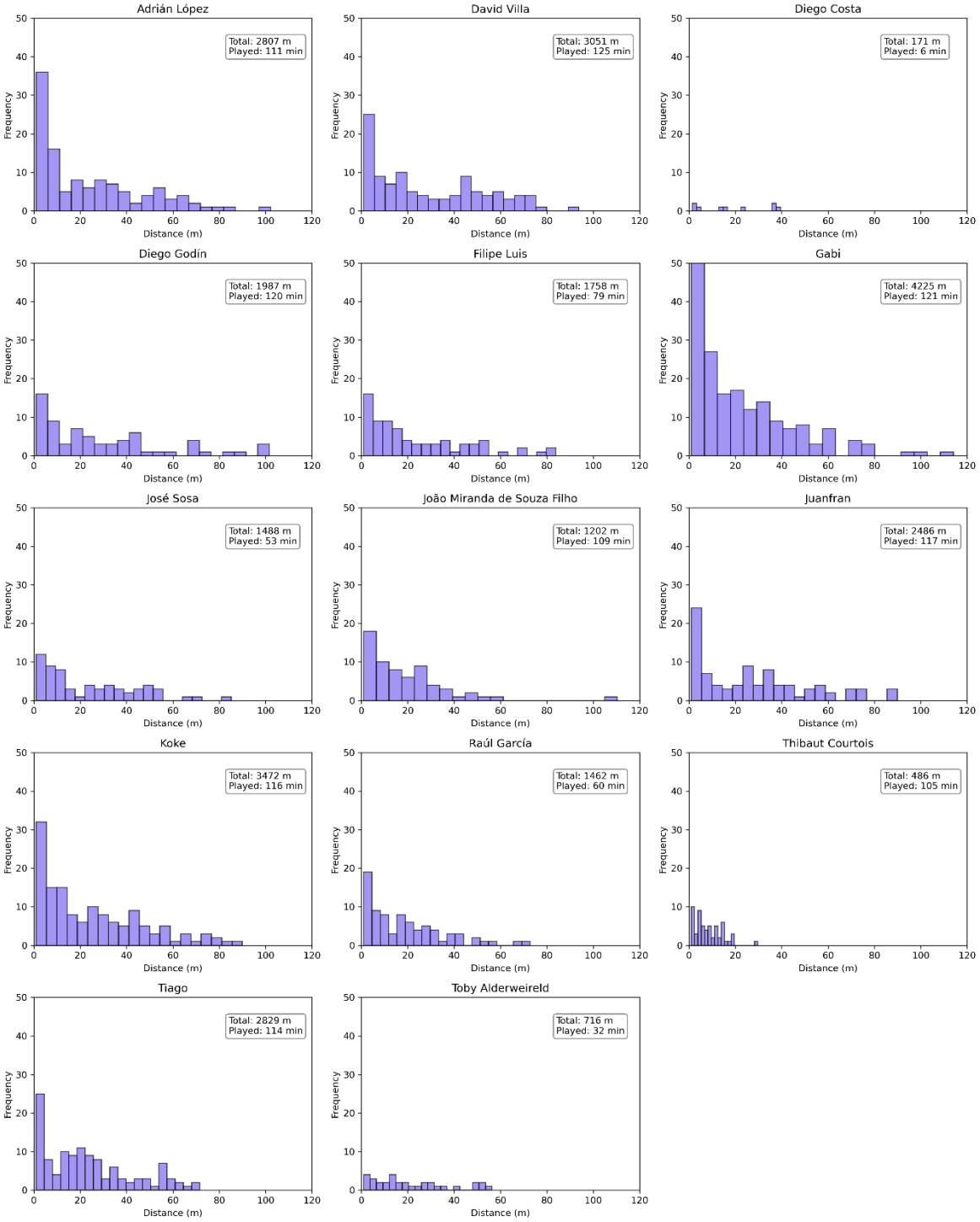
Module 7 Player Physics in Match

In final module of this paper, the player physics was discussed, as mentioned before, since the track data was very limited in Statsbomb, a lot of assumptions were made. The challenge was set up to extracted some information by the limited studies even though there were inevitable flaws.

In this study, the match 18241 was evaluated to understand the player's physics in the match. The player's physics are the major contributing factors to decide the tactical feasibility. The major investigated characteristics here include the speed, acceleration and deccelation, the total effort and metabolism cost related. The comparisons between teams and the players were plotted.

First, the comparisons of the histogram of the distances covered by players from two teams were summarized in Figure 30. Even the data was extracted from the event not the full track information, the assumption was the player's behaviors were consistent, even only pieces of their actions were presented, they were sufficient to represent the behaviors in the full match. In this comparison, it clearly showed the players from Real Madrid moved much more than their counterparts. Next, the position heatmap and directional movements of the players from both teams were summarized in Figure 31. These plots effectively described the midfield dominance of Real Madrid with Modrić as the web builder (a way to explain B2B and S2S) and Di María as the cutting knife to support left dominant tactics of attacking. Comparing the defenders of both teams revealed the game state: Atlético spent much of their resource on defending and sitting deep even the forward Villa participated heavily in the defending, while Real Madrid spent much less. The importance of two full backs of the Real Madrid Dani Cavajal and Marcelo were revealed to support both attacking flank especially on the left and defending. From attacking point of view, all attacking forces from Real Madrid Christino Ronaldo, Benzema, Isco and Bale moved relentlessly side to side to open the space. This could be the only workable tactics since Atlético was really condensed in the central.

Distance Histograms – Atlético Madrid



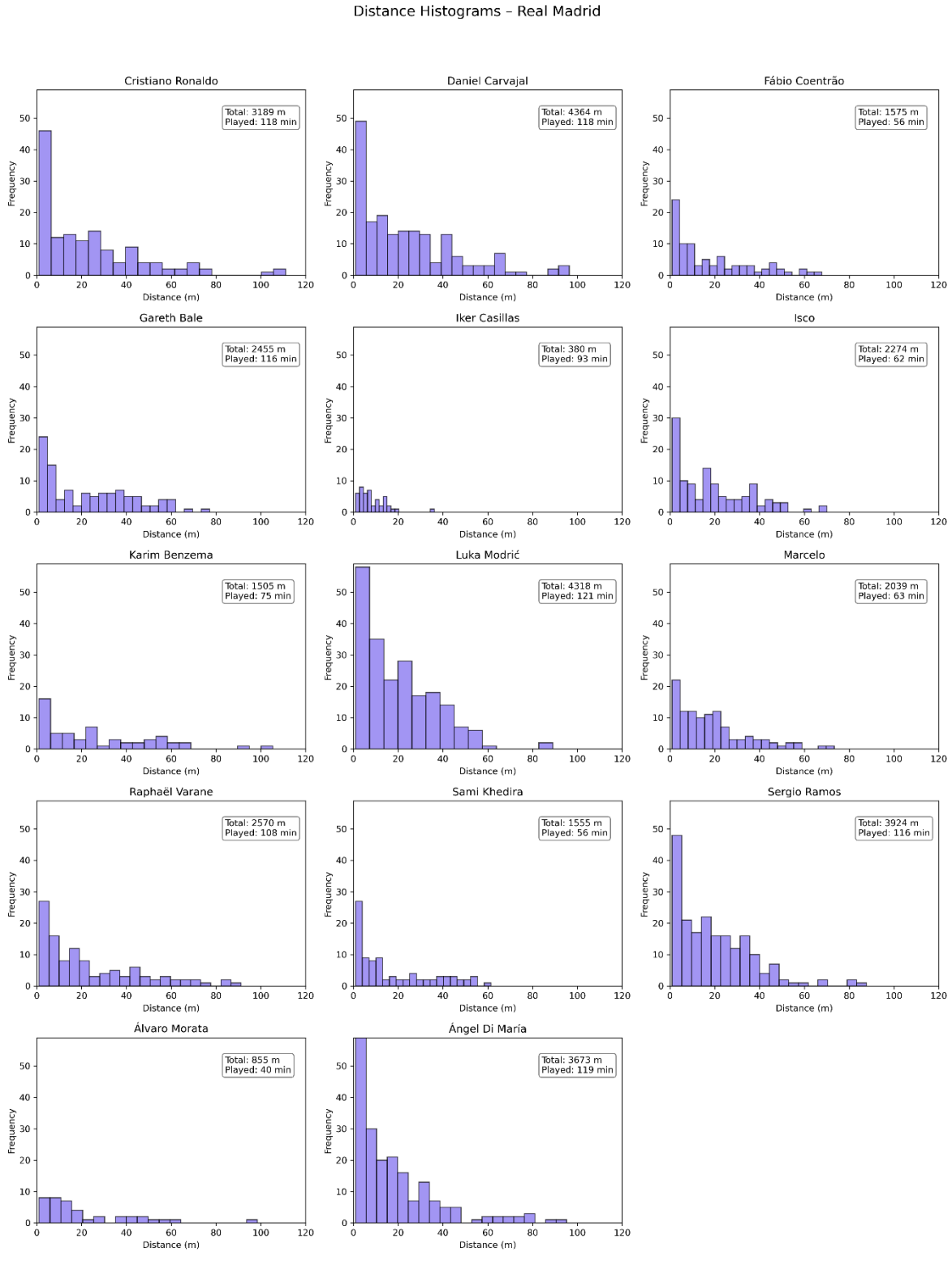
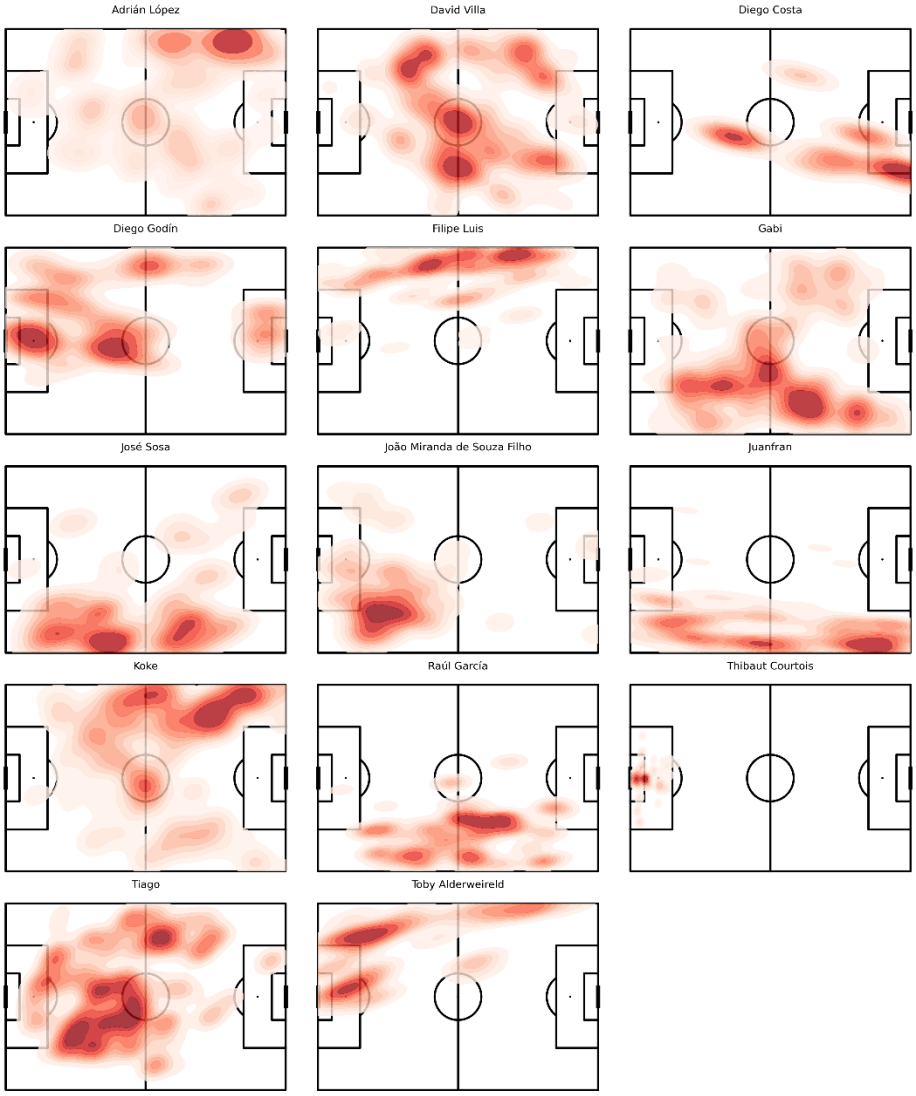
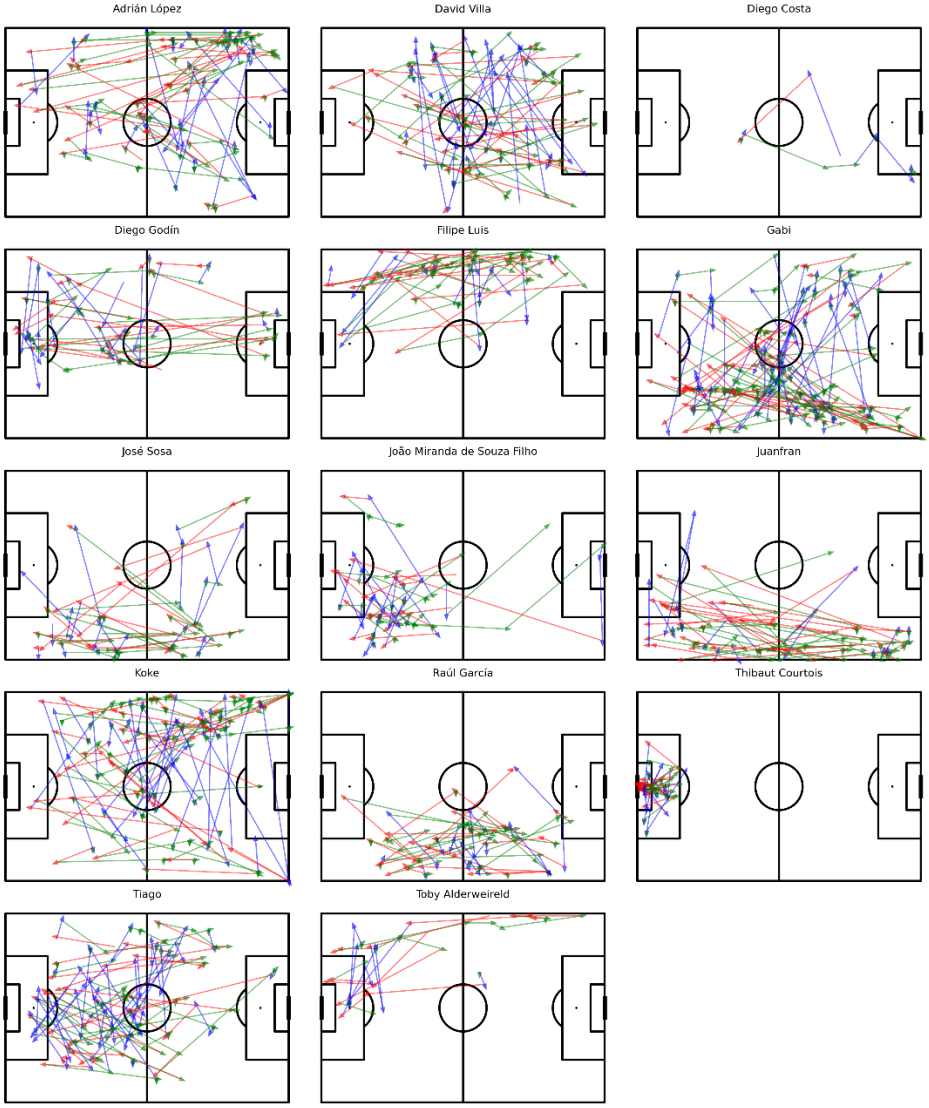


Figure 30 Distances Covered by the players in Match 18241

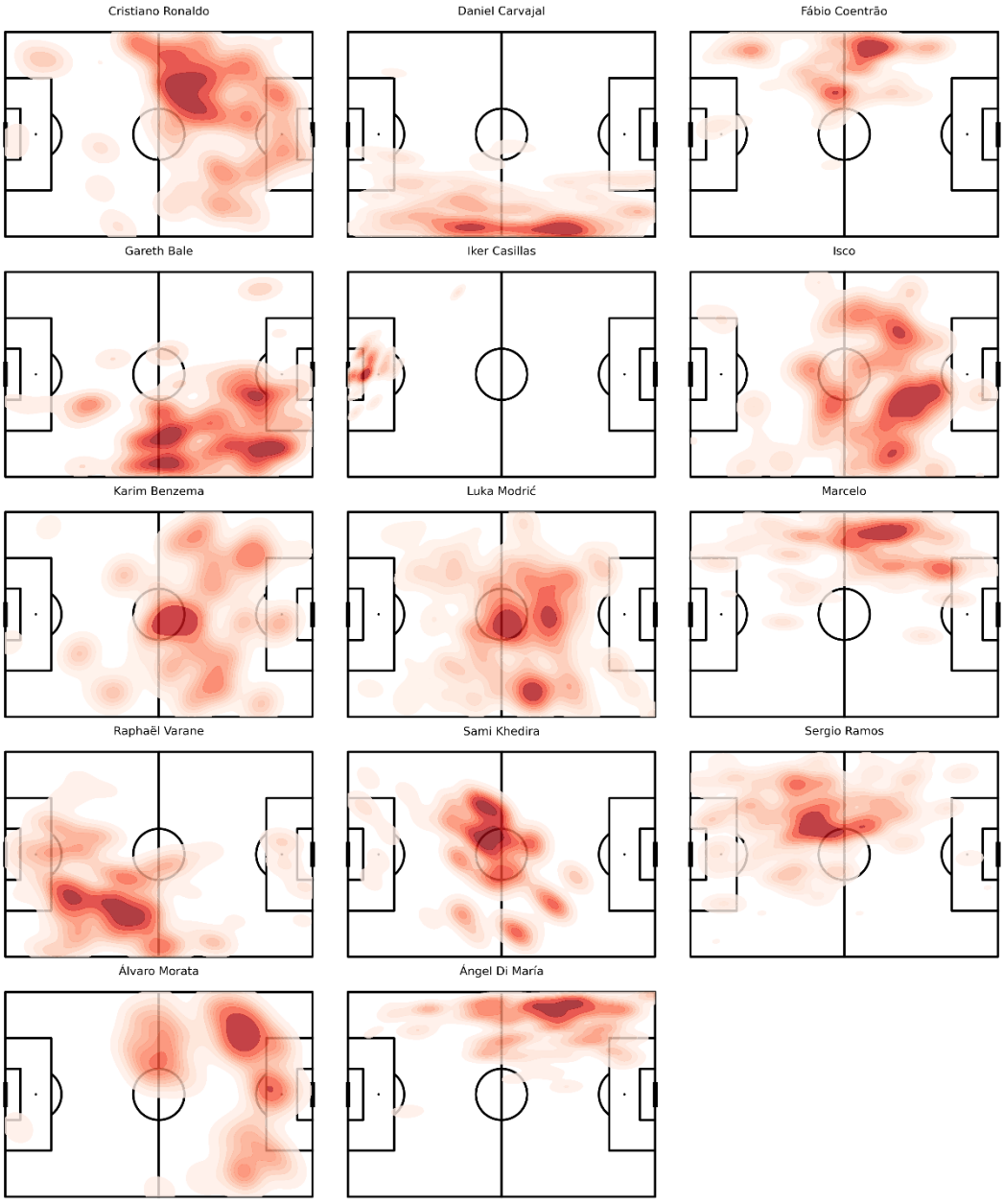
Positional Heat Maps – Atlético Madrid



Directional Movement Maps – Atlético Madrid



Positional Heat Maps – Real Madrid



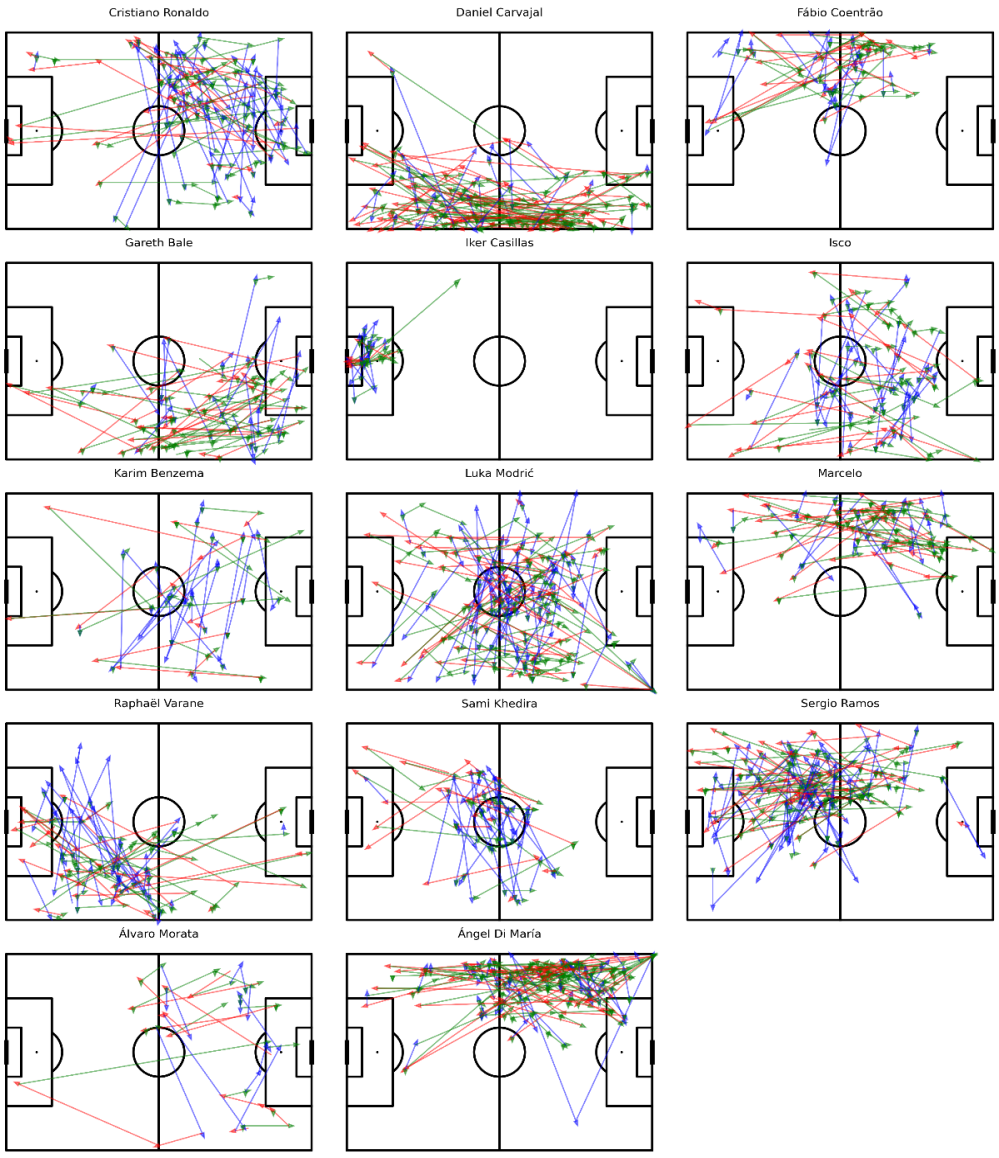


Figure 31 Positional Heatmap and Directional Movement for Match 18241

Another study was to compare the players from both teams on their efforts combined. This study helped to identify the players with high and low physics, maybe useful to for team management and rotations. Also in the long run it could be useful for the recruitment. The further analysis on the movement profile of each individual players could be plotted as well as shown in Figure 33 .

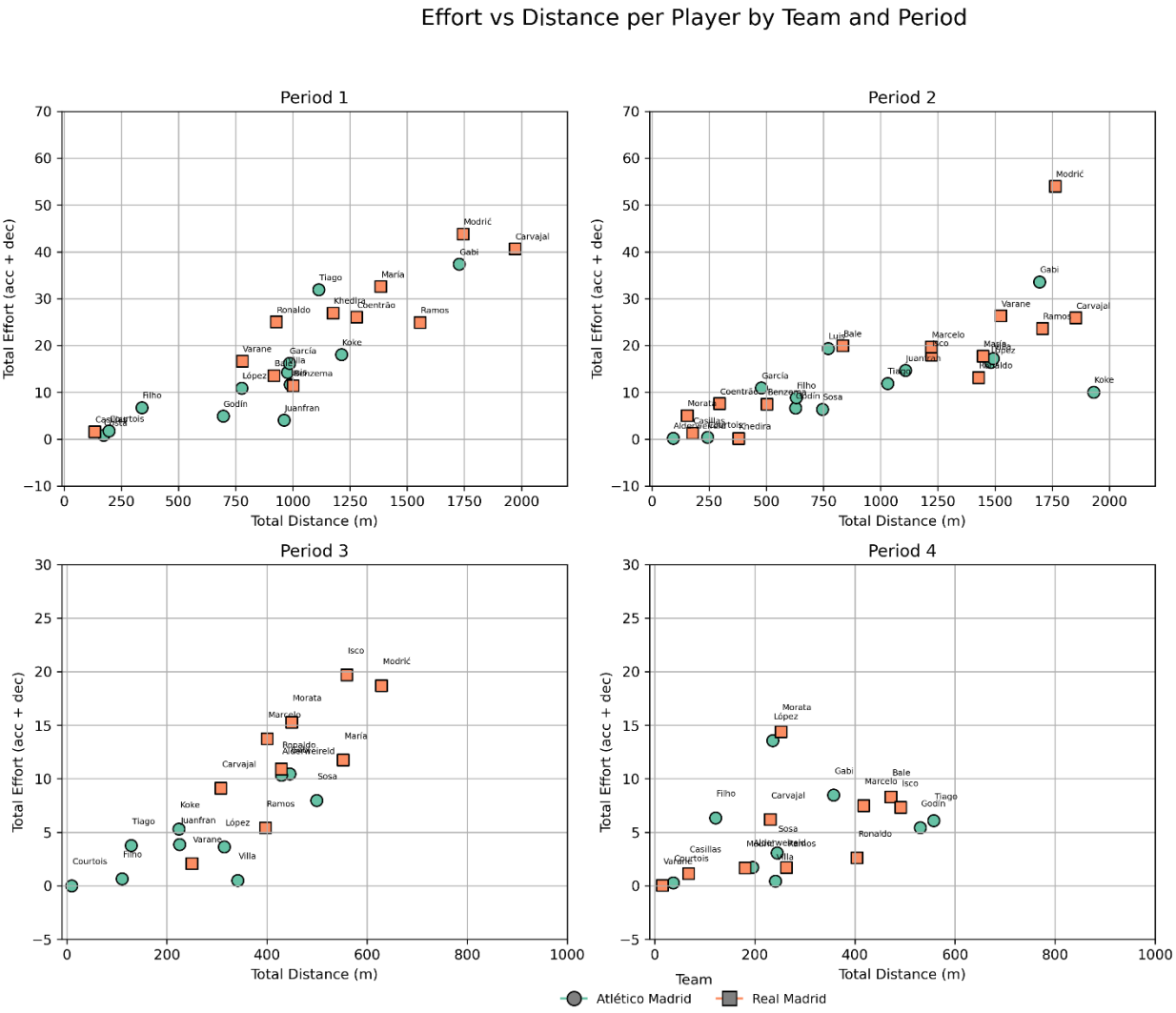


Figure 32 Evaluation of the players effort in Match 18241

Movement Profiles – Atlético Madrid

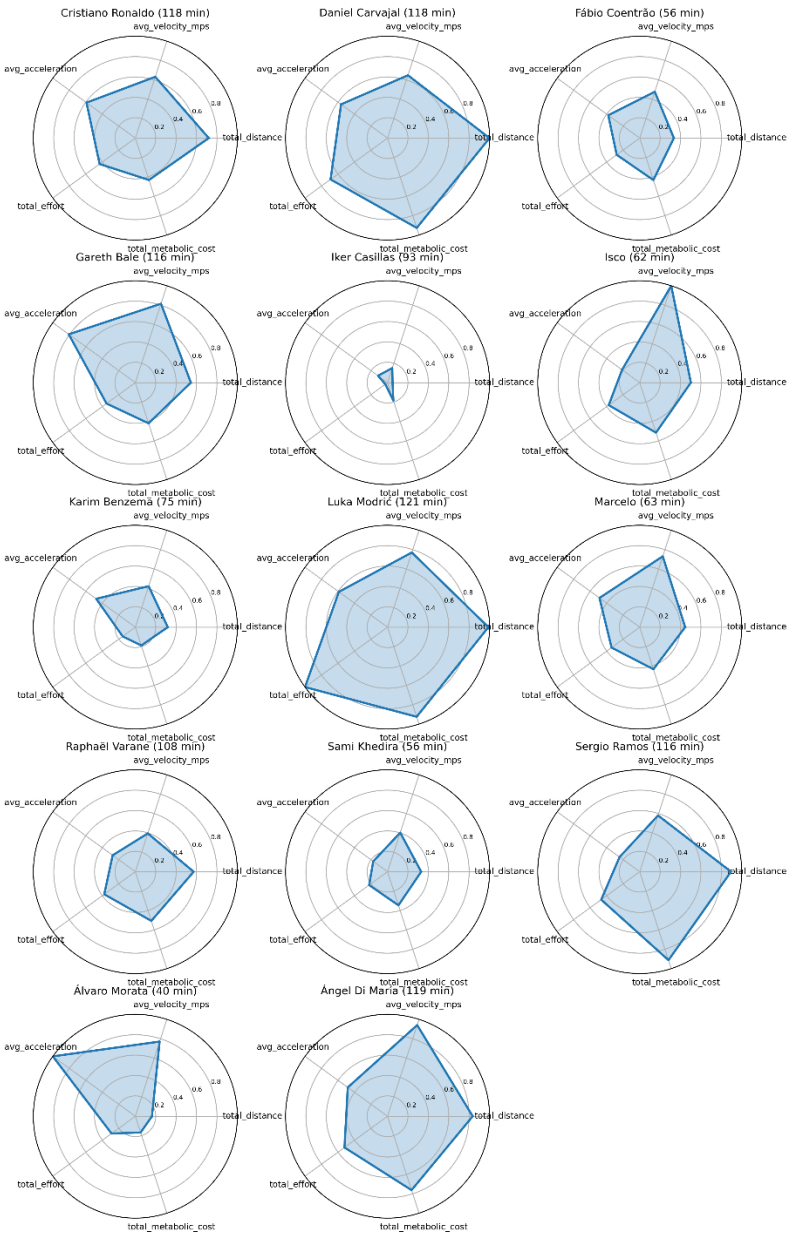


Figure 33 Movement Profiles of the Players in Match 18241

The players velocity vs. minutes could be plotted as well as showed in Figure 34. This plot could be used to explain the speed profile from both teams. And maybe in the real match, the team could adjust their tactics based on the monitoring.

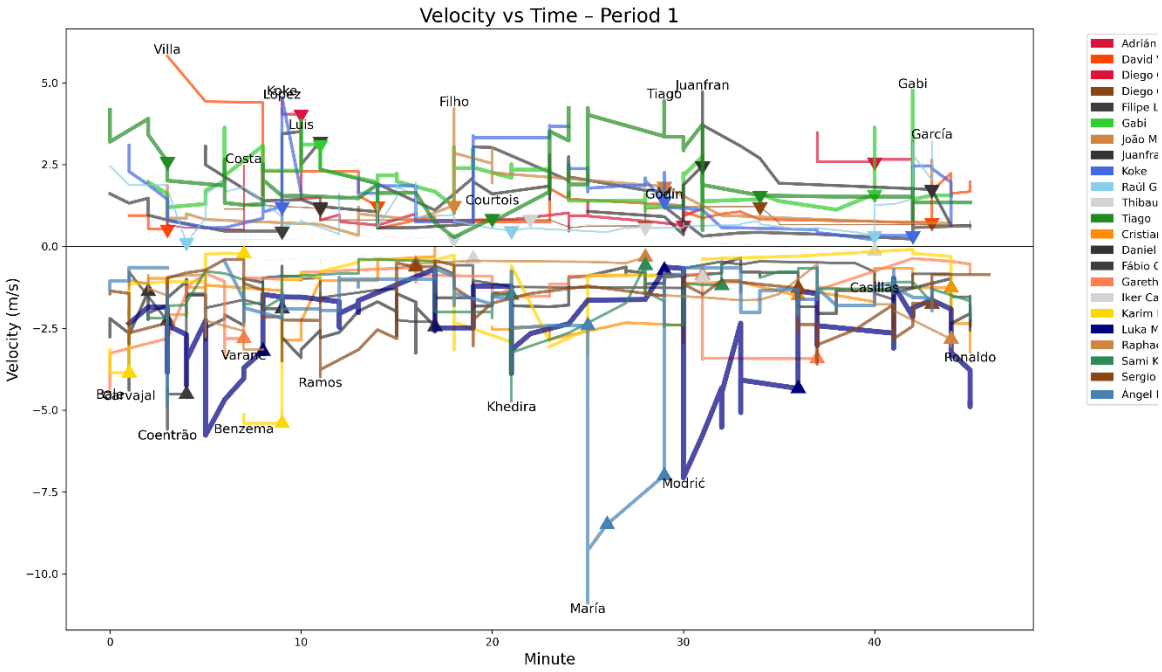
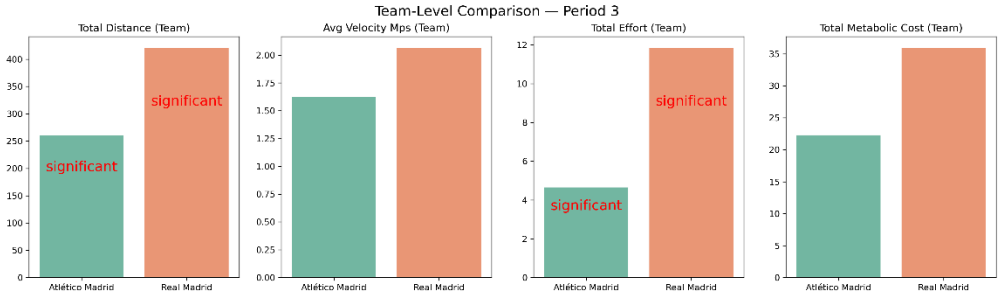


Figure 34 An example to compare the velocity from the both teams in Match 18241

The last part of the study was to compare team against team, role against role within different periods. An example of the team comparison and role comparison were summarized in



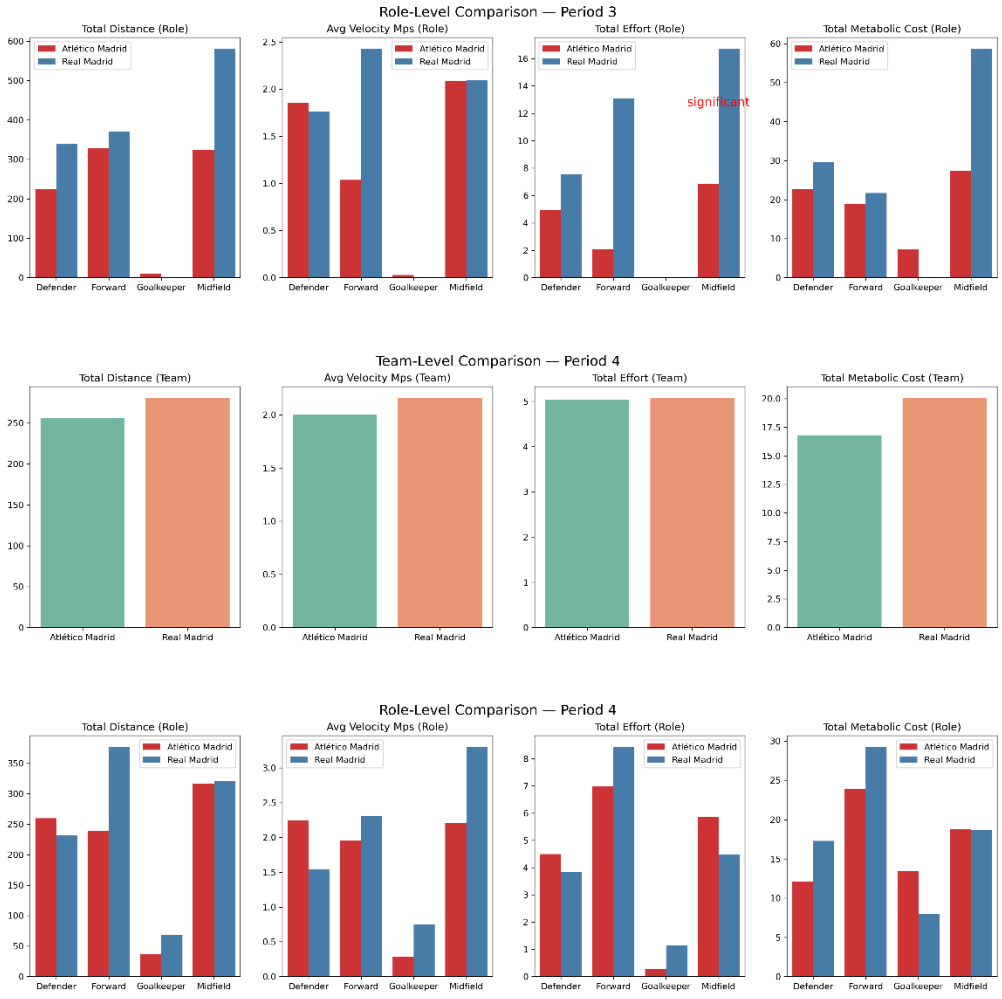
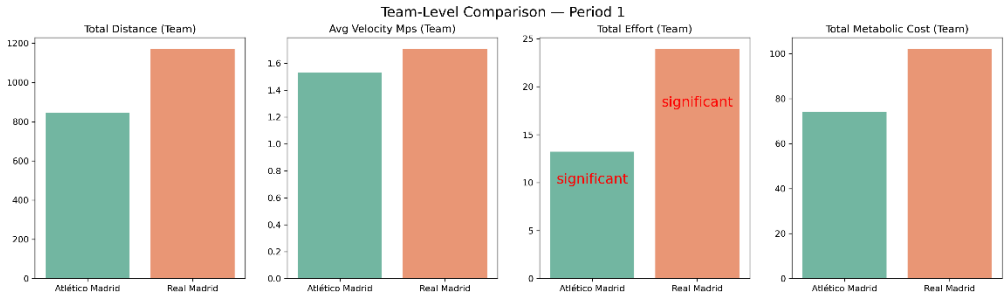
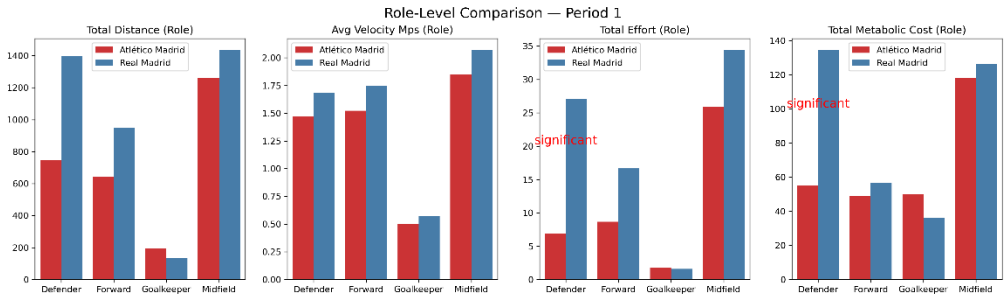
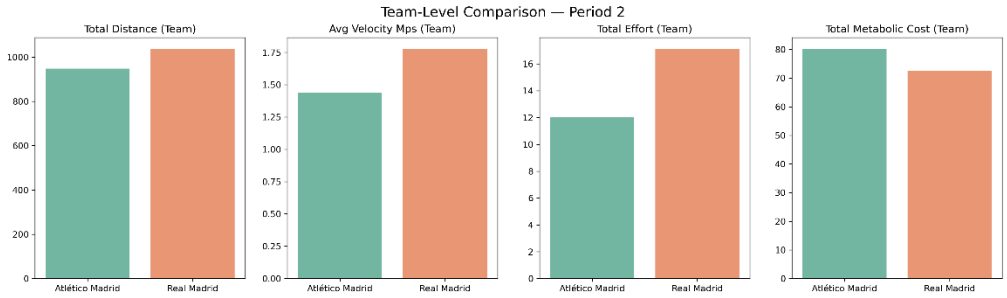
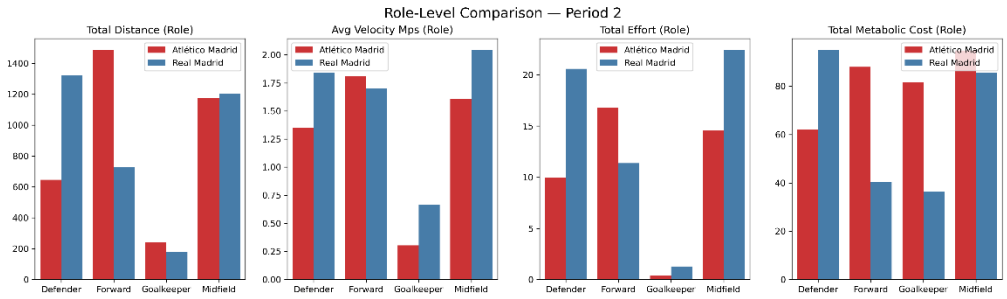


Figure 35. Even Real Madrid showed higher number in most of items , for both period 1 and 2, there were only one item was significantly different, So at least based on the player physics, Atlético Madrid was not really over powered by Real Madrid. But in period 3, it seemed the player's physics showed more gap between Atlético Madrid and Real Madrid.


Team-Level Comparison — Period 1
Total Distance (Team)
Avg Velocity Mps (Team)
Total Effort (Team)
Total Metabolic Cost (Team)
significant
significant
Atlético Madrid
Real Madrid


Role-Level Comparison — Period 1
Total Distance (Role)
Avg Velocity Mps (Role)
Total Effort (Role)
Total Metabolic Cost (Role)
significant
significant
Atlético Madrid
Real Madrid
Defender
Forward
Goalkeeper
Midfield


Team-Level Comparison — Period 2
Total Distance (Team)
Avg Velocity Mps (Team)
Total Effort (Team)
Total Metabolic Cost (Team)
Atlético Madrid
Real Madrid


Role-Level Comparison — Period 2
Total Distance (Role)
Avg Velocity Mps (Role)
Total Effort (Role)
Total Metabolic Cost (Role)
Atlético Madrid
Real Madrid
Defender
Forward
Goalkeeper
Midfield

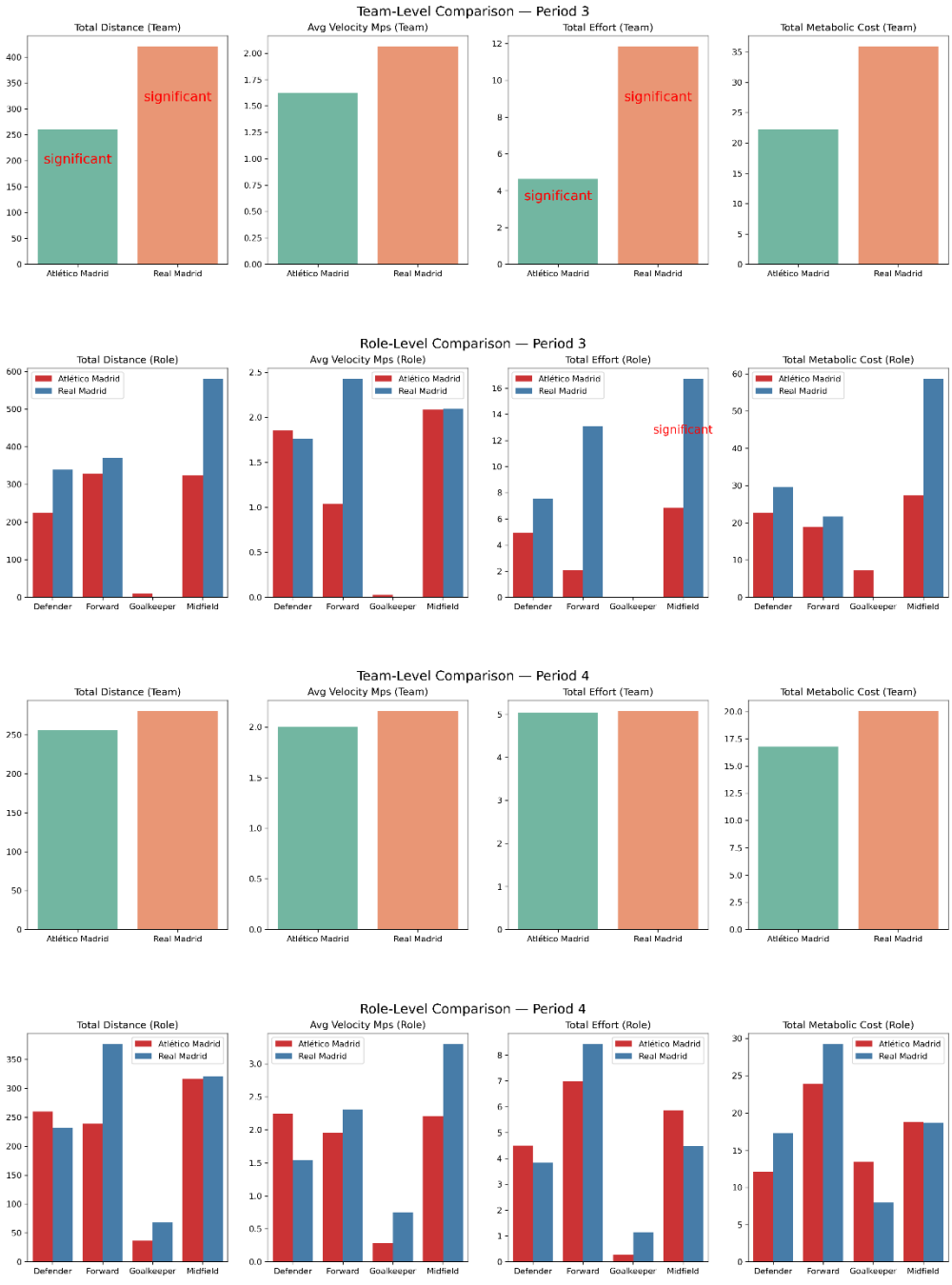


Figure 35 Team and Role Comparison in Match 18241

Module 8 Why should we run data analysis

There are many reasons why data analysis could help for the football as a business[13]. In author's view, the match is the only focused point at this stage. The most interesting/annoying feature for the football match is uncertainty, contradictory to a lot of other sports, the team who has paper advantage is not guaranteed for winning. Furthermore, from the duration of time, dominance of a team can't last long, the constant topic is to rebuild. From the normal business point of view, it is strange, in the whole football industry, each club and the associated team could be defined as a "company" since it is profit orientated, for sure, there are many "products" produced by the company, but the most important one that brings the value to the "customers" aka fans or the person who pays for the company directly or indirectly, is the match result. The inconsistence of the match results in one season and the variation of the match results over the multiple seasons just means a bad quality control, it is hard to imagine a normal business in a realistic world can sustain with a bad quality control. From the science and technology point of view, this doesn't make sense either, both ask for continuous improvements but it is hard to explain how the football industry improves over the years.

Back to the football match, how to win or at least not lose consistently, seems an impossible task. But why? Mathematically it is believed that Football is the sport that the signal is always drowned down by noise. The system was designed to build multiple filters to degrade the scoring possibility. Therefore the football is considered as most uncertain sport event. However just because of high noise ratio, treating football as "Art" is a categorization error; it is scientifically a Complex Adaptive System (CAS) that requires systems engineering, not just intuition. As the system is mapped better, the "Art" (unexplained variance) shrinks, and the "Science" (repeatable process) expands.

If we really want to generate a new generation of football with much improved quality, let's at least treat the football as science and engineering and leverage from the most powerful AI wave that is coming. The industrial can continue their current strategies for the team building, training and preparing the game, it is not asked for get ridding of the existing ways. The industrial may not have full confidence on new data science, but at least an opening mind should be provided. Football is a multi-billion dollar industry that invests negligibly in innovation compared to other high-performance sectors, creating a massive market inefficiency. High-stakes industries like Aerospace and Automotive typically reinvest 5–15% of their revenue into Research & Development (R&D) to ensure survival and reliability. In contrast, even elite football clubs often allocate less than 1% of turnover to genuine data science or systems engineering. Most "innovation" budgets in football are actually spent on basic tracking tech (GPS) rather than proprietary research. The whole industry is probably still be considered as the entertainment based on the tribal knowledge. Football management suffers from a "Peak Performance Bias," whereas engineering prioritizes Reliability and Redundancy. The system could be engineered to improve Process Protocols, the teams that strictly adhere to these probability-maximizing

protocols reduce the "Noise" of the game, making winning a matter of statistical inevitability rather than luck.

The football industries seemed moving to the wrong direction. The current market trend of hyper-inflation in transfer fees represents an unsustainable economic model based on Outcome Bias. In traditional corporate finance, companies invest in assets that generate long-term value. In football, clubs invest hundreds of millions in biological assets (players) that follow a steep depreciation curve due to aging and injury risk. Relying on "Overspending" is financially inefficient because it attempts to solve structural problems with temporary personnel solutions. While Investment in R&D (Data Science, Medical Engineering, Tactical Modeling) creates Intellectual Property. Unlike a player, a proprietary algorithm or a recruitment model does not age, does not get injured, and does not demand a transfer. It is a permanent asset that compounds in value over time.

Modern football creates a fetishization of Physical Metrics (Sprint Speed, Distance Covered) while underestimating Cognitive Speed (Decision Making), despite the latter having a higher correlation with winning elite competitions. Human physiology has a hard ceiling. A player can only run so fast (approx. 36 km/h) and cover so much ground (approx. 12km/game). Focusing recruitment on physical outliers hits a point of diminishing returns. A system optimized for Positioning and Anticipation (R&D) creates "Virtual Speed." Prioritizing athletics over "Football IQ" is an engineering error—optimizing the hardware while ignoring the software. When a club spends 30% of its budget on one superstar, the tactical system inevitably warps to accommodate them (Centralization). This creates a fragile system. If the expensive component fails (injury/form), the team has no fallback. “Over-thinking" means designing a squad where roles are defined by Statistical Profiles, not specific names. They didn't rely on one expensive savior; they relied on a *process* of chance creation. In an efficient market, the club cannot win by doing what everyone else is doing, just with more money. The club can only win by finding Information Asymmetry through R&D. Science and Engineering provide the edge that money and muscles cannot buy: Efficiency.

Module 9 Next Step

Sustainability requires process and system, not just the individual expensive assets. In a zero-sum game like football, the goal is not to be "perfect"; it is to widen the Delta between the team's performance and the opponent's.

Some investigations on the data science have been conducted[14]. The next step from the author's interest is to understand based on the following three pillars:

Control asymmetric Variance, victory is achieved when we operate within Six Sigma reliability (Order), while forcing the Opponent into High-Entropy states (Chaos). The Goa is to minimize our "Unforced Errors" while maximizing the opponent's "Forced Errors.". The internal engineering is to identify the correct CTQ that results in less variance, for an example, drill passing patterns and rest-defense structures so that players perform safely under pressure. Improve the set piece successful rate. We use data to ensure our "Floor" performance is high. DAMIC Process applied a data driven process to improve consistency. On the other side, counterpart focus to force the opponent out of their "Operational Envelope.". The data analysis can identify the specific trigger that causes the opponent's variance to spike. If the opponent's Center Back has a low "under-pressure passing score," we engineer a pressing trap specifically for him. We are not just pressing; we are targeting the high-variance node to induce a catastrophic failure.

Improve the decision-making speed/quality while slowing down or confusing the opponent's processing. Our internal objective is to improve Rationality (High Expected Value choices). We can train players (via VR or Video) to scan frequently and choose the option with the highest xT. We want players to act like algorithms—calm, calculated, and efficient. On the other side, add Cognitive Overload to the opponents**.** We inject "Noise" into the opponent's visual field, by using off ball action, we move the opponents in their least valued position and action.

Execute Systemic Game Theory (Masking vs. Exploiting), we understand our constrain very well and identify the most efficiency way to hide the liability. As for the component, we do not play our "standard" game; we tweak the system to apply maximum pressure on the opponent's weakest link.

Ultimately, this framework must be executed through a rigorous, data-driven lens. The objective of Data Science is not to displace traditional football methodology or serve as a solitary guide, but to inject a precise engineering mindset into a sport ripe for evolutionary optimization. We must move beyond the fatalistic acceptance of variance—the cliché that *'we had the best players and did our best, but football is unpredictable.'*

Set an audacious goal, refuse limits, start now, and build it step by step until it's real.

[1] The Three Types of Data in Football: Event, Tracking, and Physical - The Football Analyst

[2] open-data/doc/StatsBomb Open Data Specification v1.1.pdf at master · statsbomb/open-data · GitHub

[3] Getting Started — Soccermatics documentation

[4] Friends of Tracking - YouTube

[5] Javier M. Buldú, Javier Busquets, Johann H. Martínez, José L. Herrera-Diestra, Ignacio Echegoyen, Javier Galeano and Jordi Luque, Using Network Science to Analyse Football Passing Networks: Dynamics, Space, Time, and the Multilayer Nature of the Game, Front. Psychol., 07 October 2018, Sec. Movement Science, Volume 9 - 2018 | https://doi.org/10.3389/fpsyg.2018.01900

[6] Interactive Passing Networks

[7] Buldú, J.M., Busquets, J., Echegoyen, I. *et al.* Defining a historic football team: Using Network Science to analyze Guardiola's F.C. Barcelona. *Sci Rep* **9**, 13602 (2019). https://doi.org/10.1038/s41598-019-49969-2

[8] Kandaswamy, Suriya. 2020. Showing the Coach What He Can't See: Graphical Passing Networks as a Method of Soccer Team Analysis. Bachelor's thesis, Harvard College. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364740

[9] Introducing Expected Threat (xT)

[10] Unlocking Defensive Mastery: Introducing the Expected Disruption (xD) Model in Football Analytics | x-stats

[11] Visualizing positioning and player decisions: the innovation of Dynamic Pitch Control | Footovision

[12] https://www.researchgate.net/publication/327139841_Beyond_Expected_Goals

[13] The Growing Importance of Football Analytics - Soccerment

[14] Which Match Statistics Truly Correlate With Winning in Soccer's Top Leagues? | by Cenker Cengiz | Oct, 2025 | Medium