Methodology for Player Identification and Analysis for West Bromwich Albion

Detailed Walkthrough of the Approach and Assumptions

Written by

Marwane Hamdani

May 24, 2024

Abstract

West Bromwich Albion seeks to bolster its squad with young, versatile midfielders capable of excelling in various roles on the pitch. With a budget of £10 million for three positions, the club aims to identify players with significant resale value who can perform well in the Championship and potentially make the transition to the Premier League. This report outlines a meticulous methodology for identifying and analyzing potential targets, combining data-driven analysis with expert insights.

The process begins with player selection and data preparation, including filtering based on age and position capability, cleaning and normalizing data, and ensuring a minimum threshold of playing time in relevant positions. Ranking and scoring involve the development of robust scoring systems for both in-possession and out-of-possession performance, culminating in the calculation of overall scores for each player.

Comprehensive analysis is conducted to filter players based on percentile thresholds for in-possession and out-of-possession scores, ensuring the selection of top performers. The final selection process focuses on identifying players who demonstrate versatility across multiple positions, enhancing the team's adaptability on the field.

Contents

List of Figures

1 Introduction

West Bromwich Albion is seeking to strengthen their squad by identifying young, versatile midfielders who can excel both in and out of possession. The goal is to find players with the potential to perform at a high level in the Championship and possibly transition to the Premier League in the future. Given a total summer budget of 10 million GBP to cover three positions, it is crucial to target players with significant resale value.

This report provides a detailed walkthrough of the methodology used to identify and analyze potential player targets. The approach combines data-driven analysis and expert insights to create a shortlist of players who meet the club's strategic needs. The methodology involves several key steps:

- **Player Selection and Data Preparation:**
  - Selecting young players capable of playing multiple midfield roles.
  - Aggregating and normalizing data to ensure a comprehensive and accurate dataset.
  - Ensuring players have a minimum of 300 minutes played in at least two positions.
- **Ranking and Scoring:**
  - Developing a robust scoring system to evaluate players' performance in possession and out of possession.
  - Using weighted metrics to provide a fair and comparative analysis.
  - Calculating in-possession, out-of-possession, and overall scores for each player.
- **Comprehensive Analysis:**
  - Filtering and examining players' performance across multiple positions to ensure versatility and adaptability.
  - Setting percentile thresholds for both in-possession and out-of-possession scores.
  - Identifying top players based on their overall percentile rankings.

By following this methodology, West Bromwich Albion can make informed decisions on player acquisitions, ensuring they invest in talent that will enhance the team's performance and provide future value.

2 Data Filtering and Preparation

2.1 Initial Filtering

The initial filtering phase involved several critical steps to ensure the dataset was both comprehensive and accurate.

2.1.1 Position Capability and Youth

The first step was to select players younger than 25 years (by creating a new column for age based on the birthdate) and to focus on those capable of playing as an #8, single pivot, or double pivot, as our targets are originally midfielders.

```
# Define age threshold and relevant positions
age_threshold = 25
relevant_positions = ['single pivot', 'double pivot', '#8']

# Filter players by age and position
filtered_data = players[players['src_detailed_position'].isin(relevant_positions)]
filtered_data.head()
```

	competition_name	area_name	divisionlevel	season_name	src_detailed_position	player_name	team_name	birthdate	height	mins_played	player_match_possession	accelerations	ae
5	First League	Russia	2	2022/2023	single pivot	Khetag Khosonov	Alaniya Vladikavkaz	1998-06-18	185	2283	0.536588	8	
6	I Liga	Poland	2	2021/2022	double pivot	Josip Šoljić	Resovia Rzeszów	1987-06-18	189	1418	0.466018	2	
11	NB II	Hungary	2	2021/2022	double pivot	István Soltész	Budafoki MTE	2000-11-29	176	1517	0.522952	14	
12	1. Division	Cyprus	1	2023/2024	double pivot	Ousseynou Thioune	Anorthosis	1993-11-16	180	1662	0.560695	6	
16	Championship	England	2	2023/2024	#8	Gary Gardner	Birmingham City	1992-06-29	188	158	0.410280	2	

Figure 1: First filtering by position

2.1.2 Data Cleaning

A small percentage of `NaN` values were present in the filtered dataframe. These were dropped as they represented a tiny fraction of the data, and imputing them could have compromised data integrity, especially since some players were missing birthdates, a necessary criterion. Additionally, the column `pressing duels won` was dropped due to erroneous values of 0 throughout the dataframe.

```
[15] (filtered_data.isna().sum().sum()/len(filtered_data))*100

0.19568932252907778

[71] # Drop rows with NaN values
     cleaned_data = filtered_data.dropna()

     # Check the length of the cleaned data
     cleaned_length = len(cleaned_data)
     (cleaned_length/len(filtered_data))*100

99.84289730444848
```

Figure 2: NaN values filtering kept a high percentage of the data

2.1.3 Handling Duplicates

Several players had records across multiple seasons or the same season listed with different positions, resulting in duplicate entries. To address this issue, we adopted a strategy of aggregating player statistics from different seasons while maintaining separation based on their listed positions. For percentage-based metrics, such as aerial success rates, we calculated the mean value, whereas for cumulative statistics like total minutes played and goals scored, we computed the sum. Throughout this process, we ensured to retain all pertinent information including team affiliations, competition details, and divisions changes, ensuring the integrity of the dataset.

```
[25] young_players[young_players['player_name']=='István Soltész']
```

	competition_name	area_name	divisionlevel	season_name	src_detailed_position	player_name	team_name	birthdate	height	mins_played	player_match_possession	acceleratio
11	NB II	Hungary	2	2021/2022	double pivot	István Soltész	Budafoki MTE	2000-11-29	176	1517	0.522952	
40213	NB II	Hungary	2	2023/2024	double pivot	István Soltész	Budafoki MTE	2000-11-29	176	556	0.469360	
124224	NB II	Hungary	2	2022/2023	#8	István Soltész	Budafoki MTE	2000-11-29	176	1741	0.489839	
249977	NB I	Hungary	1	2020/2021	double pivot	István Soltész	Budafoki MTE	2000-11-29	176	163	0.492363	
360734	NB I	Hungary	1	2023/2024	double pivot	István Soltész	Zalaegerszegi TE	2000-11-29	176	128	0.645820	
364849	NB II	Hungary	2	2022/2023	single pivot	István Soltész	Budafoki MTE	2000-11-29	176	244	0.454591	
476947	NB II	Hungary	2	2021/2022	single pivot	István Soltész	Budafoki MTE	2000-11-29	176	34	0.408820	
479247	NB I	Hungary	1	2023/2024	#8	István Soltész	Zalaegerszegi TE	2000-11-29	176	21	0.277713	
479326	NB II	Hungary	2	2023/2024	#8	István Soltész	Budafoki MTE	2000-11-29	176	336	0.387685	
481541	NB I	Hungary	1	2023/2024	single pivot	István Soltész	Zalaegerszegi	2000-11-	176	56	0.491084	

Figure 3: Example of duplicate

	competition_name	area_name	division_levels	seasons_in_position	src_detailed_position	player_name	teams	birthdate	height	mins_played	player_match_possession	accele
0	NB II+NB I	Hungary	1+2	4	#8	István Soltész	Zalaegerszegi TE+Budafoki MTE	2000-11-29	176	2948	0.396588	
1	NB II+NB I	Hungary	1+2	4	double pivot	István Soltész	Zalaegerszegi TE+Budafoki MTE	2000-11-29	176	2413	0.528214	
2	NB II+NB I	Hungary	1+2	3	single pivot	István Soltész	Zalaegerszegi TE+Budafoki MTE	2000-11-29	176	334	0.451498	

Figure 4: Handling of duplicates

2.2 Final Adjustments

2.2.1 Further Filtering

After inspecting the data and its features, I deemed it appropriate to set a threshold of 300 minutes for a player's features in a specific position. Setting it higher might filter out too many players, while setting it lower could compromise data quality and include players who have not played enough in different positions, considering our goal of assessing versatility.

2.2.2 Stats Adjusted for Possession

The final list of players was set, and it was time to adjust the statistics based on possession to ensure a fair comparison. The player-match-possession was interpreted as the average possession for the player's team. We calculated the opponent possession time by taking the inverse of player-match-possession, considering 60

$$\text{effective\_playing\_time} = 60 \text{ minutes} \tag{1}$$

$$\text{opponent\_possession\_time} = \text{effective\_playing\_time} \times \left(1 - \frac{\text{player\_match\_possession}}{100}\right) \tag{2}$$

Next, we separated in-possession and out-of-possession statistics. For in-possession statistics, we normalized by dividing the values by the percentage of time the player's team had possession of the ball (player-match-possession) to account for the fact that these actions can only be performed when the player's team has the ball. This normalization ensures that players from teams with varying levels of possession are compared on a fair basis.

For out-of-possession statistics, we adjusted by multiplying the values by 30 and then dividing by the opponent possession time. This method, often referred to as Possession-adjusted (PAdj), accounts for the

fact that defensive actions can only occur when the player's team does not have the ball. Multiplying by 30 standardizes the statistics to a 50 percent possession scenario, providing a fair comparison of defensive contributions across players from different teams.

Finally, we normalized all statistics by dividing by a new column named `90s`, which is calculated as minutes-played divided by 90 to get the stats per game.

Possession-adjusted values provide insight into the frequency of actions, considering the actual time a team is in or out of possession. For instance, a player with fewer defensive actions in a possession-based team can have higher possession-adjusted values compared to a player in a team with less possession.

2.2.3 Reference for Possession-Adjustment

- Wyscout. Possession-Adjusted. Retrieved from `https://dataglossary.wyscout.com/p_adj/`

3 Ranking the Midfielders

3.1 Scores

The first step was about creating a score that from it would rank the players, an in-possession was thought of then out-of-possession score and finally an overall score.

3.1.1 In-Possession Scoring

Firstly, I separated the in-possession and out-of-possession features, organizing them into respective lists. The next step involved categorizing these features based on whether they contributed positively or negatively to a player's performance. For instance, metrics like "dangerousownhalflosses-per-90" were considered negative as they indicate potential weaknesses in a player's in-possession play.

The assignment of weights, whether positive or negative, to these features was a subjective process. However, it was guided by both intuition and logic, considering the importance of each metric in evaluating a player's overall performance. This methodology was honed during my internship, where I faced ambiguity but relied on instinct to make decisions while ensuring a logical foundation for the weighting system.

```
# Define your metrics and weights
in_possession_positive_metrics = {
    'retention': 0.3, 'shots_per_90': 0.3, 'np_goals_per_90': 0.5, 'shotsontarget_per_90': 0.3, 'np_xg_per_90': 0.5,
    'shotsblocked_per_90': 0.2, 'goals_per_90': 0.5, 'headshots_per_90': 0.2, 'np_shots_per_90': 0.2, 'assists_per_90': 0.5,
    'secondassists_per_90': 0.3, 'padj_accelerations_per_90': 0.2, 'padj_crosses_per_90': 0.2, 'padj_foulssuffered_per_90': 0.2,
    'padj_dribbles_per_90': 0.2, 'padj_forwardpasses_per_90': 0.3, 'padj_keypasses_per_90': 0.4, 'padj_lateralpasses_per_90': 0.2,
    'padj_longpasses_per_90': 0.2, 'padj_offensiveduels_per_90': 0.2, 'padj_offensiveduelswon_per_90': 0.2, 'padj_passes_per_90': 0.2,
    'padj_passestofinalthird_per_90': 0.3, 'padj_progressivepasses_per_90': 0.3, 'padj_progressiverun_per_90': 0.3,
    'padj_receivedpass_per_90': 0.2, 'padj_shotassists_per_90': 0.3, 'padj_shotontargetassists_per_90': 0.3, 'padj_smartpasses_per_90': 0.3,
    'padj_successfulattackingactions_per_90': 0.3, 'padj_successfulcrosses_per_90': 0.2, 'padj_successfuldribbles_per_90': 0.2,
    'padj_successfulkeypasses_per_90': 0.4, 'padj_successfullinkupplays_per_90': 0.2, 'padj_successfulpasses_per_90': 0.2,
    'padj_successfulpassestofinalthird_per_90': 0.3, 'padj_successfulprogressivepasses_per_90': 0.3, 'padj_successfulsmartpasses_per_90': 0.3,
    'padj_successfulthroughpasses_per_90': 0.3, 'padj_throughpasses_per_90': 0.3, 'padj_touches_inside_box_per_90': 0.3,
    'padj_xgassist_per_90': 0.3, 'padj_xgshot_per_90': 0.3, 'padj_buildup_prog_passes_per_90': 0.2, 'padj_buildup_direct_passes_per_90': 0.2,
    'padj_line_breaking_passes_per_90': 0.3, 'padj_transition_line_breaking_passes_per_90': 0.3, 'padj_line_breaking_passes_received_per_90': 0.2,
    'padj_transition_line_breaking_passes_received_per_90': 0.2, 'padj_switches_received_per_90': 0.2, 'padj_through_balls_received_per_90': 0.2,
    'padj_passes_and_touches_inside_box_per_90': 0.2, 'padj_chances_created_per_90': 0.3
}

in_possession_negative_metrics = {
    'padj_backpasses_per_90': -0.2, 'padj_dangerousownhalflosses_per_90': -0.3, 'padj_ownhalflosses_per_90': -0.5, 'padj_successfulbackpasses_per_90': -0.2
}
```

Figure 5: Positive and Negative Features' Weights

To ensure comparability across metrics, normalization was performed using the Min-Max scaling method. This step involved combining all relevant metrics and scaling them to a common range.

The calculation of the in-possession score was achieved through a weighted sum of the normalized metrics. Each metric was weighted according to its importance, as defined in the respective dictionaries of positive and negative metrics.

The resulting in-possession scores were then computed for each player.

Finally, the top players based on their in-possession scores were identified and displayed for further analysis.

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
  relevent_players_stats['in_possession_score'] += normalized_data.apply(calculate_weighted_score, axis=1, metrics_
```

	player_name	player_name_position	in_possession_score
3566	Leomar David Mosquera Niño	Leomar David Mosquera Niño_#8	9.377436
120	Adem Zorgane	Adem Zorgane_double pivot	6.989763
2345	Giorgi Moistsrapishvili	Giorgi Moistsrapishvili_double pivot	6.934698
1927	Erik Ahlstrand	Erik Ahlstrand_#8	6.813374
3543	Lazar Pavlović	Lazar Pavlović_double pivot	6.746343
6351	Łukasz Łakomy	Łukasz Łakomy_double pivot	6.580702
4901	Orkun Kökçü	Orkun Kökçü_single pivot	6.398901
4931	Othman Boussaid	Othman Boussaid_#8	6.213944
1939	Erkan Eyibil	Erkan Eyibil_double pivot	6.178716
4835	Oihan Sancet Tirapu	Oihan Sancet Tirapu_#8	6.166111

Figure 6: Top Players based on Possession Score

3.1.2 Out-of-Possession Scoring

For the out-of-possession scoring, a similar approach was undertaken. Initially, the relevant metrics and their respective weights were defined. Positive metrics such as "pct-dribbles-stopped" and "padj-defensiveduels-per-90" were considered indicative of strong defensive performance, while negative metrics like "redcards-per-90" and "yellowcards-per-90" reflected undesirable player behavior or disciplinary issues.

```
[ ] # Define your metrics and weights
    out_of_possession_positive_metrics = {
        'pct_dribbles_stopped': 0.3, 'aerials_success_rate': 0.3, 'padj_defensiveduels_per_90': 0.3, 'padj_interceptions_per_90': 0.5,
        'padj_pressingduels_per_90': 0.3, 'padj_clearances_per_90': 0.2, 'padj_counterpressingrecoveries_per_90': 0.3, 'padj_def_action_per_90': 0.3,
        'padj_blocked_shot_per_90': 0.2, 'padj_balls_won_per_90': 0.3, 'padj_tackles_and_interceptions_per_90': 0.5, 'padj_dangerousopponenthalfrecoveries_per_90': 0.5,
        'padj_defensiveduelswon_per_90': 0.5, 'padj_dribblesagainst_per_90': 0.2, 'padj_dribblesagainstwon_per_90': 0.4, 'padj_duels_per_90': 0.2,
        'padj_duelswon_per_90': 0.4, 'padj_opponenthalfrecoveries_per_90': 0.4, 'padj_recoveries_per_90': 0.3, 'padj_slidingtackles_per_90': 0.2,
        'padj_successfuldefensiveaction_per_90': 0.4, 'padj_def_action_first_third_per_90': 0.2, 'padj_def_action_middle_third_per_90': 0.3,
        'padj_def_action_final_third_per_90': 0.4, 'padj_tackle_per_90': 0.2
    }

    out_of_possession_negative_metrics = {
        'redcards_per_90': -0.5, 'directredcards_per_90': -0.5, 'yellowcards_per_90': -0.3, 'padj_fouls_per_90': -0.2
    }

    # Normalize the metrics
```

Figure 7: Positive and Negative Features' Weights for Out-of-Possession

Normalization of the metrics was conducted using the Min-Max scaling method to ensure comparability across different metrics. This involved combining all relevant metrics and scaling them to a common range.

To calculate the out-of-possession score, a weighted sum of the normalized metrics was performed. Each metric was assigned a weight based on its perceived importance in evaluating a player's out-of-possession performance, as specified in the respective dictionaries of positive and negative metrics.

The resulting out-of-possession scores were then computed for each player.

Finally, the top players based on their out-of-possession scores were identified and displayed for further analysis.

	player_name	player_name_position	out_of_possession_score
341	Alessandro Mercati	Alessandro Mercati_double pivot	5.218901
1465	David Mištrafović	David Mištrafović_single pivot	5.119354
4082	Markus Soomets	Markus Soomets_#8	5.026148
3942	Marcelencio Esajas	Marcelencio Esajas_#8	4.981153
3248	Justin Omoregie	Justin Omoregie_double pivot	4.862440
4645	Nicky Stephane Medja Beloko	Nicky Stephane Medja Beloko_single pivot	4.838861
6251	Yuto Tsunashima	Yuto Tsunashima_double pivot	4.798075
3335	Kevin Alexander Lopez Saldarriaga	Kevin Alexander Lopez Saldarriaga_single pivot	4.793958
2673	Itay Katzav	Itay Katzav_#8	4.761315
1995	Fabio Saiz Pennarossa	Fabio Saiz Pennarossa_single pivot	4.752179

Figure 8: Top Players based on Out-of-Possession Score

3.2 Ranking

The final ranking process involved:

- **Score Calculation:** Combined in-possession and out-of-possession scores to calculate an overall score for each player.
- **Percentile Thresholds:** Set filtering thresholds for percentile ranks, focusing on both in-possession and out-of-possession scores.
- **Comprehensive Analysis:** Examined top players' statistics across all positions they played to ensure versatility and adaptability.
- **Final Selection:** Sorted and identified the top players based on their overall percentile rankings and performance metrics.

3.2.1 Overall Scoring

To calculate the overall score, additional considerations beyond the in-possession and out-of-possession metrics were integrated. Specifically, the duration of a player's involvement on the field, represented by the number of 90-minute periods played ('90s'), and their age were factored in. Initially, both '90s' and 'age' were normalized to ensure uniformity in their impact on the overall score. However, to emphasize the significance of youthfulness in the scoring process, the normalized age values were inverted, giving higher scores to younger players. Weight coefficients were assigned to '90s' played and age, with both contributing equally to the final score. The overall score was then calculated as a combination of the in-possession and out-of-possession scores, along with the weighted contributions from '90s' played and age. This comprehensive approach provided a holistic evaluation of a player's performance, encompassing their effectiveness on the field, defensive capabilities, and broader demographic factors such as age and playing time.

	player_name	player_name_position	overall_score
3566	Leomar David Mosquera Niño	Leomar David Mosquera Niño_#8	6.175085
39	Abdul Razak Yusif	Abdul Razak Yusif_single pivot	5.221998
4159	Mateusz Kowalczyk	Mateusz Kowalczyk_single pivot	4.974610
2345	Giorgi Moistsrapishvili	Giorgi Moistsrapishvili_double pivot	4.969513
4070	Marko Šimun	Marko Šimun_#8	4.885051
120	Adem Zorgane	Adem Zorgane_double pivot	4.879553
6351	Łukasz Łakomy	Łukasz Łakomy_double pivot	4.831873
3543	Lazar Pavlović	Lazar Pavlović_double pivot	4.786168
4072	Markus Arsalo	Markus Arsalo_single pivot	4.777998
4566	Mylian Jiménez	Mylian Jiménez_single pivot	4.767817

Figure 9: Top Players based on Overall Score

3.2.2 Ranking

The ranking process was primarily centered around percentile ranks, ensuring a rigorous evaluation of each player's performance. However, to identify the top performers, a stringent threshold of at least 85 percent was set for both in-possession and out-of-possession scores. This meticulous approach aimed to select players who excelled in crucial aspects of the game. Furthermore, the search extended beyond mere individual prowess, targeting players who demonstrated versatility across all three positions. Despite the challenging criteria, this method ensured the selection of elite players who showcased exceptional skills and adaptability on the field.

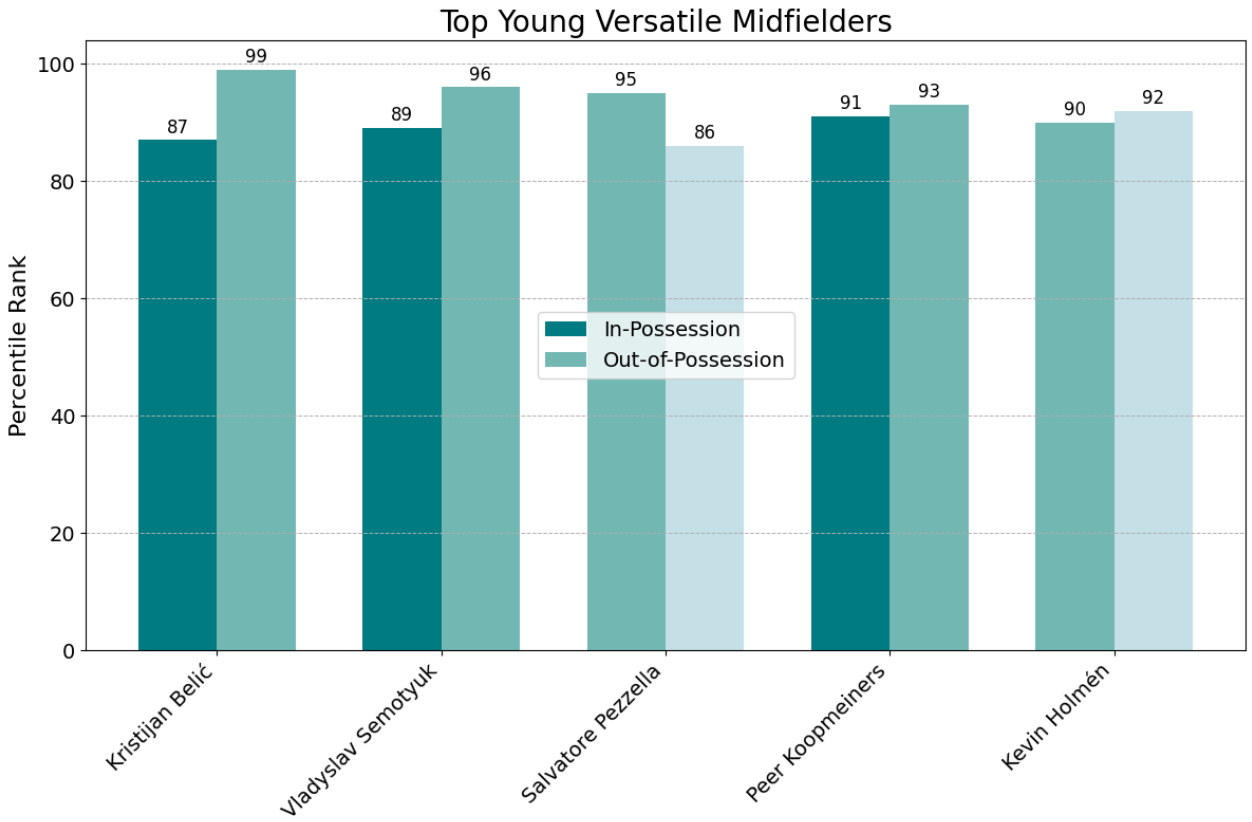

Figure 10: Top Young Versatile Midfielders

3.3 My Top Candidates Overview

In the final selection of top candidates, three players emerged as standout options for West Bromwich Albion's midfield reinforcement. Each selection was made following careful consideration and analysis of various factors. Notably, Kristijan Belić, despite his impressive performance metrics, was omitted from the shortlist due to his recent move to AZ Alkmaar, making his availability uncertain. However, this development opens the possibility of targeting Peer Koopmeiners, whose loan status and proven versatility could make him a viable option for the club.

Furthermore, Salvatore Pezzella's exclusion stemmed from a notable decrease in his transfer value, coupled with his involvement in Serie C, the third division of Italian football. These circumstances led to concerns regarding his current form and marketability, prompting the decision to focus on other prospects.

Consequently, Vladyslav Semotyuk and Kevin Holmén emerged as standout choices for their respective skill sets and potential contributions to the team. Semotyuk's promising potential in both plays, combined with Holmén's high possession ability and defensive prowess, make them valuable assets for West Bromwich Albion's midfield.

1. **Peer Koopmeiners:**
    - **Age:** 24
    - **Nationality:** Dutch
    - **Current Club:** Almere City FC (on loan from AZ Alkmaar)
    - **Positions:** Defensive Midfielder, Central Midfielder, Center-Back
    - **Height:** 1.86m
    - **Market Value:** €1.20 million
    - **Profile:** Known for high retention rates, effective passing, and strong defensive metrics. Potentially available due to AZ's acquisition of Kristijan Belić.
2. **Vladyslav Semotyuk:**
    - **Age:** 23
    - **Nationality:** Ukrainian
    - **Current Club:** Kryvbas Kryvyi Rih (on loan from FK Minaj)
    - **Positions:** Central Midfielder, Defensive Midfielder, Attacking Midfielder
    - **Height:** 1.77m
    - **Market Value:** €150,000 (rising)
    - **Profile:** Known for excellent ball control, distribution, and defensive positioning. Versatile and consistent performer with high growth potential.
3. **Kevin Holmén:**
    - **Age:** 22
    - **Nationality:** Swedish
    - **Current Club:** Degerfors IF (on loan from IF Elfsborg)
    - **Position:** Central Midfielder
    - **Height:** 1.78m
    - **Market Value:** €350,000 (rising)
    - **Profile:** Known for passing accuracy, vision, defensive duels, and interceptions. Consistent performer with potential for further development.

These players possess the versatility and skill set necessary to excel across different positions on the field.

3.4 Additional Analysis and Desired Features

3.4.1 xT

While xT (expected threat) is a powerful metric for assessing a player's on-ball value, its incorporation into the analysis was limited due to data availability constraints. However, future iterations of the methodology could benefit from integrating xT as a measure of a player's offensive contributions and goal-scoring threat.

3.4.2 Pressing Resistance Feature

The ability to resist pressing and maintain possession under pressure is a crucial attribute for modern midfielders. While pressing resistance was considered qualitatively in player evaluations, a quantitative metric for assessing this aspect was not available. Integrating a pressing resistance feature into the analysis could provide valuable insights into players' ability to thrive in high-pressure situations.

3.4.3 Carrying of the Ball

Carrying the ball effectively through midfield can disrupt opposition defenses and create goal-scoring opportunities. While this aspect was acknowledged qualitatively, a quantitative metric for evaluating players' carrying ability was not included in the analysis. Future iterations of the methodology could explore incorporating metrics such as successful dribbles or progressive carries to assess players' effectiveness in driving the ball forward.

3.4.4 Pressing Duel

The metric "pressing duel" was initially considered for inclusion in the analysis as it provides insights into a player's ability to engage in pressing situations and win possession back for their team. However, upon closer inspection, it was observed that this metric contained erroneous values that could potentially skew the evaluation of players' pressing capabilities.

While the pressing duel metric offers valuable insights into players' defensive contributions, its omission was deemed necessary to maintain the quality and trustworthiness of the analysis results. Future iterations of the methodology may explore alternative approaches for assessing players' pressing abilities or implement measures to address data quality issues associated with this metric.

3.4.5 Season-by-Season Improvement

Assessing players' improvement over multiple seasons is a valuable aspect of talent evaluation. While a direct comparison of overall scores by season may not provide a fair assessment due to varying external factors, such as changes in team tactics or surrounding player quality, leveraging TransferMarkt prices as a proxy for player development is an innovative approach. By tracking changes in market value over time, analysts can gain insights into players' progression and potential for future growth, providing a more nuanced understanding of their value to the team.

4 Conclusion

The methodology involved several key steps, starting with player selection and data preparation. This included filtering players based on age and position capability, cleaning and normalizing data, and ensuring a minimum threshold of playing time in relevant positions. The subsequent ranking and scoring phase involved developing robust scoring systems for both in-possession and out-of-possession performance, culminating in the calculation of overall scores for each player.

Comprehensive analysis was conducted to filter players based on percentile thresholds for in-possession and out-of-possession scores, ensuring the selection of top performers. The final selection process focused on identifying players who demonstrated versatility across multiple positions, thereby enhancing the team's adaptability on the field.

Among the top candidates, Peer Koopmeiners stood out for his strong defensive metrics and potential availability due to AZ Alkmaar's recent acquisition of Belic. Vladyslav Semotyuk showcased promising attributes in both of aspects, while Kevin Holmén exhibited high possession ability and defensive prowess, making them valuable assets for West Bromwich Albion.

In conclusion, these top candidates possess the versatility and skill set necessary to excel in the Championship and potentially transition to the Premier League in the future. With careful consideration of their performance metrics and strategic value, West Bromwich Albion can make informed decisions to strengthen their squad and achieve success on the field.