Performance and Value Analysis of the Premier League

Discovering Key Contributors and Emerging Talents

Ethan Homler

Table of Contents

Background and Objectives

In the last decade, the Premier League has emerged as the most dominant and competitive soccer league worldwide. In today's data-driven era, clubs rely on player statistics to make informed decisions. Advanced analytics and statistical modeling offer insights into player performance, value, and talent identification.

This project aims to enhance player scouting and valuation in the Premier League through advanced data analysis. Using a dataset containing statistics for hundreds of Premier League players, our goal is to identify key performance indicators for each position that influence market value. We also seek to uncover undervalued players and identify emerging young talents.

Our analysis combines performance metrics with market data, providing clubs with valuable insights for informed decisions in the transfer market and strengthening their competitive position. This report outlines our methodologies, presents data analysis and visualizations, and discusses key findings that can benefit Premier League clubs in player scouting and recruitment

About the Dataset

	Name	Club	Position	Age	Appearances	Wins	Losses	Goals	Goals per match	Penalties scored	...	Punches	High Claims	Catches	Sweeper clearances	Throw outs	Yellow cards	Red cards	Fouls	Offsides	Current Value
0	bernd leno	Arsenal	Goalkeeper	28.0	64	28	16	0	0.0	0.0	...	34.0	26.0	17.0	28.0	375.0	2	0	0	0.0	10000000.0
1	matt macey	Arsenal	Goalkeeper	26.0	0	0	0	0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0	0.0	0.0
2	rúnar alex rúnarsson	Arsenal	Goalkeeper	25.0	0	0	0	0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0	0	0	0.0	1500000.0
3	héctor bellerín	Arsenal	Defender	25.0	160	90	37	7	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	23	0	125	8.0	8000000.0
4	kieran tierney	Arsenal	Defender	23.0	16	7	5	1	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	2	0	9	0.0	25000000.0

The dataset contains 584 players with exactly 60 different statistical features.

Data Sources:

We obtained data from Kaggle, using two vital datasets:

Dataset 1: Detailed Premier League player statistics.
Dataset 2: Player names with their market values.

Data Integration:

We merged datasets using Excel's VLOOKUP function. This added a 'Current Value' column to the statistics dataset by matching player names.

Data Preparation:

After merging, we imported data into Python, ensuring:

- Correct data types for stats.
- Removal of missing values (NAs) for data integrity.
- Elimination of redundant features.

Dataset Exploration

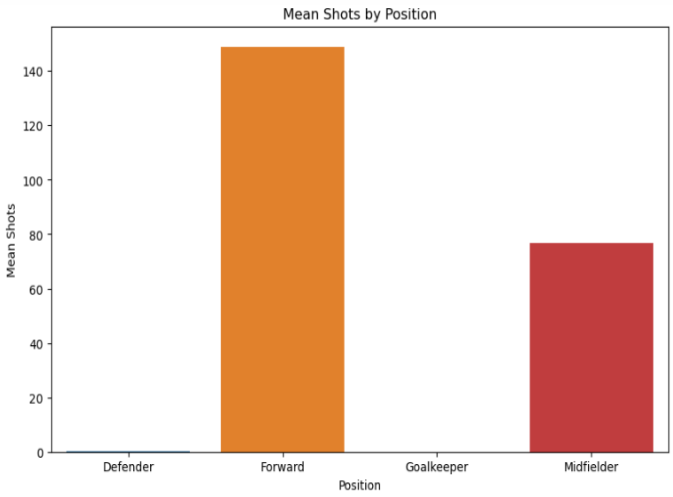
In this section, we explore how we explored and analyzed the Premier League dataset. To analyze the dataset, several python tools were implemented

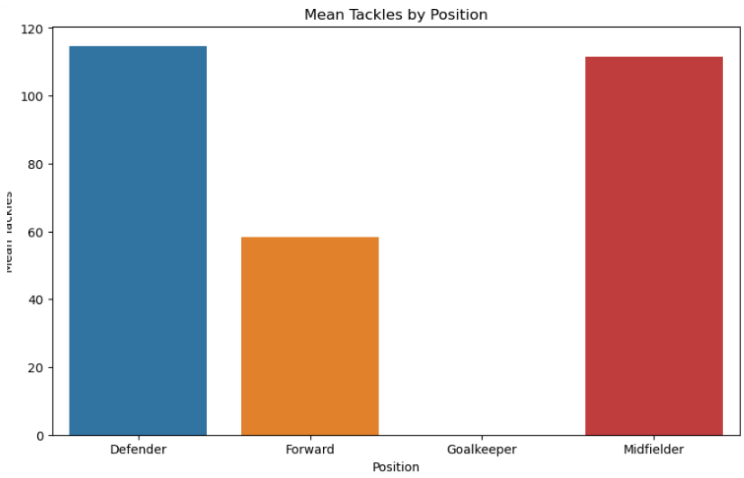
- Pandas, Seaborn, MatPlotLib, MinMaxScaler, Numpy, and RandomForestRegressor

Initial Observations:

Preliminary data exploration revealed key insights:

- Player statistics varied by position.
- Forwards excelled in goal-related metrics, midfielders in assists and passes, defenders in tackles and clearances, and goalkeepers in saves.

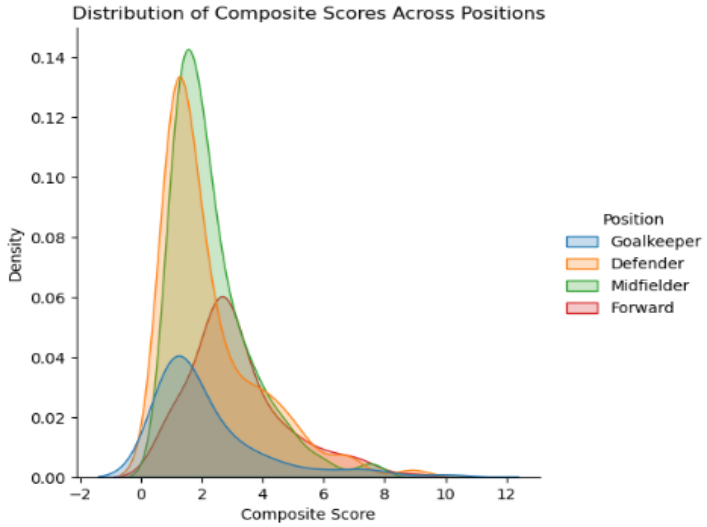




These position-specific statistics showed players can not be evaluated the same to the same statistics of players in other positions. Key Performance Indicators (KPIs) need to be identified.

Player Performance Assessment

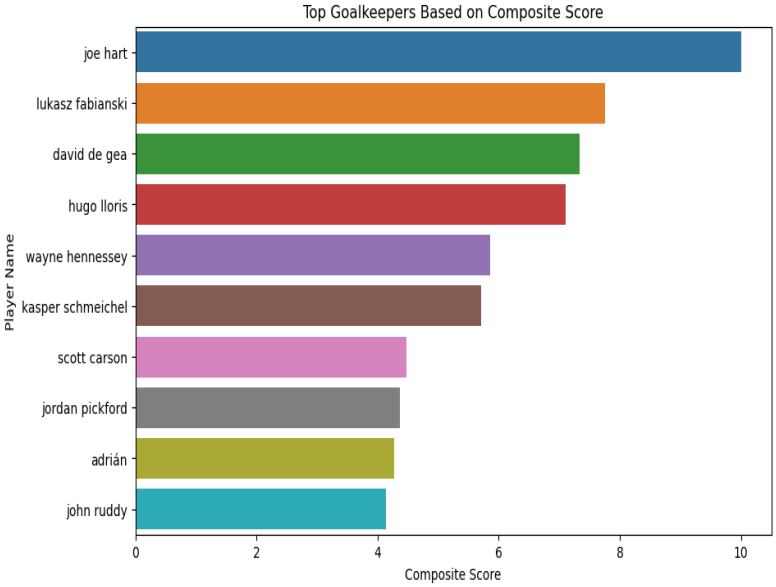
Evaluation of Premier League Player Performances based on the player's position-specific statistics. Key performance indicators (KPIs) specific to each position - such as goals and assists for forwards, tackles and interceptions for defenders - were standardized to create a uniform metric, the composite score. This assessment forms a crucial part of the talent scouting process, offering a solid base score of how to properly rate players in the league. Down below shows the distribution of scores across all positions.

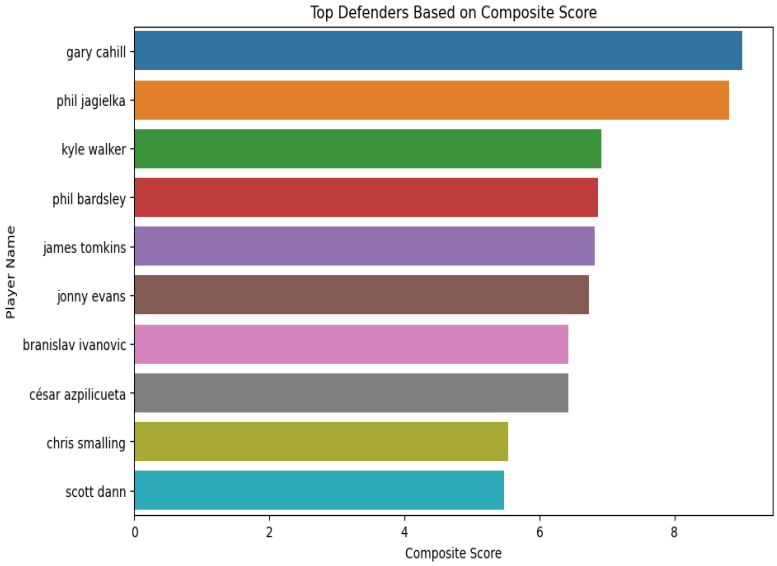


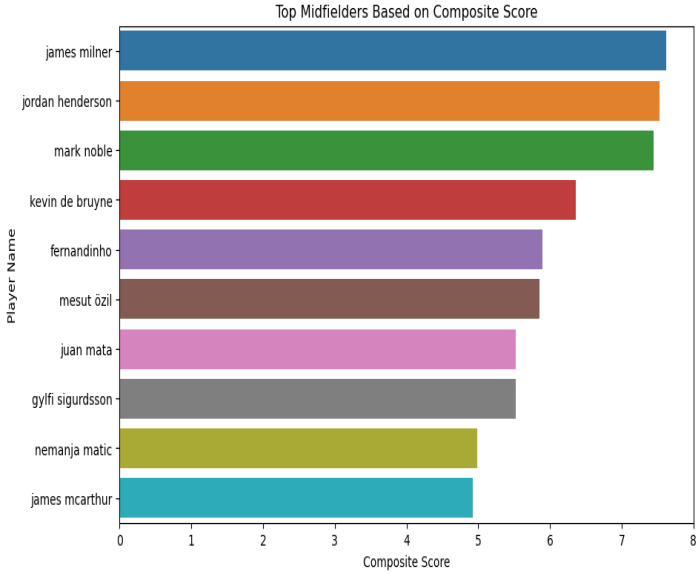
Based on the results

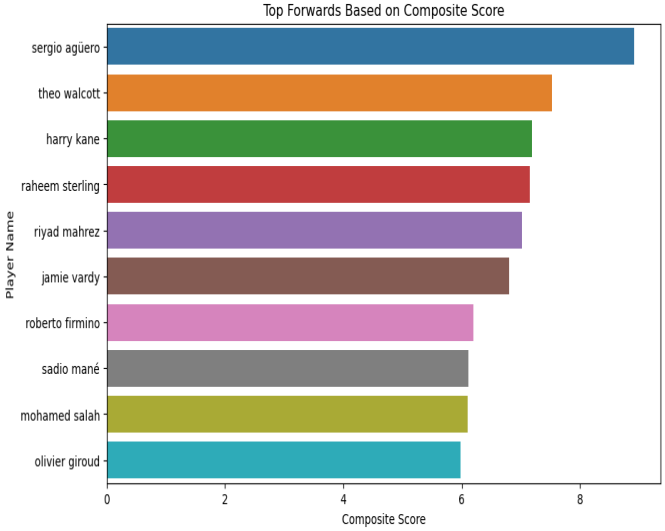
- Goalkeepers - lower composite scores based on the selected KPIs. This could be due to the nature of goalkeeper statistics
- Defenders - distribution for defenders has a wider base, suggesting more variability in the composite scores
- Midfielders - suggests a higher average composite score, possibly reflecting their involvement in various aspects of the game (defense, transition, and attack)
- Forwards - indicating higher composite scores on average. Consistent with the expectation that forwards typically contribute to more direct scoring opportunities, reflected in higher scores for KPIs like goals, shots on target

Using the KPIs and the composite score, identification of the top 10 players in each position based on their composite score was found…

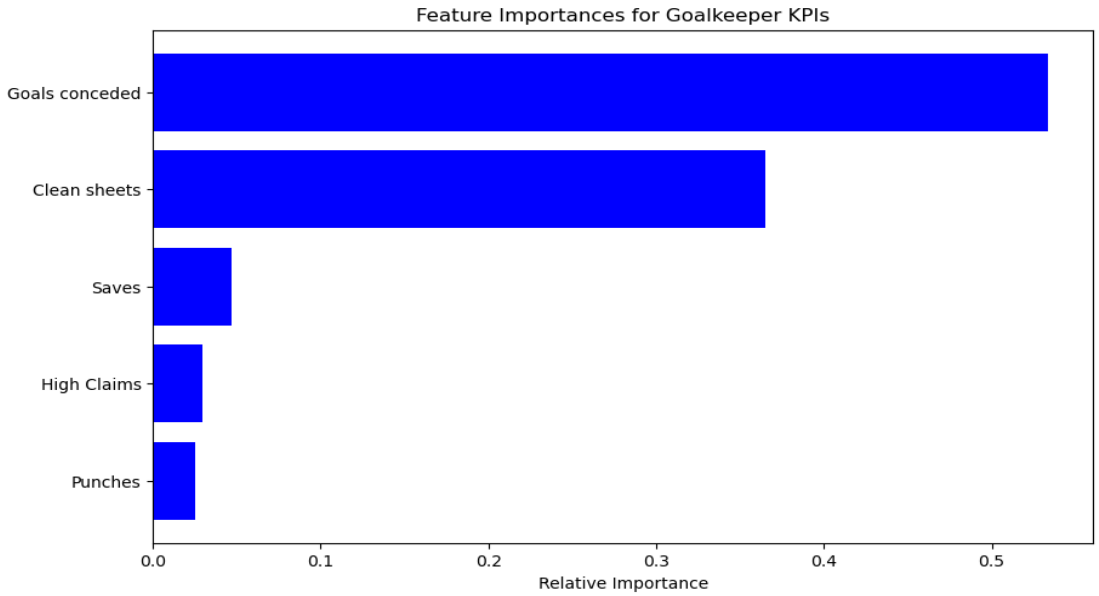


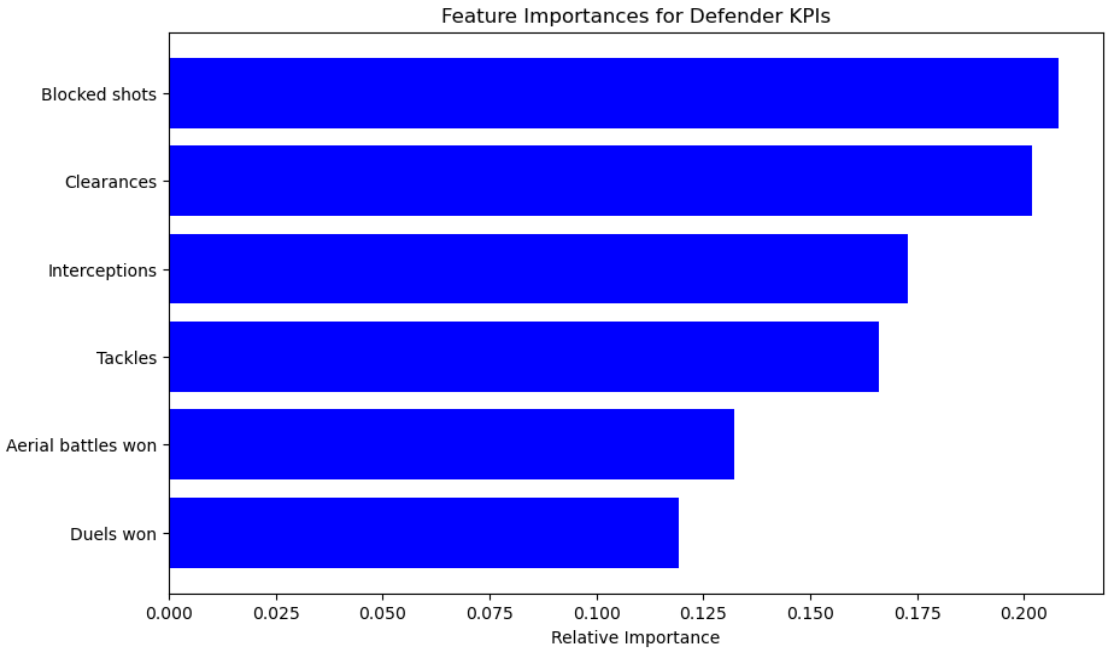


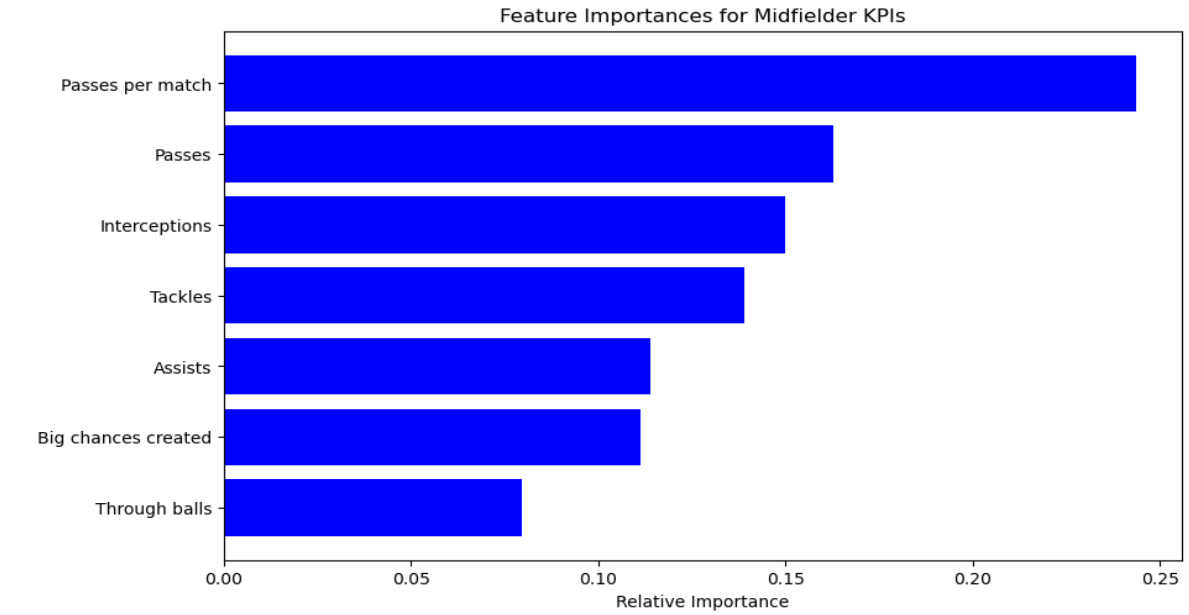


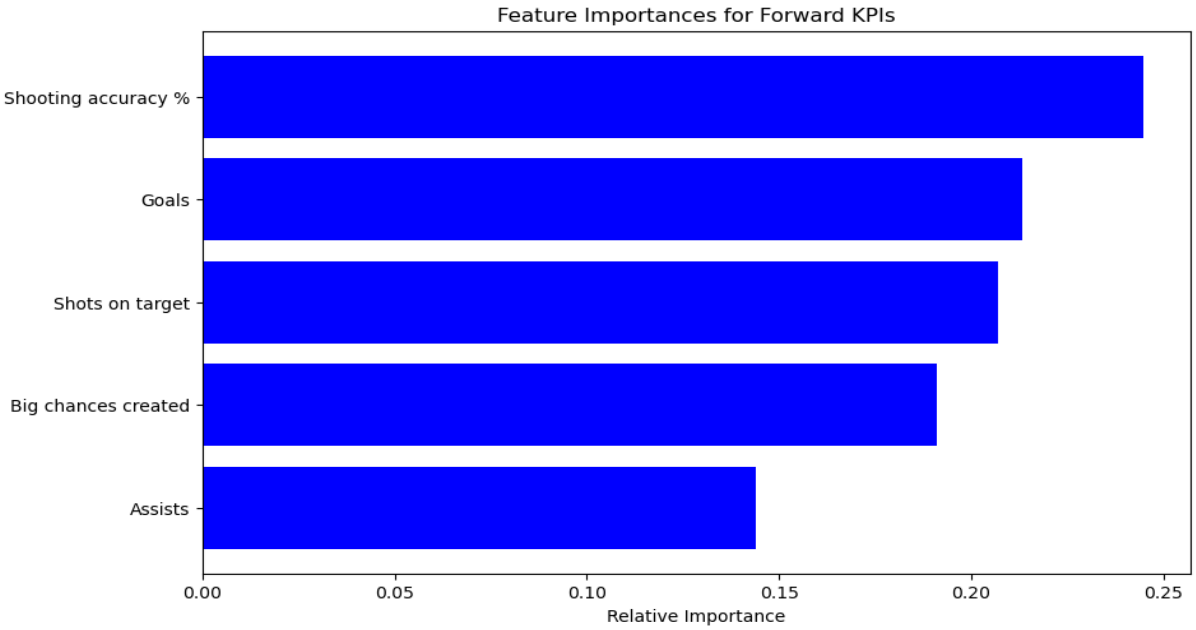


Feature importance gives us insight into which features (KPIs) per position. Using the player's current market value as the target variable, these plots show which KPIs are most important for the player's value per position.









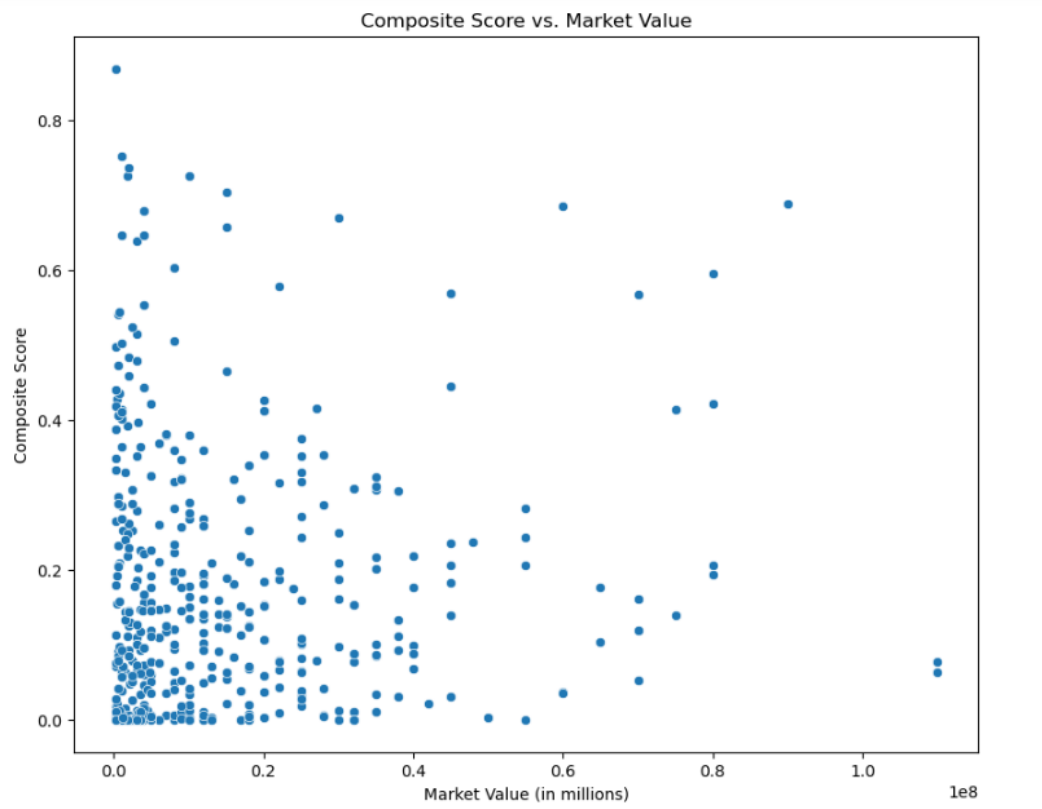
- Based on the results of plots, the most valued features tend to be those directly linked to the core responsibilities of the position
    - (e.g., preventing goals for goalkeepers, winning duels for defenders, creating chances for midfielders, and shooting accuracy for forwards).

This analysis can inform clubs and managers about which aspects of player performance might be undervalued or overvalued in the transfer market

Value Assessment

Balancing player performance with their market value is paramount. Identifying players who offer exceptional on-field contributions relative to their market price is a strategic advantage. Conversely, clubs must also address the issue of overvalued players who do not deliver expected returns.

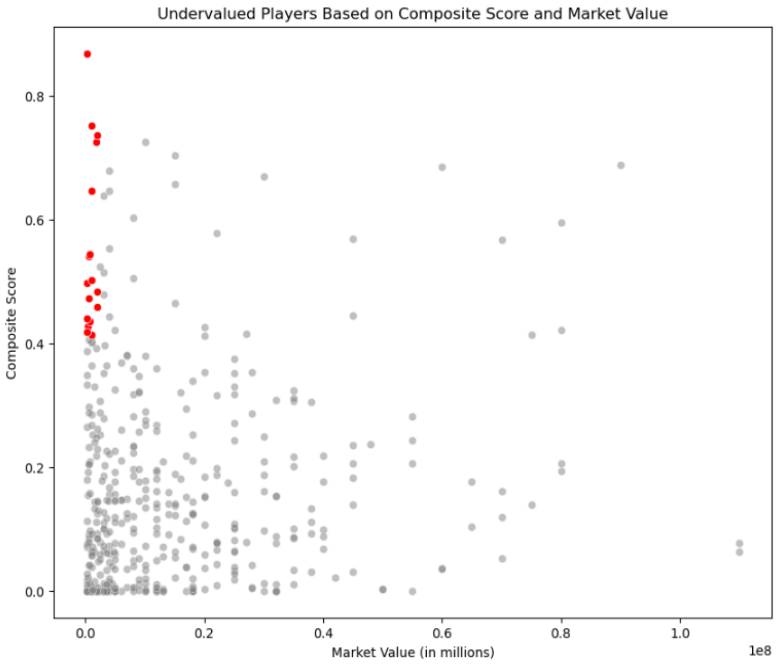
To dive into the assessment, I first wanted to see the relationship between Composite Scores and the current market value of the player.



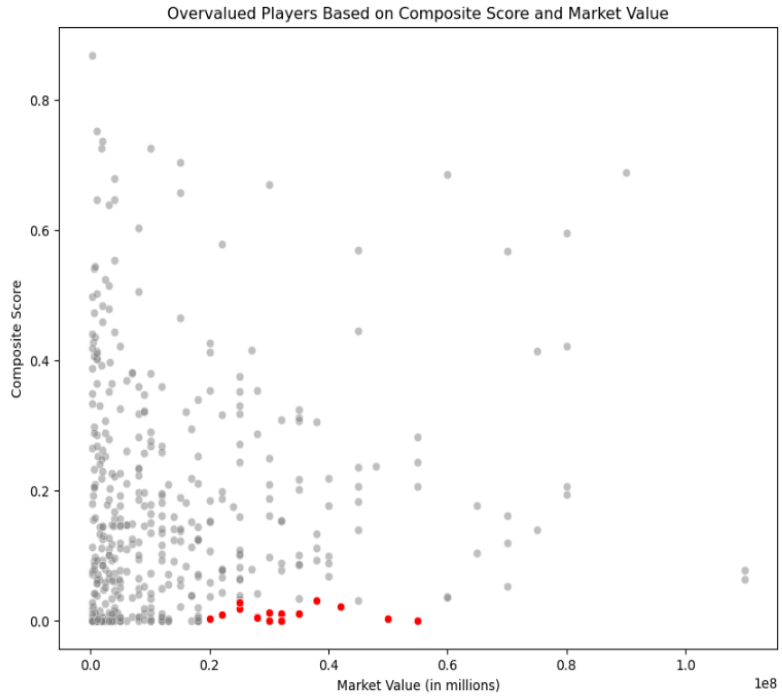
- Not many high value players (understandable)
- Wide distribution of composite scores at the lower end of the market value spectrum, which suggests that there are many players with varying performance levels who are valued relatively low in the market

Undervalued/Overvalued Player Threshold:

- Created an undervalued player threshold which are the top 10% of performers in the bottom 25% in value; vice versa for overvalued players
- Compiled a list of players who fall in each group
- Created a scatter plot. Data points are gray and points that fit within the threshold are red



Name	Position
david luiz	Defender
erik pieters	Defender
ashley westwood	Midfielder
wayne hennessey	Goalkeeper
joel ward	Defender
james tomkins	Defender
scott dann	Defender
james mcarthur	Midfielder
theo walcott	Forward
james milner	Midfielder
fernandinho	Midfielder
phil jones	Defender
juan mata	Midfielder
andy carroll	Forward
phil jagielka	Defender
shane long	Forward
lukasz fabianski	Goalkeeper

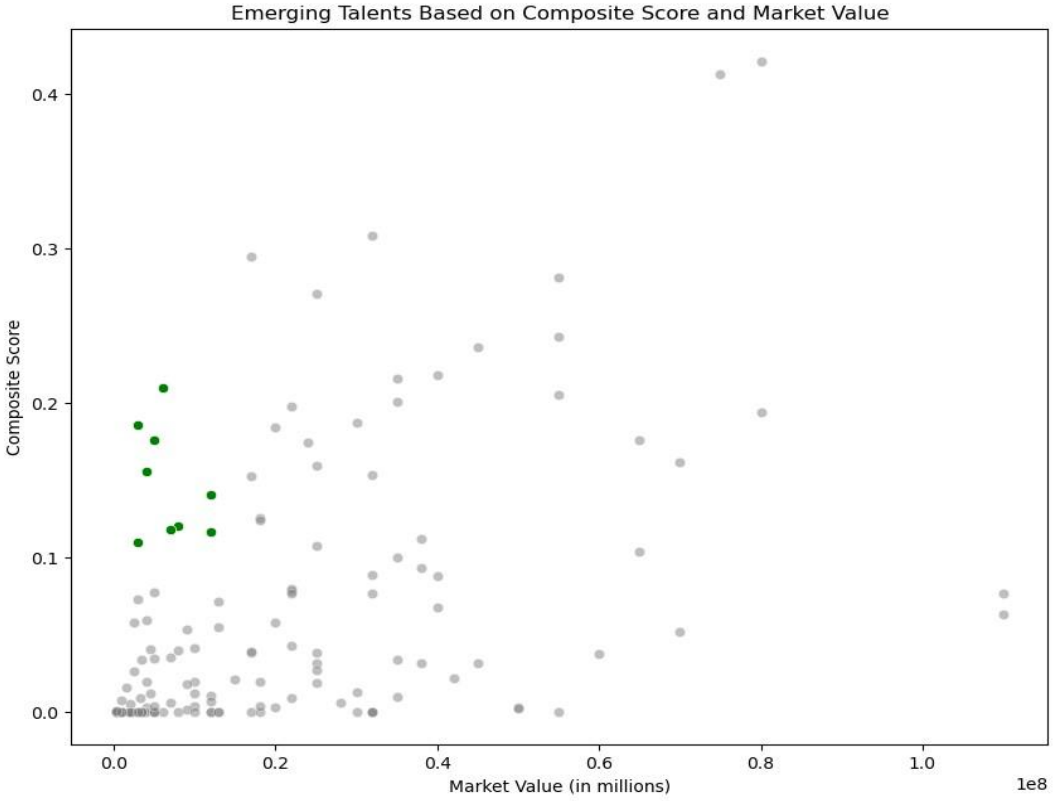


Name	Position
kieran tierney	Defender
william saliba	Defender
gabriel magalhães	Defender
emile smith rowe	Midfielder
jacob ramsey	Midfielder
ollie watkins	Forward
robert sánchez	Goalkeeper
ben white	Defender
alexis mac allister	Midfielder
tyrick mitchell	Defender
eberechi eze	Midfielder
pascal struijk	Defender
timothy castagne	Defender
harvey elliott	Midfielder
ferran torres	Forward
juan foyth	Defender
conor gallagher	Midfielder
vitinha	Midfielder

Emerging Talent Identification

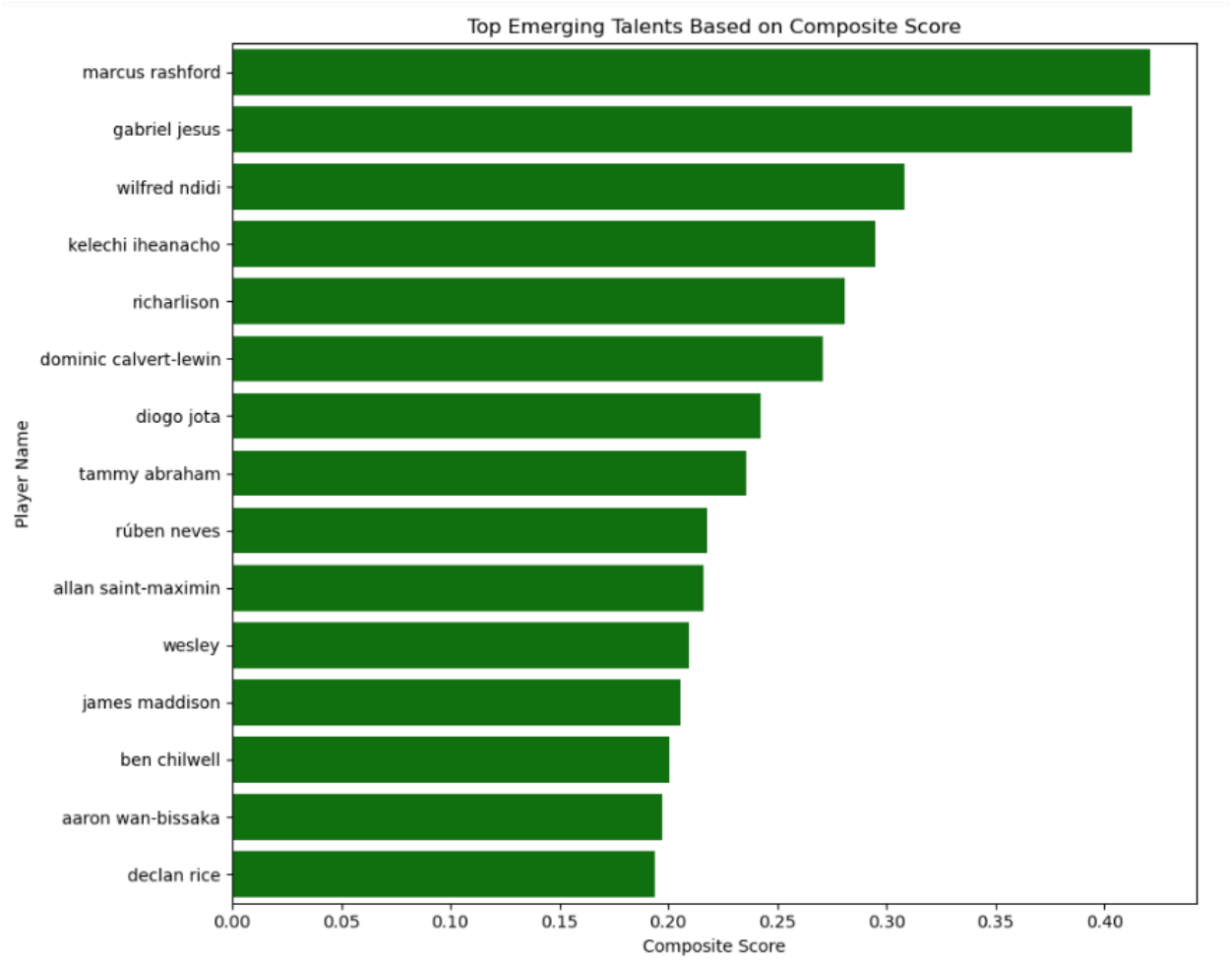
Discovering and nurturing emerging talents is crucial for long-term success. Premier League clubs are in a continuous competition to find and sign promising young players who can become future stars. Identifying these talents solely based on their statistical performance data is a critical aspect of recruitment strategy. Historical statistics/awards like the PFA Young Player of the Year show that the threshold age for the winner is 23 years of age or younger. A metric was created to find composite scores for players who fall in this age threshold.

A scatter plot was used to depict the relationship between the composite scores and market values of young Premier League players who fell in the age threshold. The gray dots represent all young players under the age of 23, providing a landscape of the current young talent in terms of performance and market value. The green dots highlight those identified as emerging talents, showcasing players with high composite scores but relatively lower market values. Next to the plot shows the names of players of the emerging talents with low values.



Name	Age	Current Value
ainsley maitland-niles	23.0	8000000.0
reiss nelson	20.0	7000000.0
wesley	23.0	6000000.0
keinan davis	22.0	3000000.0
aaron connolly	20.0	4000000.0
mason holgate	23.0	12000000.0
tom davies	22.0	12000000.0
tyler roberts	21.0	3000000.0
michael obafemi	20.0	5000000.0

After creating the scatter plot, I created a barchart that represented the top 10 emerging talents in the Premier League regardless of their current market value.



Conclusion and Recommendations

This analysis has visualized the dynamic relationship between player performances in the Premier League and their own current market values with unique attention on discovering emerging young talent within the league. This data-driven approach to analyzing player statistics and understanding player value is a tool more clubs should be using more in their squad optimization decisions.

From this analysis we saw that composite scores effectively helps analyze player performances based on their positions by introducing key performance indicators to get a true picture of how well or not a player is performing. The composite scores also help find who performances are being underappreciated in the market.

In terms of the value assessment of the players in the league, the analysis shows that some players' performance is better than what their market value might suggest. This finding offers clubs the chance to sign good players at lower prices, providing value on the field without overspending.

As a recommendation for further analysis, I believe there’s more data that could be incorporated such as player injury history, consistency of performance over multiple seasons, and psychological assessments to provide a more comprehensive player profile. Also besides just adding positive attributes to the KPI composite scores, statistics harming player performance could also be added to get a true valuation of a player like red cards for defenders or offsides count for forwards.

References

Premier League Player Statistics

https://www.kaggle.com/datasets/rishikeshkanabar/premier-league-player-statistics-updated-daily

Premier League Player Market Value

https://www.kaggle.com/datasets/khanghunhnguyntrng/football-players-transfer-fee-prediction-dataset

Emerging Talent Age Threshold

https://theathletic.com/3316248/2022/05/18/premier-league-young-player-of-the-year-award/