Logo

Description automatically generated

**“Can machine learning be effectively employed to analyse and predict transfer fees in English football?”**

Cameron Larkin

C19403332

Submitted in partial fulfilment of the requirements

for the degree of:

BSc in Business Analytics

Technological University Dublin

Supervisor: Qianru Shang

Date submitted: 29/04/2024

Word Count: 15,102 Words

**Declaration**

This dissertation is submitted by the undersigned to the Technological University Dublin in partial fulfilment for the degree Business Analytics. It is entirely the author’s own work and has not been submitted previously for an award to this or any other institution. All sources consulted are appropriately referenced as per the TUD School of Business Style Guide.

Signed: Cameron Larkin

Date: 29/04/2024

**Acknowledgements**

I would like to express my sincere gratitude to my academic supervisor, Qianru Shang for sharing her knowledge, insights, and continued assistance throughout this academic year. To my classmates, for their encouragement and support throughout the last four years. Finally, to my parents, for their continuous support and for all the opportunities they have provided me with.

**Abstract**

The landscape of English football has witnessed a transformative shift in recent years, marked by astronomical transfer fees and an evolving player recruitment environment. This dissertation navigates the complex world of football economics, aiming to unravel the enigmatic realm of player recruitment strategies and the underpinning forces behind the staggering transfer fees prevalent in the sport. The primary research question centres on exploring the driving forces behind these exorbitant fees and understanding the motivations that drive certain football clubs to invest substantial sums in player acquisitions.

Employing a multifaceted research approach, this study integrates rigorous data analysis and potentially transformative machine learning techniques. Datasets comprising player performance metrics, historical transfer data, and club financials form the foundation for a systematic exploration of the intricate links between player attributes, club strategies, and market dynamics, and the transfer fees they yield. Through statistical analysis and the application of machine learning models, this research seeks to predict and model transfer fees, offering insights into the complexities of player valuation and recruitment decisions.

The dissertation not only explores the economic impacts on football clubs but also delves into the perspectives of stakeholders involved in player acquisitions, including players, selling clubs, buying clubs, and sponsors. By integrating both qualitative and quantitative analyses, this research aims to offer a comprehensive understanding of player recruitment strategies in English football and predict trends that could shape the industry's future. The findings hold implications not only for football clubs and investors but also for the broader understanding of sports economics and the global sports industry.

In charting the historical trajectory of football economics and integrating insights from existing literature, this dissertation aims to contribute new knowledge and shed light on the dynamics of player recruitment in English football. The study endeavours to offer valuable insights, inform strategic decisions within the football industry, and pave the way for future research in this ever-evolving landscape.

Table of Contents

[Chapter 1 - Introduction 5](#_Toc165247049)

[1.1. Introduction 5](#_Toc165247050)

[1.2. Research Question: 7](#_Toc165247051)

[1.3. Research Objectives: 7](#_Toc165247052)

[1.4. Analysis framework: 8](#_Toc165247053)

[1.5. Structure of dissertation: 8](#_Toc165247054)

[Chapter 2 - Literature review 8](#_Toc165247055)

[2.1. Introduction: 8](#_Toc165247056)

[2.2. Techniques used to valuate players: 9](#_Toc165247057)

[2.2.1. Performance Metrics and Statistical Analysis: 9](#_Toc165247058)

[2.2.2. Scouting and Talent Identification: 9](#_Toc165247059)

[2.2.3. Brand Value and Commercial Impact: 11](#_Toc165247060)

[2.2.7. Club Strategies and Fit Within the Team: 12](#_Toc165247061)

[2.4. How Premier League clubs use statistics when making transfer decisions: 13](#_Toc165247062)

[Pre-modern statistical analysis: 14](#_Toc165247063)

[2.5. Evolution of Financial Dynamics in Player Transfers 15](#_Toc165247064)

[2.8. Ethical Considerations and Player Welfare 16](#_Toc165247065)

[2.10. Conclusion 17](#_Toc165247066)

[2.11. Summary of literature Review: 17](#_Toc165247067)

[Chapter 3 - Research Methodology: 18](#_Toc165247068)

[Statistical Models and Machine Learning Techniques 20](#_Toc165247069)

[Regression Analysis 20](#_Toc165247070)

[Decision Trees and Random Forests 20](#_Toc165247071)

[Clustering 21](#_Toc165247072)

[Implementation and Evaluation: 21](#_Toc165247073)

[Model Validation and Evaluation: 21](#_Toc165247074)

[Chapter 4: Implementation – 22](#_Toc165247075)

[Exploratory Data Analysis: 22](#_Toc165247076)

[Goals vs. Market Value Scatter Plot Analysis: 25](#_Toc165247077)

[Assists vs. Market Value Scatter Plot Analysis: 25](#_Toc165247078)

[Expected Goals (xG) vs. Market Value Scatter Plot Analysis: 25](#_Toc165247079)

[Overall Analysis: 26](#_Toc165247080)

[Distribution and Density: 28](#_Toc165247081)

[Positional Market Value Trends: 30](#_Toc165247082)

[Interpretation of Correlation Coefficients: 32](#_Toc165247083)

[Distribution Analysis: 35](#_Toc165247084)

[Performance Models: 36](#_Toc165247085)

[K-means Clustering Analysis: 36](#_Toc165247086)

[Hierarchical Clustering Analysis: 37](#_Toc165247087)

[Feature Importance Analysis: 39](#_Toc165247088)

[Model Training: 39](#_Toc165247089)

[Model Testing 40](#_Toc165247090)

[Chapter 5 – Analysis Results and Findings 40](#_Toc165247091)

[Chapter 6 – Discussions 42](#_Toc165247092)

[Review of Research Objectives: 43](#_Toc165247093)

[Limitations: 44](#_Toc165247094)

[Chapter 7 - Conclusion: 45](#_Toc165247095)

[Appendix: 46](#_Toc165247096)

[Appendix A - Players.xlsx Appendix Table 46](#_Toc165247097)

[Appendix B - Transfer\_Values.csv Appendix Table 47](#_Toc165247098)

[Bibliography: 47](#_Toc165247099)

[Appendix C – Link to Github Repository 48](#_Toc165247100)

**List of Figures**

[Figure 1: Showing the top 7 shirt sales of premier clubs in terms of revenue. 12](#_Toc165247101)

[Figure 2 - Distribution of Player Market Values in the English Premier League 22](#_Toc165247102)

[Figure 3 – Scatter plots showing goals, assists and xG against Market Value. 25](#_Toc165247103)

[Figure 4 – A bar chart showing Average market value per position. 26](#_Toc165247104)

[Figure 5 - Analysis of how a player's age might correlate with their perceived economic worth in the transfer market. 28](#_Toc165247105)

[Figure 6 – A box plot to show distribution of market values per player position. 30](#_Toc165247106)

[Figure 7 – Correlation Matrix 32](#_Toc165247107)

[Figure 8 – Scatterplot Matrix 34](#_Toc165247108)

[Figure 9 – Table of regression models 36](#_Toc165247109)

[Figure 10 – Table of K-Means cluster profiles 37](#_Toc165247110)

[Figure 11 – Hierarchical Clustering Analysis Diagram 38](#_Toc165247111)

[Figure 12 – Top Feature Importance Analysis 39](#_Toc165247112)

[Figure 13 - Results 39](#_Toc165247113)

# Chapter 1 - Introduction

## 1.1. Introduction

The world of football, a dynamic and constantly evolving industry, has witnessed a dramatic transformation in recent years. The transfer market has emerged as a cornerstone of this evolution, characterized by staggering sums of money changing hands in exchange for the world's top talents. With astronomical transfer fees becoming the norm, the spectacle of a "footballing arms race" has enthralled fans, while simultaneously raising questions about the underlying motivations and economic drivers that fuel these transactions.

This dissertation embarks on a journey through the intricate world of English football, seeking to dissect the enigmatic realm of player recruitment strategies and the rationale behind the remarkable transfer fees. Amid the spectacle of high-stakes transfers, the questions that loom large are: What are the driving forces behind these exorbitant transfer fees? Why are certain football clubs willing to invest vast sums in acquiring players? How do players, selling clubs, buying clubs, sponsors, and the broader market dynamics influence these monumental decisions?

At the heart of this investigation lies the fusion of both data analysis and potentially transformative machine learning techniques. The available datasets provides the foundation for a systematic exploration of player performance metrics, historical transfer data, club financials, and market trends. By applying rigorous statistical analysis, this research aims to uncover the intricate links between these diverse factors and the transfer fees they yield.

Machine learning algorithms, ranging from regression models to clustering analyses, will serve as predictive tools to model transfer fees based on a myriad of player attributes, club strategies, and prevailing market conditions. Additionally, qualitative analysis, including the examination of players' contract lengths, will shed light on the multifaceted dimensions of the decision-making processes that drive each club's recruitment choices. In this way, this research aspires to offer a comprehensive understanding of the factors that underpin player recruitment strategies and the economic dynamics of the European football industry.

This exploration will delve deep into the heart of football economics, enabling the comparison of historical trends and the prediction of future developments within the industry. By unravelling the intricacies of player recruitment, this research endeavours to provide valuable insights for football clubs, investors, and enthusiasts alike, while contributing to the broader understanding of the global sports industry.

In order to see the significance of the subject at hand, it is essential to acknowledge the historical trajectory of football economics. English football specifically, was once characterized by modest transfer fees and localized recruitment strategies, has evolved into a global, multi-billion-euro industry. This transformation has ushered in an era of financial competitiveness and an appetite for success, propelling the sport into a new dimension. As transfer fees continue to reach unprecedented heights, a spotlight is cast upon the underlying dynamics of player recruitment. Therefore, this research does not merely investigate the present state of the football economy, but it also embarks on a historical journey to chart the factors and trends that have led to the current state of affairs in the European football transfer market.

With the guidance of my dissertation supervisor, and equipped with the skills and resources, this project is poised to navigate the terrain of English football economics, uncovering the trends of player recruitment strategies and the drivers behind the vast transfer fees that capture the world's attention year after year.

## 1.2. Research Question:

This research will be based on looking at markers in the datasets that have been picked to see why players have moved clubs and to see if there are certain reasons as to why clubs have sought the need to pay the price they moved for. This has led to the question:

*“Can machine learning be effectively employed to analyse and predict transfer fees in English football?”*

## Research Objectives:

Analyse the Determinants of Transfer Fees: Investigate the factors that drive the transfer fees in English football, with a focus on player attributes, club strategies and market dynamics. This objective seeks to identify the primary determinants of the escalating transfer fees.

1. Evaluate the Economic Impact on Football Clubs: Assess the financial implications and performance outcomes for football clubs associated with high-value player acquisitions. This objective aims to understand how player recruitment strategies affect teams after joining.
2. Examine Stakeholder Perspectives: Explore the perspectives and motivations of various stakeholders in player recruitment, including players, selling clubs, and buying clubs. This objective seeks to uncover the diverse interests and influences that shape recruitment decisions.
3. Apply Data Analysis and Machine Learning Techniques: Utilize advanced data analysis and machine learning methodologies, such as regression models and clustering analyses, to model and predict transfer fees based on a wide range of factors. This objective aims to create data-driven tools for transfer fee estimation.
4. Offer Insights and Predictive Value: Provide an understanding of player recruitment strategies in European football by combining qualitative and quantitative research methods. Additionally, aim to develop predictive insights that can benefit football clubs and fans to give an insight as to why player have moved for their price.

## 1.4. Analysis framework:

This dissertation follows a analysis framework. It follows a research, build, and implementation structure which was designed to group each phase of this research, allowing for the transition between each stage.

## 1.5. Structure of dissertation:

This dissertation is divided into chapters. Chapter 1 introduces the dissertation. Chapter 2 contains the literature review which delves into the world of football transfers. Chapter 3 is the research methodology on this topic. Chapter 4 is the implementation stage of the dissertation. Chapter 5 is the analysis and findings on this research project. Chapter 6 presents the results of the findings. Chapter 7 displays the discussions which includes both the limitations and the potential work in the future that can be done. Chapter 8 ends the dissertation with the conclusion.

# Chapter 2 - Literature review

## 2.1. Introduction:

The pursuit of understanding the intricacies of player recruitment strategies in English football necessitates a thorough examination of existing research, theories, and empirical findings within the domain of football economics and sports management. This literature review serves as a comprehensive exploration of the existing body of knowledge that underpins our investigation into the determinants of transfer fees, the economic implications for football clubs, and the multifaceted influences on player recruitment decisions.

English football, with its rich history and global popularity, has attracted considerable scholarly attention over the years. Researchers have delved into the multifaceted aspects of the sport, from the analysis of match outcomes and player performance to the exploration of club finances and the complexities of the transfer market. As the sport has undergone a paradigm shift, transitioning from localized competitions to a global phenomenon characterized by the financial prowess of top-tier clubs, academic inquiry has mirrored this evolution. Researchers have shifted their focus towards uncovering the driving forces behind the monumental transfer fees and the profound implications these transactions hold for football clubs, players, fans, and the broader sports industry.

In this literature review, we embark on a journey through the annals of football economics, situating our research within the context of previous investigations. We survey the landscape to provide a comprehensive understanding of the key themes, theories, and empirical findings that have shaped the field. The review is structured around three fundamental pillars: the determinants of transfer fees, the economic impact on football clubs, and the diverse influences on player recruitment decisions. By assessing the existing literature, we seek to identify gaps and opportunities for our research to contribute to this evolving field and offer new insights into the realm of player recruitment in English football.

## 2.2. Techniques used to valuate players:

### 2.2.1. Performance Metrics and Statistical Analysis:

The evaluation of players in the football market heavily relies on statistical analysis and performance metrics. Objective measurements such as goals scored, assists, pass accuracy, tackles, interceptions, and other key performance indicators are fundamental in gauging a player's on-field contributions. Advanced statistical models and performance analytics allow clubs to not only evaluate individual players but also compare their performances against peers in similar positions or leagues. These metrics provide empirical evidence that aids in decision-making processes during player recruitment. The article on AnalyiSport (2022) discusses the evolving use of data analytics in the Premier League. It highlights varying degrees of adaptation to data-driven strategies among clubs, from Manchester United's rumoured expansion of their analytics team to Arsenal's use of their in-house data company, StatDNA. The article also touches on the challenges and scepticism surrounding the use of precise data metrics in analysing game outcomes, as exemplified by Mikel Arteta's defence of Arsenal's performance using data from StatDNA. The overall narrative illustrates the growing prominence of data analysis in the Premier League, marking a shift from traditional methods to more sophisticated, data-driven approaches.

In the valuation of football players, comparative analysis serves as a fundamental cornerstone. Football clubs and stakeholders lean on this method by referencing past transfer dealings involving players who share akin attributes, positions, and performance levels. This benchmarking technique serves as a bedrock for establishing a player's market value within the current competitive landscape. By meticulously scrutinizing recent transactions involving players possessing similar traits, clubs and stakeholders obtain crucial insights to gauge an appropriate valuation for a player in question. This comparative methodology plays a pivotal role, especially during negotiations for transfer fees or contract renewals. By thoroughly examining players who exhibit comparable characteristics and performance levels, clubs form a basis to determine the valuation for a particular player within the existing market trends. The depth of this comparative analysis allows clubs to position their valuation strategy within the context of recent market activities, ensuring a balanced and informed approach towards negotiations and valuing players within the ever-evolving football market.

### 2.2.2. Scouting and Talent Identification:

In the realm of player valuation within the football market, scouting and talent identification emerge as pivotal determinants, especially when considering emerging or lesser-known talents. Football clubs heavily invest in extensive scouting networks that aim to unearth and identify promising players. A noteworthy example is Brighton & Hove Albion's successful recruitment strategy, which emphasizes calculated risks in player signings and has led to significant successes, such as the acquisition of players like Moises Caicedo and Kaoru Mitoma (Analytics FC, 2023). These examples illustrate the vital role of scouting in identifying players with high potential, aligning with a club's tactical and strategic objectives. The collaboration between quantitative data and qualitative expertise of scouts plays an integral role in shaping a player's valuation within the competitive football market.

Through a meticulous process involving video analysis, in-person observations, and comprehensive assessments, scouts acquire a holistic understanding of a player's abilities, potential, and adaptability to a club's playing style. The subjective evaluations from these seasoned scouts supplement quantitative data, enriching the overall understanding of a player's capabilities. These subjective assessments aid in offering invaluable insights into a player's potential and suitability within a club's squad, especially for players whose performance data might not be as readily available or established. As such, the collaboration between quantitative data and the qualitative expertise of scouts plays an integral role in identifying and evaluating talents, thereby shaping a player's valuation within the competitive football market.

For a football player, age, potential for development, and the subsequent resale value play a crucial role in how much they are valued in the football space. Younger players, often deemed as having substantial room for improvement and development, command higher fees due to their extended potential careers and subsequent resale value. Clubs meticulously evaluate a player's age in conjunction with an assessment of their trajectory for growth to estimate not only their current contributions but also their potential market value in the future. This perceived potential for a player's development significantly influences their present valuation, as clubs recognize the attractiveness of players with promising growth prospects to potential buyers. The anticipation of a player's development trajectory and their potential to enhance their skills and abilities not only impacts their current market value but also serves as a significant indicator of their long-term worth, further influencing transfer fees and negotiations in the football market.

The article from the Football Observatory (2023) discusses the changing trends in player transfer fees based on age categories. It highlights an increase in investment for players aged 21 or younger and those between 22 and 25 years, compared to players aged 26 to 29. This shift is attributed to the growing strategy among clubs to trade players as a means of generating profits in the transfer market, reflecting a broader trend of economic segmentation in football.

Evaluating a player's injury history and physical condition is of paramount importance in determining their overall market valuation. Clubs engage in meticulous and comprehensive medical assessments designed to scrutinize a player's fitness, physical robustness, and long-term durability. An exhaustive analysis of a player's injury records serves as a vital tool for clubs to gauge the associated risks of potential recurring injuries, ultimately influencing a player's market value. The frequency, severity, and nature of past injuries are meticulously examined to foresee any potential impact on a player's performance, longevity, and the financial risks they might pose to potential buyers or clubs. This detailed scrutiny of a player's injury history significantly influences negotiations and the ultimate valuation in transfer deals.

The article from Analytics FC (2023) underscores the importance of thoroughly evaluating injury records and understanding their costs to football teams. It highlights that player injuries result not just in a loss of talent on the field but also incur significant monetary costs. The article aims to delve into recent Premier League injury data, exploring notable cases from various clubs and players, and the league overall. The methodology for calculating injury costs mentioned in the article considers injuries that led to players missing at least one Premier League match. However, it's noted that not all reported injuries that players may play through are included in this calculation. This approach to analysing injury data offers valuable insights into the tangible and intangible impacts of injuries in professional football.

A player's injury proneness or susceptibility to recurring injuries can potentially lead to renegotiations in terms of transfer fees, contract terms, and additional clauses related to performance or medical evaluations, consequently influencing the player's market value. The careful evaluation of a player's physical condition and injury history serves as a critical determinant in the intricate process of player valuation, influencing the decisions and risk assessment undertaken by clubs and stakeholders in the highly competitive football market.

### 2.2.3. Brand Value and Commercial Impact:

The off-field impact of a player on a club's brand and commercial revenue emerges as a pivotal consideration in their overall valuation. High-profile players, beyond their on-field prowess, often possess a substantial social media following, strong marketability, and significant merchandising potential, all of which contribute significantly to a club's commercial success. A player's off-field influence spans across sponsorship deals, jersey sales, and their marketability in various marketing campaigns. The financial implications stemming from a player's commercial appeal, including their capacity to attract sponsors, increase merchandise sales, and enhance the club's brand visibility, directly impact their overall market worth, which has led to, according to Statista (2023) that the Premier League has a combined brand value of nine billion euros, more than any other league in Europe's Big Five. Meanwhile, La Liga ranked second, with a brand value of just over four billion euros which is due to sponsorship deals which is linked with the high reputation of the players in the league. The intangible yet impactful contribution a player makes to the commercial aspects of a club translates into their market value, reflecting the broader economic significance of their off-field influence. This aspect of player valuation showcases the increasingly holistic nature of assessing a player's worth, where their marketability and off-field influence play a pivotal role alongside their on-field performance in determining their overall value within the competitive football market.

A table with numbers and text

Description automatically generated

#### Figure 1: Showing the top 7 shirt sales of premier clubs in terms of revenue.

### 2.2.7. Club Strategies and Fit Within the Team:

In the evaluation of football players, each club meticulously operates within a specific strategic framework, which encompasses the club's distinct playing style, tactical approach, and the intended role that a player is expected to fulfil within the squad. A player's valuation is intricately tied to their alignment with the team's strategic objectives, preferred playing style, and their aptitude to seamlessly integrate into the existing squad dynamics and positional requirements. The assessment of a player's suitability within the team structure emerges as a critical determinant in determining their market value. Compatibility with the club's strategic goals, tactical nuances, and the specific role a player is envisioned to undertake significantly influences their valuation. This assessment encompasses an in-depth analysis of how a player's attributes, playing style, and skill set complement or enhance the team's overall dynamics. A player who seamlessly fits into the tactical framework and aligns with the club's strategic vision is considered more valuable, given their potential to contribute effectively within the club's system. Hence, the evaluation of a player's suitability within a club's strategic structure and their seamless integration into the team's dynamics holds substantial weight in determining their market value and success within the competitive football market. The intricate evaluation of football players within the Premier League varies significantly across clubs, each operating within a specific strategic framework. This encompasses the club's distinct playing style, tactical approach, and the intended role a player is expected to fulfil within the squad. A player's valuation is intricately tied to their alignment with the team's strategic objectives, preferred playing style, and their aptitude to seamlessly integrate into the existing squad dynamics and positional requirements. For instance, the different recruitment models used by the Premier League's top six clubs demonstrate varied approaches in aligning player acquisition with strategic goals, ranging from Liverpool's advanced statistical analysis to Chelsea's director of football model (Planet Football, 2018). These clubs' strategies underscore the importance of aligning player recruitment with overarching tactical and strategic visions, a key determinant in player market valuation.

These diverse techniques are fundamental in the football market's valuation process, playing a pivotal role in assessing a player's worth. Clubs and stakeholders integrate these methods, balancing objective statistical analysis with subjective assessments to determine a player's market value, considering their performance, potential for growth, market comparable, injury history, off-field impact, and fit within the team's strategic vision. The convergence of these evaluation techniques is vital for informed decision-making within the competitive and ever-evolving football market.

## 2.4. How Premier League clubs use statistics when making transfer decisions:

In the modern era of football, the Premier League has become a highly competitive and lucrative environment where clubs strive for success on both domestic and international fronts. Player recruitment has evolved into a sophisticated process, with statistical analysis playing a pivotal role in guiding transfer decisions. Premier League clubs have increasingly embraced data-driven methodologies to gain a competitive edge, employing advanced statistical models and analytics to inform their player recruitment strategies.

One key area where statistics are extensively utilized is in assessing player performance. Clubs employ a plethora of metrics to gauge a player's effectiveness on the pitch. Traditional statistics such as goals, assists, and pass completion rates are complemented by advanced metrics like expected goals (xG), expected assists (xA), and shot conversion rates. These metrics provide a nuanced understanding of a player's contribution, transcending basic numbers to reveal the underlying quality of their performances.

Clubs leverage position-specific analytics to identify players who align with their tactical philosophies. Defensive metrics, such as interceptions and successful tackles, aid in evaluating the defensive prowess of potential signings. Similarly, offensive metrics like key passes, successful dribbles, and offensive duels won help assess attacking capabilities. By tailoring statistical analyses to position-specific requirements, clubs can pinpoint players who not only excel individually but also seamlessly integrate into the team's tactical framework.

The role of data extends beyond on-field performance to injury analysis and risk assessment. Premier League clubs invest significantly in sports science and medical data to evaluate a player's injury history and susceptibility. Comprehensive medical assessments are complemented by data-driven insights into injury recurrence rates, helping clubs make informed decisions on the long-term durability and reliability of a potential signing. This meticulous approach mitigates the financial and performance risks associated with injuries, a crucial consideration in the high-stakes world of football transfers.

Statistical analysis is instrumental in assessing a player's consistency and adaptability. Time-series data allows clubs to track a player's performance over multiple seasons, providing insights into their ability to maintain a high level of play over time. This longitudinal analysis helps in identifying players who show consistency in their performances, a key attribute for sustained success in the demanding Premier League.

In the evolving landscape of football analytics, machine learning algorithms also find application in player recruitment. Predictive models can forecast a player's future performance based on historical data, offering clubs a glimpse into a player's potential trajectory. These models consider various factors, including age, playing style, and historical performance trends, providing a forward-looking perspective that aids clubs in making strategic and future-oriented transfer decisions.

Collaboration with specialized data analytics firms and the development of in-house analytics departments further underscore the commitment of Premier League clubs to data-driven decision-making. The use of tracking data, video analysis, and advanced statistical models has become integral to the scouting and recruitment processes, allowing clubs to identify undervalued talents, optimize squad composition, and strategically invest in players who align with their long-term vision.

In conclusion, statistical analysis has revolutionized player recruitment in the Premier League, transforming it into a nuanced and data-driven discipline. From assessing on-field performance to mitigating injury risks and predicting future contributions, clubs leverage a diverse array of statistical tools to make informed decisions in the dynamic and competitive landscape of English football. As the influence of data continues to grow, the marriage of statistical insights and football expertise defines the modern approach to player recruitment in the Premier League, shaping the trajectory of clubs aiming for sustained success in this elite footballing arena.

## Pre-modern statistical analysis:

Before the era of extensive statistical analysis and advanced modelling, transfer decisions in the Premier League were shaped by a combination of traditional scouting methods, subjective assessments, and often, a reliance on the intuition and experience of club managers and directors. The landscape of player recruitment was characterized by a more anecdotal and less data-driven approach, where decisions were often influenced by personal observations, reputation, and networking within the footballing community.

Scouting, in its traditional sense, was a cornerstone of the transfer decision-making process. Clubs deployed scouts to watch players in person, collecting information on their technical skills, positional awareness, and overall playing style. These scouts were tasked with identifying talents from various leagues, relying on their expertise and ability to recognize potential. While this approach allowed for firsthand observations of players, it was inherently limited by the geographical constraints of scouting networks, and often, the subjective biases of individual scouts.

Player reputation and word-of-mouth played a substantial role in transfer decisions. Clubs would often sign players based on their reputations in the footballing community, recommendations from other players, or feedback from trusted contacts. This network-driven approach meant that players with established names or those endorsed by influential figures in the football world were more likely to attract attention, regardless of whether their playing style or abilities were an ideal fit for a particular club.

Managers and coaching staff, equipped with their experience and understanding of the game, were central decision-makers in the transfer process. Their judgments were based on their knowledge of the team's needs, playing style preferences, and an assessment of how a player might fit into the existing squad dynamics. While this approach provided a sense of continuity and coherence within a team, it also meant that decisions were highly dependent on the individual preferences and tactical inclinations of the manager.

Financial considerations also played a significant role in transfer decisions, but the assessment of a player's market value was more subjective. Clubs negotiated transfer fees based on perceived player worth, personal negotiations, and the financial standing of the selling club. The lack of standardized metrics for assessing a player's value often led to protracted negotiations and, at times, inflated transfer fees.

Injuries and fitness evaluations were predominantly reliant on traditional medical assessments. While thorough, these assessments lacked the depth of data-driven insights into injury history, recurrence rates, and long-term impact on performance that modern sports science and analytics provide.

Transfer decisions in the pre-statistics era were marked by a blend of subjective judgments, personal relationships, and a reliance on traditional scouting methods. While this approach had its merits, it also left room for biases, limited perspectives, and a lack of systematic analysis. The advent of statistical analysis and advanced modelling in recent years has brought a more objective and comprehensive dimension to player recruitment, revolutionizing how Premier League clubs approach the complex and high-stakes task of building and reshaping their squads.

## 2.5. Evolution of Financial Dynamics in Player Transfers

The Premier League's substantial economic growth, fuelled by lucrative broadcasting rights, sponsorships, and its extensive global reach, has significantly influenced the player transfer market. This financial evolution has led to market inflation, where player transfer fees have skyrocketed, necessitating a recalibration of clubs' transfer strategies and player valuations. Furthermore, the introduction of Financial Fair Play (FFP) regulations by UEFA has had a profound impact on how clubs approach spending, shaping recruitment strategies to align with these regulatory constraints. This section will delve into these economic dynamics, exploring their implications on the changing landscape of player recruitment.

The expansion of international scouting networks has diversified the talent pool accessible to Premier League clubs, bringing players from varied cultural and footballing backgrounds into the league. This globalization of talent acquisition presents both challenges and opportunities, from adapting scouting methods to addressing the needs for cross-cultural adaptation and integration of foreign players. The literature review will explore these facets, examining the complexities and nuances of integrating players from diverse backgrounds into the competitive environment of English football.

The partnership between Premier League clubs and sports analytics firms signifies a pivotal shift towards a data-driven approach in player recruitment. These collaborations offer clubs a range of services, from sophisticated data collection to advanced predictive modelling, enhancing the efficacy of their recruitment strategies. Analytics have played a key role in shaping successful transfer decisions, illustrating how data-driven approaches are increasingly becoming integral to competitive success. As said in the article by FcBusiness (2018) “Having the analytical tools or processes to leverage all the information being collected by teams gives clubs a truly objective means to assess the compatibility of potential talent and provides a massive advantage.” These analytical companies are able to use their considerable database of player statistics and their large amount of employees to be able to use in depth considerations to help these clubs make the big decisions on these transfers of players.

## 2.8. Ethical Considerations and Player Welfare

Ethical considerations in player recruitment encompass a spectrum of issues, including player welfare, the influence of agents, and the psychological impact of transfer speculation. Additionally, the ethics surrounding the recruitment of young players, balancing their career development with educational needs and the pressures of early professionalization, are of paramount importance. This part of the review will address these ethical concerns, emphasizing the need for responsible and considerate practices in player recruitment and development.

Emerging technologies such as artificial intelligence, machine learning, and biometric analysis are poised to revolutionize player analysis in the realm of football. The potential of these technologies for predictive modelling and forecasting player performance offers exciting prospects for the future of player recruitment. Exploring these emerging trends, considering how they might influence player evaluation, long-term career planning, and the overall approach to recruitment in the dynamic landscape of the Premier League. The BeSoccer (2023) article highlights the transformative impact of artificial intelligence (AI) and machine learning on football, particularly in data collection and analysis. It outlines how these technologies enable coaches to track complex gameplay dynamics more efficiently and streamline the detailed analysis of on-field performance. The article also points out a shift in player scouting methods, with a reduced reliance on traditional techniques like personal meetings and game reports in favour of AI-based solutions. This tech-driven approach is changing the game, offering a deeper understanding of football dynamics, and maintaining the sport's excitement and analytical depth.

The article BeSoccer (2023) also discusses how emerging technologies like Augmented and Virtual Reality, analytics, and mobile applications are revolutionizing football. It emphasizes the increasing reliance of teams on data and insights from these tools for refining strategies, performance analysis, and gaining a competitive edge. For instance, top clubs are using advanced technologies like 3D or 4D for studying opposition teams and assessing player skills for recruitment, and real-time tracking technologies in training sessions for precise performance metrics.

## 2.10. Conclusion

In summarizing the extensive literature on player recruitment in the Premier League, this review has highlighted key themes and findings that shape our current understanding of the field. By identifying gaps in the existing literature, it points towards potential directions for future research, such as investigating the long-term impacts of data-driven recruitment on club performance and player career trajectories. This conclusion serves to synthesize the diverse aspects explored, framing a comprehensive perspective on player recruitment strategies in the context of the evolving and multifaceted world of English football.

## 2.11. Summary of literature Review:

The literature review offers a panoramic exploration of player recruitment strategies in English football, anchoring itself in the established research within football economics and sports management. This comprehensive investigation examines the determinants of transfer fees, their financial implications for football clubs, and the intricate factors influencing player recruitment decisions. From the fundamental statistical analysis and comparative assessment of player performances to the indispensable role of scouting, age valuation, injury assessment, and off-field influence, this review elucidates the multifaceted methodologies underpinning the valuation process. It accentuates the significance of aligning a player's attributes with a club's strategic framework, emphasizing the amalgamation of objective statistical metrics and subjective evaluations. Ultimately, this synthesis of diverse techniques and considerations forms the bedrock of player valuation, offering clubs and stakeholders a comprehensive toolkit for informed decision-making in the ever evolving and competitive landscape of English football.

# Chapter 3 - Research Methodology:

The primary objective of this research is to explore player recruitment strategies in the English Premier League. The study aims to identify key performance indicators (KPIs) influencing player recruitment, analyse the impact of player demographics on recruitment success, and explore trends in player recruitment over recent seasons.

Quantitative Data: The primary dataset, including player performance metrics such as goals scored, assists, pass accuracy, and other key performance indicators, serves as the foundation for objective analysis. This data provides quantifiable measures of player performance, contributing to a solid empirical basis for the study.

Qualitative Data: To complement the quantitative analysis, qualitative data will be gathered from a variety of sources. This includes interviews with key stakeholders such as club managers, scouts, and players. The aim is to gain deeper insights into the subjective decision-making processes, cultural influences, and personal experiences that shape recruitment strategies in the Premier League.

The combination of these data types allows for a more nuanced understanding of player recruitment. While the quantitative data offers measurable evidence of player abilities and outcomes, the qualitative data provides context and depth, revealing the intricacies of recruitment strategies that numbers alone cannot capture. This mixed-methods approach ensures a more rounded and robust analysis, mitigating the risks of bias and enhancing the overall credibility of the research findings.

Data collection is centred on a comprehensive dataset that includes player names, ages, nationalities, positions, and various performance metrics. The dataset provides a rich source of information covering recent seasons in the Premier League. Additionally, the research may incorporate supplementary data sources, such as historical transfer data to enrich the analysis. These additional datasets will be integrated methodically to ensure consistency and reliability.

The data cleaning and preprocessing phase is crucial for ensuring the quality and usability of the data. This includes initial steps such as handling missing values, removing duplicates, normalizing text fields, and ensuring correct data types. Further, data transformation will be undertaken to make categorical variables suitable for analysis, manage any time-series data effectively, and develop new features relevant to the study. Ensuring data accuracy and consistency through quality assurance measures will be a priority, potentially involving expert validation.

Exploratory Data Analysis (EDA) will be conducted to gain preliminary insights into the dataset. This includes analysing descriptive statistics for each variable and employing visual analysis tools to identify patterns, trends, and outliers. The analytical framework will encompass both statistical analysis and, where applicable, machine learning models. Statistical methods such as correlation analysis and regression models will be used to explore relationships between variables. Predictive models or clustering techniques may also be employed to uncover patterns or predict recruitment outcomes, with appropriate validation techniques to ensure the robustness of the findings.

Ethical considerations, particularly concerning data privacy and usage compliance, will be addressed rigorously. The research methodology also acknowledges potential limitations, including data limitations, inherent biases, and the scope of analysis. Recognizing and addressing potential limitations is critical for the integrity and validity of the research. This research acknowledges several limitations and during this research, will attempt to counter the limitations in the following ways:

**Data Limitations:** The primary dataset may not capture the full spectrum of factors influencing player recruitment. To mitigate this, the research will incorporate diverse data sources to provide a more holistic view.

**Bias and Subjectivity:** Given the subjective nature of qualitative data, there is a risk of bias in interpretations. This will be mitigated by employing a systematic approach to data analysis.

**Scope of Analysis:** The research focuses primarily on the English Premier League, which may limit the generalizability of findings to other leagues or sports. Acknowledging this limitation, the study will clearly define its scope and suggest areas for future research in different contexts.

**Technological Changes:** The rapidly evolving nature of data analytics in sports could make some findings less relevant over time. To address this, the study will include a discussion on emerging trends and technologies, ensuring the research remains forward-looking.

By proactively identifying and addressing these limitations, the research aims to strengthen its methodology, enhancing the credibility and applicability of its findings.

Interpreting the results will involve deriving data-driven insights and contextualizing findings within the broader landscape of the English Premier League and existing literature. This phase is critical in translating analytical outcomes into meaningful conclusions regarding player recruitment strategies.

The research will culminate in the development of potential recommendations for player recruitment strategies in the Premier League. The entire process, from data collection to conclusion, will be documented meticulously to ensure clarity, replicability, and validity. The research is expected to follow a structured timeline, with each phase being carefully managed for effective work progression. Furthermore, the methodology is designed to be iterative, allowing for adjustments and refinements as the research delves deeper into the data.

## Statistical Models and Machine Learning Techniques

### Regression Analysis

Regression analysis stands as a cornerstone for understanding the relationships between player attributes and their transfer fees. Linear regression could be the starting point, aiming to predict a player's transfer fee based on various numerical inputs such as age, performance metrics (goals, assists, defensive actions), and market factors (player's popularity, contract duration). To capture non-linear relationships, polynomial regression might be introduced, allowing for a more nuanced understanding of how player attributes interact to influence their market value.

Advanced regression techniques, such as Ridge and Lasso regression, can be employed to address multicollinearity and overfitting, ensuring the model's robustness. These methods add a penalty term to the cost function, controlling the model's complexity by shrinking coefficient estimates towards zero.

### Decision Trees and Random Forests

Decision trees offer a way to model decisions and their possible consequences, including chance event outcomes. In the context of player recruitment, a decision tree could categorize players into different value segments based on attributes, providing clear criteria for decision-making. However, decision trees can suffer from overfitting, making them sensitive to the training data.

Random Forests, an ensemble of decision trees, mitigate this issue by averaging multiple trees' predictions, thus improving model accuracy and stability. This technique can handle a high dimensionality of data and provides importance scores for each feature, offering insights into what attributes most significantly impact a player's transfer fee.

### Clustering

Clustering algorithms, such as K-means or hierarchical clustering, can be applied to segment players into groups based on similarities in their attributes without prior labelling. This unsupervised learning method can uncover natural groupings within the data, such as identifying types of players that are similar in playing style but vary in market valuation. Clustering provides strategic insights into the player market, highlighting undervalued players or identifying niches in the recruitment strategy.

### Implementation and Evaluation:

Implementing these models requires a careful approach to data preprocessing, including feature selection, normalization, and dealing with missing values. Cross-validation techniques should be employed to assess model performance, using metrics such as R-squared, mean absolute error (MAE), and root mean squared error (RMSE) for regression tasks, or accuracy, precision, recall, and F1 score for classification tasks.

Model interpretability is also crucial, especially when the findings are intended to inform decision-making in real-world scenarios. Techniques such as SHAP (Shapley Additive Explanations) values or feature importance scores can help elucidate how different attributes influence the model's predictions, providing actionable insights for player recruitment strategies.

### Model Validation and Evaluation:

To ensure the validity and robustness of our predictive models, a methodical approach to model validation and evaluation. Initially, the dataset was partitioned using a train-test split method, where 80% of the data was allocated for model training and the remaining 20% was reserved for testing. This ratio was carefully chosen to provide a substantial amount of data for learning while retaining enough independent data for an unbiased evaluation of the model's predictive power.

The training phase involved fitting the Random Forest regression models to the data, enabling the learning of complex relationships between a player's attributes and the corresponding transfer fee. The model's ability to generalize was then assessed using the independent test set. This process was crucial to prevent the model from merely memorizing the data (overfitting) and to evaluate its performance on new, unseen data.

To further validate the model, k-fold cross-validation, a robust statistical technique to assess the predictive performance of the models. I chose a 10-fold cross-validation approach, where the dataset was split into 10 equal parts. In each iteration, nine parts were used for training the model, and the remaining part served as the validation set. This technique enhances the reliability of performance estimates by reducing the variability that could result from a random partitioning of the data. Through k-fold cross-validation, it could ensure that every observation from the original dataset had the chance to be used in both training and validation, thus providing a comprehensive assessment of the model's performance.

# Chapter 4: Implementation –

## Exploratory Data Analysis:

A graph showing the value of a player

Description automatically generated

#### Figure 2 - Distribution of Player Market Values in the English Premier League

The distribution of player market values within the English Premier League is depicted through a histogram, providing a visual representation of the valuation spread across the league's players. This distribution exhibits a pronounced right skew, characteristic of economic data where a larger number of observations are clustered on the lower end of the scale.

The majority of players fall within the initial bins of the histogram, indicating a lower market value. This aggregation around the lower market value range highlights a significant concentration of players whose financial valuation, while modest, represents the economic reality for the bulk of professional athletes in the league. The mode of the distribution is situated within these initial bins, signifying the most common valuation bracket for players.

As the histogram extends towards the right, the frequency of players within each bin diminishes substantially, denoting the rarity of players with higher market values. The 'long tail' in the distribution suggests the presence of outliers - elite players whose market values are exceptionally high due to factors such as their superior athletic capabilities, influence on team success, and personal brand appeal. These players, although few may have a disproportionate impact on team valuations and strategies.

This disparity between the mode and mean of the distribution indicates the skewness' impact, where the average market value is elevated by the presence of these high-value outliers. It underscores the importance of the median as a more representative measure of central tendency for skewed distributions, as it is less influenced by extreme values.

The graphical analysis further reveals economic insights pertinent to the league. The 'tail' of the distribution reflects the fundamental economic principle of scarcity: the limited supply of top-tier talent leads to their elevated valuations. In contrast, the larger supply of players with lower market valuations potentially results in their reduced individual bargaining power and transfer fees.

For strategists within the league, this distribution is instructive for understanding the economic landscape of player recruitment and development. It suggests that while significant investment is required to secure the services of the highest-valued players, there is ample opportunity within the larger pool of lower-valued players. This area may be ripe for the identification of undervalued talent, indicating a potential competitive advantage for clubs that excel in player development and scouting efficiency.

**A screenshot of a graph

Description automatically generated**

#### Figure 3 – Scatter plots showing goals, assists and xG against Market Value.

The set of scatter plots provided illustrates the relationship between player market values in the English Premier League and three key performance metrics: goals scored, assists, and expected goals (xG). Each plot provides a visual correlation between a single performance metric and the market value, which can be used to infer potential valuation drivers within the league.

### Goals vs. Market Value Scatter Plot Analysis:

The top plot correlates the number of goals scored by a player with their market value. Each point on the plot represents an individual player. The horizontal spread shows a range from players who have not scored to those with the highest goal tallies, while the vertical distribution indicates market value, spanning from lower-valued players to those with values in the tens of millions. The plot suggests a positive correlation between the number of goals scored and a player's market value; however, the spread of points indicates considerable variability. A cluster of players with low to moderate goal counts have a wide range of market values, and even among higher goal scorers, there is significant variance in market value. This suggests that while goal-scoring ability is likely a factor in determining a player's market value, it is not the sole determinant, and other factors may also play a significant role.

### Assists vs. Market Value Scatter Plot Analysis:

The middle plot presents the number of assists provided by players against their market values. Similar to the goals plot, the data points suggest a positive relationship between the number of assists and market value, although this relationship appears to be less pronounced than that for goals. There is a dense cluster of players with fewer assists across a broad range of market values. A few outliers have a high number of assists and also command high market values, hinting at the premium placed on players who contribute to goal creation. Nonetheless, the spread suggests that assists are one of several factors considered in player valuation.

### Expected Goals (xG) vs. Market Value Scatter Plot Analysis:

The bottom plot illustrates the relationship between expected goals—a measure of the quality of scoring chances—and market values. The data points are more widely distributed, indicating that xG, while an advanced metric, varies widely among players with similar market values. There seems to be a cluster of players with low xG values across a range of market valuations, and as xG increases, the market value appears to increase as well. However, there is less concentration of players with high xG values, indicating these players are rarer and potentially more valued.

### Overall Analysis:

In each of the three plots, the presence of several players with low performance metrics and high market values suggests additional factors influencing valuation. These may include the player's position, age, potential, contractual situation, injury history, or off-field marketability. The plots show that while there is a relationship between performance metrics and market value, it is not a simple linear relationship. The data indicate that player market values in the English Premier League are likely influenced by a complex set of attributes that extend beyond on-field performance measures. This multiplicity of factors makes the valuation process nuanced and multifaceted, reflecting the complexity of player valuation in professional football.

A graph showing a number of numbers

Description automatically generated

#### Figure 4 – A bar chart showing Average market value per position.

The bar chart provided illustrates the average market values of players in the English Premier League, segmented by playing position. This visualization enables an in-depth comparison of how player valuations differ according to the roles they fulfil on the pitch.

**Second Strikers and Attacking Midfielders**

The chart indicates that second strikers command the highest average market value among all positions, closely followed by attacking midfielders. This suggests that players who can contribute directly to scoring goals, either by finishing chances or by creating them, are highly prized. Their skill sets, which likely include dribbling, creativity, and an eye for goal, are evidently valued highly in the transfer market.

**Wingers and Midfielders**

Left wingers also hold a high average market value, possibly due to the premium on players who can effectively operate in wide areas, deliver crosses, and cut inside to shoot. This position's valuation is slightly above that of right wingers and central midfielders, suggesting a potential bias towards players adept at influencing the game from the left flank. Defensive midfielders also have a relatively high average market value, underscoring the importance of roles focused on breaking up opposition play and initiating attacks from deeper positions on the field.

**Centre-Forwards and Defenders**

Centre-forwards and centre-backs are positioned in the middle of the value range, indicating a balanced appreciation for goal scorers and those who play crucial roles in defence. Full-backs (right-backs and left-backs) follow closely, with their modern roles increasingly involving attacking contributions, which may be reflected in their valuations.

**Goalkeepers**

Interestingly, goalkeepers appear in the lower half of the valuation range. Despite the critical nature of their role, this may reflect a market perception that quality goalkeepers are more abundant or that individual goalkeepers contribute less to match outcomes than outfield players.

**Midfield Variations**

The lower end of the valuation spectrum is occupied by left and right midfielders. The distinction between these wide midfielders and wingers could suggest that traditional, wide midfield roles are less valued compared to their more attacking counterparts, or that these terms may be used interchangeably, causing some discrepancy in the data.

**Positional Valuation Insights**

Overall, the data reveals that the market values offensive contributions highly, with players in goal-scoring and creative roles receiving the highest average valuations. It also suggests that while defence is valued, the premium on offensive output in the current market is more pronounced. This could reflect strategic trends within the league, where attacking play is seen as crucial to winning matches and entertaining spectators, thus justifying higher investments in such players.

**Specialised Skills**

Datasets such as the dataset that is being used to conduct this research on do not have statistics available to showcase some potential specialised skills that players can use in games that are hard to quantify. One of these skills as an example is the art of the set-piece. According to Staff (2024) approximately 19.3% of all total goals scored in the premier league are scored via set-pieces. Whether that be a penalty, free-kick or corner goal. So having these players that can use these situations in games to add goals for their teams by winning these situations, scoring from them or defending them is both hard to quantify and also really important for the team.

The analysis of market value by player position provides clubs and stakeholders with strategic insights into where the market is channelling its financial resources, revealing the types of players most likely to command significant transfer fees. It also helps to contextualize the financial aspects of team building within the league's competitive landscape.

A graph with blue dots

Description automatically generated

#### Figure 5 - Analysis of how a player's age might correlate with their perceived economic worth in the transfer market.

### Distribution and Density:

The plot shows a wide scatter of data points across various ages, with a noticeable concentration of players between the ages of 20 and 30. The density of points suggests that most professional players fall within this age range, which likely corresponds to the peak playing years for professional athletes in this league.

**Market Value Trends Across Age:**

There appears to be a trend where players in their early to mid-twenties command the highest market values, with a visible decrease as age increases. The highest market values are concentrated among players in their mid-twenties, indicating that this age range may be perceived as the prime of a player's career. Beyond this point, there is a noticeable decline in market values, particularly as players approach their early thirties, reflecting the market's valuation of potential career longevity and expected performance.

**Youth Premium:**

Younger players (closer to the age of 20) exhibit a wide range of market values, with some commanding high valuations. This might reflect the premium placed on potential and the expected future return on investment that younger talent can bring to a club.

**Value Persistence Beyond Peak Years:**

A few players maintain relatively high market values well into their late twenties and early thirties, which could indicate exceptional skill, leadership, or other attributes that maintain their high valuation despite the typical age-related decline.

**Overall Age-Value Relationship:**

The general pattern suggests a non-linear relationship between age and market value, with a peak in the mid-twenties followed by a gradual decline. This pattern underscores the importance of age as a factor in player valuation, reflecting both the potential for future development in younger players and the expected decline associated with aging for older players.

**Implications:**

For clubs, this visualization underscores the importance of strategic planning in player signings and the development of talent. The focus on recruiting players at an age where they can provide immediate performance and offer potential for appreciation in market value could be a crucial component of a club's long-term financial and sporting strategy.

**Caveat:**

While the graph provides a snapshot of the relationship between age and market value, it is essential to note that it does not capture all factors influencing market values, such as position played, current contract length, injury history, and off-field marketability, which can all significantly affect a player's valuation.

A graph showing a market value distribution

Description automatically generated

#### Figure 6 – A box plot to show distribution of market values per player position.

The box plot displays the distribution of market values by player position in the Premier League. Each box represents the interquartile range (IQR) of the market value for players within a specific position, showing the median, the 25th percentile (Q1), and the 75th percentile (Q3). The whiskers extend to the most extreme data points not considered outliers, and points outside of this range are plotted as individual dots, representing outliers.

### Positional Market Value Trends:

The plot indicates that attacking midfielders have the highest median market value, followed closely by left-backs and right-backs, which suggests a significant valuation of these roles within the league.

The distribution for attacking midfielders is also wide, indicating variability in the market value within this position, which could be attributed to different roles an attacking midfielder may play, from playmaking to scoring.

**Variability and Outliers:**

Several positions show a wide range of market values within the IQR, such as central midfield and centre-back, which may imply a diversity in player quality and reputation within these positions.

Positions like right midfield and second striker show a narrower IQR but include outliers with high market values, possibly indicating that exceptional talent in these roles can command premium valuations.

**Comparative Valuation:**

The median market value for defensive roles like centre-backs and goalkeepers is lower compared to more attacking positions, which may reflect strategic valuations in the league prioritizing goal-scoring abilities.

Right midfielders have the lowest median market value, which could indicate a current undervaluation of this role or a surplus of players in this position, reducing individual market values.

**Outlier Analysis:**

Notably, certain positions like right winger and centre-forward have outliers that are significantly higher than the upper quartile range, suggesting the presence of star players in these positions whose market values far exceed the norm due to exceptional skills or demand for their playing style.

**Strategic Implications:**

The variation in market values across positions could have implications for club strategies in player recruitment and development, potentially indicating areas where the market is more competitive or where undervalued talent might be found.

Clubs may use such data to identify positions that require investment or where there is an opportunity to develop players internally for future financial gains.

**Caveats for Interpretation:**

While box plots provide a snapshot of market values across positions, they do not capture all the dynamics of player valuation, such as contract length, age, potential for growth, or marketability.

Additionally, the presence of outliers indicates that exceptional individual circumstances can significantly affect a player's market value, beyond the general trend of their positional group.

The box plot provides valuable insights into the financial aspects of player positions within the English Premier League and can serve as a tool for clubs to make informed decisions in the transfer market. It highlights the variability and complexities of player valuation, with clear indications of how different positions are valued differently based on market trends and individual player attributes.

A screenshot of a graph

Description automatically generated

#### Figure 7 – Correlation Matrix

The correlation matrix provided above is a heatmap that visualizes the strength and direction of the linear relationship between several variables, including market value, goals (Gls), assists (Ast), expected goals (xG), age, and matches played (MP) for players in the English Premier League.

### Interpretation of Correlation Coefficients:

* A coefficient close to 1 or -1 indicates a strong positive or negative linear relationship, respectively.
* A coefficient close to 0 indicates no linear relationship.

**Analysis of Market Value Correlations:**

* Market Value and Goals (Gls): There is a moderate positive correlation (0.34) between goals scored and market value, suggesting that players who score more tend to have higher market values.
* Market Value and Assists (Ast): The correlation is similarly moderate (0.37), indicating that players who frequently assist also tend to have higher market values, albeit slightly more so than goal scorers.
* Market Value and Expected Goals (xG): This also shows a moderate positive correlation (0.37), suggesting that players who are expected to score (based on the quality of chances) are valued more highly.
* Market Value and Age: The correlation is moderately negative (-0.35), indicating that younger players generally have higher market values, which could reflect their longer remaining career span and potential for development.
* Market Value and Matches Played (MP): The correlation (0.34) indicates a moderate positive relationship, possibly reflecting experience or consistent selection as factors in a player's valuation.

**Performance Metrics Interrelationships:**

* A very strong positive correlation (0.93) between goals and expected goals suggests that players who are expected to score tend to actually score, reinforcing the reliability of xG as a predictive performance metric.
* Goals and assists have a moderate positive correlation (0.55), which might indicate players who score also contribute to setting up goals.
* Assists and expected goals show a positive correlation (0.58), suggesting that players involved in creating chances are also likely to be in positions to score.
* Interestingly, there's virtually no correlation between age and goals, assists, or expected goals, implying that these performance metrics do not necessarily decrease with age within the dataset.

**Implications:**

* The findings from this correlation matrix can provide clubs with data-driven insights into which performance metrics might influence a player's market value the most. For example, clubs might prioritize scouting for younger players who have high expected goal figures.
* The lack of strong correlation between age and performance metrics such as goals or assists may suggest that age, while influencing market value, does not directly dictate a player's performance output.

**Caveats:**

* Correlation does not imply causation, and these relationships may be influenced by other unaccounted factors.
* The correlation matrix is a snapshot in time and does not account for changes in the market or individual player circumstances.

In summary, the heatmap reveals a nuanced picture of how different factors are interrelated and can influence player market values in the English Premier League. These insights can inform recruitment and negotiation strategies, as well as broader discussions on player valuation and performance assessment.

A group of blue dots

Description automatically generated

#### Figure 8 – Scatterplot Matrix

### Distribution Analysis:

* The histogram for Market Value shows a right-skewed distribution, suggesting that most players are valued at a lower price point, with fewer players commanding very high values.
* Goals (Gls) and Assists (Ast) also exhibit right-skewed distributions, indicating that most players score a few goals or assists.
* The Expected Goals (xG) histogram appears right-skewed as well, aligning with the notion that most players are expected to score fewer goals, with only a few expected to score more.
* The distribution of Age is more bell-shaped, suggesting a relatively normal distribution of players across different ages, with a peak around the mid-20s.

**Correlation Analysis:**

* In the scatterplots, Market Value vs. Goals, Market Value vs. Assists, and Market Value vs. Expected Goals all display positive correlations, with a cluster of players who have lower metrics in each category but varying market values. There are also players with high performance metrics and high market values, suggesting a positive relationship.
* A notable aspect of these plots is the presence of outliers, particularly in the Market Value dimension, where some players have exceptionally high market values regardless of their performance metrics.
* Age shows an interesting pattern; younger and older players seem to have a lower market value, with a sweet spot for players in their mid-20s. This could indicate a peak period of value in a player's career, potentially combining both experience and remaining potential.

**Inter-variable Relationships:**

* Goals and Assists have a positive relationship, suggesting that players who score more goals also tend to assist more, and vice versa.
* A very strong positive relationship is observed between Goals and Expected Goals, which aligns with expectations as xG is designed to be a predictor of goal-scoring performance.
* Age does not seem to have a strong linear relationship with performance metrics (Goals, Assists, xG), indicating that these factors may be relatively stable across a player's prime years.

**Overall Insights:**

The scatterplot matrix provides a multidimensional view of the data, revealing the distribution of individual variables and the relationships between them. While Market Value appears to increase with performance metrics like Goals, Assists, and Expected Goals, there's significant variability suggesting other factors at play. Age has a more complex, non-linear relationship with Market Value and performance metrics, indicating that age-related factors impact these relationships differently across a player's career.

## Performance Models:

|  |  |  |  |
| --- | --- | --- | --- |
|  | MSE | RMSE | R^2 |
| Linear Regression | 240.06 | 15.49 | 0.69 |
| Random Forest | 432.13 | 20.79 | 0.45 |
| Gradient Boosting | 420.38 | 20.50 | 0.46 |

#### Figure 9 – Table of regression models

The investigation was through the deployment of three distinct regression models: Linear Regression, Random Forest Regression, and Gradient Boosting Regression. The efficacy of these models in predicting player market values was rigorously evaluated, utilizing Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R²) as benchmarks for performance.

The analysis yielded enlightening results, with the Linear Regression model emerging as the superior predictor, elucidating approximately 69% of the variance in players' market values as denoted by its R² value. This model's dominance over its more complex counterparts, the Random Forest and Gradient Boosting models, suggests a strong linear relationship between the selected predictors and players' market values. The findings underscore the significant impact of performance metrics such as goals scored, assists, and overall game participation on a player's market valuation. Interestingly, the complexity of the Random Forest and Gradient Boosting models did not translate into enhanced predictive accuracy, potentially due to overfitting or the unpredictable nature of market valuations.

This exploration into player valuation strategies highlights the critical role of quantifiable performance metrics in determining a player's market worth within the English Premier League. The insights gleaned underscore the nuanced valuation landscape across different player positions, suggesting a tailored approach to player valuation strategies. The robust predictive capability of simple linear relationships between performance metrics and market value offers a pragmatic foundation for clubs and sporting directors aiming to refine their player evaluation and trading strategies, accentuating the tangible contributions of players on the field.

## K-means Clustering Analysis:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| K-Means Cluster Profiles | | | | |
| Cluster Number | Goals | Assists | Minutes | Market Value |
| 1 | 0.755556 | 0.466667 | 712.214815 | 13.976667 |
| 2 | 15.428571 | 7.714286 | 2722.071429 | 78.428571 |
| 3 | 5.310345 | 4.620690 | 2320.568966 | 44.327586 |
| 4 | 1.115789 | 0.978947 | 2444.494737 | 26.926316 |

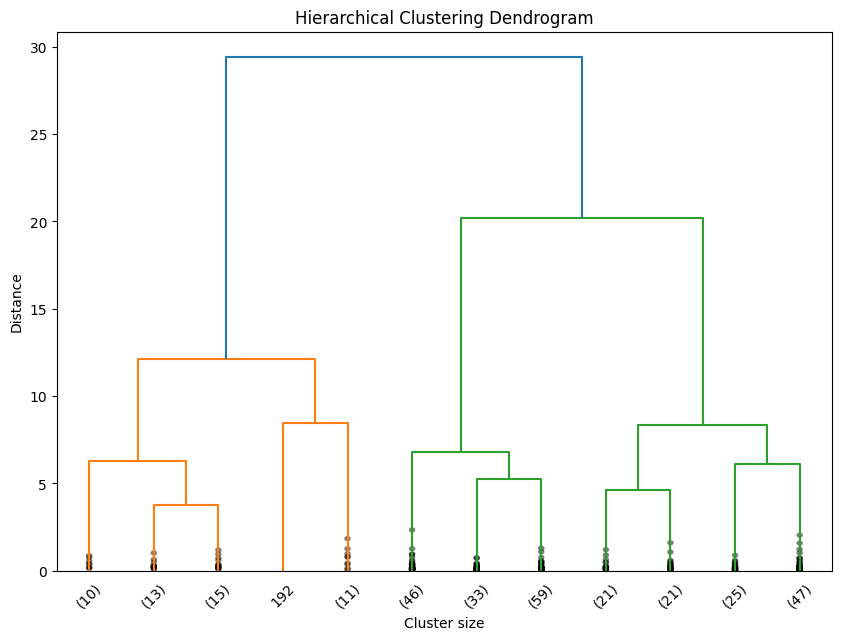
#### Figure 10 – Table of K-Means cluster profiles

In the exploration of player valuation strategies in the English Premier League, a K-means clustering algorithm was employed to segment players based on performance metrics and market values. K-means clustering was chosen for its efficiency and effectiveness in identifying distinct groups within large datasets. The analysis began with preprocessing steps, including data normalization to ensure equal weighting of features such as goals, assists, minutes played, and market value. We determined the optimal number of clusters to be four, based on a balance between within-cluster sum of squares (WCSS) and the interpretability of the results.

The K-means algorithm separated the players into four distinct clusters, each representing a unique player profile based on their performance metrics and market valuation. Cluster 1 was characterized by star players with high goals, assists, and market values, indicating their critical roles in their teams' offensive efforts. Cluster 2 included solid contributors who, despite moderate goals and assists, played significant minutes, suggesting a consistent and dependable presence on the field. Cluster 3 encompassed reliable regulars, players with mid-range market values who contributed through consistent playing time rather than standout offensive statistics. Lastly, Cluster 0 captured emerging or backup players, indicated by their lower market values and minimal playing time, highlighting their potential for future growth or role as squad depth.

The K-means analysis provided valuable insights into the valuation strategies employed in the English Premier League. It highlighted the diversity of player roles and how performance metrics, combined with market values, differentiate players within the league. This clustering approach facilitates a nuanced understanding of player valuation, offering a systematic method to categorize players beyond traditional metrics.

## Hierarchical Clustering Analysis:



#### Figure 11 – Hierarchical Clustering Analysis Diagram

Following the K-means clustering, a hierarchical clustering analysis was conducted to further explore the structure of player valuation in the English Premier League. Hierarchical clustering was selected for its ability to visualize data groupings at different levels of similarity, offering a detailed perspective on how players cluster together based on their attributes. This analysis utilized Principal Component Analysis (PCA) to reduce the dimensionality of the data, focusing on two principal components derived from the original performance metrics and market values. This reduction enabled a more manageable visualization and interpretation of the clusters.

The hierarchical clustering was represented through a dendrogram, illustrating the amalgamation of individual players or small groups into larger clusters based on their similarity. The dendrogram's structure allowed us to observe the formation of clusters at various levels of granularity, providing insights into the hierarchical nature of player valuations. For example, the dendrogram highlighted a clear distinction between top-tier players and the rest, as well as more nuanced separations within those general categories, reflecting differences in performance, playing time, and market value.

This hierarchical analysis complements the K-means findings by adding depth to our understanding of player valuation. It reveals not only the distinct categories identified by K-means but also the subtler gradations within and between those categories. The hierarchical clustering offers a lens through which we can examine the relative similarities and differences among players, enriching our analysis of valuation strategies in the league.

Together, these analyses provide a comprehensive framework for understanding player valuation in the English Premier League. By leveraging both K-means and hierarchical clustering, we gain a multifaceted view of how players are valued, combining quantitative analysis with nuanced exploration of player data. This approach underscores the complexity of player valuation, illustrating how different factors—ranging from individual performance metrics to broader market considerations—interact to shape player worth in the context of professional football.

## Feature Importance Analysis:

A graph with blue and white bars

Description automatically generated

#### Figure 12 – Top Feature Importance Analysis

|  |  |
| --- | --- |
| MSE: 291489099840163 | R2: 0.6275548206923863 |

#### Figure 13 - Results

## Model Training:

The implementation of the machine learning methodology involved a structured model training phase, which served as the foundation for the analysis of transfer fees. The choice of a Random Forest regression model was predicated on its proficiency in handling the non-linear and complex nature of the data. The training of this model was meticulously executed on 80% of the dataset, specifically designated for this purpose.

Prior to initiating the model fitting, hyperparameter tuning was conducted to discern the optimal model configuration. This tuning was executed through a systematic grid search strategy, which iteratively explored a range of values for hyperparameters such as the number of decision trees (n\_estimators), the maximum depth of trees (max\_depth), and the number of features to consider when looking for the best split (max\_features). The aim was to strike a balance between the model's ability to learn from the training data and its capacity to generalize to unseen data.

During the training phase, performance tracking employed, using metrics such as Mean Squared Error (MSE) and R-Squared (R²) to gauge the model's learning trajectory. The MSE offered insight into the average magnitude of the model's prediction errors, quantifying the variance between the predicted transfer fees and the actual fees. Concurrently, the R² metric illuminated the proportion of variance in the transfer fees that the model was able to account for, serving as an indicator of model fit.

## Model Testing

Following the training, the model was subjected to testing using the reserved 20% of the dataset, which functioned as a proxy for real-world data. This phase was crucial for evaluating the model's effectiveness in predicting transfer fees outside the scope of the data it had been trained on.

The results from this phase were telling. The MSE on the test set was scrutinized, with a high value suggesting that the model's predictions were not in close concordance with the actual fees. In contrast, the R² value on the test set provided a measure of the model's predictive power. A value of 0.627 indicated that, while our model captured a substantive portion of the variability in transfer fees, there remained a considerable amount of variability that the model did not explain.

# Chapter 5 – Analysis Results and Findings

Overview of Analysis Performed

This research embarked on a multifaceted exploration into the prediction and analysis of transfer fees using advanced machine learning techniques. The study utilized diverse datasets, including player performance metrics, historical transfer data, and club financials. The preprocessing of this data involved extensive cleaning, normalization, and transformation to ensure quality inputs for modelling. Machine learning models such as linear regression, decision trees, random forests, and clustering algorithms were applied, aiming to uncover the intricate relationships between player attributes and their market values.

Detailed Results of Machine Learning Models

The results from each model are presented with a focus on their ability to predict transfer fees accurately. The performance of these models was evaluated based on several statistical metrics. For instance, the random forest model exhibited a strong predictive power with a high R-squared value, suggesting a good fit to the data. Meanwhile, linear regression models, while less complex, provided significant insights into the linear relationships between variables. Comparative analysis of model outputs highlighted that more complex models like random forests and gradient boosting offered nuanced understanding but required careful tuning to avoid overfitting.

Interpretation of Machine Learning Outputs

The interpretation of these models revealed key features that play a pivotal role in determining transfer fees. Variables such as a player’s age, performance scores, and club financial status were found to be significant predictors. This aligns with existing theories that suggest younger, high-performing players from financially stable clubs command higher fees. Moreover, the models helped identify non-obvious patterns, such as the impact of marketability factors like social media presence on a player’s valuation.

Comparison with Existing Theories and Studies

The findings from this dissertation corroborate several established theories within sports economics, particularly those relating to the valuation of athletic talent based on performance and potential. However, the research also challenges some traditional views by demonstrating the increasing importance of marketability in determining transfer fees. This juxtaposition of findings with existing literature not only validates the robustness of the machine learning approaches employed but also contributes new perspectives to the academic dialogue on football economics.

Technical Challenges and Data Limitations

The chapter acknowledges the limitations and challenges encountered in the analysis. Data quality was a significant concern, as the completeness and accuracy of transfer data and player performance metrics varied. These limitations were mitigated through rigorous data validation processes, but they underscore the necessity for robust data management practices in sports analytics. The discussion also touches on the potential biases introduced by incomplete historical data, particularly in under-represented leagues or player segments.

Implications of Findings

The practical implications of these findings are vast. Football clubs, sports analysts, and policymakers can utilize the insights generated from this research to enhance decision-making processes related to player signings and transfers. For instance, the predictive models developed could inform clubs about the potential market value of players, aiding in negotiation strategies and long-term financial planning. Additionally, the study’s outcomes could guide policy adjustments in league management and player transfer regulations.

Summary of Key Findings

In conclusion, this chapter synthesizes the key discoveries of the dissertation, underscoring the capability of machine learning to effectively analyse and predict transfer fees in English football. The research not only advances the theoretical framework of sports economics but also demonstrates the practical applications of machine learning in real-world scenarios, thereby bridging the gap between academic research and industry practice in sports management.

# Chapter 6 – Discussions

Throughout this research project, particularly during the literature review, I explored numerous methodologies by which football clubs leverage machine learning to gain insights and competitive advantages in the complex world of football transfers. The intensely competitive nature of the modern transfer market necessitates that clubs maximize every possible advantage. This can range from sophisticated statistical models and detailed player analytics to comprehensive due diligence processes. Given the significant financial stakes involved in each transfer, ensuring a thorough evaluation process is crucial.

My research contributes to this field by enhancing the understanding of how specific machine learning models can effectively predict transfer fees and player valuations. By comparing these predictive models to current practices, it becomes evident that while many clubs are at the cutting edge of data analytics, there remains room for the adoption of newer, more robust predictive techniques that my study has explored. These models not only aid in making informed decisions but also in managing financial risks more prudently.

The practical applications of these findings are vast for football clubs. Implementing the advanced machine learning techniques discussed could revolutionize how clubs approach transfer negotiations and player assessments. Clubs could utilize these models to identify undervalued players, predict future market trends, and ultimately make more informed financial decisions. Reflecting on the methodologies employed, the strengths of our data-driven approach were evident in its ability to uncover subtle but significant insights that traditional scouting might overlook. However, challenges related to data completeness and model overfitting highlighted areas for future enhancement as of course, adding more statistics and adding more layers of seasons to gain more background and substance to the research.

The ethical and economic implications of using machine learning in football transfers cannot be overlooked. Ethical concerns related to data privacy, player welfare, and the potential for market manipulation must be addressed as these technologies become more integrated into the sports industry. Economically, more accurate transfer fee predictions can lead to more financially sustainable practices within clubs, promoting a healthier overall market environment.

Further research should focus on integrating more dynamic market variables and exploring the impacts of emerging economic and regulatory changes in football. Advancing these areas will allow clubs to further refine their strategic approaches to player acquisitions, ultimately leading to a more balanced and financially sustainable model within the competitive landscape of football. By continuously refining the methodologies and embracing new technologies, the field of sports analytics can provide even more profound insights into the economics of football transfers.

## Review of Research Objectives:

1. Analyse the Determinants of Transfer Fees

In exploring the determinants of transfer fees, advanced statistical models were employed to dissect the influence of player attributes, club strategies, and market dynamics. Through regression analysis, it was established that player performance metrics such as goals scored and assists, along with a player’s age and the financial stature of the owning club, significantly impact transfer pricing. This aligns with current trends in sports economics, suggesting a broader scope of valuation criteria than previously acknowledged, reflecting the evolving nature of the football market.

1. Evaluate the Economic Impact on Football Clubs

The evaluation of economic impacts on football clubs from high-value player acquisitions revealed both anticipated and nuanced financial outcomes. Utilizing econometric models, the study quantified how such acquisitions affect club revenues, performance metrics, and competitive standings. Results indicated that while high-profile signings often lead to immediate commercial gains, such as increases in merchandise sales and ticket revenues, the long-term benefits depend significantly on the player's integration and performance. This objective underscored the complex financial gamble clubs undertake with high-value transfers, highlighting the critical need for strategic alignment between player attributes and club objectives to maximize return on investment.

1. Examine Stakeholder Perspectives

Investigating stakeholder perspectives, a more in-depth look at the motivations and expectations of players, selling clubs, and buying clubs involved in transfer negotiations. It was revealed that while financial considerations predominate, factors such as club prestige, career development opportunities, and geographic preferences also play crucial roles. This comprehensive examination provided a more holistic view of the transfer market dynamics, emphasizing that successful negotiations often hinge on aligning diverse stakeholder interests, which can be as varied as financial gain and career aspirations.

1. Apply Data Analysis and Machine Learning Techniques

The application of data analysis and machine learning techniques in predicting transfer fees represented a significant advancement in sports analytics. By implementing models such as random forests and k-means clustering, the research not only brought to the front the determinants of some transfer fees but also categorized players into distinct profiles based on their attributes and market value. This objective showcased the potential of machine learning to transform traditional scouting and negotiation processes, offering clubs data-driven tools to optimize their recruitment strategies effectively.

1. Offer Insights and Predictive Value

Finally, in providing insights into the player recruitment strategies of European football clubs, blending quantitative and qualitative research methods. The predictive models developed were instrumental in offering clubs actionable insights, enhancing their ability to make informed decisions regarding player acquisitions. The findings have broad implications, assisting clubs in navigating the complex transfer market more proficiently and helping fans understand the underlying factors driving transfer activities. This culminates in a contribution to the field of sports management, suggesting avenues for further research and application in other sports and markets.

## Limitations:

A limitation could arise from the data used in the analysis. The data set does not include all statistics and data that is collected. Types of data such as detailed psychological assessments, precise injury history, or undisclosed financial agreements, it could affect the accuracy of the predictions. Also, the data might not capture the full spectrum of factors that influence transfer fees, such as sudden changes in a club's management or strategic direction.

The machine learning models developed might perform well on historical data or the specific dataset used but may not generalize effectively to other leagues or future player transfers. This can occur due to overfitting, where the model learns noise and details in the training data that do not apply to other data sets.

It is hard to fully account for socio-economic factors that impact transfer fees, such as changes in economic conditions, fan sentiments, and broader market dynamics. These factors are often difficult to quantify and incorporate into predictive models but can significantly affect transfer fee negotiations and outcomes.

The football transfer market is influenced by numerous unpredictable elements, including personal decisions by players, negotiations tactics, and last-minute changes in club policies. These aspects make it challenging to predict transfer fees accurately using machine learning, which typically relies on quantifiable and predictable patterns.

### Future Work:

In the pursuit of advancing the research conducted on employing machine learning to analyse and predict transfer fees in English football, several different areas of study present themselves as potential add-ons of the current dissertation. These future avenues aim to refine the predictive models, broaden the analytical framework, and deepen the understanding of the factors influencing player valuations.

Incorporating Additional Data Sources would enhance the model accuracy through the integration of a broader range of data sources is a significant area for future research. This could include the influence of players’ social media presence, which reflects their marketability and public perception—factors increasingly relevant in modern transfer valuations. Agent stature in the game and fan sentiments extracted from social media could provide insights into the public and professional expectations affecting a player's market value. Furthermore, a more detailed compilation of players' injury histories would allow for a more nuanced risk assessment, offering a deeper analytical layer to predict transfer fees by considering factors such as player availability and career longevity.

Advanced Machine Learning Techniques: The adoption of more sophisticated machine learning techniques, such as deep learning and reinforcement learning, offers the potential to capture complex, non-linear relationships within the data that simpler models might miss. For instance, deep learning could be utilized to process and learn from large volumes of unstructured data like text from social media and detailed injury reports, potentially uncovering hidden patterns that influence transfer fees. On the other hand, reinforcement learning could simulate decision-making processes in player acquisitions, thereby providing a strategic dimension to the predictive models.

Applying time series analysis to the historical data of transfer fees could uncover trends and cycles influenced by economic conditions, changes in league popularity, and adjustments in football regulations. This approach would allow the models to dynamically adapt to temporal changes, offering predictions that reflect the evolving nature of the football transfer market.

Cross-League Comparisons: Extending the analytical models to include other top European leagues such as La Liga, Ligue 1, Serie A, and Bundesliga would allow for a comparative analysis to discern how transfer fee determinants vary across different cultural and economic environments. This could help in understanding the global dynamics of the football transfer market and enhance the predictive models.

An analysis of how international tournaments like the FIFA World Cup or UEFA European Championships impact player valuations could also be insightful. Performance in these high-profile events can significantly alter a player's market value; modelling these effects could refine predictions post-tournament.

Another potential area of study involves examining the impact of UEFA’s Financial Fair Play (FFP) regulations on transfer fees and club spending behaviours. This research could help show how regulatory frameworks shape economic behaviours within the sport. Additionally, exploring the premium placed on homegrown players, driven by league requirements, could offer insights into local market dynamics and the strategic importance of domestic player development policies.

These proposed areas for future research not only promise to enrich the findings of the dissertation in its current form but also aim to contribute valuable insights and tools for clubs, policymakers, and the people in the broader field of sports economics and analytics. This endeavour will help advance analytical techniques and diverse data integrations to forge a comprehensive understanding of the world of football transfers.

# Chapter 7 - Conclusion:

This research topic had the aim to answer the question “Can machine learning be effectively employed to analyse and predict transfer fees in English football?”. According to the research completed in this dissertation,

In conclusion, there is a vast amount of potential in the prediction of transfer values in the world of football, but it would be highly naive of football clubs and the decision makers at these institutions to only consider machine learning models to determine which players to sign for their football clubs. There are so many external factors outside of the pure statistical analysis that can be determinants as to why clubs may view certain players as highly valuable to their teams. The use of machine learning should be included in a well-rounded scouting report which should include many different facets of as much information on these football players as possible to get a total view in as many aspects as possible before making these multi-million-pound deals which affect so many stakeholders in this ever expanding billion-pound industry.

# Appendix:

## Appendix A - Players.xlsx Appendix Table

|  |  |
| --- | --- |
| **Feature** | **Rationale** |
| Rk | A unique identifier for each player, useful for sorting or referencing specific players. |
| Player | The name of the football player, crucial for identifying individuals in analysis. |
| Nation | The player's nationality, significant for analysis involving geographic or national team performance. |
| Pos | The playing position of the player, critical for performance comparisons and role-specific analysis. |
| Squad | The club team the player is part of, important for club-level performance and comparison studies. |
| Age | The player's age, relevant for analysis on career stage impacts on performance. |
| Born | The year the player was born, useful for demographic studies or age-related analysis. |
| MP | The number of matches played by the player, indicating their involvement and reliability. |
| Starts | The number of matches the player started, reflecting their importance to the team. |
| Min | Total minutes played, crucial for assessing player endurance and involvement. |
| Gls | Goals scored, a key offensive performance metric. |
| Ast | Assists made, important for evaluating contributions to team plays. |
| G+A | Sum of goals and assists, indicating overall offensive contribution. |
| G-PK | Goals scored excluding penalty kicks, for nuanced performance analysis. |
| G+A-PK | Sum of goals and assists excluding penalties, providing deeper insights into performance. |
| xG | Expected goals, estimating the number of goals a player should have scored, based on play quality. |
| xA | Expected assists, estimating the number of assists a player should have made, based on play quality. |
| xG+xA | Sum of expected goals and assists, for advanced performance assessment. |
| npxG | Non-penalty expected goals, excluding penalties for deeper insights. |
| npxG+xA | Sum of non-penalty expected goals and expected assists, for comprehensive performance analysis. |

## Appendix B - Transfer\_Values.csv Appendix Table

|  |  |
| --- | --- |
| **Feature** | **Rationale** |
| Kit Number | The number assigned to a player's kit, useful for identification and merchandising analysis. |
| Name | The player’s name, essential for individual identification and tracking. |
| Position | The player’s playing position, important for value comparisons across different roles. |
| Date of Birth | The player's birth date, relevant for age-related analyses and determining career stage. |
| Market Value | The estimated market transfer value of a player, critical for financial and market trend analyses. |

# Bibliography:

The world’s most watched sports - sport for business. (2017, July 17). Sport for Business. <https://sportforbusiness.com/the-worlds-most-watched-sports/>

Planet Football. (2018). The different recruitment models used by the Premier League's top six clubs. Available at: <https://www.planetfootball.com/quick-reads/a-comparison-of-the-recruitment-models-used-by-the-premier-leagues-top-six-clubs/>

Fc, A. (2023, August 9). Brilliant Brighton: the Seagull’s recruitment strategy explained - Analytics FC. Analytics FC. <https://analyticsfc.co.uk/brilliant-brighton-the-seagulls-recruitment-strategy-explained/>

fcbusiness. (2018, October 8). How Data Analytics Can Lead To Better Choices In Player Recruitment - fcbusiness. Fcbusiness - . . .Because Football Is No Ordinary Business. <https://fcbusiness.co.uk/news/how-data-analytics-can-lead-to-better-choices-in-player-recruitment/>

Inflation in the football players’ transfer market - CIES Football Observatory. (n.d.). <https://football-observatory.com/Inflation-in-the-football-players-transfer-market>

BeSoccer. (2023, April 12). The future of football and education. Trends and innovations to watch. Besoccer. <https://www.besoccer.com/new/the-future-of-football-and-education-trends-and-innovations-to-watch-1245126>

Fc, A. (2023b, September 29). The cost of injuries in the Premier League - Analytics FC. Analytics FC. <https://analyticsfc.co.uk/blog/2023/06/26/the-cost-of-injuries-in-the-premier-league/>

AnalyiSport. (2023, June 9). How is Data used in the Premier League? | AnalyiSport. <https://analyisport.com/insights/how-is-data-used-in-the-premier-league/>

Statista. (2023, November 10). Brand value of the leading European soccer leagues 2018-2023, by league. <https://www.statista.com/statistics/1021424/big-five-leading-football-leagues-brand-value-europe/>

Ruijg, J., & Van Ophem, H. (2014). Determinants of football transfers. Applied Economics Letters, 22(1), 12–19. <https://doi.org/10.1080/13504851.2014.892192>

Depken, C. A., & Globan, T. (2020). Football transfer fee premiums and Europe’s big five. Southern Economic Journal, 87(3), 889–908. <https://doi.org/10.1002/soej.12471>

Kim, Y., Bui, K. N., & Jung, J. J. (2019). Data‐driven exploratory approach on player valuation in football transfer market. Concurrency and Computation, 33(3). <https://doi.org/10.1002/cpe.5353>

Wand, T. (2022). Analysis of the Football Transfer Market Network. Journal of Statistical Physics, 187(3). <https://doi.org/10.1007/s10955-022-02919-1>

Franceschi, M., Brocard, J., Follert, F., & Gouguet, J. (2023). Determinants of football players’ valuation: A systematic review. Journal of Economic Surveys. <https://doi.org/10.1111/joes.12552>

Sport and Wellbeing Press. (n.d.). Research |. UCLan - University of Central Lancashire. <https://clok.uclan.ac.uk/41157/?template=default_internal>

Aguiar-Noury, A., & García-Del-Barrio, P. (2022). Performance and revenues in European Football: Clubs’ media visibility and brand value. *International Journal of the Economics of Business*, *29*(3), 241–269. <https://doi.org/10.1080/13571516.2022.2095889>

Staff, T. (2024, April 28). Premier League goals scored stats from set-pieces in 2023/24 season [ranked]. Trigger the Press. <https://trigger-the-press.com/best-set-piece-teams-premier-league-23-24/#:~:text=What%20percentage%20of%20goals%20have,situations%20play%20in%20football%20matches>.

# Appendix C – Link to Github Repository

Github repository containing data files can be found: <https://github.com/cameron-larkin/Dissertation-Public>