A Work Project, presented as part of the requirements for the award of a master's degree in Business Analytics and Finance from the Nova School of Business and Economics.

The effect of injuries on football players' market values – The role of recurrent injuries in young players

Work project carried out under the supervision of:

Pedro Brinca

Abstract

Football players' market values are intensively researched as the focus gradually shifts from qualitative approaches and subjective perception to data-driven estimation. Previous work drew on player characteristics and top-level performance metrics but only few included injuries. This thesis resorts to data from Opta and Transfermarkt and leverages MLR and machine learning approaches to derive market values.

Our LightGBM and MLR models predict market values accurately and identify direct impacts of injuries. Additionally, we found that recurring injuries affect especially young athletes.

Keywords: Football, Market Values, Injuries, Machine Learning, LightGBM, MLR, Opta, Transfermarkt

ACKNOWLEDGEMENTS:

We would like to express our gratitude to our thesis supervisor, Pedro Brinca, and our thesis partners Dr. Karsten Zimmerman and Jan-Philipp Erdmann from Hamburger Sport-Verein.

1. Introduction

There is no denying an increasing economization in football (Littkemann and Kleist 2002). More and more athletes can make a living from their sports career, as TV broadcasting rights sell for billions of Euros (Carreras-Simó and García Villar 2018), and families or state funds support clubs' transfer activities with hundreds of millions (Kay 2022). This influx of money historically made transfer fees in football increase, as *Table 1* shows (SPOX 2020); the same goes for salaries (Frick 2007). Clubs sign players on long-term contracts more often than not for strategic purposes, as it helps them to justify a higher asking price in transfer negotiations with another club (SID 2011).

Decade	Name	Transfer Fee	Selling Club	Buying Club
1980s	Andy Gray	€3m	Aston Villa	Wolverhampton Wanderers
1990s	Aldair	€9m	Benfica Lisbon	AS Rome
2000s	Luis Figo	€60m	FC Barcelona	Real Madrid
2010s	Cristiano Ronaldo	€94m	Manchester United	Real Madrid
2020s	Neymar	€222m	FC Barcelona	Paris Saint-Germain

Table 1: Highest player transfer fees to date per decade

But more money on the table also means higher risks. Clubs can become highly levered institutions, ultimately risking their entire existence if they default on their debt. The case of FC Barcelona lately got lots of attention in that regard (Crafton and Ballus 2022; Doyle 2022). Fair competition is another much-discussed point. Not all clubs are compensated equally when their matches are shown on TV, nor are they all backed by investors. Higher broadcasting fees and the change in payouts from European cup competitions lead to higher revenues for domestically and internationally successful clubs, decreasing competitive balance in national leagues (Carreras-Simó and García Villar 2018; Pawlowski, Breuer, and Hovemann 2010). According to the authors, inequality will increase over time. This makes sense, as researchers have shown that sporting success impacts clubs' revenues positively (Szymanski and Smith

1997; Dobson and Goddard 1998), and investments highly influence national success in a club's squad (Szymanski and Smith 1997). Rohde and Breuer confirm this virtuous/vicious cycle, that money yields success, and success results in more money available to invest in the team (Rohde and Breuer 2016).

This has critical implications for smaller, less wealthy teams. They must be successful to stick around, but they cannot afford to outbid affluent clubs and possibly overpay on transfer fees and player wages. If a small club takes the risk and signs a new player, that player needs to perform well after that. Injuries are a possible obstacle to this, since a sidelined, injured player cannot help his team win. In addition, long-term performance degradation from more severe injuries might prevent a player from realizing his full potential. This suggests that a player's market value, which Kirschstein and Liebscher call a reflection of the financial value a club assigns to a player's performance (Kirschstein and Liebscher 2019) and approximately matches actual transfer fees paid (Herm, Callsen-Bracker and Kreis 2013), should be impacted by injuries. In recent years, much work has been done on parameters that influence market values. To our knowledge, there has been limited research on how injuries are reflected in player valuation changes. With this thesis, we want to contribute to further elaborate this area of research.

2. Finance Rationale

2.1 What are market values?

In professional football, the players are of paramount importance to a club as they are considered not only as strategic assets that ensure the successful performance of a team but also as business investments that drive value for the organization. Unless a player goes through the ranks of the club's youth academy or is available as a free agent, the only other way of acquiring an athlete is to buy them, or to be precise, the multiannual rights to exploit their performance

(Maglio and Rey 2017), from another club. Upon agreement from the buyer, seller, and player, the latter is transferred for a monetary compensation called the transfer fee. This fee must not be confused with a player’s market value, as both differ among a range of criteria. One manifestation is the remaining contract length, which significantly affects transfer fees, but not so much the market values (Trequattrini, Lombardi, and Nappo 2012). At the end of a contract, a player can switch clubs without a transfer fee being paid, but that does not make him worthless; he still holds value. Instead, the transfer fee paid for a player with an expiring contract reflects the opportunity cost to the selling club. The club must evaluate how significant the economic damage of an early departure is compared to the compensation the buyer is willing to pay. Furthermore, aspects such as the probability of a contract extension and the associated additional costs or the acquisition costs for a replacement player are also included in assessing the opportunity costs. Transfers are not that different than other business transactions. Clubs can, for example, undertake risky investments in young players at higher premia, with the desire to develop them and scoop out their potential, similar to a pharmaceutical company buying a biotech startup on the verge of a breakthrough in drug development. On the other hand, signing a more experienced player with a proven track record at a lower premium can be compared to a real estate company buying a seventh apartment house in a residential area where it already owns six.

Football clubs do not book transfer fees in full in the year a player is purchased but write them off evenly over the term of the player's contract (PriceWaterhouseCoopers 2018). If, for example, a club buys the rights for a player for a total cost of € 5m from another club, the buying club will see an increase of € 5m in intangible assets and a decrease of € 5m in cash on its balance sheet. Assuming a 5-year contract, the club will record an annual amortization expense of € 1m for the player in its income statement. At the end of the fifth year, the intangible asset's carrying amount is fully amortized and, therefore, equal to zero. Amortization is a non-cash expense and is thus added back during the cash flow calculation. On the other hand, proceeds from the sale of players are recognized immediately. Suppose a simple case where the player was sold after four years for € 3m. In that case, the club could book a € 2m profit in its income statement, the same increase in operating cash flow, and remove the intangible asset from the balance sheet, making cash go up by € 3m.

But how does the market value of a player fit into all of this? Multiple researchers tried to give a definition. Herm, Callsen-Bracker, and Kreis call it the hypothetical estimate that denotes the amount of money a club is willing to pay to sign an athlete, independent of an actual transaction (Herm, Callsen-Bracker, and Kreis 2014). As such, market values help set a common ground for transfer negotiations. However, the scope of use cases where the market value is used as a benchmark exceeds the practice of negotiating player transfers. Another definition is that market values are expert estimations of the performances and marketability of a player derived from past indicators (Ackermann and Follert 2018). First, players and clubs can use the values as proxies of how well someone plays. Secondly, external stakeholders, players, and clubs can derive fair prices for marketing campaigns and sponsorship contracts from market values (Frick 2001). Moreover, market values are solid predictors of league outcomes (Gerhards and Mutz 2017). The authors show that the larger the degree of financial inequality in a league is, the more market values become the single most important predictor of team performance and, therefore, final league standings.

A well-known source for data regarding market values in the football world is the German company Transfermarkt. In 2000, Matthias Seidel collected football data and created a website in his free time. From that state, Transfermarkt developed into employing more than 100 people, being available in 10 languages, and having 20 dedicated country-specific websites. Since 2008, it has been majority-owned by Axel Springer SE (Transfermarkt 2022d). Over the past decades, it has become the leading football-related database with publicly available figures of over 1.5m matches and 760k player profiles (Wheatley 2021). Additionally, the website has forums for its more than 680k members to discuss players' current market values (Bonacchi et al. 2021). They are updated at least twice a year for major leagues.

The process of updating is a form of crowdsourcing, intending to obtain an accurate valuation range that lets individual misjudgments disappear in the mass. Four criteria should hold for a crowd to make wise decisions: members need to be independent of one another, diverse regarding the four dimensions proposed by Gardenswartz & Rowe (Gardenswartz and Rowe 1994), knowledgeable in their domain, and finally motivated to provide accurate assessments (Simmons et al. 2011). For Transfermarkt, we regard these as fulfilled. Other studies comparing individual and group decisions found that groups get more rational results than individuals (Charness and Sutter 2012). This makes sense as groups, to come to a decision, often must compromise so that extreme opinions are filtered out. Moreover, crowds performed well at information aggregation (Wolfers and Zitzewitz 2004). Market values are many variables aggregated into one, so the Transfermarkt community should be good at its predictions.

Ultimately, decisions about values on the platform are not made on a democratic basis, choosing the value proposed most often. Instead, a few users with lots of knowledge and experience are free to either follow the majority's opinion or their view on a case-by-case basis. For that reason, Herm, Callsen-Bracker, and Kreis named them 'judges' (Herm, Callsen-Bracker, and Kreis 2014). However, this methodology can be criticized as opaque and not replicable, as judges base their decisions on no objective set of criteria. Nevertheless, market values on Transfermarkt were found to have a 93% correlation with realized transfer fees (Gerhards, Mutz, and Wagner 2014), confirming the validity of the applied process. Moreover, clubs regularly state the platform's figures in their annual reports, courts accept them as evidence (Keppel and Claessens 2020), and many researchers resort to them when examining market values (Franck and Nuesch 2012; He, Cachucho, and Knobbe 2015; Müller, Simons, and Weinmann 2017).

2.2 Which factors influence market values?

Usually, to value a good or bad project, one should consider all associated future costs and benefits and then discount them at an appropriate rate to the desired point in time. However, applying this methodology to football players would not make much sense because most costs and benefits cannot be measured as directly as product sales and costs, for example. That is why a different approach is needed. As stated above, market values are an aggregation of multiple factors. Performance on the pitch comes to mind first when thinking about a player's impact on a club. Future performances are of utmost interest here, and researchers confirmed that they could be accurately predicted via past ones (Hendricks, DeBrock, and Koenker 2003). As such, performance data helps to reduce uncertainty in the selection process of human capital. Just like students' university grades serve as signals for employers, clubs can look at past matches of a player to decide if they want to try to sign someone and how much they are willing to pay for him. Moreover, both Ziebs and Partosch (Ziebs 2004; Partosch 2013) found that next to performance, a club's achievements in (international) competitions and the resulting prestige and attention drive market values. Furthermore, far more people want to see superstars like Lionel Messi and Cristiano Ronaldo play than, e.g., watch the players in the German third division. They all play football, but on different levels. Superstars sell more jerseys, make more fans come to the stadiums, are more attractive faces for marketing campaigns for the club and partnering brands, and so on. They have a more extensive crowd-pulling power, so their popularity should also be incorporated into their market value, as clubs consequentially generate more enormous revenues through them. In addition, other factors like age, contract length, and general market conditions and trends impact player values.

Scientific literature often assesses market values based on the category's player and team performance, player and team presentation, and other player characteristics (Müller, Simons, and Weinmann 2017a). Please refer to ***Error! Reference source not found.*** in the appendix for an overview of all indicators Transfermarkt considers.

2.3 Difficulties with the valuation

Nevertheless, the valuation of football players is not straightforward. Players are not frequently traded like stocks, for example. There are only two periods, lasting a few weeks, where transfers can happen each year. Usually, players stay at one club for at least a few years. If someone switches clubs, not all (parts of the) fees will always be disclosed (Press Association 2022). There is no order book, so to speak, which makes it challenging to validate market values. In addition, there is no definite set of criteria players could be assessed on. Attacking players could be judged by the goals they score, but not everybody on a football team can be the one who scores. Another player, potentially a midfielder, first must create the opportunity for someone to shoot, either through a good pass or by distracting opposing defenders and creating room for the striker. A defender's main task could be described as preventing attackers from scoring, but only looking at the main tasks would be too one-dimensional. In modern-day football, players must participate in every game action, regardless of their position. Strikers pressure the opposing goalkeepers and defenders, helping their team to regain possession of the ball as quickly as possible. In contrast, defenders may act as supporting strikers in the closing minutes of the game (CCyler and CSmith1919 2022). And that is not all there is to it: While some players, like Mario Gómez, only play in one position during their entire career, others, like James Milner, are more versatile (Transfermarkt 2022f; 2022g). *Figure 1* confirms this.

Figure 1: Matches by position for Mario Gómez (l.) and James Milner (r.)

Moreover, players can be required to fulfill different roles while playing for the same team, under the same coach, and in very closely related positions. Angeliño and Nordi Mukiele played as fullbacks, a defensive position, for RB Leipzig in the 2020/21 and 2021/22 seasons. However, as *Table 2* shows, Angeliño contributed twice as often to his team's scoring as Mukiele did, meaning he was more integrated into the offensive play (Transfermarkt 2022e; 2022h). On another note, at 1.71m, he is 16cm shorter than 1.87m tall Mukiele, who will be considerably better at heading, an essential part of defensive play. To conclude, we can say that football is much less standardized than, e.g., baseball, where the 'Moneyball' idea originates from (MacLennan 2005). Every position in the latter has clearly defined tasks, i.e., a pitcher will never need to catch a ball. This makes the sport a lot easier to analyze.

Player	Minutes	Goals	Assists	Total contributions	Minutes per contribution
Angeliño	6809	11	24	35	195
Nordi Mukiele	5083	6	7	13	391

Table 2: Statistics for Angeliño and Nordi Mukiele in the 2020/21 and 2021/22 seasons

Objective market values can only exist when dealing with homogeneous goods without information asymmetries. We already clarified that football players are heterogeneous, like employees in other fields. Moreover, there will always be some uncertainty – if not about performances, then potentially about the player's fitness and motivation. Researchers proved the importance of motivation, discovering that highly achievement-oriented football players have a better chance of achieving an outstanding career (Zuber, Zibung, and Conzelmann 2015) and that rowers with a high drive come back stronger after weak performances (Schmid, Conzelmann, and Zuber 2020).

So, market values in football are subjective, as in other judgment-based markets, like wine, fashion, or art (Hutter 2011). Values can be off, either if the people judging lack expertise or if they try to manipulate them. In football, agents and players have been found to rig market values (Keppel and Claessens 2020). Furthermore, researchers argue that transfer values do not represent a fair human capital value. Instead, values are distorted by information asymmetries, whether a player has a professional agent, fees for this agent, potential synergies, economic conditions of the leagues, and the bargaining capacity between clubs (Oprean and Oprisor 2014; Martín, López, and Santín 2019).

In samples studied by Dimitropoulos and others, transfer fees exceeded market valuations in 85% of the cases. This spending behavior will put the financial health of most clubs to the test, as deficits lead to them taking on more and more debt (Dimitropoulos and Koumanakos 2015; Dimitropoulos, Leventis, and Dedoulis 2016). UEFA has installed a set of rules called Financial Fair Play (FFP), which aims to limit the club's spending to not more than they earn to ensure their long-term survival. If they do not comply, they get penalized (AFP 2022). This signifies that it is essential for clubs not to overpay and instead get transfer fees right, avoiding financial distress, improving the ability to generate funds to reinvest in the team, and ultimately sparing the club's fans from worry (Pantuso and Hvattum 2021).

3. Injuries in Football

Injuries can have severe consequences for football players. While someone in an office job might still be able to work with a muscle injury in their non-dominant arm or a broken ankle, injuries disrupt professional athletes in their daily business. In the following, we will discuss the negative effects of injuries on footballers.

First, injuries mean physical and psychological stress for a player. Moreover, a coach has fewer players from his team available for a match, reducing the options for the best possible line-up in pursuit of a win. Especially when star players are sidelined, a team may perform worse, making matches lose attractiveness, resulting in fewer fans attending and, therefore, a club's income going down. As results are tied to price money, both in domestic and international competitions, losing important games likewise leads to a decrease in income. Not only does it take time for players to recover, but rehabilitation also costs money. Clubs prefer their players to be available again as fast as possible, so they resort to costly measures such as having diagnostic devices, doctors, and physiotherapists on-site around the clock (FC Bayern München 2019) and sophisticated rehabilitation actions.

3.1 Injuries and Performance

According to Parry and Drust, injuries are the main factor that keeps football players away from their work. 49% of absences from games and 60% from training are due to injuries (Parry and Drust 2006). In a study by Hawkins and Fuller, an average of three to four players per team were injured in a squad of 25 players. In that study, between 86 and 100% of players suffered an injury per season, either in practice or in a game, leaving them temporarily unable to play. The authors report an average rate of 8.5 injuries per 1,000 player-hours of practice and games (Hawkins and Fuller 1999). A more recent study found a rate of 8.1 injuries per 1,000 hours of exposure, with the rate in matches (36 injuries/1,000h) being almost 10 times higher than in training (3.7 injuries/1,000h) (López-Valenciano et al. 2020). Another study reports that players suffer an average of ten to 35 injuries per year for every 1,000 hours played. This makes the injury risk factor more than 1,000 times higher than in other high-risk industries (Dvorak and Junge 2000; Junge and Dvorak 2004). A fifth study monitoring 1,743 players from 27 teams in 10 countries between 2001 and 2012 found an average of 50 injuries per team per season, lasting a combined 881 days. This results in an average downtime of 17.6 days per injury (Jan Ekstrand et al. 2013).

3.2 Causes of Injuries

Because football is a contact sport that demands a lot of players physically, for example, fast runs and changes of direction, jumps, and tackles, players are susceptible to injuries from direct contact with an opponent (e.g., bruises) but also without (e.g., strains). Players are most often injured during games, with the worst injuries occurring through contact and collisions with others (Peterson et al. 2000). Injuries in training occur most frequently in July, peaking in games in August (Hawkins et al. 2001). This is probably because the players' fitness level over the summer break is not the same as during the season, and thus the risk of injury increases. A player's position also affects his injury frequency (Hunt and Fulford 1990). Since attackers are fouled more often and fouls cause injury, they are injured more than defenders (Chomiak et al. 2000). Psychological factors can increase players' susceptibility to injury and explain approximately 15% of injuries (Junge 2000; Ivarsson and Johnson 2010). An injury can lead to subsequent injuries occurring more frequently in the same area (Chomiak et al. 2000; Arni Arnason et al. 2004). In addition, if an injury cannot heal completely, the regeneration time is prolonged in the case of secondary injuries. Even though recurring injuries generally occur less frequently than new ones (1.3 and 7.0 injuries/1000 hours of exposure) (López-Valenciano et al. 2020), they can still be of paramount impact.

3.3 Types of Injuries

Most injuries in football are traumas, 29% of which are caused by fouls (Hawkins and Fuller 1996). Between 9 and 34% are caused by excessive stress (Nielsen and Yde 1989; A Arnason et al. 1996). In outfield players, injuries occur most frequently to the lower extremities, with 6.8 injuries per 1,000 hours of exposure. The torso is the second most affected but with only 0.4 injuries per 1,000h. Especially endangered in the lower body are thighs (1.8 injuries/1,000h), knees (1.2 injuries/1,000h), and ankles (1.1 injuries/1,000h). Muscles and tendons are the most frequently injured (4.6 injuries/1,000h), followed by contusions (1.4 injuries/1,000h), joints and ligaments (0.4 injuries/1,000h), and fractures and stress injuries (0.2 injuries/1,000h) (López-Valenciano et al. 2020). Conversely, goalkeepers are more often affected in the upper extremities (Dvorak and Junge 2000). Players are most likely to miss 1-3 days after an injury (3.1 injuries/1,000h), followed by 8-28 days of absence (2.0 injuries/1,000h), then 4-7 days (1.7 injuries/1,000h), and least frequent more than 28 days (0.8 injuries/1,000h) (López-Valenciano et al. 2020). The UEFA standard classifies injuries as ‘transient’ when the recovery time is shorter than seven days, ‘mild’ (< 28 days), ‘moderate’ (< 84 days) and ‘severe’ (>= 84 days) (Kampakis 2016).

3.4 Problems with Injury Definitions

Both researchers and physicians may use different terms and definitions for the same injury (Junge and Dvorak 2000). This is primarily due to different training and their own experiences. For example, some studies use the term 'injury' as soon as a player has required medical attention (Hawkins and Fuller 1999; Hawkins et al. 2001; Andersen et al. 2004; Arni Arnason et al. 2004), while others use it only after a player has missed a game or practice (Hawkins and Fuller 1996; C W Fuller et al. 2004; A Junge, Dvorak, and Graf-Baumann 2004). Still, others call it an injury when a player has only trained with reduced load because of, e.g., a tissue injury (Peterson et al. 2000; Junge et al. 2004), and some also mix all these definitions.

Furthermore, there is no testing regimen or definition of what a full recovery requires or means. Players are often considered recovered as soon as they are cleared to return to training and play. However, this ignores misdiagnoses or the pressure on medical staff to make players available for important matches (Hägglund, Waldén, and Ekstrand 2003). Toni Kroos, for example, confessed that he did not take a break from playing football despite pubic bone inflammation but took painkillers for six months to be able to play anyway (Cortegana 2021).

3.5 Implicit Costs of Injuries

Drawer and Fuller confirmed that injuries have a negative impact on team performance. In addition, they note that small, financially constrained clubs can face major problems when key players get injured, possibly leading to the club being relegated. Larger clubs are rather capable to compensate for injuries due to their often-high-quality squads but still suffer losses in prize money if team performance deteriorates (Drawer and Fuller 2002). In Spain, players from 27 first and second-division teams missed 40,306 days, or an average of 15.8% of the 2008/09 season due to injuries, costing clubs € 188m (Fernández Cuevas et al. 2010). Catapult Sports conducted a similar study for the English Premier League in the 2018/19 season (Catapult Sports 2022). Players suffered 804 injuries that lasted a total of 18,230 days. As a result, they missed an average of 3 league games (8% of the season). The 20 clubs paid £ 166m to injured players, 14% of the total salary expenditure for players. With an average conversion factor of 1.13416 for the period, this equates to € 188m likewise. A single injury costs an average of £ 200k. Wolverhampton Wanderers had the lowest cost at £ 680k and Manchester City the highest at £ 23m. For Wolves, that meant 2.4% of their total salary budget, while Manchester City paid 20% of its budget to players unavailable through injury. The most expensive injury was that of Alexis Sánchez, who missed 128 days, or 15 games, while receiving £ 6.4m in wages during this time. There was clearly a correlation between the financial strength of clubs and the amount paid to injured players. Eliakim et al. found a statistically significant relationship between days missed by players on a team due to injury and a lower league placing. 136 days injured meant the loss of a point in the Premier League, and 271 meant the loss of a place in the Premier League table (Eliakim et al. 2020). Clubs averaged 58 injuries and 1,410 injured days, a loss of six spots in the final standings. Injuries cost a team an average of £ 45m: £ 36m due to missing out on prize money because of poorer performances, and the remaining £ 9m as salaries paid to injured players. This equates to £ 181m for salaries across all teams in the 2016/17 season. With the average conversion factor of 1.16290, this parallels to € 210m. The calculation does not include losses from the players' market values decline. However, the authors assume that these are significant.

If injuries significantly impact players' market values, an impairment test should be performed in accordance with IAS 36 and 38. UEFA has developed its own rules based on these standards that prescribe the procedure. Maglio and Rey specifically state that if a footballer, for example, suddenly ends his career due to an injury or other personal decision, his club must claim an impairment loss on the underlying intangible asset - the performance rights - on the income statement, as the club will no longer generate revenue from the player in the future (Maglio and Rey 2017). The carrying amount of the performance rights on the balance sheet will be reduced to the recoverable amount, or in case of a career end, to zero. Since impairments, like depreciation and amortization, are non-cash expenses, the impairment loss must be added back to the cash flow statement when reconciling the operating result to the cash generated from operating activities.

3.6 Injury Prevention

Football clubs are investing more in injury prevention. The focus is primarily on technology and improved return-to-play protocols (Rossi et al. 2018). Researchers are trying to delve deeper into the emergence of injuries and thus reduce their frequency (van Dyk et al. 2017; Lundgårdh, Svensson, and Alricsson 2020), but no breakthroughs have been made in this area

(Eliakim et al. 2020). Injury rates for ligament injuries decreased between 2001 and 2012, but injury rates for practice and games, as well as muscle injuries and serious injuries, remained high (Jan Ekstrand et al. 2013). Thus, the overall incidence of football injuries has not changed over the past 20 years despite massive investment and more research (Hawkins and Fuller 1999; Dai et al. 2014; Bjørneboe, Bahr, and Andersen 2014; Jan Ekstrand, Waldén, and Hägglund 2016). In other fields of employment, it has fallen between 50 and 60% in the same period (HSE 2022).

A program developed to prevent anterior cruciate ligament (ACL) tears requires players to exercise a week thrice for 20 minutes. It reduces the occurrence of cruciate ligament tears by 87% (Caraffa et al. 1996). The Nordic hamstring exercise can reduce hamstring injuries by up to 50%, but only 10% of Champions League participants resort to the exercise (Bahr, Thorborg, and Ekstrand 2015). Falling victim to the same pattern, it seems, is the 'FIFA 11+' prevention program, which has been able to reduce the injury burden among soccer players by between 20% and 50% (al Attar et al. 2016; Sadigursky et al. 2017). However, only 10% of national associations recommend the program (Bizzini and Dvorak 2015). Overall, 83% of all clubs under the UEFA umbrella do not follow evidence-based prevention programs (Bahr, Thorborg, and Ekstrand 2015). This is noteworthy, especially against the background that age and regeneration time are correlated. With increasing age, the regeneration time also increases since cells in the body need more time to renew themselves (Cloke et al. 2012). But clubs probably do not want to be without their experienced players for longer than necessary, which is why these results raise the question of why the clubs act the way they do.

The previous paragraph implies that prevention programs are either not implemented widely enough or do not work sufficiently well. Changes to programs are often made without attention to whether the desired effect is occurring or whether a player's time to invest in prevention is worth it. Investing that time is the biggest hurdle, with the biggest benefit being the increased availability that comes from reducing injuries. Therefore, the focus should be on developing prevention programs that have a good return on invested time. If programs can be integrated into normal training routines and improve both fitness and player performance, clubs and players will have sufficient incentive to implement them. For maximum efficiency in relation to the resources used, attention should be paid to players at higher risk of injury. Accordingly, screening procedures should be used to identify such players (Fuller and Hawkins 1997; Fuller, Ojelade, and Taylor 2007; Fuller 2019).

4. Machine Learning (ML) Rationale

4.1 What is Machine Learning?

The field of machine learning (ML) is a segment within the domain of artificial intelligence (AI). It is commonly understood as a machine's capability to mimic human decision-making behavior. Machines, in this context, refer to computer programs and technologies that use mathematics and statistics to absolve tasks. The field of machine learning specifically aims to give computers the ability to learn without explicitly having to be programmed (Brown 2021). Russel defines it more precisely as "machines thinking and/or acting humanly and/or rationally" (Russell 2010). However, creating a meaningful ML algorithm is a challenge that starts with what is often referred to as modern-day gold: data. The availability and quality of data are the basis for developing a purposeful algorithm. Once the data is collected and transformed into the right structure, it is used to train a model. There is a direct correlation between the amount of available data, the performance of the program, and the quality of its outcomes. A model is essentially a set of rules that are based on the so called 'training data', which is a portion of the original data set that is used for model fitting. There are descriptive models, which group and interpret data into new insightful structures, and there are models of predictive nature, which use an input to generate an output. Such predictive models can either classify the inputs into categories (*classification models*) or predict a continuous numerical value (*regression models*). Models are valuable, as they are not only efficient and cost-conscious when absolving tasks, but they are also able to process information at a scale which is beyond human capabilities. Machine learning models are therefore gaining more and more significance and influence, not only in the professional working world, but also in most of our day-to-day experiences.

4.2 Use cases and applications in professional sports & football

The range of applications of AI - an umbrella term for Machine Learning – is limitless. It is already widely incorporated in our daily activities, ranging from fraud detection in financial transactions to self-driving cars. This surge of areas of applications was driven by the convergence of advances in three main areas: With the evolution of technologies came an abundance of data. Additionally, computers have experienced and exponentially increase in their processing power over the last years, enabling users to perform much more complex computations. Thirdly, newly emerging algorithms like neural networks and deep learning have contributed to broadening the range of thinkable computations (Malone, Rus und Laubacher 2020). While the beginnings of AI date back to Alan Touring's first attempts to mimic human communication through a machine in the 1950's, the real breakthrough came in the early 2010's, with self-learning systems within robot applications, smart hubs and intelligent data analytics (Jaakkola et al. 2019). However, even though this trend made its advances rapidly, the sports- and specifically the football industry remained largely untouched for the past decades. Only recently data-driven applications have become more ubiquitous to association football. Historically, the football industry has always been an isolated environment. Despite football being the most popular sport in the world, only a tiny fraction of the world's population achieves their dream of making it into the elite circle of pro players. The same holds true for all the other stakeholders in the sports industry: from managers to agents, people in the business operate like in a long-established industrial cartel (Szymanski and Smith 1997). However, over the past decade this highly competitive realm has transformed itself into a commercialized business and hyper-professional industry, where no stone is left unturned to pursue marginal gains. Along with the increasing commercialization of association football (Littkemann and Kleist 2002) the mindset of the ever-growing range of stakeholders changed dramatically. Sporting organizations and managers started incorporating and applying business principles to the previously rigid structures of a football club. Robinson points out an observation regarding the commercialization of the sport: organizations – in this case football clubs – shifted their focus to maximizing revenue and using this goal to justify strategic and financial decision making. This cultural and organizational change is, on the one hand, highly condemned, as critics argue that it challenges the core of football. On the other hand, this commercialization was also a crucial factor in developing the sport as a business, which finally resulted in a global industry that is a driving force for boosting local and international economies by creating, for example, new forms of employment or tourism (Robinson 2008). Leading football clubs realized this early and understood that a healthy operational strategy with profitable financial statements would then in turn help them to improve their chances of sporting success. So, with industries outside of the sports sphere evolving and pioneering to find means to incorporate new technologies for a more efficient business practice, it is only logical that such data-centric advancements found their way into the football world. Such phenomenon could already be observed as early as 2002, in the most prominent case, the previously mentioned 'Moneyball' story.

Former head coach, of the baseball team Oakland A's, Billy Beane, was tasked with forming a team of baseball players on a limited budget. He conducted data mining on a multitude of available players and focused on statistics that were highly predictive of how many runs a player would score. These statistics weren't necessarily figures that baseball scouts would traditionally look out for (UW Data Science 2016). The revolutionary aspect was not the fact that he used data analytics on baseball players, but that he used the resulting insights to run his business, resulting in 20-game winning streak. Other cases soon followed, creating the term sports analytics, which is used to describe the area that leverages data to quantify fields like athletic performance and business health to streamline the operations and results of sports organizations (Schroer 2022). Within sports analytics, there are two main areas of application, namely on- and off-field. The Moneyball case falls into the latter, as the use of data analytics was applied to operations that did not occur during the game of baseball.

Over the last decades, advancements in the use of data and analytics methods have found their way into the world of football. The use of technology during a game has always been a highly controversial topic, as critics argue it reduces the essence of the game (Beiderbeck et al. 2023). Nevertheless, it is still in everyone’s interest to make the game as fair as possible for all parties involved and support human decision-making. The most prominent example is arguably the 1966 World Cup Final, where England and West Germany met. With the game tied after 90 minutes, the two opposing teams went into extra time, where a shot from England striker Geoff Hurst hit the crossbar and then bounced on the line to be cleared away after. Upon deliberation, the referee decided to award England the goal, a decision which triggered a decade-long controversy, as England went on to win the match (Goldmann and Hesselmann 2016). Over four decades later, the two nations met on the same stage, the 2010 World Cup quarterfinals. This time, the tables turned, when England midfielder Frank Lampard struck the crossbar and the ball visibly bounced back out of goal, after having crossed the line. The goal was not given, and Germany went through to the semifinals. This sparked another round of debate, upon which FIFA decided to implement the Goal Line Technology (GLT), a system that is acknowledged as one of the best innovations in football in its recent history. The technology consists of seven high-speed cameras on each side of the pitch, which detect the ball when it is within close periphery of the penalty box. A computer processes and analyzes the video images in real time and creates a 3D model of the goal and the ball. The referee then receives a signal if the ball has completely crossed the line (FIFA 2018a). This innovation opened the floodgates for other uses in technology that are being frequently tested, such as the Video Assistant Referee (VAR) or the Virtual Offside Line system that is currently being implemented at the 2022 World Cup.

To name a final example of how important data has become in football, we want to highlight the case of Kevin De Bruyne, playing for Manchester City FC and the Belgian national team. As previously discussed, football players traditionally employ agents to take care of the business aspects and communication with club officials. For his contract renewal however, De Bruyne decided to employ the UK based company Analytics FC. With their expertise in data science and available data resources in the world of football, they helped De Bruyne to gain a better understanding of how he adds value to the team by quantifying his on-field contributions and benchmarking them against other players and their respective salaries (Worville 2021). As a result, he managed to justify his salary demands by using data-driven methodologies, securing a significant wage increase, underlining his elite status at the club and his position as one of the world's best players.

Research has found strong evidence that statistical models provide better results than heuristic human judgments (Dawes, Faust, and Meehl 1989), especially for more complex tasks (Tversky and Kahneman 1974), like estimating market values of football players is one. Grove et al. assessed 136 studies that compared human and statistical estimates in various fields and found the latter to be, on average, ten percent more accurate, independent of the training or experience the human judges had (Grove et al. 2000). With respect to Transfermarkt, Müller, Simons, and Weinmann argue that with all the data on market values being publicly available, one cannot take any competitive advantage of it (Müller, Simons, and Weinmann 2017). Others also push for clubs to run their own valuations and not just blindly trust in the values from Transfermarkt (Ackermann and Follert 2018). With statistical models, these limitations can be overcome. A club can run a model at a low cost, without public announcement, and determine a player's value for transfer negotiations. Club officials do not need to wait for the next market value update on a platform; they can plug in the most recent match data and obtain values within minutes, making a model a very efficient alternative. However, of course both a model and the data need to be available to the club. As the model will draw on the same set of parameters and apply the same weightings, results are transparent and unbiased, regardless of how renowned a player is and which league or club he plays for currently. This simplifies due diligence for young or unknown players and helps to avoid the trap of potentially manipulated values on Transfermarkt. In conclusion, once one has set up a working and accurate model, it can overcome many limitations of crowd judgements.

In the past, researchers heavily relied on linear models (*see related work Chapter 4.3)* to derive market values. This type of model assumes a linear relationship between exogenous variables and the endogenous variable market value. However, this relationship may not always hold. With a more complex machine learning model, we are free to explore other kinds of relationships (quadratic, cubic, exponential, logarithmic, cosine, etc.), which should yield more accurate results. The ultimate objective therefore is to come up with a quantitative tool that serves as a complementary approach to the experts-based estimations of market value.

4.3 Related work

This work was inspired by related papers that also aim to improve decision-making processes by shifting the choice from human judgements to statistical and data-driven predictions. The first model used to price football players was developed around the turn of the millennium, considering measurable and unmeasurable parameters for players and their productivity (Carmichael, Forrest, and Simmons 1999). Multiple researchers found that players' nationalities influence their transfer values (Pedace 2007; Szymanski and Smith 1997; Schokkaert 2016). In 2012, Franck and Nüesch published one of the most influential papers in the field. They were mainly investigating the effect of talent and popularity on the market values of football players, confirming that both parameters increase them. Regarding talent, they report the highest influence for the most valuable players. According to Rosen, superior talent is the driver which will make someone stand out from others and lead to exponentially higher earnings and, therefore, value (Rosen 1981). Next to that, network effects of popularity influence values (Adler 1985). If a consumer already knows about the talent of someone, either because they already watched them before or talked to others about that artist or athlete, they will be able to appreciate the performance more. From the latter, network effects arise, helping artists and athletes to gain popularity, as more and more people will know their names. Superstars can rise where technology allows them to reach large crowds via economies of scale. This is the case in football, where more than a billion people can watch a single match (FIFA 2018b). In addition, the authors also provide evidence that, if a player presents himself as a pop icon (like Zlatan Ibrahimovic or Cristiano Ronaldo) or gets attention from the media in another way, his market value will go up, next to the popularity achieved by performing well on the pitch (Franck and Nüesch 2012). Other researchers' work confirms that popularity and public attention push players' values (Lucifora and Simmons 2003; Lehmann and Schulze 2008; Herm, Callsen-Bracker, and Kreis 2014). Recent findings even show that it is possible to compensate for missing performance metrics (e.g., when injured) with social media activity (Frenger et al. 2019). Footedness was also an area of interest for researchers, who found beneficial effects on market values for players who can pass and shoot balls equally well with the left and right foot, i.e., are two-footed (Bryson, Frick, and Simmons 2013; Herm, Callsen-Bracker, and Kreis 2014). Even though individual effort, specifically running distance, was found to improve team performance, it was not observed to affect market values, indicating potential underpricing (Wicker et al. 2013; Weimar and Wicker 2017). In 2014, Herm, Kreis, and Callsen-Bracker presented an OLS regression model based on Transfermarkt data, which identified age, passing precision, scoring, and defensive abilities to influence market values significantly. Furthermore, they also found the market value of the entire team, performance grades in football journals, and Google search hits to affect individual players' values. He, Cachucho, and Knobbe made use of lasso regression to derive performance indicators impacting market values of attacking players, finding that precise shooting and scoring, as well as assists, successful dribbles, and few committed fouls contribute positively. Interestingly, they also found that most overvalued players were generally performing excellently. As their model did not utilize popularity, they conclude that the overvaluation can be explained by transfer values rising with the marketability of a player (He, Cachucho, and Knobbe 2015). Majewski also studied forwards with OLS, GLS, and FGLS models and got similar results: the number of goals and assists, the club's value, and the player's nationality had the highest impact. He also found significant goodwill (on average €40m for the five most valuable forwards) that could neither be explained by individual nor team performance, indicating again that popularity and marketability influence market values (Majewski 2016). Müller, Simons, and Weinmann built a multilevel regression model, analyzing five European leagues over six seasons, considering more than 20 parameters – the most thorough study to this date. Compared to Transfermarkt values, their model was more accurate in predicting the lower 90 percent of all realized transfer fees, while the crowd did a better job with the top 10 percent. As the reason for this they mention, once again, a possible superstar effect. The crowd is free to price in additional factors such as additional ticket and merchandise sales generated by a star player, which the model cannot do if it does not have access to this data; it just assumes the same magnitude of effect for all players (Müller, Simons, and Weinmann 2017). With a boosted regression tree model, Richau et al. examined the relative influence of individual performance indicators on market values, finding substantial differences in the relevance of variables like goals and assists for different positions (Richau et al. 2019). Gómez et al. assessed the performance of groups of players before and after signing new contracts via magnitude-based inference, concluding that important players tend to perform better in the year they sign a new contract compared to the year before, while less important players' performance level drops of in the year after having signed a new contract compared to before (Gómez et al. 2019). Kirschstein and Liebscher found that belonging to a certain team alters players' market values via a robust MM regression model. Moreover, they state that players contribute to the merchandizing and sponsorship income of clubs, which also increases their value (Kirschstein and Liebscher 2019). Singh and Lambda applied five models: linear and ridge regression, decision trees, random forests, and gradient boost to study the influence of crowdsourcing, popularity, and previous year statistics on market values of football players, discovering that performance parameters, crowd predictions, popularity and fantasy football ratings are solid predictors (Singh and Lamba 2019). In 2019, researchers and data professionals from WyScout, a leading platform providing football match data, built the algorithm PlayeRank to rate players based on 76 performance indicators, considering specific actions players must be able to perform for their position and role. They used 31 million data points from 20,000 matches for 21,000 players over 18 competitions and four seasons. The algorithm runs in three phases and is the most sophisticated one published to date, significantly outperforming its one-dimensional competitors, Flow Centrality and Pass Shot Value (Pappalardo et al. 2019). Felipe et al. used an OLS model to derive relations between the variables position, age, and quality of the team and league and market values on Transfermarkt for players from five European leagues, finding a significant impact of team level, birth month, league, position, and age of the player on the average and maximum value (Felipe et al. 2020). With artificial neural networks and random forest models, researchers developed recommendations for building a football team and planning transfers (Ćwiklinski, Giełczyk, and Choraś 2021). Another artificial neural network was used by Inan & Cavas to estimate the market values of midfielders in the Turkish Super League. They concluded that values cannot entirely be explained by performance parameters, moreover, also hypothesizing that injuries negatively affect market values (Inan and Cavas 2021).

A more recent development is the use of machine learning models to estimate market values based on data from Sports Interactive's 'Football Manager' (Yiğit, Samak, and Kaya 2020) and Electronic Arts' 'FIFA' video game series (Behravan and Razavi 2021; Al-Asadi and Tasdemır 2022; Arrul, Subramanian, and Mafas 2022; Lee, Tama, and Cha 2022) , as that is an easy way to account for a player's ability on the pitch. The FIFA data comprises 55 features that describe a player's strengths and weaknesses, position, demographic information, monetary value, and profile. The publisher, EA Sports, employs 400 contributors to estimate the player's real-life performances, frequently updating player data. These records are reviewed by over 6,000 football experts and talent scouts, who regularly advise on improvements and modifications to the database and data structure (Lee, Tama, and Cha 2022). This type of data is a good measure for differentiating performance factors for various positions (e.g., goalkeeper, defender, midfielder, attacker). Even though it is not actual performance data, the researchers' models could thus produce accurate results.

4.4 Model methods

There is a multitude of ML algorithms to choose from. To determine the best for our use case, we tested four different modelling techniques used in academic literature before and a neural network. Given that we try to predict a value that lies on a continuous scale, we are required to use a regression model. Regression analysis is a statistical method that tries to determine a relationship between a dependent variable and a set of other factors, the independent variables. The general idea is that the independent variables explain the behavior of the dependent variable (also called Y variable), allowing to make predictions on how the Y variable behaves given certain inputs (Gallo 2015).

The simplest form of regression techniques is called Ordinary Least Square (OLS). The method compares a set of actual values (y_i) to a set of predicted values ($\hat{y}i$) and minimizes their squared residuals. The residuals are defined as the difference between y_i and $\hat{y}_i$. They are subsequently squared to avoid that positive errors are compensated by negative ones, as they are equivalently penalizing for the model. This is also referred to as the cost function, which a model commonly aims to minimize. Summing up, OLS can be considered as a strategy that obtains a 'line' or 'plane' that is as close as possible to all data points (Alto 2019). However, there are further optimization strategies that we applied to improve model performance.

To take it a step further, we also applied Ridge and Lasso regression models to our data set. In general, when developing a machine learning model, there are two phenomena that you try to avoid: under- and overfitting. Underfitting occurs when the model is too simple for the data, implying that the assumption about the distribution of the data is wrong, meaning that there is a high bias. This scenario is depicted in *Figure 2*, with the red line representing the model. Overfitting, on the other hand, refers to the scenario where the model is too complex and too tailored to the training data at hand, resulting in wrong predictions. A slight change of input data can therefore completely alter the model's output (*Figure 3*) (Jabbar and Khan 2015).

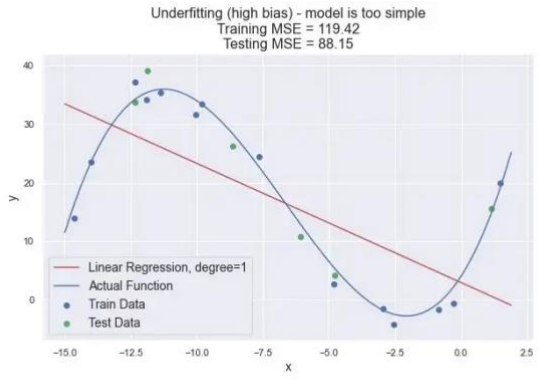

Figure 2: Example of an underfitting model

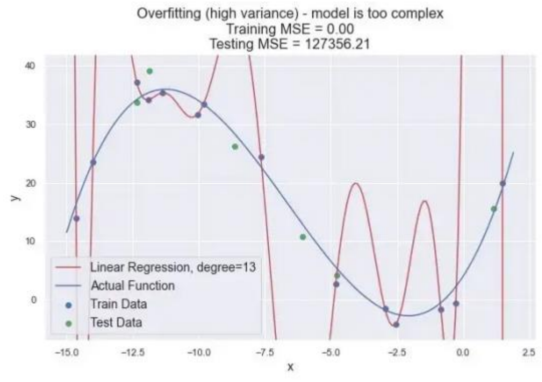

Figure 3: Example of an overfitting model

While underfitting can be avoided by augmenting the available data through several parameter transformations (e.g., feature engineering) and increasing model complexity, dealing with overfitting is not as simple. This is where regularization comes into play.

In regression, regularization refers to the process of adding additional terms to our cost function, often to introduce a preference for simpler models. The added term imposes a penalty on the complexity of the predicting function. This technique is especially important when dealing with datasets that have a lot of features, each of which contribute to predicting y. Ridge regression (L2) includes a type of regularization that penalizes coefficients of a predicting feature in proportion to the sum of the squares of the weights of the predictive features. This implies that this type of penalty helps drive outlier weights closer to zero but not quite to zero, reducing the importance of some features (Melkumova and Shatskikh 2017). Lasso regression (L1) penalizes weights in proportion to the sum of the absolute values of the weights. Hence, the difference to Ridge regression is simply that instead of taking the square of the coefficients, full magnitudes are considered. This results in the possibility of some coefficients being reduced to zero, removing the corresponding feature from the model. Consequently, in addition to reducing model complexity and avoiding overfitting, Lasso regression also helps in selecting the relevant features (Melkumova and Shatskikh 2017).

To improve model performance even further, there are several approaches that can be sought out, fitting under the umbrella term ensemble learning. The fundamental idea behind ensemble learning is that many base learners are trained as ensemble members. Their predictions are combined into a single output that, on average, should perform better than any other individual ensemble member with uncorrelated error on the target data sets (Yang 2017). A Gradient Boosting decision tree (GBDT) is an ensemble model that trains several decision trees sequentially by fitting the residual errors. However, when there is a multitude of features and the data size is large, the algorithm is limited in terms of efficiency and scalability, as every data instance needs to be scanned to approximate the information gain of every possible split point (Ke et al. 2017). A group of Microsoft researchers challenged these limitations with the LightGBM framework. It facilitates the estimation of the information gain per feature by excluding a significant part of the data from the computation and by bundling mutually exclusive features to reduce feature dimensionality. LightGBM speeds up regular GBDT training by up to 20 times while maintaining roughly the same accuracy (Ke et al. 2017).

4.5 Performance evaluation measures

To evaluate the performance of our model, we sought out common evaluation measures used for regression-based machine learning models described in literature. As mentioned in the previous section, we applied OLS regression that aims to minimize the cost function. This function here is the Mean Squared Error (MSE), which is the average of all squared differences between the actual value y_i and predicted value $\hat{y}_i$. A smaller MSE signifies a better model performance, implying that the model closely fits its predictions to the actual data points (Shcherbakov et al. 2013). A drawback for MSE is that is it sensitive to outliers in the data, which significantly increase the error term and bias for OLS models (Shcherbakov et al. 2013). A closely related performance measure is the Root Mean Square Error (RMSE), which, as its name suggests, is the square root of the sum of squared residuals. In terms of performance measurement, the RMSE conveys the same message as the MSE, as they both describe the goodness of fit of the model. However, RMSE is better in terms of interpretability, as it describes the average deviation in the same units as the target variable. To take it step further, we also used the normalized RMSE, which divides the RMSE by the range (maximum value – minimum value) of observed data points. Normalizing allows to overcome comparability problems between two models due to different scales of the data (Shcherbakov et al. 2013).

In addition to these measures, we observed the R^2 score of our model, which is a score that mathematically represents the percentage of variance in the dependent variable y that can be explained by the independent variables (Harel 2009). A high score represents a high degree of correlation between the explanatory ability of the independent variables and the actual predictions. The inherent problem of R^2 is that it mathematically cannot account for additional input parameters added to the model. Therefore, there is a chance that there is no relationship between the target variable and a new input parameter, but the R^2 score still improves when adding that variable. Hence, we also observed adjusted R^2, which allows to identify the additional explanatory effect of individual variables added to the model (Harel 2009).

5. Data

5.1 Data procurement (Web scraping & OPTA)

For the scope of this research, the time period considered ranges from January 2019 to November 2022. The dataset was gathered from two different data sources: Transfermarkt and Opta. To retrieve the publicly available data from Transfermarkt, we programmed a web scraping tool. This tool enables the user to access player-related data, such as market value or injury history, and to convert it into a .csv format suitable for further use. The Opta data was provided by our partner Hamburger Sport Verein, a renown German football club. The data covers the 1st and 2nd German Bundesliga, the Spanish LaLiga, and the French Ligue 1.

5.2 Exploratory Data Analysis

For a better understanding and context of the sourced data, we conducted an exploratory data analysis. The raw data consists of three different datasets as inputs:

- ***Dataset TM***: including 34,844 records of players, their characteristics, and their respective market values
- ***Dataset IN***: including 12,262 records of players, their respective injuries, and injury duration
- ***Dataset OPTA***: including 102,139 records of players and their respective matchday performances

5.2.1 Observed timeframe

To set a baseline of identical time intervals, we grouped all three datasets into bi-annual timeframes denoted in the column *HY* (for half-year). However, for each of the three datasets,

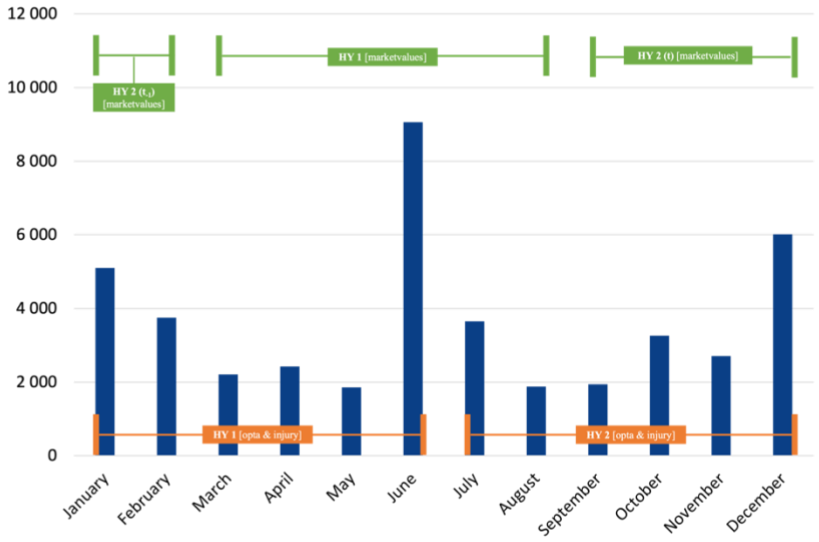

Figure 4: Amount of market value updates per month and half-year timeframe definition

the HY definition is distinctive, due to the type of data that they contain. The Opta and injury dataset include the actual performance and injury data, meaning that there are more entries during the months when matches were played. Therefore, HY = 1 represents the months from January until June, which are generally considered as one half of a season. The other half lasts from July until December and is denoted in the dataset as HY = 2. Conversely, there is a different logic that applies to the TM dataset and the market values. We assume that the market values are influenced by performance and injury factors of previous periods, i.e., a good performance or an injury in the months from January to June is reflected is reflected by a subsequent increase or decrease in valuation on Transfermarkt. Therefore, the market values that correspond to the performances or injuries between January and June are reflected slightly after they took place. We defined the HY = 1 timeframe for market values in the TM data frame as the period from March until August, as the updates in this period best explain the performances and injuries between January and June. timeframes if market value data are depicted in *Figure 4*.

The remaining months, September until February, correspond to HY = 2 for market value updates, as they best reflect the performances for the remainder of the season. Looking at the distribution of market value updates released by Transfermarkt in *Figure 4*, an uneven split can be detected, with an increased frequency in June and January/December. This is in line with the calendar of a normal football season. Consequently, because of our new split into two half years, 48% of the market values were released between March and August, as opposed to 52% between September and February, creating a more balanced set of observations. For the remaining features, the parameters were summarized in accordance with the time frame change, i.e., for a performance metric like goals scored, we added up the goals a player scored in the corresponding half year.

Another scenario that needed to be accounted for is when the duration of an injury spans across several periods. If this was the case, we split the duration of the injury at the cut-off date of the period. The first part of the injury remains in the period of when the injury occurred, while the remainder of the duration is attributed to the following period. This action ensures that a market value is accurately explained after the injury had ended and not when it occurred.

5.2.2 Target variable

The target variable, i.e., the variable that we want to predict with our model, is *Marketvalue*. *Table 4* describes the target variable. The average market value across the whole dataset is € 7.2m, the median equals € 1.2m. Market values range between € 0.0m and € 200.0m. The standard deviation amounts to € 11.8m. To get a better understanding of how the figures are distributed, we can refer to histogram (*Table 3*) with € 5m intervals. We can concur that there is a significant positive (right-) skewness, as shown in *Figure 5*. Around 48% of entries have a market value

Marketvalue	in €m
Count	43,632
Mean	5.5
Standard Deviation	11.8
Minimum	0.0
25%	0.4
50% (Median)	1.2
75%	5.0
Maximum	200.0

Table 3: Market value summary

of less than € 1.0m and only 23% of entries are above the mean. To account for this uneven distribution, the top and bottom 5 percentiles were removed for our model, resulting in a less skewed data set as depicted in *Figure 6*.

5.2.3 Parameters

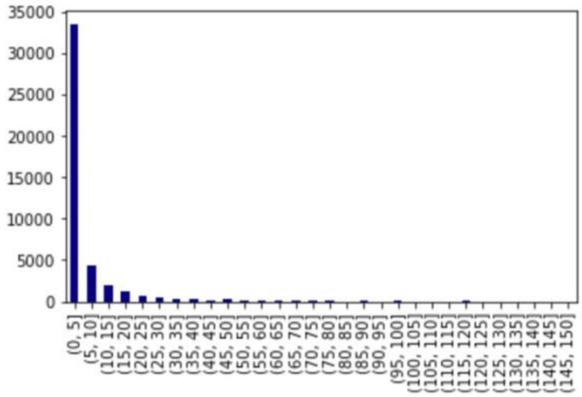

Figure 5: Market value distribution

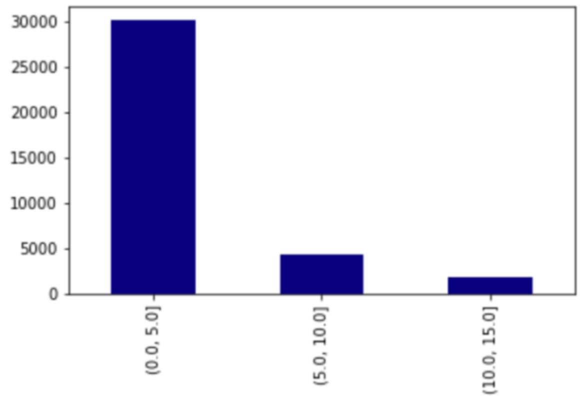

Figure 6: Market value distribution excl. top 5%

In addition to the target variable, the data set includes 231 possible explanatory variables. The parameters can be divided into four categories shown in *Table 4*:

Category	# of columns	Content
Descriptive data	36	Date, Team, League, Game, Score
Injury data	4	Date, Duration, Types
Performance data	166	Player and Team related game stats
Player related data	15	Name, Height, Age, Position, Pref. foot, Nation

Table 4: Variable categories

The descriptive data contains information about the timeframe, team, and league in which the individual played during this time. The injury data consists of the accumulated injury duration during each period and information about the type of injury. We excluded illnesses such as Covid-19 or a flu from this data set to focus on injuries. Performance data from Opta contains detailed information about game related performance measures. Finally, player-related data contains personal information about the player, such as the name, height, age, position, preferred foot, and nation of each player.

5.2.3 New parameters and transformations

To enhance the quality of our dataset, we added further parameters that scientific literature proved to be relevant for the improvement of our model performance. The additional columns and their transformation explanation can be found in *Table 5*.

Column	Description	Type
de5, es5, fr5, other	Dummy to denote if player is within a top 5 club in Germany, France, Spain or not	New variable
Last MV	Contains the market value of the previous period of observation	New variable
Nation_Argentina, Nation_Brazil, Nation_Croatia, Nation_Portugal	Dummy that indicates if the player comes from a leading country in terms of market values. The four selected countries have the highest average market value.	New variable
Pos_AT, Pos_DF	Dummy that indicates the player's on-field position. If both Dummies are 0, the player is a midfielder (MF)	Transformed from *pos*
PrefFoot_both, PrefFoot_right	Dummy to denote if player is right-, or both footed. If both Dummies are 0, the player is left-footed	Transformed from *PrefFoot*
CLT, patella, fracture, meniscus, achilles, syndesmosis	Dummy for the occurrence of a cruciate ligament tear, patella tendon tear, meniscus tear, Achilles tendon tear, syndesmosis ligament tear or bone fracture	New variable

Table 5: New parameters overview

For each player we created a Dummy variable that indicates whether the player is part of a team that is considered a top five club in either Ligue 1 (*fr5*), Bundesliga (*de5*) or LaLiga (*es5*). These three leagues, next to the English Premier League and Italian Serie A, are part of the top five European leagues, generally considered as the world's best. We identified the top five clubs per league according to accumulated league points rankings within our time frame (2019 until 2022). This information is required to test the hypothesis of whether a player's club can be an indicator of market value. Furthermore, we added the parameter *Last MV*, which includes the market value of the previous observed period, i.e., if the observed period is the first half of the year 2022, then the column *Last MV* contains the market value of the second half of 2021 for that player. By adding this variable, we intend to account for effects that cannot be explained with the performance data at hand. Such effects entail for example a player's popularity (e.g., their ability to generate merchandise sales) and other similar value-driving factors. In general, market values do not tend to fluctuate substantially within the time frame we observe, therefore the market values of the 6 months prior to the observation are assumed to be a valid and accurate indicator of what the next market value should be at if performance remains at a similar level.

However, what happens if a youth player makes it up the ranks of a club into professional football and does not have a previous market value? In this context, we set *Last MV* to zero, signifying a starting point from which a player improves through their performance. Further, we decided to transform the existing column *pos*, which represents the players on-field position, from a detailed breakdown to a more simplified split into attacking (*AT*), midfield (*MF*) and defensive (*DF*) players as observed in *Table 6*. This measure was carried out to transform a categorical variable into a Dummy and to simultaneously remove processing complexity by eliminating a multitude of columns, while retaining a good level of information about the on-pitch positioning and role of the player. Since goalkeepers have a unique position and influence within the game, it is difficult to compare them to the outfield players. As outlined in a study, different player positions require different skills (Behravan and Razavi 2021). Such skills are measured differently from each other and should therefore be accounted for. The performance measures in our Opta dataset are tailored to outfield players and therefore might lack explanatory power for the goalkeeper position. This would also impede the model's

Position	New Position
Central Midfielder	MF
Centre Forward	AT
Defensive Midfielder	MF
Left Attacking Midfielder	MF
Right Attacking Midfielder	MF
Right Midfielder	MF
Left Centre Back	DF
Right Centre Back	DF
Left Midfielder	MF
Central Midfielder	MF
Centre Attacking Midfielder	MF
Right Back	DF
Left Back	DF
Central Defender	DF
Left Winger	AT
Right Winger	AT
Second Striker	AT
Right Wing Back	DF
Left Wing Back	DF
Goalkeeper	GK

Table 6: Player position re-classification

performance, which is why all entries with goalkeepers were removed. Additionally, we decided to remove the column *Position_MF*, as this information is implied in the other two columns (*Position_AT*, *Position_DF*), which are mutually exclusive and collectively exhaustive.*Position_AT*, *Position_DF*), which are mutually exclusive and collectively exhaustive. Finally, there were discrepancies between performance variables, as some were relative to a full game of 90 minutes, whereas other variables were absolute values recorded in relation to actual minutes played. *Tables 7 and 8* and below depict the different aggregation modes and steps for relative and absolute variables and adjustments for discrepancies in minutes played:

Pre transformation		Example relative variables			Example absolute variables		
Player	Min	Pass%	1v1%	DuelsDefPer90	Touches	FwdPass	Tckl
rodrigo zalazar	66	69.2%	0.0%	5.45	42.27	6.82	1.36
rodrigo zalazar	45	60.0%	50.0%	4.00	40.00	4.00	2.00
rodrigo zalazar	71	72.2%	0.0%	3.80	38.03	10.14	0.00
rodrigo zalazar	58	72.7%	20.0%	3.10	43.45	4.66	0.00
rodrigo zalazar	71	50.0%	100.0%	13.94	36.76	5.07	2.54

Table 7: Data sample raw data

1st step		Example relative variables			Example absolute variables		
Player	Min	Pass%	1v1%	DuelsDefPer90	Touches	FwdPass	Tckl
rodrigo zalazar	311	64.8%	34.0%	6.06	200.51	30.69	5.90

Table 8: Data sample aggregation of relative and absolute variables

2nd step		Example relative variables			Example absolute variables		
Player	Min	Pass%	1v1%	DuelsDefPer90	TouchesPer90	FwdPassPer90	TcklPer90
rodrigo zalazar	311	64.8%	34.0%	6.06	58.00	8.88	1.71

Table 9: Data sample final performance data adjusted for 90 minutes

Table 7 shows a sample of raw data for the player Rodrigo Salazar for one period of observation. The data includes metrics such as percentage of completed passes (*Pass%*), which are relative, as opposed to metrics like number of touches within a game (*Touches*), which are absolute and correspond to the minutes played. The former is aggregated as a mean, while the latter is aggregated as a sum, as seen in *Table 9.* However, a player with more game time (*Min*) naturally has higher absolute values due to the increased exposure on the pitch. This translates to high correlations between those features, resulting in a loss of explanatory power of the individual variable within the model. To eliminate this effect the absolute variables were divided by the minutes played (*Min*) and multiplied by 90.

Additional dummies for the most severe injuries, as they probably have a larger impact on players' market values, have been created. In the injury dataset, which comprised 12,262 records, 304 distinct injuries were recorded. We grouped similar ones and then selected the six groups with an average duration over two months. Those were ACL, meniscus, patellar tendon, syndesmosis ligament, and Achilles tendon injuries, and fractures. Please refer to *Table 30* in the appendix for further details on the groups and individual injuries.

6. Model

The objective of this research is to provide a tool that enables the assessment of market value changes brought on by injuries. Accordingly, we sought out meticulous data cleaning and transformation processes to ensure the appropriate structure and format. Subsequently, different model techniques were applied and evaluated on their performance. Finally, injury specific features were assessed.

6.1 Modeling decisions

The regressors must be independent of one another, meaning that there cannot be multicollinearity when creating a valid model. We calculated the Variance Inflation Factor

(VIF) to rule out explanatory factors that are highly correlated. There is a significant multicollinearity if the VIF is higher than 10 (Michael Kutner et al. 2004). This limitation caused the number of features to drop from 200 to 82. A Multiple Linear Regression (MLR) was carried out on this dataset. The parameters were then condensed to the features described in *Table 10*, each of which is statistically significant at the 5% level:

Feature	Description
HY	Current half-year
Year	Current year
SeqEndsInGoal	Sequences which ended in a goal
TakeOn%A3	Take on percentage in final third
KPAf1v1	Key pass after 1v1
TouchOpBox	Touches in opponent's box
GoalOP	Goals from open play

Feature	Description
Age	Current Age
PrefFoot_both	No weak foot
Duration	Days missed due to injuries
de5	Top 5 Bundesliga team
fr5	Top 5 Ligue 1 team
Last MV	Market value from last period

Table 10: Significant features

We found that the only injury-related feature of significance was *Duration*. This model was defined as our benchmark for further approaches. A Lasso and Ridge Regression were performed to test the optimal regularization technique. With essentially two identical results, we chose to apply both L1 and L2 regularization penalties. A LightGBM model was tested fourth. To optimize the model results, a pipeline was constructed, which balances and scales the data, chooses the best combination of features and hyperparameters for LightGBM. Then, all potential model specifications are calculated using a Grid Search Cross Validation with 10 folds, and the model specification with the lowest MSE value is selected. The LightGBM configuration with the smallest MSE is shown in *Table 11* along with the balancing and scaling techniques, the hyperparameters that were examined during the cross validation, and the results of the cross validation.

Category	Possible Values	**Selected by CV**
Balancing methods	RandomUnderSampler, or RandomOverSampler	**RandomOverSampler**
Scaling methods	RobustScaler, StandardScaler, or MinMaxScaler	**RobustScaler**
LightGBM learning rate	0.1, 0.01, 0.001	**0.1**
n-Estimators	1 – 301 (in steps of 30)	**91**
Alpha (L1 penalty)	0.01, 0.26, 0.51, 0.76, or 1.01	**0.26**
Lambda (L2 penalty)	0.01, 0.26, 0.51, 0.76, or 1.01	**0.76**
Number of leaves	32, 64, 94, or 128	**32**
Max depth	4, or 8	**8**

Table 11: Pipeline LightGBM

6.2 Performance measures

Table 12 and *Table 13* summarize the performance metrics of all tested models for train and test data, respectively. The best performing model is highlighted in green.

Model	R^2	Adj. R^2	RMSE	Norm. RMSE
MLR (82 features)	0.91	0.91	2.73	0.03
MLR (13 sign. features)	0.91	0.91	2.76	0.03
Ridge	0.91	0.91	2.76	0.03
Lasso	0.91	0.91	2.85	0.03
LightGBM	0.98	0.98	1.40	0.02

Table 12: Performance metrics on train data

Model	R^2	Adj. R^2	RMSE	Norm. RMSE
MLR (82 features)	0.91	0.91	2.86	0.03
MLR (13 sign. features)	0.91	0.91	2.85	0.03
Ridge	0.91	0.91	2.85	0.03
Lasso	0.91	0.91	2.85	0.03
LightGBM	0.89	0.89	3.16	0.03

Table 13: Performance metrics on test data

All the approaches' performance metrics are extremely comparable. Only LightGBM displays results that differ. In comparison to the other models, it fits the training data the most accurately, as shown by (adjusted) R^2 and (normalized) RMSE. Surprisingly, for performances on the holdout data, it is the opposite. The findings demonstrate that, while all other models continue to perform as intended, LightGBM's accuracy outside of the fitted data declines. In general, all models are close to an adjusted R^2 of 0.9, indicating that the model can account for about 90% of the variance and, consequently, the deviations from the mean values (Wooldridge 2015).

6.3 Feature importance

To further analyze the results, permutation feature importance and Local Interpretable Model-agnostic Explanations (LIME) were obtained for the MLR and LightGBM model. These analyses help to understand machine learning models and explain their predictions. To facilitate comprehension of the LIME model explanation, the MLR and LightGBM LIME outputs are shown in *Figures 7 and 8*, respectively. The valuation example uses Rayan Cherki (random choice) from the first half of 2022, when his last market value was € 21.5m and he missed 89 days of action due to injury. Transfermarkt estimated his market value at € 18.0m.

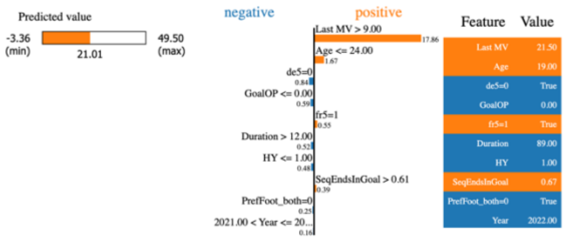

Figure 7: MLR prediction example 1 (LIME)

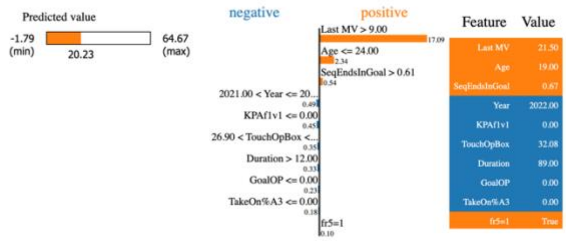

Figure 8: LightGBM prediction example 1 (LIME)

With € 21.0m for the MLR and € 20.2m for the LightGBM model, the projections are quite close to one another. Each attribute is valued differently by each model. *SeqEndsInGoal* was only the ninth most significant parameter contributing to the prediction of the MLR model, which had the third-highest impact on the market value prediction with LightGBM (+ € 0.5m). It is particularly remarkable how both models examine the effects of injuries with a lengthy duration. Both models use twelve days as the threshold for a negative effect (twelve days or less have a positive effect on market value, and a duration greater than twelve days has a negative effect on market value), however the MLR model appears to penalize an injury with a greater reduction in market value (- € 0.5m) than the LightGBM model (- € 0.3m).

Permutation feature importance is another useful tool for comprehending how the model derives its prediction. The method quantifies the increase in the model's prediction error when a feature's values are permuted. *Figures 9 and 10* illustrate the significance of permutation features for the MLR and LightGBM models, respectively. Similar weights are assigned to the features by both models, with *Last MV* having the greatest influence on the prediction (used as an anchor for market value dimension and maybe incorporating unobserved effects such as popularity).

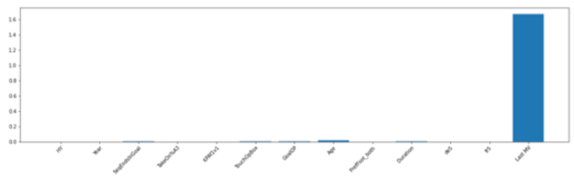

Figure 9: MLR permutation importance

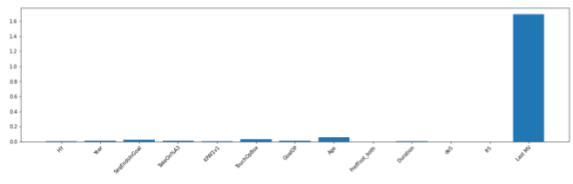

Figure 10: LightGBM permutation importance

The above-mentioned observation regarding *SeqEndsInGoal* can also be made from this approach. The LightGBM model considers this variable to be more significant than the MLR model. In general, LightGBM ranks most variables higher than MLR does.

6.4 Model conclusion

In general, the model's performance is excellent, and the supplied data provides a solid foundation for the development of an accurate model. However, there are still other restrictions that will be addressed in further detail at the final conclusion and the limitations of this paper.

The objective of this study was to determine the influence of injuries on player market prices. The model suggests that *Duration* and, hence, injuries are significant factors in estimating market values. The number of missed days due to injury appears to already have a direct effect on player values. However, there may be hidden impacts of injuries that are accounted for by other variables, such as poorer performance after recovery or the moment in a player's career when the injury occurs (i.e., young players who suffer typical recurrent injuries). Subsequent analyses will attempt to discover some of these hidden injury-related effect

7. The role of recurrent injuries in young players

Recurrent injuries are a common problem among young football players. These injuries can have long-term consequences on a player's health and ability to participate in the sport.

To acquire a better knowledge of the prevalence of recurrent injuries in young football players, as well as the causes of these injuries, it is important to review prior studies on the subject. According to recent research, recurring injuries resulted in lengthier absences than first-time injuries (Stubbe et al. 2015). A study discovered the trend that injury incidence and recurrence have increased since the previous epidemiological research of English professional football players (17% vs. 7%) (Jones et al. 2019). Another study from 2011 indicated that re-injuries constituted 12% of all injuries, and they caused longer absences (24 vs. 18 days) than one-time injuries (J Ekstrand, Hägglund, and Waldén 2011).

In order to conduct a more in-depth investigation into this matter, we will first utilize two case studies to illustrate the level of impact that injuries can have on a player's career. This is followed by a statistical study that is based on the data used in *Chapter 6*, and its purpose is to evaluate the hypothesis that young players are more susceptible to reoccurring injuries.

7.1 Qualitative Examples

The data used in the following chapter was gathered from Transfermarkt. Ousmane Dembélé and Jack Wilshere were two promising talents who suffered recurring injuries at a young age, which may have prevented them from reaching their full potential. In the following chapter, both careers and the role of recurring injuries will be presented.

7.1.1 Ousmane Dembélé

Ousmane Dembélé, born on the 15th of May 1997 in Vernon, France, began his professional football career at the age of 17, when he was promoted from the youth squad to the second team of FC Stade Rennes, a club in the French first division (Transfermarkt 2022a). Dembélé was able to advance to the first division after playing one season with the second team. He appeared in 26 of 38 games during his debut season with the first squad in 2015/16, scoring 12 goals and assisting on five others. After a great rookie campaign, he drew the attention of German powerhouse Borussia Dortmund, who ultimately paid € 35m to acquire the 19-year-old French star. According to Transfermarkt, his market value at this time was € 12m. Consequently, Borussia Dortmund saw a tremendous amount of potential in him, as they spent nearly three times his market value. Across all three tournaments (Bundesliga, German Cup, and Champions League), Dembélé made 48 appearances, scored 10 goals, and made 21 assists in 48 matches (Bundesliga, German Cup, and Champions League). FC Barcelona, one of the most prestigious clubs in the world, was impressed by his play, and acquired Dembélé, whose market value was € 33m at the time, for a staggering € 140m. Unfortunately, only a few months after his 2017 transfer to Barcelona, the French rising sensation sustained a serious thigh injury that sidelined him for 106 days. Since then, these thigh issues returned five additional times. This injury category cost Dembélé 413 days of his career in professional football. The French offensive player missed around 27.3% of his professional football career due to injuries, which amounted to 698 days during the last five years. These injuries included several muscle bundle tears and a knee injury. *Figure 11* depicts Ousmane Dembélé's previous injuries.

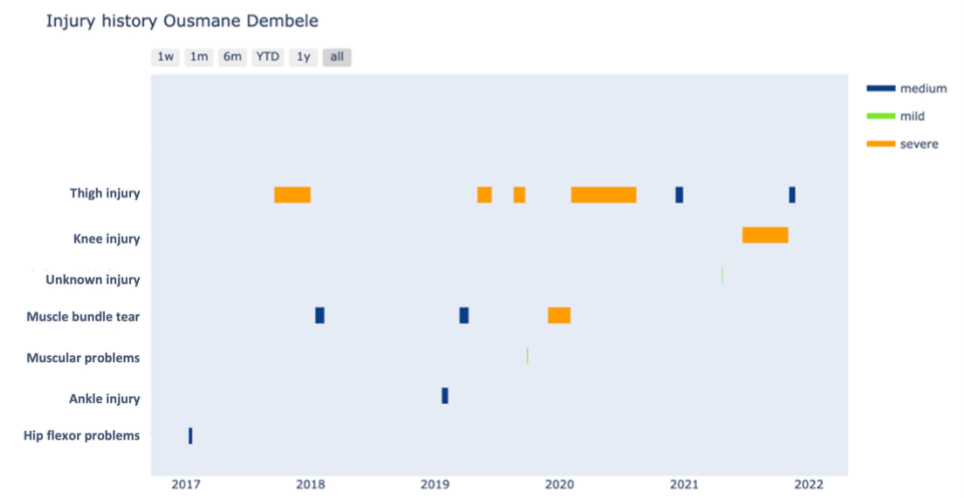

Figure 11: Injury history of Ousmane Dembélé

The orange portions represent an injury that lasted at least four weeks, the blue bar represents an injury that lasted between one and three weeks, and the green bar represents an injury that lasted less than one week. The illustration highlights the severity of his reoccurring thigh injury.

Before he suffered the second injury to his thigh, Dembélé had reached the peak of his market worth. According to Transfermarkt, his value was estimated to be € 120m in February of 2019. *Figure 12* demonstrates that after reaching its highest point, his market value continued to decline or remained unchanged until November 2022.

Figure 12: Market value history of Ousmane Dembélé

In his entire career, Ousmane Dembélé played 249 matches in club football, scored 59 goals, and made 68 assists (*Table 14*).

Category	Thigh	Knee	Muscle Bundle	Other	**Total**
Injuries of this category	6	1	3	4	**14**
Total days missed	413	134	118	33	**698**
Average days missed	69	134	39	8	**50**
Total matches missed	55	14	24	7	**100**
Average matches missed	9	14	8	2	**7**

Table 14: Injury statistics for Ousmane Dembélé

Due to his ailments, he was unable to participate in 100 matches, which suggests that he would have been capable of making 40% more appearances if he hadn't had so many injuries. In consequence of this, his development was continually hampered, and he was denied the opportunity to become one of the best players of his generation. While playing for Borussia Dortmund, the French offensive player averaged 0.20 goals scored and 0.44 assists per game. Since joining FC Barcelona, Dembélé has only been able to increase his scoring rate by 0.02

(it is currently at 0.22) but his rate of assists per game has dropped to 0.24. To put his progress in perspective, a comparison was drawn to one of Brazil's most promising talents, Vinicius Junior (Transfermarkt 2022c). The Real Madrid left winger was born in July 2000 and is currently considered to be one of the club's most prolific players. He began his professional career when he was 17 years old, and after playing for only one year with the first squad of Clube de Regatas do Flamengo, he made the move to the Spanish record champion. This was exactly one year after Ousmane Dembélé made his debut in the Primera Division. Real Madrid, also known as ‘Los Blancos’, acquired the 18-year-old Brazilian player by paying a transfer fee of € 45m. At that time, his estimated market value was € 35m. Vinicius has played 191 games for Real Madrid since the start of the 2018/19 season. During that time, he has scored 46 goals (0.24 goals per game) and assisted 48 goals (0.25 assists per game). *Table 15* presents a comparison of the progress of Dembélé and Vinicius Jr. between their first and third seasons in the Spanish La Liga, as well as a comparison of the performance of Dembélé when he was the same age. The comparison demonstrates that based on his performances prior to his first serious injury, Dembélé's future held great potential.

Category	Age	Games	Goals (per game)	Assists (per game)	**Total Contributions** (per game)
Dembélé (1st Season)	20	17	3 (*0.18*)	7 (*0.41*)	**10** **(*0.59*)**
Vinicius Jr. (1st Season)	18	18	1 (0.*06*)	2 (0.11)	**3** **(*0.17*)**
Dembélé (18 y/o)	18	26	12 (*0.46*)	5 (*0.19*)	**17** **(*0.65*)**
Dembélé (3rd Season)	23	30	6 (*0.20*)	3 (*0.10*)	**9** **(*0.30*)**
Vinicius Jr. (3rd Season)	21	35	17 (*0.49*)	13 (*0.37*)	**30** **(*0.86*)**
Dembélé (21 y/o)	21	29	8 (*0.28*)	4 (*0.14*)	**12** **(*0.41*)**

Table 15: Perfomance comparison Dembélé vs. Vinicius Jr.

Comparing the two players' performances during their first seasons in La Liga as well as when they were both 18 years old reveals that Dembélé was the more successful of the two during these times.

However, since joining Real Madrid, Vinicius Jr. has only suffered one injury: a ruptured knee ligament occurred in 2019, which sidelined him for 58 days and caused him to miss nine games. Since that time, he was spared from injuries, which enabled him to carry on with his growth process and improve as a result of the newly gained experiences he has obtained from training and competitions. After only three seasons, Vinicius Jr. has established himself as a crucial player for ‘Los Blancos’, scoring 17 goals and contributing 13 assists during the 2021/22 campaign. As a result, not only did he increase the number of scorers points he achieved by a factor of ten, but he also raised the percentage of goals he contributed to each match from 0.17 to 0.86, while Dembélé's performance continued to decline with time, most likely as a result of his history of injuries.

The instance of French super star Ousmane Dembélé exemplifies how a catastrophic injury can keep a player from realizing his full potential. Ousmane Dembélé is currently 25 years old and still has a promising career ahead of him, which will hopefully not be marred by more injuries.

7.1.2 Jack Wilshere

Today, Ousmane Dembélé is still actively pursuing a future in professional football. Jack Wilshere will serve as an example of a career that has already come to an end and was marred by a number of injuries.

Wilshere, born on the 1st of January 1992 in Stevenage, England, started his career in professional football career with Arsenal FC in London (Transfermarkt 2022i). The young English Midfielder made his first appearance for Arsenal's first team when he was only 16 years old. After two years, Wilshere was able to prove himself worthy of a spot in the starting eleven.

During the 2010/11 season, he participated in a total of 50 matches across all competitions (Premier League, FA Cup, League Cup and Champions League). In 2011, he experienced an ankle injury which required him to undergo surgery and forced him to miss a few months of playing football. This led to a series of injuries, which is depicted in *Figure 13*.

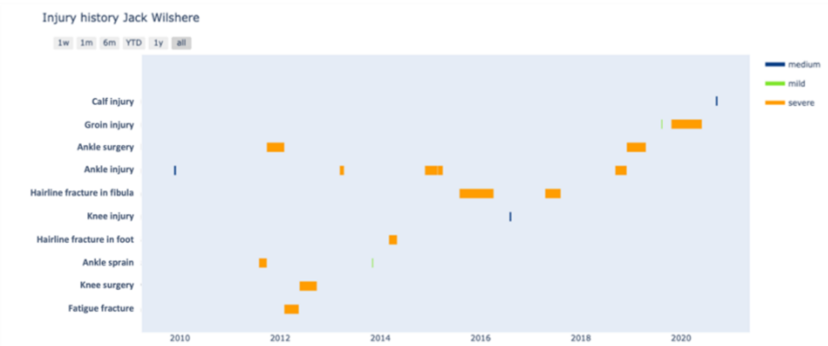

Figure 13: Injury history of Jack Wilshere

Over the course of 13 seasons, he participated in 303 games but was sidelined for 214 of them due to 18 injuries. The ankle, groin, knee, and calf bone were the areas of his body that were impacted, and each of these areas was affected more than once. According to the data presented in *Table 16*, he missed an average of 82 days and 11 matches for each injury.

Category	Ankle	Calf bone	Groin	Knee	**Total**
Injuries of this category	11	3	2	2	**18**
Total days missed	738	370	226	136	**1470**
Average days missed	67	123	113	68	**82**
Total matches missed	134	52	22	6	**214**
Average matches missed	12	17	11	3	**11**

Table 16: Injury statistics for Jack Wilshere

The injury to his left ankle was the costliest since it caused him to miss 134 matches. Wilshere was sidelined for a total of a whole season as a result of the three injuries he sustained to the bone in his left leg. On average, these injuries caused him to miss 17.3 matches, which is equivalent to around one third of the season. At the age of 21, he had reached his peak market value of € 33m in July 2013 (*Figure 14*), when 70% of his injury burden would still lay ahead of him. Within the following seven years, the value dropped to less than four million euros.

Figure 14: Market value history of Jack Wilshere

This summer (2022), he ended his career, relatively early at the age of 30, as he no longer wanted to subject his body to the great strain caused by professional football. This example demonstrates that injuries not only impede player development but can also cause a great talent to end his career early.

7.2 Quantitative Analysis

Following the findings obtained from the cases provided, a quantitative approach was used to discover a general pattern of serious reoccurring injuries encountered by young players.

7.2.1 Duration (age cluster)

First, the data used for the model in *Chapter 6* was separated into three clusters. Players classified as 'young' are those who are 23 years old or younger, 'old' players are those who are 30 years old or older, and 'middle-aged' players are those who fall in the age range in-between both other segments. *Figure 15* depicts the distribution of players among clusters. 'Middle-aged' is the largest cluster with 2,479 individuals, 'young' is the smallest with only 926 players in the dataset, and 'old' contains 1,300 persons.

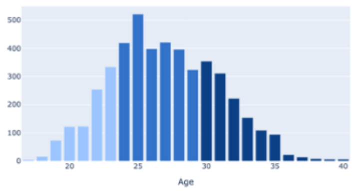

Figure 15: Player distribution per cluster

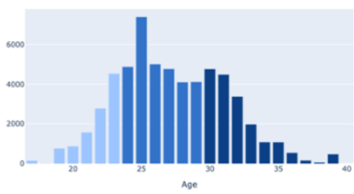

Figure 16: Total missed days due to injury per cluster

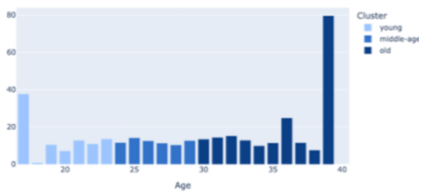

Figure 17: Average missed days per player in cluster

The total number of days missed due to injury follows a similar pattern, as shown in *Figure 16*. However, *Figure 17* demonstrates that very young and very old players, in particular, are at risk for injury. Young players are likely to be inexperienced or insufficiently strong, whereas older players are already fragile owing to their lengthy career. Aside from these age groups, the remaining age groups have nearly identical average days of injury per period.

After describing the various age groups, a linear regression was carried out on each of these clusters. The same feature selection technique as the one described in *Chapter 6.1* was used, resulting in 82 explanatory variables. The findings give credence to our working premise from the beginning.

Cluster	Age	Entries	Adj. R^2	Mean market value	p-value Duration	Duration coefficient
Young	<24	926	0.80	€ 8.4m	0.01	€ 20.7k
Middle-aged	24 - 29	2,479	0.91	€ 7.3m	0.04	€ 5.8k
Old	>29	1,300	0.91	€ 5.4m	0.12	-

Table 17: Regression results of Duration variable for age clustering

As can be seen in *Table 17*, the feature *Duration*, which refers to the total number of days that a player has missed due to an injury over the course of the most recent half year, was only among significant explanatory variables at the 5% level for young and middle-aged players. The p-value for the effect of injuries on the market value of older players is 0.12, which indicates that the effect is not statistically significant. The magnitude of the monetary damage caused, measured in terms of a loss in market value, varies widely among age groups. A young player's market worth will decrease by an average of € 20.7k for every day they miss due to injury, whereas a middle-aged player's market value will only decrease by € 5.8k for every day they miss (*Table 17*). According to the linear regression, the market value of an average young athlete would decrease by roughly € 2.1m if they suffered a serious injury that kept them out of action for one hundred days. With an average market value of € 8.3m across all young players in the data, this hypothetical injury would reduce the market value of an average player in this cluster by approximately 25%, whereas the same assumption applied to a middle-aged player only affected the player's market value with a reduction of almost 8%.

This data demonstrates that injuries can have a significant impact on players of any age, but particularly on younger athletes.

7.2.2 Recurring injuries

To further analyze the repeatability attribute in more detail, an additional data transformation was performed. To account for this feature during our analysis, the occurrence of injuries per player and their popularity were examined. *Table 18* presents the most serious injuries that have reoccurred for at least five players in our dataset, as well as the average recurrence and mean duration per repeat.

Injury	Mean Occurrence	Mean Duration	Injury	Mean Occurrence	Mean Duration
Cruciate ligament rupture	2.11	227	Meniscus injury	2.25	69
Knee surgery	2.27	124	Bone edema	2.40	61
Metatarsal fracture	2.60	95	Shoulder injury	2.29	52
Muscle bundle tear	2.88	61	Inner ligament injury	2.0	57
Knee injury	2.49	65	Back injury	2.5	41

Table 18: Most dangerous recurrent injuries

Following that, we included a dummy variable to our model that equals 1 if a player has already suffered one of the aforementioned injuries during his career. The linear regression from *Chapter 7.2.1* was repeated using this additional feature (now 83 features). Again, the hypothesis of a more significant impact on young players is supported by the results shown in *Table 19*. The newly introduced variable only was among significant features on a 5%-level for young players. The incidence of one of these ailments during a player's career has no significant effect on the market value of middle-aged and old players.

Cluster	Age	Entries	Adj. R^2	Mean market value	p-value Recurr.	Recurr. coefficient
Young	<24	926	0.81	€ 8.4m	0.02	€ 1.0m
Middle-aged	24 - 29	2,479	0.91	€ 7.3m	0.44	-
Old	>29	1,300	0.91	€ 5.4m	0.28	-

Table 19: Regression results of Recurring variable for age clustering

The model estimates a € 1.0m loss in market value, whereas the effect of one day injured (*Duration*) reduces to € 19.6k, which is € 0.9k less than the results from *Chapter 7.2.1*. Now, the first occurrence of an injury comparable to the example presented in *Chapter 7.2.1* reduces the market worth of the average young player by almost € 3.0m, or € 0.9m more than in the previous analysis.

7.3 Results

In conclusion, the qualitative and quantitative studies, which can be found in *Chapters 7.1 and 7.2*, respectively, reveal encouraging results to support the conjecture that was made at the beginning of the paper. Our research showed that injuries had a negative impact on the market prices of players of all ages, but particularly younger players. This suggests that the feature Age accounts for some of the total financial harm that an injury causes. Furthermore, there is evidence that recurring injuries are especially harmful to the careers of young players. The explanation of these inequalities across age groups may be the expected remaining career duration and the interruption of young athletes' development processe

8. Conclusion

The aim of this work was to analyze the risk of injuries for potential investors, in this case football clubs, desiring to acquire a new player. As soon as an investment decision for a club is pending and scouts have discovered a potential transfer, the target is carefully assessed, and all risks and opportunities are priced so that a competitive offer can be made. Due Diligence encompasses a broad range of criteria. The purchasing football club must evaluate the player's fitness, health status, injury risk, prior performance, potential, and popularity, among numerous other value-driving criteria. In this study, the role of injuries as a risk was examined in greater detail. Using both simple linear regression and more complex machine learning techniques, a model for estimating market value was developed. With performance data and player attributes such as height, age, and injury history, a model with an adjusted R^2 greater than 0.91 was fitted. We discovered that a direct effect of the overall injury duration over a 6-month period is a key value-driving element.

Based on these findings, additional studies were conducted to investigate potential hidden effects of injuries incorporated into other characteristics, such as age or position. These studies revealed an age-dependent distinction in the significance of injuries. The occurrence of injuries has a disproportionately negative impact on the market value of young players, whereas the reduction in market value related to injury duration is 70% less than that of young players. An old player's market worth is not significantly impacted by the number of days missed due to injury in the previous six months. Several qualitative examples were presented to support the hypothesis that reoccurring injuries have a significant effect on the market worth of developing talent. Implementing a new dummy variable that accounts for the prevalence of severe recurrent injuries in the past could provide evidence for this notion. As a result, this characteristic is only significant for the group of young athletes. These results suggest that the variable *Age* may cover a fraction of the monetary damage caused by injuries.

9. Limitations and future work

Throughout the course of working on this project, we encountered various limitations. First, it is difficult to make conclusions regarding the value of players from leagues with less coverage because the availability of data varies heavily. In addition, Transfermarkt statistics may be biased as market values could be partially manipulated and injury occurrences are not labelled accurately. For example, we saw durations of 19 to 587 days within the category ‘cruciate ligament injury’, which indicates that these may relate to dissimilar injuries. This bias may have transferred to our model. Unfortunately, there was no comparable database to validate the provided values. CIES and KPMG cover fewer participants and employ different methodologies. Besides, Hamburger SV was unable to provide us with a detailed documentation for the Opta data. Thus, there is still potential for improvement in data processing and feature selection, as it is practically impossible to identify every bias without knowing a feature's precise meaning. In addition, the supplied dataset is missing potentially essential information, such as the popularity of a player or absence owing to poor performance. We could not include popularity since we did not have access to data on the development of likes for players’ social media accounts over time. Moreover, if a player is injured for an extended length of time and does not play during a specific time period, his absence will not be visible to the model since no Opta data is available. We investigated a number of strategies, such as combining last seen performance with a dummy that compensates for long-term injuries, but this negatively affected the accuracy of our model. A possible time lag for the effect of injuries on market prices was another issue we encountered during data preparation. Due to delay in the recovery process, injuries may cause financial damage over time. Therefore, the change in market value caused by an injury may not occur entirely within the specified timeframe. A time series modelling technique could be an intriguing study strategy for the future.

Future research could include the perspective of actual transfer fees paid to review in how far clubs price in injury parameters in a deal. Moreover, with a sufficient data base, other less prominent leagues could be included in a model, to investigate whether our findings hold true there. Player positions could be regarded more differentiated, i.e., not just using the category ‘defender’, but ‘center back’, ‘left back’, ‘right back’, and so on. In addition, it would be intriguing to analyze especially goalkeepers and how injuries to the upper extremities affect them. Next to popularity, adding sentiment data to a model would be a captivating approach. Research has shown that using both the volume and sentiment of social media data can improve the accuracy of predictive models (Gayo-Avello 2013).

Bibliography

Ackermann, Phil, and Florian Follert. 2018. "Einige Bewertungstheoretische Anmerkungen Zur Marktwertanalyse Der Plattform Transfermarkt.De." *Sciamus - Sport Und Management* 9 (3): 21–41. https://dx.doi.org/10.22028/D291-32113.

Adler, Moshe. 1985. "Stardom and Talent." *The American Economic Review* 75 (1): 208–12. http://www.jstor.org/stable/1812714.

AFP. 2022. "PSG, Inter Milan and Juventus among Clubs Fined by UEFA for FFP Breaches." France24. September 2, 2022. https://www.france24.com/en/live-news/20220902-psg-inter-milan-and-juventus-among-clubs-fined-by-uefa-for-ffp-breaches.

Al-Asadi, M A, and S Tasdemır. 2022. "Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques." *IEEE Access* 10: 22631–45. https://doi.org/10.1109/ACCESS.2022.3154767.

Alto, Valentina. 2019. "Understanding the OLS Method for Simple Linear Regression." Towards Data Science. August 17, 2019. https://towardsdatascience.com/understanding-the-ols-method-for-simple-linear-regression-e0a4e8f692cc.

Andersen, T E, T W Floerenes, A Arnason, and R Bahr. 2004. "Video Analysis of the Mechanisms for Ankle Injuries in Football." *The American Journal of Sports Medicine* 32 (1 Suppl): 69S-79S. https://doi.org/10.1177/0363546503262023.

Arnason, A, A Gudmundsson, H A Dahl, and E Jóhannsson. 1996. "Soccer Injuries in Iceland." *Scandinavian Journal of Medicine & Science in Sports* 6 (1): 40–45. https://doi.org/10.1111/j.1600-0838.1996.tb00069.x.

Arnason, Arni, Stefan B Sigurdsson, Arni Gudmundsson, Ingar Holme, Lars Engebretsen, and Roald Bahr. 2004. “Risk Factors for Injuries in Football.” *The American Journal of Sports Medicine* 32 (1 Suppl): 5S-16S. https://doi.org/10.1177/0363546503258912.

Arrul, V Steve, P Subramanian, and R Mafas. 2022. “Predicting the Football Players’ Market Value Using Neural Network Model: A Data-Driven Approach.” In *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 1–7. https://doi.org/10.1109/ICDCECE53908.2022.9792681.

Attar, Wesam Saleh A al, Najeebullah Soomro, Evangelos Pappas, Peter J Sinclair, and Ross H Sanders. 2016. “How Effective Are F-MARC Injury Prevention Programs for Soccer Players? A Systematic Review and Meta-Analysis.” *Sports Medicine* 46 (2): 205–17. https://doi.org/10.1007/s40279-015-0404-x.

Bahr, Roald, Kristian Thorborg, and Jan Ekstrand. 2015. “Evidence-Based Hamstring Injury Prevention Is Not Adopted by the Majority of Champions League or Norwegian Premier League Football Teams: The Nordic Hamstring Survey.” *British Journal of Sports Medicine* 49 (22): 1466–71. https://doi.org/10.1136/bjsports-2015-094826.

Behravan, Iman, and Seyed Mohammad Razavi. 2021. “A Novel Machine Learning Method for Estimating Football Players’ Value in the Transfer Market.” *Soft Computing* 25 (3): 2499–2511. https://doi.org/10.1007/s00500-020-05319-3.

Beiderbeck, Daniel, Nicolas Evans, Nicolas Frevel, and Sascha Schmidt. 2023. “The Impact of Technology on the Future of Football – A Global Delphi Study.” *Technological Forecasting and Social Change* 187 (February): 122186. https://doi.org/10.1016/j.techfore.2022.122186.

Bizzini, Mario, and Jiri Dvorak. 2015. “FIFA 11+: An Effective Programme to Prevent Football Injuries in Various Player Groups Worldwide—a Narrative Review.” *British Journal of Sports Medicine* 49 (9): 577. https://doi.org/10.1136/bjsports-2015-094765.

Bizzini, Mario, Dave Hancock, and Franco Impellizzeri. 2012. “Suggestions from the Field for Return to Sports Participation Following Anterior Cruciate Ligament Reconstruction: Soccer.” *The Journal of Orthopaedic and Sports Physical Therapy* 42 (4): 304–12. https://doi.org/10.2519/jospt.2012.4005.

Bjørneboe, J, R Bahr, and T E Andersen. 2014. “Gradual Increase in the Risk of Match Injury in Norwegian Male Professional Football: A 6-Year Prospective Study.” *Scandinavian Journal of Medicine & Science in Sports* 24 (1): 189–96. https://doi.org/https://doi.org/10.1111/j.1600-0838.2012.01476.x.

Bonacchi, Massimiliano, Fabio Ciaponi, Antonio Marra, and Ron Shalev. 2021. “The Unintended Consequences of Accounting-Based Regulation: Real Effects on European Football Players Transfer Market.” https://doi.org/10.2139/ssrn.3978117.

Bryson, Alex, Bernd Frick, and B Simmons. 2013. “The Returns to Scarce Talent: Footedness and Player Remuneration in European Soccer.” *Journal of Sports Economics* 14 (6): 606–28. https://doi.org/10.1177/1527002511435118.

Caraffa, A, G Cerulli, M Projetti, G Aisa, and A Rizzo. 1996. “Prevention of Anterior Cruciate Ligament Injuries in Soccer.” *Knee Surgery, Sports Traumatology, Arthroscopy* 4 (1): 19–21. https://doi.org/10.1007/BF01565992.

Carmichael, Fiona, David Forrest, and Robert Simmons. 1999. “The Labour Market in Association Football: Who Gets Transferred and for How Much?” *Bulletin of Economic Research* 51 (2): 125–50. https://doi.org/https://doi.org/10.1111/1467-8586.00075.

Carreras-Simó, Miquel, and Jaume García Villar. 2018. "TV Rights, Financial Inequality, and Competitive Balance in European Football: Evidence from the English Premier League and the Spanish LaLiga." *International Journal of Sport Finance* 13 (3): 201–24. https://www.econbiz.de/Record/rights-financial-inequality-and-competitive-balance-european-football-evidence-from-the-english-premier-league-and-the-spanish-laliga-carreras-miquel/10011954633.

Catapult Sports. 2022. "The Cost of Injury." Catapult Sports. November 27, 2022. https://www.catapultsports.com/the-cost-of-injury.

CCyler, and CSmith1919. 2022. "Bayern Munich Boss Defends Use of Matthijs de Ligt at Striker in Gladbach Draw." Bavarian Football Works. August 28, 2022. https://www.bavarianfootballworks.com/2022/8/28/23324936/bayern-munich-boss-defends-use-of-matthijs-de-ligt-at-striker-in-gladbach-draw-lewandowski-haaland.

Charness, Gary, and Matthias Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26 (3): 157–76. https://doi.org/10.1257/jep.26.3.157.

Chomiak, J, A Junge, L Peterson, and J Dvorak. 2000. "Severe Injuries in Football Players. Influencing Factors." *The American Journal of Sports Medicine* 28 (5 Suppl): S58-68. https://doi.org/10.1177/28.suppl_5.s-58.

Cloke, David, Oliver Moore, Talib Shab, Steven Rushton, Mark D F Shirley, and David J Deehan. 2012. "Thigh Muscle Injuries in Youth Soccer: Predictors of Recovery." *The American Journal of Sports Medicine* 40 (2): 433–39. https://doi.org/10.1177/0363546511428800.

Cortegana, Mario. 2021. "Toni Kroos Exklusiv Über Seine Leidenszeit: 'Ich Habe Sechs Monate Unter Schmerzmitteln Gespielt.'" Goal. September 26, 2021.

https://www.goal.com/de/meldungen/real-madrid-toni-kroos-comeback-verletzung-academy/xhasnzaki0df16mpbiy070fw9.

Crafton, Adam, and Pol Ballus. 2022. “Investigation: Barcelonas Financial Crisis and What the Rest of Football Thinks of It.” The Athletic. August 3, 2022. https://theathletic.com/3468740/2022/08/03/barcelona-money-finances-crisis/.

Ćwiklinski, Bartosz, Agata Giełczyk, and Michał Choraś. 2021. “Who Will Score? A Machine Learning Approach to Supporting Football Team Building and Transfers.” *Entropy* 23 (1): 90. https://doi.org/10.3390/e23010090.

Dai, Boyi, Dewei Mao, William E Garrett, and Bing Yu. 2014. “Anterior Cruciate Ligament Injuries in Soccer: Loading Mechanisms, Risk Factors, and Prevention Programs.” *Journal of Sport and Health Science* 3 (4): 299–306. https://doi.org/https://doi.org/10.1016/j.jshs.2014.06.002.

Dawes, Robyn M, David Faust, and Paul E Meehl. 1989. “Clinical Versus Actuarial Judgment.” *Science* 243 (4899): 1668–74. https://doi.org/10.1126/science.2648573.

Dendir, Seife. 2016. “When Do Soccer Players Peak? A Note.” *Journal of Sports Analytics* 2: 89–105. https://doi.org/10.3233/JSA-160021.

Dimitropoulos, Panagiotis, and Evangelos Koumanakos. 2015. “Intellectual Capital and Profitability in European Football Clubs.” *International Journal of Accounting, Auditing and Performance Evaluation* 11 (2): 202–20. https://doi.org/10.1504/IJAAPE.2015.068862.

Dimitropoulos, Panagiotis, Stergios Leventis, and Emmanouil Dedoulis. 2016. “Managing the European Football Industry: UEFA’s Regulatory Intervention and the Impact on

Accounting Quality." *European Sport Management Quarterly* 16 (4): 459–86. https://doi.org/10.1080/16184742.2016.1164213.

Dobreff, Gergely, Alija Pašic, Balázs Sonkoly, and László Toka. 2019. "The Formation Game in Football." In *6th Workshop on Sports Analytics: Machine Learning and Data Mining for Sports Analytics (MLSA),* 1–11. https://www.researchgate.net/publication/345813301_The_formation_game_in_football.

Dobson, S M, and J A Goddard. 1998. "Performance and Revenue in Professional League Football: Evidence from Granger Causality Tests." *Applied Economics* 30 (12): 1641–51. https://doi.org/10.1080/000368498324715.

Doyle, Mark. 2022. "Lewandowski, Laporta & Levers: Barcelona Risk Bankruptcy with Biggest Bet in Club History." Goal. July 23, 2022. https://www.goal.com/en/lists/lewandowski-laporta-levers-barcelona-biggest-ever-bet-risk-bankruptcy/blt6f92a17ae964b22f#cseb6a1824ed4c0446.

Drawer, S, and C W Fuller. 2002. "An Economic Framework for Assessing the Impact of Injuries in Professional Football." *Safety Science* 40 (6): 537–56. https://doi.org/https://doi.org/10.1016/S0925-7535(01)00019-4.

Dvorak, J, and A Junge. 2000. "Football Injuries and Physical Symptoms. A Review of the Literature." *The American Journal of Sports Medicine* 28 (5 Suppl): S3-9. https://doi.org/10.1177/28.suppl_5.s-3.

Dyk, Nicol van, Roald Bahr, Angus F Burnett, Rod Whiteley, Arnhild Bakken, Andrea Mosler, Abdulaziz Farooq, and Erik Witvrouw. 2017. "A Comprehensive Strength Testing Protocol Offers No Clinical Value in Predicting Risk of Hamstring Injury: A Prospective Cohort Study of 413 Professional Football Players." *British Journal of Sports Medicine* 51 (23): 1695–1702. https://doi.org/10.1136/bjsports-2017-097754.

Ekstrand, J, M Hägglund, and M Waldén. 2011. “Injury Incidence and Injury Patterns in Professional Football: The UEFA Injury Study.” *British Journal of Sports Medicine* 45 (7): 553–58. https://doi.org/10.1136/bjsm.2009.060582.

Ekstrand, Jan, Martin Hägglund, Karolina Kristenson, Henrik Magnusson, and Markus Waldén. 2013. “Fewer Ligament Injuries but No Preventive Effect on Muscle Injuries and Severe Injuries: An 11-Year Follow-up of the UEFA Champions League Injury Study.” *British Journal of Sports Medicine* 47 (12): 732. https://doi.org/10.1136/bjsports-2013-092394.

Ekstrand, Jan, Markus Waldén, and Martin Hägglund. 2016. “Hamstring Injuries Have Increased by 4% Annually in Men’s Professional Football, since 2001: A 13-Year Longitudinal Analysis of the UEFA Elite Club Injury Study.” *British Journal of Sports Medicine* 50 (12): 731. https://doi.org/10.1136/bjsports-2015-095359.

Eliakim, Eyal, Elia Morgulev, Ronnie Lidor, and Yoav Meckel. 2020. “Estimation of Injury Costs: Financial Damage of English Premier League Teams’ Underachievement Due to Injuries.” *BMJ Open Sport &Amp; Exercise Medicine* 6 (1): e000675. https://doi.org/10.1136/bmjsem-2019-000675.

FC Bayern München. 2019. “Hände Und Hightech: Die Neuen Arztpraxen an Der Säbener Straße.” FC Bayern München. September 3, 2019. https://fcbayern.com/de/news/2019/09/haende-und-hightech-die-neuen-arztpraxen-an-der-saebener-strasse.

Felipe, Jose Luis, Alvaro Fernandez-Luna, Pablo Burillo, Luis Eduardo de la Riva, Javier Sanchez-Sanchez, and Jorge Garcia-Unanue. 2020. “Money Talks: Team Variables and Player Positions That Most Influence the Market Value of Professional Male Footballers in Europe.” *Sustainability* 12 (9): 3709. https://doi.org/10.3390/su12093709.

Fernández Cuevas, Ismael, Pedro Carmona, Manuel Quintana, Javier Salces, Javier Arnaiz-Lastras, and Antonio Barrón. 2010. "Economic Costs Estimation of Soccer Injuries in First and Second Spanish Division Professional Teams." In *15th Annual Congress of the European College of Sport Sciences ECSS*. https://www.researchgate.net/publication/258726802_Economic_costs_estimation_of_soccer_injuries_in_first_and_second_spanish_division_professional_teams.

FIFA. 2018a. "Behind the Scenes: Goal-Line Technology." FIFA. November 25, 2018. https://www.fifa.com/technical/football-technology/news/behind-the-scenes-goal-line-technology.

———. 2018b. "More than Half the World Watched Record-Breaking 2018 World Cup." FIFA. December 21, 2018. https://www.fifa.com/tournaments/mens/worldcup/2018russia/media-releases/more-than-half-the-world-watched-record-breaking-2018-world-cup.

Franck, Egon, and Stephan Nuesch. 2012. "Talent And/Or Popularity: What Does It Take To Be A Superstar?" *Economic Inquiry* 50 (1): 202–16. https://EconPapers.repec.org/RePEc:bla:ecinqu:v:50:y:2012:i:1:p:202-216.

Frenger, Monika, Florian Follert, Lukas Richau, and Eike Emrich. 2019. "Follow Me ... on the Relationship between Social Media Activities and Market Values in the German Bundesliga." Saarbrücken: Europäisches Institut für Sozioökonomie e. V. https://doi.org/10.22028/D291-32288.

Frick, Bernd. 2001. "Die Einkommen von „Superstars" Und „Wasserträgern" Im Professionellen Team-Sport – Ökonomische Analyse Und Empirische Befunde." *Journal of Business Economics : JBE* 71 (6): 701–20. https://www.econbiz.de/Record/die-einkommen-von-superstars-und-wasserträgern-im-professionellen-teamsport-ökomonische-analyse-und-empirische-befunde-frick-bernd/10001582834.

———. 2007. “The Football Players’ Labor Market: Empirical Evidence from the Major European Leagues.” *Scottish Journal of Political Economy* 54 (3): 422–46. https://doi.org/https://doi.org/10.1111/j.1467-9485.2007.00423.x.

Fuller, C W, E O Ojelade, and A Taylor. 2007. “Preparticipation Medical Evaluation in Professional Sport in the UK: Theory or Practice?” *British Journal of Sports Medicine* 41 (12): 890. https://doi.org/10.1136/bjsm.2007.038935.

Fuller, C W, G L Smith, A Junge, and J Dvorak. 2004. “The Influence of Tackle Parameters on the Propensity for Injury in International Football.” *The American Journal of Sports Medicine* 32 (1 Suppl): 43S-53S. https://doi.org/10.1177/0363546503261248.

Fuller, Colin W. 2019. “Assessing the Return on Investment of Injury Prevention Procedures in Professional Football.” *Sports Medicine* 49 (4): 621–29. https://doi.org/10.1007/s40279-019-01083-z.

Fuller, Colin W, and Richard D Hawkins. 1997. “Developing a Health Surveillance Strategy for Professional Footballers in Compliance with UK Health and Safety Legislation.” *British Journal of Sports Medicine* 31 (2): 148. https://doi.org/10.1136/bjsm.31.2.148.

Gallo, Amy. 2015. “A Refresher on Regression Analysis.” *Harvard Business Review*, November 4, 2015. https://hbr.org/2015/11/a-refresher-on-regression-analysis.

Gardenswartz, Lee, and Anita Rowe. 1994. *Diverse Teams at Work: Capitalizing on the Power of Diversity*. Chicago: Irwin Professional Pub.

Gayo-Avello, Daniel. 2013. "A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data." *Social Science Computer Review* 31 (6): 649–79. https://doi.org/10.1177/0894439313493979.

Gerhards, Jürgen, and Michael Mutz. 2017. "Who Wins the Championship? Market Value and Team Composition as Predictors of Success in the Top European Football Leagues." *European Societies* 19 (3): 223–42. https://doi.org/10.1080/14616696.2016.1268704.

Gerhards, Jürgen, Michael Mutz, and Gert G Wagner. 2014. "Die Berechnung Des Siegers: Marktwert, Ungleichheit, Diversität Und Routine Als Einflussfaktoren Auf Die Leistung Professioneller Fußballteams / Predictable Winners. Market Value, Inequality, Diversity, and Routine as Predictors of Success in European Soccer Leagues." *Zeitschrift Für Soziologie*, Zeitschrift für Soziologie, 43 (3): 231–50. https://doi.org/doi:10.1515/zfsoz-2014-0305.

Goldmann, Sven, and Markus Hesselmann. 2016. "Wolfgang Weber And The Goal That Never Was." *Tagesspiegel*, May 19, 2016. https://www.tagesspiegel.de/kultur/wolfgang-weber-and-the-goal-that-never-was-5216626.html.

Gómez, Miguel-Ángel, Carlos Lago, María-Teresa Gómez, and Philip Furley. 2019. "Analysis of Elite Soccer Players' Performance before and after Signing a New Contract." *PLOS ONE*. Vol. 14. Public Library of Science. https://doi.org/10.1371/journal.pone.0211058.

Grove, W M, D H Zald, B S Lebow, B E Snitz, and C Nelson. 2000. "Clinical versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 12 (1): 19–30. https://doi.org/10.1037/1040-3590.12.1.19.

Hägglund, Martin, Markus Waldén, and Jan Ekstrand. 2003. "Exposure and Injury Risk in Swedish Elite Football: A Comparison between Seasons 1982 and 2001." *Scandinavian*

Journal of Medicine & Science in Sports 13 (6): 364–70. https://doi.org/10.1046/j.1600-0838.2003.00327.x.

Harel, Ofer. 2009. “The Estimation of R 2 and Adjusted R 2 in Incomplete Data Sets Using Multiple Imputation.” *Journal of Applied Statistics* 36 (10): 1109–18. https://doi.org/10.1080/02664760802553000.

Hawkins, R D, and C W Fuller. 1996. “Risk Assessment in Professional Football: An Examination of Accidents and Incidents in the 1994 World Cup Finals.” *British Journal of Sports Medicine* 30 (2): 165–70. https://doi.org/10.1136/bjsm.30.2.165.

———. 1999. “A Prospective Epidemiological Study of Injuries in Four English Professional Football Clubs.” *British Journal of Sports Medicine* 33 (3): 196–203. https://doi.org/10.1136/bjsm.33.3.196.

Hawkins, R D, M A Hulse, C Wilkinson, A Hodson, and M Gibson. 2001. “The Association Football Medical Research Programme: An Audit of Injuries in Professional Football.” *British Journal of Sports Medicine* 35 (1): 43–47. https://doi.org/10.1136/bjsm.35.1.43.

He, Miao, Ricardo Cachucho, and Arno Knobbe. 2015. “Football Player’s Performance and Market Value.” In *Proceedings of the 2nd Workshop of Sports Analytics, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*. https://ceur-ws.org/Vol-1970/paper-11.pdf.

Hendricks, Wallace, Lawrence DeBrock, and Roger Koenker. 2003. “Uncertainty, Hiring, and Subsequent Performance: The NFL Draft.” *Journal of Labor Economics* 21 (4): 857–86. https://EconPapers.repec.org/RePEc:ucp:jlabec:v:21:y:2003:i:4:p:857-886.

Herm, Steffen, Hans-Markus Callsen-Bracker, and Henning Kreis. 2014. “When the Crowd Evaluates Soccer Players’ Market Values: Accuracy and Evaluation Attributes of an

Online Community." *Sport Management Review* 17 (4): 484–92. https://EconPapers.repec.org/RePEc:eee:spomar:v:17:y:2014:i:4:p:484-492.

HSE. 2022. "Historical Picture: Trends in Work-Related Ill Health and Workplace Injury in Great Britain." https://www.hse.gov.uk/statistics/history/historical-picture.pdf.

Hunt, M, and S Fulford. 1990. "Amateur Soccer: Injuries in Relation to Field Position." *British Journal of Sports Medicine* 24 (4): 265. https://doi.org/10.1136/bjsm.24.4.265.

Hutter, Michael. 2011. "Lucien Karpik: Valuing the Unique. The Economics of Singularities." *Journal of Cultural Economics* 35 (4): 315. https://doi.org/10.1007/s10824-011-9147-1.

Inan, Tugbay, and Levent Cavas. 2021. "Estimation of Market Values of Football Players through Artificial Neural Network: A Model Study from the Turkish Super League." *Applied Artificial Intelligence* 35 (13): 1022–42. https://doi.org/10.1080/08839514.2021.1966884.

Ivarsson, Andreas, and Urban Johnson. 2010. "Psychological Factors as Predictors of Injuries among Senior Soccer Players. A Prospective Study." *Journal of Sports Science & Medicine* 9 (2): 347–52. https://pubmed.ncbi.nlm.nih.gov/24149706/.

Jaakkola, H, J Henno, J Mäkelä, and B Thalheim. 2019. "Artificial Intelligence Yesterday, Today and Tomorrow." In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO),* 860–67. https://doi.org/10.23919/MIPRO.2019.8756913.

Jabbar, Haider Khalaf, and Rafiqul Zaman Khan. 2015. "Methods to Avoid Over-Fitting and Under-Fitting in Supervised Machine Learning (Comparative Study)." In *Computer*

Science, Communication and Instrumentation Devices, 70:163–72. https://doi.org/10.3850/978-981-09-5247-1_017.

Johns Hopkins Medicine. 2022. “Anterior Cruciate Ligament (ACL) Injury or Tear.” Johns Hopkins Medicine. 2022. https://www.hopkinsmedicine.org/health/conditions-and-diseases/acl-injury-or-tear.

Jones, Ashley, Gareth Jones, Neil Greig, Paul Bower, James Brown, Karen Hind, and Peter Francis. 2019. “Epidemiology of Injury in English Professional Football Players: A Cohort Study.” *Physical Therapy in Sport* 35: 18–22. https://doi.org/https://doi.org/10.1016/j.ptsp.2018.10.011.

Junge, A. 2000. “The Influence of Psychological Factors on Sports Injuries. Review of the Literature.” *The American Journal of Sports Medicine* 28 (5 Suppl): S10-5. https://doi.org/10.1177/28.suppl_5.s-10.

Junge, A, and J Dvorak. 2000. “Influence of Definition and Data Collection on the Incidence of Injuries in Football.” *The American Journal of Sports Medicine* 28 (5 Suppl): S40-6. https://doi.org/10.1177/28.suppl_5.s-40.

Junge, A, J Dvorak, and T Graf-Baumann. 2004. “Football Injuries during the World Cup 2002.” *The American Journal of Sports Medicine* 32 (1 Suppl): 23S-7S. https://doi.org/10.1177/0363546503261246.

Junge, A, J Dvorak, T Graf-Baumann, and L Peterson. 2004. “Football Injuries during FIFA Tournaments and the Olympic Games, 1998-2001: Development and Implementation of an Injury-Reporting System.” *The American Journal of Sports Medicine* 32 (1 Suppl): 80S-9S. https://doi.org/10.1177/0363546503261245.

Junge, Astrid, and Jiri Dvorak. 2004. “Soccer Injuries: A Review on Incidence and Prevention.” *Sports Medicine (Auckland, N.Z.)* 34 (13): 929–38. https://doi.org/10.2165/00007256-200434130-00004.

Kampakis, Stylianos. 2016. “Predictive Modelling of Football Injuries.” Thesis (Doctoral), London: University College London. https://arxiv.org/abs/1609.07480.

Kay, Oliver. 2022. “Newcastle’s Takeover: In Saudi Arabia, Exploring How the Club Fits a Country’s Vision.” The Athletic. October 6, 2022. https://theathletic.com/3652439/2022/10/06/newcastles-takeover-saudi-arabia/.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. “Lightgbm: A Highly Efficient Gradient Boosting Decision Tree.” In *31st Conference on Neural Information Processing Systems*, 30:1–9. https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Keppel, Pepijn, and Tom Claessens. 2020. “How the Volunteers of Data Website Transfermarkt Became Influential Players at European Top Football Clubs.” Follow the Money. December 18, 2020. https://www.ftm.eu/articles/transfermarkt-volunteers-european-football.

Kirschstein, T, and Steffen Liebscher. 2019. “Assessing the Market Values of Soccer Players – a Robust Analysis of Data from German 1. and 2. Bundesliga.” *Journal of Applied Statistics* 46 (7): 1336–49. https://doi.org/10.1080/02664763.2018.1540689.

Lee, Hansoo, Bayu Adhi Tama, and Meeyoung Cha. 2022. “Prediction of Football Player Value Using Bayesian Ensemble Approach.” https://doi.org/10.48550/arXiv.2206.13246.

Lehmann, Erik, and Günther Schulze. 2008. “What Does It Take to Be a Star? The Role of Performance and the Media for German Soccer Players.” *Applied Economics Quarterly*, Discussion Paper Series, 54 (1): 59–70. https://EconPapers.repec.org/RePEc:fre:wpaper:1.

Littkemann, Jörn, and Sebastian Kleist. 2002. “Sportlicher Erfolg in Der Fußball-Bundesliga: Eine Frage Der Auf- Oder Der Einstellung?” In *Sportökonomie*, edited by Horst Albach and Bernd Frick, 181–202. Wiesbaden: Gabler Verlag. https://doi.org/10.1007/978-3-663-07711-4_10.

Liu, Tianbiao, Lang Yang, Huimin Chen, and Antonio García-de-Alcaraz. 2021. “Impact of Possession and Player Position on Physical and Technical-Tactical Performance Indicators in the Chinese Football Super League.” *Frontiers in Psychology* 12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.722200.

López-Valenciano, Alejandro, Iñaki Ruiz-Pérez, Alberto Garcia-Gómez, Francisco J Vera-Garcia, Mark de Ste Croix, Gregory D Myer, and Francisco Ayala. 2020. “Epidemiology of Injuries in Professional Football: A Systematic Review and Meta-Analysis.” *British Journal of Sports Medicine* 54 (12): 711–18. https://doi.org/10.1136/bjsports-2018-099577.

Lucifora, Claudio, and Rob Simmons. 2003. “Superstar Effects in Sport.” *Journal of Sports Economics* 4 (1): 35–55. https://EconPapers.repec.org/RePEc:sae:jospec:v:4:y:2003:i:1:p:35-55.

Lundgårdh, Filip, Kjell Svensson, and Marie Alricsson. 2020. “Epidemiology of Hip and Groin Injuries in Swedish Male First Football League.” *Knee Surgery, Sports Traumatology, Arthroscopy* 28 (4): 1325–32. https://doi.org/10.1007/s00167-019-05470-x.

MacLennan, Tom. 2005. “Moneyball: The Art of Winning an Unfair Game.” *Journal of Popular Culture* 38 (4): 780–81. https://www.proquest.com/scholarly-journals/moneyball-art-winning-unfair-game/docview/195371313/se-2?accountid=28955.

Maglio, Roberto, and Andrea Rey. 2017. “The Impairment Test for Football Players: The Missing Link between Sports and Financial Performance?” *Palgrave Communications* 3 (1): 17055. https://doi.org/10.1057/palcomms.2017.55.

Majewski, Sebastian. 2016. “Identification of Factors Determining Market Value of the Most Valuable Football Players.” *Journal of Management and Business Administration. Central Europe* 24 (3): 91–104. https://doi.org/10.7206/jmba.ce.2450-7814.177.

Martín, Gracia Rubio, Ángel Rodríguez López, and Daniel Santín. 2019. “Valuation of Football Players in Financial Statements: The Power of the Crowd versus Transfer Fees.” In *EFMA Annual Meeting 2019 - Azores*, 1–34. https://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2019-Azores/papers/EFMA2019_0566_fullpaper.pdf.

Melkumova, L E, and S.Ya. Shatskikh. 2017. “Comparing Ridge and LASSO Estimators for Data Analysis.” *Procedia Engineering* 201: 746–55. https://doi.org/https://doi.org/10.1016/j.proeng.2017.09.615.

Michael Grubwinkler. 2020. “Kreuzbandriss.” Qualitätskliniken.De. May 15, 2020. https://www.qualitaetskliniken.de/erkrankungen/kreuzbandriss/.

Monteiro, Ricardo, Diogo Monteiro, Célia Nunes, Miquel Torregrossa, and Bruno Travassos. 2020. “Identification of Key Career Indicators in Portuguese Football Players.” *International Journal of Sports Science & Coaching* 15 (4): 533–41. https://doi.org/10.1177/1747954120923198.

Müller, Oliver, Alexander Simons, and Markus Weinmann. 2017. "Beyond Crowd Judgments: Data-Driven Estimation of Market Value in Association Football." *European Journal of Operational Research* 263 (2): 611–24. https://doi.org/https://doi.org/10.1016/j.ejor.2017.05.005.

Nielsen, A B, and J Yde. 1989. "Epidemiology and Traumatology of Injuries in Soccer." *The American Journal of Sports Medicine* 17 (6): 803–7. https://doi.org/10.1177/036354658901700614.

Oprean, Victor-Bogdan, and Tudor Oprisor. 2014. "Accounting for Soccer Players: Capitalization Paradigm vs. Expenditure." *Procedia Economics and Finance* 15 (December): 1647–54. https://doi.org/10.1016/S2212-5671(14)00636-4.

OrthoInfo. 2022. "Anterior Cruciate Ligament (ACL) Injuries." OrthoInfo. 2022. https://orthoinfo.aaos.org/en/diseases--conditions/anterior-cruciate-ligament-acl-injuries/.

Ottobock. 2022. "Kreuzbandriss - Ursachen, Symptome Und Behandlung." Ottobock. 2022. https://www.ottobock.com/de-de/situation/diagnosen-und-symptome/kreuzbandriss.

Pantuso, G, and L M Hvattum. 2021. "Maximizing Performance with an Eye on the Finances: A Chance-Constrained Model for Football Transfer Market Decisions." *TOP* 29 (2): 583–611. https://doi.org/10.1007/s11750-020-00584-9.

Pappalardo, Luca, Paolo Cintia, Paolo Ferragina, Emanuele Massucco, Dino Pedreschi, and Fosca Giannotti. 2019. "PlayeRank: Data-Driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach." *ACM Transactions on Intelligent Systems and Technology* 10 (5): 1–27. https://doi.org/10.1145/3343172.

Parry, Les, and Barry Drust. 2006. “Is Injury the Major Cause of Elite Soccer Players Being Unavailable to Train and Play during the Competitive Season?” *Physical Therapy in Sport* 7 (2): 58–64. https://doi.org/https://doi.org/10.1016/j.ptsp.2006.03.003.

Partosch, Christoph. 2013. “Der Einfluss Der Champions League Auf Den Marktwert Eines Bundesligaklubs Und Das (Transfer-)Verhalten Des Managements.” https://EconPapers.repec.org/RePEc:zbw:umiodp:92013.

Pawlowski, Tim, Christoph Breuer, and Arnd Hovemann. 2010. “Top Clubs’ Performance and the Competitive Situation in European Domestic Football Competitions.” *Journal of Sports Economics* 11 (2): 186–202. https://doi.org/10.1177/1527002510363100.

Pedace, Roberto. 2007. “Earnings, Performance, and Nationality Discrimination in a Highly Competitive Labor Market as An Analysis of the English Professional Soccer League.” *Journal of Sports Economics* 9 (2): 115–40. https://doi.org/10.1177/1527002507301422.

Peterson, L, A Junge, J Chomiak, T Graf-Baumann, and J Dvorak. 2000. “Incidence of Football Injuries and Complaints in Different Age Groups and Skill-Level Groups.” *The American Journal of Sports Medicine* 28 (5 Suppl): S51-7. https://doi.org/10.1177/28.suppl_5.s-51.

Press Association. 2022. “Everton Confirm Departure of Allan to Al Wahda for Undisclosed Fee.” FourFourTwo. September 27, 2022. https://www.fourfourtwo.com/news/everton-confirm-departure-of-allan-to-al-wahda-for-undisclosed-fee-1664283093000.

PriceWaterhouseCoopers. 2018. “Accounting for Typical Transactions in the Football Industry - Issues and Solutions under IFRS.” PriceWaterhouseCoopers. October 2018. https://www.pwc.com/gx/en/audit-services/ifrs/publications/ifrs-9/accounting-for-typical-transactions-in-the-football-industry.pdf.

Richau, Lukas, Florian Follert, Monika Frenger, and Eike Emrich. 2019. “Performance Indicators in Football: The Importance of Actual Performance for the Market Value of Football Players.” *Sciamus - Sport Und Management* 10 (4): 41–67. https://www.researchgate.net/publication/338007931_Performance_indicators_in_football_The_importance_of_actual_performance_for_the_market_value_of_football_players_in_Sciamus_-_Sport_und_Management_104_41-67.

Robinson, Leigh. 2008. “The Business of Sport.” In *Sport and Society: A Student Introduction*, 307–27. https://doi.org/10.4135/9781446278833.n14.

Rohde, Marc, and Christoph Breuer. 2016. “Europe’s Elite Football: Financial Growth, Sporting Success, Transfer Investment, and Private Majority Investors.” *International Journal of Financial Studies* 4 (2): 12. https://doi.org/10.3390/ijfs4020012.

Rosen, Sherwin. 1981. “The Economics of Superstars.” *The American Economic Review* 71 (5): 845–58. http://www.jstor.org/stable/1803469.

Rossi, Alessio, Luca Pappalardo, Paolo Cintia, F Marcello Iaia, Javier Fernàndez, and Daniel Medina. 2018. “Effective Injury Forecasting in Soccer with GPS Training Data and Machine Learning.” *PLOS ONE* 13 (7): e0201264-. https://doi.org/10.1371/journal.pone.0201264.

Russell, Stuart J. 2010. *Artificial Intelligence : A Modern Approach*. 3rd ed. Upper Saddle River, N.J.: Prentice Hall.

Sadigursky, David, Juliana Almeida Braid, Diogo Neiva Lemos de Lira, Bruno Almeida Barreto Machado, Rogério Jamil Fernandes Carneiro, and Paulo Oliveira Colavolpe. 2017. “The FIFA 11+ Injury Prevention Program for Soccer Players: A Systematic Review.” *BMC Sports Science, Medicine and Rehabilitation* 9 (1): 18. https://doi.org/10.1186/s13102-017-0083-z.

Savarez03. 2021. “Marktwertdefinition.” Transfermarkt. August 25, 2021. https://www.transfermarkt.de/-marktwertdefinition/thread/forum/67/thread_id/237454.

Schmid, Michael J, Achim Conzelmann, and Claudia Zuber. 2020. “Patterns of Achievement-Motivated Behavior and Performance as Predictors for Future Success in Rowing: A Person-Oriented Study.” *International Journal of Sports Science & Coaching* 16 (1): 101–9. https://doi.org/10.1177/1747954120953658.

Schokkaert, Jeroen. 2016. “Football Clubs’ Recruitment Strategies and International Player Migration: Evidence from Senegal and South Africa.” *Soccer & Society* 17 (1): 120–39. https://doi.org/10.1080/14660970.2014.919271.

Schroer, Alyssa. 2022. “How Sports Analytics Are Used Today, by Teams and Fans.” Built In. August 26, 2022. https://builtin.com/big-data/big-data-companies-sports.

Shcherbakov, Maxim Vladimirovich, Adriaan Brebels, Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky, and Valeriy Anatol’evich Kamaev. 2013. “A Survey of Forecast Error Measures.” *World Applied Sciences Journal* 24 (24): 171–76. http://idosi.org/wasj/wasj(ITMIES)13/28.pdf.

SID. 2011. “Verträge Interessieren Die Fußball-Söldner Nicht Mehr.” Welt. January 7, 2011. https://www.welt.de/sport/fussball/bundesliga/fc-schalke-04/article12006375/Vertraege-interessieren-die-Fussball-Soeldner-nicht-mehr.html.

Simmons, Joseph P, Leif D Nelson, Jeff Galak, and Shane Frederick. 2011. “Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds.” *Journal of Consumer Research* 38 (1): 1–15. https://doi.org/10.1086/658070.

Singh, Prabhnoor, and Puneet Singh Lamba. 2019. “Influence of Crowdsourcing, Popularity and Previous Year Statistics in Market Value Estimation of Football Players.” *Journal of*

Discrete Mathematical Sciences and Cryptography 22 (2): 113–26. https://doi.org/10.1080/09720529.2019.1576333.

Sportlexikon. 2022. “Fußballtaktik Catenaccio.” Sportlexikon. April 25, 2022. https://www.sportlexikon.com/fussball-catenaccio.

SPOX. 2020. “Cruyff, Maradona, Neymar: Transferrekorde Im Laufe Der Jahre.” SPOX. May 19, 2020. https://www.spox.com/de/sport/fussball/international/1908/Diashows/transferrekorde-historische-entwicklung/neymar-ronaldo-zinedine-zidane-gareth-bale-paul-poga-diego-maradona.html.

Stubbe, Janine H, Anne-Marie M C van Beijsterveldt, Sissi van der Knaap, Jasper Stege, Evert A Verhagen, Willem van Mechelen, and Frank J G Backx. 2015. “Injuries in Professional Male Soccer Players in the Netherlands: A Prospective Cohort Study.” *Journal of Athletic Training* 50 (2): 211–16. https://doi.org/10.4085/1062-6050-49.3.64.

Szymanski, Stefan, and Ron Smith. 1997. “The English Football Industry: Profit, Performance and Industrial Structure.” *International Review of Applied Economics* 11 (1): 135–53. https://doi.org/10.1080/02692179700000008.

Szymski, Dominik, Leonard Achenbach, Johannes Weber, Lorenz Huber, Clemens Memmel, Maximilian Kerschbaum, Volker Alt, and Werner Krutsch. 2022. “Reduced Performance after Return to Competition in ACL Injuries: An Analysis on Return to Competition in the ‘ACL Registry in German Football.’” *Knee Surgery, Sports Traumatology, Arthroscopy*. https://doi.org/10.1007/s00167-022-07062-8.

Transfermarkt. 2022a. “Profile - Ousmane Dembélé.” Transfermarkt. 2022. https://www.transfermarkt.com/schnellsuche/ergebnis/schnellsuche?query=ousmane+dembele.

———. 2022b. “Profile - Tim Leibold.” Transfermarkt. 2022. https://www.transfermarkt.com/tim-leibold/profil/spieler/185699.

———. 2022c. “Profile - Vinicius Junior.” Transfermarkt. 2022. https://www.transfermarkt.com/vinicius-junior/profil/spieler/371998.

———. 2022d. “Transfermarkt - Start.” LinkedIn. November 22, 2022. https://www.linkedin.com/company/transfermarkt-gmbh-&-co-kg/.

———. 2022e. “Detailed Stats of Angeliño.” Transfermarkt. November 23, 2022. https://www.transfermarkt.com/angelino/leistungsdatendetails/spieler/277179.

———. 2022f. “Detailed Stats of James Milner.” Transfermarkt. November 23, 2022. https://www.transfermarkt.com/james-milner/leistungsdatendetails/spieler/3333.

———. 2022g. “Detailed Stats of Mario Gómez.” Transfermarkt. November 23, 2022. https://www.transfermarkt.com/mario-gomez/leistungsdatendetails/spieler/6288.

———. 2022h. “Detailed Stats of Nordi Mukiele.” Transfermarkt. November 23, 2022. https://www.transfermarkt.com/nordi-mukiele/leistungsdatendetails/spieler/348026.

———. 2022i. “Profile - Jack Wilshere.” Transfermarkt. November 27, 2022. https://www.transfermarkt.com/jack-wilshere/profil/spieler/74223.

Trequattrini, Raffaele, Rosa Lombardi, and Fabio Nappo. 2012. “The Evaluation of the Economic Value of Long Lasting Professional Football Player Performance Rights.” *WSEAS Transactions on Business and Economics* 9 (4): 199–218. https://wseas.com/journals/bae/2012/54-741.pdf.

Tversky, Amos, and Daniel Kahneman. 1974. “Judgment under Uncertainty: Heuristics and Biases.” *Science* 185 (4157): 1124–31. https://doi.org/10.1126/science.185.4157.1124.

UW Data Science. 2016. “ The Story of Moneyball Proves Importance of Both Big Data and Big Ideas.” University of Wisconsin. August 24, 2016. https://datasciencedegree.wisconsin.edu/blog/moneyball-proves-importance-big-data-big-ideas/.

Villa, Francesco della, Bert Mandelbaum, and Lawrence Lemak. 2018. “The Effect of Playing Position on Injury Risk in Male Soccer Players: Systematic Review of the Literature and Risk Considerations for Each Playing Position.” *American Journal of Orthopedics* 47 (10). https://doi.org/10.12788/ajo.2018.0092.

Weimar, Daniel, and Pamela Wicker. 2017. “Moneyball Revisited: Effort and Team Performance in Professional Soccer.” *Journal of Sports Economics* 18 (2): 140–61. https://doi.org/10.1177/1527002514561789.

Wheatley, Chris. 2021. “The Secrets behind Transfermarkt and How Football Clubs and Players Use the Platform.” Football.London. October 24, 2021. https://www.football.london/premier-league/secrets-behind-transfermarkt-how-football-21956019?utm_source=linkCopy&utm_medium=social&utm_campaign=sharebar.

Wicker, Pamela, Joachim Prinz, Daniel Weimar, Christian Deutscher, and Thorsten Upmann. 2013. “No Pain, No Gain? Effort and Productivity in Professional Soccer.” *International Journal of Sport Finance* 8 (2): 124–39. https://EconPapers.repec.org/RePEc:jsf:intjsf:v:8:y:2013:i:2:p:124-139.

Wolfers, Justin, and Eric Zitzewitz. 2004. “Prediction Markets.” *Journal of Economic Perspectives* 18 (2): 107–26. https://doi.org/10.1257/0895330041371321.

Worville, Tom. 2021. “Inside De Bruyne’s Data Report: Sancho Comparison and Impact of Playmaker’s Possible City Exit Crucial to New Deal.” *The Athletic*, April 12, 2021. https://theathletic.com/2509349/2021/04/12/inside-de-bruynes-data-report-sancho-comparison-and-impact-of-playmakers-possible-city-exit-crucial-to-new-deal/?redirected=1.

Yang, Yun. 2017. "Chapter 4 - Ensemble Learning." In *Temporal Data Mining Via Unsupervised Ensemble Learning*, edited by Yun Yang, 35–56. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-811654-8.00004-X.

Yi, Qing, Hong Jia, Hongyou Liu, and Miguel Ángel Gómez. 2018. "Technical Demands of Different Playing Positions in the UEFA Champions League." *International Journal of Performance Analysis in Sport* 18 (6): 926–37. https://doi.org/10.1080/24748668.2018.1528524.

Yiğit, Ahmet Talha, Barış Samak, and Tolga Kaya. 2020. "Football Player Value Assessment Using Machine Learning Techniques." In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making*, edited by Cengiz Kahraman, Selcuk Cebi, Sezi Cevik Onar, Basar Oztaysi, A Cagri Tolga, and Irem Ucal Sari, 289–97. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-23756-1_36.

Ziebs, Alexander. 2004. "Ist Sportlicher Erfolg Käuflich? Eine Diskriminanzanalytische Untersuchung Der Zentralen Erfolgsfaktoren in Der Fußball-Bundesliga / Unlimited Venality of Sporting Success? A Differentiating Analysis of Success Related Factors Concerning the First German Soccer League." *Sport Und Gesellschaft*, Sport und Gesellschaft, 1 (1): 30–49. https://doi.org/doi:10.1515/sug-2004-0104.

Zuber, Claudia, Marc Zibung, and Achim Conzelmann. 2015. "Motivational Patterns as an Instrument for Predicting Success in Promising Young Football Players." *Journal of Sports Sciences* 33 (2): 160–68. https://doi.org/10.1080/02640414.2014.928827.

Appendix

Most important factors:
Future prospects
Age
Sporting achievements in club and national team
Level and importance of the league, both athletically and financially
Reputation/prestige/character traits
Development potential
League-specific characteristics
Marketing value
Number & Reputation of interested parties
Performance potential
Experience Level
Injury susceptibility
Different financial conditions of clubs and leagues
General demand and trends on the market
General development of transfer fees
External factors such as the Corona pandemic and its consequences
Individual transfer modalities:
Transfer by means of purchase option/compulsory purchase
Loan fee
Only part of transfer rights acquired
Exit clause
Repurchase option
Player swap/offset
Contract length
Resale participation
Bonuses
Embellishment of the financial balance sheet
Situational conditions:
Pressure situations such as competitive, success or financial pressure, etc.
Will/desire/interests of the player
Club does not sell to highest bidder
Player goes on strike or similar
High salary
Club wants to sell player

Value drivers of market values on Transfermarkt.com (Savarez03 2021)

Season	Injury Type	Start Date	End Date	Duration in days	Missed Games
20/21	Calf Injury	Sep 11, 2020	Sep 22, 2020	11	4
19/20	Groin Injury	Oct 23, 2019	May 31, 2020	221	22
19/20	Groin Injury	Aug 11, 2019	Aug 16, 2019	5	0
18/19	Ankle Surgery	Dec 2, 2018	Apr 19, 2019	138	22
18/19	Ankle Injury	Sep 9, 2018	Nov 30, 2018	82	11
16/17	Hairline crack in calf bone	Apr 16, 2017	Aug 6, 2017	112	1
16/17	Knee Injury	Jul 31, 2016	Aug 11, 2016	11	0
15/16	Hairline crack in calf bone	Aug 1, 2015	Apr 4, 2016	247	47
14/15	Ankle Injury	Feb 23, 2015	Mar 31, 2015	36	7
14/15	Ankle Injury	Nov 22, 2014	Feb 20, 2015	90	19
13/14	Hairline crack in the foot	Mar 5, 2014	May 1, 2014	57	11
13/14	Sprained Ankle	Nov 2, 2013	Nov 8, 2013	6	2
12/13	Ankle Injury	Mar 11, 2013	Apr 10, 2013	30	4
11/12	Knee Surgery	May 22, 2012	Sep 24, 2012	125	6
11/12	Fatigue fracture	Feb 1, 2012	May 15, 2012	104	19
11/12	Ankle Surgery	Sep 26, 2011	Jan 31, 2012	127	25
11/12	Sprained Ankle	Jul 31, 2011	Sep 25, 2011	56	10
09/10	Ankle Injury	Nov 20, 2009	Dec 2, 2009	12	4

Jack Wilshere injury overview
Categories: green = groin injury, orange = calf, blue = knee, purple = ankle

Table of Tables

Table of Figures