AMCIS 2022 Proceedings

Conference Theme Track - Innovative Research Informing Practice

Aug 10th, 12:00 AM

Towards a Reliable & Transparent Approach to Data-Driven Brand Valuation

Matthew Caron
Paderborn University, matthew.caron@uni-paderborn.de

Christian Bartelheimer
Paderborn University, christian.bartelheimer@uni-paderborn.de

Oliver Mueller
Paderborn University, oliver.mueller@uni-paderborn.de

Towards a Reliable & Transparent Approach to Data-Driven Brand Valuation

Completed Research

Matthew Caron
Paderborn University
Paderborn, Germany
matthew.caron@upb.de

Christian Bartelheimer
Paderborn University
Paderborn, Germany
christian.bartelheimer@upb.de

Oliver Müller
Paderborn University
Paderborn, Germany
oliver.mueller@upb.de

Abstract

Now accounting for more than 80% of a firm's worth, brands have become essential assets for modern organizations. However, methods and techniques for the monetary valuation of brands are still under-researched. Hence, the objective of this study is to evaluate the utility of explanatory statistical models and machine learning approaches for explaining and predicting brand value. Drawing upon the case of the most valuable English football brands during the 2016/17 to 2020/21 seasons, we demonstrate how to operationalize Aaker's (1991) theoretical brand equity framework to collect meaningful qualitative and quantitative feature sets. Our explanatory models can explain up to 77% of the variation in brand valuations across all clubs and seasons, while our predictive approach can predict out-of-sample observations with a mean absolute percentage error (MAPE) of 14%. Future research can build upon our results to develop domain-specific brand valuation methods while enabling managers to make better-informed investment decisions.

Keywords

Brand valuation, explanatory models, predictive modeling, football.

Introduction

Generally speaking, brands are defined as "a name, term, design, symbol or any other feature that identifies one seller's goods or service as distinct from those of other sellers" (American Marketing Association 2022). From being used for marking cattle to being one of the most significant drivers of shareholder value, brands have become, from a financial perspective, essential assets for modern organizations (Lindemann 2010). In point of fact, triggered by the rise of mass production at the end of the nineteenth century, the value of brands now accounts for more than 80% of a firm's worth (Lindemann 2010). However, while brands have been, for more than 30 years, high on the business agenda, (Rojas-Lamorena et al. 2022), "brands are [still] one of the most valuable yet least understood assets" (ISO 2019, p.V).

An especially challenging aspect of branding is brand valuation, that is, the process of estimating the monetary value of a brand. Although international standards like ISO 20671 (ISO 2019) and ISO 10668 (ISO 2010) specify the general principles as well as procedures and methods of brand valuation, there exist significant differences between the valuations proposed by leading brand valuation consultancies — i.e., Interbrand, Millward Brown, and Brand Finance – and the actual amounts that were paid for acquiring various brands. In fact, a comparative study comprising 162 cases found an average discrepancy of 288% between the assessments of the three leading consultancies and the actual paid amounts (Markables 2015).

Against this background, the objective of this study is to evaluate the utility of statistical modeling and machine learning approaches for explaining and predicting brand value. We pursue this objective by leveraging the empirical setting of football brands. More specifically, based on the theoretical framework of brand equity proposed by Aaker (1991), we acquired (a) the brand valuations of the most valuable football brands in the English Premier League for five consecutive seasons (2016/17 – 2020/21) and (b) a set of quantitative and qualitative features representing each dimension of the brand equity model – e.g., the total number of Champions League matches (*Brand Awareness*), the total number of fan-generated tweets (*Brand Loyalty*), or the total numbers of historical FA Cup titles (*Perceived Quality*). Employing various natural language processing techniques, we augmented our dataset with a set of extracted and generated features representing qualitative information about these brands. We then fitted and evaluated a series of explanatory statistical models for each brand equity dimension and a predictive machine learning model including all features across brand equity dimensions. Our explanatory models can explain up to 77% of the variation in brand valuations across all clubs and seasons, while our global predictive approach can predict out-of-sample observations (time-based split) with a mean absolute percentage error (MAPE) of 14%.

Overall, this study makes three main contributions: (1) we demonstrate how to operationalize the theoretical brand equity framework proposed by Aaker (1991) in the form of explanatory statistical models, (2) we empirically show that proprietary brand valuations of football clubs can largely be explained and predicted by features extracted from publicly-available data sources, (3) we provide insights about the relative importance of each brand equity dimension as well as individual features. These findings have important implications for research and practice at the intersection of Marketing, Finance, and Information Systems. They can, for instance, serve as a foundation for developing automated, transparent, and reproducible brand valuation methods and tools and, in turn, inform managers about how to best improve the value of their brands through evidence-based investment decisions.

The remainder of this paper is structured as follows: In Section 2, we position our study in the current knowledge base on brand equity and brand valuation practices. In Section 3, we describe our research and data collection process. In Section 4, we present the results of our explanatory and predictive analyses. Lastly, Section 5 discusses the implications for research and management and concludes our paper.

Theoretical Background

Brand Value and Brand Equity

Creating successful brands and strategically managing them to meet customer expectations enables organizations to gain a competitive advantage (Doyle 1989). Specifically, the strategic brand management process encompasses four phases: to identify and establish a brand, to plan and implement brand marketing programs, to measure and interpret brand performance, and to grow and sustain brand equity (Keller and Brexendorf 2019). Hence, to accomplish this goal, organizations need to measure brand performance and interpret how brand equity is built – i.e., to understand how it can be improved through strategic investments in brand equity (Keller and Brexendorf 2019; Raggio and Leone 2007).

Brand equity is a construct that is not yet conceptually agreed upon in the literature and needs further theoretical investigation (Rojas-Lamorena et al. 2022). Two often confused but conceptually different concepts are brand value and brand equity. Brand value is the monetized value an organization can gain from selling or replacing a brand (Raggio and Leone 2007). In contrast, brand equity is describing the dimensions that constitute brand value, making up "the value of a brand beyond what can be explained by a product's functional features. Brand equity leads to greater consumer preference, loyalty, and, ultimately, profits" (Bettencourt 2017). Hence, brand equity can be leveraged to generate brand value if it positively contributes to the consumer perception of a brand (Raggio and Leone 2007; Raggio and Leone 2009). However, the relationship between brand equity and brand value is not necessarily relational (Raggio and Leone 2007).

Generally speaking, there exist two motivations to study brand equity. In truth, one might study brand equity for accounting purposes – i.e., to determine the value of a brand accurately in terms of a monetized brand value – or for strategic objectives – i.e., to improve the performance of marketing investments (Keller

1993). If an organization is interested in strategic brand management, brand equity must be viewed from a customer's perspective to make investments that meet customers' expectations. "Customer-based brand equity is defined as the differential effect of brand knowledge on consumer response to the marketing of the brand" (Keller 1993, p.2) and consists of multiple dimensions, namely *Brand Associations*, *Brand Awareness*, *Brand Loyalty*, *Perceived Quality*, and *Other Proprietary* (Aaker 1991). The latter dimension refers to concepts that do not directly relate to one of the other four dimensions but also impact brand equity. These dimensions are described in Table 1.

Brand Associations	**Brand Awareness**	**Brand Loyalty**	**Perceived Quality**
Feelings, thoughts, experiences, attitudes, perceptions, beliefs, and images related to a brand	Individual's ability to recognize and recall a brand	Individual's attachment to a brand	Individual's judgement about a brand's quality that goes beyond functional and objective measurement

Table 1. Dimensions of Brand Equity According to Aaker (1991)

Brand Valuation Practices

Traditional practices for brand valuation can be classified into three general types (Salinas and Ambler 2009). First, cost-based approaches for brand valuation measure costs that have been incurred to establish a brand, respectively, those that would occur for establishing the brand again. Second, market-based approaches can be applied if data about the acquisition of a similar brand in a similar market is available and can be used to assess brand value (Salinas and Ambler 2009). While the first method is past-oriented and unsuitable for strategic brand management, market-based approaches are often hard to conduct due to missing data – i.e., detailed sales data of a similar brand. Hence, brand value is, nowadays, mostly determined by income-based approaches – i.e., approaches that aim to attribute potential future cash flows to a brand's characteristics (Salinas and Ambler 2009).

While a plethora of different methods for income-based brand valuation exists (Salinas and Ambler 2009), we focus on those that have been proposed by the ISO Project Committee for Brand Valuation Standardization and that are described in the ISO 20671 and ISO 10668 manuscripts (ISO 2010; ISO 2019). We assume that these methods are commonly used to determine customer-based brand equity in organizations while they correspond with the general requirements for brand valuation comprising transparency, validity, reliability, sufficiency, and objectivity (ISO 2010), it becomes apparent that none of these fulfils all of the requirements comprehensively.

The current knowledge base also comprises a few data-driven approaches for brand valuation. Cole (2012) developed a technique that uses publicly available data from Google Trends API. The analysis reveals that the number of searches for a particular brand on Google at time t positively correlates with its brand value at time $t+1$. Nuortimo and Harkonen (2019) perform an exploratory study by calculating the sentiment of social media posts and news articles for different brands to identify their brand image (Forbes-Brand-Index). They conclude that automated brand image calculation is possible but subject to conceptual and technical challenges that need to be addressed in future research. Concerning brand valuation in sports, Biscaia et al. (2016) analyze the differences in brand perception of fan club and non-fan club members, identifying social interaction, team success, and internalization to impact behavioral intentions positively. Wang and Tang (2018) developed a dual identification model to study the formation of team brand equity in Asian sports. Their results suggest that identifying with the team and identifying with the team brand are predictors of brand equity. Wetzel et al. (2018) investigate time-referenced aspects of building sales-based brand equity in sports and their impact on the development of market performance over time, concluding that high brand age impacts sales-based brand equity. Additionally, Huang et al. (2020) use 109 annual reports and conduct text and regression analysis to predict brand equity. They assume that, through analyzing annual reports, an organization's strategic orientations can be identified to determine its customer orientation. Their results indicate that brand equity can be predicted by identifying the degree of customer orientation.

Data & Empirical Methodology

Empirical Context

As mentioned in the introductory section of this paper, we focus on the case of brand valuation in professional football, or more precisely, the valuation of the most valuable football brands in the English Premier League (EPL)[1] for five consecutive seasons – i.e., from July 2016 to June 2021. Likely considered to be less conventional, sports brands often lack the financial fundamentals and stock market data usually relied on for valuation purposes, while, however, benefiting from high customer attention and exposure. The absence of such indicators emphasizes the need for taking a holistic view of brands, as proposed by Aaker (1991), making this case highly interesting when investigating all possible avenues for the transparent and data-driven appraisal of brands. Furthermore, a comprehensive literature review of academic research on brand equity by Rojas-Lamorena et al. (2022) underlined the recent emergence and relevance of sport-related topics in this rapidly evolving field of research, even recommending future studies to assess, specifically for sponsorships and merchandising purposes, the brand equity of sports clubs (2022).

Modeling Approach

Methodologically, we applied both explanatory and predictive modeling approaches to the above case (Shmueli 2010; Shmueli and Koppius 2011). First, for the explanatory analysis, we used linear mixed-effects models (also known as multi-level or hierarchical models) with varying intercepts for the predictors club and season to account for the panel structure of our data. In addition to these intercepts, we employed features representing a club's *Brand Association*, *Brand Awareness*, *Brand Loyalty*, *Perceived Quality*, and *Other Proprietary* factors. However, even though linear models are inherently interpretable by human decision makers, they potentially have lower predictive accuracy than more flexible machine learning methods. Hence, to assess how accurately one can predict a football club's brand value and, as a result, evaluate the practical application of the presented approach, we trained a tree-based gradient boosting model – i.e., XGBoost (Chen and Guestrin 2016) – using the first four seasons present in our dataset and evaluated its performance on the last season – i.e., time-based split. The predictive model was trained using the features exposed in the explanatory analysis while excluding the *club* and *season* predictors to approximate real-world conditions. Lastly, we optimized the model's hyperparameters using a Gaussian-based Bayesian optimization algorithm to ensure maximal performance (Bergstra et al. 2013).

Data Collection

Response Variable

As disclosed above, our analysis revolves around the most valuable football brands in the EPL for the seasons 2016/17 to 2020/21. To that end, we acquired, from the brand valuation and strategy consultancy Brand Finance[2], the so-called *Football 50* series of football brands rankings[3]. These historical brand valuations, which were calculated at the end of every season with the help of the previously-exposed valuation approaches, serve as the response variable in the present study and range from 122 million to 1.9 billion US dollars with a mean valuation of 535 million US dollars. In total, our panel sample contains 88 club-season observations, including 24 unique football clubs and five seasons. As not every one of the 24 clubs made it into the Top-50 every year, our dataset includes only 88 instead of the 120 (24 * 5) theoretically possible observations. Due to data availability and data consistency constraints, we limited our analysis to the aforementioned five-year span.

Features

Following the recommendations by Aaker (1991), we first started, as exposed in Figure 1, by defining a set of domain-specific concepts for each of the brand equity dimensions. These concepts were inspired by various

[1] https://www.premierleague.com
[2] https://brandfinance.com
[3] https://brandirectory.com/rankings/football/overview

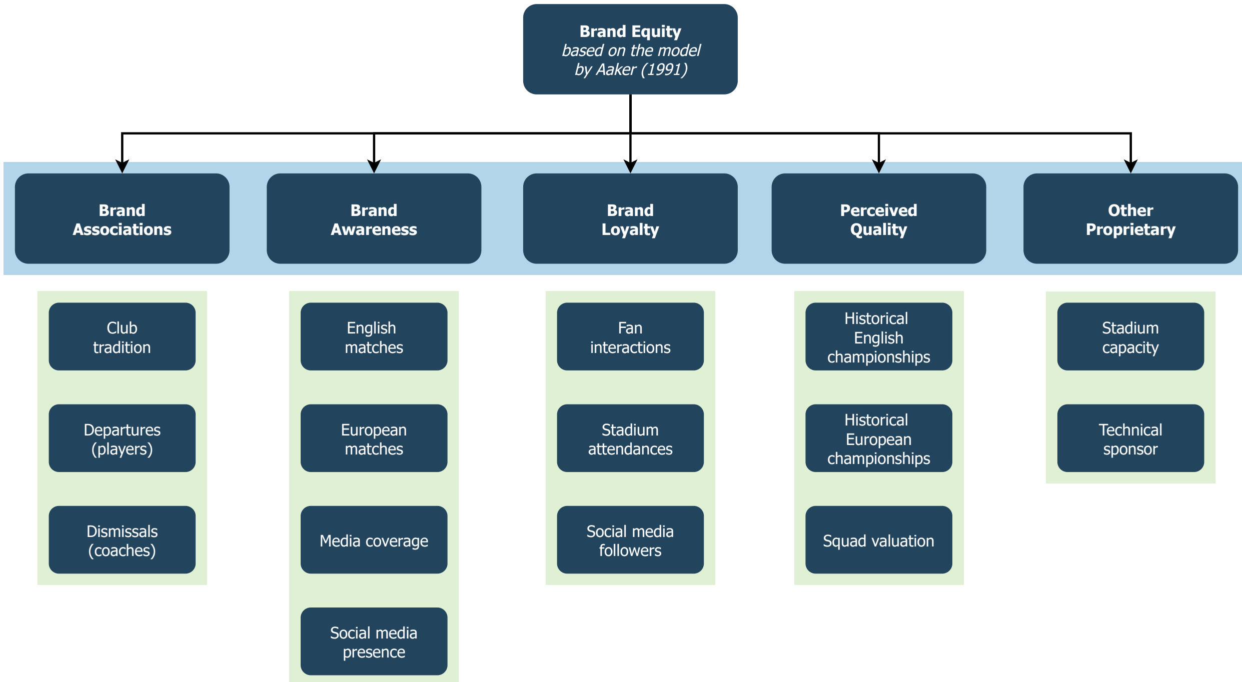

Figure 1. Brand Equity in Professional Football (Premier League)

consultancy reports as well as several academic works that investigate brand equity in the context of sports (see, for instance, Biscaia et al. 2016; Wang and Tang 2018; Wetzel et al. 2018). This first identification step is, in our opinion, crucial to bringing some structure to the data collection process and serves as a guide for all subsequent steps.

Quantitative Features

As can be observed in Table 2, we gathered, for every observation in our dataset, a set of quantitative features describing the dimensions of brand equity in professional football, as defined in the above conceptualization. These features originate exclusively from Transfermarkt[4] – i.e., one of the leading websites when it comes to football statistics, rumors, and transfers – and are all football-related – e.g., the total number of Champions League matches (*Brand Awareness*), the total number of spectators (*Brand Loyalty*), or the total numbers of historical FA Cup titles (*Perceived Quality*). Moreover, we also collected the yearly change in the number of Twitter and Reddit subscribers to the clubs' official Twitter accounts and subreddits for every season as well as the total number of likes (*Brand Awareness*).

Qualitative Features

In complement to the above-mentioned quantitative features, we also collected textual data representing the media exposure of the various brands in traditional news media and social media for the *Brand Awareness* and *Brand Loyalty* dimensions. Starting with conventional news media, we extracted, using the publicly-accessible API[5], all football-related articles from The Guardian[6] – i.e., a prominent British daily newspaper. In total, we acquired, for the previously mentioned five-year span, just short of 30,000 documents. Employing a simple disambiguation approach, we gathered, for every brand, a list of known aliases from Wikidata[7] – e.g., *Manchester United FC*, *Red Devils*, and *Man Utd* – and then matched these aliases with the obtained documents (*Brand Awareness*).

[4]https://www.transfermarkt.com
[5]https://open-platform.theguardian.com
[6]https://www.theguardian.com
[7]https://www.wikidata.org

		Feature	Measurement	Min.	Median	Mean	Max.	Source	Extracted / ML-Generated
Brand Associations	Club tradition	*Club age*	Years	101.0	133.0	129.2	153.0	Transfermarkt	
		Stadium age	Years	1.0	100.0	78.02	164.0	Transfermarkt	
	Departures (players)	*Player departures*	Total departures p.s.	4.0	17.0	17.5	34.0	Transfermarkt	
		Departures market value	Total market value of departures p.s. / 1,000,000	14.8	90.74	102.50	406.5	Transfermarkt	
	Dismissals (coaches)	*Coach dismissals*	Total dismissal p.s.	0.0	0.0	0.5227	4.0	Transfermarkt	
Brand Awareness	English matches	*EFL Cup matches*	Total matches p.s.	1.0	2.0	2.784	7.0	Transfermarkt	
		FA Cup matches	Total matches p.s.	1.0	3.0	3.227	7.0	Transfermarkt	
	European matches	*Europa League matches*	Total matches p.s.	0.0	0.0	1.5	15.0	Transfermarkt	
		Champions League matches	Total matches p.s.	0.0	0.0	2.33	15.0	Transfermarkt	
	Media coverage	*Guardian articles*	Total articles p.s. / 100	4.130	14.085	15.156	36.470	The Guardian	✓
	Social media presence	*Tweets from club*	Total tweets p.s. / 100	14.34	55.13	54.9	106.85	Twitter	
		Twitter likes from club	Total likes p.s. / 100	0.0	7.14	9.558	55.21	Twitter	
Brand Loyalty	Fan interactions	*Tweets at club*	Total tweets p.s. / 100,000	0.00132	0.42092	1.27991	6.72250	Twitter	
		Ratio of negative tweets at club	Total neg. tweets p.s. / Total tweets p.s.	0.09049	0.15716	0.16826	0.28930	Twitter	✓
		Ratio of positive tweets at club	Total pos. tweets p.s. / Total tweets p.s.	0.2766	0.3637	0.3638	0.4940	Twitter	✓
		Ratio of sensitive tweets at club	Total sens. tweets p.s. / Total tweets p.s.	0.001689	0.005120	0.005520	0.013043	Twitter	✓
	Stadium attendances	*Spectators*	Total spectators p.s. / 10,000	0.0	54.48	56.21	143.05	Twitter	
	Social media followers	*Twitter followers delta*	Total followers Δ p.s. / 10,000	-0.0072	19.8652	79.5535	670.0495	Twitter	
		Reddit subscribers delta	Total subscribers Δ p.s. / 10,000	0.0124	0.28485	1.27948	8.72010	Reddit	
Perceived Quality	Historical English championships	*Hist. Premier League titles*	Total historical championships	0.0	2.0	4.909	20.0	Transfermarkt	
		Hist. FA Cup titles	Total historical championships	0.0	4.0	4.33	14.0	Transfermarkt	
		Hist. EFL Cup titles	Total historical championships	0.0	1.0	2.125	8.0	Transfermarkt	
	Historical European championships	*Hist. Europa League titles*	Total historical championships	0.0	0.0	0.4318	3.0	Transfermarkt	
		Hist. Champions League titles	Total historical championships	0.0	0.0	0.2727	2.0	Transfermarkt	
	Squad valuation	*Squad valuation*	Total market value of squad p.s. / 1,000,000	106.0	307.7	447.0	1203.5	Transfermarkt	
Other Proprietary	Stadium capacity	*Stadium capacity*	Total seats / 10,000	1.133	3.709	4.052	7.488	Transfermarkt	
	Technical sponsor	*Technical sponsor (Adidas)*	Categorical feature		20 / 88 Observations			Misc.	
		Technical sponsor (Nike)	Categorical feature		16 / 88 Observations			Misc.	
		Technical sponsor (Puma)	Categorical feature		18 / 88 Observations			Misc	
		Technical sponsor (Umbro)	Categorical feature		20 / 88 Observations			Misc.	
		*Technical sponsor (Other)**	Categorical feature		14 / 88 Observations			Misc.	

Note: p.s. = per season; * = Reference category

Table 2. Descriptive Statistics

When it comes to social media, we extracted, using Twitter's Academic Research API[8], all available tweets posted by the clubs on their respective official accounts along with all available fan-generated tweets addressed to the clubs – e.g., *@ChelseaFC*. In total, we acquired, for the above period, 494,327 and 11,326,367 tweets, respectively (*Brand Loyalty*).

Assuming that emotions experienced and exhibited by fans could influence brand value, we also extracted, with the help of a pre-trained state-of-the-art attention-based model[9], the sentiment polarity of every fan-generated tweet (*Brand Loyalty*). This Transformer model, which is based on the BERT architecture and uses the RoBERTa pre-training approach, was developed for Twitter-specific classification tasks and was trained on over 58 million tweets (Barbieri et al. 2020; Devlin et al. 2018; Liu et al. 2019).

Categorical Features

Lastly, we gathered, from various sources, the technical sponsor of every club. Since these can, however, change over time, every season needed to be acquired separately (*Other Proprietary*).

Aggregation

As can be expected, all quantitative and qualitative features were, prior to modeling, aggregated based on club and season. An overview of the descriptive statistics can be found in Table 2.

Results

Explanatory Models

Table 3 shows the estimated regression coefficients and goodness-of-fit metrics of our explanatory models. These five models share the same response variable – i.e., the natural logarithm of brand value – but have different feature sets, one for each brand equity dimension. However, due to multicollinearity issues, we did not estimate an overall explanatory model combining all brand equity dimensions. Consequently, we present, as exposed in the previous section, a global predictive model leveraging all features.

First, as indicated by the marginal R^2 values, the *Perceived Quality* model explains, with its high R^2 value of 77%, most of the variation in brand value present in our dataset. Hence, this suggests that squad valuation and historical titles are, in fact, key predictors of brand value. Note that we deliberately focused on features representing past achievements and excluded features measuring current on-pitch performance, as this would violate the definition of perceived quality as a judgement of quality that goes beyond functional and objective measurement. Additionally, interpreting and comparing the estimated coefficients provides interesting insights into the logic of brand equity. For example, a Champions League title, arguably the most prestigious title in European football, is associated with an 8% ($e^{0.079}$)[10] change in brand value, while the national FA Cup title is only "worth" a 6% brand value premium. Further, focusing on the on-pitch potential of a team, a one million dollar increase in squad valuation is associated with a 0.1% rise in brand value.

Second, the *Brand Loyalty* model also has, with an R^2 value of roughly 48%, high explanatory power. This dimension comprises features measuring direct interactions between fans and clubs, both in the real world – e.g., visiting a game – and on social media – e.g., sending a tweet to the club or subscribing to its Reddit channel. While, according to our analysis, none of the NLP-generated features – i.e., the share of positive, negative, or sensitive tweets – seem to be predictive of brand value, the regression coefficients reveal that more genuine predictors – i.e., the number of tweets addressed to a club, the number of new subscriptions to a club's social media channels, and the attendance numbers – are all statistically significant. For instance, an increase or decrease by 100,000 tweets in the total number of tweets addressed to a club during a season is associated with a 17% increase or decrease in brand value.

Next, with a marginal R^2 value of 17%, the *Other Proprietary* model contains features that are unique to football and that are, as a result, difficult to generalize to other domains. The coefficients suggest, for example, that a sponsoring contract with Adidas is associated with a 5% increase in brand value compared to

[8] https://developer.twitter.com/en/products/twitter-api/academic-research
[9] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment
[10] Since our response variable is measured on the logarithmic scale, the coefficients can be interpreted roughly as percent changes.

clubs under contract with less popular sponsors – i.e., the reference level "Other". Furthermore, the capacity of a club's home stadium is also associated with an increased brand value – i.e., an 18% increase per 10,000 additional seats.

	Dependent variable:				
	log(brand_value)				
	Brand Associations	Brand Awareness	Brand Loyalty	Perceived Quality	Other Proprietary
Constant	5.778^{***} (1.494)	5.391^{***} (0.206)	5.349^{***} (0.412)	4.984^{***} (0.120)	5.079^{***} (0.235)
Club age	0.001 (0.011)				
Stadium age	−0.002 (0.001)				
Player departures	−0.004 (0.006)				
Departures market value	0.0001 (0.001)				
Coach dismissals	−0.005 (0.027)				
EFL Cup matches		0.028^{**} (0.013)			
FA Cup matches		0.001 (0.013)			
Europa League matches		0.013^{*} (0.007)			
Champions League matches		0.022^{***} (0.008)			
Guardian articles		−0.00003 (0.008)			
Tweets from club		0.003^{**} (0.002)			
Twitter likes from club		0.001 (0.002)			
Spectators			0.002^{***} (0.001)		
Tweets at club			0.156^{***} (0.048)		
Negative tweets at club			−0.241 (1.001)		
Positive tweets at club			0.082 (0.750)		
Sensitive tweets at club			−4.959 (10.935)		
Reddit subscribers delta			0.054^{**} (0.022)		
Twitter followers delta			0.001^{***} (0.0003)		
Squad valuation				0.001^{***} (0.0002)	
Hist. Premier League titles				0.021 (0.024)	
Hist. FA Cup titles				0.061^{**} (0.030)	
Hist. EFL titles				0.080^{*} (0.042)	
Hist. Europa League titles				0.017 (0.102)	
Hist. Champions League titles				0.079 (0.116)	
Technical sponsor (Adidas)					0.045 (0.104)
Technical sponsor (Nike)					−0.021 (0.101)
Technical sponsor (Puma)					0.015 (0.090)
Technical sponsor (Umbro)					−0.148 (0.105)
Stadium capacity					0.168^{***} (0.050)
Observations	88	88	88	88	88
R^2 (marginal)	0.011	0.041	0.478	0.774	0.155
R^2 (conditional)	0.965	0.957	0.941	0.964	0.953
Log Likelihood	-39.721	-38.416	-18.656	-16.803	-19.640
Akaike Inf. Crit.	97.442	98.831	59.313	53.605	57.279
Bayesian Inf. Crit.	119.738	126.082	86.564	78.379	79.575

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3. Empirical Results

Lastly, Table 3 reveals that the models estimating the last two remaining dimensions are, jointly, only capable of explaining about 5% of the variation in brand value present in our dataset. Surprisingly, these dimensions, which comprise features representing the tradition and consistency of a team (*Brand Association*) as well as features measuring the exposure of a club on TV, in newspapers, and on the social network Twitter (*Brand Awareness*), are, for the most part, not statistically significant. A likely explanation for these findings is our relatively small dataset and the inclusion of random intercepts for clubs. It is, however, important to note that most coefficient signs point in plausible directions.

Predictive Model

Investigating the utility of our data-driven approach for practical applications, we also evaluated the predictive accuracy of a tree-based gradient boosting model fitted using all available features except club and season. After hyperparameter optimization, the model achieved a mean absolute error (MAE) of 48 million US dollars and a mean absolute percentage error (MAPE) of 14% on out-of-sample observations (time-based split). Surprisingly, the feature importance estimates for this model revealed that four out of the five most important features, namely club age, stadium age, player departures, and departures market value, are predictors associated with the *Brand Association* dimension – i.e., the dimension with the lowest marginal R^2 value. This finding suggests that these features are, in fact, informative but were not statistically significant in the explanatory models solely because of the inclusion of club and season intercepts.

Discussion and Outlook

Organizations invest billions in building brands. However, current brand valuation practices do not sufficiently support decision makers in understanding brand equity dimensions. Hence, making evidence-based investment decisions to increase brand value is often impossible. Therefore, we evaluated the eligibility of a data-driven approach for estimating brand value with the case of the English Premier League by operationalizing Aaker's (1991) brand equity dimensions with quantitative and qualitative data from publicly available data sources and features generated using text mining and machine learning techniques. Our statistical models demonstrate that such data can explain and predict brand value and provide insight regarding their impact on brand equity, thereby offering three significant implications for theory and practice.

First, our approach is the first to align with all general requirements of financial brand valuation (ISO 2010). Thus, it provides managers with a transparent, valid, reliable, sufficient, and objective means for brand valuation. Compared to other approaches, all steps conducted to estimate brand value can be adapted to other domains and reproduced by others, thereby minimizing human bias. Second, we show how theoretical concepts can be mobilized to identify features that enable the data-driven analysis of brand value. We deduced brand equity dimensions from Aaker's (1991) framework and identified relevant features that can be used to estimate brand value. For the case of English Premier League clubs, we show that feature selection based on contextualizing theoretical frameworks by collecting publicly available data can result in valid and relevant qualitative and quantitative predictors. In this work, we based our feature selection and data collection process on works focusing on brand equity in sports (e.g., (Biscaia et al. 2016; Wang and Tang 2018; Wetzel et al. 2018). However, our approach is the first to simultaneously provide a holistic view of all brand equity dimensions, additionally enabling a drill down into single features of each dimension. Third, we provide a reliable basis for evidence-based decision-making related to marketing investments aiming at increasing brand value. Our approach can be employed to analyze the importance of the dimensions of brand equity as well as the importance of all features constituting these dimensions in a specific domain.

Even though we carefully evaluated our approach, our results are naturally subject to limitations; therefore, providing the basis for future research. Concerning generalizability, we solely assessed the eligibility of employing Aaker's (1991) dimensions of brand equity and identifying features for data-driven analysis for the case of professional football. Future research must determine if our approach can be adapted and applied to other domains. This review includes the explicit identification of more generic features in order to represent the dimensions of brand equity. However, we are confident that our results can provide other researchers with the means to develop methods and tools that managers can use for ad hoc brand valuation. Hence, we hope future research can enable better-informed decision-making regarding brand valuation.

References

Aaker, D. A. 1991. *Managing Brand Equity,* Free Press.

American Marketing Association. 2022. *Branding*. https://www.ama.org/topics/branding/.

Barbieri, F., Camacho-Collados, J., Neves, L., and Espinosa-Anke, L. 2020. "Tweeteval: Unified Benchmark and Comparative Evaluation for Tweet Classification," *arXiv preprint arXiv:2010.12421*.

Bergstra, J., Yamins, D., and Cox, D. 2013. "Making a Science of Model Search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *International conference on machine learning,* PMLR, pp. 115–123.

Bettencourt, L. A. 2017. *Are You Managing Brand Equity Incorrectly?* https://www.ama.org/2017/12/20/are-you-managing-brand-equity-incorrectly/.

Biscaia, R., Ross, S., Yoshida, M., Correia, A., Rosado, A., and Marôco, J. 2016. "Investigating the Role of Fan Club Membership on Perceptions of Team Brand Equity in Football," *Sport Management Review* 19.2, pp. 157–170.

Chen, T. and Guestrin, C. 2016. "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,* pp. 785–794.

Cole, M. S. 2012. "Is Your Band Going Out of Fashion? A Quantitative, Causal Study Designed to Harness the Web for Early Indicators of Brand Value." PhD thesis. Georgia State University.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2018. "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*.

Doyle, P. 1989. "Building Successful Brands: The Strategic Options," en. *Journal of Marketing Management* 5.1 1989, pp. 77–95.

Huang, C.-Y., Liu, P.-Y., and Xie, S.-M. 2020. "Predicting Brand Equity by Text-Analyzing Annual Reports," *International Journal of Market Research* 62.3, pp. 300–313.

ISO. 2010. *ISO 10668: Brand Valuation — Requirements for Monetary Brand Valuation*.

ISO. 2019. *ISO 20671: Brand Evaluation — Principles and Fundamentals*.

Keller, K. L. 1993. "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity," *Journal of Marketing* 57.1, pp. 1–22.

Keller, K. L. and Brexendorf, T. O. 2019. "Strategic Brand Management Process," de. in *Handbuch Markenführung,* F.-R. Esch (ed.). Wiesbaden: Springer Fachmedien Wiesbaden, pp. 155–175.

Lindemann, J. 2010. *The Economy of Brands,* Springer.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. 2019. "Roberta: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692*.

Markables. 2015. *Brand Value Rankings –- Are They Blessing or Curse?* 2015.

Nuortimo, K. and Harkonen, J. 2019. "Improving Strategic Decision Making with Big Data-Based Media Analysis – The Case of Coal Power," *International Journal of Global Environmental Issues* 18.3, pp. 187–208.

Raggio, R. D. and Leone, R. P. 2007. "The Theoretical Separation of Brand Equity and Brand Value: Managerial Implications for Strategic Planning," *Journal of Brand Management* 14.5 2007, pp. 380–395.

Raggio, R. D. and Leone, R. P. 2009. "Chasing Brand Value: Fully Leveraging Brand Equity to Maximise Brand Value," *Journal of Brand Management* 16.4 2009, pp. 248–263.

Rojas-Lamorena, Á. J., Del Barrio-García, S., and Alcántara-Pilar, J. M. 2022. "A Review of Three Decades of Academic Research on Brand Equity: A Bibliometric Approach using Co-Word Analysis and Bibliographic Coupling," *Journal of Business Research* 139, pp. 1067–1083.

Salinas, G. and Ambler, T. 2009. "A Taxonomy of Brand Valuation Practice: Methodologies and Purposes," en. *Journal of Brand Management* 17.1 2009, pp. 39–61.

Shmueli, G. 2010. "To Explain or to Predict?," *Statistical science* 25.3, pp. 289–310.

Shmueli, G. and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research," *MIS quarterly*, pp. 553–572.

Wang, M. C.-H. and Tang, Y.-Y. 2018. "Examining the Antecedents of Sport Team Brand Equity: A Dual-Identification Perspective," *Sport Management Review* 21.3, pp. 293–306.

Wetzel, H. A., Hattula, S., Hammerschmidt, M., and Heerde, H. J. van. 2018. "Building and Leveraging Sports Brands: Evidence From 50 Years of German Professional Soccer," *Journal of the Academy of Marketing Science* 46.4, pp. 591–611.