Universidade do Minho
Escola de Economia e Gestão

André Serra Campos Ferreira

Valuing Stadium Naming Rights in European Football: A Multi-Model Empirical Framework Applied to Sporting Clube de Braga

April 2026

Valuing Stadium Naming Rights in European Football: A Multi-Model Empirical Framework Applied to Sporting Clube de Braga
André Serra Campos Ferreira
UMinho | 2026

Universidade do Minho
Escola de Economia e Gestão

André Serra Campos Ferreira

Valuing Stadium Naming Rights in European Football: A Multi-Model Empirical Framework Applied to Sporting Clube de Braga

Master's Dissertation
Master's Programme in Management
General Management

Work conducted under the supervision of
Professor Doutor José António Almeida Crispim

April 2026

COPYRIGHT AND TERMS OF USE OF THE WORK BY THIRD PARTIES

This is an academic work that may be used by third parties provided that the rules and internationally accepted good practices regarding copyright and related rights are respected.

Thus, this work may be used under the terms set out in the license indicated below.

If the user requires permission to make use of the work under conditions not covered by the indicated license, they must contact the author through the RepositóriUM of the University of Minho.

License granted to users of this work:



Attribution 4.0 International
CC BY 4.0

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, José Crispim, for his guidance, expertise, and continuous support throughout the development of this thesis.

I am also deeply grateful to SC Braga for welcoming me and providing the crucial information and context required for the case study. A special word of thanks goes to José Silva, my primary point of contact and internship supervisor at the club, for his time, availability, and invaluable assistance during this process.

I also wish to extend my gratitude to Augustuna. Being part of this group provided an invaluable source of joy and stress relief, helping me maintain a healthy mental balance during the most demanding phases of this academic journey.

Finally, and most importantly, I would like to thank my family and friends for their constant support and encouragement.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE (AI) TOOLS

I hereby declare having acted this academic work with integrity and confirm that I have not resorted to any form of undue use of AI tools, nor to the falsification or omission of information, sources, data, results or processes in any of the stages leading to its elaboration.

I further declare that any use of AI tools was done in a transparent and responsible manner, being duly noted in the work, and that I critically verified the accuracy of the content produced with their support, assuming full responsibility for what I submit.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho, as well as the applicable standards of academic integrity and data protection.

Resumo

A comercialização dos naming rights de estádios afirmou-se como uma das principais fontes de receita de longo prazo para os clubes de futebol europeus, contudo a literatura empírica sobre a sua valorização permanece predominantemente concentrada no mercado norte-americano. Este estudo colmata essa lacuna através do desenvolvimento do primeiro modelo de avaliação multimodelo especificamente calibrado para o contexto do futebol europeu, aplicado ao Estádio Municipal de Braga, estádio do Sporting Clube de Braga.

Para o efeito, foi construída de raiz uma base de dados com 306 acordos de naming rights, abrangendo múltiplos desportos, continentes e ambientes comerciais, segmentada em oito subamostras analíticas que progridem da amostra global até uma configuração geograficamente focada no futebol europeu. Cinco abordagens de estimação (Mínimos Quadrados Ordinários (MQO), Ridge, LASSO, Regressão Quantílica e Random Forest) são aplicadas em paralelo a cada subamostra, tendo como variável dependente o valor anual ajustado à inflação (USD de 2025).

A análise comparativa dos resultados revela um núcleo estrutural de determinantes com consenso quase universal entre os modelos: os *Seguidores do Clube no Instagram*, o *Custo de Construção Ajustado à Inflação* e a *População da Cidade*. Os *Seguidores do Clube no Instagram* constituem a única variável a alcançar consenso pleno em todos os segmentos analíticos, emergindo como o principal fator explicativo nos contextos específicos de futebol. Este resultado confirma empiricamente que o alcance da audiência digital suplantou o ativo físico enquanto principal diferenciador comercial, uma conclusão antecipada pela literatura, mas até agora por ela não testada. A *Assistência Média* ganha peso à medida que a amostra se restringe ao futebol, ao passo que variáveis contratuais como a *Duração do Contrato* e o *Número de Acordos Anteriores* perdem capacidade discriminatória.

Aplicado ao Sporting Clube de Braga por via de três percursos de estimação independentes, o modelo gera um intervalo de valor justo de mercado triangulado entre 1,70M€ e 1,78M€ anuais em naming rights, com um limite superior de 3,23M€. A convergência de percursos metodologicamente distintos num intervalo de 5% constitui evidência robusta de solidez estrutural, demonstrando que a triangulação multimodelo é uma estratégia de valorização viável em mercados caracterizados por reduzida transparência e amostras de dimensão limitada.

PALAVRAS-CHAVE

Naming Rights; Avaliação de *Naming Rights*; Avaliação Multimodelo; Economia do Desporto; Futebol Europeu; Sporting Clube de Braga

Abstract

Stadium naming rights have established themselves as one of the primary long-term revenue sources for European football clubs, yet the empirical literature on their valuation remains concentrated in the North American market. This study addresses this gap by developing the first multi-model valuation framework specifically calibrated to the European football context and applying it to the Estádio Municipal de Braga, home of Sporting Clube de Braga.

A purpose-built database of 306 naming rights agreements spanning multiple sports, continents and commercial environments is segmented into eight analytical subsets, progressing from the full global sample to a geographically focused European football configuration. Five estimation approaches (Ordinary Least Squares, Ridge, LASSO, Quantile Regression and Random Forest) are applied in parallel to each subset, with the inflation-adjusted annual value (2025 USD) as the dependent variable.

A meta-analytical synthesis identifies a structural core of determinants: *Club's Instagram Followers*, *Inflation Adjusted Construction Cost* and *City Population*. *Club's Instagram Followers* achieves full consensus across every analytical segment and emerges as the dominant driver in football-specific contexts, empirically confirming that digital audience reach has displaced the physical asset as the primary commercial differentiator, a conclusion the prior literature had anticipated but not tested. *Average Attendance* gains relevance as the sample narrows to football, while contractual variables such as *Contract Length* and *Past Naming Deals Count* lose discriminatory power.

Applied to Sporting Clube de Braga through three independent estimation pathways, the framework produces a triangulated fair market value range of €1.70M to €1.78M in annual naming rights, with an upper bound of €3.23M. The convergence of methodologically distinct pathways within 5% provides strong evidence of structural robustness and demonstrates that multi-model triangulation is a viable valuation strategy in markets with limited transparency and small sample sizes.iable valuation strategy in markets with limited transparency and small sample sizes.

KEYWORDS

Naming Rights; Naming Rights Valuation; Multi-Model Framework; Sports Economics; European Football; Sporting Clube de Braga

Contents

1 Introduction **1**

2 Literature Review **3**

2.1 The Evolution of Sports Industry 3

2.2 Stadium Naming Rights 4

2.3 Methods for Naming Rights Valuation 5

2.4 Determinants of Naming Rights Value 7

3 Methodology **10**

3.1 Data Collection and Sample Characterization 10

3.1.1 Data Sources and Selection Criteria 10

3.1.2 Financial Normalization and Inflation Adjustments 11

3.2 Variable Definition and Operationalization 11

3.3 Database Segmentation and Refinement 13

3.3.1 Data Partitioning and Subset Definition 13

3.3.2 Deduplication Protocol 14

3.4 Limitations of the Database 14

3.5 Modeling Approach and Rationale 15

3.5.1 Hedonic Price Model (HPM) 17

3.5.2 Ridge and LASSO Regression 18

3.5.3 Quantile Regression 20

3.5.4 Random Forest 20

3.5.5 Model Evaluation and Synthesis Strategy 21

4 Results **23**

4.1 Hedonic Pricing Model (OLS) . . . 23

4.1.1 Hedonic – Full Database Analysis . . . 23

4.1.2 Hedonic – Stability and Market Dimension Analysis . . . 24

4.1.3 Hedonic – Football-Specific Market Analysis . . . 26

4.2 Ridge Regression . . . 28

4.2.1 Ridge – Full Database Analysis . . . 28

4.2.2 Ridge – Stability and Market Dimension Analysis . . . 29

4.2.3 Ridge – Football-Specific Market Analysis . . . 31

4.3 LASSO Regression . . . 33

4.3.1 LASSO – Full Database Analysis . . . 33

4.3.2 LASSO – Stability and Market Dimension Analysis . . . 34

4.3.3 LASSO – Football-Specific Market Analysis . . . 36

4.4 Quantile Regression . . . 38

4.4.1 Quantile – Full Database Analysis . . . 38

4.4.2 Quantile – Stability and Market Dimension Analysis . . . 39

4.4.3 Quantile – Football-Specific Market Analysis . . . 41

4.5 Random Forest . . . 43

4.5.1 Random Forest – Full Database Analysis . . . 43

4.5.2 Random Forest – Stability and Market Dimension Analysis . . . 45

4.5.3 Random Forest – Football-Specific Market Analysis . . . 46

4.6 Model Performance and Quality Evaluation . . . 48

5 Discussion **51**

5.1 Main Determinants of Naming Rights Value 51
5.1.1 Core Value Determinants 53
5.1.2 Impact of Sample Refinement 54
5.2 Model Performance and Selection for the Case Study 56
5.2.1 Model Evaluation and Cross-Subset Stability 56
5.2.2 Model Selection for the SC Braga Valuation 58

6 The SC Braga Case Study **61**
6.1 SC Braga's Profile and Commercial Context 61
6.2 Valuation Results 65
6.3 Interpretation and Fair Market Value Estimate 69

7 Conclusion **72**
7.1 Research Questions and Main Findings 72
7.2 Practical Implications 72
7.3 Limitations and Directions for Future Research 73

Appendix A SHAP Summary Plots: Stability and Market Dimension Analysis **78**

Appendix B SHAP Summary Plots: Football-Specific Market Analysis **80**

List of Figures

4.1 SHAP Summary Plot — Random Forest (Full Database, $n = 306$) 44

4.2 SHAP Summary Plot — Random Forest (Major Leagues Without Dups, $n = 143$) 45

4.3 SHAP Summary Plot — Random Forest (European Football, $n = 57$) 47

5.1 Model Consensus Rate — Primary Drivers of Naming Rights Value (Full Database) 51

5.2 Model Consensus Rate — Primary Drivers of Naming Rights Value (Stability & Dimension) 52

5.3 Model Consensus Rate — Primary Drivers of Naming Rights Value (Football-Specific) . . 52

6.1 SHAP Waterfall Plot — SC Braga, Random Forest (Excluding Events, $n = 285$) 66

6.2 SHAP Waterfall Plot — SC Braga, Random Forest (European Football, $n = 57$) 67

A.1 SHAP Summary Plot — Random Forest (Excluding Events, $n = 285$) 78

A.2 SHAP Summary Plot — Random Forest (Unique Agreements, $n = 219$) 79

A.3 SHAP Summary Plot — Random Forest (Major Leagues, $n = 174$) 79

B.1 SHAP Summary Plot — Random Forest (Football (Soccer) Only, $n = 95$) 80

B.2 SHAP Summary Plot — Random Forest (Soccer outside the USA, $n = 76$) 80

List of Tables

3.1 Role of Each Method in the Multi-Model Framework 17

4.1 Most Impactful Variables in Hedonic Pricing Model (Full Database) 24

4.2 Most Impactful Variables in Hedonic Pricing Model (Stability & Dimension) 25

4.3 Most Impactful Variables in Hedonic Pricing Model (Football-Specific Markets) 27

4.4 Most Impactful Variables in Ridge Regression (Full Database) 29

4.5 Most Impactful Variables in Ridge Regression (Stability & Dimension) 30

4.6 Most Impactful Variables in Ridge Regression (Football-Specific Markets) 32

4.7 Most Impactful Variables in LASSO Regression (Full Database) 34

4.8 Most Impactful Variables in LASSO Regression (Stability & Dimension) 35

4.9 Most Impactful Variables in LASSO Regression (Football-Specific Markets) 37

4.10 Most Impactful Variables in Quantile Regression (Full Database) 39

4.11 Most Impactful Variables in Quantile Regression (Stability & Dimension) 40

4.12 Most Impactful Variables in Quantile Regression (Football-Specific Markets) 42

4.13 Comparative Performance Metrics Across Models and Subsets 49

6.1 SC Braga – Input Variables for the Valuation Models 64

6.2 Summary of SC Braga Valuation Outputs Across the Three Estimation Pathways 69

List of Abbreviations and Acronyms

CBI	Crypto Blockchain Industries
CVM	Contingent Valuation Method
DCF	Discounted Cash Flow
FMV	Fair Market Value
HPM	Hedonic Pricing Model
MAE	Mean Absolute Error
MLB	Major League Baseball
MLS	Major League Soccer
NBA	National Basketball Association
NFL	National Football League
NHL	National Hockey League
NR	Naming Rights
OLS	Ordinary Least Squares
OTT	Over-the-Top
QR	Quantile Regression
RF	Random Forest
RMSE	Root Mean Square Error
SBJ	Sports Business Journal
UEFA	Union of European Football Associations
USD	United States Dollars

1 Introduction

The professional sports industry has undergone a profound transformation, evolving from a model reliant on match-day receipts to a complex commercial ecosystem centered on the monetization of intangible assets (Bauer et al., 2005). Within this landscape, corporate naming rights have emerged as one of the most prominent and high-value sponsorship properties, providing brands with a unique platform for long-term visibility and association with sporting success (Cornwell et al., 2005). These agreements represent a sophisticated marketing communication strategy for sponsors while providing sport organizations with a consistent and long-term source of revenue (Chen & Zhang, 2012).

Despite the significant financial stakes, the academic literature on naming rights valuation remains heavily concentrated in the North American context. Studies such as DeSchriver and Jensen (2003), Gerrard et al. (2007), and Centracchio et al. (2023) have produced robust hedonic pricing frameworks for the United States market, where transparency is relatively common and transaction data is more readily available. The European football market, by contrast, is characterized by substantial information asymmetry and opacity (Gerrard et al., 2007): agreements are frequently shielded by confidentiality clauses, active deals remain a minority among top-tier clubs, and no empirical study to date has developed a multi-model valuation framework specifically calibrated to the European football context. This gap is commercially significant. For growing European football organizations, the commercialization of stadium naming rights is no longer merely an opportunistic revenue stream but a critical strategic lever in an environment governed by financial sustainability regulations and increasing global competition.

This study addresses two explicit research questions. First, what are the primary determinants of naming rights value in professional sport, and to what extent do these determinants vary across sports, geographies, and competitive contexts? Second, what is the fair market value of annual naming rights for the Estádio Municipal de Braga, home of Sporting Clube de Braga? The dependent variable throughout is the inflation-adjusted annual naming rights value, expressed in 2025 USD to ensure comparability across markets, currencies, and contract vintages.

To address these questions, the study constructs a purpose-built database of 306 naming rights agreements and valuations spanning multiple sports, continents, and commercial environments. Taking as its conceptual point of departure the hedonic pricing tradition established by Rosen (1974) and applied to naming rights by DeSchriver and Jensen (2003), Gerrard et al. (2007) and Centracchio et al. (2023), which proposes that a contract's price can be decomposed into the implicit values of its constituent characteristics, the study applies a multi-model empirical framework comprising Ordinary Least Squares (OLS) regression, Ridge regression, LASSO regression, Quantile Regression (QR) and Random Forest (RF). Each model is estimated across eight analytically distinct subsets of the database, ranging from the full global sample to a geographically focused European football configuration. This multi-model, multi-subset design serves two purposes: it provides a robust meta-analytical foundation for identifying the determinants of naming rights

value, and it generates a triangulated valuation framework for the SC Braga case study that is anchored in multiple methodological perspectives rather than a single model's output.

The contribution of this study to the existing literature is threefold. First, it provides the most comprehensive empirical analysis of naming rights determinants to date, extending the hedonic pricing tradition to include digital engagement metrics, specifically Club's Instagram Followers, whose commercial relevance has been theorized but not previously tested in a multi-sport, multi-geography framework. Second, it develops the first multi-model valuation framework specifically calibrated to the European football market, addressing a gap that the prior literature has consistently acknowledged but not resolved. Third, it applies this framework to a concrete case study, Sporting Clube de Braga and the Estádio Municipal de Braga, providing a data-driven fair market value estimate (FMV) that is directly actionable for commercial negotiations and that demonstrates the practical applicability of econometric valuation methods in a context where market opacity has historically prevented rigorous benchmarking.

The remainder of the thesis is structured as follows. Section 2 reviews the existing literature on naming rights valuation, the determinants of sponsorship value, and the methodological frameworks applied in prior studies. Section 3 describes the data collection process, variable operationalization, database segmentation, and modeling approach. Section 4 presents the empirical results across all five models and eight subsets. Section 5 analyses and discusses those results, identifying the primary determinants of naming rights value and selecting the optimal modeling pathways for the SC Braga valuation. Section 6 applies the selected pathways to the SC Braga profile and derives a fair market value estimate. Section 7 concludes with a synthesis of the main findings, practical implications, and directions for future research.

2 Literature Review

This chapter establishes the theoretical and empirical foundation on which the present study builds. It begins by tracing the evolution of the professional sports industry into a commercial ecosystem in which intangible assets, sponsorship arrangements and naming rights have become central revenue drivers. It then examines the specific nature of stadium naming rights as a sponsorship property, the methodological frameworks that prior literature has applied to their valuation, and the determinants that have been identified as systematic drivers of value. Together, these four sections situate the present study within the existing body of knowledge and motivate the multi-model framework developed in the chapters that follow.

2.1 The Evolution of Sports Industry

The professional sports industry has experienced a significant structural transition over recent decades, evolving from a localized, ticket-driven enterprise into a sophisticated global commercial ecosystem (Bauer et al., 2005; Santomier, 2024). Historically, the financial viability of sports organizations was almost exclusively dependent on tangible revenue streams, such as gate receipts and stadium concessions. However, as the industry reached a state of commercial maturity, the focus shifted toward the monetization of intangible assets like intellectual property, broadcast rights, and brand associations, which now largely determine the financial hierarchy of global sport (Bauer et al., 2005). This shift has fundamentally reconfigured the business model of sports clubs, moving the emphasis away from physical attendance toward a broader, media-mediated audience where the "invisible" value of the brand often exceeds physical assets (Gratton & Solberg, 2007; Rowbottom, 2002).

Central to this evolution has been the exponential growth of the sports sponsorship market. As traditional advertising channels became increasingly saturated, corporations sought more integrated and emotionally resonant methods to engage consumers. Sponsorship emerged as a primary vehicle for this engagement, offering brands a "less cluttered" environment to establish visibility (Zoghby, 1999, as cited in Centracchio et al., 2023); (Santomier, 2024). This growth is not merely quantitative but qualitative; sponsorship has transitioned from philanthropic "donations" or simple logo placement into strategic partnerships designed to drive market share and brand equity (Crompton & Howard, 2003; DeSchriver & Jensen, 2003). For modern brands, the sports environment provides a high-impact platform to reach specific target demographics, leveraging the deep psychological identification fans maintain with their teams (Becker-Olsen, 2003; Chen & Zhang, 2012).

Within this expanded sponsorship landscape, corporate naming rights have emerged as the most strategically significant intangible asset a sports organization possesses. For clubs, the relevance of these deals is twofold.

First, they provide a consistent, long-term source of guaranteed income that remains relatively insulated from the volatility of on-field performance (Chen & Zhang, 2012). These revenues have become indispensable for capital-intensive projects, specifically the construction and modernization of stadiums and high-performance training facilities (Chen & Zhang, 2012; Crompton & Howard, 2003).

Second, naming rights serves as the cornerstone of a club's commercial portfolio, often acting as the primary driver for auxiliary sponsorship tiers. Because these agreements offer "around the clock" brand exposure and are deeply embedded in both media coverage and fan discourse, they represent a unique marketing communication strategy that surpasses the reach of standard signage or jersey branding (McCarthy & Irwin, 2000, as cited in Centracchio et al., 2023); (Crompton & Howard, 2003). As clubs continue to seek ways to maintain financial competitiveness, particularly in the European context where financial regulations and rising operational costs are prevalent, the strategic importance of successfully leveraging the venue's name has become a top priority for sport administrators (Kroll, 2024).

2.2 Stadium Naming Rights

A stadium naming rights deal is defined as a "transaction in which money or consideration changes hands in order to secure the right to name a sports facility" (Thornburg, 2003, as cited in Centracchio et al., 2023). Within the modern professional sports industry, which has transformed into a commercial ecosystem centered on the monetization of intangible assets (Bauer et al., 2005), these agreements represent one of the most significant revenue streams available to organizations. The prevailing literature suggests that corporations are primarily inclined to purchase naming rights for athletic venues to secure brand exposure and increase market share (Crompton & Howard, 2003; DeSchriver & Jensen, 2003). These agreements offer companies "around the clock" brand exposure (Crompton & Howard, 2003) and are considered a less cluttered, high-impact way to place a brand in front of the public compared to traditional media (Zoghby, 1999, as cited in Centracchio et al., 2023). Indeed, Centracchio et al. (McCarthy and Irwin, 2000, as cited in 2023) argue that sport facility corporate naming rights may represent the most cost-effective marketing communication strategy available to brands today. Beyond simple visibility, the noted benefits include increasing brand awareness, developing a positive brand image, and positioning the brand within specific target audiences (Becker-Olsen, 2003; Chen & Zhang, 2012; Clark et al., 2002).

Historically, through the early-to-mid 20th century, sports facilities were often named after the team itself, civic leaders, or notable personalities, reflecting a localized and civic-minded approach to sports infrastructure (Thornburg, 2003, as cited in Centracchio et al., 2023). The paradigm shift toward corporate involvement began in the United States during the 1970s. The New England Patriots' stadium is widely recognized as the first major sporting venue to engage in a naming rights agreement with an outside corporation, receiving $150,000 to be named "Schaefer Field" (Crompton & Howard, 2003).

This initial agreement paved the way for a rapid commercialization of the American sporting landscape.

By 1997, roughly one-third of venues in the major North American leagues (NFL, NBA, MLB, and NHL) were corporately named. This trend accelerated exponentially at the turn of the millennium, with the figure increasing to approximately 70% by 2002 (Crompton & Howard, 2003). For sport organizations, these deals provided a consistent and long-term source of income (Chen & Zhang, 2012), eventually becoming critical to the funding process for the construction and renovation of modern stadiums and arenas (Chen & Zhang, 2012; Crompton & Howard, 2003; DeSchriver & Jensen, 2003).

While the North American market has reached a state of historical maturity, the global professional sports industry continues to adapt this model. In the European context, clubs are increasingly looking to stadium naming rights to diversify income and maintain financial competitiveness in an era of rising player wages and infrastructure costs (Kroll, 2024). Recent valuations suggest that for top-tier European teams, these agreements can reach staggering values, as evidenced by landmark deals like FC Barcelona's partnership with Spotify, which can reach up to €30 million per season (Kroll, 2024).

However, the transition into Europe has not been without friction. Administrators must navigate significant cultural challenges that are often absent in the U.S. franchise model. Many European stadiums are viewed by their communities as "memory places" with deep commemorative and historical value (Boyd, 2000). Highly identified fans often perceive corporate renaming as an erosion of the club's heritage and a direct threat to its distinctiveness (Reysen et al., 2012; Woisetschläger et al., 2014). This resistance is deeply rooted in the fear that commercialization will destroy the traditional and idyllic identity associated with the venue's history (Boyd, 2000; Woisetschläger et al., 2014). Consequently, while the financial incentive for naming rights is global, the implementation remains highly dependent on local cultural sentiment.

Against this commercial and cultural backdrop, the academic literature has developed a range of methodological frameworks to address the challenge of naming rights valuation, which are reviewed in the following section.

2.3 Methods for Naming Rights Valuation

The academic literature has employed a range of methodological frameworks to study naming rights agreements, each chosen in accordance with a distinct research objective. Understanding the purpose and limitations of each approach is essential to situating the present study within the existing body of knowledge and explaining the methodological choices made here.

The earliest strand of empirical research applied Event Study Analysis to examine whether naming rights announcements create value for sponsoring firms. Rooted in the efficient market hypothesis, this methodology measures abnormal stock returns within a short window surrounding a public announcement, thereby capturing investor sentiment at the moment of disclosure (Clark et al., 2002; Leeds et al., 2007). Clark et al. (2002) found that naming rights announcements produced average abnormal returns of 1.65% for

sponsoring firms, suggesting positive investor sentiment. Leeds et al. (2007), however, challenged this finding, demonstrating that any positive effect is short-lived and that naming rights do not produce lasting improvements in firm profitability. The event study approach is well suited to measuring whether the market perceives a naming rights agreement as a value-creating investment, but it is structurally limited to the sponsor's perspective: it reveals nothing about how the characteristics of the agreement determine the price. It cannot answer the question of how much a specific naming right is worth, only whether investors believed the price paid was reasonable at the time of announcement.

A second group of approaches, commonly used in industry practice, includes Income-Based and Market-Based valuation methods. The Discounted Cash Flow (DCF) method estimates the present value of projected future benefits to the sponsor, such as increased brand awareness or cost savings relative to traditional advertising, discounting them at an appropriate rate (Kroll, 2024). Market-based comparisons benchmark a prospective deal against similar concluded agreements, adjusting for audience reach and demographic characteristics. While these approaches are consistent with standard corporate finance practice, they depend heavily on subjective assumptions about future benefits and on the availability of comparable transactions. In the European football context, where opacity is pervasive and few deals are publicly disclosed, these methods are particularly unreliable as standalone tools (Chen & Zhang, 2012).

A third methodological strand, more peripheral to naming rights but relevant to the European context, is the Contingent Valuation Method (CVM), a survey-based approach that quantifies the non-market social value of a venue's name, for instance fans' willingness to pay to preserve a traditional stadium name. This framework is not designed to produce a market price for a sponsor but rather to evaluate the social externalities and potential fan resistance that could affect the commercial viability of an agreement (Santomier, 2024).

The most directly relevant framework for the present study is the Hedonic Pricing Model (HPM), first established by Rosen (1974) and applied to naming rights by DeSchriver and Jensen (2003), Gerrard et al. (2007), and more recently Centracchio et al. (2023). This approach treats a naming rights agreement as a heterogeneous bundle of characteristics and regresses the observed transaction price against these attributes to isolate the marginal value of each. Unlike event studies, which evaluate the deal as a single undifferentiated event, and unlike income or market approaches, which require subjective projections, HPM derives implicit prices directly from observed market transactions, producing a transparent and empirically grounded decomposition of value. DeSchriver and Jensen (2003) applied this framework to 112 North American deals, identifying city size, stadium capacity, league status, and facility usage diversity as primary drivers. Gerrard et al. (2007) extended this to a larger North American sample, confirming the centrality of market size and introducing facility age as a negative driver of value. Centracchio et al. (2023) subsequently updated the HPM framework for the collegiate market, finding attendance, all-time winning percentage, and venue construction cost as the dominant explanatory variables in that context.

The present study takes this hedonic tradition as its literature starting point and departs from it in three important ways. First, it incorporates digital audience reach as an explanatory variable, a dimension that

no prior hedonic study of naming rights has tested empirically despite its growing commercial relevance (Watanabe et al., 2016). Second, it applies the analysis to a mixed-sample database that includes European football alongside North American data, directly addressing the geographic gap in the existing literature. Third, and crucially, rather than treating hedonic OLS as the definitive model, it positions OLS as one of five complementary estimators alongside Ridge regression, LASSO regression, Quantile regression and Random Forest. This multi-model design is motivated by the well-documented limitations of OLS in contexts with many correlated predictors and potential non-linear interactions, and by the recognition that a single model's output is insufficient as the basis for a commercially actionable valuation. The hedonic tradition therefore informs the variables considered and the decomposition logic applied, but the methodological anchor of the study is the five-model framework as a whole, from which the final valuation is derived through triangulation.

The variables identified in these studies as systematic drivers of naming rights value are reviewed in the following section, providing the empirical basis for the variable selection adopted in the present analysis.

2.4 Determinants of Naming Rights Value

The valuation of stadium naming rights has been shown in the literature to depend on a broad set of determinants reflecting the heterogeneous nature of the asset. Rather than being driven by a single factor, naming rights prices emerge from the interaction of sponsor characteristics, market demographics, venue attributes, team performance, and media exposure (DeSchriver & Jensen, 2003; Gerrard et al., 2007). This section reviews the main variables identified in previous studies and motivates their inclusion in the present analysis.

Sponsor and contract-related characteristics play a central role in determining the value of naming rights. The sponsor category is relevant because firms from different industries derive heterogeneous benefits from brand association and visibility, with sectors such as financial services, technology, or telecommunications often engaging more aggressively in naming-rights agreements (Clark et al., 2002). Contractual features further influence value: contract length affects both pricing and risk allocation, and both Clark et al. (2002) and Walraven et al. (2014) provide evidence that longer agreements are associated with higher overall value and greater brand awareness accumulation, reflecting the commercial premium sponsors are willing to pay for extended exclusivity. The number of past naming deals associated with a venue reflects whether naming rights are "virgin" or have been previously commercialized, a distinction shown to command significant pricing premiums (DeSchriver & Jensen, 2003; Gerrard et al., 2007). Additionally, whether the agreement coincides with a major physical renovation of the venue significantly influences its value; such structural modernizations function as a commercial reset that signals renewed media exposure opportunities and commands a premium in the sponsorship market (Crompton & Howard, 2003; DeSchriver & Jensen, 2003).

Market and demographic variables capture the scale and purchasing power of the audience reached by the sponsor. Prior studies consistently identify city population as one of the strongest predictors of naming rights value, as larger metropolitan areas provide broader exposure and media reach (DeSchriver & Jensen, 2003; Gerrard et al., 2007; Popp et al., 2016). Complementing population size, local economic conditions (proxied in this study by the inflation-adjusted county GDP per capita) serve as an indicator of consumer purchasing power and market attractiveness, consistent with the household income variable found significant by DeSchriver and Jensen (2003) and Popp et al. (2016).

Venue characteristics are also critical determinants. Venue age has been shown to negatively impact naming rights value, as older facilities are associated with lower technological standards and reduced sponsor appeal (DeSchriver & Jensen, 2003; Gerrard et al., 2007). Venue construction costs proxy for architectural quality and modernization, and have been identified as a significant predictor of naming rights value across both professional and collegiate contexts (Centracchio et al., 2023; DeSchriver & Jensen, 2003). Larger stadium capacity expands the potential in-stadium audience and is associated with higher naming rights premiums (Gerrard et al., 2007), while average attendance and average occupancy rate capture the extent to which that capacity is effectively utilized, with attendance identified as a primary driver of value in both collegiate and professional settings (Centracchio et al., 2023; Popp et al., 2016). Furthermore, whether the venue is multi-use (hosting concerts or non-sporting events) broadens exposure beyond matchdays, and the diversity of sporting usage across multiple tenants has been shown to positively influence naming rights fees (Gerrard et al., 2007; Popp et al., 2016).

The competitive and structural profile of the tenant also shapes naming rights value. The sport and competition level of the primary tenant are significant contextual determinants: Gerrard et al. (2007) demonstrated that the league status of the resident team is one of the primary drivers of naming rights value, reflecting the commercial premium associated with higher-profile competitions and greater media visibility, a finding echoed by Clark et al. (2002) in their cross-sectional analysis of naming rights announcement effects. Sporting success, measured through league performance and all-time winning percentage, further enhances team prestige and increases sponsor visibility (Centracchio et al., 2023; Clark et al., 2002).

Finally, media-related variables capture the magnitude of off-site exposure, which is a core driver of sponsor returns. The number of games broadcast on national television directly quantifies the media reach delivered to a naming rights partner (Clark et al., 2002). Beyond traditional broadcast metrics, contemporary valuations increasingly incorporate measures of digital audience reach: Watanabe et al. (2016) demonstrated that social media engagement functions as a practical proxy for the commercial audience accessible through a sponsorship contract, and the Kroll (2024) report explicitly incorporated digital reach metrics in its valuation of European football naming rights potential. These findings motivate the inclusion of Club's Instagram Followers as a novel variable in the present study, one not previously tested in the academic naming rights valuation literature but grounded in the same logic of audience delivery that underpins all media exposure variables.

The present study incorporates all of the above determinants in its modeling framework, building on the empirical consensus established by the literature while extending it to the European football context and to the digital dimension of contemporary sports sponsorship.

3 Methodology

This chapter presents the methodological framework on which the empirical analysis of this study is built. It describes the construction of the purpose-built database of 306 entries, the operationalization of the dependent variable and predictors, and the segmentation of the data into eight analytical subsets. It then discusses the constraints inherent to the global naming rights landscape and introduces the multi-model estimation framework comprising Ordinary Least Squares, Ridge regression, LASSO regression, Quantile Regression and Random Forest. Together, these sections operationalize the empirical strategy that supports the determinants identified in Section 5 and the valuation exercise developed in Section 6.

3.1 Data Collection and Sample Characterization

The empirical foundation of this research relies on a purpose-built database comprising 306 entries ($n = 306$), documenting individual clubs, teams, or sporting events that utilize venues with active naming rights (NR) agreements or valuation estimates across various sporting and geographical contexts. It is crucial to specify that while the database contains 306 observations, these correspond to 219 unique NR agreements. This distinction arises because several stadiums and arenas are shared by multiple teams or host recurring major events. As these distinct tenant entities exhibit different independent variables, such as attendance rates, sporting performance, and commercial reach, each was assigned a dedicated entry.

The construction of this sample was driven by the urgent need to overcome the current scarcity of quantitative models applicable to European football, with the ultimate goal of developing a precise valuation tool for the specific case of Sporting Clube de Braga. Given that the European NR market is still in a maturation phase and often lacks transparency regarding transaction values, it became imperative to gather a critical mass of data. This allows for the training of machine learning models and statistical regressions that seek to capture the complex relationship between stadium brand value, sporting performance metrics, visibility, and economic context, serving as a basis for extrapolating a Fair Market Value (FMV) applicable to clubs with a profile similar to SC Braga.

3.1.1 Data Sources and Selection Criteria

Regarding the North American market, data was primarily extracted from the Sports Business Journal (SBJ) specialized database, widely recognized for its reliability and depth within the sports industry. To ensure the integrity and relevance of the sample, rigorous filtering criteria were applied to the raw data. Only agreements active at the time of data collection were selected, excluding expired contracts or those scheduled to begin at a future date, thereby ensuring that the values reflect contemporary economic reality. A minimum annual threshold of $500,000 was established to focus the analysis on professional-grade

venues and eliminate the statistical noise generated by small-scale contracts or facilities intended exclusively for training and youth development. Furthermore, entries with ambiguous or incomplete data were removed, and in instances where a club utilized multiple venues, only the primary stadium was considered.

For the global and European data, the collection strategy was designed with a primary focus on European football to align the model as closely as possible with the reality of SC Braga. However, due to the aforementioned market opacity and the limited number of active NR deals on the European continent, it was necessary to expand the research to other geographical markets with similar sports consumption characteristics, such as South America, Asia, and Africa. This expansion allowed for a more diversified sample and the inclusion of variables from emerging markets with a strong footballing tradition. To bridge the information gap in Europe, data from market studies and leading financial reports were integrated, most notably the Kroll (2024) report. This source allowed for the inclusion of hypothetical valuation cases for major European clubs that, despite not having active NR deals at the moment, possess measurable market potential.

3.1.2 Financial Normalization and Inflation Adjustments

To enable a direct and scientifically valid comparison between different markets, currencies, and eras, a comprehensive normalization process was performed. All financial figures were converted into United States Dollars (USD) using historical exchange rates relevant to the date of each contract's inception. Recognizing that the value of money varies significantly over time, all annual agreement values were adjusted for inflation based on the starting year of each specific contract. In the case of hypothetical data derived from the Kroll (2024) report, the publication year was used as the reference for both currency conversion and inflation adjustment.

3.2 Variable Definition and Operationalization

The modeling process is built upon a comprehensive set of variables designed to capture the multi-dimensional nature of stadium sponsorship. The dependent variable is defined as the *Inflation Adjusted Annual Value (2025) (USD)*, representing the normalized annual payment of each agreement, adjusted to current economic terms to ensure comparability across different years and markets. Its construction draws on three supporting fields: the *Annual Value of NR Deal*, recorded in its original *Currency*, with the corresponding *Reference Year* used as the basis for inflation adjustment. Although expressed on a common inflation-adjusted monetary scale, the dependent variable is not fully homogeneous across observations, as it combines officially disclosed transaction values, market-based estimates, and hypothetical valuations. A dedicated *Confirmed Value or Estimate* flag identifies the nature of each entry: it is important to note that only 261 out of the 306 entries (85.3%) had publicly disclosed official values for their naming rights deals,

with the remaining 45 values being estimates provided by media sources and previous valuation studies.

Each observation is identified through the *Facility Name* and the *Tenant Club/Event*. A preliminary classification variable, *Real Deal or Hypothetical Deal*, formally distinguishes between active market transactions and the hypothetical valuation cases sourced from market studies and leading financial reports (most notably the Kroll, 2024 report), as mentioned before. The identity of the sponsoring entity is recorded through the *Naming Rights Holder* field, while the *Category* of that entity is also included as a variable to account for industry-specific investment patterns. The remaining contractual metrics recorded include the *Inflation Adjusted Total Value (2025) (USD)*, the *Contract Length*, and the *Expiry Year*. Additionally, a variable captures the *Past Naming Deals Count*, identifying whether the current agreement is the facility's first naming rights deal or if it follows a history of previous naming partnerships.

Regarding the physical and financial attributes of the assets, the model incorporates the *Venue Age (2025)*, calculated relative to the year 2025, and the *Inflation Adjusted Cost (2025) (USD)*. This financial metric is derived from the raw *Venue Construction Costs*, its original *Currency*, and the *Reference Year* used for inflation adjustment. It is particularly nuanced, as it encompasses both the original investment and any subsequent major renovations that significantly altered the venue's market value; in cases of such structural transformations, the year of the renovation served as the reference point for inflation adjustments. Furthermore, the model distinguishes the venue's operational profile through two separate binary indicators: whether it is a *Multi-sport facility* and whether it functions as a *Multi-use facility (Non-sport Events)*, accounting for regular non-sporting programming. Additionally, another binary variable identifies whether the *Naming Rights deal Coincides with Venue Construction/Renovation*, as the strategic timing of the agreement often dictates the commercial premium.

Sporting prestige and visibility are categorized by the *Sport*, *Competition Level*, and *Gender* of the tenant club or event. Contemporary performance is measured by the *Club's League Performance* (reflecting the last season's position in the 2024 or 2024/25 regular season), while long-term competitive consistency is captured by the *Club's All-Time Winning Percentage in Top Tier League*. Media reach and digital influence are quantified through the *Number of Games in National Television* and the *Club's Instagram Followers (February 2026)*, both of which serve as proxies for the exposure provided to the naming rights partner.

Operational metrics focus on live audience and utilization efficiency. The *Stadium Capacity* establishes the physical scale of the asset, while the *Average Attendance* (from the 2024 or 2024/25 season) tracks the actual volume of spectators. These two variables allow for the derivation of the *Average Occupancy Rate*, providing a percentage-based measure of venue utilization. For instances where clubs did not use their primary stadium during the reference period, these fields remain blank to ensure data integrity and avoid statistical skewing.

Finally, the model incorporates hierarchical geographical and macro-economic data to account for market disparities. This includes the *Continent*, *Country*, *Country Code (ISO Alpha-2)*, *State / Equivalent Region*, *County / Equivalent Region*, and *City*. The economic strength of the surrounding market is quantified by

the *City Population* and the *Inflation Adjusted County's GDP per Capita (2025) (USD)*, the latter derived from the raw *County's GDP per Capita (2024)* and its original *Currency*, adjusted for inflation to a common base year. These variables are essential for calibrating the model to different regional realities, ensuring that the final valuation tool remains sensitive to the economic context of European football and specifically applicable to the case of SC Braga.

3.3 Database Segmentation and Refinement

3.3.1 Data Partitioning and Subset Definition

To ensure the robustness of the statistical models and to account for the inherent disparities between different sporting markets, the primary database ($n = 306$) was segmented into eight distinct subsets. This partitioning strategy allows for a comparative analysis of variable significance across different competitive tiers and geographic regions, progressively narrowing the focus toward the European football context. By testing the models across these variations, the research aims to find an optimal balance between the statistical power provided by a larger sample size and the specificity required to evaluate a club with the profile of Sporting Clube de Braga.

The eight subsets are defined as follows:

1. **Full Database** ($n = 306$): The complete set of 306 entries, providing the maximum statistical breadth for the initial modeling.
2. **Excluding Events** ($n = 285$): A subset containing only resident clubs and teams, removing one-off or recurring emblematic sporting events.
3. **Unique Agreements** ($n = 219$): A version where each of the 219 unique NR agreements is represented by a single entry, eliminating potential biases from multi-tenant venues.
4. **Major Leagues** ($n = 174$): Focuses exclusively on top-tier professional leagues (e.g., NFL, NBA, MLB, football first divisions), representing the highest market value bracket.
5. **Major Leagues without Duplicates** ($n = 143$): A refinement of the previous set to ensure only one representative entity per major league venue.
6. **Football (Soccer) Only** ($n = 95$): A modality-specific subset containing all global football entries, regardless of geography.
7. **Football outside the USA** ($n = 76$): Isolates the football market from the dominant North American influence to better reflect international commercial dynamics.

8. **European Football** ($n = 57$): The most restrictive and relevant subset, consisting only of European clubs to provide the closest possible benchmark for the SC Braga case study.

3.3.2 Deduplication Protocol

A critical methodological challenge arose in the "Unique Agreements" subsets, where a single venue hosted multiple entities, such as stadiums shared by NFL and MLS teams or university arenas hosting multiple sports. To maintain objectivity and ensure that the most commercially representative entity was selected for each agreement, a rigorous hierarchical selection protocol was established. Professional clubs were prioritized over events, and a competitive tier hierarchy was strictly applied, where Major Leagues took precedence over Minor Leagues and college athletics. To maintain consistency with the sample's dominant demographics and commercial data availability, male teams were selected over female teams in multi-tenant scenarios. In cases where two entities shared the same competitive tier, the attendance rate, defined as average attendance relative to the stadium's official capacity, served as the primary tie-breaker to identify the entity with the highest relative demand. Finally, if attendance data was unavailable or inconclusive for a tie-break, the number of Instagram followers was utilized as a final proxy for brand power and commercial reach. This hierarchical structure ensures that in the deduplicated partitions, the model captures the entity with the most significant financial and media impact associated with the venue, thereby minimizing distortions in the final statistical results.

3.4 Limitations of the Database

The construction of the database faced significant challenges, primarily stemming from the structural and historical disparities between regional naming rights (NR) markets. The most prominent limitation is the geographical imbalance of the sample. The United States market, as the pioneer and most mature environment for these agreements, accounts for 221 of the 306 entries in the main database. This dominance is a direct reflection of the market's longevity and the prevalence of NR deals across a wide variety of sports beyond football, which have no direct equivalent in the European landscape.

A second critical limitation concerns data transparency and information opacity. While North American agreements are often subject to public disclosure, the European market remains significantly more opaque, with financial details frequently protected by strict confidentiality clauses. This scarcity of verified European observations directly motivated the expansion of the research to other geographical regions. To ensure a critical mass of data and the statistical relevance of the variables, it was necessary to include entries from South American, Asian, and African markets. While this expansion allows for more robust statistical modeling, it introduces a degree of "noise" due to the different sports consumption habits and economic realities of these regions, representing a necessary trade-off to achieve the volume required for machine

learning applications.

The integration of hypothetical valuation cases, notably from the Kroll (2024) report, also introduces a degree of subjectivity. While these estimates are derived from reputable financial methodologies and provide a necessary benchmark for clubs without active NR agreements, they represent market potential rather than realized commercial transactions. Such estimates may not fully capture the frictions of real-world negotiations or the specific risk-appetite of corporate sponsors at the time of signing.

Regarding digital visibility, the use of *Club's Instagram Followers (February 2026)* serves as a necessary proxy for brand reach, as it is technically impossible to retrieve historical follower counts at the exact date of inception for the vast majority of the sample. In many instances, especially for older agreements, the platform did not yet exist or had not yet matured into a commercial tool. Consequently, while these figures indicate a club's current global prestige and commercial "gravity," they do not necessarily reflect the digital landscape at the moment the contracts were brokered.

Another relevant constraint is the presence of survivorship bias. By focusing on active or widely reported agreements, the sample naturally excludes failed negotiations, bankruptcies of sponsoring partners, or venues that were unable to secure a naming rights deal despite active attempts. This may lead the resulting models to slightly overestimate the "Fair Market Value" by analyzing only those cases where the market successfully reached a financial consensus.

Ultimately, while the database is subject to geographical imbalances, market opacity, and certain temporal anachronisms, these constraints are inherent to the current state of the global naming rights industry. The systematic curation of 306 entries represents a necessary trade-off to overcome the scarcity of transparent European data, providing the critical mass required for sophisticated econometric modeling. By employing rigorous normalization procedures and universal commercial variables, this diverse sample functions as a robust proxy for identifying the underlying value drivers of stadium assets. Rather than undermining the research, these limitations reinforce the necessity of the multi-model approach adopted in the following chapters, ensuring that the final valuation for Sporting Clube de Braga is not only statistically significant but also grounded in a comprehensive, albeit complex, global reality.

3.5 Modeling Approach and Rationale

Taking as its conceptual starting point the Hedonic Pricing Model, which treats a naming rights agreement as a bundle of attributes whose marginal values can be isolated through regression analysis, this study applies a multi-model econometric framework to the 306-entry database and to each of the subsets defined in Section 3.3.1. Utilizing multiple models in parallel offers critical advantages in terms of robustness, functioning as a methodological redundancy system that mitigates the risks of bias associated with the mathematical assumptions of any single isolated model. This strategy allows for the triangulation of results,

where the convergence of different kinds of models upon a common range of values reinforces the validity of the estimated market value.

By integrating models of varying natures, ranging from the rigidity of linear regressions to the flexibility of machine learning algorithms, it becomes possible to capture both clear structural trends and complex, non-linear interactions between variables. This methodological diversity is essential for managing the heterogeneity of the global sample, ensuring that the final valuation is resilient to outliers and statistically defensible, as the limitations of one approach are offset by the analytical capabilities of the others.

Formally, the study adopts as its conceptual point of departure a general log-linear hedonic specification, expressed in Eq. (3.1).

$$\ln(V_i) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i \tag{3.1}$$

Symbol	Description
V_i	Inflation Adjusted Annual Naming Rights Value (2025, USD) for observation i
X_{ik}	Attribute k of observation i, drawn from the full vector of K structural, market, operational, competitive, digital and contractual attributes
β_0	Intercept
β_k	Implicit marginal price associated with attribute k
ε_i	Error term

Four of the five modeling approaches applied in this study estimate this specification under different objective functions and constraints. Ordinary Least Squares estimates β by minimizing the residual sum of squares under the standard Gauss-Markov assumptions. Ridge and LASSO modify this objective by adding L_2 and L_1 penalties respectively to the coefficient vector, addressing multicollinearity and performing variable selection. Quantile Regression replaces the conditional mean with conditional quantiles, relaxing the assumption that marginal effects are constant across the distribution of V_i. Random Forest, by contrast, departs from the parametric hedonic specification entirely: rather than estimating a coefficient vector, it approximates the relationship between V_i and X_i non-parametrically through an ensemble of decision trees, allowing the data to dictate both the functional form and the interaction structure without a pre-defined equation. Its inclusion in the framework therefore serves a specific diagnostic purpose: to test whether relaxing the linearity assumption materially improves predictive performance. Its results are interpreted alongside, rather than within, the hedonic parametric family.

Before proceeding to the detailed description of each method, Table 3.1 summarizes the distinct role played by each approach within this multi-model framework, clarifying how the five estimation strategies complement one another.

Table 3.1: Role of Each Method in the Multi-Model Framework

Method	**Main role in the framework**
OLS (hedonic specification)	Interpretable baseline; isolates the marginal hedonic price of each attribute
Ridge	Handles multicollinearity by shrinking correlated coefficients
LASSO	Performs variable selection through L_1-induced sparsity
Quantile Regression	Captures heterogeneity of marginal effects across the distribution of V_i
Random Forest	Captures non-linearities and interactions; non-parametric predictive benchmark, interpreted post-estimation via SHAP values

Note: OLS, Ridge, LASSO and Quantile Regression estimate the general log-linear hedonic specification presented in Eq. (3.1) under different objective functions and constraints. Random Forest departs from this parametric specification and is therefore interpreted alongside, rather than within, the hedonic family.

The subsequent subsections describe each of these five approaches in detail.

3.5.1 Hedonic Price Model (HPM)

The Hedonic Pricing Model (HPM), theoretically grounded by Rosen (1974), operates on the premise that heterogeneous goods are not valued as singular entities, but rather as a composite bundle of specific attributes that collectively dictate their market clearing price. Within the sports economics literature, this approach has been widely applied to the valuation of complex, non-market assets. As pioneered by DeSchriver and Jensen (2003), applying HPM to stadium naming rights treats the sponsorship agreement similarly to real estate valuation, where the total financial value is systematically decomposed into physical characteristics (such as stadium age and capacity, for example) and intangible market factors (such as media visibility and regional economic strength).

The robustness of this approach in the naming rights industry was further validated by Gerrard et al. (2007), who demonstrated that despite the apparent opacity of these corporate deals, the market is highly systematic and informationally efficient. Their findings confirmed that the price premium of a naming rights contract can be accurately broken down into distinct variables. More recently, Centracchio et al. (2023) successfully utilized this hedonic framework to isolate the value drivers in collegiate sports partnerships, providing a highly relevant methodological reference for the current research.

A primary advantage of the HPM is its interpretability. It provides direct, quantifiable estimates for the marginal value of each stadium attribute, allowing stakeholders to understand exactly how much a specific

variable contributes to the final valuation. Furthermore, its deep roots in consumer utility theory offer a solid academic foundation to mathematically justify these commercial valuations to potential corporate partners. To optimize this baseline model for the highly skewed financial realities of the global market, a natural logarithmic transformation was applied exclusively to the dependent variable (inflation-adjusted annual value). This log-linear (semi-log) specification effectively shrinks extreme financial outliers, improving the overall statistical precision of the standard regression.

However, the HPM is not without significant limitations. As Rosen (1974) originally noted, the model requires the researcher to pre-define a specific functional form. Although the applied semi-logarithmic transformation smooths the variance in the dependent variable, the independent predictors are still bound to strict linear assumptions. This introduces the risk of specification errors if the true relationship between variables is structurally more complex. Additionally, the model is notoriously sensitive to multicollinearity. In a comprehensive database comprising 306 entries and 30 independent variables, attributes like stadium capacity, average attendance, and city population are inherently correlated. This multicollinearity can distort the estimated coefficients, making it difficult for standard Ordinary Least Squares (OLS) regression to accurately isolate the individual impact of each factor. Finally, the HPM traditionally assumes market equilibrium and perfect information, conditions that are frequently unmet in the fragmented and confidential European naming rights landscape.

Despite these constraints, the hedonic model is employed in this study as a literature-backed starting point for identifying the structural drivers of naming rights value. Its inherent limitations regarding multicollinearity, functional form assumptions, and non-linear interactions directly motivate the subsequent application of regularization techniques (Ridge and LASSO regression), distributional modeling (Quantile Regression), and machine learning algorithms (Random Forest). The goal of this multi-model approach is not to replace the hedonic framework but to extend and stress-test it, ensuring that the final valuation for Sporting Clube de Braga is anchored in a robust and triangulated empirical foundation rather than the output of any single model.

3.5.2 Ridge and LASSO Regression

To address the inherent limitations of the Hedonic Pricing Model (specifically its vulnerability to multi-collinearity and overfitting when dealing with a high-dimensional dataset of 306 observations and 30 independent variables), this methodology incorporates Ridge and LASSO regressions. Both techniques belong to a family of machine learning methods known as regularization, which complement standard OLS by intentionally introducing a degree of bias into the coefficient estimates to significantly reduce the model's overall variance (James et al., 2023). While the traditional HPM strictly minimizes the residual sum of squares, Ridge and LASSO optimize the model by applying a mathematical penalty to the magnitude of the coefficients, thereby preventing any single highly correlated variable from disproportionately skewing the final valuation.

Ridge regression, originally formalized by Hoerl and Kennard (1970), applies an L_2 penalty, which is proportional to the square of the magnitude of the coefficients. In practical terms, this L_2 penalty acts like a mathematical constraint or "budget" that the model must respect. Because the penalty squares the coefficient values, it aggressively punishes disproportionately large weights. If the model attempts to rely too heavily on a single variable, the squared penalty grows exponentially, forcing the algorithm to distribute that weight more evenly. However, because squaring a very small decimal results in an even smaller number, this penalizing pressure naturally fades as the coefficients get close to zero. As mentioned earlier, this technique is particularly effective in resolving multicollinearity. By shrinking the coefficients of these correlated predictors asymptotically toward zero, Ridge regression stabilizes the model and ensures that the economic impact is more evenly distributed among the variables. However, a defining characteristic of Ridge is that it retains all predictors in the final model; it reduces their impact but rarely eliminates them entirely.

Conversely, the Least Absolute Shrinkage and Selection Operator (LASSO), introduced by Tibshirani (1996), employs an L_1 penalty, which is proportional to the absolute value of the coefficients. Unlike the squared approach in Ridge, the L_1 penalty applies a constant, linear pressure on all coefficients, regardless of how large or small they are. It essentially charges a "flat tax" on the model for every variable it tries to keep. Because this steady downward pressure does not weaken as the coefficients get smaller, it drives the weakest predictors completely out of the equation. While LASSO shares the fundamental goal of regularization with Ridge, its mathematical structure allows it to force the coefficients of less significant variables to exactly zero. This key difference enables LASSO to perform automatic feature selection, effectively identifying and isolating the most critical commercial drivers within the global naming rights market while discarding "noise." In a diverse sample encompassing various sports and geographies, this property is invaluable for distilling the 30 initial variables down to a more parsimonious and interpretable core.

Ultimately, both Ridge and LASSO provide a more robust and statistically defensible alternative to the standard HPM for predictive purposes. By constraining the coefficients, these regularization techniques prevent the model from over-fitting to the specific idiosyncrasies of the training data, thereby improving the reliability of the estimated Fair Market Value when applied to out-of-sample cases like Sporting Clube de Braga.

However, despite their advanced handling of multicollinearity and dimensionality, both Ridge and LASSO remain fundamentally linear models. They assume that the relationships between the selected attributes and the contract value are constant across the entire market spectrum, a constraint that necessitates the subsequent introduction of Quantile Regression and Random Forest algorithms.

3.5.3 Quantile Regression

While standard OLS and its regularized extensions (Ridge and LASSO) provide valuable analytical frameworks, they share a fundamental methodological limitation: they are designed to estimate the conditional mean of the dependent variable. In the highly asymmetrical and right-skewed global market of stadium naming rights, relying solely on mean-based predictions can be deeply problematic. A small number of astronomical corporate deals, frequently concentrated in top-tier North American leagues, can heavily distort the statistical average. Consequently, projecting a mean-centric valuation onto a medium-sized European sports property, like SC Braga, risks generating artificially inflated estimates, ultimately failing to capture the distinct economic realities of its specific market segment.

To overcome the disproportionate influence of these extreme financial outliers, this methodology incorporates Quantile Regression (QR), a technique formally introduced by Koenker and Bassett (1978). Instead of minimizing the sum of squared residuals to calculate an overall average, QR minimizes a sum of asymmetrically weighted absolute residuals to estimate specific conditional quantiles (such as the 25th, 50th, or 75th percentiles) of the response variable. This mathematical shift means the model is inherently robust to the extreme upper-tail values present in the 306-entry database, as the median (the 50th percentile) and other quantiles are not artificially pulled upward by mega-deals in the way that a standard mean is.

Furthermore, Quantile Regression provides a significantly more granular understanding of the naming rights landscape. It allows researchers to test whether the marginal impact of independent variables fluctuates across different tiers of the market. For instance, the financial premium associated with stadium capacity or market size may behave differently for bottom-tier contracts compared to elite-tier sponsorships. By evaluating these distinct percentiles, this approach offers a highly tailored valuation mechanism. It enables the final estimation for SC Braga to be aligned with the most appropriate market segment based on its specific attributes, ensuring a mathematically defensible Fair Market Value that ignores the distortions of the global average.

3.5.4 Random Forest

While standard OLS and its regularized extensions (Ridge and LASSO), as well as Quantile Regression, offer valuable analytical frameworks, they all share a central methodological limitation when applied to the 306-observation sample: they strictly assume that the relationship between the independent variables and the contract value is linear. In the complex naming rights ecosystem, value drivers rarely operate in continuous straight lines. For instance, the financial impact of expanding stadium capacity might grow exponentially up to a certain seating threshold and then stabilize. A purely linear model fails to capture this "ceiling" or other complex interactions between attributes, risking specification errors.

To break this linearity barrier, the methodology culminates with the application of the Random Forest algorithm, introduced by Breiman (2001). Unlike traditional approaches that attempt to fit a single mathematical equation to the data, Random Forest utilizes an ensemble learning technique. The model constructs a multitude of independent "decision trees" during the training phase. Each tree attempts to predict the contract value through a series of non-linear splits, evaluating the available independent variables. The final result is obtained by aggregating the predictions of all individual trees, which drastically reduces variance and increases overall predictive accuracy.

The primary advantage of this algorithm for evaluating Sporting Clube de Braga is its remarkable flexibility. Random Forest does not require the pre-definition of a functional form (such as the logarithmic transformation used in the HPM), autonomously discovering hidden non-linear relationships. Furthermore, by introducing randomness into the selection of predictors at each branch, the model avoids overfitting and organically bypasses the issue of multicollinearity. This final step ensures a statistically advanced valuation, capable of capturing the true dynamics of the European sports market without the constraints of classical models.

3.5.5 Model Evaluation and Synthesis Strategy

Having established the theoretical and mathematical foundations of the five selected algorithms, it is imperative to define the specific empirical criteria and pre-processing protocols by which their performance and explanatory power will be evaluated. The primary objective of this methodological framework is twofold: first, to isolate the specific variables that act as the most significant determinants of financial value in the naming rights market; and second, to determine the most robust and contextually appropriate strategy to estimate the Fair Market Value (FMV) for Sporting Clube de Braga.

To ensure the integrity of the 306-entry dataset prior to modeling, a rigorous feature exclusion protocol was applied. High-cardinality qualitative identifiers, such as *Facility Name*, *Tenant Club*, *City*, and *Country*, were intentionally excluded to prevent the exponential generation of dummy variables, which would deplete the models' degrees of freedom and induce severe multicollinearity. In parallel, all auxiliary variables utilized for the purpose of currency standardization and inflation adjustment (specifically the original currencies, reference years, and the nominal unadjusted figures for construction costs, annual values, and GDP) were removed to streamline the analytical test sets. Furthermore, variables presenting mathematical redundancy or data leakage risks were omitted. Specifically, the *Inflation Adjusted Total Value* was removed to prevent the models from bypassing market drivers through basic arithmetic (as Total Value is a direct product of Annual Value and Duration). Similarly, *Expiry Year* was excluded, as its temporal significance is already captured by the combination of contract duration and venue age.

To address missing data points without the severe depletion of sample size associated with listwise deletion, a dynamic data imputation protocol was implemented. Missing numerical values were generally

imputed using the median, while categorical missingness was addressed through the mode. Crucially, for observations where average attendance data was unavailable, a specific value was imputed based on 80% of the stadium's total capacity (a proxy supported by industry occupancy standards). To preserve contextual integrity, all imputations were executed dynamically within specific analytical subsets. For instance, a missing data point in the European Football subset was imputed utilizing the median of that specific cohort rather than global metrics. This isolated strategy neutralizes cross-market distortion, ensuring the algorithms are trained on robust, highly contextualized data structures.

Rather than simply fitting the models to the entire dataset, a standard train-test split approach was utilized. The data was partitioned using an 80/20 ratio, training the algorithms on the majority of the observations while withholding 20% of out-of-sample data to test their predictive accuracy. This mitigates the risk of overfitting and ensures the models are evaluated on their ability to generalize to unseen contracts. The comparative performance of the models will be judged based on three statistical metrics:

- **Root Mean Square Error (RMSE):** Employed to heavily penalize larger prediction errors, ensuring the chosen model is resilient to the financial outliers inherent in the sports market.
- **Mean Absolute Error (MAE):** Provides a linear measure of average prediction errors, offering high interpretability regarding the actual financial deviation from real contract values.
- **R-squared (R^2):** Utilized to determine the proportion of variance in naming rights values explained by the independent variables.

Ultimately, the algorithm demonstrating the highest predictive robustness and the most coherent alignment with the European market's structural drivers will be selected as the definitive valuation tool. By inputting the specific architectural, sporting, and demographic variables of Sporting Clube de Braga into this optimized framework, the study will generate a naming rights valuation that is both mathematically defensible and strictly aligned with the determinants of the club's specific market segment.

4 Results

This chapter presents the empirical findings derived from the econometric and machine learning models applied to the stadium naming rights market. Utilizing an initial comprehensive database of 306 entries, the data was systematically segmented into eight distinct subsets to account for geographical, sporting, and structural market heterogeneity. To ensure a coherent and comparative evaluation across different methodologies, the presentation of each algorithm follows a standardized, three-tier analytical structure. First, the Full Database Analysis provides a macroscopic overview of the overarching market determinants. Second, the Stability and Market Dimension Analysis explores four refined sub-samples based on venue exclusivity, event consistency, and competitive tiers. Finally, the Football-Specific Market Analysis isolates the unique dynamics of global and regional football markets. Through this systematic framework, the primary objective of these sections is to objectively expose how each model processes and weighs the physical, demographic, and economic characteristics of the venues to determine the financial value of naming rights agreements.

Following the individual presentation of each econometric and machine learning framework, the chapter concludes with a comprehensive comparative evaluation of the models' predictive quality. By utilizing a synthesized performance matrix across all analytical subsets, this final section assesses the relative accuracy, explanatory power, and robustness of each approach. This comparison is based on key statistical metrics, specifically R^2, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), allowing for a definitive assessment of model performance in different market contexts.

4.1 Hedonic Pricing Model (OLS)

The empirical analysis begins with the Hedonic Pricing Model (HPM), estimated through Ordinary Least Squares (OLS). As established in Section 3.5.1, this model assumes that the value of a naming rights agreement is not an indivisible unit, but rather the sum of specific attributes that comprise the sports venue and its market context. Utilizing this model as a baseline allows for the decomposition of the inflation-adjusted annual price into marginal components, identifying the absolute monetary impact (in USD) that each characteristic contributes individually to the final value. The following analysis focuses exclusively on the variables that demonstrated robust statistical significance with a p-value below 0.05.

4.1.1 Hedonic – Full Database Analysis

Applying this baseline framework to the comprehensive global dataset ($n = 306$) yields a macroscopic overview of the overarching determinants in naming rights pricing. At this aggregated level, architectural investment, urban scale, and social media reach emerge as the most robust predictors of contract value.

The specific financial contributions of these structural, demographic, and digital factors are detailed in Table 4.1.

Table 4.1: Most Impactful Variables in Hedonic Pricing Model (Full Database)

Variable	Coefficient (USD)	P-value
Intercept (const)	-23,506,583	0.0172
Inflation Adjusted Construction Cost (2025)	0.0057	< 0.001
Contract Length	199,604	0.0118
City Population	0.6106	0.0008
Past Naming Deals Count	1,571,507	0.0223
Club's Instagram Followers (Feb 2026)	0.0998	0.0003
Continent: North America	13,147,023	0.0301

Note: Based on the primary dataset ($n = 306$). Coefficients represent the absolute marginal impact in USD on the annual contract value. The intercept is included to maintain the full specification of the baseline model.

As shown in Table 4.1, the model establishes a negative intercept of approximately -$23.51 million, which is then adjusted by specific asset characteristics. Regarding the structural and contractual variables, the inflation-adjusted construction cost contributes a marginal increase of $0.0057 to the annual contract value for every dollar invested, while each additional year of contract length adds $199,604. Demographic scale, measured by city population, increases the annual value by $0.61 per inhabitant. The market experience of the venue, represented by the count of past naming deals, yields a premium of $1,571,507 per previous agreement. The club's digital presence contributes approximately $0.10 ($0.0998) per Instagram follower. Finally, a North American location generates a categorical premium of $13,147,023 compared to the baseline regions.

4.1.2 Hedonic – Stability and Market Dimension Analysis

Applying the OLS model to the four structural sub-samples (Excluding Events, Unique Agreements, Major Leagues, and Major Leagues Without Dups) yields the statistically significant variables consolidated in Table 4.2.

Table 4.2: Most Impactful Variables in Hedonic Pricing Model (Stability & Dimension)

Subset	Variable	Coefficient (USD)	P-value
Excluding Events	Intercept (const)	-23,702,289	0.0289
($n = 285$)	Inflation Adjusted Construction Cost (2025)	0.0052	< 0.001
	City Population	0.7199	0.0023
	Past Naming Deals Count	2,179,578	0.0189
	Club's Instagram Followers (Feb 2026)	0.0967	0.0008
	Continent: Asia	15,310,939	0.0259
	Continent: Europe	13,872,505	0.0341
	Continent: North America	15,234,074	0.0229
Unique Agreements	Intercept (const)	-11,674,732	0.2770
($n = 219$)	Inflation Adjusted Construction Cost (2025)	0.0053	< 0.001
	City Population	0.5348	0.0004
	County's GDP per Capita	8.8357	0.0433
	Club's Instagram Followers (Feb 2026)	0.1322	< 0.001
Major Leagues	Intercept (const)	-11,149,050	0.5056
($n = 174$)	Inflation Adjusted Construction Cost (2025)	0.0057	< 0.001
	City Population	0.7309	0.0023
	Club's Instagram Followers (Feb 2026)	0.1058	0.0070
	Category: Financial Services	6,328,976	0.0448
	Multi-sport facility? (Yes)	3,589,290	0.0447
Major Leagues	Intercept (const)	-2,216,921	0.7734
Without Dups ($n = 143$)	Inflation Adjusted Construction Cost (2025)	0.0054	< 0.001
	City Population	0.3973	0.0377
	Club's Instagram Followers (Feb 2026)	0.1551	0.0004

Note: Subsets aimed at isolating continuous stability and market dimension. Intercepts are reported regardless of statistical significance to preserve model equations.

Regarding the Excluding Events subset ($n = 285$), the model establishes an intercept of -$23,702,289. The inflation-adjusted construction cost contributes $0.0052, city population adds $0.7199, and the past naming deals count yields a premium of $2,179,578. The club's Instagram followers add $0.0967 per unit. Geographically, the continents of Asia, Europe, and North America present categorical premiums of $15,310,939, $13,872,505, and $15,234,074, respectively.

For the Unique Agreements subset ($n = 219$), the intercept is -$11,674,732. The inflation-adjusted construction cost coefficient is $0.0053, and city population adds $0.5348. The county's GDP per capita contributes $8.8357, while the club's Instagram followers add $0.1322 per unit.

In the Major Leagues subset ($n = 174$), the intercept stands at -$11,149,050. The inflation-adjusted construction cost adds $0.0057, and city population contributes $0.7309. The club's Instagram followers yield a coefficient of $0.1058. Categorically, Financial Services generate a premium of $6,328,976, and multi-sport facilities present an additional value of $3,589,290.

Finally, within the Major Leagues Without Dups subset ($n = 143$), the intercept is -$2,216,921. The inflation-adjusted construction cost contributes $0.0054, city population adds $0.3973, and the club's Instagram followers provide a coefficient of $0.1551.

4.1.3 Hedonic – Football-Specific Market Analysis

Applying the baseline model to the football-specific subsets (Football (Soccer) Only, Football outside the USA, and European Football) yields the statistically significant predictors consolidated in Table 4.3. Notably, within the "Football outside the USA" subset, 13 individual industry category dummies exhibited statistically significant negative coefficients. This clustering is a mathematical artifact of reference category distortion, commonly known as the dummy variable trap, which was exacerbated by the limited sample size ($n = 76$) and extreme outliers within the baseline category. Consequently, these categorical variables are grouped together in the presentation of the results to maintain analytical clarity.

Table 4.3: Most Impactful Variables in Hedonic Pricing Model (Football-Specific Markets)

Subset	Variable	Coefficient (USD)	P-value
Football (Soccer) Only ($n = 95$)	Intercept (const)	3,000,964	0.5877
	City Population	0.3375	0.0208
	Club's Instagram Followers (Feb 2026)	0.1290	< 0.001
Football outside the USA ($n = 76$)	Intercept (const)	25,648,031	0.0487
	Contract Length	316,580	0.0377
	Stadium Capacity	-242.69	0.0146
	Average Attendance	353.79	0.0023
	Club's Instagram Followers (Feb 2026)	0.2219	< 0.001
	Industry Categories (13 Dummies)[a]	**Various (Negative)**	**< 0.05**
	Continent: South America	-14,595,380	0.0448
European Football ($n = 57$)	Intercept (const)	-1,385,009	0.9144
	Club's Instagram Followers (Feb 2026)	0.1326	< 0.001

Note: Subsets isolating global and regional football environments. Intercepts are reported regardless of significance.
[a] Within this subset, 13 individual industry category dummies exhibited statistically significant negative coefficients. This clustering is an artifact of reference category distortion (dummy variable trap) exacerbated by the limited sample size and extreme outliers in the baseline category. They are grouped here for analytical clarity.

Regarding the Football (Soccer) Only subset ($n = 95$), the model establishes an intercept of $3,000,964. City population contributes $0.3375, and the club's Instagram followers add $0.1290 per unit.

For the Football outside the USA subset ($n = 76$), the intercept stands at $25,648,031. Contract length adds $316,580. Stadium capacity presents a coefficient of -$242.69, while average attendance contributes $353.79. The club's Instagram followers yield a coefficient of $0.2219. As previously noted, 13 industry categories present various significant negative coefficients. Geographically, the continent of South America presents a categorical coefficient of -$14,595,380.

Finally, within the European Football subset ($n = 57$), the intercept is -$1,385,009. The club's Instagram followers provide a coefficient of $0.1326. This is the only statistically relevant variable for this subset.

4.2 Ridge Regression

Following the baseline estimations, the analysis advances to Ridge Regression, implementing an L_2 regularization framework. As established in Section 3.5.2, this algorithmic approach introduces a penalty term to the loss function, systematically shrinking the coefficients of less influential variables to mitigate multicollinearity and reduce overall model variance. Because the algorithm requires standardized feature inputs prior to estimation, the resulting coefficients represent the penalized monetary impact (in USD) associated with a one-standard-deviation change in the continuous variables, or the adjusted categorical premium for the dummy variables. Furthermore, since Ridge regression penalizes rather than eliminates variables, the entire feature space remains mathematically active. To maintain analytical clarity and isolate the primary market drivers, the following analysis focuses exclusively on the top predictive variables that demonstrate an absolute impact greater than $500,000 on the annual contract value. Consequently, all categorical variables exceeding this threshold are grouped into unified control vectors within the presentation of the results. It should be noted that these grouped categorical features encompass various coefficients with opposing signs, which mathematically reflect each specific category's willingness to pay relative to its baseline reference.

4.2.1 Ridge – Full Database Analysis

Applying the penalized regression framework to the primary global dataset ($n = 306$) provides a regularized assessment of the overarching determinants in naming rights pricing. At this stage, the L_2 penalty adjusts the standardized weights of all features to mitigate multicollinearity and reduce overall model variance. The most impactful predictive variables for the full database, filtered by the established absolute monetary threshold, are presented in Table 4.4.

Table 4.4: Most Impactful Variables in Ridge Regression (Full Database)

Variable	Coefficient (USD)
Intercept (const)	7,325,986
Inflation Adjusted Construction Cost (2025)	3,804,560
Club's Instagram Followers (Feb 2026)	1,666,084
Contract Length	1,339,882
City Population	1,189,682
Past Naming Deals Count	1,026,752
Average Attendance	661,961
Inflation Adjusted County's GDP per Capita	560,023
Categorical Controls[a]	**Various (-728,344 to 808,250)**

Note: Based on the primary dataset ($n = 306$). The table reports only features exhibiting an absolute predictive magnitude greater than $500,000 to suppress secondary L_2 regularization noise.

[a] Categorical variables exceeding the absolute impact threshold were grouped for analytical clarity, as detailed in the section's introduction.

As shown in Table 4.4, the Ridge model establishes an intercept of $7,325,986. Evaluating the continuous variables per standard deviation increase, the inflation-adjusted construction cost contributes $3,804,560 to the annual contract value. The club's Instagram followers add a standardized premium of $1,666,084, while contract length contributes $1,339,882. City population increases the value by $1,189,682, and the past naming deals count yields a premium of $1,026,752. Average attendance adds $661,961, and the inflation-adjusted county's GDP per capita contributes $560,023. Finally, the grouped categorical controls exceeding the absolute impact threshold of $500,000 present various coefficients with opposing signs, ranging from -$728,344 to $808,250.

4.2.2 Ridge – Stability and Market Dimension Analysis

By enforcing the L_2 penalty across the four structural sub-samples, the Ridge algorithm systematically adjusts the standardized predictive weights while suppressing secondary features. The most impactful continuous predictors and grouped categorical vectors for these subsets, isolated by the absolute monetary threshold, are consolidated in Table 4.5.

Table 4.5: Most Impactful Variables in Ridge Regression (Stability & Dimension)

Variable	Coefficient (USD)
Excluding Events ($n = 285$)	
Intercept (const)	6,760,492
Inflation Adjusted Construction Cost	2,685,102
Club's Instagram Followers	1,453,074
City Population	1,032,671
Past Naming Deals Count	833,059
Contract Length	820,667
Average Attendance	633,037
Categorical Controls[a]	**Various (-689,806 to 963,584)**
Unique Agreements ($n = 219$)	
Intercept (const)	5,909,362
Inflation Adjusted Construction Cost	2,506,423
Club's Instagram Followers	1,963,004
City Population	857,000
Average Attendance	765,189
Inflation Adjusted County's GDP per Capita	682,906
Categorical Controls[a]	**Various (-597,583 to 891,498)**
Major Leagues ($n = 174$)	
Intercept (const)	9,460,705
Inflation Adjusted Construction Cost	2,986,792
Club's Instagram Followers	1,637,634
City Population	1,221,073
Average Attendance	637,130
Contract Length	561,033
Categorical Controls[a]	**Various (-1,064,940 to 1,519,775)**
Major Leagues Without Dups ($n = 143$)	
Intercept (const)	8,324,925
Inflation Adjusted Construction Cost	2,976,291
Club's Instagram Followers	1,709,357
City Population	967,453
Inflation Adjusted County's GDP per Capita	763,151
Contract Length	611,570
Average Attendance	598,794
Categorical Controls[a]	**Various (-659,522 to 758,115)**

Note: The table reports only features exhibiting an absolute predictive magnitude greater than $500,000. Continuous variables are evaluated per standard deviation increase.

[a] Categorical variables exceeding the absolute impact threshold were grouped for analytical clarity, as detailed in the section's introduction.

Regarding the Excluding Events subset ($n = 285$), the Ridge model establishes an intercept of \$6,760,492. Evaluating the continuous variables per standard deviation increase, the inflation-adjusted construction cost contributes \$2,685,102. The club's Instagram followers add \$1,453,074, city population increases the value by \$1,032,671, and the past naming deals count yields a premium of \$833,059. Contract length contributes \$820,667, and average attendance adds \$633,037. The grouped categorical controls present various coefficients with opposing signs, ranging from -\$689,806 to \$963,584.

For the Unique Agreements subset ($n = 219$), the intercept stands at \$5,909,362. The inflation-adjusted construction cost contributes \$2,506,423 per standard deviation. The club's Instagram followers add \$1,963,004, city population contributes \$857,000, average attendance yields \$765,189, and the county's GDP per capita adds \$682,906. The grouped categorical controls range from -\$597,583 to \$891,498.

In the Major Leagues subset ($n = 174$), the intercept is \$9,460,705. The continuous variables yield the following premiums per standard deviation: inflation-adjusted construction cost adds \$2,986,792, the club's Instagram followers contribute \$1,637,634, city population adds \$1,221,073, average attendance yields \$637,130, and contract length contributes \$561,033. The categorical controls exhibit coefficients ranging from -\$1,064,940 to \$1,519,775.

Finally, within the Major Leagues Without Dups subset ($n = 143$), the intercept is established at \$8,324,925. The inflation-adjusted construction cost contributes \$2,976,291 per standard deviation. The club's Instagram followers add \$1,709,357, city population increases the value by \$967,453, the county's GDP per capita yields \$763,151, contract length adds \$611,570, and average attendance contributes \$598,794. The grouped categorical controls range from -\$659,522 to \$758,115.

4.2.3 Ridge – Football-Specific Market Analysis

Applying the penalized Ridge regression framework to the football-specific subsets systematically adjusts the standardized predictive weights to account for the unique dynamics of the global football market. The most impactful continuous predictors and grouped categorical vectors for these subsets, isolated by the established absolute monetary threshold, are consolidated in Table 4.6.

Table 4.6: Most Impactful Variables in Ridge Regression (Football-Specific Markets)

Variable	Coefficient (USD)
Football (Soccer) Only ($n = 95$)	
Intercept (const)	6,866,033
Club's Instagram Followers	2,795,277
Inflation Adjusted Construction Cost	1,344,929
City Population	920,689
Stadium Capacity	733,955
Average Attendance	723,370
Categorical Controls[a]	**Various (-576,017 to 595,419)**
Football outside the USA ($n = 76$)	
Intercept (const)	5,618,462
Club's Instagram Followers	1,937,682
Inflation Adjusted Construction Cost	895,028
City Population	753,948
Inflation Adjusted County's GDP per Capita	731,039
Average Attendance	680,708
Categorical Controls[a]	**621,115**
European Football ($n = 57$)	
Intercept (const)	7,729,258
Club's Instagram Followers	4,053,429
Average Attendance	1,538,806
Inflation Adjusted Construction Cost	1,229,743
City Population	1,053,234
Number of Games in National Television	862,020
Categorical Controls[a]	**Various (-692,259 to 580,659)**

Note: The table reports only features exhibiting an absolute predictive magnitude greater than $500,000. Continuous variables are evaluated per standard deviation increase.

[a] Categorical variables exceeding the absolute impact threshold were grouped for analytical clarity, as detailed in the section's introduction.

Regarding the Football (Soccer) Only subset ($n = 95$), the Ridge model establishes an intercept of $6,866,033. Evaluating the continuous variables per standard deviation increase, the club's Instagram followers contribute $2,795,277, and the inflation-adjusted construction cost adds $1,344,929. City population increases the value by $920,689, stadium capacity yields $733,955, and average attendance adds $723,370. The grouped categorical controls present various coefficients with opposing signs, ranging from -$576,017 to $595,419.

For the Football outside the USA subset ($n = 76$), the intercept stands at \$5,618,462. The club's Instagram followers contribute \$1,937,682 per standard deviation, while the inflation-adjusted construction cost adds \$895,028. City population yields \$753,948, the county's GDP per capita contributes \$731,039, and average attendance adds \$680,708. Since only one categorical feature (Aviation) exceeded the regularization threshold, its individual premium is recorded at \$621,115.

Finally, within the European Football subset ($n = 57$), the intercept is established at \$7,729,258. The most impactful continuous variables include the club's Instagram followers, adding an overwhelming standardized premium of \$4,053,429, and average attendance, yielding \$1,538,806. The inflation-adjusted construction cost contributes \$1,229,743, city population adds \$1,053,234, and the number of games on national television yields \$862,020. The grouped categorical controls exhibit coefficients ranging from -\$692,259 to \$580,659.

4.3 LASSO Regression

Building upon the penalized framework introduced by the Ridge model, the analysis now transitions to the Least Absolute Shrinkage and Selection Operator (LASSO). Implementing an L_1 regularization approach, this algorithm introduces a distinct penalty term to the loss function. While Ridge systematically shrinks coefficients to mitigate multicollinearity, LASSO possesses the unique mathematical property of forcing the weights of less influential variables to exactly zero, effectively performing intrinsic feature selection alongside regularization. As for the Ridge model, because the algorithm requires standardized feature inputs prior to estimation, the resulting non-zero coefficients represent the penalized monetary impact (in USD) associated with a one-standard-deviation change in the continuous variables, or the adjusted categorical premium for the dummy variables.

Although LASSO mathematically eliminates redundant variables, the surviving feature space can still be extensive. Therefore, to ensure strict structural parity and methodological coherence with the preceding models, the presentation of results maintains the established absolute threshold of \$500,000 on the annual contract value. Furthermore, exactly as was done for the Ridge regression, all surviving categorical features exceeding this threshold are grouped into unified control vectors within the presentation of the results. It should be noted that these grouped categorical features encompass various coefficients with opposing signs, which mathematically reflect each specific category's willingness to pay relative to its baseline reference.

4.3.1 LASSO – Full Database Analysis

Applying the LASSO regression framework to the primary dataset ($n = 306$) intrinsically selects the most robust predictors while mathematically shrinking redundant features to zero. The most impactful predictive

variables for the full database, filtered by the established absolute monetary threshold, are consolidated in Table 4.7.

Table 4.7: Most Impactful Variables in LASSO Regression (Full Database)

Variable	Coefficient (USD)
Intercept (const)	7,325,986
Inflation Adjusted Construction Cost (2025)	4,482,620
Club's Instagram Followers (Feb 2026)	1,847,367
Contract Length	1,271,539
Past Naming Deals Count	1,173,237
City Population	1,089,180
Average Attendance	647,056
Categorical Controls[a]	**Various (-619,075 to 919,285)**

Note: Based on the primary dataset ($n = 306$). The table reports only features exhibiting an absolute predictive magnitude greater than $500,000 to suppress secondary L_1 regularization noise.

[a] Categorical variables exceeding the absolute impact threshold were grouped for analytical clarity, as detailed in the section's introduction.

As shown in Table 4.7, the LASSO model establishes an intercept of $7,325,986. Evaluating the continuous variables per standard deviation increase, the inflation-adjusted construction cost asserts dominance, contributing $4,482,620 to the annual contract value. The club's Instagram followers add a standardized premium of $1,847,367, while contract length contributes $1,271,539. The past naming deals count yields a premium of $1,173,237, city population increases the value by $1,089,180, and average attendance adds $647,056. Finally, the grouped categorical controls present surviving coefficients with opposing signs, ranging from -$619,075 to $919,285.

4.3.2 LASSO – Stability and Market Dimension Analysis

Applying the LASSO regression framework to the four structural sub-samples intrinsically selects the most robust predictors specific to market stability and competitive tiers. The most impactful continuous predictors and grouped categorical vectors for these subsets, isolated by the established absolute monetary threshold, are consolidated in Table 4.8.

Table 4.8: Most Impactful Variables in LASSO Regression (Stability & Dimension)

Variable	Coefficient (USD)
Excluding Events ($n = 285$)	
Intercept (const)	6,760,491
Inflation Adjusted Construction Cost	3,930,152
Club's Instagram Followers	1,783,368
Past Naming Deals Count	1,024,402
City Population	918,760
Contract Length	677,584
Categorical Controls[a]	**Various (509,744 to 1,140,851)**
Unique Agreements ($n = 219$)	
Intercept (const)	5,909,361
Inflation Adjusted Construction Cost	3,167,786
Club's Instagram Followers	2,394,913
Inflation Adjusted County's GDP per Capita	704,819
City Population	684,112
Average Attendance	544,116
Competition Level: Major League	1,017,886
Major Leagues ($n = 174$)	
Intercept (const)	9,460,704
Inflation Adjusted Construction Cost	4,021,342
Club's Instagram Followers	1,804,670
City Population	1,123,084
Past Naming Deals Count	845,259
Categorical Controls[a]	**Various (-806,553 to 1,943,656)**
Major Leagues Without Dups ($n = 143$)	
Intercept (const)	8,324,924
Inflation Adjusted Construction Cost	4,035,276
Club's Instagram Followers	1,919,745
Average Attendance	754,931
City Population	692,638
Inflation Adjusted County's GDP per Capita	510,929
Sport: Basketball	748,984

Note: The table reports only features exhibiting an absolute predictive magnitude greater than $500,000. Continuous variables are evaluated per standard deviation increase.

[a] Categorical variables exceeding the absolute impact threshold were grouped for analytical clarity, as detailed in the section's introduction.

Regarding the Excluding Events subset ($n = 285$), the LASSO model establishes an intercept of \$6,760,491. Evaluating the continuous variables per standard deviation increase, the inflation-adjusted construction cost contributes \$3,930,152. The club's Instagram followers add \$1,783,368, the past naming deals count yields \$1,024,402, and city population increases the value by \$918,760. Contract length adds a premium of \$677,584. The grouped categorical controls present surviving coefficients ranging from \$509,744 to \$1,140,851.

For the Unique Agreements subset ($n = 219$), the intercept stands at \$5,909,361. The inflation-adjusted construction cost contributes \$3,167,786 per standard deviation. The club's Instagram followers add \$2,394,913, the county's GDP per capita yields \$704,819, city population contributes \$684,112, and average attendance adds \$544,116. Notably, "Competition Level: Major League" emerges as the sole categorical variable to exceed the monetary threshold, yielding an individual premium of \$1,017,886.

In the Major Leagues subset ($n = 174$), the intercept is \$9,460,704. The continuous variables yield the following premiums per standard deviation: inflation-adjusted construction cost adds \$4,021,342, the club's Instagram followers contribute \$1,804,670, city population adds \$1,123,084, and the past naming deals count yields \$845,259. The grouped categorical controls exhibit surviving coefficients with opposing signs, ranging from -\$806,553 to \$1,943,656.

Finally, within the Major Leagues Without Dups subset ($n = 143$), the intercept is established at \$8,324,924. The inflation-adjusted construction cost contributes \$4,035,276 per standard deviation. The club's Instagram followers add \$1,919,745, average attendance increases the value by \$754,931, city population yields \$692,638, and the county's GDP per capita adds \$510,929. In this refined sample, only one categorical feature met the established absolute impact threshold: "Sport: Basketball." As the singular categorical driver surviving the L_1 penalty at this magnitude, it represents an adjusted premium of \$748,984.

4.3.3 LASSO – Football-Specific Market Analysis

Applying the LASSO regression framework to the football-specific subsets intrinsically selects the most robust predictors while adjusting for the unique dynamics of the global football market. The most impactful continuous predictors and grouped categorical vectors for these subsets, isolated by the established absolute monetary threshold, are consolidated in Table 4.9.

Table 4.9: Most Impactful Variables in LASSO Regression (Football-Specific Markets)

Variable	Coefficient (USD)
Football (Soccer) Only ($n = 95$)	
Intercept (const)	6,866,033
Club's Instagram Followers	3,543,891
Inflation Adjusted Construction Cost	1,361,965
Average Attendance	1,297,091
City Population	893,803
Categorical Controls[a]	**Various (-525,314 to 701,726)**
Football outside the USA ($n = 76$)	
Intercept (const)	5,618,462
Club's Instagram Followers	3,117,022
Inflation Adjusted County's GDP per Capita	903,381
Inflation Adjusted Construction Cost	842,837
City Population	825,241
European Football ($n = 57$)	
Intercept (const)	7,729,258
Club's Instagram Followers	4,442,886
Average Attendance	1,701,244
Inflation Adjusted Construction Cost	650,776
Club's League Performance	506,291

Note: The table reports only features exhibiting an absolute predictive magnitude greater than $500,000. Continuous variables are evaluated per standard deviation increase.

[a] Categorical variables exceeding the absolute impact threshold were grouped for analytical clarity, as detailed in the section's introduction.

Regarding the Football (Soccer) Only subset ($n = 95$), the LASSO model establishes an intercept of $6,866,033. Evaluating the continuous variables per standard deviation increase, the club's Instagram followers contribute an impressive $3,543,891. The inflation-adjusted construction cost adds $1,361,965, average attendance yields $1,297,091, and city population increases the value by $893,803. The grouped categorical controls present surviving coefficients with opposing signs, ranging from -$525,314 to $701,726.

For the Football outside the USA subset ($n = 76$), the intercept stands at $5,618,462. The club's Instagram followers assert dominance, contributing $3,117,022 per standard deviation. The county's GDP per capita adds $903,381, the inflation-adjusted construction cost yields $842,837, and city population contributes $825,241. Notably, within this specific subset, no categorical variable survived the L_1 penalty with an absolute magnitude exceeding the $500,000 threshold.

Finally, within the European Football subset ($n = 57$), the intercept is established at \$7,729,258. The most impactful continuous variable is the club's Instagram followers, adding a standardized premium of \$4,442,886, followed by average attendance, which yields \$1,701,244. The inflation-adjusted construction cost contributes \$650,776, and the club's league performance adds \$506,291. Similarly to the previous subset, the regularization framework shrank all categorical variables below the established monetary criterion, leaving only continuous predictors as the primary market drivers.

4.4 Quantile Regression

Building upon the conditional mean estimations provided by the baseline OLS and the penalized algorithms, the analysis now advances to Quantile Regression. Unlike traditional frameworks that evaluate the average effect of predictors, Quantile Regression assesses the impact of independent variables across different points of the dependent variable's distribution. By relying on absolute deviations rather than squared residuals, this methodological approach provides a highly nuanced understanding of the naming rights market, revealing how physical, demographic, and economic characteristics dynamically alter their premium weights across low-tier ($q = 0.25$), median ($q = 0.50$), and high-tier ($q = 0.75$) sponsorship agreements.

To maximize analytical clarity, the presentation of these results follows a specific structural logic. First, unlike the Ridge and LASSO models, the continuous variables in this section remain unstandardized. This allows for the direct interpretation of their absolute marginal economic impacts (e.g., the exact monetary value of one additional stadium seat or one local inhabitant) across the different market tiers. Second, to maintain consistency with previous models, only independent variables demonstrating strict statistical significance ($p < 0.05$) within their respective quantiles are reported. Third, to ensure complete transparency, any categorical variables that survive the significance threshold are presented individually rather than grouped into control vectors.

4.4.1 Quantile – Full Database Analysis

Applying the Quantile Regression to the primary global dataset demonstrates how the magnitude of core market drivers evolves across different sponsorship tiers. Table 4.10 presents the statistically significant and economically material predictors across the three evaluated percentiles.

Table 4.10: Most Impactful Variables in Quantile Regression (Full Database)

Variable	q = 0.25 (USD)	q = 0.50 (USD)	q = 0.75 (USD)
Inflation Adj. Construction Cost	0.0057	0.0071	0.0104
Club's Instagram Followers	0.1410	0.1682	0.2211
City Population	–	0.2441	0.4157
Inflation Adj. County's GDP per Capita	–	5.1229	19.0101

Note: Based on the primary dataset ($n = 306$). The table reports unstandardized marginal effects that achieved strict statistical significance ($p < 0.05$). Values marked with "–" indicate that the respective variable failed to meet the statistical significance threshold at the specific quantile.

As shown in Table 4.10, the full database reveals a progressive scaling of coefficients across the market distribution. At the lower tier, the inflation-adjusted construction cost yields a coefficient of 0.0057 and the club's Instagram followers contribute 0.1410 per unit increase. Moving to the median, both drivers intensify: construction cost rises to 0.0071 and Instagram followers to 0.1682, while macroeconomic variables activate for the first time, with city population yielding 0.2441 and GDP per capita contributing 5.1229. At the premium tier, the construction cost peaks at 0.0104 and Instagram followers reach 0.2211, with the macroeconomic effects becoming considerably more pronounced, with city population contributing 0.4157 and GDP per capita surging to 19.0101.

4.4.2 Quantile – Stability and Market Dimension Analysis

Applying the Quantile Regression to the specific control subsets validates the structural resilience of the primary market drivers. Table 4.11 presents the surviving variables that satisfied strict statistical significance ($p < 0.05$) across the respective quantiles.

Table 4.11: Most Impactful Variables in Quantile Regression (Stability & Dimension)

Variable	q=0.25 (USD)	q=0.50 (USD)	q=0.75 (USD)
Excluding Events			
Inflation Adj. Construction Cost	0.0056	0.0064	0.0086
Club's Instagram Followers	0.1365	0.1761	0.1662
City Population	0.2692	0.3325	0.7665
Inflation Adj. County's GDP per Capita	–	–	15.1339
Unique Agreements			
Inflation Adj. Construction Cost	0.0056	0.0067	0.0076
Club's Instagram Followers	0.1365	0.1780	0.1862
City Population	0.2692	0.3614	0.3714
Inflation Adj. County's GDP per Capita	–	–	19.5089
Major Leagues			
Inflation Adj. Construction Cost	0.0056	0.0065	0.0088
Club's Instagram Followers	0.1365	0.1596	0.1274
City Population	0.2692	0.3507	–
Inflation Adj. County's GDP per Capita	–	19.0274	22.5492
Major Leagues Without Dups			
Inflation Adj. Construction Cost	0.0056	0.0058	0.0083
Club's Instagram Followers	0.1104	0.1716	0.1665
City Population	0.2989	0.3363	–
Inflation Adj. County's GDP per Capita	–	–	22.9388

Note: The table reports unstandardized marginal effects that achieved strict statistical significance ($p < 0.05$). Values marked with "–" indicate that the respective variable failed to meet the statistical significance threshold at the specific quantile.

As shown in Table 4.11, the four structural subsets confirm the robustness of the core drivers identified in the full database analysis.

In the Excluding Events subset, the three primary drivers are significant across all tiers. The inflation-adjusted construction cost scales from 0.0056 at the lower tier to 0.0064 at the median and 0.0086 at the premium tier. Instagram followers follow a similar upward trajectory (0.1365, 0.1761, 0.1662), while city population exhibits a particularly pronounced escalation at the upper end (0.2692, 0.3325, 0.7665). GDP per capita activates exclusively at the premium tier with a coefficient of 15.1339.

The Unique Agreements subset replicates this pattern with notable consistency. Construction cost scales

from 0.0056 to 0.0067 and 0.0076 across the three tiers. Instagram followers rise progressively (0.1365, 0.1780, 0.1862), and city population remains broadly stable (0.2692, 0.3614, 0.3714). GDP per capita again emerges only at the premium tier, yielding 19.5089.

In the Major Leagues subset, the construction cost and Instagram followers maintain their significance across all tiers (0.0056–0.0088 and 0.1365–0.1274, respectively), while city population loses significance at the premium tier. GDP per capita activates from the median onwards, contributing 19.0274 and 22.5492 at the median and upper tiers.

Finally, the Major Leagues Without Dups subset follows a consistent structure: construction cost scales from 0.0056 to 0.0083, Instagram followers from 0.1104 to 0.1665, and city population is significant only at the lower and median tiers (0.2989 and 0.3363). GDP per capita is exclusively a premium-tier driver, commanding 22.9388 at the upper end of the distribution.

4.4.3 Quantile – Football-Specific Market Analysis

Applying the Quantile Regression to the specific sports and geographical subsets reveals distinct pricing dynamics within the global football market. Table 4.12 presents the surviving predictors that satisfied strict statistical significance ($p < 0.05$) across the respective quantiles.

Table 4.12: Most Impactful Variables in Quantile Regression (Football-Specific Markets)

Variable	Q 0.25 (USD)	Q 0.50 (USD)	Q 0.75 (USD)
Football (Soccer) Only			
Inflation Adj. Construction Cost	0.0033	0.0066	0.0076
Club's Instagram Followers	0.1129	0.1110	0.2386
City Population	0.3199	0.3794	0.4071
Football outside the USA			
Inflation Adj. Construction Cost	–	–	0.0051
Club's Instagram Followers	0.1498	0.1961	0.1837
Inflation Adj. County's GDP per Capita	–	–	37.7633
Stadium Capacity	–	–	-122.4279
Average Attendance	–	–	176.7837
European Football			
Club's Instagram Followers	0.0974	0.1429	0.1069
City Population	0.5157	–	–

Note: The table reports unstandardized marginal effects that achieved strict statistical significance ($p < 0.05$). Values marked with "**–**" indicate that the respective variable failed to meet the statistical significance threshold at the specific quantile.

As shown in Table 4.12, the football-specific subsets reveal increasingly concentrated driver structures as the sample narrows geographically.

In the Football (Soccer) Only subset, the three core drivers are significant across the entire distribution. The inflation-adjusted construction cost scales from 0.0033 at the lower tier to 0.0066 at the median and 0.0076 at the premium tier. Instagram followers remain broadly stable at the lower and median tiers (0.1129 and 0.1110) before surging to 0.2386 at the upper end. City population increases consistently across all tiers (0.3199, 0.3794, 0.4071).

The Football outside the USA subset presents a markedly different structure. Instagram followers are the only driver significant across all three tiers (0.1498, 0.1961, 0.1837), while all remaining variables activate exclusively at the premium tier: construction cost adds 0.0051, GDP per capita dictates a premium of 37.7633, and stadium metrics exhibit opposing dynamics, with average attendance contributing 176.7837 per unit while raw capacity subtracts 122.4279.

Finally, the European Football subset demonstrates the most concentrated driver structure of all subsets. Instagram followers are the sole significant variable across the entire distribution (0.0974, 0.1429, 0.1069), with city population contributing only at the lower tier (0.5157) and losing significance entirely at the median

and premium tiers.

4.5 Random Forest

To complement the parametric frameworks employed in the previous sections, which inherently assume linear relationships and assess marginal effects, the analysis now advances to a non-parametric machine learning approach: the Random Forest algorithm. Unlike traditional regression models, Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the average prediction of the individual trees. This methodology is particularly valuable for the current study because it does not require assumptions about data distribution, is highly resilient to outliers, and inherently captures complex, non-linear interactions between sponsorship drivers that linear models might fail to detect.

Because of its distinct architecture, the interpretation of results in a Random Forest model fundamentally differs from traditional econometric methods. Instead of yielding traditional marginal coefficients and p-values, the algorithm's decision-making process is decoded using SHAP (SHapley Additive exPlanations) values. While basic feature importance metrics only quantify the overall magnitude of a variable's contribution, SHAP values provide a more comprehensive evaluation by calculating the marginal contribution of each feature to the model's prediction. This econometric technique reveals both the magnitude and the directional impact (positive or negative) of each independent variable on the contract valuation. However, given potential correlation among venue, market, and audience variables, SHAP attributions should be interpreted as model-based contribution measures rather than isolated causal effects.

To ensure methodological rigor, analytical clarity, and to isolate the most impactful market drivers, the subsequent SHAP summary plots are configured to display only the top 10 most influential variables. This top 10 ranking is mathematically defined by calculating the mean absolute SHAP value for each feature, which aggregates the average magnitude of a variable's impact on the model's predictions across all observations. This structural criterion ensures that the analysis focuses exclusively on the core physical, demographic, economic, and digital factors dictating the valuation of naming rights across the different market segments.

4.5.1 Random Forest – Full Database Analysis

Applying the Random Forest algorithm to the primary global dataset ($n = 306$) isolates the core drivers of the naming rights market. Figure 4.1 presents the SHAP summary plot, detailing the directional impact of the top 10 variables ranked by their mean absolute SHAP value. Given potential correlation among venue, market, and audience variables, SHAP attributions should be interpreted as model-based contribution measures rather than isolated causal effects.

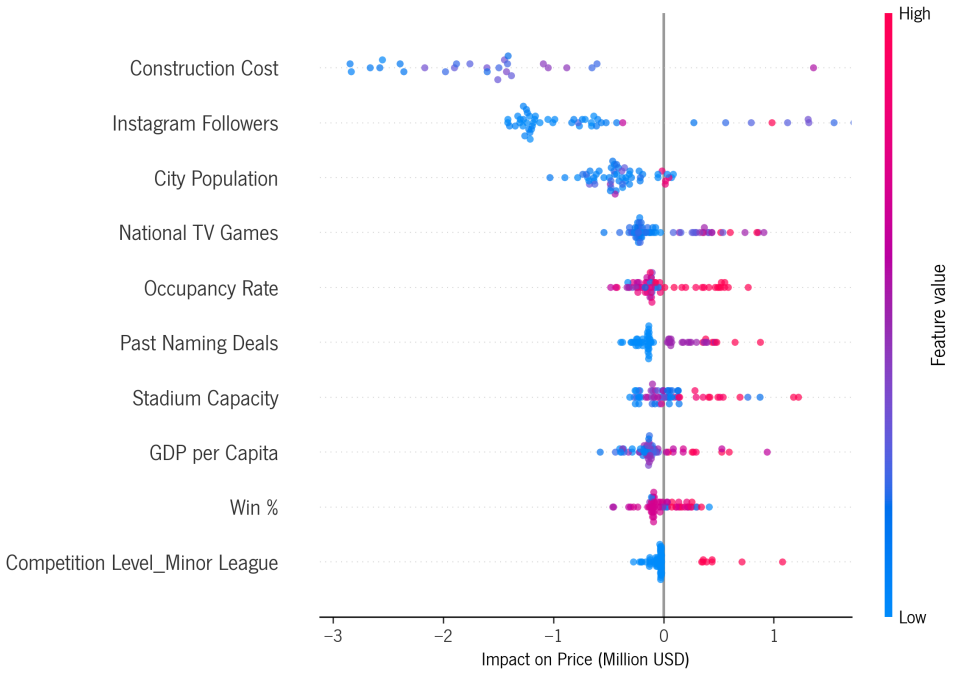


Figure 4.1: SHAP Summary Plot — Random Forest (Full Database, $n = 306$)

As shown in Figure 4.1, the global market exhibits a high concentration of model-based contributions from physical infrastructure. *Inflation Adjusted Construction Cost (2025) (USD)* accounts for 58.79% of total normalized mean absolute SHAP importance. The SHAP summary plot confirms a robust positive non-linear association: higher capital expenditures (depicted by red dots) systematically correspond to positive SHAP values, contributing to significant increases in predicted contract valuations.

Digital reach and demographics form the second tier of model contributions. *Club's Instagram Followers (February 2026)* accounts for 9.45% and *City Population* accounts for 6.37% of total normalized mean absolute SHAP importance. The SHAP summary plot indicates that robust digital followings and larger host markets consistently correlate with premium predicted contract values. Economic context directly follows, with *Inflation Adjusted County's GDP per Capita (2024 to 2025) (USD)* accounting for 3.19% of total normalized mean absolute SHAP importance, also exhibiting a positive model-based contribution.

Operational and sporting metrics occupy the remainder of the top 10. *Stadium Capacity* (2.89%) and *Club's All-Time Winning Percentage in Top Tier League* (2.80%) both contribute positively to the model's valuation. Finally, variables such as *Average Occupancy Rate (Last season - 2024 or 2024/25)* (1.95%), *Number of Games in National Televison (Last season - 2024 or 2024/25)* (1.82%), the categorical classification *Competition Level_Minor League* (1.54%), and *Club's League Performance (Last season's position - 2024 or 2024/25)* (1.40%) provide marginal refinements to the valuation. Crucially, the SHAP summary plot reveals strict directional penalties for the latter two: competing in a minor league or having a poorer league position (higher numerical rank, depicted in red) systematically generates negative SHAP values, resulting in lower predicted financial outputs. Consistent with the methodological framework, these SHAP attributions

represent model-based contribution measures within a correlated feature space rather than isolated causal effects.

4.5.2 Random Forest – Stability and Market Dimension Analysis

Applying the Random Forest algorithm to the four structural sub-samples reveals how market stability and competitive tiers shift the model-based contributions of key naming rights drivers. A comparative analysis of the SHAP summary plots across these segments, Excluding Events ($n = 285$), Unique Agreements ($n = 219$), Major Leagues ($n = 174$), and Major Leagues Without Dups ($n = 143$), demonstrates a high degree of structural convergence, with the primary model-based contribution measures remaining consistent across all models. Consistent with the methodological framework, given potential correlation among venue, market, and audience variables, these SHAP attributions should be interpreted as model-based contribution measures rather than isolated causal effects.

To avoid analytical redundancy and maintain focus on the most refined market data, the visual evaluation in this section centers on the Major Leagues Without Dups subset. This sample is prioritized as it represents the most rigorous refinement of the database, ensuring only one representative entity per major league venue is considered. By isolating the highest competitive tier and eliminating duplicate entities, this subset offers the clearest insight into the structural drivers of the global elite market. The SHAP plots for the remaining three subsets are provided in Appendix A to document the observed market stability.

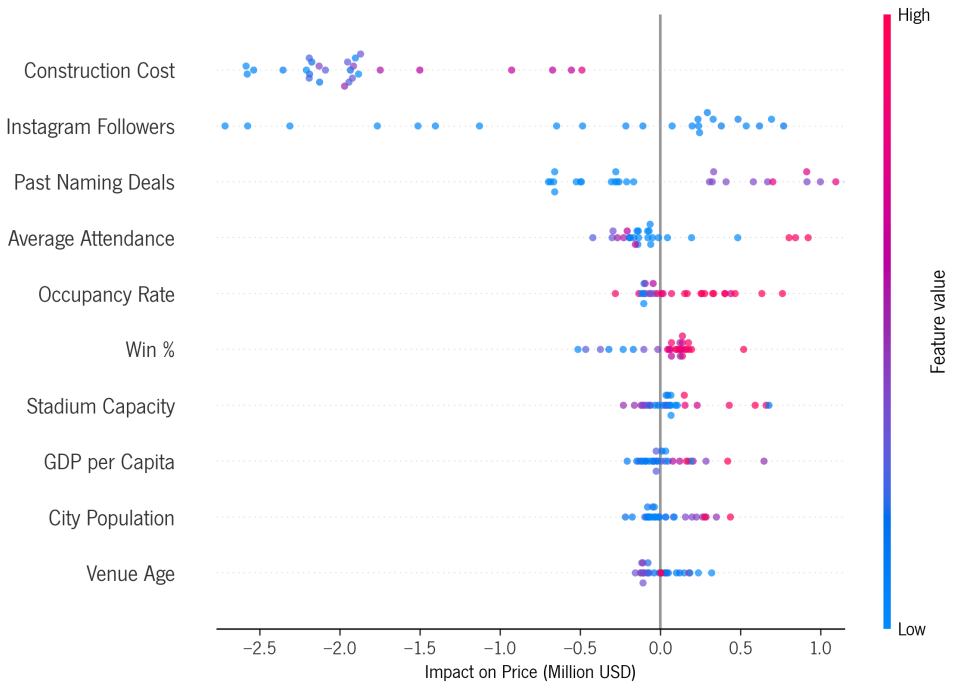


Figure 4.2: SHAP Summary Plot — Random Forest (Major Leagues Without Dups, $n = 143$)

As shown in Figure 4.2, *Inflation Adjusted Construction Cost (2025) (USD)* remains the primary contributor

within the elite market, accounting for 53.37% of the total normalized mean absolute SHAP importance. The SHAP summary plot illustrates a clear positive association, where higher venue investment systematically corresponds to positive SHAP values. *Club's Instagram Followers (February 2026)* (20.57%) follows as the secondary model-based contribution measure, also exhibiting a robust positive association with predicted contract pricing.

Operational usage and commercial history provide further positive contributions to the model's predictions. *Average Attendance (Last season - 2024 or 2024/25)* (4.33%), *Average Occupancy Rate (Last season - 2024 or 2024/25)* (3.43%), and *Past Naming Deals Count* (2.19%) all generate positive SHAP values, indicating that both actual crowd engagement and commercial maturity are associated with higher predicted naming rights revenue. Secondary factors such as *Inflation Adjusted County's GDP per Capita (2024 to 2025) (USD)* (2.16%), *Stadium Capacity* (2.15%), and *Club's All-Time Winning Percentage in Top Tier League* (2.03%) also provide positive model-based contributions to the final valuation.

Conversely, the model identifies two primary depreciating factors. *Venue Age (2025)* and *Club's League Performance (Last season's position - 2024 or 2024/25)* both exhibit negative model-based contributions; older facilities and poorer league rankings (higher numerical positions, depicted in red) systematically correlate with negative SHAP values, thereby depressing the predicted financial valuations. Consistent with the methodological framework, given potential correlation among venue, market, and audience variables, these SHAP attributions should be interpreted as model-based contribution measures rather than isolated causal effects.

4.5.3 Random Forest – Football-Specific Market Analysis

As in the previous section, the analysis is restricted to a single representative subset to avoid visual redundancy. The European Football sample ($n = 57$) is prioritized as it provides the most geographically and contextually relevant framework for the subsequent case study. The SHAP summary plots for the Football (Soccer) Only and Football outside the USA subsets are provided in Appendix B, as they demonstrate structural convergence. Consistent with the methodological framework, given potential correlation among venue, market, and audience variables, these SHAP attributions should be interpreted as model-based contribution measures rather than isolated causal effects.

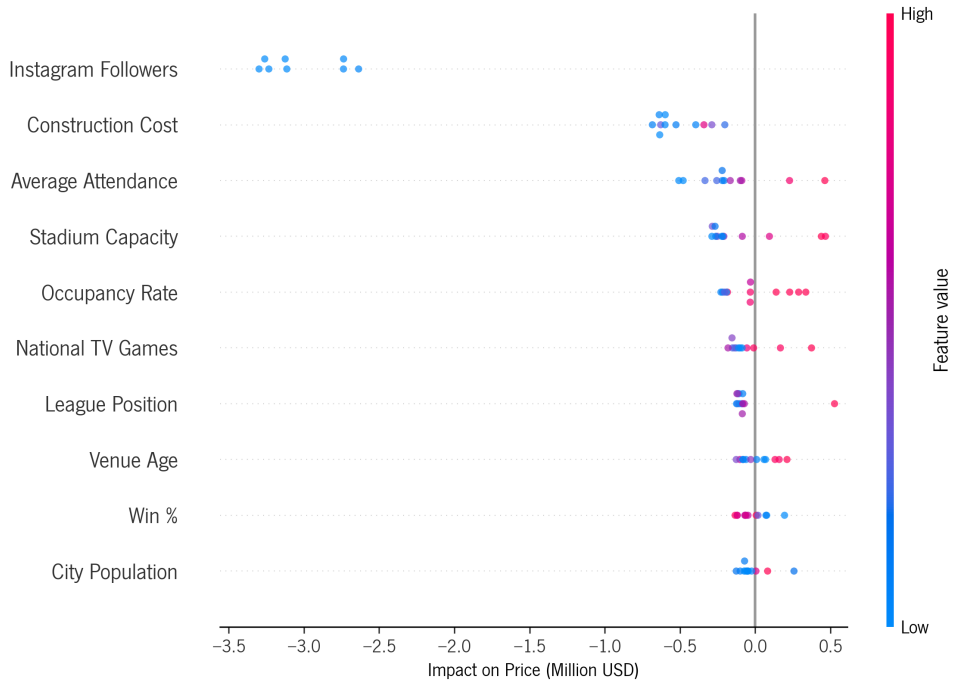


Figure 4.3: SHAP Summary Plot — Random Forest (European Football, $n = 57$)

Evaluating the European Football segment, depicted in Figure 4.3, reveals a radical concentration of model-based contributions. *Club's Instagram Followers (February 2026)* is identified as the paramount driver of valuation, accounting for 61.09% of total normalized mean absolute SHAP importance. The SHAP summary plot for this variable shows a concentrated distribution within the test set, asserting its high structural relevance in the ranking. *Inflation Adjusted Construction Cost (2025) (USD)* follows as the secondary driver, accounting for 15.32% of total normalized mean absolute SHAP importance, exhibiting a clear positive association where capital-intensive facilities systematically correspond to higher predicted naming rights premiums.

Operational and physical metrics occupy the third tier of model contributions. *Stadium Capacity* (3.78%), *City Population* (3.50%), *Average Occupancy Rate (Last season - 2024 or 2024/25)* (3.27%), and *Average Attendance (Last season - 2024 or 2024/25)* (3.17%) all maintain positive model-based contributions to the final predicted price. Additionally, *Number of Games in National Televison (Last season - 2024 or 2024/25)* contributes 2.12% to the total normalized mean absolute SHAP importance. Finally, the analysis confirms the depreciating effect of *Venue Age (2025)* (1.62%) and poor *Club's League Performance (Last season's position - 2024 or 2024/25)* (1.42%), both of which act as directional penalties that reduce predicted contract valuations in the European market. Consistent with the methodological framework, given potential correlation among venue, market, and audience variables, SHAP attributions should be interpreted as model-based contribution measures rather than isolated causal effects.

4.6 Model Performance and Quality Evaluation

To determine the predictive validity and absolute accuracy of the implemented methodologies, this section presents the out-of-sample evaluation metrics across all algorithms and analytical subsets. Table 4.13 summarizes the performance using the coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

Table 4.13: Comparative Performance Metrics Across Models and Subsets

Metric	OLS	Ridge	LASSO	Q 0.25	Q 0.50	Q 0.75	Random Forest
Full Database							
R^2	0.553	0.545	0.531	0.313	0.438	0.469	0.616
RMSE (USD)	5,957,019	6,009,224	6,101,738	7,980,720	7,309,606	8,234,198	5,520,343
MAE (USD)	3,584,613	3,591,399	3,525,292	3,934,702	3,365,260	4,310,811	3,081,589
Excluding Events							
R^2	0.630	0.632	0.639	0.284	0.389	0.410	0.848
RMSE (USD)	5,143,690	5,128,669	5,075,991	6,966,609	5,748,154	5,035,793	3,294,706
MAE (USD)	3,652,569	3,305,611	3,268,757	4,101,431	3,197,928	3,479,924	2,306,446
Unique Agreements							
R^2	0.461	0.519	0.519	0.339	0.471	0.514	0.554
RMSE (USD)	6,076,857	5,738,440	5,738,037	7,146,772	6,134,032	6,026,653	5,522,987
MAE (USD)	4,075,782	3,494,633	3,138,082	4,283,044	3,526,118	3,871,469	2,898,691
Major Leagues							
R^2	0.550	0.673	0.694	0.240	0.308	0.346	0.698
RMSE (USD)	5,676,392	4,833,500	4,678,044	6,258,109	4,740,525	7,094,379	4,644,590
MAE (USD)	4,898,350	3,990,588	3,718,379	4,672,279	3,586,925	4,553,765	3,439,908
Major Leagues Without Dups							
R^2	0.339	0.349	0.327	0.301	0.421	0.492	0.217
RMSE (USD)	7,180,688	7,128,225	7,248,363	9,228,724	8,177,674	8,184,385	7,818,069
MAE (USD)	4,496,216	4,292,509	4,277,741	6,010,566	5,177,122	5,000,632	5,178,393
Football (Soccer) Only							
R^2	0.426	0.480	0.465	0.392	0.456	0.224	0.395
RMSE (USD)	5,293,487	5,038,651	5,110,949	5,036,923	4,892,885	5,801,927	5,434,441
MAE (USD)	3,775,877	3,937,239	3,826,390	3,660,710	3,924,343	4,076,500	3,611,474
Football outside the USA							
R^2	0.042	0.620	0.731	0.373	0.442	0.640	0.439
RMSE (USD)	7,076,836	4,453,150	3,746,079	5,251,989	4,637,048	5,020,950	5,411,719
MAE (USD)	5,971,658	3,907,484	3,178,003	4,621,387	3,931,101	3,874,255	4,045,221
European Football							
R^2	0.554	0.586	0.663	0.378	0.586	0.598	0.775
RMSE (USD)	4,489,805	4,325,178	3,901,682	5,418,515	3,708,136	3,999,319	3,188,899
MAE (USD)	3,395,899	3,429,218	3,292,009	3,773,250	2,758,001	3,745,804	2,478,564

Note: The table reports out-of-sample evaluation metrics computed exclusively on the 20% test sets. For the Quantile Regression framework, R^2 values represent the corresponding Pseudo-R^2 generated for each specific quantile. Error metrics (RMSE and MAE) are reported in absolute, non-standardized United States Dollars (USD).

In the Full Database, the Random Forest achieves the highest R^2 at 0.616 (RMSE $5,520,343; MAE $3,081,589). Among the parametric models, performance converges closely: OLS returns R^2 0.553, Ridge 0.545, and LASSO 0.531, with RMSE and MAE values ranging from $5,957,019–$6,101,738 and $3,525,292–$3,591,399 respectively. Within the Quantile Regression framework, Pseudo-R^2 values scale from 0.313 at the lower tier to 0.438 at the median and 0.469 at the upper tier, with RMSE values between $7,309,606 and $8,234,198.

The Excluding Events subset produces the strongest Random Forest result in the stability group, with R^2 0.848, RMSE $3,294,706, and MAE $2,306,447, a substantial improvement over the full sample. The parametric models remain tightly clustered (OLS 0.630, Ridge 0.632, LASSO 0.639), while Quantile Regression Pseudo-R^2 ranges from 0.284 to 0.410 across the three tiers.

In the Unique Agreements subset, the Random Forest leads with R^2 0.554. Ridge and LASSO both achieve R^2 0.519, with LASSO producing a lower MAE ($3,138,082 vs $3,494,634), while OLS trails at 0.461. Quantile Pseudo-R^2 values range from 0.339 to 0.514.

The Major Leagues subset sees the closest competition between the ensemble and parametric methods: Random Forest achieves R^2 0.698 and LASSO follows closely at 0.694, with Ridge at 0.673 and OLS at 0.550. Quantile performance is more dispersed, with Pseudo-R^2 ranging from 0.240 at the lower tier to 0.346 at the upper tier.

The Major Leagues Without Dups subset produces the lowest performance levels across all models. Notably, the parametric models outperform the Random Forest here, with OLS achieving R^2 0.339 and Ridge 0.349, compared to 0.217 for the Random Forest, suggesting that the limited sample size ($n = 143$) constrains ensemble learning more than linear estimation. Q 0.75 achieves the highest R^2 in this subset at 0.492.

In the Football (Soccer) Only subset, no single model dominates. Ridge leads among the parametric models at R^2 0.480, followed by LASSO at 0.465 and OLS at 0.426. The Random Forest returns R^2 0.395, its weakest result among the football-specific subsets. Quantile Pseudo-R^2 ranges from 0.392 to 0.456, with Q 0.75 dropping to 0.224.

The Football outside the USA subset reveals the largest inter-model divergence in the analysis. LASSO achieves the highest R^2 at 0.731 (RMSE $3,746,080; MAE $3,178,004), while OLS collapses to R^2 0.042, consistent with severe multicollinearity in this geographically diverse but small subset. Ridge regularization recovers substantially at 0.620, and the Random Forest returns 0.439.

Finally, the European Football subset delivers the strongest overall performance. The Random Forest achieves R^2 0.775, RMSE $3,188,899, and MAE $2,478,564, the best absolute error metrics across all subsets and models. LASSO follows at R^2 0.663, Ridge at 0.586, and OLS at 0.554. Quantile Pseudo-R^2 ranges from 0.378 at the lower tier to 0.599 at the upper tier, with Q 0.50 matching Ridge at 0.586.

5 Discussion

This chapter provides a critical interpretation and discussion of the empirical findings presented in Section 4, bridging the gap between raw statistical output and actionable market insights. The analysis is structured to address two fundamental research inquiries: first, identifying the primary structural and digital determinants that dictate the financial valuation of naming rights, contextualizing these empirical findings within the existing sports economics literature; and second, evaluating the comparative performance and stability of the predictive models across different market subsets to isolate the optimal forecasting framework for the SC Braga profile. By synthesizing these analytical dimensions, this chapter successfully transitions from broad market dynamics to the applied empirical foundation required for the case study in Section 6.

5.1 Main Determinants of Naming Rights Value

The following analysis synthesizes the empirical results to identify the core structural drivers of naming rights valuation. To ensure a robust and objective discussion, a meta-analysis was conducted to determine the "Model Consensus Rate" of the primary market determinants. For the purpose of this analysis, a variable is considered a primary driver if it acts as a statistically significant regressor in parametric frameworks (OLS, Ridge, LASSO, Quantile) or ranks within the top 10 features defined by their mean absolute SHAP value in the Random Forest algorithm.

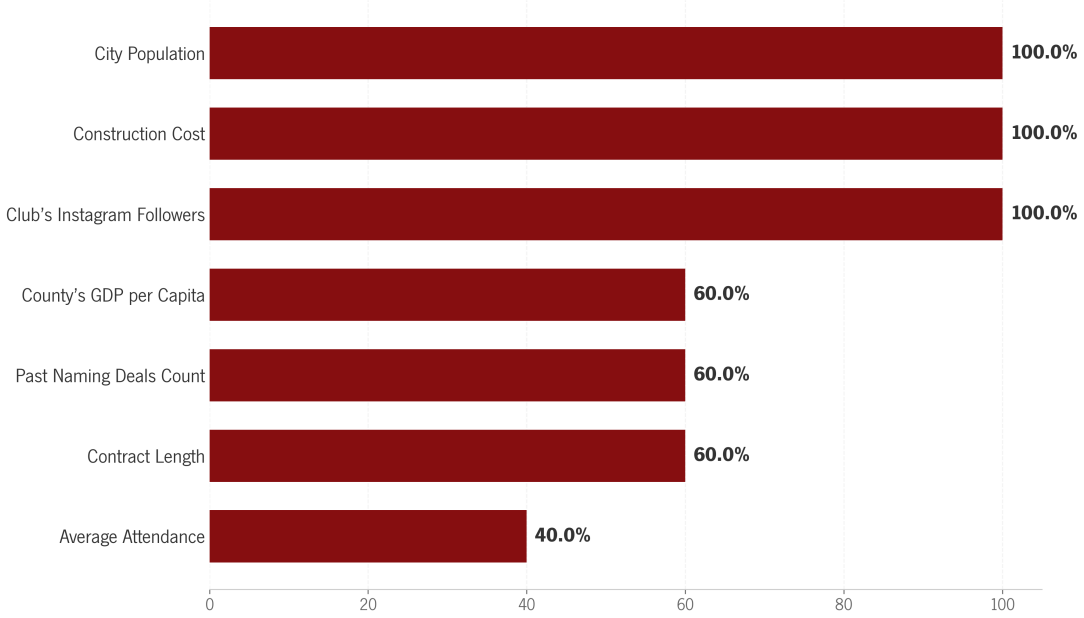


Figure 5.1: Model Consensus Rate — Primary Drivers of Naming Rights Value (Full Database)

Figure 5.1 establishes the baseline consensus structure across the full cross-sport, cross-continent sample. Three variables, *Club's Instagram Followers*, *Inflation Adjusted Construction Cost*, and *City Population*,

emerge with near-universal consensus, while a second tier of contractual and economic variables achieves moderate rates. As the sample is progressively refined to control for sport-level and structural heterogeneity, this hierarchy evolves in ways that are themselves analytically informative, as shown in Figure 5.2.

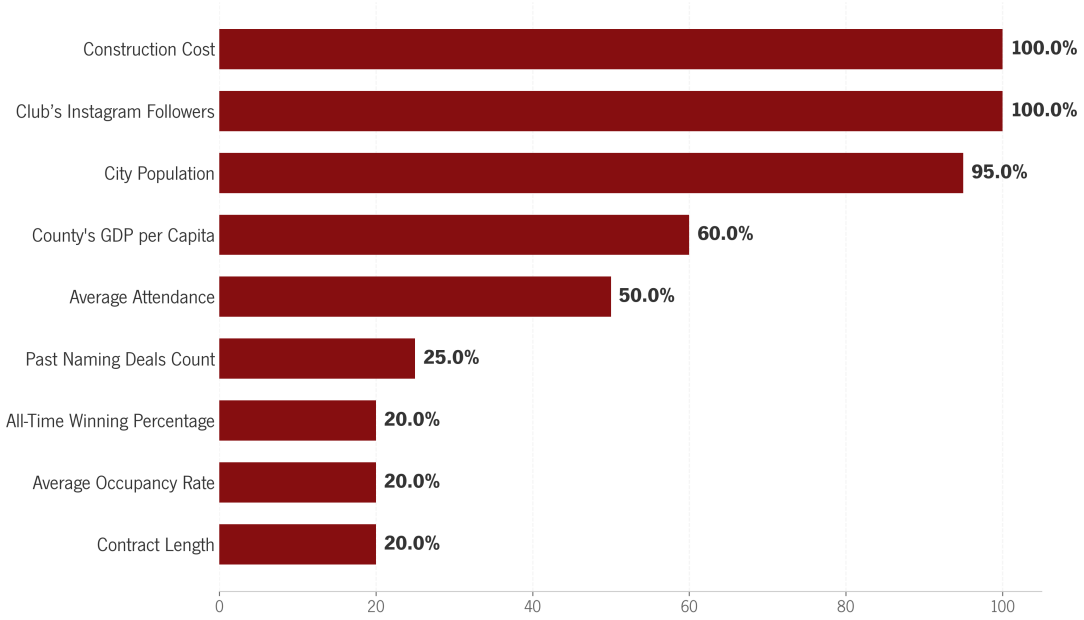


Figure 5.2: Model Consensus Rate — Primary Drivers of Naming Rights Value (Stability & Dimension)

Across the Stability and Market Dimension subsets, the structural core consolidates further while contractual variables begin to lose discriminatory power. The most pronounced shift occurs as the analysis narrows to football-specific contexts, where the commercial logic of naming rights takes on a distinctly different character, as illustrated in Figure 5.3.

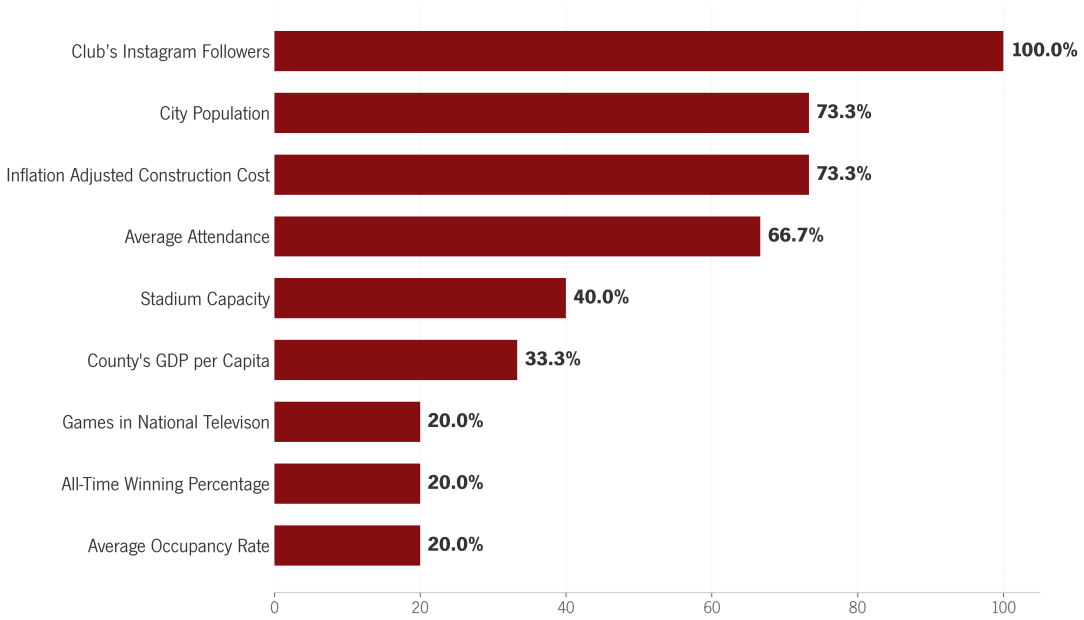


Figure 5.3: Model Consensus Rate — Primary Drivers of Naming Rights Value (Football-Specific)

The football-specific segment reveals the most structurally distinctive pattern: *Club's Instagram Followers* retains 100% consensus while *Construction Cost* declines, and operational variables such as *Average Attendance* and *Stadium Capacity* gain traction for the first time. The following subsections examine these shifts in detail, drawing on the theoretical framework established in Section 2 to contextualize the empirical findings.

5.1.1 Core Value Determinants

The most immediate finding across all three segments is the existence of a structural core of three variables that achieve near-universal model consensus: *Club's Instagram Followers*, *Inflation Adjusted Construction Cost*, and *City Population*. These variables emerge as primary drivers regardless of whether the full cross-sport, cross-continent database is considered or whether the analysis is progressively narrowed towards more homogeneous subsets, suggesting that they capture fundamental dimensions of naming rights value that transcend sport, geography, and competition level.

However, while these three variables share the same degree of consensus, they do not play the same role in the valuation process. The *Inflation Adjusted Construction Cost* and *City Population* function as what may be termed value anchors: they establish the structural ceiling of the market by capturing the scale and prestige of the physical asset and the economic dimension of its surrounding market, respectively. This finding is broadly consistent with the existing literature, though with important nuances. DeSchriver and Jensen (2003), in their seminal study of North American professional sports facilities, found metropolitan population to be a marginally significant positive predictor of naming rights price, interpreting it as a proxy for advertising reach, concluding that sponsors effectively pay more to place their name in front of a larger audience. Gerrard et al. (2007) reinforced this conclusion across a broader sample of 112 North American deals spanning both major-league and non-major-league venues, finding both population and local income to be significant positive determinants of total deal value. More recently, Centracchio et al. (2023), working exclusively in the NCAA collegiate market, found construction cost to be a statistically significant predictor, with the model implying that each additional $100 million in construction cost is associated with roughly $217,000 in additional annual naming rights value. The persistence of construction cost and market size in the present study, which extends beyond North America to include European football, a fundamentally different competitive and cultural context, suggests that these structural fundamentals are not merely artefacts of the American market but reflect genuinely universal commercial logic: sponsors pay for reach, and reach is conditioned by the scale of the asset and the size of the market it operates in.

It is worth noting, however, a meaningful divergence between the present findings and those of Gerrard et al. (2007) regarding the specific physical variable that dominates valuation. Gerrard et al. (2007) found *Facility Capacity* to carry the largest t-ratio of any variable in their model, effectively making it the single most powerful predictor of naming rights value in North American markets up to 2002. In the present study, *Construction Cost* leads with 100% consensus in both the Full Database and Stability and Market

Dimension segments, while *Stadium Capacity* only emerges in the Football-Specific subset at a considerably lower consensus rate of 40%. *Construction Cost* encapsulates not only the scale of the asset but also the quality of the investment and its renovation history, making it a richer signal of asset prestige than raw seat count. The reasons why *Stadium Capacity* behaves differently across segments, and why it only gains traction in football-specific contexts, are examined in the following section.

The *Club's Instagram Followers*, by contrast, plays a fundamentally different role. Rather than anchoring value, it acts as a commercial accelerator, amplifying the premium that structural factors alone would predict. Its behavior across the three segments is particularly telling: it is the only variable to achieve 100% consensus in all three contexts without exception, and it does so while the consensus rate of *Construction Cost* declines from 100% in the Full Database and Stability and Market Dimension subsets to 73% in Football-Specific subsets. This divergence is arguably the most significant empirical finding of this analysis. In football-specific contexts, where venue construction costs are more homogeneous and the physical differentiation between assets is narrower than in the multi-sport North American market, digital reach becomes the primary discriminating factor between valuation tiers.

This finding represents a substantive departure from the prior literature, one that the prior literature itself anticipated but could not test. Centracchio et al. (2023) explicitly acknowledge the absence of social media variables as a limitation of their study, noting that "for social media exposure, because popular social media apps such as Twitter and Instagram have been introduced fairly recently, complete and reliable data would not have been available for all the venues included in this study at the time their naming rights deals were signed." DeSchriver and Jensen (2003) and Gerrard et al. (2007), writing earlier still, operated in a market where digital audience metrics did not exist as a commercial construct. The present study, by incorporating *Club's Instagram Followers* as of February 2026, is therefore among the first to empirically demonstrate what the prior literature could only theorize: that digital influence has not merely joined the set of relevant determinants but has, in football-specific contexts, displaced the physical asset as the dominant driver of commercial value. This constitutes a finding with direct implications for how European clubs without active naming rights agreements, such as the case study at the centre of this thesis, should position themselves in the sponsorship market.

5.1.2 Impact of Sample Refinement

Beyond the structural core identified in Section 5.1.1, the meta-analytical framework reveals a second layer of determinants whose statistical relevance shifts meaningfully as the sample is progressively refined. Rather than a static set of drivers, the evidence points to a dynamic landscape in which some variables gain discriminatory power as heterogeneity is reduced, while others lose it, a pattern that is itself theoretically informative.

The most consistent pattern of decline belongs to *Contract Length* and *Past Naming Deals Count*. Both

variables achieve a 60% consensus rate in the Full Database, suggesting that, across the broad multi-sport sample, the contractual architecture of a deal and the depth of prior sponsorship activity at a venue carry meaningful explanatory power. However, both variables decline sharply in the Stability and Market Dimension subsets (*Contract Length* to 20% and *Past Naming Deals Count* to 25%) and disappear entirely from the Football-Specific segment. This trajectory is plausibly explained by the relative immaturity of the European naming rights market compared to its North American counterpart. In the football-specific subset, the majority of clubs have zero or one previous naming rights agreement, and contract durations tend to follow more uniform sector conventions than in the multi-sport North American market. When there is insufficient cross-sectional variation in a variable, it loses its capacity to explain variation in price, which is precisely what appears to occur here. This interpretation should be treated as a plausible reading of the data rather than a definitive conclusion, given that the present study does not directly test market maturity as a moderating factor.

County's GDP per Capita follows a different trajectory. It holds at 60% across both the Full Database and the Stability and Market Dimension subsets, suggesting that it functions as a stable and granular refinement of the broader market size signal captured by *City Population*: once sport-level noise is controlled for, the economic purchasing power of the immediate catchment area becomes a precise and consistent measure of commercial potential. However, its consensus rate declines to 33% in the Football-Specific segment, where *City Population* itself also falls from 95% to 73%, suggesting that in a more homogeneous competitive context, macro-economic geography becomes a less decisive differentiator.

Moving in the opposite direction, *Average Attendance* shows a consistent upward trajectory across all three segments: from 40% in the Full Database to 50% in the Stability and Market Dimension subsets, reaching 67% in the Football-Specific segment, making it the third most consistently identified driver in that context. Two further operational variables, *Average Occupancy Rate* and *Club's All-Time Winning Percentage*, emerge for the first time at 20% in the Stability subsets and maintain that level in the Football-Specific segment. *Number of Games in National Television* (20%) appears exclusively in the Football-Specific segment.

The most structurally revealing pattern concerns *Stadium Capacity*, which is absent from both the Full Database and Stability subsets but emerges at 40% in the Football-Specific segment. This behavior diverges from Gerrard et al. (2007), who found capacity to be the dominant predictor in a North American sample. In the present Full Database, which spans a wide range of venue types and sports, capacity does not emerge as a primary driver, suggesting that its explanatory power is absorbed by other structural variables that may capture asset scale more comprehensively. It is only when the sample is narrowed to football-specific contexts that *Stadium Capacity* gains traction. Within this segment, the coefficient of variation for stadium capacity among European football venues is 0.395, considerably lower than the 0.723 observed across North American venues in the full sample. In a market where the physical scale of competing assets is more compressed, the remaining variation in capacity becomes commercially legible: a stadium seating 75,000 versus one seating 25,000 within the same competitive context represents a meaningfully different advertising proposition, and sponsors appear to price this accordingly. *Stadium Capacity* therefore does

not become relevant because the sample narrows, but it becomes readable as a signal precisely because the surrounding structural noise has been removed.

The emergence of *Average Attendance* as a dominant football-specific driver is directly consistent with the findings of Centracchio et al. (2023), whose final predictive model identified attendance as the single most statistically significant predictor of collegiate naming rights value, with each additional 1,000 spectators associated with approximately $25,000 in additional annual deal value. DeSchriver and Jensen (2003) and Gerrard et al. (2007), by contrast, found attendance to be statistically insignificant in their models of professional North American markets. This divergence likely reflects the multi-sport composition of those samples, where structural variation in venue type and league format dominates the variance in naming rights value and masks the contribution of operational metrics. In the present football-specific context, *Average Attendance* emerges as the primary proxy for the intensity of the bond between club and community, and by extension for the size and loyalty of the live audience that sponsors are ultimately purchasing. This interpretation is further supported by the simultaneous relevance of *Average Occupancy Rate*: while *Average Attendance* captures the absolute volume of spectators, *Average Occupancy Rate* captures the efficiency with which that volume is generated relative to the venue's capacity. Together, the two variables provide a more complete picture of live audience delivery than either metric could offer alone.

5.2 Model Performance and Selection for the Case Study

The second dimension of the empirical analysis shifts focus from the identity of the key determinants to the predictive capacity of the models themselves. While Section 5.1 addressed what drives naming rights value, this section addresses how well that value can be estimated, and under what conditions each modeling approach performs most reliably. Table 4.13 reports the full set of out-of-sample performance metrics across all five models and eight data subsets. The discussion that follows does not reproduce those figures in detail but instead interprets the patterns they reveal, with particular attention to the conditions under which each methodology excels or deteriorates, the theoretical reasons behind those patterns, and the implications for model selection in the context of the SC Braga case study developed in Section 6.

5.2.1 Model Evaluation and Cross-Subset Stability

Across the eight subsets, three broad patterns emerge from the comparative analysis of model performance. The first concerns the general hierarchy between methodologies; the second concerns how that hierarchy shifts as the sample is refined; and the third concerns the conditions under which individual models deteriorate in ways that have direct implications for their practical application.

Regarding the general hierarchy, the Random Forest algorithm consistently achieves the highest R^2 across the majority of subsets, reflecting its capacity to capture non-linear interactions and complex variable

dependencies that linear frameworks cannot accommodate by construction. This advantage is most pronounced in the Excluding Events subset, where the Random Forest reaches its global performance peak with an R^2 of 0.848 and an MAE of $2,306,447, a level of predictive accuracy that no linear model approaches in any subset. Among the linear frameworks, the regularized models (Ridge and LASSO) consistently outperform OLS, which is theoretically expected: in a dataset with a large number of predictors relative to observations, OLS is susceptible to overfitting and coefficient instability, while L_1 and L_2 penalization shrink redundant coefficients and improve out-of-sample generalization. LASSO, through its variable selection property, tends to produce marginally lower MAE values than Ridge in most subsets, suggesting that the sparsity it imposes reflects genuine signal rather than arbitrary exclusion. The Quantile Regression framework occupies a distinct role: its Pseudo-R^2 values are systematically lower than those of the parametric linear models, though this reflects a difference in the underlying metric rather than inferior predictive capacity. The Pseudo-R^2 measures the proportional reduction in weighted absolute deviation relative to a null model, and is not directly comparable to the conventional R^2 reported for OLS, Ridge, LASSO, and Random Forest. What the framework offers instead is distributional flexibility, and its median estimator (Q 0.50) produces competitive MAE values in several subsets, most notably in European Football, where it achieves the lowest MAE of any model at $2,758,002.

The second pattern concerns how this hierarchy evolves across subsets. The most structurally informative observation is the divergence between the Random Forest and the linear models as the sample size decreases. In larger subsets, Full Database ($n = 306$), Excluding Events ($n = 285$), and Major Leagues ($n = 174$), the Random Forest's advantage over LASSO in R^2 terms is substantial, ranging from 8 to 21 percentage points. However, in the Football (Soccer) Only subset ($n = 95$), this advantage collapses entirely: the Random Forest returns an R^2 of 0.395, below Ridge (0.480), LASSO (0.465), and even the Q 0.50 quantile (0.456). This pattern is consistent with the well-documented sensitivity of ensemble tree methods to sample size: Random Forest requires sufficient observations to build diverse trees that generalize reliably, and when that condition is not met, its variance reduction advantage disappears and its bias increases. The Major Leagues Without Dups subset ($n = 143$) further illustrates this: the Random Forest reaches its lowest R^2 across all configurations at 0.217, performing worse than every linear model and even the Q 0.75 quantile, which achieves an R^2 of 0.492 in the same subset. This represents a meaningful reversal of the general hierarchy and signals that, in structurally constrained samples, the complexity of the Random Forest becomes a liability rather than an asset.

The European Football subset ($n = 57$) is particularly relevant in this regard, given its direct connection to the SC Braga case study. Here, the Random Forest recovers strongly with an R^2 of 0.775 and an MAE of $2,478,564, its second-best performance across all subsets. This suggests that, despite the small sample size, the structural patterns of the European football naming rights market are sufficiently consistent for the algorithm to generalize effectively. LASSO also performs well, achieving an R^2 of 0.663 and an MAE of $3,292,009, confirming the robustness of regularized linear approaches in this context.

However, the strong aggregate performance of the Random Forest in the European Football subset must

be interpreted alongside an important methodological caveat. With only 12 test observations (the 20% holdout of a 57-observation dataset), the SHAP values generated for the European Football model reflect an extremely limited and potentially unrepresentative sample of the true feature distribution. As illustrated in Figure 4.3, all SHAP values for *Club's Instagram Followers* in this subset are negative, not because high follower counts reduce naming rights value, but because no club with a large Instagram following appeared in the test set. The model never observed a high-follower case at evaluation, and its SHAP output therefore cannot be generalized to clubs that fall outside the observed test distribution. This limitation may be directly relevant when applying this model to SC Braga, whose digital profile may differ substantially from the clubs represented in the test partition.

The third pattern concerns the Football outside the USA subset, which produces the most extreme model divergence observed across the entire analysis. OLS returns an R^2 of just 0.042 and an MAE of \$5,971,658, while LASSO achieves an R^2 of 0.731 and an MAE of \$3,178,004 in the same configuration, a gap of nearly 69 percentage points in R^2. This divergence is the largest observed between any two models in any subset and indicates that the naming rights market outside the United States is highly sensitive to variable selection. The near-zero OLS performance suggests that the unconstrained model is capturing noise rather than signal, likely due to the presence of correlated predictors and influential outliers in a relatively small and geographically heterogeneous sample. LASSO's strong performance in this context confirms that coefficient shrinkage and variable elimination are not merely regularization conveniences but structural necessities when the data-generating process is noisy and the sample is limited.

These patterns support three conclusions that directly inform model selection for the SC Braga valuation. First, no single model dominates across all configurations, as the appropriate choice depends on sample size, data composition, and specific inferential objective. Second, the Random Forest is the most powerful model when sample conditions permit, but its reliability degrades in small samples and its interpretability is constrained by the distributional limitations of the SHAP framework in low-observation test sets. Third, regularized linear models (LASSO in particular) offer a robust and interpretable alternative that maintains competitive performance across a wider range of sample conditions, making them a valuable complement to machine learning approaches in any valuation exercise that requires transparency and generalizability.

5.2.2 Model Selection for the SC Braga Valuation

Based on the evidence presented in Section 5.2.1, rather than selecting a single model and subset combination for the SC Braga valuation, a deliberate strategic decision was made to apply three independent estimation pathways and triangulate their outputs. This approach acknowledges that estimates produced under different methodological assumptions and at different levels of geographic specificity carry distinct and complementary informational value. Convergence across pathways strengthens the robustness of the final valuation; divergence, where it occurs, is itself analytically informative and can be attributed to identifiable differences in model assumptions and sample scope.

The first pathway applies the Random Forest model to the Excluding Events subset ($n = 285$), functioning as a global market benchmark. This subset is the broadest structurally coherent configuration in the database, as it removes event-based agreements while retaining the full cross-sport, cross-continent diversity of permanent club naming rights deals, and the Random Forest achieves its peak performance here, with an R^2 of 0.848 and an MAE of \$2,306,447. The observations in this subset are, by nature, more heterogeneous than SC Braga's immediate competitive context: they span multiple sports, continents, and commercial environments that differ substantially from European professional football. This pathway therefore does not ask what SC Braga is worth relative to its peers, but rather where it sits within the full international naming rights market, a perspective that may be commercially relevant if potential sponsors operate across multiple sports and geographies, and which provides a ceiling-and-floor reference against which the more specific estimates can be benchmarked.

The second and third pathways both apply models to the European Football subset ($n = 57$), which constitutes the most geographically and competitively relevant pool of observations for SC Braga. This subset isolates professional football clubs operating in the same continental market, subject to similar competitive structures, broadcasting environments, and commercial norms. It is acknowledged that this subset carries two specific limitations: its sample size is small relative to the other configurations, and it includes hypothetical valuation cases drawn from market studies, most notably the Kroll (2024) report, alongside real naming rights agreements. These characteristics introduce uncertainty into the estimates derived from this subset. However, the out-of-sample performance metrics of both models applied here proved sufficiently strong to justify their inclusion, and the geographic and structural proximity of the observations to SC Braga's profile outweighs the costs of a reduced and partially hypothetical sample.

Within the European Football subset, the second pathway applies the Random Forest model, which achieves an R^2 of 0.775 and an MAE of \$2,478,564, the highest predictive performance of any model in this configuration. Its capacity to capture non-linear interactions between variables provides a complementary perspective to the linear estimate, though the methodological caveats identified in Section 5.2.1 apply: the small test partition of 12 observations limits the generalizability of SHAP-based attribution, and the model's output for SC Braga should be interpreted as a predictive estimate rather than a fully explainable valuation.

The third pathway applies the LASSO regression model to the European Football subset, returning an R^2 of 0.663 and an MAE of \$3,292,009. While this represents a lower predictive ceiling than the Random Forest, LASSO offers full coefficient transparency and immunity to the sample-size sensitivity that constrains ensemble methods. Its variable selection property ensures that the valuation is driven by a parsimonious and interpretable set of determinants, which is particularly valuable in a context where communicating the drivers of value to non-technical stakeholders may be as important as the point estimate itself.

Taken together, the three pathways provide a structured triangulation framework: one estimate anchored in the broadest available market context, and two estimates anchored in the most geographically relevant subset, produced by the strongest non-linear and linear models respectively. The SC Braga valuation derived

from each pathway, and the interpretation of any divergence between them, is developed in Section 6.

6 The SC Braga Case Study

This chapter applies the modeling framework developed in the preceding sections to a concrete valuation exercise: the estimation of the fair market value of naming rights for the Estádio Municipal de Braga, home stadium of Sporting Clube de Braga. While the analytical work of Section 4 and Section 5 was necessarily abstract, concerned with model construction, variable identification, and cross-subset performance, this chapter grounds those findings in a specific commercial context, translating the empirical results into an actionable valuation range for a club whose naming rights commercial landscape is examined in detail in the following section.

The chapter proceeds in three stages. It begins with a profile of SC Braga and the Estádio Municipal de Braga, contextualizing the club's competitive standing, commercial trajectory, and positioning relative to the European football observations that constitute its most directly comparable peer group. It then applies the three estimation pathways defined in Section 5.2.2: the Random Forest on the Excluding Events subset, the Random Forest on the European Football subset, and the LASSO on the European Football subset, and reports the predicted annual naming rights value generated by each. Finally, it interprets the three estimates in conjunction, identifying points of convergence and divergence, and derives a fair market value range that reflects both the central tendency and the uncertainty inherent in any model-based valuation.

6.1 SC Braga's Profile and Commercial Context

Sporting Clube de Braga was founded on 19 January 1921 (Sporting Clube de Braga, n.d. b) in the city of Braga, in the northwest of Portugal. With over a century of history, the club has grown from its regional roots into one of Portugal's most consistently competitive football clubs, maintaining uninterrupted presence in the top flight of Portuguese football since 1975, which makes it the fourth club with the longest consecutive run in the *Primeira Liga*, after FC Porto, SL Benfica, and Sporting CP (Sporting Clube de Braga, n.d. b). Beyond football, SC Braga operates across a wide range of sports disciplines, including athletics, basketball, futsal, swimming, and volleyball, though its commercial and competitive identity is anchored firmly in professional football.

On the pitch, the club's most significant period of growth came in the first quarter of the 21st century. The 2009/10 season marked a watershed moment: SC Braga finished second in the *Primeira Liga* (Zerozero, 2010), the highest league position in the club's history, which earned them qualification for the UEFA Champions League for the first time (UEFA, 2011a). In the 2010/11 season, the club competed in the Champions League group stage and, after finishing third in their group, dropped into the UEFA Europa League, ultimately reaching the final in Dublin, where they were defeated 1-0 by FC Porto (UEFA, 2011b). This run of results established the club's credentials as a genuine European competitor and significantly raised its commercial profile internationally. In subsequent years, SC Braga consolidated its position as

one of Portugal's top clubs, winning the *Taça de Portugal* in 2016 and 2021 and the *Taça da Liga* in 2013, 2020, and 2024 (Sporting Clube de Braga, n.d. c), while maintaining regular participation in UEFA competitions. In the 2024/25 season, the club finished fourth in the *Primeira Liga* (Liga Portugal, 2025) and competed in the UEFA Europa League league phase. This trajectory of competitive consistency is directly relevant to the naming rights valuation, as it underpins the club's historical competitive standing and its growing national and international media visibility.

The commercial trajectory of SC Braga accelerated markedly in October 2022, when Qatar Sports Investments (QSI), the sovereign wealth fund subsidiary that also owns Paris Saint-Germain, acquired a 21.67% stake in the club (Lusa, 2022), signaling a new phase of institutional investment and international commercial ambition. The club currently counts approximately 30,000 registered members and recorded a significant increase in season ticket renewals in 2023/24 (Sporting Clube de Braga – Futebol, SAD, 2024). On the digital front, the club launched SC Braga Next, a proprietary Over-the-Top (OTT) platform offering exclusive content, live streaming, and on-demand video (SC Braga NEXT, n.d.), and partnered with Crypto Blockchain Industries (CBI) to develop a dedicated space within the AlphaVerse platform: a Web3 environment featuring a three-dimensional recreation of the Estádio Municipal de Braga and the club's Cidade Desportiva, offering fans virtual experiences and digital collectibles (Sporting Clube de Braga, 2023).

The venue at the centre of this valuation exercise is the Estádio Municipal de Braga, commonly known as *A Pedreira* (The Quarry), a name derived from its extraordinary setting, carved directly into the granite hillside of Monte Castro in the city's northern sporting park. Designed by Portuguese architect Eduardo Souto de Moura, who was awarded the Pritzker Architecture Prize in 2011 in part for this work (Pritzker Architecture Prize, 2011), the stadium was inaugurated on 30 December 2003 (Sporting Clube de Braga, n.d. a) and served as a host venue for UEFA Euro 2004 (UEFA, 2004). Its distinctive structure features two massive parallel concrete stands connected by a cable roof, with one end open to the rocky hillside and the other to the city panorama below. The stadium has been widely recognized as one of the most architecturally significant sporting venues in the world. It has a seating capacity of 30,286 (Sporting Clube de Braga, n.d. a), and the specific operational metrics used as model inputs are presented in the following section.

The stadium's naming rights history is notable and directly relevant to the commercial context of this valuation. In July 2007, the club signed a three-year sponsorship deal with French insurance company AXA for approximately €4.5 million, making it the first stadium naming rights agreement in Portuguese football (Lusa, 2007). The deal was subsequently renewed for a further three years in 2010 (Lusa, 2007). Under this arrangement, the stadium was commercially referred to as Estádio AXA, and the name became widely used in media and among supporters.

The commercial designation of the stadium as Estádio AXA did not go uncontested within the municipal sphere. While the then-president of the Câmara Municipal de Braga, Mesquita Machado, raised no institutional objection, noting that the stadium had never ceased to be the Estádio Municipal and that he

saw no impediment to the club adopting a commercial designation under the existing occupancy agreement, the episode prompted criticism from the opposition on the municipal council (Record, 2007). Ricardo Rio, then leader of the opposition, argued that the municipality had not been consulted in the process and that, given the substantial financial burden borne by the council in maintaining the asset, the revenue from any commercial exploitation of the stadium's name should have accrued to the municipality rather than to the club (Record, 2007).

Following the expiry of the AXA agreement, the stadium reverted to its official designation, and no subsequent formal naming rights deal has been concluded. This episode illustrates a structural constraint that remains commercially relevant today: as the Estádio Municipal de Braga is the property of the Câmara Municipal de Braga rather than of the club, any naming rights agreement concluded by SC Braga without the formal involvement of the municipality as asset owner rests on the tolerance of the landlord rather than on an enforceable contractual right. Should the club seek to enter into a new naming rights agreement, this ownership constraint would need to be either navigated through a formal partnership with the municipality or resolved through a prior transfer of ownership.

In this regard, developments in 2023 are commercially significant. Ricardo Rio, by then president of the Câmara Municipal de Braga, confirmed that the municipality was in active discussions with the SC Braga SAD and with Qatar Sports Investments regarding a potential sale of the stadium. Rio stated that a formal valuation process would be initiated and that, while the stadium had cost approximately €200 million to build, the municipality did not intend to demand that amount (Silva, 2023). Should this transfer of ownership be completed, it would remove the principal structural constraint that limited the scope of the original AXA agreement and would substantially strengthen the club's position in any future naming rights negotiation, transitioning SC Braga from tenant to owner of the asset with full authority to conclude a formal and binding agreement.

Table 6.1 presents the input variables used across the three estimation pathways defined in Section 5.2.2, populated with the values recorded for SC Braga at the time of this valuation. For variables where no value is available for SC Braga (namely those that presuppose the existence of an active naming rights agreement) the dynamic imputation protocol described in Section 3.5.5 was applied, substituting the median value of the corresponding subset. This ensures full consistency with the treatment of missing data during model training and prevents the artificial penalization of the club for the absence of an agreement that the valuation exercise itself is designed to estimate.

Table 6.1: SC Braga – Input Variables for the Valuation Models

Variable	Value	EUR Equivalent
Structural		
Inflation Adjusted Construction Cost (2025, USD)	$246,861,749	€209,524,486
Stadium Capacity	30,286	–
Venue Age (2025)	22 years	–
Multi-sport Facility	No	–
Multi-use Facility (Non-sport Events)	Yes	–
NR Deal Coincided with Venue Construction/Renovation	n/a[a]	–
Market		
City Population	193,324	–
Inflation Adjusted County GDP per Capita (2025, USD)	$25,552	€21,687
Continent	Europe	–
Operational		
Average Attendance (2024/25)	12,532	–
Average Occupancy Rate (2024/25)	41.4%	–
Games in National Television (2024/25)	28	–
Competitive		
Club's All-Time Winning Percentage in Top Tier League	44.9%	–
League Position (2024/25)	4th	–
Sport	Soccer	–
Competition Level	Major League	–
Gender	Male	–
Digital		
Club's Instagram Followers (February 2026)	411,199	–
Contractual		
Sponsorship Category	n/a[a]	–
Contract Length	n/a[a]	–
Past Naming Deals Count	1	–

Note: All monetary values are inflation-adjusted to 2025 USD. EUR equivalents are calculated using the ECB reference rate of April 2026 (1 EUR = 1.1782 USD), which is applied consistently throughout all USD-to-EUR conversions in this chapter.

[a] Variables marked n/a are not applicable to SC Braga as they presuppose an active naming rights agreement. These values are imputed from the subset median during model execution, consistent with the dynamic imputation protocol described in Section 3.5.5.

Several features of SC Braga's input profile are worth contextualizing before the model estimates are presented. The club's Instagram following of 411,199 sits at approximately 19% of the European Football subset median of 2,210,838. Given the variable's role as the single most consistent driver of naming rights value across all analytical segments, this considerable gap is expected to exert meaningful downward pressure on the model estimates. The inflation-adjusted construction cost of $246,861,749 (€209,524,486 at the ECB reference rate of April 2026 of 1 EUR = 1.1782 USD, applied consistently throughout all subsequent USD-to-EUR conversions in this chapter) is, by contrast, above the subset median of $153,895,959 (€130,619,554), reflecting the architectural ambition and technical complexity of the Estádio Municipal de Braga and providing an upward counterweight. Average attendance of 12,532 and a city population of 193,324 both place the club well below the European sample medians of 41,200 and 568,996 respectively, consistent with the club's position in a smaller domestic market. The occupancy rate of 41.4% also falls considerably short of the subset median of 94.9%, though this figure must be interpreted in the context of the *Primeira Liga* as a whole, where the average stadium occupancy rate in the 2024/25 season stood at 54% (Liga Portugal, 2025), suggesting that low utilization rates are a structural feature of the Portuguese football market rather than an indicator specific to SC Braga. Taken together, these inputs suggest that the SC Braga valuation is likely to fall in the lower-to-mid range of the European Football subset, with the construction cost providing upward pressure and the digital reach and market size variables acting as moderating factors. The three estimation pathways defined in Section 5.2.2 are applied in the following section, and their outputs are interpreted in light of this profile.

6.2 Valuation Results

This section applies the three estimation pathways defined in Section 5.2.2 to the SC Braga input profile presented in Table 6.1 and reports the predicted annual naming rights value generated by each. The three estimates are presented sequentially, beginning with the global market benchmark and progressing to the geographically specific European Football estimates.

The first estimation pathway applies the Random Forest model to the Excluding Events subset, which encompasses 285 permanent club naming rights agreements spanning multiple sports, continents, and commercial environments. This configuration functions as a global market benchmark, positioning SC Braga within the full international naming rights market rather than within its immediate European football peer group. Applied to the SC Braga input profile, this model generates a predicted annual naming rights value of $1,999,365 (€1,696,795).

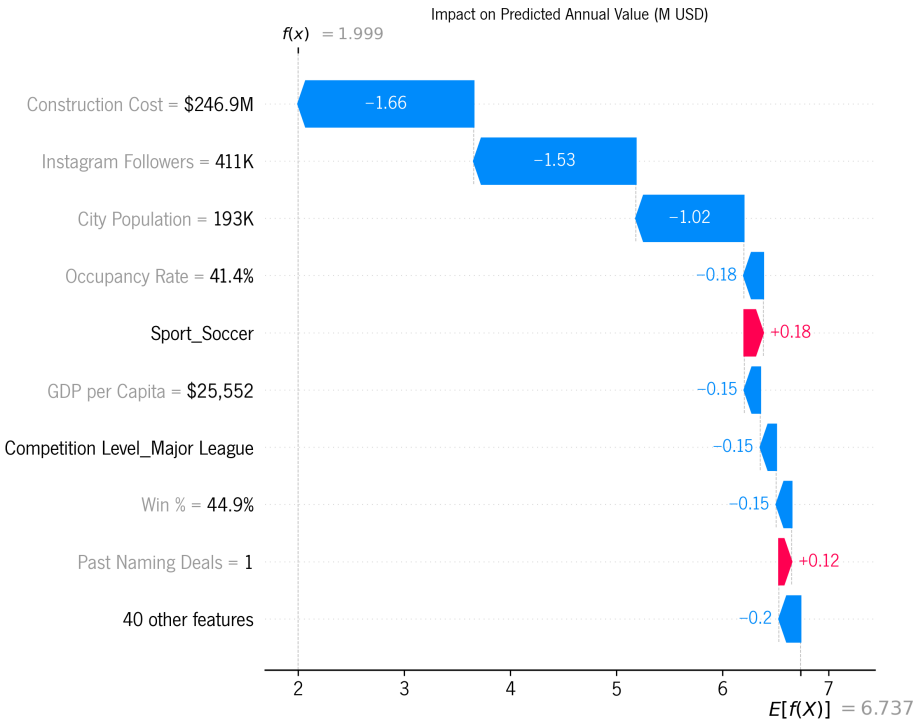


Figure 6.1: SHAP Waterfall Plot — SC Braga, Random Forest (Excluding Events, $n = 285$)

The decomposition of this estimate is illustrated in Figure 6.1, which presents the SHAP waterfall plot for SC Braga under this pathway. Starting from the model's baseline expected value of $6.737M / €5.72M (the average predicted annual value across the Excluding Events training set), the waterfall shows how each input variable adjusted the prediction upward or downward to arrive at the final estimate. The three largest downward contributions come from Construction Cost (−$1.66M / −€1.41M), Instagram Followers (−$1.53M / −€1.30M), and City Population (−$1.02M / −€0.87M). In the context of the Excluding Events subset, which is dominated by North American agreements where construction costs are associated with large-market, high-revenue venues, the Estádio Municipal de Braga's construction cost of €209.5M places it in a mid-range position that the model associates with higher expected values than SC Braga's digital and market profile can support, resulting in a net negative contribution. Smaller negative contributions from Occupancy Rate, GDP per Capita, Win Percentage, and Competition Level further compress the estimate, while minor positive adjustments from Sport classification and Past Naming Deals partially offset these effects.

Within this broader context, the result should be interpreted as a floor-level reference anchored in global market conditions: what a sponsor with cross-sport, cross-geography exposure would be expected to pay for the SC Braga naming rights based on the full breadth of comparable agreements in the database.

The second estimation pathway applies the Random Forest model to the European Football subset, the most geographically and competitively relevant configuration for SC Braga. Applied to the SC Braga input profile, this model generates a predicted annual naming rights value of $3,799,537 (€3,224,695).

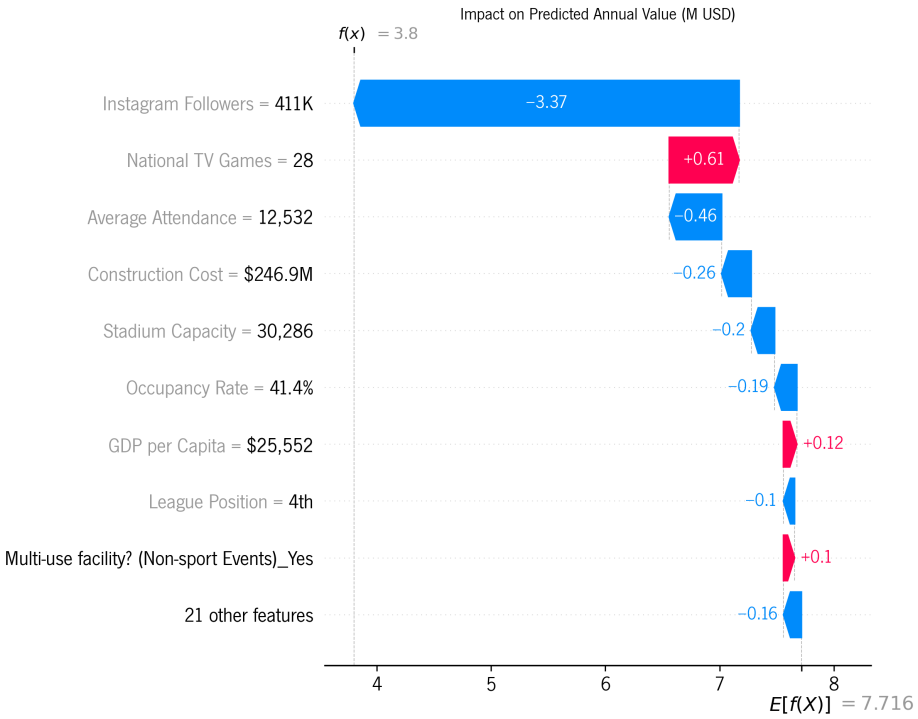


Figure 6.2: SHAP Waterfall Plot — SC Braga, Random Forest (European Football, $n = 57$)

The decomposition of this estimate is illustrated in Figure 6.2. Starting from the model's baseline expected value of $7.716M / €6.55M (the average predicted annual value across the European Football training set), the waterfall reveals a valuation structure that differs meaningfully from the global benchmark. Instagram Followers accounts for the single largest adjustment, contributing −$3.37M / −€2.86M, a figure that reflects the considerable gap between SC Braga's digital footprint of 411,199 followers and the European clubs that anchor the upper end of this subset's value distribution. National TV Games provides the most significant positive contribution ($0.61M / €0.52M), reflecting SC Braga's 28 televised matches and positioning the club favourably on this dimension relative to the subset median. Average Attendance, Construction Cost, Stadium Capacity, and Occupancy Rate each contribute modest negative adjustments, while GDP per Capita and Multi-use facility status provide small positive offsets.

It should be noted that this estimate carries a higher degree of uncertainty than the global benchmark, for reasons discussed in Section 5.2.1. A sensitivity analysis conducted across ten alternative train-test splits produced values ranging from $3,227,410 to $4,863,578 (€2,739,016 to €4,128,067), with a median of $3,943,855 (€3,347,186), indicating that while the central estimate is plausible, the small test partition of 12 observations introduces meaningful variability into the result.

The third estimation pathway applies the LASSO regression model to the European Football subset, generating a predicted annual naming rights value of $2,097,060 (€1,779,723). As the most interpretable of the three models, LASSO allows the valuation to be expressed as an explicit linear equation in terms of the original input variables, presented in Eq. (6.1) and the accompanying symbol legend.

$$\hat{y} = -1{,}177{,}938 + 0.122\,x_I + 96.26\,x_A + 0.272\,x_P + 0.00177\,x_C + 106{,}626\,x_L + 52{,}396\,x_T - 812{,}596\,x_W + \delta \tag{6.1}$$

Symbol	Variable
x_I	Club's Instagram Followers (February 2026)
x_A	Average Attendance (2024/25)
x_P	City Population
x_C	Inflation Adjusted Construction Cost (2025, USD)
x_L	Club's League Position (2024/25)
x_T	Number of Games in National Television (2024/25)
x_W	Club's All-Time Winning Percentage in Top Tier League
δ	Sponsorship Category dummy (imputed from subset mode for SC Braga)

The equation reflects the variable selection property of LASSO: of the full set of predictors available in the European Football subset, the model retained only seven continuous variables and a sponsorship category term, eliminating all remaining predictors by shrinking their coefficients to zero. The dominant positive driver is *Club's Instagram Followers*, with each additional follower associated with an increase of \$0.122 (€0.104) in annual naming rights value. *Average Attendance* contributes \$96.26 (€81.70) per additional spectator, and *Number of Games in National Television* adds \$52,396 (€44,471) per additional televised match. *Construction Cost* and *City Population* exert smaller positive effects, while *Club's All-Time Winning Percentage* carries a negative coefficient, a result that likely reflects multicollinearity with other competitive performance variables within the small European Football sample rather than a genuine negative commercial relationship. *League Position* carries a positive coefficient of \$106,626 (€90,499), which in this context means that a numerically higher position (corresponding to a lower finish in the table) is associated with a higher predicted value, a counter-intuitive result that similarly warrants caution in interpretation given the limited sample size.

Regarding the sponsorship category term δ, the LASSO retained coefficients for only four of the categories present in the subset: Financial Services (\$835,647 / €709,257), Industrials (\$1,987,300 / €1,686,556), Retail (−\$1,871,013 / −€1,588,197), and Waste Management (\$560,611 / €475,820). All remaining categories were eliminated with a zero coefficient, including Insurance, which is the most frequent in the European Football subset with nine agreements. As SC Braga does not have an active naming rights agreement, the category variable is imputed using the subset mode, which, precisely, is Insurance. Since the LASSO assigned a zero coefficient to this category, δ evaluates to zero in the SC Braga valuation, meaning the predicted value of €1,779,723 is derived entirely from the continuous variable contributions. It should be noted, however, that these category coefficients are estimated from very few observations (Industrials

from three agreements and Retail from one, for example) and should therefore not be interpreted as reliable quantitative estimates of the commercial premium associated with each sector. They are reported here for transparency but carry limited inferential value given the sample constraints of the European Football subset.

6.3 Interpretation and Fair Market Value Estimate

The three estimation pathways developed in Section 5.2.2 and applied in Section 6.2 produce a set of results that, taken together, allow for a structured and methodologically honest derivation of a fair market value range for the naming rights of the Estádio Municipal de Braga. The first pathway, the Random Forest model applied to the Excluding Events subset, produces a predicted annual value of €1,696,795. The second pathway, the Random Forest model applied to the European Football subset, produces a predicted annual value of €3,224,695. The third pathway, the LASSO regression model applied to the European Football subset, produces a predicted annual value of €1,779,723. Table 6.2 summarises these three outputs alongside the main significant variables driving each model.

Table 6.2: Summary of SC Braga Valuation Outputs Across the Three Estimation Pathways

Model (Subset)	**Main Significant Variables**	**Predicted Annual Value**
Random Forest (Excluding Events)	Construction Cost Instagram Followers City Population	€1,696,795
Random Forest (European Football)	Instagram Followers National TV Games Average Attendance	€3,224,695
LASSO Regression (European Football)	Instagram Followers Average Attendance National TV Games	€1,779,723
Fair Market Value (central estimate)		**€1.70M – €1.78M**
Upper bound		**€3.23M**

Note: The "Main Significant Variables" column lists the top three predictors by absolute contribution within each model. For the Random Forest models, these correspond to the variables with the largest mean absolute SHAP values in the SC Braga prediction (Figure 6.1 and Figure 6.2). For the LASSO model, because the variables entering Eq. (6.1) are measured on different scales, the ranking is based on the standardized coefficients of the same specification rather than read directly from the equation's unstandardized form. All values are expressed in EUR using the ECB reference rate of April 2026 (1 EUR = 1.1782 USD).

The most analytically significant finding is the strong convergence between the first and third pathways, which are methodologically distinct and applied at fundamentally different levels of geographic specificity. The gap between these two estimates is €82,928, representing a divergence of less than 5%. This degree of convergence between a non-linear ensemble method trained on a broad multi-sport global sample and a regularized linear model trained on a geographically focused European football sample constitutes strong evidence that the €1.70M–€1.78M range reflects a genuine structural signal rather than a modeling artefact. This range compares favourably with the €1.5M per annum implied by the club's own historical precedent, the AXA agreement of 2007, which was valued at approximately €4.5M over three years (Lusa, 2007).

The second pathway diverges substantially from this convergence, producing an estimate of €3,224,695. The methodological limitations of this model in the European Football context were discussed in Section 5.2.1, where the sensitivity of ensemble methods to sample size was identified as a key constraint. As mentioned in Section 6.2, a sensitivity analysis conducted across ten alternative train-test splits confirms this concern empirically: the predicted value for SC Braga ranged from €2,739,016 to €4,128,067, with a median of €3,347,186 and a standard deviation of €384,489, a level of variability that reflects the instability inherent in applying Random Forest to a subset of 57 observations with a test partition of only 12 cases. Additionally, as illustrated by the SHAP summary plot presented in Figure 4.3, the small test partition produces directionally misleading attributions for key variables, most notably *Club's Instagram Followers*, whose values in the test set are uniformly low, causing the model to attribute a negative impact to a variable that is structurally positive across the broader dataset. These limitations do not disqualify the estimate, as the model achieves an R^2 of 0.775 on its test set and captures non-linear interactions that the linear LASSO cannot accommodate, but they do preclude treating it as a point estimate with the same degree of confidence as the two converging pathways.

Within the European Football subset, the central estimate of €1.70M–€1.78M places SC Braga in the lower quartile of the distribution, with only 13 of the 57 observations recording an annual value below €1.78M. The subset median stands at €4,680,129 and the mean at €6,412,875, both substantially above the central estimate. This positioning is consistent with the input profile analysis presented in Section 6.1: SC Braga's Instagram following and city population sit well below the subset medians, and its occupancy rate is among the lowest in the sample. The gap between the club's current profile and the subset median should not be interpreted as a ceiling on future value, as it reflects the club's current commercial standing and stadium utilization, both of which are subject to change, particularly in the event of a transfer of stadium ownership as discussed in Section 6.1.

Taken together, the three pathways support the following fair market value assessment. The central estimate for the annual naming rights value of the Estádio Municipal de Braga is in the range of €1.70M to €1.78M, anchored by the convergence of the first and third pathways. An upper bound of €3.23M, produced by the second pathway, represents a plausible but methodologically constrained estimate of what the market might support under more optimistic assumptions about SC Braga's positioning within the

European football naming rights landscape. The lower estimate is the more defensible of the two as a basis for commercial negotiation, while the upper bound provides a ceiling reference that could become more relevant as the club's digital reach and stadium utilization approach the median levels of the European Football subset.

7 Conclusion

This study set out to address two interrelated research questions: what are the primary determinants of naming rights value in professional sport, and what is the fair market value of naming rights for the Estádio Municipal de Braga. The empirical framework developed across Section 3 to Section 6 provides structured and defensible answers to both.

7.1 Research Questions and Main Findings

Regarding the first research question, the meta-analytical framework developed in Section 5 identifies a structural core of three variables that achieve near-universal model consensus across all analytical segments: *Club's Instagram Followers*, *Inflation Adjusted Construction Cost*, and *City Population*. Among these, *Club's Instagram Followers* is the only variable to achieve 100% consensus across all three macro-segments, providing empirical confirmation that digital audience reach has, in football-specific contexts, displaced the physical asset as the dominant commercial differentiator between valuation tiers, a finding the prior literature anticipated but could not test. Beyond this structural core, variable relevance shifts meaningfully as the sample is refined: *Average Attendance* rises consistently as the analysis narrows to football-specific contexts, while *Contract Length* and *Past Naming Deals Count* lose discriminatory power, a pattern consistent with the relative immaturity of the European naming rights market.

Regarding the second research question, the application of three independent estimation pathways to the SC Braga input profile produces a coherent triangulated outcome. The Random Forest model applied to the Excluding Events global benchmark and the LASSO regression model applied to the European Football subset converge to within 5% of each other, at €1,696,795 and €1,779,723 respectively, providing strong evidence that the €1.70M–€1.78M range represents the fair market value of annual naming rights for the Estádio Municipal de Braga under current conditions. The Random Forest model applied to the European Football subset produces a higher estimate of €3,224,695, which constitutes a plausible upper bound subject to the methodological constraints identified in Section 5.2.1. The central estimate is broadly consistent with the club's own historical precedent, as the AXA agreement of 2007 implied approximately €1.5M per annum (Lusa, 2007).

7.2 Practical Implications

While the determinants identified in response to the first research question apply broadly to the naming rights industry, their most direct practical translation concerns the case study at the centre of this thesis. The practical implications discussed in this section therefore focus primarily on SC Braga, with brief reference to the broader industry where the findings are structurally transferable.

At the industry level, the consistent dominance of digital audience reach and the shifting relevance of North-American-anchored variables (such as *Contract Length* and *Past Naming Deals Count*) in football-specific contexts suggest that valuation frameworks developed for the American market should not be applied to European football without structural recalibration. Methodologically, the convergence of the two principal estimation pathways for SC Braga to within 5%, despite their differing assumptions and sample scopes, demonstrates that multi-model triangulation is a viable strategy for producing defensible valuations in markets where transparency and sample size are constrained.

For SC Braga specifically, the findings point to a clear ordering of commercial priorities. Growth in digital audience reach represents the highest-leverage investment the club can make in the medium term, given the substantial gap between its current Instagram following and the European Football subset median and the variable's dominant role in the valuation of clubs in this segment. Stadium utilization constitutes a secondary lever, though its potential must be interpreted against the fact that the Primeira Liga's average stadium occupancy in 2024/25 stood at 54% (Liga Portugal, 2025), indicating that low utilization is in large measure a structural feature of the Portuguese market rather than a metric the club alone can reverse. The most consequential factor is structural rather than commercial: the ownership of the Estádio Municipal de Braga by the Câmara Municipal de Braga remains the principal constraint on the commercial scope of any future agreement, as discussed in Section 6.1, and the municipality's 2023 announcement of a potential stadium sale (Silva, 2023) is therefore commercially significant, since a transfer of ownership would substantially strengthen the club's negotiating position.

7.3 Limitations and Directions for Future Research

Several limitations of this study merit acknowledgement. As discussed in Section 3.4, the sample is structurally weighted toward the North American market, which accounts for 221 of the 306 entries, and the European Football subset ($n = 57$) contains a meaningful share of hypothetical valuations drawn from market studies, most notably the Kroll (2024) report, rather than concluded agreements. While this expansion was a necessary trade-off to secure a critical mass of data for econometric modeling, it means that the European estimates are partially anchored in third-party methodologies rather than in purely transactional signal. The sample is also subject to survivorship bias, as it focuses on active or widely reported agreements and therefore excludes failed negotiations and venues that did not secure a deal, which may lead the models to marginally overestimate fair market value. The operationalization of the digital dimension through *Club's Instagram Followers* as of February 2026 constitutes a further limitation: the variable captures a single point in time and is volume-based, and therefore neither reflects the digital landscape at the moment historical agreements were brokered nor distinguishes between audience size and audience engagement, a distinction of increasing commercial relevance. Model-specific caveats identified in Section 5 and Section 6 should also be kept in mind: the small test partition of the European Football subset constrains the reliability of SHAP-based attribution for the Random Forest pathway, and several

of the categorical coefficients produced by the LASSO model in this subset are estimated from very few observations and should therefore not be read as reliable quantitative estimates.

Several directions for future research follow from these limitations. First, as additional European agreements are concluded and disclosed, the empirical foundation developed here should be revisited on a sample less dependent on hypothetical valuations, and the estimates produced in this study should be validated against the prices realized in subsequent transactions. Second, once historical follower counts become systematically recoverable through archival sources, the dynamic evolution of digital audiences over the duration of a contract could be incorporated, strengthening the temporal consistency of the digital dimension. Third, engagement-based digital metrics, such as interactions per post or audience quality indices, should be tested alongside volume-based ones to refine the operationalization of the commercial audience delivered to sponsors. Finally, future models would benefit from integrating brand equity metrics that go beyond the variables available in the current dataset, such as UEFA coefficient rankings, which capture a club's historical European performance, or global brand valuations produced by specialist firms, as these dimensions of intangible value may account for pricing differentials that structural and operational variables alone cannot explain.

Despite these constraints, this research establishes a foundational empirical framework that successfully bridges the gap between historical naming rights data and advanced predictive modeling, offering a robust starting point for future valuations in the European football landscape.

References

Bauer, H. H., Sauer, N. E., & Schmitt, P. (2005). Customer‐based brand equity in the team sport industry: Operationalization and impact on the economic success of sport teams. *European Journal of Marketing*, *39*(5), 496–513. https://doi.org/10.1108/03090560510590683

Becker-Olsen, K. (2003). Questioning the name game: An event study analysis of stadium naming rights sponsorship announcements. *International Journal of Sports Marketing and Sponsorship*, *5*(3), 9–20. https://doi.org/10.1108/IJSMS-05-03-2003-B002

Boyd, J. (2000). Selling home: Corporate stadium names and the destruction of commemoration [_eprint: https://doi.org/10.1080/00909880009365580]. *Journal of Applied Communication Research*, *28*(4), 330–346. https://doi.org/10.1080/00909880009365580

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Centracchio, B., Popp, N., & Jensen, J. A. (2023). Examining valuation of corporate naming rights partnerships in collegiate sports and their impact on consumer behavior. *International Journal of Sports Marketing and Sponsorship*, *24*(5), 871–890. https://doi.org/10.1108/IJSMS-02-2023-0015

Chen, K., & Zhang, J. (2012). To name it or not name it: Consumer perspectives on facility naming rights sponsorship in collegiate athletics. *Journal of Issues in Intercollegiate Athletics*, *5*(1). https://scholarcommons.sc.edu/jiia/vol5/iss1/7

Clark, J., Cornwell, T., & Pruitt, S. (2002). Corporate stadium sponsorships, signalling theory, agency conflicts and shareholder wealth. *Journal of Advertising Research*, *42*. https://doi.org/10.2501/JAR.42.6.16

Cornwell, T., Weeks, C., & Roy, D. (2005). Sponsorship-linked marketing: Opening the black box. *J Advert*, *34*. https://doi.org/10.1080/00913367.2005.10639194

Crompton, J., & Howard, D. (2003). The american experience with facility naming rights: Opportunities for english professional football teams [_eprint: https://doi.org/1360671032000148585]. *Managing Leisure*, *8*(4), 212–226. https://doi.org/10.1080/1360671032000148585

DeSchriver, T. D., & Jensen, P. (2003). What's in a name? price variation in sport facility naming rights. *Eastern Economic Journal*. Retrieved December 18, 2025, from https://www.semanticscholar.org/paper/What%27s-in-a-Name-Price-Variation-in-Sport-Facility-DeSchriver-Jensen/be67158230640fd0628d8b6f32889cb85fd5a570

Gerrard, B., Parent, M. M., & Slack, T. (2007). What drives the value of stadium naming rights? a hedonic-pricing approach to the valuation of sporting intangible assets. *International Journal of Sport Finance*, *2*(1), 10–24. https://doi.org/10.1177/155862350700200102

Gratton, C., & Solberg, H. A. (2007). *The economics of sport broadcasting*. Routledge.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in python*. Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33. https://doi.org/10.2307/1913643

Kroll. (2024). *Why european football clubs should look to stadium naming rights to diversify income: An analysis of potential stadium naming rights valuations* (KR24060962). Kroll. https://www.kroll.com/en/reports/valuation/european-stadium-naming-rights-report-2024

Leeds, E. M., Leeds, M. A., & Pistolet, I. (2007). A stadium by any other name: The value of naming rights. *Journal of Sports Economics*, *8*(6), 581–595. https://doi.org/10.1177/1527002506296546

Liga Portugal. (2025). *Liga portugal betclic* [Liga portugal]. Retrieved April 14, 2026, from https://www.ligaportugal.pt/competition/618/liga-portugal-betclic/round/20242025

Lusa. (2007, July 9). *Sp. braga: Cedência do nome do estádio rende 4,5 milhões de euros* [Maisfutebol]. Retrieved April 14, 2026, from https://maisfutebol.iol.pt/sp-braga-cedencia-do-nome-do-estadio-rende-4-5-milhoes-de-euros

Lusa. (2022, December 13). *Qatar sports investments consuma entrada na SAD do sp. braga* [PÚBLICO]. Retrieved April 14, 2026, from https://www.publico.pt/2022/12/13/desporto/noticia/qatar-sports-investments-consuma-entrada-sad-sp-braga-2031338

Popp, N., DeSchriver, T., McEvoy, C., & Diehl, M. A. (2016). A valuation analysis of corporate naming rights for collegiate sport venues. *Sport Marketing Quarterly*, *25*(1), 7–20. https://doi.org/10.1177/106169341602500103

Pritzker Architecture Prize. (2011). *Eduardo souto de moura | the pritzker architecture prize* [Pritzker architecture prize]. Retrieved April 14, 2026, from https://www.pritzkerprize.com/laureates/2011

Record. (2007, July 13). *Oposição contesta negócio com a axa* [Record]. Retrieved April 14, 2026, from https://www.record.pt/futebol/futebol-nacional/liga-betclic/sp–braga/detalhe/oposicao-contesta-negocio-com-a-axa

Reysen, S., Snider, J., & Branscombe, N. (2012). Corporate renaming of stadiums, team identification, and threat to distinctiveness. *Journal of Sport Management*, *26*, 350–357. https://doi.org/10.1123/jsm.26.4.350

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. Retrieved December 9, 2025, from https://www.jstor.org/stable/1830899

Rowbottom, N. (2002). THE APPLICATION OF INTANGIBLE ASSET ACCOUNTING AND DISCRETIONARY POLICY CHOICES IN THE UK FOOTBALL INDUSTRY. *The British Accounting Review*, *34*(4), 335–355. https://doi.org/10.1006/bare.2002.0215

Santomier, J. (2024, January 1). Digital transformation: The global sport industry. https://doi.org/10.1016/B978-0-443-13701-3.00209-7

SC Braga NEXT. (n.d.). *SCB next* [SCB next]. Retrieved April 14, 2026, from https://next.scbraga.pt/app/

Silva, L. F. (2023, September 27). *Câmara em conversações para vender estádio ao SC braga* [Diário do minho] [Section: Notícias]. Retrieved April 14, 2026, from https://www.diariodominho.pt/noticias/desporto/2023-09-27-camara-em-conversacoes-para-vender-estadio-ao-sc-braga-651492133bf67

Sporting Clube de Braga. (n.d. a). *Estádio* [Sporting clube de braga]. Retrieved April 14, 2026, from https://scbraga.pt/estadio/

Sporting Clube de Braga. (n.d. b). *História* [Sporting clube de braga]. Retrieved April 14, 2026, from https://scbraga.pt/historia/

Sporting Clube de Braga. (n.d. c). *Palmarés* [Sporting clube de braga]. Retrieved April 14, 2026, from https://scbraga.pt/palmares/

Sporting Clube de Braga. (2023, December 7). *SC braga associa-se à CBI e entra no AlphaVerse* [Sporting clube de braga]. Retrieved April 14, 2026, from https://scbraga.pt/sc-braga-associa-se-a-cbi-e-entra-no-alphaverse/

Sporting Clube de Braga – Futebol, SAD. (2024). *Relatório e contas 2023/24*. Retrieved April 14, 2026, from https://scbraga.pt/wp-content/uploads/2024/10/RC-2023-24-SCB-1.pdf

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

UEFA. (2004). *Época 2004 | UEFA EURO 2004* [UEFA.com]. Retrieved April 14, 2026, from https://pt.uefa.com/uefaeuro/history/seasons/2004/

UEFA. (2011a). *Época 2010/11 | UEFA champions league 2010/11* [UEFA.com]. Retrieved April 14, 2026, from https://pt.uefa.com/uefachampionsleague/history/seasons/2011/

UEFA. (2011b). *Época 2010/11 | UEFA europa league 2010/11* [UEFA.com]. Retrieved April 14, 2026, from https://pt.uefa.com/uefaeuropaleague/history/seasons/2011/

Walraven, M., Bijmolt, T. H. A., & Koning, R. H. (2014). Dynamic effects of sponsoring: How sponsorship awareness develops over time [_eprint: https://doi.org/10.1080/00913367.2013.835754]. *Journal of Advertising*, *43*(2), 142–154. https://doi.org/10.1080/00913367.2013.835754

Watanabe, N., Yan, G., & Soebbing, B. (2016). Consumer interest in major league baseball: An analytical modeling of twitter. *Journal of Sport Management*, *30*. https://doi.org/10.1123/jsm.2015-0121

Woisetschläger, D., Haselhoff, V., & Backhaus, C. (2014). Fans' resistance to naming right sponsorships – why stadium names remain the same for fans. *European Journal of Marketing*, *48*. https://doi.org/10.1108/EJM-03-2012-0140

Zerozero. (2010). *Liga sagres 2009/10 - jogos, classificações e estatísticas - portugal* [Www.zerozero.pt]. Retrieved April 14, 2026, from https://www.zerozero.pt/edicao/liga-sagres-2009-10/8838

Appendix A SHAP Summary Plots: Stability and Market Dimension Analysis

The following figures present the SHAP summary plots for the three remaining structural subsets not displayed in the main analysis. These plots confirm the structural convergence described in the main text, with *Inflation Adjusted Construction Cost (2025) (USD)* and *Club's Instagram Followers (February 2026)* consistently emerging as the dominant model-based contribution measures across all subsets.

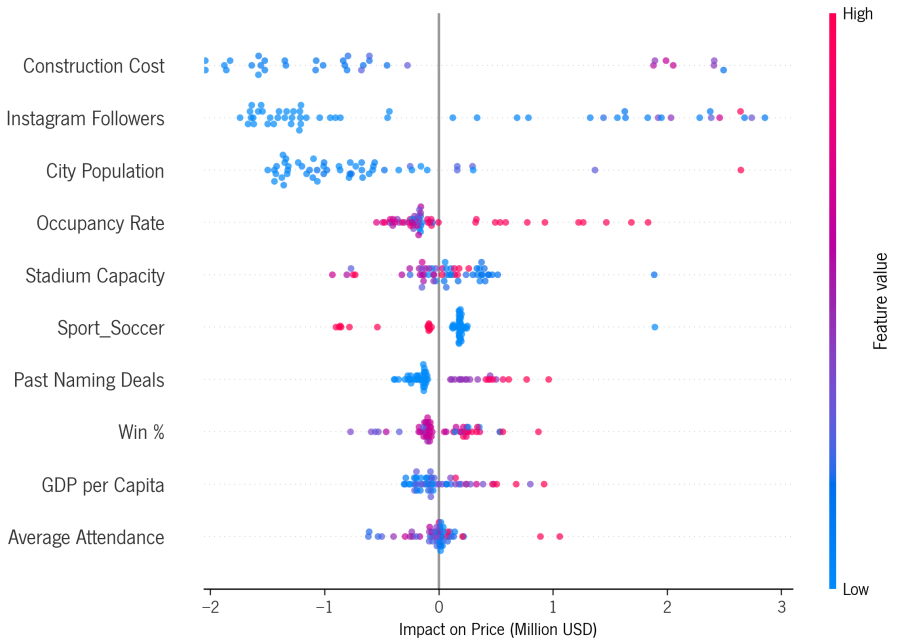


Figure A.1: SHAP Summary Plot — Random Forest (Excluding Events, $n = 285$)

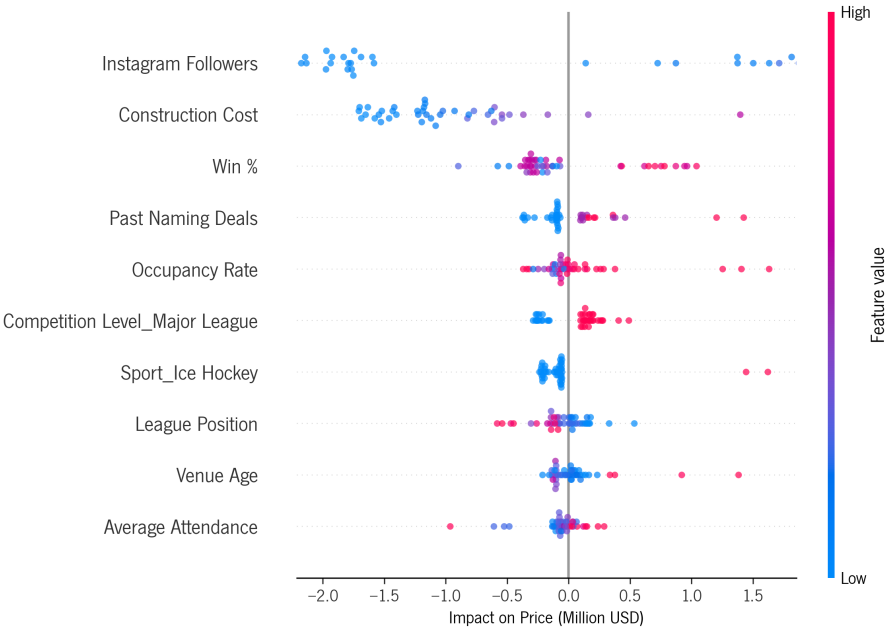


Figure A.2: SHAP Summary Plot — Random Forest (Unique Agreements, $n = 219$)

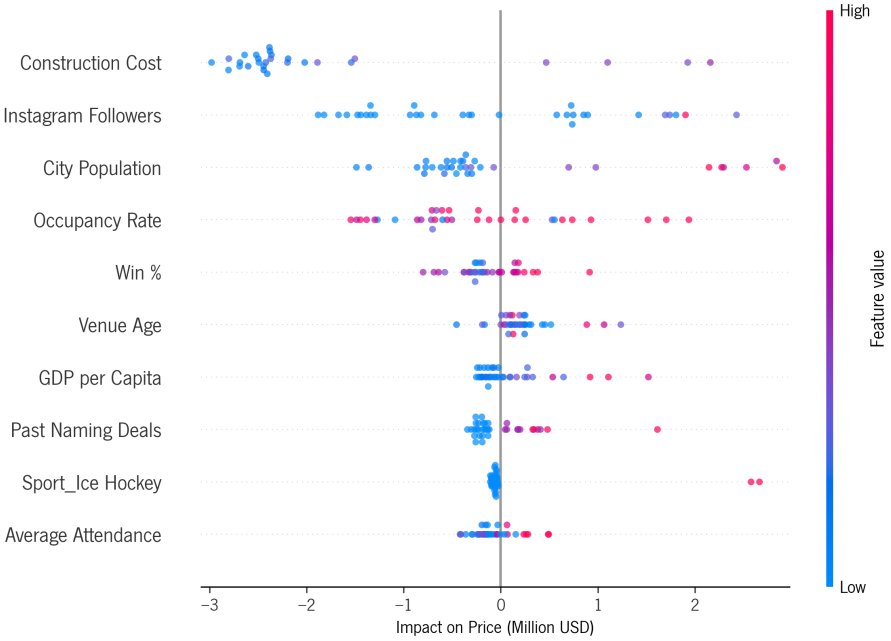


Figure A.3: SHAP Summary Plot — Random Forest (Major Leagues, $n = 174$)

Appendix B SHAP Summary Plots: Football-Specific Market Analysis

The following figures present the SHAP summary plots for the Football (Soccer) Only and Soccer outside the USA subsets not displayed in the main analysis. These plots confirm the structural convergence described in the main text.

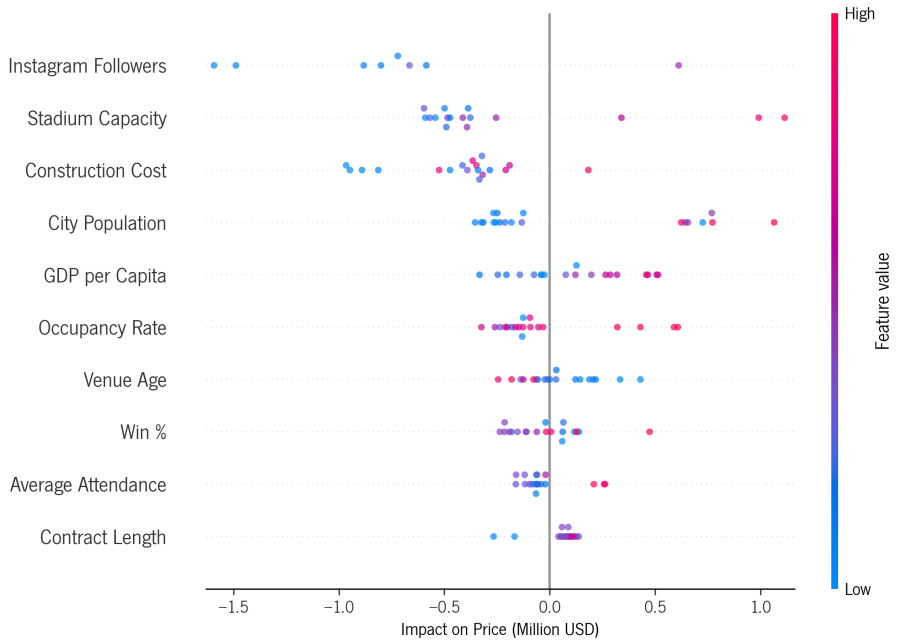


Figure B.1: SHAP Summary Plot — Random Forest (Football (Soccer) Only, $n = 95$)

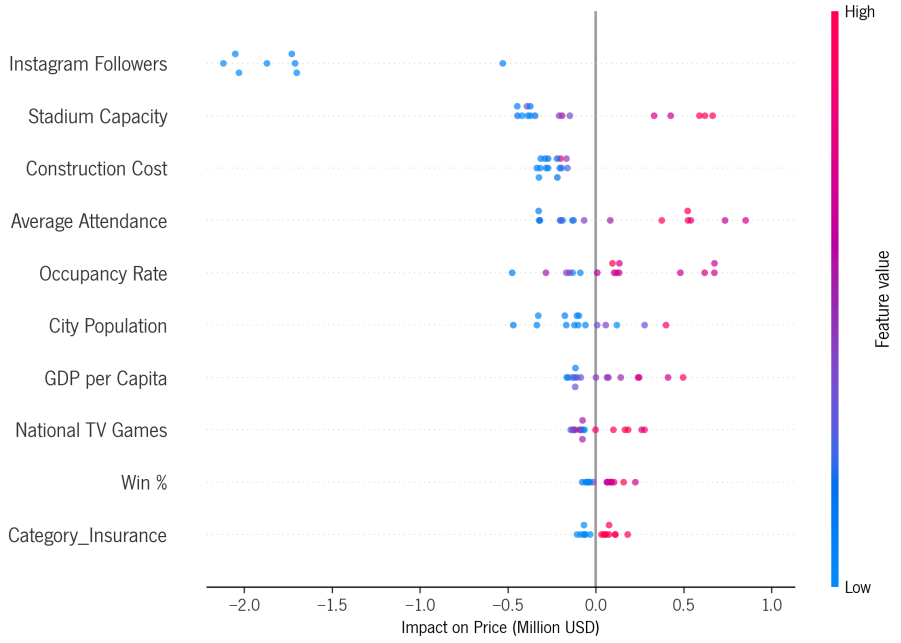


Figure B.2: SHAP Summary Plot — Random Forest (Soccer outside the USA, $n = 76$)