Football Data Pioneers: Report

Abstract

The football transfer market plays a pivotal and dynamic role in the sport, acting as the arena where clubs strategically reshape their squads to achieve success both domestically and internationally. With designated transfer windows, particularly during the summer and winter periods, the market sparks a flurry of activity as clubs engage in intricate negotiations to secure new talent or offload existing players.

At its core, the transfer market revolves around player transactions, involving financial dealings that dictate the movement of athletes between clubs. While high-profile transfers often grab headlines, the market also sees numerous loan deals and free transfers, enabling clubs to secure players temporarily or acquire free agents whose contracts have expired. Player valuation stands as a central aspect of these transactions- a complicated process influenced by factors such as a player's skill level, age, on-field performance, and market demand.

Operating within a regulatory framework established by football governing bodies like FIFA, the transfer market strives to maintain fairness, transparency, and financial stability in the sport. Despite its inherent complexities, the football transfer market reverberates globally, captivating audiences worldwide, instilling excitement, and nurturing anticipation among fans.

Beyond the financial intricacies, the market's significance lies in its ability to reshape football's competitive landscape. The addition of new talents, the departure of established players, and the constant flux of squad dynamics contribute to the ever-evolving narratives that define each football season. Essentially, the football transfer market transcends mere economic transactions; it is a dynamic force that molds the very essence of the beautiful game.

In this extensive project, our primary focus is on conducting a comprehensive analysis of the football transfer market, particularly within the context of the top 5 global leagues. Over the span of the last decade, we meticulously examine the financial intricacies, strategic decisions, and evolving trends that have shaped the transfer market. Our investigation encompasses an in-depth

exploration of the spending behaviors exhibited by the most influential clubs, unraveling the strategic considerations and patterns that underpin their player acquisition strategies.

Furthermore, we shift our attention to individual player transfers, scrutinizing the record-breaking fees associated with these transactions. This involves an examination of the multifaceted factors that drive these deals, including player performance, market demand, and the financial standing of the involved clubs. Our analysis extends to the last 10 years, allowing us to trace the trajectories of financial investments and identify key influencers, such as economic shifts, regulatory changes, and global events.

Diving into the granular details, we explore player values and transfer fees based on various parameters, including position and age. This analysis provides nuanced insights into the market's valuation criteria, shedding light on how these factors influence the financial dynamics of player transfers. The project also concludes with an intriguing investigation into the correlation between financial investments in transfers and on-field success. By examining how money spent translates into competitive achievements, we offer a holistic perspective on the football transfer market's economic, strategic, and competitive dimensions.

Motivation

Undertaking this Football TransferMarket Data Visualization project is a passionate journey fueled by the transformative events and monumental changes that have recently reshaped the footballing landscape. As a devoted football fan since childhood, the recent surge in interest surrounding high-profile transfers has ignited a deep curiosity to delve into the intricate world of football transactions and club takeovers.

The awe-inspiring transfers of Neymar, Kylian Mbappe, Dembele, Griezmann, Bale, Ronaldo, Coutinho, and others have not only captured headlines but have marked a paradigm shift in the sport. Witnessing the takeovers of iconic clubs by wealthy oligarchs, such as Sheikh Mansour's Manchester City acquisition and the Qatari and Saudi investment fund takeovers of PSG and Newcastle United, respectively, has been nothing short of fascinating.

These pivotal moments extend beyond the pitch, influencing media rights revenues, infrastructure development in nations like Greece and Egypt, and even contributing to significant global events like the hosting of the World Cup. The narrative of football legends transitioning from European powerhouses to playing in Saudi Arabia is a testament to the evolving dynamics of the beautiful game, where passion, finance, and strategic decisions intersect in unprecedented ways.

This project aims to visualize the data behind these captivating stories, offering a unique perspective on the correlation between financial investments and on-field success. It's a journey through the last decade of football, exploring how these monumental transfers and takeovers have not only reshaped individual clubs but have had a profound impact on the global footballing ecosystem. By undertaking this project, we seek to unravel the intricate threads that tie together the passion for football and the contemporary, ever-evolving narrative of the sport millions of people around the globe hold dear.

Statement of Goals

The primary goals of our Football Transfer Market Data Visualization project, guided by the overarching research question, are:

Research Question:

How accurately can we predict a football player's transfer fee based on their market value, and does this predictive ability vary across different player positions and football leagues?

Explore Financial Dynamics: Delve into the financial intricacies of the football transfer market within the top 5 European leagues, aiming to understand the patterns, trends, and strategic considerations that underpin player acquisitions and departures.

Examine Valuation Criteria: Investigate player values and transfer fees based on various parameters, such as playing position and age. Provide nuanced insights into the market's valuation criteria and how these factors influence the financial dynamics of player transfers.

Correlate Financial Investments with On-Field Success: Explore the correlation between financial investments in transfers and on-field success. Understand how money spent translates into competitive achievements, offering a holistic perspective on the economic, strategic, and competitive dimensions of the football transfer market.

Contribute to the Narrative of the Sport: Unravel the intricate threads that tie together the passion for football and the contemporary, ever-evolving narrative of the sport. Contribute valuable insights that go beyond economic transactions, capturing the essence of how player movements reshape the competitive landscape and narratives within football.

By achieving these goals, our project seeks to answer the research question comprehensively, providing statistical insights that enrich the broader understanding of the football transfer market's impact on the beautiful game.

Data Description

The dataset for this project is sourced from the Transfermarkt website, a renowned platform that serves as a comprehensive repository for football-related data, including player transfers, market values, and club information. Transfermarkt has established itself as a reliable and widely-used resource in the football analytics community, providing enthusiasts and analysts with detailed insights into the dynamic world of player movements and financial transactions within the sport.

Focus on Top 5 European Leagues:

The core of our data exploration centers on the transfer data of football players within the top 5 European leagues, each representing the pinnacle of football excellence. These leagues—Premier League (England), La Liga (Spain), Bundesliga (Germany), Serie A (Italy), and Ligue 1 (France)—host some of the most prestigious clubs and talented players globally. Our focus on

these leagues ensures a nuanced analysis of the transfer dynamics that shape the upper echelons of European football.

Variables:

The dataset encompasses a diverse array of crucial columns, each capturing essential information about every transfer. These key variables include:

- **Club**: The football club intricately involved in the transfer process.
- **Name**: The identity of the player undergoing the transfer.
- **Age**: The age of the player at the time of the transfer, a pivotal factor influencing valuation.
- **Nationality**: The player's nationality, contributing to the multicultural tapestry of the footballing landscape.
- **Position**: The playing position of the player, a defining aspect of their role on the field.
- **Short_Pos**: An abbreviated representation of the player's playing position.
- **Market_Value**: The estimated market value of the player, a crucial metric in the valuation process.
- **Dealing_Club**: The club intricately involved in the transfer deal.
- **Dealing_Country**: The country to which the dealing club belongs, adding a geopolitical dimension to transfers.
- **Fee**: The transfer fee associated with the transaction, a quantifiable measure of financial investment.
- **Movement**: Indicates whether the transfer is an incoming or outgoing movement for the player.
- **Window**: Specifies the transfer window during which the deal took place, reflecting the regulatory timeframe.
- **League**: The football league to which the clubs involved belong, providing context to the competitive landscape.
- **Season**: The specific season in which the transfer occurred, anchoring each transaction temporally.

As seen above, the key variables are highlighted and they are used extensively in our project.

Graphs

Figure 2: Transfer Distribution by Age

Description: The histogram provides a visual representation of the distribution of player transfers across different age groups from 2011 to 2020. Each bar on the histogram corresponds to a specific age range, allowing us to discern patterns in the age distribution of transferred players.

Interpretation: From **Figure 2**, we see that the concentration of bars in the 20-27 age range suggests that the majority of player transfers occur within this age bracket. This concentration gradually declines as age increases, indicating that clubs exhibit a heightened interest in acquiring players in their early to mid-20s. The decline in the histogram's tail for older age groups suggests a reduced frequency of transfers for players in their late 20s and beyond.

Figure 3: Transfer Fees Based on Age

Description: The scatter plot illustrates the relationship between player age and the corresponding transfer fees. Each data point represents a player transfer, with the x-axis denoting the player's age and the y-axis representing the transfer fee associated with the transaction.

Interpretation: From **Figure 3**, we can see that the highest transfer fees are within the age group of between 20-25. There are a couple of outliers which correspond to high value players - most of them belonging to the Premier League. These players have a high transfer fee and even they are somewhat between the ages of 19-27. We will be using age as a factor for our model as well.

Figure 4: Raw Relationship between Market Value and Transfer Fees

Description: The scatter plot visualizes the direct relationship between a player's market value and the corresponding transfer fee. Each data point on the plot represents a player transfer, with the x-axis denoting the market value and the y-axis representing the transfer fee. The plot aims to uncover any discernible patterns in how clubs assess a player's value concerning the actual transfer fees paid.

Interpretation: **Figure 4** suggests that market value and transfer fee may be linearly correlated. While each league does seem to have a slightly different slope they all generally follow a positive linear trend. It's important to note that here is where we start to really see the difference between types of players(allstars vs. 'regular' players). We can see where the data starts to get sparse on market value that there is a lot more variability in what the transfer fees may be. Suggesting that maybe a linear model isn't quite perfect for those allstars.

Figure 5: Relationship between Market Value and Transfer Fees using a Log Scale

Description: The scatter plot employs a logarithmic scale on both the x-axis (market value) and y-axis (transfer fee). Logarithmic scaling is applied to accommodate a wide range of values and improve visualization clarity. This transformation helps in identifying proportional changes in market value and transfer fees more effectively.

Interpretation: **Figure 5** is merely an extension of **Figure 4**. Where we see the same trends we saw in Figure 4. The log allows us to view all the data.

Figure 7: Relationship between Market Value and Position on Transfer Fees using a Log Scale

Description: The scatter plot visualizes the relationship between a player's market value, their playing position, and the corresponding transfer fees. Using a logarithmic scale on both the x-axis (market value) and y-axis (transfer fee), the plot aims to uncover patterns specific to different playing positions. Each data point represents a player transfer, color-coded by their position category.

Interpretation: **Figure 7** shows us that there is indeed also a trend in the general positions specifically for the market value and fee relation. We can see here the trend is still positive for all groups but it's quite shallow. Meaning position has some place in our model however it won't heavily skew the data at all.

Figure 10,11,12: Relationship between Market Value and Transfer Fees Over Years (2011-2014, 2015-2018, 2019-2021)

Description: This series of scatter plots explores the relationship between market value and transfer fees over three distinct time periods: 2011-2014, 2015-2018, and 2019-2021. Each scatter plot visualizes the market value-transfer fee dynamics within these time frames, allowing for a temporal analysis of trends and shifts in the football transfer market.

Interpretation: These figures show us that over the years all the leagues trend upwards, but some more than others in certain years. Generally the Premier League leads the way, which makes sense. In recent years second has been Serie A, and back at the start of 2011-2014 Laliga was beating out even the Premier League. The season seems to matter. We will include this as a variable in our model.

Model + Results

Linear Regression with Age Categories

model_age_cat <- lm(fee ~ market_value + short_pos + league + season + as.factor(age_category), data = filtered_data)

AIC: 315901.87, df: 27

Figure 15 below is the predictions of our model plotted against the actual transfer fees, by league.

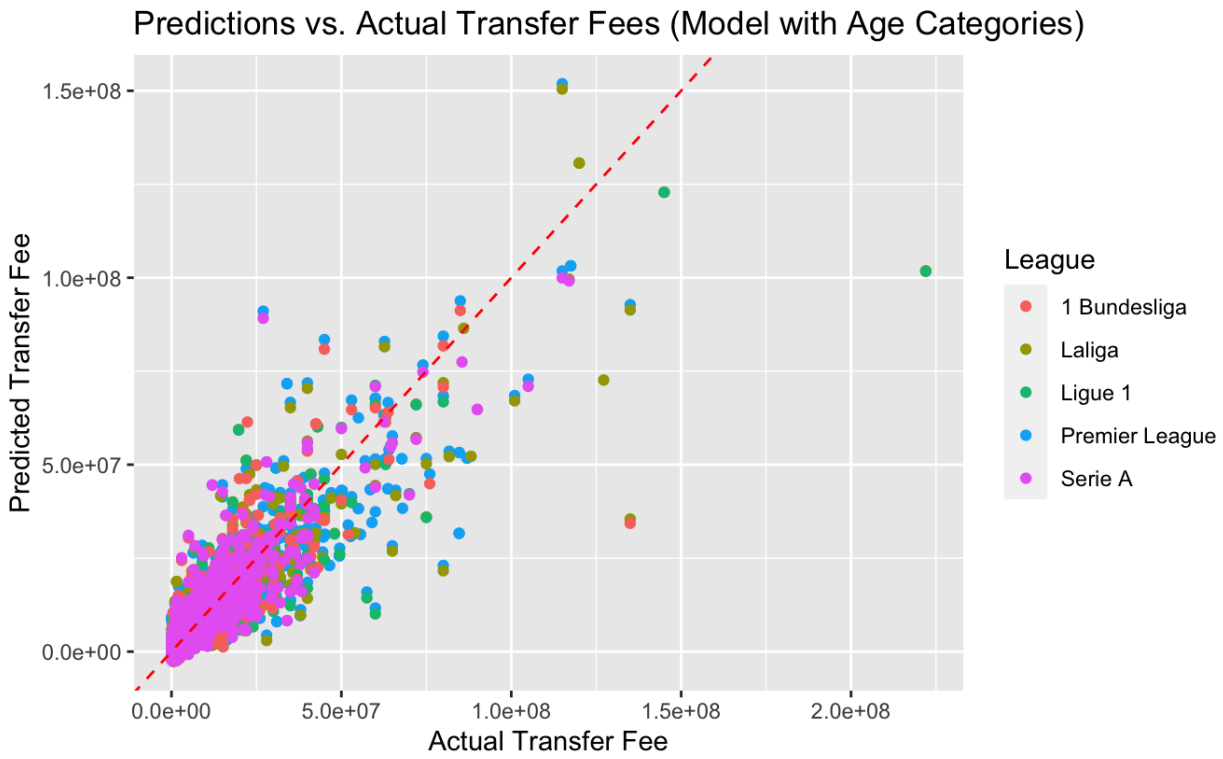


Figure 15

To further inspect the quality of our model we plotted the residuals vs. the log transformed market value. **Figure 16.**

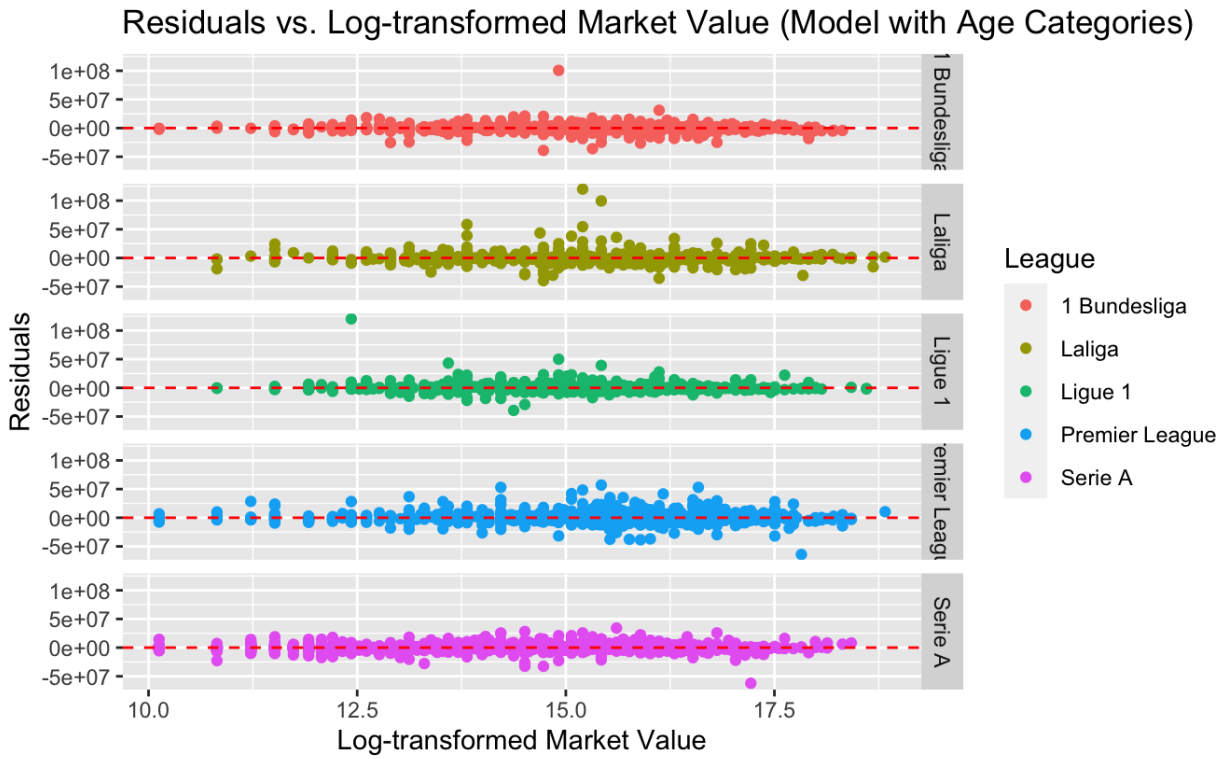

Residuals vs. Log-transformed Market Value (Model with Age Categories)
Residuals
1e+08
5e+07
0e+00
-5e+07
1 Bundesliga
Laliga
Ligue 1
Premier League
Serie A
10.0
12.5
15.0
17.5
Log-transformed Market Value
League
1 Bundesliga
Laliga
Ligue 1
Premier League
Serie A

Figure 16

Next we wanted to track the impact of position in each league. We did this in **Figure 19**.

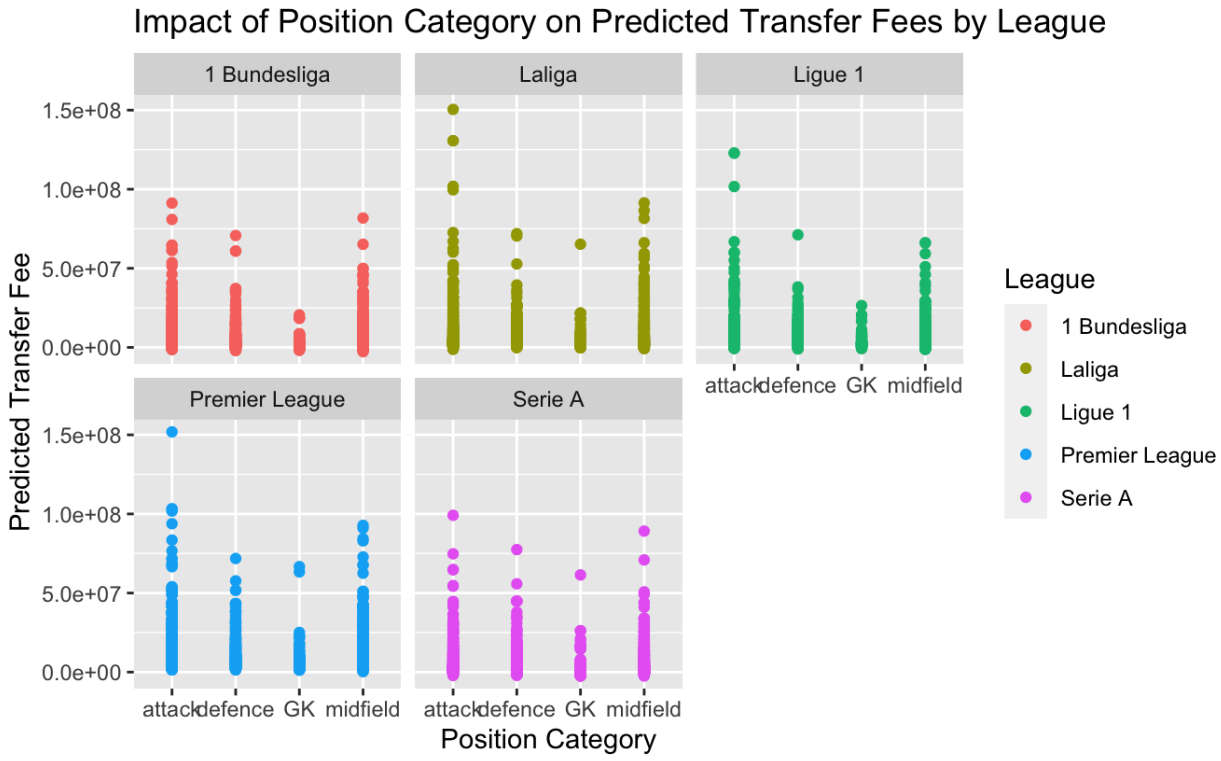


Figure 19

Attack seems to be the most well paid/highest rate players in every league. Goalkeepers tend to be the least. After viewing and experimenting with different models this model was our lowest AIC, while not overfitting the data, it allows us to explain a good deal of the trends.

Conclusions, Limitations, & Future Work

From our analysis, it's fair to say that football market trends are modelable. However, through data exploration and model building we were able to understand that even this data doesn't offer a big enough picture to truly encapsulate the scope of football, but we can still draw some conclusions and point to what may be necessary for further analysis. A linear model does actually do quite a good job describing the relationship between market value and transfer fee. The primary thing we cannot account for in our model is all-star players. Players whose names are associated with the game. All-stars are already a fraction of the fraction of football players that are professionals. This means inherently we don't have a lot of data to work with. These players also bring a certain level of confidence to any team they are on. Marketability may be another aspect of this that we could not account for. A player's popularity would absolutely play into the types of contracts they get. In other words, our linear model is not perfect for players that are a cut above the rest, but does a swell job of explaining most of the trends in the data. In terms of future work, we created a basic GAM model (**Figure 27**), that seems to overfit above a market value of about 100,000,000. If we wished to continue trying to find a better model, refining this would be the direction to go in. Other than that, figuring out some way to account for all-star players with massive market value and fees, would help to refine that model as well.

Appendix

Figure 1:

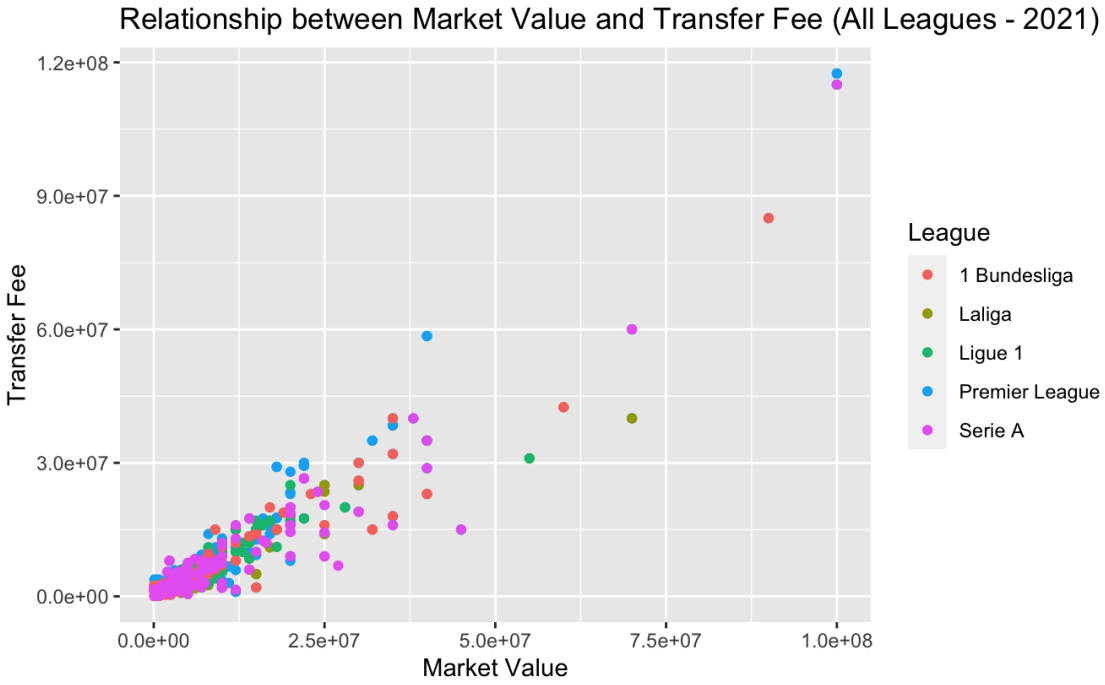


Figure 2:

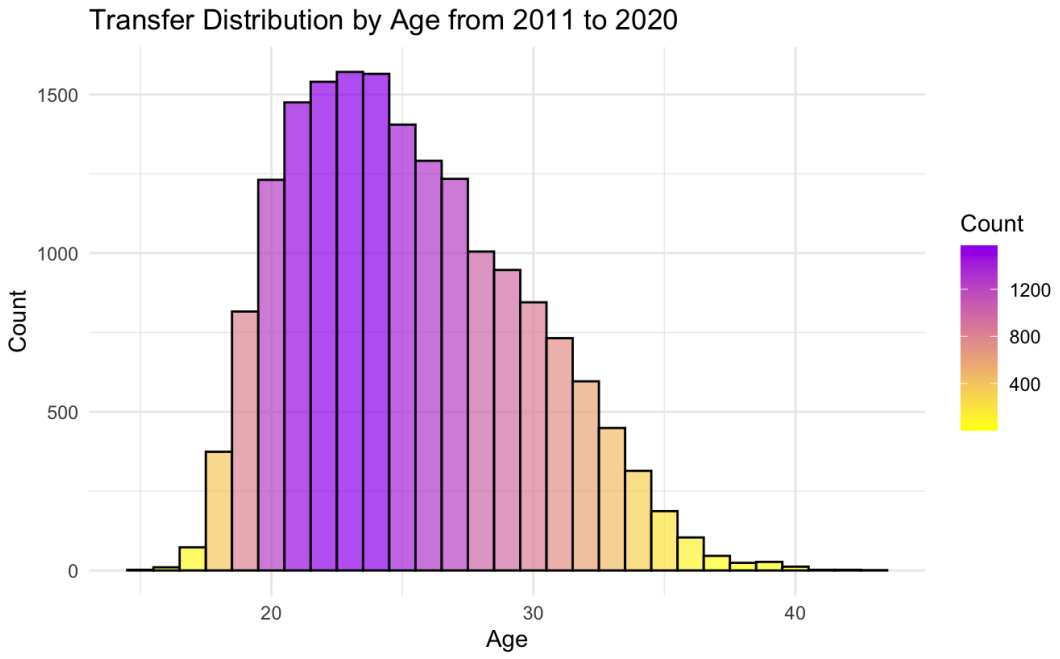


Figure 3:

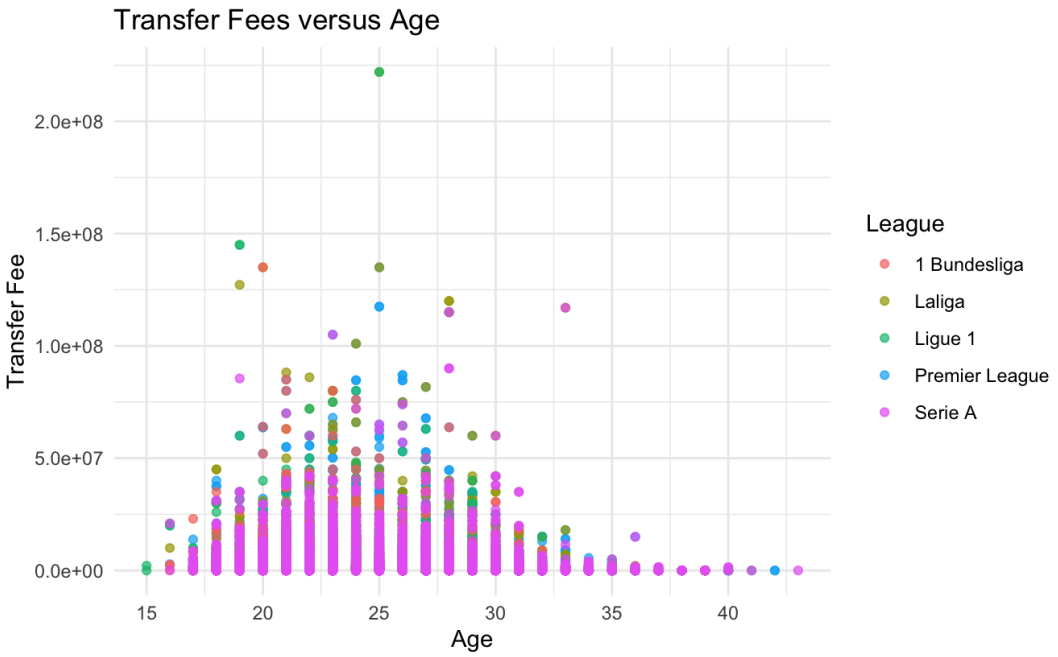


Figure 4:

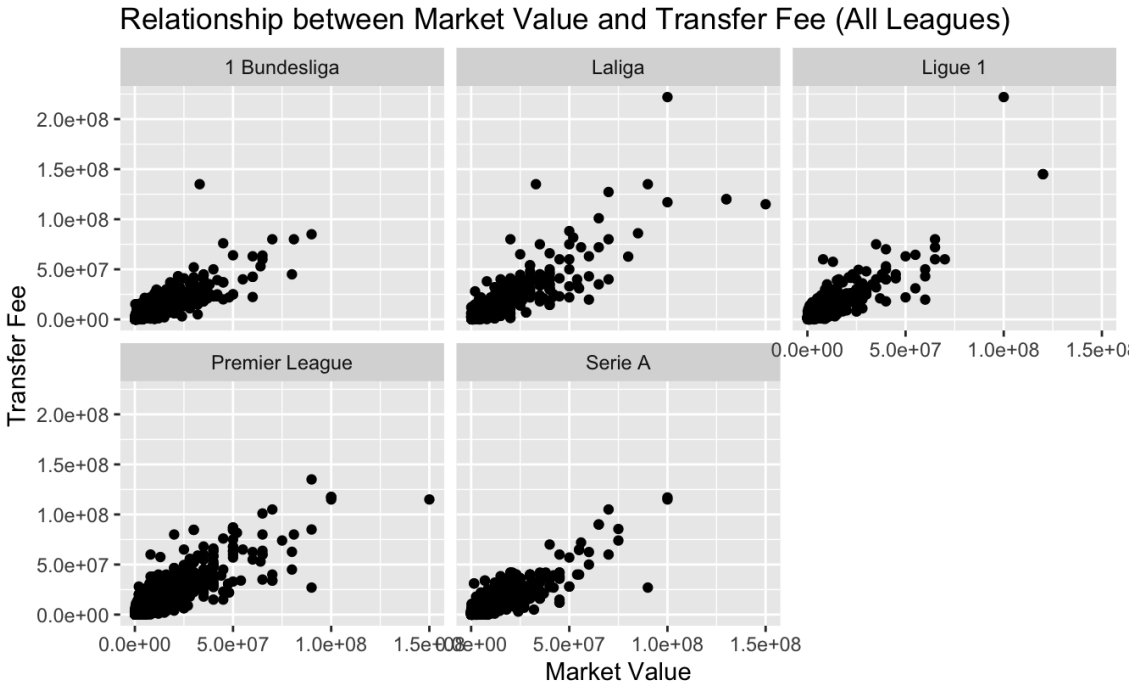


Figure 5:

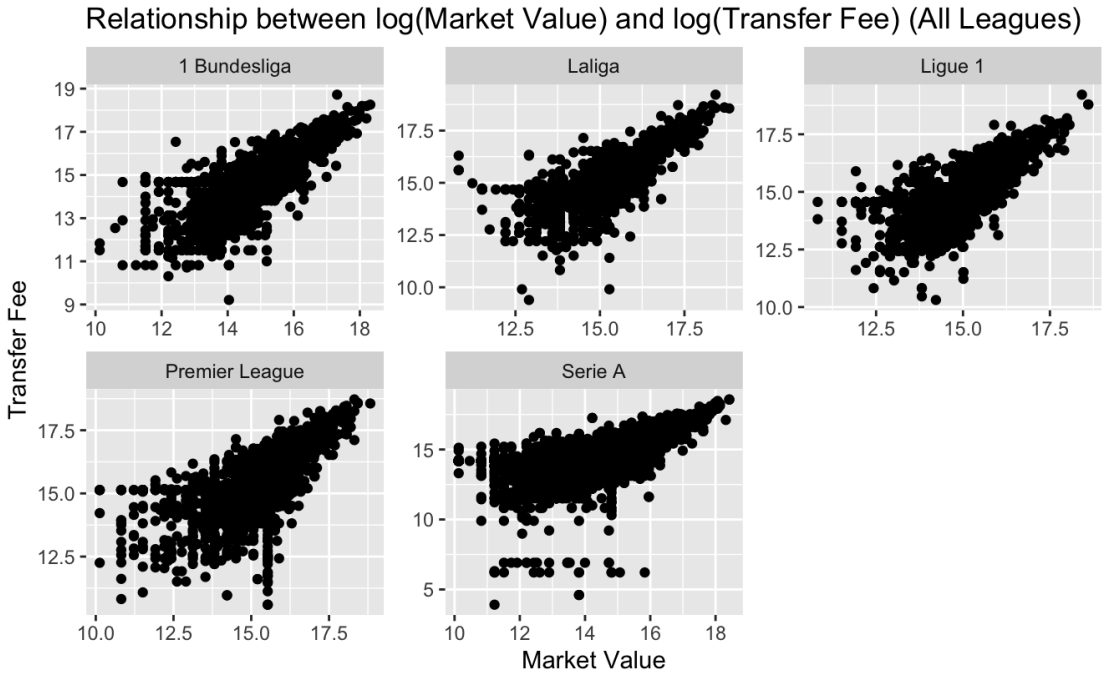


Figure 6:

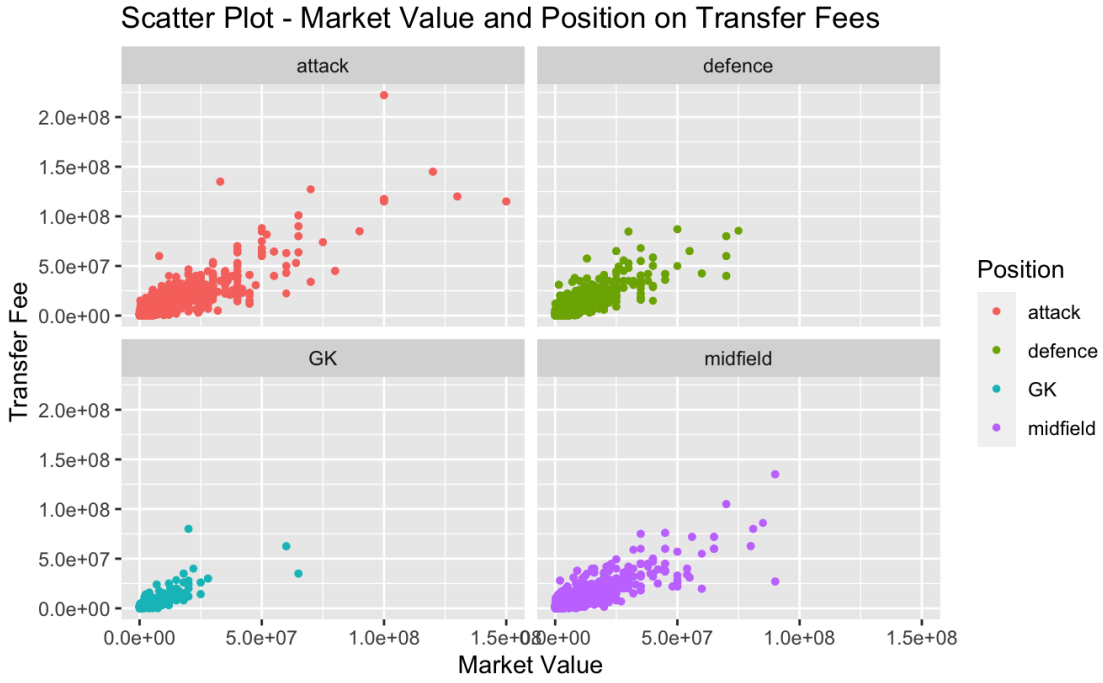


Figure 7:

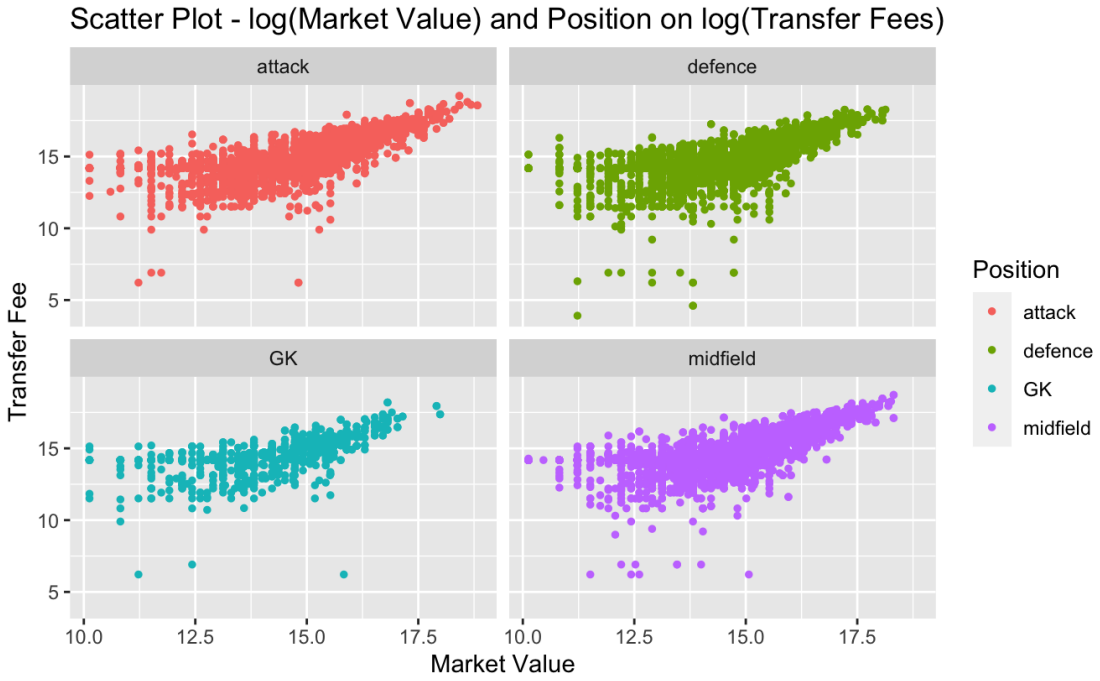


Figure 8:

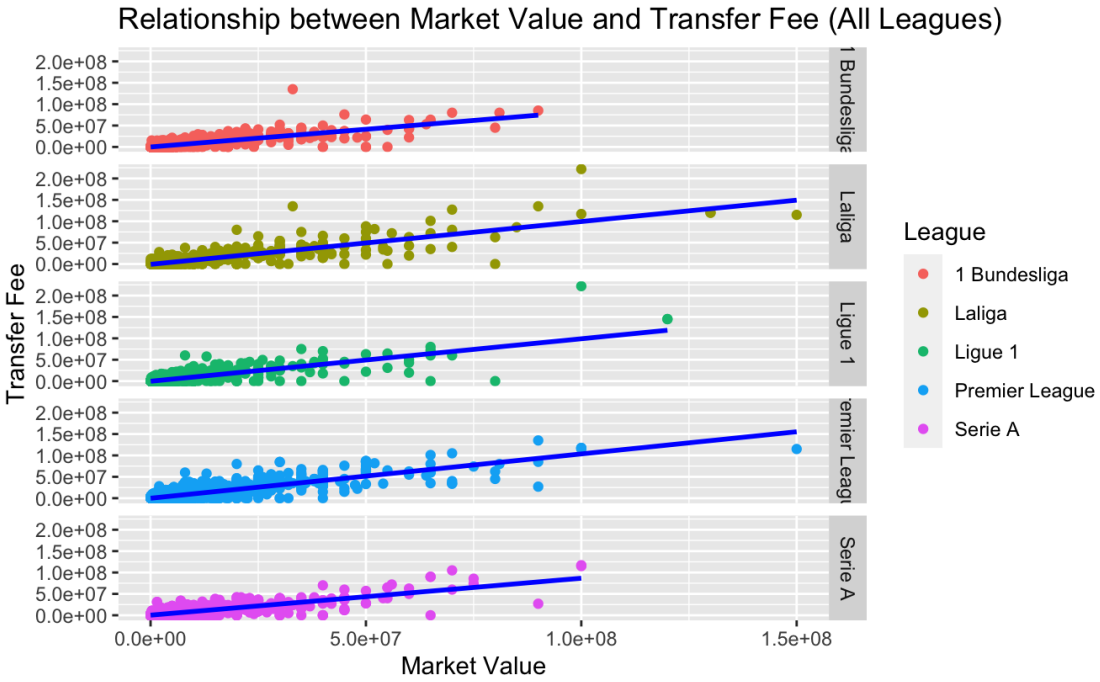


Figure 9:

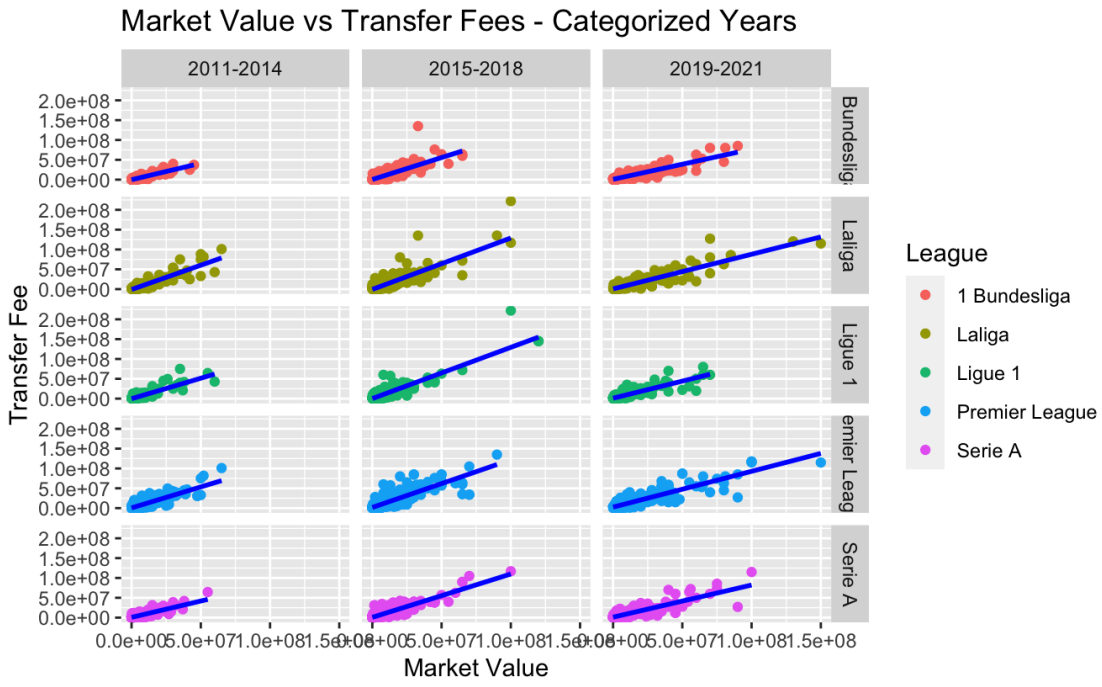


Figure 10:

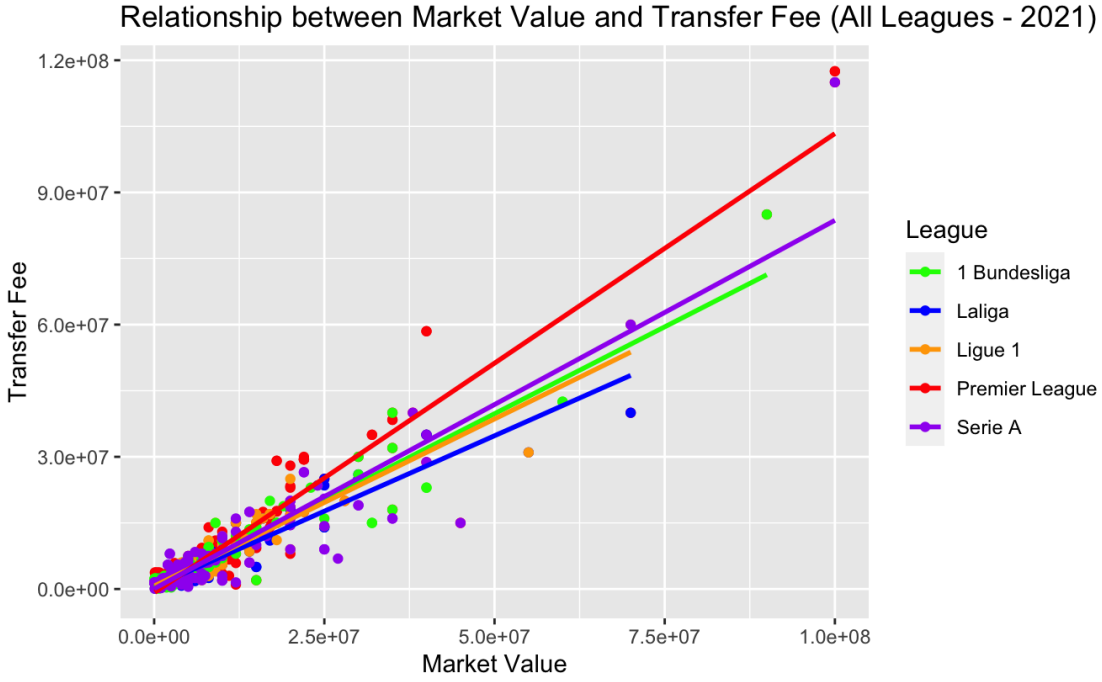


Figure 11:

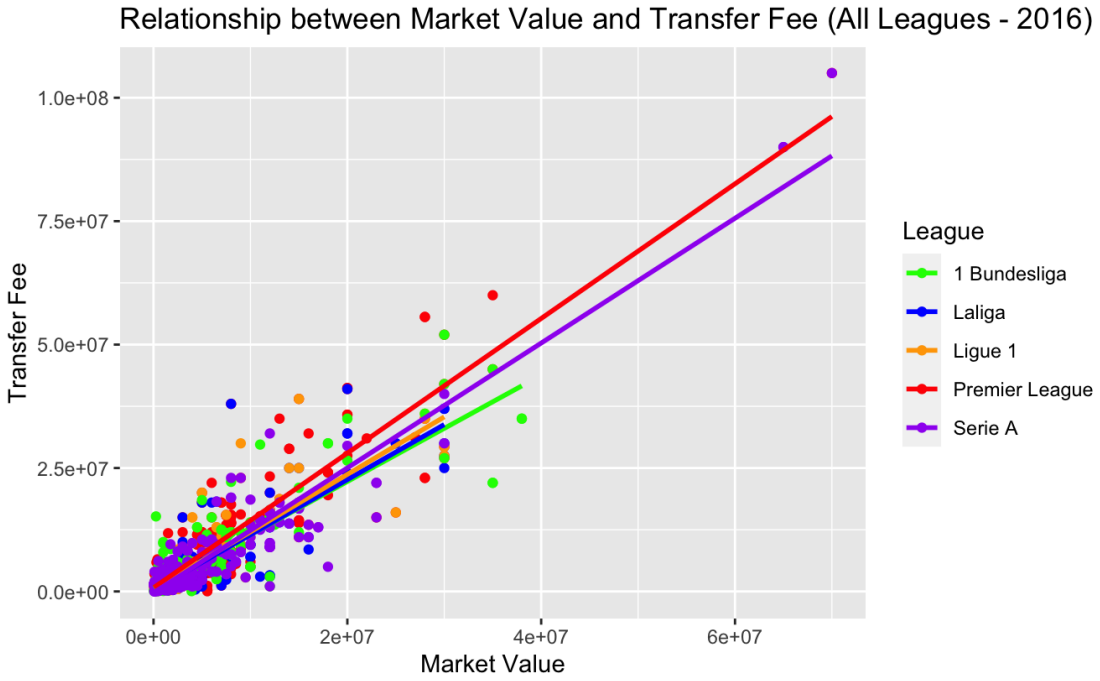


Figure 12:

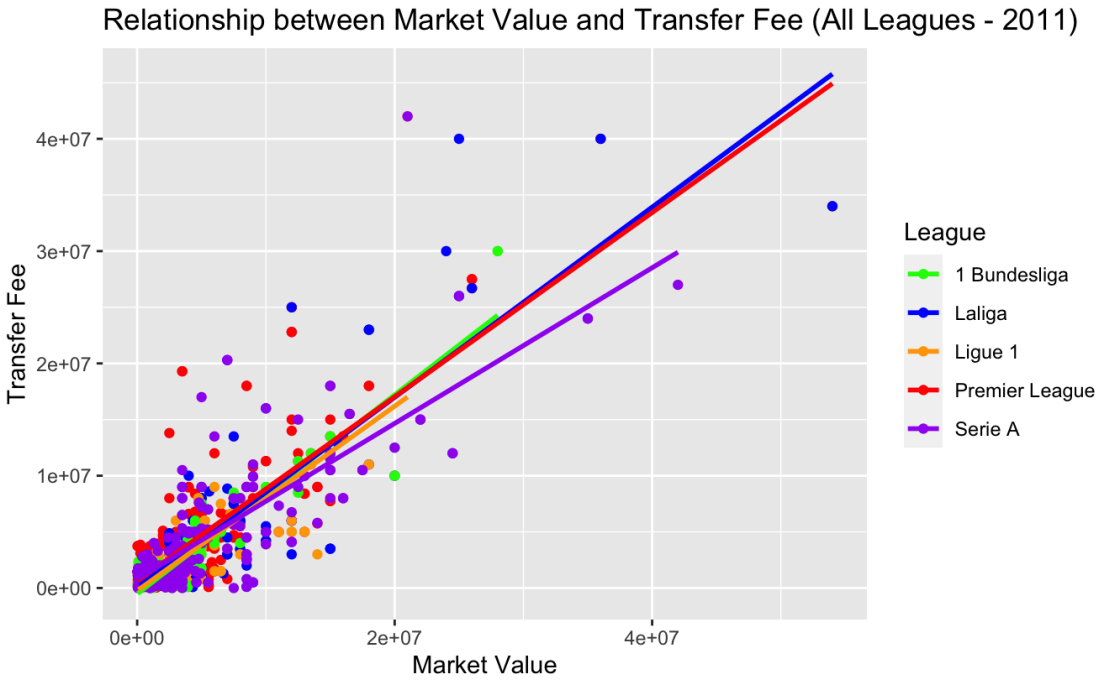


Figure 13:

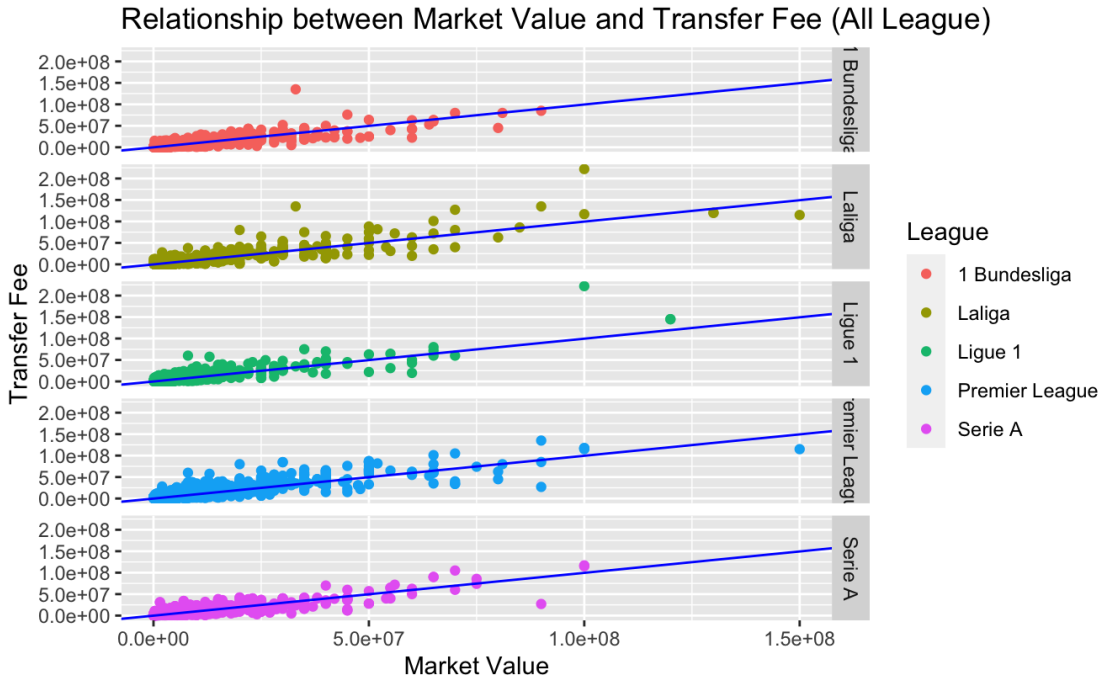


Figure 14:

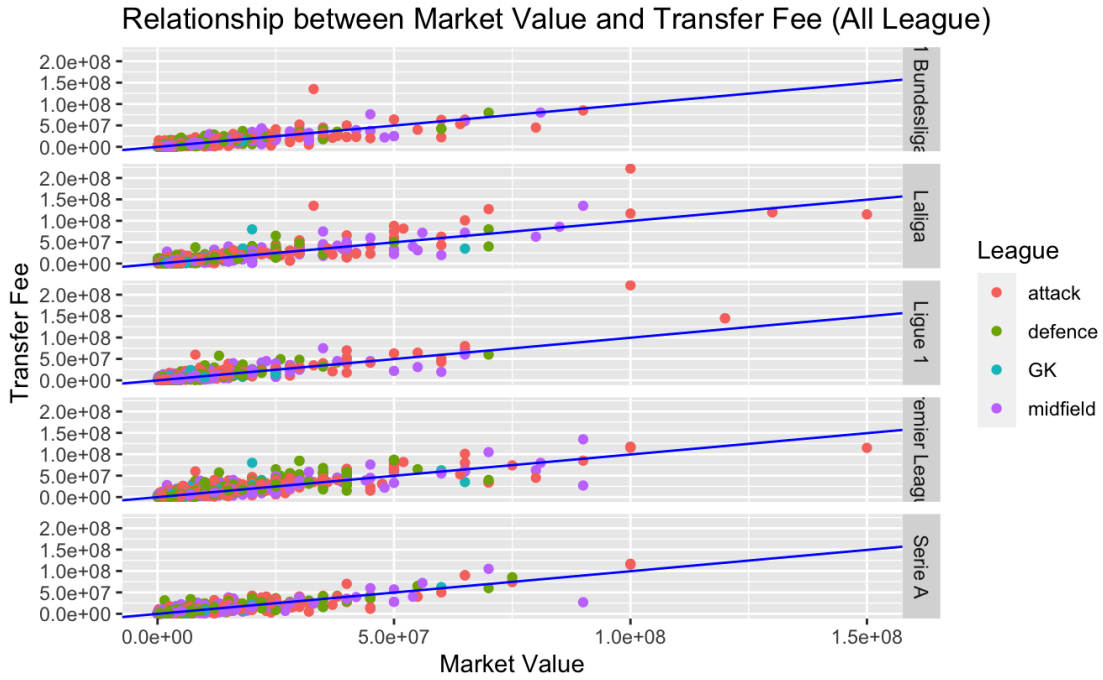


Figure 15:

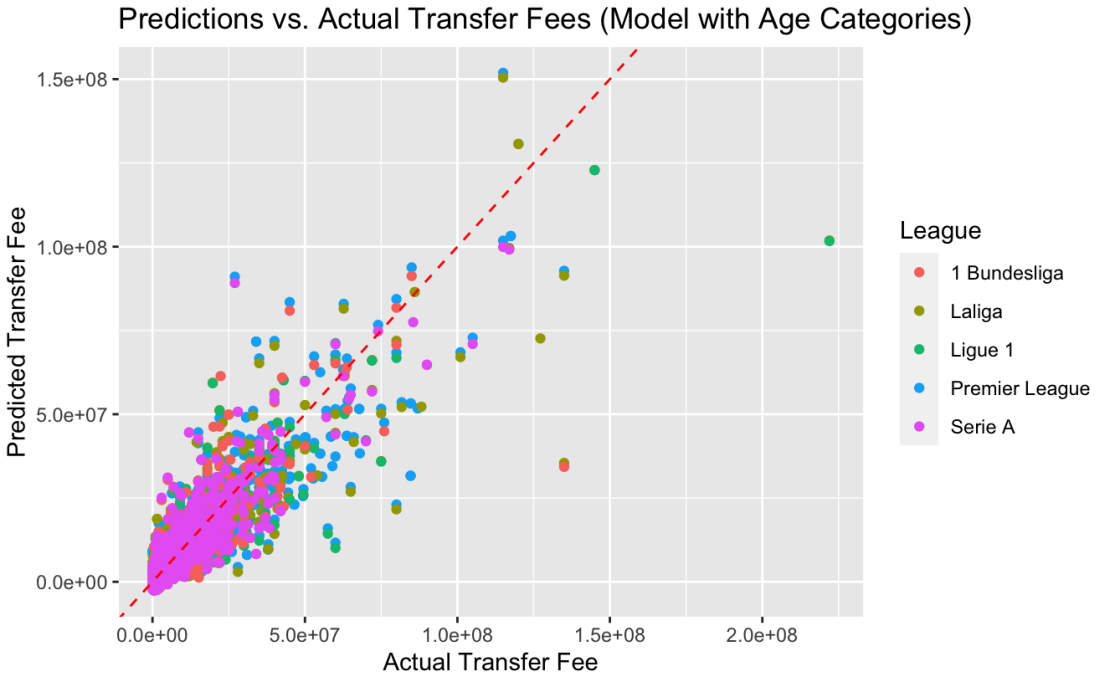


Figure 16:

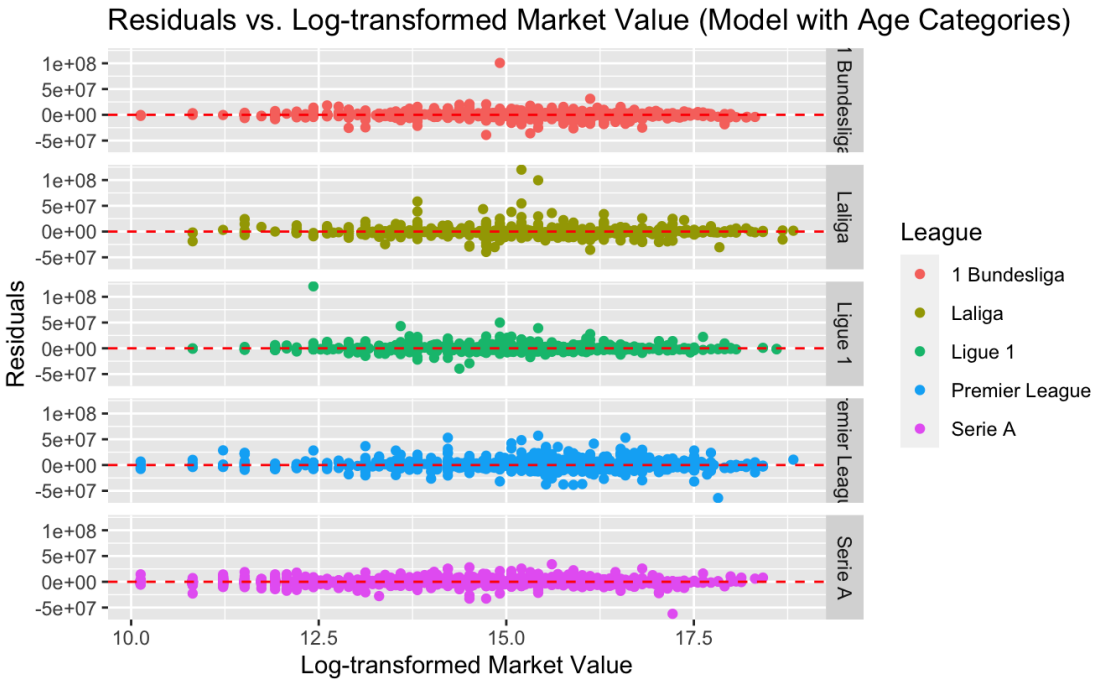


Figure 17:

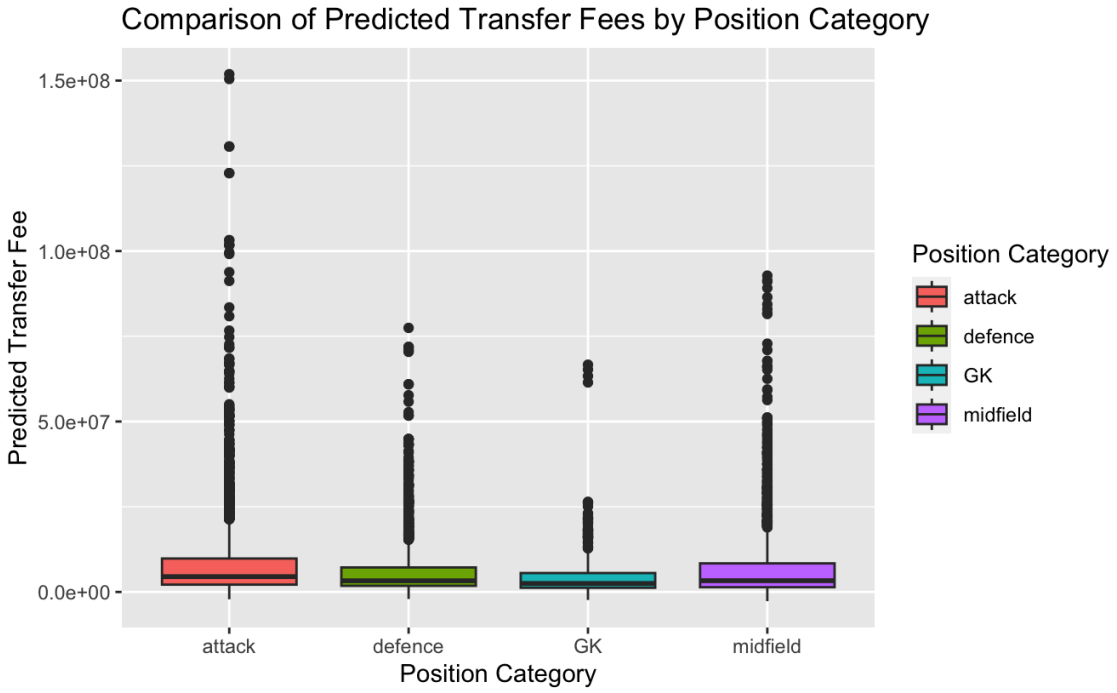


Figure 18:

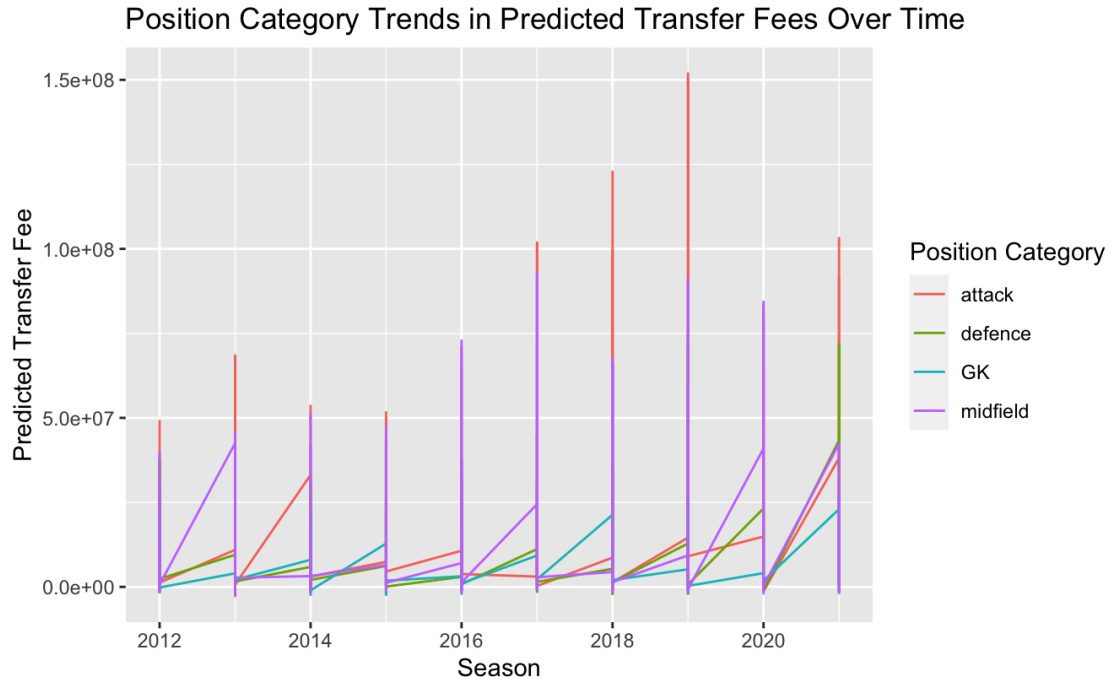


Figure 19:

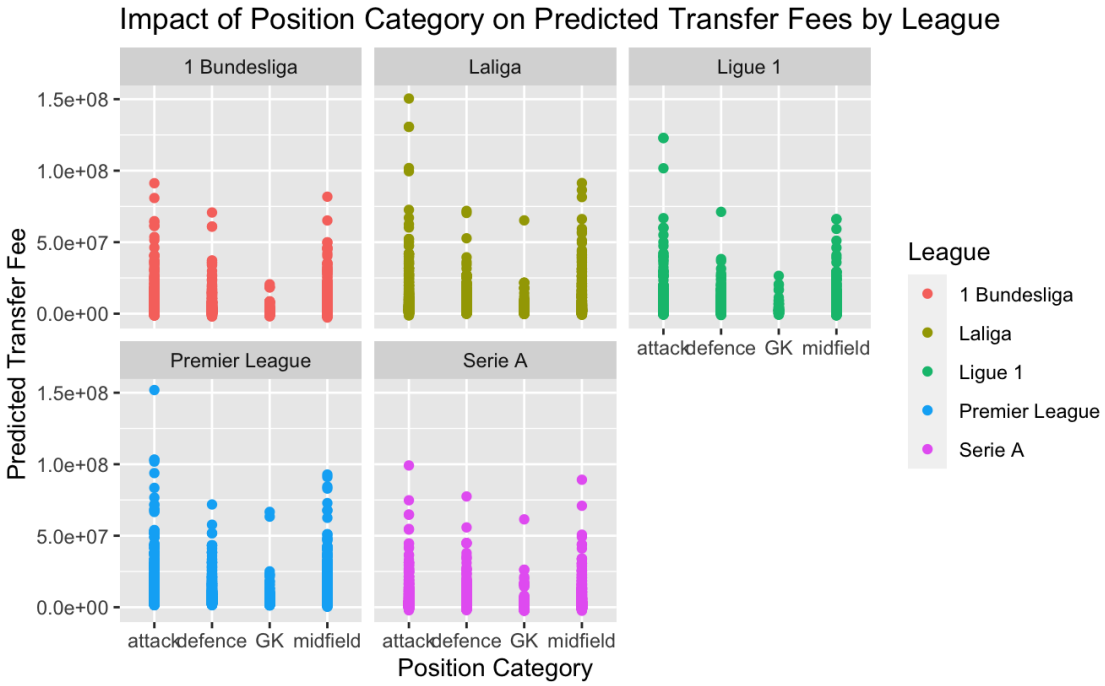


Figure 20:

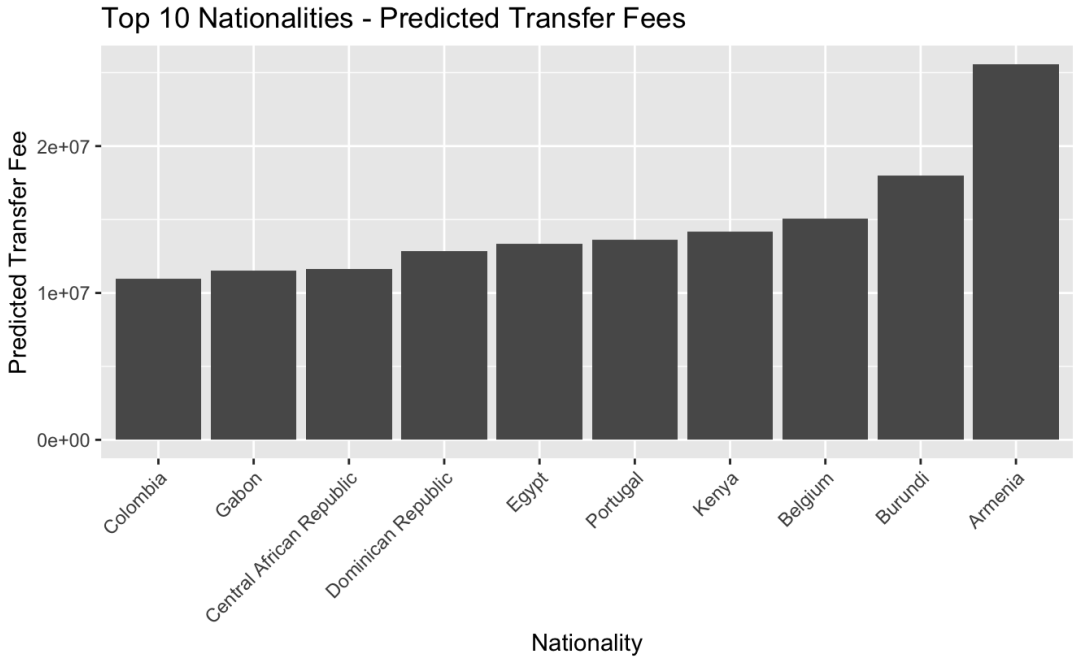


Figure 21:

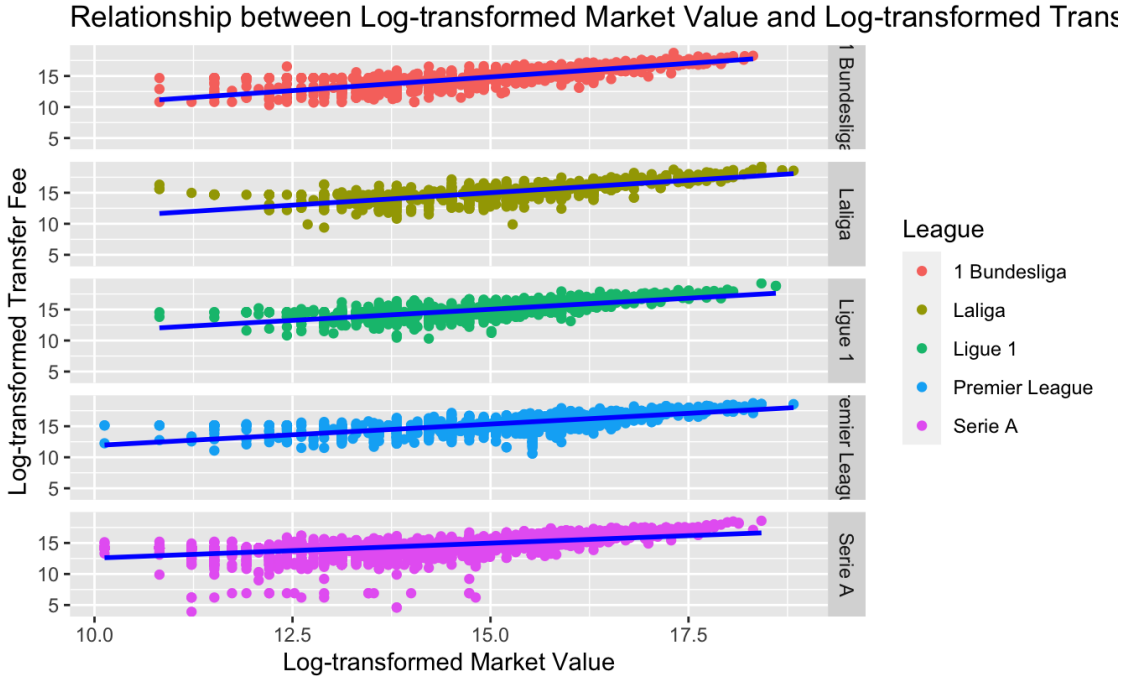


Figure 22:

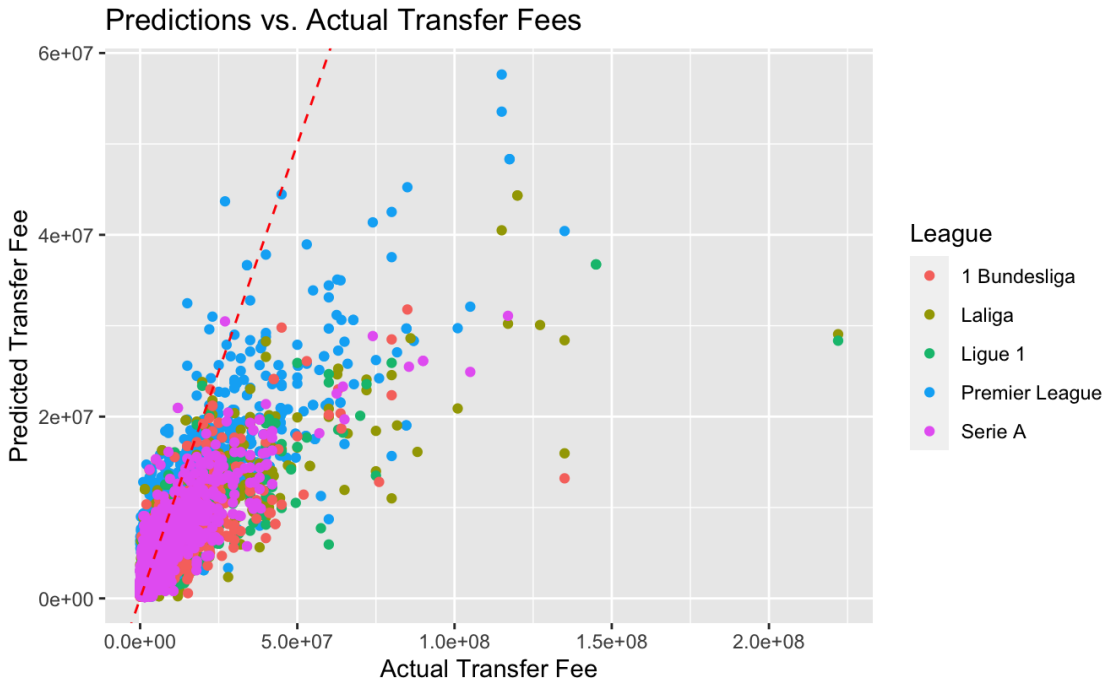


Figure 23:

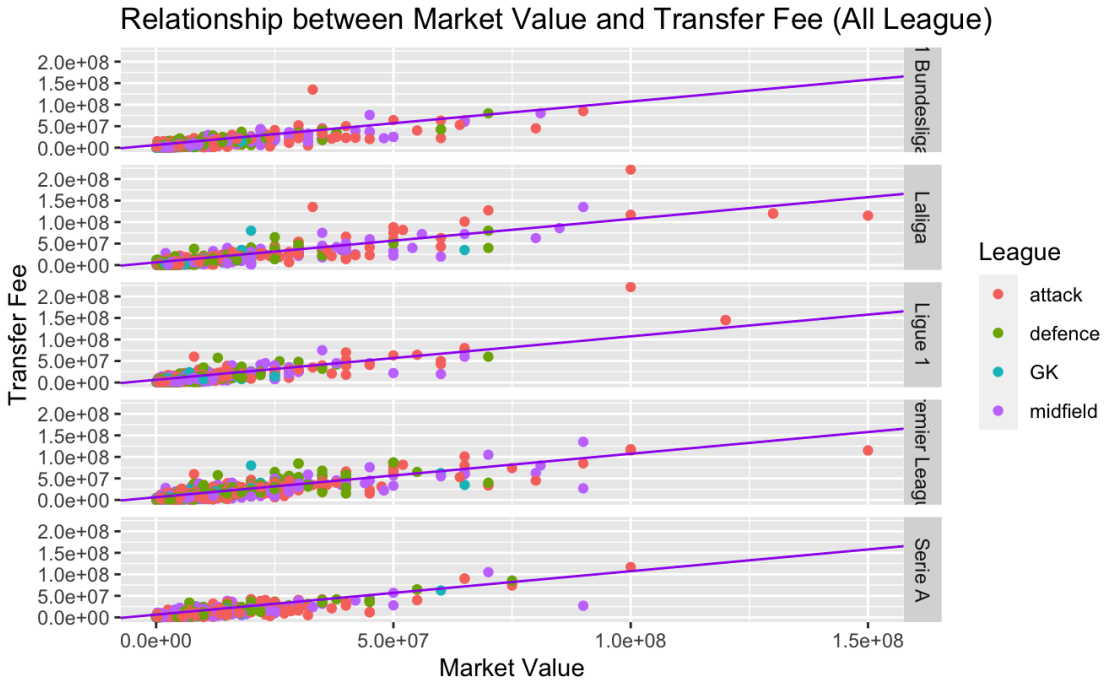


Figure 24:

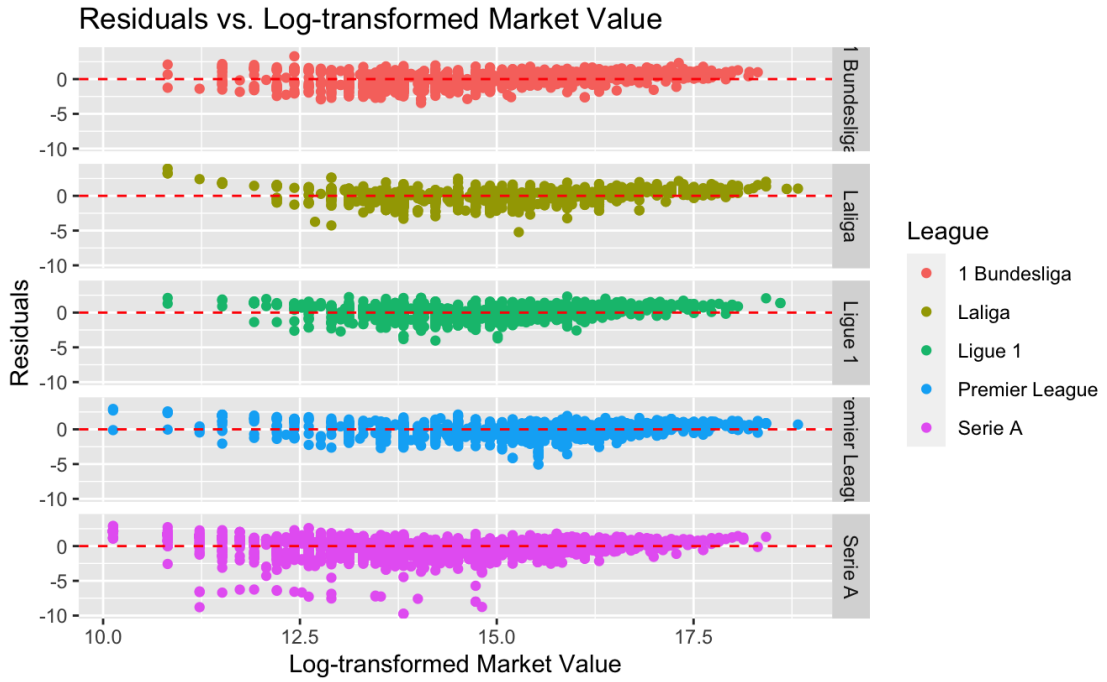


Figure 25:

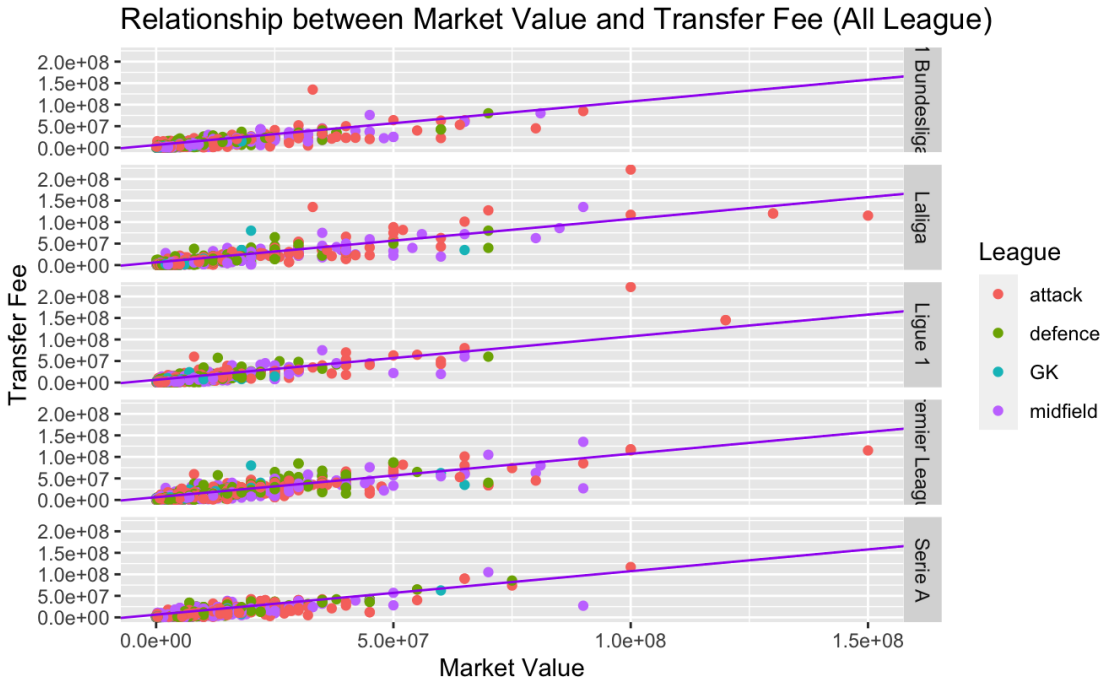


Figure 26:

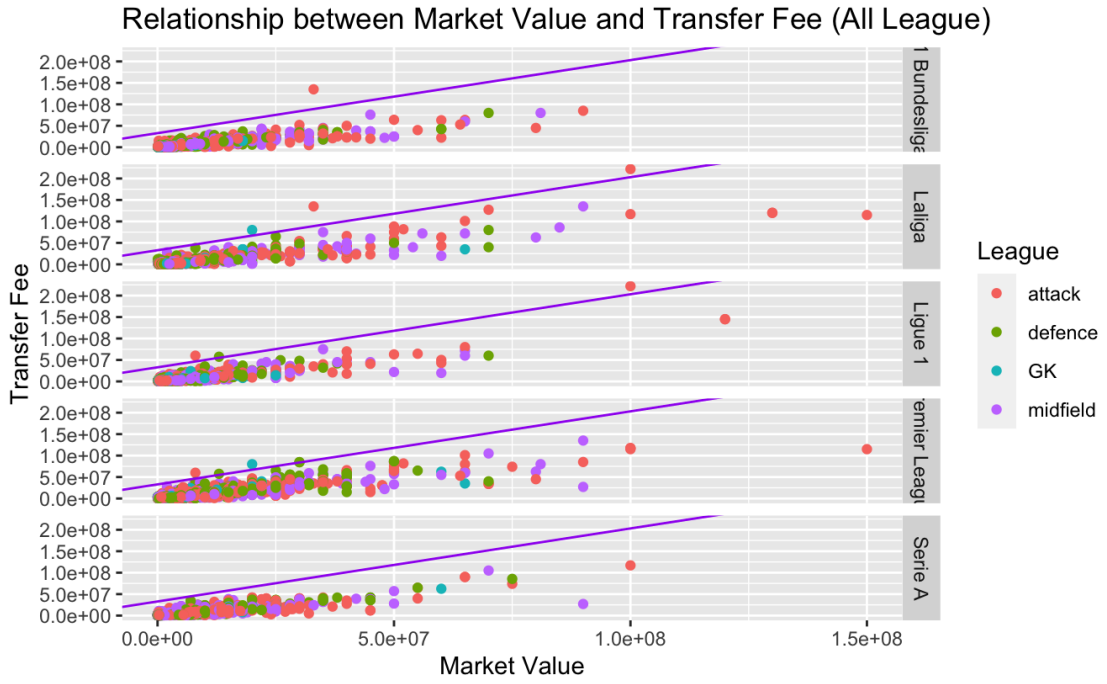
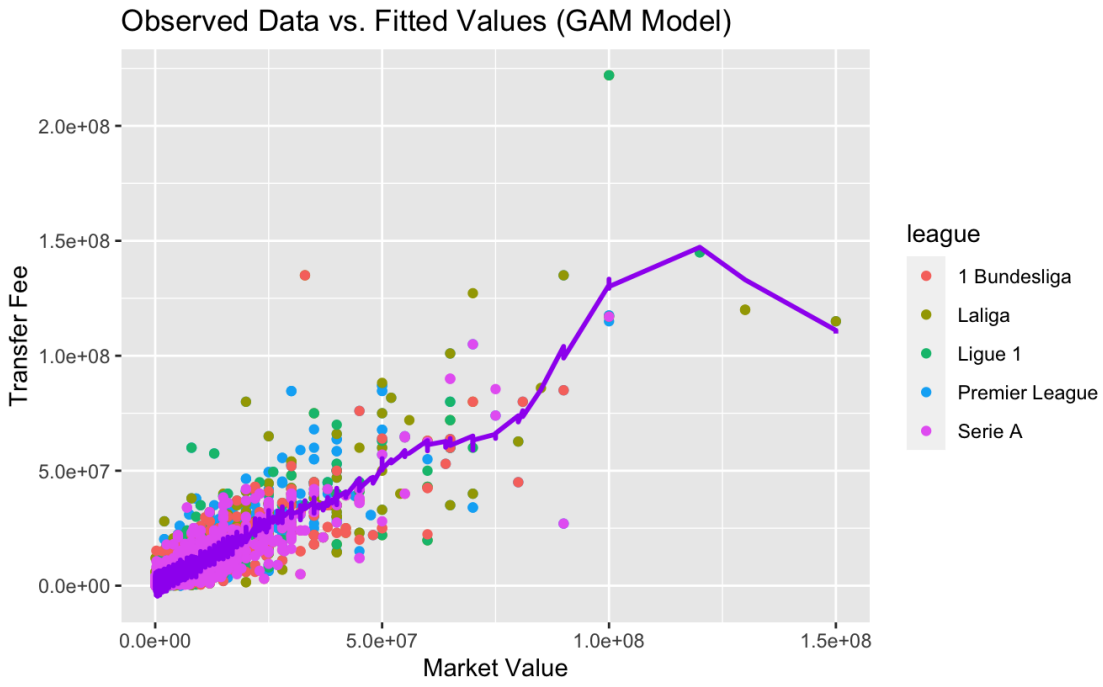


Figure 27:


Observed Data vs. Fitted Values (GAM Model)
Transfer Fee
2.0e+08
1.5e+08
1.0e+08
5.0e+07
0.0e+00
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Market Value
league
1 Bundesliga
Laliga
Ligue 1
Premier League
Serie A