Group S Final Report

Sarah Fatihi, Sophie Diop, James Street

2024-05-01

Abstract

We want to study International Federation of Association Football (FIFA) Stats in comparison to live soccer data. Our goal is to predict transfer market value from the FIFA video game stats. We looked at correlations between FIFA video game predictors including pace, shooting, passing, dribbling, defense, and physicality and transfer market value. We found that pace, shooting, passing, and physicality were significant quantitative predictors of transfer market value. The models accounted for a lot of the variability in transfer market value, although log transformations were used to make the data more linear.

Background and Meaning

In the world of soccer, a player's transfer market value holds immense significance, reflecting their monetary worth in the transfer market. This value is determined by a multitude of factors including age, skill level, performance, contract duration, market demand, and economic conditions within the sport. Understanding the predictors of transfer market value can greatly impact policy decisions and provide insights into the broader socio-economic dynamics of the sport. Our project seeks to explore the relationship between FIFA video game stats and transfer market value in soccer. Specifically, we aim to assess the accuracy of FIFA game stats in predicting transfer market value and identify the best and worst predictors of a player's market worth. FIFA game stats, which use attributes such as pace, shooting, passing, dribbling, defense, and physicality, serve as key indicators of a player's ability and performance in the virtual realm of FIFA video games. However, their real-world applicability in predicting transfer market value remains uncertain. To achieve our objective, we employed a multiple linear regression model to test the null hypothesis that there is no correlation between FIFA video game stats and transfer market value. By analyzing the relationship between these variables, we sought to determine the extent to which FIFA video game stats accurately predict a player's market worth. This approach allowed us to assess how accurate FIFA video game stats are on predicting transfer market value and identify any significant predictors.

Methods

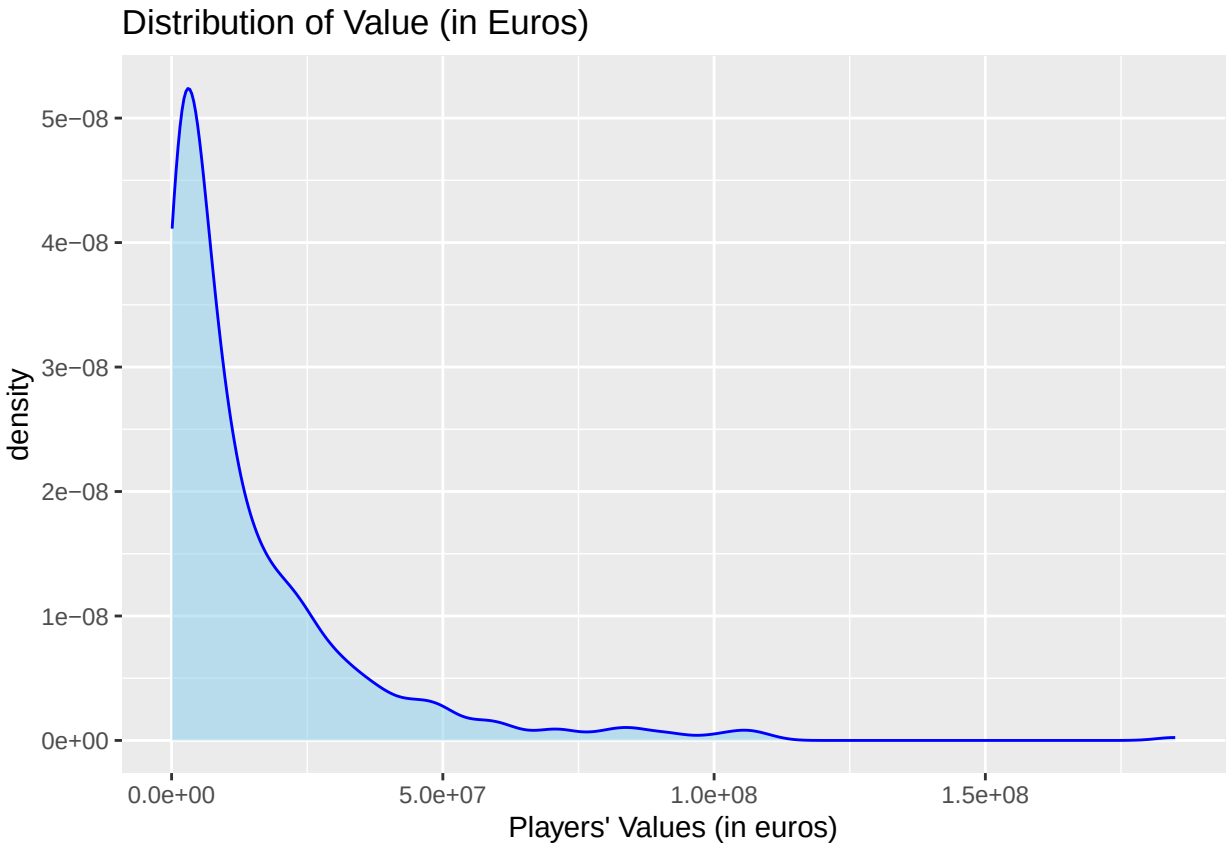
Data The FIFA Football Players dataset, https://www.kaggle.com/datasets/rehandl23/fifa-24-player-stats-dataset, is a comprehensive collection of information about football (soccer) players from around the world. This dataset offers a wealth of attributes related to each player, making it a valuable resource for various analyses and insights into the realm of football, both for gaming enthusiasts and real-world sports enthusiasts. The data is based on the game FIFA 24 which contains information on soccer players from over 19,000 fully licensed players, 700 teams, and 30 leagues in the year 2023-2024.We are planning on slimming down this data to only include players in the premier league. The Premier League is the top professional football (soccer) league in England. It consists of 20 teams, and it's widely regarded as one of the most competitive and popular football leagues in the world. Our financial dataset which contains information of players transfer value was scraped from: "https://www.transfermarkt.co.uk/premier-league/transfers/wettbewerb/GB1/plus/?saison_id=2023&s_w=& leihe=0&intern=0&intern=1". The site is made for football fans and

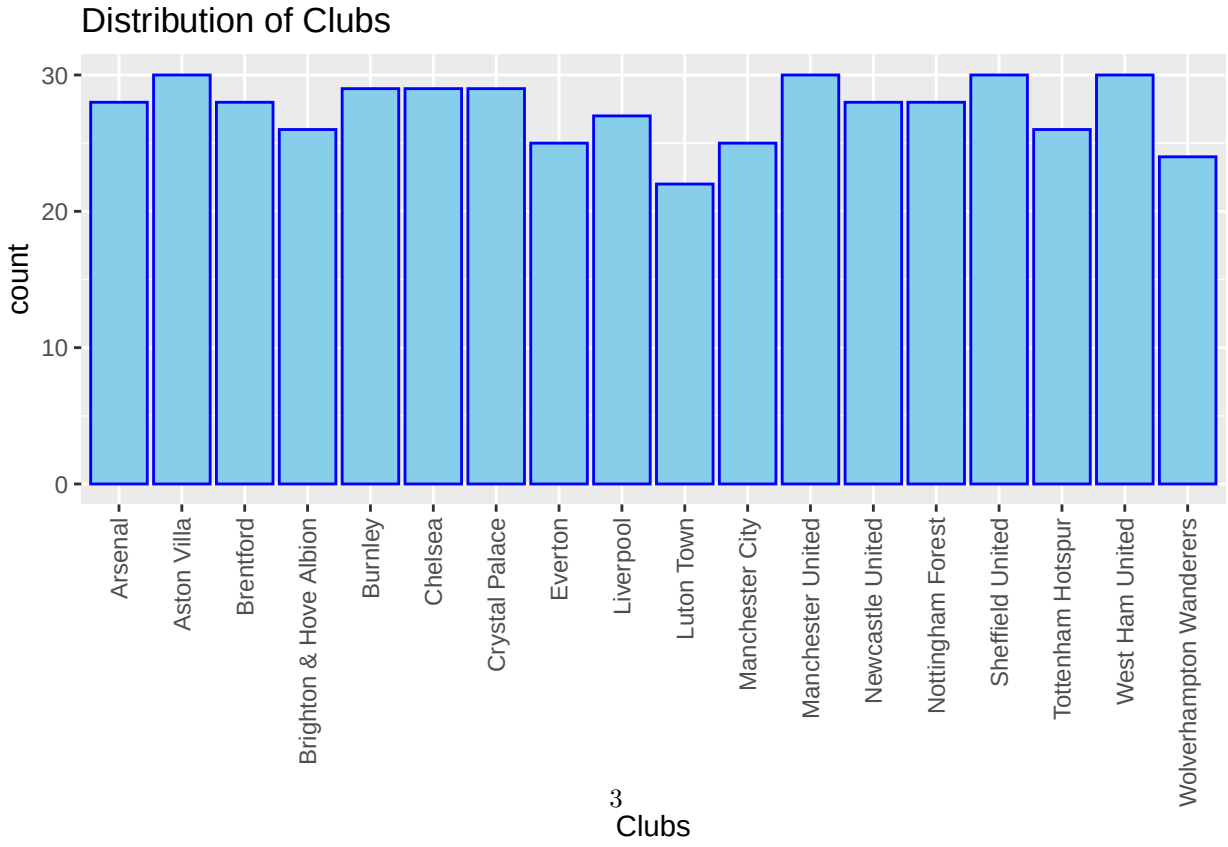
contains transfer related football stats and history statistics for over 100 men's and women's club and national team competitions. This transfer data was taken from the most recent summer transfer window before the 2023/2024 season, and included all players who were signed by premier league teams within this period for a transfer fee. A transfer fee is the amount of money one football club pays to another club in order to acquire the rights to a player. It's essentially the cost of transferring a player's registration from one club to another. In the context of the Premier League, clubs often pay significant sums of money to sign players who they believe will strengthen their squad. These transfer fees can vary widely depending on factors such as the player's skill level, age, potential, contract situation, and market demand.

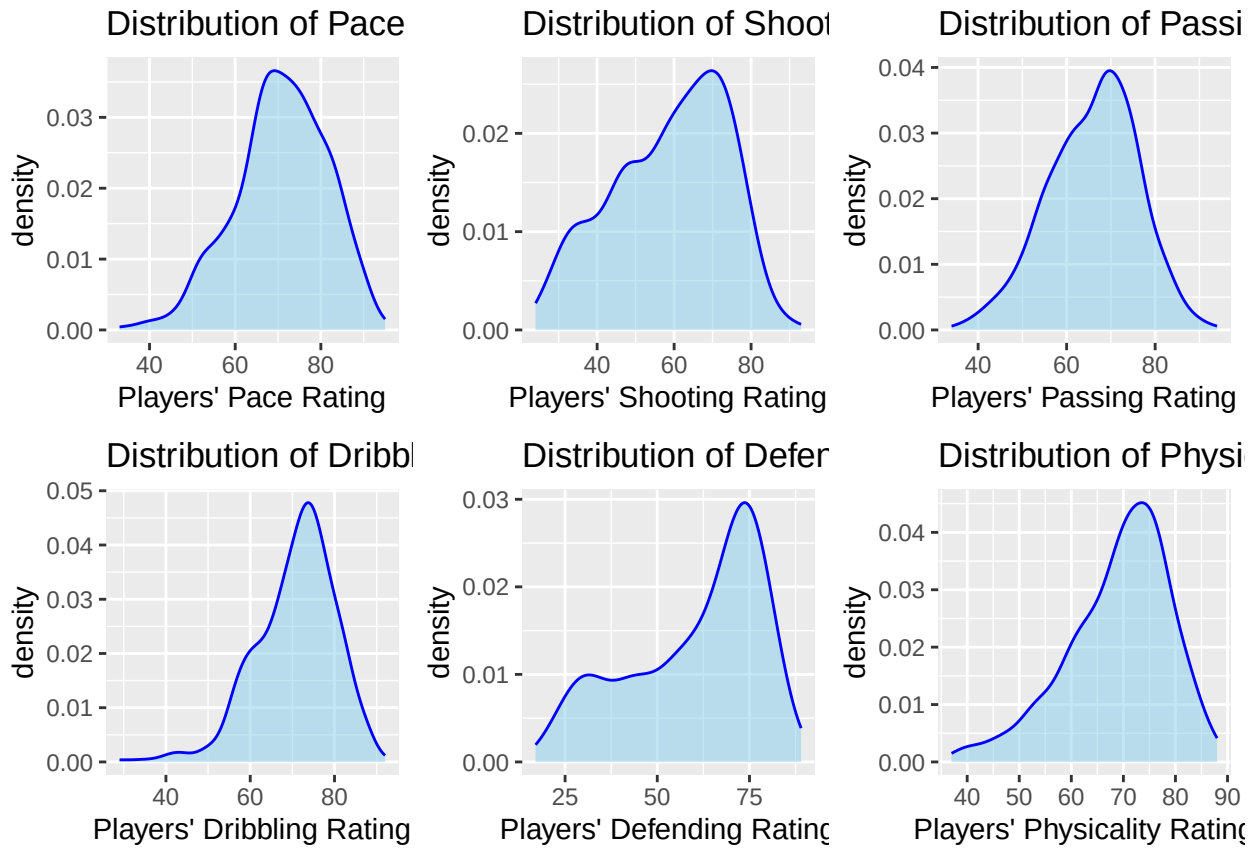
Variables: We have both categorical and quantitative variables as our explanatory variables. These are purchasing club (categorical), in-game Passing (quantitative), in-game Shooting (quantitative), in-game Dribbling (quantitative), in-game Defending (quantitative), in-game Physicality (quantitative), in-game Pace (quantitative). All of our quantitative in-game stats are on a scale of 0-100 determined by the people working at FIFA. Below is a list of the explanatory variables we intend on using: Club - This predictor is the club that the soccer player plays on. The premiere league contains 20 different clubs. These include, Aresenal, Aston Villa, Bournemouth, Brentford, Brighton, Burnley, Chelsea, Crystal Palace, Everton, Fulham, Liverpool, Luton Town, Man. City, Manchester Utd, Newcastle, Nottingham, Sheffield Utd, Tottenham, West Ham, and Wolves. Short Passing - This is a measure at how good the soccer player is at short passing in real time performance on a scale of 0-100 determined by the people working at FIFA. Shot Power - This is a measure at how powerful the soccer player's shot is in real time performance on a scale of 0-100 determined by the people working at FIFA. Dribbling - This is a measure at how good the soccer player is at dribbling in real time performance on a scale of 0-100 determined by the people working at FIFA. Tackles - This is a measure at how good the soccer player is at tackling in real time performance on a scale of 0-100 determined by the people working at FIFA. Physicality - This is a measure at how physical the soccer player is in real time performance on a scale of 0-100 determined by the people working at FIFA. Sprint Speed - This is a measure at how fast the soccer player is in real time performance on a scale of 0-100 determined by the people working at FIFA.

Results

Univariate Descriptive Stats and Figures







Pace: The ratings of players' pace is left-skewed, so we use the median of 71 as a measure of center, and the IQR of 13.75 (first quartile = 64.25, third quartile = 78) as a measure of spread.

Shooting: The ratings of players' shooting is left-skewed, so we use the median of 61 as a measure of center, and the IQR of 24 (first quartile = 47, third quartile = 71) as a measure of spread.

Passing: The ratings of players' shooting has a slight left-skew, so we use the median of 67 as a measure of center, and the IQR of 14 (first quartile = 59, third quartile = 73) as a measure of spread.

Dribbling: The ratings of players' dribbling is left-skewed, so we use the median of 72 as a measure of center, and the IQR of 12 (first quartile = 65, third quartile = 77) as a measure of spread.

Defending: The ratings of players' defending is left-skewed with a potential second peak, so we use the median of 66 as a measure of center, and the IQR of 29 (first quartile = 46, third quartile = 75) as a measure of spread.

Physicality: The ratings of players' physicality is left-skewed, so we use the median of 71 as a measure of center, and the IQR of 13 (first quartile = 63, third quartile = 76) as a measure of spread.

Bivaraite Analysis

Interpreting the Model

Conclusion