FIFA FOOTBALL FANATICS - GroupS

James Street, Sarah Fatihi, Sophie Diop

02-26-2024

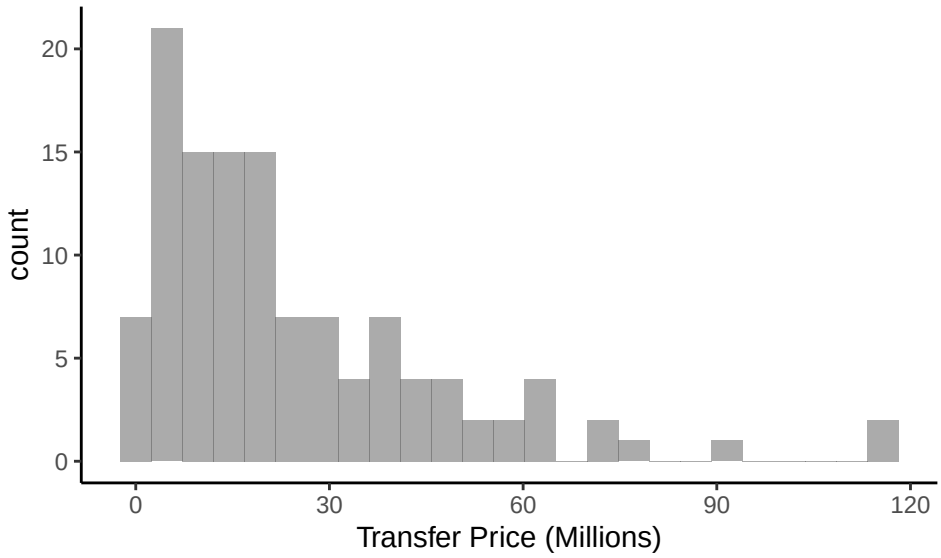
PURPOSE:

We want to study FIFA Stats in comparison to live soccer data. Our goal is to predict transfer market value from FIFA stats. Are FIFA stats accurate in predicting transfer market value? What are the best and worst predictors of transfer market value? FIFA uses six core predictors in determining a soccer players ability, are these predictors the best at determining transfer market value?

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. 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 tranfer 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.

POPULATION: We want to be able to generalize our conclusions to all soccer players in the Premier League. The Premier League is the English soccer league that is comprised of 20 clubs. In order to do this we will need to filter down our data sets to only include only players active in the league as to avoid extrapolation.

RESPONSE VARIABLE: Our response is transfer value - quantitatively measures how desirable of a player they are in terms of money in (millions of) euros. The transfer value distribution is right skewed and unimodal with a median of 17.15 and an IQR of 22.88265.



min	Q1	median	Q3	max	mean	sd	n	missing
0.87	7.95	17.3	34.85	116.6	24.19492	22.87137	120	53

EXPLANATORY VARIABLES: 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.

The short pass distribution is left skewed and unimodal with a median of 63 and an IQR of 13. The shot power distribution is relatively normal and unimodal with a potential second peak. The mean is 58.18268 and a standard deviation of 12,97192. The dribbling distribution is left skewed and bimodal with a median of 62 and an IQR of 17. The stand tackle distribution is left skewed and bimodal with a median of 57 and an IQR of 36. The sprint speed distribution is left skewed and unimodal with a potential second peak. The median is 68 and an IQR of 18. Due to the overwhelming number of clubs in our current data set we can not determining the skew or modality of the distribution and our stats appear as NA.

min	Q1	median	Q3	max	mean	sd	n	missing
11	55	63	68	93	59.33333	14.32602	5682	0

min	Q1	median	Q3	max	mean	sd	n	missing
20	49	59	68	94	58.18268	12.97192	5682	0

min	Q1	median	Q3	max	mean	sd	n	missing
5	51	62	68	95	56.12848	18.77208	5682	0

```
min Q1 median Q3 max     mean       sd    n missing
  7 30     57 66  91 48.82084 20.97597 5682       0
```

```
min Q1 median Q3 max     mean       sd    n missing
 13 57     68 75  97 64.95899 15.11404 5682       0
```