Probing the Quality of Football Leagues Through Player’s Foot Laterality: A Data Analytics Approach

Dicky Januarizky Silitonga*, Muhammad Andi Yudha Cahya Adinegara, Dwinda Mariska Putri

* Institut Teknologi Sumatera, Indonesia
b PT Petrosea, Indonesia
c AgroParisTech, France

Correspondence: dicky.silitonga@ms.itera.ac.id
Received: 1 Mar 2023  Accepted: 25 Sep 2023  Published: 30 Oct 2023

Abstract
The ability to use both sides of feet in association football, or so-called soccer in some countries, is highly valued, and therefore professional athletes rigorously train their non-dominant foot to perform effectively on the pitch. However, a subject that relates that players’ laterality aspect to the overall quality of the league of a given country has not yet been widely explored. This paper discusses the subject through a data analytics approach, leveraging the abundance of publicly available football datasets to describe interesting phenomena through data visualizations. In the present study, the foot laterality measure is represented by the percentage of left-footed scored goals and league’s overall quality by the market values of players. Results show that in general, leagues with higher market values have higher rates of left-footed goals. The findings in this work agree with those in past research conducted with other methods, hence confirming the validity of data analytics based on the crowdsourced football data. This research intends to emphasize the potential of data science in sport sector and motivate the football professionals to use big data to fine-tune their physical training methods and sport psychology approaches.

Keywords: Left-footedness; foot laterality; football analytics; data-driven approach; crowd-sourced sport data

1. Introduction

Studies revealed that a vast majority of the world population is right-handed, comprising a significantly unbalanced ratio between right- and left-handed humans of nearly 9:1 (Papadatou-Pastou et al., 2020). Furthermore, this laterality, or preference for one side of the body, works in the same way in most of the cases for handedness and footedness (Ipek et al., 2021; Sacco et al., 2018). Research shows that the right-footers are also heavily dominant in human population (Alexandre Jehan Marcori, Victor Hugo Alves Okazaki, 2020; Bondi et al., 2020) although at a lower percentage than the right-handers. In normal life for regular activities, this side preference does not really matter, or one can adapt for certain purposes without being severely handicapped. Indeed, in societies, right-handedness sometimes becomes culturally induced (Bazo et al., 2022; Ittyerah, 2019), where failing to act with that side of hand may be considered a non-compliance to traditional or common values. Thus, in general, one can still manage to live and thrive without any restrictions regardless of his or her lateral preference, however unbalanced it is.

Nevertheless, in some sports, well-balanced laterality is valuable and highly sought after (Pietsch & Jansen, 2018). In football, the ability to utilize both feet is beneficial for players to excel in their game, with a strong impact on agility (Zouhal et al., 2018). Having both feet work at nearly the same level...
of control and strength allows a player to have flexibility to operate across the width of the field, agility to anticipate the incoming ball and freedom to take a shooting decision as the momentum arises without the need to first position the ball towards the strong foot. As a remark, the term football that will be used for the rest of this paper refers to the association football or soccer, not to be confused with American football.

To improve lateral balance, one should train so as to develop the strength and control of his or her genetically non-preferred side (Akpinar, 2022). That being said, high-performing players are those who, besides other skills and instincts, strive and somehow achieve a certain level of lateral balance in terms of their feet. High-performing footballers commonly play in the top-tier leagues, where players with genetic physical advantages, hard-working ethics, and focused mentality are sharpened by the skill-intensive, well-managed competition. With that concept in mind, then one may draw a hypothesis that the overall quality of a group of players in a competition, represented by a league, can be reflected by the footedness contribution of players in decisive actions or events. Specifically, this paper discusses the footedness ratio of goals, i.e., whether they are scored by left or right foot. The study takes into account the data from multiple leagues with the objective of observing the characteristics of those leagues or the corresponding countries’ football atmosphere.

In this technology-dependent era where many sectors start to rely on big-data for analytics and decision-making processes, data science has also substantially intervened sport industry, particularly football or soccer. Nearly all aspects of players and team performance has already been subjects of analysis with the openly available data. With the publicly available data, retrieving samples for statistical analysis becomes much more feasible and less costly as compared to physical or on-site data acquisition. For instance, this research aims to provide a descriptive statistics of football data based on the geographic locations, which is classified as clustered sampling according to Dahlan (Sopiyudin Dahlan, 2009) in his book “Besar Sampel dan Cara Pengambilan Sampel”. With the presence of the publicly available big-data, enormous size of samples is accessible at the data scientists’ fingertips, and even many of them are free of charge. This advantage addresses the problems considered in the book “Biostatistics: A Methodology for the Health Sciences” (Gerald van Belle et al., 2004) regarding the issues in calculating sample size, namely the cost and sample availability.

A description of the revolutionary roles of big data is elaborated by (Daniel Memmert & Dominik Raabe, 2018) their book “Data Analytics in Football”, taking examples of Germany’s football realm to exhibit the power of big-data analytics. As in the other fields with access to big-data, any variables correlations are worth observing despite being arbitrary, since they may turn out to provide valuable information. Link (Daniel Link, 2018) provides examples of some non-traditional or interesting correlations, such as the effect of match importance to player’s activity the how the vanishing spray discourage players to violate the rules. Likewise, the authors would also like to explore uncommon correlations in the present work.

This work demonstrates a proof of concept of employing a subset of publicly available football data on the world-wide-web to deduce valuable information with the help of computational tools to visualize and uncover the obscured correlations between variables. In line with the growing philosophy in data science today, the data tells its own story through visualization. This paper offers new insights in football by correlating the players’ laterality and monetary values to draw an overall figure of leagues’ quality, by utilizing data from a crowdsourcing website. In fact, a previous study endorsed the use of information from crowdsourcing platforms like transfermarkt.com for research purposes (Prockl & Frick, 2018). That topic, combined with the stated data acquisition method, to the best of the authors’ knowledge, has not been published elsewhere.
Today’s football is a grand industry sector where business and financial consciousness are part of the game itself (Alaminos et al., 2020). Therefore, it is reasonable to consider the monetary factor, in this case the players’ prices which in turn accumulate into team valuation, as parameters that determine the quality or competitiveness. Additionally, discussing the financial impact of certain aspect in sport is expected to drive players’ motivation in training for sharpening the value-adding skills. The motivation, as reported by researchers, have a positive influence to the performance of football players (Munir et al., 2022). In this research, the foot laterality figures are correlated with those valuations to decipher the meaningful information. Note that the team value here refers to the players' total market values instead of the club's valuation as a company or business entity.

Beyond academia and researchers, this paper is aimed towards a diverse spectrum of audience such as football professionals, not only to gain insights on the subject covered in this paper, but also to spark inspirations. Extended readerships are also expected, including football enthusiasts who enjoys a fun reading while expanding their scientific knowledge, as well as data scientists and machine learning experts with curiosity in finding easter eggs in football statistics which in turn will advance the roles of big data and artificial intelligence in sports.

2. Method

This research is designed as a fully online- and computational-based work. First, the data is retrieved from transfermarkt.com, an online source that provides extensive data analytics related to football from around the world. Then, the data is processed to render a visual representation that enables easy interpretation of the correlation between variables. Keep in mind that although this research involves many numerical data, statistical parameters or jargons are not necessarily to be emphasized in the analysis, since the ultimate intention of this work is to gain a general picture of how players’ foot laterality describes overall quality of leagues based on the monetary value point of view. Again, without undermining the importance of statistical analysis, this essay mainly attempts to show the usefulness of data visualization to help understanding the phenomena merely through glancing on the data.

The subjects being sampled for this research are teams of top-tier leagues from leading countries representing all regional confederations, excluding the Oceania Football Confederation (OFC) due to its small size and least coverage. In addition, to provide richer knowledge for Indonesian readers as the main audience of this paper, the author includes Indonesia and Thailand, although those countries occupy the lower part in FIFA country ranking. That being said, this research employs cluster sampling method based on country where teams are playing in. The size of each cluster sample varies depending on the number of competing teams in the league. Table 1 lists the cluster size for analysis in this paper.

<table>
<thead>
<tr>
<th>Country</th>
<th>League</th>
<th>Cluster sample size (teams)</th>
<th>League market value (USD mio.)</th>
<th>Average team market value (USD mio.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>England</td>
<td>Premier League</td>
<td>20</td>
<td>10320</td>
<td>516</td>
</tr>
<tr>
<td>Spain</td>
<td>La Liga</td>
<td>20</td>
<td>4830</td>
<td>241.5</td>
</tr>
<tr>
<td>Germany</td>
<td>Bundesliga</td>
<td>18</td>
<td>4290</td>
<td>238.3</td>
</tr>
<tr>
<td>Italy</td>
<td>Serie A</td>
<td>20</td>
<td>4520</td>
<td>226</td>
</tr>
<tr>
<td>France</td>
<td>Ligue 1</td>
<td>20</td>
<td>3350</td>
<td>167.5</td>
</tr>
<tr>
<td>USA</td>
<td>Major League Soccer</td>
<td>29</td>
<td>1130</td>
<td>38.8</td>
</tr>
<tr>
<td>Mexico</td>
<td>Liga MX Clausura</td>
<td>18</td>
<td>812</td>
<td>45.1</td>
</tr>
<tr>
<td>Brazil</td>
<td>Camp. Bras. Série A</td>
<td>20</td>
<td>1330</td>
<td>66.26</td>
</tr>
<tr>
<td>Argentina</td>
<td>Liga Profesional de Fútbol</td>
<td>28</td>
<td>783.7</td>
<td>27.9</td>
</tr>
</tbody>
</table>
Datasets extracted for footedness analysis are statistics of goals from 2018 to 2022 that contain the number of goals scored with the right and left foot for all competing teams. The underlying reasons for the selected time frame are that within a relatively short period, the prominent players and overall style of play across leagues have not undergone a major change. Hence, the analysis will be relevant to global modern, state-of-the-art, football games. That period also considers the chance of blank dataset due to halted competitions during COVID-19 outbreak.

The primary research instrument is data observation and collection from a publicly available, online data crowdsourcing platform: transfermarkt.com. The website is a part of the German-based publishing group, Axel Springer SE, that owns several major media with extensive circulation in Europe. Practically, transfermarkt.com are accessed worldwide for various real-life purposes from entertainment and gaming, up to financially impacted activities such as player scouting and, undoubtedly, betting. Considering the scale of users and reputation of the platform, the author believes that the data available on transfermarkt.com is valid, reliable and objective to the higher extent.

To analyze the data, the percentage of parameters is taken instead of the nominal value, that way each parameter will be normalized to ensure comparability with the other parameters. Since the sample leagues have different numbers of teams, as listed in Table 1, thus the figure taken for comparison is the percentage of goals scored with either side of feet. That way, the nominal number of goals which may vary a lot from top leagues to minor leagues is normalized into percentage level, allowing facile comparison among them. Furthermore, since left-footedness is the one being rare and thus becoming the main interest, the laterality parameter taken here is in terms of left-foot goal percentage.

Choosing the right-footed percentage would work as well for the analysis yet showing the left one is more intuitive for the reader to focus on the unique skill pursued by high performing players. The left-footed goal percentage is calculated through the following straightforward formula:

\[
\text{Left-foot goals } \% = \left( \frac{\text{Left-foot goals}}{\text{Left-foot goals} + \text{Right-foot goals}} \right) \times 100 \% \tag{1}
\]

The percentages for all the studied leagues are plotted against the corresponding average team market values. A league’s market value, in this case, is the sum of the prices of players in the league in question, not the value of the team as a business entity or company that may include other non-player assets. The average team market values are taken here instead of the total league market values of the league since, again, the number of teams playing in a league is different, as shown in Table 1.

To add a broader picture on the comparison of foot utilization of players, an evaluation on the distribution of passes attempted with either foot is included in our analysis. The data is limited to the top 5 European leagues because only those leagues have extensive open-source big data availability. Limiting to only the first 5 rows in Table 1, the sample size in our cluster sampling technique becomes more uniform, since most of them has the size of 20, except Bundesliga with only 18. The data, however, takes into account all of the passes attempted by players, not only the decisive ones.
Therefore, it is more appropriate that this indicator be evaluated in terms of the frequency to provide evidence that normally players do prefer to use their right foot, even in those top-class competition.

Collecting data from crowdsourced platforms remains as the research instrument for this additional analysis. The datasets are from FBref.com for 2021-2022 Big 5 European Leagues Stats. The website was created by Sports Reference (sports-reference.com), a team behind popular sports statistics websites like Baseball-Reference.com and Basketball-Reference.com, launched in June 2018 with league coverage for six nations: England, France, Spain, Italy, Germany, and the USA. The vast audience and utilization of their data justify the validity and reliability of this research instrument. The pass attempts data is extracted from detailed player statistics throughout one full season, with the percentage of each foot is given by the following formulas:

\[
\text{Left-foot Pass \%} = \left(\frac{\text{Left-foot Pass}}{\text{Total Pass}}\right) \times 100 \%
\]

\[
\text{Right-foot Pass \%} = \left(\frac{\text{Right-foot Pass}}{\text{Total Pass}}\right) \times 100 \%
\]

Specifically for the data analysis, the datasets are drawn from multiple pages in transfermarkt.com and FBref (2021-2022 Big 5 European Leagues Stats, 2022) by either manual copying as well as scraping, i.e., a technique to extract the data contained in a webpage. Then, the raw data was processed using Python and R programming language on a shareable notebook on Google Colab (colab.research.google.com). Visualization of those processed data helps the reader to understand and perceive the characteristics of the observed variables and their relation to each other.

3. Result

The recapitulation of data drawn from transfermarkt.com pertaining to the relation of left-footedness and market values of leagues is tabulated in Table 2. For ease of interpretation, however, it is visualized in Figure 1 as a scatter plot. Presenting the table as well as the scatter plot in this section does not pose any redundancy as both of them will be useful for the readers to digest the data analytics logic discussed in the subsequent section.

<table>
<thead>
<tr>
<th>Country</th>
<th>Average team market value (USD mio.)</th>
<th>Left-footed goals (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>England</td>
<td>516</td>
<td>39.8</td>
</tr>
<tr>
<td>Spain</td>
<td>241.5</td>
<td>39.4</td>
</tr>
<tr>
<td>Germany</td>
<td>238.3</td>
<td>36.9</td>
</tr>
<tr>
<td>Italy</td>
<td>226</td>
<td>40.2</td>
</tr>
<tr>
<td>France</td>
<td>167.5</td>
<td>36.4</td>
</tr>
<tr>
<td>USA</td>
<td>38.8</td>
<td>34.7</td>
</tr>
<tr>
<td>Mexico</td>
<td>45.1</td>
<td>39.5</td>
</tr>
<tr>
<td>Brazil</td>
<td>66.26</td>
<td>38.2</td>
</tr>
<tr>
<td>Argentina</td>
<td>27.9</td>
<td>35.3</td>
</tr>
<tr>
<td>Morocco</td>
<td>7.9</td>
<td>34.3</td>
</tr>
<tr>
<td>Japan</td>
<td>15.2</td>
<td>37.7</td>
</tr>
<tr>
<td>South Korea</td>
<td>13.3</td>
<td>33.3</td>
</tr>
<tr>
<td>Australia</td>
<td>8.2</td>
<td>35.4</td>
</tr>
<tr>
<td>Indonesia</td>
<td>3.9</td>
<td>33.6</td>
</tr>
<tr>
<td>Thailand</td>
<td>4.6</td>
<td>32.7</td>
</tr>
</tbody>
</table>
In Figure 1, the horizontal and vertical axes represent the level of left-footedness and the indicative value of league, respectively. Note that the vertical axis is in logarithmic scale, while the horizontal one is linear. This logarithmic scaling style in the vertical axis is chosen considering the significant disparity between the lowest and highest valued league. Meanwhile keeping it linear would describe the extent of the disparity, it is not favorable for visualization as some of the low-valued points would be obscured due to their closeness to each other. For the interpretation of the chart in Figure 1, one may start with the understanding that the more to the right a data point is, the better-skilled player the corresponding league is, at least from the perspective of foot dominance balance. Likewise, the higher the point means the more valuable players are in a given league. Both variables then converge into the overall insights of player skill level, since the highly skilled players would expectedly have a well-balanced footedness as well as high market price.

**Figure 1. Scatter Plot Correlating Left-Foot Goal Percentage and Average Market Value of Leagues**

The distribution of passing attempts percentage with left- and right-foot, summarized from FBref.com, is visualized in Figure 2. The violin plots show that right-footed pass is at significantly higher rate with most of players have approximately 75% pass attempts with their right foot, shown in Figure 2.
The thick part of the violins indicates that the number of players with that rate of passing is heavily concentrated around that value of percentage. For example, the violin chart for Left-footed pass implies that most of the players (median value) passes with their left foot at around 15% rate. In addition, the thickest parts of violins lie around that median which signifies that the majority of all sampling point is indeed at around that median value. To evaluate the nominal numbers of pass attempts, the data is visualized in Figure 3.

![Pass Attempts Top 5 European Leagues 2021-2022](image)

**Figure 3. Comparison of the Number of Pass Attempts in Top 5 European Leagues (2021-2022)**

4. **Discussion**

The tendency shown in Figure 1 is self-explanatory in general, in the sense that the giant leagues are occupied with highly valued players, hence superior skills that include the ability to control and exert power with either foot (Jadczak et al., 2019). The leg’s muscle strength and kicking accuracy need to be optimized since they determines the success of passes and shoots (Firmansyah et al., 2021), thus, having both legs with balanced conditions would increase the chance and flexibility to take decisive actions. World’s highest-tier leagues, which are themselves the top 5 leagues in Europe, are in the upper-right regime of the chart. This position implies that they are a group where players score at a high rate with the normally-less-dominant foot, i.e., the left foot. Recall that the vertical (market value) axis is logarithmic, so the top-5-leagues group have a substantial gap in terms of valuation with the lower groups, see Table 2 for nominal values comparison.

A closer observation of the chart shows that there are countries’ leagues with left-footed goals percentage comparable to the top 5 leagues but at considerably low indicative market values, namely Brazil, Mexico, and Japan. Those countries are considered talent feeders to European leagues, notably Brazil (Lago-Peñas et al., 2019), as a long-standing fountain of exceptionally talented players. In the past decades, Japanese football has also substantially risen, with players aspiring to play in Europe (Yoshio & Horne, 2004). The serious development of football in Japan in the past 50 years (Taylor, 2006) is proven to be fruitful, as observed today, where many Japanese players have made their way to Europe (Orlowitz, 2022). As for Mexicans, their presence in European leagues is predominantly in Spain, where they speak their native tongue. However, when looking at the players migration distribution, a huge number of Mexican players are pursuing careers in the USA (Kerr, 2019). It can be linked to the socioeconomic and geographic proximity factors that are beyond the scope of this study.
Interestingly, Argentina, the 2022 World Champion does not show any outstanding tendency. An argument can be constructed by keeping in mind that the chart does not describe the national team, but the local league. That being said, it is understandable that a league is not necessarily reflecting the skills of players holding citizenship of the corresponding country. There main cause is that the country’s best players have left home for Europe. Speaking of Argentina, the FIFA World Cup 2022 champion squad consists of almost entirely European major leagues players. The players spend their prime form in Europe and only play at home during their junior and near-retirement period. Hence, European leagues benefit from those players’ peak performance during their golden age.

Similar phenomena to Argentina exist in the case of Morocco, where its position in Figure 1 leads to a view of being a lower-mediocre league, meanwhile the national team performed excellently at the World Cup 2022. A deeper look into the players’ profile, the line-up predominantly contains those who play abroad (Morocco - Detailed Squad 2023, 2022). Even more, several names that filled the regular starting squad were born and raised abroad without any experience in the domestic Moroccan professional league. Thanks to the strong diaspora tradition (Wagner, 2019) as well as linguistic advantage as a former French colony and a close tie with Spain, Moroccan-descent players can be found all over Europe bearing citizenship of European countries, Morocco, or dual. The dual citizenship policy is in their favor here (Seiberth et al., 2019), since it allows talented footballers bearing other countries’ passports, mostly of European countries, to also have Moroccan citizenship without ditching their European citizenship. That way, they can continue to gain the benefit of being European players and at the same time opportunity in the national team when Morocco calls. To this end, again, national league quality does not necessarily represent the skill level of players with the corresponding nationality, if that country has many talents playing abroad in more competitive leagues. A past study confirms this argument, stating that the migration of players to foreign leagues positively affects the performance of their home country’s national team (Allan & Moffat, 2014).

With a comparable player foot laterality percentage, the US league is seemingly much more expensive than the Australian is. In those countries, football, or soccer as they say, is not the most popular sport. Yet, the sport has stronger business and entertainment industry interest in the US (Bradbury, 2021). South Korea, although having an average league market value on a par with Japan, it is lagging in terms of laterality. Again, a quick internet survey reveals that many Korean talents compete abroad with a significant number of them joining Japanese clubs.

As for the ASEAN sides included in Figure 1, they are sitting at the low price, low foot balance, regime of the chart. A straightforward, or intuition based, reasoning can describe the fact that Indonesia and Thailand are still requiring revolutionary advancement to catch up with the world class football societies. Unlike the Morocco case, with the limited presence of ASEAN players in foreign top leagues, the performance of the national teams mimics the league quality since the national squads are mostly filled with those playing in national leagues.

Figure 2 and 3 both confirms the results of previous studies, as well as the rather common knowledge, that humans are predominantly right footed. Left-footed football players are rare, even in highly competitive professional sports arena. But for decisive actions, such as goal scoring event, the left-footedness rate is higher. It is also worth considering that players who are frequently taking decisive actions are those high-performing names, which separates them from the overall descriptive statistics of pass attempts in general.

5. Conclusion and Recommendation

This work has demonstrated the positive correlation between league quality and players’ foot laterality. That conclusion was drawn through the observation of the visualized data obtained from publicly
available, online crowdsourced platforms. As hypothesized, top leagues, with high monetary values, tend to have a high percentage of left-footed goals. This fact would reiterate the common knowledge that professional-aspiring footballers should work on their feet balance. The discussion also reveals that in the countries where many of their football talents pursue career in major world-class leagues, the quality of the domestic league does not necessarily reflect the performance of their national team in the international tournaments, since their high-skilled players do not contribute to local league, but they do to the national team.

References


