Finding the Importance of Home-field Advantage in Soccer Across the World

By: Will Hodes, Braeden Erickson, Vishal Chinthalapale

Introduction

Home-field advantage is a well-documented phenomenon in sports, where teams playing on their home ground often achieve better outcomes compared to away matches. There are many theorized confounding variables as to why home teams have this advantage. Whether it is because of travel, hostile fans, or unfamiliar environments, away teams are usually thought to be at a huge disadvantage. In soccer, the home advantage is particularly significant, influencing match results and team strategies. Despite its apparent impact, the specific factors contributing to home-field advantage and the extent of its influence remain areas of ongoing study.

Methodology

This study provides an in-depth exploratory analysis of home-field advantage in soccer, and how home-field advantage affects club soccer teams in different regions. To achieve this, we found a data set off of Kaggle that we linked above. The author of this dataset, Omri Goldstein collected it. The purpose of collecting this data is that he wanted to find a correlation between winning and being the home team.

This dataset includes data on the amount of games played and specifies how well each team did both home and away. Specifically, we will be using the goal differential statistics of each of the teams in both home and away games to compare each team's performance to see if they play better at home. Goal differential statistics specifically tell us how many goal the team is scoring compared to all of the teams they are playing. In other words, do they score more goals than their opponents, which is a good indicator of how well a team plays, so comparing the home and away goal differentials can tell us the difference of home and away team performance. We will also take into account each team's winning percentage for all of the home games, and all of the away games. The formulas as to how we calculated these formulas are shown below.

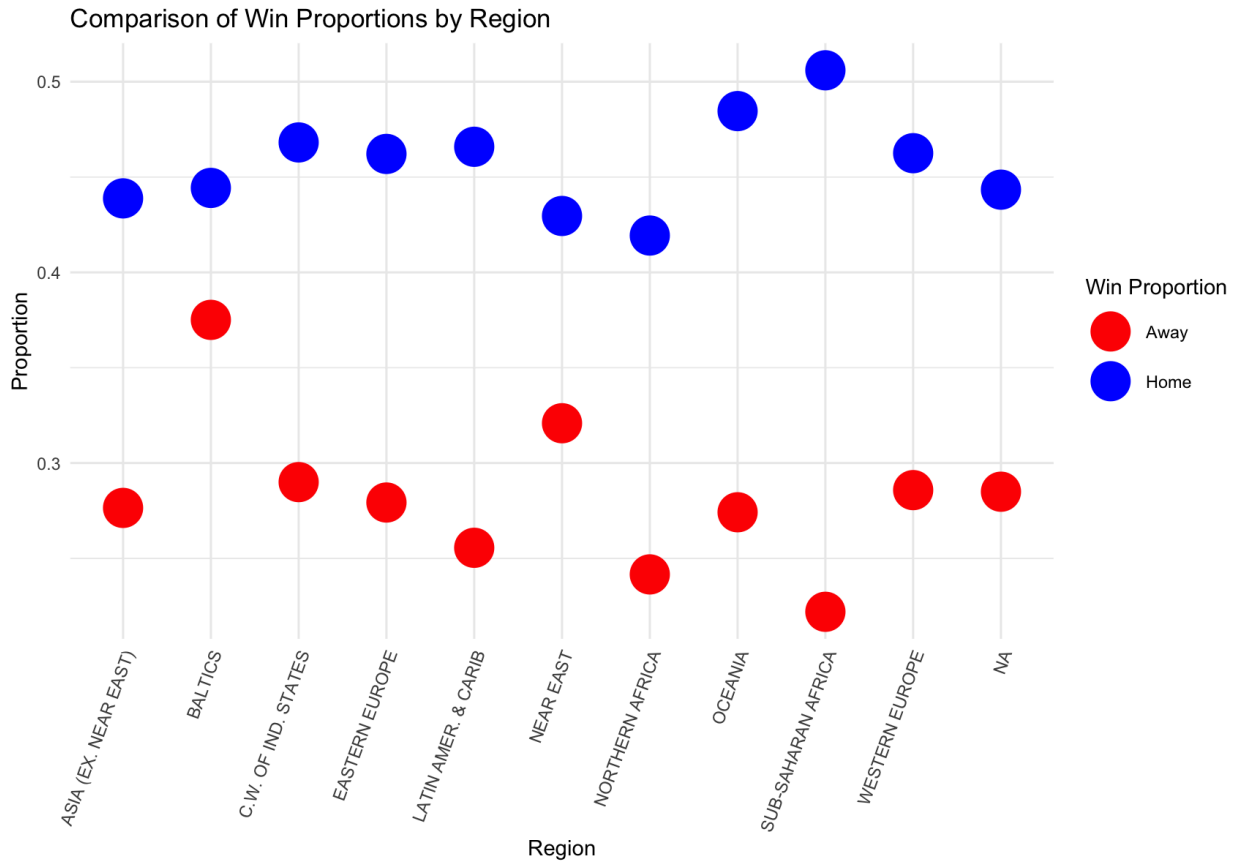
$$HomeWin\% = \frac{HomeWins}{HomeWins + HomeDraw + HomeLoss}$$

$$AwayWin\% = \frac{AwayWins}{AwayWins + AwayDraw + AwayLoss}$$

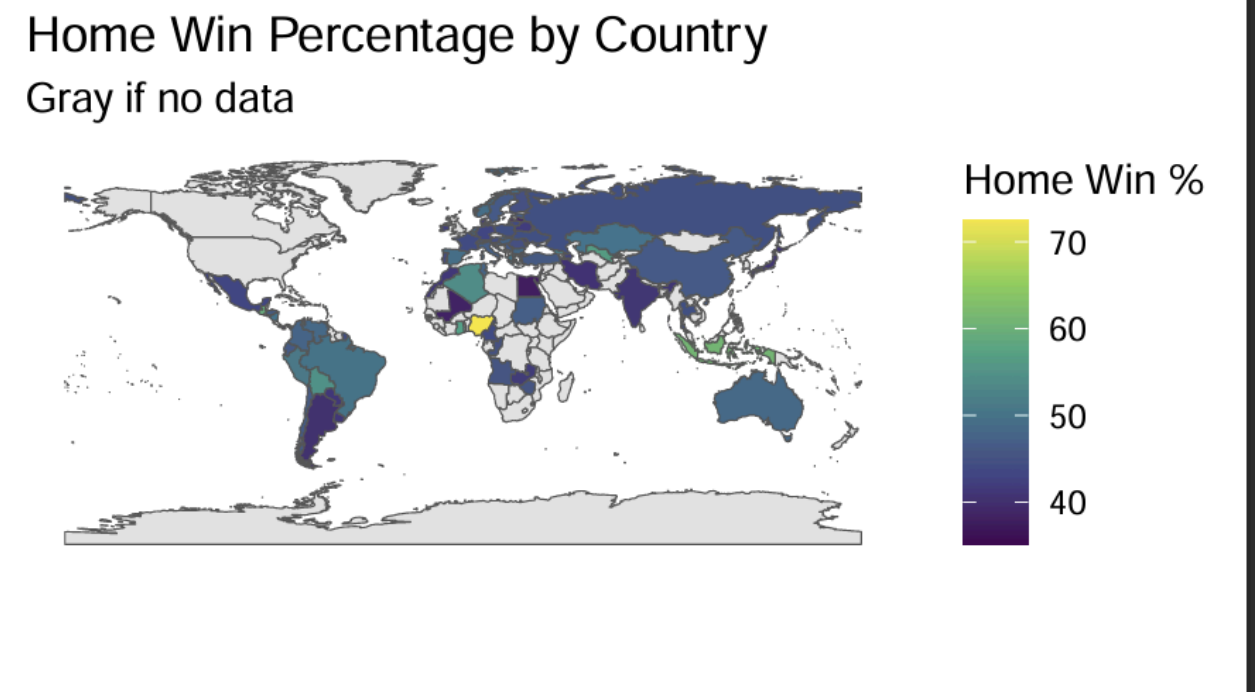
This data also Meets all of the FAIR principles, which stands for Findable, Accessible, Interoperable, and Reusable. The dataset is easily accessible on google, and can be easily downloaded, which satisfies the first three principles. We can also reuse the data because it is open to the public, therefore, it meets all of the FAIR principles.

In addition to answering the main question “Does Home Field Advantage meaningfully affect the outcome of games?” We also want to know the different effects geography has on Home-field advantage. Luckily, the same creator of the dataset mentioned previously also made another data set that lists certain facts about each country, and groups them into regions. In our data wrangling process, we merged these two datasets by country, so that each team has data on their performance, and the region their country is in. This regional data is especially useful when trying to ascertain the impact of Home Field advantage on geography because we can aggregate both the goal differential and winning percentage statistics to find how effective the advantage is in each region.

To find the effectiveness of Home Field advantage, we asked ourselves several smaller questions. These include, How does home-field advantage affect the probability of winning for the home team? When we look at the data we want to see if home-field advantage leads the home team to a higher winning percentage. In which countries is the home field the strongest? We want to know which countries do the best when given home-field advantage. In which countries is home field advantage the weakest? We made several graphs to help us answer these questions.

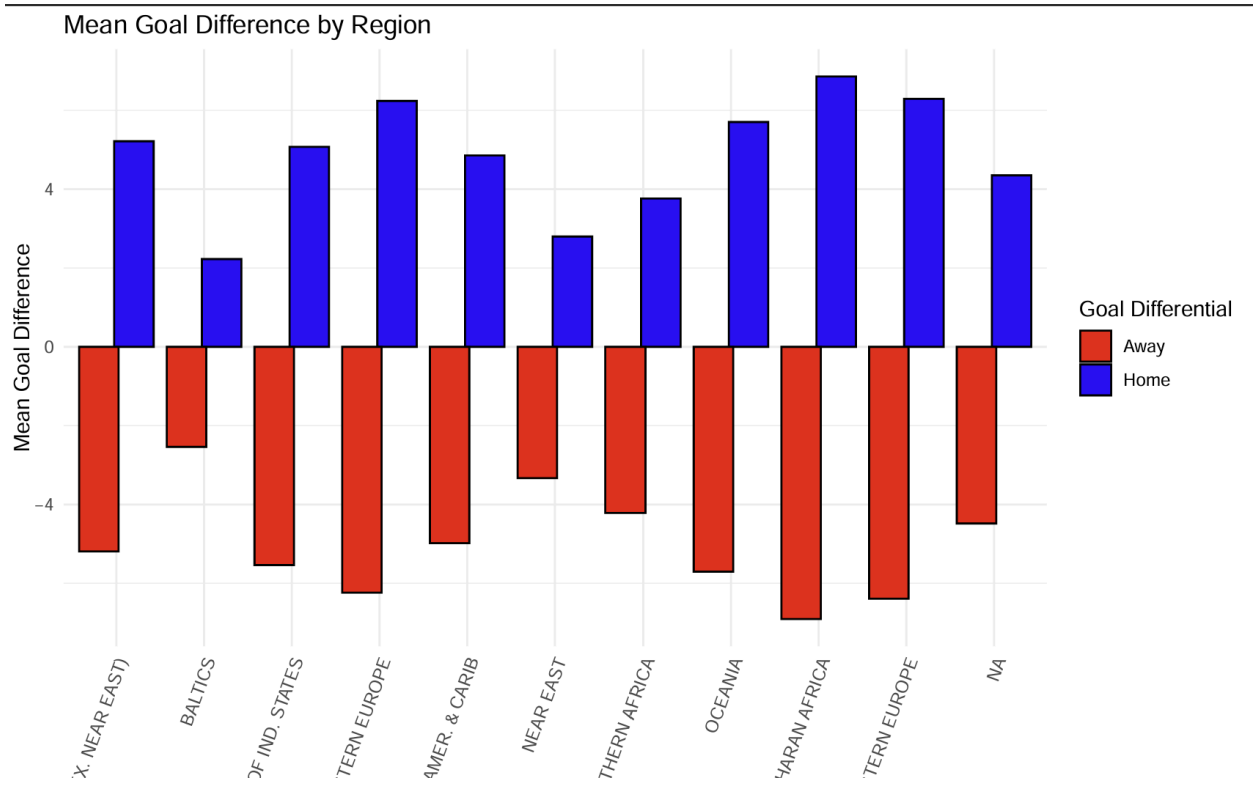


The first data visualization represents the home and away win proportion based on region. Based on the graph, we can see that there are multiple trends that are quite noticeable. We can see that countries that are in the regions of Sub-Saharan Africa, Oceania, and Western Europe, have the highest win proportion when they play at home. Some regions that have a low win percentage at home include the baltics, and Northern Africa. We also notice that there is a stark difference between the home win proportion, and the away win proportion. The home win proportion is vastly better, so much so that the highest away win proportion (the Baltics) is lower than the lowest home win proportion (Northern Africa).

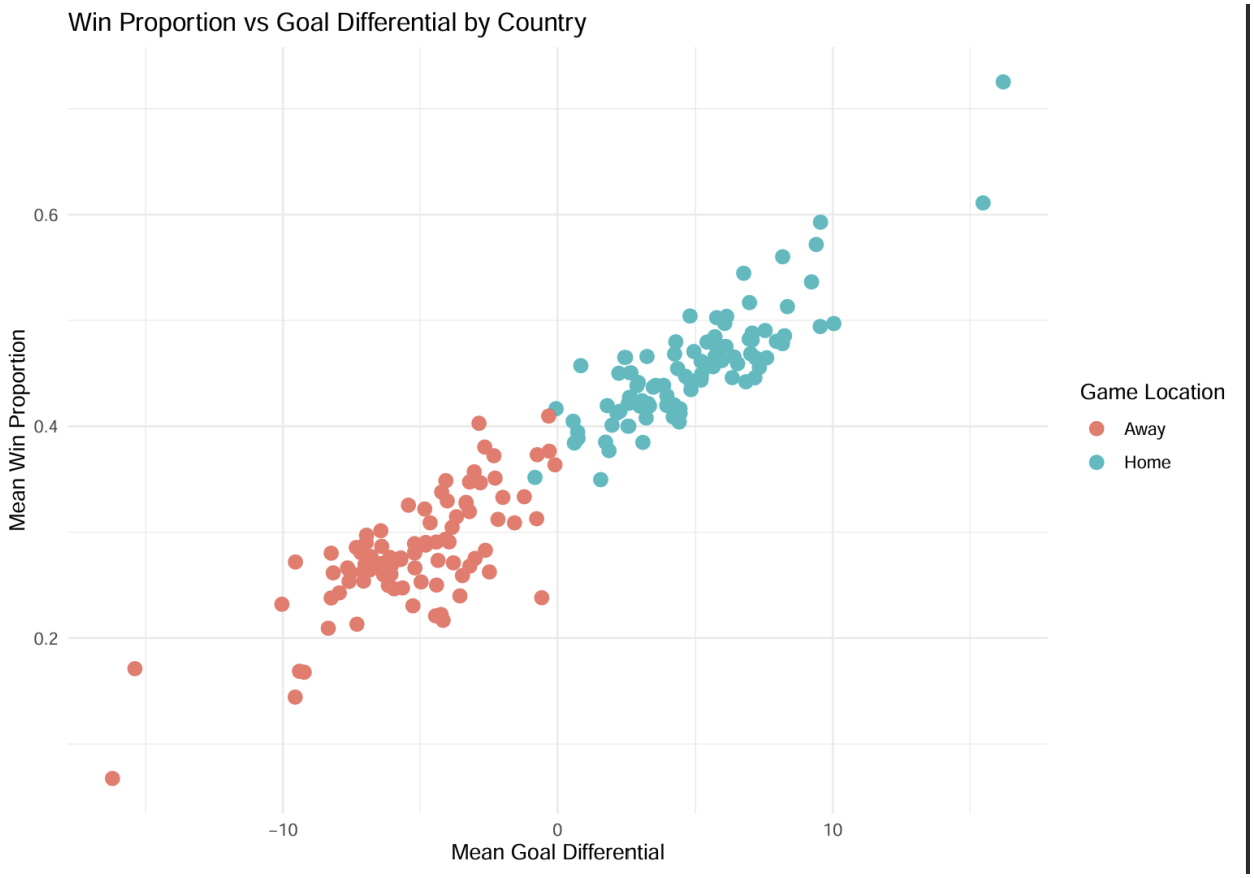


Another part of our research question includes the home win percentage by country. This data visualization below perfectly explains and displays one of our research questions. Taking a deeper dive into the visualization, we can see that there are a few countries that have a very high win percentage when they play at home. Some countries that have a high win percentage are located in southeast asia.

Countries with a lower win percentage at home include countries in western and Eastern Europe. We can see a trend in which countries with lower populations tend to win less games at home rather than playing away.



The next visualization demonstrates the mean goal difference while playing at home versus away. Looking further into it, we notice that there are a few regions that stand out when it comes to mean goal difference. For example, regions like western Europe and Sub-Saharan-Africa have a higher mean goal difference when they play games at home. Something interesting about this is that, Western Europe and Sub-Saharan Africa have the lowest mean goal difference when playing games away. That is definitely something that catches our attention when viewing this.



In this next visualization, each red dot represents a singular countries away statistics and each blue dot also represents a singular countries home statistics. The combined statistics show an obvious difference between the two. At home nearly all countries have a higher goal differential and winning percentage than even the best countries have away.

Conclusion

Our exploratory data analysis of home-field advantage, which is a phenomenon that exists in sports where sports teams tend to win more games at home rather than away, led us to many interesting points. We figured the best way to explore this idea was to take data from the most popular sport in the world, soccer. We were able to find a dataset with data from over 8000 seasons of soccer from teams all over the world and we used that data to learn about home-field advantage. We combined the soccer data to another dataset with information about the countries that the teams played in to explore ideas that in different regions of the world, home-field advantage is different.

Based on our analysis, we can see that teams have a higher winning percentage at home especially in densely populated countries, which include countries in regions such as Sub-Saharan and Western Europe. Our most important statistic was home winning percentage, which represented the proportion of times a team wins when they play at home. We compated this value to the away winning percentage and across the board, teams won a higher proportion of games at home. Our analysis can be used to dig deeper into the factors that cause home-field advantage to exist in sports in future studies.

CITATION

Gilermo, Dr. "Home Advantage in Soccer and Basketball." *Kaggle*, 2022, https://www.kaggle.com/datasets/drgilermo/home-advantage-in-soccer-and-basketball. Accessed 18 Dec. 2024.