FIFA World Cup Data Analysis

DS 311-01 Technologies in Data Analytics

Instructor:Normal Lo

Team-ds

Executive summary:

The FIFA World Cup is the most prestigious football tournament in the world. The championship has been awarded every four years since the start of the tournament in 1930. The current format involves a qualification phase, which takes place over the preceding three years, to determine which teams' quality for the tournament. In the tournament, 32 teams, including the host nation, compete for the title at different stadiums in the host country.

The reigning champion is France, which beat Croatia in the 2018 tournament in Russia. Qatar will host the 2022 tournament, for which the first match will be played in November. This dataset provides a complete overview of all international soccer matches played since the 90s. On top of that, the strength of each team is provided by incorporating actual FIFA rankings as well as player strengths based on the EA Sport FIFA video game.

Our findings have answered some of the questions that team members wanted to know at the beginning of the project which are:

- ➢ Is there really such a thing as home team advantage?
    - Do we find any evidence of home team advantage?
- ➢ Do teams with stronger offense players score more goals? And do teams with stronger goalkeepers receive fewer goals?
    - Use the data to support your hypothesis
- ➢ What team has the strongest defense, midfield, and offense players?
    - Can we quantify those statistics?
    - How we define strongest?

Introduction:

In the FIFA World Cup dataset, the records indicate all international soccer matches played since 1993 up to 2022 with its date, home and away team name, continent, rank, fifa point, score of the game, name of the tournament, name of the country that the match was played on, results, team's offense,defense,and midfield scores. These data can eaily help me revel the team with the most victories and defeats, and if the player's score have any effect to a game's result.

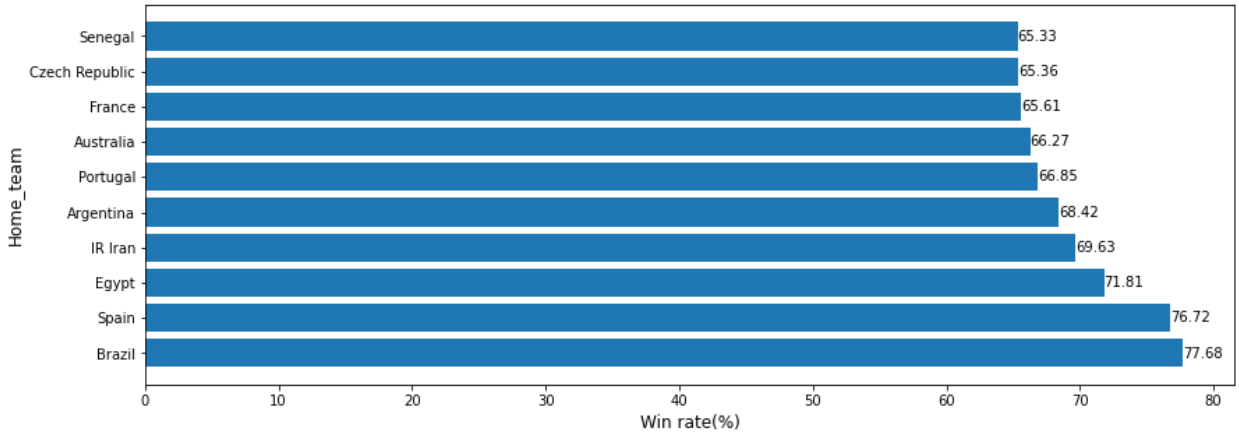
Exploratory Data Analysis:

This report will go through all three of the interesting exploratory data analysis questions that we wish to know at the beginning of the project. We were able to use the dataset that was provided to build differen visualization to support our finding by using Python in Jupyter Notebook and Tableau.

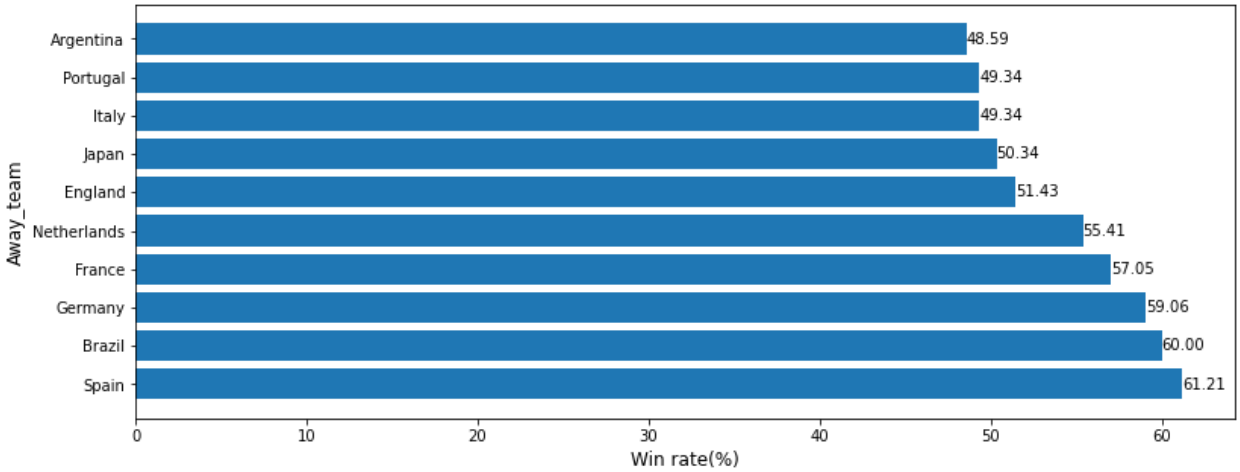
The first question we wanted to answer is "Is there really such a thing as home team advantage?"When I first saw this question I thought that home-team advantage was not exist because it was just a change of field for the players to play and was that really that big of deal for just playing in a different stadium?, but after I saw some data I realized that the home-team advantage was true.I first looked at the top ten win rate of home and away FIFA teams and then I found that for example, France, at home game their win rate is 65.61% but on the away game their win rate goes down to 57.05%, and

there is almost 8.5% difference.Other then that, we can also look at the average win rate for both top 10 teams, for Home team, they have 69.47% and for away team their average win rate is 54.17% Overall, we see that the highest win rate of the home-team, Brazil has 77.68% while the highest win rate of the away team, Spain is only 61.21%. This difference is almost 16%, and these data means a lot to the question we want to verify.

Top 10 home-team win rates:

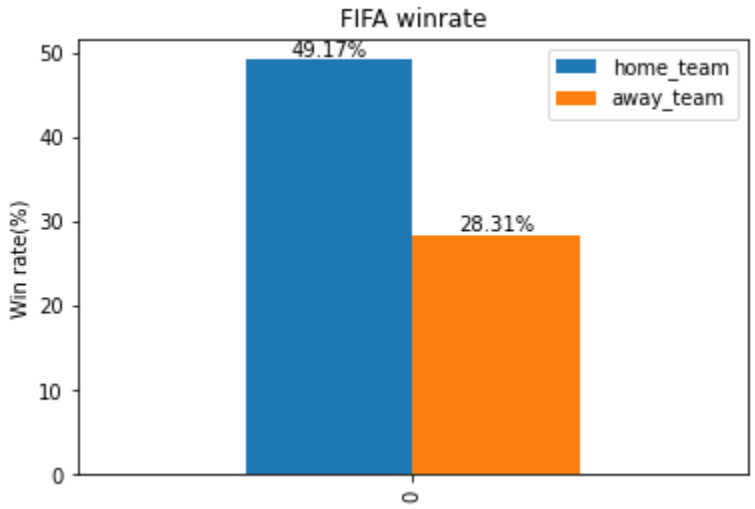


Top 10 away-team win rates:

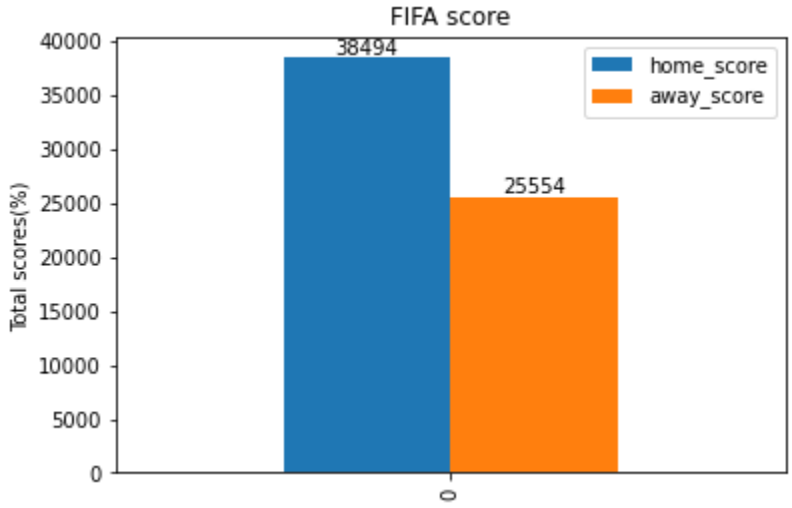


Next, I decided to check the total winning percentage and total goals scored by both sides. Throughout FIFA history, the home team had a 49.17% winning percentage while the away team only had a 28.31% winning percentage (the rest were draws), and the home team scored 38,494 goals while the away team only scored 25,554 goals. This difference is very large and this allows me to summarize the problem.

Total win rates:

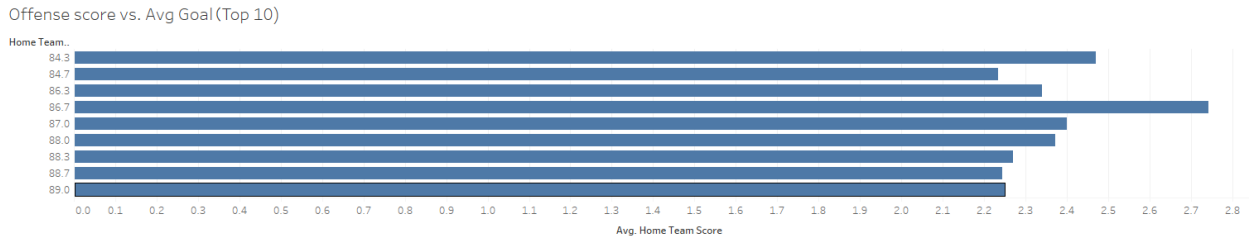


Total scores(Goals):



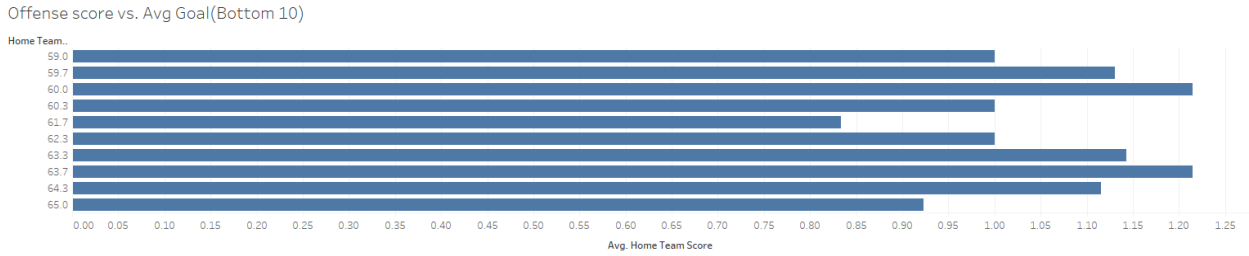
Our answer to the first question is a simple yes. With all this data, it's hard to argue that there is no such thing as a home team advantage, because when a team plays as the home team, they do win more and score more goals than when they play on the away. Of course, the home team advantage is not limited to the venue they play in, there are factors such as weather, player condition, etc. that can contribute to the home team advantage.

The second question we want to answer is “Do teams with stronger offense players score more goals? And Do teams with stronger goalkeepers receive fewer goals?” In this question, the measure of strong offense and defense player team will be the team’s offense and defense score. With this score, we thought it would be the best to match it with the average score of the matches. Here is what we found:



This chart above represents the top 10 home teams with the most player offense points and scoring averages, and from which we can see that they are all above 80 offense scores and their average goal is at least 2 and that means 2 goals per game.

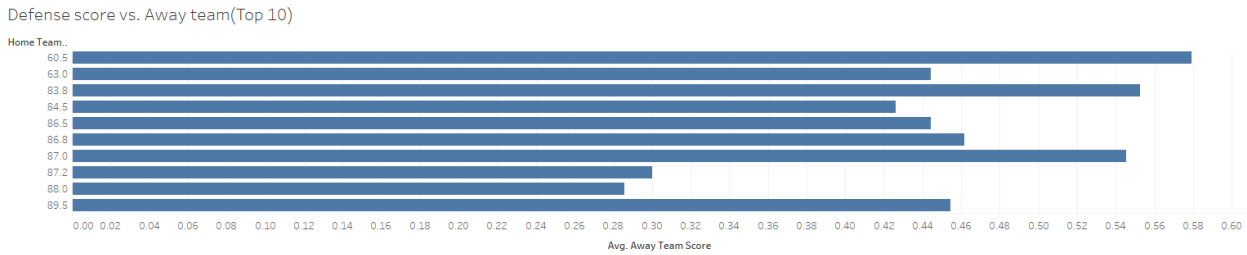
Then we look at the bottom 10 chart:



This chart represents the bottom 10 home team with the most player offense points, and all of their offense score is around 60 or below, and the average goal is

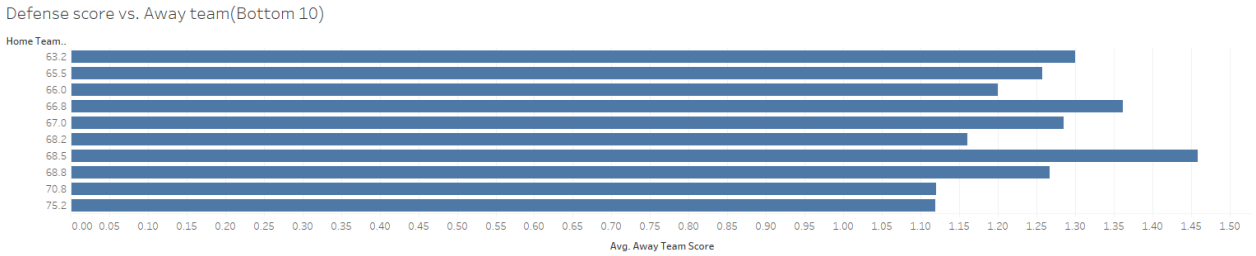
below 1. Comparing these two charts, it is quite easily to figure it out that with higher offense score could result with higher average goals per game.

Next, we look at the home team defense score vs. Avg away team score:



This is the chart of top 10 home team with defense score and its opponent's scoring averages. From that, there are two data that we didn't expected to see which are the score of 60.5 and 63.0 because rest of the defense scores of the top 10 are all above 80. All of the scores are below the average of 0.60 meaning they received less than 0.6 goal per game .

Now we look at the bottom 10 home team with defense score:



This graph is similar to the offense score graphs, where it all stay in a range of 60-70.

My answer to this question is that both offensive and defensive scores are certainly one of the most important factors in scoring more or less goals, but I don't think it's straightforward to conclude that higher offensive scores = more goals and higher defensive scores = fewer goals. For example, with the Guinea with defense scoring 60.5 points and Guinea scoring a total of 10 goals and receive only 1 goal in four different games, I checked the offensive scoring and defense of both teams and they were both close, or higher than Guinea. In this case, I believe there may be some other major factors that affect the strength of a team's offense and defense.

Conclusion:

It's a great experience to analyze some of the issues you want to know about on the sport you love. I've grown to love the sport in the process, and answering these questions allows a novice like me to know what is relatively crucial to the game of soccer. Finding the answers to the questions I wanted to know from scratch on the data provided showed me how important data is.

If we had more different data, we would be able to analyze the impact of different factors on a soccer game. For example, in the case of home advantage, if we had information about the number of local spectators, the weather, the name of the referee, etc., we would be able to use the information about these factors to give us a better idea of what causes home advantage and how to use it to win.