

11/11/22

MDA 620

FIFA WORLD CUP DATA VISUALIZATION

Capstone Project #1

Written by Jack VanDoran, Tom Flower, Zubair Hossain Mahamud

TABLE OF CONTENTS

1. Background
2. Objectives/ Goals of the project
3. Data Exploration (Coding in Python)
4. Data Visualization
5. Data Manipulation
6. Conclusions/ Recommendations
7. Bibliography/ References

Background:

The datasets used included all the primary information for every FIFA World Cup up to and including 2014. The rosters for every team from every World Cup, the year each tournament took place, host countries, the winners of each tournament and overall placements, total goals scored for each tournament and attendance numbers. With this information, you get a general scope of the popularity of the sport based on several key factors listed above. The higher the attendance, the more revenue produced so the teams participating, the cities where the games are played, and the number of goals scored all impact how large the crowds will be.

The key performance indicators (KPI's) are:

- 1) The quality of competition
- 2) The location
- 3) Recognition (teams) & Excitement (goals per game)
- 4) Which teams play as Home/ Away

Objectives/ Goals of the project:

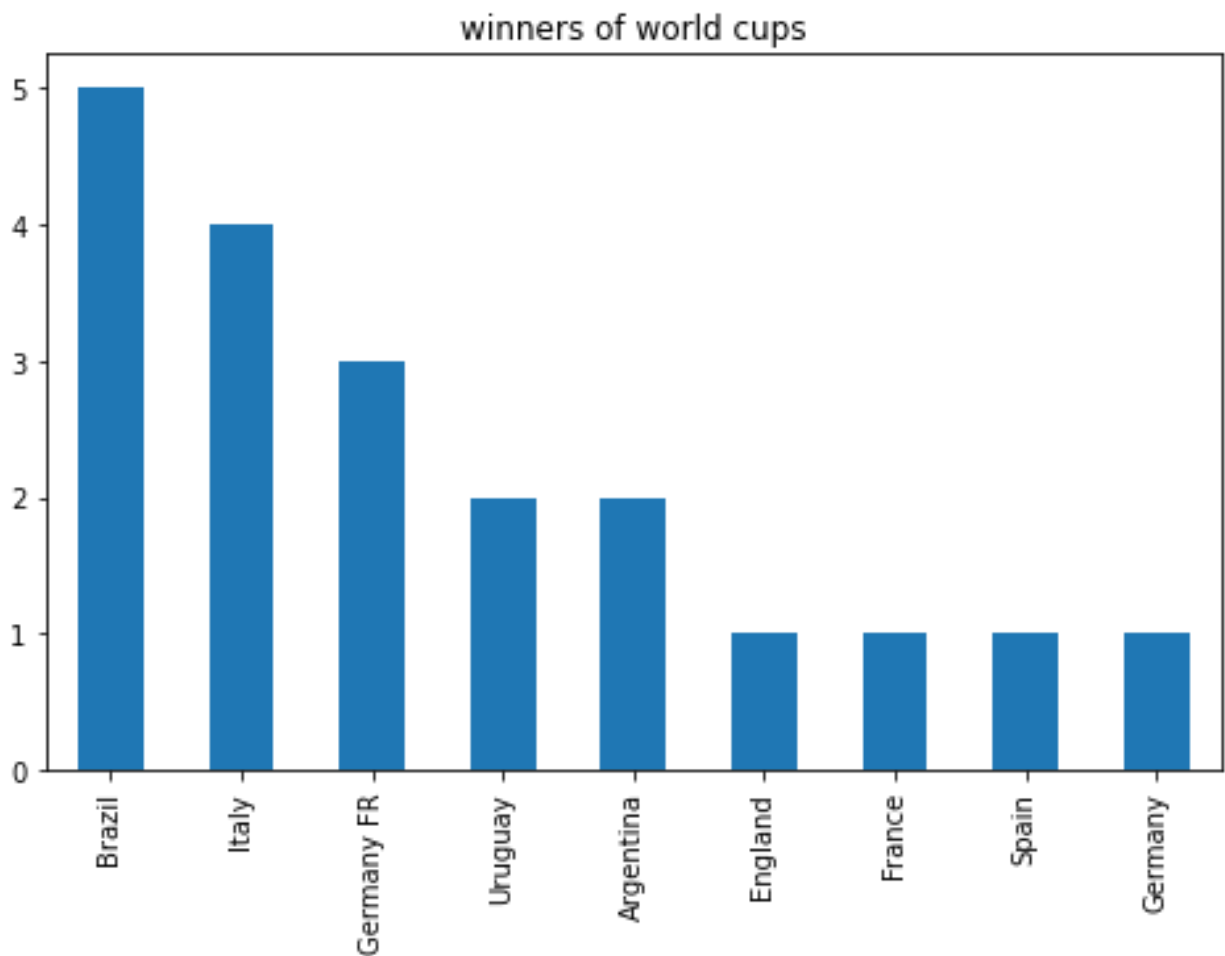
Unlike the common business/ sales breakdowns that usual portfolios have to offer, our group has decided to tackle the sports aspect directly. Our data set (<https://www.kaggle.com/code/nathanlauga/data-visualization-of-fifa-world-cup/notebook>) provides key insight on historical statistics that include goals per year, past champions, countries that participate, and most importantly attendance records. The objective for capstone project one is to focus on the key factors that drive attendance records while implementing scenarios that

could improve upon those numbers for future tournaments. As we explore, visualize, and manipulate our data using python coding we hope to grasp a better understanding of which factors influence higher attendance ratings. The group's hypothesis will support that countries that have solidified themselves as past champions will draw bigger crowds than teams that have not. Loyal fan bases always play a factor into packing a stadium. Teams such as Italy, Germany, and Brazil combine for over 12 international titles which can imply that these fanbases will travel well to the upcoming games this November in Qatar.

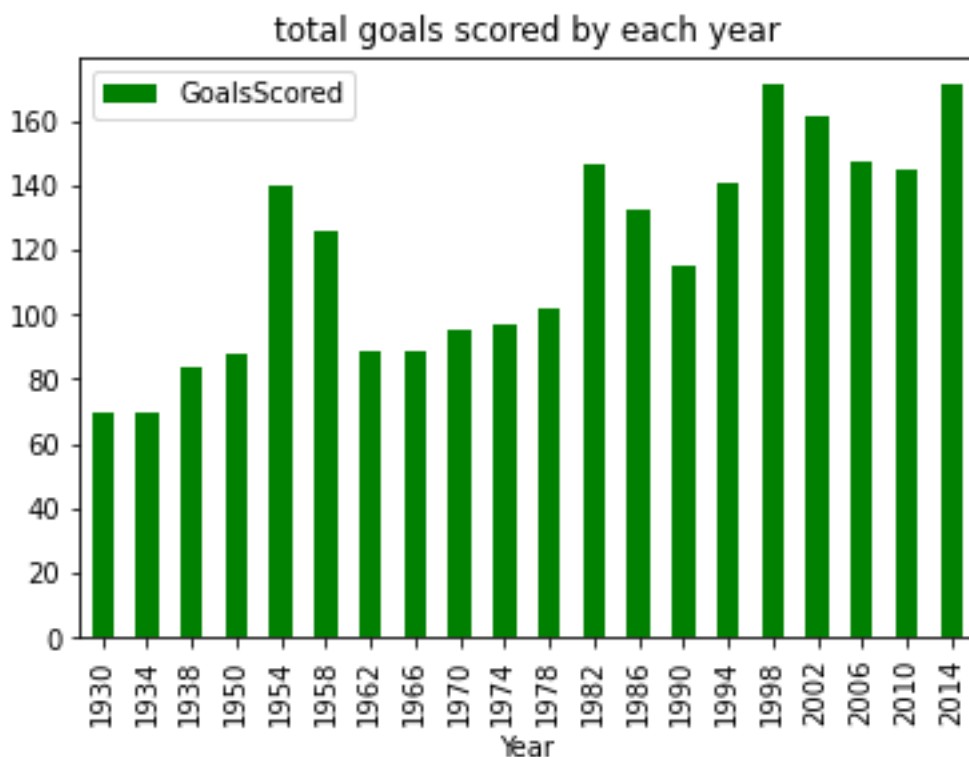
Data Exploration:

First, we imported the necessary packages and the datasets themselves. Next, we executed the head, tail, info, count, dtypes, and shape commands to get all the primary information from the datasets. This includes the first and last couple of lines of the datasets, the generic material, the datatypes of each variable, the counts of each variable, and the overall shape of the data frame (number of rows and columns). Additionally, we checked for any missing data using the null function which showed the dataset was complete. After viewing the outputs from all these snippets, you get a general idea of what the dataset includes and the type of information you're looking at. From the exploration, you can make several inferences about the variables and how they connect to our focus of attendance.

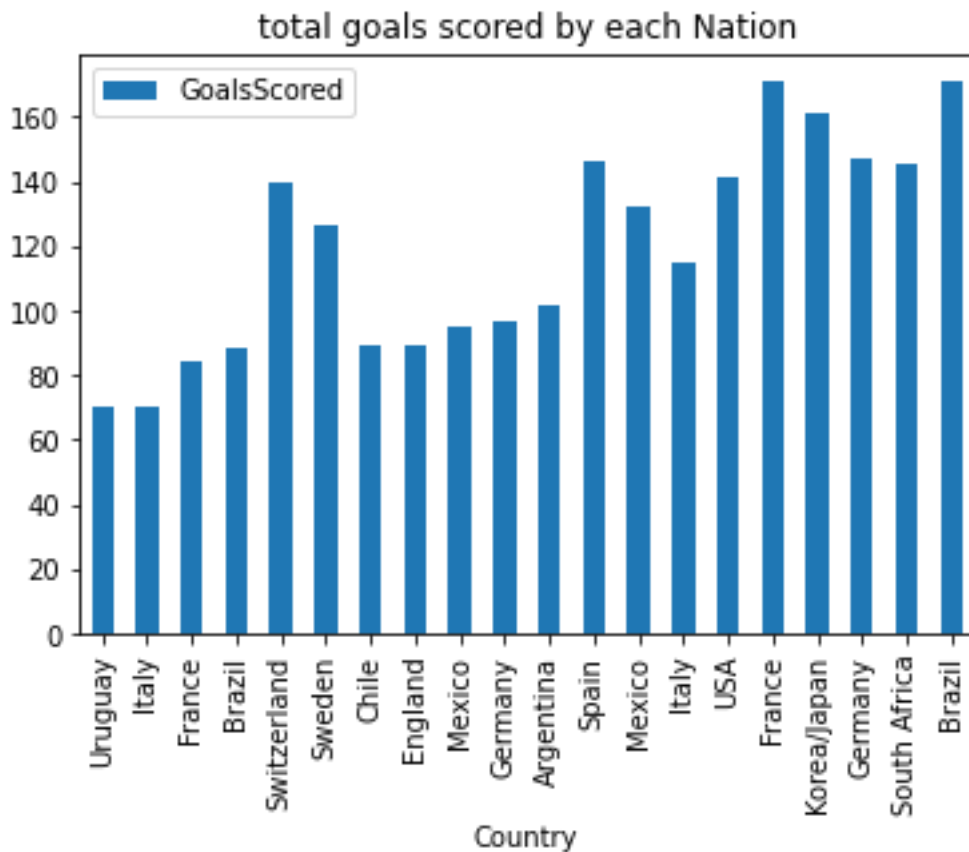
Data Visualization:



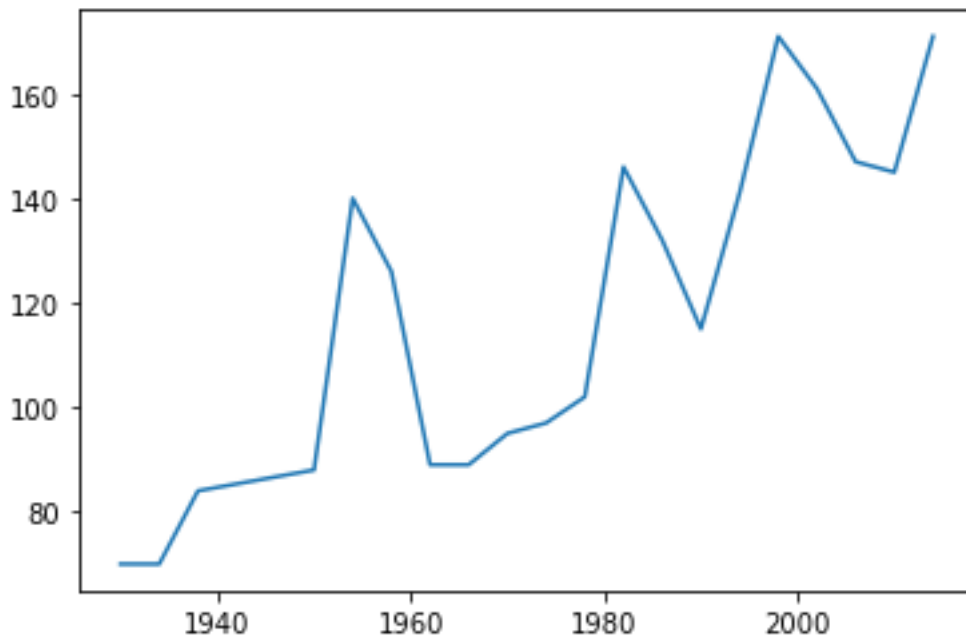
This chart is a visualization of all the countries that have won World Cups in the past and the quantity of each country. As you can see, Brazil (5), Italy (4), and Germany (3) have tallied the most titles showing they are consistently competitive and provide high quality on the field.



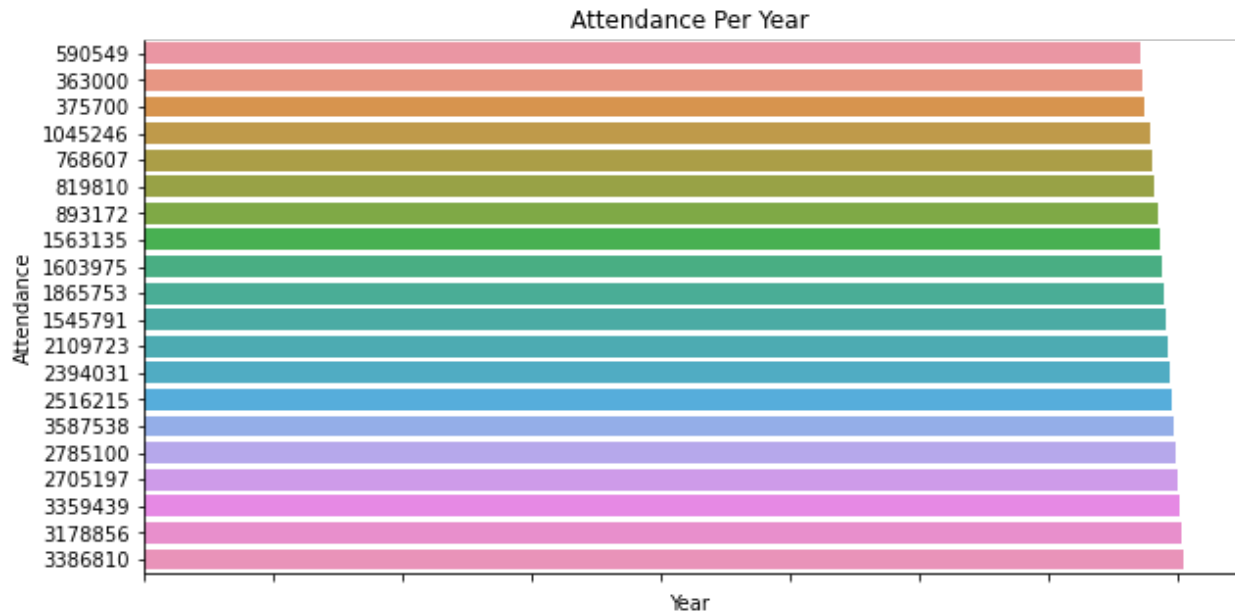
This is a visual of the total goals scored each tournament in its entirety. It includes every team that participated in each tournament. You can see a gradual upward trend. There are of course a couple of fluctuations, however, over time you can see the total's going up from 1930-2014. The implication being the style of play has changed to be more offensive minded and the quality of talent per team has improved. The more popular the game has become; the more developed players and strategies have become as well.



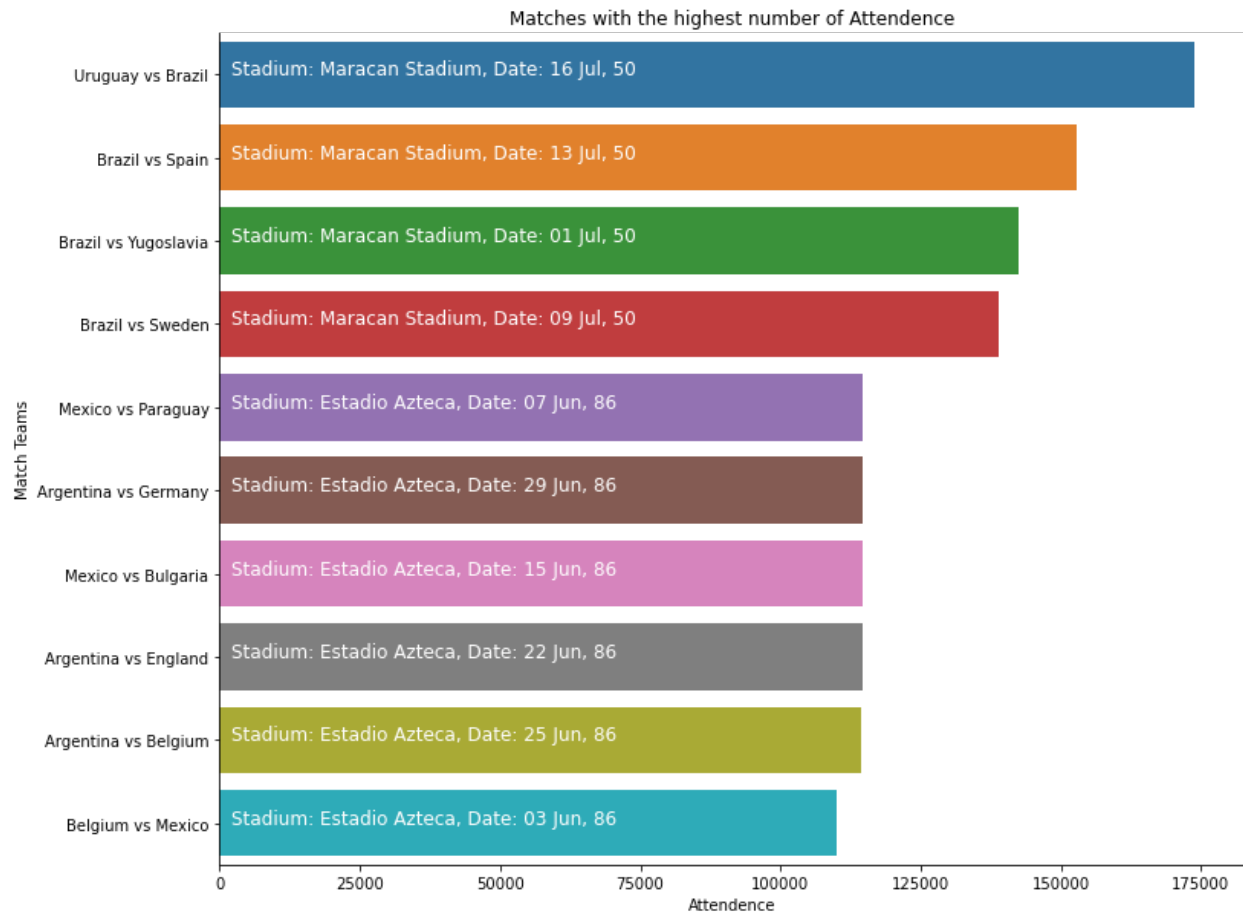
After calculating the total goals scored per tournament, the focus shifted to breaking that down per team. This shows the total amount of goals scored per nation over the course of World Cup history(20 tournaments). The teams with higher outputs show consistency because they have qualified for the most tournaments, scored more goals per game at a reliable rate and made it deeper into each World Cup. Brazil, France, Spain are some examples of nations that meet these criteria.



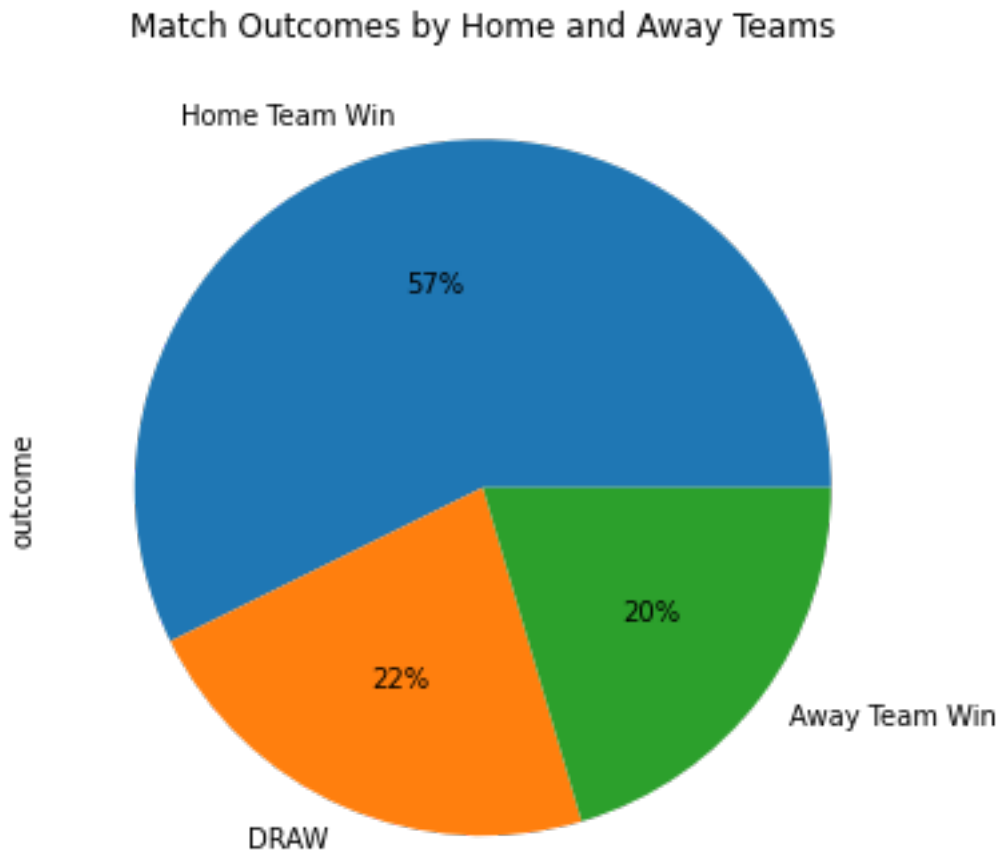
This visualization represents the same information as the total goals scored per tournament chart in a different form. This allows a free-flowing conception of the information as opposed to stiff bars. Where the other chart is based on the exact years the tournaments took place, this goes through the gaps to paint an easier picture to follow.



This shows that attendance per tournament has steadily increased over time since its inception. In addition to the key factors we've mentioned, time itself has allowed for the popularity of the sport to grow, which has resulted in attendance numbers increasing each iteration of the World Cup. Host nations are building larger stadiums, teams with previous success have fan bases willing to travel to foreign countries in support, the talent of individual players has gone up and the overall excitement level has taken on a life of its own. Since the tournament only takes place once every four years, it has created a sense of rarity which in itself has created a sense of urgency to be a part of the experience as a fan both to see the tournament unfold and show that you were involved to some degree.



This chart expresses the individual matches with the highest attendance numbers over the World Cup's entire history up until 2014. As previously expressed, when highly competitive teams match up against each other, it makes for enormous crowds. Brazil and Argentina account for seven of the ten most attended games in World Cup history. These two nations have a combined seven titles between them and have bolstered star-studded rosters for decades. This exemplifies the importance that consistent success plays in the role of attendance in not only soccer but sports in general. The sports industry is a business after all, and people want to see the highest level of competition they can.



While World Cups don't exactly format themselves to the traditional home vs away structure (only the host nation is technically home), there is always a team that represents each side even if they are both visiting nations. Historically speaking, the team that represents the home side wins at a much higher rate than the away side at World Cups. This phenomenon intersects with the attendance numbers since the larger the crowd is for a team; the more support is felt and the more competitive they will tend to play.

Data Manipulation:

For data manipulation and cleaning, we established that certain names of countries were output in an improper format, therefore, they had to be replaced with their correct names. This was done for both data sets. Additionally, we created a new column labeled outcomes that calculated whether the home or away team won each match. This was not included in the dataset, and it felt noteworthy to include because it's a significant factor in the sports world especially when a grand scale such as the World Cup is being considered.

Conclusions/ Recommendations:

The group's hypothesis which was stated in goals for the project was upheld as we dissected the FIFA world cup data set. Teams that draw the biggest crowds are indeed the teams that not only score the most goals, but also have won the most titles. Brazil, a nation with 5 titles, has drawn the top 4 biggest crowds in FIFA history. A record breaking 175,000+ fans broke the previous record held by Brazil at just about 150,000 fans. It is safe to conclude that the correlation between winning and fan attendance is directly related.

An interesting statistic found from our data was 57% of match outcomes favored the home team. Considering that all teams are technically away from their home country to visit the host site, we did not think to find such data.

As soccer continues to expand its sports empire throughout the world we believe it is important to recognize that better offensive teams generate better attendance records. We can recommend that national teams build bigger stadiums in their home countries. The more the stadium can fit, the better chances that the home team will win. The 57% home v away record

will only continue to grow which in conclusion can help increase the probabilities of a FIFA world cup championship.

Bibliography:

<https://www.kaggle.com/code/nathanlauga/data-visualization-of-fifa-world-cup/notebook>