FIFA World Cup Rankings

Ryan Meston

Bellevue University

Abstract

The purpose of this project is to analyze FIFA World Cup rankings, player rankings, and game data to determine how may points a team needs to be ranked number 1 in the world. In this project we will also look at who the top teams have been and how consistent they have been at staying at the top. I also want to see how many teams have stayed at the top and why. This project is designed to give a manager/coach and staff the information they would need to evaluate player performance and team performance to see how they would need to preform on the filed to obtain the rank of 1 in the world.

Introduction and Background

This dataset contains information Rank, Country, Country Abbreviation, Total Points, Previous Points, Rank Change, Average Previous Years Points, Average Previous Years Points Weighted (50%), Average 2 Years Ago Points, Average 2 Years Ago Points Weighted (30%), Average 3 Years Ago Points, Average 3 Years Ago Points Weighted (20%), Confederation.

- rank – team rank.
- country_full – countries full name.
- country_abrv – countries name abbreviated.
- total_points – total points country received.
- previous_points – countries previous points.
- rank_change – countries rank change per week.
- cur_year_avg – current points average.
- cur_year_avg_weighted – current points weighted average.
- last_year_avg – last year points average.
- last_year_avg_weighted – last year points weighted average.
- two_year_ago_avg – two years ago points average.
- two_year_ago_weighted – two years ago weighted average.
- three_year_ago_avg – three years ago points average.
- three_year_ago_weighted – three years ago points weighted average.
- confederation – confederation the team belongs too.
- rank_date – date team was ranked respected ranking.

The questions that I would like to answer are as follows:

- Which country has the most championships?
- Which teams are the most consistent?
- Which confederation is strongest?
- How does team strength relate to player performance?
- Can we predict the next World Cup Champions?

Methodology

I started my early analysis off as exploratory to understand the dataset and gain insight in what I wanted to do so I could portray the data in a clean manner.

Below are the six confederations that are listed in the dataset I am using.

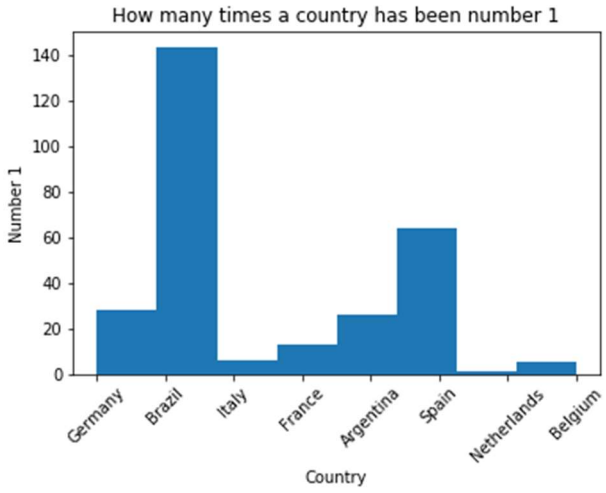
- UEFA 14933
- CAF 14876
- AFC 12617
- CONCACAF 9664
- CONMEBOL 2860
- OFC 2843

The figure below shows which confederation has had the most number 1 rankings and how many times that confederation has been ranked number 1.

- CONMEBOL 169
- UEFA 117

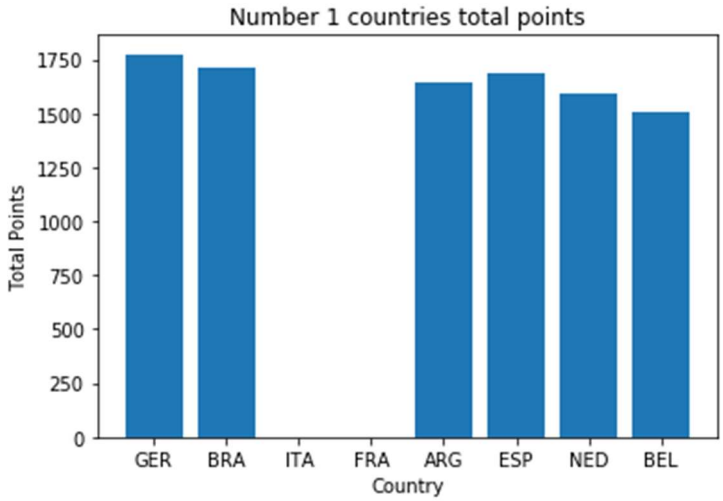
UEFA – The Union of European Football Associations has 55 national members. CONMEBOL – The South American Football Confederation is the oldest confederation, and they have 10 members.

The below chart shows how many times a country has been in the number one position in the world.



As we can see from the chart, Brazil has been the most consistent team with a number one ranking over 140 times. Brazil has always been known as a powerhouse on the world stage of soccer and the data shows that they truly are one of the greats that has performed on the field.

The below chart shows how many points the number 1 country has had in the 2018 season. Germany is the top team in the number 1 position during the 2018 season with Brazil and Spain closely behind them.



Below is showing the top five teams that have had home and away games. I wanted to do more with this data, but I did not have time to complete everything I wanted to complete with this. My goal was to combine both the home and away games played by team to get a true top 5 games played in the FIFA World cup to see who has played the most games in the FIFA World Cup. The below results are just showing the top five home and away games played and not a combine home and away total.

Home team games played:

- Brazil: 84
- Germany: 77
- Argentina: 60
- Italy: 49
- France: 40

Away team games played:

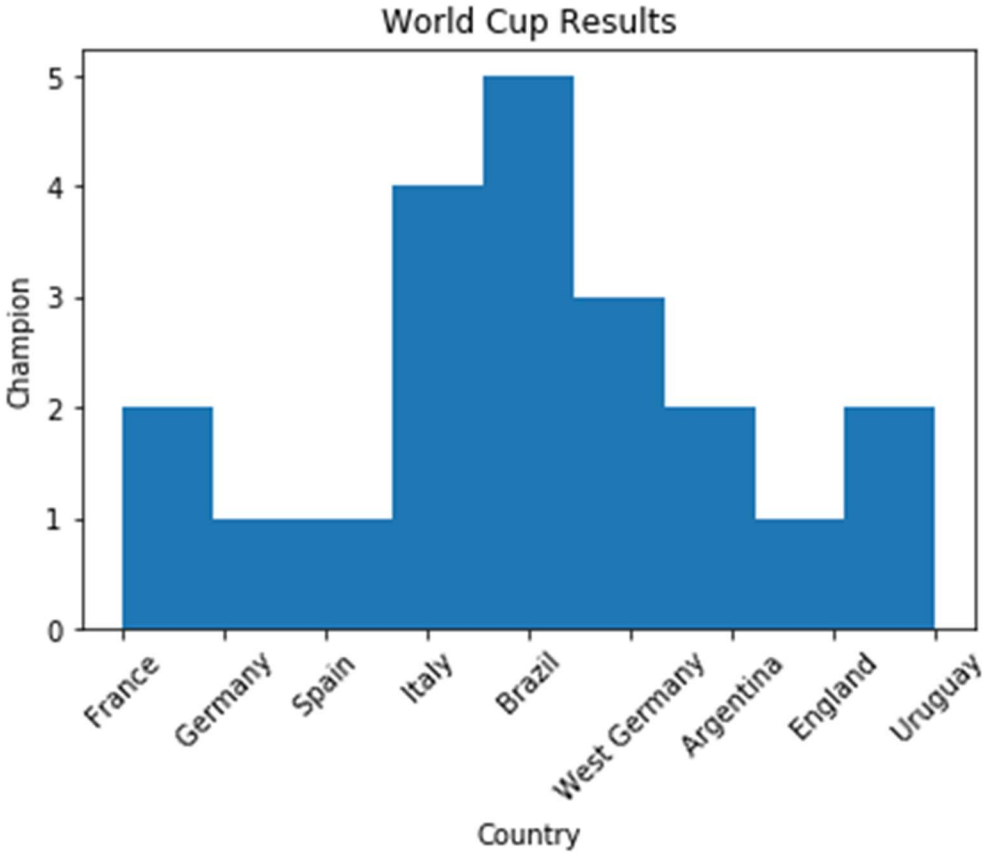
- Uruguay: 37
- Italy: 34
- Spain: 33
- England: 33
- Germany: 32

Data showing the Champion, runner-up, and third place finishes in all FIFA World Cups. The be low charts show how many times each team has placed in their respective finish.

Teams first place finishes:

- Brazil 5
- Italy 4
- West Germany 3
- France 2
- Argentina 2
- Uruguay 2

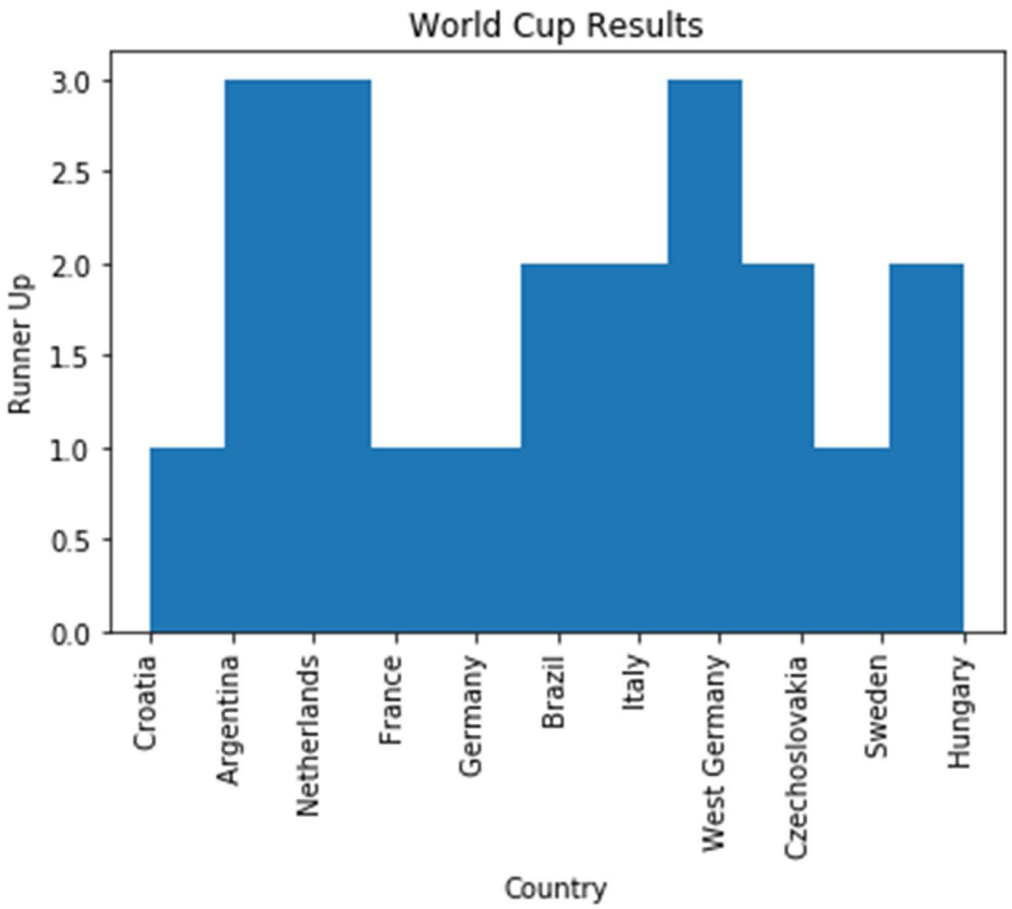
- England 1
- Germany 1
- Spain 1



Brazil has won the most FIFA world cups with 5 overall in the history of the FIFA world cup tou rnament.

Teams second place finishes:

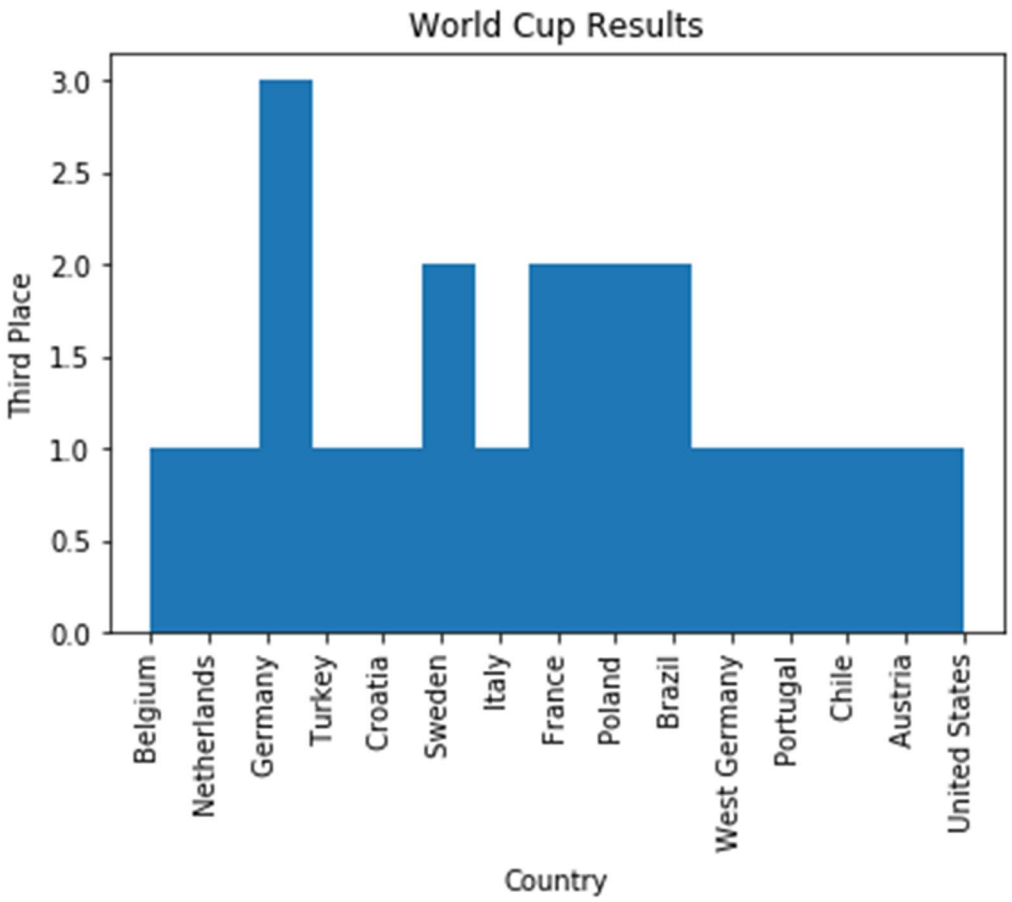
- Netherlands 3
- West Germany 3
- Argentina 3
- Italy 2
- Hungary 2
- Brazil 2
- Czechoslovakia 2
- France 1
- Sweden 1
- Germany 1
- Croatia 1



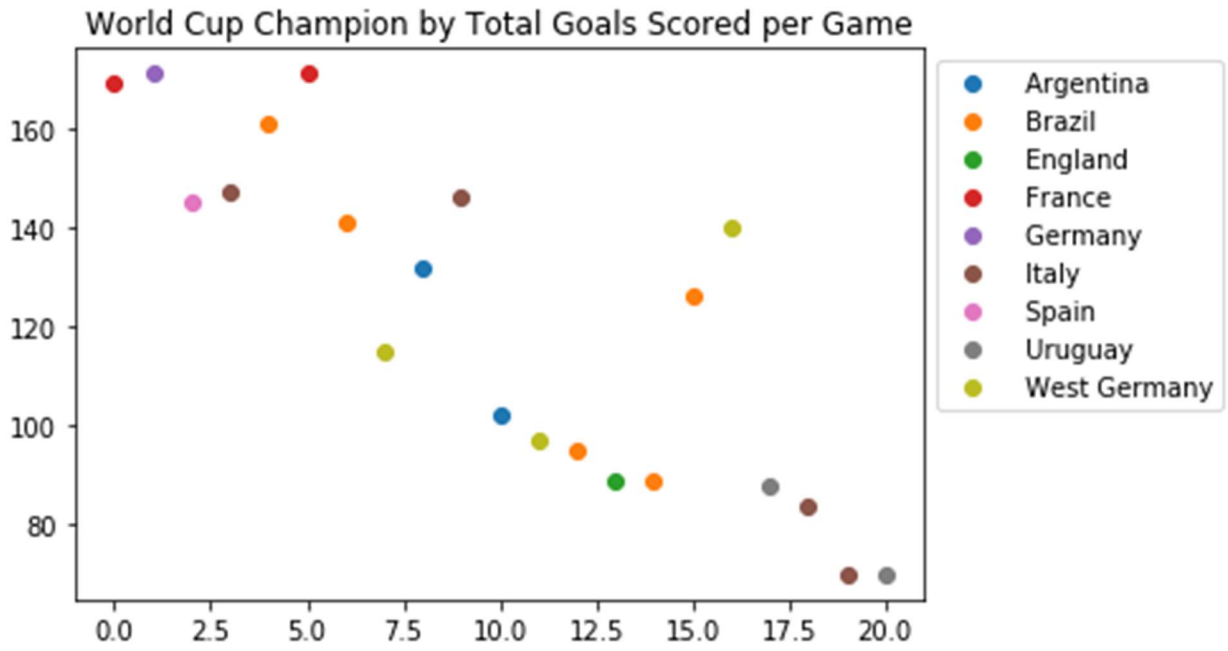
Netherlands has the most second place finishes in the history of the FIFA World Cup.

Teams third place finishes:

- Germany 3
- Poland 2
- Sweden 2
- Brazil 2
- France 2
- Chile 1
- Belgium 1
- Croatia 1
- United States 1
- Portugal 1
- Netherlands 1
- Turkey 1
- West Germany 1
- Italy 1
- Austria 1



Germany has the most third places finish in the FIFA World Cup. Brazil has shown their domin ance in the sport with 9 top 3 finishes in the history of the FIFA World Cup.



The above visual shows how many total points were scored in the championship game by the winning team.

Results

Below is a chart showing the which countries have been at number 1 in the world rankings and how many times they have been at the number 1 position.

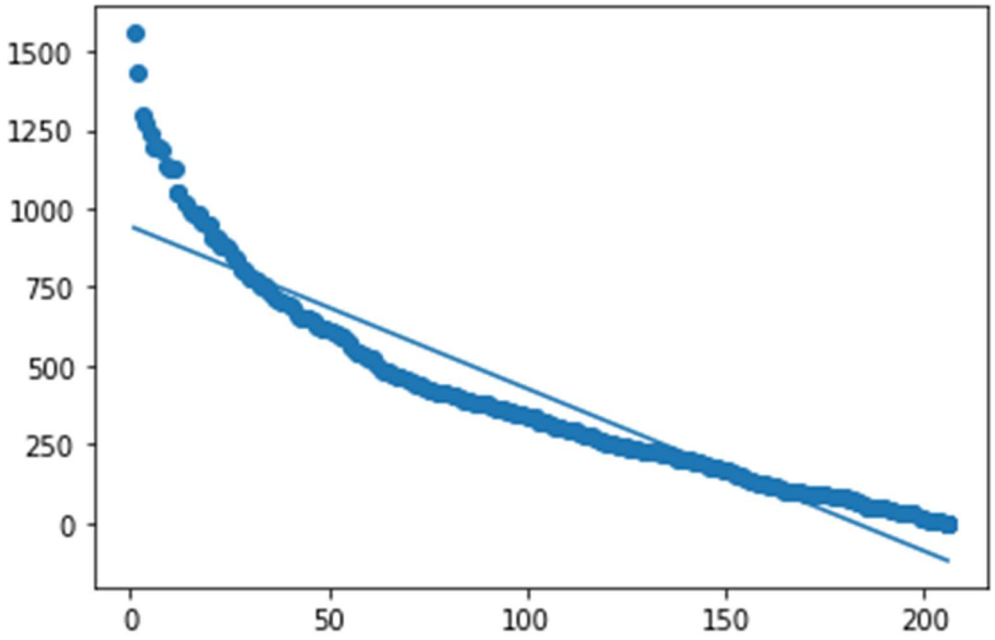
Brazil	143
Spain	64
Germany	28
Argentina	26
France	13
Italy	6
Belgium	5
Netherlands	1

The below chart shows the linear regression fit on the data I was using. I decided to use rank a nd points from the dataset so I could see what the model might look like if we wanted to predict how m any points, we would need to have a chance at achieving the rank of 1 worldwide.

```
# Seeing if my data will fit linear regression

x = df['rank']
y = df['total_points']

slope, intercept, r, p, std_err = stats.linregress(x, y)

print(r)
```

```
-0.9394426912166881
```

I was able to get a pretty good model fit of -.0939. I decided to proceed with this data, and I pl otted the data on a chart and the below output is what was created.



The code below is what we can use to input the desired rank of a team (which we can assume will always be number 1). When we enter 1 below, we will get an output of 938.22 points. This means th at a team that wants to be in the number 1 position will need to at least reach 938 points to make a num ber 1 world ranking possible.

```
x = df['rank']
y = df['total_points']

slope, intercept, r, p, std_err = stats.linregress(x, y)

def myfunc(x):
    return slope * x + intercept

# enter desired ranking
rank = myfunc(1)

# output will be points needed to reach desired ranking
print(rank)
```

```
938.2262104373618
```

This is a prediction of course and there is no guarantee that 938 points will result in a number 1 ranking. Coaches and staff could use this as a guide to see where they would need to be to be in the ru nning for a number one rank. This data could also be used to look at what other teams rank in the top te n to see what kind of points are needed to compete at the top of the field.

The questions that I wanted to answer with this project did change while I was working on the project. I wanted to do some more analysis on the players, and I could not find a dataset with player data other than a pdf file I had. My goal was to scrape the player data from the pdf file, but I was limited by time to achieve that objective. The questions below are the original and altered to show the information I found while doing this project.

- Which country has the most championships?
    - Brazil is the country with the most FIFA World Cup Championships. They have 5 total championships.
- Which teams are the most consistent/ranked 1 most often?
    - Brazil is also the most consistent team with receiving a world 1 ranking over 143 times.
- Which confederation is strongest?
    - The strongest confederation is the CONMEBOL confederation – The South American Football Confederation with 169 number world ranking.
- How does team strength relate to player performance?
    - This question I am going to comeback to at a later date. I did not have time to scrape the player data I had to analyze it to answer this question.
- Can we predict the next World Cup Champions? Changed to; how many points does a team need to reach the rank of 1 worldwide?
    - I did not get to do a machine learning algorithm like I wanted to for this project, but I was able to use linear regression to show how many points a team would need to

be in the running's for a number world ranking. For a team to reach the rank of 1 worldwide they would need to have at least 938 points.

References

1. Fitzgerald, Tadhg. FIFA Soccer Rankings. https://www.kaggle.com/tadhgfitzgerald/fifa-international-soccer-mens-ranking-1993now
2. FIFA.com. FIFA World Ranking. https://www.fifa.com/fifa-world-ranking/men?dateId=id13295
3. Jurisoo, Mart. International Football Results from 1872 – 2021. https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017/data
4. Muriuki, Gerald. June 8, 2018. Predicting FIFA World Cup 2018 Using Machine Learning. https://blog.goodaudience.com/predicting-fifa-world-cup-2018-using-machine-learning-dc07ad8dd576
5. Dar, Pranav. June 13, 2018. Data Scientists built a Random Forest Model to predict the World Cup 2018 Winner. https://www.analyticsvidhya.com/blog/2018/06/data-scientists-used-a-random-forest-model-to-predict-the-world-cup-2018-winner/
6. whoscored.com. https://www.whoscored.com/Regions/247/Tournaments/36/Seasons/5967/Stages/12763/PlayerStatistics/International-FIFA-World-Cup-2018
7. PDF. https://img.fifa.com/image/upload/hzfqyndmnqazczvc5xdb.pdf
8. FIFA.com Official Documents. https://www.record-evolution.de/en/fifa-world-cup-statistics-data-modeling-with-repods-part-i/
9. The Soccer World Cups.com https://www.thesoccerworldcups.com/stats.php
10. FBRef.com. 2014 FIFA World Cup Stats. https://fbref.com/en/comps/1/1709/2014-FIFA-World-Cup-Stats
11. Foxsports.com. FIFA World Cup. https://www.foxsports.com/soccer/fifa-world-cup/history

Appendix

For this project I was able to use linear regression to find insights into my dataset. I wanted to use machine learning to see if we could predict who the next world cup winner might be. During my analysis of this project, I think I tried to do to much and I ended up not giving myself enough time to train and test my dataset. I also noticed through my research that I was going to have to utilize other datasets to do the modeling I wanted to do.

I also wanted to try and use player data to determine team strength. I find a pdf file that had player data through FIFA, but I did not have the time to scrape the data to utilize it as a dataset. My goal was to input the pdf file into a program called Monarch. From here I was going to create a template to scrape the player data and turn it into a csv file to use with Jupyter notebook.

This is something I still want to try later to analyze player data t determine team strength to see how that compares to teams being ranked number 1.

10 Questions

1. How many games do the top ten teams play per year to maintain their dominance?
2. How many points do the top teams score in each match?
3. How many points are scored on the top teams in each match?
4. How strong is each national team's roster?
5. How do the second-string players rank nationally compared to other teams first string players?
6. Why are other conferences less dominant in the sport?
7. Who was the coach on the championship teams?
8. Who was the staff on the championship teams?
9. Which player has appeared in the most FIFA World Cups?
10. Who are the most valuable players on each national team?