

The dataset that I used was the Fifa World Cup dataset on Kaggle. This contains two datasets, an overview of the overall tournament in one, and each game in another. I was curious to see which country has won the most, and what was required of them to win. By this I mean I wanted to find out how many goals those countries scored and how many they conceded, and to see how many times they won with those stats.

I went about this by first grouping the dataset by the home_team, and using agg("sum"), to see the total numbers for each category. The two categories I was looking at are "home_score" and "away_score". What this did is it added up all of the goals that that country scored and conceded in the world cup. I then added a new column called "goal_dif" (goal differential). I could then use this to compare the countries with the most first places to see how they correlate.

```
matches_df.nlargest(5, "goaldif")
```

✓ 0.9s

	home_team	home_score	away_score	goaldif
8	Brazil	177	83	94
2	Argentina	114	61	53
78	West Germany	93	43	50
27	France	85	45	40
38	Italy	72	37	35

Above you can see that these are the top 5 countries with the highest goal differentials. Brazil is almost double the next top country which I found very surprising. I then set up a nested for loop, in order to add a new column to each country, showing the number of wins they had.

```
matches_df["WinCount"] = 0
matches_df
for i in range(len(matches_df)):
    team = matches_df.iloc[i].home_team
    count = 0
    for j in range(len(wc_df)):
        if (wc_df.iloc[j].winner) == team:
            count += 1
    matches_df.loc[matches_df.home_team == team, 'WinCount'] = count
```

The data set now looks like this.

```
matches_df.nlargest(5, "WinCount")
```

✓ 0.5s

	home_team	home_score	away_score	goaldif	WinCount
8	Brazil	177	83	94	5
38	Italy	72	37	35	4
78	West Germany	93	43	50	3
2	Argentina	114	61	53	2
27	France	85	45	40	2

As you can see, these are the top 5 countries with the most World Cup wins. At first I thought that the countries with the highest goal differentials would have the highest number of wins, which is true for Brazil, but it actually isn't the case for the other top 5 countries. It is true to an extent, as they are the same 5 countries in the highest goal differentials, but for example Italy has 15 less goal differential than West Germany, but it still has more wins. I believe that this is because of the unpredictability of soccer, and of the World Cup.

In conclusion, I found which countries had the most World Cup wins, what their goal differential was, and how having a higher goal differential gives you a better chance of winning the World Cup, all with using python I learned during this course!