



«Data Mining course»

Project 1 - Football

Dr.Farahani , Dr.Kheradpisheh

Ashkan Safavi Sohi

98422096

About dataset

This dataset includes **42,082** results of international football matches starting from the very first official match in 1972 up to 2019. The matches range from FIFA World Cup to FIFA Wild Cup to regular friendly matches. The matches are strictly men's full internationals and the data does not include Olympic Games or matches where at least one of the teams was the nation's B-team, U-23 or a league select team.

Analyze and Questions

- Who is the best team of all time

```
In [7]: ds['winner'].value_counts().head()

Out[7]: Brazil      629
        England     580
        Germany     560
        Argentina   529
        Sweden      506
        Name: winner, dtype: int64
```

As can be seen, Brazil has won the most matches

- Check the number of equal games during this period .

```
In [8]: ds['winner'].isnull().sum()

Out[8]: 9700
```

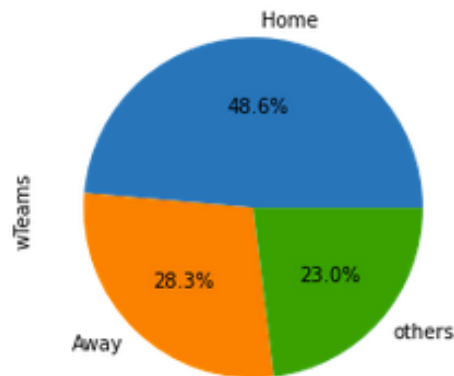
- Which team has been the worst in terms of performance?

```
In [9]: ds['loser'].value_counts().head()

Out[9]: Finland      404
        Switzerland  348
        Northern Ireland 338
        Norway       331
        Luxembourg   323
        Name: loser, dtype: int64
```

- How much does hosting a tournament help a team stand a chance of winning?

Out[40]: <AxesSubplot:ylabel='wTeams'>



- Can we say anything about geopolitics from football fixtures - how has the number of countries changed, which teams like to play each other

Out[49]:

	team1	team2	match_num
0	Argentina	Uruguay	176
1	Austria	Hungary	137
2	Belgium	Netherlands	125
3	England	Scotland	116
4	Sweden	Denmark	107
5	Norway	Sweden	107
6	Denmark	Sweden	107
7	Scotland	Wales	106
8	Brazil	Argentina	104
9	Wales	England	104
10	Argentina	Paraguay	102
11	Northern Ireland	England	99
12	Northern Ireland	Scotland	95
13	Northern Ireland	Wales	94
14	Finland	Sweden	89

According to the table above, the teams that have played the most games together are also neighbors

- Which countries host the most matches where they themselves are not participating in

Out[51]:

	country	counts
0	United States	772
1	Malaysia	428
2	France	375
3	South Africa	283
4	United Arab Emirates	273
5	England	251
6	Qatar	241
7	Spain	215
8	Thailand	207
9	Sweden	196

•

- Which teams are the most active in playing friendlies and friendly tournaments - does it help or hurt them

Out[52]:

	team	counts
0	Germany	572
1	Hungary	565
2	Sweden	544
3	Poland	529
4	France	488
5	Austria	466
6	Switzerland	457
7	Netherlands	452
8	Belgium	446
9	Mexico	441

According to the table above, it can be concluded that friendly games have been useful