In [4]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import datetime
```

In [5]:
```
df = pd.read_csv("international_matches.csv")
df.head()
```

Out[5]:

	date	home_team	away_team	home_team_continent	away_team_continent	home_team_fifa_rank	away_team_fifa_rank	home_team_total_fifa_
0	1993-08-08	Bolivia	Uruguay	South America	South America	59	22	
1	1993-08-08	Brazil	Mexico	South America	North America	8	14	
2	1993-08-08	Ecuador	Venezuela	South America	South America	35	94	
3	1993-08-08	Guinea	Sierra Leone	Africa	Africa	65	86	
4	1993-08-08	Paraguay	Argentina	South America	South America	67	5	

5 rows × 25 columns

In [5]:
```
df.tail()
```

Out[5]:

	date	home_team	away_team	home_team_continent	away_team_continent	home_team_fifa_rank	away_team_fifa_rank	home_team_tota
23916	2022-06-14	Moldova	Andorra	Europe	Europe	180	153	
23917	2022-06-14	Liechtenstein	Latvia	Europe	Europe	192	135	
23918	2022-06-14	Chile	Ghana	South America	Africa	28	60	
23919	2022-06-14	Japan	Tunisia	Asia	Africa	23	35	
23920	2022-06-14	Korea Republic	Egypt	Asia	Africa	29	32	

5 rows × 25 columns

In [6]:
```
df.describe()
```

Out[6]:

	home_team_fifa_rank	away_team_fifa_rank	home_team_total_fifa_points	away_team_total_fifa_points	home_team_score	away_team_score
count	23921.000000	23921.000000	23921.000000	23921.000000	23921.000000	23921.000000
mean	77.854688	80.797375	323.401488	315.453576	1.609214	1.068266
std	52.355225	53.232902	500.825725	490.944273	1.630127	1.263944
min	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	33.000000	36.000000	0.000000	0.000000	0.000000	0.000000
50%	71.000000	73.000000	0.000000	0.000000	1.000000	1.000000
75%	115.000000	119.000000	547.000000	523.000000	2.000000	2.000000
max	211.000000	211.000000	2164.000000	2164.000000	31.000000	21.000000

In [7]:
```
df.columns
```

Out[7]:
```
Index(['date', 'home_team', 'away_team', 'home_team_continent',
       'away_team_continent', 'home_team_fifa_rank', 'away_team_fifa_rank',
       'home_team_total_fifa_points', 'away_team_total_fifa_points',
       'home_team_score', 'away_team_score', 'tournament', 'city', 'country',
       'neutral_location', 'shoot_out', 'home_team_result',
       'home_team_goalkeeper_score', 'away_team_goalkeeper_score',
       'home_team_mean_defense_score', 'home_team_mean_offense_score',
       'home_team_mean_midfield_score', 'away_team_mean_defense_score',
       'away_team_mean_offense_score', 'away_team_mean_midfield_score'],
      dtype='object')
```

In [8]:
```
df['date'] = pd.to_datetime(df['date'])
```

In [11]:
```
#Top 10 FIFA Rank

#let's re-shape the data a bit
fifa_rank = df[['date', 'home_team', 'away_team', 'home_team_fifa_rank', 'away_team_fifa_rank',
'away_team_total_fifa_points','home_team_total_fifa_points']]
home = fifa_rank[['date', 'home_team','home_team_fifa_rank','home_team_total_fifa_points']].rename(columns = {'ho
away = fifa_rank[['date', 'away_team','away_team_fifa_rank','away_team_total_fifa_points']].rename(columns = {'aw
fifa_rank = home.append(away)

#select for each country the latest match
fifa_rank  = fifa_rank.sort_values(['team', 'date'], ascending=[True, False])
fifa_rank['row_number'] = fifa_rank.groupby('team').cumcount()+1
fifa_rank_top = fifa_rank[fifa_rank['row_number']==1].drop('row_number',axis=1).nsmallest(10, 'rank')
#fifa_points_top = fifa_rank[fifa_rank['row_number']==1].drop('row_number',axis=1).nlargest(10, 'rank_points')

#let's see the 5 strongest teams
fifa_rank_top
```

Out[11]:

	date	team	rank	rank_points
23760	2022-06-06	Brazil	1	1832
23909	2022-06-14	Belgium	2	1827
23885	2022-06-13	France	3	1789
23741	2022-06-05	Argentina	4	1765
23906	2022-06-14	England	5	1761
23907	2022-06-14	Italy	6	1723
23866	2022-06-12	Spain	7	1709
23867	2022-06-12	Portugal	8	1674
23903	2022-06-14	Mexico	9	1658
23908	2022-06-14	Netherlands	10	1658

In [12]:
```
#Top 10 Offensers

offense = df[['date', 'home_team', 'away_team', 'home_team_mean_offense_score', 'away_team_mean_offense_score']]
home = offense [['date', 'home_team', 'home_team_mean_offense_score']].rename(columns = {'home_team': 'team', 'ho
away = offense [['date', 'away_team', 'away_team_mean_offense_score']].rename(columns = {'away_team': 'team', 'aw
offense = home.append(away)
#offense
#last match that each country played
offense = offense.sort_values(['team', 'date'], ascending=[True, False])
offense ['row_number'] = offense.groupby('team').cumcount()+1
offense_top_data = offense [offense ['row_number']==1].drop('row_number',axis=1).nlargest(10, 'offense_score')
offense_top_data
```

Out[12]:

	date	team	offense_score
23741	2022-06-05	Argentina	89.0
23885	2022-06-13	France	88.3
23906	2022-06-14	England	88.0
23760	2022-06-06	Brazil	86.3
23867	2022-06-12	Portugal	86.0
23909	2022-06-14	Belgium	85.7
23907	2022-06-14	Italy	85.3
23866	2022-06-12	Spain	85.0
23909	2022-06-14	Poland	84.7
23862	2022-06-11	Uruguay	84.3

In [14]:
```
#Top 10 Defenders

defense = df[['date', 'home_team', 'away_team', 'home_team_mean_defense_score', 'away_team_mean_defense_score']]
home = defense [['date', 'home_team','home_team_mean_defense_score']].rename(columns = {'home_team': 'team', 'hom
away = defense [['date', 'away_team','away_team_mean_defense_score']].rename(columns = {'away_team': 'team', 'awa
defense  = home.append(away)
# offense
```

```
#last match that each country played
defense   = defense .sort_values(['team', 'date'], ascending=[True, False])
defense ['row_number'] = defense .groupby('team').cumcount()+1
defense_top_data = defense [offense ['row_number']==1].drop('row_number',axis=1).nlargest(10, 'defense_score')
defense_top_data
```

Out[14]:

	date	team	defense_score
23866	2022-06-12	Spain	86.5
23908	2022-06-14	Netherlands	85.2
23867	2022-06-12	Portugal	85.2
23906	2022-06-14	England	85.0
23760	2022-06-06	Brazil	84.8
23885	2022-06-13	France	84.2
23907	2022-06-14	Italy	84.2
23907	2022-06-14	Germany	84.0
23741	2022-06-05	Argentina	82.2
23879	2022-06-13	Morocco	81.2

In [17]:

```
#Top 10 Midfielders

midfield = df[['date', 'home_team', 'away_team', 'home_team_mean_midfield_score', 'away_team_mean_midfield_score'
home = midfield [['date', 'home_team','home_team_mean_midfield_score']].rename(columns = {'home_team': 'team', 'h
away = midfield [['date', 'away_team','away_team_mean_midfield_score']].rename(columns = {'away_team': 'team', 'a
midfield  = home.append(away)
# offense
#last match that each country played
midfield   = midfield .sort_values(['team', 'date'], ascending=[True, False])
midfield ['row_number'] = defense .groupby('team').cumcount()+1
midfield_top_data = midfield [offense ['row_number']==1].drop('row_number',axis=1).nlargest(10, 'midfield_score')
midfield_top_data
```

Out[17]:

	date	team	midfield_score
23907	2022-06-14	Germany	87.8
23885	2022-06-13	France	86.8
23866	2022-06-12	Spain	86.0
23909	2022-06-14	Belgium	85.5
23760	2022-06-06	Brazil	85.5
23907	2022-06-14	Italy	84.5
23867	2022-06-12	Portugal	84.5
23885	2022-06-13	Croatia	84.2
23741	2022-06-05	Argentina	84.0
23906	2022-06-14	England	84.0

In [22]:

```
#Top 10 Total Score
df['home_total_score'] = (df['home_team_mean_offense_score'] + df['home_team_mean_defense_score']+df['home_team_m
df['away_total_score'] = (df['away_team_mean_offense_score'] + df['away_team_mean_defense_score']+df['away_team_m
df['diff_total_score'] = df['home_total_score'] - df['away_total_score']
```

In [23]:

```
total_score = df[['date', 'home_team', 'away_team', 'home_total_score', 'away_total_score']]
home = total_score [['date', 'home_team','home_total_score']].rename(columns = {'home_team': 'team', 'home_total_
away = total_score [['date', 'away_team','away_total_score']].rename(columns = {'away_team': 'team', 'away_total_
total_score  = home.append(away)

#last match that each country played
total_score   = total_score .sort_values(['team', 'date'], ascending=[True, False])
total_score ['row_number'] = defense .groupby('team').cumcount()+1
total_score_top_10 = total_score [offense ['row_number']==1].drop('row_number',axis=1).nlargest(10, 'total_score'
total_score_top_10
```

Out[23]:

	date	team	total_score
23885	2022-06-13	France	86.433333
23866	2022-06-12	Spain	85.833333
23906	2022-06-14	England	85.666667
23760	2022-06-06	Brazil	85.533333

23867	2022-06-12	Portugal	85.233333
23741	2022-06-05	Argentina	85.066667
23907	2022-06-14	Germany	84.833333
23907	2022-06-14	Italy	84.666667
23909	2022-06-14	Belgium	84.000000
23908	2022-06-14	Netherlands	83.900000

In [ ]: