UEFA EURO2024 GERMANY

EXL HACKATHON 2024

By ABHISHEK THOMAS

INDEX

- ENGLISH OR SPANISH?
- DATA
- METHODOLOGY
- NEXT STEPS
- REFERENCES







ENGLISH OR SPANISH?

ITS NOT COMING HOME AGAIN…..

- After rigorous calculations and countless simulations, our model predicts that Spain will **win** the 2024 UEFA Euro by a narrow margin of **one goal**. A tough and exciting match for both sides!
- Model Results

TOTAL POINTS : 86

Phase 1		
Type	**%**	**Description**
Accuracy	47%	17 out of 36 matches correct predicted
Correct goal difference	59%	10 out of 17 matches with correct goal difference
Partial goal difference	24%	4 out of 17 with partial goal difference
Incorrect goal difference	18%	3 out of 17 with incorrect goal difference
Top goal scorer	None	Romelu Lukaku
TOTAL MATCH POINTS : 29 , TOP SCORER POINTS : 0		

Phase 2		
Type	**%**	**Description**
Accuracy	75%	9 out of 12 matches correct predicted
Correct goal difference	22%	2 out of 9 matches with correct goal difference
Partial goal difference	42%	5 out of 9 with partial goal difference
Incorrect goal difference	36%	2 out of 9 with incorrect goal difference
Top goal scorer	None	Niclas Füllkrug
TOTAL MATCH POINTS : 27 , TOP SCORER POINTS : 0		

Phase 3		
Type	**%**	**Description**
Accuracy	100%	3 out of 3 matches correct predicted
Correct goal difference	100%	3 out of 3 matches with correct goal difference
Partial goal difference	0%	0 out of 3 with partial goal difference
Incorrect goal difference	0%	0 out of 3 with incorrect goal difference
Top goal scorer	None	Jude Bellingham
TOTAL MATCH POINTS : 30 , TOP SCORER POINTS : 0		

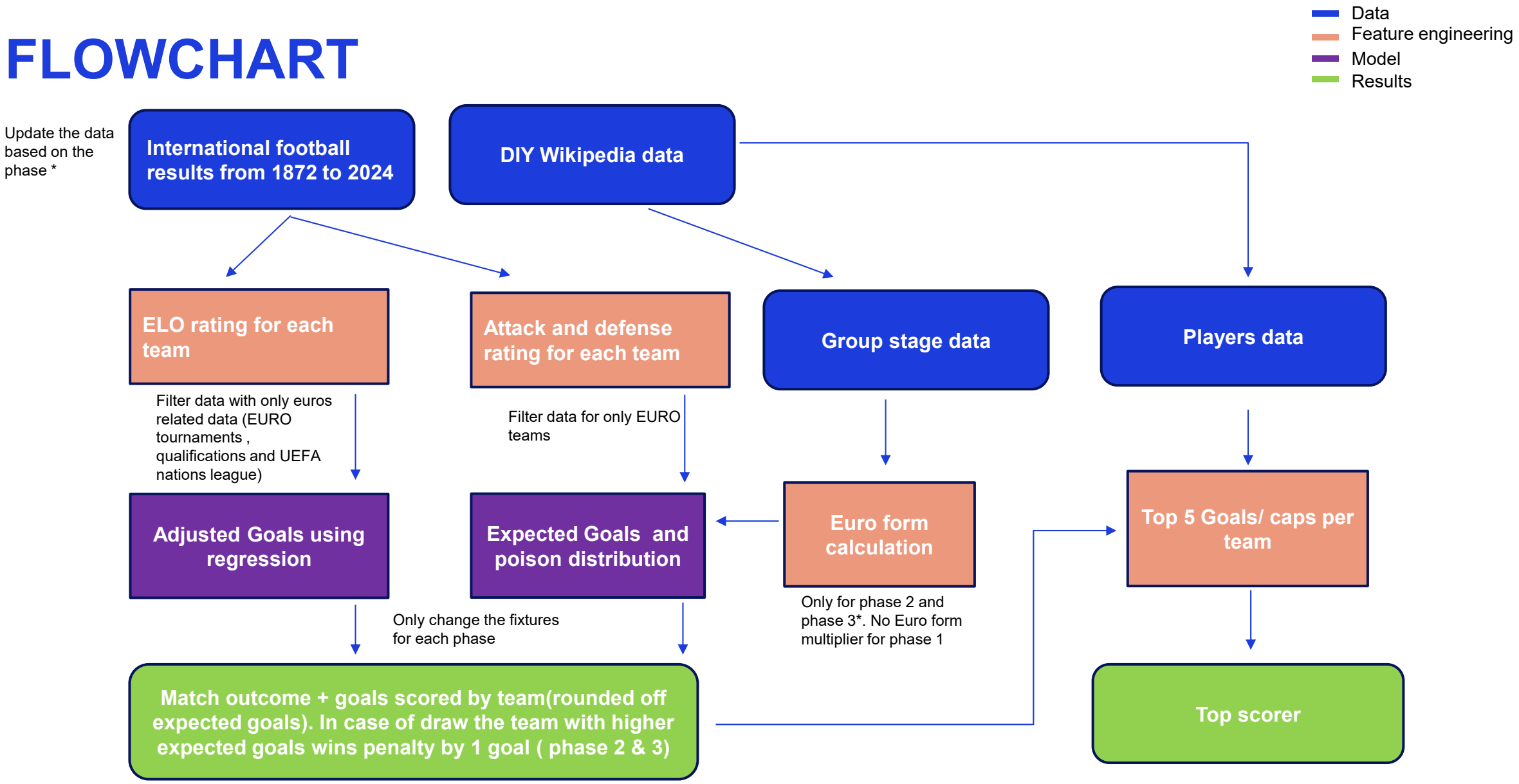
DATA USED

KAGGLE + WIKIPEDIA + GOOGLE

- **Kaggle Dataset: International match results from 1872 to 2024**
  - Contains international match results from 1872 to 2024 by **MART JÜRISOO** . updates every day!
  - Key columns: date, home_team, away_team, home_score, away_score, tournament, neutral.
- **DIY Wikipedia Data:**
  - Euro Players Data: Information on players participating in the Euros, including player_name, team, position, goals_scored, appearances. Data is static
  - Euro Form Data: Team performance metrics, including team, matches_played, wins, draws, losses, goals_for, goals_against, points. Data is static

METHODOLOGY

FLOWCHART


Data
Feature engineering
Model
Results
Update the data based on the phase *
International football results from 1872 to 2024
DIY Wikipedia data
ELO rating for each team
Attack and defense rating for each team
Group stage data
Players data
Filter data with only euros related data (EURO tournaments , qualifications and UEFA nations league)
Filter data for only EURO teams
Adjusted Goals using regression
Expected Goals and poison distribution
Euro form calculation
Top 5 Goals/ caps per team
Only change the fixtures for each phase
Only for phase 2 and phase 3*. No Euro form multiplier for phase 1
Match outcome + goals scored by team(rounded off expected goals). In case of draw the team with higher expected goals wins penalty by 1 goal (phase 2 & 3)
Top scorer

SCIENCE + FOOTBALL = WINNING FORMULA

- **Elo Calculation:**
  - All teams get ELO rating. After each match the ELO rating will get updated based on the type of match and difference of actual result and expected result. In Elo formula, the difference is between the two teams for expected result
  - Formula: $R'_a = R_a + K \times (W - E)$

    - R'_a is the new Elo rating.
    - R_a is the current Elo rating.
    - K is the weight constant (10 for friendlies, 20 for qualifiers, 40 for tournaments).
    - W is the actual result (1 for win, 0.5 for draw, 0 for loss).
    - E is the expected result:
      $E = \frac{1}{1+10^{(\frac{R_b - R_a}{400})}}$
      Default Elo rating: 1300.
- **Team attack and defense calculation:**
  - All teams get Attack and defense rating. After each match the Attack and defense will get updated based on the type of match and difference of actual goals and expected goals
  - Formula:

    $$\text{New Attack} = \text{Current Attack} + \frac{K}{2000} \times (\text{Goals Scored} - \text{Expected Goals})$$
    $$\text{New Defense} = \text{Current Defense} + \frac{K}{2000} \times (\text{Goals Conceded} - \text{Expected Goals})$$
- **Expected Go**Default values: Attack: 1.3, Defense: 1.3
  - The Poisson distribution is used to model the number of goals scored in a match. It is based on the expected number of goals a team is expected to score. The final score is rounded by expected goals
  - Incase a match ends in a draw, the team with the higher expected goal would wins the penalties with a goal difference of 1
  - Formula:

    $$ExpectedGoals = AttackingPowerTeam \times DefensivePowerOpponent$$

    $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$

    - λ is the expected goal rate.
    - k is the actual number of goals.

Did You Know?

- Copa America: No qualification rounds! All 10 teams are automatically eligible. Imagine a football tournament where everyone gets a golden ticket!

SCIENCE + FOOTBALL = WINNING FORMULA

- **Adjusted Goals Using Elo Regression with European performance :**
  - The Final goals scored by each team is determined by adding the expected goals and Adjusted goals
  - Target variable is the Adjusted goals (difference between goals scored and average goals scored by each team), the independent variable is the ELO difference between both the teams
  - Since its variable difference between goals scored and average goals, the regression can give negative results which could result in a result in a negative final score. To avoid this we add the modulus of negative value to both the teams Eg: Spain:3 – Georgia -1 would finally be Spain:4 – Georgia:0

$$FinalGoals = ExpectedGoals + AdjustedGoals$$

- **Euro form multiplier with EURO 24 group stage data:**
  - The expected goals is updated by multiplying with the current euro form. The current euro form was calculated using the points scored in group stage. The euro form is the Points scored by team by the median points scored by all European teams
  - This is only applicable for phase 2 and phase 3, this increased the accuracy from 47% to 75% from phase 1 to phase 2
  - Formula:

$$Euroform = Pointsscoredingroupstage/Medianofpointsscored$$

$$ExpectedGoals = AttackingPowerTeam \times DefensivePowerOpponent \times Euroform$$

- **Top Goal scorer**
  - All the players from each team, goals ,and cap are taken into account to calculate Goal per caps. Top 5 players Goals per caps are chosen from each. The total goals scored by each team using the match prediction model is redistributed among the top 5 players for each. This method ensures we have the top goal scorer based on the team perforo
  - dfdgg

Did You Know?

- Maximum games team can play in one euro tournament (tournament + qualifications)~19
- Maximum games team can play in one nationals ~8
- UEFA tournaments since 2010 is 3
- UEFA Nations league since 2010 is 3
- Total maximum matches = (19 x 3) + (8 x 3) =81 matches

NEXT STEPS

UNTIL NEXT TIME…

Model Improvements:

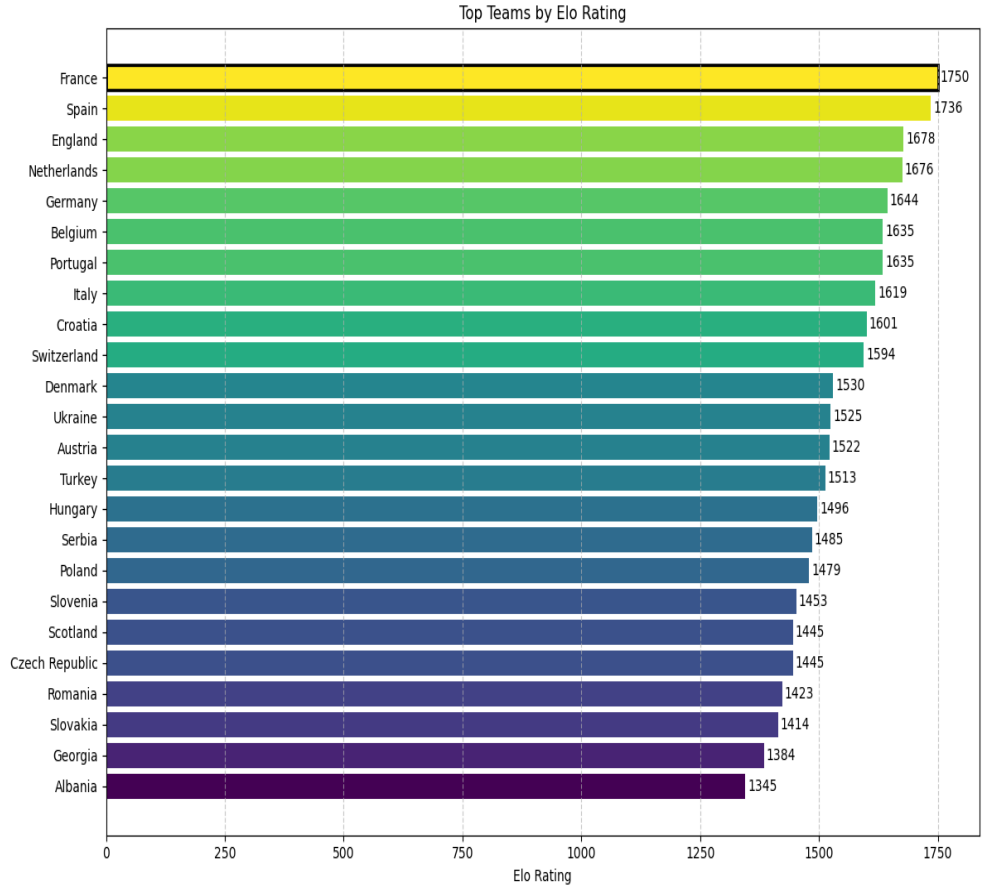
- Continuous updates and more granular data will make future predictions even more accurate. It's like fine wine, only better with time.
- Need to work on top goal scorer

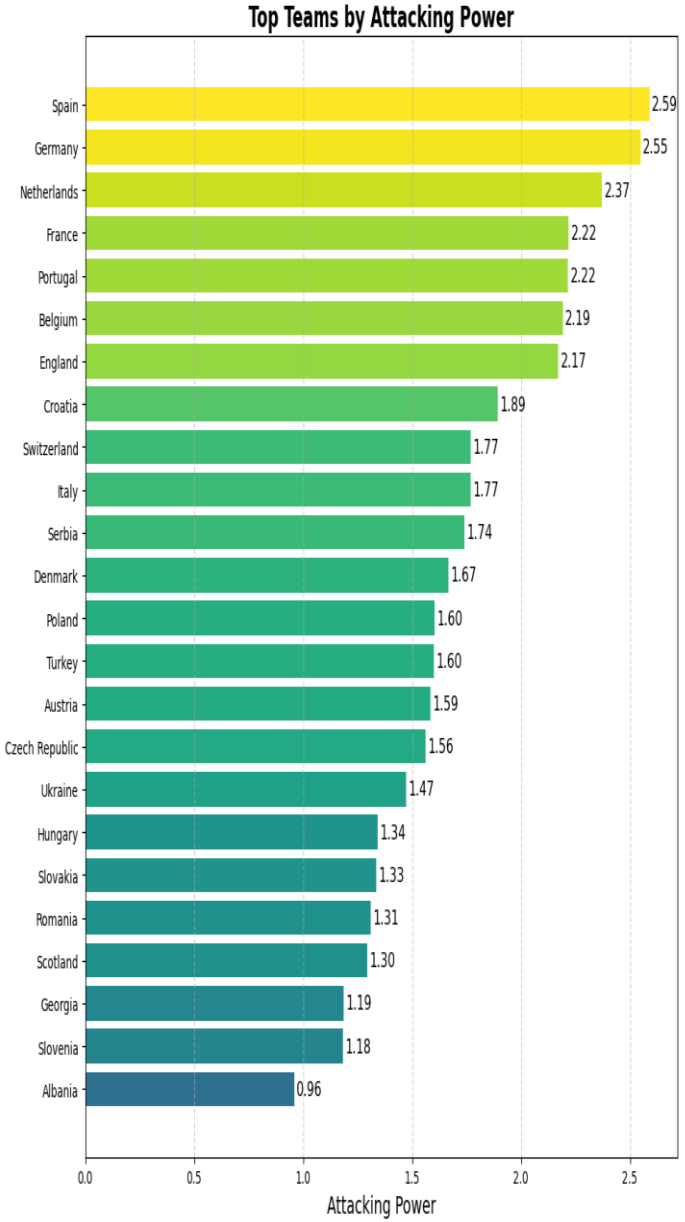
	Grupo	Home_Team	Away_Team	XGhome	XGaway	home_goals_elo	away_goals_elo
0	Phase2	Switzerland	Italy	1.063382	1.137415	1	1
1	Phase2	England	Slovakia	2.033448	0.562997	3	0
2	Phase2	Germany	Denmark	2.389509	0.796499	2	0
3	Phase2	Spain	Georgia	5.265882	0.482845	6	0
4	Phase2	France	Belgium	1.316110	0.899273	1	0
5	Phase2	Portugal	Slovenia	2.247271	0.429674	2	0
6	Phase2	Romania	Netherlands	0.720475	1.670479	0	1
7	Phase2	Austria	Turkey	1.931174	1.810637	1	1
8	Phase2	Italy	England	0.746458	1.304899	0	1
9	Phase2	Germany	Spain	1.814036	3.715199	1	3
10	Phase2	France	Portugal	1.344518	1.363729	1	1
11	Phase2	Netherlands	Austria	1.789079	1.307181	1	1
12	phase3	Spain	France	2.390389	1.128333	2	1
13	phase3	Netherlands	England	1.000482	1.492086	1	1
14	phase3	Spain	England	2.458173	1.103973	2	1

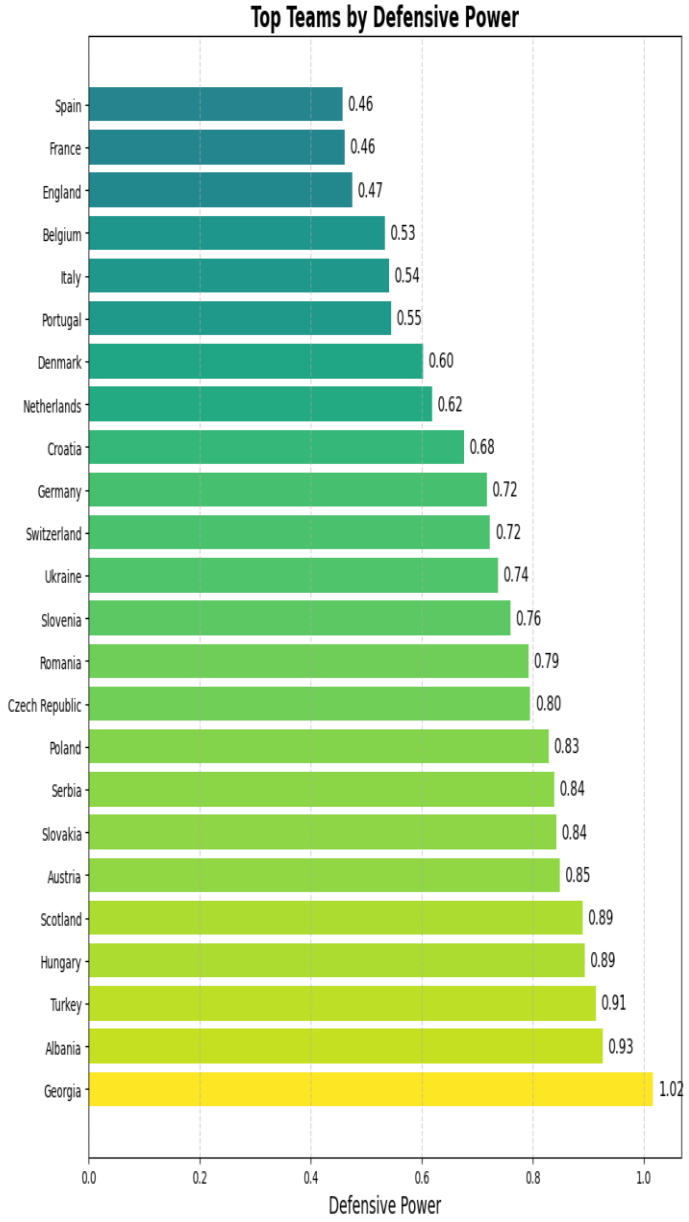
	Grupo	Home_Team	Away_Team	XGhome	XGaway	home_goals_elo	away_goals_elo
0	A	Germany	Scotland	2.213483	0.968680	2	0
1	A	Hungary	Switzerland	1.068940	1.578246	1	1
2	A	Germany	Hungary	2.341430	1.004496	2	1
3	A	Scotland	Switzerland	1.030826	1.492003	1	1
4	A	Switzerland	Germany	1.274685	2.012401	1	2
5	A	Scotland	Hungary	1.199367	1.175750	1	1
6	B	Spain	Croatia	1.597877	0.962523	1	0
7	B	Italy	Albania	1.714469	0.479096	2	0
8	B	Croatia	Albania	1.803530	0.580921	2	0
9	B	Spain	Italy	1.317798	0.914992	1	0
10	B	Albania	Spain	0.469228	2.418368	0	3
11	B	Croatia	Italy	0.982765	1.132793	0	1
12	C	Slovenia	Denmark	0.754581	1.466923	0	1
13	C	Serbia	England	0.881253	1.985955	0	1
14	C	Slovenia	Serbia	1.070937	1.526074	1	1
15	C	Denmark	England	0.847095	1.399301	0	1
16	C	England	Slovenia	1.903610	0.592786	2	0
17	C	Denmark	Serbia	1.530378	1.121783	1	1
18	D	Poland	Netherlands	1.000169	1.920270	1	1
19	D	Austria	France	0.833611	1.994929	0	2
20	D	Poland	Austria	1.384308	1.243268	1	1
21	D	Netherlands	France	1.287541	1.441345	1	1
22	D	Netherlands	Austria	2.041431	0.954944	2	0
23	D	France	Poland	1.876528	0.873090	2	0
24	E	Romania	Ukraine	0.954991	1.179193	0	1
25	E	Belgium	Slovakia	1.975744	0.746969	2	0
26	E	Slovakia	Ukraine	0.968484	1.297096	0	1
27	E	Belgium	Romania	1.796154	0.736562	2	0
28	E	Slovakia	Romania	1.040725	1.128833	1	1
29	E	Ukraine	Belgium	0.846353	1.671475	0	1

APPENDIX

ELO , ATT & DEF


Top Teams by Elo Rating
France 1750
Spain 1736
England 1678
Netherlands 1676
Germany 1644
Belgium 1635
Portugal 1635
Italy 1619
Croatia 1601
Switzerland 1594
Denmark 1530
Ukraine 1525
Austria 1522
Turkey 1513
Hungary 1496
Serbia 1485
Poland 1479
Slovenia 1453
Scotland 1445
Czech Republic 1445
Romania 1423
Slovakia 1414
Georgia 1384
Albania 1345
0 250 500 750 1000 1250 1500 1750
Elo Rating


Top Teams by Attacking Power
Spain 2.59
Germany 2.55
Netherlands 2.37
France 2.22
Portugal 2.22
Belgium 2.19
England 2.17
Croatia 1.89
Switzerland 1.77
Italy 1.77
Serbia 1.74
Denmark 1.67
Poland 1.60
Turkey 1.60
Austria 1.59
Czech Republic 1.56
Ukraine 1.47
Hungary 1.34
Slovakia 1.33
Romania 1.31
Scotland 1.30
Georgia 1.19
Slovenia 1.18
Albania 0.96
0.0 0.5 1.0 1.5 2.0 2.5
Attacking Power


Top Teams by Defensive Power
Spain 0.46
France 0.46
England 0.47
Belgium 0.53
Italy 0.54
Portugal 0.55
Denmark 0.60
Netherlands 0.62
Croatia 0.68
Germany 0.72
Switzerland 0.72
Ukraine 0.74
Slovenia 0.76
Romania 0.79
Czech Republic 0.80
Poland 0.83
Serbia 0.84
Slovakia 0.84
Austria 0.85
Scotland 0.89
Hungary 0.89
Turkey 0.91
Albania 0.93
Georgia 1.02
0.0 0.2 0.4 0.6 0.8 1.0
Defensive Power

POISSON MATRIX

Goals by France is: 2 and Goals by Albania is 0

Poission distribution matrix for France and Albania

	0	1	2	3	4	5	6	7	8	9	10
0	0.047075	0.025441	0.006875	0.001238	0.000167	0.000018	0.000002	0.000000	0.000000	0.000000	0.000000
1	0.118421	0.063999	0.017294	0.003115	0.000421	0.000045	0.000004	0.000000	0.000000	0.000000	0.000000
2	0.148948	0.080498	0.021752	0.003919	0.000529	0.000057	0.000005	0.000000	0.000000	0.000000	0.000000
3	0.124896	0.067499	0.018240	0.003286	0.000444	0.000048	0.000004	0.000000	0.000000	0.000000	0.000000
4	0.078546	0.042450	0.011471	0.002066	0.000279	0.000030	0.000003	0.000000	0.000000	0.000000	0.000000
5	0.039518	0.021357	0.005771	0.001040	0.000140	0.000015	0.000001	0.000000	0.000000	0.000000	0.000000
6	0.016568	0.008954	0.002420	0.000436	0.000059	0.000006	0.000001	0.000000	0.000000	0.000000	0.000000
7	0.005954	0.003218	0.000870	0.000157	0.000021	0.000002	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.001872	0.001012	0.000273	0.000049	0.000007	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000523	0.000283	0.000076	0.000014	0.000002	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000132	0.000071	0.000019	0.000003	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000