

FIFA World Cup Data Analysis

By: Robert Harakaly, Antoine Martinet, Zain Mirza, and Ryan Townsend

WHAT IS THE FIFA WORLD CUP?

- Most Prestigious
- Most widely viewed and followed single sporting event
 - 800 million+ viewers



Tasks Overview

- Strongest Defense? Midfield? Offense?
- Does 'Home-Team Advantage' exist?
- Longest Winning Streaks?
- Most Wins?
- Better to have a strong Defense? Midfield? Offense? (in terms of winning WC)
- Most likely to win FIFA World Cup 2022?



Data Cleansing + Processing



- Drop N/A values
- Distributions
- Drop unnecessary rows
 - i.e. 'tournament' type

`home_team_goalkeeper_score`
15542

`away_team_goalkeeper_score`
15826

`home_team_mean_defense_score`
16134

`home_team_mean_offense_score`
15411

`home_team_mean_midfield_score`
15759

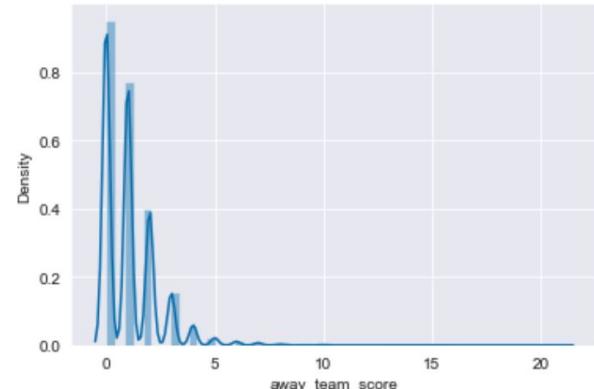
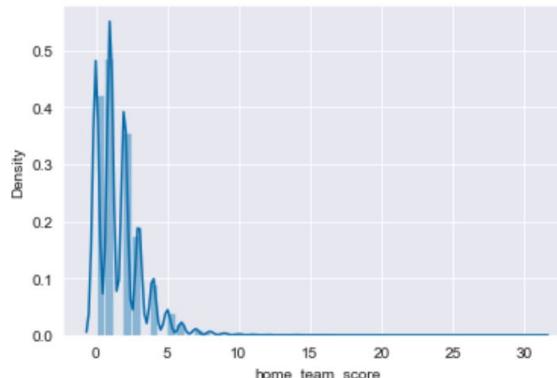
`away_team_mean_defense_score`
16357

`away_team_mean_offense_score`
15609

`away_team_mean_midfield_score`
15942

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23921 entries, 0 to 23920
Data columns (total 25 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   date             23921 non-null   object  
 1   home_team        23921 non-null   object  
 2   away_team        23921 non-null   object  
 3   home_team_continent  23921 non-null   object  
 4   away_team_continent  23921 non-null   object  
 5   home_team_fifa_rank  23921 non-null   int64  
 6   away_team_fifa_rank  23921 non-null   int64  
 7   home_team_total_fifa_points  23921 non-null   int64  
 8   away_team_total_fifa_points  23921 non-null   int64  
 9   home_team_score    23921 non-null   int64  
 10  away_team_score    23921 non-null   int64  
 11  tournament        23921 non-null   object  
 12  city              23921 non-null   object  
 13  country           23921 non-null   object  
 14  neutral_location  23921 non-null   bool    
 15  shoot_out         23921 non-null   object  
 16  home_team_result  23921 non-null   object  
 17  home_team_goalkeeper_score  8379 non-null   float64 
 18  away_team_goalkeeper_score  8095 non-null   float64 
 19  home_team_mean_defense_score  7787 non-null   float64 
 20  home_team_mean_offense_score  8510 non-null   float64 
 21  home_team_mean_midfield_score  8162 non-null   float64 
 22  away_team_mean_defense_score  7564 non-null   float64 
 23  away_team_mean_offense_score  8312 non-null   float64 
 24  away_team_mean_midfield_score  7979 non-null   float64 
dtypes: bool(1), float64(8), int64(6), object(10)
memory usage: 4.4+ MB
```

['FIFA World Cup qualification' 'Friendly'
 'African Cup of Nations qualification' 'Amílcar Cabral Cup'
 'CFU Caribbean Cup qualification'
 'United Arab Emirates Friendship Tournament'
 'Malta International Tournament' 'Lunar New Year Cup'
 'African Cup of Nations' 'CFU Caribbean Cup' 'UEFA Euro qualification'
 'Kirin Cup' 'FIFA World Cup' 'Oceania Nations Cup qualification'
 'Baltic Cup' 'Gulf Cup' 'Simba Tournament' 'CECAFA Cup'
 'Confederations Cup' 'Dynasty Cup' 'King's Cup' 'Nehru Cup' 'SAFF Cup'
 'Copa Paz del Chaco' 'Korea Cup' 'USA Cup' 'Copa América'
 'Merdeka Tournament' 'South Pacific Games' 'UNCAF Cup'
 'Oceania Nations Cup' 'Windward Islands Tournament' 'Gold Cup'
 'AFC Asian Cup qualification' 'UEFA Euro' 'AFF Championship'
 'AFC Asian Cup' 'King Hassan II Tournament'
 'Cyprus International Tournament' 'Dunhill Cup'
 'COSAFA Cup qualification' 'COSAFA Cup' 'Tournoi de France'
 'Gold Cup qualification' 'SKN Football Festival' 'Arab Cup qualification'
 'Arab Cup' 'UNIFFAC Cup' 'Nordic Championship' 'WAFF Championship'
 'Millennium Cup' 'Cup of Ancient Civilizations' "Prime Minister's Cup"
 'EAFF Championship' 'TIFOCO Tournament' 'Afro-Asian Games'
 'AFC Challenge Cup' 'Copa del Pacífico' 'AFC Challenge Cup qualification'
 'African Nations Championship' 'VFF Cup' 'Dragon Cup'
 'Nile Basin Tournament' 'Nations Cup' 'Copa Confraternidad'
 'Pacific Games' 'Superclásico de las Américas' 'ABCS Tournament'
 'Kirin Challenge Cup' 'OSN Cup' 'Copa América qualification'
 'Pacific Mini Games' 'Intercontinental Cup'
 'AFF Championship qualification' 'UEFA Nations League'
 'CONCACAF Nations League qualification'
 'African Nations Championship qualification' 'CONCACAF Nations League'
 'Three Nations Cup' 'Mahinda Rajapaksa Cup' 'Navruz Cup'
 'CONMEBOL-UEFA Cup of Champions']



Task 1: Strongest Defense? Midfield? Offense?



DEFENSE

- Metric = EA Sports' FIFA
 - 30+ statistics for player performance
- Strongest → average FIFA game score

away_team	away_team_mean_defense_score	home_team	home_team_mean_defense_score	overall defense average	
Brazil	86.086364	Brazil	85.785185	Brazil	85.94
Spain	85.427193	Spain	85.585366	Spain	85.51
Italy	85.124742	Italy	84.960902	Italy	85.04
Germany	84.671429	England	84.733607	England	84.69
England	84.640426	Germany	84.700694	Germany	84.69
France	84.132258	France	84.047887	France	84.09
Argentina	83.353636	Portugal	82.887705	Argentina	83.00
Portugal	82.896154	Argentina	82.646721	Portugal	82.89
Belgium	81.153409	Belgium	81.304717	Belgium	81.23
Netherlands	80.681250	Netherlands	80.776423	Netherlands	80.73

MIDFIELD

	away_team_mean_midfield_score
away_team	
Spain	86.978947
France	86.138710
Germany	85.801020
Brazil	85.465455
Argentina	84.720000
England	84.553191
Italy	84.228866
Netherlands	83.621875
Portugal	83.617308
Belgium	81.327273

	home_team_mean_midfield_score
home_team	
Spain	87.265854
France	86.171831
Germany	85.793750
Brazil	85.244444
England	84.478689
Italy	84.166917
Argentina	84.164754
Netherlands	83.777236
Portugal	83.567213
Belgium	81.703774

	midfield average
Spain	87.12
France	86.16
Germany	85.80
Brazil	85.35
England	84.52
Argentina	84.44
Italy	84.20
Netherlands	83.70
Portugal	83.59
Belgium	81.52

OFFENSE

away_team	away_team_mean_offense_score
Argentina	88.13
Brazil	86.70
France	85.71
Spain	85.69
Italy	85.44
Netherlands	85.43
England	84.91
Portugal	84.89
Germany	83.51
Uruguay	83.49

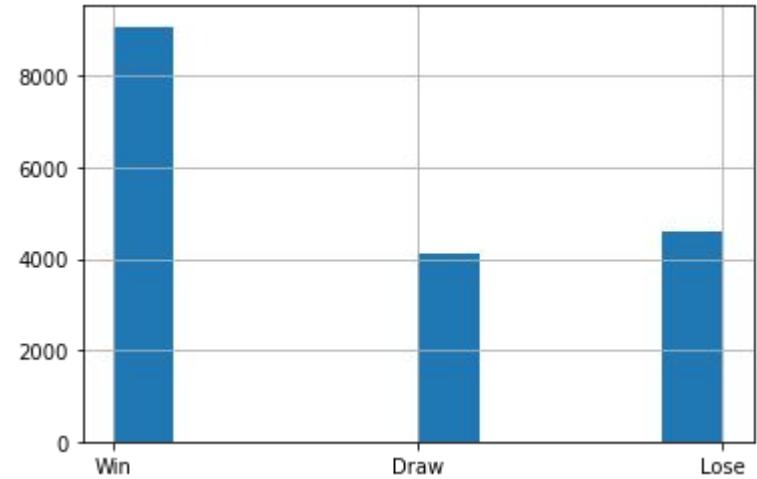
home_team	home_team_mean_offense_score
Argentina	88.38
Brazil	86.39
Spain	86.04
France	85.70
Netherlands	85.51
England	85.28
Italy	85.13
Portugal	84.97
Uruguay	83.70
Germany	83.67

overall offense average
Argentina
Brazil
Spain
France
Netherlands
Italy
England
Portugal
Uruguay
Germany



Task 2: Is there really such a
thing as home team
advantage?

Question 3: Is there really such a thing as home team advantage?



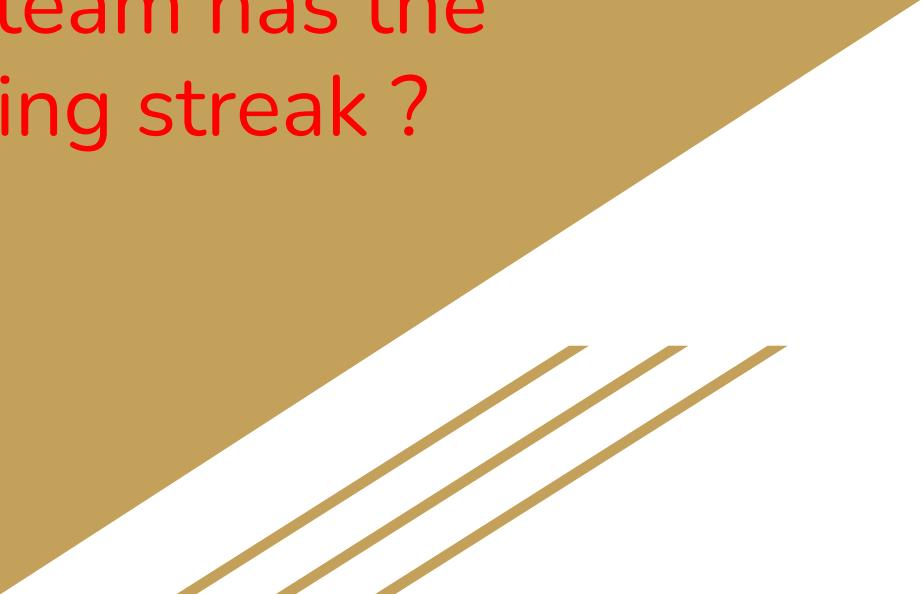
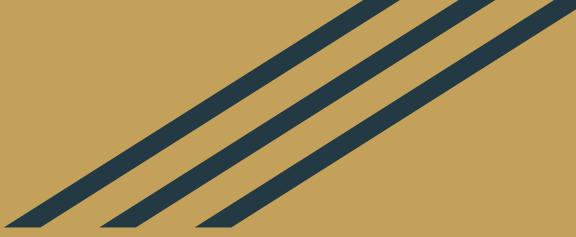
Question 3: Continued

Out[46]:

home_team	home_team_result
USA	289
Japan	207
Qatar	207
Saudi Arabia	203
France	189
Korea Republic	188
Oman	186
Thailand	186
United Arab Emirates	181
South Africa	180
Germany	173
England	167
Poland	164
Egypt	161
Tunisia	159
Estonia	159
Bahrain	158

Out[56]:

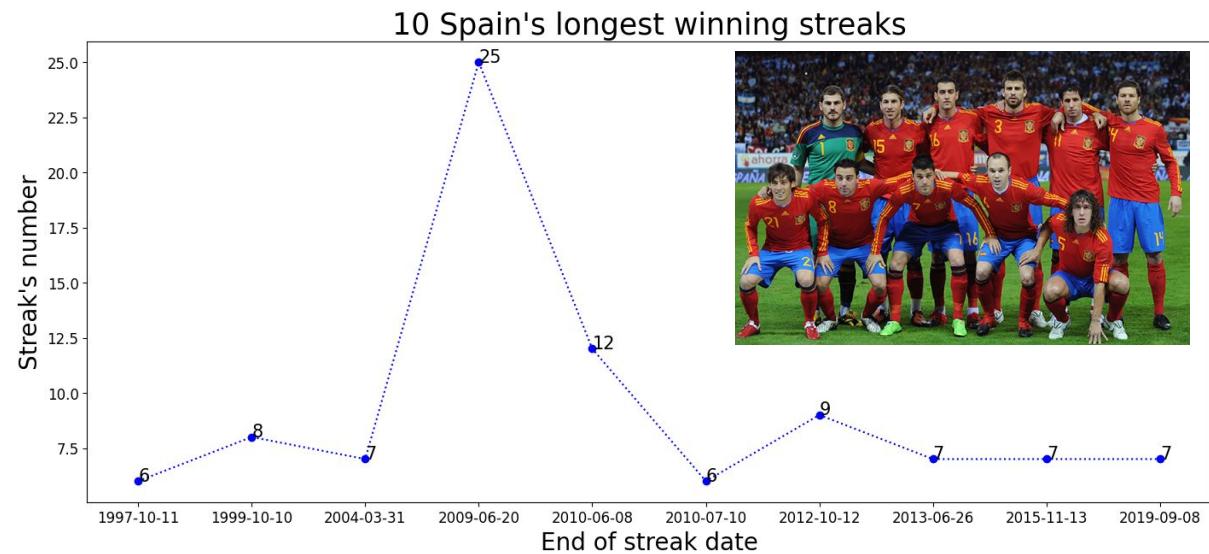
away_team	away team result
Zambia	176
Finland	175
Estonia	169
Sweden	162
Costa Rica	160
Paraguay	154
Norway	148
Uruguay	147
Lithuania	146
Jamaica	146
Syria	142
Germany	142
Brazil	140
Trinidad and Tobago	140
Turkey	139
Czech Republic	136
Algeria	135



Task 3 : What team has the
longest winning streak ?

Top 10 longest winning streak ?

Teams	Longest hot streak	Last win of the streak
Spain	25	2009-06-20
Italy	15	2021-07-10
Brazil	14	1997-12-12
France	14	2004-02-18
Australia	13	1997-07-06
Morocco	12	2021-12-07
Belgium	12	2020-09-07
Saudi Arabia	11	2001-07-23
Sweden	11	2001-10-07
Russia	11	1996-03-27





Task 4: Most Wins?

HOME

ALL MATCHES

home_team	home_team_result
USA	194
Mexico	190
Brazil	181
Saudi Arabia	164
Egypt	163
Japan	154
France	145
Spain	145
Korea Republic	140
Germany	139

FRIENDLY/EXHIBITION

home_team	home_team_result
Mexico	87
Egypt	74
Saudi Arabia	72
France	70
USA	69
Korea Republic	66
Oman	64
Brazil	62
Spain	58
Tunisia	57

WORLD CUP QUALIFICATION

home_team	home_team_result
USA	47
Australia	46
Saudi Arabia	46
Costa Rica	43
Argentina	43
Mexico	41
Uruguay	41
IR Iran	40
Ecuador	40
Brazil	39

WORLD CUP ELIMINATION

home_team	home_team_result
Brazil	24
Germany	21
Argentina	17
France	14
Netherlands	12
Belgium	10
Italy	10
Spain	9
Portugal	7
Korea Republic	7

AWAY

ALL MATCHES

	away team result
away_team	
Brazil	120
Spain	101
Germany	101
Zambia	88
Argentina	86
France	85
IR Iran	84
Sweden	83
Mexico	82
Netherlands	82

FRIENDLY/EXHIBITION

	away team result
away_team	
Brazil	68
Argentina	43
Uruguay	37
Colombia	33
Germany	33
Jamaica	31
Poland	30
Zambia	30
Sweden	30
Norway	30

WORLD CUP QUALIFICATION

	away team result
away_team	
Japan	30
IR Iran	29
Argentina	26
Spain	25
Netherlands	25
Germany	25
Mexico	24
England	23
Portugal	23
Korea Republic	22

WORLD CUP ELIMINATION

	away team result
away_team	
Croatia	10
Spain	9
Brazil	8
Germany	8
Netherlands	7
England	7
Mexico	6
France	6
Uruguay	5
Denmark	5



Task 5: Better to have a strong
Defense? Midfield? Offense? (in
terms of winning the WC)

Let's Compare our Results to grasp

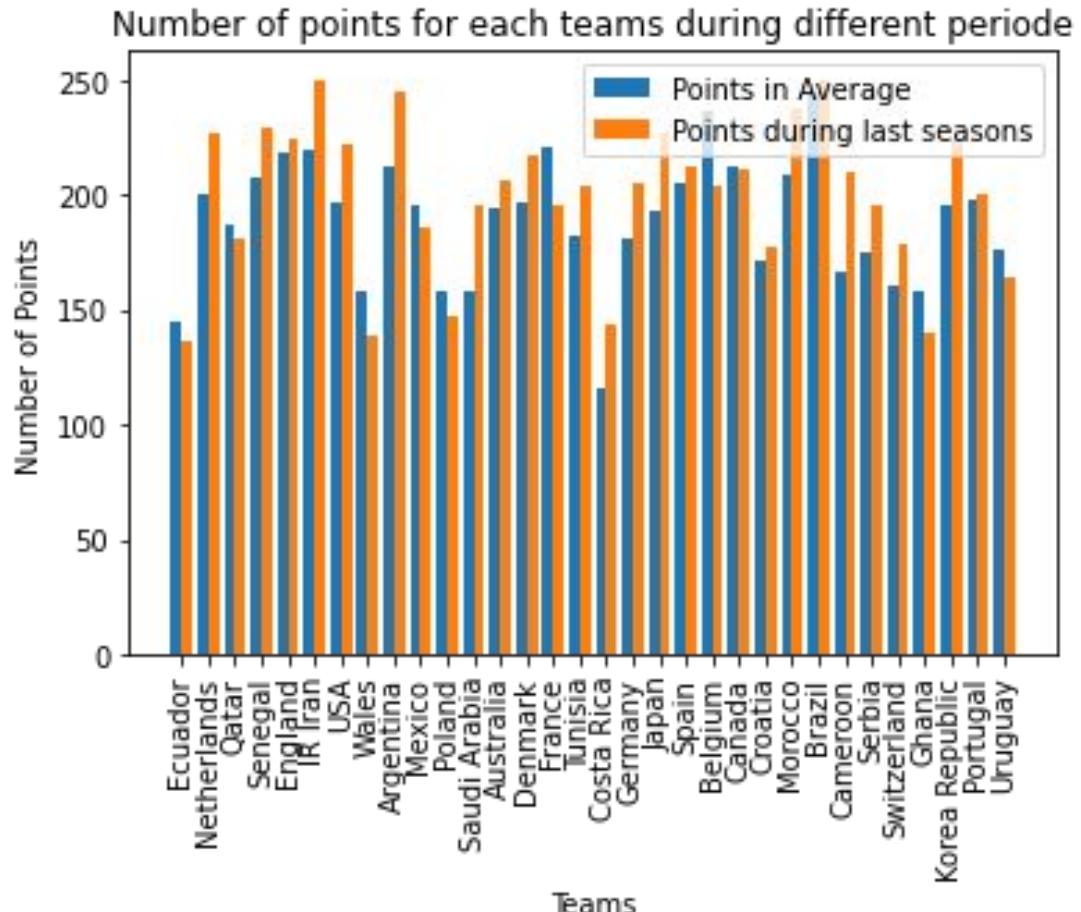
overall defense average		midfield average		overall offense average	
Brazil	85.94	Spain	87.12	Argentina	88.25
Spain	85.51	France	86.16	Brazil	86.55
Italy	85.04	Germany	85.80	Spain	85.87
England	84.69	Brazil	85.35	France	85.70
Germany	84.69	England	84.52	Netherlands	85.47
France	84.09	Argentina	84.44	Italy	85.29
Argentina	83.00	Italy	84.20	England	85.09
Portugal	82.89	Netherlands	83.70	Portugal	84.93
Belgium	81.23	Portugal	83.59	Uruguay	83.60
Netherlands	80.73	Belgium	81.52	Germany	83.59

→

away team result		home_team_result	
away_team	home_team	home_team	home_team_result
Croatia	10	Brazil	24
Spain	9	Germany	21
Brazil	8	Argentina	17
Germany	8	France	14
Netherlands	7	Netherlands	12
England	7	Belgium	10
Mexico	6	Italy	10
France	6	Spain	9
Uruguay	5	Portugal	7
Denmark	5	Korea Republic	7

Is The Team Hot/Cold/Regular ?

	TEAM	forme
59	Ecuador	Regular
134	Netherlands	Hot
154	Qatar	Regular
163	Senegal	Hot
62	England	Regular
89	IR Iran	Hot
198	USA	Hot
207	Wales	Cold
8	Argentina	Hot
124	Mexico	Regular
151	Poland	Regular
161	Saudi Arabia	Hot
11	Australia	Regular
55	Denmark	Hot
71	France	Cold
193	Tunisia	Hot
48	Costa Rica	Hot
75	Germany	Hot
97	Japan	Hot
174	Spain	Regular
19	Belgium	Cold
36	Canada	Regular
49	Croatia	Regular
129	Morocco	Hot
27	Brazil	Regular
35	Cameroon	Hot
164	Serbia	Hot
182	Switzerland	Hot
76	Ghana	Cold
102	Korea Republic	Hot
152	Portugal	Regular
202	Uruguay	Regular



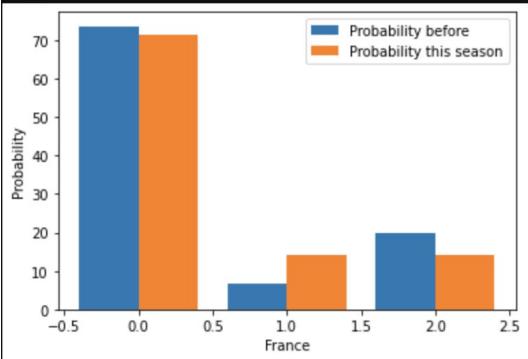
How to know who will win ?



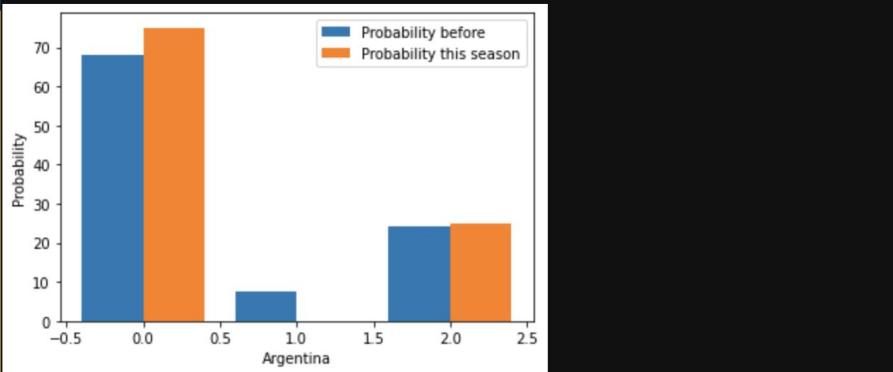
Win HOME: World class against AWAY: World class	43.626062322946176
Lose HOME: World class against AWAY: World class	32.01133144475921
Draw HOME: World class against AWAY: World class	24.362606232294617
Win HOME: World class against AWAY: Really Good	57.99648506151143
Lose HOME: World class against AWAY: Really Good	19.507908611599298
Draw HOME: World class against AWAY: Really Good	22.495606326889277
Win HOME: World class against AWAY: Average	66.66666666666666
Lose HOME: World class against AWAY: Average	12.444444444444445
Draw HOME: World class against AWAY: Average	20.888888888888889
Win HOME: World class against AWAY: Bad	74.3142144638404
Lose HOME: World class against AWAY: Bad	7.4812967581047385
Draw HOME: World class against AWAY: Bad	18.204488778054863
Win HOME: World class against AWAY: Really Bad	80.31088082901555
Lose HOME: World class against AWAY: Really Bad	6.7357512953367875
Draw HOME: World class against AWAY: Really Bad	12.953367875647666
Win HOME: World class against AWAY: Worst	93.33333333333333
Lose HOME: World class against AWAY: Worst	1.111111111111112
Draw HOME: World class against AWAY: Worst	5.555555555555555
Win HOME: Really Good against AWAY: World class	29.282868525896415
Lose HOME: Really Good against AWAY: World class	42.0318725096016
Draw HOME: Really Good against AWAY: World class	28.68525896414343
Win HOME: Really Good against AWAY: Really Good	45.774647887323944
Lose HOME: Really Good against AWAY: Really Good	26.877934272300468
Draw HOME: Really Good against AWAY: Really Good	27.347417840375588
Win HOME: Really Good against AWAY: Average	49.75124378109453
Lose HOME: Really Good against AWAY: Average	24.875621890547265
Draw HOME: Really Good against AWAY: Average	25.37313432835821
Win HOME: Really Good against AWAY: Bad	64.98829039812647
Lose HOME: Really Good against AWAY: Bad	13.700234192037472
Draw HOME: Really Good against AWAY: Bad	21.311475409836063

How to know who will win ?

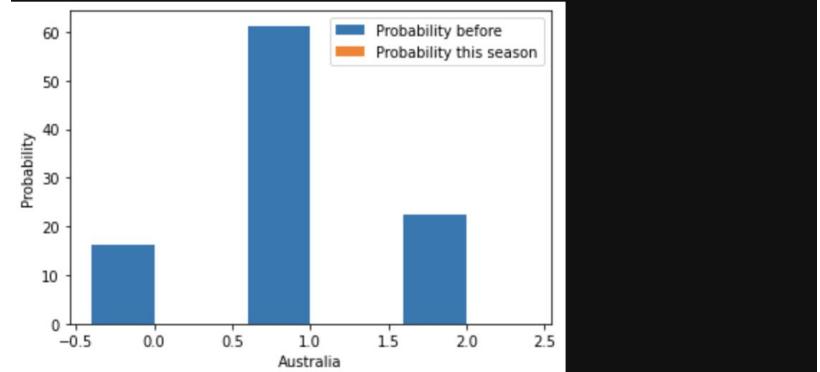
[[73.68421052631578, 6.578947368421052, 19.736842105263158], [71.42857142857143, 14.285714285714285, 14.285714285714285]]
[72.55639097744361, 10.432330827067668, 17.01127819548872] France
Bad



[[67.94871794871796, 7.6923076923076925, 24.358974358974358], [75.0, 0.0, 25.0]]
[71.47435897435898, 3.8461538461538463, 24.67948717948718] Argentina
Bad



[[16.129032258064516, 61.29032258064516, 22.58064516129032], [0, 0, 0]]
[16.129032258064516, 61.29032258064516, 22.58064516129032] Australia
World class



Start of the Simulation



Start of the Simulation





Thank you for
Listening :)