SPORTS ANALYSIS: FOOTBALL PLAYERS

This analysis would be concentrated around:-

- ### Data Cleaning and Preprocessing
- ### players and their corresponding skills and attributes that make them stand out.
- ### players and their respective clubs and
- ### players and their countries only.

Importing the Libraries we'll be using

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#Turning off the warnings
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
#To display all the columns of the dataframe
pd.set_option('display.max_columns',70)
```

In [3]:
```
#Reading the FIFA20 dataset
fifa = pd.read_csv("FIFA22_official_data.csv")
```

UNDERSTANDING THE DATA

In [4]:
```
#Dimension of the dataset
fifa.shape
```

Out[4]:
```
(16710, 65)
```

16710 datapoints or player details and 65 features/columns

In [5]:
```
#head of the dataset
```

```python
fifa.head()
```

Out[5]:

	ID	Name	Age	Photo	Nationality	
0	212198	Bruno Fernandes	26	https://cdn.sofifa.com/players/212/198/22_60.png	Portugal	https://cdn.sofifa.co
1	209658	L. Goretzka	26	https://cdn.sofifa.com/players/209/658/22_60.png	Germany	https://cdn.sofifa.cc
2	176580	L. Suárez	34	https://cdn.sofifa.com/players/176/580/22_60.png	Uruguay	https://cdn.sofifa.cc
3	192985	K. De Bruyne	30	https://cdn.sofifa.com/players/192/985/22_60.png	Belgium	https://cdn.sofifa.cc
4	224334	M. Acuña	29	https://cdn.sofifa.com/players/224/334/22_60.png	Argentina	https://cdn.sofifa.co

In [6]:

```python
#columns of the dataframe
fifa.columns
```

Out[6]:

```
Index(['ID', 'Name', 'Age', 'Photo', 'Nationality', 'Flag', 'Overall',
       'Potential', 'Club', 'Club Logo', 'Value', 'Wage', 'Special',
       'Preferred Foot', 'International Reputation', 'Weak Foot',
       'Skill Moves', 'Work Rate', 'Body Type', 'Real Face', 'Position',
       'Jersey Number', 'Joined', 'Loaned From', 'Contract Valid Until',
       'Height', 'Weight', 'Crossing', 'Finishing', 'HeadingAccuracy',
       'ShortPassing', 'Volleys', 'Dribbling', 'Curve', 'FKAccuracy',
       'LongPassing', 'BallControl', 'Acceleration', 'SprintSpeed', 'Agility',
       'Reactions', 'Balance', 'ShotPower', 'Jumping', 'Stamina', 'Strength',
       'LongShots', 'Aggression', 'Interceptions', 'Positioning', 'Vision',
       'Penalties', 'Composure', 'Marking', 'StandingTackle', 'SlidingTackle',
       'GKDiving', 'GKHandling', 'GKKicking', 'GKPositioning', 'GKReflexes',
       'Best Position', 'Best Overall Rating', 'Release Clause',
       'DefensiveAwareness'],
      dtype='object')
```

In [7]:

```python
#INFO() function
fifa.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16710 entries, 0 to 16709
Data columns (total 65 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   ID                        16710 non-null  int64 
 1   Name                      16710 non-null  object
 2   Age                       16710 non-null  int64 
 3   Photo                     16710 non-null  object
```

```
 4   Nationality               16710 non-null  object 
 5   Flag                      16710 non-null  object 
 6   Overall                   16710 non-null  int64  
 7   Potential                 16710 non-null  int64  
 8   Club                      16446 non-null  object 
 9   Club Logo                 16710 non-null  object 
 10  Value                     16710 non-null  object 
 11  Wage                      16710 non-null  object 
 12  Special                   16710 non-null  int64  
 13  Preferred Foot            16710 non-null  object 
 14  International Reputation  16710 non-null  float64
 15  Weak Foot                 16710 non-null  float64
 16  Skill Moves               16710 non-null  float64
 17  Work Rate                 16710 non-null  object 
 18  Body Type                 16681 non-null  object 
 19  Real Face                 16681 non-null  object 
 20  Position                  16684 non-null  object 
 21  Jersey Number             16684 non-null  float64
 22  Joined                    15198 non-null  object 
 23  Loaned From               1132 non-null   object 
 24  Contract Valid Until      16359 non-null  object 
 25  Height                    16710 non-null  object 
 26  Weight                    16710 non-null  object 
 27  Crossing                  16710 non-null  float64
 28  Finishing                 16710 non-null  float64
 29  HeadingAccuracy           16710 non-null  float64
 30  ShortPassing              16710 non-null  float64
 31  Volleys                   16673 non-null  float64
 32  Dribbling                 16710 non-null  float64
 33  Curve                     16673 non-null  float64
 34  FKAccuracy                16710 non-null  float64
 35  LongPassing               16710 non-null  float64
 36  BallControl               16710 non-null  float64
 37  Acceleration              16710 non-null  float64
 38  SprintSpeed               16710 non-null  float64
 39  Agility                   16673 non-null  float64
 40  Reactions                 16710 non-null  float64
 41  Balance                   16673 non-null  float64
 42  ShotPower                 16710 non-null  float64
 43  Jumping                   16673 non-null  float64
 44  Stamina                   16710 non-null  float64
 45  Strength                  16710 non-null  float64
 46  LongShots                 16710 non-null  float64
 47  Aggression                16710 non-null  float64
 48  Interceptions             16702 non-null  float64
 49  Positioning               16702 non-null  float64
 50  Vision                    16673 non-null  float64
 51  Penalties                 16710 non-null  float64
 52  Composure                 16459 non-null  float64
 53  Marking                   892 non-null    float64
 54  StandingTackle            16710 non-null  float64
 55  SlidingTackle             16673 non-null  float64
 56  GKDiving                  16710 non-null  float64
 57  GKHandling                16710 non-null  float64
 58  GKKicking                 16710 non-null  float64
 59  GKPositioning             16710 non-null  float64
 60  GKReflexes                16710 non-null  float64
 61  Best Position             16710 non-null  object 
 62  Best Overall Rating       16710 non-null  float64
 63  Release Clause            14961 non-null  object 
```

```
 64  DefensiveAwareness        15818 non-null  float64
dtypes: float64(40), int64(5), object(20)
memory usage: 8.3+ MB
```

In [8]:

```
fifa.isnull().sum().sort_values(ascending = False)
```

Out[8]:

```
Marking               15818
Loaned From           15578
Release Clause         1749
Joined                 1512
DefensiveAwareness      892
                        ...
ShortPassing              0
Name                      0
FKAccuracy                0
LongPassing               0
Dribbling                 0
Length: 65, dtype: int64
```

I can find there are some missing values for columns such as `Club,Joined` , `LoanedFrom` , `Release Clause` , `Marking` , `Joined` and a few others.

In [9]:

```
#Some initial stats for the df
fifa.describe()
```

Out[9]:

	ID	Age	Overall	Potential	Special	International Reputation	Weak Fo
count	16710.000000	16710.000000	16710.000000	16710.000000	16710.000000	16710.000000	16710.0000
mean	220560.467923	25.727409	67.646320	72.572292	1652.497307	1.169958	3.0081
std	38496.607959	5.048910	6.457695	5.688085	257.271696	0.485305	0.6817
min	27.000000	16.000000	28.000000	38.000000	571.000000	1.000000	1.0000
25%	203891.250000	22.000000	63.000000	69.000000	1525.000000	1.000000	3.0000
50%	229253.000000	25.000000	68.000000	72.000000	1687.000000	1.000000	3.0000
75%	245368.750000	29.000000	72.000000	76.000000	1826.000000	1.000000	3.0000
max	264704.000000	54.000000	93.000000	95.000000	2341.000000	5.000000	5.0000

In [10]:

```
#Copy of our pulled in df
fifa_copy = fifa.copy()
```

SOME CLEANING AND PREPROCESSING

In [11]:

```
fifa.head()
```

Out[11]:

	ID	Name	Age	Photo	Nationality

	ID	Name	Age	Photo	Nationality	
0	212198	Bruno Fernandes	26	https://cdn.sofifa.com/players/212/198/22_60.png	Portugal	https://cdn.sofifa.cc
1	209658	L. Goretzka	26	https://cdn.sofifa.com/players/209/658/22_60.png	Germany	https://cdn.sofifa.cc
2	176580	L. Suárez	34	https://cdn.sofifa.com/players/176/580/22_60.png	Uruguay	https://cdn.sofifa.cc
3	192985	K. De Bruyne	30	https://cdn.sofifa.com/players/192/985/22_60.png	Belgium	https://cdn.sofifa.cc
4	224334	M. Acuña	29	https://cdn.sofifa.com/players/224/334/22_60.png	Argentina	https://cdn.sofifa.cc

- I am going to delete some `columns` which might not be adding any value adds to our analysis going forward.
- All the url columns and index columns we will try removing right now.

In [12]:

```
#Dropping some of the columns - ID, Photo, Flag, Club Logo
fifa.drop(['ID','Photo','Flag','Club Logo'],axis=1,inplace=True)
```

In [13]:

```
#Dropping Real Face column
fifa.drop(['Real Face'],axis=1,inplace=True)
```

In [14]:

```
#Filtering for rows which have Loaned From column not NULL
fifa.loc[~fifa['Loaned From'].isnull()][:5]
```

Out[14]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Inter Re
9	A. Griezmann	30	France	85	85	Atlético de Madrid	€53M	€220K	2259	Left	
43	Saúl	26	Spain	82	85	Chelsea	€39.5M	€68K	2199	Left	
72	A. Florenzi	30	Italy	81	81	AC Milan	€22M	€58K	2178	Right	

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Inte Re
121	M. Pjanić	31	Bosnia and Herzegovina	82	82	Beşiktaş JK	€25M	€155K	2147	Right	
155	J. Kucka	34	Slovakia	77	77	Watford	€4.9M	€33K	2135	Right	

You can find some of the Loaned players in the list above.
Players like Griezmann and Pjanic who are on loan from Barcelona to Atletico de Madrid and Besiktas respectively.

I guess it would be better in this case not to handle the missing values for the columns, because we might end up missing out so many players in that case.
For columns such as `Loaned From` might be many for so many players as well, since it would be a direct contract with the player.

- ### PLAYER VALUE AND WAGE
- Checking whether all values are defined in Euros or not

In [15]:
```
fifa.loc[fifa['Value'].str.startswith('€')].shape[0]
```

Out[15]:
```
16710
```

Checking the same for Wages.

In [16]:
```
fifa.loc[fifa['Wage'].str.startswith('€')].shape[0]
```

Out[16]:
```
16710
```

So the Value and Wage of all the players are defined in Euros and also object type, which we can try converting accordingly and try making it as a numeric column.

In [17]:
```
#Splitting the value column to get just the numeric
fifa['Value'] = fifa['Value'].str.split('€')
fifa['Value'] = fifa['Value'].apply(lambda x:x[1])
```

In [18]:
```
#Splitting the wage column to get just the numeric
fifa['Wage'] = fifa['Wage'].str.split('€')
fifa['Wage'] = fifa['Wage'].apply(lambda x:x[1])
```

In [19]:
```
#Converting the player value in thousand Euros to Million Euros and then stripping the
fifa_value_K = fifa.loc[fifa['Value'].str.endswith('K')]
fifa_value_K['Value'] = fifa_value_K['Value'].apply(lambda x: x[:-1])
```

```
fifa_value_K['Value'] = fifa_value_K['Value'].astype('float64')
fifa_value_K['Value'] = fifa_value_K['Value'] / 1000
```

In [20]:
```
#Stripping the end denote for Million Euros Player value
fifa_value_M = fifa.loc[fifa['Value'].str.endswith('M')]
fifa_value_M['Value'] = fifa_value_M['Value'].apply(lambda x: x[:-1])
fifa_value_M['Value'] = fifa_value_M['Value'].astype('float64')
```

In [21]:
```
#Converting the player wage in thousand Euros to Million Euros and then stripping the e
fifa_value_K['Wage'] = fifa_value_K['Wage'].apply(lambda x: x[:-1] if x.endswith('K') e
fifa_value_K['Wage'] = fifa_value_K['Wage'].astype('float64')
fifa_value_K['Wage'] = fifa_value_K['Wage'] / 1000
```

In [22]:
```
fifa_value_M.loc[fifa_value_M['Wage'].str.endswith("M")]
```

Out[22]:

Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	W F

We can find that there are no more players who are having values and wages in euros having wages also in Millions. So we will go with thousand converted to millions

In [23]:
```
#Converting the player wage in thousand Euros to Million Euros and then stripping the e
fifa_value_M['Wage'] = fifa_value_M['Wage'].apply(lambda x: x[:-1] if x.endswith('K') e
fifa_value_M['Wage'] = fifa_value_M['Wage'].astype('float64')
fifa_value_M['Wage'] = fifa_value_M['Wage'] / 1000
```

In [24]:
```
#Concatenating both the splitted up dataframes
fifa2 = pd.concat([fifa_value_M,fifa_value_K])
```

In [25]:
```
fifa2.shape
```

Out[25]:
```
(16366, 60)
```

It looks like the resultant dataframe `fifa2` after our preprocessing, is having less number of records than our initial dataframe. These are records/rows that were having player Values provided as 0 or any other value which were neither in thousands nor Millions. We will proceed with this dataset forour analysis.

In [26]:
```
fifa2[fifa2['Value']==194]["Name"]
```

Out[26]:
```
82    K. Mbappé
Name: Name, dtype: object
```

In [27]:
```
fifa2['Value'] = pd.to_numeric(fifa2['Value'])
fifa2['Wage'] = pd.to_numeric(fifa2['Wage'])
```

CLEANING THE POSITION FEATURE

We can find some HTML code getting in(probably when the data is scrapped) with the Position column. We will clean those.

In [28]:
```
fifa2['Position'] = fifa2['Position'].str.split(">")
fifa2['Position'] = fifa2['Position'].apply(lambda x:x[1])
```

In [29]:
```
fifa2.head()
```

Out[29]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Interna Repu
0	Bruno Fernandes	26	Portugal	88	89	Manchester United	107.5	0.250	2341	Right	
1	L. Goretzka	26	Germany	87	88	FC Bayern München	93.0	0.140	2314	Right	
2	L. Suárez	34	Uruguay	88	88	Atlético de Madrid	44.5	0.135	2307	Right	
3	K. De Bruyne	30	Belgium	91	91	Manchester City	125.5	0.350	2304	Right	
4	M. Acuña	29	Argentina	84	84	Sevilla FC	37.0	0.045	2292	Left	

CLEANING THE WEIGHT COLUMN

All the player weights are mentioned in lbs and is object type. We will remove the suffixes and convert it into integer.

In [30]:
```
fifa2['Weight'] = fifa2['Weight'].apply(lambda x : x[:-3])
fifa2['Weight'] = fifa2['Weight'].astype('int64')
```

ANALYSIS ON THE PREFERED FOOT

In [31]:
```
foot = fifa2['Preferred Foot'].value_counts()
foot
```

Out[31]:
```
Right    12352
Left      4014
Name: Preferred Foot, dtype: int64
```

As we expected, the Left Footers are fewer! there are only 4014 left footers in our player list.

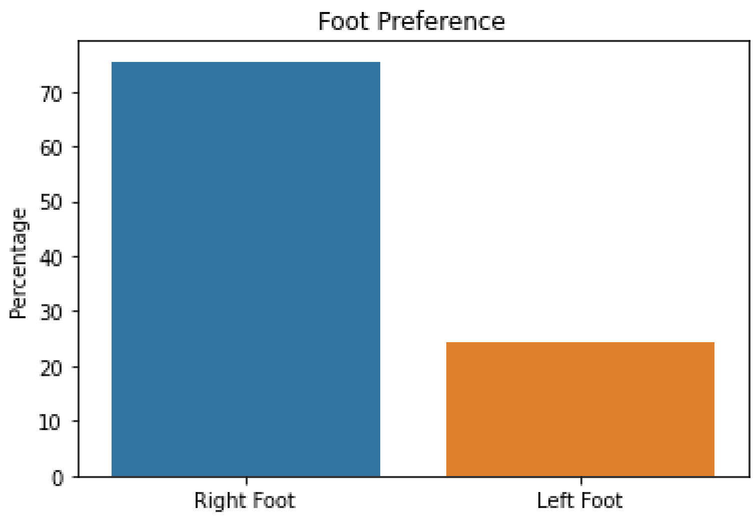
Let's check the percentage %

In [32]:
```
foot_right = foot[0]/fifa2['Preferred Foot'].count()*100
foot_left = foot[1]/fifa2['Preferred Foot'].count()*100
foot_df = pd.DataFrame({'Percentage':[foot_right,foot_left]},index=['Right Foot','Left
foot_df.style.background_gradient(cmap='Purples')
```

Out[32]:

	Percentage
Right Foot	75.473543
Left Foot	24.526457

In [33]:
```
#Barplot for the classes
plt.title("Foot Preference")
sns.barplot(x=foot_df.index,y=foot_df['Percentage'])
plt.show()
```



This clearly indicates the dominance of Right footers in the Football. There are just a small percentage of Left Footers, hence their importance.

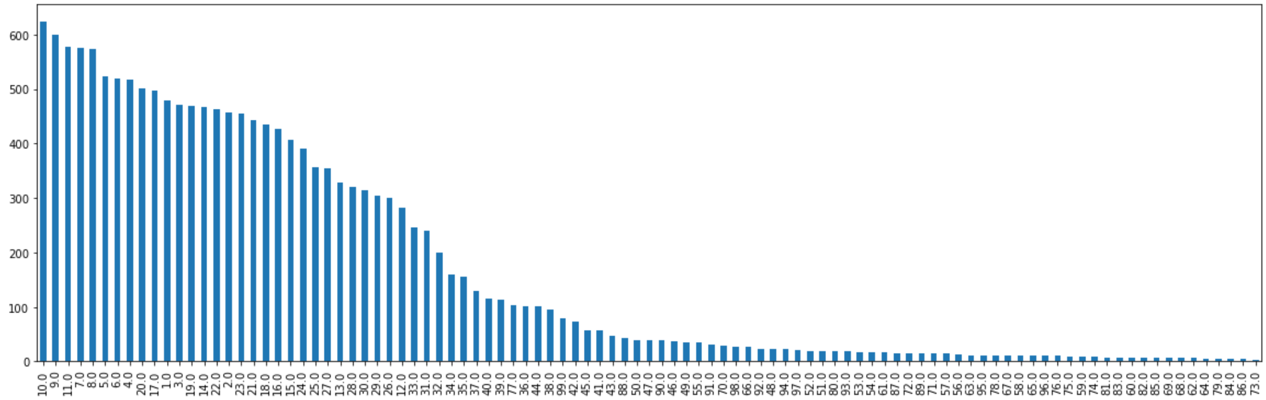
AVERAGE AGE OF ALL THE PLAYERS

In [34]:
```
fifa2.Age.mean()
```

Out[34]: 25.633752902358548

The most used Jersey Number

In [35]:
```
plt.figure(figsize = (20, 6))
fifa2["Jersey Number"].value_counts().plot(kind='bar');
```



- `Number 10` **is the most worn number!!!**

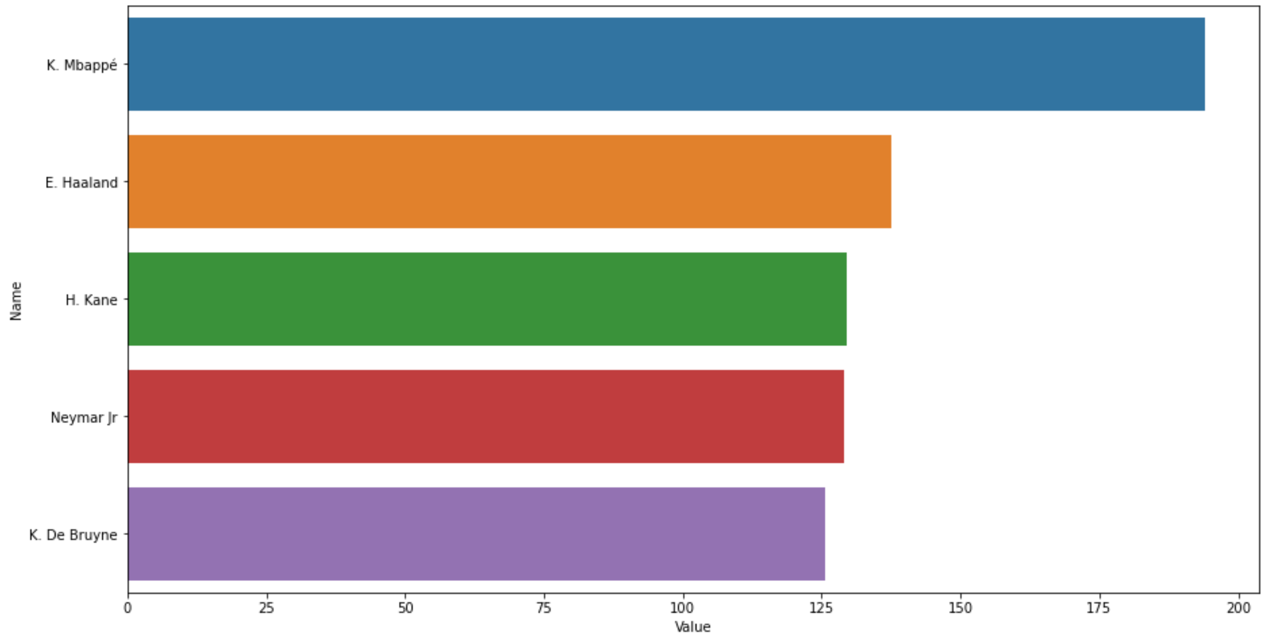
TOP 5 MOST VALUABLE PLAYERS

In [36]:
```
fifa2.sort_values(['Value'],ascending=False)[:5]
```

Out[36]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Interr Rep
82	K. Mbappé	22	France	91	95	Paris Saint-Germain	194.0	0.23	2175	Right	
251	E. Haaland	20	Norway	88	93	Borussia Dortmund	137.5	0.11	2102	Left	
39	H. Kane	27	England	90	90	Tottenham Hotspur	129.5	0.24	2205	Right	
64	Neymar Jr	29	Brazil	91	91	Paris Saint-Germain	129.0	0.27	2183	Right	
3	K. De Bruyne	30	Belgium	91	91	Manchester City	125.5	0.35	2304	Right	

In [37]:
```
top_value = fifa2.sort_values(by = ['Value'], ascending = False)
plt.figure(figsize = (15, 8))
sns.barplot(x = top_value ["Value"], y = top_value["Name"][:5]);
```



In [ ]:

TOP 10 PLAYERS WITH HIGHEST OVERALL RATINGS

In [38]:
```
fifa_overall = fifa2.sort_values(['Overall'],ascending=False)[:10]
fifa_overall[['Name','Overall','Potential','Club','Preferred Foot','Position', 'Nationa
```

Out[38]:

	Name	Overall	Potential	Club	Preferred Foot	Position	Nationality
29	L. Messi	93	93	Paris Saint-Germain	Left	RW	Argentina
33	R. Lewandowski	92	92	FC Bayern München	Right	ST	Poland
64	Neymar Jr	91	91	Paris Saint-Germain	Right	LW	Brazil
3	K. De Bruyne	91	91	Manchester City	Right	RCM	Belgium
36	Cristiano Ronaldo	91	91	Manchester United	Right	ST	Portugal
82	K. Mbappé	91	95	Paris Saint-Germain	Right	ST	France
14244	J. Oblak	91	93	Atlético de Madrid	Right	GK	Slovenia
39	H. Kane	90	90	Tottenham Hotspur	Right	ST	England
12350	M. Neuer	90	90	FC Bayern München	Right	GK	Germany
13890	M. ter Stegen	90	92	FC Barcelona	Right	GK	Germany

These are the top 10 ->

`L. Messi` , `R. Lewandowski` , `Neymar Jr` , `K. De Bruyne` , `Cristiano Ronaldo` , `K. Mbappé` , `J. Oblak` , `H. Kane` , `M. ter Stegen` , `M. Neuer`

Interesting FACT ---->>> 2 German Goalkeepers in the top 10 in `M. ter Stegen` and `M. Neuer` .

And also **I WONDER HOW PSG IS LOSING SO MANY MATCHES -> 3 OF THEIR BEST ARE IN THE TOP10 TOO**

TOP 10 PLAYERS WITH HIGHEST POTENTIAL

In [39]:
```
fifa_potential = fifa2.sort_values(['Potential'],ascending=False)[:10]
fifa_potential[['Name','Overall','Potential','Club','Preferred Foot','Position']].style
```

Out[39]:

	Name	Overall	Potential	Club	Preferred Foot	Position
82	K. Mbappé	91	95	Paris Saint-Germain	Right	ST
251	E. Haaland	88	93	Borussia Dortmund	Left	RS
29	L. Messi	93	93	Paris Saint-Germain	Left	RW
14608	G. Donnarumma	89	93	Paris Saint-Germain	Right	GK
14244	J. Oblak	91	93	Atlético de Madrid	Right	GK
17	F. de Jong	87	92	FC Barcelona	Right	RCM
24	T. Alexander-Arnold	87	92	Liverpool	Right	RB
33	R. Lewandowski	92	92	FC Bayern München	Right	ST
588	K. Havertz	84	92	Chelsea	Left	LW
13890	M. ter Stegen	90	92	FC Barcelona	Right	GK

These are the top 10:

`K. Mbappé` , `E. Haaland` , `L. Messi` , `G. Donnarumma` , `J. Oblak` , `F.de Jong` , `T.Alexander-Arnold` , `R. Lewandowski` , `K. Havertz` , `M. ter Stegen`

We've always been viewing the top guys! We will see the bottom ones now.

BOTTOM 10 IN POTENTIAL AND OVERALL

In [40]:
```
fifa2.sort_values(['Overall'],ascending=True)[:10]
```

Out[40]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot
15999	16 E. Redman	18	Wales	44	57	Newport County	0.05	0.002	1089	Right
15593	18 T. Käßemodel	28	Germany	46	46	FC Erzgebirge Aue	0.03	0.002	1174	Right
15685	15 T. Fletcher	19	England	46	52	Wycombe Wanderers	0.02	0.002	1157	Right

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot
16708	A. Shaikh	18	India	47	67	ATK Mohun Bagan FC	0.11	0.500	754	Right
15244	R. Gallagher	20	Republic of Ireland	47	61	Finn Harps	0.11	0.500	1255	Right
15283	Wang Zhen'ao	21	China PR	47	57	Dalian Professional Football Club	0.11	0.002	1246	Right
14787	Hu Xingyu	19	China PR	47	55	Chongqing Liangjiang Athletic	0.10	0.800	1352	Right
14720	N. Saliba	17	Canada	47	69	Club de Foot Montréal	0.15	0.500	1363	Right
15434	L. Rudden	19	Republic of Ireland	47	60	Finn Harps	0.11	0.500	1210	Right
15573	Wang Shilong	20	China PR	47	60	Guangzhou FC	0.11	0.002	1178	Right

In [41]:
```python
fifa2.sort_values(['Potential'],ascending=True)[:10]
```

Out[41]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot
15593	18 T. Käßemodel	28	Germany	46	46	FC Erzgebirge Aue	0.030	0.002	1174	Right
15733	21 M. Wright	33	England	47	47	Crawley Town	0.025	0.001	1148	Left
14188	J. Russell	36	Republic of Ireland	49	49	Sligo Rovers	0.015	0.500	1419	Right
15027	19 B. Barry-Murphy	39	Republic of Ireland	49	49	Rochdale	0.007	0.001	1309	Left
14304	Tan Chun Lok	25	Hong Kong	49	51	Guangzhou City	0.060	0.002	1408	Right

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot
15049	S. Haokip	28	India	51	51	SC East Bengal FC	0.060	0.500	1305	Right
16388	Ma Zhen	23	China PR	48	52	Shanghai Shenhua FC	0.050	0.001	964	Right
15685	15 T. Fletcher	19	England	46	52	Wycombe Wanderers	0.020	0.002	1157	Right
15601	M. McChrystal	37	Northern Ireland	52	52	Derry City	0.015	0.500	1172	Left
16203	K. Singh	34	India	53	53	Chennaiyin FC	0.015	0.500	1032	Right

Top Rated Players by Nationality

- ### Nigerian Players

In [42]:
```
fifa2Nigeria = fifa2.loc[fifa2.Nationality=='Nigeria']
fifa2Nigeria.sort_values(['Overall'],ascending=False)[:5]
```

Out[42]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Interı Rep
218	W. Ndidi	24	Nigeria	85	88	Leicester City	66.5	0.120	2113	Right	
1425	V. Osimhen	22	Nigeria	80	87	Napoli	43.0	0.066	1961	Right	
3795	P. Onuachu	27	Nigeria	79	80	KRC Genk	20.5	0.023	1841	Right	
1826	K. Iheanacho	24	Nigeria	78	82	Leicester City	20.5	0.083	1934	Left	
338	V. Moses	30	Nigeria	78	78	Spartak Moskva	13.5	0.056	2081	Right	

- ### German Players

In [43]:
```
fifa2_Germany = fifa2.loc[fifa2.Nationality=='Germany']
```

```
fifa2_Germany.sort_values(['Overall'],ascending=False)[:5]
```

Out[43]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Inter Re
13890	M. ter Stegen	29	Germany	90	92	FC Barcelona	99.0	0.250	1444	Right	
12350	M. Neuer	35	Germany	90	90	FC Bayern München	13.5	0.086	1534	Right	
5	J. Kimmich	26	Germany	89	90	FC Bayern München	108.0	0.160	2283	Right	
116	T. Kroos	31	Germany	88	88	Real Madrid CF	75.0	0.310	2148	Right	
125	17 P. Lahm	32	Germany	88	88	FC Bayern München	29.5	0.200	2146	Right	

- ### Portuguese players

In [44]:

```
fifa2Portugal = fifa2.loc[fifa2.Nationality=='Portugal']
fifa2Portugal.sort_values(['Overall'],ascending=False)[:11]
```

Out[44]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	I
36	Cristiano Ronaldo	36	Portugal	91	91	Manchester United	45.0	0.270	2208	Right	
0	Bruno Fernandes	26	Portugal	88	89	Manchester United	107.5	0.250	2341	Right	
2754	Rúben Dias	24	Portugal	87	91	Manchester City	102.5	0.170	1886	Right	
170	Bernardo Silva	26	Portugal	86	86	Manchester City	74.0	0.200	2128	Left	
21	João Cancelo	27	Portugal	86	87	Manchester City	71.5	0.185	2227	Right	
60	Ricardo Pereira	27	Portugal	84	84	Leicester City	40.5	0.130	2186	Right	
80	R. Guerreiro	27	Portugal	84	84	Borussia Dortmund	40.5	0.079	2175	Left	

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Ir
439	André Silva	25	Portugal	84	85	RB Leipzig	51.0	0.110	2062	Right	
421	João Félix	21	Portugal	83	91	Atlético de Madrid	82.0	0.061	2065	Right	
14615	A. Lopes	30	Portugal	82	83	Olympique Lyonnais	23.0	0.060	1376	Left	
617	Palhinha	25	Portugal	82	87	Sporting CP	41.0	0.018	2033	Right	

- ### French Players

In [45]:

```
fifa2France = fifa2.loc[fifa2.Nationality=='France']
fifa2France.sort_values(['Overall'],ascending=False)[:11]
```

Out[45]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	I
82	K. Mbappé	22	France	91	95	Paris Saint-Germain	194.0	0.230	2175	Right	
71	N. Kanté	30	France	90	90	Chelsea	100.0	0.230	2179	Right	
201	K. Benzema	33	France	89	89	Real Madrid CF	66.0	0.350	2116	Right	
27	P. Pogba	28	France	87	87	Manchester United	79.5	0.220	2222	Right	
14644	H. Lloris	34	France	87	87	Tottenham Hotspur	13.5	0.125	1372	Left	
1228	K. Coman	25	France	86	87	FC Bayern München	81.0	0.120	1974	Right	
1897	R. Varane	28	France	86	88	Manchester United	68.5	0.180	1931	Right	
9	A. Griezmann	30	France	85	85	Atlético de Madrid	53.0	0.220	2259	Left	
22	L. Digne	27	France	84	84	Everton	40.5	0.110	2227	Left	

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	I
232	W. Ben Yedder	30	France	84	84	AS Monaco	41.5	0.088	2108	Right	
199	N. Fekir	27	France	84	84	Real Betis Balompié	45.0	0.042	2117	Left	

- ### English Players

In [46]:

```
fifa2France = fifa2.loc[fifa2.Nationality=='England']
fifa2France.sort_values(['Overall'],ascending=False)[:11]
```

Out[46]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	In
39	H. Kane	27	England	90	90	Tottenham Hotspur	129.5	0.240	2205	Right	
220	R. Sterling	26	England	88	89	Manchester City	107.5	0.290	2113	Right	
24	T. Alexander-Arnold	22	England	87	92	Liverpool	114.0	0.150	2227	Right	
855	J. Sancho	21	England	87	91	Manchester United	116.5	0.150	2007	Right	
154	J. Vardy	34	England	86	86	Leicester City	33.0	0.180	2135	Right	
127	M. Rashford	23	England	85	89	Manchester United	77.5	0.150	2146	Right	
53	K. Walker	31	England	85	85	Manchester City	39.0	0.170	2194	Right	
65	K. Trippier	30	England	84	84	Atlético de Madrid	36.5	0.074	2183	Right	
444	P. Foden	21	England	84	92	Manchester City	94.5	0.125	2060	Left	
167	L. Shaw	25	England	84	86	Manchester United	48.5	0.140	2130	Left	
1591	H. Maguire	28	England	84	86	Manchester United	42.5	0.155	1950	Right	

Let's look out for some veterans

In [47]:

```
#fifa2.loc[fifa2.Age >38]
fifa_potential = fifa2.sort_values(['Age'],ascending=False)[:11]
fifa_potential[['Name','Age','Overall','Potential','Club','Preferred Foot','Position']]
```

Out[47]:

	Name	Age	Overall	Potential	Club	Preferred Foot	Position
15242	19 O. Pérez	45	71	71	Pachuca	Right	SUB
15182	G. Buffon	43	80	80	Parma	Right	GK
14932	C. Lucchetti	43	72	72	Atlético Tucumán	Right	GK
16232	17 J. Jääskeläinen	41	69	69	Wigan Athletic	Right	RES
16240	S. Lukić	41	63	63	Varbergs BoIS FC	Right	GK
16239	K. Stamatopoulos	41	56	56	AIK	Left	RES
15713	J. Pinto	41	63	63	Sport Huancayo	Right	SUB
15907	A. Boruc	41	70	70	Legia Warszawa	Right	RES
15682	S. Torrico	41	73	73	San Lorenzo de Almagro	Right	GK
15365	Bracali	40	71	71	Boavista FC	Right	GK
16311	21 H. Sogahata	40	65	65	Kashima Antlers	Right	SUB

The most represented national teams on FIFA22

- Word Count and

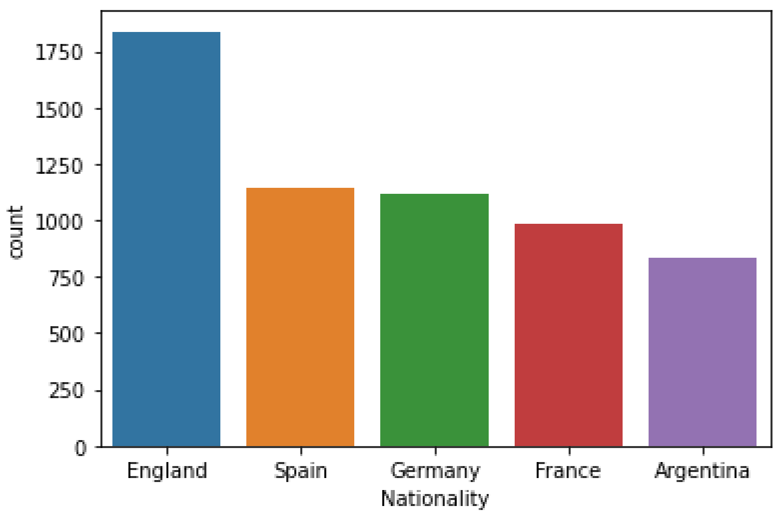
- Bar Chart

In [48]:

```
from wordcloud import WordCloud, STOPWORDS
text = " ".join(nationality for nationality in fifa2.Nationality)
word_cloud = WordCloud(collocations = False, background_color = 'white').generate(text)
plt.figure(figsize = (20, 6))
plt.imshow(word_cloud, interpolation = 'bilinear')
plt.axis("off")
plt.show()
```

In [49]:

```
sns.countplot(x="Nationality", data=fifa2, order=fifa2["Nationality"].value_counts().in
```



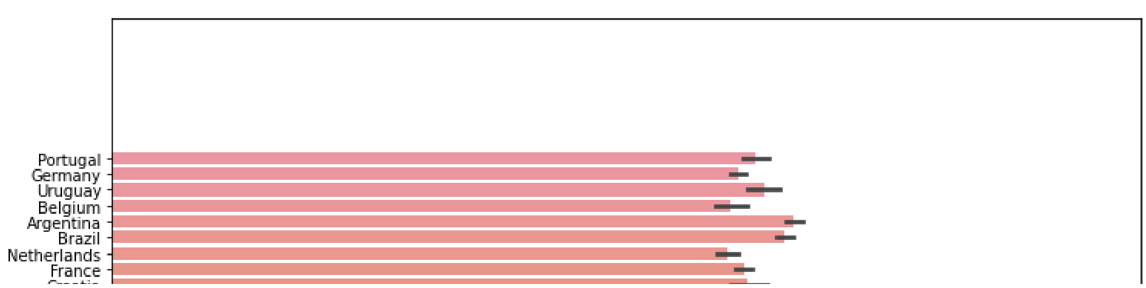
The bar chart above backs up the data

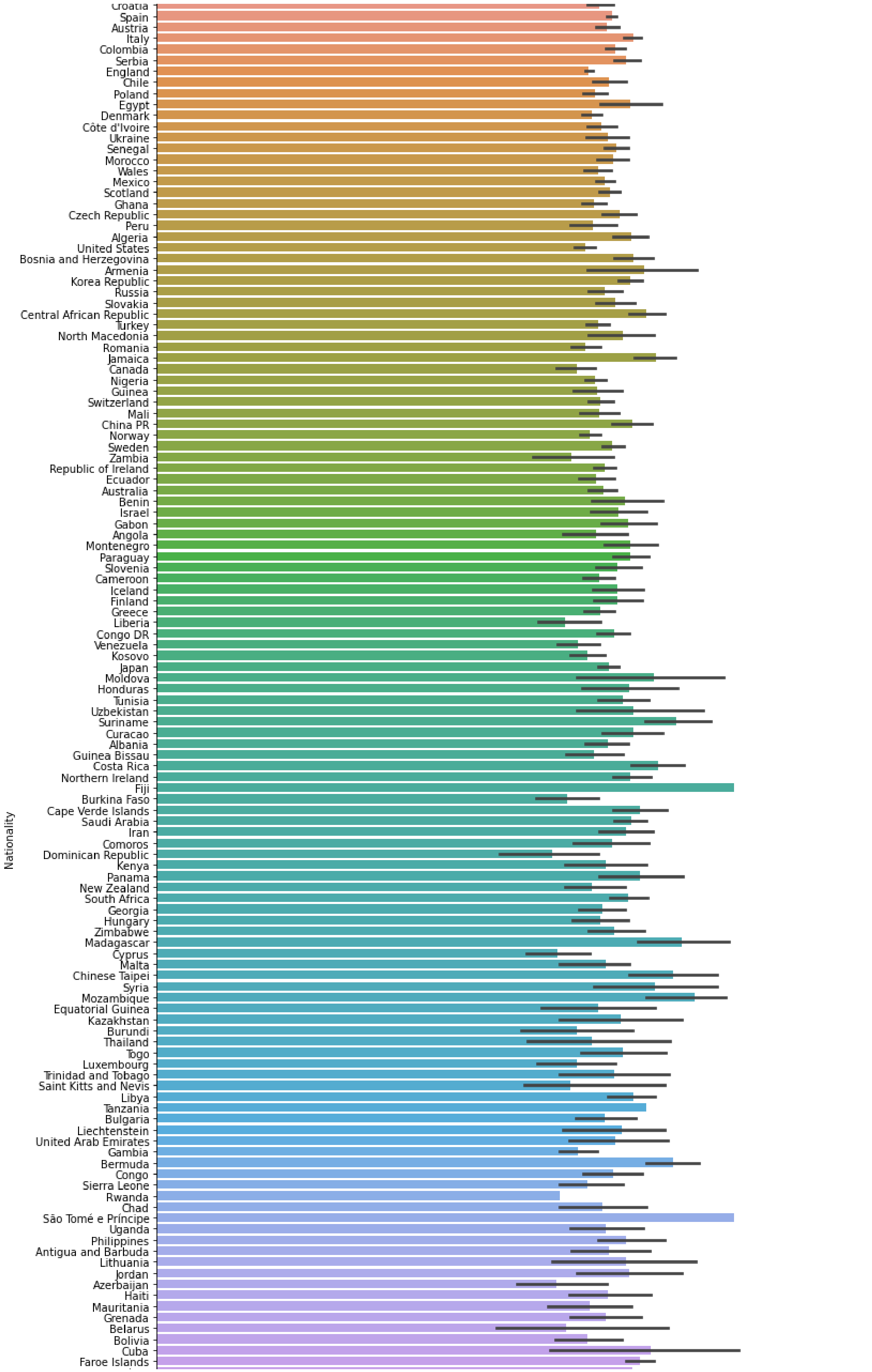
Average age of each national team in FIFA22

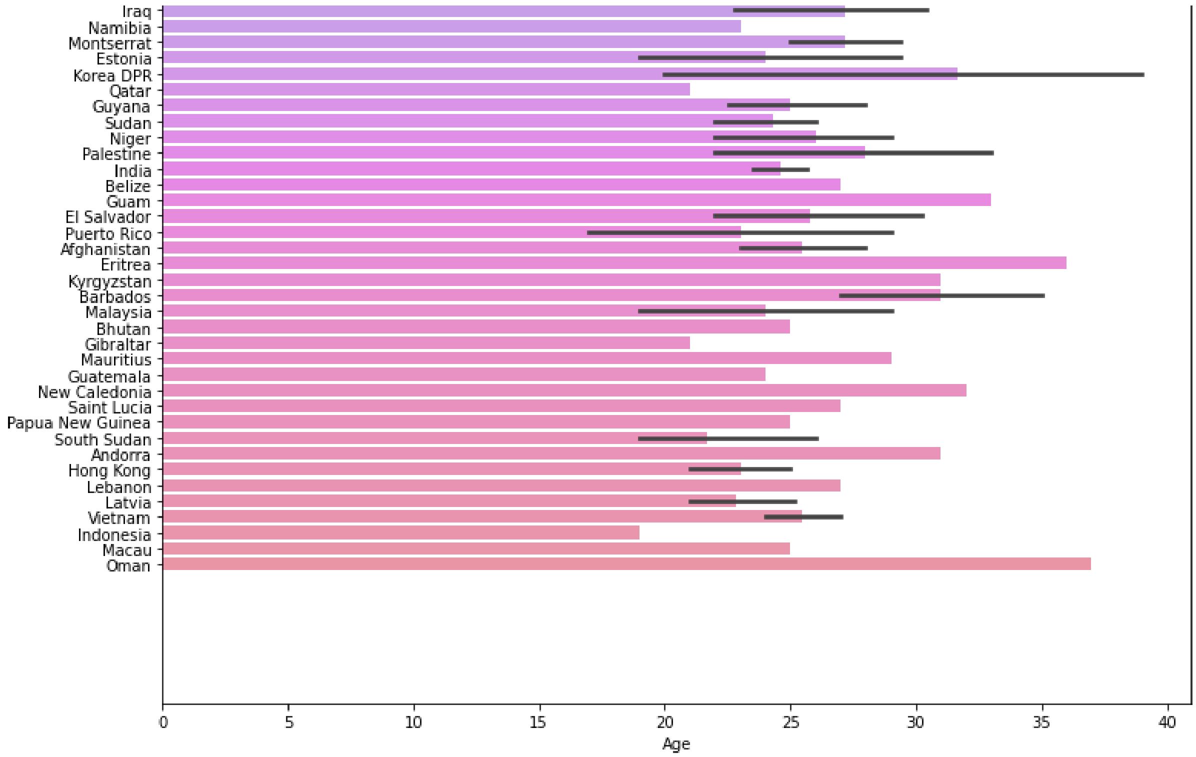
In [50]:

```
plt.figure(figsize=(12,35))
sns.barplot(y=fifa2['Nationality'],x=fifa2['Age'])
plt.plot()
```

Out[50]: []




Croatia
Spain
Austria
Italy
Colombia
Serbia
England
Chile
Poland
Egypt
Denmark
Côte d'Ivoire
Ukraine
Senegal
Morocco
Wales
Mexico
Scotland
Ghana
Czech Republic
Peru
Algeria
United States
Bosnia and Herzegovina
Armenia
Korea Republic
Russia
Slovakia
Central African Republic
Turkey
North Macedonia
Romania
Jamaica
Canada
Nigeria
Guinea
Switzerland
Mali
China PR
Norway
Sweden
Zambia
Republic of Ireland
Ecuador
Australia
Benin
Israel
Gabon
Angola
Montenegro
Paraguay
Slovenia
Cameroon
Iceland
Finland
Greece
Liberia
Congo DR
Venezuela
Kosovo
Japan
Moldova
Honduras
Tunisia
Uzbekistan
Suriname
Curacao
Albania
Guinea Bissau
Costa Rica
Northern Ireland
Fiji
Burkina Faso
Cape Verde Islands
Saudi Arabia
Iran
Comoros
Dominican Republic
Kenya
Panama
New Zealand
South Africa
Georgia
Hungary
Zimbabwe
Madagascar
Cyprus
Malta
Chinese Taipei
Syria
Mozambique
Equatorial Guinea
Kazakhstan
Burundi
Thailand
Togo
Luxembourg
Trinidad and Tobago
Saint Kitts and Nevis
Libya
Tanzania
Bulgaria
Liechtenstein
United Arab Emirates
Gambia
Bermuda
Congo
Sierra Leone
Rwanda
Chad
São Tomé e Príncipe
Uganda
Philippines
Antigua and Barbuda
Lithuania
Jordan
Azerbaijan
Haiti
Mauritania
Grenada
Belarus
Bolivia
Cuba
Faroe Islands
Nationality



Average age of players from Oman and Eritrea teams are greater than 35, while Indonesia's is less than 20!!!!!

Players with the highst rated skill moves including with their weak foot

In [51]:
```
fifa2.loc[((fifa2['Weak Foot']==fifa2['Weak Foot'].max()) &(fifa2['Skill Moves']==fifa2
```

Out[51]:

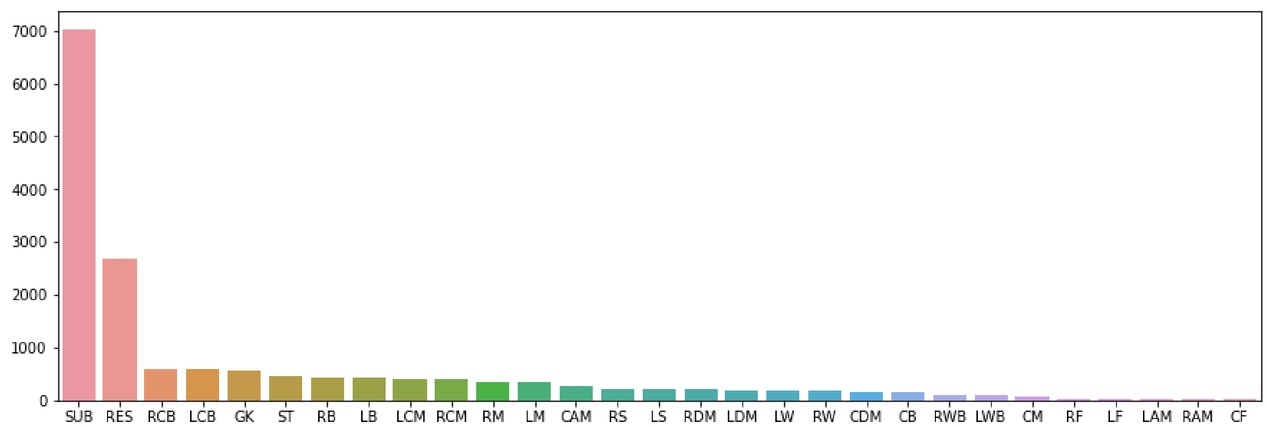
	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Inte Re
64	Neymar Jr	29	Brazil	91	91	Paris Saint-Germain	129.0	0.270	2183	Right	
91	J. Corona	28	Mexico	82	82	FC Porto	30.5	0.022	2167	Right	
1026	O. Dembélé	24	France	83	88	FC Barcelona	55.0	0.165	1990	Left	
1129	Nani	34	Portugal	81	81	Orlando City Soccer Club	12.0	0.014	1981	Right	
2402	Cesinha	31	Brazil	75	75	Daegu FC	4.7	0.009	1902	Right	
2882	F. Ribéry	38	France	79	79	US Salernitana 1919	6.0	0.041	1880	Right	

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	Inte Re
7644	R. Cherki	17	France	73	88	Olympique Lyonnais	7.0	0.009	1712	Left	

`Weak Foot` and `Skill Moves` - NEYMAR, CORONA, DEMBELE, NANI, CESINHA, RIBERY

PLAYER POSITION

In [52]:
```
fifa_pos = fifa2['Position'].value_counts()
plt.figure(figsize=(15,5))
sns.barplot(fifa_pos.index,fifa_pos.values)
plt.show()
```



No of centre forwards in the player set is the least. We can see how many of them are there in this dataset

Center Forwards are--

In [53]:
```
fifa2.loc[fifa2['Position']=='CF']
```

Out[53]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot
68	Roberto Firmino	29	Brazil	85	85	Liverpool	54.000	0.185	2180	Right
201	K. Benzema	33	France	89	89	Real Madrid CF	66.000	0.350	2116	Right
382	L. Rodríguez	36	Argentina	79	79	Gimnasia y Esgrima La Plata	6.000	0.018	2071	Right

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot
2134	Deulofeu	27	Spain	79	79	Udinese Calcio	18.500	0.022	1917	Right
2515	A. Büyük	33	Turkey	73	73	Yeni Malatyaspor	1.900	0.014	1897	Right
2547	Muriqui	35	Brazil	73	73	Shijiazhuang Ever Bright F.C.	1.300	0.007	1895	Right
8386	20 J. Ménez	32	France	70	70	Paris FC	1.400	0.006	1686	Right
9202	H. Mahou	21	France	64	71	FC Lausanne-Sport	1.200	0.003	1658	Right
5564	20 Abraham González	33	Spain	66	66	Tiburones Rojos de Veracruz	0.325	0.003	1777	Right
5773	Tan Long	33	China PR	67	67	Changchun Yatai FC	0.625	0.004	1771	Right

Top 10 Tallest palyers

In [54]:

```
data_height = fifa2.sort_values(['Height'],ascending=False)[:10]
data_height[['Name','Height', 'Club','Nationality']]
```

Out[54]:

	Name	Height	Club	Nationality
16176	T. Holý	206cm	Ipswich Town	Czech Republic
15358	20 M. Casey	203cm	Portsmouth	England
15751	21 C. Pantilimon	203cm	Denizlispor	Romania
14577	P. Ndiaye	203cm	SC Rheindorf Altach	France
12448	20 L. Traoré	203cm	CFR Cluj	Côte d'Ivoire
15678	A. Noppert	203cm	Go Ahead Eagles	Netherlands
13126	F. Mulić	203cm	Seongnam FC	Serbia
12210	21 S. Maierhofer	202cm	FC Würzburger Kickers	Austria
15204	C. Ezekwem	202cm	Sportclub Verl	Germany
15760	N. Stojišić	202cm	Vegalta Sendai	Serbia

Top 5 Shortest Players

In [55]:
```
height = fifa2.sort_values(['Height'],ascending=True)[:5]
height[['Name','Height', 'Club','Nationality']]
```

Out[55]:

	Name	Height	Club	Nationality
9791	M. García	155cm	Club Universidad Nacional	Mexico
8218	L. Paiva	156cm	Club Atlético Rentistas	Uruguay
7245	N. Barrios	156cm	San Lorenzo de Almagro	Argentina
2999	21 J. Plata	157cm	Deportivo Toluca	Ecuador
2587	Y. Soteldo	158cm	Toronto FC	Venezuela

HEAVIEST PLAYER

In [56]:
```
weight = fifa2.sort_values(['Weight'],ascending=False)[:1]
weight[['Name','Weight','Club','Nationality']]
```

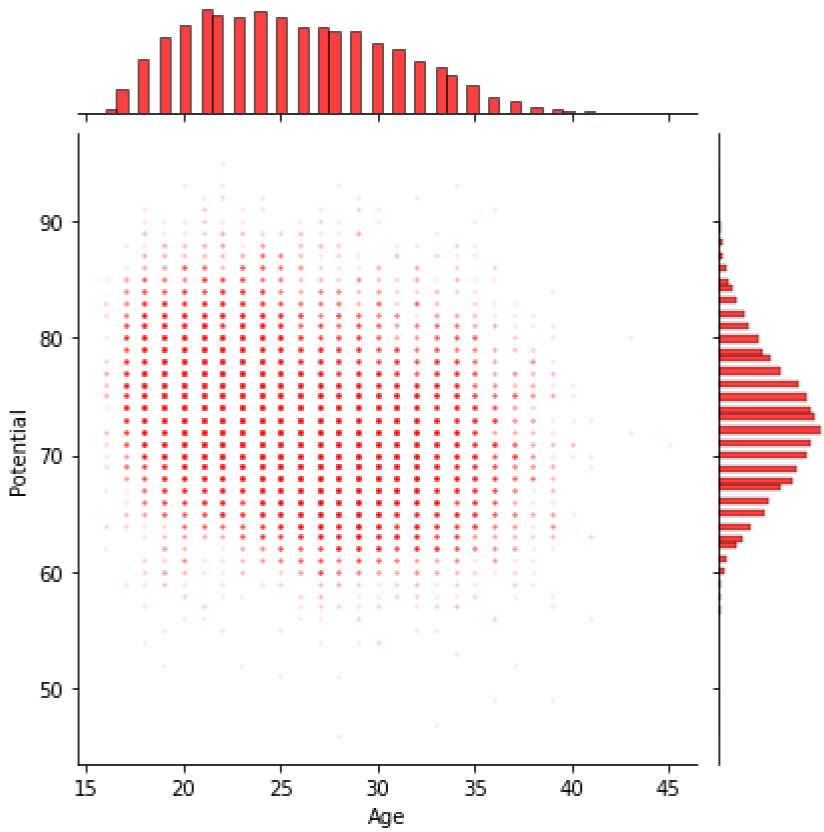
Out[56]:

	Name	Weight	Club	Nationality
11849	A. Akinfenwa	11	Wycombe Wanderers	England

Potential as related to Age

In [57]:
```
sns.jointplot(x=fifa2['Age'],y=fifa2['Potential'],
              joint_kws={'alpha':0.1,'s':5,'color':'red'},
              marginal_kws={'color':'red'})
```

Out[57]:
```
<seaborn.axisgrid.JointGrid at 0x13c0e3eb730>
```



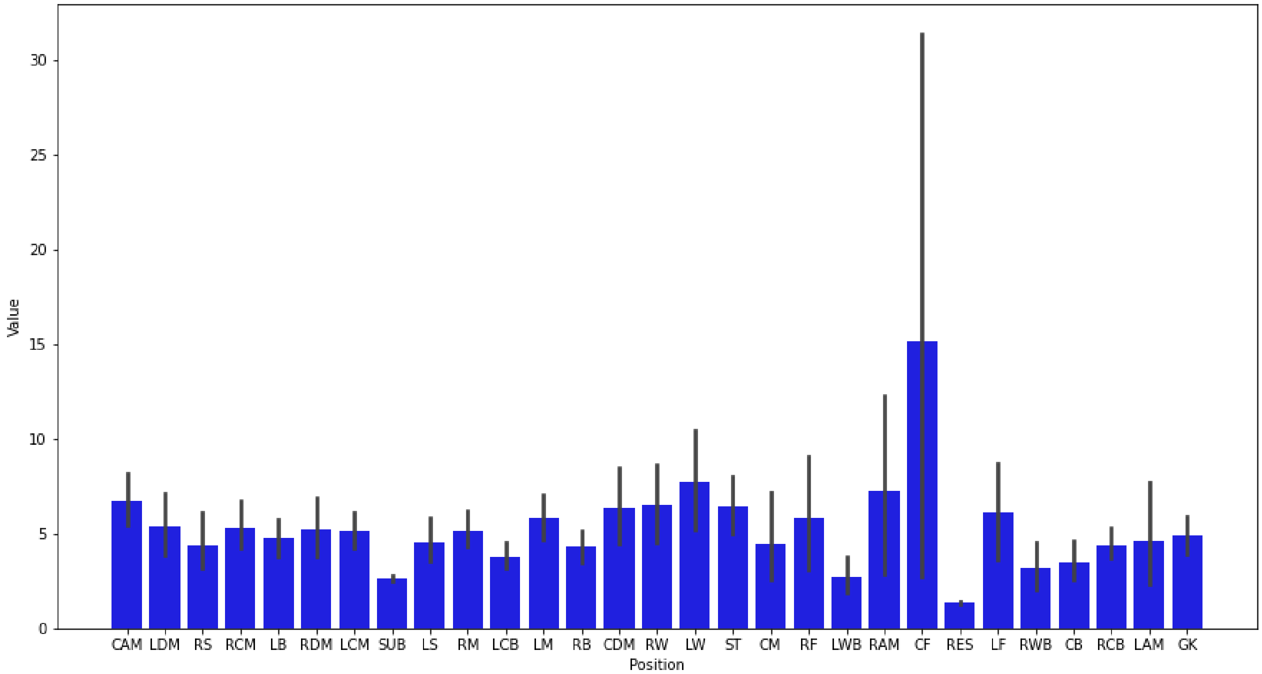
As it normally happens, players' potential increases from their young age and begins to decrease towards their retirement age.

POSITION VS PLAYER VALUE

In [58]:

```
plt.figure(figsize=(15,8))
sns.barplot(y=fifa2['Value'],x=fifa2['Position'], color = 'b')
plt.plot()
```

Out[58]: []



There are exceptions for each position. But what we can see from this is that the forwards mean values are actually at the top compared to all other positions.

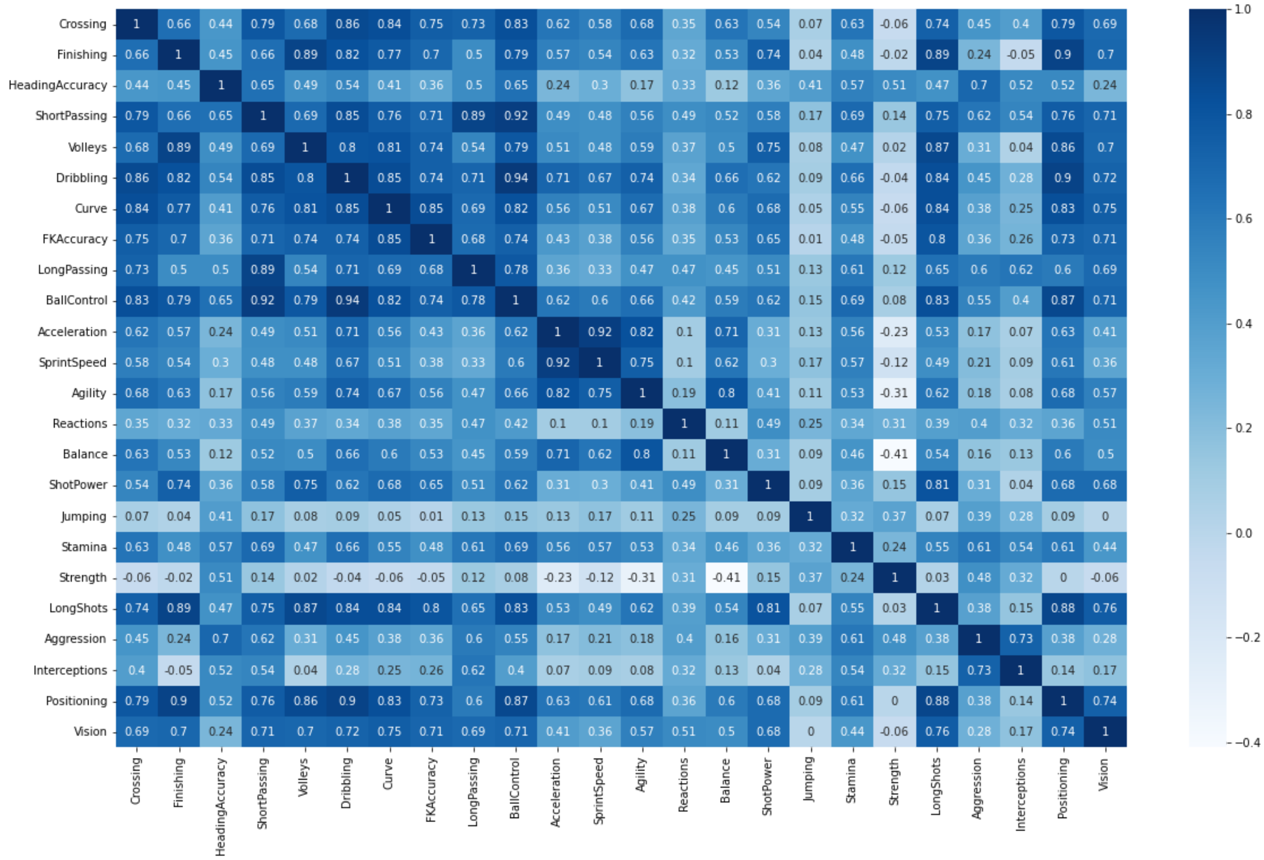
AGILITY, AGGRESSION AND THEIR CORRELATIONS

In [59]:

```
correlation = fifa2[['Crossing', 'Finishing', 'HeadingAccuracy',
       'ShortPassing', 'Volleys', 'Dribbling', 'Curve', 'FKAccuracy',
       'LongPassing', 'BallControl', 'Acceleration', 'SprintSpeed', 'Agility',
       'Reactions', 'Balance', 'ShotPower', 'Jumping', 'Stamina', 'Strength',
       'LongShots', 'Aggression', 'Interceptions', 'Positioning', 'Vision']]
```

In [60]:

```
plt.figure(figsize=(20,12))
sns.heatmap(round(correlation.corr(),2),annot=True,cmap='Blues')
plt.show()
```



Some highly correlated features in the dataset - `Positioning` and `Finishing` ; `Positioning` and `Dribbling` , `positioning` and `finishing` . Infact positioning is a highly correlated feature with many other features as well like - Longshots
`Ball Control` is highly correlated with `Short Passing` and `Dribbling` `Accelaration` is highly correlated with `Sprint Speed` .

Agility is one other feature we would be interested in. We can see that `Agility` is highly correlated with `Balance` , `dribbling` , `Sprintspeed` and `Acceleration` . So you have got an important tip to know a a great footballer.

In [61]:

```python
correlation.sort_values(['Vision'],ascending=False)[:5]
```

Out[61]:

	Crossing	Finishing	HeadingAccuracy	ShortPassing	Volleys	Dribbling	Curve	FKAccuracy	LongP
29	85.0	95.0	70.0	91.0	88.0	96.0	93.0	94.0	
3	94.0	82.0	55.0	94.0	82.0	88.0	85.0	83.0	
244	85.0	74.0	51.0	93.0	66.0	80.0	85.0	87.0	
830	69.0	77.0	49.0	92.0	76.0	85.0	86.0	79.0	
422	83.0	65.0	72.0	89.0	85.0	75.0	74.0	67.0	

Players with great `Vision` have better overalls. This is true as vision is a great technical ability which contributes to a good level of:

- Passing Accuracy
- Free kick accuracy
- positioning
- and Finishing.

In [62]:
```
correlation.sort_values(['Aggression'],ascending=False)[:5]
```

Out[62]:

	Crossing	Finishing	HeadingAccuracy	ShortPassing	Volleys	Dribbling	Curve	FKAccuracy	Lon
1040	69.0	63.0	62.0	75.0	62.0	69.0	67.0	70.0	
8508	40.0	43.0	67.0	65.0	35.0	50.0	38.0	57.0	
2899	52.0	46.0	52.0	72.0	38.0	69.0	45.0	40.0	
254	76.0	54.0	77.0	77.0	71.0	71.0	81.0	77.0	
13248	60.0	21.0	40.0	56.0	13.0	45.0	42.0	37.0	

Aggression is connected with interception and strength.