

```
In [ ]: EDA - Fifa Players Dataset
```

```
In [ ]: #Using the below dataset of fifa players dataset. Perform Exploratory data analy
#1.Which country has the most number of players
```

```
In [3]: import pandas as pd
```

```
In [5]: pd.read_csv("C:\\Users\\barbi\\Downloads\\fifa_data.csv")
```

```
Out[5]:
```

	Unnamed: 0	ID	Name	Age	Phc
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.p
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.p
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.p
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.p
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.p
...
18202	18202	238813	J. Lundstram	19	https://cdn.sofifa.org/players/4/19/238813.p
18203	18203	243165	N. Christoffersson	19	https://cdn.sofifa.org/players/4/19/243165.p
18204	18204	241638	B. Worman	16	https://cdn.sofifa.org/players/4/19/241638.p
18205	18205	246268	D. Walker-Rice	17	https://cdn.sofifa.org/players/4/19/246268.p
18206	18206	246269	G. Nugent	16	https://cdn.sofifa.org/players/4/19/246269.p

18207 rows × 89 columns

```
In [ ]: # Find the country with the most players
most_players_country = country_player_counts.idxmax()
num_players = country_player_counts.max()

print("Country with the most players:", most_players_country)
print("Number of players:", num_players)
```

```
In [8]: fifa_data = pd.read_csv("C:\\Users\\barbi\\Downloads\\fifa_data.csv")
```

```
In [9]: country_player_counts = fifa_data['Nationality'].value_counts()
```

```
In [10]: most_players_country = country_player_counts.idxmax()
num_players = country_player_counts.max()

print("Country with the most players:", most_players_country)
print("Number of players:", num_players)
```

Country with the most players: England
Number of players: 1662

In []: Insight: The country **with** the highest representation of players **in** the dataset,

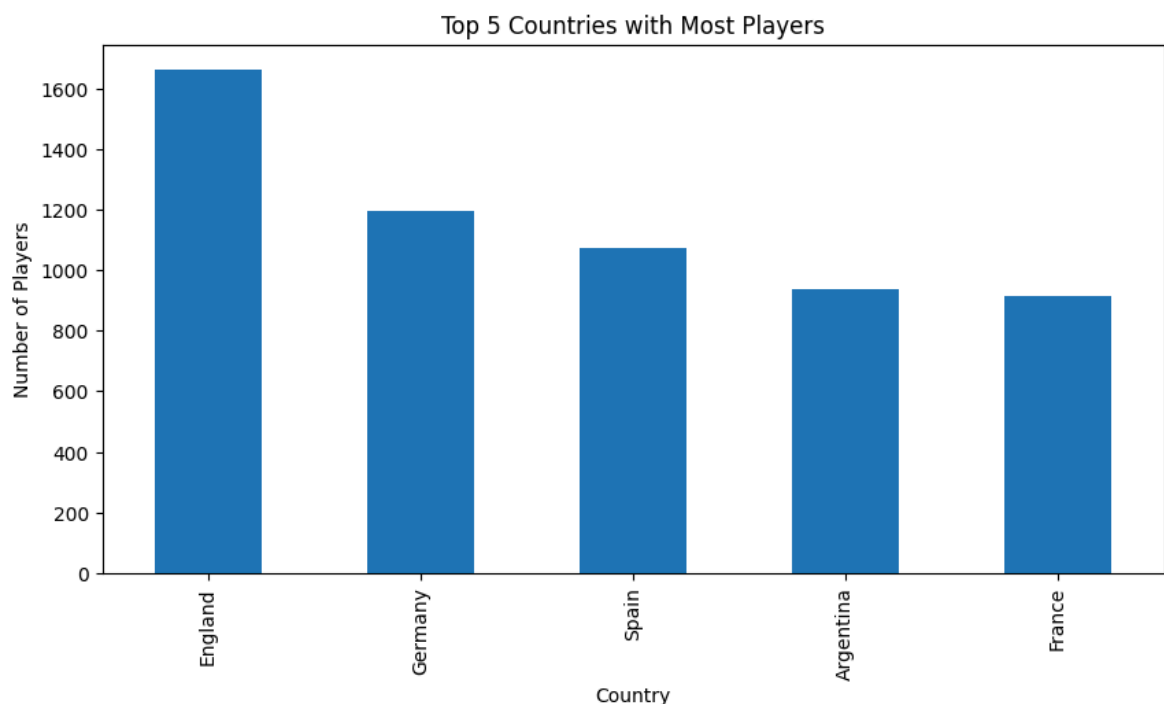
In []:

In []: *#2. Plot a bar chart of 5 top countries with the most number of players*

```
In [ ]: import matplotlib.pyplot as plt
```

```
In [11]: top_countries = country_player_counts.head(5)
```

```
In [36]: fig = plt.figure(figsize = (10, 5))
top_countries.plot(kind='bar')
xlabel=('Country')
ylabel=('Number of Players')
title=('Top 5 Countries with Most Players')
plt.show()
```



In []: Insight: The bar chart visualizes the distribution of players among the top 5 co

In []:

In []: *#3. Which player has the highest salary?*

```
In [38]: highest_salary_player = fifa_data.loc[fifa_data['Wage'].idxmax()]
print("Player with the highest Wage:")
print(highest_salary_player)
```

```

Player with the highest Wage:
Unnamed: 0      1071
ID             139668
Name           F. Marchetti
Age            35
Photo          https://cdn.sofifa.org/players/4/19/139668.png
...
GKHandling      76.0
GKKicking       65.0
GKPositioning   77.0
GKReflexes      78.0
Release Clause  €3.6M
Name: 1071, Length: 89, dtype: object

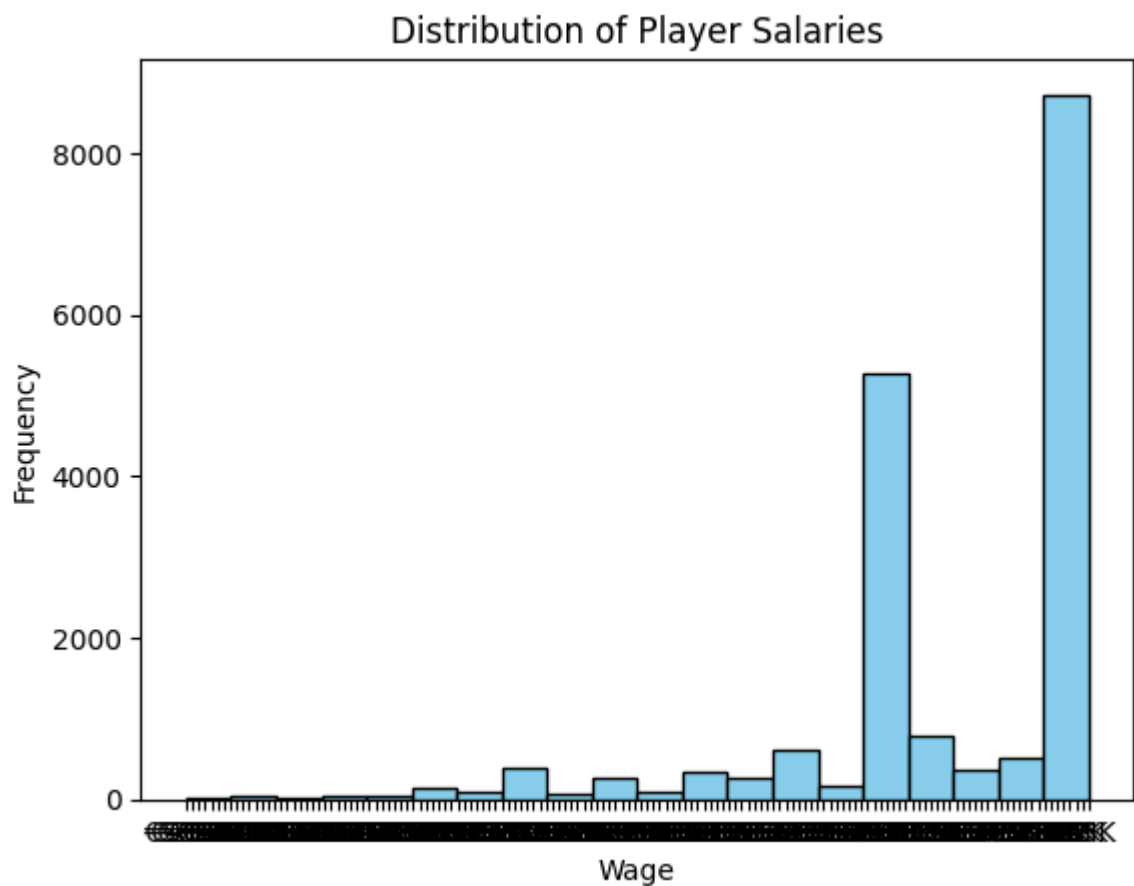
```

```
In [ ]: nsight: Identifying the player with the highest wage shows the top earner in the
```

```
In [ ]:
```

```
In [ ]: #4..Plot a histogram to get the salary range of the players
```

```
In [40]: plt.hist(fifa_data['Wage'], bins=20, color='skyblue', edgecolor='black')
plt.xlabel('Wage')
plt.ylabel('Frequency')
plt.title('Distribution of Player Salaries')
plt.show()
```



```
In [ ]: Insight: The histogram displays the distribution of player salaries, indicating
```

```
In [ ]:
```

```
In [41]: #5. Who is the tallest player in the fifa?
```

```
In [44]: print(fifa_data.head()) # View the first few rows  
print(fifa_data.info()) # Get information about the dataset
```

	Unnamed: 0	ID	Name	Age	\
0	0	158023	L. Messi	31	
1	1	20801	Cristiano Ronaldo	33	
2	2	190871	Neymar Jr	26	
3	3	193080	De Gea	27	
4	4	192985	K. De Bruyne	27	

	Photo	Nationality	\
0	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	
1	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	
2	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	
3	https://cdn.sofifa.org/players/4/19/193080.png	Spain	
4	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	

	Flag	Overall	Potential	\
0	https://cdn.sofifa.org/flags/52.png	94	94	
1	https://cdn.sofifa.org/flags/38.png	94	94	
2	https://cdn.sofifa.org/flags/54.png	92	93	
3	https://cdn.sofifa.org/flags/45.png	91	93	
4	https://cdn.sofifa.org/flags/7.png	91	92	

	Club	...	Composure	Marking	StandingTackle	SlidingTackle	\
0	FC Barcelona	...	96.0	33.0	28.0	26.0	
1	Juventus	...	95.0	28.0	31.0	23.0	
2	Paris Saint-Germain	...	94.0	27.0	24.0	33.0	
3	Manchester United	...	68.0	15.0	21.0	13.0	
4	Manchester City	...	88.0	68.0	58.0	51.0	

	GK Diving	GK Handling	GK Kicking	GK Positioning	GK Reflexes	Release Clause
0	6.0	11.0	15.0	14.0	8.0	€226.5M
1	7.0	11.0	15.0	14.0	11.0	€127.1M
2	9.0	9.0	15.0	15.0	11.0	€228.1M
3	90.0	85.0	87.0	88.0	94.0	€138.6M
4	15.0	13.0	5.0	10.0	13.0	€196.4M

[5 rows x 89 columns]

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 18207 entries, 0 to 18206

Data columns (total 89 columns):

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	18207 non-null	int64
1	ID	18207 non-null	int64
2	Name	18207 non-null	object
3	Age	18207 non-null	int64
4	Photo	18207 non-null	object
5	Nationality	18207 non-null	object
6	Flag	18207 non-null	object
7	Overall	18207 non-null	int64
8	Potential	18207 non-null	int64
9	Club	17966 non-null	object
10	Club Logo	18207 non-null	object
11	Value	18207 non-null	object
12	Wage	18207 non-null	object
13	Special	18207 non-null	int64
14	Preferred Foot	18159 non-null	object
15	International Reputation	18159 non-null	float64
16	Weak Foot	18159 non-null	float64
17	Skill Moves	18159 non-null	float64
18	Work Rate	18159 non-null	object

19	Body Type	18159	non-null	object
20	Real Face	18159	non-null	object
21	Position	18147	non-null	object
22	Jersey Number	18147	non-null	float64
23	Joined	16654	non-null	object
24	Loaned From	1264	non-null	object
25	Contract Valid Until	17918	non-null	object
26	Height	18159	non-null	object
27	Weight	18159	non-null	object
28	LS	16122	non-null	object
29	ST	16122	non-null	object
30	RS	16122	non-null	object
31	LW	16122	non-null	object
32	LF	16122	non-null	object
33	CF	16122	non-null	object
34	RF	16122	non-null	object
35	RW	16122	non-null	object
36	LAM	16122	non-null	object
37	CAM	16122	non-null	object
38	RAM	16122	non-null	object
39	LM	16122	non-null	object
40	LCM	16122	non-null	object
41	CM	16122	non-null	object
42	RCM	16122	non-null	object
43	RM	16122	non-null	object
44	LWB	16122	non-null	object
45	LDM	16122	non-null	object
46	CDM	16122	non-null	object
47	RDM	16122	non-null	object
48	RWB	16122	non-null	object
49	LB	16122	non-null	object
50	LCB	16122	non-null	object
51	CB	16122	non-null	object
52	RCB	16122	non-null	object
53	RB	16122	non-null	object
54	Crossing	18159	non-null	float64
55	Finishing	18159	non-null	float64
56	HeadingAccuracy	18159	non-null	float64
57	ShortPassing	18159	non-null	float64
58	Volleys	18159	non-null	float64
59	Dribbling	18159	non-null	float64
60	Curve	18159	non-null	float64
61	FKAccuracy	18159	non-null	float64
62	LongPassing	18159	non-null	float64
63	BallControl	18159	non-null	float64
64	Acceleration	18159	non-null	float64
65	SprintSpeed	18159	non-null	float64
66	Agility	18159	non-null	float64
67	Reactions	18159	non-null	float64
68	Balance	18159	non-null	float64
69	ShotPower	18159	non-null	float64
70	Jumping	18159	non-null	float64
71	Stamina	18159	non-null	float64
72	Strength	18159	non-null	float64
73	LongShots	18159	non-null	float64
74	Aggression	18159	non-null	float64
75	Interceptions	18159	non-null	float64
76	Positioning	18159	non-null	float64
77	Vision	18159	non-null	float64
78	Penalties	18159	non-null	float64

```

79 Composure          18159 non-null float64
80 Marking            18159 non-null float64
81 StandingTackle     18159 non-null float64
82 SlidingTackle      18159 non-null float64
83 GKDiving           18159 non-null float64
84 GKHandling          18159 non-null float64
85 GKKicking           18159 non-null float64
86 GKPositioning      18159 non-null float64
87 GKReflexes         18159 non-null float64
88 Release Clause     16643 non-null object
dtypes: float64(38), int64(6), object(45)
memory usage: 12.4+ MB
None

```

```
In [45]: print(fifa_data['Height'].describe())
```

```

count      18159
unique         21
top          6'0
freq        2881
Name: Height, dtype: object

```

```
In [46]: sorted_players_by_height = fifa_data.sort_values(by=["Height"], ascending=False)
tallest_player = sorted_players_by_height[['Name', 'Height']].head(1)
print("Tallest player:")
print(tallest_player)
```

```

Tallest player:
      Name Height
11614  T. Holý   6'9

```

```
In [ ]: Insight: Identifying the tallest player offers showcasing the diversity in physi
```

```
In [ ]:
```

```
In [ ]: #6..Which club has the most number of players?
```

```
In [50]: club_player_count = fifa_data.groupby('Club').size()
```

```
In [48]: most_players_club = club_player_count.idxmax()
num_players = club_player_count.max()
```

```
In [51]: print("Club with the most number of players:", most_players_club)
print("Number of players:", num_players)
```

```

Club with the most number of players: AS Monaco
Number of players: 33

```

```
In [ ]: Insight: Determining the club with the most players highlights the clubs popular
```

```
In [ ]:
```

```
In [ ]: #7.Which foot is most preferred by the players?Draw a bar chart for preferred fo
```

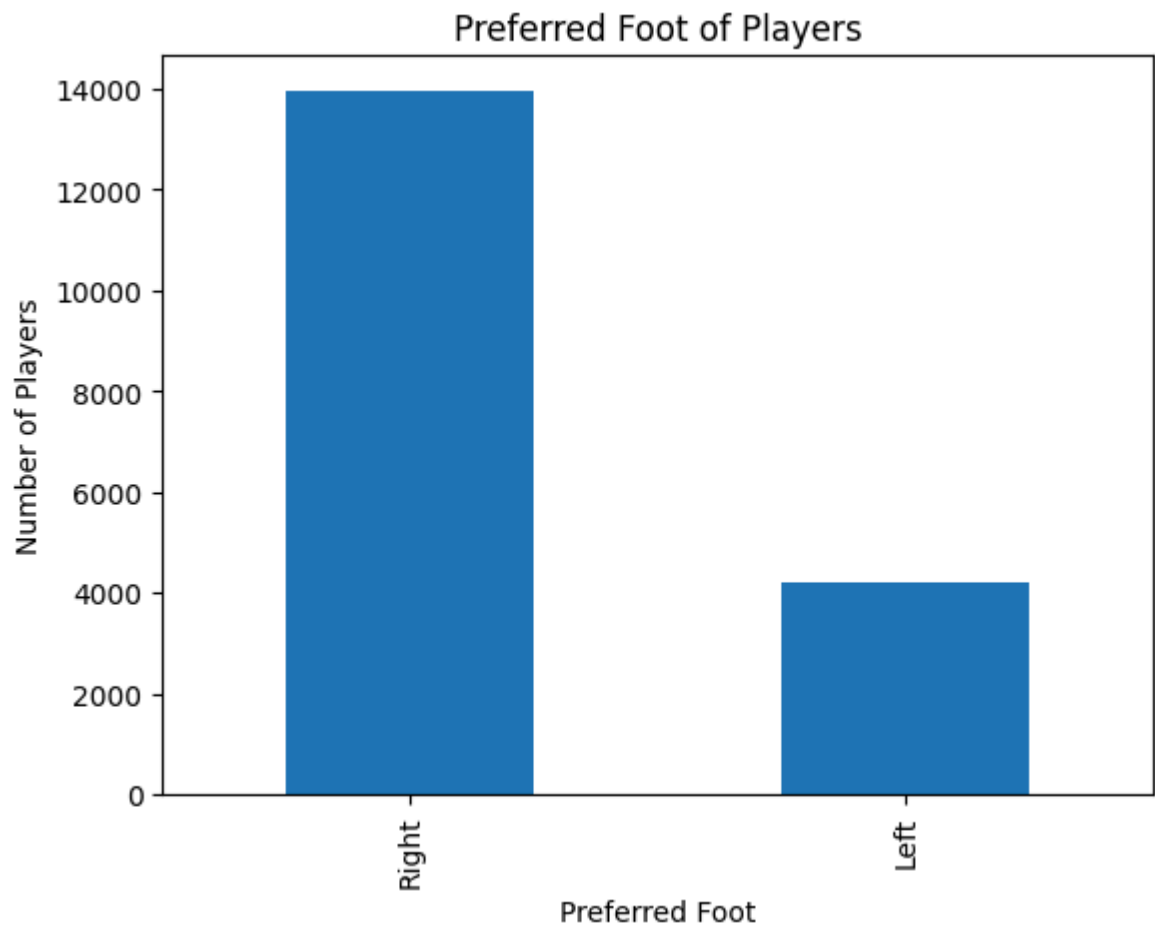
```
In [52]: preferred_foot_counts = fifa_data['Preferred Foot'].value_counts()
```

```
In [53]: most_preferred_foot = preferred_foot_counts.idxmax()
num_players = preferred_foot_counts.max()
```

```
In [54]: print("Most preferred foot:", most_preferred_foot)
print("Number of players:", num_players)
```

Most preferred foot: Right
Number of players: 13948

```
In [55]: preferred_foot_counts.plot(kind='bar')
xlabel=('Preferred Foot')
ylabel=('Number of Players')
title=('Preferred Foot of Players')
plt.show()
```



In []: Insight: The bar chart illustrates the distribution of preferred foot among play