FIFA Data Analysis

In [3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# reading fifa data csv file into dataframe
df = pd.read_csv("fifa_data.csv")
df.head()
```

Out[3]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	h
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	h
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	h
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	h
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	

5 rows × 89 columns

Details of columns in dataset

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

```
<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
```

Statistical analysis

In [57]:
```
df.describe()
```

Out[57]:

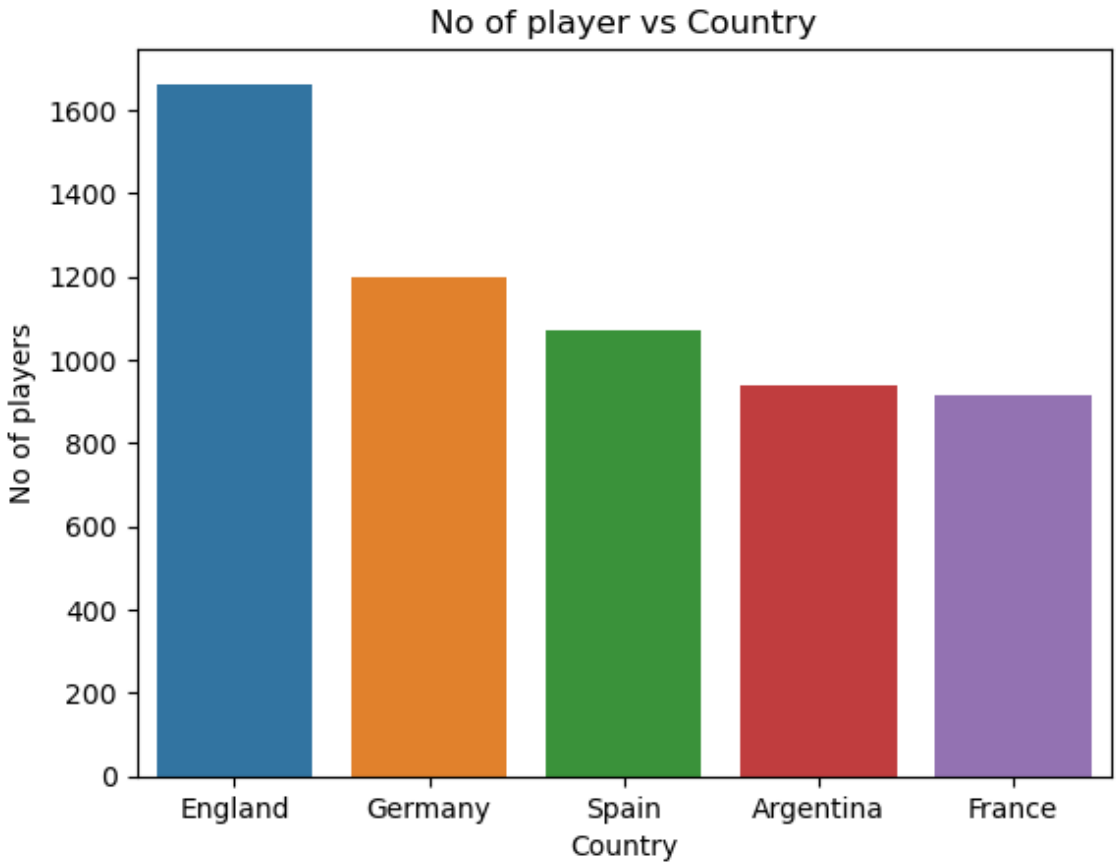
	Unnamed: 0	ID	Age	Overall	Potential	Special
count	18207.000000	18207.000000	18207.000000	18207.000000	18207.000000	18207.000000
mean	9103.000000	214298.338606	25.122206	66.238699	71.307299	1597.809908
std	5256.052511	29965.244204	4.669943	6.908930	6.136496	272.586016
min	0.000000	16.000000	16.000000	46.000000	48.000000	731.000000
25%	4551.500000	200315.500000	21.000000	62.000000	67.000000	1457.000000
50%	9103.000000	221759.000000	25.000000	66.000000	71.000000	1635.000000
75%	13654.500000	236529.500000	28.000000	71.000000	75.000000	1787.000000
max	18206.000000	246620.000000	45.000000	94.000000	95.000000	2346.000000

8 rows × 44 columns

Analysis

Which country has the most number of players?

In [71]:
```
country_count = df['Nationality'].value_counts()
country_count = country_count[:5]
sns.barplot(x = country_count.index, y = country_count.values)
plt.xlabel("Country")
plt.ylabel("No of players")
plt.title("No of player vs Country")
plt.show()
```



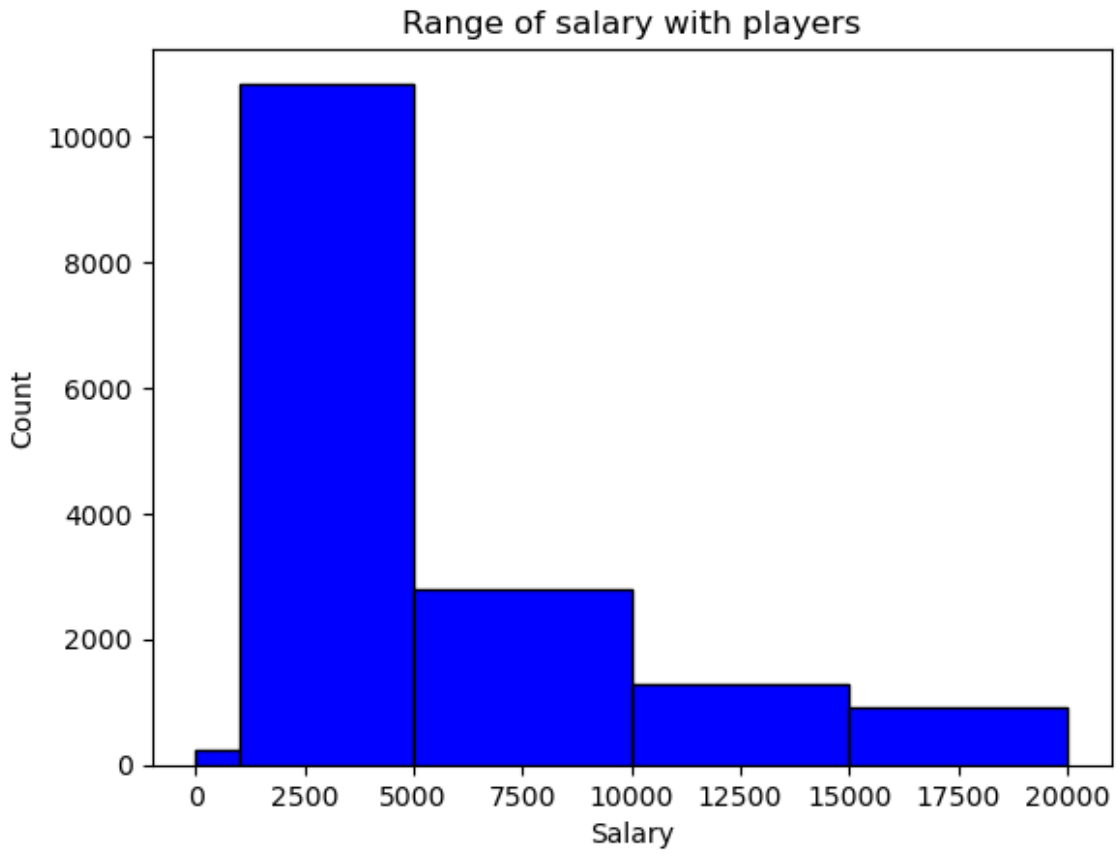
From the plotting of top 5 countries it can be understood that England has the most number of players.

The salary range and who earns highest salary

In [64]:
```
#creating a smaller dataframe for easy conversion of salary type
df_short = pd.read_csv("fifa_data.csv", usecols = ['ID','Name','Wage'])

#converting data type of salary by eliminating characters
df_short['Wage'] = df_short['Wage'].replace({'€': '','K': '000'}, regex=Tru

#plotting histogram
plt.hist(df_short['Wage'], bins = [0,1000,5000,10000,15000,20000], color =
plt.xlabel('Salary')
plt.ylabel('Count')
plt.title("Range of salary with players")
plt.show()
```



In [115]:
```
#sorting according to descending order of salary
print("The player with highest salary: ")
df_short.sort_values(by=['Wage'], ascending = False).iloc[0]
```

```
The player with highest salary:

ID        158023
Name    L. Messi
Wage    565000.0
Name: 0, dtype: object
```

Out[115]:

Who is the tallest player in the list?

In [111]:
```
#creating a smaller data for easy calculation
df_ht = pd.read_csv("fifa_data.csv", usecols = ['Name','Height'])

#converting type of height to float for comparison
df_ht['Height'] = df_ht['Height'].replace({"'": "."}, regex=True).astype(fl

# sorting based on height
print("The tallest player with height in feet:")
df_ht.sort_values(by=['Height'], ascending = False).iloc[0]
```

```
The tallest player with height in feet:

Name      T. Holý
Height        6.9
Name: 11614, dtype: object
```

Out[111]:

Which club has most number of players?

In [22]:
```
club = df['Club'].value_counts()
clb_list = club.tolist()
max_count = club[0]
counter = 0
for i in clb_list:
    if i == max_count:
        counter+=1

print("The clubs with most number of players are: ")
club.head(counter)
```

```
The clubs with most number of players are:
```

Out[22]:
```
FC Barcelona               33
Valencia CF                33
Fortuna Düsseldorf         33
Cardiff City               33
Rayo Vallecano             33
CD Leganés                 33
Frosinone                  33
Newcastle United           33
Southampton                33
Burnley                    33
Eintracht Frankfurt        33
Wolverhampton Wanderers    33
TSG 1899 Hoffenheim        33
Everton                    33
AS Monaco                  33
RC Celta                   33
Empoli                     33
Manchester City            33
Manchester United          33
Borussia Dortmund          33
Real Madrid                33
Atlético Madrid            33
Tottenham Hotspur          33
Chelsea                    33
Liverpool                  33
Arsenal                    33
Name: Club, dtype: int64
```

Which foot is preferred by players?

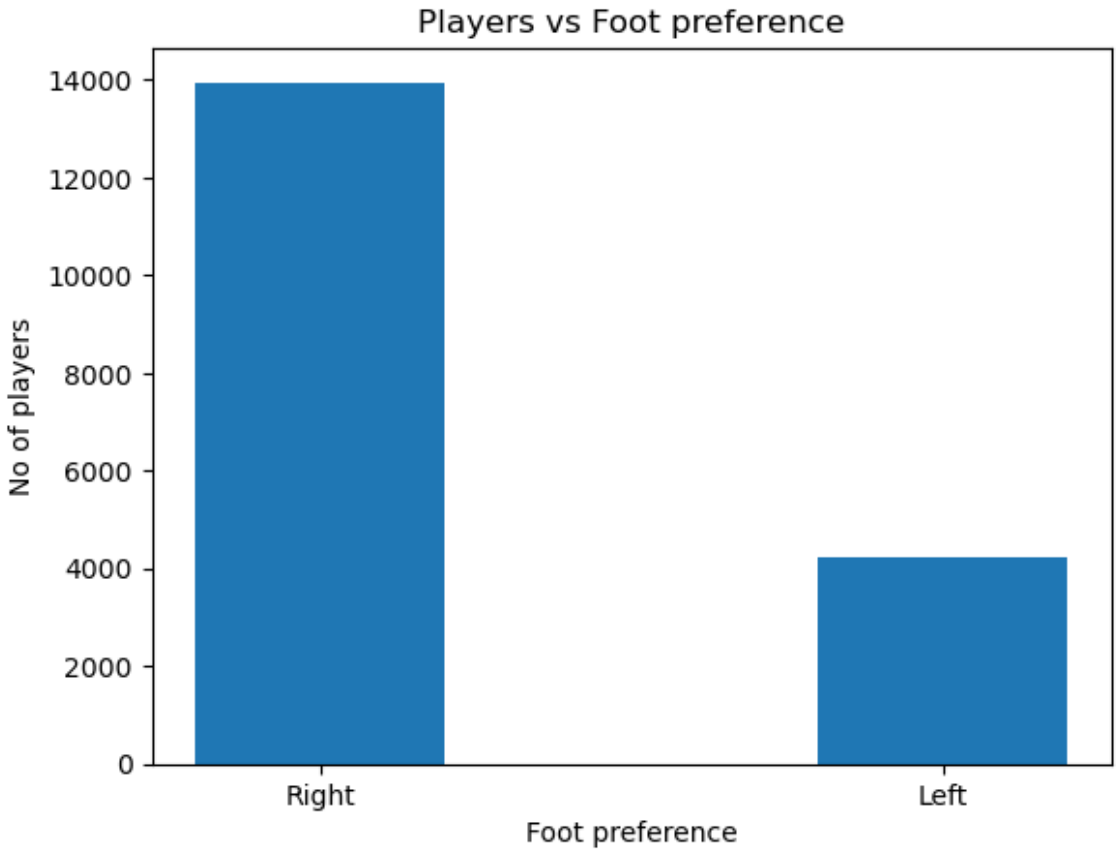
In [34]:
```
pf = df['Preferred Foot'].value_counts()
pf
```

Out[34]:
```
Right    13948
Left      4211
Name: Preferred Foot, dtype: int64
```

In [37]:
```
x=pf.index
y=pf.values

plt.bar(x,y, width = 0.4)
plt.xlabel("Foot preference")
plt.ylabel("No of players")
plt.title("Players vs Foot preference")

plt.show()
```



We can infer that most players prefer right foot

In [ ]:
```

```