EDA - Fifa Players Dataset (Formative Assessment 3)

```
In [1]: import matplotlib.pyplot as plt
        import pandas as pd
        import numpy as np

        #read csv file
        fifa = pd.read_csv("fifa_data.csv")

        #Removing the unnamed index column and saving a copy to df1
        df1=fifa.drop("Unnamed: 0",axis=1)
```

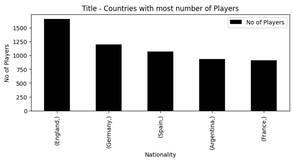
1.Which country has the most number of players (score :1):

England - 1662 Players

```
In [2]: df1.value_counts(["Nationality"]).head()

Out[2]: Nationality
        England      1662
        Germany      1198
        Spain        1072
        Argentina     937
        France        914
        Name: count, dtype: int64
```

2.Plot a bar chart of 5 top countries with the most number of players. (score :1)

```
In [3]: df1.fillna("Unknown")
        top = df1.value_counts(["Nationality"])
        plt.figure(figsize=(6.3),dpi=100)
        top_5 = top.head().plot(kind="bar",color="Black",label="No of Players")
        plt.title('Title - Countries with most number of Players')
        plt.xlabel('Nationality')
        plt.ylabel('No of Players')
        plt.legend() # shows label
        plt.show()
```

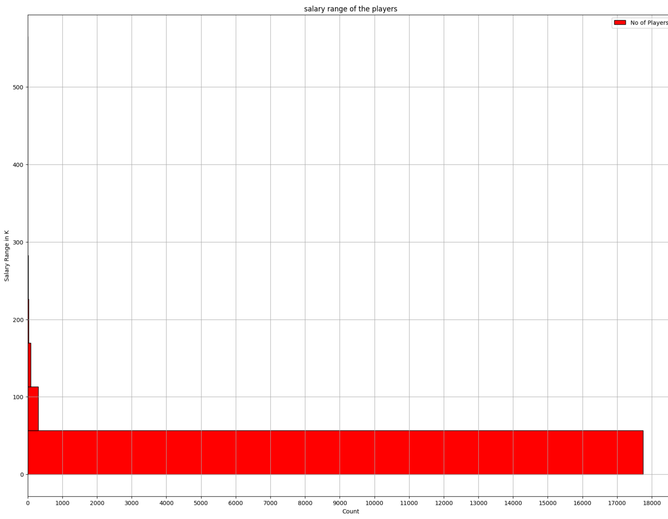


3.Which player has the highest salary? (score :1)

```
In [4]: df1['Wage']=df1['Wage'].str.replace('€','')
        df1['Wage']=df1['Wage'].str.replace('K','')
        df1['Wage']=df1['Wage'].astype(int)
        df1['Wage'].dtypes

Out[4]: dtype('int32')

In [5]: highest_salary=df1.sort_values(by="Wage",ascending=False).head(1)
        print(f"The Player with highest Salary is {highest_salary['Name']}")

        The Player with highest Salary is 0    L. Messi
        Name: Name, dtype: object
```

4. Plot a histogram to get the salary range of the players. (score :1)

```
In [6]: plt.figure(figsize=(20,15))
        plt.hist(df1['Wage'],color='red', edgecolor='black',orientation="horizontal",label="No of Players")
        plt.title('salary range of the players')  # Add a title
        plt.ylabel('Salary Range in K')  # Add x-axis label
        plt.xlabel('Count')  # Add y-axis label
        plt.xticks(range(0,19000,1000))
        plt.grid(True)
        plt.yticks(range(0,600,100))
        plt.legend()
        plt.show

Out[6]: <function matplotlib.pyplot.show(close=None, block=None)>
```



5.Who is the tallest player in the fifa? (score :1): T. Holý

```
In [7]: df1['Height']=df1['Height'].str.replace('\'','.')
        df1['Height']=df1['Height'].astype(float)
        df1['Height'].dtypes

Out[7]: dtype('float64')

In [8]: tallest_player=df1.sort_values(by="Height",ascending=False).head(1)
        print(f"The Tallest Player is {tallest_player['Name']}")

        The Tallest Player is 11514    T. Holý
        Name: Name, dtype: object
```

6.Which club has the most number of players? (score :1)

26 clubs have 33 players in each of them.The list is given below

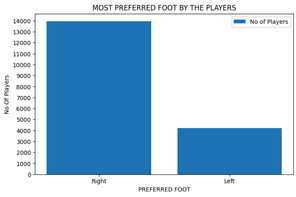
```
In [9]: df1.value_counts(["Club"]).head(26)

Out[9]: Club
        Borussia Dortmund           33
        Tottenham Hotspur           33
        Chelsea                     33
        Valencia CF                 33
        Everton                     33
        Newcastle United            33
        Real Madrid                 33
        Frosinone                   33
        Arsenal                     33
        Cardiff City                33
        Fortuna Düsseldorf          33
        Rayo Vallecano              33
        Atlético Madrid             33
        AS Monaco                   33
        Eintracht Frankfurt         33
        FC Barcelona                33
        Wolverhampton Wanderers     33
        CD Leganés                  33
        Burnley                     33
        Southampton                 33
        Manchester United           33
        Manchester City             33
        Empoli                      33
        RC Celta                    33
        TSG 1899 Hoffenheim         33
        Liverpool                   33
        Name: count, dtype: int64
```

7.Which foot is most preferred by the players?Draw a bar chart for preferred foot (score :1)

```
In [10]: df1.value_counts(["Preferred Foot"])

Out[10]: Preferred Foot
         Right            13948
         Left              4211
         Name: count, dtype: int64
```

The most preferred foot by the players is the "Right foot".

```
In [11]: x=df1.value_counts(["Preferred Foot"]).index
         y=df1.value_counts(["Preferred Foot"]).values
         x1=list(x[0])
         x2=list(x[1])
         x=x1+x2
         plt.figure(figsize=(8,5),dpi=100)
         plt.bar(x,y,label="No of Players")
         plt.title('MOST PREFERRED FOOT BY THE PLAYERS')
         plt.yticks(range(0,15000,1000))# y axis numbers
         plt.xlabel('PREFERRED FOOT')
         plt.ylabel('No Of Players')
         plt.legend() # shows label
         plt.show()
```



In addition,

Describe the insights you gained from each question. (score :2)

1. According to data given, The six players given below has highest international reputation.

The list is as given below

```
In [12]: ir=df1[df1['International Reputation']==5.0]
         ir[['Name','Age','Nationality','Potential']]
```

Out[12]:

	Name	Age	Nationality	Potential
0	L. Messi	31	Argentina	94
1	Cristiano Ronaldo	33	Portugal	94
2	Neymar Jr	26	Brazil	93
7	L. Suárez	31	Uruguay	91
22	M. Neuer	32	Germany	89
109	Z. Ibrahimović	36	Sweden	85

2. A scatter plot with Age as y axis, wage as y axis and potential as hues

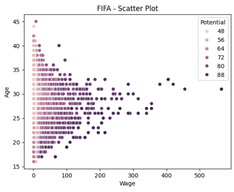
1. From the scatter plot we can summarize that the palyers with most potential and highest wages are in the age groupe of 25 to 35

2. In the age group of 15 to 25,the potential of the player increases, which results in an incline in wages.

3. But after the age of 35,As Age increases the potential of the player decreses, which results in a severe decline in wages.

4. There are also a small amount players that devaites from the above given analysis, that have huge potential in 15 to 25 and 35 to 45 age groups

```
In [13]: import seaborn as sns
         sns.scatterplot(x="Wage",y="Age",hue="Potential",data=df1)
         plt.ylabel("Age")
         plt.xlabel("Wage")
         plt.title("FIFA - Scatter Plot")
         #show
         plt.show()
```


FIFA - Scatter Plot
Potential
48
56
64
72
80
88
Age
Wage
45
40
35
30
25
20
15
0
100
200
300
400
500