Soccer Player Data Analysis

https://www.kaggle.com/datasets/maso0dahmed/football-players-data?resource=download

Step 1: Understand the dataset

The main goal is to analyse the dataset on soccer players and see who has most wins, most potential, the tallest and best international reputation.

Step 2: Importing dataset

The dataset is imported to panda datframe which is in CSV format and use matplotlib to visualise data.

Step 3: Data selection and reducing duplicacy

Will need to find rows with null values and delete them to remove dublicacy.

Step 4: Answering questions and data visualization

Now I will answer the 8 questions I made for the dataset and use matplot to visualize it.

Questions:

1. How are players distributed across different nationalities?
2. Top 10 nationalities with most players.
3. Which national teams have the most players in the dataset?
4. What is the age distribution among the players in the dataset?
5. Player International ratings top 10?
6. Best rating players from each country.
7. How does a player's potential rating compare to their current overall rating?
8. Does a player's preferred foot (left or right) have any correlation with their overall performance or specific skills?

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

data = pd.read_csv('fifa_players.csv')

data.drop_duplicates(inplace=True)

print(data.columns)

Index(['name', 'full_name', 'birth_date', 'age', 'height_cm',
'weight_kgs',
       'positions', 'nationality', 'overall_rating', 'potential',
'value_euro',
       'wage_euro', 'preferred_foot', 'international_reputation(1-5)',
       'weak_foot(1-5)', 'skill_moves(1-5)', 'body_type',
```

```
       'release_clause_euro', 'national_team', 'national_rating',
       'national_team_position', 'national_jersey_number', 'crossing',
       'finishing', 'heading_accuracy', 'short_passing', 'volleys',
       'dribbling', 'curve', 'freekick_accuracy', 'long_passing',
       'ball_control', 'acceleration', 'sprint_speed', 'agility',
'reactions',
       'balance', 'shot_power', 'jumping', 'stamina', 'strength',
'long_shots',
       'aggression', 'interceptions', 'positioning', 'vision',
'penalties',
       'composure', 'marking', 'standing_tackle', 'sliding_tackle'],
      dtype='object')
```

```
print(data.head(5))
```

```
           name                        full_name birth_date  age
height_cm  \
0      L. Messi  Lionel Andrés Messi Cuccittini  6/24/1987   31
170.18
1    C. Eriksen    Christian  Dannemann Eriksen  2/14/1992   27
154.94
2      P. Pogba                      Paul Pogba  3/15/1993   25
190.50
3    L. Insigne                 Lorenzo Insigne   6/4/1991   27
162.56
4  K. Koulibaly               Kalidou Koulibaly  6/20/1991   27
187.96

   weight_kgs  positions nationality  overall_rating
potential  ...  \
0        72.1   CF,RW,ST   Argentina              94         94  ...

1        76.2  CAM,RM,CM     Denmark              88         89  ...

2        83.9     CM,CAM      France              88         91  ...

3        59.0      LW,ST       Italy              88         88  ...

4        88.9         CB     Senegal              88         91  ...


   long_shots  aggression interceptions  positioning  vision
penalties  \
0          94          48            22           94      94
75
1          89          46            56           84      91
67
2          82          78            64           82      88
82
3          84          34            26           83      87
```

```
61
4          15          87             88           24      49
33

  composure  marking standing_tackle  sliding_tackle
0        96       33              28              26
1        88       59              57              22
2        87       63              67              67
3        83       51              24              22
4        80       91              88              87

[5 rows x 51 columns]
```

```
print(data.tail(5))
```

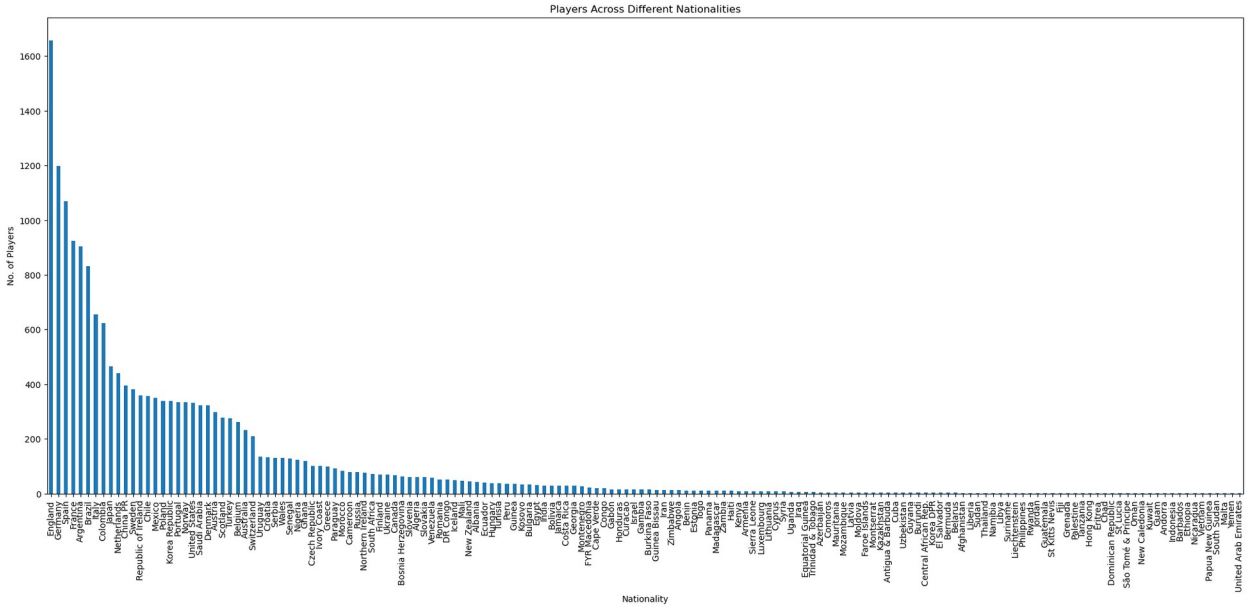
```
                 name                          full_name birth_date
age  \
17949     R. McKenzie                    Rory  McKenzie  10/7/1993
25
17950       M. Sipľak                     Michal Sipľak   2/2/1996
23
17951      J. Bekkema                       Jan Bekkema   4/9/1996
22
17952      A. Al Yami               Abdulrahman Al Yami  6/19/1997
21
17953  Júnior Brumado  José Francisco dos Santos Júnior  5/15/1999
19

       height_cm  weight_kgs  positions   nationality  overall_rating
\
17949     175.26        74.8  RM,CAM,CM      Scotland              67

17950     182.88        79.8         LB      Slovakia              59

17951     185.42        89.8         GK   Netherlands              59

17952     175.26        64.9      ST,LM  Saudi Arabia              59

17953     190.50        79.8         ST        Brazil              59


       potential  ...  long_shots  aggression interceptions
positioning  \
17949         70  ...          54          69            41
60
17950         67  ...          22          62            55
42
17951         67  ...           9          27            10
5
17952         71  ...          58          38            15
54
```

```
17953          75  ...           53           67              20
53

       vision  penalties composure  marking standing_tackle
sliding_tackle
17949      64         63        56       40              20
18
17950      39         32        52       53              64
60
17951      25         16        47        9              12
13
17952      52         50        53       16              18
17
17953      49         52        45       23              31
21

[5 rows x 51 columns]
```

```
print(data.shape)
```

```
(17954, 51)
```

```
#Q1
nat = data['nationality'].value_counts()
nat.plot(kind='bar', figsize=(25, 10))
plt.title('Players Across Different Nationalities')
plt.xlabel('Nationality')
plt.ylabel('No. of Players')
plt.show()
```



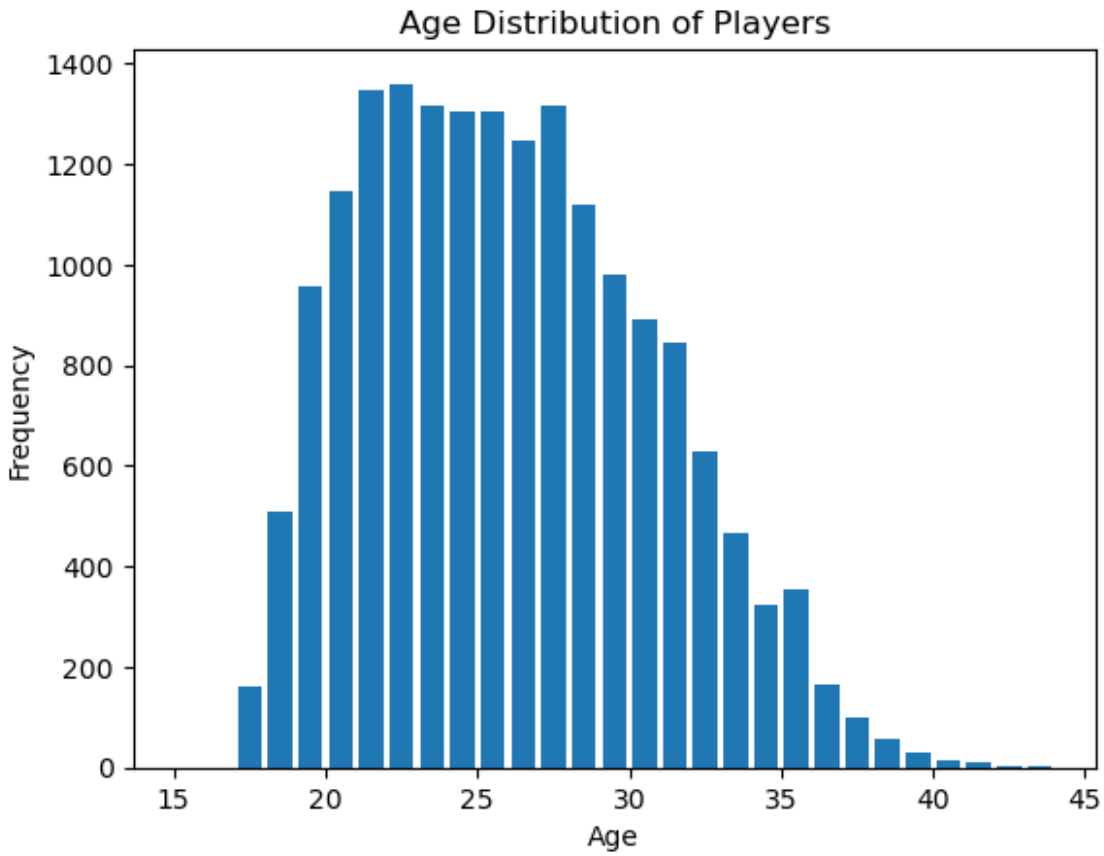
```python
#Q2
result = data.groupby('nationality')
['nationality'].count().reset_index(name='count')
result = result.sort_values(by='count', ascending=False).head(10)
print(result)
```

```
     nationality  count
44       England   1658
57       Germany   1199
133        Spain   1070
53        France    925
6       Argentina    904
18         Brazil    832
76          Italy    655
29       Colombia    624
79          Japan    466
104  Netherlands    441
```

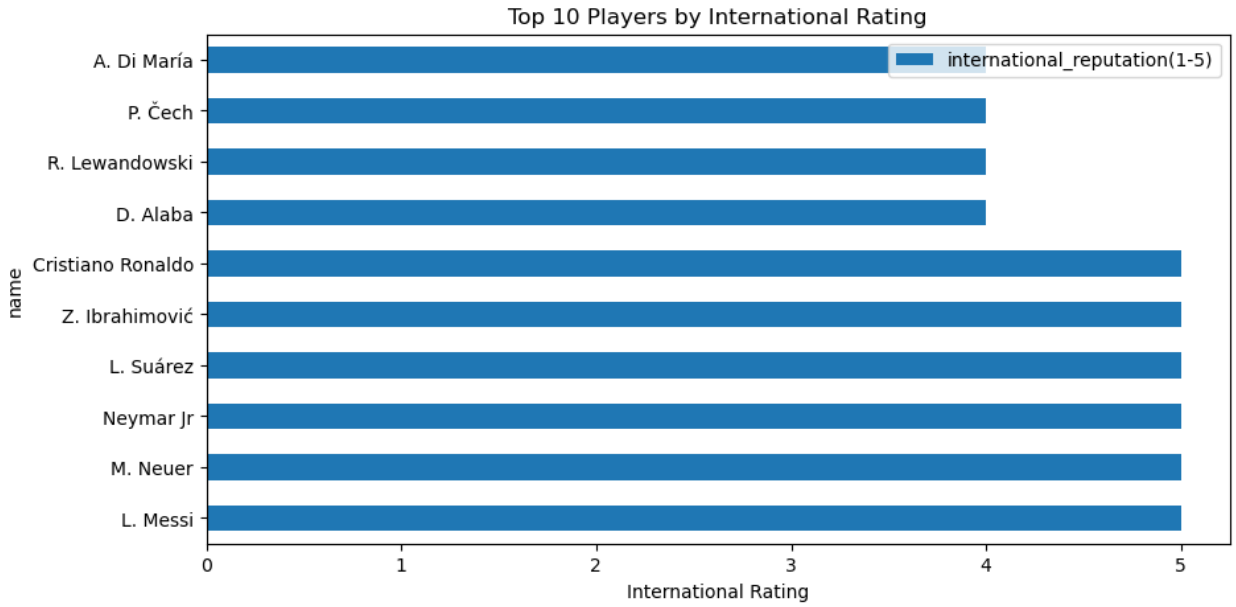
```python
#Q3
team_res = data.groupby('national_team')
['national_team'].count().reset_index(name='count')
team_res = team_res.sort_values(by='count', ascending=False).head(10)
print(team_res)
```

```
       national_team  count
31           Denmark     23
11            Brazil     23
83          Scotland     23
86             Spain     23
44           Germany     23
43            France     23
35           England     23
93     United States     23
61       Netherlands     23
65  Northern Ireland     22
```

```python
#Q4
age = data['age']
age.plot(kind='hist', bins=range(15, 45), rwidth=0.8)
plt.title('Age Distribution of Players')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



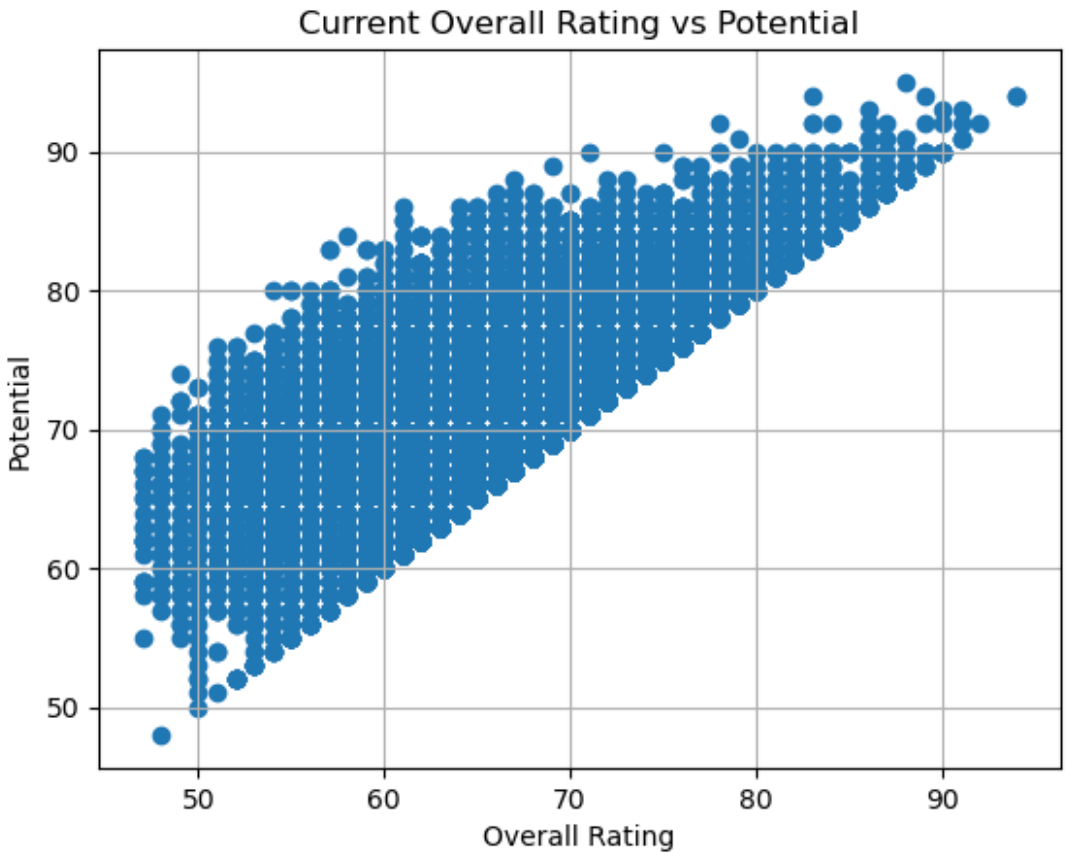

```
#Q5
inter_rat = data[['name', 'international_reputation(1-
5)']].sort_values(by='international_reputation(1-5)',
ascending=False).head(10)
inter_rat.set_index('name', inplace=True)
inter_rat.plot(kind='barh', figsize=(10, 5))
plt.title('Top 10 Players by International Rating')
plt.xlabel('International Rating')
plt.show()
```



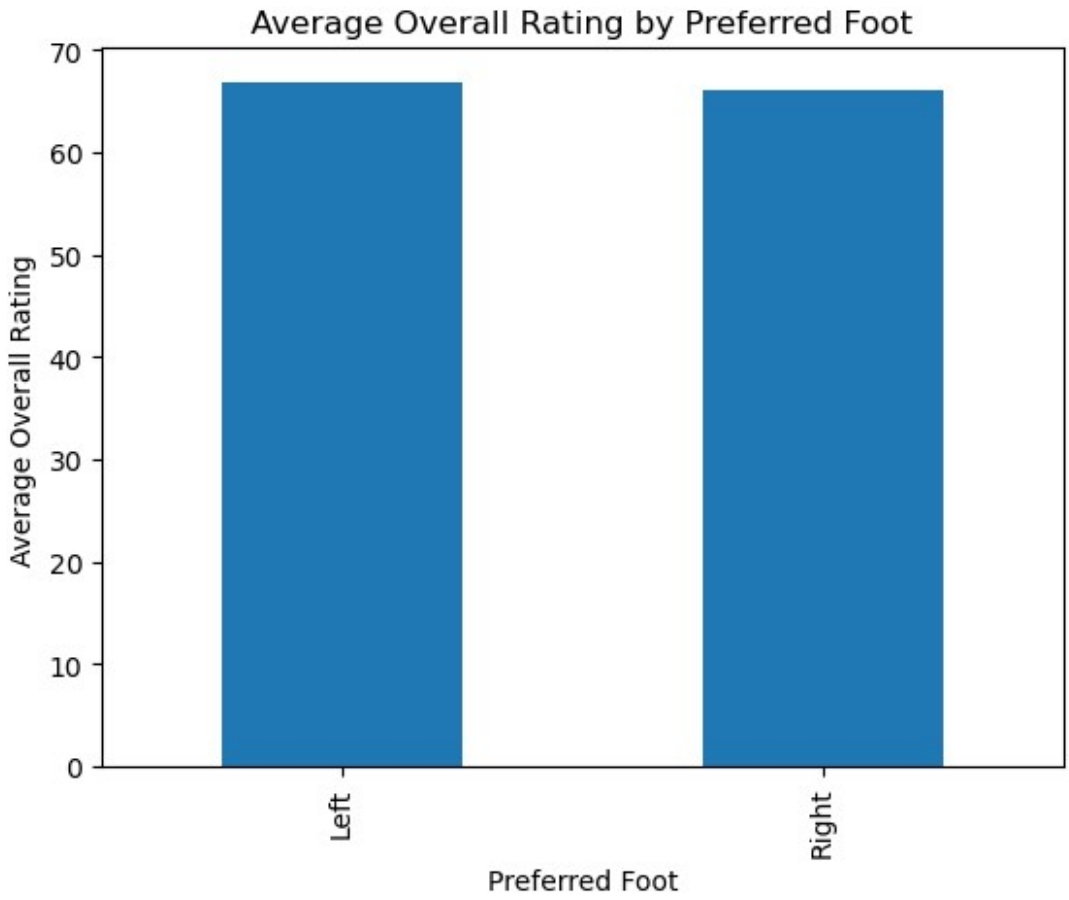

```
#Q6
best_plr = data.loc[data.groupby('nationality')
['overall_rating'].idxmax()]
best_plr = best_plr[['nationality', 'name', 'overall_rating']]
best_plr_asc = best_plr.sort_values(by='overall_rating',
ascending=False).reset_index(drop=True)
print(best_plr_asc)
```

```
          nationality                name  overall_rating
0            Portugal  Cristiano Ronaldo              94
1           Argentina           L. Messi              94
2              Brazil          Neymar Jr              92
3             Belgium          E. Hazard              91
4               Spain             De Gea              91
..                ...                 ...             ...
155         Hong Kong           F. Baise              61
156  Papua New Guinea        A. Komolong              61
157       South Sudan           K. Athiu              60
158             Yemen        A. Ba Saeed              58
159         Indonesia   E. Maulana Vikri              56

[160 rows x 3 columns]
```

```
#Q7
plt.scatter(data['overall_rating'], data['potential'])
plt.title('Current Overall Rating vs Potential')
plt.xlabel('Overall Rating')
plt.ylabel('Potential')
plt.grid(True)
plt.show()
```




```
#Q8
foot = data.groupby('preferred_foot')['overall_rating'].mean()
foot.plot(kind='bar')
plt.title('Average Overall Rating by Preferred Foot')
plt.xlabel('Preferred Foot')
plt.ylabel('Average Overall Rating')
plt.show()
```


Average Overall Rating by Preferred Foot
Average Overall Rating
70
60
50
40
30
20
10
0
Left
Right
Preferred Foot