FIFA Analysis

July 7, 2025

0.1 Importing all the required libraries

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

[2]:
```
df = pd.read_csv('players_20.csv')
```

[3]:
```
df.head()
```

[3]:
```
   sofifa_id                                     player_url  \
0     158023  https://sofifa.com/player/158023/lionel-messi/…
1      20801  https://sofifa.com/player/20801/c-ronaldo-dos-…
2     190871  https://sofifa.com/player/190871/neymar-da-sil…
3     183277  https://sofifa.com/player/183277/eden-hazard/2…
4     192985  https://sofifa.com/player/192985/kevin-de-bruy…

          short_name                             long_name player_positions  \
0          L. Messi       Lionel Andrés Messi Cuccittini       RW, CF, ST
1  Cristiano Ronaldo  Cristiano Ronaldo dos Santos Aveiro           ST, LW
2         Neymar Jr        Neymar da Silva Santos Júnior          LW, CAM
3          E. Hazard                          Eden Hazard           LW, CF
4       K. De Bruyne                       Kevin De Bruyne          CAM, CM

   overall  potential    value_eur  wage_eur  age  …    lcb    cb    rcb  \
0       94         94   95500000.0  560000.0   32  …  53+6  53+6  53+6
1       93         93   58500000.0  410000.0   34  …  54+3  54+3  54+3
2       92         92  105500000.0  290000.0   27  …  47+6  47+6  47+6
3       91         91   90000000.0  470000.0   28  …  49+6  49+6  49+6
4       91         91   90000000.0  370000.0   28  …  67+3  67+3  67+3

     rb    gk                                 player_face_url  \
0  63+6  19+6  https://cdn.sofifa.net/players/158/023/20_120.png
1  61+3  20+3  https://cdn.sofifa.net/players/020/801/20_120.png
2  61+6  20+6  https://cdn.sofifa.net/players/190/871/20_120.png
3  61+6  18+6  https://cdn.sofifa.net/players/183/277/20_120.png
4  73+3  21+3  https://cdn.sofifa.net/players/192/985/20_120.png
```

```
                              club_logo_url  \
0  https://cdn.sofifa.net/teams/241/60.png
1   https://cdn.sofifa.net/teams/45/60.png
2   https://cdn.sofifa.net/teams/73/60.png
3  https://cdn.sofifa.net/teams/243/60.png
4   https://cdn.sofifa.net/teams/10/60.png

                              club_flag_url  \
0      https://cdn.sofifa.net/flags/es.png
1      https://cdn.sofifa.net/flags/it.png
2      https://cdn.sofifa.net/flags/fr.png
3      https://cdn.sofifa.net/flags/es.png
4  https://cdn.sofifa.net/flags/gb-eng.png

                            nation_logo_url  \
0                                       NaN
1  https://cdn.sofifa.net/teams/1354/60.png
2  https://cdn.sofifa.net/teams/1370/60.png
3  https://cdn.sofifa.net/teams/1325/60.png
4  https://cdn.sofifa.net/teams/1325/60.png

                        nation_flag_url
0  https://cdn.sofifa.net/flags/ar.png
1  https://cdn.sofifa.net/flags/pt.png
2  https://cdn.sofifa.net/flags/br.png
3  https://cdn.sofifa.net/flags/be.png
4  https://cdn.sofifa.net/flags/be.png

[5 rows x 110 columns]
```

[4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18483 entries, 0 to 18482
Columns: 110 entries, sofifa_id to nation_flag_url
dtypes: float64(16), int64(44), object(50)
memory usage: 15.5+ MB
```

[5]:
```
for col in df.columns:
    print(col)
```

```
sofifa_id
player_url
short_name
long_name
player_positions
overall
```

potential
value_eur
wage_eur
age
dob
height_cm
weight_kg
club_team_id
club_name
league_name
league_level
club_position
club_jersey_number
club_loaned_from
club_joined
club_contract_valid_until
nationality_id
nationality_name
nation_team_id
nation_position
nation_jersey_number
preferred_foot
weak_foot
skill_moves
international_reputation
work_rate
body_type
real_face
release_clause_eur
player_tags
player_traits
pace
shooting
passing
dribbling
defending
physic
attacking_crossing
attacking_finishing
attacking_heading_accuracy
attacking_short_passing
attacking_volleys
skill_dribbling
skill_curve
skill_fk_accuracy
skill_long_passing
skill_ball_control
movement_acceleration

movement_sprint_speed
movement_agility
movement_reactions
movement_balance
power_shot_power
power_jumping
power_stamina
power_strength
power_long_shots
mentality_aggression
mentality_interceptions
mentality_positioning
mentality_vision
mentality_penalties
mentality_composure
defending_marking_awareness
defending_standing_tackle
defending_sliding_tackle
goalkeeping_diving
goalkeeping_handling
goalkeeping_kicking
goalkeeping_positioning
goalkeeping_reflexes
goalkeeping_speed
ls
st
rs
lw
lf
cf
rf
rw
lam
cam
ram
lm
lcm
cm
rcm
rm
lwb
ldm
cdm
rdm
rwb
lb
lcb
cb

```
rcb
rb
gk
player_face_url
club_logo_url
club_flag_url
nation_logo_url
nation_flag_url
```

[6]:
```
df.shape
```

[6]:
```
(18483, 110)
```

0.2 Visualizations

[7]:
```
player_salary = df[['short_name','wage_eur']]
```

[8]:
```
player_salary.head(10)
```

[8]:

	short_name	wage_eur
0	L. Messi	560000.0
1	Cristiano Ronaldo	410000.0
2	Neymar Jr	290000.0
3	E. Hazard	470000.0
4	K. De Bruyne	370000.0
5	J. Oblak	125000.0
6	L. Modrić	340000.0
7	M. ter Stegen	250000.0
8	V. van Dijk	200000.0
9	M. Salah	240000.0

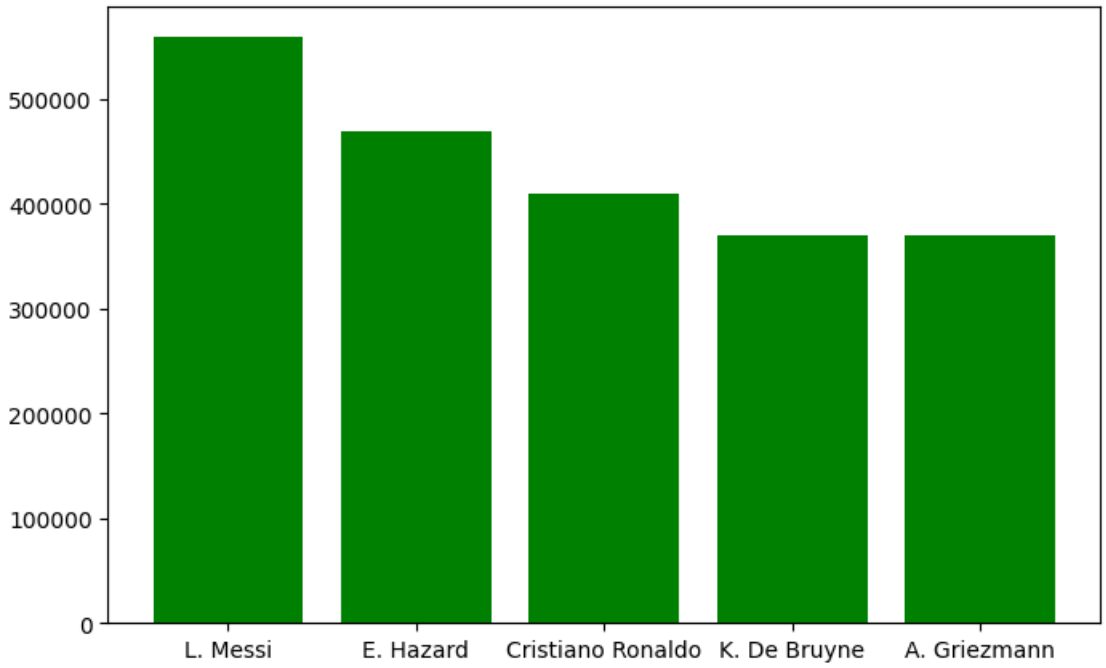
[9]:
```
player_salary = player_salary.sort_values(by=['wage_eur'],ascending = False)
```

[10]:
```
player_salary.head()
```

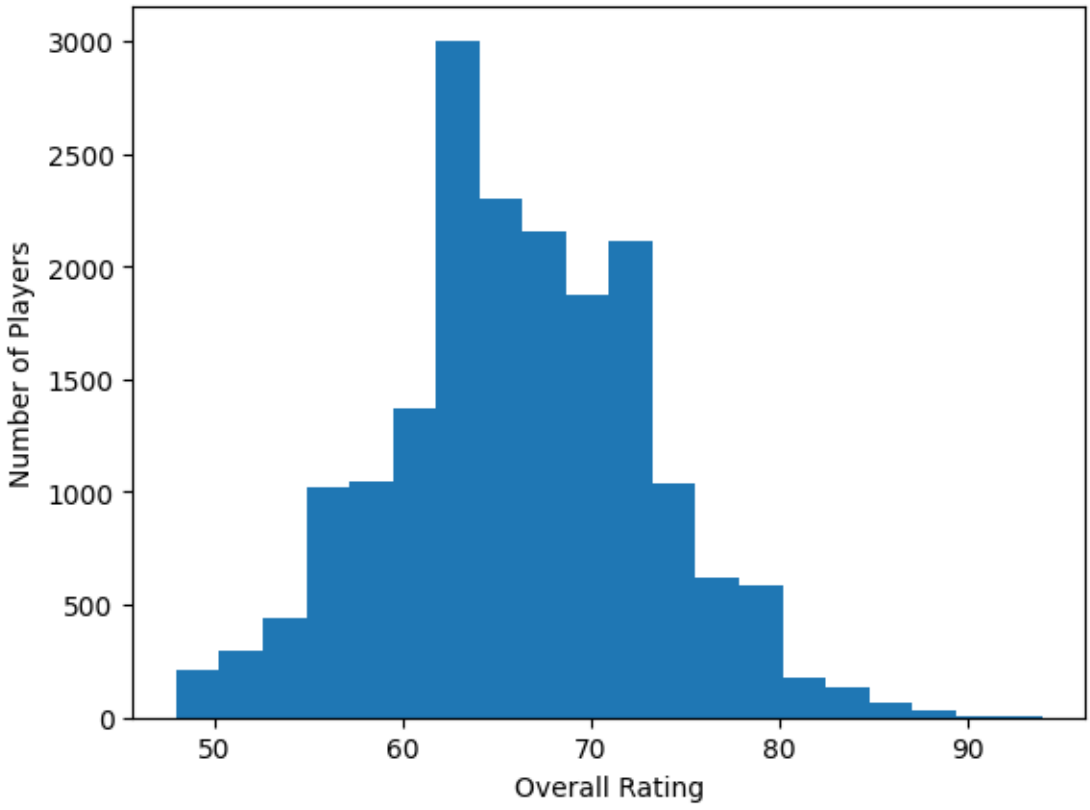
[10]:

	short_name	wage_eur
0	L. Messi	560000.0
3	E. Hazard	470000.0
1	Cristiano Ronaldo	410000.0
4	K. De Bruyne	370000.0
17	A. Griezmann	370000.0

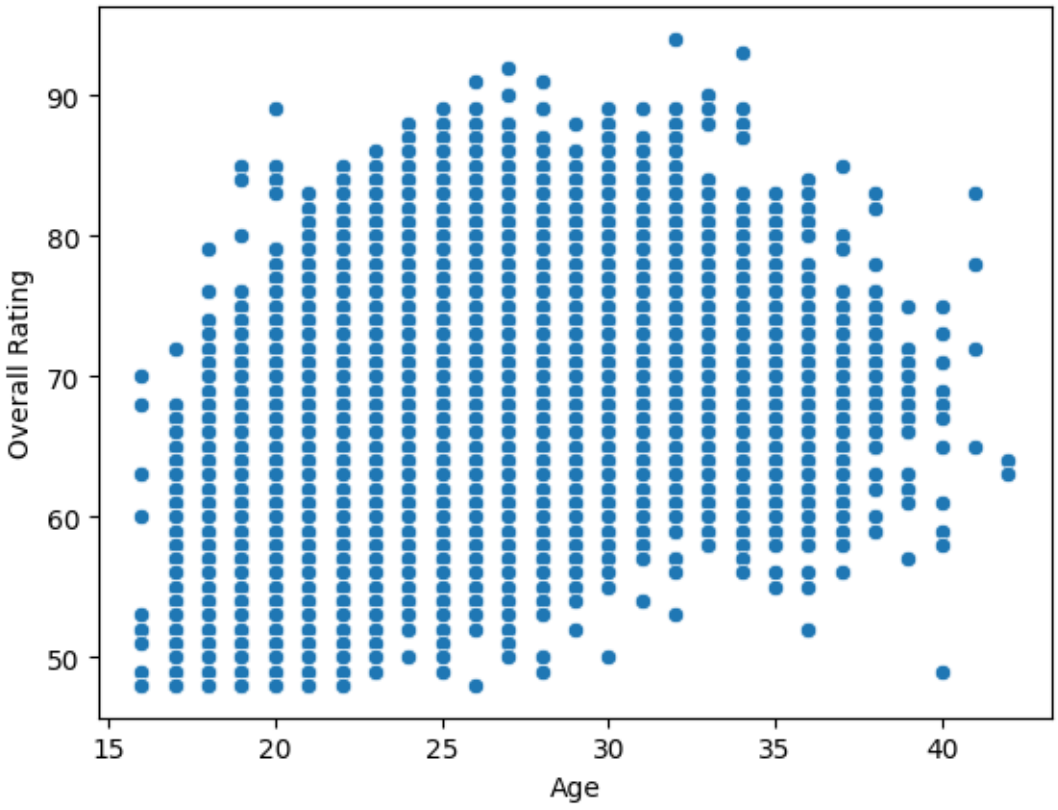
[11]:
```
plt.figure(figsize=(8,5))
plt.bar(list(player_salary['short_name'])[0:
  ↪5],list(player_salary['wage_eur'])[0:5],color = 'g')
plt.show()
```



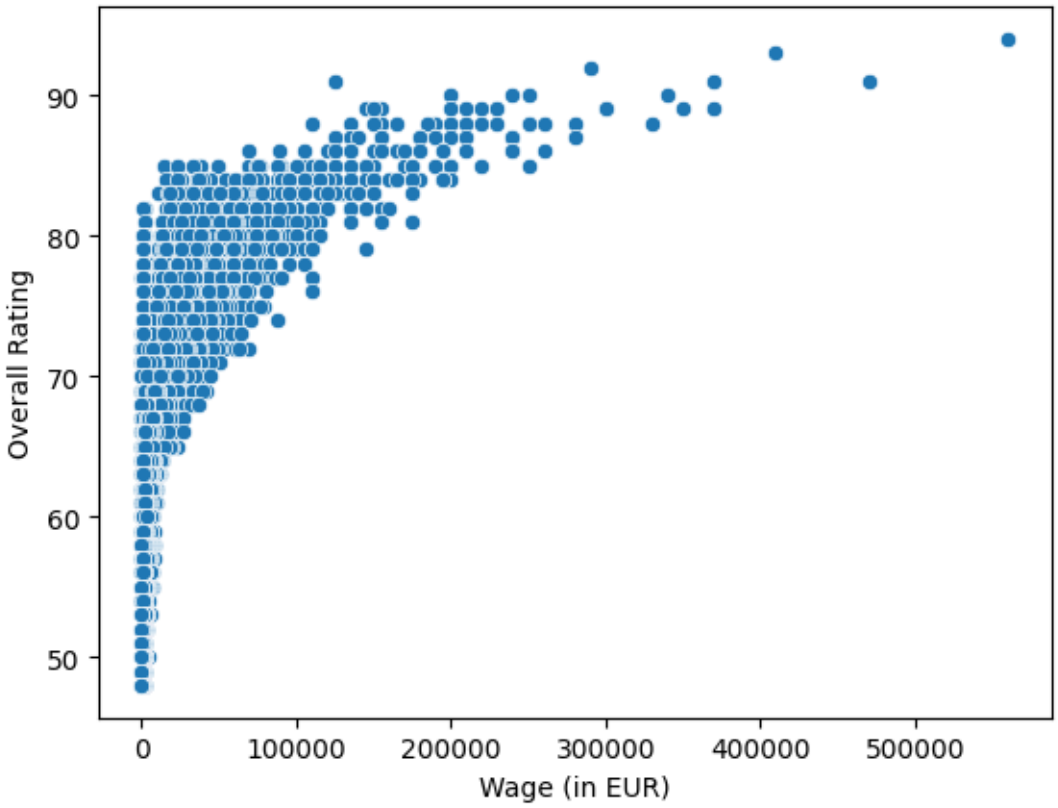
[12]:
```
plt.hist(df['overall'], bins=20)
plt.xlabel('Overall Rating')
plt.ylabel('Number of Players')
plt.show()
```



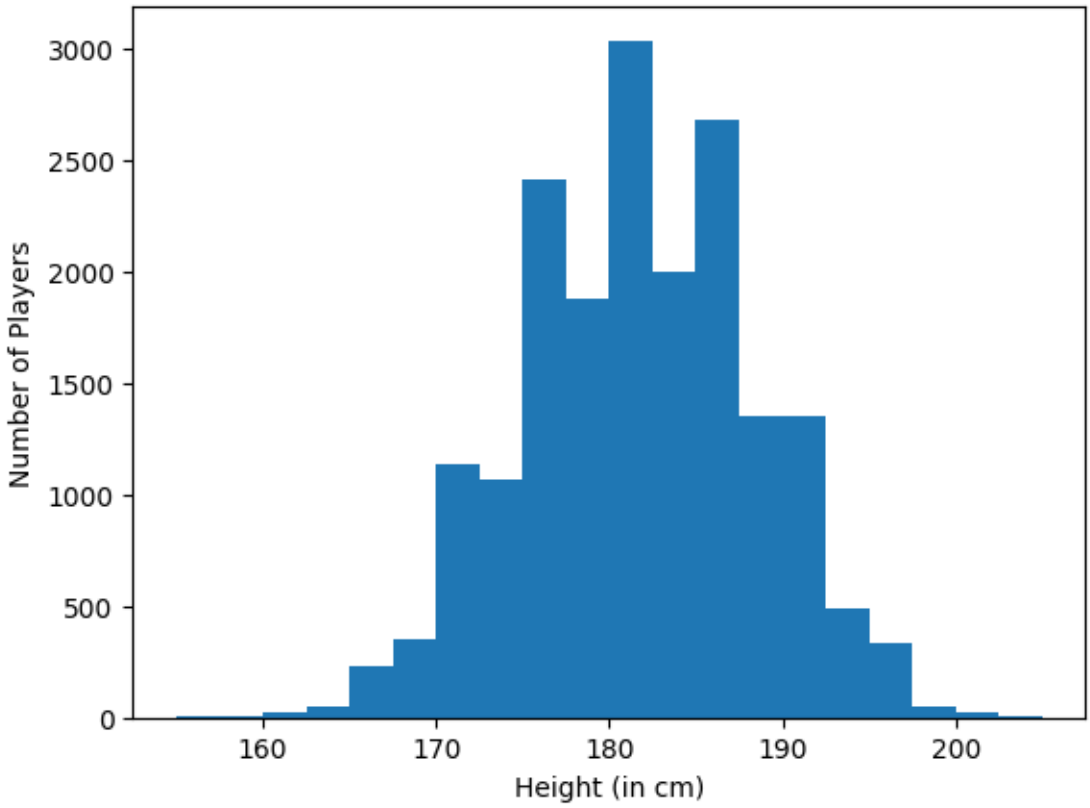
[13]:
```
sb.scatterplot(x='age', y='overall', data=df)
plt.xlabel('Age')
plt.ylabel('Overall Rating')
plt.show()
```



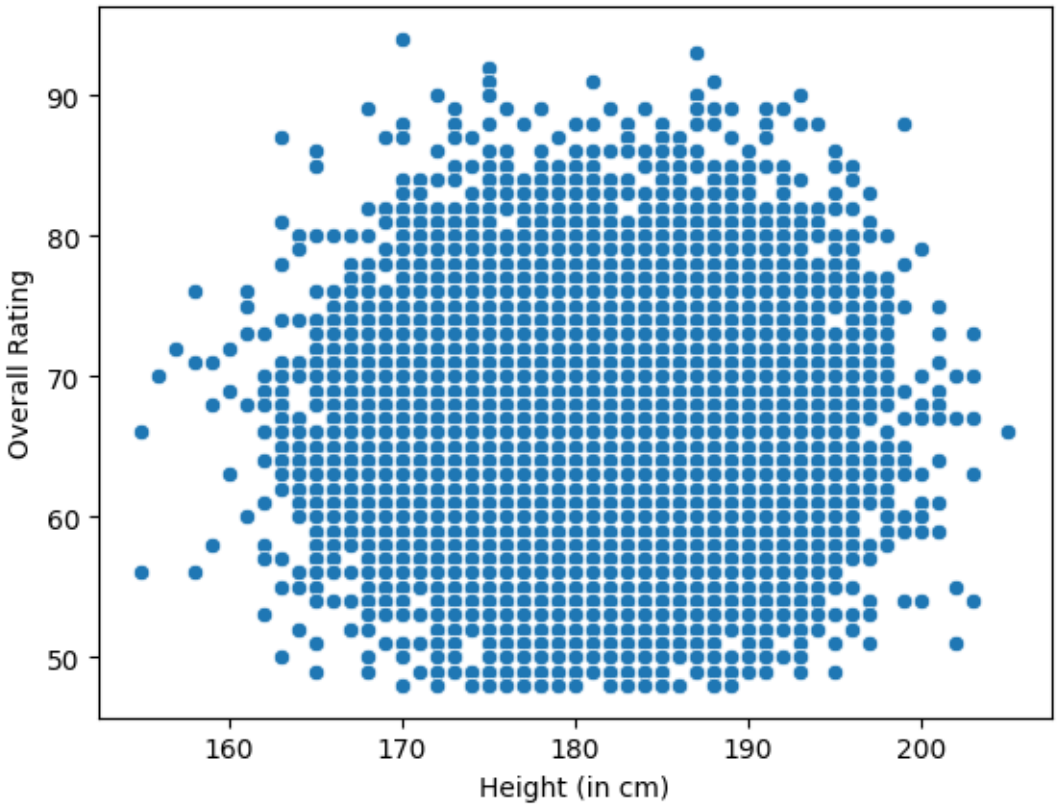
[14]:
```
sb.scatterplot(x='wage_eur', y='overall', data=df)
plt.xlabel('Wage (in EUR)')
plt.ylabel('Overall Rating')
plt.show()
```



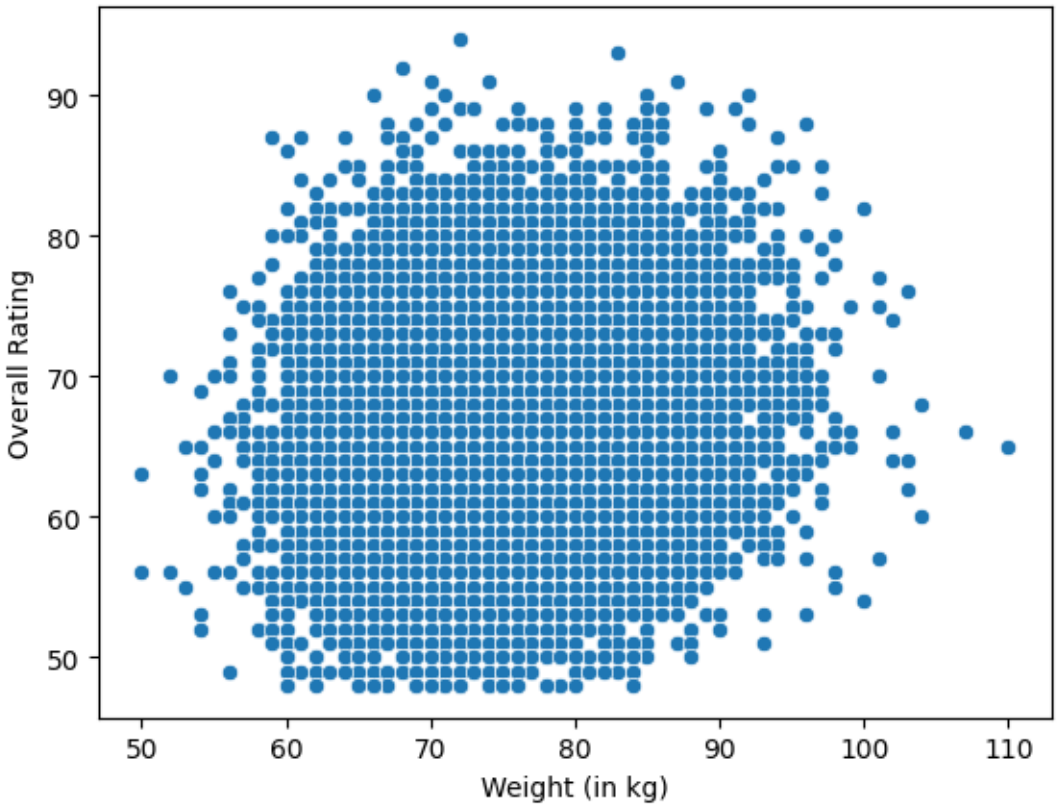
[15]:
```
plt.hist(df['height_cm'], bins=20)
plt.xlabel('Height (in cm)')
plt.ylabel('Number of Players')
plt.show()
```



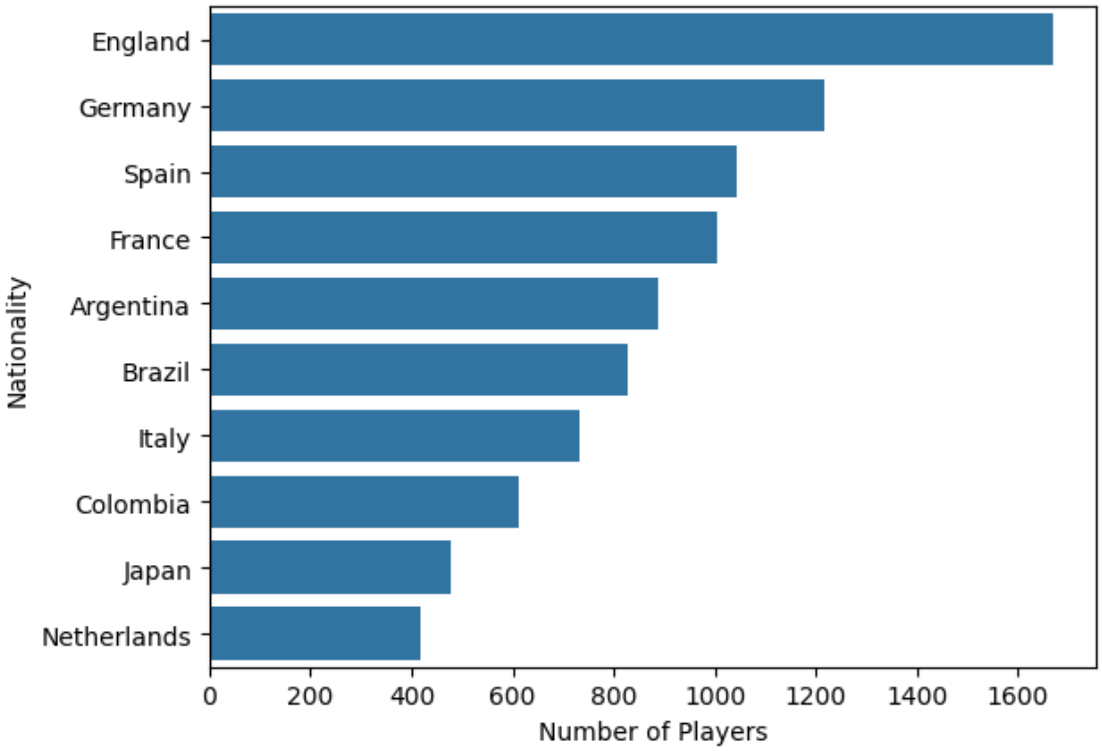
[16]:
```
sb.scatterplot(x='height_cm', y='overall', data=df)
plt.xlabel('Height (in cm)')
plt.ylabel('Overall Rating')
plt.show()
```



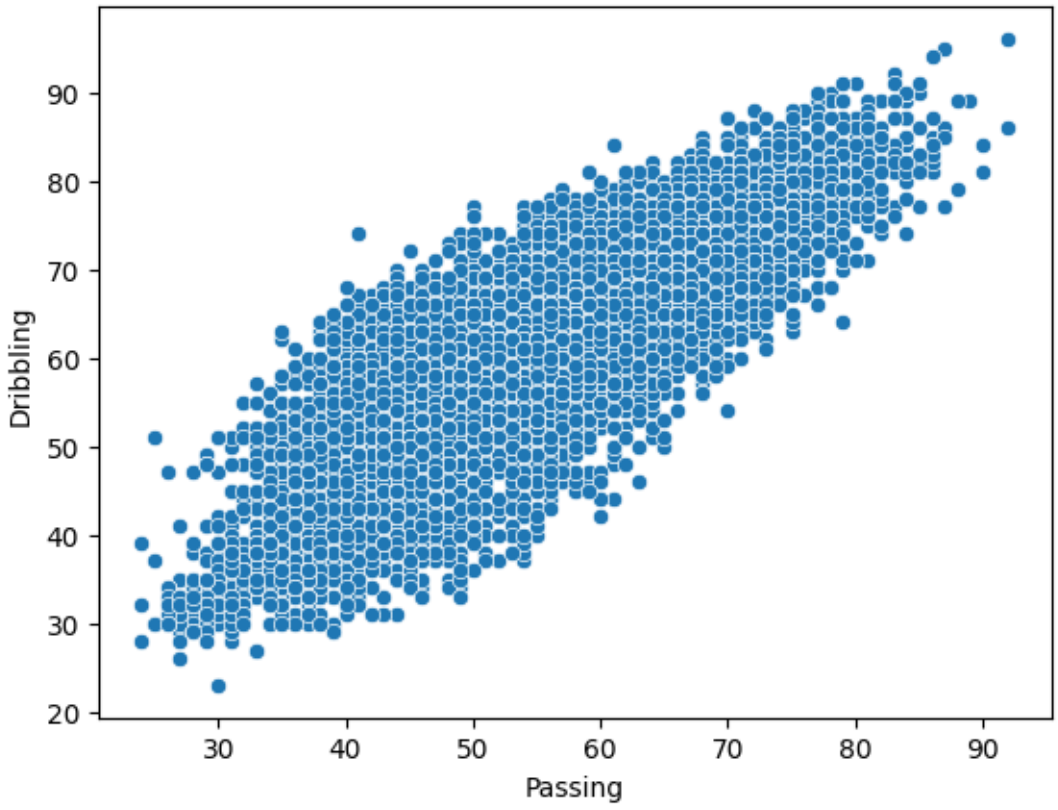
[17]:
```
sb.scatterplot(x='weight_kg', y='overall', data=df)
plt.xlabel('Weight (in kg)')
plt.ylabel('Overall Rating')
plt.show()
```



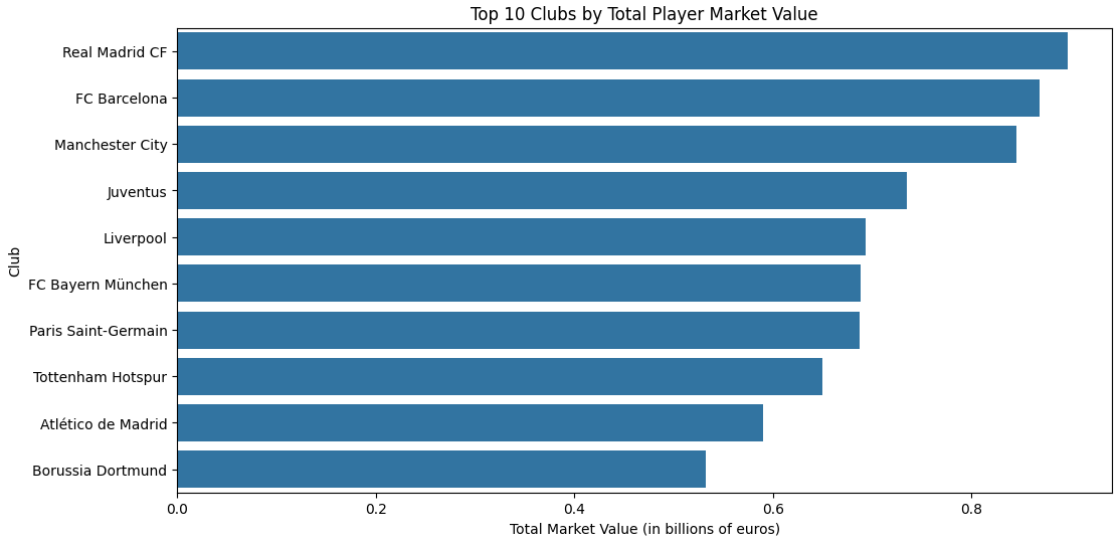
[19]:
```
sb.countplot(y='nationality_name', data=df, order=df['nationality_name'].
  ↪value_counts().iloc[:10].index)
plt.xlabel('Number of Players')
plt.ylabel('Nationality')
plt.show()
```



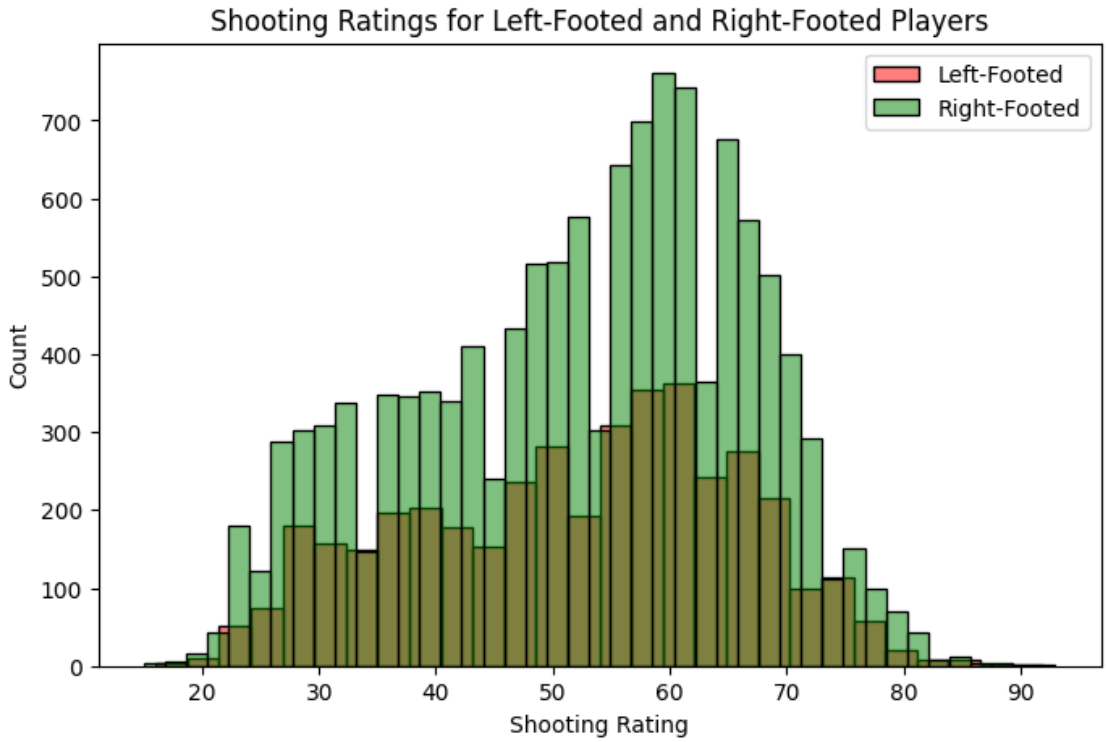
[20]:
```
sb.scatterplot(x='passing', y='dribbling', data=df)
plt.xlabel('Passing')
plt.ylabel('Dribbling')
plt.show()
```



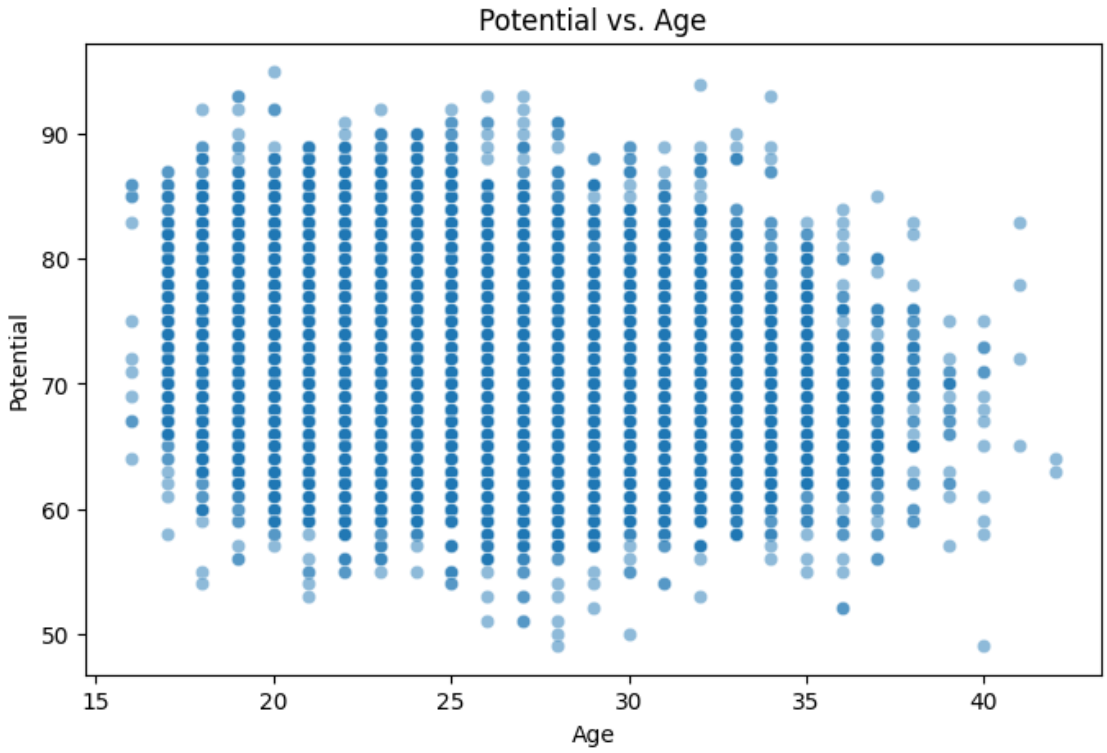
[22]:
```
club_value = df.groupby('club_name')['value_eur'].sum().
  ↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Total Player Market Value')
sb.barplot(x=club_value.values/1e9, y=club_value.index)
plt.xlabel('Total Market Value (in billions of euros)')
plt.ylabel('Club')

plt.show()
```



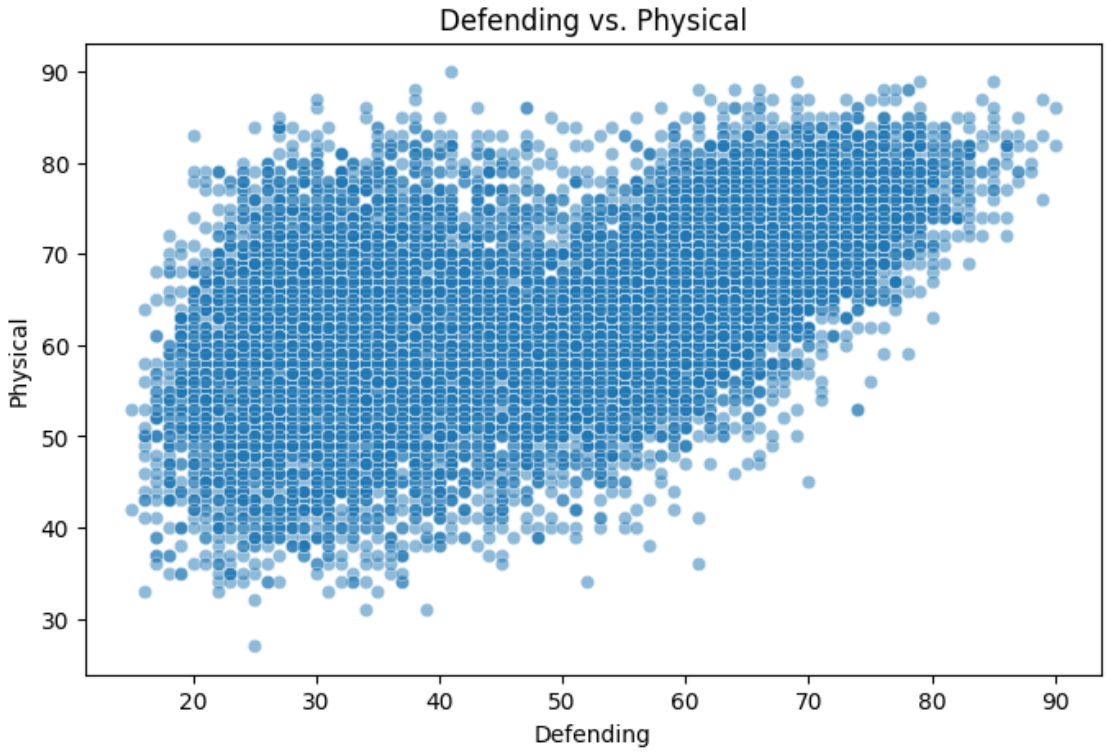
[23]:
```
left_footed_shooting = df[df['preferred_foot'] == 'Left']['shooting']
right_footed_shooting = df[df['preferred_foot'] == 'Right']['shooting']

plt.figure(figsize=(8, 5))
plt.title('Shooting Ratings for Left-Footed and Right-Footed Players')
sb.histplot(left_footed_shooting, color='r', alpha=0.5, label='Left-Footed')
sb.histplot(right_footed_shooting, color='g', alpha=0.5, label='Right-Footed')
plt.xlabel('Shooting Rating')
plt.ylabel('Count')
plt.legend()
plt.show()
```



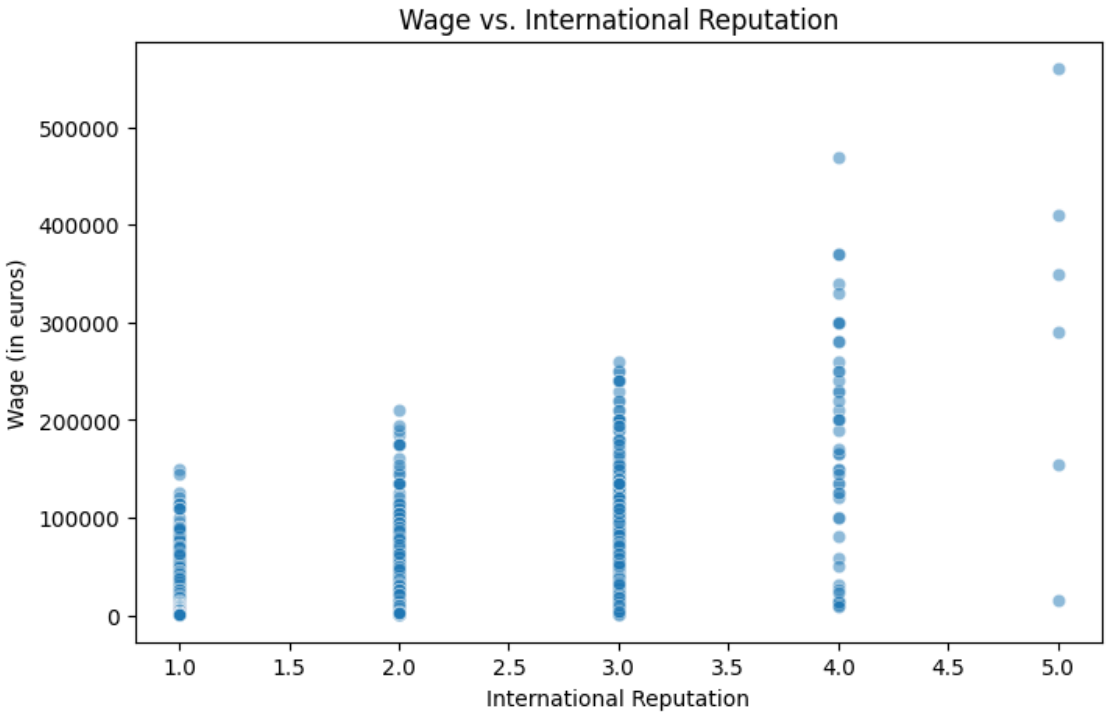
[24]:
```
plt.figure(figsize=(8, 5))
plt.title('Potential vs. Age')
sb.scatterplot(x='age', y='potential', data=df, alpha=0.5)
plt.xlabel('Age')
plt.ylabel('Potential')
plt.show()
```



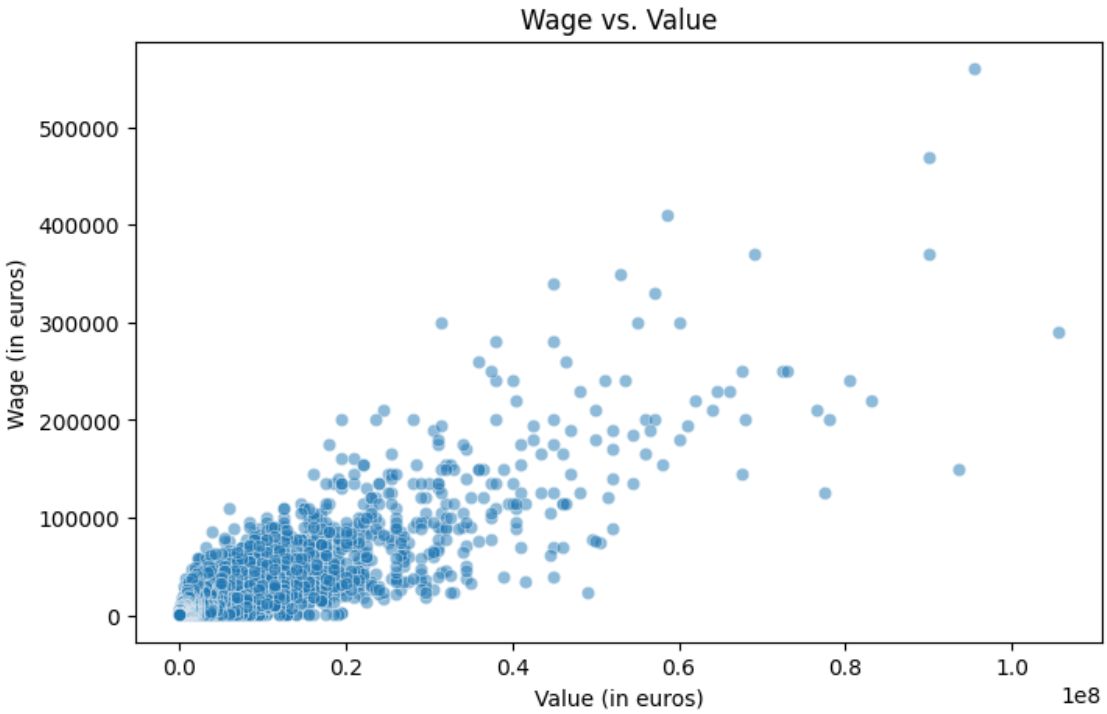
[25]:
```
plt.figure(figsize=(8, 5))
plt.title('Defending vs. Physical')
sb.scatterplot(x='defending', y='physic', data=df, alpha=0.5)
plt.xlabel('Defending')
plt.ylabel('Physical')
plt.show()
```



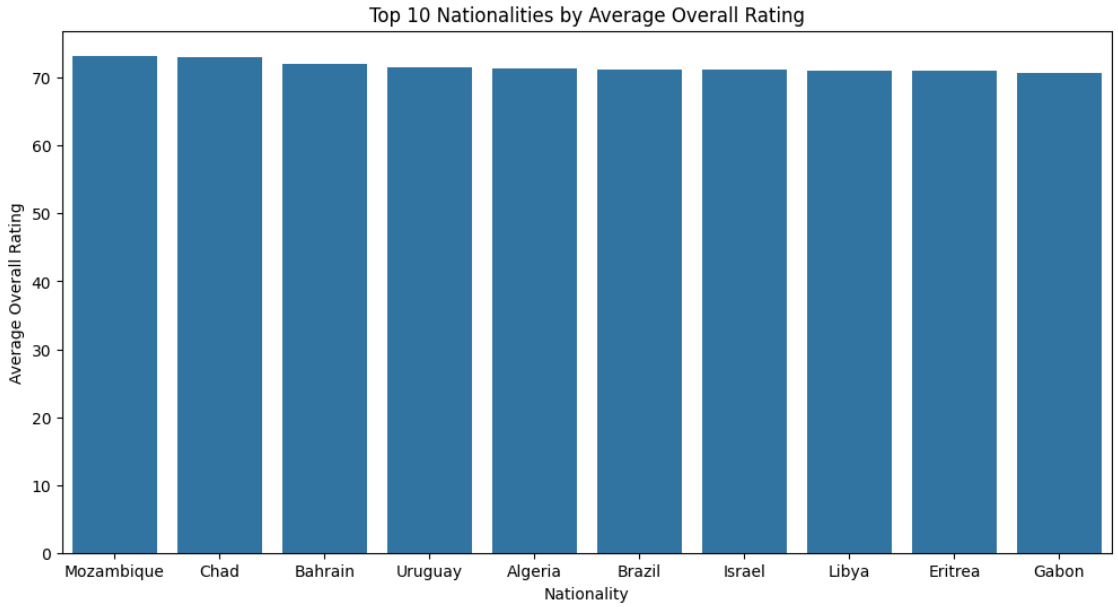
[26]:
```
plt.figure(figsize=(8, 5))
plt.title('Wage vs. International Reputation')
sb.scatterplot(x='international_reputation', y='wage_eur', data=df, alpha=0.5)
plt.xlabel('International Reputation')
plt.ylabel('Wage (in euros)')
plt.show()
```



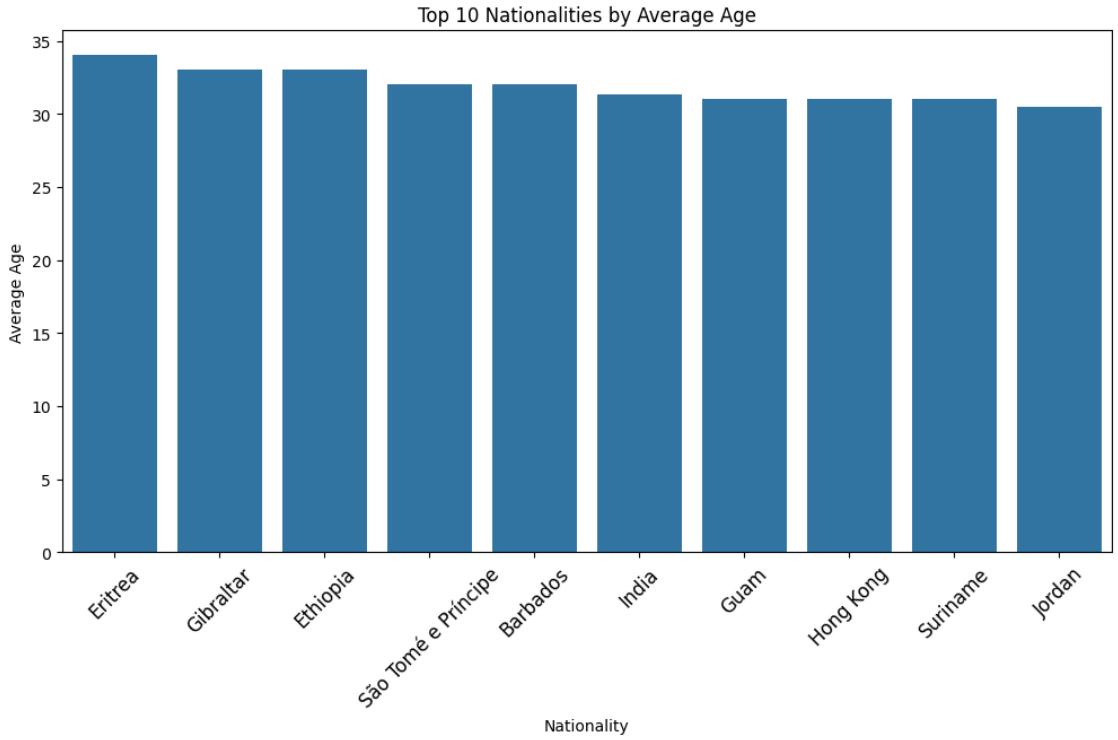
[27]:
```
plt.figure(figsize=(8, 5))
plt.title('Wage vs. Value')
sb.scatterplot(x='value_eur', y='wage_eur', data=df, alpha=0.5)
plt.xlabel('Value (in euros)')
plt.ylabel('Wage (in euros)')
plt.show()
```



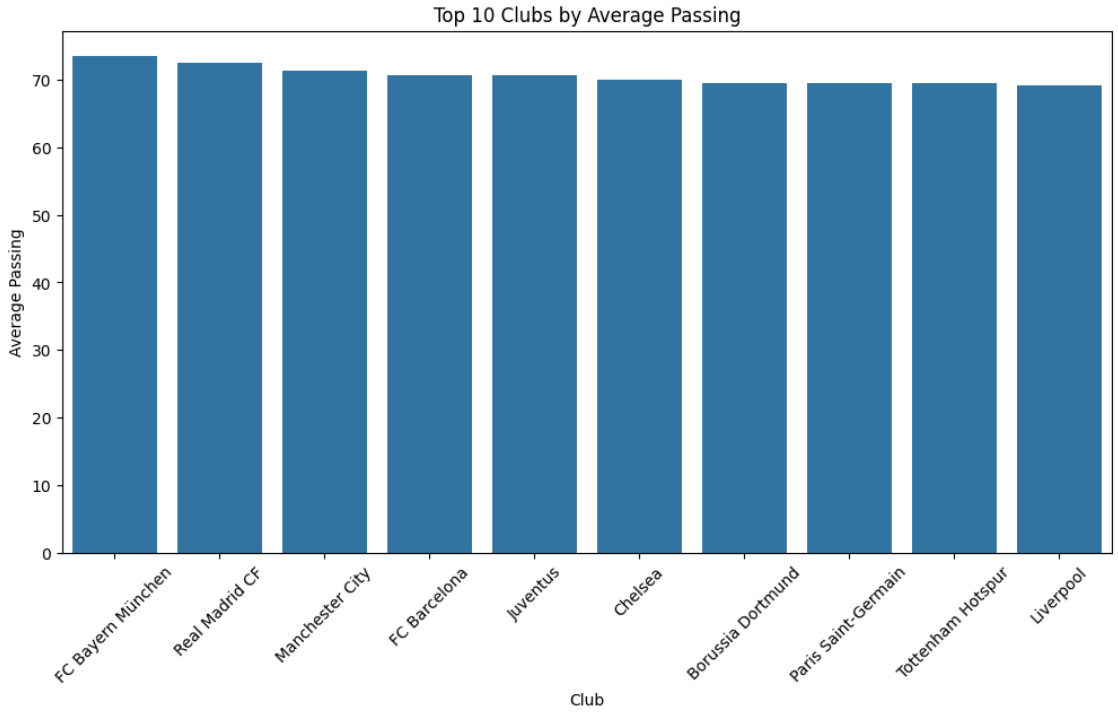
[28]:
```
nationality_overall = df.groupby('nationality_name')['overall'].mean().
↪sort_values(ascending=False).iloc[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Nationalities by Average Overall Rating')
sb.barplot(x=nationality_overall.index, y=nationality_overall.values)
plt.xlabel('Nationality')
plt.ylabel('Average Overall Rating')
plt.show()
```



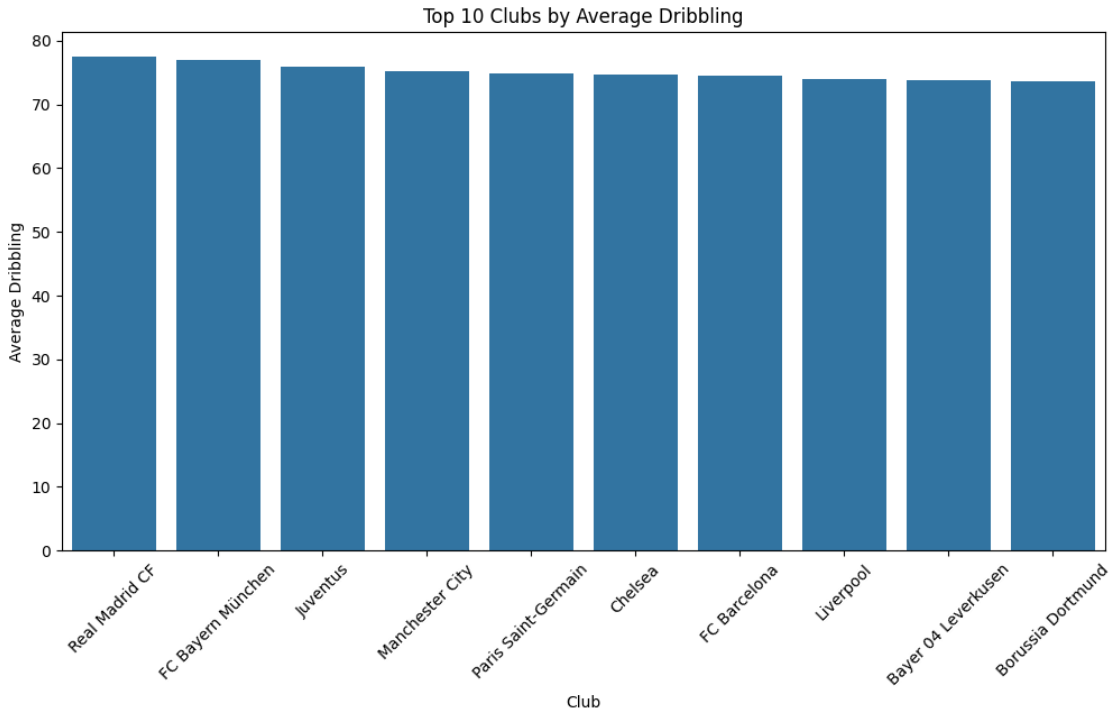
[29]:
```
nationality_age = df.groupby('nationality_name')['age'].mean().
↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Nationalities by Average Age')
sb.barplot(x=nationality_age.index, y=nationality_age.values)
plt.xlabel('Nationality')
plt.ylabel('Average Age')
plt.xticks(rotation=45, fontsize=12)
plt.show()
```



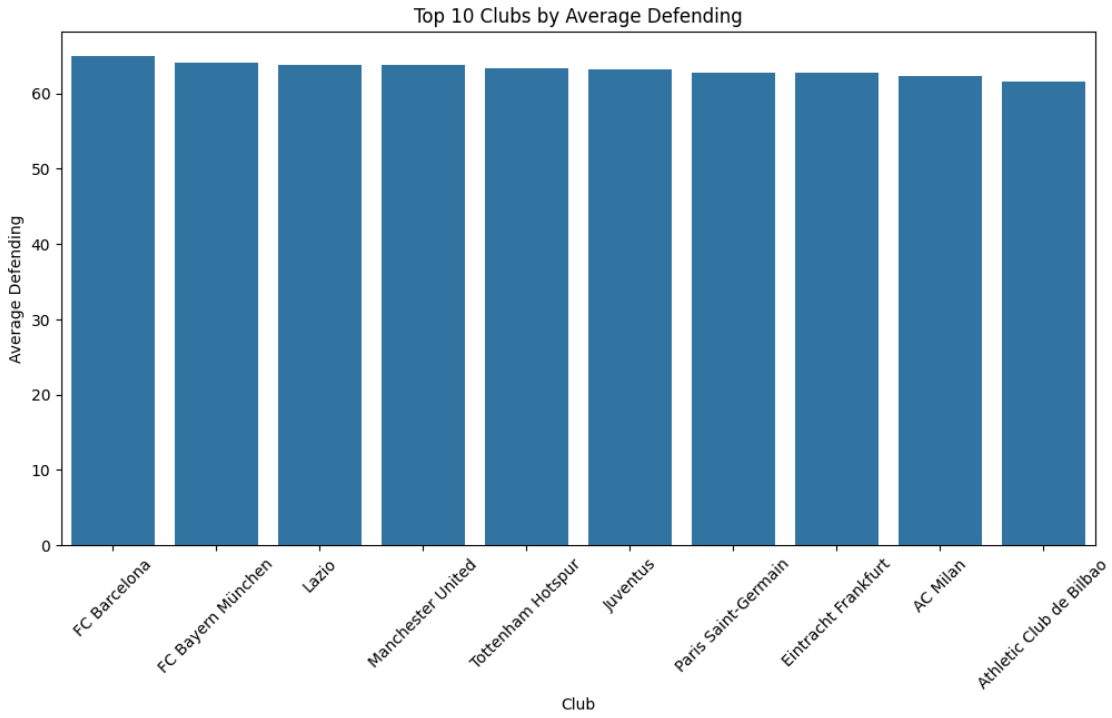
[30]:
```
passing_top10 = df.groupby('club_name')['passing'].mean().
  ↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Passing')
sb.barplot(x=passing_top10.index, y=passing_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Passing')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



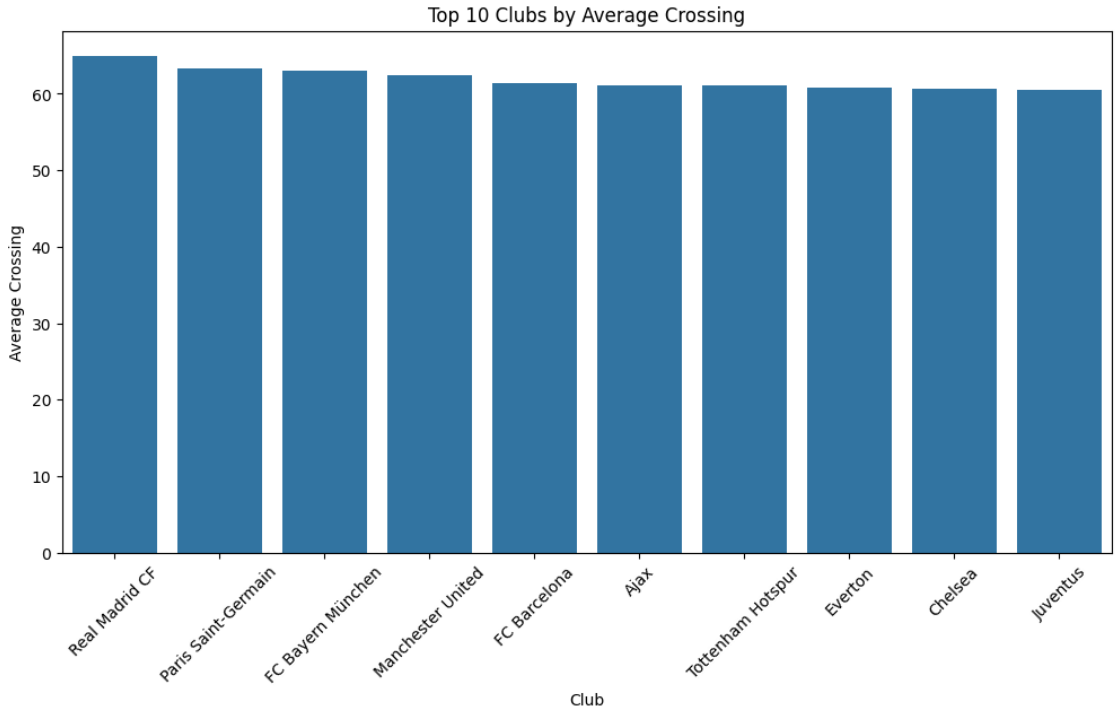
[31]:
```
dribbling_top10 = df.groupby('club_name')['dribbling'].mean().
 ↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Dribbling')
sb.barplot(x=dribbling_top10.index, y=dribbling_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Dribbling')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



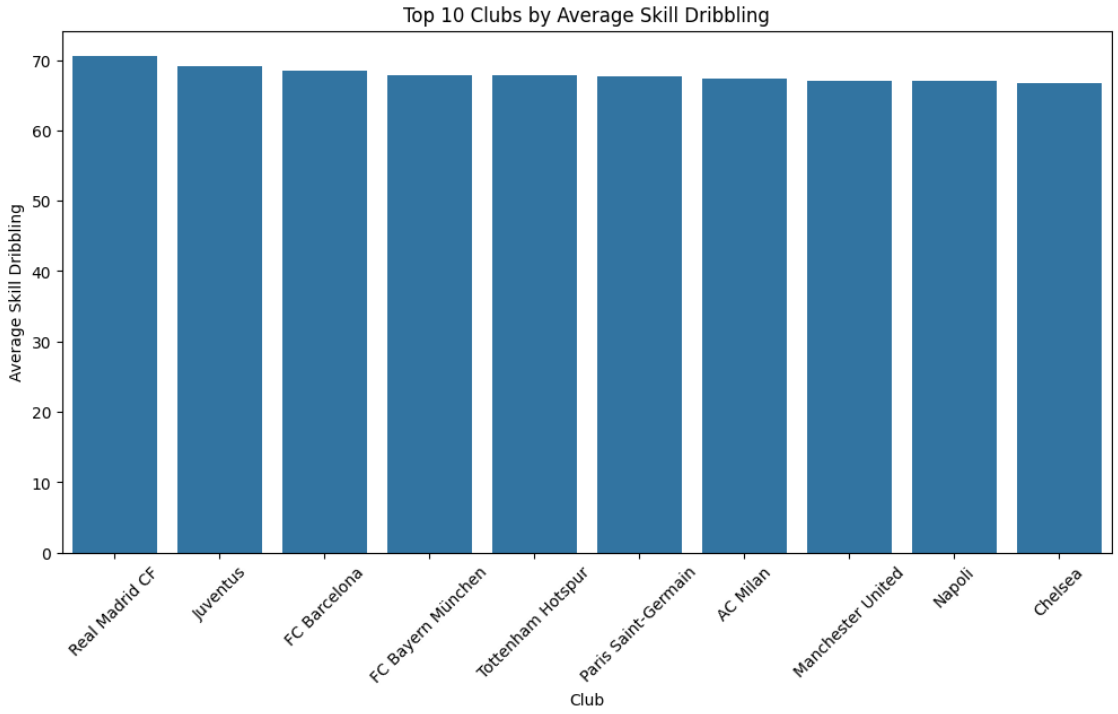
[32]:
```
defending_top10 = df.groupby('club_name')['defending'].mean().
 ↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Defending')
sb.barplot(x=defending_top10.index, y=defending_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Defending')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



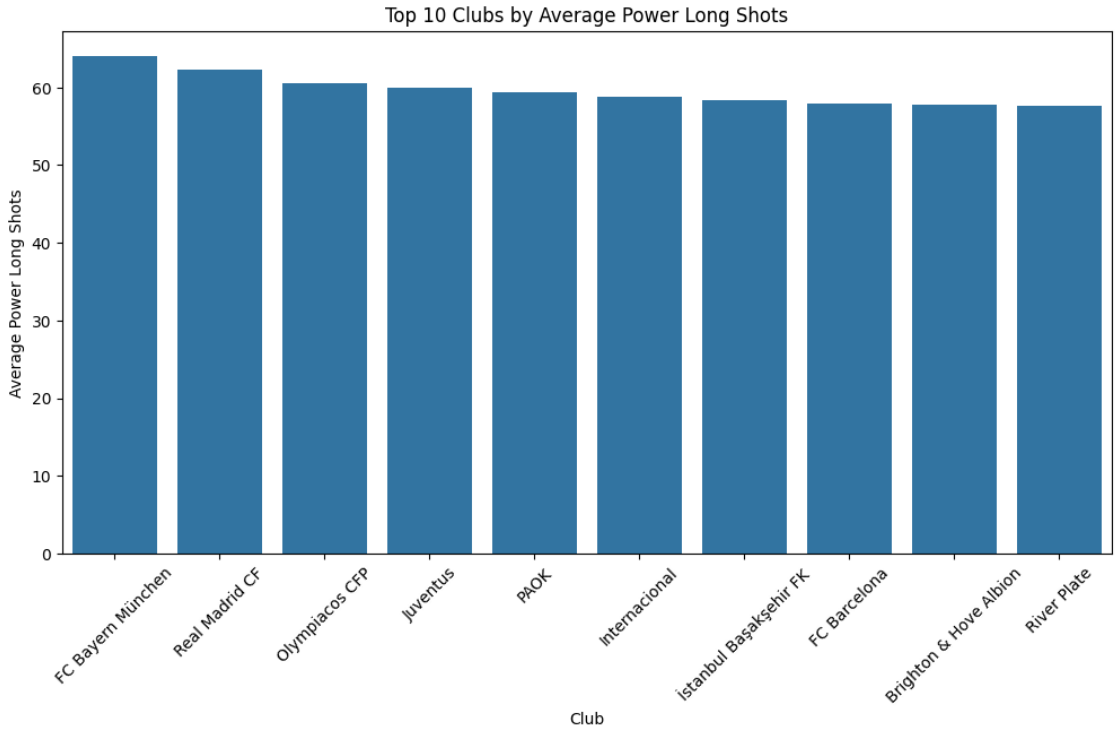
[33]:
```
crossing_top10 = df.groupby('club_name')['attacking_crossing'].mean().
↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Crossing')
sb.barplot(x=crossing_top10.index, y=crossing_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Crossing')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



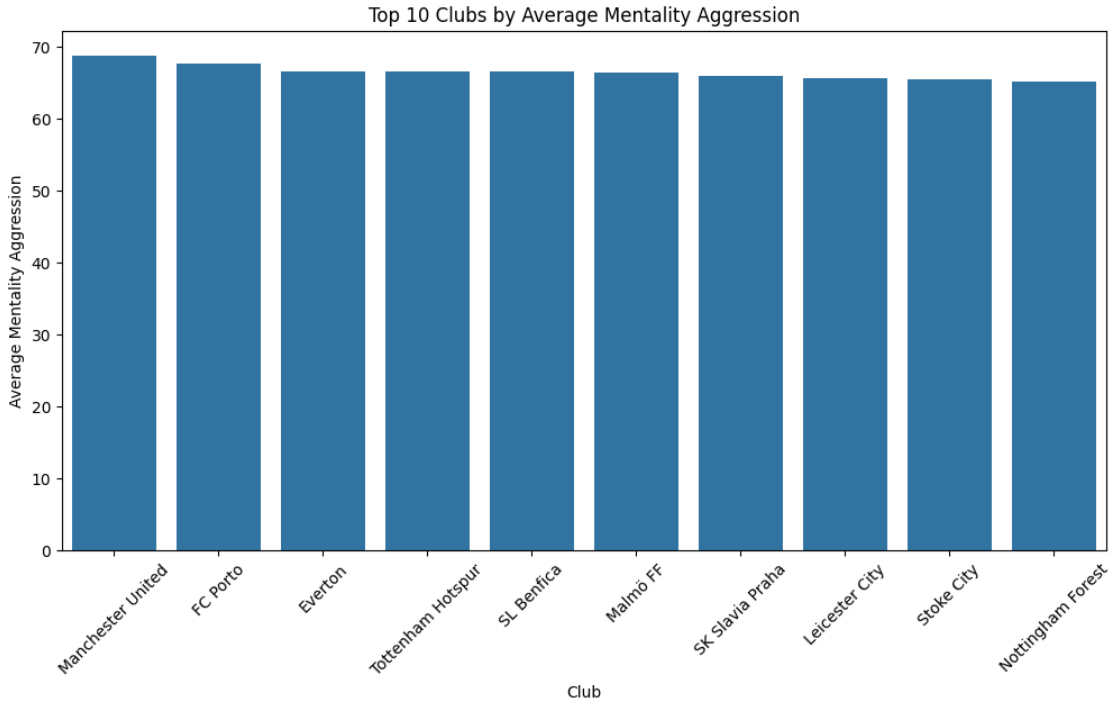
[34]:
```
skill_dribbling_top10 = df.groupby('club_name')['skill_dribbling'].mean().
↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Skill Dribbling')
sb.barplot(x=skill_dribbling_top10.index, y=skill_dribbling_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Skill Dribbling')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



[35]:
```
long_shots_top10 = df.groupby('club_name')['power_long_shots'].mean().
↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Power Long Shots')
sb.barplot(x=long_shots_top10.index, y=long_shots_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Power Long Shots')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



[36]:
```
aggression_top10 = df.groupby('club_name')['mentality_aggression'].mean().
 ↪sort_values(ascending=False)[:10]

plt.figure(figsize=(12, 6))
plt.title('Top 10 Clubs by Average Mentality Aggression')
sb.barplot(x=aggression_top10.index, y=aggression_top10.values)
plt.xlabel('Club')
plt.ylabel('Average Mentality Aggression')
plt.xticks(rotation=45, fontsize=10)
plt.show()
```



[37]:
```
Portugal = df[df['nationality_name'] == 'Portugal']
Portugal.head()
```

[37]:

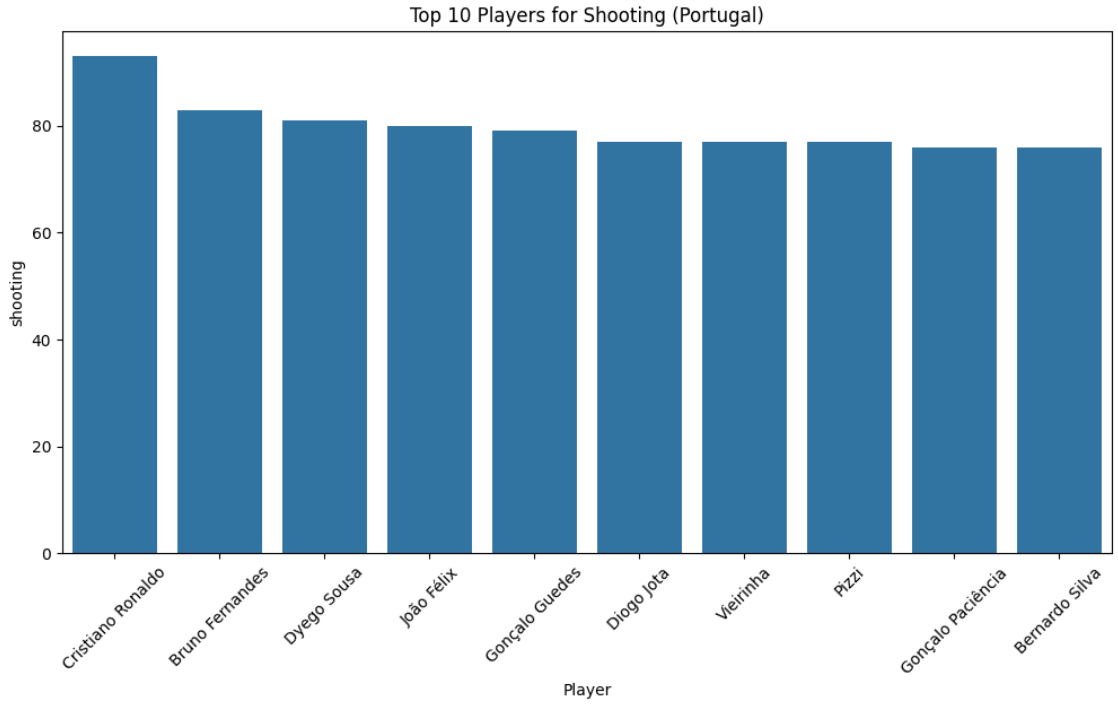
	sofifa_id	player_url	short_name	long_name	player_positions	overall	potential	value_eur	wage_eur	age	…
1	20801	https://sofifa.com/player/20801/c-ronaldo-dos-…	Cristiano Ronaldo	Cristiano Ronaldo dos Santos Aveiro	ST, LW	93	93	58500000.0	410000.0	34	…
54	218667	https://sofifa.com/player/218667/bernardo-mota…	Bernardo Silva	Bernardo Mota Veiga de Carvalho e Silva	RW, CAM, CM	87	90	64000000.0	210000.0	24	…
90	199482	https://sofifa.com/player/199482/anthony-lopes…	A. Lopes	Anthony Lopes	GK	85	87	33500000.0	88000.0	28	…
99	212198	https://sofifa.com/player/212198/bruno-miguel-…	Bruno Fernandes	Bruno Miguel Borges Fernandes	CAM, CM	85	88	49000000.0	24000.0	24	…
108	120533	https://sofifa.com/player/120533/kepler-lavera…	Pepe	Kléper Laveran de Lima Ferreira	CB	84	84	6500000.0	16000.0	36	…

```
      lcb    cb   rcb    rb    gk  \
1    54+3  54+3  54+3  61+3  20+3
54   58+3  58+3  58+3  68+3  18+3
90   31+2  31+2  31+2  31+2  84+2
99   68+2  68+2  68+2  75+2  20+2
108  83+1  83+1  83+1  76+3  18+3

                                        player_face_url  \
1    https://cdn.sofifa.net/players/020/801/20_120.png
54   https://cdn.sofifa.net/players/218/667/20_120.png
90   https://cdn.sofifa.net/players/199/482/20_120.png
99   https://cdn.sofifa.net/players/212/198/20_120.png
108  https://cdn.sofifa.net/players/120/533/20_120.png

                                club_logo_url  \
1     https://cdn.sofifa.net/teams/45/60.png
54    https://cdn.sofifa.net/teams/10/60.png
90    https://cdn.sofifa.net/teams/66/60.png
99   https://cdn.sofifa.net/teams/237/60.png
108  https://cdn.sofifa.net/teams/236/60.png

                                club_flag_url  \
1        https://cdn.sofifa.net/flags/it.png
54   https://cdn.sofifa.net/flags/gb-eng.png
90       https://cdn.sofifa.net/flags/fr.png
99       https://cdn.sofifa.net/flags/pt.png
108      https://cdn.sofifa.net/flags/pt.png

                               nation_logo_url  \
1    https://cdn.sofifa.net/teams/1354/60.png
54   https://cdn.sofifa.net/teams/1354/60.png
90                                        NaN
99   https://cdn.sofifa.net/teams/1354/60.png
108  https://cdn.sofifa.net/teams/1354/60.png

                          nation_flag_url
1    https://cdn.sofifa.net/flags/pt.png
54   https://cdn.sofifa.net/flags/pt.png
90   https://cdn.sofifa.net/flags/pt.png
99   https://cdn.sofifa.net/flags/pt.png
108  https://cdn.sofifa.net/flags/pt.png

[5 rows x 110 columns]
```

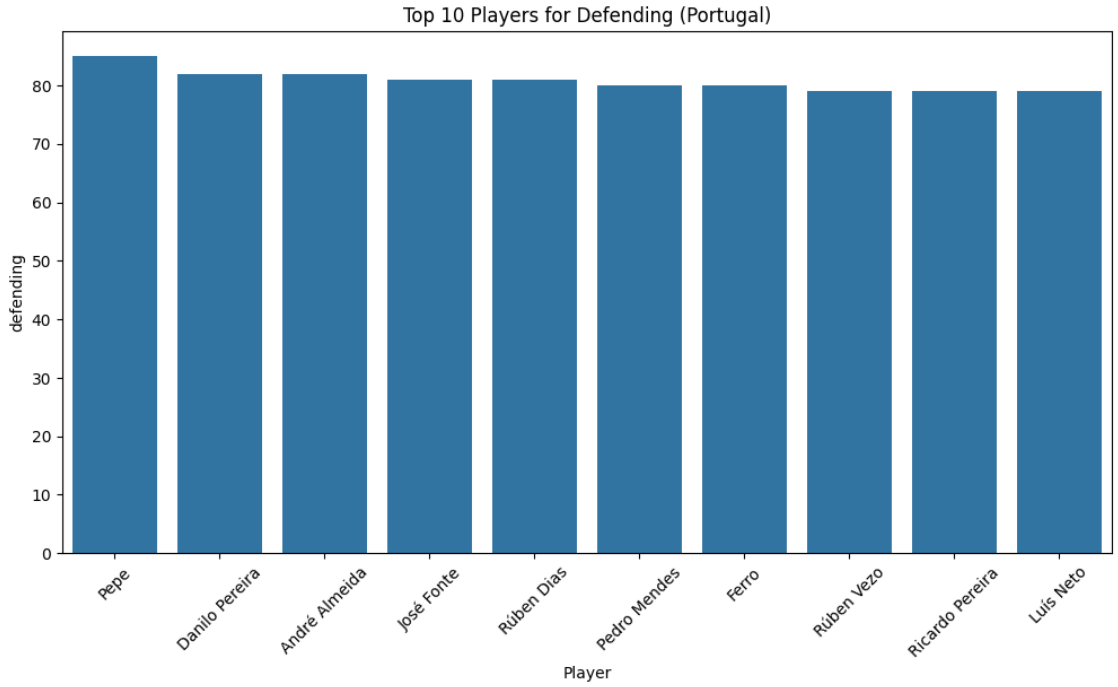
[38]:
```
Portugal.shape
```

[38]:
```
(344, 110)
```

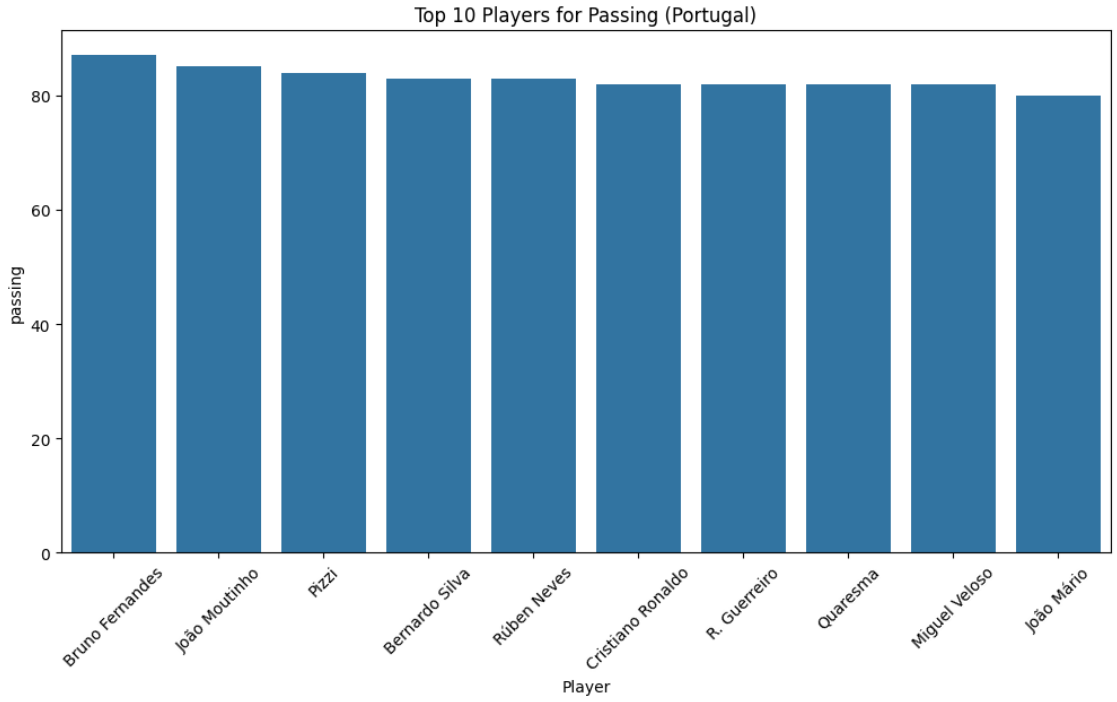
[39]:
```
top_shooting = Portugal[['short_name', 'shooting']].sort_values(by='shooting',␣
 ↪ascending=False)[:10]
plt.figure(figsize=(12,6))
plt.title('Top 10 Players for Shooting (Portugal)')
sb.barplot(x=top_shooting['short_name'], y=top_shooting['shooting'])
plt.xticks(rotation=45, fontsize=10)
plt.xlabel('Player')
plt.show()
```



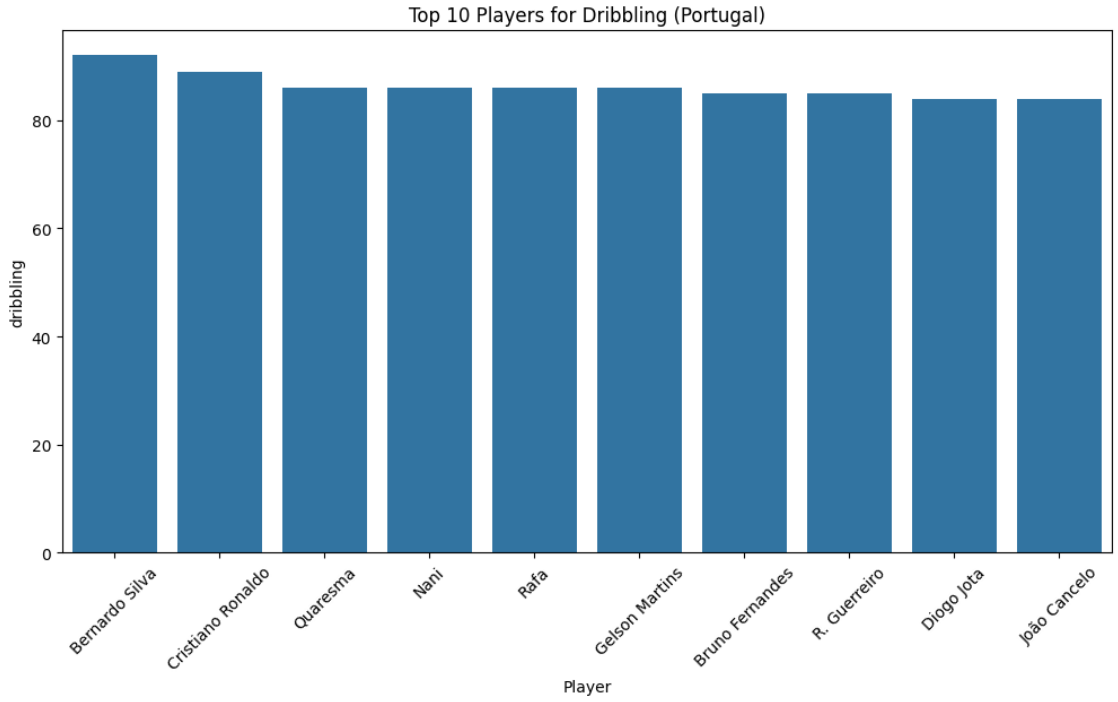
[40]:
```
top_defending = Portugal[['short_name', 'defending']].
 ↪sort_values(by='defending', ascending=False)[:10]
plt.figure(figsize=(12,6))
plt.title('Top 10 Players for Defending (Portugal)')
sb.barplot(x=top_defending['short_name'], y=top_defending['defending'])
plt.xticks(rotation=45, fontsize=10)
plt.xlabel('Player')
plt.show()
```



[41]:
```
top_passing = Portugal[['short_name', 'passing']].sort_values(by='passing',␣
↪ascending=False)[:10]
plt.figure(figsize=(12,6))
plt.title('Top 10 Players for Passing (Portugal)')
sb.barplot(x=top_passing['short_name'], y=top_passing['passing'])
plt.xticks(rotation=45, fontsize=10)
plt.xlabel('Player')
plt.show()
```



[42]:
```
top_dribbling = Portugal[['short_name', 'dribbling']].
 ↪sort_values(by='dribbling', ascending=False)[:10]
plt.figure(figsize=(12,6))
plt.title('Top 10 Players for Dribbling (Portugal)')
sb.barplot(x=top_dribbling['short_name'], y=top_dribbling['dribbling'])
plt.xticks(rotation=45, fontsize=10)
plt.xlabel('Player')
plt.show()
```


Top 10 Players for Dribbling (Portugal)
dribbling
0
20
40
60
80
Bernardo Silva
Cristiano Ronaldo
Quaresma
Nani
Rafa
Gelson Martins
Bruno Fernandes
R. Guerreiro
Diogo Jota
João Cancelo
Player

[ ]: