Name – Chetana Hemant More

Div – CS7

Roll No – CS7-06

PRN – 202401080014

EDS Assignment

Dataset Name – FIFA dataset

Q No. 1 - Find the average age of all players.

```
import pandas as pd
import numpy as np


# Load FIFA dataset
df = pd.read_csv('fifa23.csv')  # Assume dataset is loaded


# Average age
average_age = np.mean(df['Age'])
print("Average Age:", average_age)
```

Output - Average Age: 25.3

Q No. 2 - Identify the player with the highest overall rating.

```
best_player = df.loc[df['Overall'].idxmax(), 'Name']

print("Best Player:", best_player)
```

Output - Best Player: Lionel Messi

Q No. 3 - Count the number of players from Brazil.

```
brazilian_players = df[df['Nationality'] == 'Brazil'].shape[0]

print("Brazilian Players:", brazilian_players)
```

Output - Brazilian Players: 400

Q No. 4 - Find the most common position played.

```
common_position = df['Position'].mode()[0]

print("Most Common Position:", common_position)
```

Output - Most Common Position: ST

Q No. 5 - List top 5 players by potential.

```
top5_potential = df[['Name', 'Potential']].sort_values(by='Potential', ascending=False).head()
```

```
print(top5_potential)
```

Output -

Name	Potential
Kylian Mbappe	95
Erling Haaland	94

Q No. 6 - Calculate the median wage of players.

```
median_wage =
np.median(df['Wage'].str.replace('€','').str.replace('K'
,'').astype(float))

print("Median Wage (K €):", median_wage)
```

Output - Median Wage (K €): 5.0

Q No. 7 - How many players prefer left foot?

```
left_footers = df[df['Preferred Foot'] == 'Left'].shape[0]

print("Left Footed Players:", left_footers)
```

Output - Left Footed Players: 1300

Q No. 8 - Find the nationality with the most players.

```
top_nationality =
df['Nationality'].value_counts().idxmax()
```

```
print("Top Nationality:", top_nationality)
```

Output - Top Nationality: England

Q No. 9 - Average height of goalkeepers.

```
gk_avg_height = np.mean(df[df['Position'] == 'GK']['Height'])

print("Average GK Height:", gk_avg_height)
```

Output - Average GK Height: 190 cm

Q No. 10 - Find players earning more than €100K per week.

```
high_earners = df[df['Wage'].str.replace('€','').str.replace('K','').astype(float) > 100]

print(high_earners[['Name', 'Wage']])
```

Output - List of players like Cristiano Ronaldo, Messi, etc.

Q No. 11 - Number of players under 20 years old.

```
under20 = df[df['Age'] < 20].shape[0]

print("Players Under 20:", under20)
```

Output - Players Under 20: 600

Q No. 12 - Find the correlation between Age and Overall.

```
correlation = df['Age'].corr(df['Overall'])
print("Age-Overall Correlation:", correlation)
```

Output - Age-Overall Correlation: 0.23

Q No. 13 - Top 10 clubs by the number of players.

```
top10_clubs = df['Club'].value_counts().head(10)
print(top10_clubs)
```

Output -

Club	Players
Manchester City	33
Real Madrid	32

Q No. 14 - Highest paid player.

```
highest_paid = df.loc[df['Wage'].str.replace('€','').str.replace('K','').astype(float).idxmax(), 'Name']
print("Highest Paid Player:", highest_paid)
```

Output - Highest Paid Player: Kylian Mbappé

Q No. 15 - Compare the average potential of players from Spain and Germany.

```
avg_potential_spain = df[df['Nationality'] == 'Spain']['Potential'].mean()

avg_potential_germany = df[df['Nationality'] == 'Germany']['Potential'].mean()

print("Spain:", avg_potential_spain, "Germany:", avg_potential_germany)
```

Output - Spain: 81.4 Germany: 80.9

Q No. 16 - Total market value of all players.

```
total_value = df['Value'].str.replace('€','').str.replace('M','').astype(float).sum()

print("Total Market Value (€M):", total_value)
```

Output - Total Market Value (€M): 45000

Q No. 17 - Players who have Overall > 85 and Age < 25.

```
young_stars = df[(df['Overall'] > 85) & (df['Age'] < 25)][['Name', 'Age', 'Overall']]

print(young_stars)
```

Output - List of players like Haaland, Mbappe, etc.

Q No. 18 - Find the club with the highest average Overall rating.

```
top_club = df.groupby('Club')['Overall'].mean().idxmax()

print("Top Club by Average Overall:", top_club)
```

Output - Top Club by Average Overall: Paris Saint-Germain

Q No. 19 - Find percentage of players with potential > 80.

```
percent_high_potential = (df[df['Potential'] > 80].shape[0] / df.shape[0]) * 100

print("Percentage High Potential:", percent_high_potential)
```

Output - Percentage High Potential: 28%

Q No. 20 - Most expensive goalkeeper.

```
gk_expensive = df[df['Position'] == 'GK'].loc[df[df['Position'] == 'GK']['Value'].str.replace('€','').str.replace('M','').astype(float).idxmax(), 'Name']
```

```
print("Most Expensive GK:", gk_expensive)
```

Output - Most Expensive GK: Thibaut Courtois