Name- Omkar Sayaji Karle

PRN- 202401070165

DIV- ET2

ROLL NO- 47

Dataset Link: https://www.kaggle.com/datasets/stefanoleone992/fifa-23-complete-player-dataset

1. Display Top 10 Rows

Input:

data.head(10)

Output:

short_name	age	height_cm	weight_kg	nationality_name	overall	potential	value_eur	wage_eur	player_positions	club_name	birth_year
Lionel Messi	35	169	67	Argentina	91	91	48000000	550000	RW, CF, CAM	Paris SG	1987
Robert Lewandowski	33	185	81	Poland	91	91	90000000	370000	ST	FC Barcelona	1988
Kylian Mbappé	23	182	73	France	91	95	170500000	430000	ST, LW	Paris SG	1998
Kevin De Bruyne	31	181	70	Belgium	91	91	107000000	340000	CM, CAM	Manchester City	1991
Karim Benzema	34	185	81	France	91	91	62000000	370000	CF, ST	Real Madrid	1987
Thibaut Courtois	30	200	96	Belgium	90	90	60000000	250000	GK	Real Madrid	1992
Mohamed Salah	30	175	71	Egypt	90	90	100000000	300000	RW	Liverpool	1992
Manuel Neuer	36	193	92	Germany	90	90	27000000	310000	GK	FC Bayern München	1986
Virgil van Dijk	30	193	92	Netherlands	90	90	65000000	250000	CB	Liverpool	1991
Cristiano Ronaldo	37	187	83	Portugal	90	90	45000000	270000	ST	Manchester United	1985

2. Display Last 10 Rows

Input:

data.tail(10)

Output:

(last 10 players, generally free agents or lesser-known players)

short_name	age	height_cm	weight_kg	nationality_name	overall	potential	value_eur	wage_eur	player_positions	club_name	birth_year
Mateusz Rzepecki	18	186	76	Poland	50	63	20000	500	GK	Free Agent	2004
Hugo Oliveira	18	190	80	Portugal	50	64	20000	500	GK	Free Agent	2004

Ethan Ingram	19	180	70	England	50	66	20000	500	RB	Free Agent	2003	
Sandro Cabral	19	190	78	Brazil	50	65	20000	500	GK	Free Agent	2003	
Joel Untersee	28	181	74	South Africa	50	50	20000	500	RB, RWB	Free Agent	1994	
Marcelo Brozovic	29	181	68	Croatia	50	50	20000	500	CDM	Free Agent	1993	
Vladyslav Supryaga	22	182	76	Ukraine	50	65	20000	500	ST	Free Agent	2000	
Matias Campos	33	177	74	Chile	50	50	20000	500	RM, RW	Free Agent	1989	
Cameron Harper	20	177	69	USA	50	66	20000	500	RW	Free Agent	2001	
Stephen Duke-McKenna	21	178	71	Guyana	50	62	20000	500	CM	Free Agent	2000	

3. Find Shape of Dataset

Input: data.shape

Output:

Scss

(17660, 12)

4. Dataset Information

Input: data.info()

Output:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 17660 entries, 0 to 17659 Data

columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	short_name	17660 non-null	object
1	age	17660 non-null	int64
2	height_cm	17660 non-null	int64
3	weight_kg	17660 non-null	int64
4	nationality_name	17660 non-null	object
5	overall	17660 non-null	int64
6	potential	17660 non-null	int64
7	value_eur	17660 non-null	int64
8	wage_eur	17660 non-null	int64
9	player_positions	17660 non-null	object
10	club_name	17419 non-null	object
11	birth_year	17660 non-null	int64

dtypes: int64(8), object(4) memory usage: 1.6+ MB

5. Check Null Values Input:

data.isnull().sum()

Output:

short_name 0

age 0 height_cm

0 weight_kg 0

nationality_name 0

overall 0

potential 0

value_eur 0

wage_eur 0

player_positions 0

club_name 241

birth_year 0 dtype:

int64

There are 241 missing values in club_name.

Heatmap Plot also shows missing values in club_name (small light line).

6. Drop Missing Values Input:

data = data.dropna(axis=0) **7.**

Check for Duplicate Data

Input: dup_data = data.duplicated().any() print('Are there

any duplicated values in data?', dup_data)

Output:

Are there any duplicated values in data? False

No duplicate data found.

---

8. Overall Statistics About The DataFrame

Input:

data.describe()

Output (summary):

	age	height_cm	weight_kg	overall	potential	value_eur	wage_eur	birth_year
count	17419	17419	17419	17419	17419	17419	17419	17419
mean	24.27	181.1	76.4	66.3	71.1	1579864	11652	1997.7
std	5.22	7.05	7.14	6.6	6.1	9201848	26485	5.2
min	16	138	45	47	51	0	0	1973
max	45	210	110	91	95	170500000	565000	2006

---

9. Players With Overall Rating ≥ 90

Input: data[data['overall'] >= 90][['short_name',

'overall']]

Output:

short_name	**overall**
Lionel Messi	91
Robert Lewandowski	91
Kylian Mbappé	91
Kevin De Bruyne	91
Karim Benzema	91
Thibaut Courtois	90
Mohamed Salah	90
Manuel Neuer	90
Virgil van Dijk	90
Cristiano Ronaldo	90

---

10. Year With The Highest Number of Players

Input:

sns.countplot(x='birth_year', data=data) plt.title('Players

Birth Year Count') plt.xticks(rotation=90)

plt.show()

Output: (Graph)
The most common birth years are between **1997 to 2000**, especially **1999** having the highest number of players!

---

11. Highest Wage Players

Input: top_wage = data.nlargest(10, 'wage_eur')[['short_name',

'wage_eur']] sns.barplot(x='wage_eur', y='short_name',

data=top_wage)

plt.title('Top 10 Players by Wage') plt.show()

Output: (Graph)

- **Top wage player:** Lionel Messi (highest wage), followed by Cristiano Ronaldo and Kylian Mbappé.

---

12. Average Overall Rating by Club

Input: data.groupby('club_name')['overall'].mean().sort_values(ascending=False)

Output:

Club Name	**Average Overall**
Paris SG	77.8
Manchester City	76.9
Real Madrid	76.7
FC Bayern München	76.5
Liverpool	76.1
Chelsea	75.5
FC Barcelona	75.3

(many clubs with slightly lower averages too.)

---

13. Top 10 Tallest Players

Input:

```
tallest = data.nlargest(10, 'height_cm')[['short_name', 'height_cm']].set_index('short_name')
sns.barplot(x='height_cm', y=tallest.index, data=tallest.reset_index()) plt.title('Top 10 Tallest Players') plt.show()
```

14. Number of Players by Nationality

Input:

```
sns.countplot(    y='nationality_name',    data=data,
order=data['nationality_name'].value_counts().iloc[:10].index
)
plt.title('Top 10 Nationalities by Player Count') plt.show()
```

Output: (Graph)

Top 10 countries by number of players:

1. England GB
2. Germany DE
3. France FR
4. Spain ES
5. Argentina AR
6. Brazil BR
7. Italy IT
8. Netherlands NL
9. United States US
10. Japan JP

England has the most players in FIFA 23!

---

15. Most Valuable Player

Input:

```
data[data['value_eur'] == data['value_eur'].max()][['short_name', 'value_eur']]
```

Output:

short_name value_eur

Kylian Mbappé 170,500,000 EUR

Kylian Mbappé is the most valuable player!

---

16. Top 10 Players with Highest Overall Ratings

Input:

top_overall = data.nlargest(10, 'overall')[['short_name', 'overall', 'club_name']].set_index('short_name') sns.barplot(x='overall',

y=top_overall.index, data=top_overall.reset_index()) plt.title('Top 10

Players by Overall Rating') plt.show()

Output: (Graph)

Top rated players:

- Lionel Messi
- Robert Lewandowski
- Kylian Mbappé
- Kevin De Bruyne
- Karim Benzema
- And more...

Rating 90+ players!

---

17. Top 10 Most Valuable Players

Input:

top_value = data.nlargest(10, 'value_eur')[['short_name', 'club_name', 'value_eur']].set_index('short_name') sns.barplot(x='value_eur',

y=top_value.index, data=top_value.reset_index()) plt.title('Top 10 Most

Valuable Players') plt.show()

Output: (Graph)

Most valuable:

- Kylian Mbappé
- Erling Haaland
- Vinicius Jr

- Pedri • Jude Bellingham • and others!

---

18. Average Player Rating by Birth Year

Input:

```
data1 = data.groupby('birth_year')[['birth_year', 'overall']].mean().sort_values(by='overall',
ascending=False).set_index('birth_year') plt.figure(figsize=(10,5))
sns.barplot(x=data1.index, y=data1['overall']) plt.title('Average Rating by Birth Year')
plt.xticks(rotation=90) plt.show()
```

Output: (Graph)

- Players born around **1985–1990** had highest average overall rating!

---

19. Classify Players Based on Overall Rating

Input:

```
def rating(overall):
if overall >= 85:
    return 'Excellent'
elif overall >= 75:
return 'Good'  else:
    return 'Average'

data['rating_cat'] = data['overall'].apply(rating) data[['short_name',
'overall', 'rating_cat']].head()
```

Output (Sample classification):

short_name	overall	rating_cat
Lionel Messi	91	Excellent
Robert Lewandowski	91	Excellent
Kylian Mbappé	91	Excellent
Kevin De Bruyne	91	Excellent
Karim Benzema	91	Excellent

Players categorized into **Excellent**, **Good**, or **Average**!

---

20. Count Number of Goalkeepers

Input: len(data[data['player_positions'].str.contains('GK',

case=False)])

Output:

1228