**Name and Dataset Link**

**Name:** Raviraj kawle

# Roll No: ET2-92

**PRN:** 202401070195

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

**1. Display Top 10 Rows of The Dataset** data.head(10)

# 2. Display Last 10 Rows of The Dataset data.tail(10)

**3. Find Shape of Dataset** data.shape print('Number of Rows:', data.shape[0]) print('Number of Columns:', data.shape[1])

# 4. Dataset Information

data.info()

1. **Check Null Values** data.isnull().sum() sns.heatmap(data.isnull()) plt.show()
2. **Drop Missing Values** data = data.dropna(axis=0) sns.heatmap(data.isnull()) plt.show()

# 7. Check For Duplicate Data

dup\_data = data.duplicated().any() print('Are there any duplicated values in data?', dup\_data) **8. Overall Statistics About The DataFrame** data.describe()

# 9. Players With Overall Rating >= 90 data[data['overall'] >= 90][['short\_name', 'overall']] 10. Year With The Highest Number of Players

sns.countplot(x='birth\_year', data=data) plt.title('Players Birth Year Count') plt.xticks(rotation=90) plt.show()

# 11. Highest Wage Players

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()

**12. Average Overall Rating by Club** data.groupby('club\_name')['overall'].mean().sort\_values(ascending=False)

# 13. Top 10 Tallest Players

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

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()

**15. Most Valuable Player** data[data['value\_eur'] == data['value\_eur'].max()][['short\_name', 'value\_eur']]

# 16. Top 10 Players with Highest Overall Ratings

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()

# 17. Top 10 Most Valuable Players

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()

# 18. Average Player Rating by Birth Year

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()

# 19. Classify Players Based on Overall Rating

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()

**20. Count Number of Goalkeepers** len(data[data['player\_positions'].str.contains('GK', case=False)])