Recommended Dataset: FIFA 23 Players Dataset

DataSet is the FIFA 23 Players Dataset (available on Kaggle). It contains player attributes like age, nationality, club, overall rating, wage, position, etc.

1. Data Loading & Inspection

1. Load the FIFA 23 dataset (players_fifa23.csv) into a Pandas DataFrame.
2. Display the first 10 rows and last 5 rows.
3. Check the dimensions (shape) of the DataFrame.
4. Get the summary statistics (mean, min, max, etc.) for numeric columns.
5. List all column names and their data types.

2. Data Cleaning & Handling Missing Values

6. How many missing values are in each column?
7. Drop columns with more than 50% missing values.
8. Fill missing Club values with "Free Agent".
9. Replace missing Height values with the median height.
10. Remove duplicate players based on ID.

3. Data Filtering & Selection

11. Select players with an Overall rating greater than 85.
12. Filter players who are from Argentina and play as Forward (FW).
13. Find players aged between 18 and 21 with Potential > 85.
14. Select only Name, Age, Club, and Overall for the top 20 players.
15. Get players who earn more than €200K in wage.

4. Data Aggregation & Grouping

16. Calculate the average Overall rating by Nationality.
17. Find the highest-paid player (Wage) in each Club.
18. Compute the average Age of players by Position.
19. Group by Club and get the median Overall rating.
20. Find the youngest player in each Nationality.

5. Advanced Operations

21. Create a new column BMI using Weight and Height.

22.Apply a function to categorize players into Low (<70), Medium (70-80), High (>80) based on Overall.

23.Use pd.cut() to bin players into 5 equal-sized groups based on Wage.

24.Calculate the average Potential by Age Group (e.g., U20, 20-30, 30+).

25.Use pivot_table() to show average Overall by Position and Preferred Foot.

6. Merging & Reshaping Data

26.Split the Name column into First Name and Last Name.

27.Create a new DataFrame with only goalkeepers and merge it back to the original.

28.Reshape the data to show counts of players by Nationality and Club (use pd.crosstab()).

29.Stack and unstack a multi-index DataFrame (e.g., Club vs Position counts).

30.Concatenate two subsets of the dataset (e.g., players from England and Spain).

7. Visualization (Bonus with Matplotlib/Seaborn)

31.Plot a histogram of player Overall ratings.

32.Create a bar plot of the top 10 nationalities by player count.

33.Generate a box plot of Wage by Position.

34.Plot a scatter plot of Age vs Potential colored by Overall.

35.Visualize the correlation matrix between Overall, Potential, and Wage.