Name – Harshal Manohar Mekhe

PRN – 202401040051

Division – CS2

Roll no – CS2-05

Dataset link-

https://www.kaggle.com/datasets/bryanb/fifa-player-stats-database?resource=download

Google Colab link-

https://colab.research.google.com/drive/1F5SH0r0zzDIii55V2da091eqFbLwkmJB?usp=sharing

Untitled1.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text

RAM Disk

```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files FIFA23_official_data.csv
• FIFA23_official_data.csv(text/csv) - 5885314 bytes, last modified: 10/25/2022 - 100% done
Saving FIFA23_official_data.csv to FIFA23_official_data.csv
```

```
[2] import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt

    # Now the file is in the current directory
    df = pd.read_csv('FIFA23_official_data.csv')

    # Continue your code
    average_overall = np.mean(df['Overall'])
    print("Average Overall Rating:", average_overall)
```

```
Average Overall Rating: 63.36959229898075
```

```
[3] median_age = np.median(df['Age'])
    print("Median Age:", median_age)
```

```
Median Age: 22.0
```

```
[10] std_potential = np.std(df['Potential'])
     print("Std Dev of Potential:", std_potential)
```

```
Std Dev of Potential: 6.5296508075673465
```

```
[11] min_height = np.min(df['Height'])
     max_height = np.max(df['Height'])
     print("Min Height:", min_height)
     print("Max Height:", max_height)
```

```
Min Height: 149cm
Max Height: 206cm
```

```
[12] weak_foot_count = np.sum(df['Weak Foot'] > 3)
     print("Players with Weak Foot > 3:", weak_foot_count)
```

```
Players with Weak Foot > 3: 2397
```

```
[13] df['Wage(EUR)'] = pd.to_numeric(df['Wage'].str.replace('€', '').str.replace('K', 'e3').str.replace('M', 'e6'), errors='coerce')
     average_wage = np.nanmean(df['Wage(EUR)'])
     print("Average Wage (EUR):", average_wage)
```

```
Average Wage (EUR): 8189.807474518686
```

```
[16] players_20_30 = np.sum((df['Age'] >= 20) & (df['Age'] <= 30))
     print("Players aged between 20 and 30:", players_20_30)
```

```
Players aged between 20 and 30: 11898
```

```
[18] skill_moves = df['Skill Moves'] > 4
     print(skill_moves)
```

```
0        False
1        False
2        False
3        False
4        False
         ...
17655    False
17656    False
17657    False
17658    False
17659    False
Name: Skill Moves, Length: 17660, dtype: bool
```

```
[22] min_potential = np.min(df['Potential'])
     print("Minimum Potential:", min_potential)
```

```
Minimum Potential: 42
```

```
[24] max_overall = np.max(df['Overall'])
```

```
print("Maximum Overall Rating:", max_overall)
```

```
Maximum Overall Rating: 91
```

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

```
Average Age: 23.127746319365798
```

```
[26] # 1. Display first 5 rows
print(df.head())
```

```
       ID             Name  Age  \
0  209658      L. Goretzka   27
1  212198  Bruno Fernandes   27
2  224334         M. Acuña   30
3  192985     K. De Bruyne   31
4  224232       N. Barella   25

                                              Photo Nationality  \
0  https://cdn.sofifa.net/players/209/658/23_60.png     Germany
1  https://cdn.sofifa.net/players/212/198/23_60.png    Portugal
2  https://cdn.sofifa.net/players/224/334/23_60.png   Argentina
3  https://cdn.sofifa.net/players/192/985/23_60.png     Belgium
4  https://cdn.sofifa.net/players/224/232/23_60.png       Italy

                                  Flag  Overall  Potential              Club  \
0  https://cdn.sofifa.net/flags/de.png       87         88  FC Bayern München
1  https://cdn.sofifa.net/flags/pt.png       86         87  Manchester United
2  https://cdn.sofifa.net/flags/ar.png       85         85         Sevilla FC
3  https://cdn.sofifa.net/flags/be.png       91         91    Manchester City
4  https://cdn.sofifa.net/flags/it.png       86         89              Inter

                                Club Logo  ...                     Position  \
0   https://cdn.sofifa.net/teams/21/30.png  ...  <span class="pos pos28">SUB
1   https://cdn.sofifa.net/teams/11/30.png  ...  <span class="pos pos15">LCM
2  https://cdn.sofifa.net/teams/481/30.png  ...   <span class="pos pos7">LB
3   https://cdn.sofifa.net/teams/10/30.png  ...  <span class="pos pos13">RCM
4   https://cdn.sofifa.net/teams/44/30.png  ...  <span class="pos pos13">RCM

          Joined  Loaned From Contract Valid Until  Height  Weight  \
0   Jul 1, 2018          NaN                2026   189cm    82kg
1  Jan 30, 2020          NaN                2026   179cm    69kg
2  Sep 14, 2020          NaN                2024   172cm    69kg
3  Aug 30, 2015          NaN                2025   181cm    70kg
4   Sep 1, 2020          NaN                2026   172cm    68kg

   Release Clause Kit Number Best Overall Rating Wage(EUR)
0          €157M        8.0                  NaN  115000.0
1          €155M        8.0                  NaN  190000.0
2         €97.7M       19.0                  NaN   46000.0
3        €198.9M       17.0                  NaN  350000.0
4        €154.4M       23.0                  NaN  110000.0

[5 rows x 30 columns]
```

```
[27] # 2. List of unique Nationalities
print(df['Nationality'].unique())
```

```
['Germany' 'Portugal' 'Argentina' 'Belgium' 'Italy' 'Austria' 'Brazil'
 'Croatia' 'Serbia' 'Spain' 'Netherlands' 'France' 'Colombia' 'England'
 'Uruguay' 'Morocco' 'Egypt' 'Algeria' 'Ukraine' 'United States'
 "Côte d'Ivoire" 'Poland' 'Chile' 'Senegal' 'Central African Republic'
 'Denmark' 'Nigeria' 'Mexico' 'Turkey' 'Canada' 'Wales' 'Scotland'
 'Romania' 'Czech Republic' 'Ghana' 'Korea Republic'
 'Bosnia and Herzegovina' 'Mali' 'Slovakia' 'Armenia' 'Norway'
 'Switzerland' 'Cameroon' 'Peru' 'Jamaica' 'Zambia' 'Guinea' 'Sweden'
 'North Macedonia' 'Russia' 'Tunisia' 'Malta' 'Angola'
 'Republic of Ireland' 'Ecuador' 'Benin' 'Paraguay' 'Montenegro'
 'Australia' 'Comoros' 'Gabon' 'Iceland' 'Slovenia' 'Japan' 'Israel'
 'China PR' 'Venezuela' 'Liberia' 'Greece' 'Bulgaria' 'Honduras'
 'Saudi Arabia' 'Curacao' 'Northern Ireland' 'Guinea Bissau' 'Kosovo'
 'Hungary' 'Finland' 'Costa Rica' 'Albania' 'Congo DR' 'Iran' 'Mozambique'
 'Suriname' 'Cape Verde Islands' 'Bolivia' 'Madagascar' 'New Zealand'
 'Burkina Faso' 'Dominican Republic' 'Kazakhstan' 'Syria' 'Luxembourg'
 'Kenya' 'Zimbabwe' 'Haiti' 'Uzbekistan' 'South Africa' 'Cyprus' 'Qatar'
 'Equatorial Guinea' 'Libya' 'Thailand' 'Togo' 'Trinidad and Tobago'
 'Liechtenstein' 'Gambia' 'Georgia' 'Philippines' 'Burundi'
 'United Arab Emirates' 'Grenada' 'Iraq' 'Panama' 'Malaysia' 'Moldova'
 'Congo' 'India' 'Jordan' 'Kuwait' 'Antigua and Barbuda' 'Cuba' 'Vietnam'
 'Korea DPR' 'Uganda' 'Lithuania' 'Estonia' 'Montserrat' 'Sierra Leone'
 'Afghanistan' 'New Caledonia' 'Belarus' 'Laos' 'Saint Lucia' 'Bhutan'
 'Guyana' 'Mauritania' 'Faroe Islands' 'Namibia' 'Niger' 'Palestine'
 'Sudan' 'Azerbaijan' 'Hong Kong' 'Gibraltar' 'Tanzania' 'Latvia'
 'Chinese Taipei' 'Singapore' 'Lebanon' 'El Salvador' 'Indonesia'
 'Guatemala' 'Papua New Guinea' 'Puerto Rico' 'Malawi' 'South Sudan'
 'Ethiopia' 'San Marino' 'Andorra' 'Saint Kitts and Nevis']
```

```
[28] brazil_players = df[df['Nationality'] == 'Brazil']
print("Players from Brazil:", len(brazil_players))
```

```
Players from Brazil: 539
```

```
[29] best_player = df.loc[df['Overall'].idxmax()]
print("Best Player:", best_player['Name'])
```

```
Best Player: K. De Bruyne
```

```
[30] sorted_by_wage = df.sort_values('Wage(EUR)', ascending=False)
     print(sorted_by_wage[['Name', 'Wage']].head())
```

```
              Name   Wage
124      K. Benzema  €450K
41   R. Lewandowski  €420K
3      K. De Bruyne  €350K
125         T. Kroos  €310K
37          Casemiro  €300K
```

```
[31] # 6. Replace missing values in Club with "Unknown"
     df['Club'].fillna('Unknown', inplace=True)
     print(df['Club'].head())
```

```
0    FC Bayern München
1    Manchester United
2           Sevilla FC
3      Manchester City
4                Inter
Name: Club, dtype: object
<ipython-input-31-cafe7e67056c>:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to per


  df['Club'].fillna('Unknown', inplace=True)
```

```
[32] average_by_club = df.groupby('Club')['Overall'].mean()
     print(average_by_club)
```

```
Club
1. FC Heidenheim 1846        63.333333
1. FC Kaiserslautern         61.894737
1. FC Köln                   69.043478
1. FC Magdeburg              62.173913
1. FC Nürnberg               63.235294
                                ...
İstanbul Başakşehir FK       68.500000
İstanbulspor                 60.071429
İttifak Holding Konyaspor    64.214286
ŁKS Łódź                     56.000000
Śląsk Wrocław                59.750000
Name: Overall, Length: 927, dtype: float64
```

```
[33] foot_count = df['Preferred Foot'].value_counts()
     print(foot_count)
```

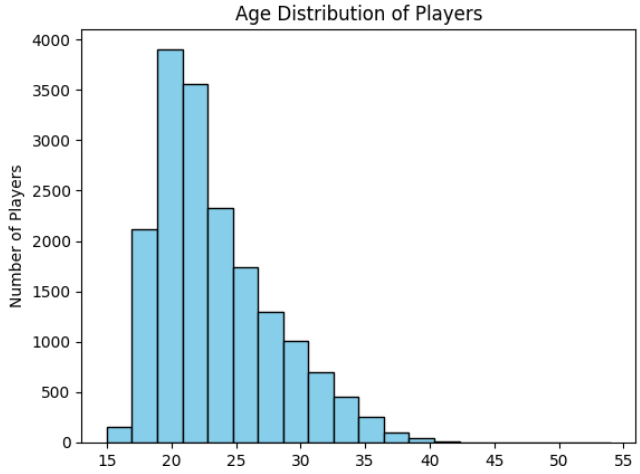
```
Preferred Foot
Right    13651
Left      4009
Name: count, dtype: int64
```

```
[34] top5_potential = df.sort_values('Potential', ascending=False).head(5)
     print(top5_potential[['Name', 'Potential']])
```

```
             Name  Potential
75      K. Mbappé         95
180    E. Haaland         94
214         Pedri         93
317      P. Foden         92
849  Vinícius Jr.         92
```

```
import matplotlib.pyplot as plt

df['Age'].plot(kind='hist', bins=20, color='skyblue', edgecolor='black')
plt.title('Age Distribution of Players')
plt.xlabel('Age')
plt.ylabel('Number of Players')
plt.show()
```



Age

✓ 0s completed at 11:20 PM