Name:-Tejas sawate

PRN:-2024010700074

Batch:-ET2

Roll no.:-ET2-06

20 . Top 5 countries by total home goals scored

```
[53] #20. Top 5 countries by total home goals scored
     print("20. Top 5 Home Team Countries (by goals):")
     print(df.groupby('Home Team Name')['Home Team Goals'].sum().sort_values(ascending=False).head(5))
```

```
20. Top 5 Home Team Countries (by goals):
Home Team Name
Brazil        180
Argentina     111
Italy          99
Germany FR     99
Hungary        73
Name: Home Team Goals, dtype: int64
```

15. Which two teams played against each other the most times?

```
[59] # 15. Which two teams played against each other the most times?
     most_common_match = df.groupby(['Home Team Name', 'Away Team Name']).size().sort_values(ascending=False).head(1)
     print("15. Teams who faced each other the most:\n", most_common_match)
```

```
15. Teams who faced each other the most:
 Home Team Name  Away Team Name
Netherlands     Argentina         5
dtype: int64
```

16. Find which city hosted matches for the maximum number of different years.

```
[60] # 16. Find which city hosted matches for the maximum number of different years.
     city_years = df.groupby('City')['Year'].nunique()
     print("16. City hosting in most different years:", city_years.idxmax())
```

```
16. City hosting in most different years: Belo Horizonte
```

17. Find the stadium that hosted the most matches.

```
# 17. Find the stadium that hosted the most matches.
top_stadium = df['Stadium'].value_counts().idxmax()
print("17. Stadium with most matches hosted:", top_stadium)
```

```
17. Stadium with most matches hosted: Estadio Azteca
```

18. Average attendance across all matches

```
[56] # 18. Average attendance across all matches
     print("18. Average Attendance:", df['Attendance'].mean())
```

```
18. Average Attendance: 45164.8
```

19. Number of matches played in each stadium

```
[55] # 19. Number of matches played in each stadium
     print("19. Matches per Stadium:")
     print(df['Stadium'].value_counts())
```

```
19. Matches per Stadium:
Stadium
Estadio Azteca              19
Jalisco                     14
Olympiastadion              14
Nou Camp - Estadio Le⊡n     11
Estadio Centenario          10
                            ..
Comunale di Cornaredo        1
Ilha do Retiro               1
Cavee Verte                  1
Friuli                       1
White City                   1
Name: count, Length: 181, dtype: int64
```

11. Find the most common match-up (team vs team) in World Cup history.

```
[24] # Solution 19
matchups = df.groupby(['Home Team Name', 'Away Team Name']).size().reset_index(name='Counts')
most_common_matchup = matchups.loc[matchups['Counts'].idxmax()]
print("Most common matchup:\n", most_common_matchup)
```

```
Most common matchup:
 Home Team Name    Brazil
Away Team Name     Chile
Counts                 5
Name: 81, dtype: object
```

12. Identify matches where attendance was missing (null values).

```
[28] # Solution 23
missing_attendance_matches = df[df['Attendance'].isnull()]
print("Matches with missing attendance:\n", missing_attendance_matches)
```

```
Matches with missing attendance:
      Year              Datetime            Stadium           City  \
823   2014  30 Jun 2014 - 17:00   Estadio Beira-Rio  Porto Alegre
841   2014  30 Jun 2014 - 17:00   Estadio Beira-Rio  Porto Alegre

    Home Team Name  Home Team Goals Away Team Name  Attendance  \
823        Germany                2        Algeria         NaN
841        Germany                2        Algeria         NaN

     Half-time Home Goals             Referee                Assistant 1
823                     0  RICCI Sandro (BRA)  DE CARVALHO Emerson (BRA)
841                     0  RICCI Sandro (BRA)  DE CARVALHO Emerson (BRA)
```

13. Team with most home goals

```
[34] print("13.", df.groupby('Home Team Name')['Home Team Goals'].sum().idxmax())
```

```
13. Brazil
```

14. Find the match with the lowest attendance.

```
[58] # 14. Find the match with the lowest attendance.
lowest_attendance_match = df.loc[df['Attendance'].idxmin()]
print("14. Match with lowest attendance:\n", lowest_attendance_match)
```

```
14. Match with lowest attendance:
 Year                                  1930
Datetime                19 Jul 1930 - 12:50
Stadium                   Estadio Centenario
City                              Montevideo
Home Team Name                         Chile
Home Team Goals                            1
Away Team Name                        France
Attendance                            2000.0
Half-time Home Goals                       0
Referee                  TEJADA Anibal (URU)
Assistant 1            LOMBARDI Domingo (URU)
Name: 9, dtype: object
```

7. Identify the team that scored the most goals as a home team across all matches.

```
# Solution 7
home_team_goals = df.groupby('Home Team Name')['Home Team Goals'].sum()
team_most_home_goals = home_team_goals.idxmax()
print("Team with most Home goals:", team_most_home_goals)
```

```
Team with most Home goals: Brazil
```

8. Find the match with the largest goal difference.

```
[32] # Solution 8
home_5_goals = df[df['Home Team Goals'] >= 5]
print("Matches where home team scored 5+ goals:\n", home_5_goals[['Year', 'Home Team Name', 'Home Team Goals', 'Away Team Name']])
```

```
Matches where home team scored 5+ goals:
      Year  Home Team Name  Home Team Goals          Away Team Name
10    1930       Argentina                6                  Mexico
15    1930       Argentina                6                     USA
16    1930         Uruguay                6                Yugoslavia
22    1934         Germany                5                 Belgium
24    1934           Italy                7                     USA
76    1938         Hungary                6          Dutch East Indies
40    1938          Brazil                6                  Poland
66    1938          Sweden                8                    Cuba
49    1938         Hungary                5                  Sweden
66    1950         Uruguay                8                 Bolivia
68    1950           Chile                5                     USA
70    1950          Brazil                7                  Sweden
71    1950          Brazil                6                   Spain
77    1954          Brazil                5                  Mexico
80    1954         Hungary                9             Korea Republic
83    1954         Uruguay                7                Scotland
84    1954         Austria                5            Czechoslovakia
87    1954         Hungary                8             Germany FR
88    1954          Turkey                7             Korea Republic
91    1954     Germany FR                 7                  Turkey
94    1954         Austria                7             Switzerland
97    1954     Germany FR                 6                 Austria
105   1958         France                 7                Paraguay
124   1958  Czechoslovakia                6               Argentina
133   1958          Brazil                5                  France
134   1958         France                 6             Germany FR
135   1958          Brazil                5                  Sweden
150   1962         Hungary                6                Bulgaria
156   1962      Yugoslavia                5                Colombia
169   1966     Germany FR                 5             Switzerland
195   1966        Portugal                5               Korea DPR
213   1970     Germany FR                 5                Bulgaria
243   1974      Yugoslavia                9                   Zaire
288   1978     Germany FR                 6                  Mexico
295   1978     Netherlands                5                 Austria
305   1978       Argentina                6                    Peru
312   1982         Hungary               10             El Salvador
313   1982        Scotland                5             New Zealand
333   1982          Poland                5                    Peru
363   1986    Soviet Union                6                 Hungary
382   1986         Denmark                6                 Uruguay
429   1990     Germany FR                 5  rn">United Arab Emirates
495   1994          Russia                6                Cameroon
542   1998     Netherlands                5             Korea Republic
544   1998       Argentina                5                 Jamaica
554   1998           Spain                6                Bulgaria
583   2002         Germany                8            Saudi Arabia
664   2006       Argentina                6  rn">Serbia and Montenegro
737   2010        Portugal                7               Korea DPR
```

6. List all matches where attendance was above 80,000.

```
# Solution 6
high_attendance_matches = df[df['Attendance'] > 80000]
print(high_attendance_matches[['Year', 'City', 'Home Team Name', 'Away Team Name', 'Attendance']])
```

```
     Year            City Home Team Name  Away Team Name  Attendance
53   1950  Rio De Janeiro         Brazil          Mexico     81649.0
63   1950  Rio De Janeiro         Brazil      Yugoslavia    142429.0
70   1950  Rio De Janeiro         Brazil          Sweden    138886.0
71   1950  Rio De Janeiro         Brazil           Spain    152772.0
74   1950  Rio De Janeiro        Uruguay          Brazil    173850.0
168  1966          London        England         Uruguay     87148.0
183  1966          London        England          Mexico     92570.0
188  1966          London        England          France     98270.0
192  1966          London        England       Argentina     98584.0
197  1966          London        England        Portugal     94493.0
198  1966          London       Portugal    Soviet Union     87696.0
199  1966          London        England  Germany FR       96924.0
200  1970     Mexico City         Mexico    Soviet Union    107160.0
207  1970     Mexico City        Belgium     El Salvador     92205.0
211  1970     Mexico City   Soviet Union         Belgium     95261.0
215  1970     Mexico City         Mexico     El Salvador    103850.0
219  1970     Mexico City   Soviet Union     El Salvador     89979.0
223  1970     Mexico City         Mexico         Belgium    108192.0
229  1970     Mexico City          Italy  Germany FR       102644.0
230  1970     Mexico City  Germany FR           Uruguay    104403.0
231  1970     Mexico City         Brazil           Italy    107412.0
233  1974     Berlin West  Germany FR            Chile     81100.0
308  1982       Barcelona      Argentina         Belgium     95000.0
351  1982          Madrid  Germany FR           Spain      90000.0
359  1982          Madrid         Italy  Germany FR       90000.0
360  1986     Mexico City       Bulgaria           Italy     96000.0
367  1986     Mexico City        Belgium          Mexico    110000.0
379  1986     Mexico City         Mexico        Paraguay    114600.0
389  1986     Mexico City           Iraq          Mexico    103763.0
397  1986     Mexico City         Mexico        Bulgaria    114580.0
402  1986     Mexico City        England        Paraguay     98728.0
407  1986     Mexico City      Argentina         England    114580.0
409  1986     Mexico City      Argentina         Belgium    114580.0
411  1986     Mexico City      Argentina  Germany FR       114600.0
468  1994     Los Angeles       Colombia         Romania     91856.0
471  1994     Los Angeles       Cameroon          Sweden     93194.0
473  1994   San Francisco         Brazil          Russia     81861.0
478  1994     Los Angeles            USA        Colombia     93869.0
483  1994   San Francisco         Brazil        Cameroon     83401.0
488  1994     Los Angeles            USA         Romania     93869.0
489  1994   San Francisco    Switzerland        Colombia     83401.0
503  1994     Los Angeles        Romania       Argentina     90469.0
505  1994   San Francisco         Brazil             USA     84147.0
511  1994   San Francisco        Romania          Sweden     83580.0
513  1994     Los Angeles         Sweden          Brazil     91856.0
514  1994     Los Angeles         Sweden        Bulgaria     91580.0
515  1994     Los Angeles         Brazil           Italy     94194.0
708  2010    Johannesburg   South Africa          Mexico     84490.0
716  2010    Johannesburg    Netherlands         Denmark     83465.0
725  2010    Johannesburg      Argentina  Korea Republic     82174.0
736  2010    Johannesburg         Brazil  Côte d'Ivoire      84455.0
746  2010    Johannesburg          Ghana         Germany     83301.0
750  2010    Johannesburg      Argentina          Mexico     84377.0
765  2010    Johannesburg        Uruguay           Ghana     84017.0
771  2010    Johannesburg    Netherlands           Spain     84490.0
```

7. Identify the team that scored the most goals as a home team across all matches.

+ Code + Text

9. Find the average attendance per tournament year.

```
[33] # Solution 14
     avg_attendance_per_year = df.groupby('Year')['Attendance'].mean()
     print("Average attendance per year:\n", avg_attendance_per_year)
```

```
Average attendance per year:
 Year
1930    32808.277778
1934    21352.941176
1938    20872.222222
1950    47511.181818
1954    29561.807692
1958    23423.142857
1962    27911.625000
1966    48847.968750
1970    50124.218750
1974    49898.763158
1978    60678.710526
1982    40571.596154
1986    66039.057692
1990    48388.750000
1994    68991.135385
1998    43517.187500
2002    42268.703125
2006    52491.234375
2010    49669.625000
2014    53174.910256
Name: Attendance, dtype: float64
```

10. Find the city that hosted the maximum number of matches.

```
[21] # Solution 16
     city_most_matches = df['City'].value_counts().idxmax()
     print("City with most matches:", city_most_matches)
```

```
City with most matches: Mexico City
```

11. Find the most common matchup (team vs team) in World Cup history.

```
[24] # Solution 19
     matchups = df.groupby(['Home Team Name', 'Away Team Name']).size().reset_index(name='Counts')
     most_common_matchup = matchups.loc[matchups['Counts'].idxmax()]
     print("Most common matchup:\n", most_common_matchup)
```

```
Most common matchup:
 Home Team Name    Brazil
Away Team Name     Chile
Counts                 5
Name: 81, dtype: object
```

5. Find the match with the maximum attendance.

```
[10] # Solution 5
     max_attendance_match = df.loc[df['Attendance'].idxmax()]
     print("Match with maximum attendance:\n", max_attendance_match)
```

```
Match with maximum attendance:
 Year                                                    1950
Datetime                                  16 Jul 1950 - 15:00
Stadium                 Maracan�  - Est�dio Jornalista M�rio Filho
City                                          Rio De Janeiro
Home Team Name                                       Uruguay
Home Team Goals                                            2
Away Team Name                                        Brazil
Attendance                                          173850.0
Half-time Home Goals                                       0
Referee                                    READER George (ENG)
Assistant 1                                 ELLIS Arthur (ENG)
Name: 74, dtype: object
```

3. Find the year with the maximum number of matches.

```
[5]  # Solution 3
     year_with_max_matches = df['Year'].value_counts().idxmax()
     print("Year with maximum matches:", year_with_max_matches)
```

```
Year with maximum matches: 2014
```

4. Calculate the average number of goals scored by home teams.

```
[6]  # Solution 4
     avg_home_goals = df['Home Team Goals'].mean()
     print("Average Home Team Goals:", avg_home_goals)
```

```
Average Home Team Goals: 1.8110328638497653
```

2. Identify all unique cities where matches were hosted.

```
[4] # Solution 2
    unique_cities = df['City'].dropna().unique()
    print("Unique cities:", unique_cities)
```

```
Unique cities: ['Montevideo ' 'Turin ' 'Naples ' 'Milan ' 'Bologna ' 'Florence ' 'Genoa '
 'Rome ' 'Trieste ' 'Paris ' 'Reims ' 'Colombes ' 'Toulouse '
 'Marseilles ' 'Strasbourg ' 'Le Havre ' 'Bordeaux ' 'Lille ' 'Antibes '
 'Rio De Janeiro ' 'Curitiba ' 'Sao Paulo ' 'Belo Horizonte '
 'Porto Alegre ' 'Recife ' 'Berne ' 'Zurich ' 'Geneva ' 'Lausanne '
 'Basel ' 'Lugano ' 'Solna ' 'Gothenburg ' 'Malm' ' 'Sandviken '
 'Norrk'Ping ' 'Vasteras ' 'Udevalla ' 'Halmstad ' 'Boras ' 'Helsingborg '
 'Eskilstuna ' 'Orebro ' 'Arica ' 'Vina Del Mar ' 'Rancagua '
 'Santiago De Chile ' 'London ' 'Sheffield ' 'Liverpool ' 'Middlesbrough '
 'Manchester ' 'Birmingham ' 'Sunderland ' 'Mexico City ' 'Puebla '
 'Leon ' 'Guadalajara ' 'Toluca ' 'Frankfurt/Main ' 'Berlin West '
 'Hamburg ' 'Dortmund ' 'Hanover ' 'D'Sseldorf ' 'Munich ' 'Stuttgart '
 'Gelsenkirchen ' 'Buenos Aires ' 'Mar Del Plata ' 'Rosario ' 'Cordoba '
 'Mendoza ' 'Barcelona ' 'Vigo ' 'Seville ' 'La Coru'A ' 'Elche '
 'Malaga ' 'Gijon ' 'Bilbao ' 'Valencia ' 'Oviedo ' 'Valladolid '
 'Zaragoza ' 'Alicante ' 'Madrid ' 'Irapuato ' 'Monterrey '
 'Nezahualcoyotl ' 'Queretaro ' 'Bari ' 'Cagliari ' 'Verona ' 'Palermo '
 'Udine ' 'Dallas ' 'Chicago ' 'Detroit ' 'New York/New Jersey '
 'Los Angeles ' 'Orlando ' 'Washington Dc ' 'San Francisco ' 'Boston '
 'Saint-Denis ' 'Montpellier ' 'Lens ' 'Nantes ' 'Lyon ' 'Saint-Etienne '
 'Seoul ' 'Ulsan ' 'Niigata ' 'Sapporo ' 'Ibaraki ' 'Busan ' 'Saitama '
 'Gwangju ' 'Kobe ' 'Suwon ' 'Daegu ' 'Jeonju ' 'Jeju ' 'Incheon '
 'Yokohama ' 'Rifu ' 'Oita ' 'Shizuoka ' 'Osaka ' 'Daejeon ' 'Leipzig '
 'Nuremberg ' 'Cologne ' 'Kaiserslautern ' 'Berlin ' 'Johannesburg '
```

Double-click (or enter) to edit

1. Find the total number of matches played.

```
[3] # Solution 1
    import pandas as pd

    df = pd.read_csv('WorldCupMatches1.csv')
    total_matches = len(df)
    print("Total matches played:", total_matches)
```

```
Total matches played: 852
```

2. Identify all unique cities where matches were hosted.

```
[4] # Solution 2
    unique_cities = df['City'].dropna().unique()
    print("Unique cities:", unique_cities)
```

```
Unique cities: ['Montevideo ' 'Turin ' 'Naples ' 'Milan ' 'Bologna ' 'Florence ' 'Genoa '
 'Rome ' 'Trieste ' 'Paris ' 'Reims ' 'Colombes ' 'Toulouse '
 'Marseilles ' 'Strasbourg ' 'Le Havre ' 'Bordeaux ' 'Lille ' 'Antibes '
 'Rio De Janeiro ' 'Curitiba ' 'Sao Paulo ' 'Belo Horizonte '
 'Porto Alegre ' 'Recife ' 'Berne ' 'Zurich ' 'Geneva ' 'Lausanne '
 'Basel ' 'Lugano ' 'Solna ' 'Gothenburg ' 'Malm' ' 'Sandviken '
```