NAME – **KAMBLE ANIKET TUKARAM**

PRN – **202401040123**

ROLL NO – **CS4-05**

Dataset Name = **FIFA Dataset**

```
import pandas as pd
import numpy as np
```

Load the dataset

+ Code + Text

```
df = pd.read_csv("/Fifa_world_cup_matches.csv")
```

1. Total number of matches played

```
total_matches = len(df)
print("1. Total number of matches played = ",total_matches)
```

```
1. Total number of matches played =  64
```

2. Total goals scored in the tournament

```
total_goals = df['number of goals team1'].sum() + df['number of goals team2'].sum()
print("2. Total goals scored in the tournament = ",total_goals)
```

```
2. Total goals scored in the tournament =  172
```

3. Match with the highest number of total goals

```
df['total_goals'] = df['number of goals team1'] + df['number of goals team2']
highest_goal_match = df.loc[df['total_goals'].idxmax()]

print("3. Match with the highest number of total goals = ",highest_goal_match)
```

```
3. Match with the highest number of total goals =  team1                              ENGLAND
team2                                IRAN
possession team1                      72%
possession team2                      19%
possession in contest                  9%
                                     ...
forced turnovers team1                63
forced turnovers team2                72
defensive pressures applied team1    139
defensive pressures applied team2    416
total_goals                            8
Name: 1, Length: 89, dtype: object
```

4. How many matches had zero goals from both teams

```
zero_goal_matches = ((df['number of goals team1'] == 0) & (df['number of goals team2'] == 0)).sum()
print("4. How many matches had zero goals from both teams = ", zero_goal_matches)
```

```
4. How many matches had zero goals from both teams =  7
```

5. Total penalties scored by both teams

```
total_penalties = df['penalties scored team1'].sum() + df['penalties scored team2'].sum()
print("5. Total penalties scored by both teams = ",total_penalties)
```

```
5. Total penalties scored by both teams =  17
```

6. Teams with most goals scored (combined)

```
goals_by_team1 = df.groupby('team1')['number of goals team1'].sum()
goals_by_team2 = df.groupby('team2')['number of goals team2'].sum()
total_goals_by_teams = goals_by_team1.add(goals_by_team2, fill_value=0).sort_values(ascending=False)

print("6. Teams with most goals scored (combined) = ",total_goals_by_teams)
```

```
6. Teams with most goals scored (combined) =  team1
FRANCE            16
ARGENTINA         15
ENGLAND           13
PORTUGAL          12
NETHERLANDS       10
SPAIN              9
CROATIA            8
BRAZIL             8
MOROCCO            6
GERMANY            6
SERBIA             5
SWITZERLAND        5
KOREA REPUBLIC     5
SENEGAL            5
GHANA              5
JAPAN              5
CAMEROON           4
IRAN               4
ECUADOR            4
AUSTRALIA          4
UNITED STATES      3
SAUDI ARABIA       3
POLAND             3
COSTA RICA         3
URUGUAY            2
MEXICO             2
CANADA             2
BELGIUM            1
DENMARK            1
QATAR              1
TUNISIA            1
WALES              1
dtype: int64
```

7. Total number of own goals

```
total_own_goals = df['own goals team1'].sum() + df['own goals team2'].sum()
print("7. Total number of own goals = ",total_own_goals)
```

```
7. Total number of own goals =  2
```

8. Average goal preventions per team

```
avg_goal_preventions_team1 = df['goal preventions team1'].mean()
avg_goal_preventions_team2 = df['goal preventions team2'].mean()
print("8. Average goal preventions per team = ")
print("avg_goal_preventions_team1 = ", avg_goal_preventions_team1)
print("avg_goal_preventions_team2 = ",avg_goal_preventions_team2)
```

```
8. Average goal preventions per team =
avg_goal_preventions_team1 =  11.59375
avg_goal_preventions_team2 =  11.359375
```

9. Total forced turnovers by both teams

```
total_turnovers = df['forced turnovers team1'].sum() + df['forced turnovers team2'].sum()
print("9. Total forced turnovers by both teams = ",total_turnovers )
```

```
9. Total forced turnovers by both teams =  9094
```

10. Average defensive pressures applied by each team

```
avg_def_pressures_team1 = df['defensive pressures applied team1'].mean()
avg_def_pressures_team2 = df['defensive pressures applied team2'].mean()
print("10. Average defensive pressures applied by each team = ")
print("avg_def_pressures_team1 = ",avg_def_pressures_team1)
print("avg_def_pressures_team2 = ",avg_def_pressures_team2)
```

```
10. Average defensive pressures applied by each team =
avg_def_pressures_team1 =  289.75
avg_def_pressures_team2 =  293.265625
```

11. Match with the lowest possession in contest (most one-sided game)

```
lowest_contest_match = df.loc[df['possession in contest'].idxmin()]
print("11. Match with the lowest possession in contest (most one-sided game) = ",lowest_contest_match)
```

```
11. Match with the lowest possession in contest (most one-sided game) =  team1                                UNITED STATES
team2                                      WALES
possession team1                             51%
possession team2                             39%
possession in contest                        10%
                                           ...
forced turnovers team1                        81
forced turnovers team2                        72
defensive pressures applied team1            242
defensive pressures applied team2            292
total_goals                                    2
Name: 3, Length: 89, dtype: object
```

12. Match where Team 1 had highest possession

```
max_possession_team1 = df.loc[df['possession team1'].idxmax()]
print("12. Match where Team 1 had highest possession = ",max_possession_team1)
```

```
12. Match where Team 1 had highest possession =  team1                                    SPAIN
team2                                  COSTA RICA
possession team1                              74%
possession team2                              17%
possession in contest                          9%
                                             ...
forced turnovers team1                        46
forced turnovers team2                        76
defensive pressures applied team1            185
defensive pressures applied team2            585
total_goals                                    7
Name: 10, Length: 89, dtype: object
```

13. Count of matches ending in a draw

```
draws = (df['number of goals team1'] == df['number of goals team2']).sum()
print("13. Count of matches ending in a draw = ",draws)
```

```
13. Count of matches ending in a draw =  15
```

14. Number of matches where Team 1 scored more than Team 2

```
team1_win_count = (df['number of goals team1'] > df['number of goals team2']).sum()
print("14. Number of matches where Team 1 scored more than Team 2 = ",team1_win_count)
```

```
14. Number of matches where Team 1 scored more than Team 2 =  29
```

15. Number of matches where Team 2 had higher possession

```
team2_higher_possession = (df['possession team2'] > df['possession team1']).sum()
print("15. Number of matches where Team 2 had higher possession = ",team2_higher_possession)
```

```
15. Number of matches where Team 2 had higher possession =  26
```

16. Top 5 matches with most defensive pressures combined

```
df['total_def_pressures'] = df['defensive pressures applied team1'] + df['defensive pressures applied team2']
top_defensive_matches = df.nlargest(5, 'total_def_pressures')
print("16. Top 5 matches with most defensive pressures combined = ",top_defensive_matches)
```

```
16. Top 5 matches with most defensive pressures combined =            team1       team2 possession team1 possession team2  \
54      MOROCCO       SPAIN              22%              68%
56      CROATIA      BRAZIL              45%              45%
42        JAPAN       SPAIN              14%              78%
10        SPAIN  COSTA RICA              74%              17%
57  NETHERLANDS   ARGENTINA              45%              44%

   possession in contest  number of goals team1  number of goals team2  \
54                   10%                      0                      0
56                   10%                      1                      1
42                    8%                      2                      1
10                    9%                      7                      0
57                   11%                      2                      2
```

```
            date     hour       category  ...  goal preventions team1  \
54  06 DEC 2022  16 : 00    Round of 16  ...                      14
56  09 DEC 2022  16 : 00  Quarter-final  ...                      21
42  01 DEC 2022  20 : 00        Group E  ...                      14
10  23 NOV 2022  17 : 00        Group E  ...                       0
57  09 DEC 2022  20 : 00  Quarter-final  ...                      15

    goal preventions team2  own goals team1  own goals team2  \
54                       6                0                0
56                       9                0                0
42                       6                0                0
10                      17                0                0
57                       5                0                0

    forced turnovers team1  forced turnovers team2  \
54                     100                      66
56                      95                      77
42                      85                      49
10                      46                      76
57                      91                      79

    defensive pressures applied team1  defensive pressures applied team2  \
54                                573                                217
56                                365                                423
42                                637                                150
10                                185                                585
57                                401                                323

    total_goals  total_def_pressures
54            0                 790
56            2                 788
42            3                 787
10            7                 770
57            4                 724

[5 rows x 90 columns]
```

17. Standard deviation of goals scored by Team 1

```
std_goals_team1 = df['number of goals team1'].std()
print("17. Standard deviation of goals scored by Team 1 = ",std_goals_team1)
```

```
17. Standard deviation of goals scored by Team 1 =  1.5512891464433753
```

18. Mean number of penalties per match

```
mean_penalties = (df['penalties scored team1'] + df['penalties scored team2']).mean()
print("18. Mean number of penalties per match = ",mean_penalties)
```

```
18. Mean number of penalties per match =  0.265625
```

19. Maximum goals scored by a team in a single match

```
df['max_goals_single_team'] = df[['number of goals team1', 'number of goals team2']].max(axis=1)
print("19. Max goals by a single team in a match:", df['max_goals_single_team'].max())
```

```
19. Max goals by a single team in a match: 7
```

20. Team with most wins (assuming Team 1 vs Team 2 result)

```
wins_team1 = df[df['number of goals team1'] > df['number of goals team2']]['team1'].value_counts()
wins_team2 = df[df['number of goals team2'] > df['number of goals team1']]['team2'].value_counts()
total_wins = wins_team1.add(wins_team2, fill_value=0).sort_values(ascending=False)
print("20. Team with most wins:")
print(total_wins.head(1))
```

```
20. Team with most wins:
FRANCE    5.0
Name: count, dtype: float64
```

21. Total number of unique teams participated

```
unique_teams = pd.concat([df['team1'], df['team2']]).nunique()
print("21. Total unique teams participated:", unique_teams)
```

```
21. Total unique teams participated: 32
```

22. Average goals per match

```
avg_goals_per_match = df['total_goals'].mean()
print("22. Average goals per match: {:.2f}".format(avg_goals_per_match))
```

```
22. Average goals per match: 2.69
```