THEORY ACTIVITY NO. 1

NAME;- SANSKAR RAJESH CHANDAWAR

PRN:- 202401040137

ROLL.NO:- CS3-49
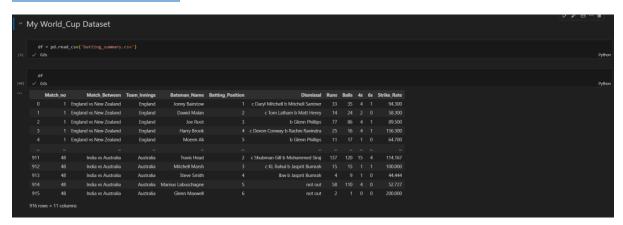
BATCH:- C33

CT_BATCH:- PE-12

PROBLEM STATEMENT:- FORMULATE 20 PROBLEM STATEMENTS FOR A GIVEN DATASET USING NUMPY AND PANDAS

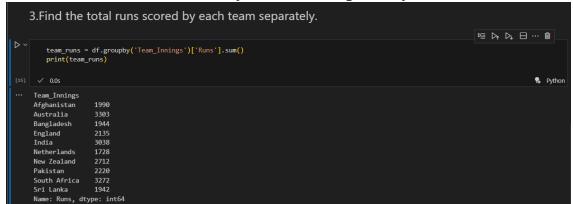DATA_SET TAKEN:- CRICKET WORLD CUP DATASET
LINK:- https://www.kaggle.com/datasets/enggbilalalikhan/odi-world-cup-2023-complete-dataset?resource=download

1. Find the total number of players in the dataset.

2. Find the player who scored the maximum runs.

3. Find the total runs scored by each team separately.

```python
3.Find the total runs scored by each team separately.

team_runs = df.groupby('Team_Innings')['Runs'].sum()
print(team_runs)
```

```
Team_Innings
Afghanistan      1990
Australia        3303
Bangladesh       1944
England          2135
India            3038
Netherlands      1728
New Zealand      2712
Pakistan         2220
South Africa     3272
Sri Lanka        1942
Name: Runs, dtype: int64
```

4. Find the batsman who hit the most sixes.

```python
4.Find the batsman who hit the most sixes

most_sixes = df[df['6s'] == df['6s'].max()]
print(most_sixes[['Batsman_Name', '6s']])
```

```
     Batsman_Name  6s
693  Fakhar Zaman  11
```

5. Find the batsman who played the most balls.

```python
5.Find the batsman who played the most balls.

most_balls = df[df['Balls'] == df['Balls'].max()]
print(most_balls[['Batsman_Name', 'Balls']])
```

```
      Batsman_Name  Balls
756  Ibrahim Zadran   143
```

6.Find the batsman who scored the highest runs.

```python
6.Find the batsman who scored the highest runs.

highest_runs = df[df['Runs'] == df['Runs'].max()]
print(highest_runs[['Batsman_Name', 'Runs']])
```

```
     Batsman_Name  Runs
767  Glenn Maxwell   201
```

7.What is the average (mean) number of balls faced by each batsman?

```python
9.What is the average (mean) number of balls faced by each batsman?

average_balls = df['Balls'].mean()
print("Average Balls Faced:", average_balls)
```

```
Average Balls Faced: 29.021834061135372
```

8. **What is the total number of runs scored in the worldcup dataset?**

8.What is the total number of runs scored in the dataset?

```python
total_runs = df['Runs'].sum()
print("Total Runs:", total_runs)
```

[21] ✓ 0.0s   Python

Total Runs: 24284

9. **Which batsman faced the most balls?**

9.Which batsman faced the most balls?

```python
most_balls = df['Balls'].max()
batsman_most_balls = df[df['Balls'] == most_balls]['Batsman_Name'].values[0]
print("Batsman Faced Most Balls:", batsman_most_balls)
```

[24] ✓ 0.0s   Python

Batsman Faced Most Balls: Ibrahim Zadran

10. **What is the highest strike rate recorded and who achieved it?**

10.What is the highest strike rate recorded and who achieved it?

```python
max_sr = df['Strike_Rate'].max()
batsman_max_sr = df[df['Strike_Rate'] == max_sr]['Batsman_Name'].values[0]
print("Highest Strike Rate:", max_sr, "by", batsman_max_sr)
```

[25] ✓ 0.0s   Python

Highest Strike Rate: 98.592 by Imam-ul-Haq

11. **What is the total number of boundaries (4s + 6s) hit by all batsmen?**

11.What is the total number of boundaries (4s + 6s) hit by all batsmen?

```python
total_boundaries = (df['4s'] + df['6s']).sum()
print("Total Boundaries:", total_boundaries)
```

[27] ✓ 0.0s   Python

Total Boundaries: 2987

12. **What is the lowest number of runs scored by any batsman?**

12.What is the lowest number of runs scored by any batsman?

```python
min_runs = df['Runs'].min()
print("Lowest Runs by Any Batsman:", min_runs)
```

[29] ✓ 0.0s   Python

Lowest Runs by Any Batsman: 0

13.How many times did a batsman get out caught (look for 'c ' in Dismissal)?

13.How many times did a batsman get out caught (look for 'c ' in Dismissal)?

```python
caught_count = df['Dismissal'].str.contains('c ').sum()
print("Caught Dismissals:", caught_count)
```

```
Caught Dismissals: 484
```

14.List all batsmen who hit at least one six.

14.List all batsmen who hit at least one six.

```python
batsmen_with_six = df[df['6s'] > 0]['Batsman_Name'].unique()
print("Batsmen with at least one 6:", batsmen_with_six)
```

```
Batsmen with at least one 6: ['Jonny Bairstow' 'Joe Root' 'Harry Brook' 'Jos Buttler' 'Adil Rashid'
 'Devon Conway' 'Rachin Ravindra' 'Saud Shakeel' 'Shadab Khan'
 'Haris Rauf' 'Vikramjit Singh' 'Bas de Leede' 'Logan van Beek'
 'Rahmanullah Gurbaz' 'Ibrahim Zadran' 'Najmul Hossain Shanto'
 'Quinton de Kock' 'Rassie van der Dussen' 'Aiden Markram'
 'Heinrich Klaasen' 'David Miller' 'Marco Jansen' 'Kusal Mendis'
 'Sadeera Samarawickrama' 'Charith Asalanka' 'Dasun Shanaka'
 'Kasun Rajitha' 'Pat Cummins' 'Mitchell Starc' 'KL Rahul' 'Hardik Pandya'
 'Will Young' 'Daryl Mitchell' 'Tom Latham' 'Mitchell Santner'
 'Matt Henry' 'Teja Nidamanuru' 'Scott Edwards' 'Aryan Dutt' 'Dawid Malan'
 'Litton Das' 'Taskin Ahmed' 'Pathum Nissanka' 'Abdullah Shafique'
 'Mohammad Rizwan' 'Hashmatullah Shahidi' 'Azmatullah Omarzai'
 'Rashid Khan' 'Rohit Sharma' 'Ishan Kishan' 'Shreyas Iyer'
 'Kagiso Rabada' 'Travis Head' 'David Warner' 'Shakib Al Hasan'
 'Mushfiqur Rahim' 'Mahmudullah' 'Kane Williamson' 'Glenn Phillips'
 'Ikram Alikhil' 'Mujeeb Ur Rahman' 'Josh Inglis' 'Glenn Maxwell'
 'Marcus Stoinis' 'Sybrand Engelbrecht' 'Roelof van der Merwe'
 'Temba Bavuma' 'Gerald Coetzee' 'Keshav Maharaj' 'Mark Chapman'
 'Tanzid Hasan' 'Shoriful Islam' 'Shubman Gill' 'Virat Kohli'
 'Mitchell Marsh' 'Iftikhar Ahmed' 'Mohammad Nawaz' "Max O'Dowd"
 'Dhananjaya de Silva' 'Reeza Hendricks' 'David Willey' 'Mark Wood'
 'Ravindra Jadeja' 'Babar Azam' 'Rahmat Shah' 'Steve Smith'
 'Marnus Labuschagne' 'Mohammad Wasim Jr' 'Jimmy Neesham' 'Trent Boult'
 'Mehidy Hasan Miraz' 'Mustafizur Rahman' 'Suryakumar Yadav'
 'Angelo Mathews' 'Maheesh Theekshana' 'Towhid Hridoy' 'Fakhar Zaman'
 'Ben Stokes' 'Chris Woakes' 'Mohammad Nabi' 'Wesley Barresi'
 'Kusal Janith Perera' 'Andile Phehlukwayo' 'Moeen Ali' 'Agha Salman'
 'Shaheen Afridi' 'Lockie Ferguson']
```

15.How many entries are there for batting position 3?

15.How many entries are there for batting position 3?

```python
pos3_count = df[df['Batting_Position'] == 3].shape[0]
print("Entries for Batting Position 3:", pos3_count)
```

```
Entries for Batting Position 3: 100
```

16.What is the average runs scored by batsmen at batting position 1?

16.What is the average runs scored by batsmen at batting position 1?

```python
avg_runs_pos1 = df[df['Batting_Position'] == 1]['Runs'].mean()
print("Average Runs at Position 1:", avg_runs_pos1)
```

```
Average Runs at Position 1: 36.84
```

17.How many times did a batsman get dismissed as 'bowled'?

```python
bowled_count = df['Dismissal'].str.contains('b ').sum()
print("Bowled Dismissals:", bowled_count)
```

[35] ✓ 0.0s   Python

```
Bowled Dismissals: 724
```

18. How many balls did "Jonny Bairstow" face?

```python
jonny_balls = df[df['Batsman_Name'] == 'Jonny Bairstow']['Balls']
print(jonny_balls)
```

[36] ✓ 0.0s   Python

```
0      35
108    59
283     4
416    12
497    31
588    23
706     1
771    17
844    61
Name: Balls, dtype: int64
```

19. Show all batsmen who scored more than 50 runs.

```python
above_50 = df[df['Runs'] > 50]
print(above_50[['Batsman_Name', 'Runs']])
```

[38] ✓ 0.0s   Python

```
        Batsman_Name  Runs
2            Joe Root    77
11      Devon Conway   152
13    Rachin Ravindra  123
17    Mohammad Rizwan   68
18       Saud Shakeel   68
..               ...   ...
891     Daryl Mitchell  134
901        Virat Kohli   54
903           KL Rahul   66
911        Travis Head  137
914  Marnus Labuschagne   58

[160 rows x 2 columns]
```

20. Find which batting position scored the maximum runs.

20.Find which batting position scored the maximum runs.

```python
best_position = df.groupby('Batting_Position')['Runs'].sum().idxmax()
print("Best Batting Position:", best_position)
```

[39] ✓ 0.0s

Best Batting Position: 3

THANK YOU