

EDS Theory Activity No:-1

Name :- Ayush Satose

Roll no :- CC-09

PRN :- 202401050038

Activity:- Formulate 20 problem statements for a **IPL** dataset using Numpy and Pandas and Apply Numpy and pandas methods to find the solution for the formulated problem statements.

1. How many matches are there?

A screenshot of a Jupyter Notebook interface with a dark theme. It shows four code cells, each with a green checkmark and '0s' execution time. The code in the cells is: 1. 'import pandas as pd', 2. 'data = pd.read_csv("/deliveries.csv")', 3. 'df = pd.DataFrame(data)', and 4. 'q1 = df['match_id'].unique()' followed by 'print(q1)'. The output of the fourth cell is '1095'. A toolbar with various icons is visible on the right side of the code area.

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q1 = df['match_id'].unique()
print(q1)
```

1095

2. How many unique batting teams?

A screenshot of a Jupyter Notebook interface with a dark theme. It shows four code cells, each with a green checkmark and '0s' execution time. The code in the cells is: 1. 'import pandas as pd', 2. 'data = pd.read_csv("/deliveries.csv")', 3. 'df = pd.DataFrame(data)', and 4. 'q2 = df['batting_team'].unique()' followed by 'print(q2)'. The output of the fourth cell is '19'. A toolbar with various icons is visible on the right side of the code area.

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q2 = df['batting_team'].unique()
print(q2)
```

19

3. How many unique bowling teams?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q3 = df['bowling_team'].nunique()
print(q3)
```

19

4. How many different batters?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q4 = df['batter'].nunique()
print(q4)
```

673

5. How many different bowlers?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q5 = df['bowler'].nunique()
print(q5)
```

530

6. How many balls were delivered in total?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q6 = len(df)
print(q6)
```

260920

7. How many runs were scored in total?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q7 = df['total_runs'].sum()
print(q7)
```

347756

8. Which batter scored the most runs?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q8 = df.groupby('batter')['batsman_runs'].sum().idxmax()
print(q8)
```

V Kohli

9. How many wickets fell in total?

```
import pandas as pd
data = pd.read_csv("/deliveries.csv")
df = pd.DataFrame(data)
q9 = df['is_wicket'].sum()
print(q9)
```

12950

10. Which bowler took the most wickets?

```
import pandas as pd
data = pd.read_csv("/deliveries.csv")
df = pd.DataFrame(data)
q10 = df[df['is_wicket'] == 1].groupby('bowler').size().idxmax()
print(q10)
```

YS Chahal

11. Which team scored the most total runs?

```
import pandas as pd
data = pd.read_csv("/deliveries.csv")
df = pd.DataFrame(data)
q11 = df.groupby('batting_team')['total_runs'].sum().idxmax()
print(q11)
```

Mumbai Indians

12. Which team conceded the most total runs?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q12 = df.groupby('bowling_team')['total_runs'].sum().idxmax()
print(q12)
```

Mumbai Indians

13. What is the most common dismissal kind?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q13 = df['dismissal_kind'].value_counts().idxmax()
print(q13)
```

caught

14. How many leg byes were there?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q14 = df[df['extras_type'] == 'legbyes'].shape[0]
print(q14)
```

4001

15. How many wides were bowled?

```
import pandas as pd
data = pd.read_csv("/deliveries.csv")
df = pd.DataFrame(data)
q15 = df[df['extras_type'] == 'wides'].shape[0]
print(q15)
```

8380

16. Which fielder made the most dismissals?

```
import pandas as pd
data = pd.read_csv("/deliveries.csv")
df = pd.DataFrame(data)
q16 = df['fielder'].value_counts().idxmax()
print(q16)
```

MS Dhoni

17. How many innings are there

```
import pandas as pd
data = pd.read_csv("/deliveries.csv")
df = pd.DataFrame(data)
q17 = df['inning'].unique()
print(q17)
```

6

18. What was the maximum runs scored in a single ball?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q18 = df['total_runs'].max()
print(q18)
```

7

19. Which match had the most total runs?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q19 = df.groupby('match_id')['total_runs'].sum().idxmax()
print(q19)
```

1426268

20. Which over had the most runs scored across all matches?

```
import pandas as pd

data = pd.read_csv("/deliveries.csv")

df = pd.DataFrame(data)

q20 = df.groupby('over')['total_runs'].sum().idxmax()
print(q20)
```

17

