MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

EDS Theory Activity 01

Name:-Siddhi Prakashrao Farkandkar

Roll No.-CS8-52

Batch-CS83

PRN-202401120029

Topic- Dataset on IPL

Code for my dataset and problem Statement -

eds.py > ...

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

data = {
    'match_id': [1, 2, 3, 4, 5, 6],
    'season': [2018, 2018, 2019, 2019, 2020, 2020],
    'team': ['Team A', 'Team B', 'Team A', 'Team B', 'Team A', 'Team B'],
    'player': ['Player 1', 'Player 2', 'Player 3', 'Player 4', 'Player 5', 'Player 6'],
    'runs': [50, 30, 40, 60, 10, 20],
    'wickets': [2, 3, 1, 4, 0, 1],
    'balls': [30, 35, 40, 38, 42, 37],
    'sixes': [3, 2, 5, 1, 0, 1],
    'fours': [5, 3, 6, 7, 2, 3],
    'result': ['win', 'loss', 'win', 'win', 'loss', 'win'],
    'total_runs_scored': [180, 160, 170, 150, 140, 130],
    'total_runs_conceded': [160, 170, 150, 140, 130, 120],
    'overs_bowled': [4, 4, 4, 4, 4, 4]
}

df = pd.DataFrame(data)

# 1. Top 5 Players with Most Runs in a Season
top_5_players_runs = df.groupby(['season', 'player'])['runs'].sum().reset_index()
top_5_players_runs = top_5_players_runs.sort_values(['season', 'runs'], ascending=[True, False]).groupby('season').head(5)

# 2. Seasonal Player Performance Summary (Total Runs, Wickets, and Sixes)
player_performance_season = df.groupby(['season', 'player']).agg({'runs': 'sum', 'wickets': 'sum', 'sixes': 'sum'}).reset_index()

# 3. Team Performance by Season (Average Runs, Total Wickets, Average Economy Rate)
df['economy_rate'] = df['total_runs_conceded'] / df['overs_bowled']
team_performance_season = df.groupby(['season', 'team']).agg(
    {'runs': 'mean', 'wickets': 'sum', 'economy_rate': 'mean'}).reset_index()

# 4. Batting Average by Player and Season
df['batting_average'] = df['runs'] / df['wickets'].replace(0, np.nan)
batting_average_by_player = df.groupby(['season', 'player'])['batting_average'].max().reset_index()

# 5. Top 3 Bowlers with Best Economy Rate in Each Season
top_3_bowlers = df.groupby(['season', 'player']).agg({'economy_rate': 'min'}).reset_index()
top_3_bowlers = top_3_bowlers.sort_values(['season', 'economy_rate']).groupby('season').head(3)

# 6. Matches with the Highest Run Chase Successfully Completed
highest_run_chase_match = df[df['result'] == 'win'].groupby(['match_id'])['total_runs_scored'].max().reset_index()

# 7. Player Performance (Runs and Wickets) Comparison
player_performance_match = df.groupby(['match_id', 'player']).agg({'runs': 'sum', 'wickets': 'sum'}).reset_index()

# 8. Player with the Most Boundaries in Each Season
df['boundaries'] = df['fours'] + df['sixes']
most_boundaries_season = df.groupby(['season', 'player'])['boundaries'].sum().reset_index()
most_boundaries_season = most_boundaries_season.sort_values(['season', 'boundaries'], ascending=[True, False]).groupby('season').head(1)

# 9. Average Runs Scored by Each Team in a Given Season
avg_runs_by_team_season = df.groupby(['season', 'team'])['runs'].mean().reset_index()

# 10. Top 3 Teams with the Most Wins and Highest Run Rates
team_performance_win = df[df['result'] == 'win'].groupby(['team']).agg({'runs': 'sum', 'total_runs_conceded': 'sum'}).reset_index()
team_performance_win['run_rate'] = team_performance_win['runs'] / team_performance_win['total_runs_conceded']
top_3_teams_win = team_performance_win.sort_values('run_rate', ascending=False).head(3)

# 11. Player with the Highest Batting Strike Rate in a Given Season
df['strike_rate'] = (df['runs'] / df['balls']) * 100
highest_strike_rate = df.groupby(['season', 'player'])['strike_rate'].max().reset_index()

# 12. Player vs Player Comparison of Runs Scored in Matches
player_comparison = df.groupby('player')['runs'].sum().reset_index()
player_comparison = player_comparison.sort_values('runs', ascending=False)

# 13. Best Batting Performance in a Match
best_batting_performance = df.groupby(['match_id'])['runs'].max().reset_index()
best_batting_performance = df[df['runs'].isin(best_batting_performance['runs'])]

```

```python
# 14. Player with the Most Ducks (0 Runs) in a Given Season
consistent_players = df[df['runs'] >= 30].groupby(['season', 'player']).size().reset_index(name='30+_run_matches')
most_consistent_players = consistent_players.sort_values(['season', '30+_run_matches'], ascending=[True, False]).groupby('season').head(1)


# 15. Average Runs Conceded by Teams in a Given Season
avg_runs_conceded_team_season = df.groupby(['season', 'team'])['total_runs_conceded'].mean().reset_index()

# 16. Players with 5 or More Wickets in a Match
all_rounders = df[(df['runs'] >= 30) & (df['wickets'] >= 2)].groupby('player').size().reset_index(name='match_winning_allrounder_performance')


# 17. Player's Contribution to Team Victory
player_contribution_victory = df[df['result'] == 'win'].groupby(['player']).agg({'runs': 'sum', 'wickets': 'sum'}).reset_index()

# 18. Players' Performance in Matches with the Highest Runs Conceded
max_runs_conceded = df.groupby('match_id')['total_runs_conceded'].max().reset_index()
players_max_runs_conceded = df[df['match_id'].isin(max_runs_conceded['match_id'])]

# 19. Season-Wise Team Ranking Based on Wickets Taken
team_wickets_season = df.groupby(['season', 'team'])['wickets'].sum().reset_index()
team_wickets_season = team_wickets_season.sort_values(['season', 'wickets'], ascending=[True, False])

# 20. Batting and Bowling Performance in a Season
performance_season = df.groupby(['season', 'player']).agg({'runs': 'sum', 'wickets': 'sum'}).reset_index()
```

```python
99
100   print("Top 5 Players with Most Runs in a Season:\n", top_5_players_runs)
101   print("\nSeasonal Player Performance Summary:\n", player_performance_season)
102   print("\nTeam Performance by Season:\n", team_performance_season)
103   print("\nBatting Average by Player and Season:\n", batting_average_by_player)
104   print("\nTop 3 Bowlers with Best Economy Rate in Each Season:\n", top_3_bowlers)
105   print("\nMatches with the Highest Run Chase Successfully Completed:\n", highest_run_chase_match)
106   print("\nPlayer Performance (Runs and Wickets) Comparison:\n", player_performance_match)
107   print("\nPlayer with the Most Boundaries in Each Season:\n", most_boundaries_season)
108   print("\nAverage Runs Scored by Each Team in a Given Season:\n", avg_runs_by_team_season)
109   print("\nTop 3 Teams with the Most Wins and Highest Run Rates:\n", top_3_teams_win)
110   print("\nPlayer with the Highest Batting Strike Rate in a Given Season:\n", highest_strike_rate)
111   print("\nPlayer vs Player Comparison of Runs Scored in Matches:\n", player_comparison)
112   print("\nBest Batting Performance in a Match:\n", best_batting_performance)
113   print("\nPlayer with the Most Ducks (0 Runs) in a Given Season:\n",consistent_players)
114   print("\nAverage Runs Conceded by Teams in a Given Season:\n", avg_runs_conceded_team_season)
115   print("\nPlayers with 5 or More Wickets in a Match:\n", all_rounders)
116   print("\nPlayer's Contribution to Team Victory:\n", player_contribution_victory)
117   print("\nPlayers' Performance in Matches with the Highest Runs Conceded:\n", players_max_runs_conceded)
118   print("\nSeason-Wise Team Ranking Based on Wickets Taken:\n", team_wickets_season)
119   print("\nBatting and Bowling Performance in a Season:\n", performance_season)
120
```

Problem Statement and their output-

1] Top 5 Players with Most Runs in a Season.

```
Top 5 Players with Most Runs in a Season:
    season    player  runs
0     2018  Player 1    50
1     2018  Player 2    30
3     2019  Player 4    60
2     2019  Player 3    40
5     2020  Player 6    20
4     2020  Player 5    10
```

2] Seasonal Player Performance Summary (Total Runs, Wickets, and Sixes)

```
Seasonal Player Performance Summary:
    season    player  runs  wickets  sixes
0     2018  Player 1    50        2      3
1     2018  Player 2    30        3      2
2     2019  Player 3    40        1      5
3     2019  Player 4    60        4      1
4     2020  Player 5    10        0      0
5     2020  Player 6    20        1      1
```

3] Team Performance by Season (Average Runs, Total Wickets, Average Economy Rate)

```
Team Performance by Season:
    season    team  runs  wickets  economy_rate
0     2018  Team A  50.0        2         40.0
1     2018  Team B  30.0        3         42.5
2     2019  Team A  40.0        1         37.5
3     2019  Team B  60.0        4         35.0
4     2020  Team A  10.0        0         32.5
5     2020  Team B  20.0        1         30.0
```

4] Batting Average by Player and Season

```
Batting Average by Player and Season:
    season    player  batting_average
0     2018  Player 1             25.0
1     2018  Player 2             10.0
2     2019  Player 3             40.0
3     2019  Player 4             15.0
4     2020  Player 5              NaN
5     2020  Player 6             20.0
```

5] Top 3 Bowlers with Best Economy Rate in Each Season

```
Top 3 Bowlers with Best Economy Rate in Each Season:
    season    player  economy_rate
0     2018  Player 1          40.0
1     2018  Player 2          42.5
3     2019  Player 4          35.0
2     2019  Player 3          37.5
5     2020  Player 6          30.0
4     2020  Player 5          32.5
```

6] Matches with the Highest Run Chase Successfully Completed

```
Matches with the Highest Run Chase Successfully Completed:
    match_id  total_runs_scored
0         1                180
1         3                170
2         4                150
3         6                130
```

7] Player Performance (Runs and Wickets) Comparison

```
Player Performance (Runs and Wickets) Comparison:
    match_id    player  runs  wickets
0         1  Player 1    50        2
1         2  Player 2    30        3
2         3  Player 3    40        1
3         4  Player 4    60        4
4         5  Player 5    10        0
5         6  Player 6    20        1
```

8] Player with the Most Boundaries in Each Season

```
Player with the Most Boundaries in Each Season:
    season    player  boundaries
0     2018  Player 1           8
2     2019  Player 3          11
5     2020  Player 6           4
```

9] Average Runs Scored by Each Team in a Given Season

```
Average Runs Scored by Each Team in a Given Season:
    season    team  runs
0     2018  Team A  50.0
1     2018  Team B  30.0
2     2019  Team A  40.0
3     2019  Team B  60.0
4     2020  Team A  10.0
5     2020  Team B  20.0
```

10] Top 3 Teams with the Most Wins and Highest Run Rates

```
Top 3 Teams with the Most Wins and Highest Run Rates:
      team  runs  total_runs_conceded  run_rate
1  Team B    80                  260  0.307692
0  Team A    90                  310  0.290323
```

11] Player with the Highest Batting Strike Rate in a Given Season

```
Player with the Highest Batting Strike Rate in a Given Season:
   season    player  strike_rate
0    2018  Player 1   166.666667
1    2018  Player 2    85.714286
2    2019  Player 3   100.000000
3    2019  Player 4   157.894737
4    2020  Player 5    23.809524
5    2020  Player 6    54.054054
```

12] Player vs Player Comparison of Runs Scored in Matches

```
Player vs Player Comparison of Runs Scored in Matches:
     player  runs
3  Player 4    60
0  Player 1    50
2  Player 3    40
1  Player 2    30
5  Player 6    20
4  Player 5    10
```

13] Best Batting Performance in a Match

```
Best Batting Performance in a Match:
   match_id  season    team    player  runs  wickets  balls  ...  total_runs_scored  total_runs_conceded  overs_bowled  economy_rate  batting_average  boundaries  strike_rate
0         1    2018  Team A  Player 1    50        2     30  ...                180                  160             4          40.0             25.0           8   166.666667
1         2    2018  Team B  Player 2    30        3     35  ...                160                  170             4          42.5             10.0           5    85.714286
2         3    2019  Team A  Player 3    40        1     40  ...                170                  150             4          37.5             40.0          11   100.000000
3         4    2019  Team B  Player 4    60        4     38  ...                150                  140             4          35.0             15.0           8   157.894737
4         5    2020  Team A  Player 5    10        0     42  ...                140                  130             4          32.5              NaN           2    23.809524
5         6    2020  Team B  Player 6    20        1     37  ...                130                  120             4          30.0             20.0           4    54.054054
```

14] Player with the Most Matches Having 30+ Runs in a Season

```
Player with the Most Ducks (0 Runs) in a Given Season:
    season    player  30+_run_matches
0     2018  Player 1                1
1     2018  Player 2                1
2     2019  Player 3                1
3     2019  Player 4                1
```

15] Average Runs Conceded by Teams in a Given Season

```
Average Runs Conceded by Teams in a Given Season:
    season    team  total_runs_conceded
0     2018  Team A                160.0
1     2018  Team B                170.0
2     2019  Team A                150.0
3     2019  Team B                140.0
4     2020  Team A                130.0
5     2020  Team B                120.0
```

16] All-Rounders with 30+ Runs and 2+ Wickets in the Same Match

```
Players with 5 or More Wickets in a Match:
      player  match_winning_allrounder_performance
0  Player 1                                     1
1  Player 2                                     1
2  Player 4                                     1
```

17] Player's Contribution to Team Victory

```
Player's Contribution to Team Victory:
      player  runs  wickets
0  Player 1    50        2
1  Player 3    40        1
2  Player 4    60        4
3  Player 6    20        1
```

18] Players' Performance in Matches with the Highest Runs Conceded

Players' Performance in Matches with the Highest Runs Conceded:

	match_id	season	team	player	runs	wickets	balls	...	total_runs_scored	total_runs_conceded	overs_bowled	economy_rate	batting_average	boundaries	strike_rate
0	1	2018	Team A	Player 1	50	2	30	...	180	160	4	40.0	25.0	8	166.666667
1	2	2018	Team B	Player 2	30	3	35	...	160	170	4	42.5	10.0	5	85.714286
2	3	2019	Team A	Player 3	40	1	40	...	170	150	4	37.5	40.0	11	100.000000
3	4	2019	Team B	Player 4	60	4	38	...	150	140	4	35.0	15.0	8	157.894737
4	5	2020	Team A	Player 5	10	0	42	...	140	130	4	32.5	NaN	2	23.809524
5	6	2020	Team B	Player 6	20	1	37	...	130	120	4	30.0	20.0	4	54.054054

19] Season-Wise Team Ranking Based on Wickets Taken

```
Season-Wise Team Ranking Based on Wickets Taken:
    season    team  wickets
1    2018  Team B        3
0    2018  Team A        2
3    2019  Team B        4
2    2019  Team A        1
5    2020  Team B        1
4    2020  Team A        0
```

20] Batting and Bowling Performance in a Season

```
Batting and Bowling Performance in a Season:
    season    player  runs  wickets
0    2018  Player 1    50        2
1    2018  Player 2    30        3
2    2019  Player 3    40        1
3    2019  Player 4    60        4
4    2020  Player 5    10        0
5    2020  Player 6    20        1
```