

Name - Atharva Thombare

EDS Activity No. 1

PRN - 202401070137

Roll No. : ET1-10

Dataset : IPL

Importing dataset into Jupyter Python Notebook

jupyter ipl Last Checkpoint: 6 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
[1]: import pandas as pd
df = pd.read_csv('matches.csv')
df.head()
```

	id	season	city	date	match_type	player_of_match	venue	team1	team2	toss_winner	toss_decision	winner	result	result_margin
0	335982	2007/08	Bangalore	2008-04-18	League	BB McCullum	M Chinnaswamy Stadium	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	14
1	335983	2007/08	Chandigarh	2008-04-19	League	MEK Hussey	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	3
2	335984	2007/08	Delhi	2008-04-19	League	MF Maharoof	Feroz Shah Kotla	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils	wickets	
3	335985	2007/08	Mumbai	2008-04-20	League	MV Boucher	Wankhede Stadium	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers Bangalore	wickets	
4	335986	2007/08	Kolkata	2008-04-20	League	DJ Hussey	Eden Gardens	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Riders	wickets	

Problem Statements:

1.Find the total number of matches played.

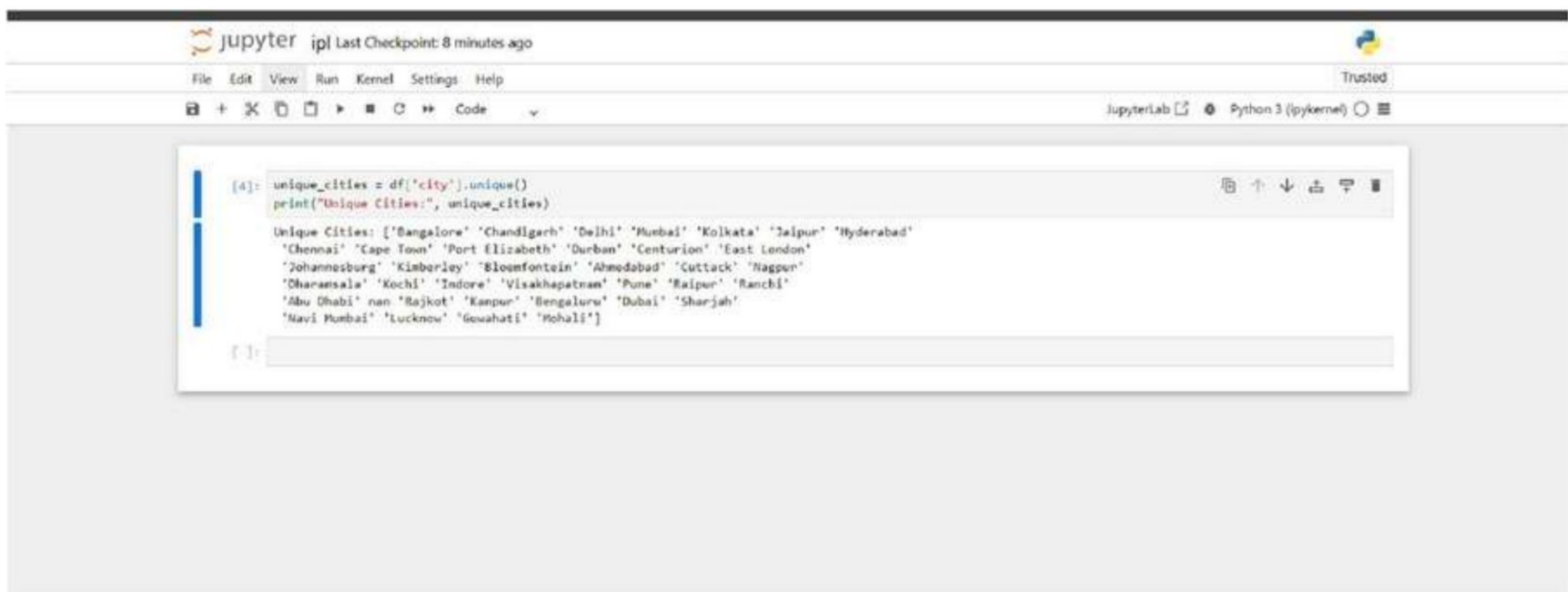
jupyter ipl Last Checkpoint: 7 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
[2]: total_matches = len(df)
print("Total Matches Played:", total_matches)
Total Matches Played: 1095
```

2.LisØng all unique ciØes where matches were played.



The screenshot shows a JupyterLab window with a code editor. The code cell contains the following Python code:

```
[4]: unique_cities = df['city'].unique()
print("Unique Cities:", unique_cities)
```

The output of the code is displayed below the code cell:

```
Unique Cities: ['Bangalore' 'Chandigarh' 'Delhi' 'Mumbai' 'Kolkata' 'Jaipur' 'Hyderabad'
'Chennai' 'Cape Town' 'Port Elizabeth' 'Durban' 'Centurion' 'East London'
'Johannesburg' 'Kimberley' 'Bloemfontein' 'Ahmedabad' 'Cuttack' 'Nagpur'
'Dharamsala' 'Kochi' 'Indore' 'Visakhapatnam' 'Pune' 'Raipur' 'Ranchi'
'Abu Dhabi' nan 'Rajkot' 'Kanpur' 'Bengaluru' 'Dubai' 'Sharjah'
'Mavi Mumbai' 'Lucknow' 'Gowahati' 'Mohali']
```

3. Finding out how many matches had a result decided by runs.



The screenshot shows a JupyterLab window with a code editor. The code cell contains the following Python code:

```
[5]: runs_wins = df[df['result'] == 'runs'].shape[0]
print("Matches won by runs:", runs_wins)
```

The output of the code is displayed below the code cell:

```
Matches won by runs: 498
```

4. Find the player with the most 'Player of the Match' awards.



The screenshot shows a JupyterLab window with a code editor. The code cell contains the following Python code:

```
[6]: top_player = df['player_of_match'].value_counts().idxmax()
print("Top Player of the Match:", top_player)
```

The output of the code is displayed below the code cell:

```
Top Player of the Match: AB de Villiers
```

5. Find all matches where a Super Over was played.


```
jupyter ipl Last Checkpoint: 11 minutes ago
File Edit View Run Kernel Settings Help
jupyterlab Python 3 (ipykernel)

[7]: super_over_matches = df[df['super_over'] == 'Y']
print("Matches with Super Over:", super_over_matches)

Matches with Super Over:
   id  season  city  date match_type player_of_match \
66  392190  2009  Cape Town  2009-04-23  League  YK Pathan
138 419121  2009/10  Chennai  2010-03-21  League  J Theron
328 598004  2013  Hyderabad  2013-04-07  League  GH Vihari
342 598017  2013  Bangalore  2013-04-16  League  V Kohli
416 729315  2014  Abu Dhabi  2014-04-29  League  JP Faulkner
475 829741  2015  Ahmedabad  2015-04-21  League  SE Marsh
610 1082625  2017  Rajkot  2017-04-29  League  KH Pandya
705 1175355  2019  Delhi  2019-03-30  League  PP Shaw
746 1178426  2019  Mumbai  2019-05-02  League  JJ Bumrah
757 1216493  2020/21  NaN  2020-09-20  League  NP Stoinis
765 1216547  2020/21  NaN  2020-09-28  League  AB de Villiers
790 1216512  2020/21  Abu Dhabi  2020-10-18  League  LH Ferguson
791 1216517  2020/21  NaN  2020-10-18  League  KI Rahul
835 1254077  2021  Chennai  2021-04-25  League  PP Shaw

   venue  team1 \
66  Newlands  Kolkata Knight Riders
138  M S Chidambaram Stadium, Chepauk  Chennai Super Kings
328  Rajiv Gandhi International Stadium, Uppal  Sunrisers Hyderabad
342  M Chinnaswamy Stadium  Royal Challengers Bangalore
416  Sheikh Zayed Stadium  Kolkata Knight Riders
475  Sardar Patel Stadium, Motera  Rajasthan Royals
610  Saurashtra Cricket Association Stadium  Gujarat Lions
705  Arun Jaitley Stadium  Kolkata Knight Riders
746  Wankhede Stadium  Mumbai Indians
757  Dubai International Cricket Stadium  Delhi Capitals
765  Dubai International Cricket Stadium  Royal Challengers Bangalore
```

6.Count how many matches each team has played as Team1.

```
jupyter ipl Last Checkpoint: 12 minutes ago
File Edit View Run Kernel Settings Help
jupyterlab Python 3 (ipykernel)

[8]: team1_match_counts = df['team1'].value_counts()
print("Matches played as Team1:\n", team1_match_counts)

Matches played as Team1:
team1
Royal Challengers Bangalore    135
Chennai Super Kings           128
Mumbai Indians                 123
Kolkata Knight Riders          121
Rajasthan Royals               101
Kings XI Punjab                92
Sunrisers Hyderabad           86
Delhi Daredevils               85
Delhi Capitals                 41
Deccan Chargers               39
Punjab Kings                   31
Lucknow Super Giants           23
Pune Warriors                  23
Gujarat Titans                 21
Gujarat Lions                  16
Royal Challengers Bengaluru     9
Kochi Tuskers Kerala           7
Rising Pune Supergiant         7
Rising Pune Supergiants        7
Name: count, dtype: int64
```

7.Find out which team won the maximum tosses.

```
jupyter ipl Last Checkpoint: 14 minutes ago
File Edit View Run Kernel Settings Help
jupyterlab Python 3 (ipykernel)

[9]: top_toss_winner = df['toss_winner'].value_counts().idxmax()
print("Top Toss Winner:", top_toss_winner)

Top Toss Winner: Mumbai Indians
```

8.Find out the number of matches each umpire officiated.


```
jupyter ipl Last Checkpoint: 16 minutes ago
File Edit View Run Kernel Settings Help
JupyterLab Python 3 (ipykernel)

[10]: umpire1_counts = df['umpire1'].value_counts()
      umpire2_counts = df['umpire2'].value_counts()
      total_umpires = umpire1_counts.add(umpire2_counts, fill_value=0)
      print("Matches umpired by each umpire:\n", total_umpires)

Matches umpired by each umpire:
A Deshmukh      12.0
A Nand Kishore  37.0
A Jotir        13.0
AG Warf         2.0
AK Chaudhary    131.0
...
UV Gandhi       31.0
VA Kulkarni     57.0
VK Sharma       65.0
Vinod Seshan    14.0
YC Barde        42.0
Name: count, Length: 70, dtype: float64
```

9. List the top 5 cities that hosted the most matches.

```
jupyter ipl Last Checkpoint: 16 minutes ago
File Edit View Run Kernel Settings Help
JupyterLab Python 3 (ipykernel)

[11]: top_cities = df['city'].value_counts().head(5)
      print("Top 5 Cities with Most Matches:\n", top_cities)

Top 5 Cities with Most Matches:
city
Mumbai      173
Kolkata      93
Delhi        90
Chennai      85
Hyderabad    77
Name: count, dtype: int64
```

10. Calculate the mean result margin when matches were won by wickets.

```
jupyter ipl Last Checkpoint: 17 minutes ago
File Edit View Run Kernel Settings Help
JupyterLab Python 3 (ipykernel)

[12]: mean_margin_wickets = df[df['result'] == 'wickets']['result_margin'].mean()
      print("Mean Result Margin (Wickets Wins):", mean_margin_wickets)

Mean Result Margin (Wickets Wins): 6.192041522491349
```

11. Find the total number of matches decided without using DLS method.


```
jupyter ipl Last Checkpoint: 18 minutes ago
File Edit View Run Kernel Settings Help
jupyterlab Python 3 (ipykernel)

[13]: non_dls_matches = df[df['method'].isna()].shape[0]
      print("Matches without DLS:", non_dls_matches)
      Matches without DLS: 1074
```

12. Find the match with the highest

'target_runs'.

```
jupyter ipl Last Checkpoint: 18 minutes ago
File Edit View Run Kernel Settings Help
jupyterlab Python 3 (ipykernel)


[14]: max_target_match = df[df['target_runs'] == df['target_runs'].max()]
      print("Match with Highest Target Runs:\n", max_target_match)
      Match with Highest Target Runs:
      id season  city  date match_type player_of_match \
1053 1426268  2024  Bengaluru  2024-04-15  League  TM Head
      venue  team1 \
1053  M Chinnaswamy Stadium, Bengaluru  Sunrisers Hyderabad
      team2  toss_winner toss_decision \
1053  Royal Challengers Bengaluru  Royal Challengers Bengaluru  field
      winner result result_margin target_runs target_overs \
1053  Sunrisers Hyderabad  runs  25.0  288.0  20.0
      super_over method  umpire1  umpire2
1053  N  NaN  AK Chaudhary  R Pandit
```

13. Find which season had the most matches played.

```
jupyter ipl Last Checkpoint: 19 minutes ago
File Edit View Run Kernel Settings Help
jupyterlab Python 3 (ipykernel)

[15]: top_season = df['season'].value_counts().idxmax()
      print("Season with most matches:", top_season)
      Season with most matches: 2013
```

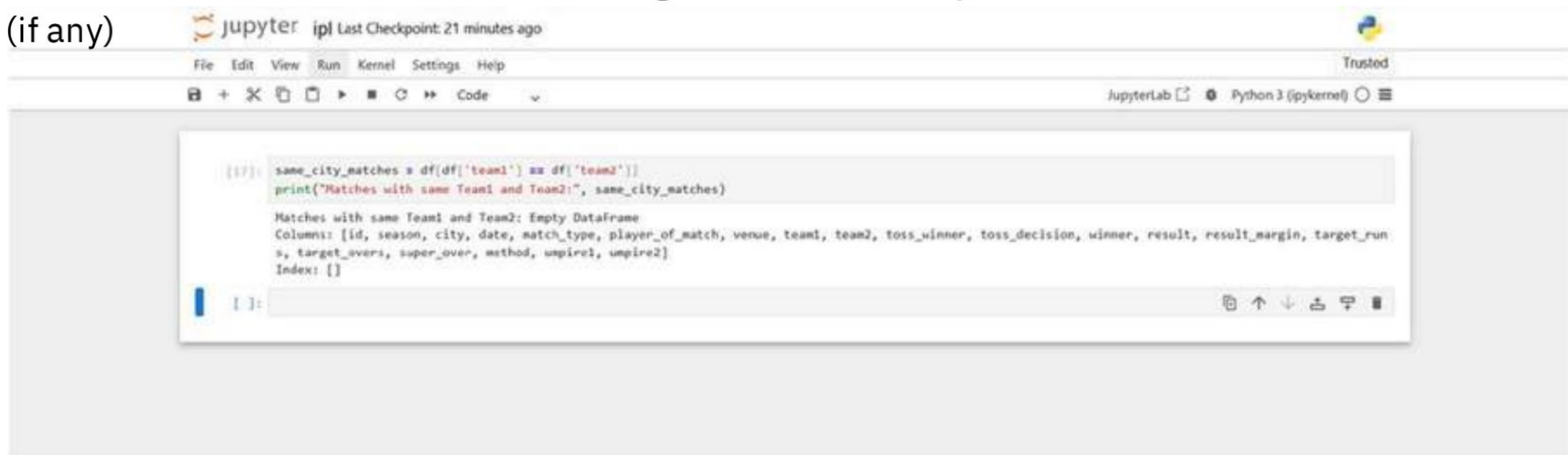
14. Find the percentage of matches where toss-winning team won the match.



```
[16]: toss_win_match_win = df[df['toss_winner'] == df['winner']].shape[0]
percentage = (toss_win_match_win / len(df)) * 100
print(f"Percentage of Toss Winners who won the match: {percentage:.2f}%")

Percentage of Toss Winners who won the match: 50.59%
```

15. List matches where both teams belonged to the same city (if any)



```
[17]: same_city_matches = df[df['team1'] == df['team2']]
print("Matches with same Team1 and Team2:", same_city_matches)

Matches with same Team1 and Team2: Empty Dataframe
Columns: [id, season, city, date, match_type, player_of_match, venue, team1, team2, toss_winner, toss_decision, winner, result, result_margin, target_runs, target_overs, super_over, method, umpire1, umpire2]
Index: []
```

16. Find top 3 players who won Player of the Match most often in finals only.



```
[18]: finals = df[df['match_type'] == 'Final']
top3_finals_players = finals['player_of_match'].value_counts().head(3)
print("Top 3 Players (Finals):\n", top3_finals_players)

Top 3 Players (Finals):
player_of_match
VK Pathan      1
A Kumble       1
SK Raina       1
Name: count, dtype: int64
```

17. Group by city and find average 'result_margin' per city.


```
jupyter ipl Last Checkpoint: 22 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

[10]: avg_margin_city = df.groupby('city')['result_margin'].mean()
      print("Average Result Margin per City:\n", avg_margin_city)

Average Result Margin per City:
city
Abu Dhabi      17.971429
Ahmedabad      16.914286
Bangalore      23.870968
Bengaluru      12.000000
Bloemfontein   10.000000
Cape Town      27.166667
Centurion      10.833333
Chandigarh     15.213115
Chennai        17.987952
Cuttack        13.285714
Delhi          18.181818
Dharamsala     31.461538
Dubai          12.000000
Durban         12.800000
East London    31.333333
Gwalhati        22.333333
Hyderabad      15.592105
Indore         9.888889
Jaipur         16.824561
Johannesburg   7.500000
Kanpur         5.500000
Kimberley      21.000000
Kochi          24.600000
Kolkata        14.946237
Lucknow        20.307692
...         ...
```

18.Using NumPy, find how many matches had a 'target_overs' of exactly 20.

```
jupyter ipl Last Checkpoint: 23 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

[10]: import numpy as np
      target_20 = np.sum(df['target_overs'] == 20)
      print("Matches with 20 overs target:", target_20)

Matches with 20 overs target: 1062

[ ]:
```

19.Find the cities which hosted matches in multiple seasons.

```
jupyter ipl Last Checkpoint: 24 minutes ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

[11]: city_season = df.groupby('city')['season'].unique()
      multi_season_cities = city_season[city_season > 1]
      print("Cities hosting matches in multiple seasons:\n", multi_season_cities)

Cities hosting matches in multiple seasons:
city
Abu Dhabi      3
Ahmedabad      7
Bangalore      9
Bengaluru      5
Chandigarh     12
Chennai        11
Cuttack        3
Delhi          14
Dharamsala     6
Gwalhati        2
Hyderabad      12
Indore         3
Jaipur         9
Kanpur         2
Kolkata        14
Lucknow        2
Mumbai        15
Pune           7
Raipur         3
Rajkot         2
Ranchi         3
Visakhapatnam  5
Name: season, dtype: int64

[ ]:
```

20.List the matches where toss decision was to field but the team lost.


```
[22]: field_loss = df[(df['toss_decision'] == 'field') & (df['toss_winner'] != df['winner'])]
print("Matches where toss winner fielded and lost:\n", field_loss)
```

Matches where toss winner fielded and lost:

	id	season	city	date	match_type	player_of_match	\
0	335982	2007/08	Bangalore	2008-04-18	league	BB McCullum	
7	335989	2007/08	Chennai	2008-04-23	league	ML Hayden	
9	335991	2007/08	Chandigarh	2008-04-25	league	KC Sangakkara	
16	335998	2007/08	Delhi	2008-04-30	league	GD McGrath	
21	336034	2007/08	Bangalore	2008-05-03	league	P Kumar	
...
1005	1426300	2024	Bengaluru	2024-05-12	league	C Green	
1006	1426302	2024	Delhi	2024-05-14	league	I Sharma	
1008	1426305	2024	Mumbai	2024-05-17	league	N Pooran	
1009	1426306	2024	Bengaluru	2024-05-18	league	F du Plessis	
1093	1426311	2024	Chennai	2024-05-24	Qualifier 2	Shahbaz Ahmed	

		venue	team1	\
0		M Chinnaswamy Stadium	Royal Challengers Bangalore	
7		MA Chidambaram Stadium, Chepauk	Chennai Super Kings	
9		Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	
16		Feroz Shah Kotla	Delhi Daredevils	
21		M Chinnaswamy Stadium	Royal Challengers Bangalore	
...	
1005		M Chinnaswamy Stadium, Bengaluru	Royal Challengers Bengaluru	
1006		Arun Jaitley Stadium, Delhi	Delhi Capitals	
1008		Wankhede Stadium, Mumbai	Mumbai Super Giants	
1009		M Chinnaswamy Stadium, Bengaluru	Royal Challengers Bengaluru	
1093		MA Chidambaram Stadium, Chepauk, Chennai	Sunrisers Hyderabad	

		team2	toss_winner	toss_decision	\
0		Kolkata Knight Riders	Royal Challengers Bangalore	field	