



PYTHON AND STATS ANALYSIS ON MATCHES DATASET

Madhura Virkar





AGENDA

1. Basic Exploration
2. Statistics & Aggregation
3. Filtering & Grouping



1. BASIC EXPLORATION



1. How many matches were played in each city?

	count
city	
Mumbai	173
Kolkata	93
Delhi	90
Chennai	85
Hyderabad	77
Bangalore	65
Chandigarh	61
Jaipur	57
Pune	51

```
# Matches were played in each city  
df1['city'].value_counts()
```





2. What is the average target score across all matches?

```
# The average target score across all matches
Avg_target_score = df1['target_runs'].mean()
print('Average target score across all macthes:',round(Avg_target_score))
```

Average target score across all macthes: 166





3. How many matches did each team win?

```
# Matches won by each team
matches_won_each_team = df1.groupby(['team1']).agg({'winner':['count']})
print(matches_won_each_team)
```

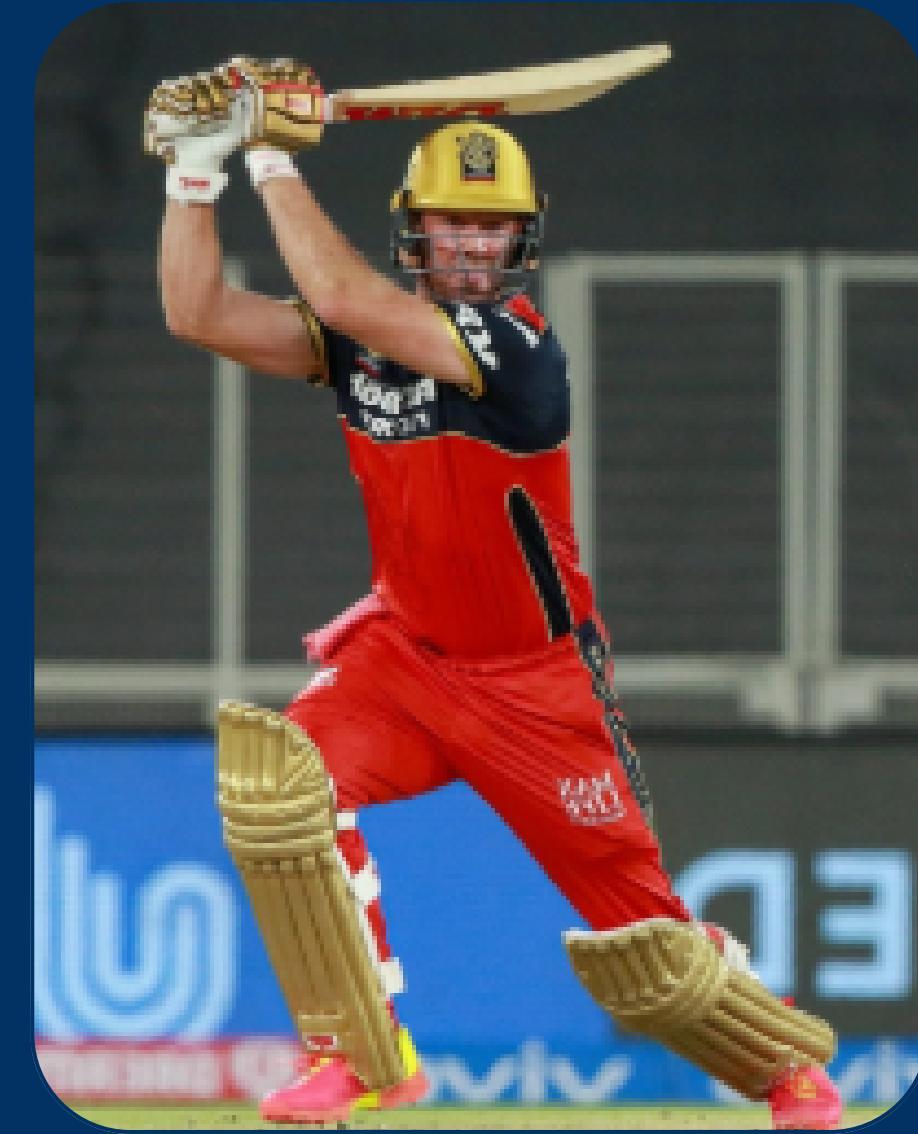
team1	winner count
Chennai Super Kings	128
Deccan Chargers	39
Delhi Capitals	41
Delhi Daredevils	84
Gujarat Lions	16
Gujarat Titans	21
Kings XI Punjab	92
Kochi Tuskers Kerala	7
Kolkata Knight Riders	121
Lucknow Super Giants	22
Mumbai Indians	123
Pune Warriors	23
Punjab Kings	31
Rajasthan Royals	101
Rising Pune Supergiant	7
Rising Pune Supergiants	7
Royal Challengers Bangalore	132
Royal Challengers Bengaluru	9
Sunrisers Hyderabad	86



4. Which player won the most "Player of the Match" awards?

```
#player who won the most "Player of the Match" awards
player_of_the_match = df1['player_of_match'].value_counts().head(1)
print(player_of_the_match)
```

```
player_of_match
AB de Villiers    25
```



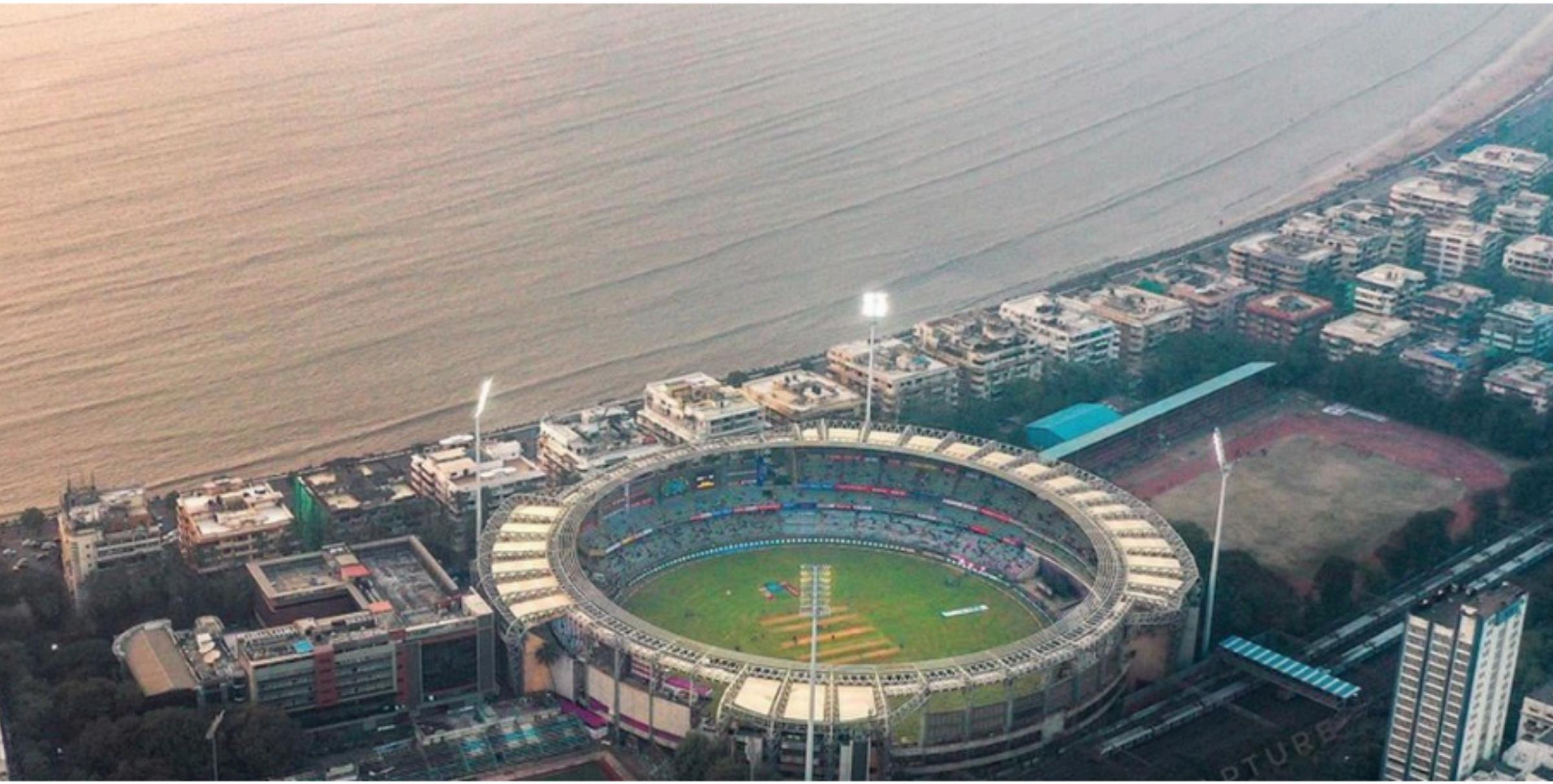


5. Which venue hosted the most matches?



```
# Venue hosted the most matches
venue_hosted_most_matches = df1['venue'].value_counts().head(1)
print(venue_hosted_most_matches)
```

```
venue
[Wankhede Stadium]    118
```



2. STATISTICS & AGGREGATION



6. What is the mean, median, and standard deviation of result_margin for matches won by runs?

```
result_margin_runs = df1[df1['result'] == 'runs']['result_margin']
print('Mean of result_margin:', round(result_margin_runs.mean()))
print('Median of result_margin::', result_margin_runs.median())
print('Standard deviation of result_margin:', round(result_margin_runs.std()))
```

Mean of result_margin: 30
Median of result_margin:: 22.0
Standard deviation of result_margin: 27





7. What is the correlation between target_runs and result_margin (for matches won by runs only)?

```
correlation = df1['target_runs'].corr(df1['result_margin'])  
print(f"Correlation = {correlation:.2f}")
```

Correlation = 0.40



CHAMPIONS

Mumbai Indians	13.15
Chennai Super Kings	12.60
Kolkata Knight Riders	11.96
Royal Challengers Bangalore	10.59
Rajasthan Royals	10.23
Sunrisers Hyderabad	8.04
Kings XI Punjab	8.04
Delhi Daredevils	6.12
Delhi Capitals	4.38
Deccan Chargers	2.65
Gujarat Titans	2.56
Punjab Kings	2.19
Lucknow Super Giants	2.19
Gujarat Lions	1.19
Pune Warriors	1.10
Rising Pune Supergiant	0.91
Royal Challengers Bengaluru	0.64
Kochi Tuskers Kerala	0.55
Rising Pune Supergiants	0.46



8. Calculate the win percentage of each team.

```
win_percentage = (df1['winner'].value_counts() / len(df1)) * 100  
print(win_percentage.map("{:,.2f}".format))
```

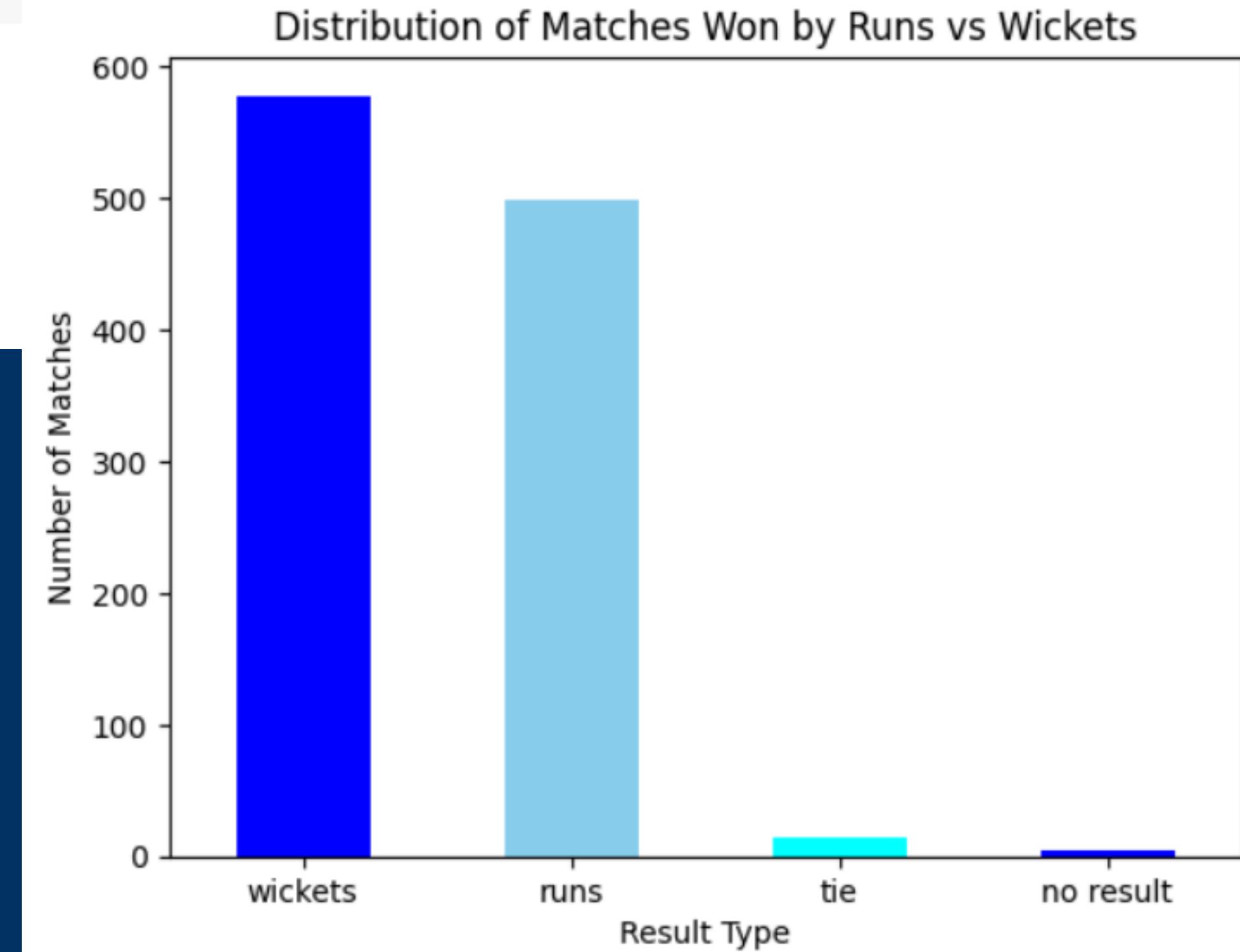


9. What is the distribution of matches won by wickets vs. by runs?

```
distribution_of_matches = df1['result'].value_counts()  
print(distribution_of_matches)
```

result	
wickets	578
runs	498
tie	14
no result	5

Maximum matches were won by wickets





10. Is there a significant difference in the average target score when the toss decision is "bat" vs "field"?

```
target_by_toss = df1.groupby('toss_decision')['target_runs'].mean().reset_index()
print(round(target_by_toss))

toss_decision_bat = df1[df1['toss_decision'] == 'bat']['target_runs'].dropna()
toss_decision_field = df1[df1['toss_decision'] == 'field']['target_runs'].dropna()
t_stat, p_value = stats.ttest_ind(toss_decision_bat, toss_decision_field)
print(f"T-statistic: {t_stat:.2f}")
print(f"P-value: {p_value:.2f}")
```



	toss_decision	target_runs
0	bat	161.0
1	field	168.0

T-statistic: -3.73
P-value: 0.00



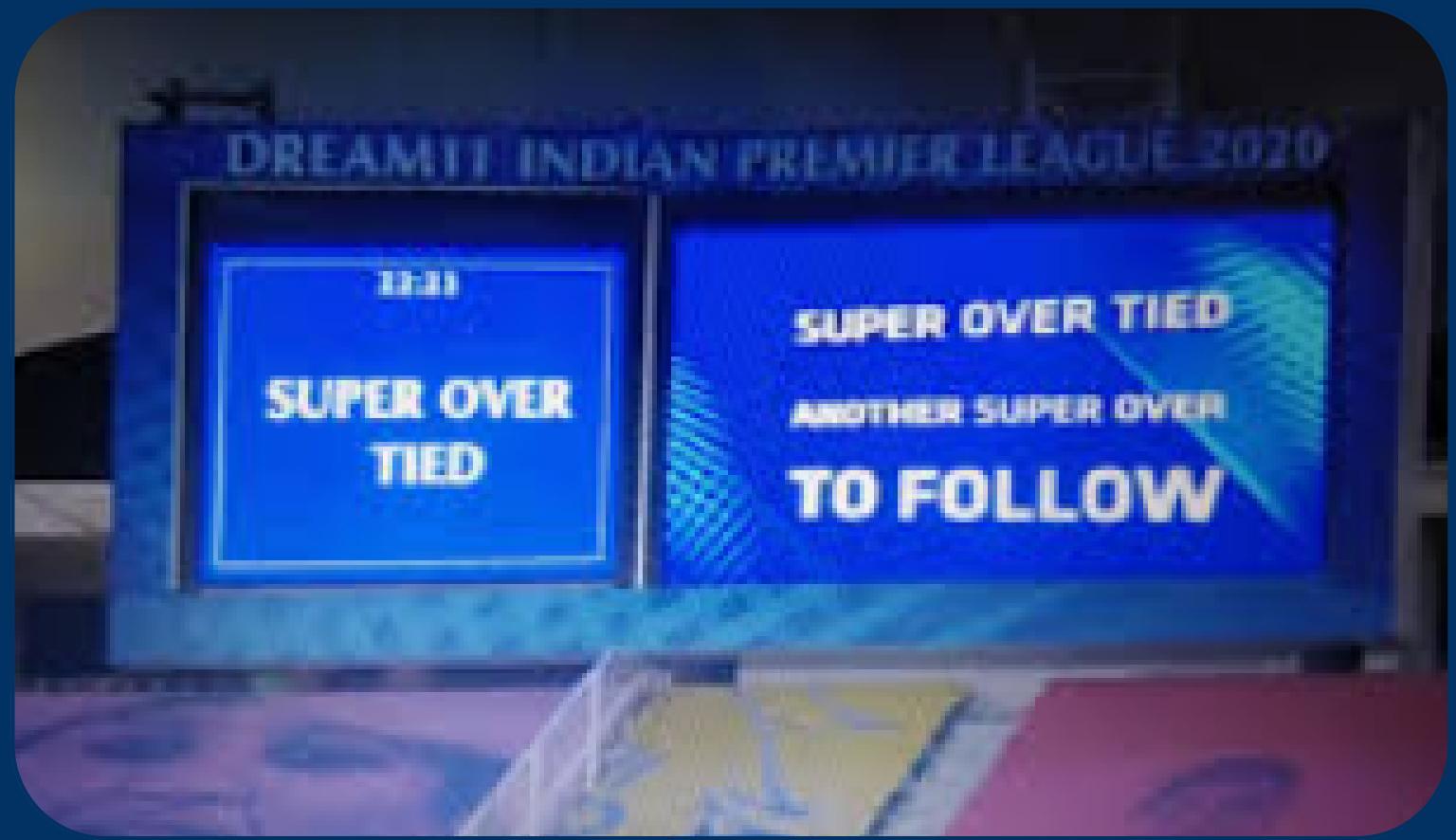
3. FILTERING & GROUPING



11. How many matches went into a super over?

```
super_over = df1[df1['super_over'] == 'Y']
print("Number of matches went into super over:",len(super_over))
```

Number of matches went into super over: 14





12. Which umpires have officiated the most matches together (umpire1 & umpire2 pair)?

```
umpires = df1.groupby(['umpire1', 'umpire2']).size().reset_index(name='count').sort_values(by='count', ascending=False).head(8)  
print(umpires)
```

umpire1	umpire2	count
S Ravi	RJ Tucker	12
AK Chaudhary	Nitin Menon	11
Asad Rauf	S Asnani	11
HDPK Dharmasena	CK Nandan	11
AK Chaudhary	CB Gaffaney	10
AK Chaudhary	NJ Llong	10
BNJ Oxenford	C Shamshuddin	9
S Ravi	C Shamshuddin	9





13. Find the match with the highest result margin (by runs or wickets).

```
highest_result_margin = df1.loc[df1['result_margin'].idxmax()]
print(highest_result_margin)
```

id	1082635
season	2017
city	Delhi
date	05/06/2017
match_type	League
player_of_match	LMP Simmons
venue	Feroz Shah Kotla
Unnamed: 7	NaN
Unnamed: 8	NaN
team1	Delhi Daredevils
team2	Mumbai Indians
toss_winner	Delhi Daredevils
toss_decision	field
winner	Mumbai Indians
result	runs
result_margin	146.0
target_runs	213.0
target_overs	20.0
super_over	N
method	NaN
umpire1	Nitin Menon
umpire2	CK Nandan





14. Which toss decision (bat/field) led to more wins?

```
toss_decision_wins = df1.groupby('toss_decision')['winner'].value_counts()  
print(toss_decision_wins)
```

bat		field		
	Chennai Super Kings	63	Mumbai Indians	90
	Mumbai Indians	54	Kolkata Knight Riders	81
	Kolkata Knight Riders	50	Royal Challengers Bangalore	79
	Rajasthan Royals	43	Chennai Super Kings	75
	Royal Challengers Bangalore	37	Rajasthan Royals	69
	Sunrisers Hyderabad	30	Kings XI Punjab	64
	Delhi Daredevils	29	Sunrisers Hyderabad	58
	Kings XI Punjab	24	Delhi Daredevils	38
	Deccan Chargers	14	Delhi Capitals	35
	Delhi Capitals	13	Punjab Kings	20
	Gujarat Titans	9	Gujarat Titans	19
	Pune Warriors	9	Lucknow Super Giants	18
	Lucknow Super Giants	6	Deccan Chargers	15
	Punjab Kings	4	Gujarat Lions	11
	Gujarat Lions	2	Rising Pune Supergiant	10
	Rising Pune Supergiants	2	Kochi Tuskers Kerala	6
	Royal Challengers Bengaluru	1	Royal Challengers Bengaluru	6
			Pune Warriors	3
			Rising Pune Supergiants	3

Toss decision for fielding led to more wins



15. Which teams have the highest win ratio when winning the toss?

```
toss_winner_wins = df1.groupby('toss_winner')['winner'].size().sort_values( ascending=False)
print(toss_winner_wins)
```

toss_winner	winner
Mumbai Indians	143
Kolkata Knight Riders	122
Chennai Super Kings	122
Rajasthan Royals	120
Royal Challengers Bangalore	113
Sunrisers Hyderabad	88
Kings XI Punjab	85
Delhi Daredevils	80
Delhi Capitals	50
Deccan Chargers	43
Punjab Kings	24
Gujarat Titans	22
Pune Warriors	20
Lucknow Super Giants	19
Gujarat Lions	15
Kochi Tuskers Kerala	8
Royal Challengers Bengaluru	8
Rising Pune Supergiants	7
Rising Pune Supergiant	6





Indian Premier League

**THANK
YOU**

Keep Playing Cricket!

