

Pandas, Matplotlib and Seaborn

```
In [1]: # fillna
# replace
# drop
# drop_duplicates
# outliers detection
# pivot table

# Matplotlib Charts
'''Bar, Pie, Line, Box, violin'''
# Seaborn Charts
'''Bar,Line,Heatmap,Box,Violin,CatPlot,Regplot,lmplot'''
#Feature Engineering
print('Topics for Today')
```

Topics for Today

```
In [2]: import kagglehub

# Download latest version
path = kagglehub.dataset_download("patrickb1912/ipl-complete-dataset-2008

print("Path to dataset files:", path)
```

Path to dataset files: /Users/utkarshtripathi24/.cache/kagglehub/datasets/patrickb1912/ipl-complete-dataset-20082020/versions/3

```
In [3]: import os
file_one,file_two = os.listdir(path)
print('Done')
```

Done

```
In [4]: file_one_path = path + '/' + file_one # also can use '\\' or '/'
file_two_path = path + '/' + file_two # also can use '\\' or '/'
```

```
In [5]: print(file_one_path)
print(file_two_path)
```

/Users/utkarshtripathi24/.cache/kagglehub/datasets/patrickb1912/ipl-complete-dataset-20082020/versions/3/matches.csv
/Users/utkarshtripathi24/.cache/kagglehub/datasets/patrickb1912/ipl-complete-dataset-20082020/versions/3/deliveries.csv

```
In [102... # Step1 : Modules Import
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
print('Done')
```

Done

```
In [7]: df_matches = pd.read_csv(file_one_path)
df_deliveries = pd.read_csv(file_two_path)

print('Done')
```

Done

```
In [8]: # EDA
df_matches.head(2)
```

```
Out[8]:
```

	id	season	city	date	match_type	player_of_match	venue
0	335982	2007/08	Bangalore	2008-04-18	League	BB McCullum	Chinnaswam Stadium
1	335983	2007/08	Chandigarh	2008-04-19	League	MEK Hussey	Punjab Cricket Association Stadium, Mohali

```
In [9]: df_deliveries.head(2)
```

```
Out[9]:
```

	match_id	inning	batting_team	bowling_team	over	ball	batter	bowler	runs
0	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	1	SC Ganguly	P Kumar	0
1	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	2	BB McCullum	P Kumar	0

```
In [10]: df_matches.shape
```

```
Out[10]: (1095, 20)
```

```
In [11]: df_deliveries.shape
```

```
Out[11]: (260920, 17)
```

```
In [12]: # Info
```

```
In [13]: df_matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    1095 non-null   int64
1   season                1095 non-null   object
2   city                  1044 non-null   object
3   date                  1095 non-null   object
4   match_type            1095 non-null   object
5   player_of_match       1090 non-null   object
6   venue                 1095 non-null   object
7   team1                 1095 non-null   object
8   team2                 1095 non-null   object
9   toss_winner           1095 non-null   object
10  toss_decision         1095 non-null   object
11  winner                1090 non-null   object
12  result                1095 non-null   object
13  result_margin         1076 non-null   float64
14  target_runs           1092 non-null   float64
15  target_overs          1092 non-null   float64
16  super_over            1095 non-null   object
17  method                21 non-null     object
18  umpire1               1095 non-null   object
19  umpire2               1095 non-null   object
dtypes: float64(3), int64(1), object(16)
memory usage: 171.2+ KB
```

```
In [14]: # columns
df_matches.columns
```

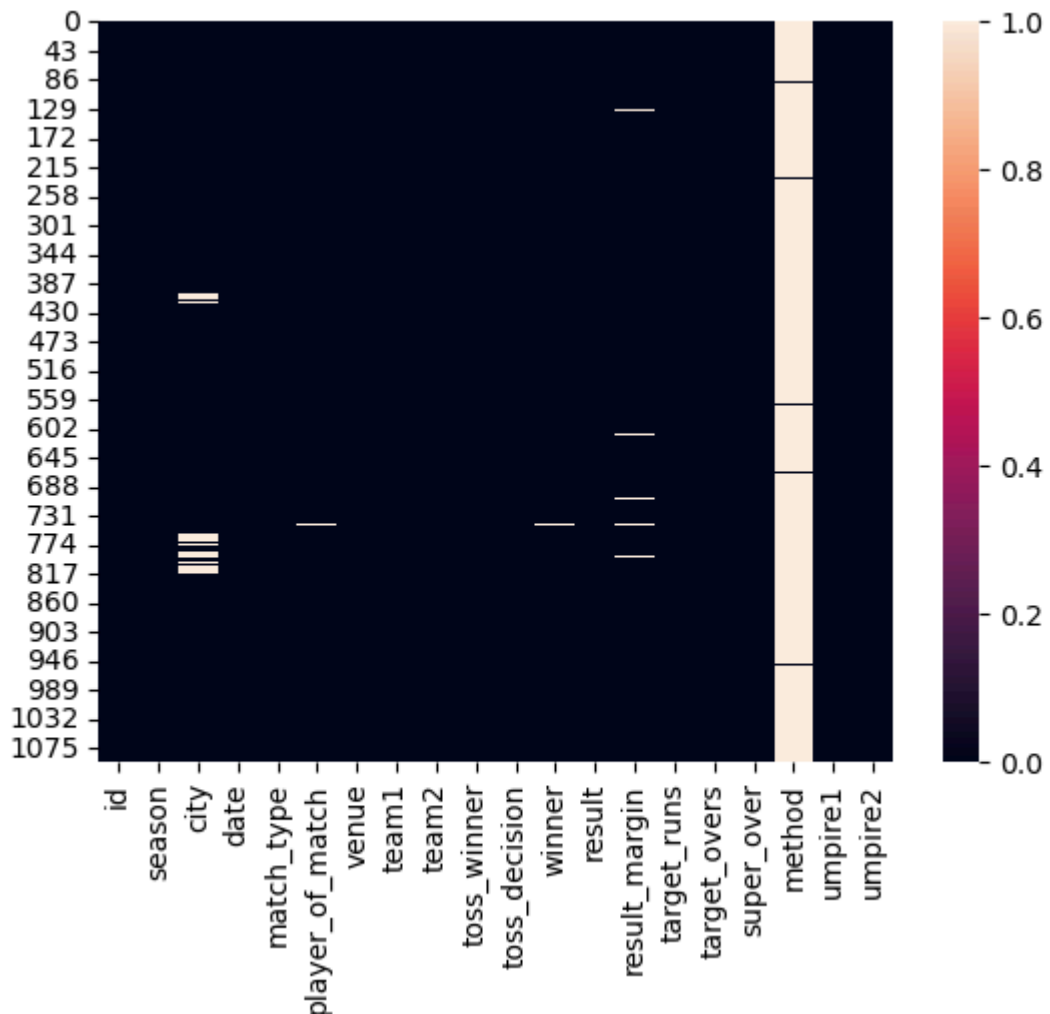
```
Out[14]: Index(['id', 'season', 'city', 'date', 'match_type', 'player_of_match',
               'venue', 'team1', 'team2', 'toss_winner', 'toss_decision', 'winne
               r',
               'result', 'result_margin', 'target_runs', 'target_overs', 'super_
               over',
               'method', 'umpire1', 'umpire2'],
              dtype='object')
```

```
In [15]: # Nullvalues count
df_matches.isna().sum().sort_values(ascending = False)
```

```
Out[15]: method          1074
city                    51
result_margin          19
winner                   5
player_of_match         5
target_overs            3
target_runs             3
id                       0
umpire1                  0
super_over              0
result                   0
toss_decision            0
season                   0
toss_winner              0
team2                    0
team1                    0
venue                    0
match_type               0
date                     0
umpire2                  0
dtype: int64
```

```
In [16]: # Heatmap

sns.heatmap(df_matches.isna())
plt.show()
```



```
In [17]: # Nullvalues count
df_matches.isna().mean().sort_values(ascending = False)*100
```

```
Out[17]: method          98.082192
city             4.657534
result_margin    1.735160
winner           0.456621
player_of_match  0.456621
target_overs     0.273973
target_runs      0.273973
id               0.000000
umpire1          0.000000
super_over       0.000000
result           0.000000
toss_decision    0.000000
season           0.000000
toss_winner      0.000000
team2            0.000000
team1            0.000000
venue            0.000000
match_type       0.000000
date             0.000000
umpire2          0.000000
dtype: float64
```

```
In [18]: df_matches['method'].value_counts()
```

```
Out[18]: method
D/L      21
Name: count, dtype: int64
```

```
In [19]: df_matches['method'].fillna('Normal method',inplace = True)
print('Done')
```

Done

/var/folders/y0/tq9y6qys5zx8m8_x74cgbd040000gn/T/ipykernel_51351/3682990747.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method. The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df_matches['method'].fillna('Normal method',inplace = True)
```

```
In [20]: df_matches['city'].value_counts()
```

```
Out[20]: city
Mumbai      173
Kolkata      93
Delhi        90
Chennai      85
Hyderabad    77
Bangalore    65
Chandigarh   61
Jaipur       57
Pune         51
Abu Dhabi    37
Ahmedabad    36
Bengaluru    29
Durban       15
Visakhapatnam 15
Lucknow      14
Dubai        13
Dharamsala   13
Centurion    12
Rajkot       10
Sharjah      10
Indore       9
Navi Mumbai  9
Johannesburg 8
Cuttack      7
Port Elizabeth 7
Ranchi       7
Cape Town    7
Raipur       6
Mohali       5
Kochi        5
Kanpur       4
Kimberley    3
East London  3
Guwahati     3
Nagpur       3
Bloemfontein 2
Name: count, dtype: int64
```

```
In [21]: # masking
df_matches[df_matches['city'].isna()]
```

Out [21]:

	id	season	city	date	match_type	player_of_match	venue	
399	729281	2014	NaN	2014-04-17	League	YS Chahal	Sharjah Cricket Stadium	I
402	729287	2014	NaN	2014-04-19	League	PA Patel	Dubai International Cricket Stadium	Cl
403	729289	2014	NaN	2014-04-19	League	JP Duminy	Dubai International Cricket Stadium	
404	729291	2014	NaN	2014-04-20	League	GJ Maxwell	Sharjah Cricket Stadium	
406	729295	2014	NaN	2014-04-22	League	GJ Maxwell	Sharjah Cricket Stadium	
407	729297	2014	NaN	2014-04-23	League	RA Jadeja	Dubai International Cricket Stadium	
408	729299	2014	NaN	2014-04-24	League	CA Lynn	Sharjah Cricket Stadium	Cl
409	729301	2014	NaN	2014-04-25	League	AJ Finch	Dubai International Cricket Stadium	F
410	729303	2014	NaN	2014-04-25	League	MM Sharma	Dubai International Cricket Stadium	
413	729309	2014	NaN	2014-04-27	League	M Vijay	Sharjah Cricket Stadium	I
414	729311	2014	NaN	2014-04-27	League	DR Smith	Sharjah Cricket Stadium	F
415	729313	2014	NaN	2014-04-28	League	Sandeep Sharma	Dubai International Cricket Stadium	
417	729317	2014	NaN	2014-04-30	League	B Kumar	Dubai International Cricket Stadium	
757	1216493	2020/21	NaN	2020-09-20	League	MP Stoinis	Dubai International Cricket Stadium	

	id	season	city	date	match_type	player_of_match	venue	
758	1216534	2020/21	NaN	2020-09-21	League	YS Chahal	Dubai International Cricket Stadium	Cl
759	1216496	2020/21	NaN	2020-09-22	League	SV Samson	Sharjah Cricket Stadium	
761	1216510	2020/21	NaN	2020-09-24	League	KL Rahul	Dubai International Cricket Stadium	
762	1216539	2020/21	NaN	2020-09-25	League	PP Shaw	Dubai International Cricket Stadium	
764	1216527	2020/21	NaN	2020-09-27	League	SV Samson	Sharjah Cricket Stadium	
765	1216547	2020/21	NaN	2020-09-28	League	AB de Villiers	Dubai International Cricket Stadium	Cl
767	1216504	2020/21	NaN	2020-09-30	League	Shivam Mavi	Dubai International Cricket Stadium	
769	1216516	2020/21	NaN	2020-10-02	League	PK Garg	Dubai International Cricket Stadium	H
771	1216515	2020/21	NaN	2020-10-03	League	SS Iyer	Sharjah Cricket Stadium	
772	1216513	2020/21	NaN	2020-10-04	League	SR Watson	Dubai International Cricket Stadium	
773	1216538	2020/21	NaN	2020-10-04	League	TA Boult	Sharjah Cricket Stadium	
774	1216519	2020/21	NaN	2020-10-05	League	AR Patel	Dubai International Cricket Stadium	
777	1216542	2020/21	NaN	2020-10-08	League	JM Bairstow	Dubai International Cricket Stadium	H

	id	season	city	date	match_type	player_of_match	venue	
778	1216500	2020/21	NaN	2020-10-09	League	R Ashwin	Sharjah Cricket Stadium	
780	1216525	2020/21	NaN	2020-10-10	League	V Kohli	Dubai International Cricket Stadium	CI
781	1216507	2020/21	NaN	2020-10-11	League	R Tewatia	Dubai International Cricket Stadium	H
783	1216540	2020/21	NaN	2020-10-12	League	AB de Villiers	Sharjah Cricket Stadium	CI
784	1216528	2020/21	NaN	2020-10-13	League	RA Jadeja	Dubai International Cricket Stadium	
785	1216543	2020/21	NaN	2020-10-14	League	A Nortje	Dubai International Cricket Stadium	
786	1216531	2020/21	NaN	2020-10-15	League	KL Rahul	Sharjah Cricket Stadium	CI
788	1216509	2020/21	NaN	2020-10-17	League	S Dhawan	Sharjah Cricket Stadium	
789	1216522	2020/21	NaN	2020-10-17	League	AB de Villiers	Dubai International Cricket Stadium	
791	1216517	2020/21	NaN	2020-10-18	League	KL Rahul	Dubai International Cricket Stadium	
793	1216546	2020/21	NaN	2020-10-20	League	S Dhawan	Dubai International Cricket Stadium	
795	1216518	2020/21	NaN	2020-10-22	League	MK Pandey	Dubai International Cricket Stadium	
796	1216521	2020/21	NaN	2020-10-23	League	TA Boult	Sharjah Cricket Stadium	

	id	season	city	date	match_type	player_of_match	venue	
798	1216498	2020/21	NaN	2020-10-24	League	CJ Jordan	Dubai International Cricket Stadium	
800	1216544	2020/21	NaN	2020-10-25	League	RD Gaikwad	Dubai International Cricket Stadium	Cl
801	1216520	2020/21	NaN	2020-10-26	League	CH Gayle	Sharjah Cricket Stadium	
802	1216524	2020/21	NaN	2020-10-27	League	WP Saha	Dubai International Cricket Stadium	H
804	1216536	2020/21	NaN	2020-10-29	League	RD Gaikwad	Dubai International Cricket Stadium	
806	1216502	2020/21	NaN	2020-10-31	League	Sandeep Sharma	Sharjah Cricket Stadium	Cl
807	1216535	2020/21	NaN	2020-10-31	League	Ishan Kishan	Dubai International Cricket Stadium	
809	1216530	2020/21	NaN	2020-11-01	League	PJ Cummins	Dubai International Cricket Stadium	
811	1216495	2020/21	NaN	2020-11-03	League	S Nadeem	Sharjah Cricket Stadium	
812	1237177	2020/21	NaN	2020-11-05	Qualifier 1	JJ Bumrah	Dubai International Cricket Stadium	
815	1237181	2020/21	NaN	2020-11-10	Final	TA Boult	Dubai International Cricket Stadium	

```
In [22]: df_matches['city'].fillna('UAE', inplace = True)
print('done')
```

done

/var/folders/y0/tq9y6qys5zx8m8_x74cgb040000gn/T/ipykernel_51351/2013551467.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method. The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always has a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df_matches['city'].fillna('UAE', inplace = True)
```

```
In [23]: # Nullvalues count
df_matches.isna().mean().sort_values(ascending = False)*100
```

```
Out[23]: result_margin      1.735160
winner      0.456621
player_of_match  0.456621
target_overs  0.273973
target_runs  0.273973
id      0.000000
umpire1      0.000000
method      0.000000
super_over  0.000000
result      0.000000
toss_decision  0.000000
season      0.000000
toss_winner  0.000000
team2      0.000000
team1      0.000000
venue      0.000000
match_type  0.000000
date      0.000000
city      0.000000
umpire2      0.000000
dtype: float64
```

```
In [24]: df_matches[df_matches['result_margin'].isna()]
```

Out [24]:

	id	season	city	date	match_type	player_of_match	
66	392190	2009	Cape Town	2009-04-23	League	YK Pathan	Ne
130	419121	2009/10	Chennai	2010-03-21	League	J Theron	Chidar S C
241	501265	2011	Delhi	2011-05-21	League	NaN	Fer
328	598004	2013	Hyderabad	2013-04-07	League	GH Vihari	Rajiv Inter Stadium
342	598017	2013	Bangalore	2013-04-16	League	V Kohli	Chinna s
416	729315	2014	Abu Dhabi	2014-04-29	League	JP Faulkner	Sheikl s
475	829741	2015	Ahmedabad	2015-04-21	League	SE Marsh	Sard S
485	829763	2015	Bangalore	2015-04-29	League	NaN	Chinna s
511	829813	2015	Bangalore	2015-05-17	League	NaN	Chinna s
610	1082625	2017	Rajkot	2017-04-29	League	KH Pandya	Sat Ass s
705	1175365	2019	Delhi	2019-03-30	League	PP Shaw	Arur s
744	1178424	2019	Bengaluru	2019-04-30	League	NaN	M.Chinna s
746	1178426	2019	Mumbai	2019-05-02	League	JJ Bumrah	We s
757	1216493	2020/21	UAE	2020-09-20	League	MP Stoinis	Inter Cricket s
765	1216547	2020/21	UAE	2020-09-28	League	AB de Villiers	Inter Cricket s
790	1216512	2020/21	Abu Dhabi	2020-10-18	League	LH Ferguson	Sheikl s

	id	season	city	date	match_type	player_of_match	
791	1216517	2020/21	UAE	2020-10-18	League	KL Rahul	Inter Cricket S
835	1254077	2021	Chennai	2021-04-25	League	PP Shaw	Chidambaram S C
994	1359519	2023	Lucknow	2023-05-03	League	NaN	Bharat Shri At Vajpaye

```
In [25]: df_matches['result_margin'].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 1095 entries, 0 to 1094
Series name: result_margin
Non-Null Count  Dtype
-----  -----
1076 non-null    float64
dtypes: float64(1)
memory usage: 8.7 KB
```

```
In [26]: df_matches['result_margin'].fillna(0, inplace = True)
print('done')
```

done

/var/folders/y0/tq9y6qys5zx8m8_x74cgb040000gn/T/ipykernel_51351/3158733585.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.

The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df_matches['result_margin'].fillna(0, inplace = True)
```

```
In [27]: # Columns with Missing values
# winner, player_of_match, target_overs, target_runs
```

```
In [28]: df_matches[df_matches['winner'].isna()]
```

Out [28]:

	id	season	city	date	match_type	player_of_match	venue
241	501265	2011	Delhi	2011-05-21	League	NaN	Feroz : Shah
485	829763	2015	Bangalore	2015-04-29	League	NaN	Chinnasv Sta
511	829813	2015	Bangalore	2015-05-17	League	NaN	Chinnasv Sta
744	1178424	2019	Bengaluru	2019-04-30	League	NaN	M.Chinnasv Sta
994	1359519	2023	Lucknow	2023-05-03	League	NaN	Bharat F Shri Atal B Vajpayee E

```
In [29]: df_matches.fillna({'player_of_match': 'No Player of the match',
                             'winner': 'No Winner'}, inplace = True)
print('done')
```

done

```
In [30]: df_matches.fillna(0, inplace = True)
print('Done')
```

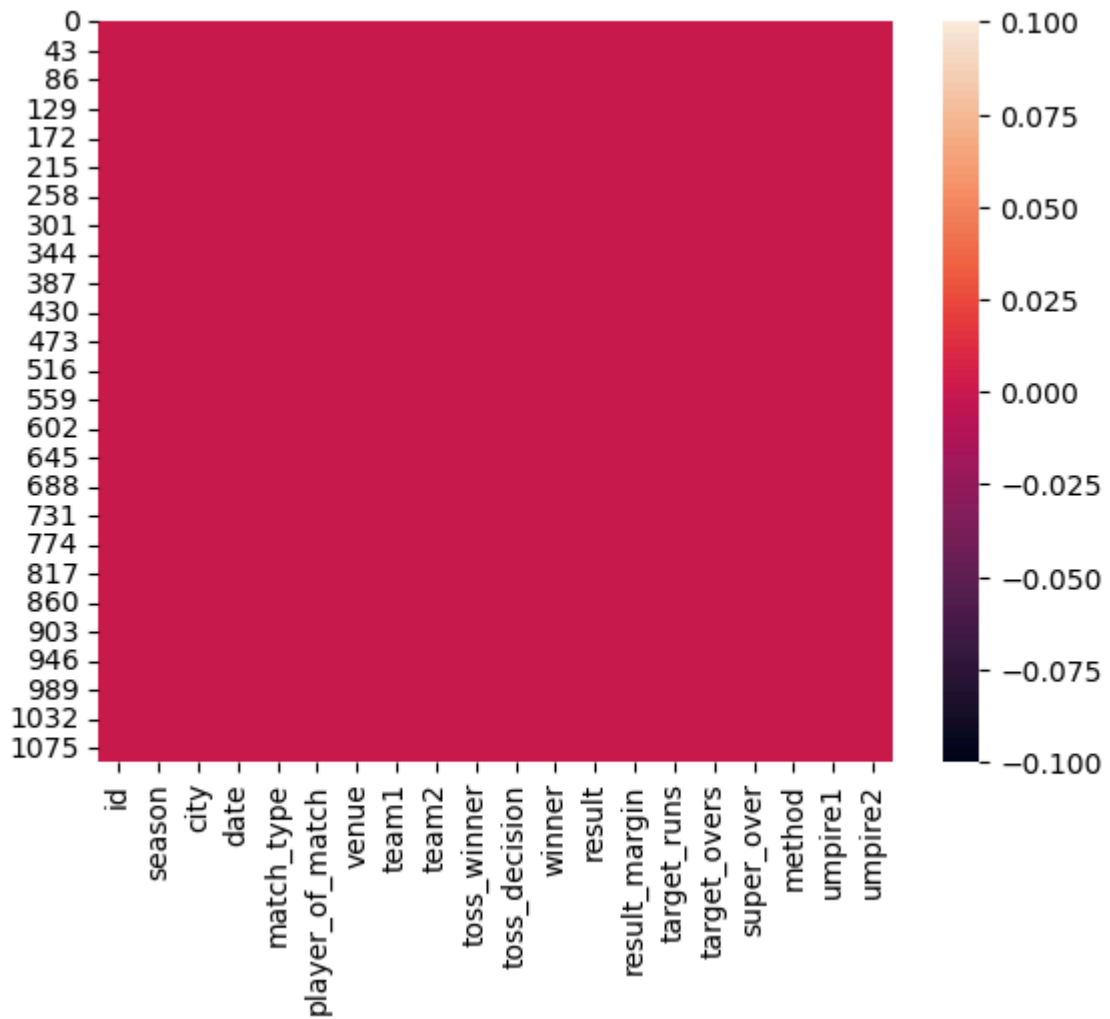
Done

```
In [31]: df_matches.isna().sum()
```

```
Out[31]: id                0
season                0
city                 0
date                 0
match_type           0
player_of_match       0
venue                0
team1                0
team2                0
toss_winner           0
toss_decision         0
winner               0
result               0
result_margin         0
target_runs           0
target_overs          0
super_over            0
method               0
umpire1              0
umpire2              0
dtype: int64
```

```
In [32]: sns.heatmap(df_matches.isna())
```

Out[32]: <Axes: >



```
In [33]: # replace
```

```
In [34]: df_matches['city'].value_counts()
```

```
Out[34]: city
Mumbai      173
Kolkata      93
Delhi        90
Chennai      85
Hyderabad    77
Bangalore    65
Chandigarh   61
Jaipur       57
UAE          51
Pune         51
Abu Dhabi    37
Ahmedabad    36
Bengaluru    29
Durban       15
Visakhapatnam 15
Lucknow      14
Dharamsala   13
Dubai        13
Centurion    12
Rajkot       10
Sharjah      10
Indore       9
Navi Mumbai  9
Johannesburg 8
Cuttack      7
Ranchi       7
Port Elizabeth 7
Cape Town    7
Raipur       6
Mohali       5
Kochi        5
Kanpur       4
Kimberley    3
East London  3
Guwahati     3
Nagpur       3
Bloemfontein 2
Name: count, dtype: int64
```

```
In [35]: data_rp = {'Navi Mumbai': 'Mumbai',
                    'Bengaluru': 'Bangalore',
                    'Sharjah': 'UAE',
                    'Abu Dhabi': 'UAE',
                    'Kimberley': 'Cape Town',
                    'Dubai': 'UAE'
                    }

df_matches.replace(data_rp, inplace = True)
print('done')
```

done

```
In [36]: # check Duplicates
```

```
In [37]: df_matches.duplicated('id').value_counts()
```

```
Out[37]: False    1095
Name: count, dtype: int64
```



```
In [38]: df_matches.drop_duplicates('id', inplace = True)
print('Done')
```

Done

```
In [39]: df_matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    1095 non-null   int64
1   season                1095 non-null   object
2   city                  1095 non-null   object
3   date                  1095 non-null   object
4   match_type            1095 non-null   object
5   player_of_match       1095 non-null   object
6   venue                 1095 non-null   object
7   team1                 1095 non-null   object
8   team2                 1095 non-null   object
9   toss_winner           1095 non-null   object
10  toss_decision         1095 non-null   object
11  winner                 1095 non-null   object
12  result                 1095 non-null   object
13  result_margin         1095 non-null   float64
14  target_runs           1095 non-null   float64
15  target_overs          1095 non-null   float64
16  super_over            1095 non-null   object
17  method                1095 non-null   object
18  umpire1               1095 non-null   object
19  umpire2               1095 non-null   object
dtypes: float64(3), int64(1), object(16)
memory usage: 171.2+ KB
```

```
In [40]: # Date column must be in date datatype
```

```
In [41]: df_matches['date'] = pd.to_datetime(df_matches['date'])
print('Done')
```

Done

```
In [42]: df_matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    1095 non-null   int64
1   season                1095 non-null   object
2   city                  1095 non-null   object
3   date                  1095 non-null   datetime64[ns]
4   match_type            1095 non-null   object
5   player_of_match       1095 non-null   object
6   venue                 1095 non-null   object
7   team1                 1095 non-null   object
8   team2                 1095 non-null   object
9   toss_winner           1095 non-null   object
10  toss_decision         1095 non-null   object
11  winner                1095 non-null   object
12  result                1095 non-null   object
13  result_margin         1095 non-null   float64
14  target_runs           1095 non-null   float64
15  target_overs          1095 non-null   float64
16  super_over            1095 non-null   object
17  method                1095 non-null   object
18  umpire1               1095 non-null   object
19  umpire2               1095 non-null   object
dtypes: datetime64[ns](1), float64(3), int64(1), object(15)
memory usage: 171.2+ KB
```

```
In [43]: df_matches['venue'].value_counts()
```

```
Out[43]: venue
Eden Gardens
77
Wankhede Stadium
73
M Chinnaswamy Stadium
65
Feroz Shah Kotla
60
Rajiv Gandhi International Stadium, Uppal
49
MA Chidambaram Stadium, Chepauk
48
Sawai Mansingh Stadium
47
Dubai International Cricket Stadium
46
Wankhede Stadium, Mumbai
45
Punjab Cricket Association Stadium, Mohali
35
Sheikh Zayed Stadium
29
Sharjah Cricket Stadium
28
MA Chidambaram Stadium, Chepauk, Chennai
28
Narendra Modi Stadium, Ahmedabad
24
Maharashtra Cricket Association Stadium
22
Dr DY Patil Sports Academy, Mumbai
20
Brabourne Stadium, Mumbai
17
Dr DY Patil Sports Academy
17
Eden Gardens, Kolkata
16
Subrata Roy Sahara Stadium
16
Arun Jaitley Stadium, Delhi
16
Rajiv Gandhi International Stadium
15
M.Chinnaswamy Stadium
15
Kingsmead
15
Arun Jaitley Stadium
14
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium, Lucknow
14
M Chinnaswamy Stadium, Bengaluru
14
Rajiv Gandhi International Stadium, Uppal, Hyderabad
13
Maharashtra Cricket Association Stadium, Pune
13
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium
```

```

13
SuperSport Park
12
Sardar Patel Stadium, Motera
12
Punjab Cricket Association IS Bindra Stadium, Mohali
11
Sawai Mansingh Stadium, Jaipur
10
Saurashtra Cricket Association Stadium
10
Brabourne Stadium
10
Punjab Cricket Association IS Bindra Stadium
10
Himachal Pradesh Cricket Association Stadium
9
MA Chidambaram Stadium
9
Holkar Cricket Stadium
9
New Wanderers Stadium
8
Zayed Cricket Stadium, Abu Dhabi
8
JSCA International Stadium Complex
7
Barabati Stadium
7
St George's Park
7
Newlands
7
Shaheed Veer Narayan Singh International Stadium
6
Nehru Stadium
5
Punjab Cricket Association IS Bindra Stadium, Mohali, Chandigarh
5
Maharaja Yadavindra Singh International Cricket Stadium, Mullanpur
5
Green Park
4
Himachal Pradesh Cricket Association Stadium, Dharamsala
4
Vidarbha Cricket Association Stadium, Jamtha
3
De Beers Diamond Oval
3
Buffalo Park
3
Barsapara Cricket Stadium, Guwahati
3
OUTsurance Oval
2
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium, Visakhapatnam
2
Name: count, dtype: int64

```

```
In [44]: df_matches['venue'].str.split(',')
```

```
Out[44]: 0 [M Chinnaswamy Stadium]
1 [Punjab Cricket Association Stadium, Mohali]
2 [Feroz Shah Kotla]
3 [Wankhede Stadium]
4 [Eden Gardens]

...
1090 [Rajiv Gandhi International Stadium, Uppal, ...]
1091 [Narendra Modi Stadium, Ahmedabad]
1092 [Narendra Modi Stadium, Ahmedabad]
1093 [MA Chidambaram Stadium, Chepauk, Chennai]
1094 [MA Chidambaram Stadium, Chepauk, Chennai]
Name: venue, Length: 1095, dtype: object
```

```
In [45]: df_matches['venue'] = df_matches['venue'].apply(lambda row: row.title().split(' '), axis=1)
print('done')
```

done

```
In [46]: df_matches['venue'].value_counts()
```

```
Out[46]: venue
Wankhede Stadium 118
Eden Gardens 93
Ma Chidambaram Stadium 85
M Chinnaswamy Stadium 79
Rajiv Gandhi International Stadium 77
Feroz Shah Kotla 60
Sawai Mansingh Stadium 57
Dubai International Cricket Stadium 46
Dr Dy Patil Sports Academy 37
Maharashtra Cricket Association Stadium 35
Punjab Cricket Association Stadium 35
Arun Jaitley Stadium 30
Sheikh Zayed Stadium 29
Sharjah Cricket Stadium 28
Brabourne Stadium 27
Punjab Cricket Association Is Bindra Stadium 26
Narendra Modi Stadium 24
Subrata Roy Sahara Stadium 16
Dr. Y.S. Rajasekhara Reddy Aca-Vdca Cricket Stadium 15
M.Chinnaswamy Stadium 15
Kingsmead 15
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium 14
Himachal Pradesh Cricket Association Stadium 13
Sardar Patel Stadium 12
Supersport Park 12
Saurashtra Cricket Association Stadium 10
Holkar Cricket Stadium 9
Zayed Cricket Stadium 8
New Wanderers Stadium 8
Jscs International Stadium Complex 7
Barabati Stadium 7
St George'S Park 7
Newlands 7
Shaheed Veer Narayan Singh International Stadium 6
Nehru Stadium 5
Maharaja Yadavindra Singh International Cricket Stadium 5
Green Park 4
De Beers Diamond Oval 3
Buffalo Park 3
Vidarbha Cricket Association Stadium 3
Barsapara Cricket Stadium 3
Outsurance Oval 2
Name: count, dtype: int64
```

```
In [47]: data_v_rp = {'M.Chinnaswamy Stadium':'M Chinnaswamy Stadium',
                    'Feroz Shah Kotla':'Arun Jaitley Stadium',
                    'Sardar Patel Stadium':'Narendra Modi Stadium',
                    'Punjab Cricket Association Is Bindra Stadium':'Punjab Crick

df_matches['venue'].replace(data_v_rp,inplace = True)
print('Done')
```

Done

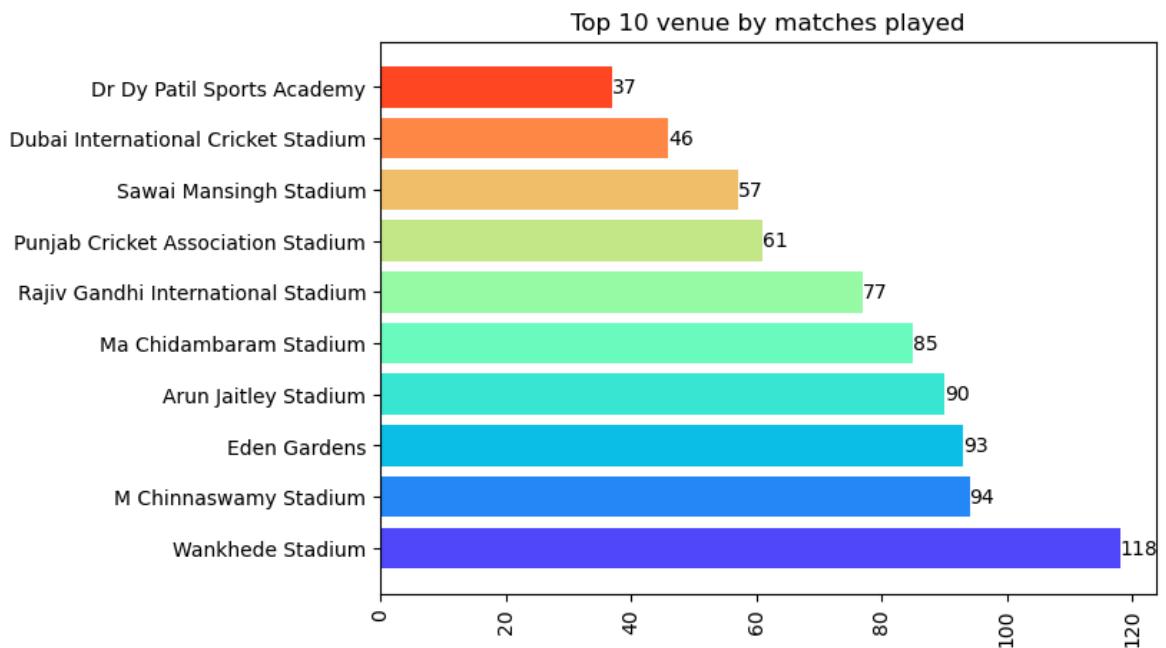
/var/folders/y0/tq9y6qys5zx8m8_x74cgbd040000gn/T/ipykernel_51351/2614453559.py:6: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method. The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always has a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df_matches['venue'].replace(data_v_rp,inplace = True)
```

```
In [48]: x = df_matches['venue'].value_counts().head(10).index
        y = df_matches['venue'].value_counts().head(10).values

        plt.figure(figsize=(7,5))
        ax = plt.barh(x,y,color = sns.color_palette('rainbow',10))
        plt.bar_label(ax)
        plt.title("Top 10 venue by matches played")
        plt.xticks(rotation = 90)
        plt.show()
```



```
In [49]: df_matches['team1'].value_counts()
```

```
Out[49]: 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
Rising Pune Supergiant          7
Rising Pune Supergiants         7
Kochi Tuskers Kerala            7
Name: count, dtype: int64
```

```
In [50]: df_matches['team2'].value_counts()
```

```
Out[50]: team2
Mumbai Indians                138
Kolkata Knight Riders          130
Rajasthan Royals              120
Chennai Super Kings           110
Royal Challengers Bangalore    105
Kings XI Punjab               98
Sunrisers Hyderabad           96
Delhi Daredevils              76
Delhi Capitals                 50
Deccan Chargers               36
Punjab Kings                  25
Gujarat Titans                 24
Pune Warriors                  23
Lucknow Super Giants           21
Gujarat Lions                  14
Rising Pune Supergiant         9
Rising Pune Supergiants        7
Kochi Tuskers Kerala           7
Royal Challengers Bengaluru     6
Name: count, dtype: int64
```

```
In [51]: import warnings
warnings.filterwarnings('ignore')

team_name_rp = { 'Royal Challengers Bangalore': 'Royal Challengers Bengaluru',
                  'Delhi Daredevils': 'Delhi Capitals',
                  'Kings XI Punjab': 'Punjab Kings',
                  'Rising Pune Supergiant': 'Rising Pune Supergiants',
                  'Pune Warriors': 'Rising Pune Supergiants' }
df_matches[['team1', 'team2']] = df_matches[['team1', 'team2']].replace(team_name_rp)
print('Done')
```

Done

```
In [52]: df_matches
```


Out [52]:

	id	season	city	date	match_type	player_of_match	
0	335982	2007/08	Bangalore	2008-04-18	League	BB McCullum	Chinna S
1	335983	2007/08	Chandigarh	2008-04-19	League	MEK Hussey	Asso S
2	335984	2007/08	Delhi	2008-04-19	League	MF Maharoo	Arun S
3	335985	2007/08	Mumbai	2008-04-20	League	MV Boucher	War S
4	335986	2007/08	Kolkata	2008-04-20	League	DJ Hussey	G
...
1090	1426307	2024	Hyderabad	2024-05-19	League	Abhishek Sharma	Rajiv Intern S
1091	1426309	2024	Ahmedabad	2024-05-21	Qualifier 1	MA Starc	Na Modi S
1092	1426310	2024	Ahmedabad	2024-05-22	Eliminator	R Ashwin	Na Modi S
1093	1426311	2024	Chennai	2024-05-24	Qualifier 2	Shahbaz Ahmed	Chidam S
1094	1426312	2024	Chennai	2024-05-26	Final	MA Starc	Chidam S

1095 rows × 20 columns

```
In [53]: team_name_rp = { 'Royal Challengers Bangalore':'Royal Challengers Bengaluru',
                        'Delhi Daredevils':'Delhi Capitals',
                        'Kings XI Punjab':'Punjab Kings',
                        'Rising Pune Supergiant':'Rising Pune Supergiants',
                        'Pune Warriors':'Rising Pune Supergiants'}
df_matches['toss_winner'].replace(team_name_rp,inplace = True)
df_matches['winner'].replace(team_name_rp,inplace = True)

print('Done')
```

Done

```
In [54]: # umpire1 and umpire2
```

```
In [55]: df_matches['umpire1'].value_counts()
```

```
Out [55]: umpire1
          AK Chaudhary      115
          HDPK Dharmasena   79
          KN Ananthapadmanabhan 59
          CB Gaffaney       53
          Asad Rauf         51
          ...
          AV Jayaprakash     2
          K Srinath          2
          AG Wharf           2
          SL Shastri         1
          SJA Taufel         1
          Name: count, Length: 62, dtype: int64
```

```
In [56]: # Drop #umpire1 and umpire2
```

```
In [57]: df_matches.drop(['umpire1','umpire2'],axis =1,inplace =True)
          print('Done')
```

Done

```
In [58]: df_matches.sample(10)
```

Out [58]:

	id	season	city	date	match_type	player_of_match	venue
957	1359482	2023	Guwahati	2023-04-05	League	NT Ellis	Barsap Cric Stadi
539	980945	2016	Delhi	2016-04-27	League	CH Morris	Arun Jait Stadi
677	1136602	2018	Delhi	2018-05-10	League	S Dhawan	Arun Jait Stadi
597	1082611	2017	Hyderabad	2017-04-19	League	KS Williamson	Rajiv Gan Internatio Stadi
1061	1426276	2024	Jaipur	2024-04-22	League	Sandeep Sharma	Sa' Mansir Stadi
381	598057	2013	Ranchi	2013-05-12	League	JH Kallis	J's Internatio Stadi Comp
645	1136570	2018	Kolkata	2018-04-14	League	B Stanlake	Ec Garde
521	980909	2016	Kolkata	2016-04-13	League	RG Sharma	Ec Garde
705	1175365	2019	Delhi	2019-03-30	League	PP Shaw	Arun Jait Stadi
458	829705	2015	Kolkata	2015-04-08	League	M Morkel	Ec Garde

In [59]: `df_matches.columns`

Out [59]: Index(['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'], dtype='object')

In [60]: `df_matches.shape`

Out [60]: (1095, 18)

In [61]: `df_matches['winner'].value_counts()`

```
Out[61]: winner
Mumbai Indians      144
Chennai Super Kings 138
Kolkata Knight Riders 131
Royal Challengers Bengaluru 123
Delhi Capitals      115
Rajasthan Royals    112
Punjab Kings        112
Sunrisers Hyderabad  88
Deccan Chargers     29
Gujarat Titans      28
Rising Pune Supergiants 27
Lucknow Super Giants 24
Gujarat Lions       13
Kochi Tuskers Kerala  6
No Winner           5
Name: count, dtype: int64
```

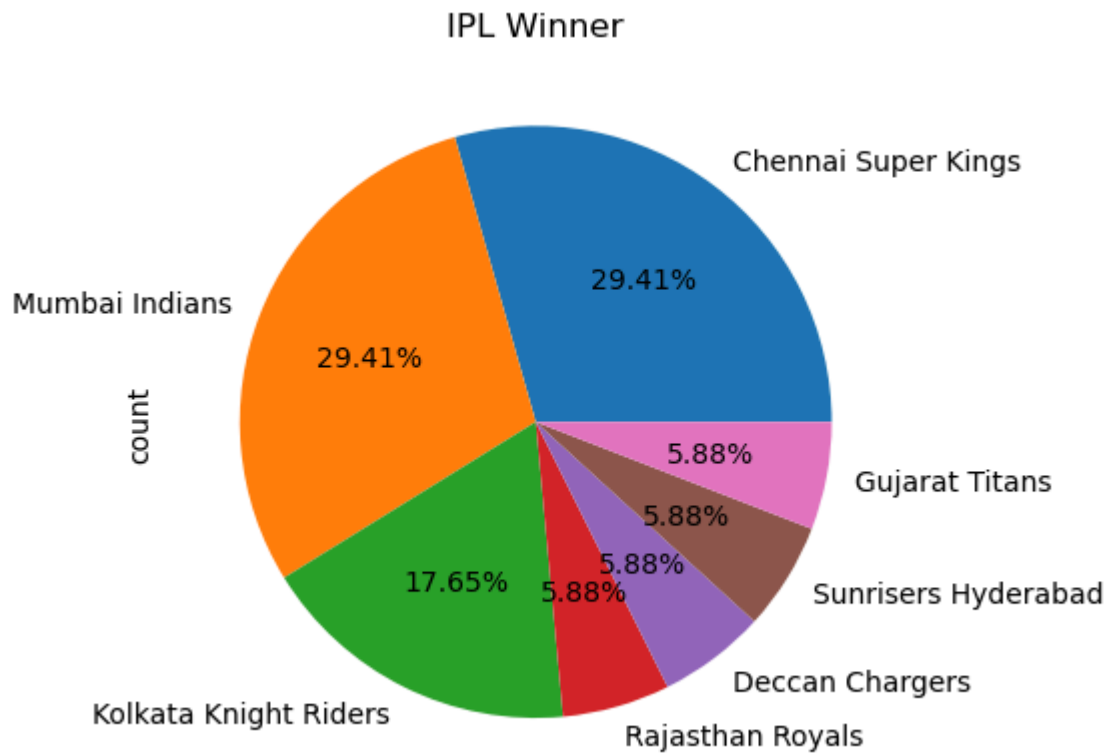
```
In [62]: df_matches['match_type'].value_counts()
```

```
Out[62]: match_type
League      1029
Final       17
Qualifier 1  14
Qualifier 2  14
Eliminator  11
Semi Final   6
Elimination Final 3
3rd Place Play-Off 1
Name: count, dtype: int64
```

```
In [63]: # List Ipl winner
df_matches[df_matches['match_type'] == 'Final']['winner'].value_counts()
```

```
Out[63]: winner
Chennai Super Kings    5
Mumbai Indians         5
Kolkata Knight Riders  3
Rajasthan Royals       1
Deccan Chargers        1
Sunrisers Hyderabad    1
Gujarat Titans         1
Name: count, dtype: int64
```

```
In [64]: plt.title("IPL Winner")
df_matches[df_matches['match_type'] == 'Final']['winner'].value_counts().
plt.show()
```



```
In [65]: df_matches['date'].dt.year
```

```
Out[65]: 0      2008
          1      2008
          2      2008
          3      2008
          4      2008
          ...
        1090    2024
        1091    2024
        1092    2024
        1093    2024
        1094    2024
        Name: date, Length: 1095, dtype: int32
```

```
In [66]: Final_matche = df_matches[df_matches['match_type'] == 'Final']
          Final_matche
```

Out [66]:

	id	season	city	date	match_type	player_of_match	
57	336040	2007/08	Mumbai	2008-06-01	Final	YK Pathan	D
114	392239	2009	Johannesburg	2009-05-24	Final	A Kumble	W
174	419165	2009/10	Mumbai	2010-04-25	Final	SK Raina	D
247	501271	2011	Chennai	2011-05-28	Final	M Vijay	Chidi
321	548381	2012	Chennai	2012-05-27	Final	MS Bisla	Chidi
397	598073	2013	Kolkata	2013-05-26	Final	KA Pollard	
457	734049	2014	Bangalore	2014-06-01	Final	MK Pandey	Chin
516	829823	2015	Kolkata	2015-05-24	Final	RG Sharma	
576	981019	2016	Bangalore	2016-05-29	Final	BCJ Cutting	Chin
635	1082650	2017	Hyderabad	2017-05-21	Final	KH Pandya	Raji Inte
695	1136620	2018	Mumbai	2018-05-27	Final	SR Watson	W
755	1181768	2019	Hyderabad	2019-05-12	Final	JJ Bumrah	Raji Inte
815	1237181	2020/21	UAE	2020-11-10	Final	TA Boult	Inte
875	1254117	2021	UAE	2021-10-15	Final	F du Plessis	Inte
949	1312200	2022	Ahmedabad	2022-05-29	Final	HH Pandya	Modi
1023	1370353	2023	Ahmedabad	2023-05-29	Final	DP Conway	Modi

	id	season	city	date	match_type	player_of_match	
1094	1426312	2024	Chennai	2024-05-26	Final	MA Starc	Chidi

```
In [67]: Final_matche['Winning year'] = Final_matche['date'].dt.year
print('Done')
```

Done

```
In [68]: Final_matche
```

Out [68]:

	id	season	city	date	match_type	player_of_match	
57	336040	2007/08	Mumbai	2008-06-01	Final	YK Pathan	D
114	392239	2009	Johannesburg	2009-05-24	Final	A Kumble	W
174	419165	2009/10	Mumbai	2010-04-25	Final	SK Raina	D
247	501271	2011	Chennai	2011-05-28	Final	M Vijay	Chidi
321	548381	2012	Chennai	2012-05-27	Final	MS Bisla	Chidi
397	598073	2013	Kolkata	2013-05-26	Final	KA Pollard	
457	734049	2014	Bangalore	2014-06-01	Final	MK Pandey	Chin
516	829823	2015	Kolkata	2015-05-24	Final	RG Sharma	
576	981019	2016	Bangalore	2016-05-29	Final	BCJ Cutting	Chin
635	1082650	2017	Hyderabad	2017-05-21	Final	KH Pandya	Raji Inte
695	1136620	2018	Mumbai	2018-05-27	Final	SR Watson	W
755	1181768	2019	Hyderabad	2019-05-12	Final	JJ Bumrah	Raji Inte
815	1237181	2020/21	UAE	2020-11-10	Final	TA Boult	Inte
875	1254117	2021	UAE	2021-10-15	Final	F du Plessis	Inte
949	1312200	2022	Ahmedabad	2022-05-29	Final	HH Pandya	Modi
1023	1370353	2023	Ahmedabad	2023-05-29	Final	DP Conway	Modi

	id	season	city	date	match_type	player_of_match
1094	1426312	2024	Chennai	2024-05-26	Final	MA Starc Chidi

```
In [69]: Final_matche[['Winning year','winner']].sort_values('winner')
```

```
Out[69]:
```

	Winning year	winner
174	2010	Chennai Super Kings
247	2011	Chennai Super Kings
1023	2023	Chennai Super Kings
695	2018	Chennai Super Kings
875	2021	Chennai Super Kings
114	2009	Deccan Chargers
949	2022	Gujarat Titans
1094	2024	Kolkata Knight Riders
321	2012	Kolkata Knight Riders
457	2014	Kolkata Knight Riders
397	2013	Mumbai Indians
516	2015	Mumbai Indians
635	2017	Mumbai Indians
755	2019	Mumbai Indians
815	2020	Mumbai Indians
57	2008	Rajasthan Royals
576	2016	Sunrisers Hyderabad

```
In [70]: Final_matche['result'].value_counts()
```

```
Out[70]: result
runs      9
wickets   8
Name: count, dtype: int64
```

```
In [71]: df_matches['result'].value_counts()
```

```
Out[71]: result
wickets    578
runs       498
tie         14
no result    5
Name: count, dtype: int64
```

```
In [72]: df_matches.groupby('venue')['result'].value_counts()
```

```
Out[72]: venue      result
Arun Jaitley Stadium runs      44
           wickets    44
           no result    1
           tie         1
Barabati Stadium  runs      4
                ..
Wankhede Stadium wickets    64
           runs      53
           tie         1
Zayed Cricket Stadium wickets    5
           runs      3
Name: count, Length: 88, dtype: int64
```

```
In [73]: Final_matche.groupby('venue')['result'].value_counts().reset_index()
```

```
Out[73]:
```

	venue	result	count
0	Dr Dy Patil Sports Academy	runs	1
1	Dr Dy Patil Sports Academy	wickets	1
2	Dubai International Cricket Stadium	runs	1
3	Dubai International Cricket Stadium	wickets	1
4	Eden Gardens	runs	2
5	M Chinnaswamy Stadium	runs	1
6	M Chinnaswamy Stadium	wickets	1
7	Ma Chidambaram Stadium	wickets	2
8	Ma Chidambaram Stadium	runs	1
9	Narendra Modi Stadium	wickets	2
10	New Wanderers Stadium	runs	1
11	Rajiv Gandhi International Stadium	runs	2
12	Wankhede Stadium	wickets	1

```
In [74]: Final_matche.groupby('venue')['result'].value_counts().reset_index().sort
```

Out [74]:

	venue	result	count
4	Eden Gardens	runs	2
7	Ma Chidambaram Stadium	wickets	2
9	Narendra Modi Stadium	wickets	2
11	Rajiv Gandhi International Stadium	runs	2
0	Dr Dy Patil Sports Academy	runs	1
1	Dr Dy Patil Sports Academy	wickets	1
2	Dubai International Cricket Stadium	runs	1
3	Dubai International Cricket Stadium	wickets	1
5	M Chinnaswamy Stadium	runs	1
6	M Chinnaswamy Stadium	wickets	1
8	Ma Chidambaram Stadium	runs	1
10	New Wanderers Stadium	runs	1
12	Wankhede Stadium	wickets	1

In [75]: `Final_matches.groupby('venue')['result'].value_counts().reset_index().sort`

Out [75]:

	venue	result	count
0	Dr Dy Patil Sports Academy	runs	1
1	Dr Dy Patil Sports Academy	wickets	1
2	Dubai International Cricket Stadium	runs	1
3	Dubai International Cricket Stadium	wickets	1
4	Eden Gardens	runs	2
5	M Chinnaswamy Stadium	runs	1
6	M Chinnaswamy Stadium	wickets	1
7	Ma Chidambaram Stadium	wickets	2
8	Ma Chidambaram Stadium	runs	1
9	Narendra Modi Stadium	wickets	2
10	New Wanderers Stadium	runs	1
11	Rajiv Gandhi International Stadium	runs	2
12	Wankhede Stadium	wickets	1

In [76]: `df_matches.groupby('venue')['result'].value_counts().reset_index().sort_v`

Out [76]:

	venue	result	count
0	Arun Jaitley Stadium	runs	44
1	Arun Jaitley Stadium	wickets	44
2	Arun Jaitley Stadium	no result	1
3	Arun Jaitley Stadium	tie	1
4	Barabati Stadium	runs	4
...
83	Wankhede Stadium	wickets	64
84	Wankhede Stadium	runs	53
85	Wankhede Stadium	tie	1
86	Zayed Cricket Stadium	wickets	5
87	Zayed Cricket Stadium	runs	3

88 rows × 3 columns

Day 13

In [77]: df_matches

Out [77]:

	id	season	city	date	match_type	player_of_match	
0	335982	2007/08	Bangalore	2008-04-18	League	BB McCullum	Chinna S
1	335983	2007/08	Chandigarh	2008-04-19	League	MEK Hussey	Asso S
2	335984	2007/08	Delhi	2008-04-19	League	MF Maharoor	Arun S
3	335985	2007/08	Mumbai	2008-04-20	League	MV Boucher	War S
4	335986	2007/08	Kolkata	2008-04-20	League	DJ Hussey	G
...
1090	1426307	2024	Hyderabad	2024-05-19	League	Abhishek Sharma	Rajiv Intern S
1091	1426309	2024	Ahmedabad	2024-05-21	Qualifier 1	MA Starc	Na Modi S
1092	1426310	2024	Ahmedabad	2024-05-22	Eliminator	R Ashwin	Na Modi S
1093	1426311	2024	Chennai	2024-05-24	Qualifier 2	Shahbaz Ahmed	Chidam S
1094	1426312	2024	Chennai	2024-05-26	Final	MA Starc	Chidam S

1095 rows × 18 columns

In [78]:

```
# Question to be solve?
# # all Season Winner team
# # Toss_dicision_value_counts
# # Result type value_counts
# # Top 10 target Runs
# # Maximum Result Margin Wins
# # Top 5 Result Margin Wins
# # Top 10 player_of_match
# # Season Player of the match
```

Q1: All Season Winner team

In [79]:

```
Final_matche[['Winning year','winner' ]].sort_values('Winning year')
```

Out [79]:

	Winning year	winner
57	2008	Rajasthan Royals
114	2009	Deccan Chargers
174	2010	Chennai Super Kings
247	2011	Chennai Super Kings
321	2012	Kolkata Knight Riders
397	2013	Mumbai Indians
457	2014	Kolkata Knight Riders
516	2015	Mumbai Indians
576	2016	Sunrisers Hyderabad
635	2017	Mumbai Indians
695	2018	Chennai Super Kings
755	2019	Mumbai Indians
815	2020	Mumbai Indians
875	2021	Chennai Super Kings
949	2022	Gujarat Titans
1023	2023	Chennai Super Kings
1094	2024	Kolkata Knight Riders

Q2:Top Match Winner team

```
In [80]: df_matches['winner'].value_counts().head(10)
```

```
Out [80]: winner
Mumbai Indians           144
Chennai Super Kings      138
Kolkata Knight Riders     131
Royal Challengers Bengaluru 123
Delhi Capitals            115
Rajasthan Royals         112
Punjab Kings             112
Sunrisers Hyderabad       88
Deccan Chargers           29
Gujarat Titans            28
Name: count, dtype: int64
```

```
In [90]: # All Univariate analysis
num_cols = df_matches.describe()
cat_cols = df_matches.describe(include = 'O').columns
```

```
In [82]: print(num_cols)
```

	id	date	result_margin \
count	1.095000e+03	1095	1095.000000
mean	9.048283e+05	2016-06-25 11:39:36.986301184	16.959817
min	3.359820e+05	2008-04-18 00:00:00	0.000000
25%	5.483315e+05	2012-04-21 00:00:00	5.000000
50%	9.809610e+05	2016-05-03 00:00:00	8.000000
75%	1.254062e+06	2021-04-13 12:00:00	19.000000
max	1.426312e+06	2024-05-26 00:00:00	146.000000
std	3.677402e+05	NaN	21.714792

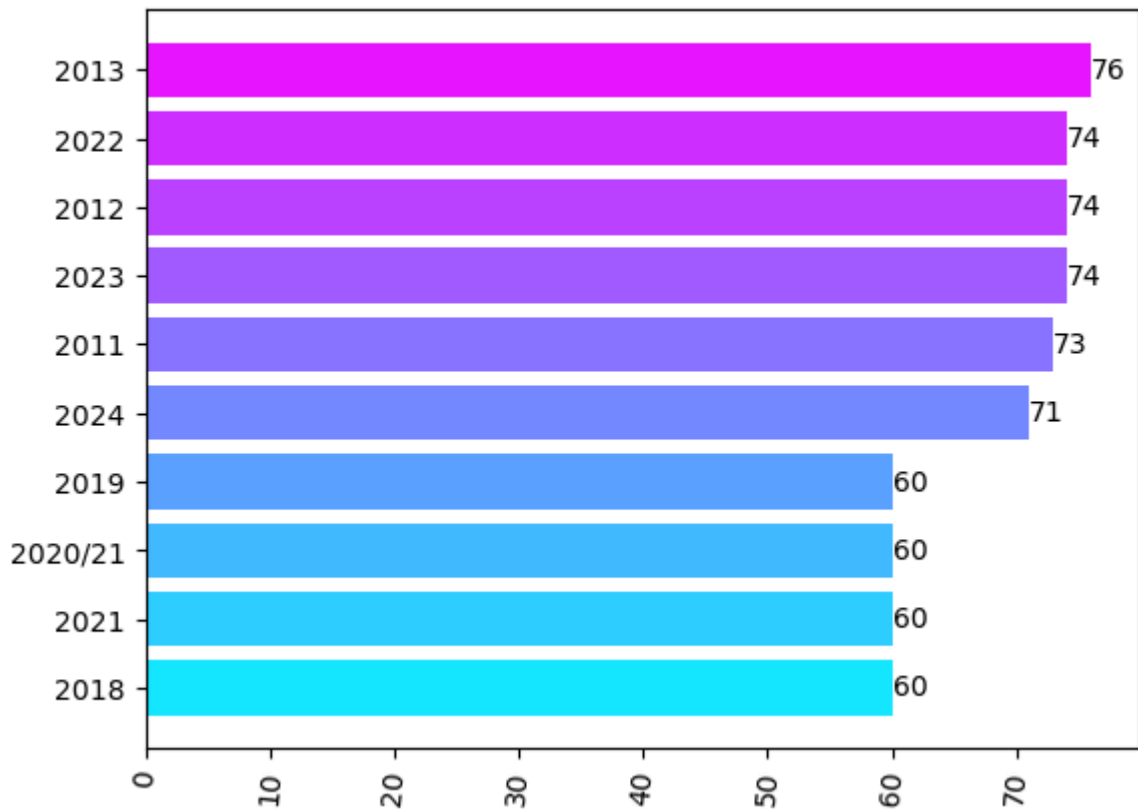
	target_runs	target_overs
count	1095.000000	1095.000000
mean	165.230137	19.705205
min	0.000000	0.000000
25%	146.000000	20.000000
50%	166.000000	20.000000
75%	187.000000	20.000000
max	288.000000	20.000000
std	34.487313	1.887000

```
In [83]: print(cat_cols)
```

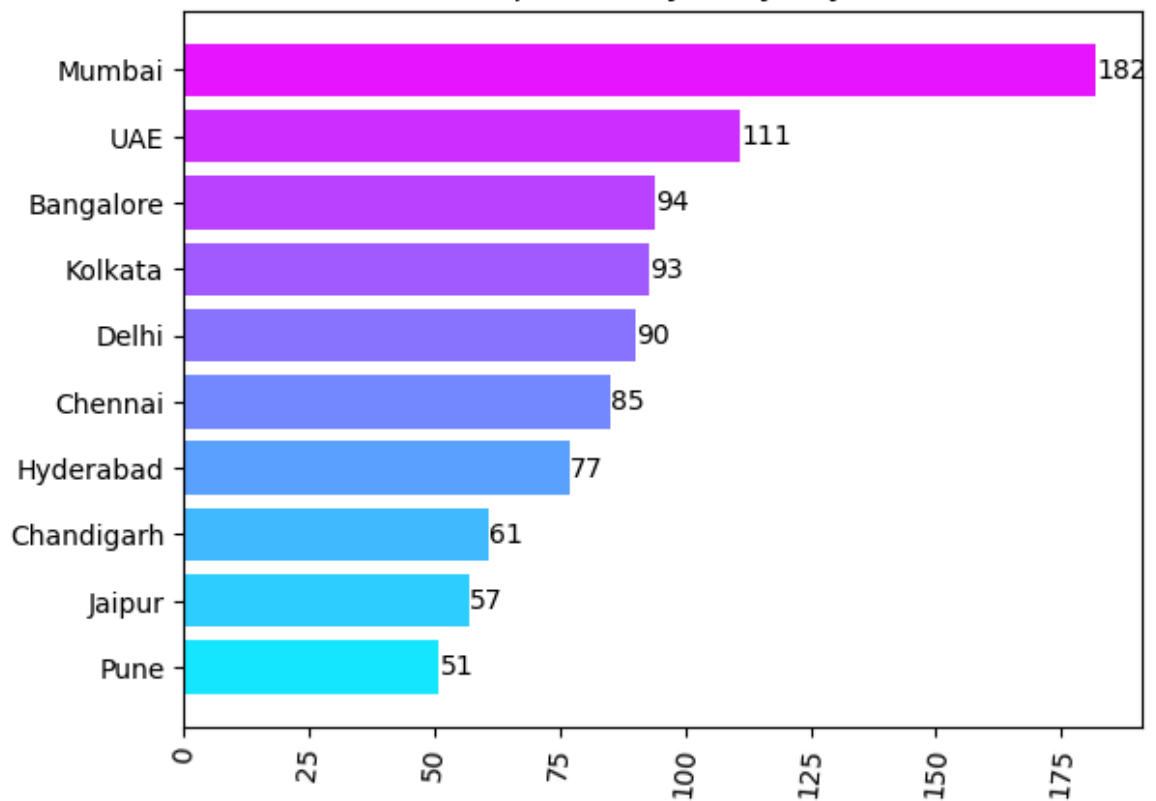
```
Index(['season', 'city', 'match_type', 'player_of_match', 'venue', 'team
1',
      'team2', 'toss_winner', 'toss_decision', 'winner', 'result',
      'super_over', 'method'],
      dtype='object')
```

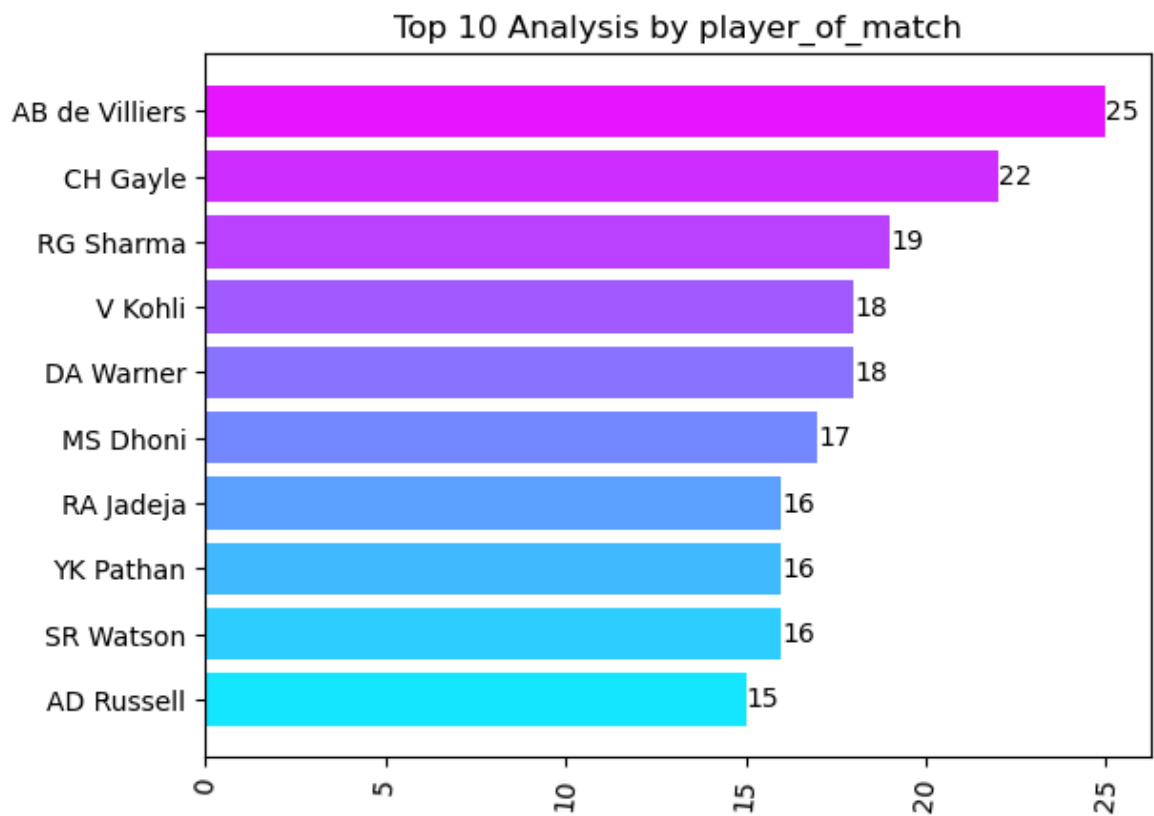
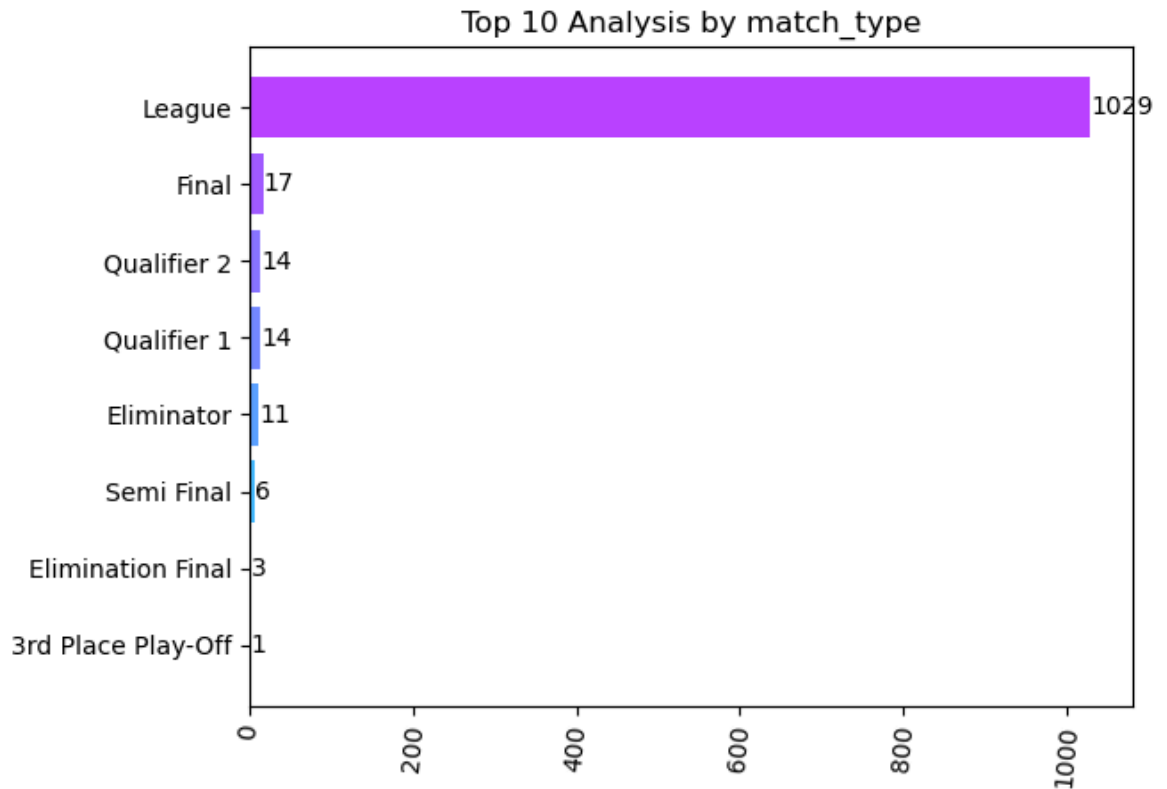
```
In [84]: for i in cat_cols:
          try:
              temp_df = df_matches[i].value_counts().head(10).sort_values(ascending=True)
              x = temp_df.index
              y = temp_df.values
              plt.title(f'Top 10 Analysis by {i}')
              ax = plt.barh(x,y,color = sns.color_palette('cool',10))
              plt.bar_label(ax)
              plt.xticks(rotation = 85)
              plt.show()
          except:
              pass
```

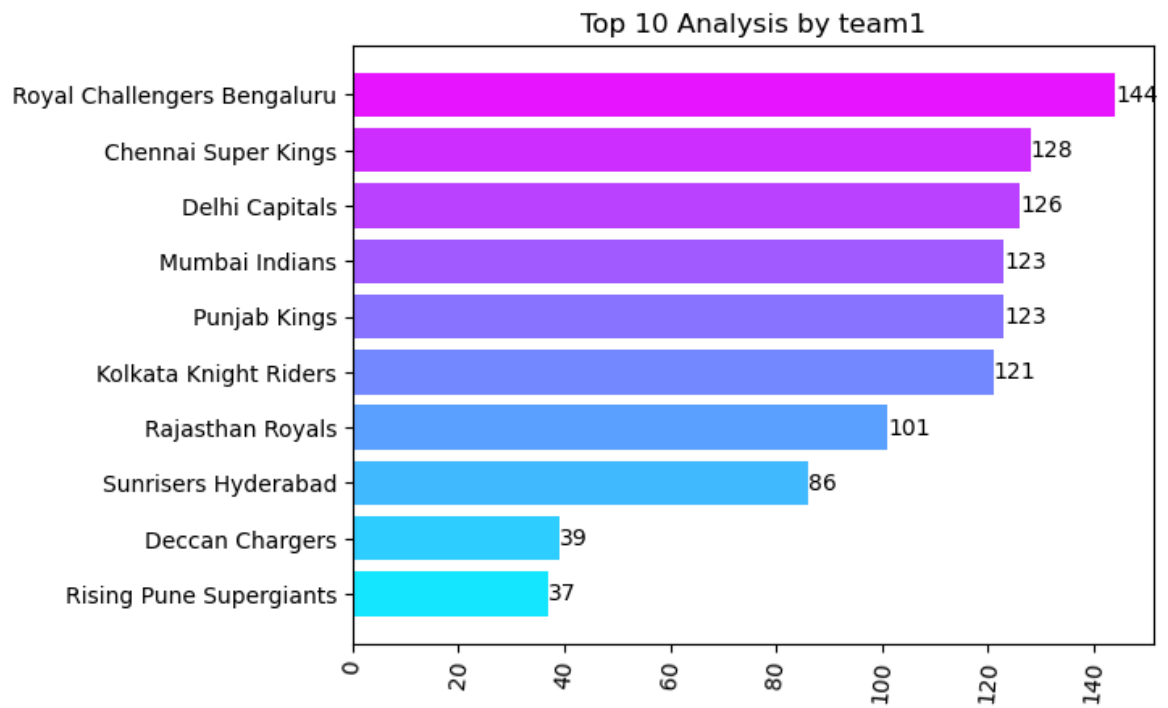
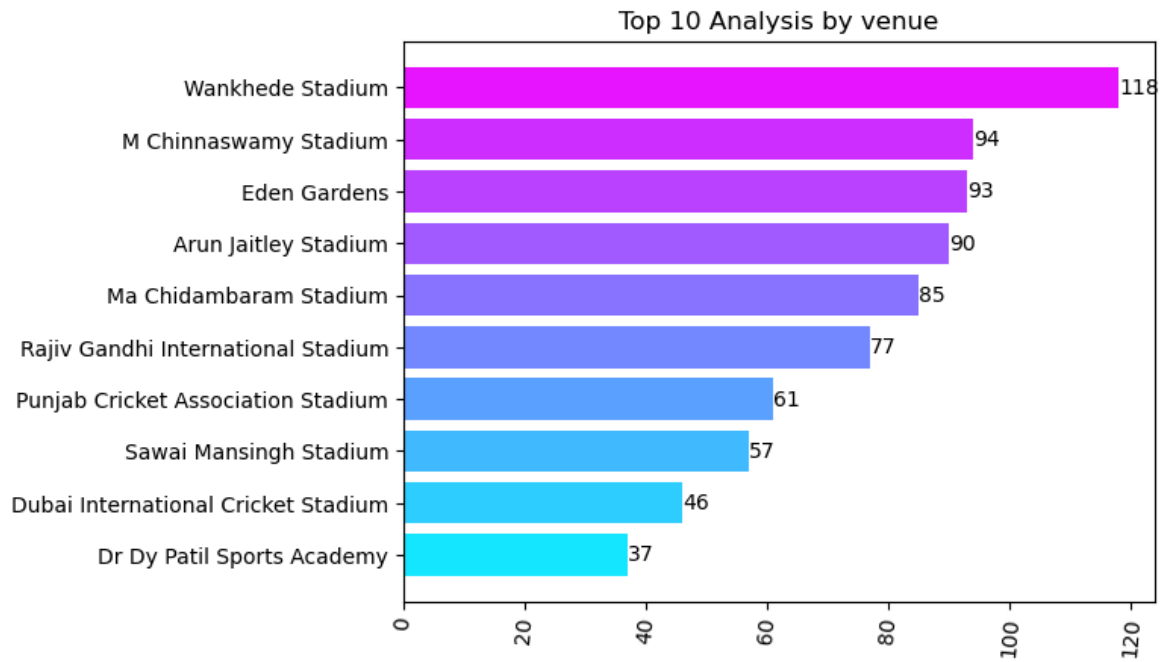
Top 10 Analysis by season

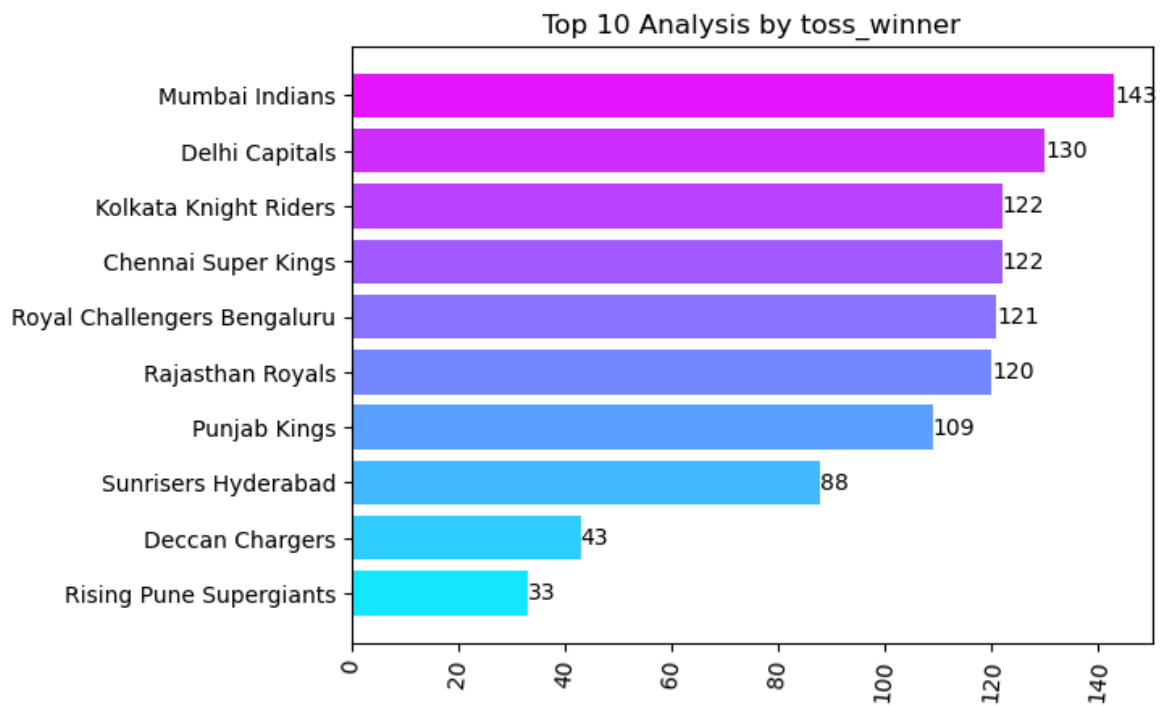
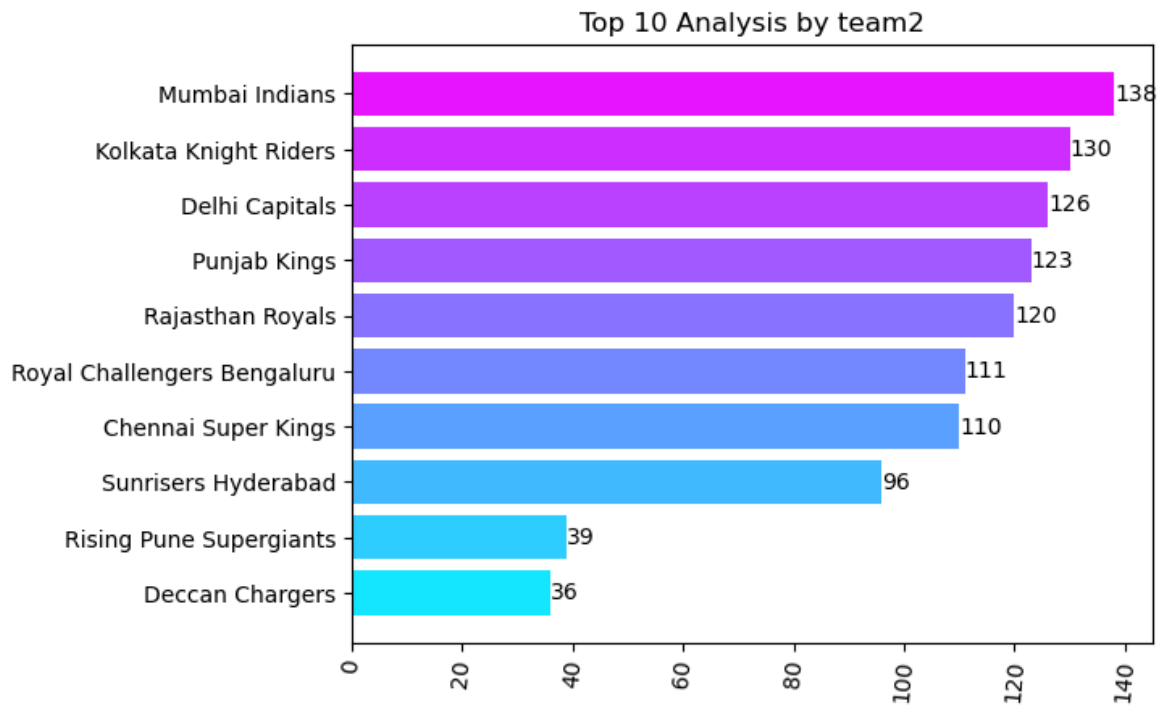


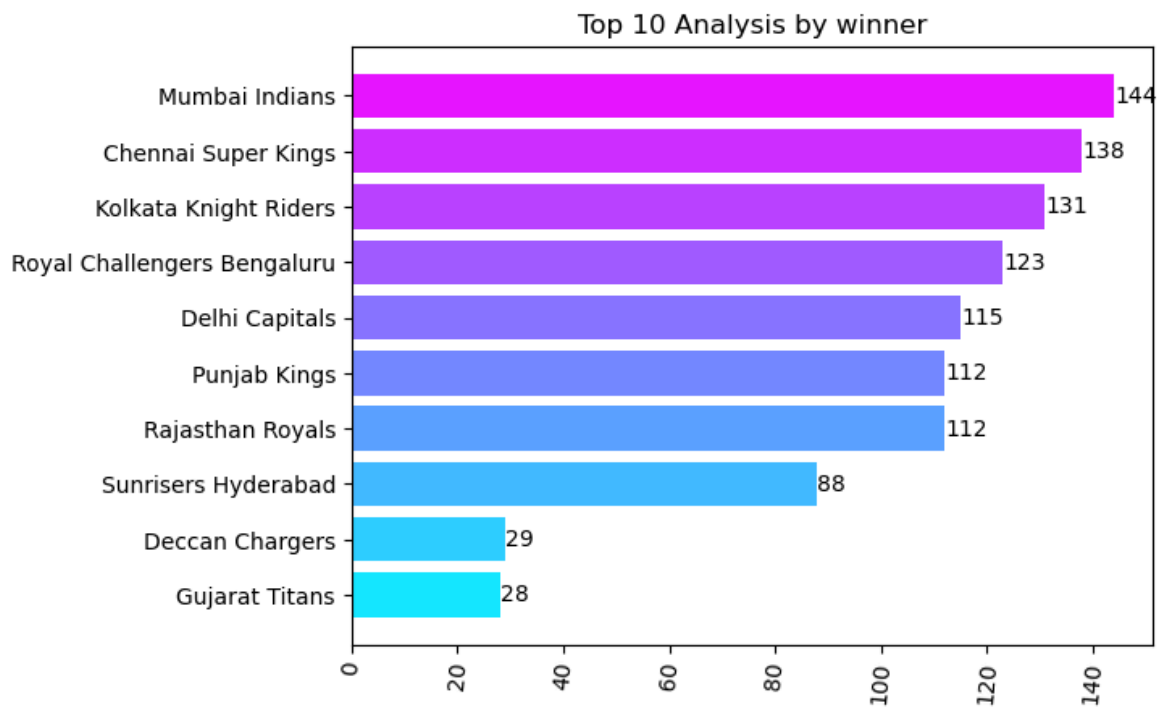
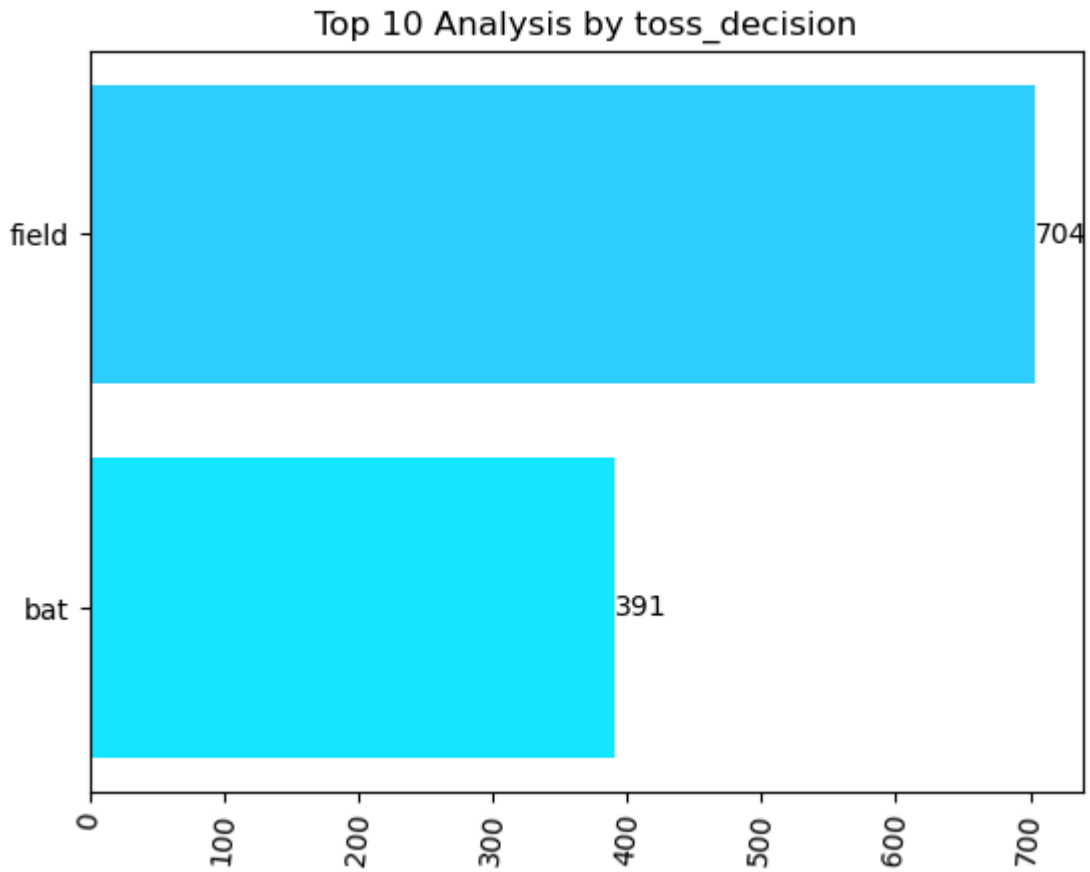
Top 10 Analysis by city

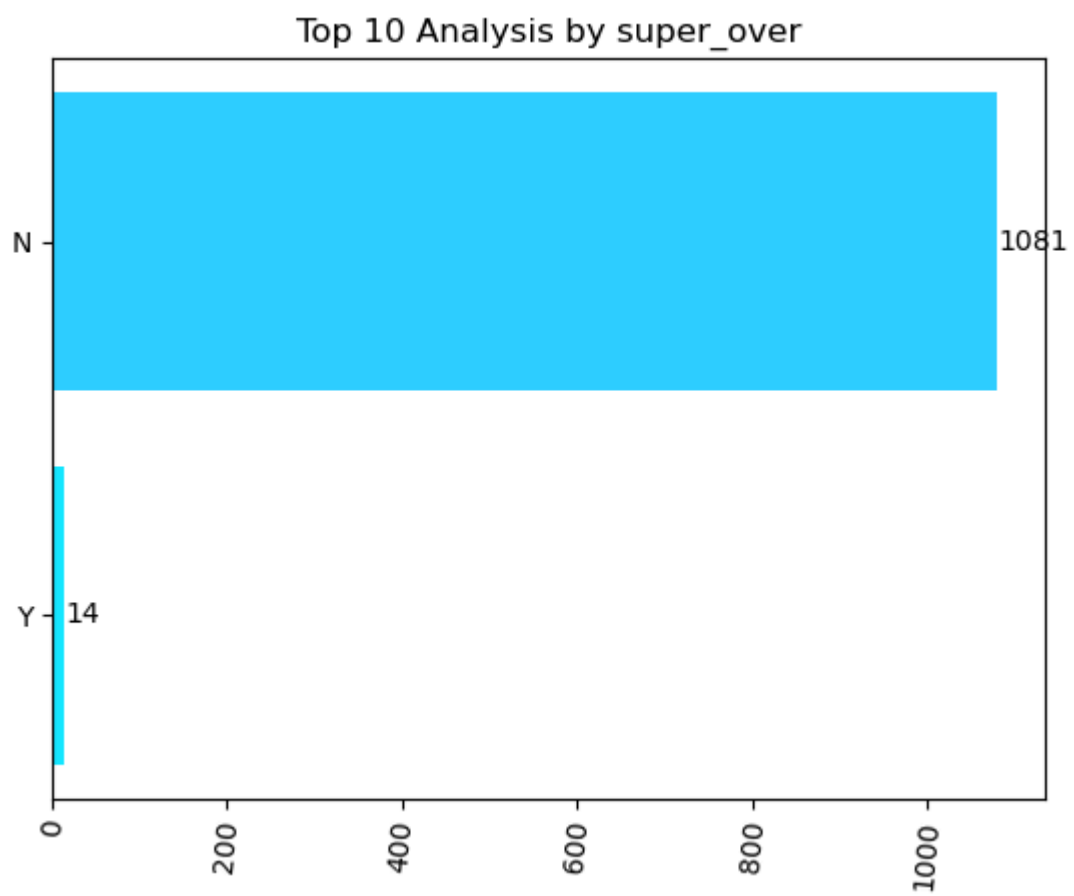
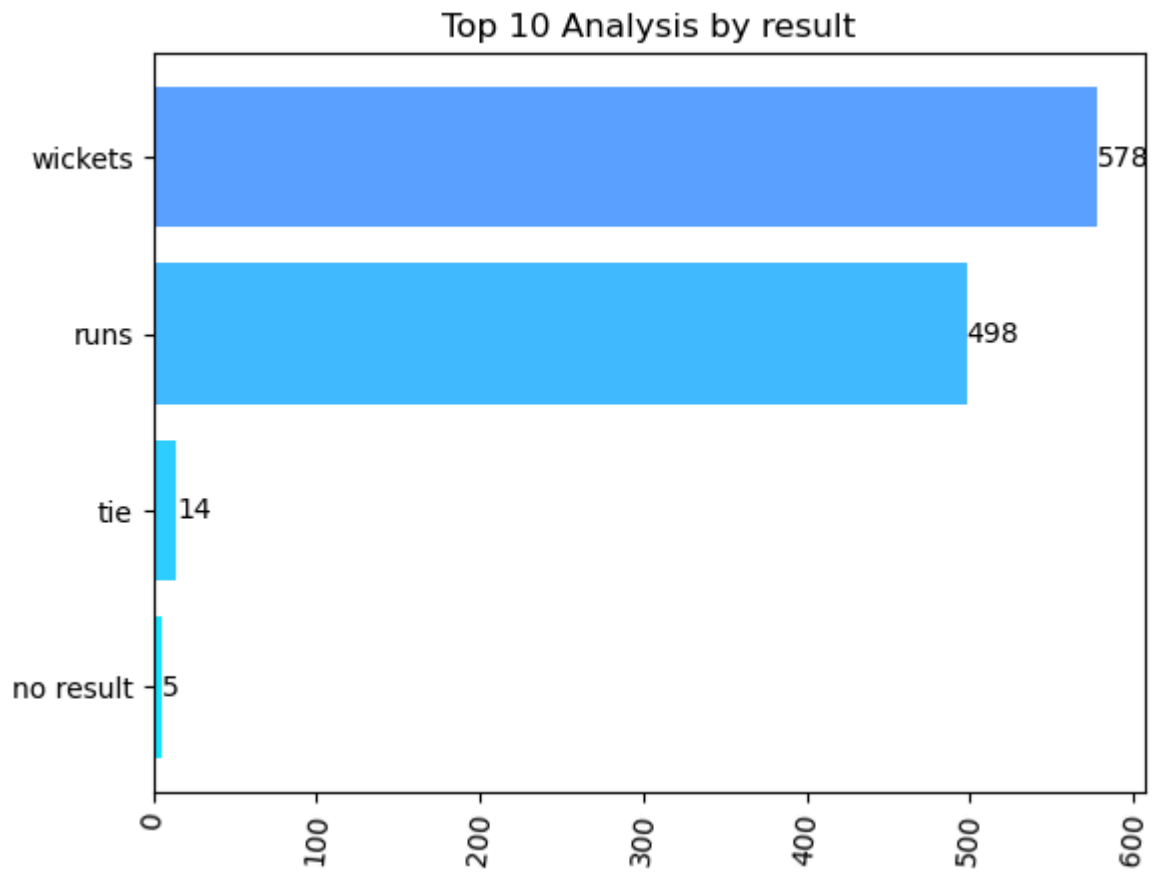


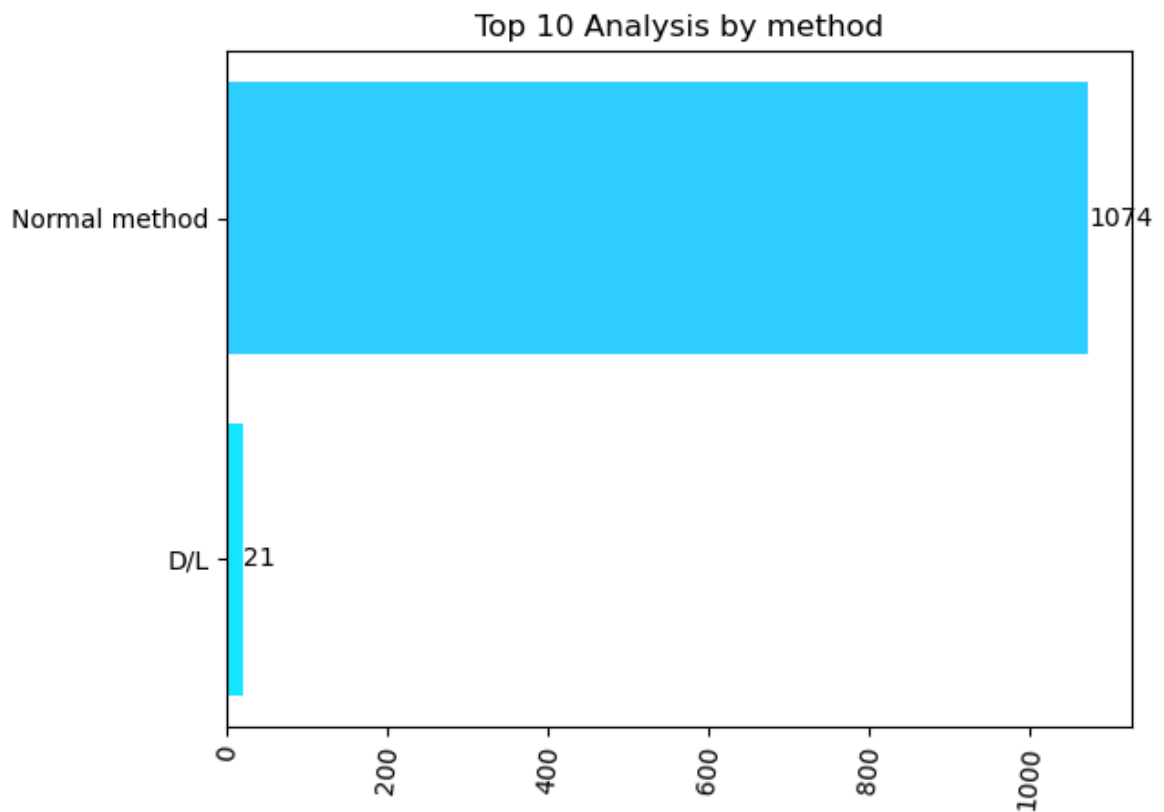






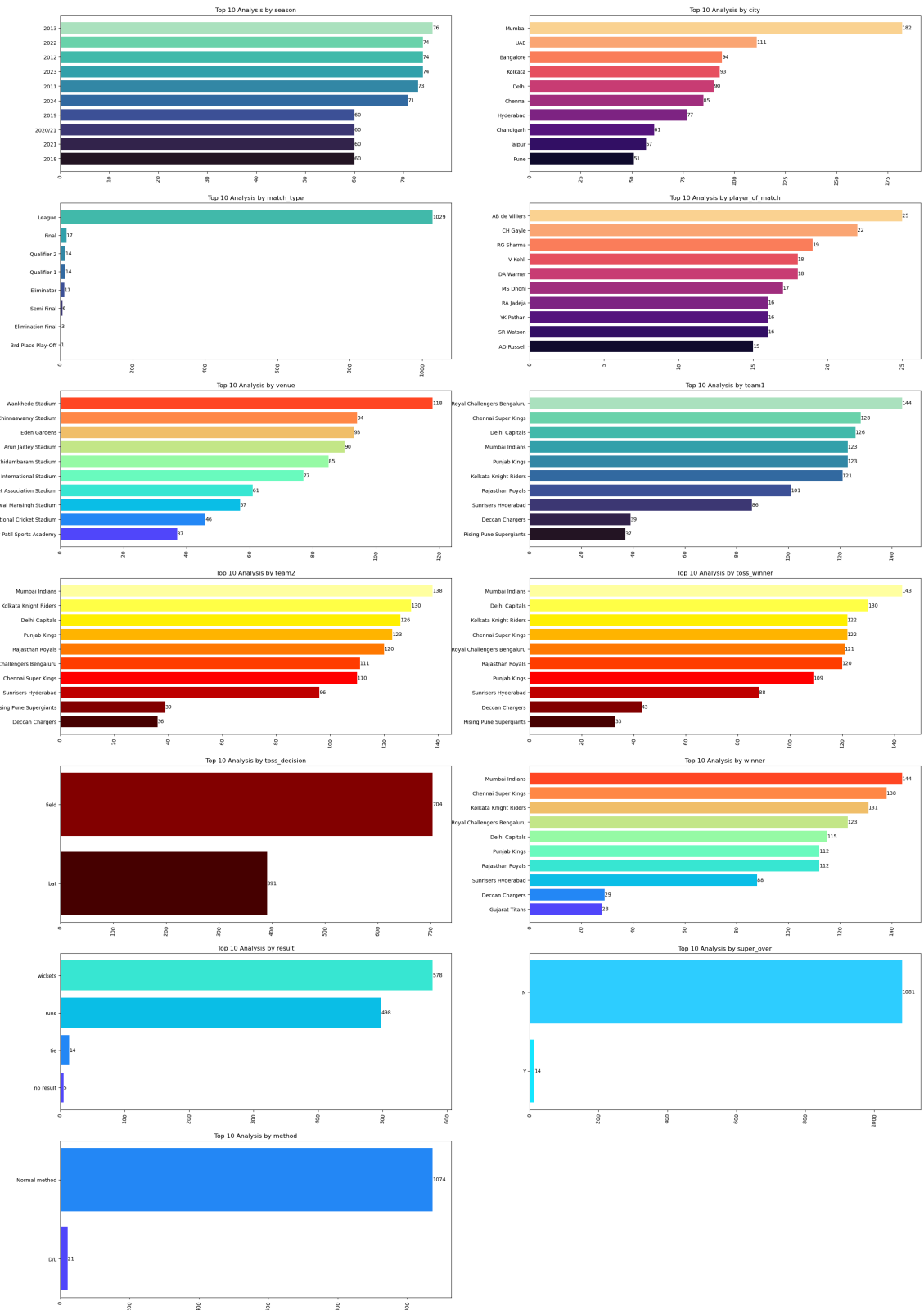






```
In [85]: plt.figure(figsize=(30,45))
for j,i in enumerate(cat_cols):
    temp_df = df_matches[i].value_counts().head(10).sort_values(ascending=True)
    x = temp_df.index
    y = temp_df.values

    plt.subplot(7,2,j+1)
    plt.title(f'Top 10 Analysis by {i}')
    colors = ['cool','mako','rainbow','magma','hot']
    import random
    ax = plt.barh(x,y,color = sns.color_palette(random.choice(colors)))
    plt.bar_label(ax)
    plt.xticks(rotation = 85)
plt.savefig('All_Categorical_Analysis.jpeg',dpi = 300)
plt.show()
print('Done')
```



Done

In []:

In []:

In []:

```
In [86]: # Seaborn dataset, sklearn dataset
# pandas
# concat
# merge, joins
```

```
# pivot table
# matplotlib
# seaborn
# Recap
# # Corr
# # outliers detection
```

```
In [87]: # df_deliveries.sample()
```

```
In [88]: # Top 10 target Runs
# Maximum Result Margin Wins
# Top 5 Result Margin Wins
# Season Player of the match
```

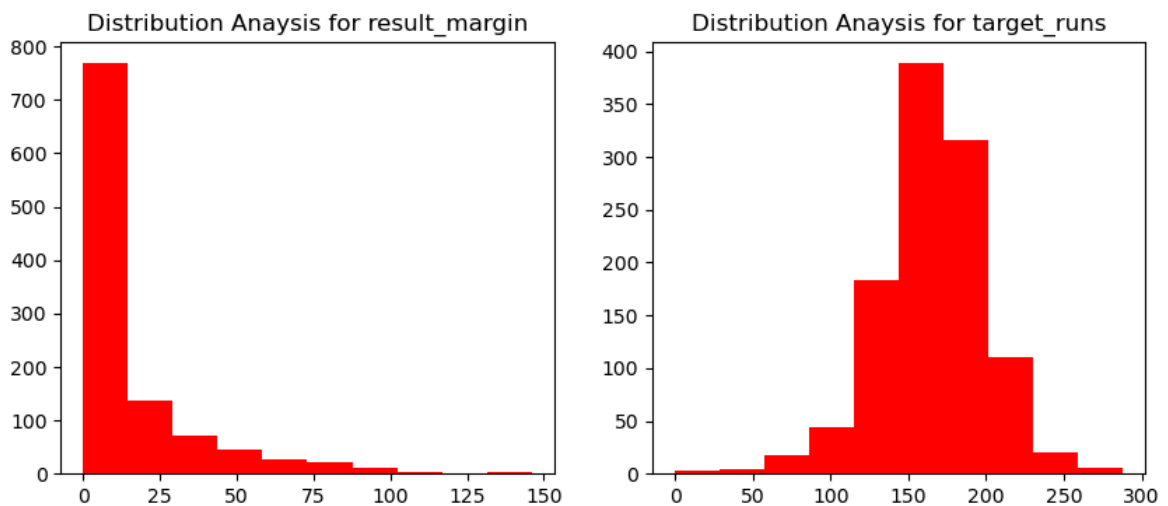
```
In [96]: num_cols = df_matches.describe().columns

num_cols = list(num_cols)
num_cols.remove('id')
num_cols.remove('date')
num_cols.remove('target_overs')
num_cols
```

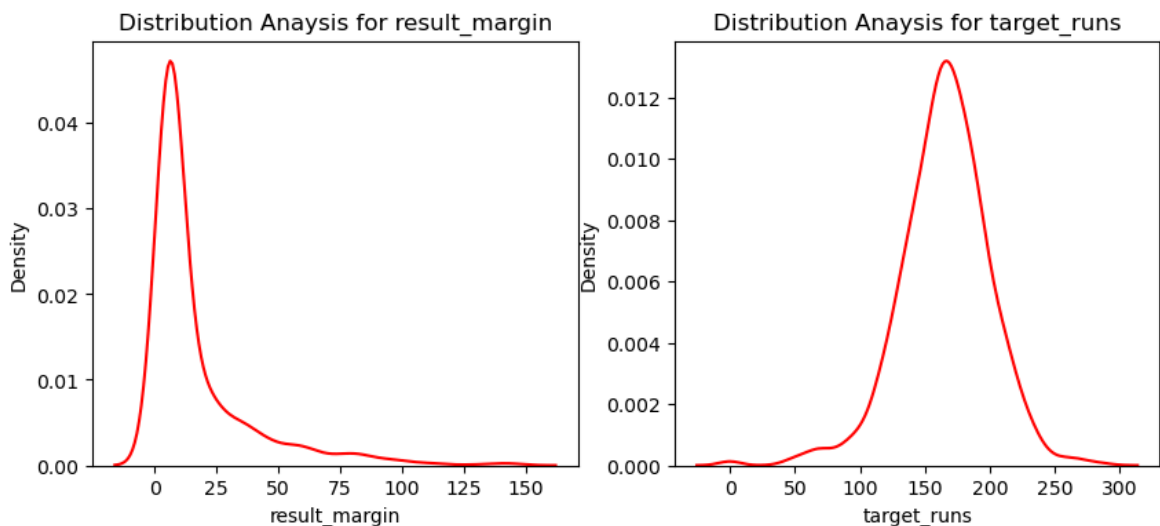
```
Out[96]: ['result_margin', 'target_runs']
```

```
In [100... # Numerical Histogram

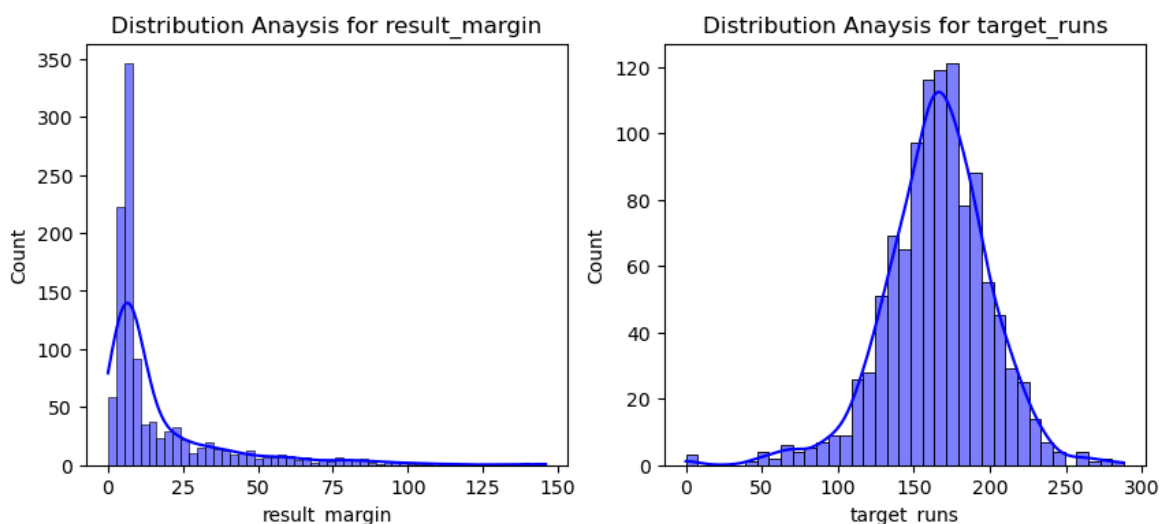
plt.figure(figsize=(10,4))
for j,i in enumerate(num_cols):
    plt.subplot(1,2,j+1)
    plt.title(f"Distribution Anaysis for {i}")
    plt.hist(df_matches[i],color = 'red')
plt.show()
```



```
In [104... plt.figure(figsize=(10,4))
for j,i in enumerate(num_cols):
    plt.subplot(1,2,j+1)
    plt.title(f"Distribution Anaysis for {i}")
    sns.kdeplot(df_matches[i],color = 'red')
plt.show()
```

```
In [106... plt.figure(figsize=(10,4))
for j,i in enumerate(num_cols):
    plt.subplot(1,2,j+1)
    plt.title(f"Distribution Anaysis for {i}")
    sns.histplot(df_matches[i],color = 'blue', kde = True)
plt.show()
```



```
In [107... # Max Target Runs which team?

df_matches[df_matches['target_runs'] == df_matches['target_runs'].max()]
```

```
Out[107...      id  season    city    date  match_type  player_of_match  ve
```

	id	season	city	date	match_type	player_of_match	ve
1053	1426268	2024	Bangalore	2024-04-15	League	TM Head Chinnaswa	Stad

```
In [118... # Min Target Runs which team?
temp_df = df_matches[(df_matches['winner'] != 'No winner') & (df_matches[
temp_df[temp_df['target_runs'] == temp_df['target_runs'].min()]
```

Out [118...

	id	season	city	date	match_type	player_of_match	venue
241	501265	2011	Delhi	2011-05-21	League	No Player of the match	Arun Jaitley Stadium
485	829763	2015	Bangalore	2015-04-29	League	No Player of the match	Chinnaswamy Stadium
994	1359519	2023	Lucknow	2023-05-03	League	No Player of the match	Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Ground

In [115...

```
df_matches[df_matches['date'].dt.year == 2017].sort_values(by = 'target_run_rate')
```

Out [115...

	id	season	city	date	match_type	player_of_match	ve
633	1082648	2017	Bangalore	2017-05-17	Eliminator	NM Coulter-Nile	Chinnaswa Stad
611	1082626	2017	Chandigarh	2017-04-30	League	Sandeep Sharma	Pun Cric Associa Stad
630	1082645	2017	Pune	2017-05-14	League	JD Unadkat	Maharas Cric Associa Stad
634	1082649	2017	Bangalore	2017-05-19	Qualifier 2	KV Sharma	Chinnaswa Stad
635	1082650	2017	Hyderabad	2017-05-21	Final	KH Pandya	Rajiv Gar Internati Stad
603	1082617	2017	Kolkata	2017-04-23	League	NM Coulter-Nile	E Gard
606	1082621	2017	Bangalore	2017-04-27	League	AJ Tye	Chinnaswa Stad
582	1082596	2017	Hyderabad	2017-04-09	League	Rashid Khan	Rajiv Gar Internati Stad
623	1082638	2017	Hyderabad	2017-05-08	League	S Dhawan	Rajiv Gar Internati Stad
618	1082633	2017	Bangalore	2017-05-05	League	Sandeep Sharma	Chinnaswa Stad
588	1082602	2017	Bangalore	2017-04-14	League	KA Pollard	Chinnaswa Stad
600	1082614	2017	Mumbai	2017-04-22	League	MJ McClenaghan	Wankh Stad
584	1082598	2017	Indore	2017-04-10	League	AR Patel	Ho Cric Stad
619	1082634	2017	Hyderabad	2017-05-06	League	JD Unadkat	Rajiv Gar Internati Stad
610	1082625	2017	Rajkot	2017-04-29	League	KH Pandya	Sauras Cric Associa Stad

	id	season	city	date	match_type	player_of_match	ve
628	1082643	2017	Kanpur	2017-05-13	League	Mohammed Siraj	Green F
616	1082631	2017	Kolkata	2017-05-03	League	RA Tripathi	E Gard
581	1082595	2017	Bangalore	2017-04-08	League	KM Jadhav	Chinnaswa Stad
609	1082624	2017	Pune	2017-04-29	League	LH Ferguson	Maharas Cric Associa Stad
586	1082600	2017	Mumbai	2017-04-12	League	JJ Bumrah	Wankh Stad
621	1082636	2017	Bangalore	2017-05-07	League	SP Narine	Chinnaswa Stad
595	1082609	2017	Hyderabad	2017-04-17	League	B Kumar	Rajiv Gar Internati Stad
607	1082622	2017	Kolkata	2017-04-28	League	G Gambhir	E Gard
604	1082618	2017	Mumbai	2017-04-24	League	BA Stokes	Wankh Stad
631	1082646	2017	Delhi	2017-05-14	League	HV Patel	Arun Jai Stad
614	1082629	2017	Pune	2017-05-01	League	BA Stokes	Maharas Cric Associa Stad
593	1082607	2017	Bangalore	2017-04-16	League	BA Stokes	Chinnaswa Stad
613	1082628	2017	Mumbai	2017-05-01	League	RG Sharma	Wankh Stad
632	1082647	2017	Mumbai	2017-05-16	Qualifier 1	Washington Sundar	Wankh Stad
580	1082594	2017	Indore	2017-04-08	League	GJ Maxwell	Ho Cric Stad
624	1082639	2017	Chandigarh	2017-05-	League	MM Sharma	Pui Cric

	id	season	city	date	match_type	player_of_match	ve
				09			Associa Stad
627	1082642	2017	Delhi	2017- 05- 12	League	KK Nair	Arun Jai Stad
594	1082608	2017	Delhi	2017- 04- 17	League	NM Coulter-Nile	Arun Jai Stad
587	1082601	2017	Kolkata	2017- 04- 13	League	SP Narine	E Gard
589	1082603	2017	Rajkot	2017- 04- 14	League	AJ Tye	Sauras Cric Associa Stad
590	1082604	2017	Kolkata	2017- 04- 15	League	RV Uthappa	E Gard
629	1082644	2017	Kolkata	2017- 05- 13	League	AT Rayudu	E Gard
592	1082606	2017	Mumbai	2017- 04- 16	League	N Rana	Wankh Stad
601	1082615	2017	Pune	2017- 04- 22	League	MS Dhoni	Maharas Cric Associa Stad
583	1082597	2017	Mumbai	2017- 04- 09	League	N Rana	Wankh Stad
605	1082620	2017	Pune	2017- 04- 26	League	RV Uthappa	Maharas Cric Associa Stad
579	1082593	2017	Rajkot	2017- 04- 07	League	CA Lynn	Sauras Cric Associa Stad
578	1082592	2017	Pune	2017- 04- 06	League	SPD Smith	Maharas Cric Associa Stad
615	1082630	2017	Delhi	2017- 05- 02	League	Mohammed Shami	Arun Jai Stad
599	1082613	2017	Kolkata	2017- 04- 21	League	SK Raina	E Gard

	id	season	city	date	match_type	player_of_match	ve
602	1082616	2017	Rajkot	2017-04-23	League	HM Amla	Sauras Cric Associa Stad
591	1082605	2017	Delhi	2017-04-15	League	CJ Anderson	Arun Jai Stad
622	1082637	2017	Chandigarh	2017-05-07	League	DR Smith	Pur Cric Associa Stad
597	1082611	2017	Hyderabad	2017-04-19	League	KS Williamson	Rajiv Gar Internatio Stad
625	1082640	2017	Kanpur	2017-05-10	League	SS Iyer	Green F
598	1082612	2017	Indore	2017-04-20	League	JC Buttler	Ho Cric Stad
585	1082599	2017	Pune	2017-04-11	League	SV Samson	Maharas Cric Associa Stad
577	1082591	2017	Hyderabad	2017-04-05	League	Yuvraj Singh	Rajiv Gar Internatio Stad
608	1082623	2017	Chandigarh	2017-04-28	League	Rashid Khan	Pur Cric Associa Stad
617	1082632	2017	Delhi	2017-05-04	League	RR Pant	Arun Jai Stad
612	1082627	2017	Hyderabad	2017-04-30	League	DA Warner	Rajiv Gar Internatio Stad
620	1082635	2017	Delhi	2017-05-06	League	LMP Simmons	Arun Jai Stad
596	1082610	2017	Rajkot	2017-04-18	League	CH Gayle	Sauras Cric Associa Stad
626	1082641	2017	Mumbai	2017-05-11	League	WP Saha	Wankh Stad

In []:

```
In [117... # LOC. , ILOC
# Index = 603
df_matches.iloc[603,1:]
```

```
Out[117... season                2017
city                Kolkata
date                2017-04-23 00:00:00
match_type          League
player_of_match      NM Coulter-Nile
venue                Eden Gardens
team1                Kolkata Knight Riders
team2                Royal Challengers Bengaluru
toss_winner          Royal Challengers Bengaluru
toss_decision        field
winner                Kolkata Knight Riders
result                runs
result_margin        82.0
target_runs          132.0
target_overs         20.0
super_over           N
method                Normal method
Name: 603, dtype: object
```

```
In [119... print(num_cols)
print(cat_cols)

['result_margin', 'target_runs']
Index(['season', 'city', 'match_type', 'player_of_match', 'venue', 'team
1',
      'team2', 'toss_winner', 'toss_decision', 'winner', 'result',
      'super_over', 'method'],
      dtype='object')
```

```
In [120... df_matches.columns
```

```
Out[120... Index(['id', 'season', 'city', 'date', 'match_type', 'player_of_match',
      'venue', 'team1', 'team2', 'toss_winner', 'toss_decision', 'winne
r',
      'result', 'result_margin', 'target_runs', 'target_overs', 'super_
over',
      'method'],
      dtype='object')
```

```
In [127... # stadium wise Team highest score
df_matches.groupby(['venue', 'team1'])['target_runs'].agg(['max', 'min', 'me
```

Out [127...

		max	min	mean
venue	team1			
Arun Jaitley Stadium	Chennai Super Kings	224.0	193.0	212.000000
	Delhi Capitals	258.0	0.0	164.971831
	Gujarat Lions	163.0	163.0	163.000000
	Kolkata Knight Riders	186.0	128.0	157.000000
	Mumbai Indians	169.0	169.0	169.000000
...
Zayed Cricket Stadium	Kolkata Knight Riders	172.0	172.0	172.000000
	Mumbai Indians	236.0	156.0	196.000000
	Punjab Kings	136.0	136.0	136.000000
	Royal Challengers Bengaluru	93.0	93.0	93.000000
	Sunrisers Hyderabad	142.0	142.0	142.000000

211 rows x 3 columns

In [125...

```
df_matches.groupby(['target_runs', 'team1'])['venue'].max().tail(10)
```

Out [125...

```
target_runs  team1
247.0        Chennai Super Kings
a Chidambaram Stadium
249.0        Royal Challengers Bengaluru
M Chinnaswamy Stadium
258.0        Delhi Capitals
Arun Jaitley Stadium
                Lucknow Super Giants
t Association Stadium
262.0        Kolkata Knight Riders
Eden Gardens
264.0        Royal Challengers Bengaluru
M Chinnaswamy Stadium
267.0        Sunrisers Hyderabad
Arun Jaitley Stadium
273.0        Kolkata Knight Riders
ca-Vdca Cricket St...
278.0        Sunrisers Hyderabad
International Stadium
288.0        Sunrisers Hyderabad
M Chinnaswamy Stadium
Name: venue, dtype: object
```

In [128...

```
# Pivot table

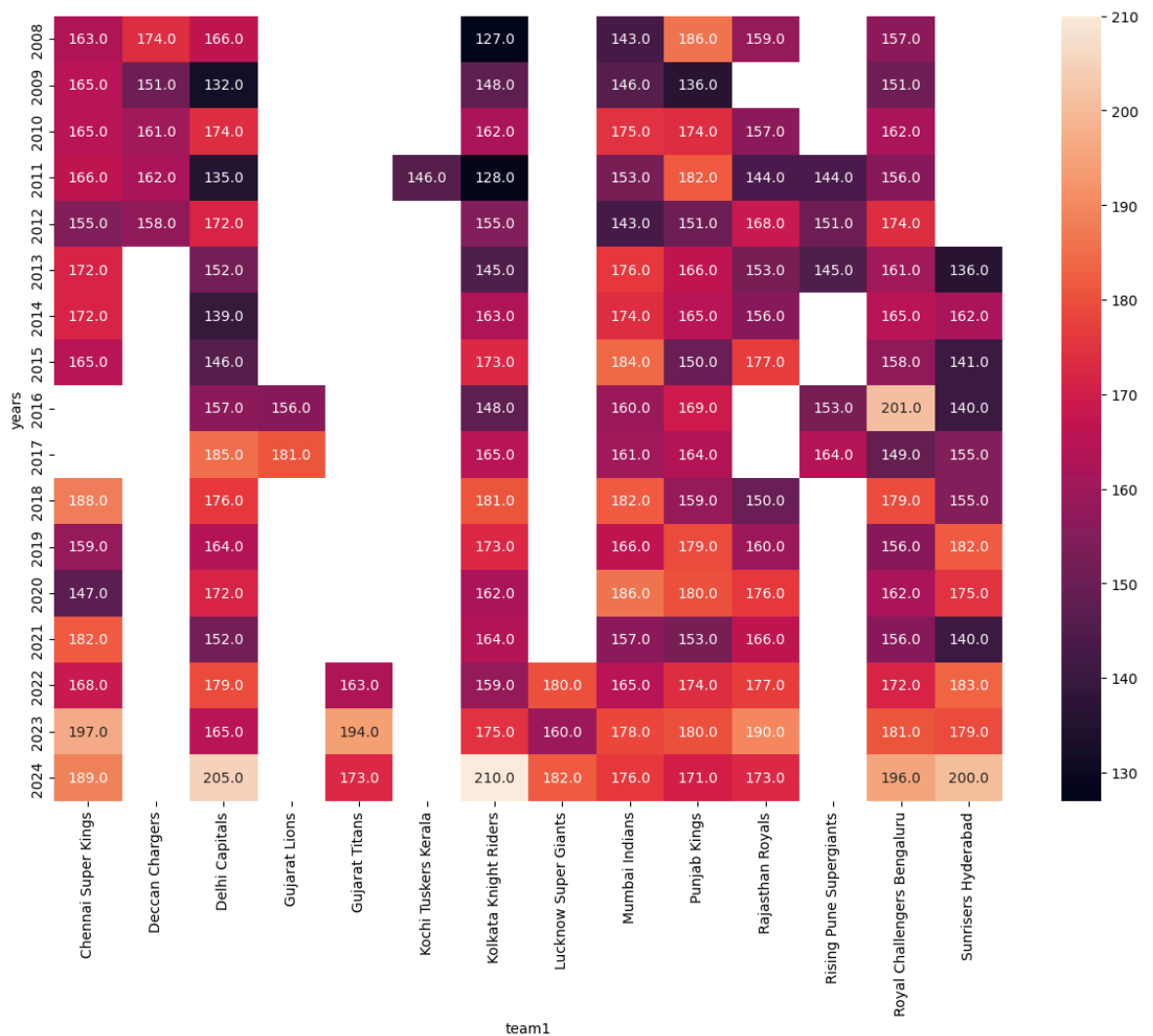
df_matches['years'] = df_matches['date'].dt.year
df_matches.columns
```



```
Out[128... Index(['id', 'season', 'city', 'date', 'match_type', 'player_of_match',
      'venue', 'team1', 'team2', 'toss_winner', 'toss_decision', 'winne
r',
      'result', 'result_margin', 'target_runs', 'target_overs', 'super_
over',
      'method', 'years'],
      dtype='object')
```

```
In [145... all_team_year_df = df_matches.pivot_table(index='years', columns='team
all_team_year_df = all_team_year_df.round(0)
```

```
In [146... plt.figure(figsize=(15,10))
# pd.options.display.float_format = '{:.2f}'.format
# pd.set_option('display.float_format', '{:.2f}'.format)
sns.heatmap(all_team_year_df,fmt = "" ,annot = True)
plt.show()
```

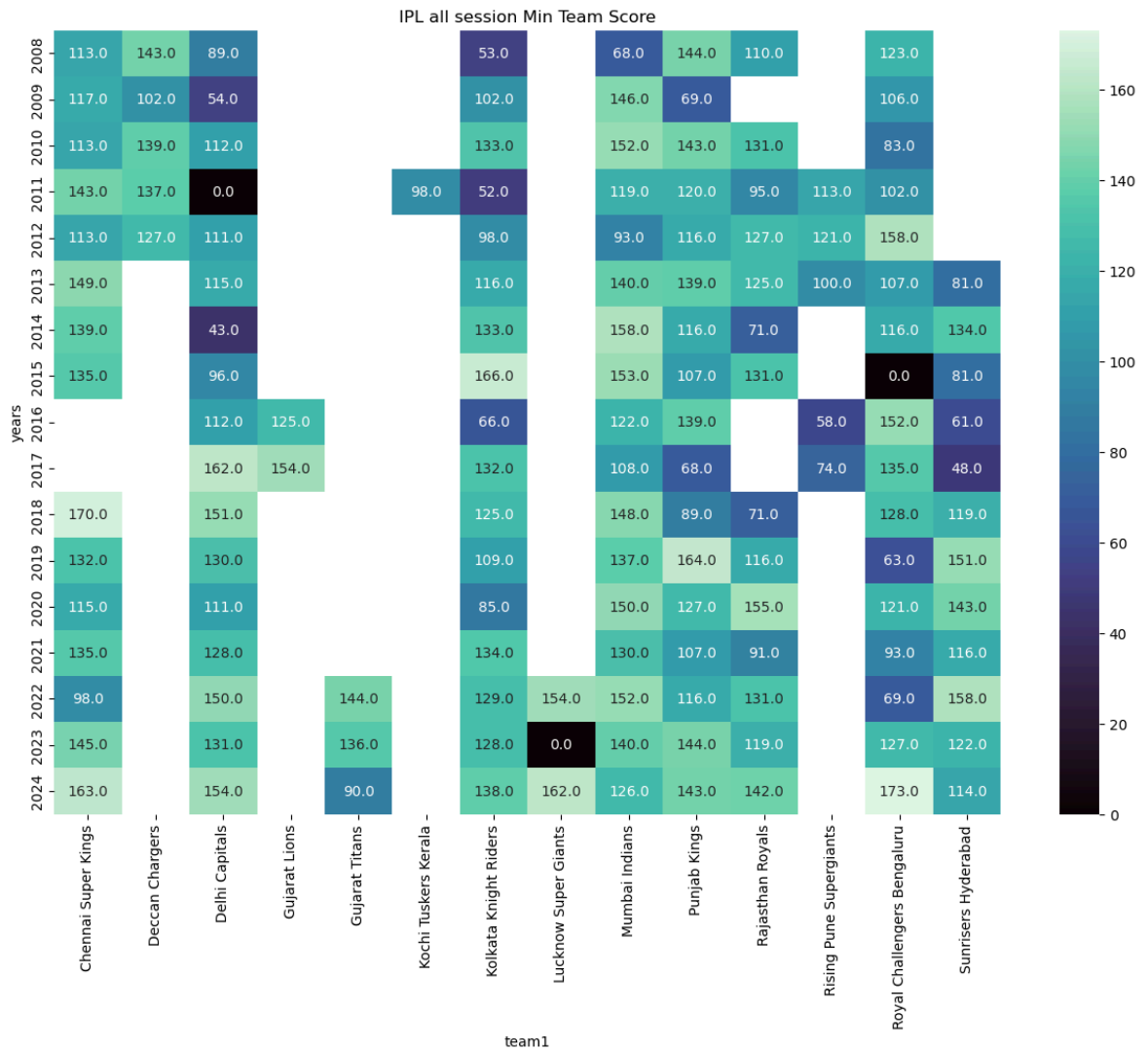


```
In [147... all_team_year_df_2 = df_matches.pivot_table(index='years', columns='te
all_team_year_df_2 = all_team_year_df_2.round(0)
```

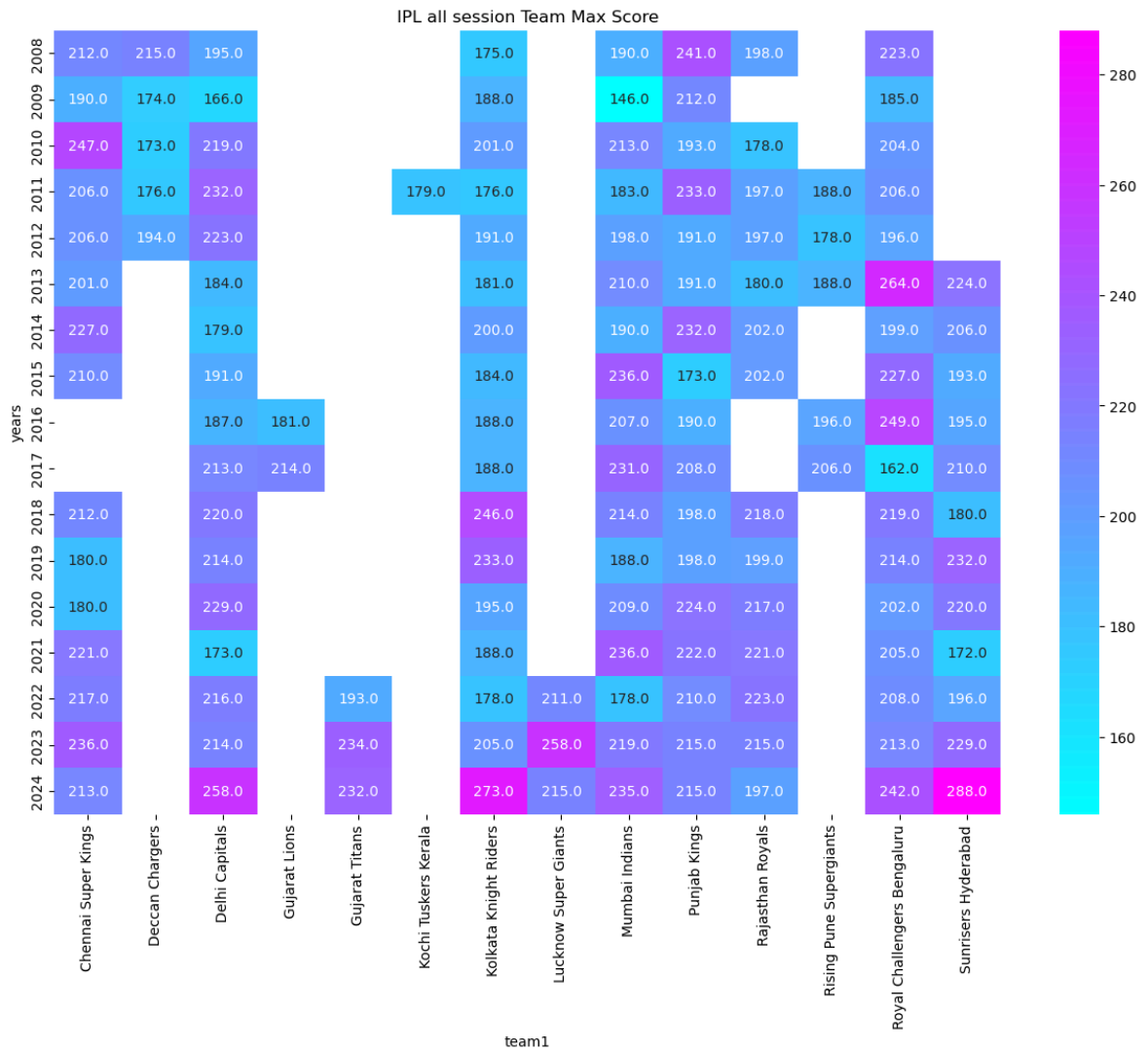
```
In [148... plt.figure(figsize=(15,10))
# pd.options.display.float_format = '{:.2f}'.format
# pd.set_option('display.float_format', '{:.2f}'.format)
sns.heatmap(all_team_year_df_2,fmt = "" ,annot = True)
plt.show()
```



```
In [155... all_team_year_df_min = df_matches.pivot_table(index='years', columns='
all_team_year_df_min = all_team_year_df_min.round(0)
plt.figure(figsize=(15,10))
# pd.options.display.float_format = '{:.2f}'.format
# pd.set_option('display.float_format', '{:.2f}'.format)
sns.heatmap(all_team_year_df_min,fmt = "" ,annot = True, cmap = 'mako')
plt.title('IPL all session Min Team Score')
plt.show()
```



```
In [159... all_team_year_df_max = df_matches.pivot_table(index='years', columns='
all_team_year_df_max = all_team_year_df_max.round(0)
plt.figure(figsize=(15,10))
# pd.options.display.float_format = '{:.2f}'.format
# pd.set_option('display.float_format', '{:.2f}'.format)
sns.heatmap(all_team_year_df_max,fmt = '' ,annot = True, cmap = 'cool')
plt.title('IPL all session Team Max Score')
plt.show()
```



In [157... `# Team2 all year analysis`

In []: `# LOC vs ILOC`
`# LOC: Location Name`
`# ILOC: Index Pass`

In [160... `df_matches.loc[50:55,['team1','team2','result']]`

Out[160...

	team1	team2	result
50	Kolkata Knight Riders	Punjab Kings	wickets
51	Rajasthan Royals	Mumbai Indians	wickets
52	Deccan Chargers	Chennai Super Kings	wickets
53	Royal Challengers Bengaluru	Mumbai Indians	wickets
54	Punjab Kings	Rajasthan Royals	runs
55	Delhi Capitals	Rajasthan Royals	runs

In [161... `df_matches.iloc[50:55,[7,8,12]]`

Out [161...

	team1	team2	result
50	Kolkata Knight Riders	Punjab Kings	wickets
51	Rajasthan Royals	Mumbai Indians	wickets
52	Deccan Chargers	Chennai Super Kings	wickets
53	Royal Challengers Bengaluru	Mumbai Indians	wickets
54	Punjab Kings	Rajasthan Royals	runs

In [164...

2016 > all rows and columns last 3

In [168...

```
rows = df_matches[df_matches['years']>=2016].index
df_matches.iloc[rows,-3:]
```

Out [168...

	super_over	method	years
517	N	Normal method	2016
518	N	Normal method	2016
519	N	Normal method	2016
520	N	Normal method	2016
521	N	Normal method	2016
...
1090	N	Normal method	2024
1091	N	Normal method	2024
1092	N	Normal method	2024
1093	N	Normal method	2024
1094	N	Normal method	2024

578 rows × 3 columns

In []:

In [170...

df_deliveries.sample()

Out [170...

	match_id	inning	batting_team	bowling_team	over	ball	batter	bowler
173626	1178412	2	Delhi Capitals	Kings XI Punjab	3	5	SS Iyer	GC Viljoen

In [173...

df_deliveries.iloc[123257,:]

```
Out[173...] match_id          980907
            inning           1
            batting_team      Royal Challengers Bangalore
            bowling_team      Sunrisers Hyderabad
            over              16
            ball              6
            batter             SR Watson
            bowler             KV Sharma
            non_striker        AB de Villiers
            batsman_runs       6
            extra_runs         0
            total_runs         6
            extras_type        NaN
            is_wicket          0
            player_dismissed   NaN
            dismissal_kind     NaN
            fielder            NaN
            Name: 123257, dtype: object
```

```
In [174...] df_deliveries.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260920 entries, 0 to 260919
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   match_id              260920 non-null  int64
1   inning                260920 non-null  int64
2   batting_team          260920 non-null  object
3   bowling_team          260920 non-null  object
4   over                  260920 non-null  int64
5   ball                  260920 non-null  int64
6   batter                260920 non-null  object
7   bowler                260920 non-null  object
8   non_striker           260920 non-null  object
9   batsman_runs          260920 non-null  int64
10  extra_runs            260920 non-null  int64
11  total_runs            260920 non-null  int64
12  extras_type           14125 non-null   object
13  is_wicket              260920 non-null  int64
14  player_dismissed      12950 non-null   object
15  dismissal_kind        12950 non-null   object
16  fielder               9354 non-null    object
dtypes: int64(8), object(9)
memory usage: 33.8+ MB
```

```
In [177...] df_deliveries.isna().sum().sort_values(ascending = False)
```

```
Out[177...] fielder          251566
dismissal_kind  247970
player_dismissed 247970
extras_type     246795
batsman_runs      0
is_wicket         0
total_runs        0
extra_runs        0
match_id          0
inning           0
bowler            0
batter           0
ball             0
over             0
bowling_team      0
batting_team      0
non_striker       0
dtype: int64
```

```
In [178...] df_deliveries.loc[:,['fielder','dismissal_kind','player_dismissed','extra
```

```
Out[178...]      fielder  dismissal_kind  player_dismissed  extras_type
```

0	NaN	NaN	NaN	legbyes
1	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	wides
3	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN
...
260915	NaN	NaN	NaN	NaN
260916	NaN	NaN	NaN	NaN
260917	NaN	NaN	NaN	NaN
260918	NaN	NaN	NaN	NaN
260919	NaN	NaN	NaN	NaN

260920 rows × 4 columns

```
In [179...] normal_cols = ['fielder','dismissal_kind','player_dismissed','extras_type
```

```
In [180...] for i in normal_cols:
    print(f'Analysis for {i} \n\n')
    display(df_deliveries[i].value_counts().head(10))
```

Analysis for fielder

```

fielder
MS Dhoni          220
KD Karthik        199
AB de Villiers    141
RV Uthappa        133
WP Saha           131
V Kohli           131
SK Raina          119
RA Jadeja         116
SV Samson         115
RG Sharma         111
Name: count, dtype: int64
Analysis for dismissal_kind

```

```

dismissal_kind
caught            8063
bowled           2212
run out          1114
lbw              800
caught and bowled 367
stumped          358
retired hurt      15
hit wicket        15
obstructing the field 3
retired out       3
Name: count, dtype: int64
Analysis for player_dismissed

```

```

player_dismissed
RG Sharma        223
V Kohli          207
S Dhawan         193
KD Karthik       184
RV Uthappa       180
SK Raina         171
DA Warner        164
AT Rayudu        155
AM Rahane        154
SV Samson        144
Name: count, dtype: int64
Analysis for extras_type

```

```

extras_type
wides           8380
legbyes         4001
noballs         1069
byes            673
penalty         2
Name: count, dtype: int64

```

```

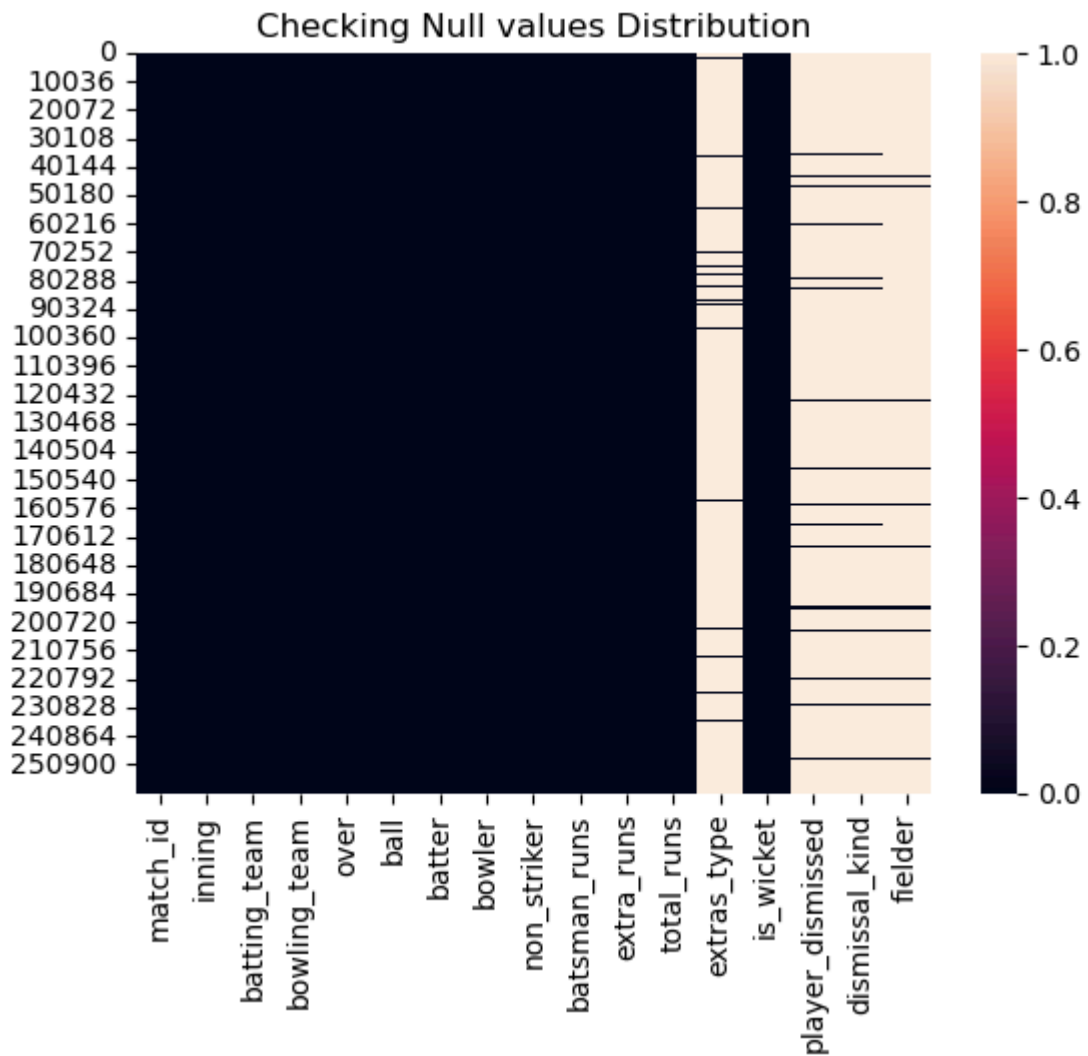
In [182... plt.title("Checking Null values Distribution")
sns.heatmap(df_deliveries.isna())

```

```

Out[182... <Axes: title={'center': 'Checking Null values Distribution'}>

```

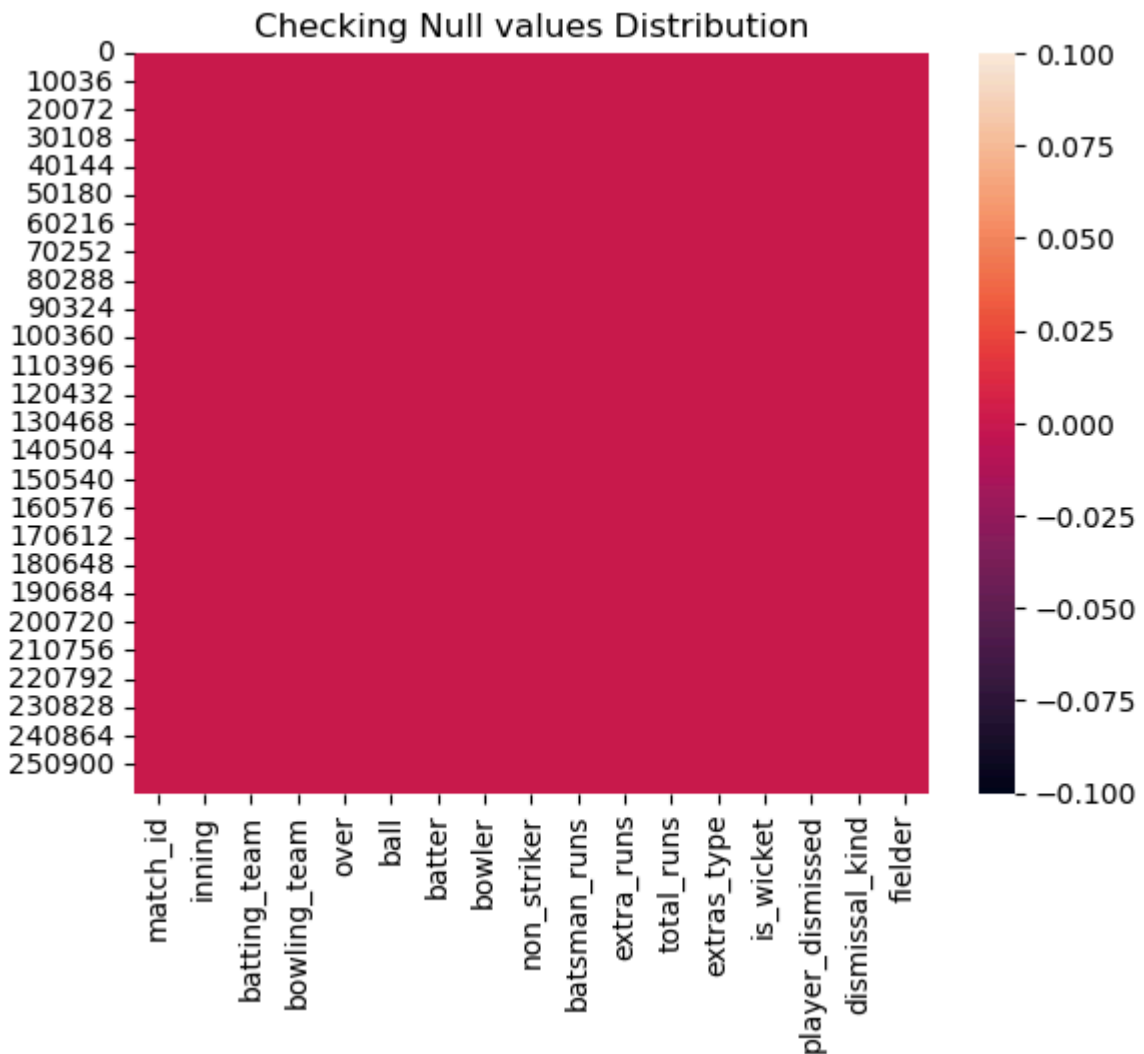



```
In [183... # Filll NAN Values

df_deliveries.fillna('Unknown',inplace = True)
print('Done')
```

Done

```
In [184... plt.title("Checking Null values Distribution")
sns.heatmap(df_deliveries.isna())
plt.show()
```



In [189... `df_deliveries.sample()`

Out[189...

	match_id	inning	batting_team	bowling_team	over	ball	batter	bowler
187127	1216525	1	Royal Challengers Bangalore	Chennai Super Kings	11	1	V Kohli	RA Jadeja

```
In [206... # batsman = 'DA Miller' #DA Miller

def get_player_performance_stats(batsman):

    player_df = df_deliveries[df_deliveries['batter'] == batsman]
    x = player_df['batsman_runs'].value_counts().sort_index().index
    y = player_df['batsman_runs'].value_counts().sort_index().values
    plt.figure(figsize=(4,2))
    plt.title(f'Analysis for {batsman}')
    ax = plt.bar(x,y,color = sns.color_palette('magma',7))
    plt.bar_label(ax)
    plt.show()

    total_runs = player_df['batsman_runs'].sum()

    print(f'{batsman} Total Runs in IPL History: {total_runs}')

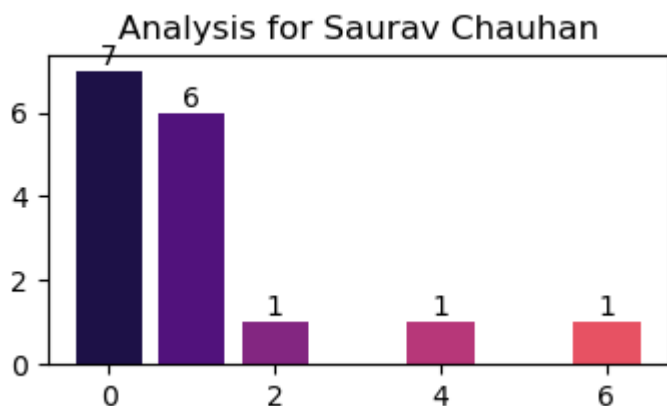
    print(f'Batsman {batsman} Top 10 Performance Against Bowlers')
```

```
performance_df = player_df.groupby('bowler')['batsman_runs'].agg(['sum'])
display(performance_df)
```

```
In [209... all_batsman = df_deliveries['batter'].unique()
print('Done')
```

Done

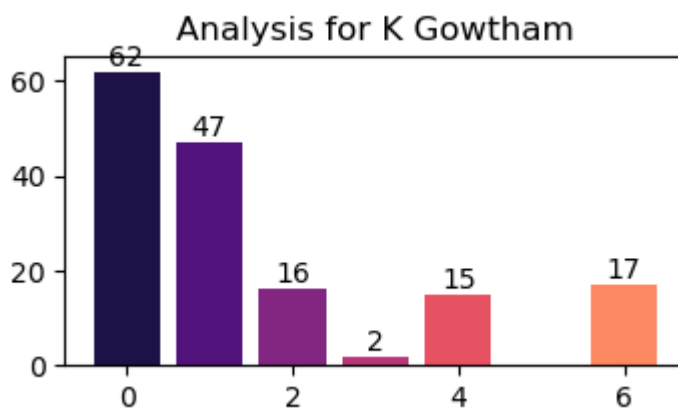
```
In [214... import random
for i in random.choices(all_batsman,k = 5):
    get_player_performance_stats(i)
```



Saurav Chauhan Total Runs in IPL History: 18

Batsman Saurav Chauhan Top 10 Performance Against Bowlers

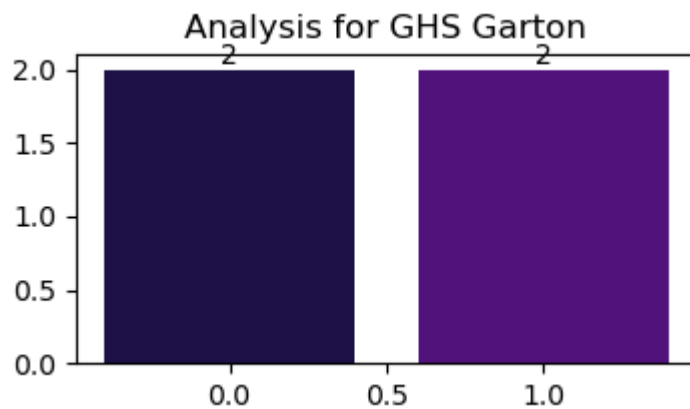
	bowler	sum	mean
5	R Ashwin	7	2.33
1	G Coetzee	6	1.20
2	JJ Bumrah	3	0.75
0	Avesh Khan	1	1.00
3	N Burger	1	1.00
4	PJ Cummins	0	0.00
6	YS Chahal	0	0.00



K Gowtham Total Runs in IPL History: 247

Batsman K Gowtham Top 10 Performance Against Bowlers

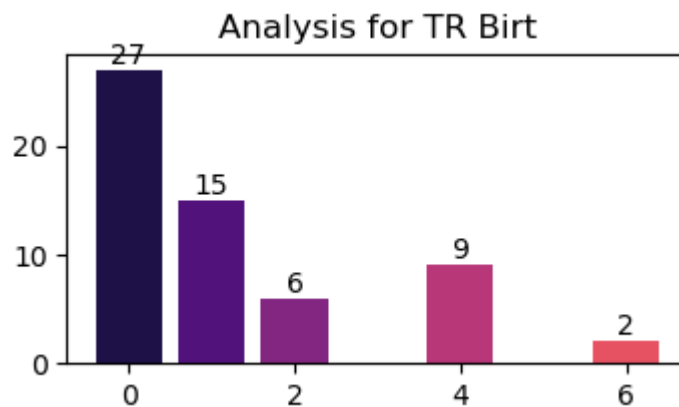
	bowler	sum	mean
38	TA Boult	22	1.69
23	MM Sharma	19	3.17
15	JJ Bumrah	17	1.70
39	TG Southee	14	2.80
11	DL Chahar	14	1.75
18	KV Sharma	13	2.17
2	AD Russell	13	2.17
27	Mustafizur Rahman	13	3.25
10	DJ Willey	13	2.60
4	AR Patel	10	0.91



GHS Garton Total Runs in IPL History: 2

Batsman GHS Garton Top 10 Performance Against Bowlers

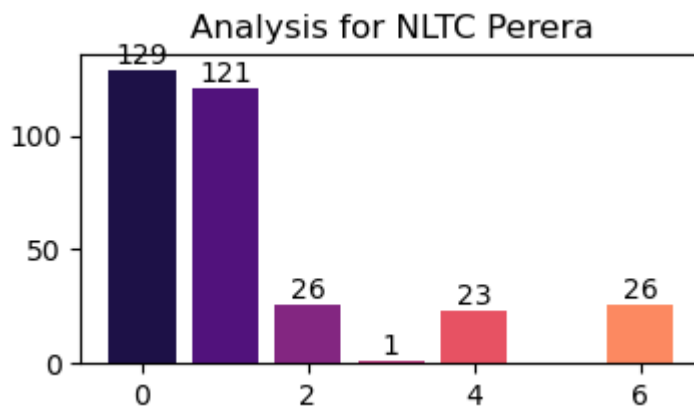
	bowler	sum	mean
0	B Kumar	1	0.50
1	JO Holder	1	1.00
2	Mohammed Shami	0	0.00



TR Birt Total Runs in IPL History: 75

Batsman TR Birt Top 10 Performance Against Bowlers

	bowler	sum	mean
7	R Vinay Kumar	26	2.36
5	P Parameswaran	9	2.25
3	Iqbal Abdulla	6	1.00
4	P Kumar	6	1.50
9	RJ Harris	6	2.00
11	S Sreesanth	6	1.50
6	PP Chawla	4	1.00
8	RA Jadeja	4	1.33
0	A Mishra	2	2.00
2	DT Christian	2	0.25



NLTC Perera Total Runs in IPL History: 424

Batsman NLTC Perera Top 10 Performance Against Bowlers

	bowler	sum	mean
56	SP Narine	32	1.33
16	HV Patel	20	3.33
6	Azhar Mahmood	20	2.50
43	P Kumar	19	2.11
36	MN Samuels	16	2.29
27	KJ Abbott	16	2.67
44	PP Chawla	15	1.25
26	JP Faulkner	15	1.36
22	JD Unadkat	15	1.36
28	KK Cooper	13	1.30

In []: `# Data Frame JOIN`

In []:

```
In [202... player_df.groupby('bowler')['batsman_runs'].agg(['sum', 'mean']).reset_ind
```

```
Out[202...
```

	bowler	sum	mean
129	RA Jadeja	111	1.66
122	R Ashwin	82	1.12
173	YS Chahal	77	1.71
1	A Mishra	68	1.31
45	HV Patel	63	1.50
14	Avesh Khan	61	1.52
118	PP Chawla	60	1.18
90	M Prasidh Krishna	59	1.69
165	UT Yadav	55	1.34
151	SP Narine	51	1.34

```
In [ ]:
```

```
In [ ]:
```