

IPL MATCH DATA ANALYSIS

importing all libraries:

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

Loading the dataset:

```
In [2]: df=pd.read_csv('match.csv')  
df
```

Out[2]:

	id	city	date	player_of_match	venue	neutral_venue
0	335982	Bangalore	2008-04-18	BB McCullum	Chinnaswamy Stadium	M 0 Cl I
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0 D I
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0
...
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0 Cl I
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0 Cl I
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0

816 rows × 17 columns

checking existing columns in dataframe:

In [3]: df.columns

```
Out[3]: Index(['id', 'city', 'date', 'player_of_match', 'venue', 'neutral_venue',
       'team1', 'team2', 'toss_winner', 'toss_decision', 'winner', 'result',
       'result_margin', 'eliminator', 'method', 'umpire1', 'umpire2'],
      dtype='object')
```

top 5 and bottom 5 data in dataset:

In [4]: `df.head()`

	id	city	date	player_of_match	venue	neutral_venue	team
0	335982	Bangalore	2008-04-18	BB McCullum	Chinnaswamy Stadium	M	0 Challe Bang
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Ki P
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Dare
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mu Ir
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Ki k I

In [5]: `df.tail()`

	id	city	date	player_of_match	venue	neutral_venue	team
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	Ro Challeng Bangal
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mum India
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	Ro Challeng Bangal
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	De Capit
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	De Capit

how many rows and columns are present in dataset:

```
In [6]: df.shape
```

```
Out[6]: (816, 17)
```

checking missing value and data type of columns in this dataset:

```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   id               816 non-null    int64  
 1   city              803 non-null    object  
 2   date              816 non-null    object  
 3   player_of_match   812 non-null    object  
 4   venue              816 non-null    object  
 5   neutral_venue     816 non-null    int64  
 6   team1             816 non-null    object  
 7   team2              816 non-null    object  
 8   toss_winner        816 non-null    object  
 9   toss_decision     816 non-null    object  
 10  winner            812 non-null    object  
 11  result            812 non-null    object  
 12  result_margin     799 non-null    float64 
 13  eliminator        812 non-null    object  
 14  method             19 non-null    object  
 15  umpire1           816 non-null    object  
 16  umpire2           816 non-null    object  
dtypes: float64(1), int64(2), object(14)
memory usage: 108.5+ KB
```

changing datatype of DATE column:

```
In [8]: df['date']=pd.to_datetime(df['date'])
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 17 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   id                816 non-null    int64  
 1   city               803 non-null    object  
 2   date               816 non-null    datetime64[ns] 
 3   player_of_match   812 non-null    object  
 4   venue               816 non-null    object  
 5   neutral_venue     816 non-null    int64  
 6   team1              816 non-null    object  
 7   team2              816 non-null    object  
 8   toss_winner         816 non-null    object  
 9   toss_decision      816 non-null    object  
 10  winner             812 non-null    object  
 11  result              812 non-null    object  
 12  result_margin      799 non-null    float64 
 13  eliminator          812 non-null    object  
 14  method              19 non-null    object  
 15  umpire1             816 non-null    object  
 16  umpire2             816 non-null    object  
dtypes: datetime64[ns](1), float64(1), int64(2), object(13)
memory usage: 108.5+ KB
```

checking null values in dataset:

```
In [9]: df.isnull().sum()
```

```
Out[9]: id                  0
city                 13
date                  0
player_of_match       4
venue                  0
neutral_venue         0
team1                  0
team2                  0
toss_winner            0
toss_decision          0
winner                 4
result                 4
result_margin          17
eliminator             4
method                 797
umpire1                 0
umpire2                 0
dtype: int64
```

deleting the 'method' column due to excess null values:

```
In [10]: df=df.drop('method',axis=1)
df.isnull().sum()
```

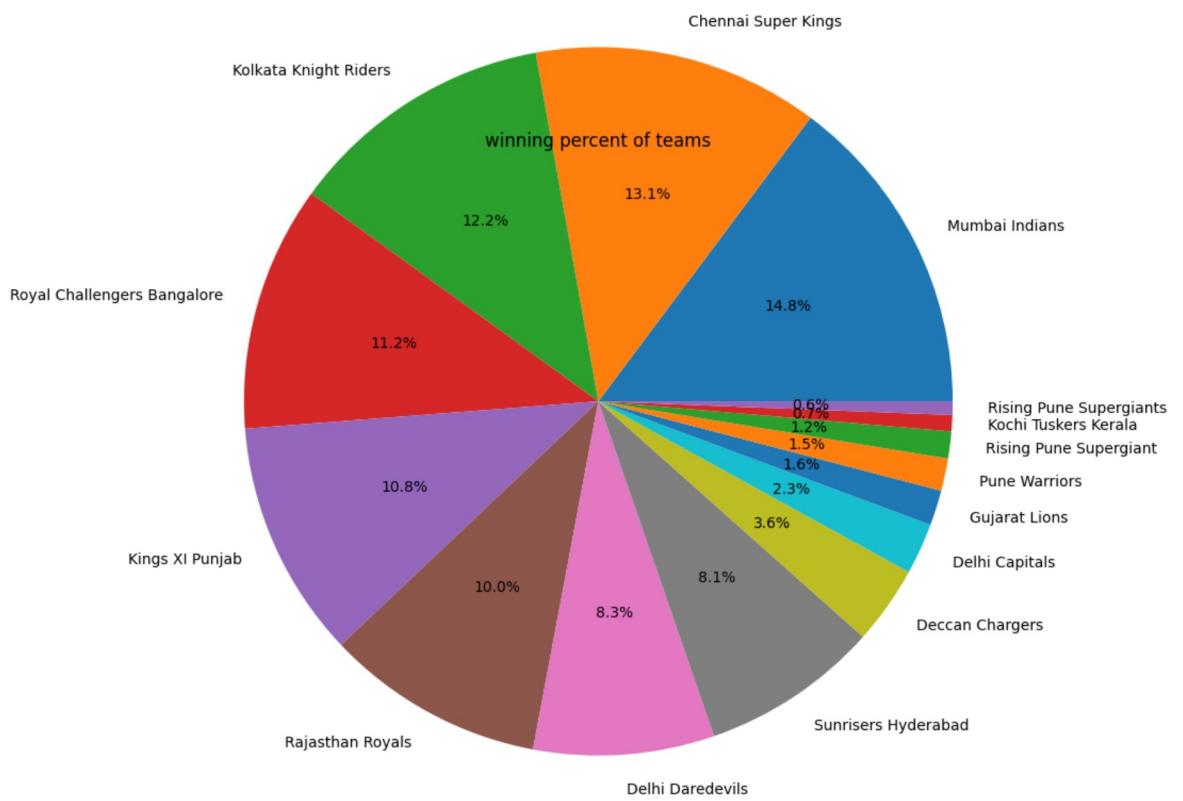
```
Out[10]: id          0  
city         13  
date          0  
player_of_match  4  
venue          0  
neutral_venue   0  
team1          0  
team2          0  
toss_winner     0  
toss_decision    0  
winner          4  
result          4  
result_margin    17  
eliminator      4  
umpire1          0  
umpire2          0  
dtype: int64
```

Data cleaning has been done now individual analysis is to be done:

```
In [11]: df['winner'].value_counts()
```

```
Out[11]: winner  
Mumbai Indians      120  
Chennai Super Kings 106  
Kolkata Knight Riders 99  
Royal Challengers Bangalore 91  
Kings XI Punjab     88  
Rajasthan Royals    81  
Delhi Daredevils    67  
Sunrisers Hyderabad  66  
Deccan Chargers      29  
Delhi Capitals       19  
Gujarat Lions        13  
Pune Warriors        12  
Rising Pune Supergiant 10  
Kochi Tuskers Kerala  6  
Rising Pune Supergiants 5  
Name: count, dtype: int64
```

```
In [12]: plt.figure(figsize=(6,6))  
plt.pie(df['winner'].value_counts(), labels=df['winner'].value_counts().keys(), ra  
plt.title('winning percent of teams')  
plt.show()
```



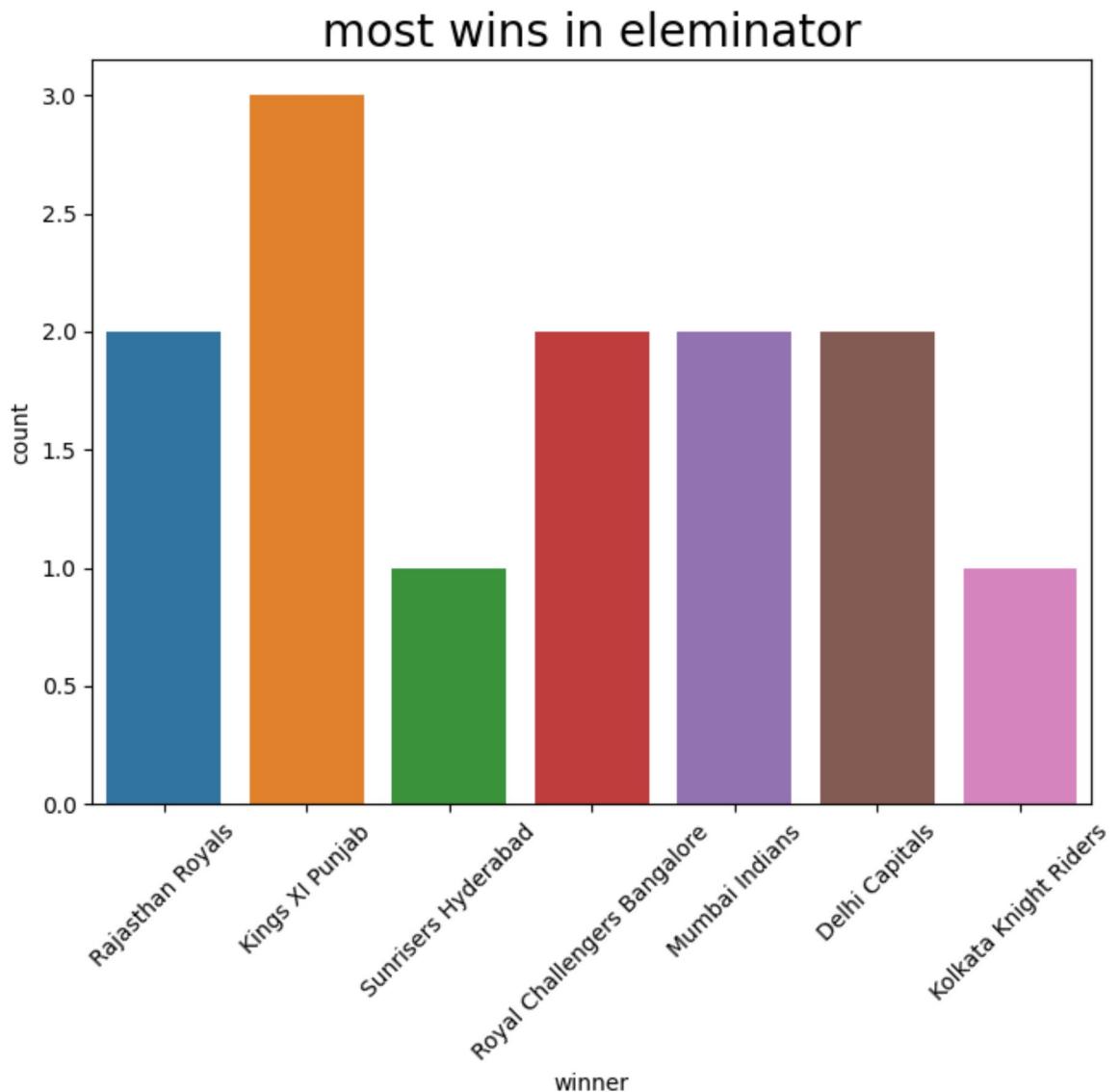
most wins in eliminator round:

```
In [13]: ydf=df.loc[df['eliminator']=="Y"]
ydf
```

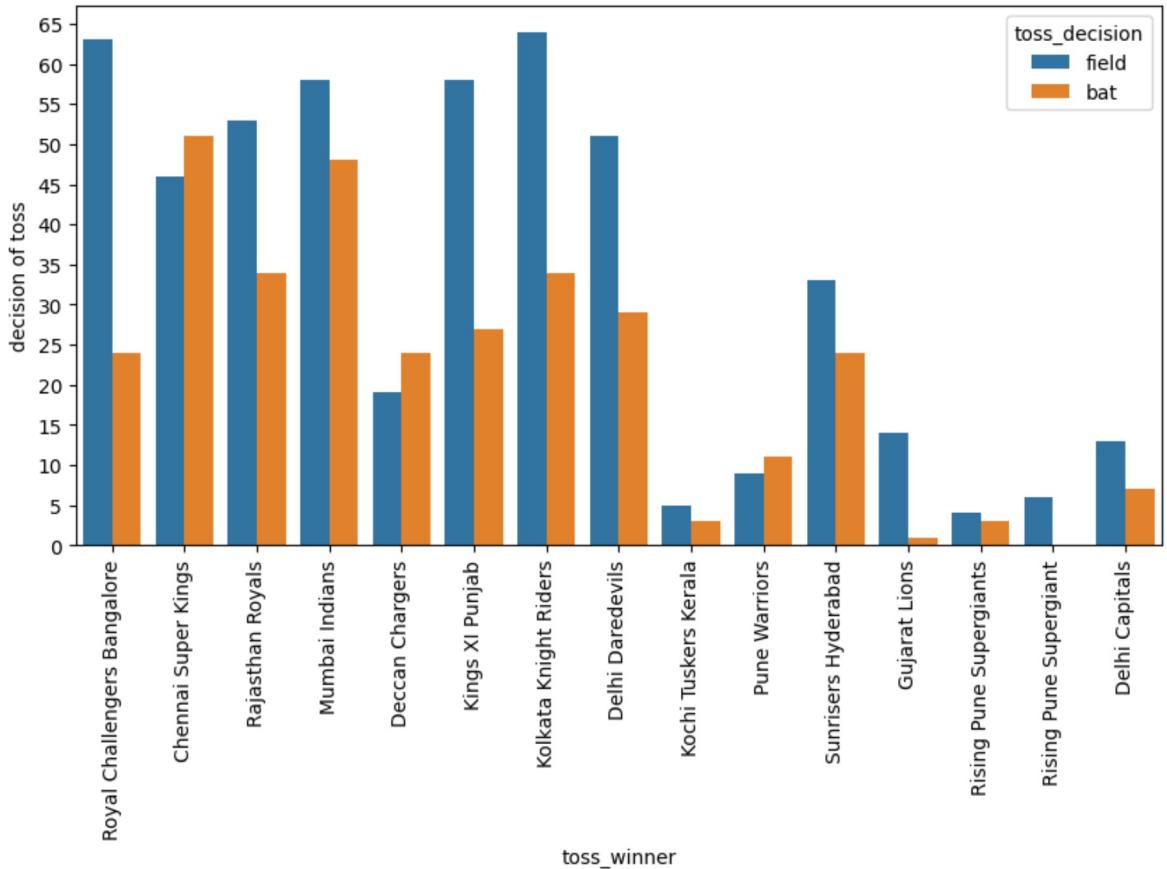
Out[13]:

	id	city	date	player_of_match	venue	neutral_venue
66	392190	Cape Town	2009-04-23	YK Pathan	Newlands	1
130	419121	Chennai	2010-03-21	J Theron	MA Chidambaram Stadium, Chepauk	0
328	598004	Hyderabad	2013-04-07	GH Vihari	Rajiv Gandhi International Stadium, Uppal	0 I
341	598017	Bangalore	2013-04-16	V Kohli	M Chinnaswamy Stadium	0 C
416	729315	Abu Dhabi	2014-04-29	JP Faulkner	Sheikh Zayed Stadium	1
476	829741	Ahmedabad	2015-04-21	SE Marsh	Sardar Patel Stadium, Motera	0
610	1082625	Rajkot	2017-04-29	KH Pandya	Saurashtra Cricket Association Stadium	0
705	1175365	Delhi	2019-03-30	PP Shaw	Feroz Shah Kotla	0
746	1178426	Mumbai	2019-05-02	JJ Bumrah	Wankhede Stadium	0
757	1216493	Dubai	2020-09-20	MP Stoinis	Dubai International Cricket Stadium	0
776	1216512	Abu Dhabi	2020-10-18	LH Ferguson	Sheikh Zayed Stadium	0
781	1216517	Dubai	2020-10-18	KL Rahul	Dubai International Cricket Stadium	0
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0 C

```
In [14]: plt.figure(figsize=(8,6))
plt.xticks(rotation=45)
sns.countplot(x=ydf[ 'winner' ],data=ydf)
plt.title('most wins in eliminator',fontsize=20)
plt.show()
```



```
In [15]: plt.figure(figsize=(10,5))
plt.yticks(np.arange(0,70,5))
plt.xticks(rotation=90)
sns.countplot(x='toss_winner',data=df,hue='toss_decision')
plt.ylabel('decision of toss')
plt.show()
```



```
In [16]: gdf=df.groupby(['toss_winner','toss_decision'])['winner'].count()
```

```
gdf
```

```
Out[16]:   toss_winner          toss_decision
Chennai Super Kings      bat             51
                      field            46
Deccan Chargers          bat             24
                      field            19
Delhi Capitals           bat              7
                      field            13
Delhi Daredevils         bat             28
                      field            51
Gujarat Lions            bat              1
                      field            14
Kings XI Punjab          bat             27
                      field            58
Kochi Tuskers Kerala     bat              3
                      field            5
Kolkata Knight Riders     bat             34
                      field            64
Mumbai Indians           bat             48
                      field            58
Pune Warriors            bat             11
                      field            9
Rajasthan Royals          bat            34
                      field            51
Rising Pune Supergiant    field            6
Rising Pune Supergiants   bat              3
                      field            4
Royal Challengers Bangalore bat             24
                      field            62
Sunrisers Hyderabad        bat             24
                      field            33
```

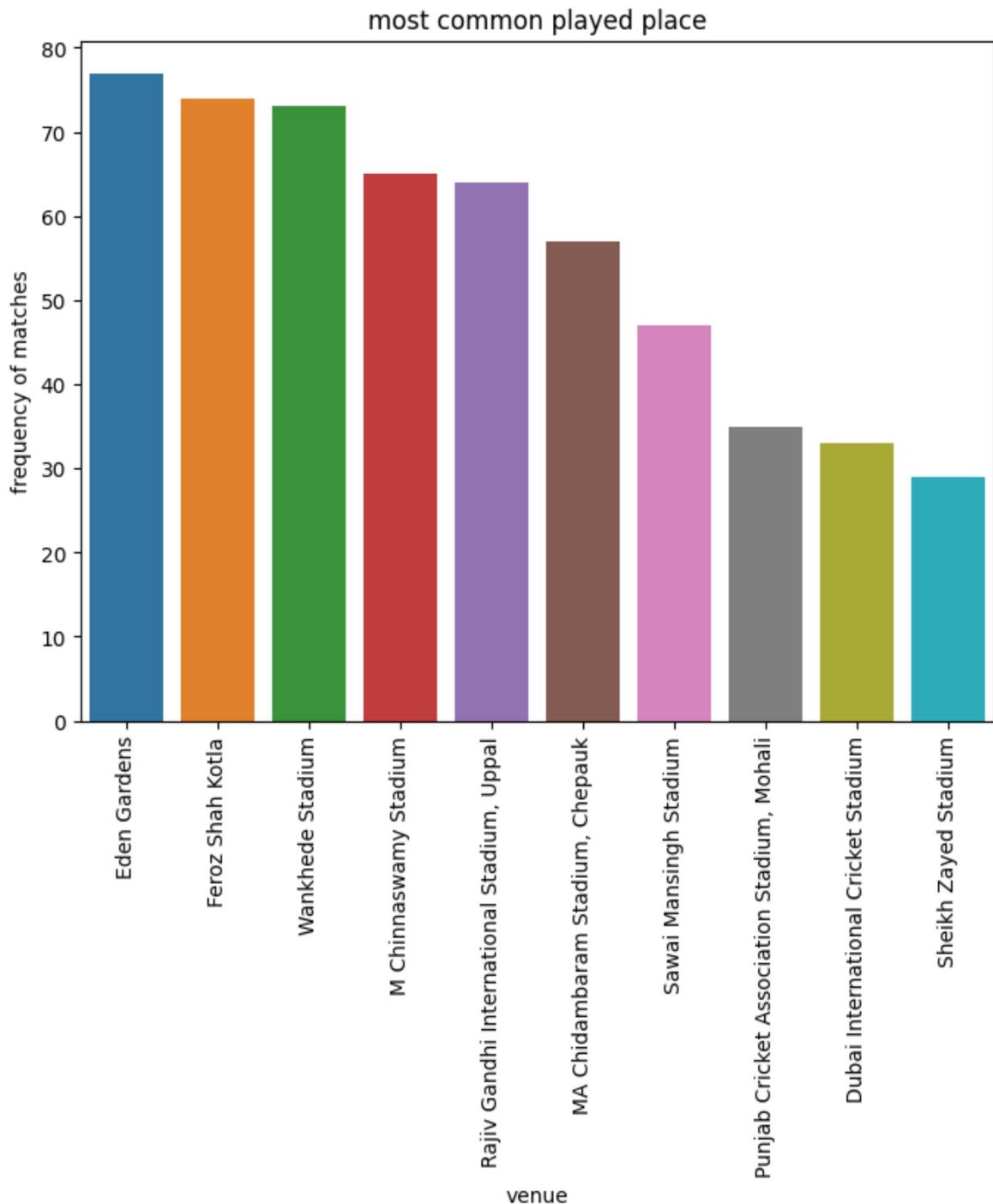
Name: winner, dtype: int64

mostly played matches at different venue:

```
In [17]: df['venue'].value_counts().head(10)
```

```
Out[17]:   venue
Eden Gardens                77
Feroz Shah Kotla            74
Wankhede Stadium             73
M Chinnaswamy Stadium       65
Rajiv Gandhi International Stadium, Uppal 64
MA Chidambaram Stadium, Chepauk 57
Sawai Mansingh Stadium      47
Punjab Cricket Association Stadium, Mohali 35
Dubai International Cricket Stadium 33
Sheikh Zayed Stadium         29
Name: count, dtype: int64
```

```
In [18]: plt.figure(figsize=(8,6))
plt.xticks(rotation=90)
sns.barplot(y=df['venue'].value_counts().head(10),x=df['venue'].value_counts().h
plt.ylabel('frequency of matches')
plt.title('most common played place')
plt.show()
```



filtering for only "csk" winning at different venue:

```
In [19]: cskw=df.loc[df['winner']=='Chennai Super Kings']
cskw.head()
```

Out[19]:

	id	city	date	player_of_match	venue	neutral_venue
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0
7	335989	Chennai	2008-04-23	ML Hayden	MA Chidambaram Stadium, Chepauk	0
11	335993	Chennai	2008-04-26	JDP Oram	MA Chidambaram Stadium, Chepauk	0
14	335996	Bangalore	2008-04-28	MS Dhoni	M Chinnaswamy Stadium	0 Chal Ba
27	336009	Delhi	2008-05-08	MS Dhoni	Feroz Shah Kotla	0 Dai

how many times csk wins at different venue:

In [20]:

```
winnings=cskw.groupby('venue').agg(
    total_count=('winner','count')
).reset_index().sort_values(by='total_count',ascending=False)
winnings.head()
```

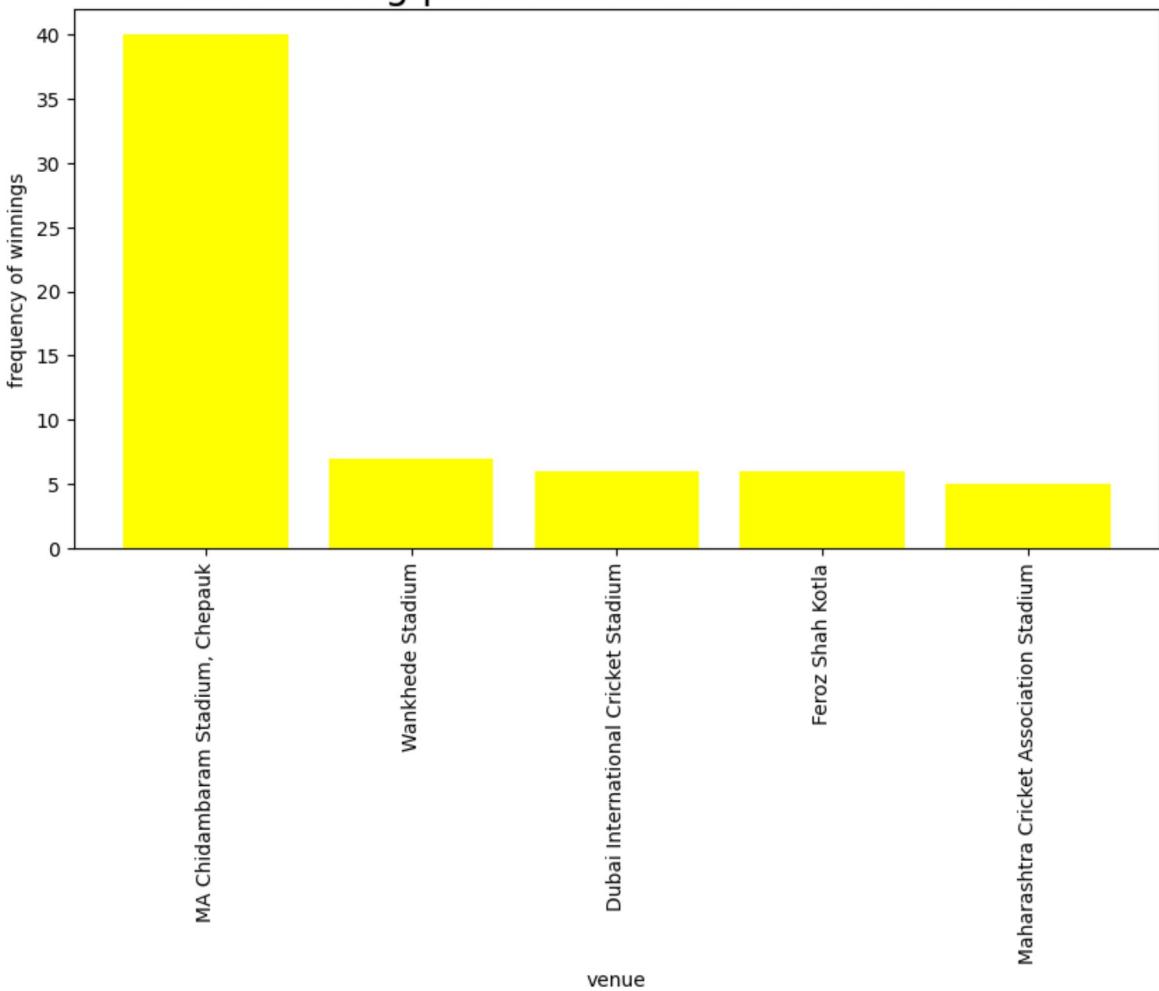
Out[20]:

	venue	total_count
13	MA Chidambaram Stadium, Chepauk	40
24	Wankhede Stadium	7
5	Dubai International Cricket Stadium	6
7	Feroz Shah Kotla	6
14	Maharashtra Cricket Association Stadium	5

In [21]:

```
plt.figure(figsize=(10,5))
plt.title('csk winning performance at different venues', fontsize=20)
plt.bar(winnings.head()['venue'],winnings.head()['total_count'],color='yellow')
plt.xlabel('venue')
plt.ylabel('frequency of winnings')
plt.xticks(rotation=90)
plt.show()
```

csk winning performance at different venues



filtering for only "mumbai indians" winning at different venue:

```
In [22]: miw=df.loc[df['winner']=='Mumbai Indians']
miw.head()
```

	id	city	date	player_of_match	venue	neutral_venue	t
15	335997	Kolkata	2008-04-29	ST Jayasuriya	Eden Gardens	0	Ko K F
22	336004	Mumbai	2008-05-04	SM Pollock	Dr DY Patil Sports Academy	0	Mu Ir
26	336008	Mumbai	2008-05-07	A Nehra	Dr DY Patil Sports Academy	0	Mu Ir
30	336012	Bangalore	2008-05-28	CRD Fernando	Chinnaswamy Stadium	0	Challe Bang
36	336018	Mumbai	2008-05-14	ST Jayasuriya	Wankhede Stadium	0	Mu Ir

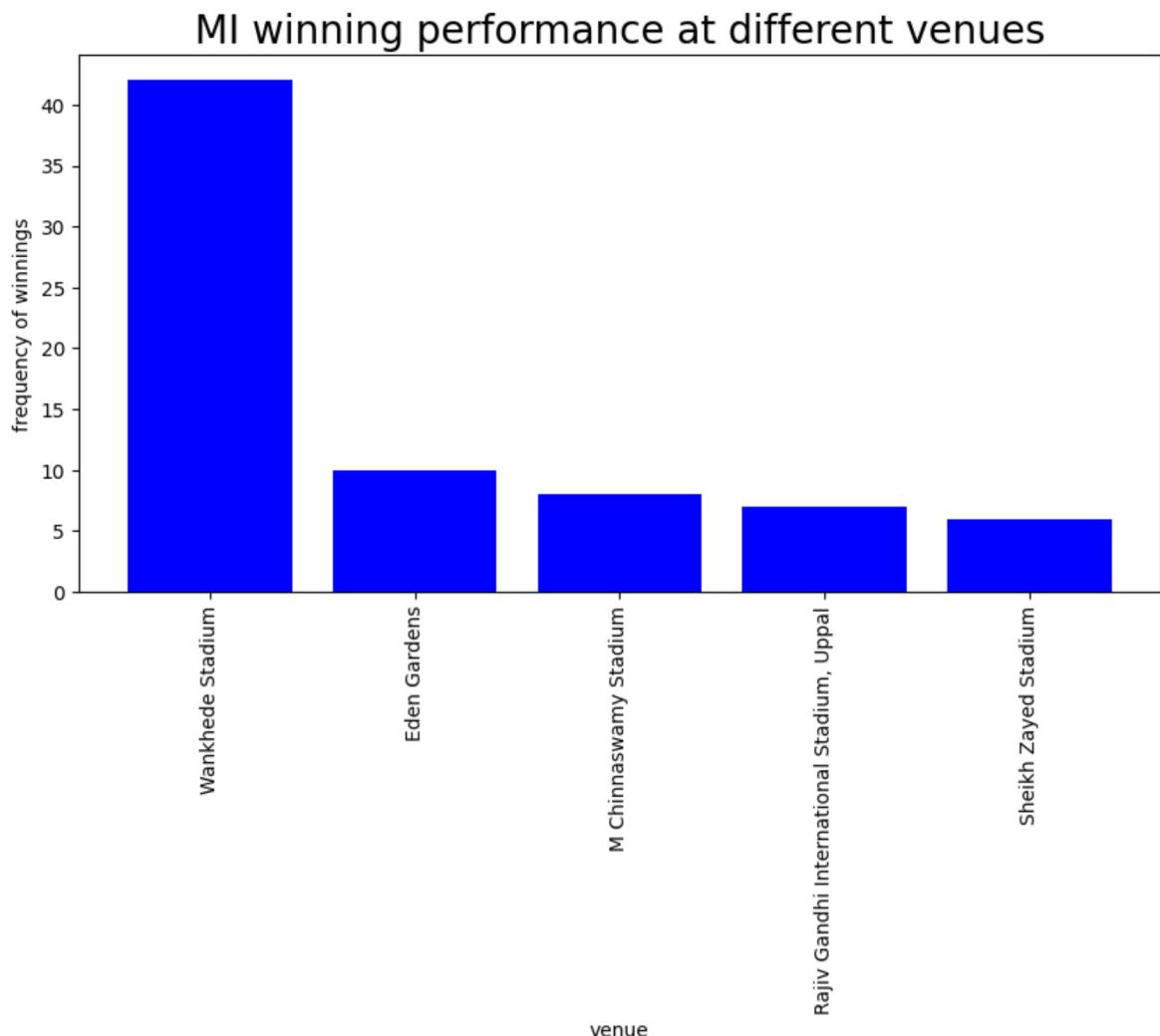
how many times "mumbai indians" wins at different venue:

```
In [23]: winnings=miw.groupby('venue').agg(  
    total_count=('winner','count')  
).reset_index().sort_values(by='total_count',ascending=False)  
winnings.head()
```

Out[23]:

	venue	total_count
24	Wankhede Stadium	42
5	Eden Gardens	10
8	M Chinnaswamy Stadium	8
15	Rajiv Gandhi International Stadium, Uppal	7
20	Sheikh Zayed Stadium	6

```
In [24]: plt.figure(figsize=(10,5))  
plt.title('MI winning performance at different venues', fontsize=20)  
plt.bar(winnings.head()['venue'],winnings.head()['total_count'],color='blue')  
plt.xlabel('venue')  
plt.ylabel('frequency of winnings')  
plt.xticks(rotation=90)  
plt.show()
```



some insights between top winning teams "csk and mi":

```
In [25]: cmdf=df.loc[((df['team1']=='Mumbai Indians') & (df['team2']=='Chennai Super King')) | ((df['team1']=='Chennai Super Kings')& (df['team2']=='Mumbai Indians'))]
cmdf.head()
```

Out[25]:

	id	city	date	player_of_match	venue	neutral_venue	tear
7	335989	Chennai	2008-04-23	ML Hayden	MA Chidamaram Stadium, Chepauk		0 Chen Sup Kir
36	336018	Mumbai	2008-05-14	ST Jayasuriya	Wankhede Stadium		0 Muml India
58	392181	Cape Town	2009-04-18	SR Tendulkar	Newlands		1 Chen Sup Kir
102	392227	Port Elizabeth	2009-05-16	ML Hayden	St George's Park		1 Chen Sup Kir
134	419125	Mumbai	2010-03-25	SR Tendulkar	Brabourne Stadium		0 Muml India

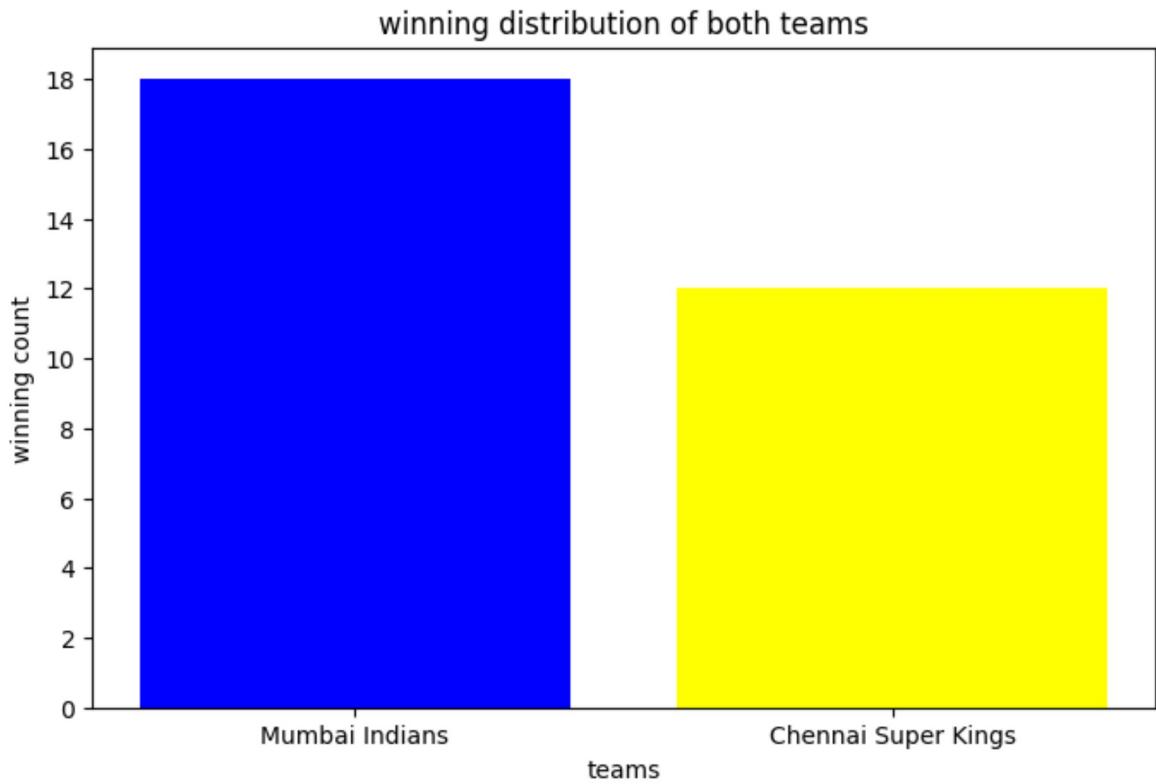
now which team wins more times 'mi' vs 'csk':

```
In [26]: cmdf['winner'].value_counts()
```

Out[26]:

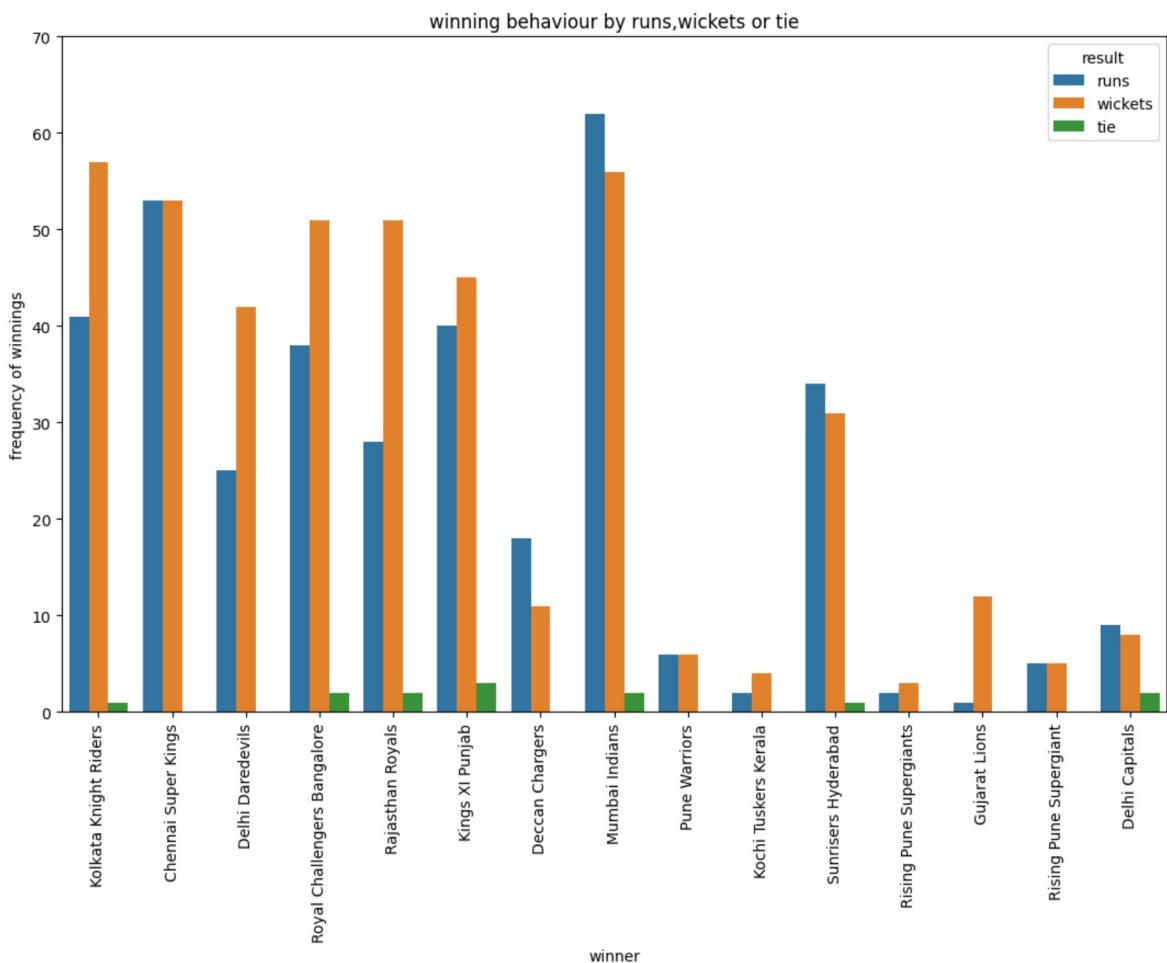
```
winner
Mumbai Indians      18
Chennai Super Kings 12
Name: count, dtype: int64
```

```
In [27]: plt.figure(figsize=(8,5))
plt.title('winning distribution of both teams')
plt.bar(cmdf['winner'].value_counts().keys(),cmdf['winner'].value_counts(),color=plt.cm.Paired.colors)
plt.xlabel('teams')
plt.ylabel('winning count')
plt.yticks(np.arange(0,19,2))
plt.show()
```



by which result, the team wins mostly:

```
In [28]: plt.figure(figsize=(13,8))
sns.countplot(x=df['winner'], data=df, hue='result')
plt.xticks(rotation=90)
plt.title('winning behaviour by runs,wickets or tie')
plt.ylabel('frequency of winnings')
plt.yticks(np.arange(0,80,10))
plt.show()
```



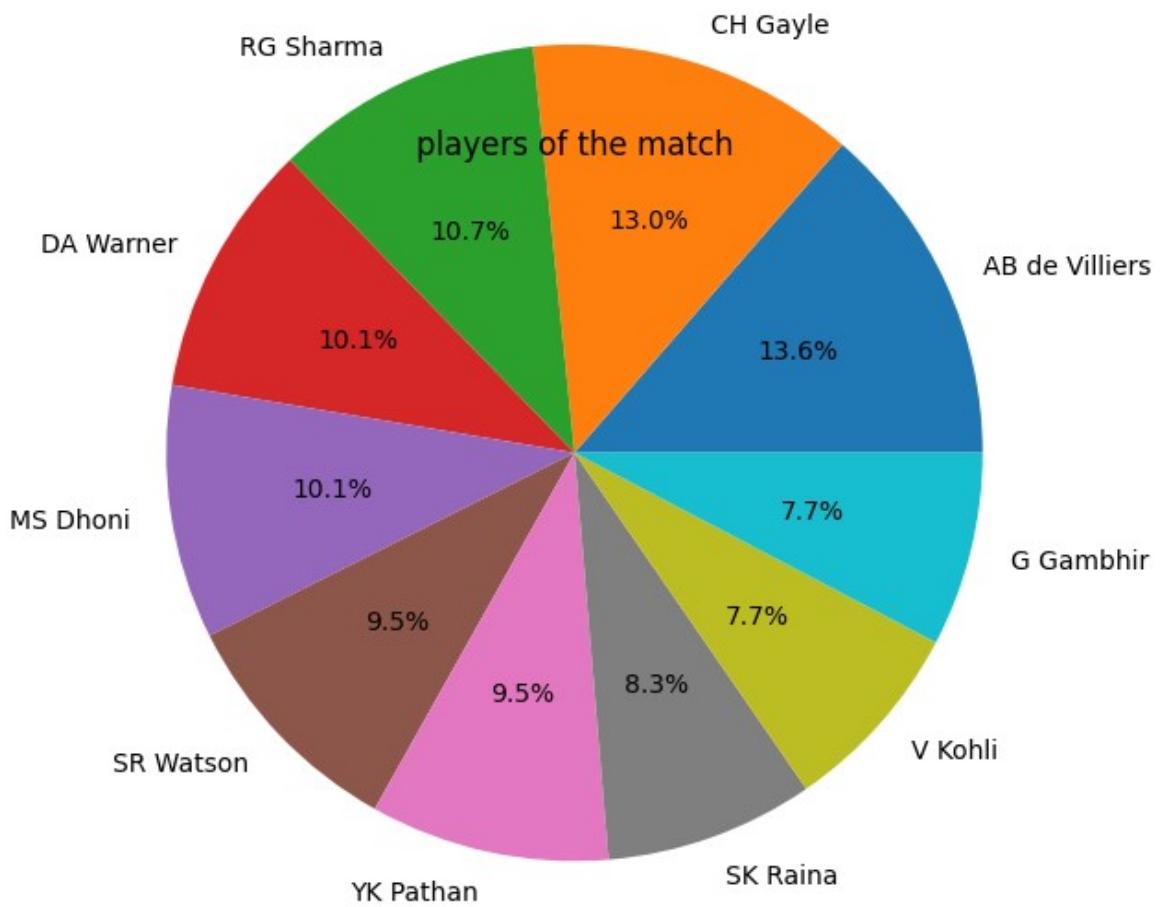
analysis on 'player of the match':

```
In [29]: df['player_of_match'].value_counts()
```

```
Out[29]: player_of_match
AB de Villiers    23
CH Gayle        22
RG Sharma       18
DA Warner       17
MS Dhoni        17
..
S Anirudha      1
M Kartik         1
R McLaren        1
RJ Harris        1
PJ Cummins      1
Name: count, Length: 233, dtype: int64
```

top 10 players who found the players of the match award:

```
In [30]: plt.figure(figsize=(6,4))
plt.pie(df['player_of_match'].value_counts().head(10),
        labels=df['player_of_match'].value_counts().head(10).keys(), autopct='%0.
plt.title('players of the match')
plt.show()
```



| THANKS |