

```
In [34]: import numpy as np
import pandas as pd
```

```
In [35]: data=pd.read_csv('matches.csv')
```

```
In [36]: type(data)
```

```
Out[36]: pandas.core.frame.DataFrame
```

```
In [37]: data.head(1)
```

```
Out[37]:
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_ap
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	

```
In [38]: data.tail(1)
```

```
Out[38]:
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal	

```
In [39]: data.shape
```

```
Out[39]: (636, 18)
```

```
In [40]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 636 entries, 0 to 635
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   id                    636 non-null   int64
1   season                636 non-null   int64
2   city                  629 non-null   object
3   date                  636 non-null   object
4   team1                 636 non-null   object
5   team2                 636 non-null   object
6   toss_winner           636 non-null   object
7   toss_decision         636 non-null   object
8   result                636 non-null   object
9   dl_applied            636 non-null   int64
10  winner                 633 non-null   object
11  win_by_runs           636 non-null   int64
12  win_by_wickets        636 non-null   int64
13  player_of_match       633 non-null   object
14  venue                  636 non-null   object
15  umpire1               635 non-null   object
16  umpire2               635 non-null   object
17  umpire3               0 non-null     float64
dtypes: float64(1), int64(5), object(12)
memory usage: 89.6+ KB
```

In [41]:

data.describe()

Out[41]:

	id	season	dl_applied	win_by_runs	win_by_wickets	umpire3
count	636.000000	636.000000	636.000000	636.000000	636.000000	0.0
mean	318.500000	2012.490566	0.025157	13.682390	3.372642	NaN
std	183.741666	2.773026	0.156726	23.908877	3.420338	NaN
min	1.000000	2008.000000	0.000000	0.000000	0.000000	NaN
25%	159.750000	2010.000000	0.000000	0.000000	0.000000	NaN
50%	318.500000	2012.000000	0.000000	0.000000	4.000000	NaN
75%	477.250000	2015.000000	0.000000	20.000000	7.000000	NaN
max	636.000000	2017.000000	1.000000	146.000000	10.000000	NaN

In [42]:

data.head()

Out[42]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_a
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	

## Series

In [43]: data['winner']

Out[43]:

0 Sunrisers Hyderabad  
1 Rising Pune Supergiant  
2 Kolkata Knight Riders  
3 Kings XI Punjab  
4 Royal Challengers Bangalore  
...  
631 Royal Challengers Bangalore  
632 Royal Challengers Bangalore  
633 Sunrisers Hyderabad  
634 Sunrisers Hyderabad  
635 Sunrisers Hyderabad  
Name: winner, Length: 636, dtype: object

In [44]: type(data['winner'])

Out[44]: pandas.core.series.Series

In [45]: data['winner'].shape

Out[45]: (636,)

## Data Frame

In [46]: data[['team1','team2','winner']]

Out[46]:

	team1	team2	winner
0	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad
1	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant
2	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders
3	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab
4	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore
...	...	...	...
631	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore
632	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore
633	Sunrisers Hyderabad	Kolkata Knight Riders	Sunrisers Hyderabad
634	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad
635	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad

636 rows × 3 columns

In [47]:

```
type(data[['team1','team2','winner']])
```

Out[47]: pandas.core.frame.DataFrame

In [48]:

```
data[['team1','team2','winner']].shape
```

Out[48]: (636, 3)

In [49]:

```
data.iloc[1:11:2]
```

Out[49]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_a
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	
7	8	2017	Indore	2017-04-10	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	normal	
9	10	2017	Mumbai	2017-04-12	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	

## Fancy Indexing

```
In [50]: data.iloc[[1,5,6]]
```

Out[50]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_appli
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant		field	normal
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad		field	normal
6	7	2017	Mumbai	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians		field	normal



```
In [51]: data.iloc[:,[4,5,10]]
```

Out[51]:

	team1	team2	winner
0	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad
1	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant
2	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders
3	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab
4	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore
...	...	...	...
631	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore
632	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore
633	Sunrisers Hyderabad	Kolkata Knight Riders	Sunrisers Hyderabad
634	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad
635	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad

636 rows × 3 columns

```
In [52]: data.head()
```

Out[52]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_a
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	

## Data Masling

In [53]:

mask=data['city']=='Hyderabad'  
data[mask].shape[0]

Out[53]:

49

In [79]:

def get\_city\_match\_count(city):  
 mask=data['city']==city  
 # return data[mask]  
 return data[mask].shape[0]

In [80]:

get\_city\_match\_count('Rajkot')

Out[80]:

10

In [82]:

mask1=data['city']=='Hyderabad'  
mask2=data['date']>'2010-01-01'  
  
data[mask1 & mask2].shape[0]

Out[82]:

42

## Value Count

In [84]:

data['winner'].value\_counts()

```
Out[84]: Mumbai Indians          92
          Chennai Super Kings    79
          Kolkata Knight Riders   77
          Royal Challengers Bangalore 73
          Kings XI Punjab         70
          Rajasthan Royals        63
          Delhi Daredevils        62
          Sunrisers Hyderabad     42
          Deccan Chargers         29
          Gujarat Lions          13
          Pune Warriors           12
          Rising Pune Supergiant   10
          Kochi Tuskers Kerala     6
          Rising Pune Supergiants  5
          Name: winner, dtype: int64
```

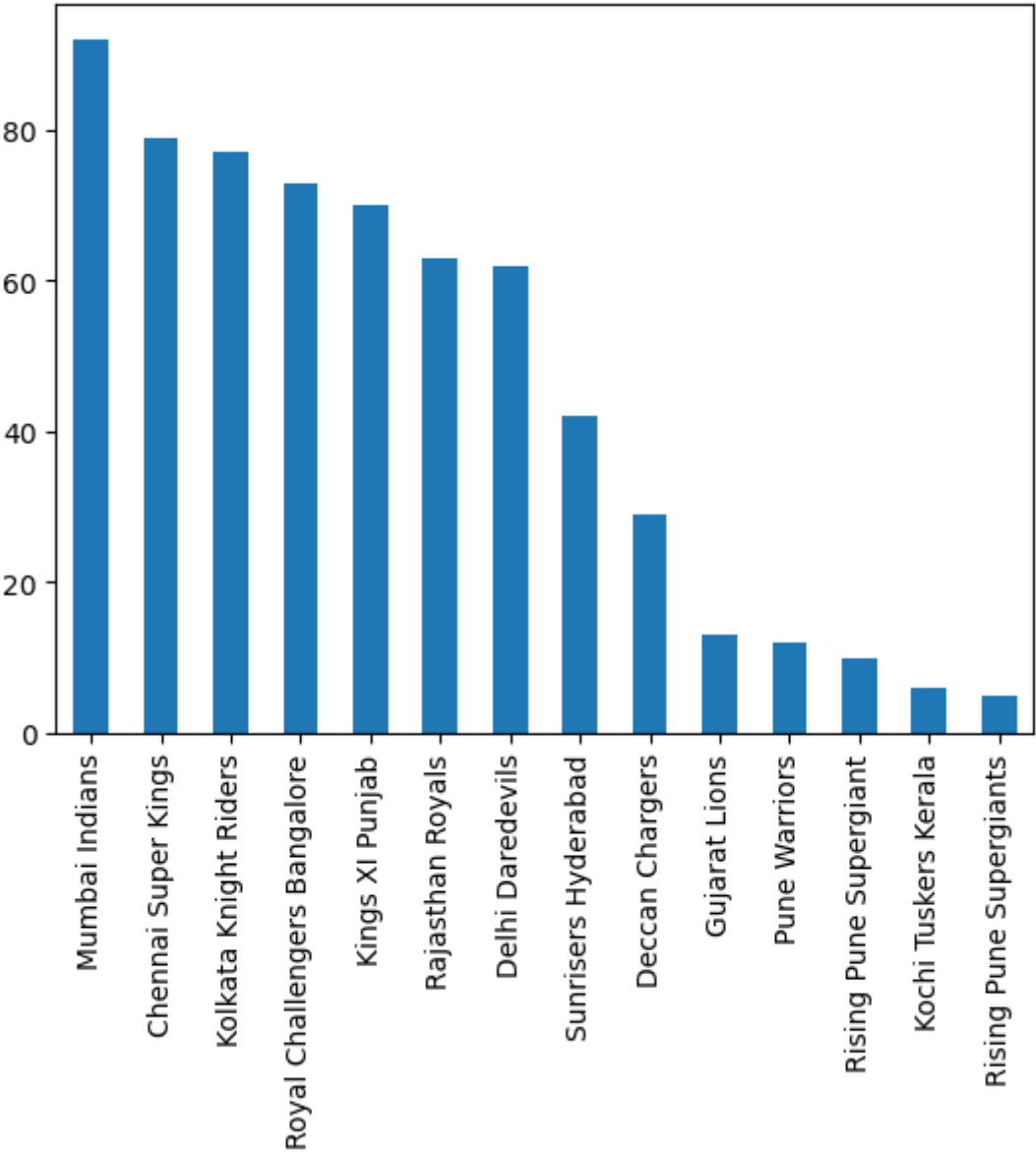
## The Pandas Plot()

### Bar Chart

```
In [87]: # The Pandas Plot()
import matplotlib.pyplot as plt
```

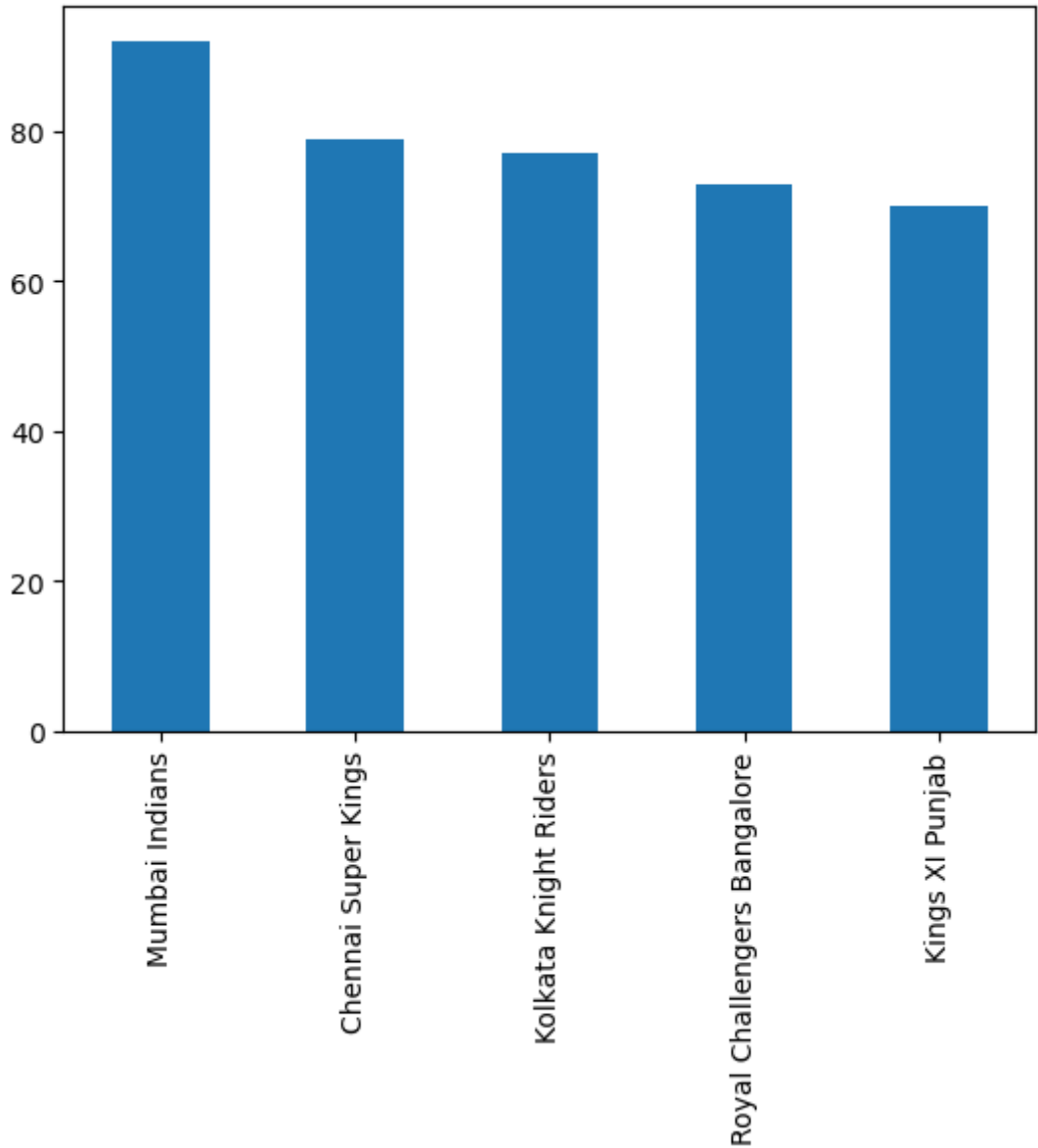
```
In [88]: data['winner'].value_counts().plot(kind='bar')
```

```
Out[88]: <Axes: >
```



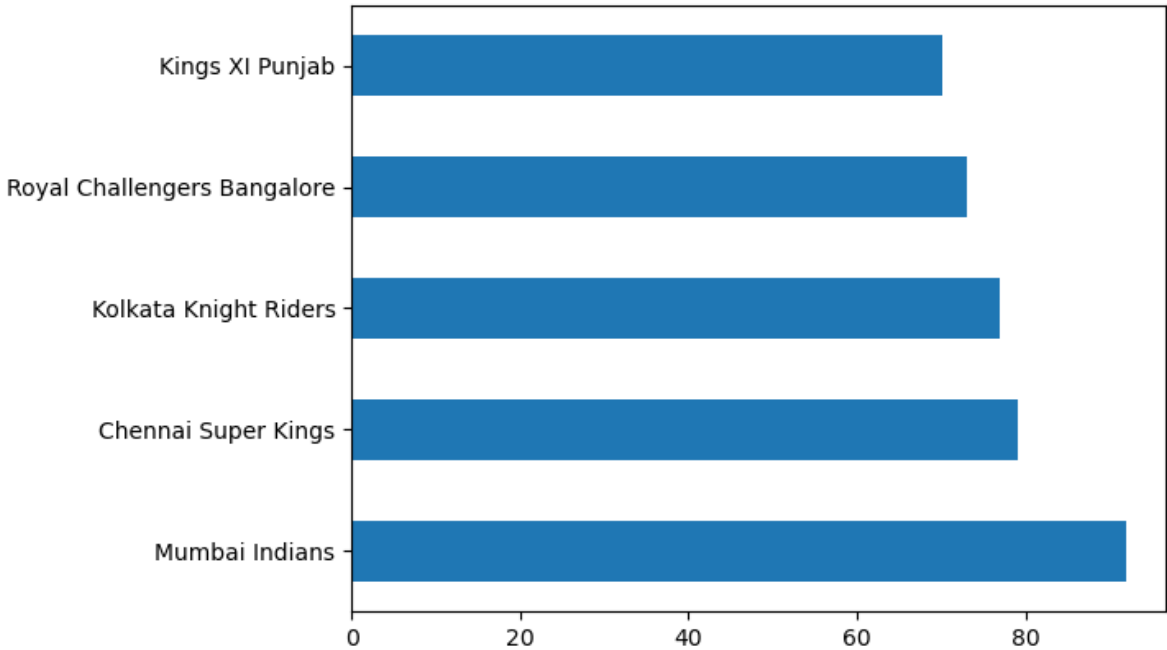
```
In [100]: data['winner'].value_counts().head().plot(kind='bar')
Out[100]: <Axes: >
```





```
In [102... data['winner'].value_counts().head().plot(kind='barh')
```

Out[102]: <Axes: >

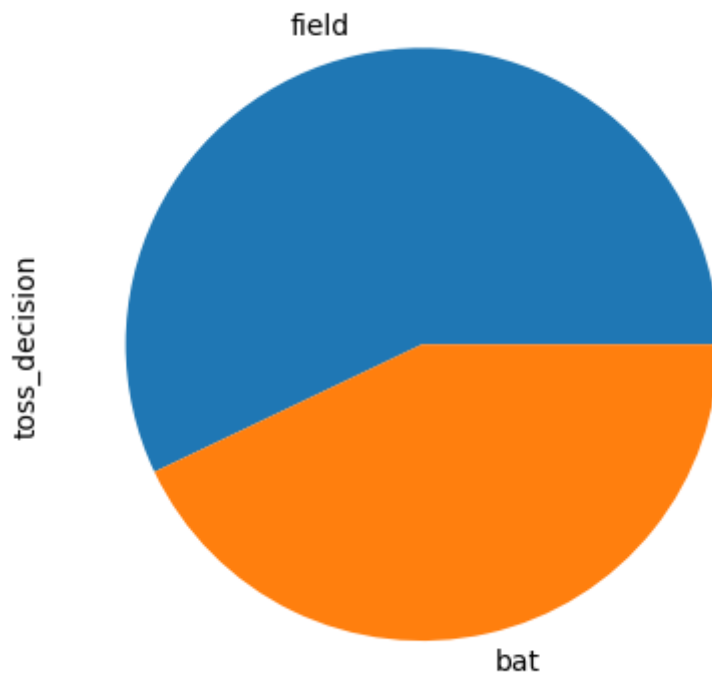


```
In [95]: data['toss_decision'].value_counts()
```

```
Out[95]: field    363  
bat        273  
Name: toss_decision, dtype: int64
```

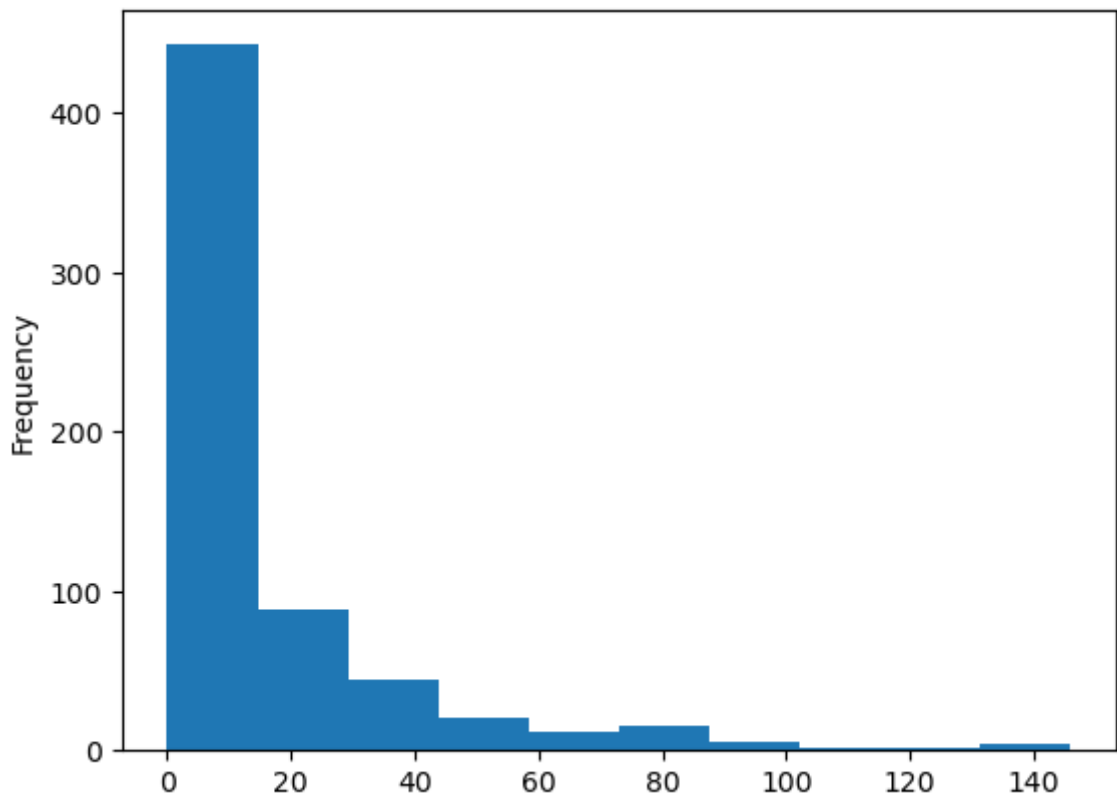
```
In [98]: data['toss_decision'].value_counts().plot(kind='pie')
```

```
Out[98]: <Axes: ylabel='toss_decision'>
```



```
In [104... data['win_by_runs'].plot(kind='hist')
```

```
Out[104]: <Axes: ylabel='Frequency'>
```



```
In [106... myseries=data['winner'].value_counts()
```

```
In [108... myseries.index
```

```
Out[108]: Index(['Mumbai Indians', 'Chennai Super Kings', 'Kolkata Knight Riders',
      'Royal Challengers Bangalore', 'Kings XI Punjab', 'Rajasthan Royals',
      'Delhi Daredevils', 'Sunrisers Hyderabad', 'Deccan Chargers',
      'Gujarat Lions', 'Pune Warriors', 'Rising Pune Supergiant',
      'Kochi Tuskers Kerala', 'Rising Pune Supergiants'],
      dtype='object')
```

```
In [110... type(data['winner'].value_counts())
```

```
Out[110]: pandas.core.series.Series
```

```
In [111... myseries.values
```

```
Out[111]: array([92, 79, 77, 73, 70, 63, 62, 42, 29, 13, 12, 10,  6,  5],
      dtype=int64)
```

```
In [113... myseries.head()
```

```
Out[113]: Mumbai Indians          92
      Chennai Super Kings       79
      Kolkata Knight Riders      77
      Royal Challengers Bangalore 73
      Kings XI Punjab           70
      Name: winner, dtype: int64
```

```
In [116... myseries['Kings XI Punjab']
```

```
Out[116]: 70
```

```
In [118... data['team2'].value_counts()
```

```
Out[118]: Delhi Daredevils      84
          Royal Challengers Bangalore  82
          Kolkata Knight Riders      80
          Mumbai Indians             75
          Kings XI Punjab            71
          Rajasthan Royals           64
          Chennai Super Kings        54
          Deccan Chargers            32
          Sunrisers Hyderabad        30
          Pune Warriors              26
          Gujarat Lions              16
          Rising Pune Supergiant      8
          Kochi Tuskers Kerala        7
          Rising Pune Supergiants     7
          Name: team2, dtype: int64
```

## Top five team who played highest match in IPL

```
In [119... data['team2'].value_counts() + data['team1'].value_counts()
```

```
Out[119]: Chennai Super Kings      131
          Deccan Chargers          75
          Delhi Daredevils        147
          Gujarat Lions           30
          Kings XI Punjab         148
          Kochi Tuskers Kerala     14
          Kolkata Knight Riders    148
          Mumbai Indians          157
          Pune Warriors           46
          Rajasthan Royals        118
          Rising Pune Supergiant   16
          Rising Pune Supergiants  14
          Royal Challengers Bangalore 152
          Sunrisers Hyderabad     76
          dtype: int64
```

## Sorted Value

```
In [125... (data['team2'].value_counts() + data['team1'].value_counts()).sort_values().tail()
```

```
Out[125]: Delhi Daredevils      147
          Kings XI Punjab      148
          Kolkata Knight Riders 148
          Royal Challengers Bangalore 152
          Mumbai Indians       157
          dtype: int64
```

```
In [126... (data['team2'].value_counts() + data['team1'].value_counts()).sort_values(ascending
```

```
Out[126]: Mumbai Indians           157
          Royal Challengers Bangalore 152
          Kings XI Punjab             148
          Kolkata Knight Riders       148
          Delhi Daredevils            147
          Chennai Super Kings         131
          Rajasthan Royals            118
          Sunrisers Hyderabad         76
          Deccan Chargers             75
          Pune Warriors               46
          Gujarat Lions               30
          Rising Pune Supergiant       16
          Kochi Tuskers Kerala        14
          Rising Pune Supergiants      14
          dtype: int64
```

```
In [129... data.sort_values('city',ascending=False)
```

Out[129]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	resu
312	313	2012	Visakhapatnam	2012-04-07	Chennai Super Kings	Deccan Chargers	Deccan Chargers	field	norm
622	623	2016	Visakhapatnam	2016-05-15	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	norm
615	616	2016	Visakhapatnam	2016-05-10	Sunrisers Hyderabad	Rising Pune Supergiants	Sunrisers Hyderabad	bat	norm
612	613	2016	Visakhapatnam	2016-05-08	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	norm
624	625	2016	Visakhapatnam	2016-05-17	Delhi Daredevils	Rising Pune Supergiants	Rising Pune Supergiants	field	norm
...	...	...	...	...	...	...	...	...	...
466	467	2014	NaN	2014-04-23	Chennai Super Kings	Rajasthan Royals	Rajasthan Royals	field	norm
468	469	2014	NaN	2014-04-25	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	norm
469	470	2014	NaN	2014-04-25	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	norm
474	475	2014	NaN	2014-04-28	Royal Challengers Bangalore	Kings XI Punjab	Kings XI Punjab	field	norm
476	477	2014	NaN	2014-04-30	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	norm

636 rows × 18 columns



## Permanently sorting in the actual datasheet

In [130...

data.sort\_values('city',ascending=False,inplace=True)

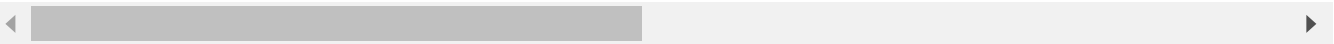
In [132...

data.sort\_values(['city','date'],ascending=[True,False])

Out[132]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_ap
475	476	2014	Abu Dhabi	2014-04-29	Rajasthan Royals	Kolkata Knight Riders	Rajasthan Royals	bat	tie	
471	472	2014	Abu Dhabi	2014-04-26	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
470	471	2014	Abu Dhabi	2014-04-26	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	normal	
464	465	2014	Abu Dhabi	2014-04-21	Chennai Super Kings	Delhi Daredevils	Chennai Super Kings	bat	normal	
460	461	2014	Abu Dhabi	2014-04-18	Sunrisers Hyderabad	Rajasthan Royals	Rajasthan Royals	field	normal	
...	...	...	...	...	...	...	...	...	...	...
468	469	2014	NaN	2014-04-25	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	
469	470	2014	NaN	2014-04-25	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	
466	467	2014	NaN	2014-04-23	Chennai Super Kings	Rajasthan Royals	Rajasthan Royals	field	normal	
461	462	2014	NaN	2014-04-19	Mumbai Indians	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
462	463	2014	NaN	2014-04-19	Kolkata Knight Riders	Delhi Daredevils	Kolkata Knight Riders	bat	normal	

636 rows × 18 columns



## Eliminate the duplicate data using drop\_duplicates() methods

In [133...

data.drop\_duplicates(subset=['city'])

Out[133]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	resu
--	----	--------	------	------	-------	-------	-------------	---------------	------

312	313	2012	Visakhapatnam	2012-04-07	Chennai Super Kings	Deccan Chargers	Deccan Chargers	field	norm
473	474	2014	Sharjah	2014-04-27	Sunrisers Hyderabad	Chennai Super Kings	Sunrisers Hyderabad	bat	norm
498	499	2014	Ranchi	2014-05-18	Chennai Super Kings	Royal Challengers Bangalore	Chennai Super Kings	bat	norm
19	20	2017	Rajkot	2017-04-18	Royal Challengers Bangalore	Gujarat Lions	Gujarat Lions	field	norm
560	561	2015	Raipur	2015-05-09	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	norm
531	532	2015	Pune	2015-04-18	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	norm
152	153	2009	Port Elizabeth	2009-05-10	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	norm
219	220	2010	Nagpur	2010-04-12	Deccan Chargers	Royal Challengers Bangalore	Royal Challengers Bangalore	field	norm
49	50	2017	Mumbai	2017-05-11	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	norm
212	213	2010	Kolkata	2010-04-07	Kolkata Knight Riders	Delhi Daredevils	Kolkata Knight Riders	bat	norm
236	237	2011	Kochi	2011-04-09	Kochi Tuskers Kerala	Royal Challengers Bangalore	Kochi Tuskers Kerala	bat	norm
154	155	2009	Kimberley	2009-05-11	Deccan Chargers	Rajasthan Royals	Deccan Chargers	bat	norm
51	52	2017	Kanpur	2017-05-13	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	norm
142	143	2009	Johannesburg	2009-05-03	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	norm
266	267	2011	Jaipur	2011-04-29	Mumbai Indians	Rajasthan Royals	Rajasthan Royals	field	norm



	id	season	city	date	team1	team2	toss_winner	toss_decision	result
293	294	2011	Indore	2011-05-15	Rajasthan Royals	Kochi Tuskers Kerala	Kochi Tuskers Kerala	field	normal
597	598	2016	Hyderabad	2016-04-26	Sunrisers Hyderabad	Rising Pune Supergiants	Rising Pune Supergiants	field	normal
137	138	2009	East London	2009-05-01	Mumbai Indians	Kolkata Knight Riders	Mumbai Indians	bat	normal
134	135	2009	Durban	2009-04-29	Kings XI Punjab	Mumbai Indians	Kings XI Punjab	bat	normal
292	293	2011	Dharamsala	2011-05-15	Kings XI Punjab	Delhi Daredevils	Delhi Daredevils	field	normal
199	200	2010	Delhi	2010-03-29	Delhi Daredevils	Kolkata Knight Riders	Delhi Daredevils	bat	normal
347	348	2012	Cuttack	2012-05-01	Deccan Chargers	Pune Warriors	Deccan Chargers	bat	normal
346	347	2012	Chennai	2012-04-30	Chennai Super Kings	Kolkata Knight Riders	Chennai Super Kings	bat	normal
578	579	2016	Chandigarh	2016-04-11	Kings XI Punjab	Gujarat Lions	Gujarat Lions	field	normal
155	156	2009	Centurion	2009-05-12	Kolkata Knight Riders	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
118	119	2009	Cape Town	2009-04-18	Royal Challengers Bangalore	Rajasthan Royals	Royal Challengers Bangalore	bat	normal
160	161	2009	Bloemfontein	2009-05-15	Delhi Daredevils	Kings XI Punjab	Kings XI Punjab	field	normal
57	58	2017	Bangalore	2017-05-19	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal
539	540	2015	Ahmedabad	2015-04-24	Rajasthan Royals	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal
457	458	2014	Abu Dhabi	2014-04-16	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	normal
461	462	2014	NaN	2014-04-19	Mumbai Indians	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal

id	season	city	date	team1	team2	toss_winner	toss_decision	resu
----	--------	------	------	-------	-------	-------------	---------------	------

```
In [134... data.drop_duplicates(subset=['city']).shape
```

Out[134]: (31, 18)

```
In [136... data.drop_duplicates(subset=['city','season']).shape
```

Out[136]: (107, 18)

```
In [141... data.drop_duplicates('season',keep='last')[['season','winner']].sort_values('season'
```

Out[141]:

	season	winner
89	2008	Mumbai Indians
164	2009	Delhi Daredevils
179	2010	Delhi Daredevils
301	2011	Royal Challengers Bangalore
316	2012	Kolkata Knight Riders
451	2013	Royal Challengers Bangalore
476	2014	Sunrisers Hyderabad
525	2015	Rajasthan Royals
635	2016	Sunrisers Hyderabad
44	2017	Kolkata Knight Riders

```
In [ ]:
```