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_ap
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
```

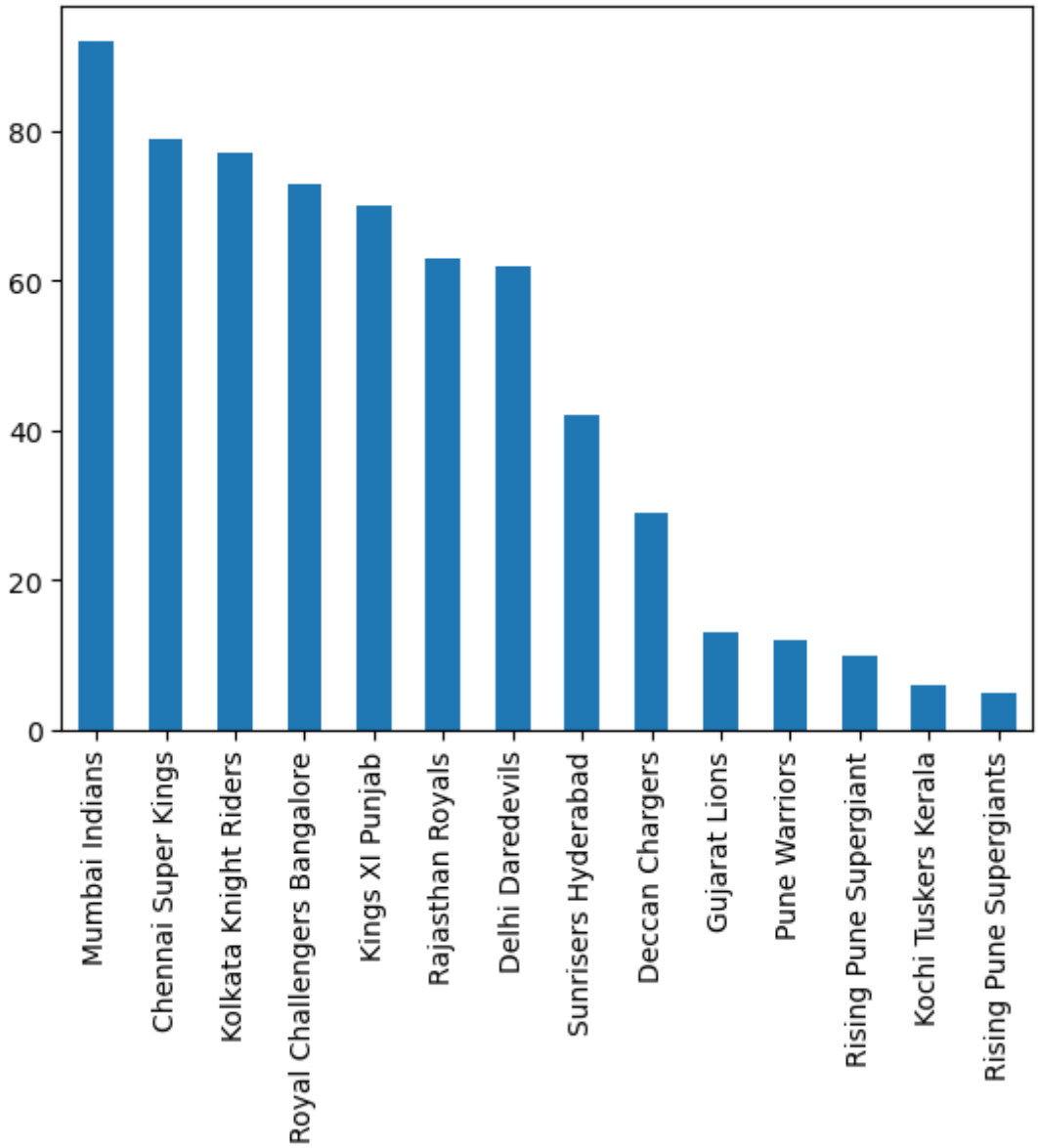
The Pandas Plot()

Bar Chart

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

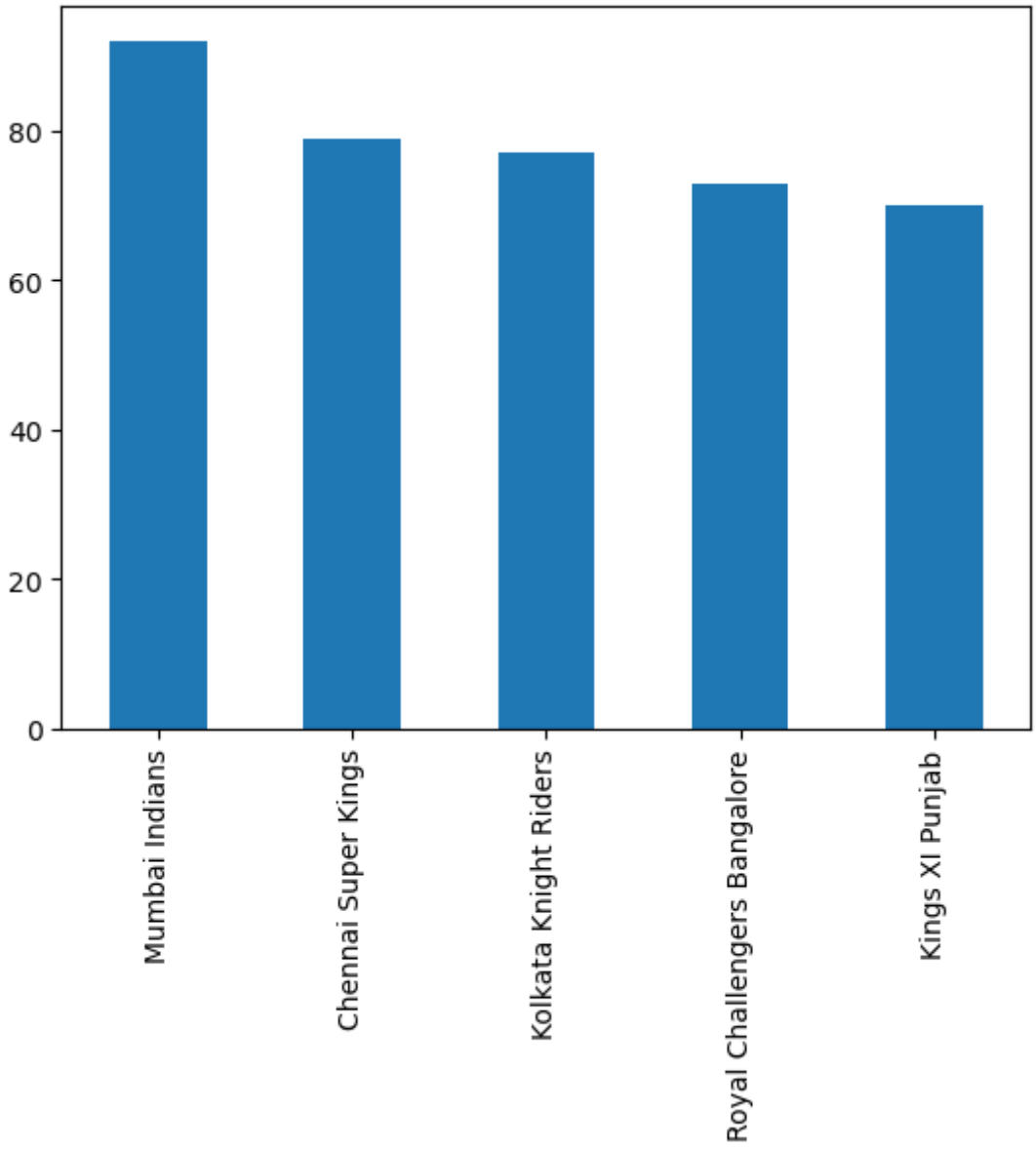
In [88]:
```
data['winner'].value_counts().plot(kind='bar')
```

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



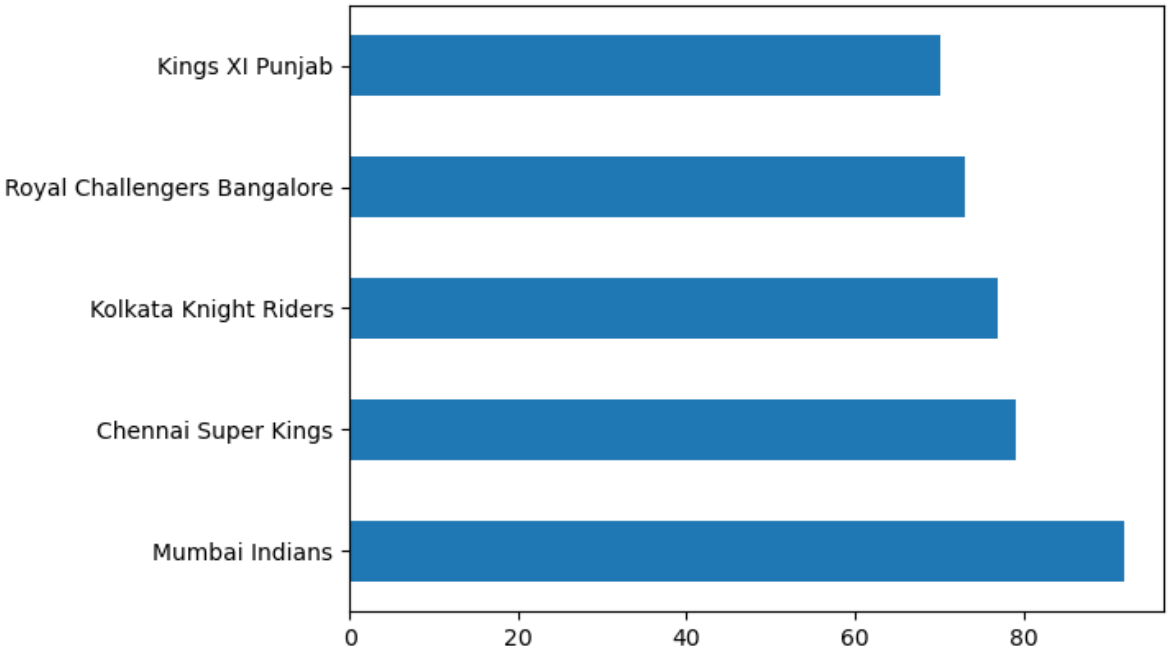
In [100…
```
data['winner'].value_counts().head().plot(kind='bar')
```

Out[100]:
```
<Axes: >
```



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

Out[102]:
```
<Axes: >
```

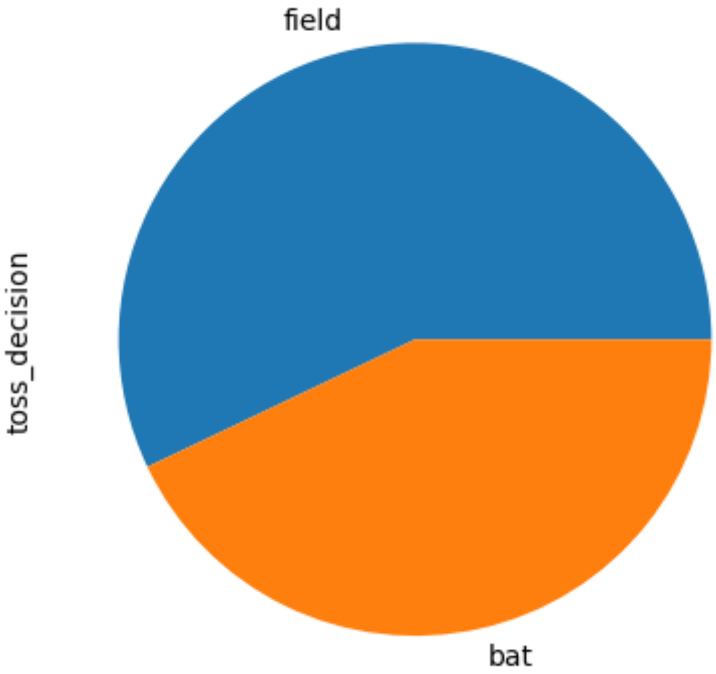


In [95]:
```
data['toss_decision'].value_counts()
```

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

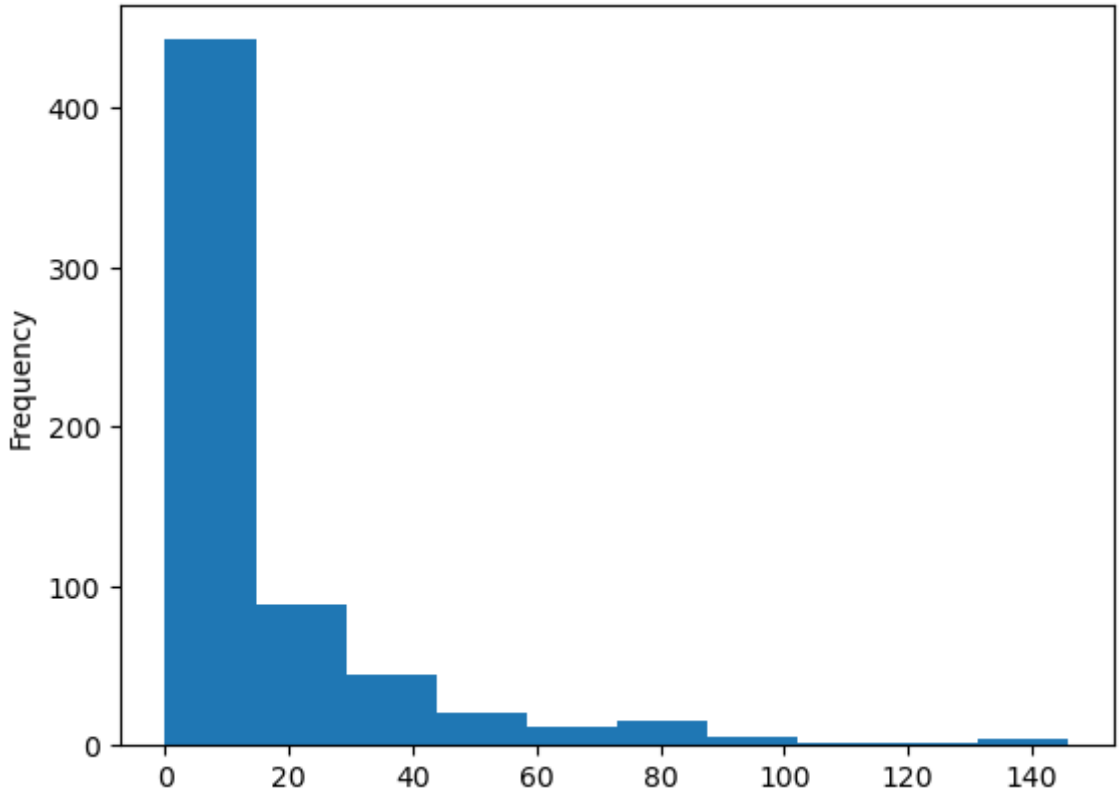
In [98]:
```
data['toss_decision'].value_counts().plot(kind='pie')
```

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



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

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



In [ ]: